วิธีการตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยใช้ฟัซซิฟิเคชัน

นายวิฒเนศ ทองมี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2545
ISBN 974-17-1457-2
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

# AN APPROACH OF SOFT PRUNING FOR DECISION TREES USING FUZZIFICATION

Mr. Wikanes Thongmee

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University

Academic Year 2002

ISBN 974-17-1457-2

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | วิธีการตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยใช้ฟัซซิฟิเคชัน |
| โดย | นายวิฒเนศ ทองมี |
| สาขาวิชา | วิทยาศาสตร์คอมพิวเตอร์ |
| อาจารย์ที่ปรึกษา | ผู้ช่วยศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล |

---

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

.....อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

.....กรรมการ
(อาจารย์ ดร.เศรษฐา ปานงาม)

.....กรรมการ
(อาจารย์ ดร.อรรถสิทธิ์ สุรฤกษ์)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิฒเนศ ทองมี : วิธีการตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยใช้ฟัซซิฟิเคชัน (AN APPROACH OF SOFT PRUNING FOR DECISION TREES USING FUZZIFICATION) อ.ที่ปรึกษา : ผศ. ดร.บุญเสริม กิจศิริกุล, 62 หน้า. ISBN 974-17-1457-2.

ต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ของเครื่องวิธีการหนึ่งที่มีการใช้งานกันอย่างแพร่หลายในการจำแนกข้อมูล แต่ในการสร้างต้นไม้ตัดสินใจจากข้อมูลสอน โดยเฉพาะข้อมูลสอนที่มีสัญญาณรบกวน ต้นไม้ตัดสินใจที่ได้อาจเกิดการโอเวอร์ฟิตข้อมูล ทำให้ต้นไม้ตัดสินใจมีขนาดใหญ่และมีความถูกต้องในการจำแนกข้อมูลน้อยลง วิธีการแก้ปัญหาการโอเวอร์ฟิตข้อมูลทำได้โดยการลดขนาดหรือรายละเอียดของต้นไม้ตัดสินใจลง เรียกว่า การตัดเล็มต้นไม้ตัดสินใจ

แต่เนื่องจากการตัดเล็มต้นไม้ตัดสินใจอาจตัดส่วนของต้นไม้ที่มีความสำคัญในการตัดสินใจจำแนกข้อมูลออกไปด้วย จึงเกิดวิธีการตัดเล็มแบบใหม่ที่เรียกว่าการตัดเล็มอย่างอ่อนขึ้น โดยการนำแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กมาช่วยกำหนดน้ำหนักความสำคัญของแต่ละโนดในต้นไม้ตัดสินใจแทนการตัดโนดออกไปจากต้นไม้ตัดสินใจโดยสิ้นเชิง

งานวิจัยนี้ได้เสนอการทำฟัซซิฟิเคชันกับวิธีการตัดเล็มอย่างอ่อน โดยใช้ฟังก์ชันความเป็นสมาชิกรูปตัวเอสกับโนดอินพุตของโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก เพื่อให้เกิดขอบเขตการตัดสินใจอย่างอ่อนกับโนดอินพุตที่พิจารณาคุณสมบัติแบบข้อมูลต่อเนื่อง ส่วนในกรณีคุณสมบัติแบบข้อมูลไม่ต่อเนื่องนั้น ก็ได้เสนอการทำฟัซซิฟิเคชันโดยใช้ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมในการเพิ่มความเป็นไปได้ให้กับค่าที่มีความใกล้เคียงกับค่าของคุณสมบัตินั้นด้วย เพื่อให้เกิดความเป็นไปได้ในเส้นทางการตัดสินใจที่ใกล้เคียงกัน ซึ่งผลการทดลองหาค่าความถูกต้องจากการจำแนกข้อมูลด้วยวิธีการดังกล่าวนี้บนชุดข้อมูลทั้งหมด 17 ชุด เปรียบเทียบกับต้นไม้ตัดสินใจจากระบบ C4.5 ที่ยังไม่ได้ตัดเล็มและที่ตัดเล็มแล้ว และเปรียบเทียบกับวิธีการตัดเล็มอย่างอ่อนนั้น สรุปได้ว่าผลของงานวิจัยนี้ให้ประสิทธิภาพที่ดีที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

| | |
|---|---|
| ภาควิชา วิศวกรรมคอมพิวเตอร์ | ลายมือชื่อนิสิต ..... |
| สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ | ลายมือชื่ออาจารย์ที่ปรึกษา ..... |
| ปีการศึกษา 2545 | ลายมือชื่ออาจารย์ที่ปรึกษาร่วม ..... |

# # 4370490721 : MAJOR COMPUTER SCIENCE

KEY WORD: DECISION TREE / PRUNING / FUZZIFICATION / NEURAL NETWORK

WIKANES THONGMEE : AN APPROACH OF SOFT PRUNING FOR DECISION TREES USING FUZZIFICATION. THESIS ADVISOR : ASST. PROF. BOONSERM KIJSIRIKUL, Ph.D., 62 pp. ISBN 974-17-1457-2.

Decision trees have been widely and successfully used in machine learning nowadays. However, they have suffered from the overfitting problem in noisy domain which causes too many details in decision trees and reduces their accuracies. This problem has been remedied by decision tree pruning. Many methods for decision tree pruning have been proposed with the same basis of removing some nodes from the tree completely. However, the removed nodes may have some significant roles in data classification. Thus, a technique of decision tree pruning without node removal, called soft-pruning, has been proposed. Soft-pruning gives weights to nodes according to their significances which are determined by a backpropagation neural network.

This thesis proposes a novel fuzzy method for soft-pruning decision trees called fuzzy soft-pruning. This method uses fuzzy membership functions to represent the decision of each node in the soft-pruning process instead of sharp boundary decision. The sigmoid membership functions were used for continuous attributes, while triangular membership functions were used for discrete attributes to give some levels of uncertainty to values around the threshold of decision. Experimental results on seventeen multi-class domains demonstrate that the novel method outperforms both C4.5's error-based pruning and soft-pruning.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

| | |
|---|---|
| Department Computer Engineering | Student's signature |
| Field of study Computer Science | Advisor's signature |
| Academic year 2002 | Co-advisor's signature |

# กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความช่วยเหลืออย่างดียิ่งจาก ผศ.ดร.บุญเสริม กิจศิริกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้เสียสละเวลาให้คำแนะนำ แนวทางการวิจัย และข้อคิดเห็นที่มีประโยชน์มากมาย เพื่อประกอบการวิจัยของข้าพเจ้ามาโดยตลอด

ขอขอบคุณอาจารย์ รศ.ดร.ประภาส จงสถิตย์วัฒนา อาจารย์ ดร.เศรษฐา ปานงาม และอาจารย์ ดร.อรรถสิทธิ์ สุรฤกษ์ กรรมการวิทยานิพนธ์ ที่ท่านได้กรุณาให้คำแนะนำและข้อชี้แนะในการตรวจและแก้ไขวิทยานิพนธ์ฉบับนี้

ขอขอบคุณคุณก้องศักดิ์ จงเกษมวงศ์ และเพื่อนๆ พี่ๆ น้องๆ ห้องปฏิบัติการอัจฉริยภาพเครื่องกลและการค้นพบความรู้ (MIND LAB) และห้องปฏิบัติการอื่นๆ ที่ได้ให้คำแนะนำ และกำลังใจแก่ข้าพเจ้าตลอดเวลาที่ศึกษาในภาควิชาวิศวกรรมคอมพิวเตอร์แห่งนี้

ท้ายที่สุด ข้าพเจ้าใคร่ขอกราบขอบพระคุณบิดา มารดา ผู้ซึ่งสนับสนุนและให้กำลังใจแก่ข้าพเจ้าเสมอมาจนสำเร็จการศึกษา

วิฆเนศ ทองมี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญ

หน้า

**บทคัดย่อภาษาไทย** ..... **ง**

**บทคัดย่อภาษาอังกฤษ** ..... **จ**

**กิตติกรรมประกาศ** ..... **ฉ**

**สารบัญ** ..... **ช**

**สารบัญตาราง** ..... **ฌ**

**สารบัญรูปภาพ** ..... **ญ**

บทที่

**1 บทนำ** ..... **1**

1.1 ความเป็นมาและความสำคัญของปัญหา ..... 1

1.2 วัตถุประสงค์ของการวิจัย ..... 2

1.3 ขอบเขตของการวิจัย ..... 2

1.4 ขั้นตอนและวิธีดำเนินงานวิจัย ..... 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ ..... 2

1.6 ผลงานที่ตีพิมพ์จากงานวิจัย ..... 3

1.7 โครงสร้างวิทยานิพนธ์ ..... 3

**2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง** ..... **4**

2.1 ทฤษฎีที่เกี่ยวข้อง ..... 4

2.1.1 ต้นไม้ตัดสินใจ ..... 4

2.1.2 แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 11

2.1.3 การตัดเล็มอย่างอ่อน ..... 13

2.1.4 เซตวิภัชนัย ..... 16

2.1.5 การเปรียบเทียบขั้นตอนวิธี ..... 18

2.2 งานวิจัยที่เกี่ยวข้อง ..... 19

2.2.1 การตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยการใช้แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 19

2.2.2 A Comparative Analysis of Methods for Pruning Decision Trees ..... 19

# สารบัญ (ต่อ)

หน้า

2.2.3 Soft Decision Trees: A New Approach Using Non-linear Fuzzification ..... 20
**3 ขั้นตอนวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน ..... 21**
3.1 แนวคิดในการใช้ฟัซซิฟิเคชันกับการตัดเล็มอย่างอ่อน ..... 21
3.2 การสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติแบบข้อมูลต่อเนื่อง ..... 22
3.3 การสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติแบบข้อมูลไม่ต่อเนื่อง ..... 25
**4 การทดลองและผลการทดลอง ..... 28**
4.1 วิธีการทดลอง ..... 28
4.1.1 ชุดข้อมูลในการทดลอง ..... 28
4.1.2 การแบ่งข้อมูลสอน และข้อมูลทดสอบ ..... 30
4.1.3 กระบวนการทดลอง ..... 30
4.2 เครื่องมือในการทดลอง ..... 31
4.3 ผลการทดลอง ..... 31
**5 สรุปผลการวิจัย และข้อเสนอแนะ ..... 36**
5.1 สรุปผลการวิจัย ..... 36
5.2 ข้อเสนอแนะ ..... 37
**รายการอ้างอิง ..... 38**
**ภาคผนวก ..... 39**
ภาคผนวก ก ผลการทดลองการจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 40
ภาคผนวก ข การใช้งานโปรแกรม ..... 42
ภาคผนวก ค รายละเอียดข้อมูลที่ใช้ในการทดลอง ..... 53
ภาคผนวก ง ตารางแจกแจงแบบ t ..... 61
**ประวัติผู้เขียนวิทยานิพนธ์ ..... 62**

# สารบัญตาราง

หน้า

ตารางที่ 2.1 ตัวอย่างข้อมูลสอนของการตัดสินใจเล่นกอล์ฟ ..... 5
ตารางที่ 4.1 ชุดข้อมูลที่ใช้ในการทดลอง ..... 29
ตารางที่ 4.2 ผลการทดสอบค่าความถูกต้องของวิธีการตัดเล็มอย่างอ่อนโดยใช้
ฟัซซิฟิเคชันเทียบกับต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็มและตัดเล็มแล้ว
และวิธีการตัดเล็มอย่างอ่อน ..... 32
ตารางที่ 4.3 การเปรียบเทียบผลการทดลองด้วยวิธีค่าระดับความมั่นใจ ระหว่าง
วิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน กับต้นไม้ตัดสินใจที่ยังไม่ได้
ตัดเล็มและตัดเล็มแล้ว และวิธีการตัดเล็มอย่างอ่อนโดยไม่ใช้ฟัซซิฟิเคชัน ..... 33
ตารางที่ ก1 ผลการทดลองการจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก
เปรียบเทียบกับวิธีการตัดเล็มอย่างอ่อน และวิธีการตัดเล็มอย่างอ่อน
โดยใช้ฟัซซิฟิเคชัน ..... 41

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญรูปภาพ

หน้า

รูปที่ 2.1 แผนภาพแสดงต้นไม้ตัดสินใจของการเล่นกอล์ฟ ..... 4
รูปที่ 2.2 ต้นไม้ตัดสินใจที่ได้หลังจากทำการเลือกคุณสมบัติสภาพแวดล้อมเป็นราก ..... 8
รูปที่ 2.3 โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 11
รูปที่ 2.4 ในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 12
รูปที่ 2.5 กฎที่ได้จากต้นไม้ตัดสินใจที่ใช้ทดสอบการเล่นกอล์ฟ ..... 14
รูปที่ 2.6 โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กในการตัดสินใจเล่นกอล์ฟ ..... 15
รูปที่ 2.7 กราฟแสดงฟังก์ชันความเป็นสมาชิก T(x) ..... 17
รูปที่ 3.1 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “ความชื้น > 75%”
จากวิธีการตัดเล็มอย่างอ่อน โดยไม่ใช้ฟัซซิฟิเคชัน ..... 22
รูปที่ 3.2 กราฟฟังก์ชันความเป็นสมาชิกรูปตัวเอส สำหรับโนดอินพุตที่เปรียบเทียบ
คุณสมบัติแบบข้อมูลต่อเนื่อง ..... 23
รูปที่ 3.3 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “ความชื้น > 75%” ..... 24
รูปที่ 3.4 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “ความชื้น ≤ 75%” ..... 25
รูปที่ 3.5 กราฟฟังก์ชันความเป็นสมาชิกที่ต้องการสำหรับโนดอินพุต
“สภาพแวดล้อม = แดดร่ม” ..... 26
รูปที่ 3.6 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “สภาพแวดล้อม = แดดร่ม” ..... 27
รูปที่ ข1 หน้าจอหลักเมื่อผู้ใช้งานเรียกโปรแกรม ..... 43
รูปที่ ข2 หน้าจอเมนูย่อยภายในเมนู Tree ..... 44
รูปที่ ข3 หน้าจอสำหรับเลือกแฟ้มข้อมูลในการสร้างต้นไม้ตัดสินใจ ..... 44
รูปที่ ข4 หน้าจอสำหรับการเลือกแฟ้มข้อมูล ..... 44
รูปที่ ข5 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของการสร้างต้นไม้ตัดสินใจ ..... 45
รูปที่ ข6 ข้อความเมื่อสร้างต้นไม้ตัดสินใจเสร็จ ..... 45
รูปที่ ข7 ตัวอย่างข้อมูลต้นไม้ตัดสินใจที่ถูกสร้างขึ้น และถูกเรียกขึ้นมาจากเมนู View Output ... 46
รูปที่ ข8 หน้าจอเมนูย่อยภายในเมนู Transform ..... 46
รูปที่ ข9 หน้าจอสำหรับเลือกแฟ้มข้อมูลในการแปลงข้อมูล ..... 47
รูปที่ ข10 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของการแปลงข้อมูล ..... 48
รูปที่ ข11 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของการทำฟัซซิฟิเคชัน ..... 49
รูปที่ ข12 หน้าจอเมื่อการแปลงข้อมูลเสร็จสมบูรณ์ ..... 49

# สารบัญรูปภาพ (ต่อ)

หน้า

รูปที่ ข13 หน้าจอเมนูย่อยภายในเมนู BNN ..... 50
รูปที่ ข14 หน้าจอสำหรับเลือกแฟ้มข้อมูลในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 50
รูปที่ ข15 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ..... 51
รูปที่ ข16 หน้าจอเมื่อสอนนิวรอลเน็ตเวิร์กเสร็จสมบูรณ์ ..... 51
รูปที่ ข17 หน้าจอแสดงค่าความถูกต้องของข้อมูลทดสอบ ..... 52

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 1

# บทนำ

## 1.1 ความเป็นมาและความสำคัญของปัญหา

ต้นไม้ตัดสินใจ (Decision Tree) เป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning) วิธีการหนึ่งที่ได้รับความนิยมอย่างมาก ในการจำแนกข้อมูลออกเป็นกลุ่ม โดยอาศัยลักษณะของข้อมูลในการตัดสินใจ อย่างไรก็ตามต้นไม้ตัดสินใจมีข้อจำกัดในเรื่องความถูกต้องสมบูรณ์ของข้อมูลที่ใช้สอน (Training Data) หากมีความผิดพลาดของข้อมูลเพียงเล็กน้อย จะทำให้ประสิทธิภาพของต้นไม้ตัดสินใจที่ได้ลดน้อยลง ความผิดพลาดของข้อมูลอาจจะเกิดได้จากสัญญาณรบกวน (Noise Data) การวัดค่าที่ไม่เที่ยงตรง การประเมินค่าโดยอาศัยความรู้สึกส่วนตัว การให้คำจำกัดความของคุณสมบัติข้อมูลไม่เพียงพอ หรือแม้กระทั่งการผิดพลาดของข้อมูลโดยทั่วไป [2]

ต้นไม้ตัดสินใจที่เกิดจากข้อมูลที่ผิดพลาดดังกล่าว จะมีขนาดใหญ่ มีรายละเอียดมากเกินความจำเป็น เนื่องมาจากการพยายามแยกแยะข้อมูลสอนที่มีความผิดพลาด เรียกว่าเกิดการโอ-เวอร์ฟิตข้อมูล (Overfit the Data) [12] ซึ่งจะได้ต้นไม้ตัดสินใจที่ละเอียดเกินไป และลดความถูกต้องของการจำแนกข้อมูลลง

จากปัญหาดังกล่าว ทำให้มีการแก้ไขเพื่อลดขนาด หรือรายละเอียดของต้นไม้ตัดสินใจลง เรียกว่าวิธีการตัดเล็ม (Pruning) ต้นไม้ตัดสินใจ เพื่อให้ต้นไม้ตัดสินใจที่ได้มีขนาดเล็กลงแต่มีประสิทธิภาพในการจำแนกข้อมูลมากขึ้น

แต่ทั้งนี้การตัดเล็มต้นไม้ตัดสินใจอาจทำให้ประสิทธิภาพของต้นไม้ตัดสินใจที่ตัดเล็มน้อยลงได้ เนื่องจากส่วนที่ทำการตัดเล็มไปอาจมีความสำคัญต่อการจำแนกข้อมูล แต่เมื่อเข้ากระบวนการตัดเล็มแล้วได้ถูกตัดออกไป ดังนั้น ก้องศักดิ์ จงเกษมวงศ์ [1] ได้วิจัยคิดหาวิธีการตัดเล็มแบบใหม่ โดยนำแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กมาใช้ในส่วนของการตัดเล็ม ซึ่งทำให้ไม่มีการตัดเล็มส่วนของต้นไม้ออกไปจริงๆ แต่ใช้นิวรอลเน็ตเวิร์กช่วยกำหนดน้ำหนักของแต่ละโนด (Node) ของต้นไม้ตัดสินใจโดยขึ้นอยู่กับความสำคัญ และเรียกวิธีการนี้ว่าการตัดเล็มอย่างอ่อน (Soft Pruning)

วิทยานิพนธ์ฉบับนี้จะเลือกใช้เซตวิภัชนัย (Fuzzy Set) ช่วยปรับปรุงการตัดเล็มอย่างอ่อนเพื่อเพิ่มความถูกต้องในการจำแนกข้อมูล

## 1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและประยุกต์ใช้เซตวิภัชนัย ในการตัดเล็มอย่างอ่อนกับต้นไม้ตัดสินใจ
2. เพื่อศึกษาและวิเคราะห์ ถึงปัจจัยที่ทำให้การตัดเล็มอย่างอ่อน และการตัดเล็มอย่างอ่อนโดยใช้เซตวิภัชนัย มีประสิทธิภาพดีกว่าหรือด้อยกว่าการตัดเล็มแบบ Error-Based Pruning

## 1.3 ขอบเขตของการวิจัย

1. งานวิจัยนี้ใช้ระบบ C4.5 รุ่น 8 [12] ในการสร้างต้นไม้ตัดสินใจ
2. งานวิจัยนี้ใช้ระบบ Aspirin/MIGRAINES Neural Network Software รุ่น 6.0 ในการตัด-เล็มอย่างอ่อน
3. ต้นไม้ตัดสินใจที่นำมาทดสอบมีขนาดไม่เกิน 1000 โนด
4. งานวิจัยนี้จะเปรียบเทียบกับวิธีการตัดเล็มอื่น 2 วิธี ได้แก่ วิธี Error-Based Pruning และวิธีการตัดเล็มอย่างอ่อน

## 1.4 ขั้นตอนและวิธีดำเนินงานวิจัย

1. ศึกษาแนวคิด และทฤษฎี การเรียนรู้ของวิธีการต้นไม้ตัดสินใจ
2. ศึกษาแนวคิด และทฤษฎี การเรียนรู้ของวิธีการแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก
3. ศึกษาแนวคิด และทฤษฎี การเรียนรู้ของวิธีการเซตวิภัชนัย
4. เลือกชุดข้อมูลสำหรับการทดลอง และปรับให้อยู่ในรูปแบบที่นำไปใช้ได้
5. พัฒนาโปรแกรมสำหรับการตัดเล็มต้นไม้ตัดสินใจ โดยประยุกต์ใช้เซตวิภัชนัยที่ได้ศึกษามา
6. ทำการทดลองตัดเล็มต้นไม้ตัดสินใจ บนชุดข้อมูลต่างๆ ด้วยโปรแกรมที่พัฒนาขึ้น และวัดความถูกต้องโดยใช้ชุดข้อมูลสำหรับทดสอบ
7. วิเคราะห์และเปรียบเทียบผลที่ได้จากต้นไม้ตัดสินใจที่ทำการตัดเล็มด้วยโปรแกรม กับต้นไม้ตัดสินใจที่ตัดเล็มด้วยวิธีการอื่น
8. แก้ไขโปรแกรม
9. สรุปผลการวิจัย ข้อเสนอแนะ และจัดทำรายงานวิทยานิพนธ์

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. งานวิจัยนี้สามารถสร้างวิธีการตัดเล็มต้นไม้ตัดสินใจที่มีความถูกต้องมากขึ้น
2. งานวิจัยนี้สามารถเป็นแนวทางในการวิจัยวิธีการตัดเล็มต้นไม้ตัดสินใจ หรือวิธีการประยุกต์ใช้เซตวิภัชนัย แบบอื่นได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

## 1.6 ผลงานที่ตีพิมพ์จากงานวิจัย

งานวิจัยในวิทยานิพนธ์ฉบับนี้ได้รับการตีพิมพ์และนำเสนอในงานประชุมวิชาการ The Third International Conference on Intelligent Technologies and Third Vietnam–Japan Symposium on Fuzzy Systems and Applications (InTech/VJFuzzy'02) วันที่ 3-5 ธันวาคม พ.ศ. 2545 ในบทความเรื่อง "Soft Pruning Decision Trees Using Fuzzification" โดย Wikanes Thongmee และ Boonserm Kijsirikul

## 1.7 โครงสร้างวิทยานิพนธ์

ในบทต่อไปของวิทยานิพนธ์นี้จะกล่าวถึงทฤษฎีที่นำมาประยุกต์ใช้ และงานวิจัยที่เกี่ยวข้อง ส่วนในบทที่ 3 จะกล่าวถึงแนวคิดและวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน ในบทที่ 4 จะกล่าวถึงรายละเอียดขั้นตอน และเครื่องมือต่างๆ ในการทดลอง พร้อมทั้งผลการทดลองหาค่าความถูกต้องในการจำแนกข้อมูลด้วยวิธีการต่างๆ พร้อมทั้งสรุปผลงานวิทยานิพนธ์ และข้อเสนอแนะต่อไปในบทสุดท้าย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 2

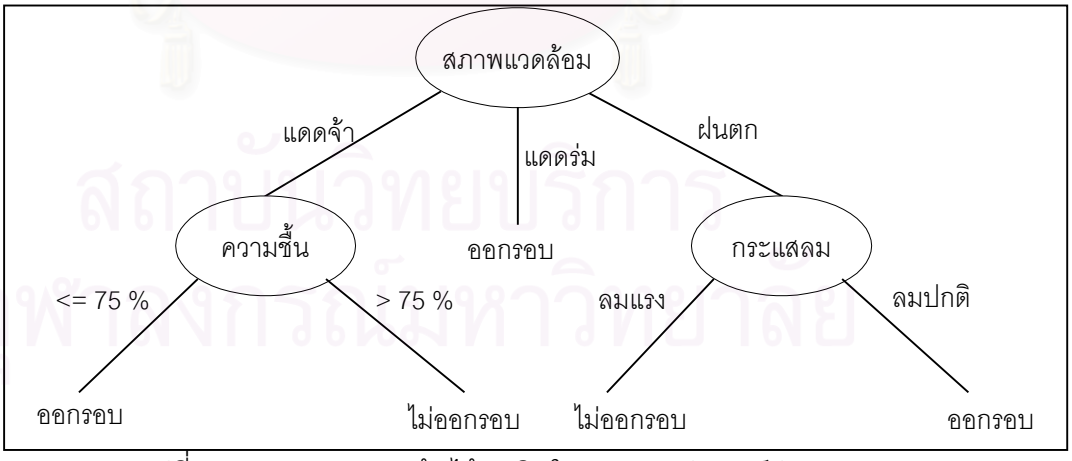
# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

## 2.1 ทฤษฎีที่เกี่ยวข้อง

### 2.1.1 ต้นไม้ตัดสินใจ (Decision Trees)

ต้นไม้ตัดสินใจเป็นโครงสร้างที่ประกอบขึ้นจาก ราก (Root) โนด (Node) กิ่ง (Branch) และ ใบ (Leaf) ช่วยในการตัดสินใจหรือตอบคำถามในเรื่องเฉพาะที่ต้นไม้นี้เก็บไว้ โดยเริ่มจากส่วนราก แล้วไล่ลงไปตามโนด กิ่ง จนกระทั่งถึงใบ ซึ่งเป็นคำตอบหรือการตัดสินใจ ตัวอย่างของต้นไม้ตัดสินใจดังแสดงในรูปที่ 2.1 ส่วนประกอบที่สำคัญของต้นไม้ตัดสินใจมีดังนี้

1. ราก เป็นจุดเริ่มต้นหรือ โนดแรกของคำถาม โดยคำตอบจะเป็นค่าที่เป็นไปได้ของคุณสมบัติไม่แบ่งกลุ่ม (Non-Category Attribute) บนข้อมูลตัวอย่าง ถ้าคำตอบตรงกับค่าใดบนโนดนี้ ก็จะวิ่งไปสู่กิ่งหรือใบต่อไป
2. โนด เป็นจุดของคำถามตามคุณสมบัติไม่แบ่งกลุ่ม
3. กิ่ง เป็นค่าที่เป็นไปได้ตามคุณสมบัติไม่แบ่งกลุ่ม ซึ่งจะนำไปสู่โนดหรือใบ
4. ใบ เป็นคำตอบหรือการตัดสินใจ โดยจะเป็นค่าที่เป็นไปได้ของคุณสมบัติแบ่งกลุ่ม (Category Attribute)
5. เส้นทาง (Tree Path) เป็นทางเดินตั้งแต่รากจนถึงใบแต่ละใบ ซึ่งจะนำไปสู่กฎต่อไป




รูปที่ 2.1 แผนภาพแสดงต้นไม้ตัดสินใจของการเล่นกอล์ฟ

ระบบต้นไม้ตัดสินใจที่ใช้ในงานวิจัยนี้ เป็นระบบ C4.5 [10,12] ซึ่งเป็นโปรแกรมเรียนรู้ต้นไม้ตัดสินใจที่รู้จักกันอย่างแพร่หลาย พัฒนาโดย J. R. Quinlan โดยพัฒนาต่อมาจากระบบ ID3

[11] เป็นวิธีการเรียนรู้จากตัวอย่างที่อาศัยแบบจำลองการจัดหมวดหมู่ (Classification Model) จากตัวอย่างเฉพาะที่เรียกว่าข้อมูลสอนแล้วสร้างเป็นต้นไม้ตัดสินใจ

ข้อมูลสอนจะมีลักษณะคล้ายกับข้อมูลในฐานข้อมูลแบบสัมพันธ์ (Relational Database) ที่ประกอบด้วย แถว (Record) หรือในที่นี้เรียกว่าตัวอย่าง (Case) และ สดมภ์ (Column) หรือในที่นี้เรียกว่า คุณสมบัติ (Attribute) ซึ่งมีด้วยกัน 2 ชนิด ดังตัวอย่างในตารางที่ 2.1 คือ

1. คุณสมบัติแบ่งกลุ่ม (Category Attribute) หรือในที่นี้จะเรียกว่า กลุ่ม (Class) เป็นคุณสมบัติที่กำหนดว่าตัวอย่างนั้น ๆ ถูกจัดอยู่ในกลุ่มไหน โดยจะมีเพียง 1 คุณสมบัติในแต่ละชุดข้อมูล และข้อมูลที่เก็บจะเป็นชนิดไม่ต่อเนื่อง (Discrete Value) เท่านั้น เช่น {ใช่, ไม่ใช่} {0, 1, 2, 3} เป็นต้น ดังในตารางที่ 2.1 จะมี “การออกรอบ” เป็นคุณสมบัติแบ่งกลุ่มของชุดข้อมูลนี้ โดยประกอบด้วย 2 กลุ่มคือ ออกรอบ และไม่ออกรอบ
2. คุณสมบัติไม่แบ่งกลุ่ม (Non-Category Attribute) หรือในที่นี่จะเรียกว่าคุณสมบัติเป็นชุดข้อมูลที่บ่งบอกถึงคุณสมบัติต่าง ๆ ของตัวอย่างแต่ละตัวอย่าง โดยแต่ละคุณสมบัติอาจจะเก็บข้อมูลได้ทั้งชนิด ค่าต่อเนื่อง (Continuous Values) เช่น ส่วนสูง น้ำหนัก เป็นต้น หรือค่าไม่ต่อเนื่อง (Discrete Values) เช่น สีผม อาชีพ เป็นต้น ดังในตารางที่ 2.1 ชุดข้อมูลมีคุณสมบัติไม่แบ่งกลุ่มทั้งหมด 4 ตัวได้แก่ สภาพแวดล้อม อุณหภูมิ ความชื้น และกระแสลม

ตารางที่ 2.1 ตัวอย่างข้อมูลสอนของการตัดสินใจเล่นกอล์ฟ

| สภาพแวดล้อม | อุณหภูมิ ($^{o}$F) | ความชื้น (%) | กระแสลม | การออกรอบ |
|---|---|---|---|---|
| แดดจ้า | 75 | 70 | ลมแรง | ออกรอบ |
| แดดจ้า | 80 | 90 | ลมแรง | ไม่ออกรอบ |
| แดดจ้า | 85 | 85 | ลมปกติ | ไม่ออกรอบ |
| แดดจ้า | 72 | 95 | ลมปกติ | ไม่ออกรอบ |
| แดดจ้า | 69 | 70 | ลมปกติ | ออกรอบ |
| แดดร่ม | 72 | 90 | ลมแรง | ออกรอบ |
| แดดร่ม | 83 | 78 | ลมปกติ | ออกรอบ |
| แดดร่ม | 64 | 65 | ลมแรง | ออกรอบ |
| แดดร่ม | 81 | 75 | ลมปกติ | ออกรอบ |
| ฝนตก | 71 | 80 | ลมแรง | ไม่ออกรอบ |
| ฝนตก | 65 | 70 | ลมแรง | ไม่ออกรอบ |
| ฝนตก | 75 | 80 | ลมปกติ | ออกรอบ |
| ฝนตก | 68 | 80 | ลมปกติ | ออกรอบ |
| ฝนตก | 70 | 96 | ลมปกติ | ออกรอบ |

การสร้างต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจจะใช้วิธีแบ่งแยกแล้วเอาชนะ (Divide and Conquer) โดยการเลือกคุณสมบัติขึ้นมา 1 ตัวจากคุณสมบัติทั้งหมด เพื่อเป็นรากของต้นไม้ จากนั้นจะแบ่งตัวอย่างเซตย่อย ๆ ตามค่าของคุณสมบัติที่เลือกมา การแบ่งนี้จะทำให้เกิดเหตุการณ์ 1 ใน 3 กรณีดังต่อไปนี้

1. หากเซตของตัวอย่างหลังจากการแบ่งที่ปลายของกิ่งหนึ่ง ๆ ประกอบด้วยตัวอย่างที่เป็นกลุ่มเดียวกันทั้งหมด โนดที่ปลายกิ่งนั้นจะกลายเป็นใบ
2. หากไม่มีตัวอย่างตกอยู่ในค่าตามกิ่งหนึ่ง ๆ หลังจากแบ่งแล้ว โนดที่ปลายกิ่งนี้ก็จะกลายเป็นใบเช่นกัน แต่ถูกจัดอยู่ในกลุ่มใดนั้นต้องตัดสินใจโดยใช้ข้อมูลอื่น โดย C4.5 จะใช้ค่าของกลุ่มที่มีความถี่สูงที่สุดของโนดก่อนหน้าเป็นค่าของใบนี้
3. หากเซตของตัวอย่างที่ปลายของกิ่งหนึ่ง ๆ ประกอบด้วยตัวอย่างหลายกลุ่ม โนดที่ปลายกิ่งนี้จะต้องถูกแบ่งต่อไป โดยแบ่งตัวอย่างออกเป็นเซตย่อย ๆ ตามค่าที่เป็นไปได้ในแต่ละกิ่งของโนดนี้ จากนั้นจึงเริ่มแบ่งตัวอย่างในแต่ละกิ่งโดยเลือกคุณสมบัติที่เหลืออยู่เพื่อแบ่งตัวอย่างต่อไป

จะเห็นได้ว่าในการสร้างต้นไม้ตัดสินใจ จะเป็นการแบ่งตัวอย่างออกเป็นเซตย่อย ๆ ตามคุณสมบัติ จนกระทั่งได้เซตของตัวอย่างที่เป็นกลุ่มเดียวกัน ตัวอย่างการตัดสินใจเล่นกอล์ฟในตารางที่ 1 ประกอบด้วยคุณสมบัติ 4 ตัว (สภาพแวดล้อม อุณหภูมิ ความชื้น และกระแสลม) และกลุ่ม (การออกรอบ) 2 กลุ่ม (ออกรอบ และไม่ออกรอบ)

การสร้างต้นไม้ตัดสินใจด้วยวิธีแบ่งแยกแล้วจัดกลุ่มจากตัวอย่างใด ๆ สามารถจะสร้างต้นไม้ตัดสินใจขึ้นมาได้หลายต้นจากข้อมูลชุดเดียวกัน ขึ้นอยู่กับการเลือกคุณสมบัติที่ใช้แบ่งที่แตกต่างกันไป ยิ่งจำนวนคุณสมบัติที่ใช้แบ่งตัวอย่างและค่าที่เป็นไปได้ในแต่ละคุณสมบัติยิ่งมาก ก็จะทำให้ได้จำนวนต้นไม้ตัดสินใจที่เป็นไปได้มากขึ้น และต้นไม้ที่ได้ก็จะมีขนาดต่าง ๆ กัน บางต้นก็มีขนาดเล็ก บางต้นก็มีขนาดใหญ่ แต่ต้นไม้ที่เราต้องการจะเป็นต้นไม้ที่มีขนาดเล็ก เพราะจะใช้จำนวนครั้งในการแบ่งตัวอย่างน้อย และเป็นต้นไม้ที่เข้าใจง่าย ดังนั้นจึงเป็นการยากที่จะสร้างต้นไม้ทั้งหมดที่เป็นไปได้ก่อน แล้วเลือกต้นไม้ที่ต้องการออกมาเมื่อมีจำนวนคุณสมบัติ หรือค่าที่เป็นไปได้ในแต่ละคุณสมบัติมีจำนวนมาก

จะเห็นได้ว่าจุดสำคัญจะอยู่ที่การเลือกคุณสมบัติที่ใช้แบ่งตัวอย่าง เนื่องจากวิธีนี้จะเป็นการทำไปข้างหน้าไม่มีการย้อนกลับ กล่าวคือเมื่อเลือกคุณสมบัติหนึ่งคุณสมบัติใดขึ้นมาแบ่งตัวอย่างแล้ว จะไม่มีการถอยหรือกลับมาเลือกคุณสมบัติอื่นเพื่อแบ่งใหม่อีก ดังนั้นจึงต้องเลือกคุณสมบัติที่ดีที่สุดในการแบ่งตัวอย่างของแต่ละกิ่ง คุณสมบัติที่ดีที่สุดควรเป็นคุณสมบัติที่เมื่อแบ่งตัว

อย่างตามคุณสมบัตินี้แล้ว จะทำให้จำนวนครั้งของการแบ่งต่อหรือจำนวนโนดต่อจากกิ่งนี้น้อยที่สุด ซึ่งจะนำไปสู่ต้นไม้ที่เล็กและเข้าใจง่าย

เมื่อโปรแกรม C4.5 ทำงานสร้างต้นไม้ตัดสินใจจากข้อมูลสอนตามตารางที่ 2.1 จะได้ต้นไม้ตัดสินใจตามรูปที่ 2.1

ค่ามาตรฐานเกน (Gain Criterion)

วิธีการสร้างต้นไม้ตัดสินใจแบบ ID3 [12] จะใช้ค่ามาตรฐานเกน (Gain Criteria) ในการตัดสินใจเลือกคุณสมบัติที่จะใช้เป็นรากหรือกิ่งในต้นไม้ โดยคำนวณค่าเกนของคุณสมบัติแต่ละตัวในการแบ่งตัวอย่าง แล้วเลือกคุณสมบัติที่มีค่าเกนสูงที่สุดมาเป็นรากหรือโนด ค่าเกนนี้คำนวณได้โดยใช้ความรู้จากทฤษฎีสารสนเทศ (Information Theory) ซึ่งมีสาระสำคัญคือ ค่าสารสนเทศของข้อมูลขึ้นอยู่กับค่าความน่าจะเป็นของข้อมูล ซึ่งสามารถวัดอยู่ในรูปของบิต (Bits) จากสูตร

ค่าสารสนเทศของข้อมูล = - $\log_2$(ความน่าจะเป็นของข้อมูล) บิต

ถ้าให้ชุดของข้อมูล M ประกอบด้วยค่าที่เป็นไปได้ คือ {$m_1$, $m_2$, …, $m_n$} และให้ความน่าจะเป็นที่จะเกิด $m_i$ เป็น $P(m_i)$ สำหรับแต่ละค่า $m_i \in M$ ดังนั้นค่าสารสนเทศของ M หรือค่าเอนโทรพี (Entropy) ของ M (เขียนแทนด้วย I(M)) จะคำนวณได้จากสูตร

$$I(M) = \sum_{i}^{n} - P(m_i) \times \log_2 (P(m_i)) \quad \text{บิต}$$

ในการเลือกคุณสมบัติที่จะมาเป็นรากของโนดใด ๆ จะอาศัยค่ามาตรฐานเกน ซึ่งคำนวณจากค่าสารสนเทศทั้งหมดของชุดข้อมูลนั้นลบด้วยค่าสารสนเทศหลังจากเลือกคุณสมบัติใดคุณสมบัติหนึ่งเป็นราก ค่าสารสนเทศหลังจากแบ่งตามคุณสมบัติที่เลือกแล้วจะคำนวณได้จาก ค่าผลรวมของผลคูณระหว่างค่าสารสนเทศของแต่ละโนดกับอัตราส่วนของตัวอย่างในแต่ละกิ่งต่อตัวอย่างทั้งหมดที่โนดนั้น ๆ หรือความน่าจะเป็นของค่าที่เป็นไปได้ของแต่ละคุณสมบัติ

กล่าวคือ ถ้าให้ข้อมูลสอนคือ T และคุณสมบัติที่เลือกเป็นราก คือ X และมีค่าทั้งหมดที่เป็นไปได้ n ค่า รากหรือโนดปัจจุบันจะแบ่งตัวอย่าง T ออกเป็นเซตย่อย ๆ {$t_1$, $t_2$, …, $t_n$} ตามค่าที่เป็นไปได้ของ X ดังนั้นจึงสามารถคำนวณค่าสารสนเทศหลังจากแบ่งตามคุณสมบัติ X ดังนี้

$$I_x(T) = \sum_{i=1}^{n} \frac{|t_i|}{|T|} \times I(t_i) \quad \text{บิต}$$

และค่ามาตรฐานเกนของคุณสมบัติ X สามารถคำนวณได้จากการลบค่าสารสนเทศทั้งหมดที่โนดนี้กับค่าสารสนเทศที่ได้หลังจากแบ่งด้วยคุณสมบัติ X ดังนี้

ค่ามาตรฐานเกน(X) = $I(T) - I_x(T)$ บิต

ในการตัดสินใจเลือกคุณสมบัติไหนเป็นรากหรือโนดนั้น จะพิจารณาจากค่ามาตรฐานเกน โดยถ้าค่ามาตรฐานเกนที่คำนวณจากการแบ่งตัวอย่างตามคุณสมบัติใดมีค่าสูงที่สุดก็จะเลือกคุณสมบัตินั้นเป็นรากหรือโนด

จากตัวอย่างการตัดสินใจเล่นกอล์ฟในตารางที่ 1 ประกอบด้วยข้อมูล 2 กลุ่ม คือ ตัวอย่างที่ตัดสินใจออกรอบ 9 ตัวอย่าง และตัดสินใจไม่ออกรอบ 5 ตัวอย่าง ดังนั้นค่าสารสนเทศทั้งหมดของข้อมูลชุดนี้จะคำนวณได้ดังนี้

$I(T) = -9/14 \times \log_2(9/14) - 5/14 \times \log_2(5/14)$

$= 0.940$ บิต

ถ้าแบ่งข้อมูลชุดนี้ตามคุณสมบัติสภาพแวดล้อม จะแบ่งตัวอย่างออกเป็น 3 เซตย่อยดังรูปที่ 2.2 โดยในรูปจะแสดงเซตย่อยด้วย [x+, y-] ซึ่งแสดงว่าข้อมูลในเซตย่อยมีข้อมูลในกลุ่มออกรอบ (แทนด้วยเครื่องหมายบวก) จำนวน x ตัว และมีข้อมูลในกลุ่มไม่ออกรอบ (แทนด้วยเครื่องหมายลบ) จำนวน y ตัว และสามารถคำนวณค่าสารสนเทศหลังจากแบ่งตัวอย่างตามคุณสมบัตินี้ได้ดังนี้

รูปที่ 2.2 ต้นไม้ตัดสินใจที่ได้หลังจากทำการเลือกคุณสมบัติสภาพแวดล้อมเป็นราก

$I_{\text{สภาพแวดล้อม}}(T)$ = 5/14 x (-2/5 x $\log_2(2/5)$ – 3/5 x $\log_2(3/5)$)
+ 4/14 x (-4/4 x $\log_2(4/4)$ – 0/4 x $\log_2(0/4)$)
+ 5/14 x (-3/5 x $\log_2(3/5)$ – 2/5 x $\log_2(2/5)$)
= 0.694 บิต

ค่ามาตรฐานเกน(สภาพแวดล้อม) = 0.940 – 0.694
= 0.246 บิต

คุณสมบัติอุณหภูมิและความชื้นเป็นข้อมูลแบบค่าต่อเนื่อง จะมีวิธีคำนวณค่ามาตรฐานเกนที่แตกต่างจากการคำนวณค่ามาตรฐานเกนของคุณสมบัติที่เป็นค่าไม่ต่อเนื่อง ซึ่งจะแสดงวิธีการคำนวณให้เห็นในหัวข้อการคำนวณสำหรับคุณสมบัติที่เป็นข้อมูลแบบต่อเนื่องต่อไป

<u>ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio Criterion)</u>

ใน ID3 จะใช้ค่ามาตรฐานเกนเป็นหลักในการเลือกคุณสมบัติที่จะใช้เป็นรากหรือโนด แต่ใน C4.5 ได้เพิ่มการใช้ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio Criterion) ในการตัดสินใจเลือกคุณสมบัติที่จะใช้เป็นรากหรือโนดอีกอย่างหนึ่ง เนื่องจากค่ามาตรฐานเกนจะมีอคติ (Bias) อย่างมากกับข้อมูลที่ประกอบด้วยคุณสมบัติที่มีค่าที่เป็นไปได้จำนวนมาก ๆ เช่นข้อมูลที่ประกอบด้วยคุณสมบัติหมายเลขประจำตัว ซึ่งปกติจะไม่ซ้ำกันในแต่ละตัวอย่าง ถ้าแบ่งข้อมูลตามคุณสมบัตินี้จะทำให้ได้จำนวนตัวอย่างเพียง 1 ตัวต่อ 1 กิ่งของต้นไม้ และชุดตัวอย่างย่อยที่ได้จะประกอบด้วยข้อมูลกลุ่มเดียว เมื่อคำนวณค่าสารสนเทศจากการแบ่งตัวอย่างบนคุณสมบัตินี้ จะได้เท่ากับ 0 เนื่องจากค่า $\log_2(1) = 0$ ทำให้ค่าเกนที่ได้ในคุณสมบัตินี้จะสูงที่สุดเสมอ แต่เป็นลักษณะการแบ่งที่ไม่เหมาะสมอย่างยิ่ง

การแก้ไขความอคติของค่ามาตรฐานเกนสามารถทำได้โดยการปรับค่ามาตรฐานเกนให้ถูกต้อง โดยใช้ค่าสารสนเทศของการแบ่งแยก (Split Information) ของคุณสมบัติแต่ละตัว ซึ่งถ้าให้ T คือชุดของตัวอย่าง เมื่อแบ่งตัวอย่างนี้ตามคุณสมบัติ X จะได้ชุดของตัวอย่างย่อยในแต่ละกิ่งคือ $\{T_1, T_2, \ldots, T_n\}$ จำนวน N ชุด ตามค่าที่เป็นไปได้ในคุณสมบัติ X เมื่อคำนวณค่าสารสนเทศของการแบ่งแยกได้ ดังนี้

ค่าสารสนเทศของการแบ่งแยก $= -\sum_{i=1}^{n} \frac{|t_i|}{|T|} \times \log_2\left(\frac{|t_i|}{|T|}\right)$

ค่าสารสนเทศของการแบ่งแยกนี้จะแสดงถึงระดับการกระจายของข้อมูล เมื่อแบ่งข้อมูลตัวอย่าง T เป็น n ชุดย่อยตามคุณสมบัติ X โดยค่านี้จะสูงสุดเมื่อ $|t_i|$ เป็น 1 เท่ากันในทุกกิ่ง และลดลงเมื่อค่า $|t_i|$ เพิ่มขึ้นเมื่อนำค่านี้ไปหารค่ามาตรฐานเกนจะได้ค่ามาตรฐานอัตราส่วนเกนดังนี้

ค่ามาตรฐานอัตราส่วนเกน = ค่ามาตรฐานเกน / ค่าสารสนเทศของการแบ่งแยก

ซึ่งค่ามาตรฐานอัตราส่วนเกนนี้จะช่วยแก้ไขความอคติของค่ามาตรฐานเกนได้ โดยทำให้ค่ามาตรฐานอัตราส่วนเกนในการแบ่งด้วยคุณสมบัติที่มีการกระจายสูงถูกปรับลดลง ดังนั้นค่ามาตรฐานอัตราส่วนเกนในคุณสมบัติของตัวอย่างที่มีการกระจายตัวของข้อมูลสูงดังที่กล่าวมาแล้วจึงไม่มีค่าสูงที่สุดเสมอ

<u>การคำนวณสำหรับคุณสมบัติที่เป็นข้อมูลต่อเนื่อง</u>

ในการคำนวณค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกนสำหรับคุณสมบัติที่ข้อมูลเป็นค่าต่อเนื่อง หรือข้อมูลตัวเลข จะกระทำได้โดยการคำนวณค่ามาตรฐานอัตราส่วนเกนหลังจากแบ่งตัวอย่างตามจุดแบ่งที่เป็นไปได้ในระดับต่าง ๆ ของคุณสมบัติที่เป็นค่าต่อเนื่อง แล้วเลือกจุดแบ่งที่มีค่ามาตรฐานอัตราส่วนเกนสูงที่สุด เป็นระดับที่จะใช้แบ่งตัวอย่าง และใช้ค่ามาตรฐานอัตราส่วนเกนที่สูงที่สุดนี้เป็นตัวแทนในการพิจารณาเลือกคุณสมบัติที่จะใช้แบ่งตัวอย่าง

สมมติว่า ตัวอย่างสอน T ประกอบด้วยคุณสมบัติต่อเนื่อง A เมื่อเรียงข้อมูลตามคุณสมบัติ A จะได้ชุดของค่าที่ไม่ซ้ำกัน m ค่า ตามลำดับ $\{v_1, v_2, \ldots, v_m\}$ จุดที่เป็นระดับที่ใช้แบ่งข้อมูลจะอยู่ระหว่างค่าของ $v_i$ กับ $v_{i+1}$ ดังนั้นจึงมีจุดที่ใช้แบ่งข้อมูลจำนวน m-1 จุดที่เป็นไปได้ ซึ่งโดยปกติจุดที่ใช้แบ่งข้อมูลจะใช้ค่า $(v_i + v_{i+1})/2$ แต่ C4.5 จะใช้ค่าจากตัวอย่างที่สูงที่สุดที่ไม่เกินจุดกึ่งกลางจากการคำนวณในแต่ละช่วง แทนที่จะใช้จุดกึ่งกลางเป็นตัวแบ่ง เพื่อรับประกันว่าค่าที่ปรากฏในต้นไม้ตัดสินใจจะปรากฏอยู่ในตัวอย่างด้วย

จากตัวอย่างการตัดสินใจเล่นกอล์ฟในตารางที่ 1 ถ้าเราแบ่งข้อมูลตามคุณสมบัติอุณหภูมิโดยใช้อุณหภูมิระหว่าง 70 ถึง 71 องศาฟาเรนไฮท์ เป็นจุดที่ใช้แบ่ง สามารถจะคำนวณค่ามาตรฐานเกนและค่ามาตรฐานอัตราส่วนเกนได้ดังนี้

$$
\begin{aligned}
I_{\text{อุณหภูมิ (ระหว่าง 70 และ 71)}}(T) &= 5/14 \times (-4/5 \times \log_2(4/5) - 1/5 \times \log_2(1/5)) \\
&\quad + 9/14 \times (-5/9 \times \log_2(5/9) - 4/9 \times \log_2(4/9)) \\
&= 0.895 \text{ บิต}
\end{aligned}
$$

ค่ามาตรฐานเกน(อุณหภูมิ) = 0.940 – 0.895

= 0.045 บิต

ค่าสารสนเทศของการแบ่งแยก (อุณหภูมิ) = $-5/14 \times \log_2(5/14) - 9/14 \times \log_2(9/14)$

= 0.940 บิต

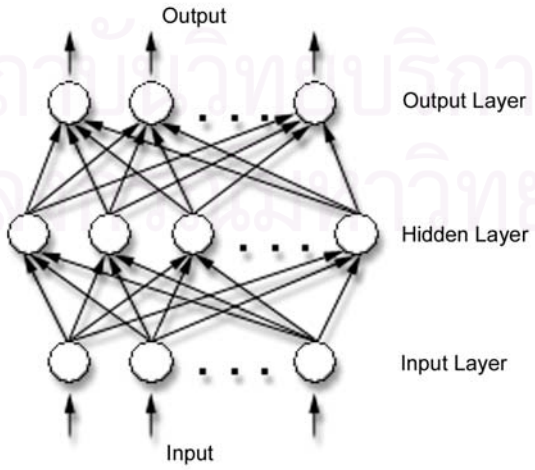
ค่ามาตรฐานอัตราส่วนเกน(อุณหภูมิ) = 0.045/0.940

= 0.0479 บิต

2.1.2 แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก (Backpropagation Neural Network) [8,10,13]

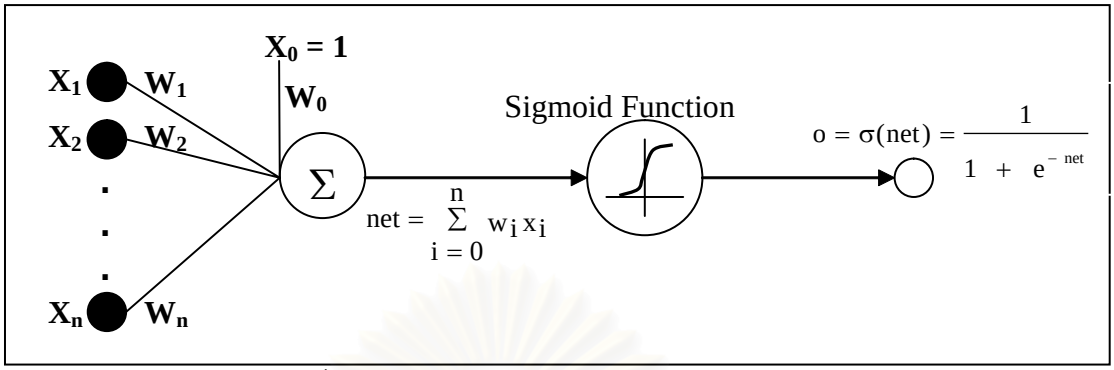
นิวรอลเน็ตเวิร์ก เป็นการเรียนรู้ของเครื่องรูปแบบหนึ่ง ซึ่งมีแนวคิดในการทำงานโดยการจำลองการทำงานบางส่วนของสมองมนุษย์ ที่ประกอบด้วยประสาท (Neural) จำนวนมากเชื่อมต่อกัน โดยนิวรอลเน็ตเวิร์กจะจำลองให้มีประสาทจำนวนหนึ่งซึ่งเชื่อมต่อถึงกัน โดยมีค่าน้ำหนัก (Weight) ของการเชื่อมต่อแต่ละแห่ง เมื่อมีการให้ตัวอย่างที่ใช้ในการเรียนรู้ นิวรอลเน็ตเวิร์กก็จะปรับค่าน้ำหนักให้เหมาะสม จนได้ผลลัพธ์ที่ถูกต้องหรือมีข้อผิดพลาดน้อยที่สุด และสามารถนำค่าน้ำหนักนี้ไปใช้ในงานที่ต้องการได้

แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กเป็นนิวรอลเน็ตเวิร์กที่ทำการเชื่อมต่อกันของเพอร์เซปตรอน โดยสามารถเชื่อมกันแบบหลายชั้น (Multilayer) ได้ และใช้ขั้นตอนวิธีแบ็กพรอพาเกชัน (The Backpropagation Algorithm) โดยในขั้นตอนการทำงานจะไม่มีการป้อนผลลัพธ์ที่ได้ในแต่ละโนดย้อนกลับไปยังโนดที่ส่งข้อมูลมาให้ โครงสร้างของแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กประกอบด้วยชั้นอินพุต (Input Layer) ชั้นฮิดเดน (Hidden Layer) และชั้นเอาต์พุต (Output Layer) แสดงดังรูปที่ 2.3 โดยจำนวนชั้นฮิดเดนสามารถมีได้มากกว่า 1 ชั้น




รูปที่ 2.3 โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

แต่ละโนดในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กจะมีลักษณะดังแสดงในรูปที่ 2.4




รูปที่ 2.4 โนดในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

ในแต่ละโนดของแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จะให้ค่าผลลัพธ์ หรือค่าเอาต์พุตออกจากโนดตามสมการ

$$o = \sigma(\vec{w} \cdot \vec{x})$$

โดยมี $\sigma$ เป็นฟังก์ชันกระตุ้น (Activation Function) ซึ่งนิยมใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function) ตามสมการ

$$\sigma(y) = \frac{1}{1+e^{-y}}$$

และมี o คือเอาต์พุตที่ออกจากโนดนั้น

$\vec{x}$ คือเวกเตอร์อินพุตที่เข้าโนดนั้น

$\vec{w}$ คือเวกเตอร์ค่าน้ำหนักของอินพุต

วิธีการแบ็กพรอพาเกชันจะเป็นการเรียนรู้เพื่อปรับค่าน้ำหนักสำหรับนิวรอลเน็ตเวิร์ก โดยที่ค่าน้ำหนักที่ได้จะเป็นค่าน้ำหนักที่ทำให้ค่าผลต่างกำลังสองที่น้อยที่สุด ระหว่างเอาต์พุตที่ได้จากเน็ตเวิร์ก และค่าเป้าหมาย โดยมีขั้นตอนสำหรับการปรับเปลี่ยนน้ำหนักดังนี้

กำหนดให้ตัวอย่างที่ใช้ในการเรียนรู้แต่ละตัวอย่างอยู่ในรูป $(\vec{x}, \vec{t})$

เมื่อ $\vec{x}$ เป็นเวกเตอร์ของอินพุตของเน็ตเวิร์ก

$\vec{t}$ เป็นเวกเตอร์ของเป้าหมายของเอาต์พุตของเน็ตเวิร์ก

$\eta$ เป็นค่าอัตราการเรียนรู้ (Learning Rate)

$x_{ji}$ เป็นอินพุตขององค์ประกอบ j ซึ่งมาจากองค์ประกอบ i

$w_{ji}$ เป็นค่าน้ำหนักขององค์ประกอบ j ซึ่งมาจากองค์ประกอบ i

1. สร้างนิวรอลเน็ตเวิร์กตามโครงสร้างที่ต้องการ ได้แก่จำนวนชั้นของเน็ตเวิร์ก จำนวนโนดในแต่ละชั้น และการเชื่อมโยงกันของโนดต่างๆ
2. กำหนดค่าน้ำหนักเริ่มต้นของแต่ละโนดแบบสุ่มให้มีค่าน้อยๆ (เช่น ระหว่าง -0.05 ถึง 0.05)
3. ทำการปรับค่าน้ำหนักด้วยขั้นตอนวิธีดังนี้ จนกระทั่งลู่เข้าหรือเกิดเงื่อนไขในการหยุดสอนตามที่กำหนด

   สำหรับ $(\vec{x}, \vec{t})$ แต่ละตัว ให้ทำดังนี้

   - ป้อนอินพุตเข้า $\vec{x}$ ในเน็ตเวิร์ก และคำนวณเอาต์พุต $o_u$ ในโนด u ทุกโนด
   - คำนวณค่าความผิดพลาด $\delta_k$ ของโนด k ในชั้นเอาต์พุต โดยที่

     $$\delta_k = o_k(1-o_k)(t_k - o_k)$$

   - คำนวณค่าความผิดพลาด $\delta_h$ ของโนด h ในชั้นฮิดเดน โดยที่

     $$\delta_h = o_h(1-o_h) \sum_{k \in outputs} w_{kh}\delta_k$$

     โดยที่ outputs หมายถึงเซตของโนดในชั้นเอาต์พุต
   - ทำการปรับค่าน้ำหนัก $w_{ji}$ โดย

     $$w_{ji} = w_{ji} + \Delta w_{ji} \quad \text{เมื่อ} \quad w_{ji} = \eta\delta_j x_{ji}$$

2.1.3 การตัดเล็มอย่างอ่อน (Soft Pruning) [1]

การตัดเล็มอย่างอ่อนประกอบไปด้วย 2 ขั้นตอนหลัก คือ (1) การสร้างกฎจากต้นไม้ตัดสินใจ และ (2) การสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กจากกฎรวมไปถึงการสร้างข้อมูลสำหรับนิวรอลเน็ตเวิร์ก และการสอนนิวรอลเน็ตเวิร์กที่ได้

การสร้างกฎจากต้นไม้ตัดสินใจ

วิธีการสร้างกฎจากต้นไม้ตัดสินใจ สามารถทำได้โดยการอ่านต้นไม้ตัดสินใจจากรากไปจนถึงใบหนึ่ง ๆ ซึ่งเส้นทางที่ได้จะต้องผ่านโนดที่เก็บคุณสมบัติ และกิ่งซึ่งเก็บค่าที่เป็นไปได้ของโนดนั้น ๆ โดยกฎที่ได้ในต้นไม้ตัดสินใจในแต่ละโนด จะมีจำนวนที่ผ่านโนดและกิ่งต่าง ๆ ไม่เท่ากัน หลังจากนั้นสร้างกฎโดยเปลี่ยนให้อยู่ในรูป “ถ้า...แล้ว” โดยนำโนดและกิ่ง อยู่ในส่วนของ “ถ้า” หากในเส้นทางของกฎที่อ่านมาได้มีโนดและกิ่งอยู่ถัดมาจากโนดที่แล้ว ก็จะเชื่อมต่อกันด้วย “และ” ต่อไป จนเส้นทางที่อ่านมาถึงใบ ซึ่งค่าที่ใบเป็นค่าของกลุ่มที่เป็นไปได้ โดยนำค่าที่ใบอยู่ในส่วนของ “แล้ว”

จากตัวอย่างต้นไม้ตัดสินใจที่ได้กล่าวถึงก่อนหน้านี้ให้รูปที่ 2.1 เราสามารถสร้างกฎตามวิธีการข้างต้นได้เป็นกฎทั้งหมด 5 ข้อดังรูปที่ 2.5

1. **ถ้า** สภาพแวดล้อมเป็นแดดจ้า **และ** ความชื้นมีค่าน้อยกว่าเท่ากับ 75 % **แล้ว** ออกรอบ
2. **ถ้า** สภาพแวดล้อมเป็นแดดจ้า **และ** ความชื้นมีค่ามากกว่า 75 % **แล้ว** ไม่ออกรอบ
3. **ถ้า** สภาพแวดล้อมเป็นแดดร่ม **แล้ว** ออกรอบ
4. **ถ้า** สภาพแวดล้อมเป็นฝนตก **และ** กระแสลมเป็นลมแรง **แล้ว** ไม่ออกรอบ
5. **ถ้า** สภาพแวดล้อมเป็นฝนตก **และ** กระแสลมเป็นลมปกติ **แล้ว** ออกรอบ

รูปที่ 2.5 กฎที่ได้จากต้นไม้ตัดสินใจที่ใช้ทดสอบการเล่นกอล์ฟ

จากกฎทั้งห้าข้อนี้ เราสามารถใช้งานได้เช่นเดียวกับต้นไม้ตัดสินใจ โดยเมื่อนำข้อมูลที่ต้องการทดสอบเข้ามาตรวจสอบกับกฎข้อต่าง ๆ ทีละข้อ หากข้อมูลที่นำมาทดสอบนี้มีค่าเป็นคุณสมบัติที่ตรงกับกฎทุกคุณสมบัติ ข้อมูลที่นำมาทดสอบชุดนั้นก็จะมีกลุ่มตามกฎข้อนั้น ๆ หากไม่ตรงทั้งหมด ก็ทำการเปรียบเทียบกับกฎข้อต่อ ๆ ไป

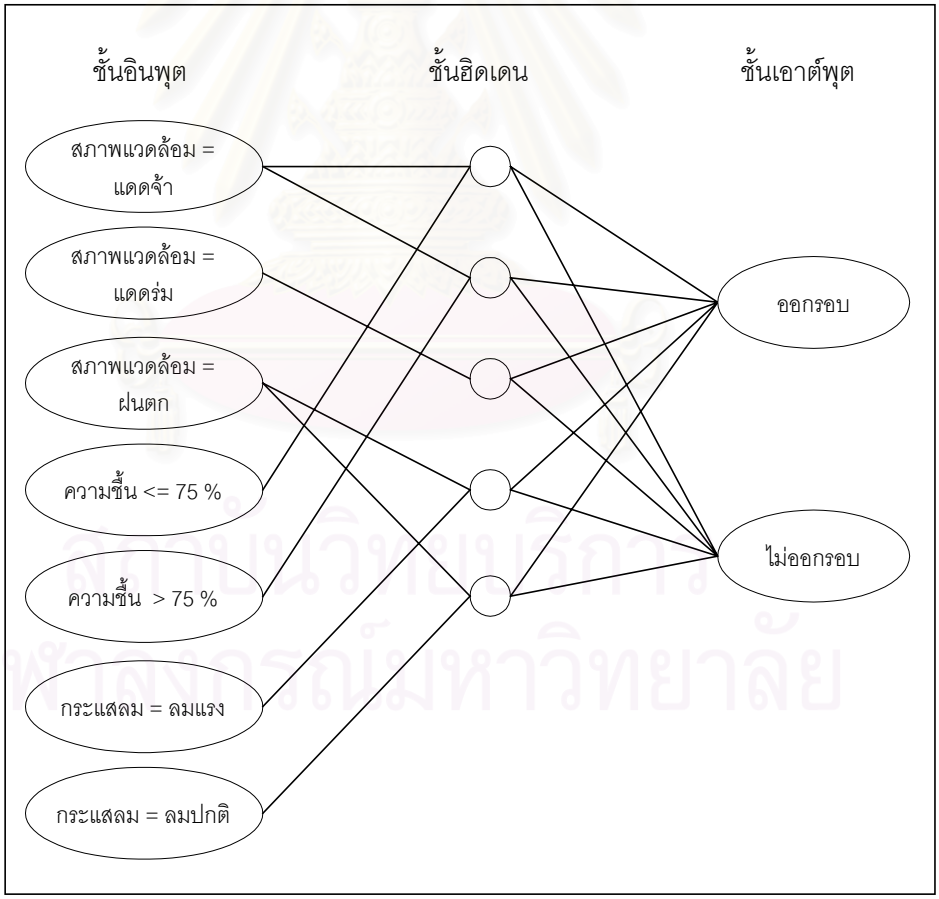
การแปลงต้นไม้ตัดสินใจเป็นกฎมีข้อดีที่สำคัญคือ โนดภายใน (Internal Node) ในต้นไม้ตัดสินใจโนดหนึ่ง ๆ ที่สนใจ จะปรากฏในกฎมากกว่าหนึ่งข้อ ต่างจากต้นไม้ตัดสินใจที่โนดนั้น ๆ จะปรากฏอยู่ครั้งเดียวและแยกตามค่าที่เป็นไปได้ของคุณสมบัติไปแต่ละกิ่ง หากเข้าสู่วิธีการตัดเล็มและโนดนี้ถูกตัดเล็มแล้ว ก็จะทำให้คุณสมบัตินั้นหายไปจากต้นไม้ตัดสินใจทันที แต่จากการแปลงเป็นกฎจะทำให้เกิดความยืดหยุ่นในการตัดเล็มมากกว่าจะกระทำโดยตรงกับต้นไม้ตัดสินใจ กล่าวคือโนดนั้นสามารถจะถูกตัดออกจากกฎข้อหนึ่ง แต่ไม่จำเป็นต้องถูกตัดออกจากกฎอีกข้อหนึ่งได้ หรือสำหรับกรณีการตัดเล็มอย่างอ่อนนี้ ก็จะทำให้การให้น้ำหนักของโนดเดียวกันในกฎแต่ละข้อ สามารถให้น้ำหนักที่ต่างกันได้ทำให้โนดนั้น ๆ ที่อยู่ในกฎหลาย ๆ ข้อสามารถถูกตัดเล็มได้อย่างอิสระไม่ขึ้นต่อกัน

<u>การสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กจากกฎ</u>

ในการสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จะต้องจำลองโครงสร้างของต้นไม้ตัดสินใจ เพื่อทำให้โครงสร้างที่อยู่ในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กนี้เปรียบเสมือนต้นไม้ตัดสินใจ

วิธีการสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กจะทำจากกฎที่สร้างขึ้นจากต้นไม้ตัดสินใจตามที่ได้กล่าวมาแล้ว โดยกฎแต่ละข้อจะนำคุณสมบัติในการทดสอบมาสร้างเป็นโนดอินพุต (Input Node) ในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จำนวนของโนดอินพุตสำหรับกฎข้อหนึ่ง ๆ จะเท่ากับจำนวนการทดสอบคุณสมบัติของกฎข้อนั้น โนดอินพุตของกฎข้อหนึ่ง ๆ ทุกโนดจะเชื่อมต่อโนดฮิดเดน (Hidden Node) หนึ่งโนดเปรียบเสมือนกับการเชื่อมโยงกันเป็นกฎข้อนั้น กล่าวคือ โนดฮิดเดนหนึ่งโนดในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จะแทนการเชื่อมกันด้วย "และ" ของการทดสอบคุณสมบัติในกฎที่ได้จากต้นไม้ตัดสินใจ ดังนั้นจำนวนโนดฮิดเดนในเน็ตเวิร์กจะมีจำนวนเท่ากับจำนวนกฎจากต้นไม้ตัดสินใจ ส่วนโนดเอาต์พุต (Output Node) ของเน็ตเวิร์กจะมีจำนวนเท่ากับกลุ่มที่มีอยู่ในต้นไม้ตัดสินใจ โดยการเชื่อมต่อระหว่างโนดฮิดเดนกับโนดเอาต์พุตจะเป็นแบบต่อกันหมด (Fully Connected)

จากตัวอย่างต้นไม้ตัดสินใจของการเล่นกอล์ฟ ซึ่งแปลงเป็นกฎได้ดังรูปที่ 2.5 เราสามารถสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จากกฎทั้ง 5 ข้อได้ดังรูปที่ 2.6




รูปที่ 2.6 โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กในการตัดสินใจเล่นกอล์ฟ

การสร้างข้อมูลสำหรับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

เมื่อได้โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กแล้ว ขั้นตอนต่อมาคือการสร้างข้อมูลสำหรับใช้ในการสอนนิวรอลเน็ตเวิร์กนี้ โดยข้อมูลที่ใช้ในการสอนและข้อมูลที่ใช้ในการทดสอบก็คือข้อมูลที่ใช้ในการสร้างและทดสอบของต้นไม้ตัดสินใจ ข้อมูลเหล่านี้จะถูกแปลงเพื่อสามารถใช้งานได้กับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก โดยข้อมูลตัวอย่างหนึ่ง ๆ จะแปลงเป็นส่วนชั้นอินพุตและชั้นเอาต์พุต ในโนดอินพุตจะพิจารณาตามการทดสอบของโนดอินพุตนั้น หากคุณสมบัติของข้อมูลมีค่าความจริงเป็นเท็จ ให้ค่าของโนดอินพุตนั้นมีค่าเป็น "0" หากมีค่าความจริงเป็นจริงให้ค่าโนดอินพุตนั้นมีค่าเป็น "1" สำหรับโนดเอาต์พุตจะกำหนดโนดที่เป็นกลุ่มของตัวอย่างนั้น ๆ ด้วย “1” และโนดที่เหลือเป็น “0”

ตัวอย่างในการสร้างข้อมูลสำหรับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กที่ได้จากการตัดสินใจเล่นกอล์ฟ จากตารางที่ 1 ในข้อมูลตัวแรก ซึ่งมีข้อมูลคือ “สภาพแวดล้อมเป็นแดดจ้า อุณหภูมิเป็น 75 $^{o}$F มีความชื้นที่ 70 % สภาพลมมีลมแรง และทำการออกรอบ” เมื่อนำมาสร้างข้อมูลจากโครงสร้างตามรูปที่ 2.6 จะได้โนดอินพุตเป็น “1 0 0 1 0 1 0” ตามลำดับ และโนดเอาต์พุตเป็น “1 0”

2.1.4 เซตวิภัชนัย (Fuzzy Sets) [7]

ทฤษฎีเซตวิภัชนัย เป็นทฤษฎีทางคณิตศาสตร์ ไม่มีความสามารถในการเรียนรู้ แต่ใช้ประมวลผลข้อมูลที่มีความไม่ชัดเจน (Fuzzy) โดยอาศัยคณิตศาสตร์พื้นฐานเกี่ยวกับเซต

มีการเรียนรู้หลายอย่างที่ได้ประยุกต์ใช้ทฤษฎีเซตวิภัชนัยร่วมด้วย เนื่องจากปัญหาโดยทั่วไปมักจะไม่สามารถแบ่งแยกได้อย่างชัดเจน แต่มีความคลุมเครือที่สามารถแทนความคลุมเครือเหล่านั้นด้วยเซตวิภัชนัยที่เหมาะสมได้ และสามารถนำไปใช้ในการเรียนรู้ต่อไปได้ แม้ข้อมูลจะมีคลุมเครือก็ตาม

เซตอวิภัชนัย (Crisp Sets)

เซตอวิภัชนัย กล่าวคือเซตตามหลักคณิตศาสตร์ทั่วไป ต่างจากเซตวิภัชนัยเนื่องจากสมาชิกทุกตัวในเซตอวิภัชนัยจะมีความเป็นสมาชิก (Membership) เท่ากัน และสิ่งที่ไม่ได้เป็นสมาชิกของเซต ก็จะไม่มีความเป็นสมาชิก

ตัวอย่างเช่น กำหนดเอกภพ (แบบอวิภัชนัย) (Crisp Universal Set) U ={a, b, c, d, e}

กำหนดเซตอวิภัชนัย A = {a, b} ภายใต้ U

ดังนั้นจะเห็นว่า $a,b \in A$ ส่วน $c,d,e \notin A$

นิยามฟังก์ชันความเป็นสมาชิก (Membership Function) A:U$\rightarrow${0, 1} ได้โดย

$$A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

ดังนั้น A(a)=1, A(b)=1 ส่วน A(b) = 0, A(d)=0 และ A(e)=0

กล่าวคือฟังก์ชันความเป็นสมาชิก ของเซตอวิภัชนัย จะให้ค่า 1 หรือ 0 เท่านั้น สำหรับสมาชิกของเอกภพแต่ละตัว

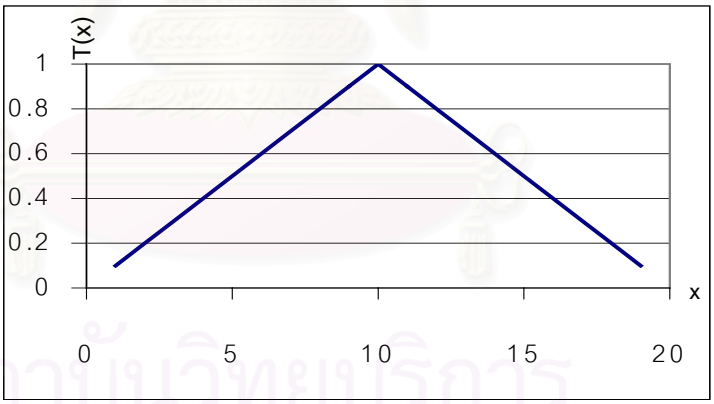
<u>เซตวิภัชนัย</u>

เซตวิภัชนัย คือเซตที่มีค่าความเป็นสมาชิกของสมาชิกในเอกภพ เป็นช่วง ซึ่งโดยทั่วไปจะนิยมให้มีค่าในช่วง [0,1]

การสร้างเซตวิภัชนัย A' บนเซตอวิภัชนัย U ทำโดย นิยามฟังก์ชันความเป็นสมาชิก A':U$\rightarrow$[0, 1]

นิยมเขียน A' ในรูป A' = {(x, A'(x)) | x$\in$U} หรือ A' = { A'(x)/x | x$\in$U}

ตัวอย่างเช่น เซตวิภัชนัย T นิยามบน U={1, 2, ..., 19} โดยต้องการให้ T(x) แทนระดับที่ x เข้าใกล้ 10 อาจจะสามารถสร้างฟังก์ชันความเป็นสมาชิก T(x) ได้ดังรูปที่ 2.7




รูปที่ 2.7 กราฟแสดงฟังก์ชันความเป็นสมาชิก T(x)

ซึ่งสามารถเขียนเซตวิภัชนัย T ได้ดังนี้

T = {(1,0.1), (2,0.2), (3,0.3), (4,0.4), (5,0.5), (6,0.6), (7,0.7), (8,0.8), (9,0.9), (10,1.0), (11,0.9), (12,0.8), (13,0.7), (14,0.6), (15,0.5), (16,0.4), (17,0.3), (18,0.2), (19,0.1)}

= { (x, 1 - |x-10|/10) | x$\in$U }

2.1.5 การเปรียบเทียบขั้นตอนวิธี

ในการเปรียบเทียบขั้นตอนวิธีการเรียนรู้สองวิธี ว่าวิธีการใดมีประสิทธิภาพที่ดีกว่านั้น Dietterich [5] ได้เสนอวิธีการเปรียบเทียบระหว่างสองขั้นตอนวิธี (ในที่นี้เปรียบเทียบระหว่าง ขั้นตอนวิธี A และ ขั้นตอนวิธี B) โดยดูจากค่าระดับความมั่นใจ (Confidence Level) ซึ่งมีวิธีการดังต่อไปนี้

1. แบ่งข้อมูลทั้งหมด ($D_0$) เป็น k ส่วน ซึ่งจะได้เซ็ตย่อย $T_1, T_2, \ldots, T_k$ มีขนาดเท่ากัน
2. ให้ i มีค่าเป็น 1 ถึง k
   - $S_i \leftarrow \{D_0 - T_i\}$ หมายความว่า ให้ $T_i$ เป็นข้อมูลในการทดสอบ และข้อมูลที่เหลือเป็นข้อมูลสอน $S_i$
   - $h_A \leftarrow L_A(S_i)$ หมายความว่า นำข้อมูลสอน $S_i$ มาเข้าสู่ขั้นตอนวิธี A (ในที่นี้แทนด้วย $L_A(S_i)$) หลังจากนั้นจะได้เป็นสมมติฐาน (hypothesis) (ในที่นี้หมายถึงต้นไม้ตัดสินใจ หรือนิวรอลเน็ตเวิร์ก) แทนด้วยสัญลักษณ์ $h_A$
   - $h_B \leftarrow L_B(S_i)$ หมายความว่า นำข้อมูลสอน $S_i$ มาเข้าสู่ขั้นตอนวิธี B (ในที่นี้แทนด้วย $L_B(S_i)$) หลังจากนั้นจะได้เป็นสมมติฐาน แทนด้วยสัญลักษณ์ $h_B$
   - $\delta_i \leftarrow error_{Ti}(h_A) - error_{Ti}(h_B)$ หมายความว่า หาค่าความผิดพลาดกับข้อมูลทดสอบของขั้นตอนวิธีแต่ละอย่าง และนำค่าความแตกต่างที่ได้เก็บเป็น $\delta_i$
3. คำนวณหาค่า $\bar{\delta}$ โดย

$$\bar{\delta} \equiv \frac{1}{k}\sum_{i=1}^{k}\delta_i$$

เมื่อได้ค่า $\bar{\delta}$ แล้วสามารถนำมาหาค่าระดับความมั่นใจว่ามีนัยสำคัญที่กี่เปอร์เซ็นต์จากสมการ

$$\bar{\delta} \pm t_{N,k-1}S_{\bar{\delta}}$$

$t_{N,k-1}$ เป็นค่าคงที่ซึ่งได้จากตารางแจกแจงแบบ t (t Distribution) (ดูภาคผนวก ง) โดยค่า N คือเปอร์เซ็นต์ระดับความมั่นใจ และ k-1 คือจำนวน degrees of freedom และ $S_{\bar{\delta}}$ มีค่าเป็น

$$S_{\bar{\delta}} \equiv \sqrt{\frac{1}{k(k-1)}\sum_{i=1}^{k}(\delta_i - \bar{\delta})^2}$$

ในขั้นตอนที่ 1 และ 2 ซึ่งเป็นขั้นตอนที่ใช้ในการแบ่งข้อมูลเป็นส่วน ๆ เพื่อหาสมมติฐานของแต่ละขั้นตอนวิธีนั้น เราเรียกว่าวิธีการครอสวาลิเดชัน (Cross Validation) [10] ซึ่งวิธีการเปรียบเทียบขั้นตอนวิธี ได้นำวิธีการครอสวาลิเดชันนี้เข้ามาเป็นส่วนหนึ่งของขั้นตอนการทำงาน

## 2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 การตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยการใช้แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก **โดย ก้องศักดิ์ จงเกษมวงศ์** [1]

งานวิจัยนี้นำเสนอวิธีการตัดเล็มต้นไม้ โดยไม่ได้ตัดส่วนของต้นไม้ออกไปจริงๆ แต่เป็นการประยุกต์ใช้แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ในการหาค่าน้ำหนัก (Weight) ให้แต่ละโนดของต้นไม้ตัดสินใจ ตามความสำคัญของแต่ละโนด และเรียกวิธีการตัดเล็มนี้ว่า การตัดเล็มอย่างอ่อน (Soft Pruning)

ผลจากการทดลองตัดเล็มอย่างอ่อนกับชุดข้อมูล 20 ชุดข้อมูล พบว่า

- เมื่อเปรียบเทียบกับต้นไม้ตัดสินใจที่ไม่ทำการตัดเล็ม ได้ผลดีกว่า 8 ชุดข้อมูล ด้อยกว่า 1 ชุดข้อมูล อย่างมีนัยสำคัญทางสถิติ
- เมื่อเปรียบเทียบกับต้นไม้ตัดสินใจที่ทำการตัดเล็มด้วยวิธี Error-Based Pruning แล้ว ได้ผลดีกว่า 9 ชุดข้อมูล และด้อยกว่า 3 ชุดข้อมูลอย่างมีนัยสำคัญทางสถิติ

2.1.2 A Comparative Analysis of Methods for Pruning Decision Trees **โดย Esposito, F., Malerba, D. และ Semeraro, G.** [6]

งานวิจัยนี้ทำการเปรียบเทียบการตัดเล็มต้นไม้ตัดสินใจ ด้วยวิธีการตัดเล็มต่างๆ ได้แก่

- วิธีการ Reduced Error Pruning (REP)
- วิธีการ Pessimistic Error Pruning (PEP)
- วิธีการ Minimum Error Pruning (MEP)
- วิธีการ Critical Value Pruning (CVP)
- วิธีการ Cost-Complexity Pruning (CCP)
- วิธีการ Error-Based Pruning (EBP)

โดยได้ดำเนินการเปรียบเทียบขั้นตอนวิธี (Algorithm) ข้อดี ข้อเสีย ตลอดจนความถูกต้องของต้นไม้ตัดสินใจสำหรับชุดข้อมูล 15 ชุด ที่ผ่านการตัดเล็มด้วยวิธีต่างๆ กัน และสรุปได้ว่าไม่มีวิธีการตัดเล็มใดที่มีประสิทธิภาพเหนือวิธีการอื่นๆ แต่วิธี Error-Based Pruning ได้ผลที่ดีโดยเฉลี่ยและมีความเสถียรกับชุดข้อมูลต่างๆ

2.1.3 Soft Decision Trees: A New Approach Using Non-linear Fuzzification **โดย Keeley A Crockett และ Zuhair Bandar [4]**

งานวิจัยนี้นำเสนอการใช้ Fuzzification กับต้นไม้ตัดสินใจ เพื่อลดขอบแบ่งของการตัดสินใจให้เกิดความคลุมเครือขึ้น โดยใช้ฟังก์ชันความเป็นสมาชิกต่างๆ 4 แบบ ได้แก่

- Sigmoid Function
- Cube Membership Function
- Convex Membership Function
- Concave Membership Function

และใช้ขั้นตอนวิธีแบบพันธุกรรม (Genetic Algorithm) ในการปรับความเหมาะสมของฟังก์ชันความเป็นสมาชิก

จากการทำการทดลองวัดผลความถูกต้องของต้นไม้ตัดสินใจที่ได้ บนชุดข้อมูล 2 ชุด พบว่ามีความถูกต้องมากขึ้นทั้ง 4 วิธีการ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 3

# ขั้นตอนวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน

ในบทนี้จะกล่าวถึงขั้นตอนวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน ซึ่งมีความแตกต่างจากวิธีการตัดเล็มอย่างอ่อนตามที่ได้อธิบายไปแล้ว ในส่วนของการสร้างข้อมูลที่ใช้กับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ซึ่งจะใช้ฟังก์ชันความเป็นสมาชิกเข้ามาช่วยในการสร้างข้อมูลนำเข้าโนดอินพุต โดยแบ่งออกได้เป็นการสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติที่เป็นข้อมูลต่อเนื่อง และการสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติที่เป็นข้อมูลไม่ต่อเนื่อง

## 3.1 แนวคิดในการใช้ฟัซซิฟิเคชันกับการตัดเล็มอย่างอ่อน

เนื่องจากในวิธีการตัดเล็มอย่างอ่อนเดิมนั้นจะเปรียบเทียบคุณสมบัติของข้อมูลกับเงื่อนไขของโนดแต่ละโนดในชั้นอินพุต หากเป็นจริงตามเงื่อนไข ค่าสำหรับโนดนั้นจะเป็น 1 มิฉะนั้นค่าสำหรับโนดนั้นจะเป็น 0 แต่ในความเป็นจริงหากค่าคุณสมบัตินั้นมีค่าที่ไม่แตกต่างจากค่าที่ใช้เปรียบเทียบมากนัก อาจจะกล่าวได้ว่าคุณสมบัตินั้นมีค่าของความถูกต้องตามเงื่อนไขของโนดและค่าของความไม่ถูกต้องตามเงื่อนไขของโนดไปพร้อมๆ กัน

ตัวอย่างเช่น จากรูปที่ 2.6 ซึ่งแสดงโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กในการตัดสินใจเล่นกอล์ฟนั้น จะมีโนดอินพุตที่เปรียบเทียบค่าคุณสมบัติความชื้นอยู่ 2 โนด ได้แก่โนดที่เปรียบเทียบ "ความชื้นน้อยกว่าหรือเท่ากับ 75%" และโนดที่เปรียบเทียบ "ความชื้นมากกว่า 75%" ในกรณีที่ความชื้นมีค่าใกล้เคียง 75% อาจจะกล่าวได้ว่าคุณสมบัติความชื้นไม่ได้โน้มเอียงไปทางโนดใดโนดหนึ่งอย่างเต็มที่ แต่ควรจะเกิดเป็นค่าความจริงในโนดทั้งสองในระดับที่มีค่าอยู่ระหว่าง 0 และ 1 ซึ่งเป็นระดับที่แทนการผิดและถูกตามเงื่อนไข ดังเช่นเมื่อความชื้นมีค่า 70% อาจจะกล่าวได้ว่ากรณีนี้มีค่า "ความชื้นมากกว่า 75%" มากกว่ากรณีที่ความชื้นมีค่า 50% แม้ว่าทั้งสองกรณีจะมีความชื้นน้อยกว่า 75% ก็ตาม

ด้วยแนวคิดนี้ วิทยานิพนธ์ฉบับนี้จึงเสนอการสร้างฟังก์ชันในการแปลงข้อมูลคุณสมบัติของชุดข้อมูล มาเป็นค่าสำหรับโนดอินพุตในโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ที่มีค่าความเป็นจริงอยู่ระหว่าง 0 ถึง 1 ซึ่งก็คือค่าความเป็นสมาชิกในเงื่อนไขของโนดอินพุตนั้นๆ ตามทฤษฎีเซตวิภัชนัย และฟังก์ชันที่ใช้ในการแปลงข้อมูลดังกล่าวก็คือฟังก์ชันความเป็นสมาชิกนั่นเอง

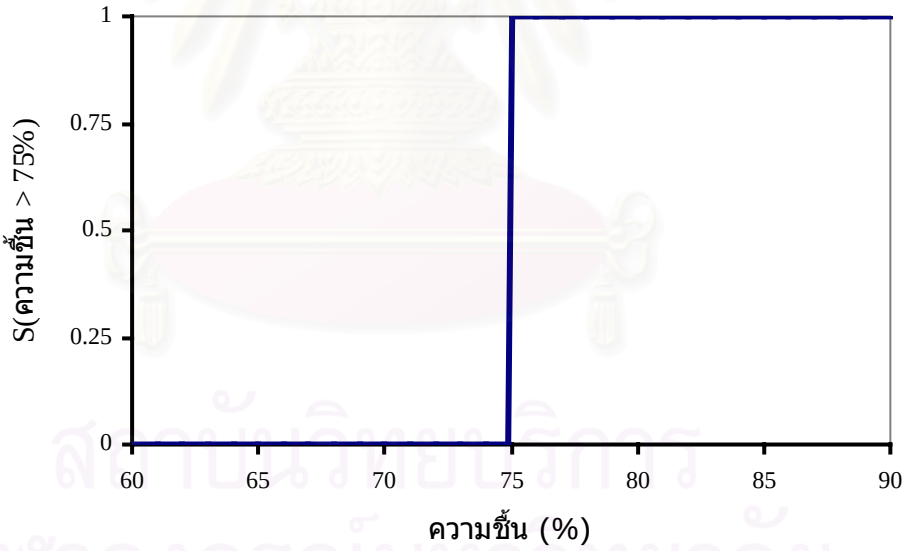
เนื่องจากต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ที่สามารถจำแนกข้อมูลได้จากทั้งคุณสมบัติแบบข้อมูลต่อเนื่อง และคุณสมบัติแบบข้อมูลไม่ต่อเนื่อง โดยจะมีลักษณะการเปรียบเทียบที่ไม่เหมือนกัน กล่าวคือในคุณสมบัติแบบข้อมูลต่อเนื่อง ต้นไม้ตัดสินใจจะหาค่าที่ใช้เป็นจุดแบ่งข้อมูล (Threshold Value) สำหรับเปรียบเทียบคุณสมบัตินั้น เช่นในกรณีคุณสมบัติความชื้นข้างต้นมีค่า

75% เป็นจุดแบ่งข้อมูลในการเปรียบเทียบว่าคุณสมบัตินั้นมีค่ามากกว่า หรือน้อยกว่า 75% แต่ในกรณีคุณสมบัติแบบข้อมูลไม่ต่อเนื่องนั้น ต้นไม้ตัดสินใจจะเปรียบเทียบค่าคุณสมบัตินั้นตรงๆ กับค่าที่เป็นไปได้ทุกกรณีว่าเป็นกรณีใด ดังนั้นจะเห็นได้ว่าเนื่องจากมีการเปรียบเทียบค่าคุณสมบัติที่แตกต่างกันตามชนิดคุณสมบัติ ทำให้การสร้างฟังก์ชันความเป็นสมาชิกที่เหมาะสมนั้นต้องมีความแตกต่างกันตามชนิดคุณสมบัติด้วย ดังต่อไปนี้

### 3.2 การสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติแบบข้อมูลต่อเนื่อง

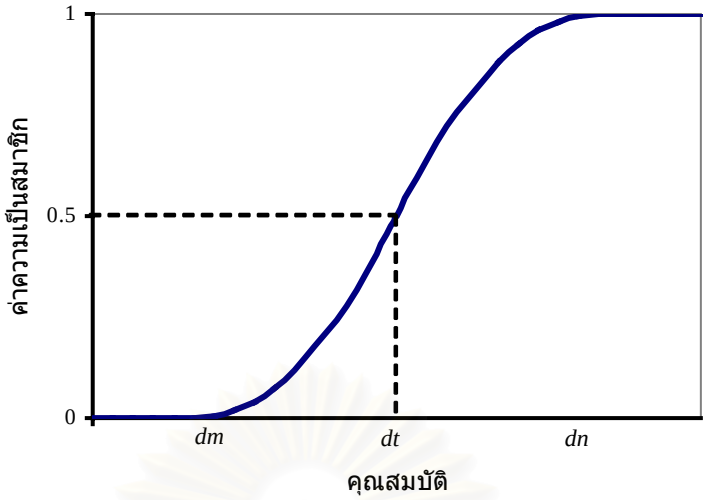
ตามที่ได้กล่าวมาแล้ว ในกรณีคุณสมบัติแบบข้อมูลต่อเนื่องนั้นการเปรียบเทียบของต้นไม้ตัดสินใจจะเปรียบเทียบค่าคุณสมบัติกับค่าที่ต้นไม้ตัดสินใจใช้เป็นจุดแบ่งข้อมูล ซึ่งจะเกิดการเปรียบเทียบขึ้น 2 กรณีคือ คุณสมบัติน้อยกว่า(หรือเท่ากับ)จุดแบ่งข้อมูล และคุณสมบัติมากกว่าจุดแบ่งข้อมูล

พิจารณากรณีที่ต้องการเปรียบเทียบว่าคุณสมบัตินั้นมีค่ามากกว่าจุดแบ่งข้อมูลหรือไม่ จากตัวอย่างรูปที่ 2.6 ที่โนดอินพุต “ความชื้นมากกว่า 75%” โดยวิธีการตัดเล็มอย่างอ่อนเดิมจะได้ฟังก์ชันระหว่างค่าคุณสมบัติ และค่าโนดอินพุตของโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กดังแสดงในรูปที่ 3.1




รูปที่ 3.1 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “ความชื้น > 75%” จากวิธีการตัดเล็มอย่างอ่อน โดยไม่ใช้ฟัซซิฟิเคชัน

แต่เนื่องจากเราต้องการให้ค่าความชื้นที่มีค่าใกล้เคียงกับ 75% ควรจะเกิดมีค่าความเป็นสมาชิกในช่วงระหว่าง 0 ถึง 1 ขึ้น ในวิทยานิพนธ์ฉบับนี้ได้เสนอให้ใช้ฟังก์ชันที่มีลักษณะเป็นรูปตัวเอส (S) ดังรูปที่ 3.2




รูปที่ 3.2 กราฟฟังก์ชันความเป็นสมาชิกรูปตัวเอส สำหรับโนดอินพุตที่เปรียบเทียบคุณสมบัติแบบข้อมูลต่อเนื่อง

จะเห็นได้จากฟังก์ชันในรูป 3.2 ว่าสามารถแบ่งช่วงของฟังก์ชันออกได้เป็น 3 ช่วง คือช่วงที่ฟังก์ชันมีค่าเป็น 0 ช่วงที่ฟังก์ชันมีค่าเป็น 1 และช่วงที่ฟังก์ชันมีค่าอยู่ในช่วง [0, 1] หรือช่วงที่เกิดการฟัซซิฟิเคชันขึ้นนั่นเอง โดย K. A. Crockett และ Z. Bandar [4] ได้เสนอวิธีการสร้างฟังก์ชันความเป็นสมาชิกในลักษณะนี้ไว้ดังนี้

การคำนวณหาช่วงที่เกิดฟัซซิฟิเคชันทำได้โดยกำหนดให้ dm และ dn เป็นค่าคุณสมบัติที่เป็นขอบเขตบนและล่างของการเกิดฟัซซิฟิเคชัน ตามลำดับ และให้ dt คือจุดแบ่งของคุณสมบัติ $\sigma$ คือค่าเบี่ยงเบนมาตรฐานของคุณสมบัตินั้นจากข้อมูลสอน จะได้ว่า

$$dm = dt - n_m\sigma \quad \text{และ} \quad dn = dt + n_n\sigma$$

โดย $n_m$ และ $n_n$ คือค่าใดๆ ที่อยู่ในช่วง $[0,\infty]$ ซึ่งแสดงถึงระดับของการทำฟัซซิฟิเคชัน โดยสามารถกำหนดให้สอดคล้องเหมาะสมกับคุณสมบัติแต่ละตัวในชุดข้อมูลหนึ่ง ๆ ได้ แต่ในการทดลองในวิทยานิพนธ์ฉบับนี้ จะใช้ฟังก์ชันความเป็นสมาชิกที่มีช่วงการเกิดฟัซซิฟิเคชันเท่ากันทั้งสองข้างของจุดเปรียบเทียบ ดังนั้นจึงใช้ $n_m=n_n=n$

เมื่อได้ช่วงของการเกิดฟัซซิฟิเคชัน dm และ dn แล้ว สามารถสร้างฟังก์ชันความเป็นสมาชิกได้ดังนี้

$$S(x, dm, dn, \beta) = \begin{cases} 0 & ; x \le dm \\ 2\left(\dfrac{x - dm}{dn - dm}\right)^2 & ; dm < x \le \beta \\ 1 - 2\left(\dfrac{x - dn}{dn - dm}\right)^2 & ; \beta < x \le dn \\ 1 & ; x > dn \end{cases}$$

โดยมีค่า x คือค่าของคุณสมบัติที่พิจารณา

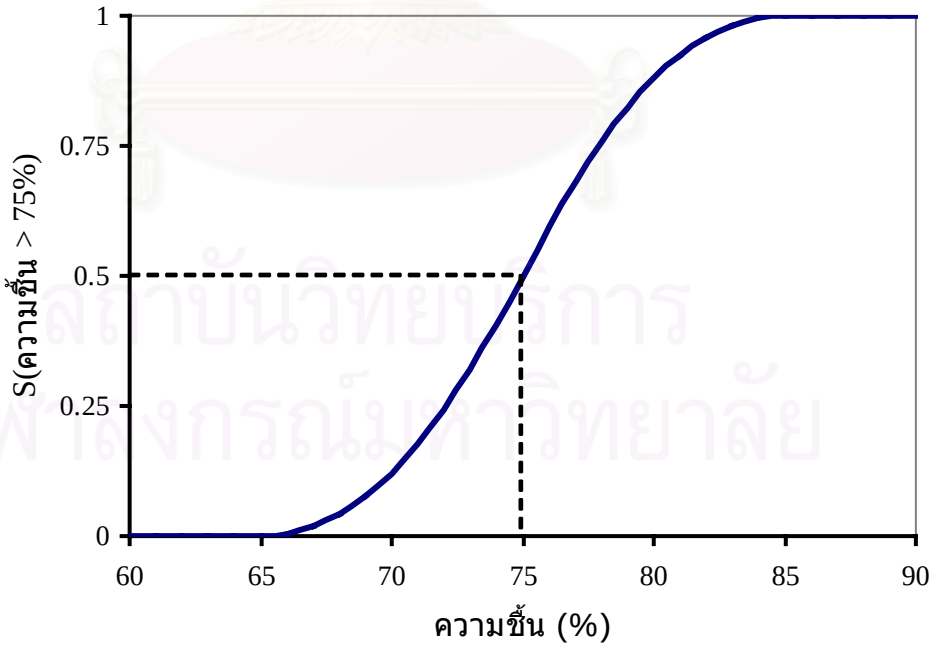
dm และ dn เป็นค่าคุณสมบัติที่เป็นขอบเขตบนและล่างของการเกิดฟัซซิฟิเคชันตามลำดับ

β คือค่ากึ่งกลางระหว่าง dm และ dn ซึ่งเรียกว่าจุดไขว้เปลี่ยน (Crossover Point) ซึ่งในกรณีที่เราใช้ $n_m$=$n_n$ ค่า β ก็คือค่าที่จุดเปรียบเทียบ dt นั่นเอง

ในกรณีของการเปรียบเทียบว่าค่าคุณสมบัติมีค่าน้อยกว่าจุดแบ่งข้อมูลหรือไม่ ก็สามารถสร้างฟังก์ชันความเป็นสมาชิกจากฟังก์ชันที่ตรงกันข้ามกับที่กล่าวมาข้างต้นนั่นเอง กล่าวคือเราสามารถใช้ค่า dm และ dn เดียวกันกับกรณีแรก และสร้างฟังก์ชันความเป็นสมาชิกดังนี้

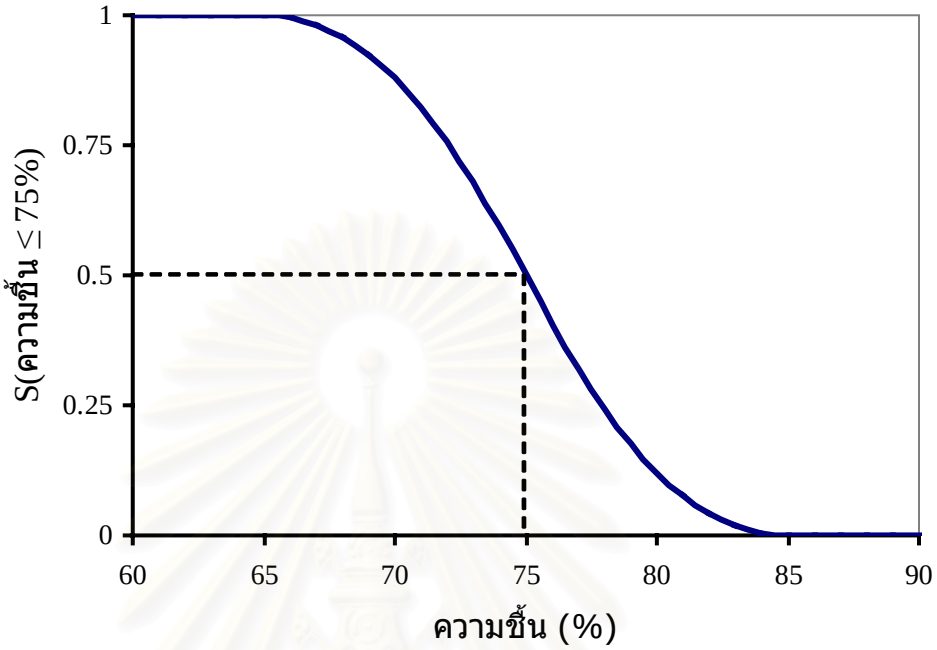
$$S(x,dm,dn,\beta)=\begin{cases}1 & ;x\le dm\\ 1-2\left(\dfrac{x-dm}{dn-dm}\right)^2 & ;dm<x\le\beta\\ 2\left(\dfrac{x-dn}{dn-dm}\right)^2 & ;\beta<x\le dn\\ 0 & ;x>dn\end{cases}$$

จากตัวอย่างที่กล่าวมาข้างต้น ในกรณีของโนดอินพุต “ความชื้นมากกว่า 75%” เราสามารถสร้างฟังก์ชันความเป็นสมาชิกด้วยวิธีการที่กล่าวมาข้างต้น โดยใช้ค่า n เป็น 1 จะได้ฟังก์ชันความเป็นสมาชิกดังแสดงในรูปที่ 3.3




รูปที่ 3.3 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “ความชื้น > 75%”

และในกรณีการเปรียบเทียบที่ตรงข้ามกัน คือโนดอินพุต “ความชื้นน้อยกว่าหรือเท่ากับ 75%” เราจะสามารถสร้างฟังก์ชันความเป็นสมาชิกได้ดังในรูปที่ 3.4 โดยใช้ค่า n เป็น 1 เช่นกัน




รูปที่ 3.4 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “ความชื้น ≤ 75%”

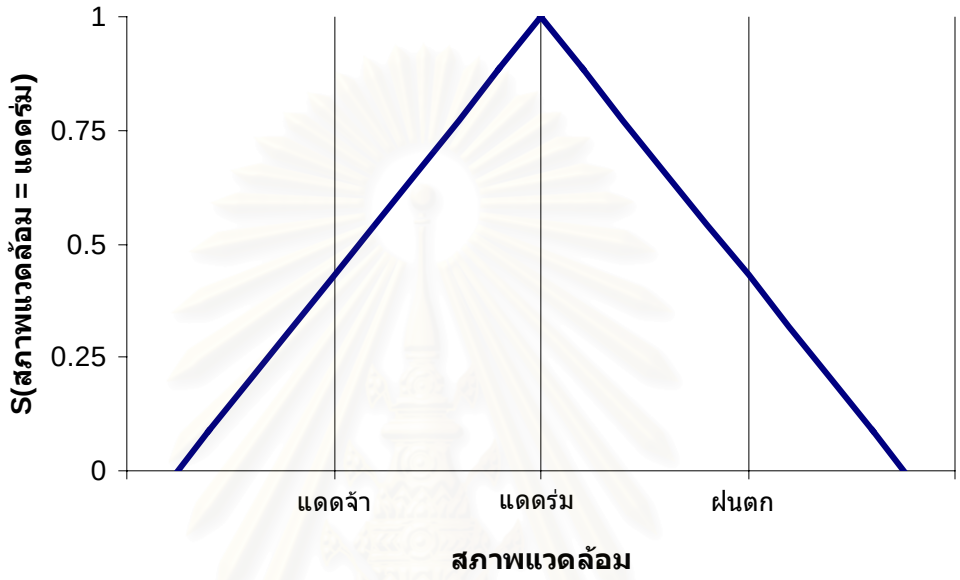
### 3.3 การสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติแบบข้อมูลไม่ต่อเนื่อง

การเปรียบเทียบคุณสมบัติแบบข้อมูลไม่ต่อเนื่องในต้นไม้ตัดสินใจ จะทำการเปรียบเทียบคุณสมบัติกับทุกค่าที่เป็นไปได้ของคุณสมบัตินั้นว่าเป็นกรณีใด หากค่าคุณสมบัตินั้นเป็นค่าใดโนดอินพุตสำหรับค่านั้นจะมีค่าเป็น 1 และโนดอินพุตอื่นๆ ของคุณสมบัติเดียวกันจะมีค่าเป็น 0

ในวิทยานิพนธ์ฉบับนี้ ได้เสนอแนวคิดในการใช้ฟัซซิฟิเคชันสำหรับคุณสมบัติแบบข้อมูลไม่ต่อเนื่องด้วย เนื่องจากหากคุณสมบัติมีค่าที่เป็นได้หลายค่า เมื่อคุณสมบัติมีค่าค่าหนึ่ง อาจจะมีความเป็นค่าอื่นที่มีความใกล้เคียงกับค่านั้นๆ อยู่เล็กน้อยด้วยก็ได้ แต่ด้วยแนวคิดนี้ แสดงให้เห็นว่าไม่ใช่คุณสมบัติแบบข้อมูลไม่ต่อเนื่องทุกตัวที่จะสามารถทำฟัซซิฟิเคชันได้ แต่ต้องเป็นคุณสมบัติที่เรารู้ถึงความใกล้เคียงกันของค่าแต่ละค่าที่เป็นไปได้

ดังเช่น จากตัวอย่างรูปที่ 2.6 สำหรับคุณสมบัติสภาพแวดล้อมนั้น จะมีค่าที่เป็นไปได้ 3 กรณีคือ *แดดจ้า แดดร่ม* และ*ฝนตก* ซึ่งในกรณีนี้เราสามารถทราบความใกล้เคียงของค่าที่เป็นไปได้ทั้งหมดได้ และสามารถเรียงลำดับเป็น *แดดจ้า แดดร่ม* และ*ฝนตก* ตามลำดับ เนื่องจาก*ฝนตก*มีความใกล้เคียงกับ*แดดร่ม* มากกว่า*แดดจ้า* และจะสังเกตได้ว่าเราสามารถจัดเรียงลำดับความใกล้เคียงในทางตรงข้ามเป็น *ฝนตก แดดร่ม* และ*แดดจ้า* ก็ได้เช่นกัน

จากที่กล่าวข้างต้น กล่าวได้ว่าการทำฟัซซิฟิเคชันสำหรับคุณสมบัติแบบข้อมูลไม่ต่อเนื่องนั้นจะเป็นการสร้างฟังก์ชันความเป็นสมาชิกที่ให้ค่าความเป็นสมาชิกสำหรับค่าที่มีความใกล้เคียงกับค่าคุณสมบัตินั่นเอง ซึ่งฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม (Triangular shape) จะมีความเหมาะสมสำหรับคุณสมบัติลักษณะนี้ ดังแสดงในรูป 3.5 ที่เป็นฟังก์ชันความเป็นสมาชิกที่ต้องการสำหรับโนดอินพุต “สภาพแวดล้อมเท่ากับแดดร่ม”




รูปที่ 3.5 กราฟฟังก์ชันความเป็นสมาชิกที่ต้องการสำหรับโนดอินพุต “สภาพแวดล้อม = แดดร่ม”

สำหรับวิทยานิพนธ์ฉบับนี้จะเสนอการสร้างฟังก์ชันความเป็นสมาชิกสำหรับคุณสมบัติแบบข้อมูลไม่ต่อเนื่องดังนี้

เมื่อทราบลำดับความใกล้เคียงของค่าที่เป็นไปได้ทั้งหมดแล้ว ต้องเปลี่ยนค่าที่เป็นไปได้ทั้งหมดให้เป็นเลขจำนวนจริง (Real Value) ตามความเหมาะสมก่อน โดยหากสมมุติให้ระดับความใกล้เคียงของค่าที่เป็นไปได้ที่ติดกันมีค่าเท่ากันทั้งหมดทุกคู่ จะสามารถเปลี่ยนค่าที่เป็นไปได้เป็นค่าเลขจำนวนเต็มที่เรียงลำดับต่อกันได้ อาทิเช่นในกรณีคุณสมบัติสภาพแวดล้อมนี้ อาจจะเปลี่ยนค่า *แดดจ้า* *แดดร่ม* และ*ฝนตก* เป็นตัวเลข 1 2 และ 3 ได้ตามลำดับ ซึ่งการแปลงค่าให้เป็นเลขจำนวนจริง หรือจำนวนเต็มนี้ จะช่วยให้สามารถทำการคำนวณต่างๆ เช่นค่าเบี่ยงเบนมาตรฐานได้

จากนั้นเราจะสามารถคำนวณช่วงที่ต้องการให้เกิดฟัซซิฟิเคชันได้ เหมือนในกรณีการคำนวณ dm และ dn สำหรับคุณสมบัติแบบข้อมูลต่อเนื่องที่ได้อธิบายไว้ข้างต้น

เมื่อทราบช่วงของการเกิดฟัซซิฟิเคชัน จะสามารถสร้างฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมได้ดังนี้

$$S(x, dm, dn, \beta) = \begin{cases} 2\left(\dfrac{x-\beta}{dn-dm}\right)+1 & ; dm < x \le \beta \\ 2\left(\dfrac{\beta-x}{dn-dm}\right)+1 & ; \beta < x \le dn \\ 0 & ; otherwise \end{cases}$$

โดยมีค่า x คือค่าของคุณสมบัติที่พิจารณา

dm และ dn เป็นค่าคุณสมบัติที่เป็นขอบเขตบนและล่างของการเกิดฟัซซิฟิเคชันตามลำดับ

β คือค่ากึ่งกลางระหว่าง dm และ dn

จากตัวอย่างที่กล่าวมาข้างต้น ในกรณีโนดอินพุต “สภาพแวดล้อมเท่ากับแดดร่ม” เราเปลี่ยนค่าที่เป็นไปได้ทั้งหมดของคุณสมบัติคือ *แดดจ้า แดดร่ม* และ*ฝนตก* เป็นตัวเลขจำนวนเต็ม 1 2 และ 3 ได้ตามลำดับ จะสามารถสร้างฟังก์ชันความเป็นสมาชิกด้วยวิธีการที่กล่าวมาข้างต้น โดยใช้ค่า n มีค่าเป็น 2 ได้ฟังก์ชันความเป็นสมาชิกดังแสดงในรูปที่ 3.6

รูปที่ 3.6 กราฟฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุต “สภาพแวดล้อม = แดดร่ม”

จากฟังก์ชันความเป็นสมาชิกที่ได้นี้ จะเห็นว่าเมื่อสภาพแวดล้อมเป็นแดดร่ม ค่าสำหรับโนดอินพุต “สภาพแวดล้อม = แดดร่ม” นี้จะเป็น 1 แต่เมื่อสภาพแวดล้อมเป็น แดดจ้า หรือ ฝนตก ค่าสำหรับโนดอินพุตนี้จะเป็น 0.43

# บทที่ 4

# การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทดลองการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน โดยการเปรียบเทียบความถูกต้องในการจำแนกข้อมูลจากวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน บนชุดข้อมูลทั้งหมด 17 ชุด กับค่าความถูกต้องที่ได้จากต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็ม และต้นไม้ตัดสินใจที่ทำการตัดเล็มโดยใช้ค่าความผิดพลาดแล้ว รวมทั้งค่าความถูกต้องที่ได้จากการตัดเล็มอย่างอ่อนเดิมด้วย หลังจากนั้นจะคำนวณเปรียบเทียบขั้นตอนวิธี ระหว่างการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน กับการตัดเล็มอย่างอ่อน ต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็ม รวมทั้งต้นไม้ตัดสินใจที่ตัดเล็มแล้ว

## 4.1 วิธีการทดลอง

### 4.1.1 ชุดข้อมูลในการทดลอง

ในการทดลองเพื่อเปรียบเทียบวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน กับวิธีการตัดเล็มต้นไม้ตัดสินใจอื่นๆ นี้ เราจะเลือกชุดข้อมูลจาก Department of Information and Computer Science, University of California, Irvine [9] ซึ่งเป็นที่รวบรวมชุดข้อมูลสำหรับเป็นเกณฑ์เปรียบเทียบสมรรถนะ (Benchmark) โดยในการเลือกชุดข้อมูลพิจารณาจาก

- จำนวนกลุ่มของชุดข้อมูล โดยในการทดลองจะเลือกชุดข้อมูลที่มีตั้งแต่ 3 กลุ่มขึ้นไป
- ข้อมูลสามารถนำมาใช้งานได้ทันทีหรือเปลี่ยนแปลงเพียงเล็กน้อย โดยข้อมูลสามารถนำไปใช้งานใน ระบบ C4.5 ซึ่งเป็นโปรแกรมที่ใช้ในการสร้างต้นไม้ตัดสินใจที่จะใช้ในการทดลองของวิทยานิพนธ์ฉบับนี้
- ลักษณะคุณสมบัติของชุดข้อมูลจะต้องเป็นคุณสมบัติแบบข้อมูลต่อเนื่อง หรือข้อมูลไม่ต่อเนื่องเหมือนกันตลอดทั้งชุดข้อมูล
- ไม่เลือกชุดข้อมูลที่มีคุณสมบัติแบบข้อมูลไม่ต่อเนื่องที่มีค่าที่เป็นไปได้แค่ 2 ค่าตลอดทั้งชุดข้อมูล เนื่องจากจะไม่ทำการฟัซซิฟิเคชันกับคุณสมบัติดังกล่าวในการทดลอง แต่เลือกชุดข้อมูลที่มีคุณสมบัติดังกล่าวปนกับคุณสมบัติแบบอื่นได้

ชุดข้อมูลที่เลือกมามีทั้งหมด 17 ชุดโดยมีรายละเอียดดังตารางที่ 4.1 โดยที่สามารถแบ่งชุดข้อมูลออกได้เป็น 2 ชนิดดังนี้

- ชุดข้อมูลแบบข้อมูลไม่ต่อเนื่อง คือชุดข้อมูลที่คุณสมบัติทั้งหมดเป็นแบบข้อมูลไม่ต่อเนื่อง ซึ่งได้แก่ชุดข้อมูล 4 ชุดแรกในตารางที่ 4.1
- ชุดข้อมูลแบบข้อมูลต่อเนื่อง คือชุดข้อมูลที่คุณสมบัติทั้งหมดเป็นแบบข้อมูลต่อเนื่องหรือเป็นแบบข้อมูลต่อเนื่อง ปนกับคุณสมบัติที่มีค่าที่เป็นไปได้แค่ 2 ค่าซึ่งไม่ทำการฟัซซิฟิเคชัน อันได้แก่ชุดข้อมูล 13 ชุดท้ายในตารางที่ 4.1

ตารางที่ 4.1 ชุดข้อมูลที่ใช้ในการทดลอง

| ชื่อชุดข้อมูล | จำนวนข้อมูลสอน | จำนวนข้อมูลทดสอบ | จำนวนกลุ่มข้อมูล | จำนวนคุณสมบัติ | | | | ข้อมูลไม่ทราบค่า |
|---|---|---|---|---|---|---|---|---|
| | | | | ทั้งหมด | ข้อมูลต่อเนื่อง | ข้อมูลไม่ต่อเนื่อง | ที่ทำฟัซ-ซิฟิเคชัน | |
| Soybean | 307 | 376 | 19 | 35 | 0 | 35 | 11 | มี |
| Balance-scale | 625 | (6CV) | 3 | 4 | 0 | 4 | 4 | ไม่มี |
| Restricted (lymphography) | 148 | (6CV) | 4 | 18 | 0 | 18 | 9 | ไม่มี |
| Restricted (primary-tumor) | 339 | (6CV) | 22 | 17 | 0 | 17 | 3 | มี |
| Thyroid-disease (allbp) | 2800 | 972 | 3 | 29 | 7 | 22 | 7 | มี |
| Thyroid-disease (allhyper) | 2800 | 972 | 5 | 29 | 7 | 22 | 7 | มี |
| Thyroid-disease (allhypo) | 2800 | 972 | 5 | 29 | 7 | 22 | 7 | มี |
| Thyroid-disease (allrep) | 2800 | 972 | 4 | 29 | 7 | 22 | 7 | มี |
| Image | 210 | 2100 | 7 | 19 | 19 | 0 | 19 | ไม่มี |
| Statlog (satimage) | 4435 | 2000 | 6 | 36 | 36 | 0 | 36 | ไม่มี |
| Statlog (shuttle) | 43500 | 14500 | 7 | 9 | 9 | 0 | 9 | ไม่มี |
| Statlog (segment) | 2310 | (6CV) | 7 | 19 | 19 | 0 | 19 | ไม่มี |
| Iris | 150 | (6CV) | 3 | 4 | 4 | 0 | 4 | ไม่มี |
| Wine | 178 | (6CV) | 3 | 13 | 13 | 0 | 13 | ไม่มี |
| Waveform | 5000 | (6CV) | 3 | 21 | 21 | 0 | 21 | ไม่มี |
| Waveform + noise | 5000 | (6CV) | 3 | 40 | 40 | 0 | 40 | ไม่มี |
| Glass | 214 | (6CV) | 6 | 9 | 9 | 0 | 9 | ไม่มี |

จากตารางที่ 4.1 นี้จะเห็นได้ว่าในการทดลองอาจไม่ได้ทำฟัซซิฟิเคชันกับคุณสมบัติทุกตัวในชุดข้อมูลหนึ่งๆ ก็ได้ โดยคุณสมบัติที่เลือกไม่ทำฟัซซิฟิเคชันนั้นได้แก่ คุณสมบัติที่มีค่าที่เป็นไปได้เพียง 2 ค่า และคุณสมบัติแบบข้อมูลไม่ต่อเนื่องซึ่งไม่ทราบลำดับความใกล้เคียงของค่าที่เป็นไปได้ทั้งหมด

4.1.2 การแบ่งข้อมูลสอน และข้อมูลทดสอบ

ชุดข้อมูลที่เลือกมาใช้ในการทดลองสามารถแบ่งออกได้เป็นสองประเภท ตามลักษณะของข้อมูลสอนและข้อมูลทดสอบดังนี้

1. ชุดข้อมูลที่แบ่งข้อมูลสอนและข้อมูลทดสอบไว้แล้ว สำหรับชุดข้อมูลชนิดนี้การทดลองจะใช้ข้อมูลสอนสร้างต้นไม้ตัดสินใจ รวมไปถึงใช้ในการสอนนิวรอลเน็ตเวิร์กสำหรับการตัดเล็มอย่างอ่อน จากนั้นจะทดสอบความถูกต้องของวิธีการต่างๆ ด้วยข้อมูลทดสอบตามที่ได้มีการแบ่งไว้แล้ว

2. ชุดข้อมูลที่ไม่มีการแบ่งข้อมูลสอนและข้อมูลทดสอบ ซึ่งชุดข้อมูลเหล่านี้จะถูกกำกับด้วยเครื่องหมาย (6CV) ในตารางที่ 4.1 โดยในหัวข้อจำนวนข้อมูลสอน จะแสดงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้น ในการทดลองในวิทยานิพนธ์ฉบับนี้จะใช้วิธีการครอสวาลิเดชัน (Cross Validation) [10] (ขั้นตอนที่ 1 และ 2 ในหัวข้อ 2.1.5) กับข้อมูลในลักษณะนี้ โดยจะแบ่งชุดข้อมูลออกเป็น 6 โฟลด์ (fold) หลังจากนั้นแต่ละโฟลด์ที่ได้จะใช้ในสร้างต้นไม้ตัดสินใจและโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กต่อไป

4.1.3 กระบวนการทดลอง

ในชุดข้อมูลแต่ละชุด จะใช้ข้อมูลสอนในการสร้างต้นไม้ตัดสินใจด้วยวิธี C4.5 ซึ่งจะได้ผลลัพธ์เป็นทั้งต้นไม้ตัดสินใจที่ยังไม่ได้รับการตัดเล็ม และต้นไม้ตัดสินที่ทำการตัดเล็มแล้วจากโปรแกรม C4.5 จากนั้นนำต้นไม้ตัดสินใจที่ยังไม่ได้รับการตัดเล็มมาสร้างเป็นแบ็กพรอพาเกชัน-นิวรอลเน็ตเวิร์กตามวิธีการตัดเล็มอย่างอ่อน และใช้ข้อมูลสอนชุดเดิมในการสร้างฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุตแต่ละโนด จากนั้นใช้ข้อมูลสอนในการสอนแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์กทั้งแบบที่ใช้ และไม่ใช้ฟังก์ชันความเป็นสมาชิก

ในการสร้างฟังก์ชันความเป็นสมาชิกสำหรับการทดลองนี้จะสร้างตามวิธีที่อธิบายในบทที่ 3 โดยใช้ค่าพารามิเตอร์ n มีค่าเท่ากับ 1 สำหรับชุดข้อมูลทุกชุด ยกเว้นชุดข้อมูล Restricted (primary-tumor) ซึ่งมีค่าเบี่ยงเบนมาตรฐานที่น้อยเนื่องมาจากคุณสมบัติต่างๆ ในชุดข้อมูลเป็นแบบข้อมูลไม่ต่อเนื่องที่มีค่าที่เป็นไปได้ 2 หรือ 3 ค่าเท่านั้น ทำให้หากใช้ค่า n เป็น 1 จะทำให้ช่วงของการเกิดฟัซซิฟิเคชันแคบจนทำให้ไม่เกิดการฟัซซิฟิเคชันขึ้น ดังนั้นจึงใช้ n มีค่าเท่ากับ 3 สำหรับชุดข้อมูลนี้เท่านั้น

จากนั้นจะทดสอบค่าความถูกต้องด้วยข้อมูลทดสอบ โดยการวัดผลการจำแนกข้อมูลจากต้นไม้ตัดสินใจด้วยวิธี C4.5 ทั้งที่ได้รับการตัดเล็มแล้ว และยังไม่ได้รับการตัดเล็ม และวัดผลการจำแนกข้อมูลจากแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ทั้งแบบที่ใช้ และไม่ใช้ฟังก์ชันความเป็นสมาชิก

วิธีการทดสอบค่าความถูกต้องบนแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จะใช้วิธีการเลือกโนดของเอาต์พุตหนึ่งโนดที่มีค่าสูงสุดเป็นโนดที่แสดงถึงกลุ่มที่ได้จากการจำแนกตัวอย่างนั้นๆ สาเหตุที่ใช้วิธีการนี้เนื่องจากการกำหนดข้อมูลจากโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กที่สร้างไว้ในส่วนของเอาต์พุต เป็นการกำหนดให้แต่ละกลุ่มที่เป็นไปได้ เป็นโนดหนึ่งโนดในชั้นเอาต์พุต ดังนั้นตัวอย่างของข้อมูลหนึ่งตัวอย่างสามารถที่จะเป็นกลุ่มใดกลุ่มหนึ่งได้เพียงกลุ่มเดียวเท่านั้น ดังนั้นวิธีการเลือกโนดที่มีค่าสูงสุดในชั้นเอาต์พุตจึงสามารถใช้ในการทดลองนี้ได้

## 4.2 เครื่องมือในการทดลอง

การทดลองในงานวิจัยนี้ใช้โปรแกรม C4.5 [12] สำหรับการสร้างต้นไม้ตัดสินใจ รวมไปถึงต้นไม้ตัดสินใจที่ทำการตัดเล็มโดยใช้ค่าความผิดพลาด และใช้โปรแกรม Aspirin/MIGRAINES Neural Network Software รุ่นที่ 6 [8] สำหรับการสร้าง การสอน และการทดสอบโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

นอกจากนี้ผู้วิจัยได้พัฒนาโปรแกรมขึ้นสำหรับงานวิจัยนี้ (ดูการใช้งานโปรแกรม ในภาคผนวก ข) โดยพัฒนาต่อจากโปรแกรมจากงานวิจัยและวิทยานิพนธ์ของ ก้องศักดิ์ จงเกษมวงศ์ [1] และเพิ่มการสร้างฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุตเพื่อใช้ในการวิจัยการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน โดยในการทดลองในงานวิจัยนี้จะอาศัยโปรแกรมนี้ในการสร้างฟังก์ชันความเป็นสมาชิก และแปลงข้อมูลสอนและข้อมูลทดสอบด้วยฟังก์ชันความเป็นสมาชิกดังกล่าวให้เป็นข้อมูลสอนและข้อมูลทดสอบสำหรับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก เพื่อนำข้อมูลนี้ไปใช้ในการสอน และการทดสอบโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ด้วยโปรแกรม Aspirin/MIGRAINES Neural Network Software ต่อไป

## 4.3 ผลการทดลอง

จากการทดลองจะได้ค่าความถูกต้องของวิธีการตัดเล็มต้นไม้ตัดสินใจวิธีต่างๆ รวมทั้งค่าความถูกต้องจากต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็ม ดังแสดงในตารางที่ 4.2 โดยในชุดข้อมูลแต่ละชุดวิธีการที่ให้ค่าความถูกต้องสูงสุดจะถูกแสดงด้วยค่าความถูกต้องตัวหนา (bold)

ตารางที่ 4.2 ผลการทดสอบค่าความถูกต้องของวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันเทียบกับต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็มและตัดเล็มแล้ว และวิธีการตัดเล็มอย่างอ่อน

| ชื่อชุดข้อมูล | ค่าความถูกต้อง (%) จากวิธี | | | |
|---|---|---|---|---|
| | C4.5 ไม่ได้ตัดเล็ม | C4.5 ตัดเล็มแล้ว | ตัดเล็มอย่างอ่อน | ตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน |
| Soybean | 85.64 | 86.70 | 89.89 | **91.49** |
| Balance-scale | 69.60 | 65.77 | 90.24 | **91.04** |
| Restricted (lymphography) | 74.36 | 78.39 | **80.42** | 79.72 |
| Restricted (primary-tumor) | 42.49 | 41.61 | 43.37 | **45.72** |
| Thyroid-disease (allbp) | 96.81 | **97.84** | 97.53 | 97.63 |
| Thyroid-disease (allhyper) | **98.97** | 98.56 | **98.97** | 98.46 |
| Thyroid-disease (allhypo) | 99.49 | 99.49 | 98.87 | **99.79** |
| Thyroid-disease (allrep) | 98.56 | **99.07** | 98.97 | 98.25 |
| Image | 89.43 | 91.00 | 90.38 | **93.76** |
| Statlog (satimage) | 85.20 | 85.90 | 87.40 | **88.20** |
| Statlog (shuttle) | 99.95 | 99.95 | **99.99** | **99.99** |
| Statlog (segment) | 96.93 | 96.93 | 96.75 | **98.05** |
| Iris | 95.33 | 94.00 | 95.33 | **98.00** |
| Wine | 93.30 | 93.30 | 93.30 | **96.63** |
| Waveform | 76.00 | 76.00 | 81.08 | **82.68** |
| Waveform + noise | 75.04 | 75.18 | 80.74 | **82.56** |
| Glass | 66.34 | 66.34 | 68.23 | **74.30** |

หลังจากทดสอบค่าความถูกต้องที่ได้จากโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กและฟังก์ชันความเป็นสมาชิกแล้ว ในงานวิทยานิพนธ์ฉบับนี้จะนำค่าความถูกต้องมาเปรียบเทียบขั้นตอนวิธีกับต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็ม ต้นไม้ตัดสินใจที่ทำการตัดเล็มแล้ว และการตัดเล็มอย่างอ่อนด้วย เพื่อเปรียบเทียบว่าวิธีการใดมีประสิทธิภาพที่ดีกว่า ในการทดลองนี้ใช้วิธีการเปรียบเทียบสองขั้นตอนวิธีโดยดูจากค่าระดับความมั่นใจ (Confidence Level) ซึ่งเป็นการวัดความสำคัญทางสถิติว่าวิธีการหนึ่งดีกว่าอีกวิธีการหนึ่งด้วยความมั่นใจเท่าใด วัดเป็นหน่วยเปอร์เซ็นต์ ผลการทดลองที่ได้ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 การเปรียบเทียบผลการทดลองด้วยวิธีค่าระดับความมั่นใจ ระหว่างวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน กับต้นไม้ตัดสินใจที่ยังไม่ได้ตัดเล็มและตัดเล็มแล้ว และวิธีการตัดเล็มอย่างอ่อนโดยไม่ใช้ฟัซซิฟิเคชัน

| ชื่อชุดข้อมูล | ค่าระดับความมั่นใจของวิธีการตัดเล็มอย่างอ่อนใช้ฟัซซิฟิเคชันเทียบกับ | | |
|---|---|---|---|
| | C4.5 ไม่ได้ตัดเล็ม | C4.5 ตัดเล็มแล้ว | ตัดเล็มอย่างอ่อน |
| Soybean | +99.43% | +98.27% | = |
| Balance-scale | +99.99% | +99.99% | +88.76% |
| Restricted (lymphography) | +94.82% | = | = |
| Restricted (primary-tumor) | +94.52% | +93.85% | +87.47% |
| Thyroid-disease (allbp) | +86.53% | = | = |
| Thyroid-disease (allhyper) | = | = | = |
| Thyroid-disease (allhypo) | +87.21% | +87.21% | +99.39% |
| Thyroid-disease (allrep) | = | -94.30% | -91.26% |
| Image | +99.99% | +99.96% | +99.99% |
| Statlog (satimage) | +99.74% | +98.49% | = |
| Statlog (shuttle) | +95.22% | +95.22% | = |
| Statlog (segment) | +97.22% | +97.22% | +97.48% |
| Iris | +94.90% | +94.47% | +94.90% |
| Wine | +94.35% | +94.35% | +94.35% |
| Waveform | +99.99% | +99.99% | +99.71% |
| Waveform + noise | +99.96% | +99.96% | +97.35% |
| Glass | +99.21% | +99.21% | +99.20% |

ในตารางที่ 4.3 นี้จะแสดงค่าระดับความมั่นใจของความถูกต้องของวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน เปรียบเทียบกับวิธีการอื่นๆ โดยสามารถแบ่งออกได้เป็น 3 กรณี

- กรณีที่วิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน ให้ความถูกต้องที่ไม่แตกต่างจากวิธีการอื่นที่เปรียบเทียบอย่างมีนัยสำคัญทางสถิติ ซึ่งในงานวิจัยนี้จะเปรียบเทียบว่าระดับความมั่นใจที่น้อยกว่า 85% ถือว่าไม่มีนัยสำคัญทางสถิติ และจะแสดงด้วยเครื่องหมายเท่ากับ (=) ในตาราง
- กรณีที่วิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน มีความถูกต้องกว่าวิธีการอื่นที่เปรียบเทียบอย่างมีนัยสำคัญทางสถิติ จะแสดงด้วยเครื่องหมายบวก (+) และตามด้วยค่าระดับความมั่นใจในหน่วยเปอร์เซ็นต์

- กรณีที่วิธีการอื่นที่เปรียบเทียบนั้น มีความถูกต้องกว่าตัดเล็มอย่างอ่อนโดยใช้ฟัซซิ-ฟิเคชัน อย่างมีนัยสำคัญทางสถิติ จะแสดงด้วยเครื่องหมายลบ (-) และตามด้วยระดับความมั่นใจในหน่วยเปอร์เซ็นต์

จากตารางที่ 4.2 จะเห็นว่าวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน จะให้ค่าความถูกต้องสูงสุดเทียบกับวิธีการอื่นๆ ในชุดข้อมูล 13 ชุดจากชุดข้อมูลทั้งหมด 17 ชุด โดยแบ่งได้เป็นชุดข้อมูลแบบต่อเนื่อง 10 ชุด และชุดข้อมูลแบบไม่ต่อเนื่อง 3 ชุด

และจากการเปรียบเทียบขั้นตอนวิธี ด้วยค่าระดับความมั่นใจตามตารางที่ 4.3 จะเห็นได้ว่าวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน ให้ความถูกต้องดีกว่าต้นไม้ตัดสินใจที่ไม่ทำการตัดเล็มอย่างมีนัยสำคัญในชุดข้อมูล 15 ชุด และเท่ากันในชุดข้อมูล 2 ชุดโดยไม่ด้อยกว่าเลย

และเมื่อเปรียบเทียบกับต้นไม้ตัดสินใจที่ตัดเล็มแล้ว วิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิ-ฟิเคชันจะให้ความถูกต้องมากกว่าอย่างมีนัยสำคัญในชุดข้อมูล 13 ชุด ให้ความถูกต้องน้อยกว่าอย่างมีนัยสำคัญในชุดข้อมูล 1 ชุด และไม่แตกต่างกันอย่างมีนัยสำคัญในชุดข้อมูล 3 ชุด

เมื่อเปรียบเทียบกับวิธีการตัดเล็มอย่างอ่อนเดิมนั้น จะพบว่าวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันจะให้ความถูกต้องมากกว่าอย่างมีนัยสำคัญในชุดข้อมูล 10 ชุด ให้ความถูกต้องน้อยกว่าอย่างมีนัยสำคัญในชุดข้อมูล 1 ชุด และไม่แตกต่างกันอย่างมีนัยสำคัญในชุดข้อมูล 6 ชุด

เมื่อพิจารณาแยกประเภทชุดข้อมูลแต่ละชนิด จะพบว่าชุดข้อมูลแบบข้อมูลต่อเนื่องให้ผลการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันดีที่สุดในชุดข้อมูลเกือบทุกชุด มีเพียงชุดข้อมูลสามชุดที่ให้ผลด้อยกว่าวิธีการอื่น ซึ่งชุดข้อมูลทั้งสามคือชุดข้อมูล Thyroid-disease ซึ่งมีลักษณะของคุณสมบัติเหมือนกันดังแสดงในตารางที่ 4.1 กล่าวคือชุดข้อมูล Thyroid-disease มีคุณสมบัติที่มีค่าที่เป็นไปได้แค่ 2 ค่าซึ่งจะไม่ถูกทำฟัซซิฟิเคชันถึง 22 ตัว และมีคุณสมบัติแบบข้อมูลต่อเนื่องที่จะทำฟัซซิฟิเคชันได้เพียง 7 ตัว จึงเป็นไปได้ว่าการทำฟัซซิฟิเคชันสำหรับคุณสมบัติเพียง 7 ตัวดังกล่าวอาจไม่เพียงพอในการเพิ่มประสิทธิภาพในการจำแนกข้อมูลกับชุดข้อมูลเหล่านี้

และเมื่อพิจารณาชุดข้อมูลแบบข้อมูลไม่ต่อเนื่อง จะเห็นว่าวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันให้ผลดีที่สุดในชุดข้อมูล Balance-scale ซึ่งเป็นชุดข้อมูลที่ทุกคุณสมบัติทุกตัวสามารถทำการฟัซซิฟิเคชันได้หมด ในขณะที่ชุดข้อมูลอื่นนั้นมีคุณสมบัติเพียงไม่กี่ตัวที่ทราบค่าความใกล้เคียงกันของค่าที่เป็นไปได้และทำการฟัซซิฟิเคชันได้ ดังนั้นการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันจึงอาจจะไม่ส่งผลให้ค่าความถูกต้องสูงขึ้นได้มากนักในชุดข้อมูลลักษณะนี้ เพราะคุณสมบัติที่มีความสำคัญต่อการจำแนกข้อมูลอาจจะไม่ได้ถูกทำการฟัซซิฟิเคชันด้วย

เมื่อพิจารณาผลการทดลองโดยแยกเป็นชุดข้อมูลที่มีข้อมูลไม่ทราบค่า กับชุดข้อมูลที่ไม่มีข้อมูลไม่ทราบค่า ในชุดข้อมูลที่มีข้อมูลไม่ทราบค่าทั้ง 6 ชุดนั้น เมื่อเปรียบเทียบกับวิธีการตัดเล็มอย่างอ่อน จะพบว่าการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันให้ค่าความถูกต้องมากกว่าอย่างมีนัยสำคัญในชุดข้อมูล 2 ชุดข้อมูล ให้ความถูกต้องน้อยกว่าอย่างมีนัยสำคัญในชุดข้อมูล 1 ชุด และไม่แตกต่างกันอย่างมีนัยสำคัญในชุดข้อมูล 3 ชุด

และเมื่อพิจารณาชุดข้อมูลที่ไม่มีข้อมูลไม่ทราบค่า 11 ชุดนั้น จะพบว่าการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันให้ค่าความถูกต้องมากกว่าวิธีการตัดเล็มอย่างอ่อนอย่างมีนัยสำคัญในชุดข้อมูล 8 ชุดข้อมูล ไม่แตกต่างกันอย่างมีนัยสำคัญในชุดข้อมูล 3 ชุด โดยไม่ด้อยกว่าเลย

นอกจากนี้ ในวิทยานิพนธ์ฉบับนี้ได้มีการทดลองเพิ่มเติม โดยการวัดผลความถูกต้องของการจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กเพียงอย่างเดียว ซึ่งดูผลการทดลองได้ในภาคผนวก ก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 5

# สรุปผลการวิจัย และข้อเสนอแนะ

## 5.1 สรุปผลการวิจัย

วิธีการตัดเล็มอย่างอ่อนได้อาศัยการให้น้ำหนักด้วยวิธีการแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์ก ในการกำหนดความสำคัญกับโนดแต่ละโนดในการจำแนกข้อมูล แทนการตัดโนดที่คาดว่าไม่มีความสำคัญออกไปโดยสิ้นเชิงตามวิธีการตัดเล็มต้นไม้ตัดสินใจโดยปกติ ด้วยแนวคิดที่ว่าโนดที่ตัดไปนั้นอาจจะมีความสำคัญในการจำแนกข้อมูลอยู่ก็ได้ อีกทั้งโนดที่ตัดไปอาจจะไม่มีความสำคัญในการจำแนกข้อมูลบางเส้นทางของต้นไม้ตัดสินใจ แต่อาจจะสำคัญกับอีกเส้นทางหนึ่งได้ ดังนั้นการสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กจากกฎที่ได้จากแต่ละเส้นทางของต้นไม้ตัดสินใจจึงปรับปรุงปัญหาดังกล่าวได้

แต่จะเห็นได้ว่าการตัดเล็มอย่างอ่อนยังคงรักษาวิธีการตัดสินใจแบบต้นไม้ตัดสินใจอยู่ สังเกตได้จากการแปลงข้อมูลเข้าโนดอินพุตที่อาศัยการตัดสินใจแบบต้นไม้ตัดสินใจว่าตรงตามเงื่อนไขของโนดอินพุตนั้นหรือไม่ ถ้าไม่ตรงค่าที่เข้าโนดอินพุตนี้จะเป็น 0 แต่ถ้าหากตรงตามเงื่อนไขค่าที่เข้าโนดอินพุตจะเป็น 1 แต่ในกรณีที่คุณสมบัติมีค่าใกล้เคียงกับค่าที่ใช้เป็นจุดแบ่งข้อมูลนั้น คุณสมบัติจะไม่เอนเอียงไปทางใดทางหนึ่งอย่างเต็มที่ แต่มีความก้ำกึ่งอยู่และสมควรจะพิจารณาคุณสมบัติอื่นประกอบในการจำแนกข้อมูล มากกว่าการจำแนกว่าคุณสมบัตินั้นเข้าสู่กรณีเงื่อนไขใดเงื่อนไขหนึ่งเพียงอย่างเดียว

วิทยานิพนธ์ฉบับนี้ได้นำเสนอวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน ซึ่งแก้ปัญหาดังกล่าวของวิธีการตัดเล็มอย่างอ่อนโดยการสร้างขอบเขตการตัดสินใจอย่างอ่อน (Soft decision boundary) ให้กับการตัดสินใจต่างๆ ของโนดอินพุตในโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก โดยอาศัยการสร้างฟังก์ชันความเป็นสมาชิกสำหรับโนดอินพุตแต่ละโนด ซึ่งทำให้คุณสมบัติที่มีความก้ำกึ่งกันระหว่างเงื่อนไขเปรียบเทียบของโนดอินพุตหลายโนด มีค่าความจริงต่างๆกันไปในโนดอินพุตแต่ละโนด ซึ่งจะต้องอาศัยคุณสมบัติอื่นประกอบในการจำแนกข้อมูล โดยในวิทยานิพนธ์ฉบับนี้ได้รวมไปถึงการสร้างฟังก์ชันความเป็นสมาชิกให้กับคุณสมบัติแบบข้อมูลไม่ต่อเนื่องด้วยเช่นกัน ซึ่งเป็นการให้ค่าความจริงกับเส้นทางการตัดสินใจที่ใช้คุณสมบัติที่มีค่าใกล้เคียงกับค่าคุณสมบัติจริงด้วย เพื่อเพิ่มความเป็นไปได้ในการจำแนกด้วยเส้นทางการตัดสินใจที่ใกล้เคียงกันให้มากขึ้น

โดยในวิทยานิพนธ์ฉบับนี้ได้ทำการทดลองตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันดังกล่าวกับชุดข้อมูล 17 ชุด โดยสร้างฟังก์ชันความเป็นสมาชิกในลักษณะรูปตัวเอสสำหรับคุณสมบัติแบบข้อมูลต่อเนื่อง และสร้างฟังก์ชันความเป็นสมาชิกลักษณะรูปสามเหลี่ยมสำหรับคุณสมบัติแบบข้อ

มูลไม่ต่อเนื่องเฉพาะคุณสมบัติที่ทราบการเรียงลำดับความใกล้เคียงกันของค่าที่เป็นไปได้ทุกค่า และต้องมีค่าที่เป็นไปได้มากกว่า 2 ค่า

ผลการทดลองกับชุดข้อมูล 17 ชุด และเปรียบเทียบขั้นตอนวิธีด้วยค่าระดับความมั่นใจระหว่างการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน กับการตัดเล็มอย่างอ่อนเดิมนั้นพบว่า การตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันให้ความถูกต้องดีกว่าอย่างมีนัยสำคัญในชุดข้อมูล 10 ชุด และด้อยกว่าในชุดข้อมูล 1 ชุด ซึ่งสามารถสรุปได้ว่าผลของการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันให้ประสิทธิภาพที่ดีกว่าการตัดเล็มอย่างอ่อนเดิม

นอกจากนั้นยังพบว่าโดยส่วนใหญ่การตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันจะให้ความถูกต้องที่ดีกว่าวิธีการอื่น สำหรับชุดข้อมูลที่ไม่มีข้อมูลไม่ทราบค่า

## 5.2 ข้อเสนอแนะ

1. การสร้างฟังก์ชันความเป็นสมาชิกมีได้หลากหลายวิธี เพื่อความถูกต้องสูงสุดจากการจำแนกข้อมูล ควรจะสร้างฟังก์ชันความเป็นสมาชิกในแต่ละคุณสมบัติตามความเหมาะสมของคุณสมบัติและลักษณะข้อมูล และปรับพารามิเตอร์ของฟังก์ชันจนได้ผลลัพธ์ที่พอใจ
2. การเรียงลำดับความใกล้เคียงของค่าที่เป็นไปได้ของคุณสมบัติแบบข้อมูลไม่ต่อเนื่อง อาจจะสามารถเรียงแบบวงกลมได้ กล่าวคือค่าสุดท้ายของการเรียงลำดับอาจจะกลับมามีความใกล้เคียงกับค่าแรกได้ เช่นการเรียงลำดับของสีเป็นต้น ซึ่งในการแปลงค่าที่เป็นไปได้ทั้งหมดเป็นจำนวนจริงเพื่อสร้างฟังก์ชันความเป็นสมาชิกสำหรับข้อมูลลักษณะนี้นั้น ควรจะประยุกต์วิธีที่สร้างความใกล้เคียงกลับมายังค่าแรกด้วย เช่นการนำค่าแรกบางส่วนไปต่อท้ายซ้ำในการเรียงลำดับเป็นต้น ทั้งนี้ในการทดลองในวิทยานิพนธ์ฉบับนี้ไม่พิจารณาความใกล้เคียงของข้อมูลลักษณะนี้
3. หากค่าความใกล้เคียงของค่าที่เป็นไปได้ของคุณสมบัติแบบข้อมูลไม่ต่อเนื่อง ไม่สม่ำเสมอเท่ากันตลอดทั้งลำดับ ควรจะเปลี่ยนการแปลงค่าที่เป็นไปได้เป็นจำนวนจริงให้สอดคล้องกับความใกล้เคียงนั้นด้วย นั่นคือควรเว้นระยะห่างให้มากขึ้นสำหรับค่าที่เป็นไปได้ที่มีความแตกต่างกันมากกว่าคู่อื่นๆ ในลำดับ หรือในทางกลับกันควรเว้นระยะห่างน้อยลงในค่าที่เป็นไปได้ที่มีความใกล้เคียงกันน้อยกว่าค่าอื่นๆ
4. จากผลการทดลองในวิทยานิพนธ์ฉบับนี้ จะพบว่าโดยส่วนใหญ่การตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันจะให้ความถูกต้องที่ดีกว่าวิธีอื่นในชุดข้อมูลที่ไม่มีข้อมูลไม่ทราบค่า ดังนั้นหากมีการศึกษาปรับปรุงวิธีการจัดการกับข้อมูลไม่ทราบค่า อาจจะช่วยปรับปรุงค่าความถูกต้องในการจำแนกชุดข้อมูลดังกล่าวให้ดีขึ้นได้

# รายการอ้างอิง


[1] ก้องศักดิ์ จงเกษมวงศ์. **การตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยการใช้แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2543.

[2] Benbrahim, H. and Bensaid, A. A Comparative Study of Pruned Decision Trees and Fuzzy Decision Trees. **19th International Conference of the North American on Fuzzy Information Processing Society** (NAFIPS 2000), pp. 227-231, 2000.

[3] Cox, E. **Fuzzy Systems Handbook**. AP Professional, 1994.

[4] Crockett, K. A. and Bandar, Z. Soft Decision Trees: A New Approach Using Non-Linear Fuzzification. **The Ninth IEEE International Conference on Fuzzy System** (FUZZ IEEE 2000), vol. 1, pp. 209-215, 2000.

[5] Dietterich, T. G. Proper Statistical Tests for Comparing Supervised Classification Learning Algorithms. **Technical Report**, Department of Computer Science, Oregon State University, Corvallis, Oregon, 1996.

[6] Esposito, F., Malerba, D., and Semeraro, G. A Comparative Analysis of Methods for Pruning Decision Trees. **IEEE Transactions** on Pattern Analysis and Machine Intelligence, vol. 19, no. 5, pp. 476-491, 1997.

[7] Klar, G. J. and Yuan, B. **Fuzzy Sets and Fuzzy Logic**. New Jersey: Prentice-Hall, 1995.

[8] Leighton, R. R. The Apirin/MIGRANES Neural Network Software User's Manual Release V6.0 [Machine readable data file]. Russell R. Leighton and the MITRE Coporation, 1992.

[9] Merz, C. J., Murphy, P. M., and Aha, D. W. UCI Repository of Machine Learning Database. Department of Information and Computer Science, University of California, Irvine, CA, 1997.
Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html

[10] Mitchell, T. **Machine Learning**. New York: McGraw Hill, 1997.

[11] Quinlan, J. R. Induction of Decision Trees. Machine Learning, 1986, pp. 81-106.

[12] Quinlan, J. R. **C4.5 Programs for Machine Learning**. California: Morgan Kaufmann, 1993.

[13] Rich, E. and Knight, K. **Artificial Intelligence**. Singapore: Prentice-Hall, 1991.

# ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ภาคผนวก ก
# ผลการทดลองการจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

งานวิจัยในวิทยานิพนธ์ฉบับนี้ได้มีการทดลองเพิ่มเติม โดยการวัดผลความถูกต้องจากการจำแนกข้อมูลโดยใช้แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กเพียงอย่างเดียว เนื่องจากเมื่อพิจารณาวิธีการตัดเล็มอย่างอ่อนจะพบว่าผลที่ได้จากการตัดเล็มก็คือโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กนั่นเอง โดยอาศัยกฎที่สร้างจากต้นไม้ตัดสินใจช่วยในการสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก จึงน่าจะได้มีการเปรียบเทียบความถูกต้องกับวิธีการจำแนกด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กเพียงอย่างเดียวด้วย

โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กที่ใช้สำหรับชุดข้อมูลแบบข้อมูลต่อเนื่องจะใช้จำนวนโนดอินพุตเท่ากับจำนวนคุณสมบัติ โดยข้อมูลเข้าสำหรับโนดอินพุตแต่ละโนดก็คือค่าจากคุณสมบัติแต่ละตัว ส่วนในชุดข้อมูลแบบข้อมูลไม่ต่อเนื่องจะใช้จำนวนโนดอินพุตเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของคุณสมบัติทุกตัว โดยข้อมูลเข้าสำหรับโนดอินพุตแต่ละโนดจะผ่านการทำฟัซซิฟิเคชันเช่นเดียวกับการทดลองในบทที่ 4 โดยคุณสมบัติที่ไม่ทำฟัซซิฟิเคชันก็จะมีข้อมูลเข้าโนดอินพุตเป็น 0 หรือ 1

และในชุดข้อมูลทั้ง 2 ชนิดนี้จะใช้จำนวนโนดเอาต์พุตเท่ากับจำนวนกลุ่มข้อมูล และทดลองใช้จำนวนโนดฮิดเดน 3 แบบคือ

- จำนวนโนดฮิดเดนเท่ากับจำนวนโนดอินพุต
- จำนวนโนดฮิดเดนเป็นครึ่งหนึ่งของจำนวนโนดอินพุต
- จำนวนโนดฮิดเดนเป็นสองเท่าของจำนวนโนดอินพุต

โดยระหว่างโนดอินพุตกับโนดฮิดเดนจะใช้การเชื่อมต่อแบบต่อกันหมด และในทำนองเดียวกันการเชื่อมต่อระหว่างโนดฮิดเดนกับโนดเอาต์พุตก็จะเชื่อมต่อกันหมด

ผลที่ได้จากการทดลองดังแสดงในตารางที่ ก1 โดยเปรียบเทียบกับค่าความถูกต้องจากวิธีการตัดเล็มอย่างอ่อน และวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน โดยค่าความถูกต้องที่สูงที่สุดในชุดข้อมูลแต่ละชุดจะถูกแสดงด้วยตัวอักษรตัวหนา

จากตารางดังกล่าวจะเห็นได้ว่าสำหรับชุดข้อมูลแบบข้อมูลต่อเนื่อง การจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก เพียงอย่างเดียวให้ผลความถูกต้องที่น้อยกว่าการตัดเล็มอย่างอ่อน ยกเว้นในชุดข้อมูล Waveform+noise ที่วิธีการจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กให้ความถูกต้องมากกว่าวิธีการตัดเล็มอย่างอ่อน แต่ยังน้อยกว่าวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน

ส่วนในชุดข้อมูลแบบข้อมูลไม่ต่อเนื่องนั้น การจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์กเพียงอย่างเดียวให้ค่าความถูกต้องที่ใกล้เคียง หรือดีกว่าการตัดเล็มอย่างอ่อน อันเนื่องมาจากโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กที่มีความใกล้เคียงกันมากเพราะมีลักษณะของโนดอินพุตและโนดเอาต์พุตเหมือนกันทุกประการ แตกต่างกันที่จำนวนโนดฮิดเดน และวิธีการเชื่อมโยงกันระหว่างโนดอินพุตและโนดฮิดเดน

จากผลการทดลองดังกล่าว จึงอาจกล่าวได้ว่าการจำแนกข้อมูลด้วยแบ็กพรอพาเกชัน-นิวรอลเน็ตเวิร์กนั้น มีความสำคัญที่การสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กอย่างมาก หากโครงสร้างที่ได้มีความเหมาะสมกับข้อมูลก็จะทำให้มีความสามารถในการจำแนกข้อมูลที่ดีได้

ตารางที่ ก1 ผลการทดลองการจำแนกข้อมูลด้วยแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก เปรียบเทียบกับวิธีการตัดเล็มอย่างอ่อน และวิธีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน

| **ชื่อชุดข้อมูล** | **ค่าความถูกต้อง(%) จากแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กที่มีจำนวนฮิดเดินโนดเป็น ... เท่าของโนดอินพุต** | | | **ค่าความถูกต้อง(%) จากวิธี** | |
|---|---|---|---|---|---|
| | **0.5 เท่า** | **1 เท่า** | **2 เท่า** | ตัดเล็มอย่างอ่อน | ตัดเล็มอย่างอ่อนใช้ฟัซซิฟิเคชัน |
| Soybean | 93.35 | **93.62** | 93.09 | 89.89 | 91.49 |
| Balance-scale | **99.04** | 98.72 | 98.88 | 90.24 | 91.04 |
| Restricted (lymphography) | **81.78** | 80.44 | 80.44 | 80.42 | 79.72 |
| Restricted (primary-tumor) | 44.15 | 46.20 | **46.49** | 43.37 | 45.72 |
| Thyroid-disease (allbp) | 96.91 | 96.91 | 97.22 | 97.53 | **97.63** |
| Thyroid-disease (allhyper) | 97.43 | 97.43 | 97.43 | **98.97** | 98.46 |
| Thyroid-disease (allhypo) | 94.65 | 96.81 | 96.30 | 98.87 | **99.79** |
| Thyroid-disease (allrep) | 96.19 | 96.19 | 96.19 | **98.97** | 98.25 |
| Image | 31.95 | 37.81 | 39.76 | 90.38 | **93.76** |
| Statlog (satimage) | 31.25 | 30.65 | 31.85 | 87.40 | **88.20** |
| Statlog (shuttle) | 79.16 | 79.18 | 84.52 | **99.99** | **99.99** |
| Statlog (segment) | 30.00 | 31.34 | 30.91 | 96.75 | **98.05** |
| Iris | 68.00 | 93.33 | 68.00 | 95.33 | **98.00** |
| Wine | 39.87 | 39.87 | 39.87 | 93.30 | **96.63** |
| Waveform | 79.68 | 79.46 | 81.00 | 81.08 | **82.68** |
| Waveform + noise | 81.96 | 82.28 | 80.70 | 80.74 | **82.56** |
| Glass | 38.84 | 42.94 | 44.88 | 68.23 | **74.30** |

# ภาคผนวก ข
# การใช้งานโปรแกรม

งานวิทยานิพนธ์ฉบับนี้ได้พัฒนาโปรแกรมต่อจากโปรแกรมในวิทยานิพนธ์ของ ก้องศักดิ์ จงเกษมวงศ์ [1] โดยโปรแกรมที่ได้พัฒนาขึ้นใหม่นี้จะครอบคลุมตั้งแต่การสร้างและใช้งานต้นไม้ตัดสินใจ การสร้างและใช้งานแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กสำหรับการตัดเล็มอย่างอ่อน และการสร้างและใช้งานฟังก์ชันความเป็นสมาชิกซึ่งใช้ในการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันด้วย

โดยโปรแกรมนี้ได้นำฟังก์ชันต่าง ๆ จากโปรแกรมที่เกี่ยวข้องมาใช้ด้วย โดยได้ใช้ส่วนของโปรแกรม C4.5 ในการทำงานกับต้นไม้ตัดสินใจ และส่วนของโปรแกรม Aspirin/MIGRAINES Neural Network Software [8] ในการทำงานกับแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก โปรแกรมนี้ถูกพัฒนาบนระบบปฏิบัติการ Microsoft Windows XP โดยเขียนด้วยภาษา Visual C++ รุ่น 6.0

โมดูล (module) หลักที่ใช้ในการพัฒนาโปรแกรมแบ่งออกเป็น 3 ส่วน คือ

**1. ต้นไม้ตัดสินใจ**

ในส่วนของฟังก์ชันที่เกี่ยวข้องกับต้นไม้ตัดสินใจนี้ จะประกอบไปด้วย การสร้างต้นไม้ตัดสินใจ พร้อมทั้งการตัดเล็มต้นไม้ ตามขั้นตอนในบทที่ 2 ข้อมูลเข้าที่ต้องการสำหรับฟังก์ชันนี้มี 3 ส่วนด้วยกัน คือ ข้อมูลสอน ข้อมูลทดสอบและรายละเอียดของคุณสมบัติและกลุ่ม โดยผลการสร้างต้นไม้ตัดสินใจที่ไม่ได้ตัดเล็ม และต้นไม้ตัดสินใจที่ทำการตัดเล็มแล้ว รวมไปถึงค่าความถูกต้องของต้นไม้ทั้งสองต้น จะถูกเก็บข้อมูลเป็นแบบข้อความ (text file)

**2. การแปลงข้อมูลต้นไม้ตัดสินใจเป็นข้อมูลสำหรับแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก**

การทำงานของโมดูลนี้ จะเป็นไปตามขั้นตอนการแปลงข้อมูลสำหรับการตัดเล็มอย่างอ่อนตามที่ได้อธิบายไว้ในบทที่ 2 คือประกอบด้วย การสร้างกฎจากต้นไม้ตัดสินใจ การสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก และการสร้างข้อมูลสำหรับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก โดยรวมไปถึงการสร้างฟังก์ชันความเป็นสมาชิกสำหรับแปลงข้อมูลให้โครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กด้วยตามในบทที่ 3 ซึ่งจะใช้สำหรับการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชัน

โดยข้อมูลเข้าที่ต้องการสำหรับโมดูลนี้ได้แก่ ข้อมูลต้นไม้ตัดสินใจที่สร้างจากโมดูลที่หนึ่ง และในกรณีการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิฟิเคชันสำหรับชุดข้อมูลแบบไม่ต่อเนื่อง จะต้องมีข้อมูลคุณสมบัติที่สามารถเรียงลำดับความใกล้เคียงของค่าที่เป็นไปได้ด้วย

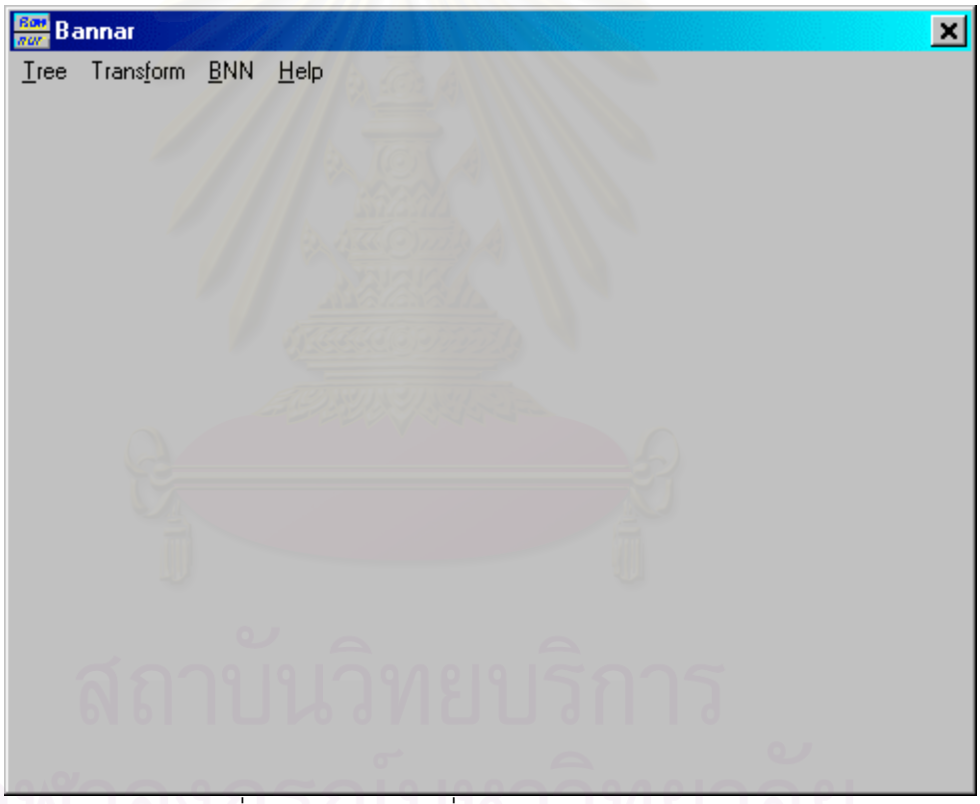
หลังจากเสร็จสิ้นการทำงานของโมดูลนี้จะได้ข้อมูลสอนและข้อมูลทดสอบ ที่นำไปใช้กับโมดูลแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กต่อไป

3. **แบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก**

โมดูลนี้จะนำข้อมูลสอนมาเข้าสู่กระบวนการสอนโครงสร้างแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์ก ตามขั้นตอนในบทที่ 2 และได้ผลเป็นค่าน้ำหนักต่าง ๆ ของโครงสร้างนิวรอลเน็ตเวิร์ก หลังจากนั้นจะทดสอบค่าความถูกต้องด้วยข้อมูลทดสอบ และนำค่าความถูกต้องที่ได้ ไปวิเคราะห์ข้อมูลต่อไป

**<u>การใช้งานโปรแกรม</u>**

เมื่อผู้ใช้งานเรียกโปรแกรมขึ้นมาจะปรากฏหน้าจอหลักในการทำงานของโปรแกรมดังรูปที่ ข1 โดยมีเมนูหลักสำหรับจัดการการทำงานต่างๆปรากฏอยู่ด้านบนของหน้าจอ ซึ่งประกอบไปด้วย 3 เมนูหลักคือ Tree Transform และ BNN โดยวิธีการใช้งานเมนูหลักต่างๆ จะได้อธิบายต่อไป



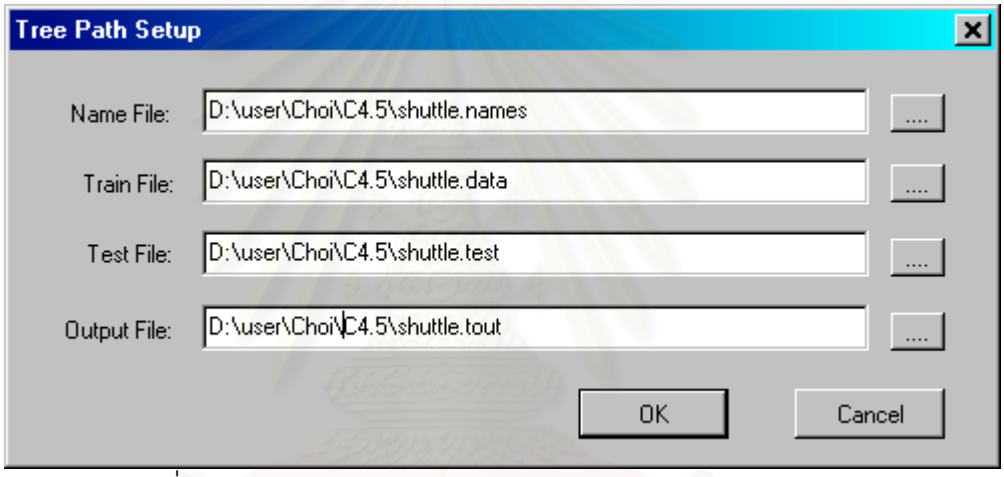

รูปที่ ข1 หน้าจอหลักเมื่อผู้ใช้งานเรียกโปรแกรม

**<u>เมนู Tree</u>**

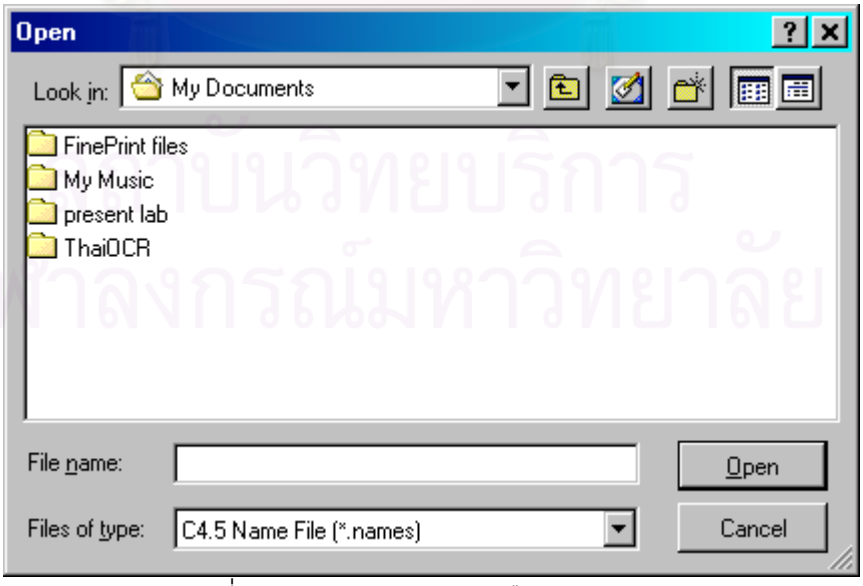
เมนู Tree เป็นเมนูที่ใช้ในจัดการกับต้นไม้ตัดสินใจทั้งที่ไม่ได้ตัดเล็ม และที่ตัดเล็มด้วยค่าความผิดพลาด โดยเมื่อเลือกที่เมนู Tree จะปรากฏเมนูย่อยดังในรูปที่ ข2 ซึ่งมีการทำงานต่างๆ ดังนี้

รูปที่ ข2 หน้าจอเมนูย่อยภายในเมนู Tree

1. **เมนู** File and Path สำหรับเลือกแฟ้มข้อมูลที่ต้องการนำมาสร้างต้นไม้ตัดสินใจ ประกอบด้วยแฟ้มข้อมูลที่เป็นรายละเอียดของชุดข้อมูล (Name File) แฟ้มข้อมูลสอน (Train File) แฟ้มข้อมูลทดสอบ (Test File) และแฟ้มข้อมูลที่เก็บต้นไม้ตัดสินใจ (Output File) ดังแสดงในรูปที่ ข3 และ ข4

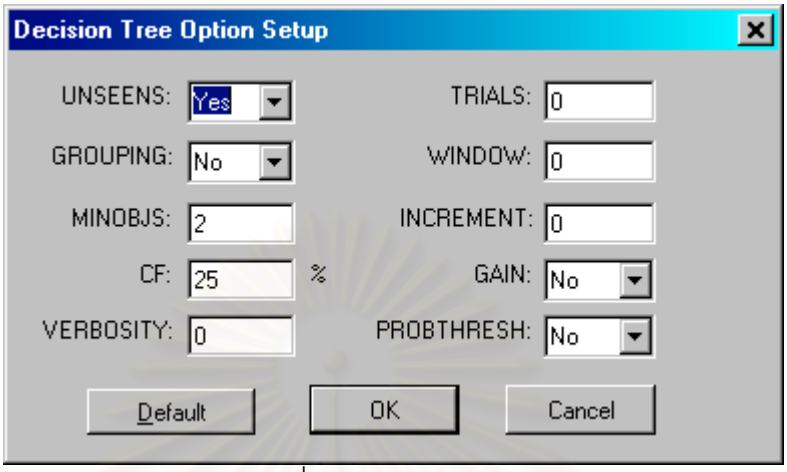



รูปที่ ข3 หน้าจอสำหรับเลือกแฟ้มข้อมูลในการสร้างต้นไม้ตัดสินใจ



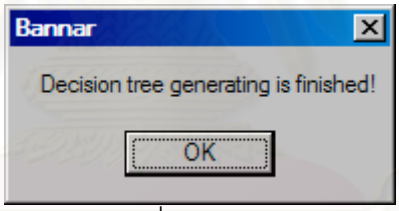

รูปที่ ข4 หน้าจอสำหรับการเลือกแฟ้มข้อมูล

2. **เมนู** Option เป็นทางเลือกต่าง ๆ สำหรับการสร้างต้นไม้ตัดสินใจ และต้นไม้ตัดสินใจที่ตัดเล็มแล้วดังแสดงในรูปที่ ข5 ทั้งนี้ค่าตั้งต้นที่ให้มาจะเป็นค่าที่ใช้สำหรับงานวิจัยนี้




รูปที่ ข5 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของการสร้างต้นไม้ตัดสินใจ

3. **เมนู** Create Tree เป็นเมนูที่ใช้สำหรับสร้างต้นไม้ตัดสินใจที่ได้จากแฟ้มข้อมูลในเมนู File and Path เมื่อสร้างเสร็จจะมีข้อความบอกดังรูปที่ ข5

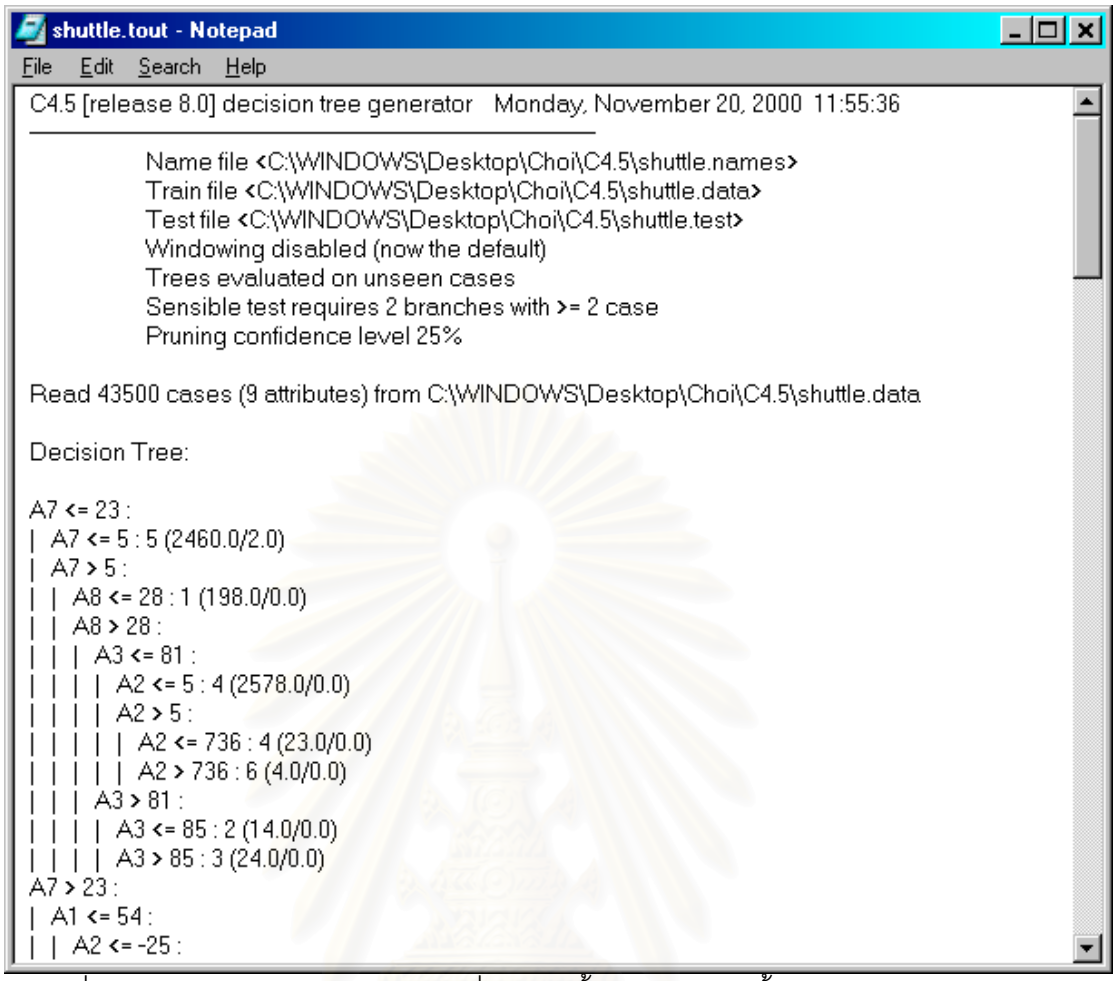



รูปที่ ข6 ข้อความเมื่อสร้างต้นไม้ตัดสินใจเสร็จ

4. **เมนู** View Output เมื่อสร้างต้นไม้ตัดสินใจเสร็จแล้ว ข้อมูลที่ได้จะเก็บลงแฟ้มข้อมูลที่ระบุไว้ใน เมนู File and Path เราสามารถดูรายละเอียดของต้นไม้ตัดสินใจดังกล่าวได้ โดยเมื่อเลือกเมนู View Output แล้ว จะเรียกใช้งานโปรแกรม notepad สำหรับเปิดแฟ้มข้อมูลที่เก็บต้นไม้ตัดสินใจ ตามรูปที่ ข7

5. **เมนู** Exit ออกจากโปรแกรม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย




รูปที่ ข7 ตัวอย่างข้อมูลต้นไม้ตัดสินใจที่ถูกสร้างขึ้น และถูกเรียกขึ้นมาจากเมนู View Output

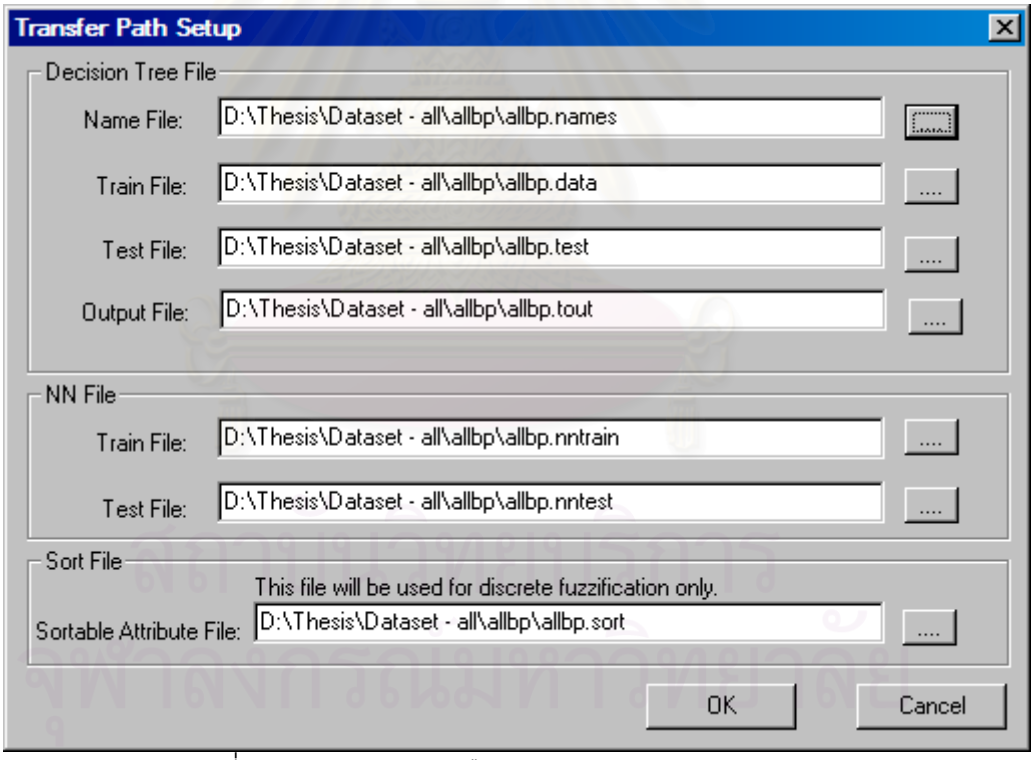
**เมนู Transform**

เมนู Transform นี้เป็นเมนูที่ใช้สำหรับการสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์ก รวมไปถึงการสร้างข้อมูลสำหรับนิวรอลเน็ตเวิร์ก โดยเมื่อเลือกเมนู Transform จะปรากฏเมนูย่อยดังรูปที่ ข8 ซึ่งมีการทำงานของเมนูย่อยดังต่อไปนี้

รูปที่ ข8 หน้าจอเมนูย่อยภายในเมนู Transform

1. **เมนู** File and Path ในเมนูย่อยนี้ จะเลือกแฟ้มข้อมูลที่ใช้สำหรับการสร้างโครงสร้างนิวรอลเน็ตเวิร์ก รวมทั้งข้อมูลสอน และข้อมูลทดสอบของนิวรอลเน็ตเวิร์ก โดยแฟ้มข้อมูลที่จะต้องใช้เป็นข้อมูลเข้าในการทำงานคือ แฟ้มข้อมูลที่เป็นรายละเอียดของชุดข้อมูล (Name File) แฟ้มข้อมูลสอน (Train File) แฟ้มข้อมูลทดสอบ (Test File) และแฟ้มข้อมูลที่เก็บต้นไม้ตัดสินใจ (Output File) และในกรณีของชุดข้อมูลแบบไม่ต่อเนื่องที่ต้องการทำการตัดเล็มอย่างอ่อนโดยฟัซซิฟิเคชันนั้น จะต้องมีแฟ้มข้อมูลที่เก็บค่าคุณสมบัติที่สามารถเรียงลำดับความใกล้เคียงกันของค่าที่เป็นไปได้ทั้งหมดด้วย (Sortable Attribute File) โดยการเรียงลำดับค่าที่เป็นไปได้นั้นจะต้องเรียงลำดับไว้แล้วในแฟ้มข้อมูล Name File

   ผลที่ได้จากการแปลงข้อมูลก็จะเก็บเป็นแฟ้มข้อมูล ซึ่งจะเป็นข้อมูลสอน และข้อมูลทดสอบของโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์กต่อไป โดยจะเก็บแฟ้มข้อมูลตามที่กำหนดในส่วน NN File ของหน้าจอนี้ตามรูปที่ ข9



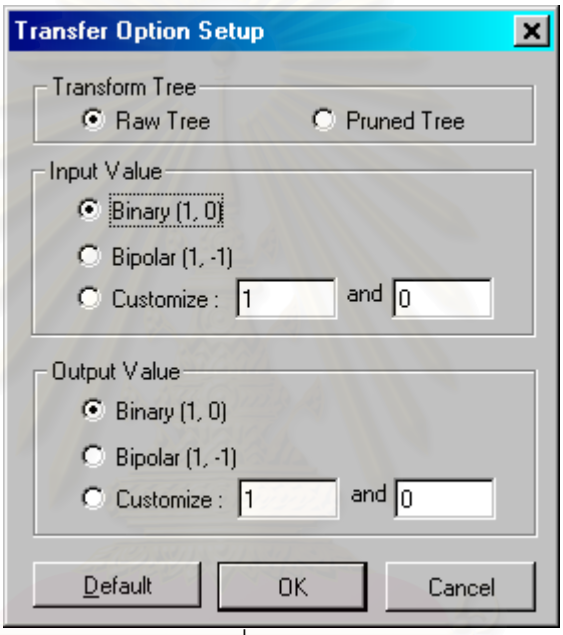

รูปที่ ข9 หน้าจอสำหรับเลือกแฟ้มข้อมูลในการแปลงข้อมูล

2. **เมนู** Option เป็นทางเลือกต่าง ๆ สำหรับการแปลงข้อมูลจากต้นไม้ตัดสินใจ เป็นข้อมูลที่ใช้กับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ทั้งนี้แบ่งออกเป็น 3 ส่วนคือ

   - Transform Tree เป็นส่วนที่ใช้สำหรับการเลือกว่าจะใช้ต้นไม้ตัดสินใจที่ยังไม่ตัดเล็ม (Raw Tree) หรือต้นไม้ตัดสินใจที่ทำการตัดเล็มแล้ว (Pruned Tree)

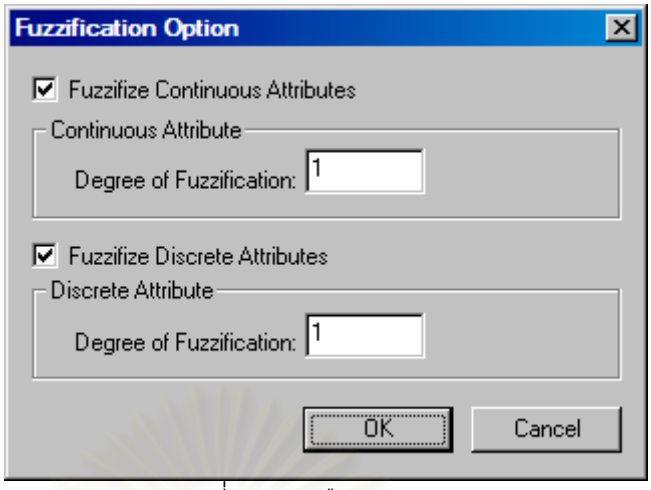
ในการสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก และข้อมูลสอนและข้อมูลทดสอบ

- Input Value เป็นทางเลือกที่ใช้สำหรับการสร้างข้อมูลในส่วนโนดอินพุตของแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก สามารถเลือกเป็นไบนารี ไบโพลาร์ หรือทำการกำหนดเองก็ได้ ทั้งนี้ค่าตั้งต้นจะเป็นไบนารี
- Output Value เป็นทางเลือกที่ใช้สำหรับการสร้างข้อมูลในส่วนโนดเอาต์พุต โดยตัวเลือกจะเป็นเช่นเดียวกันกับ Input Value



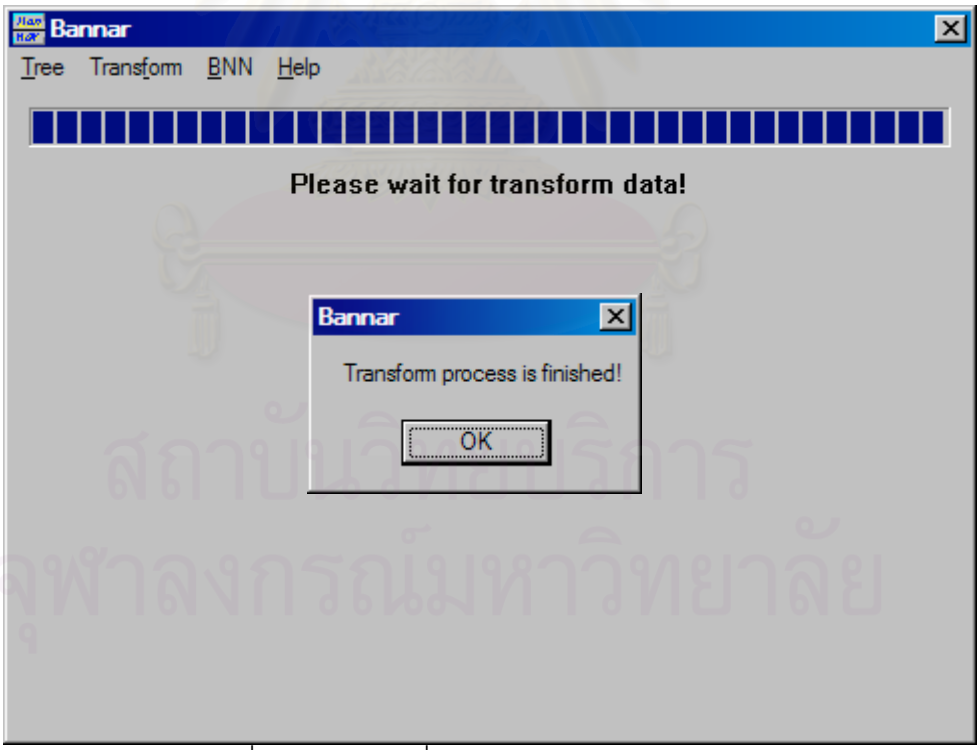

รูปที่ ข10 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของการแปลงข้อมูล

3. **เมนู** Fuzzification Option เป็นทางเลือกสำหรับการตัดเล็มอย่างอ่อนโดยใช้ฟัซซิ-ฟิเคชันดังแสดงในรูปที่ ข11 ซึ่งจะสามารถเลือกได้ว่าจะทำการฟัซซิฟิเคชันกับคุณสมบัติแบบข้อมูลต่อเนื่อง และข้อมูลไม่ต่อเนื่องหรือไม่ และสามารถกำหนดค่าตัวเลือก n ดังที่ได้อธิบายไว้ในบทที่ 3 ได้ โดยในโปรแกรมนี้จะเรียกค่าดังกล่าวว่าค่า Degree of Fuzzification




รูปที่ ข11 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของการทำฟัซซิฟิเคชัน

4. **เมนู** Transform Data เป็นเมนูที่ใช้ในการสั่งให้โปรแกรมแปลงข้อมูล ซึ่งหากมีการเลือกให้โปรแกรมทำฟัซซิฟิเคชันไว้ดังที่ได้กล่าวมาแล้วนั้น โปรแกรมก็จะสร้างฟังก์ชันความเป็นสมาชิกสำหรับแต่ละโนดอินพุต และแปลงข้อมูลโดยใช้ฟังก์ชันความเป็นสมาชิกดังกล่าว และเมื่อแปลงเสร็จสิ้นจะปรากฏหน้าจอดังรูปที่ ข12

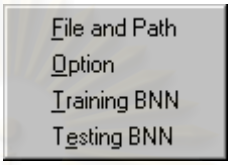



รูปที่ ข12 หน้าจอเมื่อการแปลงข้อมูลเสร็จสมบูรณ์

5. **เมนู** Transform for BNN เป็นเมนูที่ใช้ในการสั่งให้โปรแกรมแปลงข้อมูลและสร้างโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก สำหรับการจำแนกข้อมูลด้วยแบ็ก-พรอพาเกชันนิวรอลเน็ตเวิร์กเพียงอย่างเดียวดังในภาคผนวก ก.

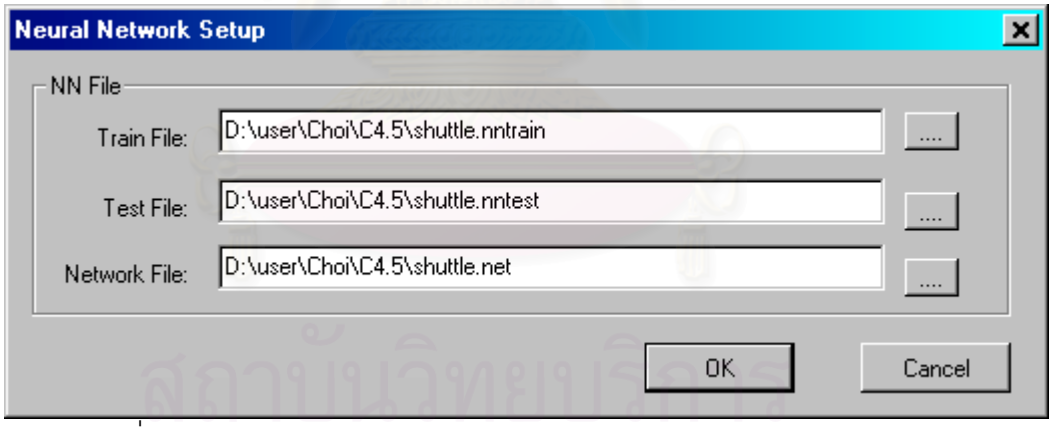
**เมนู** BNN

เมนู BNN ทำหน้าที่ในการสอน และทดสอบโครงสร้างกับแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์ก โดยเมื่อเลือกเมนูนี้ จะปรากฏเมนูย่อยดังรูปที่ ข13 ซึ่งมีการทำงานต่างๆดังนี้



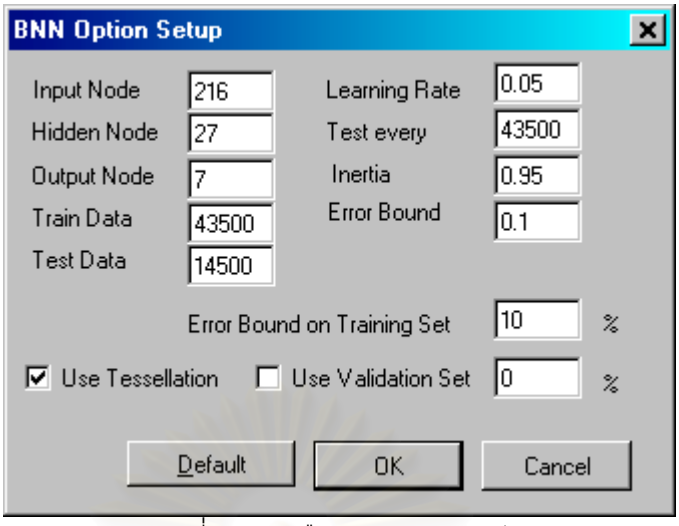

รูปที่ ข13 หน้าจอเมนูย่อยภายในเมนู BNN

1. **เมนู** File and Path สำหรับเมนูย่อยนี้ เป็นการเลือกแฟ้มข้อมูลที่ใช้กับแบ็กพรอพา-เกชันนิวรอลเน็ตเวิร์ก ประกอบไปด้วย แฟ้มข้อมูลสำหรับข้อมูลสอน แฟ้มข้อมูลสำหรับข้อมูลทดสอบ และแฟ้มข้อมูลสำหรับเก็บน้ำหนักต่าง ๆ ของโครงสร้างแบ็ก-พรอพาเกชันนิวรอลเน็ตเวิร์ก (Network File)



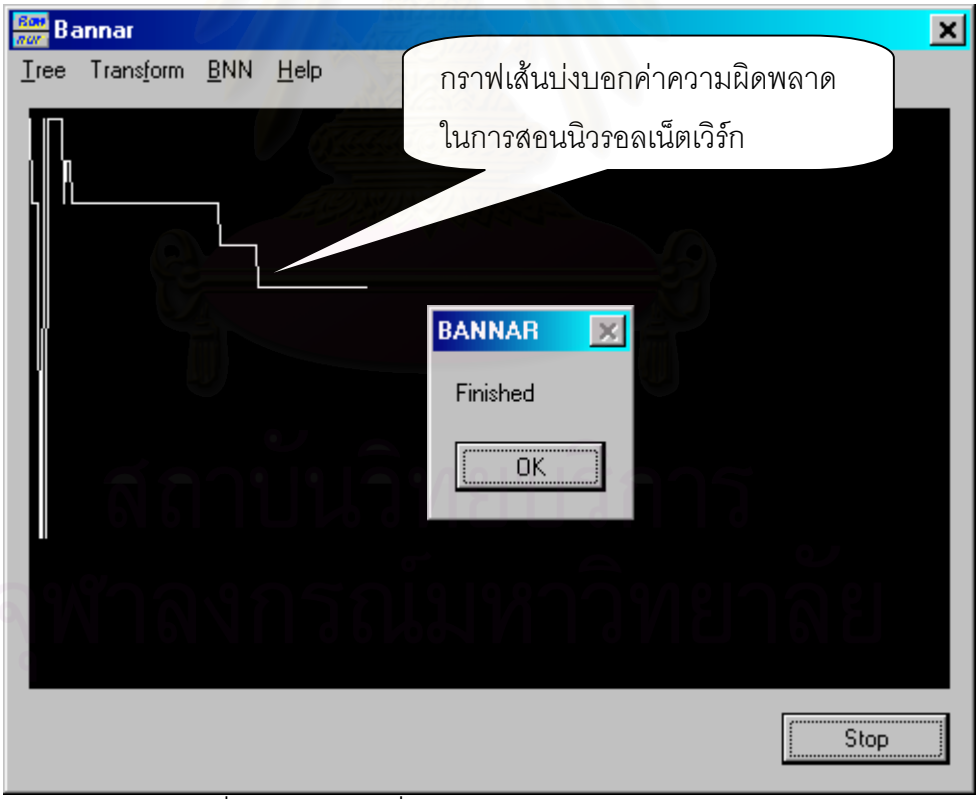

รูปที่ ข14 หน้าจอสำหรับเลือกแฟ้มข้อมูลในแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

2. **เมนู** Option เป็นทางเลือกต่าง ๆ สำหรับโครงสร้างแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ทั้งนี้หากผู้ใช้งานแปลงข้อมูลจากเมนู Transform มาก่อน ข้อมูลต่าง ๆ คือจำนวนโนดอินพุต (Input Node) จำนวนโนดฮิดเดน (Hidden Node) จำนวนโนดเอาต์พุต (Output Node) จำนวนข้อมูลสอน (Training Data) และจำนวนข้อมูลทดสอบ (Test Data) จะใส่ค่าให้อัตโนมัติ หากผู้ใช้งานไม่ได้ผ่านโมดูลแปลงข้อมูลมาก่อน จะต้องใส่ข้อมูลข้างต้นเอง สำหรับค่าต่าง ๆ ที่เหลือเป็นค่าที่เกี่ยวข้องกับการสอนนิวรอล-เน็ตเวิร์ก ซึ่งค่าตั้งต้นที่ให้ไว้เป็นค่าที่ใช้สำหรับงานวิจัยนี้



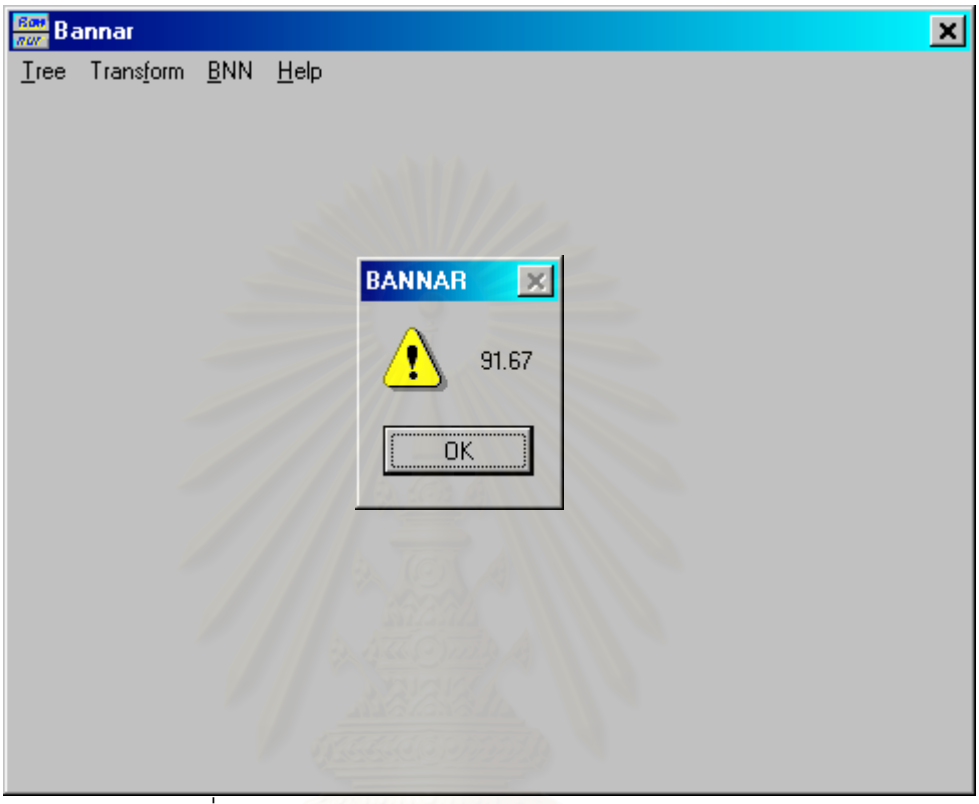

รูปที่ ข15 หน้าจอสำหรับปรับเปลี่ยนทางเลือกต่าง ๆ ของแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก

3. **เมนู** Training BNN เป็นเมนูที่ใช้สำหรับสอนแบ็กพรอพาเกชันนิวรอลเน็ตเวิร์ก ตามเงื่อนไขต่าง ๆ จากเมนู Option โดยค่าน้ำหนักต่าง ๆ ของโครงสร้างนิวรอลเน็ตเวิร์กจะถูกเก็บลงแฟ้มข้อมูลเมื่อการสอนเสร็จสิ้น และจะขึ้นข้อความบอกดังรูปที่ ข15




รูปที่ ข16 หน้าจอเมื่อสอนนิวรอลเน็ตเวิร์กเสร็จสมบูรณ์

4. **เมนู** Testing BNN สำหรับเมนูนี้จะทดสอบโครงสร้างแบ็กพรอพาเกชันนิวรอล-เน็ตเวิร์ก โดยใช้ข้อมูลทดสอบ และเมื่อทดสอบเสร็จแล้วจะขึ้นข้อความที่บอกถึงเปอร์เซ็นต์ค่าความถูกต้องที่ได้กับข้อมูลทดสอบดังรูปที่ ข16




รูปที่ ข17 หน้าจอแสดงค่าความถูกต้องของข้อมูลทดสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ภาคผนวก ค
# รายละเอียดข้อมูลที่ใช้ในการทดลอง

ชุดข้อมูลต่าง ๆ ที่ใช้ในการทดลองของวิทยานิพนธ์ฉบับนี้ได้เลือกมาจาก Department of Information and Computer Science, University of California, Irvine [9] โดยมีทั้งหมด 17 ชุด โดยมีรายละเอียดของชุดข้อมูลแต่ละชุดดังต่อไปนี้ โดยในคุณสมบัติแบบข้อมูลต่อเนื่องจะแสดงค่าที่เป็นไปได้ของคุณสมบัติ ด้วยคำว่า “continuous”

**ชื่อชุดข้อมูล** Soybean

**จำนวนคุณสมบัติ** 35

**จำนวนกลุ่ม** 19 กลุ่มคือ diaporthe-stem-canker, charcoal-rot, rhizoctonia-root-rot, phytophthora-rot, brown-stem-rot, powdery-mildew, downy-mildew, brown-spot, bacterial-blight, bacterial-pustule, purple-seed-stain, anthracnose, phyllosticta-leaf-spot, alternarialeaf-spot, frog-eye-leaf-spot, diaporthe-pod-&-stem-blight, cyst-nematode, 2-4-d-injury และ herbicide-injury

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| date | 0,1,2,3,4,5,6 | stem | 0,1 |
| plant-stand | 0,1 | lodging | 0,1 |
| precip | 0,1,2 | stem-cankers | 0,1,2,3 |
| temp | 0,1,2 | canker-lesion | 0,1,2,3 |
| hail | 0,1 | fruiting-bodies | 0,1 |
| crop-hist | 0,1,2,3 | external-decay | 0,1,2 |
| area-damaged | 0,1,2,3 | mycelium | 0,1 |
| severity | 0,1,2 | int-discolor | 0,1,2 |
| seed-tmt | 0,1,2 | sclerotia | 0,1 |
| germination | 0,1,2 | fruit-pods | 0,1,2,3 |
| plant-growth | 0,1 | fruit-spots | 0,1,2,3,4 |
| leaves | 0,1 | seed | 0,1 |
| leafspots-halo | 0,1,2 | mold-growth | 0,1 |
| leafspots-marg | 0,1,2 | seed-discolor | 0,1 |
| leafspot-size | 0,1,2 | seed-size | 0,1 |
| leaf-shread | 0,1 | shriveling | 0,1 |
| leaf-malf | 0,1 | roots | 0,1,2 |
| leaf-mild | 0,1,2 | | |

**ชื่อชุดข้อมูล** Balance-scale

**จำนวนคุณสมบัติ** 4

**จำนวนกลุ่ม** 3 กลุ่มคือ L,B และ R

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| Left-Weight | 1,2,3,4,5 | Right-Weight | 1,2,3,4,5 |
| Left-Distance | 1,2,3,4,5 | Right-Distance | 1,2,3,4,5 |

**ชื่อชุดข้อมูล** Restricted (lymphography)

**จำนวนคุณสมบัติ** 18

**จำนวนกลุ่ม** 4 กลุ่มคือ 1, 2, 3 และ 4

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| lymphatics | 1,2,3,4 | lym_nodes_enlar | 1,2,3,4 |
| block_of_affere | 1,2 | changes_in_lym | 1,2,3 |
| bl_of_lymph_c | 1,2 | defect_in_node | 1,2,3,4 |
| bl_of_lymph_s | 1,2 | changes_in_node | 1,2,3,4 |
| by_pass | 1,2 | changes_in_stru | 1,2,3,4,5,6,7,8 |
| extravasates | 1,2 | special_forms | 1,2,3 |
| regenerationof | 1,2 | dislocation_of | 1,2 |
| early_uptake_in | 1,2 | exclusion_of_no | 1,2 |
| lym_nodes_dimin | 1,2,3 | no_of_nodes_in | 1,2,3,4,5,6,7,8 |

**ชื่อชุดข้อมูล** Restricted (primary-tumor)

**จำนวนคุณสมบัติ** 17

**จำนวนกลุ่ม** 22 กลุ่มคือ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 และ 22

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| age | 1,2,3 | liver | 1,2 |
| sex | 1,2 | brain | 1,2 |
| histologic-type | 1,2,3 | skin | 1,2 |
| degree-of-diffe | 1,2,3 | neck | 1,2 |
| bone | 1,2 | supraclavicular | 1,2 |
| bone-marrow | 1,2 | axillar | 1,2 |
| lung | 1,2 | mediastinum | 1,2 |
| pleura | 1,2 | abdominal | 1,2 |
| peritoneum | 1,2 | | |

**ชื่อชุดข้อมูล** Iris

**จำนวนคุณสมบัติ** 4

**จำนวนกลุ่ม** 3 กลุ่มคือ Iris-setosa, Iris-versicolor และ Iris-virginica

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| sepal-length | continuous | petal-length | continuous |
| sepal-width | continuous | petal-width | continuous |

**ชื่อชุดข้อมูล** Thyroid-disease (allbp)

**จำนวนคุณสมบัติ** 29

**จำนวนกลุ่ม** 3 กลุ่มคือ increased binding protein, decreased binding protein และ negative

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| age | continuous | psych | f, t |
| sex | M, F | TSH measured | f, t |
| on thyroxine | f, t | TSH | continuous |
| query on thyroxine | f, t | T3 measured | f, t |
| on antithyroid medication | f, t | T3 | continuous |
| sick | f, t | TT4 measured | f, t |
| pregnant | f, t | TT4 | continuous |
| thyroid surgery | f, t | T4U measured | f, t |
| I131 treatment | f, t | T4U | continuous |
| query hypothyroid | f, t | FTI measured | f, t |
| query hyperthyroid | f, t | FTI | continuous |
| lithium | f, t | TBG measured | f, t |
| goitre | f, t | TBG | continuous |
| tumor | f, t | referral source | WEST, STMW, SVHC, SVI, SVHD, other |
| hypopituitary | f, t | | |

**ชื่อชุดข้อมูล** Thyroid-disease (allhyper)

**จำนวนคุณสมบัติ** 29

**จำนวนกลุ่ม** 5 กลุ่มคือ hyperthyroid, T3 toxic, goitre, secondary toxic และ negative

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| age | continuous | psych | f, t |
| sex | M, F | TSH measured | f, t |
| on thyroxine | f, t | TSH | continuous |
| query on thyroxine | f, t | T3 measured | f, t |
| on antithyroid medication | f, t | T3 | continuous |
| sick | f, t | TT4 measured | f, t |

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| pregnant | f, t | TT4 | continuous |
| thyroid surgery | f, t | T4U measured | f, t |
| I131 treatment | f, t | T4U | continuous |
| query hypothyroid | f, t | FTI measured | f, t |
| query hyperthyroid | f, t | FTI | continuous |
| lithium | f, t | TBG measured | f, t |
| goitre | f, t | TBG | continuous |
| tumor | f, t | referral source | WEST, STMW, SVHC, SVI, SVHD, other |
| hypopituitary | f, t | | |

**ชื่อชุดข้อมูล** Thyroid-disease (allhypo)

**จำนวนคุณสมบัติ** 29

**จำนวนกลุ่ม** 5 กลุ่มคือ hypothyroid, primary hypothyroid, compensated hypothyroid, secondary hypothyroid และ negative

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| age | continuous | psych | f, t |
| sex | M, F | TSH measured | f, t |
| on thyroxine | f, t | TSH | continuous |
| query on thyroxine | f, t | T3 measured | f, t |
| on antithyroid medication | f, t | T3 | continuous |
| sick | f, t | TT4 measured | f, t |
| pregnant | f, t | TT4 | continuous |
| thyroid surgery | f, t | T4U measured | f, t |
| I131 treatment | f, t | T4U | continuous |
| query hypothyroid | f, t | FTI measured | f, t |
| query hyperthyroid | f, t | FTI | continuous |
| lithium | f, t | TBG measured | f, t |
| goitre | f, t | TBG | continuous |
| tumor | f, t | referral source | WEST, STMW, SVHC, SVI, SVHD, other |
| hypopituitary | f, t | | |

**ชื่อชุดข้อมูล** Thyroid-disease (allrep)

**จำนวนคุณสมบัติ** 29

**จำนวนกลุ่ม** 4 กลุ่มคือ replacement therapy, underreplacement, overreplacement และ negative

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| age | continuous | psych | f, t |
| sex | M, F | TSH measured | f, t |
| on thyroxine | f, t | TSH | continuous |
| query on thyroxine | f, t | T3 measured | f, t |

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| on antithyroid medication | f, t | T3 | continuous |
| sick | f, t | TT4 measured | f, t |
| pregnant | f, t | TT4 | continuous |
| thyroid surgery | f, t | T4U measured | f, t |
| I131 treatment | f, t | T4U | continuous |
| query hypothyroid | f, t | FTI measured | f, t |
| query hyperthyroid | f, t | FTI | continuous |
| lithium | f, t | TBG measured | f, t |
| goitre | f, t | TBG | continuous |
| tumor | f, t | referral source | WEST, STMW, SVHC, SVI, SVHD, other |
| hypopituitary | f, t | | |

**ชื่อชุดข้อมูล** Image

**จำนวนคุณสมบัติ** 19

**จำนวนกลุ่ม** 7 กลุ่มคือ BRICKFACE, SKY, FOLIAGE, CEMENT, WINDOW, PATH และ GRASS

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| region-centroid-col | continuous | rawred-mean | continuous |
| region-centroid-row | continuous | rawblue-mean | continuous |
| region-pixel-count | continuous | rawgreen-mean | continuous |
| short-line-density-5 | continuous | exred-mean | continuous |
| short-line-density-2 | continuous | exblue-mean | continuous |
| vedge-mean | continuous | exgreen-mean | continuous |
| vegde-sd | continuous | value-mean | continuous |
| hedge-mean | continuous | saturatoin-mean | continuous |
| hedge-sd | continuous | hue-mean | continuous |
| intensity-mean | continuous | | |

**ชื่อชุดข้อมูล** Statlog (satimage)

**จำนวนคุณสมบัติ** 36

**จำนวนกลุ่ม** 6 กลุ่มคือ 1, 2, 3, 4, 5 และ 7

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A1 | continuous | A19 | continuous |
| A2 | continuous | A20 | continuous |
| A3 | continuous | A21 | continuous |
| A4 | continuous | A22 | continuous |
| A5 | continuous | A23 | continuous |
| A6 | continuous | A24 | continuous |
| A7 | continuous | A25 | continuous |
| A8 | continuous | A26 | continuous |
| A9 | continuous | A27 | continuous |
| A10 | continuous | A28 | continuous |
| A11 | continuous | A29 | continuous |

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A12 | continuous | A30 | continuous |
| A13 | continuous | A31 | continuous |
| A14 | continuous | A32 | continuous |
| A15 | continuous | A33 | continuous |
| A16 | continuous | A34 | continuous |
| A17 | continuous | A35 | continuous |
| A18 | continuous | A36 | continuous |

**ชื่อชุดข้อมูล** Statlog (segment)

**จำนวนคุณสมบัติ** 19

**จำนวนกลุ่ม** 7 กลุ่มคือ 1, 2, 3, 4, 5, 6 และ 7

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A1 | continuous | A11 | continuous |
| A2 | continuous | A12 | continuous |
| A3 | continuous | A13 | continuous |
| A4 | continuous | A14 | continuous |
| A5 | continuous | A15 | continuous |
| A6 | continuous | A16 | continuous |
| A7 | continuous | A17 | continuous |
| A8 | continuous | A18 | continuous |
| A9 | continuous | A19 | continuous |
| A10 | continuous | | |

**ชื่อชุดข้อมูล** Statlog (shuttle)

**จำนวนคุณสมบัติ** 9

**จำนวนกลุ่ม** 7 กลุ่มคือ 1, 2, 3, 4, 5, 6 และ 7

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A1 | continuous | A6 | continuous |
| A2 | continuous | A7 | continuous |
| A3 | continuous | A8 | continuous |
| A4 | continuous | A9 | continuous |
| A5 | continuous | | |

**ชื่อชุดข้อมูล** Wine

**จำนวนคุณสมบัติ** 13

**จำนวนกลุ่ม** 3 กลุ่มคือ 1, 2 และ 3

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A1 | continuous | A8 | continuous |
| A2 | continuous | A9 | continuous |
| A3 | continuous | A10 | continuous |
| A4 | continuous | A11 | continuous |
| A5 | continuous | A12 | continuous |
| A6 | continuous | A13 | continuous |
| A7 | continuous | | |

**ชื่อชุดข้อมูล** Waveform

**จำนวนคุณสมบัติ** 21

**จำนวนกลุ่ม** 3 กลุ่มคือ 0, 1 และ 2

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A1 | continuous | A12 | continuous |
| A2 | continuous | A13 | continuous |
| A3 | continuous | A14 | continuous |
| A4 | continuous | A15 | continuous |
| A5 | continuous | A16 | continuous |
| A6 | continuous | A17 | continuous |
| A7 | continuous | A18 | continuous |
| A8 | continuous | A19 | continuous |
| A9 | continuous | A20 | continuous |
| A10 | continuous | A21 | continuous |
| A11 | continuous | | |

**ชื่อชุดข้อมูล** Waveform + noise

**จำนวนคุณสมบัติ** 40

**จำนวนกลุ่ม** 3 คือ 0, 1 และ 2

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A1 | continuous | A21 | continuous |
| A2 | continuous | A22 | continuous |
| A3 | continuous | A23 | continuous |
| A4 | continuous | A24 | continuous |
| A5 | continuous | A25 | continuous |
| A6 | continuous | A26 | continuous |
| A7 | continuous | A27 | continuous |
| A8 | continuous | A28 | continuous |
| A9 | continuous | A29 | continuous |
| A10 | continuous | A30 | continuous |
| A11 | continuous | A31 | continuous |
| A12 | continuous | A32 | continuous |

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| A13 | continuous | A33 | continuous |
| A14 | continuous | A34 | continuous |
| A15 | continuous | A35 | continuous |
| A16 | continuous | A36 | continuous |
| A17 | continuous | A37 | continuous |
| A18 | continuous | A38 | continuous |
| A19 | continuous | A39 | continuous |
| A20 | continuous | A40 | continuous |

**ชื่อชุดข้อมูล** Glass

**จำนวนคุณสมบัติ** 9

**จำนวนกลุ่ม** 6 กลุ่มคือ 1, 2, 3, 5, 6 และ 7

| ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ | ชื่อคุณสมบัติ | ค่าที่เป็นไปได้ |
|---|---|---|---|
| RI | continuous | K | continuous |
| Na | continuous | Ca | continuous |
| Mg | continuous | Ba | continuous |
| Al | continuous | Fe | continuous |
| Si | continuous | | |

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ภาคผนวก ง

## ตารางแจกแจงแบบ t (t Distribution)

ตารางแจกแจงแบบ t นี้เป็นตารางที่ใช้ในการหาระดับความมั่นใจในหัวข้อที่ 2.1.5 ซึ่งในงานวิทยานิพนธ์ฉบับนี้ใช้ degrees of freedom (df) มีค่าเท่ากับ 5 สำหรับการทำครอสวาลิเดชันแบบ 6 โฟลด์

| df | ระดับความมั่นใจที่ N (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 60 | 70 | 80 | 90 | 95 | 96 | 98 | 99 | 99.5 | 99.8 | 99.9 |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 |
| 2 | .816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 |
| 3 | .765 | .978 | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 |
| 4 | .741 | .941 | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | .727 | .920 | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | .718 | .906 | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | .711 | .896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | .706 | .889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | .703 | .883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | .700 | .879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | .697 | .876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | .695 | .873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | .694 | .870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14 | .692 | .868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.264 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15 | .691 | .866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | .690 | .865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | .689 | .863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | .688 | .862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.611 | 3.922 |
| 19 | .688 | .861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.205 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20 | .687 | .860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.197 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21 | .663. | .859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.189 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22 | .686 | .858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.183 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |
| 23 | .685 | .858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.177 | 2.500 | 2.807 | 3.104 | 3.485 | 3.768 |
| 24 | .685 | .857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.172 | 2.492 | 2.797 | 3.091 | 3.467 | 3.745 |
| 25 | .684 | .856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.167 | 2.485 | 2.787 | 3.078 | 3.450 | 3.725 |
| 26 | .684 | .856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.162 | 2.479 | 2.779 | 3.067 | 3.435 | 3.707 |
| 27 | .684 | .855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.15 | 2.473 | 2.771 | 3.057 | 3.421 | 3.690 |
| 28 | .683 | .855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.154 | 2.467 | 2.763 | 3.047 | 3.408 | 3.674 |
| 29 | .683 | .854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.150 | 2.462 | 2.756 | 3.038 | 3.396 | 3.659 |
| 30 | .683 | .854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.147 | 2.457 | 2.750 | 3.030 | 3.385 | 3.646 |
| 40 | .681 | .851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.123 | 2.423 | 2.704 | 2.971 | 3.307 | 3.551 |
| 50 | .679 | .849 | 1.047 | 1.295 | 1.676 | 2.009 | 2.109 | 2.403 | 2.678 | 2.937 | 3.261 | 3.496 |
| 60 | .679 | .848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.099 | 2.390 | 2.660 | 2.915 | 3.232 | 3.460 |
| 80 | .678 | .846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.088 | 2.374 | 2.639 | 2.887 | 3.195 | 3.416 |
| 100 | .677 | .845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.081 | 2.364 | 2.626 | 2.871 | 3.174 | 3.390 |
| 1000 | .675 | .842 | 1.037 | 1.282 | 1.646 | 1.962 | 2.056 | 2.330 | 2.581 | 2.813 | 3.098 | 3.300 |
| inf. | .674 | .841 | 1.036 | 1.282 | 1.645 | 1.960 | 2.054 | 2.326 | 2.576 | 2.807 | 3.091 | 3.291 |

# ประวัติผู้เขียนวิทยานิพนธ์

นายวิฆเนศ ทองมี เกิดเมื่อวันที่ 14 กุมภาพันธ์ พ.ศ. 2522 สำเร็จการศึกษาหลักสูตรวิศวกรรมศาสตร์บัณฑิต (วศ.บ.) สาขาวิชาวิศวกรรมยานยนต์ จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2540 จากนั้นเข้าทำงานที่กองทุนบำเหน็จบำนาญข้าราชการในตำแหน่งนักวิจัยและพัฒนา ตั้งแต่ปี 2541 ถึง 2543 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตร์มหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2543

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย