

บทที่ 2

ทฤษฎีหรือหลักการที่เกี่ยวข้อง

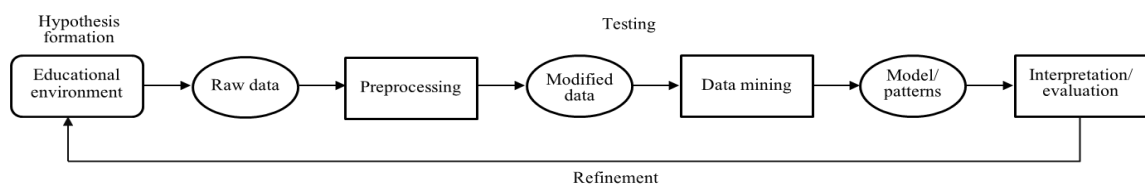
การวิเคราะห์การสำเร็จการศึกษาของนักศึกษามีกระบวนการทำงานเพื่อนำไปสู่ผลลัพธ์ที่ต้องการตามขั้นตอนการทำเหมืองข้อมูลทางการศึกษา เทคนิคที่ใช้ในการทำนายเกรดเฉลี่ยสะสมคือ เทคนิคระบบอนุมานนิเวศวิทยาแบบปรับตัวได้ โดยที่เทคนิคนี้จะมีการจัดกลุ่มข้อมูลเริ่มต้นด้วยเทคนิคแบบลบบอกและแบบพีชชีมิน มีการใช้ระบบพีชชีแบบซูเกโนเพื่ออนุมานกฎพีชชี และมีการเรียนรู้เพื่อปรับพารามิเตอร์ด้วยโครงข่ายประสาทเทียมแบบไปข้างหน้ากับแบบย้อนกลับ ส่วนเทคนิคที่ใช้ในการจำแนกกลุ่มที่มีค่าผลลัพธ์เป็นค่าความน่าจะเป็น คือ โครงข่ายรู้จำแบบ การถดถอยโลจิสติก การจำแนกแบบอย่างง่าย และต้นไม้ตัดสินใจ มีทฤษฎีหรือหลักการที่เกี่ยวข้องดังนี้

2.1 การทำเหมืองข้อมูลทางการศึกษา

การทำเหมืองข้อมูลเป็นกระบวนการกลั่นกรองสารสนเทศ ที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ ใช้ในการทำนายแนวโน้มหรือพฤติกรรมของข้อมูลจะเกิดขึ้นในอนาคต โดยอาศัยข้อมูลในอดีต เพื่อค้นหารูปแบบความสัมพันธ์ และองค์ความรู้ใหม่ จากข้อมูล ซึ่งเป็นขั้นตอนที่สำคัญในกระบวนการสืบค้นความรู้ที่เป็นประโยชน์ และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge discovery in databases) เมื่อนำมาประยุกต์ใช้กับงานทางการศึกษา เรียกว่า การทำเหมืองข้อมูลทางการศึกษา จะใช้เพื่อตรวจสอบ ประเมินระบบ ปรับปรุงบางส่วนของคุณภาพการศึกษา และเพื่อวางรากฐานกระบวนการเรียนรู้ให้มีประสิทธิภาพ มากขึ้น ปัจจุบันมีงานวิจัยที่เกี่ยวข้องกับปัญหาด้านการศึกษาในสถาบันระดับสูงจำนวนมาก และมีการแก้ปัญหาดังกล่าวได้สำเร็จ เช่น Romero และ Ventura [4] ศึกษาวรรณกรรมที่เกี่ยวข้อง จำกัดคำนิยาม และอธิบายกระบวนการทำงาน ระหว่างปี 1995 ถึง 2005 ต่อมา Baker และ Yacef [7] ได้ขยายการสำรวจจนถึงส่วนหลังปี 2005 จากนั้น Alejandro [9] ได้สำรวจช่วงปี 2010 ถึง 2013 การทำเหมืองข้อมูลทางการศึกษามีจำนวนเพิ่มขึ้น เพื่อใช้จัดการกับการลงทะเบียนเรียน การสำเร็จการศึกษา ผลการเรียนรู้ ความสามารถพิเศษ การศึกษาผ่านคอมพิวเตอร์ การรักษานักศึกษา และอื่นๆ ทำให้เกิดความเข้าใจ มีปรับปรุงประสิทธิภาพการศึกษา และประเมินกระบวนการเรียนรู้ของนักศึกษาให้ดีขึ้น

การทำเหมืองข้อมูลด้านการศึกษจะเป็นการรวมวิทยาการหลายอย่างเข้าด้วยกันประกอบด้วย สถิติ วิทยาการคอมพิวเตอร์ และการศึกษา ซึ่งจะมีส่วนที่ทับซ้อนกันอยู่คือ การศึกษานบนคอมพิวเตอร์ (Computer-based education) การทำเหมืองข้อมูลและกลไกการเรียนรู้ และการวิเคราะห์การเรียนรู้ (Learning analytics) ซึ่งกระบวนการทำงานของการทำเหมืองข้อมูลทางการศึกษาไม่ได้แตกต่างจากการทำเหมืองข้อมูลที่ประยุกต์ใช้กับงานด้านอื่น แต่จะมีความแตกต่างกันที่วัตถุประสงค์ของการทำงาน การนำมาปรับใช้กับปัญหาทางการศึกษาโดยเฉพาะ และยังมีปัจจัยด้านจิตวิทยาเข้ามาเกี่ยวข้องด้วย

กระบวนการสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานความรู้ด้านการศึกษา (Educational knowledge discovery process) จะเป็นวัฏจักรวนซ้ำของการตั้งข้อสมมติฐาน (Hypothesis formation) การทดสอบ และการกลั่นกรอง (Refinement) เพื่อใช้ในการตัดสินใจเลือกวิธีการปรับปรุงสภาพแวดล้อมด้านการศึกษา และการปรับปรุงการเรียนรู้ของนักศึกษา มีกระบวนการทำงาน ดังภาพที่ 2.1 เริ่มด้วยขั้นตอนการตั้งข้อสมมติฐานจากสภาพแวดล้อมด้านการศึกษา (Educational environment) เช่น การศึกษาในห้องเรียนแบบเดิม (Traditional classroom education) การศึกษานบนคอมพิวเตอร์ (Computer-based education) และระบบสารสนเทศที่สนับสนุน ข้อมูลที่ได้จะเป็นข้อมูลดิบที่มาจากหลายแหล่งที่แตกต่างกัน ต่อมานำข้อมูลดิบที่ได้มาทำการปรับปรุงเพื่ออยู่ในรูปแบบที่เหมาะสมกับการแก้ปัญหาทางการศึกษาที่เฉพาะและเลือกข้อมูลที่เกี่ยวข้องกับงานวิจัยเป็นขั้นตอนการเตรียมข้อมูล (Data preprocessing) ทำให้ได้ข้อมูลที่มีการปรับปรุงแล้ว (Modified data) ที่สามารถตอบคำถามได้ตรงกับปัญหา และลดตัวแปรที่ใช้ในการวิเคราะห์ ต่อมาเป็นการเลือกเทคนิคที่ใช้ในการทำเหมืองข้อมูล ส่วนมากเทคนิคการทำเหมืองข้อมูลทั่วไปจะประกอบด้วย การจำแนกประเภท การจัดกลุ่ม และการเชื่อมโยงความสัมพันธ์ ซึ่งการทำเหมืองข้อมูลด้านการศึกษาก็สามารถนำมาประยุกต์ใช้ได้อย่างประสบความสำเร็จ ส่วนเทคนิคที่แตกต่างจากการทำเหมืองข้อมูลทั่วไป เช่น วิธีการทำเหมืองข้อมูลแบบลำดับชั้น (Hierarchical data mining) และการสร้างแบบจำลองข้อมูลระยะยาว (Longitudinal data modeling) ขั้นตอนสุดท้ายจะแปลความหมายของผลลัพธ์ (Interpretation of results) เป็นการนำรูปแบบความสัมพันธ์หรือองค์ความรู้ใหม่ที่ได้มาตัดสินใจเลือกแนวทางแก้ไข ปัญหา เพื่อปรับปรุงสภาพแวดล้อมของการศึกษาหรือระบบให้ดีขึ้น



ภาพที่ 2.1 กระบวนการค้นหาคำรู้ทางการศึกษาและการทำเหมืองข้อมูล [10]

การประยุกต์ใช้งานในสภาพแวดล้อมด้านการศึกษาที่แก้ปัญหาผ่านการทำเหมืองข้อมูลมีเทคนิคที่ได้รับความนิยมในการใช้ทำเหมืองข้อมูลด้านการศึกษา เช่น การทำนาย (Prediction) การจัดกลุ่ม (Clustering) การตรวจสอบค่าผิดปกติ (Outlier detection) การทำเหมืองความสัมพันธ์ (Relationship mining) การวิเคราะห์เครือข่ายสังคม (Social network analysis) การทำเหมืองกระบวนการ (Process mining) การทำเหมืองข้อความ (Text mining) การกลั่นกรองข้อมูลการตัดสินใจของมนุษย์ (Distillation of data for human judgment) การสืบค้นด้วยแบบจำลอง (Discovery with models) การติดตามความรู้ (Knowledge tracing) และปัจจัยเมทริกซ์ที่ไม่เป็นค่าติดลบ (Nonnegative matrix factorization)

การสร้างแบบจำลองของการทำเหมืองข้อมูลทางการศึกษา แบ่งเป็น 2 ประเภท คือ แบบจำลองอธิบาย (Descriptive model) จะอธิบายชุดข้อมูลในลักษณะที่กระชับและสรุป จากนั้นจะนำเสนอคุณสมบัติที่น่าสนใจ เป็นการอธิบายรูปแบบที่ซ่อนเอาไว้เพื่อใช้เป็นแนวทางในการตัดสินใจ และแบบจำลองทำนาย (Predictive model) พยากรณ์พฤติกรรมจากข้อมูลประวัติการใช้งาน และข้อมูลของผู้ใช้งาน ซึ่งทราบผลลัพธ์อยู่แล้ว เพื่อสร้างแบบจำลองที่สามารถนำมาใช้พยากรณ์ค่าที่ชัดเจนจากข้อมูลที่แตกต่างกัน

การจำแนกประเภทและการทำนายเป็นเทคนิคการทำเหมืองข้อมูลทางการศึกษาที่สำคัญและเป็นประโยชน์ต่อการค้นพบรูปแบบของนักศึกษาที่ประสบความสำเร็จ ช่วยทำนายผลการเรียนของวิชานั้น และช่วยระบุนักศึกษาที่มีโอกาสได้เกรดไม่ดีออก โดยที่เทคนิคการทำนายจะได้ผลลัพธ์เป็นค่าตัวเลขต่อเนื่อง ส่วนเทคนิคการจำแนกประเภทจะเป็นค่าตัวเลขที่ไม่ต่อเนื่องหรือคลาสข้อมูล เทคนิคที่ได้รับความนิยมได้แก่ โครงข่ายประสาทเทียม เครือข่ายเบย์เซียน ระบบที่มีพื้นฐานอยู่บนกฎ (Rule-based systems) การวิเคราะห์ความถดถอย และการวิเคราะห์สหสัมพันธ์ เป็นต้น

2.2 การจัดกลุ่มข้อมูล

การจัดกลุ่มข้อมูล (Clustering) [11] เป็นการจำแนกข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่มย่อย ซึ่งถ้าจำนวนกลุ่มข้อมูลมีมากจะทำให้จำนวนกฎฟัซซีมีมากขึ้น เวลาที่ใช้ในการประมวลผลก็จะมากขึ้น แต่ถ้ามีจำนวนกฎน้อยจะทำให้ไม่ครอบคลุมข้อมูลที่มีอยู่ ค่าความคลาดเคลื่อนก็จะมากตาม ดังนั้นการจัดกลุ่มที่เหมาะสมกับข้อมูลจึงเป็นส่วนเริ่มต้นที่สำคัญของการสร้างโมเดล เทคนิคที่ใช้ในการจัดกลุ่มมี 2 เทคนิคคือ เทคนิคแบบลบออก และ เทคนิคฟัซซีมีน มีรายละเอียดดังนี้

2.2.1 การจัดกลุ่มแบบลบออก (Subtractive clustering) ขยายแนวคิดมาจาก Mountain Clustering จะหาค่าจุดศูนย์กลางของข้อมูลที่มีเหมือนกันจากความหนาแน่น (Density) โดยที่ไม่ต้องกำหนดจำนวนกลุ่ม มีขั้นตอนการทำงานดังนี้

- 1) คำนวณหาจุดศูนย์กลางของกลุ่มแรกด้วยการนำข้อมูลทุกจุดมาคำนวณหาค่าความหนาแน่นด้วยสมการที่ (2.1) จะเลือกจุดข้อมูลที่มีค่าความหนาแน่นสูงสุด

$$D_i = \sum_{j=1}^n \exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}\right) \quad (2.1)$$

เมื่อ D_i คือค่าความหนาแน่นสูงสุด

r_a คือค่ารัศมีระหว่างกลุ่ม

x_i คือจุดข้อมูลลำดับที่ i ที่ใช้เป็นจุดคำนวณค่าความหนาแน่น

x_j คือจุดข้อมูลลำดับที่ j

n คือจำนวนข้อมูลเรียนรู้

$\| \cdot \|$ คือการวัดระยะทางแบบยูคลิดีียน

- 2) คำนวณหาจุดศูนย์กลางของกลุ่มถัดไปด้วยสมการที่ (2.2) ทำซ้ำต่อไปจนกว่าจุดข้อมูลทุกจุดจะอยู่ในช่วงรัศมีของแต่ละกลุ่ม

$$D_i = D_i - D_{c_1} \times \exp\left(-\frac{\|x_i - x_{c_1}\|^2}{(r_b/2)^2}\right) \quad (2.2)$$

เมื่อ r_b คือค่ารัศมีระหว่างกลุ่มหาได้จาก $r_b \times \sigma$

σ คือค่าที่ใช้ลดปัญหาระยะขอบของกลุ่มที่อยู่ใกล้กัน

2.2.2 การจัดกลุ่มแบบฟัซซีซีมีน (Fuzzy c-means, FCM) เป็นการนำเอาฟัซซีลอจิกมาใช้ในการจัดกลุ่ม ทำให้ข้อมูลหนึ่งสามารถอยู่ได้มากกว่าหนึ่งกลุ่ม โดยจะแสดงเป็นค่าความเป็นสมาชิกในแต่ละกลุ่มที่แตกต่างกัน จึงทำให้สามารถจัดการกับปัญหาความคลุมเครือในการจัดกลุ่มได้ พัฒนาโดย J. C. Dunn ปี ค.ศ. 1973 หลังจากนั้น J. C. Bezdek ได้นำมาปรับปรุงในปี ค.ศ. 1981 ขั้นตอนการทำงานของฟัซซีซีมีนมีดังนี้

- 1) กำหนดกลุ่มข้อมูลที่ต้องการจัดกลุ่ม ค่าเงื่อนไขการหยุด สัมประสิทธิ์ความเป็นฟัซซี และค่าเริ่มต้นของค่าความเป็นสมาชิกด้วยการสุ่มค่าระหว่าง 0 ถึง 1 โดยค่าความเป็นสมาชิกที่ได้ต้องมีคุณสมบัติดังสมการที่ (2.3) และ (2.4)

1.1) ค่าความเป็นสมาชิกจะอยู่ระหว่าง 0 ถึง 1

$$\mu_{ij} \in [0,1], \forall i, j \quad (2.3)$$

1.2) ผลรวมของค่าความเป็นสมาชิกในฟัซซีเซตต้องเท่ากับ 1

$$\sum_{i=1}^c \mu_{ij} = 1 \quad (2.4)$$

เมื่อ μ_{ij} คือค่าความเป็นข้อมูลเรียนรู้ลำดับที่ i ในกลุ่มที่ j
 c คือจำนวนกลุ่มทั้งหมดในฟัซซีเซต

2) คำนวณจุดศูนย์กลางของกลุ่มฟัซซี c_i ด้วยสมการที่ (2.5)

$$c_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m} \quad (2.5)$$

เมื่อ c_i คือจุดศูนย์กลางของกลุ่มลำดับที่ i
 x_j คือข้อมูลเรียนรู้ลำดับที่ j
 μ_{ij}^m คือค่าความเป็นสมาชิกของข้อมูลเรียนรู้ลำดับที่ i ของกลุ่มที่ j

3) คำนวณฟังก์ชันวัตถุประสงค์ (Objection function) ในรอบแรก ด้วยสมการที่ (2.6)

ในรอบต่อไปจะหาด้วยสมการที่ (2.7) และจะหยุดทำงาน เมื่อค่า $\|J - \bar{J}\| < \varepsilon$

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d_{ij}^2 \quad (2.6)$$

$$J(U, c_1, \dots, c_c, \lambda_1, \dots, \lambda_n) = J(U, c_1, \dots, c_c) + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c \mu_{ij} - 1 \right) \quad (2.7)$$

เมื่อ J เป็นค่าฟังก์ชันวัตถุประสงค์

$\bar{J}$ เป็นค่าฟังก์ชันวัตถุประสงค์ใหม่

m เป็นสัมประสิทธิ์ความเป็นฟัซซีที่มีค่าเป็นจำนวนเต็มบวกและต้องมากกว่า 1

d_{ij} เป็นค่าระยะทางแบบยูคลิเดียน (Euclidean distance) ระหว่างจุดศูนย์กลางลำดับที่ i และจุดข้อมูลลำดับที่ j หาได้จากสมการที่ (2.8)

$$d_{ij} = \|c_i - x_j\| \quad (2.8)$$

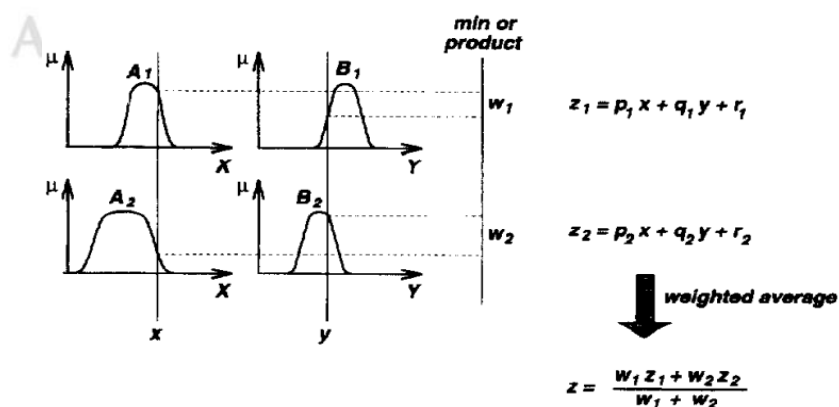
4) ปรับค่าความเป็นสมาชิกใหม่ ด้วยสมการที่ (2.9) กลับไปคำนวณข้อที่ 2 อีกครั้งจนกว่าจะครบรอบวนซ้ำ (Iteration)

