การจำแนกเสียงโกรธในบทสนทนาของศูนย์ให้บริการข้อมูล

น.ส.วิฎากร แซ่หว่อง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศทางธุรกิจ ไม่สังกัดภาควิชา/เทียบเท่า
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

CLASSIFICATION OF ANGER VOICE IN CALL CENTER DIALOG

Miss Widakorn Sae-wong

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Information Technology in Business
Common Course
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2018

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การจำแนกเสียงโกรธในบทสนทนาของศูนย์ให้บริการข้อมูล
โดย	น.ส.วิฎากร แซ่หว่อง
สาขาวิชา	เทคโนโลยีสารสนเทศทางธุรกิจ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.จันทร์เจ้า มงคลนาวิน

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.อัษฎาพร ทรัพย์สมบูรณ์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.จันทร์เจ้า มงคลนาวิน)

..... กรรมการ

(อาจารย์ ดร.ภูริพันธุ์ รุจิขจร)

..... กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.วิรัตน์ จารีวงศ์ไพบูลย์)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

วิฎากร แซ่หว่อง : การจำแนกเสียงโกรธในบทสนทนาของศูนย์ให้บริการข้อมูล . (CLASSIFICATION OF ANGER VOICE IN CALL CENTER DIALOG) อ.ที่ปรึกษาหลัก : ผศ. ดร.จันทร์เจ้า มงคลนาวิน

ศูนย์ให้บริการข้อมูลเป็นส่วนงานขององค์กรที่เกี่ยวข้องกับข้อมูลเสียงโดยตรง อารมณ์โกรธของลูกค้าที่เกิดขึ้นเป็นสิ่งที่ส่งผลในด้านลบต่อองค์กร งานในการตรวจสอบอารมณ์โกรธที่เกิดขึ้นให้ถูกต้องและรวดเร็วจึงเป็นสิ่งที่องค์กรให้ความสำคัญเนื่องจากมีจำนวนผู้ติดต่อเข้ามาใช้บริการจำนวนมากในแต่ละวัน จึงเป็นที่มาของการศึกษาการพัฒนาตัวแบบรู้จำอารมณ์โกรธของลูกค้าในบริบทของข้อมูลเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูล การพัฒนาตัวแบบรู้จำอารมณ์มีขั้นตอนสำคัญ 4 ขั้นตอน ได้ แก่ (1) การประมวลผลข้อมูลเสียงเบื้องต้น เป็นการเตรียมข้อมูลเสียงให้พร้อมสำหรับการวิเคราะห์โดยแยกเสียงของลูกค้าออกจากสายสนทนา และตัดให้เป็นรอบการโต้ตอบ (2) การระบุอารมณ์ให้กับรอบการโต้ตอบ (3) การสกัดคุณลักษณะทางคลื่นเสียงจำนวน 1,582 คุณลักษณะที่สามารถใช้จำแนกอารมณ์ที่เกิดขึ้นภายในรอบการโต้ตอบ และ (4) การพัฒนาตัวแบบ โดยผู้ศึกษาได้เปรียบเทียบประสิทธิภาพของเทคนิค 4 เทคนิค ได้แก่ แรนดอมฟอร์เรส ดิซิสชั่นทรี ลิเนียร์เอสวีซีและโลจิสติกส์รีเกรสชั่น

จากข้อมูลรอบการโต้ตอบทั้งหมดที่ใช้พัฒนาตัวแบบคือ 1,345 รอบการโต้ตอบ ที่แบ่งเป็นรอบการโต้ตอบที่เป็นอารมณ์โกรธจำนวน 517 รอบ และรอบการโต้ตอบที่ไม่พบอารมณ์โกรธหรือเป็นอารมณ์ปกติจำนวน 828 รอบ ผลการวิจัยพบว่า เทคนิคแรนดอมฟอร์เรสเป็นเทคนิคที่ให้ประสิทธิภาพที่ดีที่สุด โดยมีค่าระลึกของอารมณ์โกรธเท่ากับ 83.49 เปอร์เซ็นต์ และค่าความถูกต้องสมดุล 85.53 เปอร์เซ็นต์ นอกจากนั้นยังพบว่ารอบการโต้ตอบที่มีความยาวที่ 4 วินาทีขึ้นไปจะได้ผลลัพธ์ในการทำนายที่แม่นยำมากกว่าทั้งในรอบการโต้ตอบที่เป็นอารมณ์โกรธและอารมณ์ปกติ

สาขาวิชา	เทคโนโลยีสารสนเทศทางธุรกิจ	ลายมือชื่อนิสิต .....
ปีการศึกษา	2561	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5981539526 : MAJOR INFORMATION TECHNOLOGY IN BUSINESS
KEYWORD: Emotion Recognition, Voice, Machine Learning, Call Center
Widakorn Sae-wong : CLASSIFICATION OF ANGER VOICE IN CALL CENTER DIALOG. Advisor: Asst. Prof. Janjao Mongkolnavin, Ph.D.

Call center is a department in an organization which directly involves with audio data. Customer's angry negatively affects the organization. Thus, to monitor customers' angry precisely and timely is a task that the organization gives priority. This is because there are voluminous phone calls come in the call center every day. This research is aimed at developing an approach to detect customers' angry from voice recorded in call center dialogs. The study was divided into 4 processes : (1) Audio preprocessing, separating customer's voice from each call center conversation and splitting it into turns. (2) Emotion Annotation, labelling each turn into "ANGRY" or "NORMAL" (3) Turn feature extraction: extracting 1,582 features which are used in the modelling process from each turn (4) Model development, 4 modelling techniques: Random Forest Classifier, Decision Tree, Linear SVC, and Logistic Regression are used in the study.

From 1,345 turns, 517 angry and 828 normal turns, it is found that Random forest is the technique that has the best performance in aspects of recall and balanced accuracy which are 83.49 and 85.33 percent, respectively. We find that the model gives more accurate predictions on turns which are longer than 4 seconds for both angry and normal turns.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study:	Information Technology in Business	Student's Signature .....
Academic Year:	2018	Advisor's Signature .....

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.จันทร์เจ้า มงคลนาวิน อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำปรึกษา ชี้แนะ แนะนำ และตรวจทานแก้ไขข้อบกพร่องต่างๆในวิทยานิพนธ์ตั้งแต่เริ่มต้นจนเสร็จสมบูรณ์อย่างเต็มกำลังเสมอมาและขอขอบพระคุณ

รองศาสตราจารย์ ดร. อัษฎาพร ทรัพย์สมบูรณ์ ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร. ภูริพันธุ์ รุจิขจร กรรมการสอบวิทยานิพนธ์ และ ผู้ช่วยศาสตราจารย์ ดร. วิรัตน์ จารีวงศ์ไพบูลย์ กรรมการภายนอกมหาวิทยาลัย ที่สละเวลาชี้แนะและให้ข้อเสนอแนะเพื่อให้ผู้วิจัยสามารถนำไปปรับปรุงแก้ไขวิทยานิพนธ์ฉบับนี้จนมีความสมบูรณ์มากยิ่งขึ้น และขอขอบพระคุณผู้บริหารและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลที่อนุญาตให้ผู้วิจัยสามารถนำข้อมูลบันทึกเสียงสนทนามาใช้ในการศึกษาครั้งนี้ รวมถึงการให้การรับรองและดูแลอย่างอบอุ่นในขณะที่เข้าไปดำเนินการที่สำนักงาน

สุดท้ายนี้ ขอขอบคุณทุกท่านที่ได้กล่าวข้างต้น รวมถึง บิดา มารดา และ ญาติพี่น้อง รวมไปถึงท่านอื่นๆที่ไม่ได้กล่าวถึง ณ ที่นี้ด้วย ที่ให้คำแนะนำและกำลังใจแก่ผู้วิจัยมาโดยตลอดจนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยดี

วิฎากร แซ่หว่อง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

..... ค

บทคัดย่อภาษาไทย ..... ค

..... ง

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ซ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 3

1.3 ขอบเขตของการวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 3

1.5 ขั้นตอนในการทำวิจัยเบื้องต้น 3

บทที่ 2 การทบทวนวรรณกรรมที่เกี่ยวข้อง 5

2.1 ลูกค้าและอารมณ์โกรธ 5

2.2 การรู้จำอารมณ์ (Emotion Recognition) 5

2.3 การศึกษาและพัฒนาตัวแบบการรู้จำอารมณ์จากลักษณะทางคลื่นเสียง 8

2.4 เทคนิคในการพัฒนาและทดสอบประสิทธิภาพตัวแบบ 17

2.5 เครื่องมือที่ใช้ในการทำวิจัย 23

บทที่ 3 ระเบียบวิธีวิจัย 28

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.1 ข้อมูลที่ใช้ในการศึกษา 28
3.2 ขั้นตอนการศึกษา 29
3.3 การประมวลข้อมูลเสียงเบื้องต้น 30
3.4 การระบุอารมณ์ 34
3.5 การสกัดคุณลักษณะ 35
3.6 การพัฒนาตัวแบบ 36
3.7 การวัดประสิทธิภาพตัวแบบ (Model Measurements) 43
บทที่ 4 ผลการทดลอง 45
4.1 ผลการตรวจสอบไขว้แบบแบ่งชั้นภูมิ 45
4.2 ผลการพัฒนาตัวแบบรู้จำอารมณ์ด้วยวิธีที่นำเสนอ 55
4.3 การศึกษาเพิ่มเติม 57
บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ 59
5.1 สรุปการศึกษาการประมวลผลเบื้องต้นของข้อมูลเสียงประเภทการสนทนา 59
5.2 การศึกษาตัวแบบรู้จำทางอารมณ์จากข้อมูลเสียงที่เกิดขึ้นจริง 59
5.3 ปัจจัยที่ส่งผลต่อประสิทธิภาพตัวแบบ 61
5.4 ข้อจำกัดของงานวิจัย 61
5.5 ข้อเสนอแนะสำหรับการนำงานวิจัยไปศึกษาต่อ 62
5.6 ข้อจำกัดและข้อเสนอแนะสำหรับการนำตัวแบบประยุกต์ใช้ 63
บรรณานุกรม 65
ประวัติผู้เขียน 69

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 แสดงรายละเอียดของรอบการโต้ตอบของอารมณ์ที่เกิดขึ้น 11
ตารางที่ 2 แสดงคุณลักษณะทางคลื่นเสียงที่นำเสนอ 15
ตารางที่ 3 ตัวอย่างตารางคอนฟิวชั่นเมทริกซ์ 20
ตารางที่ 4 แสดงรายละเอียดของพารามิเตอร์ของฟังก์ชันสปริทออนไซเลนส์ 25
ตารางที่ 5 แสดงจำนวนเพศของลูกค้าและระยะเวลารวมจากข้อมูลเสียงสนทนาที่ได้รับ 28
ตารางที่ 6 แสดงฟังก์ชันและค่าพารามิเตอร์ที่ใช้ในการตัดรอบการโต้ตอบ 32
ตารางที่ 7 แสดงจำนวนข้อมูลเสียงที่ถูกทำให้อยู่ในรูปของรอบการโต้ตอบ 33
ตารางที่ 8 แสดงจำนวนรอบการโต้ตอบสุทธิ 34
ตารางที่ 9 แสดงจำนวนรอบการโต้ตอบตามอารมณ์ที่ระบุได้ 35
ตารางที่ 10 แสดงสัดส่วนระหว่างข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ 37
ตารางที่ 11 แสดงจำนวนอารมณ์โกรธและอารมณ์ปกติในแต่ละข้อมูลชุดฝึกสอน 38
ตารางที่ 12 แสดงจำนวนอารมณ์โกรธและอารมณ์ปกติในแต่ละข้อมูลชุดฝึกทดสอบ 38
ตารางที่ 13 แสดงคุณลักษณะของเทคนิคดิซิสชั่นทรี 40
ตารางที่ 14 แสดงคุณลักษณะของเทคนิคแรนดอมฟอร์เรส 41
ตารางที่ 15 แสดงค่าวัดประสิทธิภาพตัวแบบที่ใช้ 44
ตารางที่ 16 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคดีซิสชั่นทรี 46
ตารางที่ 17 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคแรนดอมฟอร์เรส 47
ตารางที่ 18 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคลิเนียร์เอสวีซี 48
ตารางที่ 19 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคโลจิสติกส์ 49
ตารางที่ 20 ตารางแสดงค่าเฉลี่ยของค่าวัดประสิทธิภาพของตัวแบบ 50
ตารางที่ 21 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคดิซิสชั่นทรี 51

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 22 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคแรนดอมฟอร์เรส 52
ตารางที่ 23 ตารางแสดงค่าเฉลี่ยของค่าวัดประสิทธิภาพตัวแปรของ 53
ตารางที่ 24 แสดงค่าเฉลี่ยของค่าวัดประสิทธิภาพของตัวแปรที่ได้จากเทคนิคทั้งหมด 54
ตารางที่ 25 แสดงจำนวนอารมณ์โกรธและอารมณ์ปกติของชุดข้อมูลฝึกและข้อมูลทดสอบ 55
ตารางที่ 26 ตารางคอนฟิวชันเมทริกซ์ที่ได้จากการพัฒนาตัวแบบ 56

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

หน้า

ภาพที่ 1 แสดงการแบ่งข้อมูลชนิดการตรวจสอบไขว้เมื่อกำหนด K=10 18
ภาพที่ 2 แสดงตัวอย่างการแบ่งข้อมูลแบบการตรวจสอบไขว้แบบแบ่งชั้นภูมิ 5 ชุด 19
ภาพที่ 3 ตัวอย่างกราฟโค้งอาร์โอซี 22
ภาพที่ 4 ตัวอย่างคลื่นเสียงของแฟ้มข้อมูลเสียงที่จัดเก็บในรูปแบบสเตอริโอ 24
ภาพที่ 5 แสดงคลื่นเสียงของข้อมูลเสียงในส่วนที่มีเสียงพูดและเสียงเงียบเป็นส่วนประกอบ 25
ภาพที่ 6 แสดงแผนภาพขั้นตอนการพัฒนาตัวแบบรู้จำอารมณ์ 29
ภาพที่ 7 แสดงหน้าต่างการทำงานของซอฟต์แวร์ออดาซิตี้ 30
ภาพที่ 8 แสดงหน้าต่างคำสั่งที่ใช้ในการแยกช่องเสียง 31
ภาพที่ 9 แสดงส่วนที่เกิดเสียง และช่องว่างเสียงเงียบในขณะที่ผู้พูดอื่นกำลังพูด 31
ภาพที่ 10 รายละเอียดของแฟ้มข้อมูลข้อกำหนดที่ใช้งานเบื้องต้น 36
ภาพที่ 11 กราฟแสดงค่า min_samples_splits ของเทคนิคดิซิสชั่นทรี 40
ภาพที่ 12 กราฟแสดงค่า min_samples_leaf ของเทคนิคดิซิสชั่นทรี 40
ภาพที่ 13 กราฟแสดงค่า min_impurity_decrease ของเทคนิคดิซิสชั่นทรี 41
ภาพที่ 14 กราฟแสดงค่า n_estimators ของเทคนิคแรนดอมฟอร์เรส 42
ภาพที่ 15 กราฟแสดงค่า min_samples_split ของเทคนิคแรนดอมฟอร์เรส 42
ภาพที่ 16 กราฟแสดงค่า min_samples_leaf ของเทคนิคแรนดอมฟอร์เรส 42
ภาพที่ 17 แสดงคุณลักษณะของเทคนิคที่ปรับแต่ง 43
ภาพที่ 18 กราฟแท่งแสดงจำนวนรอบการโต้ตอบของ 57
ภาพที่ 20 กราฟแท่งแสดงจำนวนรอบการโต้ตอบของ 58

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ปัจจุบันมีนักวิจัยจำนวนมากให้ความสนใจศึกษาเกี่ยวกับเสียงที่เกิดขึ้นจากมนุษย์ เช่น เสียงพูด และเสียงร้องเพลง รวมไปถึงอารมณ์ที่แสดงออกมาพร้อมกับการเปล่งเสียง และนำมาใช้ในการทำวิจัยเชิงลึกเกี่ยวกับคุณสมบัติของเสียง เกิดเป็นทฤษฎีที่สามารถนำไปใช้ประโยชน์ได้อย่างหลากหลาย เช่น การรู้จำเสียงพูด (Speech Recognition) การรู้จำเสียงผู้พูด (Speaker Recognition) การสังเคราะห์เสียงพูด (Speech Synthesis) การสังเคราะห์เสียงร้อง (Singing Voice Synthesis) การรู้จำรูปแบบของการพูด (Speech Pattern Recognition) และการรู้จำอารมณ์จากเสียงพูด (Speech Emotion Recognition) เป็นต้น (Chakraborty, Pandharipande, & Kopparapu, 2016a) ทฤษฎีเหล่านี้ถูกนำมาประยุกต์ใช้ร่วมกับเทคโนโลยีในปัจจุบันเกิดเป็นนวัตกรรมใหม่ที่น่าสนใจและสามารถอำนวยความสะดวกให้แก่ผู้ใช้งานได้มากขึ้น เช่น ระบบที่สามารถโต้ตอบกับผู้ใช้ผ่านทางเสียงพูดได้ ระบบจำแนกประเภทของเพลงตามอารมณ์ที่สื่อในเพลง และระบบยืนยันตัวตนโดยใช้เสียง เป็นต้น

การรู้จำอารมณ์จากเสียงพูด (Speech Emotion Recognition) ถูกนำมาประยุกต์ใช้ในงานด้านต่าง ๆ เช่น การพัฒนาระบบการสื่อสารระหว่างผู้ใช้งานกับคอมพิวเตอร์ให้มีเป็นความธรรมชาติมากขึ้น นอกจากนั้นยังมีความจำเป็นอย่างมากต่อระบบที่การตอบสนองต่อผู้ใช้งานขึ้นอยู่กับอารมณ์ของผู้ใช้งานโดยตรง (Ayadi, Kamel, & Karray, 2011) หากกล่าวถึงการนำทฤษฎีนี้ไปประยุกต์ใช้ในภาคธุรกิจ ศูนย์ให้บริการข้อมูลลูกค้า (Call center) เป็นส่วนงานขององค์กรที่เกี่ยวข้องกับข้อมูลเสียงและอารมณ์ที่เกิดขึ้นจากเสียงสนทนาโดยตรง วัตถุประสงค์หลักของศูนย์ให้บริการข้อมูล คือ การให้บริการข้อมูลเพิ่มเติมแก่ผู้บริโภค รวมไปถึงการแก้ไขปัญหาที่เกิดขึ้นหลังจากการบริโภคสินค้าหรือบริการจากองค์กร บ่อยครั้งที่พบว่าผู้บริโภคเกิดความไม่พึงพอใจ ซึ่งสาเหตุมีทั้งที่มาจากการไม่พึงพอใจในสินค้าหรือบริการที่ได้รับ และมาจากการที่เจ้าหน้าที่ของศูนย์ให้บริการข้อมูลไม่สามารถจัดการกับสถานการณ์ และแก้ไขปัญหาให้กับผู้บริโภคได้ ดังนั้น ความไม่พึงพอใจของผู้บริโภคที่เกิดขึ้นกับการใช้บริการศูนย์ให้บริการข้อมูลจึงเป็นสิ่งสำคัญที่องค์กรต้องการควบคุมและแก้ไข โดยในปัจจุบันเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลบางแห่งใช้การสุ่มตรวจสอบสายสนทนาเพื่อตรวจสอบการทำงานของเจ้าหน้าที่และระดับความพึงพอใจของผู้รับบริการ ซึ่งในแต่ละวันมีผู้รับบริการโทรศัพท์เข้ามาใช้บริการเป็นจำนวนมาก การใช้คนในการตรวจสอบสายสนทนาทุกสายจะเป็นการสิ้นเปลืองทรัพยากรเป็นเท่าตัว (Chakraborty et al., 2016a) ซึ่งหากสามารถนำความรู้ด้านการรู้จำอารมณ์มาพัฒนาเป็นระบบการรู้จำอารมณ์โดยอัตโนมัติ จะช่วยให้การตรวจสอบดังกล่าวมีประสิทธิภาพทั้งในแง่มุมของปริมาณของสายสนทนาที่สามารถตรวจสอบได้ ความถูกต้องของการตรวจสอบและ

ประสิทธิภาพการใช้ทรัพยากรมนุษย์ นอกจากนั้นยังพบว่ามีงานวิจัยเพียงส่วนน้อยเท่านั้นที่ทำการศึกษาการรู้จำอารมณ์จากเสียงพูดในบริบทของการรู้จำอารมณ์ของลูกค้าต่อการใช้บริการศูนย์ให้บริการข้อมูล

ในการศึกษาการรู้จำอารมณ์จากเสียงพูด องค์ประกอบหลักที่ส่งผลกระทบต่อระดับความถูกต้องของการรู้จำอารมณ์ที่นักวิจัยในด้านนี้ส่วนใหญ่ให้ความสำคัญ ได้แก่ ฐานข้อมูลเสียงที่นำมาใช้ (Audio Corpus), หลักการแยกเสียงผู้สนทนา (Speaker Diarization) ขนาดของเสียงที่เหมาะสมต่อการพัฒนาตัวแบบ (Window size, Audio Chunk) กลุ่มของคุณลักษณะ (Features Set) และเทคนิคที่ใช้ในการพัฒนาตัวแบบ (Classification Techniques) (Ayadi et al., 2011) ฐานข้อมูลเสียงที่นำมาใช้ในการศึกษามีความสำคัญอย่างยิ่งต่อการนำตัวแบบไปประยุกต์ใช้ ผลของงานวิจัยที่ได้จากการนำข้อมูลเสียงที่เกิดจากการตั้งใจแสดง (Acted Data) หรือการตั้งใจอ่านเพื่อแสดงอารมณ์มาใช้ในการศึกษา ย่อมไม่สามารถในการนำไปประยุกต์ใช้กับสถานการณ์จริงโดยตรง เนื่องด้วยปัจจัยต่าง ๆ เช่น การแสดงออกทางอารมณ์ของข้อมูลที่เกิดขึ้นจริงโดยธรรมชาติ (Real life Data) มักจะไม่ชัดเจนเหมือนข้อมูลเสียงที่เกิดจากการตั้งใจแสดง (Chakraborty et al., 2016a) หรือขนาดความยาวของประโยคที่สร้างขึ้นมักจะสั้นกว่าและไม่ต่อเนื่องดังเช่นข้อมูลที่เกิดขึ้นจริง เป็นต้น และยังมีประเด็นอื่นที่น่าสนใจ เช่น ขนาดของเสียงที่เหมาะสมต่อการพัฒนาตัวแบบต้องมีความยาวเท่าไร ควรมีการแยกเสียงผู้พูดออกจากกันหรือไม่ กลุ่มคุณลักษณะทางคลื่นเสียงที่เหมาะสมมากที่สุดคือกลุ่มใด รวมไปถึงเทคนิคในการสร้างตัวแบบที่เหมาะสม ประเด็นต่าง ๆ เหล่านี้ จำเป็นจะต้องทำการทบทวนและศึกษาอย่างละเอียด เพื่อให้การพัฒนาตัวแบบการรู้จำอารมณ์มีระดับความถูกต้องสูงสุด และมีประสิทธิภาพในการนำไปใช้งาน

สืบเนื่องจากผู้วิจัยได้รับความอนุเคราะห์จากผู้ให้บริการเครือข่ายโทรศัพท์เคลื่อนที่แห่งหนึ่งในประเทศไทย ในการให้ข้อมูลเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลมาใช้ในการศึกษาครั้งนี้ ร่วมกับการทบทวนวรรณกรรมในอดีตของงานวิจัยในกลุ่มของการรู้จำอารมณ์จากเสียงพูด ข้อจำกัดที่พบ คือ มีงานวิจัยส่วนน้อยเท่านั้นที่นำข้อมูลเสียงที่เกิดขึ้นจริงมาใช้ในการศึกษา (Pappas, Androutsopoulos, & Papageorgiou, 2015) จุดมุ่งหมายของงานวิจัยชิ้นนี้ คือ ต้องการศึกษาแนวทางในพัฒนาตัวแบบการรู้จำอารมณ์โกรธจากเสียงพูดโดยใช้ข้อมูลเสียงที่เกิดขึ้นจริงโดยธรรมชาติ ในบริบทของเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าศูนย์ให้บริการข้อมูลในภาษาไทย เพื่อให้สามารถนำไปต่อยอดเป็นระบบตรวจสอบสายสนทนาที่ผู้รับบริการเกิดความไม่พึงพอใจ เพื่อรวบรวมข้อมูลสำหรับผู้รับผิดชอบศูนย์ให้บริการข้อมูลในองค์กรได้อย่างมีประสิทธิภาพ

1.2 วัตถุประสงค์ของการวิจัย

1. ศึกษาเกี่ยวกับวิธีการประมวลผลเบื้องต้นของข้อมูลเสียง

2. เพื่อศึกษาและพัฒนาตัวแบบสำหรับการรู้จำอารมณ์โกรธจากบันทึกข้อมูลเสียงสนทนาที่เกิดขึ้นจริงในศูนย์ให้บริการข้อมูล

3. ศึกษาเกี่ยวกับปัจจัยที่ส่งผลต่อประสิทธิภาพของตัวแบบที่เป็นผลลัพธ์

1.3 ขอบเขตของการวิจัย

1. ใช้คลังข้อมูลเสียงจากเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลของผู้ให้บริการเครือข่ายโทรศัพท์เคลื่อนที่แห่งหนึ่ง

2. คลังข้อมูลเสียงที่นำมาศึกษาเป็นการสนทนาในภาษาไทย

3. คุณลักษณะของเสียงที่เลือกใช้ คือคุณลักษณะทางคลื่นเสียง (Acoustic Feature)

4. ตัวแบบรู้จำอารมณ์ในงานวิจัยนี้ ทำการศึกษาเพื่อจำแนกอารมณ์โกรธเพียงอารมณ์เดียว

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. ได้ศึกษาทฤษฎีและแนวคิดเกี่ยวกับข้อมูลเสียง และการประมวลผลเบื้องต้นกับข้อมูลเสียงที่เป็นเสียงสนทนาที่มีความต่อเนื่องและมีผู้พูดมากกว่าหนึ่งคน

2. ได้ศึกษาคุณลักษณะของเสียงที่เหมาะสมสำหรับการนำไปใช้กับเทคนิคการรู้จำอารมณ์จากเสียง รวมถึงวิธีการสกัดคุณลักษณะที่ต้องการ

3. ได้ศึกษาขั้นตอนและเทคนิคที่ใช้ในการพัฒนาตัวแบบรู้จำอารมณ์โกรธจากคุณลักษณะของเสียง

4. ได้ศึกษาปัจจัยเพิ่มเติมที่ส่งผลกระทบต่อประสิทธิภาพของตัวแบบการรู้จำอารมณ์

5. เพื่อเป็นแนวทางสำหรับการศึกษาการรู้จำอารมณ์จากคุณลักษณะของเสียงพูดในอนาคต

1.5 ขั้นตอนในการทำวิจัยเบื้องต้น

เนื่องจากการรู้จำอารมณ์จากเสียงสนทนาเป็นเทคนิคที่มีความซับซ้อนสูง ทำให้ผู้วิจัยจำเป็นต้องศึกษาและรวบรวมแนวคิดเกี่ยวกับงานวิจัยที่เกี่ยวข้องอย่างละเอียด โดยสามารถแสดงเป็นขั้นตอนได้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. ศึกษางานวิจัยด้านการรู้อารมณ์จากเสียง (Emotion Recognition) เพื่อให้ได้แนวคิดและสิ่งที่งานวิจัยนี้ต้องการศึกษาเพิ่มเติมต่อจากงานวิจัยในอดีต

2. ดำเนินการติดต่อกับหน่วยงานหรือองค์กรที่มีคลังข้อมูลเสียงสนทนา เช่น บันทึกเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ศูนย์ให้บริการข้อมูล เพื่อทำเรื่องขอความอนุเคราะห์ข้อมูลมาใช้ในการทำงานวิจัยชิ้นนี้

3. ศึกษางานวิจัยที่เกี่ยวข้องกับการรู้จำอารมณ์จากเสียง (Emotion Recognition) ตั้งแต่อดีตจนถึงในปัจจุบัน และสรุป ทฤษฎี แนวคิดและขั้นตอนที่พบจากงานวิจัยในอดีต รวมถึงเปรียบเทียบประสิทธิภาพของเทคนิคที่ใช้ในการรู้จำอารมณ์ของแต่ละงานวิจัย

4. ศึกษาการประมวลเสียงเบื้องต้น (Pre-processing) ได้แก่ การจำแนกเสียงสนทนาว่าเกิดอารมณ์โกรธหรือไม่ (Anger Annotation) โดยการใช้เครื่องมือในการระบุ การแบ่งส่วนของผู้สนทนา (Speaker Diarization) และการแบ่งส่วนของเสียง (Voice Segmentation)

5. ศึกษาการสกัดคุณลักษณะของเสียง (Features Extraction) เพื่อให้ได้ให้กลุ่มของคุณลักษณะของเสียงที่เหมาะสมต่อการนำไปใช้ในการรู้จำอารมณ์จากเสียงสนทนามากที่สุด

6. ศึกษาเทคนิคในการสร้างตัวแบบสำหรับการรู้จำอารมณ์จากเสียงสนทนา

7. พัฒนาตัวแบบและออกแบบวิธีการวัดประสิทธิภาพของตัวแบบที่สร้างขึ้น

8. ทดสอบระดับของความถูกต้องจากการใช้ตัวแบบการรู้จำอารมณ์จากเสียงสนทนาที่ได้พัฒนาขึ้น

9. สรุปผลของงานวิจัยชิ้นนี้และนำเสนอข้อแนะนำสำหรับการศึกษาต่อเนื่อง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 2 การทบทวนวรรณกรรมที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงวรรณกรรมที่เกี่ยวข้องกับงานวิจัย (Literature Review) และทฤษฎีที่ใช้ในงานวิจัยชิ้นนี้ เพื่อเป็นการศึกษารายละเอียดของงานวิจัยที่เกี่ยวข้องกับการรู้จำอารมณ์ เนื้อหาในบทนี้ประกอบด้วย (1) ลูกค้าและอารมณ์โกรธ (2) การรู้จำอารมณ์ (3) การศึกษาและพัฒนาตัวแบบการรู้จำอารมณ์จากคุณลักษณะทางคลื่นเสียง (4) เทคนิคในการพัฒนาและทดสอบประสิทธิภาพตัวแบบ และ (5) เครื่องมือที่ใช้ในการทำวิจัย

2.1 ลูกค้าและอารมณ์โกรธ

อารมณ์โกรธ (Anger) ของลูกค้าเป็นสิ่งที่องค์กรให้ความสำคัญและพยายามป้องกันไม่ให้เกิดขึ้น โดยอารมณ์โกรธของลูกค้ามักเกิดขึ้นหลังจากที่ลูกค้ารู้สึกว่าองค์กรเกิดความล้มเหลวในการให้บริการ โดยไม่สามารถตอบสนองความต้องการตามที่ตนเองคาดหวัง นอกจากนั้นอารมณ์โกรธที่เกิดขึ้นสามารถนำมาใช้เป็นเครื่องมือในการทำนายพฤติกรรมการตอบสนองของลูกค้าที่จะเกิดขึ้น หากองค์กรมีการตรวจสอบอารมณ์โกรธของลูกค้าอยู่สม่ำเสมอ จะทำให้ทราบปัญหาและปรับปรุงสินค้าและการให้บริการขององค์กรได้

การที่ผู้ให้บริการหรือเจ้าหน้าที่สามารถรับรู้อารมณ์โกรธของลูกค้าได้ในขณะให้บริการทันทีนั้น ถือเป็นสิ่งที่ช่วยเพิ่มประสิทธิภาพในการให้บริการเนื่องจากเมื่อเจ้าหน้าที่ตระหนักว่าลูกค้ากำลังอยู่ในอารมณ์โกรธเจ้าหน้าที่จะสามารถควบบคุมอารมณ์และตัดสินใจตอบสนองในสถานการณ์ได้อย่างเหมาะสม ในทางกลับกันหากเจ้าหน้าที่ไม่ตระหนักถึงอารมณ์ของลูกค้า ลูกค้าอาจรู้สึกความล้มเหลวจากการได้รับบริการมากขึ้น จนกลายเป็นสาเหตุหลักในการตัดสินใจเปลี่ยนไปใช้บริการกับผู้ให้บริการรายอื่น (Dallimore, Sparks, & Butcher, 2007)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2 การรู้จำอารมณ์ (Emotion Recognition)

ความสามารถในการรู้จำอารมณ์เป็นส่วนสำคัญในการสื่อสารระหว่างมนุษย์และระบบคอมพิวเตอร์ซึ่งได้รับความสนใจในการศึกษาเป็นจำนวนมากในช่วงทศวรรษที่ผ่านมา โดยเฉพาะการนำไปประยุกต์ใช้กับศาสตร์ปฏิสัมพันธ์ระหว่างคอมพิวเตอร์และมนุษย์ (Human Computer Interaction) โดยเชื่อว่าหากระบบคอมพิวเตอร์สามารถรับรู้อารมณ์ของมนุษย์ได้จะทำให้การสื่อสารระหว่างมนุษย์และคอมพิวเตอร์มีความเป็นธรรมชาติมากขึ้น (Chakraborty, Pandharipande, & Kopparapu, 2016b) มีงานวิจัยจำนวนมากที่ให้ความสนใจพัฒนาแนวคิดและขั้นตอนวิธีการพัฒนาระบบการรู้จำอารมณ์จากการแสดงออกของมนุษย์ในรูปแบบต่างๆ การแสดงออกของมนุษย์นอกจากจะเป็นการแสดงออกเพื่อสื่อสารข้อความแล้ว ยังเป็นวิธีการแสดงอารมณ์ของมนุษย์ด้วย (Paidi, Kadiri, & Yegnanarayana, 2016)

2.2.1 การรู้จำอารมณ์จากส่วนต่าง ๆ ของร่างกาย

การศึกษาการรู้จำอารมณ์ของมนุษย์ไม่ได้จำกัดอยู่ที่การพูดเท่านั้น แต่ยังมีการศึกษาการรู้จำอารมณ์จากส่วนต่าง ๆของร่างกาย เช่น การรู้จำอารมณ์จากการแสดงออกทางสีหน้า (Emotion Recognition from Facial Expression) ซึ่งเป็นกระบวนการที่ซับซ้อน ต้องการรูปหรือวีดีโอที่มีคุณภาพสูงเพื่อนำมาใช้ในการประมวลผล (Truong, van Leeuwen, & de Jong, 2012; Mariooryad & Busso, 2015; Zeng, Pantic, Roisman, & Huang, 2009) เพื่อการรู้จำอารมณ์จากท่าทางของร่างกาย (Emotion Recognition from Body Gesture) (Truong et al., 2012) เป็นต้น

2.2.2 การรู้จำอารมณ์จากเสียงพูด (Emotion Recognition from Speech)

แนวคิดในการทำให้คอมพิวเตอร์สามารถรู้จำอารมณ์จากเสียงพูดได้ถูกนำเสนอในช่วงปี ค.ศ. 1980 (Pokorny, Graf, Pernkopf, & Schuller, 2015) การพูดเป็นการสื่อสารที่สามารถบอกเล่าเรื่องราวได้อย่างเป็นธรรมชาติซึ่งถูกถ่ายทอดออกมาพร้อมกับอารมณ์ของผู้พูด (Chakraborty et al., 2016a) แม้ว่าในการสนทนาทุกครั้งมนุษย์จะเรียนรู้ที่จะทำความเข้าใจอารมณ์ของอีกฝ่าย แต่กลับพบว่ามนุษย์ไม่สามารถแปลความหมายของอารมณ์ที่ผู้อื่นถ่ายทอดออกมาได้ถูกต้องในทุกครั้ง แม้ว่าอารมณ์นั้นจะเป็นอารมณ์ที่ค่อนข้างชัดเจน อย่างเช่น อารมณ์โกรธ (Petrushin, 2000) การรู้จำอารมณ์จากเสียงพูดถูกจำแนกออกเป็น 2 ประเภท ตามคุณลักษณะของเสียงที่เลือกใช้ในการประมวลผล ได้แก่

1. การรู้จำอารมณ์จากเสียงพูดโดยใช้คุณลักษณะทางภาษาศาสตร์ (Emotion Recognition from Linguistic Feature) ซึ่งเป็นการนำความรู้ด้านการรู้จำเสียงพูด (Speech Recognition) มาใช้โดยการแปลงเสียงพูดเป็นตัวอักษร (Speech to text) จากนั้นนำข้อมูลตัวอักษรที่ได้ไปใช้ในการวิเคราะห์และพัฒนาตัวแบบ

2. การรู้จำอารมณ์จากเสียงพูดโดยใช้คุณลักษณะทางคลื่นเสียง (Emotion Recognition from Acoustic Feature) เป็นการนำคุณลักษณะทางคลื่นเสียง (Acoustic/Audio Feature) มาใช้การรู้จำอารมณ์จากเสียงประเภทนี้จะไม่คำนึงถึงความหมายของคำพูดที่เปล่งออกมา ทำให้ไม่จำเป็นต้องผ่านกระบวนการแปลงเสียงพูดให้อยู่ในรูปตัวอักษร ซึ่งเป็นการลดขั้นตอนในการประมวลผล ส่งผลให้การประมวลผลมีความรวดเร็ว และเป็นประโยชน์อย่างมากต่อการศึกษาการรู้จำอารมณ์ในภาษาที่มีระดับความถูกต้องของเครื่องมือในการแปลงเสียงพูดเป็นตัวอักษรยังไม่สูงนัก งานวิจัยนี้จะนำแนวคิดดังกล่าวมาประยุกต์ใช้ในการศึกษาและพัฒนาตัวแบบการรู้จำอารมณ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

งานวิจัยในด้านการรู้จำอารมณ์จากเสียงพูดโดยใช้คุณลักษณะทางคลื่นเสียง ถูกนำมาผนวกเข้ากับความรู้ในแขนงต่าง ๆ และนำมาประยุกต์เข้ากับข้อมูลเสียงที่ได้จากบริบทที่แตกต่างกัน

ในบริบทของข้อมูลเสียงและศูนย์ให้บริการข้อมูล วิธีการติดต่อระหว่างลูกค้าและศูนย์ให้บริการข้อมูลในอดีตเกิดขึ้นโดย ลูกค้าจะโทรศัพท์เข้ามาเพื่อฝากข้อความเสียงไว้กับระบบของศูนย์ให้บริการข้อมูลโดยส่วนมากจะเป็นเพียงข้อความเสียงขนาดสั้นเท่านั้น จากนั้นเจ้าหน้าที่จะทยอยฟังข้อความเสียงเพื่อตอบกลับ ในงานวิจัยของ Petrushin (Petrushin, 2000) ได้นำการรู้จำอารมณ์จากเสียงมาใช้ โดยการสร้างระบบเรียงลำดับความสำคัญของข้อความเสียงที่ต้องตอบกลับโดยใช้อารมณ์ที่เกิดขึ้นในข้อความเสียงนั้นร่วมกับความยาวของข้อความและข้อสังเกตอื่นเป็นเกณฑ์วัด อารมณ์ที่ระบบรู้จำได้ ได้แก่ อารมณ์ปกติ อารมณ์มีความสุข อารมณ์โกรธ อารมณ์เศร้า และอารมณ์หวาดกลัว โดยระบบจะให้ความสำคัญกับข้อความเสียงที่ถูกรู้จำว่าเกิดอารมณ์โกรธมากที่สุด และให้เจ้าหน้าที่เลือกตอบกลับข้อความเสียงกลุ่มนี้ก่อนเป็นอันดับแรก

ในบริบทของข้อมูลเสียงกับการศึกษา มีการนำเสียงที่ใช้ในการสอนมาหาความสัมพันธ์กับอารมณ์ของนักเรียนและผลการเรียนรู้ที่เกิดขึ้นว่ามีความสัมพันธ์กันอย่างไร ในปัจจุบันโรงเรียนกวดวิชามีจำนวนมากขึ้น บางสถาบันมีสาขาครอบคลุมทั่วประเทศ ทำให้การเรียนการสอนอยู่ในรูปแบบของการนำวีดีโอบันทึกการสอนมาเปิดให้นักเรียนแต่ละสาขาได้เรียนพร้อมกัน ด้วยเหตุนี้จึงทำให้ผู้สอนไม่มีโอกาสได้พูดคุยสื่อสาร สังเกตุ หรือรับรู้อารมณ์ของนักเรียนได้ งานวิจัยดังกล่าวจึงพยายามศึกษาการรู้จำทางอารมณ์ โดยแบ่งสถานะทางอารมณ์ของนักเรียนเป็น 3 อารมณ์ ได้แก่ อารมณ์ปกติ อารมณ์ด้านบวกและอารมณ์ด้านลบ หลังจากทราบอารมณ์ของนักเรียน จะนำความสัมพันธ์ที่หาได้ไปปรับใช้สำหรับการพัฒนาและออกแบบโครงสร้างการเรียนการสอนที่เหมาะสม โดยมีจุดมุ่งหมายเพื่อให้อารมณ์ของนักเรียนอยู่ในสถานะที่เหมาะสมต่อการเรียนรู้อย่างมีประสิทธิภาพมากที่สุด (Litman, Forbes, & Silliman, 2003)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ในด้านสุขภาพและอนามัย มีการนำการรู้จำอารมณ์จากเสียงมาประยุกต์ใช้เป็นเครื่องมือในการรักษาและวินิจฉัยโรคบางชนิด เช่น การประยุกต์ใช้การรู้จำอารมณ์เพื่อการตรวจสอบอาการผิดปกติทางจิต อาการของโรคซึมเศร้า ระดับของความหดหู่ รวมไปถึงการเปลี่ยนแปลงทางอารมณ์อย่างผิดปกติของผู้ป่วย โดยการใช้เสียงพูดของผู้ป่วยเป็นข้อมูลตั้งต้น นอกจากนั้นยังถูกนำไปใช้เป็นเครื่องมือเบื้องต้น ในการวินิจฉัยผู้ป่วยโรคพาร์คินสัน และโรคอัลไซเมอร์ เพื่อตรวจสอบความผิดปกติของความสามารถในการแสดงอารมณ์ของผู้ป่วย

นอกจากตัวอย่างที่ได้ยกขึ้นมาข้างต้นยังมีการนำความสามารถในการรู้จำอารมณ์ไปประยุกต์ใช้กับศาสตร์อื่นอย่างกว้างขวาง เช่น ความปลอดภัยในยานยนต์ เครื่องจับเท็จ รวมไปถึงในอุตสาหกรรมบันเทิงและเกมส์ (Alonso, Cabrera, Medina, & Travieso, 2015)

2.3 การศึกษาและพัฒนาตัวแบบการรู้จำอารมณ์จากลักษณะทางคลื่นเสียง

องค์ประกอบหลักที่สำคัญของการพัฒตัวตัวแบบรู้จำอารมณ์จากคุณลักษณะทางคลื่นเสียงมี 4 องค์ประกอบ ได้แก่ (1) การเลือกฐานข้อมูลที่จะนำมาใช้ในการศึกษาซึ่งชนิดของฐานข้อมูลจะส่งผลโดยตรงต่อการจัดการเตรียมข้อมูลให้พร้อมสำหรับการพัฒนาตัวแบบ (2) การประมวลผลข้อมูลเสียงเบื้องต้น ซึ่งสามารถแยกออกเป็นประเด็นย่อย ได้แก่ การแยกเสียงผู้พูดและการเลือกขนาดข้อมูลเสียงที่จะนำไปใช้ในการพัฒนาตัวแบบ (3) การเลือกคุณลักษณะทางคลื่นเสียงที่สามารถใช้เป็นตัวแปรในการพัฒนาตัวแบบจำแนกอารมณ์ได้ และ (4) การเลือกเทคนิคที่เหมาะสมและค่าที่ใช้ในการประเมินความถูกต้อง โดยทั้ง 4 องค์ประกอบหลัก จะถูกอธิบายโดยละเอียดในหัวข้อถัดไป

2.3.1 คลังข้อมูลเสียง (Audio Corpus)

ข้อมูลเสียงที่นักวิจัยนำมาใช้ในการศึกษาการรู้จำอารมณ์จากเสียงพูด มีความสำคัญและส่งผลต่อผลลัพธ์ของงานวิจัย ข้อมูลเสียงที่เกิดขึ้นจากสถานการณ์จริง มักมีความเป็นธรรมชาติ เป็นการพูดโดยประโยคที่ไม่ซับซ้อน อารมณ์ที่แสดงออกมาพร้อมกับเนื้อเสียงเป็นอารมณ์ที่เกิดขึ้นจริงโดยธรรมชาติ ดังนั้นหากตัวแบบที่นักวิจัยศึกษาและพัฒนาขึ้นโดยใช้ข้อมูลเสียงที่เกิดโดยธรรมชาติ ย่อมทำให้ผลลัพธ์ที่ได้จากงานวิจัย ได้รับการยอมรับด้านความน่าเชื่อถือของข้อมูลและการนำไปประยุกต์ใช้กับสถานการณ์จริง บ่อยครั้งการเข้าถึงคลังข้อมูลเสียงที่เกิดโดยธรรมชาติเป็นเรื่องยาก ปัญหาที่พบส่วนใหญ่ คือ ความกังวลเกี่ยวกับความเป็นส่วนตัวของเจ้าของข้อมูล ประเด็นด้านกฎหมายและศีลธรรม รวมไปถึงข้อห้ามต่าง ๆ สำหรับการนำข้อมูลไปใช้ในการทำวิจัย ทำให้องค์กรที่เป็นเจ้าของข้อมูลเสียงส่วนใหญ่ไม่ต้องการให้ข้อมูลถูกเผยแพร่และไม่อนุญาตให้นำข้อมูลไปใช้สำหรับการทำวิจัย ทางกลับกันคลังข้อมูลเสียงที่เกิดจากการสร้างขึ้นหรือการตั้งใจแสดง เป็นข้อมูลที่สามารถนำมาใช้ได้โดยตรง รวมทั้งมีการจำแนกอย่างชัดเจนไว้แล้วว่าแต่ละประโยคที่สร้างขึ้น คืออารมณ์ใด (Ayadi et al., 2011) แต่ Williams และ Stevens (Williams & Stevens, 1972) พบว่าข้อมูลเสียงที่เกิดจากการตั้งใจแสดงนั้นมักมีการแสดงอารมณ์ที่ค่อนข้างชัดเจน ซึ่งการแสดงอารมณ์ตามธรรมชาติของมนุษย์จะไม่ชัดเจนเท่ากับข้อมูลที่เกิดจากการตั้งใจแสดง (Cao, Verma, & Nenkova, 2015)

งานวิจัยการรู้จำอารมณ์ในอดีตได้มีการเลือกใช้ข้อมูลเสียงจากหลายแหล่งที่มา ทั้งจากการสร้างคลังข้อมูลเสียงขึ้นเอง การใช้คลังข้อมูลเสียงที่ถูกสร้างไว้แล้ว และการนำข้อมูลจริงมาใช้ พบว่างานวิจัยกลุ่มที่นำข้อมูลที่เกิดขึ้นจากการแสดงมาใช้ ได้รับการยอมรับด้านความน่าเชื่อถือของข้อมูลเช่นกัน เนื่องจากขั้นตอนวิธีในการสร้างฐานข้อมูลมีความน่าเชื่อถือ และวัตถุประสงค์ของงานวิจัยเพียงเพื่อต้องการศึกษาและพัฒนาวิธีในการสร้างตัวแบบรู้จำอารมณ์ หรือเพื่อพิสูจน์แนวคิดที่

เกี่ยวข้องที่ทำการศึกษาเท่านั้น แต่ข้อจำกัดที่เด่นชัดของการใช้ข้อมูลประเภทนี้ คือ ผลลัพธ์ของงานวิจัยจะไม่เหมาะสมกับการนำข้อสรุปที่ได้มาอ้างอิงเพื่อใช้ในสถานการณ์จริง (Truong et al., 2012)

ข้อมูลเสียงที่เกิดจากการตั้งใจแสดง (Acted Speech Data)

ข้อมูลเสียงที่เกิดจากการตั้งใจแสดง ได้จากการให้อาสาสมัครพูดและแสดงอารมณ์ตามข้อความและอารมณ์ที่ผู้วิจัยเป็นผู้กำหนด เช่น การสร้างฐานข้อมูลเสียงที่เกิดจากการตั้งใจแสดง ในงานของ Petrushin (Petrushin, 2000) นักวิจัยให้อาสาสมัครจำนวน 30 คนบันทึกเสียงประโยคสั้นๆ จำนวน 4 ประโยคในภาษาอังกฤษ ได้แก่ “This is not what I expected.”, “I’ll be right there.”, “Tomorrow is my birthday.” และ “I’m getting married next week.” แต่ละประโยคถูกบันทึกเสียงจำนวน 5 ครั้ง ครั้งละ 1 อารมณ์ ได้แก่ อารมณ์มีความสุข, อารมณ์โกรธ, อารมณ์เศร้า, อารมณ์กลัว และอารมณ์ปกติ (ไม่มีอารมณ์) และทำการบันทึกเสียงโดยการใช้ไมโครโฟนวางไว้ใกล้กับปากของอาสาสมัคร หลังจากได้ข้อมูลเสียงแล้ว นักวิจัยจะทำการทดลองให้โดยให้อาสาสมัคร 23 คน ซึ่ง 20 คนอยู่ในชุดของอาสาสมัครที่ทำการบันทึกเสียงด้วย ทำการฟังและตอบว่าประโยคที่ได้ฟังนั้นอยู่ในอารมณ์อะไร ผลการทดลองพบว่า อาสาสมัครไม่สามารถจำแนกอารมณ์ได้อย่างถูกต้องทั้งหมด และพบว่าอารมณ์โกรธและอารมณ์เศร้าเป็นอารมณ์ที่อาสาสมัครสามารถแสดงอารมณ์และรับรู้อารมณ์ได้ดีที่สุด (Truong et al., 2012)

อีกหนึ่งตัวอย่างของวิธีการสร้างฐานข้อมูลเสียงที่เกิดจากการตั้งใจแสดง นักวิจัยได้ทำการเตรียมประโยคที่ต้องการให้อาสาสมัครพูดเพื่อบันทึกเสียงจำนวน 60 ประโยค แต่ละประโยคจะถูกบันทึกเสียง 4 ครั้ง ครั้งละ 1 อารมณ์ ได้แก่ อารมณ์สนุก, อารมณ์โกรธ, อารมณ์ประหลาดใจ และอารมณ์เศร้า นักวิจัยทำการเลือกอาสาสมัครเพศชายจำนวน 10 คน การบันทึกเสียงกระทำโดยให้อาสาสมัครนั่งที่หน้าคอมพิวเตอร์หลังจากนั้นจะได้รับรายการของประโยคพร้อมกับอารมณ์ที่นักวิจัยจัดเตรียมไว้ให้แบบสุ่ม แล้วบันทึกเสียงซึ่งวิธีการในการแสดงอารมณ์ อาสาสมัครจะเป็นผู้เลือกแสดงออกเอง หลังจากนั้นนักวิจัยจะทำการเก็บเสียงที่ถูกบันทึกมาย้อนฟังอีกครั้ง หากเสียงของประโยคใดที่มีการแสดงอารมณ์ไม่ชัดเจน เสียงนั้นจะถูกลบและให้อาสาสมัครทำการบันทึกเสียงนั้นใหม่อีกครั้ง (Tang, Chu, Hasegawa-Johnson, & Huang, 2009)

มีงานวิจัยบางกลุ่มที่ไม่ได้สร้างคลังข้อมูลเสียงขึ้นเอง แต่นำใช้คลังข้อมูลเสียงที่ถูกสร้างไว้แล้วเพื่อมาใช้ในการศึกษาด้านการรู้จำอารมณ์ ซึ่ง C.K et al. (C.K et al., 2017) ได้รวบรวมไว้ ได้แก่

1. ฐานข้อมูลเสียง BES เก็บเสียงพูดจากอาสาสมัครชาวเยอรมันที่ใช้ภาษาเยอรมันเป็นภาษาหลัก มีการแสดงออกทางอารมณ์จำนวน 7 อารมณ์ ได้แก่ อารมณ์โกรธ, อารมณ์เศร้า, อารมณ์กระวน

กระวาย, อารมณ์รังเกียจ, อารมณ์ความสุข, อารมณ์เบื่อหน่าย และ อารมณ์ปกติ (Felix Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005)

2. ฐานข้อมูลเสียง SAVEE เป็นข้อมูลเสียงที่ถูกบันทึกโดยอาสาสมัครชาวอังกฤษ เพศชาย จำนวน 4 คน โดยฐานข้อมูลนี้เก็บข้อมูลเสียงทั้งหมด 7 อารมณ์ ได้แก่ อารมณ์ความสุข,อารมณ์ประหลาดใจ, อารมณ์รังเกียจ, อารมณ์กระวนกระวาย, อารมณ์กลัว, อารมณ์ปกติ และ อารมณ์เศร้า (Haq, Jackson, & Edge, 2008)

3. ฐานข้อมูล SUSAS เก็บข้อมูลเสียงที่มีแสดงออกด้านความเครียด และออกแบบวิธีการพูดหลายหลายรูปแบบ (Hansen & Bou-Ghazale, 1997)

ฐานข้อมูลเสียงทั้งสามฐานข้อมูล ถูกนำไปใช้ในงานวิจัยโดยมีวิธีการสกัดคุณลักษณะของข้อมูลและเทคนิคในการจำแนกอารมณ์ที่แตกต่างกัน ตามที่นักวิจัยแต่ละคนได้นำเสนอ

สาเหตุที่ทำให้ผลลัพธ์ของงานวิจัยที่ใช้ข้อมูลจากคลังข้อมูลเสียงที่เกิดจากการตั้งใจแสดงมาใช้ในการทำวิจัยไม่เหมาะสมในการนำไปประยุกต์กับการตรวจจับอารมณ์ในสถานการณ์จริง เพราะประโยคขนาดสั้นที่เกิดจากการแสดง เป็นเพียงเสียงพูดหนึ่งประโยคที่แสดงออกมาพร้อมกับอารมณ์ใดอารมณ์หนึ่งอย่างชัดเจนเท่านั้น แต่ธรรมชาติในการพูดของมนุษย์มีความแตกต่างออกไป ความต่อเนื่องของประโยคในบทสนทนา ความยาวและความซับซ้อนของโครงสร้างประโยค รวมถึงการแสดงออกทางอารมณ์ในสถานการณ์จริงอาจมีมากกว่าหนึ่งอารมณ์ในเสียงสนทนา และไม่ชัดเจนเท่ากับการกำหนดให้อาสาสมัครบันทึกเสียงตามที่นักวิจัยกำหนด (Pokorny et al., 2015) โดยความซับซ้อนจะมากขึ้นหากเป็นข้อมูลเสียงพูดที่มาจากบทสนทนาที่มีผู้พูดจำนวนสองคนขึ้นไป เพราะนอกจาก ความซับซ้อนของประโยคและความไม่ชัดเจนในการแสดงออกทางอารมณ์แล้ว การจัดการข้อมูลเสียงเป็นสิ่งที่ทำให้งานวิจัยมีความท้าทายมากขึ้น เช่น การแยกเสียงของผู้พูด การแบ่งความยาวของเสียง เป็นต้น การไม่สามารถเข้าถึงข้อมูลเสียงที่เกิดขึ้นจริงด้วยจำนวนที่มากพอ ถูกยกให้เป็นอุปสรรคสำคัญในการศึกษาด้านการรู้จำอารมณ์ แต่พบว่ามีหลายงานวิจัยพยายามเข้าถึงและเลือกใช้ข้อมูลบันทึกเสียงสนทนาที่เกิดขึ้นจริง

ข้อมูลเสียงที่เกิดขึ้นจริง (Real-life Data)

งานวิจัยด้านการรู้จำอารมณ์ที่ใช้ข้อมูลเสียงที่เกิดขึ้นจริง มีหลากหลายบริบท งานวิจัยของ Wang (Wang, 2015) ใช้ข้อมูลทั้งหมด 2 บริบท ได้แก่ บันทึกเสียงสนทนาในบริบทศูนย์ให้บริการข้อมูลทางโทรศัพท์ระหว่างเจ้าหน้าที่และลูกค้า และบันทึกเสียงสนทนาในบริบทศูนย์ให้บริการฉุกเฉินของโรงพยาบาล

ในคลังข้อมูลเสียงบันทึกเสียงสนทนาในบริบทศูนย์ให้บริการข้อมูลทางโทรศัพท์ระหว่างเจ้าหน้าที่และลูกค้า มีเสียงของเจ้าหน้าที่เพศหญิง 3 คน เจ้าหน้าที่เพศชาย 2 คน ให้บริการกับลูกค้าทั้งหมด 105 คน แบ่งเป็นลูกค้าเพศหญิงจำนวน 48 คน และเพศชายจำนวน 57 คน การโต้ตอบรวมทั้งเจ้าหน้าที่และลูกค้าแบ่งได้ทั้งหมดจำนวน 121 รอบการโต้ตอบ ความยาวรวมทั้งหมด 3.5 ชั่วโมง

ในฝั่งข้อมูลบริบทที่ 2 ของงานวิจัยนี้ ได้จากบันทึกเสียงสนทนาในบริบทศูนย์ให้บริการฉุกเฉินของโรงพยาบาล คลังข้อมูลนี้ประกอบด้วย บันทึกเสียงระหว่างการให้บริการของเจ้าหน้าที่เพศหญิงจำนวน 3 คน และเพศชายจำนวน 2 คน ผู้ใช้บริการจำนวน 105 คน โดยแบ่งเป็นเพศหญิง 48 คนและเพศชาย 57 คนเท่ากันกับคลังข้อมูลก่อนหน้า การโต้ตอบของเจ้าหน้าที่ และผู้ใช้บริการรวมทั้งหมดเท่ากับ 68 รอบการโต้ตอบ รวมเป็นเวลา 1.8 ชั่วโมง ซึ่งทั้งสองคลังข้อมูลสามารถแสดงสัดส่วนอารมณ์ได้ดังตารางที่ 1

ตารางที่ 1 แสดงรายละเอียดของรอบการโต้ตอบของอารมณ์ที่เกิดขึ้น

ที่มา: Wang (2015)

คลังข้อมูล	**ผู้พูด**	**อารมณ์**					
		โกรธ (%)	**กลัว (%)**	**เศร้า (%)**	**ความสุข (%)**	**อื่นๆ (%)**	**ปกติ (%)**
121 รอบการโต้ตอบ (บริบทศูนย์ให้บริการข้อมูลโทรศัพท์)	ผู้ใช้บริการ	5.70	1.50	2.40	3.50	3.05	83.85
	เจ้าหน้าที่	1.20	0.40	0.30	5.20	2.74	94.16
68 รอบการโต้ตอบ (บริบทศูนย์ให้บริการฉุกเฉินของโรงพยาบาล)	ผู้ใช้บริการ	9.23	5.8	6.80	0.30	1.64	76.23
	เจ้าหน้าที่	1.80	1.00	1.20	2.60	1.54	91.86

นอกจากนั้นยังพบงานวิจัยของ (Devillers & Vidrascu, 2006) มีการนำข้อมูลบันทึกเสียงสนทนาของศูนย์บริการทางการแพทย์ (Medical Call Center) มาใช้ในการศึกษาการตรวจจับอารมณ์ การใช้ข้อมูลประเภทนี้ ผู้วิจัยต้องให้ความสำคัญกับจริยธรรมในการเคารพและรักษาความเป็นส่วนตัวของข้อมูลที่ปรากฏในเสียงสนทนา เนื่องจากข้อมูลในนั้นเป็นข้อมูลส่วนบุคคลที่สามารถใช้ในการอ้างถึงและระบุตัวตนได้ ศูนย์บริการทางการแพทย์ในการศึกษานั้นเปิดให้บริการ 24 ชั่วโมง จุดประสงค์ของการให้บริการนี้คือ การให้คำแนะนำทางการแพทย์ เจ้าหน้าที่จะต้องให้คำปรึกษาแก่ผู้

ใช้บริการอย่างระมัดระวังเพื่อให้ผู้ใช้บริการได้รับข้อมูลที่ถูกต้อง และเกิดประสิทธิภาพมากที่สุด ขั้นตอนในการสนทนาเริ่มจากเจ้าหน้าที่จะสอบถามหัวข้อที่ผู้ใช้บริการต้องการทราบข้อมูล รวมไปถึงข้อมูลที่อยู่ เพื่อให้สามารถทำการช่วยเหลือได้ทันทีหากพบว่าเกิดเหตุฉุกเฉิน พบว่าเสียงบันทึกการสนทนาในกรณีฉุกเฉิน ผู้ป่วยจะแสดงอาการตึงเครียด เจ็บปวด ความหวาดกลัวในการอาการป่วยของตน รวมไปถึงความตื่นตระหนกที่เกิดขึ้น นอกจากนั้นยังพบว่าบางครั้งการสนทนาเป็นการกล่าวถึงบุคคลที่ 3 เช่น ในกรณีที่สมาชิกในครอบครัวของผู้ป่วยติดต่อขอคำปรึกษากับเจ้าหน้าที่แทนผู้ป่วย เมื่อนำข้อมูลบันทึกเสียงมาผ่านกระบวนการเบื้องต้น พบว่ามีเจ้าหน้าที่เพศชายจำนวน 3 คน และเพศหญิงจำนวน 4 คน จำนวนรอบในการโต้ตอบของผู้ใช้บริการรวมทั้ง 784 รอบการโต้ตอบ แบ่งเป็นเสียงเพศชาย 271 รอบ และเพศหญิง 513 รอบ โดยงานวิจัยนี้แบ่งการทดสอบการตรวจจับอารมณ์ออกเป็น 2 ชุด คือ ชุดที่ใช้คุณลักษณะข้อมูลเสียง (Acoustic Feature) และชุดที่ใช้คุณลักษณะทางภาษาศาสตร์ (Linguistic Feature) ผลปรากฏว่าตัวแบบที่เกิดจากทั้งสองชุดของคุณลักษณะ สามารถตรวจจับอารมณ์หวาดกลัวได้ดีที่สุดซึ่งเป็นอารมณ์ที่ปรากฏมากที่สุดในข้อมูลเสียงที่ใช้ในการทำวิจัยครั้งนี้ (Devillers & Vidrascu, 2006)

2.3.2 การประมวลข้อมูลเสียงเบื้องต้น

การประมวลผลข้อมูลเสียงเบื้องต้น คือ การเตรียมข้อมูลเสียงที่นำมาใช้ในการศึกษาให้พร้อมก่อนการนำไปสกัดคุณลักษณะทางคลื่นเสียง ข้อมูลเสียงแต่ละประเภทย่อมมีวิธีการจัดการที่แตกต่างกันออกไป ซึ่งผลลัพธ์จากการเตรียมข้อมูลเสียงนี้จะส่งผลกระทบต่อประสิทธิภาพของตัวแบบโดยตรง โดยทั่วไปสามารถแบ่งออกเป็นขั้นตอนย่อยได้ 2 ขั้นตอน ได้แก่ (1) การแยกเสียงผู้สนทนา (2) การกำหนดขนาดข้อมูลเสียง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การแยกเสียงผู้สนทนา (Speaker Diarization)

หากข้อมูลเสียงที่นำมาใช้ในการศึกษาเป็นข้อมูลเสียงสนทนา ย่อมมีเสียงของผู้พูดมากกว่า 1 เสียง ซึ่งข้อมูลประเภทนี้จะมีประเด็นของการแยกเสียงของผู้สนทนาออกจากกันเข้ามาเกี่ยวข้อง ตัวอย่างข้อมูลเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลซึ่งเป็นการสนทนาที่มีการโต้ตอบกันอย่างต่อเนื่อง การแยกเสียงสนทนาจะเป็นประโยชน์ต่อการวิเคราะห์ข้อมูล เนื่องจากบางครั้ง เจ้าหน้าที่ระดับสูงที่รับผิดชอบดูแลศูนย์ให้บริการข้อมูลมีความสนใจในการวิเคราะห์เสียงของลูกค้าเพียงอย่างเดียว ในทางตรงข้ามหากเจ้าหน้าที่ระดับสูงต้องการวัดผลการดำเนินงานของเจ้าหน้าที่ในด้านของการแสดงออกทางอารมณ์ เสียงของเจ้าหน้าที่จะจำเป็นในการนำไปวิเคราะห์ ประโยชน์ของการแยกเสียงของผู้สนทนาออกจากกัน คือ เพื่อให้สามารถเลือกเสียงของผู้สนทนาได้เหมาะสมกับวัตถุประสงค์ของการวิเคราะห์ในครั้งนั้นมากขึ้น (Chakraborty et al., 2016b)

นอกจากนั้นยังพบว่าบางงานวิจัย เช่น งานวิจัย Pappas et al. ได้เสนอวิธีในการวิเคราะห์ข้อมูลเสียงในบริบทของการสนทนาระหว่างลูกค้าและเจ้าหน้าของศูนย์บริการข้อมูลว่าไม่จำเป็นต้องมีการแยกเสียงของผู้สนทนาออกจากกันเพื่อให้สะดวกต่อการนำไปใช้ในสถานการณ์จริง (Pappas et al., 2015)

การกำหนดขนาดข้อมูลเสียง (Window size)

การกำหนดข้อมูลเสียงหรือความยาวของเสียง ให้อยู่ในขนาดที่เหมาะสมสำหรับนำไปใช้การศึกษาด้านการรู้จำอารมณ์เป็นสิ่งสำคัญ ปัจจุบันยังไม่มีขนาดของข้อมูลเสียงที่ได้รับการยอมรับว่าเป็นขนาดของข้อมูลเสียงที่ดีที่สุดในการเป็นตัวแทนของการนำไปวิเคราะห์อารมณ์จากเสียง (Unit of Analysis) บ่อยครั้งพบว่าการกำหนดขนาดของข้อมูลเสียงจะขึ้นอยู่กับชนิดของคลังข้อมูลเสียงที่นำมาใช้ สิ่งที่สำคัญคือ แฟ้มข้อมูลเสียง 1 แฟ้ม (Audio File) จะต้องมีขนาดหรือความยาวที่เหมาะสมสำหรับการนำใช้เป็นข้อมูลนำเข้า (Input) ในขั้นตอนการสกัดคุณลักษณะทางคลื่นเสียง โดยผลลัพธ์ (Output) หรือคุณลักษณะที่ได้จะเป็นตัวแทนของแฟ้มข้อมูลเสียงดังกล่าว จากงานวิจัยในอดีตสามารถจำแนกขนาดของข้อมูลเสียงได้ 3 รูปแบบดังนี้

- **ขนาดคงที่ (Fixed Size)**

เป็นการกำหนดขนาดข้อมูลแบบคงที่ โดย Pappas, Androutsopoulos และ Papageorgiou ทำการแบ่งเสียงของการสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลให้มีขนาด 1 วินาที แบบไม่มีการทับซ้อนกัน และนำแฟ้มข้อมูลเสียงขนาด 1 วินาทีนี้ไปใช้ในการสกัดคุณลักษณะทางคลื่นเสียง (Pappas et al., 2015) การกำหนดขนาดของข้อมูลเสียงโดยวิธีนี้ยังไม่มีการเปรียบเทียบว่าขนาดคงที่ที่เหมาะสมต่อการนำไปใช้ในการศึกษาด้านการรู้จำอารมณ์เป็นขนาดเท่าไร

- **ขนาดข้อมูลจริง (Data Size)**

งานวิจัยที่ใช้คลังข้อมูลเสียงส่วนใหญ่นิยมใช้วิธีนี้ในการกำหนดขนาดให้กับแฟ้มข้อมูลเสียง เนื่องจากคลังข้อมูลที่สร้างขึ้นจากห้องทดลอง (Tang et al., 2009; C.K et al., 2017) ดังเช่นได้อธิบายไว้ในหัวข้อคลังข้อมูลเสียง ว่านักวิจัยจะให้อาสาสมัครบันทึกเสียงเป็นประโยคขนาดสั้น โดยแต่ละประโยคไม่มีส่วนเกี่ยวข้องกัน และแสดงอารมณ์ต่างกัน ดังนั้นนักวิจัยจะใช้ความยาวเท่ากับหนึ่งประโยคในการวิเคราะห์การรู้จำอารมณ์ของประโยคนั้น ๆ งานของ เมธี เจริญดี (เมธี เจริญดี, 2016) ใช้คลังข้อมูลเสียงละครไทยมาใช้ในการศึกษา เช่นเดียวกัน แต่ละประโยคที่ได้จากบันทึกเสียงละครไทยใช้แทนหนึ่งหน่วยในการวิเคราะห์ข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

- **ขนาดรอบการโต้ตอบ (Turn Size)**

การศึกษาการรู้จำอารมณ์ที่ใช้ข้อมูลจากบันทึกเสียงสนทนาส่วนใหญ่จะใช้ รอบการโต้ตอบเป็นปัจจัยในการกำหนดขนาดของข้อมูลเสียง (Chakraborty et al., 2016a; Chakraborty et al., 2016b; Wang, 2015; Vidrascu & Devillers, 2007; F. Burkhardt, Polzehl, Stegmann, Metze, & Huber, 2009)

รอบการโต้ตอบ (Turn) เริ่มต้นเมื่อผู้สนทนาเริ่มเปล่งเสียงจนกระทั่งสิ้นสุดการพูด ซึ่งรอบการโต้ตอบถัดไปอาจเป็นการพูดซ้ำของผู้พูดเดิม หรือผู้พูดอื่นก็ได้ (Wang, 2015) ยกตัวอย่างโดยใช้บทสนทนาระหว่างนายสมชายและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูล

นายสมชาย : สวัสดีครับ ผมชื่อสมชาย ต้องการติดต่อเรื่องโปรโมชั่นโทรศัพท์ครับ

เจ้าหน้าที่ : ค่ะ คุณสมชาย ขอทราบหมายเลขโทรศัพท์ที่ต้องการดำเนินการด้วยค่ะ

นายสมชาย : เบอร์ 0801234567 ครับ

เจ้าหน้าที่ : ค่ะ ไม่ทราบว่าคุณสมชาย อยากได้โปรโมชั่นประเภทไหนคะ เน้นการใช้อินเตอร์เน็ตหรือการโทรศัพท์มากกว่ากันคะ

จากตัวอย่างข้างต้นรอบการโต้ตอบที่หนึ่งของนายสมชาย คือ “สวัสดีครับ ผมชื่อสมชาย ต้องการติดต่อเรื่องโปรโมชันโทรศัพท์ครับ” ซึ่งเมื่อนายสมชายพูดรอบการโต้ตอบนี้จบ เจ้าหน้าที่จึงเริ่มพูดรอบการโต้ตอบของเจ้าหน้าที่ถัดไป

ในการสนทนาตามธรรมชาติรอบการโต้ตอบของการสนทนาย่อมมีขนาดหรือความยาวที่แตกต่างกัน บางครั้งผู้พูดอาจกล่าวแค่คำสั้นๆ เช่น “ค่ะ” “ครับ” เท่านั้น การกำหนดขนาดข้อมูลเสียงโดยวิธีนี้จึงทำให้ขนาดของข้อมูลที่ได้มีความแตกต่างกันตามความยาวของแต่ละรอบการโต้ตอบ (Chakraborty et al., 2016a) และคุณลักษณะที่สกัดได้จะเป็นตัวแทนของรอบการโต้ตอบนั้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.3.3 คุณลักษณะทางคลื่นเสียง

ตัวแบบรู้จำอารมณ์จากเสียงพูดถูกพัฒนาโดยการใช้คุณลักษณะของเสียงหลายกลุ่มที่แตกต่างออกไป บางคุณลักษณะจะขึ้นอยู่ผู้พูดและภาษาที่ใช้พูด ดังเช่นงานวิจัยที่ใช้คุณลักษณะในกลุ่มของภาษาศาสตร์ (Linguistic Feature) ซึ่งตัวแบบที่ดีควรใช้งานได้กับผู้พูดและภาษาที่หลายหลาย ไม่ควรขึ้นอยู่กับข้อมูลส่วนบุคคลของผู้พูด (Koolagudi & Rao, 2010) นอกจากนั้นสิ่งสำคัญในการพัฒนาตัวแบบรู้จำอารมณ์จากเสียงพูด คือ การเลือกคุณลักษณะที่เหมาะสมที่สามารถใช้เป็นตัวแทนของอารมณ์จากข้อมูลเสียงนั้น ๆ เพื่อให้ตัวแบบสามารถทำนายอารมณ์ได้อย่างมีประสิทธิภาพ (Ayadi et al., 2011)

คุณลักษณะทางคลื่นเสียง (Acoustic Feature) ที่สกัดได้จากเสียงพูดหรือเสียงสนทนาเป็นปัจจัยพื้นฐานในการพัฒนาตัวแบบรู้จำอารมณ์ (Dave, 2013) เนื่องจากคุณลักษณะทางคลื่นเสียงสามารถใช้ตัวแปรในการจำแนกอารมณ์แต่ละอารมณ์ที่เกิดขึ้นพร้อมกับเสียงพูดได้ เช่น อารมณ์หวาดกลัว, อารมณ์โกรธ, อารมณ์เศร้า และ อารมณ์เพลิดเพลิน ได้อย่างชัดเจน (Murray & L. Arnott, 1993; Pittam & Scherer, 1993, Tischer, 1994) คุณลักษณะทางคลื่นเสียงที่นิยมนำมาใช้ได้แก่ พิช (Pitch หรือ Fundamental F0) พิชเป็นคุณลักษณะทางคลื่นเสียงที่ทุกงานวิจัยด้านการรู้จำอารมณ์โดยใช้คุณลักษณะทางคลื่นเสียงเห็นตรงกันว่าเป็นคุณลักษณะหลักที่สามารถใช้เป็นตัวแปรในการรู้จำอารมณ์ (Petrushin, 2000) นอกจากนั้นยังมี พลังงาน (Energy), เอ็มเอฟซีซี (MFCC) และ คุณลักษณะในคลื่นเสียงในกลุ่มของคุณภาพเสียง (Voice Quality Feature)

จากงานวิจัยในการแข่งขันงานประชุมวิชาการเกี่ยวกับคลื่นเสียง (INTERSPEECH 2010) ของ Schuller ได้นำเสนอกลุ่มของคุณลักษณะของคลื่นเสียงที่เหมาะสมต่อการนำไปใช้ในการพัฒนาตัวแบบการตรวจจับอารมณ์ของเสียงไว้จำนวนทั้งหมด 1,582 คุณลักษณะ (Schuller et al., 2010) สามารถแสดงได้ดังแสดงในตารางที่ 2

ตารางที่ 2 แสดงคุณลักษณะทางคลื่นเสียงที่นำเสนอ
ที่มา: Schuller et al., 2010

กลุ่มที่	กลุ่มของคุณลักษณะ	จำนวน	จำนวนฟังก์ชันทางสถิติ	รวม
1	Low-Level Descriptive (LLD)	34	21	1,428
2	Low-Level Descriptive (LLD) delta coefficients	34		
3	Pitch-based LLD	4	19	152
4	Pitch-based LLD delta coefficients	4		
5	Number of pitch onsets (Pseudo Syllables)	1	-	2
6	Total duration	1		
จำนวนคุณลักษณะรวม				1,582

2.3.4 เทคนิคการพัฒนาตัวแบบ

งานศึกษาการพัฒนาตัวแบบรู้จำอารมณ์ในช่วงแรก ตัวจำแนก (Classifier) ที่นิยมนำมาใช้ คือ ลิเนียร์ดิสคริมิแนนส์ Linear Discriminant (LDCs) และ เค เนียร์เรส เนเบอร์ (K-Nearest Neighbor (KNN)) อย่างไรก็ตาม เมื่อคุณลักษณะที่นักวิจัยเสนอว่า สามารถใช้ในการพัฒนาตัวแบบรู้จำอารมณ์มีจำนวนมากขึ้นซึ่งนำไปสู่ปัญหา มิติของตัวแปรที่มากเกินไป “Curse of dimensionality” นอกจากนั้น เคเอ็นเอ็น ยังมีความอ่อนไหวกับค่าที่ผิดปกติ หรือค่าที่มีค่าสูงหรือต่ำกว่าค่ากลางจนถูกจัดเป็นค่าที่ผิดปกติ (Outliers Sensitivity)

ในระยะถัดมา เอสวีเอ็ม (Support Vector Machine) กลายเป็นตัวจำแนกที่ถูกนำมาใช้อย่างแพร่หลาย เอสวีเอ็ม สามารถแก้ไขปัญหาที่ไม่ใช่สมการเส้นตรง (Non-linear) โดยใช้การแบ่งกลุ่มตำแหน่งด้วยวิธีการทางเคอร์เนลล์ (Kernel-based transformation of the feature space) ได้ ด้วยเหตุนี้ เอสวีเอ็ม จึงกลายเป็นตัวจำแนกที่ได้รับความนิยมนำมาใช้กับงานด้านรู้จำอารมณ์จนถึงปัจจุบัน (Schuller, 2011)

Pappas et al. ศึกษาการรู้จำอารมณ์ในบริบทของบันทึกเสียงสนทนาของศูนย์ให้บริการข้อมูลได้นำ แอลอาร์ (Logistic Regression) มาใช้เป็นตัวจำแนก โดยให้ตัวจำแนกเรียนรู้และทดสอบทำซ้ำทั้งหมด 21 รอบ โดยในแต่ละรอบนั้นสุ่มเลือกแฟ้มข้อมูลบันทึกเสียงสนทนา ให้ผู้พูดไม่ซ้ำกันทั้งชุดข้อมูลสำหรับเรียนรู้และทดสอบ ผลการวัดประสิทธิภาพผลพบว่าตัวจำแนกให้ค่าความถูกต้องเฉลี่ยร้อยละ 70.31% (Pappas et al., 2015)

อีก 2 เทคนิคที่นิยมนำมาใช้ในงานประเภทการจำแนก คือ ดิซิสชั่นทรี (Decision Tree) และแรนดอมฟอร์เรส (Random Forest) เทคนิคดิซิสชั่นทรี เป็นเทคนิคเบื้องต้นที่นิยมใช้ในงานที่เกี่ยวข้องกับการทำเหมืองข้อมูล เนื่องจากเป็นเทคนิคที่ไม่ซับซ้อน สามารถเข้าใจการทำงานของตัวต้นไม้ (Tree) และแผนผังของต้นไม้ได้ง่าย โดยแผนผังต้นไม้จะแสดงความสัมพันธ์ของการเลือกตัดสินใจของต้นไม้นั้น ๆ ได้ ซึ่งเทคนิคแรนดอมฟอร์เรส เป็นเทคนิคที่พัฒนามาจากดิซิสชั่นทรี โดยการเพิ่มจำนวนต้นไม้ในป่า เพื่อนำค่าที่ได้จากต้นไม้แต่ละต้นมาหาค่าเฉลี่ยเป็นผลลัพธ์รวม จุดเด่น คือ แรนดอมฟอร์เรสเป็นเทคนิคที่มีการช่วยเลือกตัวแปร การแปลงตัวแปร และการรวมตัวแปร โดยรวมเรียกว่า Feature Engineering

นอกจากเทคนิคที่ได้กล่าวไปข้างต้น ผู้วิจัยยังพบว่าในปัจจุบัน โครงข่ายประสาทเทียมเชิงลึก (Deep Neural Network) เป็นอีกหนึ่งเทคนิคที่มีประสิทธิภาพสูง เป็นการนำโครงข่ายเทียมหลายโครงข่ายมาช่วยกันประมวลผลซึ่งเป็นการเลียนแบบการทำงานของสมองมนุษย์ ข้อได้เปรียบของโครงข่ายเทียมเชิงลึก คือ จะทำการเรียนรู้จากชุดข้อมูลฝึกสอน และปรับตัวให้ทำงานได้เมื่อพบข้อมูลอื่น ๆ ที่มีลักษณะใกล้เคียงกัน

จากการทบทวนวรรณกรรมข้างต้น สามารถสรุปทฤษฎีและแนวคิดที่ผู้วิจัยนำไปศึกษาต่อได้ดังนี้ งานวิจัยนี้เป็นการพัฒนาตัวแบบรู้จำอารมณ์โกรธ ใช้ข้อมูลที่เกิดขึ้นจากการสนทนาโดยธรรมชาติในบริบทของเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลของบริษัทให้บริการเครือข่ายโทรศัพท์เคลื่อนที่แห่งหนึ่งของประเทศไทย โดยพิจารณาเฉพาะเสียงสนทนาของลูกค้าเท่านั้น ขนาดของแฟ้มข้อมูลเสียงที่ใช้ในการสกัดคุณลักษณะเท่ากับ 1 รอบการโต้ตอบ หลังจากนั้นจึงนำไปสกัดคุณลักษณะทางคลื่นเสียงโดยอ้างอิงชุดคุณลักษณะจากการแข่งขันการประชุมสัมนาเกี่ยวกับคลื่นเสียงปี 2010 ได้คุณลักษณะทั้งสิ้น 1,582 คุณลักษณะ และนำไปเรียนรู้โดยใช้ตัวจำแนก 4 ชนิด คือ แรนดอมฟอร์เรส, ดิซิสชั่นทรี, ลิเนียร์เอสวีซี และโลจิสติกส์รีเกรสชั่น

2.4 เทคนิคในการพัฒนาและทดสอบประสิทธิภาพตัวแบบ

2.4.1 วิธีในการประเมินประสิทธิภาพตัวแบบ

การพัฒนาตัวแบบในแต่ละครั้ง นักวิจัยจะพัฒนาตัวแบบโดยทดลองใช้เทคนิคหลายเทคนิคแตกต่างกัน ซึ่งหลักการที่นิยมนำมาใช้ในการตัดสินใจเลือกเทคนิคที่จะนำใช้ในการพัฒนาตัวแบบได้แก่

การตรวจสอบไขว้ (K-Fold Cross Validation)

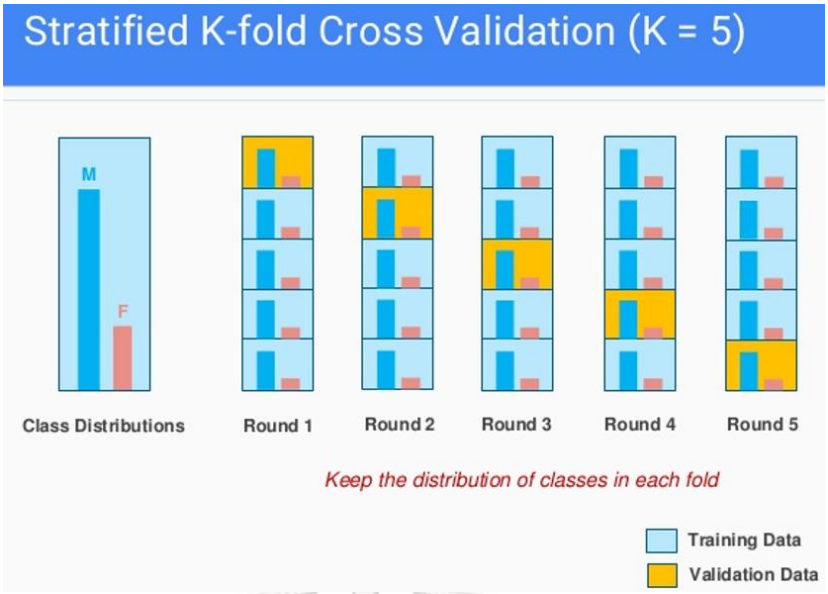
การแบ่งข้อมูลออกเป็นชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบโดยวิธีนี้ทำให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือ วิธีการ คือ แบ่งข้อมูลออกเป็น K ส่วน โดยแต่ละส่วนจะมีจำนวนข้อมูลเท่ากัน หลังจากนั้นจะนำข้อมูล 1 ส่วนจาก K ส่วนที่แบ่งได้มาใช้เป็นชุดข้อมูลทดสอบ (Testing Data) และข้อมูลอีก K-1 ส่วน จะถูกใช้เป็นชุดข้อมูลฝึกสอน (Training Data) แสดงรายละเอียดกระบวนการ K ครั้ง และนำค่าวัดประสิทธิภาพตัวแบบที่ได้จากการทดลอง K ครั้ง มาหาค่าเฉลี่ย ประโยชน์ของการตรวจสอบไขว้ คือ ทุกข้อมูลโดยการใช้เป็นข้อมูลทดสอบและกระบวนการสร้างตัวแบบและการทดสอบ มีการทำซ้ำมากกว่า 1 ครั้ง ซึ่งทำให้สามารถลดปัญหาอคติที่เกิดจากการเลือกชุดข้อมูลฝึกสอนและทดสอบได้

ค่า K ที่นิยมใช้และได้รับการยอมรับ คือ K=10 เพราะจะเหลือข้อมูลไว้สำหรับการฝึกสอน (Training) ถึง 90% ในแต่ละรอบ และอีก 10% ไว้สำหรับการทดสอบ (Testing) โดยสามารถแสดงกระบวนการได้ดังภาพที่ 1

ภาพที่ 1 แสดงการแบ่งข้อมูลชนิดการตรวจสอบไขว้เมื่อกำหนด K=10

การตรวจสอบไขว้แบบแบ่งชั้นภูมิ (Stratified K-Fold Cross Validation)

การตรวจสอบไขว้แบบแบ่งชั้นภูมิ (Stratified K-Fold Cross Validation) มีหลักการแบ่งข้อมูลสำหรับชุดฝึกสอนและชุดทดสอบเช่นเดียวกันกับการตรวจสอบไขว้ สิ่งที่ต่างกันคือ ในกรณีที่ข้อมูลที่นำใช้ในการพัฒนาตัวแบบมีจำนวนข้อมูลของแต่ละคลาสไม่สมดุลกัน (Imbalanced Data) การตรวจสอบไขว้แบบแบ่งชั้นภูมิ จะทำการแบ่งข้อมูลของแต่ละคลาสให้มีสัดส่วนเท่าเดิมสำหรับในแต่ละชุดข้อมูล ตัวอย่างเช่น ข้อมูลทั้งหมด แบ่งเป็น คลาส A จำนวน 70 เปอร์เซ็นต์และคลาส B จำนวน 30 เปอร์เซ็นต์ เมื่อนำมาใช้กับการตรวจสอบไขว้แบบแบ่งชั้นภูมิ K ส่วน ข้อมูล 1 ส่วน จาก K ส่วน จะประกอบไปด้วยข้อมูลจากคลาส A และ คลาส B จำนวน 70 เปอร์เซ็นต์และ 30 เปอร์เซ็นต์ตามลำดับ ภาพที่ 2 เป็นตัวอย่างการแบ่งข้อมูลด้วยวิธีการตรวจสอบไขว้แบบแบ่งชั้นภูมิเมื่อกำหนด K = 5



ภาพที่ 2 แสดงตัวอย่างการแบ่งข้อมูลแบบการตรวจสอบไขว้แบบแบ่งชั้นภูมิ 5 ชุด

ที่มา : https://www.slideshare.net/markpeng/general-tips-for-participating-kaggle-competition

2.4.2 ค่าวัดประสิทธิภาพตัวแบบ

การเลือกค่าวัดประสิทธิภาพของตัวแบบจะต้องสอดคล้องกับสิ่งที่ต้องการทราบจากตัวแบบ เนื่องจากค่าวัดแต่ละค่าจะบ่งบอกประสิทธิภาพของตัวแบบในแง่มุมที่ต่างกัน ค่าวัดประสิทธิภาพของตัวแบบที่นิยมใช้ในตัวแบบประเภทไบนาลีคลาส (Binary Class) หรือตัวแบบที่มีคำตอบเพียงสองทางเลือก ส่วนใหญ่คำนวณได้จากค่าการทำนายของตัวแบบ 4 ค่า ผู้วิจัยจะใช้บริบทของการทำนายรอบการโต้ตอบที่เป็นอารมณ์โกรธและอารมณ์ปกติ (ไม่โกรธ) มาใช้เป็นตัวอย่างในการอธิบาย โดยทั่วไปแล้วจะกำหนดให้ ผลบวก (Positive) เป็นคลาสที่สนใจต้องให้คลาสนี้ทำนายได้ถูกต้องมากที่สุด และผลลบ (Negative) จะแทนคลาสที่เป็นผลตรงข้ามกับคลาสที่สนใจในการศึกษาครั้งนี้ กำหนดให้ผลบวก คือ รอบการโต้ตอบที่เป็นอารมณ์โกรธ และผลลบ คือ รอบการโต้ตอบที่เป็นอารมณ์ปกติ ค่าที่ได้จากการทำนายทั้ง 4 ค่า อธิบายได้ดังนี้

1. **ผลบวกจริง (TP : True Positive)** จำนวนครั้งที่ตัวแบบสามารถทำนายรอบการโต้ตอบที่เป็นอารมณ์โกรธได้ถูกต้อง

2. **ผลลบปลอม (FN : False Negative)** จำนวนครั้งที่ตัวแบบทำนายรอบการโต้ตอบที่เป็นอารมณ์โกรธผิดเป็นอารมณ์ปกติ ซึ่งเป็นสิ่งที่ไม่ต้องการให้เกิดขึ้นมากที่สุด ถูกจัดว่าเป็นความผิดพลาดที่ร้ายแรง

3. **ผลลบจริง (TN : True Negative)** จำนวนครั้งที่ตัวแบบทำนายรอบการโต้ตอบที่เป็นอารมณ์ปกติได้ถูกต้อง

4. **ผลบวกปลอม (FP : False Positive)** จำนวนครั้งที่ตัวแบบทำนายรอบการโต้ตอบที่เป็นอารมณ์ปกติผิดเป็นอารมณ์โกรธ

โดยปกติแล้วค่าทั้ง 4 ค่านิยมนำมาแสดงในตารางคอนฟิวชันเมทริกซ์ (Confusion Matrix) ซึ่งเป็นตารางที่ใช้แสดงค่าวัดประสิทธิภาพแบบแจกแจงเป็นรายคลาสตารางที่ 3

ตารางที่ 3 ตัวอย่างตารางคอนฟิวชั่นเมทริกซ์

<table>
<tr><td></td><td></td><td colspan="2">Predicted</td><td></td><td></td></tr>
<tr><td></td><td></td><td>Angry</td><td>Normal</td><td>Total</td><td></td></tr>
<tr><td rowspan="2">Actual</td><td>Angry</td><td>True Positive</td><td>False Negative</td><td>Actual Angry</td><td></td></tr>
<tr><td>Normal</td><td>False Positive</td><td>True Negative</td><td>Actual Normal</td><td></td></tr>
<tr><td></td><td>Total</td><td>Predicted Angry</td><td>Predicted Normal</td><td>Total Turn</td><td></td></tr>
<tr><td colspan="5"></td><td></td></tr>
<tr><td colspan="2">Accuracy
(TP+TN) / Total Turn</td><td></td><td colspan="2">Balanced Accuracy
(Sensitivity / Specificity) / 2</td><td></td></tr>
<tr><td colspan="2">Recall (Angry) , Sensitivity
(TP / Actual Angry)</td><td></td><td colspan="2">Recall (Normal) , Specificity
(TN / Actual Negative)</td><td></td></tr>
<tr><td colspan="2">Precision (Angry)
(TP / Predicted Angry)</td><td></td><td colspan="2">Precision (Normal)
(TN / Predicted Normal)</td><td></td></tr>
<tr><td colspan="2">Prevalence (Angry)
(Actual Angry / Total Turn)</td><td></td><td colspan="3">F1 Score (Angry)
2 * ((Recall (Angry) * Precision (Angry) /
(Recall (Angry) + Precision (Angry))</td></tr>
</table>

ค่าวัดประสิทธิภาพตัวแบบที่แสดงในตารางที่ 3 สามารถอธิบายเพิ่มเติมได้ดังนี้

1. ค่าความถูกต้อง (Accuracy)

เป็นการวัดประสิทธิภาพโดยรวมของตัวแบบ โดยวิธีการคำนวณจากจำนวนคำตอบที่ตัวแบบตอบถูกต้องทั้งหมดทั้งทางที่เป็นอารมณ์โกรธและอารมณ์ปกติ เทียบกับจำนวนข้อมูลทั้งหมดที่ต้องทำนาย

2. ค่าความระลึก (Recall)

เป็นการวัดประสิทธิภาพว่าตัวแบบมีความครอบคลุมในการทำนายข้อมูลของให้กับแต่ละคลาส มากน้อยเพียงใด ผลลัพธ์จากคำนวณจะทำให้สามารถหาแนวโน้มว่าตัวแบบนี้เหมาะสมกับการทำนาย คลาสใด เช่น หากค่าระลึกของอารมณ์โกรธมากกว่าค่าระลึกของอารมณ์ปกติ จะหมายความว่า ตัวแบบที่พัฒนาขึ้นเหมาะสมกับการทำนายอารมณ์โกรธได้มากกว่า นอกจากนี้ค่าความระลึกยังมีชื่อเรียกเฉพาะสำหรับตัวแบบประเภทไบนาลี คือ ค่าความไว (Sensitivity) สำหรับค่าความระลึกของคลาสผลบวก และ(2) ความความเฉพาะ (Specificity) สำหรับค่าความระลึกของคลาสผลลบ

3. ค่าความแม่นยำ (Precision)

เป็นการวัดประสิทธิภาพของตัวแบบด้านความยากง่ายของตัวแบบที่จะทำนายคำตอบของแต่ละคลาสเมื่อเทียบกับจำนวนครั้งที่ทำนายทั้งหมด โดยคำตอบที่ตัวแบบทำนายผิดจะถือเป็นอุปสรรคหรือสิ่งรบกวนที่เกิดขึ้น ค่าความแม่นยำที่มากกว่าจะแสดงถึงตัวแบบสามารถทำนายคำตอบได้ถูกต้องของคลาสนั้น ๆ ได้ง่ายกว่า

4. ค่าความถูกต้องสมดุล (Balanced Accuracy)

ในตัวแบบที่ถูกพัฒนาขึ้นโดยมีจำนวนข้อมูลในแต่ละคลาสไม่เท่ากัน นิยมใช้ค่าความถูกต้องสมดุลมาเป็นหนึ่งในค่าวัดประสิทธิภาพตัวแบบ เนื่องจากค่าความถูกต้องสมดุลเป็นค่าที่แสดงถึงความแม่นยำในการทำนายเฉลี่ยของทั้งสองคลาส เกิดจากการนำค่าระลึกของทั้งสองคลาสมาหาค่าเฉลี่ย ทำให้ค่าที่ได้ไม่ขึ้นอยู่กับจำนวนข้อมูล เช่น ตัวแบบทำนายอารมณ์โกรธได้ถูกต้องเพียงเล็กน้อย แต่ทำนายอารมณ์ปกติได้จำนวนมาก เมื่อนำมาคำนวณหาค่าความถูกต้อง ย่อมทำให้ค่าความถูกต้องนั้นมีค่าสูง แต่เป็นค่าที่ถูกดึงขึ้นจากฝั่งอารมณ์ปกติ เพื่อเป็นการจัดการกับปัญหาดังกล่าวค่าความถูกต้องสมดุลจึงถูกนำมาใช้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5. ค่าความชุก (Prevalence)

แสดงถึงความชุกของข้อมูลในแต่ละคลาสเมื่อเทียบกับจำนวนข้อมูลนำเข้าทั้งหมด

6. ค่าเอฟวัน (F1 Score)

ค่าน้ำหนักเฉลี่ยระหว่างค่าความระลึกและค่าความแม่นยำของแต่ละคลาส เพื่อแสดงถึงค่าเฉลี่ยระหว่างความแม่นยำที่ทำนายได้และความง่ายในการทำนายเมื่อเทียบกับอุปสรรค เป็นการวัดประสิทธิภาพของตัวแบบโดยพิจารณาทั้งค่าความระลึก และค่าความแม่นยำโดยค่าที่ดีที่สุด คือ 1 และค่าที่แย่ที่สุด คือ 0

7. กราฟโค้งอาร์โอซี (Roc Curve: Receiver Operating Characteristic Curve)

เป็นกราฟที่นิยมใช้ในการแสดงความสามารถของการจำแนก สำหรับงานวิจัยทางด้านเหมืองข้อมูลกราฟโค้งอาร์โอซีจะเป็นกราฟที่ใช้แสดงความสามารถในการทำนายของตัวแบบประเภทที่เป็นไบนาลีคลาส

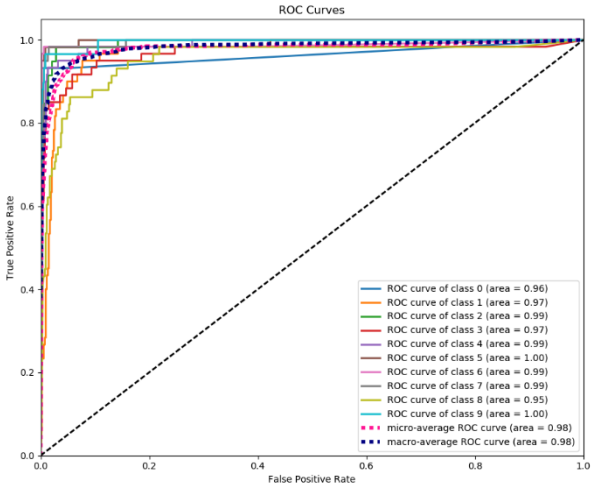
กราฟโค้งอาร์โอซีเป็นกราฟที่เกิดจากการกำหนดจุดความสัมพันธ์ระหว่าง ค่าอัตราผลบวกจริง (True Positive Rate) และค่าอัตราผลบวกปลอม (False Positive Rate) ซึ่งทั้งสองค่าสามารถคำนวณได้ดังนี้

$$Ture\ Positive\ Rate = \frac{Ture\ Positive}{Actual\ Postive\ Class}$$

และ

$$False\ Positive\ Rate = \frac{Ture\ Negative}{Actual\ Negative\ Class}$$

แกนตั้งแทนค่าอัตราผลบวกจริง และแกนนอนแทนค่าอัตราผลบวกปลอม ผลลัพธ์ในอุดมคติของตัวแบบคือ ต้องการให้ตัวแบบสามารถทำนายคลาสที่สนใจได้ถูกต้อง 100 เปอร์เซ็นต์ซึ่งจะได้ค่าอัตราผลบวกจริงเท่ากับ 1 และในทางตรงข้ามต้องการให้ตัวแบบมีอัตราผลลบจริงเท่ากับ 0 ดังนั้นจุดอุดมคติบนกราฟ คือ (0,1) (Perfect Classification) ค่าอัตราผลบวกจริงและอัตราผลบวกปลอมจะแปรผันไปเมื่อเพดานของความน่าจะเป็น (probability threshold) ในการจำแนกคลาสเปลี่ยนแปลงไป จากตัวอย่างในภาพที่ 3 เส้นประทแยงมุมแทนโอกาสของผลทำนายที่จะเกิดขึ้นของ 2 คลาสอย่างสุ่ม นั่นคือ 50 : 50 เปอร์เซ็นต์ โดยเส้นประโค้ง จะหมายถึงประสิทธิภาพของตัวแบบที่ได้จากการตรวจสอบไขว้ซึ่งแสดงให้เห็นถึงประสิทธิภาพในการทำนายที่ดีกว่าการทำนายอย่างสุ่ม



ภาพที่ 3 ตัวอย่างกราฟโค้งอาร์โอซี

8. ค่าเอยูซี (AUC Score: Area Under Curve Score)

จากกราฟโค้งอาร์โอซี พื้นที่ใต้กราฟจะแสดงถึงประสิทธิภาพในการทำนายของตัวแบบ ซึ่งพื้นที่ใต้กราฟดังกล่าว เรียกว่า ค่าเอยูซี พื้นที่ใต้กราฟขนาดใหญ่กว่าหรือค่าเอยูซีที่สูงกว่าแสดงถึงประสิทธิภาพในการจำแนกที่มากกว่าของตัวแบบที่กำลังพิจารณา ค่าเอยูซีมักได้รับการนำมาใช้ในการเปรียบเทียบประสิทธิภาพของตัวแบบหลายตัวแบบพร้อมกัน

2.5 เครื่องมือที่ใช้ในการทำวิจัย

2.5.1 เครื่องมือสำหรับการแยกเสียงสนทนา

สำหรับงานวิจัยที่นำข้อมูลเสียงที่เป็นการสนทนามาใช้ในการศึกษา จะมีประเด็นเรื่องการแยกเสียงสนทนาของผู้พูดออกจากกันก่อนการนำไปวิเคราะห์ เพื่อให้ผลลัพธ์ที่ได้สามารถอ้างอิงถึงเสียงของผู้พูดเฉพาะกลุ่มที่สนใจ ตัวอย่างเช่น การแยกเฉพาะเสียงของลูกค้าออกจากเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ ซึ่งวิธีการในการแยกเสียงผู้พูดสามารถกระทำได้หลายวิธี ขึ้นอยู่กับหลายปัจจัย สามารถยกตัวอย่างวิธีการแยกเสียงผู้พูดจากกันได้ดังนี้

การแยกเสียงสนทนาโดยการใช้ตัวแบบ (Speaker Diarization Model) จะเป็นการแยกเสียงสนทนาของผู้พูดออกจากกันโดยการใช้ตัวแบบที่ผ่านการฝึกสอนจากชุดข้อมูลการฝึกสอน หลักการทำงานเบื้องต้นคือ จะใช้แฟ้มข้อมูลเสียงที่ต้องการแยกเสียงผู้สนทนาเป็นข้อมูลนำเข้า จากนั้นตัวแบบจะสามารถให้คำตอบได้ว่า ในแฟ้มข้อมูลเสียงดังกล่าวมีจำนวนผู้สนทนาทั้งหมดกี่คน และในแต่ละช่วงเวลาผู้สนทนาคนใดเป็นผู้พูด ตัวอย่างของตัวแบบที่ได้รับความนิยมในปัจจุบัน (https://github.com/aalto-speech/speaker-diarization) ซึ่งเป็นตัวแบบที่ถูกพัฒนาขึ้นโดยภาษาไพธอน (Python) การแยกเสียงผู้สนทนาโดยวิธีนี้สามารถใช้แยกเสียงผู้สนทนาในการสนทนาที่มีผู้พูดมากกว่า 2 คนได้

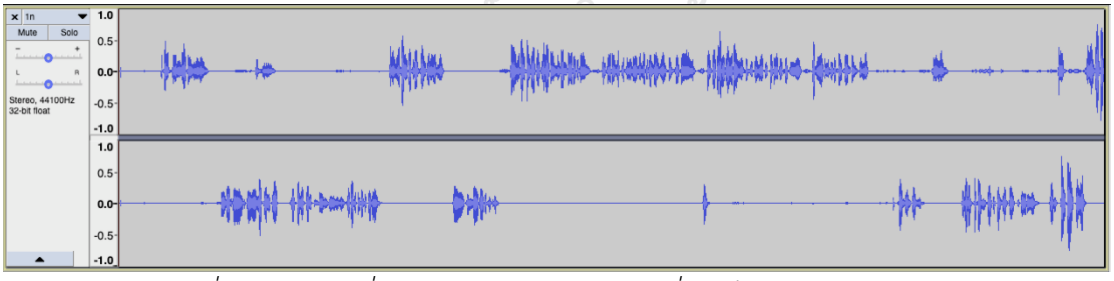
เนื่องจากตัวแบบประเภทนี้เกิดจากการฝึกสอนโดยใช้ชุดข้อมูลฝึกสอนเพื่อให้ตัวแบบเกิดการเรียนรู้ (Machine Learning) และมีความสามารถในการแยกเสียงผู้สนทนาได้ สิ่งที่ต้องคำนึงก่อนการนำมาใช้ คือ ต้องทราบว่าตัวแบบเกิดการจากนำข้อมูลอะไรมาใช้ในการทำการฝึกสอน โดยตัวแบบที่ยกตัวอย่างดังกล่าว ถูกฝึกสอนโดยชุดข้อมูลเสียงในภาษาอังกฤษ ซึ่งหากนำมาใช้กับการแยกเสียงผู้สนทนาที่มีการสนทนาในภาษาอื่นจะทำให้ระดับความถูกต้องของเสียงที่แยกได้นั้นต่ำลง หรือในบางกรณีระดับความถูกต้องอาจน้อยมากจนไม่สามารถนำเสียงที่แยกออกมาไปใช้เพื่อศึกษาในขั้นตอนถัดไปได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การแยกเสียงสนทนาโดยการแยกช่องเสียงสนทนาออกจากกัน (Splitting Channels) ในบางครั้งข้อมูลเสียงสนทนาที่นำมาใช้ในการศึกษา เป็นข้อมูลเสียงประเภทสเตอริโอ (Stereo) ข้อมูลเสียงประเภทนี้จะมีช่องเสียง 2 ช่อง ได้แก่ ช่องเสียงซ้าย และช่องเสียงขวา ข้อมูลเสียงแบบสเตอริโอจะถูกนิยมนำมาใช้เมื่อต้องการให้เสียงของทั้งสองช่องทางมีความแตกต่างกัน หากใช้หูฟังในการฟังเสียงข้อมูลประเภทนี้จะพบว่า หูฟังทั้งสองข้าง จะให้เสียงที่แตกต่างกัน เพื่อให้สามารถฟังเสียงทั้งหมดได้ครบถ้วนจำเป็นจะต้องใส่หูฟังทั้งสองข้าง หากฟังเสียงโดยใช้หูฟังข้างซ้ายเพียงข้างเดียว จะทำให้ได้ยินเฉพาะช่องเสียงซ้ายเท่านั้น และจะเป็นเช่นเดียวกันหากใส่หูฟังข้างขวาเพียงข้างเดียว

ข้อมูลเสียงบันทึกการสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูล ที่นำมาใช้ในการวิจัยครั้งนี้เป็นข้อมูลเสียงประเภทสเตอริโอ เสียงของลูกค้าและเจ้าหน้าที่ถูกบันทึกโดยใช้คนละช่องเสียง

การแยกเสียงโดยวิธีนี้จะได้ผลลัพธ์เป็นแฟ้มข้อมูลเสียงที่มีความยาวเท่ากับข้อมูลนำเข้า แต่จะทำให้บางส่วนของเสียงเกิดช่องว่างซึ่งมาจากการที่ผู้สนทนาอีกฝั่งเป็นผู้พูด มีข้อจำกัด คือ จะใช้วิธีการนี้ได้ในกรณีที่ข้อมูลเสียงถูกบันทึกโดยการแยกช่องเสียง และมีจำนวนผู้สนทนา 1 คนต่อ 1 ช่องเสียงเท่านั้น การแยกช่องเสียงออกจากกันสามารถทำได้โดยการใช้ไลบรารีจากภาษาไพธอน หรือใช้ซอฟต์แวร์สำเร็จรูปได้ ตัวอย่างซอฟต์แวร์สำเร็จรูปที่ใช้สำหรับการแยกช่องสนทนาคือ ออดาซิตี้ (Audacity 2.2.2) เป็นซอฟต์แวร์ที่มีความสามารถหลากหลายในการจัดการกับข้อมูลเสียง โดยออดซิตี้เป็นซอฟแวร์ที่เปิดให้ใช้บริการฟรี



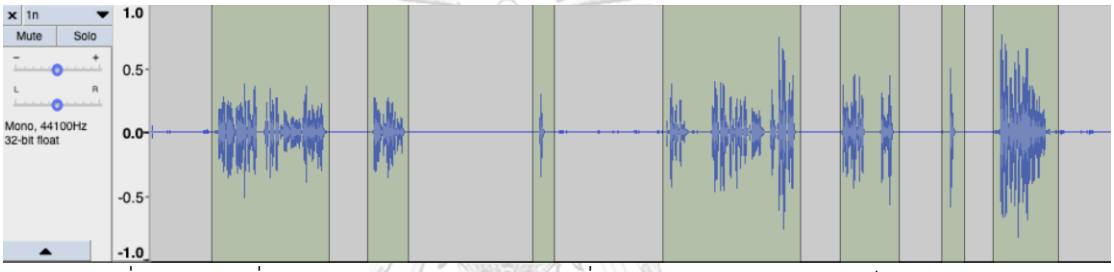
ภาพที่ 4 ตัวอย่างคลื่นเสียงของแฟ้มข้อมูลเสียงที่จัดเก็บในรูปแบบสเตอริโอ

2.5.2 เครื่องมือสำหรับการตัดข้อมูลเสียง

ในภาษาไพธอนมีไลบรารีหลายชุดที่ใช้ในการจัดการกับข้อมูลเสียง หนึ่งในนั้นคือ ไลบรารีไพดับ (pydub) มีความสามารถในการจัดการกับข้อมูลเสียงหลายด้าน ตัวอย่างเช่น ฟังก์ชันที่อยู่ภายในแฟ้มข้อมูล effect.py มีความสามารถในการแปลงข้อมูลเสียงให้เปลี่ยนไปในลักษณะที่ต้องการ เช่น การเพิ่มความเร็วให้กับข้อมูลเสียง การปรับค่าความดังของเสียง เป็นต้น ในงานวิจัยนี้จะให้ความสนใจ

กับแฟ้มข้อมูล silence.py ซึ่งความสามารถในการจัดการกับเสียงเงียบ (Silence) เช่น การเพิ่มเสียงเงียบเป็นระยะเวลาที่ต้องการเข้าไปที่ส่วนต้นและส่วนปลายของเสียง, การตรวจจับระยะเวลารวมของเสียงเงียบที่มีอยู่ภายในข้อมูลเสียง

จากความหมายของรอบการโต้ตอบที่ได้อธิบายไว้ในหัวข้อที่ 2.3.2 การกำหนดขนาดข้อมูลเสียง (Window size) กล่าวว่า รอบการโต้ตอบ เริ่มต้นเมื่อผู้สนทนาเริ่มเปล่งเสียงจนกระทั่งสิ้นสุดการพูดในประโยคนั้น ๆ หรือเมื่อหยุดพูด ซึ่งรอบการโต้ตอบถัดไปอาจเป็นการพูดซ้ำของผู้พูดเดิมหรือผู้พูดอื่นก็ได้ (Wang, 2015) ซึ่งหมายถึงรอบการโต้ตอบจะจบลงเมื่อผู้พูดเงียบเสียง ในแฟ้มข้อมูล silence.py มีฟังก์ชันสปริทออนไซเลนส์ (split_on_silence) ที่สามารถตรวจจับเสียงเงียบและตัดเฉพาะส่วนของเสียงที่เกิดความดังในที่นี้คือเกิดการพูดออกเป็นส่วนๆ ข้อมูลเสียงแต่ละส่วน จะถูกนำออกเป็นแฟ้มข้อมูลรอบการโต้ตอบนั่นเอง



ภาพที่ 5 แสดงคลื่นเสียงของข้อมูลเสียงในส่วนที่มีเสียงพูดและเสียงเงียบเป็นส่วนประกอบ

ฟังก์ชันสปริทออนไซเลนส์ มีพารามิเตอร์ที่ต้องกำหนดดังเพื่อให้เหมาะสมกับข้อมูลเสียงนำเข้าและผลลัพธ์ที่ต้องการ สามารถแสดงรายละเอียดของพารามิเตอร์ได้ ในตารางที่ 4

ตารางที่ 4 แสดงรายละเอียดของพารามิเตอร์ของฟังก์ชันสปริทออนไซเลนส์

ลำดับที่	ชื่อพารามิเตอร์	คำอธิบาย	ค่าตั้งต้น (Default Setting)
1	audio_segment	ข้อมูลเสียงนำเข้า	-
2	min_silence_len	ระยะเวลาของความเงียบที่สั้นที่สุด ที่เมื่อตรวจพบจะใช้จุดนี้เป็นจุดตัดข้อมูลเสียง หน่วยเป็น มิลลิวินาที (millisecond)	1000 มิลลิวินาที
3	silence_thresh	เสียงที่ดังน้อยกว่าเท่าไรที่จะถูกกำหนดให้เป็นความเงียบ หน่วยเป็นดีบีเอฟเอส (dBFS)	-16 ดีบีเอฟเอส
4	keep_silence	ระยะเวลาของความเงียบที่จะถูกเพิ่มให้กับข้อมูลเสียงที่ถูกตัด ทั้งในจุดเริ่มต้นและจุดสิ้นสุดของข้อมูลเสียง การเติมความเงียบเข้าไปเพื่อไม่ให้ข้อมูลเสียงที่ถูกตัดมีลักษณะคล้ายเสียงที่ถูกตัดโดยฉับพลัน หน่วยเป็น มิลลิวินาที (millisecond)	100 มิลลิวินาที

ค่าดีบีเอฟเอส (dBFS : decibel relative to Full Scale) คือ ค่าที่ใช้วัดระดับของสัญญาณเสียง เมื่อเปรียบเทียบกับค่าสัญญาณที่ดังที่สุด ค่าที่สูงที่สุดมีค่าเท่ากับ 0 ดีบีเอฟเอส เพื่อเป็นการอธิบายเพิ่มเติม ผู้วิจัยจะอธิบายโดยการเปรียบเทียบกับหน่วยวัด เดซิเบล (dB) โดยเสียงจะหมายถึงเสียงที่เดินทางในอากาศ ส่วนสัญญาณเสียง จะหมายถึงเสียงที่อยู่ในรูปดิจิตอล

หน่วยเดซิเบลเป็นการบอกความดังของเสียงหรือสัญญาณเสียงที่ได้ยินโดยการเทียบกับความดังของเสียงที่เบาที่สุดที่มนุษย์ได้ยินนั่นคือที่ 0 เดซิเบล ตัวอย่างเช่น เสียงเฮลิคอปเตอร์ที่มนุษย์ได้ยินมีค่าเท่ากับ 100 เดซิเบล ไม่ได้หมายความว่าเฮลิคอปเตอร์มีค่าความดังเท่ากับ 100 เดซิเบล แต่เสียงของเฮลิคอปเตอร์ดังกว่าเสียงที่เบาที่สุดที่มนุษย์สามารถได้ยินอยู่ 100 เดซิเบล

ส่วนค่าดีบีเอฟเอสคือค่าของสัญญาณเสียงเมื่อเทียบกับระดับสัญญาณเสียงที่มากที่สุดในแฟ้มข้อมูลที่อยู่รูปดิจิตอลตนเอง (WAV File) 0 ดีบีเอฟเอส คือ ระดับสัญญาณเสียงที่มากที่สุดของแฟ้มข้อมูลเสียงนั้น (https://www.audiomasterclass.com/newsletter/what-is-the-difference-between-0-db-and-0-dbfs)

2.5.3 โอเพ่นสมายล์ (openSmile)

โอเพ่นสมายล์ (open-Source Media Interpretation by Large feature-space Extraction) เป็นซอฟต์แวร์ที่ใช้สำหรับการสกัดคุณลักษณะทางคลื่นเสียง เพื่อใช้ในงานที่เกี่ยวข้องกับการประมวลผลทางคลื่นเสียง (Signal Processing) และแมชชินเลิร์นนิ่ง (Machine Learning) โดยเฉพาะ การสกัดคุณลักษณะทางคลื่นเสียงของโอเพ่นสมายล์สามารถใช้ทั้งในขณะที่ระบบกำลังทำงาน (Real-time) และในขณะที่ระบบหยุดทำงานแล้ว (Off-line)

การสกัดคุณลักษณะทางคลื่นเสียงของโอเพ่นสมายล์มีความยืดหยุ่น กล่าวคือ คุณลักษณะที่จะถูกสกัด ถูกกำหนดจากแฟ้มข้อมูลข้อกำหนด (Configuration File) ภายในแฟ้มข้อมูลข้อกำหนดจะเก็บคำสั่งและคุณลักษณะทางคลื่นเสียงที่ผู้ใช้งานต้องการ ซึ่งผู้ใช้งานสามารถเลือกที่จะเขียนแฟ้มข้อมูลข้อหนดนี้ขึ้นใหม่ หรือวิธีที่นิยมใช้คือ การใช้แฟ้มข้อมูลข้อกำหนดที่มีอยู่แล้วในโอเพ่นสมายล์ ซึ่งภายในแฟ้มข้อมูลข้อกำหนดแต่ละแฟ้ม จะเป็นการสกัดกลุ่มของคุณลักษณะทางคลื่นเสียงที่อ้างอิงมาจากงานวิชาการที่ได้รับการยอมรับในกลุ่มคุณลักษณะทางคลื่นเสียงนั้น ๆ เพื่อนำไปใช้กับการพัฒนาตัวแบบประเภทต่าง ๆ ตัวอย่างเช่น แฟ้มข้อมูลข้อกำหนด emo_IS09.conf เป็นแฟ้มข้อมูลข้อหนดที่ได้จาก งานประชุมการแข่งขันทางคุณลักษณะทางคลื่นเสียงที่เกี่ยวกับการแสดงอารมณ์ (The INTERSPEECH 2009 Emotion Challenge) รวบรวมคุณลักษณะไว้ 384 คุณลักษณะ สำหรับการนำไปใช้พัฒนาตัวแบบรู้จำอารมณ์ เป็นต้น

คุณลักษณะที่สกัดได้จะอยู่ในรูปของข้อมูลตัวเลข ผู้ใช้สามารถเลือกประเภทของแฟ้มข้อมูลนำออกได้ เช่น แฟ้มข้อมูลซีเอสวี (.CSV) แฟ้มข้อมูลเออาร์เอฟเอฟ (.arff) ซึ่งเป็นชนิดของแฟ้มข้อมูล

ที่สามารถนำไปใช้ในการพัฒนาตัวแบบโดยซอฟต์แวร์เวก้า (WEKA) เป็นต้น (http://opensmile.audeering.com/)

2.5.4 เอสเคเลิร์น (SKLearn, Scikit-learn)

เป็นเครื่องมือที่ใช้สำหรับงานด้านการทำเหมืองข้อมูล (Data-mining) และการวิเคราะห์ข้อมูล (Data Analysis) เป็นที่นิยมและสามารถทำงานได้อย่างมีประสิทธิภาพ ถูกพัฒนาขึ้นโดยภาษาไพธอน มีเครื่องมือครอบคลุมสำหรับการทำเหมืองข้อมูล เช่น การแบ่งข้อมูลโดยการตรวจไขว้ (K-fold Cross Validation) การตรวจสอบไขว้แบบแบ่งชั้นภูมิ (Stratified K-fold Cross Validation) ในด้านของเทคนิคที่ใช้ในการพัฒนาตัวแบบมีให้เลือกใช้งานหลายประเภท ทั้งที่เป็น การจำแนก (Classification) การถดถอย (Regression) และการจัดกลุ่มของข้อมูล (Clustering) เทคนิคที่นิยมใช้ เช่น ซัพพอร์ตเวคเตอร์แมชชิน (Support Vector Machine) แรนดอมฟอร์เรส (Random Forest) เคมีนส์ (K-means) เป็นต้น (https://en.wikipedia.org/wiki/Scikit-learn) (https://scikit-learn.org/)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3 ระเบียบวิธีวิจัย

3.1 ข้อมูลที่ใช้ในการศึกษา

จุดประสงค์หลักข้อหนึ่งของงานวิจัยนี้คือ การศึกษาการรู้จำอารมณ์จากข้อมูลเสียงที่เกิดขึ้นจริงโดยธรรมชาติ การศึกษานี้ได้รับการอนุเคราะห์ข้อมูลบันทึกเสียงสนทนาที่เกิดขึ้นจริงระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูลจากองค์กรผู้ให้บริการเครือข่ายโทรศัพท์เคลื่อนที่รายหนึ่งในประเทศไทย โดยการประสานงานกับเจ้าหน้าที่ที่รับผิดชอบหน่วยงานศูนย์ให้บริการข้อมูลโดยตรง และได้รับอนุญาตจากผู้บริการที่รับผิดชอบหน่วยงาน

ข้อมูลบันทึกเสียงสนทนาที่ได้รับมีจำนวนทั้งหมด 100 สายสนทนา ถูกแบ่งออกเป็นสายสนทนาที่ลูกค้าเกิดปัญหาจากการใช้บริการ (Complain Call) จำนวน 50 สายสนทนา และสายสนทนาปกติ (Normal Call) จำนวน 50 สายสนทนา การระบุว่าสายสนทนาใดเป็นสายสนทนาที่ลูกค้าเกิดปัญหาจากการใช้บริการหรือเป็นสายสนทนาปกติกระทำโดยเจ้าหน้าที่ และสายสนทนาทั้งหมดถูกเลือกอย่างสุ่มโดยเจ้าหน้าที่เช่นกัน

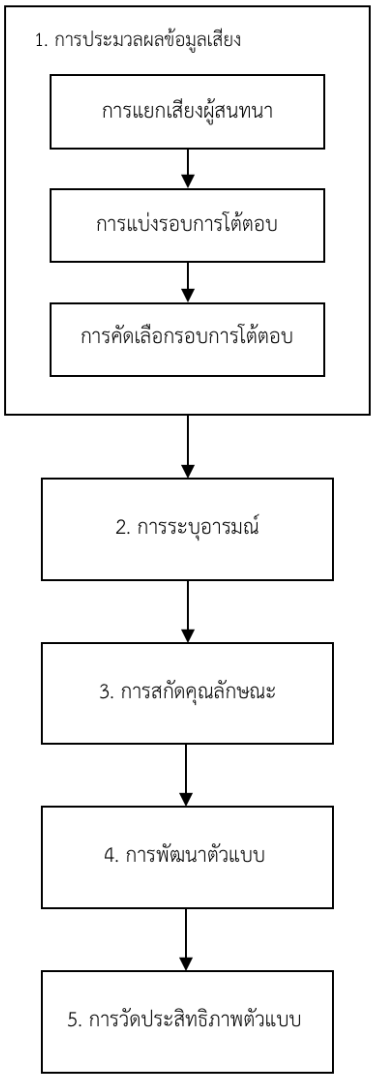
จากการสอบถามข้อมูลเพิ่มเติมพบว่า สายสนทนาประเภทที่ลูกค้าเกิดปัญหาจากการใช้บริการนั้น เป็นสายสนทนาที่ถูกคัดแยกโดยใช้เกณฑ์จากการที่ลูกค้าเกิดปัญหาจากการรับบริการจากองค์กร ซึ่งไม่ได้ใช้การเกิดอารมณ์ความไม่พึงพอใจของลูกค้าเป็นเกณฑ์การแบ่งประเภทสายสนทนาโดยตรง ดังนั้นสายสนทนาที่ลูกค้าเกิดปัญหาจากการใช้บริการอาจมีทั้งกรณีที่ลูกค้าแสดงอารมณ์โกรธและไม่แสดงอารมณ์โกรธปะปนกัน ข้อมูลเบื้องต้นของสายสนทนาที่ได้รับแสดงได้ดังตารางที่ 5

ตารางที่ 5 แสดงจำนวนเพศของลูกค้าและระยะเวลารวมจากข้อมูลเสียงสนทนาที่ได้รับ

ประเภทของสายสนทนา	เพศหญิง	เพศชาย	ระยะเวลารวม (วินาที)	ระยะเวลาเฉลี่ยต่อรอบการโต้ตอบ (วินาที)	ค่าเบี่ยงเบนมาตรฐาน (วินาที)
สายสนทนาที่ลูกค้าเกิดปัญหาจากการใช้บริการ (50 สายสนทนา)	32	18	15,562	311.24	178.45
สายสนทนาปกติ (50 สายสนทนา)	30	20	9,497	189.94	98.46
รวม 100 สายสนทนา	62	38	25,059	250.59	155.80

จากตารางที่ 5 ระยะเวลารวมของสายสนทนาที่ลูกค้าเกิดปัญหาจากการใช้บริการรวมคิดเป็น 15,562 หรือประมาณ 260 นาที และสายสนทนาปกติ 9,497 วินาที หรือประมาณ 158 นาที จะเห็นได้ว่าสายสนทนาจำนวน 50 สายเท่ากัน แต่ความยาวรวมของสายสนทนาที่ลูกค้าเกิดปัญหาจะยาวกว่า

3.2 ขั้นตอนการศึกษา



ภาพที่ 6 แสดงแผนภาพขั้นตอนการพัฒนาตัวแบบรู้จำอารมณ์

ขั้นตอนวิธีที่ผู้วิจัยนำเสนอมีทั้งหมด 5 ขั้นตอน เนื่องจากข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลที่เป็นความลับของบริษัทให้บริการเครือข่ายโทรศัพท์เคลื่อนที่ ผู้วิจัยและทางบริษัทจึงมีความเห็นตรงกันว่า ในขั้นตอน การประมวลข้อมูลเสียงเบื้องต้น การระบุอารมณ์โกรธ และการสกัด

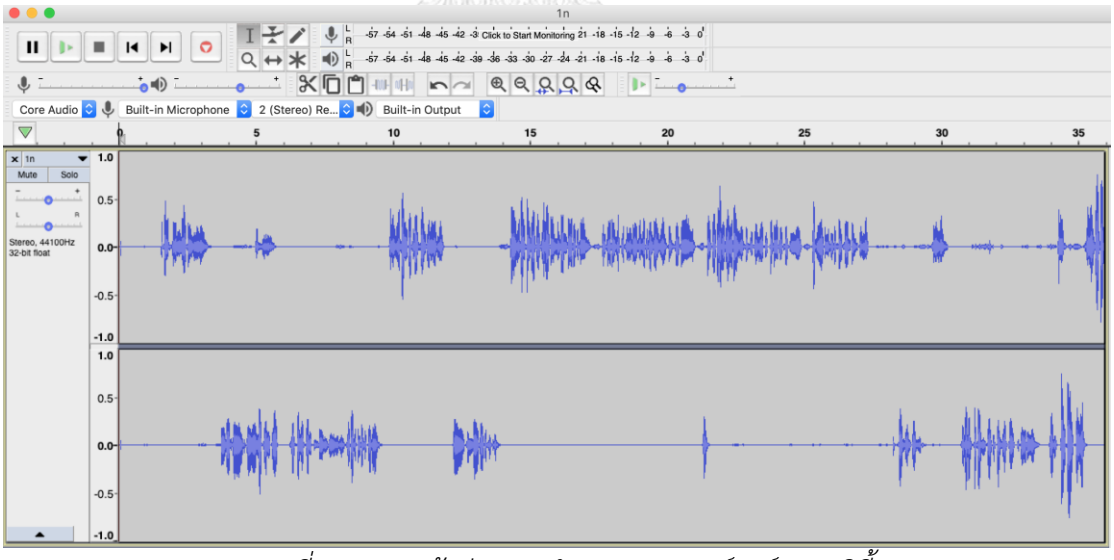
คุณลักษณะ ผู้วิจัยจะเข้าไปทำการเก็บข้อมูลและดำเนินการทั้ง 3 ขั้นตอนที่สำนักงานของบริษัทดังกล่าว ข้อมูลที่นำออกมาใช้ในการศึกษาเป็นเพียงข้อมูลตัวเลข (คุณลักษณะทางคลื่นเสียง) ที่สกัดได้จากเสียงสนทนาเท่านั้น ทั้งนี้เพื่อเป็นการรักษาความลับของข้อมูล หลังจากนั้นจึงจะนำข้อมูลตัวเลขมาใช้ในการพัฒนาตัวแบบและทำการวัดประสิทธิภาพที่ได้ต่อไป โดยในแต่ละขั้นตอนสามารถอธิบายรายละเอียดได้ดังต่อไปนี้

3.3 การประมวลข้อมูลเสียงเบื้องต้น

ในขั้นตอนการประมวลเสียงเบื้องต้น สามารถแบ่งออกเป็นขั้นตอนย่อยได้ 3 ขั้นตอน ดังนี้

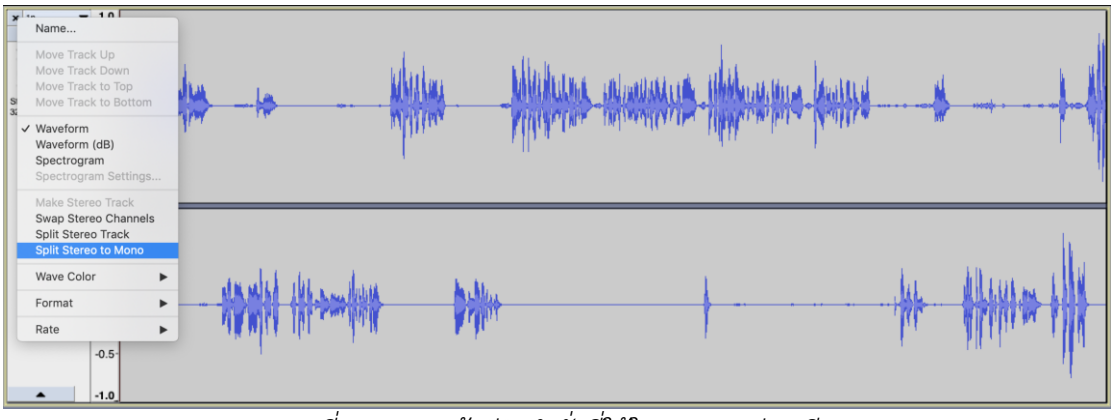
3.3.1 การแยกเสียงผู้สนทนา

เนื่องจากข้อมูลเสียงที่ผู้วิจัยได้รับเป็นข้อมูลเสียงสนทนาในภาษาไทย ถูกบันทึกเสียงแบบสเตอริโอ ที่แบ่งการบันทึกเสียงออกเป็น 2 ช่องเสียง คือ ช่องเสียงทางซ้ายจะบันทึกเสียงลูกค้า และช่องเสียงทางขวาใช้สำหรับบันทึกเสียงเจ้าหน้าของศูนย์ให้บริการข้อมูล ผู้วิจัยเลือกวิธีการแยกเสียงผู้สนทนาด้วยวิธีการใช้ซอฟต์แวร์สำเร็จรูปออดาซิตี้ โดยเริ่มจากการเลือกแฟ้มข้อมูลเสียงที่ต้องการแยกเสียงผู้สนทนา หลังเปิดแฟ้มข้อมูลเสียงแล้ว ซอฟต์แวร์จะแสดงคลื่นเสียงทั้งสองช่องทางแยกจากกันดังภาพที่ 6



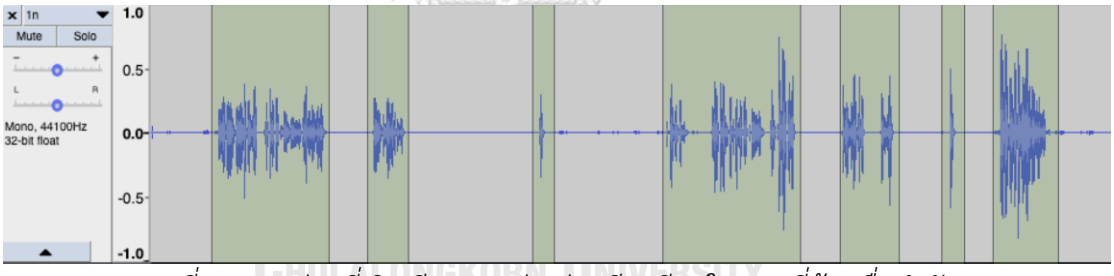
ภาพที่ 7 แสดงหน้าต่างการทำงานของซอฟต์แวร์ออดาซิตี้

หลังจากนั้นให้เลือกคำสั่ง “Split Stereo to Mono” ดังภาพที่ 7



ภาพที่ 8 แสดงหน้าต่างคำสั่งที่ใช้ในการแยกช่องเสียง

หลังจากนั้นจะได้ผลลัพธ์เป็นแฟ้มข้อมูลเสียง 2 แฟ้ม แฟ้มที่ 1 จะเก็บเสียงของเจ้าหน้าที่ และแฟ้มที่ 2 จะเก็บเสียงของลูกค้า ข้อสังเกตคือ แฟ้มข้อมูลเสียงของลูกค้าที่ได้จะมีความยาวเท่ากับแฟ้มข้อมูลเสียงตั้งต้น แต่จะประกอบไปด้วยส่วนที่เป็นเสียงพูดของลูกค้าและส่วนที่เป็นช่องว่างซึ่งจะเป็นเสียงเงียบ โดยช่องว่างที่เกิดขึ้นจะแสดงถึงการเงียบของฝั่งลูกค้า ซึ่งในขณะเดียวกันเป็นช่วงเวลาที่เจ้าหน้าของศูนย์ให้บริการข้อมูลกำลังพูดนั่นเอง ดังแสดงในภาพที่ 8



ภาพที่ 9 แสดงส่วนที่เกิดเสียง และช่องว่างเสียงเงียบในขณะที่ผู้พูดอื่นกำลังพูด

3.3.2 การแบ่งรอบการโต้ตอบ

จากขั้นตอนก่อนหน้า ผลลัพธ์ที่ได้จะเป็นแฟ้มข้อมูลเสียงของลูกค้าที่ทั้งส่วนที่เป็นพูดและเสียงเงียบปะปนกัน ขั้นตอนนี้จะเป็นการตัดแฟ้มข้อมูลเสียงออกเป็นส่วน ๆ เลือกเฉพาะส่วนที่เป็นเสียงพูดของลูกค้าเพื่อไปใช้ในการพัฒนาตัวแบบเท่านั้น

หลักเกณฑ์ที่ใช้ในการตัดแฟ้มข้อมูลเสียงออกเป็นรอบการโต้ตอบ คือ รอบการโต้ตอบจะเริ่มขึ้นเมื่อลูกค้าเริ่มเปล่งเสียงจนกระทั่งพบกับความเงียบหรือการที่ลูกค้าหยุดพูดนานติดต่อกันเป็นเวลา 2000 มิลลิวินาที จะถือว่าเป็นการจบรอบการโต้ตอบนั้น ๆ (min_selence_len) (1 วินาที = 1000 มิลลิวินาที) ในทางปฏิบัติ พบว่า ความเงียบ (Silence) ที่ได้อธิบายไปก่อนหน้าว่าเป็นช่วงเวลาที่ผู้

สนทนาอื่นกำลังพูดนั้น ยังคงมีเสียงชนิดอื่นเกิดขึ้นอยู่แม้ว่าลูกค้าจะไม่ได้เปล่งเสียงก็ตาม เช่น เสียงลม เสียงสัญญาณ หรือเสียงพื้นหลังของสิ่งแวดล้อมที่ลูกค้าอยู่ในขณะนั้น ด้วยเหตุจึงทำให้ต้องมีการกำหนดเพิ่มเติมว่า ค่าของเสียงที่ดังน้อยกว่าเท่าไรจะถูกจัดว่าเป็นเสียงเงียบ ฟังก์ชันที่ใช้ในการตัดรอบการโต้ตอบแสดงดังตารางที่ 6

ตารางที่ 6 แสดงฟังก์ชันและค่าพารามิเตอร์ที่ใช้ในการตัดรอบการโต้ตอบ

แฟ้มข้อมูลที่ใช้ : silence.py			
ฟังก์ชัน : def split_on_silence(audio_segment, min_silence_len, silence_thresh)			
ลำดับที่	**ชื่อพารามิเตอร์**	**ค่าที่กำหนด**	**หน่วย**
1	audio_segment	ข้อมูลเสียงลูกค้า	-
2	min_silence_len	2,000	มิลลิวินาที
3	silence_thresh	ค่าดีบีเอฟเอสของเสียง - 10	ดีบีเอฟเอส
4	keep_silence	* ค่าตั้งต้น = 100	มิลลิวินาที

* หากใช้ค่าตั้งต้นไม่จำเป็นต้องกำหนดตอนเรียกใช้ฟังก์ชัน

ข้อมูลเสียงแต่ละแฟ้มข้อมูลมีค่าความดังของเสียงไม่เท่ากัน ดังนั้น พารามิเตอร์ไซเลนส์เทรสช (silence_thresh) จึงเป็นพารามิเตอร์ที่ไม่สามารถใช้ค่าคงที่ใดค่าหนึ่งได้สำหรับทุกข้อมูลเสียง จะต้องเป็นค่าที่เปลี่ยนไปตามความเหมาะสมของแต่ละข้อมูลเสียงนั้น จากการทดลองพบว่า ค่าที่เหมาะสมของไซเลนส์เทรสช สามารถคำนวณได้จาก การนำค่าดีบีเอฟเอส (Decibel relative to Full Scale : dBFS) ค่าความดัง หรือกล่าวได้ว่าเป็นค่าที่ใช้อธิบายระดับของสัญญาณเสียง เมื่อเปรียบเทียบกับค่าสัญญาณที่มากที่สุด นำไปลบด้วย 10 ซึ่งค่านี้หาได้จากการทดลองปฏิบัติซ้ำในหลายครั้ง โดยค่า 10 เป็นค่าที่ทำให้ผลลัพธ์ในการตัดข้อมูลเสียงมีความถูกต้องมากที่สุด ตัวอย่างการคำนวณ ค่าไซเลนส์เทรสช

ข้อมูลเสียง ก มีค่า ดีบีเอฟเอส = -16.5

ค่าของพารามิเตอร์ไซเลนส์เทรสช = -16.5 - 10

= -26.5 ดีบีเอฟเอส

ข้อมูลเสียงตั้งต้นในขั้นตอนนี้ เป็นข้อมูลเสียงประเภทสายสนทนาที่ลูกค้าเกิดปัญหาและสายสนทนาปกติที่ถูกแยกเฉพาะเสียงของลูกค้าออกมาแล้ว จำนวนทั้งหมด 100 ข้อมูลเสียง หลังจากนำ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

แต่ละข้อมูลเสียงทั้งหมดมาทำการตัดให้อยู่ในรูปของรอบการโต้ตอบ จะได้ผลลัพธ์ดังที่แสดงในตารางที่ 7

ตารางที่ 7 แสดงจำนวนข้อมูลเสียงที่ถูกทำให้อยู่ในรูปของรอบการโต้ตอบ

ประเภทของสายสนทนา	จำนวนรอบการโต้ตอบที่ตัดได้ (รอบการโต้ตอบ)
สายสนทนาที่ลูกค้าเกิดปัญหาจากการใช้บริการ (50 สายสนทนา)	1,114
สายสนทนาปกติ (50 สายสนทนา)	852
รวม 100 สายสนทนา	1,966

3.3.3 การคัดเลือกรอบการโต้ตอบ

เมื่อนำข้อมูลรอบการโต้ตอบทั้งที่มาจากสายสนทนาของลูกค้าที่เกิดปัญหาและสายสนทนาปกติมาทำการตรวจสอบทำให้พบประเด็นที่อาจส่งผลต่อการเลือกรอบการโต้ตอบที่จะนำไปใช้ในการพัฒนาตัวแบบ ดังนี้

ประเด็นแรกพบว่า มีสายสนทนาจำนวน 1 สาย ที่มีความผิดปกติของจำนวนรอบการโต้ตอบที่ได้ โดยปกตินั้นสายสนทนาที่มีความยาวประมาณ 8 นาทีขึ้นไป เมื่อนำมาทำการตัดรอบการโต้ตอบแล้ว จะได้ค่าเฉลี่ยอยู่ที่ 40 รอบการโต้ตอบ แต่ในกรณีของสายสนทนานี้ มีความยาว 8 นาที 15 วินาที รอบการโต้ตอบที่ได้มีจำนวนเพียง 8 รอบการโต้ตอบเท่านั้น ผู้วิจัยจึงทำการตรวจสอบและพบว่า สายสนทนาดังกล่าว เป็นสายสนทนาที่ลูกค้าพูดด้วยภาษาใต้ ซึ่งเป็นภาษาท้องถิ่น มีลักษณะในการพูดด้วยหน่วยเสียงที่เร็วกว่าภาษากลาง รวมไปถึงช่องไฟหรือการเว้นจังหวะในการพูดสั้นมาก ทำให้เครื่องมือตรวจพบเงื่อนไขที่ใช้ในการตัดรอบการโต้ตอบ นั่นคือเสียงเงียบที่มีระยะเวลาอย่างน้อย 2,000 มิลลิวินาที น้อยกว่าในสายสนทนาทั่วไป ผลลัพธ์ที่ได้จึงมีเพียงแค่ 8 รอบการโต้ตอบเท่านั้น ผู้วิจัยจึงตัดสินใจนำสายสนทนานี้ออก ไม่นำมาใช้ในการพัฒนาตัวแบบเนื่องด้วยเหตุผลข้างต้น และถือเป็นการควบคุมลักษณะเฉพาะทางภาษาเพื่อไม่ให้เกิดผลกระทบต่อตัวแบบ

ประเด็นที่สอง คือ รอบการโต้ตอบบางส่วนที่ตัดได้มีความยาวน้อยกว่า 1 วินาที เมื่อทำการตรวจสอบแฟ้มข้อมูลเสียงเหล่านี้ พบว่า ผู้วิจัยไม่สามารถระบุอารมณ์ที่เกิดขึ้นภายในแฟ้มข้อมูลเสียงเหล่านี้ได้เนื่องจาก เสียงมีความยาวที่สั้นมากจนผู้วิจัยไม่สามารถจับใจความและอารมณ์ได้ จึงทำให้ผู้วิจัยจะไม่นำรอบการโต้ตอบที่มีความยาวน้อยกว่า 1 วินาทีมาใช้ในการพัฒนาตัวแบบ โดยจำนวนข้อมูลรอบการโต้ตอบที่นำออกและคงเหลือเพื่อนำไปใช้ในขั้นตอนการระบุอารมณ์ สามารถได้แสดงได้ดังตารางที่ 8

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 8 แสดงจำนวนรอบการโต้ตอบสุทธิ

ประเภทของสายสนทนา	จำนวนรอบการโต้ตอบทั้งหมด	จำนวนรอบการโต้ตอบที่ผิดปกติ	จำนวนรอบการโต้ตอบที่มีความยาวน้อยกว่า 1 วินาที	คงเหลือ
สายสนทนาที่ลูกค้าเกิดปัญหา	1,114	8	275	831
สายสนทนาปกติ	852	0	290	562
สายสนทนาที่ลูกค้าเกิดปัญหาและสายสนทนาปกติ	1,966	8	565	1,393
จำนวนรอบการโต้ตอบที่นำไปใช้ในการพัฒนาตัวแบบ				1,393

3.4 การระบุอารมณ์

ขั้นตอนนี้เป็นการระบุอารมณ์ให้กับแฟ้มข้อมูลเสียงหรือรอบการโต้ตอบของลูกค้าที่ได้จากขั้นตอนก่อนหน้าจำนวน 1,393 รอบการโต้ตอบ โดยใช้วิธีการฟังและระบุว่ารอบการโต้ตอบนั้นลูกค้าแสดงอารมณ์โกรธหรือไม่

เนื่องจากข้อมูลเสียงที่นำมาใช้ในการศึกษา เป็นข้อมูลบันทึกการสนทนาที่เกิดขึ้นจริงจากศูนย์ให้บริการข้อมูลของบริษัทให้บริการเครือข่ายโทรศัพท์เคลื่อนที่ ในบางสายสนทนานั้นลูกค้าได้มีการกล่าวถึงข้อมูลส่วนตัวที่สามารถนำไปใช้ในการยืนยันตัวตนได้ ดังนั้นเพื่อเป็นการรักษาความลับของข้อมูลสำคัญดังกล่าว ผู้วิจัยจึงเป็นเพียงคนเดียวที่ได้รับอนุญาตให้ฟังข้อมูลเสียงสนทนา โดยมีวัตถุประสงค์เพื่อนำไปใช้ในการศึกษาเท่านั้น

การจำแนกว่ารอบการโต้ตอบนั้นมีการแสดงอารมณ์โกรธหรือไม่ จะมาจากการตัดสินใจจากผู้วิจัยเพียงคนเดียว ซึ่งหากเทียบกับงานวิจัยในอดีต ข้อเสียของวิธีการนี้คือระดับของความถูกต้องของการระบุอารมณ์จะน้อยกว่าการใช้ผู้ฟังหลายคน แต่ข้อดีของการมีผู้ระบุอารมณ์เพียงคนเดียว คือ จะทำให้เกณฑ์การตัดสินว่าลูกค้าเกิดอารมณ์โกรธหรือไม่โกรธเป็นไปในทิศทางเดียวกัน ผู้วิจัยจะระบุว่ารอบการโต้ตอบนั้นเป็นรอบการโต้ตอบที่เป็นอารมณ์โกรธเมื่อพบว่ามีอารมณ์โกรธเกิดขึ้นภายในรอบการโต้ตอบนั้น และหากไม่พบอารมณ์โกรธจะระบุให้รอบการโต้ตอบนั้นเป็นอารมณ์ปกติ และเพื่อให้การระบุอารมณ์ในงานวิจัยครั้งนี้มีความถูกต้องมากขึ้นและเพื่อเป็นการตรวจทาน ผู้วิจัยจะทำการระบุอารมณ์ให้กับทุกรอบการโต้ตอบทั้งหมด 2 รอบการฟัง (รอบการฟัง หมายถึง การฟังข้อมูลเสียงทีละแฟ้มข้อมูลจนครบทั้งหมด) รอบการฟังที่ 1 ผู้วิจัยจะฟังและระบุอารมณ์ให้กับแต่ละรอบการโต้ตอบจนครบทั้งหมด หลังจากนั้นจึงทำการฟังและระบุอารมณ์ให้กับทีละแฟ้มข้อมูลเสียงจนครบทั้งหมดอีกหนึ่งรอบ โดยในการระบุอารมณ์ในรอบการฟังที่ 2 ผู้วิจัยจะระบุอารมณ์โดยไม่เห็นผลลัพธ์ของการระบุอารมณ์จากรอบที่หนึ่ง เพื่อป้องกันการนำผลลัพธ์จากรอบการฟังที่ 1 ใช้อ้างอิง หลังจาก

นั้น จะทำการเปรียบเทียบอารมณ์ที่ได้จากทั้ง 2 รอบการฟัง หากพบว่ารอบการโต้ตอบใดที่ผู้วิจัยระบุอารมณ์ไม่สอดคล้องกัน รอบการโต้ตอบนั้นจะถูกฟังอีกครั้งเป็นรอบที่ 3 ซึ่งจะถือว่าผลจากการระบุอารมณ์ที่ได้ในรอบนี้เป็นที่สิ้นสุด ผลลัพธ์ที่ได้จากการระบุอารมณ์ สามารถแสดงได้ดังตารางที่ 9

ตารางที่ 9 แสดงจำนวนรอบการโต้ตอบตามอารมณ์ที่ระบุได้

	อารมณ์โกรธ	**อารมณ์ปกติ**	***ระบุอารมณ์ไม่ได้**	**รวม**
รอบการโต้ตอบจากสายสนทนาที่ลูกค้าเกิดปัญหา	514	273	44	831
รอบการโต้ตอบจากสายสนทนาปกติ	3	555	4	562
รวม	517	828	48	1,393

* การระบุอารมณ์ไม่ได้ เกิดจาก การเกิดเสียงรบกวนจำนวนมาก เช่น เสียงสัญญาณแทรก เสียงลม เสียงสิ่งแวดล้อมพื้นหลัง รบกวนทั้งรอบการโต้ตอบจนทำให้ไม่สามารถระบุอารมณ์ได้

จากข้อมูลตารางที่ 9 ที่ได้จากขั้นตอนนี้จะได้ข้อมูลที่เป็นอารมณ์โกรธทั้งหมด 517 รอบการโต้ตอบ และอารมณ์ปกติจำนวน 828 รอบการโต้ตอบ รวมเป็น 1,345 รอบการโต้ตอบสำหรับนำไปใช้ในการพัฒนาตัวแบบ

3.5 การสกัดคุณลักษณะ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กลุ่มของคุณลักษณะทางคลื่นเสียงที่เลือกใช้ อ้างอิงตามที่กำหนดในการแข่งขันของงานประชุมวิชาการ INTERSPEECH 2010 Paralinguistic Challenge Feature Set ซึ่งได้แนะนำกลุ่มคุณลักษณะทางคลื่นเสียงที่เหมาะสำหรับนำไปใช้ในการศึกษาการรู้จำอารมณ์จำนวน 1,582 คุณลักษณะ ซึ่งได้อธิบายไว้ในบทที่ 2.3.3 คุณลักษณะทางคลื่นเสียง

ผู้วิจัยจะใช้ โอเพ่นสมายล์ 2.3.0 เป็นเครื่องมือที่ใช้ในการสกัดคุณลักษณะทางคลื่นเสียงจำนวน 1,582 คุณลักษณะ โดยใช้แฟ้มข้อมูลข้อกำหนด (Configuration File) ที่มีการกำหนดให้ข้อมูลเสียงนำเข้าถูกสกัดออกเป็นคุณลักษณะทางคลื่นเสียง รายละเอียดเบื้องต้นของแฟ้มข้อมูลข้อกำหนดที่ใช้แสดงดังภาพที่ 9

```
////////////////////////////////////////////////////////////////////////////////
///////// > openSMILE configuration file for emotion features <      //////////////////
/////////   Based on INTERSPEECH 2010 paralinguistics challenge      //////////////////
/////////   Pitch, Loudness, Jitter, MFCC, MFB, LSP and functionals  //////////////////
/////////                                                            //////////////////
/////////   1582 1st level functionals:                              //////////////////
/////////     (34 LLD + 34 delta) * 21 functionals                   //////////////////
/////////     +(4 LLD + 4 delta) * 19 functionals                    //////////////////
/////////     + 1 x Num. pitch onsets (pseudo syllables)             //////////////////
/////////     + 1 x turn duration in seconds                         //////////////////
/////////                                                            //////////////////
///////// (c) audEERING UG (haftungsbeschr&nkt),                     //////////////////
/////////     All rights reserverd.                                  //////////////////
////////////////////////////////////////////////////////////////////////////////
```

ภาพที่ 10 รายละเอียดของแฟ้มข้อมูลข้อกำหนดที่ใช้งานเบื้องต้น

ในขั้นตอนนี้ข้อมูลเสียงนำเข้าที่ใช้ในการสกัดคุณลักษณะ คือ รอบการโต้ตอบ 1,345 รอบการโต้ตอบ เป็นอารมณ์โกรธ 517 รอบการโต้ตอบ และอารมณ์ปกติ 828 รอบการโต้ตอบ ผลลัพธ์ที่ได้จากการสกัดคุณลักษณะจากโอเพ่นสมายล์ จะถูกเก็บไว้ในรูปแฟ้มข้อมูลซีเอสวี csv (Comma-separated Value) ที่ภายในมีข้อมูลจำนวน 1,345 เรกคอร์ด (Row) โดยแต่ละเรกคอร์ดประกอบไปด้วยคุณลักษณะทางคลื่นเสียงี่สกัดได้จำนวน 1,582 คอลัมน์เป็นข้อมูลประเภทตัวเลข และอีก 1 คอลัมน์เป็นตัวแปรที่ใช้บอกว่าเป็นอารมณ์ปกติหรืออารมณ์โกรธอยู่ในรูปข้อมูลตัวอักษร (String)

จะได้ข้อมูล 1,345 เรกคอร์ด แต่ละเรกคอร์ดประกอบไปด้วย 1,583 ตัวแปร สำหรับนำไปใช้ในการพัฒนาตัวแบบรู้จำอารมณ์ในขั้นตอนถัดไป

3.6 การพัฒนาตัวแบบ

เทคนิคที่นำมาใช้ในการทดลองพัฒนาตัวแบบการรู้จำอารมณ์มีทั้งหมด 4 เทคนิค ได้แก่ แรนดอมฟอร์เรส (Random Forest Classification) , ดีซิสชั่นทรี (Decision Tree), ลิเนียร์เอสวีซี (LinearSVC), และโลจิสติกส์รีเกรสชั่น (Logistic Regression)

3.6.1 การตรวจสอบไขว้แบบสุ่ม

ในการตัดสินใจเลือกตัวแบบที่มีประสิทธิภาพมากที่สุด จะใช้วิธีการตรวจสอบไขว้แบบแบ่งชั้นภูมิ โดยข้อมูลเสียงหรือรอบการโต้ตอบที่นำมาใช้ในการพัฒนาตัวแบบรู้จำอารมณ์นั้น ประกอบไปด้วย ข้อมูลเสียงจำนวน 1,345 รอบการโต้ตอบ แบ่งออกเป็น 2 คลาส ได้แก่ คลาสอารมณ์โกรธ จำนวน 517 รอบการโต้ตอบ คิดเป็นจำนวน 38.43 เปอร์เซ็นต์ และคลาสอารมณ์ปกติ จำนวน 828 รอบการโต้ตอบ คิดเป็นจำนวน 61.57 เปอร์เซ็นต์ ข้อมูลทั้งหมดนำไปใช้ในการวัดประสิทธิภาพของแต่ละเทคนิค โดยวิธีการตรวจสอบไขว้แบบแบ่งชั้นภูมิแบบ 10 ส่วน แบบสุ่ม นั่นคือ ข้อมูลจะถูกเลือกเข้าในแต่ละชุดแบบสุ่ม แต่ยังคงมีสัดส่วนของแต่ละคลาสเท่าเดิม สามารถแสดงการแจกแจงของข้อมูลได้ดังตารางที่ 10 และตารางที่ 11

ตารางที่ 10 แสดงสัดส่วนระหว่างข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

ชุดข้อมูล 1-10	จำนวนข้อมูลชุดฝึกสอน	จำนวนข้อมูลชุดทดสอบ	รวม
1	1,210 (89.96%)	135 (10.04%)	1,345
2	1,210 (89.96%)	135 (10.04%)	1,345
3	1,210 (89.96%)	135 (10.04%)	1,345
4	1,210 (89.96%)	135 (10.04%)	1,345
5	1,210 (89.96%)	135 (10.04%)	1,345
6	1,210 (89.96%)	135 (10.04%)	1,345
7	1,210 (89.96%)	135 (10.04%)	1,345
8	1,211 (90.03%)	134 (9.97%)	1,345
9	1,212 (90.11%)	133 (9.89%)	1,345
10	1,212 (90.11%)	133 (9.89%)	1,345

ตารางที่ 10 แสดงจำนวนข้อมูลฝึกสอนและข้อมูลทดสอบในชุดข้อมูลที่ 1-10 โดยตามหลักการของการตรวจสอบไขว้แบบแบ่งชั้นภูมินั้น จะแบ่งข้อมูลเป็นจำนวนชุดตามที่กำหนด ในการวิจัยนี้กำหนดให้มีชุดข้อมูลจำนวน 10 ชุด (K=10) แบ่งเป็นข้อมูลชุดฝึกสอนจำนวน 1,210 รอบการโต้ตอบ คิดเป็น 90 เปอร์เซ็นต์ และ ข้อมูลชุดทดสอบจำนวน 135 รอบการโต้ตอบ คิดเป็น 10 เปอร์เซ็นต์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 11 แสดงจำนวนอารมณ์โกรธและอารมณ์ปกติในแต่ละข้อมูลชุดฝึกสอน

ชุดข้อมูล 1-10	จำนวนข้อมูลชุดฝึกสอน	จำนวนอารมณ์โกรธ	จำนวนอารมณ์ปกติ
1	1,210 (100%)	465 (38.42%)	745 (61.58%)
2	1,210 (100%)	465 (38.42%)	745 (61.58%)
3	1,210 (100%)	465 (38.42%)	745 (61.58%)
4	1,210 (100%)	465 (38.42%)	745 (61.58%)
5	1,210 (100%)	465 (38.42%)	745 (61.58%)
6	1,210 (100%)	465 (38.42%)	745 (61.58%)
7	1,210 (100%)	465 (38.42%)	745 (61.58%)
8	1,211 (100%)	466 (38.48%)	745 (61.52%)
9	1,212 (100%)	466 (38.44%)	746 (61.56%)
10	1,212 (100%)	466 (38.44%)	746 (61.56%)

เมื่อนำข้อมูลชุดฝึกสอนชุดที่ 1-10 มาแจกแจง จำนวนอารมณ์โกรธและอารมณ์ปกติ จะสามารถแสดงได้ดัง ตารางที่ 12 สิ่งสำคัญคือ สัดส่วนของ อารมณ์โกรธ : อารมณ์ปกติ ของข้อมูลชุดฝึกสอนทั้งชุดที่ 1-10 เป็นสัดส่วนที่ใกล้เคียงกับ สัดส่วนของ อารมณ์โกรธ : อารมณ์ปกติ ของข้อมูลทั้งหมด นั่นคือ 38.43% : 61.57%

ตารางที่ 12 แสดงจำนวนอารมณ์โกรธและอารมณ์ปกติในแต่ละข้อมูลชุดฝึกทดสอบ

ชุดข้อมูล 1-10	ข้อมูลชุดทดสอบ	อารมณ์โกรธ	อารมณ์ปกติ
1	135 (100%)	52 (38.51%)	83 (61.49%)
2	135 (100%)	52 (38.51%)	83 (61.49%)
3	135(100%)	52 (38.51%)	83 (61.49%)
4	135 (100%)	52 (38.51%)	83 (61.49%)
5	135 (100%)	52 (38.51%)	83 (61.49%)
6	135 (100%)	52 (38.51%)	83 (61.49%)
7	135 (100%)	52 (38.51%)	83 (61.49%)
8	134 (100%)	51 (38.05%)	84 (61.95%)
9	133 (100%)	51 (38.24%)	84 (61.76%)
10	133 (100%)	51 (38.24%)	84 (61.76%)

CHULALONGKORN UNIVERSITY

ในทำนองเดียวกัน ตารางที่ 12 เป็นการนำข้อมูลชุดทดสอบมาแจกแจง เพื่อแสดงสัดส่วนระหว่างอารมณ์โกรธและอารมณ์ปกติ ที่สัดส่วนใกล้เคียงกับสัดส่วนของทั้ง 2 อารมณ์ของข้อมูลทั้งหมด

จากข้อมูลข้างต้นทั้งหมด จะถูกนำไปใช้ในการพัฒนาตัวแบบ โดยใช้เทคนิคที่ต่างกัน 4 เทคนิค ได้แก่ แรนดอมฟอร์เรส, ดิซิสชั่นทรี, ลิเนียร์เอสวีซี และ โลจิสติกส์รีเกรสชัน

3.6.2 การปรับแต่งคุณลักษณะของเทคนิค (Model Configuration)

เนื่องจากเทคนิคในการพัฒนาตัวแบบแต่ละเทคนิคมีลักษณะและการทำงานเฉพาะตัวที่แตกต่างกัน การศึกษาและทำความเข้าใจพฤติกรรมเหล่านี้จะทำให้สามารถปรับแต่งค่าของคุณลักษณะของเทคนิคได้เหมาะสม และสามารถดึงศักยภาพที่สูงสุดของแต่ละเทคนิคออกมาใช้ในการพัฒนาตัวแบบรู้จำอารมณ์ได้

การพัฒนาตัวแบบรู้จำอารมณ์ในการศึกษานี้ใช้ไลบราลีเอสเคเลิร์นในภาษาไพธอนเป็นเครื่องมือ คุณลักษณะของเทคนิคสามารถแบ่งได้เป็น 2 กลุ่ม คือ

1. กลุ่มของคุณลักษณะที่ปรับแต่งเพื่อให้เหมาะสมกับการพัฒนาตัวแบบตามปัจจัยต่าง ๆ ในการศึกษาครั้งนั้น ปัจจัยดังกล่าว เช่น ขนาดของข้อมูล, ประเภทของตัวแบบ เช่น แบบไบนารีคลาส เป็นต้น

2. กลุ่มของคุณลักษณะที่ปรับแต่งเพื่อให้ตัวแบบมีประสิทธิภาพสูงขึ้น โดยคุณลักษณะของเทคนิคกลุ่มนี้นิยมใช้กราฟแสดงความสัมพันธ์ระหว่างค่าของคุณลักษณะที่เปลี่ยนไปกับค่าประสิทธิภาพของตัวแบบ

เทคนิคทั้ง 4 ที่ใช้ในการพัฒนาตัวแบบ ได้แก่ ดิซิสชั่นทรี, แรนดอมฟอร์เรส, ลิเนียร์เอสวีซี และโลจิสติกส์รีเกรสชั่น เนื่องจากข้อมูลเสียงรอบการโต้ตอบทั้งหมดที่นำมาใช้ในการพัฒนาตัวแบบมีจำนวนทั้งหมด 1,345 รอบการโต้ตอบซึ่งถือเป็นข้อมูลขนาดเล็ก และค่าตั้งต้นของคุณลักษณะของเทคนิคในกลุ่มที่ 1 เอสเคเลิร์น ได้กำหนดค่าตั้งต้นให้เหมาะสมกับข้อมูลที่มีขนาดเล็กไว้แล้ว จากการศึกษาพบว่าเทคนิคที่สามารถปรับแต่งคุณลักษณะเพิ่มเติมได้และประสิทธิภาพหลังจากการปรับแต่งคุณลักษณะสูงขึ้นอย่างชัดเจนมี 2 เทคนิค คือ แรนดอมฟอร์เรสและดิซิสชั่นทรี ในส่วนของเทคนิคลิเนียร์เอสวีซีและโลจิสติกส์รีเกรสชั่นนั้น ค่าคุณลักษณะตั้งต้นที่เอสเคเลิร์นกำหนดมาเหมาะสมกับความต้องการและข้อมูลของตัวแบบแล้ว ดังนั้นจะเหลือเพียงคุณลักษณะของเทคนิคในกลุ่มที่ 2 ที่ต้องทำการปรับแต่ง สามารถแสดงผลลัพธ์ได้ดังนี้

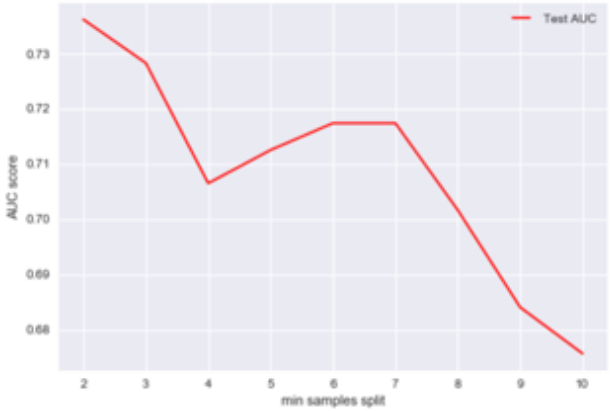
1. เทคนิคดิซิสชั่นทรี คุณลักษะของเทคนิคที่ทำการปรับแต่งสามารถแสดงได้ดังตารางที่ 13

ตารางที่ 13 แสดงคุณลักษณะของเทคนิคดิซิสชั่นทรี

คุณลักษณะ	คำอธิบาย
min_samples_splits	จำนวนข้อมูลขั้นต่ำที่ให้ต้นไม้ split ต่อ
min_samples_leaf	ระบุจำนวนข้อมูลขั้นต่ำใน leaf node ของแต่ละ decision tree หากมีจำนวนข้อมูลต่ำกว่า *min_samples_leaf* ให้หยุด split node นั้น ๆ เพื่อเป็นการลดการเกิด overfitting
min_impurity_decease	node จะทำการ split ต่อ ก็ต่อเมื่อค่าความเจือปนลดลงมากกว่าที่กำหนดใน *min_impurity_decrease*

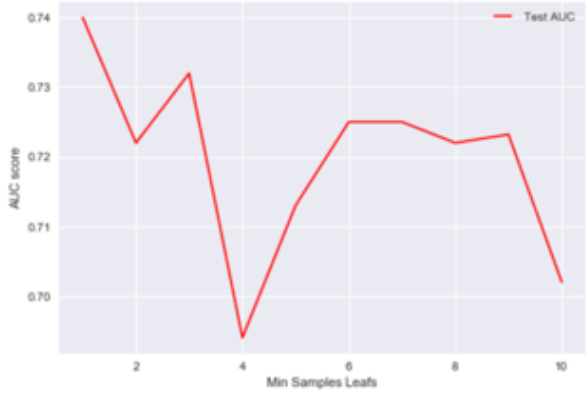
ค่าที่เหมาะสมและทำให้ประสิทธิภาพของตัวแบบสูงสุด ของแต่ละคุณลักษณะสามารถแสดงโดยใช้ค่าเอยูซีได้ดังนี้

1. min_samples_splits



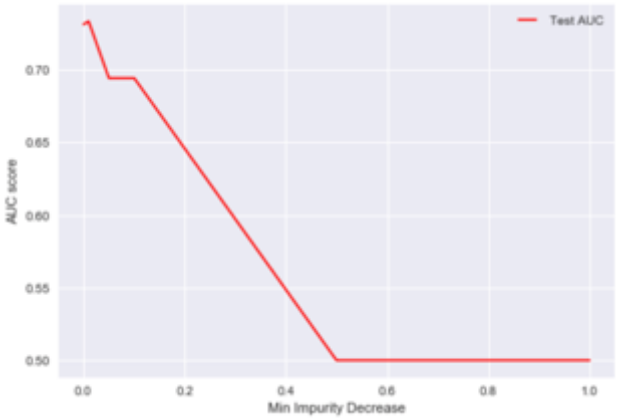
ภาพที่ 11 กราฟแสดงค่า min_samples_splits ของเทคนิคดิซิสชั่นทรี

2. min_samples_leaf



ภาพที่ 12 กราฟแสดงค่า min_samples_leaf ของเทคนิคดิซิสชั่นทรี

3. min_impurity_decrease



ภาพที่ 13 กราฟแสดงค่า min_impurity_decrease ของเทคนิคดิซิสชั่นทรี

จากกราฟที่แสดงในภาพที่ 10 – 12 สามารถสรุปค่าของคุณลักษณะที่ปรับแต่งเพิ่มเติมของเทคนิคดิซิสชั่นทรีได้ดังนี้

- Min_samples_split = 2
- Min_samples_leaf = 1
- Min_impurity_decrease = 0.01

โดยค่าของคุณลักษณะที่ได้นี้จะถูกนำไปใช้ในการพัฒนาตัวแบบโดยเทคนิคดิซิสชั่นทรี และทำการเปรียบเทียบผลลัพธ์ที่ได้ โดยค่าวัดประสิทธิภาพของตัวแบบของเทคนิคดิซิสชั่นทรีแบบปรับแต่งคุณลักษณะควรจะมากกว่าของเทคนิคดิซิสชั่นแบบไม่ปรับแต่งคุณลักษณะ

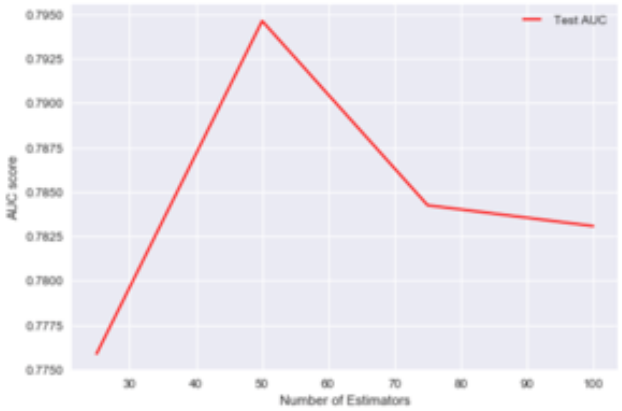
2. แรนดอมฟอร์เรส คุณลักษณะของเทคนิคที่ทำการปรับแต่งสามารถแสดงได้ดัง ตารางที่ 14

ตารางที่ 14 แสดงคุณลักษณะของเทคนิคแรนดอมฟอร์เรส

คุณลักษณะ	คำอธิบาย
n_estimators	จำนวนต้นไม้ในป่า จำนวนต้นไม้มากจะทำให้ความแปรปรวนของตัวแบบลดลง
min_samples_split	จำนวนข้อมูลขั้นต่ำที่ให้ต้นไม้ split ต่อ
min_samples_leaf	ระบุจำนวนข้อมูลขั้นต่ำใน leaf node ของแต่ละ decision tree หากมีจำนวนข้อมูลต่ำกว่า *min_samples_leaf* ให้หยุด split node นั้น ๆ เพื่อลดโอกาสในการเกิด overfitting

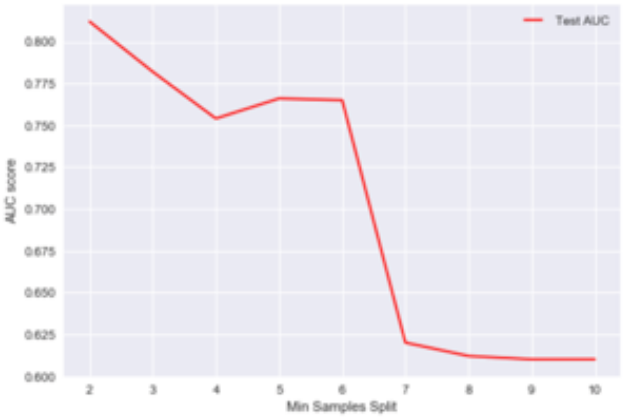
ค่าที่เหมาะสมและทำให้ประสิทธิภาพของตัวแบบสูงสุด ของแต่ละคุณลักษณะสามารถแสดงโดยค่าเอยูซีได้ดังนี้

1. n_estimators



ภาพที่ 14 กราฟแสดงค่า n_estimators ของเทคนิคแรนดอมฟอร์เรส

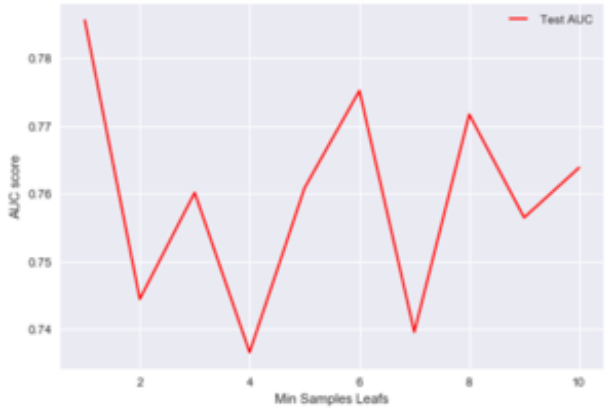
2. min_samples_splits



ภาพที่ 15 กราฟแสดงค่า min_samples_split ของเทคนิคแรนดอมฟอร์เรส

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3. min_samples_leaf



ภาพที่ 16 กราฟแสดงค่า min_samples_leaf ของเทคนิคแรนดอมฟอร์เรส

จากกราฟที่แสดงในภาพที่ 13 – 15 สามารถสรุปค่าของคุณลักษณะที่ปรับแต่งเพิ่มเติมของเทคนิคแรนดอมฟอร์เรสได้ดังนี้

- N_estimators = 50
- Min_samples_split = 2
- Min_samples_leaf = 1

โดยค่าของคุณลักษณะที่ได้นี้จะถูกนำไปใช้ในการพัฒนาตัวแบบโดยเทคนิคแรนดอมฟอร์เรส และทำการเปรียบเทียบผลลัพธ์ที่ได้ ในทำนองเดียวกับเทคนิคดิซิสชั่นทรี ค่าวัดประสิทธิภาพตัวแบบของแรนดอมฟอร์เรสที่ถูกปรับแต่งควรมีค่ามากกว่าของเทคนิคแรนดอมฟอร์เรสที่ไม่ถูกปรับแต่ง

จากการทดลองหาค่าของคุณลักษณะข้างต้นสามารถนำค่าที่ได้มาใช้สำหรับการพัฒนาตัวแบบรู้จำอารมณ์ได้ ดังภาพที่ 16

```
# models with configuration
models = [
    tree.DecisionTreeClassifier(),
    RandomForestClassifier(),
    tree.DecisionTreeClassifier(min_samples_split=2, min_samples_leaf=1
                                , min_impurity_decrease=0.01),
    RandomForestClassifier(n_estimators=50, min_samples_split=2
                           , min_samples_leaf=1),
    LinearSVC(),
    LogisticRegression()
]

# stratified 10 Fold , shuffle data (avoid sequence data in csv file)
skf = StratifiedKFold(n_splits=10,shuffle=True,random_state=1234)
```

ภาพที่ 17 แสดงคุณลักษณะของเทคนิคที่ปรับแต่ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.7 การวัดประสิทธิภาพตัวแบบ (Model Measurements)

ค่าวัดประสิทธิภาพตัวแบบที่ผู้วิจัยจะให้ความสำคัญ คือ ค่าความระลึก (Recall) ของอารมณ์โกรธ และค่าสมดุลความถูกต้อง (Balanced Accuracy) เป็นหลัก เนื่องจากงานวิจัยนี้เป็นการพัฒนาตัวแบบรู้จำอารมณ์ขึ้นเพื่อใช้ในการตรวจจับการเกิดขึ้นของอารมณ์โกรธของลูกค้า ซึ่งค่าความระลึกของอารมณ์โกรธจะเป็นค่าที่แสดงให้เห็นถึงความสามารถในการทำนายอารมณ์โกรธของตัวแบบว่าสามารถทำนายได้ครอบคลุมเพียงใด โดยสิ่งที่ต้องการให้เกิดน้อยที่สุด คือ สถานการณ์ตัวแบบไม่สามารถตรวจจับอารมณ์โกรธที่เกิดขึ้นได้และทำนายว่าเป็นเพียงอารมณ์ปกติ สถานการณ์นี้ถูกจัดเป็นความผิดพลาดร้ายแรง เพราะ ในกรณีที่รอบการโต้ตอบที่เป็นอารมณ์ปกติแต่ตัวแบบทำนายว่าเป็นอารมณ์โกรธจะถือเป็นแจ้งเตือน แต่หากการแจ้งเตือนอารมณ์โกรธนั้นไม่ถูกต้องจะยังไม่เกิดผลเสียใดขึ้น แต่หากตัวแบบไม่สามารถตรวจจับรอบการโต้ตอบที่เป็นอารมณ์โกรธได้ โดยการทำนายว่าเป็น

อารมณ์ปกติ เปรียบเสมือนกับการที่ผู้ให้บริการไม่ทราบว่าลูกค้าอยู่ในอารมณ์โกรธ และไม่ได้ดำเนินการตอบสนองอย่างเหมาะสมอาจส่งผลให้เกิดผลเสียตามมามากกว่า

นอกจากนั้นการทำนายรอบการโต้ตอบที่เป็นอารมณ์ปกติ ยังเป็นสิ่งที่ตัวแบบควรจะทำได้อย่างมีประสิทธิภาพเช่นกัน ค่าความสมดุลความถูกต้องจึงเป็นอีกหนึ่งค่าวัดหลักที่ใช้การพิจารณาและเพื่อให้เข้าใจความสามารถของตัวแบบที่จะถูกพัฒนาเพิ่มเติมในแง่มุมอื่น นอกจากค่าวัดทั้ง 2 ค่าข้างต้น ผู้ศึกษาจะนำค่าวัดอื่นมาร่วมด้วย โดยสรุปค่าวัดประสิทธิภาพตัวแบบที่นำมาใช้พิจารณาความสามารถของตัวแบบสามารถแสดงได้ดังตารางที่ 15

ตารางที่ 15 แสดงค่าวัดประสิทธิภาพตัวแบบที่ใช้

ลำดับ	ค่าวัดประสิทธิภาพตัวแบบ
1	ค่าความถูกต้อง (Accuracy)
2	ค่าความถูกต้องสมดุล (Balanced Accuracy)
3	ค่าความระลึกของรอบการโต้ตอบที่เป็นอารมณ์โกรธ (Recall Angry)
4	ค่าความแม่นยำของรอบการโต้ตอบที่เป็นอารมณ์โกรธ (Precision Angry)
5	ค่าความระลึกของรอบการโต้ตอบที่เป็นอารมณ์ปกติ (Recall Normal)
6	ค่าความแม่นยำของรอบการโต้ตอบที่เป็นอารมณ์ปกติ (Precision Normal)
7	ค่าเอฟวันของรอบการโต้ตอบที่เป็นอารมณ์โกรธ (F1 Score Angry)
8	ค่าความชุกของรอบการโต้ตอบที่เป็นอารมณ์โกรธ (Prevalence)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4 ผลการทดลอง

ในบทนี้กล่าวถึงผลการทดลองซึ่งประกอบด้วย ผลการตรวจสอบไขว้แบบแบ่งชั้นภูมิ ผลการศึกษาการพัฒนาตัวแบบรู้จำอารมณ์ที่นำเสนอ และผลการศึกษาเพิ่มเติมเกี่ยวกับพฤติกรรมของตัวแบบที่พัฒนาขึ้น

4.1 ผลการตรวจสอบไขว้แบบแบ่งชั้นภูมิ

หลังจากนำข้อมูลทั้งหมด 1,345 รอบการโต้ตอบมาแบ่งเป็น 10 ชุดข้อมูล โดยวิธีการตรวจสอบไขว้แบบแบ่งชั้นภูมิมาพัฒนาร่วมกับเทคนิคที่นำมาใช้ในการศึกษาทั้งหมด 4 เทคนิค ได้แก่ ดีซิสชันทรี, แรนดอมฟอร์เรส, ลิเนียร์เอสวีซี และโลจิสติกส์รีเกรสชัน ในหัวข้อนี้จะเป็นการแสดงการเปรียบเทียบค่าวัดประสิทธิภาพระหว่างตัวแบบที่ถูกพัฒนาโดยเทคนิค 2 กลุ่ม กลุ่มแรก คือ กลุ่มที่ถูกพัฒนาด้วยเทคนิคที่ใช้คุณลักษณะตั้งต้น (Default Configuration) และกลุ่มที่ 2 คือ กลุ่มที่ถูกพัฒนาด้วยเทคนิคที่ใช้คุณลักษณะที่มีการปรับแต่ง (Custom Configuration)

4.1.1 ผลของตัวแบบที่ถูกพัฒนาด้วยเทคนิคที่ใช้คุณลักษณะตั้งต้น

ค่าวัดประสิทธิภาพของตัวแบบที่ได้จากการพัฒนาตัวแบบโดยใช้เทคนิคทั้ง 4 เทคนิค ที่ใช้คุณลักษณะตั้งต้น สามารถแสดงได้ดังตารางที่ 16 ถึง ตารางที่ 19

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 16 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคดีซิสชั่นทรี แบบไม่ปรับแต่งคุณลักษณะ

Decision Tree										
K	TP	FP	FN	TN	Accuracy	Recall Normal	Recall Angry	Balanced Accuracy	Precision Angry	F1 Score
					Percentage (%)					
1	31	21	13	70	74.81	84.34	59.62	71.98	70.45	64.58
2	33	19	29	54	64.44	65.06	63.46	64.26	53.23	57.89
3	37	15	21	62	73.33	74.7	71.15	72.93	63.79	67.27
4	38	14	18	65	76.3	78.31	73.08	75.7	67.86	70.37
5	38	14	26	57	70.37	68.67	73.08	70.88	59.38	65.52
6	37	15	20	63	74.07	75.9	71.15	73.53	64.91	67.89
7	37	15	17	66	76.3	79.52	71.15	75.34	68.52	69.81
8	28	23	16	67	70.9	80.72	54.9	67.81	63.64	58.95
9	38	13	16	66	78.2	80.49	74.51	77.5	70.37	72.38
10	40	11	15	67	80.45	81.71	78.43	80.07	72.73	75.47
Average					73.91	76.94	69.05	73	65.49	67.01

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 17 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคแรนดอมฟอร์เรส แบบไม่ปรับแต่งคุณลักษณะ

Random Forrest										
K	TP	FP	FN	TN	Accuracy	Recall Normal	Recall Angry	Balanced Accuracy	Precision Angry	F1 Score
					Percentage (%)					
1	39	13	13	70	80.74	84.34	75	79.67	75	75
2	38	14	19	64	75.56	77.11	73.08	75.09	66.67	69.72
3	38	14	9	74	82.96	89.16	73.08	81.12	80.85	76.77
4	35	17	10	73	80	87.95	67.31	77.63	77.78	72.16
5	41	11	13	70	82.22	84.34	78.85	81.59	75.93	77.36
6	40	12	8	75	85.19	90.36	76.92	83.64	83.33	80
7	41	11	14	69	81.48	83.13	78.85	80.99	74.55	76.64
8	40	11	11	72	83.58	86.75	78.43	82.59	78.43	78.43
9	39	12	18	64	77.44	78.05	76.47	77.26	68.42	72.22
10	39	12	15	67	79.7	81.71	76.47	79.09	72.22	74.29
Average					80.887	84.29	75.446	79.87	75.318	75.259

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 18 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคลิเนียร์เอสวีซี แบบไม่ปรับแต่งคุณลักษณะ

Linear SVC										
K	TP	FP	FN	TN	Accuracy	Recall Normal	Recall Angry	Balanced Accuracy	Precision Angry	F1 Score
					Percentage (%)					
1	22	30	5	78	74.07	93.98	42.31	68.14	81.48	55.7
2	44	8	16	67	82.22	80.72	84.62	82.67	73.33	78.57
3	10	42	0	83	68.89	100	19.23	59.62	100	32.26
4	40	12	9	74	84.44	89.16	76.92	83.04	81.63	79.21
5	48	4	57	26	54.81	31.33	92.31	61.82	45.71	61.15
6	43	9	22	61	77.04	73.49	82.69	78.09	66.15	73.5
7	5	47	0	83	65.19	100	9.62	54.81	100	17.54
8	49	2	43	40	66.42	48.19	96.08	72.14	53.26	68.53
9	30	21	12	70	75.19	85.37	58.82	72.09	71.43	64.52
10	37	14	10	72	81.95	87.8	72.55	80.18	78.72	75.51
Average					73.02	79.00	63.52	71.26	75.171	60.64

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 19 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคโลจิสติกส์แบบไม่ปรับแต่งคุณลักษณะ

Logistic Regression										
K	TP	FP	FN	TN	Accuracy	Recall Normal	Recall Angry	Balanced Accuracy	Precision Angry	F1 Score
					Percentage (%)					
1	35	17	10	73	80.00	87.95	67.31	77.63	77.78	72.16
2	37	15	19	64	74.81	77.11	71.15	74.13	66.07	68.52
3	40	12	11	72	82.96	86.75	76.92	81.84	78.43	77.67
4	39	13	13	70	80.74	84.34	75.00	79.67	75.00	75.00
5	39	13	13	70	80.74	84.34	75.00	79.67	75.00	75.00
6	35	17	15	68	76.30	81.93	67.31	74.62	70.00	68.63
7	43	9	12	71	84.44	85.54	82.69	84.12	78.18	80.37
8	36	15	9	74	82.09	89.16	70.59	79.87	80.00	75.00
9	38	13	18	64	76.69	78.05	74.51	76.28	67.86	71.03
10	41	10	24	58	74.44	70.73	80.39	75.56	63.08	70.69
Average					79.32	82.59	74.08	78.33	73.14	73.40

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อนำเฉพาะค่าเฉลี่ยของค่าวัดประสิทธิของตัวแบบที่ได้จากทั้ง 4 เทคนิค ที่ไม่มีการปรับเปลี่ยนคุณลักษณะมาแสดงผลได้ดังตารางที่ 20

ตารางที่ 20 ตารางแสดงค่าเฉลี่ยของค่าวัดประสิทธิภาพของตัวแบบทั้ง 4 แบบไม่ปรับแต่งคุณลักษณะ

เทคนิค	Accuracy	Balanced Accuracy	Recall Angry / Sensitivity	Precision Angry	Recall Normal / Specificity	F1 Score
Decision Tree	73.91	76.94	69.05	73.00	65.49	67.01
Random Forest	80.88	84.29	75.44	79.87	75.31	75.25
Linear SVC	73.02	79.00	63.52	71.26	75.17	60.64
Logistic Regression	79.32	82.59	74.08	78.33	73.14	73.40

จากตารางที่ 20 พบว่าตัวแบบที่พัฒนาโดยใช้เทคนิคแรนดอมฟอร์เรสแบบไม่ปรับแต่งคุณลักษณะนั้นมีค่าวัดประสิทธิภาพ ตัวแบบสูงสุดสำหรับทุก ๆ ค่าวัด

4.1.2 ผลของตัวแบบที่ถูกพัฒนาด้วยเทคนิคที่ใช้คุณลักษณะที่มีการปรับแต่ง

จากการศึกษาทำความเข้าใจคุณลักษณะและพฤติกรรมของเทคนิคทั้ง 4 เทคนิค พบว่าสำหรับข้อมูลเสียงที่นำมาใช้ในครั้งนี้ มีเพียงเทคนิคดีซิสชั่นทรีและแรนดอมฟอร์เรสเท่านั้น ที่เหมาะสมต่อการปรับเปลี่ยนค่าคุณลักษณะ ในด้านของเทคนิคลีเนียร์เอสวีซีและโลจิสติกส์รีเกรสชันค่าคุณลักษณะตั้งต้นที่ถูกกำหนดมาจากเอสเคเลิร์นนั้นเหมาะสมกับการศึกษาในครั้งนี้แล้ว ค่าวัดประสิทธิภาพของตัวแบบที่ได้จากการใช้เทคนิคดีซิสชันทรีและแรนดอมฟอร์เรสสามารถแสดงได้ดังตารางที่ 21 และตารางที่ 22 ตามลำดับ

ตารางที่ 21 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคดิซิสชั่นทรี แบบปรับแต่งคุณลักษณะ

Decision Tree										
K	TP	FP	FN	TN	Accuracy	Recall Normal	Recall Angry	Balanced Accuracy	Precision Angry	F1 Score
					Percentage (%)					
1	30	22	13	70	74.07	84.34	57.69	71.01	69.77	63.16
2	39	13	29	54	68.89	65.06	75.00	70.03	57.35	65.00
3	34	18	15	68	75.56	81.93	65.38	73.66	69.39	67.33
4	40	12	21	62	75.56	74.70	76.92	75.81	65.57	70.80
5	44	8	25	58	75.56	69.88	84.62	77.25	63.77	72.73
6	41	11	22	61	75.56	73.49	78.85	76.17	65.08	71.30
7	39	13	16	67	78.52	80.72	75.00	77.86	70.91	72.90
8	40	11	15	68	80.60	81.93	78.43	80.18	72.73	75.47
9	41	10	20	62	77.44	75.61	80.39	78.00	67.21	73.21
10	44	7	16	66	82.71	80.49	86.27	83.38	73.33	79.28
Average					76.45	76.81	75.85	76.34	67.51	71.11

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 22 แสดงค่าวัดประสิทธิภาพของตัวแบบจากเทคนิคแรนดอมฟอร์เรส แบบปรับแต่งคุณลักษณะ

Random Forest Classifier										
K	TP	FP	FN	TN	Accuracy	Recall Normal	Recall Angry	Balanced Accuracy	Precision Angry	F1 Score
					Percentage (%)					
1	35	17	7	76	82.22	91.57	67.31	79.44	83.33	74.47
2	36	16	13	70	78.52	84.34	69.23	76.78	73.47	71.29
3	42	10	10	73	85.19	87.95	80.77	84.36	80.77	80.77
4	38	14	7	76	84.44	91.57	73.08	82.32	84.44	78.35
5	39	13	8	75	84.44	90.36	75.00	82.68	82.98	78.79
6	34	18	5	78	82.96	93.98	65.38	79.68	87.18	74.73
7	42	10	12	71	83.7	85.54	80.77	83.16	77.78	79.25
8	38	13	6	77	85.82	92.77	74.51	83.64	86.36	80.00
9	40	11	8	74	85.71	90.24	78.43	84.34	83.33	80.81
10	39	12	5	77	87.22	93.9	76.47	85.19	88.64	82.11
Average					84.02	90.22	74.09	82.16	82.83	78.06

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อนำเฉพาะค่าเฉลี่ยค่าวัดประสิทธิภาพตัวแบบที่ได้จากทั้ง 2 เทคนิค ที่มีการปรับแต่งคุณลักษณะมาแสดงจะได้ผลดังตารางที่ 23

ตารางที่ 23 ตารางแสดงค่าเฉลี่ยของค่าวัดประสิทธิภาพตัวแปรของเทคนิคดิซิสชั่นทรีและแรนดอมฟอร์เรส

เทคนิค	Accuracy	Recall Normal / Specificity	Recall Angry / Sensitivity	Balanced Accuracy	Precision Angry	F1 Score
DecisionTree	76.45	76.81	75.85	76.34	67.51	71.11
RandomForest	84.02	90.22	74.09	82.16	82.83	78.06

ข้อมูลจากตารางที่ 23 แสดงให้เห็นว่าค่าระลึกของอารมณ์โกรธที่ได้จากเทคนิคดิซิสชันทรีมีค่ามากที่สุด แต่มากกว่าแรนดอมฟอร์เรสเพียงเล็กน้อยเท่านั้น ในส่วนของค่าความถูกต้องสมดุลของเทคนิคแรนดอมฟอร์เรสมีค่าสูงกว่าของดิซิสชั่นทรีอย่างชัดเจน เมื่อเปรียบเทียบจากสองค่านี้ จึงสรุปได้ว่า แรนดอมฟอร์เรสแบบปรับแต่งคุณลักษณะสามารถทำงานได้ดีกว่าดิซิสชั่นทรีแบบปรับแต่งคุณลักษณะ

4.1.3 เปรียบเทียบค่าวัดประสิทธิภาพของตัวแบบของแต่ละเทคนิค

ผลลัพธ์ค่าเฉลี่ยค่าวัดประสิทธิภาพของตัวแบบที่ได้จากขั้นตอนก่อนหน้ามีจำนวนทั้งหมด 6 ชุด ได้แก่ 4 ชุดแรก ประกอบด้วยค่าวัดประสิทธิภาพของตัวแบบที่ได้จากเทคนิคที่ใช้คุณลักษณะตั้งต้น 4 เทคนิค ได้แก่ ดีซิสชันทรี, แรนดอมฟอร์เรส, ลิเนียร์เอสวีซี และโลจิสติกส์รีเกรสชั่น และจากเทคนิคที่ใช้คุณลักษณะที่ปรับแต่ง อีก 2 ชุด ได้แก่ ดีชันทรีและแรนดอมฟอร์เรส สามารถนำค่าเฉลี่ยค่าวัดประสิทธิภาพของตัวแบบทั้ง 6 ชุด มาแสดงได้ในตารางที่ 24

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 24 แสดงค่าเฉลี่ยของค่าวัดประสิทธิภาพของตัวแปรที่ได้จากเทคนิคทั้งหมด

เทคนิค	Accuracy	Balanced Accuracy	Recall Angry / Sensitivity	Precision Angry	Recall Normal / Specificity	F1 Score
DecisionTree (Default Configuration)	73.91	76.94	69.05	73.00	65.49	67.01
RandomForest (Default Configuration)	80.88	84.29	75.44	79.87	75.31	75.25
LinearSVC (Default Configuration)	73.02	79.00	63.52	71.26	75.17	60.64
Logistic Regression (Default Configuration)	79.32	82.59	74.08	78.33	73.14	73.40
DecisionTree (Custom Configuration)	76.45	76.81	75.85	76.34	67.51	71.11
RadomForest (Custom Configuration)	84.02	90.22	74.09	82.16	82.83	78.06

จากตารางที่ 24 ทั้งดิซิสชั่นทรีและแรนดอมฟอร์เรสที่มีการปรับแต่งคุณลักษณะจะมีค่าวัดประสิทธิภาพของตัวแบบมากขึ้นจากการใช้คุณลักษณะตั้งต้น และเมื่อนำทุกเทคนิคมาเปรียบเทียบกัน พบว่าการพัฒนาตัวแบบโดยการใช้เทคนิคแรนดอมฟอร์เรสแบบที่มีการปรับแต่งคุณลักษณะ ให้ผลลัพธ์สูงสุดในทุกค่าของค่าวัดประสิทธิภาพของตัวแบบ ยกเว้นเพียงค่าระลึกอารมณ์โกรธของเทคนิคดิซิสชั่นทรีที่มีการปรับแต่งคุณลักษณะที่มากกว่าแรนดอมฟรอ์เรสอยู่เพียงเล็กน้อยเท่านั้น จึงสรุปว่าเทคนิคที่ดีที่สุดในการพัฒนาตัวแบบในการศึกษาครั้งนี้ คือ เทคนิคแรนดอมฟอร์เรสที่มีการปรับแต่งคุณลักษณะ ซึ่งค่าความถูกต้องสมดุลและค่าความระลึกอารมณ์โกรธเป็นค่าวัดที่สำคัญที่สามารถบ่งบอกถึงประสิทธิภาพในการตรวจจับอารมณ์โกรธของตัวแบบ เนื่องจากหากลูกค้าเกิดอารมณ์โกรธแต่ตัวแบบไม่สามารถตรวจจับได้นั้น จะส่งผลเสียมากกว่าการที่ลูกค้าเกิดอารมณ์ปกติแล้วตัวแบบทำนายว่าเป็นอารมณ์โกรธ

4.2 ผลการพัฒนาตัวแบบรู้จำอารมณ์ด้วยวิธีที่นำเสนอ

หลังจากใช้การตรวจสอบไขว์แบบแบ่งชั้นภูมิในการเลือกเทคนิคที่เหมาะสมที่สุดสำหรับการนำมาพัฒนาตัวแบบรู้จำอารมณ์ในการศึกษาครั้งนี้ เทคนิคที่ดีที่สุด คือ แรนดอมฟอร์เรสที่ถูกปรับแต่งคุณลักษณะที่มีค่าความถูกต้องสมดุล (Balanced accuracy) เท่ากับ 83.54 %

เพื่อให้เข้าใจและเห็นรายละเอียดเพิ่มเติมของการพัฒนาตัวแบบโดยใช้เทคนิคแรนดอมฟอร์เรส ผู้วิจัยได้จะใช้วิธีการแบ่งข้อมูล (Split Test) เพื่อทำการฝึกสอนและการทดสอบในสัดส่วน ชุดข้อมูลฝึกสอน 80 : ชุดข้อมูลทดสอบ 20 และจะแบ่งข้อมูลโดยให้อารมณ์โกรธและอารมณ์ปกติอยู่ในสัดส่วนที่เท่ากับ สัดส่วนที่ปรากฏในชุดข้อมูลตั้งต้นดังตารางที่ 25

ตารางที่ 25 แสดงจำนวนอารมณ์โกรธและอารมณ์ปกติของชุดข้อมูลฝึกและข้อมูลทดสอบ

ชุดข้อมูล	**อารมณ์**		
	โกรธ	**ปกติ**	**รวม**
ข้อมูลฝึก	414	662	1,076
ข้อมูลทดสอบ	103	166	269
รวม	517	828	1,345

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หลังจากการทำการเรียนรู้และทดสอบกับตัวแบบ สามารถแสดงผลลัพธ์ได้ดังตารางที่ 26

ตารางที่ 26 ตารางคอนฟิวชันเมทริกซ์ที่ได้จากการพัฒนาตัวแบบ

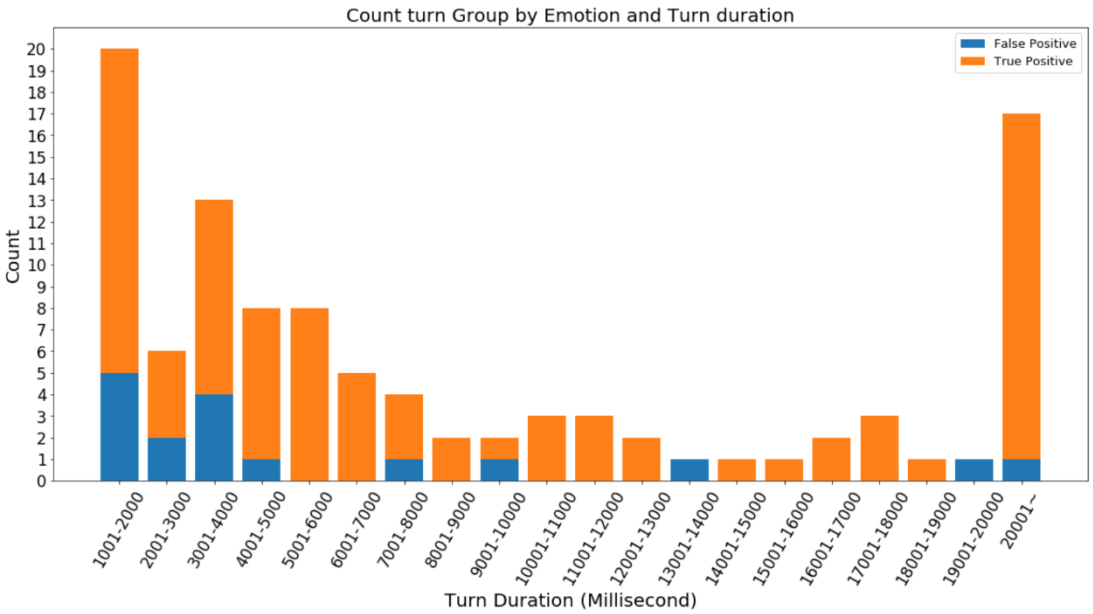
		Predicted		
		Angry	**Normal**	**Total**
Actual	**Angry**	True Positive 86	False Negative 17	103
	Normal	False Positive 20	True Negative 146	166
	Total	106	163	269

Accuracy (TP+TN) / Total Turn	86.24%	Balanced Accuracy (Sensitivity / Specificity) / 2	85.53%
Recall (Angry) , Sensitivity (TP / Actual Angry)	83.49%	Recall (Normal) , Specificity (TN / Actual Negative)	87.95%
Precision (Angry) (TP / Predicted Angry)	81.13%	Precision (Normal) (TN / Predicted Normal)	89.57%
Prevalence (Angry) (Actual Angry / Total Turn)	38%	F1 Score (Angry) 2 * ((Recall * Precision / (Recall + Precision))	82.29%

จากผลลัพธ์ของตัวแบบที่เกิดจากการเรียนรู้ด้วยข้อมูล 80 เปอร์เซ็นต์ และทำการทดสอบด้วยข้อมูลชุดทดสอบอีก 20 % พบว่า จำนวนรอบการโต้ตอบที่เป็นอารมณ์โกรธจำนวน 103 รอบการโต้ตอบ ตัวแบบสามารถทำนายได้ถูกต้อง จำนวน 86 รอบการโต้ตอบ และทำนายผิดจำนวน 17 รอบการโต้ตอบ ส่งผลให้ค่าระลึกของอารมณ์โกรธมีค่าเท่ากับ 83.49 เปอร์เซ็นต์ เมื่อนำมาเทียบกับอัตราความชุกของอารมณ์โกรธ (Prevalence) ที่มีค่าเท่ากับ 38 เปอร์เซ็นต์ พบว่าตัวแบบสามารถทำนายอารมณ์โกรธได้ คิดเป็น 2.2 เท่าของอัตราความชุกของอารมณ์โกรธ

4.3 การศึกษาเพิ่มเติม

หลังจากการพัฒนาตัวแบบรู้จำอารมณ์ในขั้นตอนที่มีการแบ่งข้อมูล 80 เปอร์เซ็นต์ เพื่อเป็นชุดข้อมูลฝึกสอน และอีก 20 เปอร์เซ็นต์ เป็นข้อมูลชุดทดสอบ ผู้วิจัยได้ศึกษาความสัมพันธ์ระหว่างจำนวนของรอบการโต้ตอบ ความยาวของรอบการโต้ตอบ และผลลัพธ์ของการทำนายอารมณ์ของตัวแบบ เมื่อนำข้อมูลรอบการโต้ตอบเฉพาะที่เป็นอารมณ์โกรธมาทำเป็นกราฟแท่งแสดงความสัมพันธ์แกนตั้งแทนจำนวนของรอบการโต้ตอบ และแกนนอนแทนความยาวรอบการโต้ตอบที่แบ่งเป็นกลุ่ม พร้อมทั้งจัดกลุ่มกราฟแท่งแทนรอบการโต้ตอบที่เป็นอารมณ์โกรธที่ตัวแบบทำนายได้ถูกต้อง และทำนายไม่ถูกต้องสำหรับกลุ่มความยาวนั้น ๆ ดังภาพที่ 18

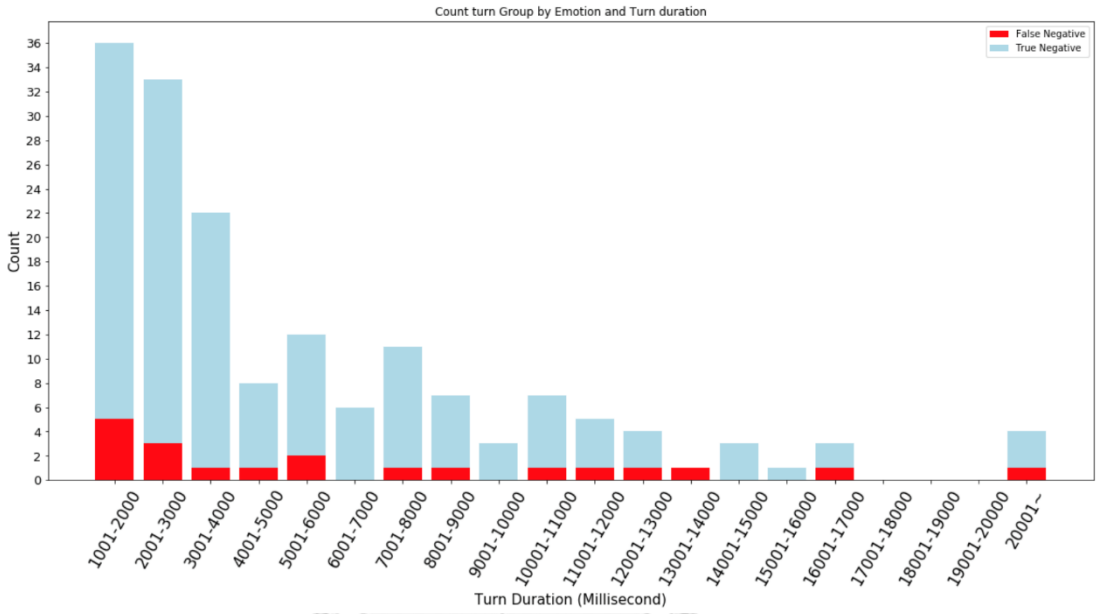


ภาพที่ 18 กราฟแท่งแสดงจำนวนรอบการโต้ตอบของอารมณ์โกรธของแต่ละช่วงความยาวของรอบการโต้ตอบ

จากภาพที่ 18 แสดงให้เห็นว่า ในรอบการโต้ตอบที่มีความยาวตั้งแต่ 4001 มิลลิวินาทีขึ้นอัตรา ที่ตัวแบบสามารถทำนายรอบการโต้ตอบที่เป็นอารมณ์โกรธได้ถูกต้องเริ่มมีความคงที่มากขึ้น แต่อย่างไรก็ตามอัตราการทำนายผิดยังคงเกิดขึ้นในรอบการโต้ตอบที่อยู่ในช่วง 7001-8000, 9001-10,000, 13,001-14,000, 19,001-20,000 และในรอบการโต้ตอบที่มีความยาวมากกว่า 20,000 มิลลิวินาที โดยมีจำนวนที่ทำนายผิดเพียงช่วงละ 1 รอบการโต้ตอบ จากข้อมูลจึงสามารถสรุปได้ว่าตัวแบบรู้จำอารมณ์ที่พัฒนาขึ้นจะมีความเหมาะสมกับการทำนายอารมณ์โกรธในรอบการโต้ตอบที่มีความยาวตั้งแต่ 4,001 มิลลิวินาที หรือนั่นคือ 4 วินาทีขึ้นไป และสำหรับรอบการโต้ตอบที่เป็น

อารมณ์โกรธที่อยู่ในช่วงความยาว 1,001 - 4,000 มิลลิวินาที ตัวแบบก็สามารถทำนายอารมณ์โกรธได้ถูกต้องเช่นเดียวกัน ระดับความถูกต้องจะน้อยกว่า

ในด้านของการทำนายอารมณ์ของรอบการโต้ตอบที่เป็นอารมณ์ปกติ สามารถนำข้อมูลที่ได้จากการทำนายของตัวแบบที่พัฒนาขึ้น มาแสดงเป็นความสัมพันธ์โดยใช้กราฟแท่ง ในลักษณะเช่นเดียวกับการทำนายรอบการโต้ตอบที่เป็นอารมณ์โกรธ ได้ดังภาพที่ 20



ภาพที่ 19 กราฟแท่งแสดงจำนวนรอบการโต้ตอบของอารมณ์ปกติของแต่ละช่วงความยาวของรอบการโต้ตอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลสรุปที่ได้จากข้อมูลการทำนายอารมณ์จากตัวแบบที่พัฒนาขึ้นทั้งด้านของรอบการโต้ตอบที่เป็นอารมณ์โกรธ และด้านของรอบการโต้ตอบที่เป็นอารมณ์ปกติ ตัวแบบจะทำนายผลลัพธ์ได้ดีและมีความคงที่มากขึ้นเมื่อรอบการโต้ตอบมีความยาวตั้งแต่ 4,001 มิลลิวินาที หรือ 4 วินาที ขึ้นไป

บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ

ในบทนี้จะกล่าวถึง ผลการศึกษา โดยการสรุปตามวัตถุประสงค์ของงานวิจัย 3 ข้อ ได้แก่ (1) สรุปการศึกษาเกี่ยวกับวิธีการประมวลผลเบื้องต้นของข้อมูลเสียง (2) เพื่อศึกษาและพัฒนาตัวแบบสำหรับการรู้จำอารมณ์โกรธจากบันทึกข้อมูลเสียงสนทนาที่เกิดขึ้นจริงในศูนย์ให้บริการข้อมูล และ (3) ศึกษาเกี่ยวกับปัจจัยที่ส่งผลต่อประสิทธิภาพของตัวแบบ หลังจากข้อสรุปที่ได้จากการศึกษา จะเป็นในส่วนของข้อจำกัดของงานวิจัย และข้อแนะนำสำหรับการศึกษาต่อเนื่อง รวมถึงข้อแนะนำสำหรับการนำตัวแบบที่ได้ไปประยุกต์ใช้

5.1 สรุปการศึกษาการประมวลผลเบื้องต้นของข้อมูลเสียงประเภทการสนทนา

เนื่องจากข้อมูลเสียงที่นำมาใช้ในการศึกษาครั้งเป็นข้อมูลเสียงที่เกิดขึ้นจริง และเป็นข้อมูลที่เป็นการสนทนา ในขั้นตอนนี้มีความท้าทายในด้านของการเตรียมความพร้อมของข้อมูลเสียงเพื่อให้เหมาะสมต่อการนำไปใช้ในขั้นตอนถัดไป จากการทบทวนวรรณกรรมร่วมกับการศึกษาข้อมูลเสียงที่ได้รับจากองค์กร สามารถสรุปขั้นตอนในการประมวลผลเบื้องต้นได้ ดังนี้

ปัจจัยทางด้านการเลือกวิเคราะห์เสียงของผู้สนทนาแต่ละคนเข้ามาเกี่ยวข้อง ในงานวิจัยนี้เลือกเฉพาะเสียงของลูกค้าเพื่อใช้ในการพัฒนาตัวแบบรู้จำอารมณ์ จึงจำเป็นจะต้องทำการแยกเฉพาะเสียงของลูกค้าเพื่อนำมาใช้ ข้อมูลที่ได้รับจากองค์กรถูกบันทึกเป็นประเภทสเตอริโอ ซึ่งช่องเสียงของลูกค้าและเจ้าหน้าที่ถูกบันทึกแยกจากกัน ผู้วิจัยจึงทำการแยกเฉพาะช่องเสียงของลูกค้ามาใช้

จากนั้นจึงนำเสียงของลูกค้ามาตัดออกให้อยู่ในรูปของรอบการโต้ตอบ โดยรอบการโต้ตอบหนึ่งมีจุดเริ่มต้นจากจุดที่ลูกค้าเริ่มเปล่งเสียงจนกระทั่งลูกค้าเงียบเสียงเป็นเวลา 2 วินาทีติดต่อกัน ซึ่งถือเป็นจุดตัดที่ใช้ในการแยกรอบการโต้ตอบออกจากกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หลังจากที่ทำให้ข้อมูลเสียงอยู่ในรูปการโต้ตอบแล้วพบว่า มีรอบการโต้ตอบจำนวนหนึ่งมีขนาดสั้นกว่า 1 วินาที จากการฟังพบว่าผู้วิจัยไม่สามารถระบุอารมณ์ให้รอบการโต้ตอบประเภทนี้ได้ จึงเลือกที่จะนำรอบการโต้ตอบที่มีความยาวน้อยกว่า 1 วินาทีออกไป จำนวน 565 รอบการโต้ตอบ ซึ่งจากการสุ่มตรวจยังพบว่าจำนวนรอบการโต้ตอบในกลุ่มนี้ส่วนใหญ่มาจากสายสนทนาปกติ บางรอบการโต้ตอบเป็นการพูดเพื่อตอบรับ หรือแสดงความเข้าใจที่ลูกค้ามีต่อข้อมูลที่เจ้าหน้าที่กำลังแจ้ง เช่น คำว่า “อ๋อ”, “อือ”, “อืม” เป็นต้น

5.2 การศึกษาตัวแบบรู้จำทางอารมณ์จากข้อมูลเสียงที่เกิดขึ้นจริง

การศึกษาและพัฒนาตัวแบบรู้จำอารมณ์จากข้อมูลเสียงที่เกิดขึ้นจริง แบ่งเป็น 4 ขั้นตอนหลัก ได้แก่ (1) การประมวลผลเบื้องต้นเกี่ยวกับข้อมูลเสียง (2) การระบุอารมณ์ (3) การสกัด

คุณลักษณะทางคลื่นเสียง และ(4) การพัฒนาตัวแบบรู้จำอารมณ์ ผู้ศึกษาจะกล่าวถึงผลสรุปของแต่ละขั้นตอนดังนี้

การประมวลผลข้อมูลเสียงผู้วิจัยได้ดำเนินการดังที่ได้สรุปไว้ในหัวข้อก่อนหน้า (5.1) ในส่วนของการระบุอารมณ์ให้แก่รอบการโต้ตอบ เนื่องจากข้อจำกัดด้านความเป็นส่วนตัวของข้อมูลเสียงสนทนา ผู้วิจัยจึงสามารถฟังเสียงและทำการระบุอารมณ์ได้เพียงคนเดียว เกณฑ์ที่ผู้วิจัยที่ใช้ในการตัดสินว่ารอบการโต้ตอบใดเป็นอารมณ์โกรธ คือ เมื่อฟังแล้วรู้สึกว่าเกิดอารมณ์โกรธ ทั้งที่เป็นการโกรธแบบแสดงออกทางน้ำเสียงและการโกรธแบบที่ผู้พูดควบคุมน้ำเสียง เพื่อให้ตัวแบบเรียนรู้อารมณ์โกรธทั้งสองประเภท ผลลัพธ์ที่ได้จากการระบุอารมณ์ แบ่งเป็น รอบการโต้ตอบที่เป็นอารมณ์โกรธจำนวน 517 รอบ และรอบการโต้ตอบที่เป็นอารมณ์ปกติอีกจำนวน 828 รอบ และเนื่องจากข้อมูลที่นำมาใช้เป็นข้อมูลที่เกิดขึ้นจริง ดังนั้น การบันทึกเสียงข้อมูลของฝั่งลูกค้าไม่สามารถทำการควบคุมปัจจัยในด้านของเสียงรบกวนได้ เช่น เสียงของสิ่งแวดล้อมพื้นหลังของลูกค้า ในบางสายสนทนาลูกค้าอยู่ในบริเวณที่เป็นพื้นที่ชุมชน ห้างสรรพสินค้า หรือท้องถนน จะมีเสียงรบกวนแทรกจำนวนมาก สำหรับในการศึกษานี้หากรอบการโต้ตอบใดที่มีเสียงรบกวนจำนวนมากและทำให้ผู้วิจัยไม่สามารถเข้าใจบริบทที่ลูกค้ากำลังพูดรวมถึงอารมณ์ที่สื่อออกมาได้ รอบการโต้ตอบดังกล่าวจะถูกดึงออกจากชุดข้อมูลที่ใช้ ซึ่งรอบการโต้ตอบที่มีเสียงรบกวนที่ถูกนำออกมีจำนวน 48 รอบ สุดท้ายจะได้รอบการโต้ตอบที่สามารถนำไปใช้การพัฒนาตัวแบบจำนวนรวมเป็นทั้งสิ้น 1,345 รอบ

การสกัดคุณลักษณะทางคลื่นเสียง คุณลักษณะที่ใช้มีจำนวนทั้งสิ้น 1,582 คุณลักษณะซึ่งเป็นกลุ่มของคุณลักษณะทางคลื่นเสียงที่เหมาะสมต่อการนำไปใช้ในการพัฒนาตัวแบบรู้จำอารมณ์ เมื่อนำรอบการโต้ตอบที่ผ่านการประมวลผลเบื้องต้นและระบุอารมณ์แล้วจำนวน 1,345 รอบการโต้ตอบ มาทำการสกัดคุณลักษณะทางคลื่นเสียง จะได้ผลลัพธ์ที่อยู่ในรูปของข้อมูลคุณลักษณะทางคลื่นเสียงจำนวน 1,345 เรกคอร์ด ซึ่ง 1 เรกคอร์ดแทน ข้อมูลเสียง 1 รอบการโต้ตอบ

ในขั้นตอนสุดท้ายการพัฒนาตัวแบบรู้จำทางอารมณ์ ผู้วิจัยได้ใช้การตรวจสอบไขว้แบบแบ่งชั้นภูมิ เป็นเครื่องมือทดสอบเทคนิคที่มีความเหมาะสมในการนำไปใช้พัฒนาตัวแบบรู้จำอารมณ์มากที่สุด โดยการแบ่งชุดข้อมูลออกเป็นจำนวน 10 ชุด โดยแต่ละชุดข้อมูลจะมีสัดส่วนรอบการโต้ตอบที่พบอารมณ์โกรธและอารมณ์ปกติเท่ากับสัดส่วนของรอบการโต้ตอบที่พบอารมณ์โกรธและอารมณ์ปกติของข้อมูลทั้งหมด ซึ่งวิธีการแบ่งข้อมูลประเภทนี้เหมาะกับข้อมูลที่มีสัดส่วนของแต่ละคลาสไม่สมดุล เช่นดังข้อมูลที่ใช้ในการศึกษาครั้งนี้ รอบการโต้ตอบที่พบอารมณ์โกรธ 38.43 เปอร์เซ็นต์ และรอบการโต้ตอบที่ไม่พบอารมณ์โกรธหรือสรุปว่าเป็นรอบการโต้ตอบที่เป็นอารมณ์ปกติ 63.57 เปอร์เซ็นต์ และผลลัพธ์ที่ได้ คือ เทคนิคแรนดอมฟอร์เรสมีค่าระลึกของอารมณ์โกรธ (Recall) และค่าความถูกต้องสมดุล (Balanced Accuracy) ดีที่สุดในบรรดาเทคนิคที่นำมาทดสอบทั้งหมด

หลังจากนั้นจึงนำเทคนิคแรนดอมฟอร์เรสมาทำการตรวจสอบ เพื่อศึกษาพฤติกรรมของตัวแบบเพิ่มเติม โดยการแบ่งข้อมูลแบบแบ่งชั้นภูมิ 80 เปอร์เซ็นต์ใช้ในการฝึกสอนและ 20 เปอร์เซ็นต์สำหรับการทดสอบ ข้อมูลที่ใช้การทดสอบ มีรอบการโต้ตอบที่เป็นอารมณ์โกรธจำนวน 103 รอบการโต้ตอบ ทำนายได้ถูกต้อง 86 รอบการโต้ตอบ และทำนายผิด 17 รอบการโต้ตอบ ทำให้ค่าระลึกของอารมณ์โกรธมีค่าเท่ากับ 83.49 เปอร์เซ็นต์ ซึ่งเมื่อเทียบกับความชุกของรอบการโต้ตอบที่เป็นอารมณ์โกรธ (Prevalence) ที่มีค่าเท่ากับ 38 เปอร์เซ็นต์ จะเห็นว่า ตัวแบบที่พัฒนาขึ้นในงานวิจัยนี้สามารถตรวจจับอารมณ์โกรธได้มากกว่าค่าความชุกดังกล่าว 2.2 เท่า และมีค่าความถูกต้องสมดุลเท่ากับ 85.53 เปอร์เซ็นต์ ในขณะที่ค่าระลึกของอารมณ์ปกติมีค่าเท่ากับ 87.95 เปอร์เซ็นต์ ซึ่งค่าระลึกของอารมณ์โกรธและอารมณ์ปกติอยู่ในเกณฑ์ที่ค่อนข้างสูง

โดยข้อสรุปสำหรับวัตถุประสงค์ในข้อนี้ คือ ผู้วิจัยได้ศึกษา ทฤษฎีที่เกี่ยวข้องการพัฒนาตัวแบบรู้จำอารมณ์จากคุณลักษณะทางคลื่นเสียง จากข้อมูลเสียงสนทนาในสภาพแวดล้อมจริง และสามารถพัฒนาตัวแบบรู้จำอารมณ์ที่มีค่าวัดประสิทธิภาพของตัวแบบอยู่ในเกณฑ์ดี

5.3 ปัจจัยที่ส่งผลต่อประสิทธิภาพตัวแบบ

นอกจากการเตรียมความพร้อมของข้อมูลเสียง และการสกัดคุณลักษณะทางคลื่นเสียงแล้วนั้น การปรับแต่งคุณลักษณะของเทคนิค (Custom Configuration) หรือการปรับแต่งค่าพารามิเตอร์ให้กับแต่ละเทคนิคในการพัฒนาตัวแบบ เป็นอีกหนึ่งประเด็นที่ส่งผลต่อประสิทธิภาพของตัวแบบ เนื่องจากเทคนิคแต่ละเทคนิคมีการทำงานที่แตกต่างกัน ดังนั้น ก่อนจะปรับแต่งคุณลักษณะของเทคนิคจำเป็นต้องมีความเข้าใจการทำงานของเทคนิคนั้น ๆ ค่าที่กำหนดให้แก่แต่ละเทคนิค จะต้องทำให้สามารถดึงศักยภาพของเทคนิคออกมาให้ได้มากที่สุด โดยจากการปรับแต่งคุณลักษณะเทคนิคให้กับดีซิสชั่นทรีและแรนดอมฟอร์เรสแล้ว ค่าวัดประสิทธิภาพของตัวแบบของทั้งสองเทคนิคมีค่าสูงขึ้นอย่างชัดเจน

5.4 ข้อจำกัดของงานวิจัย

1. การศึกษาในครั้งนี้ ผู้ที่ได้รับอนุญาติให้สามารถฟังข้อมูลบันทึกการสนทนาที่ได้รับ มีเพียงผู้วิจัยเพียงคนเดียวเท่านั้น ในบางข้อมูลเสียงที่มีความไม่ชัดเจนในการแสดงออกทางอารมณ์ของผู้พูด ผู้วิจัยอาจระบุอารมณ์ได้ไม่ถูกต้องสำหรับรอบการโต้ตอบนั้น

2. ด้านบริบทของข้อมูลเสียงที่ใช้ในการพัฒนาตัวแบบ เนื่องจากข้อมูลแต่ละบริบทที่แตกต่างกันจะไม่เหมาะในการนำตัวแบบไปใช้โดยตรง แต่สามารถนำขั้นตอนวิธีในการพัฒนาตัวแบบที่นำเสนอไปปรับให้เหมาะสมกับบริบทของข้อมูลที่ต้องการได้

5.5 ข้อเสนอแนะสำหรับการนำงานวิจัยไปศึกษาต่อ

ในหัวข้อนี้จะกล่าวถึง ประเด็นต่าง ๆ ที่ผู้วิจัยพบในขณะทำการศึกษาและพัฒนาตัวแบบรู้จำอารมณ์จากข้อมูลเสียงสนทนาที่เกิดขึ้นจริง ว่ามีประเด็นใดบ้างที่เป็นประเด็นที่จะต้องคำนึงถึงหากมีผู้วิจัยท่านอื่นต้องการที่จะศึกษาในด้านนี้เพิ่มเติม

1. ในประเด็นการประมวลผลเบื้องต้นของข้อมูลเสียง เป็นอีกประเด็นหนึ่งที่มีความสำคัญและส่งผลต่อพฤติกรรมการนำตัวแบบไปใช้อย่างมาก เนื่องจากการนำข้อมูลเสียงที่เกิดขึ้นจริงมาใช้ เสียงรบกวนที่เกิดจากเสียงของสิ่งแวดล้อมย่อมเกิดขึ้นได้ ในการศึกษาครั้งนี้ไม่ได้มีการตัดเสียงรบกวนออก (Noise Cancelling) และเมื่อนำเสียงเหล่านี้ไปใช้ในขั้นตอนการสกัดคุณลักษณะทางคลื่นเสียง คุณลักษณะทางคลื่นเสียงที่ได้ย่อมรวมทั้งเสียงพูดของลูกค้าและเสียงรบกวนที่เกิดขึ้น ดังนั้นเพื่อให้คุณลักษณะทางคลื่นเสียงที่ได้มีความถูกต้องมากยิ่งขึ้น จึงควรเพิ่มขั้นตอนในการตัดเสียงรบกวนออก ในงานวิจัยนี้เนื่องจากข้อมูลเสียงถูกตัดให้อยู่ในขนาดของรอบการโต้ตอบ รอบการโต้ตอบใดที่มีเสียงรบกวนมากและกลบเสียงของลูกค้าผู้วิจัยได้นำออกและไม่นำมาใช้ในการพัฒนาตัวแบบ

2. ประเด็นของจำนวนผู้ระบุอารมณ์ การนำข้อมูลเสียงที่เกิดขึ้นจริงมาใช้ในการพัฒนาตัวแบบการรู้จำอารมณ์นั้น ประเด็นด้านความชัดเจนทางอารมณ์เป็นสิ่งที่ต้องคำนึง ตามที่ได้กล่าวไปในบทของการทบทวนวรรณกรรม และจากการที่ผู้วิจัยได้ฟังข้อมูลเสียงสนทนาที่ได้รับ เห็นตรงกันว่า ในข้อมูลเสียงที่เกิดขึ้นจริงอย่างเป็นธรรมชาติ ในบางครั้งผู้พูดไม่ได้เจาะจงที่จะแสดงอารมณ์ใดอารมณ์หนึ่งอย่างชัดเจนดังเช่นการนำข้อมูลที่เกิดจากการแสดงมาใช้ ขั้นตอนในการระบุอารมณ์ให้กับข้อมูลเสียง จึงเป็นขั้นตอนที่จะต้องมีวิธีการที่จะทำให้การระบุอารมณ์มีความถูกต้องมากที่สุด หากเป็นไปได้ควรใช้ผู้ระบุอารมณ์มากกว่าหนึ่งคน เพื่อเป็นการสอบทานว่าอารมณ์ที่ถูกระบุโดยฟังแต่ละคนนั้น ได้ผลลัพธ์เป็นอารมณ์เดียวกันหรือไม่ และหากผลลัพธ์แตกต่างกัน จะมีวิธีในการจัดการกับข้อมูลเสียงนั้นอย่างไร อีกหนึ่งเหตุผลที่ทำให้ขั้นตอนนี้มีความสำคัญมากคือ อารมณ์ที่ระบุให้กับแต่ละข้อมูลเสียงเปรียบเสมือนการบอกว่า คุณลักษณะทางเสียงที่จะถูกสกัดได้ในขั้นตอนการสกัดคุณลักษณะนั้น เป็นตัวแทนของอารมณ์ที่ผู้ฟังระบุลงไป หากผู้วิจัยไม่สามารถระบุอารมณ์ได้ถูกต้อง ย่อมส่งผลทางลบต่อประสิทธิภาพในการทำนายอารมณ์ของตัวแบบ

3. จากข้อมูลที่ได้รับพบว่ามีสายสนทนาจำนวนหนึ่งสายที่เป็นเสียงพูดในภาษาใต้ ผู้วิจัยนำสายสนทนานี้ออกเพื่อเป็นการควบคุมในด้านภาษาที่ใช้ ด้วยเหตุผลที่ว่าแต่ละภาษาพูดมีลักษณะเฉพาะตัวแตกต่างกัน ในการพัฒนาตัวแบบรู้จำอารมณ์ในการศึกษาครั้งนี้เกิดจากการใช้คุณลักษณะทางคลื่นเสียง ไม่ได้นำคุณลักษณะทางภาษาศาสตร์มาใช้ สิ่งที่น่าสนใจคือ จะเป็นไปได้หรือไม่ที่ตัวแบบจะสามารถพัฒนาให้รองรับการใช้ข้อมูลเสียงพูดในภาษาที่มากกว่าหนึ่งภาษา เพื่อให้ตัวแบบสามารถรู้จำอารมณ์ในภาษาพูดที่หลากหลายได้

4. ในการศึกษาครั้งนี้เป็นเพียงการจำแนกระหว่างอารมณ์โกรธและอารมณ์ปกติ เท่านั้น ซึ่ง 2 อารมณ์นี้เป็นอารมณ์ที่พบในบริบทของข้อมูลเสียงการสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูล ทั้งนี้หากข้อมูลเสียงที่นำมาใช้อยู่ในบริบทอื่นจำเป็นจะต้องพิจารณาถึงจำนวนอารมณ์อื่นเพิ่มเติม เช่น อารมณ์มีความสุข หรืออารมณ์เศร้า เป็นต้น ให้เหมาะสมกับอารมณ์ที่ต้องการตรวจจับ เพื่อให้ตัวแบบเกิดประโยชน์ต่อข้อมูลเสียงในบริบทนั้นมากที่สุด

5.6 ข้อจำกัดและข้อเสนอแนะสำหรับการนำตัวแบบประยุกต์ใช้

ข้อจำกัดในการนำตัวแบบไปใช้งาน คือ จะสามารถนำตัวแบบไปใช้ได้เมื่อข้อมูลเสียงอยู่ในบริบทของเสียงสนทนาระหว่างลูกค้าและเจ้าหน้าที่ของศูนย์ให้บริการข้อมูล เนื่องจากบริบทของข้อมูลเสียงที่แตกต่างกันย่อมมีปัจจัยเฉพาะตัวของบริบทนั้น ๆ

ตัวแบบสามารถนำไปประยุกต์ใช้งานได้ทั้งระหว่างเกิดการสนทนาระหว่างเจ้าหน้าที่และลูกค้า และหลังจากการสนทนาสิ้นสุดลงแล้ว โดยหากนำตัวแบบไปประยุกต์ใช้ในระหว่างที่การสนทนากำลังเกิดขึ้นสามารถนำไปใช้เพื่อเป็นการแจ้งเตือนแก่เจ้าหน้าที่ในกรณีที่ลูกค้ากำลังเกิดอารมณ์โกรธ เพื่อให้เจ้าหน้าที่เริ่มตระหนักและระมัดระวังในการสื่อสารเพิ่มมากขึ้น ซึ่งการนำไปประยุกต์ใช้งานลักษณะนี้ จำเป็นจะต้องทำการตรวจสอบเรื่องเวลาที่ใช้ในการประมวลผล เนื่องจากตัวแบบจะต้องสามารถแจ้งเตือนได้ทันทีภายในการสนทนานั้น ๆ การนำไปประยุกต์ในลักษณะที่สองคือนำตัวแบบไปใช้สำหรับการแยกสายสนทนาที่สิ้นสุดลงแล้ว ในการแยกสายสนทนาที่ลูกค้ามีอารมณ์โกรธเกิดขึ้นในระหว่างการรับบริการศูนย์ให้บริการข้อมูลทางโทรศัพท์ ซึ่งมีจำนวนสายสนทนาเป็นจำนวนมากในแต่ละวัน เพื่อลดภาระงานในการใช้คนจำแนกสายสนทนาของลูกค้าที่เกิดอารมณ์โกรธ เพื่อนำไปศึกษาและหาวิธีการแก้ไขปัญหาการเกิดอารมณ์โกรธของลูกค้า

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตัวแบบรู้จำอารมณ์โกรธของลูกค้าที่ถูกพัฒนาขึ้นจากงานวิจัยนี้ ใช้ข้อมูลเสียงเท่ากับขนาดรอบการโต้ตอบของผู้พูดเป็นข้อมูลนำเข้าสำหรับการทำนายอารมณ์ อารมณ์ที่ตัวแบบสามารถทำนาย คือ อารมณ์โกรธและอารมณ์ปกติ ซึ่งเป็นสองอารมณ์หลักที่ลูกค้าจะแสดงออกในขณะรับบริการจากศูนย์ให้บริการข้อมูล

จากบทที่ 4 หัวข้อที่ 4.3 การศึกษาเพิ่มเติม ผู้วิจัยได้นำผลลัพธ์การทำนายอารมณ์ของตัวแบบมานำเสนอในรูปแบบของความสัมพันธ์ระหว่าง ความยาวของรอบการโต้ตอบ และผลลัพธ์ของการทำนาย พบว่ารอบการโต้ตอบที่มีความยาวมากกว่า 4 วินาทีเป็นต้นไป มีอัตราการทำนายที่ถูกต้องและคงที่มากกว่ารอบการโต้ตอบที่สั้นกว่า 4 วินาที ดังนั้นเพื่อให้การนำตัวแบบจากงานวิจัยนี้ไปประยุกต์ใช้ได้อย่างมีประสิทธิภาพมากขึ้น การทำนายอารมณ์ของรอบการโต้ตอบที่มีความยาวมากกว่า 4 วินาที อาจได้ผลการทำนายที่น่าเชื่อถือมากกว่าทั้งในรอบการโต้ตอบที่เป็นอารมณ์โกรธและอารมณ์ปกติ สำหรับรอบการโต้ตอบที่มีความยาวน้อยกว่า 4 วินาที ผลการทำนายที่ได้มีความถูกต้อง แต่ความน่าเชื่อถือทางผลลัพธ์ที่ได้อาจน้อยกว่ารอบการโต้ตอบที่ความยาวมากกว่า 4 วินาที ทั้งในด้านของรอบการโต้ตอบที่เป็นอารมณ์โกรธและอารมณ์ปกติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

Alonso, J. B., Cabrera, J., Medina, M., & Travieso, C. M. (2015). New approach in quantification of emotional intensity from the speech signal: emotional temperature. *Expert Systems with Applications, 42*(24), 9554-9564. doi:https://doi.org/10.1016/j.eswa.2015.07.062

Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition, 44*(3), 572-587. doi:https://doi.org/10.1016/j.patcog.2010.09.020

Burkhardt, F., Paeschke, A., Rolfes, M. A., Sendlmeier, W. F., & Weiss, B. (2005). *A database of German emotional speech.* Paper presented at the INTERSPEECH.

Burkhardt, F., Polzehl, T., Stegmann, J., Metze, F., & Huber, R. (2009, 19-24 April 2009). *Detecting real life anger.* Paper presented at the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing.

C.K, Y., Hariharan, M., Ngadiran, R., Adom, A. H., Yaacob, S., Berkai, C., & Polat, K. (2017). A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. *Expert Systems with Applications, 69*, 149-158. doi:https://doi.org/10.1016/j.eswa.2016.10.035

Cao, H., Verma, R., & Nenkova, A. (2015). Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech. *Computer Speech & Language, 29*(1), 186-202. doi:https://doi.org/10.1016/j.csl.2014.01.003

Chakraborty, R., Pandharipande, M., & Kopparapu, S. K. (2016a, 2016//). *MINING CALL CENTER CONVERSATIONS EXHIBITING SIMILAR AFFECTIVE STATES.* Paper presented at the Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters.

Chakraborty, R., Pandharipande, M., & Kopparapu, S. K. (2016b, 4-8 Dec. 2016). *Spontaneous speech emotion recognition using prior knowledge.* Paper presented at the 2016 23rd International Conference on Pattern Recognition (ICPR).

Dallimore, K., Sparks, B., & Butcher, K. (2007). *The Influence of Angry Customer*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Outbursts on Service Provider's Facial Displays and Affective States (Vol. 10).

Dave, N. (2013). *Feature extraction methods LPC, PLP and MFCC in speech recognition* (Vol. Volume 1).

Devillers, L., & Vidrascu, L. (2006). *Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs*.

Hansen, J. H. L., & Bou-Ghazale, S. E. (1997). *Getting started with SUSAS: a speech under simulated and actual stress database.* Paper presented at the EUROSPEECH.

Haq, S., Jackson, P. J. B., & Edge, J. D. (2008). *Audio-visual feature selection and reduction for emotion classification.* Paper presented at the AVSP.

Koolagudi, S. G., & Rao, K. S. (2010, 17-19 Dec. 2010). *Real life emotion classification using VOP and pitch based spectral features.* Paper presented at the 2010 Annual IEEE India Conference (INDICON).

Litman, D., Forbes, K., & Silliman, S. (2003). *Towards emotion prediction in spoken tutoring dialogues*. Paper presented at the Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers - Volume 2, Edmonton, Canada.

Mariooryad, S., & Busso, C. (2015). Correcting Time-Continuous Emotional Labels by Modeling the Reaction Lag of Evaluators. *IEEE Transactions on Affective Computing, 6*(2), 97-108. doi:10.1109/TAFFC.2014.2334294

Murray, I., & L. Arnott, J. (1993). *Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion* (Vol. 93).

Paidi, G., Kadiri, S., & Yegnanarayana, B. (2016). *Analysis of Emotional Speech—A Review*.

Pappas, D., Androutsopoulos, I., & Papageorgiou, H. (2015, 19-21 Oct. 2015). *Anger detection in call center dialogues.* Paper presented at the 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom).

Petrushin, V. (2000). *Emotion in Speech: Recognition and Application to Call Centers*.

Pittam, J., & Scherer, K. R. (1993). Vocal expression and communication of emotion. In *Handbook of emotions.* (pp. 185-197). New York, NY, US: Guilford Press.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Pokorny, F. B., Graf, F., Pernkopf, F., & Schuller, B. W. (2015, 21-24 Sept. 2015). *Detection of negative emotions in speech signals using bags-of-audio-words.* Paper presented at the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII).

Schuller, B. W. (2011). *Voice and Speech Analysis in Search of States and Traits.* Paper presented at the Computer Analysis of Human Behavior.

Schuller, B. W., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C. A., & Narayanan, S. (2010). *The INTERSPEECH 2010 paralinguistic challenge.* Paper presented at the INTERSPEECH.

Tang, H., Chu, S. M., Hasegawa-Johnson, M., & Huang, T. S. (2009, 28 June-3 July 2009). *Emotion recognition from speech VIA boosted Gaussian mixture models.* Paper presented at the 2009 IEEE International Conference on Multimedia and Expo.

Tischer, P. R. (1994). Self-consciousness and emotional expression. *Dissertations from ProQuest*.

Truong, K. P., van Leeuwen, D. A., & de Jong, F. M. G. (2012). Speech-based recognition of self-reported and observed emotion in a dimensional space. *Speech Communication, 54*(9), 1049-1063. doi:https://doi.org/10.1016/j.specom.2012.04.006

Vidrascu, L., & Devillers, L. (2007). *Five emotion classes detection in real-world call center data: The use of various types of paralinguistic features*.

Wang, K.-C. (2015). Time-frequency feature representation using multi-resolution texture analysis and acoustic activity detector for real-life speech emotion recognition. *Sensors (Basel, Switzerland), 15*(1), 1458-1478. doi:10.3390/s150101458

Williams, C., & Stevens, K. (1972). *Emotions and Speech: Some Acoustical Correlates* (Vol. 52).

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions %J IEEE Trans. Pattern Anal. Mach. Intell. *31*(1), 39-58. doi:10.1109/tpami.2008.52

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล	วิฎากร แซ่หว่อง
วัน เดือน ปี เกิด	15 กันยายน 2536
วุฒิการศึกษา	สำเร็จการศึกษาในระดับปริญญาตรี หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์ และเข้ารับการศึกษาต่อในระดับมหาบัณฑิต หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศธุรกิจ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ประจำปีการศึกษา 2559

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University