

บทที่ 6

บทสรุปและข้อเสนอแนะ

วิทยานิพนธ์นี้เสนอแบบจำลองในการสกัดกฎสำหรับฐานข้อมูลทางการแพทย์ โดยใช้โครงข่ายประสาทเทียม ซึ่งมีชื่อว่าแบบจำลองการสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมที่ผ่านการตัดโหนดโดยใช้การแทนค่าความถี่ของช่วงข้อมูล (Rule Extraction form Neural Networks Pruning using Frequency Interval Data Representation : RENN-P-FIDR) และพัฒนาโปรแกรมสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียมจากแบบจำลองที่ได้นำเสนอ ซึ่งโปรแกรมที่พัฒนาขึ้นผู้ใช้สามารถใช้งานได้ง่าย เนื่องจากมีการพัฒนาโปรแกรมเชื่อมต่อกับผู้ใช้ทางกราฟฟิก (Graphic User Interface)

6.1 สรุปผลการวิจัย

งานวิจัยนี้ได้บรรลุตามวัตถุประสงค์โดยสร้างแบบจำลองในการสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม ซึ่งประกอบด้วยการทำงาน 2 ขั้นตอนหลัก ดังนี้

ขั้นตอนที่หลักที่ I การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (Neural Network Pruning using Maximum Feature Weighting: NNP-MFW) เป็นการเตรียมข้อมูลก่อนที่จะสกัดกฎภาษาธรรมชาติ เนื่องจากต้องการสกัดกฎภาษาธรรมชาติที่มีค่าความถูกต้องสูง มีจำนวนกฎ มีตัวเชื่อมของเงื่อนไข และมีเวลาที่ใช้ในการสกัดกฎน้อย ดังนั้นจึงต้องมีการตัดโหนดโครงข่ายประสาทเทียมก่อนเข้าสู่ขั้นตอนการสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียม โดยการแทนค่าความถี่ของช่วงข้อมูล การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ ค่าน้ำหนักสูงสุด มีการทำงาน 3 ขั้นตอนคือ ขั้นตอนที่ 1 การเตรียมข้อมูล ขั้นตอนที่ 2 การสอนโครงข่ายประสาทเทียมด้วยการเรียนรู้แบบแพร่ย้อนกลับ และขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด

ขั้นตอนหลักที่ II การสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (Rule Extraction form Neural Networks using Frequency Interval Data Representation : RENN-FIDR) จะพิจารณาเฉพาะโหนดซ่อน และโหนดข้อมูลเข้าที่เสนอแนะจากขั้นตอนหลักที่ I (การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด: NNP-MFW) ในขั้นตอนหลักที่ II นี้มี 3 ขั้นตอนคือ ขั้นตอนที่ 4 ขั้นตอนวิธีการสร้างรูปแบบภาษาธรรมชาติ

ชาติจากความถี่ของช่วงข้อมูล ขั้นตอนที่ 5 การสอนโครงข่ายประสาทเทียมด้วยด้วย รูปแบบภาษาธรรมชาติ และขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎภาษาธรรมชาติโดยใช้ค่าน้ำหนักที่เป็นบวก

ผู้ทำการวิจัยได้พัฒนาโปรแกรมจากแบบจำลองที่นำเสนอเพื่อสกัดกฎสำหรับฐานข้อมูลทางการแพทย์โดยใช้โครงข่ายประสาทเทียม ซึ่งผู้ใช้สามารถใช้งานได้ง่ายเนื่องจากโปรแกรมมีลักษณะเป็นกราฟฟิก

มีการทดลองกับฐานข้อมูลทางการแพทย์ที่เป็นมาตรฐานจากต่างประเทศได้แก่ ฐานข้อมูลโรคมะเร็งเต้านมจากรัฐวิสคอนซิน ฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์ ฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ลือคยุโรป และฐานข้อมูลโรคเบาหวาน และฐานข้อมูลทางด้าน การแพทย์ที่เก็บตัวอย่างมาจากโรงพยาบาลในประเทศไทยได้แก่ฐานข้อมูลภาวะผู้ป่วยโรคหัวใจ พบว่า กฎภาษาธรรมชาติที่สกัดได้แพทย์สามารถเข้าใจได้เพราะกฎแทนด้วยภาษาธรรมชาติ เช่น เล็ก กลาง และใหญ่ เป็นต้น ตัวแปรที่มีผลต่อการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square จำนวนรูปแบบภาษาธรรมชาติ L และค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรมีแนวโน้มผลการทดลอง ดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square ถ้ามีการตัดโหนดโครงข่ายประสาทเทียมก่อนการสกัดกฎ ($\square > 0$) กฎที่ได้จะมีจำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขน้อยกว่าไม่มีการตัดโหนดโครงข่ายประสาทเทียมก่อนการสกัดกฎ ($\square = 0$) เวลาที่ใช้ในการสกัดกฎเมื่อมีการตัดโหนดโครงข่ายประสาทเทียมน้อยกว่ามากด้วย แต่อย่างไรก็ตามค่าความถูกต้องอาจน้อยกว่าไม่มีการตัดโหนดโครงข่ายประสาทเทียม ทั้งนี้เราสามารถกำหนดได้ว่าจะยอมรับค่าความถูกต้องน้อยกว่าได้จนถึงระดับใด ซึ่งจากผลการทดลองควรกำหนดให้ \square อยู่ในช่วง 0 ถึง 40 โดยที่ ถ้ากำหนดให้ \square มีค่าน้อยจะตัดโหนดโครงข่ายประสาทเทียมได้น้อย ในถ้ากลับกันถ้ากำหนดให้ \square มีค่ามากจะตัดโหนดโครงข่ายประสาทเทียมได้มาก

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไข และเวลาเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย ซึ่งขึ้นอยู่กับ ผู้ใช้จะเลือก

3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วงข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ค่าความ-ถูกต้องจำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน ซึ่งได้ผลซึ่งเกิดมาจากการรวมช่วง นอกจากนี้ค่าความ

ถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎ ยังขึ้นอยู่กับขั้นตอนการสอน โครงข่ายประสาทเทียม ซึ่งส่งผลมายังขั้นตอนการสร้างกฎอีกด้วย

6.2 ปัญหาและอุปสรรค

ปัญหาและอุปสรรคที่พบจากการทำวิทยานิพนธ์มีดังนี้

1. เนื่องจากได้ทดสอบแบบจำลอง และพัฒนาโปรแกรมจากแบบจำลองด้วยโปรแกรม MATLAB ซึ่งต้องการเครื่องคอมพิวเตอร์ที่มีความสามารถสูง ทำให้ใช้เวลานานมากในการประมวลผล

2. การพัฒนาโปรแกรมโดยใช้โปรแกรม MATLAB ซึ่งผู้ทำวิทยานิพนธ์ไม่มีประสบการณ์กับภาษานี้มาก่อน ทำให้ต้องใช้เวลาในการเขียนโปรแกรม

6.3 ข้อเสนอแนะ

ควรใช้เครื่องที่มีความสามารถสูงมาประมวลผลโปรแกรมที่เขียนด้วย MATLAB หรือใช้เครื่องหลายๆ เครื่องมาช่วยในการประมวลผลก็จะทำให้ได้ผลลัพธ์เร็วขึ้น