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}}} \quad (2.9)$$

2.3 ระบบฟัซซี

ฟัซซีลอจิกมาจากตรรกะพื้นฐานความเป็นจริงของมนุษย์ที่ว่า ทุกสิ่งบนโลกแห่งความเป็นจริงไม่ได้มีเฉพาะสิ่งที่มีความแน่นอนเท่านั้น แต่มีหลายเหตุการณ์ที่เกิดขึ้นอย่างไม่แน่นอน (Uncertain) อาจเป็นสิ่งที่คลุมเครือ (Fuzzy) ไม่ชัดเจน จึงนำมาประยุกต์เป็นวิธีทางคณิตศาสตร์ เพื่อช่วยตัดสินใจภายใต้ความคลุมเครือ และไม่ชัดเจน คิดค้นโดย L. A. Zadeh ในปี ค.ศ. 1965 โดยที่จะจัดการความไม่แน่นอนด้วยการกำหนดค่าความเป็นสมาชิก (Degree of Membership) ในฟัซซีเซตด้วยฟังก์ชันความเป็นสมาชิก (Membership function) โครงสร้างพื้นฐานของระบบฟัซซี [8] มีการทำงานดังนี้ ขั้นตอนที่ 1 การแปลงค่าของข้อมูลอินพุตเป็นค่าฟัซซีอินพุต (Fuzzification) ซึ่งจะคำนวณค่าความเป็นสมาชิกของข้อมูลอินพุต (Crisp Input) โดยใช้ฟังก์ชันความเป็นสมาชิก ขั้นตอนที่ 2 การอนุมานฟัซซี (Fuzzy inferencing) เป็นการนำค่าฟัซซีอินพุตไปอนุมานผ่านกฎฟัซซีที่ตั้งขึ้นมาและได้ผลลัพธ์เป็นค่าฟัซซีเอาต์พุต โดยใช้กฎฟัซซีแบบ ถ้า-แล้ว ขั้นตอนที่ 3 จะทำค่าฟัซซีให้เป็นค่าปกติ (Defuzzification) แปลงค่าฟัซซีเอาต์พุตมาเป็นค่าปกติ (Crisp output) ระบบอนุมานกฎฟัซซีมี 2 ประเภทหลัก คือแบบ แมมดานิ (Mamdani) จะรวมผลการอนุมานของกฎด้วยวิธีการซ้อนทับจากกฎหลายๆ ข้อ และแบบซูเกโน (Sugeno) มีการอนุมานแบบรวมค่าน้ำหนัก (Weight sum) จากหลายกฎ เพื่อรวมเป็นข้อสรุปสุดท้าย งานวิจัยนี้ใช้ระบบอนุมานกฎฟัซซีแบบซูเกโนที่มาจากการเลือกใช้เทคนิคระบบอนุมานนิเวโรฟัซซีแบบปรับตัวได้

โครงสร้างของฟัซซีซูเกโนจะนำข้อมูลอินพุตมาประมวลผลเพื่อหาค่าความเป็นสมาชิกส่วนของฟังก์ชันความเป็นสมาชิกด้านอินพุต (Input membership function) ผลที่ได้จะนำมาบวกกันและตีความเพื่อกำหนดค่าน้ำหนักของกฎ และใช้เป็นค่าตัวแปรของการประมวลผลในส่วนของฟังก์ชันความเป็นสมาชิกด้านเอาต์พุต (Output membership function) สามารถเลือกใช้ได้ 2 รูปแบบคือ ลำดับที่ศูนย์ (Zero order) จะกำหนดความเป็นสมาชิกด้านเอาต์พุตด้วยค่าคงที่ และลำดับที่หนึ่ง (First order) จะกำหนดแบบสมการเชิงเส้น และสามารถแสดงความสัมพันธ์ในภาพที่ 2.2



ภาพที่ 2.2 โครงสร้างของฟัซซีซูเกโน [12]

การแปลงค่าข้อมูลอินพุตเป็นค่าฟัซซีอินพุตด้วยฟังก์ชันความเป็นสมาชิกด้านอินพุต ค่าที่ได้จะอยู่ในรูปปริมาณที่แสดงค่าระดับความเป็นสมาชิกของฟัซซีเซต นอกจากนี้สามารถอธิบายในรูปคุณภาพด้วยพจน์ภาษา (Linguistic term) ที่ใช้ในการแสดงแนวคิดและองค์ความรู้ในการสื่อสารของมนุษย์ เช่น พอใช้ ปานกลาง ดี เรียกการกำหนดค่านี้ว่า ตัวแปรเชิงภาษา (Linguistic variable) ฟังก์ชันความเป็นสมาชิกที่ใช้ในการกำหนดระดับความเป็นสมาชิกของตัวแปรในงานวิจัยนี้ [13] ได้เลือกใช้ 7 ชนิด ดังนี้

- 2.3.1 ฟังก์ชันเกาส์เซียน (Gaussian membership function, gaussmf) มีพารามิเตอร์ 2 ตัว คือ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ดังสมการที่ (2.10) และลักษณะกราฟดังภาพที่ 2.3

$$\mu(x) = f(x; c, \sigma) = \exp\left(\frac{-(x-c)^2}{2\sigma^2}\right) \quad (2.10)$$

เมื่อ c คือค่าเฉลี่ย และ σ คือส่วนเบี่ยงเบนมาตรฐาน

- 2.3.2 ฟังก์ชันเกาส์เซียนแบบรวมกัน (Gaussian combination membership function, gauss2mf) เป็นการนำฟังก์ชันเกาส์เซียนสองตัวมาคูณกันโดยที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนที่แตกต่างกัน มีพารามิเตอร์ 4 ตัว คือค่าเฉลี่ย 2 ค่า และค่าส่วนเบี่ยงเบน 2 ค่า ดังสมการที่ (2.11) และลักษณะกราฟดังภาพที่ 2.4

$$\mu(x) = f(x; \sigma_1, c_1, \sigma_2, c_2) = \exp\left(\frac{-(x-c_1)^2}{2\sigma_1^2}\right) \times \exp\left(\frac{-(x-c_2)^2}{2\sigma_2^2}\right) \quad (2.11)$$

เมื่อ c_1 คือค่าเฉลี่ยขอบล่าง c_2 คือค่าเฉลี่ยขอบบน σ_1 คือส่วนเบี่ยงเบนมาตรฐานขอบบน σ_2 คือส่วนเบี่ยงเบนมาตรฐานขอบล่าง

- 2.3.3 ฟังก์ชันระฆังคว่ำ (Generalized bell-shaped membership function, gbellmf) มีพารามิเตอร์ 3 ตัว ดังสมการที่ (2.12) และลักษณะกราฟดังภาพที่ 2.5

$$\mu(x) = f(x; a, b, c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \quad (2.12)$$

- 2.3.4 ฟังก์ชันซิกมอยด์ (Sigmoidal membership function, sigmf) มีพารามิเตอร์ 2 ตัว ดังสมการที่ (2.13) และลักษณะกราฟดังภาพที่ 2.6

$$\mu(x) = f(x; a, c) = \frac{1}{1 + e^{-a(x-c)}} \quad (2.13)$$

2.3.5 ฟังก์ชันความแตกต่างระหว่างสองฟังก์ชันซิกมอยด์ (Difference between two sigmoidal functions membership function, dsigmf) เป็นการลบกันของ 2 ฟังก์ชันซิกมอยด์ มีพารามิเตอร์ 4 ตัว ดังสมการที่ (2.14) และลักษณะกราฟดังภาพที่ 2.7

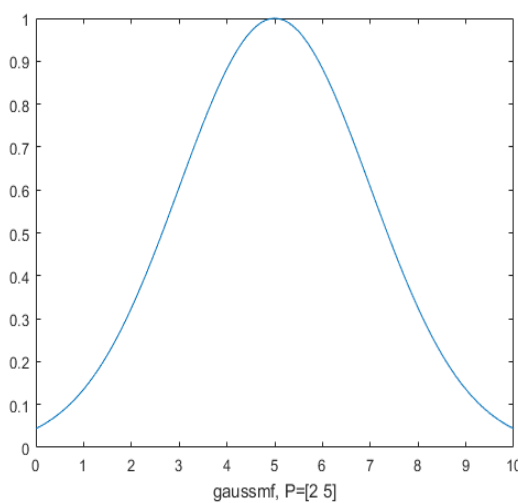
$$\mu(x) = f(x; a_1, c_1, a_2, c_2) = \left(\frac{1}{1 + e^{-a_1(x-c_1)}} \right) - \left(\frac{1}{1 + e^{-a_2(x-c_2)}} \right) \quad (2.14)$$

2.3.6 ฟังก์ชันการคูณกันระหว่างสองฟังก์ชันซิกมอยด์ (Product of two sigmoidal membership functions, psigmf) เป็นการคูณกันของ 2 ฟังก์ชันซิกมอยด์ มีพารามิเตอร์ 4 ตัว ดังสมการที่ (2.15) และลักษณะกราฟดังภาพที่ 2.8

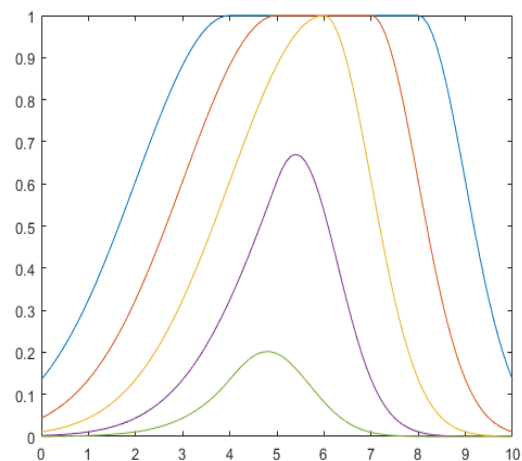
$$\mu(x) = f(x; a_1, c_1, a_2, c_2) = \left(\frac{1}{1 + e^{-a_1(x-c_1)}} \right) \times \left(\frac{1}{1 + e^{-a_2(x-c_2)}} \right) \quad (2.15)$$

2.3.7 ฟังก์ชันตัวเอส (S-shaped membership function, smf) มีลักษณะเป็นรูปตัวเอส และมีพารามิเตอร์ 2 ตัว ดังสมการที่ (2.16) และลักษณะกราฟดังภาพที่ 2.9

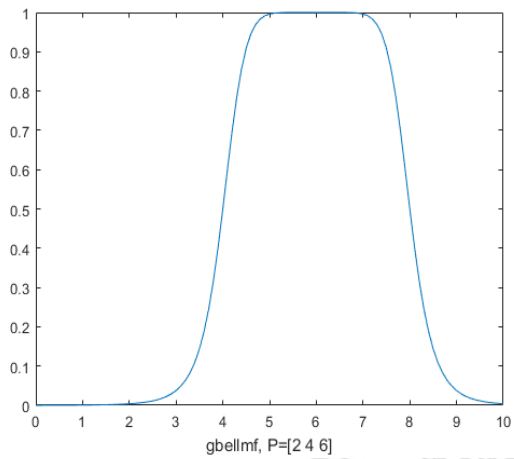
$$\mu(x) = f(x; a, b) = \begin{cases} 0 & , x \leq a \\ 2 \left(\frac{x-a}{b-a} \right)^2 & , a < x \leq \frac{a+b}{2} \\ 1 - 2 \left(\frac{x-a}{b-a} \right)^2 & , \frac{a+b}{2} < x \leq b \\ 1 & , x \geq b \end{cases} \quad (2.16)$$



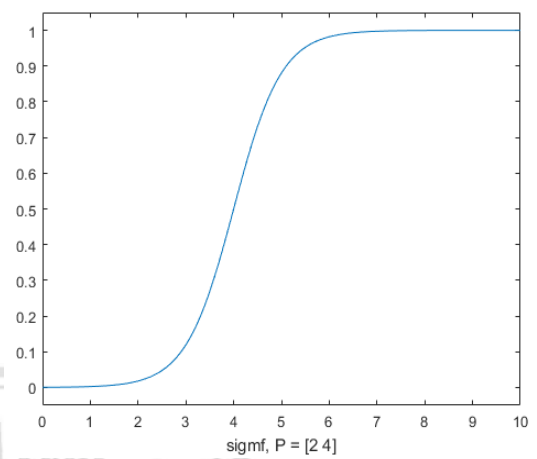
ภาพที่ 2.3 กราฟฟังก์ชันเกาส์เซียน [13]



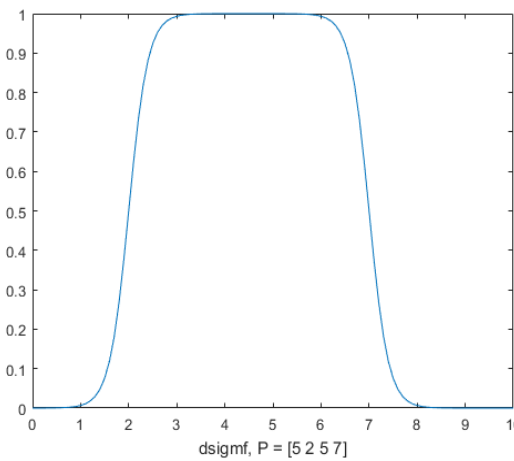
ภาพที่ 2.4 กราฟฟังก์ชันเกาส์เซียนแบบรวมกัน [13]



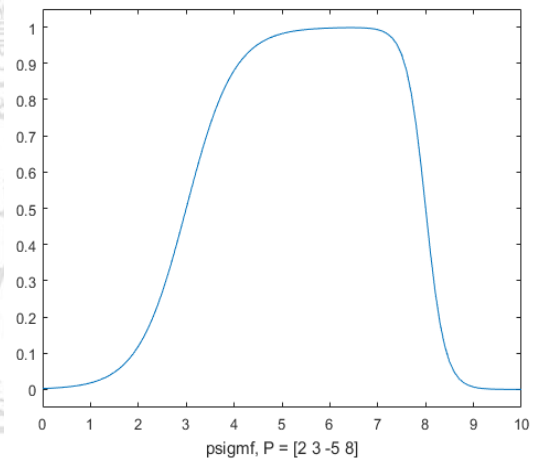
ภาพที่ 2.5 กราฟฟังก์ชันระฆังคว่ำ [13]



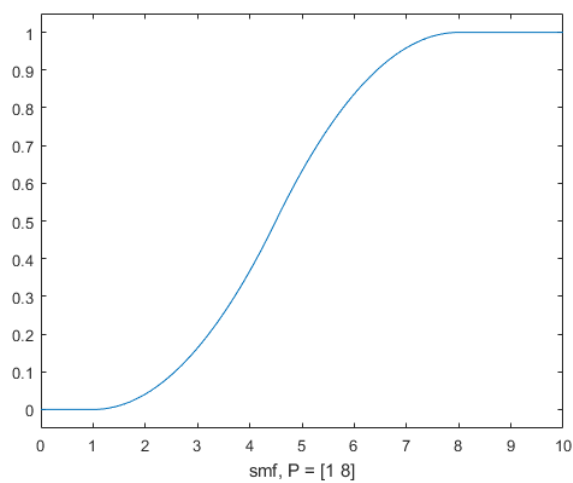
ภาพที่ 2.6 กราฟฟังก์ชันซิกมอยด์ [13]



ภาพที่ 2.7 กราฟฟังก์ชันความแตกต่าง
ระหว่างสองฟังก์ชันซิกมอยด์ [13]



ภาพที่ 2.8 กราฟฟังก์ชันการคูณกัน
ระหว่างสองฟังก์ชันซิกมอยด์ [13]



ภาพที่ 2.9 กราฟฟังก์ชันตัวเอส [13]

2.4 โครงข่ายปรับตัวได้

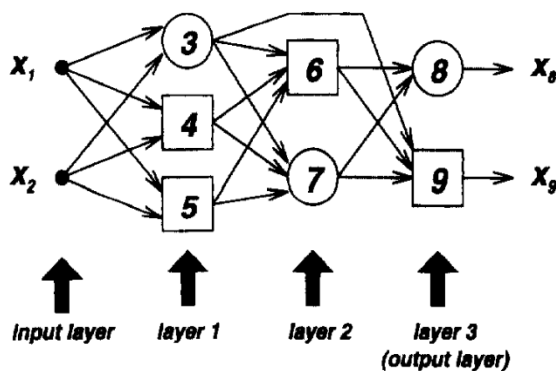
โครงข่ายประสาทเทียมมีแนวคิดมาจากการทำงานของสมองมนุษย์ที่มีการเรียนรู้และจดจำด้วยการทำงานแบบเชื่อมต่อ ข้อมูลต่างๆ มีการประมวลผล ตีความ วิเคราะห์ แล้วได้ผลลัพธ์เป็นความรู้ที่มาจากกระบวนการเรียนรู้ จึงนำมาจำลองเป็นตัวแทนทางคณิตศาสตร์ นำเสนอโดย McCulloch และ Pitts ในปี ค.ศ. 1943 จากนั้นในปี ค.ศ. 1943 Rosenblatt ได้สร้างและใช้งานเป็นคนแรก

หลักการพื้นฐานคือ เมื่อมีข้อมูลอินพุตส่งเข้ามาจะคูณกับค่าน้ำหนักเป็นการให้ความสำคัญกับข้อมูลอินพุตแต่ละตัว จากนั้นรวมค่าถ่วงน้ำหนักที่ได้ทั้งหมดเข้าด้วยกัน เพื่อนำไปวิเคราะห์ ตีความ ด้วยฟังก์ชันกระตุ้น (Activation function) ทำให้ได้ค่าเอาต์พุตออกมา สถาปัตยกรรมของโครงข่ายประสาทเทียมมีหลากหลายรูปแบบเช่น แบบป้อนไปข้างหน้า แบบย้อนกลับ ส่วนของการเรียนรู้แบ่งเป็น 2 ประเภท คือ แบบมีการสอน (Supervised learning) และแบบไม่มีการสอน (Unsupervised learning)

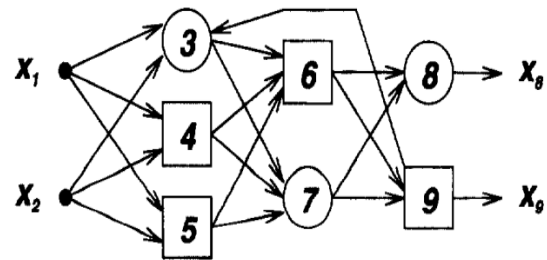
โครงข่ายปรับตัวได้ (Adaptive network) [14] มีสถาปัตยกรรมแบบโครงข่าย มีโหนดที่เชื่อมต่อกันด้วยลิงค์โดยตรง แต่ละโหนดแสดงถึงกระบวนการทำงาน และลิงค์ระหว่างโหนดแสดงถึงความสัมพันธ์ระหว่างโหนดที่เชื่อมต่อกัน โหนดทั้งหมดหรือบางส่วนที่สามารถปรับได้ เอาต์พุตที่ได้ในแต่ละโหนดขึ้นอยู่กับพารามิเตอร์ปรับได้ โดยจะมีสัญลักษณ์เพื่อหน้าที่ของแต่ละโหนดดังนี้ โหนดสี่เหลี่ยมเป็นโหนดที่มีพารามิเตอร์ปรับได้ และโหนดวงกลมเป็นโหนดที่ไม่มีพารามิเตอร์ปรับได้ โครงข่ายแบบปรับตัวได้แบ่งเป็น 2 ประเภทตามลักษณะการเชื่อมต่อ คือ แบบไปข้างหน้าที่เอาต์พุตของแต่ละโหนดจะส่งผ่านจากอินพุตไปเอาต์พุตเท่านั้นดังภาพที่ 2.10 และแบบย้อนกลับที่ส่งเอาต์พุตกลับมายังอินพุตดังภาพที่ 2.11

การสร้างโครงข่ายต้องประกอบด้วยชุดข้อมูลเรียนรู้ กฎการเรียนรู้หรืออัลกอริทึมที่ใช้ในการเรียนรู้ เพื่อเป็นกระบวนการในการปรับพารามิเตอร์ให้โครงข่ายมีประสิทธิภาพมากขึ้น การวัดความผิดพลาดเพื่อหาค่าความคลาดเคลื่อนระหว่างเอาต์พุตที่ต้องการกับเอาต์พุตของโครงข่ายภายใต้เงื่อนไขตัวแปรอินพุตเดียวกัน

กระบวนการเรียนรู้ของโครงข่ายปรับตัวได้เป็นการหาความสัมพันธ์ระหว่างอินพุตกับเอาต์พุตที่มีลักษณะแบบเชิงเส้นหรือแบบไม่เชิงเส้นขึ้นอยู่กับโครงสร้างของโครงข่ายและฟังก์ชันภายในของแต่ละโหนดและเพื่อประสิทธิผลของการสร้างโครงข่ายที่แสดงความสัมพันธ์ระหว่างอินพุตและเอาต์พุตแบบไม่เชิงเส้นจึงต้องมีการปรับพารามิเตอร์เพื่อให้ได้เอาต์พุตของโมเดลตามที่ต้องการ เรียกว่ากระบวนการเรียนรู้มีวิธีการที่ใช้ในการเรียนรู้เช่น วิธีประเมินกำลังสองน้อยสุด (Least squares estimation) การเคลื่อนลาดชันลงมา (Gradient descent method)



ภาพที่ 2.10 โครงข่ายแบบปรับตัวได้แบบไปข้างหน้า [12]



ภาพที่ 2.11 โครงข่ายแบบปรับตัวได้แบบย้อนกลับ [12]

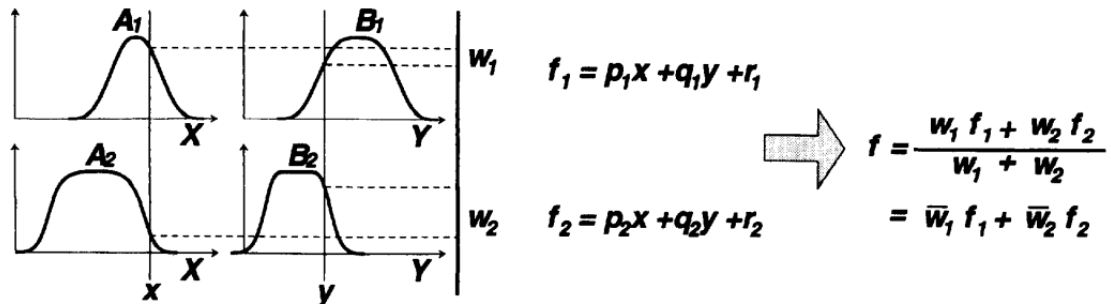
2.5 ระบบอนุมานนิเวศวิทยาแบบปรับตัวได้

ระบบนิเวศวิทยา เป็นการศึกษาความสัมพันธ์ระหว่างระบบนิเวศกับโครงสร้างประชากรที่มี แนวคิดมาจากการนำข้อดีของทั้งสองแบบมารวมกันทำให้ลดข้อเสียของแต่ละวิธีได้ โดยที่โครงข่ายประสาทเทียมที่มีคุณสมบัติการปรับแต่งความรู้ที่ซ่อนอยู่ภายในโครงข่ายด้วยการต่อเชื่อมโยงกัน และมีการส่งผ่านข้อมูลแบบขนานทำให้มีการประมวลผลที่เร็ว แต่ผลลัพธ์ที่ได้ไม่สามารถอธิบายในลักษณะของเหตุและผลได้ ส่วนระบบนิเวศวิทยาสามารถอธิบายเหตุและผลที่สอดคล้องกับตรรกะความคิดของมนุษย์ เพราะมีการตีความในภาพถ้อยคำแล้ว แต่ไม่มีกระบวนการเรียนรู้ในการปรับแต่งโครงสร้างของกฎและตัวแปรของระบบ จึงต้องอาศัยผู้เชี่ยวชาญ ระบบผสมนิเวศวิทยาที่นิยม คือ ระบบอนุมานนิเวศวิทยาแบบปรับตัวได้[8] คิดค้น โดย Jang [12]

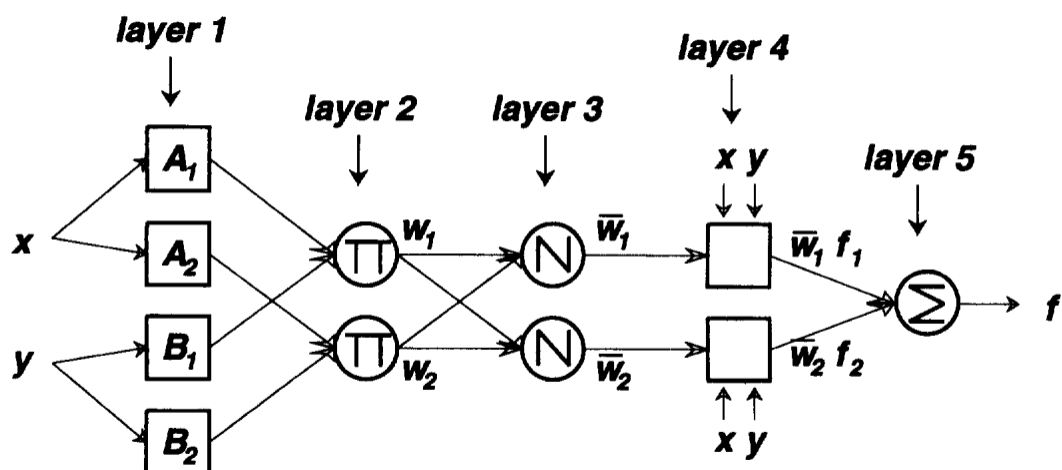
ระบบอนุมานนิเวศวิทยาแบบปรับตัวได้ หรือเรียกอีกชื่อว่า ระบบอนุมานฟัซซีบนฐานโครงข่ายที่ปรับตัวได้ (Adaptive network-based fuzzy inference system) เป็นการทำงานร่วมกันระหว่างโครงข่ายแบบปรับตัวได้กับระบบอนุมานฟัซซีแบบซูเกโน ระบบฟัซซีจะใช้สร้างกฎฟัซซีจากการเรียนรู้ของตัวแปรอินพุต และตัวแปรเอาต์พุต ดังภาพที่ 2.12 ส่วนโครงข่ายประสาทเทียมจะมีการเรียนรู้แบบไปข้างหน้า และแบบย้อนกลับเพื่อปรับพารามิเตอร์ของแต่ละกฎฟัซซี ดังภาพที่ 2.13

การเรียนรู้ของระบบอนุมานนิเวศวิทยาแบบปรับตัวได้เป็นการผสมผสานระหว่างแบบไปข้างหน้า (Forward pass) ด้วยวิธีประเมินกำลังสองน้อยสุด เพื่อปรับค่าพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านเอาต์พุต กับแบบย้อนกลับ (Backward pass) ด้วยการเคลื่อนลาดชันลงมา เพื่อปรับค่าพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านอินพุต ในแต่ละรอบการเรียนรู้จะเริ่มต้นด้วยการเรียนรู้ไปข้างหน้า และการเรียนรู้แบบย้อนกลับได้ทันทีในหนึ่งรอบการเรียนรู้ โดยจะมีรูปแบบ

การเรียนรู้แบบกลุ่ม (Batch learning) ที่มีการปรับพารามิเตอร์เพียงครั้งเดียวหลังจากข้อมูลทุกระเบียงนนำเข้าเพื่อหาค่าความผิดพลาดแล้วในแต่ละรอบการเรียนรู้ สถาปัตยกรรมของระบบอนุมานนิวโรฟิซซึแบบปรับตัวได้ประกอบด้วยการทำงาน 5 ชั้น ดังภาพที่ ซึ่งสามารถอธิบายหลักการเรียนรู้ เพื่อปรับพารามิเตอร์ได้ดังนี้



ภาพที่ 2.12 โมเดลฟิซซึแบบซูเกโนลำดับที่ 1 ที่มี 2 ตัวแปรอินพุตได้กฎฟิซซึ 2 กฎ [12]



ภาพที่ 2.13 สถาปัตยกรรมของระบบอนุมานนิวโรฟิซซึแบบปรับตัวได้ [12]

2.5.1 การเรียนรู้แบบไปข้างหน้าเป็นการเรียนรู้เพื่อปรับพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้าน เพื่อให้ได้ค่าเอาต์พุตที่มีค่าความผิดพลาดน้อยที่สุดหรือตามที่กำหนด ด้วยการ ใช้วิธีประเมินกำลังสองน้อยสุด ซึ่งจะมีการเรียนรู้ตามสถาปัตยกรรมของระบบอนุมานนิวโรฟิซซึแบบปรับตัวได้ที่แบ่งออกเป็น 5 ชั้น กำหนดให้มีตัวแปรอินพุต 2 ตัวแปรแทนค่าด้วย x และ y ตามลำดับ ตัวแปรเอาต์พุต 1 ตัวแปรแทนค่าด้วย z มีระบบอนุมานฟิซซึแบบซูเกโนลำดับที่ 1 ได้ค่าเอาต์พุตแบบเชิงเส้น และมีกฎฟิซซึ 2 กฎ คือ กฎข้อที่ 1: If x is A_1 and y is B_1 then $f_1 = p_1x + q_1y + r_1$, กฎข้อที่ 2: If x is A_2 and y is B_2 then $f_2 = p_2x + q_2y + r_2$.

ขั้นที่ 1 จะแปลงข้อมูลอินพุตเป็นค่าฟัซซีด้วยฟังก์ชันความเป็นสมาชิก ทุกโหนดในขั้นนี้สามารถปรับค่าได้ เรียกพารามิเตอร์ในโหนดนี้ว่า พารามิเตอร์ด้านอินพุต ซึ่งจะปรับเมื่อมีการเรียนรู้แบบย้อนกลับ เขียนสมการได้ดังสมการที่ (2.17) และ (2.18)

$$O_{1j} = \mu_{A_i}(x), i=1,2 \quad (2.17)$$

$$O_{1j} = \mu_{B_{i-2}}(y), i=3,4 \quad (2.18)$$

เมื่อ x, y เป็นโหนดอินพุตที่ $i, i-2$ ตามลำดับ

A_i, B_{i-2} เป็นตัวแปรเชิงภาษาเช่น พอใช้ ปานกลาง ดี

ค่า O_{1j} คือค่าระดับความเป็นสมาชิกของฟัซซีเซตประกอบด้วย A_1, A_2, B_1, B_2 หาได้จากฟังก์ชันความเป็นสมาชิก เช่น ฟังก์ชันเกาส์เซียนตามสมการที่ (2.10) ซึ่งมีพารามิเตอร์ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเป็นพารามิเตอร์ที่สามารถปรับค่าได้

ขั้นที่ 2 ทุกโหนดในขั้นนี้จะไม่สามารถปรับพารามิเตอร์ได้ ค่าเอาต์พุตของขั้นนี้มาจากการคูณกันของสัญญาณทั้งหมดที่เข้ามาจากขั้นที่ 1 ด้วยตัวดำเนินการทีนอร์ม (T-norm) แต่ละโหนดแสดงถึงค่าน้ำหนัก (Firing strength) ในกฎฟัซซี ดังสมการที่ (2.19)

$$O_{2j} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), i=1,2 \quad (2.19)$$

เมื่อ O_{2j} เป็นเอาต์พุตขั้นที่ 2 ข้อมูลระยะเบี่ยนที่ i

w_i เป็นค่าน้ำหนักของแต่ละกฎฟัซซี

$\mu_{A_i}(x), \mu_{B_i}(y)$ เป็นค่าความเป็นสมาชิกที่ได้จากเอาต์พุตขั้นที่ 1

ขั้นที่ 3 ทุกโหนดในขั้นนี้จะไม่สามารถปรับพารามิเตอร์ได้ ค่าเอาต์พุตของขั้นนี้คือค่าน้ำหนักที่เป็นมาตรฐานเดียวกัน (Normalized firing strength) ด้วยการนำค่าน้ำหนักที่ได้จากขั้นที่ 2 มารวมกัน ดังสมการที่ (2.20)

$$O_{3j} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, i=1,2 \quad (2.20)$$

ขั้นที่ 4 จะคำนวณหาเอาต์พุตของแต่ละกฎฟัซซี โดยนำค่าที่ได้จากขั้นที่ 3 คูณกับเอาต์พุตเชิงเส้น ทุกโหนดในขั้นนี้สามารถปรับค่าได้ ซึ่งการเรียนรู้แบบไปข้างหน้าจะปรับพารามิเตอร์ของขั้นนี้ ดังสมการที่ (2.21)

$$O_{4j} = \overline{w_i} f_i = \overline{w_i} (p_i x + q_i + r_i) \quad (2.21)$$

เมื่อ $\overline{w_i}$ เป็นค่าน้ำหนักที่เป็นมาตรฐานเดียวกันมาจากขั้นที่ 3

p_i, q_i, r_i เป็นเซตพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านเอาต์พุต

ชั้นที่ 5 ในชั้นนี้จะมีเพียงหนึ่งโหนดที่ไม่สามารถปรับพารามิเตอร์ได้ เป็นการรวมค่าที่ได้ในชั้นที่ 4 ทุกกฎฟัซซีเข้าด้วยกัน ดังสมการที่ (2.22)

$$O_{5,1} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (2.22)$$

วิธีประเมินกำลังสองน้อยสุดจะใช้ค่าความผิดพลาดกำลังสองต่ำสุด (Minimise squared error) ดังสมการที่ (2.23)

$$\|A\theta - B\|^2 \quad (2.23)$$

เมื่อ A เป็นเอาต์พุตที่ได้จากชั้นที่ 3

B เป็นเอาต์พุตเป้าหมาย

θ เป็นค่าพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านเอาต์พุตที่ไม่ทราบค่ามีความเกี่ยวข้องกับเซตพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านเอาต์พุตในชั้นที่ 4

จะทำการปรับแต่งพารามิเตอร์ θ ที่ไม่ทราบค่าใหม่ด้วยวิธีชูโดอินเวอร์ส (Moore-penrose pseudo inverse) ด้วยสมการที่ (2.24)

$$\hat{\theta} = (A^T A)^{-1} A^T B \quad (2.24)$$

เมื่อ A^T เป็นการทรานสโพสของ A

$(A^T A)^{-1} A^T$ เป็นชูโดอินเวอร์สของ A

$\hat{\theta}$ เป็นเซตพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านเอาต์พุตใหม่

เนื่องจากวิธีชูโดอินเวอร์สใช้เวลาประมวลผลนาน จึงมีการใช้วิธีรีเคอร์ซีฟกำลังสองน้อยสุด (Recursive least squares) โดยจะแยกเป็น 2 สมการ สมการแรกจะปรับพารามิเตอร์ $\hat{\theta}$ โดยที่หลังจากข้อมูลลำดับที่ i เข้ามาถึง สมการจะปรับค่าแบบวนซ้ำ ในสมการที่ (2.25)

$$\hat{\theta}_{i+1} = \hat{\theta}_i + P_{i+1} a_{i+1} (b_{i+1}^T - a_{i+1}^T \hat{\theta}_i) \quad (2.25)$$

ต่อมาสมการที่ 2 จะปรับค่าเมตริกซ์ความแปรปรวนร่วม (Covariance matrix) P_{i+1} ที่จะมีปรับค่าแบบวนดังสมการที่ (2.26)

$$P_{i+1} = P_i - \frac{P_i a_{i+1} a_{i+1}^T P_i}{1 + a_{i+1}^T P_i a_{i+1}}, i = 0, 1, \dots, P-1 \quad (2.26)$$

เมื่อ $\hat{\theta}_0 = 0$

a_{i+1}^T เป็นทรานสโพสของเมตริกซ์ A

b_{i+1}^T เป็นทรานสโพสของเมตริกซ์ B

$P_0 = \gamma I$ โดยที่ค่าของ γ จะเป็นจำนวนบวกมากๆ เพื่อแสดงความไม่แน่นอนในการกำหนดค่าเริ่มต้นของพารามิเตอร์ และ I เป็นเมตริกซ์ที่มีค่าเป็นหนึ่งในแถวแยง

หลังจากปรับค่าพารามิเตอร์แล้ว จะคำนวณเอาต์พุตของโครงข่ายและคำนวณค่าความผิดพลาดของข้อมูลเรียนรู้แต่ละระเบียน ด้วยสมการที่ (2.27)

$$E_n = (T_n - O_n)^2 \quad (2.27)$$

เมื่อ E_n เป็นค่าความผิดพลาดของข้อมูลเรียนรู้ระเบียนที่ n

T_n เป็นเอาต์พุตเป้าหมายที่ต้องการ

O_n เป็นเอาต์พุตที่ได้จากการเรียนรู้ด้วยระบบอนุมานนิวโรฟซีแบบปรับตัวได้

จากนั้นเมื่อต้องการทราบค่าความผิดพลาดทั้งหมดของโมเดลที่ได้จะหาด้วยค่าความผิดพลาดกำลังรากที่สอง จากสมการที่ (2.28)

$$E = \sqrt{\frac{\sum E_n}{N}} \quad (2.28)$$

เมื่อ N เป็นจำนวนระเบียนของข้อมูลเรียนรู้

2.5.2 การเรียนรู้แบบย้อนกลับเพื่อปรับพารามิเตอร์ของฟังก์ชันความเป็นสมาชิกด้านอินพุต โดยใช้ค่าความผิดพลาด ภายใต้กฎลูกโซ่ (Chain rule) ด้วยวิธีการเคลื่อนลาดชันลงมา ในการปรับแต่งค่าพารามิเตอร์ต่างๆ นั้น จะใช้ค่าความผิดพลาดกำลังสองเฉลี่ย เพื่อใช้หาค่าพารามิเตอร์ที่เหมาะสมที่สุด หาได้จากสมการที่ (2.28) และค่าอัตราความผิดพลาด $\partial E_n / \partial O_{i,n}$ สำหรับข้อมูลเรียนรู้ n ระเบียน ส่วนเอาต์พุต O คำนวณจากสมการที่ (2.29)

$$\frac{\partial E_n}{\partial O_{i,n}} = -2(T_{i,n} - O_{i,n}) \quad (2.29)$$

อัตราความผิดพลาดของโหนดตำแหน่งที่ i ในชั้นที่ n สามารถหาได้จากสมการที่ (2.30)

$$\frac{\partial E_n}{\partial O_{i,n}} = \sum_{m=1}^N \frac{\partial E_n}{\partial O_{m,n}} \frac{\partial O_{m,n}}{\partial O_{i,n}} \quad (2.30)$$

อนุพันธ์ของพารามิเตอร์ในฟังก์ชันความเป็นสมาชิกด้านอินพุตคำนวณได้ สมการที่ (2.31)

$$\frac{\partial E_n}{\partial \alpha} = \sum O^* \varepsilon_S \frac{\partial E_n}{\partial O^*} \frac{\partial O^*}{\partial \alpha} \quad (2.31)$$

ในการหาอนุพันธ์ของเซตพารามิเตอร์ในฟังก์ชันความเป็นสมาชิกด้านอินพุต (α) นั้นจะใช้กฎลูกโซ่ภายใต้หลักการทำงานตามสถาปัตยกรรมดังสมการที่ (2.32)

$$\frac{\partial E_n}{\partial \alpha} = \frac{\partial E}{\partial O_5} \times \frac{\partial O_5}{\partial O_4} \times \frac{\partial O_4}{\partial O_3} \times \frac{\partial O_3}{\partial O_2} \times \frac{\partial O_2}{\partial O_1} \times \frac{\partial O_1}{\partial \alpha} \quad (2.32)$$

สามารถหาอนุพันธ์ย่อยจากชั้นที่ 5 ไปชั้นที่ 4 ได้ดังสมการที่ (2.33)

$$\frac{\partial O_5}{\partial O_4} = \frac{\partial(\sum f_i \bar{w}_i)}{\partial(f_i \bar{w}_i)} = 1 \quad (2.33)$$

เมื่อ $\bar{w}_i$ คือค่าน้ำหนักที่ทำให้เป็นมาตรฐานเดียวกันของกฎฟuzzyลำดับที่ i

หาอนุพันธ์ย่อยจากชั้นที่ 4 ไปชั้นที่ 3 ได้ดังสมการที่ (2.34)

$$\frac{\partial O_4}{\partial O_3} = \frac{\partial(\sum f_i \bar{w}_i)}{\partial(f_i \bar{w}_i)} = f_i \quad (2.34)$$

เมื่อ f เป็นเซตพารามิเตอร์ที่ได้จากการเรียนรู้แบบไปข้างหน้าในสมการที่ (2.21)

หาอนุพันธ์ย่อยจากชั้นที่ 3 ไปชั้นที่ 2 ได้ดังสมการที่ (2.35)

$$\frac{\partial O_3}{\partial O_2} = \frac{\partial}{\partial w_i} \left(\frac{w_i}{\sum_{k=1}^r w_k} \right) = \frac{\sum_{k=1}^n w_k - w_i}{\left(\sum_{k=1}^n w_k \right)^2} \quad (2.35)$$

เมื่อ r เป็นจำนวนกฎทั้งหมดในชั้นที่ k

หาอนุพันธ์ย่อยจากชั้นที่ 2 ไปชั้นที่ 1 ได้ดังสมการที่ (2.36)

$$\frac{\partial O_2}{\partial O_1} = \frac{\partial}{\partial A_m} \left(\prod_{A_k \in R(A_m), A_k \neq A_m} A_k \right) \quad (2.36)$$

เมื่อ $A_k \in R(A_m)$ เป็นระดับความเป็นสมาชิกที่สร้างกฎส่วนข้อตั้งที่อยู่ในฟuzzyเซต A_m

อัปเดตเซตพารามิเตอร์ได้ดังสมการที่ (2.37) เมื่อใช้ฟังก์ชันความเป็นสมาชิกเกาส์เซียน

$$\frac{\partial O_1(x)}{\partial \alpha} = \frac{(x-c)^2}{\sigma^3} \mu_{A_i}(x) \quad (2.37)$$

ตัวอย่างการอัปเดตเซตพารามิเตอร์ของฟังก์ชันความเป็นสมชกด้านอินพุตเช่น เลือกใช้ฟังก์ชันความเป็นสมชกเกาส์เซียนที่มีการหาอนุพันธ์ย่อยจากสมการที่ (2.38) และ (2.39) ของแต่ละพารามิเตอร์ประกอบด้วย c และ σ

$$\frac{\partial O_1(x)}{\partial \sigma} = \frac{(x-c)^2}{\sigma^3} \mu_{A_i}(x) \quad (2.38)$$

$$\frac{\partial O_1(x)}{\partial c} = \frac{(x-c)^2}{\sigma^2} \mu_{A_i}(x) \quad (2.39)$$

เมื่อได้ค่าอนุพันธ์ที่ได้จากกฎลูกโซ่แล้ว จะนำมาหาค่าพารามิเตอร์ σ ใหม่ในฟังก์ชันความเป็นสมาชิกด้วยสมการที่ (2.40) และ (2.41)

$$\sigma_{new} = \sigma_{old} - \eta \frac{\partial E}{\partial \sigma} \quad (2.40)$$

$$c_{new} = c_{old} - \eta \frac{\partial E}{\partial c} \quad (2.41)$$

เมื่อ σ_{old}, c_{old} เป็นเซตพารามิเตอร์เดิม

η เป็นอัตราการเรียนรู้ที่เปลี่ยนแปลงได้จากค่าขนาดขั้น (Step size) ดังสมการที่ (2.42)

$$\eta = \frac{stepsize}{\sqrt{\sum \alpha \left(\frac{\partial E}{\partial \alpha} \right)^2}} \quad (2.42)$$

2.6 การจำแนกประเภทค่าความน่าจะเป็น

การจำแนกประเภทค่าความน่าจะเป็นคือการจำแนกประเภทกลุ่มข้อมูลที่มีค่าเอาต์พุตเป็นค่าความน่าจะเป็นของแต่ละกลุ่ม และเลือกกลุ่มทำนายจากกลุ่มที่มีค่าความน่าจะเป็นมากที่สุด เทคนิคที่ได้รับ ความนิยมจากการทำเหมืองข้อมูลทางการศึกษา เช่น โครงข่ายรู้จำแบบ การถดถอยโลจิสติก ตัวจำแนกประเภทเบย์อย่างง่าย และต้นไม้ตัดสินใจ [15] จึงเลือกใช้เทคนิคเหล่านี้มาทำนายค่าความน่าจะเป็นของการสำเร็จการศึกษา มีหลักการทำงานดังนี้

2.6.1 โครงข่ายการรู้จำแบบ (Pattern recognition network, PRN) เป็นการจำแนกประเภทกลุ่มข้อมูลจากการรู้จำรูปแบบข้อมูลที่มีการเรียนรู้เพื่อปรับพารามิเตอร์ด้วยโครงข่ายประสาทเทียมแบบไปข้างหน้า 2 ชั้น คือ ชั้นซ่อน (Hidden node) และชั้นเอาต์พุต (Output node) โดยที่ข้อมูลที่ประมวลผลจะถูกส่งไปในทิศทางเดียวจากโหนดอินพุต ส่งต่อมาที่โหนดซ่อน จนถึงโหนดเอาต์พุต โดยไม่มีการย้อนกลับของข้อมูลและไม่มี การเชื่อมต่อกันของโหนดในชั้นเดียวกัน ฟังก์ชันถ่ายโอนชั้นซ่อนคือ ฟังก์ชันซิกมอยด์ (Sigmoid function) และฟังก์ชันถ่ายโอนชั้นเอาต์พุตคือ ฟังก์ชันซอฟต์แมกซ์ (Softmax function) ดังสมการที่ (2.43) ค่าเอาต์พุตที่ได้จึงมีค่าเป็นค่าความน่าจะเป็น

$$P = \frac{e^x}{\sum_{k=1}^K e^x} \quad (2.43)$$

เมื่อ P คือค่าความน่าจะเป็น

x คือเอาต์พุตที่ได้จากชั้นซ่อน

K คือจำนวนข้อมูลทั้งหมด

2.6.2 การจำแนกประเภทเบย์อย่างง่าย (Naive bayes classification, NB) เป็นการจำแนกประเภทกลุ่มข้อมูลจากการใช้ค่าความน่าจะเป็นของข้อมูลก่อนหน้า โดยใช้กฎของเบย์ทำนาย ค่าความน่าจะเป็นของแต่ละกลุ่ม ดังสมการที่ (2.44) เนื่องจากงานวิจัยนี้ข้อมูลที่ใช้เป็นค่าต่อเนื่อง การหาค่าความน่าจะเป็นของข้อมูลก่อนหน้าจึงวัดจากการกระจายความน่าจะเป็นด้วยฟังก์ชันเกาส์เซียน (Gaussian function) ดังสมการที่ (2.45)

$$P(Y = y_k | X_1, \dots, X_n) = \frac{P(Y = y_k) \prod_i P(X_i | Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(X_i | Y = y_j)} \quad (2.44)$$

เมื่อ Y คือกลุ่มของตัวแปรเอาต์พุต

X คือตัวแปรอินพุต

$P(Y = y_k | X_1, \dots, X_n)$ คือความน่าจะเป็นของตัวแปรอินพุตในกลุ่ม y_k

$P(X_i | Y = y_k)$ คือความน่าจะเป็นของกลุ่ม y_k ในตัวแปรอินพุตที่ i

$P(Y = y_k)$ คือความน่าจะเป็นในกลุ่ม y_k

$$P(X_i | Y = y_k) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{1}{2} \left(\frac{x - \mu_{ik}}{\sigma_{ik}} \right)^2} \quad (2.45)$$

เมื่อ μ_{ik} คือค่าเฉลี่ยของตัวแปรอินพุตที่ i อยู่ในกลุ่มที่ k

σ_{ik}^2 คือค่าความแปรปรวนของตัวแปรอินพุตที่ i อยู่ในกลุ่มที่ k

σ_{ik} คือส่วนเบี่ยงเบนมาตรฐานของตัวแปรอินพุตที่ i อยู่ในกลุ่มที่ k

2.6.3 ต้นไม้ตัดสินใจ (Decision tree, DT) เป็นการจำแนกประเภทกลุ่มข้อมูล โดยการแบ่งข้อมูลของแต่ละตัวแปรอินพุตออกเป็นกลุ่มจากการคัดเลือกคุณสมบัติข้อมูลด้วยค่าสัมประสิทธิ์จีนิ (Gini's Diversity Index, gdi) มีขั้นตอนการทำงาน โดยเริ่มจากการคัดเลือกคุณสมบัติที่ดีที่สุดมาเป็นโหนดราก ต่อมาจะสร้างโหนดลูก และวนสร้างโหนดลูกและต้นไม้ย่อยของแต่ละกิ่งไปเรื่อยๆ จนข้อมูลที่ผ่านการแบ่งแยกนั้นจะอยู่ในกลุ่มเดียวกันหรือจำนวนข้อมูลที่ผ่านการแบ่งแยกในกิ่งหนึ่งมีค่าน้อยกว่าค่าที่กำหนด อัลกอริทึมที่ใช้คือ ซีเออาร์ที (CART, Classification and regression tree)

2.6.4 การถดถอยโลจิสติกพหุกลุ่ม (Multinomial logistic regression, MLR) เป็นการหาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามด้วยสมการโลจิสติก โดยจะนำตัวแปรอิสระมาคำนวณหาสมการ จากนั้นนำไปพยากรณ์ตัวแปรตาม ซึ่งจะใช้ในกรณีที่ตัวแปรตามที่มีค่ามากกว่า 2 ค่า ด้วยสมการที่ (2.45)

$$P(Y = y_k | X) = \frac{\exp(\omega_{k0} + \sum_{i=1}^n \omega_{ki} X_i)}{1 + \sum_{j=1}^{R-1} \exp(\omega_{j0} + \sum_{i=1}^n \omega_{ji} X_i)} \quad (2.45)$$

เมื่อ $P(Y = y_k | X)$ คือค่าความน่าจะเป็นของตัวแปรอินพุต X ในกลุ่ม y_k
 X_i คือตัวแปรอินพุตที่ i
 $\omega_{k0}, \omega_{ki}, \omega_{ji}$ คือค่าสัมประสิทธิ์ถดถอยโลจิสติก
 k, R คือจำนวนกลุ่มในตัวแปรเอาต์พุต

2.7 งานวิจัยที่เกี่ยวข้อง

การทำเหมืองข้อมูลเป็นสาขาวิชาการค้นคืนสารสนเทศและการใช้ประโยชน์จากข้อมูลขนาดใหญ่ โดยได้มีการประยุกต์ใช้กับหลายสาขางานเช่น ธุรกิจค้าขาย หุ่น เป็นต้น ในปัจจุบันมีการเพิ่มความน่าสนใจในการใช้งานของการทำเหมืองข้อมูล เพื่อตรวจสอบคำถามทางวิทยาศาสตร์ของงานวิจัยทางการศึกษา เรียกว่า การทำเหมืองข้อมูลทางการศึกษา Baker และ Yacef [7] ได้จำแนกเป้าหมายของงานวิจัยเป็น 4 กลุ่ม คือ 1) เพื่อทำนายพฤติกรรมการเรียนรู้ของนักศึกษาในอนาคตด้วยการสร้างโมเดลนักศึกษาของแต่ละคน โดยใช้ความรู้ของนักศึกษา แรงจูงใจ กระบวนการรู้คิด และทัศนคติ 2) เพื่อค้นหาหรือปรับปรุงโมเดลองค์ประกอบหลักของสถาบันการศึกษา 3) เพื่อศึกษาผลกระทบของการสนับสนุนการเรียนการสอนประเภทต่างๆ จากซอฟต์แวร์การเรียนรู้ และ 4) เพื่อแสดงความก้าวหน้าของความรู้ทางวิทยาศาสตร์ที่เกี่ยวกับการเรียนรู้และผู้เรียนรู้ด้วยการสร้างโมเดลทางคอมพิวเตอร์ ซึ่งเป้าหมายดังกล่าวครอบคลุมทั้งด้านนักศึกษา ด้านสถาบันการศึกษา ด้านซอฟต์แวร์ที่ใช้สนับสนุนการเรียนการสอน และด้านทฤษฎี นอกจากนี้มีการแบ่งเทคนิคที่ใช้ในการสร้างโมเดล 5 กลุ่มดังนี้ การทำนายล่วงหน้า การจัดกลุ่ม การกลั่นกรองข้อมูลเพื่อให้มนุษย์ตัดสินใจ การทำเหมืองความสัมพันธ์ และการพัฒนาโมเดล เทคนิคการทำนายล่วงหน้าเป็นการพัฒนาโมเดลที่สามารถอนุมานตัวแปรที่ถูกทำนายหนึ่งตัวด้วยการรวมกันของตัวแปรทำนายหลายตัวได้ ส่วนมากจะประยุกต์ใช้กับการตรวจสอบพฤติกรรมของนักศึกษา การพัฒนาโมเดลองค์ประกอบหลักของสถาบันการศึกษา การทำนายและการทำความเข้าใจเกี่ยวกับผลการศึกษานักศึกษา โดยมีวิธีสร้างโมเดลดังนี้ การจำแนกประเภท การวิเคราะห์ความถดถอย และการประมาณฟังก์ชันความหนาแน่น ซึ่งวิธีการที่ได้รับความนิยม ได้แก่ ตัดสินใจแบบต้นไม้ การวิเคราะห์ความถดถอยโลจิสติก โครงข่ายประสาทเทียม ชัฟฟอร์ตเวกเตอร์แมชชีน การวิเคราะห์ความถดถอยเชิงเส้น และการประมาณฟังก์ชันเคอร์เนล ปัจจุบันเทคนิคระบบอนุมานนิเวศวิธีแบบปรับตัวได้มีการใช้อย่างกว้างขวางในทางการศึกษาด้านการจำแนกประเภทการประมาณค่า และการทำนายล่วงหน้า [16]

ชลนิศา [17] จำแนกกลุ่มสถานภาพการสำเร็จการศึกษาที่มีตัวแปรอิสระมาจากการแปลงข้อมูลรายวิชาให้อยู่ในรูปของหมวดวิชาและกลุ่มวิชา และตัวแปรตามเป็นจบการศึกษากับไม่จบการศึกษา ด้วยแบบจำลองต้นไม้ตัดสินใจ เปรียบเทียบกับจินตિકอัลกอริทึมที่ร่วมกับโครงข่ายประสาทเทียมผลที่ได้มีค่าใกล้เคียงกัน นอกจากนี้ยังพบว่า นักศึกษาที่ได้เกรดในหมวดหมู่วิชาแกนต่ำมีโอกาสที่จะไม่สำเร็จการศึกษาตามหลักสูตร

บุษบา [18] พยากรณ์โอกาสการสำเร็จการศึกษาของนักศึกษาที่มีตัวแปรอิสระมาจากการแปลงข้อมูลรายวิชาให้อยู่ในรูปของหมวดวิชาและกลุ่มวิชา และมีตัวแปรเอาต์พุตเป็นจบการศึกษากับไม่จบการศึกษา ด้วยการใช้จินตિકอัลกอริทึมแบบหลายวัตถุประสงค์ เปรียบเทียบกับวิธีทางระบบโครงข่ายประสาทเทียม พบว่า จินตિકอัลกอริทึมแบบหลายวัตถุประสงค์ร่วมกับโครงข่ายประสาทเทียมมีประสิทธิภาพการทำนายที่ใกล้เคียงกับการใช้โครงข่ายประสาทเทียมอย่างเดียว

Han Yan [19] ใช้ระบบการอนุมานนิวโรฟัซซีแบบปรับได้ในการจำแนกประเภทสถานะคุณภาพของน้ำในแม่น้ำมีการใช้ฟังก์ชันความเป็นสมาชิกทั้ง 9 แบบ เมื่อเปรียบเทียบประสิทธิภาพของแบบจำลองแล้วพบว่า ฟังก์ชันความเป็นสมาชิกเกาส์เซียนมีประสิทธิภาพมากที่สุด และถูกเลือกให้เป็นแบบจำลองที่เหมาะสมที่สุด และยังมีประสิทธิภาพดีกว่าแบบจำลองจากโครงข่ายประสาทเทียม โดยสามารถสร้างผลลัพธ์ที่อยู่ในรูปแบบของค่าต่อเนื่องซึ่งทำให้การประเมินคุณภาพน้ำมีความเข้าใจมากขึ้น

Inyang [20] มีการประยุกต์ใช้การทำเหมืองข้อมูลทางการศึกษาร่วมกับเทคนิคระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ในการพยากรณ์เกรดเฉลี่ยสะสมของนักศึกษา ซึ่งตัวแปรอิสระที่ใช้เป็นรายวิชาที่ลงทะเบียนเรียนในปีแรกที่เข้าเรียน และตัวแปรตามเป็นเกรดเฉลี่ยสะสม หลังจากนั้นจะนำ เกรดเฉลี่ยสะสมที่ได้มาแบ่งกลุ่มความเสี่ยงด้วยเทคนิคฟัซซีซิมิน ซึ่งจะแบ่งเป็น 3 กลุ่มคือ กลุ่มอ่อน กลุ่มปานกลาง และกลุ่มดี โดยค่าที่ได้ในแต่ละกลุ่มคือค่าความเป็นสมาชิกของกลุ่มนั้น

Rusli [21] ทำนายผลการเรียนของนักศึกษาด้วยเทคนิคระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ เปรียบเทียบกับการถดถอยโลจิสติกและโครงข่ายประสาทเทียม ซึ่งวัดประสิทธิภาพของโมเดลด้วยค่าความคลาดเคลื่อนกำลังสองน้อยสุด พบว่า โมเดลที่ใช้เทคนิคระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ดีกว่าการถดถอยโลจิสติกและโครงข่ายประสาทเทียม

จากการศึกษางานวิจัยดังกล่าวที่เกี่ยวข้องกับการทำนายผลการสำเร็จการศึกษาพบว่า ผลทำนายการสำเร็จการศึกษาอยู่ในลักษณะที่มีความชัดเจน ตัวอย่างเช่น เมื่อตัวแปรเอาต์พุตคือ จบกับไม่จบ ผลลัพธ์ของนักศึกษาแต่ละคนที่ได้จะเป็นจบ หรือไม่จบ แต่งานวิจัยนี้นำเสนอผลลัพธ์ที่เป็นค่าความน่าจะเป็นของแต่ละกลุ่ม ตัวอย่างเช่น ตัวแปรเอาต์พุตคือ จบกับไม่จบ นักศึกษาแต่ละคนจะมี

ความน่าจะเป็นที่จะจบเท่าใด และมีความน่าจะเป็นที่จะไม่จบเท่าใด เมื่อรวมค่าความน่าจะเป็นทั้ง 2 กลุ่ม ต้องเท่ากับ 1 และนอกจากนี้ยังมีการทำนายเกรดเฉลี่ยสะสมด้วย

ตัวแปรอินพุตที่ใช้ในงานวิจัยส่วนใหญ่มีทั้งข้อมูลส่วนตัว เช่น เพศ อาชีพผู้ปกครอง รายได้ผู้ปกครอง เป็นต้น และข้อมูลผลการเรียน เช่น เกรดเฉลี่ยสะสมจากสถาบันเดิม เกรดเฉลี่ยสะสมทั้งหมด เกรดเฉลี่ยสะสมที่แบ่งเป็นหมวดหมู่หรือกลุ่มวิชา เป็นต้น ซึ่งงานวิจัยนี้จะใช้ข้อมูลที่เกี่ยวข้องกับการเรียนในสถาบันปัจจุบัน โดยไม่ได้พิจารณาปัจจัยอื่น จึงมีการศึกษาเอกสารเพื่อหาเงื่อนไขการสำเร็จการศึกษา จากนั้นจะนำมาเป็นตัวแปรอินพุต และจะมีการอธิบายรายละเอียดในส่วนบทต่อไป

จากงานวิจัยของ Rusli ที่มีการเปรียบเทียบการทำนายเกรดเฉลี่ยสะสมด้วยเทคนิคระบบการอนุมาน นิวโรฟัซซีแบบปรับได้ การถดถอยโลจิสติกและ โครงข่ายประสาทเทียม โดยที่เทคนิคระบบการอนุมานนิวโรฟัซซีแบบปรับได้ให้ผลลัพธ์ที่ดีที่สุด นอกจากยังมีการทำนายผลการสำเร็จการศึกษาด้วยเทคนิคจินติกอัลกอริทึม โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ แต่งานวิจัยนี้ต้องการผลลัพธ์ที่เป็นค่าความน่าจะเป็น จึงศึกษาเทคนิคการจำแนกประเภทที่มีค่าเอาต์พุตเป็นค่าความน่าจะเป็น และเทคนิคที่นิยมใช้ในการทำเหมืองข้อมูลทางศึกษา คือ โครงข่ายรู้จำแบบ การถดถอยโลจิสติก ตัวจำแนกประเภทเบย์อย่างง่าย และต้นไม้ตัดสินใจ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved