# การสกัดแนวคิดจากเสียงสนทนาโดยใช้กราฟของเซกเมนต์

นางสาวศิรินาถ ตั้งรวมทรัพย์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

# CONCEPT EXTRACTION FROM CONVERSATIONAL SPEECH USING SEGMENT GRAPH

Miss Sirinart Tangruamsub

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 | การสกัดแนวคิดจากเสียงสนทนาโดยใช้กราฟของเซกเมนต์ |
| โดย | นางสาวศิรินาถ ตั้งรวมทรัพย์ |
| สาขาวิชา | วิศวกรรมคอมพิวเตอร์ |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | อาจารย์ ดร.อติวงศ์ สุชาโต |

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาตราจารย์ ดร.บุญเสริม กิจศิริกุล)

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

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

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

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ศิรินาถ ตั้งรวมทรัพย์ : การสกัดแนวคิดจากเสียงสนทนาโดยใช้กราฟของเซกเมนต์. (CONCEPT EXTRACTION FROM CONVERSATIONAL SPEECH USING SEGMENT GRAPH) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : อ.ดร.อติวงศ์ สุชาโต, 118 หน้า.

วิทยานิพนธ์นี้ได้เสนอแนวทางในการค้นหาลำดับของแนวคิดจากกราฟของเซกเมนต์ โดยการเพิ่มความรู้ของแนวคิดในขณะที่รู้จำคำจากกราฟของเซกเมนต์ ซึ่งจะใช้แบบจำลองแนวคิด และโครงข่ายจับคู่ระหว่างคำศัพท์และแนวคิดมาช่วยในการรู้จำ ทั้งนี้เพื่อให้สามารถใช้ความรู้ในส่วนของหน่วยเสียงมาช่วยในการค้นหาลำดับของแนวคิด ในขณะเดียวกันก็สามารถใช้ความรู้ของแนวคิดมาพิจารณาค้นหาลำดับของคำ โดยได้เปรียบเทียบกับวิธีการที่ค้นหาลำดับของแนวคิดจากผลการรู้จำคำที่ได้ ซึ่งเป็นวิธีการที่แยกความรู้ของหน่วยเสียง และแนวคิดออกจากกัน โดยในวิทยานิพนธ์นี้จะเปรียบเทียบประสิทธิภาพของทั้งสองวิธี โดยพิจารณาจากผลของการรู้จำคำศัพท์ การค้นหาแนวคิด และการค้นหาจุดมุ่งหมายของประโยคจากแนวคิดที่ได้ภายใต้โดเมนของการติดต่อทางโทรศัพท์เพื่อสอบถามอัตราค่าบริการ และรหัสโทรศัพท์ทางไกลระหว่างประเทศ ผ่านทางระบบโทรศัพท์ จากผลการทดลองพบว่า เมื่อใช้ความรู้ของแนวคิดมาใช้พร้อมๆ กับการรู้จำคำนั้น จะเพิ่มค่ารีคอลในการรู้จำคำจาก 54.93% เป็น 70.59% โดยเมื่อเปรียบเทียบผลของการค้นหาลำดับของแนวคิด จะพบว่าวิธีการค้นหาแนวคิดจากกราฟของเซกเมนต์นั้นจะให้ค่ารีคอลและพรีซีชันมากที่สุด โดยจะมีค่าเป็น 81.73% และ 86.67% ตามลำดับ ในการระบุจุดมุ่งหมายนั้น ได้เปรียบเทียบวิธีการแบ่งประเภทโดยพิจารณาจากการปรากฏของแนวคิดที่รู้จำได้ และการใช้คะแนนของการรู้จำในการพิจารณา โดยวิธีการใช้กฎและโครงข่ายประสาทเทียม ผลการทดลองแสดงให้เห็นว่า การใช้กฎโดยพิจารณาจากคะแนนของการรู้จำ จะให้ความถูกต้องในการระบุจุดมุ่งหมายมากที่สุดเป็น 79.63%

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต ศิรินาถ ตั้งรวมทรัพย์

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก อติวงศ์

ปีการศึกษา 2550

# # 4970599021 : MAJOR COMPUTER ENGINEERING

KEY WORD: CONCEPT EXTRACTION / CONVERSATIONAL SYSTEM / SPEECH UNDERSTANDING / SPEECH PROCESSING

SIRINART TANGRUAMSUB : CONCEPT EXTRACTION FROM CONVERSATIONAL SPEECH USING SEGMENT GRAPH. THESIS PRINCIPAL ADVISOR : ATIWONG SUCHATO, Ph.D., 118 pp.

This thesis proposes the methodology for extracting the sequence of concept from segment graph by adding concept knowledge into the word recognition process. For this propose, concept model and the graph mapping between word and concept are used so that the acoustic-phonetic knowledge can be use for searching concepts and the concept knowledge also used for finding the sequence of words. This thesis compares the performance of this method with the concept extraction approch which tags the concepts from the sequence of words derived from speech recognizer, in the other words the approach which separates the acoustic-phonetic knowledge and the concept knowledge. In order to evaluate the performance, the result of word recognition, concept extraction and goal classification are compared. The experiment set on the telephone rate and country code inquiry system. The result of experiment show that the process which integrated concept knowledge increase recall value from 54.93% to 70.59%. Comparing extracted concept output, extracting concept through segment graph provide the highest precision and recall value, 81.73% and 86.67%, respectively. To classify goal, the accuracy of goal identification methods considereing the presence of concepts and the score of each concept are compared. We proposed rule-based and artificial neural network classifying approach. The result show that using rule-based rely on recognized concept scores provide the highest accuracy which is 79.63%

Department....Computer Engineering.....Student's signature....ศิรินาถ ตั้งอรุณทวีพร์....
Field of study....Computer Engineering.....Principal Advisor's signature.....
Academic year.....2007.....

# กิตติกรรมประกาศ

ตลอดระยะเวลาในการศึกษาและทำวิทยานิพนธ์นั้น ได้พบกับสิ่งต่างๆ มากมาย ทั้งในเรื่องดี และอุปสรรคต่างๆ จนอาจกล่าวได้ว่า หากต้องทำวิทยานิพนธ์นี้เพียงลำพัง คงไม่สามารถได้ผลในระยะเวลาเท่านี้ได้ แต่เป็นเพราะว่าได้รับการสนับสนุนและความช่วยเหลือจากคนรอบข้างต่างๆ จึงเป็นแรกผลักดันให้เกิดความก้าวหน้าในงานวิทยานิพนธ์นี้ แรกสุด ต้องขอขอบพระคุณ อาจารย์ ดร. อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาที่สละเวลาอันมีค่าในการให้คำปรึกษา ชี้แนะ และให้โอกาสต่างๆ ในระหว่างการศึกษานี้ สิ่งต่างๆ ที่อาจารย์ได้มอบให้นั้น ได้เป็นประโยชน์อย่างมากในการทำงานทั้งในปัจจุบัน และอนาคตต่อไป นอกจากนี้ยังต้องขอขอบพระคุณคณะกรรมการ อันได้แก่ ศาสตราจารย์ ดร. บุญเสริม กิจศิริกุล, รองศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ และ ดร.ชัย วุฒิวิวัฒน์ชัย ที่ได้สละเวลาให้คำแนะนำ และชี้แนะข้อบกพร่องและประเด็นต่างๆ ทั้งนี้ก็เพื่อให้วิทยานิพนธ์นี้ได้สำเร็จลงได้อย่างสมบูรณ์

นอกจากนี้ยังต้องขอขอบคุณห้องปฏิบัติการ ทั้งพี่ๆ เพื่อนๆ และน้องๆ ที่คอยให้ความช่วยเหลือ และให้กำลังใจต่างๆ ในระหว่างการศึกษานั้น มีหลายครั้งที่เกิดความเหนื่อยล้า และท้อใจ แต่กำลังใจต่างๆ ที่ได้มานั้น ก็ช่วยให้ลืมสิ่งเหล่านี้ได้หมดสิ้น และก่อให้เกิดแรงกระตุ้นในการทำงานอย่างประหลาด หากให้กล่าวชื่อของทุกคนนั้น คงไม่สามารถบอกได้ครบทุกคน แต่อย่างไรก็ตามก็ขอกล่าวคำว่าขอบคุณมา ณ ที่นี้ด้วย

บุคคลที่คงต้องกล่าวคำขอบคุณมากๆ คือ คุณพ่อ คุณแม่ และครอบครัว ที่เข้าใจ และคอยสนับสนุนในระหว่างการศึกษา มีหลายครั้งที่ได้สร้างความกังวลและเป็นห่วง หรือแม้แต่ไม่มีเวลาให้กับครอบครัว และมีหลายครั้งที่คุณพ่อคุณแม่ปริปากบ่นบ้างด้วยความเป็นห่วง แต่ท้ายที่สุดก็ได้แสดงให้เห็นว่า สิ่งที่คุณพ่อและคุณแม่ให้ความสำคัญมากที่สุดนั้น คือ การศึกษาและอนาคตของลูก ขอขอบพระคุณมา ณ ที่นี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญ

หน้า


บทคัดค่อภาษาไทย ..... ง
บทคัดค่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ญ
สารบัญรูปภาพ ..... ฎ
บทที่ 1 บทนำ ..... 1
1.1 ที่มาและความสำคัญของปัญหา ..... 1
1.2 วัตถุประสงค์ของการวิจัย ..... 2
1.3 ขอบเขตของการวิจัย ..... 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ ..... 4
1.5 ผลงานตีพิมพ์จากวิทยานิพนธ์ ..... 4
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง ..... 5
2.1 แนวคิดและทฤษฎี ..... 5
2.1.1 การรู้จำเสียงพูด (Speech Recognition) ..... 5
2.1.2 ตัวแปรสุ่มแบบต่อเนื่อง ..... 7
2.1.3 อัลกอริทึมวิเทอร์บี (Viterbi algorithm) ..... 10
2.1.4 ตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนัก (Weight Finite State Transducer-WFST) ..... 11
2.1.5 การวิเคราะห์ทางวากยสัมพันธ์ (Syntactic processing) ..... 13
2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง ..... 20
2.2.1 ส่วนที่ทำการค้นหาลำดับแนวคิดจากเสียงพูด ..... 20
2.2.2 ส่วนที่ทำการค้นหาจุดมุ่งหมายจากแนวคิดที่ได้ ..... 24
บทที่ 3 การรู้จำเสียงแบบเซกเมนต์ ..... 27
3.1 การออกแบบระบบการรู้จำเสียงแบบเซกเมนต์ ..... 27
3.2 การสร้างกราฟของเซกเมนต์ ..... 27
3.3 การสกัดค่าลักษณะสำคัญ ..... 30
3.4 การค้นหาเส้นทางของเซกเมนต์ ..... 32

3.4.1 การคำนวณค่าความน่าจะเป็น ..... 32
3.5 แบบจำลองต่างๆ ..... 36
3.5.1 แบบจำลองหน่วยเสียง ..... 36
3.5.2 แบบจำลองระยะเวลา (Duration Model) ..... 40
3.5.3 แบบจำลองการออกเสียง (Pronunciation Model) ..... 40
3.5.4 แบบจำลองภาษา ..... 41
3.6 การค้นหา ..... 42
บทที่ 4 การค้นหาแนวคิด และจุดมุ่งหมาย ..... 53
4.1 ความหมายของแนวคิด ..... 53
4.2 การค้นหาแนวคิดจากผลลัพธ์จากการรู้จำ ..... 54
4.3 การค้นหาแนวคิดโดยใช้แบบจำลองแนวคิด ..... 57
4.3.1 การค้นหาแนวคิดจากกราฟของเซกเมนต์ ..... 58
4.3.2 โครงข่ายการจับคู่คำศัพท์กับแนวคิดของคำศัพท์ ..... 59
4.3.3 แบบจำลองแนวคิด ..... 61
4.3.4 การสร้างโครงข่ายการออกเสียง ..... 62
4.4 การค้นหาจุดมุ่งหมายของประโยคจากลำดับของแนวคิด ..... 67
บทที่ 5 การทดลอง ..... 68
5.1 ระบบตัวอย่างที่ใช้ในการทดสอบ ..... 68
5.2 ตัวอย่างเสียง ..... 68
5.2.1 ตัวอย่างเสียงเพื่อใช้ในการเรียนรู้ ..... 69
5.2.2 ตัวอย่างเสียงเพื่อใช้ในการทดสอบ ..... 70
5.3 การออกแบบระบบการทดลอง ..... 70
5.3.1 การออกแบบคำศัพท์ แนวคิด และจุดมุ่งหมาย ..... 70
5.3.2 การสร้างกราฟของเซกเมนต์ ..... 73
5.3.3 การพิจารณาเวกเตอร์ลักษณะสำคัญของเซกเมนต์ ..... 73
5.3.4 แบบจำลองการออกเสียง และโครงข่ายจับคู่ระหว่างคำศัพท์และแนวคิด ..... 73
5.3.5 แบบจำลองภาษา และแบบจำลองแนวคิด ..... 74
5.4 การวัดประสิทธิภาพ ..... 75

5.5 การทดลองในการรู้จำเสียงพูด ..... 76
5.5.1 ผลการทดลอง ..... 77
5.5.2 วิเคราะห์ผลการทดลอง ..... 77
5.6 การทดลองในการค้นหาลำดับของแนวคิด ..... 80
5.6.1 ผลการทดลอง ..... 80
5.6.2 วิเคราะห์ผลการทดลอง ..... 81
5.7 การทดลองในการแบ่งประเภทจุดมุ่งหมาย ..... 85
5.7.1 ผลการทดลอง ..... 88
5.7.2 วิเคราะห์ผลการทดลอง ..... 89
บทที่ 6 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ ..... 94
6.1 ผลการวิจัย ..... 94
6.1.1 ผลการวิจัยในการเปรียบเทียบประสิทธิภาพในการรู้จำคำ ..... 94
6.1.2 ผลการวิจัยในการเปรียบเทียบประสิทธิภาพในการค้นหาแนวคิด ..... 96
6.1.3 ผลการวิจัยในการเปรียบเทียบประสิทธิภาพในการระบุจุดมุ่งหมายของประโยค .... 97
รายการอ้างอิง ..... 98
ภาคผนวก ..... 100
ภาคผนวก ก สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงพยัญชนะต้นในภาษาไทย. 101
ภาคผนวก ข สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงตัวสะกดในภาษาไทย ..... 103
ภาคผนวก ค สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงสระในภาษาไทย ..... 104
ประวัติผู้เขียนวิทยานิพนธ์ ..... 105

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญตาราง

หน้า

ตารางที่ 3-1แบบจำลองที่ใช้ในการคำนวณของพจน์ต่างๆ ในสมการที่3.6 ..... 34
ตารางที่ 3-2 การแบ่งกลุ่มของหน่วยเสียง ..... 39
ตารางที่ 3-3 สมมุติฐานในกรณีต่างๆ ของแต่ละเซกเมนต์ ..... 49
ตารางที่ 4-1 การแบ่งพิจารณาประเภทของแนวคิด ..... 53
ตารางที่ 5-1 แนวคิด และจำนวนคำศัพท์ที่ใช้ในการทดลอง ..... 71
ตารางที่ 5-2 ตัวอย่างประโยคสนทนาของแต่ละจุดมุ่งหมาย ..... 72
ตารางที่ 5-3 ผลการเปรียบเทียบประสิทธิภาพการรู้จำเสียงพูดจากกราฟของเซกเมนต์ โดยใช้ความรู้ของแนวคิด ..... 77
ตารางที่ 5-4 ค่าพรีซีชันและรีคอลในการค้นหาแนวคิดจากวิธีต่างๆ ..... 80
ตารางที่ 5-5 ความถูกต้องในการรู้จำคำ และการค้นหาแนวคิดในแต่ละแนวคิด เมื่อค้นหาแนวคิดจากลำดับของคำที่ได้จากการรู้จำ ..... 81
ตารางที่ 5-6 ความถูกต้องในการรู้จำคำ และการค้นหาแนวคิดในแต่ละแนวคิด เมื่อค้นหาแนวคิดจากลำดับจากกราฟของเซกเมนต์ ..... 81
ตารางที่ 5-7 ค่าพรีซีชันและรีคอลในการค้นหาแนวคิดจากผลลัพธ์จากการรู้จำคำของเครื่องรู้จำคำที่ดีที่สุด 10 อันดับแรก ..... 84
ตารางที่ 5-8 ค่าความถูกต้องในการระบุจุดมุ่งหมาย ..... 88
ตารางที่ 5-9 ค่าความถูกต้องในการระบุจุดมุ่งหมาย เมื่อใช้โครงข่ายประสาทเทียม และใช้กฎ ... 89
ตารางที่ 5-10 อัตราส่วนของข้อมูลทดสอบในประเภทจุดมุ่งหมายต่างๆ ..... 90
ตารางที่ 5-11 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดจากผลการรู้จำคำที่ดีที่สุด 10 ลำดับแรก ..... 90
ตารางที่ 5-12 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดจากผลการรู้จำคำที่ดีที่สุดลำดับแรก ..... 91
ตารางที่ 5-13 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดพี่พิจารณาพร้อมการรู้จำคำ ..... 91
ตารางที่ 5-14 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้กฎในการพิจารณา ..... 92
ตารางที่ 5-15 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นค่าคะแนนของแต่ละแนวคิด ..... 92

# สารบัญภาพ

หน้า

รูปที่ 2-1ส่วนประกอบของระบบการรู้จำเสียงพูดโดยทั่วไป [MIT]..... 7
รูปที่ 2-2 ลักษณะการทำงานของอัลกอริทึมวิเทอร์บี [YOUNG ET. AL., 2000]..... 10
รูปที่ 2-3 ตัวอย่างของตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนัก ..... 12
รูปที่ 2-4 ตัวอย่างการใช้ฟังก์ชันประกอบ (A), (B) ตัวอย่างของโครงข่ายสถานะ (C) ผลของการใช้ฟังก์ชันประกอบ..... 13
รูปที่ 2-5 ตัวอย่างไวยากรณ์ไม่พึ่งบริบท ..... 14
รูปที่ 2-6 ตัวอย่างของคลังคำศัพท์แบบที่ 1 [กิจศิริกุล, 2003]..... 14
รูปที่ 2-7 การบอกหมวดหมู่ของคำศัพท์โดยใช้ไวยากรณ์ [MANNING AND SCHUTZE, 1999] ..... 15
รูปที่ 2-8 ตัวอย่างการวิเคราะห์ทางวากยสัมพันธ์ ..... 15
รูปที่ 2-9 ข่ายงานเปลี่ยนสถานะ 1 ..... 17
รูปที่ 2-10 ข่ายงานเปลี่ยนสถานะ 2 ..... 18
รูปที่ 2-11 ตัวอย่างไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็น ..... 18
รูปที่ 2-12 ตัวอย่างต้นไม้แจกแจงส่วนโดยใช้ไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็นสำหรับประโยค“ผมอยากทราบอัตราค่าโทรศัพท์ของประเทศอเมริกา”..... 19
รูปที่ 2-13 ไวยากรณ์สำหรับแนวคิด “วันที่” (DATE) [GUPTA ET. AL., 2006]..... 20
รูปที่ 2-14 ตัวอย่างกฎในงานวิจัยของกลุ่มนักวิจัยของ WU [WU ET. AL., 2006] ..... 21
รูปที่ 3-1ขั้นตอนการทำงาน ..... 27
รูปที่ 3-2 การสร้างลำดับของเซกเมนต์สมมุติฐานจำนวน N แบบ ..... 29
รูปที่ 3-3 ผังงานการสร้างกราฟของเซกเมนต์จากเซตเมนต์สมมุติฐาน..... 29
รูปที่ 3-4 ตัวอย่างกราฟของเซกเมนต์ (A) ตัวอย่างสมมุติฐานของเซกเมนต์ที่เป็นไปได้ทั้งหมด (B) ตัวอย่างกราฟของเซกเมนต์..... 30
รูปที่ 3-5 การสกัดค่าลักษณะสำคัญ ..... 31
รูปที่ 3-6 ลักษณะการทำงานในการค้นหาลำดับของเซกเมนต์..... 32
รูปที่ 3-7 ตัวอย่างลำดับของเซกเมนต์ ..... 35
รูปที่ 3-8 ขั้นตอนการสร้างข้อมูลเพื่อใช้ในการเรียนรู้ค่าพารามิเตอร์การแจกแจงปกติของแต่ละเซกเมนต์..... 36
รูปที่ 3-9 ตัวอย่างการแบ่งเซกเมนต์ที่เกิดเกินขึ้นมา ..... 38

รูปที่ 3-10 ตัวอย่างการแบ่งเซกเมนต์ที่ผิดพลาดจากการที่แบ่งเซกเมนต์หายไป..... 38
รูปที่ 3-11ตัวอย่างแบบจำลองการออกเสียงของคำว่า “อเมริกา” ..... 40
รูปที่ 3-12 ตัวอย่างแบบจำลองภาษา..... 41
รูปที่ 3-13 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่าย *G* และ *L* (A) แบบจำลองภาษา *G* (B) โครงข่ายจับคู่ระหว่างคำศัพท์และหน่วยเสียง *L* (C) ผลของการทำฟังก์ชันประกอบของ *G* และ *L* ..... 43
รูปที่ 3-14 ตัวอย่างการคำนวณค่า $P(x_i|u_j)$ เมื่อเกิดเซกเมนต์ที่เกินขึ้นมา ..... 45
รูปที่ 3-15 ตัวอย่างการคำนวณค่า $P(x_i|u_j)$ เมื่อเกิดเซกเมนต์ที่เกินขึ้นมาในกรณีที่ 2 ..... 46
รูปที่ 3-16 ตัวอย่างการคำนวณค่า $P(x_i|u_j)$ เมื่อเกิดเซกเมนต์หน่วยเสียงหายไป..... 47
รูปที่ 3-17 ตัวอย่างการคำนวณค่า $P(x_i|u_j)$ เมื่อเกิดเซกเมนต์หายไปในกรณีที่ 2 ..... 48
รูปที่ 3-18 ลำดับการกระบวนการในการค้นหาลำดับของคำจากกราฟของเซกเมนต์..... 51
รูปที่ 3-19 ค้นหาลำดับของคำจากโครงข่ายหน่วยเสียง และโครงข่ายการออกเสียง ..... 52
รูปที่ 4-1 ตัวอย่างเครื่องจักรสถานะแบบถ่วงน้ำหนัก (REG) ของแนวคิด C_CODE..... 55
รูปที่ 4-2 ตัวอย่างเครื่องจักรสถานะแบบถ่วงน้ำหนัก (REG) ของแนวคิด C_CNT ..... 56
รูปที่ 4-3 การค้นหาลำดับแนวคิดเสนอโดย WUTIWIWATCHAI และ FURUI (WUTIWIWATCHAI AND FURUI, 2006) ..... 57
รูปที่ 4-4 ตัวอย่างเครื่องจักรสถานะของแนวคิด (A) เครื่องจักรสถานะของแนวคิด (L_CNT, C_CNT_NAME) (B) เครื่องจักรสถานะของแนวคิด (L_CODE, C_CODE)..... 60
รูปที่ 4-5 ตัวอย่างแบบจำลองแนวคิด..... 61
รูปที่ 4-6 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่าย *WC* และแบบจำลอง *CC* (A) ตัวอย่างโครงข่ายจับคู่คำศัพท์กับแนวคิด *WC* (B) ตัวอย่างแบบจำลองแนวคิด *CC* (C) ผลของการทำฟังก์ชันประกอบระหว่าง *WC* และ *CC*..... 63
รูปที่ 4-7 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่ายภาษา *G* และโครงข่าย $CC \circ WC$ (A) ตัวอย่างแบบจำลองภาษา *G* (B) ผลของการทำฟังก์ชันประกอบของ $CC \circ WC$..... 64
รูปที่ 4-8 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่าย *L* และโครงข่าย $CC \circ WC \circ G$ (A) ตัวอย่างโครงข่าย $CC \circ WC \circ G$ (B) ตัวอย่างของโครงข่ายจับคู่ระหว่างหน่วยเสียงกับคำ *L* (C) ผลของการทำฟังก์ชันประกอบของ $CC \circ WC \circ G \circ L$..... 66
รูปที่ 5-1ตัวอย่างระบบที่ใช้ในการทดลอง..... 68

รูปที่ 5-2ตัวอย่างแบบจำลองการออกเสียงของคำว่า “ราคา” ..... 74
รูปที่ 5-3 ตัวอย่างแบบจำลองภาษาที่ใช้ในการทดลอง เมื่อกำหนดให้มีคำศัพท์ในระบบ 4 คำ.... 75
รูปที่ 5-4 กราฟเปรียบเทียบอัตราส่วนของข้อผิดพลาดต่างๆ ที่เกิดขึ้นในการรู้จำคำ (A) เมื่อไม่ใช้ความรู้ของแนวคิดมาช่วยในการรู้จำคำ (B) เมื่อใช้ความรู้ของแนวคิดมาช่วย..... 78
รูปที่ 5-5 รูปแบบของโครงข่ายประสาทเทียมที่ใช้ในการทดลองเพื่อแบ่งประเภทจุดมุ่งหมาย ..... 85
รูปที่ 5-6 ขั้นตอนในการแบ่งประเภทเมื่อได้รับลำดับของแนวคิดมา..... 86
รูปที่ 5-7 ลักษณะการใช้กฎในการพิจารณาแบ่งประเภทจุดมุ่งหมาย ..... 87

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 1

# บทนำ

## 1.1 ที่มาและความสำคัญของปัญหา

การติดต่อสื่อสารระหว่างมนุษย์กับเครื่องคอมพิวเตอร์นั้นสามารถทำได้หลายวิธี โดยวิธีที่เป็นที่นิยมอีกวิธีหนึ่ง คือ การใช้เสียงเพื่อใช้ในการสั่งงาน ทั้งนี้เนื่องจากในการพูดเป็นวิธีการสื่อสารหนึ่งที่มนุษย์คุ้นเคย และง่ายต่อการสื่อสาร ดังจะเห็นได้จาก ในปัจจุบันได้มีการนำเอาเทคโนโลยีการรู้จำเสียงมาใช้ในระบบต่างๆ เพื่อใช้ให้มนุษย์ติดต่อสื่อสารกับเครื่องคอมพิวเตอร์ เช่น ระบบโอนสายโทรศัพท์ ระบบการบริการให้ข้อมูลต่างๆ เป็นต้น แต่เนื่องจากภาษาพูดของมนุษย์นั้นมีลักษณะของไวยากรณ์ของภาษาที่ไม่แน่นอน ทั้งนี้ขึ้นอยู่ความรู้และความเคยชินของผู้พูดแต่ละคน นอกจากนี้ยังมีปัญหาในเรื่องของการพูดซ้ำ หรือเสียงอื่นๆ ที่มิได้เป็นส่วนหนึ่งของข้อความที่ต้องการสื่อ

ในอดีตระบบที่มีการเอาเทคโนโลยีการรู้จำเสียงพูดมาใช้ในการติดต่อสื่อสารระหว่างมนุษย์กับเครื่องคอมพิวเตอร์นั้น ใช้การติดต่อในลักษณะที่ระบบจะตั้งคำถาม และให้ผู้ใช้งานตอบโดยใช้คำที่มีการกำหนดมาแล้ว ในลำดับของคำถามที่เครื่องคอมพิวเตอร์กำหนด หรือเรียกว่าเป็นระบบไอวีอาร์ (Interactive Voice Response-IVR) [iec] แต่ทั้งนี้ระบบดังกล่าวยังมีข้อจำกัดในการใช้งาน เนื่องจากผู้ใช้งานต้องพูดแต่เพียงคำที่ระบบกำหนดเท่านั้น ดังนั้นจึงได้มีการพัฒนาระบบเพื่อให้มนุษย์สามารถใช้ภาษาธรรมชาติ หรือ ภาษาพูดทั่วไป ในการติดต่อได้ รวมไปถึงระบบที่ระบบหรือผู้พูดเป็นผู้ริเริ่มหัวข้อสนทนาก็ได้ (mix-initiative system) [Zue et. al., 2000] [Fabbrizio and Lewis, 2004]

ทั้งนี้การที่จะให้เครื่องคอมพิวเตอร์รู้จักคำทั้งหมดที่ผู้พูดนั้น จะต้องให้ความรู้ของคำต่างๆ เหล่านั้นเป็นจำนวนมาก เพื่อดูความหมายและลักษณะการใช้งานของแต่ละคำ ซึ่งต้องเตรียมข้อมูลตัวอย่างเสียง และเลเบลตัวอย่างเสียงค่อนข้างเยอะ รวมทั้งยังมีปัญหาในการวิเคราะห์คำบางคำที่มีข้อมูลตัวอย่างเสียงน้อยอีกด้วย อีกทั้งในการเข้าใจความหมายจากภาษาพูดที่เป็นภาษาธรรมชาตินั้นไม่จำเป็นต้องเข้าใจทุกๆคำที่อยู่ในประโยคก็สามารถเข้าใจใจความหรือความหมายหลักๆของประโยค และสามารถตอบสนองความต้องการที่ผู้พูดต้องการสื่อได้ ดังนั้นจึงได้มีการนำเสนอการค้นหาแนวคิดจากเสียงพูดในการทำความเข้าใจภาษาพูดของมนุษย์ แทนการทำความเข้าใจจากคำศัพท์ ทั้งนี้วิธีนี้จะช่วยแก้ปัญหาดังกล่าว โดยได้มีการเสนอแนวทางต่างๆ เช่น การใช้กฎไวยากรณ์ทางภาษา [Gupta et. al., 2006] หรือการพิจารณาจากความน่าจะเป็น [Miller et. al., 1994] เพื่อค้นหาลำดับของแนวคิด ทั้งนี้ทั้ง 2 วิธีนี้จะมีข้อดีและข้อเสียที่แตกต่างกัน

การพิจารณาค้นหาแนวคิดโดยจากใช้กฎนั้น จะสามารถกำหนดกฎขึ้นมาได้เอง จึงเหมาะกับระบบที่ตัวอย่างเสียงน้อยๆ หรือระบบที่เริ่มต้นใช้งาน แต่ทั้งนี้จะมีข้อเสียคือ การใช้กฎเหล่านี้จะมีประสิทธิภาพเมื่อประโยคที่ได้รับมาตรงตามกฎที่กำหนดมา ในขณะที่การใช้ความน่าจะเป็นมาใช้ในการหาลำดับของแนวคิดนั้น จะสามารถใช้ได้แม้ว่าผู้พูดจะไม่ได้พูดตรงตามหลักไวยากรณ์ แต่ข้อเสียของวิธีนี้ก็คือ จะต้องใช้ข้อมูลจำนวนมากเพื่อใช้ในการเรียนรู้ ดังนั้น จึงได้มีนักวิจัยบางกลุ่มได้เสนอแนวทางในการความแนวคิดโดยใช้การใช้วิธีทั้ง 2 วิธีรวมกัน [Wutiwiwatchai and Furui, 2006] เพื่อรวมข้อดีและข้อเสียของวิธีดังกล่าว ทั้งนี้วิธีการที่ได้กล่าวมานี้ จะใช้ผลที่ได้จากเครื่องรู้จำมาพิจารณา

วิทยานิพนธ์นี้จึงมีวัตถุประสงค์เพื่อค้นหาแนวคิดจากเสียงพูด โดยการพิจารณานั้นจะทำในลักษณะของการค้นหาลำดับของแนวคิดจากกราฟของเซกเมนต์ซึ่งแสดงการแบ่งเซกเมนต์ของเสียงที่เป็นไปได้ตามลักษณะของหน่วยเสียง แทนการนำเอาลำดับของคำที่ได้จากการรู้เสียงพูดมาค้นหาลำดับของแนวคิด โดยใช้เหตุผลที่ว่าหากนำผลจากเครื่องรู้จำเสียงมาใช้นั้นจะได้ในลักษณะของลำดับของคำที่ดีที่สุดหรือดีที่สุดในสมมุติฐานจำนวน N ลำดับแรกมาใช้ในการค้นหาแนวคิด ลำดับของคำที่ได้เหล่านี้อาจจะไม่ถูกต้อง การค้นหาแนวคิดจากกราฟของเซกเมนต์นั้น จะเป็นการค้นหาจากลำดับของหน่วยเสียงที่เป็นไปได้ทั้งหมด ไม่ได้จำกัดแต่เพียงลำดับที่ดีที่สุดใน N ลำดับแรกเท่านั้น

### 1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอแนวทางในการค้นหาลำดับแนวคิดจากเสียงพูดในภาษาไทย
2. เพื่อนำเอาลำดับแนวคิดจากเสียงพูดที่ได้ ไปประยุกต์ใช้ในงานทำความเข้าใจเสียงพูดต่อไป เพื่อให้สามารถตอบสนองความต้องการของผู้ใช้งานได้อย่างถูกต้อง

### 1.3 ขอบเขตของการวิจัย

- ในขอบเขตของงานวิจัยนี้จะทำในส่วนของการค้นหาแนวคิดจากกราฟของเซกเมนต์ โดยที่
  - ข้อมูลขาเข้า
    - รับจากไฟล์เสียง จากนั้นจะทำการสร้างกราฟของเซกเมนต์ เพื่อทำการค้นหาลำดับของแนวคิดต่อไป
  - ผลลัพธ์ที่ได้จากการค้นหาแนวคิดนั้น จะประกอบไปด้วย
    - คู่ลำดับของแนวคิด และค่าของแนวคิดนั้นๆ
    - เป้าหมายหรือใจความที่สำคัญที่ผู้พูดต้องการในลักษณะของคำสั่ง
    - พิจารณาว่าประโยคที่พูดมานั้นอยู่ในโดเมนหรือไม่

- ในงานวิจัยนี้ไม่ได้รวมถึง
    - การหาแนวทางเพื่อปรับปรุงหรือพัฒนาประสิทธิภาพของการสร้างกราฟของเซกเมนต์จากเสียง
    - ส่วนของการจัดการบทสนทนา (Dialog Management)
    - การพัฒนาแนวทางในการค้นหาหรือแบ่งประเภทของเป้าหมายของประโยคจากลำดับแนวคิดที่ได้
- การวัดประสิทธิภาพของวิทยานิพนธ์
    - ในวิทยานิพนธ์นี้จะเปรียบเทียบประสิทธิภาพของการค้นหาแนวคิดจากเสียงพูดจากวิธีการนำเอาผลลัพธ์หรือลำดับของคำที่ดีที่สุดจากเครื่องรู้จำเสียงพูดมาค้นหาแนวคิด โดยจะเปรียบเทียบกับวิธีการในการใช้กฎร่วมกับการใช้ค่าความน่าจะเป็นในการค้นหาลำดับแนวคิดจากเสียงพูด ดังที่งานวิจัยของกลุ่มนักวิจัย Wutiwiwatchai [Wutiwiwatchai and Furui, 2006] ได้เสนอไว้
    - การประเมินผล จะพิจารณาจาก
        - ความถูกต้องของเป้าหมายหรือใจความที่ได้
        - ความถูกต้องของคู่ลำดับแนวคิดที่ค้นหาได้ ซึ่งจะพิจารณาคู่ลำดับแต่ละคู่จากแนวคิดที่เกิดขึ้นจริงในข้อมูลทดสอบ กับของแนวคิดที่สามารถทำการค้นหาได้ โดยไม่ได้พิจารณาเปรียบเทียบทั้งลำดับของแนวคิดที่ได้
- โดเมนในการค้นหาแนวคิด
    - กำหนดโดเมนในการค้นหาลำดับของแนวคิดจากโดเมนของการให้ข้อมูลโทรศัพท์ทางไกลระหว่างประเทศ โดยข้อมูลที่จะให้สอบถามนั้นประกอบด้วย
        - ข้อมูลอัตราค่าโทรศัพท์ทางไกลระหว่างประเทศในช่วงเวลาและระบบการบริการต่างๆ
        - รหัสโทรศัพท์ทางไกลระหว่างประเทศ

ซึ่งในระบบนี้จะประกอบไปด้วย ประเทศที่จะให้ข้อมูล จำนวน 231 ประเทศ และระบบบริการของอัตราค่าบริการ 2 ระบบ

- ข้อมูลที่ใช้
    - ข้อมูลที่ใช้ในการเรียนรู้
        - ข้อมูลเสียง ที่ทำการบันทึกผ่านการ์ดโทรศัพท์

- ข้อมูลตัวอักษร เพื่อศึกษาลักษณะของโครงสร้างภาษาจากภาษาเขียนจากกลุ่มตัวอย่าง โดยกำหนดให้แสดงลักษณะข้อความในลักษณะเหมือนภาษาพูด

o ข้อมูลที่ใช้ในการทดสอบ

- ข้อมูลเสียง โดยจะทำการบันทึกผ่านการ์ดโทรศัพท์ ซึ่งกลุ่มตัวอย่างเสียงทดสอบนั้นจะมาจากการสุ่ม โดยกำหนดให้กลุ่มตัวอย่างเสียงทำการพูดอย่างอิสระ ภายใต้โดเมนดังกล่าว

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

เสนอวิธีการค้นหาลำดับแนวคิดจากเสียงพูดจากกราฟของเซกเมนต์ภายใต้การใช้งานในภาษาไทย และทำการค้นหาเป้าหมายที่สำคัญจากแนวคิดที่ได้ ทั้งนี้เพื่อนำเอาเป้าหมายและลำดับของแนวคิดที่ได้ไปประยุกต์ใช้ในงานอื่นต่อไป

## 1.5 ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของงานวิทยานิพนธ์ได้รับการตีพิมพ์เป็นบทความวิชาการ ดังรายการต่อไปนี้

- หัวเรื่อง “Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling” โดย ศิรินาถ ตั้งรวมทรัพย์, โปรดปราน บุญยพุกกณะ และ อติวงศ์ สุชาโต ในงานประชุมวิชาการ “The 5th International Conference on Research, Innovation and Vision for the Future (RIVF'07)” ซึ่งจัดขึ้น ณ เมืองฮานอย ประเทศเวียดนาม ระหว่างวันที่ 5-9 มีนาคม 2550

- หัวเรื่อง “Keyword Verification using Linear Feature Combinations with MCE-trained Weights” โดย ศิรินาถ ตั้งรวมทรัพย์, โปรดปราน บุญยพุกกณะ และ อติวงศ์ สุชาโต ในงานประชุมวิชาการ “The Seventh International Symposium on Natural Language Processing (SNLP)” ซึ่งจัดขึ้น ณ เมืองพัทยา ประเทศไทย ระหว่างวันที่ 13-15 ธันวาคม 2550

- หัวเรื่อง “Phoneme Recognition from Thai Continuous Speech Using a Segment-Based Approach” โดย เกริกศักดิ์ ลิขิตสุภิณ, ศิรินาถ ตั้งรวมทรัพย์, โปรดปราน บุญยพุกกณะ และ อติวงศ์ สุชาโต ในงานประชุมวิชาการ “The 11th National Computer Science and Engineering Conference (NCSEC2007)” ซึ่งจัดขึ้น ณ กรุงเทพมหานคร ประเทศไทย ระหว่างวันที่ 19-21 พฤศจิกายน 2550

# บทที่ 2

## เอกสารและงานวิจัยที่เกี่ยวข้อง

### 2.1 แนวคิดและทฤษฎี

#### 2.1.1 การรู้จำเสียงพูด (Speech Recognition)

การรู้จำเสียงพูดนั้น เป็นการแปลงสัญญาณเสียงของคำ ให้อยู่ในรูปแบบที่คอมพิวเตอร์หรือเครื่องจักรสามารถเข้าใจได้ โดยที่สามารถพิจารณาสัญญาณเสียงได้ว่า เป็นลักษณะของลำดับของสัญลักาณ์ที่ถูกเข้ารหัสไว้ ดังนั้นในการรู้จำเสียงจะเป็นกระบวนการย้อนกลับ นั่นคือเป็นการค้นหาลำดับของสัญลักษณ์ที่ถูกเข้ารหัสนี้จากสัญญาณเสียงนั่นเอง

##### 2.1.1.1 ประเภทของการรู้จำเสียง

เราสามารถแบ่งประเภทของระบบการรู้จำเสียงพูดได้เป็นหลายประเภทตามลักษณะความสามารถในการรู้จำ โดยที่ในระบบการรู้จำเสียงใดๆ สามารถจัดให้อยู่ในหลายประเภทได้

1. จำนวนคำศัพท์ (Vocabulary)

ในแต่ละระบบการรู้จำนั้น จะมีจำนวนคำศัพท์ที่ใช้ในการรู้จำที่แตกต่างกัน เช่น ในระบบที่ทำการรู้จำตัวเลขโดดนั้น จะมีคำศัพท์จำนวน 10 ตัว หรือในระบบอาจจะมีจำนวนคำศัพท์มากจนถึง 20,000 คำ

2. วิธีการพูด (Speaking Mode)
   - การรู้จำเสียงพูดแบบคำโดดเดี่ยว (Isolated Word Recognition) เป็นการรู้จำเสียงโดยที่สัญญาณเสียงนั้นจะมีการเปล่งเสียง 1 ครั้ง และจะพยายามรู้จำคำเพียง 1 คำจากสัญญาณเสียงนั้น โดยจะค้นหาคำที่มีความใกล้คียงกับลักษณะของสัญญาณเสียงนั้นที่สุด
   - การรู้จำเสียงพูดแบบต่อเนื่อง (Continuous Speech Recognition) เป็นการรู้จำเสียง โดยที่สัญญาณเสียงนั้นจะมีการเปล่งเสียงได้หลายครั้ง โดยอาจจะเป็นในลักษณะของวลีหรือประโยค
   - การรู้จำเสียงแบบเกิดขึ้นเอง (Spontaneous Speech Recognition) เป็นการรู้จำโดยที่สัญญาณเสียงที่มีการเปล่งเสียงได้หลายครั้ง โดยที่วิธีการพูดในแบบอื่นนั้นสัญญาณเสียงจะประกอบแต่เพียงคำพูด ไม่ได้รวมถึงเสียงอื่นๆ เช่น เสียงหายใจ

โดยที่การรู้จำเสียงแบบเกิดขึ้นเองนี้ จะต้องสามารถรู้จำเสียงที่ประกอบด้วยเสียงประเภทอื่นๆ ได้ด้วย รวมทั้ง เสียง “เอ่อ” “อืม” “อ่า” ด้วย เป็นต้น

3. Speaker Generalization
   - ขึ้นกับผู้พูด (Speaker-Dependent) โดยลักษณะการรู้จำนี้ จะขึ้นอยู่กับผู้พูดด้วย ซึ่งหากเปลี่ยนผู้พูด ก็ต้องมีการเปลี่ยนแปลงแบบจำลองด้วย
   - ไม่ขึ้นกับผู้พูด (Speaker-Independent) เป็นการรู้จำแบบไม่ขึ้นกับผู้พูด นั่นคือ ระบบจะสามารถรู้จำเสียงได้ไม่ว่าเสียงที่เข้ามานั้นจะมาจากผู้พูดใดก็ตาม

### 2.1.1.2 **ส่วนประกอบของระบบการรู้จำเสียง**

ในการรู้จำเสียงนั้น จะประกอบไปด้วยหลายขั้นตอน โดยขั้นตอนหลักๆ ที่มีอยู่ในการรู้จำเสียงโดยทั่วไปนั้น จะประกอบด้วย

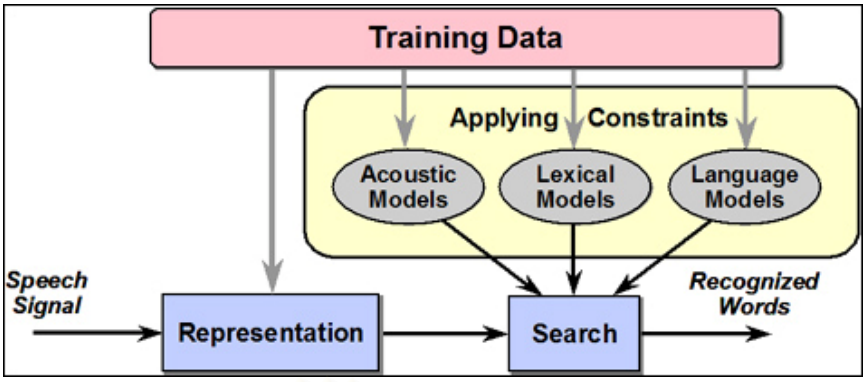
1. การสกัดค่าลักษณะสำคัญ (Speech Representation)

จะเป็นการดึงเอาค่าลักษณะสำคัญออกมาจากสัญญาณเสียงเพื่อใช้ในการรู้จำ โดยค่าลักษณะสำคัญของสัญญาณเสียงนั้นอาจจะอยู่ในรูปของลำดับของเวกเตอร์ ตัวอย่างลักษณะสำคัญมีอยู่หลายลักษณะ เช่น สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficients: LPC), สัมประสิทธิ์เซปสตรัม (Cepstral coefficient), สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepstral coefficients: MFCC), เซปสตรัมแบบหักลบค่าเฉลี่ย (Cepstral mean substraction: CMS) และเซปสตรัมแบบผ่านตัวกรองภายหลัง (Post filtered cepsturm: PFL) เป็นต้น

2. การจำแนก (Classification)

หลังจากที่ได้ลักษณะสำคัญของสัญญาณเสียงแล้ว จากนั้นจะทำการรู้จำเพื่อค้นหาผลลัพธ์ของสัญลักษณ์ โดยในขั้นตอนนี้สามารถทำได้หลายวิธี เช่น การใช้ฐานความรู้ (Knowledge-based Approach), การจับคู่แม่แบบ (Template Matching) หรือ การใช้สโทแคสติก (Stochastic Approach) การใช้โครงข่ายประสาทเทียม (Neural Network), แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model) เป็นต้น

ในระบบการรู้จำเสียงพูดนั้น โดยทั่วไปแล้วจะมีส่วนประกอบหลักๆ ดังรูปที่ 2-1




รูปที่ 2-1ส่วนประกอบของระบบการรู้จำเสียงพูดโดยทั่วไป [MIT]

จากรูปที่ 2-1 แสดงให้เห็นถึงส่วนประกอบของระบบการรู้จำโดยทั่วไป ซึ่งในตัวอย่างนี้จะเป็นการรู้จำในระดับของคำ โดยจะเห็นว่าในขั้นตอนเริ่มต้นนั้น เมื่อได้รับสัญญาณเสียงเข้ามา จะทำการสกัดค่าลักษณะสำคัญจากสัญญาณเสียง จากนั้นค้นหาลำดับของสัญลักษณ์ซึ่งอาจจะเป็นลำดับของคำที่เป็นไปได้มากที่สุด โดยจะใช้ความรู้จากแบบจำลองหน่วยเสียง (Acoustic Model), แบบจำลองคำศัพท์ (Lexical Model) และแบบจำลองภาษา (Language Model) โดยแบบจำลองทางเสียงนั้น จะเป็นแบบจำลองที่ทำการเรียนรู้ลักษณะของหน่วยเสียงต่างๆ ในเชิงสถิติ แบบจำลองการออกเสียงนั้นจะใช้ในการรู้จำในระดับของคำโดยจะบอกลำดับของหน่วยเสียงที่เป็นไปได้สำหรับแต่ละคำ และแบบจำลองภาษานั้น จะเป็นแบบจำลองที่บอกค่าทางสถิติของลำดับของคำที่จะสามารถเกิดขึ้นได้ โดยจะพิจารณาจากความถี่ของแต่ละลำดับของคำ ซึ่งแบบจำลองเหล่านี้จะช่วงในการค้นหาลำดับของคำที่เป็นไปได้มากที่สุด

### 2.1.2 ตัวแปรสุ่มแบบต่อเนื่อง

ในตัวแปรสุ่มบางตัวนั้น จะสามารถมีค่าที่เป็นไปได้ในโดเมนของค่าต่อเนื่องกัน ดังนั้นในการคำนวณหาค่าความน่าจะเป็นในการเกิดค่าคงที่ใดๆ ของตัวแปรสุ่มนั้น จะคำนวณค่าความน่าจะเป็นออกมาใกล้เคียงกับค่าศูนย์ ดังนั้นในการคำนวณหาค่าความน่าจะเป็นของตัวแปรสุ่มในลักษณะนี้ จะได้คำนวณค่าของตัวแปรนั้นในลักษณะของช่วง (interval) เช่น ในช่วงของค่า a ถึง b หรือ (a,b) เป็นต้น โดยค่าฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function) จะได้เป็น

$$P[x \in (a,b)] = \int_a^b p(x)dx$$

โดยที่หากเราพิจารณาในช่วงแคบๆ หรือในช่วงของ $(a,\ \ a+\Delta x)$ โดยที่ $\Delta x$ มีค่าน้อยมาก ค่าความหนาแน่นของความน่าจะเป็นที่ $x= a$ นั้น จะมีค่าเป็น $p(a)=P[\ x\ \varepsilon\ (a,\ a+\Delta x)]/\ \Delta x$ โดยที่ฟังก์ชันความหนาแน่นของความน่าจะเป็นนั้น จะต้องมีคุณสมบัติ คือ

$$p(x) \geq 0 \ \text{ และ } \int_{-\infty}^{\infty} p(x)dx = 1$$

โดยในการคำนวณหาค่าความคาดหมาย (Expected Value) ค่าเฉลี่ย (Mean) และค่าความแปรปรวน (Variance) นั้น จะคำนวณจาก

$$E[f(x)] = \int_{-\infty}^{\infty} f(x)\,p(x)dx$$

$$\mu = E[x] = \int_{-\infty}^{\infty} xp(x)dx$$

$$Var[x] = \sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2\,p(x)dx$$

สำหรับในกรณีที่ตัวแปร $x$ นั้นมีหลายมิติ (Dimension) เช่น มีจำนวน $d$ มิติ หรือ $\vec{x}$ เป็นเวกเตอร์สุ่มแบบต่อเนื่อง ฟังก์ชันความหนาแน่นของความน่าจะเป็น รวมถึงการคำนวณค่าความคาดหมาย ค่าเฉลี่ย และค่าความแปรปรวนยังคงคล้ายเดิม เพียงแต่ $d\boldsymbol{x}$ นั้นจะเป็นลักษณะ $d$ มิติ นั่นคือ $d\vec{x} = dx_1 dx_2 \ldots dx_d$ ดังนี้

$$\begin{aligned} E[f(\vec{x})] &= \int_{-\infty}\int_{-\infty}\ldots\int_{-\infty}^{\infty} f(\vec{x})\,p(\vec{x})dx_1 dx_2 \ldots dx_d \\ &= \int_{-\infty}^{\infty} f(\vec{x})\,p(\vec{x})d\vec{x} \\ \vec{\mu} &= E[\vec{x}] = \int_{-\infty}^{\infty} \vec{x}p(\vec{x})d\vec{x} \\ \vec{\sigma} &= E[(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^t] \\ &= \int_{-\infty}^{\infty} (\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^t\,p(\vec{x})d\vec{x} \end{aligned}$$

โดยที่ $\vec{\mu}$ จะเป็นเวกเตอร์ค่าเฉลี่ยขนาด $d$ มิติ $\vec{\sigma}$ จะเป็นเมทริกซ์ของความแปรปรวนขนาด $dxd$ มิติ หากค่าของ $x$ ในแต่ละมิตินั้นเป็นอิสระต่อกัน จะได้ว่าเมทริกซ์ความแปรปรวนนั้น จะเป็นลักษณะเมทริกซ์ทแยงมุม และฟังก์ชันความหนาแน่นของความน่าจะเป็นนั้น จะคำนวณได้ดังนี้

$$p(x) = \prod_{i=1}^{d} p_{x_i}(x_i)$$

สำหรับในกรณีของฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไขนั้น (Conditional Probability density function) จะสามารถแสดงได้เป็น

$$p(x \mid y) = \frac{p(x,y)}{p(y)}$$

โดยในกรณีนี้ เป็นการคำนวณหาค่าความหนาแน่นของ $x$ เมื่อเกิด $y$ โดยเมื่อใช้กฎของเบย์ (Bayes Rule) และฟังก์ชันความหนาแน่น จะได้ว่า

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{\int_{-\infty}^{\infty} p(y \mid x)p(x)dx}$$

**2.1.2.1 การแจกแจงปรกติ** (Normal Distribution)

การแจกแจงปรกติ หรือการแจกแจงแบบเกาส์เซียน (Gaussian Distribution) นั้น จะเป็นลักษณะหนึ่งของการแจกแจงค่าความน่าจะเป็นแบบต่อเนื่อง (Continuous Probability Distributions) โดยจะมีฟังก์ชันความหนาแน่นของความน่าจะเป็น เป็นดังนี้

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{(x-\mu)^2}{\sigma^2})}$$

การแจกแจงแบบปกตินั้น ลักษณะของความหนาแน่นจะสามารถแสดงได้อยู่ในรูปของกราฟระฆังคว่ำ (Bell-Shapsd Curve) นั่นคือ กราฟจะมีลักษณะสมมาตรโดยมีจุดยอดอยู่ที่ค่าเฉลี่ย สำหรับความกว้างของกราฟนั้น จึงขึ้นอยู่กับค่าความแปรปรวน จะเห็นได้ว่าลักษณะของการกระจายขึ้นจะพิจารณาจากค่าพารามิเตอร์เพียง 2 ค่า ได้แก่ ค่าเฉลี่ย ($\mu$) และค่าความแปรปรวน ($\sigma^2$)

การแจกแจงในลักษณะนี้ สามารถเขียนแทนได้เป็น $p(x) \sim N(\mu, \sigma^2)$ นั่นคือ ตัวแปร $x$ จะมีการแจกแจงปรกติ โดยมีค่าเฉลี่ยเป็น $\mu$ และค่าความแปรปรวนเป็น $\sigma^2$

สำหรับในกรณีที่ตัวแปร $x$ เป็นเวกเตอร์ขนาด $d$ มิติ ฟังก์ชันความหนาแน่นของความน่าจะเป็น จะสามารถแสดงได้ดังนี้

$$p(x) = \frac{1}{(2\pi)^{d/2} \mid \sum \mid^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})}$$

โดยที่ $\vec{\mu}$ จะเป็นเวกเตอร์ค่าเฉลี่ยขนาด $d$ มิติ $\Sigma$ จะเป็นเมทริกซ์ความแปรปรวนขนาด $dxd$ มิติ $\mid \Sigma \mid$ และ $\Sigma^{-1}$ จะเป็นค่าดีเทอร์มิแนนซ์ และตัวผกผันของเมทริกซ์ความแปรปรวนตามลำดับ

### 2.1.3 **อัลกอริทึมวิเทอร์บี** (Viterbi algorithm)

อัลกอริทึมวิเทอร์บีนี้ ใช้ในการค้นหาเส้นทางหรือลำดับของสถานะที่ดีที่สุด ซึ่งอาจพิจารณาจากค่าความน่าจะเป็นหรือความควรจะเป็น (likelihood) ที่มากที่สุด นั่นคือ

$$X^* = \arg\max_X P(X \mid O, \mu)$$

กำหนดให้ $X$ แทนลำดับของเส้นทาง $O$ แทน ค่าสังเกต และ $\mu$ แทนแบบจำลอง โดยอัลกอริทึมวิเทอร์บีจะมีลักษณะการทำงานเป็นแบบกำหนดการพลวัตร (Dynamic programming) โดยสามารถแสดงสมการได้ดังนี้

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

กำหนดให้ค่าสังเกต (Observation) $O = \{o_1, ...., o_p\}$ โดยที่ แทนค่าความน่าจะเป็นสูงสุดของเส้นทางที่เข้าสู่สถานะ $j$ และทำให้เกิดค่าสังเกตจาก $o_1$ ถึง $o_t$ ณ เวลา $t$ และ $a_{ij}$ แทนค่าความน่าจะเป็นในการเปลี่ยนสถานะจาก สถานะ $i$ ไปยังสถานะ $j$ สำหรับ $b_j(o_t)$ นั้นจะแทนค่าความน่าจะเป็นที่จะเกิด $o_t$ จากสถานะ $j$ ที่เวลา $t$ โดยกำหนดให้

$$\phi_1(1) = 1$$

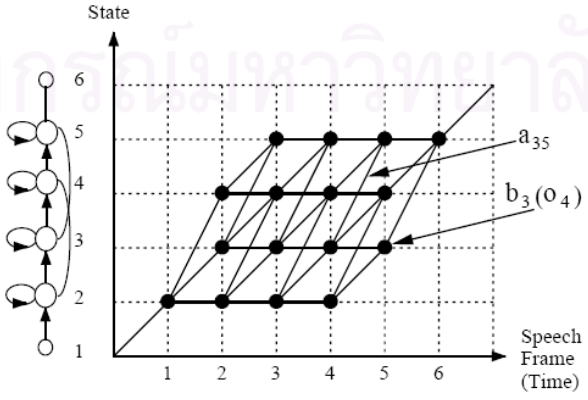
และ

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

เนื่องจากในการคำนวณค่าโดยใช้ผลคูณนั้น อาจทำให้เกิดปัญหาน้อยเกินเก็บ (Underflow) ได้ ดังนั้นจึงสามารถแสดงให้อยู่ในรูปของล็อคการิทึมได้ดังนี้

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

ในการทำงานของอัลกอริทึมวิเทอร์บีนั้น สามารถแสดงการทำงานโดยเปรียบเทียบกับรูปที่ 2-2




รูปที่ 2-2 ลักษณะการทำงานของอัลกอริทึมวิเทอร์บี [Young et. al., 2000]

จากรูปที่ 2-2 สัญลักษณ์จุดจะแทนค่าความน่าจะเป็นที่จะเกิดค่าสังเกต ณ เวลานั้น และเส้นเชื่อมระหว่างจุดจะแทนค่าความน่าจะเป็นในการเปลี่ยนสถานะระหว่างจุดที่เส้นเชื่อมนั้นเชื่อมอยู่ โดยที่ค่าความน่าจะเป็นของเส้นทางจะคือ ผลคูณหรือผลบวกของค่าล็อกการิทึมของค่าความน่าจะเป็นของผลลัพธ์กับค่าความน่าจะเป็นในการเปลี่ยนสถานะในเส้นทางนั้น

### 2.1.4 **ตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนัก** (Weight Finite State Transducer-WFST)

ตัวเปลี่ยนแปรสถานะจำกัดนั้น (ในที่นี้ของแทนด้วย WFST) จะทำหน้าที่ในการตรวจสอบอินพุตที่เข้ามาว่าสามารถยอมรับ (Accept) ได้หรือไม่ และให้เอาต์พุตออกมาในลักษณะของสายอักขระ (String) กำหนดให้ WFST $T=(\sum,\Omega,Q,E,i,F,\lambda,\rho)$ และ $\kappa$ เป็นเซมิริง (Semiring) โดยที่

$\sum$ แทน เซตของเลเบล หรือ ตัวอักษรที่เป็นอินพุต

$\Omega$ แทน เซตของเลเบล หรือ ตัวอักษรที่เป็นเอาต์พุต

$Q$ แทน เซตของสถานะโดยที่มีจำนวนจำกัด

$E$ แทน เซตของการส่งผ่าน (transition) ซึ่งมีจำนวนของสมาชิกในเซตเป็นจำนวนจำกัด โดยที่ $E \subseteq Q\times(\sum\cup\{\varepsilon\})\times(\Omega\cup\{\varepsilon\})\times\kappa\times Q$

$i$ แทน สถานะเริ่มต้น โดยที่ $i \in Q$

$F$ แทน เซตของสถานะสิ้นสุด โดยที่ $F \subseteq Q$

$\lambda$ แทน ค่าน้ำหนักเริ่มต้น

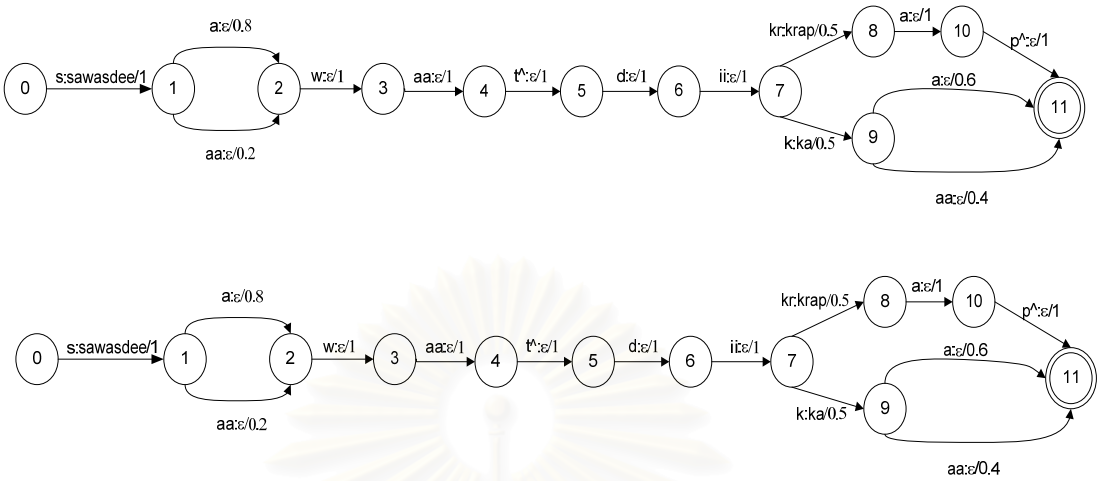
$\rho$ แทน ฟังก์ชันของค่าน้ำหนักสุดท้าย

สำหรับลำดับของสัญลักษณ์ $x$ ที่ถูกยอมรับ (accept) นั้น จะต้องมีเส้นทางที่สมบูรณ์ นั่นคือ มีเส้นทาง $\pi$ เริ่มตั้งแต่สถานะเริ่มต้น $i$ จนถึงสถานะปลายทาง $f$ โดยที่ $f \in F$ นอกจากนี้ยังมีเลเบล $\varepsilon$ ที่เป็นสัญลักษณ์ว่างเปล่า (empty symbol)

โดยสำหรับการส่งผ่าน $t$ ใดๆนั้น จะแทนด้วย $(p[t], l_i[t], l_o[t], w[t], n[t])$ โดยที่ $t \in E$ นั่นคือ $t$ จะใช้แทนเส้นเชื่อมที่เชื่อมจากสถานะ $p[t]$ ไปยังสถานะปลายทาง $n[t]$ โดยที่มีเลเบลของอินพุตเป็น $l_i[t]$ และเลเบลของเอาต์พุตเป็น $l_o[t]$ ซึ่งมีค่าน้ำหนักเป็น $w[t]$ โดยเลเบลของเอาต์พุตของเส้นทางใดๆนั้น จะเป็นเลเบลของเอาต์พุตของแต่ละเส้นเชื่อมในเส้นทางนั้นมาต่อกัน นั่นคือ

$$l[\pi] = l[t_1]....l[t_n]$$

ตัวอย่างของตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนักจะแสดงใน







รูปที่ 2-3 ตัวอย่างของตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนัก

ในรูปที่ 2-3 นั้น จะแสดงการใช้ตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนัก โดยใช้ในการแสดงของคำว่า “สวัสดีครับ” (sawasdeekrap) และ “สวัสดีค่ะ” (sawasdeeka) โดยที่สัญลักษณ์ที่อยู่บนเส้นเชื่อมนั้น จะแทน “*เลเบลของอินพุต : เลเบลของเอาต์พุต / ค่าน้ำหนัก*” เช่น ในสัญลักษณ์ “kr:krap /0.5” หมายถึง เมื่อมีเลเบลล์อินพุต *“kr”* จะให้เลเบลล์เอาต์พุตเป็น *“krap”* ซึ่งจะมีค่าน้ำหนักเป็น 0.5 โดยหากกำหนดให้น้ำหนักของเส้นเชื่อมเป็นค่าความน่าจะเป็นในการเปลี่ยนสถานะที่เชื่อมนั้นเชื่อมอยู่ ค่าความน่าจะเป็นของเส้นทางจะเป็นผลคูณของค่าน้ำหนักของแต่ละเส้นเชื่อมในเส้นทางนั้น

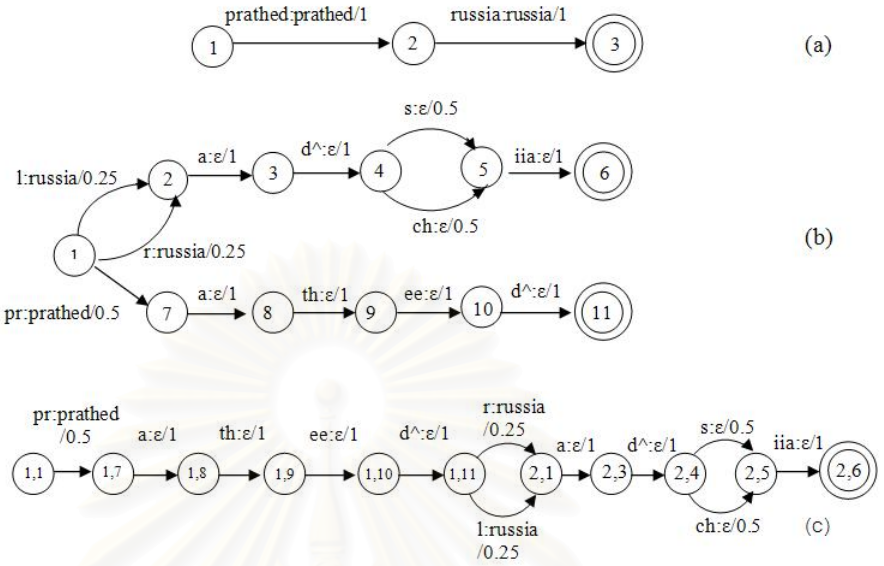
**2.1.4.1 ฟังก์ชันการประกอบ** (Composite Function)

ในการประกอบกันตัวเปลี่ยนแปรสถานะจำกัด 2 ตัวใดๆ หรือ $T = R \circ S$ ซึ่งเป็นการนำเอาตัวเปลี่ยนแปร 2 ตัว คือ $R$ และ $S$ มาประกอบกันและได้ตัวเปลี่ยนแปร $T$ ออกมา โดยตัวเปลี่ยนแปร $T$ ที่ได้จะจับคู่จากลำดับของ $u$ ยังไปลำดับของ $w$ โดยที่ตัวเปลี่ยนแปร $R$ จะจับคู่ระหว่างลำดับของ $u$ ไปยังลำดับของ $v$ และตัวเปลี่ยนแปร $S$ จะจับคู่ลำดับของ $v$ ไปยังลำดับของ $w$

ในการทำฟังก์ชันการประกอบ $R \circ S$ ของตัวเปลี่ยนแปร $R$ กับ $S$ นั้น สถานะของตัวเปลี่ยนแปรใหม่ที่ได้จะเป็นคู่ของสถานะจาก $R$ และสถานะจาก $S$ โดยที่มีเงื่อนไขว่า

1) สถานะเริ่มต้นจะเป็นคู่ของสถานะเริ่มต้นของ $R$ และ $S$
2) สถานะสิ้นสุดจะเป็นคู่ของสถานะสิ้นสุดของ $R$ และ $S$
3) จะมีการส่งผ่าน $t$ จาก $(r, s)$ ไปยัง $(r', s')$ สำหรับการส่งผ่านจาก $r$ ไป $r'$ ของ $t_r$ และจาก $s$ ไป $s'$ ของ $t_S$

โดยที่ตัวอย่างของการใช้ฟังก์ชันประกอบในการรวมตัวเปลี่ยนแปรสถานะจำกัด 2 ตัว สามารถแสดงได้ในรูปที่ 2-4




รูปที่ 2-4 ตัวอย่างการใช้ฟังก์ชันประกอบ (a), (b) ตัวอย่างของโครงข่ายสถานะ (c) ผลของการใช้ฟังก์ชันประกอบ

ในรูปที่ 2-4 จะแทนการประกอบตัวเปลี่ยนแปรสถานะจำกัด 2 ตัว เพื่อแสดงลักษณะลำดับของหน่วยเสียงสำหรับคำว่า “ประเทศรัสเซีย” โดยในรูปที่ 2-4(a) จะแสดงตัวเปลี่ยนแปรสถานะในระดับของคำ หรือ โมเดลภาษา (Language Model) ซึ่งจะแทนลำดับของคำที่เป็นไปได้ (ในที่นี้คือคำว่า “ประเทศ” และ “รัสเซีย” มาต่อกัน) ในรูปที่ 2-4(b) จะแทนลำดับของหน่วยเสียงที่ประกอบเป็นคำ โดยที่แต่ละคำจะไม่ขึ้นต่อกัน นั่นคือ ไม่กำหนดลำดับของคำ ผลของการประกอบตัวเปลี่ยนแปรสถานะจำกัดจะแสดงในรูปที่ 2-4(c) ซึ่งจะแสดงลำดับของหน่วยเสียงที่เป็นไปได้เมื่อกำหนดลำดับของคำตามรูปที่ 2-4(a)

### 2.1.5 **การวิเคราะห์ทางวากยสัมพันธ์** (Syntactic processing)

เป็นการวิเคราะห์ประโยคเพื่อดูความสัมพันธ์ของคำและโครงสร้างภายในประโยค รวมถึงการใช้ไวยากรณ์ทางภาษาในการพิจารณาว่าประโยคดังกล่าวดูต้องตามหลักของไวยากรณ์หรือไม่ ทั้งนี้การวิเคราะห์นี้จะช่วยตัดรูปแบบประโยคบางอย่างที่ไม่น่าจะเกิดขึ้น ทำให้ช่วยลดจำนวนของสมมุติฐานของประโยคก่อนจะช่วยไปประมวลผลในขั้นตอนถัดไป

ทั้งนี้การวิเคราะห์ทางวากยสัมพันธ์นั้น จะต้องใช้ส่วนประกอบดังนี้

- ไวยากรณ์ (grammar) (และคลังศัพท์ (lexicon))

- ตัวแจงส่วน (parser)

ในส่วนของไวยากรณ์นั้น จะแสดงในลักษณะของกฎของความสัมพันธ์ทางโครงสร้างของประโยค หรือกลุ่มคำที่เป็นไปได้ รวมทั้งบอกว่าประโยคดังกล่าวเป็นไปตามกฎหรือไม่ ตัวอย่างของไวยากรณ์ ดังเช่น

```
S-> NP VP
VP -> V NP
NP-> ART N
NP -> ART ADJ N
```

รูปที่ 2-5 ตัวอย่างไวยากรณ์ไม่พึ่งบริบท

จากตัวอย่างในรูปที่ 2-5 เป็นตัวอย่างของไวยากรณ์ในภาษาอังกฤษ จากกฎที่ 1 ได้ระบุว่าประโยค (แทนด้วย S) นั้นจะประกอบด้วยนามวลี (แทนด้วย NP) และกริยาวลี (แทนด้วย VP) ในกฎที่ 2 ได้บอกโครงสร้างของกริยาวลี ซึ่งจะประกอบไปด้วยคำกริยา (แทนด้วย V) และนามวลี สำหรับกฎที่ 3 และ 4 นั้น จะบอกส่วนประกอบของนามวลี ซึ่งจะประกอบไปด้วยคำนำหน้านาม (ตรงกับ article ในภาษาอังกฤษ ซึ่งเขียนแทนในกฎด้วย ART) และคำนาม (แทนด้วย N) โดยที่**ใน**กฎที่ 4 นั้น กำหนดให้สามารถมีคำขยายคำนาม (แทนด้วย ADJ) คั่นอยู่ระหว่างคำนำหน้านามและคำนามนั้นได้

ในส่วนของคลังศัพท์นั้น จะเป็นการระบุว่าคำศัพท์ต่างๆ จะอยู่ในประเภท หรือหมวดหมู่ใด เช่น เป็นคำนาม คำกริยา หรือเป็นบุพบท เป็นต้น ตัวอย่างของคลังศัพท์ เช่น

```
cried:   V
dogs:    N, V
the:     ART
old:     ADJ, N
man:     N, V
```

รูปที่ 2-6 ตัวอย่างของคลังคำศัพท์แบบที่ 1 [กิจศิริกุล, 2003]

นอกจากนี้ ยังสามารถระบุหมวดหมู่ของคำศัพท์ ในลักษณะของไวยากรณ์ ดังเช่น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

P -> *with*

V -> *saw*

NP -> *astronomers*

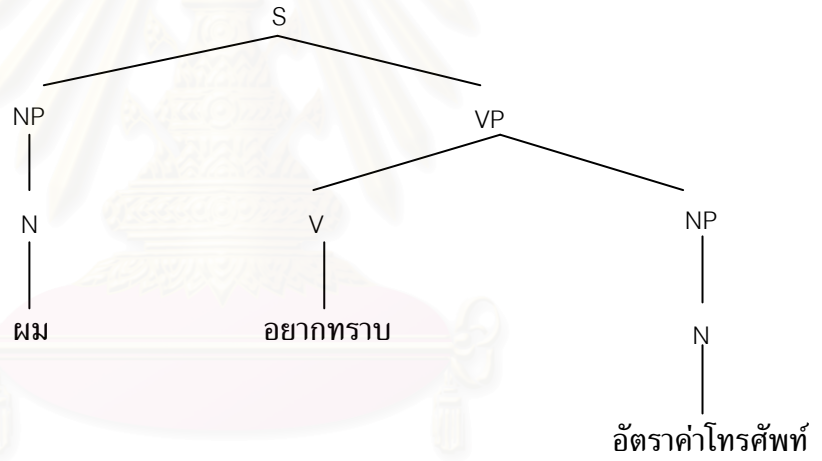
NP -> *ears*

NP -> *saw*

รูปที่ 2-7 การบอกหมวดหมู่ของคำศัพท์โดยใช้ไวยากรณ์ [Manning and Schutze, 1999]

### 2.1.5.1 **ตัวแจงส่วน**

สำหรับตัวแจงส่วนนั้น จะใช้สำหรับแจกแจงลำดับของคำในข้อความ เพื่อตรวจสอบว่าข้อความนั้นเป็นไปตามไวยากรณ์หรือไม่ รวมทั้งพิจารณาดูลักษณะของโครงสร้างของประโยคโดยใช้ไวยากรณ์และคลังคำศัพท์ เช่น เมื่อได้รับประโยคเข้ามาว่า “ผมอยากทราบอัตราค่าโทรศัพท์” จะสามารถแจงแจกในลักษณะของต้นไม้ได้ ดังรูปที่ 2-8




รูปที่ 2-8 ตัวอย่างการวิเคราะห์ทางวากยสัมพันธ์

ซึ่งการแจกแจงส่วนในลักษณะของต้นไม้นั้น จะแสดงให้เห็นถึงความสัมพันธ์ของบัพแม่และบัพลูก โดยในการแจกแจงส่วนนั้นมีวิธีการทำได้หลายแบบ โดยในที่นี้ได้เสนอการแจกแจงส่วนแบบบนลงล่าง การแจกแจงส่วนแบบล่างขึ้นบน การแจกแจงส่วนโดยใช้ไวยากรณ์ข่ายงานเปลี่ยนสถานะรวมถึงข่ายงานเปลี่ยนสถานะเรียกซ้ำ และการแจกแจงส่วนโดยใช้ไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็น

2.1.5.2 **ตัวแจงส่วนแบบบนลงล่าง** (top-down parser)

เป็นการสร้างต้นไม้แจงส่วนโดยสร้างจากส่วนบนของต้นไม้ลงมายังส่วนล่าง ซึ่งเมื่อพิจารณาจากไวยากรณ์ของภาษาในรูปที่ 1 แล้ว การแจงส่วนแบบบนลงล่างจะนำเอาด้านขวามือของกฎไปแทนที่สัญลักษณ์ด้านซ้ายมือของกฎไปเรื่อยๆจนถึงสัญลักษณ์ปลาย (Terminal symbol) ซึ่งก็คือสัญลักษณ์ที่เป็นอยู่ในบัพใบของต้นไม้นั่นเอง โดยในกรณีนี้อาจจะเป็นคำในประโยคที่ต้องการจะแจงส่วนนั่นเอง เช่น เมื่อมีประโยคตามรูปที่ 4 (“ผมอยากทราบอัตราค่าโทรศัพท์”) โดยมีกฎดังนี้

(1) S -> NP VP

(2) NP ->N

(3) VP-> V NP

(4) N->ผม | อัตราค่าโทรศัพท์

(5) V -> อยากทราบ

จากกฎดังกล่าว จะสามารถทำการแจงส่วนแบบบนลงล่างได้เป็น

| | | | | |
|---|---|---|---|---|
| S | -> | NP | VP | |
| | -> | N | VP | |
| | -> | N | V | NP |
| | -> | N | V | N |
| | -> | ผม | V | N |
| | -> | ผม | อยากทราบ | N |
| | -> | ผม | อยากทราบ | อัตราค่าโทรศัพท์ |

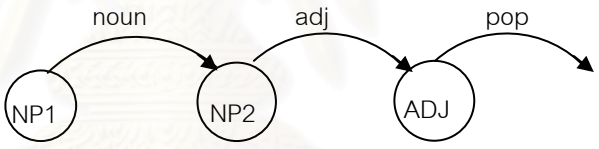
2.1.5.3 **ตัวแจงส่วนแบบล่างขึ้นบน** (bottom-up parser)

ในการแจงส่วนแบบล่างขึ้นบนนั้นจะตรงข้ามกับตัวแจงส่วนจากบนลงล่าง คือ การแจงส่วนนั้นจะเริ่มที่ส่วนล่างของต้นไม้ แล้วทำการแจงส่วนมายังส่วนบนของต้นไม้เรื่อยๆ หรือ การที่นำเอาคำในประโยคมาทำการแจงส่วน โดยเปรียบเทียบกับไวยากรณ์เพื่อรวมคำให้เป็นกลุ่มที่หน่วยใหญ่ขึ้นเรื่อยๆ โดยในการแจงส่วนนั้น จะทำการรวมด้านขวามือของกฎ โดยหากสามารถทำการแจกแจงได้ครบด้านขวามือของกฎแล้ว ก็จะทำการรวมเอาด้านขวามือของกฎนั้น ด้วยสัญลักษณ์ทางด้านซ้ายมือของกฎนั่นเอง โดยจากตัวอย่างข้างต้น สามารถแสดงการแจกแจงแบบล่างขึ้นบนได้เป็น

| | | |
|---|---|---|
| >ผม | อยากทราบ | อัตราค่าโทรศัพท์ |
| >N | อยากทราบ | อัตราค่าโทรศัพท์ |
| >N | V | อัตราค่าโทรศัพท์ |
| >N | V | N |
| >N | V | NP |
| >N | VP | |
| >S | | |

### 2.1.5.4 **ตัวแจกส่วนโดยใช้ไวยากรณ์ข่ายงานเปลี่ยนสถานะ**

ในการแสดงไวยากรณ์นั้นสามารถแสดงในรูปของข่ายงานซึ่งประกอบด้วยเส้นเชื่อม (arc) และบัพ (node) โดยเรียกไวยากรณ์ประเภทนี้ว่า ไวยากรณ์ข่ายงานเปลี่ยนสถานะ (Transition Network Grammar) โดยทำงานนั้นเริ่มจากสถานะเริ่มต้น จากนั้นจะรับอินพุตเข้ามาแล้วทำการเดินทางผ่านสถานะต่างๆ จนเมื่อเข้าสู่สถานะสุดท้ายก็จะทำการแจกแจงเสร็จ โดยพิจารณาตัวอย่างการแจกแจงโดยใช้ไวยากรณ์ข่ายงานเปลี่ยนสถานะได้จากรูปที่ 2-9




รูปที่ 2-9 ข่ายงานเปลี่ยนสถานะ 1

จากรูปที่ 2-9 แสดงข่ายงานของนามวลี (แทนด้วย NP) ในภาษาไทย ซึ่งจะเริ่มต้นการแจกแจงที่สถานะ NP1 โดยหากได้รับคำศัพท์ที่เป็นคำนาม (แทนด้วย noun) เข้ามา ก็จะสามารถผ่านยังสถานะอื่น คือ สถานะ NP2 หลังจากนั้นหากได้รับคำศัพท์ที่เป็นคำขยายคำนาม หรือ adj ก็จะย้ายไปที่สถานะ ADJ ต่อไป ซึ่งในสถานะ ADJ นั้นจะมีเส้นเชื่อม “pop” อยู่ นั่นคือ สามารถทำการแจกแจงได้สำเร็จนั่นเอง โดยจะสามารถแสดงในรูปของไวยากรณ์ไม่พึ่งบริบทได้เป็น

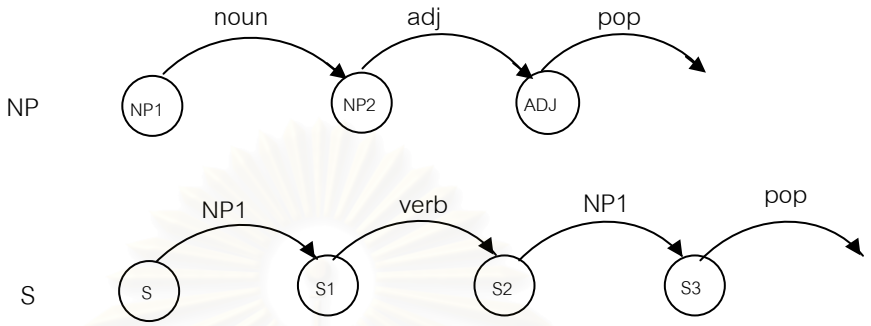
NP1 -> NP2 ADJ

NP2 -> noun

ADJ -> adj

แต่อย่างไรก็ตาม การใช้ไวยากรณ์ข่ายงานเปลี่ยนสถานะจะมีข้อจำกัดเมื่อนำมาใช้กับไวยากรณ์ที่ซับซ้อนมากขึ้น ดังนั้นจึงได้มีการเสนอ**ไวยากรณ์ข่ายงานเปลี่ยนสถานะเรียกซ้ำ**

**หรือ อาร์ทีเอ็น** (Recursive Transition Network) ซึ่งหลักการทำงานจะคล้ายกับข่ายงานเปลี่ยนสถานะ แต่สามารถอ้างอิงถึงข่ายงานอื่นได้ นั่นคือเมื่อพิจารณาจากตัวอย่างที่แล้ว นอกจากเส้นเชื่อมจะเป็นแบบหมวดคำ คือ เป็น noun, adj แล้ว เส้นเชื่อมยังสามารถอ้างอิงถึงข่ายงานอื่นได้ ดังตัวอย่างในรูปที่ 2-10




รูปที่ 2-10 ข่ายงานเปลี่ยนสถานะ 2

โดยในที่นี้ได้ใช้ตัวอักษรพิมพ์ใหญ่แทนเส้นเชื่อมที่อ้างอิงถึงข่ายงานอื่น และใช้ตัวพิมพ์เล็กแทนเส้นเชื่อมแบบเดิม หรือแบบที่เป็นหมวดคำ ซึ่งจากตัวอย่างในรูปที่ 2-10 นั้น จะมีข่ายงานย่อยอยู่ 2 ข่ายงาน คือ ข่ายงาน NP และข่ายงาน S ซึ่งการแจกแจงนั้นจะเริ่มจากสถานะ S ในข่ายงาน S โดยจะต้องรับกลุ่มคำที่ผ่านข่ายงาน NP1 เสียก่อนจึงจะสามารถข้ามไปยังสถานะ S1 ได้ นั่นคือจะต้องทำการแจกแจงในข่ายงานของ NP1 ให้สำเร็จก่อน จึงจะท่องไปยังสถานะ S1 ต่อไป

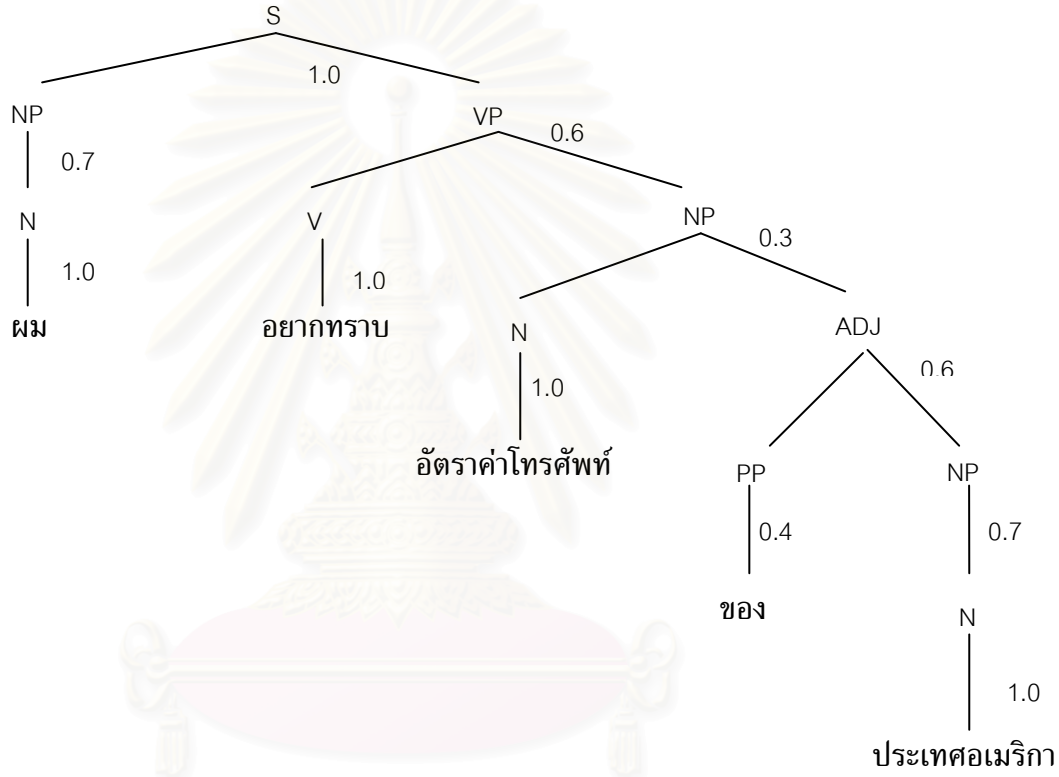
### 2.1.5.5 ตัวแจงส่วนโดยใช้ไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็น (Probability Context Free Grammar)

วิธีการนี้จะใช้ไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็น ซึ่งเป็นไวยากรณ์ที่ไม่พึ่งบริบทแต่จะมีค่าความน่าจะเป็นสำหรับแต่ละกฎ โดยวิธีการนี้จะใช้แบบจำลองทางความน่าจะเป็นในการสร้างต้นไม้สำหรับแจกแจงโครงสร้างทางวากยสัมพันธ์ ตัวอย่างของไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็นแสดงในรูปที่ 2-11

| | | | |
|---|---|---|---|
| S -> NP VP | 1.0 | ADJ-> P NP | 0.6 |
| NP-> NP ADJ | 0.3 | V -> *อยากทราบ* | 1 |
| NP -> N | 0.7 | PP -> *ของ* | 0.4 |
| VP -> V NP | 0.6 | N ->*อัตราโทรศัพท์\|ประเทศอเมริกา\|ผม* | 1 |

รูปที่ 2-11 ตัวอย่างไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็น

จากตัวอย่างที่แสดงในรูปที่ 2-11 กำหนดให้สัญลักษณ์ปลายเป็นคำศัพท์ในภาษาไทยซึ่งแสดงในลักษณะของตัวอักษรตัวเอียงและกำหนดให้สัญลักษณ์ S เป็นสัญลักษณ์เริ่มต้นในการแจกแจง สำหรับตัวเลขที่อยู่ด้านหลังกฎนั้น จะเป็นค่าความน่าจะเป็นในแต่ละกฎ นั่นคือ หากมีกฎ $N^i \rightarrow \xi_j$ ค่าความน่าจะเป็นของ $P(N^i \rightarrow \xi_j)$ จะคือค่า $P(N^i \rightarrow \xi_j \mid N^i)$ โดยจะกำหนดการกระจายของค่าความน่าจะเป็นของด้านขวามือของกฎเมื่อกำหนดด้านซ้ายมือของกฎมาให้ ตัวอย่างการแจกแจงส่วนของประโยค “ผมอยากทราบอัตราค่าโทรศัพท์ของประเทศอเมริกา” โดยใช้ไวยากรณ์ไม่พึ่งบริบทแบบอาศัยความน่าจะเป็นจะแสดงได้ 2 แบบ ดังรูปที่ 2-12




รูปที่ 2-12 ตัวอย่างต้นไม้แจกแจงส่วนโดยใช้ไวยากรณ์ไม่พึ่งบริบทแบบอาศัยค่าความน่าจะเป็นสำหรับประโยค“ผมอยากทราบอัตราค่าโทรศัพท์ของประเทศอเมริกา”

ในการคำนวณค่าความน่าจะเป็นของต้นไม้แจกแจงส่วนนั้น จะคำนวณโดยใช้ผลคูณของค่าความน่าจะเป็นในแต่ละกฎ โดยจากตัวอย่างที่แสดงในรูปที่ 2-12 นั้น จะสามารถคำนวณค่าความน่าจะเป็นของต้นไม้ ด้ดังนี้

$$P(t) = 1.0 \times 0.7 \times 1.0 \times 0.6 \times 1.0 \times 0.3 \times 1.0 \times 0.6 \times 0.4 \times 0.7 \times 1.0$$
$$= 0.021168$$

## 2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยเรื่องการค้นหาลำดับของแนวคิดจากเสียงพูด ได้มีกลุ่มนักวิจัยหลายกลุ่มทำการค้นคว้าและทดลองเพื่อหาแนววิธีการที่เหมาะสมเพื่อค้นหาลำดับของแนวคิด ซึ่งได้มีการนำเอาไปประยุกต์ใช้กับการทำความเข้าใจความหมายของภาษาพูด (Spoken Language Understanding) ทั้งนี้เราสามารถแยกการงานค้นหาแนวคิดจากเสียงพูดได้ออกเป็น 2 ส่วน ดังนี้

### 2.2.1 ส่วนที่ทำการค้นหาลำดับแนวคิดจากเสียงพูด

ในส่วนนี้ จะเป็นส่วนที่ค้นหาลำดับแนวคิดจากเสียงพูด โดยมีกลุ่มนักวิจัยต่างๆ ได้ทำการเสนอวิธีการในการค้นหาลำดับของแนวคิด ซึ่งในที่นี้ได้แบ่งวิธีการออกเป็น 3 ประเภท คือ การใช้กฎ การใช้ค่าความน่าจะเป็น และการใช้กฎและค่าความน่าจะเป็นร่วมกัน

#### 2.2.1.1 การใช้กฎในการค้นหาแนวคิดจากเสียงพูด

ได้มีกลุ่มนักวิจัยบางกลุ่มนำเสนอแนวทางในการค้นหาลำดับของแนวคิดโดยการใช้กฎมาพิจารณา เช่น กลุ่มนักวิจัยของ Gupta [Gupta et. al., 2006] ซึ่งได้แบ่งแนวคิดออกเป็น 2 ส่วน คือ ส่วนที่ไม่ขึ้นกับโดเมนในการทำงาน ซึ่งเป็นแนวคิดที่พบอยู่ทั่วไปในโดเมนต่างๆ (เช่น วัน หมายเลขโทรศัพท์ หมายเลขบัตรเครดิต เป็นต้น) และส่วนที่ขึ้นกับโดเมนในการทำงาน ซึ่งเป็นส่วนของแนวคิดที่มีลักษณะหรือความหมายเฉพาะ ขึ้นอยู่กับโดเมนในการใช้งาน (เช่น ชื่อผลิตภัณฑ์สินค้าและการบริการ เป็นต้น) ทั้งนี้ในส่วนของกฎสำหรับแนวคิดที่ไม่ขึ้นกับโดเมนในการทำงานนั้น จะสามารถนำเอากฎที่มีการใช้มาก่อนจาก คลังของกฎ (library) หรือดัดแปลงจากกฎที่มีอยู่เดิมได้ โดยจะแสดงไวยากรณ์ในรูปแบบแบกคัส-เนาร์ (Backus Naur Form -BNF) ดังแสดงในรูปที่ 2-13

```
DATE    →  ϵ:<month> MONTH ϵ:</month>
           ϵ:<day> DAY ϵ:</day>
MONTH   →  january:january | february:february | ...
DAY     →  first:first | second:second | ...
```

รูปที่ 2-13 ไวยากรณ์สำหรับแนวคิด “วันที่” (date) [Gupta et. al., 2006]

ในแต่ละกฎจะถูกแสดงในลักษณะของตัวเปลี่ยนแปรสถานะจำกัด โดยเมื่อได้ตัวเปลี่ยนแปรสถานะจำกัดสำหรับแต่ละกฎไวยากรณ์แล้ว ก็จะนำเอาตัวเปลี่ยนแปรที่ได้มารวมกัน (union) เพื่อให้ได้ตัวเปลี่ยนแปรสถานะจำกัดเพียงตัวเดียว โดยในการค้นหาแนวคิดนั้น จะนำเอาผลที่ได้จากเครื่องรู้จำเสียงซึ่งอาจจะเป็นลำดับของคำที่ดีที่สุดลำดับแรก หรือโครงข่ายคำ จากนั้นทำการแจกแจงเพื่อหาเส้นทางที่ยาวที่สุดจากตัวเปลี่ยนแปรสถานะจำกัดที่ได้จากกฎทางไวยากรณ์

อีกงานวิจัยหนึ่งที่ใช้กฎทางไวยากรณ์ในการพิจารณาหาแนวคิดจากเสียงพูด คือ งานวิจัยของกลุ่มของ Wu [Wu et. al., 2006] โดยจะทำการหาสล็อทที่เหมาะสมให้กับแนวคิด ซึ่งในงานวิจัยนี้จะแทนความสัมพันธ์ของแนวคิดในลักษณะของโครงสร้างลำดับชั้น (Hierarchical structure) โดยชื่อของสล็อทนั้นจะเป็นการนำเอาชื่อของแนวคิดจากเส้นทางของโครงสร้างลำดับชั้นเริ่มจากรากจนถึงแนวคิดที่เราสนใจมาต่อกัน เช่น [route].[origin].[location] [route].[transport_type] เป็นต้น เริ่มแรกนั้นจะมีการใช้กฎเพื่อหาแนวคิดเบื้องต้นจากประโยคคำพูด โดยจะนำเอาแนวคิดที่ได้แทนที่คำในประโยค จากนั้นจะมีการใช้ลักษณะเฉพาะ (feature) ซึ่งทำหน้าที่คล้ายกับกฎทางไวยากรณ์เพื่อทำการกำหนดสล็อทให้ในแต่ละแนวคิด โดยจะพิจารณาทั้งจากคำหรือแนวคิดที่อยู่รอบข้าง ตัวอย่างเช่น กฎสำหรับแนวคิดที่เป็นสถานที่ ([location]) ซึ่งสามารถเขียนกฎได้ดังรูปที่ 2-14

(1) *to* within the -3 windows

(2) *from_to*

(3) *ShowRoute.[route].[origin]* within the $\pm 2$ windows

รูปที่ 2-14 ตัวอย่างกฎในงานวิจัยของกลุ่มนักวิจัยของ Wu [Wu et. al., 2006]

โดยจากตัวอย่างที่แสดงในรูปที่ 2-14 ลักษณะเฉพาะที่ (1) และ (3) จะมีแนวโน้มที่เป็นการบอกว่าสถานที่นั้นเป็นสถานที่ปลายทาง และลักษณะเฉพาะที่ (2) จะบอกแนวโน้มว่าสถานที่นั้นเป็นสถานที่ต้นทาง ทั้งนี้จะเห็นว่าลักษณะเฉพาะดังกล่าวจะพิจารณาร่วมกับคำและแนวคิดที่ปรากฏอยู่รอบข้าง ซึ่งไม่จำเป็นต้องอยู่ติดกับคำหรือแนวคิดที่เราสนใจ (เช่น ในกฎข้อ (1) และ (3) ในรูปที่ 13) ซึ่งถือเป็นข้อดีอย่างหนึ่งในการเพิ่มความแม่นยำในการแบ่งประเภทของแนวคิดได้มากขึ้น โดยในลักษณะเฉพาะที่ (2) สามารถเขียนอยู่ในรูปของกฎได้เป็น [origin] -> from [location] to

ในการพิจารณาเลือกลักษณะเฉพาะมาใช้งานนั้น จะมีการกำหนดค่าคะแนนความมั่นใจในแต่ละลักษณะเฉพาะ จากนั้นจะทำการเรียงลำดับลักษณะเฉพาะตามลำดับคะแนนของแต่ละลักษณะเฉพาะจากมากไปน้อย เมื่อมีแนวคิดเข้ามาเพื่อทำการแบ่งประเภท ก็จะทำการค้นหาลักษณะเฉพาะที่ตรงกับแนวคิดนั้นตามลำดับของค่าความมั่นใจ เมื่อพบลักษณะเฉพาะแรกที่ตรงกัน ก็จะนำเอาลักษณะเฉพาะนั้นมาใช้ สำหรับค่าความมั่นใจของแต่ละลักษณะเฉพาะนั้น จะพิจารณาจากค่าความน่าจะเป็นที่จะเกิดสล็อทขึ้นเมื่อใช้ลักษณะเฉพาะนั้น ซึ่งพิจารณาจากตัวอย่างเสียงหรือข้อความที่ทำการเรียนรู้

ข้อดีของการใช้กฎทางไวยากรณ์ทางภาษาในการค้นหาแนวคิดจากเสียงพูดนั้น คือ สามารถสร้างกฎขึ้นมาได้โดยไม่จำเป็นต้องใช้ตัวอย่างข้อมูลจำนวนมาก ทั้งนี้ใช้เพียงความรู้ทางภาษาก็สามารถสร้างกฎได้ นอกจากนี้การใช้กฎยังสามารถพิจารณาจากบริบทที่อยู่ห่างกันเพื่อใช้หาแนวคิดที่เหมาะสมได้ แต่อย่างไรก็ตามข้อเสียของการตั้งกฎ คือ เมื่อพบรูปแบบประโยคที่ไม่ตรงตามกฎของไวยากรณ์ ซึ่งลักษณะของภาษาพูดก็มักจะมีโครงสร้างหรือรูปแบบของประโยคที่ไม่ค่อยแน่นอน ก็จะทำให้การค้นหาแนวคิดนั้นไม่มีประสิทธิภาพ

### 2.2.1.2 การใช้ความน่าจะเป็นในการพิจารณาหาแนวคิด

ในวิธีการนี้จะพิจารณาหาลำดับของแนวคิดจากค่าความน่าจะเป็น ซึ่งได้มีกลุ่มนักวิจัยบางกลุ่มได้ใช้เพื่อค้นหาแนวคิด เช่น ในงานวิจัยของกลุ่มนักวิจัยของ Miller [Miller et. al., 1994] ได้ทำการทดลองค้นหาแนวคิดจากตัวอย่างเสียงของ ATIS โดยใช้แบบจำลองความเข้าใจแบบซ่อน (Hidden Understanding Model) ซึ่งเมื่อกำหนดลำดับของคำ $W$ จะค้นหาความหมาย $M$ ที่ทำให้ $P(M|W)$ มีค่ามากที่สุด นั่นคือ

$$P(M \mid W) = \frac{P(W \mid M)P(M)}{P(W)}$$

ในงานวิจัยนี้ได้แบ่งแบบจำลองออกเป็น 2 ประเภท คือ แบบจำลองความหมายของภาษา (Semantic Language Models) และ แบบจำลองทางด้านคำศัพท์ (Lexical Realization Model) และได้มองแนวคิดเป็นลักษณะของแผนภูมิต้นไม้ โดยที่บัพต่อ (Nonterminal node) นั้นจะเป็นแนวคิดของความหมายในลักษณะที่เป็นนามธรรม (Abstract) สำหรับแบบจำลองความหมายของภาษาซึ่งใช้ในการดูความสัมพันธ์ของบัพต่อ จะประกอบไปด้วยข่ายงานของแต่ละแนวคิดโดยมีเส้นเชื่อมต่อหากันในระหว่างแนวคิด ซึ่งบัพในข่ายงานนั้นอาจใช้แทนข่ายงานอีกข่ายงานหนึ่งได้ โดยจะมีลักษณะเป็นข่ายงานที่มีการเรียกซ้ำ (Recursive) และมีค่าความน่าจะเป็นแต่ละเส้นเชื่อม ข่ายงานนี้จึงเปรียบเสมือนกับกฎทางไวยากรณ์ทางภาษา แต่จุดที่แตกต่างคือ โครงข่ายนี้จะกำหนดให้แต่ละแนวคิดเรียงลำดับอย่างไรก็ได้ โดยจะมีค่าความน่าจะเป็นบอกแนวโน้มของบัพหรือ แนวคิดถัดไป ในขณะที่การใช้กฎทางไวยากรณ์ จะกำหนดลำดับที่จะเกิดแนวคิดนั้นอย่างตายตัว

นอกจากนี้แบบจำลองด้านคำศัพท์ จะใช้ข่ายงานแสดงความสัมพันธ์ของบัพปลาย โดยกำหนดค่าความน่าจะเป็นของการเปลี่ยนบัพในระหว่างคำ ดังนั้น เมื่อเรามีแบบจำลองความหมายของภาษาเราก็สามารถหาค่าของ $P(M)$ ได้ และใช้แบบจำลองด้านคำศัพท์ในการพิจารณาค่า

$P(W|M)$ ดังนั้นเราสามารถหาค่าของ $P(W|M)P(M)$ ได้จากการรวมเอาแบบจำลองทั้ง 2 ประเภท จะทำการค้นหาเส้นทางที่ดีที่สุด

นอกจากนี้ยังมีกลุ่มนักวิจัย Maynard [Maynard and Lefevre, 2001] ที่ได้เสนอการใช้สโทแคสติกในการค้นหาลำดับของแนวคิดที่ต้องการ โดยจะรับลำดับของคำที่ได้มาจากเครื่องรู้จำเสียงพูด แล้วทำการค้นหาลำดับของแนวคิดที่ทำให้

$$\hat{C} = \arg\max_{C} P(C \mid W)$$
$$= \arg\max_{C} P(W \mid C)P(C)$$

กำหนดให้ $C$ เป็นลำดับของแนวคิด และ $W$ เป็นลำดับของคำที่ได้มาจากเครื่องรู้จำ โดยที่ค่าของ $P(W|C)$ และ $P(C)$ สามารถหาได้จาก

$$P(W \mid C) = P(w_1)\prod_{i=2}^{N} P(w_i \mid w_{i-1},...,w_{i-n+1},c_i)$$

และ

$$P(C) = P(c_1)\prod_{i=2}^{M} P(c_i \mid c_{i-1},...,c_{i-m})$$

การใช้ค่าความน่าจะเป็นในการพิจารณาค้นหาแนวคิดจากเสียงพูดนั้น จะมีข้อดีตรงที่มีความแม่นยำเมื่อเปรียบเทียบกับแบบการใช้กฎ โดยแต่ละแนวคิดสามารถแสดงได้อยู่ในลำดับที่ไม่จำกัดเหมือนการใช้กฎ ซึ่งลำดับของแนวคิดนั้นจะถูกควบคุมโดยค่าความน่าจะเป็นของการเกิดแต่ละแนวคิด สำหรับข้อเสียของการใช้วิธีนี้ คือ ต้องใช้ตัวอย่างเสียงจำนวนมากในการพิจารณาค่าความน่าจะเป็น ดังนั้นจึงอาจเป็นปัญหาสำหรับในกรณีที่มีข้อมูลเสียงน้อย

### 2.2.1.3 การใช้กฎทางไวยากรณ์ร่วมกับการใช้ค่าความน่าจะเป็น

เนื่องจากข้อดีและข้อเสียของวิธีการใช้กฎทางไวยากรณ์และการใช้ค่าความน่าจะเป็น ดังนั้นจึงได้มีนักวิจัยบางกลุ่มเสนอแนวทางในการรวมเอาแนววิธีทั้งสองเข้าด้วยกัน ดังในงานวิจัยของวิจัย Wutiwiwatchai และ Furui [Wutiwiwatchai and Furui, 2006] ซึ่งได้ทำการทดลองเปรียบเทียบการหาลำดับของแนวคิดโดยการใช้กฎทางไวยากรณ์ซึ่งแสดงด้วยตัวเปลี่ยนแปรสถานะจำกัดแบบถ่วงน้ำหนักในการแจกแจง โดยตัวเปลี่ยนแปรนี้จะแสดงลำดับของคำที่เป็นไปได้สำหรับแต่ละแนวคิด และจะมีค่าน้ำหนักในแต่ละเส้นเชื่อมระหว่างสถานะ

นอกจากนี้ในงานวิจัยนี้ ยังได้เปรียบเทียบกับการใช้แบบจำลองของค่าความน่าจะเป็น (Probabilistic models) โดยจะทำการค้นหาลำดับของเลเบลที่บอกว่าคำที่สนใจนั้นเป็นส่วนหนึ่งของแนวคิดใด ตัวอย่างของเลเบลเช่น กำหนดให้คำว่า “two” นั้นมีเลเบลเป็น “2:nn” นั่นคือ คำว่า

“two” ในประโยค จะเป็นส่วนหนึ่งของแนวคิดประเภทที่ 2 และมีการเลเบลของคำสำคัญเป็น “nn” โดยจะทำการค้นหาลำดับของเลเบลที่มีค่าความน่าจะเป็นสูงสุด ดังสมการ

$$\tilde{L} = \arg\max_{L} P(W, L)$$
$$\approx \arg\max_{L} \prod_{t=1}^{T} P(w_t, l_t \mid w_{t-1}, l_{t-1})$$

โดยที่ $W$ แทนลำดับของคำที่ได้จากเครื่องรู้จำ และ $L$ แทนลำดับของเลเบล โดยในกรณีนี้ได้พิจารณาในลักษณะของไบแกรม

ในการรวมเอาการใช้กฎทางไวยากรณ์ร่วมกับการใช้แบบจำลองของค่าความน่าจะเป็นนั้น ในงานวิจัยนี้ได้เสนอแนวคิดในการที่ใช้แบบจำลองของความน่าจะเป็นเพื่อหาลำดับของเลเบลที่มีค่าความน่าจะเป็นมากที่สุด นอกจากนั้นแล้วยังหาลำดับของเลเบลที่มีความน่าจะเป็นสูงสุดใน $m$ ลำดับแรก จากนั้นใช้คะแนนที่ได้จากแบบจำลองของกฎไวยากรณ์ทางภาษามาคิดร่วมกับคะแนนที่ได้จากสมมุติฐานของลำดับของเลเบลที่ดีที่สุด $m$ ลำดับที่ได้จากแบบจำลองของค่าความน่าจะเป็น เพื่อทำการเรียงลำดับ แล้วนำเอาคะแนนนั้นมาประยุกต์ใช้กับลำดับของเลเบลที่มีค่าความน่าจะเป็นสูงสุด

จากผลการทดลองพบว่าการใช้แบบจำลองของค่าความน่าจะเป็นจะให้ค่าการวัดแบบเอฟ (F-measure) มากกว่าการใช้แบบจำลองของกฎไวยากรณ์ แต่อย่างไรก็ตาม การใช้แบบจำลองค่าความน่าจะเป็นนั้นจะเกิดความสับสนกับบางคำที่สามารถเลเบลได้หลายแบบ การใช้แบบจำลองของกฎทางไวยากรณ์จะช่วยแก้ปัญหานี้ได้ ดังนั้นจึงได้นำเอาแบบจำลองของกฎทางไวยากรณ์มาใช้ร่วมกับแบบจำลองความน่าจะเป็น ซึ่งจะให้ความถูกต้องสูงสุด

นอกจากนั้นยังได้มีงานวิจัยของนักวิจัย Wang และ Acero [Wang and Acero, 2006] ได้เสนอแนวทางในการรวมการใช้กฎทางไวยากรณ์ร่วมกับการใช้ค่าทางสถิติ โดยการใช้บัพไม่ปลายในกฎไวยากรณ์มาเป็นลักษณะเฉพาะ (Feature) ในการแบ่งประเภทของโดเมนโดยใช้สถิติ จากนั้นเมื่อได้ประเภทของโดเมนแล้ว ก็ใช้กฎทางไวยากรณ์ในโดเมนนั้นในการแจกแจงเพื่อหาแนวคิดต่อไป

### 2.2.2 **ส่วนที่ทำการค้นหาจุดมุ่งหมายจากแนวคิดที่ได้**

ในงานส่วนนี้ จะเป็นการนำเอาแนวคิดที่ได้มาพิจารณาเพื่อดูจุดมุ่งหมายของประโยคนั้น เช่น ต้องการค้นหาข้อมูล ต้องการแก้ไขข้อมูล เป็นต้น ทั้งนี้ในงานวิจัยส่วนใหญ่จะใช้การแบ่งแยกแบบอย่าง (pattern classification) ในการพิจารณาค้นหาจุดมุ่งหมาย หรือการแบ่งประเภทว่าประโยคนั้นจัดอยู่ในจุดมุ่งหมายประเภทใด โดยได้มีกลุ่มนักวิจัยหลายๆกลุ่ม ได้ทำการทดลองและเปรียบเทียบ เพื่อหาแนวทางในการแบ่งกลุ่มที่เหมาะสม

ในงานวิจัยของ Wutiwiwatchai และ Furui [Wutiwiwatchai and Furui, 2006] นั้น ได้ใช้จุดมุ่งหมายในลักษณะของการกระทำของบทสนทนา (Dialogue act) โดยในการทดลอง ได้เปรียบเทียบตัวแบ่งประเภทต่างๆ ในการแบ่งประเภทของจุดมุ่งหมายของประโยคพูด โดยใช้โครงข่ายประสาทเทียม ซึ่งได้งานวิจัยนี้ได้ใช้โครงข่ายประสาทเทียมหลายชั้น ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และ Maximum a posteriori (MAP) โดยสำหรับการใช้ Support Vector Machine นั้น ในการทดลองนี้ได้ทำการทดลองโดยใช้ ฟังก์ชันเคอร์เนล (kernel) 3 แบบ ดังนี้

- แบบเชิงเส้น: $K(x, y) = x \cdot y$
- แบบพหุนาม: $K(x, y) = (x \cdot y + 1)^d$
- Radial basis function (RBF): $K(x, y) = \exp(-\gamma \mid x - y \mid^2)$

ซึ่งสำหรับการแบ่งประเภทโดยใช้โครงข่ายประสาทเทียมและซัพพอร์ตเวกเตอร์แมชชีนนั้น เวกเตอร์อินพุตจะเป็นเวกเตอร์ที่เป็นไบนารี่เพื่อบอกว่ามีแนวคิดใดที่เกิดขึ้นบ้าง โดยผลการทดลองพบว่า เมื่อใช้โครงข่ายประสาทเทียมและซัพพอร์ตเวกเตอร์แมชชีนนั้น จะให้ความถูกต้องในการแยกประเภทของเป้าหมายได้ใกล้เคียงกัน โดยเมื่อใช้โครงข่ายประสาทเทียมจะให้เปอร์เซ็นต์ความถูกต้องมากกว่าเล็กน้อย ในขณะที่ถ้าใช้ค่าหลังประสบการณ์มากสุด (Maximum a posteriori) ในการแบ่งประเภทจะให้เปอร์เซ็นต์ความถูกต้องต่ำสุด

ในกลุ่มนักวิจัยของ Wang [Wang and Acero, 2006] ได้เสนอแนวทางในการพิจารณาจุดมุ่งหมายโดยใช้ตัวแบ่งเชิงสถิติ (Statistic classification) ร่วมกับกฎไวยากรณ์ ซึ่งนำเอากฎไวยากรณ์มาร่วมใช้ในลักษณะของค่าลักษณะเฉพาะที่ใช้ในการแบ่งประเภท โดยแบ่งกลุ่มค่าลักษณะเฉพาะออกเป็น 3 ประเภท คือ แบบที่ใช้คำในการพิจารณาอย่างเดียว แบบที่ใช้คำและสัญลักษณ์ไม่ปลายที่ได้จากกฎไวยากรณ์ และแบบที่ใช้สัญลักษณ์ไม่ปลายเพียงอย่างเดียว และใช้ตัวแบ่ง 4 แบบ คือ ซัพพอร์ตเวกเตอร์แมชชีน, ทฤษฎีของเบย์ (Naïve Bayes), ยูนิแกรม (unigram) และ ไบแกรม (bigram) โดยผลการทดลองแสดงให้เห็นว่า เมื่อใช้สัญลักษณ์ไม่ปลายมาใช้แทนคำนั้น จะให้ค่าผิดพลาดน้อยที่สุดเมื่อใช้ตัวแบ่ง ซัพพอร์ตเวกเตอร์แมชชีน, ทฤษฎีของเบย์ (Naïve Bayes) และ ไบแกรม ในขณะที่เมื่อใช้ยูนิแกรมจะให้ค่าความผิดพลาดต่างกันไม่มากนักเมื่อใช้ลักษณะเฉพาะที่แตกต่างกัน

นอกจากนี้ในกลุ่มนักวิจัยของ Wang นั้น ยังได้ทำใช้ตัวแบ่งหลายตัวในการแบ่งประเภทของเป้าหมาย โดยให้วิธีการโหวต ซึ่งในงานวิจัยนี้ได้ใช้ตัวแบ่ง 3 ประเภท คือ ซัพพอร์ตเวกเตอร์แมชชีน, ทฤษฎีของเบย์ (Naïve Bayes) และ แบบจำลองเอ็นแกรม (n-gram models) โดยหากตัวแบ่งทั้ง 3 ประเภทนั้น ต่างให้ผลการแบ่งกลุ่มเป้าหมายที่แตกต่างกัน ก็จะใช้คำตอบของซัพ

พอร์ตเวกเตอร์แมชชีน ซึ่งให้ผลการแบ่งประเภทที่ดีที่สุดในการแบ่งประเภทเป้าหมาย ซึ่งจากผลการทดลองเมื่อเปรียบเทียบการใช้ตัวแบ่งหลายตัวกับการใช้ซัพพอร์ตเวกเตอร์แมชชีนเพียงอย่างเดียวในการแบ่งกลุ่มนั้น ความถูกต้องของการแบ่งทั้ง 2 ประเภทจะไม่แตกต่างกันมาก ทั้งนี้ได้ให้เหตุผลว่า เนื่องจากซัพพอร์ตเวกเตอร์แมชชีนนั้นจะค่าความผิดพลาดที่ค่อนข้างต่ำอยู่แล้วเมื่อเทียบกับตัวแบ่งประเภทอื่น

นอกจากการใช้ซัพพอร์ตเวกเตอร์แมชชีน ข่ายงานประสาทเทียม หรือ Naïve Bayes แล้ว กลุ่มนักวิจัย Gupta [Gupta et. al., 2006] ได้ใช้ อัลกอริทึมบูสติง (Boosting) โดยมีการปรับเปลี่ยนฟังก์ชันจุดประสงค์เพื่อนำเอาความรู้ก่อน (prior knowledge) มาใช้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

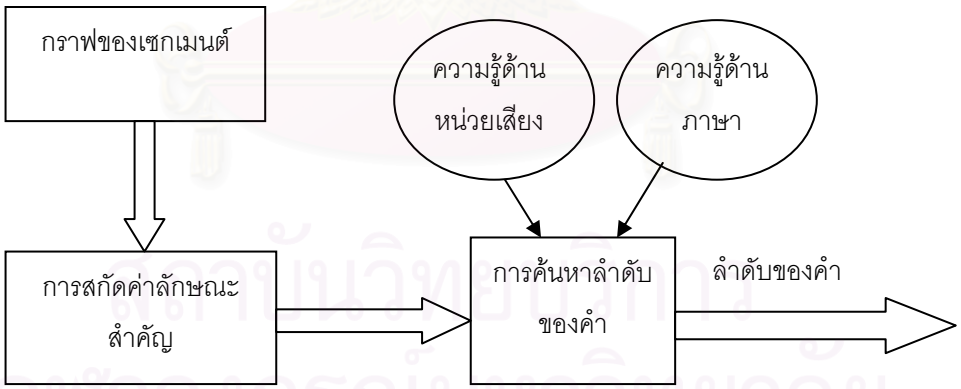
# บทที่ 3

## การรู้จำเสียงแบบเซกเมนต์

ในบทนี้จะกล่าวถึงกระบวนการในการรู้จำเสียงในระดับของคำจากกราฟของเซกเมนต์ โดยจะกล่าวถึงขั้นตอนในการสร้างกราฟของเซกเมนต์ที่ใช้ในวิทยานิพนธ์นี้ การสกัดค่าลักษณะสำคัญของแต่ละเซกเมนต์ กระบวนการในการค้นหา รวมถึงการสร้างและใช้งานแบบจำลองต่างๆ เพื่อใช้ในการพิจารณาลำดับของคำที่ดีที่สุดจากกราฟของเซกเมนต์นั้น รวมถึงแสดงการทดลองในการใช้กราฟของเซกเมนต์ในการรู้จำคำ

### 3.1 การออกแบบระบบการรู้จำเสียงแบบเซกเมนต์

ในการออกแบบระบบการรู้จำเสียงแบบเซกเมนต์นั้น เริ่มแรกจะรับสัญญาณเสียงเข้ามายังในระบบ จากนั้นจะทำการสร้างกราฟของเซกเมนต์ซึ่งเป็นกราฟที่แสดงขอบเขตเซกเมนต์ของเสียง โดยในที่นี้แต่ละเซกเมนต์นั้นจะแทนแต่ละหน่วยเสียง จากนั้นจะทำการค้นหาลำดับของคำที่มีความน่าจะเป็นมากที่สุดจากกราฟของเซกเมนต์นั้น โดยในการค้นหานั้น จะใช้ความรู้ในเรื่องของหน่วยเสียง และความรู้ทางด้านภาษาและคำศัพท์ในการพิจารณาค้นหาลำดับของคำ โดยขั้นตอนของการทำงานจะแสดงใน รูปที่ 3-1




รูปที่ 3-1ขั้นตอนการทำงาน

### 3.2 การสร้างกราฟของเซกเมนต์

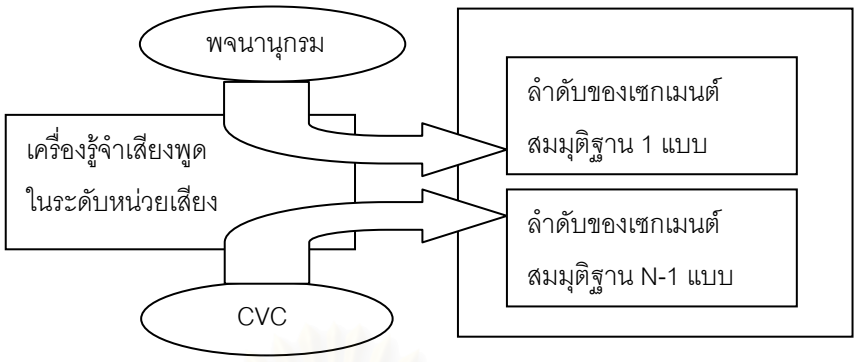
เนื่องจากเมื่อได้รับสัญญาณเสียงเข้ามานั้น เริ่มแรกนั้น เราจะทำการแบ่งขอบเขตของเสียงออกเป็นหน่วยเสียงต่างๆ ในวิทยานิพนธ์นี้จะแบ่งออกเป็นลักษณะของหน่วยเสียงภาษาไทย

จำนวน 74 หน่วยเสียง ดังแสดงในภาคผนวกที่ ก. ถึง ค. ซึ่งจะแสดงลักษณะขอบเขตของหน่วยเสียงที่เป็นไปได้ในลักษณะของกราฟ โดยที่ในแต่ละเส้นทางจากจุดเริ่มต้นจนถึงจุดสุดท้ายของกราฟนั้น จะมีลำดับของเซกเมนต์ หรือหน่วยเสียงที่แตกต่างกัน การแบ่งเซกเมนต์ออกเป็นหน่วยเสียงนี้ จะแตกต่างจากการแบ่งเซกเมนต์ออกตามช่วงเวลาที่เท่าๆ กัน (โดยทั่วไปจะแบ่งที่ประมาณ 10 มิลลิวินาที) โดยการแบ่งเซกเมนต์ตามหน่วยเสียงนั้น จะสามารถพิจารณาลักษณะสำคัญที่แตกต่างกันในแต่ละหน่วยเสียงได้ เช่น ความยาว เป็นต้น

ในวิทยานิพนธ์นี้ได้สร้างกราฟของเซกเมนต์ โดยการแบ่งสัญญาณเสียงออกเป็นเซกเมนต์จากเครื่องรู้จำแบบอาศัยกรอบเวลา โดยจะให้เครื่องรู้จำนี้ทำการรู้จำในระดับของหน่วยเสียง (Phoneme Recognition) ซึ่งจะพิจารณารู้จำลำดับของหน่วยเสียงที่ดีที่สุดจำนวน N แบบ จากนั้นจึงนำเอาผลที่ได้ทั้ง N แบบนั้นมาสร้างกราฟ โดยที่การรู้จำหน่วยเสียงทั้ง N แบบนั้น จะประกอบไปด้วย

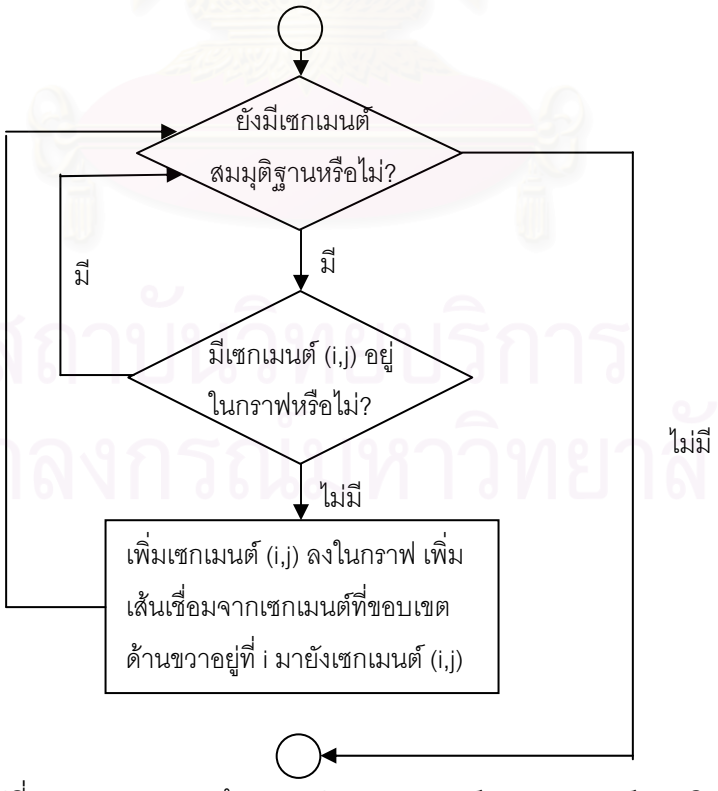
- การสร้างสมมุติฐานของเซกเมนต์โดยใช้คำศัพท์ในพจนานุกรม (Dictionary) เนื่องจากในงานวิจัยของ Glass [Glass, 2002] ที่กล่าวว่าการสร้างกราฟของเซกเมนต์นั้น จะมีประสิทธิภาพมากขึ้นเมื่อมีความรู้ทางด้านภาษาเข้ามาช่วย ดังนั้นจึงได้ใช้รายการคำศัพท์ที่มีในพจนานุกรมมาช่วยในการพิจารณาสร้างสมมุติฐานของเซกเมนต์
- การสร้างสมุติฐานของเซกเมนต์ด้วยการกำหนดลำดับของหน่วยเสียงในลักษณะของซีวีซี (CVC) เนื่องจากว่า การใช้เฉพาะรายการคำศัพท์ในการสร้างสมมุติฐานของเซกเมนต์นั้น จะเกิดข้อผิดพลาดหากในสัญญาณเสียงนั้นมีคำบางคำที่ไม่อยู่ในพจนานุกรม รวมทั้งหากการรู้จำคำนั้นผิดพลาด ก็จะส่งผลให้สมมุติฐานของเซกเมนต์นั้นผิดพลาดไปด้วย ดังนั้น จึงได้สร้างสมมุติฐานของเซกเมนต์อื่นเข้ามาร่วมกัน ซึ่งเราจะแบ่งหน่วยเสียงในภาษาไทยออกเป็น 3 ประเภท ได้แก่ กลุ่มของพยัญชนะต้น กลุ่มของสระ และกลุ่มของตัวสะกด โดยในสร้างสมมุติฐานของเซกเมนต์นั้น ได้กำหนดให้ลำดับของเซกเมนต์ที่เป็นไปได้จะประกอบด้วย พยัญชนะต้น-สระ-ตัวสะกด หรือ พยัญชนะต้น-สระ

โดยในการสร้างกราฟของเซกเมนต์นั้น จะพิจารณาจากการสร้างลำดับของเซกเมนต์สมมุติฐานจากผลที่ได้จากการรู้จำในระดับของหน่วยเสียงจำนวน N แบบ โดยพิจารณาจากผลการรู้จำหน่วยเสียงเมื่อพิจารณาคำศัพท์จากในพจนานุกรมจำนวน 1 แบบ รวมกับลำดับของสมมุติฐานของเซกเมนต์ที่ได้จากการการรู้จำในระดับของหน่วยเสียงเมื่อกำหนดลำดับของหน่วยเสียงในลักษณะของซีวีซีจำนวน N-1 แบบ โดยลักษณะการทำงานจะสามารถแสดงได้ในรูปที่ 3-2



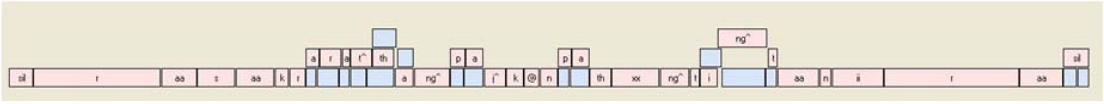

รูปที่ 3-2 การสร้างลำดับของเซกเมนต์สมมุติฐานจำนวน N แบบ

ในการสร้างกราฟของเซกเมนต์นั้น จะเริ่มจากการที่สร้างเซตของเซกเมนต์สมมุติฐานทั้งหมดจากการสร้างลำดับของเซกเมนต์จำนวน N แบบ จากนั้นจะดึงเอาเซกเมนต์แต่ละตัว โดยเซกเมนต์ (i,j) หมายถึง เซกเมนต์ที่เริ่มต้นที่เวลา i และสิ้นสุดที่เวลา j หรือ มีขอบเขตซ้ายเป็น i และขอบเขตขวาเป็น j โดยเซกเมนต์ใดๆ ที่มีขอบเขตขวาและขอบเขตซ้ายเหมือนกันนั้น จะถือว่าเป็นเซกเมนต์เดียวกัน หากยังไม่มีเซกเมนต์ (i,j) นั้นในกราฟ ก็จะทำการเพิ่มบัพของเซกเมนต์นั้นลงในกราฟ และสร้างเส้นเชื่อมจากเซกเมนต์อื่นๆ ที่มีขอบเขตขวา i ให้เชื่อมมายังเซกเมนต์ (i,j) การทำงานจะทำไปเรื่อยๆ จนหมดเซกเมนต์สมมุติฐานที่ได้ โดยการทำงานนั้นจะแสดงได้ในรูปที่ 3-3

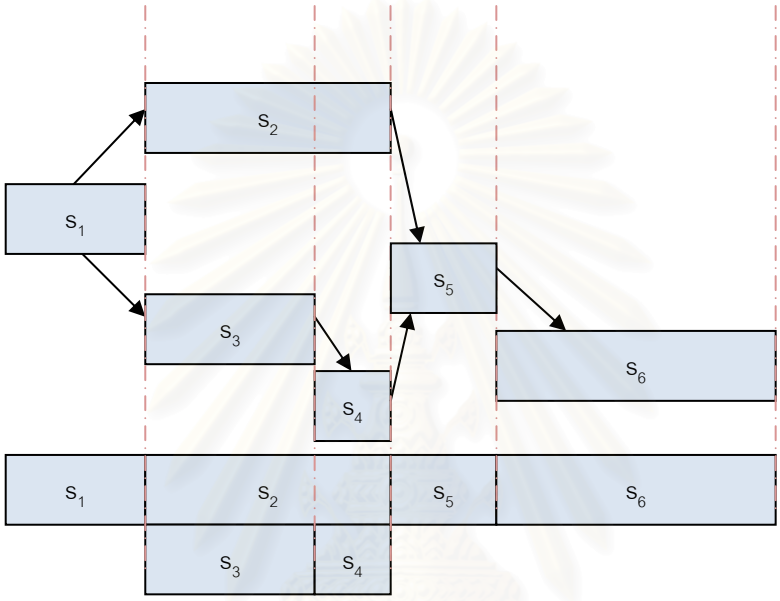



รูปที่ 3-3 ผังงานการสร้างกราฟของเซกเมนต์จากเซตเมนต์สมมุติฐาน

ตัวอย่างผลลัพธ์จากการสร้างกราฟของเซกเมนต์นั้น จะแสดงในรูปที่ 3-4



(a).




(b).

รูปที่ 3-4 ตัวอย่างกราฟของเซกเมนต์ (a) ตัวอย่างสมมุติฐานของเซกเมนต์ที่เป็นไปได้ทั้งหมด (b) ตัวอย่างกราฟของเซกเมนต์

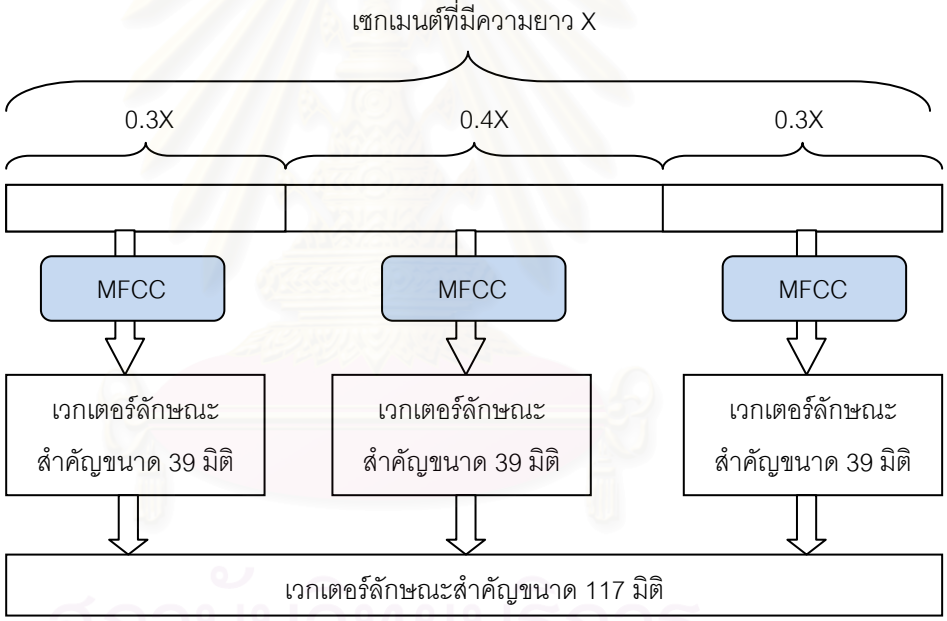
ในรูปที่ 3-4 นั้น จะแสดงตัวอย่างของกราฟของเซกเมนต์ โดยในรูปที่ 3-4(a) นั้น จะเป็นตัวอย่างของลำดับของเซกเมนต์สมมุติฐานที่เป็นไปได้ทั้งหมด และใน รูปที่ 3-4(b) นั้น จะเป็นตัวอย่างกราฟของเซกเมนต์ที่ได้จากกลุ่มของเซกเมนต์สมมุติฐาน เส้นทางของเซกเมนต์จากต้นทางไปยังปลายทางของกราฟนั้น จะเรียกว่า ลำดับของเซกเมนต์ (Segmentation) โดยที่ในกราฟนั้นจะมีลำดับของเซกเมนต์ที่แตกต่างกันหลายๆ ลำดับ

### 3.3 การสกัดค่าลักษณะสำคัญ

สัญญาณเสียงที่ได้นั้น ในเริ่มแรกจะอยู่ในลักษณะของสัญญาณดิจิตอล โดยในการนำเอาสัญญาณเสียงนั้นมาใช้ในการรู้จำ จะทำการสกัดค่าลักษณะสำคัญของเสียงนั้นเก็บอยู่ในรูปของเวกเตอร์ของค่าพารามิเตอร์ เพื่อที่จะนำมาพิจารณาเพื่อการรู้จำต่อไป

ในการสกัดค่าสำคัญในงานวิจัยนี้ จะพิจารณาแบ่งเซกเมนต์ออกเป็น 3 ช่วง ได้แก่ ช่วงต้น ช่วงกลาง และช่วงปลาย โดยหากเสียงที่รับเข้ามานั้นมีความยาว x วินาที ช่วงต้นนั้น จะเป็นช่วงของสัญญาณเสียงตั้งแต่เริ่มต้นจนถึง 0.3x วินาที ช่วงกลางจะเป็นช่วงของสัญญาณเสียงจาก 0.3x วินาที จนถึง 0.7x วินาที และช่วงปลายจะเป็นช่วงตั้งแต่ 0.7x วินาที จนถึงจุดสิ้นสุดของเซกเมนต์นั้น

ในการคำนวณหาค่าลักษณะสำคัญของเซกเมนต์นั้น จะเริ่มจากการหาค่าลักษณะสำคัญของแต่ละช่วงโดยการใช้สัมประสิทธิ์เซปสตรัมบนสเกลเมล ซึ่งมีค่าพลังงานรวมอยู่ด้วย อัตราการเปลี่ยนแปลง (Delta) และความเร่ง (Accelerations) โดยใช้กรอบสัญญาณเสียงขนาด 25 มิลลิวินาที และแต่ละกรอบจะมีระยะห่างกัน 10 มิลลิวินาที โดยจะได้เวกเตอร์ลักษณะสำคัญขนาด 39 มิติ จากนั้นจะนำเอาเวกเตอร์ลักษณะสำคัญของแต่ละช่วงที่ได้มารวมกันเป็นเวกเตอร์เดียวขนาด 117 มิติ โดยลักษณะการทำงานจะแสดงได้ในรูปที่ 3-5

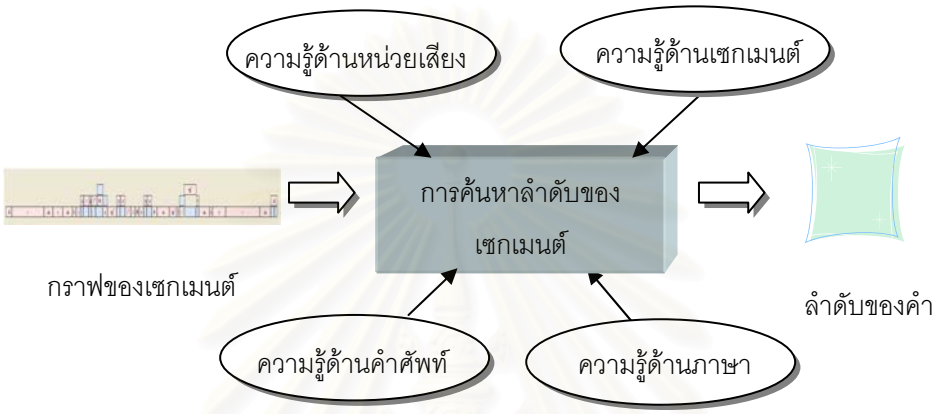



รูปที่ 3-5 การสกัดค่าลักษณะสำคัญ

ในรูปที่ 3-5 นั้น จะเป็นการแสดงการสกัดค่าลักษณะสำคัญจากเซกเมนต์ใดๆ โดยในตัวอย่างในรูปที่ 3-5 นี้ กำหนดให้เซกเมนต์มีความยาว X วินาที ซึ่งแบ่งเซกเมนต์เป็น 3 ช่วง โดยช่วงต้นนั้นจะมีขนาดความยาว 0.3X วินาที ช่วงกลางนั้นจะมีขนาด 0.4X วินาที และช่วงปลายมีขนาด 0.3X วินาที โดยผลลัพธ์สุดท้ายที่ได้นั้น จะเป็นเวกเตอร์ลักษณะสำคัญขนาด 117 มิติ

## 3.4 การค้นหาเส้นทางของเซกเมนต์

ในหัวข้อนี้ จะเป็นการค้นหาลำดับของคำที่มีค่าความน่าจะเป็นสูงสุดจากกราฟของเซกเมนต์ โดยจะอาศัยความรู้ต่างๆ เช่น ความรู้ในเรื่องของลักษณะของหน่วยเสียงต่างๆ ความรู้ด้านเซกเมนต์ ความรู้ทางภาษา และความรู้ทางด้านคำศัพท์ มาช่วยกัน เพื่อให้การค้นหาลำดับของเซกเมนต์นั้นเป็นไปอย่างถูกต้องมากยิ่งขึ้น โดยลักษณะในการทำงานจะสามารถแสดงได้รูปที่ 3-6




รูปที่ 3-6 ลักษณะการทำงานในการค้นหาลำดับของเซกเมนต์

ในรูปที่ 3-6 นั้น จะรับกราฟของเซกเมนต์ที่สร้างจากหัวข้อที่ 3.2เป็นอินพุต และให้ผลลัพธ์เป็นลำดับของคำที่มีค่าความน่าจะเป็นมากที่สุด โดยจะใช้ความรู้ทางด้านหน่วยเสียง เซกเมนต์ คำศัพท์ และภาษาในการพิจารณา โดยความรู้ทั้ง 3 ด้านนั้น จะอยู่ในรูปของแบบจำลองหน่วยเสียง (Acoustic Model) แบบจำลองระยะเวลา (Duration Model) แบบจำลองการออกเสียง (Pronunciation Model) และแบบจำลองภาษา (Language Model)

### 3.4.1 การคำนวณค่าความน่าจะเป็น

ในการหาลำดับของหน่วยเสียงที่จะเป็นผลลัพธ์นั้น เราต้องการค้นหาลำดับของคำจากกราฟของเซกเมนต์ โดยจะทำการค้นหาลำดับของคำที่ทำให้

$$W^* = \arg\max_{W} p(W \mid A) \qquad (3.1)$$

นั่นคือ หาลำดับของคำ $W$ ที่มีค่าความน่าจะเป็นที่จะเกิดขึ้นมากที่สุด ($W^*$) เมื่อเกิดค่าสังเกต (Observation) $A$ ซึ่งในที่นี่ค่าสังเกต $A$ ก็คือเวกเตอร์ลักษณะสำคัญของแต่ละเซกเมนต์นั่นเอง โดยที่ $W = \{w_1,...,w_j\}$ และ $A = \{a_1,...,a_i\}$ โดยสามารถเขียนสมการที่ (3.1) ใหม่ได้เป็น

$$W^* = \arg\max_{W} \sum_{\forall S,U} p(S,U,W \mid A) \approx \arg\max_{S,U,W} p(S,U,W \mid A) \qquad (3.2)$$

โดยที่ $W$ จะแทนลำดับของคำ $U$ แทนลำดับของหน่วยเสียง และ $S$ แทนลำดับของเซกเมนต์ในกราฟของเซกเมนต์ โดยจากสมการที่ (3.1) และ (3.2) จะเห็นว่าค่าความน่าจะเป็นที่จะเกิดลำดับของคำ $W$ เมื่อเกิดค่าสังเกต $A$ นั้น จะพิจารณาจากผลรวมค่าความน่าจะเป็นที่เป็นไปได้ทั้งหมดของทุกๆลำดับของ $S$ และ $U$ ของลำดับของคำ $W$ โดยค่าดังกล่าวนั้น ได้ประมาณให้เท่ากับค่า $p(S,U,W|A)$ ที่ดีที่สุด จากนั้นหาลำดับคำ $W^*$ ที่มีค่าความน่าจะเป็นดังกล่าวสูงที่สุด ทั้งนี้ค่าของ $p(S,U,W|A)$ สามารถเขียนใหม่ได้เป็น

$$p(S,U,W \mid A) = \frac{p(A \mid S,U,W)\,p(S \mid U,W)\,p(U \mid W)\,p(W)}{p(A)} \qquad (3.3)$$

โดยที่ $p(W)$ นั้นจะคือค่าความน่าจะเป็นที่จะเกิดลำดับของคำ $W$ ซึ่งสามารถแสดงได้ในจากสมการ

$$p(W) = \prod_{i=1}^{j} p(w_i \mid w_{i-n+1} \ldots\ldots w_{i-1}) \qquad (3.4)$$

ซึ่งในสมการที่ (3.4) จะเห็นว่า $w_i$ ใดๆนั้น จะขึ้นต่อคำที่อยู่ก่อนหน้าจำนวน $n$ คำ ($n$-gram) ในกรณีที่เราพิจารณาจากคำที่อยู่ก่อนหน้าเพียงคำเดียว หรือ ไบแกรม (Bigram) นั้น ค่าความน่าจะเป็นที่จะเกิดคำ $w_i$ ใดๆ จะสามารถหาได้จาก

$$p(w_i \mid w_{i-1}) = \frac{count(w_i, w_{i-1})}{\sum_k count(w_k, w_{i-1})} \qquad (3.5)$$

โดยที่ฟังก์ชัน $count(w_k, w_l)$ นั้น จะเป็นจำนวนของการเกิดคำ $w_k$ ขึ้นต่อจากคำ $w_l$ โดยการนับจำนวนนั้น จะนับจากฐานข้อมูลตัวอย่างที่ใช้ในการเรียนรู้

ค่า $p(U|W)$ นั้น จะบอกค่าความน่าจะเป็นที่จะเกิดลำดับของหน่วยเสียง $U$ เมื่อเกิดลำดับของคำเป็น $W$ เนื่องจากคำใดๆ นั้น สามารถประกอบจากหน่วยเสียงต่างๆ ได้หลายลักษณะ โดยที่ค่าความน่าจะเป็นในการเกิดลำดับของหน่วยเสียงในแต่ละลักษณะนั้นจะแตกต่างกัน ซึ่งจะแสดงในพจนานุกรม (Dictionary) โดยพจนานุกรมนี้จะบอกว่าคำๆ หนึ่งจะสามารถประกอบจากลำดับของหน่วยเสียงแบบใดได้บ้าง นอกจากนั้นแล้ว ยังต้องพิจารณาความคล้ายของลำดับของหน่วยเสียงในบางลักษณะที่สามารถเกิดขึ้นได้ เพื่อให้ครอบคลุมในกรณีที่ได้รับเสียง หรือเซกเมนต์ที่ไม่ตรงกับลำดับของหน่วยเสียงของคำนั้นทั้งหมด

สำหรับค่าของ $p(S|U,W)$ นั้น จะเป็นค่าความน่าจะเป็นที่จะเกิดลำดับของเซกเมนต์ $S$ เมื่อเกิดลำดับของหน่วยเสียง $U$ ลำดับของคำเป็น $W$ โดยจะพิจารณาจากแบบจำลองระยะเวลา (Duration model) ซึ่งแบบจำลองของความยาวของเซกเมนต์ในแต่ละหน่วยเสียง โดยการ

พิจารณาลักษณะของแต่ละเซกเมนต์นั้น จะขึ้นอยู่กับหน่วยเสียงและเป็นอิสระต่อคำ ดังนั้น $p(S|U,W)$ จึงสามารถคำนวณใหม่ได้เป็น $p(S|U)$

ค่า $p(A|S,U,W)$ จะบอกค่าความน่าจะเป็นที่จะเกิดค่าสังเกต $A$ โดยที่มีลักษณะของลำดับของเซกเมนต์ หน่วยเสียง และคำ เป็น $S$ $U$ และ $W$ ตามลำดับ โดยพิจารณาจากแบบจำลองเวกเตอร์ลักษณะสำคัญของเซกเมนต์สำหรับหน่วยเสียงใดๆ (หัวข้อที่ 3.3) ทั้งนี้เนื่องจากค่าสังเกต $A$ นั้นจะเป็นอิสระต่อลำดับของคำ $W$ เมื่อเกิดลำดับของหน่วยเสียง และเซกเมนต์เป็น $U$ และ $S$ อีกทั้งเรายังสามารถอนุมาน $S$ ได้จาก $A$ ดังนั้น ค่าของ $p(A|S,U,W)$ จึงสามารถหาได้จาก $p(A|U)$

ในการหาค่าของ $p(A|U)$ และ $p(S|U,W)$ นั้น ได้อ้างอิงจากงานวิจัยของ Glass [Glass, 2002] ซึ่งได้ทำการทดลองรู้จำเสียงในระดับหน่วยเสียงและคำสำหรับการรู้จำเสียงแบบเซกเมนต์

โดยสรุปแล้ว การพิจารณาหาลำดับของคำ $W^*$ หรือลำดับของคำ $W$ ที่มีค่าความน่าจะเป็นสูงสุดนั้น จะคำนวณจากผลคูณของ $p(W)$, $p(U|W)$, $p(S|U)$ และ $p(A|U)$ ซึ่งสามารถเขียนได้เป็น

$$p(S,U,W \mid A) = \frac{p(A \mid U)\,p(S \mid U)\,p(U \mid W)\,p(W)}{p(A)} \qquad (3.6)$$

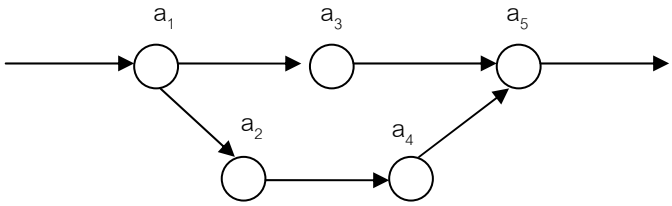
โดยที่แต่ละพจน์ในสมการที่ (3.6) นั้น จะสามารถคำนวณได้จากแบบจำลองต่างๆ ดังแสดงในตารางที่ 3-1

ตารางที่ 3-1แบบจำลองที่ใช้ในการคำนวณของพจน์ต่างๆ ในสมการที่3.6

| พจน์ | แบบจำลอง |
|---|---|
| $p(W)$ | แบบจำลองภาษา (Language Model) |
| $p(U\|W)$ | แบบจำลองการออกเสียง (Pronunciation Model) |
| $p(S\|U)$ | แบบจำลองระยะเวลา (Duration Model) |
| $p(A\|U)$ | แบบจำลองหน่วยเสียง (Acoustic Model) |

เนื่องจากเมื่อเราพิจารณาลำดับของเซกเมนต์ที่แตกต่างกันในกราฟของเซกเมนต์ ผลที่ได้ก็จะได้ลำดับของคำที่แตกต่างกัน หรือแม้จะได้ลำดับของคำเหมือนกัน ค่าความน่าจะเป็นที่จะเกิดลำดับของคำนั้นก็จะแตกต่างกัน ทั้งนี้เนื่องจากลำดับเซกเมนต์ที่แตกต่างกันนั้น จะส่งผลให้ได้ลำดับของค่าสังเกต หรือลำดับของเวกเตอร์ลักษณะสำคัญของแต่ละเซกเมนต์ต่างกัน ดังนั้นในการคำนวณใน $p(A|U)$ จึงต้องพิจารณาเซกเมนต์สมมุติฐานที่เกิดขึ้นทั้งหมด [Glass, 2002] นั่นคือ หากกำหนดให้เซตของ $M$ แทนเซตของลำดับของค่าสังเกตของลำดับของเซกเมนต์ $S$ และกำหนดให้เซต $N$ แทนเซตของของลำดับของค่าสังเกตอื่นๆ ที่เหลือ นั่นคือ $M \cap N = \phi$ และ

$M \cup N = A$ โดยที่ $A$ คือ เซตของค่าสังเกตทั้งหมดในกราฟของเซกเมนต์ พจน์ $p(A|U)$ จะคำนวณได้เป็น $p(M,N|U)$ โดยตัวอย่างของเซตของ $M$ และ $N$ จะสามารถแสดงได้ ดังรูปที่ 3-7




รูปที่ 3-7 ตัวอย่างลำดับของเซกเมนต์

ในรูปที่ 3-7 แสดงตัวอย่างลำดับของเซกเมนต์ โดยกำหนดให้เซตของค่าสังเกตทั้งหมดเป็น $\{a_1, a_2, a_3, a_4, a_5\}$ กำหนดให้ลำดับของค่าสังเกตของลำดับของเซกเมนต์ $S$ ซึ่งมีลำดับของหน่วยเสียง $U$ เป็น $a_1 \rightarrow a_3 \rightarrow a_5$ ในกรณีนี้เซต $M$ จะเป็นเซตของ $\{a_1, a_3, a_5\}$ และเซตของ $N$ จะเป็นเซตของ $\{a_2, a_4\}$

ในการสร้างแบบจำลองสำหรับ $p(M,N|U)$ นั้น ได้เพิ่มหน่วยของคำศัพท์ (lexical unit) เข้าไป เพื่อใช้ในการจัดการหน่วยของคำศัพท์อื่นๆ ที่ไม่ได้อยู่ใน $U$ ในกรณีที่สร้างแบบจำลองในระดับของหน่วยเสียง นอกจากจะต้องสร้างแบบจำลองสำหรับหน่วยเสียงทั่วไปแล้ว ยังต้องสร้างแบบจำลองสำหรับหน่วยเสียงอื่นๆ ที่ไม่ได้อยู่ใน $U$ ด้วย โดยแบบจำลองหน่วยเสียงนั้นจะแบ่งออกเป็น 2 กลุ่ม ดังนี้

- แบบจำลองหน่วยเสียงทั่วไป: $\{\alpha\}$
- แบบจำลองหน่วยเสียงที่ไม่ใช่หน่วยเสียงทั่วไป (Anti-phone: $\{\overline{\alpha}\}$)

โดยแบบจำลองหน่วยเสียงที่ไม่ใช่หน่วยเสียงทั่วไปนั้น จะใช้แทนเสียงอื่นๆ ที่ไม่ใช่หน่วยเสียงทั่วไป เช่น มีขนาดใหญ่ไป เล็กไป หรือเหลื่อมกัน

โดยเมื่อเรากำหนดว่าเซตของค่าสังเกต $M$ และ $N$ นั้นเป็นอิสระต่อกัน เมื่อกำหนดลำดับของหน่วยเสียง $U$ ดังนั้น $p(M,N|U)$ จำสามารถคำนวณได้เป็น

$$p(M,N|U) = p(M|U)p(N|\overline{\alpha})\frac{p(M|\overline{\alpha})}{p(M|\overline{\alpha})} \propto \frac{p(M|U)}{p(M|\overline{\alpha})} \qquad (3.7)$$

ดังนั้นในสมการที่ (3.6) จึงสามารถเขียนได้เป็น

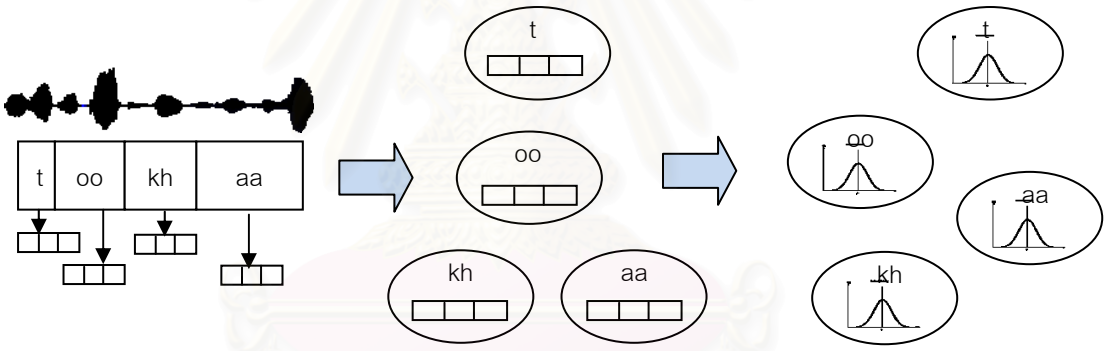
$$W^* = \arg\max_{S,U,W} \prod_{i=1}^{n} \frac{p(m_i|u_i)}{p(m_i|\overline{\alpha})} p(s_i|u_i)p(U|W)p(W) \qquad (3.8)$$

## 3.5 แบบจำลองต่าง ๆ

จากตารางที่ 3-1 จะเห็นว่าในการค้นหาลำดับของคำที่ดีที่สุด $W^*$ จากลำดับของค่าสังเกต $A$ ในสมการที่ (3.8) นั้น จะต้องอาศัยแบบจำลองต่างๆ นั่นคือ แบบจำลองหน่วยเสียง แบบจำลองระยะเวลา แบบจำลองการออกเสียง และแบบจำลองภาษา โดยในหัวนี้จะกล่าวถึงรายละเอียดของแต่ละแบบจำลองต่างๆ ดังนี้

### 3.5.1 แบบจำลองหน่วยเสียง

แบบจำลองหน่วยเสียงนั้น จะใช้ในการพิจารณาค่าความน่าจะเป็นที่จะเกิดหน่วยเสียง $k$ ณ ตำแหน่งที่เซกเมนต์ $s$ ในงานวิจัยนี้ได้ใช้การแบบจำลองเกาส์เซียนในการคำนวณหาค่าความน่าจะเป็น โดยจะทำการแบ่งกลุ่มเวกเตอร์ลักษณะสำคัญของเซกเมนต์สำหรับแต่ละหน่วยเสียง จากนั้นก็ทำการสร้างแบบจำลองเกาส์เซียนและเรียนรู้ค่าพารามิเตอร์สำหรับแต่ละกลุ่มเวกเตอร์นั้น โดยลักษณะการสร้างแบบจำลองหน่วยเสียงจะแสดงในรูปที่ 3-8




การแบ่งขอบเขตของเซกเมนต์และสกัดค่าลักษณะสำคัญ

การแบ่งกลุ่มของเวกเตอร์ลักษณะสำคัญ

การสร้างรูปแบบการกระจายของแต่ละหน่วย

รูปที่ 3-8 ขั้นตอนการสร้างข้อมูลเพื่อใช้ในการเรียนรู้ค่าพารามิเตอร์การแจกแจงปกติของแต่ละเซกเมนต์

โดยในการเรียนรู้ค่าของพารามิเตอร์สำหรับการแจกแจงแบบเกาส์เซียนของแต่ละหน่วยเสียงนั้น จะใช้อัลกอริทึมอีเอ็ม (EM Algorithm) ซึ่งเป็นอัลกอริทึมที่ใช้ในการเรียนรู้ค่าพารามิเตอร์

### 3.5.1.1 การคำนวณค่าความน่าจะเป็นหน่วยเสียง

เมื่อมีค่าสังเกตของเซกเมนต์ใดๆ เข้ามานั้น การคำนวณความน่าจะเป็นที่ค่าสังเกตของเซกเมนต์นี้จะเกิดจากหน่วยเสียงใดๆ นั้น จะใช้การคำนวณจากแบบจำลองเกาส์เซียนของหน่วยเสียงนั้นในการพิจารณา โดยในการคำนวณนั้น จะพิจารณาจากสมการดังนี้

$$p(\vec{x} \mid k) = \frac{1}{(2\pi)^{d/2} \mid \sum_k \mid^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_k)\Sigma_k^{-1}(\vec{x}-\vec{\mu}_k)} \qquad (3.9)$$

โดยในสมการที่ (3-9) นั้น เป็นความน่าจะเป็นที่จะเกิดหน่วยเสียง $k$ ที่ค่าสังเกตของเซกเมนต์ $\vec{x}$ โดยที่ $\vec{\mu}_k$ จะแทนเวกเตอร์ค่าเฉลี่ยของแบบจำลองเกาส์เซียนของหน่วยเสียง $k$ ค่า $\mid \sum_k \mid$ และ $\sum_k$ จะแทนค่าดีเทอร์มิแนนต์ และเมทริกซ์ความแปรปรวนของแบบจำลองเกาส์เซียนของหน่วยเสียง $k$ ตามลำดับ

### 3.5.1.2 หน่วยเสียงที่ใช้ในการสร้างแบบจำลอง

ในวิทยานิพนธ์นี้ได้แบ่งประเภทของหน่วยเสียงที่จะใช้ในการสร้างแบบจำลองออกเป็น 3 กลุ่มใหญ่ๆ ตามลักษณะการใช้งาน ได้แก่

- แบบจำลองหน่วยเสียงทั่วไป
- แบบจำลองของหน่วยเสียงที่เกิดจากการแบ่งเซกเมนต์ที่ผิดพลาด
- แบบจำลองหน่วยเสียงขยะ (Garbage Model)

### 3.5.1.3 แบบจำลองหน่วยเสียงทั่วไป

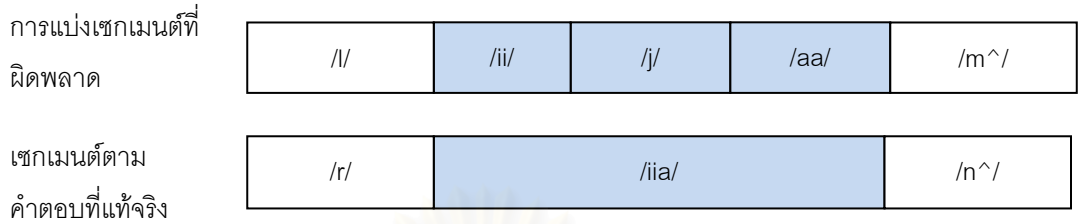
กลุ่มของแบบจำลองในประเภทนี้นั้น จะเรียนรู้จากเวกเตอร์ลักษณะสำคัญของหน่วยเสียงพยัญชนะ สระ และตัวสะกดของหน่วยเสียงภาษาไทย โดยจะมีหน่วยเสียงทั้งหมด 74 หน่วยเสียง และรวมถึงเสียงเงียบ (sil) ด้วย ดังแสดงในตาราง ก-1, ข-1 และ ค-1

### 3.5.1.4 แบบจำลองของหน่วยเสียงที่เกิดจากการแบ่งเซกเมนต์ที่ผิดพลาด

ในแบบจำลองนี้ จะทำการสร้างแบบจำลองจากค่าสังเกตของเซกเมนต์ที่เกิดจากการแบ่งเซกเมนต์ที่ผิดพลาด ซึ่งอาจเกิดจากการแบ่งเซกเมนต์ที่ใหญ่เกินไป หรือเล็กเกินไป โดยจะนำเอาผลลัพธ์ของเซกเมนต์สมมุติฐานมาเปรียบเทียบกับลำดับของเซกเมนต์ที่แบ่งขอบเขตมาแล้ว ว่าเป็นเซกเมนต์ที่เกิดจากความผิดพลาดหรือไม่ และเป็นความผิดพลาดประเภทใด แล้วจะทำการสร้างแบบจำลองของค่าสังเกตของเซกเมนต์ที่เกิดจากข้อผิดพลาดตามประเภทต่างๆนั้น โดยได้แบ่งประเภทของข้อผิดพลาดออกเป็น 2 ประเภทได้แก่

1. แบบจำลองของหน่วยเสียงของเซกเมนต์ที่เกิดเกินขึ้นมา (Insertion)

แบบจำลองนี้ใช้สำหรับเซกเมนต์ที่เกิดขึ้นเกินขึ้นมา ซึ่งอาจเกิดจากสัญญาณเสียงรบกวน โดยลักษณะของการแบ่งเซกเมนต์ที่ผิดพลาดนี้ สามารถแสดงได้ในรูปที่ 3-9

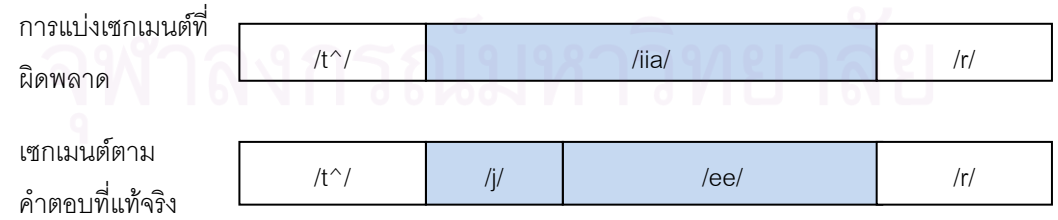



รูปที่ 3-9 ตัวอย่างการแบ่งเซกเมนต์ที่เกิดเกินขึ้นมา

รูปที่ 3-9 นั้น จะเปรียบเทียบการแบ่งเซกเมนต์ โดยในรูปบนนั้น จะเป็นผลจากการแบ่งเซกเมนต์ และในรูปล่างนั้น จะเป็นการแบ่งเซกเมนต์ที่จะนำมาเปรียบเทียบ โดยจะเห็นว่าในเซกเมนต์ของหน่วยเสียง /iia/ หรือ สระเอียในภาษาไทย จะถูกแบ่งเซกเมนต์ออกมาได้เป็น 3 เซกเมนต์ ได้แก่ เซกเมนต์ของหน่วยเสียง /ii/ /j/ และ /aa/ ตามลำดับ โดยเซกเมนต์ทั้ง 3 เซกเมนต์นี้ จะถือว่าเป็นเซกเมนต์ที่เกิดข้อผิดพลาดจากการที่มีเซกเมนต์เกินขึ้นมา และจะนำเอาเซกเมนต์ที่มีลักษณะเหล่านี้มาสร้างแบบจำลอง

2. แบบจำลองของหน่วยเสียงที่แบ่งเซกเมนต์หายไป (Deletion)

แบบจำลองนี้จะใช้สำหรับเซกเมนต์ที่หายไป ซึ่งอาจเกิดจากมีการแบ่งเซกเมนต์ครอบคลุมไปยังหน่วยเสียงอื่น ทำให้เซกเมนต์ของบางหน่วยเสียงหายไป โดยในการพิจารณานั้น จะนำเอาเซกเมนต์ในกราฟของเซกเมนต์ที่ได้ มาเปรียบเทียบกับลำดับของเซกเมนต์ที่ได้กำหนดไว้ โดยพิจารณาจากขอบเขตทางขวา และขอบเขตทางซ้าย หากมีเซกเมนต์ใดที่ครอบคลุมเกิน 1 เซกเมนต์ขึ้นไป ก็จะทำการรวบรวมค่าสังเกตของเซกเมนต์เหล่านั้น และสร้างแบบจำลองขึ้น โดยตัวอย่างของการแบ่งเซกเมนต์ที่ผิดพลาดประเภทนี้ จะแสดงในรูปที่ 3-10




รูปที่ 3-10 ตัวอย่างการแบ่งเซกเมนต์ที่ผิดพลาดจากการที่แบ่งเซกเมนต์หายไป

ในรูปที่ 3-10 นั้น จะเป็นตัวอย่างการแบ่งเซกเมนต์ที่หายไป โดยในรูปล่างนั้น จะแทนการแบ่งเซกเมนต์ที่ถูกต้อง และรูปบนนั้น จะแทนการแบ่งเซกเมนต์ที่ได้ โดยจากรูปจะเห็นว่าที่

ตำแหน่งเซกเมนต์ของหน่วยเสียง /j/ และ /ee/ ของการแบ่งหน่วยเสียงที่ถูกต้องนั้น ผลของการแบ่งเซกเมนต์กลับได้เป็นเซกเมนต์ของหน่วยเสียง /iia/ ซึ่งมีความยาวครอบคลุมเซกเมนต์ /j/ และ /ee/ โดยในที่นี้อาจกล่าวได้ว่า เซกเมนต์ของหน่วยเสียง /iia/ นั้น เป็นการแบ่งเซกเมนต์ที่ผิดพลาดโดยแบ่งเซกเมนต์หายไป และจะสร้างแบบจำลองสำหรับค่าสังเกตของเซกเมนต์เหล่านี้เข้าด้วยกัน

### 3.5.1.5 **แบบจำลองหน่วยเสียงขยะ** (Garbage Model)

แบบจำลองนี้จะให้แทนหน่วยเสียงที่ปรากฏขึ้น นอกเหนือจากคำศัพท์ที่มีในพจนานุกรมขึ้นมา รวมถึงเสียงไอ จาม หรือ เสียงที่ผู้พูดพูดขึ้นมาในขณะที่ใช้ความคิด (เช่น เอ่อ, อ่า) โดยกลุ่มของแบบจำลองประเภทนี้ ในวิทยานิพนธ์นี้ได้เสนอแนวทางการสร้างแบบจำลองโดยการแบ่งหน่วยเสียงออกเป็นกลุ่มๆ จากนั้นจึงทำการสร้างแบบจำลองของหน่วยเสียงในแต่ละกลุ่ม โดยได้เลือกแบ่งหน่วยเสียงออกเป็นกลุ่มตามงานวิจัยของกลุ่มนักวิจัย Tangruamsub [Tangruamsub et. al., 2007] ได้แก่ กลุ่มของหน่วยเสียงของพยัญชนะในภาษาไทย กลุ่มของหน่วยเสียงของสระในภาษาไทย และกลุ่มของหน่วยเสียงของตัวสะกดในภาษาไทย และสร้างแบบจำลองสำหรับแต่ละกลุ่มของหน่วยเสียงนั้น โดยที่กลุ่มของหน่วยเสียงต่างๆ จะแสดงในตารางที่ 3-2

ตารางที่ 3-2 การแบ่งกลุ่มของหน่วยเสียง

| กลุ่มของหน่วยเสียง | หน่วยเสียง |
|---|---|
| พยัญชนะต้น | b, bl, br, c, ch, d, dr, f, fl, fr, h, j, k, kh, khl, khr, khw, kl, kr, kw, l, m, n, ng, p, ph, phl, phr, pl, pr, r, s, t, th, thr, tr, w, x, xx, z; |
| กลุ่มของหน่วยเสียงสระ | @, @@, a, aa, e, ee, i, ia, ii, iia, o, oo, q, qq, u, ua, uu, uua, v, va, vv, vva |
| กลุ่มของหน่วยเสียงตัวสะกด | ch^, f^, j^, k^, l^, m^, n^, ng^, p^, s^, t^, w^ |

### 3.5.1.6 **การคำนวณค่าความจะเป็น** $p(m_i \mid \overline{\alpha})$

การคำนวณค่า $p(m_i \mid \overline{\alpha})$ ในสมการที่ (3.8) วิทยานิพนธ์นี้ได้เสนอแนวทางในการคำนวณจากแบบจำลองของหน่วยเสียงใดๆ ที่ทำให้ค่าของ $p(m_i \mid \overline{\alpha})$ มีค่ามากที่สุด โดยจะใช้แบบจำลองหน่วยเสียง ร่วมกับแบบจำลองแบบหน่วยเสียงขยะ มาช่วยในการพิจารณา แล้วเลือกค่าที่ดีที่สุดนั้นในการแทนค่าลงพจน์ $p(m_i \mid \overline{\alpha})$ ในสมการ (3.8) หรือสามารถเขียนสมการ ได้เป็น

$$W^* = \arg\max_{S,U,W} \prod_{i=1}^{n} \frac{p(m_i \mid u_i)}{p(m_i \mid \overline{\alpha}_{\max})} p(s_i \mid u_i)\, p(U \mid W)\, p(W) \qquad (3.10)$$

โดยที่ $p(m_i \mid \overline{\alpha}_{\max})$ จะคำนวณแบบจำลองของหน่วยเสียง หรือ หน่วยเสียงขยะใดๆ ที่ทำให้ $p(m_i \mid \overline{\alpha})$ มีค่ามากที่สุด ณ ที่ตำแหน่งของค่าสังเกต $m_i$

### 3.5.2 **แบบจำลองระยะเวลา** (Duration Model)

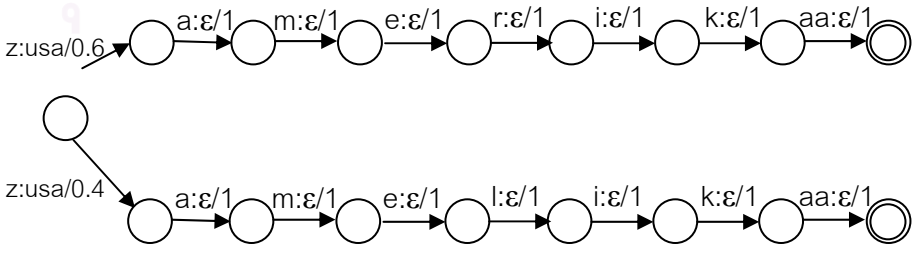
แบบจำลองระยะเวลานี้ จะเป็นแบบจำลองที่นำมาใช้เพื่อคำนวณค่าของ $p(s_i|u_i)$ ในสมการที่ (3.10) ซึ่งเป็นค่าความน่าจะเป็นที่จะเกิดเซกเมนต์ $s_i$ เมื่อเกิดหน่วยเสียง $u_i$ ขึ้น โดยในวิทยานิพนธ์นี้ ได้พิจารณาค่าความน่าจะเป็นดังกล่าว จากลักษณะของความยาวของเซกเมนต์ของแต่ละหน่วยเสียงในหัวข้อที่ 3.5.1 ซึ่งรวมถึงเซกเมนต์ที่เกิดจากข้อผิดพลาด และเซกเมนต์ของหน่วยเสียงขยะด้วย โดยจะนำเอาความยาว หรือระยะเวลาของเซกเมนต์นั้น ซึ่งจะคำนวณจากเวลาที่ตำแหน่งขอบเขตขวาลบด้วยเวลาที่ตำแหน่งขอบเขตซ้ายของเซกเมนต์นั้น มาสร้างแบบจำลองโดยใช้การแจกแจงแบบเกาส์เซียนของความยาวของแต่ละหน่วยเสียง ดังนั้นในการคำนวณ $p(s_i|u_i)$ จึงคำนวณได้จาก

$$p(s_i \mid u_i) = \frac{1}{\sqrt{2\pi\sigma_{ui}^{\,2}}} e^{-(x_{si}-\mu_{ui})^2/2\sigma_{ui}^{\,2}} \qquad (3.11)$$

โดยที่ $\mu_{ui}$ และ $\sigma_{ui}$ จะแทนค่าเฉลี่ยและค่าความแปรปรวนของแบบจำลองเกาส์เซียนหน่วยเสียง $u_i$ ตามลำดับ $x_{si}$ จะแทนค่าความยาวของเซกเมนต์ $s_i$

### 3.5.3 **แบบจำลองการออกเสียง** (Pronunciation Model)

แบบจำลองการออกเสียงนั้น จะใช้ในการแจกแจงคำศัพท์แต่ละคำในระบบว่าประกอบด้วยหน่วยเสียงใด และในลำดับใดบ้าง รวมถึงค่าความน่าจะเป็นที่จะเกิดในแต่ละลำดับ ทั้งนี้ คำศัพท์ใดๆ สามารถออกเสียงได้หลายแบบ เช่น คำว่า “อเมริกา” สามารถออกเสียงได้เป็น อะ-เม-ริ-กา (z a m ee r i k aa) หรือ อะ-เม-ลิ-กา (z a m ee l i k aa) ดังตัวอย่างในรูปที่ 3-11

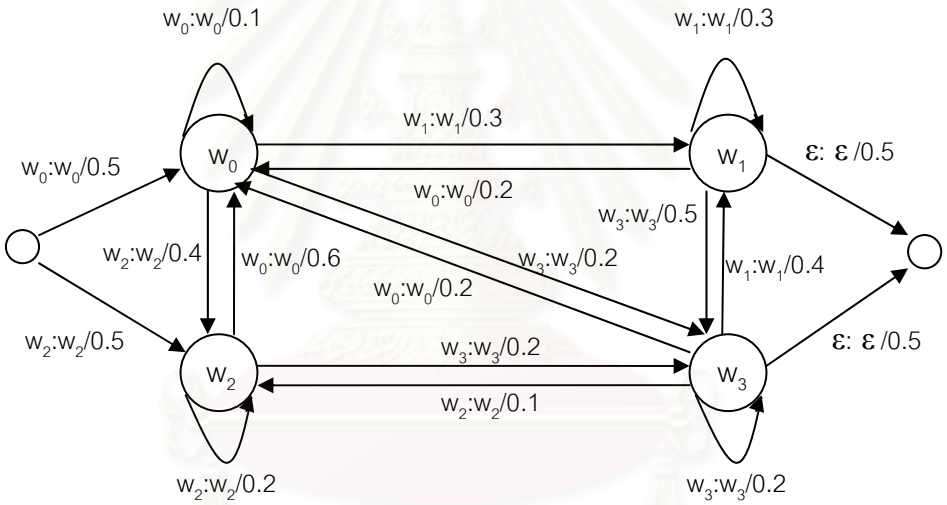



รูปที่ 3-11ตัวอย่างแบบจำลองการออกเสียงของคำว่า “อเมริกา”

ในรูปที่ 3-11 นั้น จะเป็นตัวอย่างของแบบจำลองการออกเสียงของคำว่า “อเมริกา” ในตัวอย่างนี้จะเห็นว่าคำว่า “อเมริกา” นั้น สามารถประกอบด้วยลำดับของหน่วยเสียงได้ 2 แบบ โดยแต่ละแบบจะมีค่าความน่าจะเป็นที่ต่างกัน หากมีลำดับเป็น “z a m ee r i k aa” จะมีค่าความน่าจะเป็น 0.6 และ มีลำดับเป็น “z a m ee l i k aa” มีความน่าจะเป็นเป็น 0.4 ซึ่งก็คือ ค่าความน่าจะเป็นในพจน์ของ $p(U|W)$ ในสมการที่ (3.10) นั่นเอง โดยจะทำการสร้างแบบจำลองการออกเสียงสำหรับแต่ละคำศัพท์แยกออกจากกัน

### 3.5.4 แบบจำลองภาษา

แบบจำลองภาษานั้น จะเป็นลักษณะของโครงข่ายคำ เพื่อบอกลำดับของคำที่เป็นไปได้ รวมถึงค่าความน่าจะเป็นของแต่ละลำดับของคำ โดยค่าความน่าจะเป็นนี้จะสามารถพิจารณาได้จากตัวอย่างเสียงหรือข้อความ โดยตัวอย่างของแบบจำลองภาษานั้นสามารถแสดงได้ในรูปที่ 3-12




รูปที่ 3-12 ตัวอย่างแบบจำลองภาษา

ในรูปที่ 3-12 นี้ จะแสดงแบบจำลองภาษา โดยที่ในตัวอย่างนี้จะเป็นแบบจำลองภาษาสำหรับคำ 4 คำ ได้แก่ $W = \{w_0, w_1, w_2, w_3\}$ โดยที่แต่ละบัพจะแทนคำ $w_i$ ใดๆ และจะมีเส้นเชื่อมเชื่อมถึงกันระหว่างบัพ โดยจะมีข้อความกำกับเป็น *“อินพุต: เอาท์พุท / ค่าความน่าจะเป็น”* ตัวอย่างเช่น “$w_1$: $w_1$/0.3” จะหมายถึง รับอินพุตเป็นคำ $w_1$ และให้ผลลัพธ์เป็น $w_1$ โดยที่ค่าความน่าจะเป็นในการเปลี่ยนจากสถานะที่บัพเริ่มต้นของลูกศร ไปยังบัพปลายทางของลูกศรนั้น จะมีค่าเป็น 0.3 โดยในที่นี้ได้ละเส้นเชื่อมระหว่างบัพที่มีค่าความน่าจะเป็นในการเปลี่ยนสถานะเป็น 0 เช่นในตัวอย่างในรูปที่ 3-12 นั้น ค่าความน่าจะเป็นที่จะเปลี่ยนสถานะจาก $w_2$ ไปยัง $w_1$ มีค่าเป็น 0 หรือ ค่าความน่าจะเป็นที่จะเปลี่ยนจากสถานะเริ่มต้น ไปยัง $w_1$ มีค่าเป็น 0 เป็นต้น โดยที่ค่าความ

น่าจะเป็นเหล่านี้จะได้มาจากการเรียนรู้จากตัวอย่างข้อมูล หรืออาจจะกำหนดให้ค่าความน่าจะเป็นเหล่านี้เป็น 1/n เมื่อ n แทนจำนวนเส้นเชื่อมทั้งหมดที่ออกจากบัพที่กำลังพิจารณาอยู่ เป็นต้น

### 3.6 การค้นหา

ในการค้นหานั้น จะทำการค้นหาโดยการประยุกต์ใช้อัลกอริทึมวิเทอร์บี (Viterbi Algorithm) โดยจะมองว่าเป็นการค้นหาเส้นทางที่ดีที่สุดจากโครงข่าย 2 โครงข่าย คือ

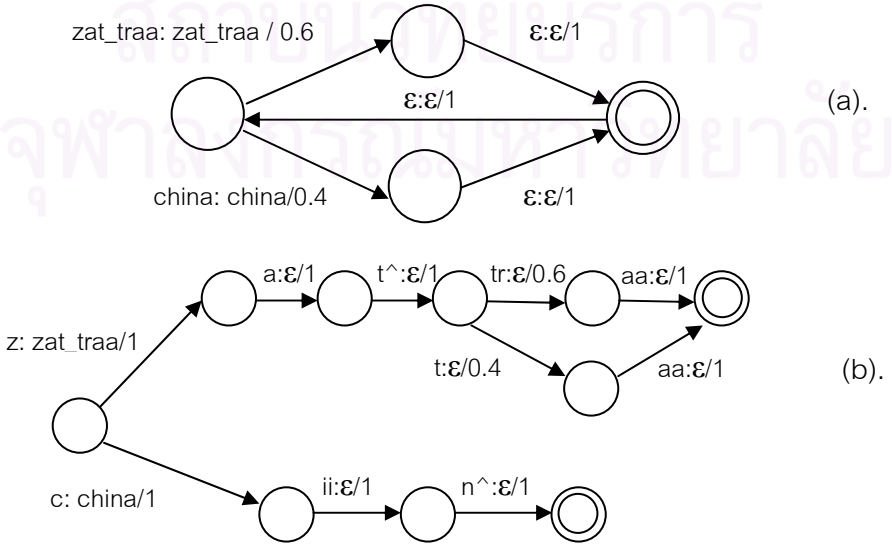
- โครงข่ายการออกเสียง (Pronunciation Graph): $U$
- โครงข่ายหน่วยเสียง (Acoustic-Phonetic Graph): $A$

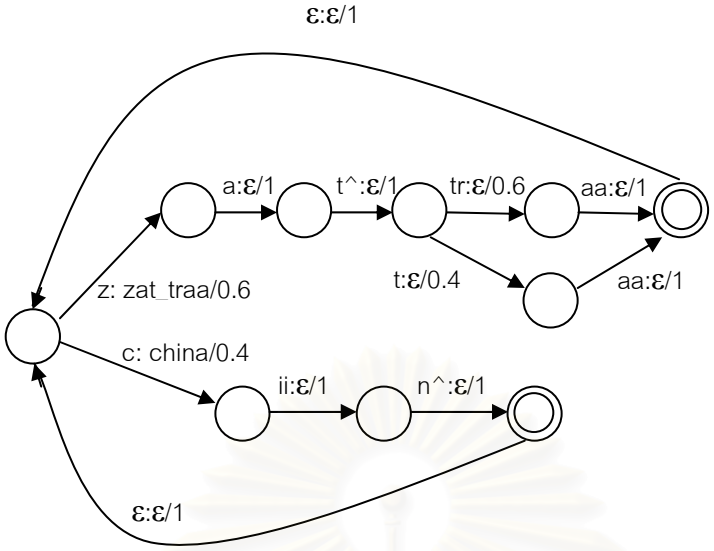
สำหรับโครงข่ายการออกเสียงนั้น จะเป็นโครงข่ายที่แสดงลำดับของคำที่เป็นไปได้ทั้งหมด รวมถึงลำดับของหน่วยเสียงที่ประกอบเป็นคำศัพท์แต่ละคำทั้งหมด โดยหากกำหนดให้ $G$ แทนแบบจำลองภาษา และ $L$ แทนโครงข่ายที่ทำการแปลลำดับของหน่วยเสียงให้อยู่ในรูปของลำดับของคำ เมื่อกำหนดให้ $U$ แทนโครงข่ายคำศัพท์ จะได้ว่า

$$U = L \circ G$$

นั่นคือ โครงข่ายคำ $U$ นั้นจะเกิดจากการนำเอาแบบจำลองภาษาและโครงข่ายที่ทำการแปลจากหน่วยเสียงให้อยู่ในรูปของคำศัพท์มาทำการประกอบ (Compose) กัน ซึ่งจะได้เป็นโครงข่ายที่ทำการแปลจากหน่วยเสียงไปสู่ลำดับของคำ โดยขั้นตอนในการสร้างโครงข่าย $L$ นั้น จะทำโดยสร้างแบบจำลองการออกเสียงสำหรับแต่ละคำศัพท์ (หัวข้อที่ 3.5.3) จากนั้นนำเอาแบบจำลองการออกเสียงทั้งหมดมาทำการรวมกัน (union) จนได้โครงข่ายขึ้น

ในการทำฟังก์ชันการประกอบกันระหว่างโครงข่าย $L$ และ $G$ นั้น จะสามารถแสดงได้ดังตัวอย่างในรูปที่ 3-13






(c).

รูปที่ 3-13 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่าย $G$ และ $L$ (a) แบบจำลองภาษา $G$ (b) โครงข่ายจับคู่ระหว่างคำศัพท์และหน่วยเสียง $L$ (C) ผลของการทำฟังก์ชันประกอบของ $G$ และ $L$

ในรูปที่ 3-13(a) นั้น จะเป็นตัวอย่างของแบบจำลองภาษา ซึ่งบอกลักษณะลำดับของคำด้วยค่าความน่าจะเป็นต่างๆ โดยที่เลเบลบอกเส้นเชื่อมระหว่างบัพนั้น จะบอก *"อินพุต: เอาท์พุต/ค่าความน่าจะเป็น"* ซึ่งค่าความน่าจะเป็นนั้น จะเป็นค่าความน่าจะเป็นในการเปลี่ยนจากบัพต้นทางไปยังปลายทาง เช่น *"zat_traa: zat_traa / 0.6"* นั้นจะหมายถึง เมื่อได้รับอินพุตเป็น *"zat_traa"* เข้ามานั้น จะให้เอาท์พุตเป็น *"zat_traa"* โดยจะมีค่าความน่าจะเป็นจากการย้ายสถานะจากบัพที่อยู่ที่ต้นเส้นเชื่อมไปยังบัพที่ปลายเส้นเชื่อมเป็น 0.6 ซึ่งค่าความน่าจะเป็นนี้สามารถเขียนได้เป็น $p(N_i|N_{i-1})$โดยที่ $N_i$ แทนบัพที่ $i$ แต่เนื่องจากโครงข่ายในรูปที่ 3-13 นั้น ในการเปลี่ยนสถานะจะพิจารณาจากอินพุตซึ่งเป็นคำเข้ามา ดังนั้นจึงสามารถพิจารณาได้ว่า ค่าความน่าจะเป็นของเส้นเชื่อมนั้น คือ $p(w_i|w_{i-1})$ หรือค่าความน่าจะเป็นที่จะเกิดคำ $w_i$ ในลำดับที่ $i$ เมื่อเกิด $w_{i-1}$ในลำดับ $i$-$1$ ในรูปที่ 3-13(a)จะเห็นว่าได้กำหนดให้มีคำศัพท์ทั้งหมด 2 คำ คือ คำว่า "อัตรา" และ "จีน" ซึ่งลำดับของคำนั้นสามารถเริ่มจากคำใดก่อนก็ได้ โดยจะมีค่าความน่าจะเป็นที่แตกต่างกัน เมื่อจบคำใดคำหนึ่งแล้ว ก็จะสามารถอยู่ในสถานะสิ้นสุด หรือสามารถกับมาเริ่มใหม่ได้ โดยเอาท์พุตที่ได้ในแต่ละเส้นทางนั้น จะเป็นการนำเอาเอาท์พุตในแต่ละเส้นเชื่อมมาต่อกัน การคำนวณค่าความน่าจะเป็นในแต่ละเส้นทางนั้น จะคือ ผลคูณของน้ำหนักของแต่ละเส้นเชื่อมที่เดินทางผ่าน

ในรูปที่ 3-13(b) นั้น จะเป็นการแสดงตัวอย่างโครงข่าย $L$ ซึ่งจะเป็นการจับคู่จากหน่วยเสียงไปยังคำศัพท์ ในตัวอย่างนี้จะเป็นการแสดงสำหรับคำศัพท์เพียง 2 คำ โดยคำว่า "อัตรา" นั้น

ได้กำหนดให้สามารถออกเสียงได้เป็น z-a-t^-tr-aa (อัด - ตรา) หรือ z-a-t^-t-aa (อัด-ตา) โดยโครงข่ายนี้จะบอกค่าความน่าจะเป็นของลำดับของหน่วยเสียงแต่ละแบบเมื่อกำหนดคำศัพท์ ซึ่งน้ำหนักของเส้นเชื่อมที่ปรากฏในข้อความบนเส้นเชื่อมนั้น จะคือค่าความน่าจะเป็น $p(u_i| u_{i-1}, w_j)$ หรือ ค่าความน่าจะเป็นที่จะเกิดหน่วยเสียง $u_i$ ที่ลำดับที่ $i$ เมื่อเกิดหน่วยเสียง $u_{i-1}$ ที่ลำดับที่ $i-1$ เมื่อกำหนดให้คำที่กำลังพิจารณาอยู่ คือ คำ $w_j$

เมื่อเรานำเอาโครงข่าย $G$ และ $L$ มาทำการประกอบกัน จะได้ผลลัพธ์สุดท้ายในรูปที่ 3-13 (c) ซึ่งจะเป็นการแปลจากหน่วยเสียงให้อยู่ในรูปของลำดับของคำศัพท์ โดยจะเห็นว่าน้ำหนักของเส้นเชื่อมได้เปลี่ยนแปลงไปจากการนำเอามาประกอบกัน โดยน้ำหนักของเส้นเชื่อมนั้นจะมาจากผลคูณของน้ำหนักของเส้นเชื่อมจาก $G$ กับ $L$ และค่าความน่าจะเป็นของเส้นทาง (path) คือ ผลคูณของน้ำหนัก หรือค่าความน่าจะเป็นในแต่ละเส้นเชื่อมของเส้นทางนั้น หรืออาจเขียนได้ว่า

$$\begin{aligned} p(path) &= \prod_{i=1,j=1}^{n,m} p(u_i \mid u_{i-1}, w_j)\, p(w_j \mid w_{j-1}) \\ &= \prod_{i=1,j=1}^{n,m} p(u_i \mid u_{i-1}, w_j) \prod_{j=1}^{m} p(w_j \mid w_{j-1}) \\ &= p(U \mid W)\, p(W) \end{aligned} \qquad (3.12)$$

โดยที่ $p(path)$ คือ ค่าความจะเป็นของเส้นทางในโครงข่ายที่ได้จากการนำมาประกอบกันระหว่าง $G$ และ $L$ โดยที่ $u_i$ นั้นจะแทนหน่วยเสียงที่ลำดับที่ $i$ และ $w_j$ จะแทนคำศัพท์ลำดับที่ $j$ หน่วยเสียง $u_i$ นั้น จะขึ้นกับคำศัพท์ $w_j$ และเป็นอิสระต่อ $w_{j-1}$

สำหรับโครงข่ายหน่วยเสียงนั้นจะแสดงลำดับของเซกเมนต์ที่เป็นได้ทั้งหมด หรือก็คือกราฟของเซกเมนต์นั่นเอง ในการรู้จำคำนั้นจะค้นหาลำดับของคำที่ดีที่สุดจากโครงข่ายการออกเสียงและโครงข่ายหน่วยเสียงนี้ หรือ ค้นหาลำดับคำที่ดีที่สุดเมื่อพิจารณาจากลำดับของเซกเมนต์ที่มีอยู่ในกราฟของเซกเมนต์ (โครงข่ายหน่วยเสียง: $A$) และค่าความน่าจะเป็นที่จะเกิดลำดับของคำต่างๆ (โครงข่ายการออกเสียง: $U$) โดยจะเริ่มจากบัพเริ่มต้นของโครงข่ายทั้งสอง และจะมีการพิจารณาค่าคะแนนเพื่อหาเส้นทางที่ให้ค่าคะแนนสูงสุด ซึ่งคะแนนนั้นจะมาจากสมการที่ (3-10) โดยที่ใช้แบบจำลองเกาส์เซียนในการพิจารณาค่า $p(x_i|u_i)$ และ $p(x_i| \bar{\alpha})$ และใช้แบบจำลองระยะเวลาในการคำนวณ $p(S|U)$ สำหรับในเทอม $p(U|W)p(W)$ นั้น จะพิจารณาจากน้ำหนักของเส้นเชื่อมในโครงข่ายคำศัพท์ $U$ ดังที่แสดงในสมการที่ (3.12)

เนื่องจากในลำดับของเซกเมนต์ที่เกิดขึ้นในกราฟของเซกเมนต์นั้น อาจมีบางเซกเมนต์ที่เกิดจากข้อผิดพลาดในการสร้างกราฟของเซกเมนต์ เช่น เซกเมนต์ที่เกิดเพิ่มขึ้นเกินมา (Insertion) หรือ การแบ่งเซกเมนต์หายไปในบางหน่วยเสียง (Deletion) ดังนั้นในการพิจารณาลำดับของคำจากกราฟของเซกเมนต์นั้น จึงได้เพิ่มสมมุติฐานในการเกิดข้อผิดพลาดเหล่านี้ในแต่ละเซกเมนต์

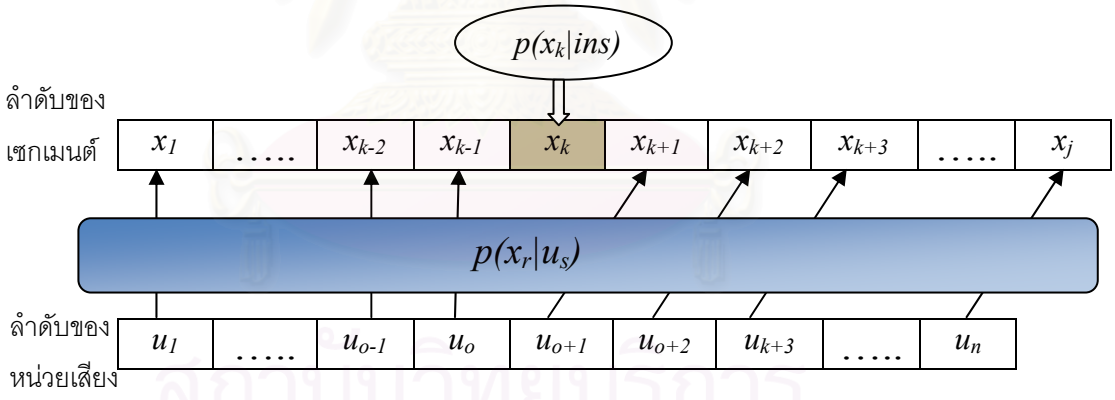
ด้วย ส่งผลให้ในการคำนวณพจน์ $p(x_i|u_i)$ ในสมการที่ (3.10) ได้มีการใช้แบบจำลองเกาส์เซียนที่ไม่เหมือนกันในแต่ละสมมุติฐาน โดยจะมีสมมุติฐานสำหรับแต่ละเซกเมนต์ดังนี้

1. แบบจำลองสำหรับคำนวณความน่าจะเป็นของเซกเมนต์ของหน่วยเสียง

ในกรณีนี้จะพิจารณาว่าเซกเมนต์นั้นเป็นหน่วยเสียงทั่วไปในภาษาไทย หรือหน่วยเสียงเงียบ ไม่ใช่เซกเมนต์ที่เกิดจากการแบ่งเซกเมนต์ที่ผิดพลาด ซึ่งในการคำนวณค่า $p(x_i|u_i)$ สามารถได้จากแบบจำลองเกาส์เซียนหน่วยเสียง $u_i$ ได้เลย

2. แบบจำลองสำหรับคำนวณความน่าจะเป็นของเซกเมนต์ของหน่วยเสียงที่เกิดข้อผิดพลาดในการที่เกิดเกินขึ้นมา (Insertion)

ในกรณีนี้จะใช้แบบจำลองของหน่วยเสียงของเซกเมนต์ที่เกิดเกินขึ้นมา (หัวข้อที่ 3.5.1.4) ในการคำนวณค่า $p(x_i|u_i)$ และ $p(s_i|u_i)$ หรือในที่นี้จะเขียนแทนด้วย $p(x_i|ins)$ และ $p(s_i|ins)$ ซึ่งข้อผิดพลาดดังกล่าวนั้น อาจเกิดจากการที่มีช่วงของเสียงเงียบแทรกขึ้นอยู่ระหว่างเซกเมนต์ หรือเกิดเซกเมนต์แทรกขึ้นมาในช่วงสั้นๆ ในกรณีนี้จะทำการข้ามเซกเมนต์นั้น แต่ยังคงอยู่ในสถานะตำแหน่งเดิมของโครงข่ายการออกเสียง $U$ นั่นคือ หากกำหนดให้ค่าสังเกต $x_k$ เป็นค่าสังเกตที่เกิดจากการแบ่งเซกเมนต์ที่เกินขึ้นมาที่ตำแหน่ง $k$ โดย ณ ขณะนั้นกำลังพิจารณาหน่วยเสียง $u_o$ จากคำ $w_u$ ดังเช่นในรูปที่ 3-14

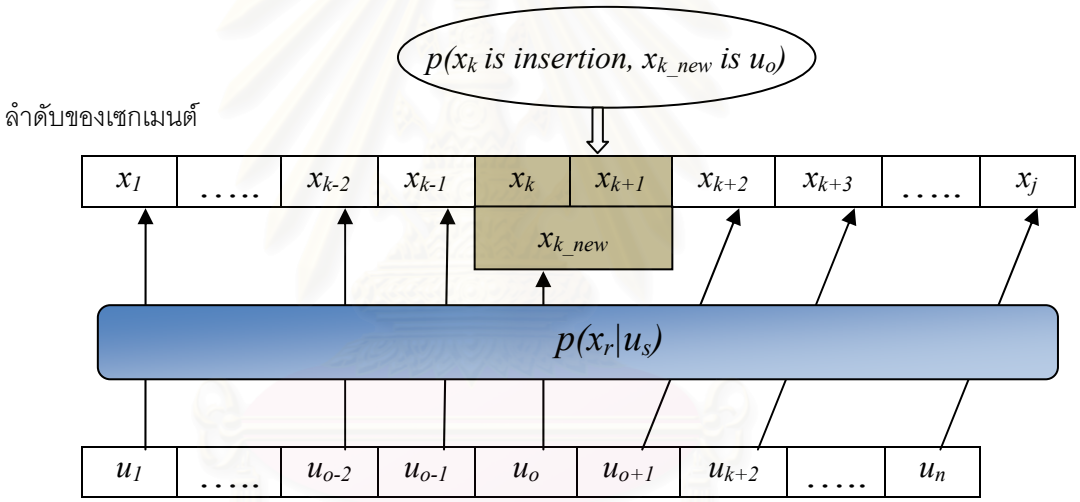



รูปที่ 3-14 ตัวอย่างการคำนวณค่า $p(x_i|u_i)$ เมื่อเกิดเซกเมนต์ที่เกินขึ้นมา

ในรูปที่ 3-14 จะเห็นว่าเซกเมนต์ที่ตำแหน่งอื่นที่ไม่ใช่เซกเมนต์ที่เกิดเกินขึ้นมานั้น จะคำนวณค่า $p(x_i|u_i)$ จากแบบจำลองเกาส์เซียนหน่วยเสียงปกติ ในขณะที่เมื่อเกิดเซกเมนต์ที่เกิดเกินขึ้นมานั้น ก็จะคำนวณเป็น $p(x_i|ins)$ แทน โดยจากรูปที่ 3-14 จะสามารถคำนวณในสมการที่ (3.10) ได้ว่า

$$p(W,U,S \mid A) = (\prod_{i=1,m=1}^{k-1,o} \frac{p(x_i \mid u_m)}{p(x_i \mid \overline{\alpha})} p(s_i \mid u_m) x \frac{p(x_k \mid ins)}{p(x_k \mid \overline{\alpha})} p(s_k \mid ins) x \prod_{i=k+1,m=o+1}^{j,n} \frac{p(x_i \mid u_m)}{p(x_i \mid \overline{\alpha})} x p(s_i \mid u_m)) p(U \mid W) p(W) \quad (3.13)$$

นอกจากนี้ข้อผิดพลาดที่เกิดเซกเมนต์เกินขึ้นมานั้น อาจเกิดมาจากการที่แบ่งเซกเมนต์ของหน่วยเสียงที่มีขนาดยาว ออกเป็นเซกเมนต์ย่อยๆ ที่มีขนาดสั้นลง เช่น คำว่า “นิวซีแลนด์” ซึ่งประกอบด้วยหน่วยเสียง “n i w^ s ii l xx n^” แต่เมื่อแบ่งเซกเมนต์กลับได้เป็น “n i w^ s ii l a xx a n^” โดยจะเห็นว่าในกรณีนี้ ได้แบ่งเซกเมนต์ของหน่วยเสียง /xx/ หรือหน่วยเสียงของสระแอ เพิ่มมาเป็น “a xx” นั่นคือ มีเซกเมนต์ของ /a/ เป็นเซกเมนต์ที่เพิ่มขึ้นมาจากเซกเมนต์ /xx/ และเป็นเซกเมนต์ที่เป็นข้อผิดพลาด ในกรณีนี้ จะทำการรวมเอาเซกเมนต์ทั้งสองเข้าด้วยกันให้กลายเป็นเซกเมนต์เดียว ดังรูปที่ 3-15



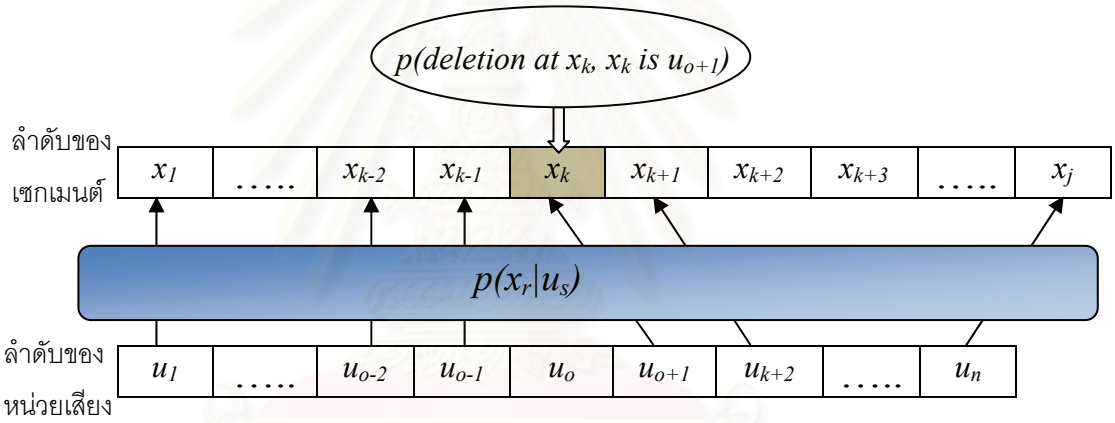

รูปที่ 3-15 ตัวอย่างการคำนวณค่า $p(x_i|u_i)$ เมื่อเกิดเซกเมนต์ที่เกินขึ้นมาในกรณีที่ 2

ในรูปที่ 3-15 นั้น เซกเมนต์ที่ $x_k$ นั้น จะเป็นเซกเมนต์ที่เกิดเกินขึ้นมาจากเซกเมนต์ที่ $x_{k+1}$ ดังนั้นจะทำการรวมเซกเมนต์ที่ $x_k$ และ $x_{k+1}$ เข้าด้วยกัน เป็นเซกเมนต์ $x_{k_new}$ โดยจะคำนวณค่าของ $p(W|A)$ ในกรณีนี้ได้เป็น

$$p(W,U,S \mid A) = (\prod_{i=1,m=1}^{k-1,o} \frac{p(x_i \mid u_m)}{p(x_i \mid \overline{\alpha})} p(s_i \mid u_m) x \frac{p(x_k \mid ins)}{p(x_k \mid \overline{\alpha})} p(s_k \mid ins) x \prod_{i=k_new,m=o+1}^{j,n} \frac{p(x_i \mid u_m)}{p(x_i \mid \overline{\alpha})} x p(s_i \mid u_m)) p(U \mid W) p(W) \quad (3.14)$$

3. แบบจำลองสำหรับคำนวณความน่าจะเป็นของเซกเมนต์ของหน่วยเสียงที่เกิดข้อผิดพลาดในการแบ่งเซกเมนต์หายไป (Deletion)

ในกรณีนี้ใช้แบบจำลองของหน่วยเสียงที่แบ่งเซกเมนต์หายไป (หัวข้อที่ 3.5.1.4) มาช่วยในการคำนวณค่าในพจน์ $p(x_i|u_i)$ โดยจะเพิ่มค่าความน่าจะเป็นจากแบบจำลองของหน่วยเสียงที่แบ่งเซกเมนต์หายไป หรือ $p(x_i|del)$ เข้าไปในการคำนวณสำหรับเซกเมนต์ที่น่าจะเกิดข้อผิดพลาดประเภทนี้ขึ้น โดยเซกเมนต์ที่หายไปนั้นอาจเกิดจากระยะเวลาของหน่วยเสียงที่เกิดขึ้นจริงๆ นั้นสั้นมาก จนทำให้ในการแบ่งเซกเมนต์ของหน่วยเสียงดังกล่าวนั้นขาดหายไป เช่น เมื่อเสียงที่เข้ามานั้นคือคำว่า “ประเทศ (pr a t ee t^)” โดยที่ถูกต้องแล้วจะต้องสามารถแบ่งเซกเมนต์ได้เป็น 6 เซกเมนต์ แต่เมื่อพิจารณาจากลำดับของเซกเมนต์ที่ได้จากกราฟของเซกเมนต์ พบว่าแบ่งเป็น 5 เซกเมนต์ ดังนี้ pr a t ee โดยเซกเมนต์ที่หายไป คือ เซกเมนต์สำหรับตัวสะกดแม่กด (t^) โดยตัวอย่างของความผิดพลาดในกรณีนี้สามารถแสดงได้ ดังรูปที่ 3-16



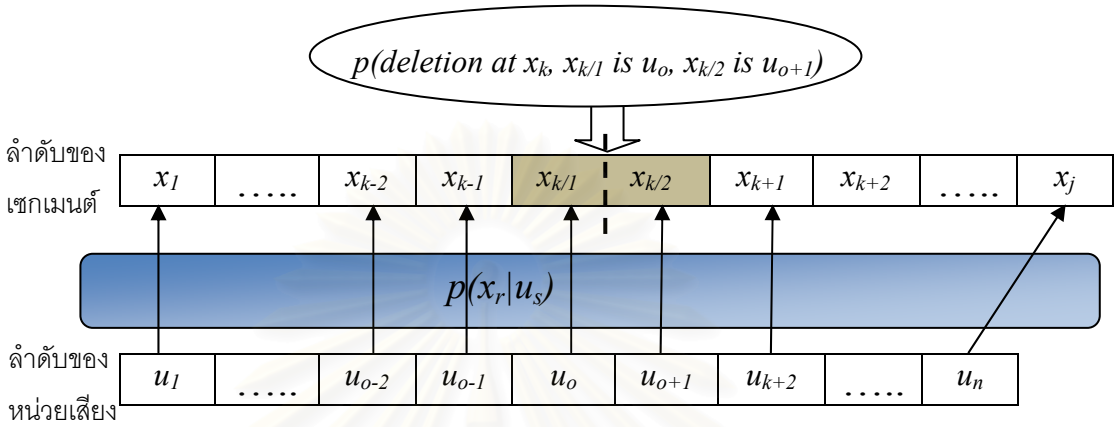

รูปที่ 3-16 ตัวอย่างการคำนวณค่า $p(x_i|u_i)$ เมื่อเกิดเซกเมนต์หน่วยเสียงหายไป

ในกรณีนี้ ได้ทำการแบ่งเซกเมนต์สำหรับหน่วยเสียง $u_o$ หายไป ในกรณีนี้จะทำการข้ามหน่วยเสียง $u_o$ จากโครงข่ายคำศัพท์ $U$ โดยสามารถเขียนเป็นสมการได้เป็น

$$p(W,U,S|A) = \prod_{i=1,m=1}^{k-1,o-1} \frac{p(x_i|u_m)}{p(x_i|\overline{\alpha})} p(s_i|u_m) x \frac{p(x_k|del)}{p(x_k|\overline{\alpha})} p(s_k|del) x \prod_{i=k,m=o+1}^{j,n} \frac{p(x_i|u_m)}{p(x_i|\overline{\alpha})} p(s_i|u_m) x p(U|W) p(W) \quad (3.15)$$

นอกจากนี้ สาเหตุอีกอย่างที่ทำให้เกิดข้อผิดพลาดในการแบ่งเซกเมนต์แบบขาดหายนั้นมาจากการที่รวมเอาหน่วยเสียงที่อยู่ใกล้เคียงกันเข้าด้วยกัน เช่น ในคำว่า “อยากทราบ (j aa k^ s aa p^)” เนื่องจากเสียง /j/ หรือ ซึ่งตรงกับเสียง “ย” ในภาษาไทยนั้น มีลักษณะเป็นแบบกึ่งสระ

(Semi-vowel) ทำให้เมื่อออกเสียงตามด้วย /aa/ หรือเสียงสระอาในภาษาไทย จะได้ลักษณะของเซกเมนต์ที่คล้ายกับหน่วยเสียงสระ /iia/ ดังนั้นจึงมีโอกาสที่กราฟของเซกเมนต์นั้น จะแบ่งเซกเมนต์ได้เป็น 5 เซกเมนต์ ดังนี้ “iia k^ s aa p^” แทนที่จะได้เป็น 6 เซกเมนต์ ในกรณีนี้เราจะแบ่งเซกเมนต์ที่สันนิษฐานว่าจะเกิดการขาดหายดังกล่าวออกเป็นส่วนย่อยๆ ดังรูปที่ 3-17




รูปที่ 3-17 ตัวอย่างการคำนวณค่า $p(x_i|u_i)$ เมื่อเกิดเซกเมนต์หายไปในกรณีที่ 2

ในรูปที่ 3-17 ได้ยกตัวอย่างให้เซกเมนต์ $x_k$ นั้นเป็นค่าสังเกตของเซกเมนต์ที่เกิดจากการรวมหน่วยเสียง $u_o$ และ $u_{o+1}$ เข้าด้วยกัน ดังนั้น จึงได้ทำการแบ่งเซกเมนต์ $x_k$ ออกเป็น 2 ส่วน คือ $x_{k/1}$ และ $x_{k/2}$ โดยในงานวิจัยนี้ จะทำการแบ่งออกเป็น 2 ส่วนที่มีขนาดเท่ากัน นั่นคือ มีความยาวเท่ากัน จากนั้นจึงคำการคำนวณค่าความน่าจะเป็นที่ค่าสังเกตที่ตำแหน่งที่ $k$ จะเป็นค่าสังเกตของเซกเมนต์ประเภทที่เกิดจากความผิดพลาดของหน่วยเสียงที่หายไป ($del$) และเมื่อทำการแบ่งเซกเมนต์เป็น $k/1$ และ $k/2$ แล้ว เซกเมนต์ที่ได้จะใช้พิจารณากับหน่วยเสียง $u_o$ และ $u_{o+1}$ โดยในกรณีนี้จะคำนวณ $p(W|A)$ ได้ดังนี้

$$p(W,U,S \mid A) = \prod_{i=1,m=1}^{k-1,o-1} \frac{p(x_i \mid u_m)}{p(x_i \mid \overline{\alpha})} p(s_i \mid u_m) x \frac{p(x_k \mid del)}{p(x_k \mid \overline{\alpha})} p(s_k \mid del) x \prod_{i=k/1,m=o}^{j,n} \frac{p(x_i \mid u_m)}{p(x_i \mid \overline{\alpha})} p(s_i \mid u_m) x p(U \mid W) p(W) \quad (3.16)$$

โดยในสมการที่ (3.16) นั้น $p(x_k \mid del)$ จะแทนค่าความน่าจะเป็นที่เซกเมนต์ $x_k$ นั้น เป็นเซกเมนต์ที่เกิดจากข้อผิดพลาดที่เกิดจากการแบ่งหน่วยเสียงหายไป พจน์ $p(x_{k/1} \mid u_o)$ และ $p(x_{k/2} \mid u_{o+1})$ จะแทนค่าความน่าจะเป็นที่จะเกิดค่าสังเกต $x_{k/1}$ เมื่อเกิดหน่วยเสียง $u_o$ ที่ตำแหน่ง $o$ และค่าความน่าจะเป็นที่จะเกิดค่าสังเกต $x_{k/2}$ เมื่อเกิดหน่วยเสียง $u_{o+1}$ ที่ตำแหน่ง $o+1$ ตามลำดับ

จากที่กล่าวมาขั้นต้น จะสามารถสรุปได้ว่า ในขณะที่พิจารณาเซกเมนต์ $s_i$ ใดๆ ว่าเซกเมนต์นั้นเป็นหน่วยเสียงใดนั้น ในวิทยานิพนธ์ได้เสนอแนวทางในการพิจารณาตามสมมุติฐานต่างๆ ในตารางที่ 3-3

ตารางที่ 3-3 สมมุติฐานในกรณีต่างๆ ของแต่ละเซกเมนต์

| สมมุติฐาน | กรณีในการพิจารณา |
|---|---|
| เซกเมนต์เป็นหน่วยเสียงทั่วไป | เซกเมนต์นั้นเป็นหน่วยเสียงของคำศัพท์ |
| เซกเมนต์เป็นเซกเมนต์ที่เกิดเกินมา | **กรณี** 1: เซกเมนต์นั้นเกิดเกินขึ้นมา ซึ่งอาจเกิดจากสัญญาณเสียงรบกวน หรือเสียงเงียบ แต่ไม่ใช่หน่วยเสียงปกติ<br><br>**กรณี** 2: เซกเมนต์นั้นเป็นส่วนหนึ่งของเซกเมนต์ข้างเคียง |
| เซกเมนต์ที่ขาดหายไป | **กรณี** 1: เซกเมนต์ของหน่วยเสียง หายไป ซึ่งอาจเกิดจากพูดเร็ว หรือสั้นเกินไปจนไม่สามารถแบ่งเซกเมนต์ได้<br><br>**กรณี** 2: เซกเมนต์นั้น ประกอบด้วยเซกเมนต์อื่นอยู่ภายใน 2 เซกเมนต์ขึ้นไป |

กระบวนการในการค้นหาลำดับของคำนั้น จะกำหนดวัตถุที่บันทึกสถานะของการค้นหา หรือ $o_l$ ซึ่งวัตถุนี้จะบันทึก สถานะในปัจจุบัน คะแนนรวม และข้อมูลของเส้นทางที่ค้นหาผ่านมา ดังนี้

- สถานะต่อไปที่จะเดินทางต่อ ซึ่งจะเขียนแทนด้วย $(s_i, (u_n, w_m)_j)$ นั่นคือ สถานะต่อไปนั้น จะพิจาณาที่สถานะของเซกเมนต์ $s_i$ ของโครงข่ายหน่วยเสียง และสถานะที่ $(u_n, w_m)_j$ ของโครงข่ายการออกเสียงร่วมกัน

- คะแนนรวมของวัตถุ $o_l$ จะเป็นผลคูณของคะแนนที่คำนวณได้ตั้งแต่เริ่มต้น จึงถึงสถานะปัจจุบัน ตามสมมุติฐานของแต่ละเซกเมนต์ต่างๆ ที่ได้เดินทางผ่านมา (หรือ ผลบวกในกรณีที่คำนวณค่าคะแนนในลักษณะของค่าล็อคการิทึมของค่า

ความน่าจะเป็น) ซึ่งแต่ละวัตถุ $o_l$ นั้นจะมีลำดับของการเดินทาง หรือลำดับของสมมุติฐานของเซกเมนต์ต่างๆ ที่แตกต่างกัน

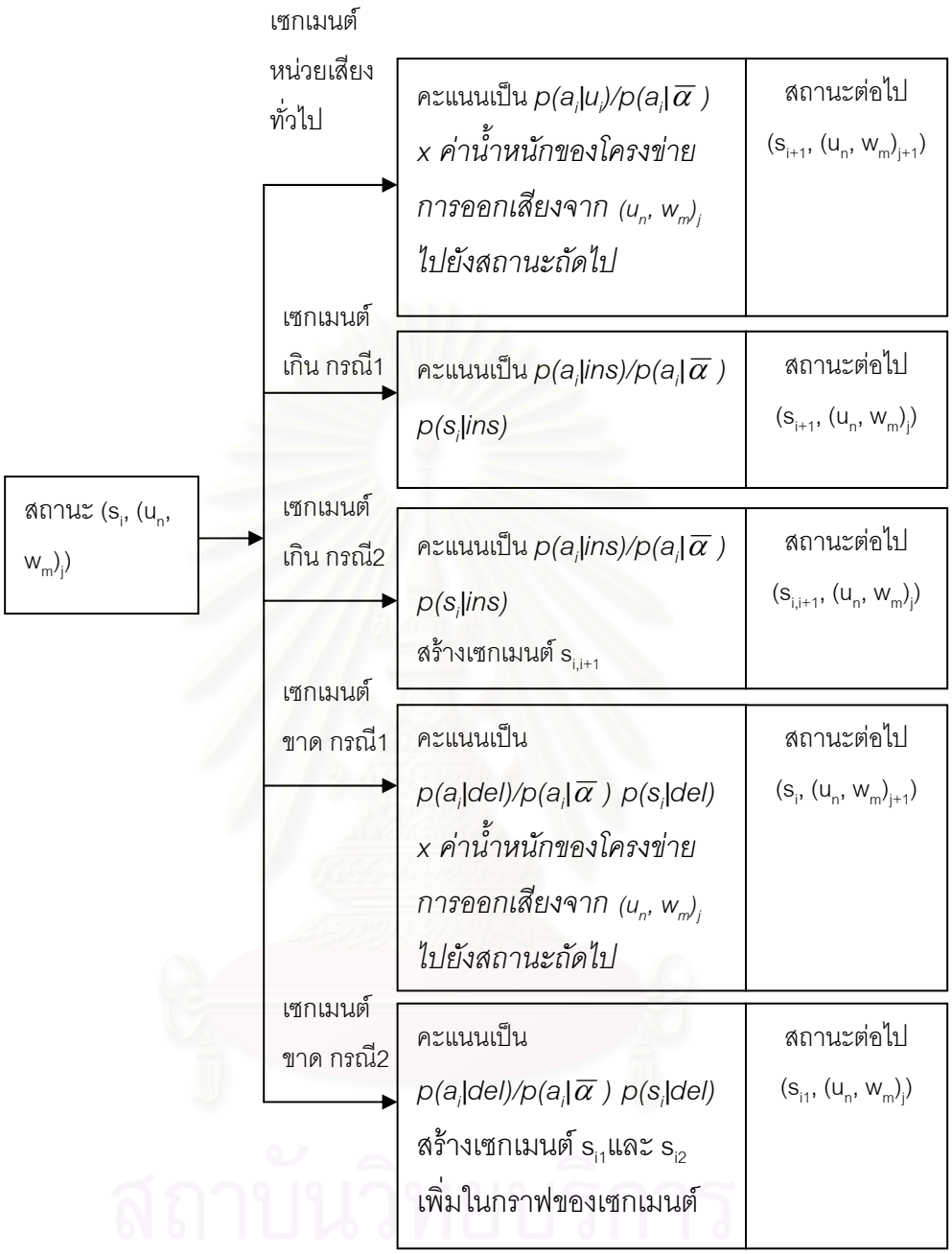
- ข้อมูลของเส้นทางที่ค้นหาผ่านมา ซึ่งก็คือ ข้อมูลของสถานะที่เดินทางตั้งแต่เริ่มต้น จึงถึงสถานะในปัจจุบัน รวมถึงลำดับของคำที่เดินทางผ่านมาด้วย

ในการค้นหาลำดับของคำนั้น จะมีกองซ้อน (Stack) ที่เก็บวัตถุ $o_l$ โดยกองซ้อนนี้จะเลือก $o_l$ ที่มีคะแนนมากที่สุดมาพิจาณาก่อน เมื่อวัตถุ $o_l$ เดินทางมาถึงสถานะ $(s_i, (u_n, w_m)_j)$ ก็จะทำการคำนวณคะแนนของสถานะ $(s_i, (u_n, w_m)_j)$ ตามสมมุติฐานต่างๆ ในตารางที่ 3-3 หรือสามารถพิจารณาได้จากรูปที่ 3-18

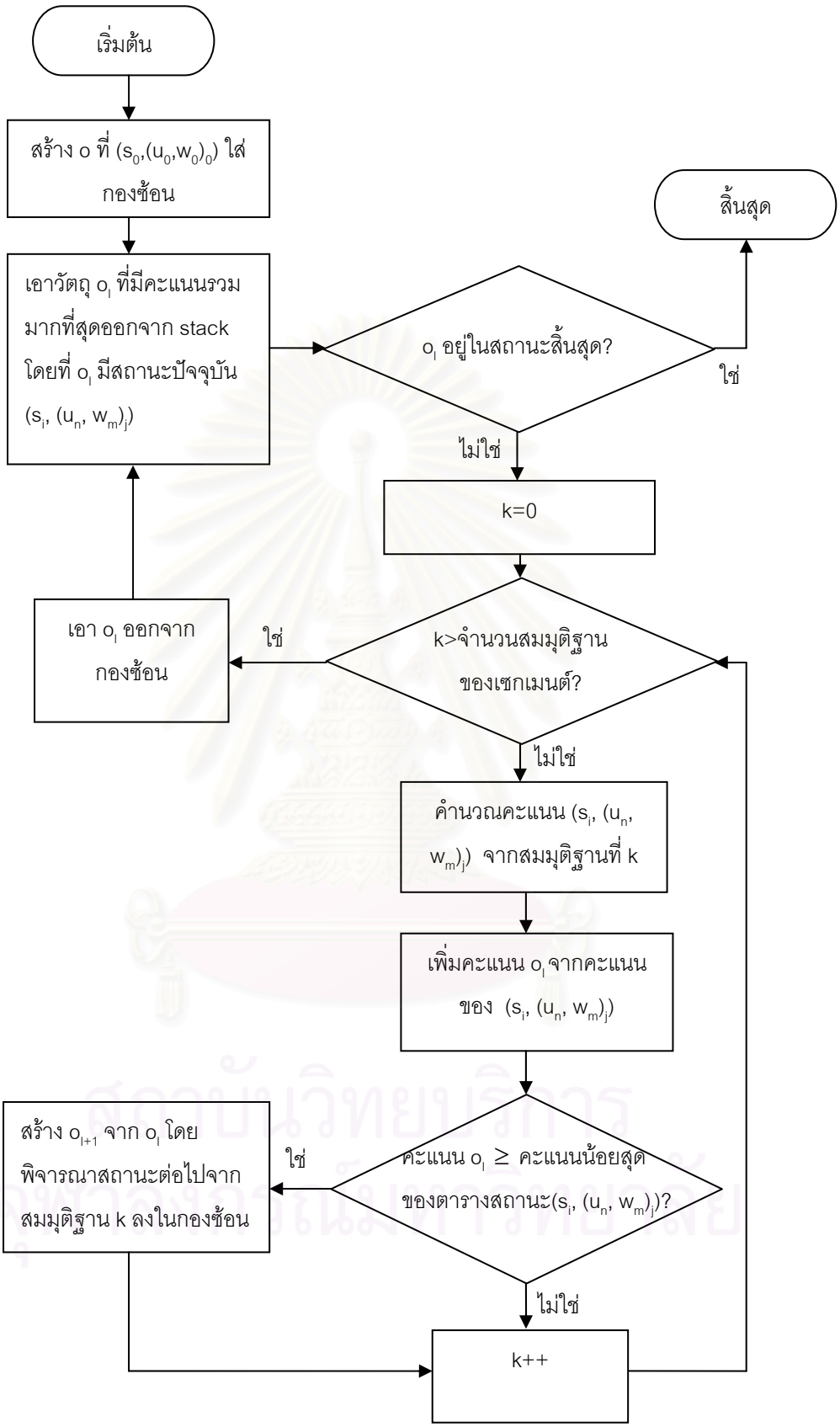
เมื่อทำการคำนวณค่าคะแนนของสถานะ $(s_i, (u_n, w_m)_j)$ แล้วจะทำการบันทึกผลรวมของคะแนน $(s_i, (u_n, w_m)_j)$ ที่ได้และคะแนนรวมของ $o_l$ ลงในตารางขนาด N มิติของของแต่ละสถานะ $(s_i, (u_n, w_m)_j)$ ซึ่งตารางนี้จะบันทึกคะแนนของเส้นทางจากสถานะเริ่มต้นจนถึง$(s_i, (u_n, w_m)_j)$ โดยที่มีการกำหนดสมมุติฐานของแต่ละเซกเมนต์ที่แตกต่างกันที่ดีที่สุด N อันดับแรก หากมีวัตถุอื่นที่เดินทางมาถึงสถานะ $(s_i, (u_n, w_m)_j)$ และมีคะแนนรวมน้อยกว่าค่าคะแนนใน N อันดับของตารางนี้ก็จะเอาวัตถุนั้นออกจากกองซ้อน และไม่นำมาค้นหาต่อ เรียกตารางนี้ว่า “ตารางสถานะ $(s_i, (u_n, w_m)_j)$” แต่หากมีคะแนนรวมมากกว่าก็จะทำการค้นหาต่อไป สามารถแสดงได้ในรูปที่ 3-19

ในการพิจารณาจุดสิ้นสุดของการค้นหานั้น จะพิจารณาเมื่อพบวัตถุ $o_l$ ใดๆ ที่อยู่ในสถานะสิ้นสุด คือ อยู่ในสถานะสุดท้าย ทั้งจากโครงข่ายหน่วยเสียง และโครงข่ายการออกเสียง โดย $o_l$ ที่สิ้นสุดตัวแรกที่พบนั้นจากกองซ้อนนั้น จะถือว่าเป็นคำตอบของการรู้จำคำที่ได้ โดยผลลัพธ์นั้นจะดูจากสถานะที่ทำการบันทึกในวัตถุ $o_l$ นั้น โดยที่คะแนนรวมของ $o_l$ ที่ได้ จะคือ ค่า $p(W|A)$ ในสมการที่ (3.10) และ สมการที่ (3.13) ถึงสมการที่ (3.16) นั่นเอง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย




รูปที่ 3-18 ลำดับการกระบวนการในการค้นหาลำดับของคำจากกราฟของเซกเมนต์




รูปที่ 3-19 ค้นหาลำดับของคำจากโครงข่ายหน่วยเสียง และโครงข่ายการออกเสียง

# บทที่ 4

## การค้นหาแนวคิด และจุดมุ่งหมาย

ในบทนี้จะกล่าวถึงวิธีการในการค้นหาลำดับของแนวคิด ซึ่งได้อธิบายลักษณะ และความหมายของแนวคิดที่ใช้ในวิทยานิพนธ์นี้ โดยได้แสดงกระบวนการในการค้นหาแนวคิดจากผลลัพธ์ของคำที่รู้จำได้ และการค้นหาแนวคิดจากกราฟของเซกเมนต์ นอกจากนี้ ยังได้กล่าวถึงวิธีการในการค้นหาจุดมุ่งหมายของประโยคจากแนวคิดที่ได้อีกด้วย

### 4.1 ความหมายของแนวคิด

แนวคิด (Concept) ของคำศัพท์ใดๆ นั้น จะหมายถึงความหมาย และลักษณะของการใช้งานของคำศัพท์นั้น โดยที่คำศัพท์ที่มีแนวคิดที่เหมือนกันจะรวมกันและเรียกว่า "*กลุ่มของแนวคิด*" ในงานวิจัยของ Wuttiwiwatchai และ Furui [Wutiwiwatchai and Furui, 2006] ได้เสนอแนวทางในการค้นหาลำดับของแนวคิด โดยได้กำหนดเป็นคู่ของแนวคิด และเลเบลของคำ (Keyword label) หรือแสดงได้เป็น c:l และทำการค้นหาพิจารณา c:l ให้สำหรับแต่ละคำ เมื่อกำหนดลำดับของคำมาให้ โดยในวิทยานิพนธ์นี้จะมีการเปลี่ยนแปลงสัญลักษณะของ c:l เล็กน้อย ด้วยการแทนให้อยู่ในรูปของคู่ลำดับ (l,c) ซึ่งในวิทยานิพนธ์นี้จะใช้ว่า *(ความหมาย, หน้าที่ของแนวคิด)* และเรียกคู่ลำดับดังกล่าวว่า "แนวคิด" โดยเป้าหมายในการค้นหาลำดับของแนวคิดนั้น คือการค้นหาคู่ลำดับของ *(ความหมาย, หน้าที่ของแนวคิด)* ที่ดีที่สุดนั่นเอง ข้อดีของการแสดงแนวคิดในลักษณะนี้ ก็เนื่องจากช่วยในการแบ่งแนวคิดให้มีความละเอียดมากขึ้นตัวอย่างเช่นตารางที่ 4-1

ตารางที่ 4-1 การแบ่งพิจารณาประเภทของแนวคิด

| คำ | สวัสดี | ค่ะ | จะ | โทรศัพท์ | ไป | ประเทศ |
|---|---|---|---|---|---|---|
| ความหมาย | l_greeting | l_gar | l_gar | l_phone | l_gar | l_cnt |
| หน้าที่ของแนวคิด | c_greeting | c_gar | c_gar | c_phone | c_gar | c _cnt_name |

| คำ | จีน | รหัส | ประเทศ | อะไร | คะ |
|---|---|---|---|---|---|
| ความหมาย | l_cnt_name | l_code | l_cnt | l_C_qw | l_gar |
| หน้าที่ของแนวคิด | c_cnt_name | c_code | c_code | c_C_qw | c_gar |

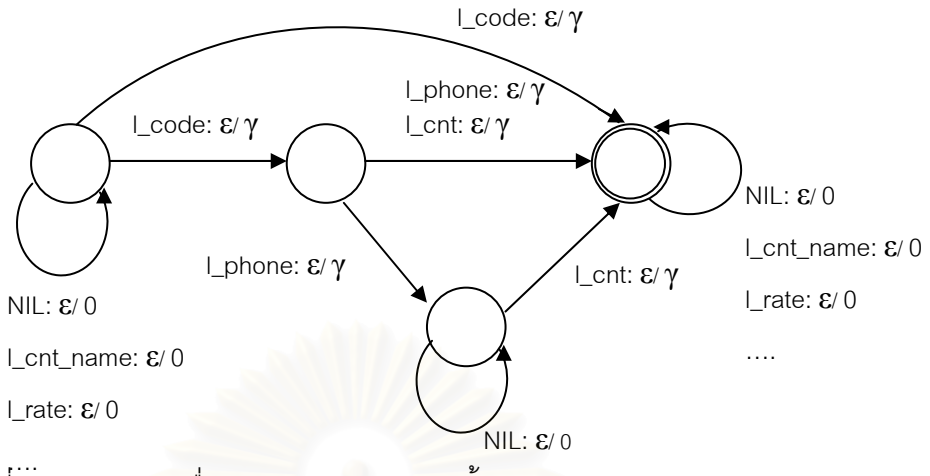
ในตารางที่ 4-1 นั้น จะแสดงการใส่แนวคิดของประโยคว่า “สวัสดีค่ะโทรไปประเทศจีนรหัสประเทศอะไรคะ” ซึ่งแนวคิดของแต่ละคำศัพท์นั้น จะเป็นคู่ลำดับของความหมายและหน้าที่ของแนวคิดของคำศัพท์นั้น โดยประเภทของความหมายของคำนั้น จะมี “l” อยู่ข้างหน้า และหน้าที่ของแนวคิดนั้น จะเขียนแทนโดยการเขียน “c” อยู่ข้างหน้า เช่น คำว่า “สวัสดี” นั้น จะมีแนวคิดเป็น (l_greeting, c_greeting) นั่นคือ คำว่า “สวัสดี” นั้น มีความหมายว่าเป็นการทักทาย(l_greeting) และอยู่ในแนวคิดประเภทของกลุ่มคำทักทาย (c_greeting) อีกตัวอย่างหนึ่ง คือคำว่า “ประเทศ” ซึ่งมีความหมายในกลุ่มของคำบอกถึงเมือง (l_cnt) แต่จะเห็นว่าเมื่อคำว่า “ประเทศ” นั้น หากอยู่ก่อนหน้าคำว่า “จีน” จะทำหน้าที่เป็นคำนำหน้าชื่อประเทศ และอยู่ในแนวคิดของชื่อประเทศ (c_cnt_name) ซึ่งจะมีแนวคิดเป็น (l_cnt, c_cnt_name) แต่เมื่อคำว่า “ประเทศ” นั้น ต่อหลังคำว่า “รหัส” ก็จะจัดอยู่ในประเภทคำที่ใช้ขยายกลุ่มคำศัพท์ที่มีความหมายว่ารหัสโทรศัพท์ หรือมีหน้าที่ของแนวคิดอยู่ในกลุ่มของรหัสโทรศัพท์ (c_code) และมีแนวคิดเป็น (l_cnt, c_code) โดยจะเห็นว่าหากแบ่งประเภทของแนวคิดจากเพียงความหมายของคำเพียงอย่างเดียว คำว่า “ประเทศ” ที่อยู่ก่อนหน้าคำว่า “จีน” และอยู่หลังคำว่า “รหัส” นั้น จะมีแนวคิดที่เหมือนกัน ทั้งที่ความหมายและหน้าที่ที่แท้จริงนั้นแตกต่างกัน

นอกจากนี้ ยังมีแนวคิดขยะ (Garbage) ซึ่งจะหมายถึงคำที่ปรากฏขึ้นนอกเหนือจากแนวคิดอื่นๆ รวมถึงคำที่ไม่ใช้ในการแปลความประโยคด้วย โดยแนวคิดประเภทนี้จะเขียนแทนด้วย (l_gar, c_gar) โดยในตารางที่ 4-1 จะเห็นว่า คำว่า “ค่ะ” “จะ” “ไป” และ “คะ” จัดอยู่ในประเภทของแนวคิดขยะ

ในบางแนวคิดนั้น จะมีค่า (Value) ของแนวคิดนั้นด้วย เช่น แนวคิด c_cnt ซึ่งมีค่าของแนวคิดเป็นชื่อประเทศ โดยจากตัวอย่างในตารางที่ 4-1 ค่าของแนวคิด c_cnt ก็คือ “จีน” โดยที่ตัวอย่างของแนวคิดอื่นๆ ที่ไม่มีค่าของแนวคิด เช่น c_code (แนวคิดของรหัสโทรศัพท์), c_greeting (แนวคิดที่เกี่ยวกับคำทักทาย) และ c_C_qw (แนวคิดเกี่ยวกับคำถามที่ใช้ถามรหัสโทรศัพท์) เป็นต้น

## 4.2 การค้นหาแนวคิดจากผลลัพธ์จากการรู้จำ

ในการค้นหาแนวคิดนั้น ได้นักวิจัยบางกลุ่มเสนอแนวทางในการค้นหาแนวคิดจากลำดับของคำที่ได้จากเครื่องรู้จำ โดยในงานวิจัยของ Wuttiwiwatchai และ Furui [Wutiwiwatchai and Furui, 2006] นั้นได้เสนอแนวทางในการค้นหาแนวคิดจากผลลัพธ์ที่ได้จากการรู้จำเสียงพูด โดยได้ใช้วิธีทางสถิติ ร่วมกับการใช้กฎทางไวยากรณ์ โดยที่การใช้กฎไวยากรณ์นั้น จะใช้เครื่องจักรสถานะแบบถ่วงน้ำหนักสำหรับแต่ละหน้าที่ของแนวคิด (Reg) ซึ่งจะแสดงลำดับของคำที่เป็นไปได้ของแนวคิดนั้น ดังตัวอย่างในรูปที่ 4-1



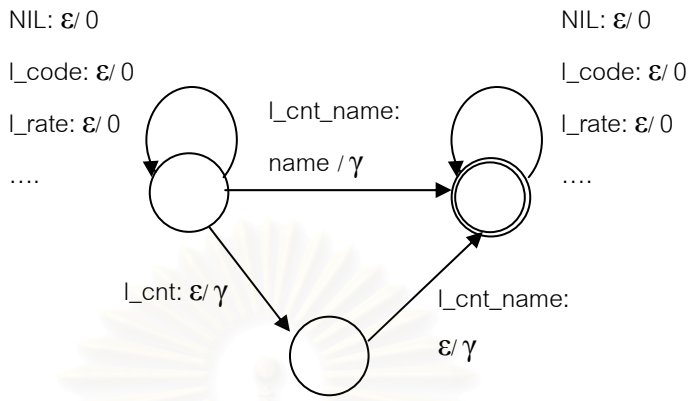

รูปที่ 4-1 ตัวอย่างเครื่องจักรสถานะแบบถ่วงน้ำหนัก (Reg) ของแนวคิด c_code

รูปที่ 4-1 นั้น จะเป็นเครื่องจักรสถานะแบบถ่วงน้ำหนักของแนวคิด c_code ซึ่งเป็นแนวคิดที่บอกถึงรหัสโทรศัพท์ระหว่างประเทศ โดยที่ข้อความที่บ่งบอกบนเส้นเชื่อมแต่ละเส้นนั้นจะแสดงในลักษณะของ “อินพุต: เอาท์พุต/ ค่าน้ำหนัก” ซึ่งอินพุตนั้นจะเป็นความหมายของคำ (l) สำหรับ NIL นั้นจะแทนกลุ่มคำที่อยู่นอกเหนือจากแนวคิดนั้น เอาท์พุตจะเป็นคำสำคัญที่จะใช้ในการบอกค่า (Value) ของแนวคิดนั้น โดยเนื่องจากแนวคิด c_code นี้ จะไม่มีค่าของแนวคิด ดังนั้นค่าเอาท์พุตจึงเป็น ε สำหรับทุกเส้นเชื่อม สำหรับค่าน้ำหนักนั้น จะกำหนดให้เป็นค่าคงที่ โดยในงานวิทยานิพนธ์นี้ กำหนดให้ค่าน้ำหนักของอินพุตที่เป็น NIL หรือกลุ่มคำขยะนั้นจะมีค่าเป็น 0 และค่า γ เป็น 1 โดยที่เครื่องจักรสถานะนี้จะสามารถแจกแจง (Parse) ได้สำเร็จนั้น จะต้องสามารถเริ่มจากสถานะแรก ไปยังสถานะสุดท้ายได้

ในรูปที่ 4-1 นั้น l_code จะหมายถึง กลุ่มคำที่มีความหมายถึงรหัสที่ใช้ในการโทรศัพท์ เช่น คำว่า “รหัส” “โค๊ด” “หมายเลข” เป็นต้น l_phone จะเป็นกลุ่มคำที่มีความหมายถึงการโทรศัพท์ เช่น “โทรไป” “โทรศัพท์” เป็นต้น l_cnt จะหมายถึง กลุ่มคำที่มีความหมายเกี่ยวกับเมือง เช่น คำว่า “ประเทศ”, “เมือง”, “ต่างประเทศ” เป็นต้น ตัวอย่างคำที่สามารถทำการแจกแจงในเครื่องจักรสถานะแบบถ่วงน้ำหนักในรูปที่ 4-1 ได้ เช่น “รหัสโทรศัพท์”, “รหัสประเทศ”, “รหัสต่างประเทศ”, “โค๊ด”, “รหัสโทรศัพท์ระหว่างประเทศ” “รหัสโทรไปต่างประเทศ”เป็นต้น (กำหนดให้คำว่า “ระหว่าง” นั้นจัดอยู่ในประเภทแนวคิดขยะ หรือ NIL)

ตัวอย่างของเครื่องจักรสถานะแบบถ่วงน้ำหนักของแนวคิดที่มีค่า (Value) เช่น แนวคิดที่ระบุประเทศ (c_cnt) ซึ่งค่าของแนวคิดประเภทนี้ก็คือ คำที่บอกถึงชื่อประเทศนั่นเอง เช่น คำว่า “ประเทศญี่ปุ่น” “สาธารณรัฐจีน” “เมืองอินเดีย” เป็นต้น ซึ่งคำเหล่านี้จัดอยู่ในแนวคิดที่ระบุ

ประเทศ โดยที่มีค่าของแนวคิดเป็น “ญี่ปุ่น” “จีน” และ “อินเดีย” ตามลำดับ โดยตัวอย่างของเครื่องจักรสถานะของแนวคิด c_cnt นี้ สามารถแสดงได้ในรูปที่ 4-2




รูปที่ 4-2 ตัวอย่างเครื่องจักรสถานะแบบถ่วงน้ำหนัก (Reg) ของแนวคิด c_cnt

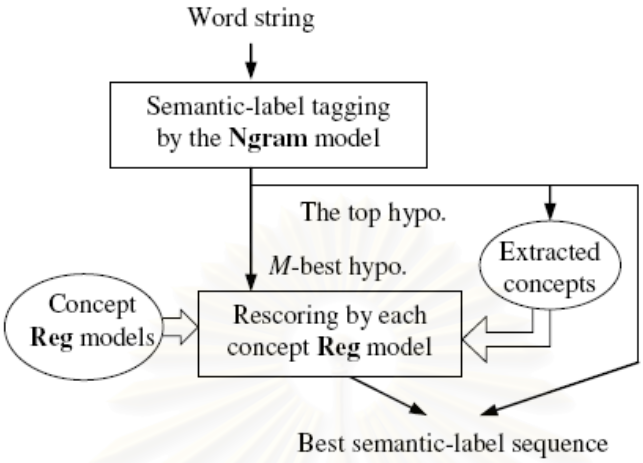
โดยในรูปที่ 4-2 จะเห็นว่า ในข้อความที่บ่งบอกบนเส้นเชื่อมที่ระบุว่า “l_cnt_name: name/ γ” เอาท์พุตของเส้นเชื่อมนี้จะเป็นการระบุว่า อินพุตที่เข้ามานั้น จะมีค่า (Value) ประเภท “name” นั่นเอง โดยค่าของแนวคิดนี้ จะนำไปใช้ในภายหลังต่อไป เช่น ใช้ในการพิจารณาการตอบกลับ หรือ ใช้เป็นข้อมูลในการค้นหาข้อมูล เป็นต้น

นอกจากจะพิจารณาลำดับของแนวคิดด้วยการใช้เครื่องจักรสถานะแบบถ่วงน้ำหนักที่สร้างจากกฎทางไวยากรณ์ (Reg) แล้ว ยังใช้แบบจำลองค่าความน่าจะเป็นในการพิจารณาอีกด้วย โดยจะแสดงแบบจำลองนี้ในลักษณะของเครื่องจักรสถานะแบบถ่วงน้ำหนัก หรือในที่นี้จะแทนว่า Ngram โดยกำหนดให้ $l_i$=c:l และ $L$ คือลำดับของ $l_i$ หรือ $L=(l_1,l_2,...,l_n)$ เมื่อกำหนดลำดับของคำ $W$ มาให้ ซึ่งแบบจำลองค่าความน่าจะเป็นนี้ จะทำการค้นหาลำดับของแนวคิด $L$ ที่มีค่า $p(W,L)$ มากที่สุด ดังในสมการที่ (4.1)

$$\begin{aligned}\widetilde{L} &= \arg\max_{L} p(W,L) \\ &\approx \arg\max_{L} \prod_{t=1}^{T} p(w_t,l_t \mid w_{t-1},l_{t-1})\end{aligned} \tag{4.1}$$

ในงานวิจัยของ Wutiwiwatchai และ Furui นั้น ได้นำเอาการใช้แบบจำลอง Reg และ Ngram มาใช้ร่วมกัน เพื่อใช้ในการพิจารณาค้นหาลำดับของ $L$ ที่ดีที่สุด โดยในเริ่มแรกนั้นจะใช้แบบจำลองความน่าจะเป็น (Ngram) ในการหาลำดับของคู่ลำดับแนวคิดที่ดีที่สุดจำนวน M อันดับแรก จากนั้นจะทำการคำนวณคะแนนของลำดับแนวคิดสมมุติฐานจำนวน M อันดับนี้ใหม่ ด้วยการใช้คะแนนจากเครื่องจักรสถานะแบบถ่วงน้ำหนักของกฎไวยากรณ์ของแนวคิด (Reg) โดยจะพิจารณาเครื่องจักรสถานะ Reg ตามแนวคิดที่ปรากฏอยู่ในลำดับแนวคิดที่ดีที่สุดเป็นอันดับแรก

จากใน M อันดับ และลำดับของแนวคิดที่จะถูกคำนวณคะแนนใหม่ ก็ต่อเมื่อสามารถแจกแจงเครื่องจักรสถานะนั้นได้ โดยขั้นตอนการทำงานจะแสดงในรูปที่ 4-3




รูปที่ 4-3 การค้นหาลำดับแนวคิดเสนอโดย Wutiwiwatchai และ Furui (Wutiwiwatchai and Furui, 2006)

โดยในการคำนวณค่าคะแนนใหม่ของลำดับแนวคิดสมมุติฐาน M อันดับแรกนั้น (เมื่อลำดับของแนวคิดนั้น สามารถแจกแจงบนเครื่องจักรสถานะของแนวคิดที่ปรากฏอยู่ในลำดับแนวคิดที่ดีที่สุดใน M อันดับได้) จะพิจารณาจากสมการ

$$Score = \sum_{t=1}^{T} (\log p(w_t, l_t \mid w_{t-1}, l_{t-1}) + \lambda_t)$$

โดยที่ $\lambda_t \in \{\lambda, 0\}$

### 4.3 การค้นหาแนวคิดโดยใช้แบบจำลองแนวคิด

ในการค้นหาลำดับของแนวคิดจากผลลัพธ์ของการรู้จำคำของเครื่องรู้จำนั้น จะเห็นว่าทั้งในการรู้จำเสียง และการค้นหาแนวคิดนั้น ได้ใช้ความรู้ที่แยกออกจากกัน ดังนั้นในงานวิจัยนี้จึงได้เสนอแนวทางในการใช้ความรู้ในส่วนของหน่วยเสียง (Acoustic-Phonetic) และส่วนของคำศัพท์ (Pronunciation) ที่ได้ใช้ในการรู้จำเสียงพูด ร่วมกับความรู้ทางด้านแนวคิด ทั้งนี้เพื่อให้ในการรู้จำเสียงนั้น ก็ได้ใช้ความรู้ด้านแนวคิดมาช่วยในการพิจารณาลำดับของคำ ในขณะเดียวกันในการพิจารณาลำดับของแนวคิด ก็จะนำความรู้ทางด้านหน่วยเสียง คำศัพท์ และภาษามาใช้ด้วย โดยที่เนื่องจากแนวคิดนั้น จะแบ่งตามความหมายและลักษณะของการใช้งานของคำศัพท์ ดังนั้น เมื่อนำความรู้ของแนวคิดมาช่วยในการรู้จำคำจากเสียงพูด ก็เหมือนเป็นการเพิ่มความรู้ในเรื่องของความหมายของคำศัพท์ในการพิจารณาค้นหาลำดับของคำด้วย

### 4.3.1 การค้นหาแนวคิดจากกราฟของเซกเมนต์

ในการค้นหาลำดับของแนวคิดจากกราฟของเซกเมนต์นั้น จะเป็นการค้นหาลำดับของแนวคิดที่มีค่าความน่าจะเป็นมากที่สุดเมื่อกำหนดลำดับของค่าสังเกตมาให้ โดยสามารถแสดงได้ดังในสมการที่ (4.2)

$$C^* = \arg\max_{C} p(C \mid A) \tag{4.2}$$

โดยที่ในสมการที่ (4.2) นั้น $C$ จะแทนลำดับของแนวคิด หรือ $C = (c_1, c_2, ..., c_k)$ และ $A$ จะแทนลำดับของค่าสังเกตที่ได้จากเวกเตอร์ลักษณะสำคัญของแต่ละเซกเมนต์ โดยที่ $C^*$ จะแทนลำดับของแนวคิดที่มีค่า $p(C|A)$ มากที่สุด ค่าของ $p(C|A)$ สามารถคำนวณได้ดังนี้

$$p(C,W,U,S \mid A) = \frac{p(A,S,U \mid W,C)\,p(W \mid C)\,p(C)}{p(A)} \tag{4.3}$$

ในสมการที่ (4.3) นั้น เมื่อพิจารณาพจน์ $p(A,S,U|W,C)$ โดยลำดับของค่าสังเกต ลำดับของเซกเมนต์ และลำดับของหน่วยเสียงนั้น จะเป็นอิสระต่อลำดับของแนวคิด ดังนั้นจึงสามารถแทนค่าพจน์ $p(A,S,U|W,C)$ ในสมการที่ (4.3) ใหม่ ได้เป็น $p(A,S,U|W)$ ซึ่งเมื่ออ้างอิงจากสมการที่ (3-10) จะสามารถเขียนสมการที่ (4.3) ใหม่ได้ว่า

$$p(C,W,U,S \mid A) = \frac{\prod_{i=1}^{n} \frac{p(m_i \mid u_i)}{p(m_i \mid \overline{\alpha}_{\max})} p(s_i \mid u_i)\,p(U \mid W)\,p(W \mid C)\,p(C)}{p(A)} \tag{4.4}$$

หรือ สามารถเขียนสมการที่ (4.2) ได้ว่า

$$C^* = \arg\max_{S,U,W,C} \prod_{i=1}^{n} \frac{p(m_i \mid u_i)}{p(m_i \mid \overline{\alpha}_{\max})} p(s_i \mid u_i)\,p(U \mid W)\,p(W \mid C)\,p(C) \tag{4.5}$$

โดยที่ในการคำนวณนั้น ยังคงใช้แบบจำลองหน่วยเสียงในการคำนวณค่าของ $p(m_i|u_i)$ และ $p(m_i|\overline{\alpha}_{\max})$ (หัวข้อที่ 3.5.1)รวมถึงใช้แบบจำลองระยะเวลาในการคำนวณ $p(s_i|u_i)$ (หัวข้อที่ 3.5.2) และใช้โครงข่ายจับคู่ระหว่างคำศัพท์และหน่วยเสียงในการคำนวณพจน์ $p(U|W)$ (หัวข้อที่ 3.5.3) สิ่งที่เปลี่ยนแปลงไปจากสมการที่ (3.10) คือ พจน์ $p(W|C)$ และ$p(C)$

สำหรับค่าของ $p(W|C)$ หรือค่าความน่าจะเป็นที่จะเกิดลำดับคำ $W$ เมื่อเกิดลำดับของแนวคิดเป็น $C$ โดยในการคำนวณนั้น จะสามารถคำนวณได้ดังนี้

$$p(W \mid C) = p(w_0 \mid c_0)\prod_{l=1}^{t} p(w_l \mid w_{l-1}, c_l)$$

$$= p(w_0 \mid c_0)\prod_{l=1}^{t} \frac{p(w_l \mid w_{l-1})p(w_l \mid c_l)}{p(w_l)} \qquad (4.6)$$

$$= p(w_0 \mid c_0)\prod_{l=1}^{t} p(w_l \mid w_{l-1})\prod_{l=1}^{t} \frac{p(w_l \mid c_l)}{p(w_l)}$$

จากสมการที่ (4.6) กำหนดให้ $w_l$ คือ คำที่ตำแหน่งที่ $l$ และ $W = (w_1, w_2, ..., w_t)$ จะเห็นว่าค่าความน่าจะเป็นที่จะเกิด $w_l$ ที่ตำแหน่ง $l$ ได้นั้น จะขึ้นอยู่กับคำที่อยู่ก่อนหน้า (ในกรณีกำหนดให้พิจารณาในลักษณะของไบแกรม) และแนวคิดที่เกิดขึ้นที่ตำแหน่ง $l$ หรือ $c_l$ โดยที่ค่า $p(w_1|w_0, c_1)$ นั้น จะเป็นค่าความน่าจะเป็นที่คำศัพท์คำแรกนั้นเริ่มต้นที่ $w_1$ และมีแนวคิดเป็น $c_1$ และค่า $p(w_0|c_0)$ มีค่าเป็น 1 โดยในที่นี้ได้กำหนดให้คำทุกคำที่ปรากฏขึ้นจากการรู้จำนั้น จะต้องสามารถจัดอยู่ประเภทของแนวคิดใดแนวคิดหนึ่งได้ โดยที่คำที่ไม่อยู่ในพจนานุกรมนั้น จะถือว่าอยู่ในประเภทแนวคิดขยะ

สำหรับผลคูณของ $p(w_l|w_{l-1})$ ในสมการที่ (4.6) นั้น จะสามารถคำนวณได้จากแบบจำลองภาษา (หัวข้อที่ 3.5.4) ในการคำนวณผลคูณของ $p(w_l|c_l)/p(w_l)$ นั้น จะคำนวณโดยใช้โครงข่ายที่ใช้ในการจับคู่คำศัพท์กับแนวคิดของคำศัพท์นั้น ซึ่งเป็นโครงข่ายที่แสดงว่า แต่ละแนวคิดนั้นจะมีคำศัพท์คำใดบ้าง และมีค่าความน่าจะเป็นในการเกิดคำศัพท์เหล่านั้นเท่าไหร่

สำหรับพจน์ $p(C)$ ในสมการที่ (4.5) นั้น จะเป็นค่าความน่าจะเป็นที่จะเกิดลำดับของแนวคิด $C$ ขึ้น โดยสามารถเขียนสมการได้ดังนี้

$$p(C) = \prod_{i=1}^{j} p(c_i \mid c_{i-n+1}.....c_{i-1}) \qquad (4.7)$$

โดยหากพิจารณาให้ความสัมพันธ์ของลำดับแนวคิดนั้น มีลักษณะเป็นไบแกรม สมการที่ (4.7) จะสามารถเขียนได้เป็น

$$p(C) = \prod_{i=1}^{j} p(c_i \mid c_{i-1}) \qquad (4.8)$$

โดยในการคำนวณ $p(C)$ นั้นจะใช้แบบจำลองแนวคิด (Concept model) ในการคำนวณ โดยแบบจำลองนี้จะบอกลำดับของแนวคิดที่สามารถเกิดได้ รวมถึงค่าความน่าจะเป็นที่จะเกิดลำดับของแนวคิดใดๆ ด้วย

### 4.3.2 โครงข่ายการจับคู่คำศัพท์กับแนวคิดของคำศัพท์

สำหรับโครงข่ายการจับคู่คำศัพท์กับแนวคิดของคำศัพท์นั้น จะเป็นโครงข่ายที่แสดงว่าแต่ละคำศัพท์นั้น จะสามารถจัดอยู่ในประเภทใดได้บ้าง รวมถึงอัตราส่วนของค่าความน่าจะเป็นที่จะ

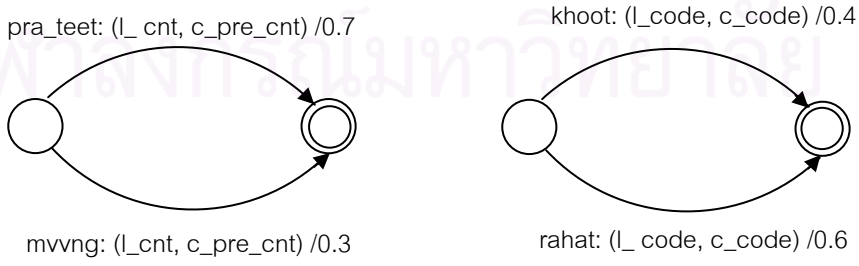
เกิดคำศัพท์นั้นขึ้นจากแนวคิดที่กำหนดต่อค่าความน่าจะเป็นที่จะเกิดคำศัพท์นั้น หรือ $p(w_l|c_l)/p(w_l)$ ซึ่งค่าความน่าจะเป็นนี้จะนำมาใช้ร่วมกับค่า $p(W)$ ที่ได้จากแบบจำลองภาษาในหัวข้อที่ 4.3.1 เพื่อนำมาใช้ในการคำนวณ $p(W|C)$ ในสมการที่ (4.5) โดยเริ่มแรก จะทำการแบ่งประเภทของคำศัพท์ตามความหมาย และหน้าที่ของแนวคิด เพื่อดูว่าแต่ละแนวคิดนั้นมีคำศัพท์อะไรอยู่บ้าง หลังจากนั้นก็ทำการคำนวณค่าความน่าจะเป็นที่จะเกิดคำศัพท์แต่ละคำในแนวคิดนั้น โดยในการพิจารณาค่าความน่าจะเป็นนี้ จะพิจารณาจากประโยคตัวอย่าง โดยจะทำการแบ่งประเภทของแนวคิดให้สำหรับแต่ละคำในประโยคตัวอย่าง จากนั้นจะทำการคำนวณค่าความน่าจะเป็น โดยใช้สมการที่ (4.9)

$$p(w_l \mid c_k) = \frac{count(w_l, c_k)}{\sum_{all\, l} count(w_l, c_k)} \qquad (4.9)$$

โดยที่ $count(w_l, c_l)$ นั้น จะเป็นจำนวนที่คำ $w_l$ ถูกจัดอยู่ในประเภทของแนวคิด $c_k$ ดังนั้นในสมการที่ (4.9) จึงอาจกล่าวได้ว่า ค่าความน่าจะเป็นที่จะเกิดคำ $w_l$ ในแนวคิด $c_k$ นั้น จะเป็นอัตราส่วนของจำนวนที่คำ $w_l$ นั้นถูกจัดอยู่ในแนวคิด $c_k$ เทียบกับจำนวนที่เกิดแนวคิด $c_k$ ทั้งหมด

ในทำนองเดียวกัน การคำนวณค่าความน่าจะเป็นที่จะเกิดคำศัพท์ $w_l$ ใดๆ นั้น สามารถคำนวณได้จากอัตราส่วนของจำนวนที่ปรากฏคำ $w_l$ ต่อจำนวนคำทั้งหมดที่มีในข้อมูลการเรียนรู้

โครงข่ายการจับคู่คำศัพท์กับแนวคิดของคำศัพท์นั้น สามารถพิจารณาได้เป็นเครื่องจักรสถานะแบบถ่วงน้ำหนัก โดยในการสร้างนั้น จะสร้างเครื่องจักรสถานะสำหรับแต่ละแนวคิดแยกออกจากกัน โดยตัวอย่างของเครื่องจักรสถานะของแนวคิด (l_cnt, l_cnt_name) ซึ่งเป็นแนวคิดของกลุ่มคำที่มีความหมายเกี่ยวกับประเทศ และมีหน้าที่เป็นส่วนหนึ่งของชื่อประเทศ เช่น คำว่า “ประเทศ” และ “เมือง” ของคำว่า “ประเทศสวีเดน” และ “เมืองจีน” และเครื่องจักรสถานะของแนวคิด (l_code, c_code) ซึ่งเป็นแนวคิดของกลุ่มคำที่มีความหมาย และหน้าที่ที่บ่งบอกถึงรหัสโทรศัพท์ จะสามารถแสดงได้ในรูปที่ 4-4




รูปที่ 4-4 ตัวอย่างเครื่องจักรสถานะของแนวคิด (a) เครื่องจักรสถานะของแนวคิด (l_cnt, c_cnt_name) (b) เครื่องจักรสถานะของแนวคิด (l_code, c_code)

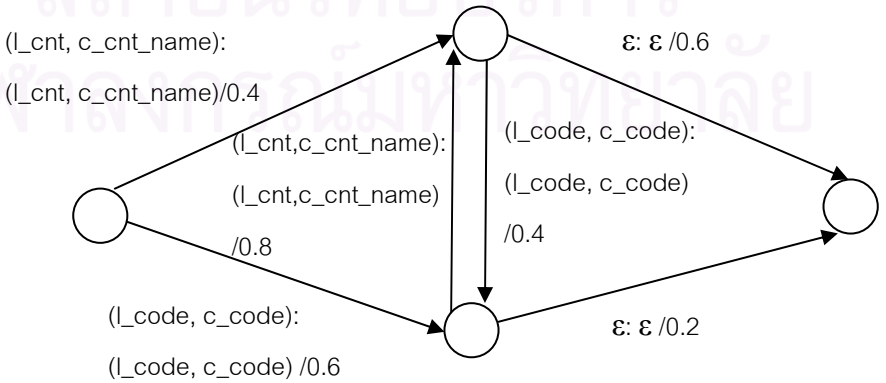
โดยในรูปที่ 4-4 นั้น จะเห็นว่าโครงข่ายสถานะนั้น จะรับอินพุตเป็นคำศัพท์ และให้เอาท์พุตเป็นแนวคิดของคำนั้น โดยที่ค่าน้ำหนักของเส้นเชื่อมนั้น จะเป็นผลหารของค่าความน่าจะเป็น $p(w_l|c_l)$ ซึ่งคำนวณได้จากสมการที่ (4.9) และค่า $p(w_l)$ เช่น ค่าน้ำหนักของเส้นเชื่อม “rahat: (l_code, c_code)/0.6” นั้น จะมาจาก $p(\text{“rahat”}|\ l_code,\ c_code))/p(\text{“rahat”})$ โดยที่ค่าน้ำหนักรวมของเส้นทางในโครงข่ายสถานะนี้ เป็นมีค่าเป็นผลคูณของน้ำหนักของเชื่อมที่เดินทางผ่าน ดังนี้

$$p(path) = \prod_{l=1}^{t} weight\ of\ edge_l$$
$$= \prod_{l=1}^{t} \frac{p(w_l \mid c_l)}{p(w_l)}$$

ในการสร้างโครงข่ายการจับคู่คำศัพท์กับแนวคิดของคำศัพท์นั้น จะนำเอาโครงข่ายสถานะของแต่ละคำศัพท์ทั้งหมดมาทำการรวมกัน (Union) ซึ่งจะทำให้ได้โครงข่ายที่ทำการจับคู่จากคำศัพท์ ไปยังแนวคิดของคำศัพท์นั้น

### 4.3.3 แบบจำลองแนวคิด

แบบจำลองแนวคิด (Concept Model) จะเป็นแบบจำลองที่บอกลักษณะของลำดับของแนวคิดที่เป็นไปได้ รวมถึงค่าความน่าจะเป็นของแต่ละลำดับแนวคิดด้วย ในการคำนวณค่า $p(C)$ ในสมการที่ (4.5) นั้น จะสามารถคำนวณได้จากแบบจำลองแนวคิดนี้ โดยจะแสดงในลักษณะของเครื่องจักรสถานะแบบถ่วงน้ำหนัก ซึ่งค่าน้ำหนักของแต่ละเส้นเชื่อมนั้น คือค่าความน่าจะเป็นของ $p(c_i|c_{i-1})$ เมื่อ $c_i$ เป็นสถานะที่ปลายทางของเส้นเชื่อม และ $c_{i-1}$ เป็นสถานะต้นทางของเส้นเชื่อม ดังรูปที่ 4-5




รูปที่ 4-5 ตัวอย่างแบบจำลองแนวคิด

ในรูปที่ 4-5 นั้น ได้แสดงตัวอย่างของแบบจำลองแนวคิด โดยกำหนดให้แบบจำลองนี้ประกอบด้วยแนวคิด 2 แนวคิด ได้แก่ (l_cnt, c_cnt_name) และ (l_code, c_code) ซึ่งน้ำหนักของเส้นเชื่อม คือ ค่าของ $p(c_i|c_{i-1})$ นั่นเอง โดยที่ค่าความน่าจะเป็นของเส้นทางในแบบจำลองแนวคิดนั้น จะเป็นผลคูณน้ำหนักของเส้นเชื่อมแต่ละเส้นของเส้นทางนั้น ดังนี้

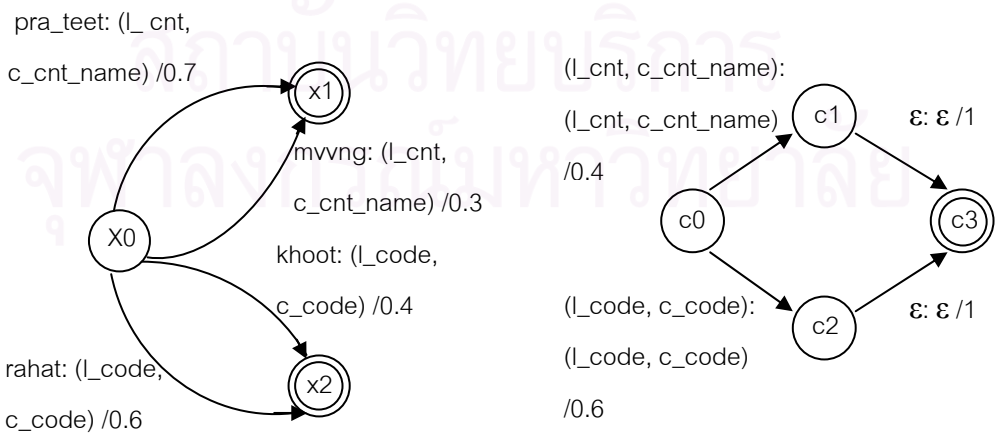
$$\begin{aligned} p(path) &= \prod_{i=1}^{n} weight\ of\ edge_i \\ &= \prod_{i=1}^{n} p(c_i \mid c_{i-1}) \\ &= p(C) \end{aligned}$$

### 4.3.4 การสร้างโครงข่ายการออกเสียง

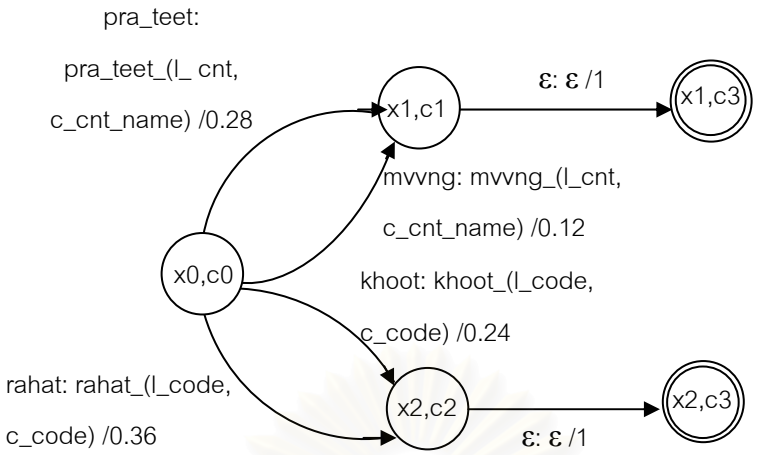
ในหัวข้อที่ 3.6 ได้กล่าวถึงการโครงข่ายการออกเสียงซึ่งใช้ร่วมกับโครงข่ายหน่วยเสียงเพื่อค้นหาลำดับของคำจากกราฟของเซกเมนต์ โดยในการค้นหาลำดับของแนวคิดจากกราฟของเซกเมนต์นั้น จะมีการเปลี่ยนแปลงโครงข่ายการออกเสียง ($U$) เป็น

$$U = CC \circ WC \circ G \circ L$$

โดยที่โครงข่าย $G$ ได้มาจากแบบจำลองภาษา ในหัวข้อที่ 3.5.4 และโครงข่าย $L$ นั้นเป็นโครงข่ายที่ทำการแปลลำดับของหน่วยเสียงให้อยู่ในรูปลำดับของคำศัพท์ ในหัวข้อที่ 3.6 สิ่งที่เพิ่มจากโครงข่ายคำศัพท์ในหัวข้อที่ 3.6 นั้น คือ การทำฟังก์ชันประกอบของโครงข่ายจับคู่คำศัพท์กับแนวคิดของคำศัพท์ $WC$ (หัวข้อที่ 4.3.2) และแบบจำลองแนวคิด $CC$ (หัวข้อที่ 4.3.3) โดยตัวอย่างของการประกอบกันระหว่างโครงข่าย $WC$ และแบบจำลอง CC จะแสดงในรูปที่ 4-6




(a). (b).




(c).

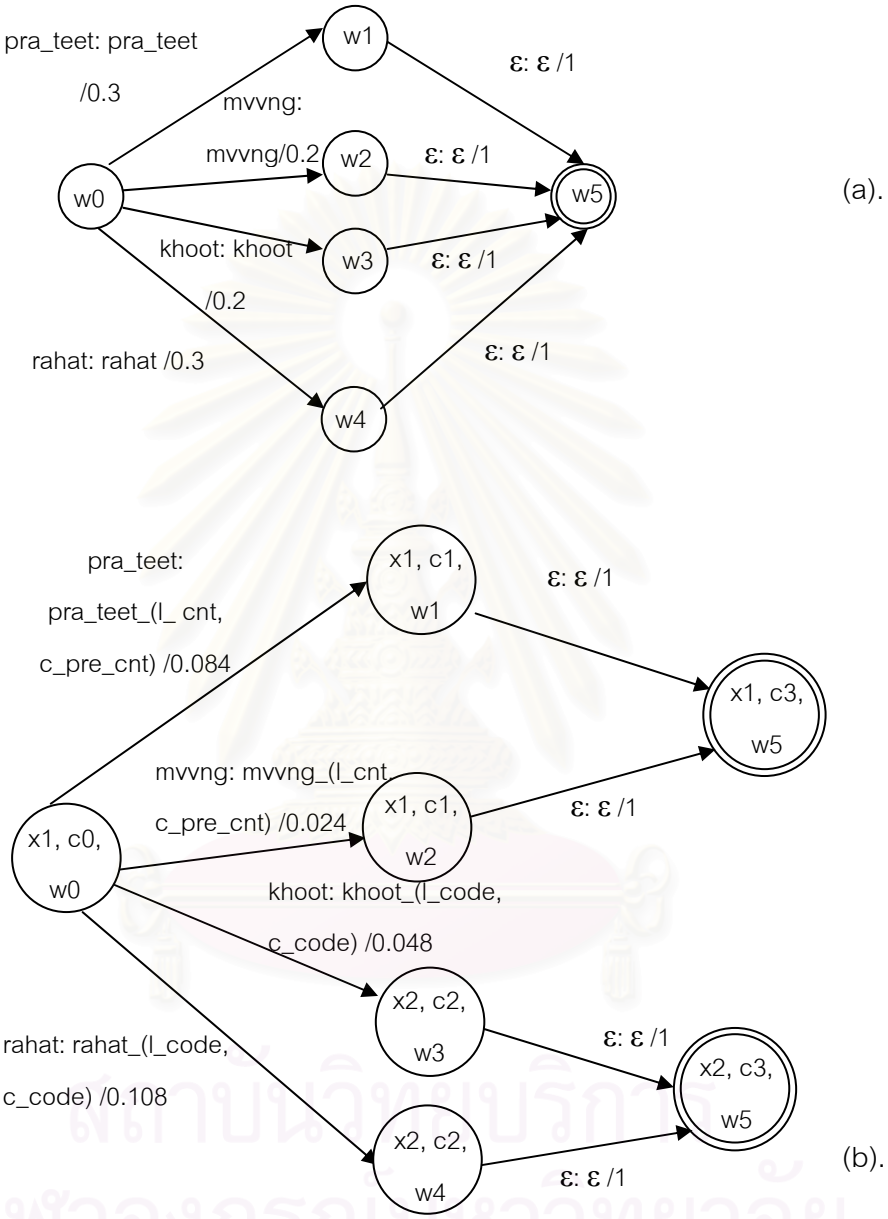
รูปที่ 4-6 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่าย $WC$ และแบบจำลอง $CC$ (a) ตัวอย่างโครงข่ายจับคู่คำศัพท์กับแนวคิด $WC$ (b) ตัวอย่างแบบจำลองแนวคิด $CC$ (c) ผลของการทำฟังก์ชันประกอบระหว่าง $WC$ และ $CC$

ในรูปที่ 4-6 นั้น จะเป็นการแสดงตัวอย่างการนำเอาโครงข่ายการจับคู่คำศัพท์กับแนวคิดของคำศัพท์ ในรูปที่ 4-6(a) มาทำการประกอบกับแบบจำลองแนวคิดในรูปที่ 4-6(b) โดยเนื่องจากเราต้องการผลลัพธ์ในลักษณะของลำดับคำ และแนวคิด ดังนั้นจึงได้กำหนดให้เอาท์พุตของแต่ละเส้นเชื่อมนั้น จะเป็นการนำเอาอินพุตจากโครงข่าย $WC$ มาต่อกับเอาท์พุตจากแบบจำลอง $CC$ ตัวอย่างเช่น ในเส้นเชื่อมของโครงข่ายในรูปที่ 4-6(c) ที่มีข้อความกำกับว่า “rahat: rahat_(l_code, c_code)/ 0.36” นั้น มีเอาท์พุตเป็น “rahat_(l_code, c_code)” ซึ่งเป็นการเอาคำว่า “rahat” ซึ่งเป็นอินพุตของเส้นเชื่อมระหว่าง x0 และ x2 ในโครงข่ายในรูปที่ 4-6(a) และเอา “(l_code, c_code)” ซึ่งเป็นเอาท์พุตของเส้นเชื่อมระหว่าง c0 และ c2 ในแบบจำลองแนวคิดในรูปที่ 4-6(b) มาต่อกัน สำหรับน้ำหนักของเส้นเชื่อมของโครงข่ายในรูปที่ 4-6(c) นั้น จะเป็นผลคูณของน้ำหนักของเส้นเชื่อมจากโครงข่ายในรูปที่ 4-6(a) และแบบจำลองในรูปที่ 4-6(b) โดยที่ค่าน้ำหนักของเส้นทางที่เดินทางในโครงข่ายในรูปที่ 4-6(c) นั้น จะเป็นผลคูณของน้ำหนักของแต่ละเส้นเชื่อมที่เดินทางผ่านมา ดังสมการที่ (4.10)

$$\begin{aligned} p(\,path\,CC \circ WC) &= \prod_{i=1}^{n} weight\ of\ edge_i \\ &= \prod_{i=1}^{n} \frac{p(w_i \mid c_i)}{p(w_l)}\, p(c_i \mid c_{i-1}) \end{aligned} \tag{4.10}$$

โดยที่ $c_i$ และ $w_i$ แทนแนวคิดและคำที่ตำแหน่งที่ $i$ ตามลำดับ โครงข่ายที่ได้จากการทำฟังก์ชันประกอบระหว่างโครงข่าย $WC$ และ $CC$ นั้น จะได้เป็นโครงข่ายที่ทำการแปลจากลำดับ

ของคำ ไปยังลำดับของคำและแนวคิดของคำ โดยค่าน้ำหนักของเส้นเชื่อมนั้น จะเป็นค่าของ $(p(w_i|c_i)/p(w_l))p(c_i|c_{i-1})$ เมื่อเรานำเอาโครงข่ายที่ทำการประกอบได้นี้ มาทำฟังก์ชันประกอบกับโครงข่ายภาษา $G$ โดยจะสามารถแสดงได้ในรูปที่ 4-7



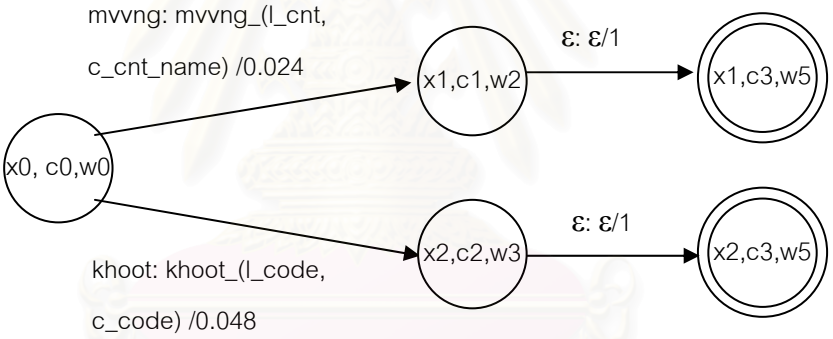

รูปที่ 4-7 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่ายภาษา $G$ และโครงข่าย $CC \circ WC$ (a) ตัวอย่างแบบจำลองภาษา $G$ (b) ผลของการทำฟังก์ชันประกอบของ $CC \circ WC$

ในตัวอย่างในรูปที่ 4-7 นั้น จะเป็นการทำการประกอบกันระหว่างโครงข่ายที่ได้จากการประกอบกันจากโครงข่าย $WC$ กับแบบจำลอง $CC$ ในรูปที่ 4-6 (c) กับแบบจำลองภาษาในรูปที่ 4-7(a) โดยโครงข่ายที่ได้นั้นจะแสดงในรูปที่ 4-7(b) ค่าน้ำหนักของเส้นเชื่อมในรูปที่ 4-7(b) นั้น จะได้จากการนำเอาค่าน้ำหนักของแบบจำลองภาษามาคูณกับค่าน้ำหนักของโครงข่ายที่ได้จากการ

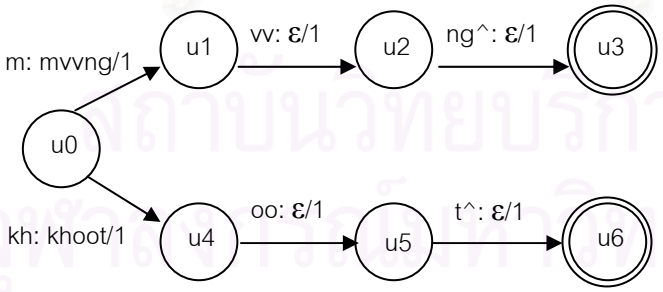
ประกอบกันจากโครงข่าย $WC$กับแบบจำลอง $CC$ โดยที่ค่าน้ำหนักของเส้นเชื่อมของแบบจำลองภาษาจะมีค่าเป็น $p(w_i|w_{i-1})$ และค่าน้ำหนักของเส้นเชื่อมของโครงข่ายที่ได้จากการประกอบกันจากโครงข่าย $WC$ กับแบบจำลอง $CC$ จะมีค่าเป็น $(p(w_i|c_i)/p(w_l))p(c_i|c_{i-1})$ ดังนั้น ค่าน้ำหนักของเส้นทางในโครงข่ายของ $CC \circ WC \circ G$ จะมีค่าเป็น

$$\begin{aligned} p(Path CC \circ WC \circ G) &= \prod_{i=1}^{n} weight\ of\ eage_i \\ &= \prod_{i=1}^{n} \frac{p(w_i \mid w_{i-1})\, p(w_i \mid c_i)\, p(c_i \mid c_{i-1})}{p(w_l)} \\ &= p(W \mid C)\, p(C) \end{aligned} \qquad (4.11)$$

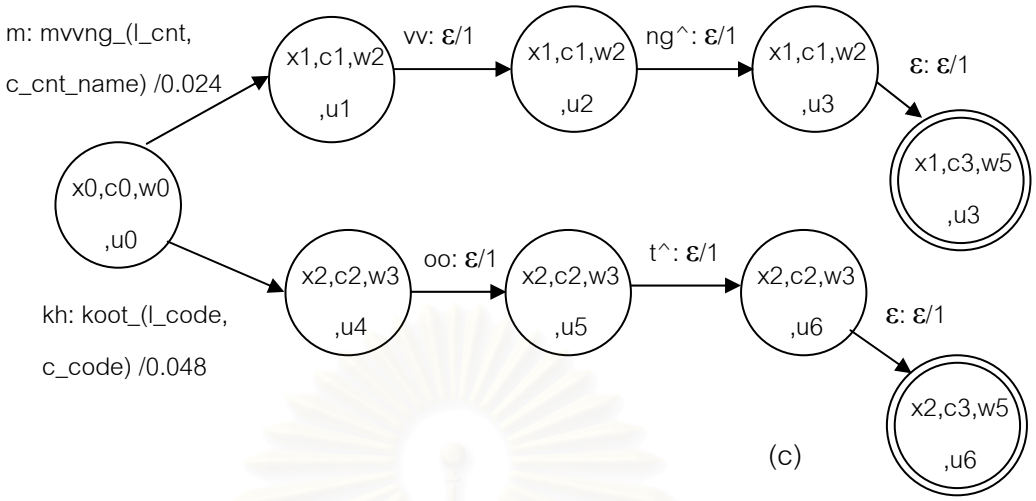
หลังจากได้โครงข่ายที่ได้จากการทำการประกอบกันระหว่าง $CC$, $WC$ และ $G$ แล้ว ก็จะนำเอาโครงข่ายที่ได้ ไปทำการประกอบกับโครงข่าย $L$ หรือโครงข่ายจับคู่ระหว่างหน่วยเสียงกับคำศัพท์ โดยผลลัพธ์สุดท้ายนั้น จะได้โครงข่ายที่ทำการจับคู่จากหน่วยเสียงไปยังคำศัพท์และแนวคิดของคำศัพท์นั้น ดังรูปที่ 4-8




(a).




(b).




รูปที่ 4-8 ตัวอย่างการทำฟังก์ชันประกอบระหว่างโครงข่าย $L$ และโครงข่าย $CC \circ WC \circ G$ (a) ตัวอย่างโครงข่าย $CC \circ WC \circ G$ (b) ตัวอย่างของโครงข่ายจับคู่ระหว่างหน่วยเสียงกับคำ $L$ (c) ผลของการทำฟังก์ชันประกอบของ $CC \circ WC \circ G \circ L$

ในรูปที่ 4-8 นั้น จะเป็นการแสดงตัวอย่างการทำการประกอบกันระหว่างโครงข่าย $CC \circ WC \circ G$ ในรูปที่ 4-8 (a) และ โครงข่าย $L$ ในรูปที่ 4-8(b) โดยที่ในรูปที่ 4-8(a) นั้น จะเป็นส่วนหนึ่งของโครงข่ายในรูปที่ 4-7(b) โดยเมื่อทำการประกอบของโครงข่าย $CC \circ WC \circ G$ และ $L$ แล้ว โครงข่ายที่ได้นั้น จะเป็นโครงข่ายของ $CC \circ WC \circ G \circ L$ ซึ่งค่าน้ำหนักของเส้นเชื่อมแต่ละเส้นนั้นจะคำนวณจากผลคูณของน้ำหนักเส้นเชื่อมของโครงข่าย $L$ ซึ่งมีค่าเป็น $p(u_i | u_{i-1,i} w_j)$ โดยที่ $u_i$ แทนหน่วยเสียงที่ตำแหน่งที่ $i$ และ น้ำหนักของเส้นเชื่อมของโครงข่าย $CC \circ WC \circ G$ ซึ่งมีค่าเป็น $p(w_j|w_{j-1})(p(w_j|c_j)/p(w_j))p(c_j|c_{j-1})$ ดังนั้นค่าความน่าจะเป็นของเส้นทางของโครงข่าย $CC \circ WC \circ G \circ L$ จะมีค่าเป็น

$$p(path\, CC \circ WC \circ G \circ L) = \prod_{i=1,j=1}^{k,n} \frac{p(u_i \mid u_{i-1}, w_j)\, p(w_j \mid w_{j-1})\, p(w_j \mid c_j)\, p(c_j \mid c_{j-1})}{p(w_j)} = p(U \mid W)\, p(W \mid C)\, p(C) \qquad (4.12)$$

โดยโครงข่าย $CC \circ WC \circ G \circ L$ นั้นจะเป็นโครงข่ายการออกเสียง ($U$) ซึ่งในการรู้จำคำและแนวคิดนั้น จะใช้โครงข่าย $U$ กับโครงข่ายหน่วยเสียง $A$ ซึ่งได้จากกราฟของเซกเมนต์ร่วมกันในการค้นหา โดยวิธีการในการค้นหานั้น ได้อธิบายในหัวข้อที่ 3.6

## 4.4 การค้นหาจุดมุ่งหมายของประโยคจากลำดับของแนวคิด

ในวิทยานิพนธ์นี้ ได้กำหนดให้จุดมุ่งหมาย (Goal) ของประโยคนั้น หมายถึงโดเมนของเรื่องที่ประโยคนั้นกล่าวถึง เช่น ในประโยค “อยากทราบอัตราค่าบริการของประเทศฟิลิปปินส์” และ “เอ่อ....อยากรู้เรื่องรหัสโทรไปประเทศอเมริกา” นั้น จะมีจุดมุ่งหมายเป็น การสอบถามอัตราค่าโทรศัพท์ และ การสอบถามเรื่องรหัสโทรศัพท์ ตามลำดับ นอกจากนี้ยังรวมถึงในบางกรณีที่ยังไม่สามารถระบุโดเมนของประโยคนั้นได้ เช่น ประโยค “ประเทศอิตาลี” ในกรณีนี้จะทราบแต่เพียงว่ามีแนวคิดของชื่อประเทศเกิดขึ้น และมีค่าของแนวคิดนั้นเป็น “อิตาลี” แต่ยังไม่ทราบว่าประโยคนี้มีจุดมุ่งหมายใด โดยในวิทยานิพนธ์นี้อาจกล่าวได้ว่าจุดมุ่งหมายนั้นเปรียบเสมือนคำสั่งที่จะทำงานต่อไป

แนวทางในการค้นหาจุดมุ่งหมายของประโยคนั้น ได้อ้างอิงมาจากงานวิจัยของกลุ่มวิจัย Wutiwiwatchai และ Furui [Wutiwiwatchai and Furui, 2006] ซึ่งได้มองปัญหาการค้นหาจุดมุ่งหมายของประโยคเป็นปัญหาการจำแนกแบบ (Pattern Classication) โดยในวิทยานิพนธ์นี้ได้เลือกการใช้โครงข่ายประสาทเทียมแบบหลายชั้นในการแบ่งแยกประเภทของจุดมุ่งหมาย ซึ่งอินพุตของโครงข่ายประสาทเทียมนั้น จะเป็นไบนารีเวกเตอร์ $V$ ขนาด n มิติ โดยที่ n คือจำนวนจุดมุ่งหมายทั้งหมดที่มีในระบบ และกำหนดให้ $v_i$ เป็นค่าของไบนารีเวกเตอร์ $V$ ที่ตำแหน่งที่ $i$ โดยที่ค่าของ $v_i$ จะมีค่าเป็น 1 เมื่อมีแนวคิด $i$ เกิดขึ้น และมีค่าเป็น 0 หากไม่เกิดแนวคิด $i$ ขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 5

## การทดลอง

ในบทนี้จะกล่าวถึงการทดลองเพื่อเปรียบเทียบประสิทธิภาพของการรู้จำคำศัพท์ การค้นหาแนวคิด และการแบ่งประเภทของจุดมุ่งหมาย ระหว่างการใช้ความรู้ของแนวคิดในการรู้จำคำศัพท์ และแนวคิดบนกราฟของเซกเมนต์ในลักษณะของแบบจำลองแนวคิด และโครงข่ายที่จับคู่ระหว่างแนวคิดกับคำศัพท์ และการค้นหาแนวคิดโดยใช้ผลจากการรู้จำคำที่ไม่ใช้ความรู้ของแนวคิดในการรู้จำคำศัพท์บนกราฟของเซกเมนต์ จากนั้นค้นหาแนวคิดจากแบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์ ทั้งนี้เพื่อวิเคราะห์ข้อได้เปรียบและเสียเปรียบของวิธีการทั้งสองวิธี

### 5.1 ระบบตัวอย่างที่ใช้ในการทดสอบ

ในวิทยานิพนธ์นี้ ได้สร้างสถานการณ์จำลองในการรู้จำคำศัพท์ ค้นหาลำดับของแนวคิดและการแบ่งประเภทของจุดมุ่งหมายจากตัวอย่างเสียงภายใต้โดเมนการสอบถามอัตราค่าบริการและรหัสโทรศัพท์ทางไกลระหว่างประเทศ ซึ่งจะใช้ภาษาธรรมชาติ (Natural Language) ในการติดต่อ และรองรับการใช้งานกับภาษาไทยเท่านั้น ในการทดลองนั้น จะให้ระบบให้ผลลัพธ์ในการรู้จำเป็นลำดับของคำ ลำดับของแนวคิด และจุดมุ่งหมายของประโยคเสียงนั้น ดังรูปที่ 5-1

รูปที่ 5-1ตัวอย่างระบบที่ใช้ในการทดลอง

### 5.2 ตัวอย่างเสียง

ในวิทยานิพนธ์นี้ ได้แบ่งตัวอย่างข้อมูลที่ใช้เป็น 2 ส่วน คือ ส่วนที่ใช้ในการเรียนรู้ และส่วนที่ใช้การทดสอบ โดยทั้ง 2 ส่วนนั้น จะบันทึกเสียงผ่านการ์ดโทรศัพท์ รุ่น Dialogic® รุ่น D4/PCI โดยกำหนดให้บันทึกเสียงในไฟล์สกุล wav ด้วยอัตราการสุ่ม (Audio Sample Rate) 8,000 เฮิร์ต ที่อัตราข้อมูลจำนวน 8 บิตต่อวินาที โดยกลุ่มตัวอย่างเสียงที่ทำการบันทึกเสียงตัวอย่างเรียนรู้ และตัวอย่างทดสอบนั้น จะแยกกลุ่มตัวอย่างเสียงเป็นคนละชุดกัน

### 5.2.1 ตัวอย่างเสียงเพื่อใช้ในการเรียนรู้

ตัวอย่างเสียงในกลุ่มนี้ จะแบ่งประเภทออกเป็น กลุ่มตัวอย่างเสียงที่ใช้ในการเรียนรู้ลักษณะไวยากรณ์ของภาษาพูดภายใต้โดเมนการสอบถามอัตราค่าบริการและรหัสโทรศัพท์ทางไกลระหว่างประเทศ และกลุ่มตัวอย่างเสียงเพื่อใช้ในการเรียนรู้ลักษณะของค่าสังเกต และเซกเมนต์ของแต่ละหน่วยเสียง ดังนี้

1. ตัวอย่างเสียงเพื่อใช้ในการเรียนรู้ลักษณะไวยากรณ์ของภาษาพูด

ตัวอย่างเสียงกลุ่มนี้ จะใช้เพื่อพิจารณาลักษณะลำดับของคำ และคำศัพท์ที่พบบ่อยภายใต้โดเมนที่กำหนด ซึ่งจะใช้ในการกำหนดคำศัพท์ในพจนานุกรม (Dictionary) การเรียนรู้แบบจำลองการออกเสียง แบบจำลองภาษา และแบบจำลองแนวคิด รวมถึงแบบจำลองทางสถิติและเครื่องจักรสถานะจำกัดแบบถ่วงน้ำหนักของไวยากรณ์ในแต่ละแนวคิด เพื่อใช้ในการค้นหาลำดับของแนวคิดจากลำดับของคำศัพท์ที่ได้จากการรู้จำ (ในหัวข้อที่ 4.3) โดยกลุ่มตัวอย่างเสียงนี้จะได้จากการบันทึกเสียงสนทนาในระหว่างการให้บริการและสอบถามข้อมูลทางโทรศัพท์ระหว่างพนักงานและลูกค้าในเรื่องของอัตราค่าบริการและรหัสโทรศัพท์ระหว่างประเทศ จำนวน 150 ประโยค โดยตัวอย่างเสียงดังกล่าว ได้จากศูนย์บริการข้อมูล บริษัท กสท โทรคมนาคม จำกัด (มหาชน)

2. ตัวอย่างเสียงเพื่อใช้ในการเรียนรู้ลักษณะของค่าสังเกต และเซกเมนต์ของแต่ละหน่วยเสียง

ตัวอย่างเสียงกลุ่มนี้ จะใช้ในการเรียนรู้ค่าพารามิเตอร์ของแบบจำลองเกาส์เซียนของเวกเตอร์ลักษณะสำคัญของหน่วยเสียงต่างๆ (ในหัวข้อที่ 3.5.1) รวมถึงแบบจำลองระยะเวลาซึ่งเป็นแบบจำลองของความยาวของเซกเมนต์ในแต่ละหน่วยเสียง กลุ่มตัวอย่างเสียงประเภทนี้จะบันทึกจากกลุ่มตัวอย่างเสียงจำนวน 30 คน แบ่งเป็นผู้ชาย 16 คน และผู้หญิง 14 คน โดยได้กำหนดข้อความที่จะทำการบันทึก แล้วให้กลุ่มตัวอย่างเสียงนั้น พูดตามข้อความที่กำหนดผ่านทางการ์ดโทรศัพท์ ซึ่งประเภทของข้อความที่ทำการบันทึกนั้น จะแบ่งเป็น

- ข้อความที่ใช้ในการบันทึกเสียงจากฐานข้อมูลโลตัส ในชุดหน่วยเสียงสมดุล Phonetically Distributed Set- PD) ซึ่งมีจำนวนข้อความทั้งหมด 801 ประโยค โดยตัวอย่างเสียงที่บันทึกนั้น จะแบ่งให้ผู้บันทึกเสียงทำการบันทึกคนละ 80-81 ประโยค ซึ่งได้กำหนดว่า ข้อความในการบันทึกเสียงของแต่ละประโยคนั้น จะต้อง

มีผู้บันทึกเสียงจำนวน 3 คนสำหรับแต่ละข้อความ ตัวอย่างเสียงประเภทนี้ จะมีจำนวนประโยคครวมทั้งสิ้น 2,403 ประโยค

- ข้อความที่เป็นคำศัพท์ที่พบบ่อยๆ ในการสอบถามอัตราค่าโทรศัพท์ และรหัสโทรศัพท์ทางไกลระหว่างประเทศ จำนวน 391 คำ โดยได้กำหนดให้ผู้บันทึกเสียงนั้น พูดคำศัพท์คนละ 40-42 คำ ตามที่ได้แบ่งไว้

### 5.2.2 ตัวอย่างเสียงเพื่อใช้ในการทดสอบ

ตัวอย่างเสียงในกลุ่มนี้ จะมีไว้เพื่อประเมินความถูกต้องของการรู้จำคำ แนวคิด และการแบ่งประเภทของเป้าหมาย โดยกลุ่มผู้บันทึกเสียงนั้น จะเป็นคนละกลุ่มกับผู้บันทึกตัวอย่างเสียงเพื่อใช้ในการเรียนรู้ ในการบันทึกนั้น จะให้ผู้บันทึกทำการสนทนากับระบบสมมุติ ซึ่งเป็นระบบที่ให้ข้อมูลทางของอัตราค่าโทรศัพท์ และรหัสโทรศัพท์ทางไกลระหว่างประเทศผ่านทางโทรศัพท์ โดยใช้ภาษาธรรมชาติในการติดต่อ โดยจะให้ผู้บันทึกทำการคิดประโยคที่จะใช้สนทนาเองและไม่มีการกำหนดข้อความที่จะให้บันทึกไว้ก่อนหน้า โดยกลุ่มตัวอย่างเสียงนี้ จะบันทึกจาก 15 คน แบ่งเป็นผู้ชาย 9 คน และผู้หญิง 6 คน โดยจะบันทึกคนละ 5-20 ข้อความเสียง รวมจำนวนข้อความเสียงทั้งหมด 200 ประโยค

## 5.3 การออกแบบระบบการทดลอง

### 5.3.1 การออกแบบคำศัพท์ แนวคิด และจุดมุ่งหมาย

ในวิทยานิพนธ์นี้ ได้มีการกำหนดคำศัพท์ แนวคิด และจุดมุ่งหมายภายใต้โดเมนของการสอบถามข้อมูลอัตราค่าบริการและรหัสโทรศัพท์ทางไกลระหว่างประเทศ โดยมีจำนวนคำศัพท์ในระบบทั้งหมด 284 คำ จำนวนแนวคิด 13 แนวคิด และจำนวนจุดมุ่งหมาย 2 ประเภท โดยที่คำศัพท์บางคำ จะสามารถจัดกลุ่มได้มากกว่า 1 แนวคิด เช่น คำว่า “ประเทศ” สามารถจัดให้อยู่ในประเภทของ (l_cnt, c_cnt_name) ซึ่งเป็นแนวคิดของชื่อประเทศ หรือ (l_cnt, c_code) ซึ่งเป็นแนวคิดที่ของกลุ่มคำที่อยู่ในประเภทรหัสโทรศัพท์ โดยจำนวนคำศัพท์ของแต่ละแนวคิดนั้น จะแสดงในตารางที่ 5-1

ตารางที่ 5-1 แนวคิด และจำนวนคำศัพท์ที่ใช้ในการทดลอง

| แนวคิด | จำนวนคำศัพท์ | ความหมาย | ตัวอย่างคำศัพท์ |
|---|---|---|---|
| (l_cnt_name, c_cnt_name) | 231 | ชื่อประเทศ | จีน, เกาหลี, ญี่ปุ่น, สหรัฐอเมริกา |
| (l_rate, c_rate) | 7 | คำศัพท์ที่เกี่ยวกับอัตราค่าบริการ | อัตรา, ราคา, ค่าโทร |
| (l_Q_question, c_Q_question) | 4 | คำถามที่ใช้ถามรหัสโทรศัพท์ | อะไร, แบบไหน, อย่างไร |
| (l_H_question, c_H_question) | 5 | คำถามที่ใช้ถามอัตราค่าบริการ | เท่าไหร่, กี่บาท, นาทีละเท่าไหร่ |
| (l_code, c_code) | 6 | คำศัพท์เกี่ยวกับรหัสโทรศัพท์ | รหัส, โค๊ด, หมายเลข |
| (l_cnt, c_cnt_name) | 5 | คำนำหน้าชื่อประเทศ | ประเทศ, เมือง, รัฐ, สาธารณรัฐ , สหรัฐ |
| (l_cnt, c_code) | 5 | คำศัพท์ที่ใช้ขยายคำในกลุ่มรหัสโทรศัพท์ โดยจะมีความหมายถึงทางไกล | ประเทศ, ต่างประเทศ, ทางไกล |
| (l_cnt, c_rate) | 4 | คำศัพท์ที่ใช้ขยายคำในกลุ่มอัตราค่าบริการ โดยจะมีความหมายถึงทางไกล | ต่างประเทศ, ทางไกล, ระหว่างประเทศ |
| (l_want, c_request) | 9 | คำขอร้อง | อยากทราบ, เรียนถาม, อยากรู้, สอบถาม |
| (l_phone, c_phone) | 6 | คำกริยา ที่หมายถึงการโทรศัพท์ | โทรศัพท์, โทร, กดไป, โทรไป , ต่อสาย |

| แนวคิด | จำนวนคำศัพท์ | ความหมาย | ตัวอย่างคำศัพท์ |
|---|---|---|---|
| (l_phone, c_code) | 6 | คำศัพท์ที่ใช้ขยายคำในกลุ่มรหัสโทรศัพท์ โดยจะมีความหมายถึงการโทรศัพท์ | โทรศัพท์, โทรไป, โทรออก, กดไป |
| (l_phone, c_rate) | 4 | คำศัพท์ที่ใช้ขยายคำในกลุ่มอัตราค่าบริการ โดยจะมีความหมายถึงการโทรศัพท์ | โทรศัพท์, โทรไป, โทรออก, กดไป |
| (l_gar, c_gar) | 8 | คำศัพท์ หรือเสียงอื่นๆ ที่อยู่นอกเหนือจากพจนานุกรม | เอ่อ, อ่า, ค่ะ, ครับ |

จากตารางที่ 5-1 นั้น แนวคิด (l_gar, c_gar) หรือแนวคิดของกลุ่มคำศัพท์ที่อยู่นอกเหนือจากแนวคิดอื่นๆ จะประกอบด้วยหน่วยเสียงขยะ (หัวข้อที่ 3.5.1.5) ซึ่งมีจำนวน 3 หน่วยเสียง (ได้แก่ หน่วยเสียงของกลุ่มของพยัญชนะต้น กลุ่มของสระ และกลุ่มของตัวสะกดในภาษาไทย) และคำอื่นๆ ที่พบบ่อยๆ จำนวน 5 คำ ได้แก่ “ค่ะ” “ครับ” “ของ” “สวัสดี” และ “ฮัลโหล” โดยในงานวิทยานิพนธ์นี้จะพิจารณาว่าหน่วยเสียงขยะนั้นเป็นคำศัพท์คำหนึ่ง ดังนั้นจำนวนคำศัพท์ของแนวคิด (l_gar, c_gar) จึงมีจำนวน 8 คำ (3 หน่วยเสียง+ 5 คำศัพท์)

ในระบบที่ใช้ในการทดลองนี้ จะมีจุดมุ่งหมาย 2 ประเภท ได้แก่ จุดมุ่งหมายในการสอบถามเรื่องอัตราค่าบริการโทรศัพท์ทางไกลระหว่างประเทศ และจุดมุ่งหมายในการสอบถามเรื่องรหัสโทรศัพท์ทางไกลระหว่างประเทศ นอกจากนี้ยังรวมถึงการพิจารณาในบางประโยคที่ยังไม่สามารถระบุจุดมุ่งหมายได้ โดยตัวอย่างประโยคของแต่ละจุดมุ่งหมายนั้น จะแสดงในตารางที่ 5-2

ตารางที่ 5-2 ตัวอย่างประโยคสนทนาของแต่ละจุดมุ่งหมาย

| จุดมุ่งหมาย | ตัวอย่างประโยค |
|---|---|
| สอบถามเรื่องอัตราโทรศัพท์ทางไกลระหว่างประเทศ | “อยากทราบอัตราค่าโทรศัพท์ไปจีน”<br>“โทรไปอินเดีย ไม่ทราบว่าราคานาทีละกี่บาทครับ” |
| สอบถามเรื่องรหัสโทรศัพท์ทางไกลระหว่างประเทศ | “จะโทรไปญี่ปุ่น ต้องกดอะไรก่อนคะ”<br>“ถ้าโทรไปประเทศฝรั่งเศส รหัสโทรศัพท์คืออะไรครับ” |

| จุดมุ่งหมาย | ตัวอย่างประโยค |
|---|---|
| ประโยคที่ยังไม่สามารถระบุจุดมุ่งหมายได้ | “ประเทศสิงคโปร์”<br>“ฮัลโหล...ฮัลโหล....” |

### 5.3.2 การสร้างกราฟของเซกเมนต์

ในการสร้างกราฟของเซกเมนต์นั้น จะสร้างจากชุดเครื่องมือ htk [HTK] เวอร์ชัน 3.3 โดยวิธีการสร้างกราฟของเซกเมนต์ ได้ระบุในหัวข้อที่ 3.2 ในการทดลองนี้ได้สร้างกราฟของเซกเมนต์จากลำดับของเซกเมนต์สมมุติฐานจำนวน 10 ลำดับ ซึ่งมาจากการสร้างสมมุติฐานของลำดับของเซกเมนต์เมื่อพิจารณาจากคำศัพท์ในพจนานุกรมจำนวน 1 สมมุติฐาน และสร้างสมมุติฐานของลำดับของเซกเมนต์เมื่อกำหนดลำดับของการแบ่งเซกเมนต์เป็น CVC (พยัญชนะต้น สระ และ/หรือ ตัวสะกด) จำนวน 9 สมมุติฐาน

### 5.3.3 การพิจารณาเวกเตอร์ลักษณะสำคัญของเซกเมนต์

ในการพิจารณาค่าเวกเตอร์ลักษณะสำคัญในหัวข้อที่ 3.3 ซึ่งจะได้เป็นเวกเตอร์ขนาด 117 มิติ ได้มีการพิจารณาค่าของสัมประสิทธิ์เซปสตรัมบนสเกลเมล โดยในการทดลองนี้ได้ใช้ชุดเครื่องมือ htk [HTK] ในการคำนวณค่าดังกล่าว โดยกำหนดให้สัมประสิทธิ์เซปสตรัมบนสเกลเมล ซึ่งมีค่าพลังงานรวมอยู่ด้วย อัตราการเปลี่ยนแปลง (Delta) และความเร่ง (Accelerations) โดยใช้กรอบสัญญาณเสียงขนาด 25 มิลลิวินาที แต่ละกรอบจะมีระยะห่างกัน 10 มิลลิวินาที ซึ่งจะได้เวกเตอร์ลักษณะสำคัญ 39 มิติ โดยที่จะนำเอาเวกเตอร์ที่ได้นี้ ไปใช้ในการคำนวณค่าเวกเตอร์ลักษณะสำคัญของแต่ละเซกเมนต์เพื่อให้ได้เวกเตอร์ขนาด 117 มิติ ต่อไป ตามที่กล่าวแล้วในหัวข้อที่ 3.3

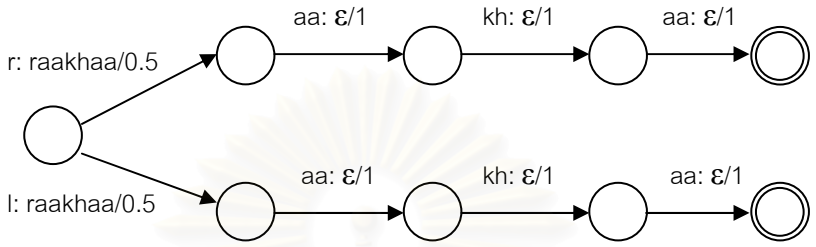
### 5.3.4 แบบจำลองการออกเสียง และโครงข่ายจับคู่ระหว่างคำศัพท์และแนวคิด

แบบจำลองการออกเสียงซึ่งใช้ในการสร้างโครงข่ายจับคู่ระหว่างหน่วยเสียงและคำศัพท์ จะแสดงลำดับของหน่วยเสียงที่ประกอบกันเป็นคำศัพท์ใดๆ โดยที่คำศัพท์แต่ละคำนั้นสามารถมีลำดับของหน่วยเสียงได้มากกว่า 1 ลำดับ ซึ่งในการทดลองนี้ ได้สร้างแบบจำลองนี้จากเครื่องจักรสถานะแบบถ่วงน้ำหนัก โดยที่อินพุตของแต่ละเส้นเชื่อมนั้นจะเป็นหน่วยเสียงที่ประกอบเป็นคำศัพท์นั้น (ในหัวข้อที่ 3.5.3)

สำหรับน้ำหนักของเส้นเชื่อมที่เชื่อมระหว่างสถานะของเครื่องจักรสถานะนั้น ในการทดลองนี้ ได้กำหนดให้น้ำหนักของเส้นเชื่อมมีค่าเป็นดังนี้

$$weight(edge_i\ of\ node\ n) = \frac{1}{\#\ all\ edges\ out\ of\ node\ n} \qquad (5.1)$$

จากสมการที่ (5.1) จะเห็นว่าค่าน้ำหนักของเส้นเชื่อมแต่ละเส้นของสถานะ n ใดๆ นั้น จะมีค่าเป็นอัตราส่วน 1 ต่อจำนวนเส้นเชื่อมทั้งหมดที่ออกจากสถานะ n นั้น ดังตัวอย่างในรูปที่ 5-2




รูปที่ 5-2ตัวอย่างแบบจำลองการออกเสียงของคำว่า “ราคา”

ในรูปที่ 5-2 นั้น จะเป็นแบบจำลองการออกเสียงของคำว่า “ราคา” โดยได้กำหนดให้คำศัพท์นี้สามารถประกอบด้วยลำดับของหน่วยเสียงเป็น “r aa kh aa” (ราคา) หรือ “l aa kh aa” (ลาคา) โดยหากพิจารณาที่สถานะเริ่มต้น หรือสถานะที่อยู่ซ้ายมือสุด จะเห็นว่ามีเส้นเชื่อมที่ออกจากสถานะเริ่มต้นอยู่ 2 เส้น ได้แก่ เส้นเชื่อม “r: raakhaa/0.5” และ “l: raakhaa/0.5” น้ำหนักของเส้นเชื่อมของทั้ง 2 เส้นนี้ จึงมีค่าเป็น 1/2 นั่นเอง

ในกรณีของโครงข่ายจับคู่ระหว่างคำศัพท์และแนวคิดนั้น จะสร้างจากเครื่องจักรสถานะแบบถ่วงน้ำหนัก โดยค่าน้ำหนักของเส้นเชื่อมนั้น จะเป็นค่าของ $p(w_l|c_l)/p(w_l)$ (ดังแสดงในหัวข้อที่ 4.3.2) ในการทดลองนี้ได้กำหนดให้ค่าของ $p(w_l|c_l)$ และ $p(w_l)$ มีค่าดังนี้

$$p(w_l \mid c_l) = \frac{1}{Number\ of\ element\ in\ c_l} \quad \text{และ} \quad p(w_l) = \frac{1}{Number\ of\ all\ word}$$

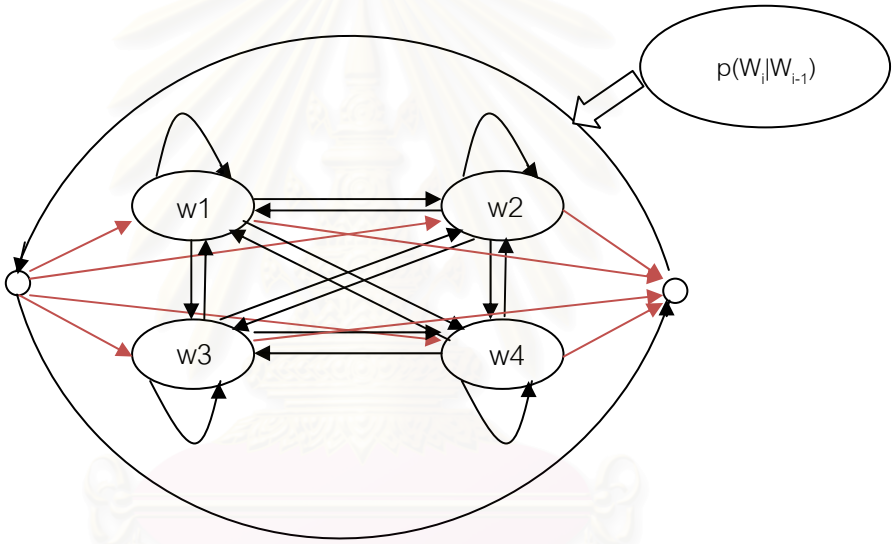
นั่นคือ กำหนดให้ คำศัพท์ทุกคำที่อยู่ในแนวคิดเดียวกันนั้น จะมีโอกาสที่จะเกิดขึ้นได้เท่าๆ กัน และค่าความน่าจะเป็นที่จะเกิดคำศัพท์ $w_l$ ใดๆ นั้น ก็จะโอกาสเกิดขึ้นได้เท่าๆ กันในทุกๆ คำศัพท์

### 5.3.5 แบบจำลองภาษา และแบบจำลองแนวคิด

การออกแบบแบบจำลองภาษานั้น ได้ใช้เครื่องจักรสถานะจำกัดแบบถ่วงน้ำหนักในการสร้างแบบจำลอง ซึ่งได้กำหนดให้สถานะแต่ละสถานะจะแทนคำศัพท์ที่มีในพจนานุกรม รวมถึงหน่วยเสียงขยะ และแต่ละสถานะจะมีเส้นเชื่อมต่อถึงสถานะอื่นๆ ทั่วทั้งหมดในแบบจำลอง (Fully

Connected Graph) นอกจากนี้ยังสามารถมีการวนซ้ำในสถานะเดิมได้ ทั้งนี้เพื่อให้สามารถรองรับลำดับของคำศัพท์ได้อย่างครอบคลุม ในการคำนวณค่าน้ำหนัก หรือค่าความน่าจะเป็นในการเปลี่ยนสถานะนั้น กรณีที่ตัวอย่างข้อมูลเสียงไม่มีข้อมูลของการเปลี่ยนจากสถานะใดไปยังสถานะหนึ่ง หรือมีค่าความน่าจะเป็นในการเปลี่ยนสถานะนั้นเป็น 0 ในกรณีนี้จะกำหนดให้น้ำหนักของเส้นเชื่อมดังกล่าวมีค่าน้อยมากๆ ที่ไม่ใช่ 0 แทน

นอกจากนี้ยังได้กำหนดให้สถานะเริ่มต้นของแบบจำลองภาษานั้น สามารถเชื่อมไปยังสถานะคำศัพท์อื่นๆ ได้ทุกสถานะ นั่นคือ สามารถให้เริ่มต้นที่คำศัพท์ใดก่อนก็ได้ เช่นเดียวกันกับสถานะสิ้นสุด ที่ทุกๆ สถานะของคำศัพท์นั้นสามารถเชื่อมไปยังสถานะสุดท้ายได้ นอกจากนี้ยังได้กำหนดให้มีเส้นเชื่อมจากสถานะสิ้นสุด มายังสถานะเริ่มต้นได้ เพื่อให้สามารถทำการวนซ้ำได้ ดังรูปที่ 5-3




รูปที่ 5-3 ตัวอย่างแบบจำลองภาษาที่ใช้ในการทดลอง เมื่อกำหนดให้มีคำศัพท์ในระบบ 4 คำ

ในแบบจำลองแนวคิดนั้น จะสร้างแบบจำลองโดยใช้เครื่องจักรสถานะจำกัดแบบถ่วงน้ำหนักเช่นกันเดียวกันกับแบบจำลองภาษา โดยที่สถานะในแบบจำลองแนวคิดนี้ จะแทนด้วยแนวคิดที่มีในระบบแต่ละแนวคิด ซึ่งรวมถึงแนวคิดขยะด้วย แบบจำลองแนวคิดนี้จะมีลักษณะเดียวกันกับแบบจำลองภาษา นั่นคือ กำหนดให้สถานะแต่ละสถานะสามารถเชื่อมต่อถึงกันได้ โดยที่ทุกๆ สถานะของแต่ละแนวคิดนั้น จะเชื่อมต่อกับสถานะเริ่มต้น และสถานะสิ้นสุด

## 5.4 การวัดประสิทธิภาพ

ในการทดลองในส่วนการวัดประสิทธิการรู้คำศัพท์ และการค้นหาลำดับของแนวคิด ได้เปรียบเทียบประสิทธิภาพของการรู้จำเสียงคำศัพท์ แนวคิด และจุดมุ่งหมายโดยใช้ค่าพรีซีชัน (Precision) และ รีคอล (Recall) โดยในการคำนวณนั้น จะมาจาก

$$\Pr ecision = \frac{A}{C} \quad (5.1)$$

$$\mathrm{Re}call = \frac{A}{O} \quad (5.2)$$

โดยที่ $A$ จะแทนจำนวนความถูกต้องที่ได้จากการรู้จำของสิ่งที่สนใจ $C$ แทนจำนวนของสิ่งที่สนใจที่ได้จากผลของการรู้จำทั้งหมด และ $O$ แทนจำนวนของสิ่งที่สนใจที่ปรากฏในข้อมูลที่ใช้ในการเปรียบเทียบ โดยสิ่งที่สนใจนั้น ได้แก่ คำศัพท์ที่ได้จากการรู้จำ แนวคิด และจุดมุ่งหมาย นั่นเอง ดังนั้นค่าพรีซีชันนั้น จึงเป็นการเปรียบเทียบความถูกต้องของผลลัพธ์ที่ได้จากการรู้จำกับจำนวนของสิ่งที่สนใจที่เกิดขึ้นจริงๆ และค่าของรีคอลนั้น จะเป็นการเปรียบเทียบความถูกต้องของสิ่งที่สนใจที่ได้จากการรู้จำกับจำนวนของสิ่งที่รู้จำได้ทั้งหมด โดยสิ่งที่เราต้องการนั้น คือ การรู้จำที่ได้พรีซีชัน และค่ารีคอลสูงๆ

## 5.5 การทดลองในการรู้จำเสียงพูด

ในการทดลองนี้ จะทำการทดลองเพื่อเปรียบเทียบประสิทธิภาพในการรู้เสียงพูดจากกราฟของเซกเมนต์ โดยจะเปรียบการรู้จำคำศัพท์ระหว่าง

- การรู้จำโดยใช้ความรู้ในเรื่องของลำดับของแนวคิดมาช่วย นั่นคือ การรู้จำจากกราฟของเซกเมนต์โดยใช้โครงข่ายการออกเสียงที่สร้างจากการทำฟังก์ชันประกอบระหว่างโครงข่ายจับคู่ระหว่างหน่วยเสียงและคำศัพท์ แบบจำลองภาษา โครงข่ายที่ใช้ในการจับคู่ระหว่างคำศัพท์และแนวคิด และแบบจำลองแนวคิดในการค้นหาคำ
- การรู้จำที่ไม่ได้ใช้ความรู้ของลำดับแนวคิดมาช่วย หรือ การรู้จำโดยใช้โครงข่ายจับคู่ระหว่างหน่วยเสียงและคำศัพท์ที่มาจากการทำฟังก์ชันประกอบระหว่างโครงข่ายจับคู่ระหว่างหน่วยเสียงและคำศัพท์ และแบบจำลองภาษา

โดยในการทดลองนี้ได้ทดลองเปรียบเทียบลำดับของคำที่ได้จากการรู้จำของเสียงในลักษณะบทพูดต่อเนื่อง (Continuous Speech) ซึ่งได้จากตัวอย่างเสียงทดสอบ (ในหัวข้อที่ 5.2.2) ซึ่งมีวัตถุประสงค์เพื่อพิจารณาบทบาทของการใช้ความรู้ของแนวคิดมาช่วยเพิ่มความถูกต้องในการรู้จำคำ

### 5.5.1 ผลการทดลอง

ประสิทธิภาพในการรู้จำเสียงพูดในระดับของคำโดยใช้ค่าพรีซีชัน (Precision) และรีคอล (Recall) นั้นจะแสดงได้ดังตารางที่ 5-3

ตารางที่ 5-3 ผลการเปรียบเทียบประสิทธิภาพการรู้จำเสียงพูดจากกราฟของเซกเมนต์ โดยใช้ความรู้ของแนวคิด

| การรู้จำเสียงพูดจากกราฟของเซกเมนต์ | ค่ารีคอล (%) | ค่าพรีซีชัน (%) |
|---|---|---|
| ไม่ใช้ความรู้แนวคิดในโครงข่ายการออกเสียง | 54.93 | 69.72 |
| ใช้ความรู้แนวคิดในโครงข่ายการออกเสียง | 70.59 | 69.84 |

โดยผลการเปรียบเทียบในตารางที่ 5-3 นั้น จะพิจารณาเฉพาะการรู้จำคำศัพท์ในแนวคิดใดๆ ที่ไม่ใช่แนวคิดขยะเท่านั้น

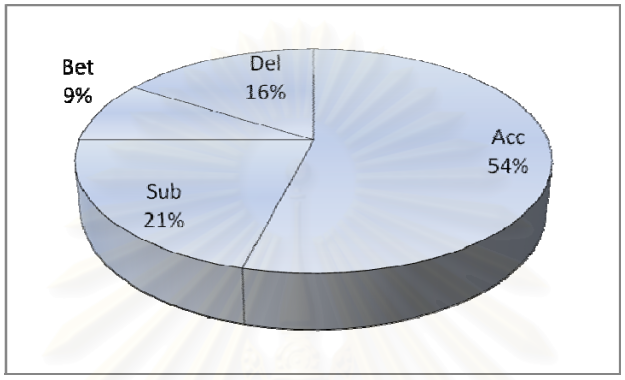
### 5.5.2 วิเคราะห์ผลการทดลอง

จากผลการเปรียบเทียบค่าพรีซีชัน และรีคอลในตารางที่ 5-3 นั้น จะเห็นว่าการใช้ความรู้ของแนวคิดมาช่วยนั้น จะได้ค่าพรีซีชัน และรีคอลมากกว่าการไม่ใช้ความรู้ของแนวคิดมาช่วย หรือกล่าวได้ว่า การใช้ความรู้ของแนวคิดมาช่วยนั้น จะรู้จำคำศัพท์ที่ถูกต้องมากกว่า (พิจารณาจากค่ารีคอล) ในขณะที่เมื่อเปรียบเทียบจำนวนผลลัพธ์ที่ถูกต้องต่อจำนวนผลลัพธ์ทั้งหมดที่ได้มา พบว่ามีอัตราส่วนที่สูงกว่า (พิจารณาจากค่าพรีซีชัน)

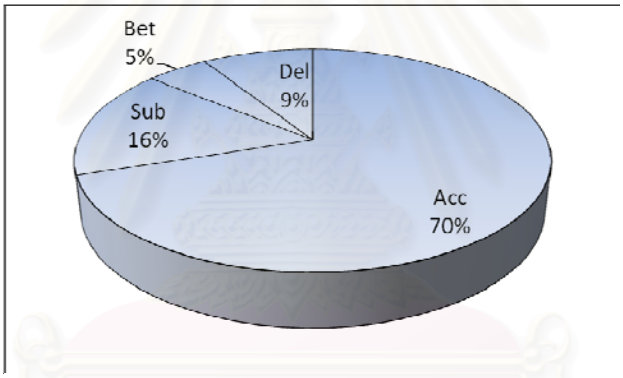
เมื่อลองพิจารณาข้อผิดพลาดที่เกิดขึ้น (Error) โดยเปรียบเทียบเปอร์เซ็นต์ของความถูกต้อง (Acc) เปอร์เซ็นต์ของข้อผิดพลาดที่เกิดจากการรู้จำคำผิด (Sub) เปอร์เซ็นต์ของข้อผิดพลาดที่เกิดจากการแบ่งคำผิด (Bet) และเปอร์เซ็นต์ของข้อผิดพลาดที่เกิดจากการรู้จำคำบางคำหายไป (Del) ของการรู้จำคำจากกราฟของเซกเมนต์ เมื่อใช้ความรู้ในเรื่องของแนวคิด และไม่ใช้ความรู้ดังกล่าวมาใช้ในการค้นหา

ในการทดลองนี้ ได้เปรียบเทียบคำที่มีการระบุลำดับของคำที่ถูกต้องแล้ว กับผลของการรู้จำคำที่ได้ ณ ตำแหน่งเดียวกัน โดยจะพิจารณาว่า คำที่เกิดขึ้นจริงในแต่ละคำนั้น จะให้ผลการรู้จำถูกต้องหรือไม่ และถ้าเกิดความผิดพลาดในการรู้จำขึ้น ข้อผิดพลาดนั้นจะเป็นข้อผิดพลาดประเภทใด โดยค่าเปอร์เซ็นต์ของความถูกต้องนั้น จะเป็นอัตราส่วนระหว่างคำศัพท์ที่รู้จำได้ถูกต้องกับจำนวนคำศัพท์ทั้งหมดที่เกิดขึ้นจริงในตัวอย่างเสียงทดสอบ เช่นเดียวกับค่าเปอร์เซ็นต์ของความผิดพลาดต่างๆ ซึ่งจะเป็นอัตราส่วนที่เกิดข้อผิดพลาดประเภทนั้นต่อจำนวนคำศัพท์ทั้งหมดที่

เกิดขึ้นจริง สำหรับตัวอย่างของข้อผิดพลาดต่างๆ นั้น ได้แก่ ตัวอย่างข้อผิดพลาดของการรู้จำคำผิด เช่น ในเสียงทดสอบนั้น พูดคำว่า “โทรไป” ที่เวลา t แต่ผลการรู้จำได้คำว่า “เท่าไหร่” ที่เวลาเดียวกัน เป็นต้น ข้อผิดพลาดของการแบ่งคำผิดนั้น จะเป็นข้อผิดพลาดที่รู้จำคำศัพท์หนึ่งคำ เป็นคำย่อยๆ หลายคำรวมกัน เช่น รู้จำคำว่า “ออสเตรเลีย” เป็นคำว่า “ขอ” และ “อะไร” เป็นต้น และข้อผิดพลาดที่เกิดจากการรู้จำคำบางคำหายไป คือ การรู้จำคำศัพท์ที่ไม่ได้อยู่ในแนวคิดขยะ ให้เป็นคำ หรือหน่วยเสียงขยะ โดยผลการเปรียบเทียบจะแสดงในกราฟในรูปที่ 5-4




(a)




(b)

รูปที่ 5-4 กราฟเปรียบเทียบอัตราส่วนของข้อผิดพลาดต่างๆ ที่เกิดขึ้นในการรู้จำคำ (a) เมื่อไม่ใช้ความรู้ของแนวคิดมาช่วยในการรู้จำคำ (b) เมื่อใช้ความรู้ของแนวคิดมาช่วย

จากกราฟในรูปที่ 5-4เป็นการเปรียบเทียบข้อผิดพลาดในการรู้จำคำที่เกิดขึ้นต่างๆ โดยที่กราฟในรูปที่ 5-4(a) นั้น จะเป็นการเปรียบเทียบข้อผิดพลาดเมื่อไม่ได้ใช้ความรู้ของแนวคิดมาช่วยในการรู้จำ และกราฟในรูปที่ 5-4(b) นั้น จะเป็นกราฟเปรียบเทียบข้อผิดพลาดของการรู้จำคำ เมื่อนำเอาความรู้ในเรื่องของแนวคิดมาใช้บนกราฟของเซกเมนต์

จากกราฟจะเห็นว่าข้อผิดพลาดที่เกิดมากที่สุด ทั้งการรู้จำคำโดยใช้ความรู้ของแนวคิดหรือไม่นั้น จะมาจากข้อผิดพลาดที่รู้จำคำผิด ซึ่งหากไม่ใช้ความรู้ของแนวคิดมาช่วยในการรู้จำคำจะมีข้อผิดพลาดในส่วนนี้ถึง 21% โดยที่ค่าความผิดพลาดนี้จะลดลงเหลือ 16% เมื่อนำเอาความรู้ของแนวคิดมาใช้ เนื่องจากการใช้ความรู้ของแนวคิดมาช่วยนั้น จะช่วยในการเลือกคำที่เหมาะสม

โดยพิจารณาจากความถี่ของลำดับของแนวคิดที่เกิดขึ้นในตัวอย่างเสียงที่ใช้ในการเรียนรู้ เช่น ในตัวอย่างเสียงทดสอบที่มีคำว่า “รหัส ทางไกล” แต่ผลของการรู้จำนั้น ได้คำว่า “รหัส เท่าไหร่” เนื่องจากคำว่า “ทางไกล” และ “เท่าไหร่” นั้น มีความคล้ายกันทำให้ได้ผลการรู้จำนั้นผิดพลาด ซึ่งเมื่อพิจารณาจากตัวอย่างข้อมูลที่ใช้ในการเรียนรู้แล้ว จะพบว่าแนวคิด (l_code, c_code) ของคำว่า “รหัส” นั้น จะมีค่าความน่าจะเป็นที่จะตามด้วยแนวคิด (l_cnt, c_code) ของคำว่า “ทางไกล” มากกว่าแนวคิด (l_H_question, C_H_question) ของคำว่า “เท่าไหร่” มากกว่า ดังนั้นเมื่อนำเอาผลคะแนน หรือค่าความน่าจะเป็นจากความรู้ของแนวคิดมาใช้ ก็จะเป็นการเพิ่มน้ำหนักให้ผลของการรู้จำไปในทางที่มีความน่าจะเป็นที่จะเกิดขึ้นมากกว่า ทำให้ได้ผลการรู้จำมีความถูกต้องมากยิ่งขึ้น

กล่าวโดยสรุปได้ว่า หากพิจารณาว่าความรู้ในด้านของหน่วยเสียง คือ ความรู้ที่ได้จากแบบจำลองเวกเตอร์ลักษณะสำคัญของค่าสังเกต และแบบจำลองระยะเวลา และพิจารณาว่าโครงข่ายจับคู่ระหว่างหน่วยเสียงและคำศัพท์ และ แบบจำลองภาษา เป็นความรู้ในด้านของภาษา โดยในการรู้จำนั้น จะนำเอาคะแนนจากทั้ง 2 ส่วนมาใช้ในการพิจารณาลำดับของคำ การเพิ่มความรู้ของแนวคิดลงไปนั้น จะเปรียบเสมือนการเพิ่มน้ำหนักในส่วนของความรู้ในด้านภาษา นั่นคือ เป็นการเพิ่มน้ำหนักให้กับลำดับของคำที่ตรงกับไวยากรณ์ หรือมีความถี่ในการเกิดขึ้นสูง ทำให้ได้ผลการรู้จำไปในทางที่ถูกไวยากรณ์มากขึ้น

เมื่อพิจารณาค่าของความผิดพลาดที่เกิดจากการแบ่งคำผิด (Bet) ข้อผิดพลาดที่เกิดจากการรู้จำคำบางคำหายไป (Del) จะเห็นว่าหากใช้ความรู้ทางด้านแนวคิดมาใช้ในการรู้จำ จะมีเปอร์เซ็นต์ความผิดพลาดของทั้ง 2 กรณี รวมกันเป็น 14% และหากไม่ใช้ความรู้ของแนวคิดมาช่วย จะมีเปอร์เซ็นต์ความผิดพลาดเป็น 25% ทั้งนี้เนื่องจากว่า ความรู้ของแนวคิด ได้แก่ โครงข่ายที่จับคู่ระหว่างคำศัพท์และแนวคิด และแบบจำลองแนวคิดนั้น จะเป็นความรู้ในระดับของคำศัพท์ ซึ่งค่าน้ำหนักของเส้นเชื่อมนั้น จะเป็นค่าความน่าจะเป็นในการเปลี่ยนคำศัพท์จากแนวคิดหนึ่ง ไปยังอีกแนวคิดหนึ่ง ดังนั้นในระหว่างการค้นหาลำดับของคำนั้น เมื่อมีการเปลี่ยนจากสถานะคำศัพท์หนึ่ง ไปยังอีกคำศัพท์หนึ่ง จึงมีการใช้ค่าน้ำหนักนี้เข้าไปพิจารณาด้วยส่งผลให้ค่าความน่าจะเป็นของลำดับของคำก็ลดลง เนื่องจากเครื่องรู้จำคำนั้นมีแนวโน้มที่จะให้ผลการรู้จำคำเป็นคำที่มีขนาดสั้น ดังนั้นการใช้ความรู้ของแนวคิดมาช่วยจะทำให้มีโอกาสที่จะได้คำศัพท์ยาวๆ มากขึ้น ในขณะที่หากไม่มีค่าน้ำหนักจากความรู้ของแนวคิดเข้ามาช่วยนั้น ก็จะมีแนวโน้มที่จะได้คำศัพท์สั้นๆ หลายๆ คำต่อกัน ทำให้เกิดข้อผิดพลาดในการแบ่งคำผิดมากขึ้น เช่น รู้จำคำว่า “ค่าบริการ” เป็นคำว่า “ค่ะ” “อเมริกา” เป็นต้น นอกจากนี้เนื่องจากในแนวคิดขยะซึ่งจะประกอบไปด้วยหน่วยเสียงเดียว หรือคำศัพท์สั้นๆ จึงมีแนวโน้มที่จะรู้จำคำบางคำเป็นคำศัพท์ในแนวคิดขยะ ซึ่งทำให้เกิดข้อผิดพลาดในการรู้จำคำหายนั่นเอง

เนื่องจากแนวคิดนั้นจะเป็นกลุ่มของคำที่มีความหมายเดียวกัน ดังนั้นข้อดีอีกอย่างในการใช้ความรู้ของแนวคิดมาช่วยนั้น คือ การช่วยแก้ปัญหาจำนวนตัวอย่างข้อมูลที่ใช้ในการเรียนรู้น้อยสำหรับคำศัพท์บางคำได้

## 5.6 การทดลองในการค้นหาลำดับของแนวคิด

ในการทดลองนี้ จะเปรียบเทียบประสิทธิภาพการค้นหาลำดับของแนวคิด โดยจะเปรียบระหว่างวิธีการนำเอาผลการรู้จำลำดับของคำจากเครื่องรู้จำเสียงพูด มาค้นหาลำดับของแนวคิดโดยใช้แบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์ ซึ่งได้อ้างอิงจากงานวิจัยของ Wuttiwiwatchai และ Furui ในหัวข้อที่ 4.2 กับการค้นหาแนวคิดไปพร้อมๆ กับรู้จำเสียงพูด ซึ่งทำโดยการนำเอาแบบจำลองแนวคิด และโครงข่ายการจับคู่จากคำศัพท์กับแนวคิด มาร่วมพิจารณาในการสร้างโครงข่ายการออกเสียง เพื่อพิจารณาความถูกต้องของผลลัพธ์ของลำดับของแนวคิดที่ได้

### 5.6.1 ผลการทดลอง

เนื่องจากกำหนดให้คำศัพท์ที่มีในพจนานุกรมทุกคำนั้น จะมีแนวคิดของคำศัพท์นั้นกำกับอยู่ โดยอาจมีบางคำที่สามารถมีแนวคิดมากกว่า 1 แนวคิด ในการวัดประสิทธิภาพนั้น ได้วัดโดยใช้ค่าพรีซีชัน (Precision) และรีคอล (Recall) ของลำดับแนวคิดที่ได้ ซึ่งในการพิจารณาค่าดังกล่าวนั้น จะเทียบจากแนวคิดแต่ละแนวคิดที่ได้ ว่าได้ผลการรู้จำแนวคิดในตำแหน่งที่อ้างอิงถูกต้องหรือไม่ โดยจะไม่นำผลของแนวคิดขยะมาร่วมพิจารณา ผลการทดลองนั้นจะแสดงในตารางที่ 5-4

ตารางที่ 5-4 ค่าพรีซีชันและรีคอลในการค้นหาแนวคิดจากวิธีต่างๆ

| การค้นหาแนวคิด | ค่ารีคอล (%) | ค่าพรีซีชัน (%) |
|---|---|---|
| ค้นหาจากผลการรู้จำจากเครื่องรู้จำ โดยการใช้แบบจำลองความน่าจะเป็นและกฎทางไวยากรณ์ | 67.94 | 83.39 |
| ค้นหาจากกราฟของเซกเมนต์ | 81.73 | 86.67 |

### 5.6.2 วิเคราะห์ผลการทดลอง

เมื่อพิจารณาข้อผิดพลาดที่เกิดขึ้นในแต่ละแนวคิด โดยจะพิจารณาจากการรู้จำคำที่ผิดพลาดของแนวคิดนั้น และ ความถูกต้องของในการระบุแนวคิด จากการค้นหาแนวคิดจากผลการรู้จำคำที่ได้จากเครื่องรู้จำโดยใช้แบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์ โดยพิจารณาจากลำดับของคำที่ให้ผลการรู้จำที่ดีที่สุดลำดับแรก และการใช้ความรู้ของแนวคิดมาใช้ในการค้นหาคำและแนวคิดจากกราฟของเซกเมนต์ไปพร้อมกัน โดยสามารถแสดงได้ในตารางที่ 5-5 และ ตารางที่ 5-6 ดังนี้

ตารางที่ 5-5 ความถูกต้องในการรู้จำคำ และการค้นหาแนวคิดในแต่ละแนวคิด เมื่อค้นหาแนวคิดจากลำดับของคำที่ได้จากการรู้จำ

| แนวคิด | ความถูกต้องของการรู้จำคำ (%) | ความถูกต้องในการค้นหาแนวคิด (%) |
|---|---|---|
| (l_cnt_name, c_cnt_name) | 27.11 | 42.37 |
| (l_Q_question, c_Q_question) | 69.23 | 76.92 |
| (l_cnt, c_cnt_name) | 23.08 | 23.08 |
| (l_code, c_code) | 77.78 | 77.78 |
| (l_rate, c_rate) | 60.71 | 60.71 |
| เฉลี่ยรวมของทุกแนวคิด | 54.93 | 67.94 |

ตารางที่ 5-6 ความถูกต้องในการรู้จำคำ และการค้นหาแนวคิดในแต่ละแนวคิด เมื่อค้นหาแนวคิดจากลำดับจากกราฟของเซกเมนต์

| แนวคิด | ความถูกต้องในการรู้จำคำ(%) | ความถูกต้องในการค้นหาแนวคิด (%) |
|---|---|---|
| (l_cnt_name, c_cnt_name) | 46.55 | 58.62 |
| (l_Q_question, c_Q_question) | 61.54 | 84.62 |

| แนวคิด | ความถูกต้องในการรู้จำคำ(%) | ความถูกต้องในการค้นหาแนวคิด (%) |
|---|---|---|
| (l_cnt, c_cnt_name) | 55.56 | 55.56 |
| (l_code, c_code) | 77.78 | 81.48 |
| (l_rate, c_rate) | 90.00 | 90.00 |
| เฉลี่ยรวมของทุกแนวคิด | 70.59 | 81.73 |

ในตารางที่ 5-5 และ ตารางที่ 5-6 นั้น ค่าความถูกต้องในการรู้จำคำของแต่ละแนวคิด เช่น หากพิจารณาค่าความถูกต้องในการรู้จำคำของแนวคิด (l_rate, c_rate) ซึ่งเป็นแนวคิดของคำศัพท์ที่บอกถึงอัตราค่าบริการ เช่น “อัตรา” “ราคา” และ “ค่าบริการ” เป็นต้น ค่าความถูกต้องนั้นจะเป็นจำนวนที่คำศัพท์ดังกล่าวให้ผลการรู้จำที่ถูกต้องเทียบกับจำนวนที่คำศัพท์ในแนวคิดนี้เกิดขึ้นจริงทั้งหมด สำหรับความถูกต้องในการค้นหาแนวคิดนั้น จะเป็นอัตราส่วนระหว่างจำนวนแนวคิดที่สามารถค้นหาได้ถูกต้อง ต่อจำนวนที่แนวคิดนั้นปรากฏขึ้นจริงทั้งหมด โดยทั้งค่าความถูกต้องในการรู้จำคำ และ ค่าความถูกต้องในการรู้จำแนวคิดนั้น จะเปรียบเทียบจากในตัวอย่างข้อมูลที่มีการระบุคำศัพท์ที่ถูกต้องแล้ว เพื่อเปรียบเทียบว่า ณ ตำแหน่งที่เกิดแนวคิดนั้นขึ้นจริง การค้นหาแนวคิดด้วยวิธีการที่เสนอนั้น ให้ผลที่คลาดเคลื่อนอย่างไร

จากตารางที่ 5-5 และตารางที่ 5-6 นั้น เมื่อพิจารณาที่แนวคิดของชื่อประเทศ หรือ (l_cnt_name, c_cnt_name) จะเห็นว่ามีความถูกต้องในการรู้จำคำศัพท์ค่อนข้างน้อย ทั้งนี้เนื่องจากคำศัพท์ในกลุ่มนี้มีจำนวนมากที่สุด และยังมีคำศัพท์ที่มีความคล้ายคลึงกับคำศัพท์ทั้งที่เป็นคำศัพท์จากกลุ่มของแนวคิดอื่น หรือภายในแนวคิดเดียวกันค่อนข้างมาก นอกจากนี้คำศัพท์ในกลุ่มแนวคิดประเภทนี้นั้น จะประกอบด้วยหน่วยเสียงที่มาจากภาษาต่างประเทศค่อนข้างมาก (เช่น /s^/ ซึ่งเป็นเสียงตัวสะกดในภาษาอังกฤษ เช่น คำว่า “ออส” จากคำว่าออสเตรเลีย) ซึ่งเป็นหน่วยเสียงที่มีข้อมูลเรียนรู้ค่อนข้างน้อย ดังนั้นจึงให้การรู้จำที่ผิดพลาดค่อนข้างสูง

จากตารางที่ 5-5 และตารางที่ 5-6 จะเห็นว่าค่าความถูกต้องของการรู้จำแนวคิดเมื่อค้นหาแนวคิดจากกราฟของเซกเมนต์ไปพร้อมกับการรู้จำคำนั้น จะให้ความถูกต้องของการรู้จำแนวคิดของแต่ละแนวคิดมากกว่าการค้นหาแนวคิดจากผลลัพธ์ของการรู้จำคำที่ดีที่สุดลำดับแรก ซึ่งหากพิจารณาจากเปอร์เซ็นต์เฉลี่ยโดยรวมแล้วจะพบว่าการค้นหาแนวคิดจากกราฟของเซกเมนต์จะให้ความถูกต้องถึง 81.73% ในขณะที่การค้นหาแนวคิดจากลำดับของคำที่ได้จากการ

รู้จำโดยการใช้แบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์นั้นจะให้ความถูกต้องเพียง 67.94%

ในบางแนวคิดที่คำศัพท์ในแนวคิดนั้นสามารถจัดอยู่ในแนวคิดประเภทเดียว ค่าผลต่างของความถูกต้องในการรู้จำแนวคิด และค่าความถูกต้องในการรู้จำคำในแต่ละแนวคิดจะเป็นการบอกอัตราส่วนของการเกิดกรณีที่ให้ผลการรู้จำผิดพลาด แต่คำที่ผิดพลาดนั้นยังคงอยู่ในประเภทของแนวคิดเดียวกันต่อจำนวนคำศัพท์ที่ปรากฏขึ้นจริงทั้งหมด ซึ่งจะเป็นการบอกได้ถึงบทบาทของการใช้ความรู้ของแนวคิดในการเลือกเส้นทาง ดังเช่น ในกรณีของแนวคิด (l_Q_question, c_Q_question) หรือ แนวคิดประเภทคำถามที่ใช้ในการสอบถามรหัสโทรศัพท์ทางไกลระหว่างประเทศ ซึ่งคำศัพท์ในแนวคิดนั้นจะสามารถจัดอยู่ในประเภทแนวคิดนี้ได้เพียงอย่างเดียว เมื่อพิจารณาผลของการรู้จำคำศัพท์ในกลุ่มคำนี้ จะพบว่าการใช้ความรู้ของแนวคิดมาช่วยนั้น จะให้อัตราส่วนความถูกต้องน้อยกว่าการที่ไม่ใช้ความรู้ของแนวคิด แต่อย่างไรก็ตามเมื่อพิจารณาจากความถูกต้องในการค้นหาแนวคิดนั้น การค้นหาแนวคิดจากกราฟของเซกเมนต์โดยใช้ความรู้ของแนวคิดเข้ามาช่วยนั้น จะมีอัตราส่วนความถูกต้องมากกว่า ทั้งนี้เนื่องจาก คำศัพท์ในกลุ่มของแนวคิดนี้ค่อนข้างจะคล้ายกัน เช่น “อะไร” “แบบไหน” หรือ “อย่างไร” รวมถึงมีความคล้ายกับแนวคิดอื่น เช่น “เท่าไหร่” ซึ่งเป็นคำศัพท์ในกลุ่มของแนวคิดที่เป็นคำถามในการถามอัตราค่าบริการโทรศัพท์ ทำให้มีแนวโน้มที่จะรู้จำคำผิดเป็นคำอื่น แต่เนื่องจากเมื่อได้เพิ่มความรู้แนวคิดเข้าไป ก็จะช่วยให้ในการรู้จำคำนั้น ไปในทางที่มีลำดับของแนวคิดที่มีความน่าจะเป็นสูง เช่น คำศัพท์คำว่า “แบบไหน” นั้น หากรู้จำจากเครื่องรู้จำที่ไม่ใช่ความรู้ของแนวคิดมาใช้นั้น อาจจะได้ผลเป็นคำว่า “เท่าไหร่” แต่หากรู้จำด้วยเครื่องรู้จำที่ใช้ความรู้ของแนวคิดเข้ามานั้น ก็จะได้ผลการรู้จำเป็น “อะไร” ซึ่งถึงแม้ว่าผลการรู้จำคำศัพท์นั้นจะผิดทั้งใน 2 กรณี แต่เมื่อใช้ความรู้แนวคิดมาช่วยในขณะการรู้จำ ก็จะมีแนวโน้มที่จะได้ผลการรู้จำเป็นคำศัพท์ที่อยู่ในแนวคิดประเภทเดียวกัน ซึ่งในกรณีนี้จะเป็นข้อดีสำหรับค้นหาแนวคิดประเภทที่ไม่ต้องการค่า เนื่องจากถึงแม้ว่าจะให้ผลการรู้จำคำที่ผิดนั้น แต่ยังคงให้ความหมายที่ได้จากแนวคิดถูกต้อง

นอกจากนี้ ในการระบุแนวคิดนั้น ความถูกต้องของการรู้จำคำศัพท์จะมีผลต่อการระบุแนวคิดด้วยเช่นกัน โดยหากมีการรู้จำคำศัพท์ถูกต้องค่อนข้างสูง ก็จะส่งผลให้การค้นหาลำดับของแนวคิดมีแนวโน้มที่จะให้ความถูกต้องสูงด้วยเช่นกัน ดังนั้นเมื่อพิจารณาความถูกต้องของการรู้จำคำในการทดลองการรู้จำเสียงพูดในหัวข้อที่ 5.5 นั้น จะเห็นว่าการค้นหาแนวคิดไปพร้อมกับการรู้จำนั้นจะให้ผลการรู้จำคำศัพท์ที่ดีกว่าเมื่อพิจารณาจากค่าพรีซีชันและรีคอล ดังนั้นจึงส่งผลให้การค้นหาลำดับแนวคิดนั้น มีความถูกต้องมากด้วยเช่นกัน

จากที่กล่าวมาขั้นต้น จึงส่งผลให้ค่าพรีซีชันและรีคอลในการใช้การค้นหาแนวคิดจากกราฟของเซกเมนต์เมื่อใช้ความรู้ของแนวคิดมาช่วยในตารางที่ 5-4 มีค่ามากกว่าการค้นหาแนวคิดจากผลของการรู้จำคำ

อย่างไรก็ตามในกรณีที่พิจารณาการค้นหาแนวคิดจากผลการรู้จำโดยใช้กฎไวยากรณ์และแบบจำลองความน่าจะเป็นนั้น เมื่อพิจารณาจากการค้นหาแนวคิดจากผลลัพธ์ของการรู้จำคำที่ดีที่สุด N อันดับแรก โดยในการทดลองนี้ได้เลือกผลการรู้จำคำที่ดีที่สุด 10 อันดับแรกแทนการนำเอาผลลัพธ์ที่ดีที่สุดเพียงลำดับเดียวมาพิจารณาดังที่แสดงในตารางที่ 5-4 นั้น ผลลัพธ์ของการค้นหาลำดับของแนวคิด จะแสดงได้ในตารางที่ 5-7

ตารางที่ 5-7 ค่าพรีซีชันและรีคอลในการค้นหาแนวคิดจากผลลัพธ์จากการรู้จำคำของเครื่องรู้จำคำที่ดีที่สุด 10 อันดับแรก

| การค้นหาแนวคิด | ค่ารีคอล (%) | ค่าพรีซีชัน (%) |
|---|---|---|
| ค้นหาจากผลการรู้จำที่ดีที่สุด 10 อันดับแรกจากเครื่องรู้จำ โดยการใช้แบบจำลองความน่าจะเป็นและกฎทางไวยากรณ์ | 72.72 | 75.77 |

จากผลการทดลองในตารางที่ 5-4 และตารางที่ 5-7 นั้น จะเห็นว่าเมื่อเปรียบเทียบผลลัพธ์ของการค้นหาแนวคิดจากลำดับของคำที่ได้จากเครื่องรู้จำเมื่อพิจารณาจากลำดับของคำที่ดีที่สุด 10 อันดับแรกนั้น จะให้ค่ารีคอลเพิ่มจาก 67.94% มาเป็น 72.72% ทั้งนี้เนื่องจากการใช้กฎทางไวยากรณ์ และแบบจำลองความน่าจะเป็นนั้น จะเลือกลำดับของคำที่มีแนวโน้มที่จะเกิดขึ้นบ่อยและถูกต้องตามไวยากรณ์มากที่สุด แต่อย่างไรก็ตาม เนื่องจากเหตุผลดังกล่าวส่งผลให้ระบบพยายามที่จะเลือกลำดับของคำที่มีมีคะแนนของแนวคิดมากที่สุด แม้ว่าจะไม่เกิดคำนั้นขึ้นจริงๆ ก็ตาม เช่น ในประโยค “เอ่อ..โทรไปลาว” โดยที่คำว่า “เอ่อ” นั้น จึงถือว่าเป็นคำศัพท์ในแนวคิดขยะ แต่ในผลการรู้จำคำในจำนวน 10 ลำดับแรกนั้น อาจจะได้ผลการรู้จำเป็น “เบอร์โทรไปลาว” โดยที่คำว่า “เบอร์” นั้น จะเป็นคำศัพท์ในแนวคิด (l_code, c_code) ซึ่งเมื่อเปรียบเทียบคะแนนของลำดับแนวคิดของประโยคทั้งสองนั้น ประโยคที่ว่า “เบอร์โทรไปลาว” จะมีค่าคะแนนของการค้นหาแนวคิดมากกว่า เนื่องจากกระบวนการในการให้คะแนนนั้นจะให้คะแนนกับความยาวของลำดับของคำที่ตรงกับไวยากรณ์มากกว่าลำดับของคำสั้นที่สั้นกว่า ดังนั้นระบบจึงมีแนวโน้มที่จะเลือกผลของการรู้จำว่า “เบอร์โทรไปลาว” ทั้งๆ ที่คำว่า “เบอร์” นั้นไม่ได้เกิดขึ้นจริง หรืออาจมองว่าคำว่า

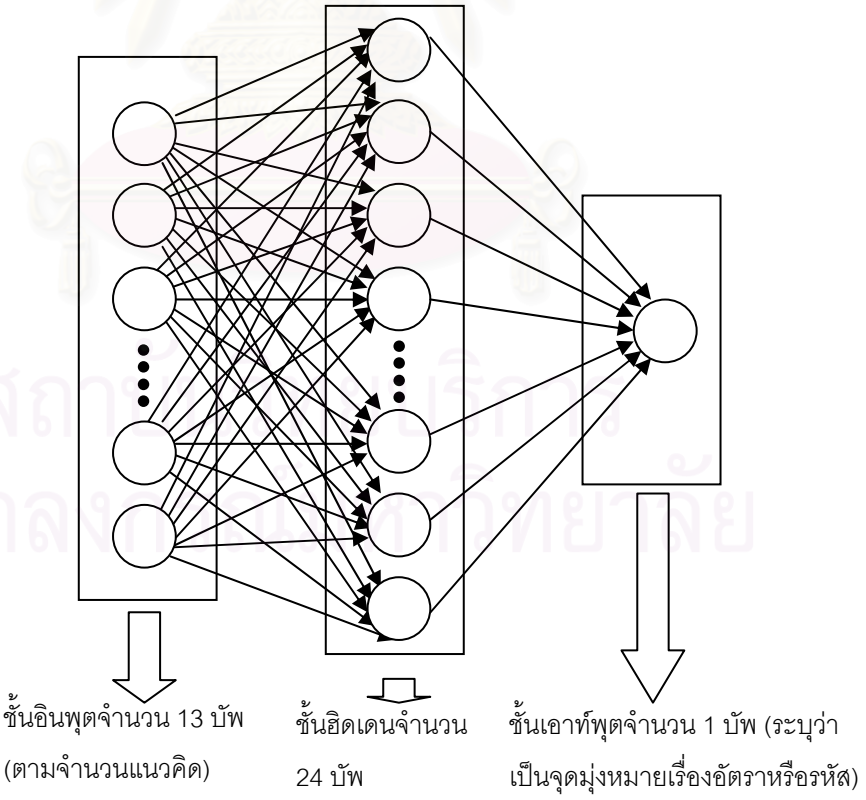
“เบอร์” นั้นเป็นข้อผิดพลาดของการรู้จำเกินมา ด้วยเหตุผลดังกล่าวจึงส่งผลให้ค่าของพรีซีชันนั้นลดลงจาก 83.39% มาเป็น 75.77%

## 5.7 การทดลองในการแบ่งประเภทจุดมุ่งหมาย

ในการทดลองนี้ จะเปรียบเทียบความถูกต้องของการระบุจุดมุ่งหมาย โดยการระบุจุดมุ่งหมายนั้น จะได้จากการนำเอาลำดับของแนวคิดที่ได้จากการทดลองในหัวข้อที่ 5.6 มาทำการแบ่งประเภทของจุดมุ่งหมาย โดยที่ลำดับของแนวคิดนั้น ได้จากการค้นหาแนวคิดจากผลการรู้จำคำศัพท์ที่ได้จากเครื่องรู้จำ โดยใช้แบบจำลองค่าความน่าจะเป็นร่วมกับกฎไวยากรณ์ และจากการใช้ความรู้ของแนวคิดมาพิจารณาเพื่อค้นหาลำดับของแนวคิดในขณะที่ทำการรู้จำคำบนกราฟของเซกเมนต์ สำหรับจุดมุ่งหมายนั้น ได้แก่

- จุดมุ่งหมายของการสอบถามเรื่องอัตราค่าบริการโทรศัพท์
- จุดมุ่งหมายของการสอบถามเรื่องรหัสโทรศัพท์ทางไกล
- การไม่ให้ผลลัพธ์ในกรณีที่ประโยคที่พิจารณานั้นยังไม่สามารถระบุจุดมุ่งหมายได้

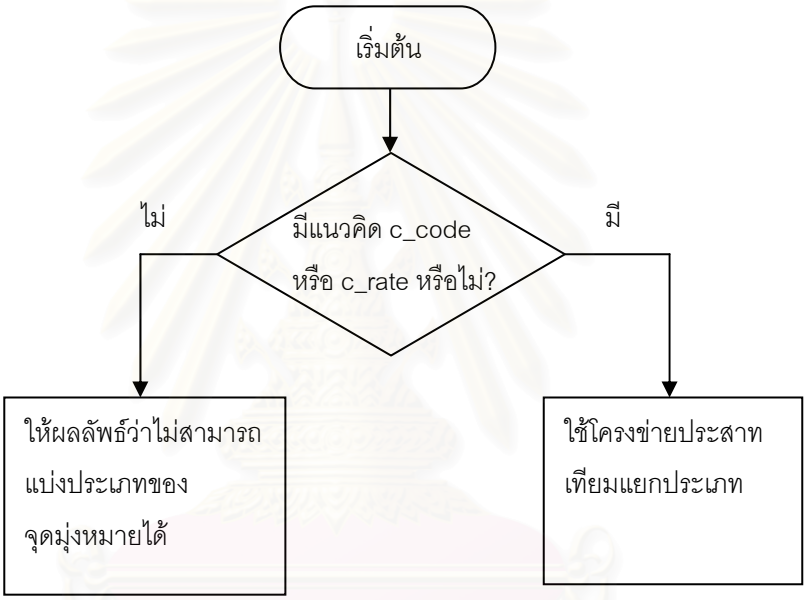
โดยลักษณะของโครงข่ายประสาทเทียมในการแบ่งประเภทจุดมุ่งหมาย 2 ประเภท (การสอบถามรหัสโทรศัพท์ และอัตราค่าบริการ) ที่ใช้ในการทดลองนั้น จะสามารถแสดงได้ดังรูปที่ 5-5




รูปที่ 5-5 รูปแบบของโครงข่ายประสาทเทียมที่ใช้ในการทดลองเพื่อแบ่งประเภทจุดมุ่งหมาย

ในรูปที่ 5-5จะแสดงรูปแบบของโครงข่ายประสาทเทียมที่ใช้ ซึ่งจะใช้เป็นโครงข่ายประสาทเทียม 3 ชั้น มีจำนวนบัพของอินพุตเป็น 13 บัพตามจำนวนของแนวคิด และจำนวนบัพของชั้นฮิดเดนมีจำนวน 24 บัพ โดยผลลัพธ์ที่ได้จากชั้นเอาท์พุตนั้น จะใช้ในการพิจารณาว่าประโยคนั้นมีประเภทของจุดมุ่งหมายในเรื่องของการสอบถามอัตราค่าบริการ หรือรหัสโทรศัพท์ทางไกล

ในการพิจารณาว่าประโยคดังกล่าว สามารถระบุจุดมุ่งหมายได้หรือไม่นั้น จะพิจารณาว่าประโยคนั้นมีแนวคิด c_code หรือ c_rate อยู่หรือไม่ หากไม่มีแนวคิดดังกล่าวอยู่ ก็จะถือว่าประโยคนั้นยังไม่สามารถระบุประเภทของจุดมุ่งหมายได้ โดยขั้นตอนการทำงานจะแสดงได้ดังรูปที่ 5-6




รูปที่ 5-6 ขั้นตอนในการแบ่งประเภทเมื่อได้รับลำดับของแนวคิดมา

นอกจากนี้ ในการทดลองนี้ยังได้เสนอแนวทางในการระบุจุดมุ่งหมาย โดยได้นำเอาคะแนนของแนวคิดที่ได้จากการรู้จำคำและแนวคิดจากกราฟของเซกเมนต์มาพิจารณาด้วย โดยได้ทำการทดลองเปรียบเทียบวิธีการแบ่งประเภทจุดมุ่งหมายโดยใช้ค่าคะแนนดังกล่าวออกเป็น 2 วิธี คือ การใช้โครงข่ายประสาทเทียม และการใช้กฎ

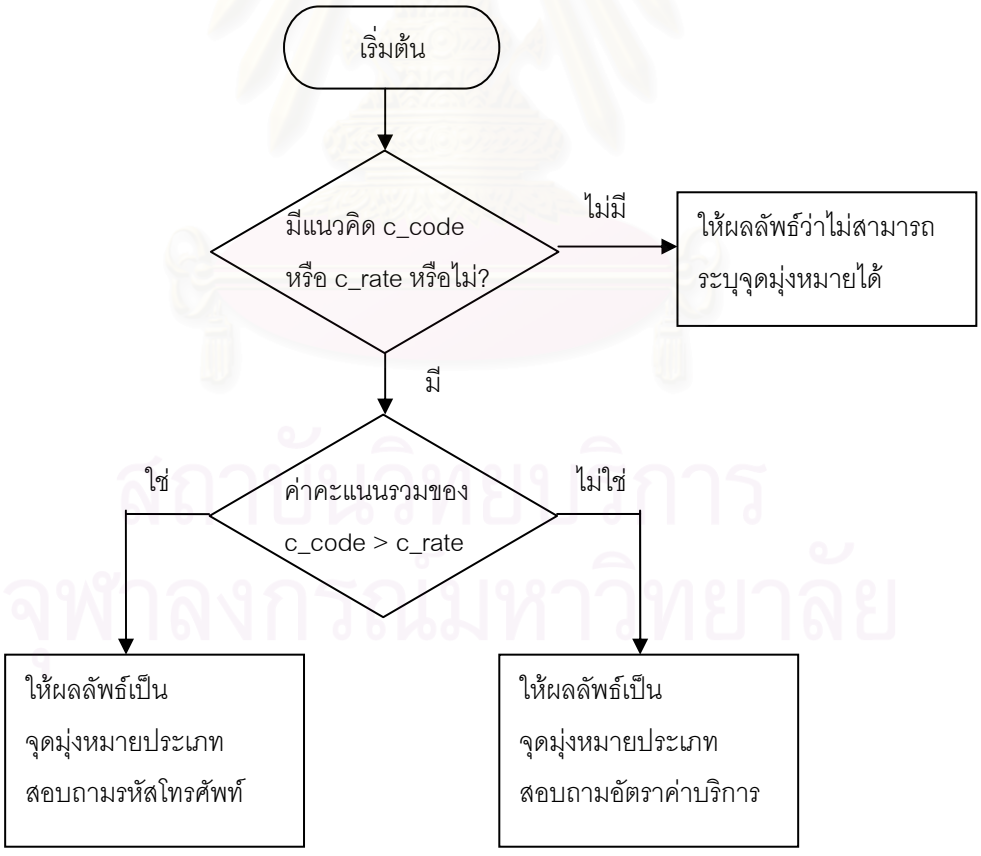
การใช้โครงข่ายประสาทเทียมโดยนำเอาคะแนนการรู้จำมาพิจารณานั้น จะพัฒนาจากเดิมที่รับอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดต่างๆ มาใช้ในการพิจารณา มาเป็นเวกเตอร์ของค่าคะแนนรวมของแต่ละแนวคิดที่ได้แทน โดยค่าคะแนนดังกล่าวนั้น จะได้มาจากผลรวมของล็อคการิทึมของคะแนนที่ได้จาก $p(C|A)$ สำหรับแต่ละแนวคิด เช่น หากกำหนดให้เกิดแนวคิด $c_n$ ของคำที่ $w_m$ ซึ่งประกอบด้วยลำดับของหน่วยเสียง $U_{w_m} = u_i, u_{i+1}, ..., u_{i+j}$ โดยกำหนดให้คำ $w_m$ นั้น เกิด

ในช่วงของค่าสังเกตุ $A_{w_m} = a_t, a_{t+1}, ..., a_{t+u}$ ของลำดับของเซกเมนต์ $S_{w_m} = s_t, s_{t+1}, ..., s_{t+u}$ ค่าคะแนนของแนวคิด $c_n$ นั้น จะมาจาก

$$score\,of\,(w_m, c_n) = \log \frac{p(A_{w_m} | U_{w_m}) p(S_{w_m} | U_{w_m}) p(U_{w_m} | w_m) p(w_n | w_{n-1}) p(w_m | c_n) p(c_n | c_{n-1})}{p(w_m)} \quad (5.1)$$

ในสมการที่ (5.1) นั้นจะอ้างอิงมาจาก การหาค่าความน่าจะเป็นของการค้นหาแนวคิด ซึ่งได้แสดงในสมการที่ (4.5) และ (4.6) ซึ่งโครงข่ายประสาทเทียมที่ใช้ในการแบ่งประเภทโดยใช้ค่าคะแนนดังกล่าว ในการทดลองนี้จะกำหนดให้มีชั้นเอาท์พุตจำนวน 3 บัพ ตามจำนวนของจุดมุ่งหมายที่มีในระบบ โดยในการพิจารณาจุดมุ่งหมายสุดท้ายนั้น จะพิจารณาโดยการนำเอาคะแนนทั้ง 3 บัพที่ได้จากชั้นเอาท์พุตมาเปรียบเทียบกัน

สำหรับการพิจารณาจุดมุ่งหมายโดยการใช้กฎจากคะแนนของการรู้จำแนวคิดนั้น จะพิจารณาโดยการเปรียบเทียบผลรวมคะแนนของแนวคิดรหัสประเทศ คือ c_code กับผลรวมของคะแนนของแนวคิดอัตราค่าบริการโทรศัพท์ ได้แก่ c_rate โดยคะแนนดังกล่าวนั้น จะอ้างอิงมาจากสมการที่ (5.1) ลักษณะการพิจารณาของกฎจะแสดงได้ดังรูปที่ 5-7




รูปที่ 5-7 ลักษณะการใช้กฎในการพิจารณาแบ่งประเภทจุดมุ่งหมาย

### 5.7.1 ผลการทดลอง

ในผลการทดลองนั้น ได้แยกการพิจารณาความถูกต้องออกเป็น 2 ส่วน ดังนี้

- ความถูกต้องในการระบุจุดมุ่งหมายของประโยค หรือ อัตราส่วนของข้อมูลทดสอบที่สามารถแบ่งประเภทจุดมุ่งหมายได้ถูกต้อง ต่อจำนวนตัวอย่างทดสอบทั้งหมด รวมถึงการระบุได้ว่าประโยคใดที่ยังไม่สามารถระบุจุดมุ่งหมายได้

- ความถูกต้องในการให้ผลลัพธ์สุดท้ายของทั้งการระบุจุดมุ่งหมายและชื่อประเทศของประโยคได้ถูกต้องในกรณีที่ประโยคนั้นสามารถระบุจุดมุ่งหมายได้ โดยกำหนดให้ค่าของจุดมุ่งหมายและชื่อประเทศนั้น จะใช้การพิจารณาการตอบกลับไปยังผู้ใช้งานของระบบการให้ข้อมูลในเรื่องของอัตราค่าบริการและรหัสโทรศัพท์ รวมถึงให้ผลลัพธ์ว่าไม่ให้จุดมุ่งหมายเมื่อโยคนั้นไม่สามารถระบุจุดมุ่งหมายได้

ในการทดลองนั้น จะทดลองกับประโยคทดสอบจำนวน 216 ประโยค ซึ่งผลการทดลองนั้นจะแสดงในตารางที่ 5-8

ตารางที่ 5-8 ค่าความถูกต้องในการระบุจุดมุ่งหมาย

| ลำดับของแนวคิดที่ใช้ในการระบุจุดมุ่งหมาย | ความถูกต้องในการระบุจุดมุ่งหมาย (%) | ความถูกต้องในการระบุจุดมุ่งหมาย และให้ค่าชื่อประเทศ (%) |
|---|---|---|
| ลำดับของแนวคิดที่ค้นหาจากผลการรู้จำคำที่ดีที่สุด 10 อันดับแรก โดยการใช้แบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์ | 61.11 | 22.22 |
| ลำดับของแนวคิดที่ค้นหาจากผลการรู้จำคำที่ดีที่สุดลำดับแรก โดยการใช้แบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์ | 66.67 | 24.07 |
| ลำดับของแนวคิดที่ค้นหาจากกราฟของเซกเมนต์ | 74.07 | 40.74 |

เมื่อเปรียบเทียบความถูกต้องของการระบุจุดมุ่งหมาย ระหว่างการใช้โครงข่ายประสาทเทียมที่รับอินพุตเป็นเวกเตอร์ของค่าคะแนนในแต่ละแนวคิดในสมการที่ (5.1) ในการพิจารณาการแบ่งประเภท เปรียบเทียบกับการใช้กฎในการแบ่งประเภท ดังแสดงในรูปที่ 5-7 ผลการทดลองจะแสดงในตารางที่ 5-9

ตารางที่ 5-9 ค่าความถูกต้องในการระบุจุดมุ่งหมาย เมื่อใช้โครงข่ายประสาทเทียม และใช้กฎ

| การค้นหาจุดมุ่งหมาย | ความถูกต้องในการระบุจุดมุ่งหมาย (%) | ความถูกต้องในการระบุจุดมุ่งหมาย และให้ค่าชื่อประเทศ (%) |
|---|---|---|
| การใช้กฎ | 79.63 | 40.74 |
| โครงข่ายประสาทเทียมแบบหลายชั้น เพื่อให้อินพุตเป็นคะแนนของแต่ละแนวคิด | 77.78 | 44.44 |

### 5.7.2 วิเคราะห์ผลการทดลอง

จากผลการทดลองในตารางที่ 5-8 จะเห็นว่าการระบุจุดมุ่งหมายหรือ การให้ผลลัพธ์สุดท้ายเป็นจุดมุ่งหมายและชื่อประเทศจากลำดับของแนวคิดที่พิจารณาพร้อมๆ กับการรู้จำนั้น จะให้ความถูกต้องมากกว่าการพิจารณาจุดมุ่งหมายจากลำดับของแนวคิดที่ได้จากการค้นหาจากผลการรู้จำคำศัพท์ที่ได้จากการรู้จำ โดยใช้วิธีที่เสนอโดย Wuttiwiwatchai และ Furui ดังที่เสนอในหัวข้อที่ 4.2 ไม่ว่าจะพิจารณาจากผลการรู้จำลำดับของคำที่ดีที่สุด 10 ลำดับแรก หรือจะพิจารณาจากแค่ลำดับของคำจากการรู้จำที่ดีที่สุดเพียงลำดับเดียว ทั้งนี้เนื่องจากเมื่อพิจารณาจากผลการค้นหาลำดับของแนวคิดในตารางที่ 5-4 และ ตารางที่ 5-7 นั้น จะเห็นว่าการรู้จำแนวคิดพร้อมกับการรู้จำคำนั้น จะให้ค่าพรีซีชัน และรีคอล มากกว่า นั่นคือ สามารถระบุแนวคิดที่ถูกต้องมากกว่า (ค่ารีคอล) รวมทั้งมีการระบุแนวคิดที่ผิดพลาดเกินมา (ค่าพรีซีชัน) น้อยกว่า เนื่องจากการแบ่งประเภทจุดมุ่งหมายนั้น จะพิจารณาจากลำดับของแนวคิดเป็นหลัก ดังนั้นการค้นหาลำดับของแนวคิดที่ผิดพลาดนั้น จะส่งผลให้การค้นหาจุดมุ่งหมายนั้นผิดพลาดด้วย เช่น หากในข้อความเสียงที่พูดว่า “ลาว ค่ะ” ซึ่งหากไม่พิจารณาแนวคิดขยะแล้ว ผลของการค้นหาลำดับของแนวคิดจะได้เป็น (l_cnt_name, c_cnt_name) แต่หากผลการรู้จำคำและแนวคิดได้เป็นคำว่า “ราคา” ซึ่งมีแนวคิดเป็น (l_rate, c_rate) ในกรณีนี้อาจจะทำให้การแบ่งประเภทจุดมุ่งหมายนั้น ได้ผลลัพธ์เป็นจุดมุ่งหมายของการสอบถามเรื่องอัตราค่าบริการโทรศัพท์ ทั้งๆ ที่ข้อความเสียงนี้ยังไม่สามารถระบุประเภทของจุดมุ่งหมายได้ เป็นต้น

เมื่อพิจาณาเปรียบเทียบผลการค้นหาจุดมุ่งหมายจากแนวคิดที่พิจารณาจากผลของการรู้จำที่ดีที่สุด 10 อันดับแรก และเลือกลำดับที่ดีที่สุดลำดับเดียวนั้น จะเห็นว่าเมื่อพิจารณาแค่ลำดับของคำที่ดีที่สุดลำดับเดียวจะให้ผลของการค้นหาจุดมุ่งหมายสูงกว่า ทั้งนี้เนื่องจากในผลการทดลองในตารางที่ 5-4 และ ตารางที่ 5-7 จะเห็นว่าเมื่อพิจารณาแค่ผลการรู้จำจากลำดับที่ดีที่สุดจะให้ค่าพรีซีชันของการค้นหาแนวคิดที่มากกว่า หรืออาจพิจารณาได้ว่าให้ผลลัพธ์ที่มีความแม่นยำ และมีข้อผิดพลาดที่เกิดเกินมาน้อยกว่า ส่งผลให้ในการแบ่งประเภทจุดมุ่งหมายโดยใช้โครงข่ายประสาทเทียมนั้น มีความถูกต้องมากขึ้น โดยเมื่อพิจารณาจากอัตราส่วนของข้อมูลทดสอบในจุดมุ่งหมายต่างๆ และข้อผิดพลาดในการแบ่งประเภทของจุดมุ่งหมายนั้น จะสามารถแสดงได้ดังตารางตารางที่ 5-10 ถึง ตารางที่ 5-13

ตารางที่ 5-10 อัตราส่วนของข้อมูลทดสอบในประเภทจุดมุ่งหมายต่างๆ

| จุดมุ่งหมาย | อัตราส่วนของข้อมูลทดสอบ (%) |
|---|---|
| รหัสโทรศัพท์ | 55.55 |
| อัตราค่าบริการ | 33.33 |
| ไม่สามารถระบุจุดมุ่งหมายได้ | 11.12 |

ตารางที่ 5-11 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดจากผลการรู้จำคำที่ดีที่สุด 10 ลำดับแรก

| ประเภทของจุดมุ่งหมายที่เกิดขึ้นจริง | ผลที่ได้จากการแบ่งประเภท | | |
|---|---|---|---|
| | รหัสโทรศัพท์ | อัตราค่าบริการ | ไม่สามารถระบุได้ |
| รหัสโทรศัพท์ | 63.33 | 30.00 | 6.67 |
| อัตราค่าบริการ | 27.78 | 66.67 | 5.55 |
| ไม่สามารถระบุได้ | 50.00 | 16.67 | 33.33 |

ตารางที่ 5-12 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดจากผลการรู้จำคำที่ดีที่สุดลำดับแรก

| ประเภทของจุดมุ่งหมายที่เกิดขึ้นจริง | ผลที่ได้จากการแบ่งประเภท | | |
|---|---|---|---|
| | รหัสโทรศัพท์ | อัตราค่าบริการ | ไม่สามารถระบุได้ |
| รหัสโทรศัพท์ | 76.67 | 20.00 | 3.33 |
| อัตราค่าบริการ | 33.33 | 66.67 | 0.00 |
| ไม่สามารถระบุได้ | 50.00 | 33.33 | 16.67 |

ตารางที่ 5-13 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นไบนารีเวกเตอร์ของแนวคิดพี่พิจารณาพร้อมการรู้จำคำ

| ประเภทของจุดมุ่งหมายที่เกิดขึ้นจริง | ผลที่ได้จากการแบ่งประเภท | | |
|---|---|---|---|
| | รหัสโทรศัพท์ | อัตราค่าบริการ | ไม่สามารถระบุได้ |
| รหัสโทรศัพท์ | 73.33 | 23.33 | 3.33 |
| อัตราค่าบริการ | 11.11 | 88.89 | 0.00 |
| ไม่สามารถระบุได้ | 62.50 | 16.67 | 20.83 |

นอกจากนี้ในการพิจารณาผลลัพธ์สุดท้ายเป็นจุดมุ่งหมาย และชื่อประเทศนั้น จะต้องมีการพิจารณาเหตุการณ์ที่ระบุชื่อประเทศได้ถูกต้องร่วมด้วย ดังนั้นเมื่อพิจารณาผลการรู้จำคำในแนวคิดของชื่อประเทศ ในตารางที่ 5-5 และ ตารางที่ 5-6 จะพบว่าการค้นหาแนวคิดจากกราฟของเซกเมนต์นั้น จะให้ค่าความถูกต้องในส่วนนี้มากกว่า ดังนั้นจึงช่วยสนับสนุนให้ความถูกต้องของผลลัพธ์สุดท้ายนี้มีค่ามากกว่าระบบที่ค้นหาแนวคิดจากผลของการรู้จำคำ

เมื่อทำการค้นหาจุดมุ่งหมายโดยนำเอาค่าคะแนนของการรู้จำแนวคิดนั้นๆ ที่ได้จากวิธีการในการค้นหาแนวคิดพร้อมกับการรู้จำคำจากกราฟของเซกเมนต์มาพิจารณาร่วมด้วย โดยได้เปรียบเทียบระหว่างวิธีการใช้กฎ และการใช้โครงข่ายประสาทเทียมที่รับอินพุตเป็นเวกเตอร์ของ

ค่าคะแนนของแนวคิดต่างๆ จากผลการทดลองในตารางที่ 5-9 จะเห็นว่า การใช้กฎนั้นจะให้ผลในการระบุจุดมุ่งหมายที่มีความถูกต้องมากกว่า โดยเมื่อพิจารณาข้อผิดพลาดในการระบุจุดมุ่งหมายจะสามารถแสดงได้ในตารางที่ 5-14 และ ตารางที่ 5-15

ตารางที่ 5-14 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้กฎในการพิจารณา

| ประเภทของจุดมุ่งหมายที่เกิดขึ้นจริง | ผลที่ได้จากการแบ่งประเภท | | |
|---|---|---|---|
| | รหัสโทรศัพท์ | อัตราค่าบริการ | ไม่สามารถระบุได้ |
| รหัสโทรศัพท์ | 66.67 | 30.30 | 3.030 |
| อัตราค่าบริการ | 4.17 | 94.44 | 1.39 |
| ไม่สามารถระบุได้ | 45.83 | 20.83 | 33.33 |

ตารางที่ 5-15 ข้อผิดพลาดของการระบุจุดมุ่งหมายเมื่อใช้โครงข่ายประสาทเทียมหลายชั้น โดยอินพุตเป็นค่าคะแนนของแต่ละแนวคิด

| ประเภทของจุดมุ่งหมายที่เกิดขึ้นจริง | ผลที่ได้จากการแบ่งประเภท | | |
|---|---|---|---|
| | รหัสโทรศัพท์ | อัตราค่าบริการ | ไม่สามารถระบุได้ |
| รหัสโทรศัพท์ | 86.67 | 4.17 | 9.17 |
| อัตราค่าบริการ | 38.89 | 55.56 | 5.56 |
| ไม่สามารถระบุได้ | 16.67 | 16.67 | 66.67 |

ทั้งนี้สาเหตุที่การแบ่งประเภทโดยใช้โครงข่ายประสาทเทียมนั้นได้ผลการแบ่งประเภทที่มีความถูกต้องน้อยกว่าเมื่อเทียบกับการใช้กฎ เนื่องจากว่าการใช้โครงข่ายประสาทเทียมในการแบ่งประเภทจะค่อนข้างขึ้นอยู่กับข้อมูลที่ใช้เรียนรู้ โดยเฉพาะในกรณีที่ข้อมูลที่ใช้ในการเรียนรู้ไม่ครอบคลุมอินพุตในทุกๆ กรณีที่สามารถเกิดขึ้นได้ ซึ่งจะส่งผลให้ผลการแบ่งประเภทเกิดความผิดพลาดได้ โดยเฉพาะเมื่ออินพุตที่เข้ามานั้นเป็นผลที่ได้จากการรู้จำ ซึ่งสามารถเกิดข้อผิดพลาดที่ไม่สามารถคาดการณ์ได้ล่วงหน้าต่างๆ ได้นั่นเอง

แต่อย่างไรก็ตาม ไม่ว่าจะเป็นการใช้กฎ หรือการใช้โครงข่ายประสาทเทียมนั้น จะเห็นว่าเมื่อนำเอาคะแนนของการรู้จำแนวคิดมาพิจารณาด้วยนั้น จะส่งผลให้การแบ่งประเภทของจุดมุ่งหมายถูกต้องมากขึ้น ดังแสดงในผลการทดลองในตารางที่ 5-8 และ ตารางที่ 5-9 จึงอาจกล่าวได้ว่า คะแนนของแนวคิดต่างๆ ที่ได้จากกระบวนการการรู้จำคำและแนวคิดพร้อมๆ กันนั้นจะมีผลต่อการตัดสินใจในการแบ่งประเภทจุดมุ่งหมาย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 6

## สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

### 6.1 ผลการวิจัย

ในวิทยานิพนธ์นี้ได้เสนอแนวทางในการค้นหาลำดับของแนวคิดพร้อมกับการรู้จำคำศัพท์จากกราฟของเซกเมนต์ โดยการเพิ่มความรู้ของแนวคิด ได้แก่ โครงข่ายที่จับคู่ระหว่างคำศัพท์และแนวคิด และแบบจำลองแนวคิด เข้าไปในโครงข่ายการออกเสียง เพื่อที่จะนำโครงข่ายการออกเสียงนี้มาใช้ร่วมกับโครงข่ายหน่วยเสียงในการค้นหาลำดับของคำ และแนวคิดต่อไป มาทำการเปรียบเทียบกับวิธีการค้นหาลำดับของแนวคิดแบบเดิม ที่แยกการค้นหาลำดับของแนวคิด และการรู้จำคำศัพท์ออกจากกัน โดยการนำเอาผลลัพธ์ของการรู้จำคำที่ได้จากเครื่องรู้จำนั้น มาค้นหาลำดับของแนวคิด จากแบบจำลองความน่าจะเป็นร่วมกับกฎทางไวยากรณ์ของภาษา ซึ่งได้อ้างอิงมาจากงานวิจัยของ Wuttiwiwatchai และ Furui

ในวิทยานิพนธ์นี้ได้ตั้งสมมุติฐานว่า การค้นหาลำดับของแนวคิดไปพร้อมๆ กับการรู้จำคำนั้น ซึ่งเป็นการนำเอาความรู้ทั้งในด้านของแนวคิด และความรู้ที่ใช้ในการรู้จำคำร่วมกันในการพิจารณาลำดับของคำศัพท์ และแนวคิด โดยในขณะที่ทำการรู้จำคำนั้น ก็จะใช้ความรู้ของแนวคิดมาช่วยในการพิจารณาลำดับของคำ ซึ่งเปรียบเสมือนการนำเอาความรู้ของความหมายมาช่วยพิจารณา มากกว่าการที่จะพิจารณาแค่เพียงหน่วยเสียง หรือคำศัพท์เพียงอย่างเดียว และการค้นหาแนวคิดนั้น ก็จะได้ใช้ความรู้ทั้งในด้านคำศัพท์ และหน่วยเสียงในการพิจารณาลำดับของแนวคิด ดังนั้นวิธีการดังกล่าวจึงน่าจะให้ได้ผลลัพธ์ในการรู้จำคำศัพท์ และค้นหาแนวคิดมีความถูกต้องมากกว่าในกรณีที่แยกพิจารณาส่วนของการรู้จำคำ และแนวคิดออกจากกัน โดยในการเปรียบเทียบนั้น ได้แบ่งหัวข้อในการเปรียบเทียบเป็นดังนี้

1. การเปรียบเทียบประสิทธิภาพในการรู้จำคำ
2. การเปรียบเทียบประสิทธิภาพในการค้นหาลำดับของแนวคิด
3. การเปรียบเทียบประสิทธิภาพในการระบุจุดมุ่งหมายของประโยค

#### 6.1.1 ผลการวิจัยในการเปรียบเทียบประสิทธิภาพในการรู้จำคำ

ในการทดลองนี้ ได้เปรียบเทียบประสิทธิภาพในการรู้จำคำ โดยใช้ค่าพรีซีชันและรีคอล ในการพิจารณา โดยจากผลการทดลองจะพบว่า การค้นหาแนวคิดไปพร้อมๆ กับการรู้จำคำนั้น จะให้ค่าพรีซีชันและรีคอล มากกว่าการแยกพิจารณาระหว่างการรู้จำคำ และค้นหาแนวคิดออกจากกัน

โดยผลของการทดลองที่ได้นั้น การค้นหาแนวคิดไปพร้อมๆ กับการรู้จำคำศัพท์นั้น จะให้ค่ารีคอลและพรีชีชันเป็น 70.59% และ 69.84% ตามลำดับ ในขณะที่การรู้จำคำศัพท์ โดยที่ไม่ใช้ความรู้ของแนวคิดเลย จะให้ค่ารีคอลและ พรีชีชัน เป็น 54.93% และ 69.72%

เมื่อพิจารณาข้อผิดพลาดที่เกิดขึ้นในการรู้จำด้วยวิธีการทั้ง 2 วิธีนั้น จะพบว่า ข้อผิดพลาดที่เกิดขึ้นมากที่สุดนั้น เกิดจากการรู้จำคำศัพท์ผิด ซึ่งมีเปอร์เซ็นต์ของข้อผิดพลาดประเภทนี้ถึง 16% และ 21% สำหรับการรู้จำโดยใช้ความรู้แนวคิดเข้ามาร่วมพิจารณา และการรู้จำที่ไม่ใช่ความรู้แนวคิดมาช่วยตามลำดับ ความผิดพลาดอื่นๆ นั้น ได้แก่ ความผิดพลาดในการแบ่งคำผิด และความผิดพลาดในการรู้จำคำหายไป ซึ่งเมื่อไม่ใช้ความรู้แนวคิดมาช่วย โดยค้นหาแนวคิดจากลำดับของคำที่ได้จากการรู้จำนั้น จะให้ความผิดพลาดในการแบ่งคำผิดเป็น 9% และความผิดพลาดในการรู้จำคำหายไป 16% เมื่อใช้ความรู้ของแนวคิดมาช่วยในการรู้จำ ความผิดพลาดในการแบ่งคำผิด และรู้จำคำหายไปจะเกิดขึ้น 6% และ 5% ตามลำดับ

จากผลการทดลองจะเห็นว่าการรู้จำคำศัพท์โดยที่มีความรู้ของแนวคิดมาช่วยนั้น จะช่วยลดข้อผิดพลาดที่เกิดขึ้นลง โดยที่เมื่อพิจาณาดูสาเหตุที่ช่วยลดการเกิดข้อผิดพลาดดังกล่าว จะสามารถสรุปได้ดังนี้

1. การใช้ความรู้ของแนวคิดมาช่วยในการรู้จำคำนั้น จะเป็นการเพิ่มน้ำหนักในการค้นหาลำดับของคำให้เป็นไปในทางที่มีความถี่ที่จะเกิดขึ้นสูง เมื่อพิจารณาดูจากแนวคิด จึงช่วยให้ลดอัตราข้อผิดพลาดในการรู้จำคำผิดสำหรับคำศัพท์บางคำที่มีลักษณะการออกเสียงที่คล้ายกัน ซึ่งในกรณีนี้ การใช้ความรู้ของแนวคิดมาช่วยนั้น จะช่วยในการเลือกคำได้ว่า คำใดที่ควรเกิดขึ้น ณ ตำแหน่งที่กำลังพิจารณามากกว่ากัน

2. หากพิจารณาว่าการรู้จำคำจากกราฟของเซกเมนต์นั้น จะใช้ความรู้ทางด้านของหน่วยเสียง และเซกเมนต์ (ได้แก่ แบบจำลองค่าเวกเตอร์ลักษณะสำคัญของหน่วยเสียง และแบบจำลองระยะเวลา) และความรู้ทางด้านภาษา (ได้แก่ แบบจำลองการออกเสียง แบบจำลองภาษา โครงข่ายที่จับคู่ระหว่างคำศัพท์และแนวคิด และแบบจำลองแนวคิด) การเพิ่มความรู้ของแนวคิดเข้าไปนั้น จะเปรียบเสมือนการเพิ่มค่าน้ำหนักในส่วนของความรู้ทางด้านภาษา ซึ่งเป็นความรู้ในระดับของคำ ดังนั้นในการพิจารณาเปลี่ยนคำเป็นคำถัดไป ก็จะนำเอาน้ำหนักส่วนนี้มาพิจารณาด้วย โดยหากยิ่งเปลี่ยนคำมาก ก็จะยิ่งเพิ่มผลคูณค่าความน่าจะเป็น ทำให้ค่าความน่าจะเป็นยิ่งลดต่ำลง ดังนั้นการเพิ่มความรู้ของแนวคิดเข้าไป จึงทำให้ลดข้อผิดพลาดการรู้จำคำศัพท์ยาวๆ เป็นคำศัพท์สั้นๆ หรือหน่วยเสียงขยะหลายคำต่อกัน

นอกจากนี้ ข้อดีอีกอย่างของการพิจารณาความรู้ของแนวคิดเข้ามาช่วยด้วยนั้น เนื่องจากแนวคิดนั้นจะพิจารณาคำเป็นกลุ่มๆ โดยคำศัพท์ที่มีหน้าที่และความหมายเดียวกัน จะจัดอยู่ในประเภทกลุ่มเดียวกัน ดังนั้นจึงเหมาะสำหรับในกรณีที่คำศัพท์บางคำที่มีจำนวนตัวอย่างข้อมูลน้อย

### 6.1.2 ผลการวิจัยในการเปรียบเทียบประสิทธิภาพในการค้นหาแนวคิด

ในการทดลองนี้ จะพิจารณาผลการพิจารณาลำดับของแนวคิด โดยเปรียบเทียบผลการค้นหาแนวคิดพร้อมกับการรู้จำเสียง และการค้นหาลำดับของแนวคิดจากผลการรู้จำคำที่ได้จากกราฟของเซกเมนต์ ซึ่งใช้แบบจำลองความน่าจะเป็น และกฎทางไวยากรณ์ในการค้นหา โดยใช้ค่าของรีคอลและ พรีซีชัน ในการเปรียบเทียบ ซึ่งผลการทดลองที่ได้นั้น พบว่าค่ารีคอลและ พรีซีชันของการค้นหาแนวคิดพร้อมกับการรู้จำนั้น ได้เป็น 83.60% และ 85.76% ตามลำดับ ในขณะที่ผลของการค้นหาแนวคิดจากผลลัพธ์ของการรู้จำคำที่ดีที่สุดลำดับแรกจะให้ค่าของรีคอลและ พรีซีชันเป็น 67.94% และ 83.39% ตามลำดับ

ข้อผิดพลาดส่วนใหญ่ของการให้ลำดับของแนวคิดผิดไปนั้น เนื่องจากผลการรู้จำคำที่ไม่ถูกต้อง ส่งผลให้เมื่อพิจารณาลำดับของแนวคิดจากคำที่รู้จำผิดนั้น จะทำให้ลำดับของแนวคิดที่ผิดไปด้วย แต่หากใช้ความรู้ของแนวคิดเข้าไปในขณะของการรู้จำคำด้วย ซึ่งจะช่วยในการพิจารณาลำดับของคำที่จะไปในทางที่จะเกิดลำดับแนวคิดที่มีความถี่ในการเกิดสูง ดังนั้นถึงแม้ว่าการรู้จำคำนั้นจะให้ผลที่ผิดพลาด แต่เมื่อนำเอาความรู้ของแนวคิดมาช่วยก็มีแนวโน้มที่จะให้คำที่รู้จำที่ผิดพลาดนั้น เป็นคำที่มีแนวคิดเดียวกันกับคำที่ถูกต้อง ดังนั้นจึงส่งผลให้ความแม่นยำในการค้นหาลำดับของแนวคิดนั้นมีค่ามาก แม้ว่าจะให้ผลการรู้จำคำผิดไปบ้างก็ตาม

เมื่อพิจารณาผลของการค้นหาแนวคิดจากผลการรู้จำลำดับของคำที่ได้ ระหว่างผลการรู้จำที่ดีที่สุด 10 ลำดับแรก และผลการรู้จำที่ดีที่สุดลำดับเดียว จากผลการทดลองแสดงให้เห็นว่าเมื่อพิจารณาค้นหาแนวคิดจากผลการรู้จำที่ดีที่สุด 10 ลำดับแรกนั้น การค้นหาแนวคิดจะพยายามเลือกเอาลำดับของคำที่ส่งผลให้ได้ลำดับของแนวคิดที่ถูกต้องตามไวยากรณ์ และมีความถี่ในการเกิดสูง แต่อย่างไรก็ตาม ด้วยสาเหตุดังกล่าวก็จะส่งผลกระทบให้ในบางกรณีนั้น ระบบจะเลือกเอาลำดับของคำที่มีคำที่เกิดเกินขึ้นมาด้วย เพื่อให้ได้คะแนนของแนวคิดที่มากที่สุด โดยจะเห็นว่าเมื่อพิจารณาค้นหาแนวคิดจากผลการรู้จำ 10 ลำดับแรก จะเพิ่มค่ารีคอลจาก 67.94% มาเป็น 72.72% ในขณะที่ค่าพรีซีชันจะลดจาก 83.39% เหลือเพียง 75.77%

### 6.1.3 ผลการวิจัยในการเปรียบเทียบประสิทธิภาพในการระบุจุดมุ่งหมายของประโยค

ในการทดลองนี้ ได้เปรียบเทียบความถูกต้องของการระบุจุดมุ่งหมายของประโยค โดยใช้โครงข่ายประสาทเทียมในการแบ่งประเภทของจุดมุ่งหมาย ซึ่งมีอินพุตเป็นไบนารีเวกเตอร์ที่ระบุว่ามีแนวคิดใดปรากฏขึ้นในประโยคบ้าง โดยได้เปรียบเทียบความถูกต้องระหว่างการนำเอาแนวคิดที่ได้จากการค้นหาแนวคิดพร้อมกับการรู้จำคำซึ่งได้ความถูกต้องเป็น 74.07% และการค้นหาแนวคิดจากผลลัพธ์ของการรู้จำที่ได้จากเครื่องรู้จำโดยวิธีที่เสนอโดย Wuttiwiwatchaiและ Furui ซึ่งผลการรู้จำคำที่นำมาพิจารณาค้นหาแนวคิดนั้น จะมาจากผลการรู้จำที่ดีที่สุด 10 ลำดับแรก และผลการรู้จำที่ดีที่สุดเพียงลำดับเดียว ผลการทดลองพบว่าค่าความถูกต้องนั้น จะได้เป็น 61.11% และ 66.67% ตามลำดับ ในกรณีที่กำหนดให้ผลลัพธ์สุดท้ายของระบบนั้น จะเป็นค่าของจุดมุ่งหมายและชื่อประเทศ ผลความถูกต้องของผลลัพธ์สุดท้ายนี้ ระบบที่ค้นหาแนวคิดจากกราฟของเซกเมนต์นั้น จะให้ความถูกต้องเป็น 40.74% และระบบที่ค้นหาแนวคิดจากผลการรู้จำนั้นจะให้ความถูกต้องของผลลัพธ์สุดท้ายเป็น 22.22% เมื่อพิจารณาค้นหาแนวคิดจากผลการรู้จำคำ 10 ลำดับแรก และความถูกต้องเป็น 24.07% เมื่อพิจารณาจากผลการรู้จำคำที่ดีที่สุดลำดับเดียว

เนื่องจากในการระบุจุดมุ่งหมายของประโยคนั้น จะพิจาณาที่ลำดับของแนวคิดที่ได้เป็นหลัก ดังนั้นหากได้ลำดับของแนวคิดที่ผิด เช่น ได้แนวคิดเกินมา หรือการระบุแนวคิดผิด ก็จะส่งผลให้การระบุจุดมุ่งหมายผิดไปด้วย ดังนั้นแนวคิดที่ได้จากการค้นหาแนวคิดไปพร้อมกับการรู้จำคำนั้น ให้ค่าความแม่นยำที่มากกว่าการค้นหาแนวคิดจากผลของการรู้จำคำจากเครื่องรู้จำ ส่งผลให้ค่าความถูกต้องในการระบุจุดมุ่งหมายจึงมากกว่าด้วย โดยเฉพาะในกรณีที่พิจารณาผลลัพธ์สุดท้าย ได้แก่ จุดมุ่งหมาย และชื่อประเทศนั้น ความถูกต้องในการรู้จำชื่อประเทศจะมีผลต่อความถูกต้องของผลลัพธ์สุดท้ายด้วยเช่นกัน

นอกจากนี้ในการทดลองยังได้เสนอแนวทางในการแบ่งประเภทของจุดมุ่งหมาย โดยใช้คะแนนของแนวคิดที่ได้จากกระบวนการการรู้จำคำและแนวคิดจากกราฟของเซกเมนต์มาพิจารณาในการแบ่งแยกประเภทของจุดมุ่งหมาย โดยได้เปรียบเทียบวิธีการแบ่งประเภทจุดมุ่งหมายระหว่างการใช้กฎ และการใช้โครงข่ายประสาทเทียมที่รับอินพุตเป็นเวกเตอร์คะแนนรวมของแนวคิดต่างๆ จากผลการทดลองแสดงให้เห็นว่า การใช้กฎในการแบ่งประเภทจุดมุ่งหมายนั้น จะให้ความถูกต้องในการแบ่งประเภทมากกว่าการใช้โครงข่ายประสาทเทียม โดยได้ความถูกต้องเป็น 79.63% เมื่อเทียบการใช้โครงข่ายประสาทเทียม ซึ่งให้ความถูกต้อง 77.78%

# รายการอ้างอิง


บุญเสริม กิจศิริกุล. ปัญญาประดิษฐ์ เอกสารคำสอนวิชา 211064. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2003.

R. O. Duda, P. E. Hart and D. G. Stork. Pattern Classification. Canada : John Wiley & Sons, 2001.

G. Di Fabbrizio and C. Lewis. Florence: A dialogue manager framework for spoken dialogue systems. Proceedings of ICSLP, pp. 3065-3068, 2004.

J. R. Glass. A Probabilistic framework for Segment-based Speech Recognition. Computer Speech and Language 17 (2003) : 137-152.

N. Gupta, G. Tur, D. Hakkani-Tur, S. Bangalore, G. Riccardi and Mazin Gilbert. The AT&T Spoken Language Understanding System. IEEE Transactions on Audio speech and language process, pp. 213-222. 2006.

HTK[Online]. Available from: http://htk.eng.cam.ac.uk/.

iec[Online] Available from: http://www.iec.org/online/tutorials/speech_enabled

C. D. Manning and H. Schutze. Foundation of Statistical Natural Language Processing. MIT Press, 1999.

H. Bonneau-Maynard and F. Lefevre. Investigating Stochastic Speech Understanding. IEEE workshop on Automatic Speech Recognition and Understanding, pp. 260-263, 2001.

S. Miller, R. Bobrow, R. Ingria and R. Schwartz. Hidden Understanding models of Natural Language. in Annaul Meeting on ACL, 1994.

J. Glass and V. Zue.(2003) MIT Open Course Ware [Online]. Available from: http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-345Automatic-Speech-RecognitionSpring2003/CourseHome/index.htm

S.Tangruamsub, P. Punyabukkana and A. Suchato. Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling. IEEE International Conference on Research Innovation and Vision for the Future, pp. 253-260, 2007.

Y.-y. Wang and A. Acero. Discriminative Models for Spoken Language Undestanding. Proceedings of ICSLP, 2006.

W.-L. Wu, R.-Z. Lu, H. Liu and F. Gao. A Spoken Language Understanding Approach Using Successive Learners. Proceedings of ICSLP, 2006.

C. Wutiwiwatchai and S. Furui. A multi-stage approach for Thai spoken language understanding. Speech Communication, 48 (2006): 305-320.

S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev and P. Woodland. The HTK Book (for HTK Version 3.3). Cambridge, 2005.

V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T. J. Hazen, and L. Hetherington. Jupiter: A telephone-Based Conversational Interface for Weather Information. IEEE transaction on speech and audio processing, pp. 85-96, 2000.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ภาคผนวก ก

## สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงพยัญชนะต้นในภาษาไทย

ตารางที่ ก.1 สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงพยัญชนะต้นในภาษาไทย

| พยัญชนะต้นในภาษาไทย | สัทอักษรสากล | ตัวอักษร ASCII |
|---|---|---|
| ป | p | p |
| ปร | pr | pr |
| ปล | pl | pl |
| ต ฏ | t | t |
| ตร | tr | tr |
| จ จร | c | c |
| ก | k | k |
| กร | kr | kr |
| กล | kl | kl |
| กว | kw | kw |
| อ | ʔ | z |
| พ ภ ผ | ph | ph |
| พร ภร ผร | phr | phr |
| พล ภล ผล | phl | phl |
| ท ธ ฒ ฑ ถ ฐ | th | th |
| ทร ธร ฑร ถร ฐร | thr | thr |
| ช ฌ ฉ | ch | ch |
| ข ค ฆ | kh | kh |
| ขร คร ฆร | khr | khr |
| ขล คล ฆล | khl | khl |
| ขว คว ฆว | khw | khw |
| บ | b | b |
| บร | br | br |
| บล | bl | bl |
| ด ฎ ฑ | d | d |

| | | |
|---|---|---|
| ดร | dr | dr |
| ม หม | m | m |
| น ณ หน | n | n |
| ง หง | ŋ | ng |
| ฝ ฟ | f | f |
| ฝร ฟร | fr | fr |
| ฟล | fl | fl |
| ส ศ ษ ซ ทร สร ศร | s | s |
| ห ฮ | h | h |
| ร หร ฤ | r | r |
| ล ฬ หล ฦ | l | l |
| ว หว | w | w |
| ย ญ หย หญ อย | j | j |

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ภาคผนวก ข

## สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงตัวสะกดในภาษาไทย

ตารางที่ ข.1 สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงตัวสะกดในภาษาไทย

| ตัวสะกด (พยัญชนะท้ายในภาษาไทย) | สัทอักษรสากล | ตัวอักษร ASCII |
|---|---|---|
| แม่กก (ก ข ค ฆ) | k˺ | k^ |
| แม่กด (จ ช ซ ฎ ฏ ฐ ฑ ฒ ด ต ถ ท ธ ศ ษ ส) | t˺ | t^ |
| แม่กบ (บ ป ภ พ) | p˺ | p^ |
| แม่กน (น ณ ญ ร ล ฬ) | n | n^ |
| แม่กง (ง) | ŋ | ng^ |
| แม่กม (ม) | m | m^ |
| แม่เกอย (ย) | j | j^ |
| แม่เกอว (ว) | w | w^ |
| ตัวสะกดลงท้ายด้วย “ช” | ch˺ | ch^ |
| ตัวสะกดลงท้ายด้วย “ฟ” | f˺ | f^ |
| ตัวสะกดลงท้ายด้วย “ล” | l˺ | l^ |
| ตัวสะกดลงท้ายด้วย “ส” | s˺ | s^ |

# ภาคผนวก ค

## สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงสระในภาษาไทย

ตารางที่ ค.1 สัทอักษรสากลและตัวอักษร ASCII ที่ใช้แทนเสียงสระในภาษาไทย

| สระในภาษาไทย | สัทอักษรสากล | ตัวอักษร ASCII |
|---|---|---|
| อิ | i | i |
| อี | iː | ii |
| เอะ | e | e |
| เอ | eː | ee |
| แอะ | ɛ หรือ æ | x |
| แอ | ɛː หรือ æː | xx |
| อึ | ɯ หรือ ɨ | v |
| อือ | ɯː หรือ ɨː | vv |
| เออะ | ɤ หรือ ə | q |
| เออ | ɤː หรือ əː | qq |
| อะ | a | a |
| อา | aː | aa |
| อุ | u | u |
| อู | uː | uu |
| โอะ | o | o |
| โอ | oː | oo |
| เอาะ | ɔ | @ |
| ออ | ɔː | @@ |
| เอียะ | ia | ia |
| เอีย | iːa | iia |
| เอือะ | ɯa หรือ ɨa | va |
| เอือ | ɯːa หรือ ɨːa | vva |
| อัวะ | ua | ua |
| อัว | uːa | uua |

## ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศิรินาถ ตั้งรวมทรัพย์ เกิดวันที่ 2 มิถุนายน พ.ศ. 2527 ที่จังหวัดนครราชสีมา สำเร็จการศึกษาในระดับมัธยมศึกษาตอนต้น และตอนปลายจากโรงเรียน บดินทรเดชา (สิงห์ สิงหเสนี) สำเร็จการศึกษาในระดับบัณฑิตศึกษาในสาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2548

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย