



การระบุและการประกอบหน่วยการสร้างในชั้นตอนวิธีประมาณการแจกแจง



นายเฉลิมทรัพย์ สังขวิจิตร

ศูนย์วิทยทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

BUILDING BLOCKS IDENTIFICATION AND COMPOSITION IN ESTIMATION OF
DISTRIBUTION ALGORITHM



Mr. Chalermsub Sangkavichitr

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2009

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การระบุและการประกอบหน่วยการสร้างในขั้นตอนวิธีประมาณการแจกแจง
โดย	นายเฉลิมทรัพย์ สังขวิจิตร
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ประภาส จงสภิตย์วัฒนา
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร.วรเศรษฐ์ สุวรรณิก

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาตรีบัณฑิต

.....
..... คนบดีคณะคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศนัทธวงค์)

คณะกรรมการสอบวิทยานิพนธ์

.....
..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.สาธิต วงศ์ประทีป)

.....
..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ศาสตราจารย์ ดร.ประภาส จงสภิตย์วัฒนา)

.....
..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ผู้ช่วยศาสตราจารย์ ดร.วรเศรษฐ์ สุวรรณิก)

.....
..... กรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

.....
..... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.ณชล ไชยรัตน์)

ศูนย์วิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นายเฉลิมทรัพย์ สังขวิจิตร : การระบุและการประกอบหน่วยการสร้างในขั้นตอนวิธี
 ประมาณการแจกแจง. (BUILDING BLOCKS IDENTIFICATION AND
 COMPOSITION IN ESTIMATION OF DISTRIBUTION ALGORITHM). อ.ที่ปรึกษา
 วิทยานิพนธ์หลัก: ศ. ดร.ประภาส จงสถิตย์วัฒนา, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม:
 ผศ. ดร.วราเศรษฐ์ สุวรรณิก, 406 หน้า.

วิทยานิพนธ์นี้นำเสนอมุมมองหน่วยการสร้างแบบใหม่มีชื่อว่า "ชิ้นส่วน" ซึ่งเป็นรูปแบบ
 ของหน่วยการสร้างที่มีโครงสร้างง่ายและชัดเจน สามารถใช้ในการศึกษากลไกการทำงานของ
 ขั้นตอนวิธีเชิงพันธุกรรม เพื่อให้เข้าใจการทำงานของหน่วยการสร้างได้ชัดเจน งานวิจัยนี้แบ่งการ
 ทำงานของขั้นตอนวิธีเชิงพันธุกรรมออกเป็นสองส่วน ได้แก่ การระบุหน่วยการสร้าง และการ
 ประกอบหน่วยการสร้าง โดยพิจารณาจากมุมมองของการแบ่งปันความรู้เป็นหลัก ภายใต้แนวทาง
 นี้ได้นำเสนอขั้นตอนวิธีใหม่ 7 วิธี เพื่อแสดงให้เห็นถึงการจัดการชิ้นส่วนในมุมมองต่าง ๆ โดยมี
 การทดสอบกับปัญหาวัดเปรียบเทียบสมรรถนะจำนวน 17 ปัญหา ผลการทดลองแสดงให้เห็นว่า
 มุมมองหน่วยการสร้างแบบชิ้นส่วนช่วยให้เข้าใจการทำงานของขั้นตอนวิธีเชิงพันธุกรรมในมุมมอง
 ของการจัดการหน่วยการสร้างได้ชัดเจนขึ้น สามารถนำไปประยุกต์ใช้ในทางปฏิบัติได้จริง และ
 ความรู้ที่ได้นำไปใช้ในการพัฒนาขั้นตอนวิธีที่ใช้แก้ปัญหาให้มีประสิทธิภาพเพิ่มขึ้นได้

ศูนย์วิทยทรัพยากร

ภาควิชา วิศวกรรมคอมพิวเตอร์
 สาขาวิชา วิศวกรรมคอมพิวเตอร์
 ปีการศึกษา 2552

ลายมือชื่อนิติ์ เฉลิมทรัพย์ สังขวิจิตร
 ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก ป.ร.ศ.ภาส
 ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์ร่วม วราเศรษฐ์ สุวรรณิก

4771810421 : MAJOR COMPUTER ENGINEERING

KEYWORDS : BUILDING BLOCKS / FRAGMENT / DATA COMPRESSION /
ESTIMATION OF DISTRIBUTION ALGORITHM

CHALERMSUB SANGKAVICHITR : BUILDING BLOCKS IDENTIFICATION
AND COMPOSITION IN ESTIMATION OF DISTRIBUTION ALGORITHM.

THESIS ADVISOR : PROF. PRABHAS CHONGSTITVATANA, Ph.D., THESIS

CO-ADVISOR : ASSIST. PROF. WORASAIT SUWANNIK, Ph.D., 406 pp.

This thesis proposed a novel way to look at Building Blocks (BB) called "Fragments". Fragments have simple and explicit structures. They can be used to study the working mechanism of Genetic Algorithms (GAs) to understand how Building Blocks actually work. The process of GA is divided into two parts: BB identification and BB composition from the point of view of knowledge sharing. Based on this concept, seven new algorithms are proposed to show different ways to manipulate Fragments. The proposed algorithms are tested with seventeen benchmark problems and the results show that Fragments can help to understand Genetic Algorithms in the perspective of Building Blocks processing. The algorithms based on Fragments can be applied in practice and this knowledge leads to development of better algorithms.

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

Department : ..Computer Engineering.....	Student's Signature .. <i>Chalermsub Sangkavichitr</i>
Field of Study : ..Computer Engineering.....	Advisor's Signature .. <i>Prabhas Chongstitvatana</i>
Academic Year : ..2009.....	Co-Advisor's Signature .. <i>Worasait Suwannik</i>

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาเป็นอย่างยิ่งจากอาจารย์ที่ปรึกษา ศ. ดร.ประภาส จงสถิตยวัฒนา ที่ได้ให้แรงจูงใจ กำลังใจ วิเคราะห์ และคำแนะนำเมื่อยามที่ผู้วิจัยมี ปัญหาหรือเกิดความย่อท้อต่ออุปสรรคต่าง ๆ จนสามารถผ่านพ้นไปได้ในที่สุด

ขอขอบคุณที่ปรึกษาร่วม ผศ. ดร.วราเศรษฐ์ สุวรรณิก ที่เป็นแรงบันดาลใจให้เกิ ดงานวิจัยนี้ และกรรมสอภวิทยานิพนธ์ทุกท่าน ได้แก่ รศ. ดร.สาธิต วงศ์ประทีป ศ. ดร.บุญเสริม กิจศิริกุล และ รศ. ดร.ณชล ไชยรัตน์ะ ในการตรวจแก้และให้คำแนะนำอันเป็นประโยชน์ยิ่งต่อ งานวิจัย

วิทยานิพนธ์นี้สำเร็จได้จากความช่วยเหลือของบุคคลหลายท่านที่คอยให้ความเมตตา ช่วยเหลือ ให้คำแนะนำ ข้อคิดเห็นต่างๆ และให้กำลังใจด้วยดีตลอดมา ดังนี้ ชัยวัฒน์ เจษฎา ปกรณ์ (พีนิ่ง) (เป็นที่ปรึกษาในทุ กเรื่องที่เข้าใจได้ยาก) และ สุนิสา ริมเจริญ (ปาล์ม) (สำหรับ คำปรึกษาในทุ ก ๆ เรื่อง) จิระเดช พลสวัสดิ์ (สำหรับความช่วยเหลือในหลายด้าน) สมาชิก หอปฏิบัติกร ISL ทุกคน ได้แก่ พีโปง พีชัช พียอด พีเดี่ยว พีซ่าง พีสุพจน์ พียุทธนา ย้ง แด่ เพชร เดย์ ก๊ก-1 ก๊ก-2 น้องมีน น้องเชษฐา น้องใจ้ และเพื่อนร่วมรุ่นได้แก่ ย้อย พีส้ม พีหน้อย พีนพ พีภค พี ใจ้ เม้า พีแจจ น้องบู่ย พู น้องใหม่ น้องพิม น้องวุฒิ เอกเทศ เบิ้ล ดิงลี น้องจุง อ้อย นัน น้องแป๊ะ น้องจิงว น้องภัทร น้องเอี่ยม น้องเอ น้องปิม พีหนู พีหน้อย นริศ น้องนิว ชิต น้องโกเมศ ต้า น้อง แมท น้องแน็ค น้องปอนด์ น้องอิฐ น้องไอ้ต น้องวิทย์ น้องอัน น้องโต้ง พีหวาน พีหญิง พีน้อย พีพะ ยัวร์ และขอขอบคุณ พี่ ๆ เพื่อน ๆ น้อง ๆ ทุกคนที่ไม่สามารถเอ่ยชื่อในที่นี้ได้หมด สำหรับน้ำใจที่มี ให้ตลอดมา

สุดท้ายนี้ ขอกราบขอบพระคุณ คุณแม่ คุณพ่อ และพี่ชายทั้งสองคน และญาติทุก ท่านที่คอยเป็นห่วงเป็นใย ให้กำลังใจ และให้การสนับสนุนในทุกด้านตลอดมา

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง	ฏ
สารบัญรูป	ฐ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ขอบเขตของการวิจัย	4
1.4 ขั้นตอนและวิธีดำเนินการวิจัย.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย.....	4
1.6 เนื้อหาในวิทยานิพนธ์นี้	5
1.7 งานตีพิมพ์	7
2 ขั้นตอนวิธีเชิงพันธุกรรม	8
2.1 ขั้นตอนวิธีเชิงวิวัฒน์.....	8
2.2 หลักการทำงานของขั้นตอนวิธีเชิงพันธุกรรม	14
3 หน่วยการสร้าง.....	19
3.1 ทฤษฎีบทเค้าร่างของฮอลแลนด์.....	19
3.1.1 ตัวดำเนินการเลือกสำหรับการเลือกสัดส่วนค่าความเหมาะสม	20
3.1.2 ตัวดำเนินการค้นหา – การไขว้เปลี่ยนแบบจุดเดียว.....	21
3.1.3 ตัวดำเนินการค้นหา – การกลายพันธุ์ (Mutation)	22
3.1.4 ทฤษฎีบท 1 – ทฤษฎีบทเค้าร่าง (Schema Theorem).....	22

บทที่	หน้า
3.2 สมมุติฐานหน่วยการสร้าง (Building Blocks Hypothesis)	24
3.3 สมมุติฐานการเชื่อมโยง (Linkage Hypothesis)	26
3.4 การมีอยู่ของหน่วยการสร้าง	27
3.5 ความยากของปัญหาในมุมมองของหน่วยการสร้าง.....	32
3.6 การออกแบบขั้นตอนวิธีเชิงพันธุกรรมในมุมมองของหน่วยการสร้าง.....	33
3.7 การแก้ปัญหาด้วยขั้นตอนวิธีเชิงพันธุกรรม	35
3.8 ทฤษฎีบทเค้ร่างและการตีความในมุมมองของตัวดำเนินการ.....	38
4 หน่วยการสร้างแบบชิ้นส่วน และวิธีการทดสอบ	42
4.1 หน่วยการสร้างแบบชิ้นส่วน.....	42
4.2 ชิ้นส่วนในมุมมองของหน่วยการสร้าง	43
4.3 ชิ้นส่วนในมุมมองของการเรียนรู้การเชื่อมโยง	44
4.4 วิธีการทดสอบขั้นตอนวิธีที่นำเสนอและการวัดผล	45
4.4.1 อักษรย่อในส่วนของตารางและรูปภาพที่ใช้แสดงผลการทดลอง.....	49
5 การไขว้เปลี่ยนแบบชิ้นส่วน.....	57
5.1 แนวคิดและหลักการ	57
5.2 การทดลองและผลการทดลอง	63
5.3 สรุปผลการทดลอง.....	78
6 ขั้นตอนวิธีประมาณการแจกแจง	79
6.1 ขั้นตอนวิธีประมาณการแจกแจง	79
6.2 ขั้นตอนวิธีประมาณการแจกแจงในโดเมนไม่ต่อเนื่อง	81
6.2.1 แบบปราศจากการฟิตฟังก์ชันระหว่างกัน.....	83
6.2.2 แบบการฟิตฟังก์ชันเชื่อมโยงแบบคู่.....	86
6.2.3 แบบการฟิตฟังก์ชันเชื่อมโยงหลายเชิง.....	90
6.3 ความแตกต่างระหว่าง GAs กับ EDAs ในมุมมองของข้อมูลสารสนเทศร่วม.....	96
6.4 หน่วยการสร้างโดยปริยายและหน่วยการสร้างแบบชัดแจ้ง	98
6.5 การเรียนรู้การเชื่อมโยง	99
7 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม	101
7.1 แนวคิดและหลักการ	101
7.2 การทดลองและผลการทดลอง	111

บทที่	หน้า
7.3 สรุปผลการทดลอง.....	126
8 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข.....	128
8.1 แนวคิดและหลักการ.....	128
8.2 การทดลองและผลการทดลอง.....	136
8.3 สรุปผลการทดลอง.....	152
9 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA).....	154
9.1 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW รูปแบบที่ 1 (LZWGA-1).....	154
9.1.1 แนวคิดและหลักการ.....	154
9.1.2 การทดลองและผลการทดลอง.....	158
9.1.3 สรุปผลการทดลอง.....	167
9.2 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW รูปแบบที่ 2 (LZWGA-2).....	168
9.2.1 แนวคิดและหลักการ.....	168
9.2.2 การทดลองและผลการทดลอง.....	172
9.2.3 สรุปผลการทดลอง.....	181
9.3 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW รูปแบบที่ 3 (LZWGA-3).....	182
9.3.1 แนวคิดและหลักการ.....	182
9.3.2 การทดลองและผลการทดลอง.....	187
9.3.3 สรุปผลการทดลอง.....	197
10 ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA).....	198
10.1 แนวคิดและหลักการ.....	198
10.2 การทดลองและผลการทดลอง.....	206
10.3 สรุปผลการทดลอง.....	222
11 ความสามารถของขั้นตอนวิธีที่นำเสนอกับปัญหาทดสอบ.....	223
11.1 ความเป็นมาของการทดสอบ.....	223
11.2 สรุปขั้นตอนวิธีที่นำเสนอทั้งหมด.....	225
11.3 ปัญหาทดสอบที่ใช้ในงานวิจัยนี้.....	227
11.4 วิธีการทดลอง.....	231
11.5 การวัดผลการทดลอง.....	233
11.6 ผลการทดลองในส่วนขอประสิทธิภาพและสมรรถนะในการแก้ปัญหา.....	235

บทที่	หน้า
11.7 สรุปผลการทดลอง.....	257
12 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ.....	258
12.1 สรุปผลการวิจัย	258
12.2 อภิปรายผลการวิจัย.....	263
12.3 ข้อเสนอแนะ	265
รายการอ้างอิง.....	267
ภาคผนวก.....	276
ภาคผนวก ก ศัพท์บัญญัติ.....	277
ประวัติผู้เขียนวิทยานิพนธ์	282



ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
5.1	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ sGA-FC	64
7.1	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ RFT	112
8.1	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA-FC และ CFT	137
8.2	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ RFT และ CFT	137
9.1	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWGA-1	159
9.2	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWGA-2	173
9.3	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ LZWGA-1 และ LZWGA-2	174
9.4	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWGA-3	188
9.5	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ LZWGA-1 และ LZWGA-3	189
9.6	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ LZWGA-2 และ LZWGA-3	190
10.1	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWEDA	207
10.2	จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ CFT และ LZWEDA	207
11.1	ค่าความซับซ้อนเชิงเวลาของแต่ละขั้นตอนวิธีที่นำเสนอ (ต่อการทำงานใน 1 รุ่นประชากร)	227
11.2	ปัญหาทดสอบ	230

ตารางที่	หน้า
11.3	ขนาดของแต่ละปัญหาทดสอบ 230
11.4	การแบ่งกลุ่มปัญหาทดสอบเพื่อใช้ในการอธิบายผลการทดลอง..... 236
11.5	(ส่วนที่ 1) จำนวนการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของแต่ละขั้นตอนวิธีกับ แต่ละปัญหาทดสอบ 243
11.6	(ส่วนที่ 1) จำนวนประชากรและประชากรย่อยของแต่ละขั้นตอนวิธีกับแต่ละ ปัญหาทดสอบ 244
11.5	(ส่วนที่ 2) จำนวนการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของแต่ละขั้นตอนวิธีกับ แต่ละปัญหาทดสอบ 245
11.6	(ส่วนที่ 2) จำนวนประชากรและประชากรย่อยของแต่ละขั้นตอนวิธีกับแต่ละ ปัญหาทดสอบ 246
11.5	(ส่วนที่ 3) จำนวนการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของแต่ละขั้นตอนวิธีกับ แต่ละปัญหาทดสอบ 247
11.6	(ส่วนที่ 3) จำนวนประชากรและประชากรย่อยของแต่ละขั้นตอนวิธีกับแต่ละ ปัญหาทดสอบ 248
11.7	(ส่วนที่ 1) ผลการทดสอบ t-Test ที่ความเชื่อมั่น 95% เพื่อทดสอบความแตกต่างระหว่าง ผลที่ได้จากแต่ละขั้นตอนวิธีในปัญหาทดสอบเดียวกัน และแสดงเฉพาะกรณีที่เกิดจาก ขั้นตอนวิธีที่ได้ไม่แตกต่างกันอย่างมีนัยสำคัญ..... 250
11.7	(ส่วนที่ 2) ผลการทดสอบ t-Test ที่ความเชื่อมั่น 95% เพื่อทดสอบความแตกต่างระหว่าง ผลที่ได้จากแต่ละขั้นตอนวิธีในปัญหาทดสอบเดียวกัน และแสดงเฉพาะกรณีที่เกิดจาก ขั้นตอนวิธีที่ได้ไม่แตกต่างกันอย่างมีนัยสำคัญ..... 251
11.8	(ส่วนที่ 1) ผลการจัดลำดับความสามารถในการแก้ปัญหาทดสอบสำหรับแต่ละ ขั้นตอนวิธี..... 252
11.8	(ส่วนที่ 2) ผลการจัดลำดับความสามารถในการแก้ปัญหาทดสอบสำหรับแต่ละ ขั้นตอนวิธี..... 253
11.8	(ส่วนที่ 3) ผลการจัดลำดับความสามารถในการแก้ปัญหาทดสอบสำหรับแต่ละ ขั้นตอนวิธี..... 254
11.9	ผลการคำนวณสมรรถนะของแต่ละขั้นตอนวิธีกับปัญหาทดสอบ 255
11.10	ผลการเปรียบเทียบความสามารถของแต่ละขั้นตอนวิธีกับปัญหาทดสอบ (นับรวมจาก ปัญหาทดสอบย่อยทั้งหมด) 256
11.11	ขั้นตอนวิธีที่มีประสิทธิภาพในการแก้ปัญหาในกลุ่มปัญหาทดสอบ 257

สารบัญรูป

รูปที่	หน้า
2.1	รหัสเทียบของชั้นตอณวิธีเชิงพันธุกรรมแบบง่าย..... 9
2.2	การไขว้เปลี่ยนแบบจุดเดี่ยวของโครโมโซมขนาด 10 บิต ณ ตำแหน่งที่ 5..... 13
3.1	ความแตกต่างระหว่างการเชื่อมโยงพันธุกรรมระหว่างสองจีนในกระบวนการ ไขว้เปลี่ยน 27
4.1	การระบุชิ้นส่วนจากเค้าร่าง..... 43
4.2	การจัดจำแนกหน่วยการสร้างโดยใช้รูปแบบคู่จีนในปัญหา Royal Road..... 51
4.3	รูปแบบหน่วยการสร้างในโครโมโซม 51
4.4	รูปแบบชิ้นส่วนในโครโมโซม 52
4.5	การระบุชิ้นส่วนด้วยวิธีการใช้ข้อมูลสารสนเทศร่วมกันต่ำสุด 53
4.6	รูปแบบชิ้นส่วนก่อนและหลังการไขว้เปลี่ยนแบบชิ้นส่วน..... 54
4.7	โครงสร้างภาพจีโนไทป์ของประชากรในหนึ่งรุ่นประชากร 55
4.8	โครงสร้างภาพจีโนไทป์ของชิ้นส่วนในหนึ่งรุ่นประชากร 56
5.1	การดำเนินการไขว้เปลี่ยนแบบชิ้นส่วนระหว่าง 2 โครโมโซม 58
5.2	รหัสเทียบของวิธีการไขว้เปลี่ยนแบบชิ้นส่วนสำหรับวิธี sGA-FC..... 58
5.3	ผลการทดสอบของ sGA ในปัญหา Royal Road 64 บิต 65
5.4	รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบ พบ..... 66
5.5	รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบ ไม่พบ 66
5.6	ผลการทดสอบของ sGA ในปัญหา Trap-5 60 บิต 67
5.7	รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบ พบ..... 68
5.8	รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบ ไม่พบ 68
5.9	ผลการทดสอบของ sGA-FC ในปัญหา Royal Road 64 บิต 70
5.10	จำนวนชิ้นส่วนเฉลี่ยต่อโครโมโซมที่เกิดการไขว้เปลี่ยนกันกับรุ่นประชากรของ sGA-FC ใน ปัญหา Royal Road 64 บิต..... 72

รูปที่	หน้า
5.11	รูปแบบจีโนไทป์ของประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ..... 72
5.12	รูปแบบจีโนไทป์ของประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ..... 73
5.13	ผลการทดสอบของ sGA-FC ในปัญหา Trap-5 60 บิต..... 74
5.14	จำนวนชิ้นส่วนเฉลี่ยต่อโครโมโซมที่เกิดการไขว้เปลี่ยนกันกับรุ่นประชากรของ sGA-FC ในปัญหา Trap-5 60 บิต..... 75
5.15	รูปแบบจีโนไทป์ของประชากรของ sGA-FC ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 77
5.16	รูปแบบจีโนไทป์ของประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ..... 77
6.1	แนวคิดของ EDA ในกระบวนการหาค่าที่เหมาะสมที่สุด 81
6.2	รหัสเทียมของแนวคิด EDA ในโดเมนไม่ต่อเนื่อง 82
6.3	การนำเสนอในรูปแบบกราฟของวิธีประมาณการแจกแจงในการหาค่าที่เหมาะสมที่สุดเชิงการจัดแบบปราศจากการพึ่งพิงระหว่างกัน ได้แก่ UMDA, PBIL และ cGA..... 83
6.4	รหัสเทียมของ UMDA 84
6.5	รหัสเทียมของ PBIL 85
6.6	รหัสเทียมของ cGA..... 86
6.7	การนำเสนอในรูปแบบกราฟของวิธีประมาณการแจกแจงในการหาค่าที่เหมาะสมที่สุดสมเชิงการจัดกับการพึ่งพิงเชื่อมโยงแบบคู่ ได้แก่ MIMIC, COMIT และ BMDA..... 87
6.8	แนวคิด MIMIC ในการประมาณการแจกแจงความน่าจะเป็นร่วม ณ รุ่น / ใด ๆ 88
6.9	รหัสเทียมของขั้นตอนวิธี COMIT 89
6.10	รหัสเทียมของขั้นตอนวิธี BMDA 90
6.11	การนำเสนอในรูปแบบกราฟของวิธีประมาณการแจกแจงในการหาค่าที่เหมาะสมที่สุดสมเชิงการจัดกับการพึ่งพิงเชื่อมโยงหลายเชิง ได้แก่ FDA, EBNA, BOA และ EcGA 92
6.12	รหัสเทียมของขั้นตอนวิธี EBNA _{pc} , EBNA _{k2+pen} , EBNA _{BIC} 93
6.13	รหัสเทียมของ EcGA 95
6.14	รหัสเทียมของ BOA 96
7.1	ชิ้นส่วนที่ระบุได้จากโครโมโซม 103

รูปที่	หน้า
7.2	รหัสเทียบของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (RFT)..... 103
7.3	รหัสเทียบของวิธีการระบุชิ้นส่วนสำหรับวิธี RFT..... 106
7.4	รหัสเทียบของวิธีการประกอบชิ้นส่วนสำหรับวิธี RFT..... 107
7.5	หลักการดำเนินงานของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม..... 108
7.6	ผลการทดสอบของ RFT ในปัญหา Royal Road 64 บิต..... 114
7.7	ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ RFT ในปัญหา Royal Road 64 บิต..... 115
7.8	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ..... 117
7.9	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ..... 118
7.10	รูปแบบจีโนไทป์ของประชากรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ..... 119
7.11	รูปแบบจีโนไทป์ของประชากรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ..... 119
7.12	ผลการทดสอบของ RFT ในปัญหา Trap-5 60 บิต..... 121
7.13	ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ RFT ในปัญหา Trap-5 60 บิต..... 122
7.14	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 123
7.15	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ..... 124
7.16	รูปแบบจีโนไทป์ของประชากรของ RFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 125
7.17	รูปแบบจีโนไทป์ของประชากรของ RFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ..... 125
8.1	ชิ้นส่วนที่ระบุได้จากคูโครโมโซม..... 131
8.2	รหัสเทียบของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (CFT)..... 132
8.3	รหัสเทียบของวิธีการระบุชิ้นส่วนสำหรับวิธี CFT..... 132
8.4	รหัสเทียบของวิธีการประกอบชิ้นส่วนสำหรับวิธี CFT..... 133

รูปที่	หน้า
8.5	หลักการดำเนินงานของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข 135
8.6	ผลการทดสอบของ CFT ในปัญหา Royal Road 64 บิต..... 140
8.7	ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ CFT ในปัญหา Royal Road 64 บิต..... 141
8.8	รูปแบบจีโนไทป์ของประชากรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ..... 143
8.9	รูปแบบจีโนไทป์ของประชากรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ 144
8.10	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ..... 145
8.11	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ..... 145
8.12	ผลการทดสอบของ CFT ในปัญหา Trap-5 60 บิต..... 147
8.13	ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ CFT ในปัญหา Trap-5 60 บิต..... 148
8.14	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ 149
8.15	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ..... 150
8.16	รูปแบบจีโนไทป์ของประชากรของ CFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 151
8.17	รูปแบบจีโนไทป์ของประชากรของ CFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ 152
9.1	การระบุชิ้นส่วนของ LZWGA-1 และการไขว้เปลี่ยนแบบจุดเดียว 155
9.2	รหัสเทียมของวิธีการไขว้เปลี่ยนแบบจุดเดียวสำหรับ LZWGA-1..... 156
9.3	รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-1..... 157
9.4	ผลการทดสอบของ LZWGA-1 ในปัญหา Royal Road 64 บิต..... 161
9.5	ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-1 ในปัญหา Royal Road 64 บิต..... 162

รูปที่	หน้า
9.6	รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Royal Road 64 บิต กรณี หาคำตอบพบ..... 162
9.7	รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Royal Road 64 บิต กรณี หาคำตอบไม่พบ..... 163
9.8	ผลการทดสอบของ LZWGA-1 ในปัญหา Trap-5 60 บิต..... 165
9.9	ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-1 ในปัญหา Trap-5 60 บิต..... 166
9.10	รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบ พบ..... 166
9.11	รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบ ไม่พบ..... 166
9.12	การระบุชิ้นส่วนของ LZWGA-2 และการไขว้เปลี่ยนแบบจุดเดียว 169
9.13	รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-2..... 170
9.14	รหัสเทียมของวิธีการไขว้เปลี่ยนแบบจุดเดียวสำหรับ LZWGA-2..... 171
9.15	ผลการทดสอบของ LZWGA-2 ในปัญหา Royal Road 64 บิต..... 176
9.16	ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-2 ในปัญหา Royal Road 64 บิต..... 177
9.17	รูปแบบจีโนไทป์ของประชากรของ LZWGA-2 ในปัญหา Royal Road 64 บิต กรณี หาคำตอบพบ..... 178
9.18	รูปแบบจีโนไทป์ของประชากรของ LZWGA-2 ในปัญหา Royal Road 64 บิต กรณี หาคำตอบไม่พบ..... 178
9.19	ผลการทดสอบของ LZWGA-2 ในปัญหา Trap-5 60 บิต..... 179
9.20	ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-2 ในปัญหา Trap-5 60 บิต..... 180
9.21	รูปแบบจีโนไทป์ของประชากรของ LZWGA-2 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบ พบ..... 180
9.22	รูปแบบจีโนไทป์ของประชากรของ LZWGA-2 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบ ไม่พบ..... 181
9.23	การระบุชิ้นส่วนของ LZWGA-3 และการไขว้เปลี่ยนแบบจุดเดียว 183
9.24	รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-3..... 185

รูปที่	หน้า
9.25 รหัสเทียบของวิธีการไขว้เปลี่ยนแบบจุดเดี่ยวสำหรับ LZWGA-3.....	186
9.26 ผลการทดสอบของ LZWGA-3 ในปัญหา Royal Road 64 บิต.....	191
9.27 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-3 ในปัญหา Royal Road 64 บิต.....	192
9.28 รูปแบบจีโนไทป์ของประชากรของ LZWGA-3 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ.....	193
9.29 รูปแบบจีโนไทป์ของประชากรของ LZWGA-3 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ.....	193
9.30 ผลการทดสอบของ LZWGA-3 ในปัญหา Trap-5 60 บิต.....	195
9.31 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-3 ในปัญหา Trap-5 60 บิต.....	196
9.32 รูปแบบจีโนไทป์ของประชากรของ LZWGA-3 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ.....	196
9.33 รูปแบบจีโนไทป์ของประชากรของ LZWGA-3 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ.....	196
10.1 หลักการทำงานของขั้นตอนวิธีประมาณการแจกแจงแบบ LZW	200
10.2 รหัสเทียบของขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA)	201
10.3 รหัสเทียบของวิธีการระบุชิ้นส่วนสำหรับ LZWEDA	201
10.4 รหัสเทียบของวิธีการประกอบชิ้นส่วนสำหรับ LZWEDA.....	202
10.5 ตัวอย่างการเข้ารหัสโครโมโซมสี่ตัวแบบ LZW	204
10.6 หน่วยเก็บถาวรเก็บเฉพาะรหัสปีบอัดที่ไม่ซ้ำกันในแต่ละตำแหน่งเท่านั้น และเรียงตามลำดับตำแหน่ง.....	205
10.7 สร้างโครโมโซมใหม่ด้วยวิธีสุ่มเลือกรหัสจากหน่วยเก็บถาวรตามลำดับตำแหน่งเริ่มจากตำแหน่งแรกไปจนถึงตำแหน่งสุดท้าย และแปลงรหัสที่ได้กลับเป็นคำ.....	205
10.8 ผลการทดสอบของ LZWEDA ในปัญหา Royal Road 64 บิต.....	209
10.9 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ LZWEDA ในปัญหา Royal Road 64 บิต.....	210
10.10 รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ.....	212

รูปที่	หน้า
10.11	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ 213
10.12	รูปแบบจีโนไทป์ของประชากรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ..... 214
10.13	รูปแบบจีโนไทป์ของประชากรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ..... 214
10.14	ผลการทดสอบของ LZWEDA ในปัญหา Trap-5 60 บิต..... 216
10.15	ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ LZWEDA ในปัญหา Trap-5 60 บิต..... 217
10.16	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 219
10.17	รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 220
10.18	รูปแบบจีโนไทป์ของประชากรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ..... 221
10.19	รูปแบบจีโนไทป์ของประชากรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ 221
11.1	กราฟเปรียบเทียบสมรรถภาพในการแก้ปัญหาทดสอบของแต่ละขั้นตอนวิธี 256
12.1	หลักการที่ใช้ในการระบุและการประกอบชิ้นส่วน 260
12.2	การแบ่งปันความรู้ในแต่ละขั้นตอนวิธีที่น่าเสนอ 262

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

จากอดีตถึงปัจจุบันได้มีการคิดค้นและพัฒนาประสิทธิภาพ และความสามารถ ขั้นตอนวิธีที่ใช้ในการแก้ปัญหาต่าง ๆ อย่างหลากหลาย แต่ก็ยังมีหลายปัญหาที่ยังไม่สามารถแก้ได้ หรือทำได้ไม่ดีเนื่องจากหลายปัจจัย เช่น ไม่สามารถนิยามปัญหาได้อย่างสมบูรณ์ มีความรู้เกี่ยวกับปัญหานั้นน้อยเกินไป หรือปัญหามีความซับซ้อนมากเกินไป เป็นต้น จึงได้มีการคิดค้น พัฒนาขั้นตอนวิธีที่มีความสามารถในการเรียนรู้และพัฒนาวิธีการแก้ปัญหาหรือหาคำตอบที่ดีขึ้น ได้ด้วยตัวเอง ข้อดีของขั้นตอนวิธีเหล่านี้ได้แก่ สามารถประยุกต์ใช้ได้กับหลากหลายปัญหาแม้แต่ ปัญหาที่มีความรู้เกี่ยวกับปัญหานั้นน้อย สามารถใช้แก้ปัญหาที่มีรูปแบบที่มีความหลากหลาย (Multimodalities) ปัญหาที่มีความไม่ต่อเนื่องและมีข้อจำกัดสูง ปัญหาที่มีฟังก์ชันการรบกวน ปัญหาที่มีเกณฑ์การตัดสินใจหลายแบบ โดยไม่ต้องมีข้อสมมุติฐานใด ๆ เกี่ยวกับปริภูมิค้นหา (Search Space) มีต้นทุนต่ำในการปรับขั้นตอนวิธีให้เข้ากับปัญหาใหม่ ๆ และสามารถปรับแต่งให้ เข้ากับแต่ละปัญหาได้ไม่ยาก รวมถึงสามารถดำเนินการเชิงตอบโต้ได้ เช่น สามารถเปลี่ยน ค่าพารามิเตอร์ได้ในระหว่างเวลาการทำงาน เป็นต้น

หลักการพื้นฐานของขั้นตอนวิธีการเรียนรู้ (Learning Algorithms) อาศัยกฎเกณฑ์ การตีความหรือการคำนวณพื้นฐานที่ไม่ซับซ้อนเกินไปนี้กร่วมกับโครงสร้างที่ออกแบบเพื่อให้ รองรับความสามารถในการปรับตัวหรือการเปลี่ยนแปลงค่าตัวแปรในระบบรวมถึงโครงสร้าง บางส่วนได้ และสิ่งสำคัญอีกส่วนหนึ่งที่ขาดไม่ได้คือข้อมูลที่ถูกป้อนเข้าสู่ระบบเพื่อการเรียนรู้ ขั้นตอนวิธีเหล่านี้พยายามที่จะวิเคราะห์และตีความเพื่อหาความสัมพันธ์ของลักษณะข้อมูล เพื่อที่จะนำไปใช้แยกแยะระหว่างสิ่งที่ต้องการและสิ่งที่ไม่ต้องการ ดังนั้นจะเห็นว่าข้อมูลที่ ถูกป้อนเข้าสู่ระบบมีผลต่อการเรียนรู้โดยตรง และถึงแม้ว่าในกรณีที่ไม่มีความรู้ว่าข้อมูลนำเข้าควรจะ มีหน้าตาหรือมีลักษณะเป็นอย่างไร บางขั้นตอนวิธีการเรียนรู้ก็ยังสามารถทำงานได้ดี เช่น ขั้นตอน วิธีเชิงพันธุกรรม (Genetic Algorithm) เป็นต้น ในภาพรวมกลไกการทำงานของขั้นตอนวิธีการ เรียนรู้เหล่านี้จะอยู่บนพื้นฐานของการสร้างความรู้ใหม่ที่สอดคล้องกับกระบวนการเรียนรู้ที่ได้ ออกแบบไว้ จากข้อมูลที่มีหรือที่ได้รับเข้าสู่ระบบ และความรู้ใหม่เหล่านี้จะค่อย ๆ ถูกสร้างขึ้นทีละ น้อย หรือกล่าวได้ว่าการความรู้ที่ได้จากการเรียนรู้จะค่อย ๆ เพิ่มขึ้นจากน้อยไปมาก โดย ความสามารถในการเรียนรู้จะช้าหรือเร็วขึ้นอยู่กับแต่ละวิธีที่ออกแบบไว้ และในที่สุดแล้วความรู้ที่ ได้เหล่านี้จะถูกปะติดปะต่อเข้าด้วยกันจนสามารถตอบคำถาม หรือหาคำตอบได้

จากการศึกษาพฤติกรรมทางชีววิทยาในการวิวัฒนาการของสิ่งมีชีวิตของฮอลแลนด์ นำไปสู่แนวคิดในการจำลองกระบวนการทำงานของโครงสร้างทางชีววิทยาให้เป็นรูปธรรม หรือเป็นที่รู้จักในนามขั้นตอนวิธีเชิงพันธุกรรมซึ่งมีความสามารถในการเรียนรู้และการปรับตัวได้ดี และเมื่อมีการนำแนวคิดนี้มาทดลองใช้แก้ปัญหา ก็ให้ผลที่น่าพึงพอใจ ทำให้งานวิจัยในสาขานี้มีการขยายวงออกไปอย่างกว้างขวางและมีการพยายามจัดหมวดหมู่ภายใต้ชื่อสาขาการคำนวณเชิงวิวัฒนาการ (Evolutionary Computation) ซึ่งได้รวบรวมวิธีจำลองการวิวัฒนาการที่ประยุกต์ใช้กับคอมพิวเตอร์ องค์ประกอบพื้นฐานของวิธีเหล่านี้ประกอบด้วยการใช้ประชากร (Population-based) การสุ่มการแปรผัน (Variation) และการคัดเลือก (Selection) ขั้นตอนวิธีที่อาศัยหลักการวิวัฒนาการจะถูกเรียกว่าขั้นตอนวิธีเชิงวิวัฒนาการ (Evolutionary Algorithm) ซึ่งได้แก่ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) กำหนดการเชิงวิวัฒนาการ (Evolutionary Programming) กลยุทธ์เชิงวิวัฒนาการ (Evolutionary Strategies) และขั้นตอนวิธีประมาณการแจกแจง (Estimation of Distribution Algorithms: EDAs) เป็นต้น

ในระหว่างการเติบโตของการศึกษาวิจัยทางด้านนี้ก็ได้มีการนำความรู้ด้านอื่นมาประยุกต์ใช้ร่วมด้วย เนื่องจากความรู้ด้านชีววิทยาเรายังมีไม่มากพอ ในบรรดาขั้นตอนวิธีเหล่านี้สามารถอธิบายกลไกการทำงานหลักที่อยู่เบื้องหลังได้ด้วยหลักการเดียวกันนั่นก็คือสมมติฐานหน่วยการสร้าง (Building Block Hypothesis) ซึ่งกล่าวถึงสมมติฐานพื้นฐานที่ได้จากการวิเคราะห์การดำเนินการในเบื้องต้นว่าอะไรเป็นปัจจัยที่ทำให้ขั้นตอนวิธีเหล่านี้สามารถหาคำตอบได้ จากสมมติฐานดังกล่าวบอกเราว่าในข้อมูลตัวที่ดี ๆ จะมีส่วนประกอบหรือโครงสร้างที่ดี ซึ่งเรียกว่าหน่วยการสร้าง (Building Blocks) รวมอยู่ด้วย ส่วนประกอบที่ดีเหล่านี้มีลักษณะเป็นข้อมูลชิ้นเล็ก ๆ ที่เมื่อนำมาประกอบเข้าด้วยกันอย่างถูกต้องแล้วมีโอกาสทำให้ได้คำตอบที่ดีขึ้น การตีความเช่นนี้ตรงกับหลักการทำงานของขั้นตอนวิธีการเรียนรู้โดยทั่วไปดังที่ได้กล่าวก่อนหน้านี้แล้วว่าความรู้ที่ได้จะค่อย ๆ ถูกสังสมจนมีมากขึ้น และสามารถนำไปใช้ประโยชน์ได้ ความแตกต่างระหว่างแต่ละขั้นตอนวิธีในสาขาการคำนวณเชิงวิวัฒนาการจะต่างกันที่วิธีการได้มาซึ่งข้อมูลหน่วยการสร้างและความสัมพันธ์ระหว่างหน่วยการสร้าง หรือเราอาจจะเรียกรวมกันว่าการระบุหน่วยการสร้าง (Building Block Identification) และวิธีการประกอบความรู้เหล่านี้เข้าด้วยกัน (Building Block Composition)

การพัฒนาประสิทธิภาพและความสามารถในการแก้ปัญหาของขั้นตอนวิธีเหล่านี้ขึ้นอยู่กับความสามารถในการระบุและการประกอบหน่วยการสร้าง ซึ่งถ้าเราสามารถหาและประกอบเข้าด้วยกันได้อย่างถูกต้องจะทำให้สามารถหาคำตอบที่เหมาะสมที่สุดได้ และเราสามารถมองกลไกการทำงานของแต่ละขั้นตอนวิธีในมุมมองการระบุและประกอบหน่วยการสร้างได้

แตกต่างกันซึ่งทำให้เกิดแนวคิดในการพัฒนาที่หลากหลาย แนวโน้มในปัจจุบันมีการนำหลักการทางสถิติและความน่าจะเป็นเข้ามาใช้ในขั้นตอนวิธีเชิงวิวัฒนาการ หรือเรียกกันว่าขั้นตอนวิธีประมาณการແຈກແຈງ ซึ่งมีความสามารถสูงในการระบุหน่วยการสร้างที่ซับซ้อนทำให้เป็นที่นิยมในการศึกษาวิจัยอย่างกว้างขวาง แต่อย่างไรก็ดียังไม่มีวิธีใดที่สามารถระบุรูปแบบหน้าตาของหน่วยการสร้างให้เห็นได้อย่างชัดเจน (Explicitly) และแน่นอน แม้ขั้นตอนวิธีประมาณการແຈກແຈງจะมีประสิทธิภาพสูงแต่หน่วยการสร้างที่ระบุได้ส่วนใหญ่จะอยู่ในรูปแบบของความน่าจะเป็นซึ่งมีค่าไม่แน่นอน

งานวิจัยนี้ได้นำเสนอมุมมองของหน่วยการสร้างแบบใหม่ซึ่งหน่วยการสร้างแบบใหม่นี้จะให้ชื่อว่า “ชิ้นส่วน” (Fragments) เพื่อไม่ให้สับสนกับคำว่าหน่วยการสร้างแบบเดิม จากมุมมองใหม่นี้จะช่วยลดความซับซ้อนของโครงสร้างแบบเดิมลง ทำให้สามารถวิเคราะห์และแสดงให้เห็นถึงการจัดการหน่วยการสร้างได้ชัดเจนมากขึ้น รวมถึงยังช่วยให้สามารถระบุรูปแบบขององค์ประกอบที่ดีที่สุดที่มีลักษณะสอดคล้องกับหน่วยการสร้างแบบเดิมได้อย่างชัดเจนว่ามีหน้าตาเป็นอย่างไร และยังสามารถแสดงให้เห็นภาพในการผสมหรือการประกอบชิ้นส่วนเหล่านี้เข้าด้วยกันได้ง่ายขึ้น โดยที่การระบุและการประกอบเข้าด้วยกันยังสามารถทำได้หลายวิธี และงานวิจัยนี้ยังได้แสดงให้เห็นถึงความเป็นไปได้และประสิทธิผลรวมถึงหลักการทำงานในมุมมองที่นำเสนอ ซึ่งจะแบ่งออกเป็นสองส่วนหลัก คือ วิธีการระบุชิ้นส่วน และวิธีการประกอบชิ้นส่วนเหล่านี้เข้าด้วยกัน โดยในส่วนวิธีการระบุชิ้นส่วนนั้นจะแสดงในสามมุมมอง ได้แก่ วิธีการระบุแบบสุ่ม (ที่ไม่พึ่งพาความรู้หรือกฎ) การระบุแบบใช้หลักการสารสนเทศร่วมกันน้อยที่สุด และการระบุแบบอาศัยหลักการบีบอัดข้อมูล ในส่วนของการประกอบนั้นจะแสดงให้เห็นในสามมุมมอง ได้แก่ การประกอบแบบอาศัยการดำเนินการพื้นฐานโดยใช้วิธีการไขว้เปลี่ยนแบบจุดเดียว และการไขว้เปลี่ยนแบบหลายจุด และการประกอบโดยเลือกจากหน่วยเก็บข้อมูลถาวรที่มีการเก็บชิ้นส่วนไว้ร่วมกัน ซึ่งมุมมองที่กล่าวทั้งหมดนี้จะถูกนำเสนอผ่านขั้นตอนวิธีที่ได้ออกแบบไว้ทั้งหมด 7 วิธีด้วยกัน และทำการวัดเปรียบเทียบสมรรถนะกับปัญหาทดสอบทั้งหมด 17 ปัญหา ในส่วนของผลการทดลองจะนำมาวิเคราะห์และแสดงให้เห็นรายละเอียดจาก 2 ปัญหาทดสอบ เพื่อช่วยให้เข้าใจกลไกการทำงานตามหลักการของแต่ละขั้นตอนวิธีที่นำเสนอได้ดีขึ้น

1.2 วัตถุประสงค์ของการวิจัย

นำเสนอมุมมองหน่วยการสร้างแบบใหม่ และขั้นตอนวิธีใหม่ในการระบุและการประกอบหน่วยการสร้างเข้าด้วยกันในรูปแบบขั้นตอนวิธีประมาณการແຈກແຈງ

1.3 ขอบเขตของการวิจัย

1. การระบุและการประกอบหน่วยการสร้างดำเนินการภายใต้เงื่อนไขของหน่วยการสร้างที่นิยามขึ้นใหม่
2. เข้ารหัสปัญหาแบบเลขฐานสองแบบเส้นตรง (ไม่เป็นรหัสวน) เท่านั้น
3. ศึกษาประสิทธิผลของขั้นตอนวิธีที่นำเสนอเกี่ยวกับปัญหาทดสอบที่นิยมใช้กันในขั้นตอนวิธีประมาณการแจกแจง เช่น ปัญหาการจับคู่ ปัญหาการหาค่าเหมาะที่สุด เป็นต้น
4. วิเคราะห์พฤติกรรมของขั้นตอนวิธีที่นำเสนอในรูปแบบหน่วยการสร้างที่นิยามขึ้นใหม่
5. เปรียบเทียบขั้นตอนวิธีที่นำเสนอเกี่ยวกับกลวิธีอื่นที่เป็นที่ยอมรับในปัญหาทดสอบเดียวกัน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้นิยามหน่วยการสร้างแบบใหม่ที่มีความซับซ้อนน้อยลง และยังคงสอดคล้องกับนิยามหน่วยการสร้างรูปแบบเดิม
2. ได้วิธีใหม่ในการระบุและประกอบหน่วยการสร้างของผลเฉลยที่เหมาะสมสำหรับขั้นตอนวิธีประมาณการแจกแจง
3. นำทฤษฎีสารสนเทศมาประยุกต์ใช้กับงานวิจัยด้านการคำนวณเชิงวิวัฒนาการ
4. ได้วิธีการปรับปรุงประสิทธิภาพขั้นตอนวิธีประมาณการแจกแจง และได้ขั้นตอนวิธีประมาณการแจกแจงแบบใหม่ที่มีประสิทธิภาพดีขึ้น
5. ช่วยทำให้เข้าใจกลไกการทำงานของการคำนวณเชิงวิวัฒนาการได้ดีขึ้น (ในส่วนของสมมติฐานหน่วยการสร้าง)

1.5 วิธีดำเนินการวิจัย

1. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์นี้ ทั้งในส่วนของขั้นตอนวิธีเชิงพันธุกรรม ขั้นตอนวิธีประมาณการแจกแจง และทฤษฎีสารสนเทศและการบีบอัดข้อมูล
2. ออกแบบและพัฒนาวิธีการระบุและการประกอบหน่วยการสร้างเข้าด้วยกัน

3. ทดลองและปรับปรุงวิธีการที่น่าเสนอ
4. ทดสอบกับปัญหาที่เป็นที่ยอมรับในสาขางานวิจัย
5. วิเคราะห์ผลลัพธ์ที่ได้จากการระบุและการประกอบหน่วยการสร้างที่ได้นำเสนอ
6. ตีพิมพ์ผลงานในวารสารวิชาการระดับนานาชาติ
7. เขียนรายงานและจัดทำวิทยานิพนธ์

1.6 ลำดับขั้นตอนในการเสนอผลการวิจัย

เนื้อหาในวิทยานิพนธ์จะมีลำดับดังต่อไปนี้

- บทที่ 2 จะพูดถึงรายละเอียดเกี่ยวกับขั้นตอนวิธีเชิงพันธุกรรมในภาพรวม เพื่อให้มีความเข้าใจถึงที่มา วิธีการดำเนินการ และหลักการทำงานโดยย่อ
- บทที่ 3 จะพูดถึงทฤษฎีบทเค้าร่าง หน่วยการสร้างในมุมมองต่าง ๆ และการจัดการหน่วยการสร้าง เพื่อให้เข้าใจถึงกลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรมในเชิงลึกได้ดีขึ้น
- บทที่ 4 นำเสนอมุมมองหน่วยการสร้างแบบใหม่ และวิธีทดสอบความเป็นไปได้ของหน่วยการสร้างในมุมมองที่น่าเสนอ รวมถึงวิธีการตีความผลการทดสอบ
- บทที่ 5 นำเสนอวิธีการไขว้เปลี่ยนแบบชิ้นส่วน เพื่อทดสอบความเป็นไปได้และผลในการประยุกต์ใช้หน่วยการสร้างในรูปแบบที่น่าเสนอ
- บทที่ 6 จะพูดถึงรายละเอียดเกี่ยวกับขั้นตอนวิธีประมาณการແจกແจง ซึ่งเป็นวิธีที่พัฒนามาจากขั้นตอนวิธีเชิงพันธุกรรม และมีจุดเด่นในเรื่องของการแบ่งปันความรู้ร่วมกัน (Shared Knowledge)
- บทที่ 7 นำเสนอขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม เพื่อทดสอบเรื่องการแบ่งปันความรู้แบบรวมศูนย์ในมุมมองของหน่วยการสร้างที่น่าเสนอว่าให้ผลต่างจากวิธีการแบ่งปันความรู้แบบกระจายหรือไม่ และให้ผลเป็นอย่างไร

- บทที่ 8 นำเสนอขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข เพื่อทดสอบความสามารถในการระบุหน่วยการสร้างในมุมมองที่นำเสนอว่าสามารถช่วยให้ประสิทธิภาพในการหาคำตอบดีขึ้นหรือไม่ และให้ผลเป็นอย่างไร
- บทที่ 9 นำเสนอขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW เพื่อทดสอบว่าหลักการบีบอัดข้อมูลที่สอดคล้องกับรูปแบบหน่วยการสร้างที่นำเสนอจะสามารถนำมาประยุกต์ใช้กับขั้นตอนวิธีเชิงพันธุกรรมอย่างง่ายได้หรือไม่ และให้ผลเป็นอย่างไร
- บทที่ 10 นำเสนอขั้นตอนวิธีประมาณการแจกแจงแบบ LZW เพื่อทดสอบความรู้ที่ได้จากการศึกษาในมุมมองของหน่วยการสร้างรูปแบบใหม่ ที่นำเสนอจากการทดลองในบทก่อนหน้าว่าถูกต้องหรือไม่ และให้ผลเป็นอย่างไร
- บทที่ 11 ทดสอบขั้นตอนวิธีที่นำเสนอทั้งหมดกับปัญหาวัดเปรียบเทียบสมรรถนะ และเปรียบเทียบผลการทดลองกับขั้นตอนวิธีอื่นในสายงานวิจัยนี้
- บทที่ 12 สรุปเนื้อหางานวิจัยที่ได้ทำมาทั้งหมดพร้อมทั้งบทวิจารณ์และข้อเสนอแนะ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.7 งานตีพิมพ์

ในระหว่างการศึกษามีการตีพิมพ์ผลงานวิจัยดังนี้

Book Chapter

- Chalermsub Sangkavichitr and Prabhas Chongstitvatana, “Fragment as a Small Evidence of Building Blocks Existence”, Exploitation of Linkage Learning in Evolutionary Algorithms, Springer, 2010. (In Progress)

Proceeding

- Chalermsub Sangkavichitr and Prabhas Chongstitvatana “Direct and Explicit Building Blocks Identification and Composition Algorithm”, In Proceedings of the 2009 IEEE Congress on Evolutionary Computation (CEC 2009), Norway, May, 2009.
- Chalermsub Sangkavichitr and Prabhas Chongstitvatana “Dictionary Based Estimation of Distribution Algorithms”, International Symposium on Communications and Information Technologies (ISCIT), 2007, Sydney, Australia.
- Chalermsub Sangkavichitr, Nattee Nipanan and Prabhas Chongstitvatana “Computing 4-fingered Force-closure Grasps from Surface Points Using Genetic Algorithm”, IEEE International Conference on Cybernetics & Intelligent Systems and Robotics, Automation & Mechatronics (CISRAM), 2006, Bangkok, Thailand.

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

ขั้นตอนวิธีเชิงพันธุกรรม

ในบทนี้จะแบ่งเนื้อหาออกเป็นสองส่วน ส่วนแรกเป็นขั้นตอนวิธีเชิงพันธุกรรมและหลักการทํางาน ส่วนที่สองเป็นที่มาของหน่วยการสร้างและการตีความรวมถึงมุมมองอื่น ๆ ที่เกี่ยวข้อง

2.1 ขั้นตอนวิธีเชิงวิวัฒนาการ

ขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัติ (Canonical GAs) จะเริ่มตั้งแต่การเข้ารหัสผลเฉลยของปัญหาให้อยู่ในรูปแบบสายอักขระซึ่งส่วนใหญ่นิยมใช้เลขฐานสอง (สายอักขระของผลเฉลยจะถูกเรียกว่าโครโมโซม (Chromosome)) เมื่อได้รูปแบบสายอักขระของผลเฉลยแล้วเราจะทำการสร้างโครโมโซมขึ้นมาจำนวนหนึ่ง (เรียกกลุ่มโครโมโซมที่สร้างว่าประชากร (Population)) โดยที่ประชากรรุ่นแรกนี้จะประกอบด้วยโครโมโซมจำนวนหนึ่งที่มีรูปแบบการเข้ารหัสตามที่ได้กำหนดไว้ (เช่น มีขนาดของประชากรเท่ากับ 100 หมายถึงมีโครโมโซม 100 ตัว) ประชากรที่ได้ในรุ่นแรกนี้แต่ละตัวจะได้จากการสุ่มสร้างขึ้นแบบเอกรูปจากเซตของยีน (Allele) ที่เป็นไปได้ซึ่งในส่วนนี้จะขึ้นกับการเข้ารหัสที่ใช้ เช่น ถ้าเราเข้ารหัสด้วยเลขฐานสอง ยีนที่เป็นไปได้จะมีแค่ 0 กับ 1 เท่านั้น (สมมติให้โครโมโซมที่เข้ารหัสแบบเลขฐานสองขนาด 5 บิต จะมีรูปแบบที่เป็นไปได้ตั้งแต่ 00000 จนถึง 11111 โดยที่ทุกรูปแบบมีโอกาสเกิดได้เท่ากันหมด) จากนั้นโครโมโซมแต่ละตัวจะถูกประเมินค่าความเหมาะสม (Fitness Value) จากฟังก์ชันค่าความเหมาะสม (Fitness Function) ซึ่งสอดคล้องกับการเข้ารหัสที่ออกแบบไว้จนครบทุกตัว หลังจากนั้นประชากรรุ่นแรกนี้จะถูกเลือกสรรมาส่วนหนึ่งเพื่อทำหน้าที่สืบพันธุ์เป็นประชากรรุ่นถัดไปโดยวิธีการเลือกสรร (Selection Method) ในรูปแบบใดรูปแบบหนึ่ง ขั้นตอนการสืบพันธุ์ (Reproduction) จะสามารถแบ่งออกได้เป็น 2 ส่วน คือ การไขว้เปลี่ยน (Crossover) และการผ่าเหล่า (Mutation) โดยการไขว้เปลี่ยนจะเป็นการแลกเปลี่ยนยีนที่มีตำแหน่งตรงกันในโครโมโซมที่ต่างกัน (โครโมโซมคนละตัว) ส่วนการผ่าเหล่าคือการเปลี่ยนยีนในโครโมโซมให้เป็นสมาชิกตัวอื่นในเซตของยีนที่เป็นไปได้ (ตัวอย่างเช่น สำหรับการเข้ารหัสแบบเลขฐานสอง ถ้าเดิมยีนมีค่าเป็น 0 เมื่อเกิดการผ่าเหล่าจะถูกเปลี่ยนค่าเป็น 1 และในทางกลับกันถ้าเดิมมีค่าเป็น 1 จะถูกเปลี่ยนค่าเป็น 0) และจะมีโอกาสเกิดขึ้นได้กับยีนทุกตัวในแต่ละโครโมโซมที่ผ่านกระบวนการไขว้เปลี่ยนมาแล้ว และกระบวนการประเมินค่า การเลือกสรร และการสืบพันธุ์จะเกิดขึ้นซ้ำหลายรอบจนกว่าจะได้ผลเฉลยที่มีค่าความ

D คือ ประชากร

P คือ ผลเฉลยที่ได้จากวิธีการเลือก

ขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย (Simple Genetic Algorithm)

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

วนซ้ำสำหรับ $i = 1, 3, 7, \dots, R-1$ {จนกระทั่งครบ R ตัว}

$P_{i,i+1} \leftarrow$ เลือกผลเฉลย 2 ตัว (ตัวที่ i กับ $i+1$) จาก D_{g-1} ที่สอดคล้องกับ
วิธีการเลือก

$P_{i,i+1} \leftarrow$ ทำการไขว้เปลี่ยนกันโดยสุ่มเลือกตำแหน่งไขว้เปลี่ยน

$D_g \leftarrow$ นำ $P_{i,i+1}$ ไปทำการผ่าเหล่า

จบการวนซ้ำ

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 2.1 รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย

เหมาะสมเป็นที่พอใจ หรือเป็นคำตอบที่ต้องการ ขั้นตอนวิธีที่เป็นพื้นฐานที่สุดและถูกใช้เป็นฐานในการเปรียบเทียบกับวิธีอื่น ๆ คือ ขั้นตอนวิธีเชิงพันธุกรรมแบบง่ายแสดงดังรูปที่ 2.1

เพื่อให้เข้าใจกระบวนการทำงานของขั้นตอนวิธีเชิงพันธุกรรมได้ชัดเจนขึ้น จะขยายความในแต่ละส่วนโดยจะเริ่มตั้งแต่การเข้ารหัสปัญหา การประเมินค่าความเหมาะสม การเลือกสรร การไขว้เปลี่ยน จนถึงการผ่าเหล่า ดังต่อไปนี้

วิธีค้นหาและเรียนรู้ทุกวิธีจะต้องอาศัยการเข้ารหัสปัญหาให้อยู่ในรูปแบบใดรูปแบบหนึ่งก่อนทำการดำเนินการซึ่งการเข้ารหัสนี้ถือเป็นปัจจัยสำคัญที่จะช่วยให้สามารถแก้ปัญหาได้สำเร็จ GAs ส่วนใหญ่จะใช้สายอักขระเลขฐานสองที่มีความยาวและลำดับที่แน่นอน แต่ก็มีการใช้การเข้ารหัสแบบอื่น ๆ หลายแบบในงานวิจัยที่ผ่านมาซึ่งสามารถจัดจำแนกออกได้เป็น 3 จำพวกหลัก ๆ ดังนี้

1. การเข้ารหัสแบบเลขฐานสองเป็นการเข้ารหัสที่ใช้กันมากที่สุดเนื่องจากหลายเหตุผล เริ่มจากงานวิจัยในช่วงต้นของฮอลแลนด์และนักศึกษาที่สนใจในการเข้ารหัสแบบนี้ และทำเป็นแบบอย่างสืบกันมา ทฤษฎีทางด้าน GA จำนวนมากอยู่บนพื้นฐานของการเข้ารหัสแบบเลขฐานสองที่มีความยาวและลำดับที่แน่นอน และมีทฤษฎีจำนวน

มากที่สามารถประยุกต์ใช้วิธีการนี้กับการเข้ารหัสแบบอื่นได้ แต่การเข้ารหัสเหล่านั้นไม่ได้ถูกใช้อย่างแพร่หลายเหมือนทฤษฎีเริ่มแรก ส่วนการตั้งค่าพารามิเตอร์โดยทั่วไป เช่น การไขว้เปลี่ยน และการผ่าเหล่า เป็นต้น ได้มีการพัฒนาให้สอดคล้องกับรูปแบบของการเข้ารหัสเลขฐานสอง แบบแผนการเข้ารหัสแบบอื่นที่อาศัยพื้นฐานจากเลขฐานสองยังคงมีให้เห็นเช่น รหัสเกรย์ (Gray Coding) (Bethke, 1980; Caruana and Schaffer, 1988) และการเข้ารหัสคู่โครโมโซมเลขฐานสองของฮิลลิส (Hillis) (การเข้ารหัสแบบคู่โครโมโซมถูกเสนอครั้งแรกในงานวิจัย (Holland, 1975) และได้มีการอภิปรายในงานวิจัย (Goldberg, 1989a) ในงานวิจัยของฮอลแลนด์ (Holland, 1975) ได้มีการให้เหตุผลในเชิงทฤษฎีสำหรับการเข้ารหัสแบบเลขฐานสองโดยการเปรียบเทียบการเข้ารหัสสองแบบอย่างคร่าวๆกับข้อมูลชุดเดียวกันในเรื่องของความสามารถในการบรรจุข้อมูลสแนท โดยตัวแรกใช้จำนวนคู่จีน (Alleles) น้อยกว่าใช้สายอักขระยาว (เช่น ใช้สายอักขระเลขฐานสองความยาว 100 บิต) และตัวที่สองใช้จำนวนคู่จีนมากกว่าสายอักขระสั้น (เช่น ใช้สายอักขระเลขฐานสิบความยาว 30 เลข) และให้เหตุผลว่าการเข้ารหัสแบบตัวแรกจะทำให้เกิดระดับของความสัมพันธ์ขนานโดยปริยาย (Implicit Parallelism) มากกว่าตัวที่สอง เพราะตัวแรกมีรูปแบบเค้าร่างมากกว่าตัวที่สอง (2^{100} เทียบกับ 10^{30}) การให้เหตุผลโดยการนับจำนวนเค้าร่างนี้เกี่ยวข้องกับพฤติกรรมของ GA เมื่อวิเคราะห์ในแง่ของทฤษฎีเค้าร่าง นอกจากนี้แล้วจะเห็นได้ว่าการเข้ารหัสแบบเลขฐานสองไม่ได้เป็นสิ่งที่เกิดขึ้นโดยปกติในธรรมชาติและไม่ได้เหมาะกับปัญหาส่วนใหญ่ เช่นการวิวัฒนาการการให้น้ำหนักในเครือข่ายประสาทเทียม เป็นต้น และลำดับการจัดเรียงของเลขฐานสองยังมีแนวโน้มที่จะเปลี่ยนแปลงได้ง่าย

2. การเข้ารหัสแบบหลายตัวอักษร (รวมถึงการเข้ารหัสแบบเลขฐานสิบด้วย) และการเข้ารหัสแบบจำนวนจริงเป็นการเข้ารหัสที่เป็นธรรมชาติที่สุด แต่จากวิธีนับเค้าร่างของฮอลแลนด์จะดูเหมือนว่า GAs ควรจะมีสมรรถนะแย่งจากการเข้ารหัสแบบหลายตัวอักษรเมื่อเทียบกับการเข้ารหัสแบบเลขฐานสอง ปัญหานี้ได้ถูกกล่าวถึงในงานวิจัย (Antonisse, 1989) จากการเปรียบเทียบการเข้ารหัสแบบเลขฐานสองกับการเข้ารหัสแบบหลายตัวอักษรหรือการเข้ารหัสแบบจำนวนจริงในทางปฏิบัติ ผลลัพธ์ที่ได้แสดงให้เห็นว่าการเข้ารหัสแบบหลายตัวอักษรหรือแบบจำนวนจริงให้ประสิทธิภาพดีกว่าการเข้ารหัสแบบเลขฐานสอง (Janikow and Michalewicz, 1991; Wright, 1991)

แต่ประสิทธิภาพที่ได้นี้ขึ้นอยู่กับปัญหาและรายละเอียดของ GA ที่ใช้เป็นอย่างมาก และในปัจจุบันก็ยังไม่มีความเห็นตรงกันว่าจะบอกว่าการเข้ารหัสแบบใดดีที่สุด

3. การเข้ารหัสแบบต้นไม้ไม่มีให้เห็นในงานของจอห์น โคซา (John Koza) เพื่อใช้แทนโปรแกรมคอมพิวเตอร์ซึ่งมีข้อดีหลายอย่างรวมถึงมีการยอมให้ปริภูมิการค้นหามีลักษณะปลายเปิดด้วย (Open-ended) (ในทางปฏิบัติตัวดำเนินการไขว้เปลี่ยนและผ่าเหล่าสามารถสร้างต้นไม้ที่มีขนาดเท่าใดก็ได้) ลักษณะปลายเปิดนี้เองทำให้มีโอกาสเกิดความผิดพลาดขึ้นได้ งานวิจัยของโคซา (Koza, 1992, 1994) ได้พูดถึงคำจำกัดความอัตโนมัติของฟังก์ชัน (Automatic Definition of Functions) ซึ่งเป็นสิ่งเดียวที่สนับสนุนให้เกิดการโปรแกรมแบบโครงสร้างลำดับชั้น (Hierarchically Structured Programs) ในสายงานวิจัยด้านการโปรแกรมเชิงพันธุกรรม (Genetic Programming: GP) แต่เมื่อต้นไม้ที่เป็นผลลัพธ์มีขนาดใหญ่ขึ้นจะทำให้การทำความเข้าใจหรือการจัดให้อยู่ในรูปแบบง่ายเป็นไปได้อย่างยากขึ้นด้วย ประชมนคมการโปรแกรมเชิงวิวัฒนาการได้เริ่มต้นการทดลองอย่างเป็นระบบเพื่อคำนวณหาประโยชน์ของการเข้ารหัสแบบต้นไม้และได้มีการเปรียบเทียบการเข้ารหัสแบบอื่น ๆ รวมถึงมีการขยายความทฤษฎีทางด้าน GA เพื่อใช้กับ GP ด้วย (Tackett, 1994; O'Reilly and Oppacher, 1995; Riccardo Poli, 1999)

คำถามที่ว่า จะเลือกใช้การเข้ารหัสแบบใดกับปัญหาดี ลอเรนซ์ เดวี (Lawrence Davis) นักวิจัยที่มีประสบการณ์อย่างสูงในการประยุกต์ใช้ GAs กับปัญหาในโลกจริง ได้ให้คำแนะนำว่าควรใช้การเข้ารหัสที่เป็นธรรมชาติกับปัญหามากที่สุด และเลือกใช้หรือสร้าง GA ที่สามารถใช้กับการเข้ารหัสแบบนั้นได้ (Davis, 1991) จนกว่าจะมีกฎเกณฑ์ในทางทฤษฎีที่เกี่ยวกับการเข้ารหัสที่ดีกว่านี้เกิดขึ้น ซึ่งแนวคิดเชิงปรัชญานี้เป็นสิ่งที่ดีที่หลายคนนำไปปฏิบัติ และแนวทางนี้มีความคล้ายคลึงกับสายงานวิจัยอื่นทางด้านการเรียนรู้ของเครื่อง (Machine Learning) เช่น การเข้ารหัสปัญหาการเรียนรู้สำหรับโครงข่ายประสาทซึ่งทำโดยวิธีการลองถูกลองผิด (Trial and Error) เป็นต้น

สำหรับแต่ละปัญหาการคำนวณที่ต้องการหาคำตอบให้เจอจะต้องมีความสามารถในการประเมินค่าความเหมาะสมของผลเฉลย ซึ่งวิธีการคำนวณค่าความเหมาะสมนี้จะเป็นส่วนประกอบสำคัญในการประเมินค่าความเหมาะสมหรือความดีของโครโมโซม ฟังก์ชันความเหมาะสมมีหน้าที่ประเมินค่าประชากรแต่ละตัวเพื่อเป็นข้อมูลส่งต่อให้วิธีการเลือกสรรใช้ในการตัดสินใจอีกทีหนึ่ง ตัวอย่างเช่น สมมติให้ปัญหามีฟังก์ชันประเมินค่าความเหมาะสมดังนี้ $f(x) =$

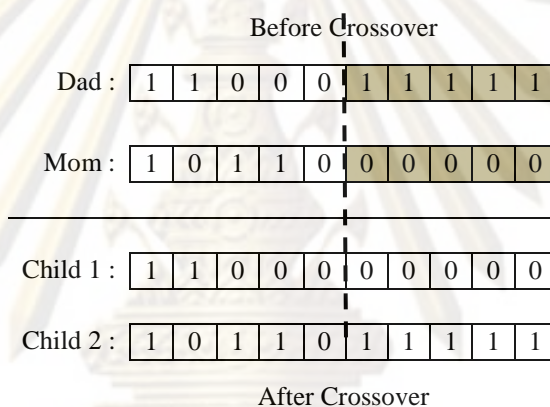
$\sum_{i=1}^n x_i$ เมื่อ n เป็นขนาดของโครโมโซม และกำหนดโครโมโซมขนาด 10 บิตจำนวน 5 ตัว ดังนี้ 11111 00000, 10101 01010, 00011 00100, 00000 00000 และ 11111 11111 จากฟังก์ชันประเมินค่าความเหมาะสม โครโมโซมที่กำหนดให้แต่ละตัวจะมีค่าความเหมาะสมเรียงตามลำดับ ดังนี้ 5, 5, 3, 0 และ 10 เป็นต้น ซึ่งจากตัวอย่างดังกล่าวจะเห็นว่าถ้าเป็นปัญหาแบบการทำให้ได้ค่ามากที่สุด (Maximization Problem) โครโมโซมตัวที่ห้าจะเป็นคำตอบ แต่ถ้าเป็นปัญหาแบบการทำให้ได้ค่าน้อยที่สุด (Minimization Problem) โครโมโซมตัวที่สี่จะเป็นคำตอบ แต่ทั้งสองแบบก็ถือว่าเป็นปัญหาการหาค่าเหมาะสมที่สุดด้วยกันทั้งคู่

หลังจากที่เรามีข้อมูลเกี่ยวกับความเหมาะสมของประชากรแต่ละตัวแล้ว ในขั้นตอนการสืบพันธุ์เราจะทำการเลือกประชากรตัวที่มีความเหมาะสมสูง หรือดีกว่าตัวอื่นนั่นเอง เพื่อใช้เป็นต้นแบบในสร้างประชากรรุ่นถัดไป การเลือกสรรที่นิยมใช้มีหลายวิธี ตัวอย่างเช่น วิธีการเลือกแบบวงล้อพันรูเล็ต (Roulette Wheel) เป็นวิธีที่โอกาสของประชากรแต่ละตัวจะถูกเลือกคิดตามสัดส่วนค่าความเหมาะสม โดยตัวที่มีความเหมาะสมสูงกว่าจะมีโอกาสถูกเลือกมากกว่าตัวที่มีความเหมาะสมต่ำกว่า (เช่น การนำค่าความเหมาะสมของแต่ละโครโมโซมมาหารด้วยค่าความเหมาะสมรวมของประชากรทั้งหมด วิธีนี้จะทำให้ได้โอกาสถูกเลือกรวมกันทั้งหมดเป็น 1) และจะใช้การสุ่มเลือกค่าขึ้นมาค่าหนึ่ง (เช่นสุ่มได้ค่า 0.7) ถ้าค่าที่สุ่มเลือกอยู่ในช่วงโอกาสถูกเลือกของประชากรตัวใด (คิดจากผลรวมโอกาสที่จะถูกเลือกของประชากรตั้งแต่ตัวแรก) ประชากรตัวนั้นจะถูกเลือกเพื่อใช้เป็นต้นแบบ วิธีการเลือกสรรแบบการแข่งขันกีฬา (Tournament Selection) จะใช้วิธีการเลือกประชากรจากประชากรทั้งหมดมาจำนวนหนึ่ง (เช่นเลือกโครโมโซมมา 4 ตัว) และเปรียบเทียบหาตัวที่มีความเหมาะสมที่สุดมากที่สุดเพื่อเลือกมาใช้เป็นต้นแบบ หรือวิธีการเลือกแบบการตัดปลาย (Truncation Selection) ทำได้โดยการนำประชากรทุกตัวมาเรียงลำดับค่าความเหมาะสมจากน้อยไปมากหรือมากไปน้อยก็ได้ (เช่นจาก 1 ถึง 100 หรือ 100 ถึง 1 ก็ได้) จากนั้นจะเลือกค่าความเหมาะสมค่าหนึ่งที่อยู่ระหว่างค่าน้อยที่สุดกับค่าที่มากที่สุด (เช่นเลือกค่า 50) และจะทำการตัดประชากรกลุ่มที่มีความเหมาะสมต่ำกว่าออกไป เหลือไว้เพียงกลุ่มที่มีความเหมาะสมดีกว่าหรือเท่ากับค่าที่กำหนด จากนั้นจะทำการเลือกต้นแบบแบบสุ่มจากประชากรที่เหลือ เป็นต้น ซึ่งการเลือกส่วนใหญ่ที่เป็นที่นิยมจะเป็นการเลือกแบบสัดส่วน (Proportional Selection) หรือ การเลือกเชิงอันดับ (Ordinal Selection) ทั้งสิ้น

ประชากรรุ่นใหม่จะเป็นผลที่ได้จากขั้นตอนการสืบพันธุ์ซึ่งจะแบ่งออกเป็นสองส่วนคือ ส่วนการไขว้เปลี่ยน และส่วนการผ่าเหล่าโดยที่จะต้องผ่านทั้งสองขั้นตอนตามลำดับ ในกระบวนการสืบพันธุ์มีการเลือกโครโมโซมมาทีละคู่ เพื่อใช้เป็นต้นแบบซึ่งเมื่อผ่านกระบวนการไขว้

เปลี่ยนแล้วจะได้โครโมโซมใหม่สองตัว โดยที่โครโมโซมแต่ละตัวจะต้องผ่านกระบวนการผ่าเหล่าอีกทีหนึ่ง หลังจากนั้นจึงจะถือว่าได้โครโมโซมที่เป็นสมาชิกของประชากรรุ่นใหม่จำนวนสองตัว (ต้นแบบสองตัวจะผลิตลูกได้ที่ละสองตัวเสมอ)

กระบวนการไขว้เปลี่ยนเป็นการแลกเปลี่ยนคู่จีนของพ่อพันธุ์กับแม่พันธุ์แบบสุ่ม โดยเลือกจุดตัดมาหนึ่งจุดจากนั้นจะทำการแลกเปลี่ยนคู่จีนระหว่างกันโดยเริ่มต้นตั้งแต่ตำแหน่งจุดตัดเป็นต้นไปจนถึงคู่จีนตำแหน่งสุดท้ายดังรูปที่ 2.2 ซึ่งจุดตัดอาจมีได้มากกว่าหนึ่งจุดขึ้นกับขั้นตอนวิธีที่ใช้ ในกรณีของขั้นตอนวิธีเชิงพันธุกรรมแบบง่ายจะมีจุดตัดเพียงจุดเดียว การไขว้เปลี่ยนไม่ได้เกิดขึ้นเสมอแต่จะมีโอกาสเกิดขึ้นสอดคล้องกับค่าความน่าจะเป็นที่จะเกิดการไขว้เปลี่ยนซึ่งเป็นค่าที่ถูกกำหนดขึ้นค่าหนึ่ง (ผู้ใช้งานเป็นผู้กำหนด) โกลด์เบิร์ก (Goldberg, 1989) ได้ให้ข้อสังเกตว่าอัตราการเกิดการไขว้เปลี่ยนมีค่ายิ่งมากยิ่งดีโดยทั่วไปควรมีค่ามากกว่า 0.8



รูปที่ 2.2 การไขว้เปลี่ยนแบบจุดเดียวของโครโมโซมขนาด 10 บิต ณ ตำแหน่งที่ 5

กระบวนการผ่าเหล่าเป็นการเปลี่ยนคู่จีนให้เป็นสมาชิกตัวอื่นในเซตของคู่จีน โอกาสที่จะเกิดการผ่าเหล่านั้นถูกกำหนดเป็นค่าความน่าจะเป็นค่าหนึ่ง (ผู้ใช้งานเป็นผู้กำหนด) โดยการผ่าเหล่ามีโอกาสเกิดขึ้นได้ทุกตำแหน่งของคู่จีนในโครโมโซม (เช่นโครโมโซมขนาด 10 บิต ทุกบิตจะมีโอกาสเกิดการผ่าเหล่าเท่ากันหมด) โดยทั่วไปแล้วเรานิยมเข้ารหัสโครโมโซมแบบเลขฐานสอง (รวมถึงขั้นตอนวิธีเชิงพันธุกรรมแบบง่ายด้วย) ดังนั้นเซตของคู่จีนที่เป็นไปได้คือ $\{0, 1\}$ ตัวอย่างเช่น สมมติให้โครโมโซมขนาด 10 บิตมีรูปแบบดังนี้ 11010 001101 ทุกบิตมีโอกาสเกิดการผ่าเหล่าเท่ากัน สมมติว่าเกิดการผ่าเหล่าที่ตำแหน่งบิตที่ 4 กับบิตที่ 9 เราจะได้โครโมโซมใหม่ที่มีรูปแบบเป็น 11000 001111 จะเห็นว่าตำแหน่งบิตที่ 4 เดิมมีค่าเป็น 1 จะถูกเปลี่ยนค่าเป็น 0 และตำแหน่งบิตที่ 9 จากเดิมเป็น 0 ก็จะถูกเปลี่ยนค่าเป็น 1 เป็นต้น โกลด์เบิร์ก (Goldberg, 1989) ได้ให้ข้อสังเกตว่าอัตราการเกิดการผ่าเหล่ายังมีค่าน้อยยิ่งดี โดยทั่วไปควรมีค่าน้อยกว่า 0.02

2.2 หลักการทำงานของขั้นตอนวิธีเชิงพันธุกรรม

ในประวัติศาสตร์การพัฒนารูปแบบการคำนวณเชิงวิวัฒนาการ (Evolutionary Algorithm: EA) จะมีการพูดถึงบทบาทสำคัญสองบทบาทที่ใช้ในการอธิบายการทำงานของขั้นตอนวิธี คือ การแสวงหาประโยชน์ (Exploitation) และการสำรวจ (Exploration) ที่เกิดขึ้นจากแต่ละกระบวนการในแต่ละขั้นตอนวิธี โดย GAs มีการบวนการหลักได้แก่ กระบวนการเลือกสรร การไขว้เปลี่ยน และการผ่าเหล่า ซึ่งจะมีการตีความในเรื่องนี้แตกต่างกันในแต่ละกระบวนการ ขึ้นกับมุมมองและการให้เหตุผลของแต่ละคน ในส่วนนี้จะพูดถึงภาพรวมอย่างคร่าวๆ เพื่อให้เห็นแนวคิดในแง่มุมต่าง ๆ งานวิจัยส่วนใหญ่แสดงให้เห็นถึงนิยามเกี่ยวกับบทบาทเหล่านี้โดยนัย หรือใช้การสื่อความหมายทางอ้อมโดยอธิบายในส่วนของกลไกการทำงานของขั้นตอนวิธีที่น่าเสนอ

มีความคิดเห็นอย่างกว้างขวางเกี่ยวกับการสำรวจและการแสวงหาประโยชน์ โดยส่วนใหญ่จะมองว่าทั้งสองบทบาทเป็นข้อตรงข้ามกัน ผู้วิจัยหลายคนได้ให้เหตุผลในเรื่องความสำคัญของการรักษาสมดุลระหว่างบทบาททั้งสอง จากงานวิจัย (Eshelman, Caruana and Schaffer, 1989) กล่าวว่า “the effectiveness of the GA depends upon the appropriate mix of exploration and exploitation” ประสิทธิภาพของ GA ขึ้นอยู่กับความเหมาะสมในการจัดสรรบทบาทการสำรวจกับการแสวงหาประโยชน์ งานวิจัย (Eshelman and Schaffer, 1993) กล่าวว่า “From this emerges a focus on the exploitation/exploration tradeoff” จากสิ่งที่แสดงให้เห็นมีจุดสนใจอยู่ที่การแลกเปลี่ยนระหว่างการแสวงหาประโยชน์และการสำรวจ งานวิจัย (Tsutsui, Ghosh, Come and Fujimoto, 1997) มีส่วนที่กล่าวว่า “...to balance the degrees of exploration and exploitation...” ในการรักษาระดับความสมดุลของการสำรวจกับการแสวงหาประโยชน์ งานวิจัย (De Jong and Spears, 1992) กล่าวว่า “This reemphasizes the need for a balance between exploration and exploitation” สิ่งนี้ให้ความสำคัญกับความต้องการในการรักษาสมดุลระหว่างการสำรวจและการแสวงหาประโยชน์ งานวิจัย (Eshelman, Caruana and Schaffer, 1989) ได้แสดงให้เห็นว่า “Increased exploitation by selection leads to decreased exploration by crossover” การเพิ่มการแสวงหาประโยชน์โดยการเลือกสรรทำให้การสำรวจที่ได้จากการไขว้เปลี่ยนลดลง และในงานวิจัย (Eshelman and Schaffer, 1993) ยังได้ให้ความคิดเห็นอีกหลายแง่มุมเมื่อเกิดความแตกต่างระหว่างการทดลองสำรวจ (Exploratory Trails) และการทดลองแสวงหาประโยชน์ (Exploitation Trials)

ได้มีการเฝ้าสังเกตพฤติกรรมการทำงานของ GAs กับค่าพารามิเตอร์ของกระบวนการไขว้เปลี่ยนและการผ่าเหล่าที่ระดับต่าง ๆ เพื่อใช้ในการอธิบายกลไกการทำงานของ ตัวอย่างเช่น

งานวิจัย (Bäck, 1992; Fogarty, 1989; Hesser and Männer, 1991; Julstrom, 1995) ได้มีหลักฐานทางทฤษฎีและการทดลองเพื่อสนับสนุนว่าการลดอัตราการทำผ่าเหล่าลงเรื่อย ๆ ตลอดการวิวัฒนาการเป็นสิ่งที่เหมาะสม สามารถตีความได้ว่าการผ่าเหล่าเป็นสิ่งที่จำเป็นในตอนเริ่มต้นของกระบวนการค้นหาแต่จะกลายเป็นความเสี่ยงในตอนหลัง หรือกล่าวอีกนัยหนึ่งได้ว่าการสำรวจควรจะค่อย ๆ เปลี่ยนเป็นการแสวงหาประโยชน์ ในขณะที่ประชาคม GA มีความเห็นว่าการผ่าเหล่ามีบทบาทสำคัญมากกว่าเป็นเพียงแค่การทำให้เกิดเงินที่หายไปในหมู่ประชากร

กระบวนการที่ได้รับความนิยมอย่างมากอีกกระบวนการหนึ่งคือการไขว้เปลี่ยนซึ่งมีความสามารถในการรวบรวม (Combination) หรือแบ่งแยก (Disruption) ขึ้นส่วนของข้อมูลได้งานวิจัย (De Jong and Spears, 1992; Eshelman, Caruana and Schaffer, 1989; Spears, 1993; Syswerda, 1989) ได้รวบรวมความรู้ที่มีสาระสำคัญเกี่ยวกับการแบ่งแยกที่เกิดจากการไขว้เปลี่ยนแบบหลายจุด (n-point Crossover) และการไขว้เปลี่ยนแบบเอกรูป (Uniform Crossover) โดยเฉพาะงานวิจัย (De Jong and Spears, 1992) ได้ให้ข้อสังเกตที่สำคัญมากในเรื่องความสัมพันธ์ระหว่างการแบ่งแยกกับความสามารถในการรวม (Recombination) เค้าวางเข้าด้วยกัน โดยกล่าวว่า "... if one operator is better than another for survival, [i.e. less disruptive] it is worse for recombination (and vice versa). This ... suggests very strongly that one cannot increase the recombination power without a corresponding increase in disruption" ถ้ามีตัวดำเนินการตัวหนึ่งดีกว่าตัวอื่น ๆ ในการทำให้เค้าวางอยู่รอดจะเป็นผลเสียสำหรับการรวมเข้าด้วยกัน เหตุการณ์นี้บ่งบอกอย่างชัดเจนว่าตัวดำเนินการหนึ่ง ๆ จะไม่สามารถเพิ่มประสิทธิภาพในการรวมเข้าด้วยกันโดยปราศจากการแบ่งแยกที่เพิ่มขึ้นด้วย การไขว้เปลี่ยนยังมีคุณลักษณะที่สำคัญคือความสามารถในการหาพื้นที่ที่มีความหลากหลายในปริภูมิค้นหา หรือกล่าวได้ว่าการไขว้เปลี่ยนสามารถทำให้เกิดโครโมโซมที่มีความแตกต่างกัน สิ่งทีกล่าวเหล่านี้ตรงกับงานวิจัย (Eshelman, Caruana and Schaffer, 1989) ที่กล่าวว่า "exploration in the form of new recombinations comes at the cost of disruption" การสำรวจในรูปแบบของการรวมเข้าด้วยกันมาพร้อมกับการทำให้เกิดการแย่งแยก อย่างไรก็ตามยังไม่มีข้อสรุปโดยทั่วไปเกี่ยวกับความสัมพันธ์ระหว่างการแบ่งแยกที่เกิดจากตัวดำเนินการกับประสิทธิภาพของ GA ทั้งนี้อาจเป็นเพราะความสัมพันธ์ดังกล่าวขึ้นอยู่กับแต่ละปัญหา การแบ่งแยกที่เกิดจากตัวดำเนินการผ่าเหล่า นั้นจะไม่ขึ้นอยู่กับโครโมโซม (ผลเฉลย) ที่ดำเนินการด้วย (ตัวดำเนินการผ่าเหล่าไม่สนใจโครโมโซมว่าเป็นตัวใด แต่จะสนใจเพียงแค่ว่าเงินเท่านั้นว่าจะถูกเปลี่ยนหรือไม่ถูกเปลี่ยน) ซึ่งต่างจากการรวมเข้าด้วยกันที่ขึ้นกับโครโมโซมที่ดำเนินการด้วย ถ้าต้นแบบมีหน้าตาเหมือนกันเมื่อผ่านกระบวนการไขว้เปลี่ยนจะได้ลูกที่มีหน้าตาเหมือนพ่อแม่ด้วย และถ้าพ่อแม่มีหน้าตาคู่คล้ายกันมาก

เมื่อผ่านกระบวนการไขว้เปลี่ยนก็จะทำให้ได้ลูกที่มีหน้าตาใกล้เคียงกับพ่อแม่มากด้วย ดังนั้นโดยทั่วไปแล้วเมื่อกระบวนการวิวัฒนาการผ่านไปได้สักระยะหนึ่งประชากรที่ได้จะเริ่มมีการลู่เข้า (เช่น ประชากรเริ่มมีหน้าตาคคล้ายคลึงกันมากขึ้น) เนื่องจากเกิดการแบ่งแยกน้อยลงในกระบวนการไขว้เปลี่ยน หรือกล่าวได้ว่าการไขว้เปลี่ยนทำให้เกิดการเปลี่ยนแปลงจากการสำรวจไปเป็นการแสวงหาประโยชน์ที่ละน้อยโดยอัตโนมัติ

นอกจากการไขว้เปลี่ยนแล้วยังมีการศึกษาเกี่ยวกับความสามารถในการค้นหาของตัวดำเนินการอื่น ๆ ด้วยเช่น จากผลการทดลองในงานวิจัย (Schaffer and Eshelma, 1991) แสดงให้เห็นถึงความสำคัญของการผ่าเหล่าและการเลือกสรรมากขึ้นเมื่อเทียบกับความเชื่อที่เคยมีมาก่อนหน้านี้ ถึงอย่างนั้นก็ยังมีข้อสังเกตว่า GA ที่มีอัตราการเกิดการแบ่งแยกที่สูงเนื่องจากตัวดำเนินการไขว้เปลี่ยนจะให้ผลที่ดีกว่า GA ที่มีแต่การผ่าเหล่าเพียงอย่างเดียวในปัญหาที่ยึดในระดับต่ำมีปฏิสัมพันธ์ระหว่างกัน นอกจากนี้ในงานวิจัย (Hordijk and Manderick, 1995; Eiben and Schippers, 1996) ยังได้มีการศึกษาถึงประโยชน์ของการรวมเข้าด้วยกันในภูมิทัศน์เอ็นเค (NK Landscape) ที่อนุญาตให้สามารถรับความหยาบของภูมิทัศน์ได้ทีละน้อย และได้ข้อสรุปว่า “recombination is useless on uncorrelated landscapes, but useful when high peaks are near one another and hence carry mutual information about their joint locations in the fitness landscape” การรวมเข้าด้วยกันไม่ก่อให้เกิดประโยชน์ในภูมิทัศน์ที่ไม่มีความสัมพันธ์เกี่ยวเนื่องกันแต่จะทำให้เกิดประโยชน์เมื่อยอดที่สูงอยู่ใกล้ยอดอื่น ๆ และมีข้อมูลสนเทศร่วมกันในส่วนของตำแหน่งจุดเชื่อมต่อไปในภูมิทัศน์ความเหมาะสม มุมมองในงานวิจัย (Caruana and Schaffer, 1989) ได้มีการเปรียบเทียบระหว่างการไขว้เปลี่ยนกับการผ่าเหล่า และได้ข้อสรุปว่าตัวดำเนินการทั้งสองมีจุดประสงค์ที่ต่างกันโดยกล่าวว่า “Mutation serves to create random diversity in the population, while crossover serves as an accelerator that promotes emergent behavior from components” การผ่าเหล่าทำให้เกิดความหลากหลายแบบสุ่มในประชากร ในขณะที่การไขว้เปลี่ยนเป็นตัวเร่งที่สนับสนุนให้เกิดการแสดงพฤติกรรมจากส่วนประกอบต่าง ๆ นอกจากนั้นการไขว้เปลี่ยนจะเป็นประโยชน์ในการผสมการชดเชยสะสม (Accumulated Payoff) ในขณะที่เดียวกันก็เป็นอันตรายในกรณีที่พบพื้นที่ที่เหมาะสมแล้ว (อาจทำให้พื้นที่เปลี่ยนไปได้)

จากที่กล่าวมาทั้งหมดจะเห็นว่านักวิจัยส่วนใหญ่ไม่ได้มีความเห็นลงรอยกันและมีการนำเสนอในหลายแง่มุมซึ่งมีทั้งมุมมองที่สอดคล้องกันและมุมมองที่ขัดแย้งกัน โดยสามารถสรุปเป็นแง่มุมต่าง ๆ ได้ดังนี้

1. การใช้ข้อมูลสันทะที่มีอยู่แล้วถือว่าเป็นการแสวงหาประโยชน์ ซึ่งรวมเอาแง่มุมที่ว่า การใช้ข้อมูลสันทะที่เลวหรือใช้ข้อมูลสันทะที่ดีในทางที่ผิดจะถูกมองว่าเป็นการแสวงหาประโยชน์ด้วยเนื่องจากมีความเป็นไปได้ในการใช้ข้อมูลสันทะที่เลวเพื่อที่จะบอกว่าพื้นที่ไหนเป็นพื้นที่ไม่ดีในปริภูมิค้นหาซึ่งควรหลีกเลี่ยง แต่โดยทั่วไปแล้วจะนิยมใช้เพียงข้อมูลสันทะที่ดีในทางที่เป็นประโยชน์
2. กระบวนการเลือกสรรถือว่าเป็นที่มาของการแสวงหาประโยชน์ ในแง่มุมนี้จะมีความสอดคล้องกับข้อที่ 1. มากน้อยเพียงใดยังไม่สามารถสรุปได้แน่ชัด
3. ตัวดำเนินการผ่าเหล่าและการไขว้เปลี่ยน (การรวมเข้าด้วยกัน) ทำให้เกิดการสำรวจมุมมองนี้อาจจะเป็นจริงส่วนหนึ่งเนื่องจากเป็นตัวดำเนินการที่มีหน้าที่สร้างความหลากหลาย แต่ทว่าการแสวงหาประโยชน์จากเค้าร่างหนึ่งสามารถนิยามเป็นการสำรวจระนาบเกิน (Hyperplane) ที่เกี่ยวข้องได้ เนื่องจากหลายจุดในระนาบเกินสามารถสร้างได้จากการผ่าเหล่าและการไขว้เปลี่ยน ในขณะที่เค้าร่างไม่ได้ถูกแบ่งแยก Eshelman และ Schaffer เรียกเหตุการณ์นี้ว่าการทดลองแสวงหาประโยชน์ (exploitive trials) (Eshelman and Schaffer, 1993)
4. การแสวงหาประโยชน์และการสำรวจเป็นข้อตรงข้ามกันที่ต้องการจุดสมดุล แง่มุมนี้ดูเหมือนว่าจะเป็นจริงเมื่อสามารถระบุการแสวงหาประโยชน์ที่เป็นประโยชน์ได้ ลองพิจารณาตัวอย่างต่อไปนี้ ในระดับของเค้าร่างถ้าเราต้องการให้มีจำนวนเค้าร่างที่ถูกใช้มากที่สุดก็จะเหมือนเป็นการคัดลอกโครโมโซม (ผลเฉลย) ที่มีอยู่แล้ว แต่ถ้าสมมติให้เค้าร่างที่ดีของโครโมโซมตัวนี้เป็นครึ่งซ้ายทั้งหมด และโครโมโซมตัวอื่นมีเค้าร่างที่ดีอยู่ครึ่งขวาทั้งหมด ถ้าตีความการแสวงหาประโยชน์ว่าเป็นการใช้ส่วนที่ดีจะทำให้เราสามารถเพิ่มการแสวงหาประโยชน์โดยการประกอบส่วนที่ดีทั้งสองส่วน (ครึ่งซ้ายกับครึ่งขวา) เข้าด้วยกันทำให้ได้โครโมโซมตัวใหม่ ด้วยวิธีนี้จะเห็นว่าเป็นการเพิ่มการสำรวจไปพร้อม ๆ กันด้วย (เนื่องจากมีส่วนที่ดีครึ่งขวาในหลายโครโมโซม)
5. จะรู้ได้อย่างไรว่าจะมีโครโมโซมใด หรือคู่จีนใด หรือเค้าร่างใด มีข้อมูลที่เป็นประโยชน์ และจะสามารถแสวงหาประโยชน์จากข้อมูลเหล่านี้ได้อย่างไร คำถามนี้เป็นเรื่องที่มีความซับซ้อนเนื่องจากขึ้นกับแต่ละปัญหา ถ้าเป็นปัญหาที่จีนไม่มีปฏิสัมพันธ์กันในรูปแบบพิเศษ (Non-epistatic) รูปแบบการจัดหมู่ (Combination) ที่ดีจะทำให้ได้ข้อมูลที่มีประโยชน์มาก (สามารถใช้รูปแบบการจัดหมู่ในส่วนย่อยนี้กับโครโมโซมตัวใหม่เพื่อให้มีความเหมาะสมที่ดีขึ้นได้) แต่ถ้าเป็นปัญหาที่จีนแต่ละตัวมีปฏิสัมพันธ์กันในลักษณะพิเศษ

(Epistatic) ความรู้เพียงอย่างเดียวที่เราต้องการคือต้องรู้ว่ารูปแบบใดบ้างที่ทำให้โครโมโซมมีค่าความเหมาะสมดีขึ้น (ต้องการความรู้พิเศษเพิ่มขึ้น)

6. รูปแบบการเข้ารหัสปัญหาที่มีผลต่อการสำรวจและการแสวงหาประโยชน์อย่างไร งานวิจัยของ (Battle and Vose, 1991) ได้แสดงให้เห็นว่าสมมติให้ฟังก์ชัน f ใด ๆ กับสายอักขระเลขฐานสอง จะสามารถทำแผนที่ตัวผกผัน (Invertible Mapping) (ที่มีความสอดคล้องกันแบบ 1-1) ได้จากการแปลง f โดยพื้นฐานไปเป็นการนับจำนวนค่าหนึ่ง (Counting Ones) ซึ่งเป็นการพิสูจน์ว่าการเข้ารหัสเป็นหัวใจสำคัญในการประกอบข้อมูลสเนทเคเข้าด้วยกัน และโดยทั่วไปค่ากล่าวที่ว่าจำนวนของคู่จีน (บิต) ที่เป็นหนึ่งในสายอักขระเลขฐานสองกับค่าความเหมาะสมของสายอักขระนั้นมีความเกี่ยวข้องกันอย่างมีนัยสำคัญทางสถิติจะเป็นค่ากล่าวที่ไม่ถูกต้อง
7. บางครั้งคู่จีนหรือการรวมกันของคู่จีนจะได้จากการแสวงหาประโยชน์โดยวิธีการสร้างโครโมโซมตัวใหม่ที่มีคู่จีนเหล่านี้ ในกรณีอื่นการแสวงหาประโยชน์เป็นเรื่องที่เข้าใจได้ยาก และค่าของจีนจะถูกผลักดันโดยคู่จีนที่มีอยู่ในโครโมโซมบรรพบุรุษ มุมมองในเรื่องความแตกต่างของกรณีเหล่านี้เป็นสิ่งที่น่าสนใจ
8. เราจะสามารถบอกถึงความแตกต่างระหว่างการสำรวจและการแสวงหาประโยชน์ในขั้นตอนวิธีที่แตกต่างกันได้อย่างไร คำตอบของคำถามในสายงานวิจัยนี้ยังเป็นสิ่งที่คลุมเครือ และยังไม่มีการให้แนวคิดหรือข้อสังเกตที่ชัดเจนได้
9. การแสวงหาประโยชน์จะเกิดขึ้นแบบปริยาย (Implicitly) ไม่สามารถระบุได้ชัดเจน จะใช้เพียงแต่ค่าความเหมาะสมเป็นตัวบ่งชี้ว่าโครโมโซมควรจะถูกแสวงหาผลประโยชน์หรือไม่ หลังจากนั้นจะเป็นความคาดหวังตามทฤษฎีเค้าร่างว่า GAs จะเลือกใช้แต่ส่วนดี ๆ ในงานวิจัย (Tsutsui, Ghosh, Corne and Fujitomo, 1997) นักวิจัยที่ศึกษาเกี่ยวกับขั้นตอนวิธีที่แสดงการแสวงหาประโยชน์ออกมาให้เห็นได้อย่างชัดเจน (Explicitly) อธิบายไว้ว่าการทำให้การแสวงหาประโยชน์มีความชัดเจนสามารถทำให้สมรรถนะของ GAs ดีขึ้น แต่แนวคิดนี้ดูเหมือนว่าจะขัดแย้งกับแนวคิดเริ่มแรกที่มีความสละสลวย เรียบง่าย และเป็นขั้นตอนวิธีที่ดีอยู่แล้ว
10. ใช้เวลานานเท่าไรถึงจะสามารถหรือควรที่จะแสวงหาประโยชน์ โดยทั่วไปแล้วจะขึ้นอยู่กับกลไกในกระบวนการเลือกสรร และอยู่บนพื้นฐานของค่าความเหมาะสม

บทที่ 3

หน่วยการสร้าง

ในบทนี้จะกล่าวถึงทฤษฎีบทเค้าร่างซึ่งเป็นที่มาของสมมติฐานหน่วยการสร้าง และมุมมองต่าง ๆ ที่เกี่ยวข้องกับหน่วยการสร้าง เช่น การตีความ และการทำความเข้าใจกลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรมในมุมมองของหน่วยการสร้าง เป็นต้น

3.1 ทฤษฎีบทเค้าร่างของฮอลแลนด์ (Holland's GA Schema Theorem)

หลังจากที่เห็นภาพขั้นตอนการดำเนินการของขั้นตอนวิธีเชิงพันธุกรรมมาแล้ว ในส่วนนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับงานวิจัยโดยจะเริ่มกล่าวถึงที่มาของสมมติฐานหน่วยการสร้างโดยอาศัยพื้นฐานทฤษฎีบทเค้าร่างของฮอลแลนด์ (Holland's Schema Theorem) และนำไปสู่การตีความการเกิดขึ้นของหน่วยการสร้างโดยอาศัยหลักการการเชื่อมโยง

ฮอลแลนด์ได้ให้กรอบของการปรับตัว (Adaptation) ว่าเป็นแรงดึงระหว่าง การสำรวจ (Exploration) (การค้นหารูปแบบใหม่ ๆ หรือการปรับตัวที่เป็นประโยชน์) และการแสวงหาประโยชน์ (Exploitation) (การใช้หรือการแพร่ลักษณะการปรับตัว) (Holland, 1998) แรงดึงเกิดขึ้นระหว่างที่เกิดการสำรวจหรือการทดสอบเพื่อหาเค้าร่างที่ไม่เคยพบมาก่อน หรือเค้าร่างที่มีค่าความเหมาะสมต่ำซึ่งห่างไกลจากการแสวงหาประโยชน์จากการลอง และเค้าร่างที่ถูกต้อง ในทุก ๆ ระบบ เช่นระบบของสิ่งมีชีวิต เราจำเป็นที่จะต้องเผชิญหน้ากับเหตุการณ์ที่ไม่รู้ล่วงหน้าดังนั้นการปรับสมดุลระหว่างการสำรวจและการแสวงหาประโยชน์จึงเป็นสิ่งจำเป็น ระบบพยายามหาวิธีปรับตัวแบบใหม่ ๆ แต่ก็พยายามปรับพฤติกรรมอย่างต่อเนื่องโดยใช้ประสบการณ์ที่ผ่านมาเป็นเครื่องนำทางสำหรับพฤติกรรมในอนาคต

ฮอลแลนด์ได้นำเสนอทฤษฎีเพื่อใช้อธิบายคุณสมบัติโครงสร้างของขั้นตอนวิธีเชิงพันธุกรรม (GAs) (Holland, 1998) ในเชิงรูปนัย (Formalization) และเป็นที่ยอมรับแพร่หลายโดยโกลด์เบิร์ก (Goldberg, 1989) โดยมุ่งเน้นไปยังการสร้างแบบจำลองสำหรับเค้าร่าง (Schema) ที่คาดว่าจะอยู่รอดต่อไปยังรุ่นถัดไป

พิจารณาจากกรณีขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัติ (Canonical GA) ซึ่งมีองค์ประกอบดังนี้

- ตัวอักษรแทนตัวเลขฐานสอง (0,1)
- ความยาวของแต่ละผลเฉลย (Individual length) ซึ่งยาว l
- วิธีเลือกค่าความเหมาะสมแบบเป็นสัดส่วน (Proportional Selection)

- การไขว้เปลี่ยนแบบจุดเดียว (One point Crossover)
- การกลายพันธุ์ของยีน (Gene Mutation)

บทนิยาม 1 – เค้าร่าง, H หมายถึงระนาบเกิน (Hyperplane)

เค้าร่างเป็นเซตย่อยของปริภูมิของผลเฉลยที่เป็นไปได้ทั้งหมดสำหรับยีนทุกตัวที่เข้ากันได้กับแผ่นแบบ (Template) ของเค้าร่าง H

ถ้าให้ A แทนตัวอักษรของยีน (Gene Alleles) แล้ว $A \cup *$ เป็นตัวอักษรเค้าร่าง (Schema Alphabet) เมื่อ $*$ เป็นสัญลักษณ์ตัวแทนทุกค่าของยีนที่เป็นไปได้ซึ่งในที่นี้คือ $\{0,1\}$

ตัวอย่างเช่น ตัวอักษรเค้าร่าง $A \in \{0,1,*\}$ โดยที่ $* \in \{0,1\}$ เป็นต้น

บทนิยาม 2 – ลำดับเค้าร่าง (Schema Order), $o(H)$

ลำดับเค้าร่าง $o(\cdot)$ คือจำนวนของยีนที่ไม่ใช่ $*$ ในเค้าร่าง H

ตัวอย่างเช่น $o(***0***) = 1$ เป็นต้น

บทนิยาม 3 – นิยามความยาวเค้าร่าง $\delta(H)$

นิยามความยาวเค้าร่าง $\delta(H)$ คือระยะทางระหว่างยีนตัวแรกและยีนตัวสุดท้ายที่ไม่ใช่ $*$ ในเค้าร่าง H

ตัวอย่างเช่น $\delta(***0***) = 4 - 4 = 0$ เป็นต้น

จำนวนของเค้าร่างที่เป็นไปได้ทั้งหมดเป็น $(k+1)^l$ เมื่อ k แทนขนาดเซตของยีนที่เป็นไปได้ และ l แทนความยาวของเค้าร่าง และจำนวนหน้าตาผลเฉลยที่เป็นไปได้ทั้งหมดเป็น k^l

3.1.1 ตัวดำเนินการเลือก (Selection Operator) สำหรับการเลือกสัดส่วนค่าความเหมาะสม (Fitness Proportional Selection)

ความน่าจะเป็นของผลเฉลย (h) ที่เข้ากันได้กับเค้าร่าง (H) นิยามเป็น $P(h \in H)$ หาได้จากสมการ

$$P(h \in H) = \frac{m(H,t)f(H,t)}{M\bar{f}(t)} \quad (3.1)$$

โดยที่

$m(H,t)$ คือจำนวนของผลเฉลยที่เข้ากันได้กับเค้าร่าง H ที่เวลา t

$f(H, t)$ คือค่าความเหมาะสมเฉลี่ยของผลเฉลยที่เข้ากันได้กับเค้าร่าง H ที่เวลา t

M คือขนาดประชากร (จำนวนผลเฉลยทั้งหมด)

$\bar{f}(t)$ คือค่าความเหมาะสมเฉลี่ยของประชากร

บทตั้ง 1 – ภายใต้การเลือกสัดส่วนค่าความเหมาะสมสำหรับจำนวนผลเฉลยที่เข้ากันได้กับเค้าร่าง H ที่เวลา t เป็น

$$E[m(H, t+1)] = M \times P(h \in H) = \frac{m(H, t) f(H, t)}{\bar{f}(t)} \quad (3.2)$$

มีความหมายว่า เค้าร่างที่มีค่าความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ยของประชากรจะมีโอกาสถูกเลือกมากกว่า เพื่อนำไปเป็นประชากรในรุ่นถัดไป และเพื่อความถูกต้องของการตีความ จำนวนประชากรควรจะเป็นอนันต์

3.1.2 ตัวดำเนินการค้นหา (Search Operators) – การไขว้เปลี่ยนแบบจุดเดียว (Single Point Crossover)

การเลือกอย่างเดียวนั้นไม่ได้ช่วยปรับปรุงค่าความเหมาะสมของผลเฉลย การไขว้เปลี่ยนเป็นตัวดำเนินการอย่างแรกที่ทำให้เกิดการตัดแปรรหัสกระจายของเค้าร่างในประชากร ในช่วงเริ่มแรก ฮอลแลนด์ให้ความสนใจเฉพาะการสร้างขอบเขตล่าง (Lower Bound) เพียงอย่างเดียว (Holland, 1998)

บทตั้ง 2 – ภายใต้การไขว้เปลี่ยนแบบจุดเดียว ขอบเขตล่างของความน่าจะเป็นที่เค้าร่าง H จะคงอยู่ในรุ่นที่ t เป็น

$$P(H \text{ ที่คงอยู่}) = 1 - P(H \text{ ที่หายไป}) \quad (3.3)$$

หรือ

$$1 - p_c \frac{\delta(H)}{l-1} P_{diff}(H, t)$$

โดยที่ $P_{diff}(H, t)$ คือความน่าจะเป็นที่ Second Parent ไม่เข้ากันกับเค้าร่าง H

p_c คือค่าความน่าจะเป็นที่จะถูกไขว้เปลี่ยน

ในกรณีพิเศษค่าขอบเขตล่างของกรณีเลวร้ายที่สุดจะเป็น $P_{diff}(H, t) = 1$

3.1.3 ตัวดำเนินการค้นหา – การกลายพันธุ์ (Mutation)

การกลายพันธุ์จะใช้กับทุกจีน (ผลเฉลย) และการที่เค้าร่าง H จะคงอยู่ได้นั้นส่วนของจีนที่ไม่ใช่ '*' จะต้องไม่เปลี่ยนแปลง โดยสามารถเขียนความน่าจะเป็นที่จีนจะไม่เปลี่ยนแปลงได้เป็น

$$(1 - p_m)$$

โดยที่ p_m คือความน่าจะเป็นในการกลายพันธุ์

และจากบทนิยาม 2 จะได้เป็น

$$(1 - p_m)^{o(H)}$$

โดยทั่วไปแล้วเรากำหนดให้ความน่าจะเป็นในการเกิดการกลายพันธุ์มีค่าต่ำมาก ดังนั้นจะได้

$$(1 - p_m)^{o(H)} \approx 1 - o(H) p_m$$

บทตั้ง 3 – ภายใต้การกลายพันธุ์ของจีน ค่าขอบเขตล่างของความน่าจะเป็นที่ลำดับเค้าร่าง $o(H)$ ของเค้าร่าง H จะยังคงอยู่ในรุ่นที่ t เป็น

$$1 - o(H) p_m$$

3.1.4 ทฤษฎีบท 1 – ทฤษฎีบทเค้าร่าง (Schema Theorem) (Goldberg, 1989 ; Holland, 1998)

จำนวนคาดหวังของเค้าร่าง H ที่รุ่น $t + 1$ เมื่อใช้ขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัติกับการเลือกค่าความเหมาะสมแบบสัดส่วนโดยใช้การไขว้เปลี่ยนแบบจุดเดียวและการกลายพันธุ์สามารถเขียนให้อยู่ในรูปสมการได้เป็น

$$E[m(H, t + 1)] \geq \frac{m(H, t) f(H, t)}{\bar{f}(t)} \left\{ 1 - p_c \frac{\delta(H)}{1 - l} p_{diff}(H, t) - o(H) p_m \right\} \quad (3.4)$$

ทฤษฎีบทถูกอธิบายในเชิงการคาดหวังและจะถูกต้องในกรณีที่จำนวนประชากรเป็นอนันต์ แต่สำหรับกรณีที่จำนวนประชากรมีจำกัด ความแตกต่างของขนาดประชากรจะแสดงบทบาทเพิ่มขึ้นอย่างมีนัยสำคัญ (Rodgers and Prugel-Bennett, 1999)

สำหรับแบบฟอร์มทั่วไปของทฤษฎีบทเค้าร่างอาจเขียนอยู่ในรูปแบบ

$$E[m(H, t + 1)] \geq m(H, t) \alpha(H, t) \{1 - \beta(H, t)\} \quad (3.5)$$

โดยที่ $\alpha(H, t)$ คือ สัมประสิทธิ์การเลือก หรืออัตราส่วนการสืบพันธุ์

$\beta(H, t)$ คือ ความผิดพลาดการถอดรหัส (Transcription Error) (Reeves and Rowe, 2003) หรือปัจจัยการแตกแยก (Disruption Factor)

และจากสมการเราจะได้ว่า

$$\alpha(H, t) \geq \{1 - \beta(H, t)\}$$

หรือ

$$\frac{f(H, t)}{f(t)} \geq \left\{ 1 - p_c \frac{\delta(H)}{1-l} p_{diff}(H, t) - o(H) p_m \right\} \quad (3.6)$$

เราจะได้ข้อสังเกตพื้นฐานว่าความยาวเค้าร่างที่สั้น (Short Defining Length) ลำดับเค้าร่างที่ต่ำ (Low Order Schema) ของประชากรที่มีค่าความเหมาะสมสูงกว่าค่าเฉลี่ยจะมีโอกาสรอดมากกว่า ซึ่งข้อสังเกตเหล่านี้จะถูกเรียกว่าสมมุติฐานหน่วยการสร้าง (Building Block Hypothesis) (Holland, 1991; Goldberg, 1989)

การวิเคราะห์เค้าร่างของฮอลแลนด์ได้แสดงให้เห็นว่าในขณะที่คำนวณค่าความเหมาะสมของผลเฉลี่ยในประชากรออกมาเป็นค่า ๆ หนึ่งนั้น กลไกภายในก็พยายามที่จะประมาณค่าความเหมาะสมเฉลี่ยของเค้าร่างจำนวนมากไปพร้อม ๆ กันซึ่งกลไกภายในนี้ไม่จำเป็นต้องใช้หน่วยความจำหรือเวลาในการคำนวณเพิ่ม ฮอลแลนด์ได้เรียกกลไกนี้ว่า “ความสัมพันธ์ขนานโดยปริยาย” (Implicit Parallelism) (Goldberg, 1989)

จากทฤษฎีบทเค้าร่างที่กล่าวไปแล้วจะพบว่าการพิจารณาเฉพาะกรณีเลวร้ายที่สุดเท่านั้น ไม่มีการพิจารณาในแง่บวกซึ่งก่อให้เกิดการพัฒนาของทฤษฎีบทเค้าร่างแบบแม่นยำ (Exact Schema Theorem) (Stephens and Waelbroeck, 1998) และจากทฤษฎีบทได้ให้ความสำคัญกับจำนวนของเค้าร่างที่รอดไม่ใช่รูปแบบของเค้าร่างที่รอดซึ่งการพิจารณาเช่นนี้สามารถใช้มาร์คอฟเชน (Markov Chains) ในการสร้างแบบจำลองของพฤติกรรมที่มีความสัมพันธ์กับผลเฉลี่ยที่ต้องการในประชากร (Vose and Nix, 1992) และการพิจารณาว่าเค้าร่างที่เหมาะสมจะมีการเพิ่มขึ้นเป็นเลขชี้กำลัง (Exponential) เช่น ถ้าเราไม่สนใจค่าคาดหวังและถือว่าตัวดำเนินการมีผลน้อยมาก จากโกลด์เบิร์ก (Goldberg, 1989) จะได้ว่า $m(H, t + 1) \geq (1 + c) \cdot m(H, t)$ โดยที่ c คือค่าคงที่ซึ่งเค้าร่างที่เหมาะสมจะมีค่าความเหมาะสมมากกว่าค่าความเหมาะสมเฉลี่ยจากประชากร การอ้างที่กล่าวมานี้ทำให้เข้าใจผิดว่าค่าเหมาะสมเฉลี่ยของประชากรจะเพิ่มขึ้นสัมพันธ์กับค่า t ดังนั้นประชากรและค่าความเหมาะสมของเค้าร่างที่ยังคงอยู่

จะถูกรับเข้าเมื่อเวลาเพิ่มขึ้น แต่การเลือกที่จะเพิ่มจำนวนการเลือกเค้าร่างที่ดีขึ้นเป็นเลขชี้กำลังนั้น โกลด์เบิร์กได้ยกตัวอย่างปัญหาเครื่องเล่นพนันแบบอัตโนมัติชนิดเอาสตางค์ใส่แบบสองแขน (Two-armed Bandit) และได้ขยายเป็นกรณีแบบหลายแขน (K-armed Bandit) (Goldberg, 1989) ซึ่งเป็นตัวอย่างที่สอดคล้องกันกับภาวะถ่วงดุลระหว่างการสำรวจและการแสวงหาประโยชน์ดังที่ได้แสดงให้เห็นในทฤษฎีบทเค้าร่างได้เป็นอย่างดี

3.2 สมมุติฐานหน่วยการสร้าง (Building Blocks Hypothesis)

จากที่ทฤษฎีบทที่กล่าวมาช่วยทำให้เข้าใจการทำงานของขั้นตอนวิธีเชิงพันธุกรรมได้ดีขึ้นในมุมมองที่ว่า “Short, low-order, and highly fit schemata are sampled, recombined, and resampled to form strings of potentially higher fitness” (Goldberg, 1989) เค้าร่างที่มีขนาดสั้น มีลำดับเค้าร่างต่ำ และมีความเหมาะสมสูงจะถูกเลือกเพื่อที่จะนำมาต่อกันเป็นผลเฉลยที่มีค่าความเหมาะสมสูงขึ้น และเรียกเค้าร่างเหล่านี้ว่า “หน่วยการสร้าง” (Building Block) เราสามารถลดความซับซ้อนของปัญหาได้โดยการหาผลเฉลยที่ดีขึ้นจากส่วนย่อยของผลเฉลยที่มีค่าความเหมาะสมสูงจากประชากรรุ่นก่อนแทนที่จะหาผลเฉลยที่ดีโดยการลองทุกรูปแบบของการจัดที่เป็นไปได้ และมีงานวิจัยที่พยายามนำเสนอทฤษฎีบทเค้าร่างใหม่โดยอยู่บนพื้นฐานของแนวคิดค่าความเหมาะสมที่มีประสิทธิภาพ (Effective Fitness) (ค่าความเหมาะสมโดยรวม) และพิสูจน์ด้วยทฤษฎีบทของเจริงเจอร์ (Geiringer's Theorem) (Stephens and Waelbroeck, 1999a, 1999b) เพื่อแสดงว่าเค้าร่างที่มีความเหมาะสมสูงกว่าค่าความเหมาะสมแบบประสิทธิภาพเฉลี่ยจะได้รับการถูกเลือกเพิ่มขึ้นเป็นเลขชี้กำลัง และได้แสดงให้เห็นว่าความยาวของเค้าร่างที่เหมาะสมไม่จำเป็นจะต้องสั้นหรือมีลำดับที่ต่ำ

การประกอบหน่วยการสร้างเข้าด้วยกันนั้นจำเป็นที่จะต้องมีการมีแบบแผนที่ถูกต้องไม่เช่นนั้นอาจจะถูกทำลายหรือไม่สามารถนำไปสู่โครโมโซมที่มีความเหมาะสมสูงขึ้นได้ แบบแผนที่ว่านี้ขึ้นอยู่กับลักษณะการดำเนินการของตัวดำเนินการ ในขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัตินั้นจะขึ้นกับตัวดำเนินการไขว้เปลี่ยน ส่วนตัวดำเนินการกลายพันธุ์เป็นการรักษาความหลากหลายของรูปแบบหน่วยการสร้าง

เกรเฟนเตท (Grefenstette, 1993) ได้กล่าวอ้างว่างานส่วนใหญ่ทางด้านขั้นตอนวิธีเชิงพันธุกรรมถือเอาสมมุติฐานหน่วยการสร้างสถิต (Static Building Block Hypothesis: SBBH) ที่กล่าวว่า “Given any low-order, short-defining-length hyperplane [i.e., schema] partition, a GA is expected to converge to the hyperplane [in that partition] with the best static average fitness (the ‘expected winner’)” การทำให้ทุก ๆ ลำดับที่ต่ำ ส่วนแบ่งกันระนาบเกินที่

นิยามความยาวไว้สั้น (เค้าร่าง) ขั้นตอนวิธีเชิงพันธุกรรมจะถูกคาดหวังให้เข้าสู่ระนาบเกิน (ใน ส่วนแบ่งกันนั้น) ด้วยค่าความเหมาะสมที่ดีที่สุดแบบสถิต (คาดหวังว่าจะเป็นคำตอบ) คำกล่าวนี้มี น้ำหนักกว่าทฤษฎีต้นฉบับ และจากส่วนที่กล่าวว่าขั้นตอนวิธีเชิงพันธุกรรมจะเข้าสู่คำตอบจริง ของแต่ละการแข่งขันในส่วนแบ่งกันลำดับต่ำและสั้น มากกว่าเค้าร่างที่มีค่าความเหมาะสมที่สุด สมมุติฐานหน่วยการสร้างสถิตนี้ไม่ใช่สิ่งที่ฮอลแลนด์นำเสนอและยังไม่มีมีการพิสูจน์หรือตรวจสอบ ความสมเหตุสมผล และนำมาซึ่งข้อสมมุติที่ว่าฟังก์ชันค่าความเหมาะสมลวง (Deceptive Fitness Function) จะเป็นปัญหาที่ยากสำหรับขั้นตอนวิธีเชิงพันธุกรรม และยังได้ให้เหตุผลที่เป็นไปได้สอง ข้อเกี่ยวกับแบบจำลองเครื่องเล่นพินแบบอัตโนมัติชนิดเอาสตาจี้ใส่แบบสองแขน ว่า SBBH จะ ล้มเหลวเมื่อ

1. เกิดการลู่เข้า ณ ตำแหน่งใดตำแหน่งหนึ่งทำให้เค้าร่างไม่กระจายตัวแบบเอกรูป ตัวอย่างเช่น สมมติให้เค้าร่าง $111*...*$ เป็นตัวที่มีค่าความเหมาะสมสูง และ ประชากรส่วนใหญ่มีลักษณะที่สอดคล้องกับเค้าร่างนี้ ดังนั้นจะเห็นว่าเค้าร่างอื่น เช่น $***000*...*$ จะมีส่วนที่เหมือนกับค่าลู่เข้าเป็น $111000*...*$ ทำให้เราไม่สามารถประมาณลักษณะที่เป็นไปได้ของเค้าร่าง $***000*...*$ ได้อย่างถูกต้อง
2. ความแปรปรวนของค่าความเหมาะสมมีค่าสูง ถ้าค่าความเหมาะสมเฉลี่ยของเค้าร่างแบบสถิตมีความแปรปรวนสูง จะทำให้ไม่สามารถประมาณค่าความเหมาะสมเฉลี่ยได้อย่างถูกต้องตัวอย่างเช่น สมมติฟังก์ชันค่าความเหมาะสมเป็น $f(x) = 2$ ถ้า $x \in 111*...*$ และ $f(x) = 1$ ถ้า $x \in 0*...*$ นอกนั้น $f(x) = 0$ จะเห็นว่าความแปรปรวนของเค้าร่าง $1*...*$ มีค่าสูง ดังนั้น GA จะลู่เข้าสู่บริเวณที่มี ค่าความเหมาะสมสูงที่สอดคล้องกับเค้าร่างดังกล่าวเป็นเหตุให้ลำดับย่อย ทั้งหมดเป็นไปตามเค้าร่างนี้ ดังนั้นจะทำให้ไม่สามารถหาค่าประมาณของค่า ความเหมาะสมแบบสถิตได้

ประเด็นเหล่านี้แสดงให้เห็นถึงพฤติกรรมของ GAs กับปัญหาการลวง เนื่องจาก ฟังก์ชันค่าความเหมาะสมของปัญหาการลวงถูกนิยามในรูปแบบของค่าความเหมาะสมเฉลี่ยแบบ สถิตของเค้าร่าง Grefenstette ยังให้ตัวอย่างของปัญหาการลวงที่ง่ายสำหรับ GAs ในการหา คำตอบ และปัญหาที่ไม่มีการลวงที่ยากสำหรับ GAs ในการหาคำตอบ และยังได้สรุปว่าการลวงไม่ เพียงพอที่จะทำให้ปัญหามีความยากสำหรับ GAs ยังไม่มีลักษณะพิเศษแบบใดที่กล่าวข้างต้นที่ จะสามารถทำให้ประสิทธิภาพของ GAs เลวร้ายลง มีแต่เพียงการสาธิตให้เห็นถึงความไม่

เหมาะสมที่จะสรุปเกี่ยวกับพฤติกรรมของ GAs จากค่าความเหมาะสมเฉลี่ยแบบสถิติของเค้าร่าง สิ่งที่ต้องคำนึงถึงมากกว่าคือความลำเอียงที่เกิดจากขั้นตอนวิธีการเลือกสรรในแต่ละรุ่นประชากร

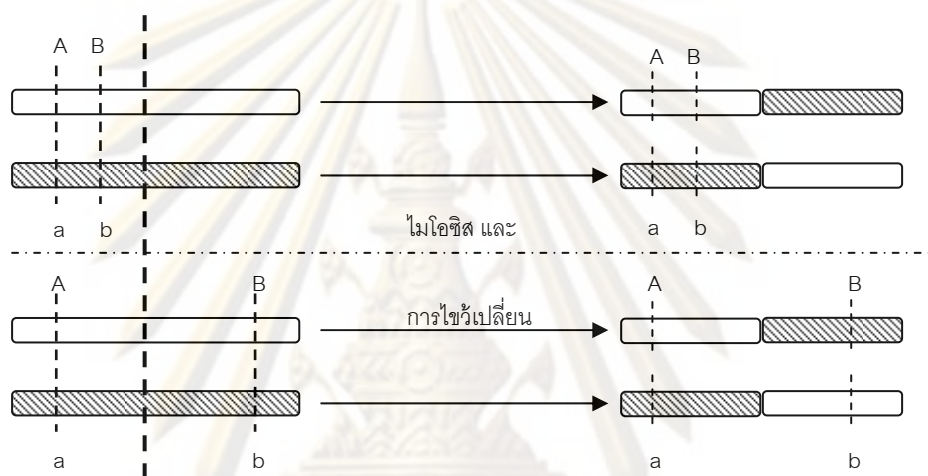
จากทฤษฎีบทเค้าร่างได้แสดงผลกระทบในทางบวกของการเลือกว่าเป็นการยอมให้มีการเพิ่มขึ้นของเค้าร่างและรักษาตัวที่ดี ๆ เอาไว้ และมุมมองในแง่ลบสำหรับการไขว้เปลี่ยนที่เป็น การทำลายเค้าร่างที่ดี แต่ไม่มีการตอบข้อสงสัยที่ว่า การไขว้เปลี่ยนสามารถประกอบเค้าร่างที่มีความเหมาะสมสูงเข้าด้วยกันเป็นดังสมมุติฐานหน่วยการสร้างจริงหรือไม่ ที่ว่าการไขว้เปลี่ยนเป็นการประกอบเค้าร่างที่สั้น และมีค่าความเหมาะสมสูงเข้าด้วยกันเพื่อทำให้ได้ผลเฉลี่ยที่มีความเหมาะสมมากขึ้น แต่ก็ไม่มีรายละเอียดในส่วนที่อธิบายว่าการประกอบควรทำอย่างไร ในการตอบข้อสงสัยนี้ได้มีการศึกษารายละเอียดเกี่ยวกับการประกอบเค้าร่างเข้าด้วยกัน (Mitchell, Forrest and Holland, 1992, 1994; Mitchell, Hraber and Crutchfield, 1996) โดยการออกแบบภูมิทัศน์ความเหมาะสม (Fitness Landscape) ที่เรียกว่า ฟังก์ชันรอยัลโรด (Royal Road Function) ซึ่งถูกออกแบบมาเพื่อตรวจจับหน่วยการสร้างในรูปแบบอุดมคติ (Idealized Form) ซึ่งผลการทดลองแสดงให้เห็นว่าการไขว้เปลี่ยนสามารถประกอบเค้าร่างที่มีความเหมาะสมสูงเข้าด้วยกันได้

3.3 สมมุติฐานการเชื่อมโยง (Linkage Hypothesis)

ในระบบชีวภาพ (Biological System) การเชื่อมโยงหมายถึงระดับของการรวมตัวในการสืบทอดของจีนตั้งแต่สองตัวขึ้นไปที่มีค่าสูงกว่าค่าคาดหวังจากการเลือกสรรอย่างอิสระ (Hartl and Jones, 1998) และในมุมมองของขั้นตอนวิธีเชิงพันธุกรรมเราจะบอกว่าจีนมีการเชื่อมโยงกันเมื่อมันมาอยู่ในโครโมโซมตัวเดียวกัน และระยะทางระหว่างจีนหมายถึงระดับของการเชื่อมโยง (Holland, 1998) และสามารถมองหน่วยการสร้างในแง่ของความสัมพันธ์กันระหว่างค่าในแต่ละตำแหน่งของเค้าร่างตั้งแต่สองตำแหน่งขึ้นไปในรูปแบบการเชื่อมโยงดังรูปที่ 3.1 ซึ่งจะเห็นได้ว่าการเชื่อมโยงแบ่งออกได้เป็นสองรูปแบบได้แก่ การเชื่อมโยงกันแบบแนบแน่น (Tightly Linkage) (รูปด้านบน) และการเชื่อมโยงกันแบบหลวม (Loosely Linkage) (รูปด้านล่าง)

ในยุคแรกของงานวิจัยในขั้นตอนวิธีเชิงพันธุกรรมเคยมีการพิจารณาปัญหาการเชื่อมโยงว่าเป็นปัญหาการเรียงลำดับ (Ordering Problem) ของรูปแบบการนำเสนอโครโมโซม และยังคงมองว่าเป็นปัญหาเดียวกับการระบุหน่วยการสร้าง (Building Block Identification) โดยถ้าขั้นตอนวิธีเชิงพันธุกรรมสามารถเรียงตำแหน่งของจีนในระหว่างกระบวนการวิวัฒนาการได้ ดังนั้นการนำเสนอแผนการเข้ารหัส (Coding Scheme) ที่ดี เป็นกุญแจสำคัญที่สามารถช่วยให้การหาคำตอบที่เหมาะสมนั้นง่ายขึ้น โดยการผลักดันกลุ่มจีนที่มีความสัมพันธ์กันให้อยู่ในหน่วยการสร้างเดียวกันในโครโมโซมจะทำให้เกิดการเชื่อมโยงที่แน่นอนในหน่วยการสร้าง และการเชื่อมโยงของ

หน่วยการสร้างจะเป็นส่วนสำคัญในกระบวนการสร้าง การระบุ การแบ่งแยก การรักษา และการผสมหน่วยการสร้าง แต่ในโลกแห่งความเป็นจริงเป็นการยากที่จะมีความรู้ว่าการเข้ารหัสแบบไหนถึงจะเหมาะสมและให้ผลดี ในมุมมองดังกล่าวงานวิจัย (Bagley, 1967) ได้ทดลองแผนการเข้ารหัสเพื่อทดสอบตัวดำเนินการผกผัน (Inversion) โดยการกลับลำดับของบางส่วนในโครโมโซม แต่ไม่ได้สรุปประโยชน์ของตัวดำเนินการแบบนี้ และงานวิจัยของฟรานซ์ (Frantz, 1972) ได้รายงานผลว่าการใช้ตัวดำเนินการผกผันทำให้กระบวนการหาคำตอบช้าลงและด้อยประสิทธิภาพอย่างมาก แต่ใน (Goldberg, 1989) ได้กล่าวว่าที่เป็นเช่นนั้นเพราะว่างานวิจัยของฟรานซ์ได้ทดสอบกับปัญหาที่ไม่ยากพอที่จะสรุปเกี่ยวกับสมมุติฐานการเชื่อมโยง



รูปที่ 3.1 ความแตกต่างระหว่างการเชื่อมโยงพันธุกรรมระหว่างสองจีนในกระบวนการไขว้เปลี่ยน รูปบนแสดงให้เห็นถึงจีนที่อยู่ใกล้ชิดกันมากกว่าทำให้โครโมโซมยังคงเดิม และรูปล่างแสดงให้เห็นว่าถ้าจีนห่างกันมากทำให้การไขว้เปลี่ยนแยกมันออกจากกันทำให้โครโมโซมเปลี่ยน

3.4 การมีอยู่ของหน่วยการสร้าง

ในการแก้ปัญหาใด ๆ ก็ตามเราจะต้องทำการแปลง (เข้ารหัส) ปัญหาให้อยู่ในรูปแบบที่สามารถคำนวณได้ก่อนที่จะดำเนินการด้วยขั้นตอนวิธีที่กำหนด ในส่วนการเข้ารหัสปัญหานี้เองที่ทำให้เกิดข้อกังขาว่าการเข้ารหัสแบบใดจะทำให้เกิดการเชื่อมโยงของโครงสร้างในผลเฉลยที่เข้ารหัสแล้วบ้าง นักวิจัยหลายคนมีความเห็นไม่ตรงกันแต่ในภาพรวมสามารถแบ่งออกได้เป็นสองกลุ่ม กลุ่มแรกมีความเชื่อว่าการเข้ารหัสที่เหมาะสมเท่านั้นจะทำให้มีความเชื่อมโยงเกิดขึ้น ส่วนกลุ่มที่สองมีความเชื่อว่าจะไม่ว่าจะเข้ารหัสแบบใดก็จะมีมีความเชื่อมโยงเกิดขึ้นทั้งสิ้น ซึ่งยังคงหาข้อสรุปที่แน่ชัดไม่ได้ แต่ถ้าใช้หลักการพิจารณาตามทฤษฎีเคาแร่งจะเห็นได้ว่าการเข้ารหัสทุกแบบล้วนทำให้เกิดการเชื่อมโยงได้ เมื่อเราพิจารณาการเชื่อมโยงคือความสัมพันธ์ระหว่างคู่จีนในเคาแร่งที่ดี ซึ่งเห็นได้จากรูปที่ 3.1 แต่ถึงอย่างไรก็ยังไม่สามารถบอกได้ว่าปัญหาหนึ่ง เมื่อเข้ารหัสแล้ว

จะมีการเชื่อมโยงเป็นแบบใด เนื่องจากการพิจารณาความสัมพันธ์ของการเชื่อมโยงนั้นขึ้นกับ ขั้นตอนวิธีที่ออกแบบ หรือแบบจำลองที่ใช้ในการหาคำตอบ ถ้าการเข้ารหัสทำให้มีการเชื่อมโยงที่ไม่สอดคล้องกับแบบจำลองที่ใช้ จะทำให้ประสิทธิภาพในการหาคำตอบลดลง หรือกล่าวได้ว่า โอกาสที่จะได้คำตอบที่ต้องการนั้นลดลง และในทางกลับกันถ้าสามารถเข้ารหัสให้มีการเชื่อมโยงเหมาะสมกับแบบจำลองที่ใช้ จะทำให้ประสิทธิภาพในการหาคำตอบเพิ่มขึ้น และมีโอกาสสูงในการที่จะพบคำตอบที่ต้องการ ขั้นตอนวิธีเชิงพันธุกรรมเป็นแบบจำลองที่ไม่มีกระบวนการเรียนรู้ ความสัมพันธ์การเชื่อมโยงโดยตรง แต่วิธีดำเนินการที่ใช้มีผลต่อการเชื่อมโยงโดยปริยาย (ในทางอ้อม) โดยแสดงให้เห็นผ่านทางกระบวนการดำเนินการกับหน่วยการสร้าง

ในส่วนนี้จะพูดถึงการมีอยู่ของหน่วยการสร้างในปัญหาใด ๆ จากทฤษฎีบทเค้ร่า และสมมติฐานหน่วยการสร้างที่อธิบายมาแล้วก่อนหน้านี้ทำให้สามารถกล่าวได้ว่าผลเฉลยของปัญหาใด ๆ เมื่อเข้ารหัสให้อยู่ในรูปแบบสายอักขระของโครโมโซมแล้วสามารถมองให้อยู่ในรูปแบบของเค้ร่าได้ ซึ่งเค้ร่าที่มีลักษณะสั้น มีลำดับต่ำ และได้มาจากโครโมโซมที่มีค่าความเหมาะสมสูงจะถูกเรียกว่าหน่วยการสร้าง ดังนั้นเราสามารถกล่าวได้ว่า ทุกปัญหามีหน่วยการสร้างอยู่จริง โดยหน่วยการสร้างที่ว่านี้จะแบ่งออกเป็น 2 จำพวกได้แก่ หน่วยการสร้างที่เป็นโครงสร้างของค่าเหมาะสมที่สุดครอบคลุม (Global Optimum) กับหน่วยการสร้างที่ไม่เป็นโครงสร้างของค่าเหมาะสมที่สุดครอบคลุม (Non Global Optimum) เพื่อให้เข้าใจได้ง่ายขึ้น สมมติให้ปัญหา OneMax ขนาด 10 บิต มีเค้ร่าของผลเฉลยตัวที่หนึ่งเป็น 11*1* ***** กับเค้ร่าของผลเฉลยตัวที่สองเป็น 1**01 ***** จากปัญหานี้เรารู้ว่าค่าคำตอบที่เหมาะสมที่สุดคือ 11111 11111 หรือมีค่าเป็นหนึ่งในสิบทุกบิตนั่นเอง ดังนั้นจะเห็นได้ว่าเค้ร่าตัวที่หนึ่ง เป็นเค้ร่าที่สอดคล้องกับคำตอบที่เหมาะสมที่สุด ซึ่งต่างจากเค้ร่าตัวที่สองซึ่งมีบิตที่มีค่าเป็น 0 ปะปนอยู่ด้วยซึ่งไม่สอดคล้องกับคำตอบที่เหมาะสมที่สุด ทำให้เราสามารถพิจารณาเค้ร่าที่หนึ่งเป็นหน่วยการสร้างที่เป็นโครงสร้างของค่าเหมาะสมที่สุดครอบคลุม (โครงสร้างของคำตอบ) และเค้ร่าที่สองเป็นหน่วยการสร้างที่ไม่เป็นโครงสร้างของค่าเหมาะสมที่สุดครอบคลุม เป็นต้น

ในส่วนคำนิยามหน่วยการสร้างนั้นจะพบปัญหาเรื่องการตีความของค่าที่ว่านี้ว่า เค้ร่าที่มีลำดับต่ำ และได้มาจากผลเฉลยที่มีค่าความเหมาะสมสูง ซึ่งทั้ง 3 ปัจจัยนี้ไม่เคยมีการระบุขอบเขตหรือเงื่อนไขที่แน่นอน ทำให้มีความกำกวมและสามารถพิจารณาได้หลากหลายรูปแบบ โดยทั่วไปแล้วในขั้นตอนวิธีเชิงพันธุกรรมเมื่อนำไปประยุกต์ใช้กับปัญหาจริงที่ไม่รู้หน้าตาของคำตอบที่เหมาะสมที่สุด รู้เพียงแต่สามารถเข้ารหัสปัญหาและสามารถวัดค่าความเหมาะสมของแต่ละผลเฉลยได้ ดังนั้นจะไม่มีโอกาสรู้ก่อนได้เลยว่าเค้ร่าที่ดี ๆ ที่สามารถเรียกได้ว่าเป็นหน่วยการสร้างนั้นควรจะหน้าตาเป็นอย่างไร จากข้อสังเกตที่ได้จากทฤษฎีบทเค้ร่าเป็นการพิจารณา

ในรุ่นประชากรหนึ่ง ๆ ในปัญหาใด ๆ และแสดงให้เห็นถึงโอกาสรอดของเค้าร่างใด ๆ ที่ไม่มีการเจาะจงเพื่อให้เห็นภาพรวม แต่เมื่อเราระบุปัญหาชัดเจนโดยธรรมชาติทั่วไปของขั้นตอนวิธีเชิงพันธุกรรมเมื่อเราคำนวณไปเรื่อย ๆ ในหลายรุ่นประชากรจะทำให้ผลเฉลยที่ได้เริ่มมีหน้าตาคล้ายกัน หรือกล่าวได้ว่าผลเฉลยจะลู่เข้าสู่ค่าความเหมาะสมหนึ่ง ๆ และหรือรูปแบบผลเฉลยหนึ่ง ๆ (Nix and Vose, 1992) ซึ่งเมื่อพิจารณาในมุมมองของเค้าร่างจะเห็นได้ว่า เค้าร่างมีการลู่เข้าสู่รูปแบบใดรูปแบบหนึ่ง และมีความยาวมากขึ้น รวมถึงมีลำดับที่สูงขึ้นเรื่อย ๆ ด้วย ดังนั้นเมื่อพิจารณาเค้าร่างในมุมมองของหน่วยการสร้างจะเห็นว่าหน่วยการสร้างก็จะมีขนาดโตขึ้นตามเค้าร่างด้วยเช่นกัน แต่ถึงอย่างไรก็ไม่สามารถหาขอบเขตที่แน่ชัดได้อยู่ดีว่าหน่วยการสร้างมีขนาดเท่าใด ถ้าพิจารณาขอบเขตที่เป็นไปได้ของหน่วยการสร้างอาจบอกได้ว่าความยาวจะเริ่มได้ตั้งแต่ 0 ถึง $l-1$ โดยที่ l คือความยาวของโครโมโซม และลำดับของโครโมโซมจะอยู่ในช่วง 1 ถึง l ที่เป็นแบบนี้เนื่องจากไม่รู้ว่าหน่วยการสร้างมีหน้าตาเป็นอย่างไร รู้แต่ว่ามีอยู่ดังที่อธิบายแล้วในส่วนก่อนหน้า อย่างไรก็ตามเรารู้ว่าหน่วยการสร้างสามารถรวมเข้าด้วยกัน (recombination) เนื่องจากตัวดำเนินการไขว้เปลี่ยน ซึ่งทำให้มีโอกาสได้หน่วยการสร้างที่มีขนาดใหญ่ขึ้นได้ ตัวอย่างเช่น สมมติให้ปัญหา OneMax ขนาด 10 บิต มีเค้าร่างสองตัวตามลำดับดังนี้ $11*1* ****$ และ $**** 1**11$ เมื่อเกิดการไขว้เปลี่ยนเค้าร่างทั้งสองตัวที่ตำแหน่งที่ห้าจะทำให้ได้เค้าร่างใหม่สองตัวที่มีหน้าตาเป็น $11*1* 1**11$ กับ $**** ****$ เห็นได้ว่าเค้าร่างใหม่ตัวแรกที่ได้มีขนาดยาวขึ้น มีลำดับที่สูงขึ้น โดยเค้าร่างใหม่นี้สามารถพิจารณาเป็นหน่วยการสร้างขนาดใหญ่ที่ประกอบขึ้นจากหน่วยการสร้างขนาดย่อยได้ เป็นต้น ซึ่งสอดคล้องกับงานวิจัย (Stephens, Waelbroeck, Aguirre, 1999) ที่แสดงให้เห็นว่าหน่วยการสร้างไม่จำเป็นต้องมีขนาดสั้นและมีลำดับที่ต่ำ เนื่องจากเราสามารถพิจารณาว่าหน่วยการสร้างขนาดใหญ่ประกอบด้วยหน่วยการสร้างขนาดย่อย ๆ ได้

ส่วนปัจจัยที่ว่าเค้าร่างต้องได้มาจากผลเฉลยที่มีค่าความเหมาะสมที่สูงนั้น ขึ้นกับขั้นตอนวิธีการเลือกที่ใช้ในการเลือกผลเฉลยซึ่งวิธีที่เป็นที่นิยมโดยทั่วไปอย่างเช่น RouletteWheel Selection, Tournament Selection หรือ Truncation เป็นต้น โครโมโซมที่มีค่าความเหมาะสมสูงกว่าจะมีโอกาสถูกเลือกมากกว่าโครโมโซมที่มีค่าความเหมาะสมต่ำกว่า ทำให้เรามีโอกาสที่จะได้โครโมโซมที่มีค่าความเหมาะสมที่สูงเมื่อเทียบกับค่าความเหมาะสมเฉลี่ยของประชากรทั้งหมดในรุ่น แต่ไม่จำเป็นจะต้องเป็นเช่นนั้นเสมอไปเนื่องจากว่าปัจจัยนี้ได้มาจากข้อสังเกตในส่วนของทฤษฎีบทเค้าร่างที่ว่าโครโมโซมที่มีค่าความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ยของประชากรจะมีโอกาสถูกเลือกสูงขึ้นไปเป็นฟังก์ชันเลขชี้กำลัง แต่ไม่ใช่ว่าตัวที่มีค่าความเหมาะสมต่ำกว่าค่าความเหมาะสมเฉลี่ยจะไม่มีโอกาสถูกเลือกเลย และหน่วยการสร้างสามารถมีอยู่ในโครโมโซมที่มีค่าความเหมาะสมที่ต่ำได้ ตัวอย่างเช่น สมมติให้ปัญหา OneMax ขนาด 10 บิต มีเค้าร่างตัวที่หนึ่ง

เป็น $1*11*$ ***** ซึ่งได้จากโครโมโซม 10110 00000 จะเห็นว่าโครโมโซมนี้มีค่าความเหมาะสมเป็น 3 ซึ่งถ้าสมมติให้ค่าความเหมาะสมเฉลี่ยเป็น 5 ดังนั้นในโครโมโซมที่มีค่าความเหมาะสมต่ำก็ สามารถมีเค้าร่างที่เป็นหน่วยการสร้างอยู่ได้ เป็นต้น และจากการตีความที่ว่าหน่วยการสร้างที่มีค่า ความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ยจะมีโอกาสถูกเลือกเพิ่มขึ้นเป็นฟังก์ชันเลขชี้กำลัง ในขณะที่หน่วยการสร้างที่มีค่าความเหมาะสมต่ำกว่าค่าความเหมาะสมเฉลี่ยมีโอกาสถูกเลือก ลดลงเป็นเลขชี้กำลัง หรือกล่าวได้ว่าเค้าร่างที่มีค่าความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ย ของประชากรมีโอกาสถูกเลือก (อยู่รอด) มากกว่า ซึ่งสามารถตีความได้ว่าค่าความเหมาะสมที่สูง หมายถึงค่าความเหมาะสมที่มีค่าไม่ต่ำกว่าค่าเหมาะสมเฉลี่ย จากมุมมองเหล่านี้เห็นได้ว่ารูปแบบ ของเค้าร่างที่เป็นหน่วยการสร้างนั้น ไม่จำเป็นต้องมีหน้าตาเหมือนรูปแบบโครงสร้างของคำตอบที่ ดีที่สุดดังตัวอย่างที่แสดงให้เห็นก่อนหน้าในปัญหา OneMax สมมุติให้โครโมโซม 11111 10000 (ขนาด 10 บิต) ในปัญหา OneMax ลองพิจารณากรณีที่เค้าร่างมีรูปแบบเป็น $**1*1$ 10*** เค้าร่าง นี้เป็นเค้าร่างที่ไม่สอดคล้องกับคำตอบที่ดีที่สุดเพราะมีบิตที่เป็น 0 ปนอยู่ด้วย แต่เนื่องจากเราไม่รู้ มาก่อนว่าหน้าตาคำตอบที่ดีที่สุดเป็นอย่างไร สิ่งที่มีเพียงอย่างเดียวคือค่าความเหมาะสม ดังนั้น เค้าร่างนี้สามารถพิจารณาให้เป็นหน่วยการสร้างได้ หรือแม้แต่กรณีที่เลวร้ายสุดจากเค้าร่าง ***** 00000 ซึ่งไม่มีแม้แต่บิตเดียวที่ตรงกับคำตอบที่ดีที่สุด แต่ก็ยังถือว่าเป็นหน่วยการสร้างได้อยู่ ดี อาจเกิดคำถามว่าแล้วถ้าเป็นเช่นนี้จะยังสามารถหาคำตอบที่ดีที่สุดได้จากหน่วยการสร้าง หรือไม่ คำตอบคือสามารถหาได้ (มีความเป็นไปได้) เนื่องจากจะเกิดการแข่งขันกันระหว่างหน่วย การสร้างที่มีเป็นโครงสร้างตรงกับรูปแบบของคำตอบที่ดีที่สุด กับหน่วยการสร้างที่มีโครงสร้างไม่ ตรงกับรูปแบบของคำตอบที่ดีที่สุด โดยกระบวนการนี้เกิดจากการแสวงหาประโยชน์และการ สำรวจดังที่ได้อธิบายไว้ในบทที่ 2 หัวข้อ 2.4 แต่ในงานวิจัยทั่วไปคำว่าหน่วยการสร้างหมายถึง หน่วยการสร้างที่เป็นโครงสร้างของรูปแบบที่เป็นคำตอบที่ดีที่สุดเท่านั้น โดยเฉพาะกับปัญหาที่ ออกแบบมาเพื่อทดสอบการมีอยู่ของหน่วยการสร้าง หรือทดสอบความสามารถในการหาหน่วย การสร้าง

งานวิจัย (Aporntewan and Chongstitvatana 2005) ได้นำเสนอวิธีการวัดระดับการ มีอยู่ของหน่วยการสร้างในปัญหาใด ๆ ทดสอบกับปัญหา Onemax ปัญหา Trap ปัญหา Parabola และปัญหาการเดินทางของนักขาย (Traveling Salesman Problem: TSP) โดยใช้ทั้ง การเข้ารหัสแบบเลขฐานสองธรรมดาและการเข้ารหัสแบบเกรย์ และเฉพาะสำหรับปัญหา TSP ได้ นำเสนอวิธีเข้ารหัส 3 รูปแบบ งานวิจัยนี้ได้ใช้หลักการของไคกำลังสอง (Chi-square) ในรูปแบบ เมทริกซ์ (Matrix) เพื่อหาระดับการเชื่อมโยงระหว่างบิตในโครงสร้างของผลเฉลย และได้กำหนด เกณฑ์ระดับความไม่ต่อเนื่อง (Degree of Discontinuity Criterion) ซึ่งพิจารณาจากปริมาณของ

หน่วยการสร้างในรุ่นประชากรที่อยู่ติดกัน เช่นรุ่นที่ 1 กับรุ่นที่ 2 และรุ่นที่ 2 กับรุ่นที่ 3 เป็นต้น เพื่อใช้ในการพิจารณาถึงความต่อเนื่องของระดับการเชื่อมโยงระหว่างรุ่นของประชากร งานวิจัยนี้แสดงให้เห็นว่าการเข้ารหัสมีผลต่อปริมาณหน่วยการสร้าง (ระดับการเชื่อมโยง) โดยวิธีการเข้ารหัสแบบเกรย์ให้ผลดีกว่าการเข้ารหัสแบบเลขฐานสองธรรมดา และสรุปได้ว่าสมมติฐานหน่วยการสร้างเป็นจริงในปัญหา OneMax ปัญหา Trap ปัญหา Parabola (เฉพาะการเข้ารหัสแบบเกรย์) และปัญหา TSP (เฉพาะการเข้ารหัสในรูปแบบที่ 3 เท่านั้น) สิ่งที่น่าสนใจเพิ่มเติมจากงานวิจัยนี้คือ แม้แต่ปัญหา OneMax ก็สามารถวัดปริมาณหน่วยการสร้างได้ ซึ่งสอดคล้องกับหลักการที่นำเสนอในส่วนก่อนหน้า และงานวิจัยนี้ได้แสดงให้เห็นถึงความเป็นไปได้ในการระบุระดับความสัมพันธ์ของการเชื่อมโยง ที่จะนำไปสู่วิธีการระบุหน่วยการสร้างได้

แนวความคิดเกี่ยวกับกลไกการทำงานของ GA อยู่บนพื้นฐานของทฤษฎีที่ว่าด้วยการแยกองค์ประกอบ (Decomposition) และการประกอบเข้าด้วยกันใหม่ (Reassembly) (Holland, 1975) ซึ่งฮอลแลนด์ได้เรียกเซตของลักษณะเฉพาะที่ทำให้เกิดส่วนประกอบย่อยของผลเฉลยที่ดีว่าหน่วยการสร้าง โดยมีแนวคิดพื้นฐานคือ การระบุหน่วยการสร้าง หรือกลุ่มย่อยของผลเฉลยที่ดี และประกอบกลุ่มย่อยเหล่านี้เข้าด้วยกันเพื่อสร้างผลเฉลยที่ดีขึ้น และการเติบโตของหน่วยการสร้างได้แนวคิดจากแข่งขันทางการตลาดที่เราต้องทำให้มันใจได้ว่าตัวที่ดีที่สุดยังคงเติบโตและมีบทบาทสำคัญใน ส่วนแบ่งตลาดของประชากร และอัตราการเติบโตควรจะไม่ช้าหรือเร็วจนเกินไป

บทบาทหนึ่งของขนาดประชากรคือการทำให้มันใจได้ว่าจะทำให้มีหน่วยการสร้างเพียงพอในประชากร (Goldberg, 1989d; Holland, 1975) การสุ่มสร้างประชากรมากขึ้นจะเป็นการเพิ่มโอกาสให้เกิดหน่วยการสร้างที่มีความซับซ้อนมากขึ้น และสำหรับปัญหายากเป็นการไม่สมเหตุผลถ้าเราสมมติว่าหน่วยการสร้างที่มีขนาดสั้นหรือลำดับต่ำเท่านั้นที่จะถูกเลือกในช่วงเริ่มต้นของการดำเนินการ เนื่องจากปัญหาต้องการหน่วยการสร้างที่มีความซับซ้อนมากขึ้นถึงจะทำให้สามารถแก้ปัญหาได้ ดังนั้นปัญหาจึงต้องการจำนวนประชากรเพิ่มขึ้นเพื่อเพิ่มโอกาสในการเกิดหน่วยการสร้างที่มีความซับซ้อนและมีคุณภาพที่ดีขึ้น ประเด็นที่สำคัญอีกประเด็นหนึ่งคือเรื่องคุณภาพและปริมาณของหน่วยการสร้างซึ่งเมื่อมีจำนวนประชากรมากขึ้น ทำให้มีโอกาสได้หน่วยการสร้างที่มีความหลากหลายขึ้น นอกจากนั้นยังช่วยเพิ่มโอกาสให้ตัวดี ๆ มีโอกาสถูกเลือกเพิ่มขึ้นเนื่องจากมีคุณภาพมากกว่าตัวอื่น หรือกล่าวได้ว่าเมื่อมีหน่วยการสร้างที่หลากหลายมากขึ้นก็จะมีโอกาสได้ประชากรที่มีคุณภาพดีขึ้นซึ่งเมื่อเข้าสู่ขั้นตอนวิธีการเลือกสรร ประชากรที่มีคุณภาพดีเหล่านี้จะมีโอกาสถูกเลือกสูงกว่าประชากรตัวอื่น ๆ ทำให้โอกาสที่จะไปถึงคำตอบที่ต้องการมีสูงขึ้นด้วย

บทเรียนที่สำคัญของงานวิจัยด้าน GAs คือเราไม่สามารถที่จะรู้ได้ว่าควรเข้ารหัสปัญหาในรูปแบบใด และการเข้ารหัสแบบนั้น ๆ มีหน่วยการสร้างเป็นอย่างไร และจะมีความเหมาะสมกับตัวดำเนินการต่าง ๆ เช่น การไขว้เปลี่ยน หรือไม่ แต่ถ้าสามารถหาทางระบุหน่วยการสร้างได้ จะมีโอกาสทำให้สามารถออกแบบหรือเลือกตัวดำเนินการประกอบหน่วยการสร้างที่มีความเหมาะสมมากขึ้นได้ แต่ก็ยังไม่สามารถรับประกันได้ว่าจะสามารถแก้ปัญหายากได้อย่างรวดเร็ว แม่นยำ และมีความน่าเชื่อถือสูง

3.5 ความยากของปัญหาในมุมมองของหน่วยการสร้าง

จากมุมมองการแบ่งปัญหาขนาดใหญ่ออกเป็นส่วนประกอบย่อย ๆ ทำให้เราสามารถอธิบายกลไกการทำงานของ GAs ได้ด้วยสมมติฐานหน่วยการสร้าง ในกรณีที่เป็นปัญหายากจากมุมมองของหน่วยการสร้างเราสามารถจำแนกคุณลักษณะความยากออกได้ดังนี้

1. ความยากภายในหน่วยการสร้าง เป็นความยากที่เกิดจากการลวง (Deceptive) (Goldberg, 1987b) ในการให้ค่าความเหมาะสมโดยจะให้ค่าความเหมาะสมของโครงสร้างย่อยภายในหน่วยการสร้างที่มีหน้าตาเป็นไปในทิศทางตรงกันข้ามกับคำตอบที่มีค่ามากกว่า และจะให้หน่วยการสร้างที่สมบูรณ์มีค่าความเหมาะสมมากที่สุด ดังนั้นการไปให้ถึงคำตอบต้องหาหน่วยการสร้างได้ครบสมบูรณ์ทั้งก้อน ซึ่งเห็นได้ว่าความยากขึ้นกับขนาดของหน่วยการสร้างโดยถ้ายังมีขนาดใหญ่จะมีความยากมากขึ้น และระดับของการลวงว่ามีความรุนแรงระดับใดโดยพิจารณาจากการให้ค่าความเหมาะสมกับรูปแบบที่ตรงข้ามกับคำตอบว่ามีค่ามากหรือน้อยเพียงใด
2. ความยากระหว่างหน่วยการสร้าง เป็นความยากที่เกิดจากการที่หน่วยการสร้างแต่ละตัวมีค่าความเหมาะสมไม่เท่ากัน ทำให้บางตัวมีความสำคัญมากกว่าตัวอื่นเป็นเหตุให้ตัวที่มีความสำคัญน้อยมีโอกาสเกิดได้ยากขึ้น เพราะอาจถูกมองว่าเป็นสัญญาณรบกวน (Noise) ในระบบซึ่งจะทำให้ถูกระบบเพิกเฉยหรือไม่ได้ให้ความสำคัญ ถ้าค่าความเหมาะสมมีความแตกต่างกันมากจะทำให้ความยากเพิ่มขึ้น ในกรณีที่พิจารณาหน่วยการสร้างมากกว่าหนึ่งตัวร่วมกันอาจทำให้เกิดความยากในภาพรวมเพิ่มขึ้น ได้แก่ ความยากที่เกิดจากการพิจารณาค่าความเหมาะสมรวมกันระหว่างหน่วยการสร้างมากกว่าหนึ่งตัว เช่นปัญหาการให้ค่าความเหมาะสมแบบลำดับชั้น (Hierarchy) เป็นต้น ความยากที่เกิดจากการทำโทษผลเฉลยทั้งตัวทำให้ค่าความเหมาะสมลดลง ก็จะส่งผลให้หน่วยการสร้างที่มีรูปแบบเหมือนกันในผลเฉลยที่ถูกทำโทษกับไม่ถูกทำโทษมีค่าไม่เท่ากัน และทำให้หน่วยการสร้างที่ดีตัวอื่น ๆ ที่อยู่ผลเฉลยที่ถูกทำโทษเสียหายไปด้วยซึ่งในกรณีนี้

สามารถจัดอยู่ในรูปแบบการลวงได้ และความยากที่เกิดจากความไม่แน่นอนของตัวฟังก์ชันการหาค่าเหมาะสมเอง เช่นฟังก์ชันการหาค่าความเหมาะสมมีการให้ค่าवलบสลับกันไปมาตลอดเวลา เช่นฟังก์ชันไซน์ เป็นต้น ในกรณีนี้ใกล้เคียงกับการมีสัญญาณรบกวนเนื่องจากมีความเปลี่ยนแปลงมากเกินไป

3. ความยากเป็นพิเศษของหน่วยการสร้าง เกิดจากการที่มีสัญญาณรบกวนในระหว่างการดำเนินการซึ่งอาจมีผลทำให้ค่าของคำตอบเปลี่ยนแปลงได้ เช่น คำตอบในช่วงเวลาหนึ่งอาจจะเป็นค่า 10 พอมีสัญญาณรบกวนคำตอบจะเปลี่ยนเป็นค่า 12 หรือค่าอื่น ๆ เป็นต้น ทำให้เกิดความไม่แน่นอนตลอดเวลา การหาคำตอบจะยากขึ้นและอาจไม่ได้คำตอบที่แท้จริงเนื่องจากถูกสัญญาณรบกวนบิดเบือนค่าความเหมาะสมที่แท้จริงให้เปลี่ยนไป

จากที่ได้กล่าวมาสามารถที่จะสรุปภาพรวมแบบจำลองของหน่วยการสร้างออกได้เป็น 8 กลุ่ม ดังนี้ (มีการกล่าวถึงรายละเอียดเพิ่มเติมในบทที่ 11)

1. หน่วยการสร้างที่มีการเชื่อมโยงแบบแนบแน่น (Tightly Linkage BBs)
2. หน่วยการสร้างที่มีการเชื่อมโยงแบบหลวม (Loosely Linkage BBs)
3. หน่วยการสร้างที่มีขนาดแปรผัน (Vary Size BBs)
4. หน่วยการสร้างที่มีการเหลื่อมกัน (Overlapping BBs)
5. หน่วยการสร้างที่มีโครงสร้างความสัมพันธ์แบบลำดับชั้น (Hierarchical BBs)
6. หน่วยการสร้างแต่ละหน่วยมีค่าความเหมาะสมไม่เท่ากัน (Vary Fitness BBs)
7. หน่วยการสร้างที่มีสัญญาณรบกวนร่วมด้วย (Noisy BBs)
8. หน่วยการสร้างแบบผสมที่มีรูปแบบ 1. ถึง 7. รวมกันมากกว่า 1 รูปแบบ (Mixed BBs)

3.6 การออกแบบขั้นตอนวิธีเชิงพันธุกรรมในมุมมองของหน่วยการสร้าง

จากงานวิจัย (Goldberg, 1991 ; Goldberg, Deb and Clark, 1992) ได้ให้หลักการ 6 ข้อในการออกแบบขั้นตอนวิธีเชิงพันธุกรรมที่ใช้ในการแก้ปัญหาได้อย่างมีประสิทธิภาพ โดยมีแนวทางดังต่อไปนี้

1. รู้ว่าขั้นตอนวิธีเชิงพันธุกรรมประมวลผลหน่วยการสร้าง
2. ทำให้มีข้อมูลเกี่ยวกับหน่วยการสร้างเพียงพอในตอนเริ่มต้นหรือในช่วงเวลาหนึ่ง
3. ทำให้หน่วยการสร้างที่จำเป็นมีปริมาณมากขึ้น หรือมีขนาดใหญ่ขึ้น
4. ทำให้หน่วยการสร้างที่จำเป็นมีการผสมหรือรวมเข้าด้วยกัน

5. ใช้กับปัญหาที่หน่วยการสร้างไม่ซับซ้อนจนเกินไป หรือสามารถเข้ารหัสใหม่ให้เหมาะสมได้
6. มีการตัดสินใจในการเลือกที่ดีท่ามกลางการแข่งขันกันของหน่วยการสร้าง

จากแนวทาง 6 ข้อที่ได้กล่าวข้างต้นทำให้เราสามารถกล่าวได้ว่าการจัดการหน่วยการสร้างเป็นสิ่งสำคัญและจำเป็น เพราะมีผลโดยตรงกับประสิทธิภาพของขั้นตอนวิธีเชิงพันธุกรรมที่ได้ออกแบบไว้ เราสามารถอธิบายหลักการทั้งหมดภายใต้ภาพรวมของกระบวนการทำงานของขั้นตอนวิธีเชิงพันธุกรรมดังต่อไปนี้ เริ่มต้นสำหรับปัญหาใด ๆ เราจำเป็นที่จะต้องแปลงปัญหาให้อยู่ในรูปที่เราสามารถคำนวณได้ นั่นก็คือการแปลงปัญหาให้กลายเป็นโครโมโซมหรือสายอักขระเสียก่อน ขั้นตอนนี้มีความสำคัญมากเนื่องจากเป็นตัวกำหนดรูปแบบความสัมพันธ์หรือการกระจายตัวของหน่วยการสร้าง (การเข้ารหัสไม่ว่ารูปแบบใดถ้าการเข้ารหัสนั้นยังสามารถแสดงถึงคำตอบที่ต้องการได้ เราเชื่อว่าจะมีความสัมพันธ์ของโครงสร้างย่อยในการเข้ารหัสแบบนั้นเสมอ) หลังจากที่เข้ารหัสปัญหาแล้ว ขั้นตอนถัดไปจะเริ่มเข้าสู่กระบวนการทำงานของขั้นตอนวิธีเชิงพันธุกรรมคือการสร้างโครโมโซมขึ้นมาจำนวนหนึ่ง (ประชากร) ซึ่งขนาดของจำนวนประชากรที่ใช้ นี้จะมีค่าคงที่ตลอดการดำเนินการ ในส่วนนี้เองที่เป็นตัวกำหนดข้อมูลเบื้องต้นให้ระบบใช้สำหรับเรียนรู้ เนื่องจากเราไม่รู้มาก่อนว่าข้อมูลที่ตีควรมีหน้าตาเป็นอย่างไร ข้อมูลที่ได้เริ่มต้นจึงเป็นข้อมูลที่ได้จากการสุ่มสร้างขึ้นในรูปแบบการแจกแจงรูปแบบใดรูปแบบหนึ่ง ดังนั้นจึงเกิดปัญหาว่าเราจะรู้ได้อย่างไรว่าข้อมูลที่ได้เพียงพอต่อการเรียนรู้ของระบบ เพราะฉะนั้นจึงต้องทำให้แน่ใจได้ว่าขนาดของประชากรที่ใช้มีปริมาณมากพอ ซึ่งค่าขนาดของประชากรนี้จะเป็นตัวแปรหลักตัวแปรหนึ่งในการออกแบบเสมอ จากนั้นประชากรแต่ละตัวจะถูกประเมินว่ามีความเหมาะสมเพียงใด (เป็นการประเมินคุณค่าของแต่ละโครโมโซมหรือผลเฉลยนั่นเอง) จากนั้นจะเป็นกระบวนการเลือกสรรที่มีหน้าที่คัดเลือกประชากรที่มีลักษณะที่ต้องการ (ลักษณะที่ว่านี้โดยทั่วไปเรามองจากค่าความเหมาะสมของประชากรเพียงอย่างเดียว) โดยปกติประชากรตัวที่มีค่าความเหมาะสมสูงจะมีโอกาสถูกเลือกเพิ่มขึ้นแบบเป็นสัดส่วนต่อกันเมื่อเทียบกับประชากรตัวที่มีค่าความเหมาะสมต่ำกว่า เรามีความเชื่อว่าหน่วยการสร้างที่ดีมีอยู่ในประชากรที่ดี หรือมีค่าความเหมาะสมที่สูง ซึ่งถ้าเลือกได้ถูกต้องก็จะได้หน่วยการสร้างที่ดี หรือเป็นที่ต้องการในการสร้างประชากรรุ่นใหม่ที่มีคุณภาพดีกว่าเดิม แต่ปัญหาอยู่ตรงที่ว่าเราจะสามารถผสมหรือประกอบหน่วยการสร้างที่อยู่ในโครโมโซมที่ดีเหล่านี้เข้าด้วยกันได้อย่างไร จึงจะมีความเหมาะสมและทำให้ได้หน่วยการสร้างใหม่ที่โตขึ้นหรือมีคุณภาพมากขึ้น สำหรับขั้นตอนนี้โดยพื้นฐานจะอาศัยตัวดำเนินการไขว้เปลี่ยนเป็นตัวจัดการซึ่งจากผลการทดลองในหลายงานวิจัยให้ผลเป็นที่น่าพอใจ แต่ยังคงมีการปรับปรุงรูปแบบและวิธีการจัดการในขั้นตอนนี้เพื่อให้ได้ประสิทธิภาพที่สูงขึ้นอยู่เสมอ ขั้นตอนสุดท้ายก่อนที่จะได้ประชากรรุ่น

ใหม่คือการทำให้ประชากรมีความหลากหลายเพิ่มขึ้นเล็กน้อย (ขั้นตอนนี้อาจจะมีหรือไม่มีก็ได้) ข้อดีของขั้นตอนนี้คือ จะทำให้มั่นใจได้ว่าไม่ว่าอย่างไรก็ยังมีโอกาสพบคำตอบที่ต้องการได้ แต่ไม่รู้ว่าจะต้องใช้เวลาขนาดไหนเท่าใด หลายคนมองว่าขั้นตอนนี้เปรียบเสมือนขั้นตอนวิธีแบบสุ่มที่ถูกใส่เข้าไปเพื่อให้เกิดความหลากหลายของคำตอบอยู่เสมอ ขั้นตอนนี้จะมีความสำคัญมากขึ้นในกรณีที่ขนาดประชากรที่ใช้มีขนาดเล็กเกินไป เพราะขั้นตอนนี้จะช่วยเพิ่มความหลากหลายหรือความเป็นไปได้ (ทำให้รูปแบบหน้าตาของโครโมโซมมีความแตกต่างกันเพิ่มขึ้น) โดยปกติแล้วจะรักษาระดับความหลากหลายนี้ไว้เพียงเล็กน้อยเพื่อไม่ให้ระบบวิวัฒนาการโดยรวมมากเกินไป หลังจากขั้นตอนนี้จะเริ่มมีการทำซ้ำดังรูปที่ 2.1 ในบทที่ 2

ข้อสังเกตเพิ่มเติมที่สำคัญอย่างหนึ่งก็คือ เราไม่รู้ว่าความสัมพันธ์ของโครงสร้างย่อยมีรูปแบบเป็นอย่างไรหลังจากการเข้ารหัสแล้ว ตัวที่ใช้ชี้วัดได้ทางอ้อมคือประสิทธิภาพหรือผลที่ได้เมื่อทำการทดลองแล้ว ถ้าได้ผลที่ดีหรือมีประสิทธิภาพสูงจะเป็นการแสดงว่ารูปแบบการจัดการหน่วยการสร้างของขั้นตอนวิธีที่ออกแบบไว้มีความเหมาะสม หรือเข้ากันได้ดีกับรูปแบบความสัมพันธ์ของหน่วยการสร้างที่มีอยู่ในโครโมโซม

3.7 การแก้ปัญหาด้วยขั้นตอนวิธีเชิงพันธุกรรม

มีหลายงานตีพิมพ์ที่ได้แสดงให้เห็นถึงความสำเร็จของ GAs ในการแก้ปัญหาที่หลากหลายนับได้จริงแต่ก็ยังคงมีกรณีที่ GAs ให้ผลที่ย่ำแย่ สมมติปัญหาขึ้นมาปัญหาหนึ่งเราจะรู้ได้อย่างไรว่า GAs เป็นวิธีที่ดีในการแก้ปัญหานี้ ถึงปัจจุบันก็ยังไม่คำตอบที่แน่นอนและชัดเจน แต่นักวิจัยในสายงานนี้หลายคนได้ให้ข้อคิดเห็นไว้ดังนี้ ถ้าเป็นปัญหาที่มีปริภูมิค้นหาขนาดใหญ่และมีความรู้เกี่ยวกับปริภูมิค้นหานั้นน้อย หรือปริภูมิค้นหาที่มีความซับซ้อน เข้าใจได้ยากและไม่ได้มีคำตอบเดียว หรือค่าฟังก์ชันความเหมาะสมมีสัญญาณรบกวน หรือเป็นงานที่ไม่ได้ต้องการคำตอบที่ดีที่สุด เช่น ต้องการให้ได้คำตอบที่น่าพอใจระดับหนึ่งในระยะเวลาอันจำกัด เป็นต้น ถ้าเข้าข่ายคุณลักษณะเหล่านี้ GAs จะมีโอกาสแข่งขันกับวิธีอื่น ๆ หรือให้ผลดีกว่าวิธีที่ไม่ได้ใช้ความรู้เกี่ยวกับโดเมนของปัญหาในกระบวนการค้นหา ถ้าปริภูมิค้นหาไม่ได้มีขนาดใหญ่สามารถใช้วิธีค้นหาแบบละเอียดถี่ถ้วน (Exhaustive Search) ซึ่งทำให้มั่นใจได้ว่าจะมีโอกาสพบคำตอบที่ดีที่สุดได้ ในขณะที่ GAs อาจลู่เข้าสู่ค่าที่เหมาะสมที่สุดท้องถิ่น (Local Optimum) แทนที่จะเจอค่าเหมาะสมที่สุดแบบครอบคลุม (คำตอบที่ต้องการ) ถ้าปริภูมิค้นหาไม่มีความสม่ำเสมอราบเรียบ (Smooth) หรือมีแค่ฐานนิยมเดียว (Unimodal) ขั้นตอนวิธีจำพวกการปีนเกรเดียนต์ (Gradient-ascent Algorithms) เช่น การปีนเขาแบบสูงชันที่สุด (Steepest-ascent Hill Climbing) จะให้ประสิทธิภาพที่ดีกว่า GAs ถ้ามีความรู้ความเข้าใจเกี่ยวกับปริภูมิค้นหาเป็นอย่างดี

ดี เช่นปัญหาที่รู้จักกันคืออย่างปัญหา TSP เป็นต้น วิธีการค้นหาสามารถใช้ความรู้เฉพาะด้านของโดเมน หรือใช้การศึกษาลำบาก (Heuristic) มาช่วยในการออกแบบเพื่อให้ได้ผลดีกว่าขั้นตอนวิธีที่ใช้โดยทั่วไปเช่น GAs ถ้าฟังก์ชันความเหมาะสมมีสัญญาณรบกวน งานที่เกี่ยวกับการวัดที่มีแนวโน้มความผิดพลาด และระบบการมองเห็นของหุ่นยนต์ เป็นต้น ขั้นตอนวิธีคู่แข่งอย่างเช่นขั้นตอนวิธีการปีนเขาอย่างง่าย (Simple Hill Climbing) อาจไม่สามารถหาค่าได้ถูกต้องเนื่องจากสัญญาณรบกวน แต่ GAs มีหลักการการทำงานโดยอาศัยการสะสมสถิติของค่าความเหมาะสมในหลายรุ่นการดำเนินการ ซึ่งมองได้ว่าเป็นกระบวนการที่ทนทานต่อสัญญาณรบกวนระดับหนึ่ง อย่างไรก็ตามปัจจัยที่กล่าวถึงเหล่านี้ยังไม่สามารถที่จะใช้บอกได้ว่า GAs มีประสิทธิภาพในกระบวนการค้นหาที่ดีกว่าเมื่อเปรียบเทียบกับวิธีอื่น เนื่องจากสมรรถนะของ GAs ขึ้นอยู่กับรายละเอียดเป็นอย่างมาก เช่น วิธีการที่ใช้ในการเข้ารหัสปัญหา ตัวดำเนินการที่เลือกใช้ค่าพารามิเตอร์ของแต่ละตัวดำเนินการ และเกณฑ์การให้คะแนน เป็นต้น งานทางด้านทฤษฎีของ GAs ที่กล่าวถึงก่อนหน้านี้ไม่ได้ให้ข้อสรุปเกี่ยวกับเรื่องเหล่านี้

ประชาคม GA ได้ใช้ทฤษฎีเค้าร่างในการนำเสนอคำถามว่าอะไรเป็นสิ่งที่ทำให้ปัญหาที่มีความยากสำหรับ GA จากคำถามดังกล่าวได้มีคำอธิบายว่าการแข่งขันกันระหว่างเค้าร่างจากเค้าร่างลำดับต่ำในช่วงเริ่มต้นไปจนถึงเค้าร่างที่มีลำดับสูงขึ้นไปในช่วงเวลาต่อมาเป็นความยากสำหรับ GA เบธเก (Bethke, 1980) ได้ให้เหตุผลว่าจะเป็นความยากสำหรับ GA ในการหาค่าที่เหมาะสมที่สุด ถ้าส่วนกันลำดับต่ำมีข้อมูลที่เขาไปผิดทางต่างจากส่วนกันลำดับสูง ตัวอย่างต่อไปนี้จะทำให้เห็นภาพได้ชัดเจนขึ้น สมมติให้เค้าร่าง H เป็นผู้ชนะ (Winner) ถ้าค่าความเหมาะสมเฉลี่ยมีค่าสูงที่สุดในส่วนกันนั้น และสมมติให้เค้าร่างที่ประกอบด้วย 0 เป็นผู้ชนะยกเว้นกรณีที่เค้าร่างที่ยาว L ประกอบด้วย 0 ทั้งหมด $000\dots 0$ มีค่าน้อยกว่า เค้าร่างที่ประกอบด้วย 1 ทั้งหมด $111\dots 1$ (เป็นผู้ชนะ) ในทางปฏิบัติเป็นการยากที่ GA จะสามารถเจอ $111\dots 1$ เนื่องจากเค้าร่างที่มีลำดับต่ำกว่าจะเป็นผู้ชนะและลงให้ไปคนละทางกับคำตอบ ฟังก์ชันค่าความเหมาะสมแบบนี้จะถูกเรียกว่าการลวงอย่างเต็มที่ (Fully Deceptive) (คำว่า การลวง (Deceptive) ถูกนิยามอย่างเป็นทางการโดยโกลด์เบิร์ก (Goldberg, 1987) เป็นคนแรก) เบธเกใช้ Walsh Transforms ในการออกแบบฟังก์ชันความเหมาะสมที่มีระดับการลวงที่หลากหลาย และได้มีการวิจารณ์ถึงฟังก์ชันนี้ในงานวิจัย (Goldberg, 1989b, 1989c; Forrest and Mitchell, 1993a) ในส่วนงานวิจัยย่อยของเบธเก โกลด์เบิร์กและผู้ร่วมงานได้ศึกษาทฤษฎีของการลวงในฟังก์ชันความเหมาะสม และทำให้การลวงได้กลายมาเป็นจุดศูนย์กลางของงานทางด้านทฤษฎีของ GAs (Das and Whitley, 1991; Deb and Goldberg, 1993; Goldberg, 1989c; Leipins and Voes, 1990, 1991; Whitley, 1991)

การศึกษาเกี่ยวกับการลวงใน GA มีความเกี่ยวข้องกับฟังก์ชันการหาค่าที่เหมาะสมที่สุด โดยการลวงในฟังก์ชันความเหมาะสมถือว่าการทำให้เกิดความยากในการหาค่าที่เหมาะสมที่สุดครอบคลุม (Global Optimum) สำหรับการนำ GAs ไปใช้แก้ปัญหาอย่างกว้างขวางนั้น การลวงถือเป็นปัจจัยหนึ่งที่สำคัญและควรทำความเข้าใจ อย่างไรก็ตามถ้ามีการมองว่า GAs ไม่ใช่เครื่องมือที่ใช้หาค่าที่เหมาะสมที่สุด แต่เป็นเครื่องมือที่ช่วยหาค่าที่ยอมรับได้ เมื่อนั้นการลวงจะมีความเกี่ยวข้องหรือมีความสำคัญน้อยลง ใงานวิจัยของเดอจอง (De Jong, 1993) ได้เขียนไว้ว่า “GA is attempting to maximize cumulative payoff from arbitrary landscapes, deceptive or otherwise. In General, this is achieved by not investing too much effort in finding cleverly hidden peaks (the risk/reward ratio is too high)” GA พยายามที่จะทำให้เกิดการชดเชยสะสมสูงสุดจากภูมิทัศน์ใด ๆ การหลอก หรือสิ่งอื่น ๆ โดยทั่วไปแล้วสิ่งนี้สามารถได้มาจากการหาค่าจุดสูงสุดที่ซ่อนอยู่อย่างชาญฉลาด โดยไม่จำเป็นต้องลงทุนมากเกินไป (อัตราความเสี่ยงต่อผลตอบแทนมีค่าสูง) ซึ่งสามารถตีความได้ว่า GA ดีเพียงพอสำหรับปัญหาที่ต้องการคำตอบที่ยอมรับได้ซึ่งอาจจะไม่ใช่คำตอบที่ดีที่สุด โดยไม่ต้องใช้การคำนวณมากเกินไป

นอกจากนี้เรายังสามารถพิจารณาความยากในมุมมองทั่วไปได้ดังนี้ ถ้าพิจารณาจากปัญหาเป็นหลัก จากงานวิจัย (Jones and Forrest, 1995) ได้นำเสนอการวัดความยากของปัญหาสำหรับขั้นตอนวิธีเชิงพันธุกรรม โดยใช้ชื่อวิธีวัดว่า สหสัมพันธ์ระยะความเหมาะสม (Fitness Distance Correlation: FDC) โดยที่ค่า FDC ยิ่งมีค่าสูงแปลว่าเป็นปัญหายาก วิธีวัดนี้อยู่บนสมมติฐานที่ว่าถ้าหน้าตาของโครโมโซมที่มีค่าความเหมาะสมสูง ๆ มีหน้าตาใกล้เคียงกับคำตอบที่ต้องการ จะถือว่าเป็นปัญหาง่าย แต่ในทางกลับกันถ้าหน้าตาของโครโมโซมที่มีค่าความเหมาะสมสูง ๆ มีหน้าตาที่แตกต่างกันมากและแตกต่างจากคำตอบที่ต้องการมาก จะถือว่าเป็นปัญหายาก การคำนวณพื้นฐานอาศัยค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าความเหมาะสมกับความแตกต่างทางด้านหน้าตาของโครโมโซม ซึ่งในงานวิจัยนี้วัดโดยใช้ระยะทางแบบแฮมมิง (Hamming Distance) และได้ทดสอบกับปัญหาที่รู้ค่าดีที่สุดแบบครอบคลุม จุดอ่อนของ FDC คือเราต้องคำนวณค่าความเหมาะสมของโครโมโซมจำนวนมากซึ่งยิ่งถ้าปัญหามีขนาดใหญ่ขึ้นเท่าไร ปริภูมิค้นหาจะมีขนาดโตขึ้นเป็นฟังก์ชันเลขชี้กำลัง ทำให้เราไม่สามารถคำนวณหาค่า FDC ได้ และมิงานวิจัย (Altenberg, 1997) ได้วิเคราะห์และให้คำแนะนำกรณีที่คำนวณกับ FDC โดยใช้กรณีที่ค่าสหสัมพันธ์ระยะความเหมาะสมมีค่าเป็นศูนย์เป็นตัวอย่าง และยังได้แสดงให้เห็นถึงความสัมพันธ์ในเชิงลึกที่มีอยู่ระหว่างระยะทางแฮมมิงกับกระบวนการประกอบเข้าด้วยกันใหม่

ถ้าเราพิจารณาจากขั้นตอนวิธีที่ได้ออกแบบไว้เป็นหลัก จากงานวิจัย (Wolpert and Macready, 1995, 1997) ได้นำเสนอทฤษฎีบทที่ชื่อว่า ไม่มีอาหารกลางวันฟรี (No Free Lunch

Theorem: NFL) ซึ่งมีเนื้อหาหลักว่า “Each deterministic non-repeating Black Box search algorithm performs on average equally well.” แต่ละขั้นตอนวิธีที่ค้นหาแบบกล่องดำแบบไม่ทวนซ้ำเชิงกำหนดโดยเฉลี่ยแล้วมีประสิทธิภาพเท่ากัน หรือกล่าวได้ว่าไม่มีขั้นตอนวิธีที่ดีที่สุดสำหรับทุกปัญหา นอกจากนี้ยังมีอีกหลายงานวิจัยที่ศึกษาเกี่ยวกับ NFL และขั้นตอนวิธีเชิงพันธุกรรม (Droste, Jansen and Wegener, 1999 ; Whitley, 2000 ; Schumacher, Vose and Whitley, 2001 ; Christensen and Oppacher, 2001) ทำให้เราสรุปได้ว่าความยากของปัญหาขึ้นกับขั้นตอนวิธีที่ออกแบบ

ในมุมมองที่พิจารณาร่วมกันระหว่างปัญหาและขั้นตอนวิธีที่ใช้แก้ปัญหา จะรู้ได้อย่างไรว่าขั้นตอนวิธีที่เราออกแบบนั้นจะเหมาะสำหรับแก้ปัญหาแบบใด ซึ่งอาจทำได้หลายวิธี เช่น ลองทดสอบกับปัญหาโดยตรง ถ้าใช้วิธีนี้จะต้องทดสอบกับปัญหาจำนวนมากระดับหนึ่งถึงจะทำให้เชื่อถือได้ เป็นต้น วิธีหนึ่งที่สามารถทำให้เราบอกได้อย่างชัดเจนและมีความน่าเชื่อถือสูงคือการวิเคราะห์ขั้นตอนวิธีที่ได้ออกแบบไว้ว่าการดำเนินการต่าง ๆ มีลักษณะหรือคุณสมบัติเป็นเช่นใด การวิเคราะห์ทำได้หลายรูปแบบ แต่จะต้องมีความชัดเจนเพียงพอที่สรุปให้เห็นจริงได้ แม้ว่าวิธีนี้จะดูเหมือนยากโดยเฉพาะอย่างยิ่งกับขั้นตอนวิธีที่มีลักษณะแบบพลวัต แต่ในหลายกรณีก็สามารถหามุมมองการนำเสนอที่มีความซับซ้อนน้อยลงได้และอาจจะตามมาด้วยข้อจำกัดจำนวนมาก อย่างไรก็ตามการวิเคราะห์ไม่ว่าเชิงคุณภาพหรือปริมาณ ก็ยังช่วยให้เราสามารถเข้าใจขั้นตอนวิธีที่ออกแบบได้ดีขึ้น หรือมีความน่าเชื่อถือเพิ่มมากขึ้น

3.8 ทฤษฎีบทเค้รำงและการตีความในมุมมองของตัวดำเนินการ

นอกจากมุมมองการตีความทฤษฎีบทเค้รำงที่เกี่ยวข้องกับหน่วยการสร้างแล้ว ยังมีงานวิจัย (Thierens, 1999) ที่กล่าวถึงทฤษฎีบทเค้รำงในมุมมองของการดำเนินการจัดการ จากหัวข้อ 3.1.4 แบบฟอร์มทั่วไปของทฤษฎีบทเค้รำงสามารถเขียนในรูปแบบ

$$E[m(H, t + 1)] \geq m(H, t) \alpha(H, t) \{1 - \beta(H, t)\} \quad (3.5)$$

เมื่อ $\alpha(H, t)$ คืออัตราส่วนการสืบพันธุ์ และ $\beta(H, t)$ คือปัจจัยความแตกแยก และ $m(H, t)$ คือจำนวนของโครโมโซมที่เข้ากันได้กับเค้รำง H ที่เวลา t จากสมการนี้เราสามารถกล่าวได้ว่าหน่วยการสร้างสามารถเติบโตได้ถ้าอัตราการสร้างที่เกิดจากการเลือกสรรมีมากกว่าอัตราการถูกทำลายโดยตัวดำเนินการทางพันธุกรรมเช่น ตัวดำเนินการประกอบเข้าด้วยกันใหม่ หรือกล่าวได้ว่า α เป็นอัตราการเติบโตภายใต้การเลือกสรรเพียงอย่างเดียว และ β เป็นค่าประมาณความน่าจะเป็นที่เค้รำงจะถูกทำลายโดยตัวดำเนินการทางพันธุกรรม จากพจน์ที่คูณกันอยู่เราสามารถสร้าง

เงื่อนไขได้สองข้อดังนี้ ถ้าปัจจัยการแตกแยก β มีค่าสูง เราสามารถเพิ่มค่าของพจน์ความกดดันในการเลือกสรร α เพื่อชดเชยการทำลายที่เกิดขึ้นได้ และ ปัจจัยความแตกแยก β สามารถควบคุมได้ผ่านค่าพารามิเตอร์ของตัวดำเนินการทำลาย ดังนั้นจากเงื่อนไขข้อแรกจะเห็นได้ว่าการมีความกดดันในการเลือกสูงไม่จำเป็นต้องทำให้เกิดการลู่เข้า เนื่องจากเราสามารถเพิ่มระดับความแตกแยกได้ และการเพิ่มความกดดันในการเลือกเป็นการทำให้มั่นใจได้ว่าหน่วยการสร้างจะเติบโตขึ้นโดยไม่จำเป็นต้องมีความรู้เกี่ยวกับการเชื่อมโยง ทำให้สามารถมองข้ามเรื่องของการเชื่อมโยงแบบแนบแน่นที่เป็นปัจจัยสำคัญในการดำเนินการกับหน่วยการสร้างของขั้นตอนวิธีเชิงพันธุกรรมได้ และจากเงื่อนไขที่สองเมื่อเจอปัญหาหลง การใช้ตัวดำเนินการไขว้เปลี่ยนทำให้เกิดความแตกแยกต่ำจะช่วยรักษาการเชื่อมโยงแบบแนบแน่นไม่ให้ถูกทำลายได้ง่าย เนื่องจากเชื่อว่าหน่วยการสร้างที่ดีเกิดขึ้นได้ยากดังนั้นจึงควรรักษาเอาไว้

อย่างไรก็ดีการเพิ่มระดับความกดดันในการเลือก α หรือการลดระดับความแตกแยก β ล้วนแล้วแต่มีค่าใช้จ่ายที่ต้องพิจารณาถึงทั้งสิ้น เช่น ถ้าเราเลือกที่จะใช้ความกดดันในการเลือกที่สูงเพื่อชดเชยความสูญเสียที่เกิดจากความแตกแยกจะทำให้ความหลากหลายของประชากรลดลงอย่างรวดเร็ว และยังทำให้การแลกเปลี่ยนหน่วยการสร้างลดลงด้วย เนื่องจากประชากรส่วนใหญ่มีหน้าตาที่เหมือนกัน ทำให้เกิดการลู่เข้าในที่สุด และถ้าความแตกแยกมีค่าน้อยเกินไป การแลกเปลี่ยนก็จะมีอัตราการลดลงด้วย ทำให้ไม่เกิดรูปแบบโครงสร้างที่หลากหลาย และเป็นเหตุให้สูญเสียโอกาสที่จะได้หน่วยการสร้างที่ดีขึ้น ซึ่งประเด็นเหล่านี้ชี้ให้เห็นว่าในทางปฏิบัติสำหรับขั้นตอนวิธีเชิงพันธุกรรม การเติบโตของหน่วยการสร้างและการผสมเข้าด้วยกันนั้นเป็นคนละส่วนกัน ทฤษฎีบทเค้าร่างไม่ได้พูดถึงเรื่องการแลกเปลี่ยนหน่วยการสร้างหรือกล่าวได้ว่า แม้ว่าจะมีหน่วยการสร้างอยู่เป็นจำนวนมากในประชากรแต่ก็ไม่ได้เป็นการรับประกันว่าจะทำให้ได้คำตอบที่ดีขึ้น เว้นเสียแต่ว่าจะสามารถประกอบหน่วยการสร้างที่ดีที่มีอยู่ในหมู่ประชากรเข้าด้วยกันได้อย่างถูกต้อง

การเพิ่มแต่ความกดดันในการเลือกสรรมากเกินไปไม่ใช่สิ่งที่ดี แต่ในงานวิจัยส่วนใหญ่ยอมรับว่าสำหรับปัญหาที่ยากขึ้นเราควรจะใช้ค่าความกดดันในการเลือกสรรที่สูงขึ้น เนื่องจากเราต้องการหน่วยการสร้างที่มีคุณภาพที่ดีขึ้น เพื่อช่วยให้สามารถไปถึงคำตอบได้อย่างมีประสิทธิภาพ โดยทั่วไปเรากำหนดให้ค่านี้เป็นค่าคงที่ซึ่งส่งผลให้อัตราการลู่เข้ามีค่าคงที่ตลอดทุกรุ่นประชากร ขั้นตอนวิธีการเลือกสรรส่วนใหญ่ที่นิยมใช้กันเกือบทั้งหมดเป็นวิธีการเลือกแบบสัดส่วน หรือการเลือกเชิงอันดับซึ่งเมื่อมีความกดดันในการเลือกสูง จะให้ผลในทิศทางเดียวกันหมดคือ จะมีแต่ประชากรที่มีค่าความเหมาะสมสูงเท่านั้นที่ถูกเลือก และยังมีระดับความกดดันในการเลือกสูงมากขึ้นเท่าไรผลของการเลือกจะยิ่งใกล้เคียงกันมากขึ้นเท่านั้น นอกจากนั้นการเลือกสรรแบบปกติจะ

พิจารณาผ่านค่าความเหมาะสมเพียงอย่างเดียวเท่านั้น ซึ่งในกรณีที่เป็นปัญหาลวงจะทำให้เกิดข้อบกพร่องในการเลือกได้ง่าย เนื่องจากโครโมโซมตัวที่มีโครงสร้างลวงมักมีค่าความเหมาะสมสูงกว่าโดยเฉลี่ยทำให้โครโมโซมเหล่านี้มีโอกาสถูกเลือกสูง ในกรณีเช่นนี้ถือว่าเป็นจุดอ่อนของวิธีการเลือกอย่างชัดเจน

กระบวนการประกอบเข้าด้วยกันใหม่เป็นกระบวนการที่เกิดขึ้นหลังจากกระบวนการเลือกสรรได้เสร็จสิ้นลงไปแล้ว ดังนั้นจะเห็นได้ว่าสองกระบวนการนี้ขึ้นแก่กันหรือส่งผลกระทบต่อกันโดยตรงดังที่กล่าวไปแล้วเบื้องต้น โดยทั่วไปแล้วจะกำหนดค่าพารามิเตอร์ของตัวดำเนินการประกอบเข้าด้วยกันเป็นค่าคงที่ตลอดทุกรุ่นของประชากรด้วยเช่นกัน ซึ่งทำให้อัตราการแลกเปลี่ยนหรือการผสมเกิดขึ้นคงที่โดยเฉลี่ย แต่ในภาพรวมนั้นการรักษาสมดุลระหว่างการสร้างรวมถึงการรักษา และการทำลายนั้นเป็นสิ่งที่จำเป็น ดังนั้นการที่กำหนดค่าความกดดันการเลือกและค่าอัตราการแลกเปลี่ยนเป็นค่าคงที่ตลอดการดำเนินการ อาจทำให้ประสิทธิภาพในการแก้ปัญหาลดลง และเป็นการยากที่จะรู้ว่าแต่ละพารามิเตอร์ควรมีค่าเท่าไรจึงจะเหมาะสมกัน และความเหมาะสมนั้นจะมีค่าคงที่ตลอดการดำเนินการหรือไม่ ถ้าเราสามารถออกแบบตัวดำเนินการที่มีความสามารถในการปรับตัวได้เองอัตโนมัติก็น่าจะเป็นตัวช่วยที่ดีในการรักษาสมดุล การจัดการน้อยสุดคือยอมให้ค่าพารามิเตอร์ของกระบวนการใดกระบวนการหนึ่งเป็นค่าคงที่ และออกแบบให้ค่าพารามิเตอร์ของอีกกระบวนการหนึ่งสามารถปรับตัวได้เองอัตโนมัติ ปัญหาอยู่ที่ว่าเราควรจะเลือกให้กระบวนการไหนมีความสามารถในการปรับตัว เมื่อพิจารณาข้อจำกัดของกระบวนการเลือกสรรที่ขึ้นกับค่าความเหมาะสมเพียงอย่างเดียวแล้วจะเห็นได้ว่า กระบวนการประกอบเข้าด้วยกันใหม่นั้นมีข้อจำกัดน้อยกว่า และในมุมมองที่ว่าขั้นตอนการประกอบเข้าด้วยกันใหม่นั้นเป็นตัวการสำคัญในการแลกเปลี่ยนหรือผสมส่วนประกอบที่ดี ๆ เข้าด้วยกันแม้ว่าประชากรมีโครงสร้างที่ดีอยู่ไม่มาก แต่ถ้ากระบวนการประกอบเข้าด้วยกันใหม่มีความสามารถหรือมีประสิทธิภาพสูงเพียงพอ เราเชื่อว่าจะสามารถประกอบข้อมูลโครงสร้างที่ดีเหล่านี้ให้มีคุณภาพดีขึ้นและมีขนาดใหญ่ขึ้นได้ ด้วยเหตุนี้จึงพิจารณาได้ว่าปัจจัยการประกอบเข้าด้วยกันใหม่นี้มีความสำคัญ และเป็นประเด็นในการพัฒนามากกว่าการการเลือกสรร ลองพิจารณาในมุมมองเพิ่มเติมต่อไปนี้

กระบวนการประกอบเข้าด้วยกันใหม่เป็นการดำเนินการพื้นฐานในขั้นตอนวิธีเชิงพันธุกรรมที่เกิดขึ้นจากตัวดำเนินการไขว้เปลี่ยน แม้ว่าจะมีการนำเสนอการไขว้เปลี่ยนไว้หลายวิธี แต่ทว่าทุกวิธีก็มีหน้าที่ย่อยเหมือนกันคือพยายามประกอบส่วนที่ดี ๆ เข้าด้วยกัน ลองพิจารณาถึงความเป็นไปได้ 2 ประการจากการไขว้เปลี่ยนดังนี้

- 1) ความเป็นไปได้ที่หน่วยการสร้างจะถูกสร้างขึ้น หรือปรากฏให้เห็นที่บางตำแหน่ง
กรณี que เห็นได้ชัดสุดเช่น โครโมโซมพอสันธุ์และแม่พันธุ์มีส่วนที่ต่ออยู่ตัวละครึ่งหนึ่ง
ของความยาวโครโมโซมจะเห็นได้ว่าการไขว้เปลี่ยนมีโอกาสช่วยให้เกิดหน่วยการ
สร้างใหม่ที่ดีขึ้นได้อย่างชัดเจน ในกรณีที่เกิดการลวงในฟังก์ชันค่าความ
เหมาะสมหรือในการเข้ารหัส เช่นปัญหาเกี่ยวกับโครงสร้างของโครโมโซมตัวที่มีค่า
ความเหมาะสมสูงจะมีองค์ประกอบที่ต่างจากหน่วยการสร้างที่แท้จริง และยังมี
การลวงขนาดใหญ่ขึ้นจำนวนบิตที่มีหน้าตาเหมือนหน่วยการสร้างที่ถูกต้องยังมี
ปริมาณลดลง ทำให้เกิดความเข้าใจผิดว่าหน้าตาของบิตที่แตกต่างจากคำตอบ
เหล่านี้เป็นหน่วยการสร้าง เป็นต้น
- 2) ความเป็นไปได้ที่หน่วยการสร้างจะถูกรวมเข้าด้วยกันหรือผสมกัน โดยทั่วไปแล้ว
แม้ว่าการไขว้เปลี่ยนมีหน้าที่ผสมหน่วยการสร้างเข้าด้วยกันเพื่อให้ได้หน่วยการ
สร้างใหม่ที่ดีขึ้น แต่การที่จะส่งต่อผลเหล่านี้ไปยังรุ่นประชากรถัดไปถูกกระทำ
ผ่านขั้นตอนการเลือกสรร ซึ่งถ้าขั้นตอนนี้ยังคงรักษาปริมาณและความ
หลากหลายของหน่วยการสร้างไว้ได้เป็นอย่างดี จะช่วยให้มีโอกาสพบคำตอบที่ดี
ขึ้นได้ การเกิดหน่วยการสร้างนั้นถือว่าเป็นความโชคดีในความบังเอิญ โดยจะเห็น
ได้ชัดเมื่อพิจารณาพร้อมกับการลวงที่เลือกร่างที่มีลำดับต่ำกว่าหน่วยการสร้าง จะถูก
ลวงให้ไกลออกจําตอบที่แท้จริง และในขั้นตอนการดำเนินการไม่มีข้อมูลอย่างอื่น
มาช่วยในการตัดสินใจ ทำให้ไม่รู้ว่หน่วยการสร้างที่ถูกต้องมีหน้าตาเป็นอย่างไร

สิ่งที่ต้องการจากตัวดำเนินการประกอบเข้าด้วยกันใหม่มีอยู่สองประการที่ขัดแย้งกัน
คือ ต้องการลดผลกระทบจากความแตกแยกของการไขว้เปลี่ยนที่เกิดขึ้นกับหน่วยการสร้าง และ
ต้องการที่จะทำให้เกิดการแลกเปลี่ยนหน่วยการสร้างในอัตราที่สูงที่สุด จากสิ่งเหล่านี้ทำให้มองได้
ว่าการเลือกสรรหน่วยการสร้างที่ดี (จากการบวนการเลือกสรร) และการประกอบหน่วยการสร้าง
เข้าด้วยกันเป็นสิ่งที่ตรงข้ามกัน

จากที่กล่าวมาสามารถสรุปเป็นประเด็นได้ว่าขั้นตอนหลักที่สำคัญในขั้นตอนวิธีเชิง
พันธุกรรมคือ การเลือกสรรและการประกอบเข้าด้วยกันใหม่ โดยทั้งสองขั้นตอนนี้มีความสัมพันธ์
กัน และเป็นารยากที่จะแสดงให้เห็นถึงการเชื่อมโยงระหว่างความสัมพันธ์นี้ได้อย่างตรงไปตรงมา
งานวิจัยส่วนใหญ่มุ่งเน้นไปที่การศึกษากระบวนการประกอบเข้าด้วยกันมากกว่าวิธีการเลือก

บทที่ 4

หน่วยการสร้างแบบชิ้นส่วน และวิธีการทดสอบ

ในบทนี้เนื้อหาหลักจะแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนแรกเป็นนิยามหน่วยการสร้างแบบชิ้นส่วน และส่วนที่สองเป็นวิธีการทดสอบและการวัดผลการทดสอบในมุมมองที่น่าเสนอ

4.1 หน่วยการสร้างแบบชิ้นส่วน

ดังที่ได้กล่าวมาก่อนหน้านี้แล้วว่าจากการศึกษากลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรมนั้น ทำให้เราได้สมมติฐานหน่วยการสร้างซึ่งใช้ในการอธิบายภาพรวมของการทำงานที่อยู่เบื้องหลังความสำเร็จ ด้วยเหตุนี้การศึกษาหน่วยการสร้างไม่ว่าจะเป็นเชิงคุณภาพหรือปริมาณ จะช่วยให้สามารถเข้าใจหลักการทำงานของขั้นตอนวิธีเชิงพันธุกรรมได้ชัดเจนขึ้น โดยหลักการพื้นฐานแล้วขั้นตอนวิธีเชิงพันธุกรรมพยายามค้นหาหน่วยการสร้างที่เหมาะสมและนำมาประกอบเข้าด้วยกันเพื่อให้ได้คำตอบที่ดีขึ้น แต่ปัญหาก็คือหลักการทำงานดังกล่าวถูกอธิบายในรูปแบบของเค้าร่างซึ่งมีความซับซ้อนและทำความเข้าใจได้ยาก ดังนั้นงานวิจัยนี้จึงได้นำเสนอมุมมองและนิยามหน่วยการสร้างแบบใหม่และให้ชื่อว่า “ชิ้นส่วน” (Fragment) เพื่อลดความซับซ้อนทางโครงสร้างของหน่วยการสร้างแบบเดิม และช่วยให้สามารถทำความเข้าใจหรือเห็นภาพกลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรม รวมถึงผลของตัวดำเนินการได้ชัดเจนมากขึ้น โดยมุมมองใหม่ที่นำเสนอสามารถนำไปประยุกต์ใช้ได้หลากหลายวิธีทั้งในส่วนของกระบวนการและการประกอบ ซึ่งสามารถทำได้ง่ายและตรงไปตรงมา โดยหลักการทั้งหมดที่ใช้ได้กับชิ้นส่วนจะสามารถใช้ได้กับหน่วยการสร้างแบบเดิมได้ เนื่องจากชิ้นส่วนในมุมมองที่น่าเสนอนี้เป็นเซตย่อยของโครงสร้างหน่วยการสร้างแบบเดิมนั่นเอง

เราสามารถนิยามชิ้นส่วนด้วยโครงสร้างย่อยของโครโมโซมใด ๆ กำหนดให้ลำดับของโครโมโซมความยาว l

$$C^k = c_1^k c_2^k \dots c_l^k : k \in I^+, c_n^k \in \{0,1\}, 1 \leq n \leq l \quad (4.1)$$

กำหนดให้เซตของดัชนีตำแหน่งอ้างอิงในโครโมโซมความยาว l จากตำแหน่งที่ f ถึงตำแหน่งที่ t

$$(f_i, t_i) = \{ z_i : 1 \leq f_i \leq z_i \leq t_i \leq l, f_i, t_i, z_i \in I^+ \} \quad (4.2)$$

and $\forall i, j \in I^+ : i \neq j, (f_i, t_i) \cap (f_j, t_j) = \emptyset$

กำหนดให้ลำดับย่อยของโครโมโซม C^{k_i} เป็น

$$F_i^k = c_{f_i}^k c_{f_i+1}^k \dots c_{t_i}^k \quad (4.3)$$

เซตของลำดับย่อยที่เรียงติดกันในโครโมโซมนิยามดังนี้

$$S^k = \{F_i^k : i \leq l\} \quad (4.4)$$

เราเรียกลำดับย่อย F ที่ติดกันในโครโมโซมว่า “ชิ้นส่วน” (Fragment)

กำหนดให้เค้าร่างของโครโมโซมยาว l

$$H^k = h_1^k h_2^k \dots h_l^k : k \in I^+, h_n^k \in \{0,1,*\}, * \in \{0,1\}, 1 \leq n \leq l \quad (4.5)$$

จากเค้าร่างเราสามารถนิยามชิ้นส่วนใหม่ได้เป็น

$$F_i^k = h_{f_i}^k h_{f_i+1}^k \dots h_{t_i}^k : h_{z_i}^k \neq * \quad (4.6)$$

สมมติให้เค้าร่างของโครโมโซมยาว 10 บิตเป็น 1*110***01 ดังนั้นจะได้ว่าเค้าร่างนี้ประกอบไปด้วยชิ้นส่วนจำนวน 3 ชิ้น คือ 1, 110 และ 01 หรือกล่าวได้ว่าชิ้นส่วนในเค้าร่างคือลำดับย่อยที่มีค่าเป็น 0 หรือ 1 ที่อยู่ติดกันโดยไม่มี * แทรกอยู่ ดังรูปที่ 4.1

Bit Position :	1	2	3	4	5	6	7	8	9	10
Schema :	1	*	1	1	0	*	*	*	0	1
Fragment :	F1			F2			F3			

รูปที่ 4.1 การระบุชิ้นส่วนจากเค้าร่าง

จากนิยามและตัวอย่างเห็นได้ว่าเค้าร่างที่ซับซ้อนถูกแบ่งออกเป็นชิ้นส่วนย่อย ๆ ที่ไม่มีความสัมพันธ์กัน (เป็นอิสระจากกัน) ซึ่งทำให้เกิดความหลากหลายในการพิจารณาเพิ่มขึ้นด้วย

4.2 ชิ้นส่วนในมุมมองของหน่วยการสร้าง

จากหัวข้อ 4.1 จะเห็นได้ว่าเค้าร่างใด ๆ สามารถแบ่งออกเป็นชิ้นส่วนย่อย ๆ ได้ หรือกล่าวในทางกลับกันได้ว่า ชิ้นส่วนย่อย ๆ สามารถประกอบเป็นเค้าร่างใด ๆ ได้เสมอ ในมุมมองนี้สามารถแย่งแยกเค้าร่างออกเป็นหน่วยย่อยที่อยู่ติดกันที่เล็กที่สุดได้ หลักการแย่งแยกนี้ยังคงมีความสอดคล้องกับสมมติฐานหน่วยการสร้างเนื่องจากชิ้นส่วนที่ได้มีลักษณะเป็นเค้าร่างที่สั้น และมีลำดับต่ำ ข้อดีของการแย่งแยกแบบนี้คือช่วยลดความซับซ้อนของเค้าร่างได้ เนื่องจากเค้าร่างแบบเดิมจะประสบปัญหาในการประกอบเข้าด้วยกันกับเค้าร่างอื่นซึ่งทำได้ยาก มุมมองของการแย่งแยกให้เค้าร่างมีขนาดเล็กลงจะช่วยทำให้การประกอบเค้าร่างหรือหน่วยการสร้างทำได้ง่ายขึ้น มีความหลากหลายเพิ่มขึ้นเพราะสามารถประกอบเข้าด้วยกันได้หลายรูปแบบ แม้จะถูกมองว่าแต่

ละชิ้นส่วนไม่มีความสัมพันธ์กันทั้ง ๆ ที่มาจากเค้าร่างเดียวกัน แต่วิธีการแบ่งแบบนี้จะช่วยรักษา ส่วนประกอบย่อยของเค้าร่างไม่ให้ถูกทำลายในขั้นตอนการรวมเข้าด้วยกัน หรือมีความเสียหาย น้อยที่สุด ในขณะที่ทำให้เกิดรูปแบบความเป็นไปได้มากที่สุดพร้อม ๆ กันด้วย อย่างไรก็ตามปัจจัย เหล่านี้ขึ้นอยู่กับวิธีการนำไปประยุกต์ใช้ เนื่องจากหลักการแบ่งแยกที่นำเสนอเป็นเพียงหลักการ พื้นฐานอย่างง่ายที่เข้ากันกับสมมติฐานหน่วยการสร้าง ซึ่งเราเชื่อว่าเป็นหลักการที่จะช่วยให้เข้าใจ กลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรมได้ดีขึ้น

4.3 ชิ้นส่วนในมุมมองของการเรียนรู้การเชื่อมโยง

หน่วยเล็กที่สุดของเค้าร่างและชิ้นส่วนคือหนึ่งคู่จีน (หรือหนึ่งบิตในกรณีที่เข้ารหัสแบบ เลขฐานสอง) ซึ่งเป็นกรณีเดียวที่ไม่มี การเชื่อมโยงเกิดขึ้น แต่ถ้ามีจำนวนมากกว่าหนึ่งคู่จีน เรา สามารถพิจารณาได้ว่าคู่จีนเหล่านั้นมีการเชื่อมโยงกัน การเชื่อมโยงมีหลายระดับและมีทั้งการ เชื่อมโยงที่ดีและไม่ดี ขึ้นกับมุมมองที่ใช้ในการตัดสินใจ โดยทั่วไปเราเชื่อว่าเค้าร่างที่ดีน่าจะมีการ เชื่อมโยงที่ดี การเกาะกลุ่มกันของคู่จีนเป็นตัวแสดงถึงระดับของการเชื่อมโยงว่ามีความแนบแน่น เพียงใด (ยิ่งอยู่ใกล้กันมากยิ่งมีการเชื่อมโยงที่แนบแน่นขึ้น) และถ้าเค้าร่างหลายตัวมีรูปแบบคู่จีน ที่เหมือนกันที่ตำแหน่งเดียวกัน เราเชื่อว่ารูปแบบนั้นจะแสดงถึงระดับความน่าเชื่อถือของการ เชื่อมโยง ซึ่งปัจจัยทั้งหลายเหล่านี้ขึ้นอยู่กับ การเข้ารหัสปัญหาเป็นหลัก (โดยทั่วไปเราไม่รู้ว่า การเข้ารหัสแบบใดถึงจะมีความเหมาะสมกับปัญหา มากที่สุด) ในกรณีของชิ้นส่วน เราสามารถ พิจารณาได้ว่ามีการเชื่อมโยงกันแบบแนบแน่นมากที่สุดรูปแบบหนึ่ง เนื่องจากในชิ้นส่วนเดียวกัน แต่ละคู่จีนจะต้องอยู่ติดกันเสมอ ขั้นตอนวิธีเชิงพันธุกรรมใน ส่วนของการไขว้เปลี่ยนจะมีการแยก โครโมโซมออกเป็น ส่วน ๆ (ตามจำนวนจุดที่เกิดการไขว้เปลี่ยน) และรวมเข้าด้วยกันใหม่ โดยจุดที่ เกิดการไขว้เปลี่ยนได้มาจากวิธีการสุ่มแบบไม่มีความลำเอียง หรือกล่าวได้ว่าไม่มีการใช้ความรู้ หรือกฎใด ๆ เพื่อที่จะหาจุดไขว้เปลี่ยน แต่สิ่งที่เราสังเกตได้คือเค้าร่างที่มีขนาดเล็ก และมีลำดับต่ำ หรือหน่วยการสร้างขนาดเล็กที่มีการเชื่อมโยงกันแบบแนบแน่นจะมีโอกาสอยู่รอดสูงกว่าหน่วยการ สร้างที่มีการเชื่อมโยงกันแบบหลวม ดังนั้นสามารถตีความได้ว่าหน่วยการสร้างที่มีขนาดเล็ก เหล่านี้มีความสำคัญที่จะทำให้ได้ผลเฉลยที่ดีขึ้น หรือนำไปสู่คำตอบได้ ซึ่งสอดคล้องกับแนวคิด เรื่องชิ้นส่วนที่เรานิยามขึ้น สำหรับปัญหาที่มีลักษณะการเชื่อมโยงของหน่วยการสร้างแบบหลวม อาจถูกมองว่าน่าจะเป็นปัญหาสำหรับขั้นตอนวิธีเชิงพันธุกรรม แต่จากศึกษาพฤติกรรมของ ขั้นตอนวิธีเชิงพันธุกรรม (Goldberg, 1989) เราพบว่าในการดำเนินการจะเกิดการประมวลเค้าร่าง ที่มีความหลากหลายพร้อม ๆ กันไปแบบขนาน และมีอัตราการเติบโตของเค้าร่างที่ดีสูงกว่าอัตรา การทำลาย ดังนั้นจึงไม่อาจสรุปได้ว่าปัญหาที่มีการเชื่อมโยงแบบหลวมเป็นปัญหาที่มีความยาก สำหรับขั้นตอนวิธีเชิงพันธุกรรม

หลักการแบ่งแยกแบบขั้นส่วนสามารถเข้าร่วมกับวิธีการเรียนรู้การเชื่อมโยงได้ โดยการเรียนรู้ว่าคู่เงินไหนควรจะอยู่ด้วยกันในขั้นส่วนเดียวกัน รวมถึงสามารถเรียนรู้ได้ว่าขั้นส่วนใดมีความสัมพันธ์กันบ้างซึ่งจะช่วยให้สามารถหาการเชื่อมโยงแบบหลวมได้ จะเห็นได้ว่าแนวความคิดเรื่องขั้นส่วนมีความง่าย ชัดเจน และสามารถนำไปประยุกต์ใช้ได้หลายวิธี รวมถึงสามารถประยุกต์ใช้กับวิธีที่มีอยู่แล้วได้ด้วย

4.4 วิธีการทดสอบขั้นตอนวิธีที่นำเสนอและการวัดผล

ทุกขั้นตอนวิธีที่นำเสนอในงานวิจัยนี้เราจะใช้ปัญหา Royal Road (Mitchell, Forrest and Holland, 1991) และปัญหา Trap-5 (G. R. Harik, 1997) เป็นปัญหาทดสอบเพื่อใช้วิเคราะห์กลไกการทำงานของขั้นตอนวิธีที่นำเสนอ การทดลองมุ่งเน้นไปที่การศึกษาระบบการระบุและประกอบหน่วยการสร้างที่เกิดขึ้นในรูปแบบขั้นส่วนเป็นหลัก ซึ่งปัญหาทดสอบทั้งสองปัญหาออกแบบมาเพื่อใช้ทดสอบหน่วยการสร้างโดยมีความแตกต่างกันตรงที่ปัญหา Royal Road ไม่ใช่ปัญหาลวงและมีขนาดหน่วยการสร้างเท่ากับ 8 บิต ส่วนปัญหา Trap-5 เป็นปัญหาลวงที่มีขนาดหน่วยการสร้างเท่ากับ 5 บิต (มีขนาดเล็กกว่า) การทดสอบกับทั้งสองปัญหาเป็นการแสดงถึงความสามารถในการจัดการกับหน่วยการสร้างที่มีรูปแบบแตกต่างกัน การทดลองสำหรับแต่ละขั้นตอนวิธีกับแต่ละปัญหาจะทำซ้ำจำนวน 30 ครั้ง และนำค่าผลที่ได้มาเฉลี่ยกัน ผลการทดลองจะแบ่งออกได้เป็นสองกรณีคือ สามารถหาคำตอบได้ติดกันครบทั้ง 30 ครั้ง และไม่สามารถหาคำตอบได้ติดกันครบ 30 ครั้ง เหตุที่ต้องแบ่งผลออกเป็นสองกรณีเนื่องจากเราต้องการเปรียบเทียบให้เห็นถึงความแตกต่างอย่างชัดเจนระหว่างลักษณะและพฤติกรรมที่มีผลต่อการหาคำตอบ การทดลองมีเงื่อนไขการทำงานดังต่อไปนี้

1. ทดสอบกับปัญหา Royal Road ขนาด 64 บิต (มีหน่วยการสร้างทั้งหมด 8 หน่วย) และทดสอบกับปัญหา Trap-5 ขนาด 60 บิต (มีหน่วยการสร้างทั้งหมด 12 หน่วย)
2. จะไม่ใช้ตัวดำเนินการผ่าเหล่ากับทุกปัญหาและทุกขั้นตอนวิธีที่นำเสนอ
3. ในการดำเนินการกับปัญหาแต่ละรอบ กำหนดให้สามารถเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมได้ไม่เกิน 1,000,000 (หนึ่งล้าน) ครั้ง
4. กำหนดให้ใช้จำนวนประชากรได้สูงสุดไม่เกิน 10,000 ตัว และถ้าประชากรทุกตัวมีความเหมาะสมเท่ากันหมดจะถือว่าสิ้นสุดการดำเนินการ สำหรับการทดสอบกรณีที่ทำคำตอบไม่พบจะแสดงผลการทดลองถึงรุ่นประชากรที่ 100 เท่านั้น

5. ใช้จำนวนประชากรน้อยที่สุด (โดยการประมาณ) เพื่อให้สามารถหาคำตอบได้ติดกันครบ 30 รอบการทดสอบ และใช้จำนวนประชากรมากที่สุด (โดยการประมาณ) โดยที่ไม่สามารถหาคำตอบได้ติดกันครบ 30 รอบ
6. วิธีที่มีการใช้หน่วยเก็บถาวรจะทำการเลือกสรรแบบตัวดีที่สุดเป็นจำนวน n ตัวไม่ซ้ำกัน (n -Best Selection) วิธีนี้จะทำการเรียงลำดับค่าความเหมาะสมของประชากร แล้วเลือกตัวที่ได้ค่าความเหมาะสมดีที่สุดมาจำนวนหนึ่ง (n ตัว) โดยไม่สนใจว่าจะมีหน้าตาซ้ำกันหรือไม่ ซึ่งในงานวิจัยนี้จะเรียกประชากรกลุ่มที่ถูกเลือกเหล่านี้ว่า ประชากรย่อย (Sub-population: sub.) ในการแก้ปัญหาเราจะทำการเพิ่มค่าประชากรย่อยจนกว่าจะสามารถแก้ปัญหาได้ แต่จะมีขนาดสูงสุดไม่เกิน 50% ของขนาดประชากร หรือ $n \leq (0.5) \cdot Pop$ เมื่อ Pop หมายถึง ขนาดประชากร เนื่องจากสมมติฐานที่ตั้งไว้ตั้งแต่ต้นว่าเค้าร่างที่ดีจะอยู่ในประชากรที่ดี (ที่ควรถูกเลือก) ซึ่งก็คือ ประชากรที่มีค่าความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ย หรือประชากรครั้งที่มีค่าความเหมาะสมสูงกว่ำนั่นเอง

ในส่วนผลการทดลองสำหรับแต่ละมุมมองในเรื่องการระบุหน่วยการสร้างและการประกอบหน่วยการสร้างจะแสดงรายละเอียดแบ่งออกเป็น 5 ส่วนดังต่อไปนี้

1. ประสิทธิภาพในการแก้ปัญหา
 - 1.1 จำนวนครั้งของการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม (Fitness Evaluations / Function Evaluations : FE) สำหรับแต่ละปัญหาทดสอบ
2. ค่าตัวแปรที่สำคัญ
 - 2.1 ขนาดประชากรที่ใช้ในการแก้ปัญหา
 - 2.2 ขนาดของประชากรย่อย¹ (Sub-population)
3. หน่วยการสร้าง²
 - 3.1 ปริมาณหน่วยการสร้างกับรุ่นของประชากร โดยแบ่งรูปแบบหน่วยการสร้างได้ดังนี้
 - 3.1.1 หน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุด (Pure BBs)
 - 3.1.2 หน่วยการสร้างที่ตรงกันข้ามกับคำตอบที่ดีที่สุด (Non BBs)
 - 3.1.3 หน่วยการสร้างที่มีบางส่วนเหมือนกับคำตอบที่ดีที่สุด (Mixed BBs)

¹ หมายถึง เป็นตัวแปรที่ใช้ในขั้นตอนวิธีที่มีการใช้หน่วยเก็บถาวร

² หมายถึง เป็นปัญหาที่ออกแบบมาเพื่อทดสอบหน่วยการสร้างเท่านั้น

3.2 จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุด (Pure BBs) โดยแบ่งออกเป็น จำนวนมากที่สุด จำนวนน้อยที่สุด และจำนวนเฉลี่ย กับรุ่นของประชากร หมายเหตุ ขนาดและรูปแบบของหน่วยการสร้างขึ้นกับนิยามในแต่ละปัญหาทดสอบ

4. ชิ้นส่วน

4.1 ขนาดของชิ้นส่วนเฉลี่ย กับรุ่นของประชากร

4.2 ปริมาณชิ้นส่วนทั้งหมดที่ระบุได้ กับรุ่นของประชากร

4.3 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวร กับรุ่นของประชากร³

หมายเหตุ ในแต่ละขั้นตอนวิธีจะใช้หลักการการระบุชิ้นส่วนไม่เหมือนกัน แต่ชิ้นส่วนที่ได้ จะมีลักษณะตรงตามนิยามเหมือนกันหมด

5. โครงสร้างจีโนไทป์

5.1 รูปแบบจีโนไทป์ของประชากร (แสดงให้เห็นจากประชากรที่เลือกมาจำนวนหนึ่ง) กับ รุ่นของประชากร

5.2 รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวร (เฉพาะวิธีที่ใช้หน่วยเก็บถาวร) กับ รุ่นของประชากร

ส่วนที่ 1 และ 2 จะแสดงรวมกันในรูปแบบตารางเพื่อให้เห็นความสัมพันธ์ได้ง่าย ส่วน ที่ 3 กับ 4 จะแสดงในรูปแบบของกราฟเพื่อให้เห็นพฤติกรรมการปรับตัวระหว่างการดำเนินการ และส่วนที่ 5 จะแสดงในรูปแบบรูปภาพเพื่อแสดงให้เห็นลักษณะการเปลี่ยนแปลงทางโครงสร้าง ของโครโมโซมในระหว่างการดำเนินการ รายละเอียดของผลการทดลองในแต่ละส่วนมีนัย ดังต่อไปนี้

1. ในงานวิจัยนี้เรากำหนดมาตรวัดประสิทธิภาพของขั้นตอนวิธีที่นำเสนอโดยใช้การนับค่า จำนวนครั้งในการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม (FEs) ถ้าจำนวน FEs มีค่าน้อยแปลว่ามีประสิทธิภาพสูง แต่ถ้า FEs มีค่ามากแปลว่ามีประสิทธิภาพที่ต่ำ โดยค่า FEs นี้จะใช้ในการเปรียบเทียบประสิทธิภาพระหว่างขั้นตอนวิธีที่ต่างกันในปัญหาเดียวกัน
2. แต่ละขั้นตอนวิธีจะมีการออกแบบที่ต่างกันทั้งในแง่ของตัวดำเนินการและตัวแปรที่ใช้ อย่างไรก็ตามขั้นตอนวิธีเหล่านี้ก็ยังคงมีส่วนที่เหมือนกันอยู่สองส่วนหลัก ๆ คือตัวดำเนินการเลือกสรร และตัวแปรที่เป็นจำนวนประชากร (ขนาดของประชากร) ในส่วนผล

³ หมายถึง เป็นปัญหาที่มีการใช้หน่วยเก็บถาวร และจะมีค่าน้อยกว่าหรือเท่ากับปริมาณชิ้นส่วนทั้งหมดที่หาได้เสมอ เนื่องจากในหน่วยเก็บถาวรจะไม่มีการเก็บชิ้นส่วนที่มีรูปแบบเหมือนกันและเกิดขึ้นซ้ำที่ตำแหน่งเดียวกัน

การทดลองจะเน้นไปที่ตัวแปรที่เป็นจำนวนประชากร เนื่องจากมีบทบาทสำคัญอย่างยิ่งในกระบวนการวิวัฒนาการ (ขาดไม่ได้) และเป็นตัวแปรที่ไม่ขึ้นอยู่กับตัวดำเนินการ ซึ่งต่างจากตัวดำเนินการเลือกสรรที่มีรายละเอียดต่างกันในแต่ละวิธี จำนวนประชากรเป็นตัวบ่งบอกถึงระดับความหลากหลาย และทรัพยากรที่ต้องใช้ในการหาคำตอบ ซึ่งส่งผลกระทบต่อโดยตรงกับประสิทธิภาพในการแก้ปัญหา

3. งานวิจัยนี้สนใจศึกษาเกี่ยวกับความสัมพันธ์ระหว่างกลไกในการแก้ปัญหา กับหน่วยการสร้างเป็นหลัก ดังนั้นจึงมีการวัดปริมาณของหน่วยการสร้าง (จำนวนหน่วยการสร้างที่มีอยู่ในประชากรทั้งหมด) และจำนวนของหน่วยการสร้างในโครโมโซมในแต่ละรุ่นประชากร เพื่อสังเกตลักษณะการเปลี่ยนแปลง (การเพิ่มขึ้นหรือลดลง) ว่ามีรูปแบบเป็นอย่างไรในแต่ละขั้นตอนวิธีที่น่าเสนอกับปัญหาทดสอบที่ต่างกัน ผลการทดลองจะช่วยทำให้เราเข้าใจพฤติกรรมการทำงานในมุมมองของหน่วยการสร้างได้ดีขึ้น
4. ได้มีการนิยามมุมมองหน่วยการสร้างในรูปแบบใหม่โดยให้ชื่อว่า ชิ้นส่วน (ซึ่งยังคงสอดคล้องกับนิยามหน่วยการสร้างแบบเดิม) เพื่อให้สามารถทำความเข้าใจกับหน่วยการสร้างได้ดีขึ้น เพราะอยู่ในรูปแบบที่ง่ายขึ้น (มีความซับซ้อนน้อยลง) การตรวจจับหรือระบุชิ้นส่วนสามารถทำได้หลายวิธี แต่ละวิธีจะได้ชิ้นส่วนที่มีรูปแบบต่างกัน ทำให้มีประสิทธิภาพในการค้นหาคำตอบไม่เท่ากัน (ในปัญหาเดียวกัน) และมีความเข้ากันได้กับลักษณะปัญหาที่ต่างกัน การทดลองจะมีการวัดขนาดของชิ้นส่วนโดยเฉลี่ย และจำนวนของชิ้นส่วนที่มีอยู่ทั้งหมดในแต่ละรุ่นประชากร เพื่อให้เห็นถึงพฤติกรรมการทำงานของขั้นตอนวิธีที่น่าเสนอกับปัญหาทดสอบในอีกมุมมองหนึ่งที่ต่างจากหน่วยการสร้างรูปแบบเดิม ผลที่ได้จะทำให้เราเห็นกลไกการทำงานในอีกมุมมองหนึ่งซึ่งช่วยส่งเสริมให้มีความเข้าใจในความรู้เดิม (ที่เกี่ยวกับหน่วยการสร้าง) ได้ดียิ่งขึ้น
5. โดยปกติขั้นตอนวิธีการเลือกสรรจะพิจารณาแต่เฉพาะค่าความเหมาะสมซึ่งใช้เป็นตัวแทนคุณภาพของโครโมโซม ในปัญหาลงเราไม่สามารถเชื่อค่าความเหมาะสมได้ทั้งหมด แม้กระนั้นก็ยังมีความสามารถหาคำตอบได้จึงเป็นสิ่งที่น่าสนใจว่าผลที่ได้จากการดำเนินการที่เหลือนอกจากการเลือกสรรให้ผลอย่างไรกับรูปแบบโครงสร้างทางพันธุกรรมของเหล่าประชากร ดังนั้นการแสดงให้เห็นถึงโครงสร้างรูปแบบคู่จิ้นในโครโมโซมแต่ละตัวในภาพรวมจะช่วยให้เราเห็นภาพได้ชัดเจนมากขึ้น วิธีหนึ่งที่จะช่วยทำให้เห็นภาพเหล่านี้ได้คือการแสดงโครงสร้างจีโนม (ค่าแต่ละบิตในโครโมโซม) ในรูปแบบของรูปภาพโดยเลือกโครโมโซมจำนวนหนึ่งในกลุ่มประชากรมาเรียงต่อกัน (ซ้อนกัน) ใน

แนวตั้งแล้วทำการกำหนดสีให้กับค่าของคู่จีนแต่ละค่า (มีค่าเป็น 0 กับ 1 สองค่าเท่านั้น) เพื่อให้เห็นรูปแบบการกระจายตัวของค่า 1 และค่า 0 ว่ามีลักษณะอย่างไรในแต่ละรุ่นของประชากร และเมื่อพิจารณาภาพเหล่านี้จากประชากรหลายรุ่น จะทำให้เราเห็นพฤติกรรมในการเปลี่ยนแปลง (การวิวัฒนาการ) ของประชากรในการหาคำตอบว่าเป็นอย่างไร (จะเห็นความแตกต่างในกรณีที่ทำคำตอบพบ และหาคำตอบไม่พบ) ซึ่งเป็นสิ่งที่สะท้อนให้เห็นถึงลักษณะเฉพาะของขั้นตอนวิธีที่นำเสนอด้วย นอกจากการแสดงรูปแบบโครงสร้างของประชากรแล้ว เราจะใช้วิธีเดียวกันนี้ในการแสดงรูปแบบโครงสร้างของชิ้นส่วนที่ถูกเก็บอยู่ในหน่วยเก็บถาวรโดยจะมีรูปแบบในการแสดงผลต่างออกไป โดยเราจะแสดงชิ้นส่วนเรียงตามตำแหน่ง (ขึ้นกับตำแหน่งแรกของชิ้นส่วนเป็นหลัก) ในแนวนอน และแนวตั้งจะแสดงชิ้นส่วนที่มีทั้งหมดในแต่ละตำแหน่ง การนำเสนอแบบนี้จะทำให้เห็นการวิวัฒนาการของชิ้นส่วนว่าเป็นไปในทิศทางและรูปแบบใด ซึ่งจะช่วยให้สามารถเข้าใจขั้นตอนวิธีที่นำเสนอได้ดียิ่งขึ้น ข้อมูลที่ใช้แสดงทั้งหมดจะถูกเลือกมาจากผลการทดลองรอบหนึ่งจากการทดลองซ้ำจำนวน 30 รอบ (เลือกรูปแบบข้อมูลที่ใกล้เคียงกับผลการทดลองส่วนใหญ่มากที่สุดเพื่อใช้เป็นตัวแทนในภาพรวม)

4.4.1 อักษรย่อในส่วนของตารางและรูปภาพที่ใช้แสดงผลการทดลอง

อักษรย่อของตารางและเส้นกราฟในส่วนรูปภาพมีความหมายดังต่อไปนี้

- #pop. คือ จำนวนประชากร ที่ใช้ในการแก้ปัญหา(Population)
- #FEs. คือ จำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม
- #sub. คือ จำนวนประชากรย่อยที่ถูกเลือก (Sub-population)
- (Pure / Mixed / Non) BBs คือหน่วยการสร้างที่เป็นคำตอบ หน่วยการสร้างที่มีส่วนประกอบของคำตอบ และหน่วยการสร้างที่ตรงข้ามกับคำตอบ
- (Max / Min / avg) Pure BBs คือจำนวนของหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดที่ปรากฏในหนึ่งโครโมโซมว่ามีจำนวนมากที่สุด น้อยที่สุด และเฉลี่ยเป็นเท่าใด
- afs: average fragment size หรือขนาดโดยเฉลี่ยของชิ้นส่วน (ความยาวของชิ้นส่วน)
- nof: number of fragment หรือปริมาณชิ้นส่วน (จำนวนชิ้นส่วนที่ระบุได้ทั้งหมดในประชากร)

- acfs: average common fragment size หรือขนาดโดยเฉลี่ยของชิ้นส่วนที่ร่วมกัน⁴ (ชิ้นส่วนที่เหมือนกันระหว่าง 2 โครโมโซมที่ตำแหน่งเดียวกัน หรือเป็นชิ้นส่วนที่ได้จากการบีบอัดแบบ LZW)
- adfs: average different fragment size หรือขนาดโดยเฉลี่ยของชิ้นส่วนที่ไม่ร่วมกัน⁴ (ชิ้นส่วนที่ต่างกันระหว่าง 2 โครโมโซมที่ตำแหน่งเดียวกัน)
- nocf: number of common fragment หรือปริมาณชิ้นส่วนที่ร่วมกัน⁴
- nodf: number of different fragment หรือปริมาณชิ้นส่วนที่ไม่ร่วมกัน⁴
- ancp: average number of crossed pieces หรือจำนวนชิ้นส่วนโดยเฉลี่ยที่ถูกไขว้เปลี่ยนระหว่างคู่โครโมโซม⁵
- soa: size of archive หรือขนาดของหน่วยเก็บข้อมูลถาวร⁶ (วัดจากจำนวนชิ้นส่วนที่มีรูปแบบไม่ซ้ำกัน ณ ตำแหน่งเดียวกันในโครโมโซมที่ถูกเก็บอยู่ในหน่วยเก็บถาวร)
- nocp: number of crossable point หรือปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซม⁷

ส่วนต่อไปนี้เป็นกรณีตัวอย่างและภาพเพื่อแสดงให้เห็นถึงนิยามคำศัพท์ต่าง ๆ ที่ใช้ใน งานวิจัยนี้ได้ชัดเจนมากขึ้น และสามารถทำความเข้าใจได้ง่ายขึ้น

กรณีตัวอย่างปัญหา Royal Road ซึ่งมีขนาดหน่วยการสร้างเท่ากับ 8 บิต (เรียงติดกัน) จากรูปที่ 4.2 หน่วยการสร้างที่ต้องการ (หน่วยการสร้างที่เป็นคำตอบ) จะต้องประกอบด้วยคู่จีนที่มีค่าเป็น '1' ทั้งหมด ซึ่งเราจะให้ชื่อว่า Pure BBs ส่วนโครงสร้างคู่จีนที่มีรูปแบบตรงกันข้ามกับคู่จีนที่ต้องการโดยสิ้นเชิงคือ มีค่าเป็น '0' ทุกบิตจะให้ชื่อว่า Non BBs และโครงสร้างที่มีส่วนประกอบของคู่จีนที่ตรงกับหน่วยการสร้างที่ต้องการอย่างน้อยหนึ่งตำแหน่ง (หนึ่งบิต) จะให้ชื่อว่า Mixed BBs

⁴ หมายถึง ใช้สำหรับขั้นตอนวิธีการไขว้เปลี่ยนแบบชิ้นส่วนในบทที่ 5 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไขในบทที่ 8 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW ในบทที่ 9 และขั้นตอนวิธีประมาณการแจกแจงแบบ LZW ในบทที่ 10

⁵ หมายถึง ใช้สำหรับขั้นตอนวิธีการไขว้เปลี่ยนแบบชิ้นส่วนในบทที่ 5 เท่านั้น

⁶ หมายถึง ใช้สำหรับขั้นตอนวิธีที่มีการใช้หน่วยเก็บถาวรได้แก่ ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่มในบทที่ 7 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไขในบทที่ 8 และขั้นตอนวิธีประมาณการแจกแจงแบบ LZW ในบทที่ 10

⁷ หมายถึง ใช้สำหรับขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW ในบทที่ 9 เท่านั้น

Desired BBs	1	1	1	1	1	1	1	1
Pure BBs	1	1	1	1	1	1	1	1
Mixed BBs	1	1	1	0	1	1	1	1
Mixed BBs	0	0	0	0	0	0	0	1
Non BBs	0	0	0	0	0	0	0	0

รูปที่ 4.2 การจัดจำแนกหน่วยการสร้างโดยใช้รูปแบบคู่จิ้นในปัญหา Royal Road

สำหรับปัญหา Trap-5 หน่วยการสร้างที่เป็นคำตอบจะมีรูปแบบเป็นคู่จิ้นที่มีค่าเป็น '1' เรียงติดกันจำนวน 5 บิต (11111) (คิดเป็น 1 หน่วย) สำหรับปัญหา Trap-5 ขนาด 20 บิต จะมีหน่วยการสร้างทั้งหมดจำนวน 4 หน่วยเรียงติดกันดังรูปที่ 4.3 ซึ่งแสดงให้เห็นถึงจำนวนของ Pure BBs ในโครโมโซม C1, C2, C3, C4 และ C5 (จะเห็นว่าหน้าตาของโครโมโซมที่เป็นคำตอบจะต้องประกอบไปด้วย Pure BBs จำนวน 4 หน่วย) ในโครโมโซม C1 มี Pure BBs จำนวน 3 หน่วย (ซึ่งมีปริมาณมากที่สุด ใน 5 โครโมโซม) ในโครโมโซม C2 มี Pure BBs จำนวน 1 หน่วย ในโครโมโซม C3 มี Pure BBs 2 หน่วย ในโครโมโซม C4 มี Pure BBs 2 หน่วย และในโครโมโซม C5 มี Pure BBs 0 หน่วย หรือไม่มีเลยจึงถือว่าเป็นโครโมโซมที่มีจำนวน Pure BBs น้อยที่สุด ถ้าเรากำหนดให้มีประชากรทั้งหมด 5 ตัว (โครโมโซมทั้งหมด 5 ตัวนี้เป็นประชากร) เราสามารถบอกได้ว่าในประชากรนี้มีจำนวน Pure BBs มากที่สุด (Max) เป็นจำนวน 3 หน่วย และมีจำนวน Pure BBs น้อยที่สุด (Min) เป็นจำนวน 0 หน่วย และมีจำนวน Pure BBs เฉลี่ย (Average: avg) เป็น $(3+1+2+2) / 5 = 1.6$ หน่วย

Desired Pattern	Answer	BBs 1	BBs 2	BBs 3	BBs 4
Max Pure BBs	C1	BBs 1	BBs 2	BBs 3	BBs 4
	C2	BBs 1	BBs 2	BBs 3	BBs 4
	C3	BBs 1	BBs 2	BBs 3	BBs 4
	C4	BBs 1	BBs 2	BBs 3	BBs 4
Min Pure BBs	C5	BBs 1	BBs 2	BBs 3	BBs 4
Definition:		Pure BBs	Non BBs	Mixed BBs	

รูปที่ 4.3 รูปแบบหน่วยการสร้างในโครโมโซม

เมื่อพิจารณาโครโมโซมในรูปแบบชิ้นส่วนจะสามารถบอกปริมาณของชิ้นส่วน (number of fragment: nof) ที่มี และขนาดโดยเฉลี่ยของชิ้นส่วน (average fragment size: afs) ได้ จากรูปที่ 4.4 โครโมโซมมีจำนวนคู่จีนเท่ากับ 6 บิต (A1-A6 ; A: Allele) ถ้าเรากำหนดให้ประชากรมีโครโมโซมจำนวน 3 ตัว (C1, C2 และ C3) สามารถนับปริมาณชิ้นส่วนที่มีในประชากรทั้งหมดได้เป็น $(5+3+6) = 14$ ชิ้นส่วน และเราสามารถคำนวณหาขนาดของชิ้นส่วนโดยเฉลี่ยได้เป็น $(6 \times 3) / 14 = 1.29$ บิต (โดยประมาณ) (คิดจากจำนวนคู่จีนที่มีทั้งหมดในประชากรหารด้วยจำนวนชิ้นส่วนทั้งหมดในประชากร)

Chromosome	A1	A2	A3	A4	A5	A6
C1	F1	F2	F3	F4	F5	
C2	F1	F2	F3			
C3	F1	F2	F3	F4	F5	F6

รูปที่ 4.4 รูปแบบชิ้นส่วนในโครโมโซม

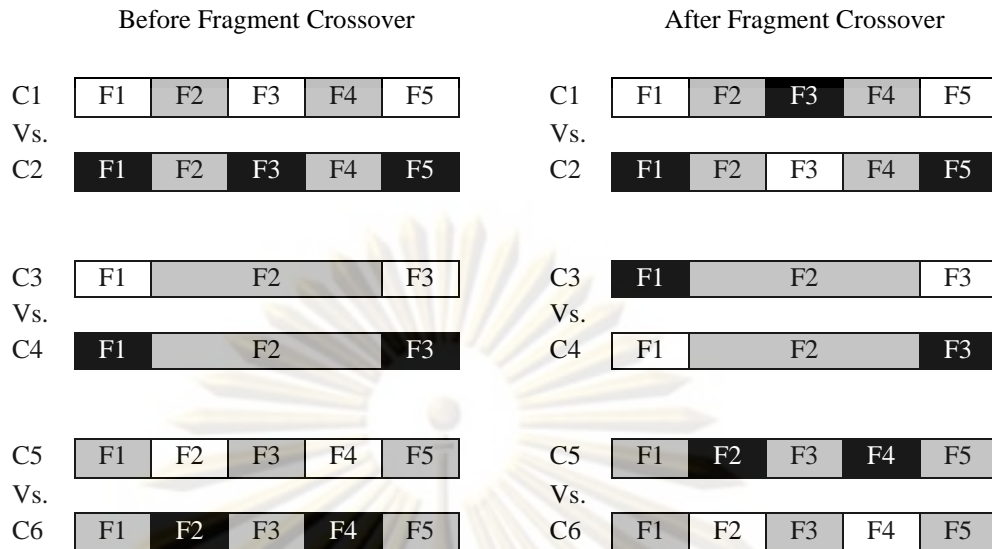
ในกรณีที่ตัวดำเนินการระบุหน่วยการสร้างมีการเปรียบเทียบรูปแบบคู่จีนระหว่างโครโมโซมสองตัวโดยใช้หลักการสารสนเทศร่วมกันต่ำสุด ซึ่งได้แก่วิธีการไขว้เปลี่ยนแบบชิ้นส่วน (บทที่ 5) และขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (บทที่ 8) จะทำให้เกิดชิ้นส่วนร่วมกัน (Common Fragment: cf) และชิ้นส่วนไม่ร่วมกัน (Different Fragment: df) (หรือเรียกว่าชิ้นส่วนเหมือนกันและชิ้นส่วนต่างกันตามลำดับ) จากรูปที่ 4.5 มีโครโมโซมทั้งหมด 6 ตัวแต่ละตัวมีความยาว 5 บิต และแบ่งออกเป็น 3 คู่ เมื่อพิจารณาแต่ละคู่เริ่มจากคู่โครโมโซม C1 กับ C2 มีจำนวนจำนวนบิตที่มีค่าเหมือนกันอยู่ 2 ช่วงเมื่อมองในมุมของชิ้นส่วนจะเท่ากับมี 5 ชิ้นส่วน (รวมทั้งช่วงที่เหมือนกันและต่างกัน) ส่วนคู่โครโมโซม C3 กับ C4 มีช่วงที่เหมือนกันอยู่ช่วงเดียวเมื่อมองในมุมมองชิ้นส่วนจะมี 3 ชิ้นส่วน และคู่โครโมโซม C5 กับ C6 จะมีช่วงที่เหมือนกันอยู่ 3 ช่วงเมื่อมองในมุมมองชิ้นส่วนจะมี 5 ชิ้นส่วน (เท่ากับคู่ C1 กับ C2) จากทั้งสามคู่จะเห็นว่าขนาดของชิ้นส่วนแปรผกผันกับจำนวนชิ้นส่วน หรือกล่าวได้ว่าถ้ามีปริมาณชิ้นส่วนมากจะได้ชิ้นส่วนที่มีขนาดเล็กและในทางกลับกันถ้ามีปริมาณชิ้นส่วนน้อยลงจะทำให้ได้ขนาดชิ้นส่วนที่ใหญ่ขึ้น ถ้าเรากำหนดให้ประชากรมีโครโมโซม 6 ตัว (C1-C6) และเกิดการเปรียบเทียบกันดังรูป 4.5 เราจะสามารถคำนวณหาปริมาณชิ้นส่วนทั้งหมดในประชากรได้เป็น $(5+3+5) = 13$ ชิ้นส่วน (จากคู่โครโมโซมที่เปรียบเทียบกันเราจะคิดจำนวนชิ้นส่วนจากโครโมโซมตัวใดตัวหนึ่งเท่านั้น เนื่องจากโครโมโซมทั้งสองตัวมีจำนวนชิ้นส่วนเท่ากันและมีหน้าตาชิ้นส่วนเหมือนกัน) และถ้าแยกเป็นกรณีชิ้นส่วนร่วมกันและไม่

ร่วมกันจะได้ว่ามีปริมาณชิ้นส่วนร่วมกันจำนวน (number of common fragment: nocf) $(2+1+3) = 6$ ชิ้นส่วน (ชิ้นส่วนสีเทา) และมีปริมาณชิ้นส่วนไม่ร่วมกันจำนวน (number of different fragment: nodf) $(3+2+2) = 7$ ชิ้นส่วน (ชิ้นส่วนสีขาว) เรายังสามารถคำนวณหาขนาดโดยเฉลี่ยของชิ้นส่วนร่วมกัน (average common fragment size: acfs) ได้เป็น $(5 \times 3) / 6 = 2.5$ บิต (คิดจากจำนวนคู่จิ้นที่มีครึ่งหนึ่งจากประชากรทั้งหมดหารด้วยจำนวนชิ้นส่วนร่วมกัน) และคำนวณหาขนาดโดยเฉลี่ยของชิ้นส่วนไม่ร่วมกัน (average different fragment size: adfs) ได้เป็น $(5 \times 3) / 7 = 2.14$ บิต (โดยประมาณ) (คิดจากจำนวนคู่จิ้นที่มีครึ่งหนึ่งจากประชากรทั้งหมดหารด้วยจำนวนชิ้นส่วนไม่ร่วมกัน)

Allele Representation		Fragment Representation											
C1	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td></tr></table>	0	1	1	1	0	C1	<table border="1"><tr><td>F1</td><td>F2</td><td>F3</td><td>F4</td><td>F5</td></tr></table>	F1	F2	F3	F4	F5
0	1	1	1	0									
F1	F2	F3	F4	F5									
Vs.		Vs.											
C2	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	1	1	0	1	1	C2	<table border="1"><tr><td>F1</td><td>F2</td><td>F3</td><td>F4</td><td>F5</td></tr></table>	F1	F2	F3	F4	F5
1	1	0	1	1									
F1	F2	F3	F4	F5									
C3	<table border="1"><tr><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	0	1	1	C3	<table border="1"><tr><td>F1</td><td>F2</td><td>F3</td></tr></table>	F1	F2	F3		
0	1	0	1	1									
F1	F2	F3											
Vs.		Vs.											
C4	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td></tr></table>	1	1	0	1	0	C4	<table border="1"><tr><td>F1</td><td>F2</td><td>F3</td></tr></table>	F1	F2	F3		
1	1	0	1	0									
F1	F2	F3											
C5	<table border="1"><tr><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr></table>	0	0	1	0	0	C5	<table border="1"><tr><td>F1</td><td>F2</td><td>F3</td><td>F4</td><td>F5</td></tr></table>	F1	F2	F3	F4	F5
0	0	1	0	0									
F1	F2	F3	F4	F5									
Vs.		Vs.											
C6	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td></tr></table>	0	1	1	1	0	C6	<table border="1"><tr><td>F1</td><td>F2</td><td>F3</td><td>F4</td><td>F5</td></tr></table>	F1	F2	F3	F4	F5
0	1	1	1	0									
F1	F2	F3	F4	F5									

รูปที่ 4.5 การระบุชิ้นส่วนด้วยวิธีการใช้ข้อมูลสารสนเทศร่วมกันล่าสุด

สำหรับตัวดำเนินการไขว้เปลี่ยนแปลงแบบชิ้นส่วนจากรูปที่ 4.5 และ 4.6 (กำหนดให้โครโมโซม C2, C4 และ C6 มีชิ้นส่วนไม่ร่วมกันเป็นสีดำเพื่อให้เห็นภาพได้ชัดเจนขึ้น) ถ้าเรากำหนดให้มีจำนวนประชากรเท่ากับ 6 (C1-C6) และเกิดการดำเนินการดังรูปที่ 4.6 (มีโอกาสเกิดการไขว้เปลี่ยนแปลงเฉพาะชิ้นส่วนไม่ร่วมกันที่เป็นสีดำ) เราจะสามารถคำนวณหาจำนวนชิ้นส่วนที่ถูกไขว้เปลี่ยนระหว่างคู่โครโมโซมในประชากรทั้งหมดได้เป็น $(1+1+2) = 4$ ชิ้นส่วน และสามารถคำนวณหาจำนวนชิ้นส่วนโดยเฉลี่ยที่ถูกไขว้เปลี่ยนระหว่างคู่โครโมโซม (average number of crossed pieces: ancp) ได้เป็น $(4 / 3) = 1.33$ ชิ้นส่วน (โดยประมาณ) (คิดจากจำนวนชิ้นส่วนที่ถูกไขว้เปลี่ยนทั้งหมดในประชากรหารด้วยจำนวนครึ่งหนึ่งของประชากร)

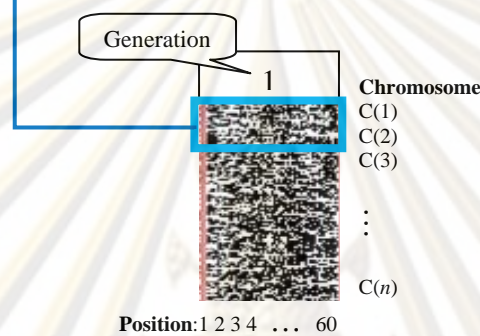
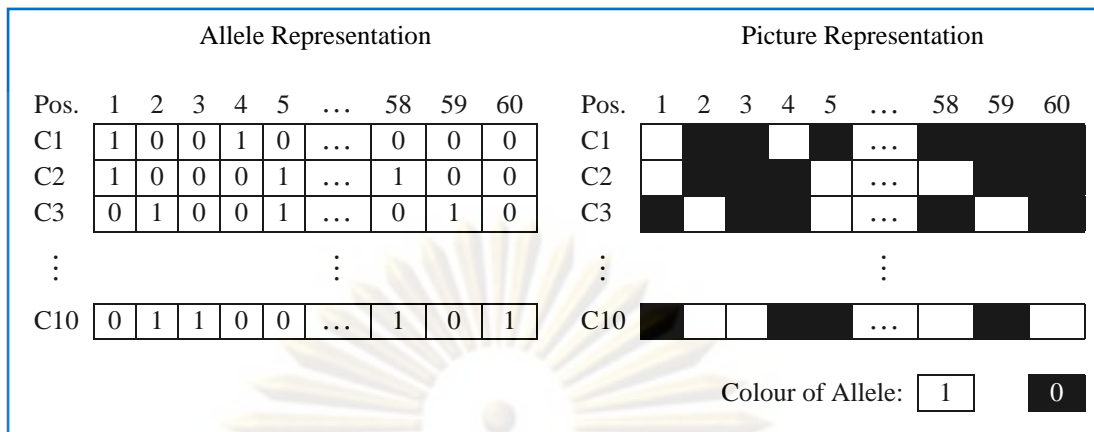


รูปที่ 4.6 รูปแบบชิ้นส่วนก่อนและหลังการไขว้เปลี่ยนแบบชิ้นส่วน

ในส่วนของการประชากรเราสามารถแสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นในแต่ละรุ่นประชากรโดยนำเสนอในรูปแบบภาพเพื่อให้เห็นภาพรวมได้ง่ายและชัดเจน จากรูปที่ 4.7 เป็นภาพแสดงรายละเอียดโครงสร้างทางพันธุกรรมของประชากรในหนึ่งรุ่นประชากร ประกอบด้วยสองส่วนได้แก่ ส่วนหัวที่มีตัวเลขอยู่ในกรอบสี่เหลี่ยมเป็นตัวบอกถึงรุ่นประชากร (บอกว่า เป็นรุ่นที่เท่าใด ในรูปที่แสดงเป็นรุ่นประชากรที่ 1) และส่วนข้อมูลจีโนไทป์ของประชากรซึ่งจะมีลักษณะเป็นภาพขาวดำ โดยตำแหน่งที่เป็นสีดำหมายถึงคู่ยีนที่มีค่าเป็น '0' และตำแหน่งที่เป็นสีขาวหมายถึงคู่ยีนที่มีค่าเป็น '1' โครงสร้างของคู่ยีนมีการจัดเรียงตามลำดับของโครโมโซมแต่ละตัวในกลุ่มประชากร ในแนวตั้งของรูปหมายถึงโครโมโซมแต่ละตัวในกลุ่มประชากร (ในรูปมีจำนวนประชากรทั้งหมด n ตัว) และในแนวนอนหมายถึงตำแหน่งของคู่ยีนในแต่ละโครโมโซม (ในรูปมีความยาวโครโมโซมเท่ากับ 60 บิต) ในภาพที่เห็นเป็นการเรียงซ้อนกันของโครโมโซมแต่ละตัวในแนวตั้งโดยไม่ได้สนใจเรื่องลำดับก่อนหลังของแต่ละโครโมโซม (โครโมโซมไหนจะอยู่บนหรือล่างก็ได้) เนื่องจากเราต้องการเห็นภาพการกระจายตัวของคู่ยีนที่เป็น '0' และ '1' ในโครโมโซมเท่านั้น

ศูนย์วิจัยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

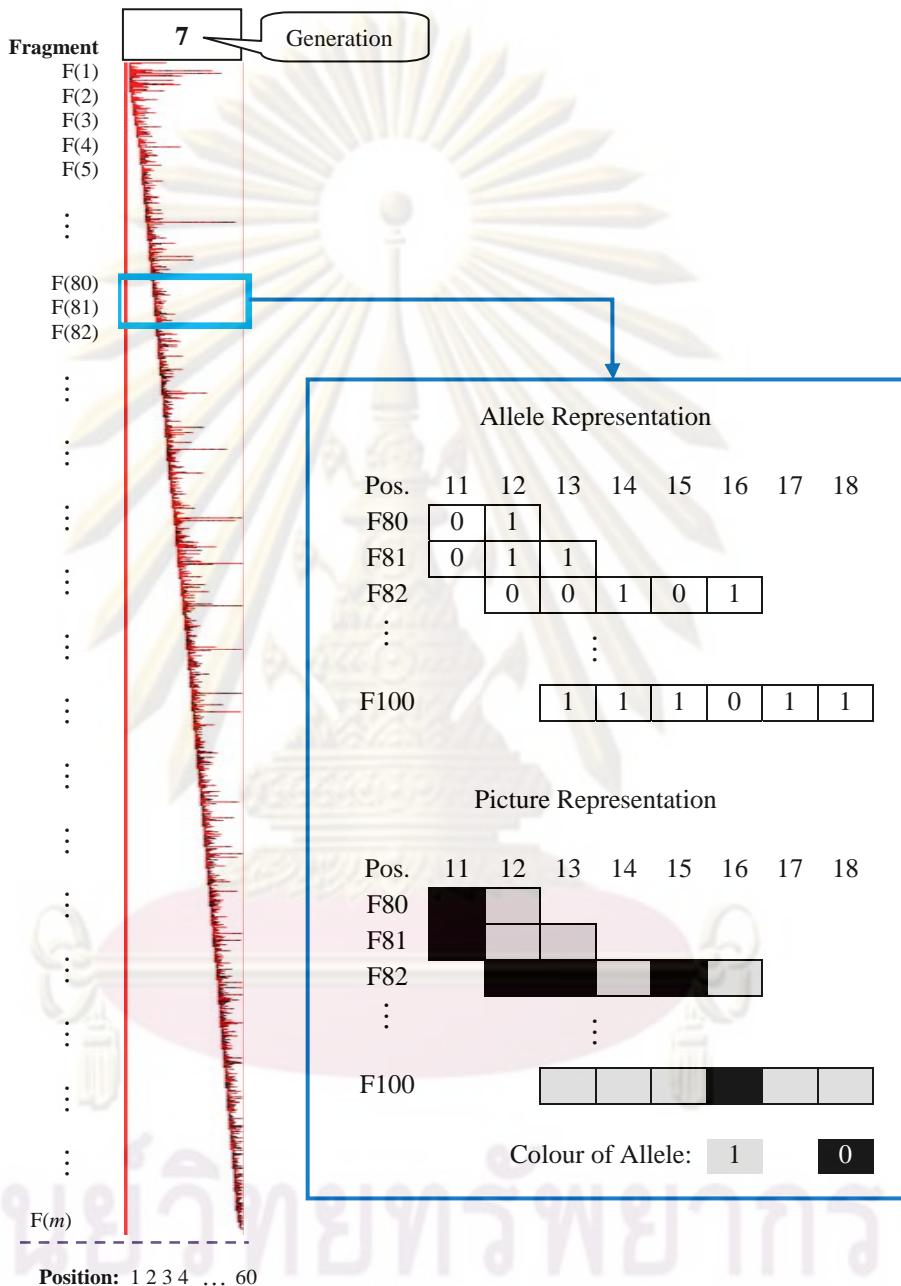


รูปที่ 4.7 โครงสร้างภาพจีโนไทป์ของประชากรในหนึ่งรุ่นประชากร

นอกจากโครงสร้างประชากรแล้วเรายังแสดงให้เห็นภาพรวมของโครงสร้างชิ้นส่วนที่สามารถระบุได้ในแต่ละรุ่นประชากรด้วย โดยใช้รูปแบบการนำเสนอแบบเดียวกับโครงสร้างประชากร จากรูปที่ 4.8 ข้อมูลโครงสร้างชิ้นส่วนจะถูกแบ่งออกเป็นสองส่วนได้แก่ ส่วนหัวที่เป็นตัวเลขในกรอบสี่เหลี่ยมหลายถึงรุ่นประชากร (บอกว่าเป็นรุ่นที่เท่าใด ในรูปที่แสดงเป็นรุ่นประชากรที่ 7) และข้อมูลจีโนไทป์ของชิ้นส่วนจะแสดงในส่วนถัดมา (ด้านล่าง) ตำแหน่งที่มีสีเทาหมายถึงคู่จีโนมที่มีค่าเป็น '1' และตำแหน่งที่มีสีดำหมายถึงคู่จีโนมที่มีค่าเป็น '0' จากรูปในแนวตั้งแสดงถึงชิ้นส่วนแต่ละชิ้นที่เรียงซ้อนกันลงมา (ในรูปมีทั้งหมด m ชิ้นส่วน) และแนวนอนเป็นตำแหน่งของชิ้นส่วนที่ระบุได้ในโครโมโซม (อ้างอิงกับตำแหน่งในโครโมโซมซึ่งมีความยาว 60 บิต) จะเห็นว่าริ้วแถบข้อมูลถูกแสดงในแนวเฉียงจากซ้ายบนลงมายังขวาล่างเนื่องจากเราต้องการแสดงโครงสร้างคู่จีโนมของชิ้นส่วนให้อยู่ในรูปแบบที่อ้างอิงกับตำแหน่งที่ระบุได้ในโครโมโซม เพื่อให้เห็นภาพว่าชิ้นส่วนที่ระบุได้ในแต่ละช่วงของโครโมโซมมีความแตกต่างกันอย่างไร (ทั้งรูปแบบและจำนวน) ข้อมูลชิ้นส่วนที่ระบุได้เหล่านี้จะถูกเก็บรวมกันไว้ในหน่วยเก็บถาวรซึ่งจะเก็บเฉพาะรูปแบบชิ้นส่วนที่ไม่ซ้ำกัน ณ ตำแหน่งเดียวกันในโครโมโซม เพื่อที่จะ

⁸ หมายถึง สำหรับขั้นตอนวิธีที่มีการใช้หน่วยเก็บถาวรเท่านั้น

นำไปใช้ต่อไปในกระบวนการประกอบชิ้นส่วนเข้าด้วยกัน โดยทั่วไปขนาดของหน่วยเก็บถาวร (size of archive: soa) จะลดลงเรื่อย ๆ ในรุ่นประชากรถัดไปเนื่องจากเกิดการสูญเสียค่าตอบ ทำให้มีความหลากหลายลดลง ส่งผลให้ชิ้นส่วนมีขนาดใหญ่ขึ้นและมีจำนวนลดลง



รูปที่ 4.8 โครงสร้างจีโนมไปป์ของชิ้นส่วนในหนึ่งรุ่นประชากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

การไขว้เปลี่ยนแบบชิ้นส่วน

(Fragment Crossover: FC)

ในบทนี้จะกล่าวถึงการออกแบบวิธีการระบุและประกอบหน่วยการสร้างในส่วนของการไขว้เปลี่ยน โดยขั้นตอนวิธีที่ใช้อยู่บนพื้นฐานของขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย เพื่อทดสอบความเป็นไปได้ของนิยามหน่วยการสร้างแบบชิ้นส่วนว่าสามารถนำไปประยุกต์ใช้ได้จริงหรือไม่ และให้ผลอย่างไร

5.1 แนวคิดและหลักการ

จากงานวิจัยที่เกี่ยวข้องกับการกับไขว้เปลี่ยนทำให้สามารถสรุปได้ว่ากระบวนการหนึ่งที่สำคัญที่สุดที่สามารถออกแบบและพัฒนาให้มีประสิทธิภาพอย่างชัดเจนคือ กระบวนการประกอบเข้าด้วยกันใหม่ หรือกระบวนการไขว้เปลี่ยนสำหรับขั้นตอนวิธีเชิงพันธุกรรมนั่นเอง การไขว้เปลี่ยนจะมีประสิทธิภาพได้ต้องประกอบด้วยความสามารถในการลดผลกระทบจากความแตกแยก (การทำลายเค้าร่างที่ดี) และความสามารถที่ทำให้เกิดการแลกเปลี่ยนกันในอัตราสูงที่สุด จากมุมมองชิ้นส่วนที่นำเสนอทำให้ออกแบบวิธีการไขว้เปลี่ยนที่มีครบทั้งสององค์ประกอบได้ด้วยวิธีดังต่อไปนี้

เรานิยามชิ้นส่วนด้วยโครงสร้างย่อยร่วมกันของสองโครโมโซมใด ๆ กำหนดให้ลำดับของโครโมโซมความยาว l

$$C^k = c_1^k c_2^k \dots c_l^k : k \in I^+, c_n^k \in \{0,1\}, 1 \leq n \leq l \quad (5.1)$$

กำหนดให้เซตของดัชนีตำแหน่งอ้างอิงในโครโมโซมความยาว l จากตำแหน่งที่ f ถึงตำแหน่งที่ t

$$(f_i, t_i) = \{ z_i : 1 \leq f_i \leq z_i \leq t_i \leq l, f_i, t_i, z_i \in I^+ \} \\ \text{and } \forall i, j \in I^+ : i \neq j, (f_i, t_i) \cap (f_j, t_j) = \emptyset \quad (5.2)$$

กำหนดให้ลำดับย่อย (ชิ้นส่วน) ต่างกันระหว่างสองโครโมโซม C^{k_1} กับ C^{k_2}

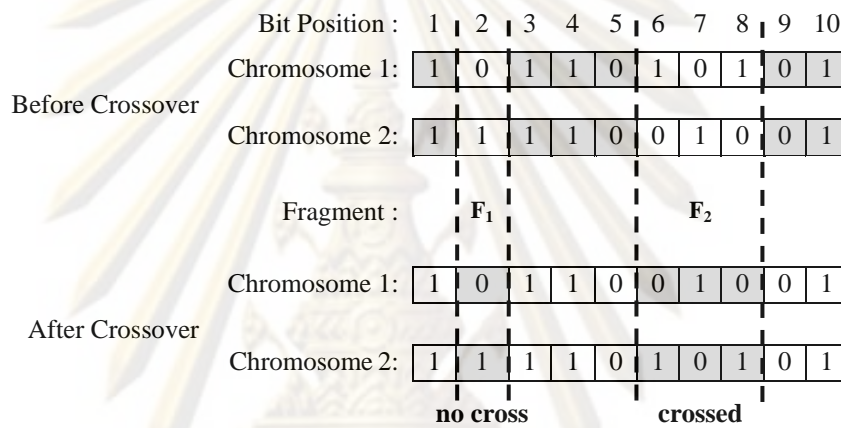
$$F_{i,1}^{k_1,k_2} = c_{f_i}^{k_1} c_{f_i+1}^{k_1} \dots c_{t_i}^{k_1} \text{ such that } c_{z_i}^{k_1} \neq c_{z_i}^{k_2} \quad (5.3)$$

$$F_{i,2}^{k_1,k_2} = c_{f_i}^{k_2} c_{f_i+1}^{k_2} \dots c_{t_i}^{k_2} \text{ such that } c_{z_i}^{k_1} \neq c_{z_i}^{k_2} \quad (5.4)$$

เซตของลำดับย่อยต่างกันที่เรียงติดกันระหว่างสองโครโมโซมนิยามดังนี้

$$S^{k_1, k_2} = \left\{ F_{i,1}^{k_1, k_2}, F_{i,2}^{k_1, k_2} : i \leq \left\lceil \frac{l}{2} \right\rceil \right\} \quad (5.5)$$

จากนิยามกำหนดให้โครโมโซมเรียงลำดับจากซ้าย (เริ่มต้น) ไปขวา สมมติให้โครโมโซมขนาด 10 บิตสองตัวมีค่าดังนี้ $C^1 = (1,0,1,1,0,1,0,1,0,1)$ และ $C^2 = (1,1,1,1,0,0,1,0,0,1)$ ดังนั้นจะได้ช่วงดัชนีตำแหน่งร่วมระหว่างโครโมโซม C^1 และ C^2 เป็น $(f_1, t_1) = (2,2)$ และ $(f_2, t_2) = (6,8)$ และจะได้ชิ้นส่วน $S^{1,2}$ ดังต่อไปนี้ $F_{1,1}^{1,2} = c_2^1 = 0$, $F_{1,2}^{1,2} = c_2^2 = 1$, $F_{2,1}^{1,2} = c_6^1 c_7^1 c_8^1 = 101$ และ $F_{2,2}^{1,2} = c_6^2 c_7^2 c_8^2 = 010$ ตามลำดับ สมมติให้เกิดการไขว้เปลี่ยนแบบชิ้นส่วนขึ้นที่ชิ้นส่วน F_2 เพียงชิ้นส่วนเดียวจะได้ผลการไขว้เปลี่ยนดังรูปที่ 5.1



รูปที่ 5.1 การดำเนินการไขว้เปลี่ยนแบบชิ้นส่วนระหว่าง 2 โครโมโซม

และสามารถแสดงให้อยู่ในรูปของรหัสเทียมได้ดังรูปที่ 5.2

P_1, P_2 คือ ผลเฉลยที่ผ่านการคัดเลือกมาแล้ว

ขั้นตอนวิธีการไขว้เปลี่ยนแบบชิ้นส่วน (Fragment Crossover)

$F_p \leftarrow$ หาช่วงส่วนที่แตกต่างกันระหว่าง P_1 กับ P_2 เมื่อนำผลเฉลยทั้งสองมาเทียบกับกันโดยกำหนดให้ตำแหน่งตรงกัน

วนซ้ำสำหรับ $m = 1, 2, \dots, k$ (สมมติมีทั้งหมด k = ชิ้นส่วนที่แตกต่างกัน)

ถ้าดำเนินการไขว้เปลี่ยน ทำ

$$F_p \leftarrow \text{swap}(F_{p_1}^m, F_{p_2}^m)$$

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 5.2 รหัสเทียมของวิธีการไขว้เปลี่ยนแบบชิ้นส่วนสำหรับวิธี sGA-FC

วิธีที่นำเสนอมาจากแนวคิดที่ได้จากขั้นตอนวิธีเชิงพันธุกรรมที่เชื่อกันว่า ระหว่างผลเฉลยตัวที่ดี ๆ น่าจะมีส่วน (ชิ้นส่วน หรือเค้กร่าง) ที่ดี ๆ ร่วมกัน และถ้าสามารถประกอบชิ้นส่วนที่ดีเหล่านั้นเข้าด้วยกันได้อย่างถูกต้องแล้วจะได้ผลเฉลยที่ดีขึ้น หลักการดังกล่าวนี้เป็นที่มาของหน่วยการสร้างและนิยามหน่วยการสร้างแบบชิ้นส่วน เราเชื่อว่าถ้าสามารถระบุชิ้นส่วนเหล่านี้ได้อย่างถูกต้องจะทำให้การประกอบเข้าด้วยกันทำได้ง่ายขึ้น จากความรู้เรื่องทฤษฎีสารสนเทศทำให้สามารถหาชิ้นส่วนที่ดีที่มีอยู่ระหว่างโครโมโซมได้ เมื่อพิจารณาโครโมโซมใด ๆ เราสามารถบอกได้ว่าทั้งสองโครโมโซมนั้นขึ้นแก่กันในระดับไหนโดยใช้หลักการสารสนเทศร่วม (Mutual Information) ซึ่งจะพิจารณาถึงความเหมือนหรือความต่างในส่วนของโครงสร้างจีโนมไทป์ (รูปแบบคู่จีนในโครโมโซม) ถ้ามีคู่จีนที่ตำแหน่งเดียวกันเหมือนกันเป็นจำนวนมากแสดงว่ายังมีสารสนเทศร่วมกันในระดับสูง ชิ้นส่วนที่เหมือนกันระหว่างสองโครโมโซมนี้เราจะพิจารณาว่าเป็นเค้กร่างที่ดีที่ไม่ควรถูกแยกจากกัน ซึ่งเป็นการลดผลกระทบในเรื่องการแตกแยก ผลที่ตามมาก็คือเราจะได้ชิ้นส่วนร่วมกัน (เหมือนกัน) (Common Fragment) และชิ้นส่วนไม่ร่วมกัน (ไม่เหมือนกัน) (Uncommon Fragment) หลายช่วงระหว่างโครโมโซมดังรูปที่ 5.1 เราจะทำการแลกเปลี่ยนชิ้นส่วนที่ไม่เหมือนกันระหว่างสองโครโมโซมแบบสุ่มด้วยอัตราการแลกเปลี่ยนค่าหนึ่ง (แต่ละชิ้นส่วนไม่ร่วมกันทุกชิ้นจะมีโอกาสเกิดการแลกเปลี่ยนเท่ากันหมด) ซึ่งจะทำให้เกิดการแลกเปลี่ยนกันในอัตราสูงสุดโดยที่ชิ้นส่วนดี ๆ จะไม่ถูกทำให้แตกแยกจากกัน วิธีที่นำเสนอนี้เป็นไปตามหลักการที่ต้องการครบทั้งสองข้อ ดังนั้นเราจึงคาดหวังว่าตัวดำเนินการแบบใหม่นี้จะมีประสิทธิภาพมากกว่าการไขว้เปลี่ยนแบบจุดเดียวในปัญหาทดสอบที่กำหนด

วิธีที่นำเสนอเมื่อพิจารณาเปรียบเทียบกับการไขว้เปลี่ยนแบบอื่น ๆ เช่น การไขว้เปลี่ยนแบบจุดเดียว การไขว้เปลี่ยนแบบสองจุด การไขว้เปลี่ยนแบบหลายจุด หรือการไขว้เปลี่ยนแบบเอกรูป เป็นต้น เราสามารถแสดงให้เห็นถึงความแตกต่างกันได้ในมุมมองดังต่อไปนี้ เมื่อพิจารณาโครโมโซมใด ๆ ที่เกิดขึ้นจากการสุ่มแบบเอกรูปและมีความยาว l โอกาสที่คู่จีนที่ตำแหน่งเดียวกันระหว่างสองโครโมโซมจะมีรูปแบบต่างกัน (ไม่เหมือนกัน) โดยเฉลี่ยมีค่าเป็น 0.5 ดังนั้นโครโมโซมยาว l จะมีจำนวนคู่จีนที่ต่างกันจำนวน $0.5 l$ โดยประมาณ (ต่างกันครึ่งหนึ่งของจำนวนคู่จีนที่มีทั้งหมดในโครโมโซม) และถ้าการกระจายตัวของคู่จีนที่ต่างกันเป็นแบบเอกรูป เมื่อใช้การไขว้เปลี่ยนแบบจุดเดียว โครโมโซมลูกที่ได้โดยเฉลี่ยจะมีคู่จีนที่มาจากพ่อครึ่งหนึ่งและแม่ครึ่งหนึ่ง ดังนั้นจะเกิดการแลกเปลี่ยนคู่จีนที่ต่างกัน (จากรุ่นพ่อแม่) ในหนึ่งโครโมโซมโดยเฉลี่ยเป็น $0.5 \times 0.5 l$ หรือ $0.25 l$ (1 ใน 4 ของจำนวนคู่จีนในโครโมโซม) โดยประมาณ ส่วนวิธีการไขว้เปลี่ยนแบบสองจุดโดยทั่วไปจะแบ่งโครโมโซมออกเป็น 3 ท่อน การแลกเปลี่ยนจะเกิดขึ้นกับท่อนกลางของโครโมโซมพ่อพันธุ์และแม่พันธุ์ ดังนั้นจะเห็นว่าจำนวนคู่จีนที่เกิดการแลกเปลี่ยนจะมีขนาด $1/3$

(0.33) ของความยาวโครโมโซมโดยเฉลี่ย และจะเกิดการแลกเปลี่ยนคู่ยีนที่ต่างกันในหนึ่งโครโมโซมโดยเฉลี่ยเป็น $(1/3) \times (1/2) I$ หรือ $(1/6) I$ (เท่ากับ $0.165 I$ โดยประมาณ) ถ้าพิจารณากรณีที่เพิ่มจุดตัดในการไขว้เปลี่ยนไปเรื่อย ๆ จะได้ว่าการไขว้เปลี่ยนที่มีจำนวนจุดตัดเป็นเลขคี่จะเกิดการแลกเปลี่ยนคู่ยีนที่ต่างกันในหนึ่งโครโมโซมโดยเฉลี่ยเป็น $0.25 I$ โดยประมาณเสมอ และถ้าจำนวนจุดตัดเป็นเลขคู่จะเกิดการแลกเปลี่ยนคู่ยีนที่ต่างกันในหนึ่งโครโมโซมโดยเฉลี่ยเป็น $\frac{(c/2)}{(c+1)} \times 0.5I$ เมื่อ c คือจำนวนจุดตัดที่เป็นเลขคู่ จะเห็นว่าเมื่อ c มีค่ามาก ๆ สมการจะลู่เข้าสู่ค่า $0.25 I$ เท่ากับกรณีที่จำนวนจุดตัดเป็นเลขคี่ (แต่โดยส่วนใหญ่ในงานวิจัยที่ใช้การไขว้เปลี่ยนแบบหลายจุดตัดจะใช้จำนวนจุดตัดไม่เกิน 4-6 จุด) ส่วนการไขว้เปลี่ยนแบบเอกรูปจะเป็นกรณีที่ต่างออกไปเนื่องจากโอกาสเกิดการไขว้เปลี่ยนในแต่ละตำแหน่งคู่ยีนจะเป็นอิสระจากกัน ดังนั้นจะเกิดการแลกเปลี่ยนคู่ยีนที่แตกต่างกันมากหรือน้อยขึ้นอยู่กับอัตราการแลกเปลี่ยนที่กำหนด ในกรณีที่กำหนดอัตราการแลกเปลี่ยนเป็น 0.5 จะทำให้โดยเฉลี่ยแล้วจะมีจำนวนคู่ยีนครึ่งหนึ่งของความยาวโครโมโซมที่ถูกแลกเปลี่ยน และจะเกิดการแลกเปลี่ยนคู่ยีนที่ต่างกันในหนึ่งโครโมโซมโดยเฉลี่ยเป็น $0.25 I$ โดยประมาณ เป็นต้น ดังนั้นจะเห็นได้ว่าวิธีการไขว้เปลี่ยนแบบเอกรูปจะสามารถกำหนดปริมาณการแลกเปลี่ยนข้อมูลได้ง่าย (ขึ้นตรงกับค่าอัตราการไขว้เปลี่ยน) ส่วนวิธีการไขว้เปลี่ยนแบบขึ้นส่วนที่ได้นำเสนอจะมีลักษณะเหมือนกับวิธีการไขว้เปลี่ยนแบบเอกรูปตรงที่แต่ละขึ้นส่วนจะมีโอกาสเกิดการไขว้เปลี่ยนเป็นอิสระจากกัน ดังนั้นในกรณีที่กำหนดอัตราการแลกเปลี่ยนเท่ากับ 0.5 จะเกิดการแลกเปลี่ยนคู่ยีนที่ต่างกันในหนึ่งโครโมโซมโดยเฉลี่ยเป็น $0.5 \times 0.5 I$ หรือ $0.25 I$ โดยประมาณ ตรงจุดนี้เองเป็นการตอกย้ำให้เห็นว่าวิธีการไขว้เปลี่ยนแบบเอกรูปมีอัตราการแลกเปลี่ยนข้อมูลสูงสุดแต่ก็เกิดอัตราการแตกแยกของข้อมูลสูงสุดเช่นกัน ซึ่งต่างกับวิธีการไขว้เปลี่ยนแบบขึ้นส่วนที่สามารถลดอัตราการแตกแยกของข้อมูลลงได้ ในระหว่างที่พยายามทำให้เกิดการแลกเปลี่ยนข้อมูลสูงสุด

จะเห็นได้ว่าการไขว้เปลี่ยนส่วนใหญ่จะมีการแลกเปลี่ยนคู่ยีนที่แตกต่างกันโดยเฉลี่ยประมาณ $0.25 I$ ในประชากรรุ่นแรก ๆ ที่มีความหลากหลายสูง แต่เมื่อเริ่มเกิดการลู่เข้าจำนวนคู่ยีนที่แตกต่างกันจะมีจำนวนลดลง ทำให้การแลกเปลี่ยนคู่ยีนที่ต่างกันมีปริมาณลดลงด้วย ถ้าพิจารณาในมุมมองนี้วิธีการไขว้เปลี่ยนส่วนใหญ่จะให้ผลเหมือนกัน แต่กลไกสำคัญที่ทำให้ไปถึงคำตอบคือ ความสามารถในการค้นหา ความสามารถในการรักษาให้คงอยู่ และความสามารถในการผสมคำตอบให้ดีขึ้น ซึ่งทั้งสามปัจจัยนี้จะส่งผลต่อประสิทธิภาพในการค้นหาคำตอบโดยตรง และแต่ละวิธีการไขว้เปลี่ยนจะมีคุณลักษณะของปัจจัยทั้งสามนี้แตกต่างกัน การไขว้เปลี่ยนแบบจุดเดียวมีความสามารถในการค้นหาต่ำ มีความสามารถในการรักษาให้คงอยู่สูง และมีความสามารถในการผสมต่ำ ส่วนการไขว้เปลี่ยนแบบหลายจุด (ตั้งแต่ 2 จุดขึ้นไป) ยังมีจุดตัดมาก

จะยังมีความสามารถในการค้นหาสูงขึ้น มีความสามารถในการรักษาให้คงอยู่ต่ำลง (เค้าร่างถูกทำลายได้ง่ายขึ้น) และมีความสามารถในการผสมที่สูงขึ้น แต่ไม่สามารถกล่าวได้ว่ามีความสามารถในการผสมที่ดีเนื่องจากการผสมเกิดขึ้นแบบสุ่ม (เริ่มตั้งแต่สุ่มจุดที่ต้องการจะตัดทุกจุดมีโอกาสถูกเลือกเท่ากันหมดทำให้ไม่สามารถกำหนดขอบเขตของการผสมได้แน่นอน) ส่วนวิธีการไขว้เปลี่ยนแบบขึ้นส่วนมีความสามารถในการค้นหาสูง มีความสามารถในการรักษาให้คงอยู่สูง และมีความสามารถในการผสมสูงและมีประสิทธิภาพ หรืออาจมองได้ว่าวิธีการไขว้เปลี่ยนแบบขึ้นส่วนที่น่าเสนอเป็นการนำเอาข้อดีด้านการสำรวจและการผสมมาจากการไขว้เปลี่ยนแบบหลายจุดรวมถึงการไขว้เปลี่ยนแบบเอกรูปด้วย และผสมเข้ากับความสามารถในการรักษาให้คงอยู่ซึ่งเป็นข้อดีของการไขว้เปลี่ยนแบบจุดเดียว แม้ว่าความสามารถบางอย่างจะมีความขัดแย้งหรือมีผลบั่นทอนกันอยู่บ้าง แต่ในภาพรวมแล้วยังคงรักษาจุดเด่นและข้อดีส่วนใหญ่เอาไว้ได้อย่างสมดุล

การไขว้เปลี่ยนส่วนใหญ่มักจะเป็นการดำเนินการเชิงควมคุม (Deterministic) ไม่มีความสามารถในการปรับตัว ในขณะที่ปัจจัยแวดล้อมอื่น ๆ ในระบบมีการปรับเปลี่ยนในรูปแบบพลวัต เช่น ความหลากหลายที่ลดลงจากประชากรรุ่นก่อน ๆ และการกระจายตัวของข้อมูลที่มีแบบแผนมากขึ้น (ไม่เป็นแบบสุ่มเหมือนประชากรรุ่นแรก ๆ) เป็นต้น แสดงให้เห็นว่าความสามารถในการปรับตัวเป็นสิ่งที่ดีและส่งผลให้ประสิทธิภาพในการหาคำตอบสูงขึ้น เมื่อพิจารณาการไขว้เปลี่ยนที่กำหนดจำนวนจุดตัดเป็นค่าคงที่ เช่นการไขว้เปลี่ยนแบบจุดเดียว การไขว้เปลี่ยนแบบหลายจุด รวมไปถึงการไขว้เปลี่ยนแบบเอกรูป จะเห็นว่าทั้งหมดเป็นตัวอย่างการที่ไม่มีความสามารถในการปรับตัว ซึ่งต่างจากวิธีการไขว้เปลี่ยนแบบขึ้นส่วนที่น่าเสนอที่มีการปรับตัวตามรูปแบบโครงสร้างและการกระจายตัวของข้อมูลอยู่ตลอดเวลา ในช่วงประชากรรุ่นแรก ๆ ที่มีความหลากหลายสูง (โครโมโซมแต่ละตัวมีรูปแบบหน้าตาแตกต่างกันมาก) การไขว้เปลี่ยนแบบขึ้นส่วนจะทำให้เกิดการสำรวจในอัตราที่สูงตามไปด้วย (พฤติกรรมคล้ายการไขว้เปลี่ยนแบบเอกรูป) และเมื่อเริ่มมีการลู่เข้า (รูปแบบคู่จิ้นในโครโมโซมเริ่มเหมือนกันมากขึ้น) ความสามารถในการสำรวจจะเริ่มถูกเปลี่ยนเป็นการแสวงหาประโยชน์มากขึ้น เนื่องจากโครงสร้างที่ดี ๆ เริ่มปรากฏชัดเจนขึ้นและสามารถรักษาให้คงอยู่ต่อไปได้ และเมื่อประชากรส่วนใหญ่เกิดการลู่เข้าสูงขึ้น (คู่จิ้นในประชากรมีความแตกต่างกันน้อยมาก) วิธีที่น่าเสนอก็ยังคงสามารถทำให้เกิดการแลกเปลี่ยนข้อมูลในส่วนต่างที่น้อยได้อยู่เสมอ เช่น ถ้าพ่อพันธุ์แม่พันธุ์มีคู่จิ้นต่างกัน 2 ตำแหน่งที่อยู่ใกล้กัน ถ้าเป็นการไขว้เปลี่ยนแบบจุดเดียวการไขว้เปลี่ยนจะทำให้คู่จิ้นทั้งสองตำแหน่งไปด้วยกันเสมอแทนที่จะแยกกัน (มีโอกาสแยกจากกันน้อยมาก) แต่ถ้าเป็นวิธีการไขว้เปลี่ยนแบบขึ้นส่วนคู่จิ้นทั้งสองมีโอกาสที่จะเกิดการแลกเปลี่ยนเป็นอิสระจากกันทำให้โอกาสที่คู่จิ้นทั้งสองจะแยกจากกันมีสูงขึ้น เป็นต้น จะเห็นว่าลักษณะการปรับตัวของวิธีการไขว้เปลี่ยนแบบขึ้นส่วนมีความสอดคล้อง

กับหลักการที่ใช้หาคำตอบ (การสำรวจและการแสวงหาประโยชน์ในช่วงเวลาต่าง ๆ) และเข้ากันได้ดีกับการเปลี่ยนแปลงแบบพลวัต

จากขั้นตอนวิธีที่ออกแบบสามารถคำนวณหาโอกาสที่ขึ้นส่วนจะอยู่รอดจากการดำเนินกรต่าง ๆ เพื่อให้เห็นภาพรวมของกลไกการทำงานอย่างคร่าว ๆ การคำนวณจะพิจารณาเพียงแค่รุ่นประชากรเดียว (จากรุ่นสู่รุ่น) จากรุ่น t ไป $t+1$ ดังต่อไปนี้

สมมติให้เค้าร่างที่หาได้จากการเปรียบเทียบสองโครโมโซมตามวิธีการไขว้เปลี่ยนแบบขึ้นส่วนเป็น 1^*01^*101 เราสามารถหาโอกาสที่เค้าร่างนี้จะเกิดได้ดังนี้

$$P_{identify}(H) = 2^{-(D(H) + o(H) + 1)} \quad (5.6)$$

เมื่อ $P_{identify}(H)$ คือ โอกาสเกิดเค้าร่าง H จากวิธีการระบุขึ้นส่วน

$D(H)$ คือ ความยาวของเค้าร่าง H

$o(H)$ คือ ลำดับของเค้าร่าง H

จากเค้าร่าง 1^*01^*101 จะมีค่า $D(H) = 10-1 = 9$ และ $o(H) = 6$ ดังนั้นโอกาสเกิดเค้าร่างจะมีค่าเป็น $2^{-(9+6+1)} = 2^{-16} = 1/65,536 = 0.0000152587890625$

จะเห็นว่าเค้าร่างนี้ประกอบด้วยขึ้นส่วนตามที่นิยามไว้แล้วจำนวน 3 ขึ้นส่วนได้แก่ (1), (01) และ (101) ดังนั้นจะสามารถหาโอกาสรอดของเค้าร่างทั้งสามนี้จากการไขว้เปลี่ยนแบบขึ้นส่วนได้เป็น

$$P_{compose}(H) = (1 - P_c)^{|S|} \quad (5.7)$$

เมื่อ $P_{compose}(H,t)$ คือ โอกาสรอดของเค้าร่าง H จากวิธีการไขว้เปลี่ยนแบบขึ้นส่วน

P_c คือ โอกาสเกิดการไขว้เปลี่ยน

S คือ เซตของขึ้นส่วนในเค้าร่าง H ที่ช่วงตำแหน่งหนึ่ง ๆ

จากสมการสามารถคำนวณโอกาสรอดของเค้าร่างเมื่อ $|S| = 3$ และกำหนดให้ $P_c = 0.5$ ได้เป็น $(1-0.5)^3 = 0.5^3 = 0.125$

และจากสองสมการด้านบนทำให้สามารถคำนวณหาโอกาสที่เค้าร่าง H ใด ๆ จะมีโอกาสเกิดขึ้นและอยู่รอดได้ดังนี้

$$P_{identify \text{ and } compose}(H) = 2^{-(D(H) + o(H) + 1)} \cdot (1 - P_c)^{|S|} \quad (5.8)$$

เมื่อ $P_{identify\ and\ compose}(H)$ คือ โอกาสที่จะเกิดขึ้นและอยู่รอดของเค้าร่าง H จากวิธีการไขว้เปลี่ยนแบบชิ้นส่วน

เมื่อพิจารณาถึงตัวดำเนินการเลือกสรรที่ใช้เป็นแบบการแข่งขันกีฬา ดังนั้นสามารถหาความสัมพันธ์ของสัดส่วนความอยู่รอดของจำนวนโครโมโซมที่มีเค้าร่าง H ได้เป็น

$$P_{select}(H) = 1 - (1 - P_{(H)})^R \quad (5.9)$$

เมื่อ $P_{select}(H)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากรที่ถูกเลือก
 $P(H)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากร
 R คือ ลำดับของการแข่งขัน (Rank) (จำนวนประชากรที่ถูกเลือกเพื่อเข้าร่วมการแข่งขัน)

ถ้าพิจารณาในมุมมองเดียวกันกับทฤษฎีบทเค้าร่างจะได้สมการเป็น

$$P(H, t+1) = P_{select}(H, t) \cdot P_{identify\ and\ compose}(H, t) \quad (5.10)$$

เมื่อ $P(H, t+1)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากรในรุ่นที่ $t+1$
 $P_{select}(H, t)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากร ที่ถูกเลือกจากรุ่นที่ t
 $P_{identify\ and\ compose}(H)$ คือ โอกาสที่จะเกิดขึ้นและอยู่รอดของเค้าร่าง H จากวิธีการไขว้เปลี่ยนแบบชิ้นส่วนจากรุ่นที่ t

5.2 การทดลองและผลการทดลอง

ในส่วนนี้จะกล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่งในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่นำเสนอ (มีการเปรียบเทียบกับผลของขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย) และมีการวิจารณ์ผลการทดลองในมุมมองที่นำเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ

ขั้นตอนวิธี: ขั้นตอนวิธีเชิงพันธุกรรม

การเลือกสรร: วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4

การไขว้เปลี่ยน: การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0

การผ่าเหล่า: ไม่มี

วิธีที่นำเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่ใช้การไขว้เปลี่ยนแบบชิ้นส่วน
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	วิธีข้อมูลสหเพศร่วมกันและไม่ร่วมกันต่ำสุดแบบชิ้นส่วน
การประกอบ:	การไขว้เปลี่ยนแบบชิ้นส่วน (การดำเนินการเหมือนการไขว้เปลี่ยนแบบหลายจุด) มีอัตราการใช้ไขว้เปลี่ยนเท่ากับ 0.5
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 5.1

ตารางที่ 5.1 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ sGA-FC

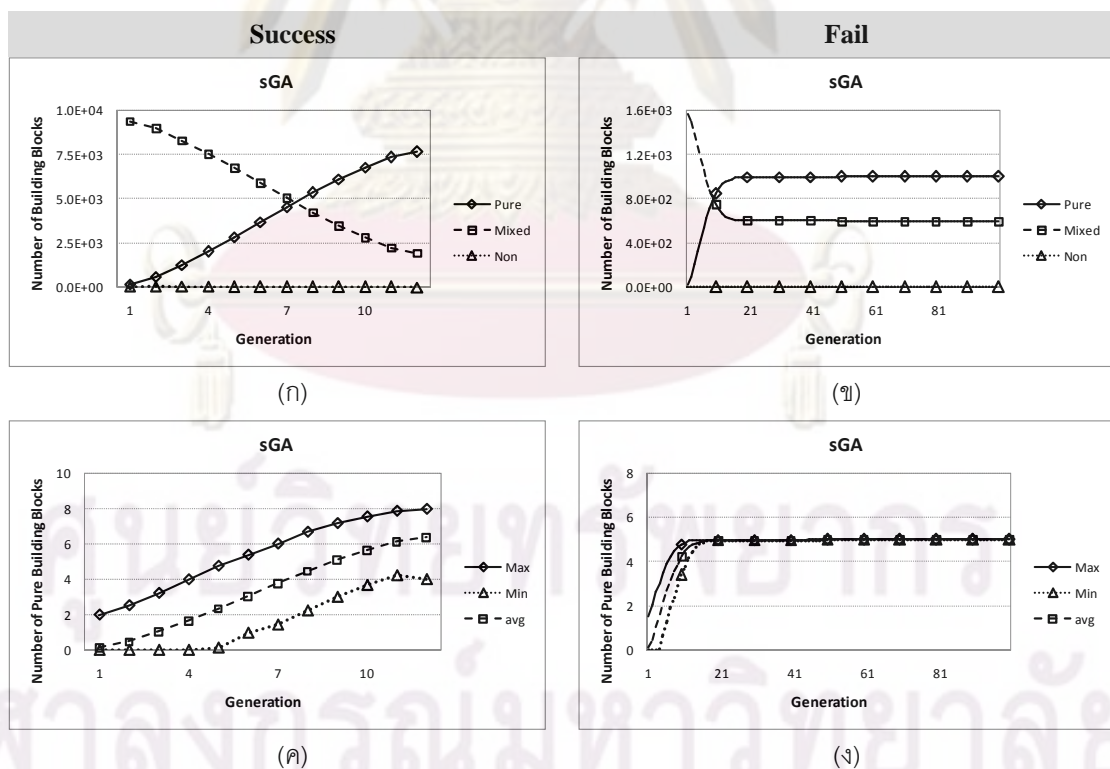
Parameters		Success 30 run				Fail 30 run	
		sGA		sGA-FC		sGA	sGA-FC
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,200	11,920	1,200	13,480	200	100
Trap-5	60	2,300	28,367	1,600	20,854	300	300

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

เมื่อพิจารณาจากผลการทดลองในส่วนของจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมจะเห็นได้ว่าวิธี sGA-FC ดีกว่า sGA ในปัญหา Trap-5 แต่แยกว่าในปัญหา Royal Road ส่วนขนาดของประชากรวิธี sGA-FC ใช้จำนวนเท่ากับ sGA ในปัญหา Royal Road และใช้จำนวนน้อยกว่าในปัญหา Trap-5 ส่วนกรณีหาคำตอบไม่พบเป็นการแสดงให้เห็นถึงประสิทธิภาพในอีกมุมมองหนึ่ง จากกรณีที่หาคำตอบไม่พบขั้นตอนวิธี sGA-FC ใช้จำนวนประชากรน้อยกว่า sGA ในปัญหา Royal Road ซึ่งหมายความว่าถ้าใช้จำนวนประชากรเท่ากับ sGA จะสามารถหาคำตอบเจออย่างน้อย 1 ครั้งจากการทดสอบ 30 ครั้ง เป็นการแสดงให้เห็นถึงประสิทธิภาพในทางอ้อมโดยใช้พิจารณาควบคู่กับกรณีที่หาคำตอบพบเพื่อเป็นการตรวจสอบการวินิจฉัยผลว่าถูกต้องหรือไม่ (เนื่องจากใช้จำนวนปัญหาทดสอบน้อย) จากกรณีนี้ถ้าพิจารณาเพียงกรณีที่หาคำตอบพบจะเห็นว่า sGA-FC ใช้จำนวนประชากรเท่ากับ sGA คือ 1,200 ตัว ซึ่งอาจทำให้เข้าใจว่าทั้งสองวิธีมีประสิทธิภาพในส่วนของการใช้ทรัพยากรเท่ากัน สำหรับปัญหาที่หน่วยการ

สร้างไม่มีการลวง หรืออาจมองได้ว่ากรณีที่ล้มเหลวทั้ง 30 ครั้งเป็นตัวบ่งบอกขอบเขตล่างของประสิทธิภาพการทำงาน ส่วนกรณีที่สำเร็จครบ 30 ครั้งเป็นตัวบ่งบอกขอบเขตบน

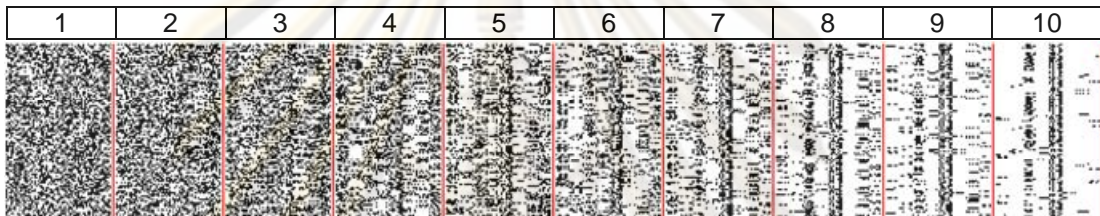
งานวิจัยนี้มุ่งเน้นศึกษาในเรื่องของหน่วยการสร้างเป็นหลักเพื่อที่จะทำความเข้าใจกับกลไกการวิวัฒนาการในขั้นตอนวิธีเชิงพันธุกรรม โดยอาศัยการเปรียบเทียบผลการดำเนินการระหว่างขั้นตอนวิธีที่ต่างกัน และกรณีที่ต่างกัน (กรณีหาค่าตอบพบ และหาค่าตอบไม่พบ) เพื่อให้เห็นภาพได้ชัดเจนมากขึ้น จากรูปที่ 5.3 เป็นผลที่ได้จาก sGA เมื่อพิจารณาปริมาณหน่วยการสร้างในกรณีที่หาค่าตอบพบ (5.3 ก) จะเห็นว่าการแข่งขันกันของ Pure BBs และ Mixed BBs เท่านั้น ส่วน Non BBs มีค่าในระดับที่ต่ำมากซึ่งเป็นการแสดงให้เห็นว่าปัญหานี้ไม่มีการลวงเกิดขึ้น กรณีที่หาค่าตอบไม่พบ (5.3 ข) จำนวน Pure BBs จะมีการเพิ่มขึ้นอย่างต่อเนื่อง และจำนวน Mixed BBs จะลดลงอย่างต่อเนื่องสวนทางกันเนื่องจากเริ่มมีการลู่เข้าสู่คำตอบที่ดีที่สุด ส่วนกรณีที่หาค่าตอบไม่พบจะเห็นว่าการแข่งขันกันระหว่าง Pure BBs และ Mixed BBs เกิดขึ้นอย่างรุนแรงในประชากรรุ่นต้น ๆ (จนถึงรุ่นที่ 10) และลดความรุนแรงลง (จนถึงรุ่นที่ 21) และมีค่าคงที่ไปตลอด ที่เป็นเช่นนี้เนื่องจากมีจำนวนประชากรน้อยดังนั้นจึงเกิดการลู่เข้าอย่างรวดเร็วเพราะขาดความหลากหลายของประชากร และเมื่อพิจารณาจำนวน Pure BBs ที่มีอยู่ในแต่ละโครโมโซมจะเห็นว่า



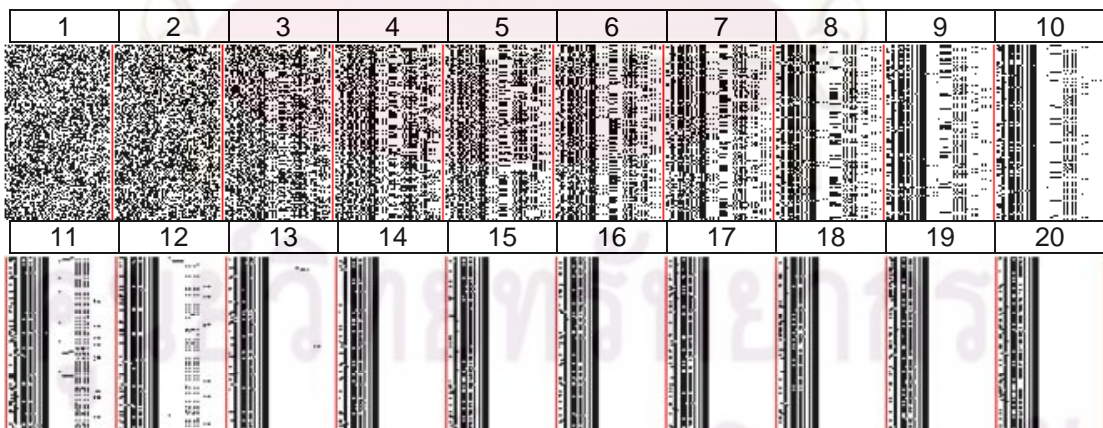
รูปที่ 5.3 ผลการทดสอบของ sGA ในปัญหา Royal Road 64 บิต ปริมาณหน่วยสร้างกับรุ่นประชากร (ก) (ข) และจำนวนหน่วยสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ค) (ง)

ในกรณีที่หาคำตอบพบ (5.3 ค) จะมีค่าเพิ่มขึ้นอย่างต่อเนื่องจนครบทุกหน่วย ส่วนกรณีที่หาคำตอบไม่พบ (3.3 ง) จำนวน Pure BBs จะเพิ่มขึ้นจนถึงระดับหนึ่ง (จนถึงรุ่นที่ 10) และจะมีค่าคงที่ซึ่งเป็นผลจากการลู่อเข้าของประชากรสอดคล้องกับรูป 5.3 (ข)

เมื่อพิจารณาถึงเรื่องการกระจายตัวของคู่จิ้นที่มีค่าเป็น 0 กับ 1 ในแต่ละโครโมโซม จากรูปที่ 5.4 กรณีที่ sGA หาคำตอบพบจะเห็นว่าจากประชากรรุ่นแรกซึ่งมีการกระจายตัวของค่า 0 กับ 1 แบบเอกรูป ซึ่งมีปริมาณของคู่จิ้นที่เป็น 0 และ 1 พอ ๆ กันเมื่อเวลาผ่านไปจะเห็นว่าการกระจายตัวเริ่มลู่อเข้าสู่รูปแบบหนึ่งอย่างเห็นได้ชัด และมีจำนวนคู่จิ้นที่เป็น 1 (พื้นที่สีขาว) มากกว่าจำนวนคู่จิ้นที่เป็น 0 (พื้นที่สีดำ) จนสามารถหาโครโมโซมที่มีค่าเป็น 1 ครบทุกคู่จิ้นซึ่งเป็นคำตอบได้



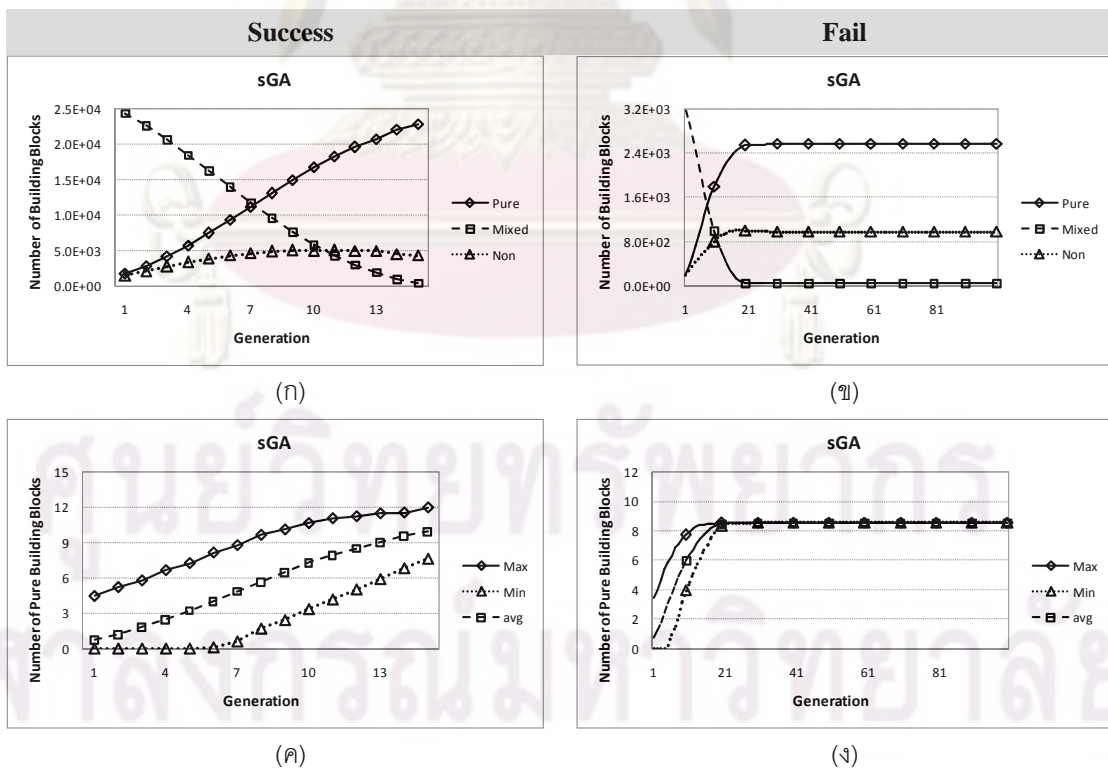
รูปที่ 5.4 รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0



รูปที่ 5.5 รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0

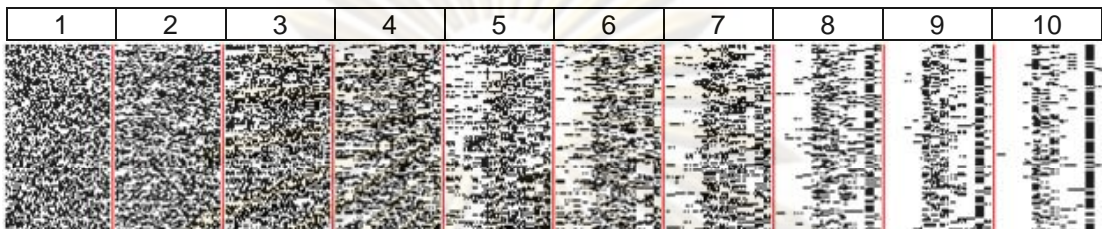
ส่วนกรณีที่ไม่สามารถหาคำตอบพบ (ใช้จำนวนประชากรน้อย) ดังรูปที่ 5.5 จะเห็นว่าเกิดการลู่เข้าสู่รูปแบบการกระจายตัวแบบหนึ่งเช่นกัน แต่มีความชัดเจนมากกว่าเห็นว่ตั้งแต่รุ่นประชากรที่ 13 เป็นต้นไปรูปแบบการกระจายตัวแทบจะไม่เปลี่ยนแปลงไปจากเดิมเนื่องจากประชากรทั้งหมดลู่เข้าสู่ค่าเหมาะสมที่สุดท้องถิ่นทำให้ไม่สามารถไปถึงคำตอบได้

สำหรับปัญหา Trap-5 ซึ่งมีการลวงในหน่วยการสร้างจากรูปที่ 5.6 จะเห็นว่ามึลักษณะโดยรวมแล้วเหมือนกับรูปที่ 5.3 แต่ต่างกันตรงที่ปริมาณของ Non BBs (5.3 ก) มีการแข่งขันกับ Pure BBs และ Mixed BBs แม้ว่าจะมีอัตราการแข่งขันไม่สูงนักก็ตามซึ่งต่างจากปัญหา Royal Road อย่างชัดเจน ส่วนนี้เป็นผลที่เกิดขึ้นจากการลวงที่มีอยู่ในหน่วยการสร้างนั่นเอง และในกรณีที่ไม่หาคำตอบไม่พบ (5.3 ข) จะเห็นว่าปริมาณของ Non BBs มีค่าสูงกว่า Mixed BBs ซึ่งแสดงให้เห็นว่าถ้ามีความหลากหลายไม่เพียงพอ (เนื่องจากจำนวนประชากรน้อย) จะทำให้ไม่สามารถแยกแยะระหว่างรูปแบบหน่วยการสร้างที่ดี และรูปแบบหน่วยการสร้างลวงได้อย่างมีประสิทธิภาพ และในปัญหา Trap-5 นี้จะเห็นว่าประชากรจะลู่เข้าชัดเจนในประชากรรุ่นที่ 20 ส่วนปัญหา Royal Road จะลู่เข้าในประชากรรุ่นที่ 10 ที่เป็นเช่นนี้เพราะปัญหา Trap-5 ใช้จำนวนประชากรมากกว่าปัญหา Royal Road จึงทำให้เกิดการลู่เข้าที่ช้ากว่า ส่วนหลักการอื่น ๆ สามารถอธิบายได้แบบเดียวกันกับรูปที่ 5.3

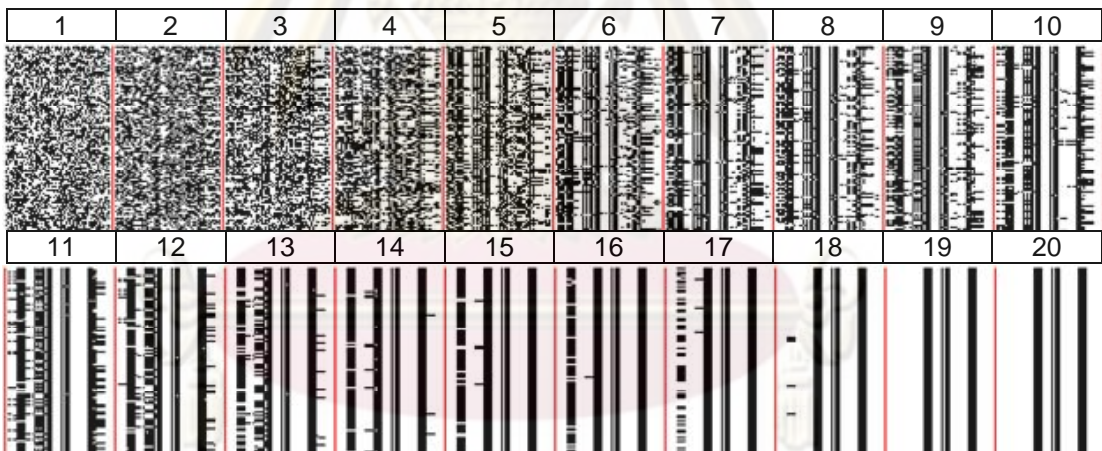


รูปที่ 5.6 ผลการทดสอบของ sGA ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) และจำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ค) (ง)

การกระจายตัวของคู้เงินที่เป็น 0 กับ 1 จากรูปที่ 5.7 แสดงให้เห็นว่ามีการลู่เข้าสู่รูปแบบหนึ่ง แต่สิ่งที่ต่างกันคือลักษณะของกลุ่มพื้นที่สีดำที่จะติดกันเป็นก้อนในแนวนอนที่ยาวกว่าเมื่อเปรียบเทียบกับรูปที่ 5.4 ซึ่งเกิดจากรูปแบบปัญหา Trap-5 ที่หน่วยการสร้างแบบวงล้อให้ติดอยู่กับค่าเหมาะสมที่สุดท้องถิ่นที่มีจำนวนคู้เงินที่เป็น 0 ติดกันจำนวน 4 บิต แต่ในท้ายสุดแล้วก็ยังสามารถหาคำตอบได้



รูปที่ 5.7 รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู้เงินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู้เงินที่มีค่าเป็น 0



รูปที่ 5.8 รูปแบบจีโนไทป์ของประชากรของ sGA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู้เงินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู้เงินที่มีค่าเป็น 0

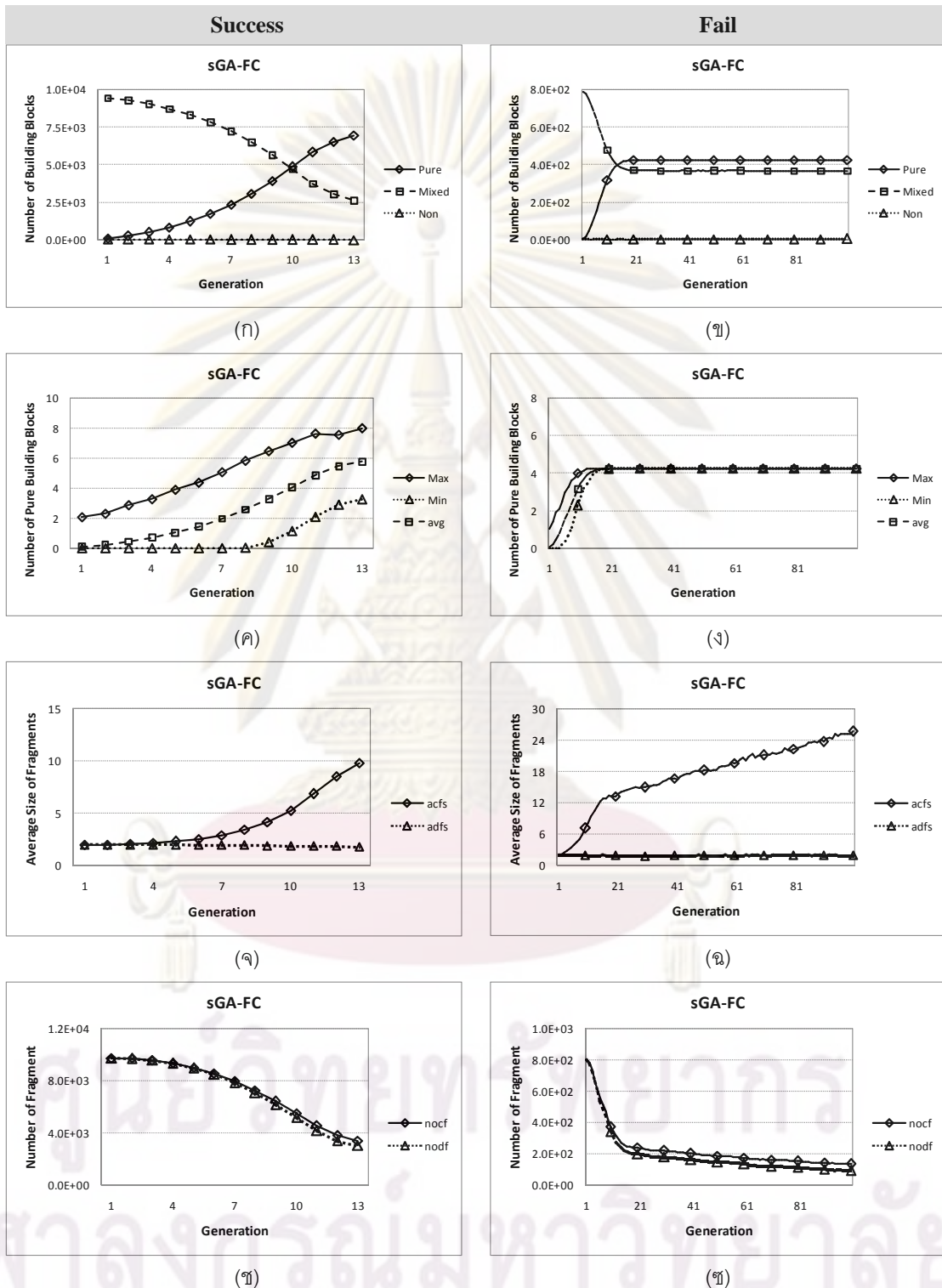
ในกรณีที่ไม่สามารถหาคำตอบพบดังรูปที่ 5.8 จะเห็นว่ารูปแบบของหน่วยการสร้างแบบวงล้อจะเกิดขึ้นจำนวนมากและเกิดขึ้นซ้ำที่ตำแหน่งเดียวกันในหลายโครโมโซม (จะเห็นเป็นริ้วสีดำในแนวตั้ง) ซึ่งเริ่มชัดเจนตั้งแต่รุ่นประชากรที่ 6 และชัดเจนขึ้นเรื่อย ๆ จนถึงรุ่นประชากรที่ 14 ซึ่งมีแถบดำถึง 4 แถบหลัก ๆ และลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 19 (เหลือแถบสีดำ 3 แถบ

หลัก) แถบดำเหล่านี้เป็นตัวแสดงถึงการลู่เข้าสู่ค่าเหมาะสมที่สุดท้องถิ่นในช่วงตำแหน่งหนึ่ง ๆ ของโครโมโซม ซึ่งถ้าแถบบ้านี้มีขนาดใหญ่และมีความชัดเจนต่อเนื่อง นั้นหมายความว่าโอกาสที่จะทำลายแถบบ้านี้เพื่อไปให้ถึงคำตอบจะลดลงอย่างมาก แถบดำเหล่านี้เป็นตัวบ่งชี้ถึงความหลากหลายที่มีอยู่ในประชากรโดยตรง การที่จะไปให้ถึงคำตอบนั้นจำเป็นที่จะต้องรักษาความหลากหลายเอาไว้ในระดับที่เพียงพอจนกว่าจะได้คำตอบที่ต้องการ (พยายามไม่ให้เกิดแถบสีดำ)

ส่วนต่อไปนี้เป็นผลการทดสอบของขั้นตอนวิธีเชิงพันธุกรรมที่ใช้การไขว้เปลี่ยนแบบชิ้นส่วน ซึ่งมีการแสดงผลในส่วนของขนาดและปริมาณของชิ้นส่วนเพิ่มเติมจาก sGA จากรูปที่ 5.9 ในส่วน (ก-ง) จะเห็นว่ามียูนิฟอร์มเดียวกันกับรูปที่ 5.3 (ก-ง) แต่มีความแตกต่างกันในรายละเอียดบางส่วน เมื่อพิจารณาถึงการแข่งขันกันระหว่าง Pure BBs กับ Mixed BBs จะเห็นว่าในรูปที่ 5.9 (ก) เส้นกราฟตัดกันในรุ่นประชากรที่ 10 ซึ่งต่างจากวิธี sGA ในรูป 5.3 (ก) ที่เส้นกราฟตัดกันในรุ่นประชากรที่ 7 และเมื่อพิจารณาร่วมกับจำนวนของ Pure BBs จากรูป 5.9 (ค) จะเห็นว่าจำนวนหน่วยการสร้างน้อยที่สุดเริ่มขยับจากค่าต่ำสุด (ค่า 0) ในรุ่นประชากรที่ 9 ซึ่งต่างจากวิธี sGA รูปที่ 5.3 (ค) ที่เริ่มขยับจากค่าต่ำสุดในรุ่นประชากรที่ 5 หรือพิจารณาจากพื้นที่ใต้กราฟระหว่างจำนวนหน่วยการสร้างมากที่สุดและน้อยที่สุด ซึ่งเป็นการแสดงให้เห็นว่าวิธี sGA-FC มีความสามารถในการรักษาระดับความหลากหลายที่สูงกว่าวิธี sGA เนื่องจากพื้นที่ใต้กราฟมีขนาดใหญ่กว่าจึงทำให้เกิดการลู่เข้าช้าลง และมีผลให้จำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมสูงขึ้น (เป็นผลที่ไม่ดี) และเมื่อพิจารณากรณีที่หาคำตอบไม่พบจากรูป 5.9 (ข) จะเห็นว่าจำนวน Pure BBs มีค่าต่างจาก Mixed BBs น้อยมากถ้าเทียบกับรูปที่ 5.3 (ข) เนื่องจาก sGA-FC มีความหลากหลายสูงกว่าจึงทำให้มีปริมาณ Mixed BBs ใกล้เคียงกับ Pure BBs มากยิ่งขึ้น (เป็นการยืนยันในส่วนของความสามารถในการรักษาความหลากหลาย)

ในส่วนของชิ้นส่วนจากรูปที่ 5.9 (จ-ฉ) กรณีที่หาคำตอบพบ (5.9 จ) จะเห็นว่าขนาดโดยเฉลี่ยของชิ้นส่วนร่วมกัน (acfs) (ชิ้นส่วนที่เหมือนกันระหว่าง 2 โครโมโซม) มีค่าเพิ่มขึ้นอย่างต่อเนื่อง ส่วนขนาดโดยเฉลี่ยของชิ้นส่วนไม่ร่วมกัน (adfs) (ชิ้นส่วนที่ต่างกันระหว่าง 2 โครโมโซม) มีค่าเกือบคงที่ ซึ่งหมายความว่าเกิดการลู่เข้าอย่างต่อเนื่องทำให้ชิ้นส่วน หรือหน่วยการสร้างมีขนาดใหญ่ขึ้น ในกรณีที่หาคำตอบไม่พบ (5.9 ฉ) ขนาดของชิ้นส่วนร่วมกันจะโตขึ้นอย่างรวดเร็วเนื่องจากการลู่เข้าเกิดขึ้นอย่างรุนแรง (สูญเสียความหลากหลายอย่างรวดเร็ว) และเมื่อพิจารณาร่วมกับรูปที่ 5.9 (ซ-ช) จะเห็นว่าปริมาณของชิ้นส่วนทั้งสองแบบ (ร่วมกันและไม่ร่วมกัน) ลดลงอย่างต่อเนื่องและมีปริมาณใกล้เคียงกันในทั้งสองกรณี แต่ในกรณีที่หาคำตอบไม่พบปริมาณชิ้นส่วนทั้งสองแบบจะลดลงอย่างรวดเร็วจนถึงรุ่นประชากรที่ 21 (และค่อย ๆ ลดอย่างต่อเนื่อง) เนื่องจากการสูญเสียความหลากหลายอย่างรวดเร็วเป็นไปตามรูปที่ 5.9 (ง) ในภาพรวมจะเห็นว่า

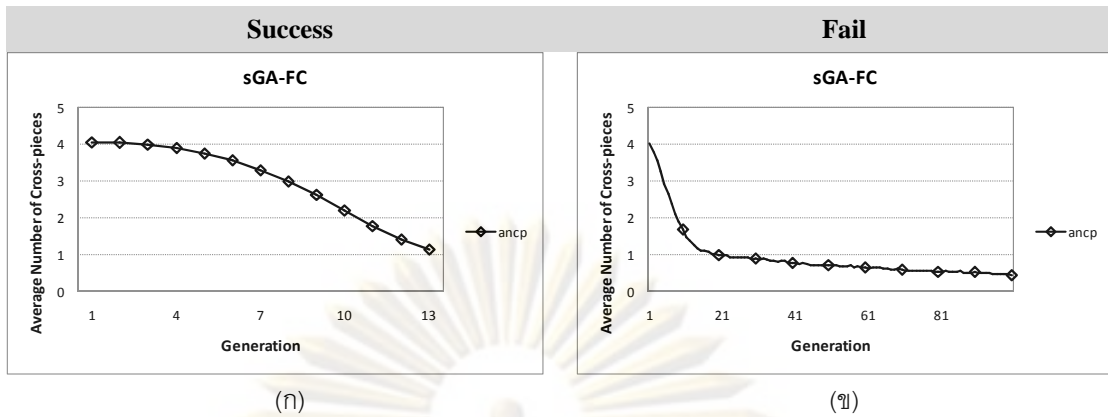
การพิจารณาลักษณะเฉพาะของขั้นตอนวิธีที่นำเสนอในรูปแบบชิ้นส่วนสอดคล้องกับการพิจารณาในรูปแบบของหน่วยการสร้าง



รูปที่ 5.9 ผลการทดสอบของ sGA-FC ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับ
 รุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง)
 ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

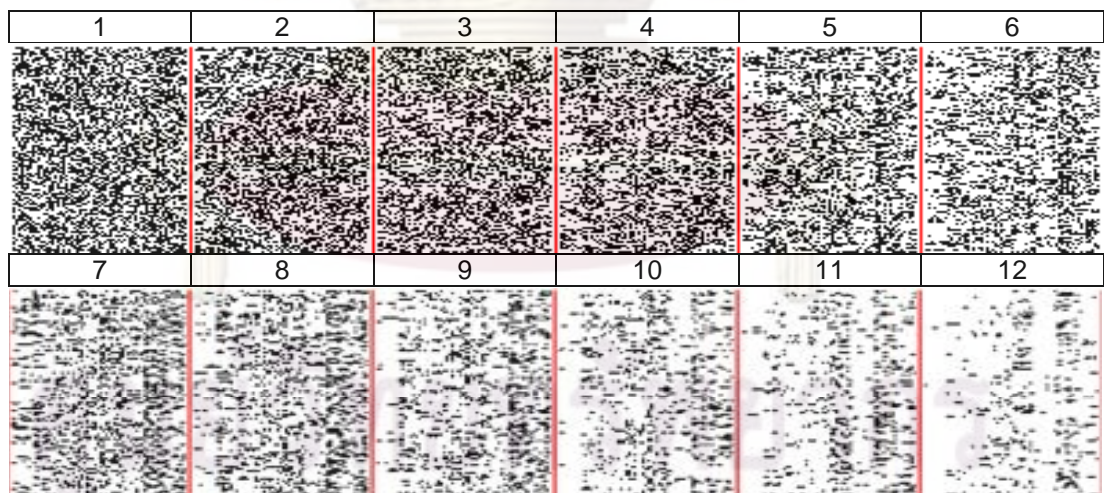
หมายเหตุ อาจมีคำถามว่าทำไมขึ้นส่วนร่วมกันมีขนาดใหญ่ขึ้นเรื่อย ๆ แต่ทำไมปริมาณของขึ้นส่วนร่วมกันและไม่ร่วมกันยังคงมีค่าใกล้เคียงกัน ที่เป็นเช่นนี้เนื่องจากการเกิดขึ้นส่วนที่ร่วมกันจำนวน 2 ชั้นในโครโมโซมเดียวกัน จะเกิดขึ้นได้ก็ต่อเมื่อมีขึ้นส่วนไม่ร่วมกันคั่นอยู่ระหว่างกลาง 1 ชั้น ในทำนองเดียวกันถ้ามีขึ้นส่วนร่วมกันจำนวน 10 ชั้นในโครโมโซมเดียวกันจะต้องมีขึ้นส่วนไม่ร่วมกันจำนวนอย่างน้อย 9 ชั้น (ชั้นระหว่างขึ้นส่วนร่วมกันแต่ละคู่) เป็นต้น

อาจมีความสงสัยว่าเมื่อใช้การไขว้เปลี่ยนแบบขึ้นส่วนซึ่งมีการปรับตัวแบบอัตโนมัติ (ปรับจำนวนจุดไขว้เปลี่ยนได้เองขึ้นกับรูปแบบคู่จีนในโครโมโซม) แล้วจะเกิดการไขว้เปลี่ยนจำนวนเท่าใดในระหว่างการดำเนินการ จากรูปที่ 5.10 เป็นกราฟแสดงถึงจำนวนขึ้นส่วนโดยเฉลี่ยต่อโครโมโซมที่มีการแลกเปลี่ยนกันในกระบวนการไขว้เปลี่ยนแบบขึ้นส่วน (ด้วยอัตราการไขว้เปลี่ยนเท่ากับ 0.5) กรณีที่หาคำตอบพบจะเห็นว่าในช่วงรุ่นประชากรแรก ๆ จะมีการแลกเปลี่ยนเกิดขึ้นสูงและค่อย ๆ ลดลง ซึ่งถ้าพิจารณาคู่กับรูปที่ 5.9 (ข) แล้วจะพบว่ามีลักษณะกราฟเหมือนกันเนื่องจากเมื่อมีขึ้นส่วนจำนวนมากโอกาสที่จะเกิดการไขว้เปลี่ยนก็จะยิ่งสูงทำให้เกิดการแลกเปลี่ยนขึ้นส่วนกันมาก ในทำนองเดียวกันเมื่อมีปริมาณขึ้นส่วนน้อยลงการแลกเปลี่ยนก็จะเกิดขึ้นน้อยลงด้วยเช่นกัน การแลกเปลี่ยนที่เกิดขึ้นในการไขว้เปลี่ยนแบบขึ้นส่วนเป็นการแลกเปลี่ยนเฉพาะขึ้นส่วนที่ต่างกันซึ่งจากรูปที่ 5.9 (จ) จะเห็นว่ามีขนาดโดยเฉลี่ยเกือบคงที่ตลอดการทำงานเนื่องจากตอนเริ่มต้นประชากรมีความแตกต่างกันมากทำให้ขึ้นส่วนร่วมกันและไม่ร่วมกันมีขนาดเล็ก และเมื่อเริ่มเกิดการลู่เข้าประชากรส่วนใหญ่จะมีรูปแบบหน้าตาเหมือนกันมากขึ้นทำให้ขึ้นส่วนร่วมกันมีขนาดใหญ่ขึ้น ในขณะที่ขึ้นส่วนไม่ร่วมกันจะยังคงมีจำนวนใกล้เคียงรุ่นก่อนหน้าแต่มีขนาดเล็กลง ถ้าพิจารณาในแง่มุมของการสำรวจและการแสวงหาประโยชน์อาจกล่าวได้ว่าจากรูปที่ 5.10 (ก) ในช่วงรุ่นประชากรแรก ๆ มีการแลกเปลี่ยนขึ้นส่วนกันมากเป็นการแสดงถึงระดับของการสำรวจที่สูง และค่อย ๆ ลดลงอย่างต่อเนื่องเพราะเกิดการลู่เข้าสู่คำตอบ ซึ่งถ้าเปรียบเทียบกับรูปที่ 5.10 (ข) กรณีหาคำตอบไม่พบจะเห็นว่าระดับของการสำรวจลดลงอย่างรวดเร็ว จากข้อเปรียบเทียบนี้ทำให้สามารถกล่าวได้ว่าการไปให้ถึงคำตอบนั้นจำเป็นที่จะต้องรักษาระดับการสำรวจเอาไว้ให้มีการเปลี่ยนแปลงแบบค่อยเป็นค่อยไปเพื่อให้สามารถแสวงหาประโยชน์ได้อย่างเต็มที่ หรือกล่าวได้อีกนัยหนึ่งว่าเราจำเป็นที่จะต้องรักษาระดับความหลากหลายเอาไว้ให้มีอยู่เพียงพอต่อการวิวัฒนาการ หรืออาจมองได้ว่าจะมีความหลากหลายมากขึ้นก็จะทำให้ยังมีสิทธิ์เลือก (รูปแบบที่ดี ๆ) มากขึ้นนั่นเอง จากการทดลองเราควบคุมความหลากหลายด้วยขนาดของประชากร กรณีที่หาคำตอบไม่พบเป็นเพราะมีจำนวนประชากรน้อยเกินไปทำให้เกิดการลู่เข้าอย่างรวดเร็ว



รูปที่ 5.10 จำนวนชิ้นส่วนเฉลี่ยต่อโครโมโซมที่เกิดการไขว้เปลี่ยนกันกับรุ่นประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต

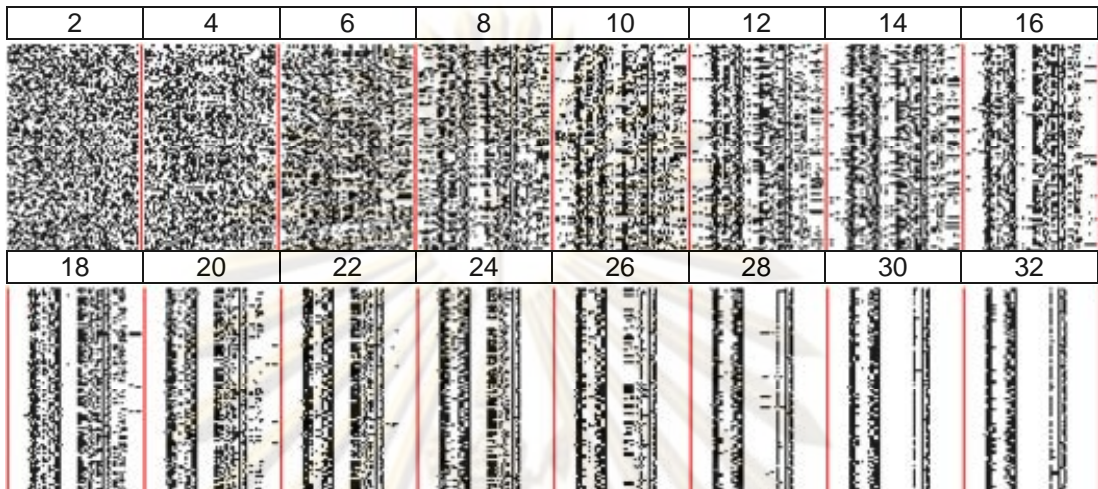
เมื่อพิจารณารูปแบบคู่จิ้นในประชากร จากรูปที่ 5.11 ซึ่งเป็นกรณีที่หาคำตอบพบจะเห็นว่าในหมู่ประชากรมีการกระจายตัวของคู่จิ้นที่เป็น 0 และ 1 ดีกว่าวิธี sGA ในรูปที่ 5.7ซึ่งเห็นเป็นริ้วแถบสีดำที่ชัดเจนกว่า การกระจายตัวที่เกิดขึ้นจากการไขว้เปลี่ยนแบบชิ้นส่วนจะส่งผลในทุกรุ่นประชากรทำให้คู่จิ้นที่เป็น 0 (สีดำ) แม้จะมีจำนวนลดลงเพราะเกิดการลู่เข้าสู่คำตอบที่เป็น 1 (สีขาว) แต่ก็ยังมีลักษณะการกระจายตัวในบริเวณกว้างซึ่งทำให้เกิดการลู่เข้าที่ช้าลง (ช้ากว่า sGA) แต่ถ้ามองในแง่ดีจะเห็นว่าส่งผลให้มีระดับการสำรวจสูงขึ้น และจะเห็นได้ชัดขึ้นเมื่อพิจารณาร่วมกับกรณีที่หาคำตอบไม่พบดังรูปที่ 5.12 จะเห็นได้ว่าการลู่เข้าอย่างสมบูรณ์จะเกิดขึ้น



รูปที่ 5.11 รูปแบบจิ้นไทป์ของประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 12 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่

สีขาว □ คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จิ้นที่มีค่าเป็น 0

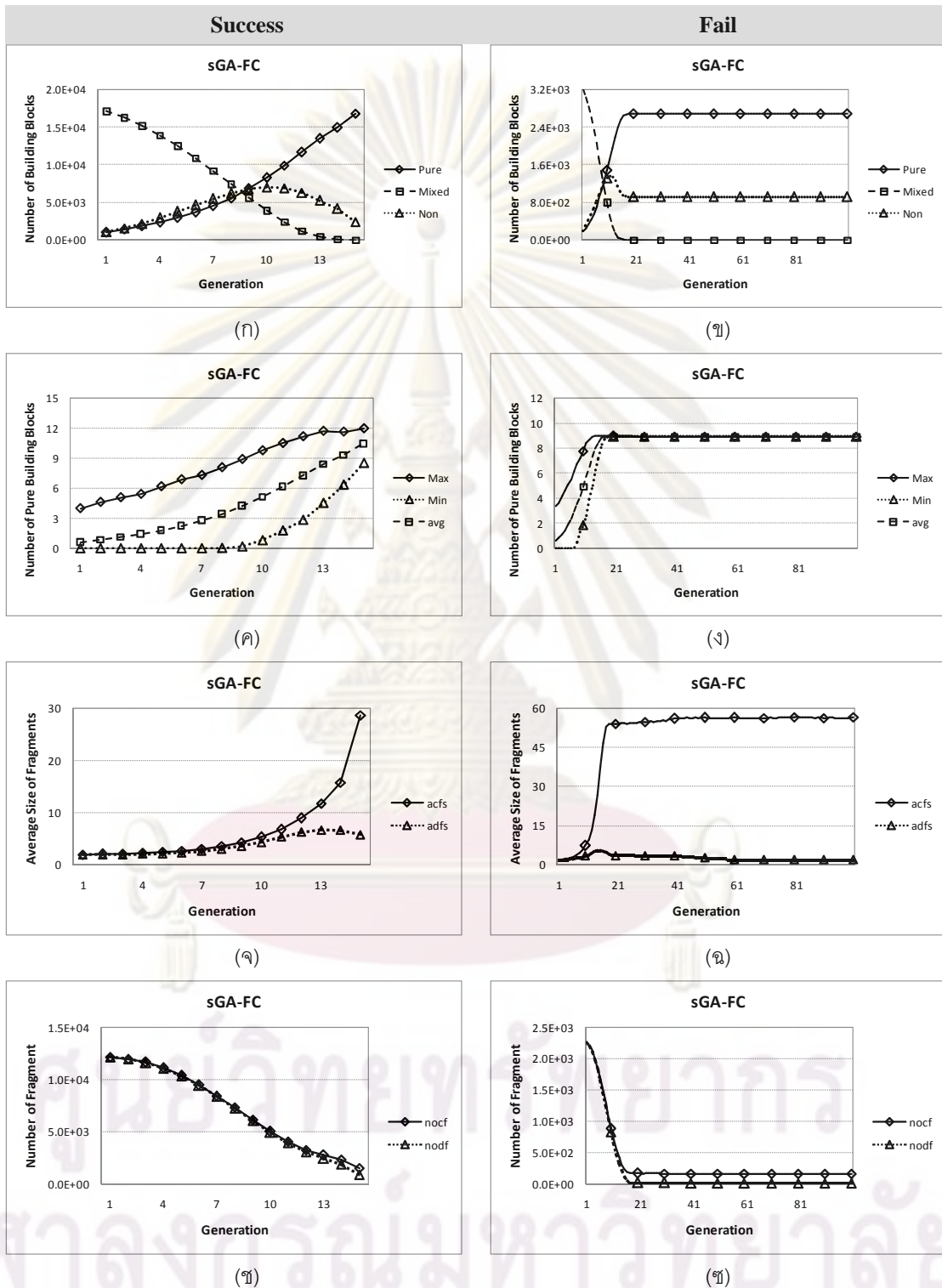
ช้ากว่ามาก (ในรุ่นประชากรที่ 30) เมื่อเปรียบเทียบกับ sGA ในรูปที่ 5.8 (ที่เกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 13 ซึ่งเป็นการยืนยันถึงความสามารถในการสำรวจของวิธีการไขว้เปลี่ยนแบบชิ้นส่วนที่มีสูงกว่การไขว้เปลี่ยนแบบจุดเดียว (อย่างเห็นได้ชัด)



รูปที่ 5.12 รูปแบบจีโนไทป์ของประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 16 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่อินที่มีค่าเป็น 0

สำหรับปัญหา Trap-5 ผลการทดลองในรูปที่ 5.13 (ก-ง) จะมีลักษณะเด่นคล้ายกับผลที่ได้จากวิธี sGA ในรูปที่ 5.6 (ก-ง) จากรูปที่ 5.13 (ก-ข) จะเห็นได้ชัดว่านอกจากจะมีการแข่งขันกันระหว่าง Pure BBs กับ Mixed BBs แล้วยังมี Non BBs เข้าร่วมแข่งขันด้วย กรณีที่หาคำตอบพบจากรูป 5.13 (ก) เมื่อพิจารณาการเพิ่มขึ้นของ Pure BBs จะเห็นว่าก่อนถึงจุดตัดระหว่าง Pure BBs และ Mixed BBs ปริมาณของ Non BBs จะมีมากกว่า Pure BBs เล็กน้อยซึ่งเป็นการแสดงให้เห็นว่าในช่วงเริ่มแรก วิธี sGA-FC มองว่าหน่วยการสร้างลวง (Non BBs) ให้ผลเฉลี่ยที่ดีจึงมีการเก็บรักษาไว้จนกระทั่งเมื่อปริมาณ Mixed BBs ลดลงถึงระดับหนึ่งทำให้ผลเฉลี่ยมีส่วนประกอบของ Pure BBs และ Non BBs มากขึ้น วิธี sGA-FC จึงสามารถตัดสินใจได้ว่า Pure BBs มีคุณค่ามากกว่า Non BBs และทำการละทิ้ง Non BBs เป็นเหตุให้สามารถไปถึงคำตอบได้ แต่จะเห็นในรายละเอียดว่าจำนวน Non BBs ค่อย ๆ ลดลงในอัตราที่ต่ำกว่า Mixed BBs หรือกล่าวได้ว่าวิธี sGA-FC พยายามรักษาความหลากหลายจึงทำให้จำนวน Non BBs ลดลงอย่างช้า ๆ ในกรณีที่หาคำตอบไม่พบจากรูปที่ 5.13 (ข) ผลที่ได้ยังคงแสดงให้เห็นว่าหน่วยการสร้างลวง

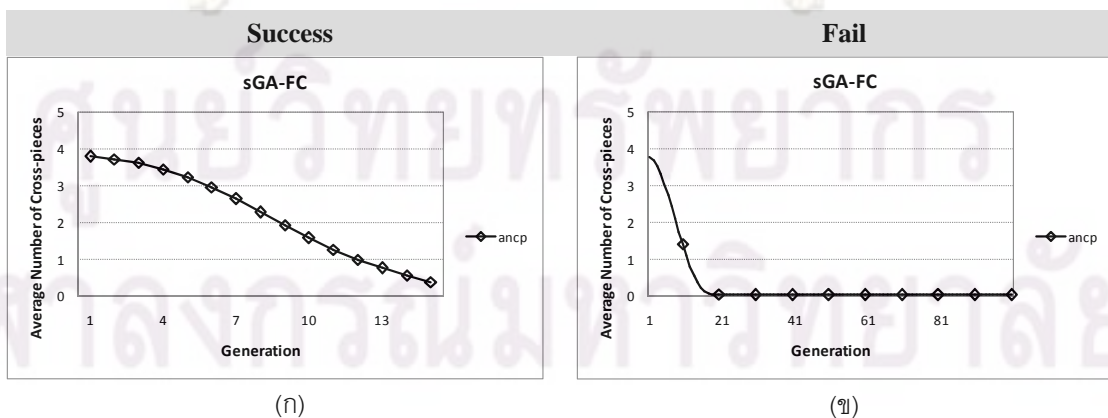
มีคุณค่าใกล้เคียงกับ Pure BBs ในช่วงเริ่มต้นแต่จะสามารถแยกแยะได้ในที่สุด เห็นได้จากการเปลี่ยนแปลงจากรุ่นประชากรที่ 11 ไปยังรุ่นประชากรที่ 21 แต่เนื่องจากจำนวนประชากร



รูปที่ 5.13 ผลการทดสอบของ sGA-FC ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (ฉ) (ช) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ซ) (ฌ)

มีน้อยจึงเกิดการลู่เข้าก่อนที่จะไปถึงคำตอบ และเมื่อพิจารณาถึงจำนวนของหน่วยการสร้างที่เป็นคำตอบ (Pure BBs) ในรูป 5.13 (ค-ง) จะเห็นได้ว่า Pure BBs จะมีจำนวนเพิ่มขึ้นอย่างรวดเร็วเมื่อสามารถตัดสินใจได้ว่า Pure BBs มีคุณค่ากว่า Non BBs ซึ่งสอดคล้องกับรูปที่ 5.13 (ก-ข) อัตราการเพิ่มขึ้นนี้เมื่อเทียบกับปัญหา Royal Road ในรูปที่ 5.9 (ค-ง) จะเห็นว่าแตกต่างกันเนื่องจากปัญหา Trap-5 ในช่วงปลายมีการแข่งขันกันแค่ Pure BBs กับ Non BBs ถ้าไม่เลือก Non BBs จำนวน Pure BBs จะเพิ่มขึ้นอย่างรวดเร็ว

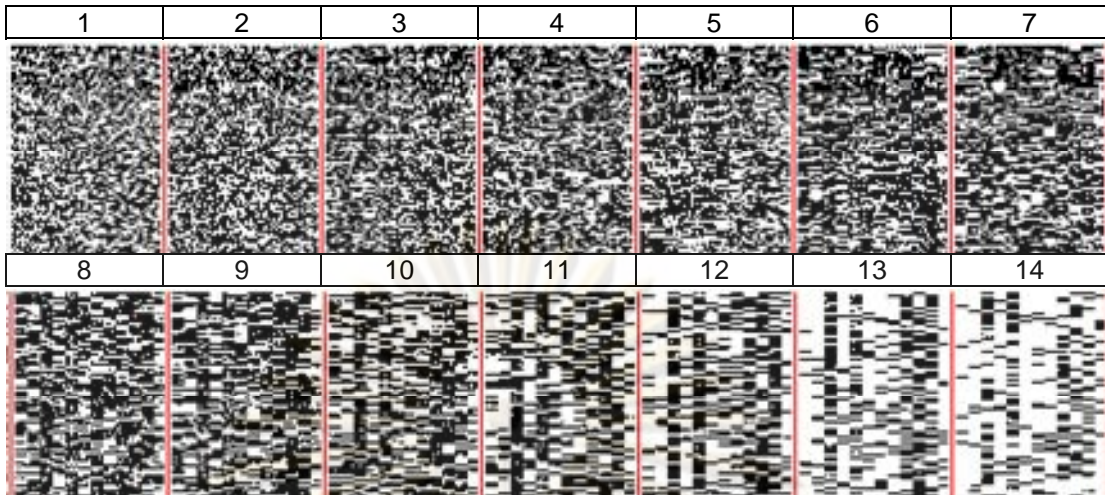
ในส่วนของขนาดชิ้นส่วนโดยเฉลี่ยจากรูป 5.13 (จ-ฉ) จะเห็นว่าขนาดของชิ้นส่วนไม่รวมกัน (adfs) มีขนาดโตขึ้นในช่วงท้าย ที่เป็นเช่นนี้เพราะจากรูป 5.13 (ก) ในช่วงท้ายจะเหลือเพียงแค่ Pure BBs กับ Non BBs ซึ่งมีรูปแบบหน้าตาที่ตรงข้ามกันโดยสิ้นเชิง จึงทำให้เกิดการลู่เข้าของ Non BBs ในระดับหนึ่งเป็นเหตุให้ชิ้นส่วนไม่รวมกันมีขนาดใหญ่ขึ้น และด้วยเหตุนี้ยังส่งผลให้ขนาดของชิ้นส่วนรวมกันในรุ่นหลัง ๆ มีขนาดโตขึ้นอย่างรวดเร็ว และจากรูป 5.13 (ข) จะเห็นว่าปริมาณของชิ้นส่วนลดลงอย่างรวดเร็วซึ่งเกิดจากการที่ขนาดของชิ้นส่วนโตขึ้นอย่างรวดเร็วนั่นเอง ส่วนกรณีที่ทำคำตอบไม่พบ (5.13 ฉ) จะเห็นว่าขนาดของชิ้นส่วนรวมกันและไม่รวมกันมีความแตกต่างกันอย่างสิ้นเชิงและคงที่ตั้งแต่รุ่นประชากรที่ 16 (โดยประมาณ) เมื่อเปรียบเทียบกับปัญหา Royal Road ในรูป 5.9 (ฉ) ที่ขนาดชิ้นส่วนรวมกันค่อย ๆ มีขนาดใหญ่ขึ้นเรื่อย ๆ ความแตกต่างนี้เกิดขึ้นจาก Mixed BBs เนื่องจากปัญหา Royal Road (5.9 ข) มีปริมาณ Mixed BBs อยู่มากจึงทำให้การลู่เข้าเป็นไปได้ช้ากว่า (เมื่อวัดในรูปแบบชิ้นส่วน) แต่สำหรับปัญหา Trap-5 (5.13 ข) จะเห็นว่าไม่มี Mixed BBs เหลืออยู่เลยหลังจากรุ่นประชากรที่ 16 (โดยประมาณ) เป็นเหตุเกิดการลู่เข้าอย่างรุนแรงทำให้ขนาดของชิ้นส่วนรวมกันและไม่รวมกันมีค่าเกือบคงที่ และด้วยเหตุเดียวกันนี้ทำให้ปริมาณของชิ้นส่วนรวมกันและไม่รวมกันในรูปที่ 5.13 (ข) ลดลงอย่างรวดเร็วจนถึงค่าหนึ่งและคงที่ตลอดการดำเนินการ



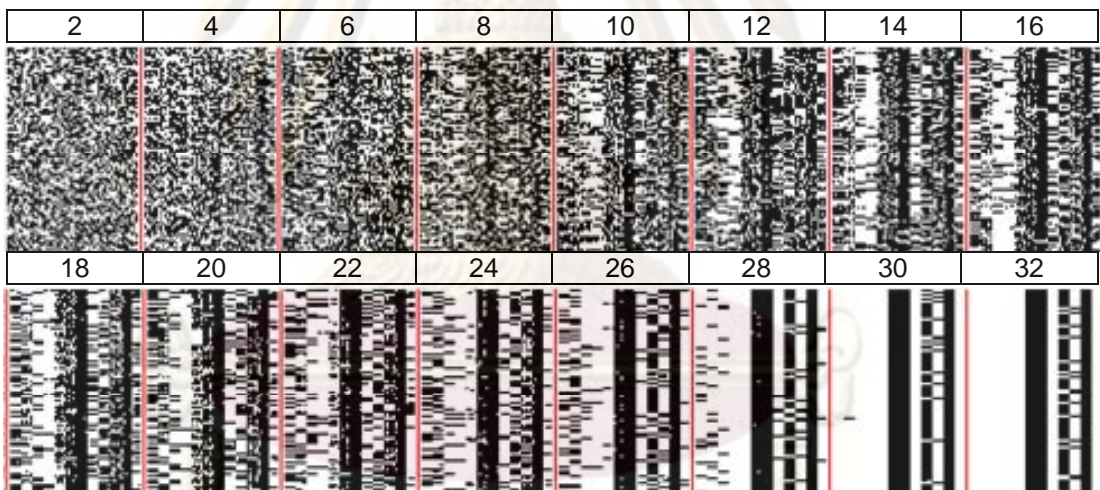
รูปที่ 5.14 จำนวนชิ้นส่วนเฉลี่ยต่อโครโมโซมที่เกิดการไขว้เปลี่ยนกันกับรุ่นประชากรของ sGA-FC ในปัญหา Trap-5 60 บิต

จากรูปที่ 6.14 กราฟของจำนวนชิ้นส่วนโดยเฉลี่ยต่อโครโมโซมที่เกิดการไขว้เปลี่ยน จะมีลักษณะเดียวกันกับปัญหา Royal Road ในรูปที่ 5.10 แต่จะลดลงรวดเร็วกว่าเนื่องจากปัญหา Trap-5 มีรูปแบบหน่วยการสร้างลวงที่ล่อให้มีรูปแบบคู่จิ้นที่มีค่าตรงข้ามกับคำตอบติดกันเป็นจำนวน 4 บิต (มีค่า 0000) ซึ่งทำให้รูปแบบค่าในคู่จิ้นมีลักษณะเป็นแบบแผนมากขึ้น (มีการกระจายตัวน้อยลง) ส่งผลให้มีจำนวนชิ้นส่วนน้อยลงเป็นเหตุให้การแลกเปลี่ยนเกิดขึ้นน้อยลงด้วย ส่วนกรณีที่หาคำตอบไม่พบในรูป 5.14 (ข) จะเห็นว่าตั้งแต่รุ่นประชากรที่ 21 แทบจะไม่มีการแลกเปลี่ยนชิ้นส่วนเกิดขึ้นเลยเป็นเพราะว่าประชากรทั้งหมดเกิดการลู่เข้าเกือบสมบูรณ์ (พิจารณาได้จากรูปที่ 5.13 (ซ))

สำหรับปัญหา Trap-5 จากรูป 5.15 และ 5.16 จะเห็นว่ามีแถบสีดำในแนวนอนที่ยาวอย่างชัดเจน เมื่อเทียบกับปัญหา Royal Road ในรูปที่ 5.11 และ 5.12 ซึ่งเป็นสิ่งที่แสดงให้เห็นถึงรูปแบบหน่วยการสร้างลวง (ล่อให้มีค่าเป็น 0 ติดกันเป็นช่วง ๆ) ได้อย่างชัดเจน วิธีการไขว้เปลี่ยนแบบชิ้นส่วนยังคงแสดงให้เห็นถึงการรักษาความหลากหลายควบคู่ไปกับความพยายามในรักษาหน่วยการสร้างที่ค้นพบ (ทั้งหน่วยการสร้างลวง และหน่วยการสร้างที่เป็นคำตอบ) และจากรูปที่ 5.15 เมื่อเทียบกับรูปที่ 5.13 (ก) จะเห็นสิ่งที่เกิดขึ้นในระดับจีโนมไทป์ได้อย่างชัดเจนว่า Pure BBs Mixed BBs และ Non BBs มีการแข่งขันกันอย่างไร ซึ่งจากรูปที่ 5.13 (ก) จะเห็นว่ามี การตัดสินใจเริ่มละทิ้ง Non BBs ตั้งแต่รุ่นประชากรที่ 10 (โดยประมาณ) ซึ่งสอดคล้องกับรูปที่ 5.15 โดยจะสังเกตเห็นว่าจำนวนแถบสีดำค่อย ๆ ลดจำนวนลงอย่างต่อเนื่องจนกระทั่งพบคำตอบ และจากรูป ถ้าสังเกตให้ดีจะเห็นว่าแถบสีดำที่เหลืออยู่ตั้งแต่รุ่นประชากรที่ 10 เป็นต้นไปจะมีขนาดเป็นท่อนใกล้เคียงกันซึ่งก็คือ Non BBs นั่นเอง ส่วน Mixed BBs (ที่มีขนาดไม่คงที่) จะมีเหลืออยู่น้อยลงเรื่อย ๆ ส่วนกรณีที่หาคำตอบไม่พบจากรูปที่ 5.16 เมื่อเปรียบเทียบกับรูปที่ 5.13 (ข) จะเห็นว่าในรุ่นประชากรที่ 21 จะเริ่มลู่เข้าอย่างสมบูรณ์ เมื่อพิจารณาจากรูปที่ 5.16 แถบสีดำซึ่งก็คือ Non BBs จะเริ่มปรากฏชัดเจนเป็นริ้วในแนวตั้งในรุ่นประชากรที่ 16 และจะชัดเจนขึ้นเรื่อย ๆ จนกระทั่งรุ่นประชากรที่ 30 ซึ่งเกิดการลู่เข้าเกือบสมบูรณ์ แถบสีดำที่ปรากฏชัดนั้นแสดงถึง Non BBs ซึ่งมีจำนวนเกือบคงที่ และแถบสีขาว (Pure BBs) ก็มีจำนวนเกือบคงที่เช่นกัน ส่วนแถบสีดำที่เป็นท่อนเล็ก ๆ ซึ่งก็คือ Mixed BBs นั้นจะแทบไม่เหลืออยู่เลยสอดคล้องกับรูปที่ 5.13 (ข) ทุกประการ



รูปที่ 5.15 รูปแบบจิงโนไทป์ของประชากรของ sGA-FC ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 14 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0



รูปที่ 5.16 รูปแบบจิงโนไทป์ของประชากรของ sGA-FC ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 16 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0

5.3 สรุปผลการทดลอง

ในบทนี้นำเสนอขั้นตอนวิธีการไขว้เปลี่ยนแบบชิ้นส่วน ซึ่งมีการใช้หลักสารสนเทศร่วมต่ำสุดในการระบุชิ้นส่วน ผลการทดสอบแสดงให้เห็นว่าวิธี sGA-FC มีประสิทธิภาพแก่กว่าวิธี sGA ในปัญหา Royal Road และดีกว่าในปัญหา Trap-5 ซึ่งเป็นการแสดงให้เห็นว่าวิธี sGA-FC สามารถจัดการกับหน่วยการสร้างที่มีการลวง (มีความซับซ้อน) ได้ดีกว่า sGA และแสดงถึงความเป็นไปได้ในการประยุกต์ใช้หน่วยการสร้างในรูปแบบที่นำเสนอ (ชิ้นส่วน) เพื่อศึกษาหลักการการทำงานของขั้นตอนวิธีเชิงพันธุกรรมในมุมมองของหน่วยการสร้างและมุมมองของชิ้นส่วน ผลการทดลองยังแสดงให้เห็นว่าการนิยามหน่วยการสร้างในรูปแบบชิ้นส่วนสามารถช่วยให้เข้าใจกระบวนการทำงานในแง่มุมมองต่าง ๆ (การระบุและการประกอบหน่วยการสร้าง) ได้ดีขึ้น



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นตอนวิธีประมาณการแจกแจง

บทนี้แนะนำขั้นตอนวิธีใหม่ในสายงานวิจัยนี้ที่มีชื่อว่าขั้นตอนวิธีประมาณการแจกแจง ซึ่งมีประสิทธิภาพในการหาคำตอบสูง และมีการวิจารณ์เปรียบเทียบความแตกต่างกับขั้นตอนวิธีเชิงพันธุกรรมในแง่มุมต่าง ๆ รวมถึงหน่วยการสร้างด้วย

6.1 ขั้นตอนวิธีประมาณการแจกแจง (EDAs: Estimation of Distribution Algorithm)

ขั้นตอนวิธีประมาณการแจกแจง (Larrañaga and Lozano, 2001) หรือขั้นตอนวิธีเชิงพันธุกรรมแบบการสร้างแบบจำลองความน่าจะเป็น (Probabilistic Model-building Genetic Algorithms) ไม่มีขั้นตอนการดำเนินการไขว้เปลี่ยนและการกลายพันธุ์ แต่อาศัยการเรียนรู้และการซักรวอย่าง (Sampling) การแจกแจงของความน่าจะเป็นของกลุ่มผลเฉลยที่มีค่าความเหมาะสมที่สุดจากผลเฉลยทั้งหมดในแต่ละรอบการทำงาน วิธีการแบบนี้สามารถตรวจจับพฤติกรรมความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องกับปัญหาได้อย่างมีประสิทธิภาพและชัดเจน

ในช่วงไม่กี่ปีนี้ได้มีการศึกษาขั้นตอนวิธีประมาณการแจกแจงอย่างจริงจังและแพร่หลาย ทำให้งานด้านนี้มีการพัฒนาอย่างรวดเร็ว วิธีนี้บางที่อาจเรียกอีกชื่อหนึ่งว่าการสร้างแบบจำลองความน่าจะเป็นเชิงสุ่มของขั้นตอนวิธีเชิงพันธุกรรม (Probabilistic Model-building GAs) หลักการทำงานได้มีการยึดหลักการพัฒนาทางด้านสถิติ ปัญญาประดิษฐ์ และการจับกลุ่ม (Clustering) มาใช้ ทำให้ได้ขั้นตอนวิธีที่มีความสามารถในการหาผลเฉลยที่ดีขึ้น มีความเร็วเพิ่มขึ้น ผลเฉลยที่ได้มีคุณภาพดีขึ้นและความน่าเชื่อถือ โดยไม่ขัดแย้งกับวิธีทางด้านการคำนวณเชิงวิวัฒนาการในอดีต ขั้นตอนวิธีประมาณการแจกแจงยังช่วยให้เข้าใจบทบาทของตัวดำเนินการทางพันธุกรรมที่หลากหลายในการสร้างแบบจำลอง การกระจายของคำตอบที่ดี ๆ ในประชากรซึ่งถือได้ว่าได้เปิดมุมมองใหม่ที่ดีกว่าในการเข้าใจกลไกการทำงานของพันธุกรรม

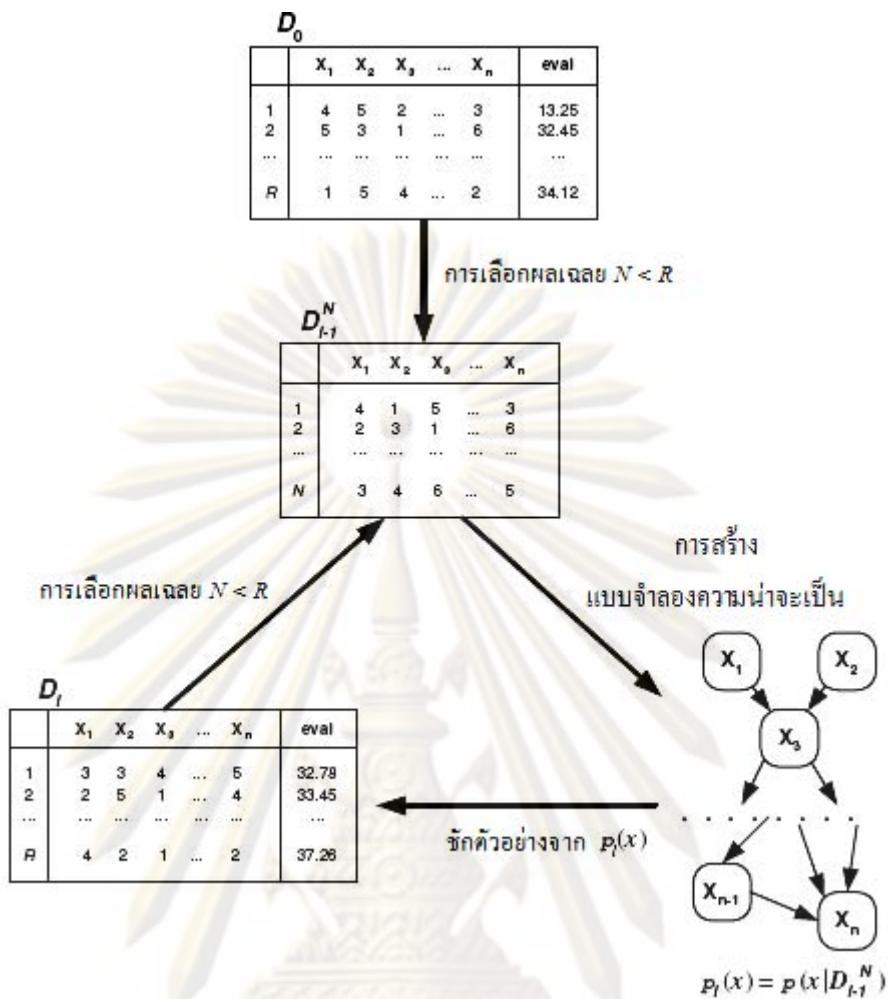
ในขั้นตอนวิธีประมาณการแจกแจงแสดงความสัมพันธ์กันของตัวแปรออกมาในรูปแบบการกระจายความน่าจะเป็นร่วม (Joint Probability Distribution) ของผลเฉลยที่ถูกเลือกในแต่ละรอบการทำงาน และต้องใช้วิธีการปรับตัวเพื่อที่จะเรียนรู้แบบจำลองจากข้อมูล โดยวิธีการเหล่านี้ถูกพัฒนามาจากโดเมนของแบบจำลองความน่าจะเป็นเชิงกราฟ (Graphical Probabilistic Model) ช่วงเริ่มแรกขั้นตอนวิธีประมาณการแจกแจงเป็นแบบโดเมนไม่ต่อเนื่อง (Discrete Domain) และได้มีการพัฒนาจนเป็นโดเมนต่อเนื่อง (Continuous Domain) ลองพิจารณาตัวอย่างที่ทำให้เห็นภาพโดยพิจารณาปัญหาที่เกิดขึ้นกับการจัดจำแนกหรือรู้จักกันในชื่อปัญหาการเลือก

ลักษณะเซตย่อย (Feature Subset Selection: FSS) (Inza, Larrañaga, Etxeberria and Sierra, 2000) โดยให้ชุดข้อมูลมากับเซตของตัวแปรทำนาย (Predictive Variables) x_1, x_2, \dots, x_n และก็ชนิดคลาสของตัวแปร ปัญหาจะเป็นการเลือกเซตย่อยของตัวแปรที่ทำให้การจัดจำแนกมีค่ามากที่สุด ในเซตทดสอบซึ่งเป็นปัญหาที่มีจำนวนปริภูมิค้นหาเป็น 2^n

รูปที่ 6.1 แสดงแผนทั่วไปของแนวคิดขั้นตอนวิธีประมาณการแจกแจงโดยมีขั้นตอนการทำงานดังนี้

1. ขั้นตอนแรกเริ่มจากสร้างประชากรรุ่นแรก D_0 ของผลเฉลย R ด้วยการสุ่มด้วยค่าความน่าจะเป็น 0.5 เพื่อให้การกระจายตัวใน R เป็นการแจกแจงเอกรูป และหลังจากนั้นจะหาค่าความเหมาะสมของผลเฉลย R แต่ละตัว
2. ขั้นตอนที่สอง ในการวิวัฒนาการจากประชากรรุ่น $l-1^{th}$ หรือ D_{l-1} ไปยังรุ่นถัดไป D_l ทำโดยการเลือกผลเฉลยจากประชากรรุ่น D_{l-1} เป็นจำนวน N ($N < R$) ซึ่งจะแทนด้วย D_{l-1}^N เป็นเซตของผลเฉลยจำนวน N ตัวจากรุ่นที่ $l-1$
3. ขั้นตอนที่สาม เป็นขั้นตอนการบวนการเรียนรู้ความสัมพันธ์ที่ขึ้นแก่กันระหว่างตัวแปรในแบบจำลองความน่าจะเป็นแบบ n มิติ เพื่อทำให้วิวัฒนาการไปในทางที่ผลเฉลยมีค่าความเหมาะสมสูงขึ้น
4. ขั้นตอนที่สุดท้าย ประชากรรุ่นใหม่ D_l ที่ประกอบด้วยผลเฉลย R จะได้จากการจำลองการเรียนรู้ความน่าจะเป็นของการกระจายตัวในขั้นตอนก่อนหน้า โดยทั่วไปยอมให้ใช้หลักการอภิชิน (Elitist Approach) (เก็บผลเฉลยที่ดีที่สุด จากรุ่น D_{l-1}^N ก่อนเอาไว้ส่วนหนึ่ง E แล้วเอามาใส่ในรุ่นปัจจุบัน D_l โดยไม่มีการเปลี่ยนแปลงใด ๆ) ได้ซึ่งจะทำให้ผลเฉลยใหม่จำนวน $R - E$ แทนที่จะเป็นจำนวน R

ขั้นตอนที่ 2, 3 และ 4 จะถูกทำซ้ำจนกระทั่งเข้าเงื่อนไขเกณฑ์การหยุด (Stopping Criteria) ตัวอย่างเช่น ค่าความเหมาะสมประชากรส่วนใหญ่เข้าสู่ค่า ๆ หนึ่ง หรือดำเนินงานครบจำนวนรุ่นที่กำหนดไว้ เป็นต้น



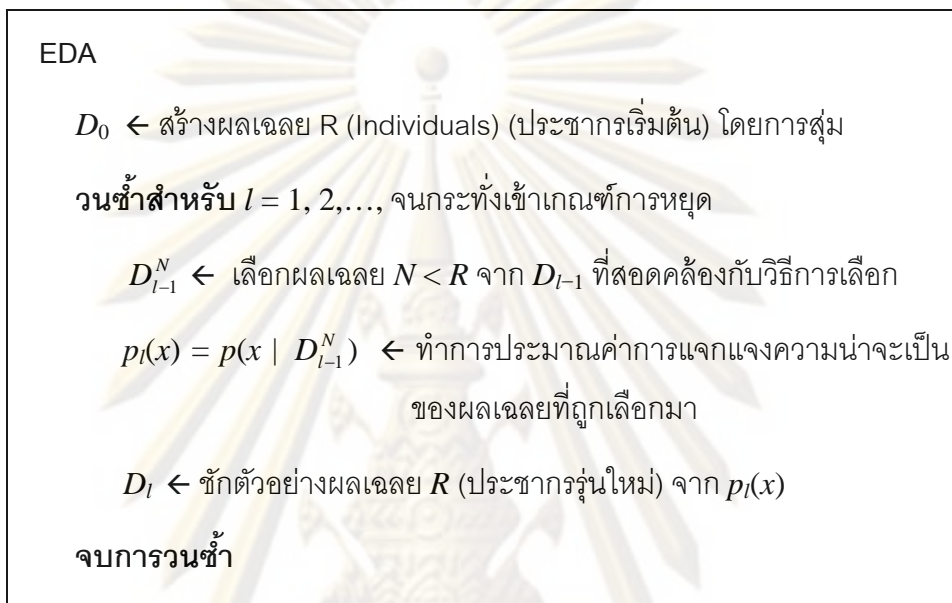
รูปที่ 6.1 แนวคิดของ EDA ในกระบวนการหาค่าเหมาะที่สุด

6.2 ขั้นตอนวิธีประมาณการแจกแจงในโดเมนไม่ต่อเนื่อง

ในส่วนนี้จะกล่าวถึงสัญกรณ์ (Notations) ที่ใช้ในการอธิบาย EDAs ในโดเมนไม่ต่อเนื่อง ซึ่งเป็นโดเมนที่นิยมศึกษากันเนื่องจากง่ายกว่าโดเมนต่อเนื่องในการนำไปปฏิบัติ

ให้ X_i ($i = 1, \dots, n$) เป็นตัวแปรสุ่ม การสร้างกรณีตัวอย่างที่เป็นไปได้ของ X_i จะแทนด้วย x_i . $p(X_i = x_i)$ หรือ $p(x_i)$ หมายถึงความน่าจะเป็นที่ตัวแปรสุ่ม X_i จะมีค่า x_i ในทำนองเดียวกัน $X = (X_1, \dots, X_n)$ จะแทนตัวแปรสุ่ม n มิติ และ $x = (x_1, \dots, x_n)$ เป็นค่าที่จะเป็นไปได้ ความน่าจะเป็นจำนวนมากของตัวแปร X จะแทนด้วย $p(X = x)$ หรือ $p(x)$ ความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability) ของตัวแปร X_i เมื่อรู้ค่า x_j ของตัวแปร X_j จะถูกเขียนในรูป $p(X_i = x_i | X_j = x_j)$ หรือ $p(x_i | x_j)$ สัญกรณ์ D แทนเซตข้อมูล เช่น เซตของตัวอย่างผลเฉลย R ของตัวแปร (X_1, \dots, X_n)

รูปที่ 6.2 แสดงรหัสเทียบของ EDAs ในปัญหาการหาค่าเหมาะสมที่สุดเชิงการจัดโดย ใช้สัญกรณ์ที่ได้กล่าวไปแล้วโดย $x = (x_1, \dots, x_n)$ แทนผลเฉลยแต่ละตัวของเงินจำนวน n ตัว และ D_l แทนประชากรทั้งหมดของผลเฉลย R ในรุ่นที่ l^{th} และให้ D_l^N แทนประชากรของผลเฉลยที่ถูกเลือกจำนวน N ตัวจาก D_{l-1} หน้าที่หลักของ EDAs จะเป็นการประมาณ $p(x | D_{l-1}^N)$ ซึ่งหมายถึงความน่าจะเป็นแบบมีเงื่อนไขร่วม (Joint Conditional Probability) ของผลเฉลย x



รูปที่ 6.2 รหัสเทียบของแนวคิด EDA ในโดเมนไม่ต่อเนื่อง

จะเห็นว่ามี การประมาณค่าความน่าจะเป็นร่วมของผลเฉลยที่ถูกเลือกจากหลาย ๆ รุ่น กำหนดให้ $p_l(x) = p_l(x | D_{l-1}^N)$ แทนความน่าจะเป็นแบบมีเงื่อนไขร่วมของรุ่นที่ l^{th}

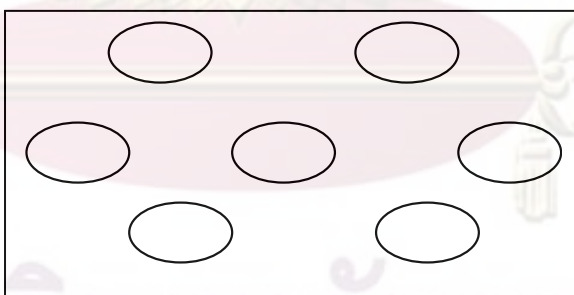
ขั้นตอนที่สำคัญที่สุดคือการหาการพึ่งพิงระหว่างกัน (Interdependencies) ระหว่างตัวแปรที่แทนจุด ๆ หนึ่งในปริภูมิค้นหา โดยหลักการพื้นฐานจะประกอบด้วย การโน้มน้ำหนักแบบจำลองความน่าจะเป็นจากกลุ่มผลเฉลยที่ดีที่สุด ในประชากรทั้งหมด สำหรับแบบจำลองความน่าจะเป็นนั้นจะเป็นแบบจำลองโดยประมาณซึ่งถูกใช้ในการสุ่มชักตัวอย่างเพื่อสร้างประชากรรุ่นใหม่ และจากประชากรรุ่นใหม่จะถูกใช้ในการสร้างแบบจำลองใหม่สำหรับสร้างประชากรรุ่นต่อไป กระบวนการนี้จะทำซ้ำจนกระทั่งเข้าเกณฑ์การหยุด ขั้นตอนที่ยากที่สุดสำหรับ EDAs คือการประมาณการแจกแจงความน่าจะเป็น $p_l(x)$ ซึ่งเป็นการคำนวณพารามิเตอร์ทั้งหมดที่จำเป็นต้องใช้ในการระบุแบบจำลองความน่าจะเป็นโดยในทางปฏิบัติไม่สามารถทำได้จริง ทำให้เกิดแนวคิดการประมาณที่หลากหลายในการแยกตัวประกอบการแจกแจงความน่าจะเป็นให้สอดคล้องกับแบบจำลองความน่าจะเป็น ในส่วนถัดไปจะเป็นการแนะนำให้รู้จักแนวคิด EDA

รูปแบบต่าง ๆ ซึ่งสามารถจัดจำแนกได้ตามจำนวนการพึ่งพิงระหว่างตัวแปร และสามารถอ่านวิจารณ์ฉบับสมบูรณ์ในเรื่องนี้ได้จาก (Larrañaga and Lozano, 2001)

6.2.1 แบบปราศจากการพึ่งพิงระหว่างกัน (Without Interdependencies)

วิธีทั้งหมดในหมวดนี้ถือว่าการแจกแจงความน่าจะเป็นร่วม n มิติ เป็นผลคูณของตัวแปรเดี่ยว n ตัว และมีการแจกแจงเป็นอิสระต่อกันแทนด้วย $p_I(x) = \prod_{i=1}^n p_i(x_i)$ การสมมุติแบบนี้ดูเหมือนไม่น่าเชื่อถือเพราะว่าธรรมชาติของปัญหาค่าความเหมาะสมที่สุดแบบยาก ตัวแปรของปัญหามักจะมีการพึ่งพิงกันระหว่างกันในระดับหนึ่ง แต่การประมาณแบบนี้ก็สามารถยอมรับได้ใน EDAs สำหรับบางปัญหาที่สามารถสมมุติว่าตัวแปรแต่ละตัวเป็นอิสระจากกัน

มีหลายงานวิจัยที่เกี่ยวข้องกับแนวคิดนี้ ตัวอย่างเช่น การจำลองการไขว้เปลี่ยนแบบพื้นฐานบิต (Bit-Based Simulated Crossover: BSC) (Syswerda, 1993) การเรียนรู้เพิ่มขึ้นแบบพื้นฐานประชากร (Population- Based Incremental Learning: PBIL) (Baluja, 1994) ขั้นตอนวิธีเชิงพันธุกรรมแบบกะชับ (compact Genetic Algorithm: cGA) (Harik, Lobo and Goldberg, 1998) และขั้นตอนวิธีการแจกแจงตามขอบหนึ่งตัวแปร (Univariate Marginal Distribution Algorithm: UMDA) (Mühlenbein, 1998) ซึ่งแนวคิดเหล่านี้แตกต่างกันที่การคำนวณ $p_i(x_i)$ ตัวอย่างเช่นใน UMDA จะใช้การประมาณความถี่ตามขอบสัมพัทธ์ (Relative Marginal Frequencies) ของตัวแปร i^{th} ในเซตย่อยของผลเฉลยที่ถูกเลือก D_{i-1}^N เป็นต้น



รูปที่ 6.3 การนำเสนอในรูปแบบกราฟของวิธีประมาณการแจกแจงในการหาค่าเหมาะสมที่สุดเชิงการจัด (Combinatorial Optimization) แบบปราศจากการพึ่งพิงระหว่างกัน ได้แก่ UMDA, PBIL และ cGA

วิธี UMDA (Univariate Marginal Distribution Algorithm) เมื่อพิจารณาจากรูปที่ 6.1 แบบจำลองที่ใช้ประมาณการกระจายการแจกแจงร่วมของผลเฉลยที่ถูกเลือกในแต่ละรุ่นจะใช้วิธีที่ง่ายที่สุดคือ การกระจายตัวประกอบในรูปการคูณกันของการแจกแจงตามขอบตามขอบหนึ่งตัวแปรที่เป็นอิสระต่อกัน ดังสมการ

$$p_l(x) = p(x | D_{l-1}^N) = \prod_{i=1}^n p_l(x_i) \quad (6.1)$$

และแต่ละการแจกแจงตามขอบหนึ่งตัวแปรจะประมาณได้จากความถี่ตามขอบ ดังสมการ

$$p_l(x_i) = \prod_{i=1}^n \left[\frac{\sum_{j=1}^N \delta_j(X_i = x_i | D_{l-1}^N)}{N} \right] \quad (6.2)$$

โดยที่

$$\delta_j(X_i = x_i | D_{l-1}^N) = \begin{cases} 1 & \text{ถ้าใน } j^{\text{th}} \text{ กรณีที่ } D_{l-1}^N, X_i = x_i \\ 0 & \text{กรณีอื่น} \end{cases} \quad (6.3)$$

UMDA

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $l = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{l-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{l-1} ที่สอดคล้องกับวิธีการเลือก

$$p_l(x) = p(x | D_{l-1}^N) = \prod_{i=1}^n p_l(x_i) = \prod_{i=1}^n \frac{\sum_{j=1}^N \delta_j(X_i = x_i | D_{l-1}^N)}{N}$$

\leftarrow ทำการประมาณค่าการแจกแจงความน่าจะเป็นร่วมของผลเฉลยที่ถูกเลือกมา

$D_l \leftarrow$ ซักตัวอย่างผลเฉลย R (ประชากรรุ่นใหม่) จาก $p_l(x)$

จบการวนซ้ำ

รูปที่ 6.4 รหัสเทียมของ UMDA

วิธี PBIL (Population- Based Incremental Learning) แทนประชากรของผลเฉลยในแต่ละรุ่นด้วยเวกเตอร์ของความน่าจะเป็น

$$p_l(x) = (p_l(x_1), \dots, p_l(x_i), \dots, p_l(x_n)) \quad (6.4)$$

โดยที่ $p_l(x_i)$ แทนค่าความน่าจะเป็นที่จะมีค่า 1 ในตัวประกอบที่ i^{th} ของ D_l ซึ่งเป็นประชากรของผลเฉลยในรุ่นที่ l^{th}

การทำงานของขั้นตอนวิธีนี้จะเริ่มจากแต่ละรุ่นจะมีเวกเตอร์ความน่าจะเป็น $p_l(x)$ จำนวนผลเฉลย R โดยที่แต่ละตัวจะถูกประเมินและจะเลือกตัวที่ดีที่สุด N ตัว ($N \leq R$) จะเขียนแทนด้วย $x'_{1:R}, \dots, x'_{k:R}, \dots, x'_{N:R}$

ซึ่งผลเฉลยที่ถูกเลือกเหล่านี้จะนำมาใช้ในการปรับค่าเวกเตอร์ความน่าจะเป็นโดยใช้กฎแนวคิดเฮบเบียน (Hebbian Inspired Rule) ดังสมการ

$$p_{l+1}(x_i) = (1 - \alpha)p_l(x_i) + \alpha \frac{1}{N} \sum_{k=1}^N x'_{i,k:R} \quad (6.5)$$

โดยที่ $\alpha \in (0,1]$ เป็นค่าพารามิเตอร์ จะถือว่า PBIL เป็นวิธีการแบบ EDA ถ้าค่า $\alpha = 1$ ซึ่งในกรณีนี้จะเทียบได้กับ UMDA

PBIL

$p_0(x) \leftarrow$ กำหนดค่าให้กับเวกเตอร์ความน่าจะเป็น

ในขณะที่ยังไม่ลู่เข้า ดำเนินการ

เริ่ม

- ใช้ $p_l(x)$ ในการสร้างผลเฉลย $x'_1, \dots, x'_k, \dots, x'_R$
- ประเมินค่าและจัดลำดับ $x'_1, \dots, x'_k, \dots, x'_R$
- เลือกผลเฉลยที่ดีที่สุด N ตัว ($N \leq R$) $x'_{1:R}, \dots, x'_{k:R}, \dots, x'_{N:R}$
- ปรับค่าของเวกเตอร์ความน่าจะเป็น $p_{l+1}(x) = (p_{l+1}(x_1), \dots, p_{l+1}(x_n))$
- สำหรับ $i = 1, \dots, n$ ทำ

$$p_{l+1}(x_i) = (1 - \alpha)p_l(x_i) + \alpha \frac{1}{N} \sum_{k=1}^N x'_{i,k:R}$$

จบ

รูปที่ 6.5 รหัสเทียมของ PBIL

cGA (compact Genetic Algorithm) วิธีนี้จะใช้เวกเตอร์ความน่าจะเป็นโดยเริ่มที่แต่ละตัวประกอบจะทำการแจกแจงแบบเบอนูลลี (Bernoulli Distribution) ด้วยค่าพารามิเตอร์เท่ากับ 0.5 และถัดจากนั้นสร้างผลเฉลยสองตัวแบบสุ่มจากเวกเตอร์ความน่าจะเป็น หลังจากนั้นผลเฉลยทั้งสองตัวจะถูกประเมินค่าแล้วก็จะปรับค่าของแต่ละเวกเตอร์ความน่าจะเป็นตามผลเฉลยที่ชนะ

โดยการปรับค่ากระทำทีละบิต (ปรับเวกเตอร์ความน่าจะเป็นประจำบิตทีละตัว) และจะปรับเฉพาะบิตที่แตกต่างกันระหว่างผลเฉลยทั้งสองตัวโดยการปรับจะเป็นดังขั้นตอนที่ 4 ในรูปที่ 6.6 ซึ่งการทำแบบนี้จะเป็นการปรับเวกเตอร์ความน่าจะเป็นไปในทิศทางเดียวกับผลเฉลยที่ชนะ และจะทำไปเรื่อย ๆ จนกว่าค่าเวกเตอร์ความน่าจะเป็นจะลู่เข้าดังเงื่อนไขในขั้นตอนที่ 5 ในรูป 6.6

cGA

ขั้นตอนที่ 1. $p_0(x) \leftarrow$ กำหนดค่าให้กับเวกเตอร์ความน่าจะเป็น

$$p_0(x) = p_0(x_1, \dots, x_i, \dots, x_n) = (p_0(x_1), \dots, p_0(x_i), \dots, p_0(x_n)) = (0.5, \dots, 0.5, \dots, 0.5)$$

ขั้นตอนที่ 2. $l = l + 1$ ชักตัวอย่าง $p_l(x)$ ด้วยค่า $l = 0, 1, 2, \dots$ จนได้ผลเฉลยครบสองตัว x'_1, x'_2

ขั้นตอนที่ 3. ประเมินค่าและจัดลำดับ x'_1, x'_2 โดย $x'_{1,2}$ (ค่าที่ดีที่สุดของทั้งคู่) และ $x'_{2,2}$ (ค่าที่แย่ที่สุดของทั้งคู่)

ขั้นตอนที่ 4. ปรับค่าของเวกเตอร์ความน่าจะเป็น $p_l(x)$ ด้วยค่า $x'_{1,2}$ สำหรับ $i = 1, \dots, n$ ทำ

ถ้า $x'_{i,1,2} \neq x'_{i,2,2}$ แล้ว

$$\text{ถ้า } x'_{i,1,2} = 1 \text{ แล้ว } p_l(x_i) = p_{l-1}(x_i) + \frac{1}{K}$$

$$\text{ถ้า } x'_{i,1,2} = 0 \text{ แล้ว } p_l(x_i) = p_{l-1}(x_i) - \frac{1}{K}$$

ขั้นตอนที่ 5. ตรวจสอบว่าเวกเตอร์ความน่าจะเป็น $p_l(x)$ ลู่เข้าหรือไม่

สำหรับ $i = 1$ ถึง n ทำ

ถ้า $p_l(x_i) > 0$ และ $p_l(x_i) < 1$ แล้ว

ไปทำขั้นตอนที่ 2.

ขั้นตอนที่ 6. $p_l(x)$ แทนค่าผลเฉลยสุดท้าย

เมื่อ l คือ ความยาวโครโมโซม

K คือ ค่าคงที่ (ใช้ในการปรับน้ำหนัก)

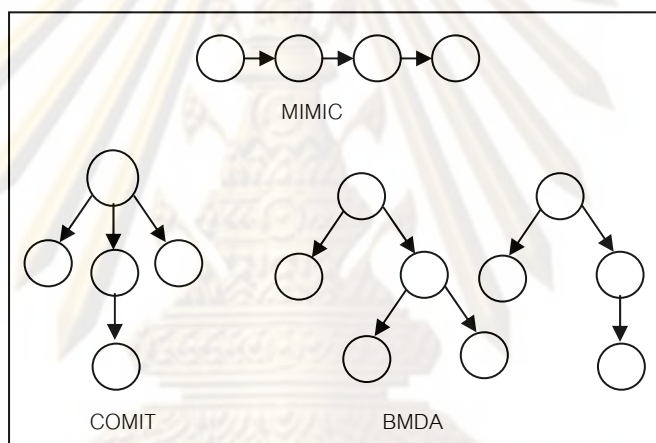
รูปที่ 6.6 รหัสเทียมของ cGA (l คือความยาวโครโมโซม)

6.2.2 แบบการพึ่งพิงเชื่อมโยงแบบคู่ (Pairwise Dependencies)

ในการพยายามที่จะอธิบายความสัมพันธ์ที่ขึ้นแก่กันของตัวแปรให้ง่ายที่สุดในหมวดนี้ จะเป็นการอ้างถึงแนวคิดการแจกแจงความน่าจะเป็นร่วมที่สามารถประมาณได้ง่ายและรวดเร็ว

โดยการใช้การฟุ้งฟุ้งเชื่อมโยงแบบคู่ หรือการขึ้นแก่กันระหว่างสองตัวแปร ในรูปที่ 6.7 แสดงตัวอย่างแบบจำลองเชิงกราฟของการฟุ้งฟุ้งเชื่อมโยงแบบคู่

ขั้นตอนวิธีในหมวดนี้ต้องการขั้นตอนการสร้างโครงสร้างที่แทนแบบจำลองความน่าจะเป็นที่ดีที่สุดซึ่งไม่มีในหมวดก่อนหน้านี้ ตัวอย่างของขั้นตอนวิธีนี้ได้แก่ ขั้นตอนวิธีเชิงละโมบ หรือเรียกว่า การทำให้ได้สารสนเทศร่วมมากที่สุดสำหรับการจัดกลุ่มข้อมูลนำเข้า (Mutual Information Maximization for Input Clustering: MIMIC) (De Bonet, Isbell and Viola, 1997), การเชื่อมตัวทำให้เหมาะที่สุดกับต้นไม้สารสนเทศร่วม (Combining Optimizers with Mutual Information Trees:COMIT) (Baluja and Davies, 1997a, 1997b) และขั้นตอนวิธีการแจกแจงตามขอบสองตัวแปร (Bivariate Marginal Distribution Algorithm: BMDA) (Pelikan and Mühlenbein, 1999)



รูปที่ 6.7 การนำเสนอในรูปแบบกราฟของวิธีประมาณการแจกแจงในการหาค่าเหมาะที่สุดสมเชิงการจัดกับการฟุ้งฟุ้งเชื่อมโยงแบบคู่ ได้แก่ MIMIC, COMIT และ BMDA

MIMIC (Mutual Information Maximization for Input Clustering) ใช้การค้นหาการเรียงสับเปลี่ยนที่ดีที่สุดระหว่างตัวแปรตามลำดับเพื่อหาการแจกแจงความน่าจะเป็น $p_l^\pi(x)$ ที่ใกล้เคียงกับการแจกแจงเชิงประจักษ์ (Empirical Distribution) มากที่สุดของเซตของจุดที่เลือก โดยใช้ตัววัดเป็นระยะทางคูลแบคเลเบลอร์ (Kullback-Leibler Distance) โดยที่

$$p_l^\pi(x) = p_l(x_{i_1} | x_{i_2}) \cdot (p_l(x_{i_2} | x_{i_3}) \cdots (p_l(x_{i_{n-1}} | x_{i_n}) \cdot p_l(x_{i_n})) \quad (6.6)$$

และ $\pi = (i_1, i_2, \dots, i_n)$ แทนการเรียงสับเปลี่ยนของดัชนี $1, 2, \dots, n$

สามารถพิสูจน์ได้ว่าการลู่ออกของแคว้นเบคเคิลเบอริ (Kullback-Leibler Divergence) ระหว่างการแจกแจงความน่าจะเป็น $p_l(x)$ และ $p_l^\pi(x)$ ที่อธิบายในสมการก่อนหน้าเป็นฟังก์ชันของ

$$H_l^\pi(x) = h_l(X_{i_n}) + \sum_{j=1}^{n-1} h_l(X_{i_j} | X_{i_{j+1}}) \tag{6.7}$$

โดยที่ $h(x) = -\sum_x p(X=x) \log p(X=x)$ แทนเอนโทรปี (Shannon Entropy) ของตัวแปร X และ $h(X|Y) = -\sum_y h(X|Y=y) p(Y=y)$ โดยที่ $h(X|Y=y) = -\sum_x h(X=x|Y=y) \log p(X=x|Y=y)$ แทนค่าเฉลี่ยของความไม่แน่นอนในตัวแปร X โดยรู้ Y

ปัญหาของการค้นหา $p_l^\pi(x)$ ที่ดีที่สุดเป็นปัญหาเดียวกันกับการค้นหาการเรียงสับเปลี่ยน π^* ที่ทำให้ $H_l^\pi(x)$ มีค่าน้อยที่สุด การที่จะหา π^* ในงานวิจัย (De Bonet, Isbell and Viola, 1997) ได้เสนอวิธีขั้นตอนวิธีเชิงละโมบแบบตรงไปตรงมาที่หลีกเลี่ยงการค้นหาการเรียงสับเปลี่ยนที่เป็นไปได้ทั้งหมด $n!$ แนวคิดเป็นการเลือก X_{i_n} เป็นตัวแปรที่มีค่าเอนโทรปีโดยประมาณน้อยที่สุดในทุกขั้นตอนต่อไปนี้ เริ่มจากหีบตัวแปรมาจากเซตของตัวแปรที่มีค่าเอนโทรปีแบบมีเงื่อนไข (Conditional Entropy) ที่เกี่ยวข้องกับตัวแปรที่ถูกเลือกในขั้นตอนก่อนหน้านี้เฉลี่ยมีค่าน้อยที่สุด รูปที่ 6.8 แสดงรหัสเทียมของการประมาณการแจกแจงความน่าจะเป็นร่วมที่ใช้ในวิธี MIMIC

MIMIC

ขั้นตอนที่ 1. เลือก $i_n = \operatorname{argmin}_j \hat{h}_l(X_j)$

ขั้นตอนที่ 2. สำหรับ $k = n - 1, \dots, 1$ ทำ

เลือก $i_k = \operatorname{argmin}_j \hat{h}_l(X_j | X_{i_{k+1}})$

โดย $j \neq i_{k+1}, \dots, i_n$

ขั้นตอนที่ 3.

$$p_l^\pi(x) = p_l(x_{i_1} | x_{i_2}) \cdot (p_l(x_{i_2} | x_{i_3}) \cdots (p_l(x_{i_{n-1}} | x_{i_n}) \cdot p_l(x_{i_n}))$$

รูปที่ 6.8 แนวคิด MIMIC ในการประมาณการแจกแจงความน่าจะเป็นร่วม ณ รุ่น l ใด ๆ กำหนดให้ $\hat{h}_l(X)$ และ $\hat{h}_l(X|Y)$ แทนเอนโทรปีเชิงประจักษ์ (Empirical Entropy) ของ X และเอนโทรปีเชิงประจักษ์ของ X เมื่อรู้ค่า Y ตามลำดับซึ่งค่าทั้งคู่ได้จากการประมาณ D_l^N

วิธี COMIT (Combining Optimizers with Mutual Information Trees) เป็นการนำเสนอขั้นตอนวิธีที่ใช้ในการเรียนรู้การแจกแจงความน่าจะเป็นลำดับสอง (Second-order) จากผลเฉลยที่ดีที่สุดที่เคยผ่านมาแล้ว และใช้หลักสถิติในการสร้างต้นไม้พึ่งพิง (Dependency Trees) ที่ดีที่สุดในเงื่อนไขค่าควรจะเป็นสูงสุด (Maximum Likelihood) วิธีนี้เป็นการผสมกันระหว่างแนวคิด EDA กับตัวหาค่าที่ดีที่สุดแบบท้องถิ่น (Local Optimizers) โดยที่การประมาณการแจกแจงความน่าจะเป็นของผลเฉลยที่ถูกเลือกในแต่ละรุ่นจะใช้โครงสร้างต้นไม้แบบโครงข่ายเบย์ (Bayesian Network) ในการเรียนรู้ และผลเฉลยจะสร้างขึ้นจากแบบจำลองความน่าจะเป็นที่ผ่านการเรียนรู้แล้ว โดยผลเฉลยที่ดีที่สุดจะถูกใช้เป็นจุดเริ่มต้นสำหรับกระบวนการค้นหาอย่างรวดเร็ว (Fast-search Procedure) และผลเฉลยที่ดีที่สุดบางตัวระหว่างกระบวนการค้นหาอย่างรวดเร็วจะถูกเพิ่มเข้าไปในประชากรที่ถูกเลือกในรอบก่อนเพื่อใช้ในการสร้างประชากรรุ่นใหม่

COMIT

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $l = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{l-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{l-1} ที่สอดคล้องกับวิธีการเลือก

$p_l(x) = p(x | D_{l-1}^N) = \prod_{i=1}^n p_l(x_i | x_{j(i)}) \leftarrow$ ทำการประมาณค่าการแจกแจงความน่าจะเป็นของผลเฉลยที่ถูกเลือกโดยใช้ขั้นตอนวิธี MWST ของไชว (Chow) และ ลิว (Liu) ในงานวิจัย (R. Etxeberria and P. Larrañaga, 1999)

$D_l^N \leftarrow$ ชักตัวอย่างผลเฉลย R_l (ประชากรรุ่นใหม่) จาก $p_l(x)$

$D_l^{F-N} \leftarrow$ ผลเฉลย $F = R - N$ ตัวจากกระบวนการค้นหาแบบเร็ว (Fast-Search Procedure) โดยเริ่มจากผลเฉลยที่ดีที่สุดของ D_l^N

$D_l \leftarrow D_l^{F-N} \cup D_{l-1}^N$

รูปที่ 6.9 รหัสเทียมของขั้นตอนวิธี COMIT

BMDA (Bivariate Marginal Distribution Algorithm) เป็นการนำเสนอแนวคิดการแจกแจงตัวประกอบของการแจกแจงความน่าจะเป็นร่วมสำหรับสถิติลำดับสอง (Second-order) โดยแนวคิดนี้อาศัยการสร้างกราฟพึ่งพิง (Dependency Graph) ที่ไม่มีวง (Acyclic) แต่ไม่จำเป็นจะต้องเป็นกราฟเชื่อมต่อ (Connected Graph) โดยทั่วไปกราฟพึ่งพิงสามารถมองในรูปเซตของ

ต้นไม้ที่ไม่มีส่วนเชื่อมต่อกัน วิธีนี้เริ่มจากการเลือกตัวแปรและเพิ่มเข้าไปเป็นปม (node) ของกราฟ โดยตัวแปรตัวแรกจะต้องมีค่าฟังก์ชันสูงสุดวัดโดยใช้ค่าสถิติของเพียสัน (Pearson's χ^2 Statistic) ขั้นตอนที่สองเป็นการเลือกปมต่อไปที่จะเพิ่มเข้าไปในกราฟโดยเลือกตัวแปรที่มีค่าฟังก์ชันรองลงมา และจะทำซ้ำจนกระทั่งไม่มีตัวแปรที่ผ่านค่าขีดแบ่งที่ตั้งไว้ ในแต่ละรุ่นการแยกตัวประกอบจะเป็นไปดังสมการ

$$p(x) = \prod_{X_r \in M_l} p_l(x_r) \prod_{X_i \in V \setminus M_l} p_l(x_i | x_{j(i)}) \tag{6.8}$$

โดยที่ V แทนเซตของตัวแปรจำนวน n ตัว และ M_l แทนเซตของตัวแปรรากในรุ่นที่ l สำหรับส่วนประกอบเชื่อมต่อกของกราฟฟังก์ชัน (Dependency Graph) และ $X_{j(i)}$ คือนค่าตัวแปรที่เชื่อมต่อกับตัวแปร X_i และมีการเพิ่มเข้าไปก่อน X_i

ความน่าจะเป็นสำหรับปมราก $p_l(x_r)$ และความน่าจะเป็นแบบมีเงื่อนไข $p_l(x_i | x_{j(i)})$ จะประมาณได้จากประชากร D_{l-1}^N ที่เป็นผลเฉลยที่ถูกเลือก

BMDA

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $l = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{l-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{l-1} ที่สอดคล้องกับวิธีการเลือก

$p_l(x) = \prod_{X_r \in M_l} p_l(x_r) \prod_{X_i \in V \setminus M_l} p_l(x_i | x_{j(i)}) \leftarrow$ ทำการประมาณการแยกประกอบ

$D_l^N \leftarrow$ ซักตัวอย่างผลเฉลย R_l (ประชากรรุ่นใหม่) จากกราฟฟังก์ชัน และ $p_l(x)$

จบการวนซ้ำ

รูปที่ 6.10 รหัสเทียมของขั้นตอนวิธี BMDA

6.2.3 แบบการฟังก์ชันเชื่อมโยงหลายเชิง (Multiple Dependencies)

มีหลายงานวิจัยได้นำเสนอแนวคิด EDA ในการแยกตัวประกอบการแจกแจงความน่าจะเป็นร่วมโดยใช้หลักสถิติของลำดับที่มากกว่าสอง ในรูปที่ 6.11 แสดงแบบจำลองความน่าจะเป็น

เป็นเชิงกราฟในหมวดนี้ ในหมวดนี้ความสัมพันธ์ของตัวแปรมีมากกว่าสองตัวที่ขึ้นแก่กันทำให้มีโครงสร้างความน่าจะเป็นที่ซับซ้อนทำให้การหาโครงสร้างที่เหมาะสมสำหรับแบบจำลองนั้นยุ่งยากขึ้น ดังนั้นจำเป็นต้องมีแนวคิดในกระบวนการเรียนรู้ที่ซับซ้อนขึ้นกว่าเดิมและจะกล่าวถึงเนื้อหาคร่าว ๆ ของแนวคิด EDA ที่สำคัญในหมวดนี้ซึ่งได้แก่

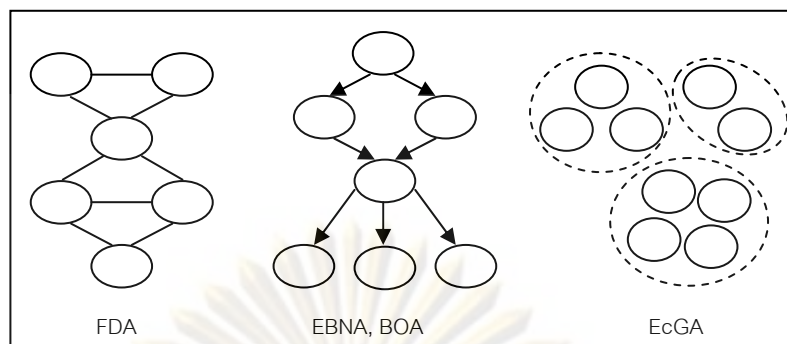
ขั้นตอนวิธีการแจกแจงตัวประกอบ (Factorized Distribution Algorithm: FDA) (Mühlenbein and Mahnig, 1999a, 1999b) ขั้นตอนวิธีนี้ได้ประยุกต์กับขั้นตอนวิธีสลายเชิงบวก (Additively Decomposed Functions) ซึ่งใช้การดำเนินการสมบัติส่วนร่วม (Running Intersection Property) เป็นการแยกของความน่าจะเป็นจำนวนมากบนพื้นฐานตกค้าง (Residuals) และตัวแบ่งแยก (Separators)

ในงานวิจัย (Etzeberria and Larrañaga, 1999) การแยกการแจกแจงความน่าจะเป็นร่วมถูกเข้ารหัสในรูปแบบของโครงข่ายเบย์ (Bayesian Network) ซึ่งใช้การเรียนรู้จากฐานข้อมูลที่บรรจุผลเฉลยที่ถูกเลือกในแต่ละรุ่น ขั้นตอนวิธีนี้มีชื่อว่า (Estimation of Bayesian Networks Algorithm: EBNA) และใช้คะแนนของเกณฑ์สารสนเทศแบบเบย์ (Bayesian Information Criterion: BIC) ในการวัดคุณภาพของโครงสร้างโครงข่ายเบย์กับขั้นตอนวิธีเชิงละโมบที่จะช่วยในการค้นหาในปริภูมิของแบบจำลอง

จากงานวิจัย (Pelikan, Goldberg and Cantú-Paz, 1999a) ผู้เขียนได้เสนอขั้นตอนวิธีที่เรียกว่า ขั้นตอนวิธีหาค่าที่ดีที่สุดแบบเบย์ (Bayesian Optimization Algorithm: BOA) ซึ่งใช้ตัววัดแบบเบย์ (Bayesian Metric) หรือเรียกว่า สมมูลตรีคัสเบย์ (Bayesian Dirichlet equivalent: BDe) (Heckman, Geiger and Chickering, 1995) ในการวัดค่าความดีของโครงสร้างทุกแบบที่พบ และแนวคิดนี้จะมีการใช้กระบวนการค้นหาแบบละโมบโดยการค้นหาจะเริ่มในแต่ละรุ่นอย่างคร่าว

ขั้นตอนวิธีเรียนรู้การแจกแจงตัวประกอบ (Learning Factorized Distribution Algorithm: LFDA) (Mühlenbein and Mahnig, 1999) จะใช้หลักการสำคัญ ๆ เหมือนกับแนวคิดของ EBNA

ขั้นตอนวิธีส่วนเพิ่มเชิงพันธุกรรมแบบกะทัดรัด (Extend compact Genetic Algorithm: EcGA) (Harik, 1999) เป็นขั้นตอนวิธีที่ใช้แนวคิดพื้นฐานในการแยกตัวประกอบของการแจกแจงความน่าจะเป็นร่วมโดยใช้ผลคูณของการแจกแจงตามขอบของขนาดตัวแปร



รูปที่ 6.11 การนำเสนอในรูปแบบกราฟของวิธีประมาณการแจกแจงในการหาค่าเหมาะที่สุดสมเชิงการจัด (Combinatorial Optimiazation) กับการพืงพืงเชื่อมโยงหลายเชิง ได้แก่ FDA, EBNA, BOA และ EcGA

EBNA (Estimation of Bayesian Networks Algorithm) วิธีนี้เรียนรู้การแจกแจงความความน่าจะเป็นผ่านโครงข่ายเบย์โดยเรียนรู้จากฐานข้อมูลที่บรรจุผลเฉลยที่ดี ๆ ในแต่ละรุ่น จากรูปที่ 6.12 จะเห็นว่ารอบการทำงานแรกโครงสร้างของโครงข่ายเบย์จะเป็นกราฟไม่มีวงแบบมีทิศทาง (Directed Acyclic Graph) ที่ไม่มีเส้นเชื่อมถึงกัน ดังนั้นการแยกตัวประกอบของการแจกแจงความน่าจะเป็นร่วมจะเป็นแบบการแจกแจงเอกรูปตัวแปรเดียวซึ่งหมายความว่าขั้นตอนเริ่มต้นโครงข่ายเบย์ BN_0 จะกำหนดความน่าจะเป็นให้กับทุกจุดในปริภูมิค้นหาที่เท่ากัน และในการที่เราจะทำการหาแบบจำลองโครงสร้างให้ได้เร็วที่สุดโดยไม่จำเป็นจะต้องเป็นโครงสร้างที่ดีที่สุดเราสามารถใช้อนุกรมวิธี B (Buntine, 1991) ซึ่งเป็นการค้นหาเชิงละโมบที่เริ่มจากโครงสร้างที่ไม่มีส่วนเชื่อมต่อกันและในแต่ละขั้นจะเพิ่มเส้นเชื่อมต่อโดยใช้ค่าที่มากที่สุดที่วัดได้ และจะหยุดเมื่อการเพิ่มเส้นเชื่อมต่อไม่ทำให้คะแนนที่ได้ดีขึ้น ส่วนวิธีอื่นนอกจากขั้นตอนวิธีนี้แล้วยังมีการใช้กลยุทธ์การค้นหาเฉพาะถิ่น (Local Search Strategies) กับแบบจำลองที่สร้างจากรุ่นก่อนหน้า เกณฑ์ที่ใช้ในการหาแบบจำลองโครงสร้างที่ดีนั้นจะต่างกัน แบบวิธีการให้คะแนนได้แก่วิธี *BIC* (Bayesian Information Criterion) เป็นวิธีค้นหาแบบจำลองที่เหมาะสมโดยใช้การทำโทษตัวแปรที่ทำให้มีต้นทุนสูง และวิธี $K2+pen$ ในส่วนของ $K2$ เป็นวิธีค้นหาแบบจำลองเชิงละโมบวิธีหนึ่ง ส่วน pen (Penalty) เป็นการทำโทษถ้าแบบจำลองที่ได้มีต้นทุนสูงกว่าค่าที่กำหนด เป็นต้น และแบบทดสอบการขึ้นแก่กันอย่างมีเงื่อนไขระหว่างตัวแปร ได้แก่วิธี *PC* (PC-algorithm) เป็นวิธีที่ใช้หาโครงสร้างของความสัมพันธ์ที่ขึ้นแก่กัน เป็นต้น แสดงดังรูปที่ 6.12 ในวิธี $EBNA_{k2+pen}$ จะมีเงื่อนไขเพิ่มขึ้นมาคือการจำกัดจำนวนของปมพ่อแม่ที่แต่ละปมจะมีได้ในโครงสร้างที่ดีที่สุดเป็นตัวเลขควบคุมความซับซ้อนของโมเดลโดยอัตโนมัติ

$EBNA_{pc}$, $EBNA_{k2+pen}$, $EBNA_{BIC}$

$BN_0 \leftarrow (S_0, \theta^0)$ โดย S_0 เป็น arcless DAG และ θ^0 เป็นเอกรูป

$$p_0(x) = \prod_{i=1}^n p_l(x_i) = \prod_{i=1}^n \frac{1}{r_i}$$

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $l = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{l-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{l-1}

$S_l^* \leftarrow$ ค้นหาโครงสร้างที่ดีที่สุดที่สอดคล้องกับเกณฑ์ต่อไปนี้

การทดสอบการขึ้นแก่กันอย่างมีเงื่อนไข $\rightarrow EBNA_{pc}$

(Conditional (in)dependent Test)

การลงโทษการให้คะแนนและค้นหาแบบเบย์ $\rightarrow EBNA_{k2+pen}$

(Penalized Bayesian Score+Search)

การลงโทษความควรจะเป็นสูงสุดและค้นหา $\rightarrow EBNA_{BIC}$

(Penalized Maximum Likelihood+Search)

$\theta^l \leftarrow$ คำนวณ θ_{ijk}^l โดยใช้ D_{l-1}^N เป็นเซตข้อมูล

$BN_l \leftarrow (S_l^*, \theta^l)$

$D_l^N \leftarrow$ ซักตัวอย่างผลเฉลย R_l (ประชากรรุ่นใหม่) จาก BN_l โดยใช้ PLS

จบการวนซ้ำ

รูปที่ 6.12 รหัสเทียมของขั้นตอนวิธี $EBNA_{pc}$, $EBNA_{k2+pen}$, $EBNA_{BIC}$

EcGA (Extend compact Genetic Algorithm) มีแนวคิดพื้นฐานมาจากการใช้แบบจำลองการคูณตามขอบ (Marginal Product Model) ในการประมาณการแจกแจงความน่าจะเป็นร่วมของผลเฉลยที่เลือกมาในแต่ละรุ่น โดยการแจกแจงตามขอบของตัวแปรจะเกี่ยวข้องกับตัวแปรหลายตัวที่อยู่ในกลุ่มเดียวกันและการแจกแจงความน่าจะเป็นของแต่ละตัวแปร การจัดกลุ่มทำได้โดยใช้ขั้นตอนวิธีล่วงหน้าเชิงละโมบ (Greedy Forward Algorithm) ในการหาการแบ่งกั้นระหว่างตัวแปรจำนวน n ตัว แต่ละกลุ่มตัวแปรจะถือว่าเป็นอิสระต่อกันแสดงดังรูปที่ 6.12 การแยกตัวประกอบของความน่าจะเป็นร่วมของตัวแปร n ตัวอยู่ในรูปสมการ

$$p_l(x) = \prod_{c \in C_l} p_l(x_c) \quad (6.9)$$

โดย C_l แทนเซตของกลุ่มในแต่ละรุ่น l^{th} และ $p_l(x_c)$ แทนการกระจายตามขอบของตัวแปร X_c ซึ่งเป็นตัวแปรของกลุ่มที่ c^{th} ในรุ่นที่ l^{th}

การแบ่งกันของเซตตัวแปรสำหรับทุกรุ่น l และทุกตัวแปรในแต่ละกลุ่ม $c, k \in C_l$ เป็นดังสมการ

$$\bigcup_{c \in C_l} X_c = \{x_1, \dots, x_n\}, X_c \cap X_k = \emptyset \quad (6.10)$$

ขั้นตอนวิธีเชิงละโมบจะทำการแบ่งกลุ่มโดยเริ่มแบ่งกันออกเป็น n กลุ่ม และจะทำการยุบเนียนตัวแปรสองตัวที่ทำให้ค่าสังยุค (Conjugate) ของผลบวกเอนโทรปีของการแจกแจงตามขอบลดลงมากที่สุดกับการทำโทษสำหรับแบบจำลองที่ซับซ้อนเกินไปโดยใช้ความยาวพรรณาน้อยสุด (Minimum Descriptive Length: MDL) (Rissanen, 1978) และเพื่อความถูกต้องมากขึ้นจะมีการลดค่าองค์ประกอบอีกสองตัวได้แก่

ความซับซ้อนของประชากรเชิงบีบอัด (Compressed Population Complexity) นิยามโดยใช้เอนโทรปีของการแจกแจงตามขอบ (Marginal Distribution) ดังสมการ

$$N \sum_{c \in C_l} h(X_c) = -N \sum_{c \in C_l} \sum_{x_c} p(X_c = x_c) \log p(X_c = x_c) \quad (6.11)$$

แบบจำลองความซับซ้อนที่คำนึงถึงมิติของแบบจำลองในสมการ

$$\log N \sum_{c \in C_l} \dim X_c \quad (6.12)$$

โดย $\dim X_c$ แทนจำนวนพารามิเตอร์ที่จำเป็นต้องใช้ในการระบุการแจกแจงตามขอบของ X_c ถ้าตัวแปรมิติเดียวทั้งหมดอยู่ในกลุ่ม c^{th} เป็นฐานสองแล้วจะได้ว่า $\dim X_c = 2^{|X_c|} - 1$ ทั้งสององค์ประกอบสามารถรวมเขียนเป็นสมการเดียวกันได้เป็น

$$-N \sum_{c \in C_l} \sum_{x_c} p(X_c = x_c) \log p(X_c = x_c) + \log N \sum_{c \in C_l} \dim X_c \quad (6.13)$$

ตัววัดนี้เรียกว่าความซับซ้อนรวม (Combine Complexity) (Harik, 1999) ในขั้นตอนการค้นหาเชิงละโมบที่ใช้ใน EcGA นั้นจะถือว่าตัวแปรเป็นอิสระต่อกันตอนเริ่มต้น และหลังจากนั้นจะทำการค้นหาการป็นขึ้นสูงชันที่สุด (Steepest Ascent Search) โดยในแต่ละขั้นนั้นจะพยายาม

รวมกลุ่มสองกลุ่มเข้าด้วยกันกลายเป็นกลุ่มที่ใหญ่ขึ้น การวัดคุณค่าของการรวมกันนี้จะใช้ค่าความซับซ้อนรวมโดยถ้ารวมกันแล้วค่าความซับซ้อนรวมลดลงจะถือว่ารวมสำเร็จ กระบวนการนี้จะทำจนกระทั่งไม่มีกลุ่มไหนสามารถรวมกันได้

EcGA

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $l = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{l-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{l-1} ที่สอดคล้องกับวิธีการเลือกแบบการแข่งขันกีฬา (Tournament Selection)

$p_l(x) = p(x | D_{l-1}^N) = \prod_{c \in C_l} p_l(x_c | D_{l-1}^N) \leftarrow$ ทำการประมาณการการแจกแจงความน่าจะเป็นของประชากรที่ถูกเลือกโดยใช้แบบจำลองการคูณตามขอบ (Marginal Product Model) แบบจำลองค้นหาโดยใช้การค้นหาการปีนขึ้นสูงชันที่สุด (Steepest Ascent Search) ลดรูปเป็น

$$-N \sum_{c \in C_l} \sum_{x_c} p(X_c = x_c) \log p(X_c = x_c) + \log N \sum_{c \in C_l} \dim X_c$$

$D_l^N \leftarrow$ ชักตัวอย่างผลเฉลย R_l (ประชากรรุ่นใหม่) จาก $p_l(x)$

จบการวนซ้ำ

รูปที่ 6.13 รหัสเทียมของ EcGA

BOA (Bayesian Optimization Algorithm) วิธีนี้จะใช้ตัววัดแบบเบย์ หรือเรียกว่า สมมูลตรีศลิเบย์ (BDe: Bayesian Dirichlet Equivalent) เป็นตัววัดคุณภาพของโครงข่ายซึ่งตัววัดนี้จะรวมเอาความรู้ที่มีอยู่ก่อนแล้ว (Prior Knowledge) เกี่ยวกับปัญหา และข้อมูลเชิงสถิติของชุดข้อมูลเข้าด้วยกัน เพื่อที่จะใช้บอกว่าข้อมูลที่มีสอดคล้องกับโครงข่ายระดับไหน ส่วนการค้นหาในแต่ละรุ่นจะใช้การค้นหาเชิงละโมภ (Greedy Search) การลดขนาดของปริภูมิค้นหาจะทำโดยการจำกัดจำนวนปมพ่อแม่ k ในโครงข่ายเบย์ และได้มีการประยุกต์แนวคิดนี้กับปัญหาเชิงลำดับชั้น (Hierarchical Problem) โดยใช้แบบจำลองผสมที่เรียกว่าโครงข่ายฮัฟฟ์แมน (Huffman Network) และมีการนำกราฟตัดสินใจ (Decision Graphs) มาใช้ในการหาโครงสร้างเครือข่ายที่เหมาะสมด้วย

BOA

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $l = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{l-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{l-1} ที่สอดคล้องกับวิธีการเลือก

$B \leftarrow$ สร้างโครงข่ายโดยใช้ตัววัดแบบเบส (BDe) และเงื่อนไขบังคับโดยใช้สมการโครงข่ายเบสเข้ารหัสการแจกแจงความน่าจะเป็นร่วม

$$p(X) = \prod_{i=0}^{n-1} p(X_i | \Pi_{X_i})$$

โดยที่ $X = (X_0, \dots, X_{n-1})$ เป็นเวกเตอร์ของตัวแปรทั้งหมดในปัญหา

Π_{X_i} เป็นเซตของบัพพ่อแม่ (Parents Node) ของตัวแปร X_i ในโครงข่าย

$p(X_i | \Pi_{X_i})$ เป็นความน่าจะเป็นมีเงื่อนไขของ X_i ต่อ Π_{X_i}

และมีการใช้ ความยาวตัววัดพรรณนาเล็กที่สุด (Minimal

Descriptive Length Metric: MDL) ร่วมด้วย

$D_l^N \leftarrow$ ชักตัวอย่างผลเฉลย R_l (ประชากรรุ่นใหม่) ที่สอดคล้องกับโครงสร้างการแจกแจงร่วมของ B

จบการวนซ้ำ

รูปที่ 6.14 รหัสเทียมของ BOA

6.3 ความแตกต่างระหว่าง GAs กับ EDAs ในมุมมองของข้อมูลสารสนเทศร่วม

สิ่งที่ทำให้ GAs มีความแตกต่างกับ EDAs อย่างเด่นชัดคือวิธีการดำเนินการโดย GAs จะอาศัยตัวดำเนินการที่สำคัญสองตัวคือ การไขว้เปลี่ยนกับการผ่าเหล่า ส่วน EDAs จะอาศัยการสร้างแบบจำลองความน่าจะเป็นของการกระจายตัว ได้มีงานวิจัย (Toussaint, 2003c) ที่ศึกษาถึงความแตกต่างระหว่าง EDAs กับ GAs ในแง่มุมมองวิธีการสำรวจโดยมีการนิยามคำว่า การสำรวจความสัมพันธ์ (Correlated Exploration) ขึ้นเพื่อใช้เป็นตัวแทนความสามารถในการสำรวจที่จะทำให้มีข้อมูลสารสนเทศร่วม (Mutual Information) เพิ่มขึ้นในประชากรรุ่นถัดไป ข้อมูลสารสนเทศร่วมนี้เป็นแหล่งข้อมูลที่สำคัญเกี่ยวกับปัญหาที่มีอยู่ในประชากรซึ่งสามารถใช้ในการแสวงหาประโยชน์สำหรับประชากรรุ่นถัดไปได้ และได้ใช้นิยามนี้เป็นเกณฑ์ในการพิจารณาความแตกต่างระหว่างขั้นตอนวิธีทั้งสอง และมีข้อพิสูจน์ทางคณิตศาสตร์ที่แสดงถึงบทบาทของตัวดำเนินการผ่าเหล่า

และการไขว้เปลี่ยนซึ่งสามารถสรุปได้ว่า การผ่าเหล่าจะเป็นตัวทำให้เอนโทรปี (Entropy) เพิ่มขึ้น และทำให้ข้อมูลสหเทศร่วมลดลง ส่วนการไขว้เปลี่ยนได้แสดงให้เห็นว่าการกระจายตัวตามขอบ ไม่ได้เปลี่ยนแปลงทั้งหมดและไม่มีการสำรวจเกิดขึ้นกับคู่จีนในแต่ละโครโมโซม การไขว้เปลี่ยนจะทำลายข้อมูลสหเทศร่วมที่มีในประชากรรุ่นพ่อแม่พันธุ์แม่พันธุ์โดยการแปลงให้อยู่ในรูปของเอนโทรปีของประชากรที่ผ่านการไขว้เปลี่ยนแล้ว ในกรณีที่ประชากรรุ่นพ่อแม่พันธุ์แม่พันธุ์ไม่มีข้อมูลสหเทศร่วมกัน การไขว้เปลี่ยนจะไม่ทำให้เกิดเอนโทรปีเพิ่มขึ้น ทฤษฎีบทนี้มีจุดประสงค์เพื่อต้องการแสดงให้เห็นถึงบทบาทของการไขว้เปลี่ยนในมุมมองที่แตกต่างออกไป จากทฤษฎีบทนี้จะเห็นได้ว่าถ้าไม่มีการไขว้เปลี่ยนใน GA จะทำให้เกิดหน่วยการสร้างที่ดีที่สุดและเป็นธรรมชาติมากที่สุด (หน่วยการสร้างที่มีข้อมูลสหเทศร่วมระหว่างคู่จีน) การไขว้เปลี่ยนจะเป็นตัวแบ่งแยกหน่วยการสร้างที่ดีที่สุดนี้ออกเป็นชิ้นเล็ก ๆ โดยการเปลี่ยนการพึ่งพิงร่วม (Mutual Dependencies) ไปเป็นเอนโทรปีผลที่ได้คือหน่วยการสร้างขนาดเล็กที่มีข้อมูลสหเทศร่วมกันน้อยลงในประชากรที่ผ่านการไขว้เปลี่ยนแล้ว หรือกล่าวได้ว่าความสัมพันธ์ระหว่างกันจะถูกทำลายและทำให้เกิดการกระจายตัวของหน่วยการสร้าง

การสำรวจหมายถึงการเพิ่มเอนโทรปีให้กับการแจกแจงค้นหา (Search Distribution) โดยทั่วไปแล้วเป็นการยากที่จะทำให้เอนโทรปีเพิ่มขึ้นโดยปราศจากการทำลายข้อมูลสหเทศร่วม การใช้การไขว้เปลี่ยนเพื่อเพิ่มเอนโทรปีจะทำให้สูญเสียข้อมูลสหเทศร่วม นอกจากความพยายามในการรักษาข้อมูลสหเทศร่วมในประชากรรุ่นพ่อแม่พันธุ์เหล่านี้ ควรจะมีวิธีในการนำข้อมูลสหเทศร่วมเหล่านี้ไปใช้ในการสำรวจครั้งต่อไปได้ วิธีที่เป็นไปได้คือการประมาณโครงสร้างของข้อมูลสหเทศเหล่านี้ที่มีอยู่ในประชากรรุ่นพ่อแม่พันธุ์ และทำการเพิ่มเอนโทรปีเข้าไปโดยที่โครงสร้างยังเหมือนเดิมซึ่งในมุมมองนี้จะตรงกับหลักการของ EDAs (ยกเว้นบางขั้นตอนวิธีที่ไม่มีการประมาณความสัมพันธ์ระหว่างคู่จีนเช่น PBIL, UMDA เป็นต้น) และเมื่อเปรียบเทียบกับ GA ที่แสดงให้เห็นว่าการไขว้เปลี่ยนเป็นตัวทำลายความสัมพันธ์ระหว่างคู่จีนในประชากร จะเห็นได้ว่า EDA สามารถประมาณการแจกแจงของประชากรได้ โดยขึ้นกับแบบจำลองความน่าจะเป็นที่ใช้ในการตรวจจับความสัมพันธ์ระหว่างคู่จีน

ถ้าการไขว้เปลี่ยนและการผ่าเหล่าไม่สามารถสำรวจความสัมพันธ์ได้ หลักการของการวิวัฒนาการในสิ่งมีชีวิตก็อาจถูกมองว่าไม่ถูกต้อง ในธรรมชาติจะมีรูปแบบพันธุกรรม (Genotype) และรูปแบบปรากฏ (Phenotype) ที่มีความสัมพันธ์กันแบบไม่ตรงไปตรงมาซึ่งแสดงให้เห็นในงานวิจัย (Toussaint, 2003b) แต่จากขั้นตอนการดำเนินการของ GAs จะอ้างอิงถึงการดำเนินการในปริภูมิรูปแบบพันธุกรรม (Genotype Space) ไม่ใช่ปริภูมิรูปแบบปรากฏ (Pheonotype Space) ในปริภูมิรูปแบบพันธุกรรมตัวดำเนินการผ่าเหล่าจะมีหน้าที่ทำให้เกิดความ

หลากหลาย และความเปลี่ยนแปลงในการสำรวจจะเกิดขึ้นอย่างจำกัดโดยที่ไม่มีความสัมพันธ์อย่างอื่นเข้ามาเกี่ยวข้อง แต่ถ้าอยู่ในปริภูมิรูปแบบปรากฏตัวดำเนินการผ่าเหล่าจะมีหน้าที่มากกว่าเพียงแค่สร้างความหลากหลายซึ่งในกรณีนี้จะมีความสัมพันธ์กับการสำรวจที่มีการเปลี่ยนแปลงรูปแบบปรากฏเข้ามาเกี่ยวข้อง หรือกล่าวได้ว่าจะมีเรื่องความเหมาะสมในการเกิดการผ่าเหล่าร่วมด้วย คำอธิบายแบบนี้สามารถใช้กับการไขว้เปลี่ยนได้เช่นกัน

นอกจากที่กล่าวมานี้ยังมีความแตกต่างระหว่าง GAs กับ EDAs ที่เป็นที่รู้จักกัน โดยทั่วไปในมุมมองของการปรับตัวเอง (Self-adaptation) ในการสำรวจการแจกแจง โดยที่ EDAs จะปรับการแจกแจงค้นหาวมถึงความสัมพันธ์อยู่เสมอเพื่อให้สอดคล้องกับการแจกแจงของผลเฉลยที่ถูกเลือก ซึ่งตรงข้ามกับ GAs ที่การไขว้เปลี่ยนจะพรางการกระจายในส่วนของความสัมพัทธ์ที่อาจถูกทำลายหรือไม่ถูกทำลายก็ได้ และในการไขว้เปลี่ยนโดยปกติจะเป็นแบบไม่ปรับตัวเอง

6.4 หน่วยการสร้างโดยปริยาย (Implicit BBs) และหน่วยการสร้างแบบชัดแจ้ง (Explicit BBs) (Day, Lamont, 2005a)

สำหรับขั้นตอนวิธีเชิงพันธุกรรมในมุมมองของการมีหรือการหาหน่วยการสร้างอาจมองว่ามันมีอยู่ภายในโครงสร้างของตัวมันเองซึ่งต่างจากขั้นตอนวิธีประมาณการแจกแจงที่สามารถหาหรือระบุได้แน่ชัดว่าหน่วยการสร้างมีหน้าตาเป็นอย่างไร ในทฤษฎีบทเคิร์ฟรังก็ได้มีการตีความว่าตัวดำเนินการอาจเป็นตัวทำลายหน่วยการสร้างที่ดี ๆ ทำให้ไม่สามารถแก้ปัญหาที่ยากได้ (Forrest and Mitchell, 1993; Khan, 2003) แต่ในขั้นตอนวิธีประมาณการแจกแจงนั้นการชัดขวางการสร้างหน่วยการสร้างที่ดี ๆ จะไม่เกิดขึ้น แต่ในทางกลับกันขั้นตอนวิธีจะพยายามหาความสัมพันธ์เหล่านี้แล้วรวมเข้าด้วยกันเพื่อให้ได้หน่วยการสร้างที่ดีขึ้น มีหลายงานวิจัยได้แสดงให้เห็นว่าขั้นตอนวิธีการค้นหาหน่วยการสร้างแบบชัดแจ้งดีกว่าขั้นตอนวิธีค้นหาหน่วยการสร้างโดยปริยาย (Day, 2005b)

ขั้นตอนวิธีแบบ EDAs สามารถอธิบายโดยใช้สมมติฐานหน่วยการสร้างแบบเดียวกับกับหน่วยการสร้างโดยปริยายได้ แต่ที่ใช้คำว่าหน่วยการสร้างแบบชัดแจ้งเนื่องจากสามารถแสดงให้เห็นเด่นชัดได้ว่าหน่วยการสร้างมีรูปแบบเป็นเช่นไร ที่เป็นเช่นนี้ก็เพราะว่า EDAs ถูกมองว่าเป็นการเรียนรู้และการเชื่อมโยง ซึ่งการเชื่อมโยงนี้จะเป็นสิ่งที่แสดงถึงหน่วยการสร้างได้ทั้งทางตรงและทางอ้อม สามารถพิจารณาได้ดังนี้

1. กรณีที่การเชื่อมโยงที่หาได้บอกได้แค่ว่าคุณจีนไหนอยู่ด้วยกันหรือไปด้วยกันบ้าง อาจมองว่าเป็นการแบ่งกัน (Partition) ถ้ามีการแบ่งแบบแยกขาดจากกัน หรือมองว่าเชื่อมให้มีการเหลื่อม (Overlap) กันระหว่างส่วนแบ่งกันได้ โดยนิยามของการแบ่งกันนั้นก็แล้วแต่ว่าใครจะนิยามว่าอย่างไร เช่น นิยามว่าส่วนแบ่งกันของคุณจีนที่เป็น 1 และคุณจีนที่เป็น 0 แยกขาดจากกัน เป็นต้น ในมุมมองนี้อาจกล่าวได้ว่าเป็นการพิจารณาหน่วยการสร้างแบบทางเชื่อม
2. กรณีที่สามารถระบุเจาะจงได้เลยว่าคุณจีนไหนมีความสัมพันธ์กันบ้าง และสามารถบอกได้ด้วยว่าคุณจีนที่สัมพันธ์กันเหล่านั้นควรจะมีความเป็นเท่าไร เช่น มีความเป็น 0 หรือเป็น 1 เป็นต้น ซึ่งมุมมองนี้อาจกล่าวได้ว่าเป็นการพิจารณาหน่วยการสร้างแบบชัดแจ้ง
3. กรณีที่แต่ละคุณจีนไม่ได้มีความสัมพันธ์กัน หรือมองว่าเป็นส่วนแบ่งกันแบบเล็กที่สุดคือแต่ละบิตแยกจากกันก็ได้ ได้แก่ EDAs จำพวกที่ปราศจากการพึ่งพิงระหว่างกัน มุมมองนี้เป็นมุมมองที่สามารถแบ่งปันความรู้ระหว่างโครโมโซมหลายตัวผ่านมาตรวัดอันหนึ่ง เพื่อที่จะบอกได้ว่าหน่วยการสร้างรูปแบบที่ดีควรจะมีรูปแบบใด ซึ่งถ้าพิจารณาจากมุมมองของเค้าร่างแล้วการกระทำแบบนี้แม้จะถูกมองว่าแต่ละคุณจีนเป็นอิสระจากกัน แต่ทว่าในที่สุดก็จะเรียนรู้เค้าร่างของโครโมโซมทั้งตัวออกมาได้ร่วมกัน หรืออาจมองว่าเป็นการเรียนรู้เค้าร่างแบบขนานเบ็ดเสร็จในโครงสร้างเดียวในขั้นตอนเดียวเหมือนตั้งตัวดำเนินการไขว้เปลี่ยนในขั้นตอนวิธีเชิงพันธุกรรม

จากมุมมองทั้งสามแบบแม้ว่าเราจะบอกว่าเป็นหน่วยการสร้างแบบชัดแจ้งแต่ในความเป็นจริงแล้วก็อาจไม่ได้แสดงให้เห็นเป็นรูปธรรมโดยตรง เนื่องจากสิ่งที่ใช้แสดงหน่วยการสร้างและเค้าร่างทั้งหมดยังอยู่ในรูปแบบของความน่าจะเป็นอีกทีหนึ่ง จึงไม่อาจกล่าวได้ว่าชัดแจ้งมากนัก จนกว่าจะมีกฎ หรือวิธีศึกษาสำนึกบางอย่างมาตัดสินค่าความน่าจะเป็นเหล่านั้นอีกทีหนึ่ง หรืออาจดูจากโครโมโซมรุ่นถัดไปที่ถูกสร้างขึ้นมาแล้ว

6.5 การเรียนรู้การเชื่อมโยง (Linkage Learning)

การเรียนรู้การเชื่อมโยง หรือเป็นที่รู้จักกันดีในนามว่าการระบุหน่วยการสร้างซึ่งได้กล่าวไปบ้างแล้วในส่วนก่อนหน้า และในส่วนนี้จะพุดให้เห็นถึงความแตกต่างระหว่างการเรียนรู้การเชื่อมโยงในแง่มุมมองอื่นซึ่งทำให้เข้าใจขั้นตอนวิธีประมาณการแจกแจงได้ดีขึ้นจากงานวิจัย (Chen, Yu, Sastry, Goldberg, 2007)

ในมุมมองของการจำแนกว่าการเชื่อมโยงไหนดีหรือไม่ดีนั้นสามารถมองได้จากตัววัดสำหรับขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัติ (Canonical GA) นั้นจะมองว่ามีเพียงตัววัดเดียว (Unimetric) นั่นก็คือค่าความเหมาะสม ต่างจากขั้นตอนวิธีประมาณการแจกแจงที่ใช้ตัววัดหลายตัว (Multimetric) นอกเหนือจากค่าความเหมาะสมโดยใช้ความรู้ด้านอื่นเข้ามาเสริมเช่น ด้านสถิติหรือความน่าจะเป็น เป็นต้น ขั้นตอนวิธีที่ใช้ตัววัดหลายตัวนั้นจะทำการเข้าถึงผลเฉลยทั้งหมดหรือบางส่วนเพื่อที่จะให้ได้ข้อมูลที่ต้องการซึ่งจะทำให้เกิดค่าใช้จ่ายในการแลกเปลี่ยนข้อมูลสนเทศระหว่างกันมาก และจะมีปัญหาคอขวดเมื่อต้องการแก้ปัญหาที่มีตัวแปรจำนวนมาก แต่ก็ทำให้สามารถเรียนรู้การเชื่อมโยงที่ซับซ้อนขึ้นได้

ในมุมมองการแทนการเชื่อมโยง (Represent Linkage) สามารถจำแนกได้เป็นสองพวกได้แก่ การเชื่อมโยงทางกายภาพ (Physical Linkage) และการเชื่อมโยงเสมือน (Virtual Linkage) สำหรับขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัติจะถูกจัดเป็นการเชื่อมโยงทางกายภาพซึ่งถูกผูกติดกับตำแหน่งของจีนตั้งแต่สองตัวขึ้นไปในโครโมโซม ดังรูปที่ 1 ส่วนขั้นตอนวิธีประมาณการแจกแจงจะอาศัยการสร้างแบบจำลองความน่าจะเป็น หรือการประมาณการแจกแจงแทนซึ่งเป็นการแทนการเชื่อมโยงทางอ้อม และสามารถนำความรู้ด้านอื่นมาช่วยในการตัดสินใจเพื่อให้ได้ผลที่ดีขึ้น

ในมุมมองการเก็บข้อมูลสนเทศการเชื่อมโยง (Store Linkage Information) สามารถจำแนกได้เป็นสองพวกได้แก่ แบบจำลองแบบกระจาย (Distributed Model) และแบบจำลองแบบรวมศูนย์ (Centralized Model) สำหรับขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัตินั้นจะไม่มีเก็บข้อมูลสนเทศการเชื่อมโยงร่วมกัน แต่จะเก็บอยู่ในโครโมโซม (ผลเฉลย) แต่ละตัวแทนซึ่งทำให้จัดเป็นแบบจำลองแบบกระจายซึ่งมีข้อดีในแง่ของการปรับมาตราส่วน (Scalability) เมื่อเจอกับปัญหาที่มีขนาดใหญ่ ต่างจากขั้นตอนวิธีประมาณการแจกแจงซึ่งเก็บข้อมูลสนเทศการเชื่อมโยงรวมไว้ด้วยกัน เช่น ตัวแปรเวกเตอร์ความน่าจะเป็นส่วนกลาง (Global Probabilistic Vector) หรือตารางพึ่งพิง (Dependency Table) เพื่อจัดกระทำและประมวลการเชื่อมโยง เป็นต้น ซึ่งข้อดีก็คือสามารถแลกเปลี่ยนข้อมูลสนเทศเกี่ยวกับการเชื่อมโยงได้เร็วเมื่อเทียบกับแบบจำลองแบบกระจาย อาจกล่าวได้ว่าขั้นตอนวิธีประมาณการแจกแจงโดยส่วนใหญ่ได้อาศัยวิธีการจัดกลุ่ม (Grouping) การเกาะกลุ่ม (Clustering) และการแบ่งส่วน (Partitioning) เป็นกลไกในการตรวจหา เรียนรู้ และใช้ประโยชน์จากการเชื่อมโยงหรือการระบุนหน่วยการสร้างจากแบบจำลองความน่าจะเป็น ซึ่งเหมาะกับปัญหาที่มีความซับซ้อนมากเกินไปสำหรับการแทนแบบธรรมดา (Simple Representation) หรือปัญหาที่ยาก

บทที่ 7

ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม

(Random Fragmentation Algorithm: RFT)

ในบทนี้จะนำเสนอวิธีการระบุหน่วยการสร้างและการประกอบเข้าด้วยกันวิธีใหม่ที่อาศัยหลักการแบ่งปันข้อมูลแบบขั้นตอนวิธีประมาณการแจกแจงและวิธีการไขว้เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์เป็นหลัก เพื่อทดสอบการแบ่งปันความรู้ (Shared Knowledge) แบบรวมศูนย์ (Centralized) ว่าให้ผลแตกต่างจากแบบกระจาย (Distributed) (ที่ใช้ใน sGA) หรือไม่อย่างไร

7.1 แนวคิดและหลักการ

กระบวนการวิวัฒนาการมีความเชื่อพื้นฐานว่าประชากรตัวที่เหมาะสม (โครโมโซมตัวที่มีความเหมาะสม) จะมีโอกาสอยู่รอดสูง ดังนั้นจึงให้ความสำคัญกับการตีความเรื่องความเหมาะสมและหาวิธีการคัดเลือกประชากรตัวที่เหมาะสมเพื่อนำไปใช้สืบพันธุ์ต่อไป เดิมทีรู้แต่ว่าตัวที่มีความเหมาะสมสูงน่าจะมีอะไรบางอย่างที่พิเศษกว่าตัวอื่น ๆ (ที่มีค่าความเหมาะสมต่ำกว่า) จนกระทั่งมีการนำเสนอทฤษฎีบทเคาท์ริงขึ้นจึงได้มีข้อสันนิษฐานเกี่ยวกับโครงสร้างหรือองค์ประกอบที่ดีที่มีอยู่ในโครโมโซมที่มีความเหมาะสมสูงว่าควรมีลักษณะหรือคุณสมบัติเป็นเช่นใด จากทฤษฎีบทเคาท์ริงที่ได้กล่าวอย่างละเอียดในบทที่ 3 เราได้ข้อสรุปสั้น ๆ เพียงว่าโครโมโซมตัวที่มีค่าความเหมาะสมสูงจะมีส่วนประกอบของเคาท์ริงที่ดีที่มีลักษณะดังนี้คือ มีความยาวเคาท์ริงที่สั้น และมีอันดับของเคาท์ริงที่ต่ำ หรือกล่าวอีกนัยหนึ่งได้ว่าประกอบไปด้วยชิ้นส่วนขนาดเล็ก ๆ นั่นเอง ซึ่งชิ้นส่วนเหล่านี้เป็นส่วนหนึ่งของโครงสร้างคำตอบที่ต้องการ หรือมีส่วนทำให้ไปถึงคำตอบที่ต้องการได้ สิ่งเหล่านี้ทำให้เราสามารถอนุมานได้ว่าเราควรเลือกเฉพาะโครโมโซมที่มีค่าความเหมาะสมสูง หรือเลือกโครโมโซมที่มีค่าความเหมาะสมสูงเป็นส่วนใหญ่ ด้วยเหตุนี้ทำให้กระบวนการเลือกสรรส่วนใหญ่มักจะให้โอกาสโครโมโซมที่มีค่าความเหมาะสมสูงกว่ามีโอกาสถูกเลือกมากกว่า เช่น วิธีการเลือกแบบสัดส่วน หรือวิธีการเลือกเชิงอันดับ เป็นต้น ผลการทดลองในงานวิจัยส่วนใหญ่ให้ผลลัพธ์ที่ตรงกับทฤษฎี (เลือกตัวที่มีค่าความเหมาะสมสูงให้ผลดีกว่าเลือกตัวที่มีค่าความเหมาะสมต่ำ) ในงานวิจัยนี้จึงเลือกใช้วิธีการเลือกสรรเชิงอันดับเป็นหลัก และสำหรับการทดลองในบทนี้ต้องการเน้นย้ำให้เห็นถึงข้อเท็จจริงในส่วนนี้ว่ายังคงเป็นไปตามทฤษฎีหรือไม่ เพื่อให้มีความชัดเจนจึงใช้วิธีการเลือกแบบตัวที่ดีที่สุดเป็นจำนวน n ตัว (n -Best Selection) (รายละเอียดเพิ่มเติมอยู่ในบทที่ 4 หัวข้อ 4.4)

การใช้พ่อพันธุ์แม่พันธุ์มากกว่าอย่างละหนึ่งตัวในการไขว้เปลี่ยน (หรือการแลกเปลี่ยนข้อมูลกันในรูปแบบใดรูปแบบหนึ่ง) แต่ละครั้งเป็นการแสดงให้เห็นถึงการแลกเปลี่ยนส่วนประกอบ

ที่อยู่ในโครโมโซมที่ถูกเลือก หรือมองได้ว่าเป็นการแลกเปลี่ยนข้อมูลหรือความรู้ทางโครงสร้างภายในกลุ่มโครโมโซมย่อยที่ดีที่สุด (มีค่าความเหมาะสมสูง) การทดลองส่วนใหญ่ให้ผลที่ดีขึ้นเมื่อเทียบกับวิธีการเดิมที่ใช้พ่อพันธุ์แม่พันธุ์รวมกันเพียงแค่ 2 ตัว ซึ่งเป็นการแสดงให้เห็นว่าโครโมโซมที่มีค่าความเหมาะสมมีส่วนประกอบของเค้าร่างที่ดี (ที่จะนำไปสู่ค่าตอบ) อยู่จริงตรงตามความเชื่อทางทฤษฎี แต่ปัญหาคือเรามีความรู้เพียงแค่ว่ามีส่วนประกอบหรือเค้าร่างที่ดี (หน่วยการสร้าง) อยู่จริง แต่เราไม่รู้ว่าจะเค้าร่างเหล่านั้นมีรูปแบบหน้าตาเป็นเช่นใด เกิดการผสมหรือแลกเปลี่ยนกันในลักษณะใดจึงทำให้ได้คำตอบออกมาดี ทั้งนี้เป็นเพราะตัวดำเนินการที่ใช้ซึ่งก็คือตัวดำเนินการไขว้เปลี่ยนแบบหลายจุด เป็นตัวดำเนินการที่ทำความเข้าใจได้ยากเนื่องจากมีความซับซ้อนสูง (เราไม่รู้ว่าเค้าร่างที่ไขว้เปลี่ยนกันมีรูปแบบหน้าตาเป็นอย่างไร และถึงแม้ว่ารู้ก็เป็นการยากที่จะติดตามการดำเนินการและทำความเข้าใจ) จากนิยามชิ้นส่วนในบทที่ 4 จะช่วยทำให้เราเห็นภาพได้ชัดเจนขึ้นด้วยวิธีดังต่อไปนี้

เราสามารถนิยามชิ้นส่วนด้วยโครงสร้างย่อยของโครโมโซมใด ๆ กำหนดให้ลำดับของโครโมโซมความยาว l

$$C^k = c_1^k c_2^k \dots c_l^k : k \in I^+, c_n^k \in \{0,1\}, 1 \leq n \leq l \quad (7.1)$$

กำหนดให้เซตของดัชนีตำแหน่งอ้างอิงในโครโมโซมความยาว l จากตำแหน่งที่ f ถึงตำแหน่งที่ t

$$(f_i, t_i) = \{ z_i : 1 \leq f_i \leq z_i \leq t_i \leq l, f_i, t_i, z_i \in I^+ \} \quad (7.2)$$

and $\forall i, j \in I^+ : i \neq j, (f_i, t_i) \cap (f_j, t_j) = \emptyset$

กำหนดให้ลำดับย่อยของโครโมโซม C^{k_i} เป็น

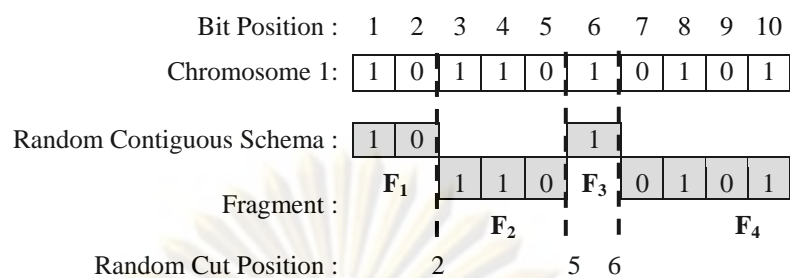
$$F_i^k = c_{f_i}^k c_{f_i+1}^k \dots c_{t_i}^k \quad (7.3)$$

เซตของลำดับย่อยที่เรียงติดกันในโครโมโซมนิยามดังนี้

$$S^k = \{ F_i^k : i \leq l \} \quad (7.4)$$

จากนิยามเราสามารถแสดงตัวอย่างได้ดังรูปที่ 7.1 กำหนดให้โครโมโซมเรียงลำดับจากซ้าย (เริ่มต้น) ไปขวา สมมติให้โครโมโซมขนาด 10 บิต มีค่าดังนี้ $C^1 = (1,0,1,1,0,1,0,1,0,1)$ และมีจำนวนจุดตัดที่สุ่มได้ทั้งหมด 3 จุดดังนี้ $(2,5,6)$ จะได้ช่วงดัชนีตำแหน่งในโครโมโซม C^1 เป็น $(f_1, t_1) = (1,2), (f_2, t_2) = (3,5), (f_3, t_3) = (6,6)$ และ $(f_4, t_4) = (7,10)$ และจะได้ชิ้นส่วน S^1 ดังต่อไปนี้ $F_1^1 = c_1^1 c_2^1 = 10, F_2^1 = c_3^1 c_4^1 c_5^1 = 110, F_3^1 = c_6^1 = 1$ และ $F_4^1 = c_7^1 c_8^1 c_9^1 c_{10}^1 = 0101$

ตามลำดับ



รูปที่ 7.1 ชิ้นส่วนที่ระบุได้จากโครโมโซม

D คือ ประชากร

F คือ ชิ้นส่วน (Fragments)

T คือ หน่วยเก็บชิ้นส่วนแบบถาวร (Archive)

L คือ ขนาดของผลเฉลย (Individual Length)

ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (Random Fragmentation)

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{g-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{g-1} ที่สอดคล้องกับวิธีการเลือก

$P \leftarrow$ ผลเฉลยจาก D_{g-1}^N แต่ละตัวจะถูกตัดออกเป็นชิ้นส่วนย่อย ๆ

แบบสุ่ม โดยขนาดของแต่ละชิ้นส่วนยาว m โดยที่ $1 \leq m \leq L$

$T \leftarrow$ สร้างหน่วยเก็บชิ้นส่วนถาวร (Archive) จาก P

$D_g^N \leftarrow$ ซักตัวอย่างผลเฉลย R_g (ประชากรรุ่นใหม่) จาก T

$D_g^N \leftarrow$ นำผลเฉลยจากทุกตัวจาก D_g^N ไปทำการผ่าเหล่า (Mutation)

ด้วยวิธีแบบ sGA (เป็นทางเลือก แต่งานวิจัยนี้ไม่ใช้)

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 7.2 รหัสเทียมของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (RFT)

สำหรับวิธี RFT ในส่วนของการระบุชิ้นส่วนจะใช้วิธีการสุ่มตัดโดยเลือกจำนวนจุดตัดจากค่าที่เป็นไปได้ทั้งหมดซึ่งมีค่าสูงสุดเท่ากับ $l-1$ เมื่อ l คือความยาวโครโมโซม และหลังจากนั้นจะทำการสุ่มเพื่อเลือกว่าจะให้ตัดที่ตำแหน่งไหนบ้าง จากรูปที่ 7.1 สมมติให้โครโมโซมมีความยาว 10 บิต ดังนั้นจำนวนจุดตัดที่เป็นไปได้มากที่สุดคือ $10-1 = 9$ จุด หลังจากนั้นจะทำการสุ่มจำนวนจุดตัดซึ่งมีค่าอยู่ในช่วง $[0, 9]$ กำหนดให้สุ่มได้ค่าเป็น 3 จากนั้นเราจะทำการสุ่มหาตำแหน่งของ

จุดตัดทั้ง 4 จุดอีกทีหนึ่ง กำหนดให้ส้มค่าตำแหน่งได้เป็น 6, 2 และ 5 จากนั้นเราจะทำการเรียงลำดับค่าตำแหน่งที่ส้มได้อีกทีหนึ่ง (จากน้อยไปมาก) มีค่าเป็น 2, 5 และ 6 มาถึงตรงนี้เราจะเห็นภาพขึ้นส่วนได้ชัดขึ้นว่ามีจำนวนทั้งหมด 4 ชิ้น ดังรูปที่ 7.1

จากวิธีการที่นำเสนอใช้วิธีการระบุชิ้นส่วนแบบส้ม หรือใช้วิธีการส้มตัดโครโมโซมที่ละตัวออกเป็นชิ้นเล็ก ๆ มีขนาดชิ้นเล็กสุดที่เป็นไปได้คือ 1 คู่จีน และใหญ่ที่สุดที่เป็นไปได้คือขนาดเท่าความยาวโครโมโซม (ไม่ถูกตัดเลยนั่นเอง) ที่เราใช้วิธีการส้มตัดแบบนี้ก็เพื่อเป็นการแสดงให้เห็นว่าในโครโมโซมที่ดี ๆ (มีค่าความเหมาะสมสูง) มีองค์ประกอบที่คืออยู่จริงแม้ว่าเราจะไม่รู้ว่าส่วนประกอบเหล่านั้นมีหน้าตาเป็นอย่างไรซึ่งตรงกับวิธีการไขว่เปลี่ยนแบบหลายจุดที่จะทำการส้มจุดตัดขึ้นมาจำนวนหนึ่งทำให้โครโมโซมถูกแบ่งออกเป็นหลายชิ้นส่วนแบบส้มเหมือนกัน แต่ส่วนสำคัญหลังจากที่ได้ชิ้นส่วนแล้วก็คือ การประกอบเข้าด้วยกันซึ่งทำได้หลายวิธี จากเนื้อหาที่นำเสนอในบทที่ 6 ที่กล่าวถึงขั้นตอนวิธีประมาณการแจกแจง และงานวิจัยเกี่ยวกับการไขว่เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์ (Eiben, 1997) มีหลักการดำเนินการที่เหมือนกันอยู่อย่างหนึ่งก็คือวิธีการเลือกของทั้งสองวิธีจะทำการเลือกประชากรตัวที่ดี ๆ มาจำนวนหนึ่ง และทำการแลกเปลี่ยนข้อมูลหรือความรู้ภายในกลุ่มประชากรที่ถูกเลือกซึ่งให้ผลแตกต่างกันดังนี้ ขั้นตอนวิธีประมาณการแจกแจงจะได้ผลในรูปแบบแบบจำลองความน่าจะเป็นของโครงสร้างที่มีโอกาสเกิดขึ้นได้ (จะถูกนำไปใช้ในการสร้างประชากรรุ่นต่อไปอีกทีหนึ่ง) ส่วนวิธีการไขว่เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์จะให้ผลเป็นโครโมโซมจำนวนหนึ่งผ่านการไขว่เปลี่ยนมาแล้ว (มีจำนวนได้ตั้งแต่ 1 ตัวจนถึงมีขนาดเท่าประชากร) หรือสามารถกล่าวได้ว่าทั้งสองวิธีมีการแบ่งปันข้อมูลหรือความรู้ระหว่างโครโมโซมที่มีจำนวนมากกว่า 2 ตัว แต่อย่างไรก็ตามทั้งสองวิธีนี้มีความแตกต่างกันอย่างมีนัยสำคัญในเรื่องการเก็บสังขมความรู้ที่ได้ ขั้นตอนวิธีประมาณการแจกแจงมีการสังขมข้อมูลแบบรวมศูนย์เพราะข้อมูลทุกอย่างที่ได้จะถูกเก็บไว้ในรูปแบบจำลองความน่าจะเป็น ซึ่งต่างจากวิธีการไขว่เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์ที่ไม่มีการสังขมความรู้เพราะเมื่อไขว่เปลี่ยนเสร็จก็จะถือว่าข้อมูลที่มีอยู่เกิดการแลกเปลี่ยนกันเรียบร้อยแล้ว (ได้ผลเป็นโครโมโซมรุ่นลูกแล้ว) จากผลการทดลองในหลายงานวิจัยทำให้เชื่อว่าขั้นตอนวิธีประมาณการแจกแจงมีประสิทธิภาพในการหาคำตอบสูงกว่าวิธีการไขว่เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์ ซึ่งน่าจะเป็นผลมาจากหลักการแบ่งปันความรู้แบบรวมศูนย์นั่นเอง เป็นเหตุให้การทดลองในบทนี้ใช้วิธีการสังขมความรู้แบบรวมศูนย์ในการประกอบเข้าด้วยกัน แต่จะไม่เก็บข้อมูลหรือความรู้ในรูปแบบความน่าจะเป็น เนื่องจากทำให้เห็นภาพกระบวนการทำงานและสิ่งที่เกิดขึ้นระหว่างกระบวนการทำงานได้ยาก (ทำให้เข้าใจยาก) และมีความไม่แน่นอนสูง แต่จะเก็บข้อมูลเหล่านี้ในรูปแบบชิ้นส่วนแทน

หลังจากที่ระบุชิ้นส่วนได้แล้วด้วยการสุ่มตัดโครโมโซมแต่ละตัวที่ถูกเลือกมาจำนวนหนึ่ง จะนำชิ้นส่วนที่ได้เหล่านี้ไปเก็บรวมกันไว้ในหน่วยเก็บถาวร (Archive) ซึ่งมีการจัดเรียงชิ้นส่วนที่ได้เหล่านี้ตามตำแหน่งที่เกิดขึ้นตรงกับตำแหน่งในโครโมโซม ทั้งนี้เป็นเพราะต้องการที่จะประกอบชิ้นส่วนที่ได้เหล่านี้กลับเป็นโครโมโซมได้อย่างถูกต้องตรงตามตำแหน่งเดิม วิธีการที่ง่ายและตรงไปตรงมาวิธีหนึ่งในการประกอบคือ วิธีการประกอบชิ้นส่วนเข้าด้วยกันตามลำดับตำแหน่งที่เกิดขึ้น โดยที่ในแต่ละตำแหน่งอาจมีชิ้นส่วนมากกว่าหนึ่งรูปแบบ (หน้าตาและความยาว) แต่จะต้องไม่มีรูปแบบที่ซ้ำกัน จะทำการประกอบโครโมโซมลูกโดยเริ่มจากตำแหน่งแรก (ซ้ายสุดของโครโมโซม) ด้วยวิธีการสุ่มเลือกชิ้นส่วนที่เริ่มต้นที่ตำแหน่งแรกมาหนึ่งชิ้นส่วน (สุ่มเลือกแบบเอกรูปคือ ทุกชิ้นส่วนมีโอกาสถูกเลือกเท่ากันหมด) และทำการเลือกชิ้นส่วนถัดไปเพื่อมาต่อท้ายชิ้นส่วนแรก และทำซ้ำจนกระทั่งมีค่าครบทุกตำแหน่งความยาวโครโมโซม ตัวอย่างเช่น โครโมโซมมีความยาว 10 บิต สุ่มชิ้นส่วนแรกขึ้นมาได้มีความยาว 3 บิต ดังนั้นชิ้นส่วนถัดไปที่จะสุ่มเลือกต้องเริ่มต้นที่ตำแหน่งที่ 4 (เพื่อไม่ให้เหลื่อมกับชิ้นส่วนแรก) สุ่มเลือกชิ้นส่วนที่สองมีความยาว 4 บิต (เมื่อรวมกับชิ้นส่วนแรกจะเห็นว่ามี ความยาวทั้งหมด 7 บิต) ชิ้นส่วนถัดไปต้องเริ่มตำแหน่งที่ 8 ถ้าสุ่มเลือกชิ้นส่วนที่สามมีความยาว 2 บิต (มีความยาวเกิน 2 บิตไม่ได้เนื่องจากโครโมโซมมีความยาวทั้งหมดเพียงแค่ว่า 10 บิต) จะถือว่าได้ชิ้นส่วนมาสร้างเป็นโครโมโซมลูกครบถ้วนสมบูรณ์ (โครโมโซมลูกตัวนี้ประกอบขึ้นจากชิ้นส่วนจำนวน 3 ชิ้นส่วน) เป็นต้น กรณีที่ตำแหน่งที่เลือกไม่มีชิ้นส่วนเกิดขึ้นเลยจะใช้วิธีการสุ่มค่ารูปแบบคู่จิ้นที่เป็นไปได้แทน โดยจะทำการสุ่มทีละตำแหน่งจนกว่าจะเจอตำแหน่งที่มีชิ้นส่วนให้เลือก หรือสิ้นสุดความยาวโครโมโซม วิธีการดำเนินการประกอบชิ้นส่วนแสดงดังรูปที่ 7.3 จะสร้างโครโมโซมรุ่นลูกขึ้นมาด้วยวิธีเดียวกันนี้จนกระทั่งมีจำนวนเท่ากับขนาดของประชากร ส่วนภาพรวมของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่มแสดงดังรูปที่ 7.4 และ 7.5

R คือ ผลเฉลย (โครโมโซม)

F คือ ชิ้นส่วนย่อย

N คือ จำนวนผลเฉลยที่ถูกเลือก

L คือ ขนาดความยาวของโครโมโซม

X คือ เซตของตำแหน่งจุดตัดในโครโมโซม

ขั้นตอนวิธีการระบุชิ้นส่วน

วนซ้ำสำหรับ $i = 1$ ถึง N ทำ

{ ทำการหาจุดตัดทั้งหมด }

$q =$ สุ่มเลือกค่าจำนวนเต็มในช่วง (1 ถึง $L-1$)

วนซ้ำสำหรับ $j = 1, 2, \dots$ ถึง q ทำ

$X \leftarrow$ สุ่มเลือกค่าจำนวนเต็ม I_j ในช่วง (1 ถึง $L-1$) { เลือกตำแหน่งของจุดตัด }

จบการวนซ้ำ

{ ทำการหาชิ้นส่วนทั้งหมด }

$X \leftarrow$ เรียงลำดับสมาชิกใน X จากน้อยไปมาก

$j = 1$ { กำหนดค่าเริ่มต้นดัชนีตำแหน่งของผลเฉลย }

วนซ้ำสำหรับ $k = 1, 2, \dots$ ถึง $|X|$ { ขนาดของเซต } ทำ

$F \leftarrow$ คัดลอก R_i จากตำแหน่งที่ j ถึงตำแหน่งที่ $(j + X_k)$

$j = j + X_k$ { กำหนดตำแหน่งเริ่มต้นถัดไป }

จบการวนซ้ำ

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 7.3 รหัสเทียมของวิธีการระบุชิ้นส่วนสำหรับวิธี RFT

R คือ ผลเฉลย (โครโมโซม)

F_i คือ เซตของชิ้นส่วนย่อยเริ่มต้นที่ตำแหน่ง i .

C คือ ชิ้นส่วนย่อย (แบบร่วมกัน และแบบไม่ร่วมกัน)

ขั้นตอนวิธีการประกอบชิ้นส่วน

{การประกอบผลเฉลยใหม่ R }

$i = 1$ {ตำแหน่งแรก}

วนซ้ำในขณะที่ $i < \text{individual_size}$ ทำ

 ถ้า $F_i \neq \emptyset$ ทำ {ถ้ามีชิ้นส่วนย่อยตั้งแต่หนึ่งชิ้น}

$C \leftarrow \text{สุ่มเลือก}(F_i)$

$R \leftarrow R + C$ {นำชิ้นส่วนมาเรียงต่อกันตามลำดับ}

$i \leftarrow i + \text{size}(C)$ {หาตำแหน่งเริ่มต้นถัดไป}

นอกจากนั้น {ถ้าไม่มีสมาชิกอยู่เลย}

$R \leftarrow \text{สุ่มเลือก}(0,1)$ {สุ่มจากค่า 0 หรือ 1}

$i \leftarrow i + 1$

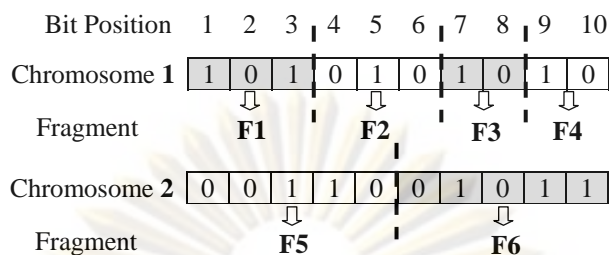
จบเงื่อนไข

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 7.4 รหัสเทียมของวิธีการประกอบชิ้นส่วนสำหรับวิธี RFT

Fragment Identification

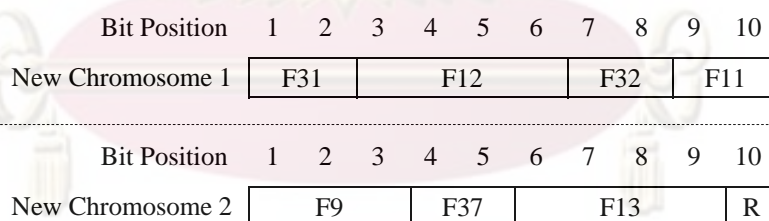


Composition Table (Archive)										
No. \ Position	1	2	3	4	5	6	7	8	9	10
1	F1			F2			F3		F4	
2	F5					F6				
...										

After random cut all individuals

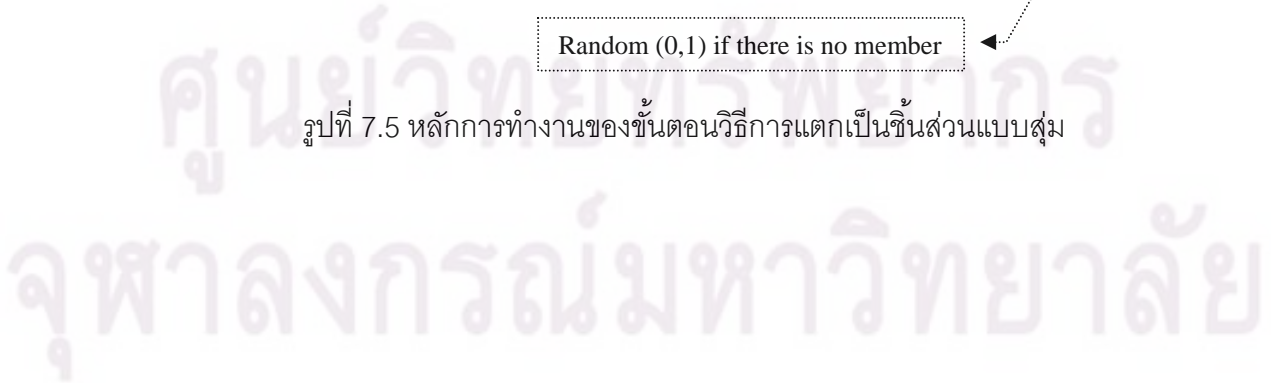
Composition Table (Archive)										
No. \ Position	1	2	3	4	5	6	7	8	9	10
1	F1	F22	F12	F2	F7	F6	F3	F8	F4	
2	F5	F27	F25	F14	F20	F13	F26	F10	F11	
3	F9		F17	F15	F23	F28	F18	F16	F21	
4	F19		F29	F37	F35		F30	F24	F34	
5	F31		F33		F39		F32	F38		
6	F36							F40		
...										

Fragment composition (new individual)



Random (0,1) if there is no member

รูปที่ 7.5 หลักการทำงานของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม



จากขั้นตอนวิธีที่ออกแบบสามารถคำนวณหาโอกาสที่ขึ้นส่วนจะอยู่รอดจากการตัวเนินการต่าง ๆ เพื่อให้เห็นภาพรวมของกลไกการทำงานอย่างคร่าว การคำนวณจะพิจารณาเพียงรุ่นประชากรเดียว (จากรุ่นสู่รุ่น) จากรุ่น t ไป $t+1$ ดังต่อไปนี้

โอกาสที่เค้าร่าง H ใด ๆ จะเกิดขึ้นเป็นดังนี้

$$P_{appear}(H) = 2^{-o(H)} \quad (7.5)$$

เมื่อ $P_{appear}(H)$ คือ โอกาสเกิดเค้าร่าง H จากวิธีการสุ่ม
 $o(H)$ คือ ลำดับของเค้าร่าง H

โอกาสที่เค้าร่าง H จะสามารถอยู่รอดได้จากการระบุขึ้นส่วนแบบสุ่มเป็น

$$P_{identify}(H) = \left(1 - \frac{D(H)}{l-1}\right)^q \quad (7.6)$$

เมื่อ $P_{identify}(H)$ คือ โอกาสที่เค้าร่าง H จะรอดจากวิธีการระบุขึ้นส่วนแบบสุ่ม
 $D(H)$ คือ ความยาวของเค้าร่าง H
 l คือ ขนาดความยาวของโครโมโซม
 q คือ จำนวนครั้งของการหาจุดตัด

โอกาสที่เค้าร่าง H จะสามารถอยู่รอดได้จากการประกอบขึ้นส่วนเข้าด้วยกัน โดยมีเงื่อนไขว่าในหน่วยเก็บขึ้นส่วนถาวรจะต้องมีขึ้นส่วนที่เป็นไปได้ทุกขนาดและรูปแบบ เป็นดังนี้

$$P_{compose}(H) = \sum_{b=1}^p \left(\frac{\sum_{j=(p-b)+(D(H)+1-o(H))}^{l-o(H)-b+1} 2^j}{\sum_{i=1+(p-b)}^{l-b+1} 2^i} \right) \quad (7.7)$$

เมื่อ $P_{compose}(H)$ คือ โอกาสเค้าร่าง H จะรอดจากวิธีการประกอบเข้าด้วยกัน

$D(H)$ คือ ความยาวของเค้าร่าง H

$o(H)$ คือ ลำดับของเค้าร่าง H

l คือ ขนาดความยาวของโครโมโซม

p คือ ตำแหน่งแรกของเค้าร่างที่มีค่าแน่นอน (fixed bit) โดยอ้างอิงจากตำแหน่งของโครโมโซม

b, i, j คือ ค่าดัชนีเพื่อใช้ในการคำนวณ

เมื่อพิจารณาถึงตัวดำเนินการเลือกสรรที่ใช้เป็นการเลือกแบบ n ตัวที่ดีที่สุด สามารถหาความสัมพันธ์ของสัดส่วนความอยู่รอดของจำนวนโครโมโซมที่มีเค้าร่าง H ได้เป็น

$$P_{select}(H) = \frac{s}{n} P_{(H)} \quad (7.8)$$

เมื่อ $P_{select}(H)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากรที่ถูกเลือก
 $P_{(H)}$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากร
 s คือ ขนาดของประชากร (จำนวนโครโมโซมที่มีทั้งหมด)
 n คือ จำนวนโครโมโซม (ที่ดีที่สุด) ที่ถูกเลือก

ถ้าพิจารณาในมุมมองเดียวกันกับทฤษฎีบทเค้าร่างจะได้สมการเป็น

$$P(H, t+1) = P_{select}(H, t) \cdot P_{identify}(H, t) \cdot P_{compose}(H, t) \quad (7.9)$$

เมื่อ $P(H, t+1)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากรในรุ่นที่ $t+1$
 $P_{select}(H, t)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากร ที่ถูกเลือกจากรุ่นที่ t
 $P_{identify}(H)$ คือ โอกาสที่จะเกิดขึ้นเค้าร่าง H จากวิธีการระบุชิ้นส่วนจากรุ่นที่ t
 $P_{compose}(H)$ คือ โอกาสที่เค้าร่าง H จะอยู่รอดจากวิธีการประกอบชิ้นส่วนจากรุ่นที่ t

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

7.2 การทดลองและผลการทดลอง

ในส่วนนี้จะกล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่งในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่นำเสนอ และวิจารณ์ผลการทดลองในมุมมองที่นำเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม (sGA)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การไขว้เปลี่ยน:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่นำเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (RFT)
การเลือกสรร:	วิธีการเลือกแบบ n ตัวที่ดีที่สุด (n -Best) ค่า n คือค่า #sub ที่แสดงในตารางที่ 7.1
การระบุ:	วิธีสุ่มตัดแบบชิ้นส่วน
การประกอบ:	วิธีชักตัวอย่างจากหน่วยเก็บถาวรตามลำดับตำแหน่งปรากฏ
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 7.1

จากผลการทดลองในตารางที่ 7.1 เป็นการเปรียบเทียบกับ sGA ซึ่งมีหลักการดำเนินการต่างจากวิธี RFT เป็นอย่างมาก ถ้าพิจารณาในมุมมองของการระบุและการประกอบหน่วยการสร้างวิธี sGA ใช้การระบุแบบสุ่มเหมือนกับวิธี RFT แต่ในการประกอบเข้าด้วยกันวิธี RFT ใช้หลักการแบ่งปันความรู้แบบรวมศูนย์ที่มีการเก็บชิ้นส่วนที่ได้ไว้ร่วมกันในหน่วยเก็บถาวรต่างจากวิธี sGA ที่มีการแบ่งปันความรู้แบบกระจายเพราะมีการแลกเปลี่ยนข้อมูลเฉพาะโครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น เมื่อพิจารณาในแง่ของประสิทธิภาพวิธี RFT มีค่า FEs มากกว่า sGA ทั้งสองปัญหา ซึ่งเป็นตัวบ่งชี้ถึงประสิทธิภาพในการแก้ปัญหาของวิธี RFT ว่าด้อยกว่าวิธี sGA อย่างชัดเจน (ในสองปัญหาทดสอบนี้) แต่สาระสำคัญของขั้นตอนวิธีที่นำเสนอในบท

นี้ (RFT) คือ เพื่อต้องการทดสอบการใช้ความรู้แบบรวมศูนย์ในการประกอบชิ้นส่วนเข้าด้วยกัน ดังนั้นจะเห็นว่าในตารางที่ 7.1 มีค่าตัวแปร #sub เพิ่มขึ้นมาซึ่งหมายถึง จำนวนประชากรย่อย (Sub-population: sub) ที่ถูกเลือกเพื่อทำการระบุชิ้นส่วนดังที่ได้อธิบายไปแล้วในตอนต้น เมื่อพิจารณาถึงจำนวนประชากรที่ใช้ในการแก้ปัญหาสำหรับวิธี RFT ในปัญหา Royal Road มีขนาดประชากรเล็กกว่าวิธี sGA ในทางตรงกันข้ามสำหรับปัญหา Trap-5 วิธี RFT ต้องการจำนวนประชากรมากกว่าวิธี sGA อย่างชัดเจนในการแก้ปัญหา และเมื่อพิจารณาถึงค่าขนาดประชากรย่อยที่ใช้แก้ปัญหาจะเห็นว่ามีค่าแปรผันตามขนาดประชากรที่ใช้ สำหรับปัญหา Royal Road วิธี RFT ใช้จำนวนประชากรย่อยเท่ากับ 200 ซึ่งคิดเป็น 20% ของขนาดประชากร และสำหรับปัญหา Trap-5 วิธี RFT ใช้จำนวนประชากรย่อยเท่ากับ 800 ซึ่งคิดเป็น 10% ของขนาดประชากร ค่าของประชากรและประชากรย่อยเป็นตัวบ่งบอกถึงระดับความหลากหลายที่วิธี RFT ต้องการเพื่อใช้ในการแก้ปัญหา โดยทั่วไปปัญหา Royal Road เป็นปัญหาที่เราถือว่ามีความยากน้อยกว่าปัญหา Trap-5 เพราะมีความต้องการระดับความหลากหลายต่ำกว่า (Trap-5) เนื่องจากใช้จำนวนประชากรและประชากรย่อยน้อยกว่าในการแก้ปัญหา และเมื่อพิจารณาในกรณีที่ไม่สามารถหาคำตอบพบจะเห็นว่าในปัญหา Royal Road วิธี RFT ใช้จำนวนประชากรเท่ากับ sGA และในปัญหา Trap-5 วิธี RFT ใช้จำนวนประชากรมากกว่า sGA สอดคล้องกับกรณีที่หาคำตอบพบ

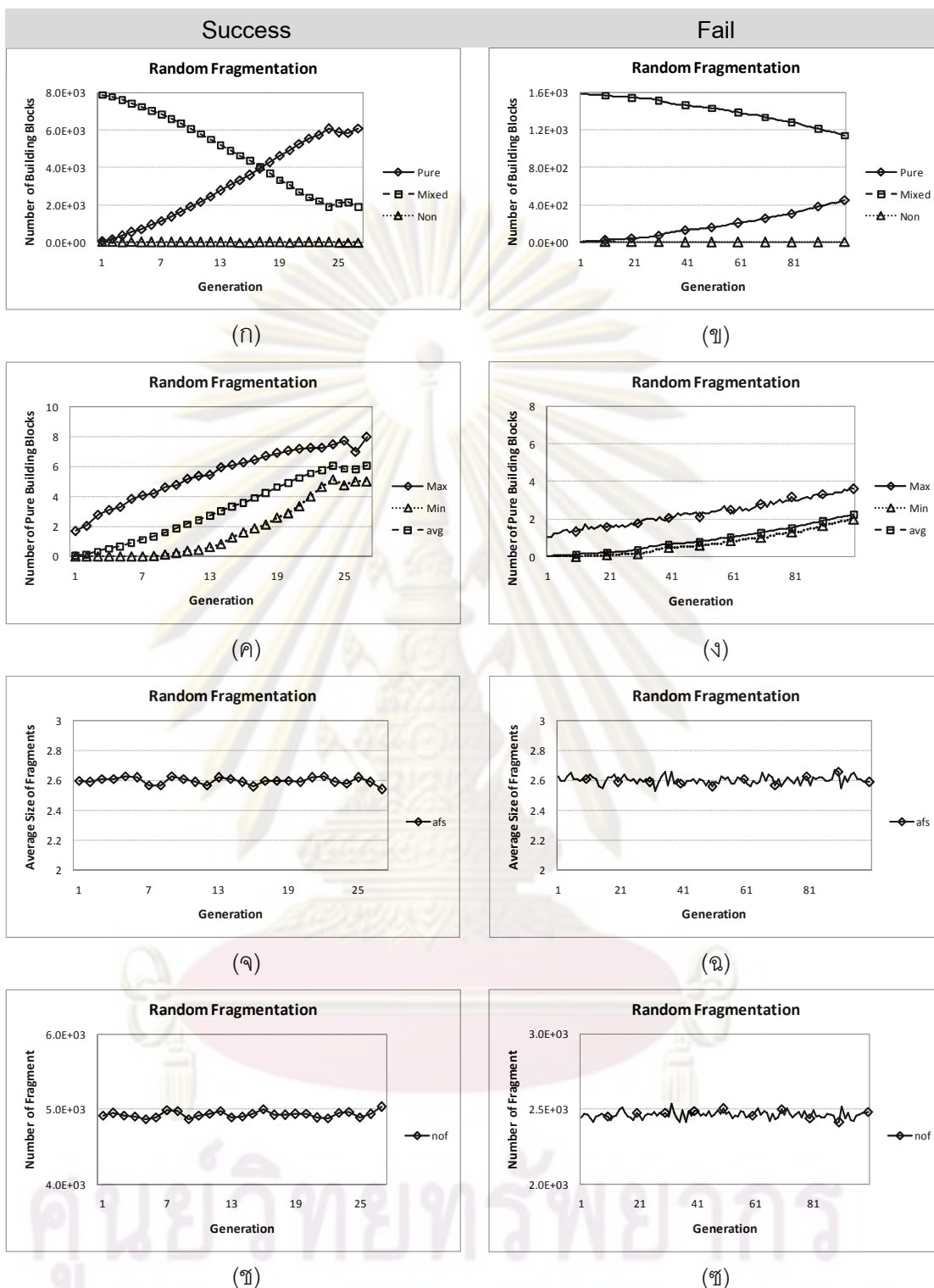
ตารางที่ 7.1 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ RFT

Parameters		Success 30 run				Fail 30 run	
		sGA		RFT		sGA	RFT
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
				#sub			#sub
RoyalRoad	64	1,200	11,920	1,000	51,734	200	200
				200			100
Trap-5	60	2,300	28,367	8,000	134,934	300	1,200
				800			200

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

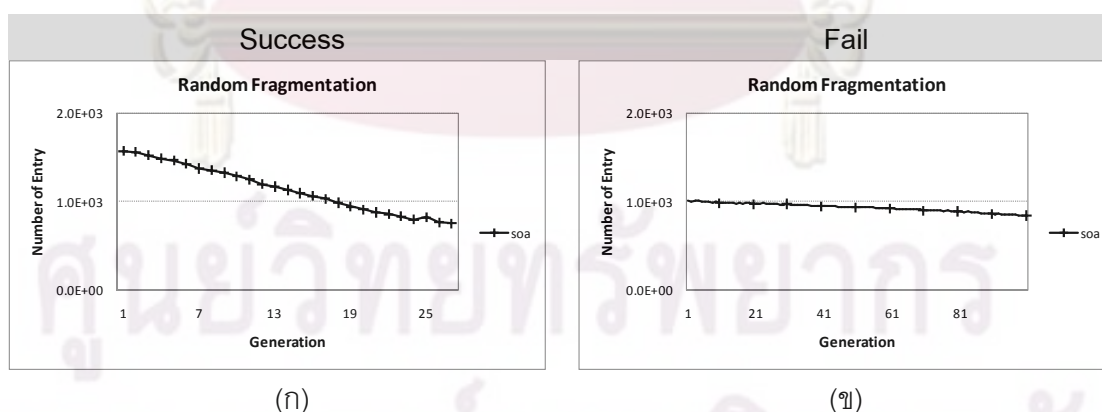
เมื่อพิจารณาผลการทดลองในมุมมองของหน่วยการสร้างสำหรับปัญหา Royal Road ดังรูปที่ 7.6 กรณีที่สามารถหาคำตอบพบในรูป 7.6 (ก) เกิดการแข่งขันกันระหว่าง Pure BBs และ Mixed BBs ในรูปแบบเชิงเส้น และในช่วงปลายจะเห็นว่ามีความใกล้เคียงที่และมีการแกว่งตัวเล็กน้อย สันนิษฐานว่าเกิดจากผลเฉลี่ยที่ได้จากการทดสอบหลายครั้ง (30 ครั้ง) เนื่องจากการทดสอบบางรอบอาจใช้จำนวนรุ่นประชากรมากเป็นพิเศษเมื่อเทียบกับรอบอื่น ๆ ดังนั้นเมื่อนำมาเฉลี่ยกันอาจได้ค่าที่มีความเบี่ยงเบนสูง ในปัญหา Royal Road เห็นชัดว่า Non BBs ไม่มีส่วนในการแข่งขันเลยเนื่องจากปัญหานี้ไม่มีการลวงในหน่วยการสร้าง และเมื่อพิจารณาต่อไปถึงจำนวน Pure BBs ที่มีอยู่ในหนึ่งโครโมโซมดังรูปที่ 7.6 (ค) มีค่าเพิ่มขึ้นโดยเฉลี่ยเป็นแบบเชิงเส้น และถ้าพิจารณาถึงค่า Max Pure BBs และ Min Pure BBs จะเห็นว่ามีความเป็นเส้นโค้งทำให้มีความเบี่ยงเบนมาตรฐานน้อยในช่วงต้นและช่วงปลาย แต่มีค่ามากในช่วงกลาง ๆ (ป่องออกเมื่อเทียบกับเส้น avg. Pure BBs) ซึ่งเป็นตัวบ่งชี้ว่าเกิดความหลากหลายของหน่วยการสร้างมากที่สุดในช่วงกลางของรุ่นประชากรที่ใช้ทั้งหมด สามารถอธิบายได้ว่าเกิดจากในช่วงรุ่นประชากรแรก ๆ ยังไม่สามารถหาหน่วยการสร้างได้ถูกต้องจึงมีจำนวน Pure BBs อยู่ต่ำ และเมื่อเริ่มเกิดการลู่เข้าจากรูปที่ 7.6 (ก) ปริมาณ Mixed BBs เริ่มลดลงและปริมาณ Pure BBs เริ่มมีมากขึ้น เป็นการแสดงให้เห็นว่าเริ่มมีการประกอบ Pure BBs ในหลากหลายรูปแบบเพิ่มขึ้น และเมื่อเวลาผ่านไปจนกระทั่งปริมาณ Pure BBs เพิ่มขึ้นมาเท่ากับ Mixed BBs ในรุ่นประชากรที่ 17 (จุดตัดในรูปที่ 7.6 (ก)) ความหลากหลายจะเริ่มลดลงเรื่อย ๆ เนื่องจากลู่เข้าสู่คำตอบมากขึ้น เหตุที่เห็นเป็นช่วงป่องได้ชัดเจนเป็นเพราะวิธี RFT มีระดับความหลากหลายสูง แต่มีความสามารถในการจัดการหน่วยการสร้างต่ำทำให้ Min Pure BBs เพิ่มขึ้นอย่างช้า ๆ ส่วนกรณีที่ไม่สามารถหาคำตอบพบจากรูปที่ 7.6 (ข) การแข่งขันกันระหว่าง Mixed BBs และ Pure BBs เกิดขึ้นอย่างช้า ๆ และจากรูปที่ 7.6 (ง) จำนวน Pure BBs ค่อย ๆ เพิ่มขึ้นอย่างช้า ๆ เช่นกัน จากข้อมูลทั้งสองภาพแสดงให้เห็นว่าวิธี RFT มีระดับความหลากหลายสูงทำให้เกิดการลู่เข้าช้าลงมากเมื่อเทียบกับวิธี sGA-FC ในบทที่ 5 และเป็นเหตุให้ใช้จำนวน FEs มากในการหาคำตอบ

สำหรับมุมมองของชิ้นส่วนดังรูปที่ 7.6 (จ-ช) ทุกกราฟมีลักษณะเหมือนกันหมดคือ มีค่าเกือบคงที่เท่ากันตลอดทุกรุ่นประชากร เป็นเพราะวิธีการระบุชิ้นส่วนของ RFT ใช้วิธีการสุ่มตัดดังที่อธิบายไปแล้วในตอนต้น (ไม่มีความลำเอียงกับรูปแบบหรือขนาดของชิ้นส่วน) ดังนั้นขนาดโดยเฉลี่ยของชิ้นส่วนที่ได้จะมีขนาดใกล้เคียงกันในทุกรุ่นประชากร และทำให้ได้ปริมาณชิ้นส่วนใกล้เคียงกันในทุกรุ่นประชากรด้วย กรณีหาคำตอบพบและหาคำตอบไม่พบต่างกันที่ปริมาณชิ้นส่วนที่ได้ รูปที่ 7.6 (ช-ช) เนื่องจากทั้งสองกรณีใช้จำนวนประชากรไม่เท่ากันจึงมีปริมาณชิ้นส่วน



รูปที่ 7.6 ผลการทดสอบของ RFT ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับค่าตอบที่ดีที่สุดกับรุ่นของประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

ต่างกัน การสุ่มตัดแบบสุ่มนี้เป็นตัวช่วยรักษาระดับความหลากหลายของชิ้นส่วนให้มีค่าสูงอยู่ตลอดเวลา แต่ก็ก็จะค่อย ๆ ลดลงเนื่องจากผลของการเลือกเชิงอันดับที่เลือกเฉพาะตัวที่มีค่าความเหมาะสมสูงเท่านั้น ถ้าพิจารณาเพียงแค่การระบุชิ้นส่วนจะเห็นว่าไม่มีการใช้หลักการหรือความรู้ทางด้านอื่นเข้ามาช่วย (เป็นแค่การสุ่มแบบเอกรูป) ดังนั้นชิ้นส่วนที่ระบุได้จึงไม่มีแบบแผนหรือโครงสร้างที่แน่ชัด ทำให้ความสัมพันธ์ของโครงสร้างคู่จิ้นในชิ้นส่วนมีโอกาสขัดแย้งหรือไม่ขึ้นแก่กันได้สูง (ไม่สอดคล้องกัน) ซึ่งจะส่งผลให้การระบุชิ้นส่วนมีประสิทธิภาพต่ำ อย่างไรก็ตามในบทนี้มุ่งเน้นไปที่การทดสอบการแบ่งปันความรู้แบบรวมศูนย์ในกระบวนการประกอบชิ้นส่วนเข้าด้วยกัน ดังนั้นจึงมีการวัดปริมาณชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรในแต่ละรุ่นประชากรแสดงดังรูปที่ 7.7 ปริมาณชิ้นส่วนที่ระบุได้ลดลงเรื่อย ๆ แบบเชิงเส้น ซึ่งเป็นผลจากการลู่เข้า อาจมีข้อสงสัยว่าทำไมปริมาณชิ้นส่วนในหน่วยเก็บถาวรลดลง แต่ปริมาณชิ้นส่วนที่ระบุได้จากรูปที่ 7.6 (ข-ช) มีค่าเกือบคงที่ ที่เป็นเช่นนี้เนื่องจากเมื่อเริ่มเกิดการลู่เข้าชิ้นส่วนจะเริ่มมีความเหมือนกันมากขึ้น (ซ้ำกัน) และหน่วยเก็บถาวรจะเก็บเฉพาะชิ้นส่วนที่มีรูปแบบไม่ซ้ำกัน ณ ตำแหน่งเดียวกันในโครโมโซมนั้นเอง แต่การวัดปริมาณชิ้นส่วนที่ระบุได้จากรูปที่ 7.6 (ข-ช) จะวัดทั้งชิ้นส่วนที่ซ้ำกันและไม่ซ้ำกัน (นับหมด) และเมื่อพิจารณาปริมาณชิ้นส่วนในหน่วยเก็บถาวรจากรูปที่ 7.7 ร่วมกับปริมาณหน่วยการสร้างในรูปที่ 7.6 (ก-ข) จะเห็นได้ว่าเป็นไปในทิศทางเดียวกัน ในกรณีที่หาค่าตอบพบปริมาณชิ้นส่วนในหน่วยเก็บถาวรลดลงเร็วกว่ากรณีที่หาค่าตอบไม่พบอย่างเห็นได้ชัด ระดับของการลดลงในกรณีที่หาค่าตอบพบมีผลทำให้เกิดการแข่งขันกันระหว่าง Pure BBs และ Mixed BBs อย่างเห็นได้ชัด (เกิดการลู่เข้าชัดเจน) ตรงกับรูปที่ 7.6 (ก) ส่วนกรณีที่หาค่าตอบไม่พบการลดลงเกิดขึ้นอย่างค่อยเป็นค่อยไปตรงกับรูปที่ 7.6 (ข)



รูปที่ 7.7 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ RFT ในปัญหา

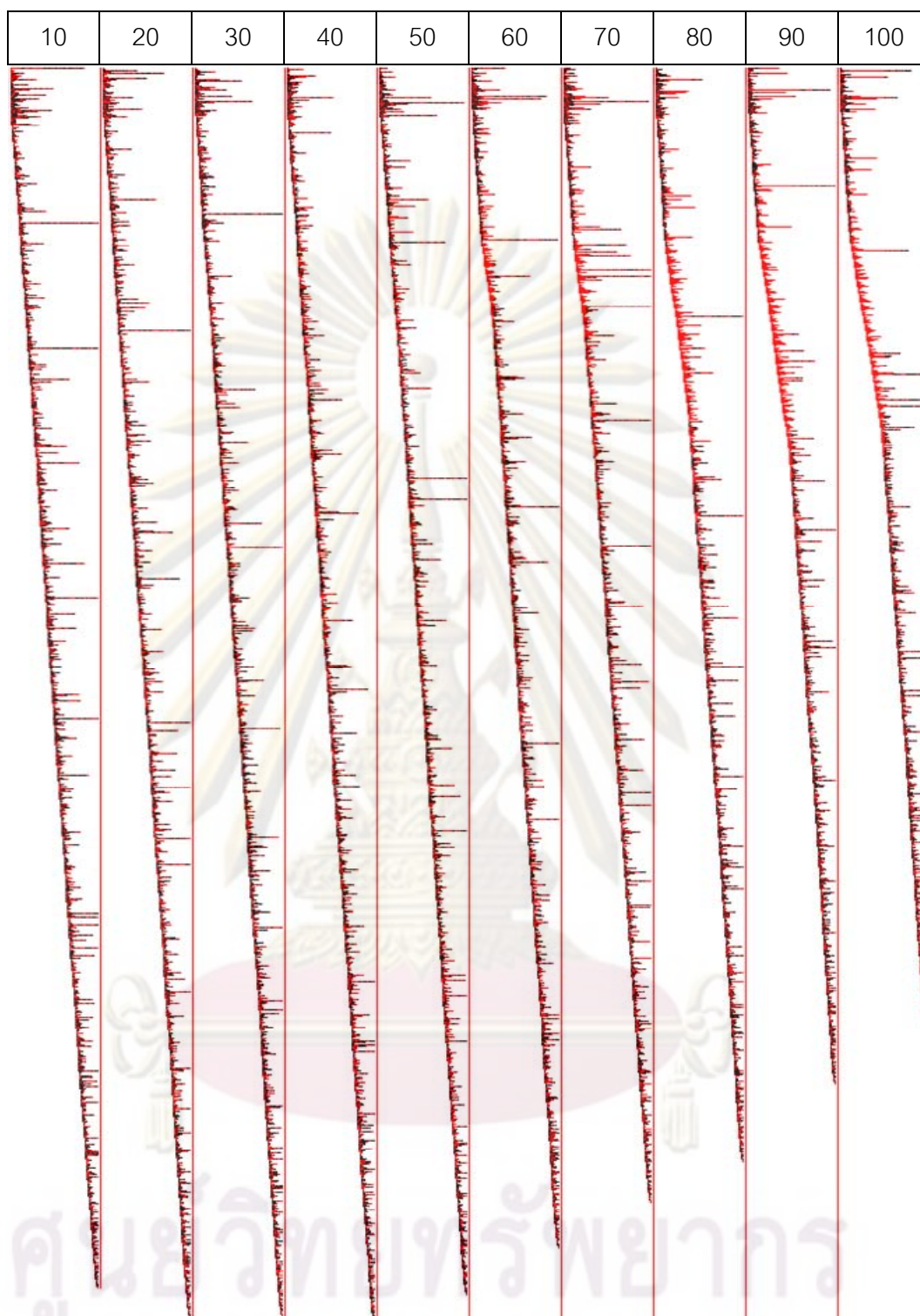
Royal Road 64 บิต

เมื่อพิจารณาในส่วนของการประกอบชิ้นส่วนเข้าด้วยกันจะเห็นว่าการเก็บข้อมูลชิ้นส่วนที่ระบุได้ทั้งหมดไว้ร่วมกันในหน่วยเก็บถาวรเดียวกัน เป็นการแบ่งปันข้อมูลแบบรวมศูนย์ซึ่งมีจุดเด่นที่ชัดเจนคือ ช่วยเพิ่มความหลากหลายให้กับรูปแบบของโครงสร้างที่เป็นไปได้ให้มีเพิ่มขึ้น เมื่อเปรียบเทียบกับวิธีการประกอบเข้าด้วยกันแบบกระจายที่มีการแลกเปลี่ยนข้อมูลกันเฉพาะโครงสร้างของโครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น ไม่มีการแลกเปลี่ยนข้ามไปยังโครโมโซมตัวอื่น เพื่อให้เห็นภาพได้ชัดเจนขึ้นว่าการแบ่งปันความรู้แบบรวมศูนย์โดยใช้หน่วยเก็บถาวรมีรูปแบบเป็นอย่างไร เราจะแสดงโครงสร้างทางพันธุกรรมของชิ้นส่วนที่ระบุได้และเก็บอยู่ในหน่วยเก็บถาวรดังรูปที่ 7.8 และรูปที่ 7.9 (วิธีการอ่านรูปอยู่ในบทที่ 4) จากรูปจำนวนชิ้นส่วนลดลงเรื่อย ๆ ในแต่ละรุ่นประชากรตรงกับรูปที่ 7.7 (ก-ข) และจำนวนชิ้นส่วนที่เริ่มต้นที่ตำแหน่งแรก (บิตแรก) มีปริมาณมากที่สุดเมื่อเทียบกับตำแหน่งอื่นเป็นเพราะวิธีการระบุชิ้นส่วนจะได้ชิ้นส่วนที่เริ่มต้นที่ตำแหน่งแรกเสมอ (เราพิจารณาชิ้นส่วนจากซ้ายไปขวา) จากรูปขนาดของชิ้นส่วนโดยเฉลี่ยมีค่าใกล้เคียงกันในทุกรุ่นประชากรสอดคล้องกับรูปที่ 7.6 (จ-ฉ) ความแตกต่างระหว่างรูปที่ 7.8 (กรณีหาค่าตอบพบ) และรูปที่ 7.9 (กรณีหาค่าตอบไม่พบ) อยู่ตรงที่รูปแบบของชิ้นส่วนซึ่งสังเกตเห็นได้ว่าในกรณีที่หาค่าตอบพบชิ้นส่วนมีส่วนประกอบของคู่จีนที่มีค่าเป็น 1 อยู่เป็นจำนวนมาก โดยจะเพิ่มขึ้นเรื่อย ๆ ในแต่ละรุ่นจากตำแหน่งแรก ๆ ไล่ไปเรื่อย ๆ ต่างจากกรณีที่หาค่าตอบไม่พบที่ยังคงเห็นคู่จีนที่มีค่าเป็น 0 ปะปนอยู่อย่างชัดเจน และอัตราการลดลงของปริมาณชิ้นส่วนในหน่วยเก็บถาวรสำหรับกรณีที่หาค่าตอบพบจะลดลงเร็วกว่ากรณีที่หาค่าตอบไม่พบ ยิ่งไปกว่านั้นภาพแสดงชิ้นส่วนในหน่วยเก็บถาวรยังสามารถสะท้อนให้เห็นถึงภาพรวมของประชากรในรุ่นถัดไป เนื่องจากประชากรในรุ่นถัดไปถูกสร้างขึ้นจากชิ้นส่วนที่มีในหน่วยเก็บถาวรเท่านั้น ลองพิจารณาจากรูปที่ 7.8 และรูปที่ 7.10 ในรุ่นประชากรที่ 20 จะเห็นได้ชัดว่าในหน่วยเก็บถาวรจะมีคู่จีนที่เป็น 0 ติดอยู่ในช่วงคู่จีนท้าย ๆ 2 ช่วง ซึ่งเห็นได้ชัดเจนขึ้นในภาพของประชากรที่มีแถบสีดำ (ในแนวตั้ง) ในช่วงคู่จีนท้าย ๆ มีตำแหน่งตรงกันกับวีวสีดำในภาพหน่วยเก็บถาวรพอดี เป็นต้น

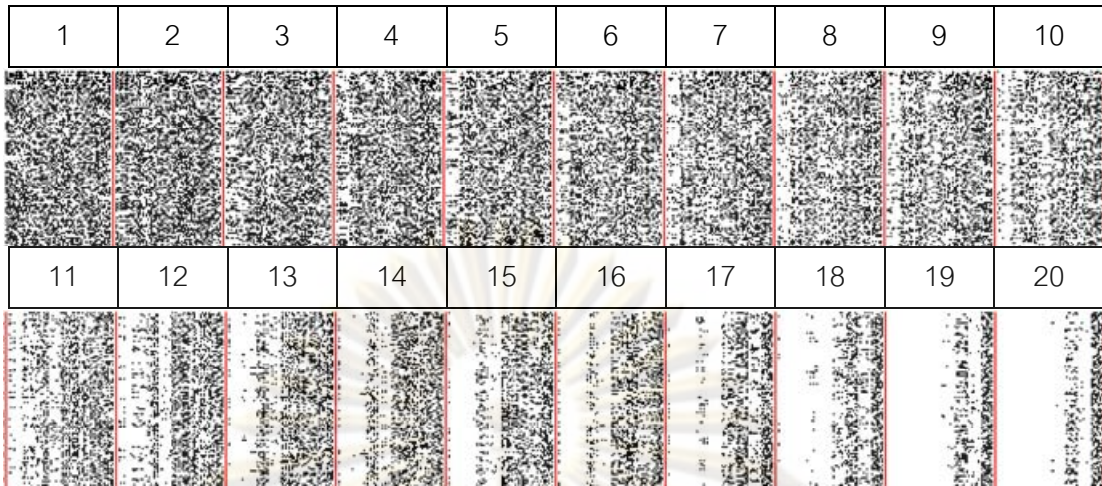
ในมุมมองของรูปแบบโครงสร้างคู่จีนของแต่ละโครโมโซมในภาพรวมในกรณีที่หาค่าตอบพบ (วิธีการอ่านรูปอยู่ในบทที่ 4) จากรูปที่ 7.10 จะเห็นว่าเกิดการลู่เข้าช้า (ความแตกต่างที่เกิดขึ้นจากรุ่นประชากรที่ 1 จนถึงรุ่นประชากรที่ 10 มีน้อย) และจะเริ่มลู่เข้าอย่างชัดเจนในรุ่นประชากรที่ 13 และในกรณีที่หาค่าตอบไม่พบจากรูปที่ 7.11 การเปลี่ยนแปลงเกิดขึ้นช้ามาก และเริ่มมีการลู่เข้าอย่างชัดเจนในรุ่นประชากรที่ 60 จากทั้งสองกรณีจะสังเกตเห็นว่าไม่มีแถบสีดำปรากฏให้เห็นชัดเจนเหมือนขั้นตอนวิธีอื่น (เช่น sGA และ sGA-FC ในบทที่ 5 เป็นต้น) ในปัญหาเดียวกัน สันนิษฐานว่าเกิดจากการระบุชิ้นส่วนแบบสุ่มทำให้ไม่มีรูปแบบของโครงสร้างที่มีความสัมพันธ์กันชัดเจน และวิธีการประกอบเข้าด้วยกันทำให้มีความหลากหลายสูง



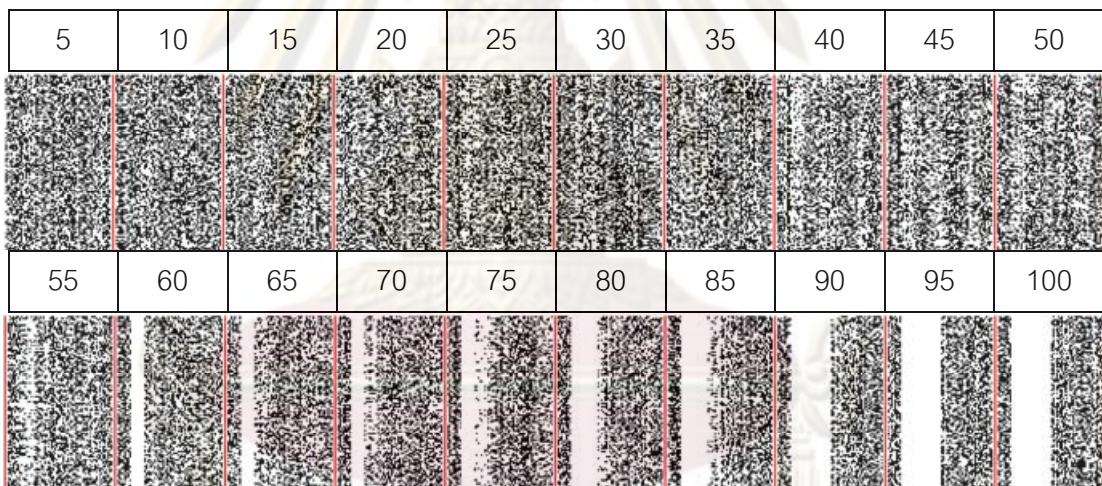
รูปที่ 7.8 รูปแบบจิ้นไทป์ของชั้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 20 รุ่น แต่เลือกมาแสดงผล 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นทึ่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 7.91 รูปแบบจิ้นไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



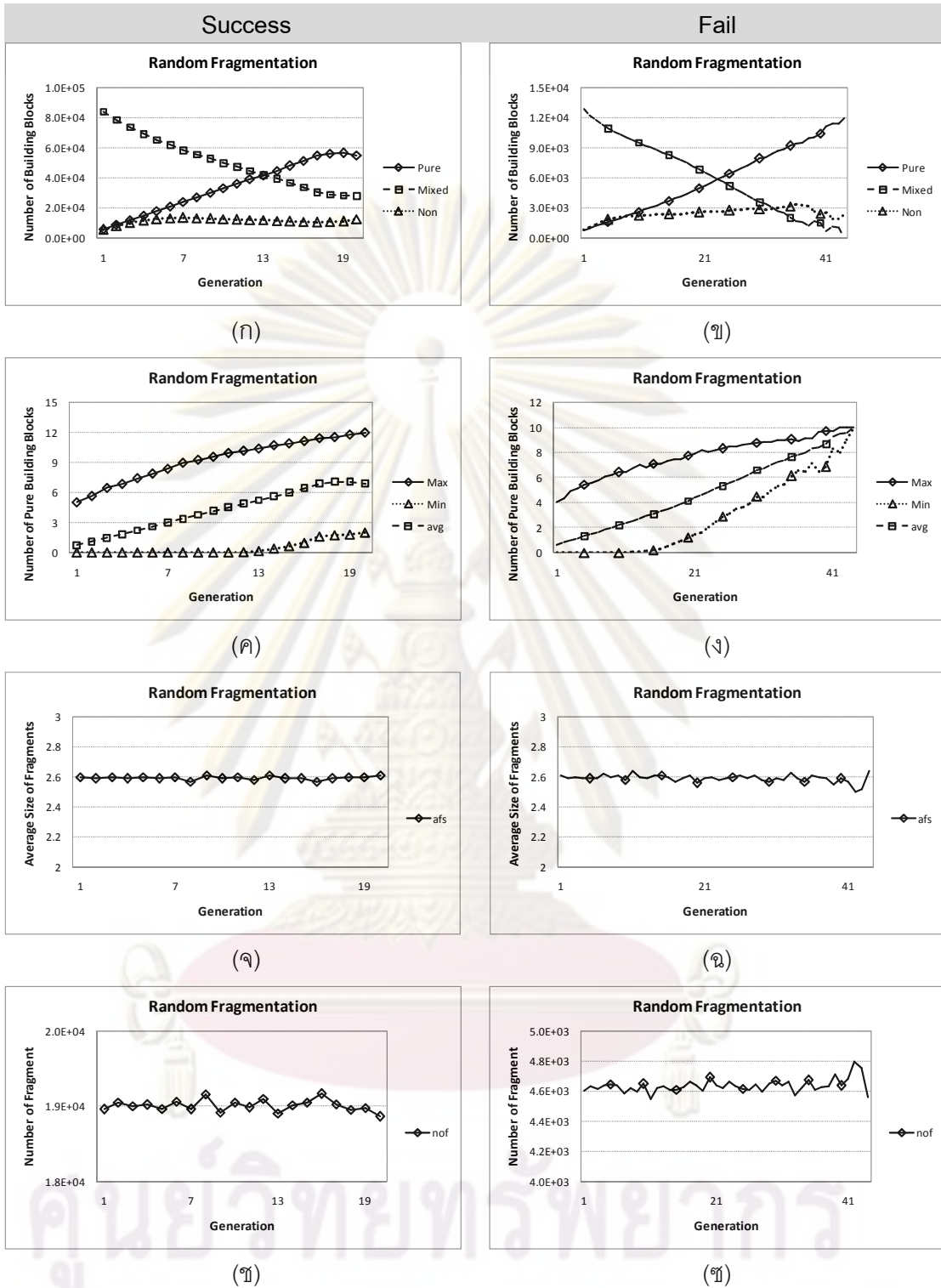
รูปที่ 7.10 รูปแบบจีโนไทป์ของประชากรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคูจีนที่มีค่าเป็น 1 และพื้นที่สีดำ คือคูจีนที่มีค่าเป็น 0



รูปที่ 7.11 รูปแบบจีโนไทป์ของประชากรของ RFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคูจีนที่มีค่าเป็น 1 และพื้นที่สีดำ คือคูจีนที่มีค่าเป็น 0

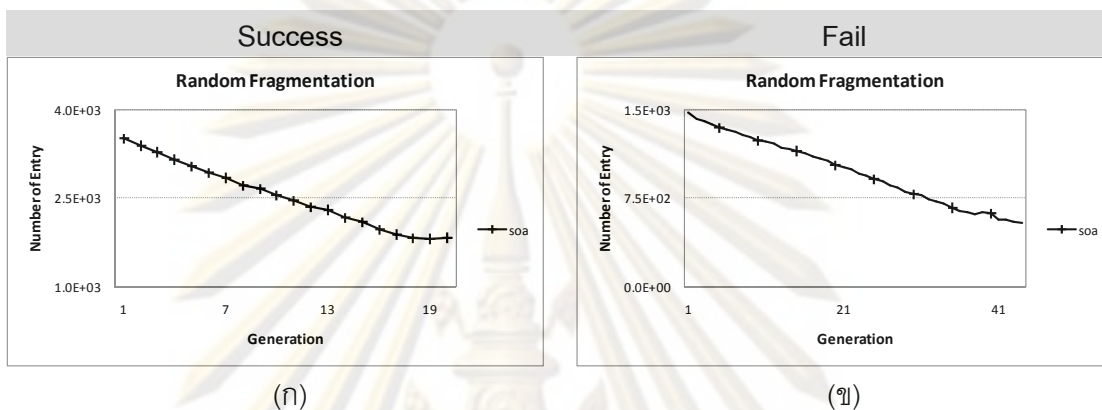
สำหรับปัญหา Trap-5 ซึ่งมีการลวงในหน่วยการสร้างผลการทดสอบจากรูปที่ 7.12 (ก-ง) มีความแตกต่างจากปัญหา Royal Road ในรูปที่ 7.6 อย่างชัดเจน เริ่มจากรูปที่ 7.12 (ก) จะเห็นว่านอกจากจะมีการแข่งขันระหว่าง Mixed BBs และ Pure BBs แล้ว Non BBs ยังมีการเพิ่มขึ้นด้วย (แต่ไม่มาก) เหตุที่เป็นเช่นนี้เพราะวิธี RFT ไม่มีความลำเอียงในการระบุและการประกอบหน่วยการสร้าง ทำให้ไม่ถูกลวงจากโครงสร้างของหน่วยการสร้างลวง การที่ Non BBs เพิ่มขึ้นระดับหนึ่งเป็นเพราะการเลือกสรรที่จะเลือกแต่ตัวที่มีค่าความเหมาะสมสูง ซึ่งอาจประกอบไปด้วยหน่วยการสร้างแบบลวง (โครงสร้างของหน่วยการสร้างลวงโดยเฉลี่ยมีค่าความเหมาะสมสูงกว่าโครงสร้างของหน่วยการสร้างที่ถูกต้อง) ในรูปที่ 7.12 (ข) กรณีที่หาคำตอบไม่พบ จะเห็นการแข่งขันกันระหว่าง Mixed BBs กับ Pure BBs อย่างชัดเจน และในช่วงรุ่นประชากรปลาย ๆ ปริมาณของ Non BBs จะสูงกว่า Mixed BBs ด้วย ต่างจากรูปที่ 7.6 (ข) (ที่เกิดการลู่เข้า) เกิดเนื่องจากระดับความกดดันในการเลือกของปัญหา Trap-5 มีสูงกว่า Royal Road มาก เมื่อคิดเป็นสัดส่วนของประชากรย่อยที่ถูกเลือกเทียบกับประชากรทั้งหมด ในปัญหา Trap-5 มีขนาดประชากรเท่ากับ 1200 ตัว เลือกมาระบุหน่วยการสร้าง 200 ตัว คิดเป็น 1/6 (ประมาณ 16.17%) ส่วนปัญหา Royal Road มีขนาดประชากรเท่ากับ 200 เลือกประชากรย่อย 100 ตัว คิดเป็น 1/2 (เท่ากับ 50%) ดังนั้นความหลากหลายในปัญหา Trap-5 จึงลดลงอย่างรวดเร็วและเกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 44 และเมื่อพิจารณาจำนวน Pure BBs เพียงอย่างเดียวจากรูปที่ 7.12 (ค) ในแต่ละรุ่นประชากรมีจำนวน Max Pure BBs และ Min Pure BBs ต่างกันมาก เป็นตัวบ่งชี้ว่ามีความหลากหลายของรูปแบบโครงสร้างคู่จิ้นสูง ต่างจากปัญหา Royal Road ในรูปที่ 7.6 (ค) อย่างชัดเจน ความหลากหลายนี้มาจากโครงสร้างคู่จิ้นในกลุ่มประชากรย่อย ที่ประกอบไปด้วยหน่วยการสร้างหลายรูปแบบ (ทั้งแบบลวงและไม่ลวง) บวกกับวิธีการประกอบชิ้นส่วนที่มีการใช้หน่วยเก็บถาวรร่วมกัน เมื่อพิจารณากรณีที่หาคำตอบไม่พบจากรูปที่ 7.12 (ง) เกิดการลู่เข้าอย่างชัดเจนในช่วงรุ่นประชากรท้าย ๆ เป็นผลจากระดับความกดดันในการเลือกสรรที่มีค่าสูง (ดังที่อธิบายแล้วก่อนหน้านี้)

ในมุมมองของชิ้นส่วนสำหรับปัญหา Trap-5 จากรูปที่ 7.12 (จ-ช) คล้ายกับผลที่ได้จากปัญหา Royal Road เนื่องจากวิธีการระบุและการประกอบชิ้นส่วนไม่มีการพิจารณา (ไม่สนใจ) ถึงรูปแบบโครงสร้างทางพันธุกรรม และในช่วงรุ่นประชากรท้าย ๆ (รุ่นที่ 41-44) มีการแกว่งตัวสูงเกิดจากจำนวนรอบของผลการทดสอบ (จากทั้งหมด 30 รอบ) ที่ใช้จำนวนรุ่นประชากรมากกว่า 40 รุ่น มีจำนวนน้อย ทำให้ค่าที่ได้ไม่ถูกหารเฉลี่ยกับรอบการทำงานอื่น



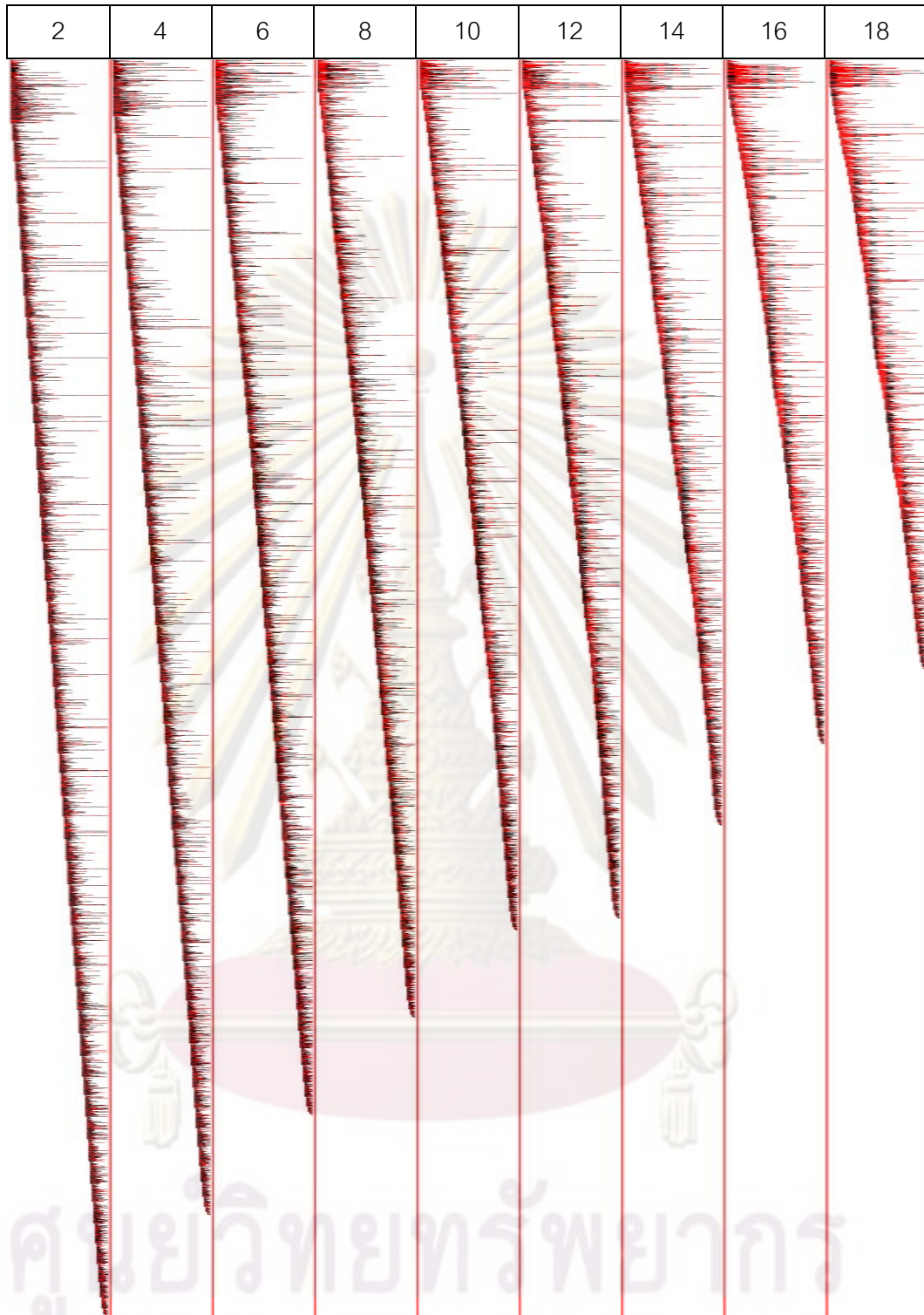
รูปที่ 7.12 ผลการทดสอบของ RFT ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

เมื่อพิจารณาถึงปริมาณชิ้นส่วนที่เกิดขึ้นอยู่ในหน่วยเก็บถาวรจากรูปที่ 7.13 (ก-ข) ปริมาณชิ้นส่วนค่อย ๆ ลดลงแบบเชิงเส้นตลอดการดำเนินการ เป็นเพราะความหลากหลายในวิธี RFT จะค่อย ๆ ลดลง เนื่องจากผลของการเลือกเชิงอันดับ และในรูปที่ 7.13 (ข) เมื่อเปรียบเทียบกับปัญหา Royal Road ในรูปที่ 7.7 (ข) จะเห็นว่ามียัตราการลดลงที่สูงกว่าอย่างชัดเจนซึ่งเป็นผลจากการลู่เข้าที่เกิดในระดับโครงสร้างย่อยของปัญหา Trap-5

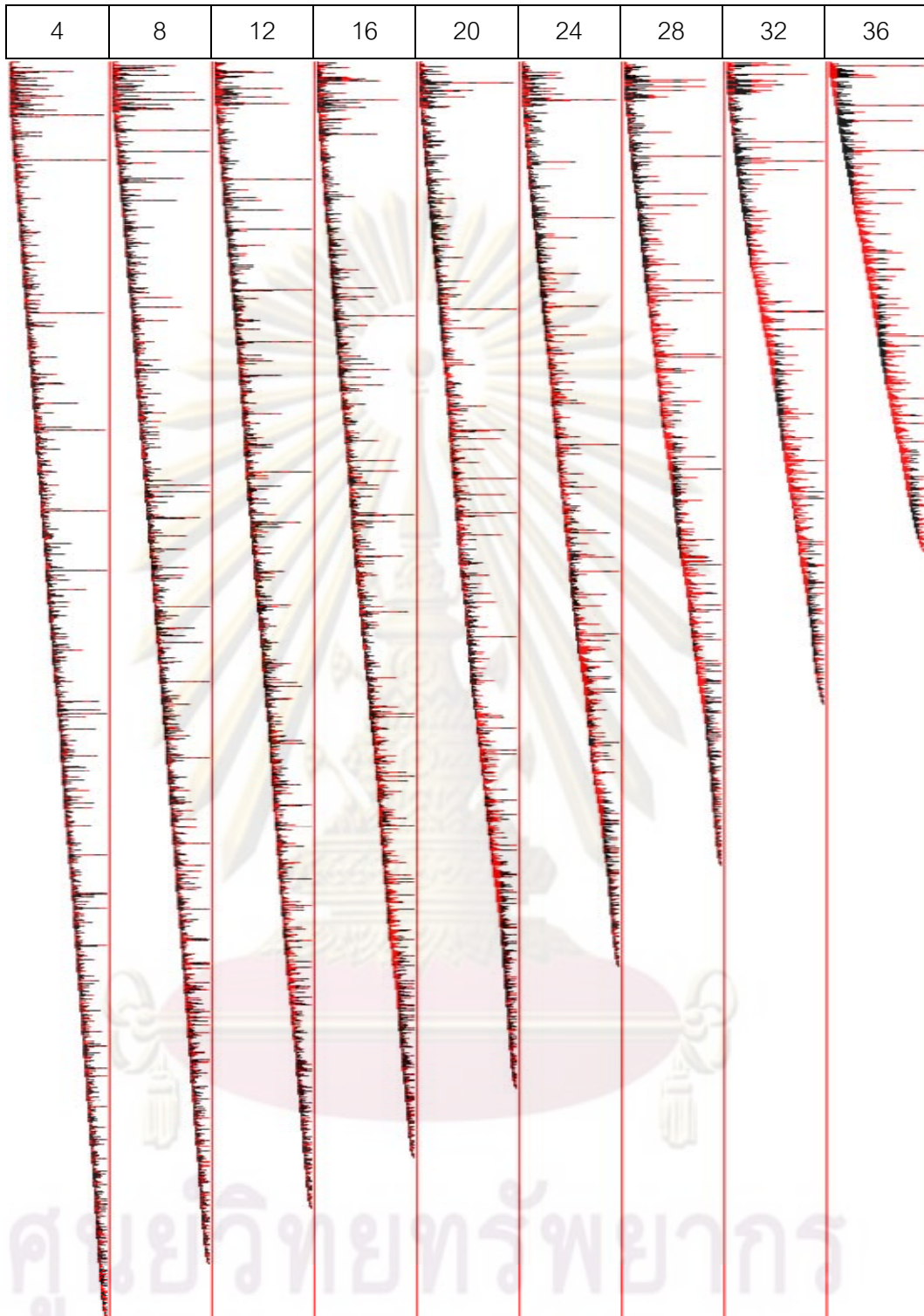


รูปที่ 7.13 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ RFT ในปัญหา Trap-5 60 บิต

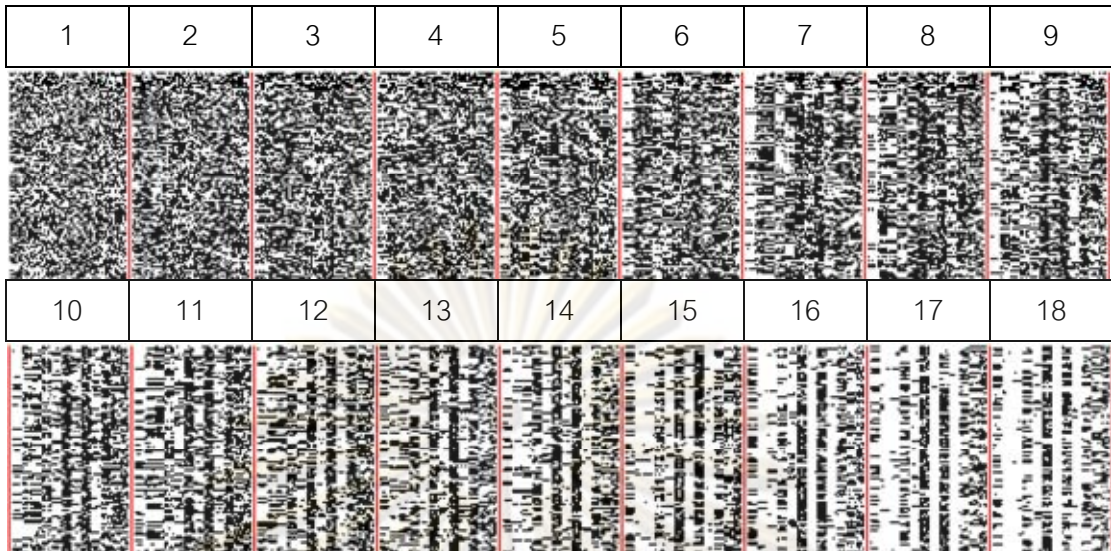
โครงสร้างทางพันธุกรรมของชิ้นส่วนในหน่วยเก็บถาวรจากรูปที่ 7.14 (กรณีที่เขา คำตอบพบ) เมื่อเทียบกับปัญหา Royal Road ในรูปที่ 7.8 จะสังเกตเห็นว่าชิ้นส่วนในปัญหา Trap-5 มีความยาวมากกว่าปัญหา Royal Road โดยเฉลี่ยในทุกช่วงตำแหน่ง ซึ่งขัดแย้งกับผลที่ได้จากรูปที่ 7.12 (ก) และรูปที่ 7.6 (ก) ที่วัดความยาวของชิ้นส่วนโดยเฉลี่ยในทุกรุ่นประชากรได้เกือบเท่ากันคือ 2.6 เป็นเพราะว่าในการวัดความยาวเฉลี่ยของชิ้นส่วนจะวัดจากทุกชิ้นส่วนที่ระบุได้ แต่ในหน่วยเก็บถาวรไม่ได้เก็บทุกชิ้นส่วนที่ระบุได้ จะเก็บเฉพาะชิ้นส่วนที่มีรูปแบบไม่ซ้ำกันในตำแหน่งเดียวของโครโมโซมเท่านั้น ชิ้นส่วนที่ยาวขึ้นนี้มาจากจำนวนประชากรย่อยที่เพิ่มขึ้นจากปัญหา Royal Road ที่ใช้เพียงแค่ 200 ตัว เพิ่มมาเป็น 800 ตัวในปัญหา Trap-5 ทำให้เกิดความหลากหลายในการสุ่มตัดในขั้นตอนการระบุชิ้นส่วนมากขึ้น หรือกล่าวได้ว่ามีจำนวนประชากรย่อยเพิ่มขึ้น ทำให้ต้องสุ่มเลือกจำนวนจุดตัดมากขึ้น (ต้องเลือกจำนวนจุดตัดให้กับประชากรย่อยแต่ละตัว) ส่งผลให้จำนวนจุดตัดมีการกระจายตัวดีขึ้น (มีค่าที่ไม่ซ้ำกันมากขึ้น) จากรูปที่ 7.14 และ 7.15 จะเห็นว่ามียัตราที่ชัดเจนเป็นช่วง ๆ ในแนวตั้งซึ่งหมายถึงตำแหน่ง ตรงกับภาพโครงสร้างทางพันธุกรรมของประชากรในรูปที่ 7.16 และรูปที่ 7.17 ตามลำดับ ในกรณีที่เขา คำตอบพบจากรูปที่ 7.14 ยัตราที่ปรากฏจะมีลักษณะสั้น และไม่ชัดเจนเท่ากับกรณีที่เขา คำตอบไม่พบ เนื่องจากยังไม่เกิดการลู่เข้าอย่างสมบูรณ์ และมีระดับความหลากหลายของชิ้นส่วนสูงกว่า



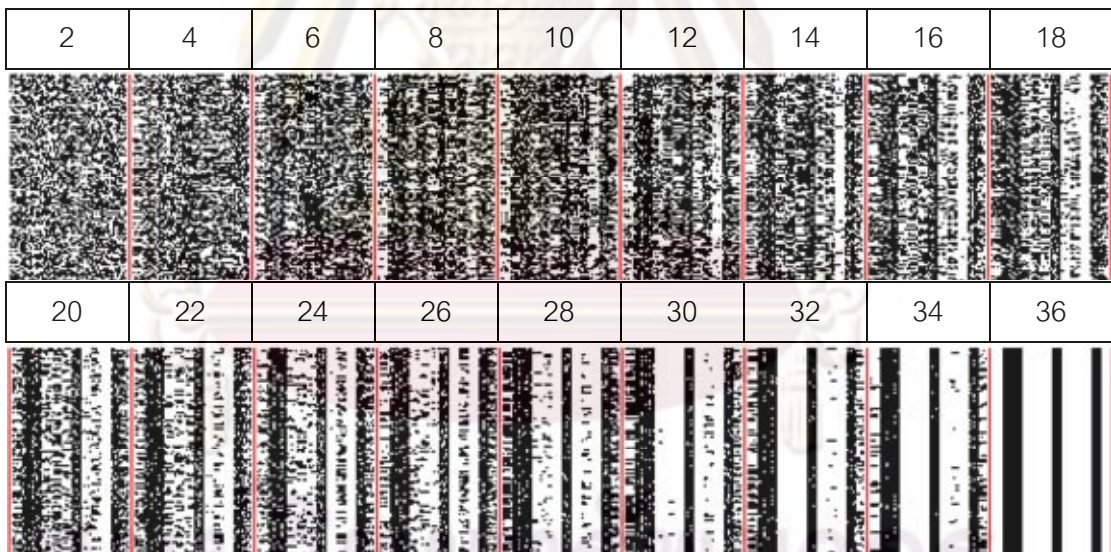
รูปที่ 7.14 รูปแบบจิ้นไทป์ของชั้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Trap-5 60 บิต
 กรณีสหาคำตอบพบ และใช้จำนวนรุ่นประชากร 18 รุ่น แต่เลือกมาแสดงผล 9 รุ่น ส่วนตัวเลขกำกับ
 ด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขต
 ข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่า
 เป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 7.15 รูปแบบจิ้นไทป์ของชั้นส่วนที่อยู่ในหน่วยเก็บถาวรของ RFT ในปัญหา Trap-5 60 บิต
 กรณีสืบค้นค่าตอบไม่พบ และใช้จำนวนรุ่นประชากร 36 รุ่น แต่เลือกมาแสดงผล 9 รุ่น ส่วนตัวเลข
 กำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอก
 ขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นี่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่
 จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 7.16 รูปแบบจีโอโทปของประชากรของ RFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 18 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จิ้นที่มีค่าเป็น 0



รูปที่ 7.17 รูปแบบจีโอโทปของประชากรของ RFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 36 รุ่น แต่เลือกมาแสดงผล 18 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จิ้นที่มีค่าเป็น 0

รูปที่ 7.16 และรูปที่ 7.17 แสดงโครงสร้างทางพันธุกรรมของประชากรในภาพรวม จะเห็นว่าลักษณะของชิ้นส่วนเป็นริ้วยาวตามลักษณะโครงสร้างของหน่วยการสร้าง (เห็นได้ชัดจากริ้วสีดำ) ต่างจากปัญหา Royal Road ในรูปที่ 7.10 และรูปที่ 7.11 อย่างชัดเจน เนื่องจากโครโมโซมตัวที่มีค่าความเหมาะสมสูงที่ถูกเลือกมีหน่วยการสร้างลงปะปนอยู่ด้วย และหน่วยการสร้างลงเหล่านี้ให้ค่าความเหมาะสมโดยเฉลี่ยสูงกว่าหน่วยการสร้างที่แท้จริง ทำให้หน่วยการสร้างลงปรากฏชัดเจน (ริ้วสีดำ) กรณีที่หาคำตอบพบในรูปที่ 7.16 จะเห็นว่าเริ่มแยกแยะหน่วยการสร้างที่ดีได้ตั้งแต่รุ่นประชากรที่ 6 และชัดเจนมากขึ้นในรุ่นประชากรที่ 10 จนกระทั่งรุ่นประชากรที่ 18 จึงหาคำตอบพบแต่ก็ยังมีหน่วยการสร้างลงปะปนอยู่มาก แต่กรณีที่หาคำตอบไม่พบในรูปที่ 7.17 การแยกแยะหน่วยการสร้างเป็นไปอย่างช้า ๆ เริ่มเห็นได้จากรุ่นประชากรที่ 8 จนเห็นได้ชัดเจนขึ้นในรุ่นประชากรที่ 20 และเกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 36 เนื่องจากระดับความหลากหลายที่เกิดขึ้นจากกลุ่มประชากรย่อยมีค่าน้อยกว่าระดับความกดดันในการเลือก หรือกล่าวได้ว่าเกิดจากการที่ใช้จำนวนประชากรและประชากรย่อยน้อยเกินไป ทำให้เกิดการลู่เข้าก่อนที่จะสามารถหาคำตอบพบ

7.3 สรุปผลการทดลอง

ในบทนี้นำเสนอขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม ซึ่งมีการใช้วิธีการระบุชิ้นส่วนแบบสุ่ม กับวิธีการประกอบเข้าด้วยกันโดยใช้หน่วยเก็บถาวรร่วมกันในการแบ่งปันความรู้แบบรวมศูนย์ ผลการทดสอบแม้ว่าจะดีด้อยกว่า sGA ในทุกปัญหา แต่ก็ยังสามารถหาคำตอบได้ภายในเงื่อนไขที่กำหนด (ในบทที่ 4) เนื่องจากวิธี RFT ไม่มีความลำเอียงกับรูปแบบโครงสร้างทางพันธุกรรมเลยทั้งในขั้นตอนการระบุและการประกอบเข้าด้วยกัน ตัวดำเนินการหลักที่ทำหน้าที่ผลักดันให้ไปถึงคำตอบคือ ตัวดำเนินการเลือกสรรเชิงอันดับ จึงทำให้มั่นใจได้ว่าหน่วยการสร้างที่ดีมีอยู่ในโครโมโซมที่ดี ๆ ที่มีค่าความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ย และการจะไปให้ถึงคำตอบขึ้นอยู่กับระดับความหลากหลายที่ได้จากจำนวนประชากรเป็นหลัก (ประชากรมากทำให้มีความหลากหลายมาก ประชากรน้อยทำให้มีความหลากหลายน้อย) ซึ่งต้องมีความสมดุลกับระดับความกดดันในการเลือกสรร ถ้าไม่สมดุลอาจทำให้เกิดการลู่เข้าก่อนที่จะหาคำตอบพบ หรือลู่เข้าช้าเกินกว่ารุ่นประชากรที่กำหนดไว้ เป็นต้น ถ้าพิจารณาในแง่ของการระบุหน่วยการสร้างจะเห็นว่า sGA เป็นวิธีที่ไม่มีความลำเอียงกับตำแหน่งในการไขว้เปลี่ยนเหมือนกับวิธี RFT ที่ไม่มีความลำเอียงกับรูปแบบหน่วยการสร้าง แต่ถ้ามองในแง่ของการประกอบหน่วยการสร้างจะเห็นว่า sGA เป็นวิธีการแบ่งปันความรู้แบบกระจาย ต่างจากวิธี RFT ที่มีกรแบ่งปันความรู้แบบรวมศูนย์ อย่างไรก็ตาม ผลการทดสอบได้แสดงให้เห็นว่า การแบ่งปันความรู้ทั้งสองแบบช่วยให้หาคำตอบได้

เหมือนกัน ต่างกันตรงที่ประสิทธิภาพที่ได้รับ สันนิษฐานว่าสาเหตุที่ทำให้วิธี RFT ให้ผลแยกว่า sGA เป็นเพราะการแบ่งปันความรู้แบบรวมศูนย์ในวิธี RFT ทำให้เกิดความหลากหลายสูงเกินไป เมื่อเทียบกับ sGA ส่งผลให้เกิดการลู่เข้าซ้ำ (ขึ้นกับระดับความกดดันในการเลือก จำนวนประชากร และจำนวนประชากรย่อย) ประกอบกับหน่วยการสร้างที่ได้จากวิธีการระบุแบบสุ่มไม่มีคุณภาพเพียงพอ ทำให้มีโอกาสน้อยที่จะประกอบเข้าด้วยกันแล้วได้คำตอบที่มีคุณภาพดีขึ้น ด้วยเหตุนี้จึงได้ขอสันนิษฐานเพิ่มเติมว่าการระบุหน่วยการสร้างที่มีประสิทธิภาพจะมีผลทำให้ได้คำตอบที่ดีขึ้น (มีการศึกษาข้อสันนิษฐานนี้ในบทถัดไป)



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 8

ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข

(Conditional Fragmentation Algorithm: CFT)

ในบทนี้จะนำเสนอวิธีการระบุหน่วยการสร้างและการประกอบเข้าด้วยกัน โดยอาศัยพื้นฐานการดำเนินการแบบเดียวกับขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (บทที่ 7) และเพิ่มเติมหลักการในการระบุชิ้นส่วนโดยใช้ความรู้ทางด้านสารสนเทศร่วมด้วย (ใช้วิธีการระบุชิ้นส่วนแบบเดียวกับการไขว้เปลี่ยนแบบชิ้นส่วนในบทที่ 5) เพื่อทดสอบสมมติฐานที่ว่าถ้าทำการระบุชิ้นส่วนได้ดีขึ้นจะทำให้ได้ผลที่ดีขึ้นด้วย เปรียบเทียบกับขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (RFT)

8.1 แนวคิดและหลักการ

จากขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่มในบทที่ 7 จะเห็นว่าได้ผลลัพธ์ที่แยกกว่า sGA เป็นอย่างมาก แต่ก็ยังแสดงให้เห็นว่าในโครโมโซมที่ดี ๆ (มีค่าความเหมาะสมสูง) ที่ถูกเลือกมีส่วนประกอบที่ดีที่จะนำไปสู่คำตอบอยู่จริง เนื่องจากยังสามารถหาคำตอบได้ตามเงื่อนไขที่กำหนด (ในบทที่ 4) ทั้งนี้เป็นเพราะวิธีการระบุชิ้นส่วนที่ใช้เป็นแบบสุ่มซึ่งเป็นวิธีที่ไม่ได้ใช้ความรู้ด้านอื่นมาพิจารณาช่วย จึงทำให้การระบุชิ้นส่วนไม่มีประสิทธิภาพ อย่างไรก็ตามเชื่อว่าปัจจัยหลักอย่างหนึ่งที่ทำให้วิธีการแตกเป็นชิ้นส่วนสามารถหาคำตอบได้ก็คือการแบ่งปันความรู้แบบรวมศูนย์ (Centralized) ดังนั้นขั้นตอนวิธีที่นำเสนอในบทนี้ยังคงใช้หลักการเดิมในการประกอบชิ้นส่วนเข้าด้วยกันเพียงแต่เปลี่ยนเฉพาะวิธีการระบุชิ้นส่วน จากความรู้ที่ได้จากวิธีการไขว้เปลี่ยนแบบชิ้นส่วนในบทที่ 5 ทำให้เชื่อว่าวิธีการระบุชิ้นส่วนที่ใช้หลักการสารสนเทศร่วม (Mutual Information) เป็นวิธีเหมาะสมสำหรับการระบุหน่วยการสร้างแบบชิ้นส่วนวิธีหนึ่งเนื่องจากทำให้ผลที่ดีขึ้น (เมื่อเทียบกับการไขว้เปลี่ยนแบบจุดเดียว) ดังนั้นในบทนี้เราจะใช้หลักการสารสนเทศร่วมในการระบุหน่วยการสร้างแต่มีรูปแบบที่เปลี่ยนไปเล็กน้อย เนื่องจากการไขว้เปลี่ยนแบบชิ้นส่วนมีการพิจารณาโครโมโซมร่วมกันที่ละคู่ (มีการแบ่งปันข้อมูลในระดับต่ำ) แต่ในบทนี้ต้องการให้มีการแบ่งปันข้อมูลในระดับที่สูงขึ้น โดยทำการเลือกโครโมโซมมาจำนวนหนึ่งด้วยวิธีการเลือกแบบตัวที่ดีที่สุดเป็นจำนวน n ตัว (n -Best Selection) ตั้งชื่อกลุ่มโครโมโซมที่ถูกเลือกเหล่านี้ว่า กลุ่มประชากรย่อย (Sub-population) หลังจากนั้นจะทำการแบ่งปันข้อมูลระหว่างประชากรที่ถูกเลือกทั้งหมด โดยใช้วิธีการเปรียบเทียบโครโมโซมทุกคู่ที่เป็นไปได้ในกลุ่มประชากรย่อยที่ถูกเลือก (โครโมโซมหนึ่งตัวจะจับคู่กับโครโมโซมทั้งหมดที่เหลือทุกตัว) เพื่อหาข้อมูลสารสนเทศร่วมกันในแต่ละคู่โครโมโซม การจับคู่กันที่ละ 2 โครโมโซมสามารถพิจารณาได้ว่าเป็นการหาข้อมูลสารสนเทศร่วมกันต่ำสุด (Minimum Mutual Information) ซึ่งเพียงพอต่อการระบุหน่วย

การสร้าง ถ้าใช้จำนวนโครโมโซมมากกว่านี้ในการหาข้อมูลสนเทศร่วมกันอาจทำให้สูญเสียความหลากหลายอย่างมีนัยสำคัญซึ่งจะส่งผลให้เกิดการลู่เข้าอย่างรวดเร็ว ตัวอย่างการจับคู่เพื่อหาข้อมูลสารสนเทศร่วมกันระหว่างโครโมโซมที่ถูกเลือก สมมติให้โครโมโซม C_1, C_2, C_3 และ C_4 ถูกเลือก สามารถหาข้อมูลสนเทศร่วมกันต่ำสุดได้โดยการจับคู่เปรียบเทียบดังนี้ C_1 กับ C_2, C_1 กับ C_3, C_1 กับ C_4, C_2 กับ C_3, C_2 กับ C_4 และ C_3 กับ C_4 สามารถจับคู่ได้ทั้งหมด 6 แบบที่ไม่ซ้ำกัน เป็นต้น ถ้ามีประชากรที่ถูกเลือก m ตัวสามารถคำนวณหารูปแบบการจับคู่ที่ไม่ซ้ำกันได้ทั้งหมด $\binom{m}{2}$ รูปแบบ ซึ่งจะเห็นว่ามิตการเพิ่มขึ้นของรูปแบบเป็น m^2

ชิ้นส่วนที่ได้จากหลักการข้อมูลสนเทศร่วมกันน้อยสุดนั้นจะแบ่งออกเป็น 2 รูปแบบ ได้แก่ ชิ้นส่วนร่วมกัน (Common Fragment) (ชิ้นส่วนที่มีรูปแบบเหมือนกัน) และชิ้นส่วนไม่ร่วมกัน (Uncommon Fragment) (ชิ้นส่วนที่มีรูปแบบต่างกัน) ในวิธีการไขว้เปลี่ยนแบบชิ้นส่วนจะเห็นว่า ชิ้นส่วนร่วมกัน (เหมือนกัน) ไม่มีบทบาทในการดำเนินการเพราะเมื่อเกิดการไขว้เปลี่ยนแล้วจะมีรูปแบบเหมือนเดิม (ไม่ทำให้เกิดความแตกต่าง) แต่สำหรับวิธีที่นำเสนอในบทนี้ ชิ้นส่วนแบบร่วมกันจะมีบทบาทสำคัญเท่ากับชิ้นส่วนไม่ร่วมกัน เนื่องจากเป็นชิ้นส่วนที่ได้จากโครโมโซมตัวเดียวกัน และชิ้นส่วนทั้งสองรูปแบบจะถูกเก็บรวมกันอยู่ในหน่วยเก็บถาวรโดยใช้หลักการเดียวกัน คือ จะเก็บเฉพาะรูปแบบชิ้นส่วนที่ไม่ซ้ำกัน ณ ตำแหน่งเดียวกันในโครโมโซม จากวิธีการจับคู่แบบพบกันหมดทุกตัวจะทำให้ได้ชิ้นส่วนจำนวนมากและมีความหลากหลายของรูปแบบชิ้นส่วนสูง เมื่อเปรียบเทียบกับวิธีการไขว้เปลี่ยนแบบชิ้นส่วนและวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม ซึ่งน่าจะเป็นผลดีสำหรับการสำรวจในช่วงรุ่นประชากรแรก ๆ และมองได้ว่าเป็นการเพิ่มการแสวงหาประโยชน์ในช่วงรุ่นประชากรหลัง ๆ จากการไขว้เปลี่ยนแบบชิ้นส่วนในบทที่ 5 ทำให้เราสามารถนิยามวิธีการระบุชิ้นส่วนได้ดังต่อไปนี้

เราสามารถนิยามชิ้นส่วนด้วยโครงสร้างย่อยร่วมกันของสองโครโมโซมใด ๆ กำหนดให้ลำดับของโครโมโซมความยาว l

$$C^k = c_1^k c_2^k \dots c_l^k : k \in I^+, c_n^k \in \{0,1\}, 1 \leq n \leq l \quad (8.1)$$

กำหนดให้เซตของดัชนีตำแหน่งอ้างอิงในโครโมโซมความยาว l จากตำแหน่งที่ f ถึงตำแหน่งที่ t

$$(f_i, t_i) = \{ z_i : 1 \leq f_i \leq z_i \leq t_i \leq l, f_i, t_i, z_i \in I^+ \} \quad (8.2)$$

and $\forall i, j \in I^+ : i \neq j, (f_i, t_i) \cap (f_j, t_j) = \emptyset$

กำหนดให้ลำดับย่อยร่วมกันระหว่างสองโครโมโซม C^{k_1} กับ C^{k_2} เป็น

$$F_i^{k_1, k_2} = c_{f_i}^{k_1} c_{f_i+1}^{k_1} \dots c_{t_i}^{k_1} \text{ such that } c_{z_i}^{k_1} = c_{z_i}^{k_2} \quad (8.3)$$

กำหนดให้ลำดับย่อยต่างกันระหว่างสองโครโมโซม C^{k_1} กับ C^{k_2} เป็น

$$F_{j,1}^{k_1, k_2} = c_{f_j}^{k_1} c_{f_j+1}^{k_1} \dots c_{t_j}^{k_1} \text{ such that } c_{z_j}^{k_1} \neq c_{z_j}^{k_2} \quad (8.4)$$

$$F_{j,2}^{k_1, k_2} = c_{f_j}^{k_2} c_{f_j+1}^{k_2} \dots c_{t_j}^{k_2} \text{ such that } c_{z_j}^{k_1} \neq c_{z_j}^{k_2} \quad (8.5)$$

เซตของลำดับย่อยที่เหมือนกันและต่างกันที่เรียงติดกันระหว่างสองโครโมโซม ณ ตำแหน่งใด ๆ นิยามดังนี้

$$S_i^{k_1, k_2} = \left\{ (f_i, F_i^{k_1, k_2}) : i \leq \left\lfloor \frac{l}{2} \right\rfloor \right\} \quad (8.6)$$

$$S_j^{k_1, k_2} = \left\{ (f_j, F_{j,1}^{k_1, k_2}), (f_j, F_{j,2}^{k_1, k_2}) : j \leq 2 \left\lfloor \frac{l}{2} \right\rfloor \right\} \quad (8.7)$$

$$S^{k_1, k_2} = \left\{ S_i^{k_1, k_2} \cup S_j^{k_1, k_2} : |S_i^{k_1, k_2}| + |S_j^{k_1, k_2}| \leq \left\lfloor \frac{l}{2} \right\rfloor + 2 \left\lfloor \frac{l}{2} \right\rfloor \right\} \quad (8.8)$$

จากนิยามเราสามารถแสดงตัวอย่างได้ดังรูปที่ 8.1 กำหนดให้โครโมโซมเรียงลำดับจากซ้าย (เริ่มต้น) ไปขวา สมมติให้โครโมโซมขนาด 10 บิตสองตัว มีค่าดังนี้ $C^1 = (1,0,1,1,0,1,0,1,0,1)$ และ $C^2 = (1,1,1,1,0,0,1,0,0,1)$ ดังนั้นจะได้ช่วงดัชนีตำแหน่งร่วมระหว่างโครโมโซม C^1 และ C^2 เป็น $(f_1, t_1) = (1,1)$, $(f_2, t_2) = (2,2)$, $(f_3, t_3) = (3,5)$, $(f_4, t_4) = (6,8)$, และ $(f_5, t_5) = (9,10)$ และจะได้ชิ้นส่วน $S^{1,2}$ ดังต่อไปนี้ $F_1^{1,2} = c_1^1 = 1$, $F_{2,1}^{1,2} = c_2^1 = 0$, $F_{2,2}^{1,2} = c_2^2 = 1$ และ $F_3^{1,2} = c_3^1 c_4^1 c_5^1 = 110$, $F_{4,1}^{1,2} = c_6^1 c_7^1 c_8^1 = 101$, $F_{4,2}^{1,2} = c_6^2 c_7^2 c_8^2 = 010$ และ $F_5^{1,2} = c_9^1 c_{10}^1 = 01$ ตามลำดับ

วิธีการที่นำเสนอในบทนี้จะมีการแบ่งปันข้อมูลร่วมกันใน 2 ระดับคือ ระดับโครงสร้างทางพันธุกรรมในขั้นตอนวิธีการระบุ และระดับชิ้นส่วนในขั้นตอนวิธีการประกอบ ซึ่งต่างจากวิธีการไขว้เปลี่ยนแบบชิ้นส่วนที่มีการแบ่งปันข้อมูลในระดับโครงสร้างทางพันธุกรรมสำหรับขั้นตอนวิธีการระบุเท่านั้น และต่างจากขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่มที่มีการแบ่งปันความรู้แค่ระดับชิ้นส่วนในขั้นตอนวิธีการประกอบเท่านั้น การแบ่งปันความรู้ใน 2 ระดับนี้เป็นหลักการที่สำคัญของขั้นตอนวิธีประมาณการแจกแจง (EDAs) เพราะประสิทธิภาพในการหาคำตอบขึ้นอยู่กับความสามารถที่ได้รับมาจากทั้งสองปัจจัย (การแบ่งปันข้อมูลในการระบุและการประกอบ) เป็น

หลัก เราจึงหวังว่าการแบ่งปันข้อมูลหรือความรู้ที่เพิ่มขึ้นจะช่วยให้ขั้นตอนวิธีที่น่าเสนอในบพนี้ (ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข) มีประสิทธิภาพที่ดีขึ้นด้วย (เมื่อเทียบกับวิธี sGA-FC และวิธี RFT)

ในส่วนของการประกอบชิ้นส่วนเข้าด้วยกันยังคงอาศัยวิธีการเดิมที่ใช้ในขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่มคือ ใช้วิธีการสุ่มเลือกชิ้นส่วนจากหน่วยเก็บถาวรตามลำดับตำแหน่งจากซ้ายไปขวา โดยแต่ละชิ้นส่วนในตำแหน่งเดียวกัน (เริ่มต้นที่ตำแหน่งเดียวกัน) มีโอกาสถูกเลือกเท่ากัน และถ้าตำแหน่งใดไม่มีชิ้นส่วนอยู่เลยจะใช้วิธีการสุ่มค่าจากรูปแบบสุ่มที่เป็นไปได้ทั้งหมดแทนที่ละตำแหน่ง การดำเนินการประกอบชิ้นส่วนแสดงดังรูปที่ 8.4

Bit Position :	1	2	3	4	5	6	7	8	9	10
Chromosome 1 (C1) :	1	0	1	1	0	1	0	1	0	1
Chromosome 2 (C2) :	1	1	1	1	0	0	1	0	0	1
Common Schema :	1	*	1	1	0	*	*	*	0	1
Fragment :	F₁			F₃			F₅			
Uncommon Schema C1:	*	0	*	*	*	1	0	1	*	*
Fragment :	F_{2,1}				F_{4,1}					
Uncommon Schema C2:	*	1	*	*	*	0	1	0	*	*
Fragment :	F_{2,2}				F_{4,2}					

รูปที่ 8.1 ชิ้นส่วนที่ระบุได้จากคูโครโมโซม

D คือ ประชากร

F คือ ชิ้นส่วน (Fragments)

T คือ หน่วยเก็บชิ้นส่วนแบบถาวร (Archive)

ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (Conditional Fragmentation)

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{g-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{g-1} ที่สอดคล้องกับวิธีการเลือก

$P \leftarrow$ หาชิ้นส่วนโดยเทียบทุกคู่ผลเฉลยจาก D_{g-1}^N

$T \leftarrow$ สร้างหน่วยเก็บชิ้นส่วนถาวร (Archive) จาก P

$D_g^N \leftarrow$ ซักตัวอย่างผลเฉลย R_g (ประชากรรุ่นใหม่) จาก T

$D_g^N \leftarrow$ นำผลเฉลยจากทุกตัวจาก D_g^N ไปทำการผ่าเหล่า (Mutation)

ด้วยวิธีแบบ sGA (เป็นทางเลือก แต่งานวิจัยนี้ไม่ใช่)

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 8.2 รหัสเทียมของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (CFT)

R คือ ผลเฉลย (โครโมโซม)

F คือ ชิ้นส่วนย่อย (แบบร่วมกัน และแบบไม่ร่วมกัน)

N คือ จำนวนผลเฉลยที่ถูกเลือก

ขั้นตอนวิธีการระบุชิ้นส่วน

{การเปรียบเทียบผลเฉลย R_i กับ R_1, \dots, R_{i-1} }

วนซ้ำสำหรับ $i = 2$ ถึง N ทำ

วนซ้ำสำหรับ $j = 1$ ถึง $(i-1)$ ทำ

$F \leftarrow$ เปรียบเทียบ R_i กับ R_j

(ค้นหาโครงสร้างย่อยแบบร่วมกัน และไม่ร่วมกัน)

จบการวนซ้ำ

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 8.3 รหัสเทียมของวิธีการระบุชิ้นส่วนสำหรับวิธี CFT

R คือ ผลเฉลย (โครโมโซม)

F_i คือ เซตของชิ้นส่วนย่อย (แบบร่วมกัน และแบบไม่ร่วมกัน)

เริ่มต้นที่ตำแหน่ง i .

C คือ ชิ้นส่วนย่อย (แบบร่วมกัน และแบบไม่ร่วมกัน)

ขั้นตอนวิธีการประกอบชิ้นส่วน

{การประกอบผลเฉลยใหม่ R }

$i = 1$ {ตำแหน่งแรก}

วนซ้ำในขณะที่ $i < \text{individual_size}$ ทำ

ถ้า $F_i \neq \emptyset$ ทำ {ถ้ามีชิ้นส่วนย่อยตั้งแต่หนึ่งชิ้น}

$C \leftarrow \text{สุ่มเลือก}(F_i)$

$R \leftarrow R + C$ {นำชิ้นส่วนมาเรียงต่อกันตามลำดับ}

$i \leftarrow i + \text{size}(C)$ {หาตำแหน่งเริ่มต้นถัดไป}

นอกจากนั้น {ถ้าไม่มีสมาชิกอยู่เลย}

$R \leftarrow \text{สุ่มเลือก}(0,1)$ {สุ่มจากค่า 0 หรือ 1}

$i \leftarrow i + 1$

จบเงื่อนไข

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 8.4 รหัสเทียมของวิธีการประกอบชิ้นส่วนสำหรับวิธี CFT

จากขั้นตอนวิธีที่ออกแบบเราสามารถคำนวณหาโอกาสที่ชิ้นส่วนจะอยู่รอดจากการดำเนินการต่าง ๆ เพื่อให้เห็นภาพรวมของกลไกการทำงานอย่างคร่าว ๆ การคำนวณจะพิจารณาเพียงแค่รุ่นประชากรเดียว (จากรุ่น t ไป $t+1$ ดังต่อไปนี้

โอกาสที่เค้าร่าง H ใด ๆ จะเกิดขึ้นได้จากวิธีการระบุหน่วยการสร้างเป็น

$$P_{\text{identify}}(H) = 2^{-(D(H) + o(H) + 1)} \quad (8.9)$$

เมื่อ $P_{\text{identify}}(H)$ คือ โอกาสเกิดเค้าร่าง H จากวิธีการระบุชิ้นส่วน

$D(H)$ คือ ความยาวของเค้าร่าง H

$o(H)$ คือ ลำดับของเค้าร่าง H

โอกาสที่เค้าร่าง H จะสามารถอยู่รอดได้จากการประกอบชิ้นส่วนเข้าด้วยกัน โดยมีเงื่อนไขว่าในหน่วยเก็บชิ้นส่วนถาวรจะต้องมีชิ้นส่วนที่เป็นไปได้ทุกขนาดและรูปแบบเป็น

$$P_{compose}(H) = \sum_{b=1}^p \left(\frac{\sum_{j=(p-b)+(D(H)+1-o(H))}^{l-o(H)-b+1} 2^j}{\sum_{i=1+(p-b)}^{l-b+1} 2^i} \right) \quad (8.10)$$

เมื่อ $P_{compose}(H)$ คือ โอกาสเค้าร่าง H จะรอดจากวิธีการประกอบชิ้นส่วนเข้าด้วยกัน

$D(H)$ คือ ความยาวของเค้าร่าง H

$o(H)$ คือ ลำดับของเค้าร่าง H

l คือ ขนาดความยาวของโครโมโซม

p คือ ตำแหน่งแรกของเค้าร่างที่มีค่าแน่นอน (fixed bit) โดยอ้างอิงจากตำแหน่งของโครโมโซม

b, i, j คือ ค่าดัชนีเพื่อใช้ในการคำนวณ

เมื่อพิจารณาถึงตัวดำเนินการเลือกสรรที่ใช้เป็นการเลือกแบบ n ตัวที่ดีที่สุด ดังนั้นเราสามารถหาความสัมพันธ์ของสัดส่วนความอยู่รอดของจำนวนโครโมโซมที่มีเค้าร่าง H ได้เป็น

$$P_{select}(H) = \frac{s}{n} P_{(H)} \quad (8.11)$$

เมื่อ $P_{select}(H)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากรที่ถูกเลือก

$P_{(H)}$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากร

s คือ ขนาดของประชากร (จำนวนโครโมโซมที่มีทั้งหมด)

n คือ จำนวนโครโมโซม (ที่ดีที่สุด) ที่ถูกเลือก

ถ้าพิจารณาในมุมมองเดียวกันกับทฤษฎีบทเค้าร่างจะได้สมการเป็น

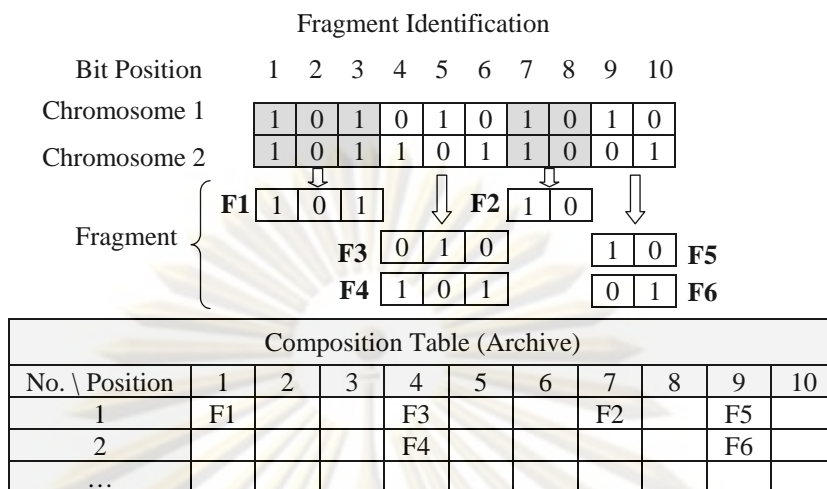
$$P(H, t+1) = P_{select}(H, t) \cdot P_{identify}(H, t) \cdot P_{compose}(H, t) \quad (8.12)$$

เมื่อ $P(H, t+1)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากรในรุ่นที่ $t+1$

$P_{select}(H, t)$ คือ สัดส่วนของโครโมโซมที่มีเค้าร่าง H ต่อประชากร ที่ถูกเลือกจากรุ่นที่ t

$P_{identify}(H)$ คือ โอกาสที่จะเกิดขึ้นเค้าร่าง H จากวิธีการระบุชิ้นส่วนจากรุ่นที่ t

$P_{compose}(H)$ คือ โอกาสที่เค้าร่าง H จะอยู่รอดจากวิธีการประกอบชิ้นส่วนจากรุ่นที่ t



After fully comparison of all individuals

Composition Table (Archive)										
No. \ Position	1	2	3	4	5	6	7	8	9	10
1	F1	F22	F12	F3	F7	F10	F2	F11	F5	
2	F8	F27	F25	F4	F20	F13	F26	F14	F6	
3	F9		F17	F15	F23	F28	F18	F16	F21	
4	F19		F29	F37	F35		F30	F24	F34	
5	F31		F33		F39		F32	F38		
6	F36							F40		
...										

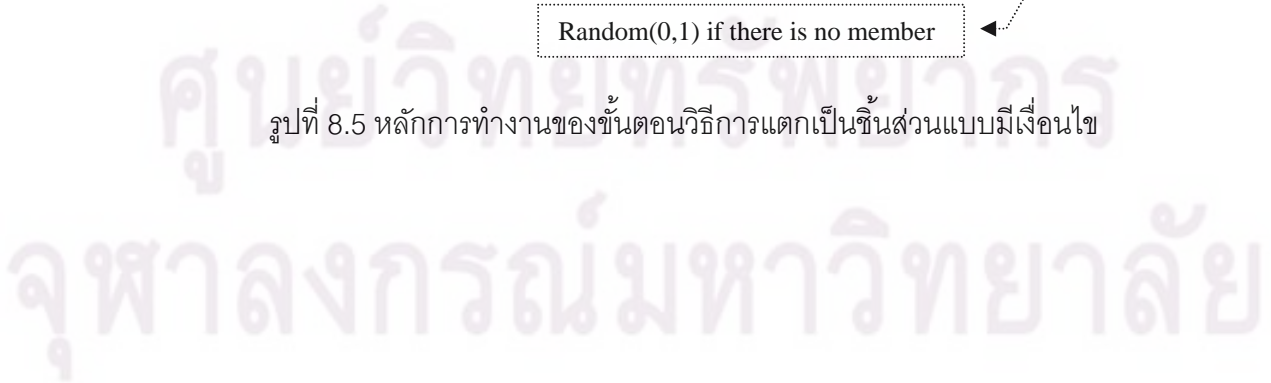
Fragment composition (new individual)

Bit Position	1	2	3	4	5	6	7	8	9	10
New Chromosome 1	F31			F12			F32		F5	

Bit Position	1	2	3	4	5	6	7	8	9	10
New Chromosome 2	F9		F37			F13			R	

Random(0,1) if there is no member

รูปที่ 8.5 หลักการทำงานของขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข



8.2 การทดลองและผลการทดลอง

ในส่วนนี้จะกล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่งในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่น่าเสนอ และมีการวิจารณ์ผลการทดลองในมุมมองที่น่าเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ-1

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม (sGA)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การไขว้เปลี่ยน:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่ใช้เปรียบเทียบ-2

ขั้นตอนวิธี:	ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (RFT)
การเลือกสรร:	วิธีการเลือกแบบ n ตัวที่ดีที่สุด (n -Best) ค่า n คือค่า #sub ที่แสดงในตารางที่ 8.2
การระบุ:	วิธีสุ่มตัดแบบชิ้นส่วน
การประกอบ:	วิธีชักตัวอย่างจากหน่วยเก็บถาวรตามลำดับตำแหน่งปรากฏ
การผ่าเหล่า:	ไม่มี

วิธีที่น่าเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีการแตกเป็นชิ้นส่วนอย่างมีเงื่อนไข (CFT)
การเลือกสรร:	วิธีการเลือกแบบ n ตัวที่ดีที่สุด (n -Best) ค่า n คือค่า #sub ที่แสดงในตารางที่ 8.1 และ 8.2
การระบุ:	วิธีข้อมูลเสนอพร้อมกันและไม่พร้อมกันต่ำสุดแบบชิ้นส่วน
การประกอบ:	วิธีชักตัวอย่างจากหน่วยเก็บถาวรตามลำดับตำแหน่งปรากฏ
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 8.1 และ 8.2

ตารางที่ 8.1 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA-FC และ CFT

Parameters		Success 30 run				Failure 30 run	
		sGA-FC		CFT		sGA-FC	CFT
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
				#sub			#sub
RoyalRoad	64	1,200	13,480	2,000	13,867	100	200
				200			100
Trap-5	60	1,600	20,854	1,000	7,834	300	100
				100			50

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ตารางที่ 8.2 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ RFT และ CFT

Parameters		Success 30 run				Failure 30 run	
		RFT		CFT		RFT	CFT
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
		#sub		#sub		#sub	#sub
RoyalRoad	64	1,000	51,734	2,000	13,867	200	200
		200		200		100	100
Trap-5	60	8,000	134,934	1,000	7,834	1200	100
		800		100		200	50

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ผลการทดสอบในตารางที่ 8.1 เมื่อเทียบกับวิธี sGA-FC ที่ใช้หลักการสารสนเทศต่ำสุดในการระบุชิ้นส่วนเหมือนกันแต่ต่างกันที่ระดับการแบ่งปันความรู้ สำหรับปัญหา Royal Road ค่า FEs ของทั้งสองวิธีมีค่าใกล้เคียงกันมาก เมื่อทดสอบด้วยวิธี t-Test ได้ผลว่าทั้งสองวิธีมีประสิทธิภาพไม่แตกต่างกันอย่างมีนัยสำคัญที่ความเชื่อมั่น 95% ส่วนปัญหา Trap-5 ค่า FEs ของวิธี CFT น้อยกว่าวิธี sGA-FC อย่างเห็นได้ชัด เป็นการแสดงให้เห็นว่าวิธี CFT มีประสิทธิภาพใน

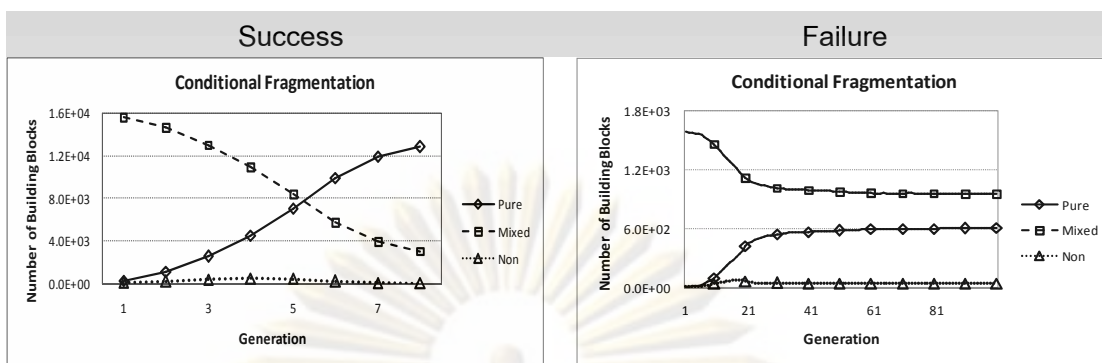
การแก้ปัญหา Trap-5 ดีกว่าวิธี sGA-FC และในมุมมองของความหลากหลายสำหรับปัญหา Royal Road วิธี CFT ใช้จำนวนประชากรมากกว่าวิธี sGA-FC แสดงให้เห็นว่าวิธี CFT ต้องการความหลากหลายเพื่อใช้ในการแก้ปัญหามากกว่าวิธี sGA-FC ทั้งในกรณีที่ทำคำตอบพบและหาคำตอบไม่พบ ส่วนปัญหา Trap-5 วิธี CFT ต้องการระดับความหลากหลายน้อยกว่าวิธี sGA-FC เนื่องจากใช้จำนวนประชากรน้อยกว่าทั้งในกรณีที่ทำคำตอบพบและหาคำตอบไม่พบ

เมื่อพิจารณาผลของวิธี CFT จากตารางที่ 8.2 เปรียบเทียบกับวิธี RFT ที่ใช้หลักการแบ่งปันความรู้แบบรวมศูนย์ในการประกอบชิ้นส่วนเข้าด้วยกันเหมือนกับวิธี CFT แต่มีความแตกต่างกันที่วิธีการระบุชิ้นส่วน โดยที่วิธี RFT ใช้วิธีการระบุชิ้นส่วนแบบสุ่มเท่านั้น ส่วนวิธี CFT ใช้หลักการสารสนเทศร่วมกันต่ำสุดและมีการแบ่งปันความรู้ในส่วนของการระบุชิ้นส่วนด้วย ผลการทดสอบแสดงให้เห็นว่าวิธี CFT มีประสิทธิภาพในการหาคำตอบสูงกว่าวิธี RFT เพราะวิธี CFT มีค่า FEs น้อยกว่าในทุกปัญหาอย่างชัดเจน ซึ่งเป็นการยืนยันข้อสันนิษฐานที่ว่ากระบวนการสร้างที่มีประสิทธิภาพจะมีผลทำให้ได้คำตอบที่ดีขึ้น ในปัญหา Royal Road วิธี CFT ต้องการระดับความหลากหลายในการแก้ปัญหามากกว่าวิธี RFT เนื่องจากใช้จำนวนประชากรมากถึงหนึ่งเท่าตัว แต่ในกรณีที่ไม่สามารถหาคำตอบได้พบใช้จำนวนประชากรเท่ากัน ซึ่งเป็นการแสดงถึงระดับความหลากหลายที่ต้องการน้อยที่สุดในการแก้ปัญหามหา Royal Road มีค่าเท่ากัน ส่วนปัญหา Trap-5 วิธี CFT ใช้จำนวนประชากรน้อยกว่าวิธี RFT เป็นอย่างมาก หรือกล่าวได้ว่าวิธี CFT ต้องการความหลากหลายในการแก้ปัญหามากกว่าวิธี RFT อย่างชัดเจน

กราฟผลการทดลองในรูปที่ 8.6 มีรูปแบบเหมือนกับผลของ sGA-FC ในบทที่ 5 รูปที่ 5.9 ทุกกราฟ (ยกเว้นรูปที่ 8.6 (ข)) เป็นเพราะในส่วนของการระบุชิ้นส่วนในวิธี CFT ใช้หลักการสารสนเทศร่วมกันต่ำสุดแบบเดียวกับวิธี sGA-FC นั่นเอง ผลที่คล้ายกันนี้เป็นตัวชี้ให้เห็นว่าวิธีการระบุชิ้นส่วนมีอิทธิพลต่อการหาคำตอบอย่างชัดเจน ในกรณีที่หาคำตอบพบใช้จำนวนรุ่นประชากรเพียงแค่ 8 รุ่นเท่านั้น ถือว่าน้อยเมื่อเทียบกับขั้นตอนวิธีอื่น เป็นเพราะว่าในกระบวนการเลือกสรรใช้กลุ่มประชากรย่อยจำนวนน้อยเมื่อเทียบเป็นอัตราส่วนกับขนาดของประชากร สำหรับปัญหา Royal Road กรณีที่ทำคำตอบพบเรากำหนดให้ขนาดประชากรเท่ากับ 2,000 ตัว เลือกกลุ่มประชากรย่อยขนาด 200 ตัว ซึ่งคิดเป็น 10% ของจำนวนประชากรทั้งหมด จึงทำให้เกิดการลู่เข้าเร็วเพราะมีระดับความหลากหลายต่ำ การเลือกประชากรย่อยจำนวนน้อยแล้วยังสามารถหาคำตอบได้เป็นการยืนยันความเชื่อที่ว่า ในกลุ่มประชากรตัวที่ดี ๆ มีข้อมูลโครงสร้างที่ดีอยู่ร่วมกันจริง ถ้าสามารถระบุหรือหาโครงสร้างเหล่านี้ได้ถูกต้องจะช่วยให้การหาคำตอบมีประสิทธิภาพ อาจเกิดข้อสงสัยว่าทำไมไม่ใช้จำนวนประชากรมาก แต่เลือกประชากรย่อยน้อย (เพียงแค่ 10%) เหตุที่ต้องเลือกมาน้อยเพราะต้องการความกดดันในการเลือกสูงเพื่อให้แน่ใจได้ว่าประชากรที่ถูกเลือก

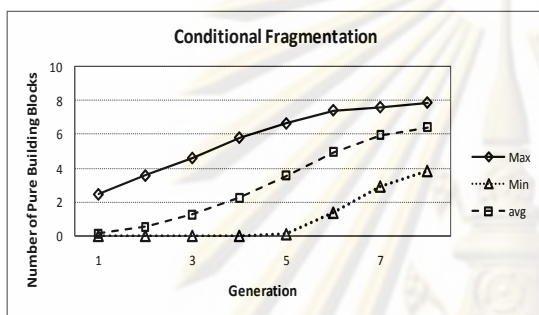
เป็นประชากรที่มีคุณภาพจริง ๆ และการใช้จำนวนประชากรมากเป็นการเพิ่มโอกาสให้เกิดประชากรที่ดี ๆ เมื่อพิจารณาถึงวิธีในการระบุชิ้นส่วนที่มีการเปรียบเทียบจากทุกคู่ของโครโมโซมในกลุ่มประชากรย่อยจะเห็นว่า โอกาสที่คู่โครโมโซมใด ๆ จะมีโครงสร้างคู่จีนที่เหมือนกันถึง 8 บิต เท่ากับขนาดหน่วยการสร้างของปัญหา Royal Road เกิดขึ้นได้น้อยเมื่อเทียบกับปัญหา Trap-5 ที่มีขนาดหน่วยการสร้างเพียง 5 บิต ด้วยเหตุนี้ทำให้ต้องใช้จำนวนประชากรในการแก้ปัญหา Royal Road มากกว่าปัญหา Trap-5 อย่างไรก็ตามวิธีวิธีการระบุชิ้นส่วนแบบเทียบทุกคู่ของโครโมโซมที่ถูกเลือกสามารถให้ความหลากหลายที่มากเพียงพอ (ทำให้เกิดชิ้นส่วนจำนวนมาก) เมื่อพิจารณาจากความแตกต่างกันของ Max Pure BBs และ Min Pure BBs ในรูปที่ 8.6 (ค) และจำนวนชิ้นส่วนที่มีทั้งหมดในแต่ละรุ่นประชากรดังรูปที่ 8.6 (ข) เทียบกับขั้นตอนวิธีอื่นที่นำเสนอ เมื่อสังเกตในรูปที่ 8.6 (ก) จะเห็นว่าปริมาณของ Non BBs มีการขยับตัวขึ้นมาเล็กน้อยเกิดจากวิธีการระบุที่มีความหลากหลายสูงนั่นเอง และแม้ว่ามีชิ้นส่วนจำนวนมากแต่ก็ยังสามารถหาคำตอบได้รวดเร็ว เป็นเพราะชิ้นส่วนทั้งหมดที่ได้เกิดจากกลุ่มประชากรย่อยที่มีขนาดเล็กมากเมื่อเทียบกับขนาดประชากร ทำให้ได้โครโมโซมมีหน้าตาใกล้เคียงกัน (เลือกตัวที่ดีที่สุดมา) ส่งผลให้ชิ้นส่วนที่ได้มีรูปแบบโครงสร้างและขนาดใกล้เคียงกัน ประกอบกับเป็นโครงสร้างที่ได้จากโครโมโซมที่ดี ทำให้สามารถไปถึงคำตอบที่ต้องการได้อย่างรวดเร็ว ในกรณีที่หาคำตอบไม่พบในรูปที่ 8.6 (ข) Mixed BBs มีปริมาณมากกว่า Pure BBs และเกิดการลู่เข้าก่อนที่ค่าทั้งสองจะตัดกัน เกิดจากการใช้ประชากรย่อยจำนวนมากเมื่อเทียบกับจำนวนประชากรทั้งหมด (ใช้ประชากรย่อยเท่ากับ 50% ของประชากรทั้งหมด) ทำให้มีความหลากหลายสูงสังเกตได้จากความแตกต่างของ Max Pure BBs และ Min Pure BBs ในรูปที่ 8.6 (ง) และส่งผลให้เกิดการลู่เข้าก่อนที่จะสามารถหา Pure BBs ได้มากขึ้นเนื่องจากมีประชากรน้อย (มีเพียง 200 ตัว)

ในมุมมองของชิ้นส่วนซึ่งแบ่งออกเป็นชิ้นส่วนที่เหมือนกันและชิ้นส่วนที่ต่างกัน ในกรณีที่หาคำตอบพบจากรูปที่ 8.6 (จ) แสดงให้เห็นว่าเกิดการลู่เข้าอย่างรวดเร็วเนื่องจากขนาดโดยเฉลี่ยของชิ้นส่วนที่เหมือนกันโตขึ้นเป็นฟังก์ชันเลขชี้กำลัง และปริมาณชิ้นส่วนที่ลดลงอย่างรวดเร็ว ในรูปที่ 8.6 (ข) ส่วนกรณีที่หาคำตอบไม่พบในรูปที่ 8.6 (ฉ) ขนาดโดยเฉลี่ยของชิ้นส่วนที่เหมือนกันโตขึ้นอย่างรวดเร็ว ในขณะที่เดียวกันขนาดโดยเฉลี่ยของชิ้นส่วนที่ไม่เหมือนกันก็มีการปรับตัวเพิ่มขึ้นเล็กน้อย เนื่องจากมี Mixed BBs จำนวนมาก (จากรูปที่ 8.6 (ข)) ทำให้ชิ้นส่วนที่ไม่เหมือนกันมีขนาดใหญ่ขึ้นด้วย และส่งผลให้การลู่เข้าเกิดขึ้นอย่างรวดเร็วในช่วงต้นจนถึงรุ่นประชากรที่ 31 และค่อย ๆ ลดลงอย่างต่อเนื่องจนถึงรุ่นสุดท้าย (รุ่นที่ 100) เนื่องจากยังคงมีความหลากหลายเหลืออยู่จากชิ้นส่วนที่ต่างกัน ดังรูปที่ 8.6 (ซ)

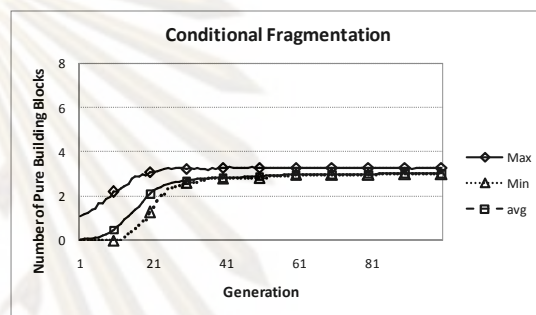


(ก)

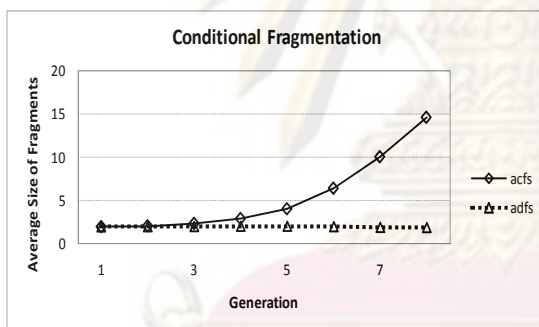
(ข)



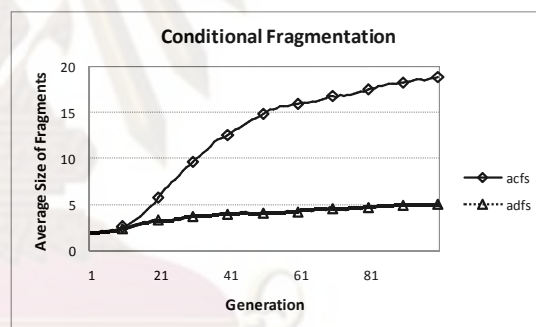
(ค)



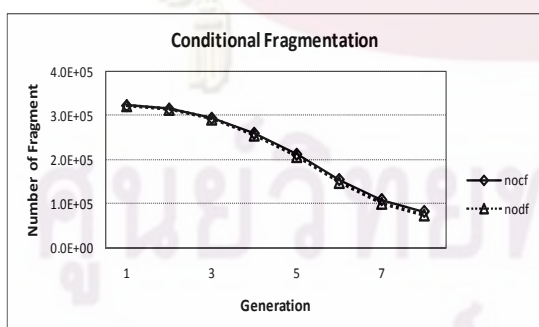
(ง)



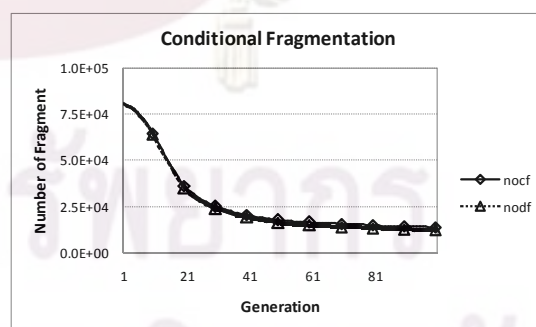
(จ)



(ฉ)



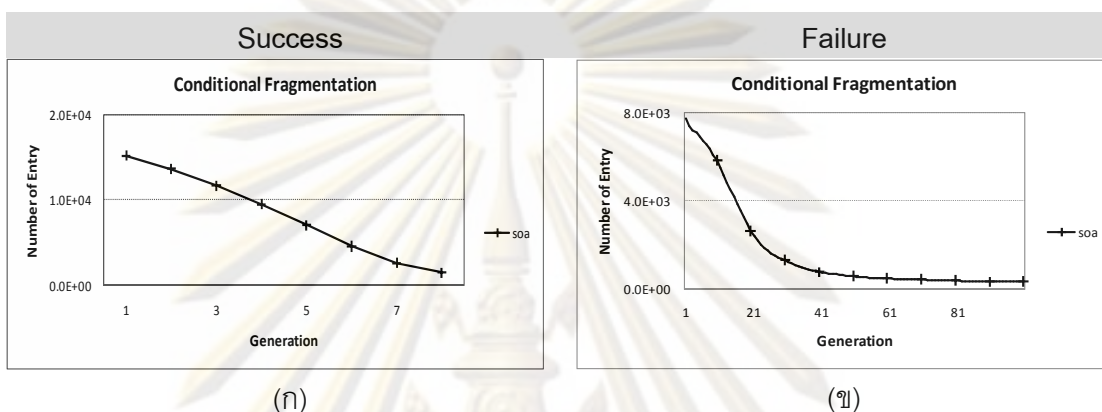
(ช)



(ซ)

รูปที่ 8.6 ผลการทดสอบของ CFT ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นของประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ซ)

เมื่อพิจารณาถึงปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวร ซึ่งเก็บเฉพาะชิ้นส่วนที่มีรูปแบบไม่ซ้ำกันในตำแหน่งเดียวกันของโครโมโซม จากรูปที่ 8.7 (ก-ข) ปริมาณของชิ้นส่วนที่มีในหน่วยเก็บถาวรถือว่าน้อยมากเมื่อเทียบกับปริมาณชิ้นส่วนที่ระบุได้ทั้งหมดในรูปที่ 8.6 (ซ-ช) ซึ่งเป็นการแสดงให้เห็นว่าชิ้นส่วนที่ระบุได้มีรูปแบบที่ซ้ำกันเป็นจำนวนมาก เกิดจากการใช้จำนวนประชากรย่อยน้อย และมีความกดดันในการเลือกสูง (ตรงกับคำอธิบายในส่วนก่อนหน้า)



รูปที่ 8.7 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ CFT ในปัญหา Royal Road 64 บิต

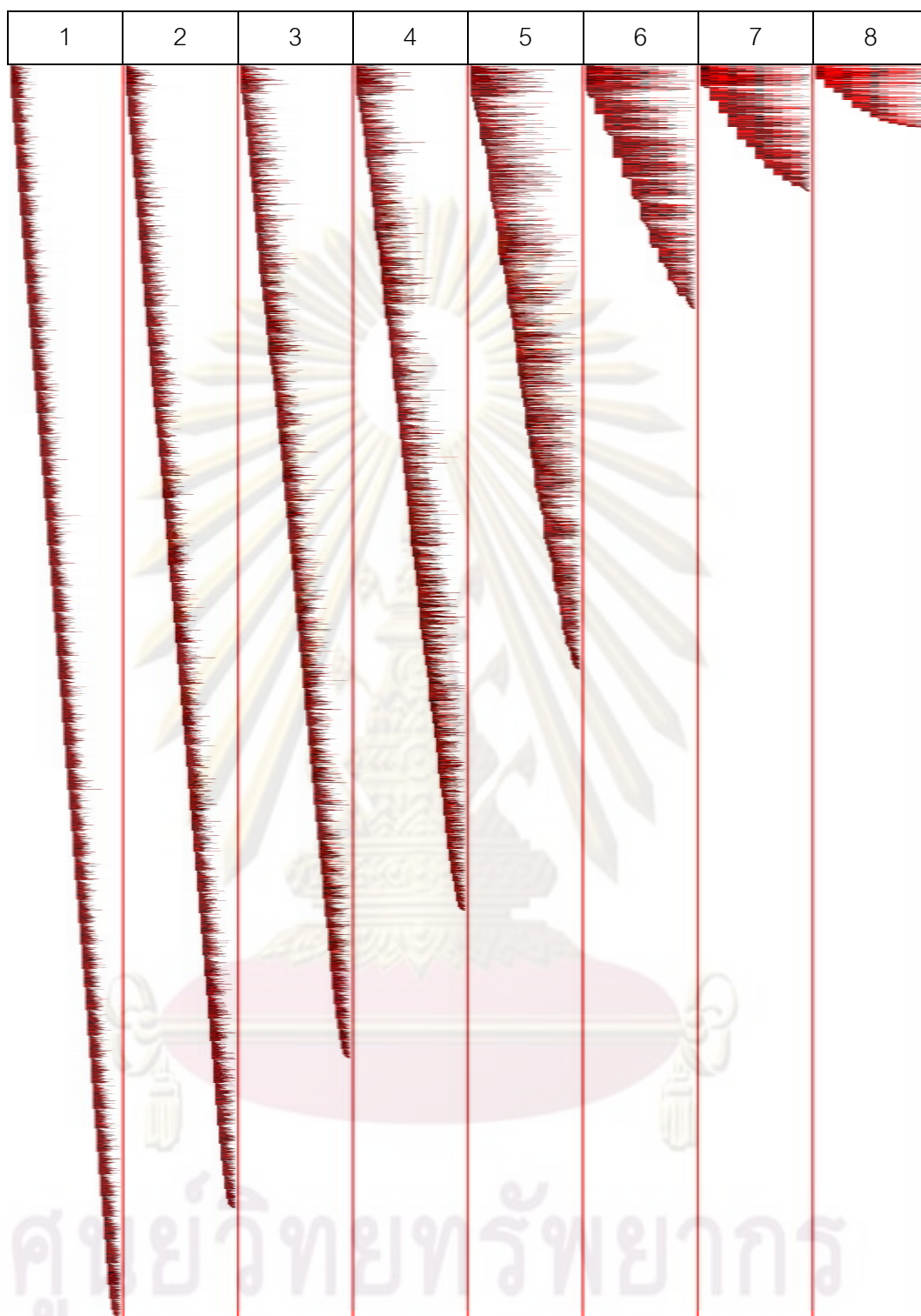
เนื่องจากวิธี CFT มีการใช้หน่วยเก็บถาวรร่วมกัน (แบ่งปันความรู้แบบรวมศูนย์) แบบเดียวกับวิธี RFT ในบทที่ 9 ดังนั้นจึงมีการแสดงถึงโครงสร้างทางพันธุกรรมของชิ้นส่วนที่เก็บอยู่ในหน่วยเก็บถาวร เพื่อให้เห็นภาพในรายละเอียดได้ชัดเจนขึ้น (วิธีการอ่านรูปอยู่ในบทที่ 4) ดังรูปที่ 8.8 และรูปที่ 8.9 จะเห็นเห็นว่าในแต่ละตำแหน่งของโครโมโซมมีปริมาณชิ้นส่วนใกล้เคียงกันในกลุ่มประชากรแรก ๆ และค่อย ๆ ลดลงจนมีเฉพาะบางตำแหน่งเท่านั้น (จะเห็นเป็นรูปขั้นบันได) และเมื่อเกิดการลู่เข้าในกลุ่มประชากรท้าย ๆ ชิ้นส่วนที่เริ่มจากตำแหน่งแรกมีปริมาณมากที่สุด เนื่องจากประชากรย่อยมีหน้าตาเหมือนกันมาก และการระบุชิ้นส่วนเริ่มจากตำแหน่งแรกเสมอ การกระจายตัวของโครงสร้างชิ้นส่วนในแต่ละตำแหน่งของโครโมโซมเป็นตัวบ่งชี้ว่า วิธีการระบุชิ้นส่วนที่มีการเปรียบเทียบทุกคู่ของโครโมโซมทำให้เกิดความหลากหลายสูง และเมื่อสังเกตลักษณะโครงสร้างทางพันธุกรรมของแต่ละชิ้นส่วนในตำแหน่งใกล้เคียงกันจะเห็นว่า มีรูปแบบคล้ายกัน เนื่องจากมีประชากรย่อยน้อย และประชากรย่อยแต่ละตัวมีโครงสร้างใกล้เคียงกันเพราะมีความกดดันในการเลือกสูง ในกรณีที่หาคำตอบพบจากรูปที่ 8.8 ขนาดของชิ้นส่วนจะค่อย ๆ โตขึ้นเนื่องจากเกิดการลู่เข้า และการกระจายตัวของชิ้นส่วนยังคงครอบคลุมทุกตำแหน่งจนกระทั่งหาคำตอบพบในกลุ่มประชากรที่ 8 ส่วนกรณีที่ไม่หาคำตอบจากรูปที่ 8.9 ขนาดของชิ้นส่วนจะโตขึ้นเรื่อย ๆ เนื่องจากการลู่เข้า และตั้งแต่วรุ่นประชากรที่ 20 ชิ้นส่วนเริ่มไม่กระจายตัวครอบคลุมทุกตำแหน่งเหมือนกรณี

ที่หาคำตอบพบ เป็นเพราะขาดความหลากหลาย และเห็นได้ชัดเจนขึ้นในรุ่นประชากรที่ 30 เป็นต้นไป

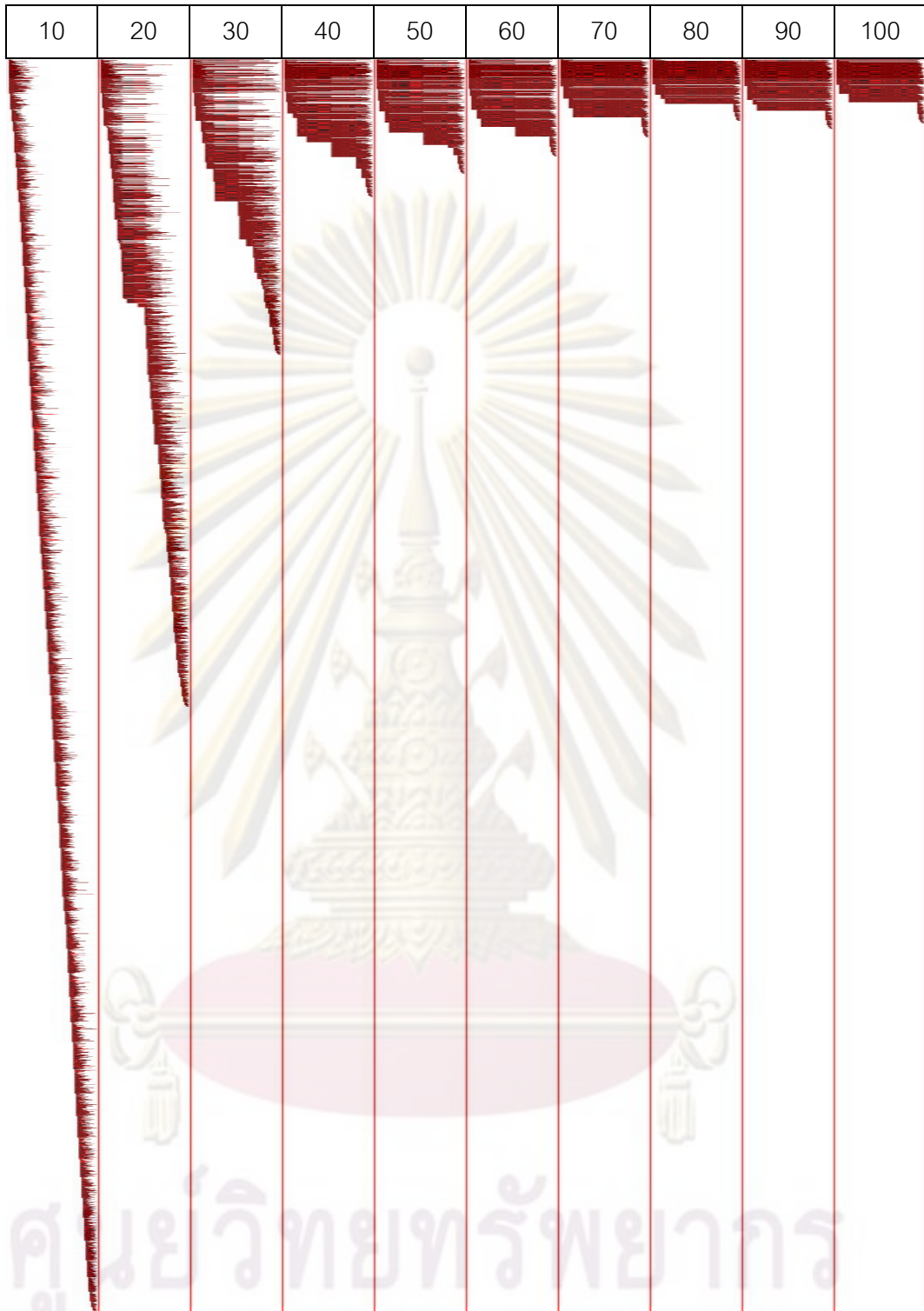
พิจารณาโครงสร้างทางพันธุกรรมในภาพรวมของประชากร (วิธีการอ่านรูปอยู่ในบทที่ 4) ในกรณีที่หาคำตอบพบจากรูปที่ 8.10 จะเห็นว่าเกิดการลู่เข้าสู่คำตอบตั้งแต่วรุ่นประชากรที่ 4 และเริ่มชัดเจนขึ้นในรุ่นประชากรที่ 6 ส่วนกรณีที่หาคำตอบไม่พบจากรูปที่ 8.11 เริ่มเกิดการลู่เข้าสู่ในรุ่นประชากรที่ 15 และชัดเจนขึ้นตั้งแต่วรุ่นประชากรที่ 25 เป็นต้นไป รูปแบบชิ้นส่วนในตำแหน่งที่เกิดการลู่เข้าแล้วสามารถสังเกตเห็นได้อย่างชัดเจน ขนาดและปริมาณของชิ้นส่วนร่วมกันและไม่ร่วมกันมีค่าใกล้เคียงกันในตำแหน่งที่เกิดการลู่เข้าแล้ว (สลับเป็นท่อนขาวดำในแนวดิ่ง) เหตุที่ทำให้เห็นเป็นท่อนชัดเจนมากกว่าขั้นตอนวิธีอื่นที่น่าเสนอ เป็นเพราะวิธีการระบุชิ้นส่วนทำให้ได้โครงสร้างที่ชัดเจน และประชากรทุกตัวสร้างขึ้นจากชิ้นส่วนที่เก็บอยู่ในหน่วยเก็บถาวรอันเดียวกัน (จึงสะท้อนให้เห็นภาพของหน่วยเก็บถาวรร่วมกัน)



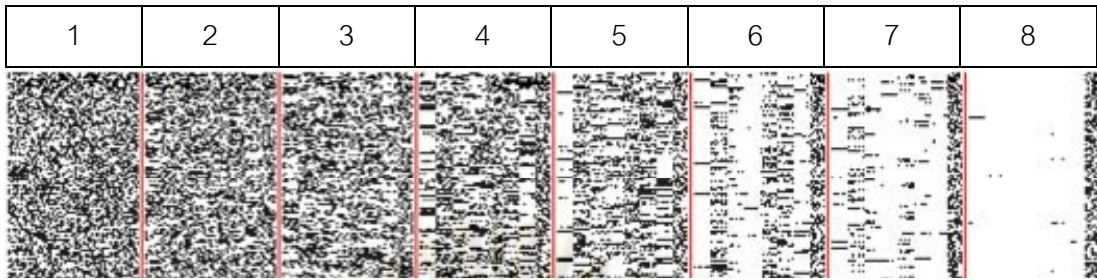
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



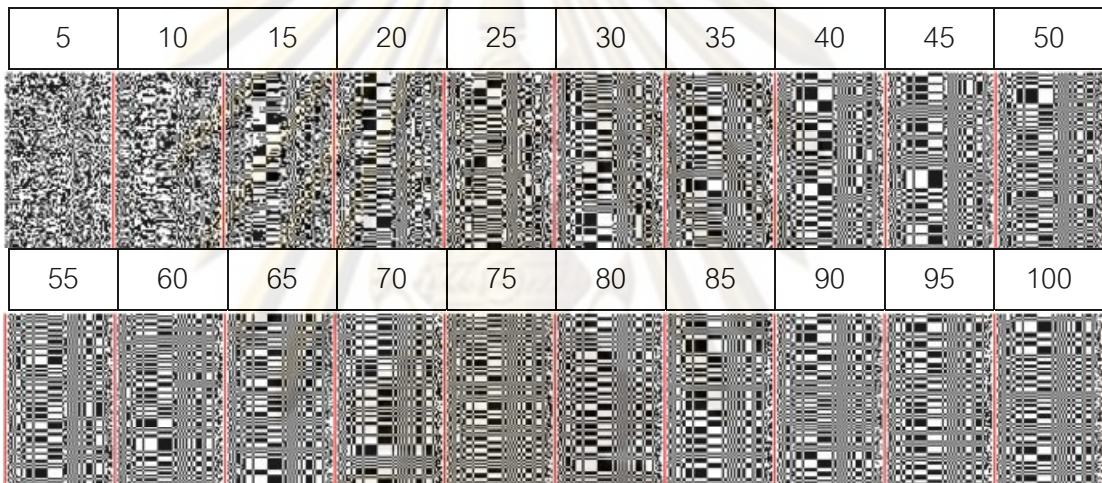
รูปที่ 8.8 รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 8 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่อง หมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จีโนที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จีโนที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 8.9 รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่เงินที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่เงินที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



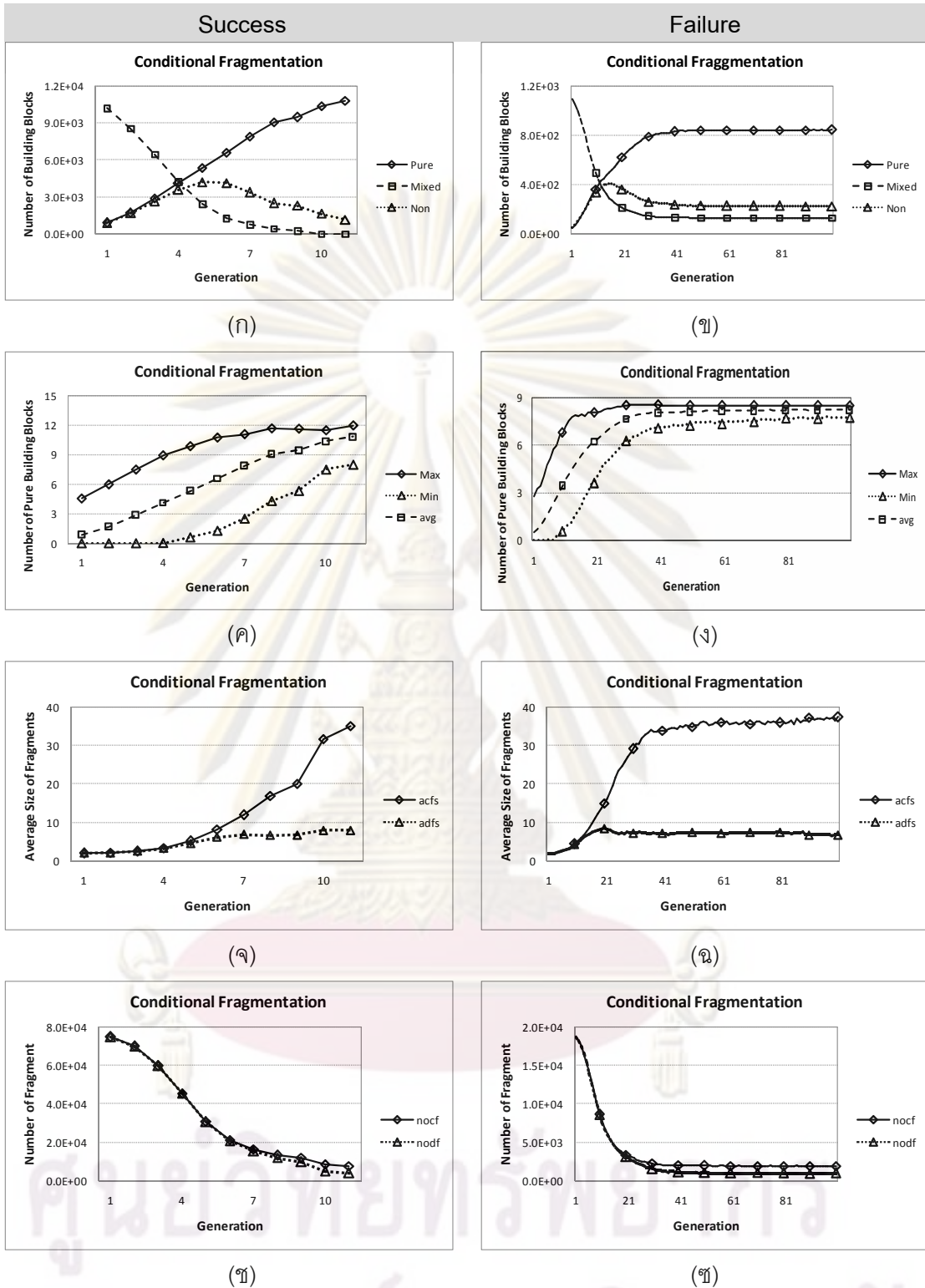
รูปที่ 8.10 รูปแบบจิงโนไทป์ของประชากรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 8 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0



รูปที่ 8.11 รูปแบบจิงโนไทป์ของประชากรของ CFT ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0

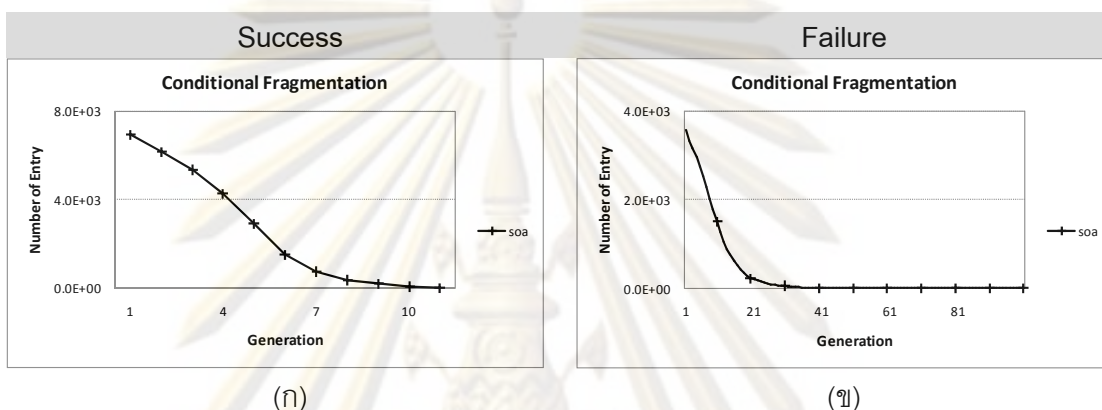
สำหรับปัญหา Trap-5 ผลการทดลองในรูปที่ 8.12 ยังคงมีรูปแบบเหมือนกับผลของ sGA-FC ในบทที่ 5 รูปที่ 5.13 ทุกกราฟ เมื่อพิจารณาในมุมมองของหน่วยการสร้างในรูปที่ 8.12 (ก) นอกจากมี Mixed BBs และ Pure BBs แข่งขันกันแล้วแล้วยังมี Non BBs ร่วมแข่งขันด้วย เป็นเพราะปัญหา Trap-5 มีการลงให้ได้โครงสร้างที่ตรงข้ามกับคำตอบ ด้วยเหตุนี้จึงทำให้ปัญหา Trap-5 ใช้จำนวนรุ่นประชากรมากกว่าปัญหา Royal Road ในการแก้ปัญหา ในกรณีที่หาคำตอบพบวิธี CFT สามารถแยกแยะระหว่างหน่วยการสร้างที่เป็นคำตอบ (Pure BBs) และหน่วยการสร้างหลง (Non BBs) ได้ ส่วนในกรณีที่หาคำตอบไม่พบจะเกิดการลู่เข้าสู่ค่าเหมาะสมที่สุดท้องถิ่น ในมุมมองของความหลากหลายจากรูปที่ 8.12 (ค-ง) แสดงให้เห็นว่าวิธี CFT มีความหลากหลายสูง เนื่องจากมีโครงสร้างของโครโมโซมที่แตกต่างกันสูง อ้างอิงจากจำนวนของหน่วยการสร้างของคำตอบที่มีอยู่ในโครโมโซมเป็นหลัก ค่า Max Pure BBs มีค่าต่างจาก Min Pure BBs มากเมื่อเปรียบเทียบกับวิธี sGA และวิธี sGA-FC (ยกเว้นวิธี RFT) และเห็นได้ชัดในกรณีที่หาคำตอบไม่พบ ค่าของ Max Pure BBs และ Min Pure BBs ยังคงมีความแตกต่างกันอยู่ระดับหนึ่งจนถึงรุ่นประชากรที่ 100 (รุ่นสุดท้ายที่กำหนด)

ในมุมมองของชิ้นส่วนในรูปที่ 8.12 (จ) ขนาดโดยเฉลี่ยของชิ้นส่วนที่ไม่เหมือนกันจะโตขึ้นทีละน้อยเป็นผลจากการลงที่ทำให้เกิดความแตกต่างในโครงสร้างโครโมโซม สังเกตได้จากปริมาณของ Non BBs ที่เพิ่มขึ้นในรูปที่ 8.12 (ก) สำหรับกรณีที่หาคำตอบไม่พบในรูปที่ 8.12 (ข) ขนาดโดยเฉลี่ยของชิ้นส่วนมีค่าคงที่ซึ่งต่างจากปัญหา Royal Road ในรูปที่ 8.6 (ข) ที่ขนาดโดยเฉลี่ยของชิ้นส่วนมีการเพิ่มขึ้นทีละน้อย เป็นเพราะว่าในปัญหา Trap-5 มีการแข่งขันกันระหว่างโครงสร้างที่เป็นคำตอบ กับโครงสร้างหลงอย่างชัดเจน จึงทำให้โครงสร้างที่ได้ส่วนใหญ่ถ้าไม่ลู่เข้าสู่โครงสร้างที่เป็นคำตอบ ก็จะลู่เข้าสู่โครงสร้างหลงอย่างชัดเจน พิจารณาได้จากรูปที่ 8.12 (ข) ปริมาณของ Non BBs ลดลงอย่างรวดเร็วและคงที่ ต่างจากปัญหา Royal Road ในรูปที่ 8.6 (ข) ที่ปริมาณของ Non BBs มีค่าสูงกว่ารูปแบบอื่นอยู่ตลอดเวลา ในเรื่องของปริมาณชิ้นส่วนจากรูปที่ 8.12 (ข-ช) มีอัตราการลดลงสอดคล้องกับขนาดโดยเฉลี่ยของชิ้นส่วนในรูปที่ 8.2 (จ-ค) และจะสังเกตเห็นได้ว่าในกรณีที่หาคำตอบไม่พบในรูป 8.12 (ข) ตั้งแต่รุ่นประชากรที่ 21 เป็นต้นไป ชิ้นส่วนเหมือนกันมีปริมาณมากกว่ารูปแบบชิ้นส่วนที่ต่างกัน เป็นเพราะว่าเริ่มเกิดการลู่เข้าสู่ค่าเหมาะสมที่สุดท้องถิ่น ทำให้ชิ้นส่วนมีขนาดยาวขึ้นเนื่องจากโครโมโซมมีความแตกต่างกันน้อย ส่งผลให้มีจำนวนชิ้นส่วนเหมือนกันมากขึ้น



รูปที่ 8.12 ผลการทดสอบของ CFT ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นของประชากร (ง) (จ) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (ค) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

รูปที่ 8.13 แสดงให้เห็นถึงปริมาณชิ้นส่วนทั้งหมดที่มีในหน่วยเก็บถาวร เมื่อเปรียบเทียบกับปัญหา Royal Road ในรูปที่ 8.7 จะเห็นว่าในปัญหา Trap-5 มีอัตราการลดลงที่สูงกว่า เป็นเพราะการลวงในปัญหา Trap-5 ส่งผลให้ความหลากหลายลดลงอย่างรวดเร็ว ต่างจากปัญหา Royal Road ที่ไม่มีการลวงและมีค่าความเหมาะสมเป็นแบบขั้นบันได และสังเกตได้จากปริมาณชิ้นส่วนทั้งหมดที่ระบุได้จากรูปที่ 8.12 (ซ-ซ) มีความแตกต่างจากปัญหา Royal Road ในรูปที่ 8.6 (ซ-ซ) เป็นอย่างมาก

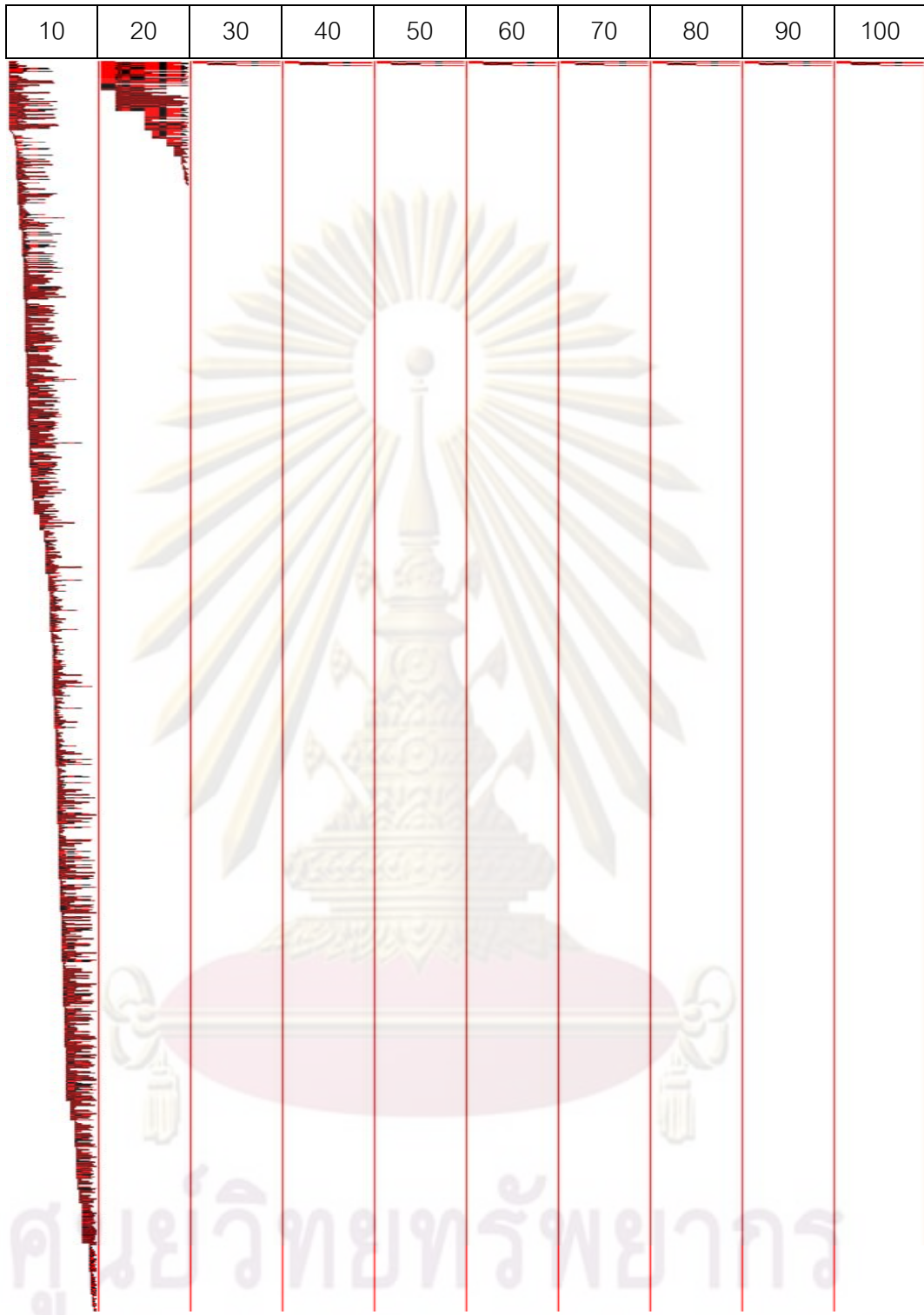


รูปที่ 8.13 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ CFT ในปัญหา Trap-5 60 บิต

ถึงแม้ว่าปัญหา Trap-5 จะมีการลวงให้เกิดการแข่งขันกันระหว่างโครงสร้างของหน่วยการสร้างที่ดี และหน่วยการสร้างลวง แต่วิธี CFT ยังคงมีความหลากหลายในการดำเนินการเหมือนกับปัญหา Royal Road ยืนยันได้จากลักษณะรูปแบบของชิ้นส่วนที่ระบุได้และเก็บอยู่ในหน่วยเก็บถาวรดังรูปที่ 8.14 เป็นกรณีที่หาคำตอบพบ ชิ้นส่วนยังคงมีการกระจายตัวอยู่ทุกตำแหน่งในปริมาณที่ใกล้เคียงกัน และในช่วงรุ่นประชากรท้าย ๆ การลู่เข้าเกิดขึ้นอย่างต่อเนื่องจนกระทั่งเหลือจำนวนชิ้นส่วนอยู่น้อยมากในรุ่นประชากรที่ 10 ซึ่งเป็นยืนยันให้เห็นว่าปัญหา Trap-5 ทำให้เกิดการลู่เข้าสู่คำตอบในระดับที่รุนแรงกว่าปัญหา Royal Road อย่างชัดเจน และจะเห็นได้ชัดขึ้นในกรณีที่หาคำตอบไม่พบดังรูปที่ 8.15 ตั้งแต่รุ่นประชากรที่ 20 เป็นต้นไปโครงสร้างของชิ้นส่วนจะลู่เข้าสู่รูปแบบหนึ่งอย่างชัดเจนต่างจากปัญหา Royal Road ในรูปที่ 8.9 เพื่อให้มีความชัดเจนมากขึ้นสามารถพิจารณาจากภาพแสดงโครงสร้างทางพันธุกรรมโดยรวมของประชากรดังรูปที่ 8.16 และรูปที่ 8.17 ซึ่งเป็นประชากรที่สร้างจากหน่วยเก็บถาวรในรูปที่ 8.14 และรูปที่ 8.15 ตามลำดับ

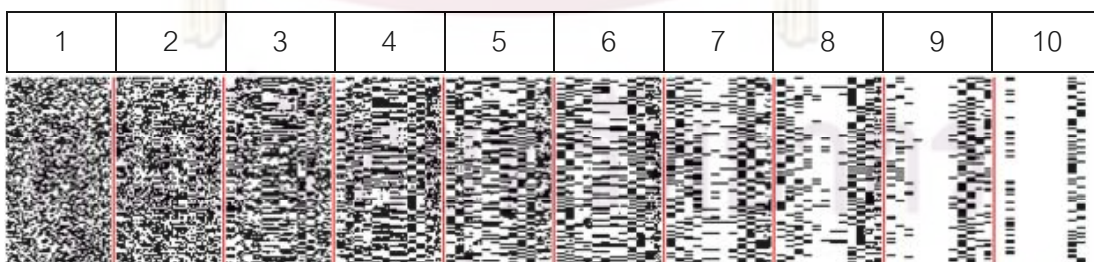


รูปที่ 8.14 รูปแบบจีโอไนท์ของชั้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Trap-5 60 บิต
 กรณีนี้อาศัยค่าตอบพบ และใช้จำนวนรูนประชากร 10 รูน ส่วนตัวเลขกำกับด้านบนในแต่ละวง
 หมายถึงรูนของประชากร เส้นขีดแบ่งรูนประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละ
 ละครูน พื้นที่สีเทา คือคูจิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคูจิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่
 ในบทที่ 4)

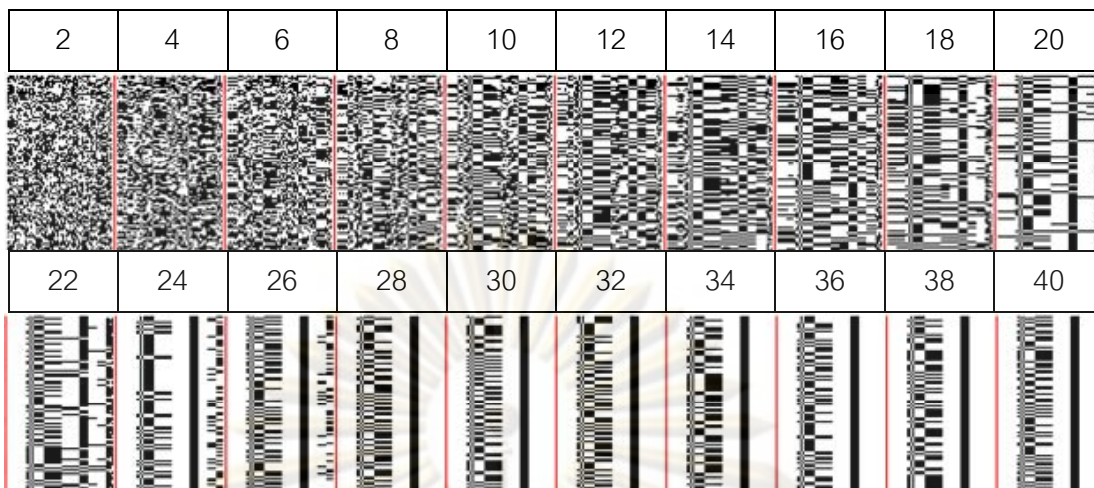


รูปที่ 8.15 รูปแบบจีโอไทป์ของชั้นส่วนที่อยู่ในหน่วยเก็บถาวรของ CFT ในปัญหา Trap-5 60 บิต
 กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 10 รุ่น ส่วนตัวเลข
 กำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอก
 ขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่
 จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)

จากภาพแสดงโครงสร้างทางพันธุกรรมของประชากรในรูปที่ 8.16 และรูปที่ 8.17 ในโครงสร้างของแต่ละโครโมโซมเมื่อผ่านไปช่วงเวลาหนึ่ง จะเริ่มมีรูปแบบที่เป็นริ้วขาวหรือดำชัดเจนขึ้น (ในแนวนอน) ริ้วดังกล่าวนี้สะท้อนให้เห็นถึงโครงสร้างของหน่วยการสร้าง โดยที่ริ้วสีดำเป็นหน่วยการสร้างลวง (Non BBs) และริ้วสีขาวเป็นหน่วยการสร้างของคำตอบ (Pure BBs) การเกิดริ้วขาวหรือดำเหล่านี้แสดงถึงการลู่เข้าที่เกิดขึ้นในระดับโครงสร้างย่อยของโครโมโซมเนื่องจากการลวงในปัญหา Trap-5 ซึ่งทำให้เกิดการสูญเสียความหลากหลายในระดับที่สูงกว่าปัญหา Royal Road จึงทำให้เกิดการลู่เข้าที่รุนแรงกว่าอย่างชัดเจน ในกรณีที่หาคำตอบพบจากรูปที่ 8.16 เริ่มมีโครงสร้างที่เป็นระเบียบมากขึ้นในรุ่นประชากรที่ 3 และชัดเจนขึ้นในรุ่นประชากรที่ 5 สอดคล้องกับรูป 8.12 ที่เกิดการแข่งขันกันระหว่าง Pure BBs กับ Non BBs อย่างเห็นได้ชัด (ส่วน Mixed BBs ลดจำนวนลงอย่างรวดเร็ว) และในที่สุดก็สามารถแยกแยะระหว่างหน่วยการสร้างลวง และหน่วยการสร้างที่เป็นคำตอบได้ ส่วนกรณีที่หาคำตอบไม่พบจากรูปที่ 8.17 โครงสร้างทางพันธุกรรมของโครโมโซมเกิดการเปลี่ยนแปลงช้ากว่า (กรณีที่หาคำตอบพบ) เนื่องจากใช้จำนวนประชากรน้อยทำให้ไม่สามารถแบ่งแยกโครงสร้างที่เป็นคำตอบกับโครงสร้างลวงได้อย่างชัดเจน การลู่เข้าเริ่มเกิดขึ้นอย่างชัดเจนในรุ่นประชากรที่ 28 เป็นต้นไป สอดคล้องกับรูปที่ 8.12 (ข และ ง) แต่อาจเกิดความสงสัยว่าทำไมรูปที่ 8.12 (ง) จำนวนของหน่วยการสร้างยังคงต่างกันอยู่ระดับหนึ่งแม้ว่าเกิดการลู่เข้าแล้วก็ตาม เป็นเพราะว่าวิธี CFT มีการใช้หน่วยเก็บถาวรร่วมกัน และประชากรทั้งหมดถูกสร้างจากหน่วยเก็บถาวรเดียวกัน ดังนั้นแม้ว่าจะมีชิ้นส่วนเหลืออยู่เป็นจำนวนน้อย แต่ทุกชิ้นส่วนก็มีโอกาสถูกเลือกมาสร้างเป็นโครโมโซม ทำให้รูปแบบชิ้นส่วนที่เหลืออยู่ปรากฏให้เห็นในโครงสร้างของประชากรในภาพรวม ส่งผลให้มีความหลากหลายหลงเหลืออยู่ในระดับหนึ่ง (วิธีการประกอบชิ้นส่วนเป็นรูปแบบปัญหาการจัดหมู่ จึงทำให้มีรูปแบบที่เป็นไปได้จำนวนมาก ส่งผลให้ความหลากหลายของโครงสร้างทางพันธุกรรมสูง)



รูปที่ 8.16 รูปแบบจีโนไทป์ของประชากรของ CFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่องหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จิ้นที่มีค่าเป็น 0



รูปที่ 8.17 รูปแบบจีโอเมทรีของประชากรของ CFT ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากร ในแต่ละรุ่น พื้นที่สีขาว □ คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จิ้นที่มีค่าเป็น 0

8.3 สรุปผลการทดลอง

ในบทนี้นำเสนอขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข ซึ่งมีการใช้หลักสารสนเทศร่วมต่ำสุดในวิธีการระบุชิ้นส่วน และวิธีการประกอบเข้าด้วยกันโดยใช้หน่วยเก็บถาวรร่วมกันในการแบ่งปันความรู้แบบรวมศูนย์ ผลการทดสอบแสดงให้เห็นว่าวิธี CFT มีประสิทธิภาพดีกว่าวิธี RFT ในทุกปัญหาเป็นการยืนยันสมมติฐานที่ว่าถ้าทำการระบุชิ้นส่วนได้ดีขึ้นจะทำให้ได้ผลที่ดีขึ้นด้วย ซึ่งเป็นการตอกย้ำให้เห็นถึงความสำคัญของกระบวนการระบุหน่วยการสร้างที่ไม่ปรากฏให้เห็นในขั้นตอนวิธีการอื่นทั่วไป จะมีเฉพาะกลุ่ม EDAs เท่านั้นที่ปรากฏอยู่ในรูปแบบจำลองความน่าจะเป็น และเมื่อเทียบกับวิธี sGA-FC ที่มีการใช้หลักการระบุหน่วยการสร้างแบบเดียวกัน จะเห็นว่ามีประสิทธิภาพเทียบเท่ากับวิธี sGA-FC ในปัญหา Royal Road และดีกว่าวิธี sGA-FC ในปัญหา Trap-5 จากผลที่เทียบเท่าหรือดีกว่าทำให้มั่นใจมากขึ้นว่าวิธีการแบ่งปันความรู้แบบรวมศูนย์ให้ผลดีกว่าวิธีการแบ่งปันความรู้แบบกระจาย ซึ่งเป็นความรู้ที่ได้ในบทที่ 6 นอกจากนี้วิธี CFT ยังแสดงให้เห็นถึงความเป็นจริงของความเชื่อในสายงานวิจัยนี้ที่ว่าโครโมโซมที่ดี ๆ มีโครงสร้างที่ดี ๆ อยู่ร่วมกัน เมื่อพิจารณาจากการเลือกกลุ่มประชากรย่อยที่จะเลือกเฉพาะโครโมโซมที่มีค่าความเหมาะสมสูงกว่าค่าเฉลี่ย (ถือว่าเป็นโครโมโซมที่ดี) และมีการเปรียบเทียบกันระหว่างทุกคู่ของโครโมโซมในกลุ่มประชากรย่อยที่ถูกเลือก โดยใช้หลักการสารสนเทศร่วมต่ำสุด ซึ่งเป็นการแสดงให้เห็นถึงโครงสร้างที่มีอยู่ร่วมกัน (โครงสร้างที่เหมือนกัน) ระหว่างคู่โครโมโซม

ผลที่ได้จากการระบุหน่วยการสร้างในบทนี้ทำให้เกิดความสงสัยต่อไปว่าจะมีวิธีอื่นที่สามารถใช้ในการระบุหน่วยการสร้างได้อย่างมีประสิทธิภาพอีกหรือไม่ จากความรู้ที่สั่งสมมาจากบทก่อนหน้าทั้งหมดทำให้สันนิษฐานได้ว่า มีอีกหลายวิธีที่สามารถนำมาประยุกต์ใช้ในระบบการระบุหน่วยการสร้างได้ ซึ่งจะแสดงให้เห็นต่อไปในงานวิจัยนี้



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA)

เนื้อหาของบทนี้เน้นเรื่องการประยุกต์ใช้หลักการบีบอัดข้อมูลในการระบุหน่วยการสร้างที่สอดคล้องกับรูปแบบชิ้นส่วน เพื่อแสดงให้เห็นถึงความเป็นไปได้ของหลักการที่นำเสนอในรูปแบบที่ต่างออกไปจำนวน 3 รูปแบบจาก 3 ขั้นตอนวิธีที่นำเสนอซึ่งมีความแตกต่างกันในเรื่องของระดับการแบ่งปันข้อมูล (ความรู้) ที่ใช้ในการบีบอัด และขั้นตอนวิธีที่นำเสนอทั้งหมดจะอาศัยตัวดำเนินการพื้นฐานเหมือนกับขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย (sGA) ผลที่ได้จากการทดลองจะช่วยให้เข้าใจถึงปัจจัยหรือกลไกที่มีผลต่อประสิทธิภาพในการหาคำตอบได้ดีขึ้นจากมุมมองที่แตกต่างกันออกไป

9.1 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW รูปแบบที่ 1 (LZWGA-1)

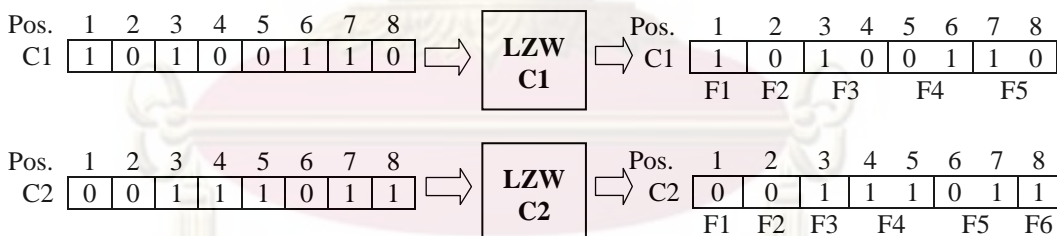
ส่วนนี้จะกล่าวถึงการประยุกต์ใช้หลักการบีบอัดข้อมูลกับโครโมโซมในรูปแบบที่ไม่มี การแบ่งปันข้อมูลระหว่างโครโมโซม โดยแยกการบีบอัดสำหรับแต่ละโครโมโซมเป็นอิสระจากกัน (ไม่บีบอัดร่วมกัน ดังนั้นหนึ่งโครโมโซมจะสร้างได้ 1 พจนานุกรม (1C-1D: 1 Chromosome – 1 Dictionary)) เพื่อต้องการศึกษาถึงผลกระทบจากการบีบอัดข้อมูลแบบ LZW กับการระบุชิ้นส่วน และการไขว้เปลี่ยนแบบจุดเดียวเปรียบเทียบผลทดสอบที่ได้กับ sGA

9.1.1 แนวคิดและหลักการ

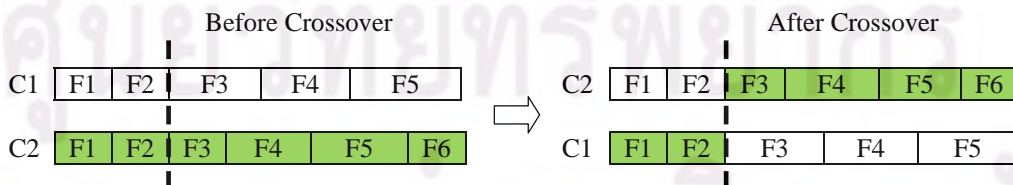
มีการประยุกต์ใช้การบีบอัดข้อมูลกับขั้นตอนวิธีเชิงพันธุกรรม (กุณาศล, สุวรรณิก และ จงสถิตย์วัฒนา, 2005) ทำให้เราเห็นภาพว่าการบีบอัดสามารถนำมาประยุกต์ใช้ได้จริง และสามารถแก้ปัญหาที่มีลักษณะคำตอบที่เป็นแบบแผน (มีรูปแบบซ้ำ ๆ กัน) ได้อย่างมีประสิทธิภาพ (แม้ปัญหาจะมีขนาดใหญ่มาก ๆ ก็ตาม) แต่วิธีการที่ใช้ส่วนใหญ่เป็นการสุ่มสร้างโครโมโซมจากรูปแบบที่มีการบีบอัดเรียบร้อยแล้ว (Compressed Chromosome) หลังจากนั้นจึงค่อยทำการคลายการบีบอัด (Decompress) เพื่อวัดค่าความเหมาะสม ซึ่งในหลายกรณีจะได้โครโมโซมที่สั้นหรือยาวกว่าขนาดโครโมโซมที่กำหนด (คลายการบีบอัดแล้วได้ขนาดเล็กลง หรือใหญ่ขึ้นกว่าขนาดที่กำหนด) ทำให้ต้องมีการเติมเต็มหรือการตัดปลายเกิดขึ้น ซึ่งอาจมองได้ว่าเป็นความไม่แน่นอนในการเข้ารหัสข้อมูลรูปแบบนี้ ด้วยเหตุนี้เราจึงได้มีแนวคิดที่จะประยุกต์ใช้หลักการบีบอัดในอีกรูปแบบหนึ่ง นั่นก็คือใช้การบีบอัดมาเป็นตัวช่วยระบุหน่วยการสร้างที่มีอยู่ในโครงสร้างโครโมโซมแทน ด้วยวิธีที่นำเสนอจะไม่กระทบกับขนาดของโครโมโซม ทำได้โดยการนำแต่ละโครโมโซมไปทำการบีบอัดด้วยวิธี LZW เพื่อทำการหารูปแบบชิ้นส่วนที่มีอยู่ในแต่ละโครโมโซม เหตุที่เลือกใช้

การบีบอัดแบบ LZW เนื่องจากเป็นวิธีการบีบอัดที่ให้ผลลัพธ์สอดคล้องกับนิยามชิ้นส่วนเป็นอย่างดี และเป็นวิธีที่ใช้กันอย่างแพร่หลายเพราะมีประสิทธิภาพในการบีบอัดที่ดี (ทั้งในแง่ของเวลาและผลการบีบอัดที่ได้) อีกทั้งยังสามารถเขียนโปรแกรมได้ง่าย (มีความซับซ้อนน้อย) เรากำหนดให้การบีบอัดแบบ LZW เริ่มดำเนินการจากคู่จิ้นตำแหน่งแรกไปจนถึงตำแหน่งสุดท้ายในแต่ละโครโมโซม (จากซ้ายไปขวา) ชิ้นส่วนที่มีอยู่ในโครโมโซม (ที่ระบุได้) เป็นรูปแบบเดียวกับคำ (รูปแบบโครงสร้างย่อยของโครโมโซม) ที่ปรากฏในพจนานุกรมที่ได้จากการบีบอัดนั่นเอง และทำซ้ำจนครบทุกโครโมโซมดังรูปที่ 9.1 เมื่อสามารถระบุชิ้นส่วนได้แล้วขั้นตอนต่อไปคือการไขว้เปลี่ยนชิ้นส่วนระหว่างสองโครโมโซมที่ถูกเลือกมา (เหมือนกับ sGA) แต่การไขว้เปลี่ยนสำหรับชิ้นส่วนเรามีกฎอยู่ว่าห้ามทำการแยกคู่จิ้นในชิ้นส่วนเดียวกันออกจากกัน (เหมือนกับการไขว้เปลี่ยนแบบชิ้นส่วน) ดังนั้นจุดตัดที่สามารถทำการไขว้เปลี่ยนได้จะมีจำนวนจำกัด (เฉพาะจุดรอยต่อของชิ้นส่วนที่ตำแหน่งตรงกัน) จากรูปที่ 9.1 โครโมโซม C1 และ C2 จะถูกบีบอัดแยกจากกัน (รูปส่วนบน) และหลังจากระบุชิ้นส่วนได้แล้วในขั้นตอนการไขว้เปลี่ยนจะเห็นว่าจุดตัดที่เป็นไปได้คือ คู่จิ้นตำแหน่งที่ 1 และ 2 เท่านั้น (ที่มีตำแหน่งตรงกัน) ชิ้นส่วนที่เหลือจะมีตำแหน่งเหลื่อมกันทั้งหมด (ถ้าตัดที่ตำแหน่งเหลื่อมกันเหล่านี้จะทำให้ชิ้นส่วนในโครโมโซมใดโครโมโซมหนึ่งถูกแยกออกจากกัน) จากรูป (ส่วนล่าง) เราสมมติว่าสุ่มเลือกจุดตัดได้เป็นตำแหน่งที่ 2 ดังนั้นคู่จิ้นที่อยู่ถัดจากตำแหน่งที่สองทั้งหมดจะถูกไขว้เปลี่ยนกันระหว่าง C1 และ C2

LZW Fragment Identification



1-Point Crossover



รูปที่ 9.1 การระบุชิ้นส่วนของ LZWGA-1 และการไขว้เปลี่ยนแบบจุดเดียว (C หมายถึง Chromosome และ F หมายถึง Fragment)

จากวิธีที่นำเสนอการใช้หลักการบีบอัดแบบ LZW ระบุชิ้นส่วนเป็นวิธีแบบปรับตัว (Adaptive) เพราะสามารถปรับให้เข้ากับโครงสร้างทางพันธุกรรมที่เปลี่ยนไปได้ การปรับตัวนี้ส่งผลโดยตรงต่อการไขว้เปลี่ยนจากเดิมใน sGA ที่ทุกตำแหน่งมีโอกาสถูกเลือกเป็นจุดตัดเท่ากันหมด กลายเป็นเหลือแค่บางตำแหน่งเท่านั้นที่สามารถทำการไขว้เปลี่ยนได้ (ตำแหน่งที่เป็นรอยต่อของชิ้นส่วน) การบีบอัดแบบ LZW โดยปกติจะส่งผลให้ชิ้นส่วนมีขนาดใหญ่ขึ้นเรื่อย ๆ เนื่องจากใช้การอ้างอิงจากรูปแบบที่เคยเกิดขึ้นแล้ว ผลที่ได้นี้จะเห็นชัดขึ้นในกรณีที่โครโมโซมมีความยาวมากขึ้น อัตราการเพิ่มของจำนวนชิ้นส่วนจะลดลงเมื่อเทียบกับขนาดความยาวโครโมโซมที่เพิ่มขึ้น ดังนั้นตำแหน่งท้าย ๆ ของโครโมโซมมีโอกาสเกิดจุดตัดขึ้นน้อยลงซึ่งส่งผลโดยตรงต่อการไขว้เปลี่ยนเพราะจำนวนจุดตัดที่ตรงกันระหว่างสองโครโมโซมใด ๆ จะเกิดขึ้นน้อยลงด้วย (เมื่อเทียบกับตำแหน่งแรก ๆ) อย่างไรก็ตามการบีบอัดที่แต่ละโครโมโซมไม่ขึ้นแก่กันนี้จะมีจุดตัดที่ตรงกันระหว่างสองโครโมโซมใด ๆ เสมอ เนื่องจากวิธีการบีบอัดแบบ LZW ในตอนเริ่มต้นจะทำการเปรียบเทียบคู่จิ้นในโครโมโซมกับค่าพื้นฐานในพจนานุกรมนั้นก็คือรูปแบบคู่จิ้นพื้นฐานที่เป็นไปได้ (การเข้ารหัสเลขฐานสองรูปแบบคู่จิ้นที่เป็นไปได้คือ 0 กับ 1) ดังนั้นตำแหน่งของคู่จิ้นจะตรงกับรูปแบบพื้นฐานในพจนานุกรมเสมอ (เป็นไปได้แค่สองค่าคือไม่ 0 ก็ 1) หรือกล่าวได้ว่าโครโมโซมใด ๆ จะมีชิ้นส่วนในตำแหน่งแรกขนาดหนึ่งบิตเท่ากันเสมอ (มีจุดตัดถัดจากตำแหน่งแรกเสมอ)

P_1, P_2 คือ ผลเฉลยที่ผ่านการคัดเลือกมาแล้ว และผ่านกระบวนการบีบอัดแบบ

LZW เพื่อหาค่าดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้แล้ว

S คือ ดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้โดยใช้วิธี LZW

ขั้นตอนวิธีการไขว้เปลี่ยนแบบจุดเดียว (One-point Crossover)

ถ้าดำเนินการไขว้เปลี่ยน ทำ

$P_1, P_2 \leftarrow$ ทำการไขว้เปลี่ยน P_1, P_2 กันแบบ sGA โดยเลือกตำแหน่งไขว้เปลี่ยนที่ตรงกันระหว่างผลเฉลยจาก S_{P_1} กับ S_{P_2} ถ้าไม่มีตำแหน่งตรงกันเลยให้ใช้ตำแหน่งที่สุ่มได้จากผลเฉลยตัวใดตัวหนึ่ง

จบเงื่อนไข

จบขั้นตอนวิธี

รูปที่ 9.2 รหัสเทียมของวิธีการไขว้เปลี่ยนแบบจุดเดียวสำหรับ LZWGA-1

ในภาพรวมเมื่อใช้หลักการบีบอัดแบบ LZW โครโมโซมจะถูกไขว้เปลี่ยนในตำแหน่ง ต้น ๆ มากกว่าตำแหน่งท้าย ๆ ซึ่งอาจมองได้ว่าเป็นการขัดขวางการสำรวจ และมีอัตราการ แสวงหาประโยชน์ที่สูง (ชิ้นส่วนมีขนาดใหญ่ขึ้นเรื่อย ๆ) ทำให้ผลที่ได้ไม่น่าจะดีเมื่อเทียบกับ sGA แต่ประเด็นสำคัญของการทดลองนี้คือ เราต้องการทดสอบความเป็นไปได้ของการระบุชิ้นส่วนใน รูปแบบปรับตัวแบบที่ไม่มีการแบ่งปันความรู้ร่วมกัน ซึ่งแตกต่างออกไปจากการไขว้เปลี่ยนแบบ ชิ้นส่วน (บทที่ 5) ว่ายังมีความสามารถในการหาคำตอบได้ตามเงื่อนไขที่กำหนดหรือไม่ แม้ว่าการ ระบุชิ้นส่วนจะมีข้อจำกัดเพิ่มขึ้นมากก็ตาม หรือมองได้ว่าเป็นการเปรียบเทียบวิธีการไขว้เปลี่ยน แบบจุดเดียวที่มีอิสระในการเลือกจุดตัด กับการไขว้เปลี่ยนแบบจุดเดียวที่อยู่บนพื้นฐานของ ชิ้นส่วน ว่ามีผลกระทบกับการหาคำตอบแตกต่างกันอย่างไร

D คือ ประชากร

S คือ ดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้โดยใช้วิธี LZW

P คือ ผลเฉลยที่ได้จากวิธีการเลือก

ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA-1) ชนิด 1C-1D

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$S_g \leftarrow$ นำผลเฉลยแต่ละตัวจาก D_{g-1} ไปทำการบีบอัดแบบ LZW โดยใช้ พจนานุกรมแยกกัน แล้วเก็บค่าดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่ บีบอัดได้สำหรับแต่ละโครโมโซม (เก็บแยกกัน)

วนซ้ำสำหรับ $i = 1, 3, 7, \dots, R-1$ {จนกระทั่งครบ R ตัว}

$P_{i,i+1} \leftarrow$ เลือกผลเฉลย 2 ตัว (ตัวที่ i กับ $i+1$) จาก D_{g-1} ที่สอดคล้อง กับวิธีการเลือก

$P_{i,i+1} \leftarrow$ ทำการไขว้เปลี่ยนกันแบบ sGA โดยเลือกตำแหน่งไขว้เปลี่ยน ที่ตรงกันระหว่างผลเฉลยจาก S_g^i กับ S_g^{i+1} ถ้าไม่มีตำแหน่ง ตรงกันเลยให้ใช้ตำแหน่งที่สุ่มได้จากผลเฉลยตัวใดตัวหนึ่ง

$D_g \leftarrow$ นำ $P_{i,i+1}$ ไปทำการผ่าเหล่า (เป็นทางเลือก แต่งานวิจัยนี้ไม่ใช้)

จบการวนซ้ำ

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 9.3 รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-1

9.1.2 การทดลองและผลการทดลอง

ในส่วนนี้จะกล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่งในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่น่าเสนอ และวิจารณ์ผลการทดลองในมุมมองที่น่าเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม (sGA)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การไขว้เปลี่ยน:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่น่าเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWGA-1)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	ใช้หลักการบีบอัดแบบ LZW โดยใช้พจนานุกรมแยกกันสำหรับแต่ละโครโมโซม
การประกอบ:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 9.1

จากผลการทดลองในตารางที่ 9.1 เป็นการเปรียบเทียบกับ sGA ซึ่งมีหลักการดำเนินการพื้นฐานเหมือนวิธี LZWGA-1 ต่างกันแค่เพียงวิธีการเลือกจุดตัดในการไขว้เปลี่ยน โดยวิธี sGA สามารถเลือกจุดตัดได้อิสระ แต่วิธี LZWGA-1 จุดตัดที่เลือกได้จะต้องเป็นตำแหน่งที่เป็นรอยต่อระหว่างชิ้นส่วนที่ตรงกันระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น ในปัญหา Royal Road วิธี LZWGA-1 ใช้จำนวน FEs สูงกว่า sGA แสดงให้เห็นว่าวิธี LZWGA-1 มีประสิทธิภาพในการแก้ปัญหาต่ำกว่าวิธี sGA และในปัญหา Trap-5 วิธี LZWGA-1 มีประสิทธิภาพสูงกว่า sGA เนื่องจากใช้จำนวน FEs น้อยกว่า sGA ถ้าพิจารณาในแง่ของระดับความหลากหลายที่ต้องการเพื่อใช้ในการแก้ปัญหา วิธี LZWGA-1 ใช้จำนวนประชากรมากกว่า sGA ในปัญหา Royal Road และใช้จำนวนประชากรใกล้เคียงกับ sGA ในปัญหา Trap-5 ซึ่งบ่งชี้ว่าวิธี LZWGA-1 ต้องการ

ระดับความหลากหลายสูงกว่า sGA ในการแก้ปัญหา Royal Road และต้องการระดับความหลากหลายใกล้เคียงกับ sGA ในการแก้ปัญหา Trap-5 เมื่อพิจารณาถึงกรณีที่หาคำตอบไม่พบ จากทั้งสองปัญหาจะเห็นว่าทั้งสองวิธีใช้จำนวนประชากรเท่ากัน หรือกล่าวได้ว่าวิธี LZWGA-1 ต้องการระดับความหลากหลายต่ำสุดเท่ากับ sGA ในการแก้ปัญหา

ตารางที่ 9.1 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWGA-1

Parameters		Success 30 run				Fail 30 run	
		sGA		LZWGA-1		sGA	LZWGA-1
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,200	11,920	1,800	18,540	200	200
Trap-5	60	2,300	28,367	2,200	25,740	300	300

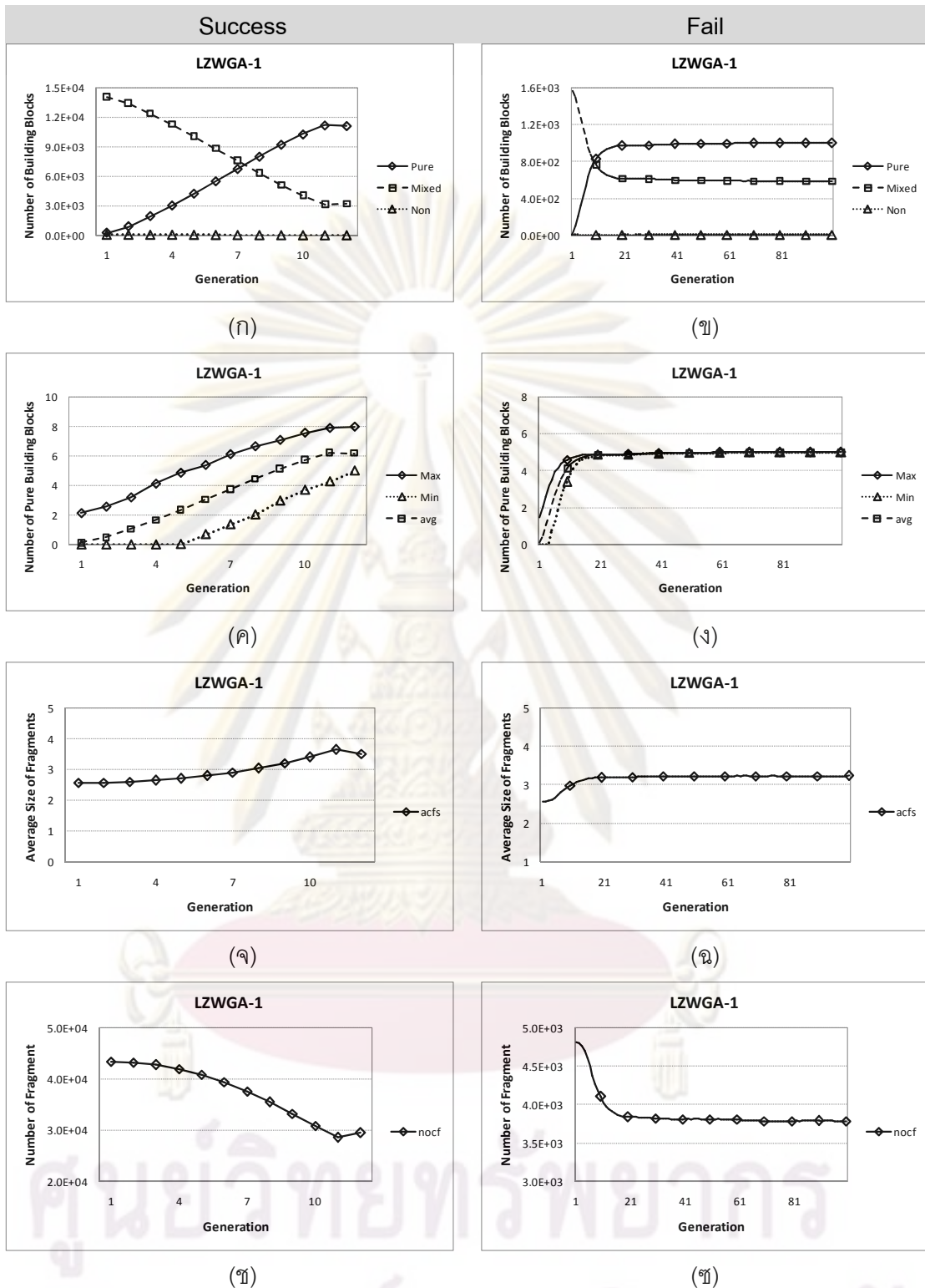
หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ผลการทดสอบกับปัญหา Royal Road ในมุมมองของหน่วยการสร้างในรูปที่ 9.4 (ก-ง) มีรูปแบบเหมือนกับผลที่ได้จากวิธี sGA ดังรูปที่ 5.3 ในบทที่ 5 ยังคงมีการแข่งขันกันเฉพาะ Mixed BBs และ Pure BBs เท่านั้นเนื่องจากไม่ใช่ปัญหาหลวง ในกรณีที่หาคำตอบพบในรูปที่ 9.4 (ก) ปริมาณของ Pure BBs เพิ่มขึ้นอย่างต่อเนื่อง ทำให้ปริมาณของ Mixed BBs ลดลงอย่างต่อเนื่องและตัดกันที่รุ่นประชากรที่ 7 และปริมาณของ Pure BBs ยังคงเพิ่มขึ้นจนกระทั่งพบคำตอบ แสดงให้เห็นว่ามีความหลากหลายเพียงพอในการค้นหาหน่วยการสร้างพิจารณาได้จากรูปที่ 9.4 (ค) จำนวนของ Max Pure BBs และ Min Pure BBs มีค่าแตกต่างกันอย่างสม่ำเสมอในทุกรุ่นประชากร (เส้นกราฟขนานกัน) ส่วนกรณีที่หาคำตอบไม่พบในรูปที่ 9.4 (ข) Pure BBs ยังคงเพิ่มขึ้นจนมีค่าสูงกว่า Mixed BBs และมีค่าคงที่เนื่องจากเกิดการลู่เข้าก่อนที่จะพบคำตอบ เป็นเพราะความหลากหลายลดลงอย่างรวดเร็วพิจารณาได้จากรูปที่ 9.4 (ง) ที่จำนวน Max Pure BBs และ Min Pure BBs เริ่มคงที่ตั้งแต่รุ่นประชากรที่ 21 สอดคล้องกับรูปที่ 9.4 (ข)

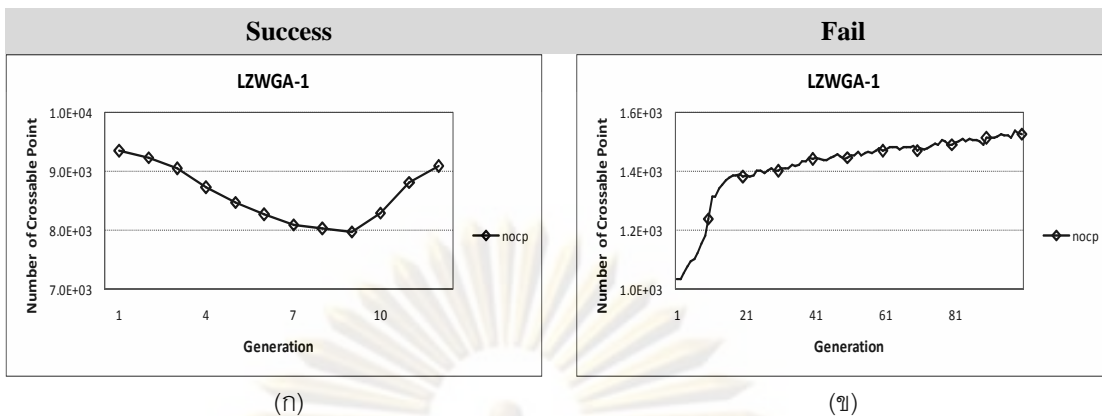
การพิจารณาในมุมมองของชิ้นส่วนจากรูป 9.4 (จ-ฉ) ขนาดโดยเฉลี่ยของชิ้นส่วนจะเพิ่มขึ้นทีละน้อย และขนาดโตสุดมีค่าไม่เกิน 4 บิต ซึ่งถือว่าเป็นชิ้นส่วนขนาดเล็ก เป็นเพราะวิธี LZWGA-1 ใช้การบีบอัดที่ละโครโมโซมแยกจากกันทำให้มีข้อมูลในการบีบอัดน้อย (ในการบีบอัดแต่ละครั้งมีข้อมูลเท่ากับความยาวของโครโมโซมเพียงตัวเดียว) ส่งผลให้ชิ้นส่วนมีขนาดสั้น ใน

กรณีที่ทำคำตอบพบจากรูปที่ 9.4 (จ) เมื่อเทียบกับกรณีที่ทำคำตอบไม่พบในรูปที่ 9.4 (ข) จะเห็นว่า มีรูปแบบต่างกันตรงที่กรณีที่ทำคำตอบพบความยาวโดยเฉลี่ยของชิ้นส่วนจะค่อย ๆ เพิ่มขึ้นทีละน้อยอย่างต่อเนื่องจนพบคำตอบ เนื่องจากมีระดับความหลากหลายสูง พิจารณาได้จากปริมาณของชิ้นส่วนที่ระบุได้จากรูปที่ 9.4 (ข) ต่างจากกรณีที่ทำคำตอบไม่พบขนาดความยาวโดยเฉลี่ยของชิ้นส่วนจะค่อย ๆ เพิ่มขึ้นและคงที่ เนื่องจากเกิดการลู่เข้าที่เป็นผลจากการใช้จำนวนประชากรน้อยทำให้ปริมาณของชิ้นส่วนที่ระบุได้มีจำนวนน้อยและขาดความหลากหลาย พิจารณาได้จากรูปที่ 9.4 (ข) ในรุ่นประชากรสุดท้ายของรูปที่ 9.4 (จ และ ข) จะสังเกตเห็นว่ามีค่าขัดแย้งกับผลจากรุ่นประชากรก่อน ๆ เป็นเพราะบางรอบการทดสอบใช้จำนวนรุ่นประชากรมากกว่ารอบอื่นทำให้ผลที่ได้ไม่ถูกต้องเหมือนกับรอบอื่น ๆ ส่งผลให้มีค่าที่แตกต่างอย่างชัดเจน

วิธี LZWGA ในบทนี้ใช้การไขว้เปลี่ยนแบบจุดเดียวในการประกอบชิ้นส่วนเข้าด้วยกัน แต่มีกฎในการไขว้เปลี่ยนเพิ่มขึ้นมาคือ จุดที่ไขว้เปลี่ยนจะต้องเป็นตำแหน่งรอยต่อของชิ้นส่วนเท่านั้นทำให้ตำแหน่งจุดตัดไม่เป็นอิสระเหมือนกับการไขว้เปลี่ยนในวิธี sGA วิธี LZWGA-1 จะทำการบีบอัดโครโมโซมแต่ละตัวแยกจากกัน ทำให้ได้ชิ้นส่วนจำนวนมากส่งผลให้มีจำนวนจุดตัดหลายจุดในแต่ละโครโมโซม เมื่อเลือกโครโมโซมมาสองตัวเพื่อทำการไขว้เปลี่ยนกันจะเลือกจุดไขว้เปลี่ยนจากเซตของจุดตัดที่มีตำแหน่งตรงกันระหว่างสองโครโมโซมเท่านั้น จำนวนจุดตัดที่เป็นไปได้ทั้งหมดระหว่างสองโครโมโซมเมื่อผ่านกระบวนการเลือกสรรแล้วแสดงดังรูปที่ 9.5 ในกรณีที่ทำคำตอบพบจากรูปที่ 9.5 (ก) ปริมาณจุดตัดที่ตรงกันระหว่างสองโครโมโซมลดลงอย่างต่อเนื่องจนถึงรุ่นประชากรที่ 9 เป็นเพราะเมื่อเริ่มเกิดการลู่เข้า ความหลากหลายของโครงสร้างทางพันธุกรรมจะลดลง ส่งผลให้เกิดการแบ่งกลุ่มระหว่างโครงสร้างที่ต่างกันออกเป็นกลุ่มประชากรย่อย ๆ หลายกลุ่ม ทำให้ผลที่ได้จากการบีบอัดแตกต่างกันตามแต่ละกลุ่มประชากรย่อย ๆ เหล่านี้ ดังนั้นในการไขว้เปลี่ยนถ้าเลือกได้โครโมโซมที่มาจากคนละกลุ่มกัน จะทำให้มีตำแหน่งจุดตัดที่ตรงกันน้อยลง และเมื่อการลู่เข้าเริ่มสูงขึ้นประชากรส่วนใหญ่เริ่มมีหน้าตาเหมือนกันมากขึ้น ส่งผลให้มีกลุ่มประชากรย่อยที่แตกต่างกันน้อยลง ทำให้มีตำแหน่งจุดตัดที่ตรงกันมากขึ้นพิจารณาได้จากรุ่นประชากรที่ 10 เป็นต้นไป ส่วนกรณีที่คำตอบไม่พบในรูปที่ 9.5 (ข) ปริมาณตำแหน่งจุดตัดที่ตรงกันเพิ่มขึ้นอย่างต่อเนื่อง เป็นเพราะมีระดับความหลากหลายในกลุ่มประชากรน้อย (มีจำนวนประชากรน้อย) ทำให้การลู่เข้าเกิดขึ้นอย่างรวดเร็ว

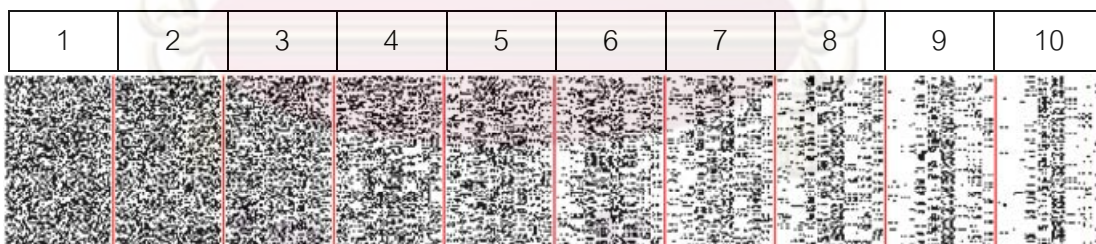


รูปที่ 9.4 ผลการทดสอบของ LZWGA-1 ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

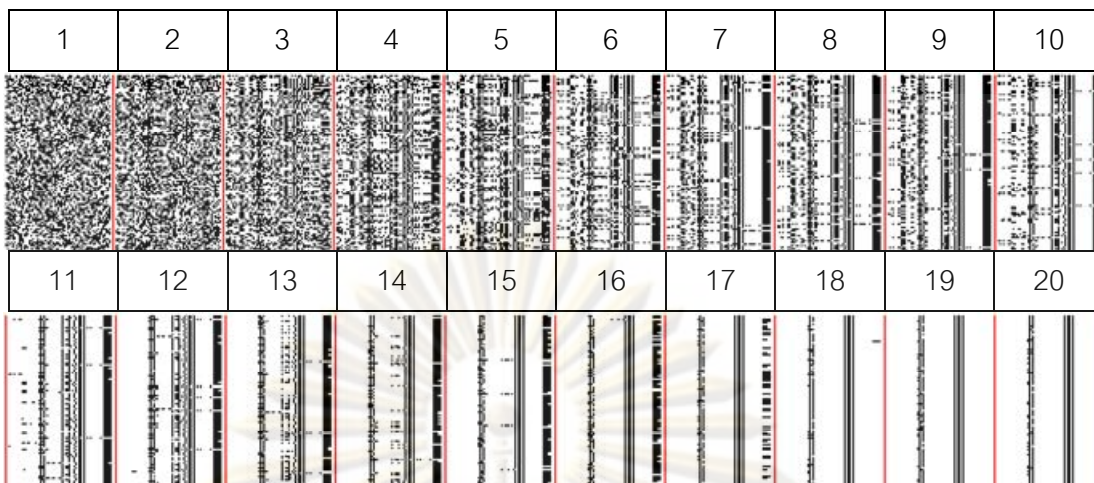


รูปที่ 9.5 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-1 ใน ปัญหา Royal Road 64 บิต

เมื่อพิจารณาโครงสร้างทางพันธุกรรมของประชากรในภาพรวมจากรูปที่ 9.6 และรูปที่ 9.7 (วิธีการอ่านรูปอยู่ในบทที่ 4) หลังจากดำเนินการผ่านไประยะหนึ่ง (ประมาณ 3 รุ่นประชากร) พื้นที่สีขาว (คู่จีนที่มีค่าเป็น 1) ค่อย ๆ เพิ่มขึ้นอย่างต่อเนื่องเป็นตัวแทนให้เห็นถึงปริมาณของ Pure BBs ที่มีอยู่ในโครงสร้างของประชากร การเพิ่มขึ้นของพื้นที่สีขาวทำให้ Mixed BBs หรือพื้นที่สีดำ (คู่จีนที่มีค่าเป็น 0) ลดลงเรื่อย ๆ (ในปัญหา Royal Road ไม่มีการลวงเกิดขึ้นดังนั้นพื้นที่สีดำจะไม่ถูกมองว่าเป็น Non BBs) กรณีที่หาคำตอบพบ (รูปที่ 9.6) ต่างจากกรณีที่ไม่หาคำตอบไม่พบ (รูปที่ 9.7) ตรงที่กรณีหาคำตอบพบจะสามารถหาคำตอบเจอก่อนที่จะเกิดการลู่เข้า เนื่องจากมีระดับความหลากหลายเพียงพอจึงทำให้การลู่เข้าเกิดขึ้นช้า สำหรับวิธี LZWGA-1 ระดับความหลากหลายขึ้นกับจำนวนประชากรเป็นหลัก



รูปที่ 9.6 รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่จีนที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จีนที่มีค่าเป็น 0



รูปที่ 9.7 รูปแบบจิ้นไทป์ของประชากรของ LZWGA-1 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู้จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู้จิ้นที่มีค่าเป็น 0

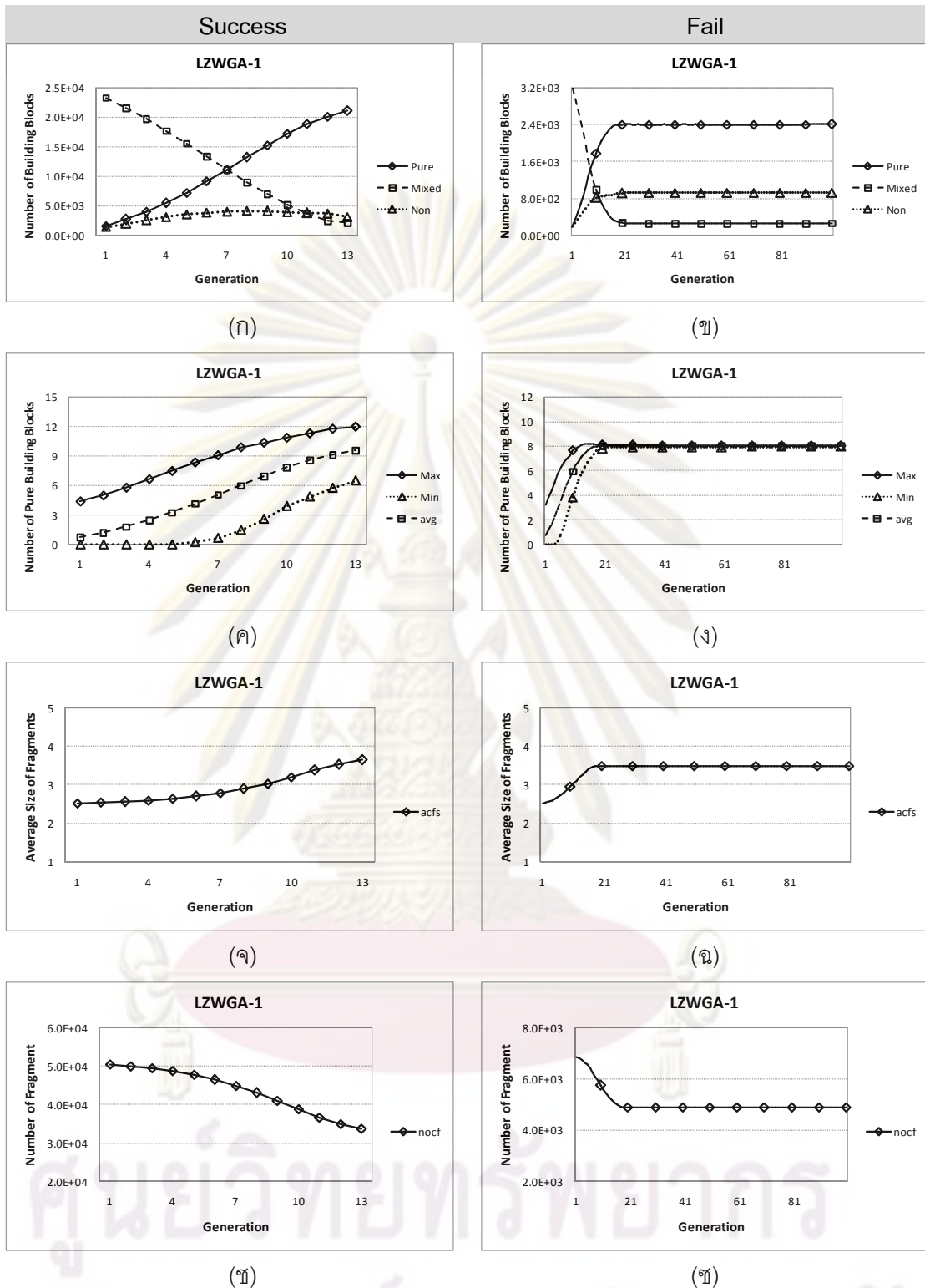
สำหรับปัญหา Trap-5 เป็นปัญหาที่มีการลวงในหน่วยการสร้าง ดังนั้นเมื่อพิจารณาในมุมมองของหน่วยการสร้างจากรูปที่ 9.7 (ก และ ข) นอกจากจะมีการแข่งขันกันระหว่าง Mixed BBs และ Pure BBs แล้วปริมาณของ Non BBs ยังเพิ่มขึ้นด้วย ในกรณีที่ทำคำตอบพบจากรูปที่ 9.7 (ก) การแข่งขันยังคงดำเนินต่อไปเรื่อย ๆ จนกระทั่งพบคำตอบ เป็นเพราะมีความหลากหลายเพียงพอในการดำเนินการ ทำให้มีโอกาสพบคำตอบที่ดีขึ้นสอดคล้องกับรูป 9.7 (ค) ที่จำนวน Max Pure BBs มีค่าต่างจาก Min Pure BBs ตลอดจนการดำเนินการ ส่วนกรณีที่ทำคำตอบไม่พบจากรูปที่ 9.7 (ข) ปริมาณของ Non BBs มีมากกว่า Mixed BBs เนื่องจากมีจำนวนประชากรน้อย (มีระดับความหลากหลายน้อย) ทำให้ไม่สามารถแยกแยะระหว่างหน่วยการสร้างที่เป็นคำตอบกับหน่วยการสร้างลวงได้ จึงเกิดการลู่เข้าในระดับโครงสร้างย่อยก่อนที่จะพบคำตอบในรุ่นประชากรที่ 20 ตรงกับรูปที่ 9.7 (ง) ลักษณะของผลในมุมมองของหน่วยการสร้างมีรูปแบบเหมือนกับวิธี sGA ที่ใช้ตัวดำเนินการแบบเดียวกันหมด ต่างกันตรงที่วิธี LZWGA-1 มีกฎสำหรับการเลือกตำแหน่งในการไขว้เปลี่ยนเพิ่มขึ้นมาเพียงอย่างเดียว

ในมุมมองของชิ้นส่วนจากรูปที่ 9.7 (จ-ข) ผลที่ได้มีรูปแบบโดยรวมเหมือนกับปัญหา Royal Road ในรูปที่ 9.4 ในกรณีที่ทำคำตอบไม่พบจากรูปที่ 9.7 (ข) จะสังเกตเห็นว่าขนาดโดยเฉลี่ยของชิ้นส่วนในปัญหา Trap-5 มีขนาดใหญ่กว่าปัญหา Royal Road (รูปที่ 9.4 (ข)) เป็นเพราะเกิดการลู่เข้าสู่โครงสร้างลวงทำให้ความหลากหลายของโครงสร้างคู้จิ้นในโครโมโซมลดลง

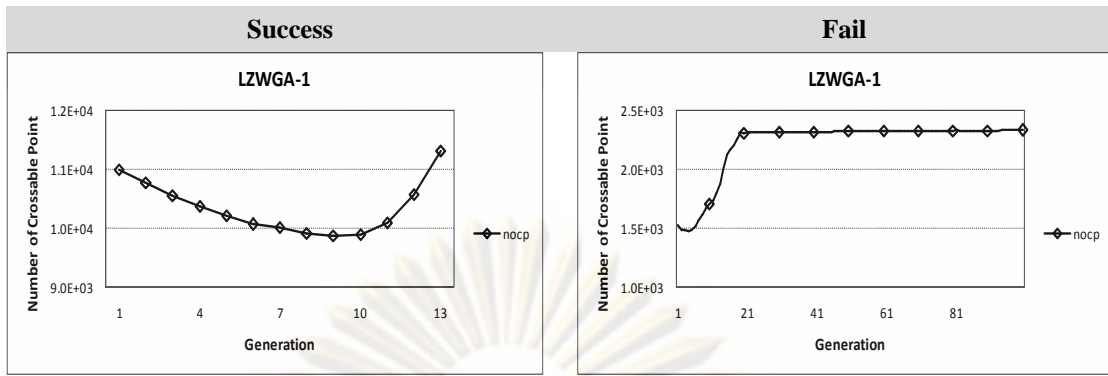
(โครงสร้างคู่อินมีระเบียบมากขึ้น) ดังนั้นเมื่อทำการบีบอัดแบบ LZW จะได้ชิ้นส่วนที่มีขนาดใหญ่ขึ้นโดยเฉลี่ย แต่อาจเกิดความสงสัยว่าทำไมในรูปที่ 9.7 (ข) จึงมีปริมาณชิ้นส่วนมากกว่าปัญหา Royal Road ในรูปที่ 9.4 (ข) ทั้งนี้เป็นเพราะในปัญหา Trap-5 กรณีที่หาคำตอบไม่พบใช้จำนวนประชากรมากกว่าปัญหา Royal Road นั่นเอง (Trap-5 ใช้จำนวนประชากร 300 ตัว ส่วน Royal Road ใช้จำนวนประชากร 200 ตัว)

ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่างสองโครโมโซมในปัญหา Trap-5 จากรูปที่ 9.9 มีลักษณะเหมือนกับปัญหา Royal Road ในรูปที่ 9.5 เนื่องจากเหตุผลเดียวกัน แต่ในกรณีที่หาคำตอบไม่พบจากรูปที่ 9.9 (ข) เมื่อเทียบกับรูปที่ 9.5 (ข) จะเห็นว่ามีความแตกต่างกัน โดยที่ปัญหา Trap-5 มีจำนวนจุดตัดระหว่างสองโครโมโซมในช่วงรุ่นประชากรแรก ๆ เกือบคงที่ เนื่องจากโครงสร้างทางพันธุกรรมของประชากรเข้าสู่โครงสร้างลงในตำแหน่งที่ต่างกัน ทำให้มีจำนวนจุดตัดเกือบคงที่ (ลดลงเล็กน้อย) และหลังจากนั้นจะเริ่มเข้าสู่รูปแบบเดียวกันเกือบหมด เพราะว่ามีระดับความหลากหลายต่ำร่วมกับการลงในปัญหา Trap-5 ทำให้เกิดการเข้าสู่ในระดับที่สูงกว่า (รุนแรงกว่า) ปัญหา Royal Road

เมื่อพิจารณาถึงโครงสร้างทางพันธุกรรมของประชากรในภาพรวมจากรูปที่ 9.10 และรูปที่ 9.11 (วิธีการอ่านรูปอยู่ในบทที่ 4) เมื่อดำเนินการผ่านไประยะหนึ่ง โครงสร้างทางพันธุกรรมของประชากรจะค่อย ๆ เริ่มเกาะกลุ่มกันในรูปของหน่วยการสร้างลง สังเกตได้จากริ้วสีดำในแนวนอน (พื้นที่สีดำคือคู่อินที่มีค่าเป็น 0) ซึ่งเริ่มปรากฏให้เห็นในรุ่นประชากรที่ 4 และชัดเจนขึ้นเรื่อย ๆ จนเห็นเป็นแถบสีดำในแนวตั้ง กรณีที่หาคำตอบพบจากรูปที่ 9.10 วิธี LZWGA-1 สามารถพบคำตอบก่อนที่จะเกิดการเข้าสู่อย่างสมบูรณ์ ซึ่งต่างจากกรณีที่หาคำตอบไม่พบในรูป 9.11 ที่เกิดการเข้าสู่เกือบสมบูรณ์ในระดับโครงสร้างย่อยในรุ่นประชากรที่ 8 เห็นได้จากแถบสีดำในช่วงตำแหน่งกลาง และตำแหน่งท้าย ๆ ในโครโมโซม



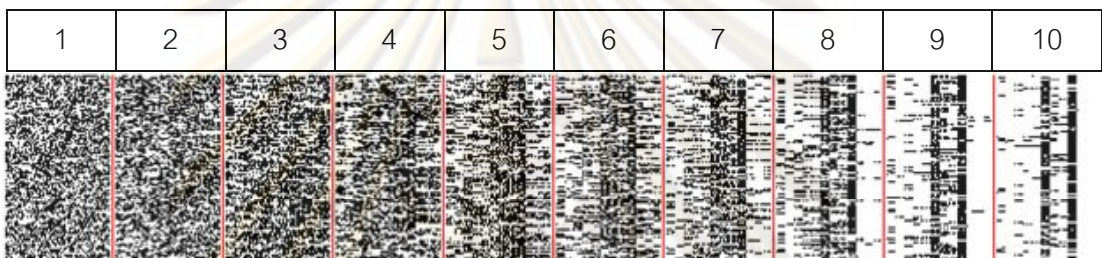
รูปที่ 9.8 ผลการทดสอบของ LZWGA-1 ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นของประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)



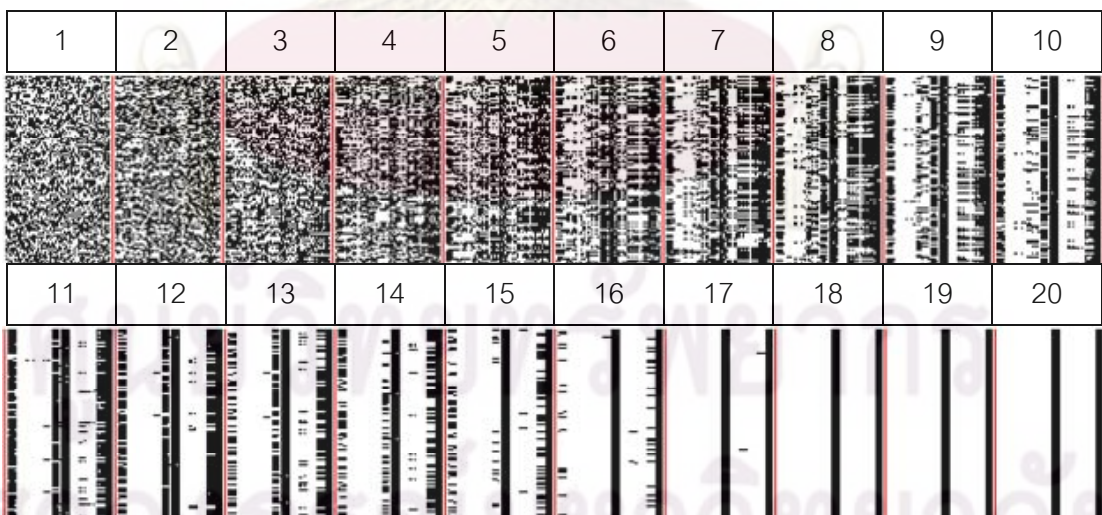
(ก)

(ข)

รูปที่ 9.9 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-1 ในปัญหา Trap-5 60 บิต



รูปที่ 9.10 รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0



รูปที่ 9.11 รูปแบบจีโนไทป์ของประชากรของ LZWGA-1 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0

9.1.3 สรุปผลการทดลอง

ในส่วนนี้นำเสนอขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-1 ที่ใช้หลักการบีบอัดข้อมูลแบบ LZW ในการระบุหน่วยการสร้างแบบไม่มีการแบ่งปันความรู้ระหว่างโครโมโซม (บีบอัดที่ละโครโมโซมแยกจากกัน) และใช้วิธีการไขว้เปลี่ยนแบบจุดเดียวในการประกอบหน่วยการสร้างเข้าด้วยกันโดย ผลการทดสอบเมื่อเทียบกับ sGA ในปัญหา Royal Road วิธี LZWGA-1 มีประสิทธิภาพในการแก้ปัญหาต่ำกว่า และในปัญหา Trap-5 วิธี LZWGA-1 มีประสิทธิภาพในการแก้ปัญหาดีกว่า เป็นการแสดงให้เห็นว่าหลักการบีบอัดแบบ LZW สามารถนำมาประยุกต์ใช้ในการระบุหน่วยการสร้างได้อย่างมีประสิทธิภาพ (เมื่อเปรียบเทียบกับ sGA) แม้ว่ามีข้อจำกัดในการไขว้เปลี่ยนสูงก็ตาม ซึ่งเป็นไปตามสมมติฐานที่ว่ามียีกหลายวิธีที่สามารถนำมาประยุกต์ใช้ในกระบวนการระบุหน่วยการสร้างได้ (ในส่วนสรุปของบทที่ 8)



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

9.2 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW รูปแบบที่ 2 (LZWGA-2)

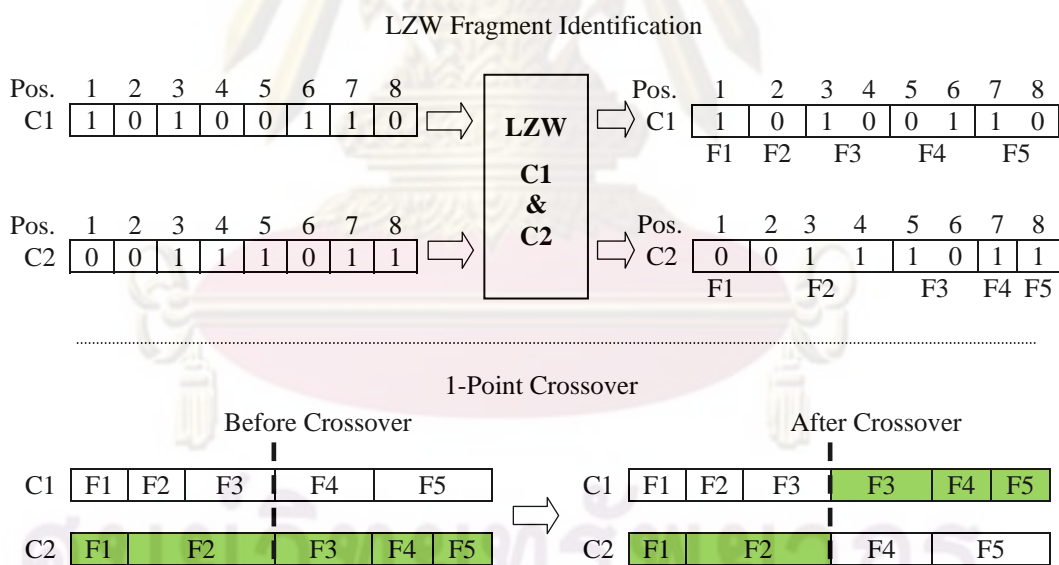
ส่วนนี้จะกล่าวถึงการประยุกต์ใช้หลักการบีบอัดข้อมูลกับโครโมโซมในรูปแบบที่มีการแบ่งปันข้อมูลระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์ (ระหว่าง 2 โครโมโซม) โดยทำการบีบอัดโครโมโซมพ่อพันธุ์และแม่พันธุ์ร่วมกัน (ดังนั้น 2 โครโมโซมจะสร้างได้ 1 พจนานุกรม (2C-1D: 2 Chromosome – 1 Dictionary)) เพื่อต้องการทดสอบว่าการแบ่งปันความรู้มีผลต่อประสิทธิภาพในการระบุหน่วยการสร้าง และการไขว้เปลี่ยนแบบจุดเดียวหรือไม่ (ให้ผลดีขึ้นหรือแย่ลง) เมื่อเทียบกับการบีบอัดข้อมูลที่ไม่มีการแบ่งปันความรู้ (LZWGA-1)

9.2.1 แนวคิดและหลักการ

จากผลการทดลองของขั้นตอนวิธีการบีบอัดแบบ LZW ที่ไม่มีการแบ่งปันความรู้ (ในหัวข้อ 9.1 LZWGA-1) ได้แสดงให้เห็นถึงความเป็นไปได้ในการประยุกต์ใช้วิธีการบีบอัดแบบ LZW กับขั้นตอนวิธีเชิงพันธุกรรม เพราะสามารถหาคำตอบพบภายใต้เงื่อนไขที่กำหนดในบทที่ 4 และมีประสิทธิภาพดีกว่า sGA ในปัญหา Trap-5 แม้ว่าหลักการระบุชิ้นส่วนด้วยวิธีการบีบอัดแบบ LZW จะสามารถเข้าใจได้ยากเนื่องจากการระบุชิ้นส่วนเป็นแบบปรับตัวที่มีความซับซ้อนกว่าการไขว้เปลี่ยนแบบชิ้นส่วน (บทที่ 5 sGA-FC) แต่จากผลการทดสอบสะท้อนให้เห็นว่ามุมมองหน่วยการสร้างในรูปแบบชิ้นส่วนยังสามารถใช้ได้ และเข้ากันได้กับตัวดำเนินการไขว้เปลี่ยนแบบจุดเดียว ดังนั้นในส่วนนี้เราจึงสนใจที่จะพัฒนาต่อโดยกำหนดให้มีการแบ่งปันข้อมูลระหว่างพ่อพันธุ์และแม่พันธุ์ที่ไขว้เปลี่ยนกัน ซึ่งเป็นความรู้ที่ได้จากการศึกษาการไขว้เปลี่ยนแบบชิ้นส่วน (ในบทที่ 5) และขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (บทที่ 8 CFT) ที่ช่วยให้การหาคำตอบมีประสิทธิภาพดีขึ้น

การแบ่งปันความรู้ในกระบวนการบีบอัดแบบ LZW ทำได้โดยการนำโครโมโซมพ่อพันธุ์และแม่พันธุ์มาบีบอัดโดยใช้พจนานุกรมร่วมกัน เริ่มจากการบีบอัดโครโมโซมพ่อพันธุ์เสร็จแล้วจึงทำการบีบอัดโครโมโซมแม่พันธุ์ต่อ (ยังใช้พจนานุกรมเดิม) ดังรูปที่ 9.10 (ส่วนบน) ผลที่ได้จะเห็นว่าชิ้นส่วนที่ระบุได้ในโครโมโซมพ่อพันธุ์ (C1) โดยเฉลี่ยจะมีขนาดเล็กกว่าชิ้นส่วนที่ระบุได้ในโครโมโซมแม่พันธุ์ (C2) ส่งผลให้จำนวนจุดตัดที่มีในโครโมโซมแม่พันธุ์ลดลงด้วย (ในรูป 9.10 อาจเห็นผลได้ไม่ชัดเนื่องจากโครโมโซมมีความยาวน้อยเกินไป) เนื่องจากวิธีการบีบอัดแบบ LZW มีการจดจำรูปแบบที่เคยเกิดขึ้นแล้วในพจนานุกรม (จดจำเฉพาะรูปแบบที่ไม่ซ้ำกัน) ดังนั้นยังมีข้อมูลที่ผ่านกระบวนการบีบอัดแบบ LZW มากขึ้นเท่าไร พจนานุกรมก็จะยิ่งมีขนาดโตขึ้นตามไปด้วย และค่า (ซึ่งก็คือรูปแบบชิ้นส่วน) ในพจนานุกรมจะมีขนาดยาวขึ้นด้วย หลังจากระบุชิ้นส่วนใน

โครโมโซมพ่อพันธุ์แม่พันธุ์ได้แล้วขั้นตอนถัดไปคือการไขว้เปลี่ยนแบบจุดเดียว ซึ่งยังคงใช้หลักการแบบเดียวกับที่ใช้ในวิธี LZWGA-1 นั่นก็คือเราจะพยายามไม่ทำให้ชิ้นส่วนแยกออกจากกัน เริ่มต้นเราจะหาจุดตัดที่เป็นไปได้ (จุดที่เป็นรอยต่อระหว่างชิ้นส่วน) ในโครโมโซมพ่อพันธุ์และแม่พันธุ์ หลังจากนั้นเราจะทำการเทียบตำแหน่งของจุดตัดว่ามีตำแหน่งที่ตรงกันหรือไม่ ถ้ามีเราจะทำการสุ่มเลือกจุดตัดที่ตรงกัน (ทุกจุดมีโอกาสถูกเลือกเท่ากันหมด) แล้วทำการไขว้เปลี่ยนกันดังรูปที่ 9.10 (ส่วนล่าง) จะเห็นว่าโครโมโซม C1 และ C2 มีจำนวนจุดตัดที่ตรงกัน 3 จุด คือคู่จิ้นตำแหน่งที่ 1 4 และ 6 จากรูปเราเลือกจุดตัดตำแหน่งที่ 4 ดังนั้นจะเกิดการไขว้เปลี่ยนกันในตำแหน่งที่ 5 – 8 แต่ในกรณีที่ไม่มีจุดตัดที่ตรงกันเลยเราจะทำการสุ่มเลือกโครโมโซมตัวใดตัวหนึ่ง (พ่อพันธุ์หรือแม่พันธุ์) แล้วทำการสุ่มเลือกจุดตัดที่เป็นไปได้ในโครโมโซมที่ถูกเลือก หลังจากนั้นเราจะทำการไขว้เปลี่ยนจากตำแหน่งที่สุ่มเลือกได้ ซึ่งจะส่งผลให้ชิ้นส่วนในโครโมโซมที่ไม่ถูกเลือกที่ตรงกับจุดตัดถูกแบ่งแยกออกจากกัน หรือกล่าวได้ว่าถ้าไม่มีจุดตัดที่ตรงกันระหว่างสองโครโมโซมที่บีบอัดร่วมกันจะทำให้มีชิ้นส่วนจำนวน 1 ชิ้นที่จะต้องถูกแบ่งแยกเสมอ การแบ่งแยกที่เกิดขึ้นนี้เราถือว่าเป็นผลกระทบที่ยอมรับได้เพราะเกิดขึ้นกับชิ้นส่วนเพียงชิ้นเดียว และถ้ามองในแง่ดีการแบ่งแยกนี้เป็นการสร้างความหลากหลายให้มีเพิ่มขึ้น



รูปที่ 9.12 การระบุชิ้นส่วนของ LZWGA-2 และการไขว้เปลี่ยนแบบจุดเดียว

(C หมายถึง Chromosome และ F หมายถึง Fragment)

ในภาพรวมการไขว้เปลี่ยนไม่ได้เกิดขึ้นบ่อยที่ตำแหน่งต้น ๆ เหมือนกับ LZWGA-1 แต่เกิดขึ้นกระจายตัวมากขึ้นกว่าเดิม เพราะว่าโครโมโซมที่ถูกบีบอัดเป็นตัวที่สองจะมีขนาดชิ้นส่วนที่ใหญ่ขึ้นโดยเฉลี่ย และถ้าเราพิจารณาถึงรูปแบบของคำในพจนานุกรมหรือรูปแบบชิ้นส่วนที่ได้จาก

การบีบอัดแบบ LZW จะมีความสัมพันธ์เป็น $Z = \sum_{i=1}^l 2^i$ เมื่อ Z คือรูปแบบชิ้นส่วนที่เป็นไปได้ทั้งหมด และ l คือขนาดความยาวของโครโมโซม เมื่อโครโมโซมมีขนาดยาวขึ้นรูปแบบชิ้นส่วนที่เป็นไปได้จะโตขึ้นเป็นฟังก์ชันเลขชี้กำลังซึ่งถือว่าเยอะมากเมื่อเทียบกับความยาว ดังนั้นรูปแบบชิ้นส่วนที่มีอยู่ (เป็นไปได้) ในโครโมโซมพ่อพันธุ์และแม่พันธุ์จะมีความหลากหลายสูง (ระดับความหลากหลายขึ้นกับรูปแบบการกระจายตัวของคู่ยีนที่เป็น 0 และ 1 ในโครโมโซมพ่อพันธุ์และแม่พันธุ์) ด้วยเหตุนี้ทำให้การวิเคราะห์การกระจายตัวของจุดตัดทำได้ยากเนื่องจากการบีบอัดแบบ LZW เป็นวิธีการบีบอัดแบบปรับตัว (ไม่แน่นอน) ขึ้นอยู่กับรูปแบบโครงสร้างของโครโมโซมเป็นหลัก และขึ้นอยู่กับลำดับก่อนหลังของข้อมูลที่ทำกรบีบอัดด้วย เช่น สมมติให้โครโมโซม C1 และ C2 มีรูปแบบโครงสร้างที่ต่างกัน ถ้าเราบีบอัด C1 ก่อนแล้วตามด้วย C2 รูปแบบชิ้นส่วนที่ระบุได้จะต่างจาก กรณีที่เราบีบอัด C2 ก่อนแล้วค่อยตามด้วย C1 เป็นต้น ใน

D คือ ประชากร

S คือ ดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้โดยใช้วิธี LZW

P คือ ผลเฉลยที่ได้จากวิธีการเลือก

ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA-2) ชนิด 2C-1D

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

วนซ้ำสำหรับ $i = 1, 3, 7, \dots, R-1$ {จนกระทั่งครบ R ตัว}

$P_{i,i+1} \leftarrow$ เลือกผลเฉลย 2 ตัว (ตัวที่ i กับ $i+1$) จาก D_{g-1} ที่สอดคล้องกับวิธีการเลือก

$S_p \leftarrow$ นำผลเฉลย $P_{i,i+1}$ ไปทำการบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกัน แล้วเก็บค่าดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้สำหรับแต่ละโครโมโซม (เก็บแยกกัน)

$P_{i,i+1} \leftarrow$ ทำการไขว้เปลี่ยนกันแบบ sGA โดยเลือกตำแหน่งไขว้เปลี่ยนที่ตรงกันระหว่างผลเฉลยจาก S_g^P กับ $S_g^{P_{i+1}}$ ถ้าไม่มีตำแหน่งตรงกันเลยให้ใช้ตำแหน่งที่สุ่มได้จากผลเฉลยตัวใดตัวหนึ่ง

$D_g \leftarrow$ นำ $P_{i,i+1}$ ไปทำการผ่าเหล่า (เป็นทางเลือก แต่งานวิจัยนี้ไม่ใช้)

จบการวนซ้ำ

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 9.13 รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-2

มุมมองของการสำรวจและการแสวงหาประโยชน์เราสามารถกล่าวได้ว่า การบีบอัดทำให้เกิดการแสวงหาประโยชน์มากกว่าการสำรวจ และการไขว้เปลี่ยนแบบจุดเดียวในวิธีที่น่าเสนอ (ไขว้เปลี่ยนจากจุดเชื่อมต่อนៃของชิ้นส่วน) แม้จะมีส่วนช่วยในการสำรวจแต่ก็ถือว่ามึผลน้อยกว่าเมื่อเทียบกับการแสวงหาประโยชน์ที่เกิดขึ้น ดังนั้นจึงทำให้เกิดความสงสัยว่าขั้นตอนวิธีที่มีการแสวงหาประโยชน์สูงจะมีความเหมาะสมกับปัญหาลักษณะใด และจากความรู้ที่มีมาก่อนหน้านี้ทำให้เราได้ข้อสันนิษฐานว่าขั้นตอนวิธีที่ใช้การบีบอัด (ซึ่งมีการแสวงหาประโยชน์สูง) น่าจะมีความสามารถในการแก้ปัญหาที่มีรูปแบบวนซ้ำ หรือมีรูปแบบตายตัวแน่นอนได้อย่างมีประสิทธิภาพ

P_1, P_2 คือ ผลเฉลยที่ผ่านการคัดเลือกมาแล้ว และผ่านกระบวนการบีบอัดแบบ

LZW เพื่อหาค่าดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้แล้ว

S คือ ดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้โดยใช้วิธี LZW

ขั้นตอนวิธีการไขว้เปลี่ยนแบบจุดเดียว (One-point Crossover)

ถ้าดำเนินการไขว้เปลี่ยน ทำ

$P_1, P_2 \leftarrow$ ทำการไขว้เปลี่ยน P_1, P_2 กันแบบ sGA โดยเลือกตำแหน่งไขว้เปลี่ยนที่ตรงกันระหว่างผลเฉลยจาก S_{P_1} กับ S_{P_2} ถ้าไม่มีตำแหน่งตรงกันเลยให้ใช้ตำแหน่งที่สุ่มได้จากผลเฉลยตัวใดตัวหนึ่ง

จบเงื่อนไข

จบขั้นตอนวิธี

รูปที่ 9.14 รหัสเทียมของวิธีการไขว้เปลี่ยนแบบจุดเดียวสำหรับ LZWGA-2

9.2.2 การทดลองและผลการทดลอง

ในส่วนนี้กล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่งในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่นำเสนอ และมีการวิจารณ์ผลการทดลองในมุมมองที่นำเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ-1

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม (sGA)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การไขว้เปลี่ยน:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่ใช้เปรียบเทียบ-2

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWGA-1)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	ใช้หลักการบีบอัดแบบ LZW โดยใช้พจนานุกรมแยกกันสำหรับแต่ละโครโมโซม
การประกอบ:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่นำเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWGA-2)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	ใช้หลักการบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกันสำหรับโครโมโซม พันธุ์พันธุ์แม่พันธุ์
การประกอบ:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 9.2 และ 9.3

ตารางที่ 9.2 เป็นการเปรียบเทียบกับ sGA ที่มีหลักการดำเนินการพื้นฐานเหมือนวิธี LZWGA-2 ต่างกันในรายละเอียดวิธีการเลือกจุดตัด โดยวิธี sGA สามารถเลือกจุดตัดได้เป็นอิสระ แต่วิธี LZWGA-2 จะเลือกได้เฉพาะตำแหน่งรอยต่อของชิ้นส่วนที่ตรงกันระหว่างโครโมโซมพ่อพันธุ์กับแม่พันธุ์เท่านั้น เมื่อพิจารณาถึงประสิทธิภาพของการหาคำตอบในปัญหา Royal Road วิธี LZWGA-2 ใช้จำนวน FEs ใกล้เคียงกับ sGA และเมื่อทดสอบโดยใช้ t-Test จะได้ผลว่าค่า FEs ของทั้งสองวิธีไม่แตกต่างกันอย่างมีนัยสำคัญที่ความเชื่อมั่น 95% หรือกล่าวได้ว่าวิธี LZWGA-2 มีประสิทธิภาพเท่ากับ sGA ในการแก้ปัญหา Royal Road ส่วนปัญหา Trap-5 วิธี LZWGA-2 ใช้จำนวน FEs น้อยกว่า sGA แสดงให้เห็นว่าวิธี LZWGA-2 มีประสิทธิภาพดีกว่า sGA ในการแก้ปัญหา Trap-5 และเมื่อพิจารณาถึงจำนวนประชากรที่ใช้ในการหาคำตอบในปัญหา Royal Road ในกรณีที่ทำคำตอบพบวิธี LZWGA-2 ใช้จำนวนประชากรเท่ากับ sGA หรือกล่าวได้ว่าวิธี LZWGA-2 ต้องการระดับความหลากหลายในการแก้ปัญหาเท่ากับ sGA แต่ในกรณีที่ทำคำตอบไม่พบวิธี LZWGA-2 ใช้จำนวนประชากรน้อยกว่า sGA แสดงให้เห็นว่าวิธี LZWGA-2 ต้องการระดับความหลากหลายต่ำสุดในการแก้ปัญหาน้อยกว่า sGA สำหรับปัญหา Trap-5 ในกรณีที่ทำคำตอบพบวิธี LZWGA-2 ใช้จำนวนประชากรน้อยกว่า sGA และในกรณีที่ทำคำตอบไม่พบวิธี LZWGA-2 ใช้จำนวนประชากรเท่ากับ sGA แสดงให้เห็นว่าวิธี LZWGA-2 ต้องการระดับความหลากหลายในการแก้ปัญหาสูงกว่า sGA และต้องการระดับความหลากหลายต่ำสุดในการแก้ปัญหาน้อยกว่า sGA (ใช้ประชากรน้อยกว่า)

ตารางที่ 9.2 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWGA-2

Parameters		Success 30 run				Fail 30 run	
		sGA		LZWGA-2		sGA	LZWGA-2
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,200	11,920	1,200	12,041	200	100
Trap-5	60	2,300	28,367	1,800	20,400	300	200

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

เมื่อพิจารณาผลจากตารางที่ 9.3 ที่มีการเปรียบเทียบกับ LZWGA-1 ซึ่งใช้หลักการบีบอัดแบบ LZW เหมือน LZWGA-2 ต่างกันที่ LZWGA-1 ไม่มีการแบ่งปันความรู้ระหว่าง

โครโมโซม แต่วิธี LZWGA-2 มีการแบ่งปันความรู้ระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์ สำหรับปัญหา Royal Road และ Trap-5 ในกรณีที่หาคำตอบพบวิธี LZWGA-2 ใช้จำนวน FEs และจำนวนประชากรในการแก้ปัญหาน้อยกว่าวิธี LZWGA-1 เป็นการแสดงให้เห็นว่าวิธี LZWGA-2 มีประสิทธิภาพสูงกว่า LZWGA-1 และต้องการระดับความหลากหลายในการแก้ปัญหาน้อยกว่า LZWGA-1 ส่วนกรณีที่หาคำตอบไม่พบในปัญหา Royal Road วิธี LZWGA-2 ต้องการระดับความหลากหลายต่ำสุดเท่ากับวิธี LZWGA-1 เนื่องจากใช้จำนวนประชากรเท่ากัน และปัญหา Trap-5 วิธี LZWGA-2 ต้องการระดับความหลากหลายต่ำสุดต่ำกว่าวิธี LZWGA-1 เนื่องจากใช้จำนวนประชากรน้อยกว่า

ตารางที่ 9.3 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ LZWGA-1 และ LZWGA-2

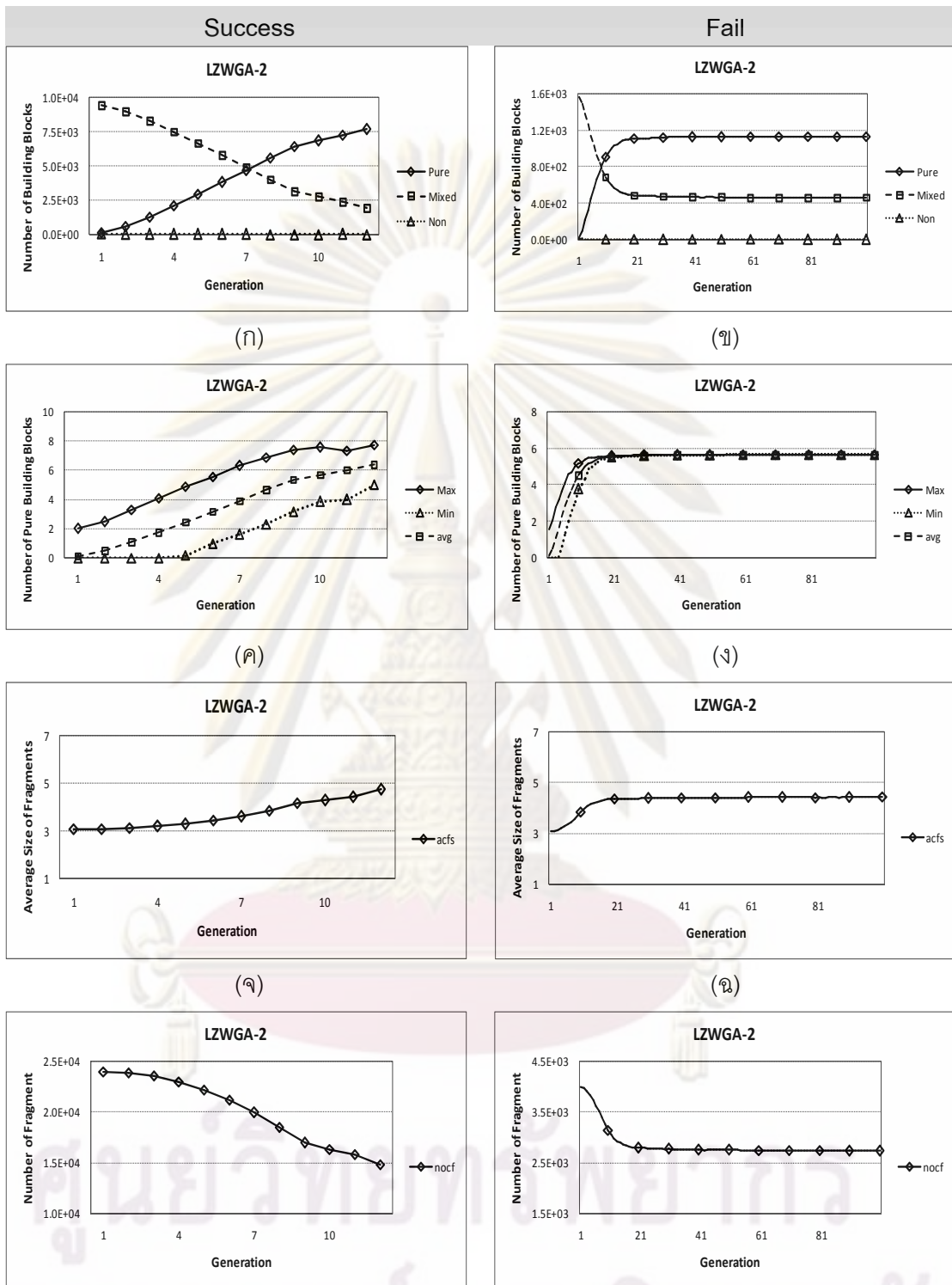
Parameters		Success 30 run				Fail 30 run	
		LZWGA-1		LZWGA-2		LZWGA-1	LZWGA-2
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,800	18,540	1,200	12,041	200	200
Trap-5	60	2,200	25,740	1,800	20,400	300	200

หมายเหตุ เส้นไขว้ใน ๓ ในการทดลองอยู่ในบทที่ 4

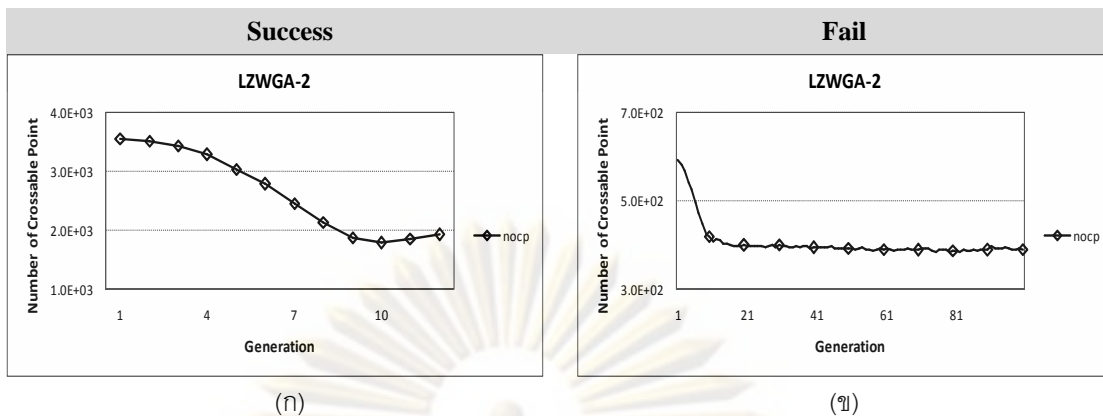
ผลทดสอบของวิธี LZWGA-2 จากรูปที่ 9.15 มีรูปแบบเหมือนกับผลทดสอบของวิธี LZWGA-1 ในรูปที่ 9.4 เนื่องจากวิธี LZWGA-2 ใช้หลักการเดียวกันในการระบุและการประกอบชิ้นส่วน ต่างกันตรงที่วิธี LZWGA-2 มีการแบ่งปันความรู้โดยจะทำการบีบอัดโครโมโซมพ่อพันธุ์และแม่พันธุ์ร่วมกัน ส่วนวิธี LZWGA-1 จะทำการบีบอัดโครโมโซมแต่ละตัวแยกกัน ความแตกต่างที่เห็นได้ชัดระหว่างวิธีทั้งสองพิจารณาได้จากมุมมองของชิ้นส่วน จากรูปที่ 9.15 (จ-ฉ) ขนาดโดยเฉลี่ยของชิ้นส่วนในวิธี LZWGA-2 มีค่ามากกว่าวิธี LZWGA-1 ในรูปที่ 9.4 (จ-ฉ) เนื่องจากวิธี LZWGA-2 ใช้โครโมโซมในการบีบอัดร่วมกันมากกว่า (มีข้อมูลในการบีบอัดมากกว่า) ขนาดของชิ้นส่วนส่งผลโดยตรงต่อปริมาณของชิ้นส่วนที่ระบุได้ ถ้าระบุได้ชิ้นส่วนขนาดใหญ่ขึ้น ปริมาณชิ้นส่วนที่ระบุได้ก็จะลดลง (แปรผกผันกัน) ทำให้ในรูปที่ 9.15 (ช-ซ) มีปริมาณชิ้นส่วนน้อยกว่าวิธี LZWGA-1 ในรูปที่ 9.4 (ช-ซ)

เมื่อพิจารณาถึงปริมาณของตำแหน่งจุดตัดที่ตรงกันระหว่างสองโครโมโซมจากรูปที่ 9.16 ปริมาณของจุดตัดจะลดลงอย่างต่อเนื่องและมีค่าเกือบคงที่ในช่วงรุ่นประชากรท้าย ๆ เหตุที่ทำให้ปริมาณของจุดตัดลดลง เป็นผลจากการบีบอัดโครโมโซมพ่อพันธุ์และแม่พันธุ์ร่วมกัน โดยปกติวิธีบีบอัดแบบ LZW อาศัยความรู้ที่มีก่อนหน้านี้ในการบีบอัดและจะทำการเพิ่มขนาดของค่า (ความรู้) ทีละหนึ่งตัว (หนึ่งบิต) ดังนั้นเมื่อเริ่มต้นบีบอัดจากโครโมโซมพ่อพันธุ์ ซึ่งในขณะนั้นพจนานุกรมยังไม่มีความรู้เพิ่มเติม หลังจากการบีบอัดเสร็จพจนานุกรมจะมีค่ามากขึ้น (มีความรู้เพิ่มขึ้น) ต่อจากนั้นจะทำการบีบอัดโครโมโซมแม่พันธุ์โดยอาศัยความรู้ที่มีอยู่ก่อนหน้านี้ในพจนานุกรมทำให้ชิ้นส่วนที่ระบุได้มีรูปแบบที่ยาวขึ้น (ได้ค่าที่มีขนาดใหญ่ขึ้น) ส่งผลให้จำนวนจุดตัดหรือรอยต่อระหว่างชิ้นส่วนในโครโมโซมแม่พันธุ์ลดลง พิจารณาได้จากรูปที่ 9.16 (LZWGA-2) เทียบกับรูปที่ 9.5 (LZWGA-1) ดังนั้นแม้ว่าโครโมโซมพ่อพันธุ์และแม่พันธุ์จะมีโครงสร้างทางพันธุกรรมเหมือนกัน แต่ผลที่ได้จากการบีบอัดในแต่ละโครโมโซมจะต่างกัน (ถ้าเป็นวิธี LZWGA-1 ผลที่ได้ในแต่ละโครโมโซมจะเหมือนกัน) หรือกล่าวได้ว่าวิธี LZWGA-2 มีตำแหน่งจุดตัดที่ตรงกันระหว่างสองโครโมโซมน้อยกว่าวิธี LZWGA-1 ส่งผลให้เมื่อเกิดการสุ่มเข้าจำนวนจุดตัดที่ตรงกันจะลดลงเรื่อย ๆ และจะมีค่าเกือบคงที่เมื่อประชากรส่วนใหญ่มีโครงสร้างทางพันธุกรรมคล้ายกัน ในกรณีที่หาคำตอบไม่พบจากรูปที่ 9.16 (ข) ปริมาณของจุดตัดลดลงอย่างรวดเร็ว แปรผันตามระดับความหลากหลายที่ต่ำลงอย่างรวดเร็ว

จากที่กล่าวมาอาจยังมีความสงสัยว่าทำไมวิธี LZWGA-2 มีประสิทธิภาพในการทำงานดีกว่าวิธี LZWGA-1 ทั้ง ๆ ที่ทั้งสองวิธีมีรูปแบบการดำเนินการคล้ายกัน สันนิษฐานว่าเป็นเพราะการแบ่งปันความรู้ร่วมกันทำให้มีการแสวงหาประโยชน์สูงขึ้น หรือมองได้ว่าการที่ชิ้นส่วนมีขนาดใหญ่ขึ้นส่งผลให้เกิดการแตกแยกน้อยลง (เกิดการทำลายเค้าร่างที่ดัดน้อยลง) และชิ้นส่วนที่ใหญ่ขึ้นนี้ส่วนหนึ่งได้มาจากโครงสร้างทางพันธุกรรมของโครโมโซมที่เกิดการบีบอัดก่อนหน้านี้ หรือมองได้ว่าเป็นการพยายามรักษาโครงสร้างส่วนที่เหมือนกันระหว่างสองโครโมโซมไม่ให้ถูกแบ่งแยกออกจากกัน ซึ่งจากความรู้ที่ได้จากบทที่ 5 (sGA-FC) และบทที่ 8 (CFT) ทำให้เชื่อได้ว่าโครงสร้างส่วนที่เหมือนกันระหว่างสองโครโมโซมเป็นชิ้นส่วนที่ดี และจะนำไปสู่คำตอบที่ต้องการได้

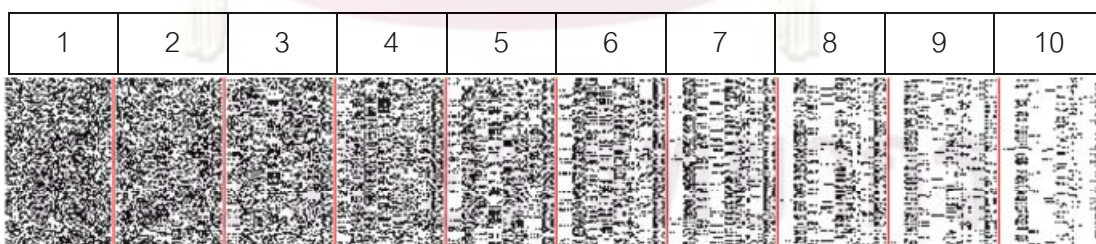


รูปที่ 9.15 ผลการทดสอบของ LZWGA-2 ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (ค) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ง) (ช)

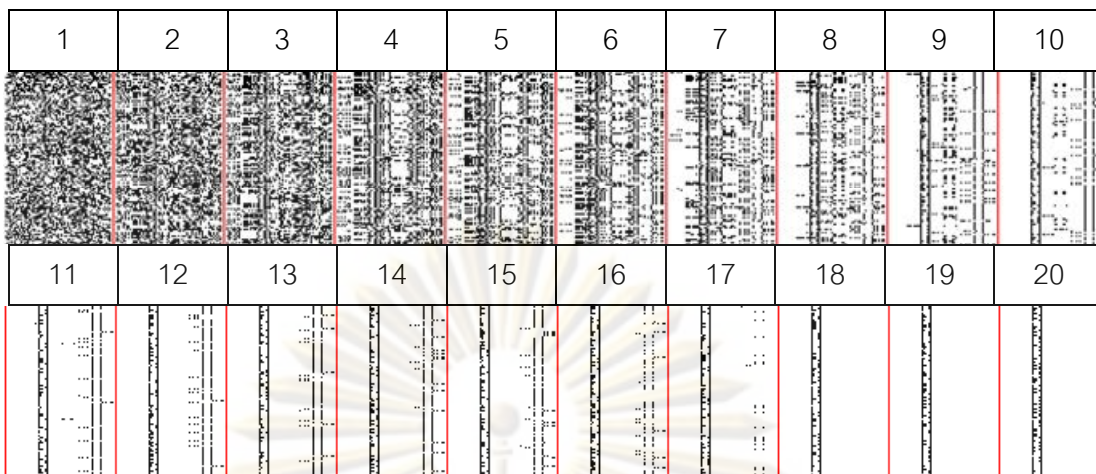


รูปที่ 9.16 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-2 ใน ปัญหา Royal Road 64 บิต

เมื่อพิจารณาโครงสร้างทางพันธุกรรมของประชากรในภาพรวมจากรูปที่ 9.17 และรูปที่ 9.18 (วิธีการอ่านรูปอยู่ในบทที่ 4) หลังจากดำเนินการผ่านไประยะหนึ่ง (ประมาณ 3 รุ่นประชากร) พื้นที่สีขาว (คู่จีนที่มีค่าเป็น 1) ค่อย ๆ เพิ่มขึ้นอย่างต่อเนื่องแสดงให้เห็นว่าวิธี LZWGA-2 สามารถแยกแยะโครงสร้างที่เป็นคำตอบ (Pure BBs) และโครงสร้างที่ไม่ใช่คำตอบ (Non BBs) ได้ กรณีที่หาคำตอบพบจากรูปที่ 9.17 เกิดการลู่เข้าในระดับโครงสร้างย่อยสังเกตได้จากแถบสีดำ (ในแนวตั้ง) ในรุ่นประชากรที่ 6 และจางลงจนกระทั่งพบคำตอบ ต่างจากกรณีที่หาคำตอบไม่พบจากรูปที่ 9.18 ที่เกิดการลู่เข้าในระดับโครงสร้างย่อยในรุ่นประชากรที่ 4 และมีความชัดเจนขึ้นเรื่อย ๆ จนถึงรุ่นประชากรที่ 20 ความแตกต่างนี้เกิดการระดับความหลากหลายที่คงเหลืออยู่ในกลุ่มประชากร ในกรณีที่หาคำตอบพบมีระดับความหลากหลายเหลืออยู่เพียงพอที่จะทำให้ค้นพบรูปแบบใหม่และแยกแยะได้ว่ารูปแบบที่ดีกว่ามีโครงสร้างเป็นอย่างไร ต่างจากกรณีที่หาคำตอบไม่พบที่ความหลากหลายเหลือน้อยเกินไปทำให้การลู่เข้าปรากฏชัดเจนขึ้นเรื่อย ๆ



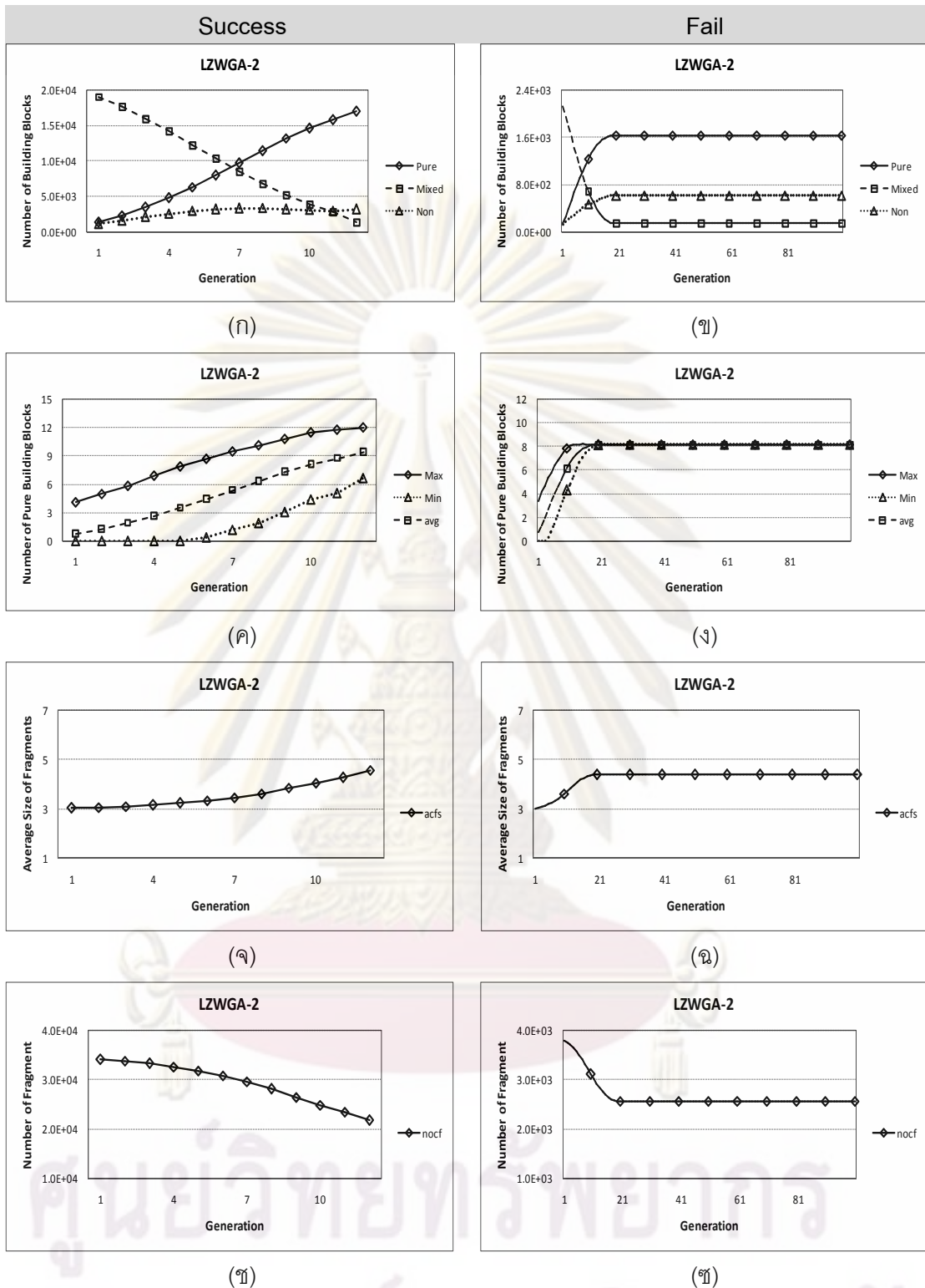
รูปที่ 9.17 รูปแบบจีโนไทป์ของประชากรของ LZWGA-2 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่จีนที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่จีนที่มีค่าเป็น 0



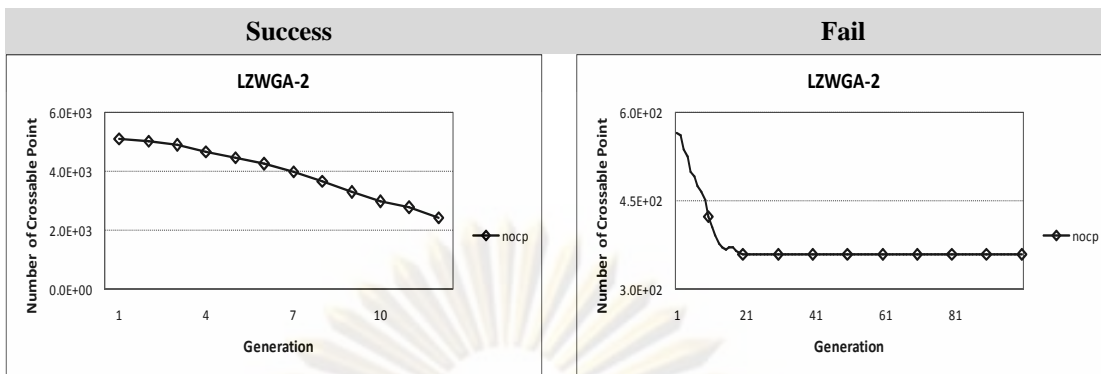
รูปที่ 9.18 รูปแบบจีโอเมทรีของประชากรของ LZWGA-2 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคูจिनที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคูจिनที่มีค่าเป็น 0

สำหรับปัญหา Trap-5 ผลการทดสอบของวิธี LZWGA-2 จากรูปที่ 9.19 ยังคงมีรูปแบบเหมือนกับผลการทดสอบของวิธี LZWGA-1 ในรูปที่ 9.8 ต่างกันในส่วนขนาดของขนาดโดยเฉลี่ยของชิ้นส่วนและปริมาณชิ้นส่วนที่ระบุได้ วิธี LZWGA-2 มีขนาดของชิ้นส่วนใหญ่กว่าเพราะมีข้อมูลในการบีบอัดมากกว่า (ใช้โครโมโซมสองตัวในการบีบอัด) ส่งผลให้มีปริมาณชิ้นส่วนที่ระบุได้น้อยกว่า (โครโมโซมมีความยาวเท่ากัน ดังนั้นถ้าขนาดชิ้นส่วนใหญ่ขึ้นจำนวนชิ้นส่วนก็จะลดลง) และต่างจากปัญหา Royal Road ตรงที่ปัญหา Trap-5 มีการลวงในหน่วยการสร้างทำปริมาณของ Non BBs เพิ่มขึ้นด้วยดังรูปที่ 9.19 (ก-ข)

ปริมาณของตำแหน่งที่ตรงกันระหว่างสองโครโมโซมในปัญหา Trap-5 จากรูปที่ 9.20 มีลักษณะเหมือนกับปัญหา Royal Road ในรูปที่ 9.16 ในกรณีที่หาคำตอบพบจากรูปที่ 9.20 (ก) เมื่อเทียบกับรูป 9.16 (ก) มีความแตกต่างกันในช่วงรุ่นประชากรท้าย ๆ ในปัญหา Royal Road ปริมาณตำแหน่งจุดมีค่าเกือบคงที่ แต่ในปัญหา Trap-5 ปริมาณตำแหน่งจุดตัดมีค่าลดลงอย่างต่อเนื่อง เป็นเพราะปัญหา Trap-5 มีการลวงในระดับหน่วยการสร้างส่งผลให้เกิดการลู่เข้าในระดับโครงสร้างย่อย ทำให้โครงสร้างทางพันธุกรรมของโครโมโซมมีความแตกต่างกันมากกว่าปัญหา Royal Road และในกรณีที่หาคำตอบพบการลู่เข้าเกิดขึ้นในระดับต่ำกว่ากรณีที่หาคำตอบไม่พบ ดังนั้นความแตกต่างระหว่างโครงสร้างทางพันธุกรรมจึงมีสูงกว่า ด้วยเหตุเหล่านี้ทำให้จำนวนจุดตัดที่ตรงกันระหว่างสองโครโมโซมยังคงลดลงอย่างต่อเนื่อง



รูปที่ 9.19 ผลการทดสอบของ LZWGA-2 ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นของประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

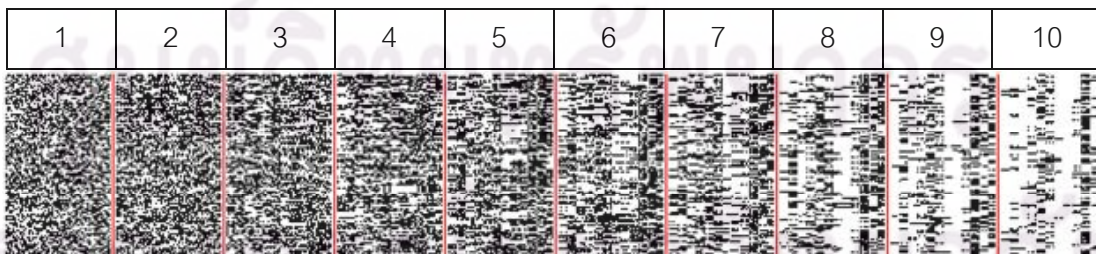


(ก)

(ข)

รูปที่ 9.20 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-2 ใน ปัญหา Trap-5 60 บิต

เมื่อพิจารณาถึงโครงสร้างทางพันธุกรรมของประชากรในภาพรวมจากรูปที่ 9.21 และรูปที่ 9.22 (วิธีการอ่านรูปอยู่ในบทที่ 4) จะเห็นว่าเกิดการลู่เข้าในระดับโครงสร้างย่อยในลักษณะรูปแบบเดียวกับหน่วยการสร้างลวง พิจารณาได้จากวิวิธสีดำ (คูจินที่มีค่าเป็น 0) ในแนวนอน ในกรณีที่หาคำตอบพบการลู่เข้าเริ่มปรากฏให้เห็นในรุ่นประชากรที่ 5 สังเกตได้จากแถบสีขาว (คูจินที่มีค่าเป็น 1) ในช่วงกลางของภาพ และมีปริมาณเพิ่มขึ้นจนพบคำตอบในรุ่นประชากรที่ 10 ส่วนกรณีที่ไม่หาคำตอบไม่พบเริ่มเกิดการลู่เข้าในรุ่นประชากรที่ 3 สังเกตได้จากแถบสีดำและแถบสีขาวที่เรียงตัวกันอย่างมีระเบียบในแนวดิ่ง การลู่เข้าชัดเจนขึ้นในรุ่นประชากรที่ 5 และเกิดขึ้นอย่างสมบูรณ์ในรุ่นประชากรที่ 18 สำหรับกรณีที่ไม่หาคำตอบพบในปัญหา Trap-5 จากรูปที่ 9.21 เมื่อเทียบกับกรณีที่ไม่หาคำตอบพบในปัญหา Royal Road ในรูปที่ 9.17 จะสังเกตเห็นว่าโครงสร้างทางพันธุกรรมในปัญหา Trap-5 มีความแตกต่างกันมากกว่าปัญหา Royal Road เป็นการยืนยันให้เห็นถึงเหตุผลที่ทำให้ปริมาณจุดตัดที่ตรงกันระหว่างสองโครโมโซมยังคงลดลงอย่างต่อเนื่องดังที่กล่าวไปแล้วในส่วนก่อนหน้า



รูปที่ 9.21 รูปแบบจีโนไทป์ของประชากรของ LZWGA-2 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคูจินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคูจินที่มีค่าเป็น 0

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20

รูปที่ 9.22 รูปแบบจีโอไทป์ของประชากรของ LZWGA-2 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0

9.2.3 สรุปผลการทดลอง

ในส่วนนี้นำเสนอขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-2 ที่ใช้หลักการบีบอัดข้อมูลแบบ LZW ในการระบุหน่วยการสร้างแบบที่มีการแบ่งปันความรู้ระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์ (บีบอัดทีละคู่) และใช้วิธีการไขว้เปลี่ยนแบบจุดเดียวในการประกอบหน่วยการสร้างเข้าด้วยกันโดย ผลการทดสอบเมื่อเทียบกับ sGA ในปัญหา Royal Road วิธี LZWGA-2 มีประสิทธิภาพในการแก้ปัญหาเทียบเท่ากับ และในปัญหา Trap-5 วิธี LZWGA-2 มีประสิทธิภาพในการแก้ปัญหาดีกว่า และเมื่อเปรียบเทียบกับ LZWGA-1 จะเห็นว่าวิธี LZWGA-2 มีประสิทธิภาพในการแก้ปัญหาดีกว่าทั้งในปัญหา Royal Road และปัญหา Trap-5 ซึ่งเป็นการแสดงให้เห็นว่าการแบ่งปันความรู้ช่วยให้การระบุหน่วยการสร้างมีประสิทธิภาพที่ดีขึ้นได้ และมีความสอดคล้องกับสมมติฐานที่ว่าขั้นตอนวิธีที่ใช้การบีบอัดน่าจะมีความสามารถในการแก้ปัญหาที่มีรูปแบบวนซ้ำ หรือมีรูปแบบตายตัวแน่นอนได้อย่างมีประสิทธิภาพ แต่อาจไม่เป็นจริงเฉพาะปัญหาที่มีรูปแบบวนซ้ำเท่านั้น เนื่องจากเรายังไม่ได้ทำการทดสอบกับปัญหาที่มีรูปแบบคำตอบไม่ตายตัว หรือไม่วนซ้ำ จึงทำให้ไม่สามารถสรุปได้ในตอนนี้ (มีการทดสอบกับปัญหาที่มีรูปแบบคำตอบแบบไม่ตายตัว หรือไม่วนซ้ำในบทที่ 11)

9.3 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW รูปแบบที่ 3 (LZWGA-3)

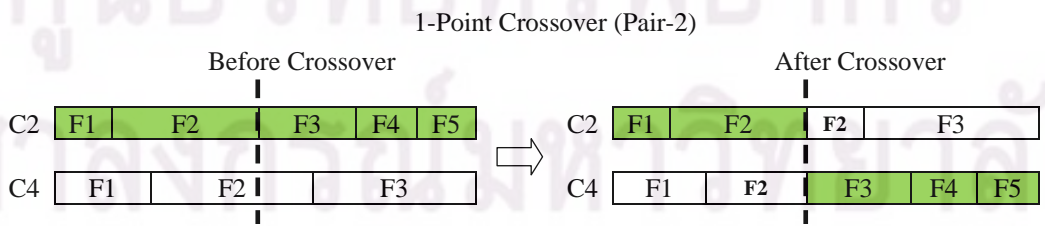
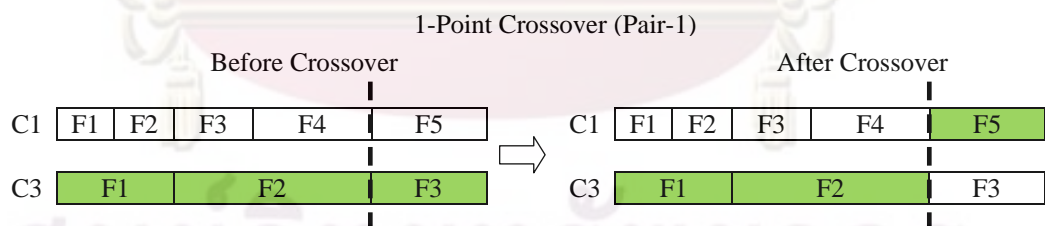
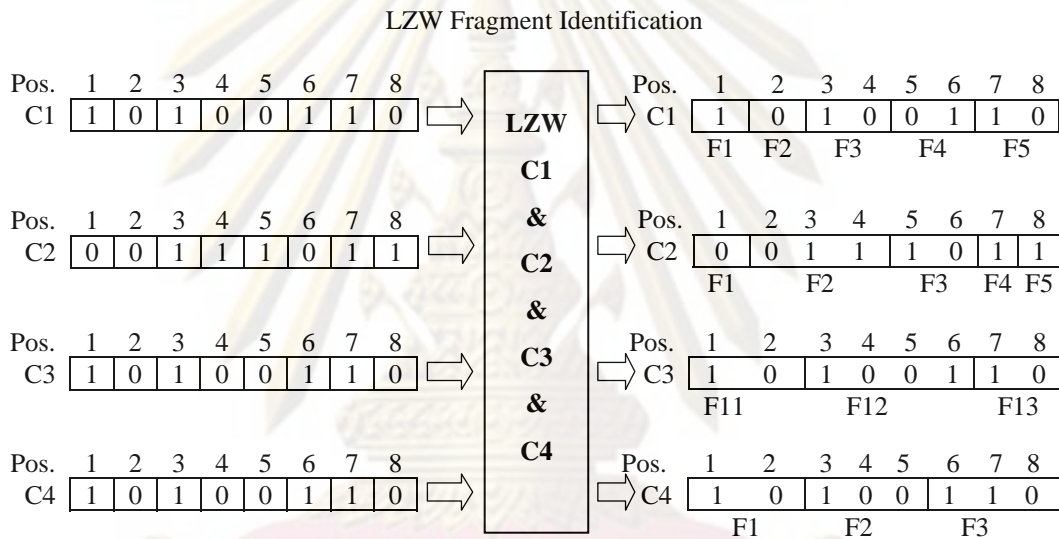
ส่วนนี้จะกล่าวถึงการประยุกต์ใช้หลักการบีบอัดข้อมูลกับโครโมโซมในรูปแบบที่มีการแบ่งปันข้อมูลระหว่างโครโมโซมที่ถูกเลือกทั้งหมด (มีจำนวนเท่ากับขนาดของประชากร) โดยทำการบีบอัดโครโมโซมที่ถูกเลือกทั้งหมดร่วมกัน (กลุ่มโครโมโซมที่ถูกเลือก (Sub-Population: SP) จะสร้างได้ 1 พจนานุกรม (1SP-1D: 1 Sub-Population – 1 Dictionary)) เพื่อต้องการทดสอบการแบ่งปันความรู้ในระดับสูงสุด (แบ่งปันระหว่างโครโมโซมที่ถูกเลือกทั้งหมด) ต่อประสิทธิภาพในการระบุหน่วยการสร้าง และการไขว้เปลี่ยนแบบจุดเดียวว่าจะมีผลอย่างไร (ดีขึ้นหรือแย่ลง) เมื่อเทียบกับวิธีการแบ่งปันความรู้ในระดับต่ำสุด (LZWGA-2 ที่ใช้การแบ่งปันความรู้ระหว่าง 2 โครโมโซม) ที่ได้นำเสนอก่อนหน้า ซึ่งจะช่วยให้เห็นภาพในระดับขอบเขตของการแบ่งปันความรู้ว่ามีผลกระทบอย่างไรกับการหาคำตอบ

9.3.1 แนวคิดและหลักการ

จากผลการทดลองของขั้นตอนวิธีการบีบอัดแบบ LZW ที่มีการแบ่งปันความรู้ในระดับต่ำสุด (ในหัวข้อ 9.2 LZWGA-2) ได้แสดงให้เห็นถึงประสิทธิภาพในการหาคำตอบเมื่อเทียบกับ sGA ว่ามีความเทียบเท่ากันในปัญหา Royal Road (ทดสอบโดยใช้หลักการทางสถิติ t-Test) และดีกว่าในปัญหา Trap-5 และเมื่อเทียบกับ LZWGA-1 จะเห็นว่า LZWGA-2 ให้ผลดีกว่าในทุกปัญหาทดสอบ ซึ่งเป็นการแสดงให้เห็นว่าการแบ่งปันความรู้ให้ผลที่ดีตรงกับความรู้ที่ได้จากเรื่องการไขว้เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์ (Eiben, 1997) แต่ก็ทำให้เกิดข้อกังขา (เกิดความสงสัย) ว่าการแบ่งปันความรู้ควรเกิดขึ้นในระดับใดจึงจะมีความเหมาะสมในการแก้ปัญหา การตอบคำถามหรือข้อสงสัยนี้ทำได้ยากเนื่องจากมีรายละเอียดและปัจจัยที่เกี่ยวข้องจำนวนมาก อย่างไรก็ตามวิธีหนึ่งซึ่งช่วยให้เห็นภาพหรือมีความรู้เพิ่มขึ้นก็คือ ทำการทดสอบการแบ่งปันความรู้ในระดับสูงสุด เพื่อให้เห็นความแตกต่างที่ชัดเจนในระดับของการแบ่งปันความรู้ (ต่ำสุดกับสูงสุด)

วิธีที่นำเสนอมีความแตกต่างจากสองวิธีที่นำเสนอก่อนหน้านี้ (LZWGA-1 และ LZWGA-2) เป็นอย่างมาก เนื่องจากวิธีนี้ต้องทำการเลือกประชากรที่จะนำไปทำการไขว้เปลี่ยนทั้งหมดก่อน (เลือกซ้ำได้) ซึ่งโดยปกติของ sGA แล้วพ่อพันธุ์แม่พันธุ์หนึ่งคู่สามารถสร้างลูกได้หนึ่งคู่ (เลือกมา 2 ตัวสร้างได้ 2 ตัวเท่ากัน) ดังนั้นจึงต้องทำการเลือกโครโมโซมจำนวนเท่ากับขนาดของประชากร เราเรียกโครโมโซมที่ถูกเลือกมาทั้งหมดเหล่านี้ว่าประชากรย่อย (Sub-Population) หลังจากนั้นจึงนำประชากรย่อยเหล่านี้ไปทำการบีบอัดแบบ LZW ร่วมกันทุกตัว (ใช้พจนานุกรมอันเดียวกัน) เพื่อทำการระบุชิ้นส่วน ขั้นตอนถัดไปเป็นขั้นตอนของการไขว้เปลี่ยนที่ทำแบบเดียวกับวิธี

ที่นำเสนอก่อนหน้านี้คือ จากโครโมโซมที่ถูกเลือกทั้งหมดเลือกมาทีละ 2 ตัวแบบสุ่ม โดยที่โครโมโซมตัวไหนเคยถูกเลือกแล้วจะไม่ถูกเลือกอีก (ไม่มีการเลือกซ้ำ) แล้วหาตำแหน่งรอยต่อของชิ้นส่วนที่ตรงกันระหว่างโครโมโซม (ทั้งสองตัว) ถ้ามีตำแหน่งที่ตรงกันจะทำการสุ่มเลือกจุดตัดจากตำแหน่งที่ตรงกันขึ้นมาหนึ่งตำแหน่งแล้วทำการไขว้เปลี่ยนในตำแหน่งที่สุ่มเลือกได้ แต่ถ้าไม่มีตำแหน่งรอยต่อชิ้นส่วนที่ตรงกันเลยจะทำการสุ่มเลือกโครโมโซมมาหนึ่งตัว (ตัวใดตัวหนึ่งจากคู่ที่จะทำการไขว้เปลี่ยนกัน) หลังจากนั้นจะสุ่มหาตำแหน่งจุดตัดจากรอยต่อของชิ้นส่วนในตัวที่ถูกเลือก แล้วจึงทำการไขว้เปลี่ยนโครโมโซมทั้งสองตรงตำแหน่งที่ถูกเลือกนี้ ตัวอย่างรูปที่ 2.19 (ส่วนบน) มีโครโมโซมที่ถูกเลือกคือ C1, C2, C3 และ C4 รวมทั้งหมด 4 ตัวและถูกบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกัน (เป็นการระบุชิ้นส่วน) เมื่อบีบอัดเสร็จจะได้ตำแหน่งและรูปแบบของชิ้นส่วนที่



รูปที่ 9.23 การระบุชิ้นส่วนของ LZWGA-3 และการไขว้เปลี่ยนแบบจุดเดียว

(C หมายถึง Chromosome และ F หมายถึง Fragment)

มีอยู่ในโครโมโซมแต่ละตัว ถัดไปเป็นขั้นตอนการไขว้เปลี่ยน (รูปส่วนล่าง) มีการจับคู่เพื่อทำการไขว้เปลี่ยนระหว่าง C1 กับ C3 และ C2 กับ C4 (มีทั้งหมด 2 คู่) คู่แรก C1 กับ C3 มีจุดตัดร่วมกันทั้งหมด 2 จุดคือคู่จีนตำแหน่งที่ 2 และ 6 ในรูปเลือกจุดตัดตำแหน่งที่ 6 ดังนั้นจึงเกิดการไขว้เปลี่ยนกันของคู่จีนตำแหน่งที่ 7-8 ส่วนคู่ที่สอง C2 กับ C4 ไม่มีจุดตัดร่วมกันเลย ดังนั้นจึงต้องทำการหาจุดตัดโดยสุ่มเลือกโครโมโซมตัวใดตัวหนึ่ง จากรูปโครโมโซม C2 ถูกเลือกและมีจุดตัดที่เป็นไปได้ทั้งหมด 4 ตำแหน่งได้แก่ ตำแหน่งคู่จีนที่ 1, 4, 6 และ 7 จากนั้นเราจะทำการสุ่มเลือกขึ้นมาหนึ่งตำแหน่ง (จากทั้งหมด 4 ตำแหน่งที่เป็นไปได้) จากรูปเลือกได้ตำแหน่งที่ 4 ดังนั้นคู่จีนตำแหน่งที่ 5-8 จะถูกไขว้เปลี่ยนกัน ซึ่งมีผลทำให้ชิ้นส่วนที่อยู่ในโครโมโซมที่ไม่ถูกเลือก (C4) ถูกแบ่งแยกออกจากกันหนึ่งชิ้นจากรูปคือชิ้นส่วน F2 จากจุดนี้อาจทำให้เกิดคำถามว่าเกิดเหตุการณ์ที่เลือกคู่โครโมโซมมาแล้วไม่มีจุดตัดร่วมกันหลายครั้งหรือไม่ (โอกาสที่จะเกิดเหตุการณ์นี้เป็นเท่าใด) คำตอบคือ เป็นการยากที่จะคำนวณหาค่าความน่าจะเป็นที่แน่นอนเนื่องจากวิธีการบีบอัดแบบ LZW มีความซับซ้อนสูง (เป็นวิธีแบบปรับตัวที่ขึ้นกับรูปแบบค่าและลำดับของข้อมูล รวมถึงปริมาณของข้อมูลด้วย) เราสามารถอธิบายเพื่อให้เห็นภาพอย่างคร่าว ๆ ได้ว่าในกรณีที่โครโมโซมมีความยาวคงที่แต่มีจำนวนไม่คงที่ (ขนาดประชากรไม่คงที่) จำนวนโครโมโซมที่เพิ่มขึ้นเมื่อผ่านกระบวนการบีบอัดแบบ LZW จะทำให้โครโมโซมตัวทำย ๆ มีขนาดของชิ้นส่วนใหญ่ขึ้นเนื่องจากใช้พจนานุกรมร่วมกันทำให้ได้คำที่มีความยาวเพิ่มขึ้น หรือกล่าวได้ว่าโครโมโซมตัวทำย ๆ ของการบีบอัดจะมีชิ้นส่วนน้อยลงเนื่องจากชิ้นส่วนมีขนาดใหญ่ขึ้นทำให้มีจุดตัดที่เป็นไปได้น้อยลง ดังนั้นการจับคู่ระหว่างโครโมโซมใด ๆ จะมีโอกาสเกิดจุดตัดที่ตรงกันน้อยลงโดยเฉลี่ย และในกรณีที่โครโมโซมมีจำนวนคงที่ (ขนาดประชากรคงที่) แต่มีขนาดความยาวของโครโมโซมไม่คงที่ (แต่ยาวเท่ากันหมดทุกตัว) ขนาดความยาวที่เพิ่มขึ้น (คู่จีนมากขึ้น) จะทำให้ขนาดของชิ้นส่วนโดยเฉลี่ยในโครโมโซมใหญ่ขึ้นด้วย และโครโมโซมตัวทำย ๆ จะมีขนาดชิ้นส่วนใหญ่กว่าโครโมโซมตัวก่อน ๆ เนื่องจากขนาดของคำที่มีอยู่ในพจนานุกรมมีขนาดใหญ่ขึ้น (ผ่านการเรียนรู้มามากขึ้นทำให้เคยเห็นรูปแบบมามากขึ้น ส่งผลให้คำมีขนาดยาวขึ้น) ดังนั้นการจับคู่ระหว่างโครโมโซมใด ๆ จะมีโอกาสเกิดจุดตัดที่ตรงกันน้อยลงโดยเฉลี่ย จากทั้งสองกรณีไม่ว่าจะเป็นเรื่องจำนวนหรือความยาวของโครโมโซมที่เพิ่มขึ้นล้วนส่งผลให้มีจุดตัดที่เป็นไปได้ (ตรงกัน) ลดลงโดยเฉลี่ยทั้งสิ้น

ถ้าใช้ประชากรขนาดใหญ่ขึ้น (มีจำนวนโครโมโซมมากขึ้น) จะทำให้เกิดการไขว้เปลี่ยนที่ไม่ตรงกับรอยต่อของชิ้นส่วนเพิ่มขึ้น ซึ่งเป็นการสะท้อนถึงจำนวนชิ้นส่วนที่ถูกแบ่งแยกจะมีเพิ่มขึ้นด้วย เมื่อพิจารณาการเพิ่มนี้จะเห็นว่ามีเพียงแค่นั้นชิ้นส่วนที่ถูกแบ่งแยกต่อการไขว้เปลี่ยนหนึ่งครั้งเท่านั้น หรือในกรณีเลวร้ายที่สุด (ทุกการไขว้เปลี่ยนที่เกิดขึ้นไม่มีจุดตัดที่ตรงกันระหว่างคู่โครโมโซมเลย) จะมีจำนวนชิ้นส่วนที่ถูกแบ่งแยกเท่ากับครึ่งหนึ่งของขนาดประชากร เช่น

มีประชากร 100 ตัว จะเกิดการไขว้เปลี่ยนขึ้น 50 ครั้ง ดังนั้นจะมีชิ้นส่วนที่ถูกแบ่งแยกเพียงแค่ 50 ชิ้นเท่านั้น เป็นต้น ซึ่งเมื่อเทียบกับจำนวนชิ้นส่วนทั้งหมดที่มีในประชากรถือว่ามีค่าน้อยมาก ผลกระทบที่เกิดขึ้นจากการแบ่งแยกชิ้นส่วนในมุมมองของการสำรวจและการแสวงหาประโยชน์ เราจะพิจารณาว่าเป็นการเพิ่มความหลากหลาย (เพิ่มการสำรวจ) ซึ่งเป็นสิ่งที่ดี เนื่องจากเมื่อชิ้นส่วนมีขนาดใหญ่ขึ้นนั้นหมายถึงเกิดการแสวงหาประโยชน์เพิ่มขึ้น แต่ถ้าเพิ่มมากเกินไปจะทำให้เกิดการลู่เข้าอย่างรวดเร็วก่อนที่จะหาค่าตอบพบ

D คือ ประชากร

S คือ ดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้โดยใช้วิธี LZW

P คือ ผลเฉลยที่ผ่านการคัดเลือกมาแล้ว

ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA-3) ชนิด 1SP-1D

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม

วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$Q_g \leftarrow$ เลือกผลเฉลยจำนวนเท่ากับขนาดประชากรโดยวิธีการเลือกจาก

D_{g-1} โดย

$S_g \leftarrow$ นำผลเฉลยแต่ละตัวจาก Q_g ที่ไม่ซ้ำกันไปทำการบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกัน แล้วเก็บค่าดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้สำหรับแต่ละโครโมโซม (เก็บแยกกัน) ถ้ามีการซ้ำกันตัวที่ซ้ำจะมีค่าเหมือนตัวแรกที่ดำเนินการไปแล้ว

วนซ้ำสำหรับ $i = 1, 3, 7, \dots, R-1$ {จนกระทั่งครบ R ตัว}

$P_{i,i+1} \leftarrow$ เลือกผลเฉลย 2 ตัว (ตัวที่ i กับ $i+1$) จาก Q_g

$P_{i,i+1} \leftarrow$ ทำการไขว้เปลี่ยนกันแบบ sGA โดยเลือกตำแหน่งไขว้เปลี่ยนที่ตรงกันระหว่างผลเฉลยจาก S_g^i กับ S_g^{i+1} ถ้าไม่มีตำแหน่งตรงกันเลยให้ใช้ตำแหน่งที่สุ่มได้จากผลเฉลยตัวใดตัวหนึ่ง

$D_g \leftarrow$ นำ $P_{i,i+1}$ ไปทำการผ่าเหล่า (เป็นทางเลือก แต่งานวิจัยนี้ไม่ใช้)

จบการวนซ้ำ

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 9.24 รหัสเทียมของขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-3

P_1, P_2 คือ ผลเฉลยที่ผ่านการคัดเลือกมาแล้ว และผ่านกระบวนการบีบอัดแบบ

LZW เพื่อหาค่าดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้แล้ว

S คือ ดัชนีตำแหน่งรอยต่อของแต่ละช่วงที่บีบอัดได้โดยใช้วิธี LZW

ขั้นตอนวิธีการไขว้เปลี่ยนแบบจุดเดียว (One-point Crossover)

ถ้าดำเนินการไขว้เปลี่ยน ทำ

$P_1, P_2 \leftarrow$ ทำการไขว้เปลี่ยน P_1, P_2 กันแบบ sGA โดยเลือก

ตำแหน่งไขว้เปลี่ยนที่ตรงกันระหว่างผลเฉลยจาก S_{P_1} กับ

S_{P_2} ถ้าไม่มีตำแหน่งตรงกันเลยให้ใช้ตำแหน่งที่สุ่มได้จาก

ผลเฉลยตัวใดตัวหนึ่ง

จบเงื่อนไข

จบขั้นตอนวิธี

รูปที่ 9.25 รหัสเทียมของวิธีการไขว้เปลี่ยนแบบจุดเดียวสำหรับ LZWGA-3

9.3.2 การทดลองและผลการทดลอง

ในส่วนนี้จะกล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่ง ในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่น่าเสนอ และมีการวิจารณ์ผลการทดลองในมุมมองที่น่าเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ-1

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม (sGA)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การไขว้เปลี่ยน:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่ใช้เปรียบเทียบ-2

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWGA-1)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	ใช้หลักการบีบอัดแบบ LZW โดยใช้พจนานุกรมแยกกันสำหรับแต่ละโครโมโซม
การประกอบ:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่ใช้เปรียบเทียบ-3

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWGA-2)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	ใช้หลักการบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกันสำหรับโครโมโซม พันธุ์พันธุ์แม่พันธุ์
การประกอบ:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่นำเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWGA-3)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การระบุ:	ใช้หลักการบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกันสำหรับโครโมโซมที่ถูกเลือกทุกตัว
การประกอบ:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 9.4 – 9.6

จากผลการทดลองในตารางที่ 9.4 เป็นการเปรียบเทียบกับ sGA ซึ่งมีหลักการดำเนินการพื้นฐานเหมือนวิธี LZWGA-3 ต่างกันแค่เพียงวิธีการเลือกจุดตัดในการไขว้เปลี่ยน โดยวิธี sGA สามารถเลือกจุดตัดได้อิสระ แต่วิธี LZWGA-3 จุดตัดที่เลือกได้จะต้องเป็นตำแหน่งที่เป็นรอยต่อระหว่างชิ้นส่วนที่ตรงกันระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น ในปัญหา Royal Road และปัญหา Trap-5 ในกรณีที่หาคำตอบพบวิธี LZWGA-3 ใช้จำนวน FEs และจำนวนประชากรสูงกว่า sGA แสดงให้เห็นว่าวิธี LZWGA-3 มีประสิทธิภาพในการแก้ปัญหาต่ำกว่าวิธี sGA และวิธี LZWGA-3 ต้องการระดับความหลากหลายในการแก้ปัญหาสูงกว่า sGA ในกรณีที่หาคำตอบไม่พบสำหรับปัญหา Royal Road วิธี LZWGA-3 ใช้จำนวนประชากรเท่ากับ sGA แสดงให้เห็นว่าต้องการระดับความหลากหลายต่ำสุดเท่ากับ sGA และสำหรับปัญหา Trap-5 วิธี LZWGA-3 ใช้จำนวนประชากรน้อยกว่า sGA ซึ่งให้เห็นว่าวิธี LZWGA-3 ต้องการระดับความหลากหลายต่ำสุดน้อยกว่าในการแก้ปัญหา

ตารางที่ 9.4 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWGA-3

Parameters		Success 30 run				Fail 30 run	
		sGA		LZWGA-3		sGA	LZWGA-3
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,200	11,920	2,500	25,667	200	200
Trap-5	60	2,300	28,367	2,500	31,417	300	200

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

เมื่อพิจารณาผลจากตารางที่ 9.5 มีการเปรียบเทียบกับ LZWGA-1 ซึ่งใช้หลักการบีบอัดแบบ LZW เหมือน LZWGA-3 ต่างกันที่ LZWGA-1 ไม่มีการแบ่งปันความรู้ระหว่างโครโมโซม แต่วิธี LZWGA-3 มีการแบ่งปันความรู้ระหว่างโครโมโซมที่ถูกเลือกทั้งหมด สำหรับปัญหา Royal Road และ Trap-5 ในกรณีที่หาคำตอบพบวิธี LZWGA-3 ใช้จำนวน FEs และจำนวนประชากรในการแก้ปัญหา มากกว่าวิธี LZWGA-1 แสดงให้เห็นว่าวิธี LZWGA-3 มีประสิทธิภาพต่ำกว่า LZWGA-1 และต้องการระดับความหลากหลายในการแก้ปัญหา มากกว่า LZWGA-1 ส่วนกรณีที่ไม่พบคำตอบในปัญหา Royal Road วิธี LZWGA-3 ต้องการระดับความหลากหลายต่ำสุดเท่ากับ LZWGA-1 เนื่องจากใช้จำนวนประชากรเท่ากัน และในปัญหา Trap-5 วิธี LZWGA-3 ต้องการระดับความหลากหลายต่ำสุดน้อยกว่าวิธี LZWGA-1 เนื่องจากใช้จำนวนประชากรน้อยกว่า

ตารางที่ 9.5 จำนวนประชากรและจำนวนครั้งการใช้ฟังก์ชันประเมินค่าความเหมาะสมของ LZWGA-1 และ LZWGA-3

Parameters		Success 30 run				Fail 30 run	
		LZWGA-1		LZWGA-3		LZWGA-1	LZWGA-3
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,800	18,540	2,500	25,667	200	200
Trap-5	60	2,200	25,740	2,500	31,417	300	200

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ผลจากตารางที่ 9.6 มีการเปรียบเทียบกับ LZWGA-2 ซึ่งใช้หลักการบีบอัดแบบ LZW เหมือน LZWGA-3 ต่างกันที่ LZWGA-2 มีการแบ่งปันความรู้ระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น ในขณะที่วิธี LZWGA-3 มีการแบ่งปันความรู้ระหว่างโครโมโซมที่ถูกเลือกทั้งหมด สำหรับปัญหา Royal Road และ Trap-5 ในกรณีที่ไม่พบคำตอบวิธี LZWGA-3 ใช้จำนวน FEs และจำนวนประชากรในการแก้ปัญหา มากกว่าวิธี LZWGA-2 แสดงให้เห็นว่าวิธี LZWGA-3 มีประสิทธิภาพต่ำกว่า LZWGA-2 และต้องการระดับความหลากหลายในการแก้ปัญหา มากกว่า LZWGA-2 ส่วนกรณีที่ไม่พบคำตอบในปัญหา Royal Road วิธี LZWGA-3 ต้องการระดับความหลากหลายต่ำสุดมากกว่า LZWGA-1 เนื่องจากใช้จำนวนประชากรมากกว่า และในปัญหา Trap-5 วิธี LZWGA-3 ต้องการระดับความหลากหลายต่ำสุดน้อยกว่าวิธี LZWGA-1 เนื่องจากใช้จำนวนประชากรน้อยกว่า

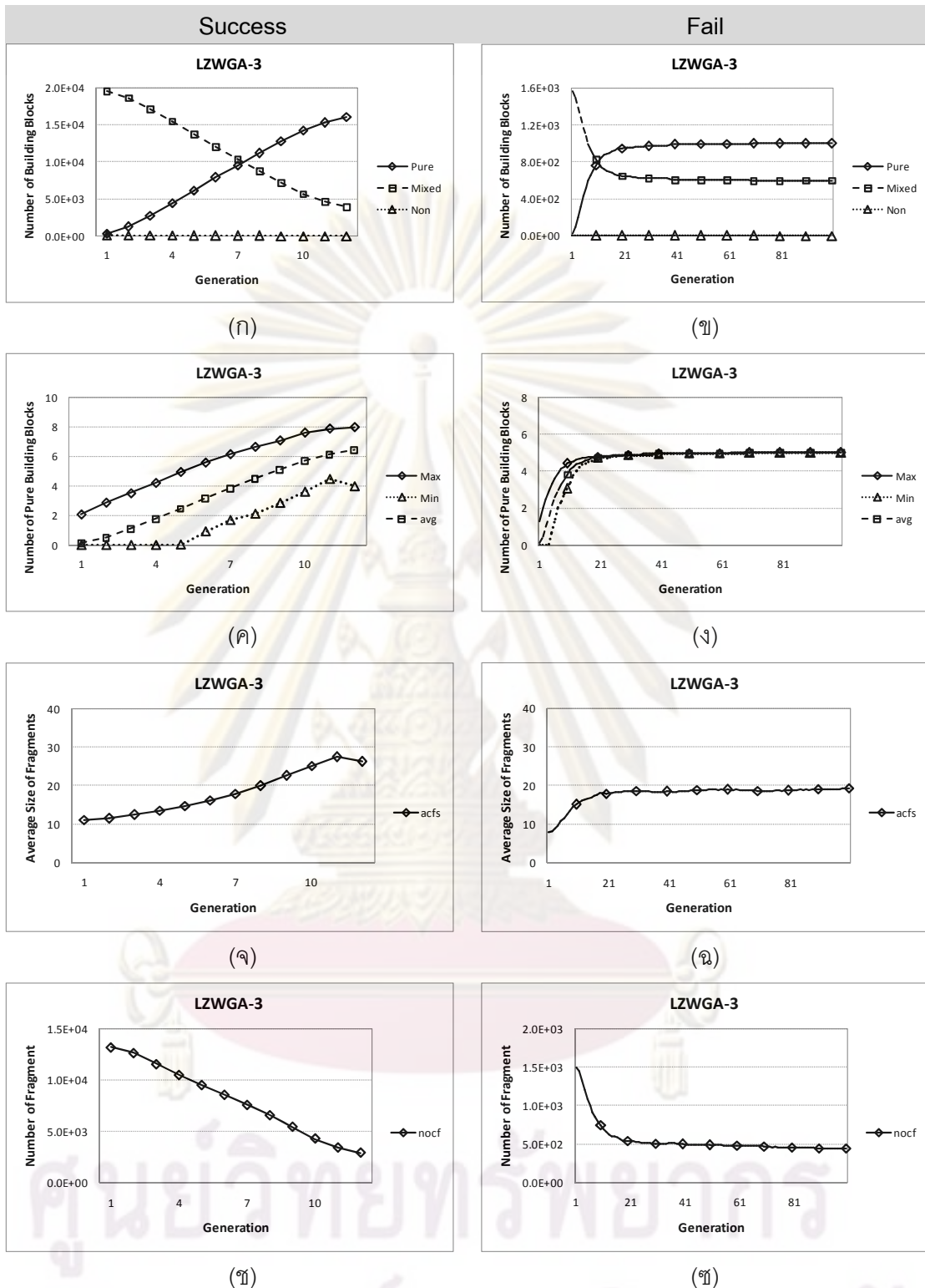
ตารางที่ 9.6 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ LZWGA-2 และ LZWGA-3

Parameters		Success 30 run				Fail 30 run	
		LZWGA-2		LZWGA-3		LZWGA-2	LZWGA-3
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
RoyalRoad	64	1,200	12,041	2,500	25,667	100	200
Trap-5	60	1,800	20,400	2,500	31,417	300	200

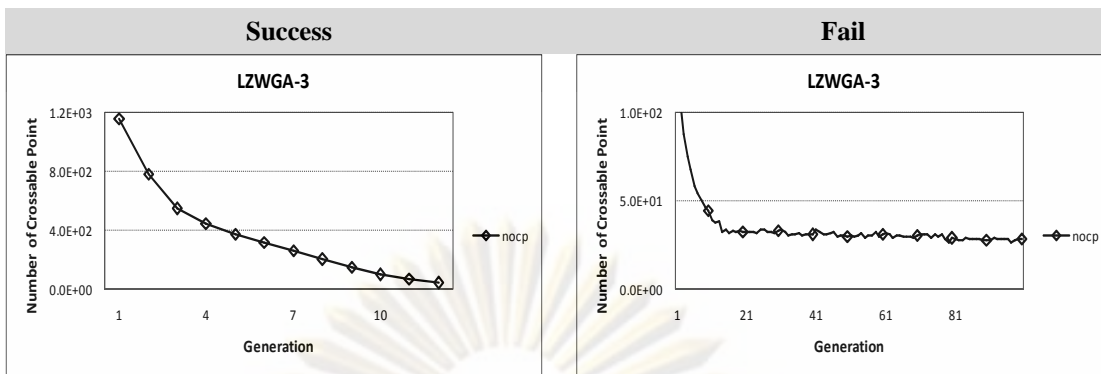
หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ผลการทดสอบของวิธี LZWGA-3 จากรูปที่ 9.26 มีรูปแบบเหมือนกับผลทดสอบในวิธี LZWGA-1 ในรูปที่ 9.4 และวิธี LZWGA-2 ในรูปที่ 9.15 เป็นเพราะใช้หลักการระบุและการประกอบชิ้นส่วนเหมือนกัน แต่ต่างกันในส่วนของการไล่เลียด วิธี LZWGA-3 มีการแบ่งปันความรู้สูงสุดพิจารณาจากการบีบอัดข้อมูลโดยใช้โครโมโซมที่ถูกเลือกมาทั้งหมดร่วมกัน (มีข้อมูลบีบอัดร่วมกันจำนวนมาก) ทำให้ได้ชิ้นส่วนที่มีขนาดใหญ่ขึ้นมากเมื่อเทียบกับสองวิธีก่อนหน้า พิจารณาได้จากรูปที่ 9.26 (จ-ฉ) ส่งผลให้มีปริมาณชิ้นส่วนน้อยลงดังรูปที่ 9.26 (ช-ซ) เมื่อเทียบกับขนาดประชากรที่ใช้ในการแก้ปัญหาซึ่งมีค่ามากกว่าวิธีอื่น

ปริมาณของตำแหน่งที่ตรงกันระหว่างสองโครโมโซมที่เลือกมาทำการไขว้เปลี่ยนจากรูปที่ 9.27 มีอัตราการลดลงเร็วกว่าวิธี LZWGA-2 เป็นเพราะวิธี LZWGA-3 ทำการบีบอัดโครโมโซมที่ถูกเลือกมาทั้งหมดก่อนทำให้โครโมโซมตัวหลัง ๆ ที่ถูกบีบมีขนาดของชิ้นส่วนใหญ่ขึ้น ส่งผลให้มีจำนวนรอยต่อระหว่างชิ้นส่วนน้อยลง และเมื่อทำการสุ่มโครโมโซมที่ผ่านการระบุชิ้นส่วนมาสองตัวเพื่อทำการไขว้เปลี่ยน โอกาสที่จะสุ่มเจอโครโมโซมที่ถูกบีบอัดในลำดับติดกันหรือใกล้เคียงกันมีน้อย ดังนั้นถ้าโครโมโซมสองตัวมีจำนวนชิ้นส่วนที่ระบุได้ต่างกันมาก โอกาสที่ตำแหน่งจุดตัดจะตรงกันมีเพิ่มขึ้นก็จริง แต่ว่าจะมีเพียงไม่กี่ตำแหน่งเท่านั้น (ชิ้นส่วนขนาดใหญ่มีจุดตัดที่เป็นไปได้น้อยอยู่แล้ว) และแม้ว่าโครโมโซมส่วนใหญ่จะมีหน้าตาคล้ายกัน แต่เนื่องจากใช้วิธีการบีบอัดแบบ LZW ดังนั้นปริมาณชิ้นส่วนที่ระบุได้ยังคงน้อยลงอยู่ดี (ชิ้นส่วนมีขนาดใหญ่ขึ้นเสมอ) ด้วยเหตุนี้ปริมาณของตำแหน่งจุดตัดที่ตรงกันระหว่างสองโครโมโซมจึงมีค่าลดลงอย่างรวดเร็ว



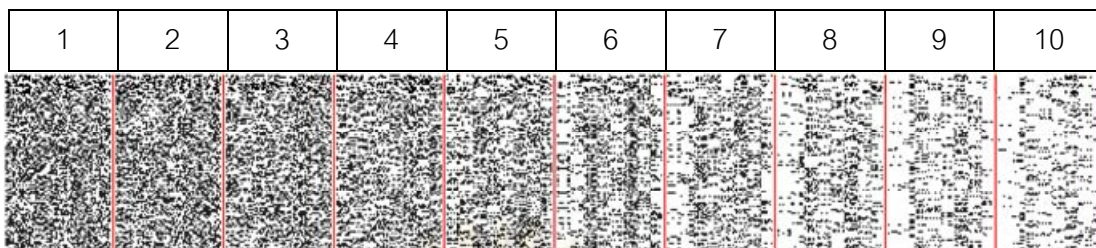
รูปที่ 9.26 ผลการทดสอบของ LZWGA-3 ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (ฉ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ซ) (ซ)



รูปที่ 9.27 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-3 ใน ปัญหา Royal Road 64 บิต

ผลการทดสอบประสิทธิภาพในการหาคำตอบจากตารางที่ 9.4 ถึงตารางที่ 9.6 วิธี LZWGA-3 มีประสิทธิภาพในการหาคำตอบต่ำกว่าวิธีอื่นทั้งหมด สันนิษฐานว่าเกิดจากการแบ่งปันความรู้ในระดับที่สูงเกินไปทำให้เกิดการแสวงหาประโยชน์มากกว่าการสำรวจมาก (ทำให้เสียสมดุล) หรือมองได้ว่าการที่มีชิ้นส่วนขนาดใหญ่เกินไปทำให้เกิดการแตกแยก (ของเค้กร่าง) ได้น้อย ดังนั้นส่วนประกอบที่ไม่ใช่คำตอบในชิ้นส่วนจะมีโอกาสคงอยู่สูง (กำจัดออกไปได้ยาก) ทำให้ประสิทธิภาพในการแก้ปัญหาลดลงเป็นอย่างมาก

โครงสร้างทางพันธุกรรมของประชากรในกรณีนี้หาคำตอบพบจากรูปที่ 9.28 (วิธีการอ่านรูปอยู่ในบทที่ 4) ในแต่ละรุ่นประชากรยังคงมีคูจีนที่เป็น 0 (พื้นที่สีดำ) ซึ่งไม่ใช่โครงสร้างของคำตอบการกระจายตัวอยู่ในหลายตำแหน่ง เป็นการยืนยันให้เห็นว่าวิธี LZWGA-3 มีการแสวงหาประโยชน์สูงทำให้โครงสร้างส่วนที่ไม่ใช่คำตอบถูกกำจัดออกไปได้ยาก ในทางกลับกันถ้าพิจารณาจากรุ่นประชากรท้าย ๆ ที่มีจำนวนจุดตัดที่ตรงกันน้อยลงทำให้เกิดการแตกแยก (ชิ้นส่วนถูกแบ่งแยก) สูงขึ้นอาจมองว่าเป็นความหลากหลายที่เพิ่มขึ้น แต่ในประชากรรุ่นท้าย ๆ มีระดับความหลากหลายต่ำเนื่องจากเกิดการลู่เข้าในระดับหนึ่ง ทำให้การแตกแยกที่เกิดขึ้นมีประโยชน์น้อยลง (ช่วยทำให้ประสิทธิภาพดีขึ้นได้ไม่มาก) ในกรณีนี้หาคำตอบไม่พบเริ่มเกิดการลู่เข้าอย่างชัดเจนในรุ่นประชากรที่ 3 สังเกตได้จากแถบสีดำและแถบสีขาวในแนวตั้งที่อยู่ติดกันในช่วงตำแหน่งกลาง ๆ ของภาพ และชัดเจนขึ้นในรุ่นประชากรที่ 4 ซึ่งเป็นการแสดงให้เห็นว่าระดับความหลากหลายของวิธี LZWGA-3 ลดลงอย่างรวดเร็วสอดคล้องกับรูปที่ 9.27



รูปที่ 9.28 รูปแบบจีโอโทป์ของประชากรของ LZWGA-3 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0



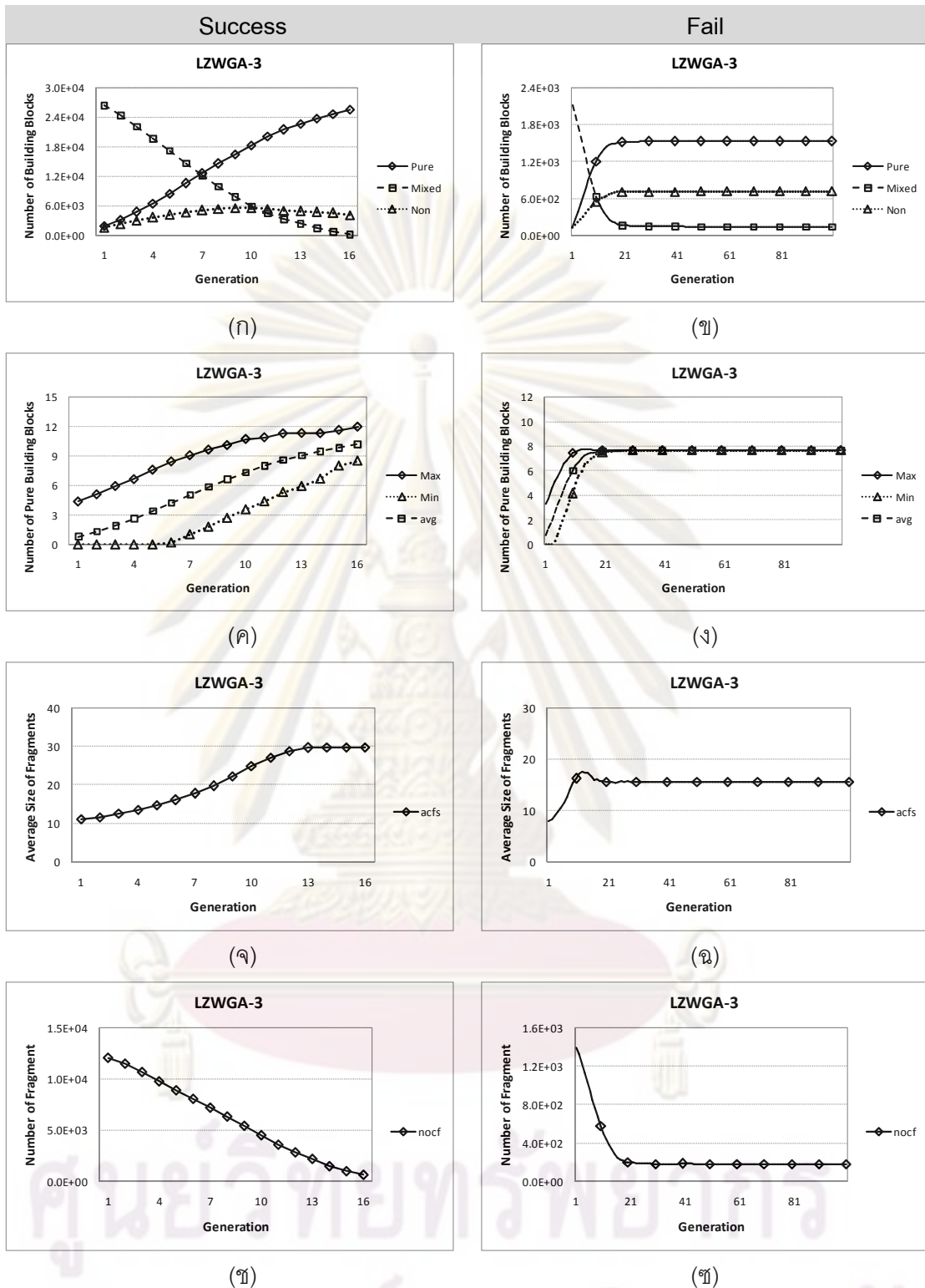
รูปที่ 9.29 รูปแบบจีโอโทป์ของประชากรของ LZWGA-3 ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0

สำหรับปัญหา Trap-5 ผลการทดสอบในภาพรวมมีรูปแบบคล้ายกับผลทดสอบของวิธี LZWGA-1 และ LZWGA-2 ส่วนที่ต่างกันคือขนาดและปริมาณของชิ้นส่วนที่ระบุได้ วิธี LZWGA-3 มีขนาดชิ้นส่วนใหญ่กว่าวิธีอื่นอย่างเห็นได้ชัดพิจารณาได้จากรูปที่ 9.30 (จ-ฉ) ทำให้มีปริมาณชิ้นส่วนน้อยเมื่อเทียบกับวิธีอื่นดังรูปที่ 9.30 (ข-ซ) ในกรณีที่ยาคำตอบพบจากรูปที่ 9.30 (จ) ในช่วงรุ่นประชากรท้าย ๆ ขนาดโดยเฉลี่ยของชิ้นส่วนมีค่าคงที่เป็นเพราะวิธี LZWGA-3 ใช้จำนวนโครโมโซมในการบีบอัดมาก (เท่ากับจำนวนประชากร) ทำให้ผลของการบีบอัดต่างกันน้อยเมื่อประชากรส่วนใหญ่มีหน้าตาคล้ายกันในระดับหนึ่ง (เกิดการลู่เข้าระดับหนึ่ง) และเนื่องจาก

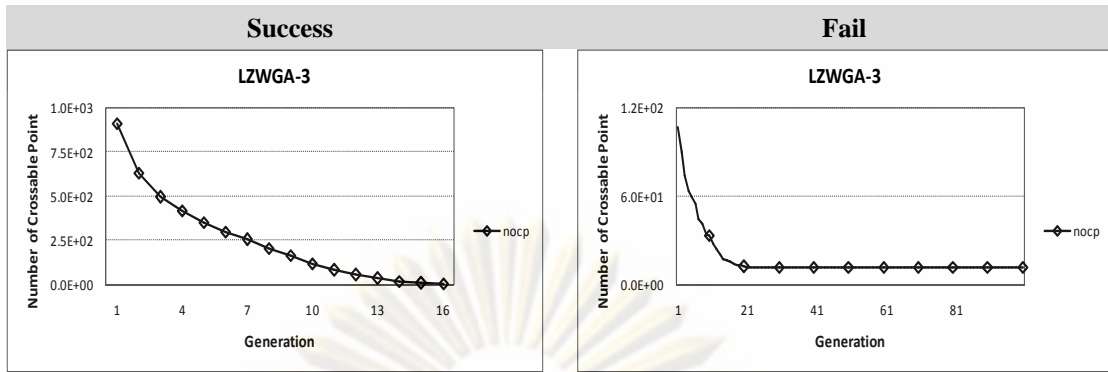
ความแตกต่างในแต่ละรุ่นประชากรเกิดขึ้นอย่างต่อเนื่องในระดับหนึ่งทำให้ปริมาณของชิ้นส่วนที่ระบุได้ลดลงแบบเชิงเส้นดังรูปที่ 9.30 (ข) ซึ่งต่างจากกรณีที่ทำค่าตอบไม่พบที่ระดับความหลากหลายลดลงอย่างรวดเร็วในรูปที่ 9.30 (ค) อย่างชัดเจน ส่วนกรณีที่ทำค่าตอบไม่พบจากรูปที่ 9.30 (ค) ในช่วงรุ่นประชากรที่ 10 – 20 ขนาดโดยเฉลี่ยของชิ้นส่วนเพิ่มขึ้นและลดลงจนมีค่าเกือบคงที่ เป็นเพราะในกรณีที่ค่าตอบไม่พบมีความหลากหลายน้อย ส่งผลให้แต่ละรอบของการทดสอบ (ทั้งหมด 30 รอบ) มีการลู่เข้าสู่รูปแบบที่ไม่เหมือนกันในช่วงรุ่นประชากรดังกล่าว และวิธีการบีบอัดแบบ LZW เมื่อใช้จำนวนประชากรในการบีบอัดมากจะยิ่งเห็นผลที่แตกต่างกันชัดเจนขึ้น วิธี LZWGA-3 ใช้จำนวนประชากรในการบีบอัดร่วมกันมากกว่าวิธีอื่น ดังนั้นการเปลี่ยนแปลงจึงเกิดชัดเจน

ปริมาณของตำแหน่งจุดตัดที่ตรงกันในการไขว้เปลี่ยนระหว่างสองโครโมโซมจากรูปที่ 9.31 มีรูปแบบเหมือนกับปัญหา Royal Road ในรูปที่ 9.27 เป็นเพราะการใช้จำนวนโครโมโซมในการบีบอัดจำนวนมากส่งผลให้การเปลี่ยนแปลงเกิดขึ้นในรูปแบบเดียวกัน (ไม่สามารถเห็นความแตกต่างได้) ปริมาณจุดตัดที่ตรงกันยังคงลดลงอย่างต่อเนื่องจนมีค่าเข้าใกล้ศูนย์ในกรณีที่ค่าตอบพบ ส่วนกรณีที่ค่าตอบไม่พบตำแหน่งจุดตัดที่ตรงกันลดลงจนถึงระดับหนึ่ง (เข้าใกล้ศูนย์) และมีค่าคงที่เป็นเพราะเกิดการลู่เข้าอย่างสมบูรณ์

เมื่อพิจารณาถึงโครงสร้างทางพันธุกรรมของประชากรในภาพรวมในกรณีที่ค่าตอบพบจากรูปที่ 9.32 (วิธีการอ่านรูปอยู่ในบทที่ 4) ยังคงเห็นว่าการกระจายตัวของโครงสร้างที่ไม่ใช่ค่าตอบ (พื้นที่สีดำ) อยู่ในหลายตำแหน่ง ซึ่งเห็นได้ชัดกว่าปัญหา Royal Road ในรูปที่ 9.28 และสามารถอธิบายได้ด้วยเหตุผลเดียวกันดังที่กล่าวไปแล้ว ส่วนกรณีที่ค่าตอบไม่พบในรูปที่ 9.33 การลู่เข้าจะเห็นได้ชัดกว่าปัญหา Royal Road ในรูปที่ 9.29 เพราะในปัญหา Trap-5 มีการวางในระดับหน่วยการสร้าง สังเกตได้จากวิธีสีดำในแนวนอน (คู่จีนที่เป็น 0 ติดกัน) การลู่เข้าเริ่มเกิดขึ้นในรุ่นประชากรที่ 3 และชัดเจนขึ้นในรุ่นประชากรที่ 4 (เห็นแถบสีดำเส้นเล็กปรากฏชัดเจนในช่วงตำแหน่งต้น ๆ) และเกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 19 สอดคล้องกับรูปที่ 9.31 (ข)



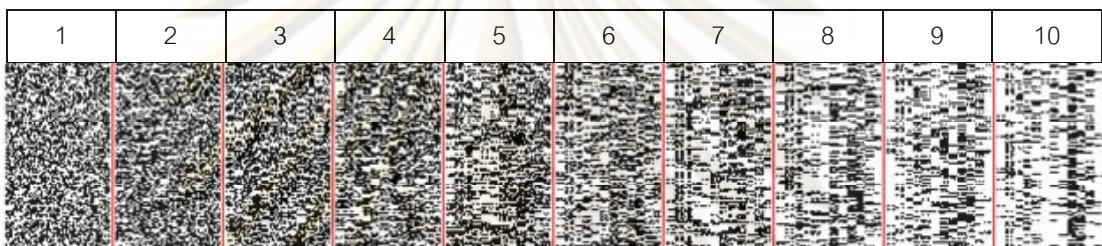
รูปที่ 9.30 ผลการทดสอบของ LZWGA-3 ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นของประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)



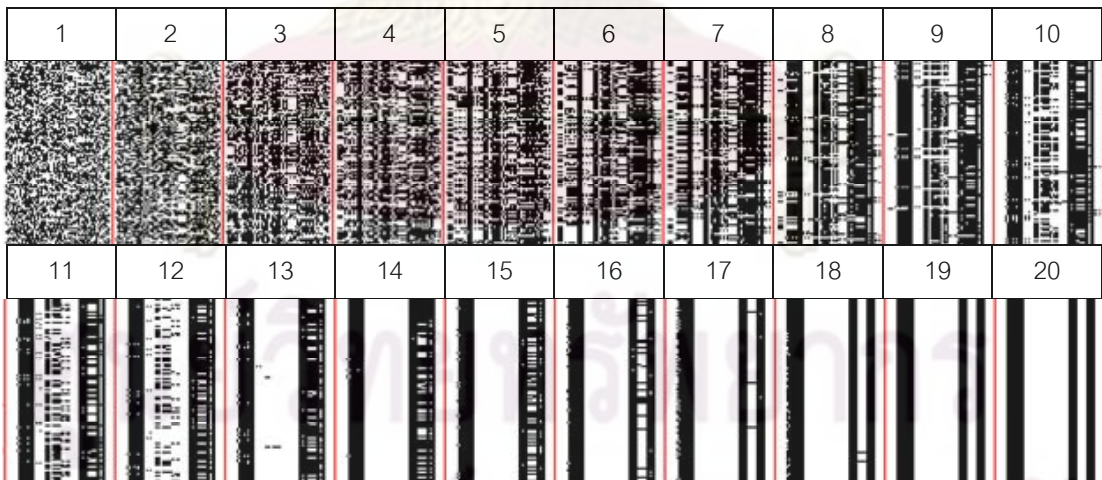
(ก)

(ข)

รูปที่ 9.31 ปริมาณของตำแหน่งที่ทำการไขว้เปลี่ยนได้ระหว่าง 2 โครโมโซมของ LZWGA-3 ใน ปัญหา Trap-5 60 บิต



รูปที่ 9.32 รูปแบบจีโนไทป์ของประชากรของ LZWGA-3 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0



รูปที่ 9.33 รูปแบบจีโนไทป์ของประชากรของ LZWGA-3 ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0

9.3.3 สรุปผลการทดลอง

ในส่วนนี้นำเสนอขั้นตอนวิธีเชิงพันธุกรรมแบบ LZWGA-3 ที่ใช้หลักการบีบอัดข้อมูลแบบ LZW ในการระบุหน่วยการสร้างแบบที่มีการแบ่งปันความรู้ระหว่างโครโมโซมที่ถูกเลือกทั้งหมด (มีจำนวนเท่ากับขนาดประชากร) และใช้วิธีการไขว้เปลี่ยนแบบจุดเดียวในการประกอบหน่วยการสร้างเข้าด้วยกัน ผลการทดสอบแสดงให้เห็นว่าวิธี LZWGA-3 มีประสิทธิภาพต่ำกว่าวิธี sGA, LZWGA-1 และ LZWGA-2 แต่ก็ยังสามารถหาคำตอบได้ภายในเงื่อนไขที่กำหนดในบทที่ 4 เหตุที่ทำให้วิธี LZWGA-3 มีประสิทธิภาพต่ำเป็นเพราะมีการแบ่งปันความรู้มากเกินไป ทำให้เกิดข้อจำกัดในการไขว้เปลี่ยนมากกว่าวิธี LZWGA-1 และวิธี LZWGA-2 เป็นอย่างมาก ซึ่งเป็นผลมาจากหลักการบีบอัดแบบ LZW จากผลการทดลองเมื่อเปรียบเทียบกับวิธี LZWGA-1 ที่ไม่มีการแบ่งปันความรู้ระหว่างโครโมโซมเลย และวิธี LZWGA-2 ที่มีการแบ่งปันความรู้ระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น ทำให้ได้ข้อสรุปว่าการแบ่งปันความรู้ในการระบุหน่วยการสร้างจะมีประโยชน์สูงสุดก็ต่อเมื่อเกิดขึ้นในระดับที่เหมาะสม ซึ่งไม่สามารถบอกได้แน่ชัดว่าควรเป็นเท่าใด เพราะขึ้นอยู่กับปัจจัยอื่น ๆ ด้วย เช่น วิธีการเข้ารหัสข้อมูล ลักษณะภูมิทัศน์ของค่าความเหมาะสม และตัวดำเนินการอื่น ๆ ที่ใช้ร่วมด้วย เป็นต้น เหมือนกับจำนวนประชากรที่เราไม่สามารถบอกได้แน่ชัดว่าควรใช้จำนวนเท่าใดนั่นเอง

ผลที่ได้จากการประยุกต์ใช้หลักการบีบอัดข้อมูลแบบ LZW ทำให้เห็นถึงความเป็นไปได้ที่จะนำหลักการนี้ไปใช้ในการระบุหน่วยการสร้าง แต่ยังคงมีความสงสัยว่าถ้านำหลักการบีบอัดแบบ LZW ไปใช้ในรูปแบบที่มีการแบ่งปันความรู้แบบรวมศูนย์ในขั้นตอนการระบุและการประกอบหน่วยการสร้างเหมือนกับวิธี CFT แล้วจะให้ผลเป็นเช่นใด จากการทดสอบที่ทำมาทั้งหมดทำให้สันนิษฐานได้ว่า เมื่อนำหลักการบีบอัดแบบ LZW ไปประยุกต์ใช้กับขั้นตอนวิธีที่มีการแบ่งปันความรู้แบบรวมศูนย์ในกระบวนการระบุและกระบวนการประกอบหน่วยการสร้าง จะได้ผลที่ดีกว่าหรือใกล้เคียงกับวิธี CFT เนื่องจากหลักการบีบอัดแบบ LZW เป็นวิธีแบบปรับตัวซึ่งจะมีส่วนช่วยให้กระบวนการระบุชั้นส่วนมีประสิทธิภาพดีขึ้น (มีการศึกษาข้อสันนิษฐานนี้ในบทถัดไป)

ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA)

ในบทนี้จะนำเสนอวิธีการระบุหน่วยการสร้างที่อาศัยหลักการบีบอัดแบบ LZW (แบบเดียวกับ LZWGA ในบทที่ 9) และการประกอบเข้าด้วยกันโดยอาศัยพื้นฐานการแบ่งปันความรู้แบบขั้นตอนวิธีประมาณการแจกแจง (แบบเดียวกับขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่มในบทที่ 7 และวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไขในบทที่ 8) เพื่อเป็นการยืนยันถึงผลของการประยุกต์ใช้ความรู้ที่ได้เรียนรู้มาจากบทก่อนหน้าทั้งหมด (ข้อดีของแต่ละขั้นตอนวิธีที่นำเสนอ) ว่าเป็นประโยชน์และสามารถนำมาใช้ร่วมกันได้อย่างมีประสิทธิภาพ (สามารถหาคำตอบได้ดีขึ้น) เปรียบเทียบผลการทดลองที่ได้กับขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (CFT)

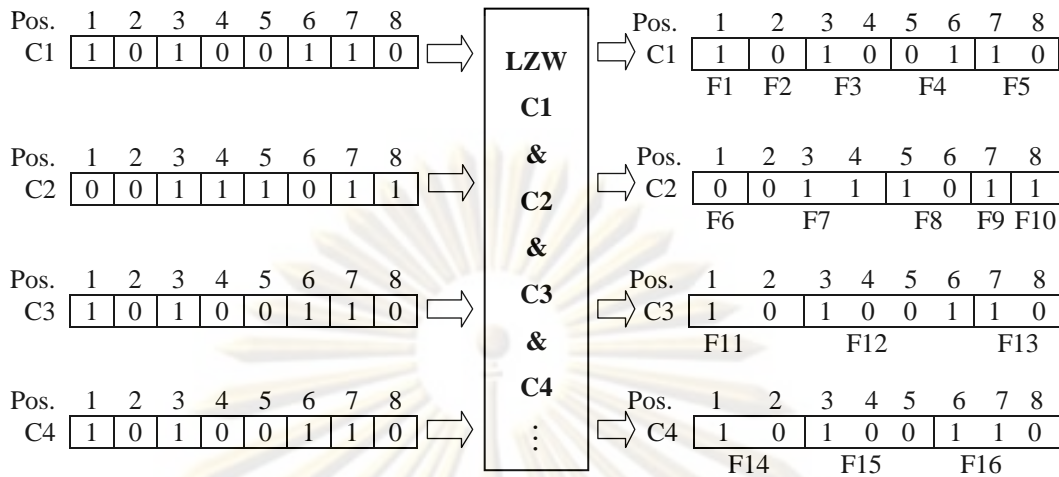
10.1 แนวคิดและหลักการ

ขั้นตอนวิธีที่นำเสนอในบทนี้ได้แรงบันดาลใจจากความรู้และจุดเด่นของแต่ละขั้นตอนวิธีที่นำเสนอไปแล้ว จากผลการทดลองในบทที่ 9 (LZWGA) เรื่องการประยุกต์ใช้หลักการบีบอัดแบบ LZW เพื่อใช้ในการระบุหน่วยการสร้างแบบชิ้นส่วนซึ่งมีจุดเด่นในเรื่องการทำงานแบบปรับตัว (Adaptive) เมื่อเปรียบเทียบกับการใช้ปรับเปลี่ยนแบบชิ้นส่วน ในบทที่ 5 (sGA-FC) ที่อยู่บนพื้นฐานการระบุชิ้นส่วนโดยใช้หลักการสารสนเทศร่วมกันต่ำสุด (Minimum Mutual Information) จะเห็นว่าวิธี LZWGA-2 มีประสิทธิภาพในการหาคำตอบเทียบเท่ากับ sGA-FC ซึ่งเป็นการแสดงให้เห็นถึงจุดเด่นหรือข้อดีของการเรียนรู้แบบปรับตัวที่ได้จากการบีบอัดแบบ LZW และจากผลการทดลองในบทที่ 7 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (RFT) ที่ใช้หลักการแบ่งปันความรู้แบบรวมศูนย์ในส่วนของประกอบชิ้นส่วนเข้าด้วยกัน ทำให้เห็นถึงข้อดีของการใช้หน่วยเก็บถาวร (Archive) ในการเก็บข้อมูลชิ้นส่วนไว้รวมกัน ต่อมาในบทที่ 8 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (CFT) ทำให้เห็นถึงประสิทธิภาพที่เพิ่มขึ้นเมื่อใช้หลักการสารสนเทศร่วมกันต่ำสุดในขั้นตอนการระบุชิ้นส่วนร่วมกับการแบ่งปันความรู้แบบรวมศูนย์ในขั้นตอนการประกอบชิ้นส่วนเข้าด้วยกัน ในบทนี้จึงได้มีแนวคิดในการประยุกต์ใช้หลักการบีบอัดข้อมูลแบบ LZW ในขั้นตอนการระบุชิ้นส่วนร่วมกับการแบ่งปันความรู้แบบรวมศูนย์ในขั้นตอนการประกอบชิ้นส่วนเข้าด้วยกัน

จากผลการทดลองในบทที่ 9 (LZWGA) ระดับของการแบ่งปันความรู้มีผลต่อประสิทธิภาพในการหาคำตอบ แต่ปัญหาคือเราไม่รู้ว่าการแบ่งปันความรู้ในระดับไหนถึงจะให้ได้ที่ดีที่สุด จากผลการทดลองในบทที่ 7 (RFT) และบทที่ 8 (CFT) ทำให้มีความเชื่อมั่นว่าโครโมโซมที่มีค่าความเหมาะสมสูง (สูงกว่าค่าความเหมาะสมเฉลี่ย) มีโครงสร้างย่อยของคำตอบอยู่จริงหรือสามารถนำไปสู่คำตอบที่ต้องการได้ ดังนั้นขั้นตอนวิธีที่นำเสนอในบทนี้จะใช้วิธีการเลือกแบบตัวดี

ที่สุดเป็นจำนวน n ตัว (n -Best Selection) เหมือนกับวิธี RFT และ CFT หลังจากเลือกโครโมโซมมาได้ n ตัวขั้นตอนถัดไปเป็นการระบุชิ้นส่วนโดยใช้หลักการบีบอัดแบบ LZW (ดังรูปที่ 10.3) จะทำการบีบอัดโครโมโซมที่ถูกเลือกมาทั้งหมดรวมกัน (ใช้พจนานุกรมอันเดียวกันแบบวิธี LZWGA-3 ในบทที่ 9) และนำชิ้นส่วนที่ระบุได้ทั้งหมดไปเก็บรวมกันไว้ในหน่วยเก็บถาวร (Archive) เรียงตามลำดับตำแหน่งที่เกิดขึ้นในโครโมโซม หลังจากนั้นเป็นขั้นตอนการประกอบชิ้นส่วนเข้าด้วยกันเพื่อสร้างโครโมโซมรุ่นใหม่ทีละตัว (ดังรูปที่ 10.4) โดยการสุ่มเลือกชิ้นส่วนจากหน่วยเก็บถาวรเริ่มจากตำแหน่งแรกของโครโมโซมชิ้นมาหนึ่งชิ้นส่วน จากนั้นจะทำการสุ่มเลือกชิ้นส่วนที่สองซึ่งมีตำแหน่งเริ่มต้นถัดจากตำแหน่งสุดท้ายของชิ้นส่วนแรก (เพื่อไม่ให้เหลื่อมกัน) และทำซ้ำจนกระทั่งครบความยาวโครโมโซม (ประกอบจากซ้ายไปขวา) (วิธีการเก็บชิ้นส่วนในหน่วยเก็บถาวรและการประกอบชิ้นส่วนเหมือนกับวิธีที่ใช้ใน CFT และ RFT) จากรูปที่ 10.1 เป็นตัวอย่างแสดงการดำเนินการของขั้นตอนวิธีที่นำเสนอ (LZWEDA) เริ่มจากกระบวนการหน่วยการสร้างในรูปมีโครโมโซมทั้งหมด 4 ตัวได้แก่ C1, C2, C3 และ C4 (ในรูปสมมติให้มีโครโมโซมมากกว่า 4 ตัวแต่แสดงให้เห็นเพียงแค่ 4 ตัว) โครโมโซมเหล่านี้จะถูกนำไปบีบอัดด้วยวิธี LZW แบบที่ใช้พจนานุกรมร่วมกัน (รูป 10.1 ส่วนบน) หลังจากผ่านการบีบอัดแล้วจะได้รูปแบบและตำแหน่งของชิ้นส่วนซึ่งจะถูกนำไปเก็บรวมกันไว้ในหน่วยเก็บถาวร (Composition Table) (รูป 10.1 ส่วนกลาง) เมื่อบีบอัดโครโมโซมที่ถูกเลือกมาจนครบแล้วจะเข้าสู่ขั้นตอนการประกอบชิ้นส่วน (รูป 10.1 ส่วนล่าง) โครโมโซมลูกตัวแรก (ตัวบน) ตำแหน่งแรกสุ่มได้ชิ้นส่วน F11 มีความยาว 2 บิต ชิ้นส่วนถัดไปจะเริ่มตำแหน่งที่ 3 สุ่มได้ชิ้นส่วน F21 มีความยาว 4 บิต ตำแหน่งที่สุ่มเลือกถัดไปคือตำแหน่งที่ 7 สุ่มได้ชิ้นส่วน F5 มีความยาว 2 บิต ครบความยาวโครโมโซมพอดี ส่วนโครโมโซมลูกตัวที่สอง (ตัวล่าง) ตำแหน่งแรกสุ่มได้ชิ้นส่วน F11 มีความยาว 1 บิต ตำแหน่งถัดไปสุ่มได้ชิ้นส่วน F24 มีความยาว 2 บิต ตำแหน่งถัดไปเป็นตำแหน่งที่ 4 ซึ่งไม่มีชิ้นส่วนเป็นสมาชิกอยู่เลยในกรณีนี้จะทำการสุ่มค่าสุ่มจากรูปแบบที่เป็นไปได้ทั้งหมด (ในกรณีการเข้ารหัสแบบเลขฐานสองรูปแบบที่เป็นไปได้คือ 0 กับ 1) เพื่อมาเติมในตำแหน่งที่ 4 เพียงตำแหน่งเดียว (ที่ว่าง) ถัดไปเป็นตำแหน่งที่ 5 สุ่มเลือกได้ชิ้นส่วน F18 มีความยาว 3 บิต ถัดไปเป็นตำแหน่งสุดท้าย (ตำแหน่งที่ 8) สุ่มเลือกได้ชิ้นส่วน F19 ครบตามความยาวโครโมโซม ในภาพรวมจะเห็นว่าวิธีที่นำเสนอแตกต่างจาก CFT ในส่วนของการระบุชิ้นส่วนเท่านั้น

LZW Fragment Identification



Composition Table (Archive)								
No. \ Position	1	2	3	4	5	6	7	8
1	F1	F2	F12		F4	F16	F5	F10
2	F6	F7	F15		F8		F9	
	F11						F13	
	F14							
⋮								

After compressed all individuals

Composition Table (Archive)								
No. \ Position	1	2	3	4	5	6	7	8
1	F1	F2	F12		F4	F16	F5	F19
2	F6	F7	F15		F8	F22	F9	F10
3	F11	F24	F21		F18		F13	
4	F14				F26		F23	
5	F17							
6	F20							
⋮								

Fragment composition (new individual)

Bit Position	1	2	3	4	5	6	7	8
New Chromosome 1	F11		F21			F5		

Bit Position	1	2	3	4	5	6	7	8
New Chromosome 2	F1	F24	R	F18	F19			

Random(0,1) if there is no member

รูปที่ 10.1 หลักการทำงานของขั้นตอนวิธีประมาณการแจกแจงแบบ LZW

(C หมายถึง Chromosome, F หมายถึง Fragment และ R หมายถึงค่าที่ได้จากการสุ่ม)

D คือ ประชากร

P คือ พจนานุกรมที่ได้จากวิธีบีบอัดแบบ LZW

T คือ หน่วยเก็บชิ้นส่วนแบบถาวร (Archive)

ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA)

$D_0 \leftarrow$ สร้างผลเฉลย R (Individuals) (ประชากรเริ่มต้น) โดยการสุ่ม
วนซ้ำสำหรับ $g = 1, 2, \dots$, จนกระทั่งเข้าเกณฑ์การหยุด

$D_{g-1}^N \leftarrow$ เลือกผลเฉลย $N < R$ จาก D_{g-1} ที่สอดคล้องกับวิธีการเลือก

$P \leftarrow$ สร้างพจนานุกรมจากข้อมูลที่ถูกเลือกโดยใช้ขั้นตอนวิธี LZW

$T \leftarrow$ สร้างหน่วยเก็บชิ้นส่วนถาวร จาก P โดยเก็บเฉพาะชิ้นส่วนที่
เกิดขึ้นจริงในผลเฉลย (ไม่ได้เก็บทั้งพจนานุกรม)

$D_g^N \leftarrow$ ชักตัวอย่างผลเฉลย R_g (ประชากรรุ่นใหม่) จาก T

$D_g^N \leftarrow$ นำผลเฉลยจากทุกตัวจาก D_g^N ไปทำการผ่าเหล่า (Mutation)
ด้วยวิธีแบบ sGA (เป็นทางเลือก แต่งานวิจัยนี้ไม่ใช้)

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 10.2 รหัสเทียมของขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA)

R คือ ผลเฉลย (โครโมโซม)

N คือ จำนวนผลเฉลยที่ถูกเลือก

P คือ พจนานุกรมที่ได้จากวิธีบีบอัดแบบ LZW

ขั้นตอนวิธีการระบุชิ้นส่วน (สร้างพจนานุกรม)

วนซ้ำสำหรับ $i = 1$ ถึง N ทำ

$P \leftarrow$ นำ R_i ไปทำการบีบอัดแบบ LZW โดยใช้
พจนานุกรมร่วมกันทั้งหมด

จบการวนซ้ำ

จบขั้นตอนวิธี

รูปที่ 10.3 รหัสเทียมของวิธีการระบุชิ้นส่วนสำหรับ LZWEDA

R คือ ผลเฉลย (โครโมโซม)
 F_i คือ เซตของชิ้นส่วนย่อย (ที่ได้จากวิธีการบีบอัดแบบ LZW)
 เริ่มต้นที่ตำแหน่ง i .
 C คือ ชิ้นส่วนย่อย (แบบร่วมกัน และแบบไม่ร่วมกัน)
 L คือ ความยาวของผลเฉลย
ขั้นตอนวิธีการประกอบชิ้นส่วน
 {การประกอบผลเฉลยใหม่ R }
 $i = 1$ {ตำแหน่งแรก}
วนซ้ำ ในขณะที่ $i < L$ ทำ
 ถ้า $F_i \neq \emptyset$ แล้ว {ถ้ามีชิ้นส่วนย่อยตั้งแต่หนึ่งชิ้น}
 $C \leftarrow$ สุ่มเลือก(F_i)
 $R \leftarrow R + C$ {นำชิ้นส่วนมาเรียงต่อกันตามลำดับ}
 $i \leftarrow i + \text{size}(C)$ {หาตำแหน่งเริ่มต้นถัดไป}
 นอกจากนั้น {ถ้าไม่มีสมาชิกอยู่เลย}
 $R \leftarrow$ สุ่มเลือก(0,1) {สุ่มจากค่า 0 หรือ 1}
 $i \leftarrow i + 1$
จบเงื่อนไข
จบการวนซ้ำ
จบขั้นตอนวิธี

รูปที่ 10.4 รหัสเทียมของวิธีการประกอบชิ้นส่วนสำหรับ LZWEDA

วิธีที่นำเสนอในบทนี้มีจุดเด่นในเรื่องการแบ่งปันความรู้ใน 2 ระดับเช่นเดียวกับวิธี CFT โดยแบ่งออกเป็น การแบ่งปันความรู้ในระดับการระบุชิ้นส่วนที่ใช้หลักการบีบอัดแบบ LZW แบบใช้พจนานุกรมร่วมกัน และการแบ่งปันความรู้ในระดับการประกอบชิ้นส่วนที่ใช้หน่วยเก็บถาวรเพื่อเก็บชิ้นส่วนรวมกัน เมื่อพิจารณาในมุมมองของการสำรวจและการแสวงหาประโยชน์จะเห็นว่าในขั้นตอนการระบุชิ้นส่วนเป็นการแสวงหาประโยชน์มากกว่าการสำรวจ เนื่องจากวิธีการบีบอัดแบบ LZW เป็นวิธีที่อาศัยข้อมูล (รูปแบบคำ) ที่เคยมีมาก่อนหน้าเป็นตัวอ้างอิง หรือมองได้ว่าเป็นการเรียนรู้จากหลายโครโมโซมร่วมกันโดยใช้พจนานุกรมอันเดียวกัน และในขั้นตอนการประกอบชิ้นส่วนเป็นการสำรวจมากกว่าการแสวงหาประโยชน์ เนื่องจากเป็นการพยายามประกอบ

ขึ้นส่วนที่มีหลายรูปแบบเข้าด้วยกันโดยไม่มีความลำเอียงในการเลือกขึ้นส่วน (ทุกขึ้นส่วนที่เริ่มต้นในตำแหน่งเดียวกันมีโอกาสถูกเลือกเท่ากันหมด) จากการถ่วงดุลที่เกิดขึ้นระหว่างสองกระบวนการนี้จะช่วยให้การค้นหาคำตอบดำเนินไปอย่างมีประสิทธิภาพ หรือมองในภาพรวมอีกมุมหนึ่งได้ว่าการบีบอัดเป็นวิธีการค้นหา (ระบุ) หน่วยการสร้างในรูปแบบขึ้นส่วนที่มีอยู่ในโครโมโซม การใช้หน่วยเก็บถาวรเป็นการรักษาขึ้นส่วนให้คงอยู่ และการประกอบขึ้นส่วนเข้าด้วยกันเป็นการพยายามแลกเปลี่ยนขึ้นส่วนที่มีความหลากหลาย (ที่ระบุได้) หรือเป็นการสำรวจโครงสร้างรูปแบบโครโมโซมที่ดีขึ้น (ใกล้เคียงโครงสร้างคำตอบมากขึ้น)

เพื่อให้มีความเข้าใจในการประยุกต์ใช้กระบวนการบีบอัดแบบ LZW ที่ใช้ในบทนี้มากขึ้น จะมีการอธิบายเพิ่มเติมในส่วนของการสร้างพจนานุกรมและการนำไปใช้ ตัวอย่างจากรูปที่ 10.5 (ก) กำหนดให้มีโครโมโซมที่ใช้ในการบีบอัดทั้งหมด 4 ตัว ได้แก่ C1, C2, C3 และ C4 โครโมโซมแต่ละตัวมีความยาว 10 บิต เราจะทำการบีบอัดโครโมโซมทั้ง 4 ตัวร่วมกันโดยใช้พจนานุกรมอันเดียวกัน การบีบอัดเริ่มจากโครโมโซม C1 ไปจนถึง C4 ตามลำดับ เมื่อบีบอัดเสร็จแล้วจะได้พจนานุกรมดังรูปที่ 10.5 (ข) พจนานุกรมประกอบไปด้วยข้อมูลรหัส (Code) และคำ (Word) ส่วนรหัสใช้อ้างถึงตำแหน่งของคำที่อยู่ในพจนานุกรม และส่วนคำเป็นข้อมูลขึ้นส่วนที่ได้จากการบีบอัดโครโมโซม (เป็นข้อมูลโครงสร้างคู่จีนของโครโมโซม) ในรูป 10.5 (ก) โครโมโซมแต่ละตัวเมื่อเข้าผ่านการบีบอัดแล้วจะได้รหัสอ้างอิงรูปแบบขึ้นส่วนสำหรับแต่ละโครโมโซมตามตำแหน่งที่เกิดขึ้น (ตำแหน่งที่ระบุได้) จะถือว่าคำที่อยู่ในพจนานุกรมทั้งหมดเป็นขึ้นส่วนที่ระบุได้ และจะนำไปเก็บไว้ในหน่วยเก็บถาวรเพื่อใช้ในการประกอบขึ้นส่วนเข้าด้วยกัน โดยมีเงื่อนไขว่าข้อมูลขึ้นส่วนที่อยู่ในหน่วยเก็บถาวรจะต้องมีรูปแบบที่ไม่ซ้ำกัน ณ ตำแหน่งเดียวกัน ดังรูปที่ 10.6 หน่วยเก็บถาวรจะประกอบไปด้วยข้อมูลตำแหน่งของขึ้นส่วนที่อ้างอิงจากตำแหน่งที่ระบุได้ในโครโมโซม และรหัสที่ได้จากพจนานุกรมซึ่งอ้างอิงคำหรือรูปแบบขึ้นส่วนที่ระบุได้ เหตุที่ไม่เก็บข้อมูลขึ้นส่วนในหน่วยเก็บถาวรโดยตรงเนื่องจากการลดการใช้พื้นที่หน่วยความจำ และเวลาที่ใช้ในการประมวลผล (เป็นเหตุผลในเรื่องการเขียนโปรแกรม) จากรูปที่ 10.6 แสดงขึ้นส่วนโดยเรียงลำดับตามตำแหน่งที่เกิดขึ้น (ก่อนหรือหลัง) ตามลำดับ และแสดงให้เห็นโดยอ้างอิงกับลำดับของโครโมโซมที่เข้าสู่กระบวนการบีบอัด (จาก C1 ถึง C4 ตามลำดับ) เพื่อให้เห็นภาพว่าขึ้นส่วนที่ได้จากโครโมโซมตัวท้าย ๆ เช่น C3 และ C4 จะมีขนาดใหญ่ขึ้น (ยาวขึ้น) เรื่อย ๆ โดยเฉลี่ย

ในส่วนวิธีการประกอบขึ้นส่วนเข้าด้วยกันเพื่อสร้างโครโมโซมรุ่นใหม่แสดงดังรูปที่ 10.7 ซึ่งเป็นตัวอย่างให้เห็นถึงการดำเนินการจริงที่เกิดขึ้น (เมื่อเทียบกับการประกอบเข้าด้วยกันในรูปที่ 10.1) วิธีการประกอบทำได้โดยการสุ่มเลือกรหัสที่เก็บอยู่ในหน่วยเก็บถาวรจากรูปที่ 10.6 เริ่มจากตำแหน่งแรกมีขึ้นส่วนที่ระบุได้จำนวน 4 ขึ้น ในรูปที่ 10.7 ตำแหน่งแรกสุ่มได้ขึ้นส่วนรหัส

หมายเลข 6 เราจะทำการย้ายไปดูในพจนานุกรมจากรูปที่ 10.5 (ข) ชั้นส่วนรหัสหมายเลข 6 มีรูปแบบเป็น '100' (ขนาด 3 บิต) ดังนั้นตำแหน่งต่อไปที่ต้องทำการสุ่มเลือกชั้นส่วนคือตำแหน่งที่ 4 จากหน่วยเก็บถาวรในรูปที่ 10.6 ตำแหน่งที่ 4 มีชั้นส่วนให้เลือกจำนวน 2 ชั้น ในรูปที่ 10.7 สุ่มได้ชั้นส่วนรหัสหมายเลข 0 เมื่อไปดูจากพจนานุกรมในรูปที่ 10.5 (ข) ชั้นส่วนรหัสหมายเลข 0 มีรูปแบบเป็น '0' (ขนาด 1 บิต) ดังนั้นตำแหน่งต่อไปที่ต้องทำการสุ่มเลือกคือตำแหน่งที่ 5 และจะทำซ้ำจนกระทั่งครบความยาวโครโมโซม จากรูปที่ 10.7 สุดท้ายแล้วจะได้โครโมโซมลูกมีหน้าตาเป็น 1000011100 (ขนาด 10 บิต)

ตัวอย่างวิธีการสร้างพจนานุกรมสำหรับวิธี LZWEDA

ข้อมูลนำเข้า (โครโมโซมขนาด 10 บิต C1-C4)								
ตำแหน่ง	C1	รหัส	C2	รหัส	C3	รหัส	C4	รหัส
1	1	1	0		1		0	
2	0	0	1		0		0	10
3	1		1	7	0	6	0	
4	0	2	0	0	0		1	
5	1	1	0		1		0	
6	1		1	3	1	7	0	12
7	0	2	0		1		1	
8	0		1		0		1	5
9	1	3	0	11	1	4	0	
10	1	1	0	0	1	1	0	10

(ก)

พจนานุกรม							
รหัส	ค่า	รหัส	ค่า	รหัส	ค่า	รหัส	ค่า
0	0	6	100	12	0100	18	000
1	1	7	011	13	0	19	01001
2	10	8	1	14	1000	20	110
3	01	9	0110	15	0111	21	00
4	101	10	00	16	1011		
5	11	11	010	17	1		

(ข)

รูปที่ 10.5 ตัวอย่างการเข้ารหัสโครโมโซมสี่ตัวแบบ LZW (ก) ข้อมูลนำเข้าของโครโมโซมแต่ละตัวเรียงจากซ้ายไปขวา และจากบนลงล่าง (ข) พจนานุกรมของข้อมูลนำเข้าที่ผ่านการบีบอัดแบบ LZW โดยใช้พจนานุกรมร่วมกันทั้งหมด

หน่วยเก็บถาวร				
ตำแหน่ง	FC-1	FC-2	FC-3	FC-4
1	1	7	6	10
2	0			
3	2			12
4		0	7	
5	1	3		
6	2			
7		11	4	5
8	3			
9				10
10	1	0		

รูปที่ 10.6 หน่วยเก็บถาวรเก็บเฉพาะรหัสปีฉบับที่ไม่ว่างในแต่ละตำแหน่งเท่านั้น และเรียงตามลำดับตำแหน่ง (FC หมายถึง ชั้นส่วนที่ระบุได้จากแต่ละโครโมโซม)

Code :	6	0	3	5	10
Code → Word	↓	↓	↓	↓	↓
Offspring :	100	0	01	11	00

รูปที่ 10.7 สร้างโครโมโซมใหม่ด้วยวิธีสุ่มเลือกรหัสจากหน่วยเก็บถาวรตามลำดับตำแหน่งเริ่มจากตำแหน่งแรกไปจนถึงตำแหน่งสุดท้าย และแปลงรหัสที่ได้กลับเป็นคำ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

10.2 การทดลองและผลการทดลอง

ในส่วนนี้จะกล่าวถึงรายละเอียดของวิธีการทดลองเพิ่มเติมจากที่ได้กล่าวไว้แล้วส่วนหนึ่งในบทที่ 4 พร้อมทั้งแสดงผลการทดลองของขั้นตอนวิธีที่นำเสนอ และวิจารณ์ผลการทดลองในมุมมองที่นำเสนอ มีรายละเอียดดังต่อไปนี้

วิธีที่ใช้เปรียบเทียบ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม (sGA)
การเลือกสรร:	วิธีการเลือกแบบการแข่งขันกีฬา มีระดับการแข่งขันเป็น 4
การไขว้เปลี่ยน:	การไขว้เปลี่ยนแบบจุดเดียว มีอัตราการไขว้เปลี่ยนเท่ากับ 1.0
การผ่าเหล่า:	ไม่มี

วิธีที่ใช้เปรียบเทียบ

ขั้นตอนวิธี:	ขั้นตอนวิธีการแตกเป็นชิ้นส่วนอย่างมีเงื่อนไข (CFT)
การเลือกสรร:	วิธีการเลือกแบบ n ตัวที่ดีที่สุด (n -Best) ค่า n คือค่า #sub ที่แสดงในตารางที่ 10.1 และ 10.2
การระบุ:	วิธีข้อมูลเสนอร่วมกันและไม่ร่วมกันต่ำสุดแบบชิ้นส่วน
การประกอบ:	วิธีชักตัวอย่างจากหน่วยเก็บถาวรตามลำดับตำแหน่งปรากฏ
การผ่าเหล่า:	ไม่มี

วิธีที่นำเสนอ

ขั้นตอนวิธี:	ขั้นตอนวิธีเชิงพันธุกรรม ที่มีการระบุขอบเขตการไขว้เปลี่ยน (LZWEDA)
การเลือกสรร:	วิธีการเลือกแบบ n ตัวที่ดีที่สุด (n -Best) ค่า n คือค่า #sub ที่แสดงในตารางที่ 10.1
การระบุ:	วิธีการบีบอัดแบบ LZW
การประกอบ:	วิธีชักตัวอย่างจากหน่วยเก็บถาวรตามลำดับตำแหน่งปรากฏ
การผ่าเหล่า:	ไม่มี

จำนวนประชากรที่ใช้ และผลการทดลองสำหรับแต่ละปัญหาแสดงดังตารางที่ 10.1 และ 10.2

ตารางที่ 10.1 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ sGA และ LZWEDA

Parameters		Success 30 run				Fail 30 run	
		sGA		LZWEDA		sGA	LZWEDA
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
				#sub			#sub
RoyalRoad	64	1,200	11,920	500	4,500	200	100
				100			50
Trap-5	60	2,300	28,367	1,100	10,487	300	200
				100			100

หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ตารางที่ 10.2 จำนวนประชากรและจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของ CFT และ LZWEDA

Parameters		Success 30 run				Fail 30 run	
		CFT		LZWEDA		CFT	LZWEDA
Problems	Problem size	#pop.	#FEs.	#pop.	#FEs.	#pop.	#pop.
		#sub		#sub		#sub	#sub
RoyalRoad	64	2,000	13,867	500	4,500	200	100
		200		100		100	50
Trap-5	60	1,000	7,834	1,100	10,487	100	200
		100		100		50	100

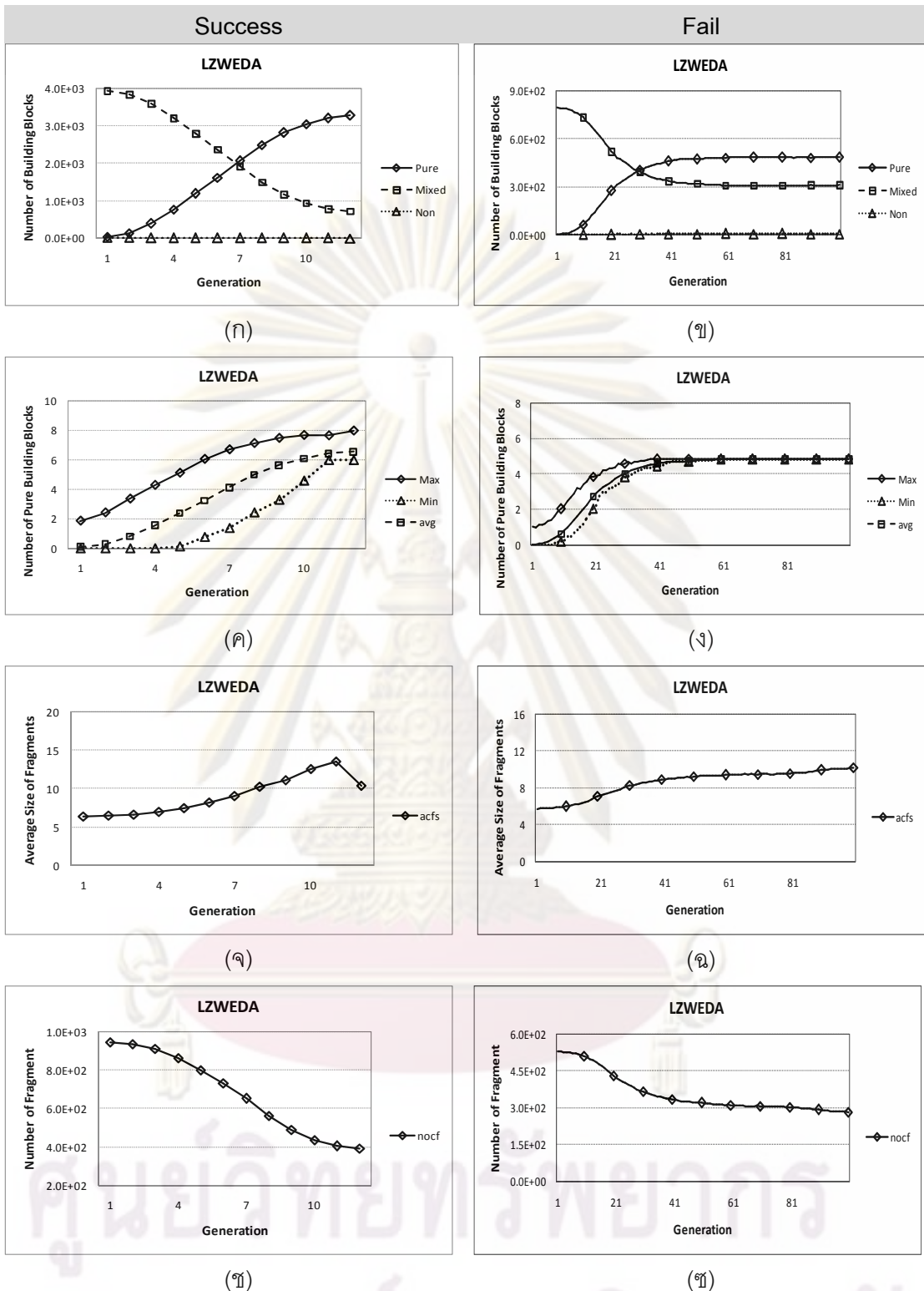
หมายเหตุ เงื่อนไขอื่น ๆ ในการทดลองอยู่ในบทที่ 4

ผลการทดสอบในตารางที่ 10.1 เมื่อเทียบกับวิธี sGA ที่มีหลักการในการดำเนินการต่างจากวิธี LZWEDA โดยสิ้นเชิงทั้งในขั้นตอนการระบุและการประกอบหน่วยการสร้าง วิธี LZWEDA มีค่า FEs น้อยกว่าทั้งในปัญหา Royal Road และปัญหา Trap-5 แสดงให้เห็นว่าวิธี LZWEDA มีประสิทธิภาพในการแก้ปัญหาดีกว่า และเมื่อพิจารณาถึงจำนวนประชากรที่ใช้ในการ

แก้ปัญหา วิธี LZWEDA ใช้จำนวนประชากรน้อยกว่า sGA ทั้งสองปัญหา เป็นตัวบ่งชี้ว่าวิธี LZWEDA ต้องการระดับความหลากหลายในการแก้ปัญหาต่ำกว่า sGA

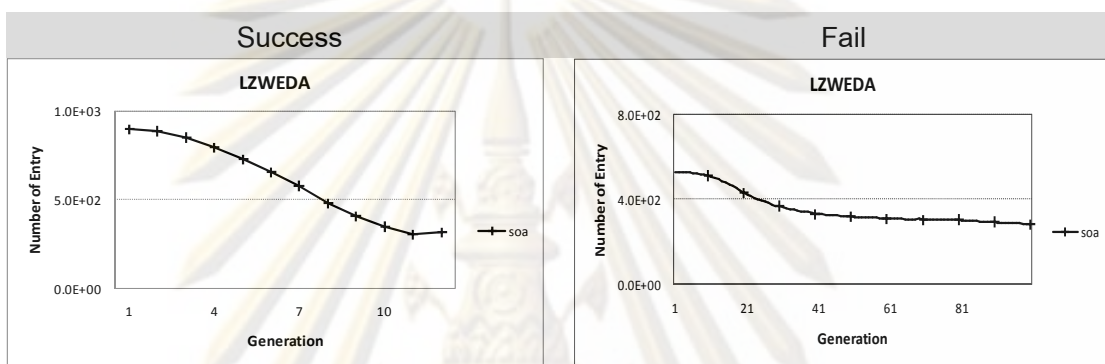
จากตารางที่ 10.2 เป็นการเปรียบเทียบผลกับวิธี CFT ที่มีการแบ่งปันความรู้แบบรวมศูนย์ในขั้นตอนการระบุชิ้นส่วนและการประกอบชิ้นส่วนเหมือนกัน แต่ต่างกันขั้นตอนการระบุชิ้นส่วนที่วิธี LZWEDA ใช้หลักการบีบอัดข้อมูลแบบ LZW แทนหลักการสารสนเทศร่วมต่ำสุดที่ใช้ในวิธี CFT เมื่อพิจารณาในมุมมองของประสิทธิภาพในการแก้ปัญหา สำหรับปัญหา Royal Road วิธี LZWEDA มีค่า FEs น้อยกว่าวิธี CFT แต่ในปัญหา Trap-5 วิธี LZWEDA มีค่า FEs มากกว่าวิธี CFT เป็นการแสดงให้เห็นว่าวิธี LZWEDA มีประสิทธิภาพในการแก้ปัญหา Royal Road ได้ดีกว่าวิธี CFT แต่มีประสิทธิภาพในการแก้ปัญหา Trap-5 ต่ำกว่าวิธี CFT และเมื่อพิจารณาถึงระดับความหลากหลายต่ำสุดที่ต้องใช้ในการแก้ปัญหา Royal Road วิธี LZWEDA ต้องการระดับความหลากหลายต่ำสุดน้อยกว่าวิธี CFT และสำหรับปัญหา Trap-5 วิธี LZWEDA ต้องการระดับความหลากหลายต่ำสุดมากกว่าวิธี CFT

ผลการทดลองในมุมมองของหน่วยการสร้างสำหรับปัญหา Royal Road ในรูปที่ 10.8 (ก-ง) มีรูปแบบกราฟเหมือนกับวิธี LZWGA ที่ใช้วิธีการระบุชิ้นส่วนเหมือนกันในบทที่ 9 แต่วิธี LZWEDA ใช้จำนวนประชากรน้อยกว่ามาก แสดงให้เห็นถึงระดับความหลากหลายที่มีในวิธี LZWEDA ที่เกิดขึ้นจากวิธีการระบุชิ้นส่วนที่มีความพิเศษตรงที่ลำดับการบีบอัดก่อนหลังให้ผลในการบีบอัดต่างกัน ภายใต้เงื่อนไขที่ว่าถ้าโครโมโซมที่ใช้ในการบีบอัดมีหน้าตาไม่เหมือนกันทุกตัว และโครโมโซมที่แตกต่างเหล่านี้มีลำดับในการบีบอัดไม่ซ้ำกันในแต่ละรอบ สมมติให้โครโมโซม C1 และ C2 มีโครงสร้างทางพันธุกรรมต่างกัน และมีขนาดความยาวเท่ากันคือ 10 บิต เมื่อนำโครโมโซมทั้งสองตัวนี้ไปบีบอัดแบบ LZW โดยเริ่มบีบอัดจาก C1 ก่อน C2 จะให้ผลในการบีบอัดต่างจากการเริ่มบีบอัดจาก C2 ก่อน C1 ในกรณีที่หาคำตอบพบและหาคำตอบไม่พบจากรูปที่ 10.8 (ก) และรูปที่ 10.8 (ข) ตามลำดับ มีการแข่งขันกันเฉพาะ Mixed BBs และ Pure BBs เท่านั้น ในกรณีที่หาคำตอบไม่พบจากรูปที่ 10.8 (ข และ ง) จะลู่อู่เข้าเกิดขึ้นซ้ำกว่าวิธี LZWGA เนื่องจากวิธีการประกอบชิ้นส่วนเข้าด้วยกันทำให้เกิดความหลากหลายสูง ในมุมมองของชิ้นส่วนจากรูปที่ 10.8 (จ) ในรุ่นประชากรสุดท้ายมีค่าลดลงเป็นเพราะว่าผลการทดลองที่นำมาแสดงได้มาจากการเฉลี่ยผลการทดสอบ 30 รอบ ดังนั้นอาจมีบางรอบที่ใช้รุ่นประชากรเกินกว่ารอบอื่นทำให้ผลที่ได้แตกต่างอย่างชัดเจนเพราะไม่ได้ถูกหารเฉลี่ยกับรอบอื่น ผลในกรณีที่หาคำตอบไม่พบจากรูปที่ 10.8 (ฉ และ ช) ยืนยันให้เห็นว่าวิธีการระบุและการประกอบเข้าด้วยกันทำให้เกิดความหลากหลายในระดับที่สูง เพราะขนาดโดยเฉลี่ยของชิ้นส่วนยังคงเพิ่มขึ้น (เล็กน้อย) แม้ว่าจะเกิดการลู่อู่เข้าตั้งแต่รุ่นประชากรที่ 40 แล้วก็ตามซึ่งต่างจากวิธี LZWGA อย่างเห็นได้ชัด



รูปที่ 10.8 ผลการทดสอบของ LZWEDA ในปัญหา Royal Road 64 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ซ)

ปริมาณชิ้นส่วนทั้งหมดที่เก็บอยู่ในหน่วยเก็บถาวรมีค่าใกล้เคียงกับจำนวนชิ้นส่วนที่ระบุได้ในรูปที่ 10.8 (ช-ช) เนื่องจากชิ้นส่วนในหน่วยเก็บถาวรของวิธี LZWEDA คือค่าในพจนานุกรมที่ได้จากการบีบอัดแบบ LZW นั่นเอง โดยปกติการบีบอัดแบบ LZW ไม่ทำให้เกิดชิ้นส่วนที่มีรูปแบบซ้ำกันในพจนานุกรม แต่ในการประยุกต์ใช้จริงจะทำการบีบอัดที่ละโครโมโซม (ไม่ได้นำโครโมโซมทุกตัวมาเชื่อมต่อเป็นเส้นเดียวกันทั้งหมด) จึงทำให้ชิ้นส่วนสุดท้ายที่ระบุได้จากแต่ละโครโมโซมมีรูปแบบซ้ำกับค่าที่มีอยู่แล้วในพจนานุกรม ดังนั้นจะไม่มีค่าใหม่เกิดขึ้นในพจนานุกรมทำให้ชิ้นส่วนที่มีอยู่ในหน่วยเก็บถาวรมีจำนวนน้อยกว่าชิ้นส่วนที่ระบุได้เสมอ



(ก)

(ข)

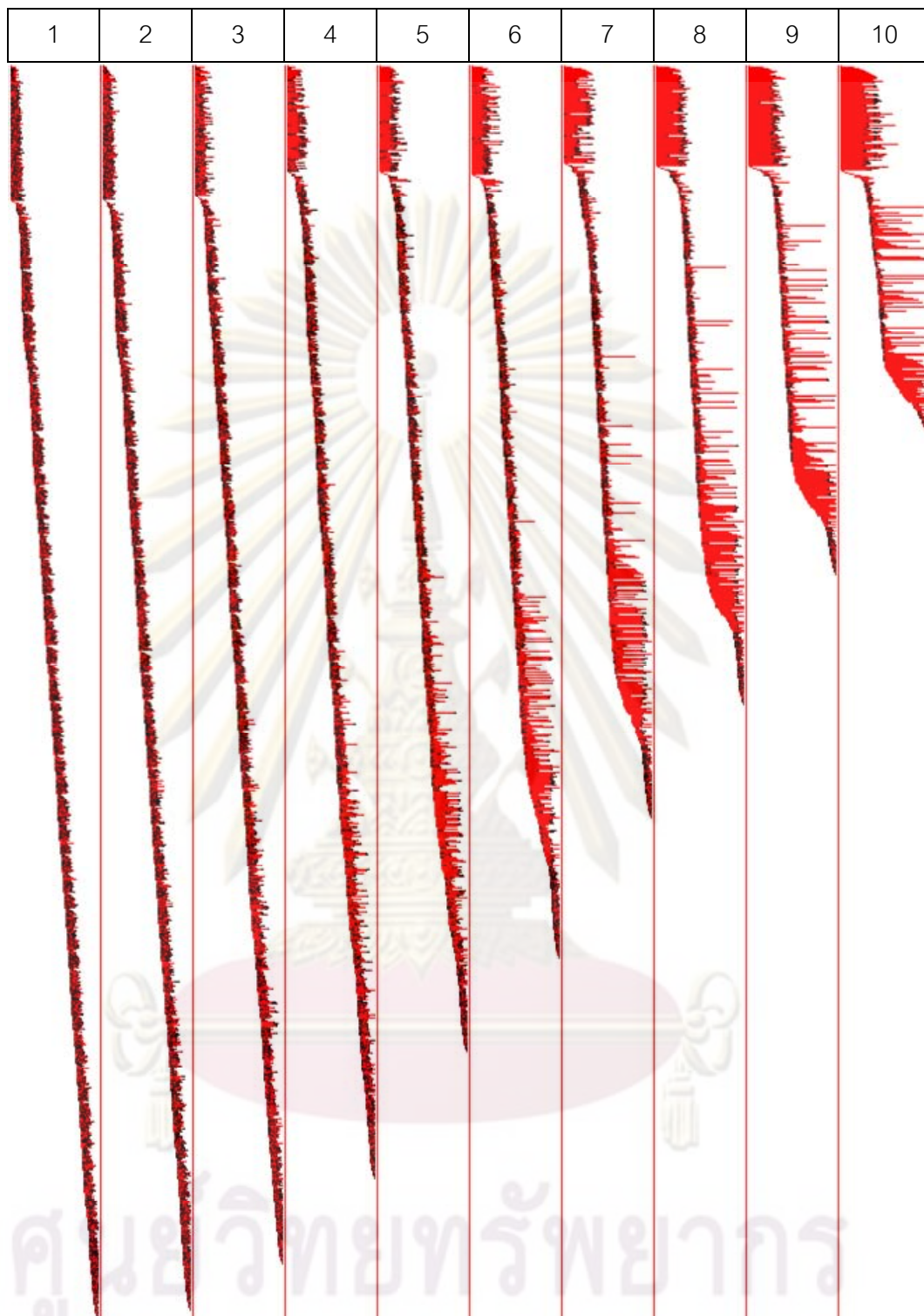
รูปที่ 10.9 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ LZWEDA ในปัญหา Royal Road 64 บิต

เมื่อพิจารณาถึงรูปแบบโครงสร้างทางพันธุกรรมของชิ้นส่วนที่เก็บอยู่ในหน่วยเก็บถาวรจากรูปที่ 10.10 และรูปที่ 10.11 (วิธีการอ่านรูปอยู่ในบทที่ 4) ในช่วงรุ่นประชากรแรก ๆ ที่ยังไม่เกิดการลู่เข้าขนาดโดยเฉลี่ยของชิ้นส่วนในแต่ละตำแหน่งจะมีค่าใกล้เคียงกัน และมีการกระจายตัวของชิ้นส่วนค่อนข้างสม่ำเสมอในทุกตำแหน่ง และเมื่อเริ่มเกิดการลู่เข้าขนาดของชิ้นส่วนในตำแหน่งแรกจะเริ่มโตขึ้นเป็นเพราะว่าวิธี LZW ทำการบีบอัดข้อมูลจากตำแหน่งแรก (บิตซ้ายสุด) ไปตำแหน่งสุดท้าย (บิตขวาสุด) ตามลำดับ และบีบอัดที่ละโครโมโซม (ต้องเริ่มใหม่ที่ตำแหน่งแรกของโครโมโซมตัวถัดไปทุกครั้ง) ดังนั้นในช่วงบิตต้น ๆ รูปแบบของคู่เงินจะซ้ำกับค่าที่มีอยู่ในพจนานุกรมเป็นส่วนใหญ่ ทำให้ชิ้นส่วนแรกมีขนาดโตขึ้นอย่างต่อเนื่อง สำหรับกรณีที่หาคำตอบพบในรูปที่ 10.10 จะสังเกตเห็นว่านอกจากตำแหน่งแรกแล้วชิ้นส่วนที่เริ่มจากตำแหน่งกลาง ๆ ของโครโมโซมเป็นต้นไปก็เริ่มมีขนาดโตขึ้นเช่นกัน เป็นเพราะเมื่อเริ่มเกิดการลู่เข้าโครงสร้างทางพันธุกรรมของโครโมโซมแต่ละตัวที่ถูกเลือกจะมีรูปแบบคล้ายกัน และเมื่อทำการบีบอัดโครโมโซมไปได้จำนวนหนึ่งรูปแบบค่าในพจนานุกรมจะเพิ่มขึ้นจนมีความหลากหลายมากพอ ซึ่งทำให้โครโมโซมตัวหลัง ๆ ที่เข้าสู่กระบวนการบีบอัดมีโอกาสที่จะมีรูปแบบของคู่เงินซ้ำกับค่าใน

พจนานุกรมสูงขึ้น หรือมองได้ว่าเมื่อทำการบีบอัดไปเรื่อย ๆ ค่าในพจนานุกรมจะเริ่มยาวขึ้น สมมติว่ายาวขึ้นจนกระทั่งมีขนาดโดยเฉลี่ยเท่ากับหนึ่งในสามของความยาวโครโมโซม ดังนั้นโครโมโซมตัวหลัง ๆ ที่เข้าสู่กระบวนการบีบอัดจะมีชิ้นส่วนที่ยาวขึ้นด้วย และเมื่อชิ้นส่วนแรกที่ระบุได้ในโครโมโซมมีขนาดยาว ดังนั้นชิ้นส่วนถัดไปที่ระบุได้ก็คือชิ้นส่วนที่ถัดจากชิ้นส่วนแรกซึ่งก็มีโอกาสซ้ำกับรูปแบบค่าที่มีในพจนานุกรมเท่ากับชิ้นส่วนแรกเช่นกัน ด้วยเหตุนี้จึงทำให้ชิ้นส่วนที่ระบุได้มีความยาวสลับกันในแต่ช่วงตำแหน่ง สำหรับกรณีที่ทำคำตอบไม่พบในรูปที่ 10.11 ไม่สามารถสังเกตเห็นขนาดชิ้นส่วนที่ต่างกันได้ในตำแหน่งอื่นนอกจากตำแหน่งแรก เป็นเพราะใช้จำนวนประชากรน้อยทำให้ค่าในพจนานุกรมมีขนาดสั้น ส่งผลให้มีความหลากหลายไม่เพียงพอเนื่องจากมีรูปแบบชิ้นส่วนที่น้อยเกินไปจึงไม่สามารถหาคำตอบพบ

ในรูปที่ 10.12 และรูปที่ 10.13 เป็นรูปแบบโครงสร้างทางพันธุกรรมของประชากรที่สุ่มสร้างขึ้นจากหน่วยเก็บถาวรในรูปที่ 10.10 และรูปที่ 10.11 ตามลำดับ (วิธีการอ่านรูปอยู่ในบทที่ 4) ซึ่งสะท้อนให้เห็นถึงโครงสร้างของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรได้เป็นอย่างดี ในกรณีที่หาคำตอบพบในรูปที่ 10.12 เริ่มมีการลู่เข้าในรุ่นประชากรที่ 3 ในช่วงตำแหน่งต้น ๆ และในรุ่นประชากรที่ 4 เริ่มมีการลู่เข้าในช่วงตำแหน่งท้าย ๆ การลู่เข้าในทั้งสองตำแหน่งเริ่มปรากฏชัดตั้งแต่ประชากรรุ่นที่ 5 เป็นต้นไป จนพบคำตอบ ส่วนกรณีที่หาคำตอบไม่พบในรูปที่ 10.13 การลู่เข้าเกิดขึ้นอย่างช้า ๆ เป็นเพราะความหลากหลายในวิธีการระบุและการประกอบชิ้นส่วนเข้าด้วยกัน การลู่เข้าเห็นได้ชัดในรุ่นประชากรที่ 30 และชัดเจนขึ้นเรื่อย ๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

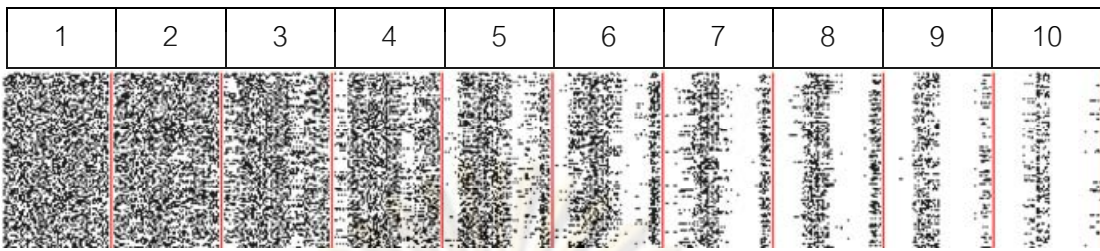


รูปที่ 10.10 รูปแบบจิ้นไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)

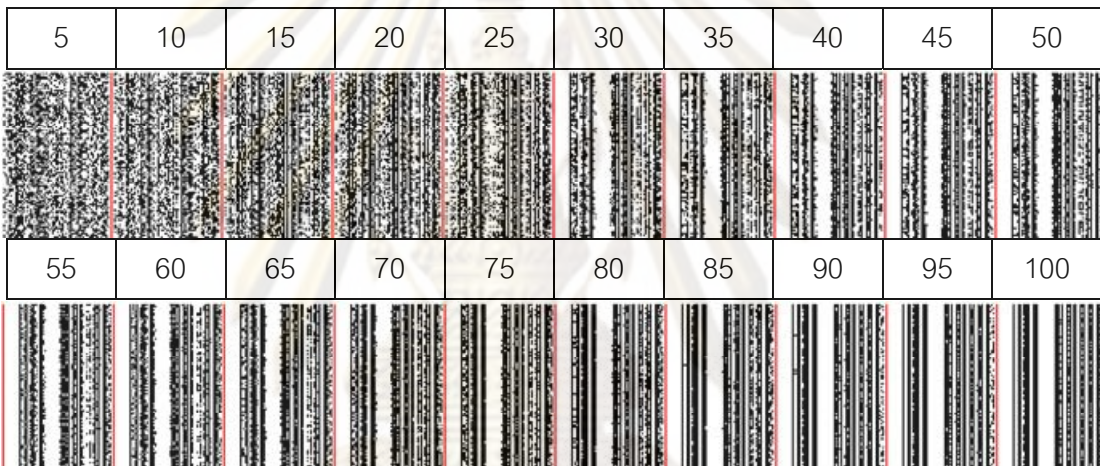


รูปที่ 10.11 รูปแบบจิ้นไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ

คือคู่จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 10.12 รูปแบบจิงโนไทป์ของประชากรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0



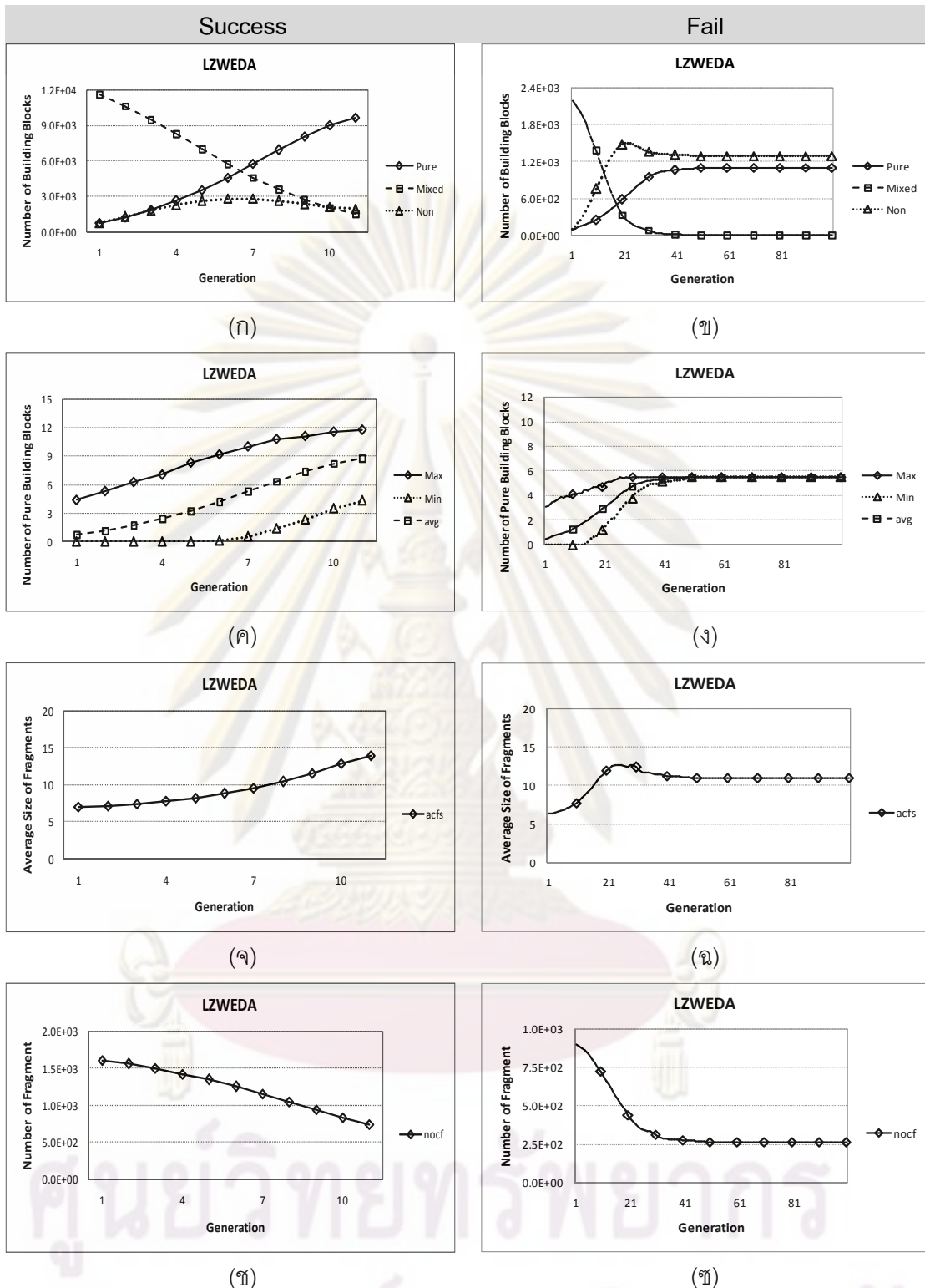
รูปที่ 10.13 รูปแบบจิงโนไทป์ของประชากรของ LZWEDA ในปัญหา Royal Road 64 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 100 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

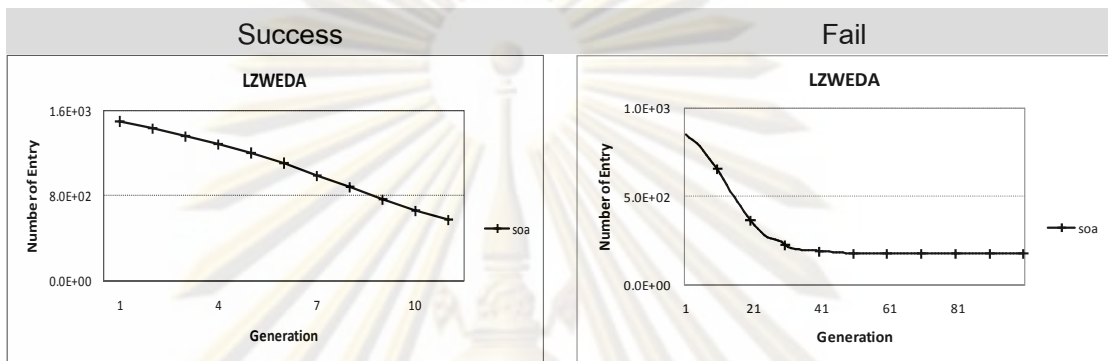
สำหรับปัญหา Trap-5 ผลการทดลองในรูปที่ 10.14 (ก) มีการแข่งกันระหว่าง Pure BBs Mixed BBs และ Non BBs เนื่องจากเป็นปัญหาลวง ในกรณีที่ทำคำตอบไม่พบจากรูปที่ 10.14 (ข) มีปริมาณของ Non BBs มากกว่า Pure BBs และ Mixed BBs ซึ่งต่างจากขั้นตอนวิธีอื่นที่นำเสนอ สันนิษฐานว่าเกิดเนื่องจากประชากรมีความหลากหลายน้อยทำให้ไม่สามารถแยกระหว่างหน่วยการสร้างที่เป็นคำตอบกับหน่วยการสร้างลวง จึงทำให้ถูกหลอกได้ง่าย และวิธีการบีบอัดแบบ LZW อาศัยรูปแบบที่เคยเกิดขึ้นแล้วเป็นหลัก ดังนั้นถ้ามีรูปแบบโครงสร้างลวงอยู่มากเมื่อบีบอัดแล้วจะได้ชิ้นส่วนที่มีโครงสร้างลวงเป็นจำนวนมาก ส่งผลให้เกิดการลู่เข้าสู่รูปแบบลวงอย่างชัดเจน และจะเห็นว่า Non BBs เพิ่มขึ้นในช่วงรุ่นประชากรที่ 20 และลดลงเรื่อย ๆ จนกระทั่งรุ่นประชากรที่ 30 และมีค่าเกือบคงที่ สันนิษฐานว่าเกิดขึ้นเนื่องมีระดับความหลากหลายต่ำ (มีจำนวนประชากรน้อย) ส่งผลให้โครงสร้างในโครโมโซมส่วนใหญ่มีส่วนประกอบของโครงสร้างลวงอยู่มากกว่า เพราะโดยเฉลี่ยโครงสร้างลวงจะมีค่าความเหมาะสมสูงกว่าโครงสร้างที่เป็นคำตอบ ยกเว้นกรณีที่โครงสร้างมีหน้าตาเหมือนคำตอบทุกบิตจึงจะมีค่าความเหมาะสมสูงกว่าซึ่งมีเพียงแค่รูปแบบเดียวเท่านั้น ทำให้มีโครงสร้างลวงในการบีบอัดสูงกว่าโครงสร้างที่เป็นคำตอบ ส่งผลให้มีปริมาณ Non BBs มากกว่า Pure BBs ชิ้นส่วนที่ระบุได้จากวิธีการบีบอัดแบบ LZW จะยาวเพิ่มขึ้นทีละหนึ่งบิต และวิธีการประกอบที่ใช้หน่วยเก็บถาวรร่วมกัน ทำให้วิธี LZWEDA มีความหลากหลายในการหาคำตอบสูงยืนยันได้จากรูปที่ 10.14 (ค) จำนวนของ Max Pure BBs มีความแตกต่างจาก Min Pure BBs มากกว่าวิธี LZWGA ในบทที่ 9 และรูปที่ 10.14 (ง) เกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 50 ซึ่งถือว่าช้ามากเมื่อเทียบกับวิธี LZWGA ที่เกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 20

จากมุมมองของชิ้นส่วนในกรณีที่ทำคำตอบพบจากรูปที่ 10.14 (จ) การเพิ่มขึ้นของขนาดชิ้นส่วนโดยเฉลี่ยคล้ายกับปัญหา Royal Road เป็นเพราะวิธีการบีบอัดแบบ LZW ที่ทำให้ชิ้นส่วนค่อย ๆ ยาวเพิ่มขึ้นทีละหนึ่งบิต ในกรณีที่ทำคำตอบไม่พบจากรูปที่ 10.14 (ฉ) ขนาดของชิ้นส่วนโดยเฉลี่ยมีค่าเพิ่มขึ้นและคงที่อยู่ในช่วงรุ่นประชากรที่ 20 ถึง 30 และค่อย ๆ ลดลงจนคงที่อีกครั้งในรุ่นประชากรที่ 40 ตรงกับรูปที่ 10.14 (ข) สันนิษฐานว่าเป็นเพราะในกรณีที่ทำคำตอบไม่พบมีความหลากหลายน้อย ส่งผลให้แต่ละรอบของการทดสอบ (ทั้งหมด 30 รอบ) มีการลู่เข้าสู่รูปแบบที่ไม่เหมือนกันในช่วงรุ่นประชากรดังกล่าว ทำให้เกิดการแกว่งตัวต่างอย่างเห็นได้ชัด ส่วนปริมาณของชิ้นส่วนที่ระบุได้ในรูปที่ 10.14 (ซ-ฅ) สอดคล้องกับรูปที่ 10.14 (จ-ฉ) อาจมีข้อสงสัยว่าทำไมปริมาณชิ้นส่วนในรูปที่ 10.14 (ซ) ในช่วงรุ่นประชากรที่ 20 ถึง 30 ยังคงลดลงต่อไปต่างจากขนาดโดยเฉลี่ยของชิ้นส่วนในรูป 10.14 (ฉ) ที่มีค่าเกือบคงที่ในช่วงรุ่นประชากรเดียวกัน เป็นเพราะการลู่เข้ายังคงเกิดขึ้นอย่างต่อเนื่อง



รูปที่ 10.14 ผลการทดสอบของ LZWEDA ในปัญหา Trap-5 60 บิต ปริมาณหน่วยการสร้างกับรุ่นประชากร (ก) (ข) จำนวนหน่วยการสร้างที่เหมือนกับคำตอบที่ดีที่สุดกับรุ่นของประชากร (ข) (ง) ขนาดของชิ้นส่วนเฉลี่ยกับรุ่นประชากร (จ) (ฉ) และปริมาณชิ้นส่วนทั้งหมดกับรุ่นประชากร (ช) (ฌ)

เมื่อพิจารณาถึงปริมาณชิ้นส่วนที่เก็บอยู่ในหน่วยเก็บถาวรในกรณีที่ไม่หาคำตอบพบ จากรูปที่ 10.15 (ก) จะเห็นว่าปริมาณชิ้นส่วนค่อย ๆ ลดลงแบบเชิงเส้นตลอดการดำเนินการ แสดงให้เห็นว่าวิธี LZWEDA มีความสามารถในการรักษาระดับความหลากหลายสูง และในกรณีที่หาคำตอบไม่พบจากรูปที่ 10.15 (ข) เมื่อเปรียบเทียบกับวิธี LZWGA ถือว่ามีอัตราการลู่เข้าที่ช้ากว่าพอสมควร เป็นผลจากความหลากหลายในวิธีการประกอบชิ้นส่วนที่ใช้หน่วยเก็บถาวรร่วมกัน



(ก)

(ข)

รูปที่ 10.15 ปริมาณชิ้นส่วนทั้งหมดที่มีอยู่ในหน่วยเก็บถาวรกับรุ่นประชากรของ LZWEDA ใน ปัญหา Trap-5 60 บิต

ภาพโครงสร้างทางพันธุกรรมของชิ้นส่วนที่เก็บอยู่ในหน่วยเก็บถาวรจากรูปที่ 10.16 และรูปที่ 10.17 (วิธีการอ่านรูปอยู่ในบทที่ 4) ชิ้นส่วนที่เริ่มต้นในตำแหน่งแรกมีปริมาณมากที่สุด เพราะการบีบอัดเริ่มจากตำแหน่งแรกของโครโมโซมเสมอ ส่งผลให้ขนาดของชิ้นส่วนที่เริ่มต้นในตำแหน่งแรกโตขึ้นเรื่อย ๆ อย่างต่อเนื่องในแต่ละรุ่นประชากรอย่างชัดเจน ในกรณีที่หาคำตอบพบจากรูปที่ 10.16 การกระจายตัวของชิ้นส่วนขนาดใหญ่มีความแตกต่างจากปัญหา Royal Road ในรูป 10.10 ที่ขนาดของชิ้นส่วนใหญ่ขึ้นในตำแหน่งเริ่มต้น และตำแหน่งกลาง ๆ ของโครโมโซมเป็นต้นไป เนื่องจากปัญหา Trap-5 มีการลวงเกิดขึ้นในโครงสร้างของหน่วยการสร้าง ส่งผลให้เกิดการลู่เข้าในระดับโครงสร้างย่อยของโครโมโซม (บางช่วงของคู่จีนมีรูปแบบเหมือนกันในหลายโครโมโซม) ทำให้ชิ้นส่วนมีขนาดยาวขึ้นในช่วงตำแหน่งคู่จีนที่เกิดการลู่เข้าดังกล่าว ด้วยเหตุนี้จึงทำให้มีชิ้นส่วนขนาดใหญ่การกระจายตัวอยู่ในหลายตำแหน่ง (เกิดการลู่เข้าในหลายตำแหน่งของโครโมโซม) ในทำนองเดียวกันสำหรับกรณีที่หาคำตอบไม่พบจากรูปที่ 10.17 ขนาดของชิ้นส่วนจะโตขึ้นเป็นช่วง ๆ เนื่องจากใช้จำนวนประชากรน้อยทำให้มีความหลากหลายต่ำ จึงเกิดการลู่เข้าในแต่ละช่วงอย่างชัดเจน

รูปที่ 10.18 และรูปที่ 10.19 แสดงโครงสร้างทางพันธุกรรมของประชากรที่ได้จากหน่วยเก็บถาวรในรูปที่ 10.16 และรูปที่ 10.17 ตามลำดับ (วิธีการอ่านรูปอยู่ในบทที่ 4) จากรูปเห็น

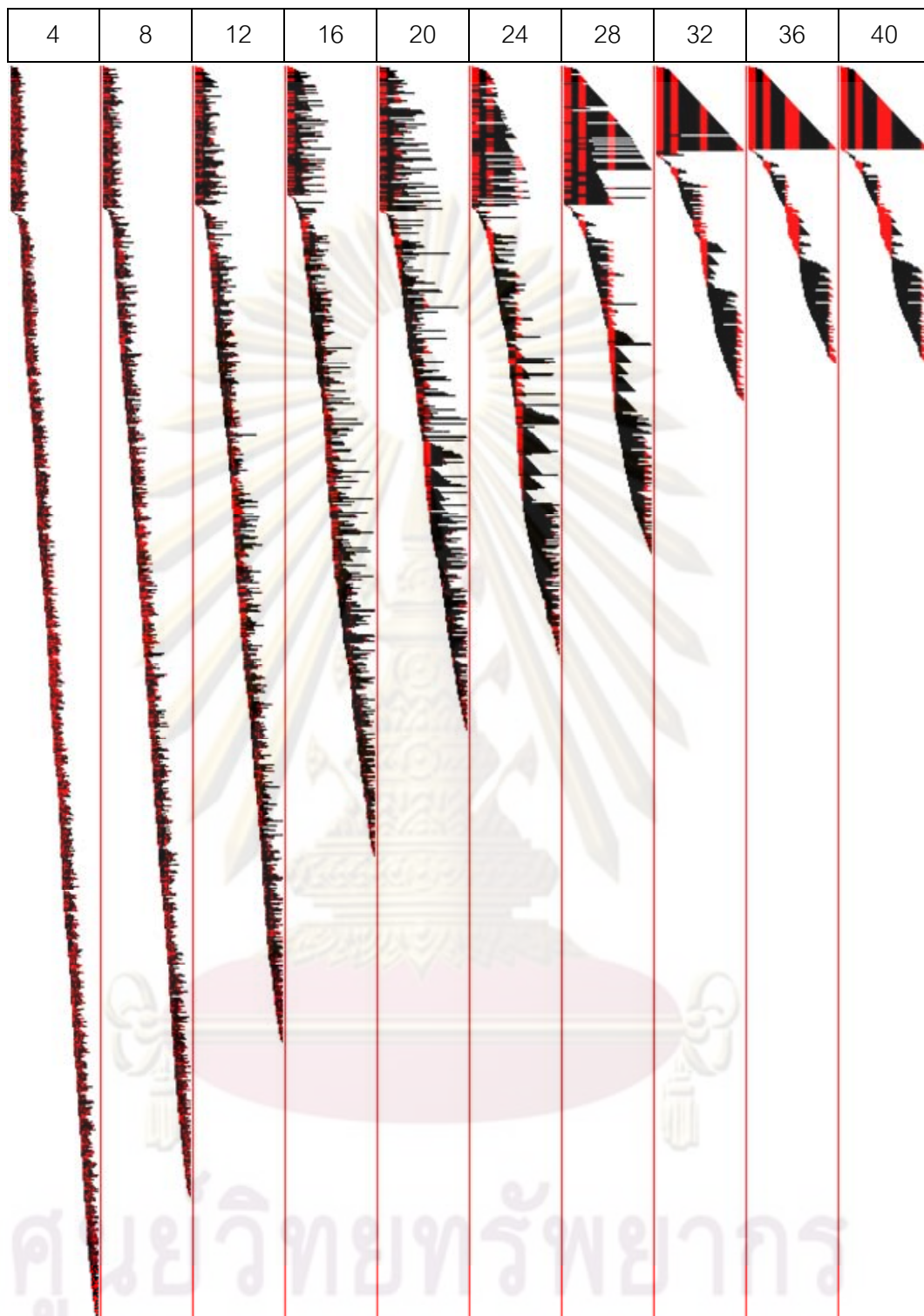
ได้ชัดเจนว่าเกิดการลู่เข้าในระดับโครงสร้างย่อยของโครโมโซมเป็นริ้วสีดำในแนวนอนซึ่งก็คือโครงสร้างลวง ในกรณีที่หาคำตอบพบจากรูปที่ 10.18 วิธี LZWEDA สามารถเอาชนะการลู่เข้าสู่โครงสร้างลวง (แถบสีดำในแนวตั้ง) ที่เกิดขึ้นตั้งแต่รุ่นประชากรที่ 6 ได้ สอดคล้องกับรูป 10.14 (ก) ที่ Non BBs มีปริมาณสูงสุดในรุ่นประชากรเดียวกัน และเริ่มลดลงจนกระทั่งพบคำตอบ กรณีที่หาคำตอบไม่พบจากรูปที่ 10.19 ประชากรลู่เข้าสู่โครงสร้างลวงในหลายช่วงตำแหน่งเริ่มสังเกตเห็นได้ตั้งแต่รุ่นประชากรที่ 10 เป็นต้นไป และเห็นได้ชัดในรุ่นประชากรที่ 18 – 20 ซึ่งโครงสร้างทางพันธุกรรมส่วนใหญ่มีค่าเป็น 0 (สีดำ) การเปลี่ยนแปลงจากค่า 0 (สีดำ) ไปเป็นค่า 1 (สีขาว) เริ่มเห็นได้ตั้งแต่รุ่นประชากรที่ 22 ไปจนถึงรุ่นประชากรที่ 34 สอดคล้องกับรูป 10.14 (ข) ที่ปริมาณของ Non BBs เริ่มลดลง (พื้นที่สีดำลดลง) และปริมาณของ Pure BBs เพิ่มขึ้น (พื้นที่สีขาวเพิ่มขึ้น) อย่างต่อเนื่อง และรูปที่ 10.14 (ค) ที่ขนาดโดยเฉลี่ยของชิ้นส่วนมีค่าเกือบคงที่ในช่วงรุ่นประชากรดังกล่าว และเกิดการลู่เข้าอย่างสมบูรณ์ในรุ่นประชากรที่ 36



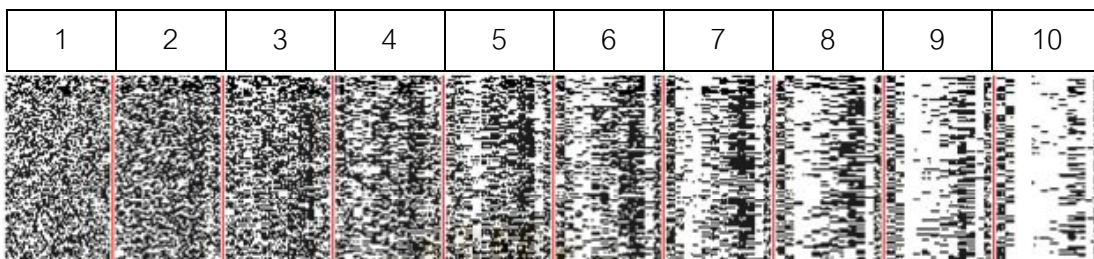
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



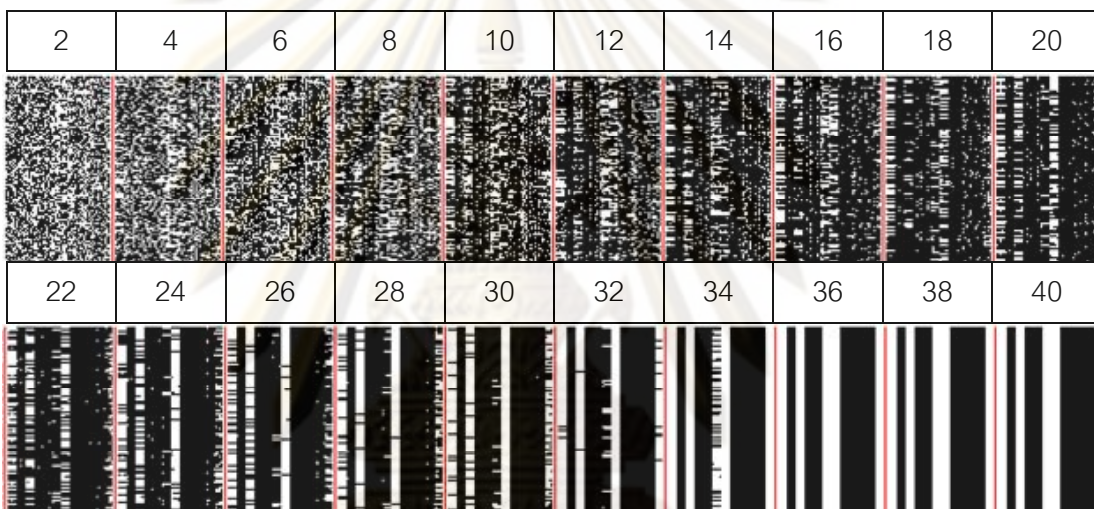
รูปที่ 10.16 รูปแบบจีโนมไทยของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จิ้นที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จิ้นที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 10.17 รูปแบบจีโนไทป์ของชิ้นส่วนที่อยู่ในหน่วยเก็บถาวรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 40 รุ่น แต่เลือกมาแสดงผล 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีเทา คือคู่จีโนไทป์ที่มีค่าเป็น 1 และพื้นที่สีดำ คือคู่จีโนไทป์ที่มีค่าเป็น 0 (วิธีการอ่านรูปอยู่ในบทที่ 4)



รูปที่ 10.18 รูปแบบจีโอเมทรีของประชากรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบพบ และใช้จำนวนรุ่นประชากร 10 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0



รูปที่ 10.19 รูปแบบจีโอเมทรีของประชากรของ LZWEDA ในปัญหา Trap-5 60 บิต กรณีหาคำตอบไม่พบ และใช้จำนวนรุ่นประชากร 40 รุ่น แต่เลือกมาแสดงผล 20 รุ่น ส่วนตัวเลขกำกับด้านบนในแต่ละช่วงหมายถึงรุ่นของประชากร เส้นขีดแบ่งรุ่นประชากรเป็นตัวบ่งบอกขอบเขตข้อมูลของประชากรในแต่ละรุ่น พื้นที่สีขาว □ คือคู่อินที่มีค่าเป็น 1 และพื้นที่สีดำ ■ คือคู่อินที่มีค่าเป็น 0

10.3 สรุปผลการทดลอง

ในบทนี้นำเสนอขั้นตอนวิธีประมาณการแจกแจงแบบ LZW ซึ่งมีการใช้หลักการบีบอัดข้อมูลแบบ LZW ในขั้นตอนการระบุชิ้นส่วน และใช้หน่วยเก็บถาวรร่วมกันในการแบ่งปันความรู้แบบรวมศูนย์ในขั้นตอนการประกอบเข้าด้วยกัน ผลการทดสอบแสดงให้เห็นว่าวิธี LZWEDA มีประสิทธิภาพดีกว่า sGA ในทุกปัญหาทดสอบและเมื่อเทียบกับวิธี CFT ที่มีการใช้หลักการสารสนเทศร่วมกันต่ำสุดในการระบุหน่วยการสร้างจะเห็นว่ามีประสิทธิภาพดีกว่าวิธี CFT ในปัญหา Royal Road และดีกว่าวิธี CFT ในปัญหา Trap-5 เป็นการแสดงให้เห็นว่าวิธี LZWEDA สามารถจัดการกับหน่วยการสร้างที่ไม่มีการลงได้ดีกว่าวิธี CFT ผลการทดสอบที่ได้สอดคล้องกับข้อสันนิษฐานในบทก่อนที่ว่า เมื่อนำหลักการบีบอัดแบบ LZW ไปประยุกต์ใช้กับขั้นตอนวิธีที่มีการแบ่งปันความรู้แบบรวมศูนย์ในกระบวนการระบุและกระบวนการประกอบหน่วยการสร้าง จะได้ผลที่ดีกว่าหรือใกล้เคียงกับวิธี CFT

ขั้นตอนวิธีที่นำเสนอทั้งหมดตั้งแต่ต้นอาจทำให้เกิดข้อกังขาว่าผลการทดสอบที่ได้มีความน่าเชื่อถือในระดับใด เนื่องจากมีการทดสอบเพียงแค่สองปัญหาเท่านั้น เพื่อให้มีความเข้าใจและเห็นภาพได้ชัดเจนขึ้น ในบทถัดไปจะแสดงผลการทดสอบกับปัญหาทั้งหมด 17 ปัญหาที่มีรูปแบบต่างกัน

บทที่ 11

ความสามารถของขั้นตอนวิธีที่นำเสนอเกี่ยวกับปัญหาทดสอบ

ในบทนี้จะพูดถึงการวัดประสิทธิภาพของขั้นตอนวิธีที่นำเสนอ โดยมุ่งเน้นไปที่ประสิทธิภาพในการระบุและประกอบหน่วยการสร้างเป็นหลัก ปัญหาทดสอบส่วนใหญ่เป็นปัญหาที่ออกแบบมาเพื่อใช้ทดสอบความสามารถในการดำเนินการกับหน่วยการสร้างโดยเฉพาะ การวัดผลใช้จำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม (Fitness Evaluation Functions: FEs) และเปรียบเทียบผลที่ได้ระหว่างขั้นตอนวิธีที่นำเสนอกับขั้นตอนวิธีอื่นในสายงานวิจัยนี้

11.1 ความเป็นมาของการทดสอบ

ปัจจัยที่มีผลต่อการแก้ปัญหาเริ่มตั้งแต่ การเข้ารหัสปัญหา (รูปแบบโครโมโซม) การคำนวณค่าความเหมาะสมจากรูปแบบการเข้ารหัสปัญหา (การประเมินค่าโครโมโซม) การเลือกสรร (การคัดเลือกโครโมโซมที่ต้องการ อาจพิจารณาที่ค่าความเหมาะสมเพียงอย่างเดียว หรือพิจารณาจากรูปแบบของโครโมโซมร่วมด้วย) การออกแบบตัวดำเนินการให้เข้ากันได้กับรูปแบบการเข้ารหัสปัญหา (การจัดการกับโครโมโซม เช่น การไขว้เปลี่ยน หรือการผ่าเหล่า เป็นต้น) การประเมินโครงสร้างรูปแบบการเข้ารหัสปัญหา (เช่น ใช้ความน่าจะเป็นแทนโครงสร้างของรูปแบบโครโมโซม เป็นต้น) การแปลงรูปแบบของการเข้ารหัสปัญหาให้อยู่ในรูปแบบอื่น (เช่น แปลงจากคู่จิ้นให้อยู่ในรูปสมการ หรือตัวแปร เป็นต้น) จนถึงเกณฑ์การตัดสินใจในการหยุดดำเนินการ เช่น จำกัดจำนวนรุ่นประชากร หรือจำกัดเวลาที่ใช้ประมวลผล เป็นต้น ปัจจัยเหล่านี้ล้วนแล้วแต่มีผลกระทบต่อประสิทธิภาพ (ความสามารถ) ในการหาคำตอบทั้งสิ้น และอาจมีความเหมาะสมกับแต่ละปัญหาแตกต่างกัน โดยปกติมักนิยมใช้ขั้นตอนวิธีเชิงพันธุกรรมในการแก้ปัญหาที่มีความรู้เกี่ยวกับปัญหาน้อย ทำให้มีความรู้ในการเข้ารหัสน้อยไปด้วยคือ ไม่รู้ว่าการเข้ารหัสรูปแบบใดจะมีประสิทธิภาพที่ดีในการแก้ปัญหา และส่งกระทบถึงปัจจัยอื่น ๆ ที่ตามมา ดังนั้นในการแก้ปัญหาส่วนใหญ่ไม่สามารถบอกได้ล่วงหน้าว่าจะเข้ารหัสปัญหาแบบใด ควรจะใช้ตัวดำเนินการแบบใด ควรจะจัดการกับปัจจัยอื่น ๆ เช่นใด จะมีความรู้เพิ่มขึ้นก็ต่อเมื่อได้มีการทดลองและมีการวัดผลการทดลองแล้ว ซึ่งไม่ต่างไปจากการแก้ปัญหาด้านอื่น ๆ ในโลกจริง

โดยทั่วไปมีความเชื่อว่าขั้นตอนวิธีที่แก้ปัญหาแบบหนึ่งได้ดี จะสามารถแก้ปัญหาอื่น ๆ ที่มีรูปแบบเดียวกันหรือใกล้เคียงกันได้ดีเช่นกัน เนื่องจากเชื่อว่าปัญหาในกลุ่มเดียวกันจะมีลักษณะหรือคุณสมบัติที่สำคัญบางประการร่วมกัน ซึ่งสามารถพิจารณาได้ในหลายระดับ เช่น ระดับที่มองว่าเป็นปัญหากลุ่ม NP (Non-deterministic Polynomial-time) หรือไม่ หรือมองถึงกลไก

ไปว่าเป็นแบบ NP-hard หรือ NP-complete เป็นต้น และหลายมุมมอง เช่น ปัญหาการเดินทางของนักขาย (TSP) กับปัญหาการจัดตาราง (Scheduling) อาจมองว่าเป็นปัญหาการเรียงสับเปลี่ยนเหมือนกัน แต่มีรายละเอียดไม่เหมือนกัน เป็นต้น ในสายงานวิจัยขั้นตอนเชิงพันธุกรรมมีการมองลึกลงไปที่ระดับการทำให้เกิดผล (Implementation) โดยมองในระดับโครงสร้างของปัญหาที่ผ่านการเข้ารหัสแล้ว โดยมีความเชื่อว่าการเข้ารหัสปัญหารูปแบบหนึ่ง ๆ เป็นการเปลี่ยนแปลงลักษณะความสัมพันธ์ของโครงสร้างที่มีต่อกัน (ความสัมพันธ์ของแต่ละคู่จิ้นที่มีต่อกัน) และไม่ว่าจะทำการเข้ารหัสแบบใดความสัมพันธ์นี้จะยังคงมีอยู่เสมอ เราเรียกความสัมพันธ์ที่มีเหล่านี้ว่าการเชื่อมโยง (Linkage) หรือเป็นที่รู้จักกันในนามของคำว่าหน่วยการสร้างนั่นเอง จากมุมมองนี้ไม่ว่าเป็นปัญหาอะไรก็ตามเมื่อทำการเข้ารหัสปัญหา (เพื่อหาคำตอบ) แล้วทุกปัญหาจะมีหน่วยการสร้างอยู่ในโครงสร้างปัญหาที่เข้ารหัสแล้วเสมอ แต่หน่วยการสร้างในปัญหาที่ต่างกันอาจมีรูปแบบต่างกัน ดังนั้นงานวิจัยในสายนี้จึงมุ่งเน้นไปที่การจัดการหน่วยการสร้างเหล่านี้ เพราะเชื่อว่าหน่วยการสร้างเหล่านี้จะนำไปสู่คำตอบที่ต้องการได้

ปัญหาทดสอบในสายงานวิจัยเชิงพันธุกรรมในภาพรวมแบ่งออกเป็น 2 กลุ่มได้แก่ กลุ่มปัญหาวัดเปรียบเทียบสมรรถนะ (Benchmark Problem) และกลุ่มปัญหาในโลกจริง (Real-world Problem) ปัญหาทั้งสองกลุ่มมีจุดประสงค์ในการวัดผลต่างกัน ปัญหาวัดเปรียบเทียบสมรรถนะนิยมใช้ในการศึกษาคุณลักษณะหรือพฤติกรรมของขั้นตอนวิธีที่ออกแบบในระดับเบื้องต้น เพื่อให้เข้าใจถึงกลไกการทำงานและผลกระทบจากตัวดำเนินการต่าง ๆ ปัญหากลุ่มนี้สามารถแบ่งเป็นกลุ่มย่อย ๆ ได้หลายวิธี เช่น แบ่งตามมุมมองความสัมพันธ์ของหน่วยการสร้างที่นำเสนอในงานวิจัยนี้ เป็นต้น โดยปกติเมื่อทำการทดสอบกับกลุ่มปัญหาวัดเปรียบเทียบสมรรถนะจนได้ผลเป็นที่น่าพอใจ (เข้าใจกระบวนการทำงาน หรือมีความเชื่อมั่นในระดับหนึ่งแล้ว) แล้วจึงนำขั้นตอนวิธีที่ออกแบบไว้ไปทดสอบกับกลุ่มปัญหาในโลกจริง (ที่มีความซับซ้อนสูง) อีกทีหนึ่ง ดังนั้นจึงไม่สามารถปฏิเสธได้ว่ากลุ่มปัญหาวัดเปรียบเทียบสมรรถนะ หรือเรียกกันในชื่อว่า ปัญหาของเล่น (Toy Problem) ไม่จำเป็น หรือไม่มีความสำคัญ หรือไม่ควรรใช้ในการทดสอบ เมื่อเปรียบเทียบกับ การแก้ปัญหาในโลกจริงหลายปัญหาจะต้องมีการทดสอบกับแบบจำลองก่อนนำไปใช้จริง เพราะมีความเสี่ยงน้อย ต้นทุนต่ำ และทำการวิเคราะห์ได้ง่าย เป็นต้น

ในบทนี้ทำการทดสอบขั้นตอนวิธีที่นำเสนอกับปัญหากลุ่มวัดเปรียบเทียบสมรรถนะ เพื่อศึกษากลไกการทำงานของขั้นตอนวิธีที่นำเสนอในมุมมองของชิ้นส่วน (การระบุและการประกอบหน่วยการสร้างแบบชิ้นส่วน) โดยใช้ปัญหาที่ออกแบบมาเพื่อทดสอบหน่วยการสร้างเป็นหลัก และหวังว่าปัญหาทดสอบเหล่านี้จะช่วยทำให้มีความเข้าใจถึงคุณลักษณะหรือพฤติกรรม

ทำงานของการระบุและการประกอบหน่วยการสร้างในมุมมองของชิ้นส่วนจากแต่ละขั้นตอนวิธีที่นำเสนอได้ดียิ่งขึ้น

11.2 สรุปขั้นตอนวิธีที่นำเสนอทั้งหมด

มุมมองเกี่ยวกับหน่วยการสร้างทั้งหมดที่นำเสนอในงานวิจัยสามารถสรุปโดยแบ่งเป็นมุมมองหลัก ๆ ได้ดังต่อไปนี้

รายละเอียดของแต่ละขั้นตอนวิธีที่นำเสนอ

1. การไขว้เปลี่ยนแบบชิ้นส่วน (Fragment Crossover)

การระบุหน่วยการสร้าง : ลำดับย่อยที่ต่างกันระหว่างสองโครโมโซม ณ ตำแหน่งเดียวกัน

การประกอบหน่วยการสร้าง : ใช้วิธีการไขว้เปลี่ยนแบบจุดเดียวเหมือนกับ sGA

การแบ่งปันความรู้ : ใช้ข้อมูลจากพ่อพันธุ์และแม่พันธุ์ร่วมกัน

แสดง : การพิจารณาหน่วยการสร้างเป็นลำดับย่อย (ก่อนติดกัน)
2. ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (Random Fragmentation)

การระบุหน่วยการสร้าง : สุ่มตัดโครโมโซมทั้งเส้นออกเป็นลำดับย่อย ๆ แล้วเก็บในหน่วยเก็บถาวร

การประกอบหน่วยการสร้าง : เลือกจากหน่วยเก็บถาวรและใช้ตำแหน่งของโครโมโซมเป็นตัวอ้างอิงในการประกอบเข้าด้วยกัน

การแบ่งปันความรู้ : ใช้ความรู้ร่วมกันจากหน่วยเก็บถาวร

แสดง : การใช้ความรู้ร่วมกันแบบรวมศูนย์
3. ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (Conditional Fragmentation)

การระบุหน่วยการสร้าง : ลำดับย่อยที่ต่างกัน และลำดับย่อยที่เหมือนกัน ระหว่างสองโครโมโซม ณ ตำแหน่งเดียวกัน

การประกอบหน่วยการสร้าง : เลือกจากหน่วยเก็บถาวรและใช้ตำแหน่งของโครโมโซมเป็นตัวอ้างอิงในการประกอบเข้าด้วยกัน

การแบ่งปันความรู้ : ใช้ข้อมูลจากพ่อพันธุ์และแม่พันธุ์ร่วมกัน และใช้ความรู้ร่วมกันจากหน่วยเก็บถาวร

แสดง : การพิจารณาหน่วยการสร้างเป็นลำดับย่อยร่วมกับการใช้ความรู้ร่วมกันแบบรวมศูนย์

4. ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA)

การระบุหน่วยการสร้าง : ลำดับย่อยที่ได้จากวิธีบีบอัดแบบ LZW

การประกอบหน่วยการสร้าง : ใช้วิธีการไขว้เปลี่ยนแบบจุดเดียวเหมือนกับ sGA

การแบ่งปันความรู้ : ใช้ข้อมูลที่ได้จากวิธีการบีบอัดแบบ LZW ร่วมกัน

แสดง : การนำหลักการเกี่ยวกับทฤษฎีสารสนเทศมาประยุกต์ใช้เป็นตัวระบุหน่วยการสร้าง

รายละเอียดปลีกย่อย : มี 3 รูปแบบ

4.1 1C-1D : 1 โครโมโซม สร้าง 1 พจนานุกรม

4.2 2C-1D : 2 โครโมโซม (พ้อพันธุกับแม่พันธุ) สร้าง 1 พจนานุกรม

4.3 1SP-1D : 1 กลุ่มประชากรย่อย สร้าง 1 พจนานุกรม

แสดง : ระดับของการใช้ความรู้ร่วมกันในการสร้างพจนานุกรมว่าให้ผลต่างกันอย่างไร

5. ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA)

การระบุหน่วยการสร้าง : ลำดับย่อยที่ได้จากวิธีบีบอัดแบบ LZW

การประกอบหน่วยการสร้าง : เลือกจากหน่วยเก็บถาวรและใช้ตำแหน่งของโครโมโซมเป็นตัวอ้างอิงในการประกอบเข้าด้วยกัน

การแบ่งปันความรู้ : ใช้ข้อมูลที่ได้จากวิธีการบีบอัดแบบ LZW และใช้ความรู้ร่วมกันจากหน่วยเก็บถาวร

แสดง : การระบุหน่วยการสร้างด้วยวิธีบีบอัดแบบ LZW ร่วมกับการใช้ความรู้ร่วมกันแบบรวมศูนย์

เมื่อพิจารณาถึงเวลาที่ใช้ในการดำเนินการสำหรับแต่ละขั้นตอนวิธีที่น่าเสนอ แสดงให้เห็นโดยใช้อัตราการเติบโตของจำนวนครั้งการเรียกใช้ตัวดำเนินการ เช่น ตัวดำเนินการไขว้เปลี่ยน หรือตัวดำเนินการบีบอัด เป็นต้น ซึ่งสะท้อนให้เห็นถึงเวลาที่ใช้ในการประมวลผลในภาพรวม ที่ไม่ขึ้นกับวิธีการเขียนโปรแกรม หรือระบบที่ใช้ในการประมวลผล สำหรับขั้นตอนวิธีเชิงพันธุกรรมสามารถพิจารณาในแง่มุมมองของผลที่ได้ต่อการเรียกใช้ตัวดำเนินการในแต่ละครั้ง โดยพิจารณาจากมุมมองของจำนวนโครโมโซมที่ผ่านการประมวลผลในแต่ละครั้งว่ามีจำนวนเท่าใด เนื่องจากการดำเนินการทั้งหมดในขั้นตอนวิธีเชิงพันธุกรรมจะกระทำกับโครโมโซมเป็นหลัก เช่น การคำนวณหาค่าความเหมาะสม การไขว้เปลี่ยน หรือ การผ่าเหล่า เป็นต้น สำหรับแต่ละขั้นตอนวิธีที่น่าเสนอสามารถแสดงอัตราการเติบโตของจำนวนครั้งการเรียกใช้ตัวดำเนินการได้ดังตารางที่ 11.1 โดยแบ่งเป็นส่วนของการระบุหน่วยการสร้าง และส่วนของการประกอบหน่วยการสร้างเข้าด้วยกัน

ตารางที่ 11.1 ค่าความซับซ้อนเชิงเวลาของแต่ละขั้นตอนวิธีที่นำเสนอ (ต่อการทำงานใน 1 รุ่นประชากร)

Big O	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
Identify	–	p	s^2	p	p	p	s
Compose	$p/2$	p	p	$p/2$	$p/2$	$p/2$	p

หมายเหตุ ตัวแปร p หมายถึง จำนวนประชากร และตัวแปร s หมายถึง จำนวนประชากรย่อย
ค่าความซับซ้อนเชิงเวลาพิจารณาจากจำนวนครั้งของการดำเนินการที่เกิดขึ้นกับ
โครโมโซมเป็นหลัก

11.3 ปัญหาทดสอบที่ใช้ในงานวิจัยนี้

ปัญหาที่สนใจในงานวิจัยนี้เป็นปัญหาแบบวัตถุประสงค์เดียว (Single Objective) มีผลเฉลยที่ดีที่สุดรูปแบบเดียวเป็นหลัก (ในกรณีที่มีผลเฉลยที่ดีที่สุดมากกว่าหนึ่งรูปแบบ จะสนใจความสามารถในการคำตอบพบเพียงรูปแบบเดียว) ไม่มีลักษณะคำตอบแบบการเรียงสับเปลี่ยน (Permutation) เหมือนปัญหาจำพวก TSP เป็นต้น ใช้วิธีเข้ารหัสผลเฉลยแบบไม่วนกลับ (Non-cyclic) และมีรายละเอียดเพิ่มเติมดังนี้

1. เป็นปัญหาค่าสูงสุด (Maximization Problem) หรือปัญหาค่าต่ำสุด (Minimization Problem) เท่านั้น ตัวอย่างปัญหาที่ไม่เข้าข่าย สมมติปัญหาหนึ่งมีค่าความเหมาะสมในช่วง 1 ถึง 10 และค่าความเหมาะสมที่เป็นคำตอบคือ 8 แทนที่จะเป็นค่า 1 หรือค่า 10 ปัญหาแบบนี้จัดว่าไม่เป็นทั้งปัญหาค่าสูงสุด หรือค่าต่ำสุด ดังนั้นจะไม่อยู่ในขอบเขตของงานวิจัยนี้
2. ไม่เป็นปัญหาแบบ Needle in a Haystack เช่น ปัญหาที่มีค่าความเหมาะสมเพียงสองค่า คือ 0 กับ 10 คือถ้าไม่เจอคำตอบก็จะมีค่าความเหมาะสมเป็น 0 และจะได้ค่า 10 ก็ต่อเมื่อเจอคำตอบที่เหมาะสมที่สุด เป็นต้น ในกรณีปัญหา RoyalRoad ซึ่งมีขนาด 64 บิต แบ่งการให้ค่าความเหมาะสมทีละ 8 บิต (คิดเป็น 1 ชุด) และเรียงต่อกัน 8 ชุด (รวมกันได้ 64 บิต) แต่ละชุด 8 บิตนั้นจะมีค่าความเหมาะสมแค่สองค่าคือ 0 กับ 8 จะได้ค่า 8 ก็ต่อเมื่อทั้ง 8 บิตมีค่าเป็น 1 ทุกบิต นอกนั้นจะมีค่าความเหมาะสมเป็น 0 หมดทุกรูปแบบ จากจุดนี้จะเห็นว่าในแต่ละชุดจะมีลักษณะเป็น Needle in a Haystack แต่เมื่อพิจารณาครบทั้ง 8 ชุดจะเห็นว่ามีความเหมาะสมเป็นขั้นบันไดดังนี้ 8, 16, 24, 32, 40, 48, 56 และ 64 (เพิ่มทีละ 8) ในกรณีปัญหาเช่นนี้ถือว่าอยู่ในขอบเขตความสนใจของงานวิจัยนี้

3. รูปแบบของโครโมโซมที่ดี ๆ (มีค่าความเหมาะสมไม่ต่ำกว่าค่าความเหมาะสมเฉลี่ยของประชากร) จะต้องมียาลดน้อย หรือโครงสร้างย่อยที่ตรงกับรูปแบบของโครโมโซมที่เป็นคำตอบ ตัวอย่างเช่น ปัญหา OneMax ขนาด 10 บิตกำหนดให้โครโมโซมตัวหนึ่งมีรูปแบบเป็น 01000 00100 และค่าความเหมาะสมเฉลี่ยของประชากรเป็น 2 จะเห็นได้ว่าโครโมโซมตัวนี้มีค่าความเหมาะสมเป็น 2 ซึ่งถือว่าต่ำเมื่อเทียบกับคำตอบที่มีค่าความเหมาะสมเป็น 10 แต่ว่ามีโครงสร้างย่อยที่ตรงกับคำตอบคือ *1*** **1** เป็นต้น ซึ่งถ้ามีคุณสมบัติตามนี้จะทำให้มีโอกาสหาคำตอบที่เหมาะสมที่สุดเจอได้ ในกรณีของปัญหาลวง เช่น ปัญหา Trap-5 ขนาด 10 บิต ซึ่งมีหน้าตาคำตอบที่ดีที่สุดเป็น 11111 11111 ถ้าโครโมโซมมีรูปแบบเป็น 00000 00000 แล้วจะได้ค่าความเหมาะสมเป็น 8 ซึ่งถือว่าสูงเมื่อเทียบกับคำตอบที่มีค่าความเหมาะสมเป็น 10 แต่ว่าโครโมโซมนี้ไม่มีโครงสร้างย่อยที่ตรงกับคำตอบที่ดีที่สุดเลยซึ่งสำหรับปัญหานี้จะเกิดขึ้นได้แค่รูปแบบเดียวเท่านั้น ส่วนรูปแบบอื่นจะมีโครงสร้างย่อยตรงกับคำตอบทุกรูปแบบ ในกรณีปัญหาแบบนี้ถือว่ายังอยู่ในขอบเขตความสนใจของงานวิจัยนี้

ถ้าปัญหามีคุณลักษณะตามที่ได้กำหนดไว้ เรามีความเชื่อว่าการระบุหน่วยการสร้างจะเกิดประโยชน์และสามารถทำให้สัมฤทธิ์ผลได้จริง

ปัญหาที่ใช้ในงานวิจัยนี้แบ่งออกเป็น 2 กลุ่มหลักได้แก่

1. ปัญหากลุ่มที่ไม่มีการลวง ประกอบด้วย
 - 1.1 ปัญหา OneMax
 - 1.2 ปัญหา RandomMax
 - 1.3 ปัญหาทดสอบจากการประชุมเชิงวิชาการ CEC-2008 ปัญหา F1 ถึง F4
 - 1.4 ปัญหา RoyalRoad
 - 1.5 ปัญหา Random RoyalRoad
 - 1.6 ปัญหา HIFF (Hierarchical If and Only If)
2. ปัญหากลุ่มลวง
 - 2.1 ปัญหา MK-Trap
 - 2.2 ปัญหา Deceptive-3
 - 2.3 ปัญหา Exponential Deceptive-3
 - 2.4 ปัญหา 6-bipolar
 - 2.5 ปัญหา HTrap (Hierarchical Trap)
 - 2.6 ปัญหา Vary Size Trap

2.7 ปัญหา Mixed Trap

2.8 ปัญหา Overlap MK-Trap

รายละเอียดเกี่ยวกับคุณลักษณะของแต่ละปัญหาแสดงไว้ในตารางที่ 11.2 และขนาดของปัญหาที่ใช้ในการทดสอบแสดงดังตารางที่ 11.3 ส่วนขั้นตอนวิธีที่ใช้เปรียบเทียบจะมีผลการทดสอบไม่ครบทุกปัญหาเนื่องจากอ้างอิงผลจากงานวิจัยที่มีการตีพิมพ์แล้ว

ในมุมมองของการทดสอบหน่วยการสร้างจากบทที่ 3 ได้พูดถึงภาพรวมแบบจำลองหน่วยการสร้างว่ามีทั้งหมด 8 ประเภทในบทนี้จะมีการจัดจำแนกปัญหาตามประเภทหน่วยการสร้างแสดงดังตารางที่ 11.2 ซึ่งหลักการที่นำเสนอในงานวิจัยนี้ได้มีการทดสอบกับหน่วยการสร้าง 7 ประเภท รูปแบบหน่วยการสร้างที่ไม่ได้ทำการทดสอบคือรูปแบบที่มีสัญญาณรบกวนร่วมด้วยเนื่องจากการออกแบบปัญหาและการวัดผลทำได้ยาก และจากการทดลองเบื้องต้นพบว่าวิธีที่นำเสนอรวมถึงขั้นตอนวิธีเชิงพันธุกรรมแบบปัญญาตินั้นมีประสิทธิภาพต่ำมากในการแก้ปัญหาหน่วยการสร้างที่มีการเชื่อมโยงแบบหลวมดังนั้นก็จึงถูกตัดออกจากส่วนการทดลองและผลการทดลอง เพราะฉะนั้นในบทนี้จะมปัญหาทดสอบหน่วยการสร้างทั้งหมด 6 รูปแบบ ได้แก่

9. หน่วยการสร้างที่มีการเชื่อมโยงแบบแนบแน่น
10. หน่วยการสร้างที่มีขนาดแปรผัน
11. หน่วยการสร้างที่มีการเหลื่อมกัน
12. หน่วยการสร้างที่มีโครงสร้างความสัมพันธ์แบบลำดับขั้น
13. หน่วยการสร้างที่แต่ละหน่วยมีค่าความเหมาะสมไม่เท่ากัน
14. หน่วยการสร้างแบบผสมที่มีรูปแบบดังข้อที่ 1. ถึง 7. รวมกันมากกว่า 1 รูปแบบ

วัตถุประสงค์ในการทดสอบปัญหาแบ่งออกได้เป็นสองส่วน ส่วนแรกต้องการให้เห็นถึงการมีอยู่ของหน่วยการสร้างในแต่ละมุมมอง (การระบุหน่วยการสร้าง) ที่นำเสนอที่มีอยู่ในแต่ละปัญหาทดสอบ ซึ่งแต่ละมุมมองอาจมีความเหมาะสมกับแต่ละปัญหาไม่เท่ากัน แต่ทุกมุมมองอยู่บนพื้นฐานนิยามหน่วยการสร้างแบบชิ้นส่วนเหมือนกันหมด ส่วนที่สองแสดงให้เห็นถึงประโยชน์ที่ได้รับหรือประสิทธิภาพในการแก้ปัญหาเมื่อสามารถระบุและประกอบหน่วยการสร้างได้อย่างเหมาะสม

ตารางที่ 11.2 ปัญหาทดสอบ

Test Problems		Non-Deceptive	Deceptive	All 0/1 optimal	Non all 0/1 optimal	Straightforward	Hierarchical	Linear Fitness	Non-linear Fitness	Unimodal	Multimodal	BBs Test	Tightly BBs	Vary size BBs	Overlapping BBs	Mixed Type BBs	Hierarchical BBs	Non-linear Fitness BBs
1	OneMax	•		•		•		•		•								
2	RandomMax	•			•	•		•		•								
3	F1 (CEC-2008)	•			•	•			•	•								
4	F2 (CEC-2008)	•			•	•			•	•								
5	F3 (CEC-2008)	•			•	•			•		•							
6	F4 (CEC-2008)	•			•	•			•		•							
7	RoyalRoad-R1	•		•		•		•		•		•	•					
8	Random RoyalRoad -R1	•		•		•		•		•		•	•					
9	MK-Trap		•	•		•		•		•		•	•					
10	Dec3 (Deceptive-3)		•	•		•		•		•		•	•					
11	6-bipolar		•	•		•			•		•	•	•					
12	Vary Size Trap		•	•		•		•		•		•	•	•				
13	Mixed Trap		•	•		•		•		•		•	•		•	•		
14	Exponential-Dec3		•	•		•			•	•		•	•					•
15	HIFF	•		•			•		•		•	•	•				•	
16	Htrap v1		•	•			•		•	•		•	•				•	
17	Overlap-MK-Trap		•	•		•		•		•		•	•		•			

ตารางที่ 11.3 ขนาดของแต่ละปัญหาทดสอบ

Test Problems		Sub-Problem Size				Details	Optimal
		ชุด 1	ชุด 2	ชุด 3	ชุด 4		
1	OneMax	100	150	200	250		All 1
2	RandomMax	100	150	200	250		A Pattern
3	F1 (CEC-2008)	40	80	—	—	8-bit * 5,10-dim.	A Pattern
4	F2 (CEC-2008)	40	—	—	—	8-bit * 5-dim.	A Pattern
5	F3 (CEC-2008)	24	—	—	—	12-bit * 2-dim.	A Pattern
6	F4 (CEC-2008)	50	—	—	—	10-bit * 5-dim.	A Pattern
7	RoyalRoad-R1	64	128	256	—		All 1
8	Random RoyalRoad -R1	64	128	256	—		A Pattern
9	MK-Trap (K=5)	100	150	200	250	Trap-5	All 1
10	Dec3 (Deceptive-3)	60	120	180	240		All 1
11	6-bipolar	60	120	180	240		All (1 or 0)
12	Vary size Trap	60	60	—	—	Trap 3,4,5	All 1
13	Mixed Trap	128	—	—	—	32-bit * 4 set	All 1
14	Exponential-Dec3	60	90	120	—		All 1
15	HIFF	32	64	128	256		All (1 or 0)
16	Htrap-1	27	81	243	—		All 1
17	Overlap-MK-Trap (K=5)	121	92	63	—	Overlap 1,2,3 bit	All 1

11.4 วิธีการทดลอง

ขั้นตอนวิธีที่นำเสนอทั้งหมดรวมถึงขั้นตอนวิธีเชิงพันธุกรรมจะถูกทดสอบโดยใช้ปัญหาทดสอบตามตารางที่ 11.2 และ 11.3 โดยมีเงื่อนไขดังต่อไปนี้

1. ทดสอบขั้นตอนวิธีที่นำเสนอกับแต่ละปัญหาซึ่งมีทั้งหมด 17 ปัญหา จำนวนปัญหาละ 30 รอบ (สำหรับแต่ละปัญหาย่อย) แต่ละรอบจะต้องหาคำตอบที่ดีที่สุดเจอ (ต้องเจอติดต่อกันทุกรอบ) และนำผลที่ได้ทั้งหมดมาเฉลี่ยกัน การทดลองยังใช้จำนวนรอบการทดสอบมากขึ้นผลที่ได้จะยิ่งมีความน่าเชื่อถือเพิ่มขึ้นด้วย (เนื่องจากขั้นตอนวิธีในสายงานวิจัยนี้อยู่บนพื้นฐานของความน่าจะเป็น) แต่ก็ต้องแลกมาด้วยเวลาที่ใช้ในการดำเนินการ ซึ่งงานวิจัยสายนี้ส่วนใหญ่ใช้จำนวนรอบการทดสอบอย่างน้อย 30 รอบ
2. จะไม่ใช้ตัวดำเนินการผ่าเหล่ากับทุกปัญหาและทุกขั้นตอนวิธีที่นำเสนอ รวมถึงขั้นตอนวิธีเชิงพันธุกรรมที่ใช้เป็นตัวเปรียบเทียบกับ ยกเว้นปัญหา OneMax และ RandomMax ซึ่งจะใช้เป็นกรณีเปรียบเทียบให้เห็นถึงผลของการใช้ตัวดำเนินการผ่าเหล่าเป็นกรณีพิเศษ เนื่องจากงานวิจัยนี้มุ่งเน้นไปที่การศึกษากระบวนการได้มาซึ่งหน่วยการสร้างและการประกอบเข้าด้วยกัน ดังนั้นเราจึงไม่ต้องการหน่วยการสร้างที่เกิดจากความบังเอิญในระหว่างการดำเนินการ (ซึ่งควบคุมและวิเคราะห์ผลกระทบได้ยาก) อันเกิดจากตัวดำเนินการผ่าเหล่า
3. ในการดำเนินการกับปัญหาแต่ละรอบ จะกำหนดให้สามารถเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมได้ไม่เกิน 1,000,000 (หนึ่งล้าน) ครั้ง ถ้ายังไม่สามารถหาคำตอบที่ดีที่สุดได้ภายในจำนวนครั้งที่กำหนด จะถือว่าขั้นตอนวิธีที่ทดสอบไม่สามารถแก้ปัญหาได้ เหตุที่ต้องกำหนดเช่นนี้เนื่องจากมีเวลาในการทดลองจำกัด ดังนั้นจึงไม่สามารถรอไปเรื่อย ๆ โดยที่ไม่มีกำหนด และการกำหนดขอบเขตการดำเนินการด้วยวิธีนี้เป็นที่ยอมรับและนิยมใช้ในงานวิจัยสายนี้ เพราะสามารถควบคุมได้ง่าย มีความยืดหยุ่นสูง (ไม่ขึ้นกับจำนวนรุ่นของประชากร ขนาดประชากร และค่าพารามิเตอร์อื่น ๆ) และเป็นส่วนของการดำเนินการที่ถือกันว่าแพงที่สุด หรือสำคัญที่สุดที่ทุกขั้นตอนวิธีต้องใช้เหมือนกันหมด
4. ในส่วนค่าพารามิเตอร์หลักที่ใช้เหมือนกันทุกวิธีนั้นก็คือขนาดของประชากร ซึ่งกำหนดค่าสูงสุดไว้ที่ 100,000 (หนึ่งแสน) ตัว เนื่องจากต้องการจำกัดการใช้

ทรัพยากรของระบบในการประมวลผล และต้องการศึกษาพฤติกรรม การวิวัฒนาการในแต่ละรุ่นของประชากร ดังนั้นถ้าใช้จำนวนประชากรมากเกินไปอาจทำให้หาคำตอบได้ภายในไม่กี่รุ่นประชากร ซึ่งจะทำให้กระบวนการวิวัฒนาการ สิ้นลงเกินกว่าที่จะสังเกตพฤติกรรมได้ ดังนั้นจึงใช้จำนวนประชากรน้อยที่สุด (โดยการประมาณ) เพื่อให้สามารถหาคำตอบได้

5. การกำหนดค่าพารามิเตอร์ของแต่ละขั้นตอนวิธีเป็นดังนี้

- วิธี sGA-FC และ LZWGA-1,2,3 (กลุ่มวิธีที่ใช้วิธีการไขว้เปลี่ยนเป็นหลัก) ใช้การเลือกสรรแบบการแข่งขันกีฬาโดยมีระดับการแข่งขันเป็น 4 และใช้การไขว้เปลี่ยนแบบจุดเดียว
- วิธี RFT, CFT และ LZWEDA (กลุ่มวิธีที่ใช้หน่วยเก็บถาวร) ใช้การเลือกสรรแบบตัวที่ดีที่สุดเป็นจำนวน n ตัว (n -Best Selection) วิธีนี้จะทำการเรียงลำดับค่าความเหมาะสมของประชากร แล้วเลือกตัวที่ได้ค่าความเหมาะสมดีที่สุดมาจำนวนหนึ่ง (n ตัว) โดยไม่สนใจว่าจะมีหน้าตาซ้ำกันหรือไม่ ซึ่งในงานวิจัยนี้จะเรียกกลุ่มโครโมโซมที่ถูกเลือกเหล่านี้ว่า “ประชากรย่อย” (sub-population: sub.)
- เฉพาะขั้นตอนวิธีที่มีการใช้ตัวดำเนินการผ่าเหล่า (กรณีพิเศษเฉพาะปัญหา OneMax และ RandomMax เท่านั้น) จะถูกกำหนดค่าความน่าจะเป็นในการผ่าเหล่าไว้ที่ 0.02 เท่ากันทุกวิธีและทุกปัญหา

หมายเหตุ วิธี LZWGA-3 จะทดสอบเฉพาะปัญหา OneMax และ RandomMax เท่านั้น เนื่องจากเป็นวิธีที่ใช้เวลาและทรัพยากรในการประมวลผลสูงมาก

คุณสมบัติของเครื่องที่ใช้ในงานวิจัยเป็นดังนี้

- CPU: 2.4 GHz. (Intel Core 2 Quad (Q6600))
- Ram: 2 GB. (DDR-2 FSB-800 MHz.)
- Harddisk: 320 GB. (SATA-II 7200 rpm.)

ศูนย์วิจัยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

11.5 การวัดผลการทดลอง

การวัดผลจะแบ่งออกเป็น 2 ประเด็นหลัก ๆ ได้แก่

ประเด็นแรกเป็นความสามารถในการแก้ปัญหาแสดงให้เห็นโดยเปรียบเทียบกับผลของขั้นตอนวิธีเชิงพันธุกรรม (แบบไม่ใช้ตัวดำเนินการผ่าเหล่า) ขั้นตอนวิธีการแจกแจงตามขอบแบบตัวแปรเดียว (Univariate Marginal Distribution Algorithm: UMDA) และขั้นตอนวิธีอื่นที่ไม่ได้ทำการทดลองเอง ได้แก่ วิธีโคสแควร์เมทริกซ์ (Chai-square Matrix: CSM) ขั้นตอนวิธีการหาค่าเหมาะสมที่สุดแบบเบย์ (Bayesian Optimization Algorithm: BOA) และขั้นตอนวิธีการแข่งขันอย่างยุติธรรมแบบลำดับขั้นอย่างรวดเร็ว (Quick Hierarchical Fair Competition: QHFC) แต่จะไม่ได้มีผลการทดลองครบทุกปัญหาเนื่องจากใช้ผลการทดลองจากงานวิจัยอื่น ๆ ที่มีการตีพิมพ์ไปแล้วเท่านั้น เหตุที่ไม่เขียนโปรแกรมเอง หรือใช้โปรแกรมที่อาจมีผู้วิจัยอื่นทำไว้ให้แล้ว เพราะไม่ต้องการให้เกิดความผิดพลาดหรือความไม่เหมาะสมจากการปรับแต่งค่าพารามิเตอร์เองโดยปราศจากความชำนาญ เนื่องจากแต่ละขั้นตอนวิธีมีความซับซ้อนค่อนข้างสูงและไม่ตรงไปตรงมาเหมือนกับ sGA และ UMDA ที่มีความซับซ้อนน้อยและสามารถเข้าใจได้ง่าย

ประเด็นที่สองเป็นการวัดผลเพื่อใช้ในการอธิบายกลไกหลักการทำงานของแต่ละขั้นตอนวิธีว่ามีลักษณะอย่างไร ทั้งในแง่ของค่าตัวแปรต่าง ๆ รวมถึงหน่วยการสร้างและชิ้นส่วนด้วย การทดสอบกับปัญหาที่หลากหลาย (มีลักษณะเด่นต่างกัน) จะช่วยทำให้ได้มุมมองที่หลากหลายขึ้น โดยในการวิจัยนี้ได้เน้นเรื่องมุมมองการมีอยู่ของหน่วยการสร้างเป็นหลัก จากตารางจำแนกลักษณะของปัญหาทดสอบก่อนหน้านี้ (ตารางที่ 11.2) จะเห็นว่าแม้มีปัญหาส่วนหนึ่งที่ไม่ได้ถูกออกแบบมาเพื่อทดสอบเกี่ยวกับหน่วยการสร้าง (ปัญหา F1–F4) แต่ก็ไม่สามารถปฏิเสธได้ว่าหน่วยการสร้างตามนิยามจะไม่มีอยู่จริง เพียงแต่ไม่สามารถบ่งบอกรูปแบบหน่วยการสร้างที่แน่นอนได้

ในส่วนผลการทดลองสำหรับแต่ละมุมมองในเรื่องการระบุหน่วยการสร้างและการประกอบหน่วยการสร้างจะแสดงรายละเอียดเฉพาะประสิทธิภาพในการแก้ปัญหา โดยวัดจากจำนวนครั้งของการเรียกใช้ฟังก์ชันการคำนวณค่าความเหมาะสม (Fitness Evaluations: FEs) สำหรับแต่ละปัญหาทดสอบ และค่าตัวแปรที่สำคัญ ได้แก่ ขนาดประชากรที่ใช้ในการแก้ปัญหา และขนาดของประชากรย่อย (Sub-population) ซึ่งเป็นตัวแปรที่ใช้ในขั้นตอนวิธีที่มีการใช้หน่วยเก็บถาวรเท่านั้น

โดยทั่วไปการวัดประสิทธิภาพในงานวิจัยสายการคำนวณเชิงวิวัฒนาการ มีมาตรวัดหลัก ได้แก่ จำนวนครั้งของการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม เพราะถือว่าการคำนวณที่แพงที่สุดที่ทุกขั้นตอนวิธีจะต้องดำเนินการเหมือนกันหมด (มีฟังก์ชันนี้ทุกขั้นตอนวิธี) ทำให้สามารถวัดเทียบประสิทธิภาพในการหาคำตอบระหว่างขั้นตอนวิธีที่ต่างกันได้ แต่รายละเอียดในการคำนวณนั้นอาจจะต่างกันเนื่องจากแต่ละวิธีอาจจะมีรูปแบบการเข้ารหัสปัญหาไม่เหมือนกัน ส่วนมาตรวัดย่อยอื่น ๆ เช่น เวลาในการประมวลผลจนได้คำตอบที่ต้องการเป็นมาตรวัดที่นิยมรองลงมา เพื่อบอกว่าขั้นตอนวิธีที่ใช้ทดสอบนั้นมีความซับซ้อนในการคำนวณมากน้อยเพียงใด เนื่องจากมีหลายกรณีที่ไม่สามารถวัดเทียบจากค่าความซับซ้อนเชิงเวลา (Time Complexity) ได้โดยตรง ตัวอย่างเช่น วิธีที่ 1 มีความซับซ้อนกว่าวิธีที่ 2 แต่สามารถหาคำตอบได้เร็วกว่า สามารถสรุปได้ว่าวิธีที่ 1 มีประสิทธิภาพในการหาคำตอบมากกว่าวิธีที่ 2 ที่เป็นเช่นนี้อาจเป็นเพราะปัจจัยดังต่อไปนี้ วิธีที่ 1 ถูกเขียนอยู่ในรูปแบบโปรแกรมหรือภาษาที่ประมวลผลได้เร็วกว่า (ส่วนนี้ขึ้นกับความรู้ความสามารถหรือความถนัดของผู้เขียนโปรแกรมทดสอบ) วิธีที่ 1 มีความเข้ากันได้กับรูปแบบปัญหาทดสอบมากกว่าวิธีที่ 2 อาจเป็นเนื่องจากผู้ออกแบบมีความตั้งใจที่จะปรับแต่งวิธีให้เข้ากับปัญหาที่ทดสอบ แต่โดยทั่วไปแล้วเรื่องเวลาก็ยังมีความสำคัญเนื่องจากเวลาที่ใช้ในการทดลองมีจำกัด ถ้าทำการทดลองวิธีที่จะเปรียบเทียบทั้งหมดบนเครื่องคอมพิวเตอร์เครื่องเดียวกัน ก็อาจจะทำให้เห็นภาพได้อย่างคร่าวๆ ว่าขั้นตอนวิธีที่ออกแบบนั้นมีความเหมาะสมหรือไม่อย่างไรกับเงื่อนไขเรื่องเวลา เป็นต้น มาตรวัดเรื่องการใช้ทรัพยากรของระบบ ซึ่งไม่ค่อยจะเห็นบ่อยนักเนื่องจากทุกวันนี้เทคโนโลยีทางด้านคอมพิวเตอร์มีการพัฒนาอย่างรวดเร็วทำให้ระบบมีทรัพยากรเพิ่มขึ้นอย่างต่อเนื่องในอัตราที่สูง เช่น ความเร็วในการประมวลผล จำนวนหน่วยประมวลผล ขนาดหน่วยความจำและพื้นที่เก็บข้อมูลในจานบันทึกแบบแข็ง (Hard Disk) เป็นต้น และราคาของทรัพยากรเหล่านี้ในปัจจุบันก็มีแนวโน้มถูกลงเรื่อย ๆ ทำให้ผู้ใช้สามารถหามาใช้ได้ง่ายขึ้น แต่อย่างไรก็ดียังมีปัญหาขนาดใหญ่มากอีกจำนวนมากที่ใช้ทรัพยากรมหาศาลในการประมวลผลซึ่งถ้าสามารถลดขนาดของทรัพยากรที่จำเป็นต้องใช้ลงได้ก็จะมีผลทำให้ต้นทุนลดลงเป็นอย่างมาก ตัวอย่างเช่นการคำนวณเกี่ยวกับโครงสร้างทางพันธุศาสตร์ หรือโครงสร้างทางเคมี เป็นต้น ปัญหาทดสอบส่วนใหญ่ต้องการทรัพยากรในการดำเนินการในระดับหนึ่งถึงจะสามารถดำเนินการได้ แต่ถ้ามีทรัพยากรเพิ่มขึ้นก็อาจทำให้ดำเนินการได้เร็วขึ้นในรูปแบบการปรับมาตรา (Scaling) ทำให้สามารถสันนิษฐานบางอย่างได้ เช่น ถ้ามีเครื่อง 1 เครื่องใช้เวลาร้อยปี แต่ถ้ามี 100 เครื่องอาจใช้เวลาเพียงแค่วันหนึ่งปี เป็นต้น ในงานวิจัยนี้เรื่องของประสิทธิภาพจะใช้เพียงมาตรวัดจำนวนครั้งของการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม เนื่องจากงานวิจัยนี้ไม่ได้เน้นเรื่องการพัฒนาประสิทธิภาพของขั้นตอนวิธี แต่เน้นเรื่องมุมมองต่าง ๆ ที่เกี่ยวข้องกับหน่วยการสร้าง

เป็นหลัก ดังนั้นจะแสดงเฉพาะจำนวนครั้งของการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมที่ใช้จนกว่าจะพบคำตอบเป็นหลัก

ในงานวิจัยนี้มีหลายมุมมองเกี่ยวกับหน่วยการสร้างดังนั้นจึงมีการนำเสนอหลายขั้นตอนวิธีเพื่อใช้ทดสอบมุมมองต่าง ๆ ว่ามีลักษณะเป็นอย่างไร ให้ผลอย่างไร จึงได้มีการแสดงค่าตัวแปรต่าง ๆ ของแต่ละขั้นตอนวิธีซึ่งเน้นที่ขนาดของประชากรที่ใช้ในการแก้ปัญหาเป็นหลัก เนื่องจากเป็นตัวแปรที่มีเหมือนกันในทุกขั้นตอนวิธี ส่วนตัวแปรอื่น ๆ ในแต่ละวิธีจะแตกต่างกันหรือใช้ค่าใกล้เคียงกันจึงไม่จัดว่าเป็นตัวแปรที่สามารถแสดงภาพรวมร่วมกันได้ ขนาดประชากรจะเป็นตัวบอกถึงขนาดของการใช้ทรัพยากรของระบบได้ด้วยเนื่องจากจำนวนประชากรเพิ่มขึ้นหมายความว่า ต้องใช้หน่วยความจำเพิ่มขึ้น เสียเวลาในการประมวลเพิ่มขึ้นรวมถึงต้องเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมมากขึ้นด้วย เป็นต้น ถ้าปัญหามีขนาดใหญ่ขึ้นหรือยากขึ้นขั้นตอนวิธีส่วนใหญ่จะใช้ขนาดประชากรที่สูงขึ้นในการแก้ปัญหา เนื่องจากต้องการความหลากหลายมากขึ้นเมื่อเทียบกับปริภูมิค้นหาที่โตขึ้น ดังนั้นอาจกล่าวได้ว่าขนาดประชากรเป็นตัวบอกได้ว่าขั้นตอนวิธีที่ออกแบบนั้นมีความต้องการความหลากหลายระดับใดเพื่อให้สามารถแก้ปัญหาได้ สำหรับวิธีที่มีการเก็บข้อมูลไว้ในหน่วยเก็บถาวรจะมีการแสดงค่าจำนวนของประชากรย่อยที่ถูกเลือก (เพื่อทำการระบุหน่วยการสร้างแล้วเก็บไว้ในหน่วยเก็บถาวร) ซึ่งจะเป็นตัวแสดงถึงระดับความต้องการความหลากหลายจากผลเฉลยที่ดี ๆ ที่ผ่านวิธีคัดเลือกมาแล้ว และสามารถใช้พิจารณาร่วมกับขนาดประชากรเพื่อแสดงถึงระดับความหลากหลายของตัวดี ๆ ในประชากรทั้งหมดที่ต้องการเพื่อใช้แก้ปัญหาได้ด้วย

นอกจากนี้ยังมีการใช้หลักการทางสถิติเพื่อทดสอบความแตกต่างระหว่างผลการทดสอบว่าแตกต่างกันอย่างมีนัยสำคัญหรือไม่ ทำให้สามารถจัดอันดับความสามารถในการแก้ปัญหาของแต่ละขั้นตอนวิธีได้น่าเชื่อถือมากขึ้น

11.6 ผลการทดลองในส่วนของประสิทธิภาพและสมรรถนะในการแก้ปัญหา

ผลการทดลองในส่วนนี้ พิจารณาจากตารางที่ 11.5–11.10 เพื่อใช้ในการเปรียบเทียบประสิทธิภาพในการแก้ปัญหาของแต่ละขั้นตอนวิธีเป็นหลัก และผลที่สังเกตได้จากลักษณะปัญหาที่ต่างกันยังช่วยทำให้เข้าใจขั้นตอนวิธีที่น่าเสนอได้ดีขึ้นว่าเข้ากันกับปัญหารูปแบบใดบ้าง ในส่วนการวิจารณ์ผลจะพิจารณาจากปัญหาทดสอบเป็นหลัก เนื่องจากแต่ละปัญหาทดสอบมีคุณสมบัติเฉพาะที่สามารถบ่งบอกถึงพฤติกรรมของขั้นตอนวิธีที่น่าเสนอได้ในทางอ้อม และการเปรียบเทียบผลจะใช้ผลจากขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย (ไม่ใช้ตัวดำเนินการผ่าเหล่า) เป็นตัวอ้างอิง จากปัญหาทดสอบทั้งหมดพอจะแบ่งออกเป็นกลุ่ม ๆ ตามคุณลักษณะเด่นได้ดังนี้

ตารางที่ 11.4 การแบ่งกลุ่มปัญหาทดสอบเพื่อใช้ในการอธิบายผลการทดลอง

No.	Problem Group	Problem	
1	Basic	OneMax,	RandomMax
2	BBs: non deceptive & additively decomposable	Royal Road R1	Random RR R1
3	BBs: deceptive & additively decomposable	Dec3	Trap-5
4	BBs: deceptive & non-linear fitness	Exponential-Dec3	
5	BBs: deceptive & multimodal	f6-bipolar	HIFF
6	BBs: deceptive & hierarchical	Htrap v1	HIFF
7	BBs: deceptive & Overlapping	Overlap-MK-Trap	
8	BBs: deceptive & Vary Size	Vary size Trap	
9	BBs: Mixed	Mixed Trap	
10	Optimization	F1- F4	

เริ่มจากปัญหา OneMax ซึ่งเป็นปัญหาพื้นฐานที่เข้าใจได้ง่ายและนิยมใช้ในการทดสอบประสิทธิภาพเบื้องต้นของขั้นตอนวิธีที่น่าเสนอ ส่วนปัญหา RandomMax (สาวิตรี ฉ่ำลัมัย และ วรเศรษฐ สุวรรณิก, 2007) เป็นปัญหาที่คล้าย OneMax ต่างกันที่หน้าตาคำตอบไม่จำเป็นต้องเป็นหนึ่งหมดทุกบิตหรือไม่ แต่ถ้าตำแหน่งใดมีหน้าตาเหมือนคำตอบจะได้ค่าความเหมาะสมตำแหน่งละหนึ่ง (ถ้าปัญหามีขนาด 100 บิตค่าความเหมาะสมสูงสุดจะมีค่าเท่ากับ 100 ด้วย) ปัญหานี้ถูกออกแบบมาเพื่อทดสอบความลำเอียงในการหาคำตอบในกรณีที่สูงสั้ยว่าขั้นตอนวิธีที่ออกแบบจะสามารถหาคำตอบได้เฉพาะปัญหาที่มีรูปแบบวนซ้ำ หรือคำตอบจะต้องเป็นหนึ่งหรือศูนย์ทั้งหมดทุกบิต และในงานวิจัยนี้ได้นำเสนอขั้นตอนวิธีที่น่าหลักการบีบอัดมาประยุกต์ใช้ร่วมด้วย ซึ่งโดยธรรมชาติของขั้นตอนวิธีการบีบอัดจะมีความลำเอียงกับรูปแบบวนซ้ำ หรือรูปแบบที่มีหน้าตาโดยรวมเหมือนกัน จึงทำให้เกิดข้อกังขาว่าขั้นตอนวิธีที่น่าหลักการบีบอัดมาใช้จะยังคงสามารถแก้ปัญหามีรูปแบบไม่ตายตัวได้หรือไม่ ในสองปัญหานี้ (OneMax และ RandomMax) จะพิจารณาแบ่งออกเป็นสองกรณีคือ กรณีที่ใช้การผ่าเหล่าร่วมด้วย และกรณีที่ไม่ใช้การผ่าเหล่า เพื่อช่วยให้เห็นภาพว่าการใช้ตัวดำเนินการผ่าเหล่ามีผลอย่างไรต่อประสิทธิภาพในการแก้ปัญหาลักษณะง่าย ในภาพรวมจะเห็นได้ว่าเมื่อมีการยอมให้เกิดการผ่าเหล่า ผลที่ได้จากแต่ละขั้นตอนวิธีสำหรับปัญหา OneMax และ RandomMax เมื่อเปรียบเทียบกันจะมีค่าใกล้เคียงกัน (ในขั้นตอนวิธีเดียวกัน) และเมื่อไม่ใช้การผ่าเหล่าจะเห็นว่าทุกขั้นตอนวิธีจะมีประสิทธิภาพแย่งลง (เนื่องจากมี

ค่า FEs เพิ่มขึ้น) และสำหรับปัญหา OneMax และ RandomMax ค่า FEs ของทั้งสองปัญหาในแต่ละขั้นตอนวิธีมีค่าใกล้เคียงกัน ขั้นตอนวิธีส่วนใหญ่เมื่อไม่มีการผ่าเหล่าจะมีประสิทธิภาพแย่งไม่มากยกเว้น LZWGA-1, LZWGA-2 และ LZWGA-3 ที่มีประสิทธิภาพแย่งอย่างมาก เหตุที่เป็นเช่นนี้สันนิษฐานว่าเกิดจากการที่ปัญหาทดสอบทั้งสองปัญหามีรูปแบบที่เป็นไปได้จำนวนมากที่มีค่าความเหมาะสมเท่ากัน ซึ่งเป็นผลเสียต่อการบีบอัดเนื่องจากมีโอกาสที่บิตที่ไม่ใช่คำตอบจะถูกบีบอัดให้อยู่รวมกับบิตที่เป็นคำตอบ ซึ่งถ้าไม่มีตัวดำเนินการผ่าเหล่ามาช่วยแก้ไขความผิดพลาดเหล่านี้ก็จะทำให้ประสิทธิภาพในการหาคำตอบแย่ง ทำให้สรุปได้ว่าตัวดำเนินการผ่าเหล่ามีผลต่อประสิทธิภาพในการหาคำตอบสำหรับปัญหา OneMax และ RandomMax รวมถึงมีผลดีต่อขั้นตอนวิธีที่ใช้วิธีการบีบอัดร่วมด้วย และขั้นตอนวิธีที่ใช้หลักการบีบอัดสามารถแก้ปัญหาที่ไม่มีรูปแบบคำตอบที่ตายตัวได้เป็นอย่างดี (แก้ปัญหา OneMax ได้มีประสิทธิภาพใกล้เคียงกับปัญหา RandomMax)

เมื่อพิจารณาปัญหาที่ใช้ทดสอบหน่วยการสร้างแบบที่ไม่มีการลวงซึ่งก็คือปัญหา Royal Road เป็นปัญหาที่ออกแบบเพื่อทดสอบความสามารถของขั้นตอนวิธีเชิงพันธุกรรมในการค้นหาหน่วยการสร้างและประกอบเข้าด้วยกัน รวมถึงการรักษาหน่วยการสร้างที่ดี ๆ ให้มีอัตราการคงอยู่และการเติบโตที่สูงกว่าการทำลาย สำหรับปัญหานี้ยังคงมีการทดสอบเรื่องความลำเอียงกับรูปแบบคำตอบที่วนซ้ำ กับขั้นตอนวิธีที่ใช้หลักการบีบอัดร่วมด้วย โดยใช้ปัญหา Random Royal Road (สวาทวีร์ ฉ่ำลัมัย และ วรเศรษฐ์ สุวรรณิก, 2007) เป็นปัญหาทดสอบ ผลที่ได้เหมือนกับสองปัญหาแรก (OneMax และ RandomMax) คือ ค่า FEs ของแต่ละขั้นตอนวิธีในปัญหา Royal Road และ Random Royal Road เมื่อเปรียบเทียบกันจะมีค่าใกล้เคียงกัน และขั้นตอนวิธีที่ใช้หลักการบีบอัดจะมีประสิทธิภาพต่ำลงเล็กน้อยในปัญหา Random Royal Road ซึ่งเป็นการแสดงให้เห็นว่าขั้นตอนวิธีที่ใช้หลักการบีบอัดร่วมด้วยแทบจะไม่มี ความลำเอียงกับปัญหาที่มีรูปแบบหน่วยการสร้างในคำตอบแบบวนซ้ำ (แก้ปัญหา Royal Road ได้มีประสิทธิภาพใกล้เคียงกับปัญหา Random Royal Road)

กรณีที่เป็นปัญหาลวงแบบที่หน่วยการสร้างสามารถแบ่งแยกได้อย่างชัดเจน (Decomposable) งานวิจัยนี้ใช้ปัญหา Dec3 และ Trap-5 เป็นตัวทดสอบความสามารถในการหาหน่วยการสร้างที่ถูกต้องจากการลวงที่เกิดขึ้น ซึ่งทั้งสองปัญหาต่างกันที่ลักษณะค่าความเหมาะสมปัญหา Dec3 จะมีอัตราความแตกต่างกันของค่าความเหมาะสมที่สูงกว่าปัญหา Trap-5 แต่มีขนาดหน่วยการสร้างที่เล็กกว่า โดยทั่วไปแล้วความยากของปัญหาขึ้นอยู่กับขนาดหน่วยการสร้างด้วย เนื่องจากหน่วยการสร้างที่มีขนาดใหญ่จะมีโอกาสถูกทำลายได้มากกว่า ดังนั้นจึงสามารถพิจารณาได้ว่าปัญหา Trap-5 มีความยากมากกว่าปัญหา Dec3 และจากผลการทดลองแสดงให้เห็น

เห็นว่าเมื่อขนาดปัญหามีค่าใกล้เคียงกันการแก้ปัญหา Trap-5 จำเป็นต้องใช้จำนวน FEs มากกว่าปัญหา Dec3 ซึ่งสอดคล้องกับการพิจารณาเบื้องต้น ผลการทดลองยังแสดงให้เห็นอีกว่าเมื่อปัญหาที่มีขนาดใหญ่ขึ้นบางขั้นตอนวิธีจะต้องใช้จำนวน FEs เพิ่มขึ้นเป็นเลขชี้กำลังด้วย และปัญหานี้ยังต้องการความสามารถในการระบุขอบเขตหรือรูปแบบของหน่วยการสร้างที่ชัดเจน สำหรับขั้นตอนวิธีที่มีประสิทธิภาพต่ำได้แก่ sGA, RFT, LZWGA-1 และ LZWGA-2 เนื่องจากไม่สามารถแก้ปัญหาย่อยบางปัญหาได้เมื่อเทียบกับขั้นตอนวิธีอื่น ได้ข้อสมมติฐานว่าเหตุที่ RFT ทำงานได้ไม่ดีเนื่องจากไม่มีความสามารถในการระบุรูปแบบหน่วยการสร้างที่ชัดเจน จึงทำให้ไม่สามารถระบุรูปแบบหน่วยการสร้างที่ถูกต้องได้ ส่วนวิธี LZWGA-1 ทำงานไม่ดีเนื่องจากไม่มีความรู้เพียงพอในการระบุรูปแบบหน่วยการสร้างที่ถูกต้อง ซึ่งต่างจาก LZWGA-2 ที่มีการแบ่งปันความรู้ระดับหนึ่ง จึงทำให้สามารถระบุหน่วยการสร้างได้ดีขึ้น และให้ผลที่ดีกว่า LZWGA-1

สำหรับกรณีปัญหาลวงที่ค่าความเหมาะสมของแต่ละหน่วยการสร้างมีค่าไม่เท่ากัน จากปัญหา Exp-Dec3 ซึ่งเป็นปัญหาที่หน่วยการสร้างมีค่าเพิ่มขึ้นเป็นฟังก์ชันเลขชี้กำลัง ใช้ในการทดสอบความอ่อนไหวของขั้นตอนวิธีต่อหน่วยการสร้างว่าจะสามารถหาหน่วยการสร้างได้ครบหรือไม่ ปัญหานี้แสดงให้เห็นถึงมุมมองที่หน่วยการสร้างแต่ละหน่วยมีความสำคัญไม่เท่ากัน แต่การที่จะไปถึงคำตอบนั้นจะต้องหาหน่วยการสร้างให้ครบถึงแม้ว่าบางหน่วยการสร้างจะมีความสำคัญน้อยมากเมื่อเทียบกับตัวอื่นก็ตาม ผลการทดลองแสดงให้เห็นว่าวิธี sGA, LZWGA-1 และ LZWGA-2 มีประสิทธิภาพต่ำเนื่องจากไม่สามารถแก้ปัญหาย่อยบางปัญหาได้ เหตุที่เป็นเช่นนี้สันนิษฐานได้ว่าเกิดจากการที่หลักการบีบอัดสามารถตรวจจับได้แต่หน่วยการสร้างที่มีความสำคัญสูงเพราะมีอยู่เป็นจำนวนมาก แต่ไม่สามารถตรวจจับหน่วยการสร้างที่มีความสำคัญต่ำซึ่งมีจำนวนน้อยกว่าได้ถูกต้อง

เมื่อปัญหาที่มีคำตอบมากกว่าหนึ่งรูปแบบ ในปัญหา f6-bipolar และ HIFF เป็นปัญหาที่มีรูปแบบคำตอบสองรูปแบบที่ตรงข้ามกันคือ เป็นหนึ่งหมดทุกบิต และเป็นศูนย์หมดทุกบิต ข้อแตกต่างที่สำคัญระหว่างปัญหาทั้งสองคือปัญหา f6-bipolar เป็นปัญหาลวง ส่วนปัญหา HIFF ไม่ใช่ปัญหาลวง แต่มีโครงสร้างของหน่วยการสร้างแบบลำดับชั้น เมื่อพิจารณาจากความสามารถในการหาคำตอบพบแล้วจะเห็นว่าในขนาดปัญหาใกล้เคียงกันขั้นตอนวิธีส่วนใหญ่สามารถหาคำตอบในปัญหา HIFF ได้มากกว่า ดังนั้นอาจกล่าวได้ว่าปัญหา f6-bipolar มีความยากมากกว่าปัญหา HIFF และถ้ามองในมุมมองเรื่องขนาดของหน่วยการสร้างจะเห็นว่าปัญหา HIFF มีหน่วยการสร้างย่อยที่สุดขนาด 2 บิต แต่ปัญหา f6-bipolar มีขนาดหน่วยการสร้างใหญ่ถึง 6 บิต ในทางกลับกันอาจมีข้อโต้แย้งว่าหน่วยการสร้างของปัญหา HIFF โตขึ้นได้เนื่องจากการพิจารณาแบบลำดับชั้น และจะมีขนาดใหญ่กว่าหน่วยการสร้างของปัญหา f6-bipolar ซึ่งเป็นการพิจารณา

ถึงความสัมพันธ์ระหว่างหน่วยการสร้างย่อยรวมกัน ทั้งนี้ไม่สามารถบอกได้อย่างชัดเจนว่าขนาดของหน่วยการสร้างมีผลต่อประสิทธิภาพในการหาคำตอบมากกว่าความสัมพันธ์ที่พิจารณาแบบลำดับชั้น (การได้มาซึ่งหน่วยการสร้างขนาดใหญ่มีความยากมากกว่าการเรียนรู้ความสัมพันธ์แบบลำดับชั้น) ในปัญหา f6-bipolar จากผลการทดลองวิธี sGA และ sGA-FC ไม่สามารถหาคำตอบได้เลยแม้แต่ปัญหาย่อยเดียว ส่วนวิธี RFT และ CFT มีประสิทธิภาพในการหาคำตอบที่ต่ำ ทั้งหมดนี้สันนิษฐานได้ว่าเป็นเพราะปัญหานี้มีรูปแบบคำตอบสองรูปแบบที่ขัดแย้งกัน ดังนั้นจึงมีหน่วยการสร้างที่ขัดแย้งกันด้วย การที่จะไปถึงคำตอบได้นั้นต้องการความสามารถพิเศษในการรักษาหน่วยการสร้างให้เป็นไปในทิศทางเดียวกัน แต่ขั้นตอนวิธีดังกล่าวไม่มีความสามารถนี้ ซึ่งต่างจากขั้นตอนวิธีที่ใช้หลักการบีบอัดที่สามารถรักษารูปแบบของหน่วยการสร้างให้เป็นไปในทิศทางเดียวกันได้ หรือกล่าวได้ว่าการบีบอัดทำให้หน่วยการสร้างรูปแบบเดียวกันอยู่รวมกันส่งผลให้อโอกาสที่หน่วยการสร้างจะถูกทำลายลดต่ำลง ซึ่งเท่ากับเป็นการรักษาทิศทางของรูปแบบหน่วยการสร้างที่จะนำไปสู่คำตอบ ส่วนปัญหา HIFF จะเห็นว่าวิธี RFT มีประสิทธิภาพต่ำอย่างชัดเจนเนื่องจากวิธี RFT ไม่มีความลำเอียงกับรูปแบบความสัมพันธ์ใดเป็นพิเศษ ดังนั้นจึงไม่สามารถเรียนรู้ความสัมพันธ์แบบลำดับชั้นได้อย่างถูกต้อง และจะเห็นได้ชัดขึ้นเมื่อปัญหามีขนาดใหญ่ขึ้น (ความสัมพันธ์มีความซับซ้อนขึ้น)

กรณีปัญหาที่มีความสัมพันธ์แบบลำดับชั้นได้แก่ปัญหา HIFF และ Htrap สำหรับปัญหา HIFF ได้มีการอธิบายก่อนหน้าไปบ้างแล้ว ความแตกต่างระหว่างสองปัญหานี้คือ ปัญหา Htrap เป็นปัญหาที่มีการลวงในหน่วยการสร้างย่อยร่วมด้วย และมีขนาดหน่วยการสร้างย่อยที่สุด 3 บิต ซึ่งใหญ่กว่าปัญหา HIFF ที่มีขนาดหน่วยการสร้างย่อยที่สุด 2 บิต ดังนั้นเมื่อพิจารณาในมุมมองของขนาดหน่วยการสร้าง รวมถึงความที่เป็นปัญหาลวงจะเห็นได้ว่าปัญหา Htrap มีความยากมากกว่าปัญหา HIFF และจากผลการทดลองก็แสดงให้เห็นว่าในขนาดปัญหาใกล้เคียงกัน ขั้นตอนวิธีส่วนใหญ่ใช้จำนวน FEs ในการหาคำตอบสำหรับปัญหา Htrap มากกว่าปัญหา HIFF เมื่อพิจารณาผลในปัญหา Htrap วิธี sGA, sGA-FC และ RFT มีประสิทธิภาพต่ำ เนื่องจากปัญหา HTrap มีลักษณะของความพันแบบลำดับชั้นร่วมกับการลวง ดังนั้นวิธีที่มีประสิทธิภาพต่ำในการแก้ปัญหาแบบลำดับชั้นเช่นปัญหา HIFF จะมีประสิทธิภาพต่ำในปัญหา HTrap ด้วย (เนื่องจากเหตุผลเดียวกัน) และสำหรับวิธี sGA และ sGA-FC แม้ว่าจะสามารถแก้ปัญหาลวงได้หลายปัญหา รวมถึงสามารถแก้ปัญหาลำดับชั้นเช่นปัญหา HIFF ได้ด้วย แต่เมื่อเจอปัญหาที่มีลักษณะสองอย่างร่วมกันกลับทำให้มีประสิทธิภาพต่ำ สันนิษฐานว่าเกิดจากในขณะที่กำลังหาหน่วยการสร้างที่ถูกต้อง ก็มีเรื่องความสัมพันธ์แบบลำดับชั้นขัดขวางเอาไว้ จึงทำให้เกิดความขัดแย้งกันเป็นเหตุให้ไม่สามารถไปสู่คำตอบที่ต้องการได้

สำหรับปัญหาวงที่หน่วยการสร้างมีการเหลื่อมกันเช่นปัญหา Overlap-MK-Trap ซึ่งในการทดลองใช้ Trap-5 เป็นหลักแล้วกำหนดให้มีการเหลื่อมกันเกิดขึ้นสามระดับคือ เหลื่อมกันหนึ่งบิต สองบิต และสามบิต แต่กำหนดให้มีจำนวนหน่วยการสร้างเท่ากันหมดนั่นคือ 30 หน่วย (ถ้าเปรียบกับปัญหา Trap-5 ปกติจะมีขนาดปัญหาเป็น 150 บิต) ปัญหานี้ใช้ทดสอบความสามารถในการแก้ปัญหาที่หน่วยการสร้างมีการทับซ้อนกันซึ่งจะต่างจากปัญหาอื่นโดยทั่วไปที่มักจะกำหนดให้หน่วยการสร้างไม่มีการเหลื่อมกัน การทับซ้อนกันของโครงสร้างย่อยอาจทำให้เกิดความกำกวมและความอ่อนไหวในการหาหน่วยการสร้าง ลองพิจารณาตามตัวอย่างต่อไปนี้ ตัวอย่างปัญหา Trap-5 มีหน่วยการสร้างเป็น 11111 เมื่อมีการเหลื่อมกัน 2 บิต จากหน่วยการสร้างสองหน่วยจะมีรูปแบบเป็น 11111111 แต่ถ้ามีตัวใดตัวหนึ่งเป็นหน่วยการสร้างที่ไม่ถูกต้องดังนี้ 11101111 จะเห็นว่าหน่วยการสร้างแรกมีค่าบิตที่สี่เป็น 0 (11101) จะทำให้หน่วยการสร้างที่สองมีบิตแรกเป็น 0 (01111) ด้วย และทำให้มีค่าความเหมาะสมเท่ากับ 0 ซึ่งถ้าเทียบกับ Trap-5 แบบปกติจะมีค่าความเหมาะสมเป็น 0 ได้จะต้องมีบิตที่เป็น 0 จำนวน 2 บิต (เช่น 11110 11011 เป็นต้น) ดังนั้นจะเห็นว่าหน่วยการสร้างทั้งสองหน่วยขึ้นแก่กัน (ไม่เป็นอิสระจากกัน) จากการทดลองเมื่อมีการเหลื่อมกันมากขึ้นจะพิจารณาว่ามีความซับซ้อนเพิ่มขึ้น แต่ในทางตรงกันข้ามการเหลื่อมกันทำให้ขนาดของปัญหา (จำนวนบิต) ลดน้อยลงซึ่งทำให้ปริภูมิค้นหาเล็กลงด้วย ด้วยเหตุนี้จึงทำให้พิจารณาได้ว่าปัญหามีความยากน้อยลงด้วย จากผลการทดลองแสดงให้เห็นว่าถ้าหน่วยการสร้างมีจำนวนเท่าเดิมแต่มีการเหลื่อมกันมากขึ้น จะทำให้ค่า FEs ที่ใช้ในการหาค่าตอบลดลงด้วย สาเหตุที่ทำให้วิธี sGA-FC และวิธี RTF มีประสิทธิภาพต่ำสันนิษฐานว่าเป็นเพราะวิธี sGA-FC ไม่สามารถแลกเปลี่ยนชิ้นส่วนที่มีการทับซ้อนกันได้อย่างถูกต้อง และวิธี RFT ไม่มีความลำเอียงเกี่ยวกับโครงสร้างของหน่วยการสร้างที่ดี หรือกล่าวได้ว่าวิธี RFT ไม่ได้สนใจว่าโครงสร้างที่ดีจะมีรูปแบบเป็นอย่างไรจึงทำให้ไม่สามารถระบุรูปแบบชิ้นส่วนที่ดีได้ ส่วนวิธี CFT ให้ผลที่มีประสิทธิภาพโดดเด่นกว่าวิธีอื่นเป็นอย่างมาก สันนิษฐานได้ว่าเกิดจากการที่ชิ้นส่วนถูกแบ่งออกเป็นส่วนย่อย ๆ โดยแบ่งออกเป็นส่วนที่ทับซ้อนกันและส่วนที่ไม่ทับซ้อนกัน หรือกล่าวอีกนัยหนึ่งได้ว่าวิธีการระบุหน่วยการสร้างของวิธี CFT เข้ากันกับรูปแบบหน่วยการสร้างที่มีการทับซ้อนกันได้ ซึ่งมีส่วนช่วยในการลดความซับซ้อนของปัญหาให้น้อยลง และยังช่วยให้กระบวนการประกอบชิ้นส่วนเป็นไปอย่างมีประสิทธิภาพ

ปัญหาทดสอบหน่วยการสร้างส่วนใหญ่จะออกแบบให้หน่วยการสร้างมีขนาดคงที่ เพื่อให้ง่ายต่อการวิเคราะห์และการทำความเข้าใจ ขั้นตอนวิธีเชิงพันธุกรรมโดยทั่วไปไม่จำเป็นต้องมีความสามารถในการระบุขนาดของหน่วยการสร้างได้อย่างชัดเจนแต่ก็ยังสามารถหาคำตอบได้ด้วยเหตุนี้จึงเกิดข้อสงสัยว่าถ้าปัญหามีขนาดหน่วยการสร้างไม่คงที่ ขั้นตอนวิธีเชิงพันธุกรรมยังจะ

สามารถหาคำตอบได้หรือไม่ ถ้าพิจารณาจากกลไกการทำงานแล้วอาจสันนิษฐานล่วงหน้าได้เลยว่าขั้นตอนวิธีเชิงพันธุกรรมน่าจะหาคำตอบได้ เป็นที่มาของปัญหา Vary Size Trap ที่ออกแบบให้มีหน่วยการสร้างขนาดต่างกัน 3 ขนาดได้แก่ขนาด 3 บิต 4 บิต และ 5 บิต ซึ่งแบ่งออกเป็นสองปัญหาย่อยที่มีขนาดเท่ากันคือ 60 บิต ต่างกันที่ลำดับการจัดเรียงขนาดของหน่วยการสร้าง ปัญหาย่อยแรกเรียงลำดับขนาดหน่วยการสร้างแบบ 3 บิต + 4 บิต + 5 บิต (คิดเป็น 1 ชุด มีขนาด 12 บิต) โดยมีจำนวนทั้งหมด 5 ชุดเรียงต่อกัน (รวมกันเป็น 60 บิต) และปัญหาย่อยที่สองเรียงลำดับแบบ (3 บิต x 5 หน่วย) + (4 บิต x 5 หน่วย) + (5 บิต x 5 หน่วย) (รวมกันเป็น 60 บิต) ปัญหาย่อยทั้งสองปัญหามีจำนวนหน่วยการสร้างแต่ละขนาดเท่ากัน คือขนาดละ 5 หน่วย ปัญหาย่อยแรกหน่วยการสร้างขนาดเดียวกันจะกระจายตัวกันอย่างสม่ำเสมอ แต่ปัญหาย่อยที่สองหน่วยการสร้างขนาดเดียวกันจะเกาะกันเป็นกลุ่ม ปัญหานี้ถูกออกแบบมาเพื่อทดสอบความสามารถในการแก้ปัญหาที่หน่วยการสร้างมีขนาดไม่คงที่ และมีการกระจายตัวในรูปแบบที่ต่างกัน ผลการทดสอบแสดงให้เห็นว่าค่า FEs ที่ใช้ในการแก้ปัญหาทั้งสองที่หนึ่งและปัญหาย่อยที่สองจากขั้นตอนวิธีเดียวกันมีค่าใกล้เคียงกัน และสามารถสรุปได้ว่าขั้นตอนวิธีที่นำเสนอรวมถึง sGA สามารถแก้ปัญหาที่หน่วยการสร้างมีขนาดไม่เท่ากันได้อย่างมีประสิทธิภาพ และรูปแบบการกระจายตัวของหน่วยการสร้างที่มีขนาดต่างกันมีผลต่อประสิทธิภาพในการหาคำตอบน้อย

โดยปกติแล้วปัญหาทดสอบจะถูกออกแบบให้มีรูปแบบโครงสร้างเป็นเนื้อเดียวกัน แต่สำหรับปัญหาในโลกจริงบางปัญหาไม่สามารถบอกได้ว่าโครงสร้างที่แน่นอนเป็นอย่างไร เพราะบางปัญหาอาจประกอบขึ้นจากปัญหาย่อยหลายแบบรวมกัน ดังนั้นจึงมีแนวคิดที่จะนำปัญหาทดสอบหลายรูปแบบมารวมให้เป็นปัญหาเดียวกัน จากปัญหา Mixed Trap ที่ใช้ในการทดสอบเป็นปัญหาที่เกิดขึ้นจากปัญหาหลายรูปแบบรวมกันทั้งปัญหาลวงและปัญหาไม่ลวง เพื่อใช้ทดสอบความสามารถในการแก้ปัญหาที่มีรูปแบบผสม (ทดสอบความสามารถในการหาหน่วยการสร้างที่มีหลายรูปแบบในปัญหาเดียว) ผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีที่นำเสนอทั้งหมดรวมถึง sGA สามารถแก้ปัญหาได้อย่างมีประสิทธิภาพ ซึ่งเป็นการแสดงให้เห็นถึงแนวโน้มหรือความเป็นไปได้ที่จะนำขั้นตอนวิธีที่นำเสนอรวมถึง sGA ไปใช้แก้ปัญหาในโลกจริงได้อย่างมีประสิทธิภาพเช่นกัน

สำหรับกลุ่มปัญหาที่ไม่สามารถระบุหน้าตาหรือรูปแบบของหน่วยการสร้างที่ชัดเจนได้แก่ปัญหา F1 ถึง F4 ที่เป็นปัญหาทดสอบความสามารถในการหาค่าเหมาะสมที่สุด สามารถแบ่งออกเป็นกลุ่มย่อยได้จากคุณสมบัติเฉพาะดังนี้ คุณสมบัติการแบ่งออกจากกันได้ (Separable) ได้แก่ปัญหา F1 กับ F4 และคุณสมบัติที่มีหลายโหมด (Multimodal) ได้แก่ปัญหา F3 กับ F4 ปัญหาในกลุ่มนี้ใช้ทดสอบประสิทธิภาพในการแก้ปัญหาเมื่อภูมิทัศน์ความเหมาะสมมีรูปแบบ

แตกต่างกัน ผลการทดลองในภาพรวมแสดงให้เห็นว่าปัญหาที่มีหลายโหมดต้องใช้จำนวน FEs ในการหาคำตอบมากกว่าปัญหาที่มีโหมดเดียว (Unimodal) เมื่อพิจารณาจากขนาดปัญหาที่มีค่าใกล้เคียงกัน และปัญหาที่มีคุณสมบัติแบ่งออกจากกันได้จะใช้จำนวน FEs น้อยกว่าปัญหาที่มีคุณสมบัติไม่สามารถแบ่งออกจากกันได้ (Non-separable) สำหรับปัญหา F1 วิธี LZWGA-1 มีประสิทธิภาพในการแก้ปัญหาต่ำ เนื่องจากไม่มีความรู้เพียงพอในการระบุหน่วยการสร้างและการบีบอัดทำให้รูปแบบการแลกเปลี่ยนที่เป็นไปได้ลดลง (ขัดขวางไม่ให้เกิดการแลกเปลี่ยนอย่างอิสระ) ส่วนปัญหา F3 เมื่อพิจารณาถึงภูมิทัศน์ความเหมาะสมจะเห็นว่าความยากอยู่ที่การไปให้ถึงคำตอบซึ่งจะต้องผ่านช่องทางแคบ ๆ ที่ผ่านไปได้ยาก จากผลการทดลองวิธี RFT และ CFT ไม่สามารถหาคำตอบภายใต้เงื่อนไขที่กำหนดได้ เป็นเพราะปัญหานี้ต้องการความสามารถในการรักษาโครงสร้างในภาพรวมสูง (ไม่เช่นนั้นจะผ่านช่องทางไปไม่ได้) แต่วิธี RFT และ CFT มีความหลากหลายที่สูงเกินไป (ทั้งในขั้นตอนการระบุและการประกอบหน่วยการสร้าง) จึงไม่สามารถรักษาโครงสร้างในภาพรวมเอาไว้ได้ และเมื่อพิจารณาผลของวิธี sGA-FC มีความโดดเด่นกว่าขั้นตอนวิธีอื่นเนื่องจากการดำเนินการจะทำการแลกเปลี่ยนกันเฉพาะส่วนที่ต่างกันเท่านั้น (ส่วนที่เหมือนกันยังคงอยู่เหมือนเดิม) ซึ่งพิจารณาได้ว่าเป็นการรักษาโครงสร้างในภาพรวมให้คงอยู่มากที่สุด และยังคงเกิดการแลกเปลี่ยนหน่วยการสร้างในระดับสูง (เพราะไม่มีการทำลายหน่วยการสร้างที่ระบุได้เลย)

ประสิทธิภาพในการแก้ปัญหาทดสอบจากตารางที่ 11.5 สำหรับกรณีที่มีค่า FEs บางค่าต่างกันไม่มากในปัญหาเดียวกัน (แต่คนละขั้นตอนวิธี) จะแน่ใจได้อย่างไรว่าความแตกต่างนี้มีผลอย่างมีนัยสำคัญ ในส่วนนี้จึงได้มีการทดสอบทางสถิติโดยการทำ paired t-Test เพื่อเปรียบเทียบประสิทธิภาพในการแก้ปัญหาของแต่ละขั้นตอนวิธีกับปัญหาทดสอบย่อยเดียวกันว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ความเชื่อมั่น 95% และแสดงเฉพาะกรณีที่ค่า FEs ที่ได้ไม่แตกต่างกันอย่างมีนัยสำคัญเพียงพอเท่านั้น ดังตารางที่ 11.7 จากผลการทดลองสามารถสรุปประสิทธิภาพในการแก้ปัญหาและปัญหาย่อย โดยจัดเรียงลำดับความสามารถตามค่าจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม (มีค่าน้อยยิ่งดี) เพื่อใช้เปรียบเทียบให้เห็นความสามารถในการแก้ปัญหาระหว่างขั้นตอนวิธีต่าง ๆ ดังตารางต่อไป นี้ หมายเหตุ จะมีกรณีที่ค่าลำดับเท่ากันเนื่องจากผลการทดสอบ paired t-Test ในตารางที่ 11.7 บ่งบอกว่าไม่มีความแตกต่างกันอย่างมีนัยสำคัญ

ตารางที่ 11.5 (ส่วนที่ 1) จำนวนการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของแต่ละขั้นตอนวิธีกับแต่ละปัญหาทดสอบ

Fitness Evaluation		sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
Problems	Problem size	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs
OneMax with Mutation	100	3,054	947	14,000	5,100	–	1,960	1,310	1,734	3,810	3,100	2,874	1,637
	150	5,034	1,657	32,500	8,300	–	2,954	1,984	2,510	6,920	5,120	4,677	2,354
	200	12,320	2,110	60,000	12,500	–	4,144	2,627	3,497	34,867	12,114	7,147	3,200
	250	38,934	2,687	80,000	15,700	–	6,127	3,177	4,607	452,000	59,627	23,781	4,337
RandomMax with Mutation	100	3,037	990	–	–	–	1,990	1,250	1,754	3,594	3,084	2,994	1,667
	150	5,060	1,534	–	–	–	2,827	2,000	2,390	6,000	5,144	4,814	2,404
	200	12,234	2,057	–	–	–	4,284	2,514	3,330	28,834	12,920	7,694	3,207
	250	37,800	2,704	–	–	–	6,447	3,380	4,584	347,000	57,841	30,294	3,434
OneMax	100	22,230	1,267	14,000	5,100	–	3,840	1,844	7,234	61,600	25,650	112,467	2,367
	150	64,307	1,654	32,500	8,300	–	6,560	4,350	14,967	290,767	77,817	198,900	3,034
	200	128,250	2,680	60,000	12,500	–	9,834	6,214	37,867	774,667	176,800	411,600	4,134
	250	271,150	3,127	80,000	15,700	–	11,500	9,080	49,467		286,500	650,000	4,720
RandomMax	100	21,970	1,070	–	–	–	3,911	1,834	7,234	71,334	39,500	84,884	2,160
	150	63,840	1,671	–	–	–	6,534	4,387	14,767	289,467	81,784	179,587	3,400
	200	137,834	2,727	–	–	–	9,784	6,107	37,867	756,934	180,484	357,500	5,820
	250	285,600	3,127	–	–	–	11,384	9,174	49,600		347,600	548,625	8,367

หมายเหตุ ช่องที่บ่งชี้หมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข ช่องว่างหมายถึงไม่มีผลการทดลอง และตัวหนาหมายถึงขั้นตอนวิธีที่ให้ผลดีที่สุด (ถ้าวิธีเปรียบเทียบให้ผลดีกว่าวิธีที่นำเสนอจะแสดงเป็นตัวหนาด้วย)

หมายเหตุ ผลลัพธ์ทั้งหมดของ sGA และ UMDA ได้มาจากโปรแกรมที่เขียนขึ้นเอง

ผลของ CSM และ BOA อ้างอิงจากงานวิจัย (Aporntewan and Chongstitvatana, 2004) เฉพาะปัญหา f6-bipolar ผลของ BOA อ้างอิงจากงานวิจัยปริญาเอก (Pelikan, 2002) และผลของ QHFC อ้างอิงจากงานวิจัย (Hu, Goodman and Rosenberg, 2004)



ตารางที่ 11.6 (ส่วนที่ 1) จำนวนประชากรและประชากรย่อยของแต่ละขั้นตอนวิธีกับแต่ละปัญหาทดสอบ

Parameters		sGA	UMDA		sGA-FC	RFT		CFT		LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA	
Problems	Problem size	#pop.	#pop.	#sub.	#pop.	#pop.	#sub.	#pop.	#sub.	#pop.	#pop.	#pop.	#pop.	#sub.
OneMax with Mutation	100	100	100	10	100	100	10	100	10	100	100	100	100	10
	150	100	100	10	100	100	10	100	10	100	100	100	100	10
	200	200	100	10	100	100	10	100	10	500	200	100	100	10
	250	500	100	10	100	100	10	100	10	6000	800	300	100	10
RandomMax with Mutation	100	100	100	10	100	100	10	100	10	100	100	100	100	10
	150	100	100	10	100	100	10	100	10	100	100	100	100	10
	200	200	100	10	100	100	10	100	10	500	200	100	100	10
	250	500	100	10	100	100	10	100	10	6,000	800	300	100	10
OneMax	100	1,300	100	30	300	100	30	1,000	50	3,500	2,500	7,000	200	50
	150	2,800	100	40	400	100	40	1,000	150	13,000	3,500	8,500	200	50
	200	4,500	200	40	500	400	30	4,000	100	28,000	6,500	14,000	200	70
	250	8,500	200	40	500	400	40	4,000	150	28,000	9,000	20,000	200	70
RandomMax	100	1,300	100	30	300	100	30	1,000	50	4,000	2,500	5,500	200	70
	150	2,800	100	40	400	100	40	1,000	150	13,000	3,500	8,000	200	70
	200	5,000	200	40	500	400	30	4,000	100	28,000	6,500	12,500	300	100
	250	9,000	200	40	500	400	40	4,000	150	28,000	11,000	16,500	500	100

หมายเหตุ ช่องที่บหมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข และตัวหนาหมายถึงขั้นตอนวิธีที่ให้ผลดีที่สุด (ถ้าวิธีเปรียบเทียบให้ผลดีกว่าวิธีที่นำเสนอจะแสดงเป็นตัวหนาด้วย)



ตารางที่ 11.5 (ส่วนที่ 2) จำนวนการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของแต่ละขั้นตอนวิธีกับแต่ละปัญหาทดสอบ

Fitness Evaluation		sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
Problems	Problem size	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs
RoyalRoad	64	11,840	-	-	-	-	13,760	21,734	13,800	16,980	11,720	-	4,650
	128	49,500	-	-	-	-	29,867	94,850	28,700	132,267	37,620	-	8,867
	256	304,400	-	-	-	-	75,167	273,000	73,334	787,000	287,467	-	22,860
Random RoyalRoad	64	12,280	-	-	-	-	13,920	21,884	13,867	23,614	15,574	-	6,600
	128	49,100	-	-	-	-	30,240	100,151	29,400	136,000	57,517	-	15,667
	256	306,621	-	-	-	-	75,750	262,167	81,250	883,300	297,367	-	34,907
Dec-3	60	25,560	-	-	27,000	2,000	14,134	107,550	8,600	47,134	22,560	-	19,520
	120	217,234	-	-	80,000	5,000	51,187		27,734	562,500	146,800	-	36,480
	180	626,000	-	-	180,000	16,500	92,100		52,600		709,167	-	56,134
	240		-	-	235,000	34,000	147,067		79,734			-	91,840
Exp-Dec-3	60	181,500	-	-	165,000	-	55,260	46,467	29,520	191,100	127,050	-	15,147
	90	732,000	-	-	480,000	-	112,584	73,134	113,100		586,134	-	27,360
	120		-	-	1,000,000	-	247,500	99,400	250,400			-	64,667
Trap-5	100	83,250	-	65,000	99,000	4,500	47,917	644,000	21,200	138,834	72,380	-	21,487
	150	305,500	-	165,000	220,000	9,200	114,000		31,334	564,167	246,400	-	38,610
	200	784,934	-	310,000	320,000	20,000	215,600		56,401		565,334	-	49,701
	250		-	750,000	490,000	37,200	375,900		102,934			-	74,620
f6-bipolar	60		-	-	31,000	2,000		104,100	53,300	28,080	20,457	-	30,334
	120		-	-	124,500	5,000		961,000		200,667	110,000	-	97,960
	180		-	-	-	15,000				540,667	322,400	-	260,001
	240		-	-	-	36,800					896,934	-	637,000

หมายเหตุ ช่องที่บ่งชี้ว่าไม่สามารถหาคำตอบได้ตามเงื่อนไข ช่องว่างหมายถึงไม่มีผลการทดลอง และตัวหนาหมายถึงขั้นตอนวิธีที่ให้ผลดีที่สุด (ถ้าวิธีเปรียบเทียบให้ผลดีกว่าวิธีที่นำเสนอจะแสดงเป็นตัวหนาด้วย)

ตารางที่ 11.6 (ส่วนที่ 2) จำนวนประชากรและประชากรย่อยของแต่ละขั้นตอนวิธีกับแต่ละปัญหาทดสอบ

Parameters		sGA	UMDA		sGA-FC	RFT		CFT		LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA	
Problems	Problem size	#pop.	#pop.	#sub.	#pop.	#pop.	#sub.	#pop.	#sub.	#pop.	#pop.	#pop.	#pop.	#sub.
RoyalRoad	64	1,200	-	-	1,200	1,000	200	2,000	200	1,800	1,200	-	500	100
	128	3,000	-	-	1,600	1,500	300	3,500	200	8,000	2,200	-	1,000	100
	256	12,000	-	-	2,500	6,000	300	5,500	300	30,000	11,000	-	1,800	200
Random RoyalRoad	64	1,200	-	-	1,200	1,300	200	2,000	200	2,200	1,600	-	1,000	100
	128		-	-	1,600	1,500	300	3,500	200	8,000	3,500	-	2,000	100
	256		-	-	2,500	5,000	300	6,500	300		11,000	-	2,200	200
Dec-3	60	1,500	-	-	1,000	4,500	400	600	150	3,500	1,600	-	800	300
	120	9,500	-	-	2,200			2,000	200	25,000	6,000	-	1,200	350
	180	20,000	-	-	3,000			3,000	300		23,000	-	2,000	350
	240	26,000	-	-	4,000			4,000	350			-	3,200	400
Exp-Dec-3	60	7,500	-	-	2,700	2,000	300	1,800	200	9,000	5,500	-	800	200
	90	18,000	-	-	3,500	2,000	300	6,500	250		16,000	-	1,200	200
	120		-	-	5,500	2,000	300	12,000	300			-	2,500	300
Trap-5	100	4,500	-	-	2,500	15,000	1,500	2,000	200	8,500	4,200	-	2,200	200
	150	13,000	-	-	4,500			2,500	200	25,000	11,000	-	3,300	200
	200	28,000	-	-	7,000			4,500	200		20,000	-	3,000	300
	250		-	-	10,500			8,000	250			-	4,200	400
f6-bipolar	60		-	-		4,500	400	13,000	100	2,400	1,700	-	2,000	300
	120		-	-		15,000	800			10,000	5,500	-	6,200	300
	180		-	-						20,000	12,000	-	15,000	300
	240		-	-							28,000	-	30,000	400

หมายเหตุ ช่องที่บหมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข ช่องว่างหมายถึงไม่มีผลการทดลอง และตัวหนาหมายถึงขั้นตอนวิธีที่ให้ผลดีที่สุด (ถ้าวิธีเปรียบเทียบให้ผลดีกว่าวิธีที่นำเสนอจะแสดงเป็นตัวหนาด้วย)

ตารางที่ 11.5 (ส่วนที่ 3) จำนวนการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสมของแต่ละขั้นตอนวิธีกับแต่ละปัญหาทดสอบ

Fitness Evaluation		sGA	UMDA	CSM	hBOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
Problems	Problem size	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs	#FEs
HIFF	32	4,800	–	3,300	2,100	–	2,634	11,834	1,160	3,687	3,180	–	2,217
	64	38,800	–	14,500	7,800	–	11,840	222,200	5,200	35,000	24,250	–	6,067
	128	584,000	–	51,000	27,000	–	45,920		17,467	761,667	305,334	–	47,017
	256		–	370,000	90,000	–	222,700		71,284			–	366,667
Htrap-1	27	11,360	–	3,000	3,400	–	8,000	36,200	2,060	7,980	6,900	–	4,200
	81		–	35,000	30,000	–			10,667	441,334	326,000	–	67,334
	243		–	310,000	225,000	–			115,000			–	
Overlap Trap	5-1	121	230,534	–	–	–	676,000		6,860	291,200	154,134	–	63,450
	5-2	92	118,934	–	–	–			4,547	130,200	89,600	–	58,667
	5-3	63	63,000	–	–	–			2,980	53,167	43,867	–	48,317
F1	40	50,034	–	–	–	–	18,650	81,400	18,000	103,734	47,367	–	31,360
	80	369,000	–	–	–	–	55,500	894,000	89,700		469,284	–	118,667
F2	40	140,000	–	–	–	–	31,750	80,500	22,917	308,934	148,634	–	63,000
F3	24	424,000	–	–	–	–	60,400			693,334	586,667	–	234,667
F4	50	106,400	–	–	–	–	80,534	474,000	30,917	242,317	126,667	–	77,867
Vary Size Trap	v1	60	21,060	–	–	–	10,470	77,600	5,620	27,600	18,984	–	6,630
	v2	60	19,050	–	–	–	14,080	76,834	5,740	22,800	18,757	–	6,467
Mixed Trap	128	25,111	–	–	–	–	5,227	12,750	3,161	12,467	14,934	–	5,680

หมายเหตุ ช่องที่บหมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข ช่องว่างหมายถึงไม่มีผลการทดลอง และตัวหนาหมายถึงขั้นตอนวิธีที่ให้ผลดีที่สุด (ถ้าวิธีเปรียบเทียบให้ผลดีกว่าวิธีที่นำเสนอจะแสดงเป็นตัวหนาด้วย)

ตารางที่ 11.6 (ส่วนที่ 3) จำนวนประชากรและประชากรย่อยของแต่ละขั้นตอนวิธีกับแต่ละปัญหาทดสอบ

Parameters		sGA	UMDA		sGA-FC	RFT		CFT		LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA		
Problems	Problem size	#pop.	#pop.	#sub.	#pop.	#pop.	#sub.	#pop.	#sub.	#pop.	#pop.	#pop.	#pop.	#sub.	
HIFF	32	900	-	-	500	2,500	200	200	50	700	600	-	500	100	
	64	4,000	-	-	1,000	33,000	1,000	500	150	3,500	2,500	-	1,000	100	
	128	40,000	-	-	2,800			2,000	200	50,000	20,000	-	6,500	300	
	256		-	-	8,500			6,500	400			-	40,000	1,000	
Htrap-1	27	2,400	-	-	1,500	6,000	1,000	300	100	1,800	1,500	-	1,500	100	
	81		-	-				1,000	200	40,000	30,000	-	10,000	700	
	243		-	-				11,500	300			-	10,000	700	
Overlap Trap	5-1	121	13,000	-	-	30,000			600	100	16,000	8,500	-	4,500	800
	5-2	92	7,000	-	-				400	100	9,000	6,000	-	5,500	800
	5-3	63	6,000	-	-				300	100	5,000	4,000	-	6,500	800
F1	40	3,800	-	-	1,500	6,000	800	2,500	200	8,000	3,500	-	4,200	200	
	80	15,000	-	-	2,500	60,000	1,500	9,000	300		18,500	-	10,000	300	
F2	40	12,000	-	-	2,500	5,000	800	2,500	300	28,000	13,000	-	10,000	400	
F3	24	60,000	-	-	6,000					100,000	100,000	-	80,000	600	
F4	50	7,000	-	-	4,000	60,000	1,500	3,500	300	15,500	8,000	-	8,000	500	
Vary Size Trap	v1	60	1,800	-	-	900	6,000	500	600	100	2,400	1,700	-	900	100
	v2	60	1,500	-	-	1,200	5,000	500	600	100	2,000	1,700	-	1,000	100
Mixed Trap	128	3,100	-	-	800	2,500	200	600	100	1,700	2,000	-	1,200	200	

หมายเหตุ ช่องที่บหมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข ช่องว่างหมายถึงไม่มีผลการทดลอง และตัวหนาหมายถึงขั้นตอนวิธีที่ให้ผลดีที่สุด (ถ้าวิธีเปรียบเทียบให้ผลดีกว่าวิธีที่นำเสนอจะแสดงเป็นตัวหนาด้วย)

จากตารางก่อนหน้าได้เห็นประสิทธิภาพในการแก้ปัญหาของแต่ละขั้นตอนวิธีในแต่ละปัญหาทดสอบ ในส่วนนี้จะทำการวัดสมรรถนะ (Performance) ในการแก้ปัญหาเพื่อช่วยให้เห็นความสามารถในการแก้ปัญหาของแต่ละขั้นตอนวิธีในภาพรวม การวัดสมรรถนะทำได้หลายวิธีขึ้นกับแต่ละมุมมอง ในงานวิจัยนี้มุ่งเน้นไปที่ประสิทธิภาพในการแก้ปัญหาทดสอบเป็นหลัก การคำนวณพิจารณาจากประสิทธิภาพในการแก้ปัญหาโดยจะคิดเฉพาะกรณีที่สามารถหาคำตอบได้ และมีความสามารถในการแก้ปัญหาเป็นอันดับที่ 1 ถึง 3 เนื่องจากบางปัญหาผลที่ได้ระหว่างอันดับที่ 1 และ 2 มีค่าต่างกันน้อยมากเมื่อเทียบเป็นอัตราส่วน ดังนั้นจึงถือว่าถ้าได้อันดับที่ 1 ถึง 3 มีประสิทธิภาพในการแก้ปัญหาที่ดีและยอมรับได้ สำหรับแต่ละปัญหาสามารถคำนวณสมรรถนะในการแก้ปัญหาได้โดยนับจำนวนปัญหาย่อยที่สามารถแก้ได้อย่างมีประสิทธิภาพอยู่ในอันดับที่ 1 ถึง 3 ตัวอย่างเช่น จากตารางที่ 11.7 (ส่วนที่ 2) จากปัญหา Royal Road จะเห็นว่าขั้นตอนวิธีเชิงพันธุกรรมมีประสิทธิภาพในการแก้ปัญหาย่อยได้ดีติดอันดับ 1 ถึง 3 อยู่เพียงปัญหาย่อยเดียว (ปัญหาขนาด 64 บิต) จากจำนวนปัญหาย่อยทั้งหมด 3 ปัญหา ดังนั้นค่าสมรรถนะในการแก้ปัญหา Royal Road สำหรับขั้นตอนวิธีเชิงพันธุกรรมมีค่าเท่ากับ $1/3$ หรือประมาณ 0.33 เป็นต้น ดังนั้นถ้าสามารถแก้ปัญหาย่อยในปัญหาหลักได้หมดจะได้ค่าสมรรถนะสูงสุดเท่ากับ 1 เกณฑ์การแบ่งปัญหาหลักและปัญหาย่อย รวมถึงผลการคำนวณแสดงดังตารางที่ 11.9 สำหรับปัญหา OneMax และปัญหา RandomMax ที่แยกคิดกรณีที่มีการใช้การผ่าเหล่าเป็นอีกปัญหาหนึ่งเพราะถือว่ามีความแตกต่าง และตารางที่ 11.10 เป็นการนับรวมประสิทธิภาพการแก้ปัญหาจากจำนวนปัญหาย่อยทั้งหมด รวมถึงแสดงจำนวนปัญหาย่อยที่ไม่สามารถหาคำตอบได้ เพื่อแสดงให้เห็นถึงมุมมองเรื่องสมรรถนะในอีกมุมมองหนึ่ง

เมื่อสรุปผลสมรรถนะในการแก้ปัญหาของแต่ละขั้นตอนวิธีจากรูปที่ 11.1 ขั้นตอนวิธีที่นำเสนอนี้ที่มีสมรรถนะดีสุดสามอันดับแรกคือ ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA) ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (CFT) และวิธีการไขว้เปลี่ยนแบบชิ้นส่วน (sGA-FC) ตามลำดับ สำหรับวิธี UMDA นั้นได้ผลดีเฉพาะกับปัญหาง่าย ซึ่งก็คือปัญหา OneMax ที่ใช้ในการทดสอบเท่านั้น (ทดสอบเพียง 4 ปัญหา) และวิธี QHFC เป็นขั้นตอนวิธีที่ออกแบบมาเพื่อแก้ปัญหาบางกลุ่มโดยเฉพาะซึ่งมีทั้งหมด 3 ปัญหา ต่างจากขั้นตอนวิธีที่นำเสนอและขั้นตอนวิธีเชิงพันธุกรรม (ใช้เป็นตัวเปรียบเทียบหลัก) ที่ทดสอบกับปัญหา ทั้งหมดจำนวน 19 ปัญหา

ตารางที่ 11.7 (ส่วนที่ 1) ผลการทดสอบ t-Test ที่ความเชื่อมั่น 95% เพื่อทดสอบความแตกต่างระหว่างผลที่ได้จากแต่ละขั้นตอนวิธีในปัญหาทดสอบเดียวกัน และแสดงเฉพาะกรณีที่เกิดจากขั้นตอนวิธีที่ได้ไม่แตกต่างกันอย่างมีนัยสำคัญ

Studen Paired t-Test		sGA	sGA	sGA	FC	FC	FC	RFT	RFT	CFT	LZWGA-1	LZWGA-2
Problems	Problem size	RFT	LZWGA-2	LZWGA-3	CFT	LZWGA-1	LZWEDA	LZWGA-1	LZWGA-2	LZWEDA	LZWGA-2	LZWGA-3
OneMax with Mutation	100	0	0.5751	0	0	0	0	0	0	0.0788	0	0
	150	0	0.3831	0	0	0	0	0	0	0	0	0
	200	0	0.5031	0	0	0	0	0	0	0	0	0
	250	0	0	0	0	0	0	0	0	0.1781	0	0
RandomMax with Mutation	100	0	0.5644	0.6164	0	0	0	0	0	0.2611	0	0.3335
	150	0	0.5362	0	0	0	0	0	0	0.8683	0	0
	200	0	0	0	0	0	0	0	0	0.3725	0	0
	250	0	0	0	0	0	0	0	0	0.7151	0	0
RR-1	64	0	0.5861	0	0	0	0	0	0	0	0	0
	128	0	0	0	0.2319	0	0	0	0	0	0	0
Random RoyalRoad	128	0	0	0	0.1230	0	0	0	0	0	0	0
Trap-5	100	0	0	0	0	0	0	0	0	0.3746	0	0
Htrap-1	27	0	0	0	0	0.4536	0	0	0	0	0	0
Vary Size Trap v2	60	0	0.8699	0	0	0	0	0	0	0	0	0
Mixed Trap	128	0	0	0	0	0	0	0.0759	0	0	0	0

หมายเหตุ แสดงค่าถึงทศนิยมตำแหน่งที่ 4

ตารางที่ 11.7 (ส่วนที่ 2) ผลการทดสอบ t-Test ที่ความเชื่อมั่น 95% เพื่อทดสอบความแตกต่างระหว่างผลที่ได้จากแต่ละขั้นตอนวิธีในปัญหาทดสอบเดียวกัน และแสดงเฉพาะกรณีที่เกิดจากขั้นตอนวิธีที่ไม่แตกต่างกันอย่างมีนัยสำคัญ

Student Paired t-Test		sGA	sGA	sGA	FC	FC	FC	RFT	RFT	CFT	LZWGA-1	LZWGA-2
Problems	Problem size	RFT	LZWGA-2	LZWGA-3	CFT	LZWGA-1	LZWEDA	LZWGA-1	LZWGA-2	LZWEDA	LZWGA-2	LZWGA-3
F1 CEC-2008	40	0	0.1284	0	0	0	0	0	0	0	0	0
F2 CEC-2008	40	0	0.1283	0	0	0	0	0	0	0	0	0
F3 CEC-2008	24	0	0	0	0	0	0	0	0	0	0.1154	0
F4 CEC-2008	50	0	0	0	0	0	0.5312	0	0	0	0	0
Vary Size Trap v2	60	0	0.8699	0	0	0	0	0	0	0	0	
Mixed Trap	128	0	0	0	0	0	0	0.0759	0	0	0	0

หมายเหตุ แสดงค่าถึงทศนิยมตำแหน่งที่ 4

หมายเหตุ ค่าที่แสดงในตารางเป็นค่า two-tail ที่ได้จากการคำนวณโดยใช้โปรแกรม Excel 2007 โดยใช้วิธี student paired t-Test ซึ่งตั้งสมมติฐาน H_0 ว่าผลเปรียบเทียบมีค่าต่างกัน และตั้งสมมติฐาน H_1 ว่าผลเปรียบเทียบมีค่าไม่ต่างกัน โดยที่ค่าความเชื่อมั่น 95% หรือ 0.95 ถ้าค่า two-tail มีค่ามากกว่า 0.05 จะแปลว่าปฏิเสธ H_0 และยอมรับ H_1 หรือกล่าวได้ว่าผลเปรียบเทียบมีค่าไม่ต่างกันที่ค่าความเชื่อมั่น 95%

ตารางที่ 11.8 (ส่วนที่ 1) ผลการจัดลำดับความสามารถในการแก้ปัญหาทดสอบสำหรับแต่ละขั้นตอนวิธี

Efficiency		Rank											
Problems	Problem size	sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
OneMax with Mutation	100	6	1	9	8	-	4	2	3	7	6	5	3
	150	7	1	10	9	-	5	2	4	8	7	6	3
	200	7	1	10	8	-	5	2	4	9	7	6	3
	250	7	1	9	5	-	4	2	3	10	8	6	3
RandomMax with Mutation	100	5	1	-	-	-	4	2	3	6	5	5	3
	150	6	1	-	-	-	4	2	3	7	6	5	3
	200	6	1	-	-	-	4	2	3	8	7	5	3
	250	6	1	-	-	-	4	2	3	8	7	5	3
OneMax	100	8	1	7	5	-	4	2	6	10	9	11	3
	150	8	1	7	5	-	4	3	6	11	9	10	2
	200	8	1	7	5	-	4	3	6	11	9	10	2
	250	8	1	7	5	-	4	3	6		9	10	2
RandomMax	100	6	1	-	-	-	4	2	5	8	7	9	3
	150	6	1	-	-	-	4	3	5	9	7	8	2
	200	6	1	-	-	-	4	3	5	9	7	8	2
	250	6	1	-	-	-	4	3	5		7	8	2

หมายเหตุ ช่องที่บหมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข และช่องว่างหมายถึงไม่มีผลการทดลอง

ตารางที่ 11.8 (ส่วนที่ 2) ผลการจัดลำดับความสามารถในการแก้ปัญหาทดสอบสำหรับแต่ละขั้นตอนวิธี

Efficiency		Rank											
Problems	Problem size	sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
RoyalRoad	64	2	-	-	-	-	3	6	4	5	2	-	1
	128	4	-	-	-	-	2	5	2	6	3	-	1
	256	6	-	-	-	-	3	4	2	7	5	-	1
Random RoyalRoad	64	2	-	-	-	-	4	6	3	7	5	-	1
	128	3	-	-	-	-	2	5	2	6	4	-	1
	256	6	-	-	-	-	2	4	3	7	5	-	1
Dec-3	60	6	-	-	7	1	3	9	2	8	5	-	4
	120	7	-	-	5	1	4		2	8	6	-	3
	180	6	-	-	5	1	4		2		7	-	3
	240		-	-	5	1	4		2			-	3
Exp-Dec-3	60	7	-	-	6	-	4	3	2	8	5	-	1
	90	7	-	-	5	-	3	2	4		6	-	1
	120		-	-	5	-	3	2	4			-	1
Trap-5	100	6	-	4	7	1	3	9	2	8	5	-	2
	150	8	-	5	6	1	4		2	9	7	-	3
	200	8	-	5	6	1	4		3		7	-	2
	250		-	6	5	1	4		3			-	2
f6-bipolar	60		-	-	5	1		7	6	3	2	-	4
	120		-	-	4	1		6		5	3	-	2
	180		-	-	-	1				4	3	-	2
	240		-	-	-	1					3	-	2

หมายเหตุ ช่องที่บ่งชี้หมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข และช่องว่างหมายถึงไม่มีผลการทดลอง

ตารางที่ 11.8 (ส่วนที่ 3) ผลการจัดลำดับความสามารถในการแก้ปัญหาดสอบสำหรับแต่ละขั้นตอนวิธี

Efficiency		Rank											
Problems	Problem size	sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
HIFF	32	8	-	6	2	-	4	9	1	7	5	-	3
	64	8	-	5	3	-	4	9	1	7	6	-	2
	128	7	-	5	2	-	3		1	8	6	-	4
	256		-	5	2	-	3		1			-	4
Htrap-1	27	7	-	2	3	-	6	8	1	6	5	-	4
	81		-	3	2	-			1	6	5	-	4
	243		-	3	2	-			1			-	
Overlap 5-1	121	4	-	-	-	-	6		1	5	3	-	2
Overlap 5-2	92	4	-	-	-	-			1	5	3	-	2
Overlap 5-3	63	5	-	-	-	-			1	4	2	-	3
F1	40	4	-	-	-	-	2	5	1	6	4	-	3
	80	4	-	-	-	-	1	6	2		5	-	3
F2	40	5	-	-	-	-	2	4	1	6	5	-	3
F3	24	3	-	-	-	-	1			4	4	-	2
F4	50	3	-	-	-	-	2	6	1	5	4	-	2
Vary Size Trap	60	4	-	-	-	-	3	6	1	5	4	-	2
	60	5	-	-	-	-	3	7	1	6	4	-	2
Mixed Trap	128	6	-	-	-	-	2	4	1	4	5	-	3

หมายเหตุ ช่องที่บหมายถึงไม่สามารถหาคำตอบได้ตามเงื่อนไข และช่องว่างหมายถึงไม่มีผลการทดลอง

ตารางที่ 11.9 ผลการคำนวณสมรรถนะของแต่ละขั้นตอนวิธีกับปัญหาทดสอบ

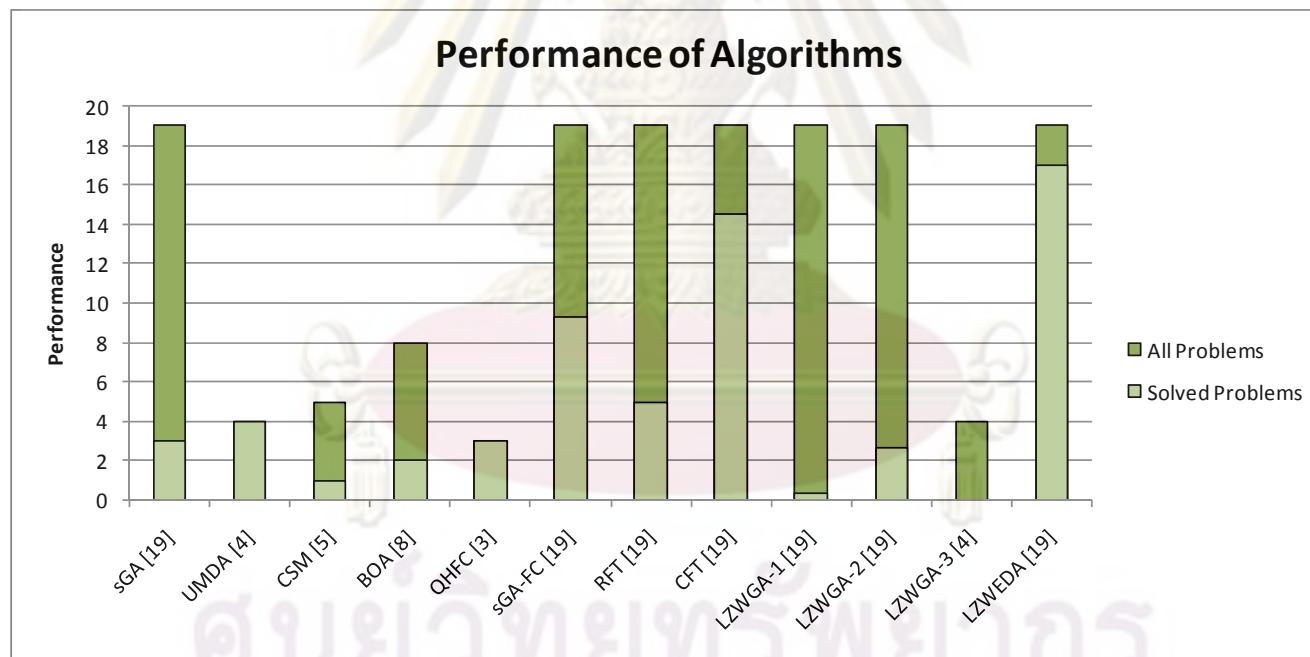
Problems		sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
1	OneMax+Mutation		1					1	0.5				1
2	RandomMax+Mutation		1					1	1				1
3	OneMax		1					1					1
4	RandomMax		1					1					1
5	RR1	0.33					1		0.67		0.67		1
6	Rand-RR1	0.67					0.67		1				1
7	Dec-3					1	0.25		1				0.75
8	Exp-Dec3						0.67	1	0.33				1
9	Trap-5					1	0.25		1				1
10	f6-bipolar					1				0.33	1		0.75
11	HIFF				1		0.5		1				0.5
12	Htrap-1			1	1				1				
13	OverlapTrap								1		1		1
14	F1						1		1				1
15	F2						1		1				1
16	F3	1					1		1				1
17	F4	1					1		1				1
18	Vary Size Trap						1		1				1
19	Mixed-Trap						1		1				1
Total		3	4	1	2	3	9.33	5	11.5	0.33	2.67	0	17
No. of problem		19	4	5	8	3	19	19	19	19	19	4	
Performance (%)		15.79	100	20	25	100	49.11	26.32	76.32	1.74	11.05	0	89.47

หมายเหตุ ค่า Performance (%) ที่คำนวณได้จะอยู่ในช่วง 0-100 และตัวหนาหมายถึงค่ามากที่สุด 3 อันดับแรกจากขั้นตอนวิธีที่นำเสนอในงานวิจัย

ตารางที่ 11.10 ผลการเปรียบเทียบความสามารถของแต่ละขั้นตอนวิธีกับปัญหาทดสอบ (นับรวมจากปัญหาทดสอบย่อยทั้งหมด)

Sub-problems	sGA	UMDA	CSM	BOA	QHFC	sGA-FC	RFT	CFT	LZWGA-1	LZWGA-2	LZWGA-3	LZWEDA
rank 1	0	16	0	0	12	2	0	16	0	0	0	9
rank 2	2	0	1	5	0	7	12	11	0	3	0	19
rank 3	3	0	2	2	0	10	7	10	1	6	0	20
Sum rank 1-3	5	16	3	7	12	19	19	37	1	9	0	48
Fail sub-problem	10	0	0	0	0	8	16	4	12	5	0	1
No. of sub-problem	55	16	19	28	12	55	55	55	55	55	16	55

หมายเหตุ ตัวหนาหมายถึงมีค่ามากที่สุดสามอันดับแรก



รูปที่ 11.1 กราฟเปรียบเทียบสมรรถภาพในการแก้ปัญหาทดสอบของแต่ละขั้นตอนวิธี (ค่าที่อยู่ในเครื่องหมายก้ามปูหมายถึงจำนวนปัญหาหลักที่ใช้ทดสอบ)

11.7 สรุปผลการทดลอง

จากผลการทดสอบกับปัญหาทดสอบทั้งหมดทำให้สามารถสรุปได้ว่าวิธีการระบุและการประกอบหน่วยการสร้างแต่ละรูปแบบมีความเหมาะสมกับปัญหาที่ต่างกัน แม้ว่าบางวิธีที่นำเสนอให้ผลแยกว่าวิธีอื่นในเกือบทุกปัญหาทดสอบก็ตามแต่ก็ไม่สามารถบอกได้ว่าจะไม่มีปัญหาที่เหมาะสมสำหรับขั้นตอนวิธีที่แยเหล่านี้เลย (อาจไม่อยู่ในชุดปัญหาทดสอบนี้) และจากการทดสอบสามารถบอกได้ในภาพรวมว่าขั้นตอนวิธีใดเหมาะสมกับปัญหากลุ่มใด โดยใช้มาตรวัดจากค่าที่คำนวณได้จากตารางที่ 11.9 เป็นเกณฑ์ตัดสิน โดยคำนวณจากค่าผลรวมของสมรรถนะในแต่ละปัญหาย่อยของแต่ละกลุ่มปัญหา และจะถือว่าขั้นตอนวิธี (ที่นำเสนอ) ที่มีค่าผลรวมสมรรถนะสูงสุดสามอันดับแรก เป็นขั้นตอนวิธีที่มีประสิทธิภาพดี (มีความเหมาะสม) ในการแก้ปัญหาในกลุ่มดังกล่าว (ดังตารางที่ 11.11)

ตารางที่ 11.11 ขั้นตอนวิธีที่มีประสิทธิภาพในการแก้ปัญหาในกลุ่มปัญหาทดสอบ

No.	Problem Group	Algorithm
1	Basic	RFT, CFT, LZWEDA
2	BBs: non deceptive & additively decomposable	sGA-FC, CFT, LZWEDA
3	BBs: deceptive & additively decomposable	sGA-FC, CFT, LZWEDA
4	BBs: deceptive & non-linear fitness	sGA-FC, RFT, LZWEDA
5	BBs: deceptive & multimodal	CFT, LZWGA-2, LZWEDA
6	BBs: deceptive & hierarchical	sGA-FC, CFT, LZWEDA
7	BBs: deceptive & Overlapping	CFT, LZWGA-2, LZWEDA
8	BBs: deceptive & Vary Size	sGA-FC, CFT, LZWEDA
9	BBs: Mixed	sGA-FC, CFT, LZWEDA
10	Optimization	sGA-FC, CFT, LZWEDA

เนื้อหาในบทนี้มุ่งเน้นไปที่การทดสอบความสามารถในการแก้ปัญหาของแต่ละขั้นตอนวิธีที่นำเสนอกับกลุ่มปัญหาทดสอบ แต่ไม่มีการวิเคราะห์ผลการทดลองในเชิงลึกเกี่ยวกับมุมมองของหน่วยการสร้างและชิ้นส่วนเหมือนกับบทที่นำเสนอขั้นตอนวิธี ทำให้ไม่สามารถกล่าวสรุปอย่างละเอียดได้ว่าแต่ละขั้นตอนวิธีมีกลไกการทำงานที่เหมาะสมกับแต่ละกลุ่มปัญหาได้อย่างไร ดังนั้นในบทถัดไปจะแสดงผลการทดสอบของแต่ละขั้นตอนวิธีที่นำเสนอในมุมมองต่าง ๆ เปรียบเทียบกัน เพื่อให้เห็นถึงภาพรวมว่าแต่ละขั้นตอนวิธีที่นำเสนอมีความแตกต่างกันอย่างไร

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

12.1 สรุปผลการวิจัย

หน่วยการสร้า่งนั้นโดยปกติจะหมายถึงหน่วยการสร้า่งที่มีหน้าตาเหมือนคำตอบที่ดีที่สุด (หรือคำตอบที่ต้องการ) เท่านั้น แต่จากการศึกษาทฤษฎีบทเค้าร่างซึ่งเป็นที่มาของสมมติฐานหน่วยการสร้า่ง (บทที่ 3) คำว่าหน่วยการสร้า่งได้มาจากการพิจารณาเค้าร่างในรૂ่นประชากรเดียวเท่านั้น ซึ่งมีคุณสมบัติเพียงแค้เป็นเค้าร่างที่สั้น มีลำดับที่ต่ำ และมีโอกาสสออยู่รอดสูงเมื่ออยู่ในโครโมโซมที่มีค่าความเหมาะสมสูงกว่าค่าความเหมาะสมเฉลี่ย ดังนั้นจากสมมติฐานนี้สามารถกล่าว (ตีความ) ได้ว่าหน่วยการสร้า่งไม่จำเป็นต้องมีหน้าตาเหมือนกับคำตอบที่ดีที่สุด (คำตอบสุดท้ายที่ต้องการ) อย่างไรก็ตามเรายังคงมีความเชื่อว่าการสร้า่งเหล่านี้จะเป็นกฎแฉที่นำไปสู่คำตอบที่ต้องการได้ในที่สุด สำหรับการวัดการมีอยู่ของหน่วยการสร้า่งหรือปริมาณหน่วยการสร้า่งในประชากรจะเกิดปัญหาว่า ไม่รู้ว่าการสร้า่งในรૂ่นประชากรหนึ่ง ๆ จะมีหน้าตาเป็นอย่างไร โดยทั่วไปนิยมวัดโดยใช้รูปแบบหน่วยการสร้า่งของคำตอบที่ดีที่สุดเป็นเกณฑ์ (นับตามรูปแบบหน่วยการสร้า่งที่ดีที่สุดว่ามีอยู่เป็นจำนวนเท่าใด) ซึ่งเป็นวิธีวัดที่ทำให้เห็นภาพได้ง่ายวิธีหนึ่งแต่ไม่ใช่วิธีวัดที่ดีเนื่องจากหน่วยการสร้า่งที่วัดได้เป็นแค่ส่วนหนึ่งของหน่วยการสร้า่งที่มีอยู่ทั้งหมด เป็นเหตุให้ไม่สามารถทำความเข้าใจกับหน่วยการสร้า่งที่มีอยู่ในแต่ละรૂ่นประชากรได้ว่ามีอยู่ในรูปแบบความสัมพันธ์ลักษณะใด (ไม่สามารถบ่งบอกหน้าตา และปริมาณได้) ส่งผลให้การพิจารณากลไกการทำงานของการจัดการหน่วยการสร้า่ง เช่น การไขว้เปลี่ยน เป็นไปอย่างยากลำบาก (เห็นภาพได้ไม่ชัด) ผลทั้งหมดนี้เกิดจากนิยามหน่วยการสร้า่งที่อยู่ในรูปแบบเค้าร่างซึ่งมีความซับซ้อนสูงนั่นเอง

ในงานวิจัยนี้นำเสนอมุมมองของหน่วยการสร้า่งในรูปแบบใหม่ (บทที่ 4) มีจุดประสงค์เพื่อลดความซับซ้อนของนิยามหน่วยการสร้า่งแบบเดิม และช่วยให้เข้าใจกลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรมได้ดีขึ้น หน่วยการสร้า่งที่นิยามขึ้นใหม่นี้มีความสอดคล้องกับนิยามหน่วยการสร้า่งแบบเดิมทุกประการ หรืออาจมองได้ว่าหน่วยการสร้า่งที่นิยามขึ้นเป็นเซตย่อยของนิยามหน่วยการสร้า่งแบบเดิม เพื่อไม่ให้เกิดความสับสนจึงมีการตั้งชื่อหน่วยการสร้า่งที่นิยามขึ้นใหม่นี้ว่า “ชิ้นส่วน” ความแตกต่างระหว่างชิ้นส่วนกับหน่วยการสร้า่งคือ หน่วยการสร้า่งจะยอมให้มีคู่จิ้นที่รู้ค่าที่แน่นอน (Unknown Bit: *) แทรกอยู่ระหว่างคู่จิ้นที่รู้ค่าแน่นอน (Fixed Bit) แต่ชิ้นส่วนจะมีเฉพาะจิ้นที่รู้ค่าแน่นอน (Fixed Bit) เท่านั้น ตัวอย่างเช่น เค้าร่างของโครโมโซมยาว 10 บิต มีรูปแบบเป็น **101**011 เค้าร่างนี้สามารถพิจารณาในมุมมองของหน่วยการสร้า่ง

ได้เป็น 101**011 และเมื่อพิจารณาในมุมมองของชิ้นส่วนจะได้ชิ้นส่วนจำนวนสองชิ้นได้แก่ 101 และ 011 แยกขาดจากกัน (คนละชิ้นกัน) (ในตัวอย่างเป็นการพิจารณาที่ไม่สนใจเรื่องตำแหน่ง) เป็นต้น จากตัวอย่างจะเห็นได้ว่ารูปแบบของชิ้นส่วนมีความซับซ้อนน้อยกว่าหน่วยการสร้างแต่ยังคงเป็นส่วนหนึ่งของหน่วยการสร้าง ความเรียบง่ายของชิ้นส่วนมีประโยชน์หลายประการได้แก่ ช่วยให้สามารถวิเคราะห์และแสดงให้เห็นถึงการจัดการหน่วยการสร้างได้ชัดเจนมากขึ้น สามารถระบุรูปแบบขององค์ประกอบที่ดีที่มีลักษณะสอดคล้องกับหน่วยการสร้างแบบเดิมได้อย่างชัดเจนว่ามีหน้าตาเป็นอย่างไร ช่วยให้เห็นภาพในการผสมหรือการประกอบชิ้นส่วนเหล่านี้เข้าด้วยกันได้ง่ายขึ้น เป็นต้น

กลไกการทำงานของขั้นตอนวิธีเชิงพันธุกรรมสามารถอธิบายได้ด้วยทฤษฎีบทเคิร์วาง ซึ่งเป็นที่มาของสมมติฐานหน่วยการสร้าง แบ่งการดำเนินการออกได้เป็นสองกระบวนการย่อยคือการระบุหน่วยการสร้าง และการประกอบ (หน่วยการสร้างที่ระบุได้) เข้าด้วยกัน กระบวนการเหล่านี้เกิดขึ้นโดยอาศัยตัวดำเนินการต่าง ๆ เช่น ตัวดำเนินการเลือกสรร และตัวดำเนินการไขว้เปลี่ยน เป็นต้น โดยทั่วไปตัวดำเนินการส่วนใหญ่ไม่ได้ให้ความสนใจกับการระบุหน่วยการสร้างหรือกล่าวได้ว่าไม่มีการระบุหน่วยการสร้างที่ชัดเจน แต่มุ่งเน้นที่การประกอบหรือผสมเคิร์วางเข้าด้วยกันเป็นหลัก เนื่องจากตัวดำเนินการส่วนใหญ่ได้แรงบันดาลใจมาจากกระบวนการทำงานที่มีอยู่ในธรรมชาติ เช่น การไขว้เปลี่ยนในเชิงพันธุกรรม เป็นต้น งานวิจัยนี้จึงตั้งข้อสันนิษฐานว่าถ้าสามารถระบุหน่วยการสร้างได้ดีขึ้น (แม่นยำมากขึ้น) จะส่งผลให้สามารถหาคำตอบได้ดีขึ้นด้วย (มีประสิทธิภาพเพิ่มขึ้น) โดยมีเงื่อนไขว่าการประกอบเข้าด้วยกันจะต้องไม่ทำลายหรือตัดขวางหน่วยการสร้างที่ระบุได้ หรือพยายามรักษาหน่วยการสร้างที่ระบุได้ให้คงอยู่มากที่สุด งานวิจัยนี้มุ่งเน้นไปที่การศึกษาการระบุหน่วยการสร้างเป็นหลัก เนื่องจากการประกอบเข้าด้วยกันต้องสอดคล้องกับหน่วยการสร้างที่ระบุได้ ไม่สามารถออกแบบเป็นอิสระแยกจากกันได้ และต้องการแสดงให้เห็นถึงผลกระทบที่มาจากการระบุหน่วยการสร้างว่ามีนัยสำคัญไม่น้อยไปกว่าการประกอบหรือผสมเคิร์วางเข้าด้วยกันที่นักวิจัยส่วนใหญ่ให้ความสนใจตั้งแต่ยุคเริ่มแรกของสายงานวิจัยนี้ (ส่วนใหญ่ให้ความสนใจกับการไขว้เปลี่ยน หรือการผสมเคิร์วางเข้าด้วยกันเป็นหลัก) งานวิจัยนี้พิจารณาหน่วยการสร้างแบบชิ้นส่วน เพื่อให้สามารถระบุหน่วยการสร้างได้ง่ายและชัดเจน ยิ่งไปกว่านั้นยังได้นำเสนอวิธีระบุหน่วยการสร้างแบบชิ้นส่วนที่แตกต่างกันถึง 3 วิธี และวิธีประกอบหน่วยการสร้างเข้าด้วยกัน 2 วิธี เพื่อเป็นแนวทางให้เห็นถึงความเป็นไปได้ และผลที่ได้รับจากการระบุหน่วยการสร้าง

ขั้นตอนวิธีที่นำเสนอในงานวิจัยนี้ได้แก่

1. การไขว้เปลี่ยนแบบชิ้นส่วน (Fragment Crossover: FC)

2. ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม (Random Fragmentation: RFT)
3. ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข (Conditional Fragmentation: CFT)
4. ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW (LZWGA)
5. ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW (LZWEDA)

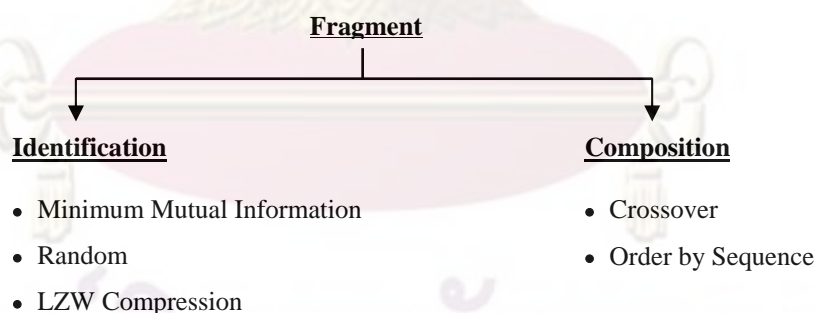
หมายเหตุ รายละเอียดของความแตกต่างระหว่างแต่ละขั้นตอนวิธีอยู่ในบทที่ 11

วิธีที่นำเสนอทั้งหมดสามารถแบ่งตามวิธีการระบุหน่วยการสร้างได้ดังนี้

1. วิธีการระบุหน่วยการสร้างที่ใช้หลักการสารสนเทศร่วมกันน้อยที่สุด (Minimum Mutual Information) ได้แก่วิธี FC และ CFT
2. วิธีการระบุหน่วยการสร้างแบบสุ่มตัด ได้แก่วิธี RFT
3. วิธีการระบุหน่วยการสร้างที่ใช้หลักการบีบอัดแบบ LZW ได้แก่วิธี LZWGA และ LZWEDA

วิธีที่นำเสนอทั้งหมดสามารถแบ่งตามวิธีการประกอบหน่วยการสร้างเข้าด้วยกันได้ดังนี้

1. วิธีประกอบเข้าด้วยกันโดยใช้การไขว้เปลี่ยน ได้แก่วิธี FC และ LZWGA
2. วิธีประกอบเข้าด้วยกันแบบเรียงต่อกันโดยใช้ตำแหน่งคู่จิ้นในโครโมโซมเป็นตัวอ้างอิง ได้แก่วิธี RFT, CFT และ LZWEDA



รูปที่ 12.1 หลักการที่ใช้ในการระบุและการประกอบชิ้นส่วน

การออกแบบวิธีการระบุชิ้นส่วนและการประกอบเข้าด้วยกันพิจารณาจากมุมมองของการแบ่งปันความรู้ (Shared Knowledge) ซึ่งได้จากการศึกษาในเชิงทฤษฎีที่มีหลายงานวิจัยในกลุ่มของขั้นตอนวิธีประมาณการแจกแจง (บทที่ 6) รวมถึงการไขว้เปลี่ยนแบบหลายพ่อพันธุ์แม่พันธุ์ ที่ได้แสดงให้เห็นถึงข้อดีหรือจุดเด่นเรื่องการแบ่งปันความรู้ที่มีการเรียนรู้แบบรวมศูนย์ (Centralized Learning) ที่ให้ผลดีกว่า (มีประสิทธิภาพในการหาคำตอบสูงกว่า) การเรียนรู้แบบกระจาย (Distributed Learning) ซึ่งมีการแบ่งปันข้อมูลกันอย่างจำกัดระหว่างพ่อพันธุ์กับแม่พันธุ์

เท่านั้น ขั้นตอนวิธีที่นำเสนอเมื่อแบ่งตามมุมมองการแบ่งปันความรู้ในส่วนของการระบุหน่วยการสร้างสามารถแบ่งได้ดังนี้

1. วิธีระบุหน่วยการสร้างที่มีการแบ่งปันความรู้แบบรวมศูนย์ได้แก่วิธี CFT⁹, LZWGA-3¹⁰ และ LZWEDA (ใช้พจนานุกรมร่วมกันระหว่างโครโมโซมที่ถูกเลือกทั้งหมด)
2. วิธีระบุหน่วยการสร้างที่มีการแบ่งปันความรู้แบบกระจายได้แก่วิธี FC, CFT¹ และ LZWGA-2² (ใช้พจนานุกรมร่วมกันระหว่างพ่อพันธุ์และแม่พันธุ์)
3. วิธีระบุหน่วยการสร้างแบบที่ไม่มีการแบ่งปันความรู้ระหว่างโครโมโซมเลยได้แก่วิธี RFT และ LZWGA-1² (ไม่มีการใช้พจนานุกรมร่วมกันเลยเพราะหนึ่งโครโมโซมจะบีบอัดได้หนึ่งพจนานุกรม)

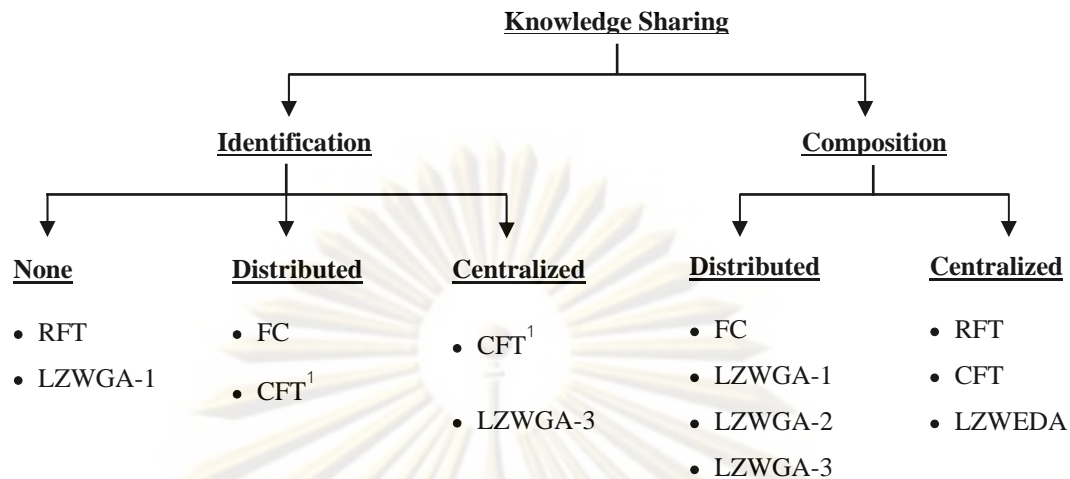
และเมื่อแบ่งตามมุมมองการแบ่งปันความรู้ในส่วนของการประกอบหน่วยการสร้างเข้าด้วยกันสามารถแบ่งได้ดังนี้

1. วิธีประกอบเข้าด้วยกันที่มีการแบ่งปันความรู้แบบรวมศูนย์ได้แก่วิธี RFT, CFT และ LZWEDA (มีการใช้หน่วยเก็บถาวรในการเก็บชิ้นส่วนที่ระบุได้ไว้ร่วมกัน)
2. วิธีประกอบเข้าด้วยกันที่มีการแบ่งปันความรู้แบบกระจายได้แก่วิธี FC และ LZWGA (ใช้วิธีการไขว้เปลี่ยนซึ่งมีการแลกเปลี่ยนชิ้นส่วนระหว่างคู่โครโมโซมพ่อพันธุ์และแม่พันธุ์เท่านั้น)

การทดสอบผลของขั้นตอนวิธีที่นำเสนอทั้งหมด (มี 5 ขั้นตอนวิธี) ใช้ปัญหาวัดเปรียบเทียบสมรรถนะ (Benchmark Problem) ซึ่งเป็นที่ยอมรับและนิยมใช้กันในสายงานวิจัยนี้ โดยให้ความสนใจกับกลุ่มปัญหาทดสอบหน่วยการสร้างเป็นพิเศษ เนื่องจากเป็นกลุ่มปัญหาที่มีรูปแบบของหน่วยการสร้างชัดเจน ตรงกับความต้องการในการศึกษาถึงผลของการระบุและการประกอบหน่วยการสร้าง ผลการทดลองที่ได้เป็นตัวสะท้อนให้เห็นถึงคุณลักษณะเฉพาะและประสิทธิภาพในการแก้ปัญหา (แต่ละกลุ่ม) ของแต่ละขั้นตอนวิธีที่นำเสนอได้เป็นอย่างดี งานวิจัยนี้แบ่งผลการทดลองของแต่ละขั้นตอนวิธีที่นำเสนอออกเป็นสองส่วนได้แก่ ส่วนแรกเป็นการแสดง

⁹วิธี CFT สามารถพิจารณาได้ว่ามีการแบ่งปันความรู้แบบกระจายเมื่อมองจากวิธีการระบุชิ้นส่วนที่ใช้หลักสารสนเทศร่วมกับโครโมโซมทีละคู่ (พิจารณาร่วมกันเพียงแค่ 2 โครโมโซมเท่านั้น) และสามารถพิจารณาได้ว่ามีการแบ่งปันความรู้แบบรวมศูนย์เมื่อมองจากการระบุชิ้นส่วนที่ดำเนินการกับโครโมโซมทุกคู่ที่เป็นไปได้ในกลุ่มของประชากรย่อย (กลุ่มโครโมโซม) ที่ถูกเลือก (เป็นการเปรียบเทียบกันระหว่างโครโมโซมทุกตัวที่ถูกเลือกทีละคู่)

¹⁰เฉพาะ LZWGA มีการทดสอบการแบ่งปันความรู้ในหลายระดับจึงแบ่งออกเป็น LZWGA รูปแบบที่ 1, 2 และ 3 (ไม่มีการแบ่งปันความรู้เลย มีการแบ่งปันความรู้ระหว่างคู่โครโมโซม และมีการแบ่งปันความรู้ระหว่างโครโมโซมที่ถูกเลือกทั้งหมด ตามลำดับ)



รูปที่ 12.2 การแบ่งปันความรู้ในแต่ละขั้นตอนวิธีที่น่าเสนอ

ถึงมุมมองในการจัดการหน่วยการสร้างและชิ้นส่วนของแต่ละขั้นตอนวิธีที่น่าเสนอ โดยทดสอบกับปัญหา Royal Road และ Trap-5 (รายละเอียดเกี่ยวกับการทดลองและการตีความผลการทดลองอยู่ในบทที่ 4) และแสดงผลในรูปแบบกราฟความสัมพันธ์ระหว่างหน่วยการสร้าง (ชนิด ขนาด และปริมาณ) กับรุ่นประชากร และกราฟความสัมพันธ์ระหว่างชิ้นส่วน (ชนิด ขนาด และปริมาณ) กับรุ่นประชากร เป็นหลัก รวมถึงมีการแสดงให้เห็นถึงพฤติกรรมในการจัดการหน่วยการสร้างในระดับโครงสร้างโครโมโซม โดยแสดงผลในรูปแบบภาพเพื่อให้เห็นความสัมพันธ์ระหว่างโครโมโซม (รูปแบบคู่จิ้นในแต่ละโครโมโซม) กับรุ่นประชากร ในส่วนที่สองแสดงถึงประสิทธิภาพในการแก้ปัญหาหรือความสามารถในการหาคำตอบโดยวัดจากจำนวนครั้งการเรียกใช้ฟังก์ชันประเมินค่าความเหมาะสม และทรัพยากรที่ใช้ในการแก้ปัญหาโดยวัดจากจำนวนของประชากร สำหรับส่วนที่สองนี้มีการทดสอบกับปัญหาวัดเปรียบเทียบสมรรถนะทั้งหมด และเปรียบเทียบผลกับขั้นตอนวิธีอื่นในสายงานวิจัยนี้ รายละเอียดเกี่ยวกับการทดลองและผลการทดลองอยู่ในบทที่ 11

การพิจารณาผลการทดลองในมุมมองของชิ้นส่วนร่วมกับหน่วยการสร้าง (ของคำตอบที่ดีที่สุด) จะทำให้เห็นภาพกลไกการทำงานของแต่ละขั้นตอนวิธีได้ดีขึ้น สามารถอธิบายหลักการทำงานของขั้นตอนวิธีที่น่าเสนอในแง่ของการจัดการหน่วยการสร้างได้อย่างชัดเจน และจากการเปรียบเทียบผลระหว่างขั้นตอนวิธีที่น่าเสนอซึ่งมีหลักในการระบุและการประกอบชิ้นส่วน

¹ วิธี CFT สามารถพิจารณาได้ว่ามีการแบ่งปันความรู้แบบกระจายเมื่อมองจากวิธีการระบุชิ้นส่วนที่ใช้หลักสารสนเทศร่วมกับโครโมโซมทีละคู่ (พิจารณาร่วมกันเพียงแค่ 2 โครโมโซมเท่านั้น) และสามารถพิจารณาได้ว่ามีการแบ่งปันความรู้แบบรวมศูนย์เมื่อมองจากการระบุชิ้นส่วนที่ดำเนินการกับโครโมโซมทุกคู่ที่เป็นไปได้ในกลุ่มของประชากรย่อย (กลุ่มโครโมโซม) ที่ถูกเลือก (เป็นการเปรียบเทียบกันระหว่างโครโมโซมทุกตัวที่ถูกเลือกทีละคู่)

ต่างกัน จะยิ่งทำให้เข้าใจถึงผลกระทบจากปัจจัยต่าง ๆ (ข้อจำกัดในการระบุและการประกอบ) ที่เกิดขึ้นกับหน่วยการสร้างได้ดียิ่งขึ้น ความรู้ที่ได้เหล่านี้สามารถนำไปใช้ประโยชน์ในการออกแบบขั้นตอนวิธีการแก้ปัญหาให้มีประสิทธิภาพสูงขึ้นได้

12.2 อภิปรายผลการวิจัย

ในส่วนนี้จะพูดถึงภาพรวมกลไกการทำงานของแต่ละขั้นตอนวิธี เพื่อให้เห็นภาพการจัดการหน่วยการสร้างในมุมมองของปัจจัยต่าง ๆ โดยอ้างอิงกับผลการทดลองในบทที่ 11 และพิจารณาในมุมมองของการสำรวจ (Exploration) การรักษาให้คงอยู่ (Retention) และการแลกเปลี่ยน (Exchange) ของชิ้นส่วน

เมื่อพิจารณาจากปัจจัยเรื่องระดับของการแบ่งปันความรู้จากรูปที่ 12.2 สามารถแบ่งออกได้เป็น 2 ระดับได้แก่ ระดับการระบุชิ้น และระดับการประกอบชิ้นส่วน จากผลการทดลองแสดงให้เห็นว่าการแบ่งปันความรู้ในระดับการระบุชิ้นส่วนมีความสำคัญเทียบเท่ากับการแบ่งปันความรู้ในระดับการประกอบชิ้นส่วน และจะให้ได้ผลที่ดีก็ต่อเมื่อวิธีการแบ่งปันความรู้ของทั้งสองระดับสามารถเข้ากันได้ดี (สอดคล้องกัน) จากผลของแต่ละขั้นตอนวิธีที่น่าเสนอจะเห็นว่าวิธีที่มีการแบ่งปันความรู้แบบรวมศูนย์ในสองระดับซึ่งได้แก่วิธี CFT และ LZWEDA ให้ผลดีที่สุด (จากการจัดอันดับความสามารถในการแก้ปัญหาในบทที่ 11 ได้อันดับที่ 2 และ 1 ตามลำดับ) เป็นผลจากการสำรวจ การรักษาให้คงอยู่ และการแลกเปลี่ยนที่เพิ่มขึ้นอย่างสมดุล เมื่อเปรียบเทียบกับวิธีแบบกระจายความรู้ และเมื่อพิจารณาผลจากวิธี RFT จะเห็นได้ชัดว่าแม้ขั้นตอนการระบุหน่วยการสร้างจะเกิดขึ้นแบบสุ่ม (ไม่มีการแบ่งปันความรู้) แต่ก็ยังสามารถหาคำตอบได้ภายใต้เงื่อนไขที่กำหนด (ในหลายปัญหา) เนื่องจากการแบ่งปันความรู้แบบรวมศูนย์ในส่วนของประกอบเข้าด้วยกัน

sGA-FC ได้ผลดีเป็นอันดับที่ 3 (จากการทดสอบในบทที่ 11) รองจาก CFT และ LZWEDA แม้ว่าจะไม่มีการแบ่งปันความรู้แบบรวมศูนย์อยู่เลยทั้งในการระบุและการประกอบชิ้นส่วน ผลที่ได้นี้เป็นตัวบ่งชี้ให้เห็นถึงความสำคัญของวิธีระบุหน่วยการสร้างว่า ถ้าสามารถระบุหน่วยการสร้างได้ดีจะส่งผลให้ประสิทธิภาพในการคำตอบดีขึ้นด้วย ภายใต้เงื่อนไขที่การประกอบเข้าด้วยกันไม่ขัดขวางหรือทำลายหน่วยการสร้างที่ระบุได้ เมื่อพิจารณาการสำรวจเราสามารถมองได้ว่าหลักการสารสนเทศร่วมกันน้อยสุดเป็นวิธีการสำรวจรูปแบบหนึ่ง (ไม่ได้เป็นตัวบ่งชี้ระดับการสำรวจที่เกิดขึ้นแต่อย่างใด) และมองได้ว่าการไขว่เปลี่ยนแบบชิ้นส่วนมีการระบุหน่วยการสร้างที่ดีและหน่วยการสร้างที่ระบุได้ไม่ถูกทำลาย เป็นการเพิ่มระดับการรักษาให้คงอยู่สูงขึ้น ในขณะที่เดียวกันการไขว่เปลี่ยนยังช่วยเพิ่มความสามารถในการแลกเปลี่ยนให้สูงขึ้นด้วย

ขั้นตอนวิธีที่ใช้การบีบอัดแบบ LZW เป็นตัวระบุชิ้นส่วนร่วมกับการใช้การไขว้เปลี่ยนแบบจุดเดียวเป็นตัวประกอบชิ้นส่วนเข้าด้วยกัน เพื่อศึกษาระดับของการแบ่งปันข้อมูลในการระบุชิ้นส่วนซึ่งแบ่งออกเป็น 3 ระดับดังรูปที่ 12.2 จากผลการทดลองแสดงให้เห็นว่าการบีบอัดที่ไม่มี การแบ่งปันความรู้ระหว่างโครโมโซมในวิธี LZWGA-1 ให้ผลที่ไม่ดีเพราะการระบุชิ้นส่วนไม่มี ประสิทธิภาพ (ไม่มีความรู้เพียงพอในการระบุชิ้นส่วน) และทำให้เกิดข้อจำกัดในการไขว้เปลี่ยน ส่งผลให้ไม่สามารถไขว้เปลี่ยนได้อย่างอิสระ (เกิดการไขว้เปลี่ยนได้เฉพาะรอยต่อของชิ้นส่วน เท่านั้น) ซึ่งเป็นการขัดขวางกระบวนการสำรวจ (ทำให้ระดับการสำรวจต่ำลง) มีผลให้ ประสิทธิภาพในการทำงานลดลงด้วย ส่วนการบีบอัดแบบที่มีการแบ่งปันความรู้แบบรวมศูนย์ใน วิธี LZWGA-3 ก็ประสบกับปัญหาในทำนองเดียวกันกับวิธี LZWGA-1 เนื่องจากระดับการแบ่งปัน ความรู้แบบ LZW ที่สูงเกินไปทำให้ได้ชิ้นส่วนที่ยาวมากขึ้นจนกลายเป็นการขัดขวางการสำรวจ เนื่องจากเมื่อชิ้นส่วนยาวขึ้นจะทำให้จำนวนชิ้นส่วนที่มีอยู่ในแต่ละโครโมโซมลดลง ส่งผลให้ ตำแหน่งที่สามารถทำการไขว้เปลี่ยนได้ยังมีน้อยลง ด้วยเหตุนี้จึงทำให้ได้ผลการทดลองที่ไม่ดี แต่ สำหรับการบีบอัดที่มีการแบ่งปันความรู้กันระหว่างโครโมโซมพ่อพันธุ์และแม่พันธุ์ (2 โครโมโซม) ในวิธี LZWGA-2 ให้ผลดีกว่า LZWGA-1, LZWGA-3 และ sGA เป็นเพราะการแบ่งปันความรู้ที่ได้ จากการบีบอัดเกิดขึ้นในระดับที่เหมาะสม ทำให้ได้ชิ้นส่วนที่มีขนาดไม่ใหญ่จนเกินไป และ พิจารณาได้ว่าเป็นการสำรวจในรูปแบบหนึ่ง ส่วนการบีบอัดทำให้เกิดการรักษาชิ้นส่วนให้คงอยู่ (ไม่ให้แยกจากกัน) ในระหว่างที่เกิดการไขว้เปลี่ยน

จากที่กล่าวมาทั้งหมดสรุปได้ว่าการแบ่งปันความรู้ในกระบวนการทำงานไม่ว่าจะเป็นการระบุชิ้นส่วนหรือการประกอบชิ้นส่วนเข้าด้วยกันเป็นสิ่งสำคัญที่ส่งผลโดยตรงต่อประสิทธิภาพ ในการหาคำตอบ รวมถึงความสอดคล้องกันระหว่างกระบวนการระบุและการประกอบชิ้นส่วนก็มี ผลกับประสิทธิภาพในการหาคำตอบเช่นเดียวกัน ถ้ามองกระบวนการเหล่านี้เป็นการสำรวจและ การแสวงหาประโยชน์จะเห็นว่าหัวใจสำคัญคือ ความสมดุลในการรักษาระดับความหลากหลาย ไม่ให้มันน้อยเกินไปและมากเกินไประหว่างการดำเนินการ สิ่งเหล่านี้มองเห็นได้จากมุมมองของ หน่วยการสร้างรูปแบบเดิม ดังนั้นงานวิจัยนี้จึงได้นำเสนอมุมมองหน่วยการสร้างแบบชิ้นส่วนซึ่ง เป็นมุมมองหนึ่งที่ช่วยให้เข้าใจหน่วยการสร้างในภาพรวมได้ดีขึ้น เพราะมีรูปแบบที่ง่าย ตรงไปตรงมา สามารถนำไปประยุกต์ใช้ได้หลายรูปแบบ และทำความเข้าใจได้ง่าย งานวิจัยนี้เป็น เพียงแค่จุดเริ่มต้นเท่านั้น ผู้ที่สนใจแนวคิดที่นำเสนอในงานวิจัยนี้สามารถนำไปศึกษาวิจัยต่อยอด ได้ในอีกหลายประเด็นซึ่งจะได้กล่าวถึงในส่วนถัดไป

12.3 ข้อเสนอแนะ

จุดด้อยที่เด่นชัดของการนิยามหน่วยการสร้างแบบขึ้นส่วนอยู่ที่โครงสร้างซึ่งมีรูปแบบเป็นคู่จิ้นที่อยู่ติดกัน และแต่ละขึ้นส่วนไม่ขึ้นแก่กัน (เป็นอิสระจากกัน) ด้วยเหตุนี้ทำให้ไม่สามารถแก้ปัญหาที่หน่วยการสร้างมีความสัมพันธ์กันแบบหลวม (Loosely Linkage) ได้ แนวทางการแก้ไขวิธีหนึ่งคือการพิจารณาให้ขึ้นส่วนมีความสัมพันธ์กัน (ขึ้นแก่กัน) เช่น มีความสัมพันธ์กันแบบลำดับชั้นหรือโครงสร้างต้นไม้ เป็นต้น หรือพิจารณาขึ้นส่วนให้กลับไปอยู่ในรูปแบบเค้าร่างดั้งเดิมและถ้ามองในรูปแบบของการเชื่อมโยง (แบบหลวมและแบบแนบแน่น) ในระดับต่ำสุดจะเป็นการเชื่อมโยงกันระหว่างสองคู่จิ้น ดังนั้นจะเห็นได้ว่าหลักการพิจารณาแบบขึ้นส่วน เมื่อพิจารณาในรูปแบบของการเชื่อมโยงจะสามารถเกิดขึ้นได้ทุกขนาดและรูปแบบของคู่จิ้นที่เป็นไปได้ (ขึ้นส่วนมีขนาดเล็กที่สุดที่เป็นไปได้อาจเป็นหนึ่งคู่จิ้น จึงทำให้ไม่มีข้อจำกัดในส่วนนี้) ด้วยเหตุนี้จึงทำให้เกิดข้อสันนิษฐานว่าแนวทางที่นำเสนอสามารถนำไปประยุกต์ใช้แก้ปัญหาที่มีการเชื่อมโยงกันแบบหลวมได้

ทุกปัญหาทดสอบในงานวิจัยนี้ใช้การเข้ารหัสปัญหาแบบเลขฐานสองเท่านั้น เพื่อให้เห็นภาพการทำงานในระดับล่างสุดได้ชัดเจน จึงอาจทำให้เกิดข้อสงสัยว่าถ้านำไปประยุกต์ใช้กับปัญหาที่มีการเข้ารหัสแบบจำนวนเต็ม หรือจำนวนจริง จะสามารถทำได้หรือไม่ และจะให้ผลเป็นเช่นใด เมื่อพิจารณาจากหลักการขึ้นส่วนแล้วจะเห็นว่าไม่ขึ้นอยู่กับการเข้ารหัสแต่อย่างใด ข้อจำกัดอาจเกิดขึ้นได้กับกระบวนการระบุหรือประกอบขึ้นส่วนเข้าด้วยกัน แต่ในความเห็นส่วนตัวของผู้วิจัยมีความเชื่อว่าจะต้องมีรูปแบบการดำเนินการที่เหมาะสมและสามารถใช้งานได้จริง (กับทุกปัญหา) อย่างไรก็ตามข้อสงสัยเหล่านี้ยังคงรอการพิสูจน์ให้เห็นต่อไป

ในงานวิจัยนี้ไม่มีการทดสอบกับกลุ่มปัญหาการเรียงสับเปลี่ยน (Permutation) เช่น ปัญหาการเดินทางของนักขาย (TSP) เป็นต้น ปัญหาในกลุ่มนี้สามารถออกแบบการเข้ารหัสได้หลายวิธี แต่ผลที่ต่างกันจากหลักการขึ้นส่วนที่นำเสนอผู้วิจัยเชื่อว่ามีความเป็นไปได้สูงที่จะแก้ปัญหากลุ่มนี้ได้อย่างมีประสิทธิภาพ (ขึ้นกับรูปแบบการเข้ารหัสปัญหาด้วย) ปัญหาในกลุ่มนี้เป็นปัญหาที่พบได้ในโลกจริงจึงทำให้น่าสนใจที่จะทำการศึกษาวิจัย

ปัจจุบันมีการให้ความสนใจกับกลุ่มปัญหาที่มีหลายจุดประสงค์ (Multiple Objective) ในงานวิจัยนี้มีการทดสอบกับปัญหาหลายจุดประสงค์จำนวนสองปัญหาได้แก่ ปัญหา 6-bipolar และปัญหา HIFF การทดลองมุ่งเน้นไปที่การหารูปแบบคำตอบให้พบเพียงรูปแบบเดียว ผลการทดลองแสดงให้เห็นถึงความสามารถในการแก้ปัญหาในรูปแบบนี้ได้ดีในระดับหนึ่ง (พบ

คำตอบหนึ่งจุดประสงค์) จากผลที่ได้ทำให้ผู้วิจัยมีความเชื่อว่าหลักการที่นำเสนอสามารถนำไปประยุกต์ใช้เพื่อแก้ปัญหาที่มีหลายจุดประสงค์ได้อย่างมีประสิทธิภาพ

เนื่องจกงานวิจัยนี้นำเสนอหลักการในมุมมองพื้นฐานที่สุด (รูปแบบหนึ่ง) ของหน่วยการสร้าง ซึ่งเป็นหลักการพื้นฐานที่ใช้อธิบายกลไกการทำงานของขั้นตอนวิธีในสายงานวิจัยนี้ ดังนั้นถ้าเรามีความเชื่อว่าหน่วยการสร้างมีอยู่จริง หลักการที่นำเสนอในงานวิจัยนี้จะมีประโยชน์ และสามารถนำไปประยุกต์ใช้จริงได้



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

นริศ กุณาศล, วรเศรษฐ สุวรรณิก และ ประภาส จงสถิตย์วัฒนา, การเข้ารหัสแบบ LZW ในขั้นตอนวิธีเชิงพันธุกรรม. Proceedings of Electrical Engineering Conference (EECON) 2005 : 861-864.

ภาษาอังกฤษ

Altenberg, L. Fitness Distance Correlation Analysis: An Instructive Counterexample. Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA) 1997 : 57-64.

Antonisse, J. A new interpretation of schema notation that overturns the binary encoding constraint. Proceedings of the Third International Conference on Genetic Algorithms 1989 : 86-91.

Aporntewan, C. Building-Block Identification by Simultaneity Matrix. Ph.D. Dissertation Chulalongkorn University, 2004.

Aporntewan, C. and Chongstitvatana, P. A quantitative approach for validating the building-block hypothesis. Congress on Evolutionary Computation 2005: 1403-1409.

Bäck, T. Applications of Evolutionary Algorithms Technical Report of the University of Dortmund Department of Computer Science Systems Analysis Research Group 1992.

Baluja, S. Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning. Technical Report CMU-CS-95-163 Carnegie Mellon University 1994.

Baluja, S. and Davies, S. Combining multiple optimization runs with optimal dependency trees. Technical Report CMU-CS-97-157 Carnegie Mellon University 1997.

Baluja, S. and Davies, S. Using optimal dependency-trees for combinatorial optimization: Learning the structure of the search space. Technical Report CMU-CS-97-107 Carnegie Mellon University 1997.

- Battle, D. L. and Vose, M. D. Isomorphisms of genetic algorithms. Rowling J. E., Foundations of Genetic Algorithms, 242-251. San Mateo, CA, USA : Morgan Kaufmann, 1991.
- Buntine, W. L. Theory refinement of Bayesian networks. Proceedings of the Uncertainty in Artificial Intelligence (UAI) 1991 : 52-60.
- Carlos, A. A study on synergy of multiple crossover operators in a hierarchical genetic algorithm applied to structural optimization. Structural and Multidisciplinary Optimization 2009 : 117-135.
- Caruana, R. A., and Schaffer, J. D. Representation and hidden bias: Gray vs. binary coding for genetic. Proceedings of the fifth International Conference on Machine Learning 1988 : 153-161.
- Caruana, R. A. and Schaffer, J. D. Using Multiple Representations to Control Inductive Bias: Gray and Binary Codes for Genetic Algorithms. Proceedings of the Sixth International Workshop on Machine Learning 1989 : 375-378.
- Chen, Y. P., Yu, T. L., Sastry, K. and Goldberg, D. E. A Survey of Linkage Learning Techniques in Genetic and Evolutionary Algorithms. IlliGAL Report No. 2007014. University of Illinois at Urbana-Champaign 2007.
- Christensen, S. and Oppacher, F. What can we learn from No Free Lunch? A First Attempt to Characterize the Concept of a Searchable Function. Genetic and Evolutionary Computation Conference (GECCO) 2001 : 1219-1234.
- Davis, L. and Kaufman, M. Genetic Algorithms and Simulated Annealing. Southport (UK) : Pitman Publisher, 1996.
- Day, R. O. Explicit Building Block Multiobjective Evolutionary Computation:Methods and Applications. Ph.D. Thesis Air Force Institute of Technology Department of the Air Force Air University Wright-Patterson, 2005.
- Day, R. O. and Lamont, G. B. An effective explicit building block MOEA, the MOMGA-IIa. Congress on Evolutionary Computation 2005 : 17-24.
- Deb, K., Horn, J. and Goldberg, D. E. Multimodal deceptive functions. Complex System 1993 : 131-153.

- De Bonet, J. S., Isbell, C. L. and Viola, P. MIMIC: Finding optima by estimating probability densities. Advance in Neural Information Processing Systems 1997 : 424-430.
- DeJong, K. A. and Spears, W. M. A Formal Analysis of the Role of Multi-Point Crossover Annals of Mathematics and Artificial Intelligence 1992 : 1-26.
- DeJong, K. A. Genetic algorithms are NOT function optimizers. Whitley, L. D., Foundations of Genetic Algorithm, 1993.
- Doerr, B., Neumann, F., Sudholt, D. and Witt, C. On the runtime analysis of the 1-ANT ACO algorithm. Proceedings of the 9th annual conference on Genetic and evolutionary computation 2007.
- Droste, S., Jansen, T. and Wegener, I. Perhaps not a free lunch but at least a free appetizer. Genetic and Evolutionary Computation Conference (GECCO) 1999 : 833-839.
- Eiben, A. E. Multi-parent Recombination. Handbook of Evolutinoary Computation : Oxford University Press, 1997.
- Eshelman, L. J., Caruana, R. A. and Schaffer, J. D. Biases in the Crossover Landscape. Proceedings of the 3rd International Conference on Genetic Algorithms 1989 : 10-19.
- Eshelman, L. J., and Schaffer, J. D. Crossover's Niche. Proceedings of the 5th International Conference on Genetic Algorithms (ICGA) 1993 : 9-14.
- Etxeberria, R. and Larrañaga, P. Global optimization with Bayesian networks. Symposium on Artificial Intelligence (CIMAF) 1999 : 332-339.
- Fogarty, T. C. Varying the Probability of Mutation in the Genetic Algorithm. International Conference on Genetic Algorithms (ICGA) 1989 : 104-109.
- Forrest, S. and Mitchell, M. Relative building-block fitness and the building-block hypothesis. Foundations of Genetic Algorithms (FOGA) 1993.
- Goldberg, D. E. Simple genetic algorithms and the minimal deceptive problem. Davis, L. D., Genetic Algorithms and Simulated Annealing, 74-88. London : Pitman Publisher, 1987.

- Goldberg, D. E., and Richardson, J. Genetic algorithms with sharing for multimodal function. Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application 1987 : 41-49.
- Goldberg, D. E. Genetic algorithms and Walsh functions: Part I, A gentle introduction. Journal of Complex Systems 1989 : 129–152.
- Goldberg, D. E. Genetic algorithms and Walsh functions: Part II, Deception and its analysis. Journal of Complex Systems 1989 : 153–171.
- Goldberg, D. E. Genetic Algorithms in Search Optimization and Machine Learning. : Addison Wesley, 1989.
- Goldberg, D. E., Deb, K. and Clark, J. H. Genetic Algorithms, Noise, and the Sizing of Populations. Applications of Artificial Intelligence in Engineering VI 1991 : 3-16.
- Goldberg, D. E., Deb, K. and Clark, J. H. Genetic algorithms, noise, and the sizing of populations. Journal of Complex Systems 1992 : 333-362.
- Goldberg, D. E. The design of innovation: Lessons from and for competent genetic algorithms. Boston, MA : Kluwer Academic Publishers, 2002.
- Grefenstette, J. J., Deception considered harmful. In L. D. Whitley, ed., Foundations of Genetic Algorithms 2. Morgan Kaufmann, 1993.
- Harik, G. R. Learning linkage. Foundation of Genetic Algorithms (FOGA), 247–262, 1997.
- Harik, G. R., Lobo, F. G. and Goldberg, D. E. The compact genetic algorithm. In Proceeding of IEEE Conf. on Evolutionary Computation 1998 : 523-528.
- Harik, G. R. Linkage Learning via Probabilistic Modeling in the Extended Compact Genetic Algorithm (ECGA). IlliGAL Technical Report no. 99010, 1999.
- Hartl, D. L., and Jones, E. W. Genetics: principles and analysis. Sudbury, MA : Jones and Bartlett Publisher, 1998.
- Heckman, D., Geiger, D. and Chickering, D. M. Learning Bayesian Networks: the combination of knowledge and statistical data. Machine Learning, 197-243, 1995.
- Hesser, J. and Männer, R. Towards an Optimal Mutation Probability for Genetic Algorithms. Proceedings of the 1st Workshop on Parallel Problem Solving from Nature 1991 : 23-32.

- Holland, J. H. Adaptive plans optimal in payoff-only environments. In Proceedings of the Second Hawaii International Conference on Systems 1969 : 917-920.
- Holland, J. H. Non-linear environments permitting efficient adaptation. In Computer and Information Sciences 1967 : 147-64.
- Holland, J. H. Processing and processors for schemata. In Associative Information Techniques 1971 : 127-46.
- Holland, J. H. Genetic algorithms and the optimal allocation of trials. SIAM Journal of Computing 1973 : 88-105.
- Holland, J. H. Adaptation in Natural and Artificial Systems. Michigan : University of Michigan Press, 1975
- Holland, J. H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to biology, control and artificial intelligence. MIT Press, 1998
- Hordijk, W. and Manderick, B. The Usefulness of Recombination. European Conference on Artificial Life (ECAL) 1995 : 908-919.
- Inza, I., Larranaga, P., Etxeberria, R. and Sierra, B. Feature subset selection by Bayesian network-based optimization. Journal of Artificial Intelligence 2000 : 157-184.
- Janikow, C. Z. and Michalewicz, Z. An experimental comparison of binary and floating point. International Conference on Genetic Algorithms 1991 : 31-36.
- Jessica, A. C., Ignacio, P. and nélida B. MCPC and MCMP Evolutionary Algorithms for the TSP. Argentine Symposium on Artificial Intelligence (ASAI) 2003.
- Jones, T. and Forrest, S. Fitness Distance Correlation as a Measure of Problem Difficulty for Genetic Algorithms. Proceedings of the Sixth International Conference on Genetic Algorithms 1995 : 184-192.
- Julstrom, B. A. What Have You Done for Me Lately? Adapting Operator Probabilities in a Steady-State Genetic Algorithm. Proceedings of the 6th International Conference on Genetic Algorithms (ICGA) 1995 : 81-87
- Khan, N. Bayesian optimization algorithms for multiobjective and hierarchically difficult problems. Master thesis University of Illinois at Urbana-Champaign, 2003.
- Koza, J. R. Genetic Programming II: Automatic Discovery of Reusable Programs. Cambridge, MA : MIT Press, 1994

- Larrañaga, P. and Lozano, J. A. Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Boston : Kluwer Academic Publishers, 2001.
- Liepins, G. E., and Vose, M. D. Representational issues in genetic optimization. Journal of Experimental and Theoretical Artificial Intelligence 1990 : 101-115.
- Liepins, G. E., and Vose, M. D. Deceptiveness and genetic algorithm dynamics. Foundations of Genetic Algorithms (FOGA) 1991 : 221-241.
- Mitchell, M., Forrest, S. and Holland, J. H. The royal road for genetic algorithms: Fitness landscapes and GA performance. Proceedings. of first European Conference on Artificial Life 1992 : 245-254.
- Mitchell, M., Holland, J. H., and Forrest, S. When will a genetic algorithm outperform hill climbing?. Advances in Neural Information Processing Systems 1994 : 51-58.
- Mitchell, M., Hraber, P. T. and Crutchfield, J. P. Revisiting the edge of chaos: Evolving cellular automata to perform computations. Complex Systems 1996 : 89-130.
- Mühlenbein, H. The equation for response to selection and its use for prediction, IEEE International Conference on Evolutionary Computation 1998 : 303-346.
- Mühlenbein, H. and Mahnig, T. FDA - A scalable evolutionary algorithm for the optimization of additively decomposed functions. IEEE International Conference on Evolutionary Computation 1999 : 353-376.
- Mühlenbein, H. and Mahnig, T. Convergence theory and applications of the factorized distribution algorithm. Journal of Computing and Information Technology 1999 : 1999.
- Mühlenbein, H. and Mahnig, T. The Factorized Distribution Algorithm for the optimization of additively decomposed functions. IEEE International Conference on Evolutionary Computation 1999 : 353-376.
- O'Reilly, U. M., and Oppacher, F. The troubling aspects of a building block hypothesis for genetic programming. Foundations of Genetic Programming 1995 : 73-88.
- Pelikan, M., Goldberg, D. E. and Cant'u-Paz, E. Linkage problem, distribution estimation, and Bayesian networks. IlligAL Report No. 98013 University of Illinois at Urbana-Champaign 1998.
- Pelikan, M. and Mühlenbein, H. The bivariate marginal distribution algorithm. Advance in Soft Computing-Engineering Design and Manufacturing 1999 : 521-535.

- Pelikan, M., Goldberg, D.E. and Cantú-Paz, E. BOA: The Bayesian optimization algorithm. In Proceeding of the Genetic and Evolutionary Computation Conference (GECCO) 1999 : 525-532.
- Poli, R. Schema Theory without Expectations for GP and GAs with One-Point Crossover. Foundations of Genetic Programming 1999.
- Reeves, C. R. and Rowe, J. E. Genetic Algorithms – Principles and Perspectives. Kluwer Academic Publisher, 2003.
- Rissanen, J. Modeling by shortest data description. Automatica 1978 : 465-471.
- Rodgers, A., Prugel-Bennett A. Genetic Drift in GA Selection Schemes. IEEE Trans. on Evolutionary Computation 1999 : 298-303.
- Rudnick, W. M. Genetic algorithms and fitness variance with an application to the automated design of artificial neural networks. Doctoral Dissertation Oregon Graduate Institute of Science & Technology Beaverton, USA, 1992
- Schaffer, J. D., Eshelman, L. J. and Offut, D. Spurious correlations and premature convergence in genetic algorithms. Foundations of Genetic Algorithms (FOGA) 1991 : 102-122.
- Schumacher, C., Vose, M. and Whitley, D. The No Free Lunch and Problem Description Length. In Proceeding of the Genetic and Evolutionary Computation Conference (GECCO) 2001.
- Spears, W. M. Crossover or mutation? Foundations of Genetic Algorithms (FOGA) 1993 : 221-237.
- Stephens, C. and Waelbroeck, H. Effective Degrees of Freedom in Genetic Algorithms. Physical review 1998 : 3251-3264.
- Stephens, C. R., Waelbroeck, H. and Aguirre, R. Schemata as Building Blocks: Does Size Matter?. Foundations of Genetic Algorithms (FOGA) 1999 : 117-133.
- Stephens, C. R. and Waelbroeck, H. Schemata Evolution and Building Blocks. IEEE Congress on Evolutionary Computation 1999 : 109-124.
- Syswerda, G. Uniform crossover in genetic algorithms. Proceedings of the Third International Conference on Genetic Algorithms 1989 : 2-9.
- Syswerda, G. Simulated crossover in genetic algorithms. Foundation of Genetic Algorithm 2 (FOGA), 239-255, 1993.

- Tackett, W. A. Recombination, Selection, and the Genetic Construction of Computer Programs. Ph.D. Thesis University of Southern California Department of Computer Engineering, 1994.
- Tang, K., Yao, X., Suganthan, P. N., MacNish, C., Chen, Y. P., Chen, C. M. and Yang, Z. Benchmark Functions for the CEC'2008 Special Session and Competition on Large Scale Global Optimization, Congress on Evolutionary Computation (CEC) 2008.
- Thierens, D. Scalability Problems of Simple Genetic Algorithms. Journal of Evolutionary Computation 1999 : 331-352.
- Toussaint, M. On the evolution of phenotypic exploration distributions. Foundations of Genetic Algorithms (FOGA) 2003 : 169-182.
- Toussaint, M. The structure of evolutionary exploration: On crossover, building blocks, and Estimation-Of-Distribution Algorithms. Genetic and Evolutionary Computation Conference (GECCO) 2003 : 1444-1456.
- Tsutsui, S., Ghosh, A., Corne, D. and Fujimoto, Y. A Real Coded Genetic Algorithm with an Explorer and an Exploiter Populations. Proceedings of the 9th International Conference on Genetic Algorithms (ICGA) 1997 : 238-245.
- Vose, M D. and Nix, A. Modeling Genetic Algorithms with Markov Chains Annals of Mathematics and Artificial Intelligence 1992 : 79-98.
- Watson, R. A., Hornby, G. S. and Pollack, J. B. Modeling Building Block Interdependency. Proceeding of Parallel Problem Solving from Nature (PPSN) 1998 : 97-106.
- Watson, R. A. and Pollack, J. B. Hierarchically consistent test problems for genetic algorithms. Proceeding of Congress on Evolutionary Computation 1999 : 292-297.
- Whitley, L. D. Fundamental principles of deception in genetic search. Foundations of Genetic Algorithms (FOGA) 1991 : 221-241.
- Whitley, D. Functions as permutations regarding no free lunch, walsh analysis and summary statistics. Parallel Problem Solving from Nature 2000 : 169-178.
- Wolpert, D. H. and Macready W. G. No free lunch theorems for search. Technical Report SFI-TR-95-02-010 Santa Fe Institute 1995.

Wolpert, D. H. and Macready, W. G. No free lunch theorems for optimization. IEEE Trans Evolutionary Computation 1997 : 67–82.

Wright, A. H. Genetic algorithms for real parameter optimization. Foundations of Genetic Algorithms (FOGA) 1991 : 205-218.

Yu, T. L., Sastry, K. and Goldberg, D. E. Linkage learning, overlapping building blocks, and systematic strategy for scalable recombination. Conference on Genetic and evolutionary computation 2005 : 1217–1224.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก ก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

ศัพท์บัญญัติ

ภาษาไทย

กระแสข้อมูลนำเข้า

กระแสข้อมูลนำออก

กระแสรหัส

การกวาดตรวจ

การไขว้เปลี่ยน

การไขว้เปลี่ยนแบบจุดเดียว

การไขว้เปลี่ยนแบบชิ้นส่วน

การไขว้เปลี่ยนแบบหลายจุดตัด

การไขว้เปลี่ยนเอกรูป

การจับคู่

การเชื่อมโยง

การเชื่อมโยงแบบแนบแน่น

การเชื่อมโยงแบบหลวม

การดัดแปร

การบีบอัด

การแบ่งปัน

การแบ่งแยก, การแตกแยก

การประกอบชิ้นส่วนเข้าด้วยกัน

การผกผัน

การผ่าเหล่า

การรวมเข้าด้วยกัน, การประกอบเข้าด้วยกันใหม่

การระบุชิ้นส่วน

การรักษาให้คงอยู่

การลวง

การเลือก, วิธีการเลือก, การเลือกสรร

ภาษาอังกฤษ

input stream

output stream

coding stream

scanning

crossover

one point crossover

fragment crossover (FC)

multiple point crossover, multi-point crossover

uniform crossover

mating

linkage

tightly linkage

loosely linkage

modify, modification

compression

share

disruption

fragment composition

inversion

mutation

recombination

fragment identification

retention

deception, deceptive

selection

ภาษาไทย

การลู่เข้า
 การสำรวจ
 การสุ่มแบบเอกรูป
 การแสวงหาประโยชน์
 การเหลื่อม
 ข้อมูลนำเข้า
 ข้อมูลนำออก
 ข้อมูลสนเทศร่วม, สารสนเทศร่วม
 ขั้นตอนวิธี, วิธี
 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบมีเงื่อนไข
 ขั้นตอนวิธีการแตกเป็นชิ้นส่วนแบบสุ่ม
 ขั้นตอนวิธีเชิงพันธุกรรม
 ขั้นตอนวิธีเชิงพันธุกรรมแบบ LZW
 ขั้นตอนวิธีเชิงพันธุกรรมแบบง่าย
 ขั้นตอนวิธีเชิงพันธุกรรมแบบบัญญัติ
 ขั้นตอนวิธีประมาณการแจกแจง
 ขั้นตอนวิธีประมาณการแจกแจงแบบ LZW
 ขั้นตอนวิธีเชิงวิวัฒนาการ
 ความกดดันในการเลือก
 ความซับซ้อนเชิงเวลา
 ความหลากหลาย
 ค่าความเหมาะสม
 ค่าเหมาะสมที่สุดครอบคลุม
 ค่าเหมาะสมที่สุดท้องถิ่น
 คำตอบ
 คู่จีน
 เครื่องเล่นพนันแบบอัตโนมัติชนิดเอาสตาจิสต์

ภาษาอังกฤษ

convergence
 exploration
 uniform random
 exploitation
 overlap, overlapping
 input
 output
 mutual information
 algorithm
 conditional fragmentation (CFT)
 random fragmentation (RFT)
 Genetic Algorithms (GAs)
 LZW Genetic Algorithm (LZWGA)
 simple Genetic Algorithm (sGA)
 canonical Genetic Algorithm
 Estimation of Distribution
 Algorithms (EDAs)
 LZW Estimation of Distribution
 Algorithm (LZWEDA)
 Evolutionary Algorithms (EAs)
 selection pressure
 time complexity
 diversity
 fitness value
 global optimum
 local optimum
 desired solution
 allele
 bandit

ภาษาไทย

เค้าร่าง

จีโนไทป์, รูปแบบพันธุกรรม

ชิ้นส่วน, หน่วยการสร้างแบบชิ้นส่วน

ชิ้นส่วนไม่ร่วมกัน

ชิ้นส่วนร่วมกัน

เชิงตอบโต้

แบบกระจาย

แบบรวมศูนย์

ประชากรย่อย

ปริภูมิค้นหา

ปัญหาค่าเหมาะสมที่สุด

ปัญหาวัดเปรียบเทียบสมรรถนะ

ผลเฉลย, คำตอบ

พ่อพันธุ์แม่พันธุ์

ฟังก์ชันประเมินค่าความเหมาะสม

ฟังก์ชันลวง

ฟีโนไทป์, รูปแบบปรากฏ

ภูมิทัศน์ความเหมาะสม

ระนาบเกิน

รุ่นประชากร

ลำดับย่อย

วิธีการเลือกเชิงอันดับ

วิธีการเลือกแบบการแข่งขันกีฬา

วิธีการเลือกแบบตัดปลาย

วิธีการเลือกแบบตัวดีที่สุดเป็นจำนวน n ตัวไม่ซ้ำกัน

วิธีการเลือกแบบสัดส่วน

วิธีการเลือกแบบเอกรูป

หน่วยการสร้าง

ภาษาอังกฤษ

schema

genotype

fragments

uncommon fragment

common fragment

Interactive

distributed

centralized

sub-population

search space

optimization problems

benchmark problems

solutions

parent

fitness evaluations, function

evaluation (FEs)

deceptive function

phenotype

fitness landscape

hyper-plane

generation

sub-string

ordinal selection

tournament selection

truncation selection

 n -best selection

proportional selection

uniform selection

building blocks

ภาษาไทย

หน่วยเก็บถาวร

อภิสิทธิ์ชน

ภาษาอังกฤษ

archive

elitism

หมายเหตุ ในบางกรณีจะใช้คำว่าหน่วยการสร้าง แทนชั้นส่วน (หรือหน่วยการสร้างแบบชั้นส่วน) เนื่องจากต้องการสื่อให้เห็นถึงหน่วยการสร้างในภาพรวม เพราะชั้นส่วนสามารถสะท้อนให้เห็นถึงหน่วยการสร้างในทางอ้อมได้แม้ว่าจะเป็นเพียงโครงสร้างย่อยรูปแบบหนึ่งของหน่วยการสร้างก็ตาม



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นายเฉลิมทรัพย์ สังขวิจิตร เกิดเมื่อวันที่ 3 กันยายน พ.ศ. 2522 ที่กรุงเทพมหานคร สำเร็จการศึกษาปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้า จากมหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2544 และสำเร็จการศึกษาในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาศาสตร์คอมพิวเตอร์ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2546 หลังจากนั้นได้เข้าศึกษาในหลักสูตรวิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2547



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย