

การตรวจจับการตัดต่อเสียงโดยใช้การเรียนรู้ของเครื่อง

EDITED AUDIO DETECTION USING MACHINE LEARNING



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2558

KMITL-2015-SC-M-002-047

การตรวจจับการตัดต่อเสียงโดยใช้การเรียนรู้ของเครื่อง

EDITED AUDIO DETECTION USING MACHINE LEARNING



ทักษ์ดนัย สุวรรณ
TAKDANAI SUWAN

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2557

KMITL-2015-SC-M-002-047

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EDITED AUDIO DETECTION USING MACHINE LEARNING



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE
DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE
FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2014

KMITL-2015-SC-M-002-047

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2014

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ “การตรวจจับการติดต่อเสียงโดยใช้การเรียนรู้แบบรวมกลุ่ม”
“EDITED AUDIO DETECTION USING MACHINE LEARNING”
ชื่อนักศึกษา นายทักษ์ดนัย สุวรรณ
รหัสประจำตัว 53650804
ปริญญา วิทยาศาสตรมหาบัณฑิต (สาขาวิชาวิทยาการคอมพิวเตอร์)
ภาควิชา วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ ดร.สายชล ใจเย็น
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์

คณะกรรมการสอบวิทยานิพนธ์	ลายมือชื่อ
ผศ.ดร.ศรัณย์ อินทโกสุม ประธานกรรมการ ผศ.ดร.อนันตพร หารราชคุณาตย์ อาจารย์บัณฑิตประจำ (ในสาขาวิชาที่เกี่ยวข้อง) ดร.ธนส์นี เพียรตระกูล ผู้ทรงคุณวุฒิจากภายนอกสถาบันฯ ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.สายชล ใจเย็น อาจารย์ที่ปรึกษาวิทยานิพนธ์	 อนันตพร หารราชคุณาตย์  ธนส์นี เพียรตระกูล  P.W.  ส.ช.

วัน/ เดือน/ ปี ที่สอบ วันที่ 30 มิถุนายน พ.ศ.2558
สถานที่สอบ ห้อง 306 อาคารปฏิบัติการใหม่ ชั้น 3

คณะวิทยาศาสตร์รับรองแล้ว

(รองศาสตราจารย์ ดร.ศุภณี ธนะบริพัทธ์)

คณบดีคณะวิทยาศาสตร์

วันที่ ๒๕ เดือน ๖ ปี พ.ศ. ๒๕๕๘

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การตรวจจับการตัดต่อเสียงโดยใช้การเรียนรู้ของเครื่อง
ชื่อนักศึกษา	ทักษ์ดนัย สุวรรณ
รหัสประจำตัว	53650804
ปริญญา	วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์
ภาควิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2557
อาจารย์ที่ปรึกษาวิทยานิพนธ์	อาจารย์ ดร.สายชล ใจเย็น
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	อาจารย์ ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์

บทคัดย่อ

การตรวจสอบการตัดต่อข้อมูลเสียงเป็นปัญหาท้าทายของนักนิติวิทยาศาสตร์ ซึ่งสามารถช่วยในการพิสูจน์และแยกข้อมูลเสียงว่าข้อมูลเสียงไหนเป็นข้อมูลเสียงต้นฉบับ และข้อมูลเสียงไหนเป็นข้อมูลเสียงที่มีการตัดต่อ วิทยานิพนธ์นี้นำเสนอเทคนิคการตรวจสอบข้อมูลเสียงโดยใช้การเรียนรู้ของเครื่อง โดยทำการจำแนกประเภทข้อมูลเป็นสองกลุ่มคือ ข้อมูลเสียงต้นฉบับและข้อมูลเสียงที่มีการตัดต่อ วิทยานิพนธ์นี้ทำการทดสอบเปรียบเทียบประสิทธิภาพความถูกต้อง โดยใช้การเรียนรู้ของเครื่องในรูปแบบการจำแนกประเภทข้อมูลมาทำการเปรียบเทียบกัน ได้แก่ การเรียนรู้แบบรวมกลุ่ม ซัพพอร์ตเวกเตอร์แมชชีน ตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน โครงข่ายประสาทเทียมแบบโพรบาบิลิสติก ผลการทดลองแสดงให้เห็นว่า การเรียนรู้แบบรวมกลุ่มมีความเหมาะสมสำหรับใช้ในการตรวจสอบข้อมูลเสียงเพราะ การเรียนรู้แบบรวมกลุ่มให้ผลประสิทธิภาพความถูกต้องดีที่สุด

คำสำคัญ : การตัดต่อเสียง การตรวจจับการตัดต่อเสียง การเรียนรู้แบบรวมกลุ่ม เอตาบυσต์

Thesis Title	Edited Audio Detection Using Machine Learning
Student Name	Takdanai Suwan
Student ID	53650804
Degree	Master of Science Computer
Department	Computer Science
Year	2014
Thesis Advisor	Dr. Saichon Jaiyen
Thesis Co-advisor	Dr. Rungrat Wiangsripanawan

Abstract

Detecting edited audios is the challenge problem that can help forensic scientists to separate genuine, unedited recording from edited recordings. This research proposes the technique for detecting edited audios using machine learning. This problem can be considered as two-class classification problem which audio data are classified into two classes including edited and unedited audios. The performance of machine learning is compared to Ensemble learning, Support Vector Machine, Naïve Bayes, Radial Basis Function Neural Network, and Probabilistic Neural Networks. The experimental results demonstrate that the Ensemble learning is appropriated for detecting the edited audios because Ensemble learning give best accuracy

Keywords : Edited Audio, Edited Audio Detection, Ensemble Learning, Adaboost

กิตติกรรมประกาศ

วิทยานิพนธ์นี้มีอาจสำเร็จลุล่วงไปได้ หากมิได้รับคำแนะนำ คำชี้แนะ ความรู้และความเอาใจใส่จาก ดร.สายชล ใจเย็นและ ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์ผู้เป็นอาจารย์ที่ปรึกษา ซึ่งท่านได้สละเวลาและความรู้ให้กับข้าพเจ้าอย่างเต็มที่ จึงใคร่ขอขอบพระคุณอย่างสูง

ขอขอบคุณ ผศ.ดร.ศรัณย์ อินทโกสม ผศ.ดร.อนันตพร หารราชคุณาฒย และดร.ธนัสณี เพียรตระกูล คณะกรรมการสอบหัวข้อและโครงร่างวิทยานิพนธ์ ที่กรุณาให้คำแนะนำและข้อชี้แนะจนวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วง

ขอขอบพระคุณบิดา มารดาและญาติทุกคน ที่สนับสนุนด้านค่าใช้จ่ายในการเรียนและคอยเคียงข้างชี้แนะ และให้กำลังใจข้าพเจ้าจนสำเร็จการศึกษา

ขอขอบคุณพี่ๆ เพื่อนๆ น้องๆ ทุกคนที่ศูนย์อิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ที่คอยช่วยเหลืองาน ให้คำปรึกษา ให้ความรู้ และอำนวยความสะดวกในการติดต่อขอข้อมูลแก่ข้าพเจ้า จนทำวิทยานิพนธ์นี้สำเร็จลุล่วง

ขอขอบคุณเพื่อนๆ น้องๆ ร่วมรุ่นอันได้แก่ นายมงคล ทองไกรแก้ว นายอนุสรณ์ เจริญนาน นางสาวบุญหทัย เครือแก้ว นายจันิสธา ศรีสรवल นายวริษฐ์ เพ็ชรประสิทธิ์ นายนิพัทธ์ คล้ายโพธิ์ และเพื่อนพี่น้องทุกคนที่ไม่ได้กล่าวถึงที่ให้คำปรึกษา ให้ความรู้ ช่วยเหลืออำนวยความสะดวกในด้านต่างๆ

สำหรับคุณงานความดีและประโยชน์อันใดที่เกิดขึ้นจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับทุกคนที่กล่าวถึงในวิทยานิพนธ์นี้

ทักษ์ดนัย สุวรรณ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	จ
สารบัญรูป	ฉ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	1
1.3 สมมติฐานของงานวิจัย	2
1.4 ขอบเขตของงานวิจัย	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 ความรู้เบื้องต้นของเสียง	4
2.1.1 ประเภทของข้อมูลเสียง	5
2.1.2 สเปคโตรแกรม	6
2.2 การเรียนรู้ของเครื่อง	7
2.2.1 การเรียนรู้แบบมีผู้สอน	8
2.2.2 การเรียนรู้แบบไม่มีผู้สอน	9
2.2.3 การเรียนรู้แบบกึ่งมีผู้สอน	10
2.3 โครงข่ายประสาทเทียม	11
2.3.1 เพอร์เซ็ปตรอนหลายชั้น	14
2.3.2 โครงข่ายประสาทเทียมแบบเรเดียลเบสซิสฟังก์ชัน	16
2.4 ชับพอร์ตเวคเตอร์แมชชีน	19
2.4.1 เคอร์เนล ฟังก์ชัน	20
2.5 ตัวจำแนกแบบเบย์อย่างง่าย	21
2.6 การเรียนรู้ต้นไม้ตัดสินใจ	23
2.7 การเรียนรู้แบบรวมกลุ่ม	25
2.7.1 เอดาบуст อัลกอริทึม	27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8 วิจัยที่เกี่ยวข้อง.....	35
บทที่ 3 วิธีการดำเนินงานวิจัย.....	36
3.1 ขั้นตอนการเตรียมข้อมูล.....	37
3.1.1 การรวบรวมข้อมูลเสียง.....	38
3.1.2 การเตรียมข้อมูลเสียง.....	38
3.1.3 การแปลงข้อมูลเสียง.....	40
3.1.4 การแปลงข้อมูลในรูปแบบข้อมูลภาพ.....	42
3.1.5 Single Value Decomposition (SVD).....	43
3.2 ขั้นตอนการเรียนรู้และทดสอบโมเดล.....	45
3.3 ขั้นตอนการการวัดประสิทธิภาพ.....	47
บทที่ 4 การทดลองและผลการทดลอง.....	48
4.1 การทดลอง.....	48
4.1.1 การปรับค่า Kernel Function.....	48
4.2 ผลการทดลอง.....	49
บทที่ 5 สรุปและข้อเสนอแนะ.....	57
5.1 สรุป.....	57
5.2 ข้อเสนอแนะ.....	58
เอกสารอ้างอิง.....	59
ภาคผนวก ก.....	64
ประวัติผู้เขียน.....	70

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
3.1 เทคนิคการเรียนรู้ของเครื่องและฟังก์ชันที่ใช้ในการจำแนกประเภท.....	46
4.1 Kernel Function และ ค่าการปรับตัวแปรที่ดีที่สุด	49
4.2 การทดลองโมเดลการเรียนรู้แบบรวมกลุ่ม (ข้อมูลรูปแบบตัวเลข).....	49
4.3 การทดลองโมเดลซัพพอร์ตเวกเตอร์แมชชีน (ข้อมูลรูปแบบตัวเลข).....	50
4.4 การทดลองโมเดลตัวจำแนกแบบเบย์อย่างง่าย (ข้อมูลรูปแบบตัวเลข)	50
4.5 การทดลองโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน (ข้อมูลรูปแบบตัวเลข).....	51
4.6 การทดลองโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก (ข้อมูลรูปแบบตัวเลข).....	51
4.7 การเปรียบเทียบโมเดล 5 วิธีของข้อมูลเสียงรูปแบบตัวเลข.....	52
4.8 การทดลองโมเดลการเรียนรู้แบบรวมกลุ่ม (ข้อมูลรูปแบบภาพ).....	53
4.9 การทดลองโมเดลซัพพอร์ตเวกเตอร์แมชชีน (ข้อมูลรูปแบบภาพ).....	53
4.10 การทดลองโมเดลตัวจำแนกแบบเบย์อย่างง่าย (ข้อมูลรูปแบบภาพ).....	54
4.11 การทดลองโครงข่ายประสาทเทียมแบบ เรเดียลเบสิส ฟังก์ชัน (ข้อมูลรูปแบบภาพ).....	54
4.12 การทดลองโมเดลโครงข่ายประสาทเทียมแบบ โพรบาบิลิสติก (ข้อมูลรูปแบบภาพ).....	55
4.13 การเปรียบเทียบโมเดล 5 วิธีของข้อมูลเสียงรูปแบบภาพ.....	55
4.14 ตารางเปรียบเทียบค่าประสิทธิภาพของข้อมูล 2 รูปแบบ	56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างคลื่นเสียง.....	4
2.2 สเปกโตรแกรม.....	6
2.3 การเรียนรู้ของเครื่อง.....	7
2.4 การเรียนรู้แบบมีผู้สอน.....	8
2.5 การเรียนรู้แบบไม่มีผู้สอน.....	10
2.6 การเรียนรู้แบบเสริมแรง.....	11
2.7 รูปแบบของนิเวรอนในสมองมนุษย์.....	12
2.8 เปรียบเทียบนิเวรอน.....	12
2.9 รูปแบบของนิเวรอนในคอมพิวเตอร์.....	13
2.10 โครงข่ายประสาทเทียมแบบหลายชั้น.....	14
2.11 วิธีการการประมาณค่าในช่วง (Interpolation).....	16
2.12 โครงสร้างของเรเดียลเบสิสฟังก์ชัน (Radial basis neural network).....	18
2.13 ซัพพอร์ตเวกเตอร์แมชชีน.....	20
2.14 เคอเนล RBF (Kernel RBF).....	21
2.15 โครงสร้างต้นไม้ตัดสินใจ (Decision tree).....	23
2.16 ขั้นตอนการสร้างต้นไม้ตัดสินใจ.....	24
2.17 การเรียนรู้แบบรวมกลุ่ม.....	26
2.18 อัลกอริทึม เอดาบัสต์ (Adaboost Algorithm).....	27
2.19 การกำหนดค่า Weight (w).....	29
2.20 สร้างโมเดลในการเรียนรู้แบบกลุ่มตามค่า Iteration (T) ที่กำหนด.....	29
2.21 เริ่มต้นสอนโมเดลการจำแนกประเภทแบบพื้นฐาน.....	30
2.22 ค่าสัมประสิทธิ์ความผิดพลาด (α_t).....	31
2.23 ปรับปรุงค่าน้ำหนัก (Weight).....	32
2.24 สอนโมเดลการจำแนกประเภทแบบพื้นฐานถัดไป.....	33
2.25 ขั้นตอนการทำงานของการเรียนรู้แบบรวมกลุ่มโดยใช้เอดาบัสต์.....	34
3.1 ขั้นตอนการทำงานของโมเดลพิสูจน์ข้อมูลเสียง.....	36
3.2 ขั้นตอนการเตรียมข้อมูล.....	37
3.3 การนำข้อมูลเสียงเข้าโปรแกรม Adobe Audition.....	39
3.4 การแทนที่ข้อมูลเสียง.....	40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 ตัวอย่างของข้อมูลเสียงในรูปแบบตัวเลข	41
3.6 ตัวอย่างของข้อมูลภาพแบบสเปกโตรแกรม	42
3.7 ตัวอย่างของข้อมูลภาพแบบสเปกโตรแกรมสีเทา	43
3.8 ตัวอย่างเมตริกซ์ขนาด 4x5	43
3.9 เมตริกซ์ U.....	44
3.10 เมตริกซ์ Σ	44
3.11 เมตริกซ์ V	44
3.12 ขั้นตอนการเรียนรู้และทดสอบโมเดล	45



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันในประเทศไทยมีการใช้ข้อมูลเสียงกันอย่างแพร่หลาย เพื่อใช้ประโยชน์ในด้านต่างๆ ตัวอย่าง เช่น ในด้านความบันเทิงก็ใช้คลิปเสียงในการสร้างความสุขเป็นเสียงเพลง หรือในด้านธุรกรรมทางธุรกิจที่เก็บคลิปเสียงเพื่อเป็นหลักฐานในการอ้างอิงข้อมูลของการเงิน การใช้เป็นหลักฐานในการกู้หรือทำบัตรเครดิตต่างๆ แต่ในทางตรงกันข้ามก็มีการปลอมแปลงข้อมูลเสียงกันอย่างแพร่หลาย ทั้งการตัดต่อ การแทนที่เพื่อเปลี่ยนเนื้อหาให้ต่างจากเดิม โดยของจุดประสงค์การตัดแปลง คือ การแอบอ้างข้อมูลเสียงเพื่อโครงการมทรัพย์สิน การปลอมแปลงข้อมูลเสียงเพื่อลดความน่าเชื่อถือของบุคคล การกล่าวร้าย การละเมิดสิทธิ และการหลอกลวงเพื่อให้ได้ประโยชน์มาในทางมิชอบ การกระทำนี้ทำให้เกิดความเสียหายเสื่อมเสีย ในประเทศไทยถือว่าเป็นการผิดต่อกฎหมายอย่างรุนแรง ในประเทศไทยมีวิธีการพิสูจน์การตัดต่อโดยวิธีใช้ภาพสเปกโตรแกรมและให้ผู้เชี่ยวชาญด้านเสียงในการวิเคราะห์และพิสูจน์ข้อมูลเสียง โดยสังเกตจากลักษณะกราฟที่ผิดปกติ ซึ่งในการวิเคราะห์ต้องใช้ผู้เชี่ยวชาญจำนวนมากมาช่วยในการวิเคราะห์

ดังนั้นวิทยานิพนธ์นี้จึงเสนอวิธีการพิสูจน์ข้อมูลเสียงโดยใช้เทคนิคการเรียนรู้ของเครื่อง ได้แก่ การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) ตัวจำแนกแบบเบย์อย่างง่าย (Naïve Bayes) โครงข่ายประสาทเทียมแบบเรเดียลเบสฟังก์ชัน (Radial basis function neural network) และ โครงข่ายประสาทเทียมแบบโพรบาบิลิสติก (Probabilistic neural network) เข้ามาช่วยในการจำแนกประเภทของข้อมูลเสียง โดยทำการจำแนกประเภทข้อมูลเสียงแบ่งออกเป็น กลุ่มข้อมูลเสียงที่มีการตัดต่อ กลุ่มข้อมูลเสียงที่ไม่มีการตัดต่อ และเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) และเวลา ระหว่างแต่ละวิธีการจำแนกข้อมูลว่าวิธีไหนให้ค่าประสิทธิภาพความถูกต้องสูงสุด เพื่อช่วยในการพิสูจน์หลักฐานจากการแอบอ้างผ่านทางข้อมูลเสียง

1.2 วัตถุประสงค์

1. เพื่อนำเสนอวิธีการจำแนกประเภทข้อมูลเสียงโดยทำการจำแนกข้อมูลเสียงออกเป็นสองประเภทคือ ข้อมูลเสียงปกติและข้อมูลเสียงที่มีการตัดต่อโดยใช้เทคนิคการเรียนรู้ของเครื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เพื่อศึกษาและเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลเสียงแบบปกติกับข้อมูลเสียงในรูปแบบของภาพสเปกโตรแกรม

1.3 สมมติฐานของงานวิจัย

ปัญหาการระบุว่าข้อมูลเสียงเป็นแบบตัดต่อหรือไม่ตัดต่อ สามารถถูกพิจารณาว่าเป็นปัญหาการจำแนกประเภทข้อมูลได้ ดังนั้นการประยุกต์ใช้การเรียนรู้แบบรวมกลุ่ม สามารถจำแนกประเภทของข้อมูลเสียงได้อย่างถูกต้องและการแปลงข้อมูลเสียงเป็นภาพแบบสเปกโตรแกรมสามารถช่วยให้การจำแนกประเภทข้อมูลของตัวจำแนกประเภทมีประสิทธิภาพมากขึ้น

1.4 ขอบเขตการวิจัย

1. ข้อมูลเสียงเป็นข้อมูลเสียงผู้หญิงพูดภาษาไทยในภาวะห้องเงียบสนิทยะยะเวลา 3 วินาทีโดยทำการแบ่งข้อมูลเสียงเป็นสองส่วนคือ ข้อมูลเสียงแบบปกติและข้อมูลเสียงตัดต่อโดยข้อมูลเสียงตัดต่อใช้การแทนที่แทรกเข้าไปในข้อมูลเสียง
2. การเรียนรู้ของเครื่องรูปแบบการจำแนกประเภทข้อมูลมีหลายประเภท วิทยานิพนธ์นี้เลือกการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) มาใช้ในการทดสอบและเปรียบเทียบผลประสิทธิภาพความถูกต้องและเวลาเพื่อหาค่าดีที่สุดกับการจำแนกประเภทข้อมูลอีก 4 วิธีคือ ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) ตัวจำแนกแบบเบย์อย่างง่าย (Naïve Bayes) โครงข่ายประสาทเทียมแบบเรเดียลเบสิคฟังก์ชัน (Radial basis function neural network) และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก (Probabilistic neural network)
3. เปรียบเทียบข้อมูลเสียงปกติกับข้อมูลเสียงในรูปแบบภาพสเปกโตรแกรมในการทดสอบการเรียนรู้ของเครื่องว่า ข้อมูลรูปแบบเสียงปกติกับข้อมูลเสียงในรูปแบบภาพสเปกโตรแกรมว่า ข้อมูลรูปแบบไหนให้ผลดีกว่ากัน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถจำแนกประเภทของข้อมูลเสียงได้ว่าข้อมูลเสียงไหนเป็นข้อมูลเสียงตัดต่อ และข้อมูลเสียงไม่ตัดต่ออย่างถูกต้อง
2. ทราบถึงขั้นตอนวิธีการประยุกต์ใช้ขั้นตอนวิธีการเรียนรู้แบบรวมกลุ่มกับเทคนิคการแปลงข้อมูลเสียงเป็นภาพ เพื่อใช้ในการจำแนกประเภทข้อมูลเสียง
3. สามารถนำผลการเปรียบเทียบประสิทธิภาพและเวลาไปใช้เป็นแนวทางในการเลือกวิธีการจำแนกประเภทข้อมูลที่เหมาะสมที่สุด
4. สามารถนำไปประยุกต์กับนิติวิทยาศาสตร์เพื่อใช้ในการพิสูจน์หลักฐานด้านกฎหมายได้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

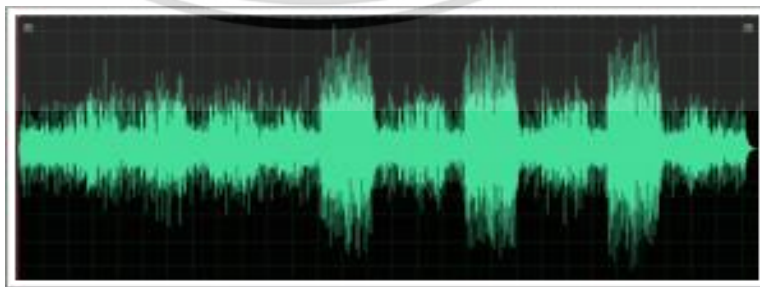
บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องได้แก่ ความรู้เบื้องต้นเกี่ยวกับข้อมูลเสียง ประเภทของไฟล์ข้อมูลเสียง ความรู้เบื้องต้นเกี่ยวกับการเรียนรู้ของเครื่อง การจำแนกประเภทข้อมูล ในส่วนที่สองจะกล่าวถึงทฤษฎีที่เกี่ยวข้องในการนำมาประยุกต์ใช้ในงานวิจัยได้แก่ การจำแนกประเภทข้อมูลด้วยซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) โครงข่ายประสาทเทียม (Neural Network) การเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ตัวจำแนกแบบเบย์อย่างง่าย (Naive Bayes) โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน (Radial Basis Function Neural Network) ส่วนสุดท้ายจะกล่าวถึงงานวิจัยที่เกี่ยวข้องโดยแบ่งประเภทของงานวิจัยที่เกี่ยวข้องดังนี้ การเรียนรู้ของเครื่อง การจำแนกประเภทข้อมูล ได้แก่ การจำแนกประเภทข้อมูลโดยใช้การเรียนรู้ของเครื่อง และการนำการเรียนรู้แบบรวมกลุ่มโดยใช้อัลกอริทึมเอดาบัสต์ไปประยุกต์ใช้งาน

2.1 ความรู้เบื้องต้นเกี่ยวกับข้อมูลเสียง

เสียง [1] เกิดจากการสั่นสะเทือนของวัตถุ โดยสามารถเกิดได้หลายวิธี ได้แก่ ตีดีด สีตีและเป่า เมื่อแหล่งกำเนิดเสียงหรือวัตถุเกิดการสั่นสะเทือน โมเลกุลที่อยู่ในอากาศจะสั่นตามไปด้วยทำให้เกิดเป็นช่วงอัดและช่วงขยายของโมเลกุลในอากาศ ซึ่งความถี่ที่เกิดขึ้นจะเท่ากับการสั่นของแหล่งกำเนิดเสียง พลังงานของการสั่นจะแผ่ขยายออกไปรอบๆ แหล่งกำเนิดเสียงทำให้ได้ยินเสียงโดยทั่วจากแหล่งกำเนิดเสียง เนื่องจากคลื่นเสียงเป็นคลื่นตามยาวเพราะโมเลกุลของอากาศ จะสั่นในทิศทางเดียวกับทิศทางที่เสียงเคลื่อนที่ไป การเปลี่ยนความดันอากาศนี้สามารถเคลื่อนที่ไปข้างหน้าจนถึงหูของผู้ฟังทำให้ได้ยินเสียง โดยความดังของเสียงจะขึ้นอยู่กับช่วงกว้างของการสั่น (แอมพลิจูด) แอมพลิจูดสูงเสียงจะดังมาก



รูปที่ 2.1 ตัวอย่างคลื่นเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในปัจจุบันมีการนำเสียงมาทำเป็นข้อมูลในเชิงดิจิทัลอย่างแพร่หลาย ซึ่งจะอยู่ในรูปแบบของไฟล์ข้อมูลและมีการพัฒนารูปแบบของไฟล์ข้อมูลเสียงอย่างต่อเนื่อง ซึ่งสามารถเก็บและเปิดได้กับอุปกรณ์อิเล็กทรอนิกส์หลากหลายประเภทอาทิเช่น คอมพิวเตอร์ โทรศัพท์ โทรทัศน์ วิทยุ หรือเครื่องเล่นเสียงแบบพกพา ในทำนองเดียวกันเทคโนโลยีการบันทึกเสียงก็พัฒนาจนสามารถนำไปใช้งานอย่างง่ายดายเช่นเดียวกัน สามารถบันทึกเสียงให้อยู่ในรูปแบบของข้อมูลเสียงดิจิทัล โดยใช้แค่เพียงโทรศัพท์มือถือหรือเครื่องบันทึกเสียงคุณภาพของไฟล์ข้อมูลเสียงที่บันทึกจะมีความแตกต่างกันออกไปดังนี้

2.1.1 ประเภทของข้อมูลเสียง

ประเภทของข้อมูลเสียง [2] มีได้หลายรูปแบบ เช่น แบบไม่บีบอัด (Uncompressed) คือ การอัดเสียงหรือเข้ารหัสเสียงอยู่ในรูปแบบไฟล์ดิจิทัลดั้งเดิมที่ได้จากโปรแกรมอัดเสียง และแบบทำการบีบอัด (Compressed) คือ การนำเสียงที่อัดดั้งเดิมมาทำการลดขนาดไฟล์ข้อมูลเสียงให้มีขนาดเล็กลงเพื่อประหยัดเนื้อที่ในการเก็บประเภทของข้อมูลเสียงแบ่งได้ดังนี้

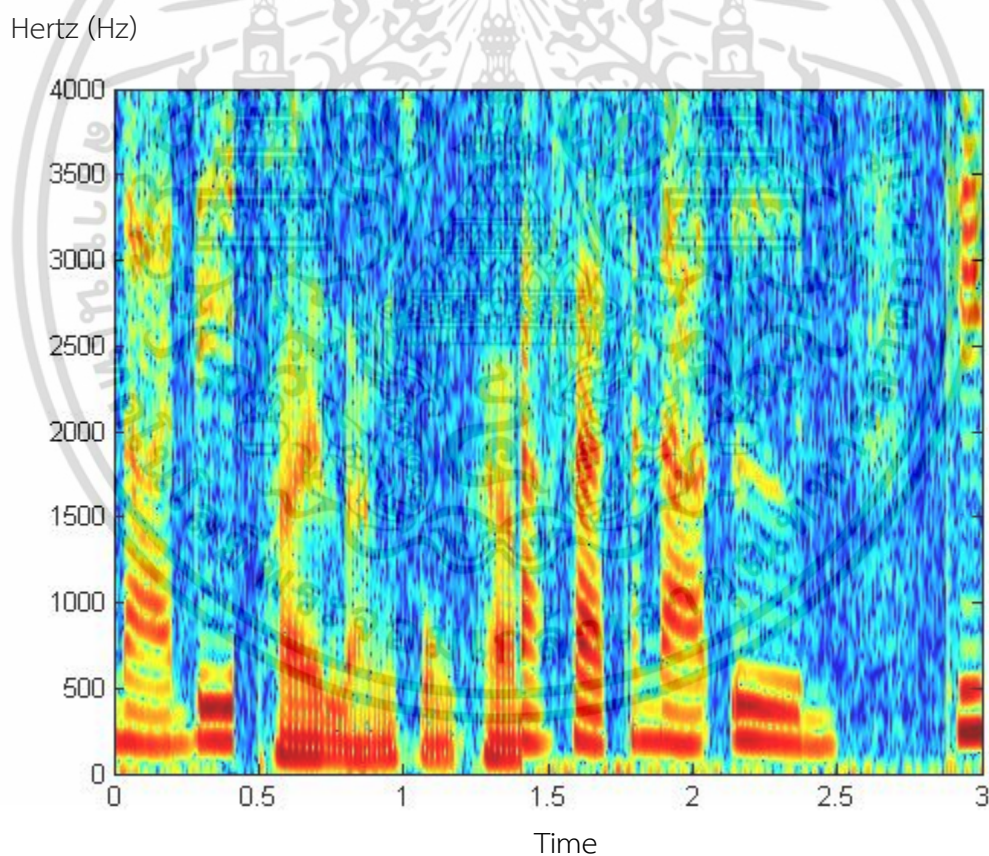
1.) ไฟล์ข้อมูลเสียงที่ไม่มีการบีบอัด (Uncompressed audio format) เป็นไฟล์ข้อมูลเสียงต้นฉบับที่ไม่มีการบีบอัดและใช้หลักการแปลงสัญญาณอนาล็อกเป็นดิจิทัลในรูปแบบเชิงเส้น (Linear pulse-code modulation: LPCM) ในการเก็บเสียงให้อยู่ในรูปแบบของข้อมูลเสียงบนคอมพิวเตอร์เพื่อให้ได้เสียงที่มีคุณภาพสูงสุด นามสกุลของไฟล์ประเภทนี้ได้แก่ WAV, AIFF, AU เป็นต้น

2.) ไฟล์ข้อมูลเสียงที่มีการบีบอัดแบบไม่มีการสูญเสียข้อมูล (Lossless compressed audio format) เป็นการบีบอัดเสียงเพื่อลดขนาดไฟล์ โดยบีบอัดให้สูญเสียข้อมูลน้อยที่สุดและยังคงคุณภาพของเสียงเหมือนไฟล์ต้นฉบับที่ไม่มีการบีบอัด นามสกุลของไฟล์ประเภทนี้ได้แก่ FLAC, WavPack เป็นต้น ทาง Microsoft ได้ทำไฟล์ข้อมูลเสียง Window Lossless ขึ้นมาเองซึ่งมีนามสกุลของไฟล์คือ WMA เช่นเดียวกัน Apple ได้สร้างรูปแบบไฟล์ข้อมูลเสียงซึ่งมีนามสกุลของไฟล์คือ ALAC

3.) ประเภทไฟล์ข้อมูลเสียงที่มีการบีบอัดและสูญเสียข้อมูลมาก (Lossy compressed audio format) เป็นการลดขนาดไฟล์ของข้อมูลเสียงอย่างมาก โดยทำการนำบางส่วนของเสียงออกไปทำให้ไฟล์มีรายละเอียดของเสียงต้นฉบับไม่ครบถ้วน แต่ก็ยังคงใจความเนื้อหาหลักเอาไว้ ตัวอย่างเช่นไฟล์ข้อมูลเสียงประเภทที่เพลงมีความคมชัดแต่รายละเอียดของดนตรีจะลดลง แต่ยังคงเสียงของคนร้องและดนตรีโดยในภาพรวมที่เป็นใจความหลักของเพลงไว้ นามสกุลของไฟล์ประเภทนี้ได้แก่ MP3, AAC เป็นต้น

2.1.2 สเปกโตรแกรม (Spectrogram)

สเปกโตรแกรม (Spectrogram) [3] คือ กราฟแสดงความถี่ของข้อมูลเสียงที่เปลี่ยนแปลงตามเวลา ซึ่งจะมีอยู่สองแกน ได้แก่ แกน X คือ เวลาของข้อมูลเสียง เวลาในที่นี้หมายถึงความยาวของข้อมูลเสียง ซึ่งแต่ละข้อมูลเสียงจะมีความยาวของคลิปแตกต่างกัน โดยแกน X จะแปรผันตามเวลาของข้อมูลเสียงนั้น ส่วนแกน Y คือ ความถี่ของคลื่นเสียง ถ้าภาพภายในกราฟขึ้นสูงแสดงว่ามีความถี่ของเสียงมาก ความถี่ของเสียงมีหน่วยเป็น เฮิรตซ์ (Hertz) รูปที่ 2.2 สีแดงและเหลืองหมายถึง ส่วนที่มีการเปล่งเสียง หรือมีเสียงพูดภายในข้อมูลเสียง สีฟ้าคือ สีพื้นหลังเป็นสีที่แตกต่างจากส่วนของกราฟหลักทำให้มองเห็นความแตกต่างได้ง่าย แต่ในขณะเดียวกันสีฟ้าก็อาจจะเป็นเสียงรบกวน (Noise) ได้เช่นเดียวกันเพราะเมื่อทำการแปลงข้อมูลเสียงเป็นภาพสเปกโตรแกรมแล้ว เสียงข้อมูลหลัก (คนพูดหลัก) ในภาพจะถูกแทนด้วยสีแดงและเหลืองเพราะเป็นความถี่หลัก ส่วนเสียงรบกวนหรือเสียงรบกวน (Noise) จะแทนที่ด้วยสีฟ้า

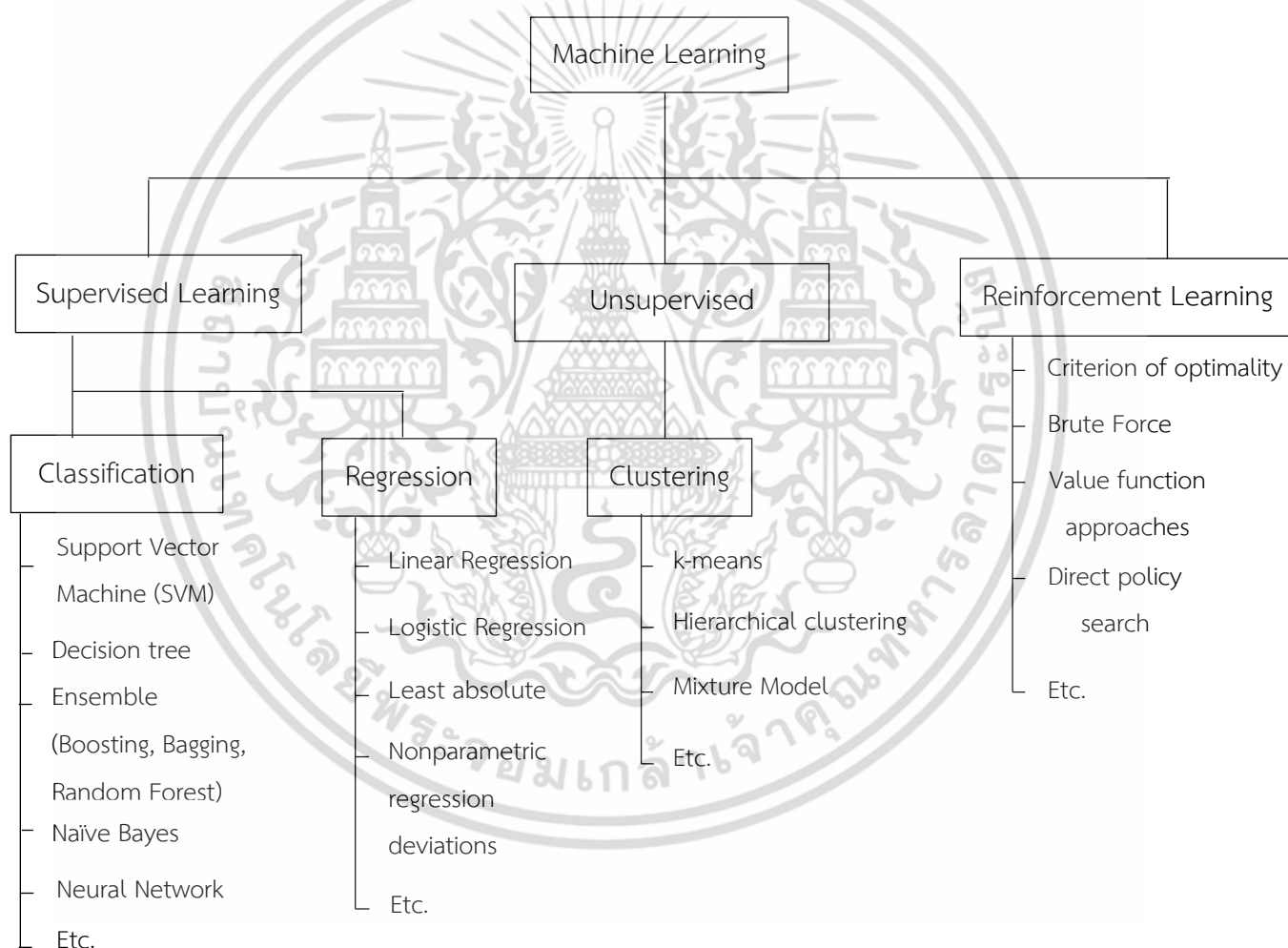


รูปที่ 2.2 สเปกโตรแกรม

2.2 การเรียนรู้ของเครื่อง (Machine learning)

การเรียนรู้ของเครื่อง (Machine learning) [4] เป็นการพัฒนาเครื่องให้มีการเรียนรู้จากข้อมูล ตัวอย่างหรือจากสภาพแวดล้อมได้ จุดมุ่งหมายคือ พัฒนาหรือปรับปรุงประสิทธิภาพการทำงานของระบบให้ดีขึ้น เมื่อทำการเรียนรู้ จะนำความรู้ที่เรียนเก็บไว้ในฐานความรู้ในรูปแบบสิ่งแทนความรู้บางอย่างใดอย่างหนึ่ง เช่น กฎ ฟังก์ชัน ฯลฯ โดยการเรียนรู้ของเครื่องได้นำความรู้ด้านสถิติศาสตร์และคณิตศาสตร์เข้ามาประยุกต์ เนื่องจากทั้งสองสาขาศึกษาการวิเคราะห์ข้อมูลเช่นเดียวกัน การเรียนรู้ของเครื่องแบ่งได้ดังรูปที่

2.3



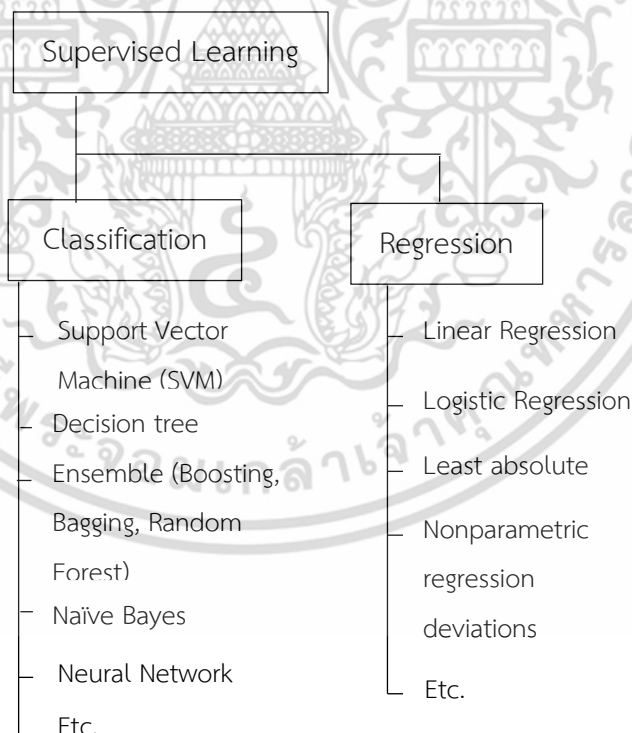
รูปที่ 2.3 การเรียนรู้ของเครื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.3 การเรียนรู้ของเครื่องสามารถแบ่งได้เป็น 3 ประเภท คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และ การเรียนรู้แบบเสริมแรง (Reinforcement Learning) อธิบายได้ดังนี้

2.2.1 การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน (Supervised Learning) [5] คือ การสอนระบบโดยบอกว่าข้อมูลอยู่ในกลุ่มอะไร ข้อมูลในการสอนระบบจะอยู่ในรูปแบบเวกเตอร์เสมอ ไม่ว่าข้อมูลจะมาในรูปแบบไหนก็จะทำการแปลงเป็นเวกเตอร์ก่อนถึงจะสามารถใช้ในการสอนระบบได้ โดยข้อมูลที่ใช้สอนทุกข้อมูลจะมีการใส่ผลลัพธ์เป้าหมาย (Target) ไว้เพื่อบอกว่าข้อมูลนี้อยู่ในกลุ่มไหน ในการทดสอบจะใช้ข้อมูลคู่กับผลลัพธ์เป้าหมาย (Target) เช่นเดียวกันและจะทดสอบว่าข้อมูลนี้อยู่ในกลุ่มตามผลลัพธ์เป้าหมาย (Target) หรือไม่ โดยมีความถูกต้องที่เปอร์เซ็นต์ การประยุกต์ใช้การเรียนรู้แบบมีผู้สอน (Supervised Learning) แบ่งเป็น 2 ประเภทคือ การจำแนกประเภทข้อมูล (Classification) และการถดถอย (Regression)



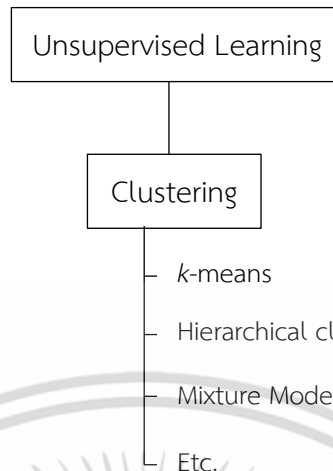
รูปที่ 2.4 การเรียนรู้แบบมีผู้สอน (Supervised Learning)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. การจำแนกประเภทข้อมูล (Classification) คือ การจำแนกประเภทข้อมูลเป็นกลุ่มจำนวน 2 กลุ่มขึ้นไป ตัวอย่างการจำแนกประเภทของข้อมูลเสียงโดยการจำแนกข้อมูลเป็น 2 กลุ่ม ได้แก่ ข้อมูลเสียงแบบปกติกับข้อมูลเสียงแบบตัดต่อและทำการกำหนดผลลัพธ์เป้าหมายให้กับข้อมูลทั้งสองประเภท จากนั้นใช้อัลกอริทึมในการจำแนกประเภทข้อมูลโดยอัลกอริทึมการจำแนกประเภทข้อมูลมีอยู่หลายอัลกอริทึม ยกตัวอย่างจากรูป 2.4 ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector machine) การเรียนรู้ต้นไม้ตัดสินใจ (Decision tree) การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ตัวจำแนกแบบเบย์อย่างง่าย (Naive Bayes) โครงข่ายประสาทเทียม (Neural network) เป็นต้น
2. การถดถอย (Regression) คือ การใช้ความรู้ทางสถิติมาศึกษาความสัมพันธ์ของข้อมูล เพื่อพยากรณ์หรือทำนายข้อมูลที่มีความสัมพันธ์กันในอนาคต ตัวอย่างเช่น ข้อมูลมูลค่าของหุ้นในอนาคต ความสัมพันธ์คือซื้อหุ้นและมูลค่าของหุ้น เช่น หุ้น A ทำการทำนายเพื่อให้ทราบถึงมูลค่าของหุ้น A ว่ามูลค่าของหุ้น A ในอนาคตจะเป็นอย่างไร เป็นต้น

2.2.2 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ใช้ข้อมูลฝึกสอนหรือชุดตัวอย่างที่ไม่มีการใส่ผลลัพธ์เป้าหมาย (Target) ให้กับข้อมูลและระบบจะทำการเรียนรู้โดยการนำข้อมูลไปผ่านกระบวนการหาความคล้ายคลึงกับข้อมูลตัวอย่างที่สอน จนกระทั่งได้กลุ่มตัวอย่างที่จัดเป็นประเภทต่างๆอย่างเหมาะสมหรือเมื่อนำข้อมูลนำเข้าในอัลกอริทึมการเรียนรู้ของเครื่อง อัลกอริทึมจะทำการแบ่งประเภทข้อมูลโดยอัตโนมัติ เช่น ข้อมูลเสียงที่มีความแตกต่างกันตัวอย่างเช่นเสียงผู้ชายและผู้หญิง อัลกอริทึมจะทำการแบ่งกลุ่มจากคุณลักษณะของข้อมูลที่แตกต่างกันคือ กลุ่มเสียงผู้ชายและกลุ่มเสียงผู้หญิง เทคนิคประเภทนี้ได้แก่ การแบ่งกลุ่ม (clustering)



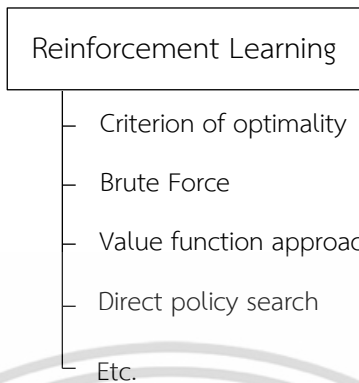
รูปที่ 2.5 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

จากรูป 2.5 ตัวอย่างอัลกอริทึมที่ใช้ในการเรียนรู้แบบไม่มีผู้สอนได้แก่ *k-mean*, Hierarchical clustering, Mixture model เป็นต้น

2.2.3 การเรียนรู้แบบเสริมแรง (Reinforcement)

การเรียนรู้แบบเสริมแรง (Reinforcement) คือ การกระทำให้ผู้แสดงพฤติกรรมเกิดความพึงพอใจเมื่อทำพฤติกรรมใดพฤติกรรมหนึ่งและต้องการให้ทำพฤติกรรมนั้นซ้ำๆ อีก เช่น เมื่อนักเรียนตอบคำถามถูกต้อง ครูให้รางวัล (นักเรียนพอใจ) นักเรียนจะตอบคำถามอีกหากครูถามคำถามครั้งต่อไป ประเภทของการเสริมแรง แบ่งออกเป็น 2 ประเภท ได้แก่ การเสริมแรงบวก คือ การให้ตัวเสริมแรงบวก เมื่อทำพฤติกรรมที่กำหนด เช่น ทำงานเสร็จตามเป้าหมายจะได้รับค่าจ้าง การทำงานคือพฤติกรรมที่กำหนด ส่วนเงินค่าจ้างเป็นตัวเสริมแรงบวกให้อยากทำพฤติกรรมนั้นต่อและการเสริมแรงลบ คือ การให้ตัวเสริมแรงลบ เมื่อทำพฤติกรรมที่กำหนด เช่น เมื่ออยู่ในห้องที่อบอ้าวเราจะเปิดหน้าต่าง เปิดหน้าต่างเป็นพฤติกรรมที่กำหนด หายอบอ้าวเป็นตัวเสริมแรงลบ หรือนักเรียนที่ตอบคำถามครูถูกต้องจะได้รับการยกเว้นไม่ต้องทำรายงานมาส่ง เป็นต้น อัลกอริทึมของการเรียนรู้แบบเสริมแรง (Reinforcement) ได้แก่ Criterion of optimality, Brute force, Value function approaches, Direct policy search เป็นต้น ดังรูป 2.6

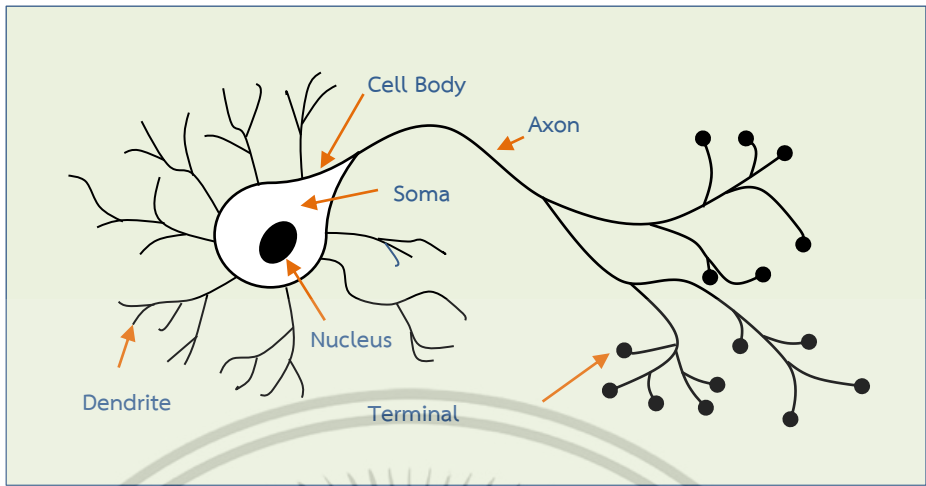
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 การเรียนรู้แบบเสริมแรง (Reinforcement)

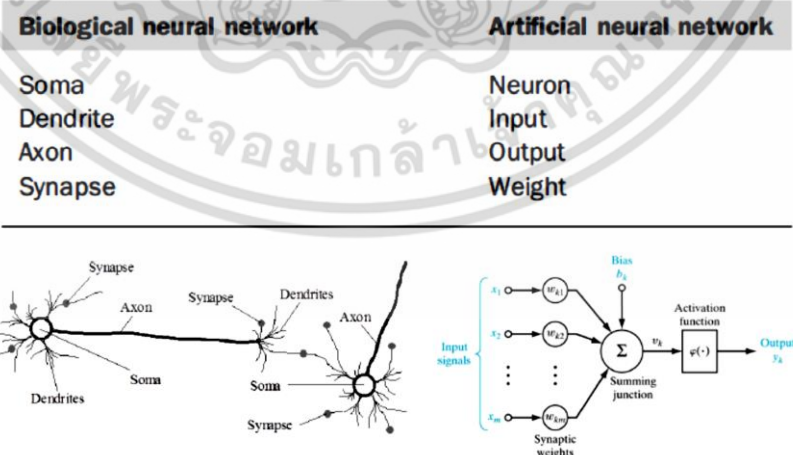
2.3 โครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม [6][7][8] (Artificial neural network) หรือที่มักจะเรียกสั้นๆ ว่าข่ายงานประสาท (neural network หรือ neural net) คือ โมเดลทางคณิตศาสตร์หรือโมเดลการคำนวณที่จำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์เพื่อวัตถุประสงค์ในการสร้างเครื่องมือให้มีความสามารถในการเรียนรู้การจดจำรูปแบบ (Pattern Recognition) การอุปมานและการเรียนรู้จดจำข้อมูลความรู้เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ แนวคิดเริ่มต้นของโครงข่ายประสาทเทียมเริ่มจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาทหรือ “นิวรอน” (neurons) และจุดประสานประสาท (synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรต์”(Dendrite) ซึ่งเป็นข้อมูลขาเข้า (input) และปลายในการส่งกระแสประสาทเรียกว่า “แอกซอน” (Axon) ซึ่งเป็นเหมือนข้อมูลขาออก (output) ของเซลล์เหล่านี้ การทำงานด้วยปฏิกิริยาไฟฟ้าเคมีจะเกิดขึ้นเมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกันกระแสประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่นๆ ต่อไปผ่านทางแอกซอน ดังรูปที่ 2.7



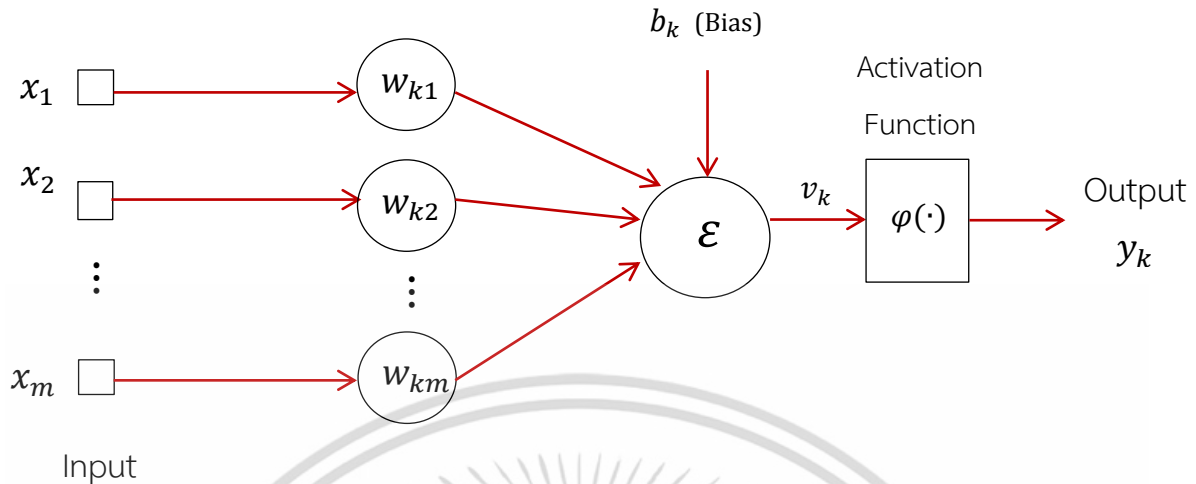
รูปที่ 2.7 รูปแบบของนิวรอนในสมองมนุษย์

สำหรับในคอมพิวเตอร์นิวรอน (Neurons) ประกอบด้วยข้อมูลนำเข้า (Input) และข้อมูลส่งออก (Output) โดยจำลองให้ข้อมูลนำเข้า (Input) แต่ละข้อมูลมีค่าน้ำหนัก (Weight) เป็นตัวกำหนดน้ำหนักของข้อมูลนำเข้า (Input) โดยนิวรอน (Neurons) แต่ละโหนดจะมีค่าเทรชโฮลด์ (Threshold) เป็นตัวกำหนดว่าค่าน้ำหนักรวมของข้อมูลนำเข้า (Input) ต้องมากขนาดไหนจึงจะสามารถส่งข้อมูลส่งออก (Output) ไปยังนิวรอน (Neurons) ตัวอื่น เมื่อนำนิวรอน (Neurons) แต่ละหน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ ในทางตรรกะแล้วก็จะเหมือนกับปฏิกิริยาไฟฟ้าเคมีที่เกิดขึ้นในสมอง แต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลขเท่านั้น ดังรูป 2.8 เปรียบเทียบระหว่างนิวรอนในสมองมนุษย์กับนิวรอนในคอมพิวเตอร์



รูปที่ 2.8 เปรียบเทียบนิวรอน [8]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

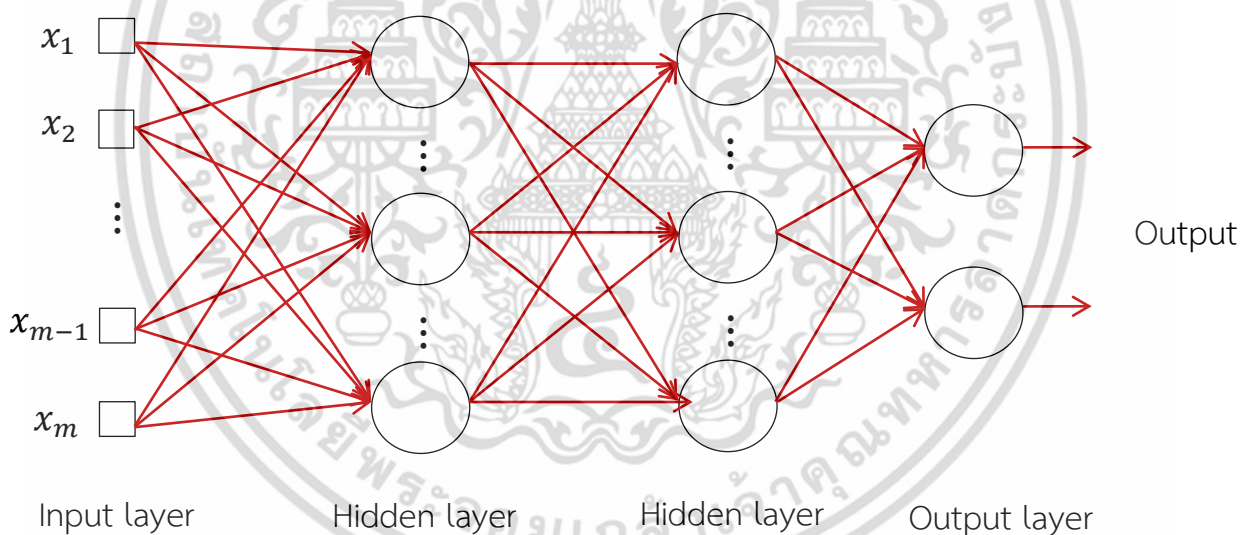


รูปที่ 2.9 รูปแบบของนิวรอนในคอมพิวเตอร์

จากรูป 2.9 คือตัวอย่างโครงสร้างของนิวรอน (Neuron) ในคอมพิวเตอร์ซึ่งมีอยู่ 5 ส่วน ได้แก่ ส่วนแรกคือข้อมูลนำเข้า (Input) แทนโดยค่า x ซึ่งข้อมูลนำเข้าจะมีหลายจำนวนขึ้นอยู่กับจำนวนแอดทริบิวต์ของข้อมูลที่จะนำมาสอนและค่า m แทนจำนวนแอดทริบิวต์ของข้อมูลนำเข้า ส่วนที่สองคือน้ำหนักหรือไซแนป (Synaptic weights) หรือค่าน้ำหนักของข้อมูลแทนโดยค่าน้ำหนัก (w) เมื่อได้ค่าข้อมูลนำเข้า (Input) และค่าน้ำหนัก (Weight) ค่าทั้งสองจะถูกนำมาคูณกันและส่งต่อไปยังส่วนของนิวรอน (Neurons) ซึ่งเป็นส่วนที่สาม หรือส่วนผลรวมเชิงเส้น (Summing Junction) ส่วนนี้จะเป็นส่วนรวบรวมผลคูณและไบแอส (Bias) เพื่อจำแนกประเภทข้อมูลในนิวรอน (Neurons) เมื่อทำงานเสร็จเรียบร้อย จะส่งต่อไปยังส่วนฟังก์ชันการกระตุ้น (Activation Function) ในส่วนนี้จะใช้ฟังก์ชันการบีบค่าเช่นฟังก์ชันซิกมอย (Sigmoid Function) มาช่วยในการปรับค่าผลรวมเชิงเส้นให้เป็นค่าเป้าหมาย (Target) ฟังก์ชันการกระตุ้นที่นิยมใช้กันมากคือฟังก์ชันซิกมอย (Sigmoid Function) จะมีการรับค่าที่กว้างมากคือค่าเป็นบวกไม่จำกัด ($+\infty$) ถึงค่าติดลบไม่จำกัด ($-\infty$) จากนั้นจะค่อยๆทำการบีบค่าให้แคบลงตามลักษณะ curve ของฟังก์ชันซิกมอย (Sigmoid Function) จนได้ค่าที่อยู่ในช่วงที่ต้องการ

2.3.1 เพอร์เซ็ปตรอนหลายชั้น (Multi-layer Perceptron)

โครงข่ายประสาทเทียมนี้เป็นรูปแบบหนึ่งที่มีโครงสร้างแบบหลายชั้น ใช้สำหรับปัญหาที่มีความซับซ้อนของโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น มีกระบวนการฝึกสอนเป็นแบบมีผู้สอน (Supervise) และใช้ขั้นตอนวิธีการส่งค่าย้อนกลับ (Back propagation) สำหรับการฝึกสอนซึ่งประกอบด้วย 2 ส่วนย่อย คือ การส่งผ่านไปข้างหน้า (Forward Pass) และการส่งผ่านย้อนกลับ (Backward Pass) โครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้นมีส่วนประกอบคือ ชั้นข้อมูลนำเข้า (Input layer) ชั้นข้อมูลส่งออก (Output layer) และชั้นฮิดเดน (Hidden layer) ซึ่งลักษณะของโครงข่ายประสาทเทียมรูปแบบนี้สามารถมีชั้นฮิดเดนชั้นเดียวหรือหลายชั้นก็ได้ การส่งผ่านไปข้างหน้า ข้อมูลจะผ่านเข้าโครงข่ายประสาทเทียมในชั้นข้อมูลนำเข้า (Input) และส่งผ่านเข้าส่วนของฮิดเดน (Hidden layer) ต่อกันไปเรื่อยๆ จนถึงข้อมูลส่งออก (Output) รูปที่ 2.10 คือตัวอย่างสถาปัตยกรรมโครงข่ายประสาทเทียมแบบหลายชั้น



รูปที่ 2.10 โครงข่ายประสาทเทียมแบบหลายชั้น [9]

อัลกอริทึม Back Propagation ใช้สำหรับสอนโครงข่ายประสาทเทียมโดยจะคำนวณจากค่าความผิดพลาด (Error rate) ค่าความผิดพลาดน้อยแสดงว่ามีการเรียนรู้ที่ดี โดยโครงข่ายประสาทเทียมจะทำการปรับค่าน้ำหนัก (Weight) ให้สัมพันธ์กับการแก้ไขข้อผิดพลาด (Error-Correction) ทำการปรับค่าน้ำหนักไปเรื่อยๆ จนกระทั่งได้ผลคำตอบที่แท้จริงคือ โกล์ผลลัพธ์เป้าหมายหรือที่มีค่าความผิดพลาดน้อย

ที่สุด โดยจะทำการปรับค่าน้ำหนัก (Weight) ดังสมการ 2.1 และ 2.2

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad (2.1)$$

โดยที่
$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (2.2)$$

โดยที่ค่า w แทนค่าน้ำหนัก และค่า E แทนค่าความผิดพลาด สมการ 2.1 ใช้ในการคำนวณเพื่อปรับค่าน้ำหนัก (Weight) ในการปรับการเรียนรู้ของโครงข่ายประสาทเทียมโดยมีสมการ 2.2 ใช้ในการคำนวณปรับค่าการเรียนรู้ในการปรับค่าน้ำหนัก (Weight) ตัวอย่างเช่นถ้าค่าน้ำหนักตัวแรก (w_0) มีค่าความผิดพลาด (Error rate) ที่ยังสูง สมการ 2.2 จะทำการปรับค่าลงมาที่ w_1 จะได้ค่าความผิดพลาด (Error rate) ที่ลดต่ำลงเข้าใกล้ 0 ผลที่ได้ดีขึ้น โดยหาค่าความผิดพลาดได้จากสมการ 2.3 และ 2.4

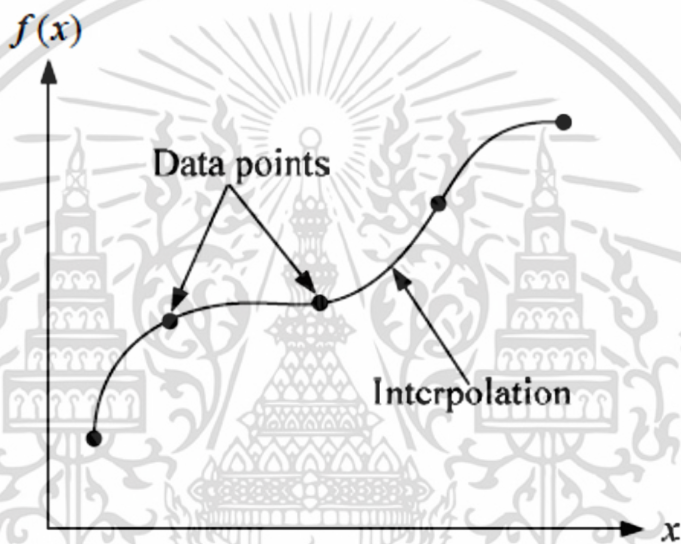
$$e_j = d_j - y_j \quad (2.3)$$

$$E = \sum_j e_j^2 \quad (2.4)$$

โดยที่ค่า e แทนค่าความผิดพลาดของแต่ละนิวรอน ค่า d แทนค่าผลลัพธ์เป้าหมาย ค่า y แทนผลลัพธ์ที่ได้จากโครงข่าย ค่า i แทนดัชนีของข้อมูล ค่า j แทนดัชนีของนิวรอน ค่า E แทนค่าความผิดพลาดรวมจากสมการ 2.4 ทำการหาค่าความผิดพลาดของแต่ละนิวรอนในชั้นข้อมูลส่งออกโดยการนำค่าผลลัพธ์ลบกับค่าเป้าหมายจะได้ค่าความผิดพลาดของนิวรอนออกมา เมื่อทำการหาค่าความผิดพลาดครบทุกนิวรอน จะนำค่าความผิดพลาดของนิวรอนทุกตัวมาทำการหาค่าความผิดพลาดรวมโดยสมการ 2.4 ซึ่งค่าความผิดพลาดรวมจะนำไปใช้ในการปรับค่าน้ำหนัก (Weight) ในสมการ 2.2

2.3.2 โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน (Radial basis Function Neural Network)

โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน (Radial basis function neural network) (RBF) [9][10] เป็นโครงข่ายประสาทเทียมที่ถูกคิดค้นขึ้นมาเพื่อแก้ไขปัญหาการประมาณค่าในช่วง (Interpolation Problem) ของข้อมูลที่มีมิติสูง วิธีการการประมาณค่าในช่วง (Interpolation) คือ การหาฟังก์ชันที่ลากผ่านทุกจุดของข้อมูลที่มีดังรูปที่ 2.11



รูปที่ 2.11 วิธีการการประมาณค่าในช่วง (Interpolation)

ตัวอย่างกำหนดเซตของจุด N จุด $\{x_i \in R^n | i = 1, 2, \dots, N\}$ และกำหนดเซตของผลลัพธ์เป้าหมายที่เป็นจำนวนจริง $\{d_i \in R^1 | i = 1, 2, \dots, N\}$ จากนั้นจะหาฟังก์ชัน $F: R^n \rightarrow R^1$ ที่สอดคล้องตามเงื่อนไขของการประมาณค่าในช่วง

$$F(x_i) = d_i \quad i = 1, 2, \dots, N \quad (2.5)$$

โดยที่ $F(\mathbf{x}_i)$ แทนฟังก์ชัน และ d แทนค่าผลลัพธ์เป้าหมาย

การประมาณค่าในช่วง (Interpolation) ตามที่ได้กล่าวไปแล้วจะเลือกใช้ฟังก์ชันที่เรียกว่า เรเดียลเบซิสฟังก์ชัน มาช่วยในการประมาณค่าในช่วงดังสมการ 2.6

$$F(x) = \sum_{i=1}^N w_i \cdot \varphi(\|x - x_i\|) \quad (2.6)$$

โดยที่ $\varphi(\cdot)$ คือ เบสิคฟังก์ชัน

w_i คือ ค่าน้ำหนัก

$\|\cdot\|$ คือ นอร์มแบบยูคลิด (Euclidean Norm)

$x_i \in R^n$ คือ จุดศูนย์กลาง (Centers) ของเรเดียลเบสิคฟังก์ชัน

จากนั้นเริ่มการประมาณค่าในช่วงจากสมการ 2.5 และ 2.6 โดยใช้ข้อมูลทั้งหมด จะได้สมการเชิงเส้นของตัวแปรที่เป็นค่าน้ำหนัก (Weight) ดังสมการที่ 2.7

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \quad (2.7)$$

โดยที่

$$\varphi_{ji} = \varphi(\|x_j - x_i\|), (j, i) = 1, 2, \dots, N$$

$$d = [d_1, d_2, \dots, d_N]^T \quad (2.8)$$

$$w = [w_1, w_2, \dots, w_N]^T$$

จากสมการ 2.8 d คือ เวกเตอร์ผลลัพธ์ที่ต้องการ (desired response vector) และ w คือ เวกเตอร์น้ำหนักเชิงเส้น (linear weight vector) โดยที่ N คือขนาดของตัวอย่างในการสอนระบบให้ Φ แทนเมทริกซ์ขนาด $N \times N$ ประกอบด้วยสมาชิก φ_{ji} ดังสมการ 2.9

$$\Phi = \{\varphi_{ji} | (j, i) = 1, 2, \dots, N\} \tag{2.9}$$

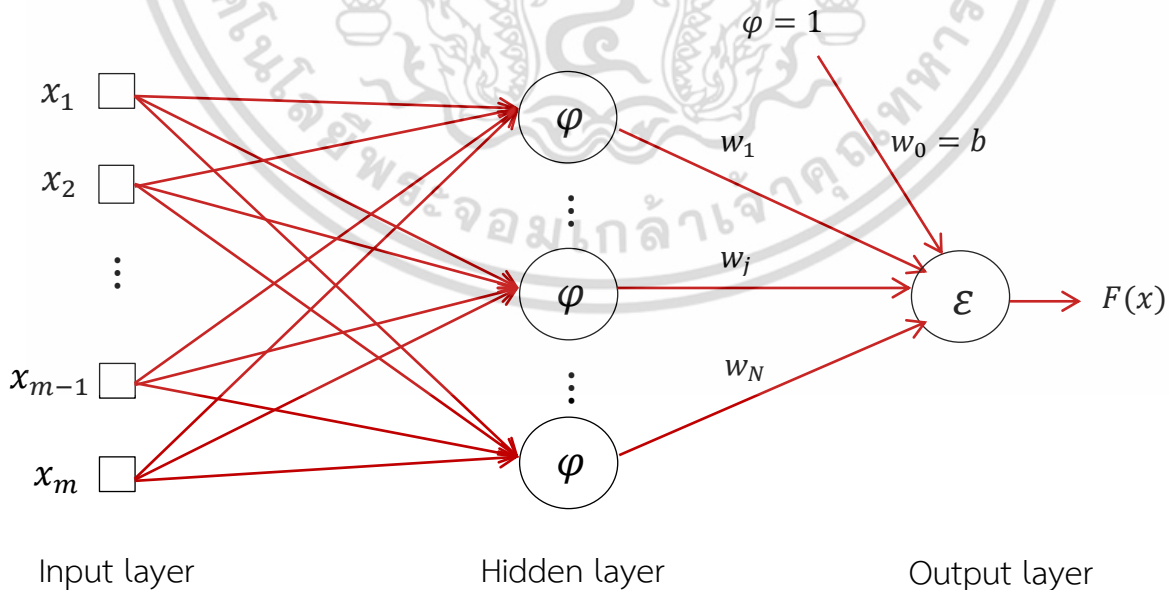
โดยเรียกเมทริกซ์นี้ว่าเมทริกซ์การประมาณค่าในช่วง (Interpolation matrix) จากนั้นทำการลดรูปสมการ 2.7 ให้อยู่ในรูปเมทริกซ์ ดังสมการที่ 2.10

$$\Phi w = x \tag{2.10}$$

สมมติว่า Φ เป็นเมทริกซ์ที่สามารถหาเมทริกซ์ผกผันได้ (inverse matrix) เมื่อทำการแทนค่าในสมการ 2.10 ค่าเวกเตอร์น้ำหนัก (w) จะสามารถคำนวณได้ดังสมการ 2.11

$$w = \Phi^{-1}x \tag{2.11}$$

จากปัญหาการประมาณค่าในช่วง (Interpolation problem) สามารถนำเรเดียลเบสซิสฟังก์ชันมาใช้ในการสร้างโครงข่ายประสาทเทียมแบบเรเดียลเบสซิสฟังก์ชันได้ โดยโครงข่ายประสาทเทียมแบบนี้จะทำการแมพข้อมูลจากปริภูมิข้อมูลนำเข้าไปยังปริภูมิคุณลักษณะเพื่อจำแนกประเภทข้อมูลบนปริภูมิคุณลักษณะ (Feature Space) ข้อเสียของโครงข่ายประสาทเทียมแบบเรเดียลเบสซิสฟังก์ชัน (Radial basis Function neural network) คือ ถ้าข้อมูลมีการเกาะกลุ่มกันสูง เมื่อนำฟังก์ชันเชิงรัศมีมาครอบข้อมูลจะทำให้เกิดความผิดพลาดสูงเพราะข้อมูลต่างกันอยู่ในฟังก์ชันเชิงรัศมีเดียวกันทำให้ข้อมูลปะปนกันทำให้เกิดความผิดพลาด (Error) รูปที่ 2.12 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบเรเดียลเบสซิสฟังก์ชัน (Radial Basis Neural Network)



รูปที่ 2.12 โครงสร้างของเรเดียลเบสซิสฟังก์ชัน (Radial basis neural network)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM)

ซัพพอร์ตเวกเตอร์แมชชีน [11][12][13] เป็นตัวจำแนกประเภทข้อมูลแบบหนึ่งที่สร้างขึ้นเพื่อแก้ไขปัญหาของโครงข่ายประสาทเทียมแบบดั้งเดิม โดยอาศัยพื้นฐานความรู้ดั้งเดิมมาจากวิชาทางสถิติและกระบวนการลดความเสี่ยงเชิงโครงสร้างให้ต่ำที่สุด (Statistical Learning Theory and Structure Risk Minimization Principle) มาใช้เป็นกระบวนการเลือกแบบจำลองที่เหมาะสม โดยนำซัพพอร์ตเวกเตอร์แมชชีนไปใช้ในการจำแนกประเภทข้อมูลเป็น 2 กลุ่มที่อยู่ในปริภูมิคุณลักษณะ (Feature Space) โดยทำการสร้างระนาบในมิติสูง (Hyperplane) ที่อยู่กึ่งกลางของข้อมูลเพื่อแยกข้อมูลออกจากกัน เริ่มจากมีข้อมูลสำหรับใช้ในการสอนระบบ $\{x_i, d_i\}_{i=1}^N$ โดยที่ x_i คือข้อมูล และ d_i คือค่าเป้าหมาย จากนั้นทำการแบ่งค่าเป้าหมายเป็น 2 กลุ่มคือ $d_i = +1$ และ $d_i = -1$ ซัพพอร์ตเวกเตอร์แมชชีนจะทำการสร้าง Optimal Hyperplane ที่สามารถจำแนกข้อมูลทั้งสองกลุ่มออกจากกัน ดังสมการ 2.12

$$w^T x + b = 0 \quad (2.12)$$

โดยที่ x คือ เวกเตอร์ข้อมูล w คือ เวกเตอร์ค่าน้ำหนัก และ b คือ ไบแอส (Bias) ภายใต้งื่อนไข
ดังนี้

$$w^T x + b \geq 0 \text{ for } d_i = +1 \quad (2.13)$$

$$w^T x + b < 0 \text{ for } d_i = -1 \quad (2.14)$$

สำหรับเวกเตอร์น้ำหนัก w และไบแอส b เป็นค่าเดียวกันกับไฮเปอร์เพลนตามสมการ 2.12 และจุดข้อมูลที่สุดดคล้องกับสมการ 2.13 และ 2.14 เรียกว่า ซัพพอร์ตเวกเตอร์ โดยซัพพอร์ตเวกเตอร์จะถูกใช้สำหรับหาระยะห่างระหว่างขอบทั้งสองด้านของไฮเปอร์เพลนที่ทำให้ระยะห่างระหว่างขอบกว้างที่สุดซึ่งไฮเปอร์เพลนที่ได้จะอยู่กึ่งกลางระหว่างขอบทั้งสอง เรียกไฮเปอร์เพลนนี้ว่า ออพติมัล ไฮเปอร์เพลน (Optimal Hyperplane) รูป 2.13 แสดงให้เห็นถึงการสร้างออพติมัลไฮเปอร์เพลน (Optimal Hyperplane) สำหรับปริภูมิข้อมูลนำเข้าขนาดสองมิติ (two-dimension) โดย w_0 และ b_0 คือค่าเวกเตอร์น้ำหนักและไบแอสที่เหมาะสมที่สุด ดังนั้น ออพติมัลไฮเปอร์เพลนจะเป็นตัวที่ใช้ในการตัดสินใจแบ่งแยกข้อมูลแบบหลายมิติในปริภูมิข้อมูลนำเข้าดังสมการ 2.15

$$w_0^T x + b_0 = 0 \quad (2.15)$$

จากสมการที่ 2.15 เราสามารถเขียนฟังก์ชันการจำแนกได้ดังสมการที่ 2.16

$$g(x) = w_0^T x + b_0 \quad (2.16)$$

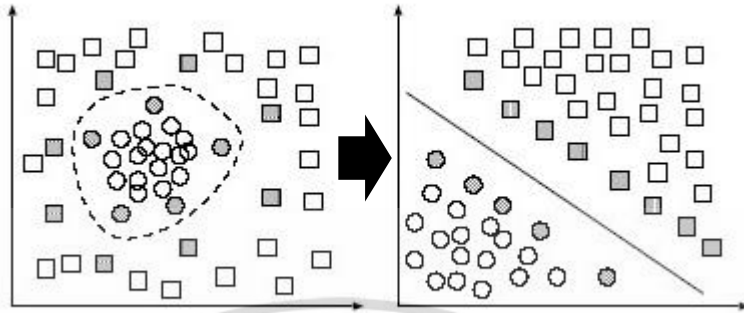


รูปที่ 2.13 ซีฟพอร์ตเวกเตอร์แมชชีน [11]

2.4.1 เคอร์เนล ฟังก์ชัน (Kernel function)

Kernel function เป็นฟังก์ชันที่ได้มาจากผลคูณภายใน (Inner product) ของฟังก์ชันที่ทำหน้าที่ในการแมพข้อมูลจากปริภูมิข้อมูลนำเข้า (Input Space) ไปยังปริภูมิคุณลักษณะ (Feature Space) เคอร์เนลฟังก์ชันจะช่วยให้การจำแนกประเภทข้อมูลทำให้การจำแนกประเภทข้อมูลมีประสิทธิภาพดียิ่งขึ้น ตัวอย่างของเคอร์เนลฟังก์ชัน เช่น Polynomial Function และ Gaussian Function ซึ่งมีลักษณะเป็นกราฟทรงระฆังคว่ำมาช่วยในการครอบกลุ่มข้อมูลเพื่อทำการแมพข้อมูล และแบ่งกลุ่มข้อมูลในปริภูมิคุณลักษณะ (Feature space) โดยใช้ไฮเปอร์เพลนในการแบ่งข้อมูลออกเป็นสองกลุ่ม ดังรูปที่ 2.14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.14 เคอเนล RBF (Kernel RBF) [12]

2.5 ตัวจำแนกแบบเบย์อย่างง่าย (Naïve Bayes Classifier)

ตัวจำแนกแบบเบย์อย่างง่าย [14] จะใช้หลักการของทฤษฎีความน่าจะเป็นเข้ามาช่วยในการทำงาน โดยทำการคำนวณหาค่าความน่าจะเป็นแบบมีเงื่อนไข กล่าวคือ การคำนวณหาความน่าจะเป็นของเหตุการณ์ใดเหตุการณ์หนึ่ง โดยใช้เงื่อนไขที่ว่าได้มีอีกเหตุการณ์หนึ่งเกิดขึ้นอยู่ก่อนแล้ว ซึ่งจะทำให้การเขียนเป็นสัญลักษณ์ทางคณิตศาสตร์เป็น $P(A|B)$ โดยที่ค่าของ A และ B แทนเหตุการณ์ที่เกิดขึ้นมีสูตรการคำนวณความน่าจะเป็นแบบมีเงื่อนไข ดังสมการที่ 2.16

$$P(A|B) = \frac{P(A \cap B)P(B|A)}{P(B)} \quad (2.17)$$

โดยที่ $P(A|B)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ A เมื่อเกิดเหตุการณ์ B ก่อน
 $P(A \cap B)$ คือ ความน่าจะเป็นของการเกิดเหตุการณ์ A และ B เกิดขึ้นร่วมกัน
 $P(B)$ คือ ค่าความน่าจะเป็นของการเกิดเหตุการณ์ B

การประยุกต์ใช้ทฤษฎีแบบเบย์สามารถปรับให้สอดคล้องกับการจำแนกประเภทข้อมูลได้ดังสมการที่ 2.18

$$P(\text{class}|\text{attribute}) = \frac{P(\text{class})P(\text{attribute}|\text{class})}{P(\text{attribute})} \quad (2.18)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ $P(\text{class}|\text{attribute})$ คือ ความน่าจะเป็นในการเกิดคลาส (class) เมื่อมีคุณลักษณะ (attribute) เกิดขึ้น

$P(\text{class})$ คือความน่าจะเป็นในการเกิดคลาส

$P(\text{attribute})$ คือความน่าจะเป็นในการเกิดคุณลักษณะ

จากสมการที่ 2.18 เมื่อข้อมูลแต่ละตัวอย่างมีจำนวนคุณลักษณะมากกว่าหนึ่งตัวจะได้สมการที่ 2.19 ดังนี้

$$P(\text{class}|a_1, a_2, \dots, a_n) = \frac{P(\text{class})P(a_1, a_2, \dots, a_n|\text{class})}{P(a_1, a_2, \dots, a_n)} \quad (2.19)$$

ในกรณีที่การเกิดคุณลักษณะ a_i ใดๆไม่ขึ้นตรงต่อกันสามารถคำนวณ $P(a_1, a_2, \dots, a_n|\text{class})$ ได้ดังสมการที่ 2.20

$$P(a_1, a_2, \dots, a_n|\text{class}) = \prod_{i=1}^n P(a_i|\text{class}) \quad (2.20)$$

โดยที่ $\prod_{i=1}^n P(a_i|\text{class}) = P(a_1|\text{class})P(a_2|\text{class}) \dots P(a_n|\text{class}) \quad (2.21)$

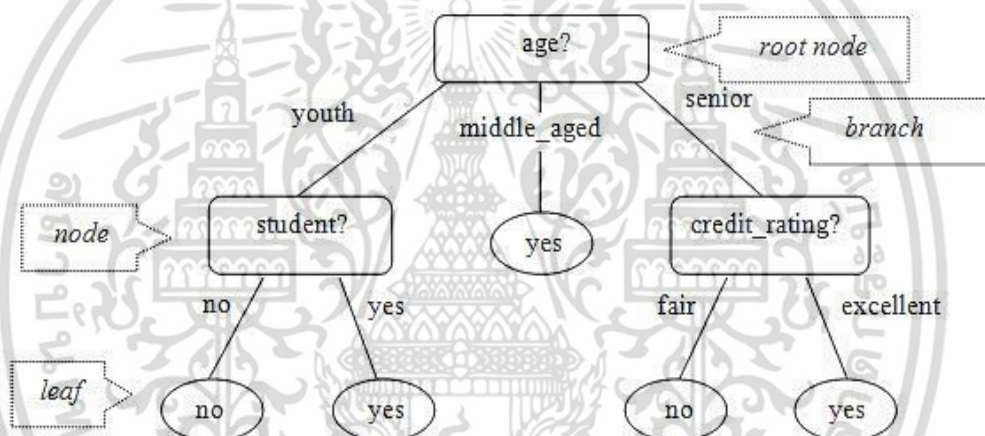
ดังนั้นการจำแนกแบบเบย์อย่างง่ายจึงสามารถทำได้โดยการหาค่าความน่าจะเป็น $P(a_1, a_2 \dots a_n|\text{class})$ ซึ่งขึ้นอยู่กับประเภทข้อมูลที่มีการเรียนรู้ในชุดข้อมูลนั้น เมื่อ c_j เป็นสมาชิกของ Class ต่างๆ โดยผลลัพธ์ของการจำแนกประเภทจะกำหนดคลาสที่มีค่าความน่าจะเป็นมากที่สุดให้กับตัวอย่างที่ต้องการทำนาย ดังสมการที่ 2.22

$$y = \underset{c_j \in C}{\text{argmax}} P(c_j) \prod_{i=1}^n P(a_i|c_j) \quad (2.22)$$

โดยที่ y คือผลลัพธ์ที่ได้จากตัวจำแนกแบบเบย์อย่างง่าย

2.6 การเรียนรู้ต้นไม้ตัดสินใจ (Decisions tree Learning)

ต้นไม้ตัดสินใจ (Decision Tree) [15] เป็นโครงสร้างข้อมูลชนิดแบบลำดับชั้น (hierarchy) โดยใช้ช่วยเหลือในการตัดสินใจ โดยจะมีลักษณะคล้ายต้นไม้กลับหัวที่มีรากอยู่ด้านบนบนสุดและมีใบอยู่ล่างสุดของต้นไม้ ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีคุณลักษณะ (attribute) เป็นตัวทดสอบ กิ่งของต้นไม้ (branch) จะแสดงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล (class) นั่นก็คือผลลัพธ์ที่ได้จากการทำนาย โดยโหนดที่อยู่บนสุดของต้นไม้เรียกว่า โหนดราก (root node) รูปที่ 2.15 แสดงโครงสร้างของต้นไม้ตัดสินใจ



รูปที่ 2.15 โครงสร้างต้นไม้ตัดสินใจ (Decision tree) [15]

จากรูปเป็นต้นไม้ที่ใช้ในการตัดสินใจว่าจะเลือกซื้อคอมพิวเตอร์หรือไม่ โดยมีคุณลักษณะที่พิจารณาคืออายุ (age) นักเรียน (student) และอัตราเครดิต (credit_rating) โดยที่โหนดสี่เหลี่ยมมุมโค้งจะเป็นการทดสอบคุณลักษณะของข้อมูล และสุดท้ายจะได้ผลลัพธ์ของการทำนายว่าจะซื้อคอมพิวเตอร์ (yes) หรือไม่ซื้อคอมพิวเตอร์ (no) จากการทดสอบตามเส้นทางของต้นไม้ตัดสินใจตั้งแต่โหนดรากไปจนถึงใบ การสร้างต้นไม้ตัดสินใจจะสร้างในลักษณะจากบนลงล่าง (top-down) นั่นก็คือ เริ่มจากการหาคุณลักษณะที่เหมาะสมที่สุด เพื่อนำมาเป็นรากของต้นไม้แล้วจึงแตกกิ่งไปจนถึงใบโดยขั้นตอนการสร้างต้นไม้ตัดสินใจจะมีดังนี้

Algorithms Generate_decision_tree.Generate a decision tree from the training tuples of data partition D .

Input:

- Data partition, D , which is a set of training tuples and tuples and their associated class labels;
- *attribute_list*, the set of candidate attributes;
- Attribute_selection_method, a procedure to determine the splitting criterion that “best” partitions the data tuples into individual classes. This criterion consists of a *splitting_attribute* and, possibly, either a split point or splitting subset.

Output: A decision tree

Method:

- (1) Create a node N ;
- (2) **if** tuples in D are all of the same class, C **then**
- (3) return N as a leaf node labeled with the class C ;
- (4) **if** *attribute_list* is empty **then**
- (5) return N as a leaf node labeled with the majority class in D ; //majority voting
- (6) apply Attribute_selection_method (D , *attribute_list*) to find the “best” *splitting_criterion*;
- (7) label node N with *splitting_criterion*;
- (8) **if** *splitting_attribute* is discrete-value **and**
 multiway splits allowed **then** // not restricted to binary trees
- (9) *attribute_list* \leftarrow *attribute_list* - *splitting_attribute*; // remove *splitting_attribute*
- (10) **for each** outcome j of *splitting_criterion*
 // partition the tuples and grow subtrees for partition
- (11) let D_j be the set of data tuples in D satisfying outcome j ; // a partition
- (12) **if** D_j empty **then**
- (13) attach a leaf labeled with the majority class in D to node N ;

รูปที่ 2.16 ขั้นตอนการสร้างต้นไม้ตัดสินใจ (Decision tree) [16]

- 1 เริ่มต้นสร้างโหนดขึ้นมาหนึ่งโหนด
- 2 ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกัน ให้โหนดที่สร้างขึ้นใหม่เป็นโหนดใบและ กำหนดค่าด้วยกลุ่มของข้อมูลนั้น เช่น กลุ่ม 1 เป็นต้น
- 3 ถ้าข้อมูลไม่มีคุณลักษณะใดที่เหมาะสมในการแบ่งกลุ่มให้โหนดที่สร้างขึ้นใหม่นั้นเป็นโหนดใบและกำหนดค่าด้วยกลุ่มที่มีข้อมูลสนับสนุนมากที่สุดหรือมีข้อมูลอ้างอิงมากที่สุด
- 4 ถ้าข้อมูลมีหลากหลายกลุ่มปะปนกันจะทำการเลือกคุณลักษณะที่มีความเหมาะสมที่สุด

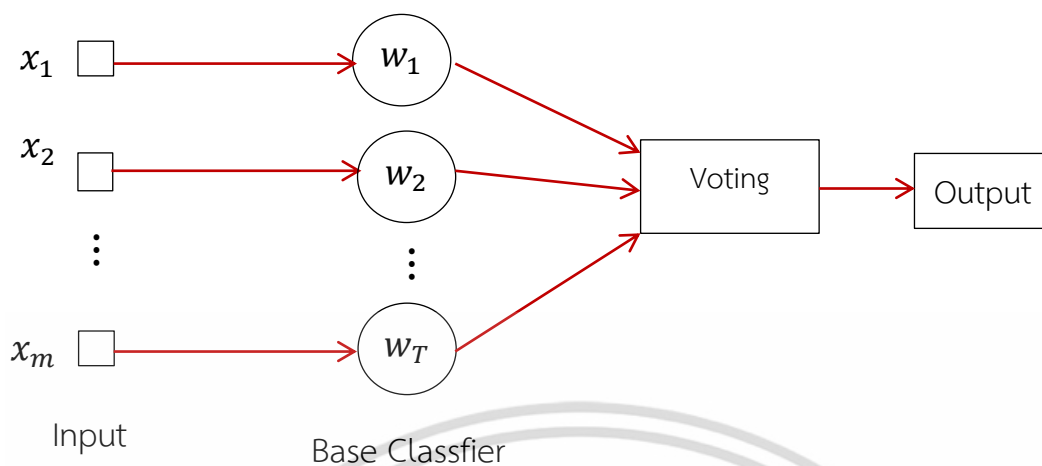
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นตัวทดสอบการตัดสินใจ โดยการวัดจากค่าเกน (gain) ของแต่ละคุณลักษณะ และกำหนดค่าให้โหนดที่สร้างขึ้นด้วยตัวทดสอบการตัดสินใจ

- 5 เมื่อได้ตัวทดสอบการตัดสินใจแล้ว หลังจากนั้นให้สร้างกิ่งของต้นไม้ด้วยค่าต่างๆ ที่เป็นไปได้ของตัวทดสอบ และแบ่งข้อมูลออกตามกิ่งต่างๆที่สร้างขึ้น
- 6 พิจารณาข้อมูลแต่ละกิ่ง หากพบว่าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันให้ต่อกิ่งด้วยโหนดใบ และกำหนดค่าด้วยกลุ่มของข้อมูลนั้น แต่ถ้าพบว่าข้อมูลมีหลากหลายกลุ่มปะปนกัน ให้ทำการวนซ้ำการหาตัวทดสอบการตัดสินใจที่เหมาะสมต่อไป
- 7 ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อยๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้เป็นจริง
 - ข้อมูลทั้งหมดในโหนดอยู่ในกลุ่มเดียวกันให้โหนดที่สร้างขึ้นนั้นเป็นโหนดใบและกำหนดค่าด้วยกลุ่มของข้อมูลนั้น
 - ไม่มีคุณลักษณะใดที่เหมาะสมในการแบ่งกลุ่ม ให้โหนดที่สร้างขึ้นนั้นเป็นโหนดใบ และกำหนดค่าด้วยกลุ่มที่มีข้อมูลสนับสนุนมากที่สุด

2.7 การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)

การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) [17][18] เป็นการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Learning) รูปแบบหนึ่ง โดยการเรียนรู้แบบรวมกลุ่มเป็นการนำโมเดลการจำแนกประเภทข้อมูล (Classification) ในรูปแบบต่างๆ มาทำงานร่วมกันจำนวนหลายโมเดล เพื่อช่วยเพิ่มประสิทธิภาพในตัวจำแนกประเภทข้อมูลให้มีประสิทธิภาพสูงขึ้น



รูปที่ 2.17 การเรียนรู้แบบรวมกลุ่ม

จากรูปที่ 2.17 การเรียนรู้แบบรวมกลุ่มแบ่งได้เป็น 4 ส่วนคือ ข้อมูลนำเข้า (Input) โมเดลการจำแนกประเภทข้อมูล (Base Classifier) อัลกอริทึมที่ใช้ในการโหวต (Voting) และข้อมูลส่งออก (Output) อธิบายดังนี้

1. ข้อมูลนำเข้า (Input) คือ ส่วนของข้อมูลสำหรับการสอน และทดสอบระบบ
2. โมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier) คือ ส่วนของอัลกอริทึมการจำแนกประเภทข้อมูล (Classification) ที่นำมาใช้ในการเรียนรู้ โดยจะทำการสร้างอัลกอริทึมเดียวกันจำนวนหลายๆ โมเดลเข้ามาช่วยในการเรียนรู้เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ให้ผลออกมาดียิ่งขึ้น ตัวอย่างของอัลกอริทึมที่ใช้ได้แก่ ต้นไม้ตัดสินใจ (Decision tree) การวิเคราะห์ตัวจำแนกประเภท (Discriminant Analysis) k -Nearest neighbor หรือแม้แต่ซัพพอร์ตเวกเตอร์แมชชีน ก็สามารถนำมาประยุกต์ใช้งานร่วมกับการเรียนรู้แบบรวมกลุ่มได้เช่นกัน
3. การโหวต (Voting) คือ อัลกอริทึมในการโหวต โดยจะทำการรวบรวมผลลัพธ์ที่ได้จากโมเดลการจำแนกประเภทแบบพื้นฐานทั้งหมดมาทำการโหวตหาผลลัพธ์
4. ข้อมูลส่งออก (Output) คือ ส่วนของผลลัพธ์ที่ได้จากการเรียนรู้แบบรวมกลุ่ม

การทำงานของ การเรียนรู้แบบรวมกลุ่ม เริ่มจากทำการเลือกหรือกำหนดโมเดลการจำแนกประเภทข้อมูล (Base Classifier) และอัลกอริทึมที่ใช้สร้างโมเดลแบบรวมกลุ่ม จากนั้นกำหนดจำนวนโมเดลการเรียนรู้ว่าจะใช้จำนวนเท่าไรในการเรียนรู้ เมื่อทำการกำหนดทุกอย่างเรียบร้อยแล้ว นำข้อมูลนำเข้ามาทำการ

สอนระบบ จากนั้นอัลกอริทึมที่ใช้สร้างโมเดลแบบรวมกลุ่มจะนำผลที่ได้จากโมเดลการเรียนรู้มาทำการปรับปรุงค่าการแจกแจงของข้อมูลนำเข้า ทำซ้ำจนครบจำนวนโมเดลที่กำหนด ในวิทยานิพนธ์นี้ประยุกต์นำเอาดาบูสต์มาใช้สำหรับสร้างโมเดลการเรียนรู้แบบรวมกลุ่มเพื่อจำแนกประเภทข้อมูลเสียง

2.7.1 อัลกอริทึม เอาดาบูสต์ (Adaboost Algorithm)

อัลกอริทึมเอาดาบูสต์ (Adaboost Algorithm) เป็นอัลกอริทึมที่ใช้สร้างโมเดลการเรียนรู้แบบรวมกลุ่ม ขั้นตอนการทำงานของการเรียนรู้แบบรวมกลุ่มโดยใช้อัลกอริทึมเอาดาบูสต์แสดงดังรูป 2.18

Adaboost Algorithm

Input: A weak learning algorithm *WeakLearn*, an integer T specifying number of iterations, and N training data $\{(x_1, y_1), \dots, (x_N, y_N)\}$.

Output: A strong classifier F

1. Initially the weight vector $w_i^1 = \frac{1}{N}$, for $i = 1, \dots, N$.
2. For $t \leftarrow 1, 2, \dots, T$ do
3. Call *WeakLearn*, providing it with the distribution on w^t ; get back a weak Learner $h_t : X \rightarrow \pm 1$.
4. Calculate the weight error of h_t :

$$\varepsilon_t = \sum_{i=1}^N w_i^t I [h_t(x_i) \neq y_i]$$
5. $\alpha_t = \frac{1}{2} \ln \left(\frac{1-\varepsilon_t}{\varepsilon_t} \right)$
6. Update weight $w_i^{t+1} \leftarrow w_i^t e^{(-\alpha_t h_t(x_i) y_i)}$, for $i = 1, 2, \dots, N$.
7. Normalize weight $w_i^t \leftarrow w_i^t / \sum_{i=1}^N w_i^t$
8. Output the final strong classifier

$$F(x) = \begin{cases} 1, & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq 0, \\ -1, & \text{otherwise} \end{cases}$$

รูปที่ 2.18 อัลกอริทึม เอาดาบูสต์ (Adaboost Algorithm) [17]

จากรูปที่ 2.18 การเรียนรู้แบบรวมกลุ่มโดยใช้อัลกอริทึมเอดาบู้สต์จะมีการใช้ตัวแปรต่างๆ มากมายเพื่อใช้ในการคำนวณหาผลลัพธ์ออกมา ตัวแปรต่างๆ มีความหมายดังนี้

1. ค่า T คือ จำนวนโมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier) ทั้งหมด
2. ค่า N คือ จำนวนข้อมูลทั้งหมด
3. ค่า x คือ ข้อมูลนำเข้า
4. ค่า y คือ ผลลัพธ์ที่ต้องการ (Target)
5. ค่า F คือ ผลลัพธ์ทั้งหมดของโมเดลการเรียนรู้แบบรวมกลุ่ม
6. ค่า w คือ ค่าน้ำหนัก (Weight)
7. ค่า i คือ ลำดับของข้อมูล
8. ค่า t คือ ลำดับของโมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier)
9. ค่า w^t คือ ค่าการแจกแจง (Distribution) หรือ ค่าความน่าจะเป็น โดยค่านี้จะเป็นค่าที่แสดงความเป็นไปได้ที่ข้อมูลจะถูกเลือกเพื่อมาทำการสอนโมเดลการจำแนกประเภท ถ้าข้อมูลที่ใช้สอนมีจำนวน N ข้อมูล ความน่าจะเป็นที่จะเลือกข้อมูลแต่ละตัวเข้ามาทำการสอนเริ่มต้นจะเท่ากับ $1/N$ ค่าการแจกแจง หรือค่าความน่าจะเป็นนี้จะใช้ในการคำนวณเพื่อการปรับปรุงการสอนระบบให้ดีขึ้น เมื่อทำการปรับค่าการแจกแจงแล้ว ถ้าข้อมูลที่มีการจัดกลุ่มผิดพลาด หรือยากต่อการจำแนกประเภทข้อมูลก็จะปรับค่าการแจกแจงเพื่อให้ข้อมูลมีโอกาสถูกเลือกมากขึ้นเพื่อนำไปสร้างโมเดลการเรียนรู้โมเดลใหม่
10. ค่า h_t คือ ผลลัพธ์ของโมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier) ซึ่งจะบอกว่าข้อมูลที่เข้ามาอยู่ประเภทอะไร
11. ค่า ϵ_k คือ ค่าความผิดพลาด (Error rate)
12. α_t คือ ค่าสัมประสิทธิ์ความผิดพลาด
13. \exp คือ Exponential Function ใช้ในการช่วยปรับค่าแจกแจงของข้อมูล ทำให้ค่าที่ออกมา มีความแตกต่างกันค่อนข้างชัดเจน
14. I คือ ฟังก์ชันที่มีเงื่อนไขว่า ถ้าค่าที่ออกมาตรงตามเงื่อนไขของฟังก์ชันให้ค่าเป็น 1 ถ้าไม่ตรงตามเงื่อนไขให้ค่าเป็น 0

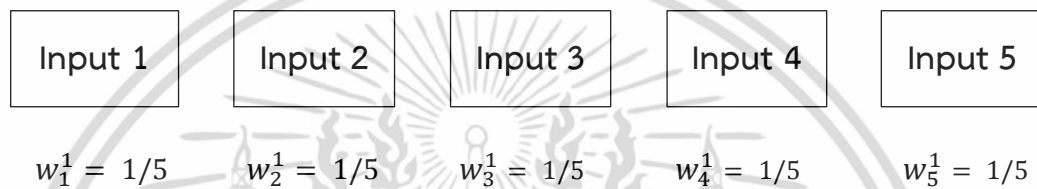
การทำงานของการเรียนรู้แบบรวมกลุ่มโดยใช้เอดาบู้สต์ สามารถอธิบายได้ดังนี้

- 1.) จากรูปที่ 2.20 เริ่มจากขั้นตอนที่ 1 ทำการแทนค่าการแจกแจงเริ่มต้น

$$w_i^1 = \frac{1}{N}, \text{ for } i = 1, \dots, N \quad (2.23)$$

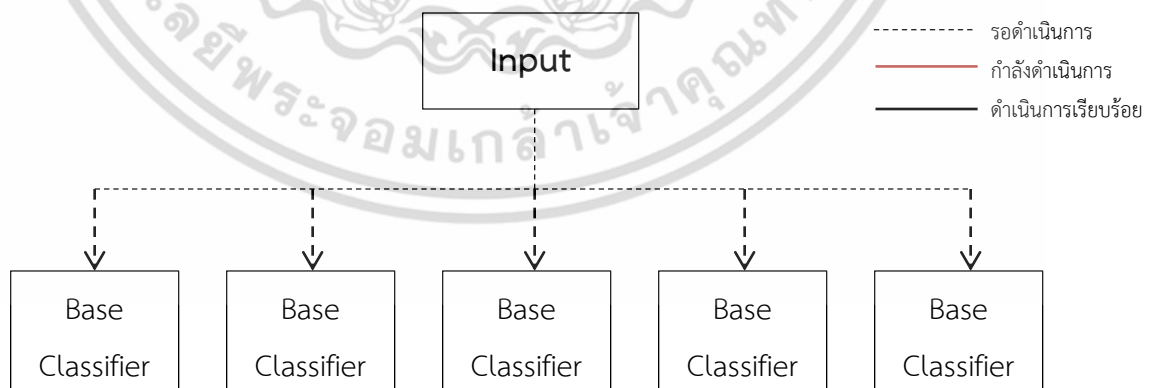
โดยที่ค่า N คือ จำนวนข้อมูลทั้งหมด

ในที่นี้จะกำหนดค่าการแจกแจง (Distribution) เริ่มต้นให้กับข้อมูลทุกตัวที่ใช้ในการสอนระบบ ด้วยค่า $w_i^1 = \frac{1}{N}$ ตัวอย่างเช่น มีข้อมูลที่ใช้ในการสอนระบบทั้งหมด $N = 5$ ดังนั้นค่าของ $w_i^1 = 1/5$ ซึ่งจะถูกใช้เป็นค่าเริ่มต้น โดยจะกำหนดค่านี้ให้กับข้อมูลที่ใช้ในการสอนทุกข้อมูลดังรูป 2.19



รูปที่ 2.19 การกำหนดค่า Weight (w)

2) จากนั้นกำหนดโมเดลการเรียนรู้ที่จะใช้ในการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) โดยโมเดลการเรียนรู้ช่วยทำงานมีหลายรูปแบบเช่น ต้นไม้ตัดสินใจ (Decision Tree) การวิเคราะห์ตัวจำแนกประเภท (Discriminant Analysis) เป็นต้น เมื่อกำหนดรูปแบบโมเดลการเรียนรู้เรียบร้อยแล้ว จากนั้นทำการกำหนดจำนวนโมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier)

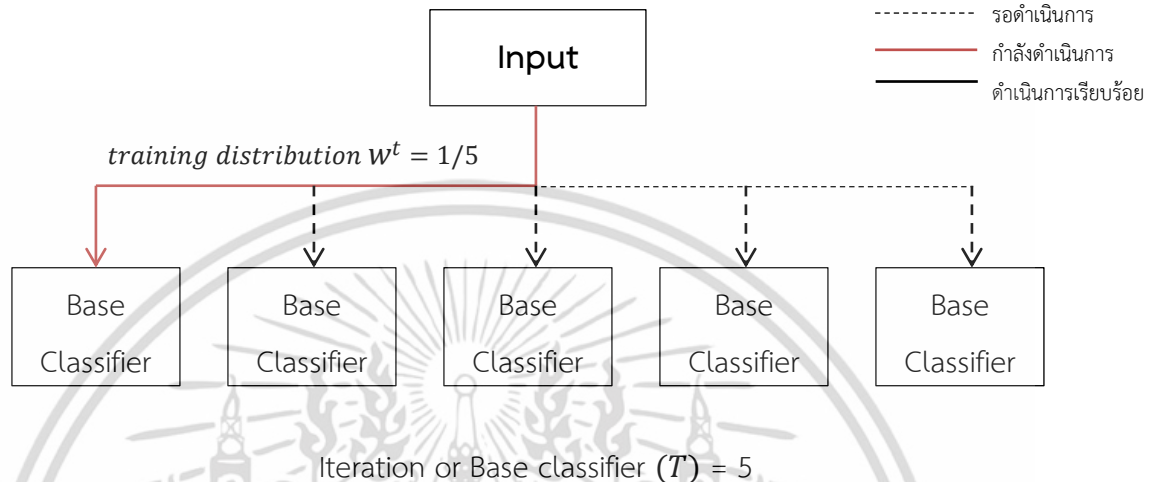


Iteration or Base classifier (T) = 5

รูปที่ 2.20 สร้างโมเดลในการเรียนรู้แบบกลุ่มตามค่า Iteration (T) ที่กำหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การเรียนรู้แบบรวมกลุ่มจะเริ่มทำการสอนโมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier) โมเดลแรก ($t = 1$) โดยใช้ค่าการแจกแจง (Distribution) เริ่มต้น ดังรูปที่ 2.21



รูปที่ 2.21 เริ่มต้นสอนโมเดลการจำแนกประเภทแบบพื้นฐาน

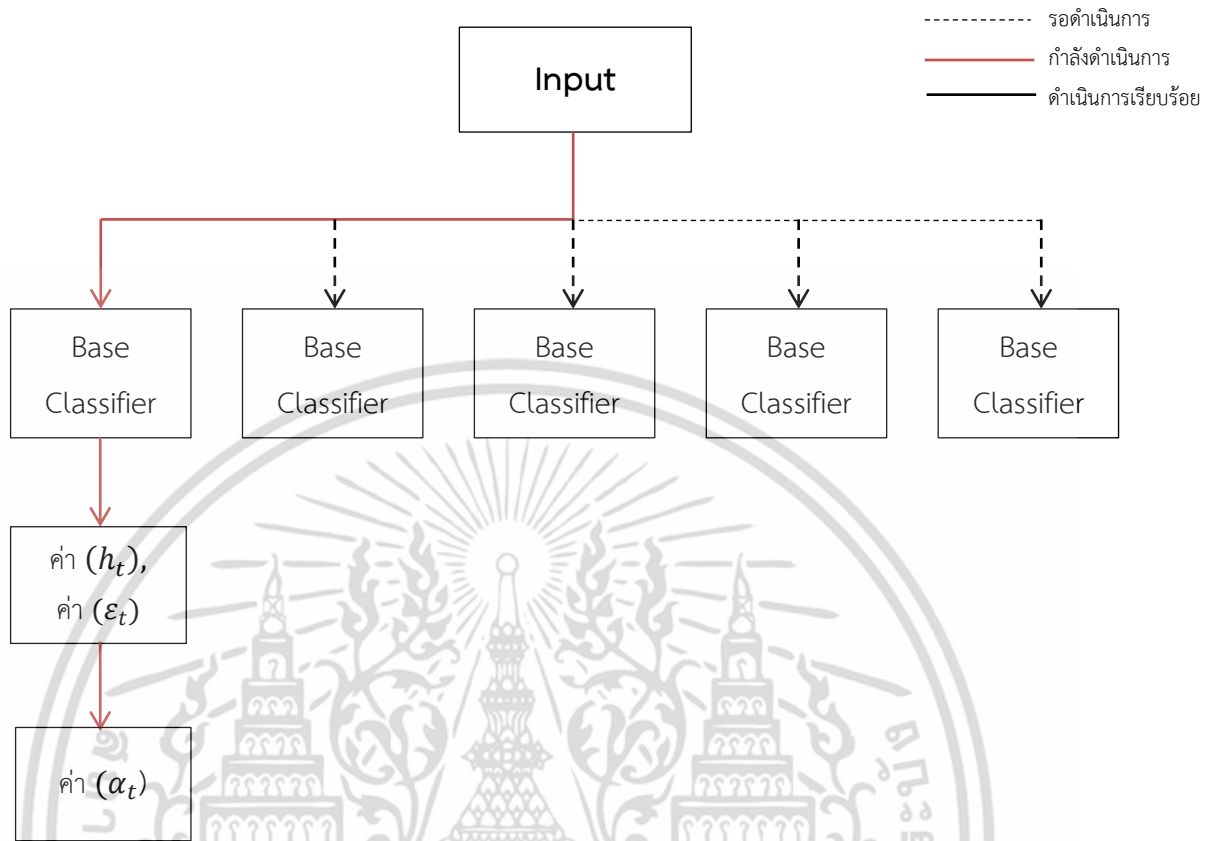
4) เมื่อสร้างโมเดลการจำแนกประเภทแบบพื้นฐาน (Base classifier) เสร็จเรียบร้อยแล้ว จะได้ผลลัพธ์การเรียนรู้ของโมเดลการจำแนกประเภทแบบพื้นฐาน (h_t) จากนั้นทำการคำนวณหาค่าความผิดพลาด (ε_t) ที่เกิดขึ้นในโมเดลแรก โดยใช้สมการ 2.24

$$\varepsilon_t = \sum_{i=1}^N w_i^t I [h_t(x_i) \neq y_i] \quad (2.24)$$

5) นำค่าความผิดพลาด (ε_t) มาคำนวณหาค่าสัมประสิทธิ์ความผิดพลาด (α_t) โดยใช้สมการ

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t} \quad (2.25)$$

จากสมการ 2.25 เมื่อทำการคำนวณจากสมการ 2.24 ผลที่ออกมาคือ ค่าความผิดพลาด จากนั้นคำนวณหาค่าสัมประสิทธิ์ความผิดพลาด (α_t) ซึ่งเป็นค่าที่บอกว่า ผลการเรียนรู้ที่ออกมา มีความถูกต้องมากหรือถูกต้องน้อย ดังรูปที่ 2.22



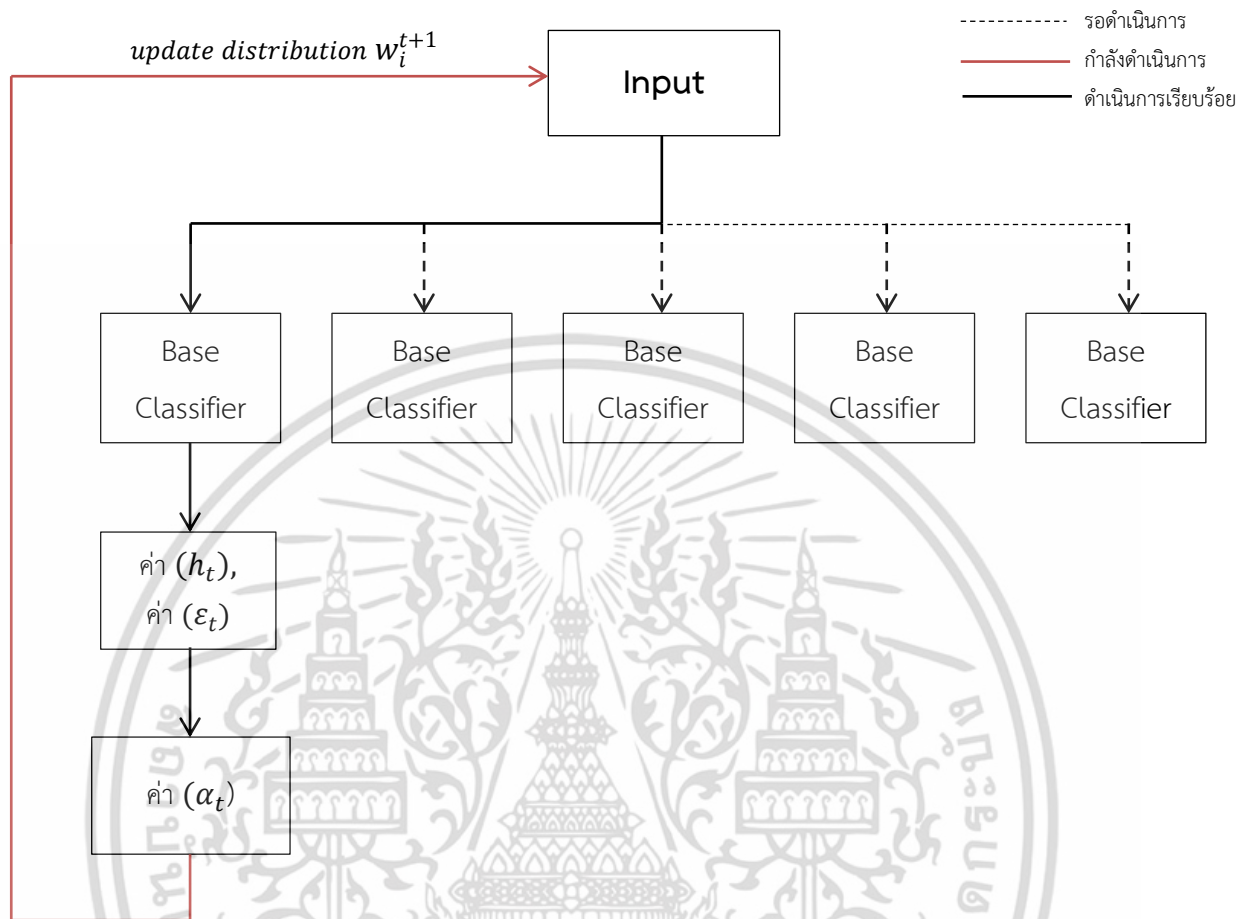
รูปที่ 2.22 ค่าสัมประสิทธิ์ความผิดพลาด (α_t)

6) เมื่อทำการคำนวณหาค่าสัมประสิทธิ์ความผิดพลาด (α_t) เรียบร้อยแล้วขั้นตอนต่อไปคือทำการคำนวณเพื่อปรับค่าน้ำหนัก (Weight) ใหม่ โดยนำค่าสัมประสิทธิ์ความผิดพลาด (α_t) มาคำนวณหาค่าน้ำหนักใหม่จากสมการ 2.26

$$w_i^{t+1} \leftarrow w_i^t e^{-\alpha_t h_t(x_i) y_i} \quad (2.26)$$

สมการ 2.26 หาค่าน้ำหนัก (w_i^{t+1}) โดยนำค่าผลลัพธ์การเรียนรู้ของโมเดลการจำแนกประเภทแบบพื้นฐาน (h_t) คูณกับค่าเป้าหมาย (y_i) คูณกับค่าสัมประสิทธิ์ความผิดพลาด (α_t) แล้วใช้ฟังก์ชัน Exponential เพื่อทำให้ค่าที่ออกมามีความแตกต่างกันอย่างเห็นได้ชัด คือ ถ้าค่าออกมาสูงกราฟก็จะขึ้นสูงแบบเห็นได้ชัด ถ้าค่าน้อยกราฟก็จะลงต่ำมากอย่างเห็นได้ชัดเช่นกัน เมื่อได้ค่าน้ำหนัก (Weight) ใหม่สำหรับนำไปใช้ในการเรียนรู้และทดสอบโมเดลการจำแนกประเภทแบบพื้นฐานถัดไป ดังรูป 2.23

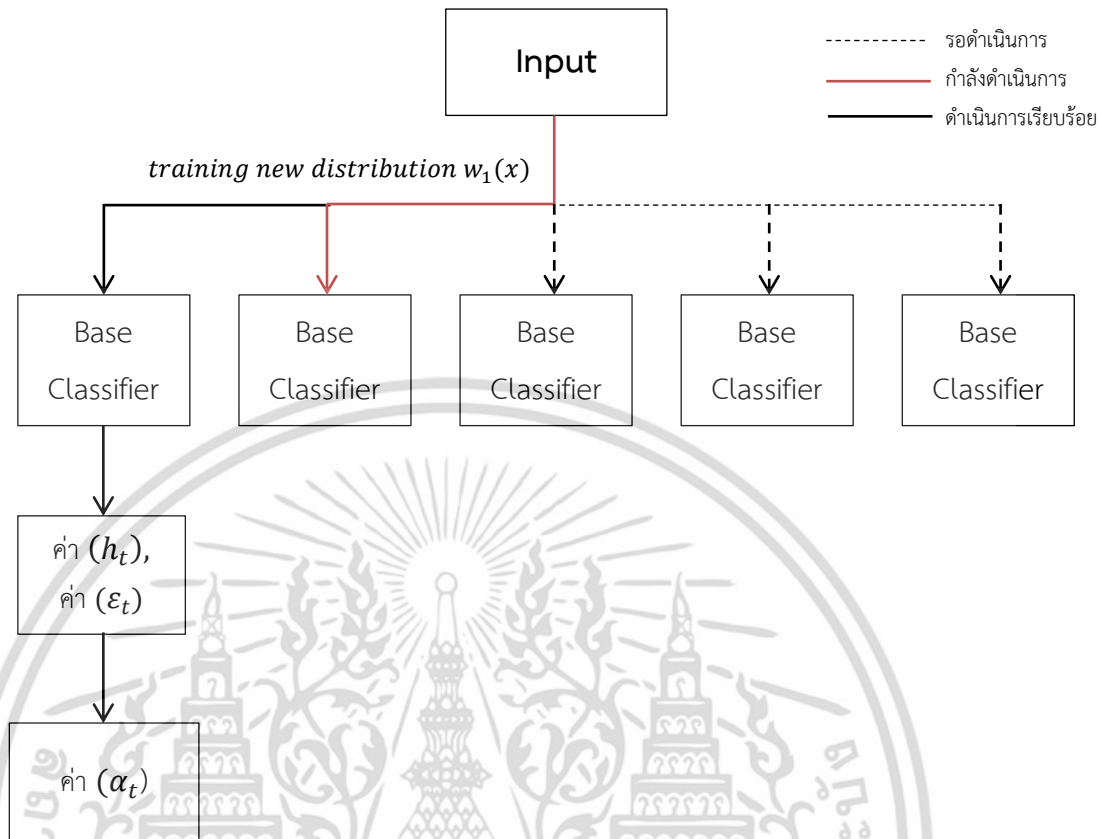
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.23 ปรับปรุงค่าน้ำหนัก (Weight)

เมื่อทำการปรับปรุงค่าน้ำหนัก (Weight) เรียบร้อย ก็จะทำให้การนำข้อมูลที่ปรับปรุงค่าน้ำหนักแล้ว มาทำการสอนโมเดลการจำแนกประเภทแบบพื้นฐานถัดไปดังรูป 2.24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.24 สอนโมเดลการจำแนกประเภทแบบพื้นฐานถัดไป

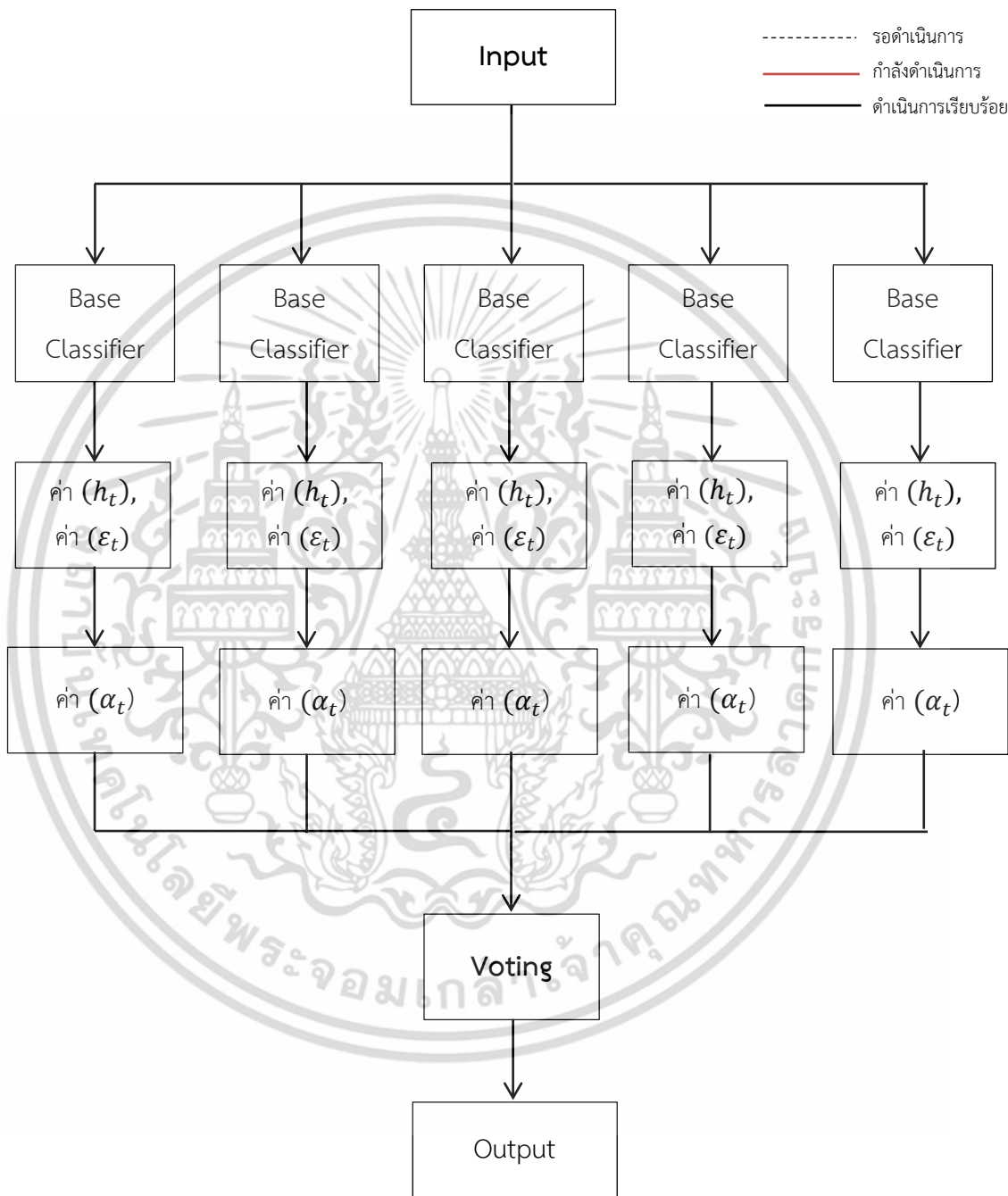
ทำการสอนโมเดลการจำแนกประเภทแบบพื้นฐานที่ 2 โดยทำตามขั้นตอนเดียวกับโมเดลการจำแนกประเภทแบบพื้นฐานแรกทำจนครบตามจำนวนโมเดลการจำแนกประเภทแบบพื้นฐานที่กำหนด

7) เมื่อทำการสร้างโมเดลการจำแนกประเภทแบบพื้นฐาน (Base Classifier) ตามค่าที่กำหนดขึ้นมาเรียบร้อยแล้ว จากนั้นทำการหาผลลัพธ์ (Output) ของการเรียนรู้แบบรวมกลุ่ม โดยใช้สมการ 2.27

$$F(x) = \begin{cases} 1 & , \text{ if } \sum_{t=1}^T \alpha_t h_t(x) \geq 0, \\ -1 & , \text{ otherwise} \end{cases} \quad (2.27)$$

จากสมการที่ 2.27 ทำการหาผลลัพธ์ของการเรียนรู้แบบรวมกลุ่ม ($F(x)$) โดยนำค่าผลลัพธ์การเรียนรู้ของโมเดลการจำแนกประเภทแบบพื้นฐาน (h_t) และค่าสัมประสิทธิ์ความผิดพลาด (α_t) มาคูณกัน

เพื่อหาผลลัพธ์ และทำการตรวจเงื่อนไขโดยถ้าผลลัพธ์ที่ออกมามีค่ามากกว่าเท่ากับ 0 การจำแนกประเภทได้กลุ่ม 1 แต่ถ้าไม่ตรงเงื่อนไขแรกการจำแนกประเภทได้กลุ่ม -1



รูปที่ 2.25 ขั้นตอนการทำงานของ การเรียนรู้แบบรวมกลุ่มโดยใช้เอตาบวสต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8 งานวิจัยที่เกี่ยวข้อง

Ensemble Data Classification Based on Decision Tree, Artificial Neuron network and Support Vector Machine Optimized by Genetic Algorithm [19]

งานวิจัยนี้นำเสนอการใช้โมเดลการจำแนกประเภทข้อมูลมาช่วยในการจัดกลุ่มข้อมูลของการล้มละลาย โดยการใช้การจำแนกประเภทข้อมูลหลากหลายประเภท และในงานวิจัยนี้ได้ประยุกต์ใช้การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ที่ใช้โมเดลการจำแนกประเภทข้อมูลแบบต้นไม้ตัดสินใจ ทำการเปรียบเทียบผลประสิทธิภาพความถูกต้องกับการจำแนกประเภทข้อมูลอีกสองรูปแบบคือ โครงข่ายประสาทเทียม (Artificial Neuron network) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector machine) สรุปผลการทดลองออกมาว่า การเรียนรู้แบบรวมกลุ่ม ทำงานมีประสิทธิภาพกว่าอีกสองแบบ ผลประสิทธิภาพความถูกต้องดีที่สุดจากงานวิจัยนี้ได้ทำให้มีความคิดที่จะนำการจำแนกประเภทข้อมูลมาช่วยในการจำแนกประเภทข้อมูลเสียงและทำการเปรียบเทียบประสิทธิภาพเพื่อหาการจำแนกประเภทข้อมูลที่มีประสิทธิภาพในการจัดประเภทของข้อมูลเสียงที่ดีที่สุด และให้ผลประสิทธิภาพที่สูงที่สุด

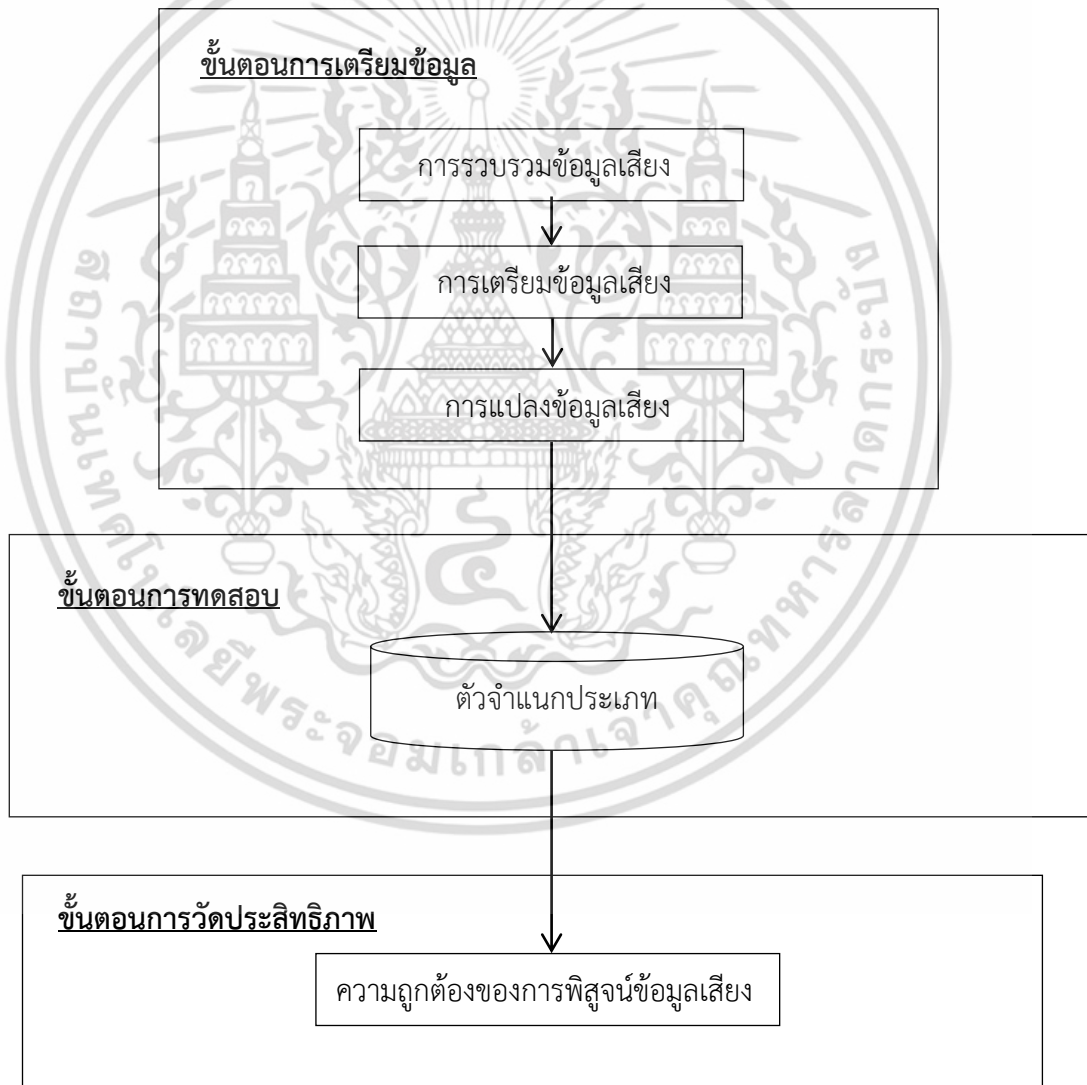
A Visual Activity Detection Method with Adaboosting [20]

งานวิจัยนี้แนะนำการนำ การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) มาประยุกต์ใช้ในการเดาคำพูดจากริมฝีปากหรือการอ่านริมฝีปาก และทำการเปรียบเทียบระหว่างเสียงกับการขยับริมฝีปากว่า มีเนื้อหาข้อมูลที่ตรงกันหรือไม่ เริ่มจากทำการสอนระบบด้วยข้อมูลเสียงที่ไม่มีข้อมูลรบกวนเข้าสู่การเรียนรู้แบบรวมกลุ่ม ข้อมูลอีกชุดหนึ่งจะเป็นข้อมูลวิดีโอที่มีคนพูด จากนั้นทำการกำหนดคลาสในการสอนข้อมูลให้ตรงกัน ทั้งข้อมูลเสียงและวิดีโอ เพื่อทำการสอนการเรียนรู้แบบรวมกลุ่ม ทำการทดสอบโดยการใช้อudioในการทดสอบระบบ โดยเพิ่มในส่วนของคุณลักษณะเสมือน (Virtual Feature) เข้ามาช่วยทำการอ่านริมฝีปากในวิดีโอ คือ ทำการวัดองศาของการขยับปาก เพื่อทำการแปลความหมาย รูปแบบปากขยับลักษณะแบบนี้ องศาของมุมปากและริมฝีปากขยับองศาเท่าไร พูดคำว่าอะไร ทำการเก็บค่าองศาของการขยับปากครบถ้วน นำค่านี้มาทำการทดสอบระบบคู่กับข้อมูลวิดีโอ จากผลการทดลองการใช้คุณลักษณะเสมือน (Virtual Feature) เข้ามาทำงานร่วมกับการเรียนรู้แบบรวมกลุ่ม ทำให้ค่าความผิดพลาด (Error rate) ในการทดสอบลดลงจากการทดสอบข้อมูลวิดีโอโดยใช้การเรียนรู้แบบแบบปกติ

บทที่ 3

โมเดลการพิสูจน์ข้อมูลเสียง

ในบทนี้กล่าวถึงขั้นตอนการสร้างโมเดลพิสูจน์ข้อมูลเสียงว่า ข้อมูลเสียงไหนเป็นข้อมูลเสียงติดต่อหรือข้อมูลเสียงไม่ติดต่อ ประกอบด้วย 4 ขั้นตอน ได้แก่ ขั้นตอนการเตรียมข้อมูล ขั้นตอนการทดสอบโมเดลการจำแนกประเภทหลายโมเดล ขั้นตอนการทดสอบโดยใช้การคัดเลือกคุณลักษณะร่วมกับโมเดลการจำแนกประเภทหลากหลายรูปแบบ การทดสอบโมเดลการจำแนกประเภทเพื่อวัดประสิทธิภาพ โดยการทำงานภาพรวมของการสร้างโมเดลพิสูจน์ข้อมูลเสียงแสดงได้ดังรูปที่ 3.1

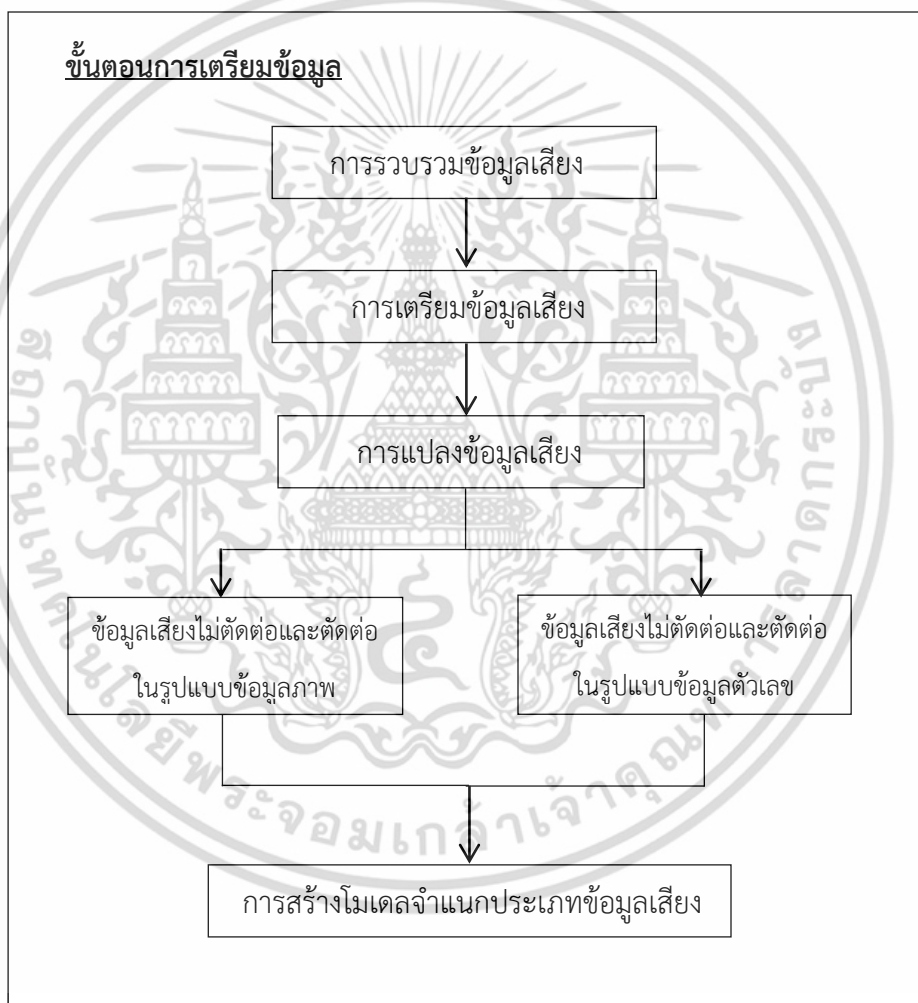


รูปที่ 3.1 ขั้นตอนการทำงานของโมเดลพิสูจน์ข้อมูลเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1 ขั้นตอนการเตรียมข้อมูล

หัวข้อนี้กล่าวถึงการเตรียมข้อมูล ประเภทของข้อมูลเสียง และการแปลงข้อมูลเสียงให้อยู่ในรูปแบบตัวเลข (Numerical Format) เพื่อนำไปสร้างโมเดลการเรียนรู้ และการทดสอบโมเดล ขั้นตอนการเตรียมข้อมูลแบ่งเป็น 3 ขั้นตอน ได้แก่ การรวบรวมข้อมูลเสียง (Data collection) การเตรียมข้อมูลเสียง การแปลงข้อมูลเสียง (Data Transformation) โดยขั้นตอนของการเตรียมข้อมูลแสดงดังรูป 3.2



รูปที่ 3.2 ขั้นตอนการเตรียมข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 การรวบรวมข้อมูลเสียง

ในขั้นตอนการรวบรวมข้อมูลเสียงได้ขอความอนุเคราะห์ข้อมูลเสียงพูดภาษาไทย (Thai speech corpus) จากห้องปฏิบัติการวิจัยเทคโนโลยีเสียง ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (Speech and Audio Technology Laboratory , National Electronics and Computer Technology Center) ข้อมูลเสียงสกุลอยู่ในรูปแบบไฟล์ .wav ซึ่งมีความละเอียดเสียงสูง และไม่มีการตัดแปลง ลดขนาดของข้อมูลเสียง หรือความละเอียดของข้อมูลเสียง (Bit rate) ทำให้มีคุณภาพ และความชัดเจนของเสียงที่ดี โดยภายในข้อมูลเสียงมีเนื้อหาเป็นเสียงผู้หญิง จำนวนหนึ่งคน พูดภาษาไทยในภาวะห้องที่เงียบสนิทไม่มีเสียงรบกวนจากภายนอกและรอบข้าง ข้อมูลที่ได้รับเป็นข้อมูลเสียงที่มีความยาวของข้อมูลไม่แน่นอน เช่น 5, 10, 15 นาที เป็นต้น

3.1.2 การเตรียมข้อมูลเสียง

ข้อมูลเสียงที่ได้มานั้นมีความยาวของเสียงในแต่ละข้อมูล หรือเวลาภายในข้อมูลเสียง ที่ไม่เท่ากัน จึงทำการจัดการให้ความยาวของข้อมูลเสียงมีความยาวเท่ากันก่อน ถ้าเกิดไม่ทำให้ข้อมูลเสียงมีเวลาที่เท่ากันจะทำให้เกิดปัญหาเรื่องการนำเข้าข้อมูล เพราะต้องทำการแปลงข้อมูลเสียงให้อยู่ในรูปแบบข้อมูลตัวเลข (Numerical Format) ให้ความยาวของข้อมูลตัวเลขมีความยาวที่ไม่เท่ากันเช่นกัน ในวิทยานิพนธ์นี้ นำโปรแกรม Adobe Audition CS มาช่วยจัดการข้อมูลเสียงให้มีความยาวเท่ากัน และได้ใช้โปรแกรม Adobe Audition CS ในการเตรียมข้อมูลเสียงที่มีการตัดต่อเช่นเดียวกัน รูปที่ 3.3 คือ ตัวอย่างการนำข้อมูลเสียงเข้ามาภายในโปรแกรม Adobe Audition CS



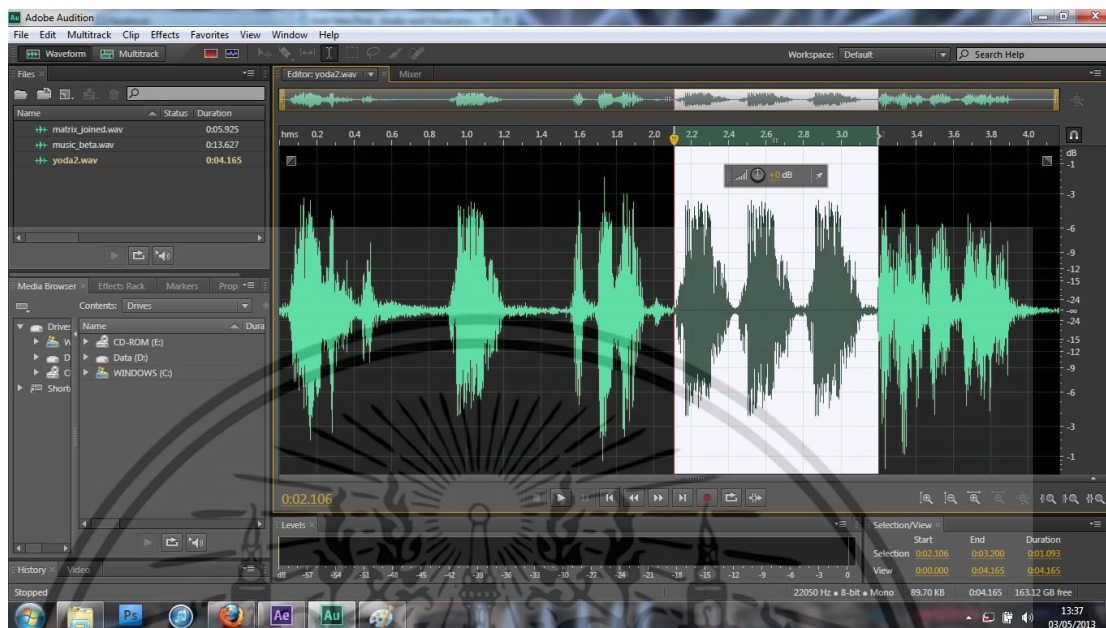
รูปที่ 3.3 การนำข้อมูลเสียงเข้าโปรแกรม Adobe Audition

จากนั้นทำการลดข้อมูลเสียงให้เหลือข้อมูลละ 3 วินาที เพราะข้อมูลเสียงความยาวเกิน 3 วินาที เมื่อทำการแปลงเป็นข้อมูลตัวเลขจะมีความยาวของจำนวนคอลัมน์ (Column) จำนวนมาก ทำให้เครื่องมือในการทดลองใช้เวลานาน เกิดความล่าช้า เพิ่มภาระของเครื่องมือในการทดลอง จึงลดความยาวของข้อมูลเสียงเหลือ 3 วินาที ตัวอย่างเช่น 1 นาที จำนวนคอลัมน์ราว 100,000 คอลัมน์ เพื่อลดภาระของเครื่องมือที่ใช้ในการทดสอบ จึงทำการลดข้อมูลเสียงเหลือ 3 วินาที คอลัมน์ที่เหลือมีราว 15,000 คอลัมน์ เราได้ทำการทดลองข้อมูลเสียงที่มีความยาวมากกว่า 3 วินาที ให้ผลใกล้เคียงกันกับข้อมูลเสียง 3 วินาทีจึงเลือกข้อมูลเสียง 3 วินาทีมาใช้ในการสอนและทดสอบ เมื่อทำการลดข้อมูลเสียงและทำการบันทึกด้วยความละเอียด 6000 Hz ก็จะได้ข้อมูลเสียงจำนวน 1 ข้อมูลเสียง ทำอย่างวิธีนี้จึงได้ข้อมูลเสียงครบ จำนวน 450 ข้อมูลเสียง คือ ใช้ในการสอนระบบ 300 ข้อมูล และทดสอบ 150 ข้อมูล การสอนและทดสอบจะใช้หลักการ Holdout Method ซึ่งเป็นวิธีในเทคนิคการเรียนรู้ของเครื่องในการสอนและทดสอบการเรียนรู้ของเครื่อง

เมื่อได้ข้อมูลเสียงที่ไม่มีการตัดต่อครบทั้ง 450 ข้อมูลเสียง ขั้นตอนถัดมาทำการเตรียมข้อมูลเสียงที่มีการตัดต่อโดยข้อมูลเสียงที่มีการตัดต่อนั้น ใช้วิธีการเตรียมโดยการนำข้อมูลเสียงอีกข้อมูลหนึ่งที่มีความยาว 1 วินาที มาทำการแทนที่ข้อมูลเสียงที่ไม่มีการตัดต่อโดยการสุ่มตำแหน่งในการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แทนที่แต่ละข้อมูลเสียงแบบไม่แน่นอน โดยความยาวหลังจากการแทนที่แล้วจะมีความยาว 3 วินาทีเท่าเดิม ถ้าลองเปิดฟังดูเนื้อหาที่พูดในข้อมูลเสียงจะเปลี่ยนไป



รูปที่ 3.4 การแทนที่ข้อมูลเสียง

จากรูปที่ 3.4 เป็นตัวอย่างการแทนที่ของข้อมูลเสียง ในส่วนของแถบสีขาวคือ ส่วนของข้อมูลเสียงที่ทำการแทนที่ข้อมูลเสียงของเดิมลงไป ทาวิธีนี้จนครบจำนวน 450 ข้อมูลเสียง เป็นข้อมูลเสียงที่มีการตัดต่อในการสอนระบบ 300 ข้อมูลเสียง และใช้ในการทดสอบระบบ 150 ข้อมูลเสียง เช่นเดียวกัน เท่ากับว่า เตรียมข้อมูลที่ใช้ในการสอนระบบทั้งข้อมูลเสียงที่มีการตัดต่อและข้อมูลเสียงที่ไม่มีการตัดต่อ จำนวนทั้งหมด 600 ข้อมูลเสียง และข้อมูลเสียงที่ใช้ในการทดสอบระบบจำนวนทั้งหมด 300 ข้อมูลเสียง

3.1.3 การแปลงข้อมูลเสียงเป็นข้อมูลตัวเลข

วิทยานิพนธ์นี้เลือกใช้โปรแกรม Matlab เวอร์ชัน 2014a ในการสร้างโมเดลในการทดสอบระบบ เนื่องจากในโปรแกรม Matlab มีเครื่องมือการเรียนรู้ของเครื่อง (Machine learning) อยู่มากและเครื่องมือใน Matlab มีการรับข้อมูลเป็นข้อมูลตัวเลขในรูปแบบเวกเตอร์เท่านั้น ดังนั้น จึงต้องทำการแปลงข้อมูลจากข้อมูลคลื่นเสียงให้อยู่ในรูปแบบของข้อมูลตัวเลข (Numerical Format) ในรูปแบบเวกเตอร์ก่อน โดยทำการอ่านไฟล์ข้อมูลเสียงเข้ามาในโปรแกรม Matlab ก่อน จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลตัวเลข (Numerical Format) โดย 1 ไฟล์ข้อมูลเสียงจะได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เท่ากับข้อมูล 1 แถวมีความยาว 15,000 คอลัมน์และบันทึกไฟล์นามสกุล .csv โดยข้อมูลที่ได้จะเป็นข้อมูลแบบตารางตัวเลขนามสกุลไฟล์ .csv ซึ่งจะทำการแปลงเป็น CSV ทั้งหมดสี่ไฟล์ ได้แก่

1. ไฟล์ข้อมูลเสียงตัดต่อเพื่อใช้ในการสอนระบบ (ขนาดตารางเท่ากับ 300x15000)
2. ไฟล์ข้อมูลเสียงไม่ตัดต่อเพื่อใช้ในการสอนระบบ (ขนาดตารางเท่ากับ 300x15000)
3. ไฟล์ข้อมูลเสียงตัดต่อเพื่อใช้ในการทดสอบระบบ (ขนาดตารางเท่ากับ 150x15000)
4. ไฟล์ข้อมูลเสียงไม่ตัดต่อเพื่อใช้ในการทดสอบระบบ (ขนาดตารางเท่ากับ 150x15000)

ในรูป 3.5 จะเป็นตัวอย่างของข้อมูลเสียงที่ทำการแปลงมาเป็นรูปแบบข้อมูลตัวเลขสี่จريبร้อย

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.03125	0.054688	0.039063	0.046875	0.039063	0.046875	0.023438	0.03125	0.0078125	0	-0.0078125	-0.015625	-0.039063	-0.039063	-0.0625	-0.078125	-0.0625	-0.054688	-0.0625
0	0.0078125	0	0	0	0	0	0	0	0	-0.0078125	0	-0.0078125	0	0	0	0	0	-0.0078125
-0.015625	-0.039063	-0.054688	0.039063	0.09375	0.13281	0.023438	0.015625	-0.046875	-0.13281	-0.21875	-0.17188	-0.22656	0.20313	0.34375	0.015625	0.11719	-0.125	-0.26563
0	0.0078125	0	0.0078125	0	0.0078125	0.0078125	0.0078125	0	0.0078125	0	0.0078125	0	0	0.0078125	0	0	0.0078125	0
-0.0078125	-0.039063	-0.09375	-0.054688	-0.09375	-0.054688	-0.039063	-0.0078125	0.03125	0.023438	0.078125	0.046875	0.078125	0.03125	0.039063	0	0.0078125	-0.023438	0
0	0	0	0	0	0	0	0	0	0.0078125	0	0	0	0	0	0	0	0	0
-0.015625	-0.054688	-0.039063	-0.03125	-0.023438	-0.039063	-0.0625	-0.125	-0.070313	0	-0.046875	0.125	0.0078125	0.10156	0.0625	-0.0078125	0.03125	-0.0625	-0.039063
0	0	-0.0078125	0	0	0	0	0	0	0	0	0	0	-0.0078125	0	0	0	0	0
0	0	0	0	0	0	0	0.0078125	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	-0.0078125	0	-0.0078125	0	0	0	0	0	0	0	0
0.023438	0.078125	0.085938	0.10156	0.10938	0.11719	0.10938	0.11719	0.085938	0.070313	0.023438	0	-0.03125	-0.0625	-0.10156	-0.10938	-0.11719	-0.11719	-0.09375
0	0	0	0	0.0078125	0	0	0	0	0	0	0	0	0	0	0	0.0078125	0	0
0.0078125	-0.0078125	0	0	0	0	0	0	0	0	-0.0078125	0.0078125	0	0	0	0	0	0	0
0.0078125	0	0	0	0	0	0	0	0	0	0	0.0078125	0	0	0.015625	0	0	0	0
-0.085938	-0.13281	-0.03125	-0.023438	0.078125	0.023438	0.054688	0	-0.0078125	-0.023438	-0.046875	-0.0078125	0	0.039063	0.0625	0.046875	0.070313	0.015625	0.03125
0.0078125	0	0	0.0078125	0	-0.0078125	-0.0078125	0.0078125	-0.0078125	-0.0078125	0	0	0	0	0	0	0	0	0
-0.03125	-0.039063	-0.0078125	-0.023438	0.03125	0.0078125	0.046875	0.023438	0.023438	0.015625	0	0.0078125	-0.0078125	0.0078125	0	0.015625	0.03125	0.023438	0.023438
0	0.0078125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.03125	-0.03125	-0.0078125	-0.015625	0.015625	0	0.023438	0	0.023438	0.0078125	0.023438	0.023438	0.03125	0.03125	0.023438	0.0078125	0	0	-0.015625
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0078125
0.015625	0.039063	0.023438	0.03125	0.015625	0.015625	0.0078125	0	0	0	0.015625	0.015625	0.023438	0	0.0078125	0	0.015625	-0.0078125	-0.0078125
-0.0078125	-0.0078125	0	-0.0078125	0	-0.0078125	-0.0078125	0	-0.0078125	0	-0.0078125	-0.0078125	-0.015625	0	0	-0.0078125	-0.015625	-0.0078125	-0.0078125
0.03125	0.039063	0	-0.0078125	-0.0078125	0.015625	0	0	-0.023438	-0.0078125	-0.0078125	-0.023438	-0.070313	-0.09375	-0.078125	0.0078125	0.054688	0.023438	-0.03125
0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0078125	0	0	0	0
0	0	0	-0.0078125	0	0	0	0	0	0	0	0	0	0	0	0	0.0078125	0	0
0	0	0	0.0078125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.0078125	0.015625	0.0078125	0.0078125	0	0	-0.0078125	0	0	-0.015625	-0.0078125	-0.015625	-0.015625	-0.039063	-0.023438	-0.039063	-0.023438	0.015625	0
0	0	0	0	0	0	-0.0078125	-0.0078125	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0078125	0	0
0.0078125	0	-0.03125	-0.023438	-0.0078125	0.0078125	0.070313	0.039063	0.11719	0.046875	0.023438	0.03125	-0.054688	-0.054688	-0.085938	-0.14844	-0.11719	-0.0625	-0.0078125

รูปที่ 3.5 ตัวอย่างของข้อมูลเสียงในรูปแบบตัวเลข

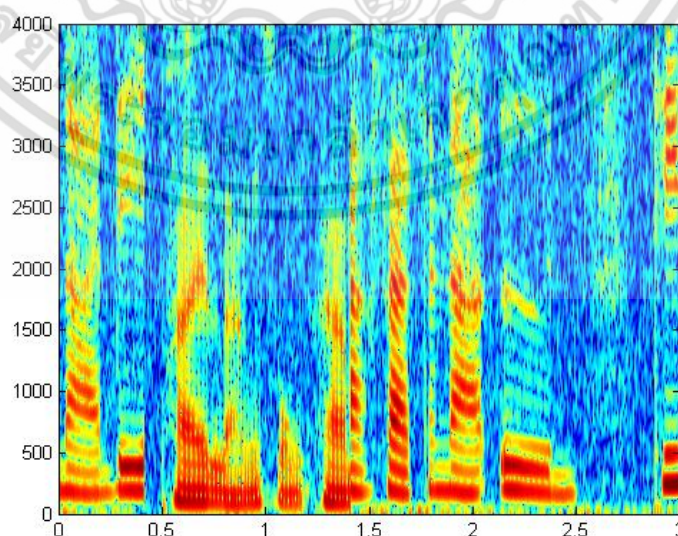
จำนวนแถวคือ จำนวนของข้อมูลเสียงในที่นี้คือ 150 และ 300 ข้อมูลแล้วแต่ไฟล์ที่ได้เตรียมไว้จำนวนคอลัมน์คือ ความยาวต่อ 1 ข้อมูลเสียง ความยาวในที่นี้คือ 15,000 คอลัมน์ ข้อมูลแต่ละคอลัมน์จะเก็บค่าความสูงของกราฟคลื่นเสียง รูปที่ 3.4 ค่าบวกในตารางคือกรณีที่กราฟขึ้นสูง และค่าติดลบคือ กรณีที่กราฟต่ำลง เมื่อทำการแปลงข้อมูลและรวบรวมข้อมูลครบถ้วนแล้วจะได้ไฟล์ข้อมูลสำหรับนำไปใช้ในการเตรียมการสอนและทดสอบระบบต่อไป

3.1.4 การแปลงข้อมูลเสียงในรูปแบบข้อมูลภาพ

เมื่อได้ข้อมูลเสียงตัดต่อและไม่ตัดต่อที่ใช้ในการสอนระบบและทดสอบระบบในรูปแบบไฟล์ CSV แล้ว จากนั้นทำการเตรียมข้อมูลในอีกรูปแบบหนึ่งซึ่งเป็นข้อมูลประเภทภาพ เนื่องจากข้อมูลเสียงนั้นเราสามารถแปลงให้อยู่ในรูปแบบรูปภาพได้ จากนั้นจะทำการเปรียบเทียบประสิทธิภาพความถูกต้องกับข้อมูลเสียงในรูปแบบตัวเลข ว่าถ้าทำการแปลงข้อมูลเสียงเป็นภาพจะให้ผลที่ดีหรือไม่ ซึ่งในงานวิจัยนี้จะแปลงเป็นภาพแบบสเปกโตรแกรม (Spectrogram) การแปลงข้อมูลเสียงให้อยู่ในรูปแบบภาพจะใช้โปรแกรม Matlab เช่นเดียวกับการแปลงข้อมูลตัวเลข ในการแปลงไฟล์ข้อมูลรูปภาพจะทำการแบ่งกลุ่มข้อมูลเป็น 4 กลุ่ม คือ

1. ไฟล์ข้อมูลเสียงตัดต่อเพื่อใช้ในการสอนระบบ (จำนวน 300 ข้อมูล)
2. ไฟล์ข้อมูลเสียงไม่ตัดต่อเพื่อใช้ในการสอนระบบ (จำนวน 300 ข้อมูล)
3. ไฟล์ข้อมูลเสียงตัดต่อเพื่อใช้ในการทดสอบระบบ (จำนวน 150 ข้อมูล)
4. ไฟล์ข้อมูลเสียงไม่ตัดต่อเพื่อใช้ในการทดสอบระบบ (จำนวน 150 ข้อมูล)

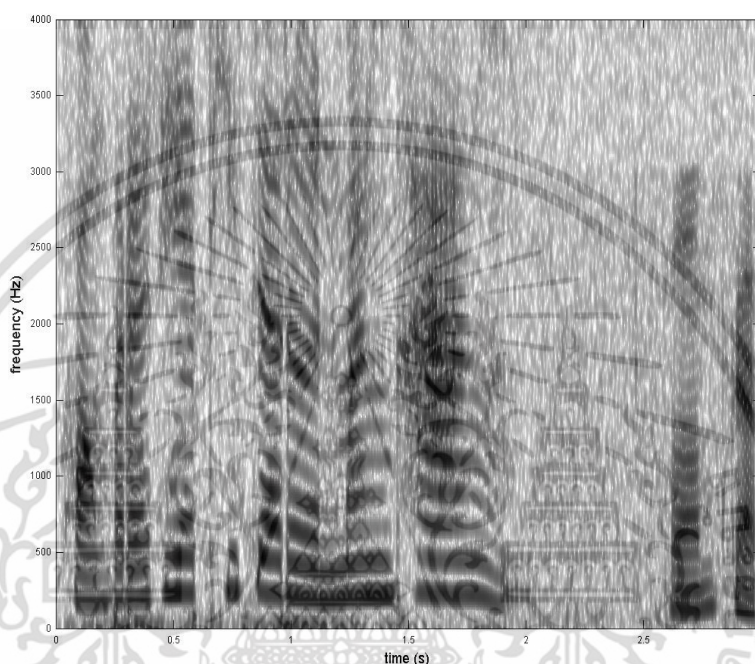
การแปลงข้อมูลเสียงเป็นข้อมูลภาพในโปรแกรม Matlab เริ่มจากการอ่านข้อมูลเสียงเข้ามาในโปรแกรม จากนั้นทำการแปลงข้อมูลเสียงให้กลายเป็นรูปภาพแบบสเปกโตรแกรม เมื่อได้รูปภาพแบบสเปกโตรแกรมแล้วก็จะทำการบันทึกข้อมูลภาพที่มีนามสกุลไฟล์เป็น PNG ดังรูปที่ 3.6



รูปที่ 3.6 ตัวอย่างของข้อมูลภาพแบบสเปกโตรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่ไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อได้ข้อมูลรูปภาพแล้ว ต้องทำการอ่านข้อมูลเข้าไปในโปรแกรม Matlab ภาพที่อ่านเข้ามา จะเป็นภาพแบบ Spectrogram ซึ่งเป็นภาพสีแบบ R G B (Red Green Blue) ซึ่งแต่ละแม่สีจะเก็บ ข้อมูลในแบบของตัวเองจึงทำให้เห็นเสมือนว่ามีภาพซ้อนกัน 3 ภาพ ดังนั้นเพื่อความสะดวกในการ คำนวณ จึงแปลงข้อมูลภาพสีให้อยู่ในรูปแบบภาพสีเทา (Grayscale) ดังรูปที่ 3.7 และทำการ บันทึกรูปภาพ



รูปที่ 3.7 ตัวอย่างของข้อมูลภาพแบบสเปกโตรแกรมสีเทา

เมื่อได้ภาพสีเทามาเรียบร้อยแล้ว เนื่องจากเทคนิคการเรียนรู้ของเครื่องจะอ่านข้อมูลเข้าไป ในรูปแบบเวกเตอร์เท่านั้น แต่ข้อมูลภาพในโปรแกรมเป็นข้อมูลรูปแบบเมทริกซ์ 400×768 จึงเลือกใช้วิธี Singular Value Decomposition (SVD) เพื่อทำการแปลงข้อมูลเมทริกซ์เป็นเวกเตอร์

3.1.5 Singular Value Decomposition (SVD)

SVD จะทำการแยกแฟกเตอร์ (Factor) เมทริกซ์ให้กลายเป็น 3 เมทริกซ์คูณกัน ก่อนที่จะ นำมาใช้เป็นข้อมูล สมมติว่าเมทริกซ์ A มีขนาด $m \times n$ โดย $m=4$ และ $n=5$ ดังรูปที่ 3.8

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \end{bmatrix}$$

รูปที่ 3.8 ตัวอย่างเมทริกซ์ขนาด 4×5
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ในเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นจะทำการแยกแวกเตอร์เมทริกซ์ A ออกเป็น 3 เมทริกซ์คูณกันคือ $A = U \times \Sigma \times V$ โดย U คือเมทริกซ์ของไอเกนเวกเตอร์ซึ่งมีขนาดเป็น $m \times m$ ค่า m ในที่นี้คือ จำนวนแถวของเมทริกซ์ ในตัวอย่างนี้คือ 4 จะได้เมทริกซ์ของไอเกนเวกเตอร์ดังรูปที่ 3.9

$$U = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

รูปที่ 3.9 เมตริกซ์ U

เมทริกซ์ที่สองคือ Σ ซึ่งเป็นเมทริกซ์ไอเกนแวลู ที่จะมีขนาดเท่ากับ A เมทริกซ์นี้จะมีค่าข้อมูลเป็นแนวเฉียงดังรูปที่ 3.10

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

รูปที่ 3.10 เมตริกซ์ Σ

เมทริกซ์สุดท้ายคือ V ซึ่งเป็นเมทริกซ์ขนาด $n \times n$ โดย n ในที่นี้คือ จำนวนคอลัมน์ของเมทริกซ์ในตัวอย่างคือ 5 จะได้เมทริกซ์ V ขนาด 5×5 ดังรูปที่ 3.11

$$V = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

รูปที่ 3.11 เมตริกซ์ V

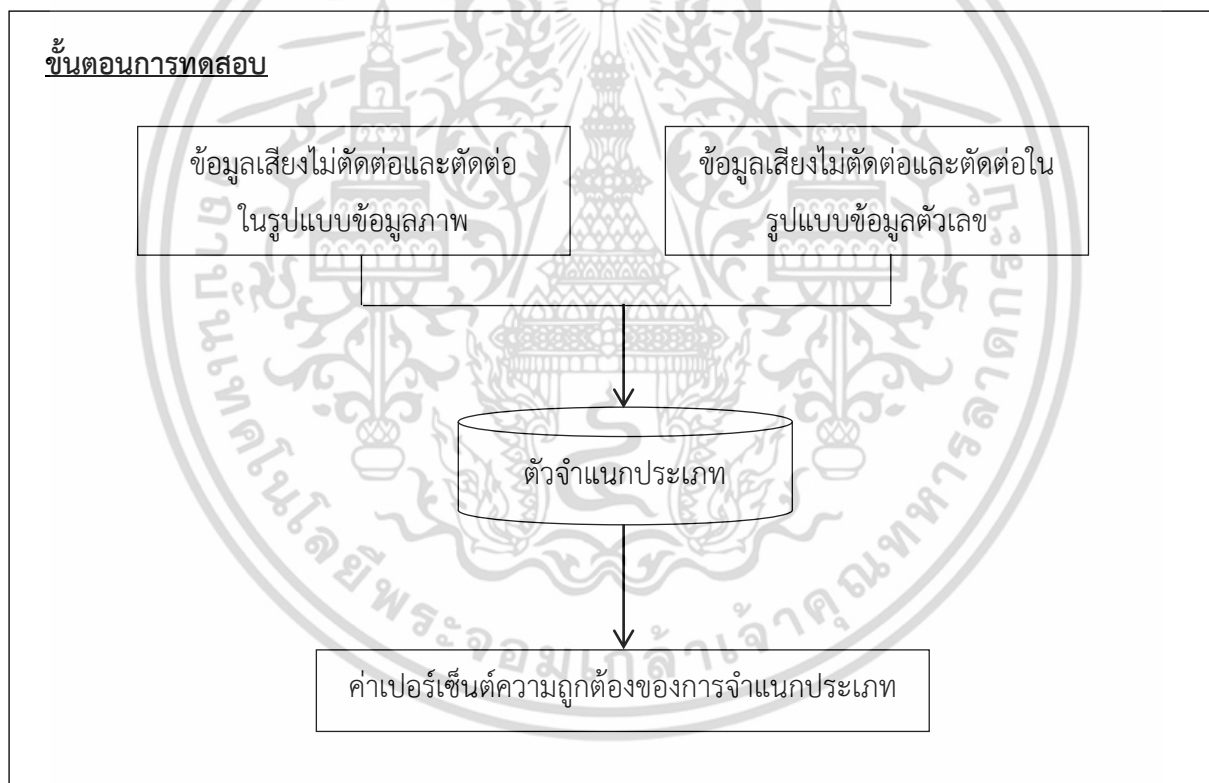
เมทริกซ์ Σ จะเป็นเมทริกซ์ที่นำไปใช้งานซึ่งค่าในเมทริกซ์นี้จะเป็นค่าข้อมูลแนวเฉียง (Diagonal) ซึ่งสามารถแปลงเป็นข้อมูลแบบเวกเตอร์ได้ จึงเลือกค่านี้ไปใช้ในการสอนและทดสอบระบบที่ไม่เลือก เมทริกซ์ในส่วนของ U และ V มาใช้งานก็เพราะว่า ข้อมูลยังอยู่ในรูปแบบเมทริกซ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อยู่ซึ่งเทคนิคการเรียนรู้ของเครื่องไม่สามารถนำมาใช้งานได้ และที่นำวิธี SVD มาช่วยเพราะว่าลดขนาดของข้อมูลในการสอนและทดสอบ ถ้านำข้อมูลเสียงรูปแบบภาพปกติของแต่ละแถมมาต่อกัน ข้อมูลต่อหนึ่งภาพจะมีจำนวนมากในการทำเป็นเวกเตอร์

3.2 ขั้นตอนการทดสอบ

เมื่อข้อมูลสำหรับการสอนและทดสอบโมเดลครบถ้วนแล้ว จากนั้นทำการทดสอบโมเดลโดยใช้อัลกอริทึมหลักคือ การเรียนรู้แบบรวมกลุ่มโดยนำมาเปรียบเทียบกับ อัลกอริทึมต่างๆ อีกจำนวน 4 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน ตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบโพรบาบิลิสติก โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน โดยขั้นตอนการทดสอบ แสดงดังรูป 3.12



รูปที่ 3.12 ขั้นตอนการเรียนรู้และทดสอบโมเดล

จากรูปที่ 3.8 นำข้อมูลทั้งสองประเภทคือ ข้อมูลเสียงไม่ตัดต่อและตัดต่อในรูปแบบข้อมูลภาพและข้อมูลเสียงไม่ตัดต่อและตัดต่อในรูปแบบข้อมูลตัวเลข มาทำการสร้างโมเดลของการเรียนรู้แบบรวมกลุ่มและสร้างโมเดลการเรียนรู้ของเครื่องอีก 4 โมเดลดังตารางที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 เทคนิคการเรียนรู้ของเครื่องและฟังก์ชันที่ใช้ในการจำแนกประเภท

Machine Learning	Function
Ensemble Learning	-
Naïve Bayes	Gaussian Function
Support Vector Machine	RBF
Radial basis function neural network	Gaussian Function
Probabilistic neural network	Gaussian Function

จากตารางที่ 3.1 จะเห็นว่าเทคนิคการเรียนรู้ของเครื่องจะใช้ฟังก์ชันช่วยในการจำแนกประเภทข้อมูล (Classification) โดยแต่ละการจำแนกประเภทข้อมูลจะใช้ฟังก์ชันในการช่วยเพิ่มประสิทธิภาพที่ต่างกันไปในวิทยานิพนธ์นี้ทำการเลือกฟังก์ชันดังต่อไปนี้ ตัวจำแนกแบบเบย์อย่างง่ายเลือกใช้ Gaussian Function ซึ่งเป็นฟังก์ชันพื้นฐานที่ช่วยในการหาค่าประมาณความน่าจะเป็นเนื่องจากถ้าข้อมูลเป็นแบบไม่ต่อเนื่องสามารถหาค่าความน่าจะเป็นได้ทันที แต่ถ้าต้องการจะหาความน่าจะเป็นของข้อมูลที่เป็นแบบต่อเนื่องจะต้องใช้ Gaussian Function มาช่วยในการประมาณค่าความน่าจะเป็น ต่อมาในส่วนของซัพพอร์ตเวกเตอร์แมชชีนเลือกใช้ RBF เป็น Kernel Function โดยการกำหนดค่าพารามิเตอร์ความกว้างของ Gaussian Function ที่มีลักษณะเป็นระฆังคว่ำซึ่งจะครอบคลุมข้อมูลและแมพข้อมูลเพื่อทำการจำแนกประเภทข้อมูลต่อไป ในส่วนของโครงข่ายประสาทเทียมทั้งสองแบบได้แก่ โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก เลือกใช้ Gaussian Function มาช่วยในการเพิ่มประสิทธิภาพของโครงข่ายประสาทเทียมทั้งสองรูปแบบ ซึ่งในแต่ละโมเดลจะกำหนดค่าพารามิเตอร์ของแต่ละฟังก์ชันแตกต่างกันออกไปตามความเหมาะสมของแต่ละโมเดลการเรียนรู้ เพื่อให้เปอร์เซ็นต์ความถูกต้องเพิ่มสูงขึ้น เพราะบางฟังก์ชันปรับค่าพารามิเตอร์น้อยให้ผลไม่ดี แต่พอปรับค่าสูงขึ้นก็ให้ผลที่ดีขึ้น เมื่อได้ค่าที่ดีที่สุดของแต่ละการจำแนกประเภทข้อมูลก็นำผลที่ได้นั้นมาทำการเปรียบเทียบเพื่อวัดประสิทธิภาพว่าการจำแนกประเภทข้อมูลไหนดีที่สุดต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ขั้นตอนการวัดประสิทธิภาพ

เมื่อผ่านขั้นตอนการทดสอบแล้ว จากนั้นจะทำการรวบรวมค่าประสิทธิภาพความถูกต้อง โดยทำการเก็บค่าประสิทธิภาพความถูกต้องและเวลาของแต่ละการปรับค่าพารามิเตอร์ในตัวจำแนกประเภทข้อมูลว่า ในการจำแนกประเภทข้อมูลค่าไหนที่ให้ประสิทธิภาพความถูกต้องสูงสุดในข้อมูล รูปแบบตัวเลขและค่าที่ให้ประสิทธิภาพความถูกต้องสูงสุดในรูปแบบภาพในแต่ละตัวจำแนกประเภทข้อมูล และทำการวัดผล ตัวอย่างเช่น เก็บค่าประสิทธิภาพความถูกต้องของการจำแนกประเภทข้อมูล รูปแบบตัวเลขจากการปรับค่าพารามิเตอร์ของฟังก์ชันในตัวจำแนกประเภทข้อมูล และหาค่าประสิทธิภาพความถูกต้องสูงสุด จากนั้นนำค่าประสิทธิภาพความถูกต้องสูงสุดที่ได้ในแต่ละตัวจำแนกประเภทข้อมูลมาทำการเปรียบเทียบว่าตัวจำแนกประเภทข้อมูลแบบไหนให้ประสิทธิภาพความถูกต้องสูงที่สุดจากตัวจำแนกประเภทข้อมูลที่เลือกมาทั้งหมด 5 ประเภท และเก็บประสิทธิภาพความถูกต้องของการจำแนกประเภทข้อมูลของข้อมูลเสียงรูปแบบภาพจากการปรับค่าพารามิเตอร์ของฟังก์ชันในตัวจำแนกประเภทข้อมูล และหาค่าประสิทธิภาพความถูกต้องสูงสุด จากนั้นนำค่าประสิทธิภาพความถูกต้องสูงสุดที่ได้ในแต่ละตัวจำแนกประเภทข้อมูลมาทำการเปรียบเทียบว่า ตัวจำแนกประเภทข้อมูลแบบไหนให้ประสิทธิภาพความถูกต้องสูงที่สุดจากตัวจำแนกประเภทข้อมูลที่เลือกมาทั้งหมด 5 ประเภทเช่นเดียวกัน และทำการเปรียบเทียบข้อมูล ว่าผลประสิทธิภาพความถูกต้องสูงสุดในทุกตัวจำแนกประเภทข้อมูลแบบไหนดีกว่ากันระหว่างข้อมูลเสียงรูปแบบตัวเลข และข้อมูลเสียงรูปแบบภาพ

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทดลองและผลการทดลอง โดยใช้การเรียนรู้แบบรวมกลุ่มเป็นโมเดลหลักในพิสูจน์ข้อมูลเสียง และนำมาเปรียบเทียบประสิทธิภาพความถูกต้องกับโมเดลการเรียนรู้ของเครื่องอีกจำนวน 4 โมเดล โดยได้รับความอนุเคราะห์ข้อมูลเสียงจากห้องปฏิบัติการเทคโนโลยีเสียง ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ จากนั้นทำการเตรียมข้อมูลเสียงแบบปกติจำนวน 300 ข้อมูลเสียงที่ทำการตัดต่อจำนวน 300 ข้อมูลเสียง รวม 600 ข้อมูลเสียง สำหรับการสอนโมเดล และ ข้อมูลเสียงแบบปกติจำนวน 150 ข้อมูลเสียงที่ทำการตัดต่อจำนวน 150 ข้อมูลเสียง รวม 300 ข้อมูลเสียงสำหรับทดสอบโมเดล วิทยานิพนธ์นี้ทำการแปลงข้อมูลเป็น 2 ลักษณะ คือ การแปลงข้อมูลเสียงให้อยู่ในรูปแบบข้อมูลตัวเลข และการแปลงข้อมูลเสียงให้อยู่ในรูปแบบภาพ ข้อมูลทั้งสองรูปแบบจะถูกใช้ในการสอนและทดสอบระบบเท่ากับจำนวนที่เตรียมไว้ข้างต้น เพื่อทำการเปรียบเทียบประสิทธิภาพความถูกต้อง

4.1 การทดลอง

ในการทดลองนี้ได้ทำการสร้างโมเดลการเรียนรู้แบบรวมกลุ่มเป็นโมเดลหลัก และทำการเลือกโมเดล การเรียนรู้ของเครื่องมาอีก 4 โมเดล คือ ซัพพอร์ตเวกเตอร์แมชชีน ตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน โครงข่ายประสาทเทียมแบบโพรบาบิลิสติก ซึ่งในแต่ละวิธีจะทำการปรับค่าพารามิเตอร์ที่แตกต่างกัน เพื่อให้ประสิทธิภาพดีที่สุดและนำผลลัพธ์ที่ได้ไปทำการเปรียบเทียบผลลัพธ์กับวิธีการเรียนรู้แบบรวมกลุ่ม และเปรียบเทียบประสิทธิภาพระหว่างข้อมูลเสียงรูปแบบตัวเลข และข้อมูลเสียงรูปแบบภาพ เพื่อวัดประสิทธิภาพความถูกต้องระหว่างข้อมูลทั้งสอง

4.1.1 การปรับค่าพารามิเตอร์

ในการทดลองนี้ได้กำหนดฟังก์ชันและการปรับค่าพารามิเตอร์ที่แตกต่างกัน เพื่อใช้ในการทดสอบและวัดประสิทธิภาพของโมเดล ตารางที่ 4.1 แสดงโมเดลการเรียนรู้ของเครื่องที่ทำการทดลอง ฟังก์ชันที่ใช้ในแต่ละโมเดล และค่าพารามิเตอร์ที่ปรับเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

ตารางที่ 4.1 ฟังก์ชันและการปรับค่าพารามิเตอร์

โมเดลการเรียนรู้ของเครื่อง	ฟังก์ชัน	ค่าพารามิเตอร์
Ensemble Learning	-	1-1000
Support Vector Machine	RBF	0.1 - 1000
Naïve Bayes	Gaussian Function	1-1000
Radial basis Function Neural Network	Gaussian Function	0.1 - 1000
Probabilistic Neural network	Gaussian Function	1 - 1000

4.2 ผลการทดลอง

ในขั้นตอนนี้จะทำการเปรียบเทียบผลลัพธ์ของแต่ละโมเดลระหว่างการเรียนรู้แบบรวมกลุ่ม และ 4 โมเดลที่เหลือ ผลลัพธ์ที่ได้จากการทดลองแสดงดังตารางที่ 4.2 ซึ่งเป็นส่วนของการทดสอบโมเดลการเรียนรู้แบบรวมกลุ่มโดยการปรับค่าต่างๆ ของข้อมูลเสียงในรูปแบบตัวเลข

ตารางที่ 4.2 การทดลองโมเดลการเรียนรู้แบบรวมกลุ่ม (ข้อมูลรูปแบบตัวเลข)

ค่าจำนวนรอบ	ความถูกต้อง	เวลา (วินาที)
1 - 100	98.833	15.932
101 - 200	99.667	29.101
201 - 300	99.667	148.954
301 - 400	99.667	307.352
401 - 500	100	1585.290

จากผลการทดลองในตารางที่ 4.2 คือ ค่าความถูกต้องที่ได้จากการปรับค่าพารามิเตอร์ของการเรียนรู้แบบรวมกลุ่ม โดยค่าจำนวนรอบจะเป็นค่าจำนวนโมเดลย่อยในการเรียนรู้แบบรวมกลุ่ม ถ้าปรับมากจำนวนโมเดลย่อยก็ยิ่งเพิ่มขึ้นมีการปรับค่าความผิดพลาดในโมเดลย่อยหลายรอบ ทำให้ผลออกมาดีขึ้น ในตารางนี้แสดงการปรับค่าเป็นช่วงเช่นปรับตั้งแต่ค่า 1 – 100 จนได้ค่าประสิทธิภาพความถูกต้องสูงสุด ค่าจำนวนรอบที่ทำให้ความถูกต้องสูงสุดคือ 500 ต่อไปเป็นการทดลองโมเดลซัพพอร์ตเวกเตอร์แมชชีน ดังตารางที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 การทดลองโมเดลซัพพอร์ตเวกเตอร์แมชชีน (ข้อมูลรูปแบบตัวเลข)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
0.1 - 1	94.167	4.206
1 - 50	99.667	4.156
50 - 100	99.667	4.025
101 - 300	95.333	4.116
301 - 500	64	4.267

จากผลการทดลองในตารางที่ 4.3 คือ ค่าความถูกต้องที่ได้จากการปรับค่า σ ของโมเดลซัพพอร์ตเวกเตอร์แมชชีน เนื่องจากซัพพอร์ตเวกเตอร์แมชชีนเป็นการแบ่งข้อมูลเป็นสองกลุ่มโดยใช้ไฮเปอร์เพลนตัดกลางระหว่างข้อมูล ในการทดลองนี้เลือกใช้ Kernel Function เป็น RBF ซึ่งจะใช้ฟังก์ชันเชิงรัศมีครอบคลุมของข้อมูล โดยค่าที่ทำการปรับคือ ค่าความกว้างเชิงรัศมี σ ซึ่งจะปรับตามความเหมาะสมอยู่ที่แต่ละประเภทข้อมูล เพราะข้อมูลบางประเภทมีการกระจายตัวกว้างสูง จึงต้องปรับค่าพารามิเตอร์ความกว้างให้กว้างเพื่อครอบคลุมข้อมูลเพื่อทำการแบ่งกลุ่มข้อมูลทำให้ค่าออกมาดีขึ้น แต่ข้อมูลบางประเภทมีการเกาะกลุ่มกัน จึงต้องใช้ค่าพารามิเตอร์ความกว้างที่น้อยจึงจะให้ผลที่ดี ค่าพารามิเตอร์ความกว้างที่ปรับแล้วให้ผลดีที่สุด คือ 51 ให้ความถูกต้องคือ 99.667% และใช้เวลาในการประมวลผลน้อยกว่าค่าอื่นที่ผลเท่ากัน เมื่อปรับค่าสูงขึ้นทำให้ความถูกต้องลดลง แสดงว่าข้อมูลมีการเกาะกลุ่มกันจึงใช้ค่าพารามิเตอร์ความกว้างเชิงรัศมีน้อยในการครอบคลุม ต่อไปเป็นการทดลองตัวจำแนกแบบเบย์อย่างง่ายซึ่งผลการทดลองแสดงดังตารางที่ 4.4

ตารางที่ 4.4 การทดลองโมเดลตัวจำแนกแบบเบย์อย่างง่าย (ข้อมูลรูปแบบตัวเลข)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
1 - 100	97.833	123.408
101 - 200	97.833	129.847
201 - 300	96	138.369
301 - 400	95.833	151.308
401 - 500	95.833	112.872

จากผลการทดลองในตารางที่ 4.4 คือ ค่าความถูกต้องที่ได้จากการปรับค่า σ เนื่องจากข้อมูลเสียงเป็นข้อมูลในรูปแบบค่าต่อเนื่องทำให้ไม่สามารถหาค่าความน่าจะเป็นได้ จึงนำ Gaussian Function มาทำการช่วยในการประมาณค่าเพื่อหาความน่าจะเป็นของข้อมูล ค่า σ ของโมเดลตัวจำแนกแบบเบย์อย่างง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่ขึ้นต้นการคำนวณว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำแนกแบบเบ้อย่างง่ายที่ให้ผลความถูกต้องดีที่สุดคือ 105 ซึ่งได้ผลความถูกต้อง 97.833% เมื่อปรับค่าสูงขึ้นทำให้ความถูกต้องไม่มีการเพิ่มไปมากกว่านี้และเมื่อทำการปรับค่าถึง 200 ค่าความถูกต้องเริ่มตกลง ต่อไปเป็นการทดลองโมเดลโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน ได้ผลการทดลองดังตารางที่ 4.5

ตารางที่ 4.5 การทดลองโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน (ข้อมูลรูปแบบตัวเลข)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
1 - 100	47.833	248.289
101 - 200	53.667	309.445
201 - 300	50.667	317.927
301 - 400	49.833	341.9517
401 - 500	49	341.071

จากผลการทดลองในตารางที่ 4.5 คือ ค่าความถูกต้องที่ได้จากการปรับค่า σ ของโมเดลโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน ค่าที่ทำการปรับคือ ค่า σ ในสมการ Gaussian Function หรือความกว้างของฟังก์ชันเชิงรัศมีเพื่อมาทำการแมพข้อมูลเพื่อแบ่งข้อมูลเป็นสองกลุ่มโดยไฮเปอร์เพลนที่ทำการตัดแบ่งข้อมูลเป็นสองกลุ่มนั้น จะทำการตัดแบ่งจากข้อมูลทำการแมพมาจากฟังก์ชันเชิงรัศมีซึ่งครอบคลุมข้อมูลที่มีคลาสต่างกันกันเยอะทำให้เกิดค่าความผิดพลาดสูง การปรับค่าพารามิเตอร์ที่ให้ผลดีที่สุดคือ 150 ให้ผลความถูกต้องคือ 53.667% เมื่อปรับค่าสูงขึ้นทำให้ผลความถูกต้องเริ่มลดลง และการปรับค่าในช่วงแรกให้ค่าความถูกต้องที่น้อยกว่า 150 จึงเลือก 150 เป็นค่าที่ดีที่สุด ต่อไปเป็นการทดลองโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก ได้ผลการทดลองดังตารางที่ 4.6

ตารางที่ 4.6 การทดลองโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก (ข้อมูลรูปแบบตัวเลข)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
1 - 100	50	318.218
101 - 200	50	286.923
201 - 300	50	325.488
301 - 400	50	306.491
401 - 500	50	323.405

จากผลการทดลองตารางที่ 4.6 คือ ค่าความถูกต้องที่ได้จากการปรับค่า σ ของ Gaussian Function ที่ครอบคลุมข้อมูล โดยถ้าข้อมูลอยู่ในระยะใกล้เคียงกับจุดศูนย์กลางของ Gaussian Function ก็จะถือว่าข้อมูลนั้นอยู่ในกลุ่มเดียวกัน ค่า σ ของโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติกฟังก์ชัน ที่ให้ผลดีที่สุดคือ 110 ให้ค่าความถูกต้อง 50% เมื่อปรับค่าให้สูงขึ้นทำให้ผลความถูกต้องไม่มีการเปลี่ยนแปลง เมื่อทำการทดลองครบทั้ง 5 โมเดลเรียบร้อยแล้ว ทำการเก็บค่าความถูกต้องดีที่สุดของแต่ละอัลกอริทึม เพื่อมาทำการเปรียบเทียบผลดังตารางที่ 4.7

ตารางที่ 4.7 การเปรียบเทียบโมเดล 5 วิธีของข้อมูลเสียงรูปแบบตัวเลข

โมเดลการเรียนรู้ของเครื่อง	ฟังก์ชัน	ค่าพารามิเตอร์ที่ดีที่สุด	ความถูกต้อง	เวลา (วินาที)
Ensemble Learning	-	500	100	1585.290
Support Vector Machine	RBF Sigma	51	99.667	4.156
Naïve Bayes	Gaussian Function	105	97.833	123.408
Radial basis Function Neural Network	Gaussian Function	150	53.667	309.445
Probabilistic Neural network	Gaussian Function	110	50	318.218

จากผลการทดลองในตารางที่ 4.7 คือ ค่าความถูกต้องของข้อมูลเสียงในรูปแบบตัวเลข โมเดลการเรียนรู้แบบรวมกลุ่มให้ความถูกต้องสูงที่สุดคือ 100% รองลงมาคือ ซัพพอร์ตเวกเตอร์แมชชีน 99.667% และตัวจำแนกแบบเบย์อย่างง่าย 97.833 ตามลำดับ ในการทดลองได้ทำการลองปรับค่าพารามิเตอร์หลายค่าเพื่อให้ประสิทธิภาพความถูกต้องออกมาสูงที่สุด ซึ่งค่าพารามิเตอร์ที่ให้ผลดีที่สุดอยู่ในคอลัมน์ค่าพารามิเตอร์ดีที่สุด ต่อไปเป็นการทดลองกับข้อมูลเสียงในรูปแบบภาพกับโมเดลการเรียนรู้แบบรวมกลุ่ม และ การเรียนรู้ของเครื่องอีก 4 โมเดล การทดลองแต่ละอัลกอริทึมจะทำได้เช่นเดียวกับข้อมูลเสียงในรูปแบบตัวเลข และทำการเปรียบเทียบผลเช่นเดียวกัน โดยข้อมูลนำเข้ามาจากข้อมูลเสียงในรูปแบบภาพ เมื่ออ่านค่าเข้ามาจะเป็นค่าของความเข้มสีในข้อมูลภาพ ซึ่งเป็นสีเทา (Gray scale) จากการทดลองข้อมูลเสียงในรูปแบบภาพของการเรียนรู้แบบรวมกลุ่มจะได้ผลการทดลองดังตารางที่ 4.8

ตารางที่ 4.8 การทดลองโมเดลการเรียนรู้แบบรวมกลุ่ม (ข้อมูลรูปแบบภาพ)

ค่าจำนวนรอบ	ความถูกต้อง	เวลา (วินาที)
1 - 100	97.833	14.180
101 - 200	99.833	31.265
201 - 300	99.833	126.612
301 - 400	99.833	264.370
401 - 500	99.833	1237.750

จากตารางที่ 4.8 การทดลองการเรียนรู้แบบรวมกลุ่มกับข้อมูลเสียงในรูปแบบภาพ โดยทำการปรับค่าพารามิเตอร์เริ่มต้นที่ 1 และทำการปรับเพิ่มขึ้นเรื่อยๆ ตามตารางแสดงการปรับค่าเป็นช่วง เช่นปรับค่าเริ่มตั้งแต่ 1 ถึง 100 จนได้ค่าประสิทธิภาพความถูกต้องและใช้เวลาที่ดีที่สุด ค่าพารามิเตอร์ที่ให้ความถูกต้องดีที่สุดคือ 120 ต่อไปเป็นการทดลองโมเดลซัพพอร์ตเวกเตอร์แมชชีน ได้ผลการทดลองดังตารางที่ 4.9

ตารางที่ 4.9 การทดลองโมเดลซัพพอร์ตเวกเตอร์แมชชีน (ข้อมูลรูปแบบภาพ)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
0.1 - 1	95.333	1.953
1 - 50	99.667	2.023
50 - 100	99.667	1.923
101 - 300	97.333	1.843
301 - 500	73	1.943

จากผลการทดลองตารางที่ 4.9 คือ ค่าความถูกต้องที่ได้จากการปรับค่าพารามิเตอร์ของโมเดลซัพพอร์ตเวกเตอร์แมชชีน ค่า σ ที่ให้ผลดีที่สุดคือ 52 ให้ความถูกต้องคือ 99.667% เมื่อปรับค่าสูงขึ้นทำให้ความถูกต้องลดต่ำลง แสดงว่าข้อมูลเสียงรูปแบบตัวเลขและรูปแบบภาพ การกระจายตัวของข้อมูลไม่ต่างกันมาก ข้อมูลมีการเกาะกลุ่มกันสูงจึงปรับค่าความกว้างเชิงรัศมีที่มีขนาดไม่ใหญ่มาก ซึ่งเมื่อปรับค่าที่ใหญ่เกินไปทำให้เกิดการแบ่งข้อมูลที่ยาก ทำให้ค่าความถูกต้องลดลง ต่อไปเป็นการทดลองตัวจำแนกแบบเบย์อย่างง่าย ได้ผลการทดลองดังตารางที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 การทดลองโมเดลตัวจำแนกแบบเบย์อย่างง่าย (ข้อมูลรูปแบบภาพ)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
1 - 100	97.833	82.840
101 - 200	98.333	84.211
201 - 300	99	88.668
301 - 400	96.833	85.393
401 - 500	96.833	83.740

จากผลการทดลองในตารางที่ 4.10 คือ ค่าความถูกต้องที่ได้จากการปรับค่า σ ในโมเดลตัวจำแนกแบบเบย์อย่างง่าย โดยใช้ Gaussian Function มาช่วยในการประมาณค่าความน่าจะเป็นของข้อมูลแบบต่อเนื่อง ค่า σ ที่ให้ผลดีที่สุดคือ 220 ให้ความถูกต้อง 99% เมื่อทำการปรับค่าสูงขึ้นทำให้ค่าความถูกต้องไม่มีการเพิ่มไปมากกว่านี้และเมื่อทำการปรับค่าถึง 290 ค่าประสิทธิภาพเริ่มลดลงต่อไปเป็นการทดลองโมเดลโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน ได้ผลการทดลองดังตารางที่ 4.11

ตารางที่ 4.11 การทดลองโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน (ข้อมูลรูปแบบภาพ)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
1 - 100	62.333	127.913
101 - 200	53.833	125.480
201 - 300	50.667	133.432
301 - 400	47.333	120.754
401 - 500	47.333	117.259

จากผลการทดลองในตารางที่ 4.11 คือ ค่าความถูกต้องที่ได้จากการปรับค่าความกว้างเชิงรัศมี σ ของโมเดลโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน โดยค่า σ ที่ให้ผลดีที่สุดคือ 10 ให้ความถูกต้อง 62.333% เมื่อปรับค่าสูงขึ้นจะทำให้ความถูกต้องเริ่มลดลงต่อไปเป็นการทดลองโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก ได้ผลการทดลองดังตารางที่ 4.12

ตารางที่ 4.12 การทดลองโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก (ข้อมูลรูปแบบภาพ)

ค่า σ	ความถูกต้อง	เวลา (วินาที)
1 - 100	53.667	91.341
101 - 200	53.667	90.42
201 - 300	53.667	92.263
301 - 400	53.667	90.581
401 - 500	53.667	90.359

จากผลการทดลองตารางที่ 4.12 คือ ค่าความถูกต้องที่ได้จากการปรับค่า σ ของฟังก์ชัน Gaussian Function ที่นำมาครอบกลุ่มข้อมูล ถ้าข้อมูลอยู่ในระยะใกล้เคียงกันกับจุดศูนย์กลางของ Gaussian Function ก็จะถือว่าข้อมูลนั้นอยู่กลุ่มเดียวกัน ค่า σ ของโมเดลโครงข่ายประสาทเทียมแบบโพรบาบิลิสติกฟังก์ชันที่ให้ผลดีที่สุดคือ 1 ให้ความถูกต้อง 53.667% เมื่อปรับค่าสูงขึ้นทำให้ความถูกต้องไม่มีการเปลี่ยนแปลง หลังจากทำการทดลองครบทั้ง 5 โมเดลเรียบร้อยแล้ว จากนั้นทำการเก็บค่าความถูกต้องดีที่สุดในแต่ละอัลกอริทึมมาทำการเปรียบเทียบผล ดังตารางที่ 4.13

ตารางที่ 4.13 การเปรียบเทียบโมเดล 5 วิธีของข้อมูลเสียงรูปแบบภาพ

โมเดลการเรียนรู้ของ เครื่อง	Function	ค่าพารามิเตอร์ ที่ดีที่สุด	ความถูกต้อง	เวลา (วินาที)
Ensemble Learning	-	120	99.833	31.265
Support Vector Machine	RBF Sigma	52	99.667	2.023
Naïve Bayes	Gaussian Function	220	99	88.668
Radial basis Function Neural Network	Gaussian Function	10	62.333	127.913
Probabilistic Neural network	Gaussian Function	410	53.667	90.359

จากผลการทดลองในตารางที่ 4.13 ความถูกต้องของข้อมูลเสียงในรูปแบบภาพของโมเดลการเรียนรู้แบบรวมกลุ่มให้ความถูกต้องสูงที่สุดคือ 99.833% ซัพพอร์ตเวกเตอร์แมชชีน 99.667% เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวจำแนกแบบเบย์อย่างง่าย 99% โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน 62.333% และ โครงข่ายประสาทเทียมแบบโพรบาบิลิสติก 53.667% ในการทดลองนี้ได้ทำการปรับค่าพารามิเตอร์หลายค่าเพื่อหาค่าที่ดีที่สุดเช่นเดียวกับการทดลองกับข้อมูลเสียงในรูปแบบตัวเลข เมื่อได้ผลความถูกต้องของข้อมูลเสียงทั้งสองรูปแบบแล้ว จากนั้นจะนำผลความถูกต้องของข้อมูลทั้งสองรูปแบบมาทำการเปรียบเทียบ ดังตารางที่ 4.14

ตารางที่ 4.14 ตารางเปรียบเทียบค่าประสิทธิภาพของข้อมูล 2 รูปแบบ

โมเดลการเรียนรู้ของเครื่อง	Function	ความถูกต้อง (ข้อมูลแบบตัวเลข)	ความถูกต้อง (ข้อมูลแบบภาพ)
Ensemble Learning	-	100	99.833
Support Vector Machine	RBF	99.667	99.667
Naïve Bayes	Gaussian Function	97.833	99
Radial basis Function Neural Network	Gaussian Function	53.667	62.333
Probabilistic Neural network	Gaussian Function	50	53.667

จากผลการทดลองในตารางที่ 4.14 คือ การเปรียบเทียบความถูกต้องของข้อมูลเสียงทั้งสองรูปแบบ จากผลการทดลอง การเรียนรู้แบบรวมกลุ่มในรูปแบบข้อมูลตัวเลขให้ผลที่ดีกว่ารูปแบบภาพ ส่วนของซัพพอร์ตเวกเตอร์แมชชีนให้ผลที่เท่ากัน ในส่วนของตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก ให้ผลความถูกต้องเพิ่มขึ้นทั้งสามโมเดล เพราะฉะนั้นการแปลงข้อมูลเป็นภาพสามารถช่วยเพิ่มความถูกต้องของ 3 โมเดลคือ ตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก แต่อย่างไรก็ตามโมเดลการเรียนรู้แบบรวมกลุ่มให้ค่าความถูกต้องสูงสุดทั้งข้อมูลแบบตัวเลขและข้อมูลแบบรูปภาพ

บทที่ 5

สรุปและข้อเสนอแนะ

5.1 สรุป

งานวิจัยนี้จัดทำขึ้นสำหรับการระบุข้อมูลเสียงว่าเป็น ข้อมูลเสียงตัดต่อหรือข้อมูลเสียงแบบปกติ โดยทำการแปลงข้อมูลเสียงเป็น 2 ลักษณะคือ ข้อมูลเสียงในรูปแบบตัวเลข และข้อมูลเสียงในรูปแบบภาพเพื่อใช้สำหรับการสอนและทดสอบตัวจำแนกประเภท 5 โมเดล และทำการเปรียบเทียบค่าประสิทธิภาพความถูกต้องทั้ง 5 โมเดล ได้แก่ การเรียนรู้แบบรวมกลุ่ม ซัพพอร์ตเวกเตอร์แมชชีน ตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน โครงข่ายประสาทเทียมแบบโพรบาบิลิสติก และทำการเปรียบเทียบค่าประสิทธิภาพความถูกต้องระหว่างการแปลงข้อมูลทั้งสองรูปแบบว่าใช้ข้อมูลแบบไหนให้ผลความถูกต้องดีกว่ากัน

จากผลการทดลองเปรียบเทียบประสิทธิภาพโดยใช้ข้อมูลเสียงรูปแบบตัวเลข การเรียนรู้แบบรวมกลุ่มให้ประสิทธิภาพความถูกต้องดีที่สุดคือ 100% รองลงมาคือ ซัพพอร์ตเวกเตอร์แมชชีน ให้ผลความถูกต้องเท่ากับ 99.667% และตัวจำแนกแบบเบย์อย่างง่าย ให้ผลความถูกต้องเท่ากับ 97.833% ส่วนโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก ให้ผลความถูกต้องเท่ากับ 53.667% และ 50% ตามลำดับ ในการทดสอบกับข้อมูลเสียงในรูปแบบตัวเลขการเรียนรู้แบบรวมกลุ่มให้ผลความถูกต้องสูงที่สุด

จากผลการทดลองเปรียบเทียบประสิทธิภาพโดยใช้ข้อมูลเสียงรูปแบบภาพ การเรียนรู้แบบรวมกลุ่มให้ผลความถูกต้องดีที่สุด โดยให้ผลเท่ากับ 99.833% รองลงมาคือ ซัพพอร์ตเวกเตอร์แมชชีน ให้ผลเท่ากับ 99.667% และตัวจำแนกแบบเบย์อย่างง่าย ให้ผลเท่ากับ 99% ส่วนโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติกให้ผลความถูกต้องเท่ากับ 62.333% และ 53.667% ตามลำดับ การทดสอบกับข้อมูลเสียงในรูปแบบภาพ การเรียนรู้แบบรวมกลุ่มให้ผลความถูกต้องสูงที่สุดเช่นเดียวกัน แต่การใช้ข้อมูลเสียงรูปแบบภาพ ช่วยให้ผลความถูกต้องของตัวจำแนกแบบเบย์อย่างง่าย โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และโครงข่ายประสาทเทียมแบบโพรบาบิลิสติก มีผลดีขึ้นกว่าการทดสอบโดยใช้ข้อมูลเสียงแบบตัวเลข

เมื่อเปรียบเทียบประสิทธิภาพระหว่างข้อมูลเสียงทั้งสองรูปแบบ การเรียนรู้แบบรวมกลุ่มให้

ค่าความถูกต้องดีที่สุดในทั้งสองรูปแบบข้อมูล เพราะการเรียนรู้แบบรวมกลุ่มเป็นการใช้โมเดลการเอกซตรัคชันเป็นเอกซตรัคชันที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรียนรู้ต้นไม้ช่วยตัดสินใจ ทำงานซ้ำหลายๆ รอบตามจำนวนโมเดลที่เรากำหนดและมีการปรับค่าความผิดพลาดอยู่เสมอ จึงทำให้ผลที่ออกมามีความถูกต้องสูงทั้งสองรูปแบบข้อมูล แต่ใช้เวลาในการประมวลผลนานกว่าการจำแนกประเภทข้อมูลรูปแบบอื่น เพราะมีการสร้างโมเดลจำแนกประเภทข้อมูลย่อยในการเรียนรู้แบบรวมกลุ่มจำนวนมาก

5.2 ข้อเสนอแนะ

งานวิจัยนี้เป็นการระบุข้อมูลเสียงตัดต่อโดยใช้การเรียนรู้แบบรวมกลุ่มโดยทำการแปลงข้อมูลเสียงเป็น 2 ลักษณะ คือ ข้อมูลรูปแบบตัวเลขและข้อมูลภาพ และทำการเปรียบเทียบประสิทธิภาพความถูกต้องกับโมเดลการเรียนรู้อีก 4 โมเดล โดยในการทดลองได้ทำการทดสอบกับอัลกอริทึมการเรียนรู้ของเครื่องเพียง 5 อัลกอริทึม ซึ่งอัลกอริทึมที่มีอยู่หลากหลายทั้งรองรับข้อมูลในลักษณะนี้ และไม่รองรับ ดังนั้นการใช้อัลกอริทึมอื่นมาทดสอบ อาจจะให้ผลที่ดีกว่าหรือเทียบเท่ากับการเรียนรู้แบบรวมกลุ่ม นอกจากนี้ยังสามารถนำข้อมูลเสียงในรูปแบบตัวเลขและรูปแบบภาพไปทำการพัฒนาเพื่อพิสูจน์ข้อมูลเสียงในรูปแบบอื่น

เอกสารอ้างอิง

- [1] สาขาวิชาฟิสิกส์ประยุกต์. คณะวิทยาศาสตร์. มหาวิทยาลัยเทคโนโลยีราชมงคล ธัญบุรี. 2540. **การสั่นและคลื่นเสียง**. [Online]. Available: <http://www.atom.rmutphysics.com/physics/oldfront/89/sound.html>.
- [2] Wikimedia Foundation Inc. 2015. **Audio file format**. [Online]. Available: https://en.wikipedia.org/wiki/Audio_file_format.
- [3] Wikimedia Foundation Inc. 2015. **Spectrogram**. [Online]. Available: <https://en.wikipedia.org/?title=Spectrogram>.
- [4] Wikimedia Foundation Inc. 2015. **Machine Learning**. [Online]. Available: https://en.wikipedia.org/?title=Machine_learning.
- [5] Wikimedia Foundation Inc. 2015. **Supervised Learning**. [Online]. Available: https://en.wikipedia.org/wiki/Supervised_learning.
- [6] Wikimedia Foundation Inc. 2015. **Neural network**. [Online]. Available: https://en.wikipedia.org/wiki/Artificial_neural_network.
- [7] ต่าย บัณฑิตศักดิ์. 2552. “การศึกษาการวัดค่าบิลิรูบินในทารกแรกเกิดทางผิวหนังโดยใช้โครงข่ายประสาทเทียม.” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า, มหาวิทยาลัยสงขลานครินทร์.
- [8] Simon O. Haykin. 1998. Chapter 2. **Neural Networks: A Comprehensive Foundation**.
- [9] Wikimedia Foundation Inc. 2015. **Radial basis function neural network**. [Online]. Available: https://en.wikipedia.org/wiki/Radial_basis_function_network.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [10] Simon O. Haykin. 1998. Chapter 5. **Neural Networks: A Comprehensive Foundation**.
- [11] Wikimedia Foundation Inc. 2015. **Support Vector Machine**. [Online]. Available: https://en.wikipedia.org/wiki/Support_vector_machine.
- [12] Shanshan Li, Bing Zhang, Dongmei Chen, Lianru Gao, Man Peng. 2011. “Adaptive support vector machine and Markov random field model for classifying hyper spectral imagery.” *Journal of Applied Remote Sensing*. Vol.5. Issue 1.
- [13] DTREG. 2014. **SVM – Support Vector Machines**. [Online]. Available: <https://www.dtreg.com/solution/view/20>.
- [14] จันิสรา ศรีสรवल. 2557. “การค้นหาโรงแรมด้วยเทคนิคการรวมกลุ่มแบบเบย์อย่างง่าย.”
วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย,
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- [15] Brigham Young University. 2010. **Decision tree Learning**. [Online]. Available: <http://wiki.cs.byu.edu/cs-677sp2010/decision-theoretic-graphical-models>
- [16] Jiawei Han, Micheline Kamber. 2006. **Data Mining: Concepts and Techniques: 2006**
- [17] Scholarpedia. 2009. **Ensemble Learning Adaboost**. [Online]. Available: http://www.scholarpedia.org/article/Ensemble_learning.
- [18] Mikel Galar, Alberto Fernández, Edurne Barrenechea, Humberto Bustince, Francisco Herrera. 2012. “A Review on Ensemble for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches.” *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS*. 42(4): 463-482.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [19] Dech Thammasiri, Phayung Meesad. 2010. "Ensemble Data Classification Based on Decision Tree, Artificial Neuron network and Support Vector Machine Optimized by Genetic Algorithm." The 6th National Conference on Computing and Information Technology (NCCIT).
- [20] Qingju Liu, Wenwu Wang, Philip Jackson. 2011. "A Visual Activity Detection Method with Adaboosting." Sensor Signal Processing for Defense (SSPD).
- [21] Dech Thammasiri, Vathinee Nuipean, Patrawut Sangsiri, Paranya Aummurkkarat, Narong Pothi, Phayung Meesad. 2009. "Credit Scoring with a Data Mining Approach based on Support Vector Machine and Feature Selection also Setting Model Parameters Using Grid Search." The 5th National Conference on Computing and Information Technology (NCCIT).
- [22] Yogachandran Rahulamathavan, Raphael C.-W. Phan, Suresh Veluru, Kanapathippillai Cumanan, Muttukrishnan Rajarajan. 2013. "Privacy-Preserving Multi-Class Support Vector Machine for Outsourcing the Data Classification in Cloud." IEEE Transactions on Dependable and Secure Computing.
- [23] Duc A. Tran, Tinh Nguyen. 2008. "Localization in Wireless Sensor Network based on Support Vector Machine." IEEE Transactions on Parallel and Distributed Systems. 19(7): 168-175.
- [24] Guilherme Andrade, Felipe Viegas, Gabriel Spada Ramos, Jussara Almeida, Leonardo Rocha, Marcos Goncalves, Renato Ferreira. 2013. "GPU-NB: A Fast CUDA-Based Implementation of Naive Bayes." 25th International Symposium on

Computer Architecture and High Performance Computing. 168.175

- [25] Chris Tseng, Nishant Patel, Hrishikesh Paranjape, T Y Lin. 2012. "Classifying twitter data with Naive Bayes Classifier." IEEE International Conference on Granular Computing. 294-299.
- [26] Dech Thammasiri , Phayung Meesad. 2011. "Adaboost Artificial Neural network Ensemble Learning for Classification." Vol.7 No.14, King Mongkut University of Technology Thonburi Information Technology Journal. 7(4).
- [27] Fang Chu, Carlo Zaniolo, 2004. "Fast and Light Boosting for Adaptive Mining of Data Streams." Lecture Notes in Computer Science, Springer Berlin Heidelberg. 3065: 282-292.
- [28] Xiaoying Jiang, Rong Wei, Yanjun Zhao. 2008. "Using Chou's pseudo acid composition based on approximate entropy and ensemble of AdaBoost classifiers to predict protein subnuclear location." Amino Acids. 34: 669-675.
- [29] Donald F. Specht. 1990. "Probabilistic Neural networks." Neural Networks. 3: 109-118
- [30] Hojjat Adeli, Ashif Panakkat. 2010. "A probabilistic neural network for earthquake magnitude prediction." Neural Network 22: 1018-1024.
- [31] M.R. Mosavi. 2010. "Use of Accurate GPS Timing Based on Radial Basis Probabilistic Neural Network in Electric System." 2010 International Conference on Electrical and Control Engineering

- [32] Shemeey Baluja, Henry A. Rowley. 2007. "Boosting Sex Identification Performance." International Journal of Computer Vision. 71(1): 111-119.
- [33] J. Wang, X. Liao, Z. Yi. 2015. "Robust Precise Eye Location by Adaboost and SVM Techniques." Second International Symposium on Neural Networks (ISNN).
- [34] Roro Fitri Sari, Prasetyawidi Indrawan. 2013. "Cloud Computing Services in Mobile Devices Using Android Face Detector API and Rest Communication." The Third International Conference on Digital Information Processing and Communications (ICDIPC 2013). p. 529-537.
- [35] Walter Cerroni, Gianluca Moro, Tommaso Pirini, Marco Ramilli. 2013. "Peer-to-Peer Data Mining Classifiers for Decentralized Detection of Network Attacks." 24th Australasian Database Conference (ADC 2013). 137: 101-107.
- [36] Sumei Liang, Xinhua Fan. 2014. "Audio Content Classification Method Research Based on Two-Step Strategy." International Journal of Advanced Computer Science and Applications. 5(3).
- [37] Tetsuya Takiguchi, Jun Adachi, Yasuo Arika. 2010. "Video Editing Based on Situation Awareness from Voice Information and Face Emotion." Digital Video book edited by Floriano De Rango.



ภาคผนวก ก
งานวิจัยที่ตีพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**Proceedings of the 2015-7th
International Conference on
Knowledge and Smart Technology**

(KST)



January 28-31, 2015

**Faculty of Informatics, Burapha University
Chonburi, THAILAND.**

ISBN 978-1-4799-6049-1



ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Edited Audio Detection using Ensemble Learning

Takdanai Suwan, Saichon Jaiyen, and Rungrat Wiangsripanawan

Faculty of Science, King Mongkut's Institute of Technology Ladkrabang (KMITL)
Bangkok, Thailand. 10520

Email: takdanai.suwan@hotmail.com, kjsaicho@kmitl.ac.th, kwrunga@kmitl.ac.th

Abstract— Detecting edited audios is the challenge problem that can help forensic scientists to separate genuine, unedited recording from edited recordings. This paper proposes the technique for detecting edited audios using ensemble learning. This problem can be considered as two-class classification problem which audio data are classified into two classes including edited and unedited audios. The performance of the proposed model is compared to Support Vector Machine, Naive Bayes, Radial Basis Function Neural Network, and Probabilistic Neural Networks. The experimental results demonstrate that the proposed model is appropriated for detecting the edited audios.

Keywords- *Boosting Ensemble; Adaboost; Support Vector Machine; SVM; Classification; Audio; Naive Bayes;*

I. INTRODUCTION

Currently, the computers can learn to recognize some useful data in order to classify them into classes. Machine learning techniques are widely used in various ways to teach the systems to improve its accuracy and efficiency. Machine learning method has been used in many ways but most commonly used of this system is classifications. There are many classification problems that can be solved by using machine learning. For example, the data for female characteristics can be skin, hair or age. These characteristic data can be classified into white skin, long or short hair, or colored skin. It depend upon how the system learns. There are three types of learning algorithms including supervised learning, reinforcement learning, and unsupervised learning. Supervised learning will be mentioned and tested in this paper. There are a lot of learning algorithms for classification problems such as support vector machine (SVM) [1], Naive Bayes, decision trees, artificial neural networks, and so on. SVM is one of supervised learning algorithms in machine learning. SVM is widely applied to solve classification problems. For example, SVM has been used for fraud detection by classifying the customers into two primary classes. SVM has also been used in security purposes [3], it can improve the encrypted algorithm by adopting the algorithm as a hash algorithm for data encryption before sending the data to secure the data from being stolen. Additionally, another research have used SVM to locate the Wi-Fi signal [4] and check the distance of the signal source to remember it and use it for better Wi-Fi efficiency used in the next time. Naive Bayes is another supervised learning algorithm that have been used to classify the data on websites and programs [5] to improve GPU (Graphic Processing Unit) and CPU (Central Processing Unit) efficiency. This algorithm can be applied to classify Twitter

data [6] and other social media data. Boosting Ensemble (Decision Tree) algorithm has been used for data stream [8] management. The algorithm will classify data before sending through the data stream in order to make the data transferring faster and more efficient. Moreover, Boosting Ensemble has been used for protein gene data classification [9] to improve the efficiency and accuracy of the classification results. In this research, two algorithms in neural network are used which are probabilistic neural network [10] and radial basis function neural network (RBFNN). These algorithms can be used to classify, analyze and predict the data. M.R. Mosavi applied radial basis neural network algorithm to improve GPS (Global Positioning System) [12] locating. The system will be taught to locate the GPS faster and more accurate in order to improve its efficiency. Many researches have applied Adaboost for classification problems. In image processing, Adaboost has been used for deciding the gender from any image [13]. In visual activity detection, Adaboost has been used for analyzing the movement of the lips to predict the words [14]. Moreover, Adaboost has been applied for locating eyes in the image [15]. In the cloud computing application, Adaboost and image processing have been used for face detection in Android phone in order to connect with the cloud computing to unlock the phone if it is the face of the owner [16]. In the network security, Adaboost has been applied to learn log files from network attacking and identifying attacking types [17]. Sumei Liang and Xinhua Fan's [18] applied SVM and Naive Bayes to the audio content classification. Tetsuya Takiguchi [19] applied Adaboost to learn and investigate the video data whether they were modified or not by using voice recognition and face detection.

In this research, we apply the ensemble learning algorithm to detect the edited audios. The performance of the proposed model will be compared to Support Vector Machine, Naive Bayes, Probabilistic Neural Network, and Radial Basis Neural Network. The audio data used in this research is the talking without any outside noises. The data will be categorized into two types which are unedited audios and edited audios. All algorithms will be conducted using MATLAB software.

II. ENSEMBLE LEARNING

Ensemble learning is the learning algorithm which multiple models are combined to solve a particular problem. It is widely used to improve the performance of a classification model. The ensemble classifier is obtained by combining diverse classification models. In this research we adopt the ensemble learning algorithm to classify the sound clips into unedited and edited sound clips. The ensemble classifier used in this paper is

an ensemble of decision tree classifiers. In this paper, we apply the boosting algorithm to build ensembles of decision trees. In addition, we apply Adaboost algorithm to build the ensemble classifier. Given $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ be a set of N training data where $y_i \in \{-1, +1\}$. The algorithm of Adaboost is demonstrated as follows.

Adaboost Algorithm

1. Initially assign uniform weights $w_0(i) = 1/N$
2. For each iteration k
3. train a base classifier $h_k(x)$ using weights w_i
4. calculate the error rate

$$\varepsilon_k = \sum_i w_k(i) I[h_k(x_i) \neq y_i]$$
5. calculate the weight of the classifier $h_k(x)$:

$$\alpha_k = \frac{1}{2} \log \frac{1 - \varepsilon_k}{\varepsilon_k}$$
6. update weights

$$w_{k+1}(i) = w_k(i) \exp(-\alpha_k I[h_k(x_i) = y_i])$$
 and then normalize weights
7. The final classifier $H(x) = \text{sign}(\sum_k \alpha_k h_k(x))$

The function $[\cdot]$ is the indication function which outputs 1 if the inner expression is true and -1 otherwise.

III. EXPERIMENT AND RESULTS

In this paper, we propose the model for detecting the edited audios using ensemble learning algorithm. For the experiment, the performance of the proposed model is compared to support vector machine (SVM), Naïve Bayes, probabilistic neural network (PNN), and radial basis function neural network (RBFNN). The data set used in this experiment is sound clips provided by Thai Speech Corpus, Speech and Audio Technology Laboratory, National Electronics and Computer Technology Center Thailand. The conversations in the sound clips are in Thai language and were record in the silent room without noises. All experiments are conducted using MATLAB software.

A. Sound Data Preparation

The input data of all algorithms used in this paper must be in the numeric format. Therefore, these data are transferred from .wav format to numerical format in order to feed into all learning algorithms. The wav file format is used because its quality is not compressed. The inputs of all learning algorithms must be the vectors with the same dimensions. However, the sound data length in each clip is not the same and they are in 32 bits rate. Therefore, these data need to be modified to make the dimensions equally and the bits rate is also decreased in order to reduce the dimensions of the input data. The data used in the experiment are categorized into two groups which are unedited sound data and edited sound data. In unedited sound data, each clip is cut into 3 minutes. In addition, the bit rate of the data is decreased from 32 to 8 bit rate in order to reduce the dimensions of these data. In the training set, there are 300 unedited sound clip were

prepared in 8 bits rate. These sound clips are cut and then are shuffled and rearranged in order to make new 300 edited clips. Therefore, there are totally 600 sound clips in the training set. Both unedited and edited sound clips are transformed from .wav format to numerical format. The groups of data mentioned above are taught to machine learning as the target for categorizing the data. The target of unedited sound data, which is the first group, is assigned as 1 and another is assigned as -1. All sound data and targets is used to teach all learning algorithms. In the testing set, there are 300 sound clips which are 150 unedited clips and 150 edited clips.

B. Experimental Results

In the experiment, the performance of the proposed model is evaluated from the testing set and compared with SVM, Naive Bayes, PNN, and RBFNN. For the proposed model, Adaboost algorithm is applied to build ensembles of decision trees. The number of iterations are set as 5, 10, 50, 100, 500 and 1,000. The experimental result is shown in Figure 1. From the result, when the number of iterations is increasing, it gives more accurate results. It give the best accuracy of 100 percent when the number of iterations is 1000.

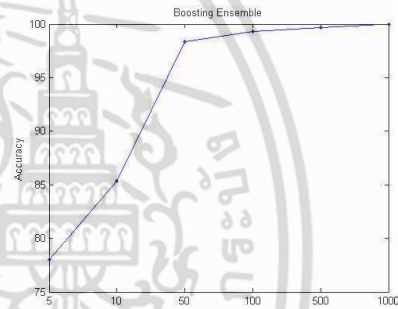


Figure 1. Boosting Ensemble Results

For SVM, the kernel function of SVM is radial basis function. The spread parameters, RBF Sigma, of radial basis function are set to several values and the value that give the maximum accuracy is selected. These values are set as 5, 10, 50, 100, 500 and 1,000. From Figure 2, it can be seen that different values of RBF Sigma can give different accuracies. Since RBF Sigma values of 5, 10 and 50 give the best accuracy of 99.67 percent, the value of 50 is selected to evaluate the performance of SVM.

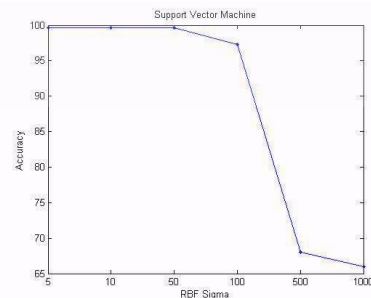


Figure 2. The comparative result of RBF Sigma using SVM

For Naïve Bayes, the multivariate multinomial distribution is applied and the kernel smoothing window bandwidth (KSWidth) parameter is set as 5, 10, 50, 100, 500 and 1,000. From Figure 3, it can be seen that the different KSWidth values give the different accuracies. The best accuracy is 97.33 percent at 1000 KSWidth value. Therefore, the KSWidth value of 1000 is selected to conduct the experiment. For RBFNN, the spread parameters of radial basis functions are set as 5, 10, 50, 500 and 1000. The experimental result for these values is shown in Figure 4. From this Figure, it can be seen that the spread value of 1000 can give the best accuracy at 53.67 percent. Thus, the spread value of 1000 is selected to conduct the experiment.

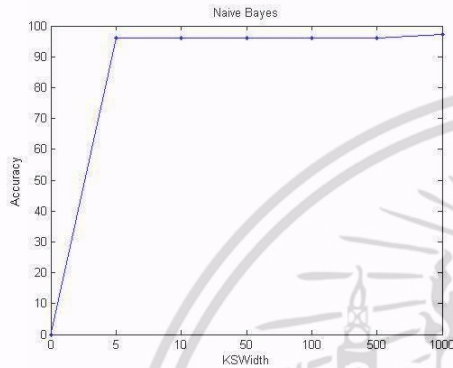


Figure 3. The comparative result of KSWidth using Naïve Bayes

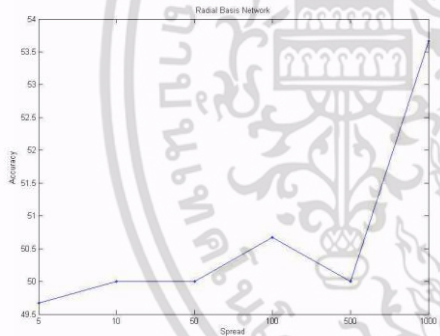


Figure 4. The comparative result of Spread values using RBFNN

For PNN, the spread parameters of PNN is set the same as RBFNN. The experimental result is shown in Figure 5. From Figure 5, it can be seen that the accuracies is not changed whenever the spread values are changed. It give the same accuracy. The best accuracy for PNN is 50 percent.

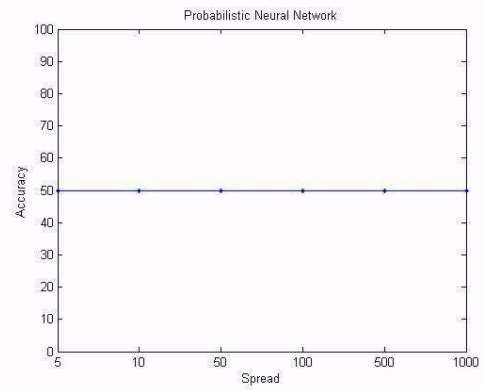


Figure 5. The comparative result of Spread values using PNN

After the parameter value that give the best result in each model is selected, the experiments are conducted with these values. The experimental result is shown in Table 1. Form this table, it can be seen that the ensemble learning with boosting ensemble algorithm gives the best accuracy of a 100 percent. For SVM, the best accuracy comes from the RBF Sigma value of 50 and the best accuracy is 99.67 percent which is higher than Naïve Bayes, FBRNN and PNN. The best accuracy of Naïve Bayes is 97.33 percent which is higher than FBRNN and PNN as shown in Table I. The best accuracy of Naïve Bayes comes from the KSWidth value of 1000. From the experimental result, the accuracies of RBFNN and PNN are not high so it might not be appropriated for sound data categorization. The best accuracy of RBFNN is 53.67 percent which is higher than PNN. This value comes from the Spread value of 1000. The best accuracy of PNN comes from the spread value of 1000 and gives the best accuracy at 50 percent. From the experimental result shown in Table 1, it can be concluded that the proposed model gives the best accuracy at 100 percent. This result show that the ensemble learning algorithm is suitable for detecting the edited sound audios. For further works, we may adopt the algorithm to detecting the edited sound audios with noises. In addition, the feature extraction methods should be applied to reduce the dimensions of the data in order to reduce the computational times.

Table I: The comparative accuracies of all models

Algorithms	Accuracy
Ensemble Learning	100.00
SVM	99.67
Naïve Bayes	97.33
RBFNN	53.67
PNN	50.00

IV. CONCLUSION

In this paper, we propose the model to detect the edited sound audios using ensemble learning. There are 300 unseen sound clips used to test all models in the experiment. These clips are divided into two classes which are unedited and edited sound clips. The performance of the proposed model is evaluated and compared to Support Vector Machine, Naïve Bayes, Radial Basis Function Neural Network, and Probabilistic Neural Network. From the experiential result, the proposed model, Support Vector Machine, and Naïve Bayes give the high accuracy at 100, 99.67, and 97.33, respectively. So, these algorithms can be applied to detect the edited sound audios. However, the experimental result shows that the proposed model gives the best accuracy at 100 percent. This result can identify that the ensemble learning is appropriated for classifying sound clips into unedited and edited sound clips more than Support Vector Machine and Naive Bayes. Therefore, the proposed model can help forensic scientists to separate genuine, unedited recording from edited recordings.

REFERENCES

- [1] Dech Thammasiri, Phayung Meesad, "Ensemble Data Classification Based on Decision Tree, Artificial Neuron network and Support Vector Machine Optimized by Genetic Algorithm," The 6th National Conference on Computing and Information Technology (NCCIT), 2010
- [2] Dech Thammasiri, Vathinee Nuipean, Pattrawut Sangsiri, Paranya Aummurkkarat, Narong Pothi, Phayung Meesad, "Credit Scoring with a Data Mining Approach based on Support Vector Machine and Feature Selection also Setting Model Parameters Using Grid Search," The 5th National Conference on Computing and Information Technology (NCCIT), 2009.
- [3] Yogachandran Rahulamathavan, Raphael C.-W. Phan, Suresh Veluru, Kanapathippillai Cumanan, Mutukrishnan Rajarajan, "Privacy-Preserving Multi-Class Support Vector Machine for Outsourcing the Data Classification in Cloud," IEEE Transactions on Dependable and Secure Computing, 17 Dec 2013.
- [4] Duc A. Tran, Thinh Nguyen, "Localization in Wireless Sensor Network based on Support Vector Machine," Vol.19 No.7, pp. 981-994, IEEE Transactions on Parallel and Distributed Systems, pp.168.175, July 2008
- [5] Guilherme Andrade, Felipe Viegas, Gabriel Spada Ramos, Jussara Almeida, Leonardo Rocha, Marcos Goncalves, Renato Ferreira, "GPU-NB: A Fast CUDA-Based Implementation of Naive Bayes," 2013 25th International Symposium on Computer Architecture and High Performance Computing, pp.168.175, 2013.
- [6] Chris Tseng, Nishant Patel, Hrishikesh Paranjape, T Y Lin, "Classifying twitter data with Naive Bayes Classifier," 2012 IEEE International Conference on Granular Computing, pp. 294-299, 2012.
- [7] Dech Thammasiri, Phayung Meesad, "Adaboost Artificial Neural network Ensemble Learning for Classification," Vol.7 No.14, King Mongkut University of Technology Thonburi Information Technology Journal, July - December 2011
- [8] Fang Chu, Carlo Zaniolo, "Fast and Light Boosting for Adaptive Mining of Data Streams," Volume 3056, Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 282-292, 2004
- [9] Xiaoying Jiang, Rong Wei, Yanjun Zhao, "Using Chou's pseudo acid composition based on approximate entropy and ensemble of AdaBoost classifiers to predict protein subnuclear location," pp. 669-675, Vol.34, Amino Acids, 2008
- [10] Donald F. Specht, "Probabilistic Neural networks," Vol.3, Neural Networks, pp. 109-118, 1990
- [11] Hojjat Adeli, Ashif Panakkt "A probabilistic neural network for earthquake magnitude prediction," 2009 Neural Network 22, pp. 1018-1024, 2010
- [12] M.R. Mosavi, "Use of Accurate GPS Timing Based on Radial Basis Probabilistic Neural Network in Electric System," 2010 International Conference on Electrical and Control Engineering, pp. 2572-2575, 2010
- [13] Shemeey Baluja, Henry A. Rowley, "Boosting Sex Identification Performance", Vol.71, Issue 1, International Journal of Computer Vision, p.111-119, Jan 2007.
- [14] Qingju Liu, Wenwu Wang, Philip Jackson "A Visual Activity Detection Method with Adaboosting", Sensor Signal Processing for Defence (SSPD), 2011.
- [15] J. Wang, X. Liao, Z. Yi, "Robust Precise Eye Location by Adaboost and SVM Techniques", Second International Symposium on Neural Networks (ISNN), 2015.
- [16] Roro Fitri Sari, Prasetyawidi Indrawan, "Cloud Computing Services in Mobile Devices Using Android Face Detector API and Rest Communication", p. 529-537, The Third International Conference on Digital Information Processing and Communications (ICDIPC 2013), Feb 2013.
- [17] Walter Cerroni, Gianluca Moro, Tommaso Pirini, Marco Ramilli, "Peer-to-Peer Data Mining Classifiers for Decentralized Detection of Network Attacks", Vol. 137, p. 101-107, 24th Australasian Database Conference (ADC 2013), 2013
- [18] Sumei Liang, Xinhua Fan, "Audio Content Classification Method Research Based on Two-Step Strategy," Vol.5, No.3, International Journal of Advanced Computer Science and Applications, 2014
- [19] Tetsuya Takiguchi, Jun Adachi, Yasuo Arika, "Video Editing Based on Situation Awareness from Voice Information and Face Emotion", Chapter 18, Digital Video book edited by Floriano De Rango, Feb 2010

ประวัติผู้เขียน

ชื่อ นายทักษ์ดนัย สุวรรณ
วัน เดือน ปีเกิด 16 เมษายน 2531
ที่อยู่ปัจจุบัน 41/253 หมู่บ้านพรธิสาร 8 ม.1 ต.คลองเจ็ด อ.คลองหลวง จ.ปทุมธานี
ประวัติการศึกษา (2553) วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ เกเรดเฉลี่ย 2.60
มหาวิทยาลัยเทคโนโลยีราชมงคล ธัญบุรี
(2557) วิทยาศาสตรมหาบัณฑิต สาขา วิทยาการคอมพิวเตอร์
เกเรดเฉลี่ย 3.25
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ผลงานทางวิชาการ 1. 7th International Conference on Knowledge and Smart Technology
2015 (KST)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้