

บทที่ 1

บทนำ

1.1 ความเป็นมา และความสำคัญของงานวิจัย

ปัจจุบันการเรียนรู้โดยวิธีการโปรแกรมตรรกะเชิงอุปนัย (Inductive logic programming: ILP) หรือ ไอแอลพี (S. Muggleton, 1992; S. Muggleton, L. De Raedt, 1994; S. Muggleton and C. Feng, 1990) นั้นเป็นที่นิยมและถูกนำไปใช้งานอย่างแพร่หลาย เช่น การวิเคราะห์ไฟไนต์เอลิเมนต์ (Finite Element) การวิเคราะห์โครงสร้างโปรตีน การวิเคราะห์สาเหตุของการเกิดอุบัติเหตุทางจราจร การวิเคราะห์ความสามารถก่อกลายพันธุ์ (Mutagenesis) ของโมเลกุล ฯลฯ อันเนื่องมาจากผลลัพธ์ที่เป็นกฎลำดับที่หนึ่ง (first-order rules) ซึ่งง่ายต่อการทำความเข้าใจ สามารถอธิบายแนวคิดของข้อมูลที่มีความซับซ้อนและสามารถนำความรู้ภูมิหลังมาใช้ในการเรียนรู้ได้

แต่การใช้วิธีการโปรแกรมตรรกะเชิงอุปนัยในการเรียนรู้นั้นก็มีข้อเสียในเรื่องของการใช้เวลาในการเรียนรู้ที่ค่อนข้างสูงเมื่อต้องทำการเรียนรู้กับเซตเรียนรู้ที่มีปริมาณข้อมูลขนาดใหญ่เมื่อเปรียบเทียบกับวิธีการเรียนรู้โดยวิธีการแบบอื่นๆ หลายงานวิจัยได้ถูกนำเสนอขึ้นเพื่อช่วยลดเวลาในการเรียนรู้ของวิธีการโปรแกรมตรรกะเชิงอุปนัย เช่น การใช้การประมวลผลแบบขนานหรือการใช้ตัวแทนหลายตัวร่วมกันเพื่อแก้ไขปัญหา เป็นต้น ซึ่งหลักการในการทำงานโดยส่วนใหญ่ของวิธีการที่ถูกนำเสนอ จะทำการแบ่งข้อมูลตัวอย่างออกเป็นส่วนย่อยทั้งแบบที่แต่ละเซตย่อยมีข้อมูลซ้ำกับและไม่ซ้ำกัน แล้วนำไปทำการเรียนรู้ จากนั้นจึงทำการรวมหรือคัดเลือกกฎที่ด้วยวิธีการต่างๆ แตกต่างกัน ซึ่งวิธีการเหล่านี้สามารถที่จะลดเวลาในการประมวลผลได้ อย่างไรก็ตามก็มีความถูกต้องของเซตของกฎที่ได้ก็อาจจะลดลงได้เนื่องจากกฎที่ได้มาจากเรียนรู้กับข้อมูลตัวอย่างเพียงบางส่วนซึ่งอาจจะไม่สามารถครอบคลุมเซตตัวอย่างทั้งหมดได้หรือกฎที่ได้มีการวางนัยโดยทั่วไป (General) สูงกว่าการเรียนรู้กับเซตข้อมูลตัวอย่างทั้งหมด วิธีในการรวมกฎที่ได้จากหน่วยประมวลผลย่อยหรือตัวแทนหลายตัวร่วมกันเรียนรู้จึงมีความสำคัญ (Nuno A. Fonseca, Fernando Silva, and Rui Camacho, 2005)

เพื่อที่จะสามารถสร้างเซตของกฎจากที่สามารถให้ค่าความถูกต้องได้สูง หลายวิธีการในการรวมและคัดเลือกกฎได้ถูกนำเสนอขึ้น ซึ่งวิธีการส่วนใหญ่จะใช้การคัดเลือกและรวมกฎจากค่า

