บทที่ 1

บทนำ

การวินิจฉัยโรคและฐานข้อมูลทางการแพทย์ในปัจจุบันได้มีข้อมูลเพิ่มขึ้นเป็นจำนวนมาก ส่งผลให้แพทย์เข้าใจหรือวิเคราะห์ข้อมูลเหล่านั้นได้ยาก เพื่อให้แพทย์สามารถเข้าใจข้อมูลเหล่านั้นได้ง่าย โดยเฉพาะอย่างยิ่งข้อมูลที่มีลักษณะเป็นตัวเลข นอกจากนี้แล้วเพื่อให้ประหยัดเวลาขึ้น จึงต้องมีการนำข้อมูลเหล่านั้นมาค้นหาความรู้ หรือสกัดกฎที่แพทย์เข้าใจได้ โดยที่กฎอยู่ในลักษณะ “ถ้า-แล้ว” (“If-Then” Rules) มี 2 แบบคือกฎทั่วไป (Crisp Rules/ Conventional Rules) ซึ่งแทนด้วยค่าจำนวนจริง และกฎภาษาธรรมชาติ (Linguistic Rules) ซึ่งแทนด้วยภาษาธรรมชาติ เช่น เล็ก กลาง และใหญ่ เป็นต้น ทั้ง 2 แบบมีองค์ประกอบ 2 ส่วน โดยที่ส่วนแรกที่อยู่หลัง “ถ้า” หรือ “If” จะหมายถึงเหตุ และส่วนที่สองที่อยู่หลัง “แล้ว” หรือ “Then” หมายถึงผลลัพธ์ที่ตามมา รายละเอียดของแต่ละแบบมีดังนี้ [1, 2]

1. กฎทั่วไป (Crisp Rules/Coventional Rules) มีรูปแบบดังสมการ (1.1)

$$\text{If } x_1 \text{ op } t_1 \text{ and/or } \ldots \text{ and/or } x_i \text{ op } t_j \text{ then } c_k \qquad (1.1)$$

โดยที่ x_i คือตัวแปรข้อมูลเข้าตัวที่ i

t_j คือค่าของตัวแปรข้อมูลเข้าที่เป็นจำนวนจริงที่ j

op คือตัวดำเนินการ =, <, >, <=, >= หรือ ≠

c_k คือกลุ่มข้อมูล (Class) ที่ k

ตัวอย่างกฎทั่วไปจากฐานข้อมูลโรคมะเร็งเต้านม (Wisconsin Breast Cancer Database) เช่น If x_2 (Uniformity of Cell Size) < 6 and x_4 (Marginal Adhesion) < 3 and x_8 (Mitoses) < 8 then Malignant (เป็นโรคมะเร็งเต้านม) [3] เป็นต้น

2. กฎภาษาธรรมชาติ (Linguistic Rules) มีรูปแบบดังสมการ (1.2)

$$\text{If } x_1 \text{ is } l_1 \text{ and/or } \ldots \text{ and/or } x_i \text{ is } l_j \text{ then } c_k \qquad (1.2)$$

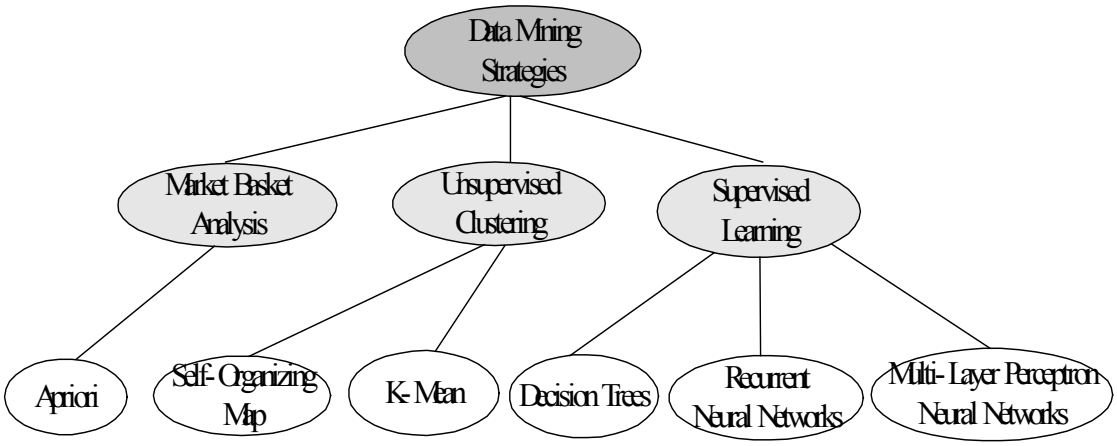
โดยที่ x_i คือตัวแปรข้อมูลเข้าตัวที่ i

l_j คือค่าของตัวแปรข้อมูลเข้าที่ j มีลักษณะเป็นรูปแบบภาษาธรรมชาติ (Linguistic Terms) เช่น เล็ก กลาง และใหญ่ เป็นต้น

c_k คือกลุ่มข้อมูล (Class) ที่ k

ตัวอย่างกฎภาษาธรรมชาติจากฐานข้อมูลดอกไม้ไอริช (Iris Flower Database) เช่น If x_3 (Petal Length) is small and x_4 (Petal Width) is small then Vericolor [3] เป็นต้น

การทำเหมืองข้อมูล (Data Mining) หมายถึงวิธีการที่ใช้เทคนิคการเรียนรู้ของคอมพิวเตอร์เพื่อวิเคราะห์ค้นหา และสกัดความรู้หรือกฎโดยอัตโนมัติ จากข้อมูลที่อยู่ใน ฐานข้อมูลจำนวนมาก [4, 5] ซึ่งวิธีการในการทำเหมืองข้อมูลถูกแบ่งออกเป็น 3 ประเภทใหญ่ๆ คือ การวิเคราะห์ทางการตลาด (Market Basket Analysis) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Clustering) และการเรียนรู้แบบมีผู้สอน (Supervised Learning) ดังภาพประกอบ 1.1 [5]



ภาพประกอบ 1.1 วิธีการในการทำเหมืองข้อมูล

1. การวิเคราะห์ทางการตลาด (Market Basket Analysis) เป็นการหา กฎความสัมพันธ์ (Association Rules) จากพฤติกรรมการซื้อสินค้าของลูกค้า ตัวอย่างเช่น ถ้าลูกค้าซื้อผ้าอ้อมแล้วลูกค้ามีโอกาสที่จะซื้อเบียร์ด้วยค่าสนับสนุน(Support) และค่าความเชื่อมั่น (Confidence) เป็นเท่าใด สามารถเขียนเป็นกฎความสัมพันธ์ได้คือ {ผ้าอ้อม} $\rightarrow$ {เบียร์} [23]

เป็นต้น โดยมีการนำกฎความสัมพันธ์มาปรับปรุงการตลาดให้สามารถขายสินค้าได้เพิ่มขึ้น หรือวางชั้นสินค้าได้เหมาะสมขึ้น ตัวอย่างอัลกอริทึมที่ใช้ในการหากฎความสัมพันธ์ เช่น อัลกอริทึม Apriori [6]

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Clustering) ระบบจะทำการเรียนรู้เพื่อแบ่งกลุ่มข้อมูลเอง ตัวอย่างเทคนิคที่อยู่ในประเภทนี้ได้แก่ แผนที่การจัดกลุ่มเอง (Self-Organizing Map: SOM) และการแบ่งกลุ่ม K กลุ่มด้วยค่าเฉลี่ย (K-Mean) [5] เป็นต้น

3. การเรียนรู้แบบมีผู้สอน (Supervised Learining) ระบบจะทำการเรียนรู้เพื่อแบ่งกลุ่มข้อมูลจากชุดข้อมูลตัวอย่างที่ใช้ในการสอน ตัวอย่างเทคนิคที่มีการเรียนรู้แบบมีผู้สอน เช่น ต้นไม้การตัดสินใจ (Decision Trees) โครงข่ายประสาทเทียมแบบย้อนกลับ (Recurrent Neural Networks) และโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (Multilayer Perceptron Neural Networks: MLP) เป็นต้น

โครงข่ายประสาทเทียมถูกสร้างขึ้นเพื่อเลียนแบบการทำงานของสมองมนุษย์ ซึ่งมีการทำงานแบบขนานจำนวนมาก หลักการทำงานประกอบด้วยหน่วยประมวลผลย่อยหลายๆ หน่วยทำงานเชื่อมต่อกัน แต่ละหน่วยสามารถปรับค่าพารามิเตอร์ประจำหน่วยได้จากกระบวนการเรียนรู้ โครงข่ายประสาทเทียมที่ผ่านการฝึกฝนจะสามารถนำไปใช้แก้ปัญหาจริงได้ [7] ข้อดีของโครงข่ายประสาทเทียมคือ สามารถทำนายได้ค่าความถูกต้องสูง เหมาะกับข้อมูลที่มีลักษณะเป็นตัวเลข [5] โดยที่ข้อมูลไม่จำเป็นต้องมีการกระจาย อีกทั้งทนทานต่อการทำงานล้มเหลว (Fault Tolerance) [8] แต่อย่างไรก็ตามเป็นการยากมากที่จะอธิบายการทำงานของโครงข่ายประสาทเทียมให้ผู้ใช้เข้าใจได้ ดังนั้นผู้ใช้จะไม่สามารถเข้าใจได้ว่าโครงข่ายประสาทเทียมมีการตัดสินใจในการทำนายอย่างไร เพื่อแก้ไขปัญหานี้จึงได้คิดค้นการสกัดกฎจากโครงข่ายประสาทเทียมเพื่อให้ ผู้ใช้สามารถเข้าใจได้ง่าย [9, 10] วิทยานิพนธ์ฉบับนี้ได้นำเสนอแบบจำลองโดยใช้เทคนิค โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้นในการสกัดกฎสำหรับฐานข้อมูล ทางการแพทย์ โดยกฎที่ได้จะอยู่ในลักษณะ “ถ้า-แล้ว” ในลักษณะภาษาธรรมชาติที่แพทย์สามารถเข้าใจได้ง่าย

1.1 การตรวจเอกสาร

เทคนิคที่ใช้ในการสกัดกฎมี 4 เทคนิคคือ โครงข่ายประสาทเทียม (Neural Networks) ต้นไม้การตัดสินใจ (Decision Trees) ฟัซซี่เซต (Fuzzy Set) และขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) รายละเอียดของแต่ละวิธีมีดังนี้

1.1.1 โครงข่ายประสาทเทียม (Neural Networks)

แนวทางการสกัดกฎจากโครงข่ายประสาทเทียมมี 2 ประเภทคือ แนวทางกล่องเปิด (Open Box) และแนวทางกล่องดำ (Black Box) [2, 11]

1. แนวทางกล่องเปิด (Open-Box/Nonsearch-Based) จะสกัดกฎด้วยการแปลความหมายจากค่าน้ำหนัก (Weight) ที่ได้จากการเชื่อมต่อภายในโครงข่ายประสาทเทียมที่ได้มาจากการสอน ตัวอย่างวิธีการที่ใช้แนวทางกล่องเปิดในการสกัดกฎ คือวิธีการ Knowledgetron (KT Method) [12] วิธีการสกัดกฎโดยใช้ค่าความเชื่อมั่น (Rule Extraction Certainty Factor: RECF) [9] วิธีการเปลี่ยนเพอร์เซฟตรอนหลายชั้นไปเป็นโครงข่ายตรรกะ (Changing MLP into Logical Network: C-MLP2LN) [3] วิธีการแปลค่าแบบเต็มหน่วยกับเพอร์เซฟตรอนหลายชั้น (Discretized Interpretable Multilayer Perceptron: DIMLP) [1] เป็นต้น

2. แนวทางกล่องดำ (Black-Box/Search-Based) จะใช้หลักการค้นหา ซึ่งปกติแล้วจะมีการคำนวณที่สลับซับซ้อนมาก ทำให้เวลาที่ใช้ในการประมวลผลเพิ่มขึ้นตามจำนวนข้อมูลเข้าและโหนดซ่อน ซึ่งแนวทางนี้ต้องการใช้คอมพิวเตอร์ที่มีความสามารถในการประมวลผลสูง ตัวอย่างวิธีการที่ใช้แนวทางกล่องปิดในการสกัดกฎ คือ วิธีการ REX (Rule Extractor) [13] การจัดกลุ่มข้อมูลด้วยขั้นตอนวิธีเชิงพันธุกรรม (Clustering Genetic Algorithm) [14] เป็นต้น

โครงข่ายประสาทเทียมที่นำมาใช้ในการสกัดกฎ โดยแบ่งตามสถาปัตยกรรม มี 3 แบบคือ แผนที่การจัดกลุ่มเอง (Self-Organizing Map: SOM) โครงข่ายประสาทเทียมแบบ ย้อนกลับ (Recurrent Neural Networks) และโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอน หลายชั้น (Multilayer Perceptron Neural Networks: MLP) มีรายละเอียดดังต่อไปนี้

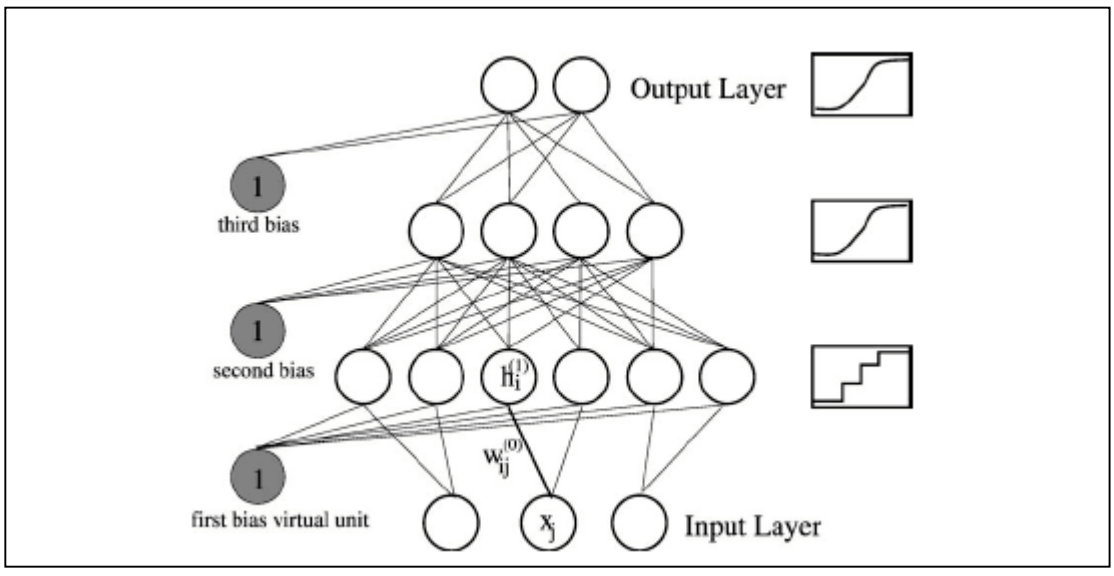
1. แผนที่การจัดกลุ่มเอง (Self-Organizing Map: SOM) เป็นโครงข่ายประสาทเทียมที่มีการเรียนรู้แบบไม่มีผู้สอน โดยจะมีการจัดกลุ่มข้อมูลที่มีลักษณะเหมือนกันให้อยู่ในกลุ่มเดียวกัน โดยการลดมิติของข้อมูลเข้า (Dimensional Input Spaces) ให้มีจำนวนมิติเป็น 2 มิติ [15] เป็นต้น

2. โครงข่ายประสาทเทียมแบบย้อนกลับ (Recurrent Neural Networks) เป็นโครงประสาทเทียมที่มีการเรียนรู้แบบมีผู้สอน โดยสัญญาณจากข้อมูลออก (Output) สามารถวนกลับไปเป็นข้อมูลเข้า (Input) ได้ [16] ข้อเสียของโครงข่ายประสาทเทียมแบบย้อนกลับ คือใช้เวลาในการประมวลผลนาน เมื่อเทียบกับโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น [3]

3. โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (Multilayer Perceptron Neural Networks: MLP) เป็นโครงประสาทเทียมที่มีการเรียนรู้แบบมีผู้สอน นิยมใช้ในการแบ่งกลุ่มข้อมูล ในการสกัดกฎจากโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น มีทั้งได้กฎทั่วไป (Crisp Rules/Conventional Rules) และกฎภาษาธรรมชาติ (Linguistic Rules)

ตัวอย่างวิธีการที่สกัดกฎที่ได้กฎทั่วไป (Crisp Rules/Conventional Rules) เช่น วิธีการสกัดกฎโดยใช้ค่าความเชื่อมั่น (Rule Extraction Certainty Factor: RECF) [9] และ วิธีการแปลค่าแบบเต็มหน่วยกับเพอร์เซฟตรอนหลายชั้น (Discretized Interpretable Multilayer Perceptron: DIMLP) [1] เป็นต้น

วิธีการแปลค่าแบบเต็มหน่วยกับเพอร์เซฟตรอนหลายชั้น (Discretized Interpretable Multilayer Perceptron: DIMLP) เป็นวิธีการสกัดกฎที่นำมาใช้ในการแก้ปัญหา ข้อมูลด้านการแพทย์ ในการวินิจฉัยโรค เช่น โรคตับ โรคมะเร็งเต้านม และโรคหัวใจ เป็นต้น [1] วิธีการแปลค่าแบบเต็มหน่วยกับเพอร์เซฟตรอนหลายชั้นมีโครงสร้างประกอบด้วยชั้นข้อมูลเข้า (Input Layer) 1 ชั้น ชั้นซ่อน (Hidden Layer) 1 ชั้นหรือมากกว่า 1 ชั้นก็ได้ และชั้นผลลัพธ์ (Output Layer) 1 ชั้น ฟังก์ชันกระตุ้นที่ใช้ระหว่างชั้นข้อมูลเข้าและชั้นซ่อนคือฟังก์ชันกระตุ้นแบบขั้นบันได (Staircase Activation Function) ฟังก์ชันกระตุ้นระหว่างชั้นอื่นๆ ใช้ฟังก์ชัน ซิกมอยด์ (Sigmoid Function) ตัวอย่างโครงข่ายประสาทเทียมที่มีชั้นซ่อน 2 ชั้นแสดงดังภาพประกอบ 1.2 [1] โดยมีการกำหนดเส้นตรงในการแบ่งกลุ่มข้อมูล (Discriminant Hyper-Planes) จากข้อมูลชุดสอน (Training Set) และสร้างต้นไม้การตัดสินใจจากค่าน้ำหนักที่ได้จากการสอนโครงข่ายประสาทเทียม และมีการตัดกฎที่ไม่จำเป็นทิ้งไป



ภาพประกอบ 1.2 โครงสร้างโครงข่ายประสาทเทียมของวิธีการ DIMLP

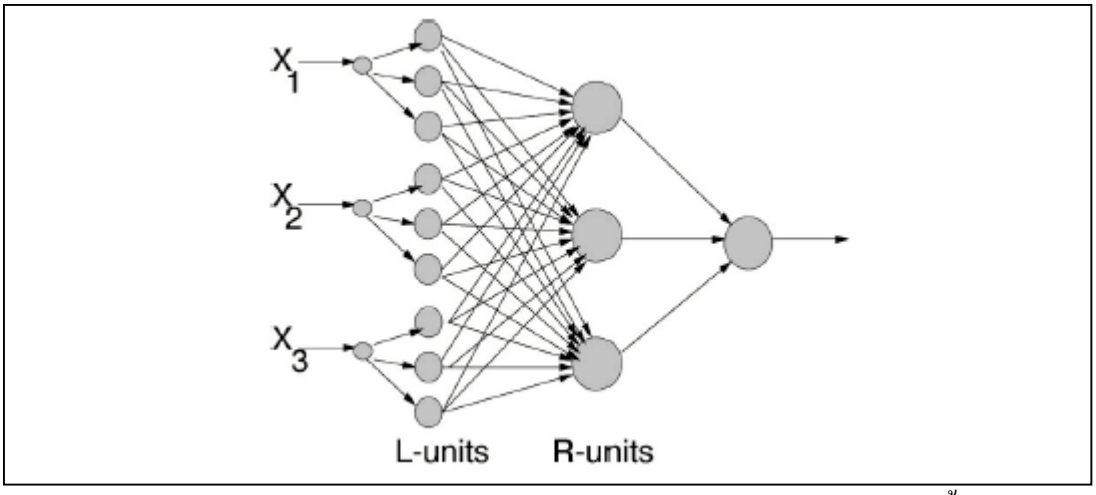
การสกัดกฎที่ได้กฎภาษาธรรมชาติ (Linguistic Rules) จากโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น Duch Huang และ Fan [3, 8, 17] ได้แปลงข้อมูลที่มีลักษณะเป็น

เลขจำนวนจริงให้มีลักษณะเป็นรูปแบบภาษาธรรมชาติ (Linguistic Terms) ก่อนที่จะทำการสอนโครงข่ายประสาทเทียม จากนั้นจึงนำน้ำหนักที่ได้จากการสอนโครงข่ายประสาทเทียมมาสกัดกฎการสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมที่ต้องมีการแทนรูปแบบ ภาษาธรรมชาติในโครงข่ายประสาทเทียม ถ้ามีจำนวนภาษาธรรมชาติน้อยจะสามารถสกัดกฎ ภาษาธรรมชาติได้อย่างมีประสิทธิภาพ แต่ในทางตรงกันข้ามถ้าจำนวนภาษาธรรมชาติมีมากจะทำให้จำนวนโหนดข้อมูลเข้ามีมากด้วย ส่งผลให้สกัดกฎภาษาธรรมชาติได้ประสิทธิภาพต่ำ [17] กล่าวคือใช้เวลาในการประมวลผลนาน จำนวนกฎและตัวเชื่อมของเงื่อนไขมาก [18] เพื่อทำให้เวลาที่ใช้ในการสกัดกฎภาษาธรรมชาติเร็วขึ้น Nii [19] ได้เสนอการสกัดกฎภาษาธรรมชาติโดยใช้คอมพิวเตอร์หลายเครื่องมาช่วยกันประมวลผลแบบขนานกัน (Parallel Computers) ซึ่งถึงแม้ว่าจะสกัดกฎภาษาธรรมชาติได้เร็วขึ้น แต่มีค่าใช้จ่ายสูงขึ้นตามจำนวนเครื่องคอมพิวเตอร์ที่ใช้ [17] นอกจากนี้การตัดโหนดในโครงข่ายประสาทเทียมเป็นอีกแนวทางหนึ่งในการแก้ปัญหานี้ ตัวอย่างเช่น Nii [18] ได้เสนออัลกอริทึมการเลือกตัวแปรข้อมูลเข้าด้วยการนับจำนวนข้อมูลที่แบ่งได้ในแต่ละกลุ่มที่ได้จากการสอนโครงข่ายประสาทเทียม และ Setiono [20] ได้เสนออัลกอริทึมตัดโหนดในโครงข่ายประสาทเทียมโดยใช้ฟังก์ชั่นการประมาณค่า (Neural Network Pruning for Function Approximation : N2PFA) ซึ่งพบว่าได้ค่าความผิดพลาดในการสอนโครงข่าย ประสาทเทียมลดลง แต่ต้องมีการสอนโครงข่ายประสาทเทียมซ้ำหลายครั้ง โดยจำนวนครั้งในการสอนโครงข่ายประสาทเทียมขึ้นอยู่กับจำนวนตัวแปรข้อมูลเข้า และโหนดซ่อนที่ตั้งไว้ [21]

ตัวอย่างวิธีการที่สกัดกฎที่ได้กฎภาษาธรรมชาติ (Linguistic Rules) เช่น วิธีการเปลี่ยนเพอร์เซฟตรอนหลายชั้นไปเป็นโครงข่ายตรรกะ (Changing MLP into Logical Network: C-MLP2LN) [3] เป็นต้น

วิธีการเปลี่ยนเพอร์เซฟตรอนหลายชั้นไปเป็นโครงข่ายตรรกะ (Changing MLP into Logical Network: C-MLP2LN) [3] เป็นวิธีการสกัดกฎที่ใช้โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้นมาช่วยในการสกัดกฎที่มีโครงสร้างดังภาพประกอบ 1.3 ประกอบด้วยชั้นข้อมูลเข้า (Input Layer) 1 ชั้น ซึ่งแทนหน่วยของรูปแบบภาษาธรรมชาติ (L-units) ชั้นซ่อน 1 ชั้น (Hidden Layer) ซึ่งแทนหน่วยของกฎ (R-units) และชั้นผลลัพธ์ 1 ชั้น (Output Layer) โหนดข้อมูลเข้าถูกแบ่งออกเป็น L โหนด ตามจำนวนรูปแบบภาษาธรรมชาติ (Linguistic Terms) เพื่อแทนภาษาธรรมชาติที่ต้องการ เช่นแบ่งตัวแปรข้อมูลเข้า x_1 เป็น 3 โหนดคือ [x_{11} x_{12} x_{13}] แทนตำแหน่งด้วย [เล็ก กลาง ใหญ่] ถ้าค่าข้อมูลเข้ามีค่าตรงกับค่าของโหนดข้อมูลเข้าจะแทนเป็น 1 ถ้าไม่ตรงจะแทนเป็น -1 ดังนั้น การแทนข้อมูลเล็กจะแทนด้วย [1 -1 -1] การแทนข้อมูลกลางจะแทนด้วย [-1 1 -1]

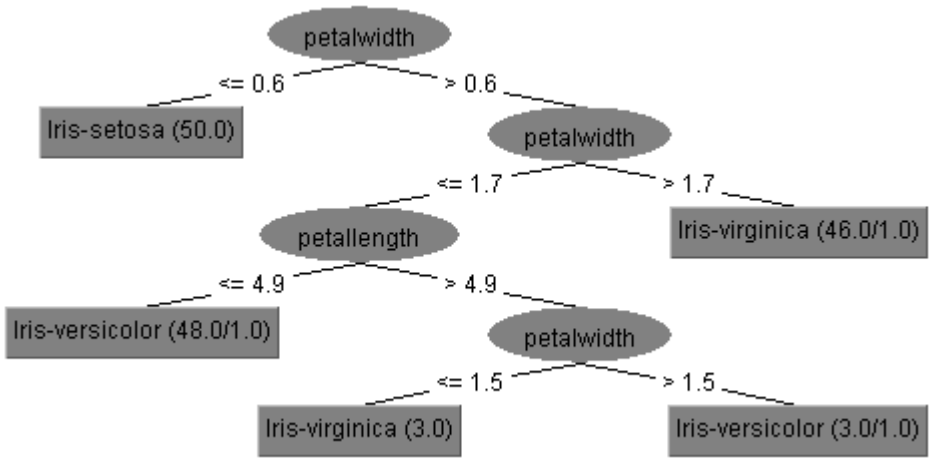
และการแทนข้อมูลใหญ่จะแทนด้วย [-1 -1 1] เป็นต้น จากนั้นทำการสอนโครงข่ายประสาทเทียมและนำน้ำหนักมาสร้างเป็นกฎ



ภาพประกอบ 1.3 โครงสร้างโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้นของวิธีการ C-MLP2LN

1.1.2 ต้นไม้การตัดสินใจ (Decision Trees)

ต้นไม้การตัดสินใจเป็นการเรียนรู้ที่ใช้การแทนความรู้ให้อยู่ในรูปของต้นไม้ การตัดสินใจ ใช้สำหรับจำแนกประเภทของข้อมูล [22] วิธีการสกัดกฎที่ใช้ต้นไม้การตัดสินใจ เช่น วิธีการ C4.5 และ วิธีการ J48 เป็นต้น ซึ่งวิธีการ J48 ได้พัฒนามาจากวิธีการ C4.5 [6] ในการทำงานของวิธีการ C4.5 จะมีการสร้างต้นไม้การตัดสินใจจากบนลงล่าง โดยเริ่มจากการเลือก ตัวแปรข้อมูลเข้า (Input Attributes) ที่สามารถแบ่งกลุ่มได้ดีที่สุดในขณะนั้นมาสร้างเป็น โหนดราก เมื่อข้อมูลผ่านการแบ่งแยกที่โหนดรากตามค่าตัวแปรข้อมูลเข้าของโหนดรากแล้ว ก็จะหาตัวแปรข้อมูลเข้าที่ดีที่สุดของข้อมูลที่ผ่านการแบ่งแยกนั้นมาสร้างเป็นโหนดลูกของโหนดรากนั้นต่อไป และจะวนสร้างโหนดลูกและต้นไม้ย่อยของแต่ละกิ่งไปเรื่อยๆ จนกว่าข้อมูลที่ผ่านการแบ่งแยกนั้นจะจัดอยู่ในกลุ่มเดียวกัน หรือจำนวนข้อมูลที่ผ่านการแบ่งแยกในกิ่งหนึ่งๆ มีค่าน้อยกว่าค่าที่กำหนดไว้ [23] ตัวอย่างต้นไม้การตัดสินใจของฐานข้อมูลดอกไม้ไอริชจากวิธีการ J48 [24] แสดงดังภาพประกอบ 1.4



ภาพประกอบ 1.4 ตัวอย่างต้นไม้การตัดสินใจของฐานข้อมูลดอกไม้ไอริช

1.1.3. ฟัซซี่เซต (Fuzzy sets)

ฟัซซี่เซตเป็นแนวทางในการแสดงถึงลักษณะความคลุมเครือ หรือไม่ชัดเจนของข้อมูล [25] ซึ่งจะมีการกำหนดขอบเขตในการแบ่งความเป็นสมาชิกและไม่เป็นสมาชิกของข้อมูลที่เราสนใจโดยใช้ฟังก์ชันความเป็นสมาชิก (Membership Fuction) [26] เทคนิคฟัซซี่เซตจะนำมาใช้ในการแปลงข้อมูลเดิมที่มีลักษณะเป็นตัวเลขจำนวนจริงให้มีลักษณะเป็นรูปแบบภาษา ธรรมชาติ (Linguistic Terms) เช่น เล็ก กลาง และใหญ่ กฎที่เกิดจากการสกัดกฎโดยใช้เทคนิคฟัซซี่เซตเข้าร่วมด้วยจะเรียกว่ากฎฟัซซี่ (Fuzzy Rules) มีลักษณะเหมือนกับกฎภาษาธรรมชาติ (Linguisitc Rules) มีการนำเทคนิคฟัซซี่เซตมาสกัดกฎร่วมกับเทคนิคโครงข่ายประสาทเทียม พบว่าได้กฎที่กระชับและมีความแม่นยำสูง ตัวอย่างเช่น อัลกอริทึมการสกัดกฎโดยใช้ฟัซซี่เซต (Rule Extraction Fuzzy Set : REFS) [27] อัลกอริทึมกฎนิวโร (NeuroRule Algorithm) [28] และ วิธีการนิวโร-ฟัซซี่สำหรับแบ่งกลุ่มข้อมูล (A Neuro-Fuzzy Approach for The Classification of Data: NEFCLASS) [29] เป็นต้น

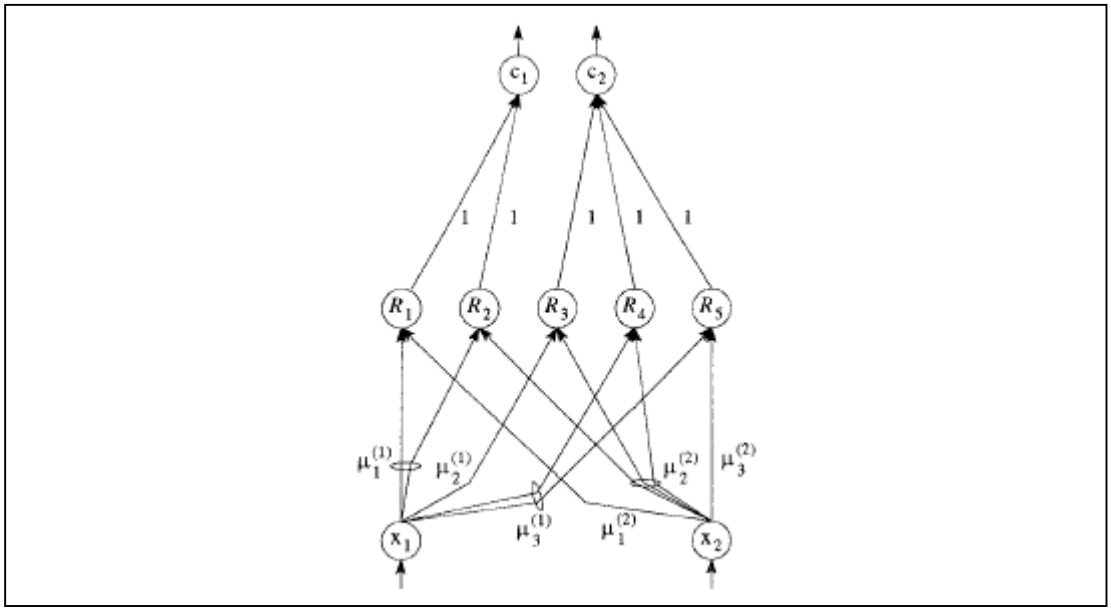
วิธีการนิวโร-ฟัซซี่สำหรับแบ่งกลุ่มข้อมูล (A Neuro-Fuzzy Approach for The Classification of Data: NEFCLASS) [29] รูปแบบของกฎมีดังสมการที่ (1.3) และมีการสกัดกฎจากโครงข่ายประสาทเทียมที่มีโครงสร้างดังภาพประกอบ 1.5

$$\text{If } x_1 \text{ is } \mu_1 \text{ and } x_2 \text{ is } \mu_2 \text{ and } \ldots \text{ and } x_i \text{ is } \mu_j$$
$$\text{Then the pattern } (x_1, x_2, \ldots, x_i) \text{ belong to } c_k \qquad (1.3)$$

โดยที่ x_i คือตัวแปรข้อมูลเข้าตัวที่ i

μ_j คือรูปแบบภาษาธรรมชาติที่ j

c_k คือกลุ่มข้อมูล (Class) ที่ k



ภาพประกอบ 1.5 โครงสร้างโครงข่ายประสาทเทียมของวิธีการ NEFCLASS

1.1.4 ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm)

ขั้นตอนวิธีเชิงพันธุกรรมเป็นการเรียนรู้ที่จำลองการวิวัฒนาการของสิ่งมีชีวิต เช่น การไขว้เปลี่ยนของโครโมโซม (Chromosome Crossover) และการกลายพันธุ์ของยีน (Gene Mutation) ดังนั้นขั้นตอนวิธีเชิงพันธุกรรมเป็นกระบวนการค้นหาประเภทหนึ่ง ขั้นตอนวิธีเชิงพันธุกรรมได้ถูกขยายขึ้นเป็นการโปรแกรมเชิงพันธุกรรม (Genetic Programming) แม้ว่ากฎที่เกิดจากขั้นตอนวิธีเชิงพันธุกรรมจะมีค่าความแม่นยำที่สูง แต่ก็ใช้เวลานานใน การประมวลผล [18] มีการนำขั้นตอนวิธีเชิงพันธุกรรมมาสกัดกฎร่วมกับเทคนิคโครงข่ายประสาทเทียม เช่นการจัดกลุ่มข้อมูลด้วยขั้นตอนวิธีเชิงพันธุกรรม (Clustering Genetic Algorithm) [14] เป็นต้น

1.2 วัตถุประสงค์ของโครงการ

1.2.1 เพื่อสร้างแบบจำลองในการสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม

1.2.2 พัฒนาโปรแกรมสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม

1.3 ขอบเขตของการดำเนินงาน

1.3.1 พัฒนาโปรแกรมเพื่อสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น

1.3.2 ข้อมูลที่นำมาสกัดกฎเป็นข้อมูลที่ได้มาจากฐานข้อมูลทางการแพทย์โดยแบ่งเป็น 2 กลุ่มดังนี้

1. กลุ่มฐานข้อมูลทางการแพทย์ที่เป็นมาตรฐานเป็นที่ยอมรับในระดับสากล ซึ่งทำการดาวน์โหลดมาจาก University of California at Irvine (UCI) [30]
2. กลุ่มฐานข้อมูลทางด้านการแพทย์ที่เก็บตัวอย่างมาจากโรงพยาบาลในประเทศไทย [39]

1.4 ขั้นตอนและระยะเวลาการดำเนินงาน

1.4.1 ขั้นตอนการดำเนินงาน

1. ศึกษางานวิจัยและเอกสารที่เกี่ยวข้อง
2. ศึกษาเทคนิคการสกัดกฎด้วยโครงข่ายประสาทเทียม และเทคนิคอื่นๆ ที่ใช้ในงานวิจัย
3. ศึกษาเครื่องมือและซอฟต์แวร์สำหรับทำงานวิจัย
4. พัฒนาอัลกอริทึมและสร้างแบบจำลองที่สามารถสกัดกฎสำหรับฐาน ข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม
5. เก็บรวบรวมข้อมูลทางการแพทย์จากโรงพยาบาลในประเทศไทย
6. พัฒนาโปรแกรมเพื่อสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม
7. ทดสอบโปรแกรม
8. เขียนผลงานวิจัยและนำเสนองานวิจัย
9. จัดทำเอกสารวิทยานิพนธ์

1.4.2 ระยะเวลาดำเนินการ

ตุลาคม 2547 – มีนาคม 2549

1.4.3 แผนการดำเนินการวิจัย

ตารางที่ 1.1 แสดงระยะเวลาดำเนินการวิจัย

กิจกรรมขั้นตอนการดำเนินงาน	เดือน																	
	2547			2548												2549		
	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
1. ศึกษางานวิจัยและเอกสารที่เกี่ยวข้อง																		
2. ศึกษาเทคนิคการสกัดกฎด้วยโครงข่ายประสาทเทียม และเทคนิคอื่นๆ ที่ใช้ในงานวิจัย																		
3. ศึกษาเครื่องมือและซอฟต์แวร์สำหรับทำงานวิจัย																		
4. พัฒนาอัลกอริทึมและสร้างแบบจำลองที่สามารถสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม																		
5. เก็บรวบรวมข้อมูลทางการแพทย์จากโรงพยาบาลในประเทศไทย																		
6. พัฒนาโปรแกรม เพื่อสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม																		
7. ทดสอบโปรแกรม สรุป และประเมินกฎ																		
8. เขียนผลงานวิจัย																		
9. จัดทำเอกสารวิทยานิพนธ์และสอบวิทยานิพนธ์																		

1.5 สถานที่และเครื่องมือที่ใช้

1.5.1 สถานที่

ห้องปฏิบัติการคอมพิวเตอร์ M105 ภาควิชาวิทยาการคอมพิวเตอร์ ตึกคณิตศาสตร์ และห้องวิจัยปัญญาประดิษฐ์ CS207 ภาควิชาวิทยาการคอมพิวเตอร์ ตึกวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

1.5.2 เครื่องมือที่ใช้

ฮาร์ดแวร์

เครื่องไมโครคอมพิวเตอร์หน่วยความจำ 256 เมกะไบต์ ฮาร์ดดิสก์ 40 กิกกะไบต์หน่วยประมวลผลกลางรุ่นเพนเทียมอาร์โฟร์ 3.00 กิกกะเฮิร์ต จำนวน 1 เครื่อง

ซอฟต์แวร์

1. Microsoft Windows XP เป็นระบบปฏิบัติการ
2. MATLAB 7.0 สำหรับพัฒนาโปรแกรมจากแบบจำลอง และทดสอบแบบจำลอง
3. Java (TM) 2 SDK, Standard Edition สำหรับประมวลผลโปรแกรม Weka
4. Weka สำหรับเปรียบเทียบผลการทดลอง
5. Microsoft Excel 2000 สำหรับเตรียมข้อมูล

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 ได้แบบจำลองในการสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น

1.6.2 ได้โปรแกรมเพื่อสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้ โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น