การรวมโหนดในนิวรอลเน็ตเวิร์กหลายโครงข่ายบนชุดข้อมูลขนาดใหญ่

นายต้า เกียรติไกรวัลศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

COMBINING NODES IN MULTIPLE NEURAL NETWORKS ON LARGE DATASETS

Mr. Ta Kiatkaiwansiri

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2013
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การรวมโหนดในนิวรอลเน็ตเวิร์กหลายโครงข่ายบนชุดข้อมูลขนาดใหญ่
โดย	นายต้า เกียรติไกรวัลศิริ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.นัทที นิภานันท์)

.....กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.ชลวิช นัทธี)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ต้า เกียรติไกรวัลศิริ : การรวมโหนดในนิวรอลเน็ตเวิร์กหลายโครงข่ายบนชุดข้อมูลขนาดใหญ่. (COMBINING NODES IN MULTIPLE NEURAL NETWORKS ON LARGE DATASETS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.สุกรี สินธุภิญโญ, 52 หน้า.

การสอนนิวรอลเน็ตเวิร์กด้วยชุดข้อมูลขนาดใหญ่จำเป็นต้องใช้เวลามาก และเทคนิคที่นิยมใช้ในลดเวลาในการสอนลงเช่นการลดขนาดชุดข้อมูลให้มีขนาดเล็กลงก่อนนำไปสอนนั้น ไม่สามารถรับประกันได้ว่านิวรอลเน็ตเวิร์กที่ได้รับการสอนนั้นจะมีประสิทธิภาพเท่ากับนิวรอลเน็ตเวิร์กที่สอนด้วยชุดข้อมูลขนาดใหญ่โดยตรงหรือไม่ ในการวิจัยนี้จะนำเสนอเทคนิคในการลดเวลาในการสอนนิวรอลเน็ตเวิร์กบนชุดข้อมูลขนาดใหญ่ลงด้วยการแบ่งชุดข้อมูลขนาดใหญ่ออกเป็นชุดข้อมูลขนาดเล็กหลายชุด จากนั้นนำชุดข้อมูลย่อยเหล่านั้นไปสอนนิวรอลเน็ตเวิร์กที่มีโครงสร้างคล้ายกันหลายโครงข่าย จากนั้นจึงนำองค์ความรู้ในนิวรอลเน็ตเวิร์กเหล่านั้นมารวมกันอีกครั้ง จากผลการทดลองแสดงให้เห็นว่าเทคนิคนี้สามารถลดเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กลงได้อย่างมาก โดยที่ประสิทธิภาพยังคงเทียบเท่ากับนิวรอลเน็ตเวิร์กที่สอนด้วยชุดข้อมูลขนาดใหญ่ทั้งชุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2556

5570488021 : MAJOR COMPUTER SCIENCE
KEYWORDS: ARTIFICIAL NEURAL NETWORK / LARGE DATESET / MULTIVARIATE REGRESSION

TA KIATKAIWANSIRI: COMBINING NODES IN MULTIPLE NEURAL NETWORKS ON LARGE DATASETS. ADVISOR: ASST. PROF. SUKREE SINTHUPINYO, Ph.D., 52 pp.

Training a neural network on large dataset needs a long training time. And a technique to reduce a training time by resampling dataset cannot guarantee a neural network's performance. In this research, we propose a technique to reduce a training time using multiple networks. Our approach divides a large dataset into n subsets. We use those subsets to train multiple same-structure neural networks. Finally, we combine knowledge in multiple neural networks into one network. Results from experiments show that our technique can reduce a training time and preserve performance as a single network trained by the whole dataset.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department: Computer Engineering Student's Signature .....
Field of Study: Computer Science Advisor's Signature .....
Academic Year: 2013

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบพระคุณ ท่านผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ซึ่งกรุณารับข้าพเจ้าเข้าเป็นศิษย์ในที่ปรึกษา ผู้ซึ่งสละเวลา ให้ความรู้ ให้คำแนะนำ ให้คำปรึกษา ทำให้การจัดทำวิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดี ข้าพเจ้าขอขอบคุณคณาจารย์ทุกท่านที่ให้การสั่งสอนข้าพเจ้าตลอดระยะเวลาการศึกษา

ข้าพเจ้าขอขอบพระคุณ ท่านศาสตราจารย์ ดร.บุญเสริม กิจศิริกุญ ประธานกรรมการสอบวิทยานิพนธ์ ท่านผู้ช่วยศาสตราจารย์ ดร.นัทที นิภานันท์ กรรมการ และ ท่านผู้ช่วยศาสตราจารย์ ดร.ชลวิช นัทธี กรรมการภายนอกมหาวิทยาลัย ที่กรุณาสละเวลา ให้คำแนะนำ ชี้แนะ ทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ข้าพเจ้าขอขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัว ทีให้การสนับสนุน และให้กำลังใจข้าพเจ้าเสมอมา

สุดท้ายนี้ ข้าพเจ้าขอขอบคุณทุกๆท่านที่คอยให้กำลังใจ ให้ความช่วยเหลือ และสนับสนุนข้าพเจ้าเสมอมา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญภาพ ..... ฌ

สารบัญตาราง ..... ฎ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ 1

1.3 ขอบเขตของการวิจัย 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 2

1.5 ขั้นตอนการดำเนินงาน 2

1.6 ผลงานตีพิมพ์ 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 ทฤษฎีที่เกี่ยวข้อง 4

2.1.1 โครงข่ายประสาทเทียม (Artificial Neural Network) 4

2.1.2 การหาค่าต่ำสุดด้วยวิธี Nelder–Mead 11

2.1.3 การทดสอบสมมติฐาน (Tests of hypothesis) 14

2.2 งานวิจัยที่เกี่ยวข้อง 15

บทที่ 3 แนวคิดและวิธีดำเนินงานวิจัย 17

3.1 แบ่งชุดข้อมูล 17

3.2 สอนนิวรอลเน็ตเวิร์ก 18

3.3 หาโครงข่ายที่ดีที่สุด 19

3.4 รวมกลุ่มโหนดซ่อน 20

3.4.1 การวิเคราะห์ความคล้าย 20

3.4.2 การรวมกลุ่ม 20

3.5 รวมโหนดซ่อน 22

หน้า

3.6 เพิ่มโหนดซ่อน 23
3.4.1 เพิ่มโหนดซ่อน 23
3.6.2 ปรับค่าถ่วงน้ำหนัก 24
3.6.3 สอนชั้นแสดงผลใหม่ 24
3.7 วัดผล 25
3.7.1 วัดความแม่นยำ 25
3.7.2 วัดผลกระทบของการแบ่งกลุ่ม 25
3.8 เปรียบเทียบผล 26
บทที่ 4 การทดลองและผลการทดลอง 27
4.1 ชุดข้อมูลที่ใช้ 27
4.2 เครื่องมือที่ใช้ 27
4.3 ขั้นตอนการทดลอง 27
4.4 ผลการทดลอง 28
4.4.1 ความถูกต้อง 28
4.4.2 เวลาที่ใช้ในการสอน 39
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 45
5.1 สรุป 45
5.2 แนวทางการพัฒนา 46
รายการอ้างอิง 47
ประวัติผู้เขียนวิทยานิพนธ์ 52

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

หน้า

รูปที่ 1 แสดงเซลล์ประสาท[1] 4
รูปที่ 2 แสดงลักษณะของเพอร์เซปตรอน[2] 4
รูปที่ 3 แสดงระนาบตัดสินใจของเพอร์เซปตรอน 5
รูปที่ 4 แสดงระนาบของเพอร์เซปตรอนในปริภูมิ n มิติ 6
รูปที่ 5 ตัวอย่างฟังก์ชันกระตุ้น 6
รูปที่ 6 แสดงโครงสร้างของโครงข่ายเพอร์เซปตรอนหลายชั้น[2] 7
รูปที่ 7 แสดงโหนดซ่อนของโครงข่ายเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน 1 ชั้น 7
รูปที่ 8 แสดงเอาต์พุตของขั้นซ่อนก่อนผ่านฟังก์ชันกระตุ้น 8
รูปที่ 9 แสดงเอาต์พุตหลังจากผ่านฟังก์ชันกระตุ้น 8
รูปที่ 10 แสดงเอาต์พุตจากเอาต์พุตโหนด 9
รูปที่ 11 แสดงการเปรียบเทียบสัญญาณจากเอาต์พุตโหนด 9
รูปที่ 12 แสดงค่าความผิดพลาดเทียบกับค่าน้ำหนักและค่าขีดแบ่งของนิลรอลเน็ตเวิร์ก 10
รูปที่ 13 แสดงการกำหนดการหยุดสอนเมื่อค่าความผิดพลาดต่ำกว่าจุดที่กำหนด 10
รูปที่ 14 แสดงการกำหนดการหยุดสอนเมื่อค่าความผิดพลาดบนชุดทดสอบไม่ลดลงต่อเนื่อง 10
รูปที่ 15 แสดงการขยับรูปเรขาคณิตด้วยการสะท้อน 12
รูปที่ 16 แสดงการขยายรูปทรงเรขาคณิต 12
รูปที่ 17 แสดงการปรับโดยยึดจุดใหม่ และยึดจุดเดิมเป็นหลัก 13
รูปที่ 18 แสดงขั้นตอนในงานวิจัย 17
รูปที่ 19 แสดงการแบ่งชุดข้อมูล 18
รูปที่ 20 แสดงขั้นตอนการสอนนิวรอลเน็ตเวิร์กของกลุ่มหลักที่ 1 18
รูปที่ 21 แสดงการหาโครงข่ายที่ดีที่สุด 19
รูปที่ 22 แสดงการแตกตัวอย่างของโหนดซ่อน 20
รูปที่ 23 แสดงการรวมกลุ่มโหนดซ่อน 21
รูปที่ 24 แสดงการทำงานของฟังก์ชันระหว่างค่าถ่วงน้ำหนักและค่าความผิดพลาด 22
รูปที่ 25 แสดงการรวมโหนดซ่อนและสร้างนิวรอลเน็ตเวิร์กโครงข่ายใหม่ 23

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 26 แสดงการเพิ่มโหนดซ่อน 23
รูปที่ 27 แสดงขั้นตอนการชั้นแสดงผล 25
รูปที่ 28 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล Covertype 30
รูปที่ 29 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล Synthesized 32
รูปที่ 30 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล MiniBooNE 34
รูปที่ 31 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล Statlog 36
รูปที่ 32 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล King-Rook vs. King 38
รูปที่ 33 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Covertype 40
รูปที่ 34 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Synthesized 41
รูปที่ 35 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล MiniBooNE 42
รูปที่ 36 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Statlog 43
รูปที่ 37 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล King-Rook vs. King .. 44

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 ตัวอย่างความคล้ายของการวางตัวของโหนดซ่อนระหว่างโครงข่ายหลักและรอง 21

ตารางที่ 2 แสดงชุดข้อมูล 27

ตารางที่ 3 แสดงค่าความผิดพลาดบนชุดข้อมูล Covertype ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 28

ตารางที่ 4 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Covertype 29

ตารางที่ 5 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Covertype 30

ตารางที่ 6 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล Covertype 30

ตารางที่ 7 แสดงค่าความผิดพลาดบนชุดข้อมูล Synthesized ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 31

ตารางที่ 8 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Synthesized 31

ตารางที่ 9 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Synthesized 32

ตารางที่ 10 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล Synthesized 32

ตารางที่ 11 แสดงค่าความผิดพลาดบนชุดข้อมูล MiniBooNE ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 33

ตารางที่ 12 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล MiniBooNE 33

ตารางที่ 13 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล MiniBooNE 34

ตารางที่ 14 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล MiniBooNE 34

ตารางที่ 15 แสดงค่าความผิดพลาดบนชุดข้อมูล Statlog ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 35

ตารางที่ 16 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Statlog35

ตารางที่ 17 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล MiniBooNE 36

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 18 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล Statlog 36

ตารางที่ 19 แสดงค่าความผิดพลาดบนชุดข้อมูล King-Rook vs. King ของนิวรอลเน็ตเวิร์ก โครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 37

ตารางที่ 20 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล King-Rook vs. King 37

ตารางที่ 21 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล King-Rook vs. King 38

ตารางที่ 22 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล King-Rook vs. King 38

ตารางที่ 23 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Covertype ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 39

ตารางที่ 24 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Synthesized ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 40

ตารางที่ 25 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล MiniBooNE ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 41

ตารางที่ 26 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Statlog ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 42

ตารางที่ 27 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล King-Rook vs. King ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม 43

ตารางที่ 28 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล Covertype 49

ตารางที่ 29 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล Synthesized 49

ตารางที่ 30 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล MiniBooNE 50

ตารางที่ 31 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล Statlog 50

ตารางที่ 32 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล King-Rook vs. King 51

