**บทที่ 5**

**ผลการวิเคราะห์การทดลอง**

เป้าหมายหลักของบทนี้คือการวิเคราะห์ผลการทดลองจากการนำชุดข้อมูลฝึกสอนที่สร้างจากกฎของไฟร์วอลล์ที่ออกแบบให้ตรงตามจุดประสงค์ของสมมติฐาน เพื่อหาชุดข้อมูลฝึกสอนที่สามารถทำให้โมเดลมีประสิทธิภาพในด้านความแม่นยำในการทำนายและเวลาที่ใช้ได้ดีที่สุด จึงจำเป็นต้องมีการวิเคราะห์ในเชิงเปรียบเทียบ ปรับรูปแบบกราฟเพื่อหาความสัมพันธ์ของตัวแปรต่างๆ

**5.1. การวิเคราะห์กลไกการทำงานโดยรวมของโมเดล**

5.1.1 วิเคราะห์ความสัมพันธ์ เวลาที่ใช้ในการฝึกสอนโมเดลและจำนวนข้อมูลฝึกสอน

จากผลการทดลองในบทก่อนหน้าพบว่า ทั้ง 2 ผลการทดลอง เมื่อมีจำนวนชุดข้อมูลฝึกสอนในระบบมากขึ้น โมเดลจะใช้เวลาในการเรียนรู้ชุดข้อมูลฝึกสอนในอัตราคงที่ สังเกตได้จากกราฟที่ออกมามีลักษณะใกล้เคียงกับกราฟเส้นตรงมาก

**รูปที่ 5.1** กราฟผลลัพธ์ เวลาที่ใช้ในการฝึกสอนโมเดล : จำนวนชุดข้อมูลฝึกสอนที่ใช้

5.1.2. วิเคราะห์ความสัมพันธ์ เวลาที่ใช้ในการประมวลผลและจำนวนข้อมูลที่ใช้ฝึกสอน

จากผลการทดลองในบทก่อนหน้าพบว่า ทั้ง 2 ผลการทดลอง จะเห็นได้ว่าจำนวนชุดข้อมูลฝึกสอนแทบไม่ส่งผลกับเวลาที่ใช้ในการประมวลผล นั่นหมายความว่าถ้าหากเราใช้ชุดกฎไฟร์วอลล์ที่มีเงื่อนไขมากขึ้นก็ยังใช้เวลาในการประมวลผลเท่าเดิม จากภาพ 5.2 จะเห็นได้ว่าทั้ง 2 รูป ถึงจะใช้เวลามากขึ้นหรือน้อยลงบ้าง แต่ค่าความแตกต่างจะอยู่ในเสี้ยววินาทีเท่านั้น

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

**รูปภาพที่ 5.2** เปรียบเทียบกราฟผลลัพธ์เวลาที่ใช้ในการประมวลของ N Sample และ Ratio

**5.2.** **การวิเคราะห์ประสิทธิภาพการทำงานของโมเดล**

5.2.1. อัตราการเรียนรู้ของโมเดล

ตัวแปรสำคัญในการวัดผลในเชิงประสิทธิภาพ ได้แก่ ความแม่นยำในการทำนายผล เวลาที่ใช้ในการฝึกโมเดล และ จำนวนชุดข้อมูลทดสอบ ซึ่งทั้ง 3 ค่านี้ให้ความหมายในเชิงประสิทธิภาพได้ดังนี้

* โมเดลที่มีความแม่นยำสูงกว่า เป็นโมเดลที่มีประสิทธิภาพมากกว่า
* โมเดลที่ใช้เวลาในการเรียนรู้น้อยกว่าย่อมดีกว่าโมเดลที่ใช้เวลาในการเรียนรู้มากกว่า ถ้าหากโมเดลทั้งสองให้ผลลัพธ์ความถูกต้องในการทำนายผลเท่ากัน
* โมเดลที่ใช้จำนวนชุดข้อมูลฝึกสอนน้อยกว่าจะดีกว่าโมเดลที่ใช้จำนวนชุดข้อมูลฝึกสอนมากกว่า ถ้าหากโมเดลทั้งสองให้ผลลัพธ์ความถูกต้องในการทำนายผลเท่ากัน ซึ่งจำนวนชุดข้อมูลฝึกสอนจะมีผลโดยตรงกับเวลาที่ใช้ นั่นหมายความว่า เราจะต้องใช้เวลาในการสร้างชุดข้อมูลและฝึกฝนนานขึ้น

ซึ่ง 3 ตัวแปรนี้ ทำให้ได้กราฟอีกตัวหนึ่งคือกราฟ ความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอน

Chart, line chart

Description automatically generated

**รูปที่ 5.3** กราฟความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอนของ N Sample

Chart, line chart

Description automatically generated

**รูปที่ 5.4** กราฟความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอนของอัตราส่วน Ratio

Chart, line chart

Description automatically generated

**รูปที่ 5.5** กราฟความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอนของ N Sample (2)

จะเห็นได้ว่ากราฟทั้งสองรูปแบบ ทั้ง N Sample และอัตราส่วน Ratio มีอัตราการเรียนรู้ที่มีลักษณะคล้ายกันคือ โมเดลที่มีการเรียนรู้จากชุดข้อมูลฝึกสอนน้อยยังสามารถคำนวณหาผลลัพธ์ได้ถูกต้องบ้างอยู่ แต่การที่จะเพิ่มความแม่นยำได้นั้นจะต้องเพิ่มจำนวนชุดข้อมูลฝึกสอนไปอีกเกือบเท่าตัวหรือหลายเท่า นั่นหมายความว่าอัตราการเรียนรู้จะเริ่มน้อยลงไปเรื่อยๆแปรผกผันกับจำนวนชุดข้อมูลฝึกสอนที่ป้อนเข้าไป

และเมื่อเปรียบเทียบอัตราการเรียนรู้ของทั้ง 2 รูปแบบระหว่าง N Sample และ Ratio โดยมีข้อมูลฝึกสอนที่ใช้ทั้งหมดเป็นตัวตั้ง จะเห็นว่าแบบที่ไม่ได้มีการนำ Default มาคิด การแบ่งชุดฝึกสอนแบบ Ratio จะให้ผลดีกว่าเล็กน้อย แต่เมื่อมีการนำ Default Rule มาคิดด้วย แบบ N Sample ให้ผลที่ดีกว่าอย่างเห็นได้ชัด ซึ่งถ้ามองตามหลักความเป็นจริงการนำ Default Rule มาคิด เป็นสิ่งที่จำเป็นมาก เพราะกฎไฟร์วอลล์ที่ใช้จริงจะมีจำนวนความเป็นไปได้ที่มากกว่านี้มาก และการแบ่งแบบ Ratio ให้ผลลัพธ์ที่แย่กว่า

**รูปภาพที่ 5.6** การเปรียบเทียบอัตราการเรียนรู้ของแบบ N Sample และแบบ Ratio

**5.2.2. การเลือกหาจุดจำนวนชุดข้อมูลฝึกสอนที่เหมาะสมที่สุดหรือจุด optimum ของโมเดล**

Chart, line chart

Description automatically generated

**รูปภาพที่ 5.7** กราฟผลลัพธ์ความแม่นยำของการแบ่งชุดข้อมูลฝึกสอน N Sample

Chart, line chart

Description automatically generated

**รูปภาพที่ 5.8** กราฟผลลัพธ์ความแม่นยำของการแบ่งชุดข้อมูลฝึกสอนแบบอัตราส่วน Ratio

ในการหาจุด optimum นั้นจะต้องเลือกจุดที่โมเดลทำนายผลได้แม่นยำมากที่สุดและใช้จำนวนชุดข้อมูลฝึกสอนที่น้อยที่สุด ซึ่งหมายความว่าเวลาที่ใช้ก็จะน้อยลงด้วยเช่นกัน (จากการวิเคราะห์ที่ 5.1.1) และประเด็นที่คาดไม่ได้เลยคือ ควรเป็นจุดที่ทุก rule set ยอมรับได้ ถ้าหากโมเดลสามารถตัดสิน rule set ที่ 1 ได้มีความแม่นยำสูง แต่ไปตัดสิน rule set อื่นได้ความแม่นยำที่ต่ำมาก จำนวนชุดข้อมูลฝึกสอนในช่วงนั้นก็จะเป็นจำนวนชุดข้อมูลฝึกสอนที่ยกจำนวนมาไม่ได้มาตรฐาน ซึ่งเป้าหมายของเราคือการแบ่งด้วยจำนวนเท่าใด จะให้ความแม่นยำและเหมาะสมกับทุก rule set มากที่สุด

จากกราฟ 5.6 และ 5.7 จะเห็นได้ว่า ถ้าหากเป็น N Sample ค่าความแม่นยำที่ได้จะค่อยๆเพิ่มขึ้น จะมีจุดหนึ่งที่ทุก rule set มีแนวโน้มความแม่นยำที่เพิ่มขึ้น ซึ่งตัวแปรที่ใช้ในการทดลองจุดนั้นคือ 600 ชุดข้อมูลฝึกสอนต่อหนึ่งกฎไฟร์วอลล์ และเมื่อให้จำนวนข้อมูลฝึกสอนที่มากกว่านั้น ความแม่นยำในทุกชุดกฎไฟร์วอลล์จะเริ่มตกลงเล็กน้อย โดยเฉพาะชุดกฎไฟร์วอลล์ที่มีเงื่อนไขที่มากกว่าจะสังเกตได้ชัดเจน ในขณะเดียวกันถ้าหากเป็นแบบอัตราส่วน Ratio หากสังเกตจะมีช่วงอัตราส่วน Ratio ที่ 0.01 มีจุดค่าความแม่นยำที่สูงที่สุดและเริ่มมีอัตราความแม่นยำตกลงมาเมื่อมีจำนวนชุดข้อมูลฝึกสอนมากขึ้น โดยเราได้นำจุด optimum ทั้ง 2 จุดที่ของแต่ละแบบมาเปรียบเทียบเพื่อหาความแตกต่างด้านเวลา ซึ่งแต่ละชุดเงื่อนไขใช้จำนวนข้อมูลฝึกสอนดังนี้

|  |  |  |
| --- | --- | --- |
|  | Without Default | With Default |
| Rule set 1 without Default – N Sample | 1200 | 1800 |
| Rule set 1 with Default – Ratio | 818 | 2628 |
| Rule set 2 without Default – N Sample | 2400 | 3000 |
| Rule set 2 with Default – Ratio | 822 | 2617 |
| Rule set 3 without Default – N Sample | 3600 | 4200 |
| Rule set 3 with Default - Ratio | 1148 | 2943 |

ตารางเทียบข้อมูลฝึกสอนที่ใช้ทั้งหมดระหว่าง N Sample (600) และ Ratio (0.01)

จะเห็นได้ว่าชุดข้อมูลฝึกสอนที่ใช้นั้นแบบอัตราส่วนใช้จำนวนชุดข้อมูลฝึกสอนที่น้อยกว่าแทบทุกช่วง ยกเว้นแค่ Rule set 1 With Default ที่มีจำนวนชุดข้อมูลฝึกสอนมากกว่า ซึ่งสรุปได้ว่า แบบอัตราส่วน Ratio ใช้ชุดข้อมูลฝึกสอนน้อยกว่าแบบ N Sample ในการไปถึงจุดเหมาะสมที่ให้ความแม่นยำสูงสุด ซึ่งนั่นหมายความว่าจำเวลาฝึกสอนที่ใช้จะน้อยกว่าเช่นกัน แบบอัตราส่วน Ratio จึงเหมาะแก่การนำมาใช้มากกว่า

แม้เราจะเห็นได้ว่าจุด optimum ของการแบ่งแบบอัตราส่วน Ratio ที่ให้ค่าความแม่นยำได้พอๆกันกับแบบ N Sample และมีการใช้เวลารวมที่น้อยกว่า แต่การแบ่งแบบอัตราส่วน Ratio มีข้อจำกัดอย่างหนึ่งคือ ถ้าหาก โมเดลมีความแตกต่างทางด้านชุดข้อมูลฝึกสอนมากๆ เราไม่สามารถสร้างจำนวนชุดข้อมูลฝึกสอนจากกฎที่น้อยเกินไปได้ เพราะถ้าหากอ้างอิงตามเลขที่ได้จากการทดลองคือ 0.01 ถ้าหากความเป็นไปได้ของเงื่อนไขที่เรากำหนดขึ้นใหม่มีขนาดน้อยกว่า 100 เราจะไม่สามารถออกแบบชุดข้อมูลฝึกสอนที่มีขนาดต่ำกว่า 1 ได้เลย ซึ่งนั่นอาจทำให้เป็นปัญหาและจะเป็นประเด็นสำคัญที่เราจะนำมาพิจารณาในเทอมการศึกษาถัดไป