| | | | |
|---|---|---|---|
| **หัวข้อโครงงาน** | การศึกษาวิจัยเพื่อพัฒนาสร้างชุดข้อมูลในการฝึกสอนไฟร์วอลล์ปัญญาประดิษฐ์ด้วยเทคโนโลยีโครงข่ายประสาทเทียมจากกฎของไฟร์วอลล์ | | |
| **นักศึกษา** | ฐิติโชติ | ใจเมือง | รหัสนักศึกษา 60070019 |
| | พิพัฒน์บุญ | พุทธคุณ | รหัสนักศึกษา 60070065 |
| **ปริญญา** | วิทยาศาสตรบัณฑิต | | |
| **สาขาวิชา** | เทคโนโลยีสารสนเทศ | | |
| **ปีการศึกษา** | 2563 | | |
| **อาจารย์ที่ปรึกษา** | ผู้ช่วยศาสตราจารย์ อัครินทร์ คุณกิตติ | | |


# บทคัดย่อ


ในงานทำวิจัยนี้เราได้พัฒนาโปรแกรมสร้างชุดข้อมูลฝึกสอนจากกฎไฟร์วอลล์และโมเดลประสาทเทียมเชิงลึก เพื่อสังเกตุและวิเคราะห์การทดลองศึกษาหาผลลัพธ์หรือแนวทางที่จะนำไปประยุกต์ใช้กับการสร้างชุดข้อมูลฝึกสอนที่มีประสิทธิภาพ มีการใช้จำนวนข้อมูลฝึกสอนและเวลาที่ใช้น้อยแต่ได้ความแม่นยำสูง โดยออกแบบชุดข้อมูลฝึกสอนที่แตกต่างกันในเรื่องของจำนวนและกฎไฟร์วอลล์ สามารถแบ่งออกได้เป็น 2 รูปแบบ คือ แบบ N Sample จำนวนของชุดข้อมูลฝึกสอนในกฎไฟร์วอลล์แต่ละข้อมีจำนวนเท่ากันทั้งหมด และแบบ Ratio ที่จำนวนของชุดข้อมูลฝึกสอนในกฎไฟร์วอลล์แต่ละข้อจะแตกต่างกันโดยจำนวนที่มีมากหรือน้อยเป็นไปตามอัตตราส่วนที่กำหนดขึ้น หลังจากนั้นทำการทดสอบในแต่ละแบบโดยกำหนดค่าที่แตกต่างกัน 8 ค่า ในแต่ละแบบเพื่อเปรียบเทียบและวิเคราะห์ ในส่วนทำการทดลองจะสังเกตได้ว่าเมื่อมีจำนวนชุดข้อมูลฝึกสอนมากขึ้น เวลาที่ใช้ก็จะมากขึ้นตาม ในส่วนของความถูกต้องนั้นในแต่ละชุดกฎไฟร์วอลล์ ยิ่งกฎมีความซับซ้อนมากเท่าใดค่าความถูกต้องก็จะลดลง แต่ในส่วนของ N Sample จะไม่ได้ลดลงมากเมื่อเทียบกับ Ratio ซึ่งคาดว่าเกิดจากจำนวนชุดข้อมูลฝึกสอนที่แตกต่างในกฎแต่ละของ Ratio และจำนวน False positive และ False negative จำนวนชุดข้อมูลฝึกสอนที่มี Allow และ Deny ไม่เท่ากัน และจำนวนชุดข้อมูลฝึกสอนแต่ละกฎที่ต่างโดยในเฉพาะในแบบของ Ratio ยิ่งถ้าหากมีการนำ Default Rule เข้ามาเทรนจะเห็นได้ชัดว่า False negative มีจำนวนเพิ่มขึ้นอย่างมาก

จากการวิเคราะห์และทดลองสังเกตได้ว่าจุดเหมาะสมของการแบ่งอัตราส่วน Ratio และการแบ่งด้วยจำนวนที่เท่ากันมีการให้ความแม่นยำที่เท่าๆกัน แบบอัตราส่วน Ratio จะมีการใช้เวลาในการฝึกโมเดลที่น้อยกว่าเพราะต้องการจำนวนชุดข้อมูลฝึกสอนน้อยกว่า

ในงานวิจัยถัดไปจะเป็นการลงลึกรายละเอียดเกี่ยวกับการพัฒนาแบ่งชุดข้อมูลฝึกสอนด้วยอัลกอริทึ่มแบบใหม่ ซึ่งเราได้คาดเดาว่าวิธีนี้จะเป็นการแก้ไขปัญหาวิธีการแบ่งชุดข้อมูลที่เป็นแบบอัตราส่วน โดยประเด็นปัญหาที่สามารถเห็นได้ชัดคือ การแบ่งข้อมูลฝึกสอนที่มีความแตกต่างกันทางด้านกฎของไฟร์วอลล์มากเกินไปจนทำให้ไม่สามารถทำนายชุดข้อมูลที่มีความเป็นไปได้ภายในเงื่อนไขน้อยเกินไป หรืออาจเพิ่มประเด็นวิจัยเพื่อเพิ่มความแม่นยำในการทำนายผล เช่น การปรับโมเดลหรือเปลี่ยนแปลงโครงสร้างของชุดข้อมูลฝึกสอน เป็นต้น

| | | | |
|---|---|---|---|
| **Project Title** | Researching for developing training set with artificial neural network technology based on firewall rules | | |
| **Student** | Thitichote | Chaimuang | Student ID 60070019 |
| | Pipatboon | Buddhakul | Student ID 60070065 |
| **Degree** | วิทยาศาสตรบัณฑิต | | |
| **Program** | เทคโนโลยีสารสนเทศ | | |
| **Academic Year** | 2020 | | |
| **Advisor** | ผู้ช่วยศาสตราจารย์ อัครินทร์ คุณกิตติ | | |

# ABSTRACT

This researching project we create DNN model and packer generator for development of a train set which was designed based on firewall rules. We are mainly focused to create most efficient training set that assess our train sets are the less packet, the less train time, and more accuracy. We have created train set by 8 values and made hypotheses under different condition consist classifying equal train set classification and equal ratio classification, then we evaluate and analysis the result of the model. In the accuracy term we found that if there are multiple rules or the more packet we used. the learning rate will decrease overtime, but the classifying Equal train set have less fall rate than the Equal ratio classification. we guess that the reason is each rule divided by ratio has too much different on allow or deny and will cause the learning factor model to become worse, so the false positive and false negative on the classifying by ratio has very high.

In the term of analysis, we considered the most appropriate point of classifying by ratio use less packet which can provided the same accuracy as classifying by equal sample, and less packet mean the less training time model used.

Next researching we will focus on third train set classifying algorithm which can avoid the problem of the classifying by ratio. The threat we found is the vary of the rule set, if the number of possible packets is not enough to generate ratio, so we cannot provide the packet based on the rule.

# กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สาเร็จลุล่วงไปได้ด้วยดี ทางผู้จัดทาขอขอบพระคุณเป็นอย่างสูงกับความกรุณาช่วยเหลือและการให้คาปรึกษาของ ผู้ช่วยศาสตราจารย์อัครินทร์ คุณกิตติ ที่ช่วยชี้แนะแนวทาง ตั้งแต่วันแรกถึงวันสุดท้าย และขอบพระคุณอาจารย์ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระ จอมเกล้าเจ้าคุณทหารลาดกระบังทุกๆท่าน ที่ให้ความรู้อันเป็นประโยชน์ยิ่ง ต่อการพัฒนาต่อยอดองค์ความรู้

ขอขอบคุณครอบครัวที่ให้การสนับสนุนอย่างดีเสมอมา

ขอขอบคุณคู่โครงงานที่อดทนและร่วมแรงร่วมใจช่วยกันมาจนถึงทุกวันนี้

ฐิติโชติ ใจเมือง

พิพัฒน์บุญ พุทธคุณ

# สารบัญ

**หน้า**

# สารบัญ (ต่อ)

**หน้า**

# สารบัญตาราง

**หน้า**

**ตารางที่**

# สารบัญรูป

**หน้า**

**รูปที่**

# สารบัญรูป (ต่อ)

**หน้า**

**รูปที่**