สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ต้นแบบวิธีการจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน

A PROTOTYPE METHOD FOR CLASSIFICATION OF OVERLAPPING DATA

กุลวลัญช์ วรุณสิน
KULWARUN WARUNSIN

วพ.
ก7728ต
2548
เลขหมู่.....
เลขทะเบียน....60859.
วัน,เดือน,ปี....6 ก.ค. 2549

.b 11590944
.i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2548

ISBN 974-15-1424-7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2005

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	ต้นแบบวิธีการจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน
นักศึกษา	นางสาวกุลวลัญช์ วรุณสิน
รหัสประจำตัว	44061619
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
พ.ศ.	2548
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.บุญธีร์ เครือตราชู

บทคัดย่อ

วิทยานิพนธ์นี้ นำเสนอแนวความคิดต้นแบบสำหรับใช้ในการจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน(ข้อมูลที่มีคุณลักษณะ (feature) เหมือนกันบางส่วนแต่ถูกจัดอยู่ในกลุ่มที่ต่างกัน) ซึ่งต้นแบบที่นำเสนอนี้ใช้หลักการการจัดกลุ่มด้วยวิธี Neural Network Classifier (NNC)[1] โดยเริ่มต้นจากการนำข้อมูลที่ระบุกลุ่มแน่นอนแล้วจำนวนหนึ่ง มาผ่านขั้นตอนการเรียนรู้ (training) ด้วยหลักการที่นำเสนอในวิทยานิพนธ์นี้เรียกว่าหลักการแบ่งกลุ่มด้วยวิธี NNC ได้ต้นแบบสำหรับกลุ่มข้อมูลตามจำนวนกลุ่มที่มีก่อนขั้นตอนการเรียนรู้ โดยต้นแบบแต่ละกลุ่มประกอบด้วยจุดศูนย์กลางหรือน้ำหนักของกลุ่มหนึ่งค่า รัศมีของกลุ่มสองค่า และต้นแบบกลุ่มต่างๆ ซึ่งจะมีจุดศูนย์กลางและขอบเขตของกลุ่มต่างกันออกไป สำหรับการนำไปใช้งานคือเมื่อมีข้อมูลที่ยังไม่ได้ระบุกลุ่มเข้ามาทดสอบจะทำการวัดระยะทางของข้อมูลนี้กับจุดศูนย์กลางของแต่ละกลุ่ม ถ้าระยะทางอยู่ภายในค่ารัศมีที่สั้นกว่า (ค่า lower) ของกลุ่มนั้น แสดงว่าข้อมูลนั้นมีความเป็นไปได้มากที่จะจัดอยู่ในกลุ่มนี้ ถ้าระยะทางอยู่ระหว่างค่ารัศมีที่สั้นกว่าและค่ารัศมีที่ยาวกว่า (lower และ upper) แสดงว่าข้อมูลนั้นมีความเป็นไปได้ที่จะจัดอยู่ในกลุ่มนี้หรือกลุ่มอื่น ๆ แต่ถ้าระยะทางอยู่นอก upper แสดงว่าข้อมูลนั้นไม่ได้จัดอยู่ในกลุ่มนี้ ซึ่งเราขอยืมความหมายของคำศัพท์คำว่า lower และ upper จากทฤษฎี rough set [3] มาใช้อธิบายการกำหนดพื้นที่ต่างๆ ของต้นแบบกลุ่มที่นำเสนอในวิทยานิพนธ์นี้ สำหรับผลการทดลองได้จากการนำข้อมูลตัวอักษรภาษาไทยมาทดสอบด้วยหลักการที่นำเสนอในวิทยานิพนธ์และนำผลการทดลองมาเปรียบเทียบกับหลักการแบ่งกลุ่มด้วย NNC[1] พบว่าการแบ่งกลุ่มที่นำเสนอในวิทยานิพนธ์นี้สามารถแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกันได้เหมาะสมกว่าการแบ่งกลุ่มด้วยหลักการของ NNC [1] เมื่อพิจารณาจากการแบ่งกลุ่มด้วยขอบเขตของข้อมูล (data clustering boundary:density)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	A Prototype method for classification of overlapping data
Student	Miss. Kulwarun Warunsin
Student ID	44061619
Degree	Master of Engineering
Programme	Computer engineering
Year	2005
Thesis Advisor	Assoc. Prof. Dr.Boontee Kruatrachue

ABSTRACT

In this thesis we introduce a new prototype based classifiers for overlapping data, where training pattern are overlap on the feature space. The proposed classifier is based on the prototype from neural network classifier (NNC)[1] for overlap data. The method automatically chooses the initial center and two radiuses for each class. The center is used as a mean representative of training data for each class. The unclassified pattern is classified by measure distance from the class center. If the distance is in the lower (shorter radius) the unknown pattern has the high percentage of being in this class. If the distance is between the lower and upper (further radius), the pattern has the probability of being in this class or others. But if the distance is outside the upper, the pattern is not in this class. We borrow the words upper and lower from the rough set to represent the region of certainty [3]. The training algorithm to find number of cluster and their parameters (center, lower, upper) is presented. The clustering result is tested using patterns from Thai handwritten letter and the clustering result is very similar to human eyes clustering.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้จะไม่ลุล่วงหากปราศจากแรงผลักดันและคำแนะนำที่มีประโยชน์ของ รศ.ดร.บุญธีร์ เครือตราชู ผู้ควบคุมวิทยานิพนธ์ ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ รศ.ดร.เอื้อน ปิ่นเงิน ดร.สมศักดิ์ วลัยรัชต์ ผศ.กฤตวัน ศิริบูรณ์ ดร.อรฉัตร จิตต์โสภักตร์ กรรมการสอบหัวข้อและโครงร่างวิทยานิพนธ์ที่ได้กรุณาให้คำแนะนำตลอดจนข้อชี้แนะ จนในที่สุดทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้

ขอขอบคุณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยรามคำแหง ที่มีส่วนให้การสนับสนุนงานวิจัยนี้

ขอขอบคุณบริษัท อาร์ แอนด์ ดี คอมพิวเตอร์ ซิสเท็ม จำกัด ที่มอบโอกาสให้ข้าพเจ้าได้พัฒนาความรู้ความสามารถของข้าพเจ้า จนสามารถเข้าศึกษาและสำเร็จการศึกษาระดับบัณฑิตศึกษา

ข้าพเจ้าขอขอบคุณสำหรับกำลังใจ คำแนะนำ และประสบการณ์ที่ดีจากเพื่อนๆ น้องๆ นักศึกษา ป.โททุกท่าน ขอขอบคุณ คุณณรงค์ชัย มุ่งแฝงกลาง ที่ช่วยเหลือให้คำแนะนำต่างๆ พร้อมทั้งช่วยตรวจเทียบและแก้ไขทฤษฎีอื่นๆ ที่ผิดพลาดจนสมบูรณ์ยิ่งขึ้น และขอขอบคุณ นางสาวกนิษฐา ปัญญาภรณ์ เพื่อนแท้ของข้าพเจ้าที่เป็นทั้งผู้ให้คำปรึกษาและมีส่วนในการสนับสนุนงานวิจัยนี้

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูอาจารย์ที่เคารพทุกท่าน หากวิทยานิพนธ์ฉบับนี้มีข้อผิดพลาดประการใด ข้าพเจ้าขอน้อมรับเพียงผู้เดียว

กุลวลัญช์ วรุณสิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความหมายของข้อมูลที่มีการซ้อนทับกัน 1

1.3 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.4 สมมุติฐานของการศึกษา 3

1.5 ทฤษฎีหรือแนวความคิดที่ใช้ในงานวิจัย 3

1.5.1 การกำหนดขอบเขตของกลุ่มข้อมูลต้นแบบ 5

1.5.2 การนำต้นแบบมาจัดกลุ่มข้อมูลที่ซ้อนทับกัน 5

1.5.3 การนำหลักการ NNC มาจัดกลุ่มข้อมูลต้นแบบ 6

1.5.4 การนำงานวิจัยในวิทยานิพนธ์ไปใช้งาน 6

1.6 ขอบเขตของงานวิจัย 7

1.7 ขั้นตอนของการศึกษา 7

1.8 รายละเอียดในแต่ละบท 7

บทที่ 2 การจัดกลุ่มข้อมูล 9

2.1 บทนำ 9

2.2 ขั้นตอนการจัดกลุ่ม 10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

2.3 ทฤษฎีที่ใช้ในการจัดกลุ่ม....11

2.3.1 ทฤษฎีต่าง ๆ ของการจัดหมวดหมู่ (Classification)....12

2.3.1.1 ดีซีชันทรี (Decision Tree)....12

2.3.1.2 Bayesian Classification....12

2.3.1.3 นิวรอนเน็ดเวิร์ก (Neuron Network: NN)....14

2.3.1.4 จีเนติก อัลกอริทึม (Genetic Algorithm:GA)....15

2.3.2 ทฤษฎีต่าง ๆ ของการรวมตัว (Clustering)....15

2.3.2.1 Hierarchical....15

2.3.2.2 Partition....16

2.3.2.3 Density-base....17

บทที่ 3 NNC Architecture (Neural Network Classification Architecture)....19

3.1 สถาปัตยกรรมของ NNC....19

3.2 อัลกอริทึมของ NNC (NNC training algorithm)....20

3.3 ผลการจัดกลุ่มโดยใช้อัลกอริทึมของ NNC กับข้อมูลที่ไม่มีการซ้อนทับกัน....24

บทที่ 4 การนำเสนอต้นแบบของการจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน....25

4.1 ต้นแบบของกลุ่มข้อมูล....25

4.1.1 การปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน....27

4.1.2 การปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน....27

4.2 อัลกอลิทึมที่ใช้ในงานวิจัย....32

4.3 ปัญหาและการปรับปรุงงานวิทยานิพนธ์....33

บทที่ 5 ผลการทดลอง....36

5.1 การทดลองที่ 1....36

5.2 การทดลองที่ 2....38

5.3 การทดลองที่ 3....46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ 49
6.1 สรุปผลการวิจัย 49
6.2 ข้อเสนอแนะ 50

เอกสารอ้างอิง 52

ภาคผนวก ก ข้อมูลในไฟล์ data012_4.txt 54
ภาคผนวก ข ข้อมูลในไฟล์ data011.txt 59
ภาคผนวก ค งานวิจัยที่ได้รับการตีพิมพ์ 66

ประวัติผู้เขียน 72

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 แสดงข้อมูลการเรียนรู้ (Training) กลุ่มลูกค้าที่ตัดสินใจซื้อคอมพิวเตอร์ 13

4.1 แสดงผลการทดลองจัดกลุ่มข้อมูลกลุ่ม ฒ ของลายมือเขียนตัวอักษรภาษาไทย 33

4.2 แสดงผลการทดลองจากโปรแกรม NNC version 1.14.....34

4.3 แสดงการทดลองจากโปรแกรม NNC version 1.14 สำหรับกลุ่มที่ 7 34

5.1 แสดงตัวอย่างกลุ่มข้อมูล 2 กลุ่ม จำนวน 13 ข้อมูล 2 คุณลักษณะ.....37

5.2 แสดงข้อมูลเหมือนตารางที่ 5.1 แต่มีการเปลี่ยนลำดับ 38

5.3 แสดงการทดลองการจัดกลุ่มข้อมูลของลายมือเขียนตัวอักษรภาษาไทย 354 ข้อมูล 39

5.4 ผลการนำข้อมูลชุดที่ 2 (test data) มาทดสอบกับผลการทดลองที่ 2.....43

5.5 ตัวอย่างพยัญชนะที่ใช้ในการทดสอบ.....46

5.6 ผลการทดลองการแบ่งกลุ่มพยัญชนะภาษาไทยด้วย MNNC.....47

ก.1 ข้อมูลในไฟล์ data012_4.txt.....55

ข.1 ข้อมูลในไฟล์ data011.txt.....60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

1.1 แสดงข้อมูลซ้อนทับกันของกลุ่มผู้หญิงและผู้ชาย 2

1.2 แสดงการแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกัน 2

1.3 แสดงการเปรียบเทียบ prototype ต้นแบบในงานวิจัยของ M.A. ABOU-NASR (ซ้าย) และ prototype ต้นแบบของงานวิจัยนี้ (ขวา) 4

1.4 แสดงการกำหนดพื้นที่ของส่วน lower และ upper สำหรับข้อมูลซ้อนทับกัน 4

1.5 แสดงการนำข้อมูลตัวอย่างไปใช้สร้าง Classifier ต้นแบบ 4

1.6 แสดงการกำหนดพื้นที่ของ lower และ upper สำหรับข้อมูลซ้อนทับกัน 5

1.7 แสดงการแบ่งข้อมูลที่ซ้อนทับกันเป็นกลุ่มด้วยต้นแบบงานวิจัยของงานวิจัยนี้ 5

1.8 แสดงการนำข้อมูลตัวอย่างไปใช้สร้าง Classifier ต้นแบบ 6

1.9 แสดงการนำ Classifier ต้นแบบไปใช้งาน 7

2.1 แสดงโครงสร้างของการจัดกลุ่ม 10

2.2 แสดงการนำข้อมูลตัวอย่างไปใช้สร้าง Classifier ต้นแบบ 11

2.3 แสดงการนำ Classifier ต้นแบบไปใช้งาน 11

2.4 แสดงดิซีชั่นทรีสำหรับการตัดสินใจซื้อเครื่องคอมพิวเตอร์ของลูกค้ากลุ่มหนึ่ง 12

2.5 แสดงแผนภาพเดนโดแกรมของการจัดกลุ่มแบบ hierarchical 16

2.6 แสดงแผนภาพกราฟ xy ของข้อมูล 1-5 และค่าเริ่มต้นของกลุ่มเป็นรูปกากบาท 17

2.7 แสดงข้อมูลที่ต้องการจัดกลุ่ม 18

2.8 แสดงการนำข้อมูลจากรูปที่ 2.7 มาทำการจัดกลุ่มข้อมูลแบบ Density-base 18

3.1 สถาปัตยกรรมของ NNC 19

3.2 NNC training algorithm 21

3.3 NNC training algorithm กรณีที่ 1 22

3.4 NNC training algorithm กรณีที่ 2 23

3.5 NNC training algorithm กรณีที่ 3 23

3.6 NNC training algorithm กรณีที่ 4 24

3.7 ตัวอย่างการจัดข้อมูลด้วย NNC สำหรับข้อมูลที่ไม่มีการซ้อนทับกัน 24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่ หน้า

4.1 แสดงการเปรียบเทียบ prototype ต้นแบบในงานวิจัยของ M.A. ABOU-NASR (ซ้าย) และ prototype ต้นแบบของงานวิจัยนี้ (ขวา) 25
4.2 แสดงการกำหนดพื้นที่ของส่วน lower และ upper สำหรับข้อมูลซ้อนทับกัน 26
4.3 แสดงข้อมูลที่กำลังทดสอบเป็นกลุ่มเดียวกับกลุ่มทดสอบ 27
4.4 แสดงข้อมูลที่กำลังทดสอบต่างกลุ่มกับกลุ่มทดสอบ 27
4.5 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Lower 28
4.6 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Upper 28
4.7 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่นอกพื้นที่ Upper (ระยะทางน้อยกว่า 1.2 เท่าของ Upper) 29
4.8 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่นอกพื้นที่ Upper (ระยะมากกว่า 1.2 เท่าของ Upper) 29
4.9 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Lower 30
4.10 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Upper 31
4.11 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน กรณีข้อมูลใหม่อยู่นอกพื้นที่ Upper 31
4.12 อัลกอริทึมการจัดข้อมูลที่มีการซ้อนทับกัน 32
5.1 แสดงข้อมูลก่อนการจัดกลุ่ม และการจัดกลุ่มด้วยวิธี MNNC เปรียบเทียบกับ NNC 37
5.2 แสดงผลการจัดกลุ่มด้วยวิธี MNNC เฉพาะกลุ่มที่มีค่า Cetatin = 100 %บางส่วน 42
5.3 แสดงผลการจัดกลุ่มด้วยวิธี MNNC เฉพาะกลุ่มที่มีค่า Cetatin < 100 %บางส่วน 42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่ หน้า

5.4 ตัวอย่างข้อมูลตัวอักษรภาษาไทย 6 กลุ่มที่ใช้ในการจัดกลุ่ม....47

5.5 ผลการทดลองวิธีการจัดกลุ่มด้วยหลักการ NNC....48

5.6 ผลการทดลองวิธีการจัดกลุ่มด้วยหลักการ NNC และต้นแบบที่ได้นำเสนอในวิทยานิพนธ์...48

6.1 ภาพแสดงข้อมูลสำหรับใช้ในการจัดกลุ่ม....50

6.2 ภาพแสดงการจัดกลุ่มที่นำเสนอในงานวิจัยนี้....51

6.3 ภาพแสดงการจัดกลุ่มแบบที่ 1....51

6.4 ภาพแสดงการจัดกลุ่มแบบที่ 2....51

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

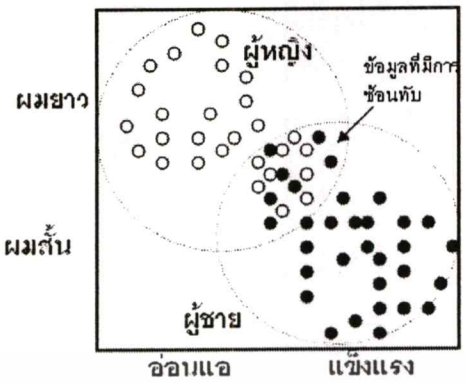
1.1 ความเป็นมาและความสำคัญของปัญหา

วิธีการจัดกลุ่มข้อมูลเป็นสิ่งสำคัญในการประยุกต์ใช้งานในหลายๆ ด้านไม่ว่าจะเป็นงานทางด้านภูมิศาสตร์ เช่น การวิเคราะห์การสั่นสะเทือนของแผ่นดิน คลื่นเสียง การตรวจจับข้อมูลด้วยสัญญาณเรดาร์ งานทางด้านวิทยาศาสตร์ วิศวกรรมศาสตร์ เช่น การรู้จำคำพูดและตัวอักษร ระบบการตรวจจับ และเทคนิคนิวรอน (neural) เป็นเทคนิคหนึ่งที่ถูกนำมาวิจัยประยุกต์ใช้งานเพื่อลดอัตราความผิดพลาดของการจัดกลุ่ม โดยวิธีการจัดกลุ่มวิธีนี้ถูกปรับปรุงให้มีประสิทธิภาพของการรู้จำมากขึ้น และมีงานวิจัยหลายๆ งานที่นำเสนอวิธีการจัดกลุ่มด้วยเทคนิคนิวรอน [7-12] และงานวิจัยเกี่ยวกับการใช้เทคนิคนิวรอนมาใช้ในการจัดกลุ่มที่น่าสนใจคือ Neural Network Classifier (NNC) โดย M.A. ABOU-NASR [1] แต่งานวิจัยดังกล่าวยังไม่สามารถใช้เทคนิค NNC ในการจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน ประกอบกับการศึกษางานวิจัยด้านการจัดกลุ่มข้อมูลที่มีการซ้อนทับกันยังมีไม่มากนัก ดังนั้นในวิทยานิพนธ์ฉบับนี้จึงนำเทคนิคของ NNC และเสนอต้นแบบการจัดกลุ่มรูปแบบหนึ่งมาใช้ในการจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน(สำหรับความหมายของข้อมูลที่มีการซ้อนทับกันสามารถอธิบายเพิ่มเติมในหัวข้อ 1.2) สามารถแบ่งขั้นตอนการจัดกลุ่มข้อมูลแบ่งเป็นสองขั้นตอน โดยขั้นตอนแรกเป็นการสร้างต้นแบบของกลุ่มเพื่อกำหนดลักษณะและขอบเขตของข้อมูลตามข้อมูลที่ใช้ทดสอบเรียกว่าขั้นตอนการเรียนรู้ (training) ส่วนขั้นตอนที่สองเป็นการนำผลลัพธ์ที่ได้จากขั้นตอนแรกมาใช้งานนั่นคือการหากลุ่มให้กับข้อมูลที่ไม่ทราบกลุ่มแต่ทราบลักษณะของข้อมูลที่เรียกว่าขั้นตอนการทำนาย (prediction)

1.2 ความหมายของข้อมูลที่มีการซ้อนทับกัน

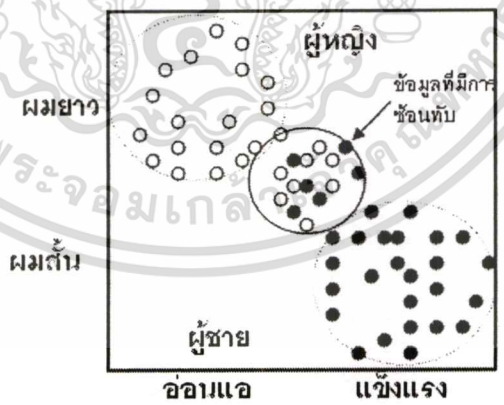
การจัดกลุ่มข้อมูลสามารถแยกได้ 2 ประเภทคือ การจัดหมวดหมู่ (Classification) และการรวมตัว (Clustering) การจัดหมวดหมู่คือการจำแนกประเภทการจัดกลุ่มแบบมีผู้สอน (Supervised) ซึ่งกลุ่มข้อมูลจะถูกกำหนดไว้ล่วงหน้าแล้ว ส่วนการรวมตัวคือการจำแนกประเภทของการจัดกลุ่มแบบไม่มีผู้สอน (Unsupervised) คือไม่มีข้อมูลใดๆ ของกลุ่มที่ถูกกำหนดไว้ก่อน สำหรับข้อมูลที่มีการซ้อนทับกันสามารถอธิบายในประเภทของการจัดหมวดหมู่คือข้อมูลที่มีคุณลักษณะ (feature) เหมือนกันทั้งหมดหรือบางส่วนแต่ถูกจัดอยู่ในกลุ่มที่ต่างกัน เช่น การจัดกลุ่มผู้หญิง และกลุ่มผู้ชาย ตามลักษณะภายนอก โดยสามารถใช้คุณลักษณะแยกได้ดังนี้ ผู้ชายมีคุณลักษณะ แข็งแรง ผมสั้น ส่วนผู้หญิงมีคุณลักษณะ อ่อนแอ ผมยาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 1.1 แสดงข้อมูลซ้อนทับกันของกลุ่มผู้หญิงและผู้ชาย

พิจารณารูปที่ 1.1 เป็นการแบ่งกลุ่มข้อมูลจากคุณลักษณะที่กล่าวมาเบื้องต้นได้ 2 กลุ่ม คือกลุ่มผู้ชายและกลุ่มผู้หญิง และในรูปที่ 1.1 ส่วนที่ไม่สามารถระบุได้ว่าเป็นกลุ่มผู้หญิงหรือผู้ชาย เนื่องจากข้อมูลปนกันอยู่เราเรียกว่าส่วนของข้อมูลที่มีการซ้อนทับกัน และถ้าต้องการจัดกลุ่มส่วนที่มีการซ้อนทับกันนี้จะได้ดังรูปที่ 1.2 คือแยกเป็นกลุ่มเพิ่มอีกหนึ่งกลุ่ม ดังนั้นถ้าเรามีข้อมูลตกในพื้นที่กลุ่มที่เพิ่มใหม่นี้เราสามารถระบุได้ว่าเป็นกลุ่มผู้หญิงหรือกลุ่มผู้ชายก็ได้ แต่ถ้าข้อมูลตกในพื้นที่กลุ่มผู้หญิงอย่างเดียวเราก็สามารถระบุได้ว่าเป็นผู้หญิงแน่นอน ทำนองเดียวกันกลุ่มผู้ชายก็เช่นกัน



รูปที่ 1.2 แสดงการแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกัน

1.3 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

วิทยานิพนธ์นี้ มีความประสงค์ที่จะนำเสนอแนวความคิดต้นแบบการจัดกลุ่มใหม่กับหลักการการจัดกลุ่มด้วยวิธี Neural Network Classifier (NNC) เพื่อจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน โดยมีการศึกษาต้นแบบเพื่อกำหนดพื้นที่ต่างๆ ของต้นแบบที่ใช้ในการจัดกลุ่มข้อมูล

เอกสารนี้มีลิขสิทธิ์ ห้ามนำไปใช้เพื่อการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และศึกษาการทำงานของ Neural Network เพื่อนำมาใช้จัดกลุ่มข้อมูลที่มีการซ้อนทับกัน นำผลการจัดกลุ่มไปเปรียบเทียบกับการจัดกลุ่มด้วย NNC [1]

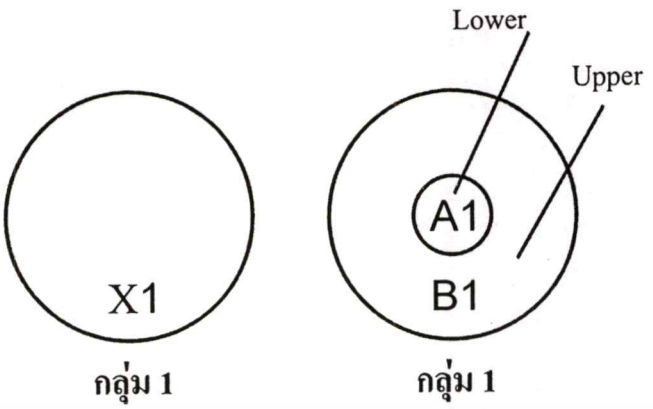
1.4 สมมุติฐานของการศึกษา

ฐานข้อมูลต่างๆ ที่มีอยู่ในปัจจุบันมีความเป็นไปได้สูงที่จะมีข้อมูลที่มีการซ้อนทับกัน ดังนั้นเพื่อเป็นประโยชน์ต่อการนำไปใช้งานจริง จึงคาดหวังให้การนำเสนอวิทยานิพนธ์ฉบับนี้สามารถนำข้อมูลตัวอย่างบางส่วนไปสร้างต้นแบบของกลุ่มได้อย่างเหมาะสม สามารถอธิบายขอบเขตของกลุ่มข้อมูลแต่ละกลุ่มได้อย่างชัดเจนและครอบคลุม อธิบายขอบเขตของข้อมูลที่มีการซ้อนทับกันได้ และสามารถนำไปใช้ประโยชน์ต่อขั้นตอนการทำนายเพื่อวิเคราะห์และวางแผนงานต่างๆ ได้ไม่ว่าจะเป็นงานทางด้านวิทยาศาสตร์ วิศวกรรมศาสตร์ หรืองานในด้านอื่นๆ

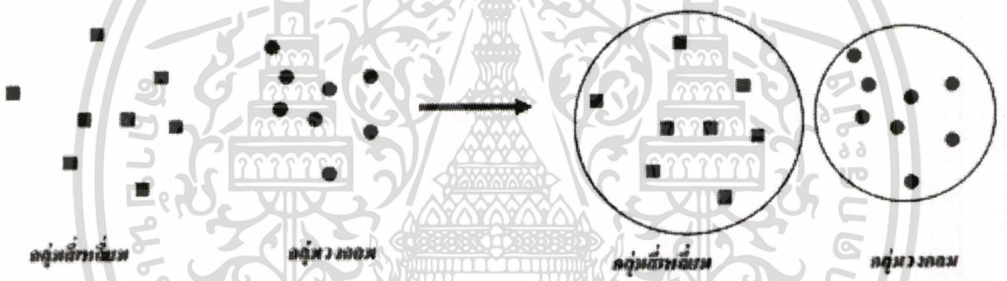
1.5 ทฤษฎีหรือแนวความคิดที่ใช้ในงานวิจัย

วิทยานิพนธ์นี้นำเสนอแนวความคิดในการจัดกลุ่มข้อมูลที่มีการซ้อนทับเรียกว่า Modified Neural Network Classifier (MNNC) โดยมีการนำเสนอต้นแบบการจัดกลุ่มใหม่กับหลักการการจัดกลุ่มด้วยวิธี Neural Network Classifier (NNC) เพื่อแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกัน สำหรับแนวความคิดต้นแบบได้มีการปรับปรุงจากต้นแบบการจัดกลุ่มของ M.A. ABOU-NASR [1] ที่นำเสนอต้นแบบกลุ่มหนึ่งประกอบด้วยจัดศูนย์กลางหนึ่งค่าและรัศมีหนึ่งค่า ดังรูปที่ 1.3 ด้านซ้าย หมายความว่าข้อมูลที่อยู่ในพื้นที่ X1 เป็นข้อมูลของกลุ่ม 1 และในกรณีที่ข้อมูลไม่ทราบกลุ่มตกในพื้นที่ของรัศมีของกลุ่มใดๆ แสดงว่าเป็นข้อมูลนั้นเป็นสมาชิกของกลุ่มนั้น ในรูปที่ 1.4 แสดงการแบ่งกลุ่มข้อมูลที่แบ่งด้วยวิธีของ M.A. ABOU-NASR ด้านซ้ายของรูปที่ 1.4 เป็นข้อมูลตัวอย่างและด้านขวาเป็นกลุ่มที่ถูกแบ่งโดยวิธีของ M.A. ABOU-NASR จะเห็นว่าข้อมูลที่ใช้แบ่งในรูปที่ 1.4 ไม่มีการซ้อนทับกัน แต่ถ้าข้อมูลมีการซ้อนทับกันในรูปที่ 1.5 ด้านซ้าย การแบ่งกลุ่มข้อมูลที่ซ้อนทับกันด้วยวิธีของ M.A. ABOU-NASR จะถูกแสดงดังรูปที่ 1.5 ด้านขวา พบว่าการแบ่งกลุ่มในรูปที่ 1.5 มีการกระจายเป็นกลุ่มย่อยในการนำไปใช้งานในขั้นตอนการทำนายอาจมีความผิดพลาดได้ ดังนั้นวิทยานิพนธ์นี้นี้จึงนำเสนอต้นแบบดังแสดงในรูปที่ 1.3 ด้านขวา

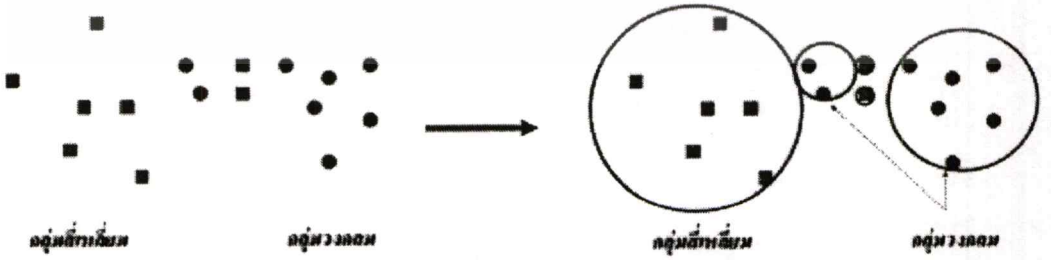
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 1.3 แสดงการเปรียบเทียบ ต้นแบบการจัดกลุ่มในงานวิจัยของ M.A. ABOU-NASR (ซ้าย) และ ต้นแบบการจัดกลุ่มของงานวิจัยนี้ (ขวา)



รูปที่ 1.4 แสดงการแบ่งข้อมูลเป็นกลุ่มด้วยต้นแบบงานวิจัยของ M.A. ABOU-NASR



รูปที่ 1.5 แสดงการแบ่งข้อมูลที่ซ้อนทับกันด้วยต้นแบบงานวิจัยของ M.A. ABOU-NASR

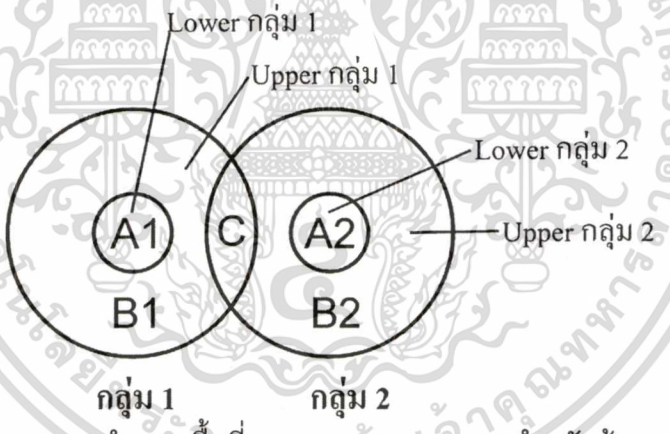
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5.1 การกำหนดขอบเขตของกลุ่มข้อมูลต้นแบบ

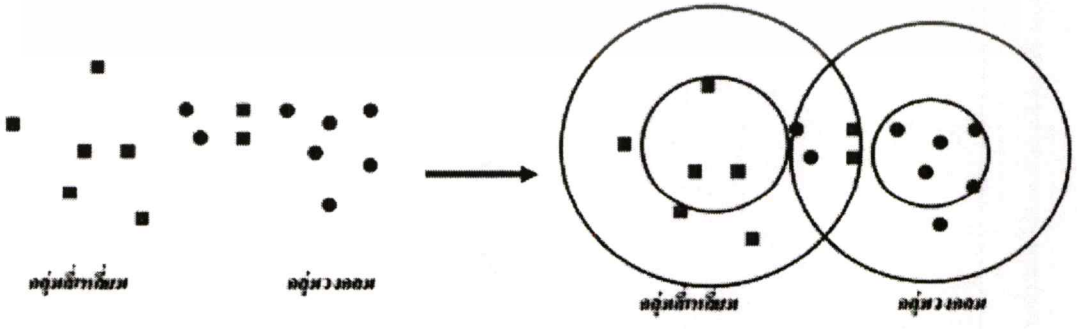
จากรูปที่ 1.3 ด้านขวาสามารถแบ่งกลุ่มของข้อมูลออกเป็น 2 ส่วน คือส่วน lower และส่วน upper ใช้อธิบายการกำหนดพื้นที่ต่างๆ โดยส่วน lower คือพื้นที่ A1 หมายความว่าข้อมูลที่อยู่ในพื้นที่ A1 เป็นข้อมูลที่มีความแน่นอนว่าเป็นข้อมูลของกลุ่ม 1 ในส่วน upper คือพื้นที่ B1 หมายความว่า ข้อมูลที่อยู่ในส่วนของ B1 เป็นส่วนที่มีโอกาสเป็นข้อมูลของกลุ่ม 1 และกลุ่มอื่นๆ

1.5.2 การนำกลุ่มข้อมูลต้นแบบมาจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน

เมื่อพิจารณาต้นแบบใหม่ที่ได้นำเสนอในรูปที่ 1.3 ด้านขวามาอธิบายข้อมูลที่มีการซ้อนทับกันจะแสดงในรูปที่ 1.6 โดยพื้นที่ A1 เป็นข้อมูลของกลุ่มที่ 1 แน่นอน และ พื้นที่ A2 เป็นข้อมูลของกลุ่มที่ 2 แน่นอนเราเรียกว่าส่วน lower ส่วนพื้นที่ B1 เป็นส่วนที่มีโอกาสเป็นข้อมูลของกลุ่มอื่นๆ หรือ กลุ่มที่ 1 ทำนองเดียวกันพื้นที่ B2 เป็นส่วนที่มีโอกาสเป็นข้อมูลของกลุ่มอื่นๆ หรือกลุ่มที่ 2 เราเรียกว่าส่วน upper และพื้นที่ C เป็นส่วนที่มีโอกาสเป็นข้อมูลของกลุ่มที่ 1 หรือ กลุ่มที่ 2 ก็ได้ เราเรียกว่าส่วนการซ้อนทับกัน จากการกำหนดต้นแบบนี้สามารถนำมาแบ่งข้อมูลที่มีการซ้อนทับกันได้ผลลัพธ์แสดงในรูปที่ 1.7



รูปที่ 1.6 แสดงการกำหนดพื้นที่ของ lower และ upper สำหรับข้อมูลซ้อนทับกัน



รูปที่ 1.7 แสดงการแบ่งข้อมูลที่ซ้อนทับกันเป็นกลุ่มด้วยต้นแบบงานวิจัยของงานวิจัยนี้

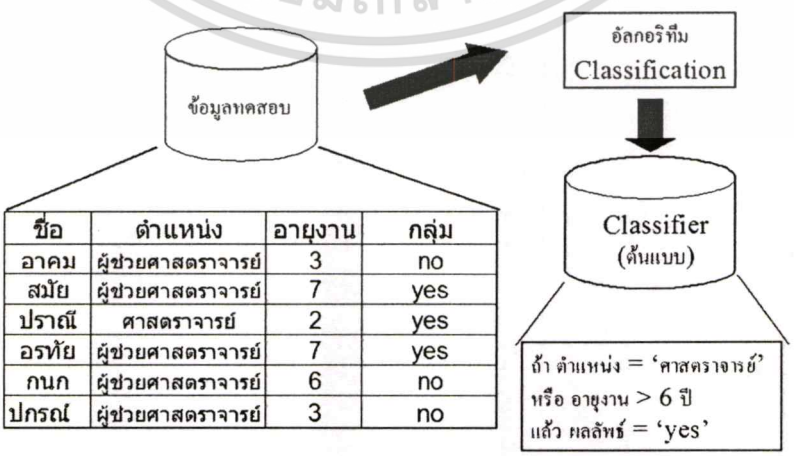
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5.3 การนำหลักการจัดกลุ่มด้วยวิธี Neural Network Classifier (NNC) มาจัดกลุ่มข้อมูลต้นแบบ

ที่กล่าวในส่วนต้นเป็นเรื่องของต้นแบบการแบ่งกลุ่ม ส่วนถัดมาเป็นเรื่องเกี่ยวกับหลักการที่ใช้ในการแบ่งกลุ่มซึ่งใช้หลักการเดียวกับ M.A. ABOU-NASR ได้เสนอ Neural Network Classifier (NNC) เป็นการแบ่งกลุ่มด้วยวิธีของ Neural เริ่มต้นด้วยการนำข้อมูลตัวแรกเข้ามากำหนดขอบเขตของกลุ่ม(จุดศูนย์กลางหรือน้ำหนักของกลุ่มและกำหนดรัศมีหรือค่า threshold ของกลุ่มสองค่า) จากนั้นนำข้อมูลที่เหลือทีละค่า มาทดสอบกับกลุ่มที่มีอยู่ โดยสามารถแบ่งการพิจารณาออกเป็น 2 กรณีด้วยกัน คือ การเปรียบเทียบข้อมูลกับกลุ่มข้อมูลกลุ่มเดียวกัน และการเปรียบเทียบข้อมูลกับกลุ่มข้อมูลต่างกัน และจะมีการพิจารณาปรับค่ารัศมีทั้งสองและค่าน้ำหนักของแต่ละกลุ่ม เพื่อจัดให้ข้อมูลที่นำมาทดสอบอยู่ในกลุ่มต่างๆ โดยจะอธิบายทฤษฎีของนิวรอนในบทที่ 3 และอัลกอริทึมที่ใช้ปรับค่ารัศมีทั้งสองในบทที่ 4 ต่อไป

1.5.4 การนำงานวิจัยในวิทยานิพนธ์ไปใช้งาน

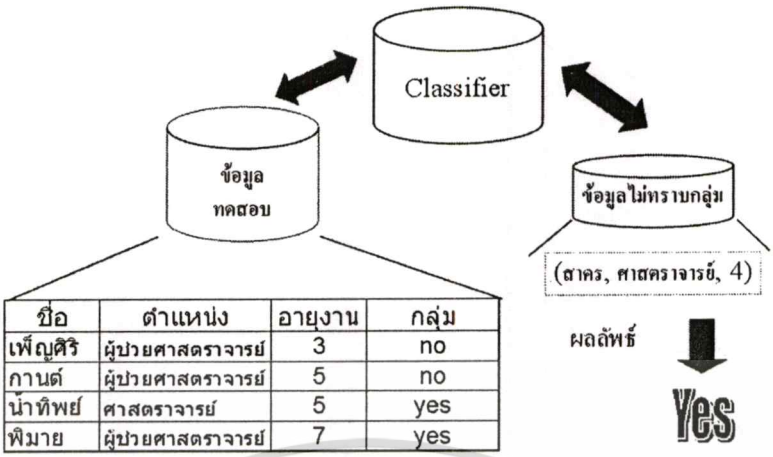
สำหรับขั้นตอนการจัดกลุ่มข้อมูลแบ่งเป็นสองขั้นตอน โดยขั้นตอนแรกเป็นการสร้างต้นแบบของกลุ่มเพื่อกำหนดลักษณะและขอบเขตของข้อมูลตามข้อมูลที่ใช้ทดสอบเรียกว่าขั้นตอนการเรียนรู้ (training) ตามรูปที่ 1.8 ส่วนขั้นตอนที่สองเป็นการนำผลลัพธ์ที่ได้จากขั้นตอนแรกมาใช้งานนั่นคือการหากลุ่มให้กับข้อมูลที่ไม่ทราบกลุ่มแต่ทราบลักษณะของข้อมูลที่เรียกว่าขั้นตอนการทำนาย (prediction) ตามรูปที่ 1.9 สำหรับการทำงานในวิทยานิพนธ์นี้อยู่ในส่วนการเรียนรู้ของ Classification คือส่วนที่ใช้หลักการในการแบ่งกลุ่ม และรูปที่ 1.9. จะพบว่าเราสามารถนำประโยชน์จากงานวิจัยไปใช้ในการอธิบายขอบเขตของกลุ่ม เพื่อใช้ในการทำนายหรือคาดเดากลุ่มของข้อมูลใหม่ได้ จะมีการกล่าวถึงรายละเอียดของขั้นตอนการจัดกลุ่มเพิ่มเติมอีกครั้งหนึ่งในบทที่ 2



ชื่อ	ตำแหน่ง	อายุงาน	กลุ่ม
อาคม	ผู้ช่วยศาสตราจารย์	3	no
สมัย	ผู้ช่วยศาสตราจารย์	7	yes
ปราณี	ศาสตราจารย์	2	yes
อรทัย	ผู้ช่วยศาสตราจารย์	7	yes
กนก	ผู้ช่วยศาสตราจารย์	6	no
ปกรณ์	ผู้ช่วยศาสตราจารย์	3	no

รูปที่ 1.8 แสดงการนำข้อมูลตัวอย่างไปใช้สร้าง Classifier ต้นแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ชื่อ	ตำแหน่ง	อายุงาน	กลุ่ม
เพ็ญศิริ	ผู้ช่วยศาสตราจารย์	3	no
กานต์	ผู้ช่วยศาสตราจารย์	5	no
น้ำทิพย์	ศาสตราจารย์	5	yes
พิมาย	ผู้ช่วยศาสตราจารย์	7	yes

รูปที่ 1.9 แสดงการนำ Classifier ต้นแบบไปใช้งาน

1.6 ขอบเขตของการวิจัย

ศึกษาการทำงานของนิวรอนเพื่อใช้ในการแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกัน นำเสนอต้นแบบของกลุ่มข้อมูลพร้อมทั้งกำหนดพื้นที่ต่างๆ ของต้นแบบกลุ่มที่นำเสนอ จากนั้นทดสอบแบ่งกลุ่มข้อมูลพยัญชนะภาษาไทยที่มีข้อมูลซ้อนทับกัน และนำผลที่ได้จากการทดลองเปรียบเทียบกับงานวิจัยเดิมที่นำเทคนิค NNC [1]

1.7 ขั้นตอนของการศึกษา

นำข้อมูลพยัญชนะภาษาไทยที่เป็นลายมือเขียนที่มีข้อมูลการซ้อนทับกันมาทำการทดสอบแบ่งกลุ่มตามต้นแบบและใช้หลักการของ NNC ที่ได้นำเสนอ เปรียบเทียบผลการแบ่งกลุ่มกับวิธีการแบ่งกลุ่มแบบ NNC [1] สรุปผลการทดลองพร้อมจัดทำบทความตีพิมพ์และวิทยานิพนธ์

1.8 รายละเอียดในแต่ละบท

ในวิทยานิพนธ์นี้แบ่งเนื้อหาการนำเสนอออกเป็น 6 บทดังนี้

บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหา แนวความคิดที่นำเสนอเพื่อแก้ไขปัญหา วัตถุประสงค์ ขอบเขตของงานวิจัย ขั้นตอนการศึกษา และรายละเอียดของเนื้อหาต่างๆ ในวิทยานิพนธ์

บทที่ 2 กล่าวถึงความหมายของการจัดกลุ่ม ประเภทของการจัดกลุ่มแต่ล่ะวิธีว่ามีหลักการและแนวคิดอย่างไรที่ได้มีผู้วิจัยแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3 กล่าวถึงทฤษฎีหลักการแบ่งกลุ่มโดยใช้ Neural network Classifier (NNC) ซึ่งเป็นหลักการที่ใช้ในงานวิจัยของ M.A. ABOU-NASR และมีการนำมาใช้ในวิทยานิพนธ์นี้ด้วย

บทที่ 4 กล่าวถึงการนำเสนอต้นแบบใหม่ที่ใช้ในการแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกัน โดยนำเอาทฤษฎีของ NNC จากบทที่3 มาควบคุมการทำงานของต้นแบบในวิทยานิพนธ์โดยเรียกวิธีนี้ว่า Modified Neural Network Classifier (MNNC) ภายในบทจะอธิบายถึงการกำหนดพื้นที่ในกลุ่มที่นำเสนอ และอัลกอริทึมของ Modified Neural network Classifier (MNNC) และ การปรับค่า threshold (รัศมี) ทั้งสองค่าของกลุ่มเพื่อจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน

บทที่ 5 กล่าวถึงผลการทดลอง

บทที่ 6 กล่าวถึงการสรุป วิเคราะห์ผลการทดลอง รวมทั้งข้อเสนอแนะและแนวทางการทำวิจัยต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
การจัดกลุ่มข้อมูล

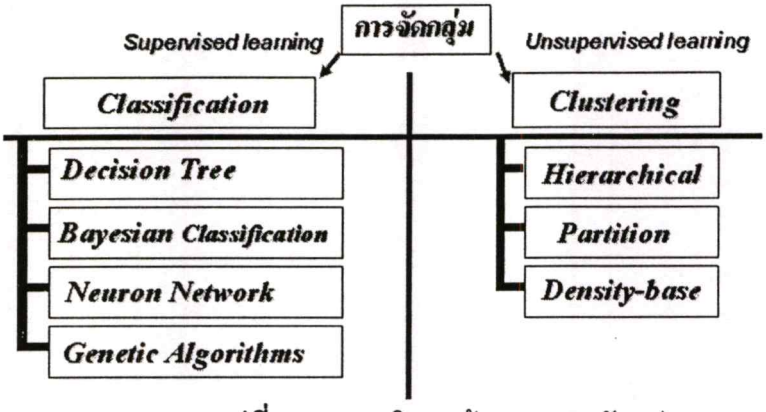
ในบทนี้จะกล่าวถึงภาพรวมของการจัดกลุ่มข้อมูล เริ่มต้นจากความหมายของการจัดกลุ่มข้อมูล ขั้นตอนการจัดกลุ่ม ประโยชน์ที่ได้รับจากการการจัดกลุ่ม การแบ่งประเภทของการจัดกลุ่มข้อมูลโดยแบ่งเป็น การจัดหมวดหมู่ (Classification) และ การรวมตัว (Clustering) อธิบายทฤษฎีต่างๆ ที่นิยมนำมาใช้ในการจัดหมวดหมู่ เช่น Decision Tree, Bayesian, Neural Network, Genetic, และทฤษฎีที่นิยมใช้ในการรวมตัว (Clustering) เช่น Partition, Hierarchical, Densitiy-base สำหรับงานวิจัยนี้ใช้ทฤษฎี Neural Network จะมีการกล่าวถึงรายละเอียดเพิ่มเติมอีกครั้งหนึ่งในบทที่ 3

2.1 บทนำ

การจัดกลุ่มข้อมูลคือ การแบ่งข้อมูลออกเป็นกลุ่มๆ โดยข้อมูลที่อยู่กลุ่มเดียวกันจะมีความคล้ายคลึงกัน ส่วนข้อมูลที่อยู่ต่างกลุ่มกันจะมีความแตกต่างกัน จากรูปที่ 2.1 สามารถแยกประเภทของการจัดกลุ่มได้เป็น การจัดหมวดหมู่ (Classification) และ การรวมตัว (Clustering) โดยการจัดหมวดหมู่เป็นการจำแนกประเภทของการจัดกลุ่มแบบมีผู้สอน (Supervised Classification) ซึ่งกลุ่มของข้อมูลจะถูกกำหนดไว้ล่วงหน้าแล้ว จากนั้นข้อมูลจะถูกจัดอยู่ในกลุ่มที่กำหนดไว้ ส่วนการรวมตัว เป็นการจำแนกประเภทของการจัดกลุ่มแบบไม่มีผู้สอน (Unsupervised Classification) คือไม่มีข้อมูลใดๆ ของกลุ่มที่ถูกจัดไว้ก่อน สำหรับทฤษฎีต่าง ๆ ในรูปที่ 2.1 อธิบายเพิ่มเติมในหัวข้อ 2.3

การจัดกลุ่มข้อมูลถูกนำมาประยุกต์ใช้ในงานสนับสนุนการตัดสินใจของงานวิจัยในหน่วยงานหรือองค์กรต่างๆ เช่นงานวิจัยการรู้จำตัวอักษร การรู้จำใบหน้าคน การสืบค้นสารสนเทศ ในโลกของธุรกิจการตลาดมีการนำการจัดกลุ่มมาใช้สร้างความสัมพันธ์กับลูกค้า โดยสังเกตความต้องการและความสนใจของลูกค้า และอาจมีการเรียนรู้ได้จากผลสะท้อนในอดีตว่าทำอย่างไรให้การบริการลูกค้ามีประสิทธิภาพดีขึ้นในอนาคต เช่นบริษัทผู้ออกบัตรเครดิตและธนาคารต่างๆ จะใช้ประโยชน์ของการจัดกลุ่มในการตัดสินว่าลูกค้ากลุ่มใดเป็นกลุ่มที่ดี อย่างไรก็ตามการจัดกลุ่มต้องอาศัยการเรียนรู้และเก็บสะสมข้อมูลอย่างตรงไปตรงมาจึงจะทำให้การจัดกลุ่มข้อมูลเป็นไปอย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



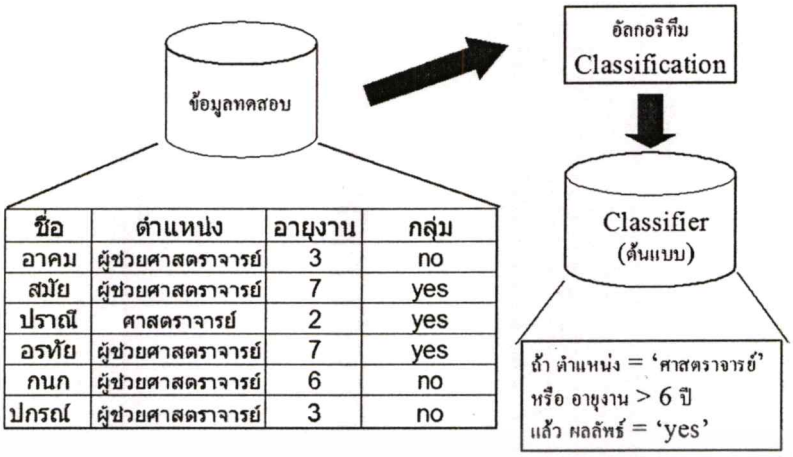
รูปที่ 2.1 แสดงโครงสร้างของการจัดกลุ่ม

2.2 ขั้นตอนการจัดกลุ่ม

สำหรับขั้นตอนการจัดกลุ่ม Jiawei Han [2] ได้เสนอการจัดกลุ่มข้อมูลแบ่งเป็นสองขั้นตอน โดยขั้นตอนแรกเป็นการสร้างต้นแบบของกลุ่มเพื่อกำหนดลักษณะและขอบเขตของข้อมูลตามข้อมูลที่ใช้ทดสอบเรียกว่าขั้นตอนการเรียนรู้ (training) ตามรูปที่ 2.2 ส่วนขั้นตอนที่สองเป็นการนำผลลัพธ์ที่ได้จากขั้นตอนแรกมาใช้งานนั่นคือการหากลุ่มให้กับข้อมูลที่ไม่ทราบกลุ่มแต่ทราบลักษณะของข้อมูลที่เรียกว่าขั้นตอนการทำนาย (prediction) ตามรูปที่ 2.3 สำหรับการทำงานของวิทยานิพนธ์นี้อยู่ในส่วนการเรียนรู้ของ Classification คือส่วนที่ใช้หลักการในการแบ่งกลุ่ม และรูปที่ 2.3. จะพบว่าเราสามารถนำประโยชน์จากวิทยานิพนธ์นี้ไปใช้ในการอธิบายขอบเขตของกลุ่มเพื่อใช้ในการวิเคราะห์การจัดกลุ่มของข้อมูลใหม่ได้

ในรูปที่ 2.2 เป็นการนำข้อมูลตัวอย่างส่วนหนึ่งที่มีการระบุกลุ่มและคุณลักษณะของกลุ่ม (feature) ไว้ก่อนแล้ว ในตารางด้านซ้ายประกอบด้วยชื่อ ตำแหน่ง อายุงาน(ปี) คือค่าคุณลักษณะของข้อมูล และผลลัพธ์เป็นการเลื่อนตำแหน่งคือกลุ่มที่มีการระบุแน่นอนในรูปมี 2 กลุ่มคือ yes/no ค่า yes คือการได้เลื่อนตำแหน่ง ค่า no คือการไม่ได้เลื่อนตำแหน่ง จากนั้นนำข้อมูลนี้มาทดสอบกับอัลกอริทึมที่ได้นำเสนอในวิทยานิพนธ์นี้ จะได้ผลลัพธ์เป็นขอบเขตของข้อมูลกลุ่มต่างๆ เป็นต้นแบบ ผลลัพธ์ในการจัดกลุ่มข้อมูลที่ได้จากการเรียนรู้คือถ้าบุคคลนั้นๆ มีตำแหน่งเป็นศาสตราจารย์ หรือ บุคคลนั้นมีอายุงานมากกว่า 6 ปี บุคคลนั้นมีสิทธิได้รับพิจารณาขึ้นเงินเดือน

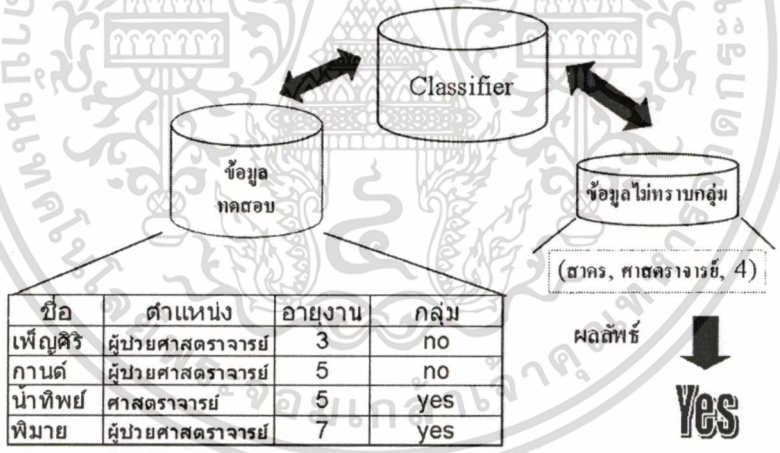
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ชื่อ	ตำแหน่ง	อายุงาน	กลุ่ม
อาคม	ผู้ช่วยศาสตราจารย์	3	no
สมัย	ผู้ช่วยศาสตราจารย์	7	yes
ปราณี	ศาสตราจารย์	2	yes
อรทัย	ผู้ช่วยศาสตราจารย์	7	yes
กนก	ผู้ช่วยศาสตราจารย์	6	no
ปกรณ์	ผู้ช่วยศาสตราจารย์	3	no

รูปที่ 2.2 แสดงการนำข้อมูลตัวอย่างไปใช้สร้าง Classifier ต้นแบบ

จากนั้นจะเข้าสู่ขั้นตอนที่ 2 คือ การนำ Classifier ต้นแบบในรูป 2.2 ไปใช้งาน ในรูปที่ 2.4 เราสามารถทดสอบ Classifier ต้นแบบก่อน และในการใช้งานจริงเราสามารถทำนายผลลัพธ์ว่าบุคคลที่เราไม่ทราบว่าจัดอยู่ในกลุ่มใด แต่ทราบค่าคุณลักษณะของบุคคลนั้น เช่นในภาพ 2.3 นายสาครมีตำแหน่งเป็นศาสตราจารย์ และมีอายุงาน 4 ปี เมื่อนำไปวิเคราะห์กับ Classifier พบว่านายสาครจัดอยู่ในกลุ่มการได้เลื่อนตำแหน่ง(yes)



ชื่อ	ตำแหน่ง	อายุงาน	กลุ่ม
เพ็ญศิริ	ผู้ช่วยศาสตราจารย์	3	no
กานต์	ผู้ช่วยศาสตราจารย์	5	no
นาทิพย์	ศาสตราจารย์	5	yes
พิมาย	ผู้ช่วยศาสตราจารย์	7	yes

รูปที่ 2.3 แสดงการนำ Classifier ต้นแบบไปใช้งาน

2.3 ทฤษฎีที่ใช้ในการจัดกลุ่ม

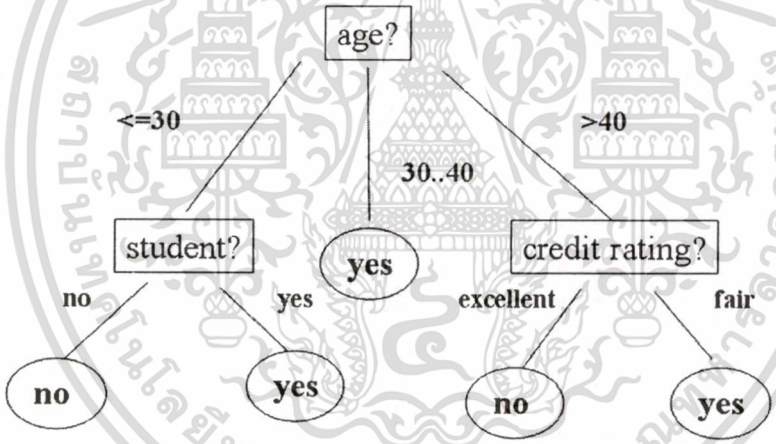
จากรูปที่ 2.1 สามารถแยกประเภทของการจัดกลุ่มได้เป็น การจัดหมวดหมู่ (Classification) และ การรวมตัว (Clustering) ในแต่ละประเภทของการจัดกลุ่มมีทฤษฎีที่อธิบายการแบ่งกลุ่มหลายทฤษฎีมีกล่าวถึงในงานวิจัยของ Anil K. Jain [4],[5] สามารถอธิบายได้ในหัวข้อ 2.3.1 และ 2.3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1 ทฤษฎีต่าง ๆ ของการจัดหมวดหมู่ (Classification)

2.3.1.1 ดีซีชันทรี (Decision Tree)

ดีซีชันทรี เป็นการจัดกลุ่มโดยอาศัยแบบจำลองที่มีลักษณะคล้ายกับต้นไม้ จะมีการสร้างกฎต่างๆ เพื่อใช้ในการตัดสินใจ โดยภายในแบบจะลองจะประกอบด้วยโหนดภายใน (Internal node) แสดงคุณลักษณะที่ใช้จำแนกกลุ่ม แต่ละกิ่ง(Branch) แสดงผลลัพธ์ของการทดสอบ และลีฟโหนด (Leaf node) คือโหนดที่ไม่มีโหนดย่อยแสดงการแบ่งกลุ่ม สามารถแสดงกลุ่มข้อมูลตามลักษณะแผนภาพที่กล่าวมานี้สามารถแสดงได้ดังรูปที่ 2.4 โดยในรูปนำข้อมูลกลุ่มลูกค้าที่ซื้อเครื่องใช้ไฟฟ้ามาวิเคราะห์และศึกษาความเป็นไปได้ที่ลูกค้ากลุ่มนี้จะซื้อเครื่องคอมพิวเตอร์โดย ในรูปที่ 2.4 สัญลักษณ์สี่เหลี่ยมผืนผ้าแสดงคุณลักษณะที่ใช้จำแนกกลุ่มคือโหนดภายใน เช่นคุณลักษณะของอายุ (age), สถานะภาพนักศึกษา(student), ความน่าเชื่อถือเพื่อขอสินเชื่อ (credit rating) และสัญลักษณ์วงรีแสดงข้อมูลกลุ่มคือลีฟโหนด ในรูปที่ 2.4 แสดง 2 กลุ่ม กลุ่มลูกค้าที่ซื้อคอมพิวเตอร์ (yes) และ กลุ่มลูกค้าที่ไม่ซื้อคอมพิวเตอร์ (no)



รูปที่ 2.4 แสดงดิซีชั่นทรีสำหรับการตัดสินใจซื้อเครื่องคอมพิวเตอร์ของลูกค้ากลุ่มหนึ่ง

2.3.1.2 Bayesian Classification

Bayesian เป็นวิธีการจัดหมวดหมู่อีกแบบหนึ่งที่อาศัยทฤษฎีพื้นฐานของ Bayes และนำอัลกอริทึมชื่อ ทฤษฎี Bayesian พื้นฐานสำหรับการจัดหมวดหมู่ (Native Bayesian Classifier) โดยมีพื้นฐานการคำนวณทางสถิติคือความน่าจะเป็น (Probabilities) มาศึกษาข้อมูลจากการเรียนรู้ (training) สำหรับใช้ในการจัดหมวดหมู่ของและวิเคราะห์ข้อมูลได้

ทฤษฎี Bayesian พื้นฐานสำหรับการจัดหมวดหมู่ กำหนดให้ X เป็นข้อมูลตัวอย่างซึ่งไม่ทราบกลุ่ม $P(C_i)$ คือความน่าจะเป็นของแต่ละกลุ่ม $P(X|C_i)P(C_i)$ คือค่าความน่าจะเป็นของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X ในกลุ่ม C_i ข้อมูล X จะจัดอยู่ในกลุ่ม C_i ที่มีค่าความน่าจะเป็นสูงที่สุดในเงื่อนไขคุณลักษณะของ X เมื่อ m คือจำนวนกลุ่มเช่น $C_1, C_2, C_3,\ldots,C_m$

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \quad \text{สำหรับ } 1 \leq j \leq m, \quad j \neq i. \quad (2.1)$$

ตัวอย่างการวิเคราะห์ข้อมูลโดยการใช้การจัดหมวดหมู่ด้วยวิธี Bayesian เพื่อวิเคราะห์ข้อมูล X ที่มีคุณลักษณะดังนี้ อายุ = '<=30', รายได้ = 'medium', สถานะภาพนักศึกษา = 'yes', ข้อมูลเครดิต = 'fair' โดยมีข้อมูลการเรียนรู้ตามตารางที่ 2.1 คือมีข้อมูล 14 ข้อมูล มีคุณลักษณะ 4 แบบ คือ อายุ รายได้ สถานะภาพการเป็นนักศึกษา และเครดิต มีกลุ่มที่สามารถตัดสินใจได้ 2 กลุ่มคือกลุ่มตัดสินใจซื้อเครื่องคอมพิวเตอร์ (yes) และกลุ่มที่ตัดสินใจไม่ซื้อเครื่องคอมพิวเตอร์ (no)

ตารางที่ 2.1 แสดงข้อมูลการเรียนรู้ (Training) กลุ่มลูกค้าที่ตัดสินใจซื้อคอมพิวเตอร์

ข้อมูลที่	อายุ	รายได้	นักศึกษา	เครดิต	กลุ่ม
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31...40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31...40	medium	no	excellent	yes
13	31…40	high	yes	fair	yes
14	>40	medium	no	excellent	no

จากข้อมูลการเรียนรู้ตามตารางที่ 2.1 ต้องการคำนวณค่า $P(X|C_i)P(C_i)$ สำหรับ i= 1, 2. คือค่าความน่าจะเป็นของข้อมูล X ในกลุ่ม C_i และ $P(C_i)$ คือความน่าจะเป็นของแต่ละกลุ่ม สามารถคำนวณได้ดังนี้ คำนวณหาค่า $P(C_i)$ ก่อน

P(กลุ่มที่ซื้อคอมพิวเตอร์ = "yes") = 9/14 =0.643

P(กลุ่มที่ซื้อคอมพิวเตอร์ = "no") = 5/14 = 0.357

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับต่อไปคำนวณหาค่า $P(X|C_i)$, สำหรับ i = 1, 2, และสามารถคำนวณความน่าจะเป็นของแต่ละคุณลักษณะได้ดังนี้

P(อายุ= '<30' | กลุ่มที่ซื้อคอมพิวเตอร์ = "yes") = 2/9 = 0.222

P(อายุ= '<30' | กลุ่มที่ซื้อคอมพิวเตอร์ = "no") = 3/5 = 0.600

P(รายได้= 'medium' | กลุ่มที่ซื้อคอมพิวเตอร์ = "yes") = 4/9 = 0.444

P(รายได้= 'medium' | กลุ่มที่ซื้อคอมพิวเตอร์ = "no") = 2/5 = 0.400

P(สถานะภาพนักศึกษา= 'yes' | กลุ่มที่ซื้อคอมพิวเตอร์ = "yes") = 6/9 = 0.667

P(สถานะภาพนักศึกษา= 'yes' | กลุ่มที่ซื้อคอมพิวเตอร์ = "no") = 1/5 = 0.200

P(ข้อมูลเครดิต= 'fair' | กลุ่มที่ซื้อคอมพิวเตอร์ = "yes") = 6/9 = 0.667

P(ข้อมูลเครดิต= 'fair' | กลุ่มที่ซื้อคอมพิวเตอร์ = "no") = 2/5 = 0.400

จากข้อมูลความน่าจะเป็นด้านบนสามารถพิจารณาต่อได้

P(X| กลุ่มที่ซื้อคอมพิวเตอร์ = "yes") = 0.222 x 0.444 x 0.667 x 0.667 = 0.044

P(X| กลุ่มที่ซื้อคอมพิวเตอร์ = "no") = 0.600 x 0.400 x 0.200 x 0.400 = 0.019

P(X| กลุ่มที่ซื้อคอมพิวเตอร์ = "yes")P(กลุ่มที่ซื้อคอมพิวเตอร์ = "yes")=0.044x0.643= 0.028

P(X| กลุ่มที่ซื้อคอมพิวเตอร์ = "no")P(กลุ่มที่ซื้อคอมพิวเตอร์ = "no") =0.019x0.357= 0.007

ผลลัพธ์การจัดหมวดหมู่จากการวิเคราะห์ด้วยวิธี Bayesian สำหรับข้อมูล X ที่มีคุณลักษณะ age = '<=30', income = 'medium', student = 'yes', credit_rating = 'fair' พิจารณาตามสมการที่ 2.1 ได้ว่าสามารถจัดข้อมูล X อยู่ในกลุ่มที่ซื้อคอมพิวเตอร์

2.3.1.3 นิวรอนเน็ตเวิร์ก (Neuron Network: NN)

นิวรอนเน็ตเวิร์ก ถูกพัฒนาขึ้นโดยโมเดลทางคณิตศาสตร์ที่อาศัยกระบวนการเรียนรู้ของมนุษย์ (เลียนแบบการทำงานของสมอง) และจะเรียนรู้จากชุดข้อมูลการเรียนรู้ (Training) นิวรอนเน็ตเวิร์ก ประกอบด้วยหน่วยความจำจำนวนมาก เรียกว่า นิวรอน (Neurons) เซล (Cells) หรือโหนด (Nodes) แต่ละนิวรอนต่อกันโดยคอนเน็กชันลิงค์ (Connection Link) ที่มีค่าน้ำหนักของมันอยู่ ในแต่ละการเชื่อมต่อ โดยค่าน้ำหนักจะแสดงรายละเอียดที่เน็ตเวิร์กใช้ในการแก้ปัญหา นิวรอนเน็ตเวิร์กถูกใช้ในการแก้ปัญหาอย่างกว่างขวาง เช่น การเก็บและการเรียกข้อมูล การแยกประเภทของข้อมูล การจัดหมวดหมู่ ความสามารถในการตรวจสอบรูปแบบของข้อมูลที่คล้ายคลึงกับความคิดของมนุษย์ เป็นต้น สำหรับทฤษฎี Neural Network จะมีการกล่าวถึงรายละเอียดเพิ่มเติมอีกครั้งหนึ่งในบทที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1.4 **จีเนติก อัลกอริทึม** (Genetic Algorithm:GA)

จีเนติก อัลกอริทึม เป็นทฤษฎีการจำลองกระบวนการทางธรรมชาติ โดยอาศัยพื้นฐานความคิดทางพันธุศาสตร์หรือพันธุวิศวกรรมในการถ่ายทอดลักษณะต่าง ๆ ไปยังรุ่นถัดไป จึงสามารถนำมาพัฒนาใช้ในการหาคำตอบที่เหมาะสมที่สุดของแต่ละปัญหา จีเนติก อัลกอริทึม เริ่มต้นโดยการสร้างกฎเพื่อใช้แทนข้อมูลที่เรียกว่าประชากร (Population) ด้วยรูปแบบของเลขฐานสองคือข้อมูล 0 และ 1 ตัวอย่างเช่นในขั้นตอนการเรียนรู้ สามารถใช้รูปแบบของเลขฐานสองหรือเปรียบเสมือนโครงสร้างโครโมโซม (Cromosomes) อธิบายฐานข้อมูลที่มีคุณลักษณะ A1, A2 และมีกลุ่มข้อมูลสองกลุ่มคือ C1 และ C2 สามารถสร้างกฎเพื่ออธิบายได้เป็น "ถ้าข้อมูลมีคุณลักษณะ A1 และไม่มีคุณลักษณะ A2 แล้วจัดว่าข้อมูลอยู่ในกลุ่ม C2" สามารถแทนได้ด้วยรูปแบบของเลขฐานสองคือ 100 โดยที่ข้อมูลเลขฐานสอง 2 บิตจากด้านซ้ายแทนข้อมูลคุณลักษณะ A1 และ A2 ตามลำดับ และข้อมูลทางด้านขวาสุดแทนกลุ่ม (กลุ่ม C1=1, C2=0) ในทำนองเดียวกันสร้างกฎว่า "ถ้าข้อมูลไม่มีคุณลักษณะ A1 และ A2 แล้วจัดว่าข้อมูลอยู่ในกลุ่ม C1" สามารถแทนได้ด้วยรูปแบบของเลขฐานสองคือ 001 จากนั้นทำการสร้างประชากรใหม่ (Offsprings) โดยการผ่านกระบวนการสับเปลี่ยนค่าโครโมโซม (Crossover) และ การกลายพันธุ์ (Mutation) ต่อไปทำการทดสอบประชากรใหม่ด้วยสมการทางคณิตศาสตร์เพื่อใช้วัดประสิทธิภาพของขั้นตอนการเรียนรู้สำหรับการจัดกลุ่มที่เรียกว่ากระบวนการ fitness กระบวนการจีเนติกอัลกอริทึม จะทำการสร้างประชากรใหม่และกระบวนการ fitness ซ้ำจนกระทั่งได้เงื่อนไขของสมการทางคณิตศาสตร์ในกระบวนการ fitness ที่เหมาะสมที่สุด

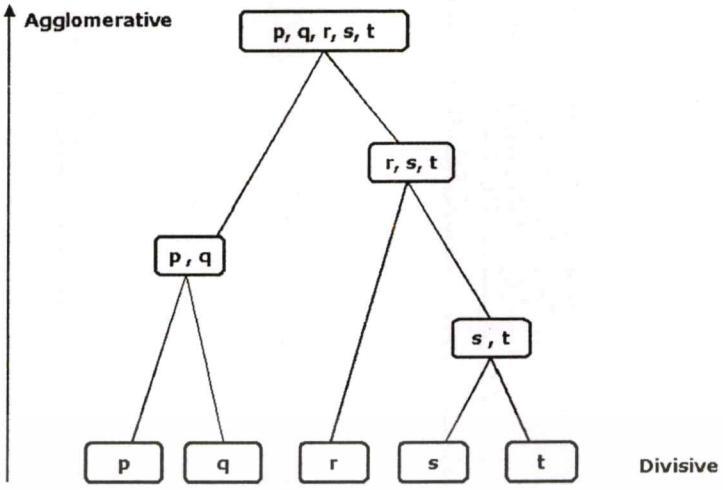
2.3.2 **ทฤษฎีต่าง ๆ ของการรวมตัว** (Clustering)

สำหรับทฤษฎีการรวมตัวแบบ Hierarchical และ Partition ในวิทยานิพนธ์ของพรเทพ โรจนวสุ [6] ได้อธิบายไว้ดังนี้

2.3.2.1 Hierarchical

เป็นการจัดกลุ่มแบบ hierarchical ได้อธิบายว่า เป็นการจัดกลุ่มแบบโครงสร้างแบบลำดับชั้นโดยชั้นบนสุดจะมองเห็นเป็นกลุ่มเดียวกันหมด และชั้นล่างสุดข้อมูลแต่ละอันจะถือว่าเป็นหนึ่งกลุ่ม ส่วนชั้นตรงกลางเกิดจากการรวมกันของกลุ่มในชั้นที่ต่ำกว่าหรือจะมองว่าเกิดจากการแตกออกเป็นสองกลุ่มของชั้นที่สูงกว่าก็ได้ แผนภาพที่นิยมใช้แสดงการจัดกลุ่มแบบ hierarchical เรียกว่าแผนภาพเดนโดแกรม (dendogram) ซึ่งมีลักษณะคล้ายโครงสร้างแบบต้นไม้ดังรูปที่ 2.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 แสดงแผนภาพเดนโดแกรมของการจัดกลุ่มแบบ hierarchical

จากรูปที่ 2.5 เทคนิค agglomerative จะได้รับความนิยมมากกว่าเทคนิค division เนื่องจากสามารถทำได้สะดวกกว่า เริ่มต้นจากการกำหนดฟังก์ชันที่ใช้วัดระยะห่างของข้อมูลซึ่งนิยมใช้ฟังก์ชันการวัดระยะห่างแบบยูคลิด(Euclidean distance) สามารถแสดงได้ดังสมการที่ 2.2

$$d(i,j)=\sqrt{\left|x_{i1}-x_{j1}\right|^2+\left|x_{i2}-x_{j2}\right|^2+\ldots+\left|x_{ip}-x_{jp}\right|^2}, \qquad (2.2)$$

โดยที่ $i=(x_{i1},x_{i2},...,x_{ip})$ และ $j=(x_{j1},x_{j2},...,x_{jp})$ และค่า p คือจำนวนคุณลักษณะของข้อมูล

อัลกอลริทึมของ agglomerative hierarchical สามารถแสดงได้ดังนี้

สร้างเมตริกซ์ความต่าง (dissimilarity matrix) NxN ของข้อมูลทั้งหมด

รวมสองกลุ่มที่เหมือนกันที่สุด (ในกรณีใช้ฟังก์ชันยูคลิดจะพิจารณาสองกลุ่มที่มีค่าระยะห่างน้อยที่สุด)

ทำการปรับปรุงเมตริกซ์ความไม่เหมือน โดยคิดรวมสองกลุ่มก่อนหน้าที่เหมือนกันมากที่สุดเป็นกลุ่มเดียวกัน

ทำซ้ำกระบวนการ 2 และ 3 จนกระทั่งเหลือกลุ่มเดียว

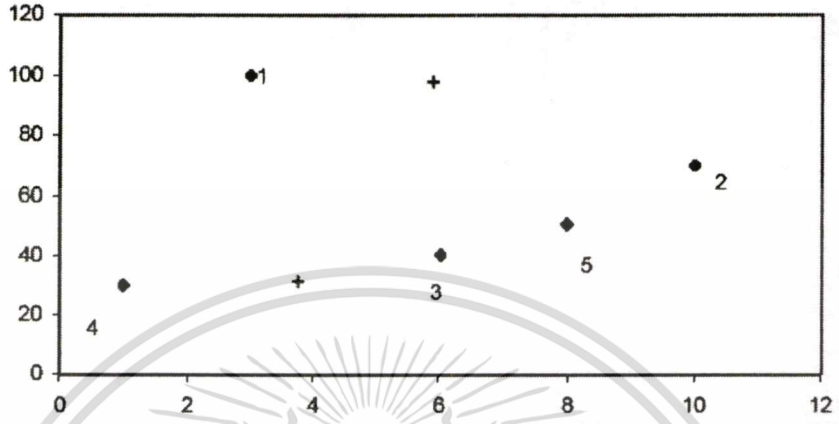
2.3.2.2 Partition

การจัดกลุ่มแบบ partition [6] จะไม่มีการสร้างเป็นโครงสร้างต้นไม้โดยแต่ละกลุ่มจะแยกจากกันและมีแค่ระดับเดียว งานวิจัยที่นิยมใช้และจัดอยู่ในทฤษฎีนี้คือ K-means จะเริ่มต้นด้วยการกำหนดจำนวนกลุ่มที่ต้องการและค่าเริ่มต้นคุณลักษณะของกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

จากนั้นพิจารณาข้อมูลแต่ละตัวลงในกลุ่ม โดยทำการวัดค่าว่าข้อมูลนั้นใกล้กับกลุ่มใดมากที่สุดฟังก์ชันที่นิยมใช้คือฟังก์ชันการวัดระยะห่างแบบยูคลิดตามสมการที่ 2.2 สามารถแสดงตัวอย่างการจัดกลุ่มแบบ K-means ได้ดังรูปที่ 2.6



รูปที่ 2.6 แสดงแผนภาพกราฟ xy ของข้อมูล 1-5 และค่าเริ่มต้นของกลุ่มเป็นรูปกากบาท

จากรูปที่ 2.6 สมมุติมีข้อมูล 5 ค่าดังต่อไปนี้ (3,100), (10,70), (6,40), (1,30), (8,50) และกำหนดจำนวนกลุ่ม K=2 มีคุณลักษณะเริ่มต้นที่ (4,30) และ (6,100) อัลกอลิทึมเริ่มทำงานโดยการคำนวณระยะห่างระหว่างข้อมูลกับค่าเริ่มต้นคุณลักษณะของกลุ่มพบว่าข้อมูลที่ 1-5 เมื่อคำนวณระยะห่างกับคุณลักษณะเริ่มต้นของกลุ่มที่ 1 (4,30) ทีละข้อมูลได้ 70.0, 40.4, 10.2, 3.0, 20.4 ตามลำดับ และข้อมูลที่ 1-5 เมื่อคำนวณระยะห่างกับคุณลักษณะเริ่มต้นของกลุ่มที่ 2 (6,100) ทีละข้อมูลได้ 3.0, 30.3, 60.0, 70.2, 50.0 จะได้ว่าในกลุ่มที่ 1 มีสมาชิกคือข้อมูลที่ 3, 4, 5 ส่วนกลุ่มที่ 2 มีสมาชิกคือข้อมูลที่ 1, 2 และทำการคำนวณคุณลักษณะของกลุ่มใหม่คือใช้จุดศูนย์กลางมวลของสมาชิกจะได้ ตัวแทนใหม่ของกลุ่มที่ 1 เป็น (5, 40) ตัวแทนใหม่ของกลุ่มที่ 2 เป็น (6.5, 85)

หลังจากนั้นทำการหาสมาชิกกลุ่มใหม่และหาตัวแทนกลุ่มใหม่ ทำต่อไปจนกระทั่งสมาชิกในกลุ่มไม่เปลี่ยนแปลงหรือตัวแทนของกลุ่มไม่เปลี่ยนแปลง

2.3.2.3 Density-base

การจัดกลุ่มแบบ Density-base เป็นการจัดกลุ่มโดยอาศัยพื้นฐานของความหนาแน่นข้อมูล มีข้อดีคือสามารถสร้างกลุ่มโดยไม่จำกัดรูปร่างของกลุ่มได้ อัลกอริทึมทำงานเพียงครั้งเดียวมีหลักการและต้องการพารามิเตอร์ดังต่อไปนี้

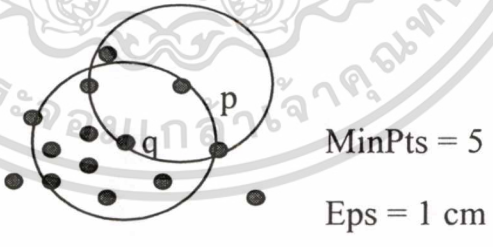
การจัดกลุ่มแบบ Density-base ใช้พารามิเตอร์ 2 ตัวคือ Eps: คือค่ารัศมีสูงสุดของกลุ่มไปยังกลุ่มใกล้เคียง MinPts: คือจำนวนข้อมูลที่น้อยที่สุดในกลุ่ม การจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

60859

แบบ Density-base สามารถอธิบายหลักการทำงานโดยพิจารณารูปที่ 2.7 คือข้อมูลที่เตรียมไว้สำหรับการจัดกลุ่ม และในรูปที่ 2.8 กำหนดให้จุด p เป็นจุดศูนย์กลางและกำหนดรัศมี Eps = 1 เซนติเมตร (cm.) พบว่ามีมีสมาชิกข้อมูลของกลุ่ม MinPts = 5 ดังนั้นจะได้พื้นที่วงกลม p เป็นขอบเขตของกลุ่ม ต่อมาทำการขยายพื้นที่กลุ่มโดยการเลือกสมาชิกหนึ่งค่าจากห้าค่าในวงกลม p เป็นจุดศูนย์กลาง ในรูปที่ 2.8 เลือกจุด q เป็นจุดศูนย์กลางและกำหนดรัศมี Eps = 1 เซนติเมตร (cm.) พบว่ามีสมาชิกของกลุ่ม MinPts > 5 ดังนั้นวงกลม q เป็นขอบเขตของกลุ่มเดียวกับวงกลม p ด้วย และการจัดกลุ่มแบบ Density-base จะทำการสร้างกลุ่มใหม่ไปเรื่อยๆ (โดยมีเงื่อนไขว่ากลุ่มที่สร้างใหม่จะมีรัศมีคือค่า Eps และในพื้นที่ของกลุ่มที่สร้างใหม่นี้จะต้องมีสมาชิกของกลุ่มอย่างน้อย Minpts ข้อมูล) จนกระทั่งไม่สามารถสร้างกลุ่มได้อีก และกลุ่มที่สร้างใหม่ทั้งหมดนี้ Density-base จะถือว่าเป็นขอบเขตของกลุ่มเดียวกันหมด

รูปที่ 2.7 แสดงข้อมูลที่ต้องการจัดกลุ่ม



รูปที่ 2.8 แสดงการนำข้อมูลจากรูปที่ 2.7 มาทำการจัดกลุ่มข้อมูลแบบ Density-base

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

NNC Architecture

(Neural Network Classifier Architecture)

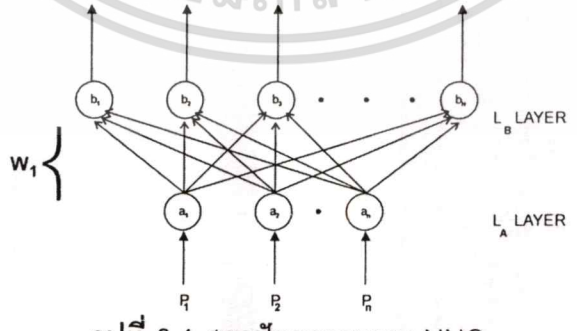
บทนี้กล่าวถึงหลักการแบ่งกลุ่มวิธีหนึ่งที่ใช้หลักการ Neural Network โดยงานวิจัยของ M.A. ABOU-NASR [1] ได้นำหลักการนี้ไปใช้จัดกลุ่มข้อมูล และในวิทยานิพนธ์นี้ก็ใช้หลักการเดียวกัน หัวข้อที่ 3.1 กล่าวถึงสถาปัตยกรรมของนิวรอนเน็ตเวอร์ก หัวข้อ 3.2 แสดงอัลกอริทึมที่ใช้ในขั้นตอนการเรียนรู้ และหัวข้อที่ 3.3 แสดงตัวอย่างข้อมูลที่ใช้อัลกอริทึมนี้จัดหมวดหมู่ของข้อมูลสำหรับข้อมูลที่ไม่มีการซ้อนทับกันเท่านั้น ส่วนในบทที่ 4 จะเป็นการนำเสนอต้นแบบใหม่ที่ใช้แบ่งกลุ่มข้อมูลที่มีการซ้อนทับกัน

3.1 สถาปัตยกรรมของ NNC (Neural Network Classifier Architecture)

NNC (Neural Network Classifier) แสดงดังรูปที่ 3.1 มี ชั้นอินพุท(input layer) $L_A = (a_1, a_2, a_3,\ldots,a_n)$ ประกอบด้วย n นิวรอน เพื่อรับข้อมูลอินพุท P ที่มี n คุณลักษณะ แต่ละนิวรอนในชั้น L_A จะต่อกับนิวรอนในชั้น L_B คือ $L_B = (b_1, b_2, b_3, \ldots,b_n)$ และน้ำหนักจะถูกกำหนดจากนิวรอนในชั้น L_A ไปยังนิวรอนในชั้น L_B ตามสมการ (3.1)

$$W_1 = (w_1 w_2 \ldots w_N)^T \tag{3.1}$$

โดยที่ $W_i = (W_{b_i a_1}, W_{b_i a_2}, \ldots, W_{b_i a_n})$ เป็นน้ำหนักของเวกเตอร์ตัวที่ i ในนิวรอนชั้น L_B



รูปที่ 3.1 สถาปัตยกรรมของ NNC

นิวรอนในชั้น L_B จะใช้แทนกลุ่มๆ หนึ่ง ที่สามารถสร้างขึ้นในระหว่างขั้นตอนการเรียนรู้ (learning) โดยสามารถที่จะมีมากกว่าหนึ่งนิวรอนสำหรับกลุ่มเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอาท์พุทที่ได้คือข้อมูล P_k สำหรับใช้ระบุจำนวนกลุ่มทั้งหมดที่สามารถเป็นไปได้จากขั้นตอนการเรียนรู้ตามสมการที่ (3.2)

$$c_1, c_2, ...c_k, c_m \tag{3.2}$$

c_k เป็นนิวรอนในชั้น L_B ของ NNC สำหรับกลุ่ม k

m เป็นจำนวนกลุ่มทั้งหมดที่เกิดจากการเรียนรู้การแบ่งกลุ่มด้วย NNC

3.2 อัลกอริทึมของ NNC (NNC Training Algorithm)

อัลกอริทึมของ NNC มี $p^k = \{p_1^k, p_2^k, ...p_n^k\}$ เป็นอินพุท และสามารถแสดงการทำงานได้ดังนี้

1. สร้างต้นแบบนิวรอนในชั้น L_B ขึ้นใหม่สำหรับข้อมูลอินพุท P ที่ไม่สามารถระบุกลุ่มได้
2. ปรับน้ำหนักระหว่างชั้นของนิวรอน ที่วัดระยะห่างของข้อมูลแบบยูคลิด(Euclidean distance) ในสมการที่ 3.3 แล้วได้ระยะทางน้อยที่สุด โดยถ้าระยะทางดังกล่าวน้อยกว่าค่า threshold ของกลุ่มนั้น ให้จัดข้อมูลนี้อยู่ในกลุ่มเดียวกัน ระยะทางยูคลิดจากข้อมูล P ไปยังนิวรอน b_j คือ

$$b_{jT} = \sqrt{\sum_{s=1}^{n} (W_{b_j a_s} - P_s)^2} \tag{3.3}$$

$W_{b_j a_s}$ เป็นค่าน้ำหนักระหว่างนิวรอน b_j ในชั้นของ L_B และ นิวรอน a_s ในชั้น L_A

ปรับน้ำหนักเฉลี่ยของนิวรอนตัวที่ข้อมูล P อยู่ใกล้ที่สุดตามสมการ (3.4)

$$W_{b_j a_{inew}} = \frac{W_{b_j a_{iold}} + P_i}{M+1}, \forall i = 1,2,...,n \tag{3.4}$$

$W_{b_j a_{inew}}$ เป็นค่าน้ำหนักใหม่ระหว่างนิวรอน j ในชั้น L_B และนิวรอน i ในชั้น L_A

$W_{b_j a_{iold}}$ เป็นค่าน้ำหนักเก่าระหว่างนิวรอน j ในชั้น L_B และนิวรอน i ในชั้น L_A

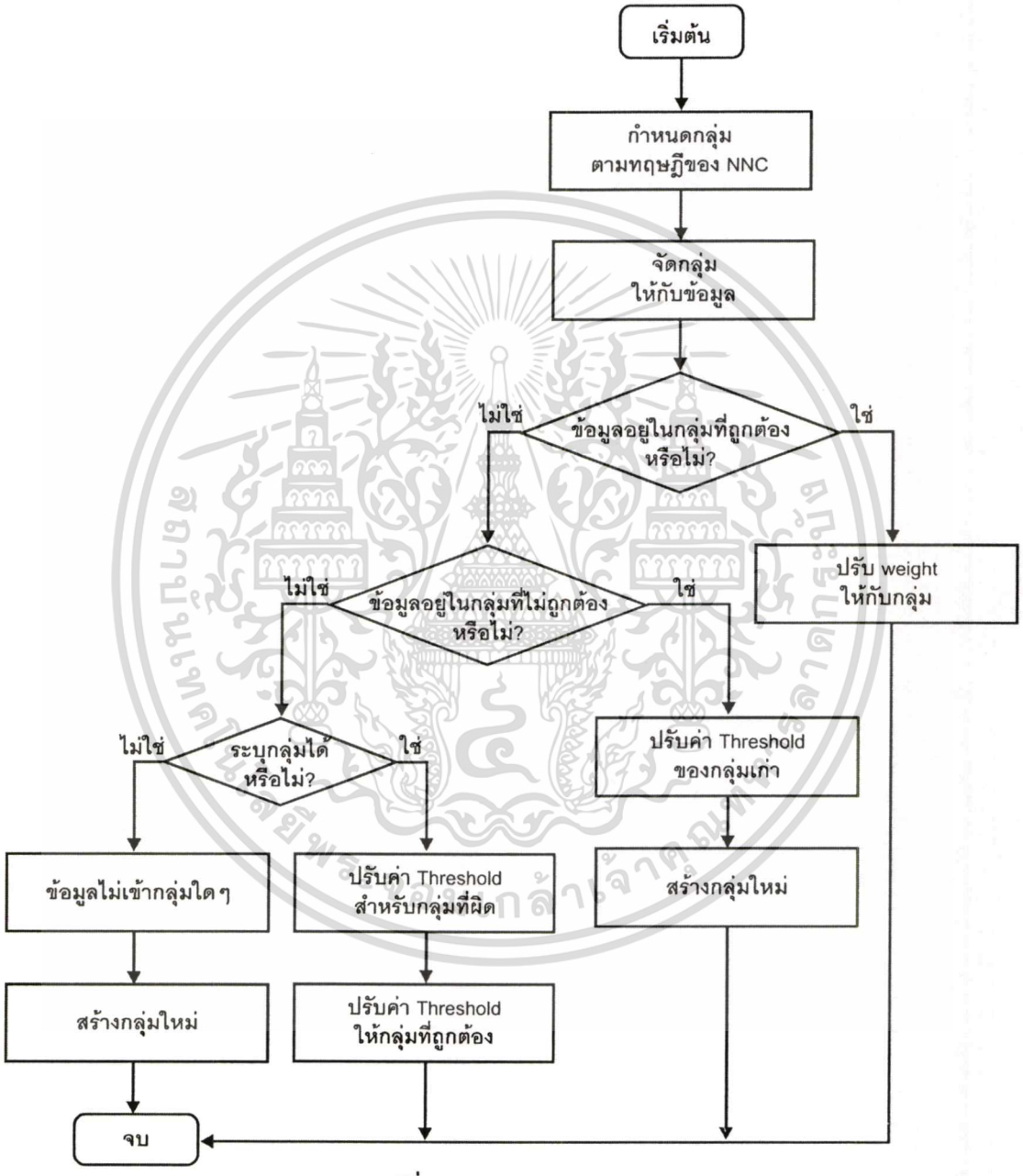
p_i เป็นค่าคุณลักษณะ(feature)ของข้อมูล P

M เป็นค่าจำนวนข้อมูลที่เข้ามาทดสอบในอัลกอริทึม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. สำหรับกรณีที่มีข้อมูลอินพุท P วัดระยะทางยูคลิดจากข้อมูล P ไปยังนิวรอนใดๆ ที่อยู่ต่างกลุ่มกับอินพุท P แล้ว น้อยกว่าค่าระยะทางนิวรอน threshold จะทำการลดนิวรอน threshold เพื่อกำจัดข้อมูล P ให้อยู่นอกกลุ่มเนื่องจากเป็นข้อมูลต่างกลุ่มกัน

สามารถแสดงอัลกอริทึมของ NNC ในรูปแผนผังการทำงาน (flowchart)ได้ดังรูปที่ 3.2

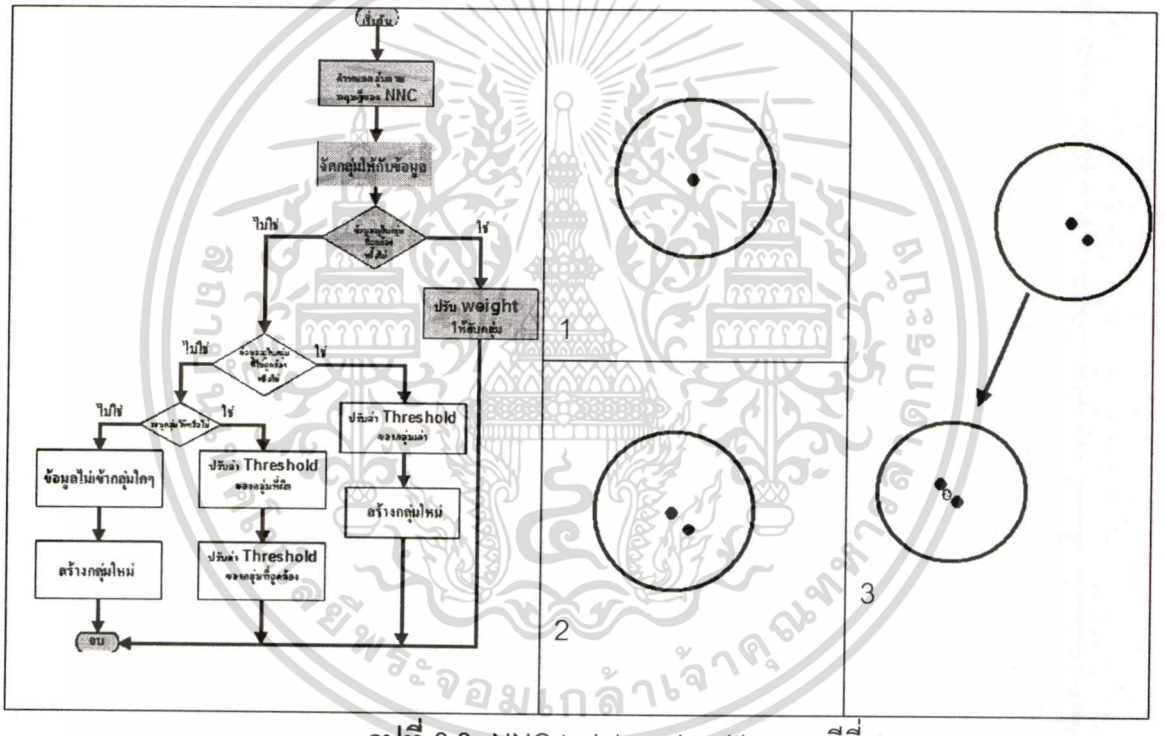


รูปที่ 3.2 NNC training algorithm

จากรูปที่ 3.2 อัลกอริทึมของ NNC สามารถแบ่งการทำงานออกเป็น 4 รูปแบบด้วยกันคือ กรณีที่ 1 รูปที่ 3.3 เมื่อมีข้อมูลตัวแรกเข้ามา NNC จะทำการสร้างกลุ่มใหม่ให้กับข้อมูลนั้น กำหนดให้ข้อมูลนั้นเป็นน้ำหนักเฉลี่ยของกลุ่มและรัศมีวงกลมให้กว้างที่สุดเท่าที่จะทำได้ ต่อมา

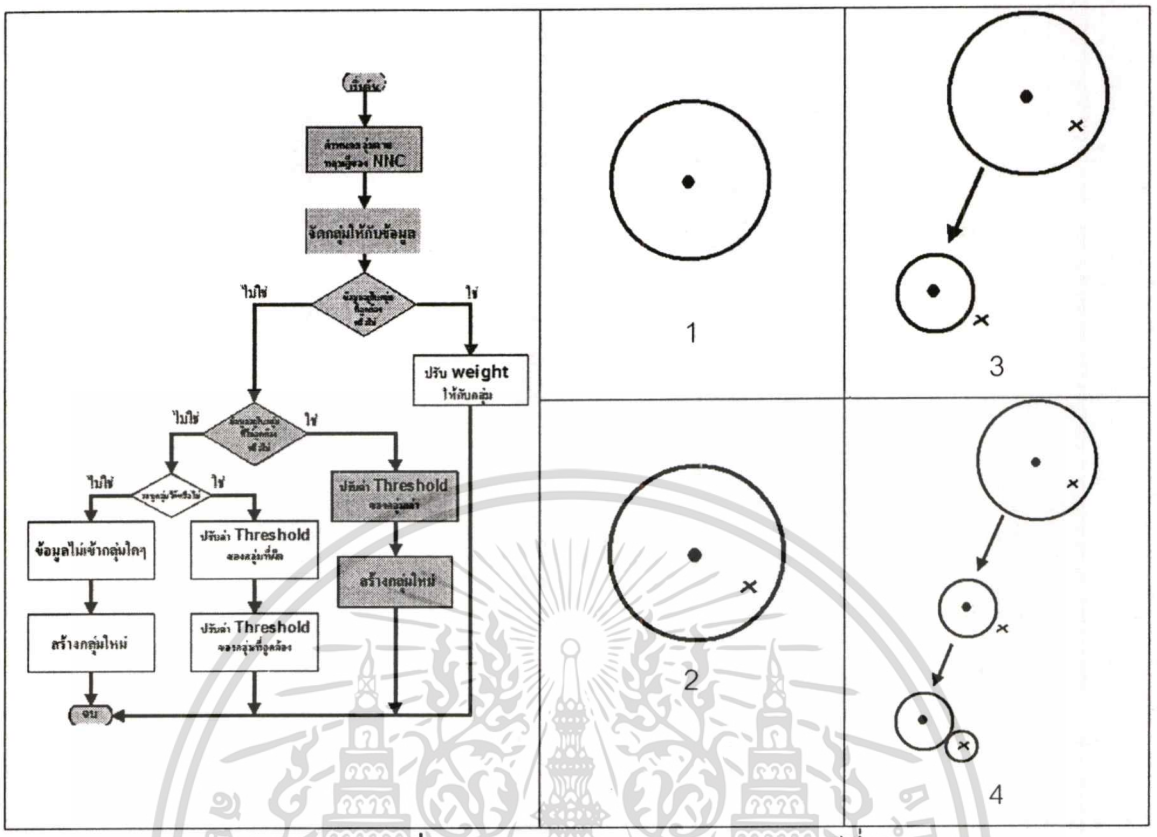
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อมีข้อมูลกลุ่มเดียวกันตกในรัศมีกลุ่มเดิม NNC จะทำการปรับน้ำหนักเฉลี่ยของกลุ่มด้วยสมการที่ (3.3) ในกรณีที่ 2 รูปที่ 3.4 เมื่อมีข้อมูลต่างกลุ่มกันตกในรัศมีกลุ่มข้อมูลอื่น NNC จะทำการลดค่ารัศมีหรือ threshold ของกลุ่มแรกและสร้างกลุ่มใหม่คือการกำหนดน้ำหนักเฉลี่ยและรัศมีของกลุ่มใหม่ กรณีที่ 3 รูปที่ 3.5 กรณีนี้มีกลุ่ม 2 กลุ่มอยู่ก่อนแล้ว และมีข้อมูลตัวใหม่เข้ามาตกผิดกลุ่ม NNC จะทำการลด threshold กลุ่มที่ข้อมูลตกผิดและเพิ่ม threshold กลุ่มที่ถูกพร้อมทั้งปรับน้ำหนักกลุ่มที่ข้อมูลที่เพิ่มค่า threshold และกรณีสุดท้ายรูปที่ 4 รูป 3.6 เมื่อมีข้อมูลใหม่เข้ามาแต่ไม่สามารถระบุกลุ่มได้ NNC จะถือเป็นข้อมูลกลุ่มใหม่โดยจะทำการสร้างน้ำหนักเฉลี่ยและรัศมีของกลุ่มใหม่

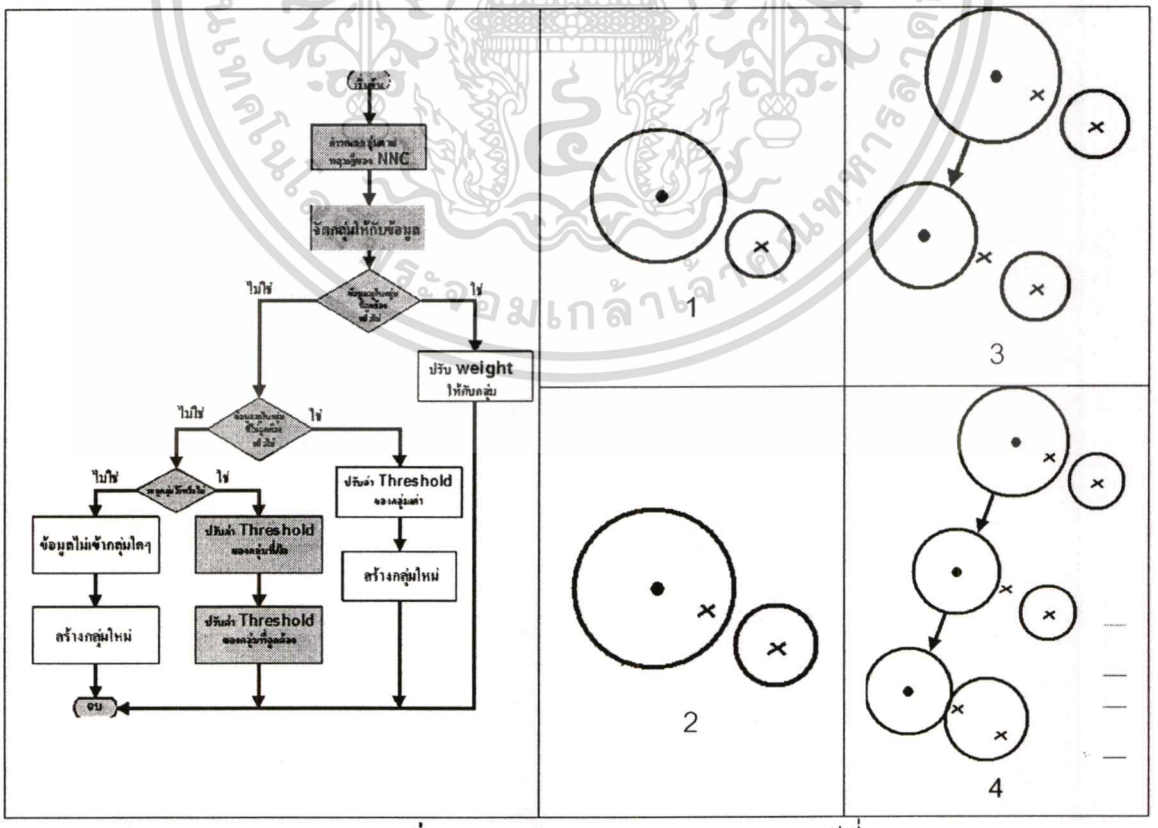


รูปที่ 3.3 NNC training algorithm กรณีที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

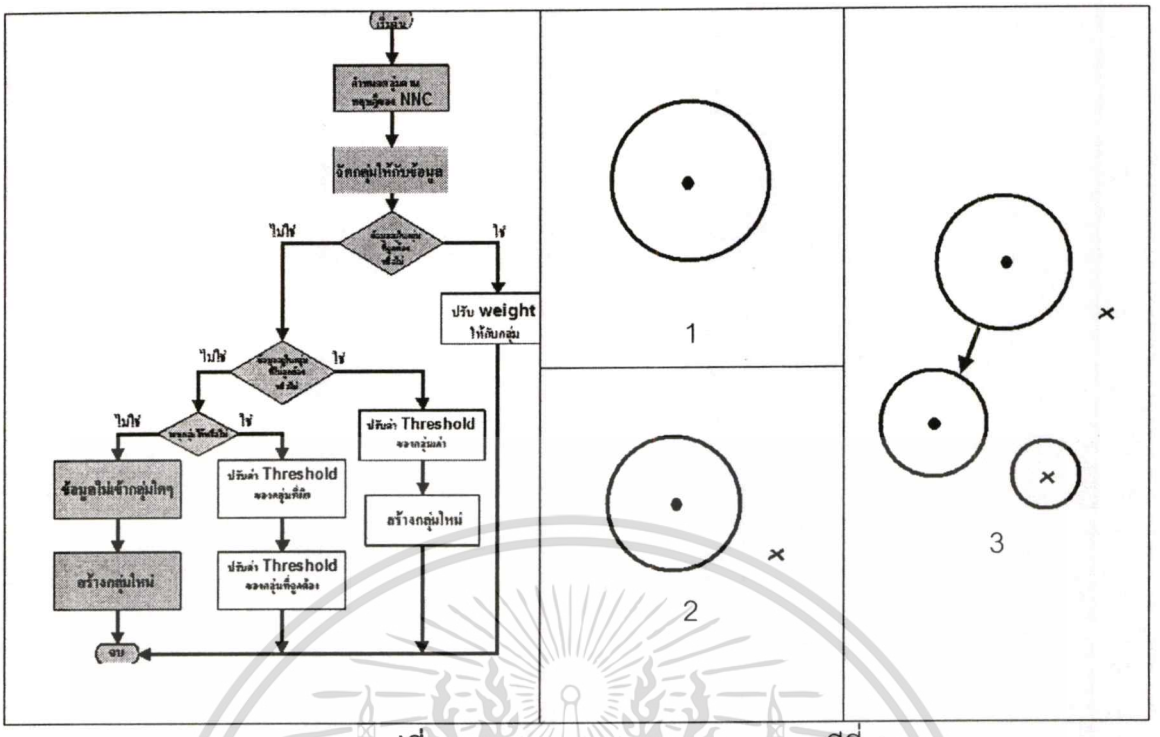


รูปที่ 3.4 NNC training algorithm กรณีที่ 2



รูปที่ 3.5 NNC training algorithm กรณีที่ 3

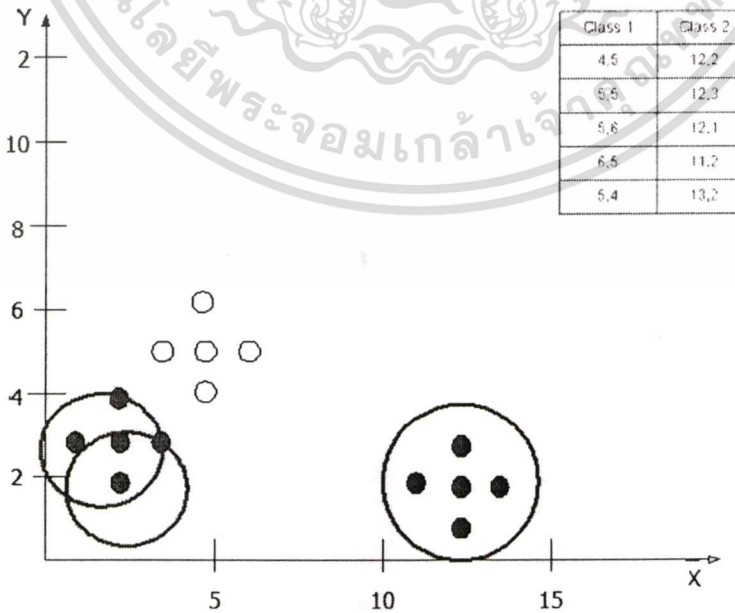
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 NNC training algorithm *กรณีที่* 4

3.3 ผลการจัดกลุ่มโดยใช้อัลกอริทึมของ NNC กับข้อมูลที่ไม่มีการซ้อนทับกัน

จากรูปที่ 3.7 แสดงตัวอย่างการจัดกลุ่มอัลกอริทึม NNC โดยมีข้อมูลที่ใช้ในขั้นตอนการเรียนรู้ 15 ข้อมูล 3 กลุ่ม ตามตารางในภาพที่ 3.7 โดยแต่ละค่ามี 2 คุณลักษณะคือค่า x และ y สามารถแบ่งกลุ่มข้อมูลได้ 4 กลุ่ม และในบทที่ 5 จะนำอัลกอริทึมเดียวกันนี้จัดกลุ่มข้อมูลที่มีการซ้อนทับกัน เปรียบเทียบผลการทดลองกับวิทยานิพนธ์นี้



Class 1	Class 2	Class 3
4.5	12.2	3.3
5.5	12.3	3.4
5.8	12.1	3.2
6.5	11.2	2.3
5.4	13.2	4.3

รูปที่ 3.7 ตัวอย่างการจัดข้อมูลด้วย NNC สำหรับข้อมูลที่ไม่มีการซ้อนทับกัน

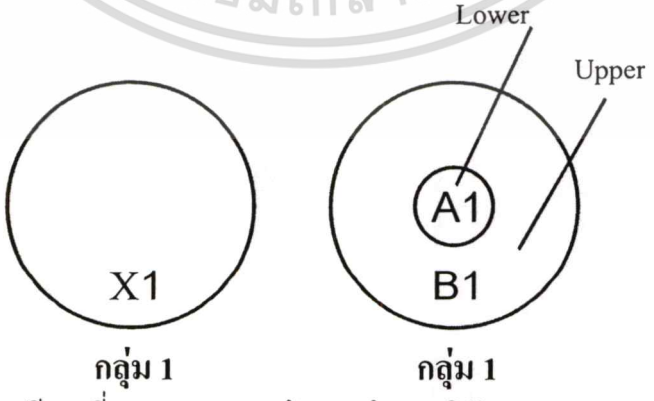
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การนำเสนอต้นแบบของการจัดกลุ่มข้อมูล ที่มีการซ้อนทับกัน

บทนี้จะกล่าวถึงรายละเอียดของต้นแบบการจัดกลุ่มสำหรับข้อมูลที่มีการซ้อนทับกัน โดยนำเอาทฤษฎีของ NNC จากบทที่3 มาควบคุมการทำงานของต้นแบบในวิทยานิพนธ์โดยเรียกวิธีนี้ว่า Modified Neural Network Classifier (MNNC) จากนั้นมีการกำหนดกำหนดพื้นที่ต่างๆ ของต้นแบบที่นำเสนอ โดยแบ่งพื้นที่ของกลุ่มเป็นส่วนของ lower upper และส่วนที่ซ้อนทับกัน (Overlap) กัน โดยขอยืมความหมายของคำศัพท์คำว่า lower และ upper จากทฤษฎี rough set ในงานวิจัยของ Sarah Coppock [3] จากนั้นจะทำการแสดงอัลกอริทึมที่ใช้ในงานวิจัยและอธิบายการทำงานของอัลกอริทึม ในส่วนท้ายของบทจะเป็นการการดำเนินผลการทดลอง และจะรายงานผลการทดลองในบทที่ 5 ต่อไป

4.1 ต้นแบบของกลุ่มข้อมูล

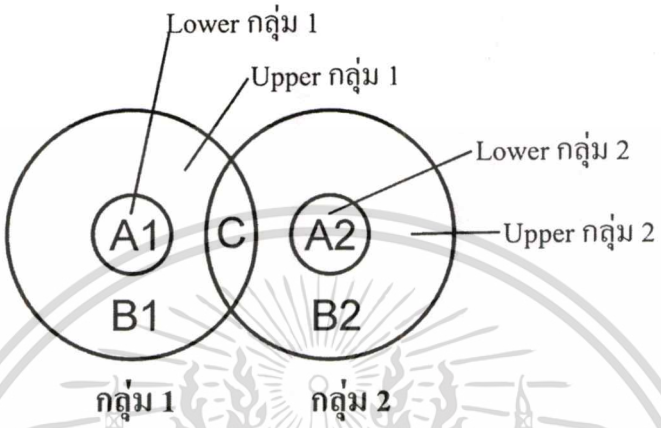
เป็นการนำเสนอต้นแบบสำหรับข้อมูลที่มีการซ้อนทับกัน (Overlap data pattern presentation) ดังรูปที่ 4.1 ด้านขวา จากรูปสามารถแบ่งกลุ่มของข้อมูลออกเป็น 2 ส่วน คือส่วน lower และ ส่วน upper โดยส่วน lower คือพื้นที่ A1 หมายความว่าข้อมูลที่อยู่ในพื้นที่ A1 เป็นข้อมูลที่มีความแน่นอนว่าเป็นข้อมูลของกลุ่ม 1 ส่วน upper คือพื้นที่ B1 หมายความว่า ข้อมูลที่อยู่ในส่วนของ B1 เป็นส่วนที่มีโอกาสเป็นข้อมูลของกลุ่ม 1 และกลุ่มอื่นๆ สำหรับงานวิจัยของ M.A. ABOU-NASR [1]นำเสนอต้นแบบกลุ่มหนึ่งจะมีข้อมูลแค่ส่วนเดียวดังรูปที่ 4.1 ซ้าย หมายความว่าข้อมูลที่อยู่ในพื้นที่ X1 เป็นข้อมูลของกลุ่ม 1



รูปที่ 4.1 แสดงการเปรียบเทียบ prototype ต้นแบบในงานวิจัยของ M.A. ABOU-NASR (ซ้าย) และ prototype ต้นแบบของงานวิจัยนี้ (ขวา)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากต้นแบบที่ได้นำเสนอในรูปที่ 4.1 เมื่อนำไปใช้จัดกลุ่มข้อมูลที่มีการซ้อนทับกันจะได้ดังรูปที่ 4.2 ส่วน A1 และ A2 เป็นข้อมูลของกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ ส่วน B1 เป็นพื้นที่ที่มีโอกาสเป็นข้อมูลของกลุ่มอื่นๆ หรือ กลุ่มที่ 1 ส่วน B2 เป็นพื้นที่ที่มีโอกาสเป็นข้อมูลของกลุ่มอื่นๆ หรือ กลุ่มที่ 2 ส่วน C เป็นพื้นที่ที่มีโอกาสเป็นข้อมูลของกลุ่มที่ 1 หรือ กลุ่มที่ 2 ก็ได้



รูปที่ 4.2 แสดงการกำหนดพื้นที่ของส่วน lower และ upper สำหรับข้อมูลซ้อนทับกัน

การทำงานของอัลกอริทึมในการจัดกลุ่มใช้หลักการของ NNC [1] จะประกอบไปด้วยการปรับค่าจุดศูนย์กลางหรือน้ำหนักของกลุ่มและกำหนดรัศมีหรือค่า threshold ของกลุ่มทั้งสองค่า สำหรับการปรับค่าน้ำหนักของกลุ่มสามารถคำนวณได้ตามสมการ (4.1)

$$W_{b_j a_{inew}} = \frac{W_{b_j a_{iold}} + P_i}{M+1}, \forall i = 1,2,...,n \qquad (4.1)$$

$W_{b_j a_{inew}}$ เป็นค่าน้ำหนักเฉลี่ยใหม่ของกลุ่ม

$W_{b_j a_{iold}}$ เป็นค่าน้ำหนักเฉลี่ยเดิมของกลุ่ม

p_i เป็นค่าคุณลักษณะ(feature)ของข้อมูล

M เป็นค่าจำนวนข้อมูลที่เข้ามาทดสอบในอัลกอริทึม

สำหรับการปรับค่า threshold (lower และ upper) ใช้ฟังก์ชันการวัดระยะห่างแบบยูคลิด(Euclidean distance) โดยสามารถคำนวณได้ตามสมการ (4.2) ซึ่งเป็นการวัดระยะห่างของข้อมูลใหม่กับค่าจุดศูนย์กลางหรือน้ำหนักของกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$b_{jT} = \sqrt{\sum_{s=1}^{n} (W_{b_j a_s} - P_s)^2} \qquad (4.2)$$

$W_{b_j a_s}$ เป็นค่าน้ำหนักของกลุ่มที่กำลังพิจารณา

P_s เป็นค่าคุณลักษณะ(feature)ของข้อมูล

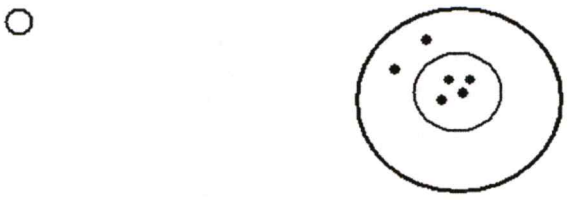
b_j เป็นค่าระยะห่างของกลุ่มที่กำลังพิจารณากับข้อมูลใหม่

เริ่มต้นอัลกอริทึมจะทำการพิจารณาข้อมูลของการเรียนรู้ทีละค่า เมื่อข้อมูลตัวแรกที่เข้ามาอัลกอริทึมจะกำหนดข้อมูลนี้เป็นน้ำหนักของกลุ่ม และกำหนด ค่า threshold ของกลุ่ม 2 ค่าคือ lower threshold และ upper threshold โดยกำหนดครั้งแรกให้มีค่าเท่ากัน จากนั้นเมื่อข้อมูลตัวที่สองเข้ามาอัลกอริทึมจะทำการหาวัดระยะห่างของข้อมูลกับทุกกลุ่มเพื่อปรับค่า threshold ของกลุ่มโดยสามารถพิจารณาภาพรวมได้ 2 หัวข้อดังต่อไปนี้

4.1.1 การปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกันดังรูปที่ 4.3 ใช้สัญลักษณ์จุดทึบสีดำแทนข้อมูลใหม่และกลุ่มข้อมูลที่กำลังพิจารณาคือจุดสีดำทึบเหมือนกัน

รูปที่ 4.3 แสดงข้อมูลที่กำลังทดสอบเป็นกลุ่มเดียวกับกลุ่มทดสอบ

4.1.2 การปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกันดังรูปที่ 4.4 ดังรูปที่ 4.4 ใช้สัญลักษณ์จุดสีดำขาวแทนข้อมูลใหม่และกลุ่มข้อมูลที่กำลังพิจารณาคือจุดสีดำทึบซึ่งต่างกลุ่มกัน

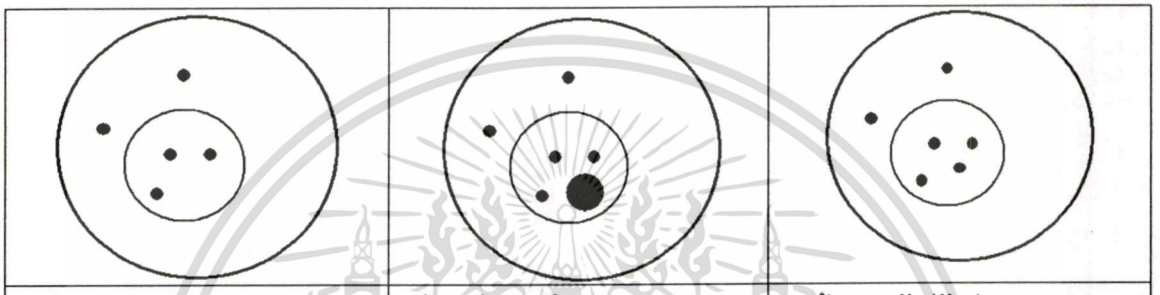


รูปที่ 4.4 แสดงข้อมูลที่กำลังทดสอบต่างกลุ่มกับกลุ่มทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

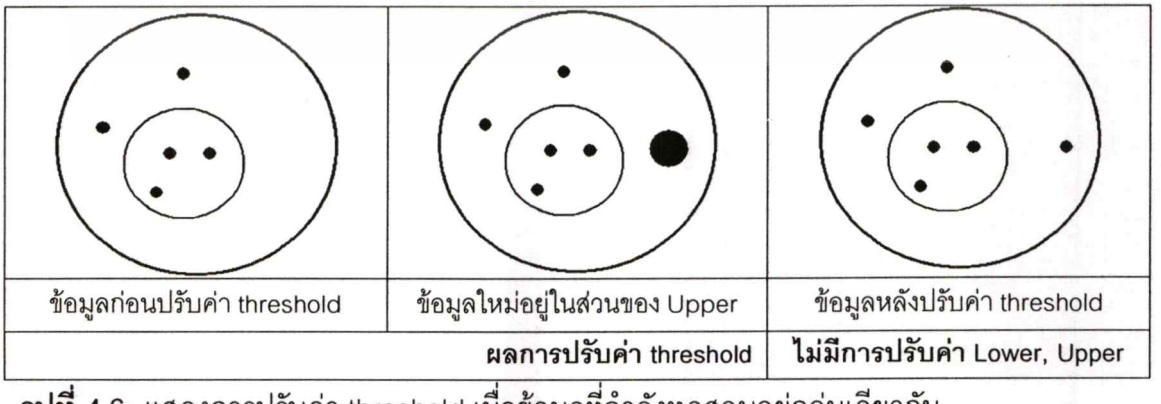
จากหัวข้อที่ 4.1.1 จะพบว่าการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน เป็นไปตามรูปที่ 4.5-4.8 และสามารถอธิบายรายละเอียดเป็นหัวข้อย่อยดังนี้

4.1.1.1 การปรับค่าเมื่อข้อมูลที่กำลังทดสอบอยู่ใน lower threshold ดังรูปที่ 4.5 ด้านซ้ายคือกลุ่มข้อมูลเดิม รูปที่ 4.5 กลางคือข้อมูลใหม่(วงกลมทึบใหญ่) ที่เมื่อคำนวณระยะห่างจากข้อมูลถึงกลุ่มที่กำลังทดสอบแล้วอยู่ในพื้นที่ lower ของกลุ่ม อัลกอริทึมจะไม่มีการปรับค่า threshold ใดๆ ของกลุ่ม ดังรูปที่ 4.5 ด้านขวา เพราะข้อมูลอยู่ในพื้นที่ที่ถูกต้องแล้ว อัลกอริทึมทำการคำนวณเฉลี่ยน้ำหนักของกลุ่มเท่านั้น



รูปที่ 4.5 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Lower

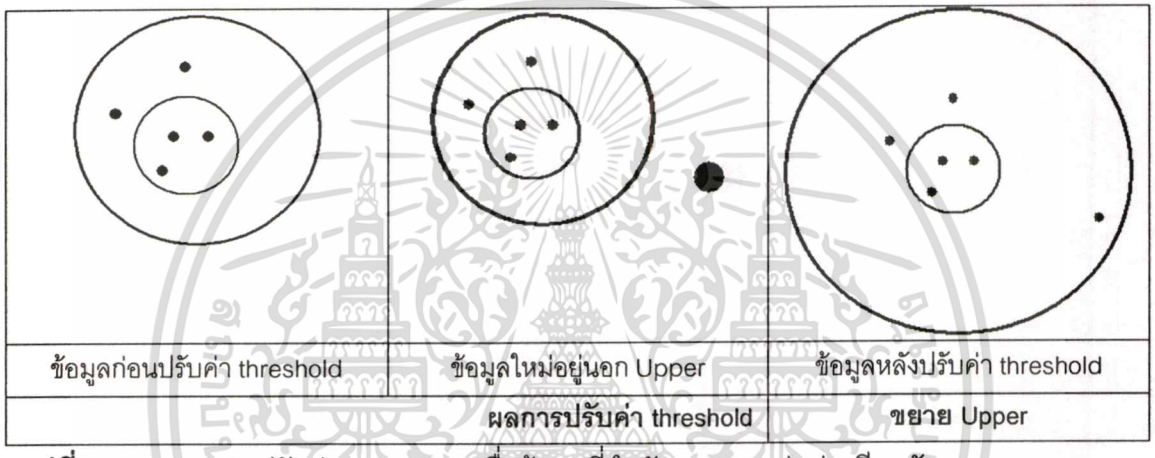
4.1.1.2 การปรับค่าเมื่อข้อมูลที่กำลังทดสอบอยู่ระหว่าง lower threshold และ upper threshold ดังรูปที่ 4.6 ด้านซ้ายคือกลุ่มข้อมูลเดิม รูปที่ 4.6 กลางคือข้อมูลใหม่ที่เมื่อคำนวณระยะห่างจากข้อมูลถึงกลุ่มที่กำลังทดสอบแล้วอยู่ในพื้นที่ ระหว่างค่า lower และค่า upper ของกลุ่ม อัลกอริทึมจะไม่มีการปรับค่า threshold ใดๆ ของกลุ่ม ดังรูปที่ 4.6 ด้านซ้าย เพราะข้อมูลอยู่ในพื้นที่ที่ถูกต้องแล้ว อัลกอริทึมทำการคำนวณเฉลี่ยน้ำหนักของกลุ่มเท่านั้น



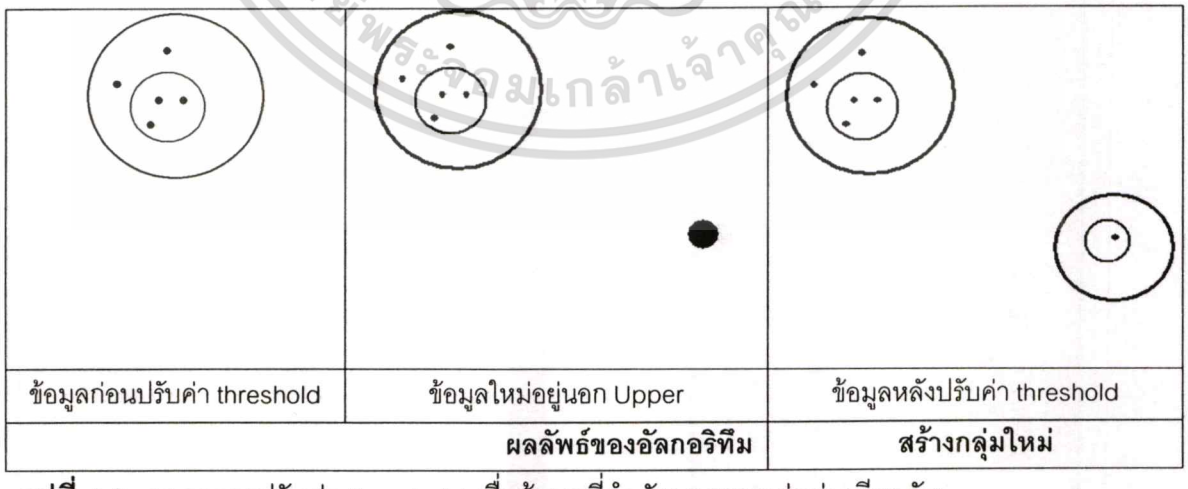
รูปที่ 4.6 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Upper

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1.3 การปรับค่าเมื่อข้อมูลที่กำลังทดสอบอยู่นอก upper threshold ดังรูปที่ 4.7 และรูปที่ 4.8 สำหรับรูปที่ 4.7 ด้านซ้ายคือกลุ่มข้อมูลเดิม รูปที่ 4.7 กลางคือข้อมูลใหม่ ที่เมื่อคำนวณระยะห่างจากข้อมูลถึงกลุ่มที่กำลังทดสอบแล้วอยู่นอกพื้นที่ของกลุ่ม อัลกอริทึมจะทำการปรับค่า threshold โดยแบ่ง 2 กรณีคือ กรณีที่ข้อมูลใหม่อยู่นอก upper threshold มีระยะทางน้อยกว่า 1.2 เท่าของ upper threshold ของกลุ่ม อัลกอริทึมจะขยาย upper threshold ของกลุ่มนั้นดังรูปที่ 4.7 ด้านขวา กรณีที่ข้อมูลใหม่อยู่นอก upper threshold มากกว่า 1.2 เท่า อัลกอริทึมจะสร้างกลุ่มใหม่ดังรูปที่ 4.8



รูปที่ 4.7 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่นอกพื้นที่ Upper (ระยะทางน้อยกว่า 1.2 เท่าของ Upper)

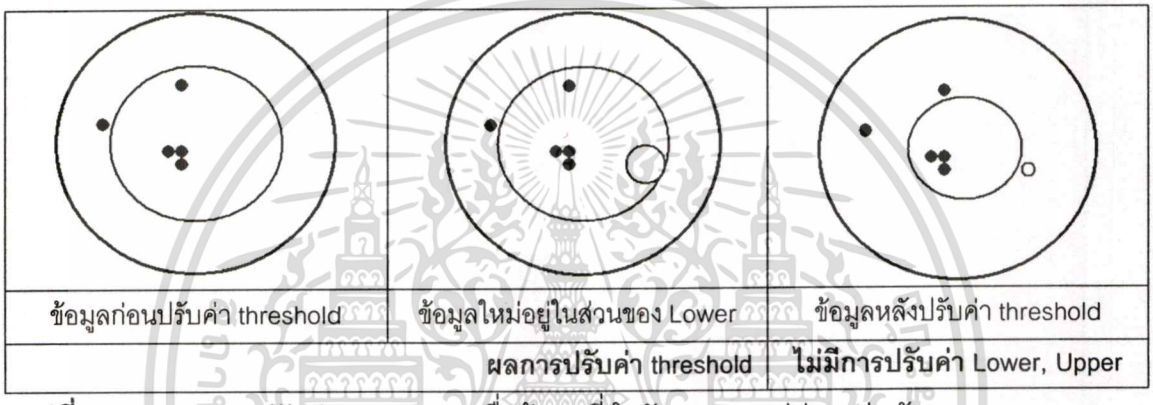


รูปที่ 4.8 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่กลุ่มเดียวกัน กรณีข้อมูลใหม่อยู่นอกพื้นที่ Upper (ระยะมากกว่า 1.2 เท่าของ Upper)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากหัวข้อที่ 4.1.2 จะพบว่าการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน เป็นไปตามรูปที่ 4.9-4.11 และสามารถอธิบายรายละเอียดเป็นหัวข้อย่อยดังนี้

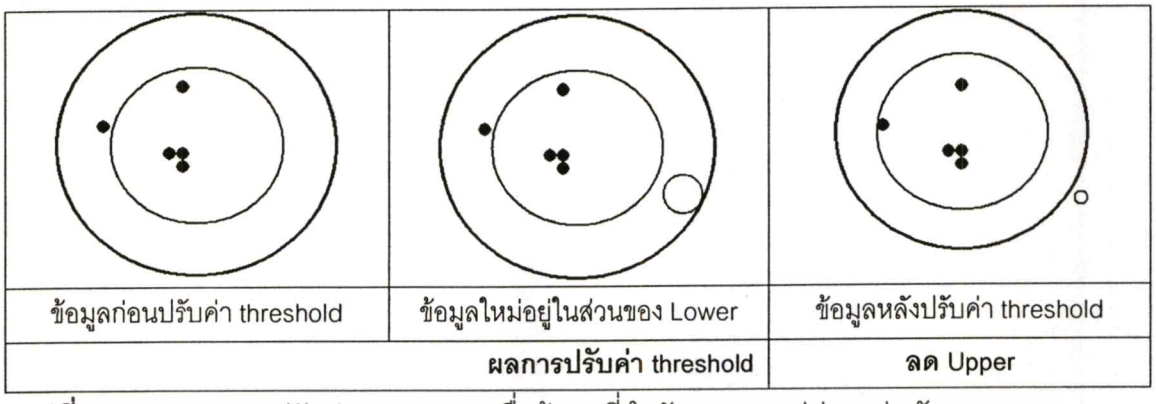
4.1.2.1 การปรับค่าเมื่อข้อมูลใหม่ที่กำลังทดสอบอยู่ใน lower threshold ดังรูปที่ 4.9 ด้านซ้ายคือกลุ่มข้อมูลเดิม รูปที่ 4.9 กลางคือข้อมูลใหม่(วงกลมใหญ่สีขาว) ที่เมื่อคำนวณระยะห่างจากข้อมูลถึงกลุ่มที่กำลังทดสอบแล้วข้อมูลอยู่ในพื้นที่ lower ของกลุ่ม อัลกอริทึมจะปรับลดค่า lower threshold เพื่อให้ข้อมูลของกลุ่มถูกต้อง ดังรูปที่ 4.9 ด้านขวา



รูปที่ 4.9 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Lower

4.1.2.2 การปรับค่าเมื่อข้อมูลที่กำลังทดสอบอยู่ระหว่าง lower threshold และ upper threshold ดังรูปที่ 4.10 ด้านซ้ายคือกลุ่มข้อมูลเดิม รูปที่ 4.10 กลางคือข้อมูลใหม่ ที่เมื่อคำนวณระยะห่างจากข้อมูลถึงกลุ่มที่กำลังทดสอบแล้วข้อมูลอยู่ในพื้นที่ระหว่าง lower และ upper ของกลุ่ม อัลกอริทึมจะทดสอบปรับลดค่า upper threshold แล้วตรวจสอบว่าข้อมูลเดิมของกลุ่มหายไปน้อยกว่า 20% จะลด upper threshold ดังรูปที่ 4.10 ขวา แต่ถ้าข้อมูลเดิมหายไปมากกว่า 20 % จะไม่ปรับลดค่า upper threshold

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน กรณีข้อมูลใหม่อยู่ในพื้นที่ Upper

4.1.2.3 การปรับค่าเมื่อข้อมูลที่กำลังทดสอบอยู่นอก upper threshold ดังรูปที่ 4.11 ด้านซ้ายคือกลุ่มข้อมูลเดิม รูปที่ 4.11 กลางคือข้อมูลใหม่ ที่เมื่อคำนวณระยะห่างจากข้อมูลถึงกลุ่มที่กำลังทดสอบแล้วข้อมูลอยู่นอกพื้นที่ upper threshold ของกลุ่ม อัลกอริทึมจะไม่ปรับลดค่า threshold ใดๆ เพื่อให้ข้อมูลของกลุ่มถูกต้อง ดังรูปที่ 4.11 ด้านขวา

รูปที่ 4.11 แสดงการปรับค่า threshold เมื่อข้อมูลที่กำลังทดสอบอยู่ต่างกลุ่มกัน กรณีข้อมูลใหม่อยู่นอกพื้นที่ Upper

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 อัลกอริทึมที่ใช้ในงานวิจัย

```
CONSTANT SMALL  is value for adjust radius.

while ( (radius reduce) || (new neuron) ) {
 for each pattern p in the training patterns{
  read pattern p;
  for(node=0; node<total neuron; node++){
      compute distance = Euclidean distance (weight[node], p);
      if (class of p == class [node]){
        if(distance < lower[node])
            mark fall flag for adjust weight;
        else if (distance < upper[node])
            mark fall flag for adjust weight;
        if (distance < (1.2 * upper[node]))
            upper[node] = distance – SMALL;          // expand upper
        else
            clear fall flag for create new node
      }// if (class of p == class [node])

      if (class of p != class [node]){
        if(distance < lower[node])
            lower[node] = distance - SMALL;          //reduce lower
        else if (distance < upper[node])
            if (reduce upper[node] at least 80 % of the pattern)
                upper[node] = distance – SMALL;
      }// if (class of p != class [node])
  }// for node
  if (pattern not fall in any node with the same class)
      create new node (p);
  else
      average pattern and weight in the nearest node with the same class;
 }// for each pattern p
}// while ( (radius reduce) || (new neuron) )
```

รูปที่ 4.12 อัลกอริทึมการจัดข้อมูลที่มีการซ้อนทับกัน(MNNC)

จากอัลกอริทึมในรูปที่ 4.12 จะนำค่าข้อมูลที่ใช้สำหรับการเรียนรู้มาทดสอบทีละค่าตามลำดับโดยจะทำการทดสอบทั้งในส่วน lower และ upper ของทุกกลุ่มที่ถูกสร้างขึ้นในอัลกอริทึม ถ้ามีข้อมูลใดตกผิดกลุ่มในพื้นที่ lower หรือ upper ของกลุ่มอื่นๆ จะทำการลดรัศมีเพื่อให้ขอบเขตของกลุ่มสามารถอธิบายข้อมูลได้ถูกต้อง ถ้าข้อมูลใหม่ตกอยู่ในพื้นที่ของกลุ่มที่ถูกต้องแล้วจะทำการปรับค่าน้ำหนักเฉลี่ยของกลุ่มที่ถูกต้องนั้น และถ้าข้อมูลไม่ได้อยู่ในพื้นที่ของ lower และ upper กลุ่มใดๆ จะมีการสร้างกลุ่มใหม่ให้กลับข้อมูลนี้ และจะทำการเรียนรู้นี้ซ้ำจนกระทั่งไม่มีการสร้างกลุ่มใหม่ และค่า threshold ของทุกกลุ่มไม่มีการเปลี่ยนแปลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ปัญหาและการปรับปรุงงานวิทยานิพนธ์

สำหรับหัวข้อ 4.3 นี้ จะกล่าวถึงที่มาของการทดลองในบทที่ 5 ของงานวิทยานิพนธ์นี้ ซึ่งได้นำเสนอผลการทดลอง 3 การทดลอง จะกล่าวถึงปัญหาที่พบในระหว่างการทดลองและการปรับปรุงการทดลอง

สำหรับการทดลองทั้ง 3 เป็นการนำข้อมูลของลายมือเขียนตัวอักษรภาษาไทยมี 4 คุณลักษณะ และระบุกลุ่มแน่นอนแล้วมาทดสอบ การจัดกลุ่มในการทดลองที่ 1 ใช้ข้อมูลจำนวน 70 ข้อมูล 3 คุณลักษณะ 7 กลุ่ม ทำการทดสอบกับโปรแกรม NNC ได้ผลลัพธ์การทดลองตามตารางที่ 5.2 ในบทที่ 5 สามารถจัดกลุ่มได้ 7 กลุ่ม และเพื่อให้ง่ายในการวิเคราะห์ผลการทดลองจึงแสดงผลการทดลองเป็นรูปภาพด้วยโปรแกรม MATLAB ดังรูปที่ 5.1 ในบทที่ 5 เนื่องจากเป็นข้อมูล 3 มิติข้อมูลในส่วนที่ซ้อนทับกัน อาจจะวิเคราะห์ยาก จึงเป็นที่มาของการทดลองที่ 2 คือลดคุณลักษณะลงเหลือ 2 คุณลักษณะ และใช้จำนวนข้อมูลมากขึ้นเป็น 354 ข้อมูล มี 22 กลุ่ม และแสดงผลในตาราง 5.3 และแสดงผลในรูปที่ 5.3 พบว่าสามารถจัดกลุ่มได้ 121 กลุ่ม ในการทดลองที่ 2 การปรับเงื่อนไขต่าง ๆ ของอัลกอริทึมโดยถ้าพิจารณาในรูปที่ 4.2 รูปอัลกอริทึมการจัดข้อมูลที่มีการซ้อนทับกันคือส่วนที่อยู่ในบรรทัดที่ 14, 21, 24 คือส่วนการปรับค่าพารามิเตอร์ และเนื่องจากข้อมูลที่ใช้ทดสอบในการทดลองนี้มีการซ้อนทับกันมากเพราะใช้ข้อมูลแค่ 2 คุณลักษณะเท่านั้น พบว่าในส่วนข้อมูลที่มีการซ้อนทับกันมากจะมีการกระจายของกลุ่มมาก ดังนั้นในการทดลองนี้จึงใช้เวลาทดลองเพื่อปรับค่าพารามิเตอร์มากที่สุดใน 3 การทดลองเพื่อให้ได้ข้อมูลที่เหมาะสมที่สุด โดยสามารถสรุปปัญหาต่างๆ และการแก้ไขในหัวข้อ 4.3.1-4.3.4

4.3.1 การแก้ปัญหาให้แบ่งกลุ่มได้ถูกต้องมากขึ้น

ในการทดลองที่ 2 หากพิจารณากลุ่ม ฒ รูปที่ 5.2 และ รูปที่ 5.3 ในบทที่ 5 โปรแกรม NNC version 1.13 จัดกลุ่มได้ตามตาราง ข.1 ถ้าพิจารณาจากความเหมาะสมตามรูปแล้ว กลุ่ม ฒ ค่า lower threshold ควรจะเท่ากับ upper threshold แต่ใน โปรแกรม NNC version 1.13 นี้ ทั้งสองค่าไม่เท่ากันดังแสดงด้านล่าง

ตาราง 4.1 แสดงผลการทดลองจัดกลุ่มข้อมูลกลุ่ม ฒ ของลายมือเขียนตัวอักษรภาษาไทยในการทดลองที่ 2

กลุ่ม ฒ lower threshold = 0.0868605 upper threshold = 0.0380890 และมีสมาชิกดังนี้
Lower ฒ055 ฒ070 ฒ047 ฒ041 ฒ018
Upper ฉ22 ฉ 90

สาเหตุเนื่องจากลำดับการเข้าทำงานของข้อมูล ในการทำงานรอบใหม่ทุกรอบ ฉ22 และ ฉ90 มาก่อน ฒ ทุกตัวทำให้ไม่สามารถลดค่า upper threshold ได้การแก้ไข เมื่อจบการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำงานแล้วให้โปรแกรมทำงานทั้งหมดอีกครั้งเพื่อทดสอบข้อมูลก่อนจบการทำงานและกำหนดให้เป็นโปรแกรม NNC version 1.14

4.3.2 ปัญหาการจัดกลุ่มในพื้นที่ lower ไม่ถูกต้อง

โปรแกรม NNC version 1.14 มีผลการทดลองเป็นดังตารางที่ 4.2 เมื่อพิจารณาแล้วพบว่ามีบางกลุ่มที่มีสมาชิกในกลุ่ม ไม่ถูกต้อง เช่น กลุ่มที่ 7 (ง), กลุ่ม 17(จ10), กลุ่ม 18(จ11) พบว่ามีสมาชิกกลุ่มอื่นอยู่ในพื้นที่ lower (LO=1) ซึ่งจากการนำเสนอต้นแบบในบทที่ 4 กำหนดให้พื้นที่ lower เป็นข้อมูลของกลุ่มใดกลุ่มหนึ่งเพียงอย่างเดียว เมื่อพิจารณาข้อมูลกลุ่มที่ 7 ได้ผลลัพธ์ดังตารางที่ 4.3 โดยในพื้นที่ lower มีค่า ข091 อยู่ในพื้นที่ lower ของกลุ่ม ง

ตารางที่ 4.2 แสดงผลการทดลองจากโปรแกรม NNC version 1.14

	Grp	X	Y	U	L	LS	US	LO	UO	Total	Certain	uncertain	
ก	1	-0.411380	0.256229	0.156472	0.156472	4	0	0	0	4	100.00	0.00	O.K.
ข1	2	-0.556615	0.426433	0.073230	0.007454	8	36	0	53	97	8.25	91.75	US>UO
ข2	3	-0.630272	0.426375	0.015316	0.015316	3	0	0	0	3	100.00	0.00	
ข3	4	-0.517096	0.422658	0.005183	0.005183	3	0	0	0	3	100.00	0.00	
ข4	5	-0.654192	0.142830	0.028268	0.028268	1	0	0	0	1	100.00	0.00	
ข5	6	-0.637535	0.405177	0.027620	0.027620	3	0	0	0	3	100.00	0.00	
ง	7	-0.688806	0.122466	0.153874	0.040508	2	2	1	3	8	25.00	62.50	
จ1	8	-0.582019	0.568932	0.054244	0.011029	6	35	0	37	78	7.69	92.31	
จ2	9	-0.525351	0.564705	0.014381	0.014381	2	0	0	0	2	100.00	0.00	
จ3	10	-0.587936	0.626602	0.015189	0.015189	3	0	0	0	3	100.00	0.00	
จ4	11	-0.608506	0.555483	0.005684	0.005684	3	0	0	0	3	100.00	0.00	
จ5	12	-0.678229	0.552867	0.011564	0.011564	1	0	0	0	1	100.00	0.00	
จ6	13	-0.690199	0.383379	0.005544	0.005544	1	0	0	0	1	100.00	0.00	
จ7	14	-0.517364	0.547296	0.010620	0.010620	2	0	0	0	2	100.00	0.00	
จ8	15	-0.637535	0.612194	0.006291	0.006291	1	0	0	0	1	100.00	0.00	
จ9	16	-0.629887	0.547121	0.002594	0.002594	1	0	0	0	1	100.00	0.00	
จ10	17	-0.517364	0.584182	0.000000	0.000000	1	0	1	0	2	50.00	0.00	
จ11	18	-0.517364	0.596017	0.000000	0.000000	1	0	1	0	2	50.00	0.00	
ฉ	19	-0.807844	0.349357	0.098763	0.035492	1	34	0	0	35	2.86	97.14	US>>UO
ช1	20	-0.566144	0.544380	0.044885	0.000000	1	49	1	0	50	2.00	96.00	US>>UO

สาเหตุ เกิดจากลำดับการเข้ามาของข้อมูลและการเฉลี่ยน้ำหนักของกลุ่ม เพราะเมื่อมีข้อมูลเข้ามาโปรแกรมจะทำการเฉลี่ยข้อมูลเลยทำให้ จุดศูนย์กลางของกลุ่มเปลี่ยน ข้อมูลกลุ่มอื่นที่เดิมอยู่ในกลุ่มนี้อาจจะไม่ได้อยู่ในกลุ่ม หรือมีข้อมูลใหม่เข้ามาในกลุ่มแทน การแก้ไข สิ้นสุดการทำงานโปรแกรมจะตรวจสอบสมาชิกในกลุ่มอีกครั้ง ว่าอยู่กลุ่มถูกต้องหรือไม่ก่อนจบโปรแกรม

ตารางที่ 4.3 แสดงการทดลองจากโปรแกรม NNC version 1.14 สำหรับกลุ่มที่ 7

กลุ่มที่ 7 กลุ่ม ง -0.688806 0.122466 member = 3 LOWER 0.040508

ง004 -0.6956620 0.0963480 distance 0.027003

ข091 -0.6541920 0.1428300 distance 0.040160 ****** ผิด

ง077 -0.6819500 0.1485840 distance 0.027003

UPPER 0.153874

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.3 การเพิ่มประสิทธิภาพของการแบ่งกลุ่มในพื้นที่ lower

โปรแกรม NNC version 1.14 มีผลการทดลองเป็นดังตารางที่ 4.2 เมื่อพิจารณาแล้วพบว่ามีบางกลุ่มที่น่าจะปรับปรุงให้การแบ่งกลุ่มมีประสิทธิภาพมากขึ้น เช่น กลุ่มที่ 2 (ข1), กลุ่ม 19(ฉ), กลุ่ม 20(ช1) พบว่ามีจำนวนสมาชิกของ US(#upper self)< UO(#upper other) หมายความว่าในพื้นที่ upper ของกลุ่ม มีจำนวนสมาชิกที่เป็นกลุ่มอื่นอยู่มากกว่าจำนวนสมาชิกของกลุ่มเอง โปรแกรม NNC จะทำการตรวจสอบค่า US และ UO ถ้าพบว่ามี US<UO จะทำการเปลี่ยนพารามิเตอร์ของกลุ่มโดยการย้ายน้ำหนักของกลุ่มใหม่เท่ากับข้อมูลสมาชิกในกลุ่มที่ใกล้น้ำหนักเฉลี่ยเดิมที่สุด และใช้ค่า lower threshold , upper threshold เท่าเดิม การลบกลุ่มนี้ทำให้เกิดการสร้างกลุ่มย่อยขึ้นอีกมากมาย ต่อมาได้ยกเลิกโปรแกรมส่วนนี้

4.3.4 การจัดกลุ่มที่มีการกระจายสูง

จากรูปที่ 5.3 ในบทที่ 5 พิจารณากลุ่ม ป โดยกลุ่ม ป มีลักษณะการกระจายของข้อมูลไม่ได้อยู่รวมกันเป็นกลุ่ม มีแนวคิดในการพิจารณาแบ่งกลุ่มออกเป็น 3 แนวทางแนวทางแรกหาจุดศูนย์กลางของข้อมูล ป ทั้งหมดแล้วกำหนดให้เป็น lower ข้อมูล ป ที่เหลือจะอยู่ในส่วนของ upper แนวคิดนี้จะใช้ประโยชน์ได้น้อยเพราะพื้นที่ upper มีขนาดใหญ่กว่า พื้นที่ lower มาก แนวทางที่สองจะแบ่งกลุ่มข้อมูล ป ออกเป็นกลุ่มย่อยๆ มีการกระจายสูงเกือบเป็นข้อมูล 1 ข้อมูลต่อ 1 กลุ่ม สำหรับประโยชน์ในการใช้งานกลุ่มดีขึ้นกว่าแนวความคิดแรกเพราะส่วน lower ของแต่ละกลุ่มมีมากขึ้นกว่าแนวความคิดแรก ส่วนแนวความคิดที่สามจะพิจารณาแบ่งกลุ่ม ป เป็น 3 กลุ่ม กลุ่มแรกอยู่ด้านซ้ายของรูปที่ 5.3 เป็นส่วนของข้อมูล ป ที่มีการซ้อนทับกันสูง กลุ่มที่สองเป็นกลุ่มที่อยู่ส่วนกลางของข้อมูล ป มีการซ้อนทับกันน้อย และกลุ่มที่สามเป็นกลุ่มที่อยู่ด้านขวาของรูปที่ 5.3 แนวความคิดที่สามนี้ดีกว่าแนวความคิดสองแบบแรกเพราะจะใช้ประโยชน์ได้มากขึ้นจากกลุ่มที่สองและกลุ่มที่สามของแนวความคิดที่สาม

สำหรับการทดลองในวิทยานิพนธ์ ทำการแบ่งกลุ่มได้ตามแนวความคิดที่สอง เพราะการปรับค่าพารามิเตอร์ของโปรแกรม NNC คือค่าโค้ดบรรทัดที่ 14, 21, 24 ในรูปที่ 4.2 รูปอัลกอริทึมการจัดข้อมูลที่มีการซ้อนทับกัน มีผลต่อจำนวนกลุ่มทั้งหมดจึงต้องปรับค่าให้เหมาะสมที่สุด ถ้าปรับค่า lower หรือ upper ให้ลดด้วยค่าพารามิเตอร์ที่น้อยเกินไปจะได้จำนวนกลุ่มมาก การแบ่งกลุ่มจะไม่มีประสิทธิภาพในการใช้งานเพราะพื้นที่ส่วน lower น้อยแต่พื้นที่ upper มาก แต่ถ้าปรับค่า lower หรือ upper ให้ลดด้วยค่าพารามิเตอร์ที่มากเกินไปจะได้จำนวนกลุ่มมาก การแบ่งกลุ่มต่างๆ จะกระจาย แต่น่าจะดีกว่าแบบแรก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
ผลการทดลอง

ในวิทยานิพนธ์นี้ได้นำต้นแบบที่ใช้สำหรับจัดกลุ่มข้อมูลที่มีการซ้อนทับกันและอัลกอริทึมในบทที่ 4 มาเขียนโปรแกรมชื่อ MNNC(Modified Neural Network Classification) ด้วยภาษา C version 1.18 บนเครื่อง Notebook Compaq Presario 1500 ใช้ CPU Pentium 4 หน่วยความจำ 256 MB โดยในวิทยานิพนธ์นี้ได้นำเสนอ 3 การทดลอง โดยใช้ข้อมูลลายมือเขียนตัวอักษรภาษาไทยที่มี 4 คุณลักษณะ ในการทดลองที่ 1 เป็นการทดสอบการทำงานของ MNNC ว่าสามารถนำมาแบ่งกลุ่มข้อมูลได้จริง และลำดับของอินพุทมีผลต่อผลการทดลอง ด้วยการนำข้อมูลกลุ่มตัวอย่าง 2 กลุ่ม จำนวน 13 ข้อมูล 2 คุณลักษณะ มาทดสอบแบ่งกลุ่มด้วยวิธี MNNC และแสดงผลในหัวข้อที่ 5.1 พบว่าการแบ่งกลุ่มด้วย MNNC ให้ลำดับของอินพุทที่ต่างกันจะให้ผลลัพธ์ต่างกัน การทดลองที่ 2 เป็นการนำ MNNC มาแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกันมากโดยใช้ข้อมูลจำนวน 354 ข้อมูล 2 คุณลักษณะ 22 กลุ่ม เป็น training data แสดงผลในหัวข้อที่ 5.2 และนำข้อมูลอีกชุดหนึ่งจำนวน 570 ข้อมูล มาเป็น test data พบว่าสามารถระบุกลุ่มให้กับข้อมูลได้ 272 ข้อมูล คิดเป็น 47.72 % ของข้อมูลทั้งหมด และการทดลองที่ 3 นำข้อมูลจำนวน 198 ข้อมูล 2 คุณลักษณะ 6 กลุ่ม มาแบ่งกลุ่มด้วยวิธี MNNC เปรียบเทียบกับวิธี NNC ในหัวข้อที่ 5.3 พบว่าการแบ่งกลุ่มด้วย MNNC ให้จำนวนกลุ่มที่น้อยกว่าคือ 6 กลุ่ม ขณะที่ NNC แบ่งกลุ่มได้ 17 กลุ่ม เมื่อต้นแบบไปใช้งาน MNNC จะมีประสิทธิภาพกว่าเพราะพื้นที่ Lower มากกว่า (หมายถึงความแน่นอนในการระบุกลุ่มมีมากกว่า)

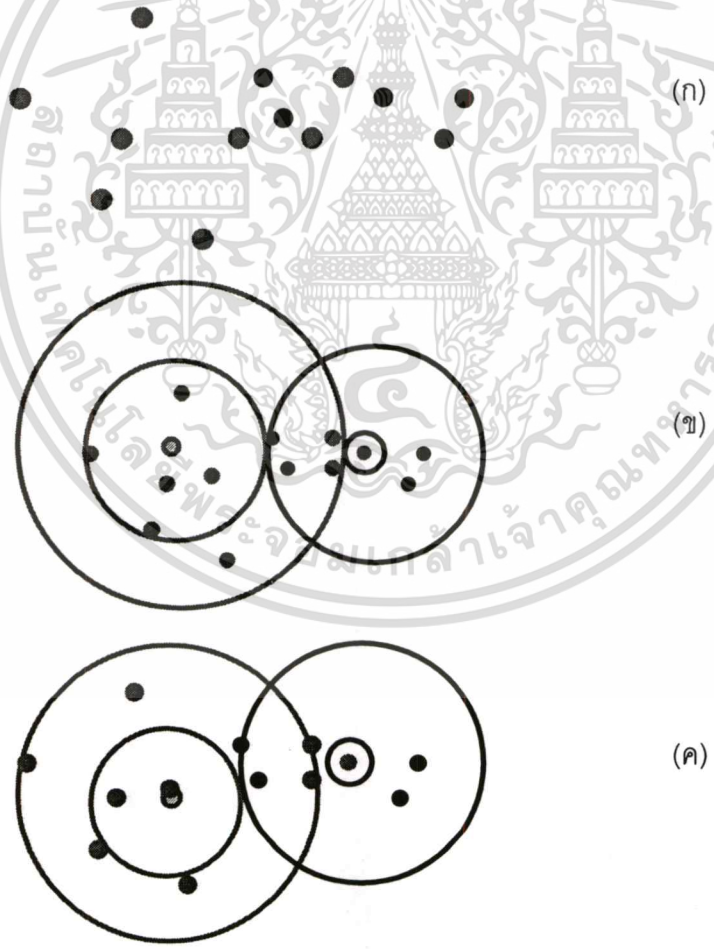
5.1 การทดลองที่ 1

ทดสอบโดยการนำข้อมูลกลุ่มตัวอย่าง 2 กลุ่ม จำนวน 13 ข้อมูล 2 คุณลักษณะดังตารางที่ 5.1 สามารถแสดงได้ดังรูปที่ 5.1 (ก) มาทดสอบแบ่งกลุ่มด้วยวิธี MNNC ได้ดังรูปที่ 5.1 (ข) จากนั้นนำข้อมูลในตารางที่ 5.1 มาจัดลำดับในการจัดกลุ่มใหม่เป็นตารางที่ 5.2 แล้วนำมาทดสอบแบ่งกลุ่มด้วยวิธี MNNC อีกครั้งได้ดังรูปที่ 5.1 (ค) และพบว่าสามารถจัดกลุ่มได้ดีกว่ารูปที่ 5.1 (ข) แสดงว่าลำดับที่ใช้เป็นอินพุทในการจัดกลุ่มแบบ MNNC มีผลต่อการจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 แสดงตัวอย่างข้อมูลกลุ่มตัวอย่าง 2 กลุ่ม จำนวน 13 ข้อมูล 2 คุณลักษณะ

กลุ่ม	x	y
R001	9.19	11.76
R002	10.61	11.76
R003	9.66	8.91
R004	14.41	11.29
R005	14.41	10.34
R006	11.09	14.14
R007	8.71	13.18
R008	6.82	10.81
B001	17.26	10.84
B002	15.36	10.81
B003	16.78	11.79
B004	15.51	10.34
B005	12.99	11.29



รูปที่ 5.1 (ก) แสดงข้อมูลก่อนการจัดกลุ่ม (ข) การจัดกลุ่มด้วยวิธี MNNC มีอินพุทตามตารางที่ 5.1 (ค) การจัดกลุ่มด้วยวิธี MNNC มีอินพุทตามตารางที่ 5.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.2 แสดงข้อมูลเหมือนตารางที่ 5.1 แต่มีการเปลี่ยนลำดับ

กลุ่ม	x	y
R005	14.41	10.34
R008	6.82	10.81
B004	15.51	10.34
R001	9.19	11.76
R002	10.61	11.76
R003	9.66	8.91
R004	14.41	11.29
R006	11.09	14.14
R007	8.71	13.18
B001	17.26	10.84
B002	15.36	10.81
B003	16.78	11.79
B005	12.99	11.29

5.2 การทดลองที่ 2

นำข้อมูลลายมือเขียนตัวอักษรภาษาไทยที่มี 4 จำนวน 354 ข้อมูล 2 คุณลักษณะ 22 กลุ่ม ตามตารางที่ 5.1 แต่ใช้ 2 คุณลักษณะแรก มาแสดงข้อมูลในรูปที่ 5.3 และผลการทดลองดังรูปที่ 5.4 ตามลำดับ ผลการทดลองพบว่า กลุ่ม ก ด ช ฒ ว2 ง1 ง2 ง3 มี lower threshold เท่ากับ upper threshold และกลุ่มตัวอักษร ง มี 3 กลุ่ม กลุ่มตัวอักษร ว มี 2 กลุ่ม ส่วนข้อมูลกลุ่มอื่นๆ นอกจากนี้ เมื่อพิจารณาในรูปที่ 5.4 พบว่ามีการซ้อนทับกันของข้อมูลมากทำให้การแบ่งกลุ่มที่กระจายเป็นกลุ่มย่อย โดยจำนวนข้อมูลทั้งหมดมี 121 กลุ่ม ซึ่งกลุ่มที่ไม่ได้แสดงให้เห็นในรูปที่ 4.8 คือข้อมูลที่แบ่งกระจัดกระจายเป็นกลุ่มย่อยสามารถแสดงผลการทดลองได้ในตารางที่ 5.3

จากตารางที่ 5.3 no .คือจำนวนกลุ่มที่งานวิจัยนี้สามารถแบ่งได้โดยสามารถแบ่งได้ทั้งหมดจำนวน 121 กลุ่ม, ch คือชื่อของกลุ่ม ค่า X, Y คือค่าน้ำหนักเฉลี่ยของกลุ่ม, Max คือ upper threshold (ค่าระยะทางจากจุดศูนย์กลางหรือน้ำหนักเฉลี่ยของกลุ่มถึงค่า upper threshold), Min คือค่า lower threshold (ค่าระยะทางจากจุดศูนย์กลางถึงค่า lower threshold), TOT คือค่าจำนวนข้อมูลทั้งหมดของกลุ่มนั้น, LS (lower self) คือจำนวนข้อมูลที่ตกในกลุ่มของตัวเองภายในพื้นที่ lower US (upper self) คือจำนวนข้อมูลที่ตกในกลุ่มของตัวเองภายในพื้นที่ upper, LO(lower other) คือจำนวนข้อมูลอื่นๆ ตกในพื้นที่ lower ของกลุ่ม, UO (upper other) คือจำนวนข้อมูลอื่นๆ ตกในพื้นที่ upper ส่วนค่า certain และ uncertain คือค่าที่ใช้วัดประสิทธิภาพของกลุ่มมีหน่วยเป็นเปอร์เซ็นต์ สามารถคำนวณได้ตามสมการที่ (5.1) และ (5.2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$certain = (LS/(LS+LO+US+UO))*100 \quad (5.1)$$

$$uncertain = ((US+UO)/(LS+LO+US+UO))*100 \quad (5.2)$$

ตารางที่ 5.3 แสดงการทดลองจัดกลุ่มข้อมูลของลายมือเขียนตัวอักษรภาษาไทยจำนวน 354 ข้อมูล มี 2 คุณลักษณะจำนวน 22 กลุ่ม โดยโปรแกรม NNC.C version 1.18 (เป็นโปรแกรมที่ได้นำเสนออัลกอริทึมในวิทยานิพนธ์นี้) สามารถแบ่งกลุ่มได้ 121 กลุ่ม

no.	ch	index0	X	Y	MAX	MIN	LS	US	LO	UO	Certain	Uncertain
1	ก	1	-0.411380	0.256229	0.156472	0.156472	4	0	0	0	100.00	0.00
2	ข1	2	-0.557989	0.425899	0.007608	0.007608	8	0	0	0	100.00	0.00
3	ข8	94	-0.508909	0.393344	0.032912	0.032912	6	0	0	0	100.00	0.00
4	ข9	95	-0.580620	0.421940	0.013462	0.013462	6	0	0	0	100.00	0.00
5	ข11	109	-0.549512	0.416264	0.013022	0.013022	6	0	0	0	100.00	0.00
6	ข6	63	-0.636994	0.419134	0.026029	0.026029	4	0	0	0	100.00	0.00
7	ข4	20	-0.654192	0.142830	0.028268	0.028268	1	0	0	0	100.00	0.00
8	ข7	64	-0.578790	0.400331	0.001781	0.001781	1	0	0	0	100.00	0.00
9	ข10	96	-0.552867	0.476380	0.003214	0.003214	1	0	0	0	100.00	0.00
10	ข12	110	-0.519655	0.443995	0.003827	0.003827	1	0	0	0	100.00	0.00
11	ข14	121	-0.591505	0.440945	0.013789	0.013789	1	0	0	0	100.00	0.00
12	ง1	3	-0.688806	0.122466	0.040080	0.040080	2	0	0	0	100.00	0.00
13	ง2	53	-0.624342	0.240138	0.089068	0.089068	1	0	0	0	100.00	0.00
14	ง3	65	-0.578065	0.131036	0.077025	0.077025	1	0	0	0	100.00	0.00
15	จ4	4	-0.582296	0.568035	0.010595	0.010595	5	0	0	0	100.00	0.00
16	จ12	68	-0.555559	0.602255	0.023279	0.023279	4	0	0	0	100.00	0.00
17	จ18	114	-0.579331	0.582691	0.007665	0.007665	4	0	0	0	100.00	0.00
18	จ6	22	-0.587944	0.626615	0.015189	0.015189	3	0	0	0	100.00	0.00
19	จ13	69	-0.612403	0.584182	0.007295	0.007295	2	0	0	0	100.00	0.00
20	จ16	104	-0.582336	0.542466	0.006247	0.006247	2	0	0	0	100.00	0.00
21	จ19	115	-0.584182	0.552867	0.002279	0.002279	2	0	0	0	100.00	0.00
22	จ8	24	-0.678229	0.552867	0.011565	0.011565	1	0	0	0	100.00	0.00
23	จ9	25	-0.690199	0.383379	0.005544	0.005544	1	0	0	0	100.00	0.00
24	จ10	66	-0.584182	0.533286	0.000790	0.000790	1	0	0	0	100.00	0.00
25	จ11	67	-0.637535	0.612194	0.006291	0.006291	1	0	0	0	100.00	0.00
26	จ14	70	-0.629887	0.547121	0.002594	0.002594	1	0	0	0	100.00	0.00
27	จ15	97	-0.622869	0.584182	0.001924	0.001924	1	0	0	0	100.00	0.00
28	ช2	54	-0.572157	0.681950	0.040606	0.040606	3	0	0	0	100.00	0.00
29	ช5	72	-0.517364	0.669594	0.058320	0.058320	3	0	0	0	100.00	0.00
30	ช3	55	-0.552867	0.517364	0.019246	0.019246	2	0	0	0	100.00	0.00
31	ช6	6	-0.552867	0.760297	0.079454	0.079454	2	0	0	0	100.00	0.00
32	ช1	26	-0.637535	0.517364	0.024468	0.024468	1	0	0	0	100.00	0.00
33	ช4	71	-0.584182	0.517364	0.015112	0.015112	1	0	0	0	100.00	0.00
34	ฌ	7	-0.849690	0.376056	0.044649	0.044649	5	0	0	0	100.00	0.00
35	ด	8	-0.754115	0.205668	0.079948	0.079948	2	0	0	0	100.00	0.00
36	ท	119	-0.601073	0.326123	0.020912	0.020912	1	0	0	0	100.00	0.00
37	น	116	-0.787137	0.132847	0.069500	0.069500	1	0	0	0	100.00	0.00
38	บ1	9	-0.488689	0.427806	0.015776	0.015776	6	0	0	0	100.00	0.00
39	บ13	93	-0.416541	0.466979	0.060985	0.060985	6	0	0	0	100.00	0.00
40	บ4	73	-0.542270	0.476011	0.010530	0.010530	2	0	0	0	100.00	0.00

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.3 (ต่อ)

no.	ch	index0	X	Y	MAX	MIN	LS	US	LO	UO	Certain	Uncertain
41	บ5	74	-0.601073	0.454525	0.016532	0.016532	2	0	0	0	100.00	0.00
42	บ9	78	-0.568970	0.457577	0.009100	0.009100	2	0	0	0	100.00	0.00
43	บ15	99	-0.538891	0.427905	0.010430	0.010430	2	0	0	0	100.00	0.00
44	บ19	107	-0.538891	0.457577	0.014391	0.014391	2	0	0	0	100.00	0.00
45	บ2	27	-0.123350	0.457577	0.033150	0.033150	1	0	0	0	100.00	0.00
46	บ7	76	-0.554872	0.443995	0.003663	0.003663	1	0	0	0	100.00	0.00
47	บ8	77	-0.566144	0.427905	0.004973	0.004973	1	0	0	0	100.00	0.00
48	บ10	79	-0.644969	0.443995	0.017644	0.017644	1	0	0	0	100.00	0.00
49	บ11	80	-0.570956	0.440945	0.009168	0.009168	1	0	0	0	100.00	0.00
50	บ12	92	-0.292091	0.457577	0.027015	0.027015	1	0	0	0	100.00	0.00
51	บ14	98	-0.535157	0.435745	0.004786	0.004786	1	0	0	0	100.00	0.00
52	บ16	100	-0.499939	0.476380	0.009247	0.009247	1	0	0	0	100.00	0.00
53	บ17	105	-0.557051	0.434569	0.001842	0.001842	1	0	0	0	100.00	0.00
54	บ18	106	-0.535444	0.443995	0.005113	0.005113	1	0	0	0	100.00	0.00
55	บ20	120	-0.515053	0.427905	0.002231	0.002231	1	0	0	0	100.00	0.00
56	ป2	57	-0.217701	0.458301	0.063764	0.063764	3	0	0	0	100.00	0.00
57	ป3	58	-0.091288	0.448843	0.033220	0.033220	2	0	0	0	100.00	0.00
58	ป5	82	-0.555836	0.481971	0.006250	0.006250	2	0	0	0	100.00	0.00
59	ป8	85	-0.577989	0.476380	0.019004	0.019004	2	0	0	0	100.00	0.00
60	ป11	88	-0.319802	0.450931	0.021041	0.021041	2	0	0	0	100.00	0.00
61	ป6	83	-0.528955	0.431108	0.003244	0.003244	1	0	0	0	100.00	0.00
62	ป7	84	-0.532486	0.439812	0.002847	0.002847	1	0	0	0	100.00	0.00
63	ป9	86	-0.566144	0.448843	0.006923	0.006923	1	0	0	0	100.00	0.00
64	ป10	87	-0.550508	0.437343	0.005778	0.005778	1	0	0	0	100.00	0.00
65	ป12	89	-0.524832	0.454456	0.006245	0.006245	1	0	0	0	100.00	0.00
66	ป13	108	-0.514667	0.454098	0.003496	0.003496	1	0	0	0	100.00	0.00
67	ป14	113	-0.517364	0.457577	0.003124	0.003124	1	0	0	0	100.00	0.00
68	พ2	59	-0.737155	0.417364	0.016221	0.016221	2	0	0	0	100.00	0.00
69	พ1	12	-0.746788	0.405177	0.006941	0.006941	1	0	0	0	100.00	0.00
70	ม5	31	-0.727348	0.376139	0.013394	0.013394	13	0	0	0	100.00	0.00
71	ม4	30	-0.579396	0.374005	0.026253	0.026253	8	0	0	0	100.00	0.00
72	ม7	33	-0.622810	0.369141	0.037503	0.037503	7	0	0	0	100.00	0.00
73	ม3	29	-0.586846	0.392416	0.001711	0.001711	1	0	0	0	100.00	0.00
74	ม6	32	-0.612194	0.427905	0.005812	0.005812	1	0	0	0	100.00	0.00
75	ม10	101	-0.568027	0.405776	0.012052	0.012052	1	0	0	0	100.00	0.00
76	ม11	102	-0.596017	0.427905	0.007940	0.007940	1	0	0	0	100.00	0.00
77	ล2	35	-0.458766	0.565686	0.051573	0.051573	5	0	0	0	100.00	0.00
78	ล5	38	-0.728104	0.559494	0.041757	0.041757	4	0	0	0	100.00	0.00
79	ล8	41	-0.785014	0.519552	0.095423	0.095423	4	0	0	0	100.00	0.00
80	ล4	37	-0.587219	0.548093	0.004756	0.004756	3	0	0	0	100.00	0.00
81	ล11	44	-0.637535	0.637535	0.020661	0.020661	2	0	0	0	100.00	0.00
82	ล17	103	-0.575827	0.550369	0.008640	0.008640	2	0	0	0	100.00	0.00
83	ล3	36	-0.541354	0.566144	0.002946	0.002946	1	0	0	0	100.00	0.00
84	ล6	39	-0.517364	0.536596	0.007704	0.007704	1	0	0	0	100.00	0.00
85	ล10	43	-0.627818	0.621098	0.005232	0.005232	1	0	0	0	100.00	0.00
86	ล13	46	-0.561091	0.534775	0.010773	0.010773	1	0	0	0	100.00	0.00
87	ล14	47	-0.571796	0.584182	0.005742	0.005742	1	0	0	0	100.00	0.00
88	ล15	48	-0.612194	0.637535	0.018060	0.018060	1	0	0	0	100.00	0.00
89	ล16	91	-0.622869	0.582248	0.001854	0.001854	1	0	0	0	100.00	0.00
90	ว3	50	-0.091288	0.589466	0.104609	0.104609	2	0	0	0	100.00	0.00

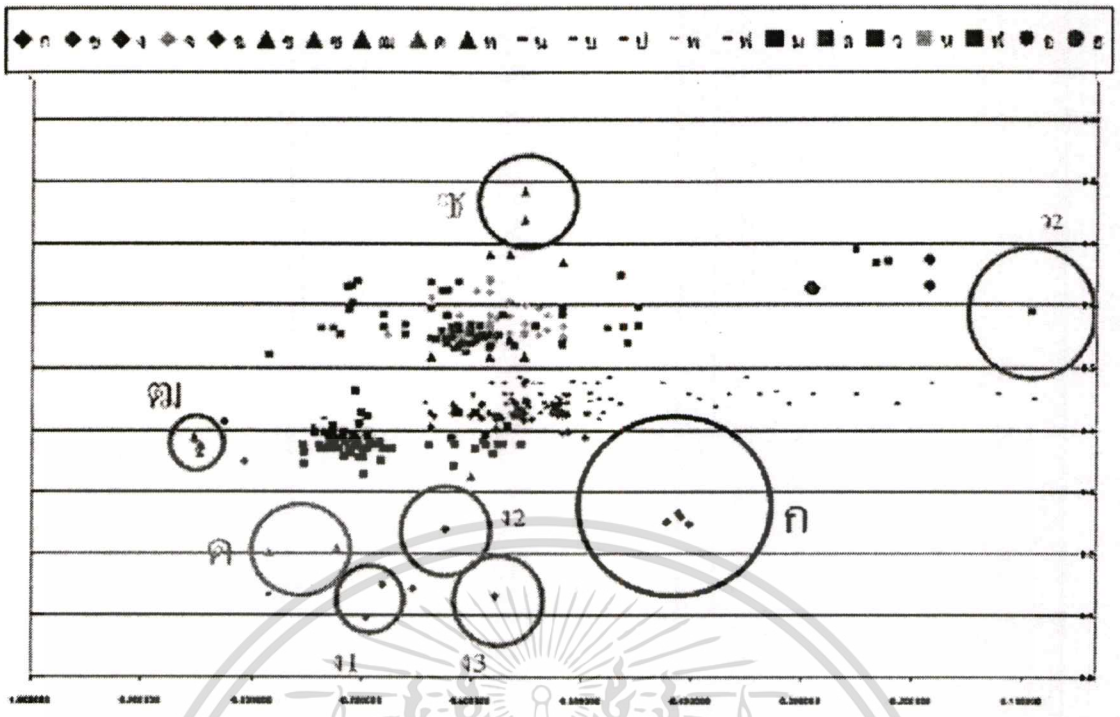
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.3 (ต่อ)

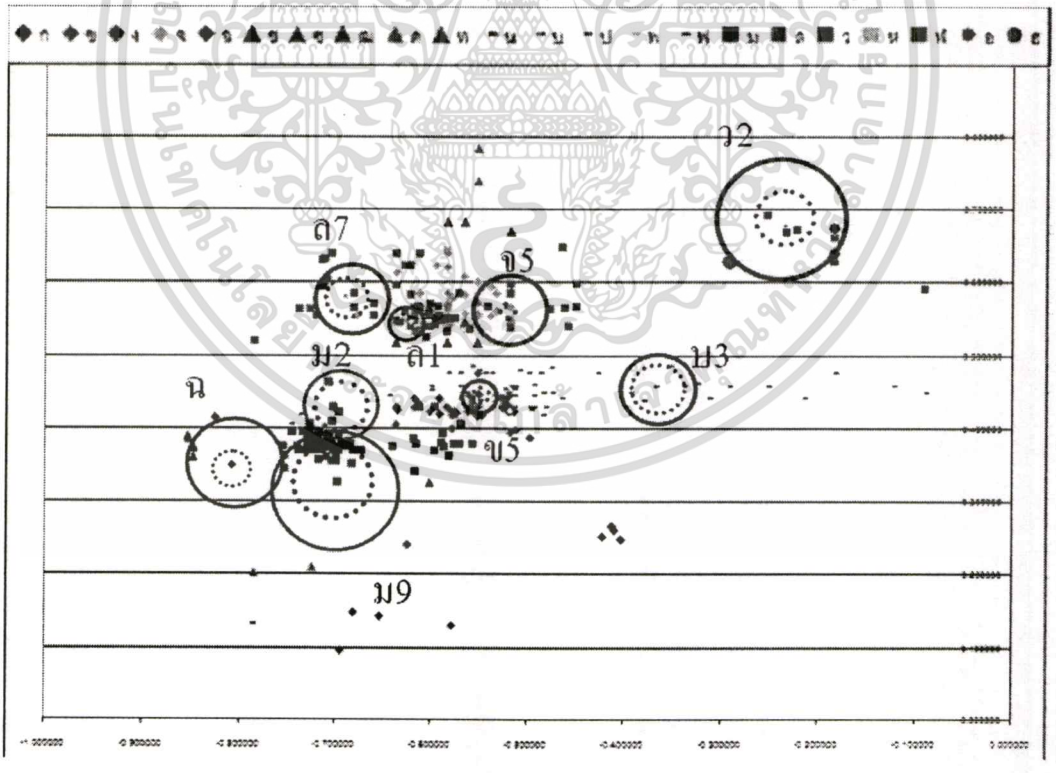
no.	ch	index0	X	Y	MAX	MIN	LS	US	LO	UO	Certain	Uncertain
91	ห	117	-0.584182	0.641334	0.003719	0.003719	1	0	0	0	100.00	0.00
92	ห4	60	-0.710089	0.633123	0.030028	0.030028	3	0	0	0	100.00	0.00
93	ห1	16	-0.622869	0.622869	0.004557	0.004557	1	0	0	0	100.00	0.00
94	ห2	51	-0.584182	0.532486	0.000720	0.000720	1	0	0	0	100.00	0.00
95	ห3	52	-0.463747	0.646625	0.052678	0.052678	1	0	0	0	100.00	0.00
96	อ	17	-0.184289	0.637535	0.005104	0.005104	1	0	0	0	100.00	0.00
97	อ	61	-0.184289	0.673436	0.003832	0.003832	1	0	0	0	100.00	0.00
98	ฮ	118	-0.292091	0.627444	0.070334	0.070334	1	0	0	0	100.00	0.00
99	ม9	90	-0.698200	0.326123	0.057802	0.053968	35	5	0	1	85.37	14.63
100	บ3	56	-0.366938	0.457129	0.039649	0.035902	5	1	0	0	83.33	16.67
101	ป1	10	-0.490925	0.450884	0.022063	0.011572	3	1	0	0	75.00	25.00
102	ล12	45	-0.598891	0.565627	0.016563	0.009452	3	1	0	0	75.00	25.00
103	จ	5	-0.807844	0.349357	0.066899	0.035493	1	1	0	0	50.00	50.00
104	พ	11	-0.730259	0.399455	0.027330	0.005764	1	1	0	0	50.00	50.00
105	ว2	49	-0.236786	0.676369	0.059299	0.052510	3	3	0	1	42.86	57.14
106	จ7	23	-0.608499	0.555494	0.012766	0.005684	3	5	0	0	37.50	62.50
107	ม8	34	-0.753769	0.360114	0.054831	0.039275	17	28	0	4	34.69	65.31
108	ข5	62	-0.551410	0.446887	0.012823	0.004463	2	3	0	2	28.57	71.43
109	ล9	42	-0.609584	0.533205	0.024889	0.012931	6	11	0	5	27.27	72.73
110	ล7	40	-0.691612	0.580099	0.059003	0.025546	3	7	0	3	23.08	76.92
111	บ6	75	-0.514230	0.457577	0.020042	0.000954	1	4	0	1	16.67	83.33
112	ม1	13	-0.696998	0.368618	0.054518	0.005544	9	48	0	3	15.00	85.00
113	ข3	19	-0.517680	0.423243	0.018519	0.003657	2	9	0	4	13.33	86.67
114	ข2	18	-0.617267	0.427148	0.029167	0.003722	1	6	0	2	11.11	88.89
115	ป4	81	-0.487723	0.476714	0.034577	0.009177	1	6	0	2	11.11	88.89
116	ม2	28	-0.701975	0.430558	0.068187	0.037557	6	44	0	9	10.17	89.83
117	จ5	21	-0.525337	0.564668	0.046330	0.014381	2	16	0	5	8.70	91.30
118	ล1	14	-0.621711	0.546793	0.033005	0.002524	2	27	0	9	5.26	94.74
119	ข13	111	-0.526743	0.440945	0.000010	0.000000	0	1	0	0	0.00	100.00
120	จ17	112	-0.566144	0.544380	0.000010	0.000000	0	1	0	0	0.00	100.00
121	ว1	15	-0.184289	0.644521	0.065749	0.003762	0	6	0	2	0.00	100.00
					TOT=354		323	235	0	53		

จากตารางที่ 5.3 แสดงผลตามลำดับกลุ่มที่มี Certain มากไปหาน้อยเป็นคีย์หลัก (primary key) และ LS เป็นคีย์รอง (secondary key) พบว่าโปรแกรมทำงานวนลูปซ้ำกันเป็นจำนวน 29 รอบ กลุ่มที่มี Certain = 100% จำนวน 98 กลุ่ม 50-80 % จำนวน 6 กลุ่ม และมี Certain น้อยกว่า 50 % 17 กลุ่ม จะเห็นว่าข้อมูลที่มีการซ้อนทับกันสูง มีโอกาสของการแบ่งกลุ่มกระจายออกเป็นกลุ่มย่อยๆ มากเนื่องจากข้อมูลมีการซ้อนทับกันการแบ่งกลุ่มจึงเห็นกลุ่มที่มีสมาชิกหนึ่งตัวต่อหนึ่งกลุ่มสูงจำนวน 53 กลุ่ม (จาก Certain = 100% จำนวน 98 กลุ่ม) สามารถแบ่งข้อมูลทั้งหมด 354 ข้อมูล จัดอยู่ใน Lower ของกลุ่มได้ 323 ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 แสดงผลการจัดกลุ่มด้วยวิธี MNNC เฉพาะกลุ่มที่มีค่า Cetatin = 100 %บางส่วน



รูปที่ 5.3 แสดงผลการจัดกลุ่มด้วยวิธี MNNC เฉพาะกลุ่มที่มีค่า Cetatin < 100 %บางส่วน

จากตารางที่ 5.3 จะได้กลุ่มข้อมูลต้นแบบ 121 กลุ่ม จากนั้นนำข้อมูลอีกชุดหนึ่ง data011.txt ที่มีจำนวน 570 ข้อมูลมาทำการทดสอบ ได้ผลลัพธ์ดังแสดงในตารางที่ 5.4 พบว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มต้นแบบสามารถจัดกลุ่มข้อมูลได้ 272 (ผลรวมจำนวนข้อมูลใน LS ของ data011.txt ในตารางที่ 5.4)ข้อมูลคิดเป็น 47.72 % ของข้อมูลทั้งหมด

ตารางที่ 5.4 แสดงผลการทดลองนำข้อมูลชุดที่ 2 (test data) มาทดสอบกับต้นแบบที่ได้ในการทดลองที่ 2 (ตารางที่ 5.3)

		data012_4.txt (train data)						data011.txt (test data)					
no.	ch	LS	US	LO	UO	Certain	Uncertain	LS	US	LO	UO	Certain	Uncertain
1	ก	4	0	0	0	100.00	0.00	7	0	5	0	58.33	0.00
2	ข1	8	0	0	0	100.00	0.00	5	0	1	0	83.33	0.00
3	ข8	6	0	0	0	100.00	0.00	6	0	5	0	54.55	0.00
4	ข9	6	0	0	0	100.00	0.00	6	0	2	0	75.00	0.00
5	ข11	6	0	0	0	100.00	0.00	6	0	1	0	85.71	0.00
6	ข6	4	0	0	0	100.00	0.00	4	0	1	0	80.00	0.00
7	ข4	1	0	0	0	100.00	0.00	1	0	2	0	33.33	0.00
8	ข7	1	0	0	0	100.00	0.00	2	0	0	0	100.00	0.00
9	ข10	1	0	0	0	100.00	0.00	6	0	1	0	85.71	0.00
10	ข12	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
11	ข14	1	0	0	0	100.00	0.00	2	0	0	0	100.00	0.00
12	ง1	2	0	0	0	100.00	0.00	4	0	3	0	57.14	0.00
13	ง2	1	0	0	0	100.00	0.00	8	0	22	0	26.67	0.00
14	ง3	1	0	0	0	100.00	0.00	2	0	5	0	28.57	0.00
15	จ4	5	0	0	0	100.00	0.00	3	0	1	0	75.00	0.00
16	จ12	4	0	0	0	100.00	0.00	7	0	4	0	63.64	0.00
17	จ18	4	0	0	0	100.00	0.00	3	0	1	0	75.00	0.00
18	จ6	3	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
19	จ13	2	0	0	0	100.00	0.00	3	0	2	0	60.00	0.00
20	จ16	2	0	0	0	100.00	0.00	1	0	1	0	50.00	0.00
21	จ19	2	0	0	0	100.00	0.00	0	0	1	0	0.00	0.00
22	จ8	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
23	จ9	1	0	0	0	100.00	0.00	0	0	0	0	-	-
24	จ10	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
25	จ11	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
26	จ14	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
27	จ15	1	0	0	0	100.00	0.00	0	0	0	0	-	-
28	ช2	3	0	0	0	100.00	0.00	2	0	2	0	50.00	0.00
29	ช5	3	0	0	0	100.00	0.00	2	0	9	0	18.18	0.00
30	ช3	2	0	0	0	100.00	0.00	1	0	3	0	25.00	0.00
31	ช6	2	0	0	0	100.00	0.00	2	0	7	0	22.22	0.00
32	ช1	1	0	0	0	100.00	0.00	0	0	3	0	0.00	0.00
33	ช4	1	0	0	0	100.00	0.00	0	0	3	0	0.00	0.00
34	ฌ	5	0	0	0	100.00	0.00	3	0	2	0	60.00	0.00
35	ด	2	0	0	0	100.00	0.00	2	0	9	0	18.18	0.00
36	ท	1	0	0	0	100.00	0.00	0	0	0	0	-	-
37	น	1	0	0	0	100.00	0.00	1	0	5	0	16.67	0.00
38	บ1	6	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
39	บ13	6	0	0	0	100.00	0.00	5	0	4	0	55.56	0.00
40	บ4	2	0	0	0	100.00	0.00	2	0	2	0	50.00	0.00
41	บ5	2	0	0	0	100.00	0.00	1	0	3	0	25.00	0.00
42	บ9	2	0	0	0	100.00	0.00	3	0	0	0	100.00	0.00
43	บ15	2	0	0	0	100.00	0.00	2	0	0	0	100.00	0.00

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.4 (ต่อ)

no.	ch	data012_4.txt (train data)						data011.txt (test data)					
		LS	US	LO	UO	Certain	Uncertain	LS	US	LO	UO	Certain	Uncertain
44	บ19	2	0	0	0	100.00	0.00	0	0	3	0	0.00	0.00
45	บ2	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
46	บ7	1	0	0	0	100.00	0.00	0	0	0	0	-	-
47	บ8	1	0	0	0	100.00	0.00	0	0	2	0	0.00	0.00
48	บ10	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
49	บ11	1	0	0	0	100.00	0.00	1	0	1	0	50.00	0.00
50	บ12	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
51	บ14	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
52	บ16	1	0	0	0	100.00	0.00	0	0	0	0	-	-
53	บ17	1	0	0	0	100.00	0.00	1	0	1	0	50.00	0.00
54	บ18	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
55	บ20	1	0	0	0	100.00	0.00	0	0	0	0	-	-
56	ป2	3	0	0	0	100.00	0.00	2	0	0	0	100.00	0.00
57	ป3	2	0	0	0	100.00	0.00	0	0	0	0	-	-
58	ป5	2	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
59	ป8	2	0	0	0	100.00	0.00	2	0	4	0	33.33	0.00
60	ป11	2	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
61	ป6	1	0	0	0	100.00	0.00	0	0	0	0	-	-
62	ป7	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
63	ป9	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
64	ป10	1	0	0	0	100.00	0.00	3	0	0	0	100.00	0.00
65	ป12	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
66	ป13	1	0	0	0	100.00	0.00	0	0	0	0	-	-
67	ป14	1	0	0	0	100.00	0.00	0	0	0	0	-	-
68	ฟ2	2	0	0	0	100.00	0.00	0	0	2	0	0.00	0.00
69	ฟ1	1	0	0	0	100.00	0.00	0	0	0	0	-	-
70	ม5	13	0	0	0	100.00	0.00	9	0	0	0	100.00	0.00
71	ม4	8	0	0	0	100.00	0.00	5	0	0	0	100.00	0.00
72	ม7	7	0	0	0	100.00	0.00	4	0	1	0	80.00	0.00
73	ม3	1	0	0	0	100.00	0.00	0	0	0	0	-	-
74	ม6	1	0	0	0	100.00	0.00	0	0	1	0	0.00	0.00
75	ม10	1	0	0	0	100.00	0.00	1	0	3	0	25.00	0.00
76	ม11	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
77	ล2	5	0	0	0	100.00	0.00	7	0	11	0	38.89	0.00
78	ล5	4	0	0	0	100.00	0.00	2	0	2	0	50.00	0.00
79	ล8	4	0	0	0	100.00	0.00	4	0	26	0	13.33	0.00
80	ล4	3	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
81	ล11	2	0	0	0	100.00	0.00	2	0	0	0	100.00	0.00
82	ล17	2	0	0	0	100.00	0.00	1	0	1	0	50.00	0.00
83	ล3	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
84	ล6	1	0	0	0	100.00	0.00	0	0	0	0	-	-
85	ล10	1	0	0	0	100.00	0.00	0	0	0	0	-	-
86	ล13	1	0	0	0	100.00	0.00	0	0	0	0	-	-
87	ล14	1	0	0	0	100.00	0.00	0	0	0	0	-	-
88	ล15	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
89	ล16	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
90	ว3	2	0	0	0	100.00	0.00	2	0	5	0	28.57	0.00

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.4 (ต่อ)

		data012_4.txt (train data)						data011.txt (test data)					
no.	ch	LS	US	LO	UO	Certain	Uncertain	LS	US	LO	UO	Certain	Uncertain
91	ห	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
92	ห4	3	0	0	0	100.00	0.00	1	0	1	0	50.00	0.00
93	ห1	1	0	0	0	100.00	0.00	0	0	0	0	-	-
94	ห2	1	0	0	0	100.00	0.00	1	0	0	0	100.00	0.00
95	ห3	1	0	0	0	100.00	0.00	0	0	19	0	0.00	0.00
96	อ	1	0	0	0	100.00	0.00	0	0	0	0	-	-
97	อ	1	0	0	0	100.00	0.00	0	0	0	0	-	-
98	ฮ	1	0	0	0	100.00	0.00	2	0	3	0	40.00	0.00
99	ม9	35	5	0	1	85.37	14.63	37	4	11	2	68.52	11.11
100	บ3	5	1	0	0	83.33	16.67	2	0	0	1	66.67	33.33
101	ป1	3	1	0	0	75.00	25.00	1	1	1	0	33.33	33.33
102	ล12	3	1	0	0	75.00	25.00	1	5	2	8	6.25	81.25
103	จ	1	1	0	0	50.00	50.00	1	1	2	13	5.88	82.35
104	พ	1	1	0	0	50.00	50.00	0	2	0	10	0.00	100.00
105	ว2	3	3	0	1	42.86	57.14	3	1	1	1	50.00	33.33
106	จ7	3	5	0	0	37.50	62.50	2	2	0	6	20.00	80.00
107	ม8	17	28	0	4	34.69	65.31	24	18	3	7	46.15	48.08
108	ข5	2	3	0	2	28.57	71.43	2	1	0	2	40.00	60.00
109	ล9	6	11	0	5	27.27	72.73	3	3	1	6	23.08	69.23
110	ล7	3	7	0	3	23.08	76.92	3	7	1	8	15.79	78.95
111	บ6	1	4	0	1	16.67	83.33	0	5	0	6	0.00	100.00
112	ม1	9	48	0	3	15.00	85.00	5	41	0	6	9.62	90.38
113	ข3	2	9	0	4	13.33	86.67	0	8	1	6	0.00	93.33
114	ข2	1	6	0	2	11.11	88.89	1	2	0	5	12.50	87.50
115	ป4	1	6	0	2	11.11	88.89	1	4	0	3	12.50	87.50
116	ม2	6	44	0	9	10.17	89.83	7	29	3	17	12.50	82.14
117	จ5	2	16	0	5	8.70	91.30	3	14	0	13	10.00	90.00
118	ล1	2	27	0	9	5.26	94.74	0	13	0	15	0.00	100.00
119	ข13	0	1	0	0	0.00	100.00	0	0	0	0	-	-
120	จ17	0	1	0	0	0.00	100.00	1	0	1	0	50.00	0.00
121	ว1	0	6	0	2	0.00	100.00	0	4	0	5	0.00	100.00
TOT=354		323	235	0	53			272	165	223	140	TOT=570	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 การทดลองที่ 3

เราได้ทดสอบต้นแบบการจัดกลุ่มที่ได้นำเสนอในบทที่ 4 กับข้อมูลลายมือเขียนตัวอักษรภาษาไทย 6 กลุ่ม คือ ก ส ฟ บ ม จ โดยมี 198 ข้อมูล ข้อมูลแต่ละตัวมี 2 คุณลักษณะ สามารถแสดงข้อมูลที่ใช้ในกระบวนการเรียนรู้ได้ดังตารางที่ 5.5 และแสดงข้อมูลในรูปที่ 5.5 โดยอัลกอริทึมจะทำการจัดข้อมูลตามลำดับข้อมูลที่เข้ามา ผลการทดลองสามารถจัดกลุ่มข้อมูลด้วยวิธีการจัดกลุ่มที่ได้นำเสนอในวิทยานิพนธ์นี้ซึ่งจะสามารถแบ่งกลุ่มได้ 6 กลุ่ม ดังแสดงได้ในตารางที่ 5.6 และนำผลลัพธ์ของตารางมาแสดงดังรูปที่ 5.6 จะพบว่าเมื่อเปรียบเทียบกับการจัดกลุ่มด้วยวิธี NNC แสดงดังรูปที่ 5.5 พบว่าผลการทดลองที่นำเสนอ (MNNC) สามารถจัดกลุ่มได้จำนวนกลุ่มที่เหมาะสมกว่าเมื่อเทียบกับการแบ่งกลุ่มด้วยสายตาของมนุษย์

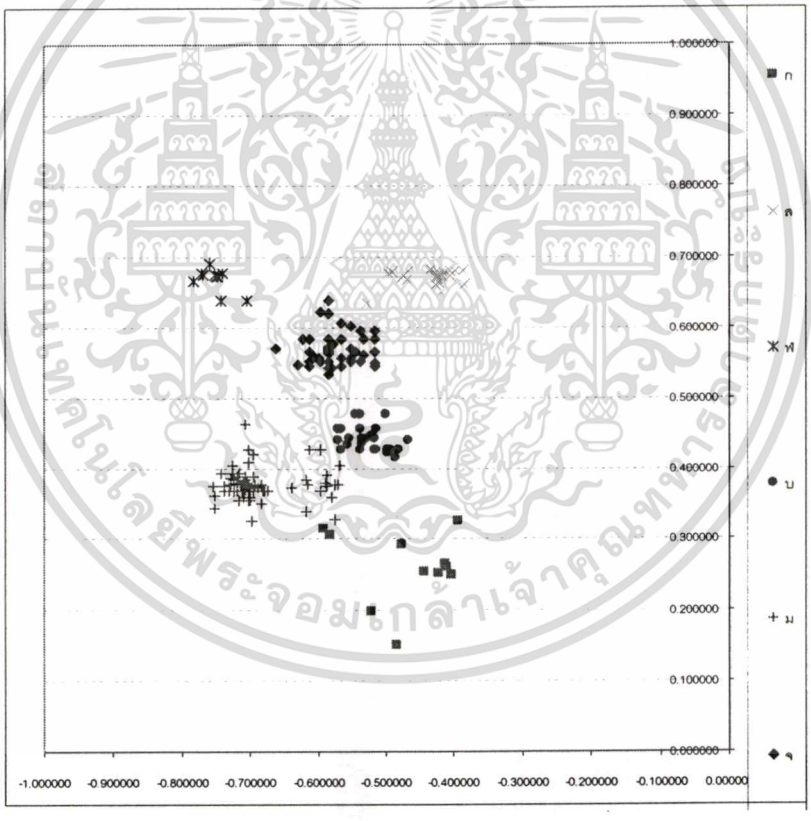
ตารางที่ 5.5 แสดงตัวอย่างข้อมูลพยัญชนะที่ใช้ในการทดสอบ

กลุ่ม	feature1	feature2
ก013	-0.402433	0.248172
ก021	-0.409239	0.260210
ก010	-0.409239	0.260210
ก048	-0.412089	0.265182
ส001	-0.528631	0.633205
ส002	-0.418121	0.649524
ส003	-0.469269	0.668075
...		
ฟ120	-0.748883	0.672864
บ017	-0.535157	0.435745
ม040	-0.727160	0.404487

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

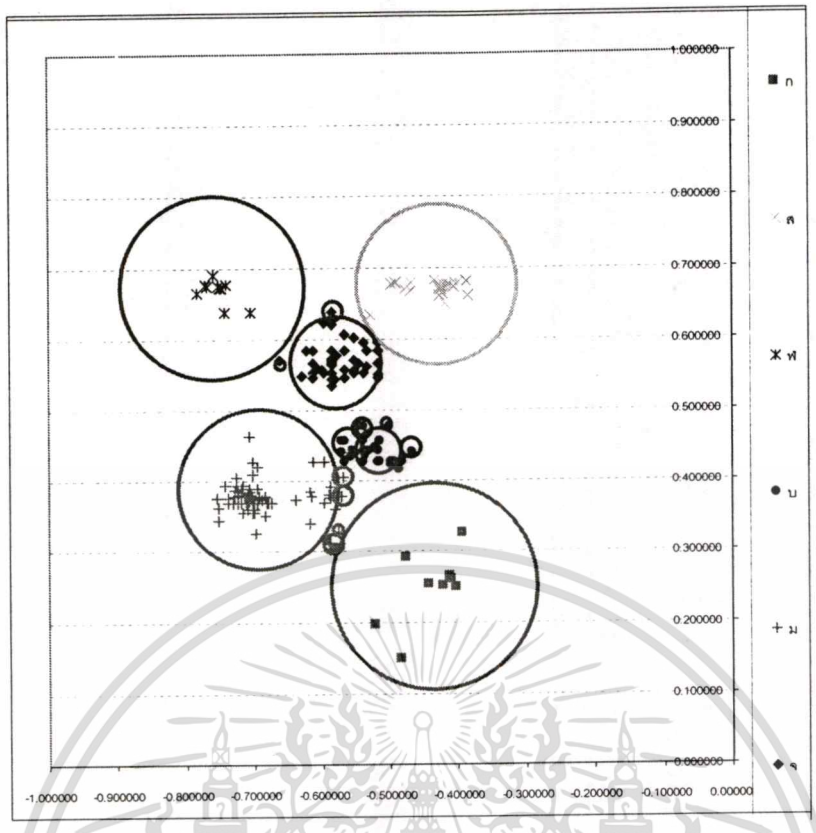
ตารางที่ 5.6 ผลการทดลองการแบ่งกลุ่มพยัญชนะภาษาไทยด้วย MNNC

กลุ่มที่	กลุ่ม	weight (จุดศูนย์กลางกลุ่ม)		Threshold(รัศมี)	
		feature 1	feature 2	Upper	lower
1	ก	-0.427978	0.2503020	0.1674217	0.142593
2	ส	-0.4102137	0.6829049	0.1264682	0.1080466
3	จ	-0.5864160	0.5769214	0.1268253	0.0978127
4	ม	-0.6641089	0.3833587	0.1300036	0.1079414
5	บ	-0.5447880	0.4436110	0.0576791	0.0301523
6	ฟ	-0.7597150	0.6716670	0.1557313	0.1394818

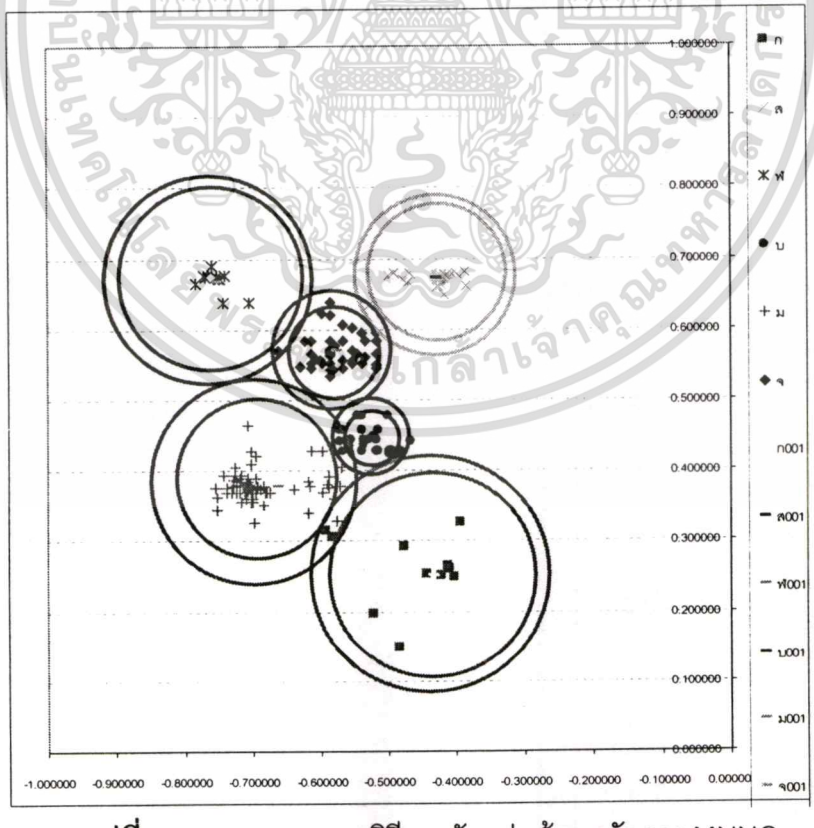


รูปที่ 5.4 ตัวอย่างข้อมูลตัวอักษรภาษาไทย 6 กลุ่มที่ใช้ในการจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.5 ผลการทดลองวิธีการจัดกลุ่มด้วยหลักการ NNC



รูปที่ 5.6 ผลการทดลองวิธีการจัดกลุ่มด้วยหลักการ MNNC

และต้นแบบที่ได้นำเสนอในวิทยานิพนธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผลการวิจัยและข้อเสนอแนะ

6.1 สรุปผลการวิจัย

การจัดกลุ่มข้อมูลให้ประโยชน์สองด้านคือ หนึ่งอธิบายลักษณะและขอบเขตของข้อมูลในแต่ละกลุ่ม สองทำนายความเป็นไปได้ของข้อมูลในอนาคตที่ไม่ทราบกลุ่มแต่ทราบลักษณะของข้อมูล สำหรับขั้นตอนการจัดกลุ่มข้อมูลแบ่งเป็นสองขั้นตอน โดยขั้นตอนแรกเป็นการสร้างต้นแบบของกลุ่มเพื่อกำหนดลักษณะและขอบเขตของข้อมูลตามข้อมูลที่ใช้ทดสอบเรียกว่าขั้นตอนการเรียนรู้ (training) ส่วนขั้นตอนที่สองเป็นการนำผลลัพธ์ที่ได้จากขั้นตอนแรกมาใช้งานนั่นคือการหากลุ่มให้กับข้อมูลที่ไม่ทราบกลุ่มแต่ทราบลักษณะของข้อมูลที่เรียกว่าขั้นตอนการทำนาย (prediction)

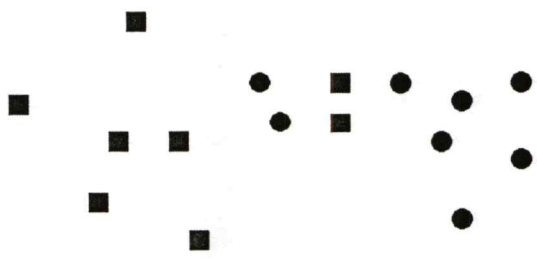
วิทยานิพนธ์นี้เป็นขั้นตอนแรกของการจัดกลุ่มข้อมูล คือเป็นการสร้างต้นแบบของกลุ่มข้อมูลโดยนำเสนอแนวความคิดต้นแบบสำหรับใช้ในการจัดกลุ่มข้อมูลที่มีการซ้อนทับกันและต้นแบบที่นำเสนอนี้ปรับปรุงจากหลักการการจัดกลุ่มด้วยวิธี Neural Network Classifier (NNC)[1] เริ่มต้นจากการนำข้อมูลที่ระบุกลุ่มแน่นอนแล้วจำนวนหนึ่ง มาผ่านการกระบวนการเรียนรู้ ด้วยหลักการที่นำเสนอในวิทยานิพนธ์นี้ นั่นคือหลักการแบ่งกลุ่มด้วยวิธี MNNC (คือใช้หลักการของ Neural network มาปรับปรุงการจัดกลุ่มข้อมูลให้สอดคล้องกับต้นแบบที่นำเสนอ) และ การนำเสนอต้นแบบสำหรับจัดกลุ่มข้อมูลที่มีการซ้อนทับกัน (สำหรับต้นแบบในการจัดกลุ่มที่นำเสนอในวิทยานิพนธ์นี้ประกอบด้วยจุดศูนย์กลางหรือน้ำหนักของกลุ่มหนึ่งค่า และรัศมีของกลุ่มสองค่า) ได้ผลลัพธ์ของงานวิจัยเป็นต้นแบบของกลุ่มต่างๆ ที่มีจุดศูนย์กลางและขอบเขตของกลุ่มต่างกันออกไป ตามจำนวนกลุ่มที่มีก่อนการ training สำหรับการนำไปใช้งานคือเมื่อมีข้อมูลที่ยังไม่ได้ระบุกลุ่มเข้ามาทดสอบจะทำการวัดระยะทางของข้อมูลนี้กับจุดศูนย์กลางของแต่ละกลุ่ม ถ้าระยะทางอยู่ภายในค่า lower (ค่ารัศมีที่สั้นกว่า) ของกลุ่มนั้น แสดงว่าข้อมูลนั้นมีความเป็นไปได้มากที่จะจัดอยู่ในกลุ่มนี้ ถ้าระยะทางอยู่ระหว่างค่า lower และ upper (ค่ารัศมีที่ยาวกว่า) แสดงว่าข้อมูลนั้นมีความเป็นไปได้ที่จะจัดอยู่ในกลุ่มนี้หรือกลุ่มอื่น ๆ แต่ถ้าระยะทางอยู่นอก upper แสดงว่าข้อมูลนั้นไม่ได้จัดอยู่ในกลุ่มนี้ สำหรับผลการทดลองได้นำข้อมูลตัวอักษรภาษาไทยมาทดสอบด้วยหลักการที่นำเสนอในวิทยานิพนธ์และนำผลการทดลองมาเปรียบเทียบกับหลักการแบ่งกลุ่มด้วย NNC[1] พบว่าการแบ่งกลุ่มที่นำเสนอในวิทยานิพนธ์นี้สามารถแบ่งกลุ่มข้อมูลที่มีการซ้อนทับกันได้เหมาะสมกว่าการแบ่งกลุ่มด้วยหลักการของ NNC [1] เมื่อพิจารณาจากการแบ่งกลุ่มด้วยขอบเขตของข้อมูล (data clustering boundary:density) โดยลักษณะของการจัดกลุ่มข้อมูลที่ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ควรจะมีจำนวนกลุ่มมากกว่าหรือเท่ากับจำนวนกลุ่มที่มีในขั้นตอนการเรียนรู้เล็กน้อย หมายถึง ข้อมูลกลุ่มเดียวกันจะไม่มีการแตกออกเป็นกลุ่มย่อยมากนัก แต่ละกลุ่มควรจะมีพื้นที่ส่วน lower มากกว่าพื้นที่ระหว่าง lower และ upper เพื่อให้ต้นแบบมีประสิทธิภาพสูงในการระบุกลุ่มของข้อมูลได้แม่นยำสำหรับการนำไปใช้ในขั้นตอนการทำนาย

6.2 ข้อเสนอแนะ

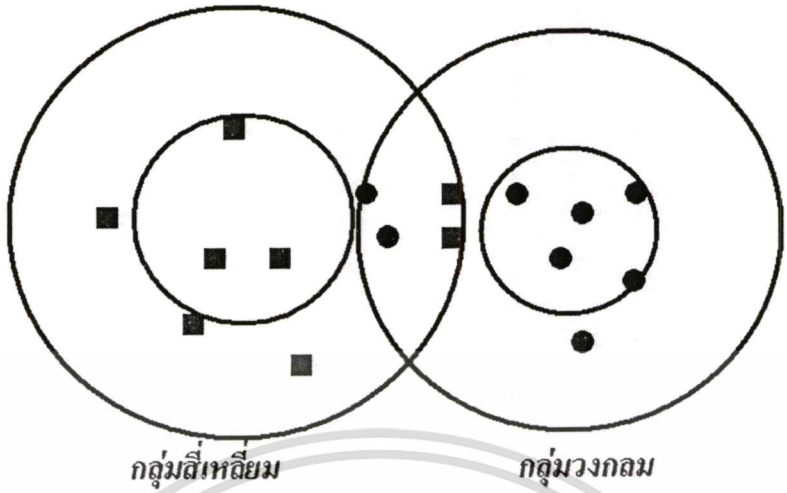
สำหรับการปรับปรุงงานวิจัยต่อไปคือ เพิ่มความประสิทธิภาพการจัดกลุ่มฐานข้อมูลขนาดใหญ่ หรือฐานข้อมูลที่มีหลายมิติได้ (สำหรับงานวิทยานิพนธ์นี้ใช้ข้อมูล 2 มิติ) จากงานวิทยานิพนธ์นี้สามารถทำได้ 2 ส่วนคือ ส่วนที่หนึ่งโดยการเปลี่ยนวิธีการจัดกลุ่มจากนิวรอนเน็ตเวิร์ก เป็นหลักการอื่น เช่น เทคนิค Fuzzy เสนอโดย P. K. Simpson [13-14], เทคนิค ART (Adaptive Resonance Theory) เสนอโดย G.A. Carpenter [15-16] ซึ่งเทคนิคทั้งสองนี้ได้ถูกนำมาใช้ในการจัดกลุ่มข้อมูลที่ไม่ซ้อนทับกัน, เทคนิค Density-Base เนื่องจากเทคนิคนี้ใช้ความหนาแน่นของข้อมูลในการจัดกลุ่มทำให้รูปทรงของกลุ่มไม่จำกัดรูปร่างโดยไม่เป็นวงกลมหรือทรงกลมเหมาะสำหรับข้อมูลที่เกาะกลุ่มกัน หรือใช้เทคนิคจากทฤษฎีอื่นๆ ซึ่งใช้หลักการของความหนาแน่นของข้อมูล ส่วนที่สองอาจปรับปรุงต้นแบบของกลุ่มข้อมูลเพื่อให้การจัดกลุ่มมีประสิทธิภาพมากขึ้นสำหรับข้อมูลที่มีการซ้อนทับกัน จากรูปที่ 6.1 แสดงข้อมูลสำหรับใช้ในการจัดกลุ่ม ในรูปประกอบด้วย 2 กลุ่มคือกลุ่มข้อมูลที่มีลักษณะสี่เหลี่ยมและกลุ่มข้อมูลที่เป็นวงกลม รูปที่ 6.2 แสดงการจัดกลุ่มข้อมูลที่ได้จากการทดลองนี้ซึ่งสามารถแบ่งกลุ่มได้ 2 กลุ่มและมีส่วนที่ซ้อนทับกันอยู่ส่วนกลาง ส่วนรูปที่ 6.3 และ รูปที่ 6.4 เป็นการนำเสนอ ต้นแบบที่ปรับปรุงประสิทธิภาพให้ดีขึ้น โดยภาพที่ 6.3 แสดงการจัดกลุ่มแบบที่ 1 จะให้พื้นที่ในส่วนที่เป็นของกลุ่มสี่เหลี่ยมแน่นอน กลุ่มวงกลมแน่นอนมากขึ้นเมื่อเปรียบเทียบกับรูปที่ 6.2 ส่วนรูปที่ 6.4 แสดงการจัดกลุ่มแบบที่ 2 โดยแบ่งกลุ่มข้อมูลที่การซ้อนทับกันออกมาเป็นอีกกลุ่มหนึ่งทำให้ข้อมูลนี้สามารถแบ่งกลุ่มตามการแบ่งกลุ่มในรูปที่ 6.4 ได้ทั้งหมด 3 กลุ่ม ทั้งนี้ต้นแบบที่นำเสนอทั้งสองแบบจะต้องสร้างอัลกอริทึมมาควบคุมการทำงานเพิ่มเติมต่อไป



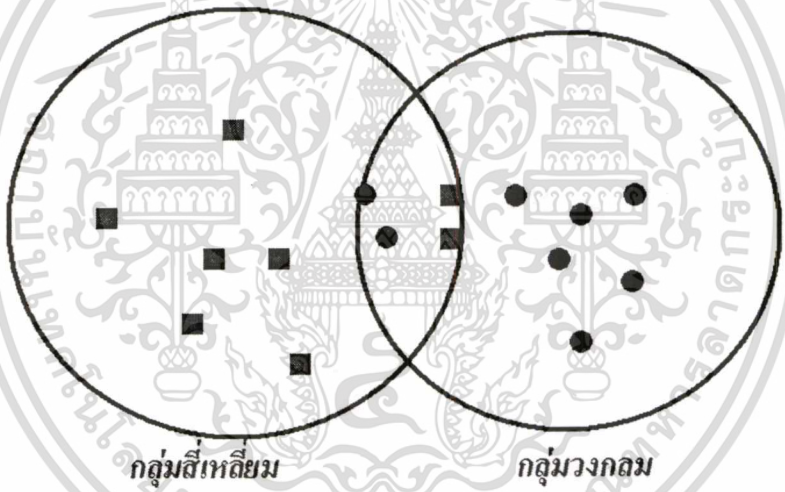
กลุ่มสี่เหลี่ยม *กลุ่มวงกลม*

รูปที่ 6.1 ภาพแสดงข้อมูลสำหรับใช้ในการจัดกลุ่ม

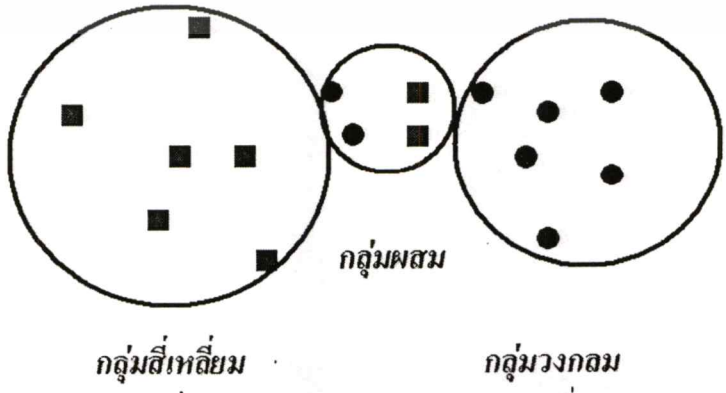
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 6.2 ภาพแสดงการจัดกลุ่มที่นำเสนอในงานวิจัยนี้



รูปที่ 6.3 ภาพแสดงการจัดกลุ่มแบบที่ 1



รูปที่ 6.4 ภาพแสดงการจัดกลุ่มแบบที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] M.A. ABOU-NASR and M.A. SID-AHMED. 1995. "Fast learning and efficient memory utilization with a prototype based neural classifier." **Pattern Recognition**. 28(4): 581-593.

[2] Jiawei Han, Micheline Kamber. 2001. **Data Mining Concepts and Techniques**. Morgan Kaufmann.

[3] Sarah Coppock, Lawrence Mazlack. 2003. "Rough Sets Used In The Measurement Of Similarity Of Mixed Mode Data." 197-201. in **22nd International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2003) Proceedings**. Chicago.

[4] A.K. Jain, M.N. Murty, P.J. Flynn. 1999. "Data Clustering: A Review." **ACM Computing Surveys**. 31(3): .

[5] Anil K. Jain, Fellow IEEE, Robert P.W. Dulin, and Jaincahang Mao. 2000. "Statistical Pattern Recognition: A Review." **IEEE Transactions on pattern analysis and machine intelligence**. 22(1): .

[6] พรเทพ โรจนวสุ และ เอื้อน ปิ่นเงิน. "**การจัดกลุ่มเอกสารโดยใช้ Self-Organizing Map แบบความเร็วสูง**." วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2547.

[7] R. Hecht-Nielson. 1990. **Neurocomputing**. Massachusetts : Addison-Wesley.

[8] Y.Pao. 1989. **Adaptive Patterns Recognition and Neural Networks**. Massachusetts : Addision-Wesley.

[9] B.Kosko. 1991. "Stochastic competitive learning." **IEEE Trans**. Neural Networks 2 : 522-529.

[10] R.Duda and P.Hart. 1973. **Pattern Classification and Science Analysis**. New York: Weilay-Interscience.

[11] D.E.Rumellhart and J.L. McClelland. 1987. **Parallel Distributed Processing**. Vol. 1 MIT Press. Massachusetts : Cambridge.

[12] D.L. Reilay, L.N. Cooper and CElbaum. 1982. "A neural model for category learning." **Biol**. Cybernetics 45: 35-41.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[13] P.K. Simpson. 1993. "Fuzzy Min-Max neural networks-part 2: Clustering." **IEEE trans.** Fuzzy Systems 1: 32-45.

[14] P.K. Simpson. 1992. "Fuzzy Min-Max neural networks-part 1: Classification." **IEEE trans.** Neural Networks 3: 776-786.

[15] G.A. Carpenter and S. Grossberg. 1987. "A massive parallel architecture for a self-organizing neural pattern recognition machine." **Computer Vision Graphics Image Process.** 37: 54-115.

[16] G.A. Carpenter and S. Grossberg. 1987. "ART 2: self-organization of stable category recognition codes for analog input patterns." **Applied Optics.** 26: 4919-4930.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก
ข้อมูลในไฟล์ data012_4.txt

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลในไฟล์ data012_4.txt เป็นข้อมูลที่ใช้ในขั้นตอนการเรียนรู้ (training data) ของการทดลองที่ 2 ในบทที่ 5 ในไฟล์ประกอบด้วย ข้อมูลจำนวน 354 ข้อมูล มี 4 คุณลักษณะ (feature) และในการทดลองที่ 2 ใช้ 2 คุณลักษณะแรก

ตารางที่ ก.1 แสดงข้อมูลในไฟล์ data012_4.txt

no.	ch	feature1	feature2	feature3	feature4
1	ก013	-0.402433	0.248172	0.753737	0.246263
2	ก021	-0.409239	0.260210	0.875250	0.124750
3	ข001	-0.517364	0.427905	0.498553	0.501447
4	ข004	-0.618794	0.440945	0.403409	0.596591
5	ง004	-0.695662	0.096348	0.723025	0.269866
6	ง035	-0.624342	0.240138	0.872033	0.109290
7	จ001	-0.577896	0.577896	0.297348	0.689103
8	จ002	-0.580406	0.542443	0.265239	0.726844
9	ฉ022	-0.807844	0.349357	0.693878	0.294290
10	ฉ090	-0.825803	0.413790	0.657422	0.336948
11	ช010	-0.584182	0.681950	0.422911	0.577089
12	ช013	-0.552867	0.517364	0.529136	0.455482
13	ซ021	-0.552867	0.737155	0.189191	0.793438
14	ซ027	-0.552867	0.783439	0.129096	0.870904
15	ฌ018	-0.847727	0.360006	0.774637	0.194581
16	ฌ041	-0.852833	0.388542	0.777028	0.206377
17	ด052	-0.784810	0.202388	0.894435	0.105565
18	ด070	-0.723419	0.208947	0.879382	0.120618
19	ท042	-0.601073	0.326123	0.962913	0.037087
20	น031	-0.787137	0.132847	0.863555	0.118964
21	บ005	-0.292091	0.457577	0.288745	0.711255
22	บ007	-0.427905	0.476380	0.342092	0.657907
23	ป005	-0.184289	0.476380	0.246709	0.753291
24	ป006	-0.091288	0.448843	0.218321	0.781679
25	พ019	-0.713336	0.378008	0.625184	0.374816
26	พ054	-0.730259	0.399455	0.658730	0.332416
27	ฟ052	-0.746788	0.405177	0.553711	0.446289
28	ฟ080	-0.737155	0.417364	0.667911	0.320392
29	ม001	-0.737155	0.368578	0.771071	0.228929
30	ม002	-0.729878	0.378899	0.639982	0.349322
31	ล002	-0.596016	0.549325	0.567568	0.432432
32	ล004	-0.588818	0.546793	0.377069	0.608569
33	ว006	-0.184289	0.669594	0.030794	0.969206
34	ว007	-0.233832	0.666991	0.047571	0.952429
35	ห039	-0.584182	0.641334	0.464387	0.518348
36	ห035	-0.704591	0.638929	0.607496	0.392504
37	ห042	-0.711267	0.630553	0.431602	0.555794
38	อ096	-0.184289	0.673436	0.054616	0.945384
39	อ098	-0.184289	0.637535	0.030698	0.969302
40	ฮ021	-0.292091	0.627444	0.135477	0.864523
41	ก050	-0.421758	0.251352	0.841478	0.152506
42	ก098	-0.412089	0.265182	0.903737	0.096263
43	ข005	-0.551352	0.446599	0.281926	0.718074
44	ข006	-0.591505	0.419350	0.288157	0.686149
45	ข007	-0.591505	0.440945	0.367687	0.632313
46	ข008	-0.615721	0.434913	0.463466	0.531467
47	ข009	-0.524646	0.438045	0.356723	0.643277
48	ข010	-0.562497	0.422703	0.426171	0.573829
49	ข011	-0.521355	0.433476	0.257484	0.742516
50	ข013	-0.546156	0.417086	0.250715	0.736783

no.	ch	feature1	feature2	feature3	feature4
178	ป083	-0.644969	0.443995	0.426866	0.573134
179	ป084	-0.528955	0.443995	0.320912	0.679088
180	ป085	-0.570956	0.440945	0.417015	0.548919
181	ป088	-0.538891	0.457577	0.336773	0.663227
182	ป090	-0.352904	0.484267	0.262160	0.728445
183	ป091	-0.545785	0.476380	0.440455	0.559545
184	ป092	-0.554872	0.443995	0.420203	0.574506
185	ป093	-0.485538	0.417086	0.339600	0.660400
186	ป096	-0.405177	0.457577	0.302309	0.690465
187	ป100	-0.601073	0.451472	0.385608	0.579980
188	ป007	-0.487290	0.476380	0.346334	0.636628
189	ป012	-0.552867	0.479674	0.382375	0.617625
190	ป023	-0.493940	0.444908	0.365976	0.634024
191	ป024	-0.528955	0.431108	0.364567	0.635433
192	ป029	-0.215761	0.440945	0.164193	0.822029
193	ป036	-0.526743	0.440945	0.300274	0.699726
194	ป039	-0.571796	0.476380	0.318950	0.681050
195	ป040	-0.514667	0.454098	0.305945	0.694055
196	ป044	-0.474851	0.448843	0.321059	0.678941
197	ป049	-0.584182	0.476380	0.395483	0.604517
198	ป050	-0.253054	0.457577	0.201613	0.798387
199	ป057	-0.532486	0.439812	0.304463	0.695537
200	ป063	-0.566144	0.448843	0.404352	0.595648
201	ป069	-0.509878	0.478460	0.449842	0.543282
202	ป076	-0.550508	0.437343	0.306036	0.693964
203	ป080	-0.485538	0.448843	0.281835	0.718165
204	ป081	-0.558804	0.484267	0.266422	0.718334
205	ป082	-0.326123	0.460917	0.255795	0.744205
206	ป084	-0.517364	0.457577	0.343606	0.641117
207	ป085	-0.313481	0.440945	0.206259	0.777143
208	ป087	-0.524832	0.454456	0.255938	0.744062
209	ป088	-0.091288	0.448843	0.139384	0.860616
210	ป092	-0.493940	0.458981	0.303362	0.696638
211	ป093	-0.476380	0.483305	0.329882	0.640438
212	พ071	-0.744932	0.398406	0.714017	0.272475
213	ม003	-0.704052	0.376882	0.611599	0.388401
214	ม004	-0.703836	0.376142	0.634450	0.352013
215	ม005	-0.708014	0.390206	0.719056	0.280944
216	ม006	-0.716775	0.394501	0.740531	0.249960
217	ม007	-0.734136	0.394501	0.714291	0.281555
218	ม010	-0.701653	0.427905	0.739317	0.260683
219	ม011	-0.727160	0.394501	0.721893	0.252225
220	ม012	-0.684565	0.378723	0.706492	0.293508
221	ม013	-0.707512	0.463747	0.753824	0.219965
222	ม015	-0.692478	0.376142	0.630151	0.361383
223	ม016	-0.698200	0.326123	0.698892	0.293254
224	ม017	-0.568027	0.405776	0.586526	0.413474
225	ม018	-0.693757	0.378308	0.797070	0.202930
226	ม019	-0.555806	0.377781	0.540509	0.459491
227	ม020	-0.688736	0.374152	0.614324	0.385676

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
51	ข015	-0.525233	0.438858	0.322299	0.677701
52	ข017	-0.522862	0.427263	0.327043	0.672956
53	ข019	-0.551022	0.446110	0.321300	0.667837
54	ข020	-0.517364	0.422656	0.507852	0.492148
55	ข023	-0.561091	0.443995	0.410270	0.574125
56	ข024	-0.637535	0.405177	0.375103	0.624897
57	ข025	-0.557051	0.427160	0.401672	0.598328
58	ข026	-0.496632	0.387449	0.362532	0.637468
59	ข029	-0.584182	0.427905	0.619848	0.366381
60	ข031	-0.635911	0.424321	0.376132	0.623868
61	ข033	-0.512732	0.398081	0.428701	0.571299
62	ข035	-0.580012	0.423645	0.368756	0.591281
63	ข036	-0.576121	0.423919	0.435755	0.564245
64	ข037	-0.552867	0.476380	0.465807	0.529153
65	ข040	-0.552867	0.427905	0.331062	0.648049
66	ข045	-0.517364	0.394501	0.365700	0.634300
67	ข047	-0.578377	0.417364	0.333483	0.652483
68	ข049	-0.573974	0.419350	0.346127	0.643065
69	ข051	-0.519655	0.443995	0.208891	0.791109
70	ข053	-0.561091	0.427905	0.336559	0.650302
71	ข055	-0.513625	0.422656	0.489382	0.510618
72	ข057	-0.618214	0.427905	0.560980	0.391241
73	ข061	-0.637535	0.427905	0.404860	0.595140
74	ข065	-0.578790	0.400331	0.520912	0.463662
75	ข067	-0.517364	0.427905	0.608227	0.391773
76	ข072	-0.517364	0.427905	0.290140	0.709860
77	ข073	-0.552867	0.415441	0.591942	0.408058
78	ข074	-0.558077	0.436194	0.374708	0.625292
79	ข080	-0.526743	0.427905	0.498952	0.501048
80	ข082	-0.520756	0.423033	0.459908	0.530110
81	ข084	-0.552867	0.427905	0.443486	0.537875
82	ข086	-0.559765	0.422445	0.498427	0.484932
83	ข090	-0.601892	0.422445	0.436247	0.530570
84	ข091	-0.654192	0.142830	0.931569	0.051251
85	ข092	-0.558077	0.436194	0.339955	0.644953
86	ข093	-0.556362	0.422294	0.388407	0.605209
87	ข096	-0.526743	0.440945	0.599329	0.400671
88	ข097	-0.561091	0.427905	0.608393	0.391607
89	ง077	-0.681950	0.148584	0.873068	0.126932
90	ง093	-0.578065	0.131036	0.867099	0.115470
91	จ003	-0.552867	0.552867	0.202720	0.761784
92	จ005	-0.612194	0.545949	0.412695	0.563287
93	จ006	-0.517364	0.550212	0.392640	0.593040
94	จ009	-0.552867	0.601073	0.373763	0.615620
95	จ012	-0.584182	0.620361	0.427455	0.572545
96	จ013	-0.607347	0.556362	0.439954	0.538338
97	จ014	-0.584182	0.584182	0.445621	0.542767
98	จ015	-0.584182	0.584182	0.377939	0.614507
99	จ016	-0.517364	0.566144	0.320390	0.665470
100	จ019	-0.584182	0.637535	0.390285	0.597662
101	จ020	-0.584182	0.552867	0.463887	0.536113
102	จ021	-0.566144	0.544380	0.526342	0.473658
103	จ023	-0.566144	0.582786	0.424924	0.567751
104	จ025	-0.584182	0.533286	0.367813	0.603729
105	จ027	-0.532486	0.584182	0.263034	0.720696
106	จ028	-0.596016	0.622869	0.368796	0.618839

no.	ch	feature1	feature2	feature3	feature4
228	ม021	-0.754798	0.374997	0.727934	0.272066
229	ม022	-0.712634	0.368578	0.730429	0.269571
230	ม023	-0.576121	0.377781	0.628341	0.371659
231	ม024	-0.703070	0.373504	0.688516	0.311484
232	ม026	-0.738775	0.374997	0.632356	0.363416
233	ม027	-0.737155	0.368578	0.597392	0.402608
234	ม028	-0.703070	0.409475	0.657929	0.342071
235	ม029	-0.596017	0.427905	0.719528	0.280472
236	ม030	-0.729797	0.386866	0.690955	0.305069
237	ม031	-0.724414	0.379762	0.682855	0.313296
238	ม032	-0.587570	0.391902	0.616108	0.371732
239	ม033	-0.618794	0.384779	0.633095	0.366905
240	ม034	-0.705062	0.380323	0.802019	0.197981
241	ม035	-0.587054	0.374997	0.590991	0.409009
242	ม036	-0.743877	0.394501	0.671647	0.328353
243	ม037	-0.709321	0.383379	0.776462	0.223538
244	ม038	-0.612194	0.427905	0.777137	0.216129
245	ม039	-0.719996	0.368578	0.767172	0.232828
246	ม040	-0.727160	0.404487	0.544036	0.439042
247	ม041	-0.723179	0.384148	0.802858	0.181315
248	ม042	-0.753274	0.343916	0.611863	0.384392
249	ม043	-0.706384	0.377461	0.741702	0.258298
250	ม044	-0.703656	0.354516	0.658682	0.341318
251	ม045	-0.707081	0.372063	0.674337	0.325663
252	ม047	-0.699369	0.354557	0.721168	0.278832
253	ม048	-0.701653	0.368578	0.621361	0.378639
254	ม050	-0.681950	0.368578	0.717445	0.282555
255	ม051	-0.681950	0.368578	0.742446	0.257554
256	ม052	-0.706384	0.384779	0.616270	0.370591
257	ม053	-0.639575	0.374152	0.627151	0.359183
258	ม054	-0.615914	0.377781	0.655559	0.344441
259	ม055	-0.696837	0.368578	0.616837	0.383163
260	ม056	-0.719996	0.368578	0.732556	0.267444
261	ม058	-0.701653	0.368578	0.681730	0.318270
262	ม060	-0.701653	0.368578	0.613078	0.373225
263	ม061	-0.681950	0.368578	0.759787	0.227325
264	ม062	-0.699452	0.360792	0.731542	0.268458
265	ม063	-0.701653	0.368578	0.720695	0.279305
266	ม065	-0.730259	0.368578	0.743202	0.241416
267	ม066	-0.580750	0.360792	0.675285	0.324715
268	ม069	-0.678229	0.368578	0.766158	0.233842
269	ม070	-0.696312	0.388542	0.616926	0.368781
270	ม071	-0.701653	0.368578	0.730932	0.269068
271	ม072	-0.701653	0.368578	0.785513	0.206683
272	ม073	-0.683186	0.351324	0.685298	0.314702
273	ม076	-0.701653	0.368578	0.698608	0.293155
274	ม080	-0.724414	0.368578	0.644476	0.321151
275	ม081	-0.719996	0.368578	0.751991	0.239242
276	ม082	-0.596016	0.368578	0.745501	0.254499
277	ม083	-0.716775	0.356289	0.743050	0.256950
278	ม085	-0.695820	0.421695	0.778260	0.221740
279	ม086	-0.589466	0.380323	0.731648	0.256699
280	ม087	-0.701653	0.427905	0.661492	0.338508
281	ม090	-0.709321	0.360285	0.698971	0.301029
282	ม091	-0.701653	0.368578	0.772521	0.227479
283	ม092	-0.753274	0.362000	0.782427	0.199422

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
107	จ037	-0.637535	0.612194	0.370974	0.611271
108	จ039	-0.566144	0.607347	0.284941	0.699402
109	จ040	-0.552867	0.570729	0.443668	0.556332
110	จ045	-0.538891	0.552867	0.232757	0.756019
111	จ046	-0.612194	0.566144	0.387447	0.594416
113	จ048	-0.517364	0.544380	0.448145	0.551855
114	จ050	-0.612194	0.560098	0.536548	0.457237
115	จ051	-0.584182	0.584182	0.387947	0.588376
116	จ052	-0.552867	0.552867	0.484956	0.502069
117	จ057	-0.612194	0.584182	0.473869	0.518528
118	จ059	-0.532486	0.561091	0.405797	0.573782
119	จ060	-0.566144	0.556675	0.369751	0.630249
120	จ061	-0.517364	0.584182	0.426114	0.573886
121	จ062	-0.552867	0.557314	0.339533	0.660467
122	จ064	-0.517364	0.596017	0.211418	0.780642
123	จ065	-0.584182	0.563036	0.448410	0.551590
124	จ067	-0.629887	0.547121	0.422070	0.572016
125	จ069	-0.612194	0.584182	0.484006	0.503587
126	จ071	-0.584182	0.571796	0.226711	0.773289
127	จ073	-0.678229	0.552867	0.449949	0.516557
128	จ076	-0.605427	0.558077	0.375468	0.615725
129	จ077	-0.538891	0.594012	0.379836	0.588288
130	จ078	-0.584182	0.566144	0.470892	0.529108
131	จ083	-0.544380	0.566144	0.467551	0.527817
132	จ084	-0.584182	0.568499	0.444867	0.520215
133	จ085	-0.612194	0.552867	0.489814	0.503925
134	จ088	-0.584182	0.552867	0.500207	0.499793
135	จ089	-0.584182	0.569470	0.398722	0.601278
136	จ092	-0.622869	0.584182	0.288074	0.701531
137	จ096	-0.690199	0.383379	0.753511	0.246489
138	จ097	-0.584182	0.542488	0.541923	0.458077
139	จ100	-0.596017	0.552867	0.470860	0.524839
140	ช014	-0.552867	0.517364	0.553030	0.446970
141	ช017	-0.566144	0.681950	0.341379	0.639906
142	ช023	-0.566144	0.681950	0.308540	0.671640
143	ช028	-0.637535	0.517364	0.512367	0.481958
144	ช030	-0.584182	0.517364	0.532232	0.467768
145	ช034	-0.517364	0.669594	0.235413	0.746163
146	ช041	-0.566144	0.544380	0.647480	0.345129
147	ฌ047	-0.852181	0.384779	0.767636	0.213560
148	ฌ055	-0.847340	0.374152	0.797514	0.191809
149	ฌ070	-0.848370	0.372802	0.712674	0.256864
150	บ009	-0.538891	0.476380	0.455808	0.544192
151	บ017	-0.535157	0.435745	0.433066	0.566934
152	บ019	-0.601073	0.457577	0.474847	0.525153
153	บ024	-0.514230	0.457577	0.429160	0.570840
154	บ030	-0.515748	0.443995	0.443503	0.556497
155	บ037	-0.123350	0.457577	0.228399	0.771601
156	บ040	-0.495194	0.427905	0.448756	0.551244
157	บ043	-0.368578	0.457577	0.209542	0.782557
158	บ044	-0.498295	0.427905	0.465081	0.534919
159	บ045	-0.517364	0.427905	0.500836	0.499164
160	บ047	-0.394501	0.448843	0.327278	0.672722

no.	ch	feature1	feature2	feature3	feature4
284	ม095	-0.718258	0.388542	0.734497	0.265503
285	ม097	-0.715504	0.378899	0.736470	0.263530
286	ม098	-0.570562	0.377781	0.549808	0.450192
287	ม099	-0.616955	0.339850	0.692394	0.307606
288	ม100	-0.673437	0.368578	0.758457	0.241543
290	ล005	-0.603025	0.553849	0.459149	0.540851
291	ล006	-0.596017	0.541354	0.507799	0.477449
292	ล007	-0.476380	0.562294	0.406590	0.564914
293	ล008	-0.448843	0.596016	0.421315	0.578685
294	ล009	-0.457577	0.538891	0.459035	0.540965
295	ล010	-0.601073	0.538891	0.451154	0.548847
296	ล011	-0.462186	0.565083	0.425911	0.574089
297	ล012	-0.448843	0.566144	0.343158	0.642673
298	ล013	-0.541354	0.566144	0.341883	0.658117
299	ล014	-0.575827	0.550369	0.455225	0.539929
300	ล017	-0.719996	0.552867	0.579430	0.414059
301	ล018	-0.622869	0.546793	0.406423	0.593577
302	ล020	-0.517364	0.536596	0.402611	0.580057
303	ล021	-0.616955	0.530746	0.400338	0.599662
304	ล022	-0.612194	0.532486	0.310320	0.657285
305	ล023	-0.709321	0.603025	0.508202	0.466735
306	ล024	-0.737155	0.563036	0.578858	0.421142
307	ล027	-0.785014	0.519552	0.551915	0.392337
308	ล030	-0.712634	0.592788	0.536059	0.458490
309	ล031	-0.591505	0.552867	0.459123	0.540877
310	ล032	-0.612194	0.563036	0.408737	0.563648
311	ล035	-0.681950	0.563901	0.501237	0.481653
312	ล037	-0.660669	0.569470	0.529927	0.451911
313	ล040	-0.622869	0.538891	0.532481	0.452934
314	ล042	-0.606035	0.524139	0.532755	0.448649
315	ล043	-0.622869	0.582248	0.530511	0.458054
316	ล047	-0.517364	0.584182	0.417292	0.582708
317	ล049	-0.517364	0.596017	0.434533	0.565467
318	ล050	-0.614066	0.566144	0.521514	0.464130
319	ล051	-0.603025	0.561091	0.379806	0.620194
320	ล052	-0.612194	0.563538	0.510592	0.489408
321	ล053	-0.627444	0.558804	0.514605	0.477795
322	ล055	-0.681950	0.584182	0.541769	0.433925
323	ล056	-0.627444	0.622112	0.493685	0.500111
324	ล057	-0.637535	0.637535	0.505562	0.463716
325	ล059	-0.584182	0.550508	0.496566	0.503434
326	ล061	-0.637535	0.594729	0.516972	0.474472
327	ล064	-0.618794	0.552867	0.537838	0.452224
328	ล071	-0.727160	0.562578	0.472262	0.456440
329	ล074	-0.660669	0.552867	0.550944	0.415653
330	ล076	-0.637535	0.546793	0.396980	0.580462
331	ล077	-0.618794	0.563901	0.532221	0.429374
332	ล079	-0.612194	0.561091	0.382683	0.617317
333	ล080	-0.588818	0.546793	0.555150	0.444850
334	ล081	-0.612194	0.561091	0.572741	0.427259
335	ล083	-0.609107	0.538891	0.499520	0.478582
336	ล085	-0.618794	0.546793	0.583460	0.416540
337	ล086	-0.592735	0.566144	0.537374	0.462626

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
161	บ048	-0.380323	0.437343	0.230566	0.754299
162	บ051	-0.538891	0.427905	0.464327	0.527190
163	บ052	-0.339850	0.457577	0.272737	0.727263
164	บ054	-0.515053	0.427905	0.421284	0.572299
165	บ055	-0.521917	0.448843	0.286588	0.713412
166	บ056	-0.493940	0.427905	0.437799	0.562201
167	บ057	-0.521051	0.448843	0.401109	0.598891
168	บ058	-0.497025	0.424121	0.514926	0.485074
169	บ060	-0.481944	0.427905	0.451337	0.548663
170	บ061	-0.557051	0.434569	0.382397	0.611644
171	บ064	-0.566144	0.427905	0.432948	0.552950
172	บ065	-0.517364	0.427905	0.478761	0.521239
173	บ068	-0.466284	0.440945	0.387127	0.612873
174	บ069	-0.566144	0.457577	0.498828	0.501172
175	บ072	-0.499939	0.476380	0.435218	0.564782
176	บ077	-0.571796	0.457577	0.441375	0.558625
177	บ078	-0.535444	0.443995	0.453876	0.512724

no.	ch	feature1	feature2	feature3	feature4
338	ล088	-0.601073	0.569470	0.331983	0.624461
339	ล090	-0.561091	0.534775	0.496484	0.503516
340	ล092	-0.633132	0.544380	0.487506	0.497786
341	ล097	-0.612194	0.534172	0.531079	0.468921
342	ล098	-0.571796	0.584182	0.573658	0.397003
343	ล103	-0.612194	0.637535	0.368281	0.611012
344	ล106	-0.620360	0.546793	0.510792	0.441381
345	ว057	-0.222822	0.671364	0.073347	0.926653
346	ว059	-0.184289	0.628186	0.101619	0.889555
347	ว063	-0.184289	0.660669	0.078502	0.921498
348	ว066	-0.184289	0.669594	0.021816	0.960111
349	ว093	-0.253054	0.690199	0.020435	0.979565
350	ว099	-0.091288	0.589466	0.063107	0.936893
351	ห082	-0.622869	0.622869	0.548048	0.446340
352	ห083	-0.584182	0.532486	0.555664	0.433404
353	ห084	-0.463747	0.646625	0.388358	0.603532
354	ห092	-0.714409	0.629887	0.506347	0.493653

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

ข้อมูลในไฟล์ data011.txt

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลในไฟล์ data011.txt เป็นข้อมูลที่ใช้ในขั้นตอนการทดสอบ (testing data) ของการทดลองที่ 2 ในบทที่ 5 ในไฟล์ประกอบด้วย ข้อมูลจำนวน 570 ข้อมูล มี 4 คุณลักษณะ (feature) และในการทดลองที่ 2 ใช้ 2 คุณลักษณะแรก

ตารางที่ ข.1 แสดงข้อมูลในไฟล์ data011.txt

no.	ch	feature1	feature2	feature3	feature4
1	ก010	-0.409239	0.260210	0.875250	0.124750
2	ก048	-0.412089	0.265182	0.903737	0.096263
3	ก067	-0.483305	0.148584	0.819063	0.180937
4	ก075	-0.520790	0.196092	0.881577	0.103226
5	ก077	-0.394501	0.326123	0.847159	0.125990
6	ก110	-0.476380	0.292091	0.923169	0.076832
7	ก115	-0.443995	0.253054	0.874709	0.125291
8	ข001	-0.517364	0.427905	0.498553	0.501447
9	ข003	-0.551352	0.446599	0.281926	0.718074
10	ข004	-0.591505	0.440945	0.367687	0.632313
11	ข005	-0.524646	0.438045	0.356723	0.643277
12	ข006	-0.521355	0.433476	0.257484	0.742516
13	ข007	-0.546156	0.417086	0.250715	0.736783
14	ข008	-0.525233	0.438858	0.322299	0.677701
15	ข009	-0.522862	0.427263	0.327043	0.672956
16	ข010	-0.551022	0.446110	0.321300	0.667837
17	ข012	-0.561091	0.443995	0.410270	0.574125
18	ข013	-0.557051	0.427160	0.401672	0.598328
19	ข015	-0.584182	0.427905	0.619848	0.366381
20	ข016	-0.635911	0.424321	0.376132	0.623868
21	ข017	-0.512732	0.398081	0.428701	0.571299
22	ข018	-0.580012	0.423645	0.368756	0.591281
23	ข019	-0.552867	0.476380	0.465807	0.529153
24	ข023	-0.517364	0.394501	0.365700	0.634300
25	ข024	-0.578377	0.417364	0.333483	0.652483
26	ข025	-0.573974	0.419350	0.346127	0.643065
27	ข026	-0.519655	0.443995	0.208891	0.791109
28	ข027	-0.561091	0.427905	0.336559	0.650302
29	ข028	-0.513625	0.422656	0.489382	0.510618
30	ข029	-0.618214	0.427905	0.560980	0.391241
31	ข031	-0.637535	0.427905	0.404860	0.595140
32	ข033	-0.578790	0.400331	0.520912	0.463662
33	ข034	-0.517364	0.427905	0.608227	0.391773
34	ข036	-0.566144	0.427905	0.467805	0.532195
35	ข037	-0.552867	0.415441	0.591942	0.408058
36	ข046	-0.654192	0.142830	0.931569	0.051251
37	ข047	-0.556362	0.422294	0.388407	0.605209
38	ข049	-0.561091	0.427905	0.608393	0.391607
39	ข058	-0.596017	0.405177	0.516713	0.464640
40	ข060	-0.584182	0.427905	0.607691	0.392309
41	ข061	-0.622869	0.394501	0.472375	0.527625
42	ข062	-0.806540	0.151958	0.812408	0.167919
43	ข063	-0.660669	0.427905	0.148773	0.843021
44	ข073	-0.601073	0.184289	0.976728	0.023272
45	ข076	-0.517364	0.476380	0.511654	0.468284
46	ข082	-0.558405	0.233832	0.916397	0.050753
47	ข086	-0.416482	0.466895	0.515143	0.484857
48	ข087	-0.130467	0.413215	0.273351	0.726649
49	ข088	-0.509516	0.388542	0.497943	0.502057

no.	ch	feature1	feature2	feature3	feature4
286	บ022	-0.368578	0.457577	0.209542	0.782557
287	บ023	-0.517364	0.427905	0.500836	0.499164
288	บ024	-0.394501	0.448843	0.327278	0.672722
289	บ026	-0.538891	0.427905	0.464327	0.527190
290	บ028	-0.521917	0.448843	0.286588	0.713412
291	บ029	-0.521051	0.448843	0.401109	0.598891
292	บ031	-0.557051	0.434569	0.382397	0.611644
293	บ033	-0.517364	0.427905	0.478761	0.521239
294	บ035	-0.566144	0.457577	0.498828	0.501172
295	บ039	-0.571796	0.457577	0.441375	0.558625
296	บ042	-0.644969	0.443995	0.426866	0.573134
297	บ043	-0.570956	0.440945	0.417015	0.548919
298	บ046	-0.545785	0.476380	0.440455	0.559545
299	บ047	-0.485538	0.417086	0.339600	0.660400
300	บ065	-0.584182	0.517364	0.470794	0.512266
301	บ067	-0.656635	0.517364	0.494192	0.487353
302	บ070	-0.552867	0.476380	0.400255	0.599745
303	บ073	-0.584182	0.476380	0.452829	0.538044
304	บ074	-0.571796	0.457577	0.438329	0.554791
305	บ075	-0.612194	0.476380	0.502162	0.497838
306	บ076	-0.584182	0.584182	0.449879	0.535311
307	บ078	-0.637535	0.552867	0.451549	0.526732
308	บ079	-0.607347	0.476380	0.468693	0.499599
309	บ080	-0.571796	0.476380	0.465711	0.526992
310	บ082	-0.584182	0.476380	0.466681	0.514846
311	บ085	-0.501330	0.545145	0.446475	0.527123
312	บ088	-0.584182	0.517364	0.478121	0.505678
313	บ091	-0.552867	0.517364	0.450032	0.543016
314	บ097	-0.438858	0.427905	0.361493	0.638507
315	บ099	-0.512880	0.440587	0.455933	0.544067
316	บ104	-0.515800	0.417086	0.312144	0.687856
317	บ106	-0.584182	0.427905	0.225079	0.774921
318	บ111	-0.535466	0.441595	0.380005	0.619995
319	บ112	-0.603025	0.416749	0.517156	0.482844
320	บ113	-0.432957	0.423360	0.332182	0.667818
321	บ114	-0.491635	0.446110	0.420770	0.579230
322	บ115	-0.523403	0.446940	0.459863	0.540137
323	บ117	-0.512972	0.440353	0.415678	0.584322
324	ป003	-0.184289	0.476380	0.246709	0.753291
325	ป004	-0.487290	0.476380	0.346334	0.636628
326	ป012	-0.493940	0.444908	0.365976	0.634024
327	ป015	-0.215761	0.440945	0.164193	0.822029
328	ป020	-0.571796	0.476380	0.318950	0.681050
329	ป023	-0.538891	0.457577	0.325644	0.666435
330	ป025	-0.584182	0.476380	0.395483	0.604517
331	ป029	-0.532486	0.439812	0.304463	0.695537
332	ป032	-0.566144	0.448843	0.404352	0.595648
333	ป035	-0.509878	0.478460	0.449842	0.543282
334	ป041	-0.558804	0.484267	0.266422	0.718334

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
51	ข094	-0.552867	0.427905	0.608645	0.380993
52	ข095	-0.719996	0.476380	0.438149	0.504522
53	ข097	-0.584182	0.476380	0.592132	0.382466
54	ข099	-0.584182	0.517364	0.400863	0.599137
55	ข101	-0.552867	0.476380	0.317084	0.682916
56	ข103	-0.538891	0.457577	0.480764	0.519236
57	ข105	-0.552867	0.476380	0.410098	0.589902
58	ข106	-0.584182	0.552866	0.415597	0.574454
59	ข107	-0.517364	0.427905	0.620649	0.379351
60	ข108	-0.552867	0.476380	0.296924	0.703076
61	ข109	-0.538891	0.457577	0.352246	0.647754
62	ข112	-0.538891	0.538891	0.423719	0.557884
63	ข113	-0.552867	0.476380	0.330380	0.669620
64	ข114	-0.552867	0.476380	0.521880	0.478120
65	ข116	-0.552867	0.552867	0.481533	0.509249
66	ข118	-0.601073	0.457577	0.444170	0.555830
67	ข119	-0.517364	0.517364	0.567512	0.416878
68	ข122	-0.601073	0.457577	0.383246	0.598561
69	ข124	-0.557193	0.434794	0.447370	0.546208
70	ข126	-0.575580	0.412215	0.492218	0.507782
71	ข127	-0.541498	0.479674	0.546871	0.453129
72	ข128	-0.584040	0.437799	0.474175	0.525825
73	ข129	-0.566776	0.428967	0.442992	0.557008
74	ข130	-0.578313	0.401372	0.248141	0.751859
75	ข131	-0.566406	0.421607	0.489578	0.510422
76	ข132	-0.549677	0.420833	0.552549	0.447451
77	ข134	-0.550035	0.427905	0.478628	0.521372
78	ข136	-0.497665	0.402572	0.287026	0.712974
79	ข137	-0.557051	0.402045	0.475679	0.524321
80	ข139	-0.522388	0.402433	0.432665	0.567335
81	ค090	-0.601073	0.184289	0.929658	0.054865
82	ค091	-0.581579	0.134909	0.965917	0.030322
83	ฆ130	-0.801987	0.362198	0.744133	0.245882
84	ง018	-0.624342	0.240138	0.872033	0.109290
85	ง039	-0.681950	0.148584	0.873068	0.126932
86	ง047	-0.578065	0.131036	0.867099	0.115470
87	ง060	-0.664642	0.102329	0.755774	0.244226
88	ง076	-0.782846	0.292091	0.685214	0.314786
89	ง084	-0.772157	0.134909	0.973096	0.026904
90	ง109	-0.566144	0.517364	0.697486	0.276877
91	ง116	-0.637535	0.292091	0.674308	0.306299
92	ง117	-0.552867	0.566144	0.604237	0.355743
93	ง122	-0.637535	0.292091	0.752924	0.247076
94	ง125	-0.690199	0.184289	0.789751	0.193808
95	ง129	-0.572071	0.271833	0.767363	0.232637
96	ง131	-0.603492	0.207283	0.923945	0.076055
97	ง134	-0.580790	0.177948	0.758412	0.241588
98	ง141	-0.688254	0.128892	0.835871	0.164129
99	ง144	-0.447010	0.240434	0.911815	0.088185
100	ง145	-0.662911	0.144434	0.884294	0.115706
101	ง146	-0.630280	0.233832	0.979829	0.020171
102	จ001	-0.577896	0.577896	0.297348	0.689103
103	จ002	-0.552867	0.552867	0.202720	0.761784

no.	ch	feature1	feature2	feature3	feature4
336	ป044	-0.524832	0.454456	0.255938	0.744062
337	ป047	-0.476380	0.483305	0.329882	0.640438
338	ป058	-0.405177	0.457577	0.122385	0.870159
339	ป066	-0.612194	0.584182	0.444827	0.520966
340	ป070	-0.601073	0.517364	0.366674	0.620902
341	ป071	-0.627444	0.457577	0.371843	0.621321
342	ป074	-0.622869	0.517364	0.356473	0.627545
343	ป080	-0.637535	0.552867	0.402156	0.597844
344	ป081	-0.660669	0.544380	0.417150	0.567283
345	ป082	-0.612194	0.571796	0.409016	0.583946
346	ป084	-0.681950	0.292091	0.569364	0.402862
347	ป085	-0.627444	0.457577	0.383438	0.585357
348	ป089	-0.476380	0.476380	0.328933	0.671067
349	ป090	-0.612194	0.448843	0.367456	0.632544
350	ป093	-0.660669	0.517364	0.446316	0.543171
351	ป096	-0.576651	0.544380	0.330857	0.655767
352	ป097	-0.719996	0.326123	0.400883	0.599118
353	ป098	-0.442313	0.433053	0.237625	0.762375
354	ป101	-0.576797	0.417086	0.258981	0.741019
355	ป104	-0.552867	0.432142	0.443455	0.556545
356	ป105	-0.548468	0.437343	0.405955	0.594045
357	ป106	-0.593240	0.412672	0.356566	0.643434
358	ป107	-0.512375	0.446110	0.412858	0.587142
359	ป108	-0.525419	0.443588	0.376043	0.623957
360	ป110	-0.552867	0.437343	0.417973	0.582027
361	ฬ010	-0.713336	0.378008	0.625184	0.374816
362	ฬ036	-0.744932	0.398406	0.714017	0.272475
363	ฬ053	-0.394501	0.727160	0.224671	0.771475
364	ฬ055	-0.637535	0.753274	0.148360	0.784900
365	ฬ058	-0.427905	0.709321	0.236596	0.754007
366	ฬ059	-0.493940	0.719996	0.122073	0.848452
367	ฬ062	-0.589466	0.712634	0.186618	0.773878
368	ฬ063	-0.601073	0.730259	0.196738	0.795333
369	ฬ095	-0.737155	0.427905	0.763875	0.217827
370	ฬ115	-0.647256	0.418121	0.663741	0.336259
371	ฬ068	-0.601073	0.715906	0.185375	0.748888
372	ฬ069	-0.538891	0.712634	0.195881	0.793627
373	ฬ072	-0.762350	0.427905	0.606269	0.341885
374	ฬ073	-0.753274	0.427905	0.650242	0.349758
375	ฬ081	-0.753274	0.476380	0.609686	0.382485
376	ฬ112	-0.612194	0.427905	0.528727	0.471273
377	ม001	-0.737155	0.368578	0.771071	0.228929
378	ม002	-0.704052	0.376882	0.611599	0.388401
379	ม003	-0.708014	0.390206	0.719056	0.280944
380	ม004	-0.734136	0.394501	0.714291	0.281555
381	ม006	-0.727160	0.394501	0.721893	0.252225
382	ม007	-0.707512	0.463747	0.753824	0.219965
383	ม008	-0.692478	0.376142	0.630151	0.361383
384	ม009	-0.568027	0.405776	0.586526	0.413474
385	ม010	-0.555806	0.377781	0.540509	0.459491
386	ม011	-0.754798	0.374997	0.727934	0.272066
387	ม012	-0.576121	0.377781	0.628341	0.371659
388	ม014	-0.737155	0.368578	0.597392	0.402608

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
104	৭003	-0.612194	0.545949	0.412695	0.563287
105	৭005	-0.552867	0.601073	0.373763	0.615620
106	৭007	-0.607347	0.556362	0.439954	0.538338
107	৭008	-0.584182	0.584182	0.377939	0.614507
108	৭010	-0.584182	0.637535	0.390285	0.597662
109	৭011	-0.566144	0.544380	0.526342	0.473658
110	৭012	-0.566144	0.582786	0.424924	0.567751
111	৭013	-0.584182	0.533286	0.367813	0.603729
112	৭014	-0.532486	0.584182	0.263034	0.720696
113	৭019	-0.637535	0.612194	0.370974	0.611271
114	৭020	-0.566144	0.607347	0.284941	0.699402
115	৭023	-0.538891	0.552867	0.232757	0.756019
116	৭024	-0.598764	0.556095	0.265152	0.723839
117	৭026	-0.584182	0.584182	0.387947	0.588376
118	৭029	-0.612194	0.584182	0.473869	0.518528
119	৭030	-0.532486	0.561091	0.405797	0.573782
120	৭031	-0.517364	0.584182	0.426114	0.573886
121	৭033	-0.584182	0.563036	0.448410	0.551590
122	৭034	-0.629887	0.547121	0.422070	0.572016
123	৭035	-0.612194	0.584182	0.484006	0.503587
124	৭036	-0.584182	0.571796	0.226711	0.773289
125	৭037	-0.678229	0.552867	0.449949	0.516557
126	৭039	-0.538891	0.594012	0.379836	0.588288
127	৭042	-0.544380	0.566144	0.467551	0.527817
128	৭043	-0.612194	0.552867	0.489814	0.503925
129	৭045	-0.584182	0.569470	0.398722	0.601278
130	৭049	-0.584182	0.542488	0.541923	0.458077
131	৭051	-0.696837	0.401053	0.663382	0.336618
132	৭053	-0.767425	0.479051	0.636800	0.359437
133	৭055	-0.678229	0.513313	0.589095	0.394518
134	৭065	-0.681950	0.612194	0.308343	0.691657
135	৭066	-0.681950	0.457577	0.342107	0.657893
136	৭072	-0.762350	0.457577	0.460390	0.539610
137	৭073	-0.782846	0.493940	0.574920	0.421739
138	৭074	-0.731462	0.620727	0.361005	0.621273
139	৭075	-0.712634	0.594449	0.349321	0.639564
140	৭076	-0.517364	0.639887	0.274961	0.713824
141	৭077	-0.612194	0.571796	0.277512	0.696694
142	৭081	-0.427905	0.651433	0.206891	0.793109
143	৭082	-0.681950	0.487290	0.442923	0.523716
144	৭084	-0.552867	0.633132	0.235614	0.728260
145	৭085	-0.759605	0.448843	0.376656	0.623344
146	৭086	-0.668075	0.533456	0.367692	0.632308
147	৭087	-0.666180	0.561962	0.371406	0.573416
148	৭088	-0.549598	0.608025	0.355074	0.644926
149	৭089	-0.427905	0.609583	0.170736	0.829264
150	৭106	-0.612194	0.584182	0.393507	0.585810
151	৭107	-0.517364	0.603769	0.421496	0.563968
152	৭108	-0.612194	0.603769	0.401157	0.532986
153	৭113	-0.517364	0.552866	0.381203	0.618797
154	৭116	-0.561091	0.614781	0.294627	0.649555
155	৭117	-0.601073	0.584182	0.465622	0.511269

no.	ch	feature1	feature2	feature3	feature4
389	৯015	-0.596017	0.427905	0.719528	0.280472
390	৯016	-0.724414	0.379762	0.682855	0.313296
391	৯017	-0.618794	0.384779	0.633095	0.366905
392	৯018	-0.587054	0.374997	0.590991	0.409009
393	৯019	-0.709321	0.383379	0.776462	0.223538
394	৯020	-0.719996	0.368578	0.767172	0.232828
395	৯021	-0.723179	0.384148	0.802858	0.181315
396	৯022	-0.706384	0.377461	0.741702	0.258298
397	৯023	-0.707081	0.372063	0.674337	0.325663
398	৯024	-0.699369	0.354557	0.721168	0.278832
399	৯026	-0.681950	0.368578	0.742446	0.257554
400	৯027	-0.639575	0.374152	0.627151	0.359183
401	৯028	-0.696837	0.368578	0.616837	0.383163
402	৯031	-0.681950	0.368578	0.759787	0.227325
403	৯032	-0.701653	0.368578	0.720695	0.279305
404	৯033	-0.730259	0.368578	0.743202	0.241416
405	৯035	-0.678229	0.368578	0.766158	0.233842
406	৯036	-0.701653	0.368578	0.730932	0.269068
407	৯037	-0.683186	0.351324	0.685298	0.314702
408	৯041	-0.719996	0.368578	0.751991	0.239242
409	৯042	-0.716775	0.356289	0.743050	0.256950
410	৯043	-0.695820	0.421695	0.778260	0.221740
411	৯044	-0.701653	0.427905	0.661492	0.338508
412	৯046	-0.701653	0.368578	0.772521	0.227479
413	৯048	-0.718258	0.388542	0.734497	0.265503
414	৯049	-0.715504	0.378899	0.736470	0.263530
415	৯050	-0.616955	0.339850	0.692394	0.307606
416	৯051	-0.724414	0.313481	0.702383	0.278913
417	৯053	-0.685168	0.378899	0.748044	0.227603
418	৯054	-0.747918	0.368578	0.696485	0.239947
419	৯056	-0.552867	0.368578	0.692758	0.307242
420	৯061	-0.724414	0.313481	0.771839	0.228161
421	৯064	-0.758523	0.351324	0.829735	0.153353
422	৯065	-0.715504	0.351324	0.789203	0.199935
423	৯066	-0.705062	0.339850	0.723767	0.257717
424	৯070	-0.709321	0.326123	0.760661	0.221665
425	৯071	-0.730259	0.339850	0.717179	0.282821
426	৯072	-0.696837	0.351324	0.722955	0.277045
427	৯073	-0.719996	0.368578	0.818596	0.181404
428	৯074	-0.753274	0.368578	0.719605	0.273123
429	৯075	-0.719996	0.427905	0.782586	0.194569
430	৯076	-0.732946	0.351324	0.733276	0.266724
431	৯077	-0.743877	0.394501	0.794110	0.205890
432	৯078	-0.737155	0.427905	0.785406	0.214594
433	৯079	-0.746788	0.339850	0.781508	0.200489
434	৯081	-0.730259	0.339850	0.693251	0.306749
435	৯084	-0.753274	0.427905	0.750920	0.249080
436	৯085	-0.732946	0.351324	0.720781	0.279219
437	৯086	-0.732946	0.351324	0.718520	0.281480
438	৯087	-0.758523	0.405177	0.719431	0.258741
439	৯089	-0.727160	0.394501	0.713816	0.276017
440	৯090	-0.715504	0.465003	0.702710	0.274164

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
156	จ118	-0.552867	0.605427	0.285653	0.677149
157	จ124	-0.496216	0.548142	0.334097	0.665903
158	จ125	-0.170966	0.571796	0.215237	0.784763
159	จ128	-0.496328	0.557988	0.292212	0.707788
160	จ129	-0.480055	0.596017	0.343694	0.656306
161	จ131	-0.247669	0.563901	0.113999	0.886001
162	จ132	-0.199832	0.561540	0.133706	0.866294
163	จ134	-0.243900	0.581579	0.073968	0.926032
164	จ135	-0.446110	0.561962	0.269699	0.730301
165	จ137	-0.228850	0.566071	0.107045	0.892955
166	จ139	-0.518633	0.572830	0.306122	0.693878
167	จ140	-0.125981	0.557193	0.222338	0.777662
168	จ143	-0.470211	0.570229	0.221623	0.778377
169	จ144	-0.215761	0.577896	0.257246	0.742754
170	จ145	-0.207283	0.578377	0.250404	0.749596
171	จ146	-0.450955	0.588312	0.253903	0.746097
172	จ147	-0.146069	0.580406	0.174490	0.825510
173	จ148	-0.460480	0.578659	0.290719	0.709281
174	จ151	-0.504647	0.572169	0.307334	0.692666
175	จ152	-0.481837	0.580406	0.313101	0.686899
176	จ153	-0.221629	0.615389	0.230053	0.769947
177	จ154	-0.223156	0.583860	0.291935	0.708065
178	จ155	-0.279765	0.615246	0.209588	0.790412
179	ฉ050	-0.824162	0.468021	0.677790	0.318273
180	ฉ055	-0.806254	0.458405	0.493877	0.478308
181	ฉ056	-0.854122	0.457577	0.661341	0.301726
182	ฉ057	-0.872208	0.468251	0.616192	0.361352
183	ฉ059	-0.753274	0.507643	0.645325	0.354675
184	ฉ060	-0.759605	0.463747	0.657904	0.342096
185	ฉ063	-0.937563	0.422656	0.718208	0.268464
186	ฉ065	-0.879000	0.419350	0.595189	0.384722
187	ฉ069	-0.766965	0.525233	0.502357	0.475354
188	ฉ070	-0.804513	0.510170	0.706695	0.290282
189	ฉ072	-0.713244	0.476380	0.595304	0.391077
190	ฉ073	-0.731921	0.462186	0.553857	0.430413
191	ฉ074	-0.780305	0.394501	0.564590	0.405791
192	ฉ075	-0.729871	0.526180	0.531316	0.461217
193	ฉ076	-0.861238	0.453995	0.657101	0.337584
194	ฉ077	-0.807920	0.469269	0.550754	0.443933
195	ฉ078	-0.775028	0.495585	0.549299	0.448652
196	ฉ080	-0.858235	0.483305	0.646246	0.327122
197	ฉ082	-0.872813	0.394990	0.673969	0.310972
198	ฉ083	-0.846951	0.460195	0.709824	0.283188
199	ฉ095	-0.833642	0.427905	0.803222	0.170313
200	ฉ097	-0.796483	0.476380	0.760120	0.231440
201	ฉ105	-0.830721	0.427905	0.726507	0.263824
202	ฉ111	-0.782846	0.493940	0.709530	0.284815
203	ฉ112	-0.821823	0.356289	0.725571	0.197938
204	ฉ113	-0.809455	0.427905	0.794292	0.205708
205	ฉ119	-0.686676	0.528225	0.549931	0.450069
206	ฉ120	-0.598713	0.601073	0.509194	0.490806
207	ฉ123	-0.551800	0.595025	0.386404	0.613596
208	ฉ131	-0.557559	0.603025	0.466184	0.533816

no.	ch	feature1	feature2	feature3	feature4
441	ม091	-0.701653	0.368578	0.678332	0.321668
442	ม093	-0.715504	0.507643	0.732694	0.256704
443	ม095	-0.730259	0.339850	0.771093	0.228907
444	ม096	-0.725938	0.305765	0.577922	0.400651
445	ม097	-0.732946	0.351324	0.751465	0.248535
446	ม099	-0.693757	0.405177	0.714396	0.285604
447	ม105	-0.658099	0.326123	0.600817	0.399183
448	ม106	-0.573063	0.357840	0.533261	0.466739
449	ม107	-0.569914	0.335326	0.538485	0.461516
450	ม108	-0.546479	0.345032	0.552955	0.447045
451	ม109	-0.568053	0.372574	0.597968	0.402032
452	ม110	-0.553846	0.335200	0.561450	0.438550
453	ม111	-0.739773	0.320555	0.644252	0.355748
454	ม114	-0.749314	0.335976	0.840043	0.159957
455	ม115	-0.556834	0.257679	0.534738	0.465262
456	ม118	-0.573120	0.344638	0.544216	0.455784
457	ม119	-0.550508	0.292091	0.452922	0.547078
458	ม120	-0.572092	0.348264	0.601658	0.398342
459	ม122	-0.565607	0.341544	0.555042	0.444958
460	ล001	-0.618794	0.553608	0.470362	0.529638
461	ล003	-0.603025	0.553849	0.459149	0.540851
462	ล004	-0.476380	0.562294	0.406590	0.564914
463	ล005	-0.457577	0.538891	0.459035	0.540965
464	ล006	-0.462186	0.565083	0.425911	0.574089
465	ล007	-0.541354	0.566144	0.341883	0.658117
466	ล009	-0.719996	0.552867	0.579430	0.414059
467	ล011	-0.616955	0.530746	0.400338	0.599662
468	ล012	-0.709321	0.603025	0.508202	0.466735
469	ล014	-0.785014	0.519552	0.551915	0.392337
470	ล016	-0.591505	0.552867	0.459123	0.540877
471	ล018	-0.681950	0.563901	0.501237	0.481653
472	ล019	-0.660669	0.569470	0.529927	0.451911
473	ล022	-0.622869	0.582248	0.530511	0.458054
474	ล024	-0.517364	0.584182	0.417292	0.582708
475	ล025	-0.517364	0.596017	0.434533	0.565467
476	ล026	-0.603025	0.561091	0.379806	0.620194
477	ล027	-0.627444	0.558804	0.514605	0.477795
478	ล028	-0.681950	0.584182	0.541769	0.433925
479	ล029	-0.637535	0.637535	0.505562	0.463716
480	ล030	-0.584182	0.550508	0.496566	0.503434
481	ล031	-0.637535	0.594729	0.516972	0.474472
482	ล036	-0.727160	0.562578	0.472262	0.456440
483	ล039	-0.618794	0.563901	0.532221	0.429374
484	ล040	-0.612194	0.561091	0.382683	0.617317
485	ล041	-0.612194	0.561091	0.572741	0.427259
486	ล042	-0.609107	0.538891	0.499520	0.478582
487	ล043	-0.618794	0.546793	0.583460	0.416540
488	ล049	-0.612194	0.534172	0.531079	0.468921
489	ล052	-0.612194	0.637535	0.368281	0.611012
490	ล061	-0.819403	0.422445	0.576943	0.369659
491	ล063	-0.785567	0.382849	0.700209	0.256637
492	ล064	-0.753274	0.432375	0.665430	0.317994
493	ล077	-0.612194	0.569470	0.414473	0.573194

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
209	จ134	-0.568027	0.440887	0.403375	0.596625
210	จ142	-0.552522	0.398579	0.446374	0.553626
211	ช007	-0.552867	0.517364	0.529136	0.455482
212	ช009	-0.566144	0.681950	0.341379	0.639906
213	ช012	-0.566144	0.681950	0.308540	0.671640
214	ช021	-0.566144	0.544380	0.647480	0.345129
215	ช047	-0.552867	0.427905	0.646770	0.338670
216	ช056	-0.591505	0.215761	0.904119	0.089808
217	ช011	-0.552867	0.737155	0.189191	0.793438
218	ช014	-0.552867	0.783439	0.129096	0.870904
219	ฌ056	-0.908715	0.300950	0.717480	0.247239
220	ฌ059	-0.957084	0.363839	0.816212	0.180383
221	ฌ067	-0.959180	0.339850	0.769588	0.214652
222	ฌ070	-0.945270	0.346992	0.742038	0.257963
223	ฌ074	-0.695136	0.329305	0.676064	0.306910
224	ฌ075	-0.953114	0.353806	0.750242	0.234365
225	ฌ081	-0.752850	0.345032	0.772979	0.222543
226	ฌ109	-0.905446	0.388542	0.810454	0.133913
227	ฌ117	-0.730856	0.313889	0.710729	0.289271
228	ฌ119	-0.777439	0.289347	0.770437	0.229563
229	ฌ120	-0.750433	0.311648	0.780474	0.219526
230	ฌ121	-0.651244	0.300950	0.708879	0.291121
231	ฌ127	-0.637535	0.233832	0.722570	0.277430
232	ฌ130	-0.764732	0.351324	0.754025	0.245975
233	ฌ131	-0.830012	0.326123	0.813105	0.183396
234	ฌ132	-0.646820	0.310102	0.703780	0.296220
235	ฌ134	-0.660669	0.339850	0.736671	0.263329
236	ฌ135	-0.641866	0.307639	0.783817	0.216183
237	ฌ137	-0.665785	0.311990	0.765233	0.234767
238	ฌ138	-0.650040	0.282801	0.732586	0.267414
239	ฌ139	-0.653371	0.339850	0.808962	0.191038
240	ฌ140	-0.644969	0.256470	0.808352	0.191648
241	ฌ141	-0.641866	0.241960	0.782108	0.217892
242	ฌ142	-0.649866	0.310102	0.698193	0.301807
243	ญ003	-0.492691	0.385582	0.476711	0.523289
244	ญ004	-0.516054	0.425387	0.569079	0.430921
245	ญ005	-0.500850	0.414777	0.535023	0.464977
246	ญ007	-0.513625	0.443995	0.614756	0.385244
247	ญ008	-0.536749	0.483305	0.544603	0.455397
248	ญ009	-0.503653	0.420833	0.455862	0.544138
249	ญ001	-0.648506	0.195147	0.766739	0.233261
250	ญ014	-0.665287	0.233832	0.792040	0.207960
251	ฏ069	-0.612194	0.566144	0.580431	0.419569
252	ฏ075	-0.622869	0.571796	0.578206	0.399570
253	ฏ079	-0.596017	0.566144	0.568122	0.426370
254	ฏ116	-0.598122	0.535157	0.656273	0.343727
255	ฏ119	-0.600924	0.570562	0.574564	0.425436
256	ฏ120	-0.570956	0.528955	0.593525	0.406475
257	ฏ121	-0.572710	0.534964	0.593287	0.391363
258	ฏ122	-0.442313	0.535970	0.417879	0.555178
259	ฏ126	-0.545823	0.531813	0.665218	0.334782
260	ฒ021	-0.852833	0.388542	0.777028	0.206377

no.	ch	feature1	feature2	feature3	feature4
494	ล078	-0.647256	0.625354	0.368166	0.597517
495	ล080	-0.660669	0.552867	0.498056	0.501944
496	ล082	-0.674924	0.517364	0.471312	0.504608
497	ล096	-0.681950	0.566144	0.505595	0.479946
498	ล097	-0.618794	0.620620	0.362829	0.607842
499	ล100	-0.579463	0.561091	0.384656	0.615344
500	ล102	-0.416482	0.565207	0.317583	0.682417
501	ล103	-0.664642	0.538891	0.422662	0.577338
502	ล105	-0.411366	0.562294	0.369599	0.630401
503	ล106	-0.425218	0.578225	0.301247	0.698753
504	ล107	-0.407402	0.572987	0.388327	0.601188
505	ล108	-0.394501	0.566144	0.316389	0.683611
506	ล109	-0.422371	0.569134	0.334117	0.665883
507	ล110	-0.404618	0.566144	0.320484	0.679516
508	ว004	-0.233832	0.666991	0.047571	0.952429
509	ว029	-0.222822	0.671364	0.073347	0.926653
510	ว030	-0.184289	0.628186	0.101619	0.889555
511	ว032	-0.184289	0.660669	0.078502	0.921498
512	ว047	-0.253054	0.690199	0.020435	0.979565
513	ว050	-0.091288	0.589466	0.063107	0.936893
514	ว062	-0.292091	0.709321	0.052077	0.947923
515	ส001	-0.598631	0.653205	0.280521	0.719479
516	ส002	-0.418121	0.649524	0.315184	0.684816
517	ส003	-0.469269	0.668075	0.329864	0.670136
518	ส004	-0.490819	0.678229	0.263788	0.736212
519	ส005	-0.424664	0.670173	0.301786	0.698214
520	ส006	-0.466895	0.677614	0.309630	0.690370
521	ส007	-0.497889	0.676465	0.358494	0.641506
522	ส008	-0.422556	0.675749	0.334657	0.665343
523	ส009	-0.421070	0.672820	0.297656	0.702344
524	ส010	-0.405810	0.672964	0.264524	0.735476
525	ส011	-0.415441	0.675493	0.370574	0.629426
526	ส012	-0.476380	0.670602	0.382242	0.617758
527	ส013	-0.148584	0.675493	0.150265	0.849735
528	ส014	-0.427905	0.676759	0.329146	0.670854
529	ส016	-0.386103	0.681950	0.294575	0.705424
530	ส017	-0.383523	0.661281	0.280341	0.719659
531	ส018	-0.418568	0.671852	0.287605	0.706411
532	ส019	-0.426327	0.671050	0.317582	0.682418
533	ส020	-0.434488	0.681344	0.334337	0.665663
534	ส022	-0.398760	0.679052	0.289354	0.710646
535	ส023	-0.405177	0.677614	0.276711	0.715978
536	ส024	-0.426461	0.669594	0.275247	0.724753
537	ส025	-0.422420	0.663095	0.319638	0.680362
538	ส026	-0.427905	0.660669	0.341719	0.658281
539	ห020	-0.584182	0.641334	0.464387	0.518348
540	ห057	-0.538891	0.601073	0.288573	0.707869
541	ห064	-0.184289	0.660669	0.032122	0.946778
542	ห100	-0.546793	0.601073	0.254218	0.745782
543	ห102	-0.251830	0.676977	0.227065	0.772935
544	ห103	-0.161168	0.678521	0.168593	0.831407
545	ห104	-0.454316	0.620360	0.402094	0.597906

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.1 (ต่อ)

no.	ch	feature1	feature2	feature3	feature4
261	ฌ024	-0.852181	0.384779	0.767636	0.213560
262	ฌ028	-0.847340	0.374152	0.797514	0.191809
263	ฌ111	-0.726887	0.264646	0.797083	0.198084
264	ฌ114	-0.876273	0.184289	0.835531	0.164469
265	ค051	-0.912557	0.148584	0.909063	0.074609
266	ค054	-0.817695	0.134909	0.932328	0.067672
267	ค056	-0.690199	0.233832	0.953231	0.039167
268	ค066	-0.698200	0.159117	0.978269	0.021731
269	ค112	-0.655041	0.184289	0.921258	0.078742
270	ค116	-0.587904	0.184289	0.966285	0.033715
271	ถ056	-0.806060	0.190263	0.734995	0.242637
272	ท078	-0.728841	0.176222	0.899988	0.100012
273	ท099	-0.737155	0.184289	0.975431	0.024569
274	ท117	-0.692444	0.286090	0.898236	0.101764
275	ท119	-0.762350	0.162361	0.899143	0.100857
276	ท130	-0.710881	0.152234	0.945924	0.054076
277	ท134	-0.584595	0.145919	0.963620	0.036380
278	ท139	-0.680620	0.147488	0.966098	0.033902
279	น016	-0.787137	0.132847	0.863555	0.118964
280	บ003	-0.292091	0.457577	0.288745	0.711255
281	บ004	-0.427905	0.476380	0.342092	0.657907
282	บ005	-0.538891	0.476380	0.455808	0.544192
283	บ009	-0.535157	0.435745	0.433066	0.566934
284	บ010	-0.601073	0.457577	0.474847	0.525153
285	บ019	-0.123350	0.457577	0.228399	0.771601

no.	ch	feature1	feature2	feature3	feature4
546	ห110	-0.491353	0.491353	0.462095	0.537905
547	ห118	-0.612980	0.584182	0.461880	0.538120
548	ห119	-0.150906	0.594449	0.147033	0.852967
549	ห120	-0.623612	0.593240	0.432957	0.563294
550	ห121	-0.642078	0.589724	0.311711	0.688289
551	ห122	-0.526743	0.591505	0.223120	0.776880
552	ห018	-0.704591	0.638929	0.607496	0.392504
553	ห042	-0.584182	0.532486	0.555664	0.433404
554	ห056	-0.558804	0.800159	0.207224	0.755234
555	ห075	-0.743877	0.637535	0.503131	0.466537
556	ห076	-0.759605	0.690199	0.415150	0.549187
557	ห081	-0.771832	0.675445	0.454173	0.479481
558	ห085	-0.768471	0.673436	0.450278	0.499193
559	ห091	-0.601073	0.815903	0.281089	0.707082
560	ห097	-0.782846	0.666180	0.524827	0.456162
561	ห100	-0.753274	0.674925	0.501403	0.486480
562	ห117	-0.741299	0.675493	0.426832	0.573168
563	ห120	-0.748883	0.672864	0.495886	0.504114
564	ห122	-0.747042	0.671192	0.525546	0.454499
565	อ079	-0.184289	0.757176	0.045699	0.909435
566	อ082	-0.292091	0.714409	0.038332	0.961668
567	ฮ011	-0.292091	0.627444	0.135477	0.864523
568	ฮ064	-0.292091	0.627444	0.135477	0.864523
569	ฮ150	-0.154755	0.768471	0.020721	0.971661
570	ฮ181	-0.154755	0.768471	0.020721	0.971661

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค
งานวิจัยที่ได้รับการตีพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ICCAS '04

ICCAS 2004

Final Program & Abstract Book

International Conference on Control, Automation and Systems

August 25-27, 2004

The Shangri-La Hotel, Bangkok, Thailand

ed by:

Center for Communications and Information Technology, King Mongkut's Institute of Technology Ladkrabang

ute of Control, Automation, and Systems Engineers, Korea

ICASE

In association with: SICE CAA AAI ECTI

ICCAS2004 August 25-27, The Shangri-La Hotel, Bangkok, THAILAND

The classified method for overlapping data

Boontee Kruatrachue*, Kulwarun Warunsin*, and Kritawan Siriboon*
*Computer Engineering Department, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang,
Chalongkrung Road, Bangkok 10520, Thailand
(Tel : +66-1-840-0405; E-mail: booontee@yahoo.com, kulwarun@hotmail.com, kritawan@ce.kmitl.ac.th)

Abstract: In this paper we introduce a new prototype based classifiers for overlapping data, where training pattern can be overlap on the feature space. The proposed classifier is based on the prototype from neural network classifier (NNC)[1] for overlap data. The method automatically chooses the initial center and two radiuses for each class. The center is used as a mean representative of training data for each class. The unclassified pattern is classified by measure distance from the class center. If the distance is in the lower (shorter radius) the unknown pattern has the high percentage of being in this class. If the distance is between the lower and upper (further radius), the pattern has the probability of being in this class or others. But if the distance is outside the upper, the pattern is not in this class. We borrow the words upper and lower from the rough set to represent the region of certainty [3]. The training algorithm to find number of cluster and their parameters (center, lower, upper) is presented. The clustering result is tested using patterns from Thai handwritten letter and the clustering result is very similar to human eyes clustering.

Keywords: overlapping data, classifier , supervised learning, Neural prototype

1. INTRODUCTION

A class is a collection of data object that are similar to one another within the same class and are dissimilar to the objects in other class. The process of grouping a set of physical or abstract objects into classes of similar objects is called classification. Classification is very close to clustering except that the classification is a supervised learning where classifier is trained with known class. Classification has wide applications including market or customer segmentation, pattern recognition, biological studies, spatial data analysis. Web document classification, and many others. Classification can be used as a standalone data mining tool to gain insight into the data distribution, or serve as a preprocessing step for other data mining algorithms operating on the detected clusters. Classification is a dynamic field of research in data mining. Many clustering algorithms have been develops. These can be categorized into hierarchical methods, partitioning methods, density-based methods, grid-based methods, and model-based methods [2].

The rest of the paper is organized as follows. In section 2 describe the NNC architecture. Section 3 we present our new prototype based classifiers for overlap data. Section 4, we report the experimental tested using patterns from Thai handwritten letter and the clustering result compare with human eyes clustering. Section 5 concludes with a summary and some directions for future research.

2. NNC ARCHITECTURE [1]

The neural network that implements the NNC is shown in Fig. 1. The input layer L_A=(a_1, a_2, a_3,..., a_n) consists of n neurons, one for each dimension of the input pattern P. Each input neuron is connected to all the N prototype neurons in layer L_B=(b_1, b_2, b_3,..., b_n). The connection weights from neurons in layer L_A to neurons in layer L_B are represented by:

$$W_1 = (w_1, w_2, ..., w_N)^T \quad (1)$$

Where $W_i = (w_{b_i a_1}, w_{b_i a_2}, ..., w_{b_i a_N})$ is the weight vector of the ith neuron in layer L_B.

A neuron in layer L_B representing only one pattern class. Each prototype neuron is added during learning, where there can be more than one neurons for a single class.

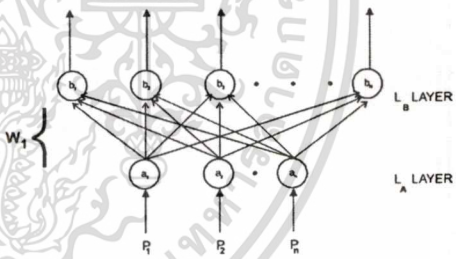


Fig. 1 NNC architecture.

In response to a pattern P_k the output of the NNC can be classified into one of the following class:

$$c_1, c_2, ...c_k, c_m \quad (2)$$

Where

c_k is the class of neuron in layer L_B of the NNC corresponding to class k,

m is the total number of classes in the training pattern.

If an input pattern P causes neuron b_k to fire, and b_k belong to class c_k. Then P is classified as class c_k:

In the context of this work, training the NNC means ensuring that its can correctly classify any training input pattern $P^k = \{p^k_1, p^k_2, ..., p^k_n\}$. This can be achieved by doing the following:

- creating new prototype neuron in layer L_B for any unclassified P,

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- adjusting the connection weights of the nearest neurons according to the Euclidean distance in equation (3), if the distance < that neuron threshold and the class of neuron is the same as the class of pattern P, Euclidean distance from P to neuron b_j is

$$b_{jT} = \sqrt{\sum_{s=1}^{n} (w_{b_j a_s} - p_s)^2} \qquad (3)$$

$w_{b_j a_s}$ is the connection weight between neuron b_j in L_B and neuron a_s in L_A

The weights are adjust to average the new pattern P into the nearest neurons as follow:

$$W_{b_j a_{inew}} = \frac{W_{b_j a_{iold}} + p_i}{M+1} \quad \forall i = 1,2,\ldots,n \qquad (4)$$

M is a popularity measure corresponding to the number of patterns

- For any neurons that has different class from pattern P and the distance between P and those neurons < that neurons threshold, reduce the firing threshold conditions of those neurons in order to exclude the pattern P from them.

A flowchart for training NNC is described in Fig. 2.

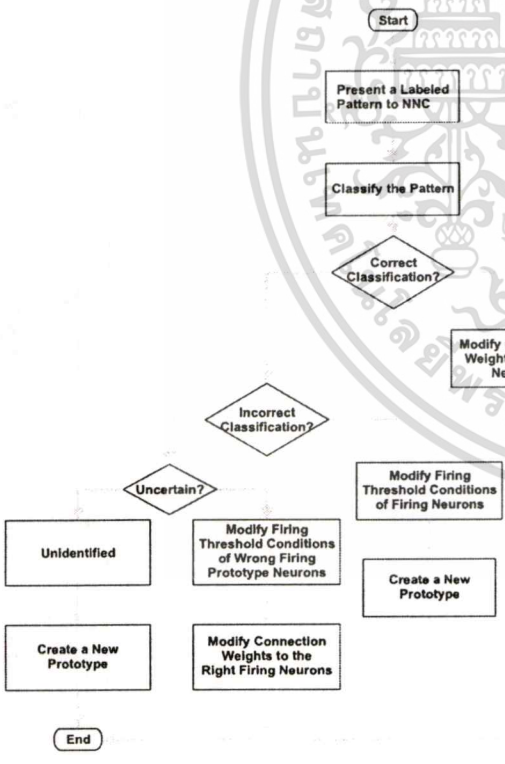


Fig. 2 NNC training algorithm.

3. PROTOTYPE BASED CLASSIFIERS FOR OVERLAP DATA

The NNC classifier described in previous section will keep reducing threshold and create new prototype neuron if the training Patterns has overlap as shown in Fig 3. These training loops start by the threshold reduction of wrong classified pattern. Then, any the pattern that used to be classified by that neuron become unclassified and finally new neuron is created for the wrong classified pattern. The final network will have large number of neuron, one for each training pattern in the highly overlap area.

In order to reduce number of neuron in the overlap area, each neuron has two thresholds (radius), shorter and the longer one as shown in Fig 3. The shorter radius is to define the perimeter of the neuron in the same way as the threshold in the NNC. The longer one is to define the conditional perimeter where any pattern located in this area can be the same or different class as this neuron. Hence the pattern in the shorter radius can be classified but the pattern between shorter and longer radius is uncertain but has some probability to be in the neuron class. But the pattern outside the longer radius is certainly not in the same cluster (neuron). These shorter and longer radiuses are analogous to the idea of lower and upper in the rough set [3].

From Fig. 3 A1 and A2 is area of class 1 and 2. B1 is area of probability data of class 1 and other, B2 is area of probability data of class 2 and other, C is area of data class 1 and class 2.

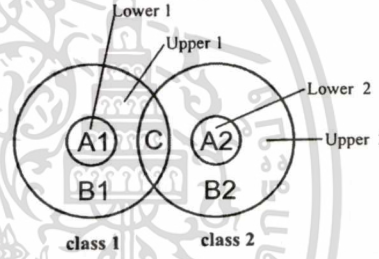


Fig. 3 area of lower and upper in class 1 and class 2.

The algorithm start by initial lower equal to upper radius and first data is the center of class (neuron). Then each training pattern is used to adjust the neuron parameter (center, lower, upper). The adjustment split into two cases depends on whether the pattern and the neuron have the different class (Fig. 4), or same class (Fig. 5). The distance used here is the same Euclidean distance as NNC.

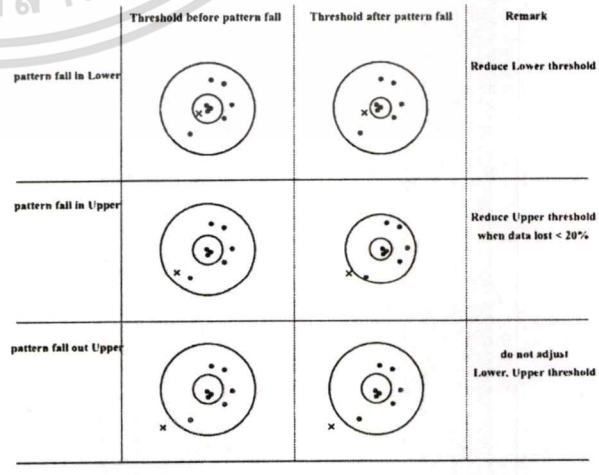


Fig. 4 adjustment of radius when training data is different class as the train cluster.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

For each training pattern, the distance is calculated for each neuron. For all neurons that have different class from the pattern, the radius adjustments depend on location of the pattern as shown in fig 4. If the pattern is in lower (shorter radius) then reduce the lower to exclude that pattern out of the lower region. If the pattern is between upper and lower, we don't need to change any radius. But in order to maintain high class probability of the upper region, the upper is reduce to exclude the wrong pattern if at least 80 % of the pattern that used to be in the upper still remain. Certainly, if the pattern is outside the upper, no change to the radius. The weight of the neuron is unchanged from the pattern of different class.

For the neuron that has the same class with the pattern, only the closet one is consider, others neurons are unchanged. The radius adjustment is as shown in fig 5. If the pattern is in the lower area no change to the radius since the correct pattern is already in the neuron lower area. If the pattern is between lower and upper, still the lower and upper is unchanged. Here the lower is not expanding to include the pattern because there may be other class pattern that will be in the lower. If the pattern is outside the upper area of the closet neuron with the same class, the upper is expanded to include that pattern if the distance is < 1.2 upper. Otherwise new neuron is constructed at that training pattern with the lower equal to the upper. And the upper extend to the center of the closet neuron.

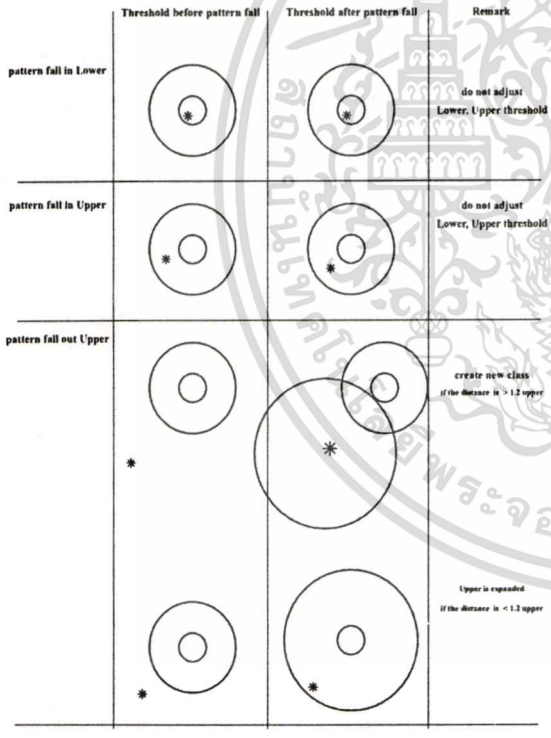


Fig. 5 adjustment of radius when training data is the same class as the train cluster.

As shown in Fig. 6, the distance between a training pattern and each neuron is calculated in order to reduce the lower or upper radius of each neuron with different class to the pattern if the wrong pattern is in the lower or upper region. For the neuron with the same pattern only the closet neuron weight is changed by average the new pattern with its weight. But if the pattern is not in any lower or upper region of any neuron with the same class, the new neuron node is created at the pattern. The training is repeated for each training pattern until all patterns are tested. If any neuron has radius change or new neurons are created, the training process is repeated until no change in the network.

```
CONSTANT SMALL  is value for adjust radius.

While ( (radius reduce) || (new neuron) ) {
  for each pattern p in the training patterns{
    read pattern p;
    for(node=0; node<total neuron; node++){
      o   compute distance = Euclidean distance (weight[node], p);
      o   if (class of p == class [node]){
            if(distance < lower[node])
                mark fall flag for adjust weight;
            else if (distance < upper[node]){
                mark fall flag for adjust weight;
                if (distance < (1.2 * upper[node]))
                  upper[node] = distance – SMALL;
                                              // expand upper
                else
                  clear fall flag for create new node
            }
          }// if (class of p == class [node])
      o   if (class of p != class [node]){
            if(distance < lower[node])
                lower[node] = distance - SMALL;
                                              //reduce lower
            else if (distance < upper[node]){
                if (reduce upper[node] at least 80 % of the pattern)
                    upper[node] = distance – SMALL;
            }
          }// if (class of p != class [node])
      }// for node
      if (pattern not fall in any node with the same class)
          create new node (p);
      else
          average pattern and weight
          in the nearest node with the same class;

    }// for each pattern p
}// While ( (radius reduce) || (new neuron) )
```

Fig. 6 Algorithm for overlap classifier.

4. EXAMPLE OF EXPERIMENT

We test new algorithm with clustering of 198 patterns from 6 letter THAI handwritten letter. Each pattern has only 2 dimension features in order to create some overlap as shown in Fig. 7. The clustering result of data into 6 clusters as shown in Fig. 9. It is hard to compare the classification results for overlap data since the classification accuracy depend on how much data overlap, especially for large number of class and patterns. Hence, in this paper we intend to use small number of classes (6 letter) and only 198 patterns in order to compare the clustering results with human approximate cluster.

As expected, the NNC create higher number of neurons in the overlap area. On the contrary overlap classifier create less number of neuron but with uncertainty ring region (between upper and lower).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

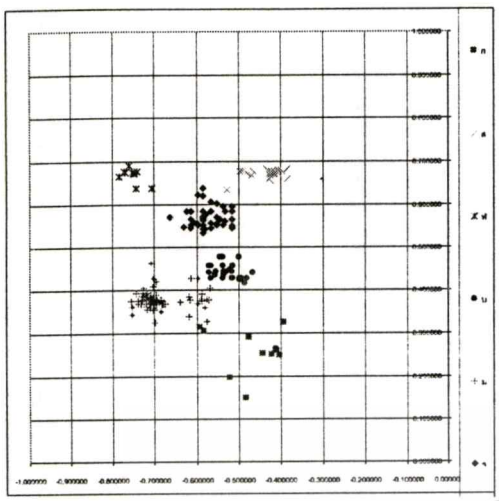


Fig. 7 Sample two feature THAI alphabet 198 data.

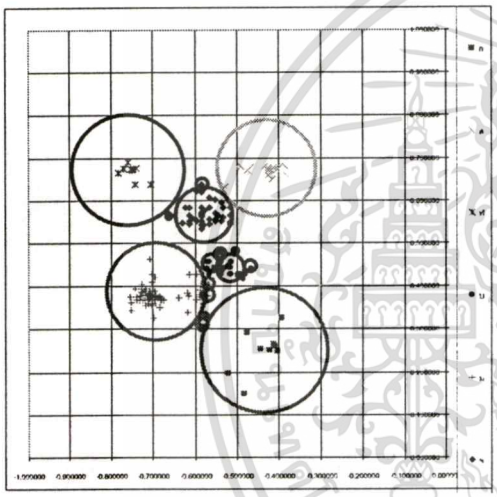


Fig. 8 Result is the NNC classifier described in section 2.

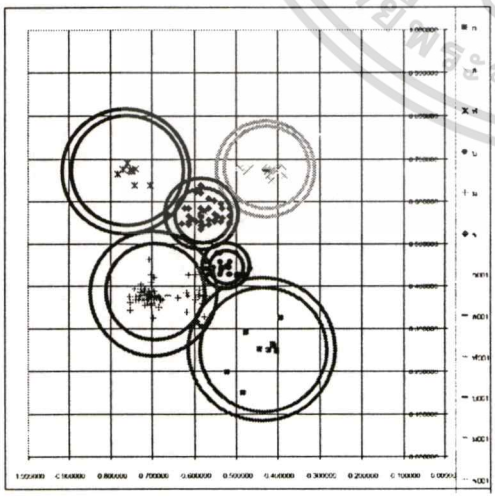


Fig. 9 Result of prototype based classifiers for overlap data.

5. CONCLUSIONS

In this paper, we propose a prototype based classifiers and algorithm for overlap data. Our algorithm is an extension of neural classifier, NNC [1] where the concept of class boundary is extend to have lower and upper boundary instead of just one radius to represent class boundary. This clustering has the same fast training time as neural classifier [1] which is much faster than normal gradient descent based back propagation neural classifier. Example has been provided that demonstrates the ability of overlap classifier in continuous classification problems with overlapping data. The technique used to set and modify the threshold conditions of its prototype is described in this paper. Future work improves algorithm for efficient and effective cluster analysis in large databases. Active themes for research focus on scalability of clustering methods, the effectiveness of methods for clustering high-dimensional techniques.

REFERENCES

[1] M.A. ABOU-NASR and M.A. SID-AHMED, Fast learning and efficient memory utilization with a prototype based neural classifier., Pattern Recognition, Vol.28, No.4, pp. 581-593, 1995.

[2] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann, pp. 335-393, 2001.

[3] Sarah Coppock, Lawrence Mazlack, "Rough Sets Used In The Measurement Of Similarity Of Mixed Mode Data," 22nd International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2003) Proceedings, Chicago, pp. 197-201, July, 2003.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล	กุลวลัญช์ วรุณสิน
วัน เดือน ปี	25 พฤษภาคม 2515 จังหวัดกรุงเทพมหานคร
ที่อยู่	103/24 โครงการบ้านสวนพงษ์เพชร ซอยแจ้งวัฒนะ 43
	ถ.แจ้งวัฒนะ ต.คลองเกลือ อ.ปากเกร็ด จ.นนทบุรี โทร. 0-1840-0405

ประวัติการศึกษา

พ.ศ. 2526 สำเร็จการศึกษาชั้นประถมศึกษาจากโรงเรียนดลวิทยา จังหวัดกรุงเทพมหานคร

พ.ศ. 2529 สำเร็จการศึกษาชั้นมัธยมศึกษาตอนต้นจากโรงเรียนดลวิทยา จังหวัดกรุงเทพมหานคร

พ.ศ. 2532 สำเร็จการศึกษาชั้นมัธยมศึกษาตอนปลายจากเตรียมอุดมศึกษาน้อมเกล้า จังหวัดกรุงเทพมหานคร

พ.ศ. 2536 วิทยาศาสตร์บัณฑิต สาขาฟิสิกส์ มหาวิทยาลัยศิลปากร จังหวัดนครปฐม

ประสบการณ์การทำงานและผลงานวิจัย

พ.ศ. 2536-2537	ตำแหน่ง Technical support บริษัท สหวิริยา โอเอ จำกัด - ในคำปรึกษาด้าน technique อุปกรณ์ต่อพ่วงกับคอมพิวเตอร์
พ.ศ. 2537-2544	ตำแหน่ง Researcher บริษัท อาร์ แอนด์ ดี คอมพิวเตอร์ ซิสเท็ม จำกัด - ให้คำปรึกษาเกี่ยวกับไวรัสคอมพิวเตอร์ การกู้ข้อมูล - พัฒนาการโปรแกรมป้องกันไวรัสคอมพิวเตอร์ลงใน hardware - วิทยากรให้การอบรมเกี่ยวกับไวรัสคอมพิวเตอร์ และการกู้ข้อมูลใน harddisk
พ.ศ. 2544-ปัจจุบัน	อาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยรามคำแหง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้