**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

5.1.3. วิเคราะห์ความสัมพันธ์ ความแม่นยำการทำนายต่อจำนวนของข้อมูลฝึกสอน

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

**รูปภาพที่ 5.3** เปรียบเทียบกราฟผลลัพธ์ความแม่นยำของการแบ่งชุดข้อมูลฝึกสอนแต่ละแบบ

จะเห็นได้ว่า ถ้าหากเป็น N Sample ค่าความแม่นยำที่ได้จะค่อยๆเพิ่มขึ้น และขึ้นเกือบจุดสูงสุดที่จุดหนึ่ง ซึ่งตัวแปรที่ใช้ในการทดลองจุดนั้นคือ 600 ชุดข้อมูลฝึกสอนต่อหนึ่งกฎไฟร์วอลล์ แต่เมื่อให้จำนวนข้อมูลฝึกสอนที่มากกว่านั้น ความแม่นยำในทุกชุดกฎไฟร์วอลล์จะเริ่มตกลงเล็กน้อย โดยเฉพาะชุดกฎไฟร์วอลล์ที่มีเงื่อนไขที่มากกว่า ยกเว้นชุด 2 กฎไฟร์วอลล์ที่มี Default Rule ด้วย

จึงสรุปได้ว่า ความซับซ้อนและเงื่อนไขของชุดข้อมูลฝึกสอนส่งผลต่อ การเรียนรู้ของโมเดล หมายความว่า ถ้าหากเงื่อนไขกฎไฟร์วอลล์ที่ใช้นั้นมีจำนวนเงื่อนไขและกฎการประเมินที่มากขึ้น จำเป็นต้องหาจำนวนของข้อมูลที่จะใช้ฝึกสอนที่เหมาะสมที่ทำให้โมเดลมีประสิทธิภาพมากที่สุด

**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 5.4** กราฟความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอนของ N Sample

Chart, line chart

Description automatically generated

**รูปที่ 5.5** กราฟความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอนของอัตราส่วน Ratio

Chart, line chart

Description automatically generated

**รูปที่ 5.6** กราฟความแม่นยำ / เวลาฝึกโมเดล : จำนวนชุดข้อมูลฝึกสอนของ N Sample (2)

จะเห็นได้ว่ากราฟทั้งสองรูปแบบ ทั้ง N Sample และอัตราส่วน Ratio มีอัตราการเรียนรู้ที่มีลักษณะคล้ายกันคือ โมเดลที่มีการเรียนรู้จากชุดข้อมูลฝึกสอนน้อยยังสามารถคำนวณหาผลลัพธ์ได้ถูกต้องบ้างอยู่ แต่การที่จะเพิ่มความแม่นยำได้นั้นจะต้องเพิ่มจำนวนชุดข้อมูลฝึกสอนไปอีกเกือบเท่าตัวหรือหลายเท่า นั่นหมายความว่าอัตราการเรียนรู้จะเริ่มน้อยลงไปเรื่อยๆแปรผกผันกับจำนวนชุดข้อมูลฝึกสอนที่ป้อนเข้าไป

5.2.2. การเลือกหาจุดจำนวนชุดข้อมูลฝึกสอนที่เหมาะสมที่สุดในการพัฒนาโมเดล

ในการเลือกจำนวนชุดข้อมูลฝึกสอนมาใช้พัฒนาโมเดล ผลลัพธ์สุดท้ายที่สามารถตอบโจทย์ประสิทธิภาพทั้ง 3 ค่าได้ โดยประกอบไปด้วย จำนวนชุดข้อมูลฝึกสอน เวลาที่ใช้ในการฝึกสอน และความแม่นยำในการทำนายผล จะต้องให้ผลลัพธ์ได้ดีที่สุด ซึ่งในหัวข้อการวิเคราะห์นี้เราได้ใช้ผลลัพธ์การทดลองและหลักการมาอ้างอิง ดังนี้

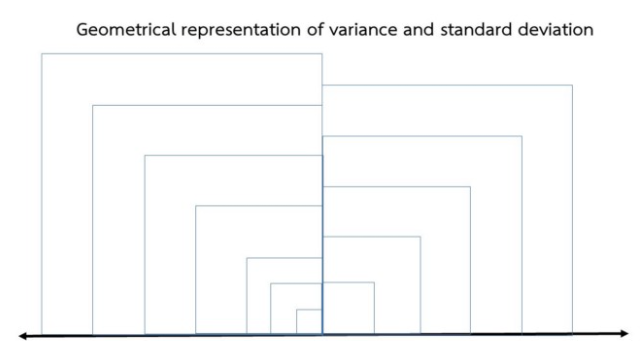
5.2.2.1. การอ้างอิงความสัมพันธ์ของเวลาฝึกสอนและจำนวนชุดข้อมูลฝึกสอน

จากการนำกราฟที่ 5.1 มาวิเคราะห์ พบว่าจำนวนชุดข้อมูลฝึกสอนโมเดลและเวลาที่ใช้ในการฝึกสอนโมเดลมีความสัมพันธ์แปรผันตรงด้วยอัตราการเพิ่มค่าที่คงที่ หมายความว่าเราสามารถหาเวลาที่ใช้ในการฝึกสอนโมเดลได้ถ้าหากเราทราบจำนวนชุดข้อมูลฝึกสอนที่ใช้ หมายความว่าการเลือกจุดจำนวนชุดข้อมูลฝึกสอนที่ให้ค่าความถูกต้องได้เยอะที่สุด เหมาะสมที่จะเป็นจำนวนที่ควรหยิบมาใช้มากที่สุดนั่นเอง ซึ่งเป็นไปตามสูตร

* n เป็นจำนวนชุดข้อมูลฝึกสอนที่ต้องการ
* t เป็นเวลาที่โมเดลใช้ในการเรียนรู้ต่อ 1 ข้อมูล
* T เป็นเวลาทั้งหมดที่โมเดลใช้ในการเรียนรู้จนเสร็จ

5.2.2.2. การใช้หลักการหาค่าความแปรปรวนในการเปลี่ยนมิติการเปรียบเทียบ

ความแปรปรวนคือความแกว่งของข้อมูลหรือความห่างของการกระจายตัวของชุดข้อมูล โดยในทางสถิติหรือการเปรียบเทียบเชิงคณิตศาสตร์มักมีการนำความแปรปรวนและส่วนเบี่ยงเบนมาใช้ในการหาค่าเฉลี่ยของกราฟ ซึ่งหลักการหนึ่งที่ทำให้เห็นผลชัด คือการนำค่ามายกกำลังสองตามสูตรความแปรปรวน เราจะสามารถหาจุดแตกต่างหรือมีความแปรปรวณได้ชัดเจนขึ้น โดยสมการมีรูปแบบสูตร ดังนี้

  
**รูปที่ 5.7** การเปรียบเทียบเมทริกซ์ Geometrical ก่อนและหลังยกกำลังสอง

จากประเด็นก่อนหน้าที่นำมาพิจารณาคือจำนวนชุดข้อมูลฝึกสอนกับเวลาในการเรียนรู้มีทิศทางความสัมพันธ์แบบแปรผันตรงแบบคงที่ จึงสามารถนำชุดข้อมูลฝึกสอนมาเป็นเกณฑ์ในการวัดเรื่องเวลาได้เลย และตัวแปรอีกตัวหนึ่งคือความแม่นยำ โดยเราจะอ้างอิงหลักการหาความแปรปรวนที่มีการยกกำลังสอง มาเพิ่มมิติให้กับค่าความแม่นยำเพื่อให้เห็นความแตกต่างของความแม่นยำได้ดีขึ้น สาเหตุมาจากความแม่นยำมีค่าระยะหรือ Range ที่ต่ำคืออยู่ระหว่าง 0 ถึง 1 แน่นอน (หรือ 0% - 100% ) ทำให้สรุปได้เป็นกราฟใหม่ คือค่าความแม่นยำกำลังสองต่อจำนวนข้อมูลฝึกสอนที่ใช้

**กราฟแสดงผลลัพธ์ที่ได้จากการคำนวณใหม่แบบ N Sample**

Chart, line chart

Description automatically generated**รูปที่ 5.8** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (N Sample, R1, Without Default)

Chart, line chart

Description automatically generated**รูปที่ 5.9** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (N Sample, R1, With Default)

Chart, line chart

Description automatically generated**รูปที่ 5.10** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (N Sample, R2, Without Default)

Chart, line chart

Description automatically generated

**รูปที่ 5.11** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (N Sample, R2, With Default)

Chart, line chart

Description automatically generated

**รูปที่ 5.12** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (N Sample, R3, Without Default)

Chart, line chart

Description automatically generated

**รูปที่ 5.13** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (N Sample, R3, With Default)

**กราฟแสดงผลลัพธ์ที่ได้จากการคำนวณใหม่แบบอัตราส่วน Ratio**

Chart, line chart

Description automatically generated**รูปที่ 5.14** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (Ratio, R1, Without Default)

Chart, line chart

Description automatically generated**รูปที่ 5.15** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (Ratio, R1, With Default)

Chart, line chart

Description automatically generated**รูปที่ 5.16** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (Ratio, R2, Without Default)

Chart, line chart

Description automatically generated**รูปที่ 5.17** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (Ratio, R2, With Default)

Chart, line chart

Description automatically generated**รูปที่ 5.18** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (Ratio, R3, Without Default)

Table

Description automatically generated**รูปที่ 5.19** กราฟแม่นยำยกกำลังสอง: จำนวนข้อมูลฝึกสอนที่ใช้ (Ratio, R3, With Default)

จากกราฟที่ได้ทำการปรับค่าตัวแปรใหม่ จะเห็นได้ว่าทั้งกราฟการแบ่งจำนวนชุดฝึกสอนแบบ N Sample และแบบอัตราส่วน Ratio เมื่อมีกฎเดียวจะสามารถหาค่าจุดสูงสุดที่คุ้มค่าทั้งความแม่นยำและเวลาที่ใช้ได้ในกราฟที่มีเงื่อนไขกฎไฟร์วอลล์ 2 กฎ แต่ถ้าหากคิดในเรื่องจำนวนชุดข้อมูลฝึกสอนที่ใช้แล้ว วิธีแบบ N Sample ที่จำนวน 600 ได้ค่ายกกำลังที่ 0.7 ซึ่งเท่ากับการแบ่งแบบอัตราส่วนที่ 0.11 แต่มีการใช้ Packet จำนวนมากถึง 9,010 ชุด นั่นหมายความว่าวิธีแบบ N Sample ใช้เวลาได้ดีกว่าแบบอัตราส่วน Ratio ดีกว่าถึง 15 เท่า ในขณะที่มีการใช้กฎไฟร์วอลล์เหมือนกัน

เมื่อมาถึงเงื่อนไขกฎไฟร์วอลล์ที่มี 4 กฎพบว่าแบบ N Sample ยังสามารถหาจุดเหมาะสมได้อยู่แต่ความแม่นยำของ Ratio ที่ Without Default ให้ผลได้ดีกว่าในช่วงอัตราส่วน 0.01 ซึ่งมีจำนวนชุดข้อมูลทั้งหมดเพียง 800 และให้ผลลัพธ์ที่ดีกว่า แต่เมื่อนำ Default Rule มาใช้คิดคำนวณด้วยปรากฎว่าแบบ N Sample ให้ค่าความแม่นยำได้ถึง 0.8 ในขณะที่การแบ่งอัตราส่วน Ratio มีอัตราความแม่นยำที่ตกลงเรื่อยๆและมีความแม่นยำน้อยกว่าแบบ N Sample มาก

ต่อมาคือเงื่อนไขที่มี 6 กฎไฟร์วอลล์ จาการทดลองพบว่าแบบอัตราส่วน Ratio ให้ผลความแม่นยำที่ต่ำกว่าแบบ N Sample ในทุกจุด ซึ่งหากดูจากการทดลองที่ผ่านมาแล้ว เมื่อมีเงื่อนไขมากขึ้น แบบอัตราส่วน Ratio แทบจะด้อยกว่าแบบ N Sample ในทุกด้าน ไม่ว่าจะเป็นทั้งเรื่องความแม่นยำและเวลาที่ใช้ จึงสรุปได้ว่าเราไม่ควรนำวิธีการแบ่งจำนวนชุดฝึกสอนแบบอัตราส่วน Ratio ที่เท่ากันมาใช้ในการหาจำนวนชุดข้อมูลฝึกสอนที่จะเอาไปนำเข้าโมเดล