ความครอบคลุมและความถูกต้องของกฎ อาจจะแตกต่างกันในส่วนของเงื่อนไขในการเลือกกฎ เช่น เลือกกฎโดยการเปรียบเทียบค่าความครอบคลุมและค่าความถูกต้องตามลำดับ (Nuno A. Fonseca, Fernando Silva and Rui Camacho, 2005) หรือเลือกกฎตามค่ากำหนด (Ines de Castro Dutra, David Page, Vitor Santos Costa and Jude Shavlik, 2002) เป็นต้น แต่อาจจะมีบางงานวิจัยที่แปลงกฎให้อยู่ในรูปแบบของกราฟและใช้ในการตรวจสอบและคัดเลือกกฎ (Ronaldo C. Prati, and Peter A. Flach. Roccer, 2005) ซึ่งวิธีการที่นำเสนอขึ้นให้ค่าความถูกต้องของเซตของกฎได้ดีและสามารถลดจำนวนของกฎที่ซ้ำซ้อนกันได้ดี

อย่างไรก็ดีวิธีการเหล่านี้ไม่ได้มีปรับปรุงที่ตัวกฎหรือมีการรวมกฎ 2 กฎหรือมากกว่าเข้าด้วยกัน กฎที่มีในเซตกฎยังคงมาจากการเรียนรู้กับเซตตัวอย่างเรียนรู้เพียงบางส่วน แต่ใช้การรวมกฎจากตัวเรียนรู้หลายๆ ตัวเข้าด้วยกันเพื่อให้สามารถครอบคลุมเซตตัวอย่างทั้งหมดได้ ซึ่งอาจจะเป็นไปได้ที่ตัวอย่างเรียนรู้บางตัวอาจจะไม่ถูกครอบคลุมหรือเซตของกฎที่ได้ยังคงมีการวางนัยโดยทั่วไปสูงเนื่องจากไม่สามารถตัดตัวอย่างลบที่ไม่เคยเห็นออกจากเซตของกฎได้ จึงเป็นที่มาของแนวคิดที่ว่า เพื่อให้กฎที่ได้สามารถครอบคลุมเซตตัวอย่างทั้งหมดได้ดียิ่งขึ้นหรือเพื่อให้เซตของกฎที่ได้มีความเฉพาะเจาะจงมากขึ้น บางส่วนของกฎที่ถูกตัดออกหรือไม่ได้ถูกนำมารวมในเซตของกฎที่เป็นคำตอบ น่าจะสามารถใช้ในการปรับปรุงกฎให้ดีขึ้นได้ นอกเหนือจากการรวมและคัดเลือกกฎด้วยวิธีการตามปกติ

สำหรับวิทยานิพนธ์ฉบับนี้ได้ศึกษาเกี่ยวกับวิธีการเรียนรู้โดยใช้วิธีการโปรแกรมตรรกะเชิงอุปนัยกับข้อมูลที่มีปริมาณขนาดใหญ่ และศึกษาหาวิธีในการลดระยะเวลาในการเรียนรู้โดยยังคงความถูกต้องของกฎที่ได้ วิธีการที่นำมาใช้ในวิทยานิพนธ์จะใช้วิธีการแบ่งเซตตัวอย่างเรียนรู้ให้ตัวเรียนรู้แต่ละตัวไปทำการเรียนรู้ด้วยวิธีการโปรแกรมตรรกะเชิงอุปนัย โดยเซตตัวอย่างที่จะใช้ในการเรียนรู้สำหรับตัวเรียนรู้แต่ละตัวจะไม่เหมือนกัน เมื่อตัวเรียนรู้แต่ละตัวได้เซตของกฎ ก็จะมีการเลือกเซตของกฎที่ให้ค่าความถูกต้องสูงสุดในเซตตัวอย่างเรียนรู้ที่แบ่งให้ จากนั้นจะนำมาทำการปรับปรุงโดยการเพิ่มอนุประโยคจากเซตของกฎที่ได้จากตัวเรียนรู้อื่นๆ และนำไปทดสอบกับเซตตัวอย่างเรียนรู้ของตัวเรียนรู้อื่นๆ จากนั้นจะทำการเปรียบเทียบค่าความถูกต้องที่ได้เซตของกฎที่ได้กับการรวมกฎอย่างง่ายแบบต่างๆ และเซตของกฎที่ได้จากการเรียนรู้กับเซตตัวอย่างเรียนรู้ทั้งหมด ซึ่งผลที่ได้น่าจะให้กฎที่มีความถูกต้องที่ดีกว่าวิธีการรวมกฎอย่างง่าย และใกล้เคียงกับการเรียนรู้ด้วยวิธีการโปรแกรมตรรกะเชิงอุปนัยกับเซตตัวอย่างเรียนรู้ทั้งเซต

1.2 วัตถุประสงค์

1.2.1 เพื่อศึกษาวิธีการในการเรียนรู้ของวิธีการโปรแกรมตรรกะเชิงอุปนัยกับข้อมูลที่มีปริมาณขนาดใหญ่

1.2.2 เสนอวิธีการในการลดเวลาในการเรียนรู้ของวิธีการโปรแกรมตรรกะเชิงอุปนัยมาใช้กับข้อมูลที่มีปริมาณขนาดใหญ่ โดยที่ยังคงความถูกต้องของกฎที่ได้ให้ใกล้เคียงกับการเรียนรู้ด้วยวิธีการโปรแกรมตรรกะเชิงอุปนัยกับข้อมูลทั้งหมด

1.3 ขอบเขตของงานวิจัย

1.3.1 ชุดข้อมูลประกอบด้วยข้อมูลชุดตัวอย่างการหาจำนวนองค์ประกอบจำกัดของโครงข่าย (Finite-Element Mesh) เฉพาะคลาสที่ 1 และ 2 (Mesh1 และ Mesh2) ความสามารถก่อกลายพันธุ์ของโมเลกุล (Mutagenesis) และ ข้อมูลการก่อมะเร็ง (Carcinogenesis)

1.3.2 งานวิจัยทดลองบนเครื่องคอมพิวเตอร์ที่มีหน่วยประมวลผลที่ความเร็ว 1.73 GHz ขนาดหน่วยความจำที่ 1 GB

1.3.3 เครื่องมือในการสร้างกฎด้วยวิธีการโปรแกรมตรรกะเชิงอุปนัย Aleph

1.3.4 ระบบปฏิบัติการที่ใช้เป็น Microsoft Windows XP professional SP2

1.3.5 การทำงานในระบบ off-line และไม่ได้ทำงานแบบขนาน

1.3.6 การวัดเปรียบเทียบเวลาที่ใช้จะทำการแบ่งข้อมูลตัวอย่างเป็นส่วนๆ แล้วจึงนำมาทำการเรียนรู้ทีละรอบจากนั้นจึงทำการบันทึกและรวมเวลาที่ได้ทั้งหมดมาเปรียบเทียบกับการเรียนรู้กับชุดข้อมูลทั้งหมด

1.3.7 ใช้วิธีทดสอบความถูกต้องแบบไขว้ข้ามหากลุ่มในการประมาณค่าความผิดพลาด (5-fold Cross-Validation)

1.4 ผลที่คาดว่าจะได้รับ

สามารถสร้างเซตของกฎที่มีความถูกต้องใกล้เคียงและใช้เวลาน้อยกว่าการเรียนรู้กับเซตข้อมูลทั้งหมด

1.5 รายละเอียดของวิทยานิพนธ์

ในส่วนของวิทยานิพนธ์แบ่งออกเป็นส่วนต่างๆ ดังนี้

บทที่ 1 บทนำ ประกอบด้วย ความเป็นมาและความสำคัญของงานวิจัยนี้ วัตถุประสงค์ขอบเขตของงานวิจัย ผลที่คาดว่าจะได้รับ และรายละเอียดของโครงร่างวิทยานิพนธ์

บทที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง ประกอบด้วย ความรู้พื้นฐานที่เกี่ยวข้องกับวิทยานิพนธ์ฉบับนี้ และงานวิจัยที่เกี่ยวข้องกับการเรียนรู้ด้วยวิธีโปรแกรมตรรกะเชิงอุปนัยกับการแก้ปัญหาต่างๆ

บทที่ 3 วิธีการวิจัย ประกอบด้วยหลักการและวิธีในการวิจัย คำนึงเกี่ยวกับสายโซ่สัญญาณ

บทที่ 4 ผลที่ได้จากการทดลอง เปรียบเทียบกับวิธีการรวมกฎอย่างง่ายแบบอื่นๆ และการเรียนรู้โดยวิธีการโปรแกรมตรรกะเชิงอุปนัยกับเซตตัวอย่างเรียนรู้ทั้งหมด

บทที่ 5 สรุปผลการทดลองและข้อเสนอแนะ