การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงภาษาไทยโดยใช้ขั้นตอนวิธีเชิงพันธุกรรม

นายพัฒนะ ภูริยากร

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

HIDDEN MARKOV MODEL TOPOLOGY ESTIMATION FOR THAI PHONEME RECOGNITION USING GENETIC ALGORITHM

Mr. Pattana Bhuriyakorn

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic year 2007

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงภาษาไทยโดยใช้ขั้นตอนวิธีเชิงพันธุกรรม
โดย	นายพัฒนะ ภูริยากร
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.อติวงศ์ สุชาโต

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.อติวงศ์ สุชาโต)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ชัย วุฒิวิวัฒน์ชัย)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พัฒนะ ภูริยากร : การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงภาษาไทยโดยใช้ขั้นตอนวิธีเชิงพันธุกรรม. (HIDDEN MARKOV MODEL TOPOLGY ESTIMATION FOR THAI PHONEME RECOGNITION USING GENETIC ALGORITHM) อาจารย์ที่ปรึกษา : อาจารย์ ดร.อติวงศ์ สุขาโต, 103 หน้า.

ในปัจจุบันมีการนำแบบจำลองฮิดเดนมาร์คอฟมาประยุกต์ใช้ในการรู้จำรูปแบบต่างๆ เป็นจำนวนมาก การรู้จำเสียงก็เป็นการรู้จำแบบหนึ่งที่นิยมใช้แบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองทางเสียง การรู้จำด้วยเทคนิคนี้ประสิทธิภาพส่วนหนึ่งของการรู้จำเสียงจะขึ้นอยู่กับทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่เลือกใช้ ในวิทยานิพนธ์นี้นำเสนอวิธีการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงในภาษาไทยโดยวิธีที่นำเสนอนั้นประกอบด้วย 2 ขั้นตอน ในขั้นตอนแรกเป็นการสร้างเซตของทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟด้วยการวิธีการประมาณทอพอโลยีที่เกิดจากการจับคู่กันระหว่างฟังก์ชันวัตถุประสงค์กับวิธีการผลิตทอพอโลยี ขั้นตอนต่อมาจะนำขั้นตอนวิธีเชิงพันธุกรรมมาประยุกต์ใช้เป็นขั้นตอนวิธีในการเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่มีความเหมาะสมสำหรับแต่หน่วยเสียงโดยคำนึงถึงความเหมาะสมโดยรวมและการเลือกทอพอโลยีนั้นจะเลือกจากเซตของทอพอโลยีที่สร้างขึ้นในขั้นตอนแรก จากผลการทดลองพบว่า ทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่ประมาณได้จากวิธีที่นำเสนอนั้นสามารถลดค่าความผิดพลาดในการรู้จำเสียงลงได้ 4.36% เมื่อเปรียบเทียบกับทอพอโลยีที่ใช้การเชื่อมต่อของสถานะแบบซ้ายขวา นอกจากนั้นทอพอโลยีดังกล่าวยังสามารถใช้งานได้ดี เมื่อสภาพแวดล้อมของเสียงที่ใช้เป็นฐานข้อมูลเพื่อประมาณทอพอโลยีแตกต่างจากสภาพแวดล้อมของเสียงในการรู้จำ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2550

4970467921 : MAJOR COMPUTER ENGINEERING

KEY WORD: HMM TOPOLOGY / PHONEME RECOGNITION / GENETIC ALGORITHM / MAXIMUM LIKELIHOOD / VITERBI ALGORITHM / BAUM-WELCH RE-ESTIMATION

PATTANA BHURIYAKORN : HIDDEN MARKOV MODEL TOPOLOGY ESTIMATION FOR THAI PHONEME RECOGNITION USING GENETIC ALGORITHM. THESIS ADVISOR : ATIWONG SUCHATO, Ph.D., 103 pp.

The use of Hidden Markov Models (HMM) in many pattern recognition tasks is very common. Like other pattern recognitions, most Automatic Speech Recognition systems rely on HMM acoustic models. In such systems, recognition performances are significantly affected by their topologies. In this research, we proposed an HMM topology estimation approach for Thai phoneme recognition tasks whose process is divided into 2 stages. Firstly, a set of suitable topologies are constructed by combinations of different objective functions and topology generation methods. Second, a Genetic Algorithm is deployed as the topology selection algorithm which considers global fitness and selects the most suitable topology from the candidates proposed in the previous stage for each phoneme. As a result, the well-trained topology yields a maximum of 4.36% error reduction over predefined left-to-right models. The estimated topologies still work well when the topology estimation was performed on speech utterances whose recording environments differ from ones recognized.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department...COMPUTER ENGINEERING... Student's signature.....

Field of study.COMPUTER ENGINEERING... Advisor's signature.....

Academic year.....2007.....

กิตติกรรมประกาศ

ตลอดระยะเวลาการศึกษาและวิจัย ความเกียจคร้าน ความเหนื่อยล้า ความเบื่อหน่าย และสิ่งอันเป็นนิวรณ์อื่นๆ ที่ขัดขวางการทำวิทยานิพนธ์ ผู้เขียนนับว่าโชคดีที่มีอาจารย์อติวงศ์และอาจารย์โปรดปรานที่ช่วยปกป้อง ปัดเป่านิวรณ์ต่างๆ เหล่านั้นให้หายไป ทำให้วิทยานิพนธ์นี้สำเร็จขึ้นมาได้ ไม่เพียงเท่านั้นอาจารย์ทั้งสองท่านยังให้ความรักและความเอ็นดูแก่ผู้เขียนเสมอมา ตั้งแต่เป็นนิสิตปริญญาตรี และขอกราบขอบพระคุณอาจารย์บุญเสริม กิจศิริกุล และดร.ชัย วุฒิวิวัฒน์ชัย ที่สละเวลาอันมีค่ามาเป็นคณะกรรมการสอบวิทยานิพนธ์และให้คำแนะนำหรือข้อเสนอแนะเพิ่มเติมเพื่อให้ผู้ทำวิทยานิพนธ์เห็นประเด็นที่ควรศึกษาเพิ่มเติม ทำให้ผู้เขียนมีแนวทางในการทำวิทยาพนธ์มากขึ้น อีกทั้งยังเป็นการเติมเต็มข้อบกพร่องและประคับประคองผู้เขียนผ่านความลำบากมาด้วยดี

ขอขอบพระคุณอาจารย์วีระ เมืองสิน และกลุ่มปฏิบัติการทดลอง SPACE (Scientific Parallel Computer Engineering) ที่กรุณาเอื้อเฟื้อเครื่องคลัสเตอร์สำ หรับการทำวิจัย ทำให้ผู้เขียนวิทยานิพนธ์สามารถทำงานวิจัยได้อย่างมีสะดวกสบายและประหยัดเวลาในการทำการทดลองได้เป็นอย่างมาก สุดท้ายต้องขอบขอคุณเพื่อนๆ พี่ๆ ชั้น 18 และชั้น 20 ที่เป็นเพื่อนร่วมสนุกและช่วยเหลือยามมีปัญหาทั้งในงานวิจัยและชีวิตประจำวัน ทำให้ช่วงชีวิตที่ใช้ไปในการทำวิทยานิพนธ์เต็มไปด้วยความสนุกสนานและไม่เบื่อหน่าย

สุดท้ายนี้ ที่ขาดไม่ได้คือ บิดามารดาที่คอยให้ความสนับสนุนในทุกด้าน ไม่ว่าจะเป็นเรื่องการเงิน การใช้ชีวิตประจำวัน และการทำวิจัย เป็นต้น อีกทั้งยังอบรมสั่งสอนให้ผู้เขียนวิทยานิพนธ์อยู่ในศีลธรรมอันดี ถึงแม้บางครั้งอาจจะออกนอกลู่นอกทางไปบ้าง จนทำให้ผู้เขียนวิทยานิพนธ์สำเร็จการศึกษาในวันนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 2

1.3 ขอบเขตของการวิจัย 2

1.4 ขั้นตอนและวิธีการดำเนินงานวิจัย 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์ 3

บทที่ 2 ทฤษฎีแและงานวิจัยที่เกี่ยวข้อง 5

2.1 ทฤษฎีที่เกี่ยวข้อง 5

2.1.1 ฟังก์ชันความเป็นไปได้ 5

2.1.2 ความเป็นไปได้สูงสุด 5

2.1.3 การอนุมานแบบเบส์ 6

2.1.4 แบบจำลองฮิดเดนมาร์คอฟ 8

2.1.4.1 กระบวนการมาร์คอฟแบบวิตยุต 8

2.1.4.2 แบบจำลองมาร์คอฟไปสู่แบบจำลองฮิดเดนมาร์คอฟ 10

2.1.4.3 องค์ประกอบของแบบจำลองฮิดเดนมาร์คอฟ 13

2.1.5 การสกัดลักษณะสำคัญ 15

2.1.5.1 สัมประสิทธิ์เซปตรัมบนสเกลเมล 15

2.1.6 หลักการทั่วไปของการรู้จำด้วยเทคนิคแบบจำลองฮิดเดนมาร์คอฟ 16

2.1.7 การประมาณซ้ำแบบบอมเวลช์ 18

2.1.8 การรู้จำและการถอดรหัสวิเทอร์บิ 22

2.1.9 การลดจำนวนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ 23

2.1.9.1 การเลือกใช้แบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่อง..... 23
2.1.9.2 การผูกสถานะ 26
2.1.9.3 การโคลนสถานะ..... 32
2.1.10 หน่วยเสียงในภาษาไทย 33
2.1.10.1 พยัญชนะ..... 33
2.1.10.2 สระ..... 35
2.1.11 ขั้นตอนวิธีเชิงพันธุกรรม 37
2.2 งานวิจัยที่เกี่ยวข้อง..... 43
2.2.1 ฐานข้อมูลโลตัส 43
2.2.2 การประมาณทอพอโลยีแบบหลายทางเดิน..... 44
2.2.3 การประมาณทอพอโลยีโดยใช้ฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยี 45
2.2.3.1 ฟังก์ชันวัตถุประสงค์ 46
2.2.3.2 วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ..... 47
บทที่ 3 การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ 50
3.1 การออกแบบระบบ 50
3.2 วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ 52
3.2.1 การกำหนดรูปแบบการเชื่อมต่อสถานะล่วงหน้า..... 52
3.2.2 วิธีแบ่งสถานะต่อเนื่อง 53
3.2.3 ขั้นตอนวิธีการลดสถานะ..... 53
3.3 ฟังก์ชันวัตถุประสงค์ 54
3.3.1 ความเป็นไปได้มากสุด..... 55
3.3.2 หลักการมีดโกนออคแคม..... 55
3.3.2.1 รายละเอียดประกอบความยาวน้อยสุด 56
3.3.2.2 เกณฑ์สารสนเทศแบบเบส์..... 57
3.3.3 เกณฑ์สารสนเทศแบบเบส์สำหรับแบบจำลองฮิดเดนมาร์คอฟ 57
3.4 การทดลองและผลการทดลอง..... 59
3.4.1 ภาพรวมของการทดลอง..... 59
3.4.1.1 การกำหนดจำนวนสถานะมากสุด..... 59
3.4.1.2 การนับจำนวนพารามิเตอร์อิสระ 60
3.4.1.3 การวัดประสิทธิภาพ..... 60

3.4.2 การทดลองและผลการทดลอง 61
3.4.2.1 การทดลองเมื่อใช้ ML เป็นฟังก์ชันวัตถุประสงค์ 62
3.4.2.2 การทดลองโดยใช้ BIC เป็นฟังก์ชันวัตถุประสงค์ 64
3.4.2.3 การทดลองเมื่อใช้ MDL เป็นฟังก์ชันวัตถุประสงค์ 66
3.4.2.4 การทดลองเมื่อใช้ HBIC เป็นฟังก์ชันวัตถุประสงค์และ
การเปรียบเทียบจำนวนพารามิเตอร์ของทอพอโลยีที่ประมาณได้ 67
บทที่ 4 การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟโดยใช้ขั้นตอนวิธีเชิงพันธุกรรม . 71
4.1 การทดลองคัดเลือกทอพอโลยีโดยใช้จีเอเพียงอย่างเดียว 74
4.1.1 รายละเอียดการทดลอง 74
4.1.2 ผลการทดลอง 75
4.1.3 วิเคราะห์ผลการทดลอง 76
4.2 การทดลองคัดเลือกทอพอโลยีโดยใช้จีเอและทอพอโลยีที่ประมาณได้ในบทที่ 3 76
4.2.1 รายละเอียดการทดลอง 77
4.2.2 ผลการทดลอง 77
4.2.3 วิเคราะห์ผลการทดลอง 78
บทที่ 5 การทดลองประสิทธิภาพของทอพอโลยีที่ประมาณได้ในสภาพแวดล้อมอื่นๆ 79
5.1 รายละเอียดการทดลอง 80
5.2 ผลการทดลอง 81
5.2.1 ผลการทดลองในสภาพแวดล้อมแบบ CO 82
5.2.2 ผลการทดลองในสภาพแวดล้อมแบบ UC 82
5.2.3 ผลการทดลองในสภาพแวดล้อมแบบ UO 83
5.3 วิเคราะห์ผลการทดลอง 83
บทที่ 6 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 85
6.1 สรุปผลการวิจัย 85
6.2 ข้อเสนอแนะ 85
รายการอ้างอิง 87
ภาคผนวก 91
ภาคผนวก ก หน่วยเสียงในภาษาไทยที่ใช้ในฐานข้อมูลโลตัส 91

ภาคผนวก ข แผนภูมิสายงานของขั้นตอนวิธีเชิงพันธุกรรม..... 94

ภาคผนวก ค การเรียนรู้ การรู้จำและการวัดประสิทธิภาพเครื่องรู้จำเสียงพูด 95

ประวัติผู้เขียนวิทยานิพนธ์ 103

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 1 การเปรียบเทียบระหว่าง CHMM และ DHMM 24
ตารางที่ 2 ความแตกต่างระหว่าง CHMM และ ScHMM.....24
ตารางที่ 3 ตัวอย่างกลุ่มข้อมูล 4 กลุ่มเริ่มต้น 29
ตารางที่ 4ตัวอย่างในรอบที่ 2 30
ตารางที่ 5 ขั้นตอนการโคลนสถานะ 32
ตารางที่ 6 ตัวอักษรที่ใช้แทนเสียงพยัญชนะต้นในภาษาไทย.....35
ตารางที่ 7 ตัวอักษรที่ใช้แทนเสียงพยัญชนะท้ายในภาษาไทย.....35
ตารางที่ 8 ตัวอักษรที่ใช้แทนเสียงสระเดี่ยวในภาษาไทย.....37
ตารางที่ 9 ตัวอักษรที่ใช้แทนเสียงสระผสมในภาษาไทย 37
ตารางที่ 10 ตัวอย่างของแผนการแทน 38
ตารางที่ 11 ตัวอย่างประชากรเริ่มต้น (รุ่นที่ 0).....39
ตารางที่ 12 ตัวอย่างการสร้างแมตติงพูล.....40
ตารางที่ 13 ตัวอย่างของสตริงที่ถูกแยกเป็นส่วนที่แตกออกมาไขว้เปลี่ยนและส่วนที่เหลืออยู่ 41
ตารางที่ 14 ตัวอย่างของรุ่นที่ 1 ที่เป็นไปได้ 41
ตารางที่ 15 รายละเอียดของชุดข้อมูลในโลตัส.....44
ตารางที่ 16 องค์ประกอบที่ใช้ในการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ 59
ตารางที่ 17 องค์ประกอบที่ในการสร้างเครื่องรู้จำเพื่อวัดประสิทธิภาพ.....61
ตารางที่ 18 จำนวนสถานะมากสุดของแต่ละหน่วยเสียง.....62
ตารางที่ 19 จำนวนตัวอย่างหน่วยเสียงของแต่ละหน่วยเสียง 64
ตารางที่ 20 องค์ประกอบของเครื่องรู้จำที่ใช้วัดผลของการทดลองที่ 4 74
ตารางที่ 21 องค์ประกอบของเครื่องรู้จำที่ใช้วัดผลในการทดลองที่ 5 81

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 1 ตัวอย่างของแบบจำลองฮิดเดนมาร์คอฟ....1
รูปที่ 2 ตัวอย่างสายโซ่มาร์คอฟที่มี 5 สถานะ....9
รูปที่ 3 ตัวอย่างแบบจำลองที่สร้างขึ้นสำหรับการโยนเหรียญ....12
รูปที่ 4 ขั้นตอนการคำนวณค่าสัมประสิทธิ์เซปตรัมบนสเกลเมล....15
รูปที่ 5 การสร้างเวคเตอร์เสียง....16
รูปที่ 6 ตัวอย่างลักษณะทั่วไปของแบบจำลองฮิดเดนมาร์คอฟ....17
รูปที่ 7 ส่วนผสมเกาส์เซียน....18
รูปที่ 8 ขั้นตอนวิธีวิเทอร์บิสำหรับหน่วยเสียงย่อย....22
รูปที่ 9 ตัวอย่างการผูกสถานะ....26
รูปที่ 10 ระดับการผูก....26
รูปที่ 11 การจำแนกหน่วยเสียง....27
รูปที่ 12 ขั้นตอนวิธีการแบ่งกลุ่มแบบข้อมูลหน่วยขับ....29
รูปที่ 13 ขั้นตอนวิธีต้นไม้ตัดสินใจ....31
รูปที่ 14 แบบจำลองฮิดเดนมาร์คอฟแบบหลายทาง....45
รูปที่ 15 ความสัมพันธ์ระหว่างฟังก์ชันวัตถุประสงค์และวิธีการคัดเลือกทอพอโลยี....46
รูปที่ 16 โครงงานการทดลอง....50
รูปที่ 17 ระบบการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ....51
รูปที่ 18 ตัวอย่างการเชื่อมต่อของสถานะแบบต่างๆ....52
รูปที่ 19 แผนภูมิสายงานขั้นตอนการแบ่งสถานะต่อเนื่อง....53
รูปที่ 20 ขั้นตอนวิธีการลดสถานะโดยเริ่มจากการเชื่อมต่อแบบทั่วถึง....54
รูปที่ 21 ประสิทธิภาพของทอพอโลยีที่ประมาณได้จาก ML....63
รูปที่ 22 ประสิทธิของทอพอโลยีที่ประมาณได้จาก BIC....65
รูปที่ 23 ประสิทธิภาพของทอพอโลยีที่ประมาณได้จาก MDL....66
รูปที่ 24 ความแม่นยำของทอพอโลยีที่ประมาณได้จาก HBIC....68
รูปที่ 25 จำนวนสถานะของทอพอโลยีที่ประมาณได้....68
รูปที่ 26 จำนวนพารามิเตอร์ของทอพอโลยีที่ประมาณได้....69
รูปที่ 27 การลงจุดระหว่างจำนวนพารามิเตอร์และความแม่นยำ....69
รูปที่ 28 แผนการแทนสำหรับทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ....72
รูปที่ 29 ขั้นตอนวิธีเชิงพันธุกรรมสำหรับการเลือกทอพอโลยี....73

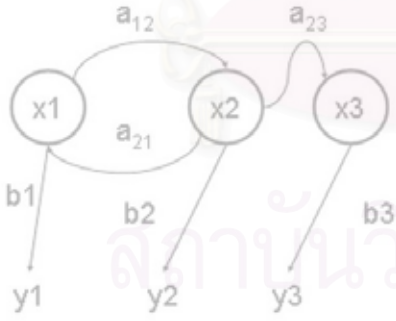
รูปที่ 30 ประสิทธิภาพของทอพอโลยีที่ได้จากการรันจีเอเทียบกับ 5-LTR 7-LTR และ ML-LTR...75
รูปที่ 31 การลงจุดระหว่างความแม่นยำและจำนวนพารามิเตอร์ของการทดลองที่ 4.1 76
รูปที่ 32 ประสิทธิภาพของทอพอโลยีที่ได้จากการรันจีเอเทียบกับวิธีการประมาณในบทที่ 3 77
รูปที่ 33 การลงจุดระหว่างจำนวนพารามิเตอร์กับความแม่นยำของการทดลองที่ 4.2.....78
รูปที่ 34 ขั้นตอนของการทดลองที่ 5 80
รูปที่ 35 ความแม่นยำของการรู้จำในสภาพแวดล้อม CO.....82
รูปที่ 36 ความแม่นยำของการรู้จำในสภาพแวดล้อม UC.....82
รูปที่ 37 ความแม่นยำของการรู้จำในสภาพแวดล้อม UO.....83

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

เทคนิคการรู้จำเสียงพูดที่กำลังเป็นที่นิยมในปัจจุบันสามารถแบ่งออกได้เป็น 2 ประเภท คือ การรู้จำเสียงพูดแบบอาศัยเซกเมนต์ (Segment-based Speech Recognition) ตัวอย่างเช่น ระบบการรู้จำเสียงพูด SUMMIT ของ MIT [1] การทำงานของการรู้จำแบบนี้จะประกอบไปด้วย 2 ขั้นตอนต่อเนื่องกัน คือ ขั้นตอนการแบ่งเสียงพูดออกเป็นเซกเมนต์เพื่อหาขอบเขตของหน่วยเสียงแล้วนำหน่วยเสียงที่ได้นั้นมาประกอบเป็นกราฟของเซกเมนต์ และขั้นตอนการรู้จำเสียงพูดจะนำกราฟของเซกเมนต์ที่ได้เป็นข้อมูลขาเข้าเพื่อค้นหาลำดับของหน่วยเสียงที่เหมาะสมที่สุด ประเภทที่ 2 คือ การรู้จำเสียงพูดแบบอาศัยเฟรม (Frame-based Speech Recognition) ซึ่งเทคนิคที่รู้จักกันอย่างแพร่หลายได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) ดังเช่น ระบบโอนสายโทรศัพท์จากเสียงพูดชื่อไท ย [2] และแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model: HMM) การรู้จำเสียงพูดแบบอาศัยเฟรมมีขั้นตอนการทำงาน 2 ขั้นตอน คือ ขั้นแรกเป็นการสร้างและฝึกฝนหน่วยเสียงซึ่งสามารถทำได้โดยระบุชุดของหน่วยเสียงเพื่อใช้ในการจำแนกกลุ่ม จากนั้นเก็บตัวอย่างเสียงของหน่วยเสียงที่ระบุในขั้นตอนแรก เมื่อมีตัวอย่างเสียงแล้วก็ใช้ตัวอย่างเสียงฝึกฝนแบบจำลองเสียงพูดและรู้จำเสียงพูดโดยใช้แบบจำลองของหน่วยเสียงที่ฝึกฝนมาอย่างดีแล้ว



x — สถานะซ่อน (Hidden State)
y — เอาพุตที่สังเกตได้ (Observable Output)
a — ความน่าจะเป็นของการเปลี่ยนสถานะ (Transition Probability)
b — ความน่าจะเป็นของเอาพต (Output Probability)

รูปที่ 1 ตัวอย่างของแบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองแบบสถิติ โดยที่ระบบที่กำลังจำลองจะถูกสมมุติว่ามีการทำงานเป็นกระบวนการมาร์คอฟ (Markov Process) ด้วยพารามิเตอร์ที่ยังไม่ทราบค่าและสร้างความท้าทายที่จะค้นหาพารามิเตอร์ซ่อน (Hidden Parameter) หากมองถึงแบบจำลองมาร์คอฟทั่วไปแล้ว ผู้สังเกตจะสามารถทราบถึงสถานะในขณะใดขณะหนึ่งได้โดยตรง และความน่าจะเป็นในการเปลี่ยนสถานะ (State Transition Probability) จะเป็นพารามิเตอร์เพียง

อย่างเดียวแต่ในแบบจำลองฮิดเดนมาร์คอฟสถานะในขณะใดขณะหนึ่งจะไม่สามารถทราบได้โดยตรง แต่ละสถานะจะมีการแจกแจงความน่าจะเป็น (Probability Distribution) บนความน่าจะเป็นของเอาพุตที่เป็นไปได้ ตัวอย่างการเปลี่ยนแปลงสถานะ (State Transition) ของแบบจำลองฮิดเดนมาร์คอฟสามารถแสดงได้ดังรูปที่ 1

ขั้นตอนหนึ่งในการสร้างระบบรู้จำโดยอาศัยเทคนิคแบบจำลองฮิดเดนมาร์คอฟ คือ การเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟให้แก่แต่ละหน่วยเสียงที่กำหนด ซึ่งการเลือกทอพอโลยีที่เหมาะสมนั้นจะทำให้การรู้จำเสียงพูดมีประสิทธิภาพมากยิ่ง ขึ้น วิทยานิพนธ์นี้เสนอวิธีการประมาณทอพอโลยีของ แบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมสำหรับการรู้จำหน่วยเสียงภาษาไทยโดยนำฟังก์ชันวัตถุประสงค์ (Objective Function) วิธีการผลิตทอพอโลยี (Topology Generation Method) และขั้นตอนวิธี เชิงพันธุกรรมมาประกอบกันมาประยุกต์ใช้เพื่อค้นหาทอพอโลยีที่เหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

เป้าหมายของวิทยานิพนธ์นี้ต้องการพัฒนาวิธีการค้นหาทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมสำหรับหน่วยเสียงในภาษาไทยเพื่อทำให้ประสิทธิภาพการรู้จำเสียงหน่วยเสียงภาษาไทยมีความถูกต้องแม่นยำมากยิ่งขึ้น

1.3 ขอบเขตของการวิจัย

ในงานวิจัยนี้จะศึกษาการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับหน่วยเสียงในภาษาไทยเพียงอย่างเดียวเท่านั้น ทอพอโลยีที่ได้จากการประมาณนั้นจะขึ้นอยู่กับฐานข้อมูลเสียงที่นำมาใช้ฝึกฝนเท่านั้น ไม่สามารถนำไปใช้กับฐานข้อมูลอื่นได้ และมุ่งเน้นการพัฒนาประสิทธิภาพของการรู้จำหน่วยเสียงในภาษาไทย

1.4 ขั้นตอนและวิธีการดำเนินงานวิจัย

ขั้นตอนการดำเนินงานวิจัยสามารถแบ่งออกเป็นขั้นต่างๆ ได้ดังนี้

1. ขั้นตอนการเตรียมตัว
    - ศึกษาข้อมูลและทฤษฎีต่างๆ ที่เกี่ยวข้องกับงานวิจัย เช่น การสร้างและการทำงานของแบบจำลองฮิดเดนมาร์คอฟ การดึงค่าสำคัญของเสียง แบบจำลองทางสถิติ ทฤษฎีความน่าจะเป็น ขั้นตอนวิธีเชิงพันธุกรรม และความรู้อื่นๆ

- ศึกษาวิธีการใช้งาน HTK (รายละเอียดวิธีการใช้ HTK จะกล่าวในภาคผนวก ค)
- ศึกษางานวิจัยที่เกี่ยวข้อง

2. ขั้นตอนการออกแบบระบบการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ
   - ออกแบบระบบเพื่อให้การประมาณทอพอโลยีโดยใช้ฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยีแบบต่างๆ เป็นไปอย่างมีระบบแบบแผนเพื่อให้ง่ายและมีความถูกต้อง
   - ออกแบบขั้นตอนการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ
   - ออกแบบขั้นตอนวิธีเชิงพันธุกรรมเพื่อนำมาประยุกต์ใช้กับการประมาณทอพอโลยี
3. ขั้นตอนการพัฒนาและทดสอบระบบการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ
   - จัดทำโปรแกรมในแต่ละส่วนตามที่ได้ออกแบบไว้
   - ทดสอบประสิทธิภาพของการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงในภาษาไทยด้วยฐานข้อมูลเสียงภาษาไทย
4. ขั้นตอนวิเคราะห์และสรุปผล
   - วิเคราะห์และสรุปผลการทดลอง
   - เรียบเรียบวิทยานิพนธ์ พร้อมทั้งนำเสนองานวิจัยทั้งหมด

1.5 ประโยชน์ที่คาดว่าจะได้รับ

สามารถนำทอพอโลยีที่ได้ไปสร้างเครื่องรู้จำหน่วยเสียงในภาษาไทยที่ใช้เทคนิคแบบจำลองฮิดเดนมาร์คอฟที่มีประสิทธิภาพและสามารถนำวิธีการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่นำเสนอในงานวิจัยไปใช้ค้นหาทอพอโลยีที่เหมาะสมสำหรับการรู้จำภาษาไทยในสภาพแวดล้อมอื่นๆ

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของงานวิทยานิพนธ์ได้รับการตีพิมพ์เป็นบทความวิชาการในหัวเรื่อง “Comparison of HMM Topology Estimation Approaches for Thai Phoneme Recognition”

โดยพัฒนะ ภูริยากร โปรดปราน บุณยพุกกณะ และอติวงศ์ สุชาโต ในงานประชุมวิชาการ “The Seventh International Symposium on Natural Language Processing” ซึ่งจัดขึ้นที่จังหวัดชลบุรี ประเทศไทย วันที่ 13 – 15 ธันวาคม พ.ศ. 2550 และบทความวิชาการในหัวเรื่อง “A Genetic Algorithm-aided Hidden Markov Model Topology Estimation for Phoneme Recognition of Thai Continuous Speech” โดยพัฒนะ ภูริยากร โปรดปราน บุณยพุกกณะ และอติวงศ์ สุชาโต ในงานประชุมวิชาการ “The Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing” ซึ่งจัดขึ้นที่จังหวัดภูเก็ต ประเทศไทย วันที่ 6-8 สิงหาคม พ.ศ. 2551

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
ทฤษฎีแและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ฟังก์ชันความเป็นไปได้ (Likelihood Function: LF)

ฟังก์ชันความเป็นไปได้ $L(\theta \mid x)$ มีค่าเท่ากับ $f(x:\theta)$ เมื่อ f คือ ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function – PDF) และ θ คือ พารามิเตอร์ ดังนั้น $f(x:\theta)$ จึงคือ ความน่าเป็นที่ x จะเกิดขึ้นบน f เมื่อทราบพารามิเตอร์ θ ไม่เพียงเท่านั้น ฟังก์ชันความเป็นไปได้ยังมีค่าเท่ากับ $P(x \mid \theta)$ ในกรณีที่เป็นการแจกแจงแบบวิยุต (Discrete Distribution) ซึ่งในกรณีนี้ ฟังก์ชันความหนาแน่นของความน่าจะเป็นจะถูกเปลี่ยนเป็นฟังก์ชันมวลของความน่าจะเป็น (Probability Mass Function) แทนทำให้เขียนเป็นสมการได้ว่า

$$L(\theta) = L(\theta \mid x) = f(x:\theta) = P(x \mid \theta) \tag{2.1}$$

$L(\theta) = P(x \mid \theta)$ เมื่อ x มีการแจกแจงแบบวิยุต

2.1.2 ความเป็นไปได้สูงสุด (Maximum Likelihood: ML)

ความเป็นไปได้สูงสุดหรือวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Method) คือ กระบวนการหาค่ามากสุดของความเป็นไปได้ของพารามิเตอร์หนึ่งหรือมากกว่าสำหรับการแจกแจงทางสถิติที่กำหนดให้ เราจะใช้สัญลักษณ์ $\hat{\mu}$ แทนการประมาณความเป็นไปได้สูงสุดของพารามิเตอร์ μ ยกตัวอย่างเช่น สำหรับ การแจกแจงแบบปกติ (Normal Distribution)

$$f(x_1,...,x_n \mid \mu,\sigma) = \prod \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2} = \frac{(2\pi)^{-n/2}}{\sigma^n} \exp\left[-\frac{\sum (x-\mu)^2}{2\sigma^2} \right] \tag{2.2}$$

เมื่อใช้คุณสมบัติของลอการึทึมจะได้ว่า

$$\ln f = -\frac{1}{2} n \ln(2\pi) - n\ln\sigma - \frac{\sum (x_i-\mu)^2}{2\sigma^2} \tag{2.3}$$

เพื่อหาค่าสูงสุดของโดยการหาอนุพันธ์เทียบกับ μ ดังสมการ

$$\frac{\partial(\ln f)}{\partial\mu} = \frac{\sum (x_i-\mu)}{\sigma^2} = 0 \tag{2.4}$$

จึงได้

$$\hat{\mu} = \frac{\sum x_i}{n} \tag{2.5}$$

และในทำนองเดียวกัน เพื่อหาค่ามากสุดของ σ จะได้ว่า

$$\frac{\partial(\ln f)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum (x_i - \mu)^2}{\sigma^3} = 0 \tag{2.6}$$

จึงได้

$$\hat{\sigma} = \sqrt{\frac{\sum (x_i - \hat{\mu}^2)}{n}} \tag{2.7}$$

2.1.3 **การอนุมานแบบเบส์** (Bayesian Inference)

การอนุมานแบบเบส์ คือ การอ้างอิงทางสถิติที่ใช้หลักฐานหรือผลการสังเกตใหม่ปรับค่าความน่าเชื่อถือของสมมุติฐานที่ตั้งเอาไว้ โดยใช้แนวคิดแบบวิทยาศาสตร์ซึ่งเกี่ยวข้องกับการเก็บตัวอย่างที่สอดคล้องหรือขัดแย้งกับสมมุติฐาน เมื่อมีตัวอย่างมากขึ้น โดยมากความเชื่อมั่นในสมมุติฐานที่ตั้งไว้จะมีค่าที่น้อยหรือสูงมากๆ ดังนั้น ทำให้เราสามารถคิดไปเองได้ว่า สมมุติฐานนั้นควรจะเป็นจริงหรือไม่ กล่าวคือ สมมุติฐานที่มีความเชื่อมั่นมากก็ควรจะบอกว่าเป็นจริง และสมมุติฐานที่มีความเชื่อมั่นต่ำมากก็ควรจะเป็นเท็จ ในทางปฏิบัติ จำเป็นที่จะต้องกำหนดค่า ความน่าจะเป็นเริ่มต้นสำหรับสมมุติฐานที่เรากำลังสนใจก่อน เช่น

“สำหรับหลายพันล้านปีที่ผ่านมา พระอาทิตย์ตกแล้วก็ขึ้นในวันรุ่งขึ้น คืนนี้พระอาทิตย์ตก เราจึงมีความเชื่อมั่นสูงมากว่า พรุ่งนี้พระอาทิตย์จะขึ้น และมีความเชื่อมั่นต่ำมากว่าพรุ่งนี้พระอาทิตย์จะไม่ขึ้นพรุ่งนี้”

การอนุมานแบบเบส์ใช้วิธีทางตัวเลข (Numerical Method) เพื่อประมาณค่าความเชื่อมั่นที่มีต่อสมมุติฐานก่อนที่หลักฐานจะปรากฏและใช้วิธีเดียวกันเพื่อประมาณค่าความเชื่อมั่นหลังจากที่หลักฐานปรากฏ โดยใช้แนวคิด ดังสมการต่อไปนี้

$$P(H_0 \mid E) = \frac{P(E \mid H_0)P(H_0)}{P(E)} \tag{2.8}$$

เมื่อ H_0 คือ สมมุติฐานที่เราตั้งไว้ก่อนที่หลักฐาน E จะเกิดขึ้น $P(H_0)$ คือ ความน่าจะเป็นมีมาก่อน (Prior Probability) ของ H_0 ส่วน $P(E \mid H_0)$ คือ ความน่าจะเป็นแบบมี

เงื่อนไข (Conditional Probability) ของการเกิดหลักฐาน E เมื่อสมมุติฐานเป็นจริง เรียกอีกอย่างหนึ่งว่า ฟังก์ชันความเป็นไปได้ของ E ที่เขียนอยู่ในรูปของ H_0 พจน์ $P(H_0 \mid E)$ คือ ความน่าจะเป็นหลัง (Posterior Probability) ของ H_0 เมื่อ E เกิดขึ้น และ $P(E)$ คือ ความน่าเป็นที่หลักฐาน E จะเกิดขึ้น ซึ่งสามารถหาได้โดย

$$P(E) = \sum P(E \mid H_i)P(H_i) \tag{2.9}$$

เมื่อ $H_1, H_2, ..., H_n$ คือ สมมุติฐานแยกออกจากกันเป็นอิสระ (Mutually Exclusive Hypotheses) และจากสมการ (2.8) พจน์ $\frac{P(E \mid H_0)}{P(E)}$ บ่งบอกถึงผลกระทบที่หลักฐาน E ส่งผลต่อความเชื่อมั่นในสมมุติฐาน กล่าวคือ ถ้ามีแนวโน้มว่าหลักฐาน E จะเกิดขึ้นได้บ่อยเมื่อสมมุติฐานเป็นจริง เทอมนี้ก็จะมีค่ามาก เมื่อนำไปคูณกับ ความน่าจะเป็นมีมาก่อนก็จะทำให้ค่าความน่าจะเป็นหลังมีค่ามากขึ้น สำหรับการอนุมานแบบเบส์ ทฤษฎีของเบส์เป็นการบอกว่า หลักฐานใหม่มีผลกระทบต่อความเชื่อมั่นในสมมุติฐานเดิมเท่าใด

หากกำหนดให้มีหลักฐาน E_1 และ E_2 ไม่ขึ้นต่อกัน เราสามารถใช้หลักฐาน E_1 คำนวณหาความน่าจะเป็นมีมาก่อนแล้วใช้ค่าความน่าจะเป็นหลังนี้เป็นคว ามน่าจะเป็นมีมาก่อนเพื่อคำนวณหาความน่าจะเป็นหลังของหลักฐาน E_2 ได้โดย

$$P(E_1, E_2 \mid H_0) = P(E_1 \mid H_0) \times P(E_2 \mid H_0) \tag{2.10}$$

เนื่องจาก E_1 และ E_2 ไม่ขึ้นต่อกัน จึงได้ว่า

$$P(E_1, E_2) = P(E_1) \times P(E_2) \tag{2.11}$$

เมื่อนำไปประยุกต์ในทฤษฎีของเบส์จะได้ว่า

$$P(H_0 \mid E_1, E_2) = \frac{P(E_1 \mid H_0) \times P(E_2 \mid H_0)P(H_0)}{P(E_1) \times P(E_2)} \tag{2.12}$$

และหากกำหนดให้

$$\Lambda = \frac{P(E \mid H_0)}{P(E \mid not\, H_0)} \tag{2.13}$$

และเราทราบว่า

$$P(E) = P(E \mid H_0)P(H_0) + P(E \mid not\ H_0)P(not\, H_0) \tag{2.14}$$

จึงสามารถเขียนสมการใหม่ได้เป็น

$$P(H_0 \mid E_1, E_2) = \frac{\Lambda_1 \Lambda_2 P(H_0)}{\left[\Lambda_1 P(H_0) + P(not\, H_0)\right]\left[\Lambda_2 P(H_0) + P(not\, H_0)\right]} \qquad (2.15)$$

ซึ่งจะเห็นว่า ด้วยวิธีนี้ทำให้สามารถใช้กับจำนวนหลักฐานที่ไม่ขึ้นต่อกันมากกว่า 2 หลักฐานขึ้นไปได้และค่า Λ ก็คือ อัตราส่วนความเป็นไปได้ (Likelihood Ratio) นั่นเอง

2.1.4 แบบจำลองฮิดเดนมาร์คอฟ [3]

ทฤษฎีพื้นฐานของแบบจำลองฮิดเดนมาร์คอฟถูกตีพิมพ์ครั้งแรกโดย Baum และคณะ [4-8] ในช่วงปลายทศวรรษ 1960 ถึงต้นทศวรรษ 1970 หลังจากนั้น Baker [9] กับ Jelinek และคณะ [10-15] ได้นำแบบจำลองฮิดเดนมาร์คอฟมาประยุกต์ใช้กับการประมวลผลเสียงพูด อย่างไรก็ตามแบบจำลองฮิดเดนมาร์คอฟถูกใช้อย่างแพร่หลายในการประมวลผลเสียงพูดในราวทศวรรษที่ 1980 เนื่องจากทฤษฎีของแบบจำลองฮิดเดนมาร์คอฟถูกตีพิมพ์ในวารสารทางคณิตศาสตร์และในงานวิจัยแรกๆ ก็อธิบายถึงการนำแบบจำลองฮิดเดนมาร์คอฟไปประยุกต์กับการประมวลเสียงพูดไม่ละเอียดเพียงพอ ทำให้ผู้อ่านนั้นไม่สามารถเข้าใจและไม่สามารถนำแบบจำลองฮิดเดนมาร์คอฟไปประยุกต์ใช้ในงานวิจัยของตนได้ ด้วยปัญหาที่เกิดขึ้นทำให้มีงานวิจัยมากมายอธิบายถึงขั้นตอนและวิธีการประยุกต์ใช้แบบจำลองฮิดเดนมาร์คอฟกับการประมวลเสียงพูด

2.1.4.1 กระบวนการมาร์คอฟแบบวิตยุต (Discrete Markov Process)

พิจารณาระบบที่มีเซตของสถานะที่แตกต่างกัน N สถานะ $S_1, S_2, \cdots, S_N$ เช่นในรูปที่ 2 $N = 5$ ทุกๆ ช่วงเวลาหนึ่งระบบจะเปลี่ยนสถานะจากสถานะหนึ่งไปยังอีกสถานะหนึ่งตามเซตของความน่าจะเป็นในการเปลี่ยนสถานะ ดังแสดงในรูปที่ 2 ให้ $t = 1,2,\cdots$ แทนเวลาที่มีการเปลี่ยนสถานะและ q_t แทนสถานะ ณ เวลา t สำหรับ กระบวนการสายโซ่มาร์คอฟ (Markov Chain Process) ความน่าจะเป็นในการเปลี่ยนสถานะจะขึ้นอยู่กับเวลา ณ ตอนนั้นและสถานะก่อนหน้าเท่านั้น นั่นคือ

$$P[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \cdots] = P[q_t = S_j | q_{t-1} = S_i] \qquad (2.16)$$

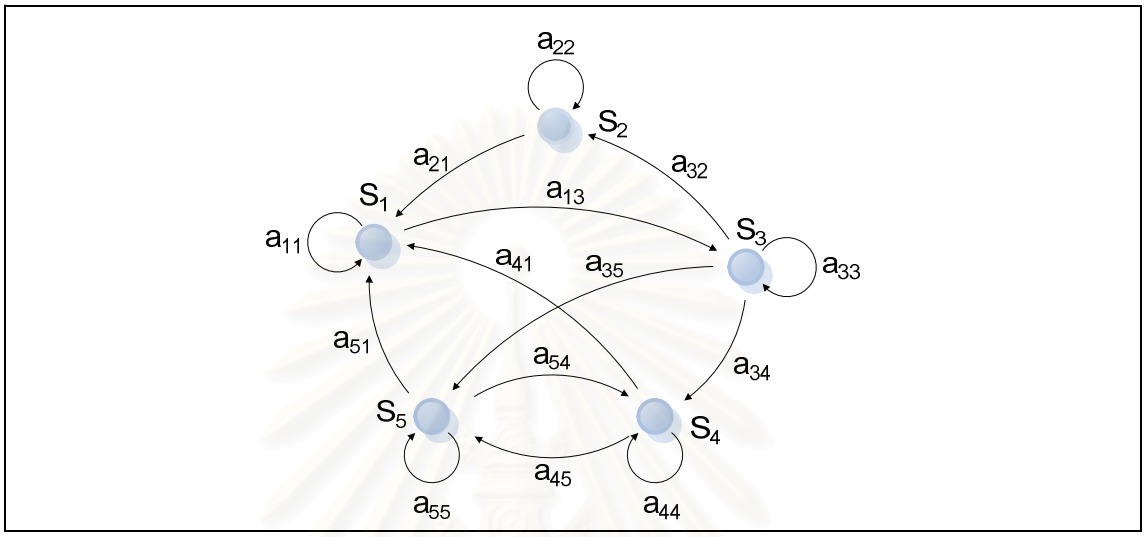
หากพิจารณาพจน์ทางขวามือของสมการ (2.16) โดยไม่สนใจเวลา เราจะได้เซตของความน่าจะเป็นในการเปลี่ยนสถานะ ดังสมการ (2.17)

$$a_{ij} = P[q_t = S_j \mid q_{t-1} = S_i], \quad 1 \le i, j \le N \qquad (2.17)$$

ค่าสัมประสิทธิ์การเปลี่ยนแปลงสถานะ a_{ij} มีคุณสมบัติดังนี้

$$a_{ij} \geq 0 \tag{2.18a}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{2.18b}$$



รูปที่ 2 ตัวอย่างสายโซ่มาร์คอฟที่มี 5 สถานะ (ดัดแปลงจาก [3])

ระบบที่กล่าวมานี้ คือ แบบจำลองมาร์คอฟสังเกตได้ (Observable Markov Model) เนื่องจากข้อมูลขาออกของกระบวนการคือ เซตของสถานะ ณ เวลาใดเวลาหนึ่ง เมื่อ แต่ละสถานะเกี่ยวพันกับเหตุการณ์ใดเหตุการณ์หนึ่ง ยกตัวอย่างเช่น พิจารณาการจำลองเหตุการณ์ของอากาศโดยอาศัยแบบจำลองมาร์คอฟที่มี 3 สถานะ โดยสมมุติว่า ในหนึ่งวันอากาศเป็นหนึ่งอย่างใดต่อไปนี้

State 1 : หิมะตก

State 2 : ฝนตก

State 3 : แดดออก

กำหนดให้ A คือ เมตริกการเปลี่ยนสถานะ (Transition Matrix) ของการเหตุการณ์นี้เป็นดังแสดง

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \tag{2.19}$$

ให้อากาศในวันที่ 1 ($t = 1$) เป็นแดดออก (State 3) เราสามารถบอกได้ว่า ความน่าจะเป็นที่อากาศ 7 วันข้างหน้าจะเป็น แดดออก แดดออก หิมะตก หิมะตก แดดออก ฝนตก แดด

ออก เป็นเท่าใด ซึ่งก็คือ $O=\{S_3,S_3,S_3,S_1,S_1,S_3,S_2,S_3\}$ เป็นลำดับการสังเกต (Observation sequence) ตามเวลา $t=1,2,\cdots,8$ เราต้องการหาความน่าจะเป็นของ O เมื่อทราบจำลองซึ่งสามารถคำนวณได้ดังนี้

$$\begin{aligned}P(O\mid Model) &= P(S_3,S_3,S_3,S_1,S_1,S_3,S_2,S_3\mid Model)\\ &= P[S_3]\cdot P[S_3\mid S_3]\cdot P[S_3\mid S_3]\cdot P[S_1\mid S_3]\\ &\quad\cdot P[S_1\mid S_1]\cdot P[S_3\mid S_1]\cdot P[S_2\mid S_3]\cdot P[S_2\mid S_3]\\ &= \pi_3\cdot a_{33}\cdot a_{33}\cdot a_{13}\cdot a_{11}\cdot a_{13}\cdot a_{32}\cdot a_{23}\\ &= 1.536\times 10^{-4}\end{aligned}$$

เมื่อ π แทนความน่าจะเป็นของสถานะเริ่มต้น

$$\pi_i = P[q_1=S_i],\qquad 1\le i\le N \tag{2.20}$$

อีกปัญหาหนึ่งที่น่าสนใจและสามารถตอบได้คือ เมื่อทราบแบบจำลองและทราบว่า ณ เวลานั้นอยู่ที่สถานะใดความน่าจะเป็นที่จะคงอยู่ ณ สถานะนั้นเป็นเวลา d คือเท่าใด คำถามนี้สามารถคำนวณได้จากลำดับการสังเกต

$$O=\{\underset{1}{S_i},\underset{2}{S_i},\underset{3}{S_i},\cdots,\underset{d}{S_i},\underset{d+1}{S_k}\},\qquad S_k\neq S_i$$

เนื่องจากเราทราบแบบจำลองทำให้ได้ว่า

$$P(O\mid Model,q_1=S_i)=(a_{ii})^{d-1}(1-a_{ii})=p_i(d) \tag{2.21}$$

ปริมาณ $p_i(d)$ เป็นค่าเฉพาะของแต่ละสถานะในสายโซ่มาร์คอฟ (Markov Chain) นอกจากนี้เรายังสามารถคำนวณ ค่าคาดหวังของระยะเวลาที่สถานะจะไม่เปลี่ยนแปลงได้โดย

$$\bar{d}_i=\sum_{d=1}^{\infty}dp_i(d)=\sum_{d=1}^{\infty}d(a_{ii})^{d-1}(1-a_{ii})=\frac{1}{1-a_{ii}} \tag{2.22}$$

จากสมการ 6 ทำให้ทราบว่า ค่าคาดหวังของการที่แดดจะออกติดต่อกันตามแบบจำลองที่กล่าวไปในตอนต้น เท่ากับ $\frac{1}{0.2}=5$ สำหรับหิมะตกเท่ากับ $\frac{1}{0.6}=1.67$ และสำหรับฝนตกเท่ากับ $\frac{1}{0.4}=2.5$ เป็นต้น

2.1.4.2 แบบจำลองมาร์คอฟไปสู่แบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองมาร์คอฟเป็นแบบจำลองที่สามารถนำไปใช้ได้ประโยชน์ได้ แต่อย่างไรก็ดีแบบจำลองนี้มีข้อจำกัดมากเกินไปทำให้ไม่สามารถนำไปประยุกต์ใช้กับการประมวลเสียงพูดได้ ดังนั้น จึงจำเป็นที่จะต้องขยายแนวคิดของแบบจำลองมาร์คอฟเพิ่มขึ้นเพื่อให้ครอบคลุม กรณีที่ผลการสังเกต (Observation) เป็นฟังก์ชันของความน่าจะเป็นสำหรับแต่ละสถานะ แบบจำลองลักษณะนี้คือแบบจำลองฮิดเดนมาร์คอฟซึ่งเราจะไม่ทราบถึงการเปลี่ยนแปลงของสถานะ มีเพียงแต่ลำดับการสังเกตเท่านั้นที่สามารถทราบได้ เพื่อให้เกิดเห็นภาพมากขึ้น พิจารณาปัญหาการโยนเหรียญต่อไปนี้

แบบจำลองการโยนเหรียญ (Coin Toss Model) สมมุติว่า เราอยู่ในห้องซึ่งมีม่านกั้น ทำให้ไม่สามารถทราบได้ว่าเกิดอะไรขึ้น อีกฝั่งของม่านมีคนอีกคนหนึ่งกำลังทำการทดลองโยนเหรียญ (อาจใช้เหรียญมากกว่า 1 เหรียญ) เราไม่ทราบวิธีการทดลองและคนที่ทำการทดลองจะบอกเพียงแค่ผลลัพธ์ของการโยนเหรียญในแต่ละครั้ง ซึ่งจะเป็น หัว (H) หรือก้อย (T) เท่านั้น ดังนั้นลำดับการสังเกตที่เราทราบได้จึงเป็น

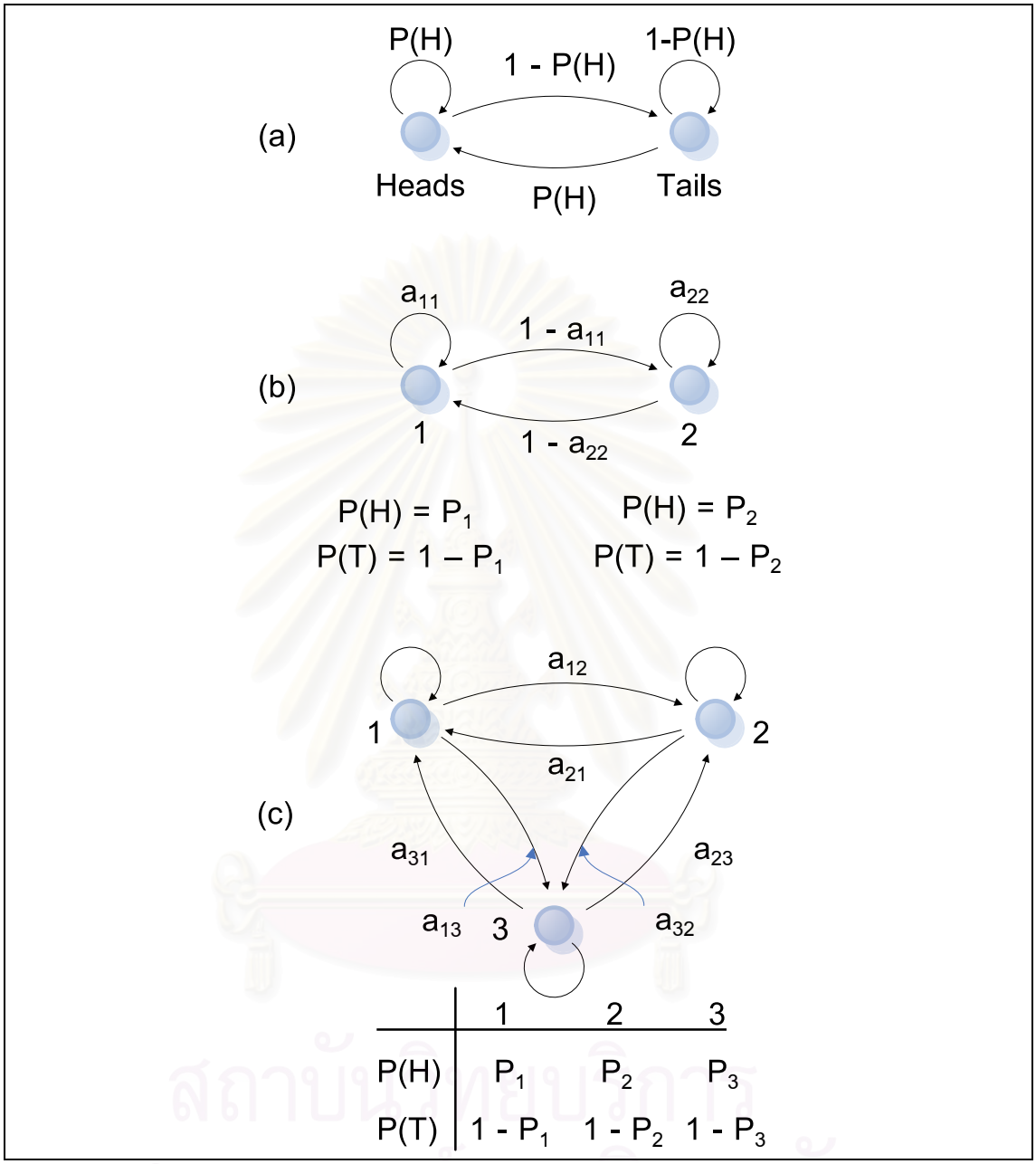
$$O = O_1, O_2, O_3, \cdots, O_t \tag{2.23a}$$

$$= H, H, T, H, T, \cdots, T \tag{2.23b}$$

(2.23b) เป็นรูปแบบหนึ่งของลำดับการสังเกตที่เป็นไปได้ การสร้างแบบจำลองฮิดเดนมาร์คอฟเพื่อที่จะอธิบายผลการสังเกตที่เห็น ต้องกำหนดว่าสถานะแต่ละสถานะของแบบจำลองฮิดเดนมาร์คอฟจะใช้แทนสิ่งใดและกำหนดว่าจะมีทั้งหมดกี่สถานะ ตัวเลือกหนึ่งที่เป็นไปได้สำหรับปัญหาการโยนเหรียญ คือ ตั้งสมมุติฐานว่า การโยนเหรียญนั้นใช้เหรียญเพียงแค่เหรียญเดียว ในกรณีนี้ เราจะสามารถจำลองเหตุการณ์ได้ด้วยแบบจำลอง 2 สถานะ แต่ละสถานะแทนหน้าของเหรียญ ดังแสดงในรูปที่ 3a ซึ่งแบบจำลองนี้ก็คือ แบบจำลองมาร์คอฟนั่นเอง

อีกวิธีหนึ่งที่เป็นไปได้คือ ตั้งสมมุติฐานว่าการโยนเหรียญใช้เหรียญมากกว่าหนึ่งเหรียญ เช่น พิจารณาว่ามีเหรียญทั้งหมด 2 เหรียญ ในสถานการณ์แบบนี้ เราไม่สามารถประยุกต์ใช้แบบจำลองมาร์คอฟได้ เนื่องจากเราทราบว่า ผลการสังเกตที่เกิดขึ้นนั้นเกิดจากโยนเหรียญอันใด แต่แบบจำลองฮิดเดนมาร์คอฟสามารถจำลองเหตุการณ์นี้ได้โดย ให้แต่ละสถานะคือ เหรียญแต่ละเหรียญที่จะถูกทอยและแต่ละสถานะจะมีการแจกแจงของความน่าจะเป็น (Probability distribution) ของการเกิดหัวและก้อยเป็นของตัวเอง เราใช้เม ตริกการเปลี่ยนสถานะเพื่อกำหนดการเปลี่ยนแปลงระหว่างสถานะ ซึ่งในที่นี้ก็คือ การที่ผู้ทำการทดลองจะปลี่ยนเหรียญที่

ใช้โยน การจำลองการโยนเหรียญโดยตั้งสมมุติฐานว่ามีเหรียญ 2 เหรียญและ 3 เหรียญสามารถแสดงได้รูปที่ 3b และรูปที่ 3c ตามลำดับ



รูปที่ 3 ตัวอย่างแบบจำลองที่สร้างขึ้นสำหรับการโยนเหรียญ (ดัดแปลงจาก [3])

ถึงแม้จะสามารถสร้างแบบจำลองได้แล้ว ปัญหาที่พบต่อมาก็คือ เราสามารถสร้างแบบจำลองได้หลายรูปแบบ แล้วแบบจำลองแบบใดเป็นแบบจำลองที่เหมาะสมที่สุดกับผลการสังเกตที่เกิดขึ้นจริง และหากพิจารณาถึงจำนวนพารามิเตอร์จะได้ว่า แบบจำลอง รูปที่ 3a มีพารามิเตอร์ที่ไม่ทราบค่า 1 พารามิเตอร์ แบบจำลอง รูปที่ 3b มี 4 พารามิเตอร์ที่ไม่ทราบค่าและ 9 พารามิเตอร์ที่ไม่ทราบค่าในแบบจำลองรูปที่ 3c ในทางปฏิบัติการมีจำนวนพารามิเตอร์มากขึ้น

ไม่ได้ส่งผลให้แบบจำลองดีขึ้น แบบจำลองที่ดีที่สุดคือ แบบจำลองที่มีความใกล้เคียงกับเหตุการณ์ที่เกิดขึ้นจริงมากที่สุด ดังเช่น ในตัวอย่างการโยนเหรียญ หากผู้ทดลองใช้เหรียญเพียงเหรียญเดียวเพื่อสร้างผลการสังเกตแต่เราตั้งสมมุติฐานว่ามี 3 เหรียญ ในกรณีนี้แบบจำลองที่สร้างขึ้นไม่สัมพันธ์เหตุการณ์ที่เกิดขึ้นจริง ทำให้การจำลองเหตุการณ์ไม่เหมาะสม เป็นต้น

2.1.4.3 องค์ประกอบของแบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนมาร์คอฟประกอบด้วยองค์ประกอบต่างๆ ดังต่อไปนี้

1. N คือ จำนวนสถานะในแบบจำลอง หากพิจารณาปัญหาการโยนเหรียญที่กล่าวไป N คือ จำนวนเหรียญที่เราตั้งสมมุติฐานให้ผู้ทำทดลองใช้โยน เราใช้สัญลักษณ์ $S = \{S_1, S_2, \cdots, S_N\}$ แทนสถานะที่ $1,2,\cdots,N$ และใช้สัญลักษณ์ q_t แทนสถานะ ณ เวลา t
2. M คือ **จำนวนสัญลักษณ์ของผลการสังเกต**ที่แตกต่างกัน เช่น การโยนเหรียญสัญลักษณ์ของผลการสังเกตมี 2 แบบ คือ H และ T เราใช้สัญลักษณ์ $V = \{V_1, V_2, \cdots, V_M\}$ แทนสัญลักษณ์ของผลการสังเกต
3. การแจกแจงความน่าจะเป็นในการเปลี่ยนสถานะ (State Transition Probability Distribution) $A = \{a_{ij}\}$

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i], \qquad 1 \le i, j \le N \tag{2.24}$$

4. การแจกแจงของความน่าจะเป็นของสัญลักษณ์การสังเกต (Observation Symbol Probability Distribution) สำหรับสถานะ S_i, $B = \{b_i(k)\}$ โดยที่

$$b_i(k) = P[V_k \; at \; t \mid q_t = S_i], \qquad 1 \le i \le N, 1 \le k \le M \tag{2.25}$$

5. การแจกแจงของสถานะเริ่มต้น $\pi = \{\pi_i\}$ เมื่อ

$$\pi_i = P[q_1 = S_i], \quad 1 \le i \le N \tag{2.26}$$

เมื่อมีค่าที่เหมาะสมของ N, M, A, B และ π แบบจำลองฮิดเดนมาร์คอฟก็สามารถใช้เป็นเครื่องผลิตลำดับการสังเกต

$$O = O_1, O_2, \cdots, O_T \tag{2.27}$$

เมื่อผลการสังเกต O_t แต่ละตัวเป็นสมาชิกของเซต V และ t คือ เวลา ณ ขณะใดขณะหนึ่งได้ด้วยขั้นตอนต่อไปนี้

1. กำหนดสถานะเริ่มต้น $q_1 = S_i$ จาก π
2. ตั้งค่า $t = 1$
3. เลือก $O_t = V_k$ โดยใช้การแจกแจงของความน่าจะเป็นของสัญลักษณ์การสังเกตสำหรับสถานะ S_i ($b_i(k)$)
4. เปลี่ยนสถานะไปสู่สถานะ $q_{t+1} = S_j$ ตามเมตริกการเปลี่ยนสถานะของสถานะ S_i (a_{ij})
5. ค่า $t = t + 1$ และกลับไปสู่ขั้นตอนที่ 3 จนกระทั่ง $t = T$ จึงหยุด

จะเห็นว่า การระบุแบบจำลองฮิดเดนมาร์คอฟจำเป็นต้องมีองค์ประกอบของแบบจำลอง (N และ M) สัญลักษณ์ของผลการสังเกต และ A, B, π แต่ทั่วไปแล้วนิยมใช้สัญลักษณ์โดยย่อ

$$\lambda = (A, B, \pi) \tag{2.28}$$

เพื่อระบุองค์ประกอบที่สมบูรณ์ของแบบจำลองฮิดเดนมาร์คอฟและการที่จะนำไปประยุกต์ใช้ได้นั้น จำเป็นจะต้องแก้ปัญหา 3 ข้อ ดังต่อไปนี้ก่อน คือ

1. หากมีลำดับการสังเกต $O = O_1, O_2, \cdots, O_T$ และแบบจำลอง $\lambda = (A, B, \pi)$ เราจะสามารถคำนวณหา $P(O|\lambda)$ ได้อย่างไร ซึ่งปัญหานี้จะอธิบายในหัวข้อ 2.1.6
2. หากมีลำดับการสังเกต $O = O_1, O_2, \cdots, O_T$ และแบบจำลอง $\lambda = (A, B, \pi)$ เราจะสามารถเลือกลำดับของสถานะ $Q = q_1, q_2, \cdots, q_T$ ที่สัมพันธ์กับลำดับการสังเกตดังกล่าวได้อย่างไร ปัญหานี้จะอธิบายในหัวข้อ 2.1.8
3. เราจะปรับแต่งพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ เพื่อให้ได้ค่า $P(O|\lambda)$ มากที่สุดได้อย่างไร ปัญหานี้จะอธิบายในหัวข้อ 2.1.7

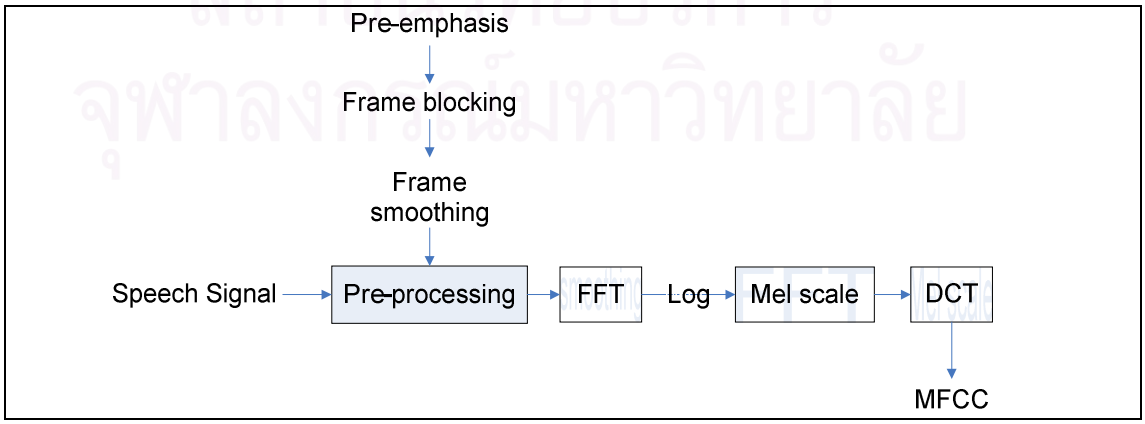
ถึงแม้ว่า แบบจำลองฮิดเดนมาร์คอฟจะเป็นเทคนิคที่มีประสิทธิภาพในการจำแนกรูปแบบที่ดีที่สุดวิธีหนึ่ง [16] แต่อย่างไรก็ดี แบบจำลองฮิดเดนมาร์คอฟก็มีข้อจำกัดอยู่หลายประการสำหรับการประมวลผลเสี ยงพูด เช่น การตั้งสมมุติฐานที่ว่าผลการสังเกตที่ติดกันไม่ขึ้นต่อกัน การตั้งสมมุติฐานที่ว่า ผลการสังเกตสามารถแสดงให้อยู่ในรูปส่วนผสมของเกาส์เซียน (Mixture of Gaussian) หรือความหนาแน่นถดถอยอัตโนมัติ (Autoregressive Density) ได้เป็นอย่างดี และสมมุติฐานของแบบจำลองมาร์คอฟที่ว่า ความน่าจะเป็นของสถานะที่เวลา t ขึ้นอยู่กับสถานะก่อนหน้าที่เวลา $t-1$ เท่านั้น ซึ่งทั้งสามสมมุติฐานนี้เห็นได้ชัดว่าไม่เหมาะสมกับการประมวลผลเสียงพูด

2.1.5 การสกัดลักษณะสำคัญ [17]

การสกัดลักษณะสำคัญ คือ การแปลงข้อมูลขาเข้าที่มีขนาดใหญ่และมีข้อมูลที่ไม่จำเป็นมากเกินไปให้ไปอยู่ในรูปแบบที่สามารถใช้ประโยชน์ได้และมีขนาดเล็กลงโดยที่ยังไม่สูญเสียลักษณะที่สำคัญไป สำหรับการรู้จำเสียงนั้นนิยมใช้การสกัดค่าสำคัญแบบเอนเวโลบเชิงสเปกตรัม (Spectral Envelop Feature) [18] เพราะว่าสามารถดึงค่าสำคัญที่จำเป็นที่ต้องใช้ในการรู้จำเสียงออกมาได้เกือบทั้งหมดและยังสามารถคำนวณได้ง่ายอีกด้วย การสกัดลักษณะสำคัญในกลุ่มนี้ได้แก่ สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Prediction Coefficients: LPC) [19] สัมประสิทธิ์การสะท้อน (Reflection Coefficients: RC) การทำนายเชิงเส้นแบบรับรู้ (Perceptual Linear Prediction: PLP) [20] สัมประสิทธิ์เซปตรัม (Cepstral Coefficient) ซึ่งมีพัฒนาออกไปมากมาย เช่น สัมประสิทธิ์เซปตรัมบนสเกลเมล (Mel Frequency Cepstral Coefficients: MFCC) สัมประสิทธ์เซปตรัมเชิงเส้น (Linear Frequency Cepstrum Coefficients: LFCC) สเปกตรัมโมดูเลชันบนสเกลเมล (Mel-Cepstrum Modulation Spectrum : MCMS) [21] เป็นต้น

2.1.5.1 สัมประสิทธิ์เซปตรัมบนสเกลเมล

การรู้จำเสียงพูดและผู้พูดนิยมนำค่าสัมประสิทธิ์เซปตรัมเป็นค่าลักษณะสำคัญ โดยการคำนวณนั้น เริ่มจากนำสัญญาณไปทำการแปลงฟูริเยร์แบบเร็ว (Fast Fourier Transform: FFT) และนำค่าลอการิธึม (Logarithm) ของแอมพลิจูดของสเปกตรัมที่ได้ไปผ่านตัวกรอง (Filter bank) หากเป็นสัมประสิทธิ์บนเกสเมลก็จะนำไปผ่านสเกลเมล (Mel-scale) ซึ่งเป็นตัวกรองที่ออกแบบมาให้เหมาะสมกับการรับฟังของหู MFCC คือ ค่าของแอมพลิจูดของสเปตรัมผลลัพธ์จากการแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete Cosine Transformation) ของสเปกตรัมที่ผ่านตัวกรองสเกลเมล ซึ่งสามารถเขียนได้ดังรูปที่ 4

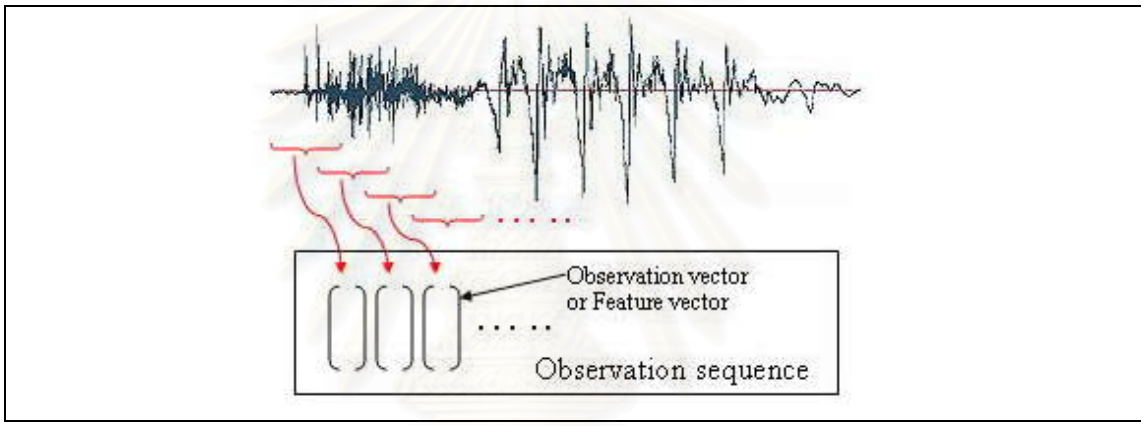


รูปที่ 4 ขั้นตอนการคำนวณค่าสัมประสิทธิ์เซปตรัมบนสเกลเมล

MFCC วิเคราะห์สัญญาณเสียงทีละหนึ่งกรอบ แต่ในความเป็นจริงสัญญาณเสียงมีการเปลี่ยนแปลงอยู่ตลอดเวลา ดังนั้น Furui จึงเสนอให้เพิ่มเดลต้าฟีเจอร์ (Delta Feature) เพื่อสกัดลักษณะการเปลี่ยนแบบพลวัตของสัญญาณเสียง [22] ไม่เพียงเท่านั้น วิธีการคำนวณ เดลต้าฟีเจอร์ยังสามารถนำไปใช้คำนวณความเร่งหรือเดลต้า-เดลต้าได้อีกด้วย

2.1.6 หลักการทั่วไปของการรู้จำด้วยเทคนิคแบบจำลองฮิดเดนมาร์คอฟ [23]

ในการรู้จำนั้น เราจะนำเสียงที่ได้มาเทียบกับเสียงที่เรามีอยู่ ซึ่งก่อนที่เราจะนำมาเทียบนั้นเราต้องทำการสกัดค่าสำคัญของเสียงนั้นๆออกมาก่อน ซึ่งเราเรียกค่าสำคัญนี้ว่าเวคเตอร์เสียง (Speech Feature Vector) โดยค่าสำคัญที่นิยมในปัจจุบันคือ สัมประสิทธิ์เซปตรัมบนสเกลเมล การสกัดค่าสำคัญนี้สามารถแสดงได้ดังรูปที่ 5



รูปที่ 5 การสร้างเวคเตอร์เสียง (รูปจาก http://thaispeech.longdo.org)

เมื่อเราสกัดค่าสำคัญ (เวคเตอร์เสียง) ออกมาจากเสียงใดเสียงหนึ่งได้แล้ว เราก็จะทำการเทียบเสียงที่ได้มานี้กับรูปแบบของเสียงที่เรามีอยู่โดยการที่ จะตอบว่าเป็นคำใดนั้นจะทำได้จากการนำลำดับการสังเกตที่ได้มานั้นไปเทียบกับที่เรามีอยู่แล้วตัวใดให้ความน่าจะเป็นมากสุดก็ตอบตัวนั้น ซึ่งหลักการดังกล่าวสามารถสรุปได้ดังนี้

ให้คำที่เข้ามาสามารถแสดงได้ด้วยลำดับการสังเกตหรือเวคเตอร์เสียง (O) ซึ่งเขียนได้ดังนี้

$$O = o_1, o_2, ..., o_T$$

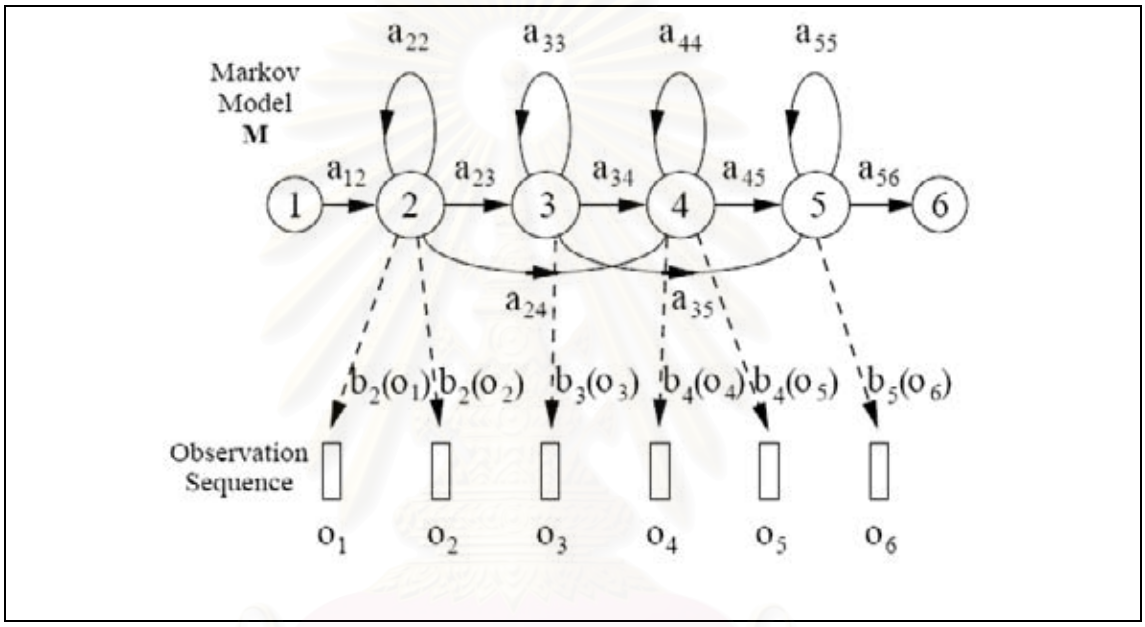
เมื่อ o_t คือ เวคเตอร์เสียงที่สังเกตเมื่อเวลา t สำหรับคำหนึ่งคำเราสามารถตอบได้โดยการคำนวณหาความน่าจะเป็นโดย

$$\arg\max\left[P(w_i \mid O)\right] \qquad (2.29)$$

เมื่อ w_i คือ คำที่ i ในพจนานุกรม เนื่องจากความน่าจะเป็นไม่สามารถคำนวณได้โดยตรงแต่สามารถคำนวณได้จากกฎของเบส์ดังนี้

$$P(w_i \mid O) = \frac{P(O \mid w_i)P(w_i)}{P(O)} \tag{2.30}$$

แบบจำลองฮิดเดนมาร์คอฟเป็นวิธีหนึ่งที่ช่วยทำให้เราหา $P(w_i \mid O)$ โดยแบบจำลองฮิดเดนมาร์คอฟนี้จะสมมุติว่าลำดับของเวคเตอร์เสียงที่สัมพันธ์กับแต่ละคำจะถูกสร้างขึ้นมาจากเครื่องจักรสถานะ (State machine)



รูปที่ 6 ตัวอย่างลักษณะทั่วไปของแบบจำลองฮิดเดนมาร์คอฟ (ดัดแปลงจาก [23])

แบบจำลองมาร์คอฟเป็นเครื่องจักรสถานะซึ่งจะเปลี่ยนสถานะทุกๆ เวลา t ใหม่เข้ามา ซึ่งแต่ละสถานะของแบบจำลองฮิดเดนมาร์คอฟจะสร้างลำดับการสังเกต ณ เวลานั้นด้วยความน่าจะเป็น $b_j O(t)$ (j คือ สถานะ) และมีการเปลี่ยนสถานะได้ 2 แบบ คือ เปลี่ยนจากสถานะ i ไปสถานะ j หรืออยู่ที่เดิม ด้วยความน่าจะเป็น a_{ij}

รูปที่ 6 เป็นตัวอย่างของแบบจำลองฮิดเดนมาร์คอฟซึ่งมี 6 สถานะ โดยเริ่มจากซ้ายไปขวา ให้ X คือ ลำดับการเปลี่ยนสถานะ (State Sequence) X = 1, 2, 2, 3, 4, 4, 5, 6 เพื่อสร้างเวคเตอร์เสียง o_1 ถึง o_6 (สถานะเริ่มต้นและสถานะสุดท้ายจะไม่สร้างผลการสังเกต) จากเครื่องจักรสถานะดังกล่าวทำให้เราสามารถหาความน่าจะเป็นสำหรับ X ดังกล่าวได้คือ

$$P(O, X \mid M) = o_{12} b_2(o_1) a_{22} b_2(o_2) a_{23} b_3(o_3) \ldots \tag{2.31}$$

แต่อย่างไรก็ตามค่า X เราไม่สามารถทราบได้ว่าจริงๆแล้ว แบบจำลองฮิดเดนมาร์คอฟมีการเปลี่ยนสถานะอย่างไร (ทำให้เราเรียกแบบจำลองนี้ว่า ฮิดเดน) ดังนั้น จึงมีวิธีจัดการได้ 2 วิธีคือ

1. หาผลรวมของวิธีทั้งหมดที่ X จะเป็นไปได้ คำนวณดังสมการที่ 2.32

$$P(O \mid M) = \sum_{X} a_{x(0)x(1)} \prod_{t=1}^{T} b_{x(t)}(o_t) a_{x(t)x(t+1)} \tag{2.32}$$

2. หาค่าที่น่าจะเป็นมากที่สุดของการเปลี่ยนสถานะดังสมการที่ 2.33

$$\hat{P}(O \mid M) = \arg\max \left[a_{x(0)x(1)} \prod_{t=1}^{T} b_{x(t)}(o_t) a_{x(t)x(t+1)} \right] \tag{2.33}$$

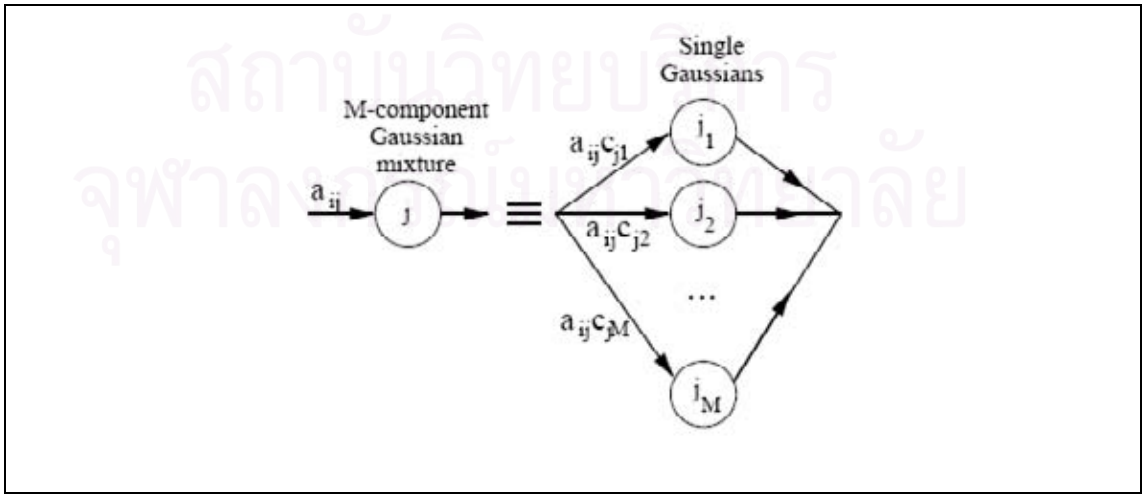
จากวิธีทั้งสองถ้ามีแบบจำลอง (M_i) ให้กับทุกๆ w_i จะได้ว่า

$$P(O \mid w_i) = P(O \mid M_i) \tag{2.34}$$

2.1.7 **การประมาณซ้ำแบบบอมเวลช์** (Baum-Welch Re-Estimation) [23]

ในการที่จะหาค่าของพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟจะต้องทำการเดาหาค่าใกล้เคียงก่อนแล้วค่าที่ถูกต้องมากขึ้นสามารถหาได้จากสูตรของบอมเวลช์

เนื่องจากการแบ่งกระแสของข้อมูล (Data Stream) ไม่ได้ส่งผลต่อการประมาณ เพราะว่า สายของข้อมูลแต่ละอันถูกมองเหมือนสถิต (Static) และยิ่งไปกว่านั้นส่วนประกอบส่วนผสมเกาส์เซียน (Mixture Component) ก็สามารถที่จะคิดในรูปแบบของ สถานะย่อยๆได้ โดยความน่าจะเป็นในการเปลี่ยนสถานะคือ น้ำหนักส่วนผสมเกาส์เซียน (Mixture Weight) ดังรูปที่ 7



รูปที่ 7 ส่วนผสมเกาส์เซียน (รูปจาก [23])

ปัญหาหลักของเราคือการที่จะประมาณซ้ำ (Re-estimate) ค่ากลาง (Mean) และค่าความแปรปรวน (Variance) ของแบบจำลองฮิดเดนมาร์คอฟโดยการแจกแจงของเอาพุต (Output Distribution) ในแต่ละสถานะซึ่งก็คือ ส่วนประกอบเกาส์เซียนเชิงเดี่ยว (Single Component Gaussian)

$$b_j(o_t) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp(-\tfrac{1}{2}(o_t - \mu_j)'\Sigma_j^{-1}(o_t - \mu_j)) \qquad (2.35)$$

ถ้าหากแบบจำลองฮิดเดนมาร์คอฟมีเพียงสถานะเดียวการประมาณค่าพารามิเตอร์ก็สามารถทำได้ $\hat{\mu}_j$ และ $\hat{\Sigma}_j$ จะสามารถคำนวณได้โดย

$$\hat{\mu}_j = \frac{1}{T}\sum_{t=1}^{T} o_t \qquad (2.36)$$

และ

$$\hat{\Sigma}_j = \frac{1}{T}\sum_{t=1}^{T} (o_t - \mu_j)(o_t - \hat{\mu}_j)' \qquad (2.37)$$

แต่ในความเป็นจริงแบบจำลองฮิดเดนมาร์คอฟมีหลายสถานะและไม่มีการกำหนดเวคเตอร์สังเกตว่าตัวใดถูกสร้างขึ้นมาจากสถานะใดเนื่องจากเราไม่ทราบลำดับการเปลี่ยนแปลงสถานะ แต่อย่างไรก็ดีถ้าเรามีการกำหนดว่าเวคเตอร์สังเกตตัวใดถูกสร้างมาจากสถานะใดก็สามารถที่จะคำนวณหาค่าของ μ_j และ Σ_j ได้เพื่อที่จะใช้เป็นค่าเริ่มต้นของตัวแปร จากนั้นเราจะใช้ตัวแปรที่ได้มานี้หาลำดับของการเปลี่ยนแปลงสถานะที่น่าจะเป็นมากสุดโดยใช้ขั้นตอนวิธีวิเทอร์บิ (Viterbi Algorithm) เมื่อหาได้ก็ทำการกำหนดค่าที่ได้ให้ μ_j และ Σ_j แล้วก็หาค่าของ μ_j และ Σ_j ใหม่จนกระทั่งเราพอใจหรือค่าไม่เปลี่ยนแปลง

เนื่องจากความเป็นไปได้ของแต่ละ ลำดับการสังเกตถูกคิดมาจากการรวมของความน่าจะเป็นทั้งหมดที่เป็นไปได้ของลำดับการเปลี่ยนสถานะ ดังนั้นแทนที่เราจะกำหนดให้เวคเตอร์สังเกต (o) ถูกสร้างออกมาจากสถานะหนึ่งๆ จึงเปลี่ยนเป็นเวคเตอร์สังเกต (o_j) แต่ละตัวจะถูกจัดให้ถูกสร้างมาจากทุกสถานะ โดยถ่วงน้ำหนักด้วยความน่าจะเป็นที่เวคเตอร์สังเกตนั้นถูกสังเกตที่สถานะและเวลานั้น ให้ $L_j(t)$ คือ ความน่าจะเป็นที่สถานะ j และเวลา t ดังนั้นสมการ (2.36) และ (2.37) จึงกลายเป็น

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} L_j(t) o_t}{\sum_{t=1}^{T} L_j(t)} \tag{2.38}$$

และ

$$\Sigma_j = \frac{\sum_{t=1}^{T} L_j(t)(o_j - \mu_j)(o_t - \hat{\mu}_j)'}{\sum_{t=1}^{T} L_j(t)} \tag{2.39}$$

ทั้งสองสมการนี้คือ สูตรการหาค่าประมาณของบอมเวลช์เพื่อหาค่ากลางและความน่าจะเป็นของแบบจำลองฮิดเดนมาร์คอฟซึ่งความน่าจะเป็นในการเปลี่ยนสถานะก็สามารถทำได้ในทำนองเดียวกัน

อีกปัญหาหนึ่งที่เราพบต่อมาคือ ต้องคำนวณหาค่า $L_j(t)$ ซึ่งสามารถทำได้โดยขั้นตอนวิธีที่เรียกว่า ขั้นตอนวิธีแบบไปข้างหน้าและแบบกลับหลัง (Forward - Backward Algorithm) โดยความน่าจะเป็นแบบไปข้างหน้า (Forward Probability) สำหรับแบบจำลอง M ซึ่งมี N สถานะ สามารถเขียนได้เป็น

$$\alpha_j(t) = P(o_1, ..., o_t, x(t) = j \mid M) \tag{2.40}$$

ซึ่งก็คือความน่าจะเป็นร่วม (Joint Probability) ของการสังเกตเวคเตอร์ในเวลา t แรกและมาหยุดที่สถานะ j ดังนั้น ความน่าจะเป็นไปข้างหน้าจึงสามารถคำนวณได้ด้วยการวนซ้ำตามสมการที่ 2.41

$$\alpha_j(t) = \left[\sum_{i=2}^{N-1} \alpha_j(t-1) a_{ij} \right] b_j(o_t) \tag{2.41}$$

เนื่องจากสถานะแรกและสถานะสุดท้ายเป็นสถานะที่เรียกว่า สถานะไม่ สร้างผลการสังเกต (Non-emitting State) จึงได้ว่า

$$\alpha_1(1) = 1 \tag{2.42a}$$

$$\alpha_j(1) = \alpha_1 b_j(o_1) \tag{2.42b}$$

สำหรับทุกๆ สถานะ i ซึ่ง 1 < i < N เงื่อนไขสุดท้าย คือ

$$\alpha_N(T) = \sum_{i=2}^{N-1} \alpha_i(T) a_{iN} \tag{2.43}$$

และจากคำนิยามของความน่าจะเป็นไปข้างหน้า สามารถเขียนเป็นสมการได้ว่า

$$P(O \mid M) = \alpha_N(T) \tag{2.44}$$

จะเห็นว่า ผลการคำนวณหาความน่าจะเป็นไปแบบข้างหน้าก็จะให้คำตอบเป็น $P(O \mid M)$ และความน่าจะเป็นแบบกลับหลัง (Backward Probability) สามารถนิยามได้เป็น

$$\beta_i(t) = P(o_{t+1},...,o_T, x(t) = j \mid M) \tag{2.45}$$

ซึ่งก็คล้ายกับความน่าจะเป็นแบบไปข้างหน้า ความน่าจะเป็นแบบกลับหลังสามารถคำนวณได้ด้วยการวนซ้ำดังนี้

$$\beta_i(t) = \sum_{j=2}^{N-1} a_{ij} b_j(o_{t+1}) \beta_j(t+1) \tag{2.46}$$

เมื่อเงื่อนไขเริ่มต้นเป็นตามสมการที่ 2.47

$$\beta_i(t) = a_{iN} \tag{2.47}$$

สำหรับทุกๆ สถานะ i ซึ่ง 1 < i < N เงื่อนไขสุดท้ายก็คือ

$$\beta_1(1) = \sum_{j=2}^{N-1} a_{1j} b_j(o_1) \beta_j(1) \tag{2.48}$$

จากคำนิยามของความน่าจะเป็นแบบไปข้างหน้าและความน่าจะเป็นแบบกลับหลังซึ่งจะเห็นว่า ความน่าจะเป็นไปข้างหน้าก็คือความน่าจะเป็นร่วม ส่วนความน่าจะเป็นไปข้างหลังคือความน่าจะเป็นแบบมีเงื่อนไข หากเรานำทั้งสองตัวนี้มาคูณกันจะได้ว่า

$$\alpha_j(t)\beta_j(t) = P(O, x(t) = j \mid M) \tag{2.49}$$

ดังนั้น

$$L_j(t) = P(x(t) = j \mid O, M) \tag{2.50a}$$

$$= \frac{P(O, x(t) = j \mid M)}{P(O \mid M)} \tag{2.50b}$$

$$= \frac{\alpha_j(t)\beta_j(t)}{P(O \mid M)} \tag{2.50c}$$

2.1.8 **การรู้จำและการถอดรหัสวิเทอร์บิ** (Viterbi Decoding) [23]

วิธีการรู้จำเสียงสามารถทำได้โดยการใช้วิธีเดียวกับความน่าจะเป็นไปข้างหน้า เพียงแต่เปลี่ยนจากหาผลรวมเป็นการหาความเป็นไปได้มากสุดแต่ละแบบจำลองโดย

$$\phi_j(t) = \max_i \{\phi_i(t-1)a_{ij}\}b_j(o_t) \tag{2.51}$$

$\phi_j(t)$ คือ ความเป็นไปได้มากสุดของ o_1 ถึง o_t ซึ่งอยู่ที่ สถานะ j ในเวลา t และ

$$\phi_1(1) = 1 \tag{2.52a}$$

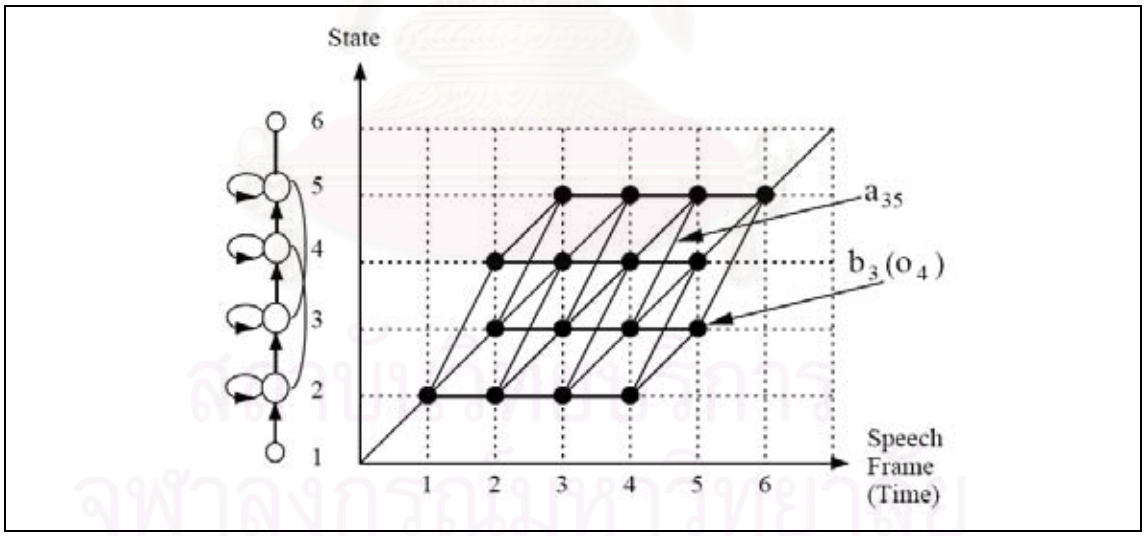
$$\phi_j(1) = a_{1j}b_j(o_1) \tag{2.52b}$$

สำหรับ 1 < j < N, ความเป็นไปได้มากสุดของ $P(O \mid M)$ จะเท่ากับ

$$\phi_N(T) = \max_i \{\phi_i(T)a_{iN}\} \tag{2.53}$$

ซึ่งการคำนวณเราควรใช้ฟังก์ชันลอการิธึมช่วยจึงได้เป็น

$$\psi_j(t) = \max_i \{\psi_i(t-1) + \log(a_{ij})\} + \log(b_j(o_t)) \tag{2.54}$$



รูปที่ 8 ขั้นตอนวิธีวิเทอร์บิสำหรับหน่วยเสียงย่อย (รูปจาก [23])

การวนซ้ำรูปแบบนี้เรียกว่า ขั้นตอนวิธีวิเทอร์บิ สามารถมองได้ในอีกมุม ดังรูปที่ 8 คือ การหาเส้นทางที่ดีที่สุดจากเมตริกซึ่งแนวตั้งเป็น สถานะของแบบจำลองฮิดเดนมาร์คอฟและแนวนอนเป็นเวลาของคำพูด โดยจุดต่างๆ แสดงถึง ความน่าจะเป็นในการสังเก ตกรอบที่เวลานั้นๆ และเส้นระหว่างจุดแสดงถึงความน่าเป็นในการเปลี่ยนสถานะ ซึ่งการคำนวณความน่าจะเป็นของ

เส้นทางสามารถทำได้โดยการรวมความน่าจะเป็นในการเปลี่ยนสถานะและความน่าจะเป็นใน การสร้างเอาพุตตามเส้นทางนั้นๆ รูปที่ 8 เครื่องจักรสถานะทางซ้ายสามารถแปลงให้อยู่ในรูปเมตริกทางด้านขวา ซึ่งการหาความน่าจะเป็นที่ดีที่สุดก็สามารถทำได้โดยหาเส้นทางที่ดีที่สุดตั้งแต่จุดกำเนิดไปจนถึงคู่ลำดับ (6,5) ซึ่งก็คือ สถานะที่ 5 และเวคเตอร์สังเกตที่ 6 ของเมตริกนั่นเอง

2.1.9 การลดจำนวนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ [24],[25]

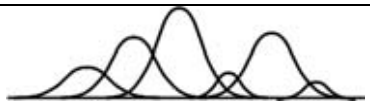
การรู้จำเสียงพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบบริบทไม่อิสระ (Context-dependent Hidden Markov Model) จะมีจำนวนพารามิเตอร์จำนวนมาก เช่น หากลองพิจารณาการสร้างเครื่องรู้จำโดยใช้ทอพอโลยีแบบ 3 สถานะที่มีการเชื่อมต่อของสถานะแบบซ้ายขวาซ้าย โดยแต่ละสถานะมี 4 ส่วนผสมเกาส์เซียนและใช้การสกัดลักษณะสำคัญที่มีพลังงาน (Energy) การเปลี่ยนแปลง (Delta) และความเร่ง (Delta-Delta) เป็นค่าลักษณะสำคัญ อย่างละ 13 มิติ รวมเป็น 39 มิติ โดยมีขนาดเฟรม (Frame size) เท่ากับ 10 มิลลิวินาที ดังนั้น สำหรับหนึ่งหน่วยเสียงจะมีพารามิเตอร์ เท่ากับ (3 x 2 x 39 + 2) x 3 = 708 สำหรับภาษาไทยโดยใช้หน่วยเสียงตามที่กำหนดในฐานข้อมูลเสียงโลตัสซึ่งมีจำนวนเท่ากับ 74 หน่วย รวมกับ sil (สัญลักษณ์แทนเสียงเงียบ) และ sp (สัญลักษณ์แทนการหยุดระยะสั้น) เมื่อพิจารณาไตรโฟน (Triphone) ที่เกิดขึ้นได้จริงจะมีประมาณ 100,000 ไตรโฟน ทำให้การสร้างเครื่องรู้จำเสียงพูดในภาษาไทยจำต้องมีพารามิเตอร์ประมาณ 70 ล้านพารามิเตอร์ และหากสมมุติว่า เราต้องการ 10 ตัวอย่างต่อหนึ่งมิติของฟีเจอร์ต่อส่วนผสมเกาส์เซียนเพื่อฝึกฝนแบบจำลองเสียงพูด ผลที่ได้คือ เราจำต้องใช้ฐานข้อมูลขนาด 38 ชั่วโมง ที่มีการแจกแจงของหน่วยเสียงอย่างสม่ำเสมอและจำต้องผ่านทุกสถานะของแบบจำลองฮิดเดนมาร์คอฟของทุกหน่วยเสียง ซึ่งในความเป็นจริงแล้ว ไตรโฟนบางตัวเกิดขึ้นน้อย บางตัวเกิดขึ้นบ่อย ดังนั้น เพื่อแก้ปัญหานี้ จึงจำเป็นต้องลดจำนวนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟลง เพื่อให้ฐานข้อมูลเสียงที่ใช้ฝึกฝนมีขนาดเล็กลง ในที่ นี้จะนำเสนอ 3 วิธี คือ การเลือกใช้แบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่อง (Semi-continuous Hidden Markov Model: ScHMM) การผูกสถานะ (State Tying) และการโคลนสถานะ (State Cloning)

2.1.9.1 การเลือกใช้แบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่อง

แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง (Continuous Hidden Markov Model: CHMM) และแบบจำลองฮิดเดนมาร์คอฟแบบวิตยุต (Discrete Hidden Markov Model: DHMM) มีข้อดีและคุณสมบัติที่แตกต่างกัน ดังแสดงในตารางที่ 1 และแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องมีข้อดี คือ ไม่มีการทำการแบ่งนับ (Quantization) ทำให้มีความแม่นยำมากกว่าแบบจำลองฮิดเดนมาร์คอฟแบบวิตยุต เมื่อแบบจำลองได้รับการฝึกฝนมาเป็นอย่างดี แต่

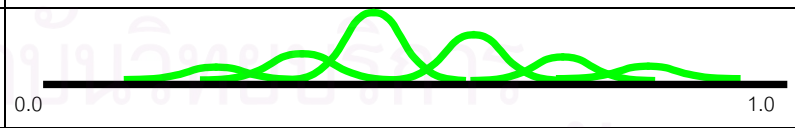
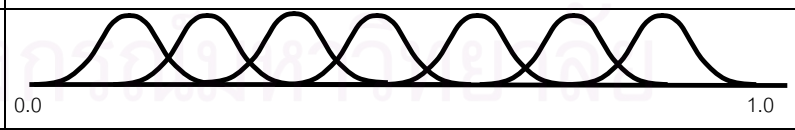
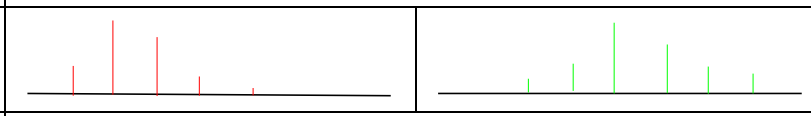
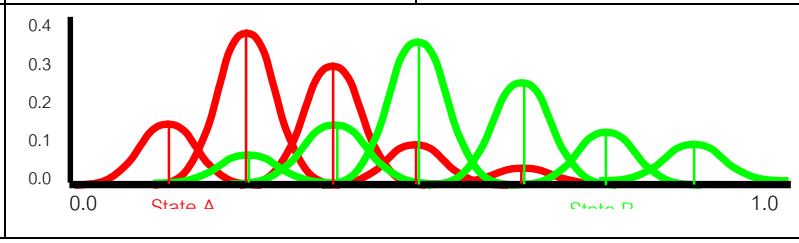
แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องก็มีข้อเสียคือ ใช้ทรัพยากรในการคำนวณและมีจำนวนพารามิเตอร์มาก

ตารางที่ 1 การเปรียบเทียบระหว่าง CHMM และ DHMM

	CHMM	DHMM
ฟังก์ชันความหนาแน่นของความน่าจะเป็น		
ความแม่นยำ	สูง	ต่ำ
ความเร็วในการคำนวณ	ต่ำ	สูง
จำนวนพารามิเตอร์	มาก	น้อย

แบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องหรือการผูกส่วนผสมเกาส์เซียน (Tie Mixture) เป็นแบบจำลองฮิดเดนมาร์คอฟที่รวมข้อดีของแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องและแบบจำลองฮิดเดนมาร์คอฟแบบวิตยุตเข้าด้วยกัน โดยยังคงใช้ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function) ที่เป็นแบบต่อเนื่องอยู่ แต่ว่ามีจำนวนพารามิเตอร์น้อยกว่าแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องมาก เนื่องจากสิ่งที่แตกต่างออกไประหว่างแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องกับแบบจำลองฮิดเดนมาร์คอฟแบบอื่นๆ คือทุกสถานะจะมีน้ำหนักส่วนผสมเกาส์เซียน (Mixing Weight) เป็นของตัวเองแต่ใช้ส่วนผสมเกาส์เซียนร่วมกัน ดังแสดงในตารางที่ 2

ตารางที่ 2 ความแตกต่างระหว่าง CHMM และ ScHMM

CHMM	State A	
	State B	
ScHMM	All states	
	Mixing Weight	
	GMM	

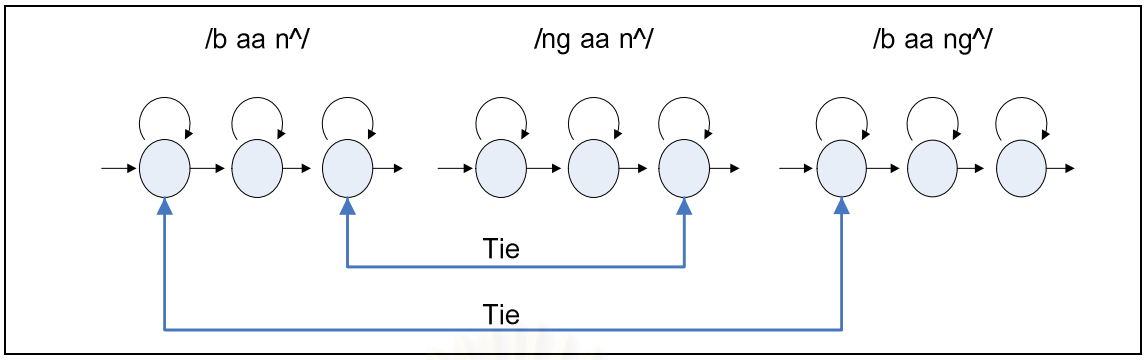
จากตารางที่ 2 จะเห็นว่าสำหรับแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องแต่ละสถานะจะมีแบบจำลองส่วนผสมเกาส์เซียน (Gaussian Mixture Model: GMM) เป็นของตัวเอง เมื่อมีผลการสังเกตเข้ามา แต่ละสถานะก็จะคำนวณ ค่าความน่าจะเป็นของผลการสังเกตที่สถานะนั้นๆ (Probability of Observation Data given State) หรือค่า $b_j(O_t)$ จากแบบจำลองส่วนผสมเกาส์เซียนของสถานะนั้นๆ แต่สำหรับแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องจะมีองค์ประกอบเกาส์เซียน (Gaussian Component) เพียง 1 ตัวเท่านั้น นั่นคือ ทุกสถานะจะมีค่ากลางและความแปรปรวนค่าเดียวกันโดยการคำนวณค่า $b_j(O_t)$ แต่ละสถานะจะนำค่าน้ำหนักส่วนผสมเกาส์เซียนของตัวเองมาคูณกับองค์ประกอบเกาส์เซียนนี้ก่อน เพื่อสร้างองค์ประกอบเกาส์เซียนของตัวเองแล้วจึงคำนวณตามปกติ

เครื่องมือที่นำแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องไปประยุกต์เพื่อสร้างเครื่องรู้จำเสียงที่เป็นที่รู้จักกันอย่างกว้างขวางคือ Sphinx [26] ซึ่งถูกพัฒนาโดยมหาวิทยาลัยคาร์เนกีเมลลอน (Carnegie Mellon University) และหากพิจารณาจำนวนพารามิเตอร์ที่ใช้ในการสร้างเครื่องรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องโดยมีทอพอโลยีและวิธีสกัดค่าลักษณะเหมือนกับแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องจะพบว่า มีจำนวนพารามิเตอร์น้อยกว่ามาก คือ หากสมมุติว่าองค์ประกอบเกาส์เซียนมีจำนวนส่วนผสมเกาส์เซียน เท่ากับ 256 และหากเราต้องการใช้ส่วนผสมเกาส์เซียนเป็น 2 สำหรับแต่ละสถานะจำนวนพารามิเตอร์จะเท่ากับ (39 x 2 x 256) + (100,000 x 2 x 2) = 462768 เมื่อ 100,000 คือ จำนวนไตรโฟนและในการเก็บน้ำหนักส่วนผสมเกาส์เซียนนั้นต้องเก็บด้วยว่าสถานะใดไม่เป็นศูนย์และมีค่าเป็นเท่าใดจึงทำให้ต้องใช้พารามิเตอร์ในส่วนนี้เป็น 2 เท่าของจำนวนส่วนผสมเกาส์เซียน

เมื่อเปรียบเทียบกับแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องพบว่าแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องใช้พารามิเตอร์ไม่ถึง 1% ของแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องซึ่งแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องน่าจะมีความแม่นยำดีกว่า ในกรณีที่ตัวอย่างเสียงมีจำกัด โดยสรุปแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องมีข้อดีดังนี้

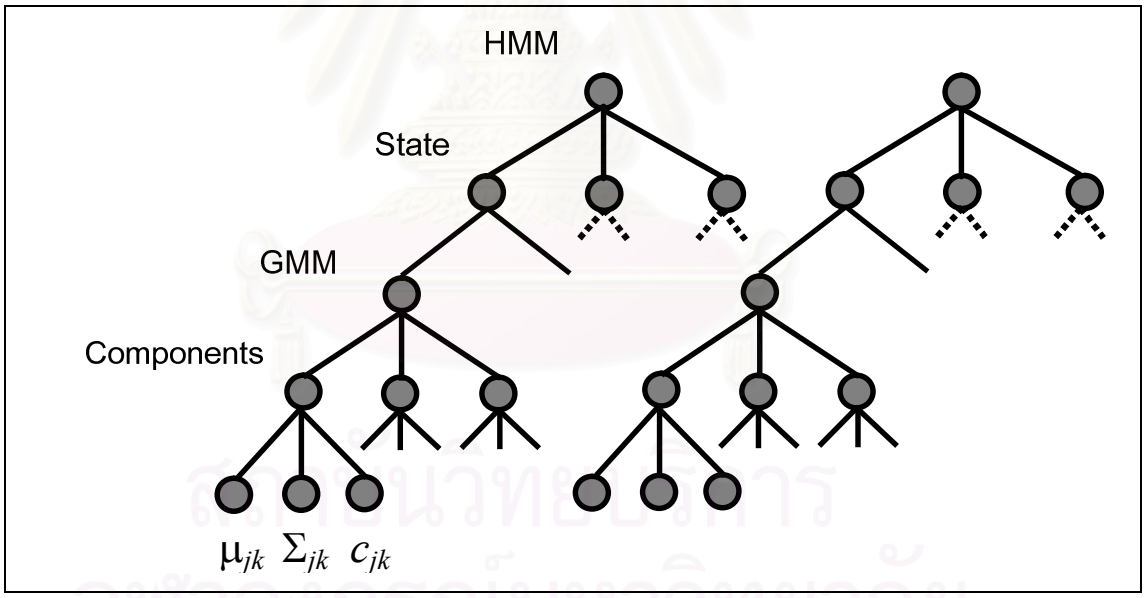
1. สูญเสียข้อมูลน้อยกว่าแบบจำลองฮิดเดนมาร์คอฟแบบวิตยุต
2. ลดจำนวนพารามิเตอร์ได้เป็นจำนวนมากเมื่อเปรียบเทียบกับแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องดังที่ได้ยกตัวอย่างไป
3. ลดจำนวนข้อมูลฝึกฝน
4. เนื่องจากมีจำนวนพารามิเตอร์น้อยกว่าแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องทำให้สามารถคำนวณได้รวดเร็วทำให้สามารถขยายขนาดลำค้นหา (Search Beam) เมื่อทำการรู้จำได้

2.1.9.2 การผูกสถานะ



รูปที่ 9 ตัวอย่างการผูกสถานะ

การผูกสถานะหรือการจัดกลุ่มสถานะ (State Clustering) เป็นอีกขั้นตอนวิธีหนึ่งที่ช่วยลดจำนวนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องโดยมีแนวคิดว่า หาก 2 สถานะใดๆ มีแบบจำลองหน่วยเสียงที่มีลักษณะคล้ายกัน กล่าวคือ มีพารามิเตอร์ของแบบจำลองส่วนผสมเกาส์เซียนคล้ายกัน เพื่อลดจำนวนพารามิเตอร์ก็ให้ 2 สถานะนั้นๆ ยุบรวมกันเป็นสถานะเดียวและใช้พารามิเตอร์ร่วมกัน ดังเช่น แสดงในรูปที่ 9



รูปที่ 10 ระดับการผูก

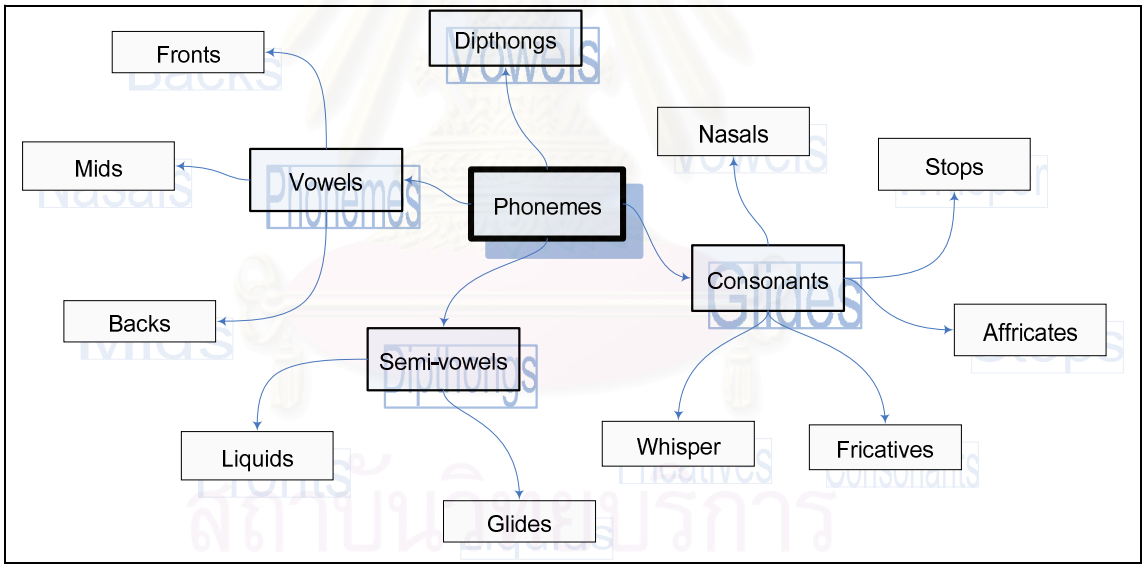
พารามิเตอร์ที่ของสถานะที่คล้ายกัน เมื่อถูกผูกเข้าด้วยกันก็จะกลายเป็นพารามิเตอร์ที่เหมือนกัน ส่งผลให้ความสามารถในการแยกแยะหน่วยเสียงลดลง แต่จำนวนพารามิเตอร์ก็ลดลงเช่นกัน ซึ่งโดยทั่วไปแล้วสามารถผูกสถานะใดๆ เข้าด้วยกันได้มากกว่า 2 สถานะ แต่อย่างไรก็ดี การผูกสถานะนั้นทำให้เกิดคำถามว่า เราควรผูกสถานะใดเข้าด้วยกัน เมื่อมีจำนวนสถานะที่คล้ายกันมากกว่า 2 สถานะ และหากผูกสถานะเข้าด้วยกันไปแล้วแบบจำลองฮิด

เดนมาร์คอฟจะมีประสิทธิภาพดีขึ้น เนื่องจากจำนวนพารามิเตอร์ลดลง เมื่อจำนวนพารามิเตอร์ลดลงแต่ว่าเรามีข้อมูลที่ใช้สำหรับฝึกฝนเท่าเดิมจึงทำให้มีตัวอย่างสำหรับการฝึกฝนต่อจำนวนพารามิเตอร์มากขึ้นหรือมีประสิทธิภาพลดลง เพราะว่าการผูกสถานะทำให้ความสามารถในการแยกแยะหน่วยเสียงลดลง โดยการผูกสถานะสามารถทำได้หลายระดับดังรูปที่ 10

การแบ่งกลุ่ม (Clustering) คือ การรวมสถานะของแบบจำลองฮิดเดนมาร์คอฟเข้าด้วยกันและเรียกขั้นตอนวิธีการในการตัดสินใจเลือกคู่ของสถานะหรือกลุ่มใดๆ เพื่อผูกเข้าด้วยกันว่า ขั้นตอนวิธีการแบ่งกลุ่ม (Clustering Algorithm) โดยทั่วไปแล้วนิยมผูกในระดับสถานะหรือองค์ประกอบส่วนผสมเกาส์เซียนซึ่งมีทั้งหมด 3 วิธี ดังนี้

2.1.9.2.1 การแบ่งกลุ่มแบบฐานความรู้ (Knowledge-based Clustering)

การผูกสถานะแบบการแบ่งกลุ่มแบบฐานความรู้เป็นการใช้ความรู้ทางเสียงเพื่อกำหนดว่า สถานะใดควรผูกกับสถานะใด เนื่องจากเราทราบว่า หน่วยเสียงในสามารถแบ่งออกเป็นกลุ่มๆ ได้ดังรูปที่ 11



รูปที่ 11 การจำแนกหน่วยเสียง

การใช้ความรู้ทางเสียงจะช่วยให้เราสามารถกำหนดสถานะที่จะผูกกันได้ ยกตัวอย่าง เช่น ไตรโฟน /b-aa+n^/ และ /t-aa+n^/ สามารถผูกทุกสถานะเข้าด้วยกันได้ เพราะว่า /b/ และ /t/ มีคุณลักษณะทางเสียงคล้ายกัน คือ เป็นหน่วยเสียงจำพวก พยัญชนะหยุดเหมือนกัน ซึ่งเมื่อผูกเข้าด้วยกันแล้วก็จะทำให้มีตัวอย่างของหน่วยเสียงสำหรับแบบจำลองใหม่ที่เกิดขึ้นจากการผูก /b-aa+n^/ และ /t-aa+n^/ เข้าด้วยกัน ในกรณีที่ตัวอย่างเสียงของทั้งสองไตรโฟนนี้มีน้อยการผูกจะช่วยเพิ่มจำนวนตัวอย่างเสียงต่อแบบจำลองได้ และเมื่อพิจารณาไตรโฟน /s-

i+p^/ และ /s-i+t^/ เราสามารถที่จะผูก 2 ใน 3 ของสถานะของทั้งสองไตรโฟนนี้เข้าด้วยกันได้ เพราะว่า ทั้งสองไตรโฟนมีคุณลักษณะทางเสียงช่วงแรก คือ /s/ และ /i/ ที่น่าจะคล้ายกันมาก เป็นต้น

2.1.9.2.2 การแบ่งกลุ่มแบบข้อมูลหน่วยขับ (Data-driven Clustering)

การผูกสถานะแบบการแบ่งกลุ่มแบบข้อมูลหน่วยขับเป็นการเปรียบเทียบความคล้ายของไตรโฟนด้วยเมตริกระยะทาง (Distance Matrix) เพื่อรวมสถานะที่มีระยะทางใกล้กันเข้าด้วยกัน ขั้นตอนวิธีการแบ่งกลุ่มแบบข้อมูลหน่วยขับจำต้องกำหนดสิ่งต่างๆ ต่อไปนี้ก่อน เพื่อเป็นพารามิเตอร์เริ่มต้น

1. ให้ทุกสถานะมีกลุ่มเป็นของตัวเอง นั่นคือ จำนวนกลุ่มเริ่มต้นจะเท่ากับจำนวนสถานะทั้งหมด
2. เลือกใช้วิธีการวัดเมตริกระยะทาง เช่น เมื่อ D คือ มิติของปริภูมิฟีเจอร์ (Feature space) i คือ จุดข้อมูล (Data Point) ที่อยู่ในเซตข้อมูลฝึกฝน I และ x กับ y คือ กลุ่มข้อมูล
   a. ระยะทางแบบยูคลิด (Euclidean Distance) คำนวณได้ดังสมการ (2.55)

$$UD = \left[\sum_{d=1}^{D} (\mu_{x_d} - \mu_{y_d})^2 \right]^{\frac{1}{2}} \quad (2.55)$$

   b. ระยะทางแบบมาฮาลาโนบิส (Mahalanobis Distance) หรือระยะทางแบบยูคลิดถ่วงน้ำหนัก (Weighted Euclidean Distance) ดังแสดง (2.56)

$$MD = \left[\sum_{d=1}^{D} \frac{(\mu_{x_d} - \mu_{y_d})^2}{\sigma_{x_d}\sigma_{y_d}} \right]^{\frac{1}{2}} \quad (2.56)$$

   c. ระยะทางสมมาตรของ Kullback-Liebler (Symmetric Kullback-Liebler Distance) สามารถคำนวณได้จากสมการ (2.57)

$$SKLD = \sum_{i \in I} p(i \mid x) \log \frac{p(i \mid x)}{p(i \mid y)} + \sum_{i \in I} p(i \mid y) \log \frac{p(i \mid y)}{p(i \mid x)} \quad (2.57)$$

3. มาตรวัดขนาดของกลุ่ม (Measure of cluster size) ซึ่งค่านี้จะกำหนดระยะทางมากที่สุดระหว่าง X และ Y ในกลุ่ม

4. ขีดแบ่ง (Threshold) สำหรับขนาดของกลุ่มที่ใหญ่ที่สุด (Largest Cluster Size) และจำนวนกลุ่มน้อยสุด

เมื่อกำหนดค่าพารามิเตอร์ต่างๆ ตามที่กล่าวไปแล้ว ขั้นตอนการแบ่งกลุ่มแบบข้อมูลหน่วยขับสามารถทำได้ดังแสดงในรูปที่ 12

ขั้นตอนวิธีการแบ่งกลุ่มแบบข้อมูลหน่วยขับ

1. หาคู่ของกลุ่ม A และ B ที่มีระยะทางน้อยสุด เมื่อ A และ B คือ กลุ่มใดๆ ที่แตกต่างกัน
2. รวม A และ B เป็นกลุ่มเดียวกัน
3. ผูกทุกสถานะที่อยู่ในกลุ่ม A กับทุกสถานะที่อยู่ในกลุ่ม B
4. วนซ้ำขั้นตอนที่ 1 – 3 จนกระทั่งถึงขีดแบ่ง (Threshold)

รูปที่ 12 ขั้นตอนวิธีการแบ่งกลุ่มแบบข้อมูลหน่วยขับ

เพื่อให้เห็นภาพชัดเจนมายิ่งขึ้น พิจารณาปริภูมิฟีเจอร์ขนาด 1 มิติ และเลือกใช้ระยะทางแบบมาฮาลาโนบิสเป็นเมตริกระยะทางโดยที่ M_{xy} คือ ระยะทางระหว่างกลุ่ม x กับกลุ่ม y โดยกำหนดให้ ในตอนแรกมีกลุ่มข้อมูลทั้งหมด 4 กลุ่มดังแสดงในตารางที่ 3

ตารางที่ 3 ตัวอย่างกลุ่มข้อมูล 4 กลุ่มเริ่มต้น

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
จุดข้อมูล	0.10 0.15 0.05 0.10	0.40 0.45 0.30 0.40	0.65 0.60 0.50 0.70	0.95 0.80 0.85 0.95
ค่ากลาง	0.10	0.36	0.61	0.89
ส่วนเบี่ยงเบนมาตรฐาน	0.0408	0.075	0.0854	0.075
เมตริกระยะทาง	M_{11} = 0.0	M_{12} = 4.7 M_{22} = 0.0	M_{13} = 8.64 M_{23} = 3.12 M_{33} = 0.00	M_{14} = 14.28 M_{24} = 7.07 M_{34} = 3.50 M_{44} = 0.00

จากตารางที่ 3 จะเห็นว่า M_{23} มีค่าน้อยที่สุด ดังนั้น จึงยุบรวม Cluster 2 และ Cluster 3 เข้าด้วยกัน ทำให้ในรอบต่อไปกลุ่มข้อมูลที่ต้องพิจารณาจึงเหลือเพียง 3 กลุ่มดังแสดงในตารางที่ 4

ตารางที่ 4ตัวอย่างในรอบที่ 2

	Cluster 1	Cluster 23	Cluster 4
จุดข้อมูล	0.10 0.15 0.05 0.10	0.40, 0.65 0.45, 0.60 0.30, 0.50 0.40, 0.70	0.65 0.60 0.50 0.70
ค่ากลาง	0.10	0.49	0.61
ส่วนเบี่ยงเบนมาตรฐาน	0.0408	0.1529	0.0854
เมตริกระยะทาง	$M_{1\,1} = 0.0$	$M_{1\,23} = 4.94$ $M_{23\,23} = 0.0$	$M_{1\,4} = 14.28$ $M_{23\,4} = 3.73$ $M_{4\,4} = 0.00$

ดังนั้นในรอบที่ 2 นี้ cluster 23 จะถูกรวมกับ cluster 4 เป็นต้น

2.1.9.2.3 การแบ่งกลุ่มแบบต้นไม้ตัดสินใจ (Decision-tree Clustering)

การแบ่งกลุ่มแบบต้นไม้ตัดสินใจ คือ การนำเอาการแบ่งกลุ่มแบบฐานความรู้รวมกับการแบ่งกลุ่มแบบข้อมูลหน่วยขับซึ่งการทำแบบนี้มีข้อดีคือ เมื่ อเราไม่มีตัวอย่างสำหรับสถานะของไตรโฟนแบบบริบทไม่อิสระที่กำหนด สถานะนี้ก็ควรที่จะถูกรวมเข้าไปกับสถานะในกลุ่มที่มีลักษณะทางเสียงที่คล้ายกัน ซึ่งการตัดสินใจว่าจะรวมกับ สถานะใดนั้น เราไม่สามารถใช้วิธีข้อมูลหน่วยขับได้เพราะว่า เราไม่มีตัวอย่างเสียง ดังนั้นจึงจำต้ องใช้ความรู้ทางเสียงเพื่อตัดสินใจว่า ไตรโฟนเหล่านั้นควรจะรวมกับกลุ่มใดและสำหรับไตรโฟนที่มีตัวอย่างเสียงแล้วการตัดสินใจรวมสถานะด้วยวิธีข้อมูลหน่วยขับก็ถือว่าเป็นวิธีที่ยอมรับได้ เพราะว่า วิธีข้อมูลหน่วยขับวัดความคล้ายด้วยเมตริกระยะทางโดยคิดคำนวณจากปริภูมิฟีเจอร์ในแต่ละมิติ ซึ่งค่า เมตริกระยะทางนี้สามารถบ่งชี้ถึงความคล้ายกันทางสัญญาณได้

ในการที่จะสร้างต้นไม้ตัดสินใจ จำเป็นต้องมีองค์ประกอบดังต่อไปนี้

1. เซตของคำถามเพื่อการตัดสินใจ นั่นคือ แนวทางที่ข้อมูลจะถูกแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มที่ตอบว่า ‘ใช่’ และกลุ่มที่ตอบว่า ‘ไม่ใช่’ โดยใช้ฐานความรู้
2. เกณฑ์การวัดความดีของการแบ่งกลุ่ม เมื่อข้อมูลถูกแบ่งออกเป็น 2 ส่วนด้วยเซตของคำถามใช่-ไม่ใช่แล้วจำต้องมีเกณฑ์เพื่อตัดสินว่า การแบ่งกลุ่มด้วยคำถามใดให้ผลดีที่สุดและดีกว่าจากเดิมที่อยู่กลุ่มเดียวกันหรือไม่
3. เงื่อนไขในการหยุดโปรแกรม

เมื่อมีองค์ประกอบครบทั้ง 3 อย่างแล้วการแบ่งกลุ่มแบบต้นไม้ตัดสินใจสามารถทำได้ดังนี้

ขั้นตอนวิธีการแบ่งกลุ่มแบบต้นไม้ตัดสินใจ

1. ให้ทุกสถานะอยู่ที่รากของต้นไม้
2. แบ่งต้นไม้ออกเป็นสองส่วนโดยใช้คำถามแบบ 'ใช่' 'ไม่ใช่' จากเซตของคำถามโดยเลือกแบ่งตามคำถามที่ให้ค่าความเป็นไปได้ (Likelihood) กับข้อมูลมากที่สุด
3. แบ่งข้อมูลเป็น 2 ส่วน คือ ส่วนที่ตอบว่า 'ใช่' และส่วนที่ตอบว่า 'ไม่ใช่'
4. สำหรับทั้งสองกลุ่มใหม่ กลับไปขั้นตอนที่ 2
5. สำหรับทุกใบของต้นไม้ คำนวณค่าความเป็นไปได้เมื่อนำใบ 2 ใบใดๆ รวมเข้าด้วยกัน หากค่าความเป็นไปได้ลดลงน้อยกว่าขีดแบ่งที่กำหนด ก็ให้ทำการยุบรวมใบ 2 ใบนั้นๆ

รูปที่ 13 ขั้นตอนวิธีต้นไม้ตัดสินใจ

การคำนวณค่าคาดหวังของความเป็นไปได้ (L) ของใบ S เมื่อ S คือ เซตของสถานะที่ถูกผูกเข้าด้วยกันและเป็นส่วนประกอบเกาส์เซียนเชิงเดี่ยว ที่มีค่ากลางคือ μ และมีเมตริกความแปรปรวนร่วมเกี่ยว คือ $\sum$ หากให้ $O = \{O_1, O_2, \cdots, O_T\}$ คือ ลำดับการสังเกต L จะมีค่าเท่ากับ

$$L = \sum_{t=1}^{T} [\sum_{s \in S} \log(p(O_t \mid \mu, \Sigma)) \cdot \gamma_t(s)] \tag{2.58}$$

โดยที่ s คือ สถานะที่เป็นสมาชิกในเซตของใบ S และ $\gamma_t(s)$ คือความน่าจะเป็นที่จะอยู่ที่สถานะ s ณ เวลา t หากต้องการทราบความน่าจะเป็นที่ จะอยู่ที่ s ซึ่งก็คือ ความน่าจะเป็นที่จะอยู่ที่สถานะผูก s ณ เวลา t ก็สามารถคำนวณได้โดยคำนวณ ผลรวมทั้งหมดของ $\gamma_t(s)$ เมื่อ $s \in S$

สำหรับค่าลอการิธึมของปริภูมิฟีเจอร์ขนาด n มิติสามารถคำนวณได้โดย

$$\log p(O_t \mid \mu, \Sigma) = -\frac{1}{2}[n \log(2\pi) + \log(|\Sigma|) + (O_t - \mu)^{tr} \Sigma^{-1} (O_t - \mu)] \tag{2.59}$$

จากสมการ (2.58) และ (2.59) Zhao และคณะ [27] ได้พิสูจน์ว่า

$$\sum_{t=1}^{T} \left[(O_t - \mu)^{tr} \Sigma^{-1} (O_t - \mu) \cdot \sum_{s \in S} \gamma_t(s) \right] = n \sum_{t=1}^{T} \sum_{s \in S} \gamma_t(s) \tag{2.60}$$

ดังนั้นค่าลอการิธึมของความเป็นไปได้จึงเขียนใหม่ได้เป็นดังสมการ (2.61)

$$L = -\frac{1}{2}[n(\log(2\pi)+1)+\log(|\sum|)]\cdot\sum_{s\in St=1}^{T}\gamma_t(s) \quad (2.61)$$

และเมตริกความแปรปรวนร่วมเกี่ยวของสถานะผูกก็สามารถคำนวณได้โดย

$$\sum = E(x^2) - E(x)^2 \quad (2.62)$$

โดยที่

$$E(x^2) = \frac{\sum_{s\in S}\left[\left(\sum_{t=1}^{T}\gamma_t(s)\right)\cdot\left(\sum_s + \mu_s\mu_s^{tr}\right)\right]}{\sum_{s\in St=1}^{T}\gamma_t(s)} \quad (2.63a)$$

$$E(x)^2 = \left(\frac{\sum_{s\in St=1}^{T}\gamma_t(s)\mu_s}{\sum_{s\in St=1}^{T}\gamma_t(s)}\right)\cdot\left(\frac{\sum_{s\in St=1}^{T}\gamma_t(s)\mu_s}{\sum_{s\in St=1}^{T}\gamma_t(s)}\right)^{tr} \quad (2.64b)$$

เมื่อ μ_s และ $\sum_s$ คือ ค่ากลางและเมตริกความแปรปรวนร่วมเกี่ยวของสถานะ s

2.1.9.3 การโคลนสถานะ

แบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่องและการผูกสถานะเป็นวิธีที่มีจุดมุ่งหมายเพื่อให้สามารถฝึกฝนแบบจำลองเสียงพูดของแบบจำลองฮิดเดนมาร์คอฟได้อย่างมีประสิทธิภาพโดยการลดจำนวนพารามิเตอร์ลง การโคลนสถานะก็เป็นอีกวิธีหนึ่งที่มุ่งเน้นให้สามารถฝึกฝนแบบจำลองเสียงพูดที่มีประสิทธิภาพ แต่แทนที่จะลดจำนวนพารามิเตอร์เหมือน 2 วิธีแรก การโคลนสถานะใช้วิธีสร้างไตรโฟนจากหน่วยเสียงเดี่ยวเพื่อเพิ่มประสิทธิให้ค่าพารามิเตอร์เริ่มต้นของการฝึกฝนไตรโฟนดังนั้นการโคลนสถานะจึงสามารถสรุปได้ดังตารางที่ 5

ตารางที่ 5 ขั้นตอนการโคลนสถานะ

เริ่มต้น	เซตของหน่วยเสียงเดี่ยว (Monophone) ที่มีพารามิเตอร์ที่เหมาะสม
ขั้นตอนวิธี	1. โคลนหน่วยเสี่ยงเดี่ยวทั้งหมดเพื่อสร้างไตรโฟนที่ต้องการ โดยให้โคลนหน่วยเสียงเดี่ยวตามหน่วยเสียงที่อยู่ตรงกลางของไตรโฟน 2. ฝึกฝนไตรโฟนที่ได้ด้วยวิธีการฝึกฝนกำหนด

2.1.10 หน่วยเสียงในภาษาไทย [28]

หน่วยเสียงในภาษาไทยสามารถแบ่งได้ดังนี้

2.1.10.1 พยัญชนะ

พยัญชนะ หมายถึง การที่อากาศผ่านจากปอดขึ้นมายังกล่องเสียงแล้วตรงขึ้นมาปะทะกับอวัยวะต่างๆ ในช่องปากซึ่งอยู่ในระดับเหนือช่องของเส้นเสียง ทำให้ลมที่ผ่านขึ้นมานั้นพบกับอุปสรรคที่เกิดจากการทำงานประสานกันของอวัยวะในช่องปาก เช่น การกักลมปล่อยออกมาจากปอดไว้ชั่วครู่ แล้วผลักออกมาโดยแรง การบีบลมให้ผ่านช่องแคบๆ เป็ นต้น ก่อให้เกิดเสียงที่มีลักษณะเฉพาะตัวของเสียงพยัญชนะซึ่งแตกต่างจากเสียงสระ พยัญชนะในภาษาไทยสามารถแบ่งออกเป็นประเภทใหญ่ๆ ได้ 6 ประเภท คือ

2.1.10.1.1 **เสียงพยัญชนะหยุด**

เสียงพยัญชนะหยุด เสียงพยัญชนะประเภทนี้จะแบ่งเป็น 2 ลักษณะย่อยๆ คือ เสียงพยัญชนะระเบิด (Plosive) เกิดจากการที่ลมถูกกักเอาไว้ใน ณ ตำแหน่งใดตำแหน่งหนึ่งในช่องปาก แล้วจึงเปิดช่องนั้นเพื่อให้ลมพุ่งออกมาโดยแรงและเสียงพยัญชนะกัก (Unreleased Stop) มีลักษณะการเกิดในระยะแรกเหมือนกับเสียงพยัญชนะระเบิด แต่ว่าแทนที่ลมนั้นจะถูกปล่อยออกมากลับถูกกลืนเข้าไปใหม่หรือถูกแปรสภาพเป็นลมหายใจธรรมดาออกทางจมูก สำหรับในภาษาไทยนั้น เสียงพยัญชนะกักจะเป็นพยัญชนะท้ายหรือตัวสะกดเท่านั้นและก่อให้เกิดเสียงหยุดระยะสั้น (Short Pause) เช่น คำว่า “กับ” เป็นต้น หน่วยเสียงพยัญชนะหยุดมีทั้งหมด 11 หน่วยเสียง คือ /p/ /t/ /c/ /k/ /ʔ/ /p^h/ /t^h/ /c^h/ /k^h/ /b/ และ /d/

2.1.10.1.2 เสียงพยัญชนะเสียดแทรก

เสียงพยัญชนะเสียดแทรก คือ เกิดการจากที่ลมที่ผ่านขึ้นมาจากหลอดลมถูกบีบให้ผ่านช่องแคบๆ ณ ที่ใดที่หนึ่งในช่องปาก เช่น ริมฝีปากบนกับฟันล่าง ริมฝีปา กล่างกับฟันบน ริมฝีปากบนกับริมฝีปากล่างหรือ ลิ้นส่วนหลังกับลิ้นไก่ เป็นต้น การสร้างเสียงเสียดแทรกต่างจากการสร้างเสียงหยุดตรงที่ เราสามารถสร้างเสียงเสียดแทรกให้นานเท่าใดก็ได้จนสุดลมหายใจ แต่เสียงหยุดหรือเสียงระเบิดนั้นไม่สามารถที่จะถูกยืดให้ยาวออกไปได้ เสียงพยัญชนะเสียดแทรกนั้นมีทั้งแบบที่เป็นเสียงก้องและแบบไม่ก้อง แต่สำหรับในภาษาไทยจะมีเฉพาะเสียงพยัญชนะเสียดแทรกแบบไม่ก้องเท่านั้น ซึ่งจะมีทั้งหมด 3 หน่วยด้วยกัน คือ /f/ /s/ และ /h/

2.1.10.1.3 เสียงพยัญชนะนาสิก

เสียงพยัญชนะนาสิกหรือเสียงพยัญชนะขึ้นจมูก มีวิธีเปล่งเสียงคล้ายกับพยัญชนะกัก แต่เมื่อปล่อยให้ลมมากักอยู่ในช่องปากแล้วจะต้องลดเพดานและลิ้นไก่ลงเพื่อเปิดช่องจมูกให้ลมผ่านออกทั้งทางจมูกและปาก ถ้าลมถูกกักเอาไว้ที่ช่องปากตำแหน่งใด ตำแหน่งนั้นก็จะถูกเรียกว่าเป็นที่เกิดของพยัญชนะนาสิกนั้น ในภาษาไทยมีเสียงพยัญชนะนาสิก 3 หน่วยเสียง คือ /m/ เกิดที่ริมฝีปาก /n/ เกิดที่บริเวณปุ่มเหงือกหรือฟันและ /ŋ/ เกิดที่เพดานอ่อน

2.1.10.1.4 เสียงพยัญชนะข้างลิ้น (Lateral)

เสียงพยัญชนะข้างลิ้น (Lateral) เป็นเสียงพยัญชนะที่มักจะเป็นเสียงก้อง การเปล่งเสียงจะต้องเอาลิ้นปิดบริเวณปุ่มเหงือกและเพดานแข็งส่วนกลางเอาไว้ แล้วปล่อยให้ลมผ่านออกมาทางข้างลิ้น เสียงพยัญชนะข้างลิ้นในภาษาไทย มีเพียงหน่วยเสียงเดียวคือ /l/ สำหรับหน่วยเสียง /l/ นี้จะปรากฏเป็นพยัญชนะเดี่ยวหรือเกิดตามหน่วยเสียง /p/ /ph/ /k/ /k^h/ กลายเป็นพยัญชนะต้น /pl/ /p^hl/ /kl/ /k^hl/ ได้

2.1.10.1.5 เสียงพยัญชนะรัว (Trill)

เสียงพยัญชนะรัว (Trill) เกิดจากการที่ลมซึ่งขับออกมาจากปอดผ่านอวัยวะหนึ่งในช่องปากที่กำลังกระทบกับอีกอวัยวะหนึ่งอย่างรุนแรง ลมอาจจะเกิดจากการขยับของอวัยวะบางส่วนซึ่งไม่ใช่ปอดก็ได้ อวัยวะที่ใช้ออกเสียงพยัญชนะรัวมีอยู่หลายที่ด้วยกัน เช่น ริมฝีปากทั้งสอง ปลายลิ้นกับปุ่มเหงือก เป็นต้น ในภาษาไทยมีการใช้เสียงพยัญชนะรัวอยู่เสียงเดียว คือ หน่วยเสียง /r/ เกิดบริเวณปุ่มเหงือกหรือฟัน (Voiced Alveolar or Dental) ในการออกเสียงจะใช้ส่วนปลายของลิ้นสั่นสะบัดไปมาบริเวณหลังฟันด้านบน หน่วยเสียง /r/ สามารถเกิดตามหลังหน่วยเสียง /p/ /p^h/ /t/ /k/ /k^h/ เกิดเป็นพยัญชนะต้นควบ /pr/ /p^hr/ /tr/ /kr/ /k^hr/

2.1.10.1.6 เสียงพยัญชนะกึ่งสระ

เสียงพยัญชนะกึ่งสระหรืออรรธ สระ (Semi-vowel, Approximant) การเปล่งเสียงพยัญชนะกึ่งสระ อวัยวะที่ใช้เปล่งเสียงจะต้องอยู่ในตำแหน่งการออกเสียงของสระใดสระหนึ่งก่อน แล้วจึงเปล่งเสียงออกมาพร้อมกับเปลี่ยนตำแหน่งของอวัยวะไปสู่ตำแหน่งการออกเสียงของอีกสระหนึ่ง ในภาษาไทยมีหน่วยเสียงกึ่งสระอยู่ 2 หน่วยเสียง คือ /j/ เกิดจากทำปากและลิ้นเหมือนกับจะออกเสียง /ia/ หรือ /i:a/ แล้วเปล่งเสียงออกมาในขณะที่เปลี่ยนอวัยวะไปสู่ตำแหน่งของสระอื่น และ /w/ เกิดจากการเปล่งเสียงออกมาขณะที่เปลี่ยนอวัยวะจากตำแหน่งของสระ /ua/ หรือ /u:a/ ไปสู่ตำแหน่งของสระอื่น

ตารางที่ 6 และตารางที่ 7 แสดงการเปรียบเทียบพยัญชนะในภาษาไทยกับสัญลักษณ์ต่างๆ ที่ใช้แทนเสียงพยัญชนะเหล่านั้นตามตัวอักษรที่กำหนดโดยสมาคมสัทศาตร์สากล (International Phonetic Alphabet : IPA)

ตารางที่ 6 ตัวอักษรที่ใช้แทนเสียงพยัญชนะต้นในภาษาไทย

พยัญชนะต้น	สัทอักษรสากล	พยัญชนะต้น	สัทอักษรสากล
ป	/p/	ม หม	/m/
ต ฏ	/t/	น ณ หน	/n/
จ จร	/c/	ง หง	/ŋ/
ก	/k/	ฝ ฟ	/f/
อ	/ʔ/	ส ศ ษ ซ ทร สร ศร	/s/
พ ภ ผ	/p^h/	ห ฮ	/h/
ท ธ ฒ ฑ ถ ฐ	/t^h/	ร หร ฤ	/r/
ช ฌ ฉ	/c^h/	ล ฬ หล ฦ	/l/
ข ค ฆ	/k^h/	ว หว	/w/
บ	/b/	ย ญ หญ อย	/j/
ด ฎ	/d/		

ตารางที่ 7 ตัวอักษรที่ใช้แทนเสียงพยัญชนะท้ายในภาษาไทย

ตัวสะกดในภาษาไทย	สัทอักษร	ตัวสะกดในภาษาไทย	สัทอักษร
แม่กก (ก ข ค ฆ)	/k̚/	แม่กง (ง)	/ŋ̚/
แม่กด (จ ช ซ ฏ ฎ ฑ ฒ ด ต ถ ท ธ ศ ษ ส)	/t̚/	แม่กม (ม)	/m̚/
แม่กบ (บ ป ภ พ ฟ)	/p̚/	แม่เกอย (ย)	/j̚/
แม่กน (น ณ ญ ร ล ฬ)	/n̚/	แม่เกอว (ว)	/w̚/

2.1.10.2 สระ

เสียงสระเป็นเสียงที่เปล่งออกมาทางช่องปากหรือจมูก โดยไม่มีอุปสรรคใดๆ กั้นไว้เลย เสียงสระทุกเสียงเป็นเสียงก้องตามธรรมชาติ เพราะว่าเสียงสระเกิดจากการที่ลมผ่านเส้นเสียงทั้งสองที่อยู่ชิดกันมาก ทำให้ลมต้องดันออกมาอย่างรุนแรงจนเส้นเสียงเกิดการสั่นสะเทือน ลักษณะของเสียงสระเกิดจากการเคลื่อนไหวของลิ้นไปยังตำแหน่งต่างๆ ภายในช่องปาก โดยรายละเอียดของแต่ละคุณลักษณะที่ทำให้เสียงสระแตกต่างกันและใช้บ่งชี้ถึงสระมีดังนี้

1. ส่วนของลิ้นที่ใช้เปล่งเสียง (Place of Articulation) ลิ้นมีส่วนทำให้เสียงสระแตกต่างกันออกไปขึ้นอยู่กับส่วนที่ของลิ้นที่ยกขึ้น โดยถ้าลิ้นส่วนหน้าถูกยกขึ้น

ใกล้เพดานแข็งก็จะเรียกเสียงสระนั้นว่าเสียงสระส่วนเพดานแข็งหรือสระหน้า แต่ถ้าสระใดออกเสียงโดยยกลิ้นส่วนหลังให้จุดสูงสุดอยู่ที่เพดานอ่อน ก็จะถูกเรียกว่าสระส่วนเพดานอ่อนหรือสระหลัง และหากการออกเสียงสระใดลิ้นส่วนกลางถูกยกขึ้นไปยังส่วนกลางของเพดานปากก็จะเรียกสระนั้นว่าสระกลาง

2. ระยะห่างระหว่างลิ้นและเพดานปาก (Degree of Stricture) ระยะห่างระหว่างลิ้นกับเพดานปากในตำแหน่งที่ลิ้นยกตัวสูงสุดจะเป็นตัวกำหนดว่าสระนั้นๆ เป็นสระเปิดหรือสระปิด กล่าวคือ ถ้าหากลิ้นอยู่ห่างจากเพดานปากมากก็จะทำให้โพรงในปากเปิดกว้าง เสียงสระที่ได้ก็จะเป็นสระเปิด ในทางตรงกันข้ามหากลิ้นอยู่ใกล้กับเพดานปากมากช่องโพรงจะแคบ เสียงสระที่เปล่งออกมาจะเป็นสระปิด เป็นต้น
3. การห่อริมฝีปาก (Labialization) หมายถึง การที่ริมฝีปากทั้งสองเคลื่อนไหวโดยยื่นตัวไปข้างหน้าแล้วห่อกลมมากน้อยเพียงใด ถ้าริมฝีปากห่อกลมมากสระที่ได้ก็จะเป็นสระกลม (Round Vowel) แต่หากในขณะที่เปล่งเสียง ริมฝีปากทั้งสองฉีกออกก็จะเป็นสระไม่กลม (Un-rounded Vowel)
4. ลักษณะเสียงนาสิก (Nasalization) เป็นการออกเสียงให้ขึ้นจมูก ซึ่งจะทำให้แตกต่างจากสระโอษฐชะ (Oral Vowel) เนื่องจากการเปล่งเสียงสระโอษฐชะนั้นเพดานอ่อนจะยกขึ้นปิดโพรงจมูกทำให้อากาศไม่สามารถผ่านออกทางจมูกได้ แต่การออกเสียงสระนาสิกนั้น เพดานอ่อนละลดต่ำลงและปล่อยให้อากาศผ่านออกทางช่องปากและช่องจมูกด้วยในเวลาเดียวกัน สำหรับภาษาไทยไม่มีการออกเสียงสระนาสิก
5. ความยาวในการออกเสียง สระในภาษาไทยหากแบ่งตามความสั้นยาวจะถูกแบ่งออกเป็น 2 ประเภท คือ สระเสียงสั้นและสระเสียงยาว ซึ่งความยาวในการออกเสียงนั้นมีความสำคัญมากในภาษาไทยเพราะว่า เป็นคุณลักษณะที่ใช้แยกความหมายของคำในภาษาไทย เช่น คำว่า “ลุก” กับ “ลูก” เป็นต้น

หน่วยเสียงสระต่างๆ ที่ใช้อยู่ในภาษาไทย สามารถแบ่งออกเป็น 2 ประเภท คือ เสียงสระเดี่ยว ได้แก่ /i/ /iː/ /e/ /eː/ /æ/ /æː/ /ɨ/ /ɨː/ /ɜ/ /ɜː/ /a/ /aː/ /u/ /uː/ /o/ /oː/ /ɔ/ /ɔː/ และเสียงสระประสม ได้แก่ /ia/ /iːa/ /ua/ /uːa/ /ɨa/ /ɨːa/ ซึ่งการเปรียบเทียบสระในภาษาไทยกับสัญลักษณ์ต่างๆที่ใช้แทนเสียงสระเหล่านั้นตามตัวอักษรที่กำหนดโดยสมาคมสัทศตร์สากลสามารถแสดงได้ดังตารางที่ 8 และตารางที่ 9

ตารางที่ 8 ตัวอักษรที่ใช้แทนเสียงสระเดี่ยวในภาษาไทย

สระในภาษาไทย	สัทอักษรสากล	สระในภาษาไทย	สัทอักษรสากล
/i/	อิ	/iː/	อี
/ɜ/	เออะ	/ɜː/	เออ
/e/	เอะ	/eː/	เอ
/a/	อะ	/aː/	อา
/æ/	แอะ	/ æː/	แอ
/u/	อุ	/uː/	อู
/ɨ/	อึ	/ɨː/	อือ
/o/	โอะ	/oː/	โอ
/ɔ/	เอาะ	/ɔː/	ออ

ตารางที่ 9 ตัวอักษรที่ใช้แทนเสียงสระผสมในภาษาไทย

สระในภาษาไทย	สัทอักษรสากล	สระในภาษาไทย	สัทอักษรสากล
/ia/	เอียะ	/iːa/	เอีย
/ua/	เอือะ	/uːa/	เอือ
/ɨa/	อัวะ	/ɨːa/	อัว

2.1.11 **ขั้นตอนวิธีเชิงพันธุกรรม** [25],[29]

ขั้นตอนวิธีเชิงพันธุกรรมจำลองกระบวนการวิวัฒนาการของชารลส์ ดาร์วิน และการดำเนินเกี่ยวกับยีน (Genetic Operation) ต่างๆ ที่เกิดขึ้นตามธรรมชาติบนโครโมโซม โครโมโซมคือ สายอักขระที่จำลองด้วยอักษร 4 ตัว คือ Adenine (A), Cytosine (C), Guanine (G) และ Thymine (T) เรียงกัน ดีเอ็นเอโมเลกุลสามารถคัดลอกตัวเองได้อย่างแม่นยำ นอกจากนั้น สายอักขระย่อยที่มีความยาวมากกว่า 1,000 ตัวอักษร จากดีเอ็นเอโมเลกุลจะถูกแปลด้วยรหัสพันธุกรรมไปเป็นโปรตีนและเอ็นไซม์เพื่อสร้างและควบคุมพฤติกรรมในเซล เมื่อโครงสร้างและพฤติกรรมถูกสร้างขึ้นก็ทำให้แต่ละเซลสามารถทำงานในสภาพแวดล้อมรอบๆ ตัวได้ โครโมโซมของลูกจะประกอบด้วยสายอักขระของผู้ให้กำเนิดผสมกัน ดังนั้น พฤติกรรมหรือลักษณะที่มีความเด่นและมีประสิทธิภาพจะถูกส่งต่อไปยังรุ่นต่อในโอกาสที่สูง กว่า โดยส่วนใหญ่แล้วการกลายพันธุ์เกิดที่โครโซม ขั้นตอนวิธีเชิงพันธุกรรม คือ ขั้นตอนวิธีที่เปลี่ยนประชากรวัตถุทางคณิตศาสตร์แต่ละอัน (โดยปกติใช้เป็นสายอักขระความยาวคงที่) ไปสู่ประชากรรุ่นใหม่โดยใช้รูปแบบตามหลักการของชารลส์ ดาร์วินเกี่ยวกับการสืบพันธุ์และการวิวัฒนาการโดยผ่านการคัดเลือกตามธรรมชาติ

ปัญหาร้านขายแฮมเบอเกอร์

ขั้นตอนวิธีเชิงพันธุกรรมสามารถยกตัวอย่างได้ด้วยปัญหาการหาค่าเหมาะสม ซึ่งในที่นี้คือ การหากลยุทธ์ทางธุรกิจที่ดีที่สุดสำหรับร้านขายแฮมเบอร์เกอร์ โดยกลยุทธ์จะขึ้นอยู่กับการตัดสินใจเลือกค่าต่างๆ ที่เป็น ทวิภาค 3 อย่าง คือ

1. ราคา จะขายแฮมเบอร์เกอร์ที่ราคา 50 เซนต์ หรือว่า $10
2. เครื่องดื่ม ควรจะเสิร์ฟไวน์หรือโค้กพร้อมกับแฮมเบอร์เกอร์
3. ความรวดเร็วในการบริการ ควรจะเร็วหรือช้า

โดยวัตถุประสงค์ คือ การหาส่วนผสมของการตัดสินใจทั้งสามอย่างนี้เพื่อให้มีผลกำไรมากสุด เนื่องจากว่ามี 3 ตัวแปร ตัวแปรแต่ละมีค่าที่เป็นไปได้ 2 ค่า ดังนั้น เราจึงจะใช้สายอักขระความยาวเท่ากับ 3 และขนาดชุดตัวอักษรเท่ากับ 2 แต่ละตัวแปรจะใช้เลข 0 และ 1 แทนค่าที่เป็นไปได้สำหรับแต่ละการตัดสินใจ การกำหนดความยาวของสายอักขระและการส่ง (Mapping) แบบนี้ เรียกว่า แผนการแทน (Representation Scheme) ซึ่งถือเป็นขั้นตอนแรกเพื่อเตรียมพร้อมแก้ปัญหาที่สนใจด้วยขั้นตอนวิธีเชิงพันธุกรรม ซึ่งในปัญหานี้สามารถแสดงตัวอย่างของแผนการแทนได้ด้วยตารางที่ 10

ตารางที่ 10 ตัวอย่างของแผนการแทน

ร้านค้าเลขที่	ราคา	เครื่องดื่ม	ความเร็ว	สายอักษระทวิภาคที่ใช้แทน
1	สูง	โค้ก	เร็ว	011
2	สูง	ไวน์	เร็ว	001
3	ต่ำ	โค้ก	ช้า	110
4	สูง	โค้ก	ช้า	010

โดยปกติแล้วผู้จัดการภัตตาคารจะไม่รู้ว่าตัวแปรใดสำคัญที่สุด ไม่รู้ว่าผลกำไรมากที่สุดเป็นเท่าใด ไม่รู้ว่าเขาจะสูญเสียกำไรไปเท่าใดหากเลือกกลยุทธ์ทางการค้าที่ไม่ถูกต้อง และเขาก็ยังไม่รู้อีกว่า ตัวแปรใดหากเปลี่ยนค่าเพียงตัวเดียวจะทำให้ได้ผลกำไรมากขึ้นมาก ความจริงแล้ว ผู้จัดการไม่ทราบอะไรเลยเกี่ยวกับความสัมพันธ์ของทั้งสามตัวแปรด้วยกันเองและความสัมพันธ์ของทั้งสามตัวแปรกับผลกำไรที่ได้

เนื่องจากผู้จัดการไม่ทราบถึงสภาพแวดล้อมที่เขากำลังจะพบ ดังนั้นเขาจึงตัดสินใจที่จะใช้การสุ่มเลือกกลยุทธ์สำหรับแต่ละร้านทั้ง 4 ของเขา เพื่อทดลองใช้ในเวลา 1 อาทิตย์ ซึ่งสมมุติให้สุ่มเลือกกลยุทธ์ได้ตามตารางที่ 11 ซึ่งกลยุทธ์ที่สุ่มเลือกได้นี้เป็นประชากรเริ่มต้น

ตารางที่ 11 ตัวอย่างประชากรเริ่มต้น (รุ่นที่ 0)

ประชากร	สายอักขระ	ค่าความเหมาะ
i	Xi	f(Xi)
1	011	3
2	001	1
3	110	6
4	010	2
	รวม	12
	แย่ที่สุด	1
	เฉลี่ย	3.00
	ดีที่สุด	6

การทำแบบนี้ ผู้จัดการร้านอาหารกำลังดำเนินตามวิธีขั้นตอนวิธีเชิงพันธุกรรม เนื่องจาก การเริ่มต้นของขั้นตอนวิธีเชิงพันธุกรรมจะเริ่มจากการพยายามที่จะเรียนรู้สิ่งต่างๆ เกี่ยวกับสภาพแวดล้อมโดยทดสอบจากการสุ่มจุดใดๆ ขึ้นมาจำนวนหนึ่งจากปริภูมิค้นหา (Search Space) นั่นก็คือ ขั้นตอนวิธีเชิงพันธุกรรมเริ่มจากรุ่นที่ 0 (รุ่นที่เกิดจากการสุ่ม) ซึ่งในตัวอย่างนี้ขนาดของประชากร (M) มีค่าเป็น 4

สำหรับแต่ละรุ่นที่ขั้นตอนวิธีเชิงพันธุกรรมกำลังทำงาน ประชากรแต่ละตัวจะถูกทดสอบกับสภาพแวดล้อมเพื่อหาค่าความเหมาะ (ค่าความเหมาะของประชากรแต่ละตัว) ในตัวอย่างนี้ค่าความเหมาะของประชากรเริ่มต้น ทั้ง 4 ตัว ดังแสดงในตารางที่ 10 โดยกำหนดว่า ค่าความเหมาะของแต่ละกลยุทธ์จะมีค่าเท่ากับค่าในเลขฐาน 10 เช่น ค่าความเหมาะของสายอักขระ 110 จะเท่ากับ 6 และ 011 จะเท่ากับ 3 เป็นต้น ดังนั้น ค่าความเหมาะสมมากสุดจึงเท่ากับ 7 สิ่งที่ผู้จัดการเรียนรู้จากการทดสอบ 4 กลยุทธ์ที่สุ่มขึ้นมา ก็คือ เขาพบว่ากลยุทธ์แบบ 110 ให้ผลกำไร $6 ใน 1 สัปดาห์ ซึ่งกลยุทธ์นี้เป็นกลยุทธ์ที่ดีที่สุดในรุ่นและกลยุทธ์ 001 ให้ผลกำไรเพียง $1 ใน 1 สัปดาห์ ซึ่งเป็นกลยุทธ์ที่แย่ที่สุด

ขั้นตอนวิธีเชิงพันธุกรรมจะใช้ค่าความเหมาะของแต่ละประชากรเพื่อเปลี่ยนประชากรจากกลุ่มหนึ่งไปเป็นประชากรอีกกลุ่มหนึ่งตามหลักการของชารลส์ ดาร์วินเกี่ยวกับการสืบพันธุ์และการวิวัฒนาการโดยผ่านการคัดเลือกตามธรรมชาติและการดำเนินการเกี่ยวยีนตามธรรมชาติ เราจะเริ่มด้วยการสืบพันธุ์โดยจะเลือกปร ะชากรที่จะอยู่รอดในรุ่นถัดไ ปด้วยความน่าจะเป็นตามสัดส่วนของค่าความเหมาะนั่นคือ ผลรวมของค่าความเหมาะของรุ่นที่ 0 นี้เท่ากับ 3+1+6+2 = 12 ประชากรที่ดีที่สุด (110) มีค่าความเหมาะ เท่ากับ 6 จึงมีโอกาสที่จะถูกเลือกให้

อยู่ในรุ่นถัดไปเท่ากับ 0.5 ดังนั้น เราจึงคาดหวังว่า สายอักขระ 110 จะจองพื้นที่ครึ่งหนึ่งของประชากรในรุ่นถัดไป (ในที่นี้คือจำนวนประชากรในแต่รุ่นทั้งหมดเป็น 4 ดังนั้น สายอักขระน่าจะมี 2 ตัวในประชากรรุ่นที่ 1) แต่เนื่องจาก ขั้นตอนวิธีเชิงพันธุกรรมใช้ความน่าจะเป็นจึงมีความเป็นไปได้ที่จะพบสายอักขระ 110 ถึง 3 ครั้ง ในประชากรรุนต่อไปและก็มีความเป็นไปได้ที่จะพบ 4 ครั้ง หรือไม่พบเลย ในทำนองเดียวกันสายอักขระ 011 ก็ความน่าจะเป็น 0.25 ที่จะถูกเลือกให้อยู่รอด นั่นคือ เราคาดหวังว่า สายอักขระ 011 จะจองที่ 1 ที่ ในประชากรรุ่นต่อไป เป็นต้น

หากสมมุติว่าสายอักขระ 110 ถูกเลือก 2 ครั้ง 011 และ 010 ถูกเลือกอย่างละ 1 ครั้งดังตารางที่ 12 และเราเรียกผลลัพธ์ที่ได้นี้ว่า แมตติงพูลที่สร้างขึ้นหลักจากการสืบพันธุ์ (Mating Pool Created after Reproduction) ซึ่งผลที่เกิดขึ้นจากขั้นตอนนี้คือ ค่าความเหมาะโดยเฉลี่ยสูงขึ้นจาก 3.0 เป็น 4.25 และค่าน้อยสุดเพิ่มขึ้นจาก 1 เป็น 2

ตารางที่ 12 ตัวอย่างการสร้างแมตติงพูล

	ประชากรรุ่นที่ 0			แมตติงพูลที่สร้างขึ้นหลักจากการสืบพันธุ์	
i	Xi	f(Xi)	f(Xi)/sum(f(Xi))	แมตติงพูล	f(Xi)
1	011	3	0.25	011	3
2	001	1	0.08	110	6
3	110	6	0.50	110	6
4	010	2	0.17	010	2
รวม		12			17
แย่ที่สุด		1			2
เฉลี่ย		3.00			4.25
ดีที่สุด		6			6

การดำเนินการเกี่ยวกับยีนต่อมาคือ การไขว้เปลี่ยน ด้วยวิธีนี้จะทำให้ จุดใหม่ๆ ของปริภูมิถูกนำมาทดสอบ การไขว้เปลี่ยนจะต้องเริ่มด้วยผู้ให้กำเนิดโดยเลือกตามค่าความเหมาะโดย การไขว้เปลี่ยนจะผลิตลูก 2 ตัวซึ่งลูกทั้งสองจะแตกต่างกับผู้ให้กำเนิดและแตกต่า งกันเอง ลูกแต่ละตัวจะได้รับยีนจากผู้ให้กำเนิดทั้งสอง การไขว้เปลี่ยนจะเริ่มด้วยการสุ่มเลือกจำนวนตั้งแต่ 1 ถึง L-1 (ในตัวอย่างนี้ L = 3 ดังนั้น ค่าที่เป็นไปได้ คือ 1 และ 2) เพื่อเลือกตำแหน่งที่จะไขว้เปลี่ยน สมมุติว่า เลือกตำแหน่งที่ 2 ผู้ให้กำเนิดจะถูกแยก ณ ตำแหน่งไขว้เปลี่ยน สมมุติให้สายอักขระ 110 และ 011 ถูกเลือกเพื่อเป็นผู้ให้กำเนิด ดังนั้นสายอักขระทั้งสองจึงถูกแยกออกเป็น 2 ส่วน คือ ส่วนที่แตกออกมาไขว้เปลี่ยน (Crossover Fragment) และส่วนที่เหลืออยู่ (Remainder) ดังตารางที่ 13

ตารางที่ 13 ตัวอย่างของสตริงที่ถูกแยกเป็นส่วนที่แตกออกมาไขว้เปลี่ยนและส่วนที่เหลืออยู่

สายอักขระ	ส่วนที่แตกออกมาไขว้เปลี่ยน	ส่วนที่เหลืออยู่
110	11-	--0
011	01-	--1

ตัวอย่างที่เป็นไปได้หลังจากการประยุกต์ใช้การดำเนินการเกี่ยวกับยีนกับประชากรรุ่นที่ 0 เพื่อสร้างประชากรรุ่นที่ 1 สามารถแสดงได้ดังตารางที่ 14

ตารางที่ 14 ตัวอย่างของรุ่นที่ 1 ที่เป็นไปได้

	ประชากรรุ่นที่ 0			แมตติงพูล		ประชากรรุ่นที่ 1		
i	Xi	f(Xi)	f/sumf	Xi	f(Xi)	จุดแยก	Xi	f(Xi)
1	011	3	0.25	011	3	2	111	7
2	001	1	0.08	110	6	2	010	2
3	110	6	0.50	110	6	-	110	6
4	010	2	0.17	010	2	-	010	2
รวม		12			17			17
แย่สุด		1			2			2
เฉลี่ย		3.00			4.25			4.25
ดีสุด		6			6			7

ทั้งการสืบพันธุ์และการไขว้เปลี่ยนจำเป็นต้องมีขั้นตอนการเลือกตามสัดส่วนของค่าเหมาะสม เราสามารถทำให้ง่ายขึ้นโดยตอนแรกประยุกต์ใช้การสืบพันธุ์กับประชากรทั้งหมดเพื่อสร้างแมตติงพูลจากนั้นค่อยประยุกต์ใช้การไข้วเปลี่ยนโดยกำหนดเปอร์เซ็นต์ของประชากรในแมตติงพูลที่จะนำไปมีส่วนร่วมในการไขว้เปลี่ยน ในกรณีใช้ Pc = 50% หมายความว่า 50% ของประชากรจะมีส่วนร่วมในการไขว้เปลี่ยนเพื่อสร้างรุ่นต่อไป และ 50% ที่เหลือก็จะมีส่วนร่วมแต่เฉพาะในการสืบพันธุ์นั่นคือ Pr = 100-50 = 50% สำหรับตัวอย่างนี้

ตารางที่ 14 แสดงการไขว้เปลี่ยนซึ่งกระทำกับแมตติ งพูล ประชากร 2 ตัวที่มีส่วนร่วมใน การไขว้เปลี่ยนถูกเลือกโดยใช้สัดส่วนตามค่าเหมาะสม เนื่องจากว่าเราสร้างแมตติงพูลตามสัดส่วนของค่าเหมาะสม ดังนั้นการเรียกประชากรเพื่อนำมาไขว้เปลี่ยน จึงสามารถเลือกใช้โดยใช้การสุ่มแบบเอกรูป (Uniform) ได้ และในสดมภ์สุดท้ายของตารางที่ 14 ก็คือ ประชากรใหม่ซึ่งเป็นผลลัพธ์จากการดำเนินการเกี่ยวกับยีนทั้งสอง

เมื่อคำนวณค่าความเหมาะของรุ่นที่ 1 จะได้ว่า ค่าความเหมาะที่ดีที่สุดของประชากรกลุ่มนี้คือ 7 ซึ่งในขณะที่รุ่นแรกมีค่าความเหมาะสมสูงสุดเป็น 6 เนื่องจาก การไขว้เปลี่ยนได้สร้างประชากรใหม่ขึ้นและในตัวอย่างนี้ประชากรใหม่มีค่าความความเหมาะสูงกว่าผู้ให้กำเนิดทั้งสอง และเมื่อเปรียบเทียบประชากรรุ่นที่ 1 กับรุ่นที่ 0 เราพบว่า

1. ค่าความเหมาะโดยเฉลี่ยสูงขึ้นจาก 3 เป็น 4
2. ค่าความเหมาะดีที่สุดของประชากรในรุ่นเพิ่มจาก 6 เป็น 7
3. ค่าความเหมาะแย่ที่สุดของประชากรในรุ่นเพิ่มจาก 1 เป็น 2

เนื่องจาก เราพบกลยุทธ์ที่แทนด้วยสายอักขระ 111 กลยุทธ์นี้เป็นกลยุทธ์ที่ดีสุดตามค่าความเหมาะ (โดยส่วนมากเราจะไม่รู้ว่าค่าความเหมาะมีค่ามากสุดเป็นเท่าใด) เราก็สามารถที่จะหยุดการทำงานได้ในรุ่นที่ 1 สำหรับตัวอย่างนี้

การกลายพันธุ์เป็นอีกการดำเนินการหนึ่งที่นิยมใช้ในขั้นตอนวิธีเชิงพันธุกรรม ความถี่ในการประยุกต์ใช้การกลายพันธุ์จะกำหนดโดยพารามิเตอร์ที่เรียกว่า ค่าความน่า จะเป็นในการกลายพันธุ์ (Mutation Probability - Pm) การกลายพันธุ์ คือ การดำเนินการที่ไม่เกี่ยวกับเพศ จึงกระทำกับประชากรเพียงตัวเดียว และเริ่มจากสุ่มเลือกประชากรจากแมตติ้งแล้วค่อยสุ่มเลือก ก ตำแหน่งที่ต้องการจะให้กลายพันธุ์ หากสายอักขระเป็นสายอักขระทวิภาคจุดที่กลายพันธุ์ ก็จะเป็นส่วนเติมเต็ม (Complement) ของค่าเดิม เราใช้การดำเนินการนี้เพื่อเพิ่มความหลากหลายให้แก่ประชากรในรุ่นต่อไป ขั้นตอนวิธีเชิงพันธุกรรมไม่ได้ทำงานโดยการเปลี่ยนสายอักขระสุ่มจากประชากรเริ่มต้นไปสู่จุดดีสุดวงกว้าง (Global Optimum) โดยการกลายพันธุ์เพียงครั้งเดียว ซึ่งการวิวัฒนาการตามแบบของชารลส์ ดาร์วิน ก็ไม่ได้เปลี่ยนคาร์บอน ไนโตรเจน ออกซิเจนและไฮโดรเจนอิสระไปเป็นกบภายในการกลายพันธุ์เพียงครั้งเดียว

โดยสรุปขั้นตอนวิธีเชิงพันธุกรรมจะมี 4 ขั้นตอนหลักเพื่อเตรียมพร้อมก่อนที่จะแก้ปัญหา

1. ระบุแผนการแทน
2. ระบุวิธีการวัดค่าความเหมาะ
3. ระบุตัวแปรและพารามิเตอร์ต่างๆ สำหรับควบคุมขั้นตอนวิธีเชิงพันธุกรรม
4. ระบุลักษณะของผลที่ต้องการและเกณฑ์สำหรับหยุดการทำงาน

และขั้นตอนการกระทำการขั้นตอนวิธีเชิงพันธุกรรมที่กระทำบนสายอักขระความยาวคงที่สามารถสรุปได้ดังนี้

1. สุ่มสร้างประชากรเริ่มต้น

2. วนซ้ำขั้นตอนต่อไปนี้กับประชากรจนกระทั่งเข้าสู่เงื่อนไขหยุดการทำงาน
    a. คำนวณค่าความเหมาะของประชากรแต่ละตัว
    b. สร้างประชากรใหม่โดยประยุกต์ใช้ 2 ใน 3 วิธีต่อไปนี้ โดยเลือกประชากรตามสัดส่วนความน่าจะเป็นที่คำนวณจากค่าความเหมาะ
        i. คัดลอกประชากรที่มีอยู่ไปสู่ประชากรรุ่นใหม่
        ii. สร้างประชากรใหม่โดยการผสมสายอักขระที่สุ่มได้จากประชากรที่มีอยู่
        iii. สร้างประชากรใหม่จากสายอักขระที่มีอยู่โดยการกลายพันธุ์
3. ประชากรที่ดีที่สุดที่ปรากฏในรุ่นใดๆ ก็ตาม คือ ผลลัพธ์ในของขั้นตอนวิธีเชิงพันธุกรรมและก็คือคำตอบสำหรับปัญหา (อาจจะเป็นคำตอบโดยประมาณ)

ขั้นตอนวิธีเชิงพันธุกรรมมีความหลากหลายในขั้นตอนมากมายในภาคผนวก ข แสดงแผนภูมิสายงานตัวอย่างของขั้นตอนทางวิธีพันธุกรรมอีกแบบหนึ่ง

2.2 งานวิจัยที่เกี่ยวข้อง

สำหรับส่วนนี้จะมีการเรียงลำดับเนื้อหาโดยขอกล่าวถึงฐานข้อมูลเสียงโลตัสซึ่งเป็นฐานข้อมูลเสียงภาษาไทยแบบต่อเนื่องขนาดใหญ่ และจากนั้นจะกล่าวถึงวิธีการประมาณทอพอโลยีในภาพรวม 2 วิธี คือ การประมาณทอพอโลยีแบบหลายทางเดินและการประมาณทอพอโลยีโดยใช้ฟังก์ชันวัตถุประสงค์ควบคู่กับการผลิตทอพอโลยีเพื่อนำมาทดสอบกับวัตถุประสงค์

2.2.1 ฐานข้อมูลโลตัส [38]

ฐานข้อมูลโลตัส (Large Vocabulary Thai Continuous Speech Recognition LOTUS) เป็นฐานข้อมูลเสียงที่เป็นภาษาไทยขนาดใหญ่ที่มีคำศัพท์เป็นจำนวนมากแบบเสียงพูดต่อเนื่องโดยโลตัสได้ถูกพัฒนาขึ้นจากความร่วมมือของ 3 สถาบัน คือ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ มหาวิทยาลัยสงขลานครินทร์และมหาวิทยาลัยเทคโนโลยีมหานคร ประกอบไปด้วยชุดประโยค 4 ชุดคือ ชุดหน่วยเสียงสมดุล (Phonetically Distributed Set: PD) ใช้สำหรับการฝึกฝนแบบจำลองทางเสียงเริ่มต้น ชุดฝึกฝน (Training Set: TR) ใช้สำหรับฝึกฝนแบบจำลองทางเสียงเพิ่มเติมและฝึกฝนแบบจำลองภาษา (Language Model) ชุดทดสอบเพื่อพัฒนา (Development Test Set: DT) ที่ใช้ในขั้นตอนการวิจัยระบบรู้จำเสียงพูดและชุดทดสอบเพื่อประเมิน (Evaluation Test Set: ET) โลตัสนั้นใช้ผู้พูดทั้งหมด 248 คน เป็นชาย 124 คน และหญิง 124 คน ครอบคลุมคำศัพท์ในภาษาไทยกว่า 5000 คำและครอบคลุมหน่วยเสียงคู่

(Bi-phone) ถึง 90.9% ของหน่วยเสียงคู่ทั้งหมดที่เกิดได้ภาษาไทย โดยในแต่ละชุดมีรายละเอียดตามตารางที่ 15

ตารางที่ 15 รายละเอียดของชุดข้อมูลในโลตัส

รายละเอียด	PD set	TR set	DT set	ET set
จำนวนประโยค	801	3,007	500	500
จำนวนคำศัพท์	2,269	5,000	1,622	1,630
จำนวนคำ	7,847	55,504	8,076	8,290

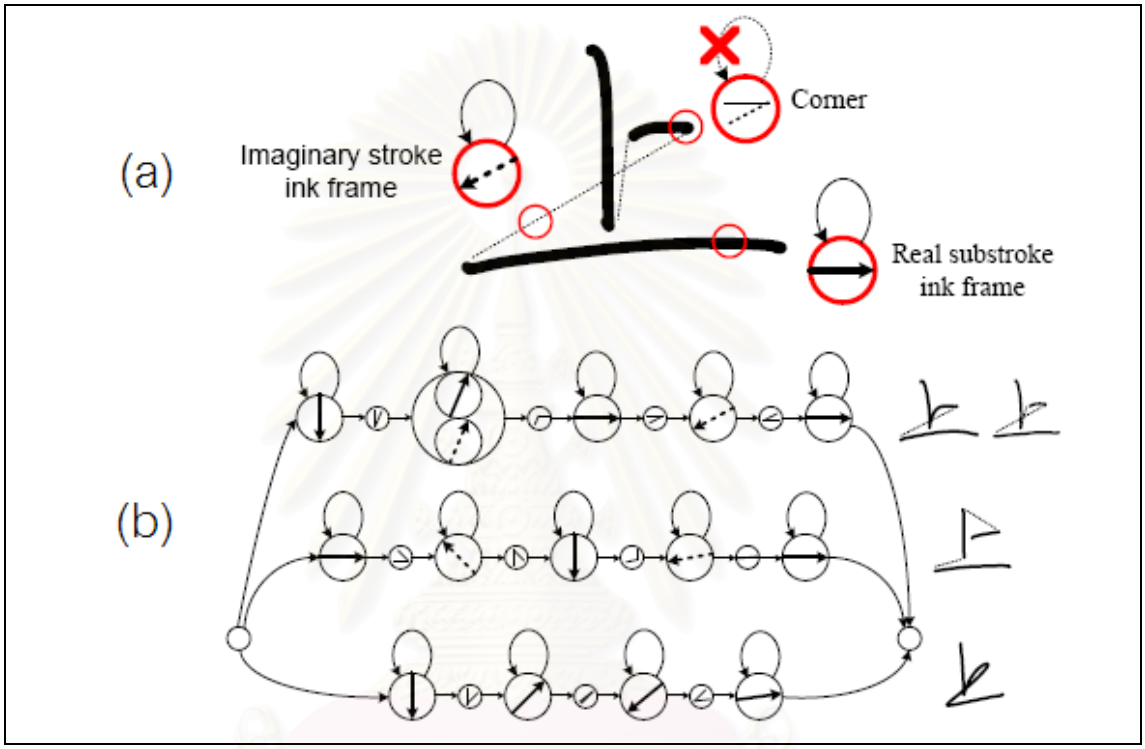
ฐานข้อมูลโลตัสมีการบันทึกเสียง 2 ประเภท โดยประเภทแรกจะอัดเสียงในสภาพแวดล้อมเงียบ ประเภทที่สองเป็นการอัดเสียงในสภาพแวดล้อมสำนักงานทั่วไป ทั้งสองประเภทจะบันทึกเสียงผ่านไมโครโฟน 2 ตัวพร้อมกัน คือ แบบไดนามิกโคลสทอล์ค (Dynamic Close-Talk) คุณภาพระดับสูงและแบบไดนามิกยูนิไดเรคชันแนล (Dynamic-Unidirectional) คุณภาพระดับปานกลาง ข้อมูลที่ได้จะถูกเก็บในรูปของแถบบันทึกเสียงดิจิทัล (Digital Audio Tape - DAT) ก่อนนำมาเชื่อมต่อกับคอมพิวเตอร์เพื่อแปลงสัญญาณเป็นไฟล์อิเล็กทรอนิกพื้นฐาน (Standard Electronic File)

การพูดในฐานข้อมูลเสียงโลตัสจะใช้วิธีพูดทีละประโยคตามรูปอ่าน (Orthography) ของประโยค ในแต่รอบผู้พูดจะได้รับคำแนะนำและทดสอบเพื่อปรับวิธีการพูดความดังในการพูดพร้อมทั้งได้รับคำแนะนำวิธีการพูดเพื่ออัดเสียง

2.2.2 การประมาณทอพอโลยีแบบหลายทางเดิน (Multi-path)

ในงานวิจัยของ Han และคณะ [39] ซึ่งเป็นการค้นหาทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมสำหรับการรู้จำตัวหนังสือตะวันออกลายมือเขียนแบบออนไลน์ (Online Handwriting Recognition of East Asian Character) โดยในงานวิจัยนี้มีความแนวคิดว่า ตัวหนังสือมีวิธีการเขียนได้หลายรูปแบบ ดังตัวอย่างในรูปที่ 14 สมมุติว่า ตัวหนังสือจีนที่แสดงในรูปที่ 14a สามารถเขียนได้ 3 วิธี ดังนั้น Han จึงกำหนดให้แบบจำลองฮิดเดนมาร์คอฟที่ใช้ในการรู้จำตัวหนังสือตัวนี้มี 3 ทางเดิน โดยที่แต่ละทางเดินเป็นวิธีการเขียนและกำหนดท อพอโลยีให้กับแต่ละทางเดินตามความเหมาะสมโดยอาศัยการเปลี่ยนทิศของพู่กัน เช่น สำหรับกรณีวิธี การเขียนในทางเดินกลางของ รูปที่ 14b เป็นวิธีการเขียนที่มีการเปลี่ยนทิศของพู่กัน 4 ครั้ง (เส้นประคือ เส้นการเขียนเสมือน) ดังนั้น Han จึงให้มีจำนวนสถานะทั้งหมดในทางเดินเป็น 9 สถานะ(5 สถานะสำหรับเส้นการเขียน และ 4 สถานะสำหรับการเปลี่ยนทิศของพู่กัน) สำหรับวิธีการเขียนแบบนี้ โดยกำหนดให้สถานะเป็นที่เป็นการเปลี่ยนทิศของพู่กันเป็นสถานะที่ไม่มีการเปลี่ยนสถานะ

อัตตา (Self Transition) และให้สถานะที่เป็นการ เส้นการเขียนมีการเปลี่ยนสถานะอัตตา เป็นต้น นอกจากนั้น Han ยังกำหนดจำนวนส่วนผสมเกาส์เซียนโดยสังเกตจากวิธีการเขียนว่า หากมีเส้นการเขียนใดสามารถเป็นได้ทั้งเส้นการเขียนจริงและเส้นการเขียนเสมือนก็จะให้สถานะที่แทนเส้นการเขียนนั้นมีจำนวนส่วนผสมเกาส์เซียนเป็น 2 แต่ถ้าหากเป็นเส้นการเขียนได้เพียงแบบเดียวก็ให้มีจำนวนส่วนผสมเกาส์เซียนเป็น 1 และสำหรับการวิธีการเชื่อมต่อของสถานะนั้น Han เลือกใช้การเชื่อมต่อแบบซ้ายขวา



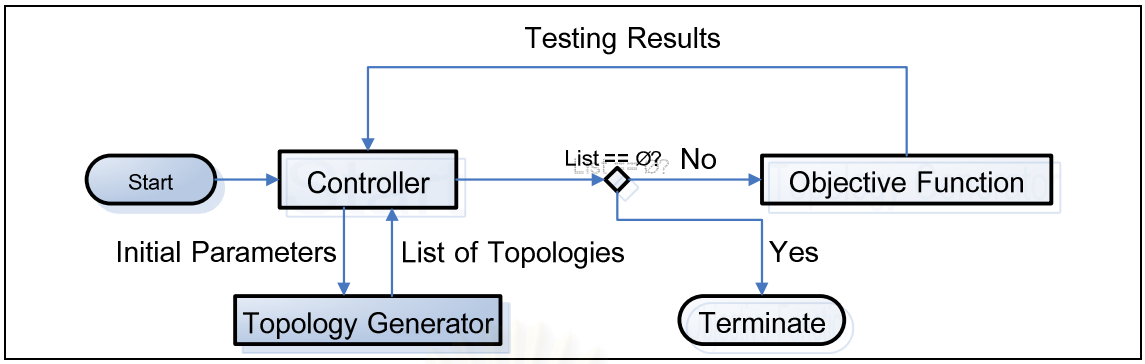
รูปที่ 14 แบบจำลองฮิดเดนมาร์คอฟแบบหลายทาง (รูปจาก [39])

2.2.3 การประมาณทอพอโลยีโดยใช้ฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยี

งานวิจัยที่เกี่ยวข้องกับการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟแบบนี้โดยส่วนใหญ่ [30-36] แล้วจะประกอบด้วย 2 ส่วน คือ วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ (Topology Generation Method) และเกณฑ์ที่ใช้วัดความเหมาะของทอพอโลยีหรือฟังก์ชันวัตถุประสงค์ (Objective Function) กล่าวคือ ในงานวิจัยประเภทนี้จะระบุถึงวิธีการผลิตทอพอโลยีที่จะนำมาทดสอบกับฟังก์ชันวัตถุประสงค์ที่นำเสนอ ดังนั้น ในหัวข้อนี้จึงแบ่งการนำเสนอออกเป็น 2 ส่วน คือ

- ฟังก์ชันวัตถุประสงค์
- วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ

ซึ่งทั้งสองส่วนนี้มีความสัมพันธ์กันดังแสดงในรูปที่ 15



รูปที่ 15 ความสัมพันธ์ระหว่างฟังก์ชันวัตถุประสงค์และวิธีการคัดเลือกทอพอโลยี

ขั้นตอนการประมาณทอพอโลยีเริ่มต้นด้วย Controller ส่งพารามิเตอร์เริ่มต้นให้ Topology Generator สร้างเซตของทอพอโลยีที่น่าสนใจ จากนั้นส่งเซตนี้ให้แก่ Objective Function เพื่อคำนวณค่าความเหมาะสมของทอพอโลยีแต่ละทอพอโลยี และ Controller จะใช้ค่าที่คำนวณได้นี้เพื่อคำนวณหาค่าพารามิเตอร์สำหรับการวนซ้ำในรอบต่อไป การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟจะจบการทำงาน เมื่อเซตของทอพอโลยีที่น่าสนใจเป็นเซตว่าง

ทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟประกอบด้วย 3 องค์ประกอบคือ

1. จำนวนสถานะ
2. จำนวนส่วนผสมเกาส์เซียน (Gaussian Mixture) ต่อสถานะ
3. รูปแบบการเชื่อมต่อของสถานะ (Connectivity Architecture)

ดังนั้นการประมาณทอพอโลยีจึงให้ความสนใจแก่พารามิเตอร์ทั้งสามนี้ ซึ่งการกำหนดทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟให้กับแต่ละหน่วยเสียงนั้น ไม่มีทฤษฎีหรือวิธีการเลือกที่สามารถพิสูจน์ได้ว่า แบบจำลองรูปแบบใดเหมาะสม [3] ดังนั้นการประมาณทอพอโลยีจึงเป็นวิธีการกำหนดทอพอโลยีแบบฮิวริสติก (Heuristic)

2.2.3.1 ฟังก์ชันวัตถุประสงค์

เมื่อพิจารณาการรู้จำเสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ ทอพอโลยีที่เหมาะสมนั้นจะช่วยเพิ่มประสิทธิภาพการรู้จำเสียงให้มีความแม่นยำและถูกต้องมากยิ่งขึ้น เนื่องด้วยแบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองทางสถิติ ดังนั้น เราจึงไม่สามารถระบุได้อย่างชัดเจนว่า ทอพอโลยีที่เหมาะสมนั้น ควรมีจำนวนสถานะและส่วนผสมเกาส์เซียนต่อสถานะเป็นเท่าใด และมีรูปแบบของการเชื่อมต่อของสถานะเป็นอย่างไร ฟังก์ชันวัตถุจึงเปรียบเสมือนแนวทางบนสมมุติฐานที่ตั้งขึ้นเพื่อวัดความเหมาะสมของทอพอโลยีของหน่วยเสียงที่กำหนด ฟังก์ชันวัตถุตัว

แรกที่ถูกนำมาใช้ คือ ความเป็นไปได้มากสุด (Maximum Likelihood: ML) โดยที่ ML วัดความเหมาะสมจากความสามารถในการปรับพารามิเตอร์ของทอพอโลยีให้เหมาะกับข้อมูลฝึกฝน ดังเช่นในงานวิจัย Siriboon และคณะ [33] และงานวิจัยของ Singer [30] และคณะได้นำ ML ไปใช้เป็นฟังก์ชันวัตถุประสงค์เพื่อประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ แต่อย่างไรก็ดี เนื่องจาก ML วัดความเหมาะสมของทอพอโลยีโดยอาศัยเพียงความเหมาะกับข้อมูลฝึกฝน ทำให้อาจจะเกิดปัญหาการปรับเหมาะเกินไป (Overfit) เพื่อแก้ปัญหานี้จึงได้มีเสนอฟังก์ชันวัตถุประสงค์ที่นำการนำหลักการมีดโกนออคแคมมาประยุกต์ใช้ คือ เกณฑ์สารสนเทศแบบเบส์ (Bayesian Information Criterion: BIC) และรายละเอียดประกอบความยาวน้อยสุด (Minimum Description Length :MDL) โดยที่ BIC ถูกนำเสนอโดย Schwarz [40] หลังจากนั้น Dafang และคณะ [32] ได้นำ BIC มาประยุกต์ใช้เป็นฟังก์ชันวัตถุประสงค์ Jitsushiro และคณะ [34] ได้เสนอฟังก์ชันวัตถุประสงค์ MDL ซึ่งฟังก์ชันวัตถุประสงค์ทั้งสองจะแบ่งการคำนวณทอพอโลยีออกเป็น 2 ส่วน คือ ความเป็นไปได้สูงสุดและส่วนหักลบที่คำนวณจากจำวนพารามิเตอร์อิสระของทอพอโลยีเพื่อถ่วงดุลไม่ให้ทอพอโลยีที่ประมาณได้มีความเหมาะกับข้อมูลฝึกฝนจนเกินไป แต่อย่ างไรก็ดี Biem และคณะ [36] ให้ข้อสังเกตว่า แบบจำลองฮิดเดนมาร์คอฟนั้นเป็นแบบจำลองที่มีจำนวนพารามิเตอร์มาก ดังนั้น เราจึงควรให้ความสนใจถึงความสำคัญของพารามิเตอร์แต่ละพารามิเตอร์อีกทั้งยังควรพิจารณาถึงลักษณะเฉพาะของพาร ามิเตอร์แต่ละพารามิเตอร์อีกด้วย Biem จึงเสนอเกณฑ์สารสนเทศแบบเบส์สำหรับแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model Oriented Bayesian Information Criterion: HBIC) ซึ่งเป็นฟังก์ชันวัตถุประสงค์ที่พัฒนาต่อจาก BIC โดย HBIC จะมีการคำนวณความเหมาะจาก BIC และเพิ่มพจน์อีก 2 พจน์ ที่คำนวณจากพารามิเตอร์แต่ละตัวในทอพอโลยีโดยตั้งสมมุติฐานว่า พารามิเตอร์ทุกตัวของทอพอโลยีไม่ขึ้นต่อกันแล้วจำลองพารามิเตอร์แต่ละตัวให้มีการแจกแจงแบบต่างๆ ผลที่ได้ คือ เมื่อพิจารณาเฉพาะการเชื่อมต่อของสถานะแบบ LTR แล้ว ทอพอโลยีที่ประมาณได้จาก HBIC มีจำนวนพารามิเตอร์น้อยกว่า BIC แต่กลับมีความแม่นยำมากกว่า นอกจากนี้ยังมีฟังก์ชันวัตถุแบบอื่นอีก เช่น การแปรผันสำหรับเบส์ (Variational Bayesian: VB) [34] เกณฑ์สารสนเทศเลือกปฏิบัติ (Discriminative Information Criterion: DIC) [37] เป็นต้น

2.2.3.2 วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ

เนื่องจากรูปแบบของทอพอโลยีมีจำนวนไม่จำกัด ดังนั้น เราจึงจำเป็นที่จะต้องเลือกทอพอโลยีที่คาดว่าน่าสนใจมาทดสอบ ซึ่งวิธีการผลิตทอพอโลยีแบบที่นิยมใช้กันคือ กำหนดรูปแบบการเชื่อมต่อของสถานะเป็นรูปแบบใดรูปแบบหนึ่ง ดังในงานวิจัยของ Siriboon และคณะ [33] Danfang และคณะ [32] และ Biem และคณะ [36] ประมาณทอพอโลยีโดยกำหนดรูปแบบ

ของการเชื่อมต่อเป็นแบบ ซ้ายขวา (Left-to-right: LTR) แบบซ้ายขวาซ้าย (Left-right-left: LRL) แบบทั่วถึงจากซ้ายไปขวา (Fully-connected Left-to-right: FLTR) และแบบทั่วถึง (Fully-connected: FC) โดยสำหรับงานวิจัยของ Siriboon เลือกใช้การเชื่อมต่อของสถานะแบบ LTR LRL และ FC โดยคำนวณความเหมาะของทอพอโลยีโดยใช้ ML ชี้ให้เห็นว่า สำหรับการรู้จำตัวอักษรภาษาไทยแบบออนไลน์ทอพอโลยีแบบ LTR ไม่ใช่ทอพอโลยีที่ดีสุด ส่วนงานวิจัย Danfang ประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟโดยสนใจเฉพาะการเชื่อมต่อสถานะแบบ FLTR เพื่อเปรียบเทียบ BIC กับฮิสโตแกรม Biem ได้เลือกรูปแบบการเชื่อมต่อสถานะแบบ LTR เพื่อทดสอบฟังก์ชันวัตถุประสงค์ BIC และ HBIC แต่อย่างไรก็ดี เราก็ไม่ควรเลือกผลิตเฉพาะทอพอโลยีโดยการกำหนดรูปแบบการเชื่อมต่อของสถานะ เนื่องจาก การเปลี่ยนสถานะของทอพอโลยีถูกกำหนดโดยเมตริกการเปลี่ยนสถานะ ดังนั้น หากเรามีวิธีการปรับเปลี่ยนเมตริกสถานะแล้วการเชื่อมต่อของสถานะก็สามารถทำได้ โดย Singer และคณะ [30] Takatoshi และคณะ [35] และ Jitsuhiro และคณะ [34] ได้ใช้วิธีการผลิตทอพอโลยีแบบแบ่งสถานะต่อเนื่อง (Successive State Splitting: SSS) ซึ่งเป็นการผลิตทอพอโลยีโดยเริ่มจาก 1 สถานะ แล้วค่อยๆ เพิ่มจำนวนสถานะไปเรื่อยๆ ด้วยการแบ่งสถานะตามวิธีที่กำหนดมาใช้ในการประมาณทอพอโลยีควบคู่ไปกับการใช้ฟังก์ชันวัตถุประสงค์แบบ ML MDL และ VB ตามลำดับ ส่วน Vasko และคณะ [31] ได้เสนอวิธีการผลิตในทางตรงข้ามกับ SSS โดยให้เริ่มต้นจากทอพอโลยีแบบ FC แล้วค่อยๆ ลดการเปลี่ยนสถานะและสถานะลงเรื่อยๆ จนกระทั่งไม่สามารถลดสถานะและการเปลี่ยนสถานะได้อีก ในวิทยานิพนธ์นี้ขอใช้คำว่า ขั้นตอนวิธีการลดสถานะ (State Reduction Algorithm: SRA) เพื่อกล่าวถึงวิธีการผลิตทอพอโลยีที่ Vasko เสนอ วิธีการผลิตทอพอโลยีทั้งสองแบบนี้ เป็นวิธีการผลิตทอพอโลยีที่ยอมให้ฟังก์ชันวัตถุประสงค์ปรับเปลี่ยนสมาชิกในเมตริกการเปลี่ยนสถานะได้อย่างอิสระ

จากงานวิจัยที่เกี่ยวข้องได้มีการนำเสนอฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยีในแบบต่างๆ ซึ่งจะเห็นว่า ฟังก์ชันวัตถุประสงค์จะคำนวณความเหมาะสมของหน่วยเสียงที่เรากำลังสนใจเท่านั้น แต่เมื่อนำไปใช้จริงแล้วทอพอโลยีเหล่านี้จะถูกนำมาต่อกันเพื่อรู้จำเสียงที่ต่อเนื่องซึ่งเราไม่สามารถบอกได้ว่า การประมาณทอพอโลยีเพื่อค้นหาทอพอโลยีที่เหมาะสมทีละหน่วยเสียงจะให้ประสิทธิภาพดีเมื่อนำไปใช้รู้จำเสียงต่อเนื่อง ในวิทยานิพนธ์นี้จึงจะนำเสนอวิธีการค้นหาทอพอโลยีที่คำนึงถึงความเหมาะสมโดยรวมของทอพอโลยีที่มีต่อหน่วยเสียงทั้งหมด โดยจะแบ่งขั้นตอนออกเป็นสองส่วน คือ ส่วนแรกจะเป็นการประมาณทอพอโลยีเพื่อสร้างเซตของทอพอโลยีที่มีแนวโน้มว่า เมื่อนำไปประกอบเป็นเครื่องรู้จำแล้วจะมีโอกาสทำให้ความแม่นยำโดยรวมดีขึ้นสำหรับแต่ละหน่วยเสียงและส่วนที่สองคือ การเลือกทอพอโลยีจากเซตของทอพอโลยีที่

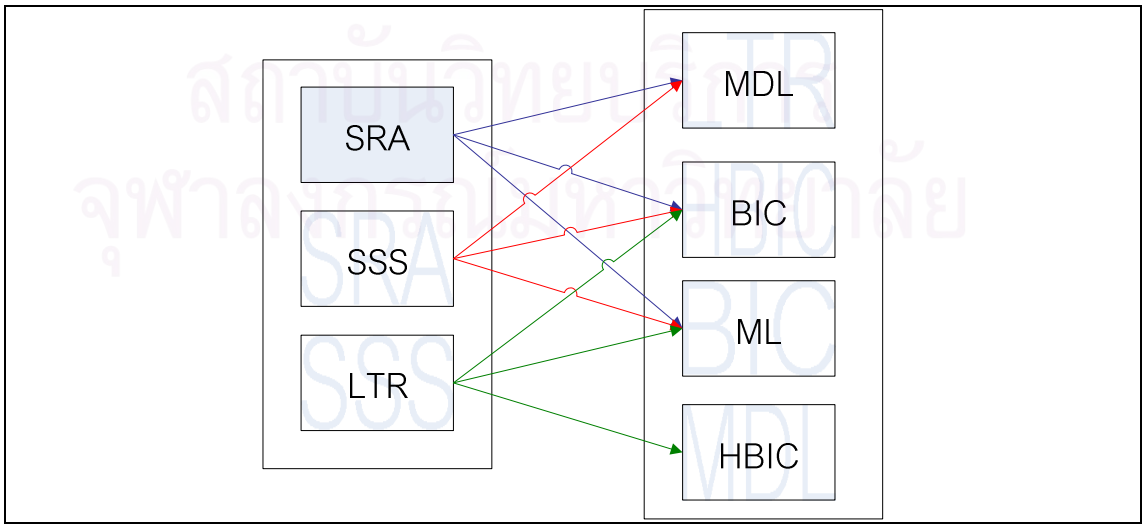
ประมาณได้ในขั้นตอนแรกเพื่อนำมาสร้างเครื่องรู้จำโดยขั้นตอนที่สองนี้จะประยุกต์ใช้ขั้นตอนวิธีเชิงพันธุกรรมเพื่อเลือกทอพอโลยีที่เหมาะสมสำหรับแต่ละหน่วยเสียงโดยคำนึงถึงความเหมาะสมโดยรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3
การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ

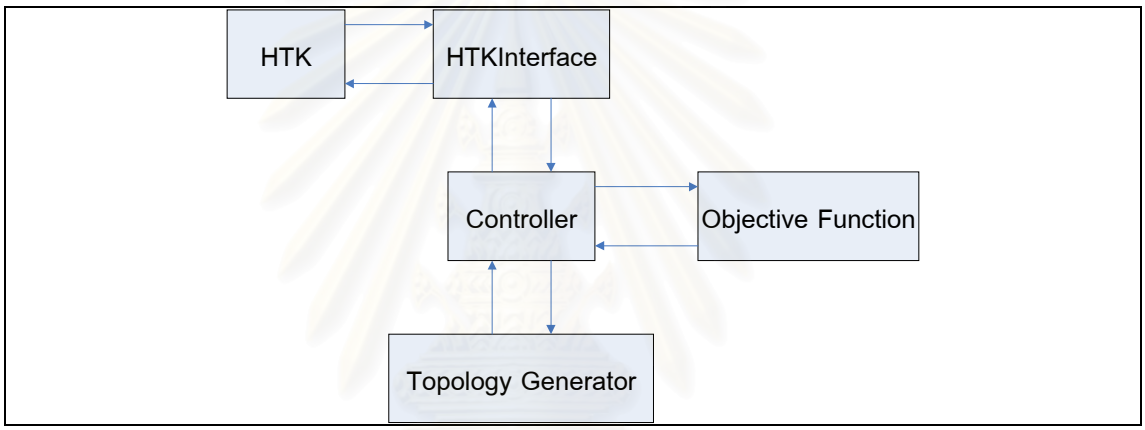
3.1 การออกแบบระบบ

ดังจะเห็นได้จากหัวข้องานวิจัยที่เกี่ยวข้องว่า การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟโดยการใช้ฟังก์ชันวัตถุประสงค์ควบคู่ไปกับวิธีการผลิตทอพอโลยีนั้น ฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยีหลายแบบได้ถูกเสนอขึ้นมาเพื่อนำมาใช้ในการประมาณทอพอโลยีที่เหมาะสมในการรู้จำต่างๆ เช่น การรู้จำเสียงพูดในภาษาญี่ปุ่นหรือภาษาอังกฤษ การรู้จำตัวอักษรลายมือเขียนแบบออนไลน์ แต่ในวิทยานิพนธ์นี้สนใจที่จะนำฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยีมาประยุกต์ใช้ในการรู้จำหน่วยเสียงภาษาไทย เนื่องจากเรายังไม่ทราบว่า ฟังก์ชันวัตถุประสงค์และวิธีการผลิตทอพอโลยีแบบใดมีความเหมาะสมกับการรู้จำหน่วยเสียงในภาษาไทยมากที่สุด ดังนั้น ในบทนี้จึงจะทำการเปรียบเทียบวิธีการประมาณทอพอโลยีโดยเลือกฟังก์ชันวัตถุประสงค์ 4 แบบคือ ความเป็นไปได้มากสุด (Maximum Likelihood: ML) เกณฑ์สารสนเทศแบบเบส์ (Bayesian Information Criterion: BIC) รายละเอียดประกอบความยาวน้อยสุด (Minimum Description Length :MDL) และเกณฑ์สารสนเทศแบบเบส์สำหรับแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model Oriented Bayesian Information Criterion: HBIC) และวิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ 3 แบบด้วยกัน คือ การแบ่งสถานะต่อเนื่อง (Successive State Splitting: SSS) ขั้นตอนวิธีการลดสถานะ (State Reduction Algorithm: SRA) และการกำหนดรูปแบบของทอพอโลยีไว้ล่วงหน้าโดยจะใช้การเชื่อมต่อแบบซ้ายไปขวา (Left-to-right Architecture: LTR) เป็นตัวแทนของวิธีนี้ มาจับคู่กันเพื่อประมาณทอพอโลยีที่เหมาะสมสำหรับการรู้จำหน่วยเสียงในภาษาไทย ดังแสดงในรูปที่ 16



รูปที่ 16 โครงงานการทดลอง

จากรูปที่ 16 เราจะมีวิธีการประมาณทอพอโลยีทั้งหมด 9 แบบ คือ ML-LTR ML-SSS ML-SRA BIC-LTR BIC-SSS BIC-SRA MDL-LTR MDL-SSS MDL-SRA และ LTR-HBIC เนื่องจาก HBIC ไม่สามารถนำไปใช้กับ SRA และ SSS ได้ ซึ่งจะกล่าวในภายหลัง และไม่ประมาณทอพอโลยีด้วยวิธี MDL-LTR เนื่องจากทั้ง MDL และ BIC เป็นฟังก์วัตถุประสงค์ที่คล้ายกันอีกทั้งการผลิตทอพอโลยีโดยกำหนดการเชื่อมต่อของสถานะเป็น LTR เป็นการผลิตทอพอโลยีที่ค่อนจำกัดความอิสระของฟังก์ชันวัตถุประสงค์ ดังนั้นจึงขอเลือกใช้ BIC เพื่อเป็นตัวแทนของฟังก์ชันวัตถุประสงค์ที่ประยุกต์ใช้หลักการมีดโกนออคแคม เพื่อให้การทำงานระหว่างฟังก์ชันวัตถุประสงค์และการผลิตทอพอโลยีเพื่อให้ระบบมีขนาดเล็กและง่ายต่อประยุกต์ใช้ อีกทั้งยังสามารถรองรับวิธีการผลิตทอพอโลยีหรือฟังก์ชันวัตถุประสงค์ที่ต้องการทดสอบในภายหน้าได้ จึงออกแบบระบบที่มีส่วนประกอบต่างๆ ตามรูปที่ 17



รูปที่ 17 ระบบการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ

ระบบมีขั้นตอนการทำงานสำหรับการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟดังนี้

1. Controller ส่งพารามิเตอร์เริ่มต้นต่างๆ เช่น จำนวนสถานะมากสุด จำนวนส่วนผสมเกาส์เซียน จำนวนผลการสังเกตของลำดับการสังเกตเฉลี่ย จำนวนตัวอย่างเสียงของแต่ละหน่วยเสียงและหน่วยเสียงที่ต้องการประมาณทอพอโลยีเพื่อให้ Topology Generator ผลิตทอพอโลยีที่น่าสนใจ
2. Topology Generator ส่งรายการของทอพอโลยีกลับมาที่ Controller
3. Controller ส่งรายการของทอพอโลยีไปที่ HTKInterface เพื่อสร้างชุดคำสั่งในการฝึกฝนแบบจำลองเสียงพูดโดยใช้ทอพอโลยีที่กำหนด
4. HTKInterface สั่งงานให้ HTK ทำการฝึกฝนแบบจำลองเสียงพูด และส่งรายการของแบบจำลองเสียงพูดที่ได้รับการฝึกฝนแล้วกลับมาที่ Controller

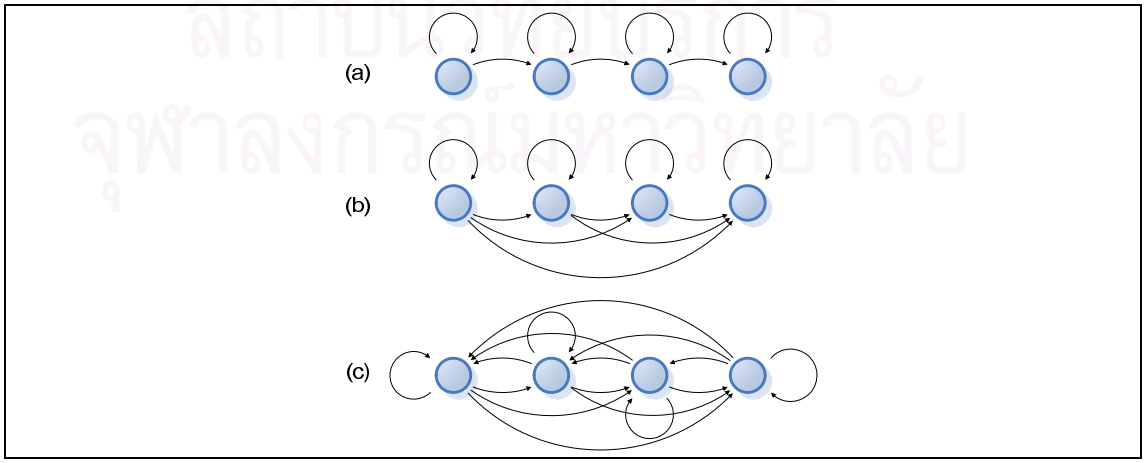
5. Objective function รับรายการแบบจำลองเสียงพูดที่ฝึกฝนแล้วไปคิดคำนวณค่าความเหมาะสม
6. Controller รับผลการประมาณความเหมาะสมของทอพอโลยี จากนั้นสร้างพารามิเตอร์เริ่มต้นเพื่อประมาณทอพอโลยีในการวนซ้ำรอบต่อไป
7. วนซ้ำข้อ 1 – 6 จนกระทั่งเข้าสู่เงื่อนไขการหยุดการทำงาน เช่น เซตของทอพอโลยีที่ผลิตจาก Topology Generator เป็นเซตว่างหรือจำนวนสถานะเท่ากับจำนวนสถานะมากที่สุดที่กำหนดไว้ เป็นต้น

3.2 วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ

เนื่องจากรูปแบบของทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟมีจำนวนไม่จำกัด ดังนั้น จึงเป็นไปไม่ได้ที่จะนำทอพอโลยีทุกรูปแบบมาทดสอบด้วยฟังก์ชันวัตถุประสงค์ วิธีการ ผลิตทอพอโลยีจึงเป็นวิธีที่นำมาใช้เพื่อคัดเลือกทอพอโลยีเพื่อนำไปทดสอบ ด้วยสมมุติฐานที่ว่า จะสามารถพบพอทอโลยีที่เหมาะสมในระหว่างกระทำการขั้นตอนวิธี วิธีการผลิตทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟมีดังนี้

3.2.1 การกำหนดรูปแบบการเชื่อมต่อสถานะล่วงหน้า

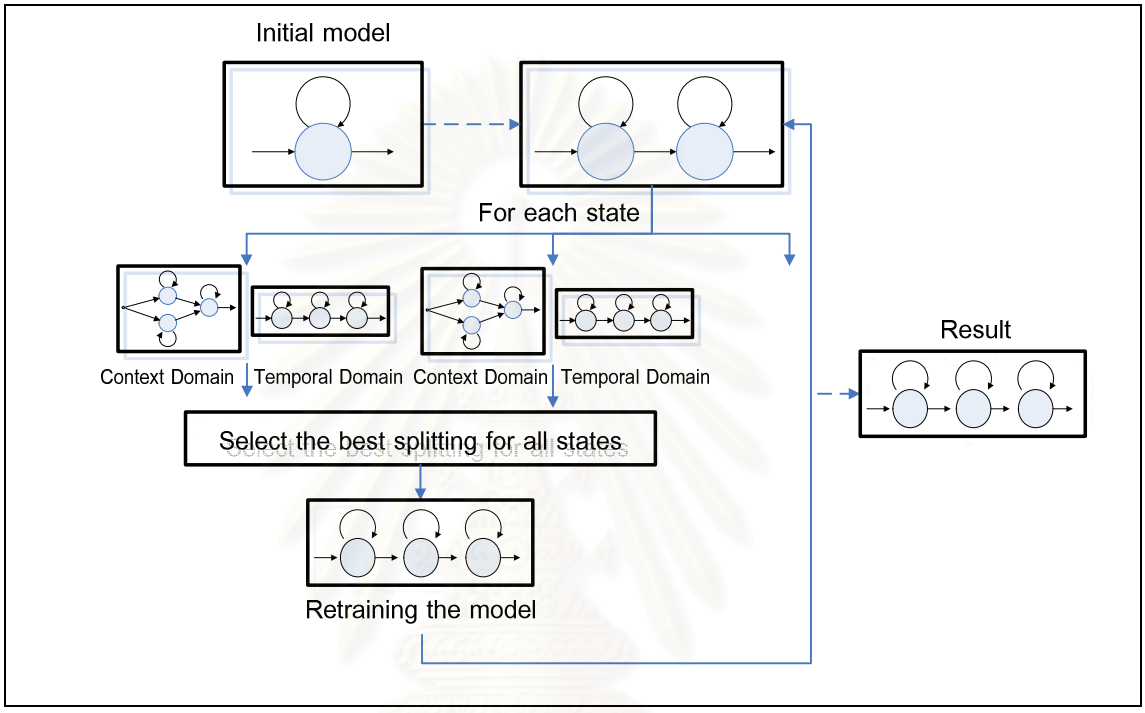
รูปแบบการเชื่อมต่อระหว่างสถานะที่นิยมใช้มี 3 แบบ คือ การเชื่อมต่อแบบซ้ายไปขวา (Left-to-right Architecture) รูปที่ 18a การเชื่อมต่อแบบทั่วถึงจากซ้ายไปขวา (Fully-left-to-right Architecture) รูปที่ 18b และ การเชื่อมต่อแบบทั่วถึง (Fully-connected Architecture) รูปที่ 18c ดังแสดงในรูปที่ 18 ซึ่งเมื่อรูปแบบการเชื่อมต่อของสถานะถูกกำหนดแล้ว ปัญหาการประมาณพอโลยีจึงเหลือเพียงการประมาณจำนวนสถานะกับจำนวนส่วนผสมเกาส์เซียน ในที่นี้จะใช้ LTR เป็นตัวแทนของวิธีการผลิตทอพอโลยีแบบนี้



รูปที่ 18 ตัวอย่างการเชื่อมต่อของสถานะแบบต่างๆ

3.2.2 วิธีแบ่งสถานะต่อเนื่อง

วิธีการกำหนดรูปแบบการเชื่อมต่อไว้ล่วงหน้าเป็นวิธีหนึ่งที่สามารถนำมาใช้คัดเลือกทอพอโลยีเพื่อนำมาทดสอบได้ แต่อย่างไรก็ดี ทอพอโลยีที่เหมาะสมนั้นอาจจะมีการเชื่อมต่อของสถานะในรูปแบบที่ไม่สามารถกำหนดได้ชัดเจนก็เป็นได้ ดังนั้น Singer และคณะ [30] จึงเสนอวิธีการผลิตทอพอโลยีที่มีการทำงานดังรูปที่ 19



รูปที่ 19 แผนภูมิสายงานขั้นตอนการแบ่งสถานะต่อเนื่อง (รูปดัดแปลงจาก [30])

การผลิตทอพอโลยีโดยวิธีแบ่งสถานะต่อเนื่องจะเริ่มจากทอพอโลยีที่มีหนึ่งสถานะและมีเพียงการเปลี่ยนสถานะอัตตา (Self Transition) ของสถานะดังกล่าว ในแต่ละรอบการวนซ้ำก็ให้ทอพอโลยีเพิ่มจำนวนสถานะและการเปลี่ยนสถานะขึ้นเรื่อยๆ โดยให้แต่ละสถานะแบ่งตัวแบบเทมพอรัล (Temporal) และแบบบริบท (Contextual) จากนั้น คำนวณความเหมาะสมของทอพอโลยีทั้งหมดที่ผลิตมา แล้วเลือกทอพอโลยีที่ดีสุดตามฟังก์ชันวัตถุประสงค์ที่เลือกใช้ เป็นทอพอโลยีเริ่มต้นของการวนซ้ำในรอบต่อไป

3.2.3 ขั้นตอนวิธีการลดสถานะ

วิธีการผลิตทอพอโลยีแบบแบ่งสถานะต่อเนื่องเป็นการผลิตทอพอโลยีโดยเริ่มจากรูปแบบของทอพอโลยีที่มีขนาดเล็กที่สุดเท่าที่จะเป็นได้ แล้วในแต่ละรอบการทำงานทอพอโลยีจะค่อยๆ เพิ่มจำนวนสถานะเรื่อยๆ จนกระทั่งได้จำนวนสถานะตามต้องการ ในทางกลับกัน Vasko

และคณะ [31] ก็ได้เสนอวิธีการผลิตทอพอโลยีในทางตรงข้าม นั่นคือ ให้เริ่มต้นการผลิตทอพอโลยีจากทอพอโลยีที่มีขนาดใหญ่ที่สุดแล้วในแต่ละรอบการวนซ้ำก็ให้ทดลองลบ การเปลี่ยนสถานะแต่ละอันออกแล้วเลือกทอพอโลยีที่มีความเหมาะสมมากสุดเป็นทอพอโลยีเริ่มต้นในรอบการวนซ้ำต่อไป ซึ่งวิธีของ Vasko นั้นสามารถสรุปได้ดังแสดงในรูปที่ 20

ขั้นตอนวิธีการลดสถานะโดยเริ่มจากการเชื่อมต่อแบบทั่วถึง

1. ฝึกฝนโมเดลทั่วไปโดยใช้ทอพอโลยีแบบติดต่อทั่วถึงด้วยการประมาณซ้ำตามสูตรบอมเวลช์
2. สำหรับแต่ละการวนซ้ำ ทุกๆ การเปลี่ยนสถานะจะถูกตั้งค่าความน่าจะเป็นของการเปลี่ยนสถานะให้เป็น 0 เพื่อที่จะตรวจสอบว่า การเปลี่ยนแปลงของสถานะใดจะถูกลบออกในรอบวนซ้ำนั้นๆ โดยใช้ฟังก์ชันวัตถุประสงค์ที่กำหนดเป็นเกณฑ์ในการเลือก ซึ่งการลบการเปลี่ยนแปลงของสถานะออกนั้นจะเป็นไปได้ 2 กรณี คือ
    a. ถ้าการเปลี่ยนสถานะนั้นเป็นการเปลี่ยนสถานะอัตตาก็จะลบสถานะนั้นออกไปด้วย นั่นคือ ลบทุกการเปลี่ยนสถานะที่เกี่ยวข้องกับสถานะนั้นออกไปและเพิ่มการเปลี่ยนสถานะ หากสถานะที่เกี่ยวข้องมีวิถีเชื่อมถึงกันโดยผ่านสถานะที่ถูกลบออก
    b. ถ้าเป็นการเปลี่ยนแปลงของสถานะที่ไม่ใช่การเปลี่ยนสถานะอัตตาก็ลบออกไปธรรมดา (ตั้งค่าความน่าจะเป็นของการเปลี่ยนสถานะให้เป็น 0)
3. ประมาณซ้ำทอพอโลยีที่ได้ด้วยการประมาณซ้ำตามสูตร Baum-Welch
4. ทำซ้ำในข้อ 2 และ 3 จนกระทั่งลบการเปลี่ยนสถานะครบทุกอัน
5. เลือกทอพอโลยีที่ให้ค่าความเป็นไปได้สูงสุดจากผลในการวนซ้ำนั้น
6. กลับไปทำขั้นตอนที่ 2 จนกระทั่งไม่สามารถเลือกลบการเปลี่ยนสถานะได้

รูปที่ 20 ขั้นตอนวิธีการลดสถานะโดยเริ่มจากการเชื่อมต่อแบบทั่วถึง

3.3 **ฟังก์ชันวัตถุประสงค์**

ฟังก์ชันวัตถุประสงค์ คือ ฟังก์ชันที่ถูกตั้งสมมุติฐานว่า สามารถบอกความเหมาะสมของทอพอโลยีได้เมื่อใส่พารามิเตอร์ตามที่ฟังก์ชันวัตถุประสงค์นั้นๆ ต้องการ ฟังก์ชันวัตถุประสงค์ที่ใช้โดยส่วนมากจะคำนวณโดยวัดจากค่าความเป็นไปได้ของแบบจำลองฮิดเดนมาร์คอฟที่มีต่อชุดข้อมูลฝึกฝนเป็นหลัก โดยค่านี้จะบ่งบอกถึงความสามารถที่แบบจำลองฮิดเดนมาร์คอฟที่สนใจสามารถปรับค่าพารามิเตอร์ต่างๆ ให้เข้ากับฐานข้อมูลที่นำมาใช้ฝึกฝนมากเพียงไร รวมกับส่วนอื่นๆ ที่เพิ่มเข้าตามแต่ละงานวิจัย ฟังก์ชันวัตถุประสงค์ที่นิยมใช้กันมีดังนี้

3.3.1 ความเป็นไปได้มากสุด

ความเป็นไปได้หรือความคล้ายมากสุดเป็นฟังก์ชันวัตถุประสงค์ที่วัดความเหมาะสมของทอพอโลยีโดยอาศัยค่าความเป็นไปได้มากสุด เมื่อนำแบบจำลองฮิดเดนมาร์คอฟที่สนใจไปฝึกฝนด้วยฐานข้อมูลที่กำหนดไว้ โดยความ ความเหมาะสมของทอพอโลยีจะแปรผันตรงกับค่าความเป็นไปได้มากสุด กล่าวคือ หากค่าแบบจำลองใดมีค่า ความเป็นไปได้มากสุดมาก ก็จะถือว่าแบบจำลองนั้นมีความเหมาะสม การคำนวณความเหมาะสมโดยยึดถือตามความเป็นไปได้มากสุดสามารถคำนวณได้ดังสมการ

$$ML(\theta_i) = \log P(X_n \mid \theta_i) \tag{3.1}$$

เมื่อมีเซตของของแบบจำลอง $\{\theta_i \mid i = 1,2,\cdots,I\}$ และเซตของลำดับการสังเกต $\{X_n \mid n = 1,2,\cdots,N\}$

ความเป็นไปได้มากสุด เป็นฟังก์ชันวัตถุประสงค์ที่จะมีประสิทธิภาพ ในทางทฤษฎี เมื่อข้อมูลที่นำมาใช้ฝึกฝนกับข้อมูลที่ต้องการรู้จำมีความใกล้เคียงกัน แต่ในทางปฏิบัติแล้ว ทอพอโลยีที่ได้จากการใช้ความเป็นไปได้มากสุดเป็นฟังก์ชันวัตถุประสงค์จะพบกับปัญหา การปรับเหมาะเกินไป (Overfitting) เนื่องจาก ทอพอโลยีที่ประมาณได้นั้นมีความเหมาะกับข้อมูลที่นำมาฝึกฝนมากจนเกินไป ดังนั้น การใช้วัตถุประสงค์แบบนี้จำเป็นต้องกำหนดข้อบังคับเพิ่มเติม เช่น จำกัดรูปแบบของทอพอโลยีที่ต้องการค้นหาหรือกำหนดรูปแบบการเชื่อมของสถานะต่อไว้ล่วงหน้า เพื่อหลีกเลี่ยงปัญหาการปรับเหมาะเกินไป

3.3.2 หลักการมีดโกนออคแคม (The Occam's Razor Principle)

มีดโกนออคแคมเป็นหลักการที่ถูกเสนอโดย Franciscan friar William of Ockham เมื่อศตวรรษที่ 14 กล่าวว่า "*Entia non sunt multiplicanda praeter necessitatem*" แปลเป็นภาษาอังกฤษได้ว่า "Entities should not be multiplied beyond necessity" เป็นหลักการเพื่ออธิบายเกี่ยวกับปรากฏการณ์ทางธรรมชาติ โดยเราควรที่จะตั้งสมมุติฐานเกี่ยวกับปรากฏการณ์นั้นๆ ให้น้อยที่สุดเท่าที่จะเป็นไปได้ และควรกำจัดสมมุติฐา นที่มองไม่เห็นในข้อมูลที่สังเกตออกหรือหากมองอีกมุมหนึ่ง มีดโกนออคแคมเสนอว่า หากมีสิ่งที่ให้ผลลัพธ์เหมือนกันสิ่งที่มีความเรียบง่ายกว่าน่าจะดีกว่า

หลักการนี้สามารถนำประยุกต์ใช้กับการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟได้ ดังที่กล่าวไปแล้วว่า การจำลองเหตุการ ณ์ด้วยแบบจำลองฮิดเดนมาร์คอฟนั้น ควร

จะตั้งสมมุติฐานให้ใกล้เคียงกับความเป็นจริงมากที่สุดและในทางปฏิบัติแบบจำลองที่มีจำนวนพารามิเตอร์มากจนเกินไป เนื่องจาก เราตั้งสมมุติฐานเกินความเป็นจริงก็ไม่ได้สามารถจำลองเหตุการณ์ที่เกิดขึ้นจริงได้ดีไปกว่าการตั้งสมมุติฐานที่ใกล้เคียงความเป็นจริง อีกทั้งแบบจำลองที่ มีขนาดเล็ก นั่นคือ มีจำนวนพารามิเตอร์น้อยก็มีผลดี เช่น เมื่อนำไปใช้ก็สามารถคำนวณได้อย่างรวดเร็วและประหยัดเนื้อที่จัดเก็บ เป็นต้น เราคาดหวังได้ว่า การที่ทอพอโลยีไม่มีความซับซ้อนจนเกินไปจะทำให้ทอพอโลยีที่ประมาณได้นั้นสามารถรับมือกับข้อมูลรบกวน (Noise) ที่ผสมกับมาเสียงพูดได้ นั่นคือ ฟังก์ชันวัตถุประสงค์ที่ประยุกต์ใช้หลักการมีดโกนออคแคมน่าจะนำพาการค้นหาทอพอโลยีไปสู่ทอพอโลยีที่เหมาะในรูปแบบที่ไม่เหมาะกับข้อมูลฝึกฝนมากนัก เพราะว่า ทอพอโลยีที่เรียบง่ายย่อมมีความสามารถในการแยกแยะความแตกต่างได้น้อยกว่าทอพอโลยีที่มีความซับซ้อน แต่อย่างไรก็ดีความแตกต่างบางอย่าง เช่น ผู้พูด วิธีการพูด อารมณ์ของผู้พูด เป็นต้น ความแตกต่างเหล่านี้ไม่ได้เป็นตัวกำหนดว่าหน่วยเสียงที่กำลังรู้จำนี้คือ หน่วยเสียงใด ฉะนั้นเราก็ไม่มีความจำเป็นที่จะต้องแยกแยะความแตกต่างเหล่านี้ ทำให้การค้นหาทอพอโลยีโดยตั้งสมมุติฐานว่า ทอพอโลยีที่เหมาะสมควรเป็นทอพอโลยีที่ไม่ซับซ้อน น่าจะทำให้เราค้นพบทอพอโลยีที่เหมาะกับหน่วยเสียงและสามารถรองรับความหลากหลายของเสียงที่ต้องการรู้จำได้มากกว่าการคิดคำนวณจากความสามารถในการปรับตัวของทอพอโลยีให้เข้ากับข้อมูลชุดฝึกฝนเพียงอย่างเดียวแบบความเป็นไปได้มากสุด ฟังก์ชันวัตถุประสงค์ที่นำหลักการมีดโกนของออคแคมไปประยุกต์ใช้ มี 2 ฟังก์ชัน คือ

3.3.2.1 รายละเอียดประกอบความยาวน้อยสุด

MDL เป็นฟังก์ชันวัตถุประสงค์ที่นำหลักการมีดโกนออคแคม มาประยุกต์ใช้ MDL ถูกเสนอในงานวิจัยของ Jitsushiro และ คณะ [34] โดยคำนวณค่าความเหมาะสมของทอพอโลยีจาก 2 ส่วน คือ ส่วนความเป็นไปได้มากสุดและส่วนหักลบซึ่งคำนวณโดยใช้จำนวนพารามิเตอร์และจำนวนผลการสังเกตของลำดับการสังเกตของหน่วยเสียงที่สนใจดังแสดงในสมการที่ 3.2

$$MDL(\theta_i) = -\log P(X_n \mid \hat{\theta}_i) + \frac{\alpha_i}{2}\log(N_T) + \log I \qquad (3.2)$$

เมื่อกำหนดให้มีเซตของแบบจำลอง $\{\theta_i \mid i = 1,2,\cdots I\}$ $\hat{\theta}_i$ คือ เซตของพารามิเตอร์ที่มีค่าความเป็นไปได้ (Likelihood) สูงสุดสำหรับแบบจำลอง θ_i เทียบกับข้อมูลชุดฝึกฝนที่ใช้ในการฝึกฝน X_n คือ เซตของลำดับการสังเกต α_i คือ จำนวนพารามิเตอร์อิสระทั้งหมดของแบบลอง θ_i และ N_T คือ จำนวนผลการสังเกตของลำดับการสังเกตที่เราสนใจ จะเห็นว่าพจน์ $\frac{\alpha_i}{2}\log(N_T)$ เปรียบเสมือนพจน์หักลบ ซึ่งเมื่อแบบจำลองมีจำนวนพารามิเตอร์มากขึ้น ค่า

ของพจน์นี้ก็จะมากขึ้นด้วย ทำให้เกิดการถ่วงดุลระหว่างค่าความเป็นไปได้สูงสุดและค่าทำโทษเมื่อจำนวนพารามิเตอร์มากขึ้น กล่าวคือ เมื่อทอพอโลยีมีขนาดใหญ่ขึ้นก็จะสามารถเรียนรู้รายละเอียดของตัวอย่างเสียงได้มากขึ้น นั่นคือทอพอโลยีที่มีขนาดใหญ่ขึ้นจะมีความเหมาะกับข้อมูลที่ใช้ฝึกฝนมากขึ้น แต่ก็จะถูกหักลบคะแนนความเหมาะจากพจน์ทำโทษมากขึ้นเช่นเดียวกัน ในทางทฤษฎีแล้วฟังก์ชันวัตถุประสงค์แบบ MDL จึงน่าจะมีโอกาสเกิดปัญหา การปรับเหมาะเกินไปน้อยกว่า ML ที่คำนวณความเหมาะสมของทอพอโลยีจากค่าความสามารถในการปรับตัวของแบบจำลองให้เข้ากับข้อมูลที่นำมาใช้ฝึกฝนเพียงอย่างเดียว

3.3.2.2 เกณฑ์สารสนเทศแบบเบส์

Schwarz [40] ได้เสนอฟังก์ชันวัตถุประสงค์ชื่อว่า BIC และ Danfeng และคณะ [32] ได้นำมาประยุกต์ใช้กับการประมาณทอพอโลยี BIC เป็นฟังก์ชันวัตถุประสงค์อีกฟังก์ชันหนึ่งที่ประยุกต์ใช้หลักการมีดโกนออคแคมโดย BIC นั้นมีคล้ายกับ MDL เป็นอย่างมาก เพียงแต่ในการคำนวณพจน์หักลบนั้นแทนที่จะใช้จำนวนผลการสังเกตของลำดับการสังเกตดังที่ MDL ใช้ BIC กลับนำจำนวนของลำดับการสังเกตมาคิดคำนวณแทนทำให้ BIC สามารถคำนวณได้ดังนี้

$$BIC(\theta_i) = \log P(X_n \mid \hat{\theta}_i) - \frac{\alpha_i}{2}\log(N) \qquad (3.3)$$

เมื่อ กำหนดให้ $\hat{\theta}_i$ คือ เซตของพารามิเตอร์ที่มีค่าความเป็นไปได้สูงสุดสำหรับแบบจำลอง θ_i X_n คือ เซตของลำดับการสังเกต α_i คือ จำนวนพารามิเตอร์อิสระทั้งหมดของแบบลอง θ_i และ N คือ จำนวนลำดับการสังเกตที่เราสนใจ

3.3.3 เกณฑ์สารสนเทศแบบเบส์สำหรับแบบจำลองฮิดเดนมาร์คอฟ

ถึงแม้ว่า MDL และ BIC จะเพิ่มพจน์หักลบเพื่อนำมาช่วยในการวัดความเหมาะสมของทอพอโลยี แต่การคำนวณพจน์หักลบนั้นใช้เพียงจำนวนพารามิเตอร์ข องแบบจำลองเท่านั้น จึงมีข้อด้อยที่ว่า พารามิเตอร์ทุกตัวมีความสำคัญเท่ากันกล่าวคือ ทั้ง MDL และ BIC ไม่ได้คำนึงว่า พารามิเตอร์เป็นพารามิเตอร์ชนิดใด และมีการแจกแจงอย่างไร ซึ่งแบบจำลองฮิดเดนมาร์คอฟนั้นเป็นแบบจำลองที่มีพารามิเตอร์หลายชนิดและมีพารามิเตอร์เป็นจำนวนมาก ทำให้ BIC และ MDL อาจจะเป็นฟังก์ชันวัตถุประสงค์ที่ ยังไม่เหมาะสม เพื่อแก้ปัญหานี้ Biem และคณะ [36] จึงได้เสนอฟังก์ชันวัตถุประสงค์ฟังก์ชันใหม่ โดยตั้งชื่อว่า HMM-Oriented BIC หรือ HBIC ซึ่งสามารถคำค่าความเหมาะสมของทอพอโลยีได้ดังสมการ (3.4)

$$HBIC(\theta_i) = BIC(\theta_i) + \log P(\tau) + \log P(\hat{\theta}_i \mid \tau) + \frac{\alpha_i}{2}\log(2\pi) \quad (3.4)$$

เมื่อกำหนดให้ $\hat{\theta}_i$ คือ เซตของพารามิเตอร์ที่มีค่าความเป็นไปได้สูงสุดสำหรับแบบจำลอง θ_i α_i คือ จำนวนพารามิเตอร์อิสระทั้งหมดของทอพอโลยี θ_i τ คือ เซตของจำนวนสถานะ จำนวนส่วนผสมเกาส์เซียนต่อหนึ่งสถานะ และรูปแบบการเชื่อมต่อของสถานะ สำหรับ θ และ τ สามารถเขียนเป็นสมการได้ คือ

$$\tau = \{A, \omega, \mu, \Sigma\} \quad (3.5a)$$

$$\theta = \{Q, L, C\} \quad (3.5b)$$

เมื่อ A คือ เมตริกการเปลี่ยนสถานะ ω คือ น้ำหนักส่วนผสมเกาส์เซียน μ คือ ค่ากลาง Σ คือ เมตริกความแปรปรวนร่วมเกี่ยว Q คือ จำนวนสถานะ L คือ จำนวนส่วนผสมเกาส์เซียนต่อหนึ่งสถานะ และ C คือ รูปแบบการเชื่อมต่อของสถานะ

สำหรับ HBIC นั้น ต้องการสมมุติฐานเพิ่มเติมเพื่อทำให้สามารถที่จะคำนวณพจน์ $\log P(\tau)$ และ $\log P(\hat{\theta}_i \mid \tau)$ ได้ โดยหากตั้งสมมุติฐานว่า Q และ L ไม่ขึ้นต่อกันและหากกำหนดรูปแบบการเชื่อมต่อของสถานะเอาไว้ล่วงหน้าจะทำให้สามารถคำนวณพจน์ $\log P(\tau)$ ได้โดย

$$\log P(\tau) = \log P(Q) + \log P(L) \quad (3.6)$$

และหากสมมุติว่า A, ω, μ และ Σ เป็นอิสระต่อกันพจน์ $\log P(\hat{\theta}_i \mid \tau)$ ก็สามารถคำนวณได้ ดังนี้

$$\log P(\hat{\theta}_i \mid \tau) = \log P(\hat{A}_i \mid \tau) + \log P(\hat{\omega}_i \mid \tau) + \log P(\hat{\mu}_i \mid \tau) + \log P(\hat{\Sigma}_i \mid \tau) \quad (3.7)$$

ฟังก์ชันวัตถุประสงค์ถือเป็นส่วนสำคัญในการประมาณทอพอโลยีที่เหมาะสมของแบบจำลองฮิดเดนมาร์คอฟ เนื่องจาก หากฟังก์ชันวัตถุประสงค์ใดสามารถบอกความเหมาะสมของทอพอโลยีได้ตามความเป็นจริงแล้ว การประมาณทอพอโลยีก็สามารถทำได้โดยง่าย แต่ในทางปฏิบัติแล้ว ฟังก์ชันวัตถุประสงค์ไม่สามารถบอกความเหมาะสมของทอพอโลยีได้ตรงตามความเป็นจริงทั้งหมด เพียงแต่บอกถึงค่าระดับความความเหมาะตามสมมุติฐานที่กำหนดเท่านั้น ดังนั้นการเลือกฟังก์ชันวัตถุประสงค์ให้เหมาะสมกับข้อมูลฝึกฝนและข้อมู ลขาเข้าเมื่อนำไปใช้งานจริง จึงเป็นข้อประเด็นที่มีน่าสนใจทดลองเพื่อเพิ่มประสิทธิภาพในการรู้จำเสียง

จากฟังก์ชันวัตถุประสงค์และวิธีการผลิตแบบจำลองเพื่อนำมาทดสอบ เราสามารถแบ่งการทำงานออกเป็น 2 ส่วนได้อย่างชัดเจน คือ ส่วนของ การผลิตทอพอโลยีและส่วนของการคำนวณความเหมาะสมของทอพอโลยี ดังนั้น การออกแบบระบบเพื่อให้มีการทำงานโดยทั้งสองส่วนนี้แยกกันชัดเจนก็จะทำให้การทดลองทำได้ง่ายขึ้นดังที่ได้กล่าวไปแล้วในตอนต้น

3.4 การทดลองและผลการทดลอง

3.4.1 ภาพรวมของการทดลอง

การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงภาษาไทยแบบบริบทอิสระในวิทยานิพนธ์นี้ใช้ข้อมูลเสียงชุดหน่วยเสียงสมดุลในโลตัสเพื่อประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมและฝึกฝนแบบจำลองทางเสียง โดยที่การประมาณทอพอโลยีทุกการทดลองจะใช้องค์ประกอบแบบเดียวกันแบบเดียวกันตามที่แสดงในตารางที่ 16

ตารางที่ 16 องค์ประกอบที่ใช้ในการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ

	รายละเอียด	
การสกัดค่าสำคัญ	วิธีการสกัดค่าสำคัญ	สัมประสิทธิ์บนสเกลเมลที่มีเวคเตอร์ลักษณะสำคัญขนาด 39 มิติที่คำนวณจากค่าพลังงงาน (Energy) อัตราการเปลี่ยนแปลง (Delta) และความเร่ง (Accelerations หรือ Delta-delta) อย่างละ 13 มิติ
	ขนาดเฟรม	10 มิลลิวินาที
	ความกว้างวินโดว์	25 มิลลิวินาที
หน่วยเสียง	74 หน่วยเสียงตามโลตัสรวมกับ sil (เสียงเงียบ) และ sp (การหยุดระยะสั้น)	
ตัวอย่างเสียง	ข้อมูลเสียงชุดหน่วยเสียงสมดุลในโลตัส	
ส่วนผสมเกาส์เซียน	ประมาณทอพอโลยีที่ 1 ส่วนผสมเกาส์เซียน แล้วจากนั้นเพิ่มเป็น 2 3 และ 4 ตามลำดับ	

3.4.1.1 การกำหนดจำนวนสถานะมากสุด

การกำหนดจำนวนสถานะมากสุดสำหรับการผลิตทอพอโลยีเพื่อนำมาทดสอบนั้นสามารถเลือกใช้ได้วิธี เช่น

1. กำหนดขึ้นเอง
2. ใช้ค่าเฉลี่ยของจำนวนผลการสังเกตในแต่ละลำดับการสังเกตของหน่วยเสียง
3. ค่ามากสุดของจำนวนผลการสังเกตในแต่ละลำดับการสังเกตของหน่วยเสียง

4. ค่าน้อยสุดของจำนวนผลการสังเกตในแต่ละลำดับการสังเกตของหน่วยเสียง

ในวิทยานิพนธ์นี้เลือกใช้ค่าเฉลี่ยจำนวนของผลการสังเกตแต่ในละลำดับการสังเกตสำหรับหน่วยเสียงที่สนใจในชุดข้อมูลหน่วยเสียงโดยจำนวนผลการสังเกตของลำดับการสังเกตสามารถคำนวณได้ดังสมการ

$$\mathrm{T} = \frac{\text{Duration}}{\text{Frame size}} \tag{3.7}$$

เมื่อ T คือ จำนวนผลการสังเกตของลำดับการสังเกต Duration คือ ระยะการเปล่งเสียงของตัวอย่างหน่วยเสียง และ Frame size คือ ขนาดของกรอบที่ใช้ซึ่งกำหนดในขั้นตอนการสกัดค่าลักษณะสำคัญ

3.4.1.2 การนับจำนวนพารามิเตอร์อิสระ (α)

MDL BIC และ HBIC นั้นต้องการจำนวนพารามิเตอร์อิสระของทอพอโลยีเพื่อนำไปใช้คำนวณค่าความเหมาะสมของทอพอโลยี เมื่อพิจารณาพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ เราสามารถเขียนได้ดังสมการ (3.8)

$$\lambda = \{A, B, \pi\} \tag{3.8}$$

เมื่อ A คือ ความน่าจะเป็นเปลี่ยนสถานะ B คือ การแจกแจงของความน่าจะเป็นของสัญลักษณ์การสังเกต หากเลือกใช้การแจกแจงแบบเกาส์เซียนก็จะประกอบไปด้วยค่ากลางและค่าความแปรปรวนของแต่ละส่วนผสมเกาส์เซียนและน้ำหนักส่วนผสม ดังนั้น จำนวนพารามิเตอร์อิสระของแบบจำลองฮิดเดนมาร์คอฟที่ทุกสถานะมีจำนวน ส่วนผสมเกาส์เซียน เท่ากันจึงสามารถคำนวณได้ตามสมการ (3.9)

$$\alpha = |A| + (L \times Q \times (|\Sigma| + |\mu|)) + (Q \times |\omega|) \tag{3.9}$$

เมื่อกำหนดให้ $|X|$ แทนขนาดของ X A คือ เมตริกการเปลี่ยนสถานะ ω คือ น้ำหนักส่วนผสมเกาส์เซียน μ คือ ค่ากลาง Σ คือ เมตริกความแปรปรวนร่วมเกี่ยว Q คือ จำนวนสถานะ L คือ จำนวนส่วนผสมเกาส์เซียนต่อสถานะ

3.4.1.3 การวัดประสิทธิภาพ

ในวิทยานิพนธ์นี้สร้างเครื่องรู้จำหน่วยเสียงภาษาไทยแบบบริบทอิสระเพื่อนำมาใช้วัดประสิทธิภาพของทอพอโลยีที่ประมาณได้ โดยอาศัยชุดเครื่องมือ HMM-Toolkit [39]

ซึ่งเป็นชุดเครื่องมือสำหรับสร้างเครื่องรู้จำเสียงพูดที่ใช้เทคนิดแบบจำลองฮิดเดนมาร์คอฟ รายละเอียดเกี่ยวกับขั้นตอนการเรียนรู้และการรู้จำจะอยู่ที่ภาคผนวก ค และเครื่องรู้จำเสียงพูดที่ใช้ในการทดลองมีองค์ประกอบดังตารางที่ 17

ตารางที่ 17 องค์ประกอบที่ในการสร้างเครื่องรู้จำเพื่อวัดประสิทธิภาพ

	รายละเอียด	
การสกัดค่าสำคัญ	วิธีการสกัดค่าสำคัญ	สัมประสิทธิ์บนสเกลเมลที่มีเวคเตอร์ลักษณะสำคัญขนาด 39 มิติที่คำนวณจากค่าพลังงงาน อัตราการเปลี่ยนแปลง และความเร่งอย่างละ 13 มิติ
	ขนาดเฟรม	10 มิลลิวินาที
	ความกว้างวินโดว์	25 มิลลิวินาที
หน่วยเสียง	74 หน่วยเสียงตามโลตัสรวมกับ sil และ sp	
แบบจำลองหน่วยเสียง	หน่วยเสียงแบบบริบทอิสระที่มีเมตริกความแปรปรวนร่วมเกี่ยวตามแนวทแยง (Context Independent Phoneme with Diagonal Covariance Gaussian Distribution)	
แบบจำลองภาษา	ไบแกรม (Bi-gram) ที่ฝึกฝนด้วยไฟล์กำกับหน่วยเสียงของชุดฝึกฝนในโลตัส	
ตัวอย่างเสียง	ชุดข้อมูลฝึกฝน	ข้อมูลเสียงชุดหน่วยเสียงสมดุลจากโลตัส
	ชุดข้อมูลประเมินผล	ข้อมูลเสียงชุดทดสอบเพื่อพัฒนาจากโลตัส

การวัดประสิทธิภาพของเครื่องรู้จำนั้น สามารถคำนวณได้ตามสูตร

$$\%\text{Correct} = \frac{H}{N} \times 100\% \tag{3.10a}$$

$$\text{Accuracy} = \frac{H - I}{N} \times 100\% \tag{3.10b}$$

เมื่อ H คือ จำนวนของหน่วยเสียงที่ตรงกับไฟล์กำกับหน่วยเสียงอ้างอิง I คือ จำนวนหน่วยเสียงแทรก (Insertion) ที่เกิดขึ้นเมื่อเทียบกับไฟล์กำกับหน่วยเสียงอ้างอิง S คือ จำนวนหน่วยเสียงแทนที่ (Substitution) ที่เกิดขึ้นเมื่อเทียบกับไฟล์กำกับหน่วยเสียงอ้างอิงและ N คือ จำนวนหน่วยเสียงทั้งหมด

3.4.2 การทดลองและผลการทดลอง

ในปัจจุบันประสิทธิภาพของเครื่องรู้จำเสียงพูดในภาษาไทยแบบไม่ขึ้นกับบริบทรอบข้าง (Context independent phoneme Model) ที่ใช้การแจกแจงของความน่าจะเป็นแบบ

เกาส์เซียนที่มีเมตริกความแปรปรวนร่วมเกี่ยวตามแนวทแยง (Diagonal Covariance) เมื่อใช้ข้อมูลชุดหน่วยเสียงสมดุลในฐานข้อมูลเ สียงโลตัส มีความแม่นยำเพียง 48% [17] ซึ่งถือว่าอยู่ในระดับที่ค่อนข้างต่ำ หากเปรียบเทียบประสิทธิภาพของเครื่องรู้จำภาษาอื่นๆ ในแบบเดียวกัน เช่น ภาษาญี่ปุ่นมีความ ซึ่งมีความแม่นยำประมาณ 55-60%

การทดลองการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟในที่นี้จะทดลองโดยการจับคู่ฟังก์ชันวัตถุประสงค์กับวิธีการ ผลิตทอพอโลยีตามที่ได้นำเสนอไปในส่วนก่อนหน้านี้ เมื่อแบ่งการทดลองตามฟังก์ชันวัตถุประสงค์แล้วจะสามารถแบ่งการทดลองออกเป็น 4 การทดลอง ซึ่งแต่ละการทดลองมีรายละเอียดดังต่อไปนี้

3.4.2.1 การทดลองเมื่อใช้ ML เป็นฟังก์ชันวัตถุประสงค์

(a) รายละเอียดการทดลอง

เนื่องจาก ML คำนวณความเหมาะสมของทอพอโลยีที่กำหนด ตามสมการที่ 3.1 ซึ่งเป็นการคำนวณโดยอาศัยค่าความสามารถที่ทอพอโลยีสามารถปรับตัวให้เข้ากับชุดข้อมูลฝึกฝนและไม่มีการตั้งสมมุติฐานใดเพิ่มเติม ดังนั้น ML จึงสามารถที่จะนำไปใช้กับการเลือกทอพอโลยีเพื่อทดสอบได้ทั้ง 3 แบบ คือ LTR SSS และ SRA

สำหรับ LTR นั้น การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟจะพิจารณาเพียงการประมาณจำนวนสถานะและจำนวนส่วนผสมเกาส์เซียนต่อสถานะ โดยการประมาณนั้นจะทำโดยเปลี่ยนแปลงจำนวนสถานะตั้งแต่ 4 (สถานะไม่สร้างผลการสังเกต 2 สถานะ และ สถานะปลดปล่อย 2 สถานะ) ถึง N เมื่อ N คือ สถานะมากสุด ซึ่งคำนวณจากค่าเฉลี่ยจำนวนผลการสังเกตของลำดับการสังเกตของหน่วยเสียงที่สนใจตามตารางที่ 18

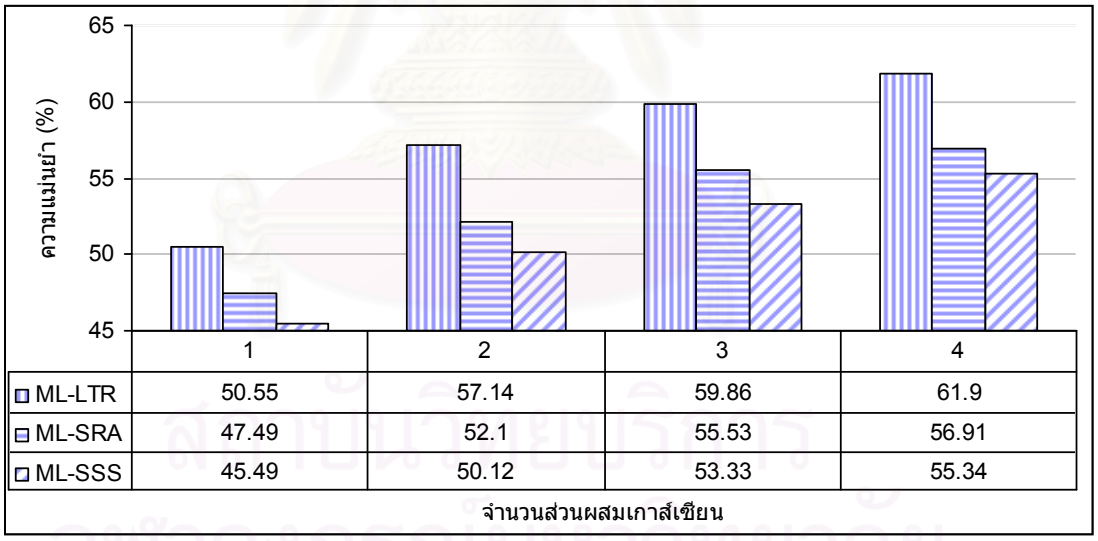
ตารางที่ 18 จำนวนสถานะมากสุดของแต่ละหน่วยเสียง

หน่วยเสียง	N	หน่วยเสียง	N	หน่วยเสียง	N	หน่วยเสียง	N
p	5	w	7	a	9	iia	19
t	5	j	8	aa	19	va	0
c	5	pr	7	i	8	vva	18
k	5	phr	10	ii	20	ua	0
z	6	tr	8	v	10	uua	18
ph	8	kr	8	vv	22	p^	5
th	8	khr	12	u	9	t^	5

ch	11	pl	6	uu	17	k^	5
kh	9	phl	10	e	9	n^	13
b	8	thr	12	ee	18	m^	13
d	6	kl	7	x	14	ng^	12
m	8	khl	11	xx	19	j^	12
n	8	kw	7	o	10	w^	11
ng	8	khw	9	oo	16	f^	10
l	8	br	12	@	13	l^	15
r	8	bl	11	@@	17	s^	13
f	12	fr	14	q	15	ch^	11
s	12	fl	14	qq	25	sil	31
h	10	dr	11	ia	20	sp	20

จำนวนสถานะเริ่มต้นของการผลิตทอพอโลยีแบบ SRA และเกณฑ์ในการหยุดการทำงานของ SSS ก็ใช้ตามตารางนี้เช่นเดียวกัน

(b) ผลการทดลอง



รูปที่ 21 ประสิทธิภาพของทอพอโลยีที่ประมาณได้จาก ML

รูปที่ 21 แสดงประสิทธิภาพของเครื่องรู้จำที่ได้จากการประมาณทอพอโลยีโดยใช้ ML เป็นฟังก์ชันวัตถุประสงค์ ควบคู่กับการใช้ LTR SSS และ SRA เป็นวิธีการผลิตทอพอโลยี

(c) วิเคราะห์ผลการทดลอง

จะเห็นว่าการผลิตทอพอโลยีโดยใช้วิธีกำหนดรูปแบบการเชื่อมต่อเป็น LTR ให้ผลดีที่สุดในทุกส่วนผสมเกาส์เซียน เนื่องจาก SRA และ SSS เป็นวิธีการผลิตโดยการพยายามปรับโครงสร้างของการเชื่อมต่อของสถานะให้มีความใกล้เคียงกับชุดข้อมูลฝึกฝน ดังนั้น เมื่อ ML วัดความเหมาะสมของทอพอโลยีจากความสามารถที่ทอพอโลยีปรับตัวให้เหมาะกับชุดข้อมูลฝึกฝนได้มากน้อยเพียง จึงทำให้วิธี การผลิตทอพอโลยีทั้ง SRA และ SSS ซึ่งยอมให้ ML ปรับโครงสร้างของการเชื่อมต่อของสถานะเกิดปัญหาการปรับเหมาะเกินไป แต่สำหรับ LTR นั้น การประมาณทอพอโลยีจะประมาณเพียงจำนวนสถานะ การกำหนดรูปแบบการเชื่อมต่อเอาไว้ล่วงหน้า ในที่นี้คือ LTR ทำให้ ML ไม่สามารถปรับโครงสร้างของแบบจำลองได้ จึงช่วยลดปัญหาการปรับเหมาะเกินไปจึงทำให้ LTR ได้ผลดีกว่า SSS SRA เมื่อใช้ ML เป็นฟังก์วัตถุประสงค์

3.4.2.2 การทดลองโดยใช้ BIC เป็นฟังก์ชันวัตถุประสงค์

(a) รายละเอียดการทดลอง

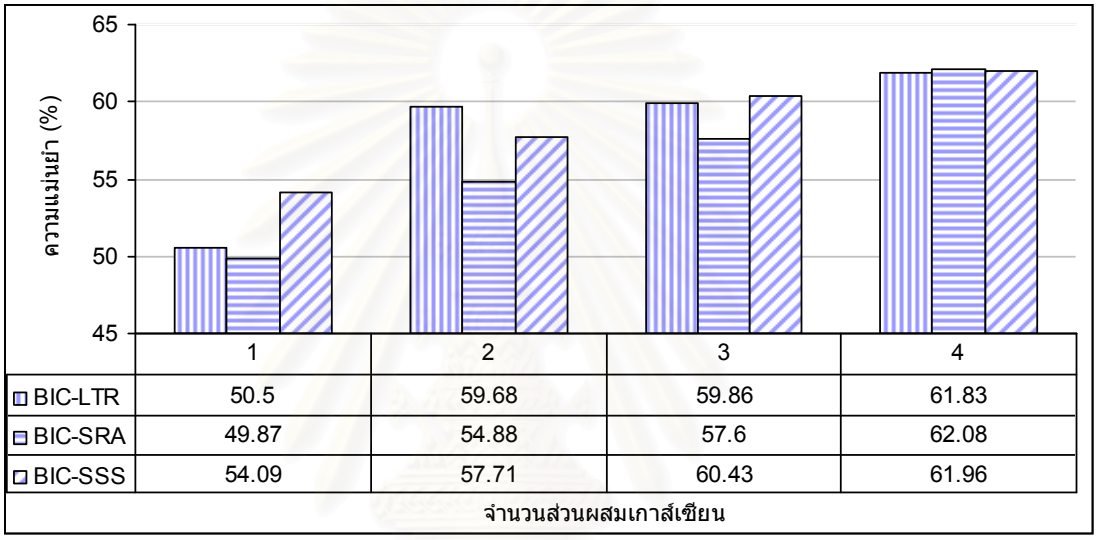
BIC เป็นอีกวิธีหนึ่งที่อาศัยหลักการมีดโกนออคแคมมาประยุกต์ใช้ในการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟโดย BIC ต้องการจำนวนพารามิเตอร์อิสระและจำนวนตัวอย่างของหน่วยเสียง เพื่อใช้ในการคำนวณ ความเหมาะสมของทอพอโลยี ซึ่งวิธีการนับพารามิเตอร์ได้กล่าวไปใน 3.4.1.2 เนื่องจาก การประมาณทอพอโลยีใช้ข้อมูลเสียงชุดสมดุลจากฐานข้อมูลโลตัสดังนั้น จำนวนตัวอย่างของแต่ละหน่วยเสียงจึงเป็นไปดังตารางที่ 19

ตารางที่ 19 จำนวนตัวอย่างหน่วยเสียงของแต่ละหน่วยเสียง

หน่วยเสียง	#ตัวอย่าง	หน่วยเสียง	#ตัวอย่าง	หน่วยเสียง	#ตัวอย่าง	หน่วยเสียง	#ตัวอย่าง
aa	5288	ii	1546	n	1801	thr	63
a	6503	i	1124	n^	3877	t	1253
bl	37	j	683	o	1168	tr	140
b	729	j^	2592	oo	523	t^	1865
br	8	khl	105	phl	69	u	588
ch	763	kh	1466	ph	1209	ua	0
ch^	8	khr	380	phr	177	uu	737
c	1121	khw	237	pl	117	uua	770
d	1042	kl	141	p	708	v	559
dr	16	k	1662	pr	341	va	0
ee	430	kr	254	p^	1185	vva	592
e	763	kw	61	q	29	vv	427

fl	10	k^	1515	qq	407	w	681
f	211	l	1405	r	1299	w^	692
fr	6	l^	10	sil	3195	x	513
f^	51	m	1364	s	2230	xx	800
h	599	m^	1640	sp	12997	@	483
ia	2	ng	216	s^	50	@@	1688
iia	364	ng^	2948	th	1943	z	807

(b) ผลการทดลอง



รูปที่ 22 ประสิทธิของทอพอโลยีที่ประมาณได้จาก BIC

รูปที่ 22 แสดงประสิทธิภาพของเครื่องรู้จำที่ได้จากการประมาณทอพอโลยีโดยใช้ BIC เป็นฟังก์ชันวัตถุประสงค์ควบคู่กับการใช้ LTR SSS และ SRA

(c) วิเคราะห์ผลการทดลอง

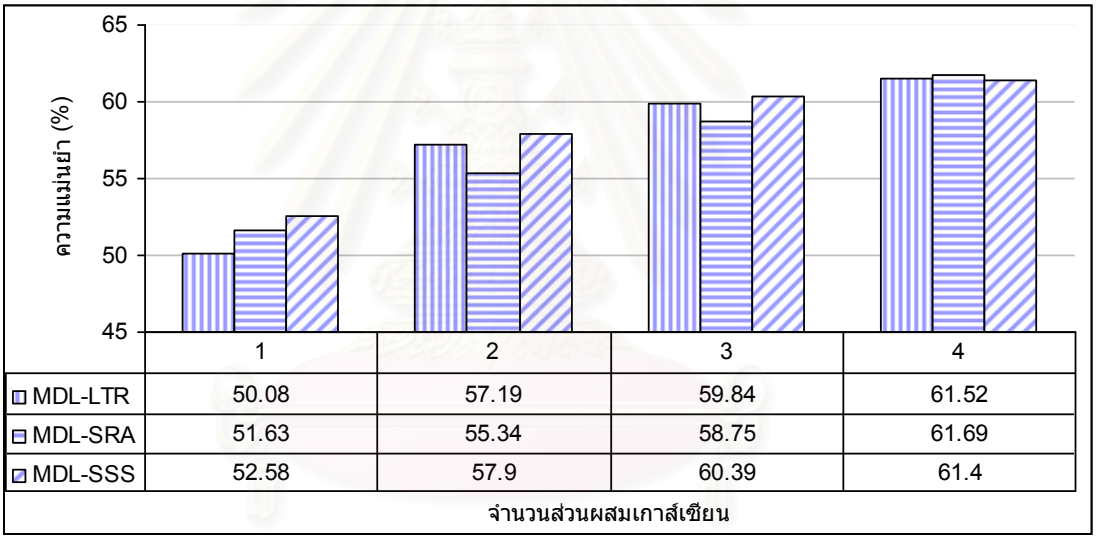
เมื่อเปรียบเทียบที่จำนวนส่วนผสมเกาส์เซียนเท่ากับ 1 จะเห็นว่า วิธีการ ผลิตทอพอโลยีแบบ SSS เมื่อนำไปสร้างเครื่องรู้จำแล้วให้ค่าความแม่นยำถึง 54.09% และเมื่อเปรียบเทียบที่จำนวนส่วนผสมเกาส์เซียนเท่ากับ 4 SRA ให้ค่าความแม่นยำสูงสุด คือ 62.08% หากเปรียบเทียบกับการใช้ ML ฟังก์ชันวัตถุประสงค์แล้ว พบว่า เครื่ องรู้จำที่สร้างจากวิธีผลิต แบบ SRA และ SSS ที่ใช้ BIC เป็นฟังก์ชันวัตถุจะมีความแม่นยำสูงกว่า ML ทั้งนี้เนื่องจาก BIC มีพจน์เพิ่มติมที่เป็นค่าหักลบเพื่อถ่วงดุลกับความสามารถในการปรับตัว ให้เข้ากับชุดข้อมูลฝึกฝน ดังนั้นจึงทำให้ทอพอโลยีที่ประมาณจาก SSS และ SRA ไม่ยึดติดกับข้อมูลฝึกฝนมากจนเกินไป

3.4.2.3 การทดลองเมื่อใช้ MDL เป็นฟังก์ชันวัตถุประสงค์

(a) รายละเอียดการทดลอง

MDL เป็นฟังก์ชันวัตถุประสงค์ที่คล้ายกับ BIC มาก แต่ฟังก์ชันวัตถุประสงค์ทั้งสองมีการคำนวณที่แตกต่างกันตรงการคำนวณพจน์หักลบ โดย BIC นั้นใช้จำนวนตัวอย่างเสียงมาเป็นส่วนประกอบในการคำนวณ แต่ MDL จะใช้ค่าเฉลี่ยของจำนวนผลการสังเกตของลำดับการสังเกตมาคิดคำนวณ ดังนั้น เราจึงสามารถนำวิธีการคัดเลือกทอพอโลยีทั้ง 3 แบบมาใช้กับ MDL ได้ โดยที่ความยาวของลำดับการสังเกตของหน่วยเสียงที่สนใจนั้น ก็คือ จำนวนสถานะมากสุดที่คำนวณโดยจากจำนวนผลการสังเกตเฉลี่ยของลำดับการสังเกตของหน่วยเสียงที่เราสนใจ ดังแสดงในตารางที่ 18 โดยไม่รวมสถานะไม่สร้างผลการสังเกต

(b) ผลการทดลอง



รูปที่ 23 ประสิทธิภาพของทอพอโลยีที่ประมาณได้จาก MDL

รูปที่ 23 แสดงประสิทธิภาพของทอพอโลยีที่ประมาณได้จากการนำ MDL มาเป็นฟังก์ชันวัตถุประสงค์และใช้วิธีการผลิตทอพอโลยีเพื่อนำมาทดสอบ 3 แบบ คือ LTR SSS และ SRA

(c) วิเคราะห์ผลการทดลอง

เมื่อเปรียบเทียบผลการทดลองที่ได้กับการทดลองที่ 3.4.2.2 ซึ่งใช้ BIC เป็นฟังก์ชันวัตถุประสงค์ พบว่า ผลการทดลองของฟังก์ชันวัตถุประสงค์ทั้งสองมีค่าใกล้เคียงกัน โดยที่เมื่อพิจารณาที่ส่วนผสมเกาส์เซียนเท่ากับ 4 แล้ว วิธีการผลิตทอพอโลยีแบบ SRA ให้ผลสูงสุด

เหมือนกัน และ MDL ก็ให้ผลที่ดีสำหรับการผลิตทอพอโลยีแบบ SSS เช่นเดียวกับ BIC ทั้งนี้เนื่องจากฟังก์ชันวัตถุประสงค์ทั้งสองนำหลักการมีดโกนออคแคมมาประยุกต์ใช้เหมือนกัน ดังนั้นผลการทดลองจึงไม่แตกต่างกันมากนัก

3.4.2.4 การทดลองเมื่อใช้ HBIC เป็นฟังก์ชันวัตถุประสงค์และการเปรียบเทียบจำนวนพารามิเตอร์ของทอพอโลยีที่ประมาณได้

(a) รายละเอียดการทดลอง

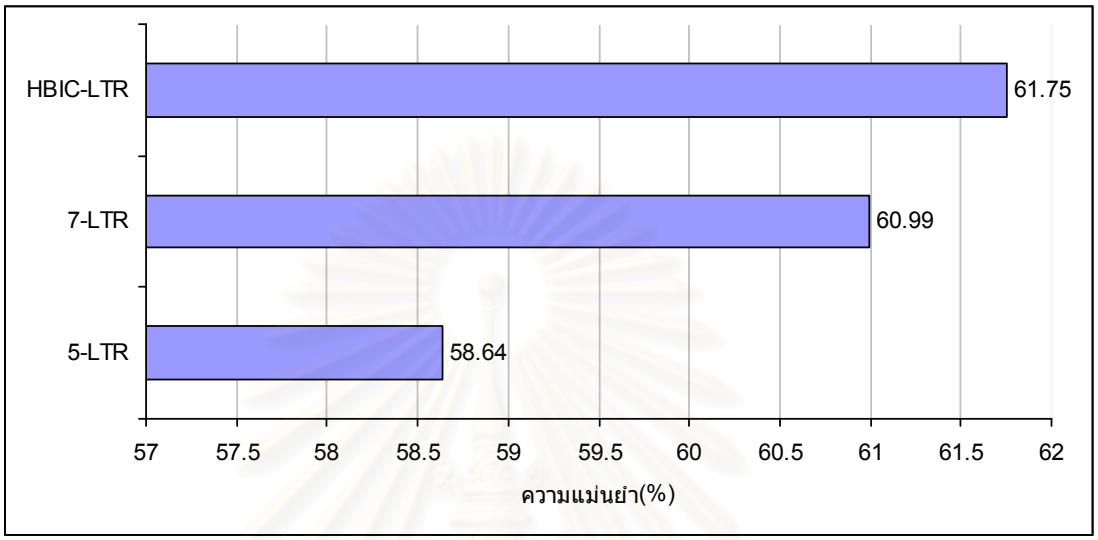
การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟโดยใช้ HBIC และ LTR จำนวนส่วนผสมเกาส์เซียนที่เหมาะสมสำหรับแต่ละหน่วยเสียงจะถูกรวมเข้าไปในการประมาณด้วยและการคำนวณความเหมาะสมของทอพอโลยีสามารถคำนวณได้ดังสมการ (3.4)

การคำนวณพจน์ $\log P(\tau)$ หากกำหนดรูปแบบการเชื่อมต่อของสถานะเป็นแบบ LTR และตั้งสมมุติฐานไว้ล่วงหน้าว่า จำนวนสถานะและจำนวนส่วนผสมเกาส์เซียนไม่ขึ้นต่อกันจะทำให้พจน์นี้สามารถเขียนได้เป็น $\log P(\tau) = \log P(Q) + \log P(L)$ โดยที่ค่า $P(Q)$ และ $P(L)$ คือ ค่าความน่าจะเป็นที่ได้จากฟังก์ชันความหนาแน่นของความน่าจะเป็นซึ่งในการทดลองนี้สมมุติให้ Q และ L มีการแจกแจงแบบเกาส์เซียนโดยการสร้างฟังก์ชันความหนาแน่นของความน่าจะเป็นสำหรับ Q และ L นั้น ทำโดยสุ่มตัวอย่างจำนวน 10% ของหน่วยเสียงที่สนใจจากข้อมูลชุดหน่วยเสียงสมดุลเป็นจำนวน 50 ครั้ง และในแต่ละครั้งที่สุ่มตัวอย่างขึ้นมา นำตัวอย่างที่ได้ไปเป็นข้อมูลเสียงเพื่อประมาณหาจำนวนสถานะและจำนวนส่วนผสมเกาส์เซียนที่เหมาะสมโดยใช้ ML เป็นฟังก์ชันวัตถุประสงค์ จากนั้นนำผลลัพธ์ที่ได้จากการประมาณในแต่ละครั้งไปเป็นข้อมูลตัวอย่างเพื่อสร้างฟังก์ชันความหนาแน่นของความน่าจะเป็นของสถานะและส่วนผสมเกาส์เซียน

ในทำนองเดียวกัน เมื่อตั้งสมมุติฐานว่า A ω μ และ Σ ไม่ขึ้นต่อกัน จะทำให้ $\log P(\hat{\theta}_i \mid \tau) = \log P(\hat{A}_i \mid \tau) + \log P(\hat{\omega}_i \mid \tau) + \log P(\hat{\mu}_i \mid \tau) + \log P(\hat{\Sigma}_i \mid \tau)$ และเมื่อกำหนด τ เอาไว้ล่วงหน้า และสมมุติให้พารามิเตอร์ต่างๆ มีการแจกแจงแบบเกาส์เซียน การคำนวณพจน์ทั้ง 4 นี้สามารถทำได้โดยสร้างฟังก์ชันความหนาแน่นของความน่าจะเป็นสำหรับ A ω μ และ Σ ด้วยวิธีการเดียวกับสถานะและส่วนผสมเกาส์เซียน คือ สุ่มตัวอย่างเสียงของหน่วยเสียงจำนวน 10% ของตัวอย่างหน่วยเสียงนั้นๆ ที่มีในข้อมูลเสียงชุดสมดุล และใช้ตัวอย่างเสียงที่สุ่มได้เป็นฐานข้อมูลเสียงเพื่อฝึกฝนแบบจำลองเสียงพูด จากนั้นใช้แบบจำเสียงพูดที่ฝึกฝนแล้วทั้ง 50 แบบจำลองสร้างฟังก์ชันความหนาแน่นของความน่าจะเป็นสำหรับ A ω μ และ Σ

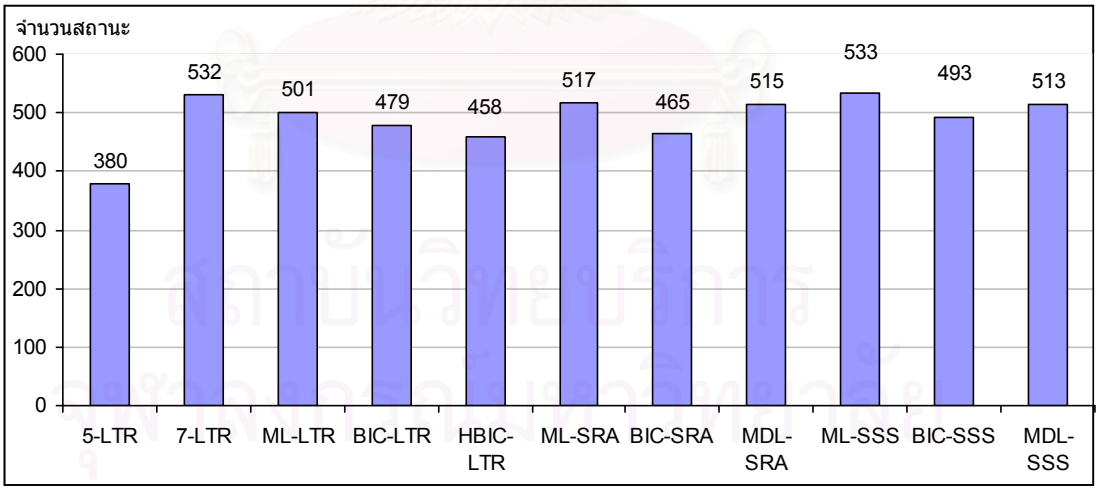
(b) ผลการทดลอง

การทดลองนี้เปรียบเทียบความแม่นยำของเครื่องรู้จำที่สร้างจากทอพอโลยีที่ได้จากการประมาณด้วย HBIC กับการเลือกใช้ทอพอโลยีแบบ 5 และ 7 สถานะที่มีรูปแบบการเชื่อมต่อของสถานะเป็น LTR และใช้จำนวนส่วนผสมเกาส์เซียนเป็น 4 ได้ผลการทดลอง ดังแสดงในรูปที่ 24



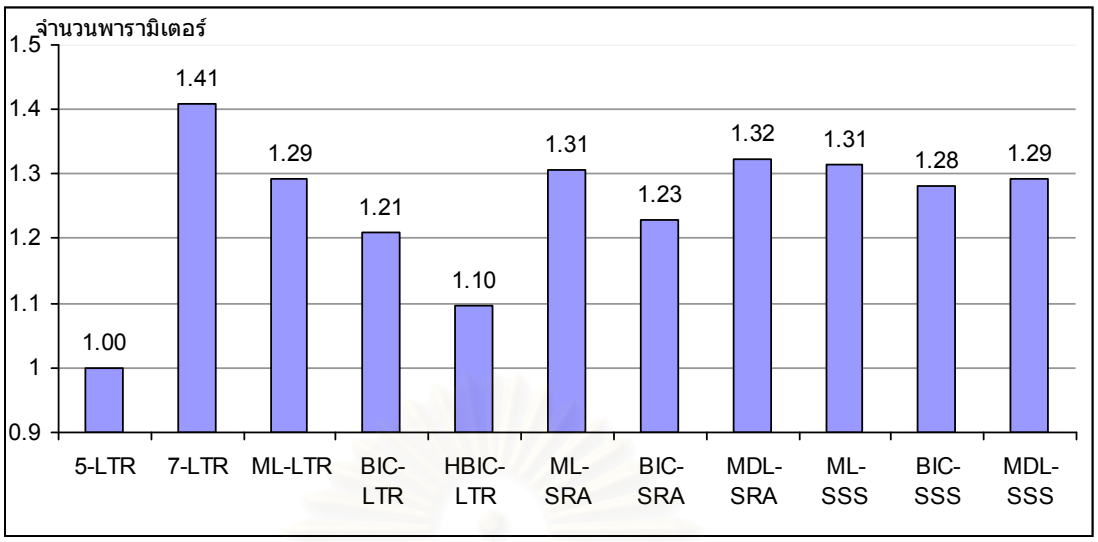
รูปที่ 24 ความแม่นยำของทอพอโลยีที่ประมาณได้จาก HBIC

รูปที่ 25 แสดงจำนวนสถานะทั้งหมดที่ประมาณได้จากการทดลองที่ 3.4.2.1-3.4.2.4



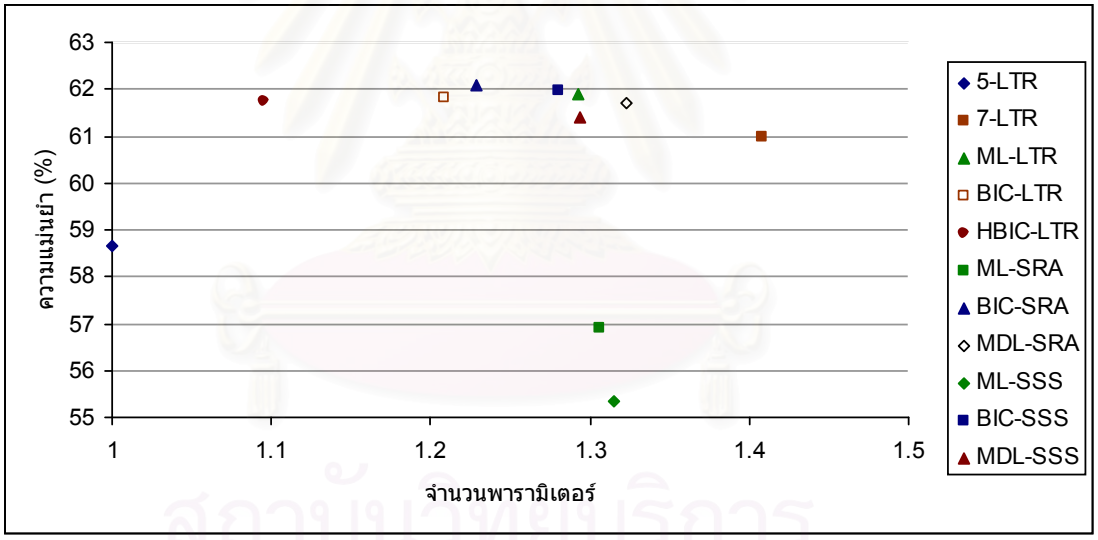
รูปที่ 25 จำนวนสถานะของทอพอโลยีที่ประมาณได้

รูปที่ 26 แสดงจำนวนพารามิเตอร์ของทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่ประมาณได้จากการทดลองที่ 3.4.2.1-3.4.2.4 โดยคิดเป็นจำนวนเท่าของทอพอโลยีแบบ LTR จำนวน 5 สถานะ 4 ส่วนผสมเกาส์เซียน



รูปที่ 26 จำนวนพารามิเตอร์ของทอพอโลยีที่ประมาณได้

รูปที่ 27 แสดงการลงจุด (Plot) เมื่อแกน y คือ ความแม่นยำของเครื่องรู้จำที่สร้างจากการประมาณทอพอโลยีที่เลือกใช้ และแกน x คือ จำนวนพารามิเตอร์



รูปที่ 27 การลงจุดระหว่างจำนวนพารามิเตอร์และความแม่นยำ

(c) วิเคราะห์ผลการทดลอง

เมื่อเปรียบเทียบความแม่นยำของ HBIC-LTR กับ 5-LTR และ 7-LTR พบว่า HBIC-LTR มีความแม่นยำมากกว่า 2.09% และ 0.76% ตามลำดับ ซึ่งแสดงให้เห็นว่า การประมาณแบบ HBIC ให้ทอพอโลยีที่เหมาะสมกว่าการเลือกใช้ทอพอโลยีแบบ 5 และ 7 สถานะ นอกจากนั้นทอพอโลยีที่ประมาณได้จาก HBIC-LTR ยังมีจำนวนสถานะและจำนวนพารามิเตอร์น้อยที่สุดเป็นอันดับ 2 แต่กลับให้ค่าความแม่นยำเป็นอันดับที่ 4 ทำให้เมื่อทดลองลงจุดตามรูปที่

11 แล้ว HBIC-LTR จึงอยู่ทางมุมซ้ายบนของกราฟ นั่นคือ หากพิจารณาความเหมาะสมโดยคำนึงถึงจำนวนพารามิเตอร์ ด้วยแล้วทอพอโลยีที่ประมาณได้จากวิธี HBIC-LTR จึงมีความเหมาะสมมากที่สุด แต่อย่างไรก็ดี การประมาณทอพอโลยีแบบ HBIC จำเป็นต้องสร้างฟังก์ชันความหนาแน่นของความน่าจะเป็นของพารามิเตอร์ ซึ่งใช้เวลาค่อนข้างมาก ทำให้การประมาณด้วยวิธีนี้ใช้เวลามากกว่าวิธีอื่น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4
การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟโดยใช้ขั้นตอนวิธีเชิงพันธุกรรม

ในบทที่ 3 เครื่องรู้จำหน่วยเสียงภาษาไทยแบบบริบทอิสระที่สร้างขึ้นเพื่อวัดประสิทธิภาพมีความแม่นยำประมาณ 55-62% เมื่อวัดผลด้วยชุดข้อมูล DT แต่อย่างไรก็ดีเราสามารถเพิ่มประสิทธิภาพของเครื่องรู้จำดังกล่าวให้มากกว่านี้ได้ เนื่องจากการวัดความเหมาะสมของฟังก์ชันวัตถุประสงค์ที่นำเสนอไปในบทที่ 3 เป็นการวัดความเหมาะสมของทอพอโลยีสำหรับหน่วยเสียงที่เราสนใจเท่านั้น แต่ในการวัดผลนั้นทอพอโลยีที่ประมาณได้ทั้งหมดจะถูกนำมารวมกันแล้วสร้างเป็นเครื่องรู้จำเสียงแบบต่อเนื่อง ดังนั้นในวิทยานิพนธ์นี้จึงมีแนวคิดว่า หากเปลี่ยนปัญหาการประมาณทอพอโลยีที่เหมาะสมสำหรับแต่ละหน่วยเสียง เป็นปัญหาการเลือกทอพอโลยีของแต่ละหน่วยเสียงที่มีอยู่แล้วมาประกอบกันเพื่อเป็นเครื่องรู้จำที่มีประสิทธิภาพ กล่าวคือ หากกำหนดให้ $P = \{P_1, P_2, \cdots, P_I\}$ คือเซตของหน่วยเสียงที่เราสนใจ ซึ่งในวิทยานิพนธ์ยึดถือหน่วยเสียงตามฐานข้อมูลเสียงโลตัส จึงมีหน่วยเสียงทั้งหมด 74 หน่วยเสียง รวมกับ sil และ sp ดังนั้น I จึงเท่ากับ 76 และหากกำหนดให้ $T_i = \{t_1^i, t_2^i, \cdots, t_K^i\}$ คือ เซตของทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับหน่วยเสียง P_i สามารถเลือกใช้ได้ ปัญหาการเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟจึง คือ การเลือกสมาชิก t_k^i สำหรับ P_i เมื่อ $1 \le i \le 76$ ตัวใดตัวใดหนึ่งเพื่อให้เมื่อนำมาสร้างเครื่องรู้จำแล้วมีประสิทธิภาพ

หากพิจารณา T_i ซึ่งคือเซตของทอพอโลยีสำหรับหน่วยเสียง P_i นั้นเป็นเซตที่มีขนาดใหญ่มาก เนื่องจาก ทอพอโลยีที่เป็นไปได้มีจำนวนไม่จำกัด ทำให้เราจำเป็นต้องมีการคัดเลือกทอพอโลยีที่คิดว่าน่าสนใจ และมีแนวโน้มที่เมื่อถูกเลือกใช้เพื่อสร้างเครื่องรู้จำแล้วจะให้ความแม่นยำที่สูงขึ้น เพื่อชี้แจงให้เกิดความเข้าใจตรงกันระหว่างผู้อ่านและผู้เขียนวิทยานิพนธ์ จึงขอยกตัวอย่างดังนี้ เนื่องจากทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟประกอบด้วยส่วนประกอบ 3 ส่วนหลัก คือ รูปแบบการเชื่อมต่อของสถานะ จำนวนสถานะและจำนวนส่วนผสมเกาส์เซียน ดังนั้น การเลือกเซตของทอพอโลยีที่คิดว่าน่าสนใจสำหรับแต่ละหน่วยเสียงก็คือการเลือกองค์ประกอบของทอพอโลยีตามที่กล่าว โดยสามารถเลือกได้ ดังตัวอย่างต่อไปนี้

1. สร้าง T_i สำหรับทุกหน่วยเสียงโดยกำหนดให้สมาชิกของ T_i มีเงื่อนไขดังต่อไปนี้
    a. มีจำนวนสถานะตั้งแต่ 4 – 25
    b. มีการเชื่อมต่อของสถานะแบบ LTR
    c. มีจำนวนส่วนผสมเกาส์เซียนตั้งแต่ 1 – 4
2. สร้าง T_i สำหรับทุกหน่วยเสียงโดยกำหนดให้สมาชิกของ T_i มีเงื่อนไขดังต่อไปนี้

a. มีการเชื่อมต่อของสถานะแบบ LTR, LTRF, FC
b. มีจำนวนสถานะเท่ากับจำนวนสถานะที่ประมาณได้จาก ML-LTR
c. มีจำนวนส่วนผสมเกาส์เซียนเป็น 1 เท่านั้น

วิธีการเลือกเซตของทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่มีอยู่แล้วสำหรับแต่ละหน่วยเสียงนั้น เป็นขั้นตอนที่สำคัญเนื่องจาก หากเลือกเซตของทอพอโลยีที่ไม่สามารถนำไปสู่การสร้างเครื่องรู้จำที่มีความแม่นยำแล้ว เครื่องรู้จำที่สร้างได้ก็จะไม่มีประสิทธิภาพอย่างแน่นอน

ปัญหาอีกปัญหาหนึ่งที่ต้องคำนึงถึง เมื่อ $T = \{T_1, T_2, \cdots T_{76}\}$ คือ เซตของทอพอโลยีที่เลือกมาสำหรับ $P = \{P_1, P_2, \cdots, P_{76}\}$ ทั้งหมดตามโลตัส หากพิจารณาถึงจำนวนของ T ที่เป็นไปได้ จะพบว่า เราไม่สามารถทดสอบเซต T ที่เป็นไปได้ทั้งหมด เช่น หากให้แต่ละหน่วยเสียงสามารถเลือกแบบจำลองได้ 3 แบบ รูปแบบทั้งหมดที่เป็นไปได้จะมี 3^{76} รูปแบบ และ นอกจากนั้นเรายังไม่ทราบถึงปัจจัยต่างๆที่ส่งผลต่อความแม่นยำโดยรวมเมื่อเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับแต่ละหน่วยเสียง แม้แต่ฟังก์ชันวัตถุประสงค์ก็ไม่ได้บอกถึงความเหมาะสมอย่างถูกต้อง เพียงแต่บอกแนวโน้มหรือความน่าจะเป็นเท่านั้น ดังนั้น ในวิทยานิ พนธ์นี้จึงเลือกใช้ขั้นตอนวิธีเชิงพันธุกรรม ซึ่งเป็นการค้นหาคำตอบในกลุ่มการค้นหาครอบคลุมแบบฮิวริสติก (Global Search Heuristic) ซึ่งมีข้อดีคือ เป็นวิธีการค้นหาคำตอบที่เหมาะสมโดยประมาณ ขั้นตอนวิธีนี้ไม่จำเป็นต้องทราบถึงปัจจัยที่ส่งผลต่อวัตถุประสงค์ที่ต้องการ และหากเลือกใช้ขั้นตอนวิธีเชิงพันธุกรรมจำเป็นที่จะต้องกำหนดแผนการแทนและค่าความเหมาะมาตรฐานซึ่งจะทำดังนี้

a) แผนการแทน

แผนการแทน คือ วิธีแสดงประชากรในรูปของอาเรย์ของบิต (Array of bits) ในที่นี้ปัญหาของเราคือ การเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ สำหรับหน่วยเสียงที่สนใจ ซึ่งประกอบด้วย จำนวนสถานะ รูปแบบการเชื่อมต่อของสถานะและจำนวนส่วนผสมเกาส์เซียนต่อสถานะ ดังนั้น แผนการแทนสำหรับแต่ละหน่วยเสียงจึงเป็นดังรูปที่ 28

จำนวนสถานะ (1 – N)	LTR,FLTR หรือ FC	Mixture (1 – 4)

รูปที่ 28 แผนการแทนสำหรับทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟ

b) ค่าความเหมาะมาตรฐาน

ในวิทยานิพนธ์นี้วัดประสิทธิภาพทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่ประมาณได้จากความแม่นยำของเครื่องรู้จำหน่วยเสียงที่สร้างขึ้นจากการประมาณเพื่อ ให้ฟังก์ชันค่าความเหมาะมาตรฐานตรงกับวัตถุประสงค์ จึงนำความแม่นยำของเครื่องรู้จำหน่วยเสียงที่สร้างขึ้นมาเป็นฟังก์ชันค่าความเหมาะมาตรฐานโดยใช้ข้อมูลเสียงชุด DT เป็นตัวอย่างเสียงเพื่อคำนวณ

เมื่อกำหนดแผนการแทนและฟังก์ชันค่าความเหมาะมาตรฐานแล้ว ขั้นตอนวิธีเชิงพันธุกรรมสามารถเขียนเป็นขั้นตอนดังนี้

ขั้นตอนวิธีเชิงพันธุกรรมสำหรับการประมาณทอพอโลยี

1. เลือกประชากรเริ่มต้นโดยสุ่มเลือกจำนวนสถานะ รูปแบบการเชื่อมต่อของสถานะและจำนวนส่วนผสมเกาส์เซียนตามขอบเขตที่กำหนด
2. คำนวณค่าความเหมาะมาตรฐานของประชากรเริ่มต้น
3. วนซ้ำ
   ก. คัดเลือกประชากรตามวิธีการคัดเลือกตามธรรมชาติที่กำหนด
   ข. สร้างประชากรรุ่นใหม่ด้วยการกลายพันธุ์และการไขว้เปลี่ยน
   ค. คำนวณค่าความเหมาะมาตรฐานของประชากรรุ่นใหม่
   ง. คัดเลือกประชากรตามวิธีการคัดเลือกตามธรรมชาติที่กำหนด
   จ. แทนที่ประชากรรุ่นเก่าด้วยประชากรรุ่นใหม่
4. จบการทำงาน

รูปที่ 29 ขั้นตอนวิธีเชิงพันธุกรรมสำหรับการเลือกทอพอโลยี

การทดลองการเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับหน่วยเสียงภาษาไทยโดยใช้จีเอจะถูกแบ่งออกเป็น 2 การทดลองดังนี้

- การทดลองเลือกทอพอโลยีโดยเลือกจากการเชื่อมต่อแบบ LTR FLTR และ FC โดยมีจำนวนสถานะตั้งแต่ 4–N เมื่อ N คือ จำนวนสถานะที่คำนวณจากจำนวนผลการสังเกตของลำดับการสังเกตของหน่วยเสียงที่สนใจ และมีจำนวนส่วนผสมเกาส์เซียนเป็น 1–4
- การทดลองเลือกทอพอโลยีโดยเลือกจากทอพอโลยีที่ประมาณได้ในบทที่ 3

ในการทดลองนั้นนำโปรแกรม JGAP (J-Gap) [41] ซึ่งเป็นโปรแกรมและโครงร่างการทำงาน (Framework) สำหรับจีเอที่ถูกพัฒนาด้วยภาษาจาวา โดยผู้ใช้จะต้องกำหนดแผนการ

แทนและฟังก์ชันค่าความเหมาะมาตรฐานไว้ล่วงหน้า ในส่วนของโปรแกรมจะเตรียมวิธีการคัดเลือกทางธรรมชาติ การกระทำทางยีน (Genetic operation) ต่างๆ และการวิวัฒนาการไว้ให้ หรือหากต้องการสร้างสิ่งเหล่านี้เองก็สามารถทำได้โดยง่าย โดยการสืบทอดคลาสแม่แบบที่กำหนดไว้ในโครงร่างการทำงาน

การวัดประสิทธิภาพนั้นจะใช้เหมือนกับในบทที่ 3 คือ สร้างเครื่องรู้จำจากทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่ประมาณได้แล้วทำการวัดความแม่นยำ แต่ในบทนี้ จะนำข้อมูลเสียงชุดทดสอบประเมินผล (ET) เป็นข้อมูลเสียงทดสอบ โดยองค์ประกอบขอ งเครื่องรู้จำจะเป็นไปตามตารางที่ 20

ตารางที่ 20 องค์ประกอบของเครื่องรู้จำที่ใช้วัดผลของการทดลองที่ 4

	รายละเอียด		
การสกัดค่าสำคัญ	วิธีการสกัดค่าสำคัญ	สัมประสิทธิ์บนสเกลเมลที่มีเวคเตอร์ลักษณะสำคัญขนาด 39 มิติที่คำนวณจากค่าพลังงงาน อัตราการเปลี่ยนแปลงและความเร่งหรืออย่างละ 13 มิติ	
	ขนาดเฟรม	10 มิลลิวินาที	
	ความกว้างวินโดว์	25 มิลลิวินาที	
หน่วยเสียง	74 หน่วยเสียงตามโลตัส รวมกับ sil และ sp		
แบบจำลองหน่วยเสียง	หน่วยเสียงแบบบริบทอิสระที่มีเมตริกความแปรปรวนร่วมเกี่ยวตามแนวทแยง		
แบบจำลองภาษา	ไบแกรมที่ฝึกฝนด้วยไฟล์กำกับหน่วยเสียงของชุดฝึกฝนในโลตัส		
ตัวอย่างเสียง	ตัวอย่างเสียงสำหรับแบบจำลองทางเสียงและการประมาณทอพอโลยี		PD
	ตัวอย่างเสียงสำหรับการคำนวณค่าความเหมาะ		DT
	ตัวอย่างเสียงสำหรับการประเมิน		ET

เนื่องจากการวัดประสิทธิภาพของทอพอโลยีที่ประมาณได้เหมือนกับการทดลองในบทที่ 3 เพียงแต่เปลี่ยนชุดข้อมูลเสียงทดสอบจากชุดทดสอบเพื่อพัฒนาเป็นชุดทดสอบเพื่อประเมินผลดังนั้น ในส่วนของผลการทดลองจึงจะแนบการวัดผลการทดลองในบทที่ 3 โดยใช้ชุดทดสอบเพื่อประเมินผลเป็นข้อมูลเสียงทดสอบเข้าไปด้วย

4.1 การทดลองคัดเลือกทอพอโลยีโดยใช้จีเอเพียงอย่างเดียว

4.1.1 รายละเอียดการทดลอง

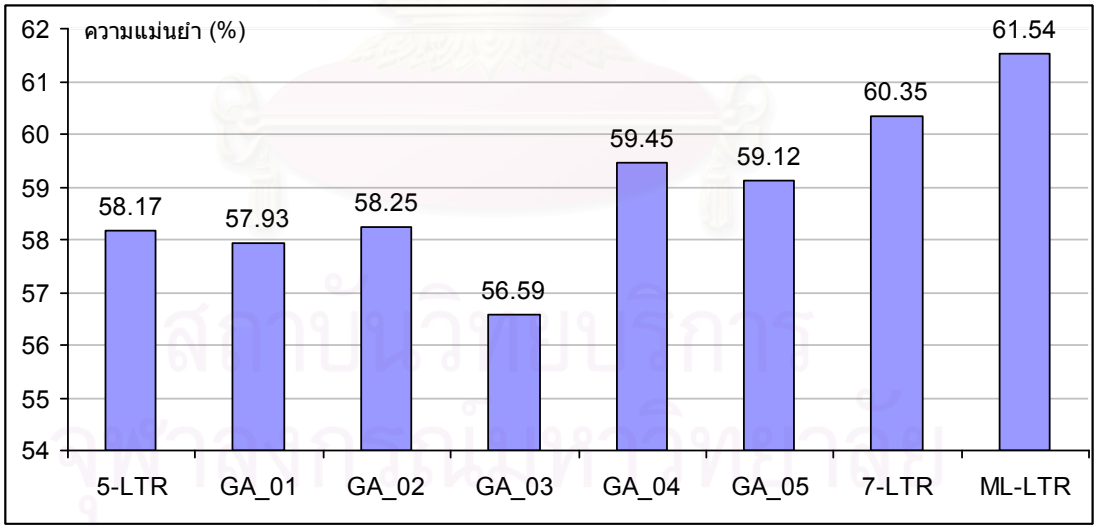
การทดลองนี้ กำหนดให้จำนวนประชากร คือ 10 ประชากรคงที่ ใช้จำนวนรอบในการวิวัฒนาการเท่ากับ 100 รอบ เลือกใช้การคัดเลือกตามธรรมชาติแบบรัสเซียนรูเล็ตต์ถ่วง

น้ำหนัก (Weighted Russian Roulette Selector) ซึ่งประชากรจะได้โควตา ช่อง (Slot Quota) ในรูเล็ตต์ตามสัดส่วนค่าความเหมาะสมของตน จากนั้นการคัดเลือกจะทำการโยนลูกเหล็กเพื่อเลือกทอพอโลยี โดยมีอัตราการกลายพันธุ์ (Mutant Rate) เป็น 50% และมีอัตราการไขว้เปลี่ยนเป็น 30% ของจำนวนประชากร แต่ละหน่วย เสียงจะสามารถเลือกการเชื่อมต่อของสถานะได้ 3 แบบ คือ LTR FLTR และ FC สำหรับส่วนผสมเกาส์เซียนนั้นสามารถเป็นได้ตั้งแต่ 1–4 และมีจำนวนสถานะเป็นได้ตั้งแต่ 4–N เมื่อ N คือ จำนวนสถานะที่คำนวณ จากจำนวนผลการสังเกต ของลำดับการสังเกตของหน่วยเสียงที่สนใจ ดังนั้นในการทดลองนี้จึงใช้แผนการแทนของแต่ละหน่วยเสียงตามรูปที่ 28

ในวิทยานิพนธ์นี้ใช้หน่วยเสียงสำหรับภาษาไทยตามโลตัสซึ่งมีทั้ง 74 หน่วยเสียง รวมกับ sil และ sp ดังนั้น แผนการแทนสำหรับการทดลองนี้ จึงคือ สายอักขระความยาว 228 หน่วยโดยที่ละแต่ 3 หน่วยจะแทนหน่วยเสียง 1 หน่วยเสียง

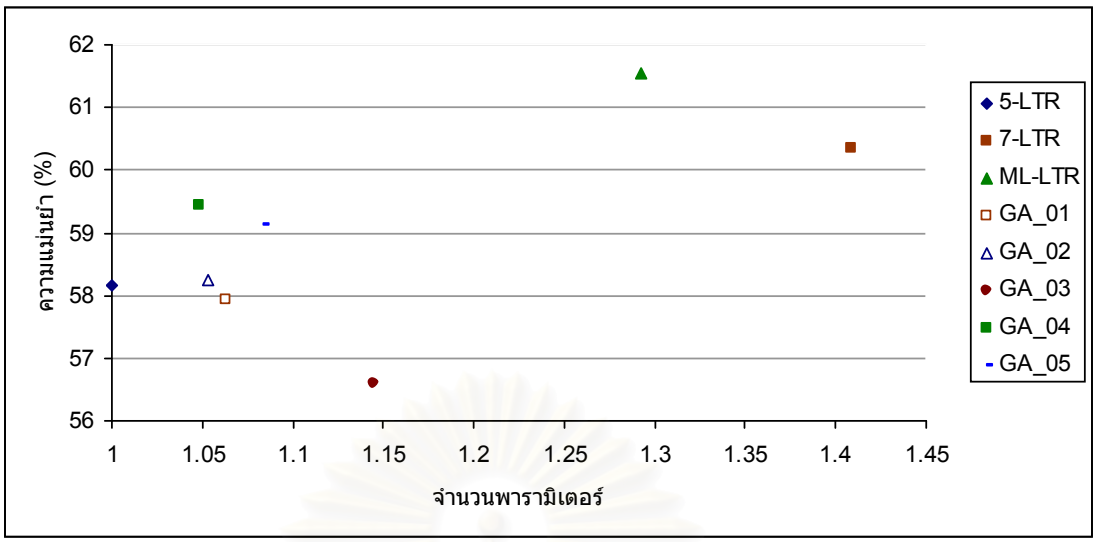
4.1.2 ผลการทดลอง

รูปที่ 30 แสดงแผนภูมิแท่งของความแม่นยำ เครื่องรู้จำที่สร้างขึ้นจาก ทอพอโลยีที่ประมาณได้ จากการรันจีเอตามเงื่อนไขที่กล่าวในรายละเอียดการทดลองจำนวน 5 ครั้ง เทียบกับทอพอโลยีที่ประมาณได้จาก 5-LTR 7-LTR และ ML-LTR ที่เพิ่มจำนวนส่วนผสมเกาส์เซียนเป็น 4



รูปที่ 30 ประสิทธิภาพของทอพอโลยีที่ได้จากการรันจีเอเทียบกับ 5-LTR 7-LTR และ ML-LTR

รูปที่ 31 แสดงจำนวนพารามิเตอร์ของทอพอโลยีที่ประมาณได้โดยเทียบเป็นสัดส่วนกับ 5-LTR ที่มีจำนวนส่วนผสมเกาส์เซียนเป็น 4



รูปที่ 31 การลงจุดระหว่างความแม่นยำและจำนวนพารามิเตอร์ของการทดลองที่ 4.1

4.1.3 วิเคราะห์ผลการทดลอง

จะเห็นว่า เมื่อใช้ข้อมูลเสียงชุดทดสอบเพื่อประเมินผลเป็นข้อมูลทดสอบ ML-LTR ก็ยังให้ความแม่นยำสูงถึง 61.54% ซึ่งเมื่อเปรียบเทียบกับการรันจีเอทั้ง 5 ครั้งแล้ว มีความแม่นยำมากกว่า แต่เมื่อนำจำนวนพารามิเตอร์มาเข้ามาพิจารณาด้วยแล้วจะเห็นว่าการรันจีเอทั้ง 5 ครั้งมีจำนวนพารามิเตอร์ที่น้อยกว่า คือ มีประมาณ 1.05 – 1.15 เท่าของ 5-LTR ที่มีจำนวนส่วนผสมเกาส์เซียนเป็น 4 แต่ ML-LTR นั้นมีจำนวนจำนวนพารามิเตอร์ประมาณ 1.3 เท่า ดังจะเห็นได้ชัดจากรูปที่ 31 ข้อมูลจะถูกแบ่งออกเป็น 2 กลุ่มอย่างชัดเจน คือ กลุ่มของการรันจีเอและกลุ่มของที่ประกอบด้วย ML-LTR และ 7-LTR

ความแม่นยำของการรันจีเอทั้ง 5 ครั้งนั้น เมื่อเปรียบเทียบกับจำนวนพารามิเตอร์ที่ใช้แล้ว ถือว่าอยู่ในเกณฑ์ที่พอใช้ได้ แต่อย่างไรก็ดีจุดประสงค์ของวิทยานิพนธ์นี้ มุ่งประเด็นที่ความแม่นยำของเครื่องรู้จำที่สร้างขึ้น ดังนั้น จึงถือว่าการประมาณด้วยวิธีนี้จึงต้องการปรับปรุงเพื่อให้มีความแม่นยำมากกว่านี้

4.2 การทดลองคัดเลือกทอพอโลยีโดยใช้จีเอและทอพอโลยีที่ประมาณได้ในบทที่ 3

ข้อบกพร่องที่คาดว่าเกิดขึ้นในการทดลองที่ 4.1 คือ แต่ละหน่วยเสียงมีจำนวนทอพอโลยีให้เลือกมากจนเกินไป ทำให้การรันจีเอทั้ง 5 ครั้งไม่สามารถค้นหาทอพอโลยีที่เหมาะสมให้แก่แต่ละหน่วยเสียงได้ ดังนั้น เพื่อแก้ปัญหานี้ในการทดลองที่ 2 จึงเลือกใช้ทอพอโลยีที่ประมาณได้จากบทที่ 3 และใช้จำนวนส่วนผสมเกาส์เซียนเท่ากับ 4 เท่านั้น โดยเหตุผลที่เลือกใช้ทอพอโลยีที่ประมาณได้จากบทที่ 3 เพราะว่า ทอพอโลยีที่ประมาณได้นั้นมีแนวโน้มที่จะเป็นทอ

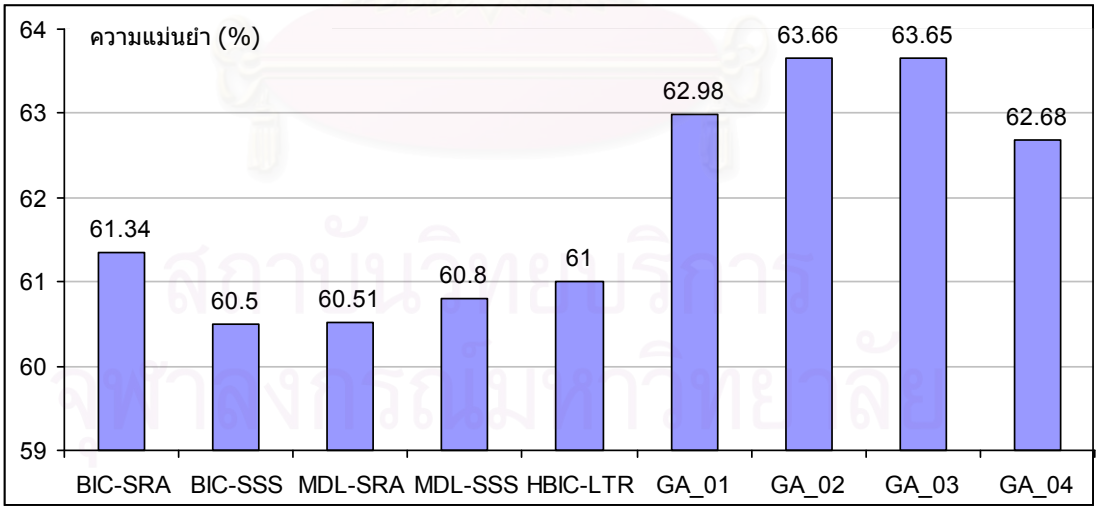
พอโลยีที่มีความเหมาะสมสำหรับแต่ละหน่วยเสียง เพียงแต่ว่า ความเหมาะสมนั้นเป็นความเหมาะสมเมื่อพิจารณาทีละหน่วยเสียง ไม่ใช่ความเหมาะสมโดยรวม

4.2.1 รายละเอียดการทดลอง

ทอพอโลยีจะถูกเลือกมาจากทอพอโลยีที่ประมาณได้จาก ML-SSS MDL-SSS BIC-SSS BIC-SRA ML-SRA MDL-SRA และ HBIC-LTR เนื่องด้วยการประมาณทั้งหมดที่เลือกใช้ได้ประมาณทั้งจำนวนสถานะและรูปแบบการเชื่อมต่อของสถานะแล้ว ดังนั้น การเลือกทอพอโลยีจึงเลือกเพียงว่าจะใช้ทอพอโลยีจากการประมาณแบบใด ในการทดลองนี้จึงใช้แผนการแทนสำหรับแต่ละหน่วยเสียงเป็นเลข 0 – 6 ที่แทนรูปแบบของทอพอโลยีที่ประมาณได้จากการทดลองที่ 3

ในการทดลองนี้กำหนดจำนวนประชากรเป็น 10 ประชากรคงที่ และใช้จำนวนรอบในการวิวัฒนาการ 100 รอบ โดยใช้การคัดเลือกทางธรรมชาติ 2 แบบ คือ การคัดเลือกแบบรัสเซียนรูเล็ตต์ถ่วงน้ำหนักและการคัดเลือกแบบโครโมโซมดีที่สุด (Best Chromoshome Selector) มีอัตราการกลายพันธุ์เป็น 50% และมีอัตราการไขว้เปลี่ยนเป็น 30% ของจำนวนประชากร โดยจะทำการรันจีเออย่างละ 2 ครั้ง

4.2.2 ผลการทดลอง

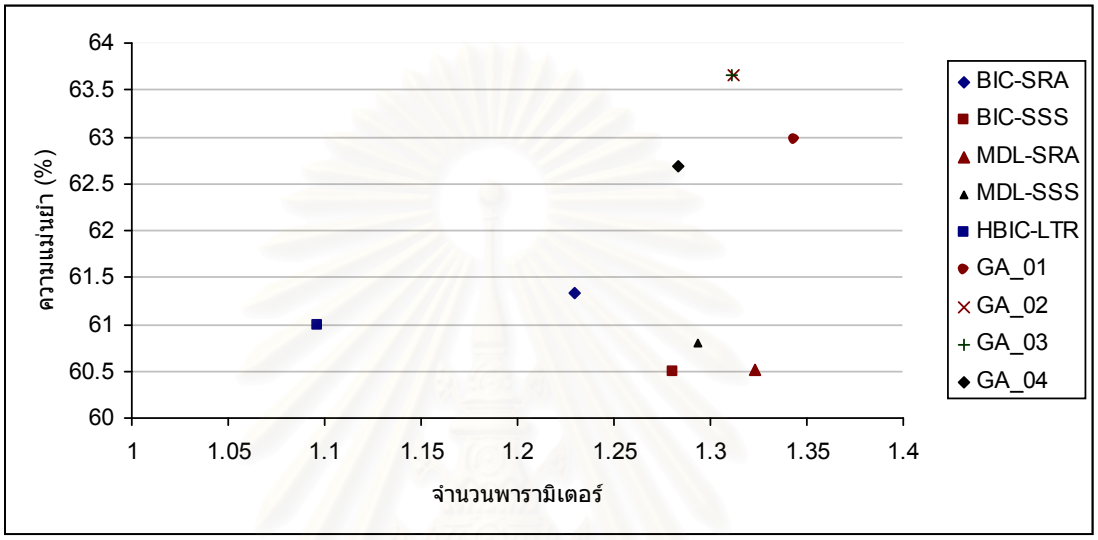


รูปที่ 32 ประสิทธิภาพของทอพอโลยีที่ได้จากการรันจีเอเทียบกับวิธีการประมาณในบทที่ 3

รูปที่ 32 แสดงแผนภูมิแท่งของความแม่นยำของทอพอโลยีที่ประมาณได้ จากการรันจีเอตามเงื่อนไขที่กล่าวในรายละเอียดโดยที่ GA_01 กับ GA_02 ใช้การคัดเลือกตามธรรมชาติแบบรัสเซียนรูเล็ตต์ถ่วงน้ำหนัก และ GA_03 กับ GA_04 ใช้การคัดเลือกตามธรรมชาติแบบ

โครโมโซมดีที่สุด เปรียบเทียบกับ MDL-SSS BIC-SSS BIC-SRA MDL-SRA และ HBIC-LTR ที่มีส่วนผสมเกาส์เซียนเป็น 4 โดยที่จะไม่นำ ML-SRA และ ML-SSS มาแสดงเนื่องจากในบทที่ 3 การประมาณทั้งสองแบบนี้ เมื่อนำมาสร้างเครื่องรู้จำแล้วมีความแม่นยำต่ำ

รูปที่ 33 การลงจุดระหว่างจำนวนพารามิเตอร์ของทอพอโลยีที่ประมาณได้กับความแม่นยำของทอพอโลยีเมื่อนำมาสร้างเครื่องรู้จำ



รูปที่ 33 การลงจุดระหว่างจำนวนพารามิเตอร์กับความแม่นยำของการทดลองที่ 4.2

4.2.3 วิเคราะห์ผลการทดลอง

ความแม่นยำของทอพอโลยีที่ได้จากการรันจีเอที่ใช้การคัดเลือกตามธรรมชาติแบบรัสเซียนรูเล็ตต์ถ่วงน้ำหนักนั้น มีความแม่นยำ 63.66% เมื่อเปรียบเทียบกับทอพอโลยีที่ประมาณได้จากบทที่ 3 แล้ว สามารถลดความผิดพลาดลงไปได้ 3.78% และ 4.36% เมื่อเปรียบเทียบกับ HBIC-LTR อีกทั้งเมื่อพิจารณาการลงจุดระหว่างความแม่นยำและจำนวนพารามิเตอร์แล้ว การรันจีเอทั้ง 4 ก็มีจำนวนพารามิเตอร์ใกล้เคียงกับวิธีการประมาณทอพอโลยีแบบอื่น ดังนั้น เราจึงพบว่า การใช้จีเอมาช่วยเลือกทอพอโลยีที่คัดเลือกมาแล้ว มีส่วนช่วยเพิ่มความแม่นยำของเครื่องรู้จำขึ้น

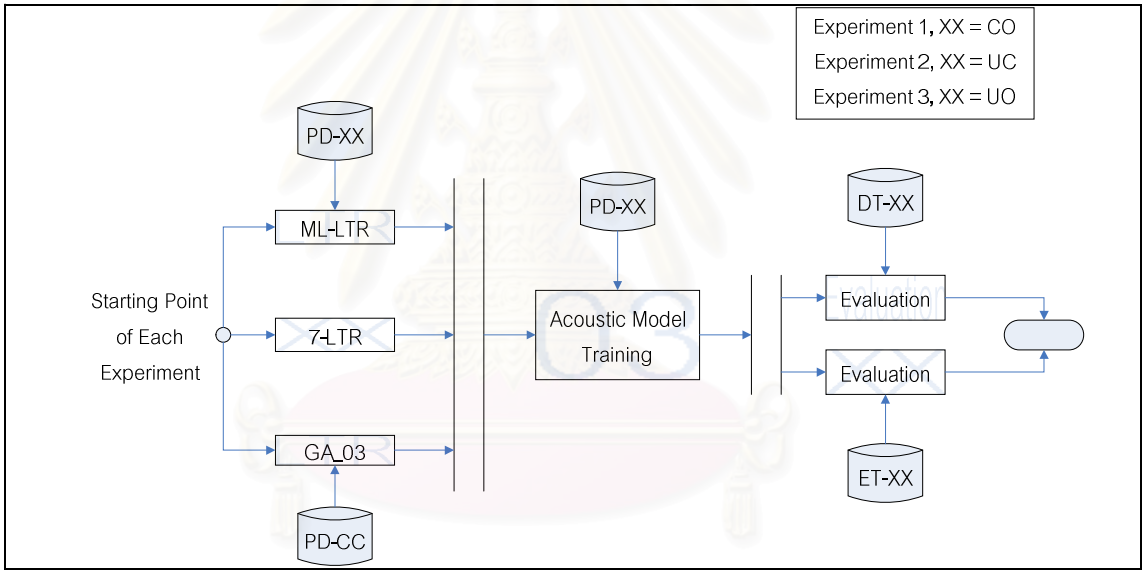
บทที่ 5
การทดลองประสิทธิภาพของทอพอโลยีที่ประมาณได้ในสภาพแวดล้อมอื่น ๆ

การประมาณทอพอโลยี บทที่ 3 และ 4 ได้นำเสนอวิธีการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟเพื่อใช้ในรู้จำหน่วยเสียงภาษาไทย โดยในการประมาณทอพอโลยีและฝึกฝนแบบจำลองทางเสียงใช้ฐานข้อมูลที่มีสภาพแวดล้อมเหมือนกัน จะเห็นว่า เมื่อนำไปสร้างเครื่องรู้จำหน่วยเสียงแบบบริบทอิสระโดยใช้ทอพอโลยีที่ประมาณได้ตามวิธีที่นำเสนอมีความแม่นยำในการรู้จำมากกว่าเครื่องรู้จำที่ทุกหน่วยเสียงใช้ทอพอโลยีแบบเดียวกัน และการนำขั้นตอนวิธีเชิงพันธุกรรมมาประยุกต์ใช้เพื่อบ่งบอกถึงความเหมาะสมโดยรวมก็ทำให้เราสามารถค้นพบทอพอโลยีที่มีความเหมาะสมมากยิ่งขึ้น แต่ปัญหาหนึ่งที่น่าสงสัย คือ ทอพอโลยีที่ประมาณได้นั้นเป็นทอพอโลยีที่มีความเหมาะกับข้อมูลที่ใช้ฝึกฝนมากน้อยเพียงใด นั่นคือ เราจำต้องประมาณทอพอโลยีใหม่ทุกครั้งเมื่อมีการเปลี่ยนสภาพแวดล้อมหรือไม่หรือความแตกต่างของสภาพแวดล้อมแบบใดบ้างที่ไม่ส่งผลกระทบต่อทอพอโลยีที่ประมาณได้ตามวิธีที่เสนอ ซึ่งคำถามนี้เป็นคำถามที่น่าสนใจเช่นกัน เนื่องจาก การประมาณทอพอโลยีนั้นค่อนข้างใช้ทรัพยากรมาก ดังนั้น ในบทนี้จึงทำการทดลองเพื่อทดสอบความเหมาะสมของทอพอโลยีที่ประมาณจากวิธีที่ได้นำเสนอโดยใช้ฐานข้อมูลในการฝึกฝนแบบจำลองทางเสียงและฐานข้อมูลสำหรับการประมาณทอพอโลยีต่างกัน เนื่องจากในการเก็บเสียงของโลตัสนั้นผู้พูดแต่ละคนจะทำการอัดเสียง 2 รอบ โดยรอบแรกเป็นการอัดเสียงในสภาพแวดล้อมเงียบผ่านไมโครโฟน 2 ตัวพร้อมกัน คือ แบบไดนามิกโคลสทอล์ค (TELEX H-41) และแบบยูนิไดเรคชันแนล (SONY F-720) ข้อมูลเสียงที่ได้แรกรอบนี้เป็นเสียงสะอาด ส่วนในรอบที่ 2 เป็นการอัดเสียงในสภาพแวดล้อมสำนักงานทั่วไปโดยผ่านไมโครโฟน 2 ตัวเช่นเดียวกับในรอบแรก ข้อมูลเสียงชุดนี้เรียกว่า เสียงสภาพแวดล้อมสำนักงาน เพื่อความสะดวกในการอ้างอิงถึงสภาพแวดล้อมและชนิดของไมโครโฟนที่ใช้บันทึกเสียงในฐานข้อมูลโลตัส วิทยานิพินธ์นี้จะใช้ตัวอักษรภาษาอังกฤษ 2 ตัวเขียนติดกัน โดยที่ตัวแรกจะ ระบุสภาพแวดล้อมของการบันทึกเสียง โดยใช้ตัวอักษณ C แทนการบันทึกในสภาพแวดล้อมเงียบ และสัญลักษณ์ O แทนการบันทึกเสียงในสภาพแวดล้อมสำนักงาน ตัวอักษรที่สองจะเป็ นการระบุถึงชนิดของไมโครโฟนที่ใช้ โดยหากบันทึกเสียงผ่านไมโครโฟนไดนามิกโคลสทอล์คจะใช้ตัวอักษร C และตัวอักษร U แทนแบบยูนิไดเรคชันแนล ดังนั้น เราจึงมีชุดข้อมูลเสียง 4 ชุด ที่มีสภาพแวดล้อมใกล้เคียงกันเพื่อใช้ในการทดลอง คือ CC (การบันทึกเสียงในสภาพแวดล้อมเงียบผ่านไมโคโฟนแบบไดนามิกโคลสทอล์ค) UC (การบันทึกเสียงในสภาพแวดล้อมเงียบผ่านไมโคโฟนแบบยูนิไดเรคชันแนล) CO (การบันทึกเสียงในสภาพแวดล้อมสำนักงานผ่านไมโคโฟนแบบไดนามิกโคลสทอล์ค)และ UO (การบันทึกเสียงในสภาพแวดล้อมสำนักงานผ่านไมโคโฟนแบบยูนิไดเรคชันแนล)

5.1 รายละเอียดการทดลอง

เนื่องจากเราต้องการทดลองว่า เมื่อสภาพแวดล้อมเปลี่ยนไปเราจำเป็นต้องประมาณทอพอโลยีตามวิธีที่ได้นำเสนอในบทที่ 3 และ 4 ใหม่หรือไม่ ดังนั้น เราจะทำการทดลองโดยเปรียบเทียบทอพอโลยี 3 แบบ ดังนี้

1. ทอพอโลยีที่ประมาณได้ในการทดลองที่ 4.2 (GA_03) ซึ่งเป็นทอพอโลยีที่ประมาณโดยใช้ชุดข้อมูลเสียง PD ในสภาพแวดล้อม CC
2. ทอพอโลยีแบบ 7 สถานะที่มีการรูปแบบการเชื่อมต่อเป็น LTR และมีจำนวนส่วนผสมเกาส์เซียนเป็น 4
3. ทอพอโลยีที่ประมาณได้จาก ML-LTR และเพิ่มจำนวนส่วนผสมเกาส์เซียนเป็น 4 โดยกำหนดสภาพแวดล้อมให้เหมือนกับสภาพแวดล้อมในการฝึกฝนแบบจำลองทางเสียงและการรู้จำ



รูปที่ 34 ขั้นตอนของการทดลองที่ 5

เหตุผลที่เลือกใช้ทอพอโลยี 3 แบบในการทดลองนี้ คือ ทอพอโลยีแบบแรกจะเป็นตัวแทนของทอพอโลยีที่ถูกประมาณจากสภาพแวดล้อมหนึ่ง แล้วนำมาฝึกฝนแบบจำลองทางเสียงและทดสอบในสภาพแวดล้อมที่แตกต่างออกไป ส่วนทอพอโลยีแบบที่สองจะเป็นตัวแทนของทอพอโลยีที่ไม่ได้ทำการประมาณใดๆ ทั้งสิ้น และทอพอโลยีแบบที่ 3 จะเป็นตัวแทนของทอพอโลยีที่ประมาณในสภาพแวดล้อมเดียวกับการฝึกฝนแบบจำลองทางเสียงและการทดสอบ เหตุผลที่เลือกใช้การประมาณ ML-LTR เพราะว่า การประมาณแบบนี้สามารถทำได้ง่ายและรวดเร็ว ซึ่งการตอบคำถามว่าเราต้องประมาณทอพอโลยีใหม่ เมื่อมีการเปลี่ยนสภาพแวดล้อมหรือไม่นั้น สามารถสังเกตได้จากผลการทดลองว่า หากทอพอโลยีแบบแรกยังคงให้ผลดี เมื่อเปรียบเทียบกับทอพอโลยี

ที่เหลือแล้ว เราก็มีความมั่นใจระดับหนึ่งว่า เราไม่จำเป็นต้องประมาณทอพอโลยีใหม่ทุกครั้ง แต่หากทอพอโลยีแบบแรกมีประสิทธิภาพต่ำกว่าทอพอโลยีแบบที่ 2 และ 3 แล้ว เราก็เชื่อมั่นได้ว่า เราจำเป็นต้องประมาณทอพอโลยีใหม่ทุกครั้ง เมื่อมีการเปลี่ยนสภาพแวดล้อม

จากรูปที่ 34 ทอพอโลยีทั้ง 3 แบบตามที่กล่าวไปข้างต้น จะถูกนำไปทดสอบกับสภาพแวดล้อม 3 แบบ คือ CO UC และ CO โดยจะฝึกฝนแบบจำลองทางเสียงด้วยข้อมูลเสียงชุด PD ของสภาพแวดล้อมที่กำหนดและวัดประสิทธิภาพด้วยชุดข้อมูลเสียง DT และ ET ตามสภาพแวดล้อมที่ใช้ฝึกฝนแบบจำลองทางเสียง สำหรับทอพอโลยีแบบที่ 1 นั้นก็คือ ทอพอโลยีที่ประมาณได้จากการทดลองที่ 4.2 นั่นเอง และทอพอโลยีแบบที่ 3 จะถูกประมาณทอพอโลยีโดยใช้ชุดข้อมูลเสียง PD ตามสภาพแวดล้อมที่กำหนดเพื่อค้นหาจำนวนสถานะและจำนวนส่วนผสมเกาส์เซียนเหมาะสมโดยใช้วิธีการประมาณเหมือนกับการทดลองในบทที่ 3

วิธีการวัดผลในการทดลองนี้ก็จะ วัดผลตามการทดลองในบทที่ 3 และบทที่ 4 คือ นำแบบจำลองทางเสียงของแต่ละหน่วยเสียงที่ฝึกฝนจากทอพอโลยีที่กำหนดมาสร้างเป็นเครื่องรู้จำหน่วยเสียงภาษาไทยแบบบริบทอิสระที่มีเมตริกความแปรปรวนร่วมเกี่ยวตามแนวทแยงโดยมีแบบจำลองทางภาษาเป็น ไบแกรมที่ฝึกฝนจากไฟล์กำกับหน่วยเสียงของชุดข้อมูล TR ดังแสดงในตารางที่ 21

ตารางที่ 21 องค์ประกอบของเครื่องรู้จำที่ใช้วัดผลในการทดลองที่ 5

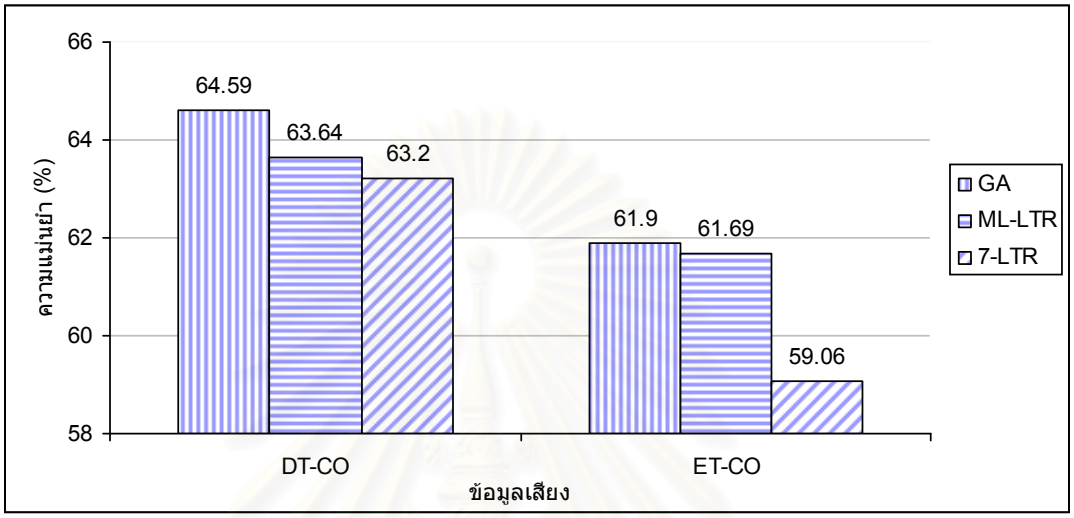
	รายละเอียด		
การสกัดค่าสำคัญ	วิธีการสกัดค่าสำคัญ	สัมประสิทธิ์บนสเกลเมลที่มีเวคเตอร์ลักษณะสำคัญขนาด 39 มิติที่คำนวณจากค่าพลังงงาน อัตราการเปลี่ยนแปลงและความเร่งหรืออย่างละ 13 มิติ	
	ขนาดเฟรม	10 มิลลิวินาที	
	ความกว้างวินโดว์	25 มิลลิวินาที	
หน่วยเสียง	74 หน่วยเสียงตามโลตัส รวมกับ sil และ sp		
แบบจำลองหน่วยเสียง	หน่วยเสียงแบบบริบทอิสระที่มีเมตริกความแปรปรวนร่วมเกี่ยวตามแนวทแยง		
แบบจำลองภาษา	ไบแกรมที่ฝึกฝนด้วยไฟล์กำกับหน่วยเสียงของชุดฝึกฝนในโลตัส		
ตัวอย่างเสียง	ตัวอย่างเสียงสำหรับการฝึกฝนแบบจำลองทางเสียง		PD
	ตัวอย่างเสียงสำหรับการประเมิน		DT,ET

5.2 ผลการทดลอง

สำหรับผลการทดลองจะแสดงเป็น 3 การทดลองย่อยโดยแบ่งตามสภาพแวดล้อมที่ใช้ในการฝึกฝนแบบจำลองทางเสียงและทดสอบ ซึ่งผลการทดลองเป็นดังนี้

5.2.1 ผลการทดลองในสภาพแวดล้อมแบบ CO

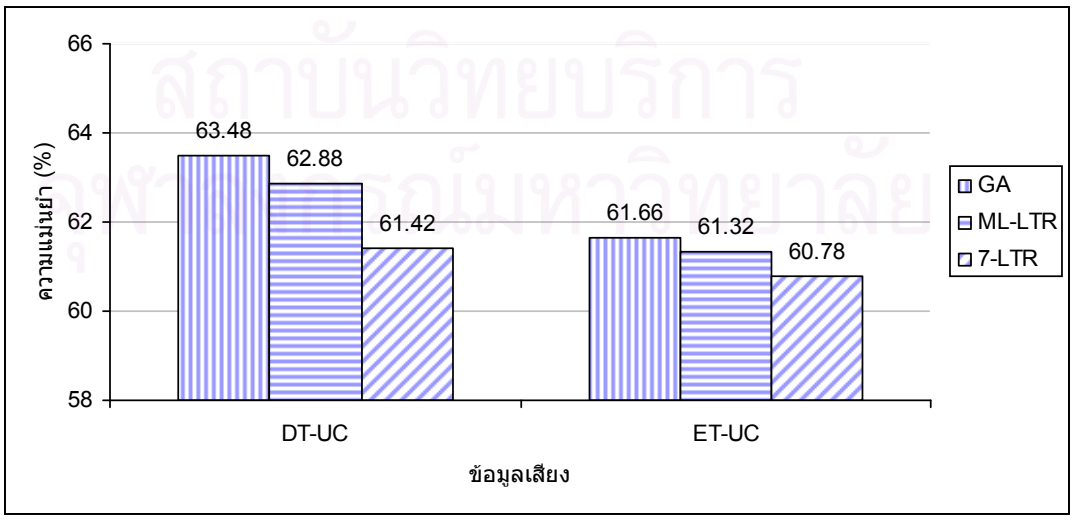
สภาพแวดล้อมแบบ CO คือ การบันทึกเสียงในสภาพแวดล้อมสำนักงานโดยใช้ไมโครโฟนแบบไดนามิกโคลสทอล์คซึ่งเป็นไมโครโฟนที่มีคุณภาพสูง โดย CO จะมีความแตกต่างกับ CC คือ เสียงรบกวนที่เกิดขึ้นระหว่างการอัดเสียง ซึ่งผลการทดลองเป็นดังรูปที่ 35



รูปที่ 35 ความแม่นยำของการรู้จำในสภาพแวดล้อม CO

5.2.2 ผลการทดลองในสภาพแวดล้อมแบบ UC

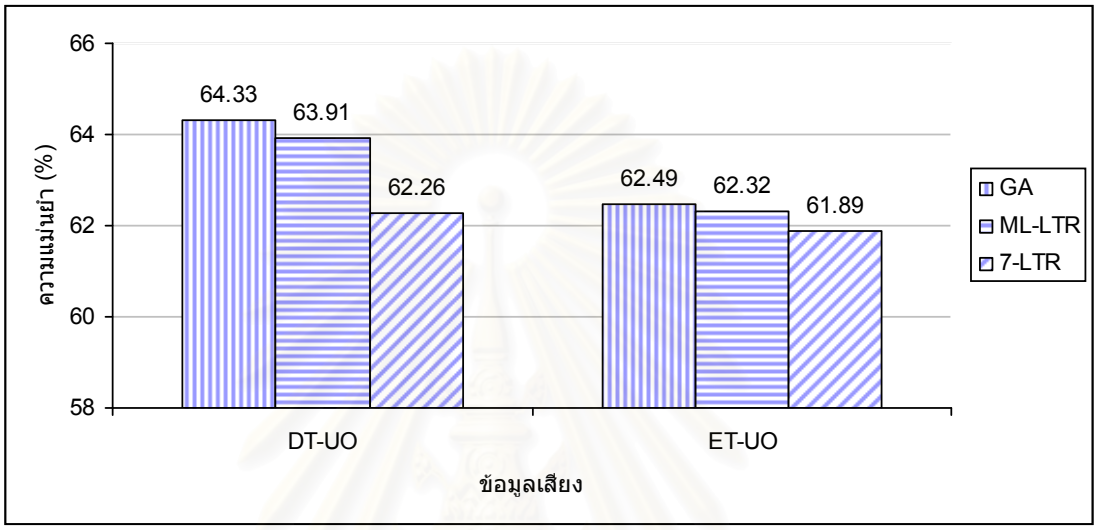
สภาพแวดล้อมแบบ UC คือ การอัดเสียงในห้องเงียบโดยจะเป็นการอัดเสียงพร้อมกับสภาพแวดล้อม CC ดังนั้น UC จึงมีความแตกต่างกับ CC เพียงช่องทางการบันทึกเสียง ซึ่งก็คือ ความแตกต่างของไมโครโฟนนั่นเอง โดย UC ใช้ไมโครโฟนแบบยูนิไดเรคชันแนล แต่ CC จะใช้ไมโครโฟนแบบไดนามิกโคลสทอล์ค ซึ่งผลการทดลองเป็นดังรูปที่ 36



รูปที่ 36 ความแม่นยำของการรู้จำในสภาพแวดล้อม UC

5.2.3 ผลการทดลองในสภาพแวดล้อมแบบ UO

สภาพแวดล้อมแบบ UO คือ การอัดเสียงใน สภาพแวดล้อมสำนักงานโดยจะอัดเสียงพร้อมกับสภาพแวดล้อม CO แต่ผ่านไมโครโฟนยูไดเรคชันแนล ดังนั้น UO จึงมีความแตกต่างกับ CC 2 ประการ คือ ช่องทางการบันทึกเสียงและเสียงรบกวนที่เกิดขึ้นระหว่างการอัดเสียง ซึ่งผลการทดลองสามารถแสดงได้รูปที่ 37



รูปที่ 37 ความแม่นยำของการรู้จำในสภาพแวดล้อม UO

5.3 วิเคราะห์ผลการทดลอง

เนื่องจากทุกสภาพแวดล้อมที่ใช้ทดลองเป็นผู้พูดชุดเดียวกันและพูดประโยคเดียวกัน สิ่งที่แตกต่างกันจึง คือ ช่องการบันทึกเสียง (ชนิดของไมโครโฟน) หรือสภาพแวดล้อมการบันทึกเสียง (สภาพแวดล้อมเงียบหรือสภาพแวดล้อมสำนักงาน) ซึ่งจากการทดลองย่อยทั้งสามสามารถวิเคราะห์ได้ดังนี้

1. จากผลการทดลองในสภาพแวดล้อมแบบ CO สามารถวิเคราะห์ได้ว่า เราไม่เห็นข้อบกพร่องหรือสิ่งผิดปกติ เมื่อนำทอพอโลยีที่ประมาณได้จากสภาพแวดล้อม CC มาฝึกฝนแบบจำลองทางเสียงและทดสอบในสภาพแวดล้อมแบบ CO เนื่องจาก ทอพอโลยีที่ประมาณได้จากจีเอก็ยังคงให้ค่าความแม่นยำสูงกว่า
2. เมื่อทดลองเปลี่ยนช่องทางของการบันทึกเสียง โดยข้อมูลเสียงที่ใช้ประมาณทอพอโลยีจะเป็นการบันทึกเสียงด้วยไมโครโฟนแบบไดนามิกโคลสทอล์ค และใช้ข้อมูลเสียงที่บันทึกด้วยไมโครโฟนโครโฟนแบบยูนิไดเรคชันแนลเป็นข้อมูลฝึกฝนแบบจำลองทางเสียงและทดสอบ จากผลการทดลอง เราก็ยังไม่เห็นข้อบกพร่อง

หรือสิ่งผิดปกติที่เกิดขึ้น กล่าวคือ ทอพอโลยีที่ประมาณได้จากจีเอก็ยังให้ความแม่นยำที่สูงสุด

3. เมื่อข้อมูลเสียงที่ใช้ในการประมาณทอพอโลยีและข้อมูลเสียงที่ใช้ในการฝึกฝนแบบจำลองทางเสียงและทดสอบ มีความแตกต่างกันทั้งชองทางการบันทึกการเสียงและสภาพแวดล้อมการบันทึก ผลการทดลองที่ได้ก็ยังมีแนวโน้มเป็นแบบการทดลองย่อย 1 และ 2 คือ ทอพอโลยีที่ประมาณได้จากจีเอมีความแม่นยำสูงสุด

จากการวิเคราะห์ผลการทดลองย่อยทั้งสาม ซึ่งการเปลี่ยนสภาพแวดล้อมของข้อมูลเสียงที่ใช้การฝึกฝนแบบจำลองทางเสียงโดยที่ผู้พูดและประโยคที่พูดยังคงเดิม สรุปได้ว่า เรายังไม่เห็นข้อผิดพลาดที่เกิดขึ้น เมื่อประมาณทอพอโลยีด้วยข้อมูลเสียงที่บันทึกเสียงด้วยไมโครโฟนแบบหนึ่ง แต่ฝึกฝนแบบจำลองทางเสียงและทดสอบด้วยข้อมูลเสียงที่บันทึกด้วยไมโครโฟนอีกชนิดหนึ่งและการบันทึกเสียงในสภาพแวดล้อมที่แตกต่างไม่มากนักระหว่างการบันทึกเสียงเพื่อประมาณทอพอโลยีและการฝึกฝนแบบจำลองทางเสียง เราก็ยังไม่เห็นข้อผิดพลาดที่เกิดขึ้นเช่นเดียวกัน

อีกประเด็นหนึ่งที่น่าสนใจ คือ ในฐานข้อมูลทดสอบชุด DT ทอพอโลยีที่ประมาณได้จากจีเอจะมีความแม่นยำสูงกว่าทอพอโลยีแบบ 7-LTR และทอพอโลยีที่ประมาณได้จาก ML-LTR อย่างเห็นได้ชัดเจนกว่าการวัดประสิทธิภาพโดยใช้ฐานข้อมูลชุด ET นั้นน่าจะมีสาเหตุมาจากการประมาณทอพอโลยีโดยประยุกต์ใช้จีเอได้นำข้อมูลเสียงชุด DT ที่บันทึกเสียงในสภาพแวดล้อม CC (บันทึกเสียงในสภาพแวดล้อมเงียบโดยใช้ไมโครโฟนแบบไดนามิกโคลสทอล์ค) เป็นข้อมูลเสียงเพื่อคิดคำนวนค่าความเหมาะของโครโมโซมและการบันทึกเสียงข้อมูลชุด DT ในสภาพแวดล้อมอื่นๆ นั้นก็ยังคงใช้ผู้พูดและประโยคเหมือนแบบเดิม จึงทำให้ทอพอโลยีที่ประมาณได้จากจีเอเปรียบเสมือนได้รับการปรับให้มีความเหมาะกับผู้พูดและประโยคไว้แล้วล่วงหน้าจึงทำให้ผลการวัดประสิทธิภาพโดยใช้ข้อมูลชุด DT ในสภาพแวดล้อมต่างๆ มีความแม่นยำสูง

บทที่ 6
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

6.1 สรุปผลการวิจัย

งานวิจัยได้นี้นำเสนอวิธีการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟสำหรับการรู้จำหน่วยเสียงภาษาไทยแบบบริบทอิสระ โดยใช้ขั้นตอนเชิ งพันธุกรรมเลือกทอพอโลยีที่ประมาณได้จากฟังก์ชันวัตถุประสงค์และวิธีการคัดเลือกทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟในแบบต่างๆ มาประกอบกัน เพื่อค้นหาทอพอโลยีที่เหมาะสมสำหรับหน่วยเสียงภาษาไทย เพื่อให้การรู้จำหน่วยเสียงภาษาไทยมีความแม่นยำมากยิ่งขึ้น จากผลการทดลองแล ะวิจัยสามารถสรุปได้ดังนี้

1. ทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟเป็นปัจจัยหนึ่งที่ส่งผลต่อความแม่นยำของการรู้จำเสียงพูดในภาษาไทย ซึ่งทอพอโลยีที่เหมาะสมนั้นจะช่วยให้มีความแม่นยำมากขึ้นโดย ที่ประมาณทอพอโลยีนั้นมีวิธีต่างๆ มากมาย แต่ละวิธีจะมีแนวคิดในการวัดความเหมาะสมแตกต่างกันออกไป โดยการเลือกวิธีการคัดเลือกทอพอโลยีจำต้องเลือกให้เหมาะสมกับฟังก์ชันวัตถุประสงค์ ดังเช่น การคัดเลือกทอพอโลยีแบบ SSS และ SRA ไม่เหมาะสมที่จะนำมาใช้กับ ML แต่เหมาะสมกับ BIC และ MDL แต่ในทางกลับกัน ML กลับเหมาะสมกับการคัดเลือกทอพอโลยีแบบกำหนดรูปแบบการเชื่อมต่อของสถานะไว้ล่วงหน้า ดังเช่น LTR มากกว่า BIC และ MDL

2. ทอพอโลยีที่มีการเชื่อมต่อแบบ LTR นั้น ถือเป็นทอพอโลยีที่มีประสิทธิภาพ เมื่อพิจารณาถึงความแม่นยำที่ได้และคำนึงถึงทรัพยากรที่ใช้ในการประมาณ โดยจะเห็นว่าทอพอโลยีแบบนี้ เมื่อนำไปสร้างเครื่องรู้จำเพื่อวัดประสิทธิภาพแล้วมีความแม่นยำอยู่ในเกณฑ์ที่น่าพอใจ

3. การค้นหาทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมนั้น ควรพิจารณาความเหมาะสมโดยรวม มิใช่คำนึงถึงเพียงความเหมาะสมของทอพอโลยีสำหรับแต่ละหน่วยเสียง เนื่องจากในการรู้จำเสียงพูดนั้นจะเป็นรู้จำเสียงพูดที่ต่อเนื่อง ซึ่งทำให้หน่วยเสียงแต่ละหน่วยเสียงจะถูกนำมาต่อกันเพื่อรู้จำ

6.2 ข้อเสนอแนะ

การประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่ได้ทดลองวิจัยนั้นเป็นวิธีการประมาณที่ได้จากการเรียนรู้โดยใช้ตัวอย่างเสียง ดังนั้น ทอพอโลยีที่ได้จึงขึ้นอยู่กับตั วอย่างที่นำมาใช้ในการเรียนรู้ นั่นคือ ทอพอโลยีที่ได้จะขึ้นอยู่กับฐานข้อมูลที่นำมาใช้ และในการประมาณทอพอโลยีนั้นใช้ทรัพยากรมาก กล่าวคือ หากไม่ใช้ เครื่องคอ มพิวเตอร์ประเภทคลัสเตอร์คิดคำนวณแล้วจะใช้เวลานานมาก เมื่อพิจารณาถึงฐานข้อมูลเสียงแล้ว ฐานข้อมูลเสียงที่นำมาใช้

ในการประมาณ (โลตัส) ก็เป็นฐานข้อมูลเสียงที่ได้รับการออกแบบมาเป็นอย่างดีและมีความหลากหลายของตัวอย่างเสียง อีกทั้งในการวัดผลในที่สร้างเครื่องรู้จำหน่วยเสียงแบบบริบทอิสระ ซึ่งในระบบการรู้จำเสียงทั่วไปนิยมใช้เครื่องรู้จำหน่วยเสียงแบบบริบทไม่อิสระซึ่งในวิทยานิพนธ์นี้ยังไม่ได้ค้นคว้าถึงการประมาณทอพอโลยีสำหรับหน่วยเสียงที่เป็นแบบบริบทไม่อิสระดังนั้น จึงมีข้อเสนอแนะดังนี้

1. หากมีเวลาที่จำกัด ควรเลือกใช้การประมาณที่ประหยัดทรัพยากร เช่น การประมาณที่กำหนดรูปแบบการเชื่อมต่อของทอพอโลยีไว้ล่วงหน้า โดยอาจกำหนดเป็น LTR LTRF หรือ FC เพื่อประมาณจำนวนสถานะและจำนวนส่วนผสมเกาส์เซียนหรือเลือกใช้วิธีแบบ SSS ซึ่งเป็นการเพิ่มจำนวนสถานะโดยการแยกสถานะที่สนใจออกเป็น 2 สถานะ ซึ่งเมื่อพิจารณาถึงจำนวนครั้งในการประมาณแล้วจะน้อยกว่าการประมาณแบบ SRA ค่อยๆ ลด การเปลี่ยนสถานะและจำนวนสถานะลงเรื่อยๆ มาก

2. การประมาณทอพอโลยีสำหรับหน่วยเสียงแบบบริบทไม่อิสระนั้นมีความเป็นไปได้ โดยอาศัยการโคลนหน่วยเสียงและการผูกในระดับที่ต่ำลงไปกว่าการผูกสถานะ เช่น การผูกส่วนประกอบของแบบจำลองส่วนผสมเกาส์เซียน (Gaussian Mixture Model: GMM)

3. ผลการวัดประสิทธิภาพการประมาณทอพอโลยีในบทที่ 3 นั้น การที่ ML ให้ผลคะแนนอยู่ในระดับเกณฑ์ดีเพราะว่าชุดข้อมูลเสียงในโลตัสมีความคล้ายกัน เนื่องจาก ข้อมูลเสียงทุกชุดจะถูกบันทึกเสียงจากสภาพแวดล้อมเดียวกัน อีกทั้งในแต่ละรอบ ผู้พูดจะได้รับคำแนะนำเพื่อปรับวิธีการพูด ความดังในการพูดและวิธีการพูดตามรูปอ่าน แต่หากเป็นฐานข้อมูลที่เก็บจากการพูดในชีวิตประจำวันแล้ว ML อาจจะให้ผลที่ไม่ดีนัก

4. วิธีการประมาณทอพอโลยีของแบบจำลองฮิดเดนมาร์คอฟที่ได้นำไปเสนอนั้นสามารถนำไปประยุกต์ใช้กับการรู้จำอย่างอื่น เช่น การรู้จำอักษขระแบบออนไลน์และการรู้จำอื่นๆ ที่อาศัยเทคนิคแบบจำลองฮิดเดนมาร์คอฟช่วยในการรู้จำได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] Zue, V., Glass, J.R., Phillips, M., and Sene, S. (1989). The MIT SUMMIT speech recognition system: a progress report. Proceedings of the Speech and Natural Language Workshop, 179–189.

[2] กิจศิริกุล, บ., and ทับทอง, ณ. (2548). การพัฒนาระบบรู้จำเสียงพูดภาษาไทย. รายงานวิจัยฉบับสมบูรณ์ โครงการวิจัยร่วมภาครัฐและเอกชน ปีงบประมาณ 2546: ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

[3] Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceeding of the IEEE 77: 257-286.

[4] Baum, L.E., and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. Annals of Mathematical Statistics 37: 1554-1563.

[5] Baum, L.E., and Egon, J.A. (1967). An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. Bulletin of the American Mathematical Society 73: 360-363.

[6] Baum, L.E., Petrie, T., and Sell, G.R. (1968). Growth functions for transformations on manifolds. Pacific Journal of Mathematics 27: 211-227.

[7] Baum, L.E., Petrie, T., Soules, G., and weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics 41: 164-171.

[8] Baum, L.E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov process. Inequalities 3: 1-8.

[9] Baker, J. (1975). The DRAGON system--An overview. IEEE Transactions on Acoustics, Speech, and Signal Processing 23: 24-29.

[10] Jelinek, F. (1969). A fast sequential decoding algorithm using a stack. IBM Journal Reseach and Development, 675-685.

[11] Bahl, L., and Jelinek, F. (1975). Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition. IEEE Transactions on Information Theory 21: 404-411.

[12] Jelinek, F., Bahl, L., and Mercer, R. (1975). Design of a linguistic statistical decoder for the recognition of continuous speech. IEEE Transactions on Information Theory 21: 250-256.

[13] Jelinek, F. (1976). Continuous speech recognition by statistical methods. Proceedings of the IEEE, 532-556.

[14] Jelinek, F., Bahl, L.R., and Mercer, R.L. (1982). Continuous speech recognition: Statistical methods. Amsterdam, The Netherlands: North Holland:

[15] Bahl, L.R., Jelinek, F., and Mercer, R.L. (1983). A maximum likelihood approach to continuous speech recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-5: 179-190.

[16] Roe, D.B., and Wilpon, J.G. (1993). Whither speech recognition: the next 25 years. Communications Magazine, IEEE 31: 54-62.

[17] ลีลาภัทรกิจ, พ. (2549). การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยแบบอาศัยเซกเมนต์โดยใช้สารสนเทศสวนสัทศาสตร์. วิศวกรรมศาสตรมหาบัณฑิต. สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

[18] Rabiner, L.R., and Juang, B.H. (1993). Fundamentals of Speech Recognition. Fundamentals of Speech Recognition. ed. Oppenheim, A, Englewood Cliffs, NJ:Prentice-Hall:

[19] Mammone, R.J., Xiaoyu, Z., and Ramachandran, R.P. (1996). Robust speaker recognition: a feature-based approach. Signal Processing Magazine, IEEE 13: 58.

[20] Hermansky, H. (1990). Perceptual Linear Predictive (PLP) Analysis of Speech. Journal of Acoustic Society of Americca 87: 1738-1752.

[21] Tyagi, V., McCowan, I., Misra, H., and Bourlard, H.A.B.H. (2003). Mel-cepstrum modulation spectrum (MCMS) features for robust ASR. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '03), 399-404.

[22] Furui, S. (1986). Speaker-independent isolated word recognition using dynamic features of speech spectrum. IEEE Transactions on Acoustics, Speech, and Signal Processing 34: 52-59.

[23] Young, S., Jansen, J., Odell, J. Ollason, D., and Woodland, P. (2005). The HTK Book (for HTK Version 3.3). Cambridge.

[24] Hosom, J.-P. (2006). Hidden Markov Models for Speech Recognition [Lecture Notes on CSE552/652]. Beaverton, Oregon, USA: Department of Computer Science & Engineering, Oregon Health&Science University.

[25] Kijsirikul, B. (2003). Artificial Intelligence [Lecture Notes on 2110654]. Chulalongkorn University.

[26] Lee, K.-F. (2000). CMU Sphinx [Computer program]. Carnegie Mellon University (Producer). Available from: http://cmusphinx.sourceforge.net/html/cmusphinx.php

[27] Zhao, J., Zhang, X., Ganapathiraju, A., Deshmukh, N., and Picone, J. (1999). Decision Tree-Based State Tying For Acoustic Modeling.

[28] นาคสกุล, ก. (2541). ระบบเสียงภาษาไทย. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

[29] Holland, J.H., Langton, C., and Wilson, S.W. (1998). Genetic Programming. Genetic Programming On the Programming of Computers by Means of Natural Selection, London, England.

[30] Singer, H., and Ostendorf, M. (1996). Maximum likelihood successive state splitting. Proceedings of the IEEE International Conferences on Acoustics, Speech and Signal Processing (ICASSP'96) 602: 601-604.

[31] Vasko, R.C., Jr., El-Jaroudi, A., and Boston, J.R. (1996). An algorithm to determine hidden Markov model topology. Proceedings of the IEEE International Conferences on Acoustics, Speech and Signal Processing (ICASSP'96), 3577-3580

[32] Danfeng, L., Biem, A., and Subrahmonia, J. (2001). HMM topology optimization for handwriting recognition. Proceedings of the IEEE International Conferences on Acoustics, Speech and Signal Processing (ICASSP'01), 1521-1524.

[33] Siriboon, K., Jirayusakul, A., and Kruatrachue, B. (2002). HMM topology selection for on-line Thai handwriting recognition. Proceedings of the First International Symposium on Cyber Worlds, 142-145.

[34] Jitsuhiro, T., and Nakamura, S. (2003). Variational Bayesian approach for automatic generation of HMM topologies. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '03), 77-82.

[35] Takatoshi, J., Tomoko, M., and Nakamura, S. (2003). Automatic generation of non-uniform context-dependent HMM topologies based on the MDL criterion. EUROSPEECH-2003, 2721-2724. Geneva, Switzerland.

[36] Biem, A., Ha, J.-Y., and Subrahmonia, J. (2002). A Bayesian model selection criterion for HMM topology optimization. Proceedings of the IEEE International Conferences on Acoustics, Speech and Signal Processing (ICASSP'02) 981: I-989-I-992

[37] Biem, A. (2003). A model selection criterion for classification: application to HMM topology optimization. Proceedings of the 7th International Confconference on Document Analysis and Recognition 101: 104-108

[38] Kasuriyam, S., Sornlertlamvanich, V., Cotsomrong, P., Kanokphara, S., and Thatphithakkul, N. (2003). Thai Speech Corpus for Thai Speech Recognition. Proc. of the COCOSDA'03.

[39] Han, S., Chang, M., Zou, Y., Chen, X.A., and Zhang, D.A. (2007). Systematic Multi-Path HMM Topology Design for Online Handwriting Recognition of East Asian Characters. Proceeding of the 9th International Conference on Document Analysis and Recognition (ICDAR'07), 604-608.

[40] Schwarz, G. (1978). Estimating the dimension of a model. Annals of Mathematical Statistics 6: 461-464.

[41] Meffert, K., Meseguer, J., Marti, E.D., Meskauskas, A., Vos, J., and Rotstan, N. (2007). Java Genetic Algorithms Package: JGAP [Computer program]. (Producer). Available from: http://jgap.sourceforge.net/

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก
หน่วยเสียงในภาษาไทยที่ใช้ในฐานข้อมูลโลตัส

หน่วยเสียงเดี่ยว (Monophone) มีจำนวนทั้งหมด 74 หน่วยเสียง ตารางที่ ก-1 แสดงตัวอย่างของคำที่ใช้กับสัญลักษณ์ของหน่วยเสียงที่กำหนดในฐานข้อมูลเสียงโลตัส

ตาราง ก-1 หน่วยเสียงในภาษาไทยที่กำหนดในโลตัส

พยัญชนะต้น (Ci)				สระ (V)			ตัวสะกด (Cf)		
เดี่ยว	ตัวอย่าง	ผสม	ตัวอย่าง	เดี่ยว	ตัวอย่าง	ผสม	ตัวอย่าง	เดี่ยว	ตัวอย่าง
p	ปาก	pr	ประสาน	a	อะ	ia	เอียะ	p^	พบ
t	เต้น,กฎิ	phr	พราน	aa	อา	iia	เอีย	t^	เกร็ด
c	จะ	tr	เตรียม	i	อิ	va	เอือะ	k^	ปาก
k	ก่อน	kr	กราบ	ii	อี	vva	เอือ	n^	หาร
z	อ่าน	khr	คร่า	v	อึ	ua	อัวะ	m^	ลม
ph	พบ,ภัย,ผ่าน	pl	ปลา	vv	อือ	uua	อัว	ng^	ฟาง
th	ทิ้ง,ธง,เฒ่า, ฐาน,มณโฑ	phl	พลาด	u	อุ	6 หน่วย		j^	ยาย
ch	ชอบ,เฌอ	thr	จันทรา	uu	อู			w^	กาว
kh	คน,เงิน,ฆ่า	kl	เกลอ	e	เอะ				
b	บอก	khl	เคลื่อน	ee	เอ			เสียงทับศัพท์	
d	ด้าน,ชฎา	kw	กวาง	x	แอะ			f^	กราฟ
m	ไม่	khw	ขวา	xx	แอ			l^	แอล
n	นาน,เณร			o	โอะ			s^	เอส
ng	เงิน	เสียงทับศัพท์		oo	โอ			ch^	คลัช
l	เล่น,กีฬา	br	เบรน	@	เอาะ			12 หน่วย	
r	รอ,ฤทัย	bl	บลู	@@	ออ				
f	ฝน,ฟัน	fr	ฟราย	q	เออะ				
s	สาย,ศิลา, รักษา,ซ่อน	fl	เฟลม	qq	เออ				
h	หิน,เฮฮา	dr	ดรากอน	18 หน่วย					
w	ว่า	17 หน่วย							
j	ย้อน,หญิง								
21 หน่วย									

โดยแบ่งหน่วยเสียงดังกล่าวออกเป็น 4 ประเภท คือ

1. เสียงพยัญชนะต้นเดี่ยว

ตาราง ก-2 เสียงพยัญชนะต้นเดี่ยว

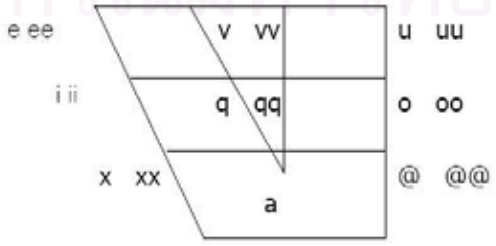
	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p ph b		t th d			k kh	z
Nasal	m		n			ng	
Fricative		f	s				h
Affricate				c ch			
Trill			r				
Approximant					j	w	
Lateral Approximant			l				

2. เสียงพยัญชนะท้ายเดี่ยว

ตาราง ก-3 เสียงพยัญชนะท้ายเดี่ยว

	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p^		t^			k^	
Nasal	m^		n^			ng^	
Fricative		f^	s^				
Affricate				ch^			
Trill			r^				
Approximant					j^	w^	
Lateral Approximant			l^				

3. เสียงสระเดี่ยว (Vowel)



รูปที่ ก-1 Vowel Chart

4. หน่วยเสียงผสม

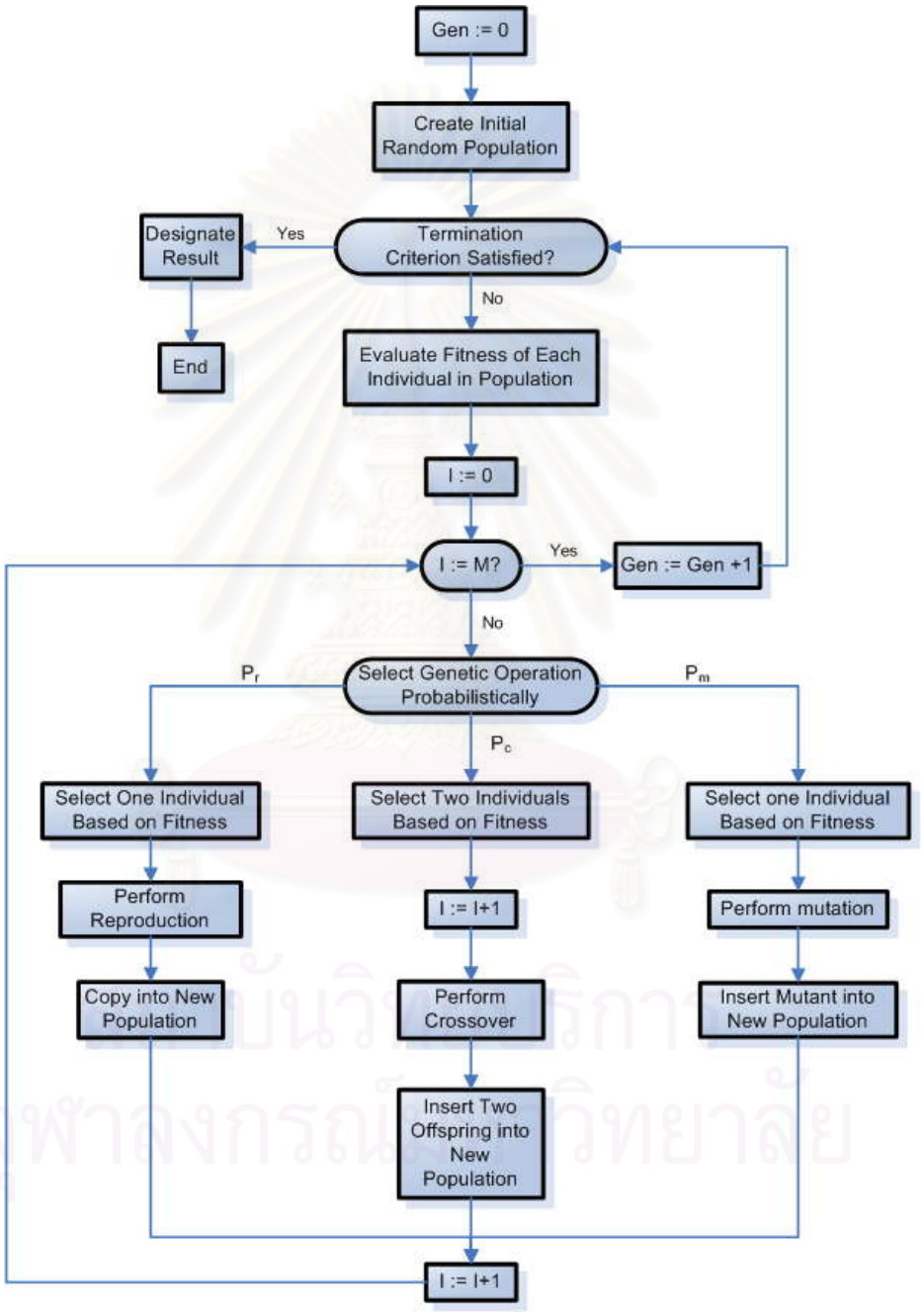
   4.1 เสียงควบกล้ำ (Cluster Consonant) ได้แก่ pr, phr, tr, kr, khr, pl, phl, kl, khl, kw และ khw

   4.2 เสียงสระผสม (Dipthong) ได้แก่ ia, iia, va, vva, ua และ uua

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข
แผนภูมิสายงานของขั้นตอนวิธีเชิงพันธุกรรม

รูปที่ ข-1 แสดงแผนภูมิสายงานของขั้นตอนวิธี เชิงพันธุกรรมแบบหนึ่ง โดยที่ I ใช้อ้างถึงประชากรตัวที่ I, M คือ ขนาดของประชากร และ Gen คือ รุ่นของประชากร



รูปที่ ข-1 แผนภูมิสายงานของขั้นตอนวิธีเชิงพันธุกรรม

ภาคผนวก ค
การเรียนรู้ การรู้จำและการวัดประสิทธิภาพเครื่องรู้จำเสียงพูด

การสร้างเครื่องจำเสียงพูดในวิทยานิพนธ์นี้ใช้โปรแกรม HMM-Toolkit: HTK [23] ซึ่งเป็นงานวิจัยของมหาวิทยาลัยเคมบริดจ์ (Cambridge University) โดยในส่วนนี้จะนำเสนอเกี่ยวกับการสร้างและรู้จำเครื่องรู้จำเสียงพูดด้วย HMM-Toolkit

1. การเรียนรู้ของเครื่องรู้จำ

การเรียนรู้ของเครื่องรู้จำ ประกอบด้วยการสกัดลักษณะสำคัญของเสียงพูด การเรียนรู้แบบจำลองทางเสียงและเรียนรู้แบบจำลองทางภาษา ซึ่งแต่ละส่วนมีกระบวนการดังนี้

(a) การสกัดลักษณะสำคัญของเสียงพูด

ลักษณะสำคัญที่วิทยานิพนธ์นี้เลือกใช้ คือ สัมประสิทธิ์เซปตรัมบนสเกลเมล การสกัดลักษณะสำคัญนี้สามารถทำได้โดยใช้โปรแกรม HCopy โดยกำหนดค่าพารามิเตอร์ต่างๆ ในไฟล์โครงแบบ (Configuration File) ดังรูปที่ ค-1

```
      # HTK Configuration Parameters for Generating MFCC_D_A_E
SOURCEFORMAT=WAV
SOURCEKIND = WAVEFORM
SOURCERATE = 625    # source sampling frequency is 16 [kHz]
TARGETKIND = MFCC_D_A_E
TARGETRATE=100000.0 # frame interval is 10 [msec]
SAVECOMPRESSED=F    # set T, if you like to save disk storage
SAVEWITHCRC=F
WINDOWSIZE=250000.0 # window length is 25 [msec]
USEHAMMING=T        # use HAMMING window
PREEMCOEF=0.97      # apply highpass filtering
NUMCHANS=24         #  number of filterbank for MFCC is 24
NUMCEPS=12          # number of parameters for MFCC presentation
ZMEANSOURCE=T
      # Rather local Parameters
ENORMALISE=F
ESCALE=1.0
RAWENERGY=F
     # CAUTION !! Do not use following option for nist encoded data.
# BYTEORDER=SUN
BYTEORDER=VAX
```

รูปที่ ค-1 ไฟล์โครงแบบสำหรับโปรแกรม HCopy

เมื่อเรียกใช้โปรแกรม HCopy และใช้ค่าปรับแต่งตามที่กำหนด ผลลัพธ์ที่ได้จะเป็นการสกัดลักษณะสำคัญของเสียงแบบสัมประสิทธิ์เซปตรัมบนสเกลเมลที่รวมเอาค่าพลังงาน อัตราการเปลี่ยนแปลง (Delta) และความเร่ง (Acceleration) คำนวณจากสัญญาณเสียงทุกๆ 10

มิลลิวินาที โดยมีขนาดความกว้าง 25 มิลลิวินาที โดยการเรียกใช้โป รแกรม HCopy นั้นสามารถเรียกได้ด้วยคำสั่ง

```
HCopy -C config -S scpList
```

โดยที่

- –C config เป็นการกำหนดให้ HCopy ใช้ไฟล์โครงแบบที่ชื่อว่า config
- –S scpList เป็นการกำหนดให้ HCopy ทำการสกัดลักษณะคำสั่งของไฟล์เสียงที่อยู่ในที่ถูกระบุใน scpList การระบุนั้นจำเป็นต้องบอกด้วยว่า จะนำผลลัพธ์ไปเก็บไว้ที่ใด โดยใช้เว้นวรรคคั่นระหว่างข้อมูลขาเข้าและข้อมูลขาออก เช่น หากระบุใน scpList ดังนี้

  CCF001_Pa001_001.wav CCF001_Pa001_001.mfc

  CCF001_Pa002_002.wav CCF002_Pa002_001.mfc

เมื่อเรียกคำสั่ง HCopy จะเป็นสกัดลักษณะสำคัญของไฟล์เสียง 2 ไฟล์ คือ CCF001_Pa001_001.wav กับ CCF001_Pa002_002.wav และผลที่ได้จะถูกเขียนอยู่ในไฟล์ชื่อ CCF001_Pa001_001.mfc CCF001_Pa002_001.mfc ตามลำดับ

(b) การเรียนรู้แบบจำลองเสียงพูด

การที่จะเรียนรู้แบบจำลองเสียงพูดนั้นจำเป็นที่จะเตรียมข้อมูลเพื่อนำไปใช้ในการเรียนรู้ ดังนี้

1. ไฟล์รายการหน่วยเสียง คือ ไฟล์ที่กำหนดหน่วยเสียงทั้งหมดที่ต้องการจะเรียนรู้แบบจำลอง โดยมีข้อกำหนดคือ แต่ละบรรทัดจะแทนหน่วยเสียงหนึ่งหน่วยเสียง ซึ่งต่อจากนี้จะใช้สัญลักษณ์ phnList แทนไฟล์รายการหน่วยเสียง
2. ไฟล์พจนานุกรม คือ ไฟล์ที่กำหนดคำศัพท์ของเครื่องรู้จำ โดยมีข้อกำหนดว่ารายการคำศัพท์จะต้องเรียงตามแบบพจนานุกรม แต่ละบรรทัดจะแทนคำศัพท์หนึ่งคำ พร้อมกับสตริงที่ต้องการให้ปรากฏในไฟล์ผลการรู้จำและหน่วยเสียงที่ใช้ประกอบเป็นคำนั้นๆ โดยหน่วยเสียงดังกล่าวต้องปรากฏอยู่ในไฟล์รายการด้วย ดังตัวอย่าง

   baan [home] b aa n^

   - baan คือ คำศัพท์ที่ถูกกำหนดขึ้น

- home คือ สตริงที่ต้องการแสดงเมื่อผลการรู้จำได้คำว่า baan โดยในส่วนนี้ต้องถูกเขียนอยู่ในเครื่องหมาย []
- b aa n^ คือ หน่วยเสียงที่ใช้ประกอบเป็นคำนั้นๆ

เนื่องจากในวิทยานิพนธ์นี้สร้างเครื่องรู้จำหน่วยเสียงบริบทอิสระ ไฟล์พจนานุกรมจึงเป็นรายการของหน่วยเสียงทั้งหมดที่ปรากฏอยู่ใน phnList โดยมีรูปแบบ

$phn [$phn] phn

เมื่อ $phn คือ หน่วยเสียงที่ปรากฏอยู่ใน phnList และต่อจากนี้จะใช้สัญลักษณ์ dict แทนพจนานุกรมคำศัพท์

3. ไฟล์ต้นแบบของแบบจำลองฮิดเดนมาร์คอฟ เป็นไฟล์ที่ระบุวิธีการสกัดค่าสำคัญของเสียง จำนวนสถานะ จำนวนส่วนผสมเกาส์เซียนที่ใช้ น้ำหนักของแต่ละส่วนผสมและเมตริกการเปลี่ยนสถานะ (Transition matrix) ดังแสดงในรูปที่ ค-2

```
~h "hmm2"
<BeginHMM>
<VecSize> 4 <MFCC>
<NumStates> 4
   <State> 2 <NumMixes> 2
      <Mixture> 1 0.4
      <Mean> 4
         0.3 0.2 0.2 1.0
      <Variance> 4
         1.0 1.0 1.0 1.0
      <Mixture> 2 0.6
      <Mean> 4
         0.1 0.0 0.0 0.8
      <Variance> 4
         1.0 1.0 1.0 1.0
   <State> 3 <NumMixes> 2
      <Mixture> 1 0.7
      <Mean> 4
         0.1 0.2 0.6 1.4
      <Variance> 4
         1.0 1.0 1.0 1.0
      <Mixture> 2 0.3
      <Mean> 4
         2.1 0.0 1.0 1.8
      <Variance> 4
         1.0 1.0 1.0 1.0
<TransP> 4
   0.0 1.0 0.0 0.0
   0.0 0.5 0.5 0.0
   0.0 0.0 0.6 0.4
   0.0 0.0 0.0 0.0
<EndHMM>
```

รูปที่ ค-2 ตัวอย่างของไฟล์ต้นแบบของแบบจำลองฮิดเดนมาร์คอฟ

สำหรับไฟล์ต้นแบบนั้นต่อจากนี้จะใช้สัญลักษณ์ proto_$phn แทนไฟล์ต้นแบบของแบบจำลองฮิดเดนมาร์คอฟของหน่วยเสียง $phn และ {proto_$phn} แทนไฟลต้นแบบของแบบจำลองฮิดเดนมาร์คอฟของหน่วยทั้งหมดที่ปรากฏอยู่ใน phnlist

4. ไฟล์ลักษณะสำคัญ คือ ไฟล์ผลลัพธ์ที่ได้ขั้นตอนการสกัดลักษณะสำคัญ ซึ่งในวิทยานิพนธ์นี้ข้อมูลเสียงจากชุดข้อมูลหน่วยเสียงสมดุลจึงมีทั้งหมด 1680 ไฟล์
5. ไฟล์รายการของไฟล์ลักษณะสำคัญ คือ รายการของไฟล์ผลลัพธ์ที่ได้ขั้นตอนการสกัดลักษณะสำคัญ โดยแต่ละบรรทัดให้เขียนชื่อไฟล์ลักษณะสำคัญเพียงหนึ่งไฟล์เท่านั้นหรือไฟล์รายการของไฟล์ลักษณะสำคัญ ก็คือ คอลัมน์ที่สองของ scpList ในการเรียกคำสั่ง HCopy นั่นเอง โดยจะขอใช้สัญลักษณ์ mfcList แทนไฟล์รายของไฟล์ลักษณะสำคัญ
6. ไฟล์กำกับหน่วยเสียง (Label file) คือ ไฟล์ที่กำกับหน่วยเสียงของไฟล์ลักษณะสำคัญแต่ละไฟล์ โดยมีข้อกำหนดว่า ชื่อของไฟล์กำกับหน่วยเสียงจะต้องเหมือนกับชื่อไฟล์ลักษณะสำคัญ เพียงแต่ต่างกันที่ชนิดของไฟล์โดยให้เปลี่ยนนามสกุลของไฟล์เป็น lab เช่น สำหรับไฟล์ลักษณะสำคัญ CCF001_Pa001_001.mfc ไฟล์ที่กำกับหน่วยเสียงของ CCF001_Pa001_001.mfc จะต้องชื่อว่า CCF001_Pa001_001.lab

(c) ขั้นตอนการเรียนรู้แบบจำลองเสียงพูด

เมื่อเตรียมไฟล์ที่จำเป็นสำหรับการเรียนรู้แบบจำลองเสียงพูดเรียบร้อยแล้วการเรียนรู้แบบจำลองเสียงพูดสามารถทำได้ โดยแบ่งเป็น 2 ขั้นตอน คือ

(i) การคำนวณค่าพารามิเตอร์เริ่มต้น

การประมาณค่าพารามิเตอร์เริ่มต้นสามารถทำได้โดยใช้โปรแกรมที่เรียกว่า HInit โดยเรียกใช้ด้วยคำสั่ง

```
HInit -C trainConfig -I $phn -m 1 -L labDir -o output proto_$phn -S scpTrainList
```

โดยที่

- -C trainConfig เป็นการกำหนดให้ HInit อ่านค่า โครงแบบจากไฟล์ที่ชื่อว่า trainConfig ซึ่งในไฟล์นี้จำเป็นต้องระบุค่าพารามิเตอร์ 3 ตัว คือ SOURCEKIND

TARGETKIND และ RAWENERGY โดยที่ค่าพารามิเตอร์ทั้งสามนี้จำเป็นที่จะต้องเหมือนกันที่กำหนดใน config ตอนแรกคำสั่ง HCopy

- -l $phn เป็นการกำหนดให้ HInit ค้นหา label $phn ในไฟล์กำกับหน่วยเสียงเพื่อนำมาใช้เป็นตัวอย่างหน่วยเสียงในการฝึกฝน
- -m N กำหนดให้หน่วยเสียงที่กำลังฝึกฝนจะต้องมีอย่างน้อย N ตัวอย่างขึ้นไป
- -L labDir เป็นการกำหนดให้ค้นหาไฟล์ลักษณะสำคัญที่ไดเรกทอรี่ labDir
- -o output คือ การกำหนดให้เขียนข้อมูลขาออกในไฟล์ที่ชื่อ output
- -S scpTrainList กำหนดให้ HInit ค้นหารายชื่อของไฟล์ลักษณะสำคัญเพื่อใช้ประมาณค่าพารามิเตอร์จากไฟล์ scpTrainList

ผลลัพธ์ที่ได้คือไฟล์ที่ชื่อ output ซึ่งเป็นไฟล์แบบจำลองเสียงพูด โดยหากต้องกา ระบุ ไดเรกทอรีเพื่อให้เขียนข้อมูลขาออกก็สามารถทำได้โดยใช้ -M เพื่อระบุไดเรกทอรีได้

(ii) การประมาณค่าพารามิเตอร์

เมื่อประมาณค่าพารามิเตอร์เริ่มต้นแล้ว เราสามารถประมาณค่าพามิเตอร์ได้โดยใช้ HRest ซึ่งเป็นโปรแกรมสำหรับประมาณค่าพารามิเตอร์ทีละหน่วยเสียง โดยการเรียกใช้ด้วยคำสั่ง

```
HRest -C trainConfig -l $phn -m 1 -L labDir output -S scpTrainList
```

โดยที่ความหมายของอาร์กิวเมนต์แต่ละตัวของ HRest เหมือนกับ HInit และ เมื่อเรียกคำสั่ง HRest จะเป็นการประมาณค่าพารามิเตอร์ที่อ่านได้ จาก $output ซึ่งเป็นไฟล์ข้อมูลขาออกของขั้นตอนการประมาณค่าเริ่มต้น โดยจะเขียนค่าพารามิเตอร์ที่ประมาณใหม่ในแต่ละรอบเก็บไว้ที่ไฟล์นี้ ดังนั้น ผลลัพธ์ที่ได้จากการเรียกคำสั่ง HRest จึงถูกเก็บไว้ใน output เช่นเดิม

(b) ขั้นตอนการเรียนรู้แบบจำลองทางภาษา

ในวิทยานิพนธ์ใช้แบบจำลองทางภาษาที่เรียกว่า ไบแกรม (Bigram หรือ Bigram Languge Model) ซึ่งอาศัยความน่าจะเป็นของการที่หน่วยเสียงหนึ่งจะปรากฏต่อจากอีกหน่วยเสียงหนึ่ง การเรียนรู้แบบจำลองทางภาษานั้นจำต้องใช้โปรแกรมของ HTK 2 โปรแกรมด้วยกันคือ HLStats เพื่อเก็บรวบรวมค่าความน่าจะเป็นของการปรากฏหน่วยเสียง 2 หน่วยเสียงใด ๆ ต่อกัน และ HBuild เพื่อสร้างแบบจำลองทางภาษา โดยทั้งสองโปรแกรมนี้สามารถเรียกได้ด้วยคำสั่งดังต่อไปนี้

(i) การเรียกใช้โปรแกรม HLStats

```
HLStats -C config -b phnBig phnList -S scpLMTrainList
```

โดยที่

- -C config คือ การกำหนดให้ HLStats ใช้ค่าโครงแบบตาม config ซึ่งคือ ค่าเดียวกับที่ใช้ในการเรียกคำสั่ง HCopy
- -b phnBig กำหนดให้เขียนข้อมูลขาออก HLStats ไปที่ phnBig
- phnList คือ ไฟล์รายการของหน่วยเสียงที่ใช้ทั้งหมด โดยจำต้องเหมือนกับ phnList ที่ใช้ในขั้นตอนการเรียนรู้แบบจำลองเสียงพูด
- -S scpLMTrainList คือ การกำหนดให้ HLStats อ่านไฟล์ที่ใช้ในการฝึกฝนจาก scpLMTrainList ซึ่งในวิทยานิพนธ์นี้ใช้ไฟล์กำกับหน่วยเสียงของข้อมูลเสียงชุดฝึกฝนในโลตัสเป็นข้อมูลฝึกฝน ดังนั้น scpLMTrainList เป็นไฟล์รายการของไฟล์กำกับหน่วยเสียงทั้งหมดของข้อมูลเสียงชุดฝึกฝน

(ii) การเรียกใช้โปรแกรม HBuild

```
HBuild -C config -n phnBig phnList phnNet
```

โดยที่

- -C config คือ การกำหนดให้ HBuild ใช้ค่าโครงแบบตาม config ซึ่งคือ ค่าเดียวกับที่ใช้ในการเรียกคำสั่ง HLStats
- -n phnBig เป็นการกำหนดให้ HBuild อ่าน phnBig ซึ่งเป็นไฟล์ที่เก็บรวมสถิติของไบแกรมสำหรับคู่หน่วยเสียงทั้งหมด
- phnList คือ ไฟล์รายการของหน่วยเสียงที่ใช้ทั้งหมดโดยจำต้องเหมือนกับ phnList ที่ใช้ในขั้นตอนการเรียนรู้แบบจำลองเสียงพูด
- phnNet คือ ไฟล์ข้อมูลขาออกที่เป็นแบบจำลองทางภาษา

2. **การรู้จำเสียงพูด**

การรู้จำเสียงพูดเป็นการนำแบบจำลองทางภาษาและแบบจำลองเสียงพูดมาใช้เพื่อรู้จำเสียงพูดที่เราไม่ทราบ การรู้จำเสียงพูดนั้นสามารถทำได้โดยใช้โปรแกรม HVite ด้วยคำสั่งดังนี้

```
HVite -S scpTest -d  phnDir -w phnNet -i outMLF phnDict phnList
```

โดยที่

- -S scpTest คือ การกำหนดให้รู้จำไฟล์ลักษณะสำคัญที่ถูกกำหนดอยู่ใน scpTest
- -d phnDir คือ กำหนดให้ HVite ค้นหาแบบจำลองเสียงพูดในไดเรกทอรี phnDir
- -w phnNet เป็นการกำหนดให้ใช้แบบจำลองทางภาษา phnNet
- -i outMLF เป็นการกำหนดให้เขียนผลการรู้จำไว้ที่ outMLF
- phnDict คือ การระบุไฟล์พจนานุกรมที่ต้องการใช้
- phnList คือ การระบุไฟล์รายการของหน่วยเสียงที่มีทั้งหมด

3. การประเมินผล

การวัดประสิทธิภาพของเครื่องรู้จำเสียงพูดสามารถทำได้ด้วยโปรแกรม HResults โดยจะเป็นการเปรียบเทียบไฟล์กำกับหน่วยเสียงที่ได้จากการรู้จำกับไฟล์กำกับหน่วยเสียงอ้างอิง โดยการเปรียบเทียบนั้นจะใช้วิธีการปรับแนวสายอักขระโดยใช้กำหนดการพลวัต (Dynamic Programming-based String Alignment) การวิเคราะห์นั้นจะอยู่ในรูปแบบ ดังแสดงในรูปที่ ค-3

```
---------------------------------------------------- Overall Results ----------------------------------------------------
SENT: %Correct=13.00 [H=13, S=87, N=100]
WORD: %Corr=53.36, Acc=44.90 [H=460,D=49,S=353,I=73,N=862]
================================================================================
```

รูปที่ ค-3 ผลการวิเคราะห์ประสิทธิภาพเมื่อใช้โปรแกรม HResults

ในบรรทัดแรก จะบอกถึงเปอร์เซ็นต์ความแม่นยำระดับประโยคโดยหนึ่งไฟล์กำกับหน่วยเสียง คือ หนึ่งประโยค ส่วนบรรทัดที่สองบอกความแม่นยำระดับคำ โดย H คือ จำนวนของหน่วยเสียงที่ตรงกับไฟล์กำกับหน่วยเสียงอ้างอิง I คือ จำนวนหน่วยเสียงแทรก (Insertion) ที่เกิดขึ้นเมื่อเทียบกับไฟล์กำกับหน่วยเสียงอ้างอิง D คือ จำนวนหน่วยเสียงที่หายไป (Deletion) ที่เกิดขึ้นเมื่อเทียบกับไฟล์กำกับหน่วยเสียงอ้างอิง S คือ จำนวนหน่วยเสียงแทนที่ (Substitution) ที่เกิดขึ้นเมื่อเทียบกับไฟล์กำกับหน่วยเสียงอ้างอิง และ N คือ จำนวนหน่วยเสียงทั้งหมดโดยการเรียกใช้โปรแกรม HResults สามารถทำได้ดังนี้

```
HResults -I refMLF phnList recoutMLF
```

โดยที่

- -I refMLF คือ กำหนดให้ใช้ refMLF เป็นไฟล์กำกับหน่วยเสียงอ้างอิง

- phnList คือ ไฟล์รายการหน่วยเสียง
- recoutMLF คือ ไฟล์ผลลัพธ์ของการรู้จำที่ต้องการนำมาเปรียบเทียบกับไฟล์กำกับหน่วยเสียงอ้างอิง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นายพัฒนะ ภูริยากร เกิดเมื่อวันที่ 12 ธันวาคม พ.ศ. 2527 ที่จังหวัดจันทบุรี สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นจากโรงเรียนเบญมราชูทิศจังหวัดจันทบุรีและสำเร็จการศึกษาระดับมัธยมตอน ปลายจากศูนย์การศึกษานอกโรงเรียนจังหวัดจันทบุรี สำเร็จการศึกษาระดับปริญญาบัณฑิต ในสาขาวิชาวิศวกรรมคอมพิวเตอร์ จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2548

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย