



ใบรับรองวิทยานิพนธ์
บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

วิทยาศาสตร์มหาบัณฑิต (วิทยาการคอมพิวเตอร์)

ปริญญา

วิทยาการคอมพิวเตอร์

วิทยาการคอมพิวเตอร์

สาขา

ภาควิชา

เรื่อง การเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิก

Membership Functions Selection for Fuzzy Logic

นามผู้วิจัย นางสาวภัทรวดี เดียวเดชะเดชา

ได้พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ นवलวรรณ สุนทรภิชัย, Ph.D.)

หัวหน้าภาควิชา

(อาจารย์เสกฐวิทย์ เกิดผล, Ph.D.)

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์รับรองแล้ว

(รองศาสตราจารย์ กัญญา ชีระกุล, D.Agr.)

คณบดีบัณฑิตวิทยาลัย

วันที่ เดือน พ.ศ.

สืบสีทงี มหาวิทยาลัยเกษตรศาสตร์

วิทยานิพนธ์

เรื่อง

การเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิก

Membership Functions Selection For Fuzzy Logic

โดย

นางสาวภัทรวดี เตียวเตชะเดชา

เสนอ

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

พ.ศ. 2553

ลิขสิทธิ์ มหาวิทยาลัยเกษตรศาสตร์

ภัทรวดี เตียวเตชะเดชา 2553: การเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzyลอจิก
ปริญญาวิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์) สาขาวิทยาการคอมพิวเตอร์
ภาควิชาวิทยาการคอมพิวเตอร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์
นวลวรรณ สุนทรภิชช์, Ph.D. 82 หน้า

งานวิจัยนี้เสนอขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzyลอจิก สำหรับการ
การจำแนกประเภทข้อมูลที่มีแอททริบิวต์เป็นตัวเลขต่อเนื่อง ซึ่งกราฟฟังก์ชันความเป็นสมาชิก
เป็นตัวกำหนดค่าระดับความเป็นสมาชิกให้กับแอททริบิวต์ ซึ่งมีความสำคัญต่อกระบวนการคิด
และแก้ไขปัญหาของฟuzzyลอจิกอย่างมาก โดยปกติจะต้องอาศัยผู้เชี่ยวชาญเฉพาะด้านของข้อมูล
นั้นๆ เป็นผู้กำหนดว่าจะใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบใด และต้องอาศัยทั้ง
ประสบการณ์ เวลา การลองผิดลองถูก เพื่อให้ได้ชนิดของกราฟฟังก์ชันความเป็นสมาชิกที่ทำให้
ได้ประสิทธิภาพในการจำแนกประเภทข้อมูลสูงที่สุด

งานวิจัยนี้ใช้การกระจายตัวของแอททริบิวต์ และทำการสร้างสมการการถดถอยเชิงเส้นหลายตัว
แปรเพื่อหาค่าความชัน และเลือกรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกให้สอดคล้องกับแต่
ละแอททริบิวต์ โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม และรูปประฆังคว่ำได้อย่าง
อัตโนมัติ เพื่อทำการจำแนกประเภทข้อมูลของไวน์และข้อมูลของผู้ป่วยโรคมะเร็งเต้านม ผลการ
ทดลองแสดงให้เห็นว่าขั้นตอนวิธีที่เสนอสามารถจำแนกประเภทข้อมูลไวน์โดยใช้ประสิทธิภาพ
เฉลี่ยเท่ากับ 95.01% เมื่อเทียบกับการจำแนกประเภทข้อมูลโดยใช้กราฟฟังก์ชันความเป็นสมาชิก
รูปสามเหลี่ยมเพียงอย่างเดียว (91.08%) หรือรูปประฆังคว่ำเพียงอย่างเดียว (93.30%) สำหรับการ
จำแนกประเภทข้อมูลผู้ป่วยโรคมะเร็งเต้านมเมื่อเทียบกับการจำแนกประเภทข้อมูลโดยใช้กราฟ
ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมเพียงอย่างเดียวได้ค่าความถูกต้องเป็น 93.85% รูปประฆัง
คว่ำเพียงอย่างเดียวได้ค่าความถูกต้องเป็น 94.91% และมีประสิทธิภาพเฉลี่ยเท่ากับ 95.47%

ลายมือชื่อนิติสด

ลายมือชื่อประธานกรรมการ

Phattarawadee Tiewtechadecha 2010: Membership Functions Selection for Fuzzy Logic. Master of Science (Computer Science), Major Field: Computer Science, Department of Computer Science. Thesis Advisor: Assistant Professor Nuanwan Soonthornphisaj, Ph.D. 82 pages.

This research proposes an algorithm that has an ability to determine a suitable form of the membership function in fuzzy logic. The classification is used for data sets that have continuous value attributes. The graph of membership function defines the membership value for an attribute which is significant to fuzzy logic. Generally, the process needs to be analyzed by specialists who can decide the suitable form of membership function. This requires experience, time, and experiment to receive the appropriate membership function that can achieves the highest performance of the classification.

Our approach applies data distribution and uses multi-linear regression techniques to evaluate the slopes for choosing a suitable form of membership function in each attribute. We investigate two kinds of membership functions which are triangular shape and bell-curve in order to automatically classify data for Wines and Breast Cancer data set. Considering the performance on Wines data set, the result shows that the average performance of our approach is 95.01%. whereas the classification uses only triangular membership function gets 91.08%, bell-curve gets 93.30%. For Breast Cancer data set, the performance on triangular membership function is 93.85%, and 94.91% in bell-curve. The average performance is 95.47%

Student's signature

Thesis Advisor's signature

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ ผศ.ดร.นवलวรรณ สุนทรภิชช์ ประธานกรรมการที่ปรึกษา
วิทยานิพนธ์ ที่ได้ให้คำแนะนำ วิธีการและขั้นตอนในการดำเนินงาน ซึ่งแนะแนวแก้ปัญหาและ
ตรวจทานข้อบกพร่องต่างๆ ในงานวิจัยตลอดมา ขอกราบขอบพระคุณ ผศ.ดร.วรเศรษฐ์ สุวรรณิก
และ ดร.ดวงดาว วิชาดากุล ที่กรุณาให้คำแนะนำ และข้อเสนอแนะในการทำวิทยานิพนธ์นี้

ขอกราบขอบพระคุณบิดาและมารดา ที่ส่งเสริม สนับสนุนและคอยช่วยเหลืออย่างดีที่สุด
เสมอมา ขอขอบพระคุณคณาจารย์ทุกท่านที่ให้การอบรม สั่งสอนวิชาความรู้ต่างๆ ของผู้วิจัย

ขอขอบคุณพี่แอร์ นายธนรัฐ รัตนอุบล ที่คอยแนะนำ เอื้อเฟื้อข้อมูล และให้กำลังใจในการ
ทำวิจัยครั้งนี้ ขอขอบคุณนายศักดิ์สิทธิ์ ชูติภาสวงศ์กุล ที่คอยให้กำลังใจ ให้คำปรึกษา และให้ความ
ช่วยเหลือที่ดีตลอดมา และขอขอบคุณเพื่อนๆ พี่ น้องทุกท่านที่ช่วยให้วิทยานิพนธ์นี้สำเร็จลุล่วงไป
ได้ด้วยดี

ขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์ ที่ได้สนับสนุนด้านอุปกรณ์ และสถานที่ตลอด
ระยะเวลาในการทำวิทยานิพนธ์

ภัทรวดี เตียวเตชะเดชา

เมษายน 2553

สารบัญ

	หน้า
สารบัญ	(1)
สารบัญตาราง	(2)
สารบัญภาพ	(3)
คำอธิบายสัญลักษณ์และคำย่อ	(5)
คำนำ	1
วัตถุประสงค์	3
การตรวจเอกสาร	4
อุปกรณ์และวิธีการ	21
อุปกรณ์	21
วิธีการ	21
ผลและวิจารณ์	35
ผล	35
วิจารณ์	43
สรุปและข้อเสนอแนะ	45
สรุป	45
ข้อเสนอแนะ	46
เอกสารและสิ่งอ้างอิง	47
ภาคผนวก	49
ภาคผนวก ก รายละเอียดผลการทดลอง	50
ภาคผนวก ข ผลงานตีพิมพ์	63
ประวัติการศึกษา และการทำงาน	82

สารบัญตาราง

ตารางที่		หน้า
1	แอททริบิวต์ต่างๆของข้อมูลโรคมะเร็งเต้านม	21
2	แอททริบิวต์ต่างๆของข้อมูลไวน์	22
3	ตารางค่าความชันของข้อมูลไวน์	35
4	ผลการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลไวน์	36
5	ประสิทธิภาพเฉลี่ยของการจำแนกประเภทของข้อมูลไวน์โดยใช้ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก	37
6	ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลไวน์โดยใช้ขั้นตอนวิธีฟuzzyลอจิกของกราฟฟังก์ชันความเป็นสมาชิกที่รูปร่างต่างกัน	38
7	ผลการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลโรคมะเร็งเต้านม	39
8	ประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก	40
9	ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีฟuzzyลอจิกของกราฟฟังก์ชันความเป็นสมาชิกที่รูปร่างต่างกัน	41
10	ประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลของโรคกล้ามเนื้ออ่อนแรงและโรคมะเร็งเต้านม โดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียม	42
11	ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลของไวน์ที่ได้จากการใช้กราฟฟังก์ชันความเป็นสมาชิก	43
12	ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลของผู้ป่วยโรคมะเร็งเต้านมที่ได้จากการเปลี่ยนรูปแบบกราฟฟังก์ชันความเป็นสมาชิก	44

สารบัญตาราง (ต่อ)

ตารางผนวกที่		หน้า
ก1	ประสิทธิภาพจากการจำแนกประเภทข้อมูลสอนของข้อมูลไวน์โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก	51
ก2	ประสิทธิภาพจากการจำแนกประเภทข้อมูลทดสอบของข้อมูลไวน์โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก	52
ก3	ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม	53
ก4	ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ	54
ก5	ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปเกาส์เซียน	55
ก6	ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู	56
ก7	ประสิทธิภาพจากการจำแนกประเภทข้อมูลสอนของผู้ป่วยที่เป็นโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก	57
ก8	ประสิทธิภาพจากการจำแนกประเภทข้อมูลทดสอบของผู้ป่วยที่เป็นโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก	58
ก9	ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม	59
ก10	ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ	60
ก11	ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปเกาส์เซียน	61
ก12	ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู	62

สารบัญภาพ

ภาพที่		หน้า
1	หลักการงานของการจำแนกประเภทข้อมูล	6
2	ข้อมูลสอนระบบ (ก) และต้นไม้ตัดสินใจใช้ทำนายว่าจะเล่นกีฬาหรือไม่ (ข)	9
3	กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม เมื่อ $a=0$, $b=5$ และ $c=10$	12
4	ฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน เมื่อ $m=5$ และ $\sigma=1$	13
5	ฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ เมื่อ $a=2$, $b=4$ และ $c=5$	14
6	การยูเนียนของฟังก์ชันความเป็นสมาชิกของ A และ B	15
7	การอินเตอร์เซกชันของฟังก์ชันความเป็นสมาชิกของ A และ B	15
8	การคอมพลิเมนต์ของฟังก์ชันความเป็นสมาชิกของ A	16
9	ตัวอย่างการจำแนกประเภทด้วยฟัซซี่ลอจิก	17
10	กระบวนการทำงานของฟัซซี่ลอจิก	17
11	แผนภาพแสดงขั้นตอนการทำงานของงานวิจัย	24
12	ต้นไม้ตัดสินใจของข้อมูลไวน์	26
13	ต้นไม้ตัดสินใจของข้อมูลลักษณะของเซลล์ของผู้ป่วยที่เป็นโรคมะเร็งเต้านม	27
14	ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก	29
15	ขั้นตอนวิธีการหาประสิทธิภาพของรูปแบบกราฟฟังก์ชันความเป็นสมาชิก	30
16	กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำของแอททริบิวต์ Flavanoids	30
17	กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำของแอททริบิวต์ Color intensity	31
18	กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมของแอททริบิวต์ OD280/OD315 of diluted wines	31
19	ขั้นตอนวิธีการสร้างฟังก์ชันความเป็นสมาชิก	32

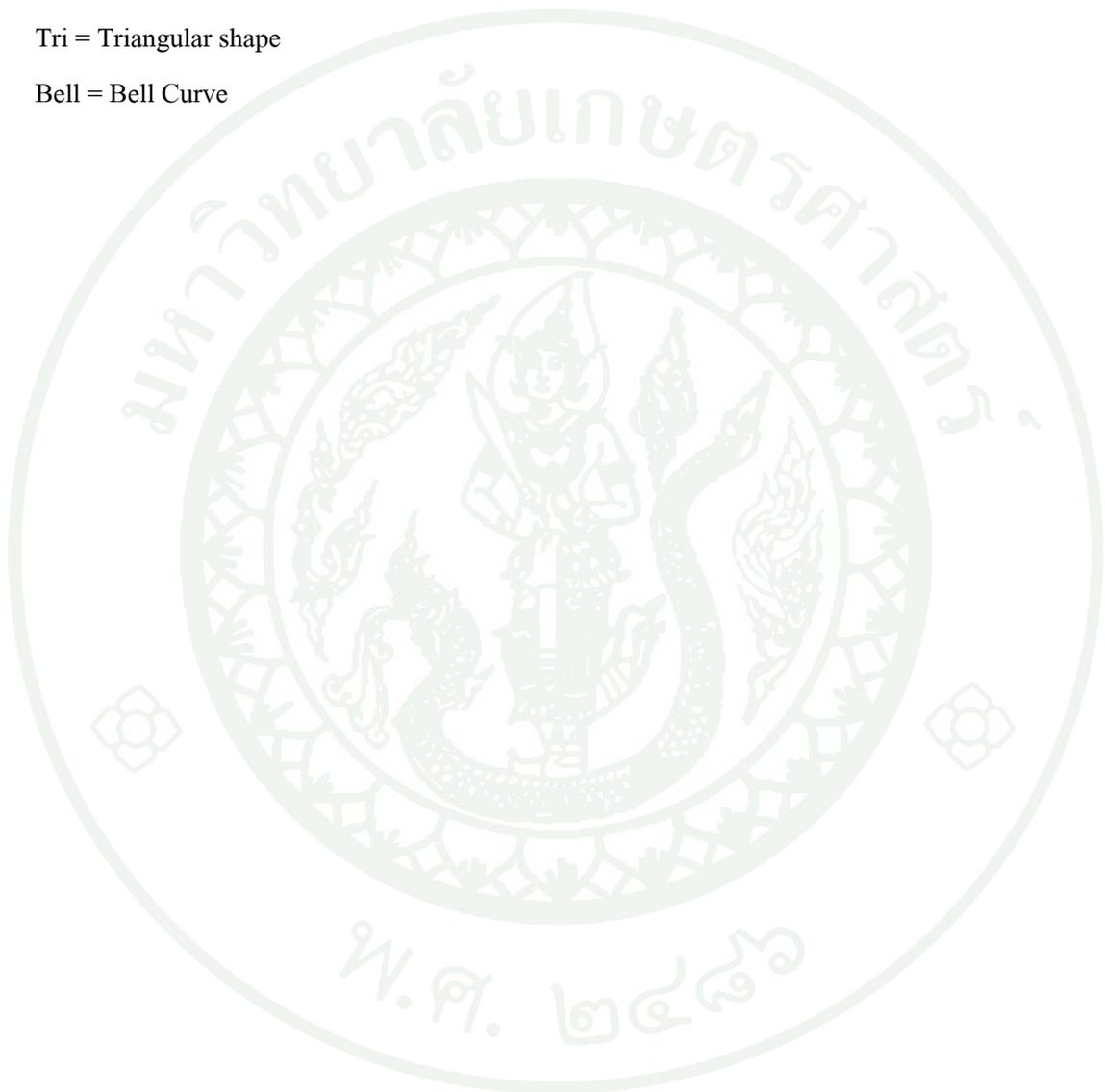
คำอธิบายสัญลักษณ์และคำย่อ

OD280/OD315 = OD280/OD315 of diluted wines

CFS = Chronic Fatigue Syndrome

Tri = Triangular shape

Bell = Bell Curve



การเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzy ลอจิก

Membership Functions Selection for Fuzzy Logic

คำนำ

ขั้นตอนวิธีฟuzzy ลอจิกเป็นเทคนิคหนึ่งในการจำแนกประเภทข้อมูลที่มีความคลุมเครือที่ได้รับ ความสนใจอย่างมาก เพื่อช่วยในการตัดสินใจในสิ่งที่มีความคลุมเครือกัน ฟuzzy ลอจิกเป็นการวิเคราะห์เชิงตรรกะในข้อมูลที่มีความคลุมเครือกัน ยกต่อการตัดสินใจว่าสิ่งนั้นใช่หรือไม่ใช่ อาจเป็นสิ่งที่คลุมเครือ ไม่ชัดเจน เช่น อุณหภูมิของน้ำ ความสูง เป็นต้น เซตของเหตุการณ์ที่ไม่แน่นอน เรียกว่าฟuzzy เซต (Fuzzy Set) โดยฟuzzy เซตก็จะขึ้นกับฟังก์ชันความเป็นสมาชิก (Membership Function) ซึ่งจะทำให้การแบ่งระดับความเป็นสมาชิกให้มีค่าอยู่ระหว่าง 0 กับ 1 จากนั้นฟuzzy เซตต่างๆ ถูกนำมาสร้างความสัมพันธ์และแปลความหมายออกมาอีกครั้งหนึ่ง เพื่อช่วยในการวิเคราะห์ข้อมูล

ฟังก์ชันความเป็นสมาชิกนี้เป็นฟังก์ชันที่มีการกำหนดระดับความเป็นสมาชิกให้กับ แอททริบิวต์ที่ต้องการใช้งาน โดยเริ่มจากการแทนที่กับตัวแทนที่มีความไม่ชัดเจน ไม่แน่นอน และคลุมเครือ ดังนั้นจึงเป็นส่วนที่สำคัญต่อคุณสมบัติหรือการดำเนินการของฟuzzy เพราะรูปร่างของ ฟังก์ชันความเป็นสมาชิกมีความสำคัญต่อกระบวนการคิดและแก้ปัญหา โดยฟังก์ชันความเป็นสมาชิก จะไม่สมมาตรหรือสมมาตรกันก็ได้

การเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzy ลอจิกนี้เพื่อแก้ปัญหาการขาดแคลน ผู้เชี่ยวชาญที่มีความรู้ความสามารถเกี่ยวกับข้อมูลที่ต้องการนำไปใช้ ซึ่งผู้เชี่ยวชาญมีความสำคัญในการกำหนดรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกให้กับข้อมูลนั้นๆ เพื่อนำไปหาค่าระดับ ความเป็นสมาชิก เพื่อให้ได้ประสิทธิภาพของการจำแนกประเภทข้อมูลที่มีความถูกต้อง แม่นยำมากขึ้น และช่วยสนับสนุนการตัดสินใจต่างๆ ได้สะดวกและรวดเร็วมากยิ่งขึ้น

งานวิจัยนี้ได้นำเสนอการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzy ลอจิก ซึ่งเป็นการอาศัยการกระจายตัวของแอททริบิวต์ และทำการสร้างสมการการถดถอยเชิงเส้นหลายตัวแปร เพื่อหาค่าความชัน และเลือกรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกให้สอดคล้องกับแต่ละแอททริบิวต์ โดยใช้กราฟฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมและรูประฆังคว่ำได้อย่างอัตโนมัติ เพื่อทำ

การจำแนกประเภทข้อมูลของไวน์และข้อมูลของผู้ป่วยโรคมะเร็งเต้านม ซึ่งมีการเปรียบเทียบผล
การจำแนกประเภทข้อมูลโดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมเพียงอย่างเดียว และ
รูปประฆังคว่ำเพียงอย่างเดียว และเปรียบเทียบผลการจำแนกประเภทกับขั้นตอนวิธีโครงข่าย
ประสาทเทียมอีกด้วย



วัตถุประสงค์

1. เพื่อสร้างกราฟฟังก์ชันความเป็นสมาชิกโดยใช้ขั้นตอนวิธีการเรียนรู้
2. เพื่อศึกษาถึงประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกให้เหมาะสมกับข้อมูล

ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถสร้างฟังก์ชันความเป็นสมาชิกได้ โดยไม่จำเป็นต้องให้ผู้เชี่ยวชาญที่มีความรู้เกี่ยวกับข้อมูลนั้นเป็นผู้สร้าง
2. เพื่อให้ประสิทธิภาพการจำแนกประเภทข้อมูลมีความถูกต้องแม่นยำมากยิ่งขึ้น
3. ได้ขั้นตอนวิธีในการสร้างกราฟฟังก์ชันความเป็นสมาชิกสำหรับข้อมูลต่างๆ เพื่อช่วยสนับสนุนการตัดสินใจในอนาคต

ขอบเขตและข้อจำกัด

1. กราฟฟังก์ชันความเป็นสมาชิกที่ทำการศึกษาอยู่ในรูปแบบสามเหลี่ยม และรูปประฆังคว่ำ
2. ใช้ค่า Sensitivity, Specificity และ Accuracy ในการวัดประสิทธิภาพการทำงาน
3. ข้อมูลที่นำมาทดลองต้องเป็นข้อมูลประเภทตัวเลขเท่านั้น ถ้าไม่ใช่ต้องทำการแปลง
 อย่างเช่น ชื่อคลาสต้องแปลงให้อยู่ในรูปของตัวเลข 0-1 ก่อน เนื่องจากการทำงานของฟัซซี่ลอจิกใช้
 ข้อมูลที่เป็นตัวเลขเท่านั้น
4. ทดสอบกับปัญหาการจำแนกประเภทข้อมูลของไวน์และผู้ป่วยที่เป็นโรคมะเร็งเต้านม

การตรวจเอกสาร

1. งานวิจัยที่เกี่ยวข้อง

ในปัจจุบันมีงานวิจัยต่างๆ ที่นำเอาขั้นตอนวิธีฟัซซี่ลอจิกซึ่งเป็นเทคนิคหนึ่งในการจำแนกประเภทข้อมูลที่มีความคลุมเครือที่ได้รับความนิยมอย่างมาก เพื่อช่วยในการตัดสินใจในสิ่งที่มีความคลุมเครือกัน โดยทฤษฎีฟัซซี่ลอจิกซึ่งเป็นงานวิจัยของ Zadeh (1994) ได้กล่าวถึงฟัซซี่ลอจิกไว้ว่าเป็นการวิเคราะห์เชิงตรรกะในข้อมูลที่มีความคลุมเครือกัน ยกต่อการตัดสินใจว่าสิ่งนั้นใช่หรือไม่ใช่ อาจเป็นสิ่งที่คลุมเครือ ไม่ชัดเจน เช่น อุณหภูมิของน้ำ ความสูง เป็นต้น เซตของเหตุการณ์ที่ไม่แน่นอนเรียกว่าฟัซซี่เซต (Fuzzy Set) โดยฟัซซี่เซตก็จะขึ้นกับฟังก์ชันความเป็นสมาชิก (Membership Function) ซึ่งจะทำการแบ่งระดับความเป็นสมาชิกให้มีค่าอยู่ระหว่าง 0 กับ 1 จากนั้นฟัซซี่เซตต่างๆ ถูกนำมาสร้างความสัมพันธ์และแปลความหมายออกมาอีกครั้งหนึ่ง เพื่อช่วยในการวิเคราะห์ข้อมูล

งานวิจัยของ Pedrycz (1994) ได้ศึกษาว่าเพราะเหตุใดจึงมีการใช้ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมแทนค่าข้อมูลโดยสร้าง โมเดลขึ้นมาเป็น โครงสร้างที่ประกอบด้วย 3 ส่วนคือ ส่วนของข้อมูลนำเข้า ส่วนการประมวลผล และส่วนของข้อมูลผลลัพธ์ ซึ่งข้อมูลนำเข้าได้จากการสร้างโดยอาศัยสิ่งแวดล้อม อย่างในกรณีข้อมูลที่เป็นตัวเลขที่มาจากการแมพของเครื่องจักรกล เช่น ความน่าจะเป็นและเครื่องมือวัดที่จำเป็น จากนั้นทำการแปลงค่าให้เป็นค่าความเป็นสมาชิกที่มีค่าระหว่าง 0 ถึง 1 และทำให้อยู่ในรูปของเวกเตอร์ ในส่วนของตัวประมวลผลจะทำการกำหนดความสัมพันธ์ระหว่างส่วนของข้อมูลนำเข้าและส่วนของข้อมูลผลลัพธ์ของระบบ และบทบาทของส่วนข้อมูลผลลัพธ์เป็นการแปลเวกเตอร์ความเป็นสมาชิกไปเป็นค่าจำนวนจริง ผลที่ได้ภายใต้ข้อสมมติฐานของฟัซซี่เซตสามเหลี่ยมนี้ได้รับความพอใจของการที่ไม่มีข้อผิดพลาดของเกณฑ์สำหรับข้อมูลผลลัพธ์

งานวิจัยของ Lotfi and Tsoi (1994) ได้ทำการสำรวจถึงความสำคัญของฟังก์ชันความเป็นสมาชิก โดยศึกษาการเปรียบเทียบบนความแตกต่างของวิธีการเรียนรู้สำหรับระบบการอนุมานฟัซซี่ ซึ่งได้ทดสอบผลกระทบของฟังก์ชันความเป็นสมาชิกบนกระบวนการที่มีเหตุผล เมื่อจำนวนกฎถูกกำหนดแน่นอน โดยใช้รูปแบบฟังก์ชันความเป็นสมาชิกสามรูปแบบ คือ รูปสามเหลี่ยม รูปแบบเกาส์เซียน และรูปประฆังคว่ำ (Cauchy bell shape membership function) จากการศึกษาพบว่าฟังก์ชัน

ความเป็นสมาชิกมีผลกระทบต่อระบบการอนุมานฟัซซี่มากกว่าจำนวนกฎหรือโครงสร้างการอนุมาน

งานวิจัยที่นำฟัซซี่ลอจิกมาประยุกต์ใช้กับการจำแนกประเภทข้อมูลของ Yang (Yang et al., 1997) ได้นำเสนอวิธีการจำแนกประเภทข้อมูลด้วยฟัซซี่ ซึ่งใช้ข้อมูล synthesizes overlapping data และ Electromyogram (EMG) และใช้ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม ในแต่ละกลุ่มข้อมูลจะมีกฎหนึ่งกฎ เอาท์พุทที่ได้ในการสรุปกฎถูกกำหนดค่าเริ่มต้นด้วยการเลือกปกติในกลุ่มที่ตรงกัน ฟัซซี่เซตและเอาท์พุทจะถูกปรับค่าด้วยวิธีการ error back-propagation ในภายหลัง ผลที่ได้ในการนำเสนอระบบฟัซซี่คือ ได้โครงสร้างที่ดูง่าย สามารถถูกสอนอย่างรวดเร็วและง่ายในการนำไปใช้ประโยชน์ ประสิทธิภาพในการจำแนกของระบบฟัซซี่โดยทั่วไปจะดีกว่าแบบโครงข่ายประสาทเทียมแบบสามชั้น

งานวิจัยหลายงานที่นำเสนอวิธีการเพิ่มประสิทธิภาพในการจำแนกข้อมูลโดยใช้ฟัซซี่ลอจิก อย่างเช่นงานวิจัยของ ธนรัฐ (2007) นำเสนอวิธีการจำแนกประเภทข้อมูลโรคกล้ามเนื้ออ่อนแรง โดยการทดลองได้วิเคราะห์จากข้อมูลของผู้ป่วยจำนวน 187 ราย ซึ่งข้อมูลของผู้ป่วยแบ่งออกเป็น 2 ส่วน คือ ส่วนที่หนึ่งเป็นข้อมูลผลการตรวจเลือดของผู้ป่วย และส่วนที่สองเป็นผลประเมินอาการเจ็บป่วย ผู้วิจัยได้เสนอถึงวิธีการจำแนกประเภทข้อมูลออกเป็น 2 วิธีด้วยกัน คือ 1) ใช้ฟัซซี่ลอจิก 2) ใช้ฟัซซี่ลอจิกร่วมกับวิธีการรวมผลจำแนก โดยผลที่ได้จากการทดลองพบว่าประสิทธิภาพที่ได้จากการใช้ฟัซซี่ลอจิกร่วมกับวิธีการรวมผลการจำแนกให้ค่าความถูกต้อง 92.96% ส่วนวิธีการใช้ฟัซซี่ลอจิกเพียงอย่างเดียวให้ค่าความแม่นยำที่ 83.37%

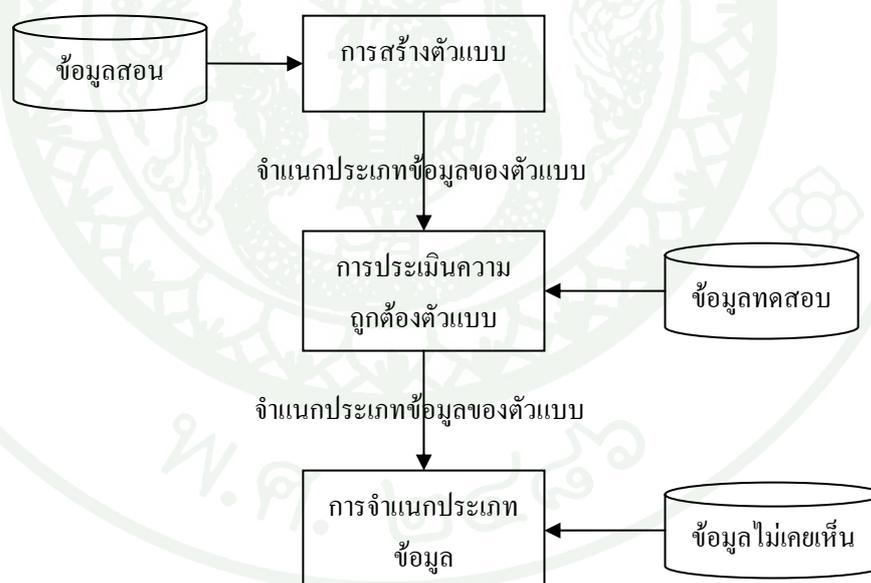
งานวิจัยของพิทักษ์และคณะ (2005) นำเสนอการจัดกลุ่มโรคตับแบบอัตโนมัติด้วยขั้นตอนวิธีนิวโร-ฟัซซี่ โดยใช้ข้อมูลโรคตับจำนวน 536 รายสำหรับการเรียนรู้และการทดสอบภายใต้ 3-fold Cross-validation โดยใช้ข้อมูล 70% จากข้อมูลทั้งหมดและใช้ข้อมูล 30% จากข้อมูลที่เหลือทดสอบ ประสิทธิภาพของโมเดล ซึ่งในขั้นแรกของการนำข้อมูลเข้าสู่การสร้างโมเดลการจัดกลุ่มข้อมูลโรคตับด้วยนิวโร-ฟัซซี่เป็นการจัดกลุ่มข้อมูลในรูปแบบของฟังก์ชันความเป็นสมาชิกรูปแบบสามเหลี่ยมโดยมีคลาสจำนวน 4 คลาส ได้ประสิทธิภาพโดยรวมของระบบเท่ากับ 94%

2. ทฤษฎีที่เกี่ยวข้อง

ในส่วนของทฤษฎีที่เกี่ยวข้องประกอบด้วย การจำแนกประเภทข้อมูล ขั้นตอนวิธีการในการจำแนกประเภทข้อมูล ต้นไม้ตัดสินใจ ฟิชชี่ลอจิก และโครงข่ายประสาทเทียม

2.1 การจำแนกประเภทข้อมูล (Data classification)

ทฤษฎีการจำแนกประเภทข้อมูล (กฤษณะ, 2549) เป็นเทคนิคที่ใช้ในการสกัดหา กฎความสัมพันธ์ในฐานข้อมูล เพื่อจำแนกประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่างๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ เป็นต้น ซึ่งการจำแนกประเภทข้อมูลมีขั้นตอนการทำงานอยู่ 3 ขั้นตอน ดังนี้



ภาพที่ 1 แสดงถึงหลักการทำงานของ การจำแนกประเภทข้อมูล

2.2 ขั้นตอนวิธีในการจำแนกประเภทข้อมูล

2.2.1 การเรียนรู้แบบมีการสอน (Supervised Learning)

การเรียนรู้แบบมีการสอนเป็นกระบวนการเรียนรู้เพื่อหาความสัมพันธ์ของข้อมูลนำเข้า และข้อมูลผลลัพธ์ที่ทราบคำตอบให้ระบบทำการเรียนรู้ ชุดข้อมูลที่ใช้ฝึกระบบจะมีคำตอบไว้เพื่อตรวจสอบว่าระบบให้คำตอบที่ถูกต้องหรือไม่ ถ้าให้คำตอบไม่ถูกต้อง ระบบก็ทำการปรับตัวเองเพื่อให้ได้คำตอบที่ดีขึ้น เสมือนเป็นผู้สอนและสร้างเป็นระบบการเรียนรู้ เพื่อใช้ในการประมาณค่าผลลัพธ์จากข้อมูลนำเข้าชุดใหม่ โดยใช้การเรียนรู้ประเภทการจำแนกประเภท ซึ่งใช้จำแนกประเภทของข้อมูลผลลัพธ์ที่ต้องการ

2.2.2 การเรียนรู้แบบกึ่งมีการสอน (Semi-Supervised Learning)

การเรียนรู้แบบกึ่งมีการสอนเป็นกระบวนการเรียนรู้ที่ใช้ข้อมูลแบบไม่ทราบคำตอบ และข้อมูลบางส่วนทราบคำตอบควบคู่กันไปในกระบวนการเรียนรู้ เพราะบางครั้งขั้นตอนการสร้างชุดข้อมูลโดยเพิ่มคำตอบเข้าไปนั้น อาจจะยุ่งยากเพราะการกำหนดคำตอบต้องอาศัยผู้ที่มีความรู้ในระดับหนึ่งเกี่ยวกับข้อมูลนั้นๆ และในบางครั้งขั้นตอนการกำหนดคำตอบอาจจะต้องใช้เวลาอันยาวนานหากข้อมูลนั้นมีจำนวนมาก ในขณะที่ขั้นตอนการสร้างชุดข้อมูลที่ไม่มีคำตอบอาจจะทำได้ง่ายกว่า แต่เราไม่สามารถใช้ข้อมูลเหล่านี้ในการเรียนรู้แบบมีการสอนได้

2.3 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Quinlan, 1986) เป็นการเรียนรู้ที่แทนความรู้ในรูปต้นไม้ตัดสินใจ ใช้สำหรับจำแนกประเภทของข้อมูล ซึ่งอาศัยค่าความไร้ระเบียบของข้อมูล (Entropy) และค่าเกน (Gain) ในการเลือกแอททริบิวต์ที่มีอำนาจในการจำแนกข้อมูลสูง โดยแสดงผลลัพธ์ในรูปของโครงสร้างต้นไม้ซึ่งสามารถแปลงเป็นกฎที่เข้าใจได้ง่าย ต้นไม้ตัดสินใจประกอบด้วย โหนด (Node) และกิ่ง (Link) ที่ต่อกับโหนด โดยที่แต่ละโหนดแสดงถึงคุณสมบัติ และกิ่งแสดงถึงค่าของคุณสมบัตินั้น โหนดปลายสุดเรียกว่า โหนดใบ (Leaf node) แสดงถึงประเภทข้อมูลที่กำหนดไว้ การสร้างต้นไม้ตัดสินใจทำโดยสร้างโหนดทีละโหนด เพื่อตรวจสอบคุณสมบัติของข้อมูล แล้วแยก

ประเภทลงตามค่าของกึ่ง ทำจนกระทั่งข้อมูลแต่ละใบอยู่ในประเภทเดียวกันทั้งหมด สูตรในการคำนวณค่าความไร้ระเบียบและค่าเกน คือ

$$Entropy(S) = -p_+ \log_2 p_+ - p_- \log_2 p_- \quad (1)$$

โดยที่ S คือ ชุดข้อมูลสอนระบบ S

p_+ คือ ค่าความน่าจะเป็นที่พบคลาสที่สนใจ (Positive)

p_- คือ ค่าความน่าจะเป็นที่พบคลาสที่ไม่สนใจ (Negative)

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

โดยที่ A คือ ข้อมูลแอททริบิวต์ A

$|S_v|$ คือ จำนวนคลาสที่พบในแอททริบิวต์ A ที่มีข้อมูลเป็น v

$|S|$ คือ จำนวนคลาสทั้งหมดที่พบในชุดข้อมูลสอนระบบ S

$Entropy(S)$ คือ ค่าความไร้ระเบียบของชุดข้อมูลสอนระบบ S

$Entropy(S_v)$ คือ ค่าความไร้ระเบียบของแอททริบิวต์ A ที่มีข้อมูลเป็น v

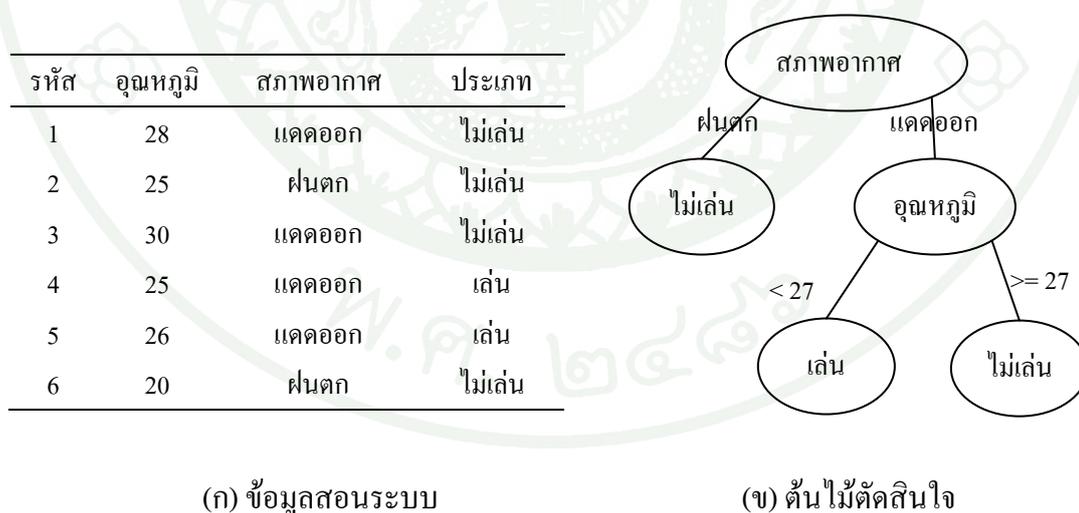
ขั้นตอนวิธีพื้นฐานที่ง่ายที่สุดของการสร้างต้นไม้ตัดสินใจ คือ ขั้นตอนวิธีเชิงละโมบ (Greedy Algorithm) โดยการสร้างต้นไม้จากบนลงล่างแบบวนซ้ำ (Recursive) ด้วยวิธีการแบ่งปัญหาใหญ่เป็นปัญหาย่อย (Divide-And-Conquer) ซึ่งต้นไม้ตัดสินใจนี้ถูกสร้างขึ้นโดยการเรียนรู้จากข้อมูลสอนระบบเป็นหลัก วิธีการทำงาน คือ ในขั้นตอนแรกต้นไม้ทำการสร้างโหนดแรกขึ้น จากนั้นจะทำการทดสอบข้อมูล ถ้าข้อมูลทั้งหมดเป็นข้อมูลประเภทเดียวกัน โหนดนั้นจะกลายเป็นโหนดใบ ถ้าไม่ใช่จะมีการคำนวณหาค่าเกน ของแต่ละแอททริบิวต์ ขึ้นต่อมาจะทำการเลือกแอททริบิวต์ที่มีค่าเกนสูงสุด เพื่อใช้แอททริบิวต์นั้นเป็นโหนดที่แบ่งแยกข้อมูล จากนั้นทำการสร้างกิ่งของต้นไม้และแบ่งข้อมูลออกเป็นส่วน แล้ววนทำงานอย่างนี้ซ้ำไปเรื่อยๆ จนกว่าจะไม่สามารถแยกความแตกต่างของข้อมูลได้อีก

ขั้นตอนการตัดเล็มต้นไม้ (Pruning) เป็นขั้นตอนการตัดโหนดในต้นไม้ตัดสินใจ เพื่อลดความซับซ้อนและการรู้จำจากข้อมูลสอนระบบมากเกินไป ทำให้สามารถจำแนกประเภทข้อมูลได้เร็วขึ้น ซึ่งการตัดเล็มต้นไม้ตัดสินใจแบ่งออกเป็น 2 รูปแบบด้วยกัน ดังนี้

1) การตัดเล็มาก่อนการสร้างต้นไม้เสร็จเต็มรูป (Pre-Pruning) เป็นขั้นตอนการหยุดสร้างโหนดเพิ่ม เมื่อคลาสที่พบในข้อมูลนั้นเป็นคลาสประเภทเดียวกันหมด และหยุดสร้างโหนดเพิ่มเมื่อพบว่าค่าของแอททริบิวต์นั้นเป็นค่าเดียวกันทั้งหมด

2) การตัดเล็มหลังจากสร้างต้นไม้เสร็จแล้วเต็มรูป (Post-Pruning) คือ ขั้นตอนการตัดเล็มโหนดจากล่างขึ้นบน โดยเลือกต้นไม้ย่อยที่มีความลึกมากที่สุดแล้วทำการตัดต้นไม้ย่อยที่เลือกออก จากนั้นแทนต้นไม้ย่อยนั้นด้วยโหนดใบ ซึ่งการเลือกคลาสที่ใส่ในโหนดใบนั้นได้มาจากการหาคลาสที่ปรากฏอยู่ในต้นไม้ย่อยที่ตัดทิ้งไปมากที่สุด (Majority Vote)

ตัวอย่างของต้นไม้ตัดสินใจ ซึ่งจะแสดงผลลัพธ์ในรูปของโครงสร้างต้นไม้ ประกอบด้วย โหนดและกิ่งที่ต่อกับโหนด โดยที่แต่ละโหนดแสดงถึงคุณสมบัติ และกิ่งแสดงถึงค่าของคุณสมบัตินั้น โหนดปลายสุดเรียกว่า โหนดใบ แสดงถึงประเภทข้อมูลที่กำหนดไว้ ดังภาพที่ 2



ภาพที่ 2 ข้อมูลสอนระบบ (ก) และต้นไม้ตัดสินใจใช้ทำนายว่าจะเล่นกีฬาหรือไม่ (ข)

การเปลี่ยนต้นไม้เป็นกฎ (บุญเสริม, 2546) เมื่อสร้างต้นไม้ตัดสินใจแล้วสามารถเปลี่ยนต้นไม้ให้อยู่ในรูปของกฎ เพื่อใช้ในกรณีที่ระบบใช้การแทนความรู้ของกฎเป็นหลัก วิธีการแปลง

ต้นไม้เป็นกฎ IF-THEN ทำได้โดยแสดงทุกเส้นทางเริ่มต้นจากโหนดรากไปยังโหนดใบ และทุกครั้งที่พบโหนดทดสอบก็ให้เพิ่มโหนดทดสอบกับค่าของการทดสอบไว้ในส่วนของ IF และเมื่อพบโหนดใบก็ให้ใส่ประเภทไว้ในส่วนของ THEN ตัวอย่างเช่น จากภาพที่ 2 จะได้กฎดังนี้

- 1) IF สภาพอากาศ เป็น ฝนตก THEN ไม่เล่นกีฬา
- 2) IF สภาพอากาศ เป็น แดดออก AND อุณหภูมิ < 27
THEN เล่นกีฬา
- 3) IF สภาพอากาศ เป็น แดดออก AND อุณหภูมิ >= 27
THEN ไม่เล่นกีฬา

การเปลี่ยนต้นไม้เป็นกฎมีข้อดีที่สำคัญคือ โหนดภายในในต้นไม้ตัดสินใจโหนดหนึ่งๆ ที่สนใจ จะปรากฏในกฎมากกว่าหนึ่งข้อ ต่างจากต้นไม้ตัดสินใจที่โหนดนั้นๆ จะปรากฏอยู่ครั้งเดียว และแยกตามค่าที่เป็นไปได้ของคุณสมบัติไปแต่ละกิ่ง หากเข้าสู่วิธีการตัดเล็มและโหนดนี้ถูกตัดเล็มไปแล้ว ก็จะทำให้คุณสมบัตินั้นหายไปจากต้นไม้ตัดสินใจทันที แต่ต่างจากการเปลี่ยนต้นไม้เป็นกฎจะทำให้เกิดความยืดหยุ่นในการตัดเล็มมากกว่าจะกระทำโดยตรงกับต้นไม้ตัดสินใจ กล่าวคือโหนดนั้นสามารถจะถูกตัดออกจากกฎข้อหนึ่ง แต่ไม่จำเป็นต้องถูกตัดออกจากกฎอีกข้อหนึ่งได้

2.4 ฟัซซีลอจิก (Fuzzy Logic)

ฟัซซีลอจิก คือ ศาสตร์ด้านการคำนวณที่เข้ามามีบทบาทมากขึ้นในวงการวิจัยด้านคอมพิวเตอร์ และได้ถูกนำไปประยุกต์ใช้ในงานต่างๆ มากมาย เช่น ด้านการแพทย์ ด้านการทหาร ด้านธุรกิจ ด้านอุตสาหกรรม เป็นต้น ซึ่งมีความสามารถในการปรับเปลี่ยนระบบได้โดยอัตโนมัติตามสภาพแวดล้อมที่เปลี่ยนไป มีการตัดสินใจแบบชาญฉลาดเทียบมนุษย์ได้มากขึ้น ซึ่งมนุษย์สามารถแก้ปัญหาต่างๆ ที่ไม่เคยพบได้โดยอาศัยความรู้เท่าที่ได้เรียนรู้มาก่อนนำมาประยุกต์ในการแก้ปัญหาได้อย่างมีประสิทธิภาพ ฟัซซีลอจิกถูกคิดค้นโดย L. A. Zadeh ในปี ค.ศ. 1965 (Zadeh, 1994) ฟัซซีลอจิกเป็นรูปแบบหนึ่งของตรรกศาสตร์ที่นำเอาบูลีนลอจิก (Boolean Logic) มาใช้ในการนิยามเหตุการณ์หรือสถานการณ์ที่ไม่สามารถอธิบายได้อย่างชัดเจน ความสำคัญของการใช้ฟัซซีลอจิกอยู่ที่โดยธรรมชาติแล้วการให้เหตุผล หรือการระบุค่าโดยอาศัยเหตุผลตามสามัญสำนึกของคนเราส่วนใหญ่มักจะเป็นไปในลักษณะของการประมาณ อย่างเช่น เมื่อเวลาเราเปิดก๊อกน้ำแล้ว

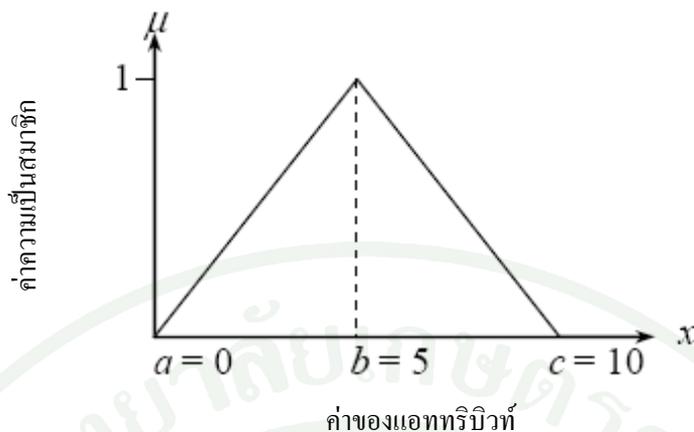
ทำการตรวจสอบอุณหภูมิของน้ำว่าเป็นอย่างไร ถ้าน้ำร้อนมากเราก็จะหมุนลูกบิดมากเพื่อให้น้ำเย็นขึ้น แต่ถ้าน้ำร้อนไม่มากนักเราก็หมุนลูกบิดเล็กน้อย แต่ถ้าน้ำร้อนกำลังพอดีเราก็ไม่จำเป็นต้องปรับลูกบิด ซึ่งเราไม่เคยนิยามว่า ร้อนมาก ร้อนน้อย หรือว่าพอดี ต้องมีค่าเท่าไร สิ่งเหล่านี้เรียนรู้ได้จากประสบการณ์และค่อนข้างจะแตกต่างกันไปขึ้นอยู่กับแต่ละบุคคล ซึ่งเรียกเหตุการณ์ที่ไม่แน่นอนนี้ว่า ฟัซซีเซต

ในทฤษฎีฟัซซีเซตนั้นสามารถแก้ปัญหาข้อจำกัดของเซตแบบดั้งเดิมที่มีค่าความเป็นสมาชิกเป็น 0 หรือ 1 ถูก หรือ ผิด เพียงอย่างเดียวเท่านั้น แต่ในฟัซซีเซตยอมให้มีค่าความเป็นสมาชิก (degree of membership - μ) แสดงค่าตัวเลขตั้งแต่ 0 ถึง 1 หรือเขียนเป็นสัญลักษณ์ $[0, 1]$ โดยที่เลข 0 หมายถึง ไม่เป็นสมาชิกในเซต และเลข 1 หมายถึง เป็นสมาชิกในเซต และค่าระหว่าง 0 กับ 1 เป็นสมาชิกบางส่วนในเซต

องค์ประกอบที่สำคัญในการสร้างฟัซซีลอจิกประกอบไปด้วย 3 ส่วนดังนี้

1) ฟังก์ชันความเป็นสมาชิก คือ ฟังก์ชันที่กำหนดระดับค่าความเป็นสมาชิกของแอททริบิวท์ ซึ่งค่าความเป็นสมาชิกของแอททริบิวท์นั้นจะขึ้นอยู่กับรูปแบบของกราฟฟังก์ชันความเป็นสมาชิก ดังนั้นฟังก์ชันความเป็นสมาชิกจึงเป็นส่วนที่สำคัญต่อกระบวนการคิดและแก้ปัญหา โดยที่กราฟฟังก์ชันความเป็นสมาชิกจะสมมาตรหรือไม่สมมาตรกันทุกประการก็ได้ แล้วแต่ชนิดของฟังก์ชันความเป็นสมาชิก ซึ่งฟังก์ชันความเป็นสมาชิกที่ใช้งานทั่วไปมีหลายชนิดเช่น

ก) ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม (Triangular Membership Function)
การสร้างฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมสามารถทำได้โดยการกำหนดตำแหน่งจุดมุมทั้ง 3 มุม โดยจุดแรกเป็นจุดที่บอกตำแหน่งของจุดมุมที่ฐานด้านซ้ายของสามเหลี่ยม (a) จุดที่สองบอกถึงจุดยอดของสามเหลี่ยม (b) และจุดที่สามบอกถึงจุดมุมที่ฐานด้านขวาของสามเหลี่ยม (c)
ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ $a = 0$, $b = 5$ และ $c = 10$ จะได้กราฟดังที่แสดงในภาพที่ 3



ภาพที่ 3 กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม เมื่อ $a=0$, $b=5$ และ $c=10$

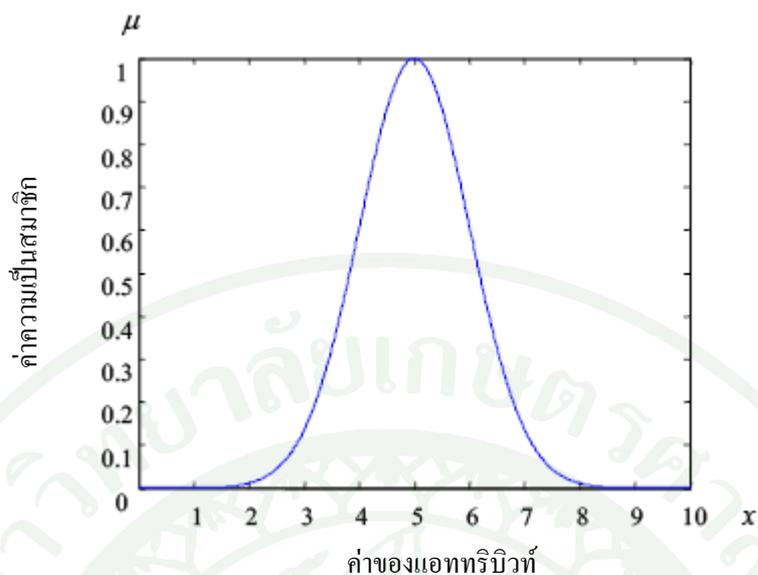
ที่มา: พยุง (2550)

สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

$$Triangular(x: a,b,c) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ \frac{c-x}{b-a} & b \leq x \leq c \\ 0 & x > c \end{cases} \quad (3)$$

โดยที่ x แทน ค่าของแอททริบิวต์

ข) ฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียน (Gaussian membership function) เป็นฟังก์ชันความเป็นสมาชิกที่รับค่าพารามิเตอร์ทั้งหมด 2 ค่า คือ ค่า m หมายถึงค่าเฉลี่ย และ σ หมายถึง ค่าเบี่ยงเบนมาตรฐาน ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ให้ $m = 5$ และ $\sigma = 1$ จะได้กราฟดังแสดงในภาพที่ 4



ภาพที่ 4 ฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน เมื่อ $m=5$ และ $\sigma=1$

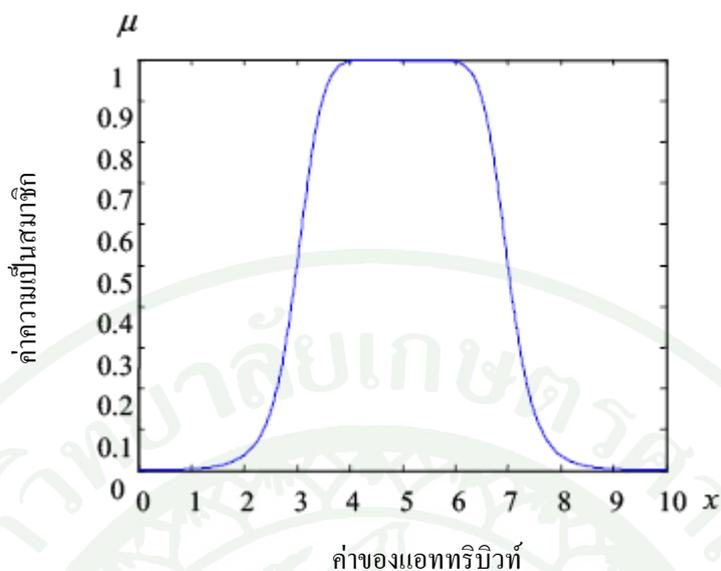
ที่มา: พยุง (2550)

สมการคำนวณค่าความเป็นสมาชิกมีดังนี้

$$Gaussian(x : m, \sigma) = \exp\left(-\frac{(x-m)^2}{\sigma^2}\right) \quad (4)$$

โดยที่ x แทน ค่าของแอมทริบิวท์

- ค) ฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำ (Bell-shaped membership function)
ฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำที่รับค่าพารามิเตอร์ทั้งหมด 3 ค่าคือ (a, b, c) ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ $a = 2$, $b = 4$ และ $c = 5$ จะได้กราฟดังแสดงในภาพที่ 5



ภาพที่ 5 ฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำ เมื่อ $a=2$, $b=4$ และ $c=5$

ที่มา: พยุ่ง (2550)

สมการคำนวณค่าความเป็นสมาชิกมีดังนี้

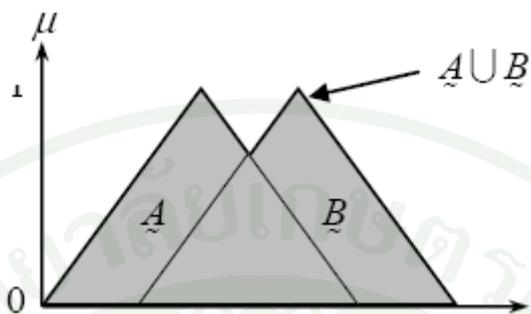
$$\text{Bell-shaped}(x: a, b, c) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}} \quad (5)$$

โดยที่ x แทน ค่าของเอททริบิวท์

2) ตัวดำเนินการของฟัซซี่ลอจิก ประกอบด้วย คำสั่งยูเนียน (Union) อินเตอร์เซกชัน (Intersection) และคำสั่งคอมพลิเมนต์ (Complement) ทั้ง 3 คำสั่งนี้มีคุณสมบัติที่เหมือนกับคำสั่งในเซตต่างๆไป ซึ่งมีรายละเอียดดังนี้

ก) คำสั่งยูเนียน (Union) เป็นคำสั่งในการเชื่อมฟังก์ชันความเป็นสมาชิกของเซตทั้ง 2 เซตรวมเข้าด้วยกัน โดยเขียนให้อยู่ในรูปสมการของฟัซซี่ลอจิกได้ดังนี้

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (6)$$

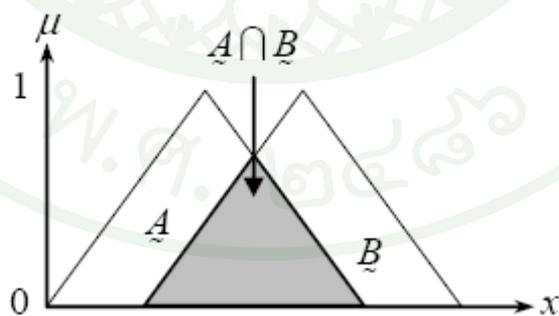


ภาพที่ 6 การยูเนียนของฟังก์ชันความเป็นสมาชิกของ A และ B

ที่มา: พยุง (2550)

ข) คำสั่งอินเตอร์เซกชัน (Intersection) เป็นคำสั่งในการเลือกฟังก์ชันความเป็นสมาชิกของเซตทั้ง 2 เซตที่มีค่าเหมือนกันเท่านั้น เขียนให้อยู่ในรูปสมการของพีชคณิตได้ดังนี้

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \quad (7)$$

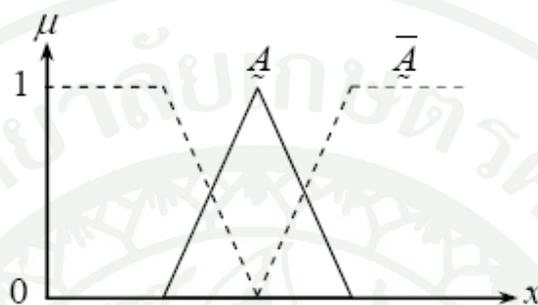


ภาพที่ 7 การอินเตอร์เซกชันของฟังก์ชันความเป็นสมาชิกของ A และ B

ที่มา: พยุง (2550)

ค) คำสั่งคอมพลิเมนต์ (Complement) เป็นคำสั่งที่เลือกฟังก์ชันความเป็นสมาชิกที่ไม่อยู่ในเซตนั้น โดยเขียนให้อยู่ในรูปสมการของฟัซซี่ลอจิกได้ดังนี้

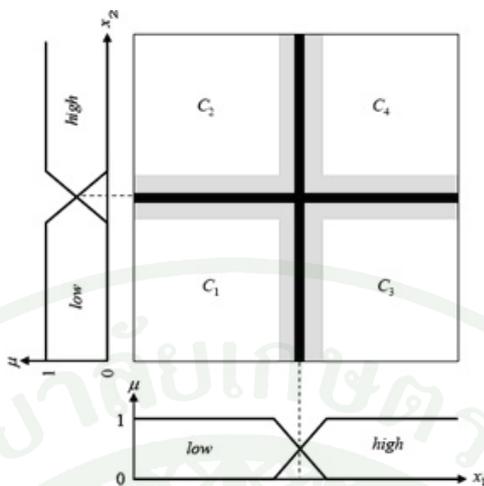
$$\mu_{\bar{A}} = 1 - \mu_A(x) \quad (8)$$



ภาพที่ 8 การคอมพลิเมนต์ของฟังก์ชันความเป็นสมาชิกของ A

ที่มา: พยุง (2550)

3) กฎฟัซซี่ (Fuzzy Rules) เป็นส่วนที่สำคัญในการทำงานที่มีไว้สำหรับช่วยตัดสินใจเลือกข้อมูล ซึ่งกฎฟัซซี่ลอจิกนี้จำเป็นต้องอาศัยผู้เชี่ยวชาญในการใส่กฎลงในฟัซซี่ลอจิก ซึ่งกฎของฟัซซี่ลอจิกอยู่ในรูปคำสั่งเงื่อนไข ถ้า...แล้ว (If-Then) ตัวอย่างการใส่กฎฟัซซี่ลอจิกมีดังนี้



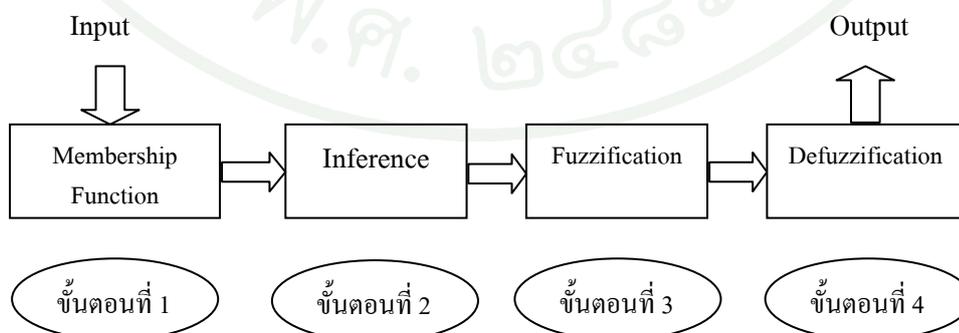
ภาพที่ 9 ตัวอย่างการจำแนกประเภทด้วยฟัซซี่ลอจิก

ที่มา: พยุง (2550)

จากภาพที่ 9 สามารถเขียนเป็นกฎเงื่อนไขได้ดังนี้

- ก) ถ้า x_1 มีค่า low และ x_2 มีค่า low แล้ว ข้อมูล (x_1, x_2) เป็นกลุ่ม C_1
- ข) ถ้า x_1 มีค่า low และ x_2 มีค่า high แล้ว ข้อมูล (x_1, x_2) เป็นกลุ่ม C_2
- ค) ถ้า x_1 มีค่า high และ x_2 มีค่า low แล้ว ข้อมูล (x_1, x_2) เป็นกลุ่ม C_3
- ง) ถ้า x_1 มีค่า high และ x_2 มีค่า high แล้ว ข้อมูล (x_1, x_2) เป็นกลุ่ม C_4

กระบวนการทำงานของฟัซซี่ลอจิกประกอบด้วย 4 ขั้นตอน แสดงดังภาพที่ 10



ภาพที่ 10 กระบวนการทำงานของฟัซซี่ลอจิก

ขั้นตอนที่ 1 เป็นการแปลงข้อมูลนำเข้าเป็นค่าความเป็นสมาชิก โดยจะสร้างฟังก์ชันความเป็นสมาชิก ซึ่งแต่ละข้อมูลไม่จำเป็นต้องมีลักษณะเดียวกัน ขึ้นกับคุณลักษณะของแต่ละข้อมูลนำเข้า โดยมีกราฟฟังก์ชันความเป็นสมาชิกเป็นตัวกำหนดค่า ซึ่งมีค่าได้ตั้งแต่ 0 ถึง 1

ขั้นตอนที่ 2 เป็นการสร้างความสัมพันธ์ระหว่างข้อมูลนำเข้าทั้งหมดที่เกี่ยวข้องกับผลลัพธ์ อาศัยหลักการของการหาเหตุผล โดยเขียนเป็นกฎพีชซึ่งเชื่อมกฎแต่ละกฎด้วยคำสั่ง และ (and) หรือ คำสั่งหรือ (or) ใดๆอย่างหนึ่ง

ขั้นตอนที่ 3 เป็นการนำค่าความเป็นสมาชิกของข้อมูลนำเข้ามาประมวลผลกับผลที่ได้จากขั้นตอนที่ 2 จะได้กราฟฟังก์ชันความเป็นสมาชิกผลลัพธ์ของทุกกฎ

ขั้นตอนที่ 4 เป็นขั้นตอนการจำกัดความคลุมเครือ ทำการคำนวณหาค่าคำตอบ โดยวิธีการหาจุดศูนย์กลาง คือคำนวณหาจุดมวลกลางของพื้นที่ภายใต้กราฟผลลัพธ์ที่ได้จากการรวมกันของแต่ละกฎ มีสูตรดังนี้

$$COG = \frac{\sum_{i=1}^N \alpha_i w_i}{\sum_{i=1}^N \alpha_i} \quad (9)$$

โดยที่ COG แทน จุดมวลกลางของวัตถุ

N แทน ค่าตั้งแต่ตำแหน่งที่ 1 ถึงตำแหน่งที่ i

α_i แทน ค่าความเป็นสมาชิกผลลัพธ์ในเซตพีชซึ่งตำแหน่งที่ i

w_i แทน พื้นที่ใต้กราฟของเซตพีชซึ่งตำแหน่งที่ i

2.5 โครงข่ายประสาทเทียม (Artificial Neural Networks)

ขั้นตอนวิธีการแพร่กระจายย้อนกลับ (Back-propagation algorithm) เป็นขั้นตอนวิธีที่ใช้ในการเรียนรู้ค่าน้ำหนักสำหรับโครงข่ายงานประสาทเทียมวิธีหนึ่งที่นิยมใช้ในเพอร์เซ็ปตรอนหลายชั้น เพื่อปรับค่าน้ำหนักในเส้นเชื่อมต่อระหว่างโหนดให้เหมาะสม โดยการปรับค่านี้อาจขึ้นกับ

ความแตกต่างของค่าผลลัพธ์ที่คำนวณได้กับค่าผลลัพธ์ที่ต้องการ ซึ่งมี 2 ส่วนในการส่งค่าย้อนกลับ คือ การส่งผ่านไปข้างหน้า (Forward pass) การส่งผ่านย้อนกลับ (Backward pass) สำหรับการส่งผ่านไปข้างหน้า ข้อมูลจะผ่านเข้าโครงข่ายงานประสาทเทียมที่ชั้นข้อมูลเข้าเข้า และจะส่งผ่านจากชั้นหนึ่ง ไปยังอีกชั้นหนึ่ง จนกระทั่งถึงชั้นข้อมูลออก ส่วนการส่งผ่านย้อนกลับค่าน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด (Error-Correction) คือ ผลต่างของผลตอบที่แท้จริง (Actual Response) กับผลตอบเป้าหมาย (Target Response) เกิดเป็นสัญญาณผิดพลาด (Error Signal) ซึ่งสัญญาณผิดพลาดนี้จะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงกันข้ามกับการเชื่อมต่อ และค่าน้ำหนักของการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบที่แท้จริงเข้าใกล้ผลตอบเป้าหมาย

การทำงานเพอร์เซ็ปตรอนหลายชั้น คือ ในแต่ละชั้นของชั้นซ่อนตัว (Hidden Layer) จะมีฟังก์ชันสำหรับคำนวณเมื่อได้รับข้อมูลผลลัพธ์จากโหนดในชั้นก่อนหน้านี้ เรียกว่า ฟังก์ชันกระตุ้น (Activation Function) โดยในแต่ละชั้นไม่จำเป็นต้องเป็นฟังก์ชันเดียวกันก็ได้ ชั้นซ่อนตัวมีหน้าที่คือ จะพยายามแปลงข้อมูลที่เข้ามาในชั้นนั้นๆ ให้สามารถแยกแยะความแตกต่างโดยใช้เส้นตรงเส้นเดียว และก่อนที่ข้อมูลจะถูกส่งไปยังชั้นข้อมูลออก (Output Layer) ในบางครั้งอาจจำเป็นต้องใช้ชั้นซ่อนตัวมากกว่า 1 ชั้นในการแปลงข้อมูลให้อยู่ในรูปเส้นตรงเส้นเดียว โดยมีการคำนวณดังต่อไปนี้

ค่าผลลัพธ์ของแต่ละโหนด

$$\begin{aligned} y_i &= f(w_i^1 x_1 + w_i^2 x_2 + w_i^3 x_3 + \dots + w_i^m x_m) \\ &= f\left(\sum_j w_i^j x_j\right) \end{aligned} \quad (10)$$

โดยที่ X_j คือ input จากโหนดอื่นๆ

w_{ij} คือ น้ำหนัก (weight) ของแต่ละแขน (connection)

ขั้นตอนของอัลกอริทึมการแพร่กระจายย้อนกลับมีดังนี้

1. กำหนดค่าอัตราการเรียนรู้ (Learning rate : r)

2. สำหรับแต่ละตัวของข้อมูลนำเข้าให้ทำตามขั้นตอนต่อไปนี้จนกว่าจะได้ประสิทธิภาพที่ต้องการ

1) คำนวณหาค่าผลลัพธ์โดยใช้ค่าน้ำหนักเริ่มต้นจากการสุ่ม

2) คำนวณหาค่า β : แทนประโยชน์ที่จะได้รับสำหรับการเปลี่ยนค่าผลลัพธ์ของแต่ละโหนด

3) ในชั้นข้อมูลออก (Output layer)

$$\beta_z = d_z - o_z \quad (11)$$

โดยที่ d_z คือ ค่าผลลัพธ์ที่ต้องการ

o_z คือ ค่าผลลัพธ์ที่คำนวณได้

4) ในชั้นซ่อนตัว (Hidden layer)

$$\beta_j = \sum_k w_{jk} o_k (1 - o_k) \beta_k \quad (12)$$

โดยที่ w_{jk} คือ น้ำหนักของเส้นเชื่อมระหว่างชั้นที่ j กับ k

o_k คือ ค่าผลลัพธ์ที่คำนวณได้ของชั้น k

5) คำนวณค่าน้ำหนักที่เปลี่ยนแปลงไปสำหรับในทุกน้ำหนัก ด้วยสมการนี้

$$\Delta w_{ij} = r o_i o_j (1 - o_j) \beta_j \quad (13)$$

6) เพิ่มค่าน้ำหนักที่เปลี่ยนแปลงสำหรับข้อมูลนำเข้าทั้งหมด และเปลี่ยนค่าน้ำหนัก

อุปกรณ์และวิธีการ

อุปกรณ์

1. เครื่องคอมพิวเตอร์ตั้งโต๊ะ Intel Core i7, 1.6 GHz, 6 GB of RAM, HD 400 GB
2. โปรแกรม MATLAB version 7.0.1
3. โปรแกรม WEKA version 3.6.2
4. ระบบปฏิบัติการ Microsoft Windows XP Professional

วิธีการ

แผนการดำเนินงานของงานวิจัยนี้ คือ ทำการเลือกกราฟฟังก์ชันความเป็นสมาชิกในพีชชีลอจิก สำหรับการจำแนกประเภทข้อมูลที่มีแอททริบิวต์เป็นตัวเลข ซึ่งอาศัยวิธี Discretize ในการแบ่งช่วงข้อมูลของแต่ละแอททริบิวต์ เพื่อทำการหาค่าความชันของข้อมูลในแต่ละช่วงจากสมการการถดถอยเชิงเส้นหลายตัวแปร ในการทำนายชนิดของรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของขั้นตอนวิธีพีชชีลอจิกที่ได้จากการเรียนรู้ของต้นไม้ตัดสินใจ เพื่อจำแนกประเภทข้อมูล ข้อมูลที่ใช้ในการทดลองประกอบไปด้วย ข้อมูลไวน์ และข้อมูลโรคมะเร็งเต้านม (UCI Machine Learning Repository, 1997) ซึ่งมีรายละเอียดดังต่อไปนี้

- 1) ข้อมูลของผู้ป่วยที่เป็น โรคมะเร็งเต้านม เป็นข้อมูลที่ได้มาจากการเก็บลักษณะของเซลล์ว่ามีลักษณะรูปร่างแบบใด ซึ่งประกอบไปด้วยแอททริบิวต์จำนวน 9 แอททริบิวต์ มีรายละเอียดข้อมูลดังตารางที่ 1

ตารางที่ 1 แอททริบิวต์ต่างๆ ของข้อมูลโรคมะเร็งเต้านม

ลำดับ	แอททริบิวต์
1	Clump Thickness
2	Uniformity of Cell Size

ตารางที่ 1 (ต่อ)

ลำดับ	แอททริบิวต์
3	Uniformity of Cell Shape
4	Marginal Adhesion
5	Single Epithelial Cell Size
6	Bare Nuclei
7	Bland Chromatin
8	Normal Nucleoli
9	Mitoses

ข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านม แบ่งออกเป็น 2 คลาส คือ

ก) คลาสผู้ป่วยปกติ (Benign) ประกอบด้วยผู้ป่วยที่ไม่เป็นโรคมะเร็งเต้านม มีจำนวน 444 ราย

ข) คลาสผู้ป่วยที่เป็นมะเร็งเต้านม (Malignant) ประกอบด้วยผู้ป่วยที่เป็นโรคมะเร็งเต้านม มีจำนวน 239 ราย

2) ข้อมูลไวน์ เป็นข้อมูลการวิเคราะห์ทางเคมีขององุ่นที่นำมาบ่มเพาะเพื่อให้ได้เป็นเครื่องดื่มไวน์ แบ่งออกเป็น 3 คลาส คือ คลาสที่ 1 มีจำนวน 59 ข้อมูล คลาสที่ 2 มีจำนวน 71 ข้อมูล คลาสที่ 3 มีจำนวน 48 ข้อมูล ซึ่งประกอบไปด้วยแอททริบิวต์จำนวน 13 แอททริบิวต์ มีรายละเอียดข้อมูลดังตารางที่ 2

ตารางที่ 2 แอททริบิวต์ต่างๆ ของข้อมูลไวน์

ลำดับ	แอททริบิวต์
1	Alcohol
2	Malic acid
3	Ash
4	Alcalinity of ash

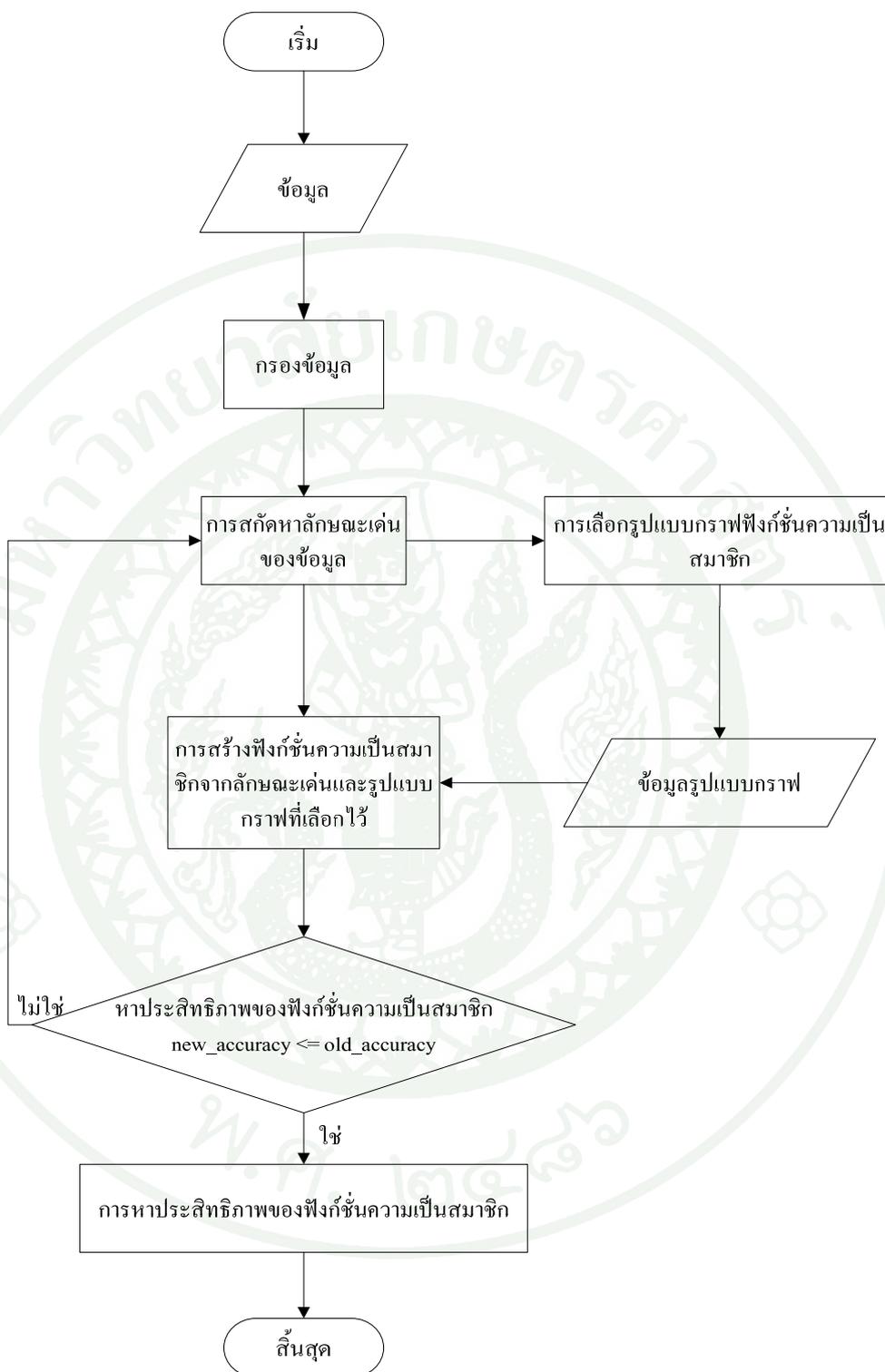
ตารางที่ 2 (ต่อ)

ลำดับ	แอททริบิวต์
5	Magnesium
6	Total Phenols
7	Flavanoids
8	Nonflavanoid phenols
9	Proanthocyanins
10	Color intensity
11	Hue
12	OD280/OD315 of diluted wines
13	Proline

ขั้นตอนการทำงานของงานวิจัยนี้ประกอบไปด้วย 5 ขั้นตอน คือ

- 1) ขั้นตอนการกรองข้อมูล
- 2) ขั้นตอนการสกัดหาลักษณะเด่นของข้อมูล
- 3) ขั้นตอนการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก
- 4) ขั้นตอนการสร้างฟังก์ชันความเป็นสมาชิกของฟuzzyลอจิก
- 5) ขั้นตอนการหาประสิทธิภาพของการจำแนกประเภทข้อมูล

ดั่งภาพที่ 11



ภาพที่ 11 แผนภาพแสดงขั้นตอนการทำงานของงานวิจัยนี้

1. ขั้นตอนการกรองข้อมูล (Pre-Processing)

ขั้นตอนการกรองข้อมูลเป็นขั้นตอนการตรวจสอบข้อมูล เพื่อกรองข้อมูลที่มีค่าของข้อมูล บางส่วนขาดหายไป (Missing Value) จะทำการตัดข้อมูลรายนั้นออกไปหรือกรองข้อมูลผู้ป่วยที่มีสถานะอื่นแทรกซ้อนอยู่ อาจจะมีการกินยาบางอย่างก่อนการตรวจเลือด เพราะยาบางตัวมีผลต่อค่าของผลเลือด ซึ่งทำให้ค่าของการตรวจผลเลือดผิดไปจากความเป็นจริง จึงทำการตัดข้อมูลของผู้ป่วยรายนั้นออกไป

จากการทำงานของพีชคณิตนั้นต้องใช้ข้อมูลที่เป็นตัวเลขเท่านั้น จึงได้ทำการเปลี่ยนชื่อคลาสแต่ละคลาสให้อยู่ในรูปแบบของตัวเลขก่อน โดยตัวเลขที่กำหนดขึ้นในการแทนค่าคลาสจะเป็นตัวเลขที่มีค่าระหว่าง 0 ถึง 1 (Normalization) เช่น ข้อมูลชุดใดมีค่าตอบ 2 คลาส จะกำหนดค่าให้มีค่าเป็น 0.4 และ 0.6 เป็นต้น

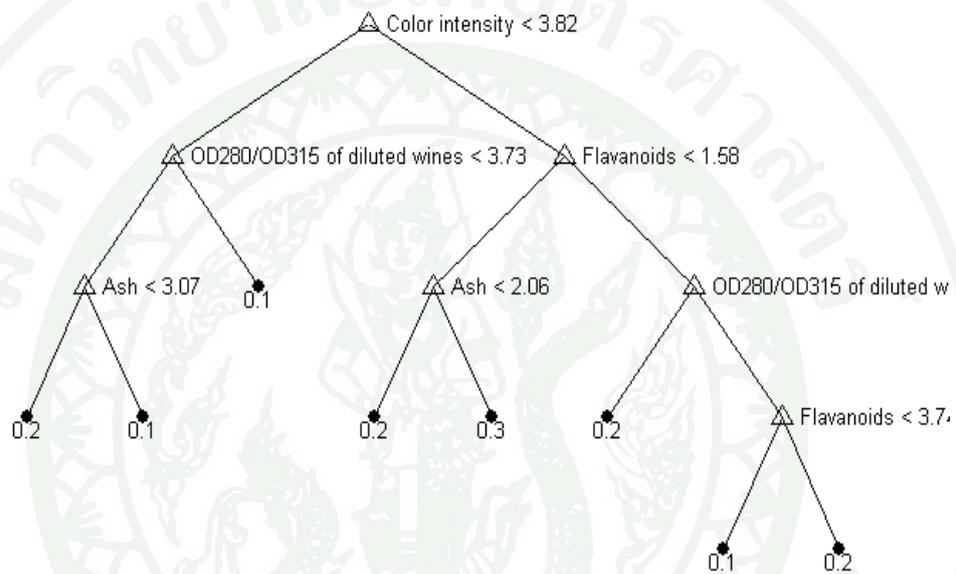
งานวิจัยนี้ทำการทดลองโดยทำการแบ่งข้อมูลออกเป็นส่วนๆ เพื่อใช้ในการทดสอบชุดข้อมูล โดยใช้หลักวิธีการเค-โฟลด์ คลอสวิลดิเคชั่น (K-Fold Cross Validation) คือ แบ่งข้อมูลออกเป็นส่วนๆ ทั้งหมด k ส่วนเท่าๆ กัน โดยจะใช้ $k-1$ ส่วนเป็นข้อมูลที่ใช้ในการสอนระบบ อีก 1 ส่วนใช้เป็นข้อมูลทดสอบ ในงานวิจัยนี้กำหนดให้ $k=10$ ดังนั้นจึงจะต้องทำการทดสอบทั้งหมด 10 รอบ

2. ขั้นตอนการสกัดลักษณะเด่นของข้อมูล (Feature Extraction Algorithm)

ขั้นตอนการสกัดลักษณะเด่นของข้อมูล (ธนรัฐ, 2007) ใช้ต้นไม้ตัดสินใจเป็นเครื่องมือในการสกัดลักษณะเด่นของข้อมูล เพื่อสกัดหาปัจจัยที่มีผลต่อประสิทธิภาพของข้อมูล ข้อมูลนำเข้าคือ ชุดข้อมูลที่ประกอบด้วยแอททริบิวต์ต่างๆ จากนั้นจะตรวจสอบค่าความถูกต้องแม่นยำที่ได้ ว่าเป็นค่าที่ยอมรับได้หรือไม่ ถ้ายังจะมีการตัดแอททริบิวต์ออกทีละตัวแล้วจะทำการสร้างต้นไม้ใหม่ จากนั้นจะตรวจสอบว่าค่าความถูกต้องเพิ่มขึ้นจากเดิมหรือไม่ ถ้าเพิ่มขึ้นจะทำการตัดแอททริบิวต์อื่นต่อ ถ้าลดลงจะทำการใส่แอททริบิวต์ตัวล่าสุดที่เพิ่งตัดออกไปใส่กลับเข้ามาใหม่ แล้วทำการวนตัดแอททริบิวต์ตัวอื่นต่อไป จะทำงานวนอย่างนี้ไปเรื่อยๆ จนกว่าจะได้ค่าความถูกต้องแม่นยำเป็นที่น่าพอใจ และแอททริบิวต์ที่ผ่านการสกัดลักษณะเด่นของข้อมูลจะถูกแสดงอยู่ในรูปของต้นไม้ตัดสินใจ โดยที่แอททริบิวต์ในแต่ละโหนดของต้นไม้ตัดสินใจแสดงถึงแอททริบิวต์ที่

เป็นลักษณะเด่นของข้อมูล และจะถูกนำมาใช้ในการสร้างกราฟฟังก์ชันความเป็นสมาชิก และกฎความสัมพันธ์ของฟuzzyลอจิก ข้อมูลที่ได้จากขั้นตอนนี้จะถูกนำไปใช้ในขั้นตอนการสร้างกราฟฟังก์ชันความเป็นสมาชิกของฟuzzyลอจิกต่อไป

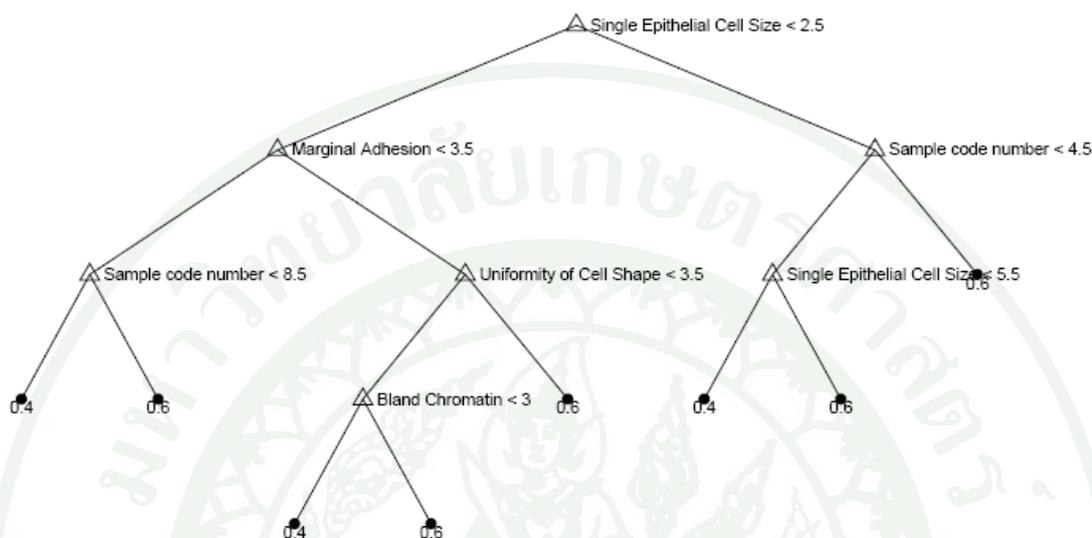
ตัวอย่างแอททริบิวต์ที่ผ่านขั้นตอนการสกัดหาลักษณะเด่นของข้อมูลไวน์ ดังภาพที่ 12



ภาพที่ 12 ต้นไม้ตัดสินใจของข้อมูลไวน์

จากภาพที่ 12 พบว่าแอททริบิวต์ที่ผ่านขั้นตอนการสกัดหาลักษณะเด่นของข้อมูลไวน์ คือ Color intensity, Ash, OD280/OD315 of diluted wines, และ Flavanoids

ตัวอย่างเอทริบิวท์ที่ผ่านขั้นตอนการสกัดหาลักษณะเด่นของข้อมูลลักษณะของเซลล์ของผู้ป่วยที่เป็นโรคมะเร็งเต้านม ดังภาพที่ 13



ภาพที่ 13 ต้นไม้ตัดสินใจของข้อมูลลักษณะของเซลล์ของผู้ป่วยที่เป็นโรคมะเร็งเต้านม

จากภาพที่ 13 พบว่าเอทริบิวท์ที่ผ่านขั้นตอนการสกัดหาลักษณะเด่นของข้อมูลลักษณะของเซลล์ของผู้ป่วยที่เป็นโรคมะเร็งเต้านม คือ Single Epithelial Cell Size, Marginal Adhesion, SampleCode Number, Bland Chromatin และ Uniformity of Cell Shape ตามลำดับ

3. ขั้นตอนการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

ขั้นตอนการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก ประกอบไปด้วย ขั้นตอนการสร้างและเก็บค่าต่างๆ ของโหนด ขั้นตอนการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก มีรายละเอียดดังต่อไปนี้

3.1 ขั้นตอนการสร้างและเก็บค่าโหนดของต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจโดยใช้ข้อมูลที่ผ่านขั้นตอนการสกัดหาลักษณะเด่น และทำการเก็บค่าข้อมูลของแต่ละโหนดของต้นไม้ตัดสินใจลงในโครงสร้างข้อมูล ซึ่งภายในโครงสร้าง

ข้อมูลจะประกอบไปด้วย ลำดับที่ของโหนด ชื่อแอททริบิวต์ของโหนด ค่าแอททริบิวต์ของโหนด โหนดลูก โหนดพ่อ เป็นต้น จากนั้นนำค่าที่ได้จากโครงสร้างข้อมูลของต้นไม้ตัดสินใจนี้ไปหาค่าทางสถิติซึ่งคือค่าเฉลี่ย เพื่อนำไปใช้ในขั้นตอนการทำนายรูปแบบกราฟฟังก์ชันความเป็นสมาชิก โดยมีสูตร (สุทธิชัย, 2536) ดังนี้

ค่าเฉลี่ย (Mean) คือ ค่าที่ได้จากการคำนวณผลบวกของข้อมูลหารด้วยจำนวนของข้อมูลทั้งหมด

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (15)$$

โดยที่ x_i = ค่าของข้อมูล n จำนวนจาก x_1 ถึง x_n
 n = จำนวนข้อมูลทั้งหมด

การหาช่วงข้อมูลของแอททริบิวต์ที่ได้หลังจากขั้นตอนการสกัดหลักคุณลักษณะเด่นจะนำมาทำ Discretize หาช่วงของข้อมูลแต่ละแอททริบิวต์ แล้วนำแต่ละช่วงที่ได้มาหาค่าเฉลี่ย จากนั้นทำการหาค่าความชันโดยใช้วิธีของสมการเชิงเส้นทั้งช่วงบนและช่วงล่างของค่าเฉลี่ยในแต่ละช่วง ข้อมูลที่ได้นับจากค่าเฉลี่ยไป 10% ของข้อมูล โดยสูตรของความชันมีดังนี้

$$\text{Slope} = \frac{y_2 - y_1}{x_2 - x_1} \quad (16)$$

โดยที่ y_2, y_1 = ค่าข้อมูลในแกน y
 x_2, x_1 = ค่าข้อมูลในแกน x

3.2 ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก เป็นการอาศัยการกระจายตัวของแอททริบิวต์โดยเรียงลำดับจากน้อยไปหามาก ทำการหาช่วงของข้อมูลแบบ discretize และนำค่าทางสถิติที่ได้จากขั้นตอนข้างต้นซึ่งคำนวณจากข้อมูลจำนวน 10% จากค่าเฉลี่ยเพื่อให้ได้ค่าความชันมาประมวลผลโดยใช้กฎที่ว่า ถ้าความชันของแอททริบิวต์ใดมีค่าความชันของแต่ละช่วง

เป็น 0 และ 0 จะให้แอททริบิวต์นั้นเป็นกราฟฟังก์ชันความเป็นสมาชิกประฆังคว่ำ เนื่องจากรูปร่างของกราฟประฆังคว่ำจุดสูงสุดจะมีค่าความชันเป็นศูนย์ซึ่งให้เป็นกฎการเลือกรูปแบบของกราฟ แต่ถ้าแอททริบิวต์ใดไม่มีค่าความชันเป็น 0 และ 0 ของแต่ละช่วงข้อมูลแล้วจะให้มันเป็นกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม เพื่อทำนายว่าควรใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิกรูปแบบใดจากรูปสามเหลี่ยม และรูปประฆังคว่ำ มีขั้นตอนดังภาพที่ 14

INPUT: TrainingSet, FeatureSet

For Each Distinct Value in DataSetFromTree

For Each Discretize Attribute

If slope == (0, 0) Then

MF = Bell-shape

Else

MF = Triangular

Next Discretize Attribute

Next DataSetFromTree

ภาพที่ 14 ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

ภาพที่ 14 แสดงขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกโดยใช้ค่าทางสถิติที่ได้จากขั้นตอนข้างต้นมาประมวลผล มีข้อมูลนำเข้าคือชุดข้อมูลทดสอบและชุดของแอททริบิวต์ โดยกำหนดกฎที่ว่า ถ้าค่าความชันของแอททริบิวต์ใดมีค่าความชันของช่วงมีค่าความชันเป็น 0 และ 0 จะให้แอททริบิวต์นั้นเป็นกราฟฟังก์ชันความเป็นสมาชิกประฆังคว่ำ แต่ถ้าแอททริบิวต์ในช่วงใดไม่มีค่าของความชันเป็น 0 และ 0 แล้วจะให้มันเป็นกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม ผลที่ได้คือรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมหรือรูปประฆังคว่ำ

4. ขั้นตอนการสร้างฟังก์ชันความเป็นสมาชิกของพีชชีลอจิก

ขั้นตอนการสร้างฟังก์ชันความเป็นสมาชิกของพีชชีลอจิก เป็นการนำผลที่ได้จากขั้นตอนที่ 2 และขั้นตอนที่ 3 คือขั้นตอนการสกัดหลักขณะเด่นของข้อมูล และขั้นตอนการเลือกรูปแบบ

กราฟฟังก์ชันความเป็นสมาชิก มาเป็นข้อมูลนำเข้าในการหาประสิทธิภาพของรูปแบบของกราฟฟังก์ชันความเป็นสมาชิก ที่ทำการสร้างพีชชีลอจิก โดยรับชุดข้อมูลแอททริบิวท์จากต้นไม้ตัดสินใจ เพื่อให้ได้ประสิทธิภาพที่ได้จากการจำแนกประเภทข้อมูล โดยมีขั้นตอนวิธีการดังภาพต่อไปนี้

Input: DataSet, FeatureSet, $MemFn = \text{Triangular OR Bell-shape}$

Begin

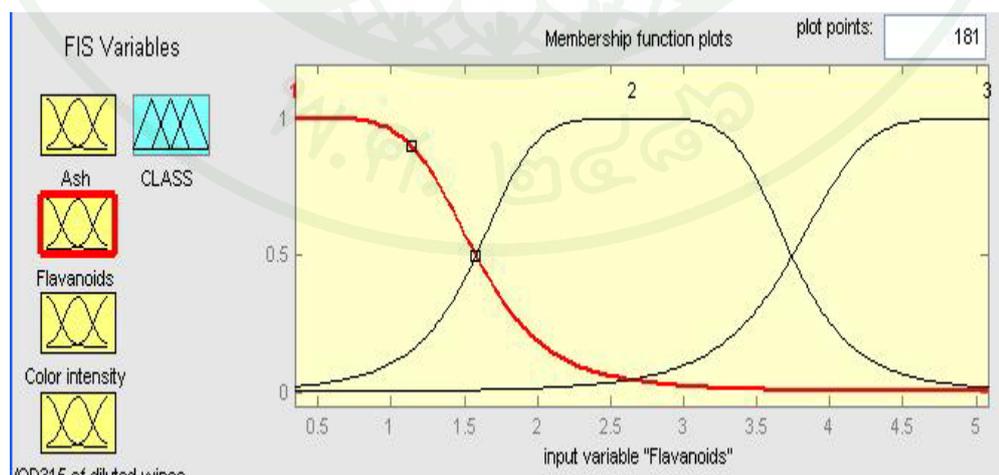
Performance = Classify(AutoGenFn($MemFn$))

End

ภาพที่ 15 ขั้นตอนวิธีการหาประสิทธิภาพของรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

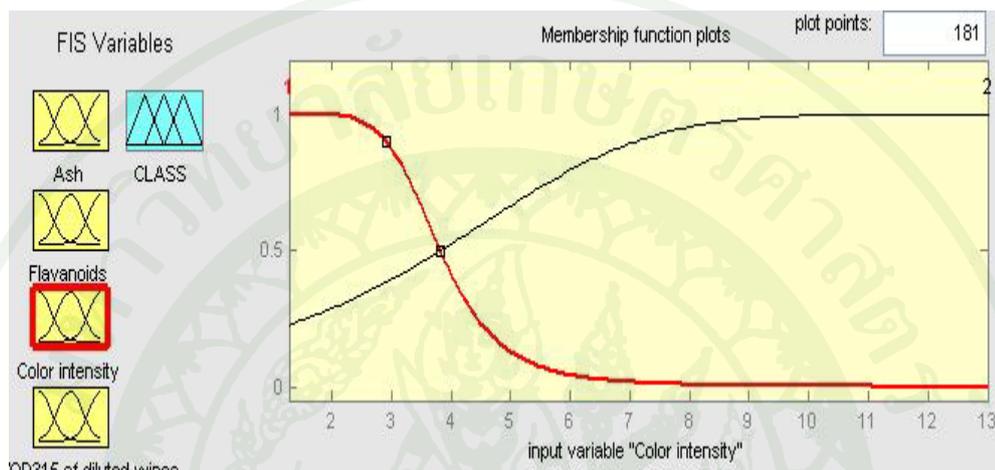
ภาพที่ 15 แสดงขั้นตอนวิธีการหาประสิทธิภาพของรูปแบบกราฟฟังก์ชันความเป็นสมาชิก โดยรับข้อมูลนำเข้าเป็นชุดข้อมูลและชุดข้อมูลของแอททริบิวท์ที่ได้จากขั้นตอนการสกัดคุณลักษณะเด่นของข้อมูล และ $MemFn$ เป็นรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกคือ รูปสามเหลี่ยม และรูปประฆังคว่ำ ซึ่งถูกกำหนดรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกจากขั้นตอนการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก เพื่อทำการหาประสิทธิภาพในการจำแนกประเภทข้อมูลจากการใช้พีชชีลอจิก โดยการสร้างกราฟฟังก์ชันความเป็นสมาชิก $AutoGenFn(MemFn)$

ตัวอย่างของกราฟฟังก์ชันความเป็นสมาชิกที่ได้บางส่วนของข้อมูลในเวนด์ภาพที่ 16-18



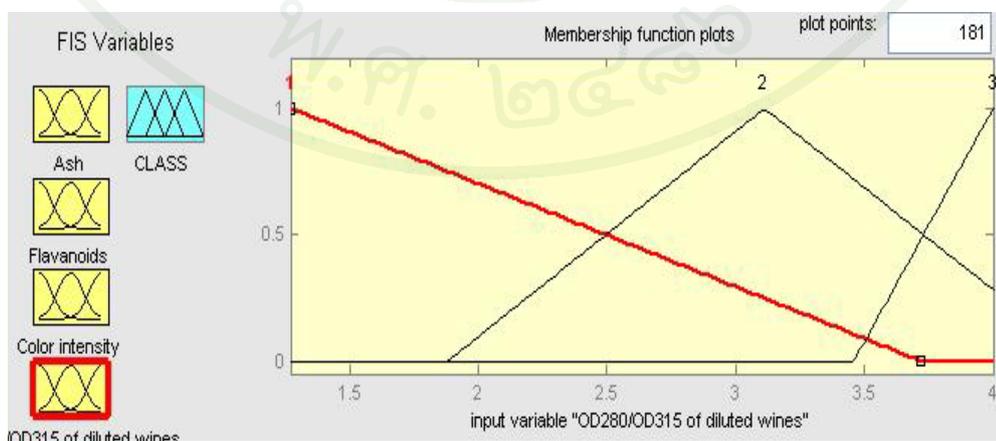
ภาพที่ 16 กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำของแอททริบิวท์ Flavanoids

ภาพที่ 16 เป็นกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำของแอททริบิวต์ Flavanoids ซึ่งเป็นผลมาจากการหาค่าความชันในขั้นตอนวิธีการทำนารูปแบบกราฟฟังก์ชันความเป็นสมาชิกของตารางที่ 6 เนื่องจากแอททริบิวต์ Flavanoids มีค่าความชันในช่วงที่ 3 มีค่าความชันเป็น 0 และ 0 จึงมีกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ



ภาพที่ 17 กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำของแอททริบิวต์ Color intensity

ภาพที่ 17 เป็นกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำของแอททริบิวต์ Color intensity ซึ่งเป็นผลมาจากการหาค่าความชันในขั้นตอนวิธีการทำนารูปแบบกราฟฟังก์ชันความเป็นสมาชิกของตารางที่ 6 เนื่องจากแอททริบิวต์ Color intensity มีค่าความชันในช่วงที่ 3 มีค่าความชันเป็น 0 และ 0 จึงมีกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ



ภาพที่ 18 กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมของแอททริบิวต์ OD280/OD315

ภาพที่ 18 เป็นกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมของแอททริบิวต์ OD280/OD315 of diluted wines ซึ่งเป็นผลมาจากการหาค่าความชันในขั้นตอนวิธีการทำนายรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของตารางที่ 6 เนื่องจากแอททริบิวต์ OD280/OD315 of diluted wines มีค่าความชันในแต่ละช่วงไม่มีค่าความชันเป็นคู่ 0 และ 0 จึงมีกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม

โดยมีขั้นตอนการสร้างฟังก์ชันความเป็นสมาชิกดังภาพที่ 19

ขั้นตอนวิธีการสร้างฟังก์ชันความเป็นสมาชิก

```

AutoGenFn(MF)
Input: DecisionTree, m
Atr = getTreeNode(DecisionTree)
FOR i=1 to TreeSize
    Count = NoOfoccurence(attribute,)
    IF Count > 1 THEN
        FOR j=1 to Count+1
            IF j = 1 THEN
                b = min(Atr, value)
                c = locateIntersect(b, Atr, value[j], m)
                a = b-c
                DrawGraph(a,b,c)
            ELSE IF j = Count+1
                b = max(Atr, value)
                c = locateIntersect(b, Atr, value[j-1], m)
                c = b+a
                DrawGraph(a,b,c)

```

ภาพที่ 19 ขั้นตอนวิธีการสร้างฟังก์ชันความเป็นสมาชิก

ขั้นตอนวิธีการสร้างฟังก์ชันความเป็นสมาชิก (ต่อ)

```

ELSE
    
$$b = \frac{Atr_i.value[j-1] + Atr_i.value[j]}{2}$$

     $a = \text{locateIntersect}(b, Atr_i.value[j], m)$ 
     $c = \text{locateIntersect}(b, Atr_i.value[j], m)$ 
    DrawGraph(a,b,c)
END FOR
ELSE
     $b = \min(Atr_i)$ 
     $c = \text{locateIntersect}(b, \frac{\min(Atr_i.value) + \max(Atr_i.value)}{2}, m)$ 
     $a = b - c$ 
    DrawGraph(a,b,c)
     $b = \max(Atr_i)$ 
     $a = \text{locateIntersect}(b, \frac{\min(Atr_i.value) + \max(Atr_i.value)}{2}, m)$ 
     $c = b + a$ 
    DrawGraph(a,b,c)
END IF
END FOR
RETURN MembershipFunctions

```

ภาพที่ 19 (ต่อ)

ภาพที่ 19 เป็นขั้นตอนการสร้างฟังก์ชันความเป็นสมาชิกของแอททริบิวต์ตัวที่ i โดยรับค่าเป็นโครงสร้างต้นไม้ตัดสินใจ และค่าบอกตำแหน่งจุดตัด ณ ค่าความเป็นสมาชิกเท่ากับ m โดยจากการอ่านค่าต่างๆ ในต้นไม้แล้วเก็บค่าไว้ที่ตัวแปร Art จากนั้นวนทำการสร้างฟังก์ชันความเป็นสมาชิกตามแอททริบิวต์ที่ปรากฏอยู่ในต้นไม้ช่วยตัดสินใจตั้งแต่ตัวแรกไปจนถึงตัวสุดท้าย โดยที่ค่า $\min(Atr_i.value)$ และค่า $\max(Atr_i.value)$ คือ ค่าต่ำสุดและค่าสูงสุดของแอททริบิวต์ตัวที่ i ซึ่งได้จากการเรียนรู้ในข้อมูลสอนระบบ และ $\text{locateIntersect}(p1, p2, m)$ คือ การหาจุดตัดของสมการเส้นตรงตัดกับแกนนอนที่แสดงถึงค่าของแอททริบิวต์ โดยฟังก์ชันนี้จะรับค่า 3 ค่า ในค่าแรก $p1$ ซึ่งบอกถึงตำแหน่ง $(p1, 1)$ ส่วนค่า $p2$ ซึ่งบอกถึงตำแหน่ง $(p2, m)$ โดยที่ค่า m เป็นค่าที่กำหนดขึ้น

เพื่อให้กราฟฟังก์ชันความเป็นสมาชิกของแอททริบิวต์นั้น ตัดกัน ณ ตำแหน่งที่ค่าสมาชิกเท่ากับ m เมื่อได้ค่า a, b, c จากนั้นทำการสร้างฟังก์ชันความเป็นสมาชิก ($\text{DrawGraph}(a, b, c)$) โดยใช้ค่า a, b, c ที่ได้เป็นจุดบอกตำแหน่งของกราฟ

5. ขั้นตอนการหาประสิทธิภาพของการจำแนกประเภทข้อมูล

ขั้นตอนการหาประสิทธิภาพของการจำแนกประเภทข้อมูลเป็นขั้นตอนการหาประสิทธิภาพของข้อมูล โดยใช้ค่า Sensitivity, Specificity และ Accuracy

การวัดประสิทธิภาพมีสูตรดังต่อไปนี้

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{Specificity} = \frac{TN}{TP + TN} \quad (18)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (19)$$

โดยกำหนดให้

True Positives (TP) = จำนวนข้อมูลที่ขั้นตอนวิธีทำนายว่าเป็นคลาสบวกและมีค่าเฉลี่ยเป็นคลาสบวก

True Negatives (TN) = จำนวนข้อมูลที่ขั้นตอนวิธีทำนายว่าเป็นคลาสลบและมีค่าเฉลี่ยเป็นคลาสลบ

False Positives (FP) = จำนวนข้อมูลที่ขั้นตอนวิธีทำนายว่าเป็นคลาสบวกแต่ค่าเฉลี่ยเป็นคลาสลบ

False Negatives (FN) = จำนวนข้อมูลที่ขั้นตอนวิธีทำนายว่าเป็นคลาสลบแต่ค่าเฉลี่ยเป็นคลาสบวก

ผลและวิจารณ์

เนื้อหาในส่วนนี้จะกล่าวถึงผลการวิจัย และวิจารณ์ปัญหาในการทดลองต่างๆ ซึ่งประกอบไปด้วย ประสิทธิภาพในการจำแนกประเภทข้อมูลไวน์ และข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านม โดยใช้ข้อมูลลักษณะเซลล์

ผล

1. ผลการวิจัยการจำแนกประเภทข้อมูลไวน์

ผลการวิจัยของขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก ในส่วนของการหาค่าความชันของข้อมูลไวน์บางส่วนเพื่อทำนายว่าควรใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิก รูปสามเหลี่ยมหรือรูปประฆังคว่ำ ดังตารางที่ 3 โดยช่วงข้อมูลที่ 1-4 แบ่งแบบ discretize และหาค่าความชันโดยใช้ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกซึ่งอยู่ในขั้นตอนการทำงานขั้นที่ 3 โดยคู่วิธีการหาค่าความชันได้จากภาพที่ 14

ตารางที่ 3 ตารางค่าความชันของข้อมูลไวน์

แอททริบิวต์	ช่วงข้อมูล							
	1		2		3		4	
Ash	0	0	-50	5.71				
Flavanoids	-57.14	0	0	-33.33	0	0	0	50
Color intensity	-20	18.42	-3.29	0	0	0		
OD280/OD315	0	-50	33.33	0	11.64	0		

ตารางที่ 3 แสดงให้เห็นว่าแอททริบิวต์ Ash มีค่าความชันของช่วงที่ 1 เท่ากับ 0, 0 และช่วงที่ 3 ของแอททริบิวต์ Flavanoids และ Color intensity มีค่าความชันเป็น 0, 0 ดังนั้นแอททริบิวต์ Ash, Flavanoids และ Color intensity มีกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ และแอททริบิวต์ OD280/OD315 of diluted wines มีกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม

ผลการวิจัยของขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลไวน์ เพื่อทำนายว่าควรใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมหรือรูประฆังคว่ำดังตารางที่ 4

ตารางที่ 4 ผลการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลไวน์

ชุดข้อมูล	Ash	Flavanoids	Color intensity	OD280/OD315	Hue	Malic acid
1	Bell Curve	Bell Curve	Bell Curve	Triangular	-	-
2	-	-	Bell Curve	Triangular	-	-
3	Bell Curve	Bell Curve	Triangular	-	-	-
4	-	-	Bell Curve	Triangular	-	-
5	-	Bell Curve	Bell Curve	-	Triangular	-
6	-	Bell Curve	Bell Curve	-	-	-
7	-	Bell Curve	Bell Curve	-	-	-
8	-	-	Bell Curve	Bell Curve	-	-
9	-	Bell Curve	Bell Curve	Bell Curve	-	Triangular
10	-	Bell Curve	Bell Curve	-	Triangular	Bell Curve

ตารางที่ 4 แสดงผลการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลไวน์ทั้ง 10 ชุดข้อมูล เพื่อทำนายว่าควรใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมหรือรูประฆังคว่ำของแต่ละแอททริบิวต์ในแต่ละชุดข้อมูล ซึ่งเป็นผลมาจากตัวอย่างในตารางที่ 3 ทำการหาค่าความชันของช่วงข้อมูลจากสมการการถดถอยเชิงเส้นหลายตัวแปร โดยที่ - คือแอททริบิวต์ที่ผ่านการสกัดหาลักษณะเด่นออกไปแล้ว จึงไม่มีแอททริบิวต์ดังกล่าวอยู่ในชุดข้อมูล

ผลการวิจัยของขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิก ทำการจำแนกประเภทข้อมูลของข้อมูลไวน์ มีประสิทธิภาพที่ได้จากการทดลองดังตารางที่ 5

ตารางที่ 5 ประสิทธิภาพเฉลี่ยของการจำแนกประเภทของข้อมูลไวน์โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก

ประสิทธิภาพ	ข้อมูลสอนระบบ			ข้อมูลทดสอบ		
	class 1	class 2	class 3	class 1	class 2	class 3
Sensitivity	94.26	95.02	96.05	82.62	89.84	89.07
Specificity	97.75	94.99	99.32	94.59	85.17	99.33
Accuracy	96.63	95.07	98.44	89.93	86.57	96.63

ผลการทดลองในตารางที่ 5 เห็นว่าส่วนของข้อมูลสอนระบบสามารถให้ประสิทธิภาพในการจำแนกประเภทข้อมูลที่ดี แต่ในส่วนของข้อมูลทดสอบไม่ว่าจะเป็นข้อมูลในคลาสใดๆ ให้ประสิทธิภาพที่ต่ำกว่าเล็กน้อย ซึ่งเห็นได้จากค่า Sensitivity, Specificity, Accuracy ของการจำแนกประเภทของข้อมูลไวน์ในคลาสที่ 3 เท่ากับ 89.07%, 99.33% และ 96.63% ตามลำดับ จากผลการวิจัยนี้จึงแสดงให้เห็นว่าการจำแนกประเภทของขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิกของข้อมูลไวน์ให้ประสิทธิภาพเฉลี่ยเท่ากับ 95.01% สามารถดูรายละเอียดผลการทดลองเพิ่มเติมได้ในตารางในภาคผนวกที่ ก1-ก2

ผลการวิจัยโดยการใช้ขั้นตอนวิธีฟuzzyลอจิกจำแนกประเภทข้อมูลของไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกที่ต่างกันดังนี้ ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม ฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ และฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียน ประสิทธิภาพที่ได้จากการทดลองดังตารางที่ 6

ตารางที่ 6 ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลไวน์โดยใช้ขั้นตอนวิธีฟuzzyลอจิกของกราฟฟังก์ชันความเป็นสมาชิกที่รูปร่างต่างกัน

รูปร่าง	ข้อมูลของไวน์		
	คลาสที่ 1	คลาสที่ 2	คลาสที่ 3
Bell Curve	89.38	85.46	93.30
Triangular shaped	86.05	82.07	91.08
Gaussian Curve	87.16	83.24	92.74

ตารางที่ 6 แสดงประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์ที่ได้จากกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำเท่ากับ 93.30% รองลงมาเป็นประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียนเท่ากับ 92.74% และสุดท้ายเป็นประสิทธิภาพที่ได้จากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมเท่ากับ 91.08% ผลการวิจัยนี้จึงแสดงให้เห็นว่า ประสิทธิภาพที่ได้จากฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำให้ประสิทธิภาพดีที่สุด โดยสามารถดูรายละเอียดผลการทดลองเพิ่มเติมได้ในตารางภาคผนวกที่ ก3-ก5

จากการศึกษาการจำแนกประเภทข้อมูลไวน์โดยใช้ขั้นตอนวิธีฟuzzyลอจิกจำแนกประเภทโดยใช้กราฟฟังก์ชันความเป็นสมาชิกที่ต่างกันทั้ง 3 รูปแบบ คือ รูปสามเหลี่ยม รูปประฆังคว่ำ และรูปเกาส์เซียน พบว่าประสิทธิภาพของการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำเพียงอย่างเดียวให้ประสิทธิภาพที่ดีกว่าการใช้รูปแบบสามเหลี่ยมและเกาส์เซียน แต่ยังไม่ให้ประสิทธิภาพดีพอว่าการใช้ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

2. ผลการวิจัยการจำแนกประเภทข้อมูลโรคมะเร็งเต้านม

ผลการวิจัยของขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลผู้ป่วยที่เป็นโรคมะเร็งเต้านม เพื่อทำนายว่าควรใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมหรือรูปประฆังคว่ำดังตารางที่ 7

ตารางที่ 7 ผลการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลโรคมะเร็งเต้านม

แอททริบิวต์	ชุดข้อมูล									
	1	2	3	4	5	6	7	8	9	10
Uniformity of Cell Shape	Tri	Bell	-	-	Tri	Bell	-	-	-	-
Uniformity of Cell Size	-	-	Bell	Bell	Bell	-	-	Bell	Bell	-
Marginal Adhesion	Bell	Bell	-	-	-	Bell	-	-	-	Bell
Single Epithelial Cell Size	Bell	Bell	Bell	Bell	Bell	Bell	Bell	Bell	-	Bell
Bare Nuclei	-	Tri	Tri	Bell	Bell	Bell	Tri	Tri	Bell	Bell
Bland Chromatin	Bell	Bell	Bell	-	-	Bell	Bell	-	Bell	Bell
Normal Nucleoli	-	-	-	Tri	Tri	Tri	Tri	Tri	Tri	-
Mitoses	Tri	Bell	-	-	-	-	-	-	-	-
Clump Thickness	-	-	-	-	-	-	Tri	-	Tri	Bell

ตารางที่ 7 แสดงผลการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกของข้อมูลโรคมะเร็งเต้านมทั้ง 10 ชุดข้อมูล เพื่อทำนายว่าควรใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมหรือรูปประฆังคว่ำของแต่ละแอททริบิวต์ในแต่ละชุดข้อมูล ซึ่งเป็นผลมาจากการหาค่าความชันของช่วงข้อมูลจากสมการการถดถอยเชิงเส้นหลายตัวแปร

ผลการวิจัยของขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิก ทำการจำแนกประเภทข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านม มีประสิทธิภาพที่ได้จากการทดลองดังตารางที่ 8

ตารางที่ 8 ประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านม โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก

ประสิทธิภาพ	ข้อมูลสอนระบบ		ข้อมูลทดสอบ	
	Benign	Malignant	Benign	Malignant
Sensitivity	98.38	98.05	96.41	93.75
Specificity	98.05	98.38	93.76	96.41
Accuracy	98.26	98.26	95.47	95.47

ผลการทดลองในตารางที่ 8 แสดงถึงประสิทธิภาพในการจำแนกประเภทข้อมูลสอนของผู้ป่วยที่ไม่เป็นโรคมะเร็งเต้านมมีประสิทธิภาพเฉลี่ยได้จากค่า Sensitivity, Specificity, Accuracy เท่ากับ 98.38%, 98.05%, 98.26% ตามลำดับ ประสิทธิภาพของการจำแนกข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านมเท่ากับ 98.05%, 98.38%, 98.26% ตามลำดับ ส่วนประสิทธิภาพในการจำแนกประเภทข้อมูลทดสอบของผู้ป่วยที่ไม่เป็นโรคมะเร็งเต้านมเท่ากับ 96.41%, 93.76%, 95.47% ตามลำดับ และประสิทธิภาพของการจำแนกข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านมเท่ากับ 93.75%, 96.41%, 95.47% ตามลำดับ สรุปแล้วประสิทธิภาพโดยรวมที่ได้จากการจำแนกประเภทข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านมสูงถึง 95.47% โดยสามารถดูรายละเอียดผลการทดลองเพิ่มเติมได้ในตารางในภาคผนวกที่ ก7-ก8

ผลการวิจัยโดยการใช้ขั้นตอนวิธีฟuzzyลอจิกจำแนกประเภทข้อมูลของผู้ป่วยที่เป็นโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชันความเป็นสมาชิกที่แตกต่างกันดังนี้ ฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยม ฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำ และฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียน ประสิทธิภาพที่ได้จากการทดลองดังตารางต่อไปนี้

ตารางที่ 9 ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีฟuzzyลอจิกของกราฟฟังก์ชันความเป็นสมาชิกที่รูปร่างแตกต่างกัน

รูปร่าง	ข้อมูลของผู้ป่วยโรคมะเร็งเต้านม	
	Malignant	Benign
Bell Curve	94.91	94.91
Gaussian Curve	93.85	93.85
Triangular shaped	94.03	94.03

ตารางที่ 9 แสดงให้เห็นถึงประสิทธิภาพจากการจำแนกประเภทข้อมูลผู้ป่วยโรคมะเร็งเต้านมที่ได้จากกราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำสูงถึง 94.91% รองลงมาเป็นประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมเท่ากับ 94.03% และสุดท้ายเป็นประสิทธิภาพที่ได้จากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียนเท่ากับ 93.85% ผลการวิจัยนี้จึงแสดงให้เห็นว่าประสิทธิภาพที่ได้จากฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำให้ประสิทธิภาพดีที่สุด โดยสามารถฉายละเอียดผลการทดลองเพิ่มเติมได้ในตารางในภาคผนวกที่ ก9-ก11

จากการศึกษาการจำแนกประเภทข้อมูลของผู้ป่วยโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีฟuzzyลอจิกจำแนกประเภทโดยใช้กราฟฟังก์ชันความเป็นสมาชิกที่แตกต่างกันทั้ง 3 รูปแบบ คือ รูปสามเหลี่ยม รูปประฆังคว่ำ และรูปเกาส์เซียน พบว่าประสิทธิภาพของการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำเพียงอย่างเดียวให้ประสิทธิภาพที่ดีกว่าการใช้รูปสามเหลี่ยมและรูปแบบเกาส์เซียน แต่ยังไม่ให้ประสิทธิภาพดีพอกว่าการใช้ขั้นตอนวิธีการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

3. ผลการวิจัยการจำแนกประเภทผู้ป่วยที่เป็นโรคลำไส้เนื้ออ่อนแรงโดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียม

ผลการวิจัยโดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียมชนิด Multi-Layer Perceptron ซึ่งใช้ค่าอัตราการเรียนรู้เท่ากับ 0.2 และได้ทำการทดลองโดยใช้หลักการเค-โฟลด์ คลอสวาติเคชันในการทดสอบ ซึ่งในงานวิจัยนี้ได้กำหนดให้ $K = 10$ ประสิทธิภาพที่ได้จากการทดลองดังตารางที่ 10

ตารางที่ 10 ประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลของโรคลำไส้เนื้ออ่อนแรงและโรคมะเร็งเต้านม โดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียม

สัปดาห์ลักษณะเด่นของข้อมูล	CFS				Breast Cancer	
	ข้อมูลสอน		ข้อมูลทดสอบ		ข้อมูลสอน	ข้อมูลทดสอบ
	ผลเลือด	อาการป่วย	ผลเลือด	อาการป่วย		
ผ่าน	55.60%	79.16%	42.50%	64.61%	96.52%	95.19%
ไม่ผ่าน	55.57%	78.41%	46.22%	80.33%	95.28%	

ตารางที่ 10 แสดงถึงตารางสรุปประสิทธิภาพจากการจำแนกประเภทข้อมูลของโรคลำไส้เนื้ออ่อนแรงด้วยขั้นตอนวิธีโครงข่ายประสาทเทียมที่ผ่านขั้นตอนการสกัดลักษณะเด่นของข้อมูลและการแบ่งข้อมูลโดยใช้หลักการเคโฟลด์ คลอสวาติเคชันในการทดสอบ ซึ่งในงานวิจัยนี้ได้กำหนดให้ $K = 10$ เพื่อให้ได้ประสิทธิภาพใกล้เคียงกับการใช้ขั้นตอนวิธีพีชชีลोजิก จึงต้องมีการแบ่งข้อมูลให้เหมือนกัน ประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลโรคลำไส้เนื้ออ่อนแรงโดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียมที่ผ่านขั้นตอนการสกัดลักษณะเด่นของข้อมูลมีประสิทธิภาพเฉลี่ยของข้อมูลผลเลือดเท่ากับ 42.5% และข้อมูลอาการป่วยเท่ากับ 64.61% และไม่ผ่านขั้นตอนการสกัดลักษณะเด่นของข้อมูลมีประสิทธิภาพเฉลี่ยของข้อมูลผลเลือดเท่ากับ 46.22% และข้อมูลอาการป่วยเท่ากับ 80.33%

ประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียมที่ผ่านขั้นตอนการสกัดลักษณะเด่นของข้อมูลมีประสิทธิภาพเฉลี่ยเท่ากับ 95.19% และไม่ผ่านขั้นตอนการสกัดลักษณะเด่นของข้อมูลมีประสิทธิภาพเฉลี่ยเท่ากับ 95.28%

วิจารณ์

1. วิจารณ์ประสิทธิภาพจากการจำแนกประเภทข้อมูล

สำหรับการวิจัยนี้ผู้วิจัยเห็นว่ากรจำแนกประเภทข้อมูลโดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมและรูปประฆังคว่ำนั้นให้ผลดีเป็นที่น่าพอใจ เพราะใช้เวลาในการประมวลผลน้อย สำหรับข้อมูลของไวน์นี้เหมาะสำหรับการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู แต่ในงานวิจัยนี้เลือกใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมและรูปประฆังคว่ำนั้น เพราะใช้เวลาในการประมวลผลน้อยและยังคำนวณง่าย ไม่ยุ่งยาก

ตารางที่ 11 ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลของไวน์ที่ได้จากการใช้กราฟฟังก์ชันความเป็นสมาชิก

รูปร่างของกราฟ	ข้อมูลของไวน์		
	คลาสที่ 1	คลาสที่ 2	คลาสที่ 3
Bell Curve	89.38	85.46	93.30
Triangular shaped	86.05	82.07	91.08
Gaussian Curve	87.16	83.24	92.74
Trapezoidal shaped	90.49	86.57	96.08

จากตารางที่ 11 ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์ที่ได้จากกราฟฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมูสูงถึง 96.08% รองมาเป็นประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำเท่ากับ 93.30% รองลงมาเป็นประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียนเท่ากับ 92.74% และสุดท้ายเป็นประสิทธิภาพที่ได้จากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมเท่ากับ 91.08% ผลการวิจัยนี้จึงแสดงให้เห็นว่า ประสิทธิภาพที่ได้จากฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมูให้ประสิทธิภาพดีที่สุด และเวลาที่ใช้ในการประมวลผลรวดเร็วกว่าขั้นตอนวิธีของการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzy-ลอจิก แต่ประสิทธิภาพเฉลี่ยยังดี้อยกว่าการใช้ขั้นตอนวิธีของการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟuzzyลอจิก โดยสามารถดูรายละเอียดผลการทดลองเพิ่มเติมได้ในตารางในภาคผนวกที่ ก3-ก6

ตารางที่ 12 ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลของผู้ป่วยโรคมะเร็งเต้านมที่ได้จากการเปลี่ยนรูปแบบกราฟฟังก์ชันความเป็นสมาชิก

รูปร่างของกราฟ	ข้อมูลของผู้ป่วยโรคมะเร็งเต้านม		
	Malignant	Benign	Time (sec.)
Bell Curve	94.91	94.91	37.06
Gaussian Curve	93.85	93.85	40.70
Triangular shaped	94.03	94.03	34.91
Trapezoidal shaped	95.47	95.47	39.08

จากตารางที่ 12 ประสิทธิภาพเฉลี่ยจากการจำแนกประเภทข้อมูลผู้ป่วยโรคมะเร็งเต้านมที่ได้จากกราฟฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมูสูงถึง 95.47% รองมาเป็นประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำเท่ากับประฆังคว่ำเท่ากับ 94.91% รองลงมา เป็นประสิทธิภาพจากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมเท่ากับ 94.03% และสุดท้ายเป็นประสิทธิภาพที่ได้จากการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียนเท่ากับ 93.85% ผลการวิจัยนี้จึงแสดงให้เห็นว่าประสิทธิภาพที่ได้จากฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมูให้ประสิทธิภาพดีที่สุด และเวลาที่ใช้ในการประมวลผลรวดเร็วกว่าขั้นตอนวิธีของการเลือกกราฟฟังก์ชันความเป็นสมาชิกในพีชชีลลอจิก แต่ประสิทธิภาพเฉลี่ยยังน้อยกว่าการใช้ขั้นตอนวิธีของการเลือกกราฟฟังก์ชันความเป็นสมาชิกในพีชชีลลอจิก โดยสามารถดูรายละเอียดผลการทดลองเพิ่มเติมได้ในตารางในภาคผนวกที่ ก9-ก12

ทั้งนี้เวลาของการประมวลผลจะขึ้นอยู่กับขนาดของข้อมูลที่นำมาทดลองและประสิทธิภาพของเครื่องคอมพิวเตอร์ที่นำมาประมวลผลด้วย เนื่องจากข้อมูลที่นำมาทดลองนี้มีจำนวนแอททริบิวต์มาก จึงทำให้เสียเวลาในการประมวลผลนานพอสมควร

สรุปและข้อเสนอแนะ

สรุป

วัตถุประสงค์ของงานวิจัยนี้คือ นำเสนอถึงวิธีการจำแนกประเภทข้อมูลโดยใช้ขั้นตอนวิธีฟัซซี่ลอจิก แต่ขั้นตอนวิธีของฟัซซี่ลอจิกนี้เป็นกระบวนการที่ต้องอาศัยผู้เชี่ยวชาญที่มีความรู้เกี่ยวกับข้อมูลด้านนั้นๆ ในการสร้างกราฟฟังก์ชันความเป็นสมาชิก เพื่อแก้ปัญหาดังกล่าว ผู้วิจัยจึงได้นำเสนอขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิก ซึ่งไม่จำเป็นต้องอาศัยความรู้จากผู้เชี่ยวชาญในการออกแบบกราฟฟังก์ชันความเป็นสมาชิก โดยใช้ต้นไม้ตัดสินใจช่วยในการเรียนรู้ข้อมูลและสกัดหาลักษณะเด่นของข้อมูล นำข้อมูลที่ได้มาจัดแบ่งช่วงข้อมูลเพื่อหาค่าความชันจากสมการการถดถอยเชิงเส้นหลายตัวแปร เพื่อเป็นเกณฑ์ในการเลือกรูปแบบกราฟฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมหรือรูปประฆังคว่ำของแต่ละแอททริบิวท์ ซึ่งเป็นการผสมรูปแบบของกราฟฟังก์ชันความเป็นสมาชิก เพื่อให้ได้ประสิทธิภาพในการจำแนกประเภทที่ดี และสมเหตุสมผลกับข้อมูลมากที่สุด เพื่อช่วยในการสนับสนุนการตัดสินใจเรื่องต่างๆ ได้สะดวก และรวดเร็วมากยิ่งขึ้น

ในการวิจัยนี้ได้ทำการทดลองกับข้อมูลไวน์และข้อมูลโรคมะเร็งเต้านม ซึ่งจากผลการวิจัยการจำแนกประเภทข้อมูลของไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกแบบผสมนั้นมีประสิทธิภาพการจำแนกที่ดีกว่าการใช้กราฟฟังก์ชันความเป็นสมาชิกแบบประฆังคว่ำเพียงรูปแบบเดียว ซึ่งรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกประฆังคว่ำให้ประสิทธิภาพการจำแนกประเภทได้ดีกว่ารูปสามเหลี่ยมเพียงรูปแบบเดียว

จุดเด่นของงานวิจัยนี้ คือ การสร้างกราฟฟังก์ชันความเป็นสมาชิกในฟัซซี่ลอจิกโดยที่ไม่จำเป็นต้องใช้องค์ความรู้จากผู้เชี่ยวชาญ และการเลือกกราฟฟังก์ชันความเป็นสมาชิกสามารถทำได้โดยอัตโนมัติ ใช้เวลาในการทำงานไม่นานมาก สามารถนำไปปรับใช้กับการจำแนกประเภทข้อมูลที่มีความหลากหลายอื่นๆ ได้ ทั้งยังให้ประสิทธิภาพในการจำแนกประเภทข้อมูลดีเป็นที่น่าพอใจ

จุดด้อยของงานวิจัยนี้ คือ ใช้ได้เฉพาะข้อมูลชนิดตัวเลข และขั้นตอนวิธีที่ผู้วิจัยเสนอยังคงจำกัดใช้ได้ด้วยฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยม และรูปประฆังคว่ำเท่านั้น

ข้อเสนอแนะ

ขั้นตอนวิธีนี้เหมาะกับข้อมูลที่มีความคลุมเครือกันและต้องใช้ข้อมูลที่เป็นตัวเลขเท่านั้น และกราฟฟังก์ชันความเป็นสมาชิกเลือกใช้เฉพาะฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม และรูปประฆังคว่ำเท่านั้น ดังนั้นแนวทางการพัฒนาครั้งต่อไป ควรที่จะสร้างกราฟฟังก์ชันความเป็นสมาชิกให้มีความหลากหลายรูปแบบ รวมไปถึงกำหนดน้ำหนักที่เหมาะสมให้กับแต่ละกฎ เพื่อให้มีประสิทธิภาพที่ได้มีประสิทธิภาพในการทำงานมากยิ่งขึ้น และนำไปทดลองบนชุดข้อมูลที่มีความหลากหลายมากยิ่งขึ้น

เอกสารและสิ่งอ้างอิง

กฤษณะ ไวยมัย. 2549. เอกสารคำสอน วิชา 214554 คลังข้อมูลและการทำเหมืองข้อมูล. ครั้งที่ 1. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ. (เอกสารไม่ตีพิมพ์)

ธนรัฐ รัตนอุบล และนวนวรรณ สุนทรภิชช. 2007. การจำแนกประเภทข้อมูลของโรคกล้ามเนื้ออ่อนแรงโดยใช้ฟัซซี่ลอจิกและวิธีการรวมผลการจำแนก, **Proceedings of Joint Conference on Computer Science and Software Engineering (JCSSE 2007)**, May 2-4, 2007

บุญเสริม กิจศิริกุล. 2546. เอกสารคำสอนวิชา 2110654 ปัญญาประดิษฐ์. จุฬาลงกรณ์มหาวิทยาลัย. (เอกสารไม่ตีพิมพ์)

พยุง มีสังข์. 2548. แนวคิดพื้นฐานเกี่ยวกับฟัซซี่ลอจิก. ระบบฟัซซี่และโครงข่ายประสาทเทียม. แหล่งที่มา: <http://suanpalm3.kmitnb.ac.th/teacher/phayung/powerpoint.asp?pno=1>, 13 มิถุนายน 2550.

พิทักษ์ ศรีแก้ว, ขนิษฐา ชันงาม, จันทิมา พลพินิจ, และอุมารภรณ์ สายแสงจันทร์. 2005. นิวโร-ฟัซซี่กับการจัดกลุ่มข้อมูลโรคตับแบบอัตโนมัติ. **The Joint Conference on Computer Science and Software Engineering**, Nov 17-18, 2005

สุทธิชัย ไ้วศิริ. 2536. หลักสถิติ **Principles of Statistics**. พิมพ์ครั้งที่ 9. กรุงเทพมหานคร: โรงพิมพ์มหาวิทยาลัยรามคำแหง. ภาควิชาสถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง, 79-101.

Lotfi, A. and Tsoi, A. C. 1994. Importance of Membership Functions: A Comparative Study on Different Learning Methods for Fuzzy Inference Systems. **Proceedings of the Third IEEE Conference on Fuzzy Systems** 3: 1791-1796.

CAMDA. **CAMDA 2006 Conference Contest Datasets**. Critical Assessment of Microarray Data Analysis. Available Source: <http://www.camda.duke.edu/camda06/datasets/index.html>, June 13, 2007.

L.A. Zadeh. 1994. Fuzzy Logic, Neural Networks, and Soft Computing. **Communications of the ACM** 37: 77-84.

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://archive.ics.uci.edu/ml/datasets.html>, 13 มิถุนายน 2552.

Quinlan, J. R. 1986. Induction of decision trees. **Machine Learning**, 1(1): 81-106

The MathWorks. 2007. Foundations of Fuzzy Logic. **Fuzzy Logic Toolbox User's Guide**. Available Source: <http://www.mathworks.com/access/helpdesk/help/toolbox/fuzzy/>, May 12, 2007.

Wikanes Thongmee and Boonserm Kijirikul. 2002. Soft-Pruning Decision Trees Using Fuzzification *In Proc. of the 3rd International Conference on Intelligent Technologies* (InTech'2002), Hanoi, Vietnam, December 3-5

William H. Wolberg and O. L. Mangasarian. 1990. Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology, pp. 9193-9196. *In Proceedings of the National Academy of Sciences* 87. ed. U.S.A.

W. Pedrycz. 1994. Why triangular membership functions?, **Fuzzy Sets Syst.** 64: 21-30

Yong-Sheng Yang, Francis H. Y. Chan, F. K. Lam and Hung Nguyen. 1997. A New Fuzzy Classifier with Triangular Membership Functions. *In Proceedings of International Conference on Neural Network*. 97(1): 479-484. U.S.A.



ภาคผนวก



ตารางผนวกที่ ก1 ประสิทธิภาพจากการจำแนกประเภทข้อมูลสอนของข้อมูลไวน์โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก

ชุดข้อมูล										
คลาส 1	1	2	3	4	5	6	7	8	9	10
Sensitivity	84.31	94.23	90.39	98.18	100.00	96.43	100.00	96.23	90.39	92.45
Specificity	99.08	97.22	99.07	92.38	98.15	99.04	98.10	96.26	98.17	100.00
Accuracy	94.38	96.25	96.25	94.38	98.75	98.13	98.75	96.25	95.65	97.52
คลาส 2	1	2	3	4	5	6	7	8	9	10
Sensitivity	98.46	95.24	93.75	84.38	96.77	98.33	96.83	91.05	96.97	98.46
Specificity	89.47	94.85	93.75	95.88	100.00	97.00	100.00	96.77	87.37	94.79
Accuracy	93.13	95.00	93.75	91.88	98.75	97.50	98.75	94.38	91.30	96.27
คลาส 3	1	2	3	4	5	6	7	8	9	10
Sensitivity	95.46	95.56	97.73	95.12	100.00	97.73	100.00	97.50	83.72	97.67
Specificity	100.00	100.00	97.41	98.32	100.00	100.00	100.00	98.33	100.00	99.15
Accuracy	98.75	98.75	97.50	97.50	100.00	99.38	100.00	98.13	95.65	98.76

ตารางผนวกที่ ก2 ประสิทธิภาพจากการจำแนกประเภทข้อมูลทดสอบของข้อมูลไวน์โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก

ชุดข้อมูล										
คลาส 1	1	2	3	4	5	6	7	8	9	10
Sensitivity	75.00	100.00	57.14	75.00	85.71	100.00	100.00	66.67	100.00	66.67
Specificity	90.00	100.00	100.00	100.00	90.91	73.33	100.00	91.67	100.00	100.00
Accuracy	83.33	100.00	83.33	94.44	88.89	77.78	100.00	83.33	100.00	88.24
คลาส 2	1	2	3	4	5	6	7	8	9	10
Sensitivity	83.33	87.50	100.00	100.00	88.89	63.64	100.00	75.00	100.00	100.00
Specificity	83.33	90.00	72.73	81.82	88.89	85.71	90.00	85.71	91.67	81.82
Accuracy	83.33	88.89	83.33	88.89	88.89	72.22	94.44	83.33	94.12	88.24
คลาส 3	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	66.67	100.00	85.71	100.00	75.00	83.33	100.00	80.00	100.00
Specificity	100.00	93.33	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Accuracy	100.00	88.89	100.00	94.44	100.00	94.44	94.44	100.00	94.12	100.00

ตารางผนวกที่ ก3 ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม

ชุดข้อมูล										
คลาส 1	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	100.00	57.14	75.00	85.71	100.00	0.00	66.67	100.00	66.67
Specificity	80.00	100.00	100.00	100.00	90.91	73.33	71.43	91.67	100.00	100.00
Accuracy	88.89	100.00	83.33	94.44	88.89	77.78	55.56	83.33	100.00	88.24
คลาส 2	1	2	3	4	5	6	7	8	9	10
Sensitivity	66.67	87.50	100.00	100.00	77.78	63.64	37.50	75.00	80.00	100.00
Specificity	100.00	90.00	72.73	81.82	88.89	85.71	50.00	85.71	100.00	81.82
Accuracy	88.89	88.89	88.33	88.89	83.33	72.22	44.44	83.33	94.12	88.24
คลาส 3	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	66.67	100.00	85.71	100.00	75.00	16.67	100.00	100.00	100.00
Specificity	100.00	93.33	100.00	100.00	93.75	100.00	58.33	100.00	91.67	100.00
Accuracy	100.00	88.89	100.00	94.44	94.44	94.44	44.44	100.00	94.12	100.00

ตารางผนวกที่ ก4 ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกูประนังคว่ำ

ชุดข้อมูล										
คลาส 1	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	100.00	57.14	75.00	85.71	100.00	100.00	66.67	100.00	66.67
Specificity	90.00	100.00	100.00	100.00	90.91	73.33	100.00	91.67	100.00	100.00
Accuracy	94.44	100.00	83.33	94.44	88.89	77.78	100.00	83.33	100.00	88.24
คลาส 2	1	2	3	4	5	6	7	8	9	10
Sensitivity	83.33	87.50	100.00	100.00	88.89	63.64	100.00	75.00	100.00	100.00
Specificity	100.00	90.00	72.73	81.82	88.89	85.71	90.00	85.71	91.67	81.82
Accuracy	94.44	88.89	83.33	88.89	88.89	72.22	94.44	83.33	94.12	88.24
คลาส 3	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	66.67	100.00	85.71	100.00	75.00	83.33	100.00	80.00	100.00
Specificity	100.00	93.33	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Accuracy	100.00	88.89	100.00	94.44	100.00	94.44	94.44	100.00	94.12	100.00

ตารางผนวกที่ ก5 ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน์โดยใช้กราฟฟังก์ชันความเป็นสมาชิกรูปเกาส์เซียน

ชุดข้อมูล										
คลาส 1	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	100.00	57.14	75.00	75.00	100.00	0.00	66.67	100.00	66.67
Specificity	80.00	100.00	100.00	100.00	100.00	73.33	78.57	91.67	100.00	100.00
Accuracy	88.89	100.00	83.33	94.44	94.44	77.78	61.11	83.33	100.00	88.24
คลาส 2	1	2	3	4	5	6	7	8	9	10
Sensitivity	66.67	87.50	100.00	100.00	100.00	63.64	75.00	75.00	100.00	100.00
Specificity	100.00	90.00	72.73	81.82	81.82	85.71	40.00	85.71	91.67	81.82
Accuracy	88.89	88.89	83.33	88.89	88.89	72.22	55.56	83.33	94.12	88.24
คลาส 3	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	66.67	100.00	85.71	85.71	75.00	16.67	100.00	80.00	100.00
Specificity	100.00	93.33	100.00	100.00	100.00	100.00	83.33	100.00	100.00	100.00
Accuracy	100.00	88.89	100.00	94.44	94.44	94.44	61.11	100.00	94.12	100.00

ตารางผนวกที่ ก6 ประสิทธิภาพจากการจำแนกประเภทข้อมูลไวน้ทั้ง 10 ชุดข้อมูลโดยใช้กราฟ
ฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู

ชุดข้อมูล										
คลาส 1	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	100.00	57.14	75.00	85.71	100.00	100.00	66.67	100.00	66.67
Specificity	80.00	100.00	100.00	100.00	90.91	73.33	100.00	91.67	100.00	100.00
Accuracy	88.89	100.00	83.33	94.44	88.89	77.78	100.00	83.33	100.00	88.24
คลาส 2	1	2	3	4	5	6	7	8	9	10
Sensitivity	66.67	87.50	100.00	100.00	77.78	63.64	100.00	75.00	80.00	100.00
Specificity	100.00	90.00	72.73	81.82	88.89	85.71	90.00	85.71	100.00	81.82
Accuracy	88.89	88.89	83.33	88.89	83.33	72.22	94.44	83.33	94.12	88.24
คลาส 3	1	2	3	4	5	6	7	8	9	10
Sensitivity	100.00	66.67	100.00	85.71	100.00	75.00	83.33	100.00	100.00	100.00
Specificity	100.00	93.33	100.00	100.00	93.75	100.00	100.00	100.00	91.67	100.00
Accuracy	100.00	88.89	100.00	94.44	94.44	94.44	94.44	100.00	94.12	100.00

ตารางผนวกที่ ก7 ประสิทธิภาพจากการจำแนกประเภทข้อมูลสอนของผู้ป่วยที่เป็นโรคมะเร็งเต้านม โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก

ชุดข้อมูล										
คลาส	1	2	3	4	5	6	7	8	9	10
Benign										
Sensitivity	98.50	98.25	97.25	98.50	98.25	99.00	98.75	97.50	98.75	99.00
Specificity	99.07	98.61	98.14	97.21	98.15	96.28	98.14	99.07	97.67	98.14
Accuracy	98.70	98.37	97.56	98.05	98.21	98.05	98.53	98.05	98.37	98.70
Malignant										
Sensitivity	99.07	98.61	98.14	97.21	98.14	96.28	98.14	99.07	97.67	98.14
Specificity	98.50	98.25	97.25	98.50	98.25	99.00	98.75	97.50	98.75	99.00
Accuracy	98.70	98.37	97.56	98.05	98.21	98.05	98.53	98.05	98.37	98.70

ตารางผนวกที่ ๓ ประสิทธิภาพจากการจำแนกประเภทข้อมูลทดสอบของผู้ป่วยที่เป็นโรคมะเร็ง
เต้านม โดยใช้ขั้นตอนวิธีการเลือกกราฟฟังก์ชันความเป็นสมาชิก

ชุดข้อมูล										
คลาส	1	2	3	4	5	6	7	8	9	10
Benign										
Sensitivity	97.78	95.46	95.46	95.56	97.73	97.73	95.56	100.00	97.73	91.11
Specificity	95.93	100.00	95.83	95.83	95.83	79.17	100.00	87.50	91.67	95.83
Accuracy	97.10	97.06	95.59	95.65	97.06	91.18	97.10	95.59	95.59	92.75
Malignant										
Sensitivity	95.83	100.00	95.83	95.83	95.83	79.17	100.00	87.50	91.67	95.83
Specificity	97.78	95.46	95.46	95.56	97.73	97.73	95.56	100.00	97.73	91.11
Accuracy	97.10	97.06	95.59	95.65	97.06	91.18	97.10	95.59	95.59	92.75

ตารางผนวกที่ ๑๑ ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟิงก์ชัน
ความเป็นสมาชิกกลุ่มสามเหลี่ยม

ชุดข้อมูล										
คลาส	1	2	3	4	5	6	7	8	9	10
Malignant										
Sensitivity	95.24	83.33	100.00	96.00	85.71	86.67	88.89	86.67	100.00	95.46
Specificity	88.89	89.74	100.00	100.00	93.02	100.00	90.00	95.24	93.33	97.06
Accuracy	91.23	87.72	100.00	98.25	91.23	96.49	89.47	92.98	96.49	96.43
คลาส	1	2	3	4	5	6	7	8	9	10
Benign										
Sensitivity	88.89	89.74	100.00	100.00	93.02	100.00	90.00	95.24	93.33	97.06
Specificity	95.24	83.33	100.00	96.00	85.71	86.67	88.89	86.67	100.00	95.46
Accuracy	91.23	87.72	100.00	98.25	91.23	96.49	89.47	92.98	96.49	96.43

ตารางผนวกที่ 10 ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชัน
ความเป็นสมาชิกูประนังคว่ำ

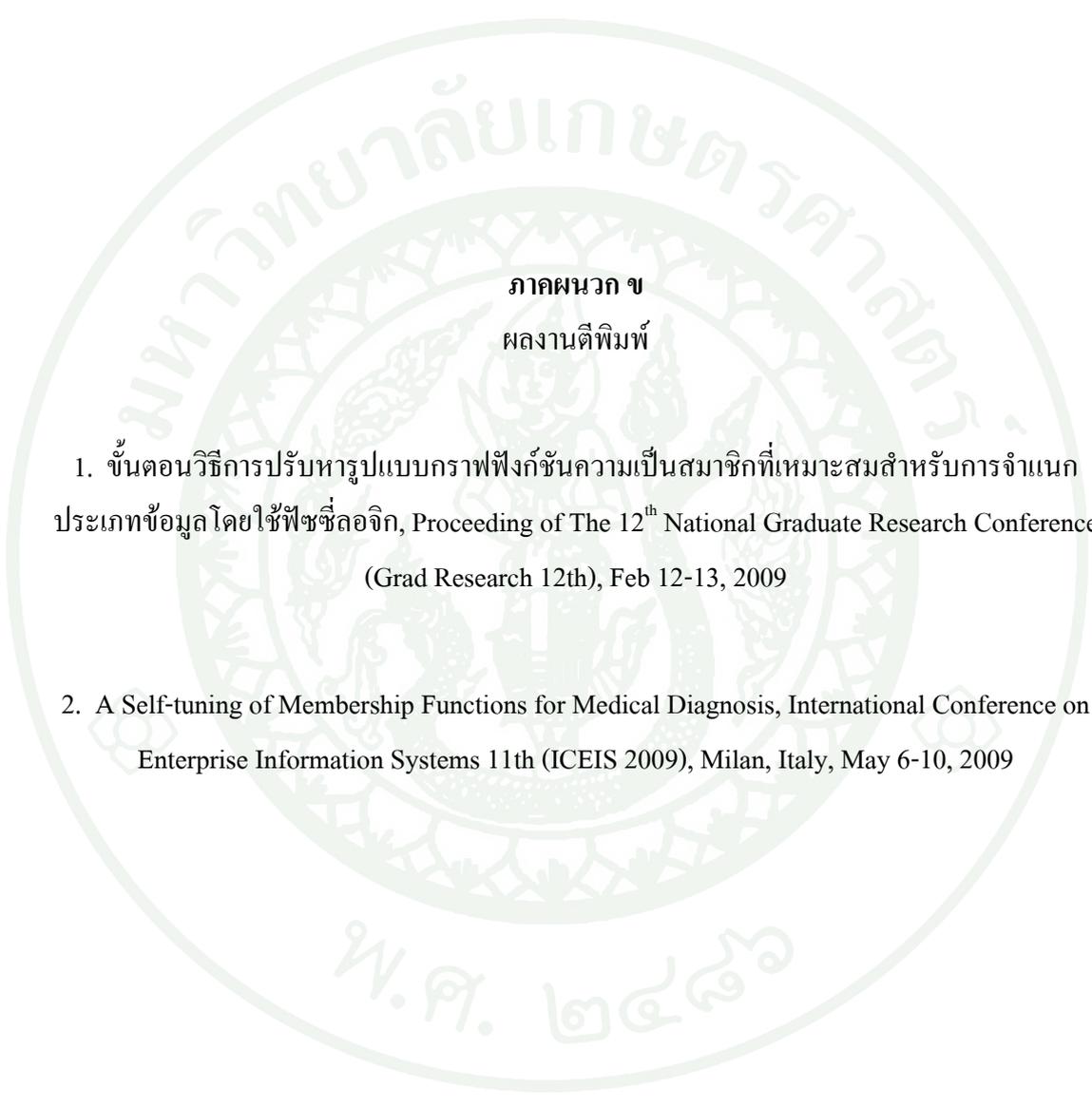
ชุดข้อมูล										
คลาส	1	2	3	4	5	6	7	8	9	10
Malignant										
Sensitivity	95.24	88.89	100.00	96.00	85.71	86.67	88.89	86.67	100.00	95.46
Specificity	88.89	97.44	100.00	100.00	93.02	100.00	96.67	92.86	93.33	97.06
Accuracy	91.23	94.74	100.00	98.25	91.23	96.49	92.98	91.23	96.49	96.43
คลาส	1	2	3	4	5	6	7	8	9	10
Benign										
Sensitivity	88.89	97.44	100.00	100.00	93.02	100.00	96.67	92.86	93.33	97.06
Specificity	95.24	88.89	100.00	96.00	85.71	86.67	88.89	86.67	100.00	95.46
Accuracy	91.23	94.74	100.00	98.25	91.23	96.49	92.98	91.23	96.49	96.43

ตารางผนวกที่ ก11 ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชัน
ความเป็นสมาชิกรูปเกาส์เซียน

ชุดข้อมูล										
คลาส	1	2	3	4	5	6	7	8	9	10
Malignant										
Sensitivity	95.24	77.78	100.00	96.00	85.71	86.67	85.19	93.33	100.00	95.46
Specificity	88.89	89.74	100.00	100.00	93.02	97.62	93.33	97.62	93.33	94.12
Accuracy	91.23	85.97	100.00	98.25	91.23	94.74	89.47	96.49	96.49	94.64
คลาส	1	2	3	4	5	6	7	8	9	10
Benign										
Sensitivity	88.89	89.74	100.00	100.00	93.02	97.62	93.33	97.62	93.33	94.12
Specificity	95.24	77.78	100.00	96.00	85.71	86.67	85.19	93.33	100.00	95.46
Accuracy	91.23	85.97	100.00	98.25	91.23	94.74	89.47	96.49	96.49	94.64

ตารางผนวกที่ ก12 ประสิทธิภาพจากการจำแนกประเภทข้อมูลโรคมะเร็งเต้านมโดยใช้กราฟฟังก์ชัน
ความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู

ชุดข้อมูล										
คลาส	1	2	3	4	5	6	7	8	9	10
Benign										
Sensitivity	97.78	95.46	95.46	95.56	97.73	97.73	95.56	100.00	97.73	88.89
Specificity	95.83	100.00	95.83	95.83	100.00	79.17	100.00	87.50	91.67	95.83
Accuracy	97.10	97.06	95.59	95.65	98.53	91.18	97.10	95.59	95.59	91.30
Malignant										
Sensitivity	95.83	100.00	95.83	95.83	100.00	79.17	100.00	87.50	91.67	95.83
Specificity	97.78	95.46	95.46	95.56	97.73	97.73	95.56	100.00	97.73	88.89
Accuracy	97.10	97.06	95.56	95.65	98.53	91.18	97.10	95.59	95.59	91.30



ภาคผนวก ข

ผลงานตีพิมพ์

1. ขั้นตอนวิธีการปรับหารูปแบบกราฟฟังก์ชันความเป็นสมาชิกที่เหมาะสมสำหรับการจำแนกประเภทข้อมูลโดยใช้ฟuzzyลอจิก, Proceeding of The 12th National Graduate Research Conference (Grad Research 12th), Feb 12-13, 2009
2. A Self-tuning of Membership Functions for Medical Diagnosis, International Conference on Enterprise Information Systems 11th (ICEIS 2009), Milan, Italy, May 6-10, 2009

ขั้นตอนวิธีการปรับหารูปแบบกราฟฟังก์ชันความเป็นสมาชิกที่เหมาะสม

สำหรับการจำแนกประเภทข้อมูลโดยใช้ฟัซซี่ลอจิก

A Self-Tuning Algorithm for Data Classification Using Fuzzy Logic

ภัทรวดี เตียวเตชะเดชา (Phattarawadee Tiewtechadecha)* นวลวรรณ สุนทรภิชัย (Nuanwan Soonthornphisaj)**

บทคัดย่อ

การจำแนกประเภทของข้อมูลที่มีความคลุมเครือกันอย่างเช่น ข้อมูลทางการแพทย์ จำเป็นต้องอาศัยความชำนาญในการตรวจ วิเคราะห์โรค จากแพทย์ผู้เชี่ยวชาญเฉพาะทาง ซึ่งแพทย์แต่ละท่านย่อมมีความชำนาญที่แตกต่างกันขึ้นอยู่กับประสบการณ์ของการวินิจฉัยโรค งานวิจัยนี้จึงได้นำเทคนิคการจำแนกประเภทข้อมูลซึ่งเป็นเทคนิคในการจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดไว้ให้ เพื่อทำนายว่าข้อมูลนั้นควรจัดอยู่ในกลุ่มใด ด้วยการอ้างอิงจากโมเดลที่ได้จากการวิเคราะห์จากกลุ่มข้อมูลสอนระบบ ซึ่งใช้เทคนิคของต้นไม้ตัดสินใจเพื่อสกัดหาลักษณะเด่นของข้อมูล และฟัซซี่ลอจิกที่ใช้กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม รูปแบบเกาส์เซียน และรูประฆังคว่ำ เพื่อทำการจำแนกประเภทผู้ป่วยโรคกล้ามเนื้ออ่อนแรง และข้อมูลของผู้ป่วยโรคมะเร็งเต้านม โดยจากผลการทดลองแสดงให้เห็นว่าสามารถจำแนกประเภทข้อมูลผู้ป่วยโรคกล้ามเนื้ออ่อนแรงให้ประสิทธิภาพเฉลี่ยเท่ากับ 85.11% และประสิทธิภาพเฉลี่ยเท่ากับ 94.91% ในการจำแนกข้อมูลผู้ป่วยโรคมะเร็งเต้านม

ABSTRACT

A classification of ambiguous data like medical data needs to be analyzed by specialists. However, the physicians may have different expertises that affect the medical diagnosis result. This research proposes the classification techniques that can arrange data to a set of predefined classes. We use decision tree to learn from training set in order to extract a feature set, then our approach investigates three kinds of membership functions which are Triangular shape, Bell-shape and Gaussian curve to find the most suitable model for the classification. The data sets used in our experiment are Chronic Fatigue Syndrome (CFS) and Breast Cancer data set. Considering the performance on CFS data set, we found that the performances of our approach is higher than Neural Network (85.11%). For Breast Cancer data set, our algorithm gets 94.91% on correctness.

คำสำคัญ : การจำแนกประเภทข้อมูล ฟัซซี่ลอจิก ฟังก์ชันความเป็นสมาชิก

Keywords : Data Classification, Fuzzy Logic, Membership Function

* มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

** ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

เทคนิคการจำแนกประเภทข้อมูลเป็นเทคนิคหนึ่งที่สำคัญในการจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดไว้ให้ เพื่อแสดงให้เห็นถึงความแตกต่างกันระหว่างกลุ่มของข้อมูล หรือเพื่อทำนายว่าข้อมูลนั้นควรจัดอยู่ในกลุ่มใด ด้วยการอ้างอิงจากโมเดลที่ได้จากการวิเคราะห์จากกลุ่มข้อมูลสอนระบบ ซึ่งใช้เทคนิคของต้นไม้ตัดสินใจ เทคนิคฟิชชีลอจิก และเทคนิคโครงข่ายประสาทเทียม ซึ่งเป็นเทคนิคในการจำแนกประเภทข้อมูล เนื่องจากการจำแนกประเภทของข้อมูลที่มีความคลุมเคลือกันอย่างเช่น ข้อมูลทางการแพทย์ จำเป็นต้องอาศัยความชำนาญในการตรวจวิเคราะห์โรคจากแพทย์ผู้เชี่ยวชาญเฉพาะทางด้านต่างๆ ซึ่งแพทย์แต่ละท่านย่อมมีความชำนาญการณที่แตกต่างกัน ขึ้นอยู่กับประสบการณ์ของการวินิจฉัยโรค

งานวิจัย (ธนรัฐ, 2551) ได้นำเสนอการวิธีการจำแนกประเภทข้อมูลด้วยการใช้เทคนิคฟิชชีลอจิก และเทคนิคฟิชชีลอจิกร่วมกับวิธีการรวมผลจำแนกแต่งงานวิจัยนี้ได้ทำการปรับปรุงวิธีการให้มีความหลากหลายมากยิ่งขึ้น โดยนำเทคนิคต้นไม้ตัดสินใจ เพื่อทำการสกัดหาลักษณะเด่นของข้อมูลมาใช้ และเทคนิคฟิชชีลอจิกทำการจำแนกประเภทข้อมูล ซึ่งใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบสามเหลี่ยม รูปแบบเกาส์เซียน และรูปประฆังคว่ำ และได้ทำการทดลองเปรียบเทียบกับเทคนิคโครงข่ายประสาทเทียม โดยใช้ข้อมูลโรคกล้ามเนื้ออ่อนแรงและโรคกล้ามเนื้ออ่อนแรง แบ่งออกเป็น 3 คลาสดังนี้

- 1) คลาสผู้ป่วยปกติ (Non-fatigued: NF) หมายถึงสภาวะปกติ
- 2) คลาสผู้ป่วยที่มีอาการผิดปกติแต่ไม่ใช่เป็นโรคกล้ามเนื้ออ่อนแรง (Insufficient number of symptoms or fatigue severity: ISF) หมายถึงผู้ป่วยที่มี

อาการของโรคไม่ชัดเจนว่าป่วยเป็นโรคกล้ามเนื้ออ่อนแรงหรือไม่

3) คลาสผู้ป่วยที่เป็นโรคกล้ามเนื้ออ่อนแรง (Chronic Fatigue Syndrome: CFS) หมายถึงผู้ที่เป็โรคกล้ามเนื้ออ่อนแรง

โรคมะเร็งเต้านม แบ่งออกเป็น 2 คลาสดังนี้

1) คลาสผู้ป่วยปกติ (Benign: B) แสดงถึงผู้ป่วยที่ไม่เป็นโรคมะเร็งเต้านม

2) คลาสผู้ป่วยที่เป็นมะเร็งเต้านม (Malignant: M) แสดงถึงผู้ป่วยที่เป็นโรคมะเร็งเต้านม

ทฤษฎีฟิชชีลอจิก (Zadeh, 1994) เป็นการใชเหตุผลแบบประมาณซึ่งแตกต่างจากการใช้เหตุผลแบบเด็ดขาดในลักษณะถูกหรือผิด ใช่หรือไม่ใช่ โดยมีค่าความเป็นสมาชิกของเซตระบุเป็นค่าระดับความเป็นจริงแทนค่าความน่าจะเป็นในแบบของเซตทั่วไป และกฎเกณฑ์ในการบอกร่องเงินไข มักใช้กับข้อมูลที่มีความคลุมเคลือ ยกต่อการตัดสินใจ ซึ่งมีผู้ที่ได้สนับสนุนถึง (Lotfi, 1994) ความสำคัญของฟังก์ชันความเป็นสมาชิกมีผลกระทบต่อระบบการอนุมานฟิชชีมากกว่าจำนวนกฎ หรือโครงสร้างของการอนุมาน โดยส่วนมากมักนิยมใช้ฟังก์ชันความเป็นสมาชิกรูปแบบสามเหลี่ยม (Pedrycz, 1994) เนื่องจากใช้เวลาในการคำนวณรวดเร็ว เป็นโครงสร้างที่ดูง่าย (Yang, 1997) และยังมีการนำเทคนิคฟิชชีไปผสมกับเทคนิคโครงข่ายประสาทเทียมอีกด้วย (พิทักษ์ และคณะ, 2548) เป็นการจัดกลุ่มโรคด้วยนิวโร-ฟิชชี ได้ประสิทธิภาพโดยรวมของระบบเท่ากับ 94% และ (Wikanes, 2002) เสนอวิธีการตัดเล็มอย่างอ่อนสำหรับต้นไม้ตัดสินใจโดยใช้ฟิชชีฟิเคชันได้ประสิทธิภาพดีกว่าวิธีการตัดเล็มอย่างอ่อน โดยอาศัยการให้น้ำหนักด้วยวิธีการแบ็กพรอพาเกชันนิรอลเน็ตเวิร์ก

ทฤษฎีที่เกี่ยวข้อง

1. ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Quinlan, 1986) เป็นการเรียนรู้ที่ใช้แทนความรู้ที่อยู่ในรูปต้นไม้ตัดสินใจ ใช้สำหรับจำแนกประเภทของข้อมูล ซึ่งได้อาศัยค่าความไร้ระเบียบของข้อมูล (Entropy) และค่าเกน (Gain) ในการเลือกแอททริบิวต์ที่มีอำนาจในการจำแนกข้อมูลสูง สูตรในการคำนวณค่าความไร้ระเบียบคือ

$$Entropy(S) = \frac{-P}{P+N} \log_2 \frac{P}{P+N} - \frac{N}{P+N} \log_2 \frac{N}{P+N} \quad (1)$$

โดยที่ S คือ ชุดข้อมูลสอนระบบ S

P คือ จำนวนข้อมูลสอนระบบคลาส

บวกที่พบ (Positive)

N คือ จำนวนข้อมูลสอนระบบคลาสลบ

ที่พบ (Negative)

สูตรในการคำนวณค่าเกนคือ

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

โดยที่ A คือ ข้อมูลแอททริบิวต์ A

$|S_v|$ คือ จำนวนคลาสที่พบในแอททริบิวต์ A

ที่มีข้อมูลเป็น v

$|S|$ คือ จำนวนคลาสทั้งหมดที่พบในชุด

ข้อมูลสอนระบบ S

$Entropy(S)$ คือ ค่าความไร้ระเบียบของชุด

ข้อมูลสอนระบบ S

$Entropy(S_v)$ คือ ค่าความไร้ระเบียบของ

แอททริบิวต์ A ที่มีข้อมูลเป็น v

โดยจะแสดงผลลัพธ์ในรูปของโครงสร้างต้นไม้ และสามารถแปลงเป็นกฎที่เข้าใจได้ง่าย ซึ่งได้ทำการตัดเล็มต้นไม้ (Pruning) เป็นขั้นตอนการตัดโหนดในต้นไม้ตัดสินใจ เพื่อลดความซับซ้อนและการรู้จำจากข้อมูลสอนระบบมากเกินไป ด้วยการตัดเล็มแบบการตัดเล็มหลังจากสร้างต้นไม้เสร็จแล้วเติมรูป (Post-Pruning) ทำให้สามารถจำแนก

ประเภทข้อมูลได้เร็วขึ้น และการเปลี่ยนต้นไม้ให้อยู่ในรูปของกฎ จะใช้คำสั่ง IF-THEN ทำได้ โดยแสดงทุกเส้นทางเริ่มต้นจากโหนดรากไปยังโหนดใบ และทุกครั้งที่พบโหนดทดสอบก็ให้เพิ่มโหนดทดสอบกับค่าของการทดสอบในส่วน

ของ IF และเมื่อพบโหนดใบก็ใส่

2. ฟัซซี่ลอจิก

ฟัซซี่ลอจิก (Zadeh, 1994) เป็นทฤษฎีทางคณิตศาสตร์แขนงหนึ่ง ซึ่งเป็นการใช้เหตุผลแบบประมาณที่แตกต่างจากการใช้เหตุผลแบบเด็ดขาด ในลักษณะถูกหรือผิด ใช่หรือไม่ใช่ ไม่มีความสามารถในการเรียนรู้ แต่ใช้การประมวลผลข้อมูลที่มีความไม่ชัดเจน โดยมีค่าความเป็นสมาชิกของเซตระบุเป็นค่าระดับความเป็นจริง แทนค่าความน่าจะเป็นในแบบของเซตทั่วไป และให้กฎเกณฑ์ในการบอกเงื่อนไข มักใช้กับข้อมูลที่มีความคลุมเครือ ไม่ชัดเจน ยากต่อการตัดสินใจ เช่น ค่าของผลเลือด เซตของอายุคน (พุง, 2550) อาจแบ่งเป็นวัยทารก วัยเด็ก วัยรุ่น วัยกลางคน และวัยชรา เราไม่สามารถระบุได้แน่ชัดว่าวัยเด็กกับวัยรุ่น ช่วงต่ออายุควรจะอยู่ในช่วงใด เป็นต้น เซตของเหตุการณ์ที่ไม่แน่นอนนี้เรียกว่า ฟัซซี่เซต

3. โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมสำหรับงานวิจัยนี้จะใช้แบบเพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron) (Haykin, 1999) เพื่อทำการปรับค่าน้ำหนักในเส้นเชื่อมต่อระหว่างโหนดให้เหมาะสม ซึ่งเพอร์เซ็ปตรอนหลายชั้นจะมีลักษณะของโครงสร้างเป็นชั้นๆ เมื่อมีข้อมูลที่ให้นำเข้ามาเรียนรู้จะถูกใส่เข้าไปในชั้นแรกสุดคือชั้นรับข้อมูล (Input Layer) เมื่อผ่านการคำนวณจากชั้นแรกแล้วผลลัพธ์ที่ได้ในชั้นนี้จะถูกส่งไปยังชั้นกลางต่อไปหรือเรียกอีกอย่างว่า ชั้นซ่อนตัว

(Hidden Layer) โดยที่แต่ละชั้นของชั้นซ่อนตัว จะมีฟังก์ชันสำหรับคำนวณเมื่อได้รับข้อมูลมาจาก โหนดของชั้นก่อนหน้ามาคำนวณ แล้วส่งต่อไปยัง ชั้นถัดไป และเมื่อข้อมูลถูกส่งต่อกันมาจนถึงชั้น สุดท้ายหรือชั้นผลลัพธ์ (Output Layer) ก็ได้ ผลลัพธ์ออกมาจากระบบ ซึ่งการส่งผ่านข้อมูลต่อกันไปเรื่อยๆ แบบนี้เรียกว่า การส่งผ่าน ไปข้างหน้า (Forward pass) หลังจากนั้นจะต้องมีการตรวจสอบ ผลลัพธ์ที่ได้จากระบบว่ามีความคลาดเคลื่อนจาก เป้าหมายมากน้อยอย่างไร เพื่อนำค่าคลาดเคลื่อนนี้ ไปปรับน้ำหนัก (Weight) ใหม่แบบการส่งผ่าน ย้อนกลับ (Backward pass) การปรับน้ำหนักนี้ จะถูกปรับไปเรื่อยๆ จนกระทั่งได้ผลลัพธ์ตามที่ กำหนดหรือค่าคลาดเคลื่อนน้อยพอที่จะยอมรับได้

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาผลของประสิทธิภาพที่ได้จากการ ใช้ฟังก์ชันความเป็นสมาชิกรูปแบบสามเหลี่ยม รูปแบบ เกาส์เซียน และรูปประฆังคว่ำ
2. เพื่อเปรียบเทียบประสิทธิภาพของการใช้ เทคนิคฟuzzyลอจิกกับเทคนิคโครงข่ายประสาทเทียม เนื่องจากเทคนิคโครงข่ายประสาทเทียมใช้กับข้อมูล แทรกซ้อน และข้อมูลคลุมเคลือได้เช่นเดียวกับเทคนิค ฟuzzyลอจิก

อุปกรณ์และวิธีการวิจัย

- เครื่องคอมพิวเตอร์ตั้งโต๊ะ หน่วยประมวลผลความเร็ว 1.73 กิกะเฮิร์ต, หน่วยความจำหลัก 512 เมกกะไบต์, หน่วยความจำสำรอง 60 กิกะไบต์ ความเร็วรอบ 7200 รอบต่อนาที
- โปรแกรม WEKA version 3.5.6
- โปรแกรม MATLAB version 7.0.4

- ระบบปฏิบัติการ Microsoft Windows XP Professional SP2

วิธีการวิจัย

ทำการทดลองการจำแนกประเภทข้อมูลของผู้ป่วยที่เป็น โรคกล้ามเนื้ออ่อนแรง (CAMDA, 2006) และ โรคมะเร็งเต้านม (Merz, 1997) โดยใช้วิธีฟuzzyลอจิก โดยเลือกใช้กราฟฟังก์ชันความเป็นสมาชิก รูปแบบสามเหลี่ยม รูปแบบเกาส์เซียน และรูปประฆังคว่ำให้เหมาะสมกับข้อมูล เพื่อจำแนกประเภทข้อมูลของผู้ป่วยที่เป็น โรคกล้ามเนื้ออ่อนแรง หรือ โรคมะเร็งเต้านมกับคนปกติ



รูปที่ 1 แผนภาพแสดงขั้นตอนการทำงาน

จากรูปที่ 1 เป็นการแสดงขั้นตอนการทำงาน โดยแบ่ง ขั้นตอนการทำงานแบ่งออกเป็น 5 ขั้นตอน ดังนี้

1. ขั้นตอนการกรองข้อมูล

ข้อมูลของผู้ป่วยโรคกล้ามเนื้ออ่อนแรงที่ใช้ในงานวิจัยครั้งนี้แบ่งออกเป็น 2 ประเภทด้วยกัน คือ 1) ข้อมูลผลเลือด (Blood Test Data) ประกอบด้วย ประวัติและผลการวิเคราะห์เลือดของผู้ป่วย 2) ข้อมูลผลประเมินอาการเจ็บป่วย (Symptom Data)

ประกอบด้วยค่าประเมินอาการเจ็บป่วยที่ได้จากการลงความเห็นของแพทย์ และข้อมูลของผู้ป่วยโรคมะเร็งเต้านม ประกอบด้วย ลักษณะของเซลล์ว่ามีลักษณะรูปร่างแบบใด จากนั้นนำข้อมูลทั้งสองโรคมมาผ่านการกรองข้อมูลเพื่อให้เหลือเฉพาะข้อมูลที่สมบูรณ์ โดยจะมีการตัดข้อมูลของผู้ป่วยที่มีค่าข้อมูลบางค่าขาดหายไปทำให้มีไม่ครบถ้วนและข้อมูลของผู้ป่วยที่มีการกินยาบางประเภทก่อนการตรวจผลเลือด เนื่องจากยาบางชนิดอาจจะส่งผลทำให้ค่าของผลเลือดมีความผิดเพี้ยนไปจากความเป็นจริงได้ จำนวนข้อมูลของผู้ป่วยโรคลำไส้เนื้ออ่อนแรงที่ผ่านการกรองแล้วมีจำนวน 187 ราย ประกอบไปด้วยผู้ป่วยที่อยู่ในคลาส NF จำนวน 64 ราย คลาส ISF จำนวน 69 ราย และคลาส CFS จำนวน 54 ราย และโรคมะเร็งเต้านมมีจำนวน 569 ราย ประกอบด้วยผู้ป่วยที่อยู่ในคลาส M จำนวน 212 ราย และคลาส B จำนวน 357 ราย

2. ขั้นตอนการสกัดหาลักษณะเด่นของข้อมูล

ขั้นตอนนี้ได้นำวิธีการของชนรัฐ ซึ่งเป็นการสกัดหาลักษณะเด่นของข้อมูลจะใช้ต้นไม้ช่วยในการตัดสินใจเพื่อสกัดหาปัจจัยที่มีผลต่อการบ่งบอกถึงการเป็นโรคลำไส้เนื้ออ่อนแรงและโรคมะเร็งเต้านม

3. ขั้นตอนการสร้างฟัซซี่ลอจิก

การสร้างฟังก์ชันความเป็นสมาชิกของแอททริบิวต์แต่ละตัว โดยรับค่าเป็นโครงสร้างต้นไม้ตัดสินใจ และอ่านค่าต่างๆ ในต้นไม้แล้วเก็บค่าไว้ในตัวแปร จากนั้นทำการวนสร้างฟังก์ชันความเป็นสมาชิกตามแอททริบิวต์ที่ปรากฏอยู่ในต้นไม้ตัดสินใจตั้งแต่ตัวแรงไปจนถึงตัวตัวสุดท้าย

การสร้างฟัซซี่ลอจิกโดยใช้ต้นไม้ตัดสินใจด้วยวิธีการปรับเปลี่ยนฟังก์ชันความเป็นสมาชิกตามต้นไม้ตัดสินใจ แล้วทำการวัดประสิทธิภาพ จากนั้นทำการตัดเล็มต้นไม้ตัดสินใจทีละลำดับชั้น เพื่อลดระดับความลึกของต้นไม้ ซึ่งการสร้างฟัซซี่ลอจิกจากต้นไม้ตัดสินใจที่มีระดับความลึกมากๆ นั้นจะทำให้

เกิดปัญหาการเรียนรู้จำข้อมูลสอนระบบมากเกินไป ซึ่งไม่ได้ช่วยให้ประสิทธิภาพการจำแนกประเภทดีขึ้นยังทำให้เกิดโอกาสที่จะคำนวณค่าความเป็นสมาชิกผิดพลาดได้ และยังทำให้ประสิทธิภาพของฟัซซี่ลดลงอีกด้วย

4. ขั้นตอนการเลือกกราฟฟังก์ชันความเป็นสมาชิก

จะทำการสร้างกราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมแล้วหาประสิทธิภาพก่อน จากนั้นทำการสร้างกราฟฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียน และกราฟฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำ และหาประสิทธิภาพ เพื่อเลือกใช้รูปแบบกราฟฟังก์ชันความเป็นสมาชิกกับข้อมูลทดสอบ มีขั้นตอนดังตารางที่ 1

ตารางที่ 1 แสดงถึงขั้นตอนการเลือกฟังก์ชันความเป็นสมาชิก

ขั้นตอนวิธีการเลือกใช้กราฟฟังก์ชันความเป็นสมาชิก

```
Algorithm LearnFuzzy
Input: TrainingSet, FeatureSet
Memfn = {triangular, gaussian, bell-shape}
begin
  FOR i=1 TO 3
    Performance[i] = Accuracy(AutoGenFn(Memfn[i]))
  END FOR
  IF Performance[1] = Max(Performance[1],Max(
    Performance[2],Performance[3])) THEN
    MF = Memfn[1]
  ELSEIF Performance[2] = Max(Performance[1],Max(
    Performance[2],Performance[3])) THEN
    MF = Memfn[2]
  ELSE
    MF = Memfn[3]
  END IF
End
```

จากตารางที่ 1 เป็นขั้นตอนการเลือกกราฟฟังก์ชันความเป็นสมาชิก โดยรับข้อมูลนำเข้าเป็นชุดข้อมูล

สอนและชุดของแอททริบิวท์ มี Memfn เป็นชุดของรูปแบบกราฟฟังก์ชันความเป็นสมาชิกสามเหลี่ยม รูปแบบเกาส์เซียน และรูประฆังคว่ำ จากนั้นทำการวนหาประสิทธิภาพที่ได้จากการหาค่าความถูกต้องจากการสร้างฟังก์ชันความเป็นสมาชิก AutoGenFn(Memfn[i]) เป็นการสร้างกราฟฟังก์ชันความเป็นสมาชิกของแอททริบิวท์ตั้งแต่ตัวแรกไปจนถึงตัวสุดท้ายที่ปรากฏในต้นไม้ตัดสินใจ โดยรับค่าเป็นโครงสร้างต้นไม้ที่ได้จากต้นไม้ตัดสินใจ และค่าบอกตำแหน่งจุดตัด ค่าความเป็นสมาชิก โดยทำการคำนวณรูปร่างของกราฟตามสมการของแต่ละชนิดของกราฟฟังก์ชันความเป็นสมาชิก ตั้งแต่กราฟรูปสามเหลี่ยม กราฟรูปแบบเกาส์เซียน และกราฟรูประฆังคว่ำ จากนั้นทำการเปรียบเทียบประสิทธิภาพ (Performance[i]) ที่ได้จากการใช้ฟังก์ชันความเป็นสมาชิกรูปแบบต่างๆ ประสิทธิภาพของการใช้กราฟฟังก์ชันความเป็นสมาชิกรูปแบบใดมีค่าสูงสุด จะนำ MF คือรูปแบบของกราฟฟังก์ชันความเป็นสมาชิกนั้น (Memfn[i]) ไปใช้กับชุดข้อมูลทดสอบต่อไป

5. ขั้นตอนการจำแนกประเภทข้อมูล

ขั้นตอนการประมวลผลการจำแนกประเภท

ข้อมูลจะทำการหาประสิทธิภาพโดยใช้ค่า Sensitivity, Specificity และ Accuracy โดยใช้สูตรดังนี้

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TP + TN} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

โดยที่

True Positives (TP) = จำนวนผู้ป่วยที่จำแนกประเภทถูกต้องว่าเป็นโรค

True Negatives (TN) = จำนวนผู้ป่วยที่จำแนกประเภทถูกต้องว่าไม่ได้เป็นโรค

False Positives (FP) = จำนวนผู้ป่วยที่ไม่เป็นโรคแต่จำแนกประเภทผิดว่าเป็นโรค

False Negatives (FN) = จำนวนผู้ป่วยที่เป็นโรคแต่จำแนกประเภทผิดว่าไม่ได้เป็นโรค

การทดลองวัดประสิทธิภาพนี้ ใช้วิธีการเคโฟลด์คลอสวาเลชัน (K-Fold Cross Validation) คือ การแบ่งข้อมูลออกเป็นส่วนๆ ทั้งหมด K ส่วนเท่าๆกัน โดยจะใช้ K-1 ส่วนเป็นข้อมูลที่ใช้ในการสอนระบบ อีก 1 ส่วนใช้เป็นข้อมูลทดสอบ และทำการวนทดสอบใหม่โดยสลับส่วนข้อมูลทดสอบกับส่วนข้อมูลสอนระบบ เพื่อให้ทุกส่วนนั้นได้มีโอกาสเป็นข้อมูลทดสอบ โดยงานวิจัยนี้กำหนดให้ K =10 ดังนั้นจึงมีการเปลี่ยนข้อมูลทดสอบทั้งหมด 10 รอบ

ผลการวิจัยและการอภิปรายผล

ผลการวิจัยแบ่งออกเป็น 2 ส่วน คือ 1) ผลการจำแนกประเภทผู้ป่วยด้วยพีชชีลอจิกของการใช้ฟังก์ชันความเป็นสมาชิกสามเหลี่ยม รูปแบบเกาส์เซียนและรูประฆังคว่ำ 2) ผลการจำแนกประเภทผู้ป่วยด้วยขั้นตอนวิธีโครงข่ายประสาทเทียม

1) ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยของการจำแนกประเภทผู้ป่วยโรคกล้ามเนื้ออ่อนแรงด้วยพีชชีลอจิกของการใช้ฟังก์ชันความเป็นสมาชิกสามเหลี่ยม รูปแบบเกาส์เซียนและรูประฆังคว่ำ

ตารางที่ 2 ผลการจำแนกประเภทผู้ป่วยโรคกล้ามเนื้ออ่อนแรง

Accuracy	ข้อมูลผลเลือด			ข้อมูลอาการป่วย		
	NF	ISF	CFS	NF	ISF	CFS
ระฆังคว่ำ	78.47	56.50	69.39	92.28	81.28	84.00
เกาส์เซียน	81.44	57.11	68.22	100.00	83.72	83.72
สามเหลี่ยม	77.61	59.28	65.67	100.00	85.11	85.11

จากผลการวิจัยตารางที่ 2 แสดงว่าประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลผู้ป่วยโรคกล้ามเนื้ออ่อนแรงด้วยพีชชีลอจิกของการใช้ฟังก์ชัน

ความเป็นสมาชิกูประนังคว่ำของข้อมูลอาการป่วยเท่ากับ 84% ของการใช้ฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียนของข้อมูลอาการป่วยเท่ากับ 83.72% และของการใช้ฟังก์ชันความเป็นสมาชิกูประนังคว่ำของข้อมูลอาการป่วยเท่ากับ 85.11% พบว่าหากใช้กราฟฟังก์ชันความเป็นสมาชิกูประนังคว่ำในการจำแนกประเภทข้อมูลผู้ป่วยโรคลำไส้เนืออ่อนแรงจะให้ประสิทธิภาพในการจำแนกประเภทที่ดีที่สุดจากทั้งสามรูปแบบของกราฟฟังก์ชันความเป็นสมาชิก

2) ผลการเปรียบเทียบประสิทธิภาพเฉลี่ยของการจำแนกประเภทผู้ป่วยโรคมะเร็งเต้านมด้วยพีชชีลอจิกของการใช้ฟังก์ชันความเป็นสมาชิกูประนังคว่ำ รูปแบบเกาส์เซียนและรูประนังคว่ำ

ตารางที่ 3 ผลการจำแนกประเภทผู้ป่วยโรคมะเร็งเต้านม

Accuracy	ข้อมูลของผู้ป่วยโรคมะเร็ง		
	Malignant	Benign	เฉลี่ย
ระนังคว่ำ	92.35	95.93	94.91
เกาส์เซียน	91.54	94.77	93.85
สามเหลี่ยม	91.80	94.73	94.03

จากผลการวิจัยตารางที่ 3 แสดงว่าประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลผู้ป่วยโรคมะเร็งเต้านมด้วยพีชชีลอจิกของการใช้ฟังก์ชันความเป็นสมาชิกูประนังคว่ำเท่ากับ 94.91% ประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลผู้ป่วยโรคมะเร็งเต้านมด้วยพีชชีลอจิกของการใช้ฟังก์ชันความเป็นสมาชิกูประนังคว่ำเท่ากับ 93.85% และประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลผู้ป่วยโรคมะเร็งเต้านมด้วยพีชชีลอจิกของการใช้ฟังก์ชันความเป็น

สมาชิกูประนังคว่ำเท่ากับ 94.03% พบว่าหากใช้กราฟฟังก์ชันความเป็นสมาชิกูประนังคว่ำในการจำแนกผู้ป่วยโรคมะเร็งเต้านมจะให้ประสิทธิภาพในการจำแนกประเภทที่ดีที่สุดจากทั้งสามรูปแบบของกราฟฟังก์ชันความเป็นสมาชิก

3) ผลการจำแนกประเภทข้อมูลผู้ป่วยของโรคลำไส้เนืออ่อนแรง และโรคมะเร็งเต้านมด้วยขั้นตอนวิธีโครงข่ายประสาทเทียม

ตารางที่ 4 ผลการจำแนกประเภทข้อมูลด้วยขั้นตอนวิธีโครงข่ายประสาทเทียม

Accuracy	CFS	Breast
	ผลเลือด	อาการป่วย
ข้อมูลที่ทำให้การสกัดหาลักษณะเด่น	42.50%	64.61%
ข้อมูลที่ไม่ได้ทำการสกัดหาลักษณะเด่น	46.22%	80.33%
		95.19%
		95.28%

จากผลการวิจัยตารางที่ 4 สรุปได้ว่าประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลของผู้ป่วยโรคลำไส้เนืออ่อนแรงโดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียมซึ่งมีการทำการสกัดหาลักษณะเด่นของข้อมูลผลเลือดเท่ากับ 42.50% ของข้อมูลอาการป่วยเท่ากับ 64.61% และข้อมูลจากผู้ป่วยโรคมะเร็งเต้านมมีประสิทธิภาพเท่ากับ 95.19% ประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลของผู้ป่วยโรคลำไส้เนืออ่อนแรงโดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียมซึ่งไม่มีการทำการสกัดหาลักษณะเด่นของข้อมูลผลเลือดเท่ากับ 46.22% ของข้อมูลอาการป่วยเท่ากับ 80.33% และข้อมูลจากผู้ป่วยโรคมะเร็งเต้านมมีประสิทธิภาพเท่ากับ 95.28%

สรุปผลการวิจัย

การใช้เทคนิคต้นไม้ตัดสินใจร่วมกับเทคนิคฟัซซี่ลอจิก ซึ่งเลือกใช้กราฟฟิงก์ชันความเป็นสมาชิกในรูปสามเหลี่ยม รูปแบบเกาส์เซียน และรูปประฆังคว่ำ ทำให้ได้ประสิทธิภาพที่ดีกว่าใช้กราฟฟิงก์ชันความเป็นสมาชิกเพียงรูปแบบเดียว ดังเห็นได้ชัดจากการทดลองข้อมูลของผู้ป่วยโรคกล้ามเนื้ออ่อนแรง โดยใช้ข้อมูลอาการป่วยเท่ากับ 85.11% ที่ให้ประสิทธิภาพที่สูงกว่าการใช้ขั้นตอนวิธีโครงข่ายประสาทเทียมเท่ากับ 80.33% ประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลของผู้ป่วยโรคกล้ามเนื้ออ่อนแรงด้วยฟัซซี่ลอจิกโดยใช้ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมของข้อมูลอาการป่วยเท่ากับ 85.11% ซึ่งให้ประสิทธิภาพที่สูงที่สุดจากทั้งสามรูปแบบของกราฟฟิงก์ชันความเป็นสมาชิก และประสิทธิภาพเฉลี่ยของการจำแนกประเภทข้อมูลของผู้ป่วยโรคมะเร็งเต้านมด้วยฟัซซี่ลอจิกโดยใช้ฟังก์ชันความเป็นสมาชิกรูปประฆังคว่ำเท่ากับ 94.91% ซึ่งให้ประสิทธิภาพที่สูงที่สุดจากทั้งสามรูปแบบของกราฟฟิงก์ชันความเป็นสมาชิก

ข้อเสนอแนะ

ในอนาคตอันใกล้นี้ผู้วิจัยมีแผนจะทำการทดลองเปรียบเทียบผลการวิจัยนี้กับโรคอื่นๆ หรือชุดข้อมูลอื่น เพื่อศึกษาถึงผลของประสิทธิภาพที่จะได้รับ และอาจทำการปรับปรุงให้มีกราฟฟิงก์ชันความเป็นสมาชิกให้หลายรูปแบบยิ่งขึ้น เพื่อประสิทธิภาพที่ดียิ่งขึ้น

เอกสารอ้างอิง

Thanarat Rattana-ubol. 2007. Classification of Chronic Fatigue Syndrome by Fuzzy Logic and Fusion Method. Proceedings of Joint Conference on Computer Science and Software Engineering (JCSSE 2007). 4: 363-368

พยุง มีสัจ. 2550. แนวคิดพื้นฐานเกี่ยวกับฟัซซี่ลอจิก.

ระบบฟัซซี่และโครงข่ายประสาทเทียม. ค้นเมื่อ 13 มิถุนายน 2550, จาก

<http://suanpalm3.kmitnb.ac.th/teacher/phayung/powerpoint.asp?pno=1>

พิทักษ์ ศรีแก้ว, ขนิษฐา ชันงาม, จันทิมา พลพิณิช, และอุมาภรณ์ สายแสงจันทร์. 2548. นิวโร-ฟัซซี่กับการจัดกลุ่มข้อมูลโรคตับแบบอัตโนมัติ. The Joint Conference on Computer Science and Software Engineering. Nov 17, 2005.

A. Lotfi and A. C. Tsoi. 1994. Importance of Membership Functions: A Comparative Study on Different Learning Methods for Fuzzy Inference Systems. Proceedings of the Third IEEE Conference on Fuzzy Systems. 3: 1791-1796.

CAMDA. 2006. Conference Contest Datasets. Critical Assessment of Microarray Data Analysis. Retrieved June 13, 2007, from <http://www.camda.duke.edu/camda06/datasets/index.html>;

Haykin S. 1999. Neural Network a comprehensive foundation. International Edition. 2: 50-312.

L.A. Zadeh. 1994. Fuzzy Logic, Neural Networks, and Soft Computing. Communications of the ACM 37: 77-84.

Merz, C. J., Murphy, P. M., and Aha, D. W. 1997. UCI Repository of Machine Learning Database. Department of Information and Computer Science. University of California. Irvine. CA. Retrieved June 13, 2007, from [http://archive.ics.uci.edu/ml/datasets/Breast%20Cancer%20Wisconsin%20\(Diagnostic\)](http://archive.ics.uci.edu/ml/datasets/Breast%20Cancer%20Wisconsin%20(Diagnostic))

Quinlan, J. R. 1986. Induction of decision trees Machine Learning. 1(1): 81-106.

Wikanes Thongmee and Boonserm Kijsirikul. 2002. Soft-Pruning Decision Trees Using Fuzzification. In Proc. of the 3rd International Conference on Intelligent Technologies (InTech 2002). Hanoi. Vietnam. December

W. Pedrycz. 1994. Why triangular membership functions? Fuzzy Sets Syst. 64: 21-30.

A Self-tuning of Membership Functions for Medical Diagnosis

Nuanwan Soonthornphisaj , Phattarawadee Tiewtechadecha

Department of Computer Science, Faculty of Science, Kasetsart University,
Bangkok, Thailand
{fscinws@ku.ac.th, iq_phat@hotmail.com}

Abstract. In this paper, a self-tuning of membership functions for fuzzy logic is proposed for medical diagnosis. Our algorithm uses decision tree as a tool to generate three kinds of membership functions which are triangular, bell shape and Gaussian curve. The system can automatically select the best form of membership function for the classification process that can provide the best classification result. The advantage of our system is that it doesn't need the expert to create membership functions for each feature. But the system can create various membership functions using learning algorithm that learns from the training set. In some domains, user can provide prior knowledge that can be used to enhance the performance of the classifier. However, in medical domain, we found that some diseases are difficult to diagnose. It would not be a problem if that disease has been completely explored in medical area. In order to rule out the patient, we need a domain expert to provide the membership functions for many attributes obtained from the laboratory test. Since the disease has not been completely explored in medical area, the membership function provided by the expert might be biased and lead to the poor classification performance. The performance of our proposed algorithm has been investigated on 2 medical data sets. The experimental results show that our approach can effectively enhance the classification performance compare to neural networks and the traditional fuzzy logic.

Keywords: fuzzy logic, Gaussian Curve, Triangular Membership function, Bell Shape membership function, medical diagnosis.

1 Introduction

In medical informatics field, concepts such as symptoms, signs, interpreted results obtained from the laboratory or from clinical investigations, usually identified by a linguistic term and can be formalized by fuzzy sets. In order to directly model these linguistic expressions, fuzzy sets allow us to assign degrees of compatibility of what is observed in the patient to what the term stands for may be applied.

Fuzzy logic is mainly contributed in two areas of medical fields which are medical control systems and expert systems. Fuzzy logic is very well suited for medical control systems because the parameters involved are mostly uncertain. Hence a domain expert must provide the concept of symptoms that are obtained from the laboratory test in the form of ranged values. Many researches apply fuzzy logic as a key technology to deal with uncertainty and found that it provides a promising result. By the way, there are some diseases that haven't had enough knowledge about the symptom, yet. Therefore, using the traditional way of membership functions created by physician seems not to be a good idea. Because these membership functions created by human might be biased and lead to the wrong diagnosis.

Our hypotheses is that for the disease that is difficult to diagnose, we should do some kinds of data mining using machine learning technique to learn from set of attributes and used the knowledge to generate the membership functions automatically. The advantage of our idea is that the learning process can bring new knowledge from data that physicians may never know before and this will contribute to the body of knowledge for domain expert as well.

The rest of this paper is organized as follows: in Section 2, we briefly present the basic concepts of fuzzy logic and its contribution in many research papers. Section 3 provides the background of decision tree learning. In Section 4 we introduce the framework of our system. Section 5 shows all experimental results. We finally conclude the paper with an outlook to our future work in Section 6.

2 Basic Concept of Fuzzy Logic

In this section, we briefly review basic concepts of fuzzy sets which were introduced by Zadeh [17]. Fuzzy logic is the theory of fuzzy sets that calibrate vagueness. It is based on the idea that some attributes are obtained in degrees such as temperature, blood pressure, etc. These attributes are in the form of continuous values. Therefore, each attribute need a function called *membership function* to convert the continuous value to the degree of nominal value (ie. high, low). Fuzzy inference can be defined as a process of mapping from a given input to an output, using the theory of fuzzy sets. The most commonly used fuzzy inference technique is the so-called Mamdani method [8]. The Mamdani-style fuzzy inference process is performed in four steps which are fuzzification of the input variables, rule evaluation, aggregation of the rule outputs and finally the defuzzification process. The process of fuzzy inference consists of 4 steps which are as follows

- Step1: Fuzzify inputs:
- Step2: Apply Fuzzy Operator:
- Step3: Apply Implication Method:
- Step4: Aggregate All Outputs

Fuzzy logic has broadly been used in many medical informatics researches. An internal blood glucose monitoring system was developed by [13]. The system is able to monitor the glucose levels and adjusts the level of insulin using an expert fuzzy logic algorithm. A specific software and hardware design can provide a high level of reliability. Their implementation uses multiple sensors to monitor blood glucose levels. The data will be compared against each other to control the amount of insulin to supply.

The severity of patients suffered from asthma, using a fuzzy decision-making analysis was determined by [7]. The data set is consists of two parts which are the objective severity (OS) (the standard tool for doctor) and the data obtained from questionnaire (PS). Both OS (rated by doctors) and PS (rated by patients) were rated as mild intermittent, mild persistent, moderate, or severe. These variables were pooled, and considered as potential variables patients might use to determine their PS. They were tested against the PS measurement using FDMA.

Four applications of fuzzy logic theory in epidemic problems were presented by [9], using linguistic fuzzy models, possibility measure, probability of fuzzy events and fuzzy decision making techniques. The results demonstrate that the application of fuzzy sets in epidemiology is a very promising area of research. Furthermore, fuzzy logic is used to enhance mammographic features for breast cancer diagnosis [3]. N.H. Phuong and V. Kreinovich build a fuzzy expert system for syndromes differentiation in oriental traditional medicine and combine with the disease diagnosis of Western Medicine [11]. Lesmo.L. et al. combines fuzzy production rules with frame-like structures in order to assess the liver function and to the diagnosis of hepatic disease [6].

Many hybrid systems combine fuzzy logic with neural networks such as the research done by [1]. They implemented a hybrid neuro-fuzzy prognosis system for the prediction of breast cancer relapse patients. The membership functions used in their works are the tumor sizes, grade of tumor and number of axillary nodes. Neuro-fuzzy was also implemented as a multi-sensor fusion system for control of dept of desflurane anesthesia. In this study, depth of desflurane anesthesia was examined through cardiovascular-based an adaptive neuro-fuzzy system according to the changing in the blood pressure and heart rate taken from the patient. The membership functions are heart rate and blood pressure which are provided by the physician [16].

An application that can induce atracurium block during surgery was developed by [10]. They observed improved control over complex numerical techniques. The self-learning fuzzy control technique shows much promise for other medical applications such as post-operative blood pressure management, intra-operative control of anaesthetic depth, and multivariable circulatory management of intensive care patients.

Fuzzy logic was also used in Stuttering Therapy in order to correct the pronunciation of speech. A system called *Orator* can adjust all the therapy parameters automatically and tune them adaptively to the patient [5].

3 Decision Tree Learning

Decision tree learning [12] is a popular algorithm for supervised learning. The tree is constructed using only those best attributes that are able to differentiate the concepts of the target class. Each node in the tree is an attribute selected from the training set using *gain ratio* (see Equation 2). The gain ratio measures the different between the *entropy* (see Equation 1) of the training set before and after selecting the attribute. The attribute which has the highest value of the gain ratio is selected to be a node in the tree.

$$\text{Entropy}(S) = \frac{-P}{P+N} \log_2 \frac{P}{P+N} - \frac{N}{P+N} \log_2 \frac{N}{P+N} \quad (1)$$

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (2)$$

S = training set,
 A = attribute,
 S_v = subset of training data that has value v on attribute A,
 P = training set with Positive class,
 N = training set with Negative class.

4 The Self-tuning Algorithm

We extend our previous work [14] to make our algorithm more flexible in order to get higher performance by using variety styles of membership functions. The classification's performance is optimized by the most suitable form of membership functions. An approach called Self-tuning, which consists of the following steps:

- 1) **The AutoGenFuzzy algorithm generates three kinds of membership function which are triangular, bell shape and Gaussian curve.**
- 2) **The membership tuning process that aims to locate the best degree of membership function belong to the intersection point (m).**
- 3) The fuzzy inference is the final step that classifies the patients into classes.

4.1 Feature Selection

The feature selection algorithm is based on decision tree learning. Since the decision tree's mechanism aims to select the best attribute that has the highest information gain. Therefore, we develop a greedy feature selection algorithm as shown in Table 1.

We found that the selected features obtained from blood test are Intake, ALT, MCH, HCT, BUN, Mononuclearcells, CO2, AST and MCHC. The selected features obtained from symptoms are General Fatigue, Role Physical, Social Function, Bodily Pain, General Health and Vitality. Note that, the value found in all leaf nodes are the class value.

The feature extraction obtained from WBCD data set are Single Epithelial Cell Size, Marginal Adhesion, Sample Code Number, Bland Chromatin and Uniformity of Cell Shape, respectively.

Table 1. Feature Selection Algorithm.

Algorithm: Feature Selection

```

Initialize  $i = 1, tmp = 0$ 
FeatureSet =  $\{a_1, a_2, \dots, a_n\}$ 
 $n = \text{sizeofAttr}(\text{FeatureSet})$ 
While  $i \leq n$ 
     $acc = \text{accuracy}(\text{DecisionTree}(\text{TrainData}, \text{FeatureSet}))$ 
    IF  $tmp \leq acc$  THEN
         $tmp = acc$ 
         $\text{FeatureSet} = \text{FeatureSet} - \{a_i\}$ 
    IF  $tmp \geq \text{accuracy}(\text{DecisionTree}(\text{TrainData}, \text{FeatureSet}))$ 
    THEN
         $\text{FeatureSet} = \text{FeatureSet} + \{a_i\}$ 
     $i++$ 
END WHILE
RETURN  $\text{FeatureSet}, \text{DecisionTree}$ 

```

4.2 Automatic Membership Functions Generation

The AutoGenFuzzy algorithm generates a set of membership functions for each attribute using the decision tree. In this study, three kinds of membership functions are automatically generated which are triangular, bell and Gaussian curve.

Considering any $attribute_i$ founded in the decision tree, the AutoGenFuzzy creates the number of membership functions (graph) of that attribute using Equation 3. That means the number of membership functions for $attribute_i$ depends on the number of occurrences of this attribute appears in the tree plus one.

$$G(attribute_i) = \text{NumberofNode}(attribute_i) + 1 \quad (3)$$

where as $G(attribute_i)$ is the number of graphs

$\text{NumberofNode}(attribute_i)$ is the number of occurrences of $attribute_i$.

Next, the membership function of that attribute is generated using the lowest value of that attribute founded in the decision tree and we assign the degree of membership as 1. For the highest value of this attribute, the AutoGenFuzzy algorithm generates another graph and assigns the degree of membership as 1.

4.2.1 The attribute occurs only once in the decision tree

For any $attribute_i$ occurs only once in the tree, there are 2 graphs for such attribute (see Figure 1 and 2). Suppose that the graph's line is connected in triangular form (see Figure 1)

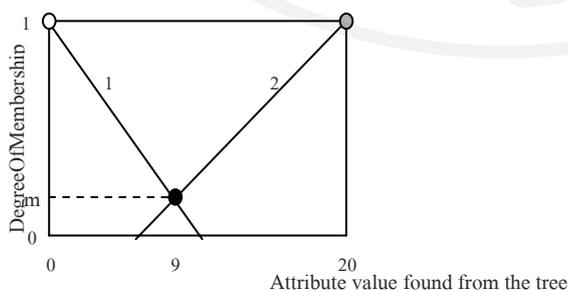


Fig. 1. The membership functions for $attribute_i$ which has one occurrence in the decision tree.

Where \circ = The lowest value of attribute_i
 as \bullet = The highest value of attribute_i
 \bullet = The value of attribute_i found in the decision tree

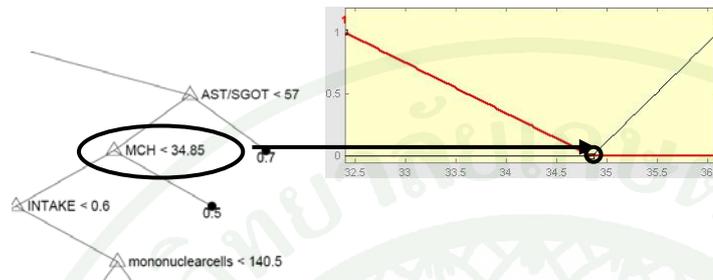


Fig. 2. The membership function of attribute *MCH*.

4.2.1 The attribute occurs twice in the decision tree

In case that the attribute founds in the decision tree twice. The algorithm generates three membership functions (see Figure 3-4). The AutoGenFuzzy algorithm is shown in Table 3. The algorithm scan the decision tree in order to get a set of features and count the number of their occurrences in order to determine the number of membership functions for each attribute.

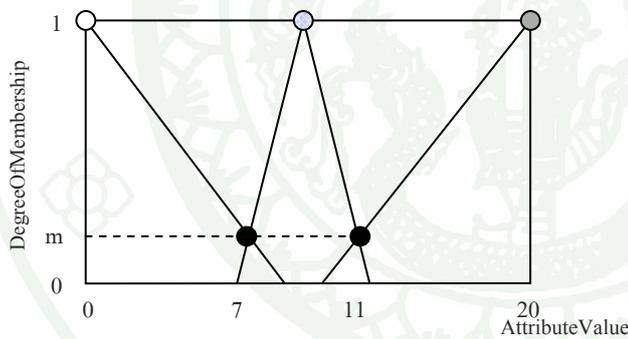


Fig. 3. The membership function of attribute which has 2 occurrences founded in the decision tree

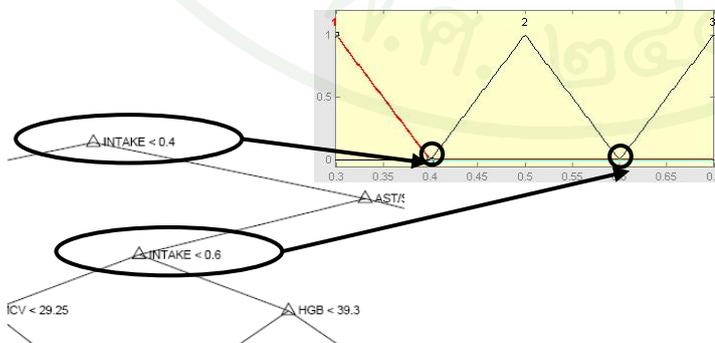


Fig. 4. The membership function of attribute INTAKE which has 2 occurrences founded in the decision tree

Table 2: Self-tuning algorithm

Algorithm: Self-tuning

```

INPUT: TrainingSet, FeatureSet
MemFn = {Triangular, Bell-shape Gaussian }
BEGIN
FOR i=1 TO 3
  Performance[i] = Classify(AutoGenFuzzy(Memfn[i]))
END FOR
IF MAX(Performance) == Performance[1] THEN
  MF = MemFn[1]
IF MAX(Performance) == Performance[2] THEN
  MF = MemFn[2]
IF MAX(Performance) == Performance[3] THEN
  MF = MemFn[3]

END

```

Table 3: Membership function generation algorithm

Algorithm: AutoGenFuzzy

```

Input: DecisionTree
Atr = getTreeNodees(DecisionTree)
m = 0.2
FOR i=1 to TreeSize

  Count = NoOfoccurence(attributei)
  IF Count > 1 THEN
    FOR j=1 to Count+1
      IF j = 1 THEN
        b = min(Atri.value)
        c = locateIntersect(b, Atri.value[j], m)
        a = b - c
        DrawGraph(a, b, c)
      ELSE IF j = Count+1
        b = max(Atri.value)
        c = locateIntersect(b, Atri.value[j-1], m)
        a = b + c
        DrawGraph(a, b, c)
      ELSE
        
$$b = \frac{Atr_i.value[j-1] + Atr_i.value[j]}{2}$$

        a = locateIntersect(b, Atri.value[j], m)
        c = locateIntersect(b, Atri.value[j], m)
        DrawGraph(a, b, c)
    END FOR
  ELSE
    b = min(Atri)
    
$$c = \text{locateIntersect}(b, \frac{\min(Atr_i.value) + \max(Atr_i.value)}{2}, m)$$

    a = b - c
    DrawGraph(a, b, c)
    b = max(Atri)
    
$$a = \text{locateIntersect}(b, \frac{\min(Atr_i.value) + \max(Atr_i.value)}{2}, m)$$


```

```

c = b+a
DrawGraph(a,b,c)
END IF
END FOR
RETURN MembershipFunctions

```

4.3 Membership Function Tuning

The intersection point tuning is an optional step since the AutoGenFuzzy has already set the default of m to be 0.2 which provide the promising results. This step makes fine tuning for membership function. The process aims to locate the intersection point (m) between each graph line of membership function. The range of value m is between 0-0.9. We evaluate the affect of different values of m by doing fuzzy classification.

5 Experiment

5.1 Data sets

Two data sets were used in our experiments which are the Chronic Fatigue Syndrome and the Wisconsin Breast Cancer data set.

5.1.1 The Chronic Fatigue Syndrome Data Set

This data set was obtained from CDC Chronic Fatigue Syndrome Research Group. There are 3 classes of patients as follows:

a) Patients who suffered from Chronic Fatigue Syndrome disease (54 cases).

The class is labeled as CFS.

b) Patients who were ruled out as normal ones (64 cases). The class is labeled as NF.

c) Patients who have insufficient number of symptoms or fatigue severity (69 cases). The class is labeled as ISF.

We separate patients' records into two groups which are the blood test data and the symptom data obtained from patient's self-report. The blood test data contains 34 blood test values. The symptom data obtained from self-report consists of 70 attributes.

5.1.2 The Wisconsin Breast Cancer Data Set (WBCD)

The data set consists of 683 samples that were collected by Dr. W.H. Wolberg at the University of Wisconsin-Madison Hospitals taken from needle aspirates from human breast cancer tissue. The WBCD consists of nine features obtained from fine needle aspirates. 444 of the data set with 683 samples belong to benign, and the remaining 239 samples are of malignant class. The malignant means that the patient gets cancer pathology whereas the benign implies that the patient is healthful.

5.2 Experimental Results

A comprehensive performance study has been conducted to evaluate our method. In order to make the experimental results more reliable, we use 10-fold cross validation method. All experimental results are evaluated using sensitivity, specificity and accuracy, respectively. In this section, we describe those experiments and the results. We run our algorithm on 2 datasets which are CFS and WBCD to test its classification performance against traditional Fuzzy logic and Neural Networks.

5.2.1 Performance of AutoGenFuzzy using Different Feature Set

The objective of this experiment is to evaluate the potential of our feature selection algorithm. In order to test the feature selection algorithm, we set up the experiment as follows

- 1) We used 2 data sets which are CFS and WBCD data set.
- 2) We run two different learning algorithms which are AutoGenFuzzy and Backpropation Neural Networks.

From Table 4, we can see that feature selection provides satisfactory results. Considering the CFS data set, we found that feature selection can enhance the performance of both Neural Networks and Self-tuning. However, Neural Network gets less performance compared to Self-tuning on both data sets.

Table 4: Performance obtained from AutoGenFuzzy on CFS and WBCD data set

Data Set	Algorithm	FeatureSelection	Accuracy
CFS	NeuralNetworks	yes	67.50
		no	46.22
	AutoGenFuzzy	yes	87.40
		no	83.37
WBCD	NeuralNetworks	yes	95.19
		no	95.19
	AutoGenFuzzy	yes	97.10
		no	95.47

5.2.2 Performance of Self-tuning using Different Forms of Membership Function

We compare the performance of Self-tuning algorithm using different forms of membership functions which are triangular, bell shape and Gaussian curve. Classification performance on CFS and WBCD data set are shown in Table 5 and 6.

Table 5: Performance obtained from Self-tuning on CFS data set

Membership functions		Class Label		
		NF	ISF	CFS
Triangular	Sensitivity	100.00	78.57	70.33
	Specificity	100.00	86.43	88.79
	Accuracy	100.00	83.50	83.50
Bell shape		NF	ISF	CFS
	Sensitivity	93.57	67.14	75.67
	Specificity	96.73	85.39	85.77
GaussianCurve	Accuracy	95.67	79.06	82.89
		NF	ISF	CFS
	Sensitivity	100.00	70.00	72.67
	Specificity	97.57	87.48	86.65
	Accuracy	98.39	81.00	82.61

Table 6: Performance obtained from Self-tuning on WBCD data set

Membership functions		Class Label	
		Malignant	Benign
Triangular	Sensitivity	91.80	94.73
	Specificity	94.73	91.80
	Accuracy	94.03	94.03
Bell shape		Malignant	Benign
	Sensitivity	92.34	95.93
	Specificity	95.93	92.35
GaussianCurve		Malignant	Benign
	Sensitivity	91.54	94.77
	Specificity	94.77	91.54
	Accuracy	93.85	93.85

Considering the results shown in Table 6, we found that Self-tuning using triangular membership function can effectively rule out the patients who are not suffered from the disease (NF). Self-tuning gets 100% of sensitivity, specificity and accuracy.

For the CFS patients, we found that AutoGenFuzzy using bell shape membership function gets 75.67% on sensitivity which is more than that of the traditional fuzzy which gets only 59.67%. Considering the specificity performance, Self-tuning using triangular gets 88.79% which is lower than that of traditional fuzzy logic which gets 100%. However, this performance is acceptable because in medical domain the specificity means the ability to rule out the CFS patient which is less important than the ability to detect the real CFS patient measured by the sensitivity. Therefore, the overall performance of Self-tuning is more satisfiable than the traditional one since the Self-tuning is more sensitive than the traditional fuzzy logic.

Considering the sensitivity in Table 6, we found that the Self-tuning using Bell shape membership function outperforms other kinds of membership function (sensitivity = 92.34%). The ability of Self-tuning to rule out the malignant patient is 95.93%

6 Conclusion

In this study, we propose an approach that can automatically generate membership functions without domain expert intervention. The advantage of our system is that it can enhance the classification performance because of the contribution of the decision tree learning. The learning algorithm creates the concept of CFS data set by selecting informative feature sets and partitions each feature value that maximize the information gain. Therefore, we build an automatic system that gets the learned decision tree as input and generate the set of membership functions as output. We test the potential of these membership functions and compare to the traditional fuzzy logic using the well-known membership function obtained from domain expert. We compared our classification performance in terms of sensitivity, specificity and accuracy to the traditional fuzzy logic algorithm. We found that Self-tuning algorithm is a high potential algorithm since it outperforms the traditional fuzzy logic and Neural Networks. We plan to investigate the self-tuning algorithm on other data sets in the near future.

References

1. Aragonés, J.M.J., Sánchez, J.I.P., Doña, J.M. and Alba, E.: A Neuro-fuzzy decision model for prognosis of breast cancer Relapse. LNCS, vol. 3040, pp. 638-648. Springer, Heidelberg (2004)
2. CFS Data set, <http://www.cdc.gov/ncidod/diseases/cfs/>
3. Cheng H.D. and Xu, H.: A novel fuzzy logic approach to mammogram contrast enhancement. Information Sciences, vol. 148, issues 1-4, pp. 167-184 (2002)
4. Cox, E.: The Fuzzy systems Handbook: A practitioner's Guide to building, using and Maintaining Fuzzy Systems. Academic Press, Cambridge. (1994)
5. Kwasnicka, H. and Zak, B.: Fuzzy logic in stuttering therapy. LNCS, vol. 4029, pp. 925-930. Springer, Heidelberg (2006)
6. Lesmo, L., Saitta, L. and Torasso, P.: Dealing with uncertain knowledge in medical decision-making: A case study in hepatology. Artificial Intelligence in Medicine, vol.1, issue 3, pp.105-116 (1989)
7. Lurie, A., Marsala, C., Hartley, S., Bouchon-Meunier, B. and Dusser, D.: Patients perception of asthma severity, Respiratory Medicine, Available online 23 July 2007.
8. Mamdani, E.H. and Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. of Man-Machine Studies, vol.7 issue 1, pp.1-13 (1975)
9. Massad, E., Regina, N., Ortega, S., Struchiner, C.J. and Burattini, M.N.: Fuzzy epidemics. Artificial Intelligence in Medicine, vol. 29, issue 3, pp. 241-259 (2003)
10. Mason, D.G., Linkens, D.A., Edwards, N.D.: Self-learning fuzzy logic control in medicine. LNCS, vol. 1211, pp. 300-304. Springer, Heidelberg (1997)
11. Phuong N.H. and Kreinovich, V.: Fuzzy logic and its applications in medicine. International Journal of Medical Informatics, vol. 62, issues 2-3, pp.165-173(2001)
12. Quinlan, J.R.: Introduction of Decision Trees. Machine Learning 1: 81-106. (1986)
13. Shereck D. and Jabur, F.: VHDL implementation of a fuzzy logic based expert system to control insulin-pump doses. McGill University, ECE Department pp. 304-487(2005)
14. Soonthornphisaj, N. and Kijisirikul B.: An Automatic Membership Functions Generation using Decision Tree for Fuzzy Logic. In Proc of the 8th Int. Conf. Intelligent Technologies. pp. 87-94 (2007)
15. WBCD Data set: Blake, C.L., and Merz, C.J.: UJI Repository of Machine Learning Databases, <ftp://ftp.ics.uci.edu/pub/machine-learning-databases> (last accessed: 15 June 2006)
16. Yardimci, A., Hadimioglu, N., Bigat, Z. and Ozen, S.: Depth Control of Desflurane Anesthesia with Cardiovascular-based an Adaptive Neuro-Fuzzy System. Advances in Soft Computing, 2, pp.787-796 (2005)
17. Zadeh, L.A.: Fuzzy sets. Inf Control. vol.8, pp.338-353 (1965)

ประวัติการศึกษาและการทำงาน

ชื่อ	นางสาวภัทรวดี เตียวเตชะเดชา
เกิดวันที่	10 มกราคม พ.ศ. 2527
สถานที่เกิด	อำเภอเมือง จังหวัดขอนแก่น
ประวัติการศึกษา	วท.บ. (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยขอนแก่น
ตำแหน่งปัจจุบัน	-
สถานที่ทำงานปัจจุบัน	-
ผลงานดีเด่นและรางวัลทางวิชาการ	1. ขั้นตอนวิธีการปรับหารูปแบบกราฟฟังก์ชันความเป็นสมาชิกที่เหมาะสมสำหรับการจำแนกประเภทข้อมูลโดยใช้ฟัซซี่ลอจิก 2. A Self-tuning of Membership Functions for Medical Diagnosis
ทุนการศึกษาที่ได้รับ	-