ตารางที่ 33 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กที่มีโหนดซ่อนเท่ากับนิวรอลเน็ตเวิร์กที่ได้จากการรวมและสอนบนชุดข้อมูล King-Rook vs. King โดยตรง 51

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การทำเหมืองข้อมูลกลายเป็นส่วนหนึ่งในหลายๆอุตสาหกรรม ประกอบกับแนวโน้มที่ข้อมูลในปัจจุบันมีขนานใหญ่ขึ้น ทำให้การทำเหมืองข้อมูลกับข้อมูลขนาดใหญ่เป็นเรื่องที่สำคัญ โดยการทำเหมืองข้อมูลกับข้อมูลขนาดใหญ่สามารถทำได้ด้วยหลายๆเทคนิค และหนึ่งในเทคนิคที่เป็นที่นิยมใช้คือนิวรอลเน็ตเวิร์ก เนื่องจากนิวรอลเน็ตเวิร์กมีความยืดหยุ่นในการทำงานกับข้อมูลที่มีความซับซ้อน และสามารถทำงานได้อย่างมีประสิทธิภาพแม้ข้อมูลนั้นๆจะมีสัญญาณรบกวน แต่การใช้นิวรอลเน็ตเวิร์กในการเรียนรู้ชุดข้อมูลขนาดใหญ่นั้นมีข้อเสียคือ ใช้เวลาในการเรียนรู้และทรัพยากรในการเรียนรู้มาก ดังนั้นจึงมีผู้เสนอเทคนิคต่างๆเพื่อทำให้การเรียนรู้ชุดข้อมูลขนาดใหญ่ด้วยนิวรอลเน็ตเวิร์กนั้น ทำได้รวดเร็วและประหยัดทรัพยากรมากขึ้น เช่น เทคนิคการประมวลผลข้อมูลให้มีมิติและขนาดเล็กลงก่อนการเรียนรู้ เทคนิคการจัดกลุ่มข้อมูลก่อนทำการแยกประมวลผลด้วยนิวรอลเน็ตเวิร์ก ซึ่งแต่ละเทคนิคก็มีข้อดีข้อเสียเป็นของตัวเอง และมีเทคนิคหนึ่งที่น่าสนใจคือ การใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายช่วยกันเรียนรู้บนชุดข้อมูลขนาดใหญ่ ซึ่งมีข้อดีคือได้เรียนรู้ชุดข้อมูลขนาดใหญ่ทั้งชุดไม่ใช่แค่เพียงส่วนใดส่วนหนึ่ง และ การเรียนรู้เป็นการแบ่งข้อมูลขนาดใหญ่เป็นข้อมูลขนาดย่อยก่อนการเรียนรู้ ทำให้สามารถเรียนรู้ได้เร็วขึ้น โดยขั้นตอนการเรียนรู้ข้อมูลขนาดใหญ่โดยใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายทำได้โดย แบ่งข้อมูลขนาดใหญ่เป็นชุดข้อมูลย่อยๆเพื่อให้นิวรอลเน็ตเวิร์กหลายโครงข่ายช่วยกันเรียนรู้ จากนั้นทำการรวมนิวรอลเน็ตเวิร์กหลายๆโครงข่ายเข้าด้วยกัน โดยงานวิจัยที่เกี่ยวข้องกับการประยุกต์ใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายบนชุดข้อมูลขนาดใหญ่นั้นยังมีข้อเสียอยู่คือ ยังขาดขั้นตอนที่เหมาะสมในการวิเคราะห์และการรวมนิวรอลเน็ตเวิร์กเข้าด้วยกัน

ในงานวิจัยนี้จะนำเสนอขั้นตอนวิธีที่เหมาะสมในการสอน การวิเคราะห์และการรวมโหนดซ่อนในนิวรอลเน็ตเวิร์กหลายโครงข่ายเข้าด้วยกัน โดยจะทำการวิเคราะห์โหนดซ่อนเพื่อค้นหาโหนดที่มีลักษณะคล้ายกันในนิวรอลเน็ตเวิร์กแต่ละโครงข่าย และจะทำการศึกษาหาวิธีการที่เหมาะสมในการรวมโหนดซ่อนเหล่านั้นเข้าด้วยกันเพื่อเพิ่มประสิทธิภาพให้การประยุกต์ใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายสามารถเรียนรู้ชุดข้อมูลขนาดใหญ่ได้รวดเร็วขึ้นโดยที่ประสิทธิภาพความถูกต้องยังคงไม่ด้อยกว่าการที่นิวรอลเน็ตเวิร์กเรียนรู้ชุดข้อมูลขนาดใหญ่โดยตรง

1.2 วัตถุประสงค์

งานวิจัยนี้นำเสนอวิธีวิเคราะห์โหนดและการรวมโหนดในนิวรอลเน็ตเวิร์กหลายโครงข่ายเพื่อลดเวลาในการสอนนิวรอลเน็ตเวิร์กบนชุดข้อมูลขนาดใหญ่ โดยการวิเคราะห์โหนดจะวิเคราะห์จากความคล้ายคลึงของการแตกกลุ่มตัวอย่าง จากนั้นจึงทำการรวมโหนดซ่องของนิวรอลเน็ตเวิร์กโครงข่ายที่มีการวางตัวของโหนดซ่อนคล้ายกันเข้าด้วยกัน และทำการวัดความถูกต้องและเวลาที่ใช้ในการสอนของนิวรอลเน็ตเวิร์กที่ได้จากการรวมกับความถูกต้องและเวลาที่ใช้การสอนของนิวรอลเน็ตเวิร์กที่ได้จากการเรียนชุดข้อมูลขนาดใหญ่โดยตรงเพื่อเปรียบเทียบประสิทธิภาพ

1.3 ขอบเขตของการวิจัย

งานวิจัยนี้เป็นงานวิจัยที่นำเสนอวิธีการวิเคราะห์การวางตัวของโหนดของนิวรอลเน็ตเวิร์ก และวิธีการรวมโหนดในนิวรอลเน็ตเวิร์กหลายโครงข่ายเพื่อการประยุกต์บนชุดข้อมูลขนาดใหญ่ โดยขอบเขตของการดำเนินงานมีดังนี้

1.3.1 นิวรอลเน็ตเวิร์กหลายโครงข่ายที่ใช้ในงานวิจัยนี้เป็นนิวรอลเน็ตเวิร์กที่มีโครงสร้างเหมือนกัน

1.3.2 นิวรอลเน็ตเวิร์กที่ใช้ในงานวิจัยนี้เป็นนิวรอลเน็ตเวิร์กที่ประกอบด้วย 3 ชั้น ได้แก่ ชั้นอินพุต ชั้นซ่อน 1 ชั้น และชั้นเอาต์พุต

1.3.3 โหนดที่จะทำการวิเคราะห์และนำมารวมคือโหนดที่มาจากชั้นของชั้นซ่อน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้ทราบถึงวิธีการในการวิเคราะห์การวางตัวของโหนดของนิวรอลเน็ตเวิร์กภายในชุดข้อมูล

1.4.2 ได้ทราบถึงวิธีการรวมโหนดภายในนิวรอลเน็ตเวิร์กหลายตัวเข้าด้วยกัน

1.4.3 ได้นิวรอลเน็ตเวิร์กที่เกิดจากการรวมนิวรอลเน็ตเวิร์กหลายโครงข่ายที่ประสิทธิภาพความถูกต้องไม่ด้อยกว่านิวรอลเน็ตเวิร์กที่สอนโดยชุดข้อมูลขนาดใหญ่โดยตรง

1.4.4 เป็นแนวทางในการศึกษาวิจัยทางด้านการประยุกต์ใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายบนชุดข้อมูลขนาดใหญ่ต่อไป

1.5 ขั้นตอนการดำเนินงาน

1.5.1 ศึกษางานวิจัยเกี่ยวกับการประยุกต์ใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายบนชุดข้อมูลขนาดใหญ่

1.5.2 ศึกษาวิธีการและทดลองการรวมค่าน้ำหนักจากนิวรอลเน็ตเวิร์กหลายโครงข่าย

1.5.3 ทดสอบความถูกต้องของวิธีรวมค่าน้ำหนักจากนิวรอลเน็ตเวิร์กหลายโครงข่าย

1.5.4 จัดทำรายงาน สรุป และวิทยานิพนธ์

CHULALONGKORN UNIVERSITY

1.6 ผลงานตีพิมพ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้นำเสนอในการประชุมวิชาการ ดังนี้

Ta Kiatkaiwansiri and Sukree Sinthupinyo, Combining Nodes in Multiple Neural Network on Large Datasets, 2014 Fourth International Conference on Digital Information and Communication Technology and it's Applications (DICTAP), University of the Thai Chamber of Commerce, Thailand, May 6-8, 2014

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

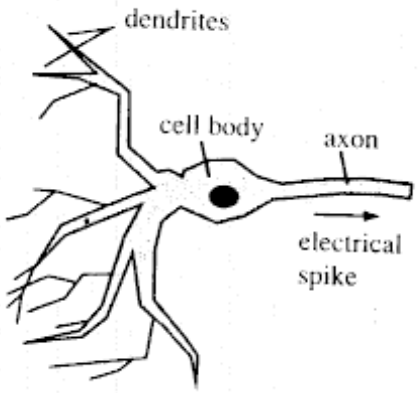
บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

ในงานวิจัยนี้มีทฤษฎีที่สำคัญที่เกี่ยวข้องดังนี้

2.1.1 โครงข่ายประสาทเทียม (Artificial Neural Network)

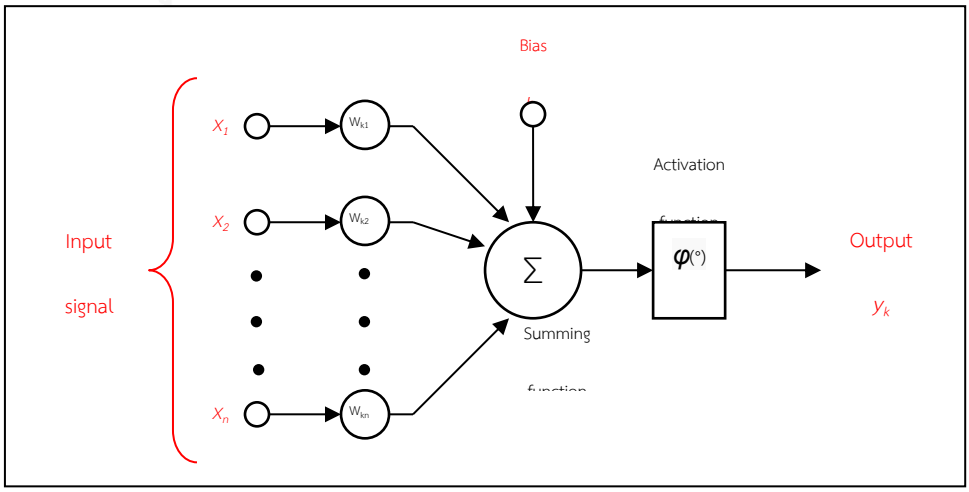
โครงข่ายประสาทเทียมหรือนิวรอลเน็ตเวิร์ก เป็นการจำลองการทำงานบางส่วนของสมองมนุษย์ เซลล์ประสาท (neuron) ในสมองของคนเราประกอบด้วยนิวเคลียส (nucleus) ตัวเซลล์ (cell body) ใยประสาทนำเข้า (dendrite) แกนประสาทนำออก (axon) ดังที่แสดงแสดงในรูปที่ 1



รูปที่ 1 แสดงเซลล์ประสาท[1]

2.1.1.1 เพอร์เซปตรอน (Perceptron)

เพอร์เซปตรอนเป็นข่ายงานประสาทเทียมที่มีเซลล์ประสาทเพียงโหนดเดียวดังที่แสดงในรูปที่ 2

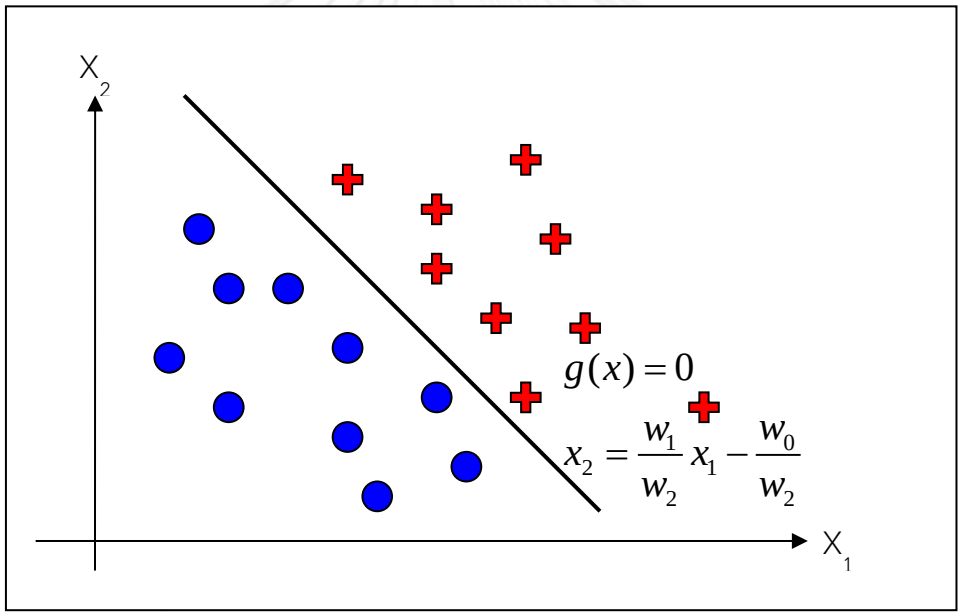


รูปที่ 2 แสดงลักษณะของเพอร์เซปตรอน[2]

เพอร์เซปตรอนจะรับค่าอินพุตเข้ามาแล้วคำนวณหาผลรวมเชิงเส้น (linear combination) โดยการนำค่าอินพุต (x1, x2, ..., xn) มาคูณกับค่าถ่วงน้ำหนัก w1, w2, ...,wn จากนั้นนำมาบวกด้วยค่ากำหนด (bias) จากนั้นจึงนำน่าที่คำนวณได้ไปผ่านฟังก์ชันกระตุ้น (activate function) โดยเอาต์พุตของเพอร์เซปตรอนจะเป็น 1 เมื่อมีการกระตุ้น (ผลรวมเชิงเส้นมีค่ามากกว่า 0) และเป็น 0 เมื่อไม่มีการกระตุ้น โดยเอาต์พุตของเพอร์เซปตรอนสามารถแสดงได้ดังในสมการที่ 1

$$y(x_1, x_2, ..., x_n) = \begin{cases} 1 if\ b + w_1x_1 + w_2x_2 + ... + w_nx_n > 0 \\ 0 if\ b + w_1x_1 + w_2x_2 + ... + w_nx_n < 0 \end{cases} \quad (1)$$

หากมีอินพุตเข้าสู่เพอร์เซปตรอน 2 ตัว เราสามารถมองเพอร์เซปตรอนเป็นเส้นตรงแบ่งอินพุตเป็น 0 และ 1 ได้ แต่ในกรณีที่อินพุตมากกว่าสองมิติเพอร์เซปตรอนจะเป็นระนาบตัดสินใจหลายมิติ (hyperplane decision surface)



รูปที่ 3 แสดงระนาบตัดสินใจของเพอร์เซปตรอน

Chulalongkorn University

และนอกจากนั้นเราสามารถมองเพอร์เซปตรอนเป็นระนาบในปริภูมิ n มิติได้ โดยหากสมการของเพอร์เซปตรอนเป็นดังสมการที่ 2

$$y(x_1, x_2, ..., x_n) = w_1x_1 + w_2x_2 + ... + w_nx_n + w_0 \quad (2)$$

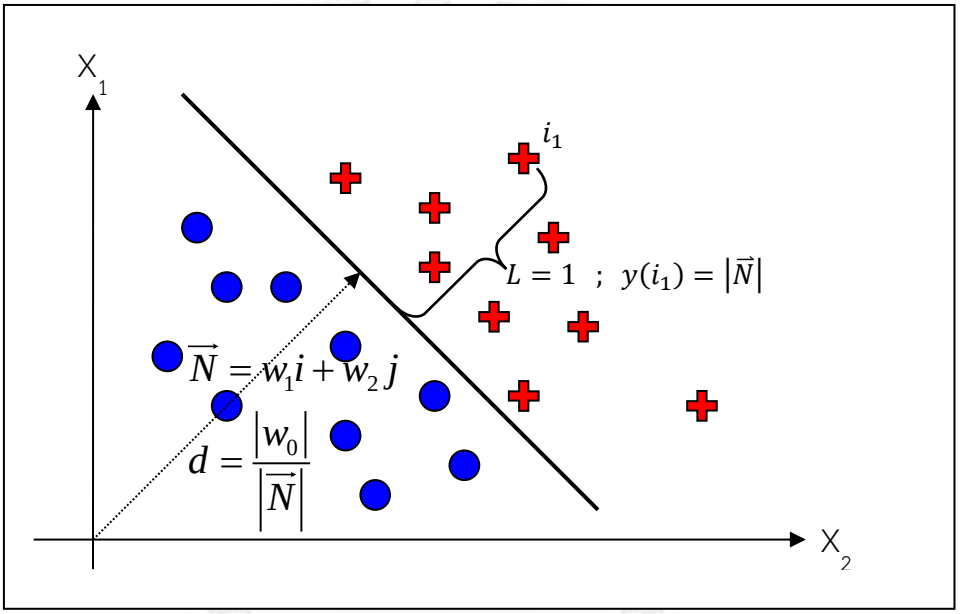
โดยเวกเตอร์ปกติ (Normal Vector) ที่ลากจากจุดกำเนิดจนถึงระนาบและตั้งฉากกับระนาบสามารถแสดงได้ดังในสมการที่ 3[3]

$$\vec{N} = w_1 i + w_2 j + w_3 k + \ldots \quad (3)$$

และระยะห่างจากระนาบของเพอร์เซปตรอนและจุดกำเนิดสามารถแสดงได้ดังในสมการที่ 4

$$d = \frac{|w_0|}{|\vec{N}|} \quad (4)$$

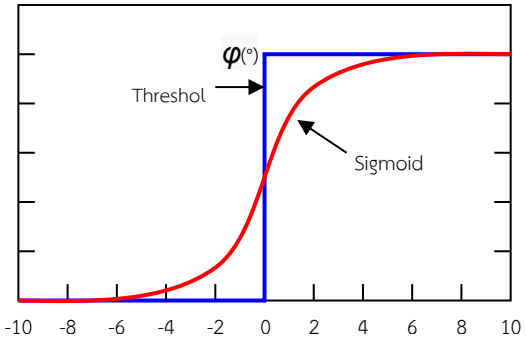
โดยขนาดของเวกเตอร์ปกตินั้นแสดงถึงค่าที่ได้จากคำนวณผลรวมเชิงเส้นของเพอร์เซปตรอนเมื่อข้อมูลตัวนั้นอยู่ห่างจากระนาบของเพอร์เซปตรอน 1 หน่วย



รูปที่ 4 แสดงระนาบของเพอร์เซปตรอนในปริภูมิ n มิติ

2.1.1.2 ฟังก์ชันกระตุ้น (activation function)

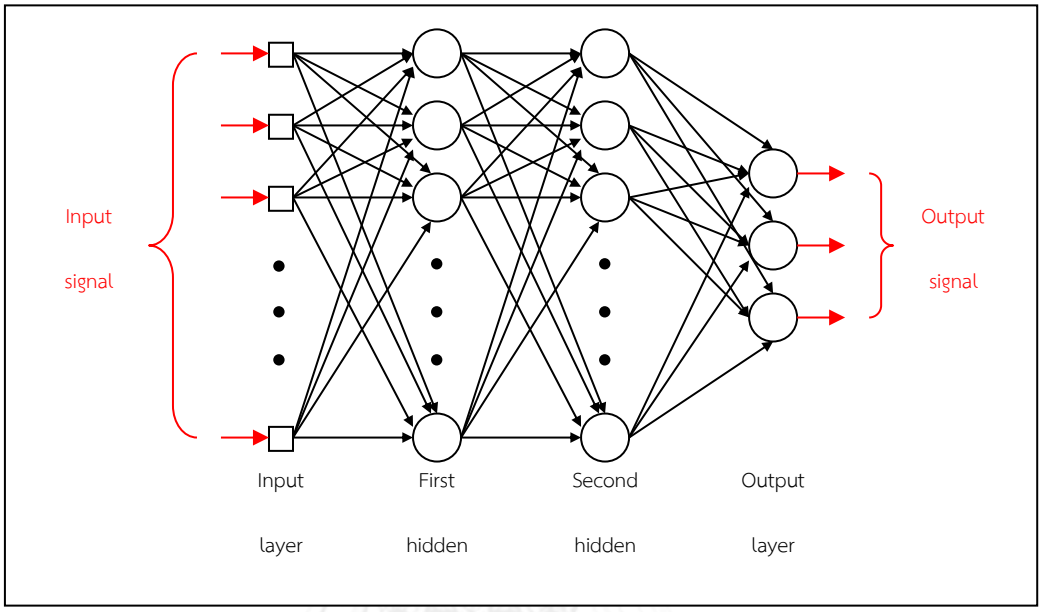
ฟังก์ชันกระตุ้นเป็นฟังก์ชันที่กำหนดลักษณะของเอาต์พุตที่ออกจากเพอร์เซปตรอนตัวนั้นๆ โดยตัวอย่างลักษณะของฟังก์ชันกระตุ้นสามารถแสดงได้ดังรูปที่ 5



รูปที่ 5 ตัวอย่างฟังก์ชันกระตุ้น

2.1.1.3 โครงข่ายเพอร์เซปตรอนหลายชั้น (Multilayer perceptrons)

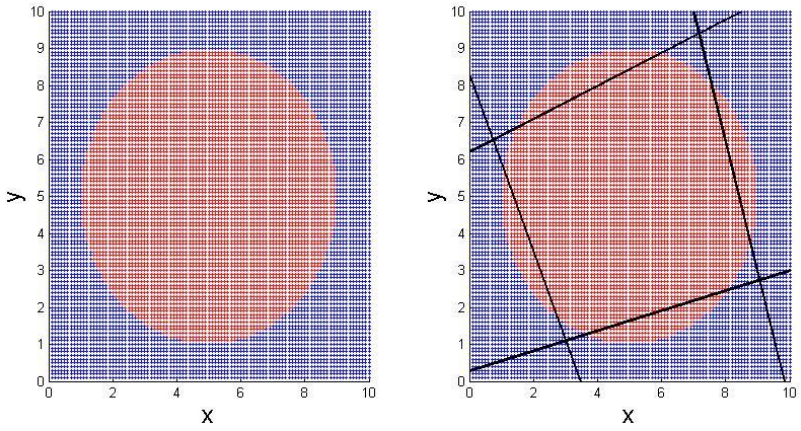
โครงข่ายเพอร์เซปตรอนหลายชั้นเป็นนิวรอลเน็ตเวิร์กที่ประกอบด้วยชั้นของเพอร์เซปตรอนหลายชั้น โดยในแต่ละชั้นจะประกอบด้วยเพอร์เซปตรอนจำนวนหลายโหนด ดังรูปที่ 6



รูปที่ 6 แสดงโครงสร้างของโครงข่ายเพอร์เซปตรอนหลายชั้น[2]

โดยเอาต์พุตของโหนดซ่อนในชั้นก่อนหน้าจะกลายมาเป็นอินพุตของโหนดในชั้นปัจจุบัน การทำงานของนิวรอลเน็ตเวิร์กจะทำงานโดยการเริ่มคำนวณค่าตั้งแต่ชั้นแรกและส่งค่าไปตามชั้นซ่อนไปเรื่อยๆทีละชั้นจนถึงชั้นเอาต์พุตซึ่งเป็นชั้นสุดท้ายในที่สุด

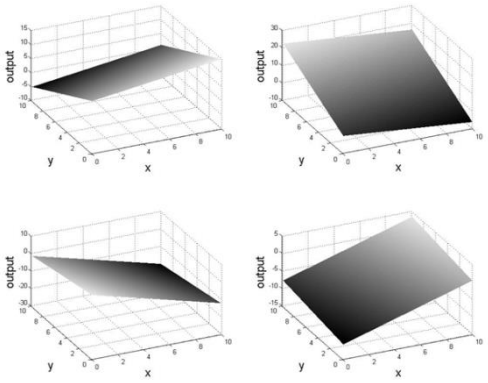
เพอร์เซปตรอนที่อยู่ในชั้นซ่อนของโครงข่ายเพอร์เซปตรอนหลายชั้นสามารถแสดงได้ดังในตัวอย่างในรูปที่ 7



รูปที่ 7 แสดงโหนดซ่อนของโครงข่ายเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน 1 ชั้น

รูปที่ 7 ทางด้านซ้ายแสดงตัวอย่างชุดข้อมูลที่อยู่ภายในระนาบ 2 มิติซึ่งข้อมูลอยู่ในช่วง [0 10] โดยข้อมูลประกอบด้วย 2 ประเภท ข้อมูลประเภทที่ 1 เป็นข้อมูลที่อยู่ภายในช่วงวงกลมตรงกลาง และประเภทที่ 2 เป็นข้อมูลที่อยู่รอบนอก รูปที่ 7 ทางด้านขวาแสดงการวางตัวของโหนดซ่อนของนิวรอลเน็ตเวิร์กที่มีชั้นซ่อน 1 ชั้นและประกอบด้วยโหนดซ่อน 4 โหนดซึ่งเรียนรู้ชุดข้อมูลนี้

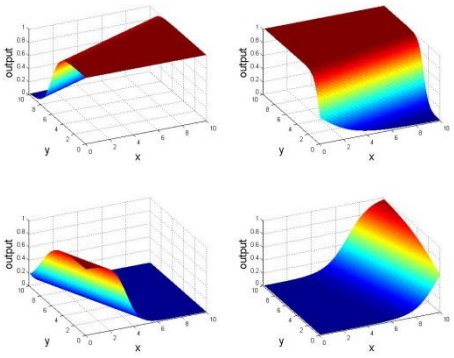
การทำงานของโครงข่ายเพอร์เซปตรอนหลายชั้นหรือนิวรอลเน็ตเวิร์กที่มีชั้นซ่อน 1 ชั้นนั้น จะเริ่มจากการส่งอินพุตเข้าสู่ชั้นซ่อนเพื่อคำนวณหาเอาต์พุต จากนั้นส่งเอาต์พุตของชั้นซ่อนนั้นไปให้เอาต์พุตโหนดเพื่อคำนวณหาผลลัพธ์ของนิวรอลเน็ตเวิร์กอีกครั้ง โดยเอาต์พุตของชั้นซ่อนสามารถแสดงได้ดังในรูปที่ 8



รูปที่ 8 แสดงเอาต์พุตของขั้นซ่อนก่อนผ่านฟังก์ชันกระตุ้น

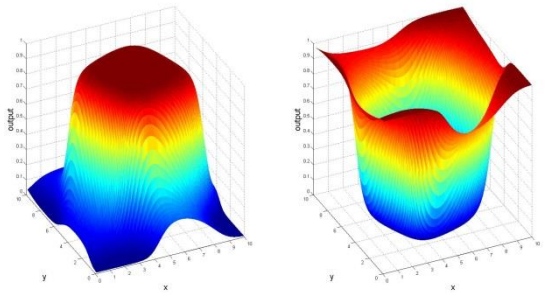
เมื่อชั้นซ่อนคำนวณค่าของอินพุตกับค่าถ่วงน้ำหนักแล้ว เมื่อพล็อตเอาต์พุตทั้งหมดออกมาจะได้เป็นระนาบตามสมการของเพอร์เซปตรอนที่มีลักษณะเป็น linear

หลังจากคำนวณค่าเสร็จแล้ว ก่อนที่จะส่งค่าไปยังเอาต์พุตโหนดนั้น จะทำการเอาเอาต์พุตที่ได้ไปผ่านฟังก์ชันกระตุ้นก่อน โดยเอาต์พุตที่ผ่านฟังก์ชันกระตุ้นแล้วสามารถแสดงได้ดังรูปที่ 9



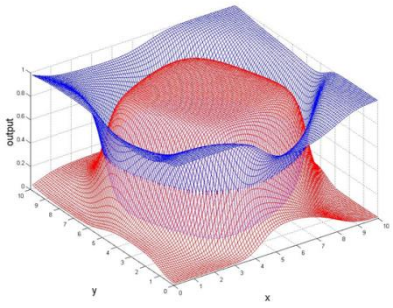
รูปที่ 9 แสดงเอาต์พุตหลังจากผ่านฟังก์ชันกระตุ้น

ฟังก์ชันกระตุ้นจะทำหน้าที่กำหนดลักษณะของเอาต์พุตและเพิ่มความซับซ้อนของเอาต์พุตก่อนที่จะทำการส่งไปยังเอาต์พุตโหนด เอาต์พุตโหนดจะทำการคำนวณค่าผลลัพธ์สุดท้ายจากค่าได้รับมา โดยเอาต์พุตโหนดสามารถแสดงได้ดังในรูปที่ 10



รูปที่ 10 แสดงเอาต์พุตจากเอาต์พุตโหนด

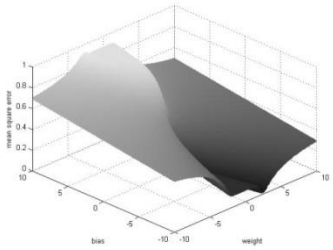
สำหรับนิวรอลเน็ตเวิร์กที่ใช้สำหรับการจำแนกประเภท (Classification Neural Network) การตรวจสอบผลลัพธ์ว่าข้อมูลตัวนั้นๆเป็นประเภทใดนั้น จะทำการตรวจสอบสัญญาณว่าสัญญาณเอาต์พุตจากเอาต์พุตโหนดโหนดใดให้สัญญาณแรงกว่ากัน ดังที่แสดงในรูปที่ 11



รูปที่ 11 แสดงการเปรียบเทียบสัญญาณจากเอาต์พุตโหนด

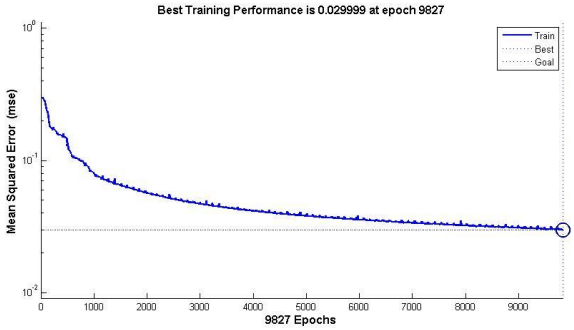
2.1.1.4 การสอนนิวรอลเน็ตเวิร์ก

ในการใช้งานนิวรอลเน็ตเวิร์ก เราจำเป็นต้องทำการปรับค่าน้ำหนักและค่าขีดแบ่งของโหนดทุกโหนดภายในนิวรอลเน็ตเวิร์กให้วางตัวและสร้างพื้นผิวตัดสินใจได้อย่างเหมาะสม โดยในขั้นตอนการสอนนิวรอลเน็ตเวิร์กจะกินเวลาและทรัพยากรมากน้อยตามขนาดของข้อมูลที่จะให้นิวรอลเน็ตเวิร์กเรียนรู้ โดยขั้นตอนการสอนนิวรอลเน็ตเวิร์กเริ่มจากการผ่านชุดข้อมูลสอนทั้งหมดเข้าไปยังนิวรอลเน็ตเวิร์ก จากนั้นคำนวณหาค่าความผิดพลาดของเอาต์พุตของนิวรอลเน็ตเวิร์กกับค่าเป้าหมาย



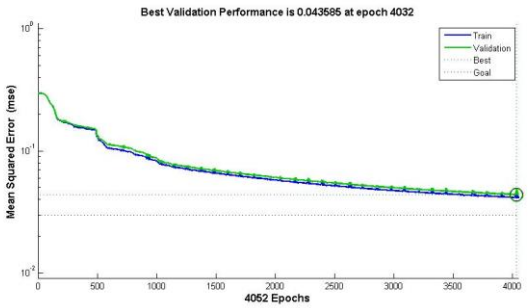
รูปที่ 12 แสดงค่าความผิดพลาดเทียบกับค่าน้ำหนักและค่าขีดแบ่งของนิลรอลเน็ตเวิร์ก

หลังจากทำการคำนวณหาค่าความผิดพลาดได้แล้ว จะทำการคำนวณหาทิศทางการปรับค่าน้ำหนักและค่าขีดแบ่งเพื่อทำให้ค่าความผิดพลาดน้อยลง จากนั้นทำการผ่านชุดข้อมูลสอนอีกครั้งและทำซ้ำกระบวนการจนได้ผลที่พอใจ โดยการกำหนดการหยุดสอนนั้นสามารถกำหนดได้หลายวิธี เช่น การกำหนดเวลาสอนทีจะใช้สอน การกำหนดจำนวนรอบที่จะใช้สอน หรือการกำหนดให้หยุดขั้นตอนการสอนเมื่อค่าความผิดพลาดลดลงต่ำกว่าค่าที่กำหนด



รูปที่ 13 แสดงการกำหนดการหยุดสอนเมื่อค่าความผิดพลาดต่ำกว่าจุดที่กำหนด

แต่หากเรากำหนดให้หยุดสอนเมื่อค่าความผิดพลาดมีค่าต่ำมากๆ นิวรอลเน็ตเวิร์กจะพยายามปรับค่าถ่วงน้ำหนักและค่าขีดแบ่งให้เข้ากับข้อมูลสอนชุดนี้จนอาจทำให้นิวรอลเน็ตเวิร์กที่ได้นั้นสามารถใช้ได้ดีกับเฉพาะข้อมูลที่อยู่ในชุดสอนชุดนี้เท่านั้น ทำให้ไม่สามารถใช้งานได้ดีบนชุดข้อมูลที่ยังไม่เคยพบ (Overfitting) โดยเราสามารถหลีกเลี่ยงการเกิดเหตุการณ์นี้ได้โดยใช้เทคนิคการแบ่งชุดสอนออกเป็นชุดสอนจริงและชุดวัดผล (Validate) โดยเราสามารถกำหนดให้หยุดสอนนิวรอลเน็ตเวิร์กเมื่อค่าความผิดพลาดบนชุดวัดผลไม่ลดลงต่อเนื่องเป็นเวลา n รอบ



รูปที่ 14 แสดงการกำหนดการหยุดสอนเมื่อค่าความผิดพลาดบนชุดทดสอบไม่ลดลงต่อเนื่อง

2.1.2 การหาค่าต่ำสุดด้วยวิธี Nelder–Mead

การหาค่าต่ำสุดด้วยวิธีแบบ Nelder–Mead[4] เป็นวิธีการที่นิยมใช้ในการหาค่าต่ำสุดบนฟังก์ชันที่ไม่ใช่เชิงเส้น โดยการรูปเรขาคณิตขึ้นมา จากนั้นทำการขยับรูปเรขาคณิตนั้นไปในปริภูมของอินพุตเรื่อยๆเพื่อทำการค้นหาบริเวณที่ค่าของฟังก์ชันมีค่าน้อยที่สุด โดยหากอินพุตของฟังก์ชันประกอบด้วย n มิติ จะทำการสร้างรูปเรขาคณิตที่ประกอบด้วย n+1 จุดขึ้นมา เช่น หาอินพุตมี 1 มิติ จะทำการสร้างเส้นตรงที่ประกอบด้วย 2 จุดขึ้นมา หากอินพุตมี 2 มิติจะทำการสร้างรูปสามเลี่ยมที่ประกอบด้วย 3 จุดขึ้นมา หากอินพุตมี 3 มิติ จะทำการสร้างพีระมิดฐานสามเหลี่ยมที่ประกอบด้วย 4 จุดขึ้นมา เป็นต้น

โดยขั้นตอนการทำงานของการหาค่าต่ำสุดด้วยวิธี Nelder–Mead จะประกอบด้วยขั้นตอนดังนี้

1. Order

ทำการเรียงลำดับค่าที่ได้จากการแทนค่าจุดในรูปเรขาคณิตที่สร้างขึ้นเข้าไปในฟังก์ชันจากน้อยไปหามาก

$$f(x_1) < f(x_2) < ..... < f(x_{n+1}) \tag{5}$$

เมื่อ $f(x_i)$ คือค่าที่ได้จากการแทนค่าจุด x_i เข้าไปยังฟังก์ชัน f

2. Reflect

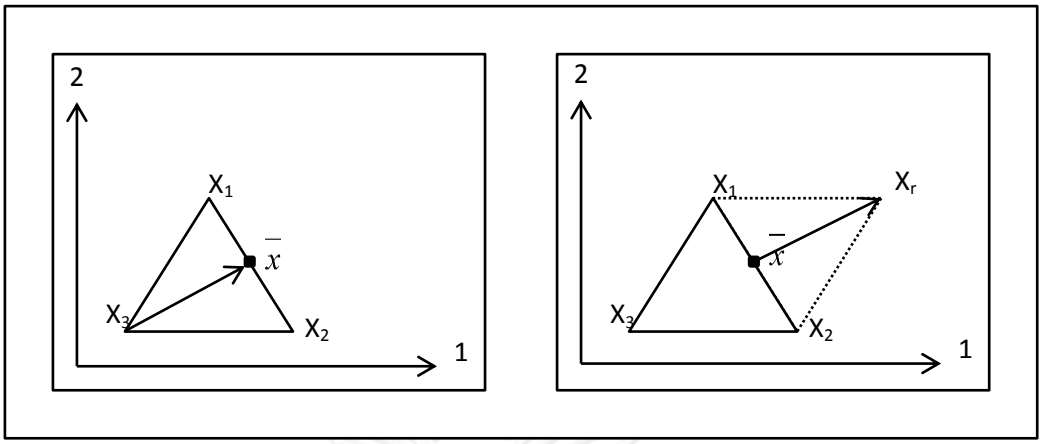
คือการขยับรูปทรงเรขาคณิตที่สร้างขึ้นด้วยการสะท้อนรูปทรงเรขาคณิตจากด้านหนึ่งไปยังอีกด้านหนึ่ง โดยเริ่มจาก

- คำนวณจุดศูนย์กลางของจุดที่ 1 ถึงจุดที่ n
- คำนวณจุดที่ที่จะทำการสะท้อนด้วยการหาทิศทางจากจุด n+1 (จุดที่เมื่อแทนค่าเข้าสู่ฟังก์ชันแล้วมีค่ามากที่สุด) ไปยังจุดศูนย์กลางที่คำนวณได้ ดังสมการที่ 6

$$x_r = \bar{x} + \rho(\bar{x} - x_{n+1}) \tag{6}$$

เมื่อ x_r คือจุดใหม่ที่ได้จากการสะท้อน $\bar{x}$ คือจุดศูนย์กลางชองจุดที่ 1 ถึง n และ ρ คือค่าพารามิเตอร์ที่ใช้ในการปรับการสะท้อน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 15 แสดงการขยับรูปเรขาคณิตด้วยการสะท้อน

- หาก $f(x_1) < f(x_r) < f(x_n)$ ให้ทำการยอมรับจุดใหม่แทนจุดเดิม

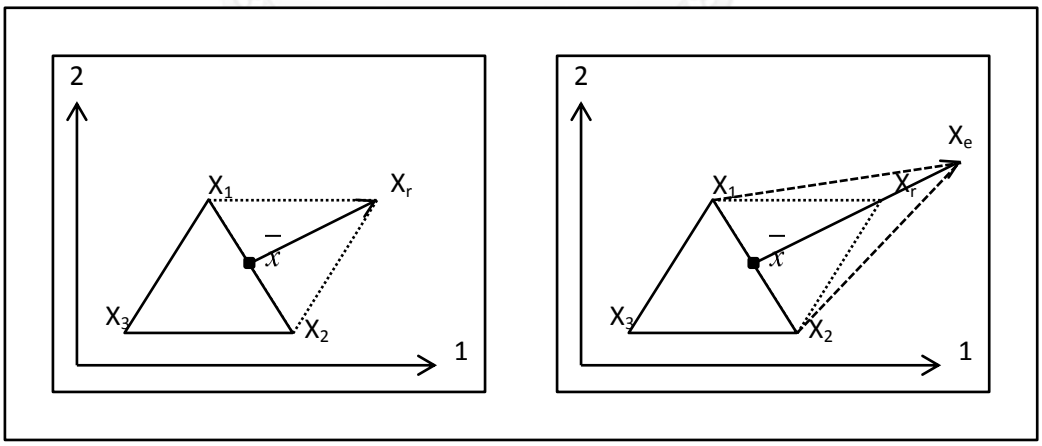
3. Expand

คือการขยายรูปเรขาคณิตไปยังทิศทางที่ดีขึ้น หากจุดที่จะทำการขยับที่คำนวณในขั้นตอน Reflect มีค่าดีกว่าจุดที่มีอยู่ทั้งหมด เราจะทำการปรับรูปเรขาคณิตไปในทิศทางนั้นมากขึ้น

- หาก $f(x_r) < f(x_1)$ จะทำการคำนวณหาค่าจุดที่จะขยายด้วยสมการที่ 7

$$x_e = \bar{x} + \chi(x_r - \bar{x}) \tag{7}$$

เมื่อ x_e คือจุดที่จะทำการขยายไป และ χ คือพารามิเตอร์ที่ใช้สำหรับการขยาย



รูปที่ 16 แสดงการขยายรูปทรงเรขาคณิต

- หาก $f(x_e) < f(x_r)$ ให้ยอมรับจุด x_e เป็นจุดใหม่ หากไม่ ให้ยอมรับ x_r เป็นจุดใหม่แทน

4. Contract

หากจุดใหม่ที่จะทำการปรับไม่ดีกว่าจุดที่แย่ที่สุดในทุกจุดในปัจจุบัน จะมีลักษณะการปรับ 2 แบบดังนี้

4.1 ปรับโดยอิงจากจุดใหม่

- ให้ทำการคำนวณหาจุดที่อยู่ระหว่างจุดศูนย์กลางและจุดที่ใหม่

$$x_c = \bar{x} + \gamma(x_r - \bar{x}) \tag{8}$$

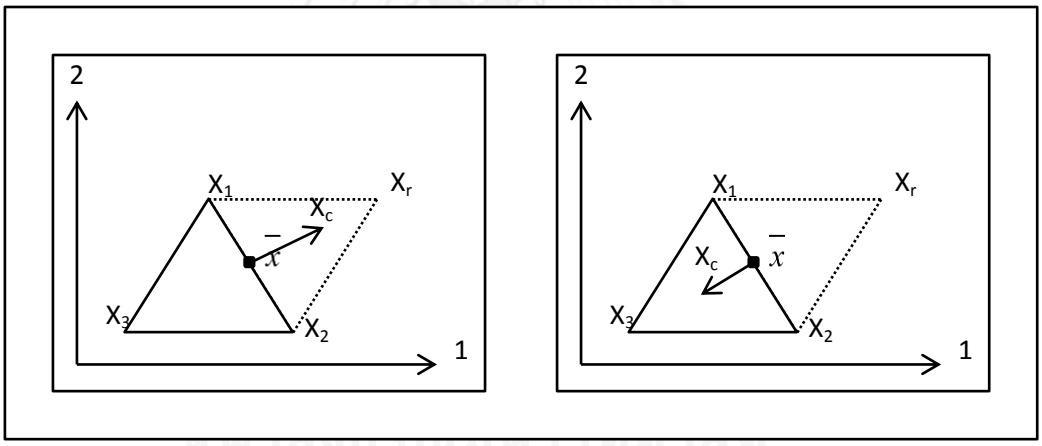
- หาก $f(x_c) < f(x_n)$ ให้ยอมรับ x_c เป็นจุดใหม่ หากไม่ ให้เข้าสู่ขั้นตอนที่ 5

4.1 ปรับโดยอิงจากจุดเดิม

- ให้ทำการคำนวณหาจุดที่อยู่ระหว่างจุดศูนย์กลางและจุดเดิม

$$x_c = \bar{x} - \gamma(\bar{x} - x_{n+1}) \tag{9}$$

- หาก $f(x_c) < f(x_n)$ ให้ยอมรับ x_c เป็นจุดใหม่ หากไม่ ให้เข้าสู่ขั้นตอนที่ 5



รูปที่ 17 แสดงการปรับโดยยึดจุดใหม่ และยึดจุดเดิมเป็นหลัก

5. Perform a shrink step

หากจุดที่ได้จากการคำนวณไม่สามารถทำให้ค่าของฟังก์ชันลดลงได้ จะการปรับทุกจุดใหม่ด้วยสมการที่

$$v_i = x_i + \sigma(x_1 - x_i) \tag{10}$$

2.1.3 การทดสอบสมมติฐาน (Tests of hypothesis)

การทดสอบสมมติฐานเป็นการทดสอบว่าสมมติฐานของลักษณะประชากรเป็นจริงหรือไม่โดยใช้การวิเคราะห์กลุ่มตัวอย่างของประชากรด้วยวิธีทางสถิติ[5]

2.1.3.1 สมมติฐานทางสถิติ

สมมติฐานทางสถิติเป็นสมมติฐานที่อยู่ในรูปแบบของคณิตศาสตร์ โดยสัญลักษณ์ที่ใช้เขียนในสมมติฐานจะเป็นพารามิเตอร์เสมอ เช่น μ (ตัวกลางเลขคณิต หรือ ค่าเฉลี่ยของประชากร) σ (ส่วนเบี่ยงเบนมาตรฐานของกลุ่มประชากร) โดยสมมติฐานทางสถิติมี 2 ชนิด คือ

2.1.3.1.1 สมมติฐานที่เป็นกลางหรือสมมติฐานไร้นัยสำคัญ (Null hypothesis) เขียนแทนด้วยสัญลักษณ์ H0 เป็นสมมติฐานที่แสดงให้เห็นว่าไม่มีความแตกต่างระหว่างตัวแปร เช่น

H0: $\mu1=\mu2$ (ค่าเฉลี่ยของประชากรกลุ่มที่ 1 และ 2 ไม่มีความแตกต่างกัน)

2.1.3.1.2 สมมติฐานอื่นๆ (Alternative hypothesis) เขียนแทนด้วยสัญลักษณ์ H1 เป็นสมมติฐานที่แสดงให้เห็นว่ามีความแตกต่างระหว่างตัวแปร เช่น

H0: $\mu1 \neq \mu2$ (ค่าเฉลี่ยของประชากรกลุ่มที่ 1 และ 2 มีความแตกต่างกัน)

2.1.3.2 ขั้นตอนการทดสอบสมมติฐาน

ขั้นที่ 1 ตั้งสมมติฐาน H0 และ H1

ขั้นที่ 2 กำหนดระดับนัยสำคัญทางสถิติ (α)

ขั้นที่ 3 คำนวณค่าสถิติ

ขั้นที่ 4 หาค่าวิกฤติ (Critical value) จากตารางสถิติ

ขั้นที่ 5 เปรียบเทียบค่าสถิติที่คำนวณได้กับค่าวิกฤติ

ขั้นที่ 6 สรุปผลโดย

ยอมรับ H0 เมื่อค่าที่คำนวณได้ตกอยู่ในเขตยอมรับ

ปฏิเสธ H0 ยอมรับ H1 หากค่าที่คำนวณได้ตกอยู่ในเขตวิกฤติ

CHULALONGKORN UNIVERSITY

2.2 งานวิจัยที่เกี่ยวข้อง

Aaron J. Owens [6] ได้ทำการศึกษาเกี่ยวกับวิธีการเรียนรู้ชุดข้อมูลขนาดใหญ่ด้วยนิวรอลเน็ตเวิร์ก โดยการทำการลดมิติของชุดข้อมูลขนาดใหญ่ลงด้วยการใช้ PLS (Partial Least Squares) ทำการฉายข้อมูลในมิติเดิมลงไปยังมิติใหม่ที่มีขนาดน้อยลง และใช้ BNN (Bottleneck Neural Network) เพื่อลดค่าน้ำหนักที่จะต้องสนใจลงด้วย และยังได้ทำการสุ่มตัวอย่างที่จะใช้ในการสอนนิวรอลเน็ตเวิร์กจากชุดตัวอย่างขนาดใหญ่เพื่อลดเวลาและทรัพยากรที่ใช้ลงด้วย

Rahul Kala, Anupam Shulla และ Ritu Tiwari [7] ได้ทำการศึกษาเกี่ยวกับวิธีการเรียนรู้ชุดข้อมูลขนาดใหญ่ด้วยนิวรอลเน็ตเวิร์กโดยการใช้การคัสเตอร์จัดกลุ่มของชุดข้อมูลขนาดใหญ่ออกเป็นกลายๆกลุ่มก่อนจากนั้นจึงนำชุดข้อมูลย่อยๆนั้นไปสอนนิวรอลเน็ตเวิร์กหลายๆตัว เมื่อมีชุดข้อมูลใหม่เข้ามา หากเราต้องการที่จะแบ่งกลุ่มของขุดข้อมูลที่เข้ามา เราจะต้องนำชุดข้อมูลนั้นไปทำการจัดกลุ่มกลุ่มด้วยการคลัสเตอร์ก่อนเพื่อให้ทราบว่าชุดข้อมูลนั้นอยู่กลุ่มใด ต้องใช้นิวรอลเน็ตเวิร์กตัวใดในการแบ่ง จากนั้นจึงนำชุดข้อมูลไปผ่านนิวรอลเน็ตเวิร์กตามกลุ่มที่ข้อมูลนั้นถูกจัดไว้

Osama Badawy และ Ahmad Almotwaly [8] ได้ทำการศึกษาเกี่ยวกับการรวมองค์ความรู้ที่อยู่ภายในนิวรอลเน็ตเวิร์กที่ใช้ในระบบหลายตัวแทน โดยทำการสร้างและส่งนิวรอลเน็ตเวิร์กต้นแบบไปสอนอย่างง่ายในโฮสต่างๆที่มีชุดข้อมูลสอนต่างๆกันตามลำดับจากโฮสแรกไปยังโฮสสุดท้าย จากนั้นจึงทำการคัดลอกนิวรอลเน็ตเวิร์กที่ผ่านการสอนอย่างง่ายนั้นเพื่อกระจายไปสอนยังโฮสต่างๆอีกครั้งหนึ่ง โดยครั้งนี้จะทำการสอนจนกว่านิวรอลเน็ตเวิร์กจะเรียนรู้ชุดข้อมูลทั้งหมด จากนั้นจึงทำการรวมนิวรอลเน็ตเวิร์กที่ผ่านการสอนทั้งหมดเข้าด้วยกันเพื่อจะนำไปใช้ต่อไป

Sherif Hashem และ Bruce Schmeiser [9] ได้ทำการศึกษาเกี่ยวกับการรวมผลลัพธ์ของนิวรอลเน็ตเวิร์กหลายโครงข่ายที่มีโครงสร้างหรือวิธีการสอนที่แตกต่างกันเข้าด้วยกัน โดยการสร้างฟังก์ชันระหว่างค่าถ่วงน้ำหนักของผลลัพธ์ของนิวรอลเน็ตเวิร์กกับค่าความผิดพลาดที่เกิดจากการรวมผลลัพธ์ด้วยค่าน้ำหนักนั้น โดยในงานวิจัยนี้นี้ได้นำเสนอวิธีการหาค่าน้ำหนักที่เหมาะสม 2 วิธี ได้แก่ 1 หาอณุพันธ์บนฟังก์ชันที่ได้สร้างขึ้นเพื่อหาชุดของค่าถ่วงน้ำหนักที่ทำให้ผลลัพธ์จากการรวมมีค่าความผิดพลาดน้อยที่สุด และ 2 ใช้การประมาณค่าถ่วงน้ำหนักโดยการคำนวณ

Nadeem Qazi และ Hoi Yeung [10] ได้ทำการศึกษาการเกี่ยวกับการรวมผลลัพธ์ของนิวรอลเน็ตเวิร์กหลายโครงข่าย โดยก่อนจะทำการสอนนิวรอลเน็ตเวิร์ก คุณลักษณะที่อยู่ภายในชุดข้อมูลจะถูกทำการเลือกเพื่อให้เหลือเพียงคุณลักษณะที่ดีโดยการใช้ MI (Mutual Information) จากนั้นชุดข้อมูลจะถูกทำการนอร์มอลไลซ์ จากนั้นจึงนำชุดข้อมูลไปสอนนิวรอลเน็ตเวิร์กหลายๆโครงข่าย ผลลัพธ์ของนิวรอลเน็ตเวิร์กหลายๆโครงข่ายจะถูกรวมเข้าด้วยกันเพื่อเพิ่มความถูกต้อง โดยในงานวิจัยนี้ได้ทำการศึกษาวิธีการรวม 3 รูปแบบ ได้แก่ 1 รวมโดยให้ค่าถ่วงน้ำหนักของนิวรอลเน็ตเวิร์กทุกๆโครงข่ายมีค่าเท่ากัน 2 คำนวณหาค่าถ่วงน้ำหนักโดยการใช้ Liner Regression เพื่อคำนวณหาค่าถ่วงน้ำหนักที่ทำให้ผลลัพธ์จากการรวมใกล้เคียงกับผลลัพธ์ที่ถูกต้องมากที่สุด 3 คำนวณหาค่าถ่วงน้ำหนักโดยการใช้ Principle Component Regression เพื่อคำนวณหาค่าถ่วงน้ำหนักที่ทำให้ผลลัพธ์จากการรวมใกล้เคียงกับผลลัพธ์ที่ถูกต้องมากที่สุด โดยในงานวิจัยนี้ได้สรุปว่า

การรวมโดยใช้ค่าถ่วงน้ำหนักที่ได้จากการคำนวณด้วย Principle Component Regression ให้ค่าประสิทธิภาพของความถูกต้องมากที่สุด

Mohammad Bahrami [11] ได้ทำการศึกษาเกี่ยวกับการรวมองค์ความรู้ที่อยู่ภายในนิวรอลเน็ตเวิร์กที่ผ่านการสอนแล้ว 2 ตัวเข้าด้วยกัน โดยในงานวิจัยนี้ได้เสนอวิธีการรวมต่างๆไว้ดังนี้ 1 สร้างนิวรอลเน็ตเวิร์กใหม่ที่มีจำนวนโหนดซ่อนเท่ากับผลรวมของจำนวนโหนดซ่อนของสองโครงข่ายรวมกัน ทำการสุ่มค่าน้ำหนักของชั้นโหนดซ่อน และทำการสอนเฉพาะชั้นแสดงผลใหม่ 2 ทำการสร้างนิวรอลเน็ตเวิร์กโดยนำค่าน้ำหนักของโหนดซ่อนจากโหนดซ่อนของนิวรอลเน็ตเวิร์กตั้งต้นทั้งสองโครงข่ายมาเป็นค่าน้ำหนักของชั้นซ่อน จากนั้นทำการสอนชั้นแสดงผลใหม่ 3 ทำการสร้างนิวรอลเน็ตเวิร์กโดยนำค่าน้ำหนักของโหนดซ่อนจากโหนดซ่อนของนิวรอลเน็ตเวิร์กตั้งต้นทั้งสองโครงข่ายมาเป็นค่าน้ำหนักของชั้นซ่อน และทำการเพิ่มโหนดพิเศษในชั้นซ่อนเพื่อใช้ผสานผลลัพธ์ จากนั้นทำการสอนโหนดพิเศษและชั้นแสดงผลใหม่ โดยผลการทดลองจากงานวิจัยนี้สรุปได้ว่า การนำค่าน้ำหนักจากนิวรอลเน็ตเวิร์กตั้งต้นมาสร้างเป็นโหนดซ่อน จากนั้นทำการสอนชั้นแสดงผลใหม่ใช้เวลาในการสอนน้อยที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

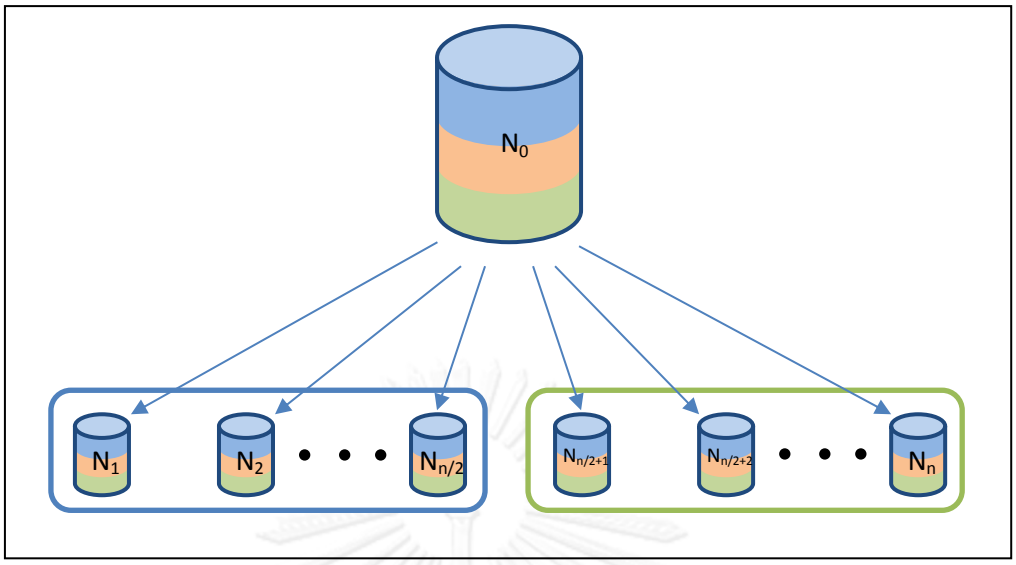
บทที่ 3
แนวคิดและวิธีดำเนินงานวิจัย

ในงานวิจัยนี้มีเป้าหมายเพื่อศึกษาขั้นตอนที่เหมาะสมในการวิเคราะห์โหนดในนิวรอลเน็ตเวิร์ก และวิธีการรวมนิวรอลเน็ตเวิร์กเข้าด้วยกัน โดยจะแบ่งขั้นตอนการทำงานออกเป็น 8 ขั้น ได้แก่ แบ่งชุดข้อมูลออกเป็นชุดย่อยๆ ทำการสอนนิวรอลเน็ตเวิร์ก วิเคราะห์หานิวรอลเน็ตเวิร์กโครงข่ายที่ดีที่สุด รวมกลุ่มโหนดซ่อน รวมโหนดซ่อน เพิ่มโหนดซ่อน วัดผลความถูกต้อง เปรียบเทียบความถูกต้อง โดยขั้นตอนการทำงานสามารถแสดงได้ดังรูปที่ 18

รูปที่ 18 แสดงขั้นตอนในงานวิจัย

3.1 แบ่งชุดข้อมูล

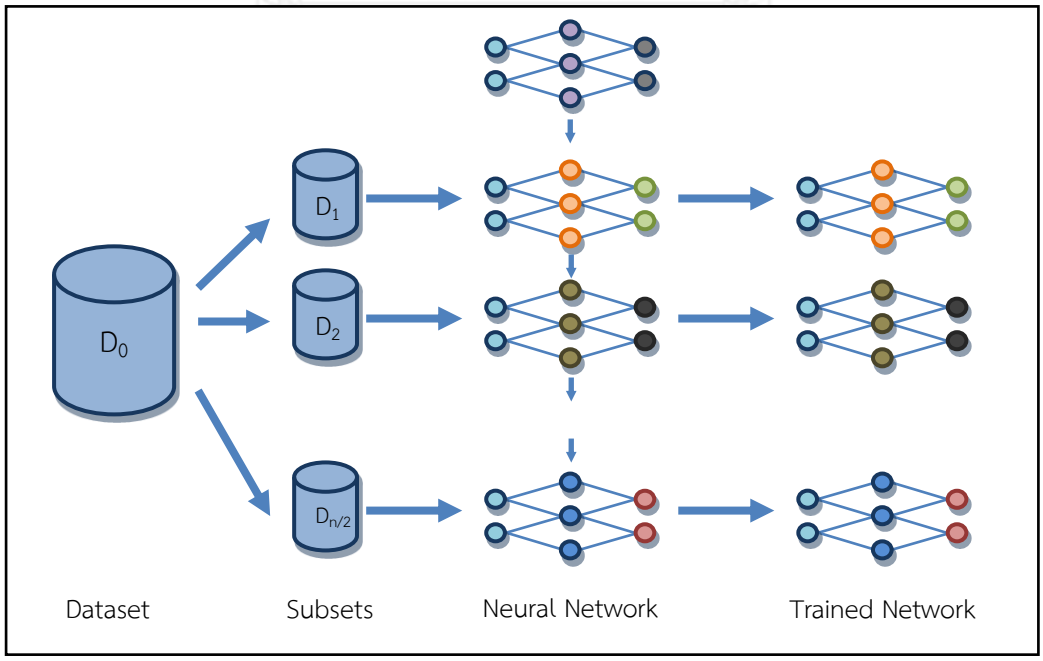
ก่อนจะทำการสอนด้วยเทคนิคการใช้นิวรอลเน็ตเวิร์กหลายโครงข่ายนั้น จะต้องเริ่มจากการทำการแบ่งชุดข้อมูลที่จะนำมาสอนนิวรอลเน็ตเวิร์กออกเป็นกลุ่มย่อยๆ n กลุ่ม และจัดกลุ่ม n กลุ่มนั้นเป็น 2 กลุ่มหลัก ดังที่แสดงในรูป 19 โดยที่ภายในชุดข้อมูลย่อยแต่ละชุดจะมีจำนวนตัวอย่างแต่ละประเภทเท่าๆกัน



รูปที่ 19 แสดงการแบ่งชุดข้อมูล

3.2 สอนนิวรอลเน็ตเวิร์ก

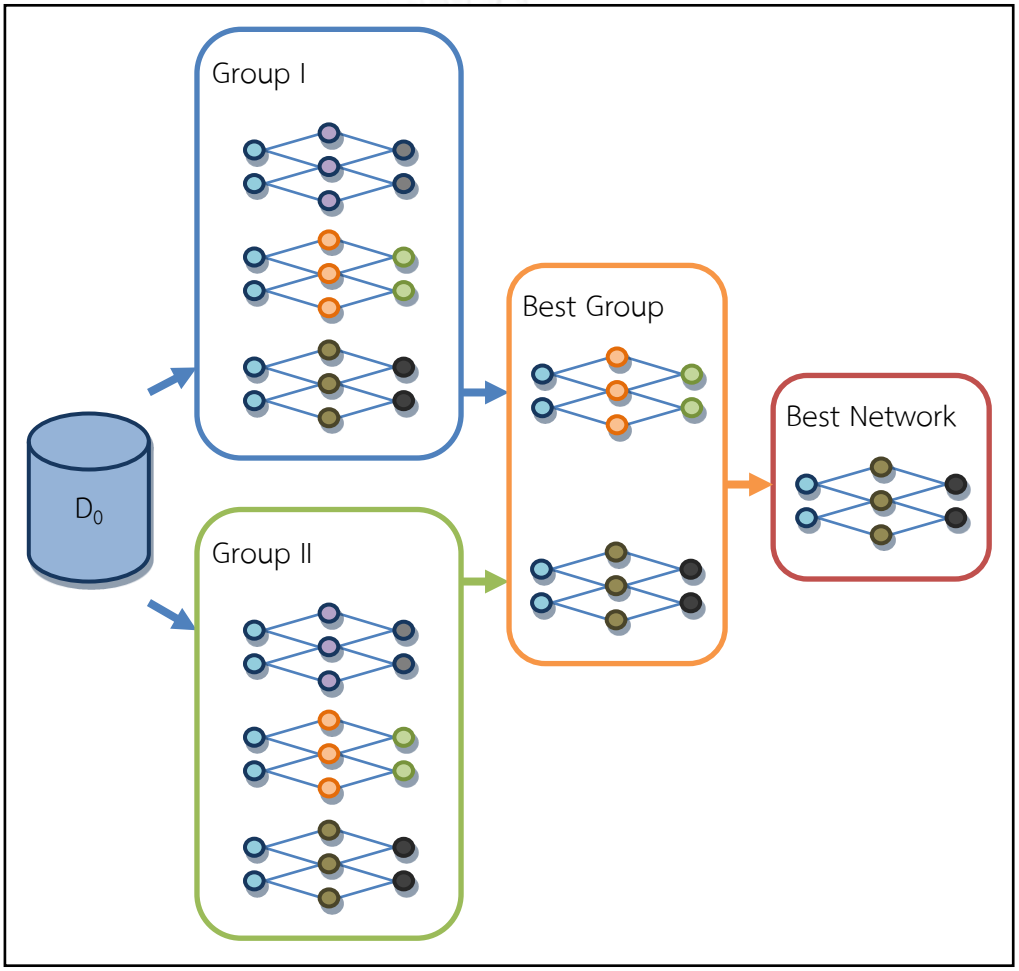
เมื่อได้ชุดข้อมูล 2 กลุ่มหลักแล้ว เราจะนำกลุ่มหลักแต่ละกลุ่มไปทำการสอนนิวรอลเน็ตเวิร์กแบบเป็นลำดับ โดยเริ่มจากนำนิวรอลเน็ตเวิร์กที่ได้สร้างมาเป็นนิวรอลเน็ตเวิร์กเริ่มต้น จากนั้นจะนำชุดตัวอย่างย่อยๆแต่ละชุดไปทำการสอนนิวรอลเน็ตเวิร์กตามลำดับ โดยเมื่อนิวรอลเน็ตเวิร์กเรียนรู้ชุดข้อมูลหนึ่งๆแล้ว จะทำการเก็บค่าน้ำหนักของนิวรอลเน็ตเวิร์กโครงข่ายนั้นไว้ จากนั้นจึงนำนิวรอลเน็ตเวิร์กนั้นไปเรียนรู้ชุดข้อมูลอื่นๆจนครบทุกชุดข้อมูลย่อยในกลุ่มหลักนั้นดังที่แสดงในรูปที่ 20



รูปที่ 20 แสดงขั้นตอนการสอนนิวรอลเน็ตเวิร์กของกลุ่มหลักที่ 1

3.3 หาโครงข่ายที่ดีที่สุด

หลังจากผ่านขั้นตอนการสอนแล้ว เราจะได้นิวรอลเน็ตเวิร์ก 2 กลุ่ม กลุ่มละ n/2 โครงข่าย ขั้นตอนนี้เป็นขั้นตอนการหานิวรอลเน็ตเวิร์กโครงข่ายที่มีความถูกต้องบนชุดสอนมากที่สุด เพื่อทีจะใช้เน็ตเวิร์กนี้เป็นฐานที่จะนำองค์ความรู้จากนิวรอลเน็ตเวิร์กตัวอื่นมารวม โดยเริ่มจากการหานิวรอลเน็ตเวิร์กที่ดีที่สุดในแต่ละกลุ่มก่อน จากนั้นหานิวรอลเน็ตเวิร์กที่ดีที่สุดจากตัวแทนของกลุ่มนั้นอีกครั้งหนึ่ง โดยการค้นหานิวรอลเน็ตเวิร์กโครงข่ายที่ดีที่สุดทำโดยการผ่านชุดข้อมูลสอนเข้าไปในทุกๆ โครงข่าย จากนั้นวัดความถูกต้องของการจำแนกตัวอย่างของแต่ละโครงข่าย ดังที่แสดงในรูปที่ 21



รูปที่ 21 แสดงการหาโครงข่ายที่ดีที่สุด

เมื่อได้นิวรอลเน็ตเวิร์กโครงข่ายหลักแล้ว เราจะใช้นิวรอลเน็ตเวิร์กนี้เป็นโครงข่ายหลัก และจะใช้นิวรอลเน็ตเวิร์กโครงข่ายที่ดีที่สุดจากอีกกลุ่มหนึ่งเป็นโครงข่ายรองและจะทำการรวมโครงข่ายรองเข้าสู่โครงข่ายหลักต่อไป

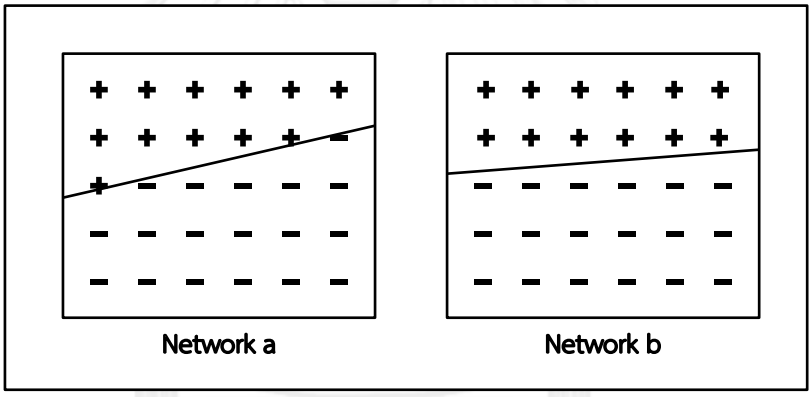
3.4 รวมกลุ่มโหนดซ่อน

เมื่อเราได้นิวรอลเน็ตเวิร์กโครงข่ายหลักแล้ว เราจะทำการรวมโหนดซ่อนที่วางตัวคล้ายกันจากนิวรอลเน็ตเวิร์กโครงข่ายรองเข้าสู่โหนดในโครงข่ายนี้ โดยจะเริ่มจากการวิเคราะห์ความคล้ายของโหนดซ่อน จากนั้นรวมหลุ่มโหนดที่คล้ายกันไว้ด้วยกันเพื่อจะทำการรวมต่อไป

3.4.1 การวิเคราะห์ความคล้าย

ภายในนิวรอลเน็ตเวิร์กแต่ละโครงข่ายนั้นจะประกอบด้วยชั้นอินพุต ชั้นซ่อน และชั้นเอาต์พุต และภายในชั้นซ่อนนั้นจะประกอบด้วยโหนดซ่อนจำนวนหลายโหนด ซึ่งโหนดซ่อนแต่ละโหนดนั้นคือเพอร์เซปตรอน 1 ตัว ลักษณะของเพอร์เซปตรอนนั้นจะเป็นสมการเชิงเส้นซึ่งสามารถมองเป็นระนาบตัดสินใจได้ โดยเราสามารถบอกตำแหน่งที่ระนาบวางตัวอยู่บนข้อมูลได้โดยดูจากเอาต์พุตของเพร์เซปตอนเมื่อผ่านข้อมูลเข้าไปว่าเป็น + หรือ –

ในงานวิจัยนี้จะใช้วิธีวัดความคล้ายด้วยการวัดความเหมือนของการแตกข้อมูลเป็น + หรือ – ของโหนดที่อยู่ในตำแหน่งเดียวกันของนิวรอลเน็ตเวิร์กโครงข่ายหลักและโครงข่ายอื่นๆ



รูปที่ 22 แสดงการแตกตัวอย่างของโหนดซ่อน

จากรูปที่ 22 เราสามารถคำนวณความคล้ายของโหนดซ่อนทั้ง 2 ตัวได้โดยสมการ 11

$$similarity = 1 - \frac{1}{2n}\sum_{i=1}^{n}|a_i - b_i| \qquad (11)$$

เมื่อ similarity คือความคล้ายคลึงในการแตกข้อมูลของโหนดซ่อน a_i และ b_i คือเอาต์พุตข้อมูลตัวที่ i ของโหนดซ่อน a และ b ซึ่งมีค่าเป็น -1 หรือ 1 และ n คือจำนวนของข้อมูลทั้งหมด

3.4.2 การรวมกลุ่ม

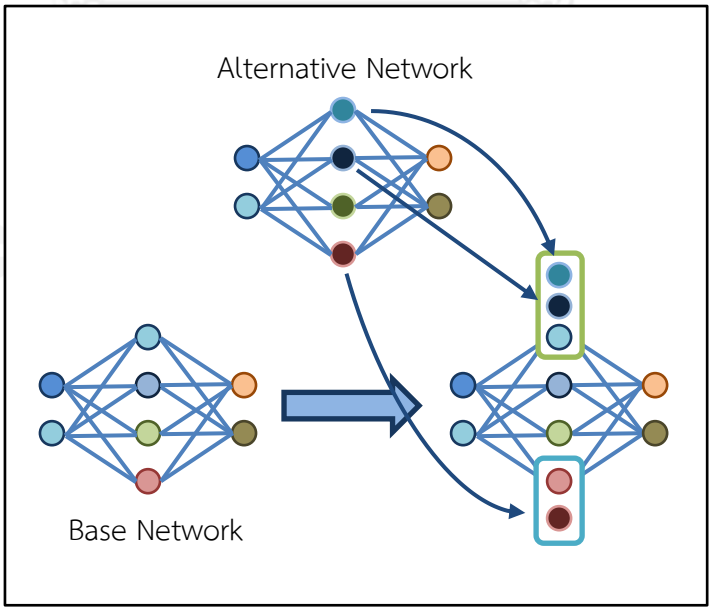
ในขั้นตอนนี้จะเป็นการรวมกลุ่มโหนดซ่อนที่วางตัวคล้ายกันไว้ด้วยกัน โดยจะทำการวัดความคล้ายของโหนดซ่อนระหว่างนิวรอลเน็ตเวิร์กหลักและโครงข่ายรอง ดังที่แสดงในตารางที่ 1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 1 ตัวอย่างความคล้ายของการวางตัวของโหนดซ่อนระหว่างโครงข่ายหลักและรอง

โครงข่ายหลัก	โครงข่ายรอง									
	1	2	3	4	5	6	7	8	9	10
1	0.34	0.76	0.77	0.38	0.42	0.06	0.75	0.70	0.54	0.48
2	0.62	0.41	0.35	0.88	0.56	0.58	0.49	0.44	0.66	0.45
3	0.35	0.82	0.98	0.23	0.48	0.21	0.52	0.86	0.50	0.42
4	0.37	0.79	0.95	0.27	0.51	0.23	0.52	0.88	0.53	0.38
5	0.48	0.30	0.43	0.49	0.75	0.59	0.54	0.52	0.62	0.37
6	0.42	0.56	0.55	0.52	0.47	0.46	0.49	0.58	0.53	0.52
7	0.61	0.38	0.33	0.61	0.44	0.81	0.17	0.33	0.39	0.63
8	0.42	0.62	0.53	0.48	0.26	0.30	0.70	0.47	0.44	0.65
9	0.76	0.57	0.43	0.50	0.35	0.55	0.48	0.30	0.23	0.77
10	0.42	0.70	0.70	0.32	0.51	0.19	0.80	0.63	0.51	0.49

โดยในงานวิจัยนี้จะทำการรวมโหนดซ่อนที่แตกกลุ่มตัวอย่างคล้ายกัน 80 เปอร์เซ็นต์ขึ้นไปไว้ในกลุ่มเดียวกัน เช่น จัดกลุ่มโหนดซ่อนที่ 2 3 และ 8 ของนิวรอลเน็ตเวิร์กรองไว้ในกลุ่มเดียวกับโหนดซ่อนที่ 3 ของนิวรอลเน็ตเวิร์กหลัก ดังตัวอย่างที่แสดงในรูปที่ 23



รูปที่ 23 แสดงการรวมกลุ่มโหนดซ่อน

3.5 รวมโหนดซ่อน

ในขั้นตอนนี้เราจะทำการรวมโหนดซ่อนที่ถูกจัดให้อยู่ในกลุ่มเดียวกันเข้าด้วยกัน โดยเราจะใช้วิธีการหาค่าถ่วงน้ำหนักที่เหมาะสมที่จะใช้ในการรวมด้วยวิธีการ optimization โดยเริ่มจากการสร้างฟังก์ชันระหว่างค่าถ่วงน้ำหนักที่จะใช้ในการรวมกับค่าความผิดพลาดในการจำแนกข้อมูลของนิวรอลเน็ตเวิร์กที่ผ่านการรวมโหนดซ่อนด้วยค่าถ่วงน้ำหนักชุดนี้ดังสมการที่ 12

$$error = f(w_1, w_2, ..., w_h) \tag{12}$$

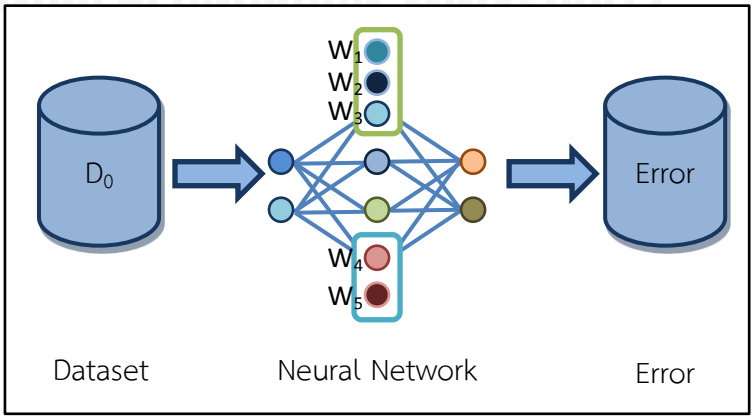
โดยฟังก์ชัน f ในสมการที่ 13 นั้นมีการทำงานดังนี้

- นำค่าน้ำหนักที่ได้รับมาไปทำการรวมค่าน้ำหนักของโหนดซ่อน

$$\begin{bmatrix} b & w_1 & w_2 & \dots & w_n \end{bmatrix}_i = [weight] \begin{bmatrix} \begin{bmatrix} b & w_1 & w_2 & \dots & w_n \end{bmatrix}_A \\ \begin{bmatrix} b & w_1 & w_2 & \dots & w_n \end{bmatrix}_B \\ \dots \\ \begin{bmatrix} b & w_1 & w_2 & \dots & w_n \end{bmatrix}_N \end{bmatrix} \tag{13}$$

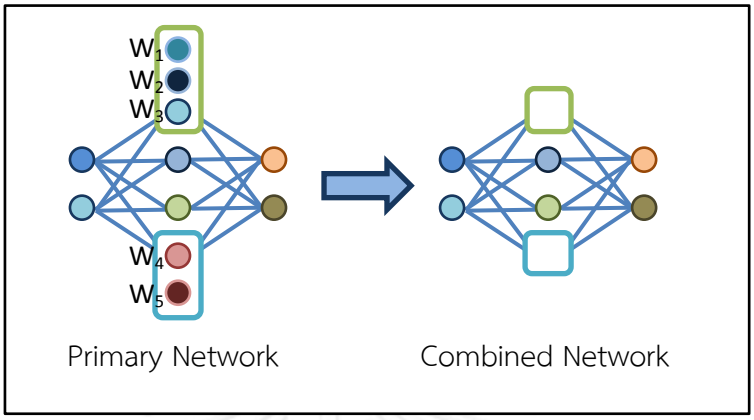
- นำค่าน้ำหนักที่ได้ไปสร้างโหนดซ่อนและนำไปรวมกับชั้นแสดงผลของนิวรอลเน็ตเวิร์กหลักเพื่อสร้างนิวรอลเน็ตเวิร์กตัวใหม่
- ผ่านชุดข้อมูลเข้าไปในนิวรอลเน็ตเวิร์กที่ได้และคำนวณค่าความผิดพลาดในการจำแนกตัวอย่าง
- ส่งคืนค่าความผิดพลาดที่คำนวณได้

โดยการทำงานของฟังก์ชันในสมการที่ 12 สามารถแสดงได้ดังรูปที่ 24



รูปที่ 24 แสดงการทำงานของฟังก์ชันระหว่างค่าถ่วงน้ำหนักและค่าความผิดพลาด

เมื่อคำนวณค่าถ่วงน้ำหนักที่เหมาะสมที่สุดได้แล้ว จะนำค่าถ่วงน้ำหนักนี้ไปทำการรวมโหนดซ่อนที่ถูกจัดอยู่ในกลุ่มเดียวกันเข้าด้วยกัน เพื่อสร้างเป็นนิวรอลเน็ตเวิร์กโครงข่ายใหม่ดังรูปที่ 25



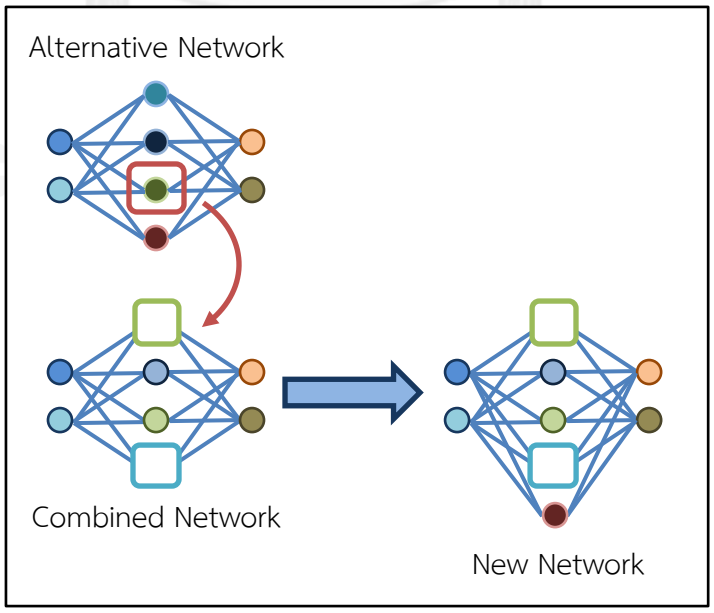
รูปที่ 25 แสดงการรวมโหนดซ่อนและสร้างนิวรอลเน็ตเวิร์กโครงข่ายใหม่

3.6 เพิ่มโหนดซ่อน

ในขั้นตอนนี้จะทำการเพิ่มโหนดซ่อนของนิวรอลเน็ตเวิร์กโครงข่ายรองที่วางตัวแตกต่างจากโหนดซ่อนของโครงข่ายหลักเราสู่โครงข่ายหลัก โดยในขั้นตอนนี้จะประกอบด้วยขั้นตอนย่อย 3 ขั้นตอน ได้แก่ เพิ่มโหนดซ่อน ปรับค่าน้ำหนักที่ต่อกับชั้นแสดงผลของโหนดที่จะทำการเพิ่ม และสอนชั้นแสดงผลใหม่

3.4.1 เพิ่มโหนดซ่อน

ในขั้นตอนนี้จะทำการเพิ่มโหนดซ่อนของนิวรอลเน็ตเวิร์กโครงข่ายรองที่วางตัวแตกต่างจากโครงข่ายหลักเข้าสู่โครงข่ายหลักดังรูปที่ 26



รูปที่ 26 แสดงการเพิ่มโหนดซ่อน

3.6.2 ปรับค่าถ่วงน้ำหนัก

ค่าน้ำหนักของโหนดแสดงผลของนิวรอลเน็ตเวิร์กโครงข่ายหลักและโครงข่ายรองนั้นมีขนาดที่แตกต่างกัน ในขั้นตอนนี้จะทำการปรับค่าน้ำหนักของโหนดแสดงผลให้เท่าเทียมกัน โดยเริ่มจากการหาขนาดของเวกเตอร์ของโหนดซ่อนโดยใช้สมการที่ 14

$$\left|\overrightarrow{V_i}\right| = \sqrt{\sum_{j=1}^{n} w_{ij}^{\ 2}} \tag{14}$$

เมื่อ V_i คือขนาดของเวกเตอร์ของค่าน้ำหนักของโหนดแสดงผลที่ i w_{ij} คือค่าน้ำหนักของโหนดแสดงผลตัวที่ i ซึ่งเชื่อมต่ออยู่กับโหนดซ่อนที่ j และ n คือจำนวนของโหนดซ่อน

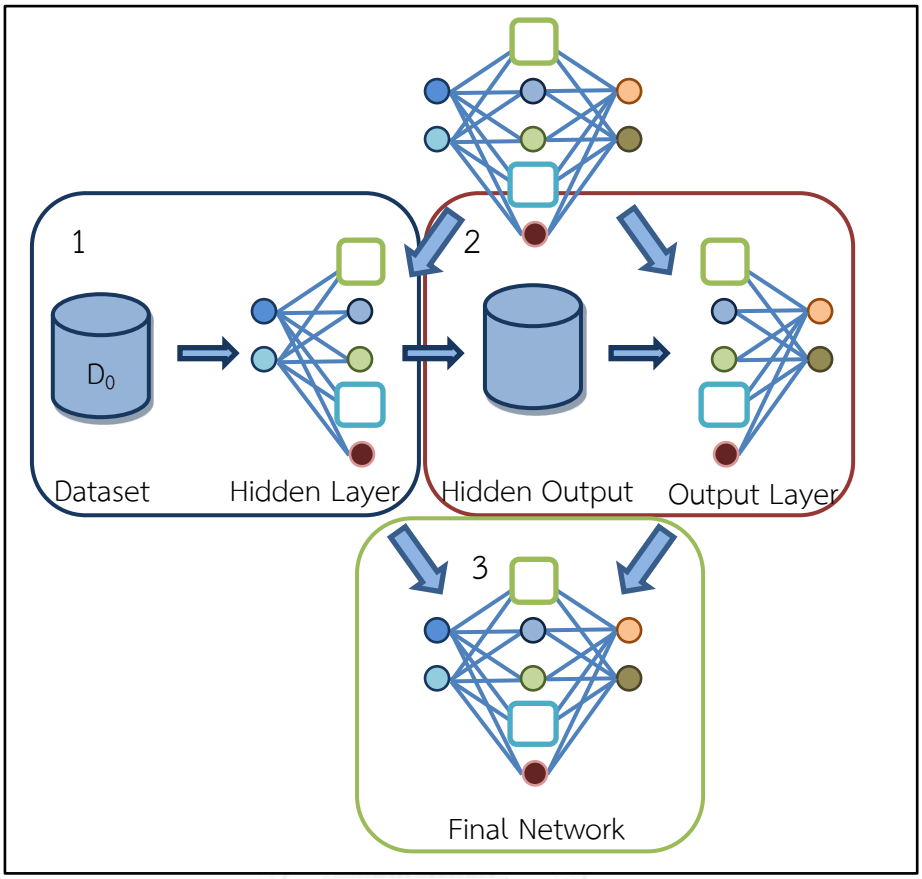
เมื่อเราคำนวณขนาดของเวกเตอร์ของค่าน้ำหนักของโหนดซ่อนของนิวรอลเน็ตเวิร์กโครงข่ายและโครงข่ายรองทุกโหนดแล้ว เราจะนำค่าที่คำนวนได้ไปทำการปรับค่าน้ำหนักของโหนดแสดงผลที่เชื่อมต่อกับโหนดซ่อนที่ถูกเพิ่มเข้ามาในขั้ตอนก่อนหน้าด้วยสมการที่ 15

$$W_{ij} = \frac{V_{1i}}{V_{2i}} \times w_{ij} \tag{15}$$

เมื่อ W_{ij} คือค่าน้ำหนักใหม่ของโหนดแสดงผลที่ i ที่เชื่อมต่ออยู่กับโหนดที่เพิ่มเข้าไปที่ j V_{1i} คือขนาดของเวกเตอร์ของโหนดแสดงผลที่ i ของนิวรอลเน็ตเวิร์กหลัก V_{2i} คือขนาดของเวกเตอร์ของโหนดแสดงผลที่ i ของนิวรอลเน็ตเวิร์กรอง w_{ij} คือค่าน้ำหนักของโหนดแสดงผลที่ i ที่เชื่อมต่ออยู่กับโหนดที่เพิ่มเข้าไปที่ j

3.6.3 สอนชั้นแสดงผลใหม่

ในขั้นตอนนี้จะเป็นการสอนโหนดแสดงผลใหม่เพื่อปรับค่าน้ำหนักให้เข้ากับโหนดซื่อที่ได้ทำการเพิ่มเข้าไปจากขั้นตอนก่อนหน้านี้ โดยเริ่มจากการผ่านชุดข้อมูลเข้าสู่ชั้นโหนดซ่อนจากกนั้นเก็บผลลัพธ์จากชั้นโหนดซ่อนไว้ และใช้ข้อมูลที่เก็บไว้นี้ทำการสอนชั้นแสดงผล และเมื่อสอนชั้นแสดงผลเสร็จแล้วจะนำชั้นแสดงผลนั้นไปรวมกับชั้นซ่อนเพื่อสร้างเป็นนิวรอลเน็ตเวิร์กใหม่ดังที่แสดงในรูปที่ 27



รูปที่ 27 แสดงขั้นตอนการชั้นแสดงผล

3.7 วัดผล

ในงานวิจัยนี้การวัดผลจะประกอบด้วย 2 ส่วนได้แก่

3.7.1 วัดความแม่นยำ

การวัดความแม่นยำจะใช้การทดสอบแบบ k-fold Cross-Validation โดยจะนำชุดข้อมูลขนาดใหญ่ผ่านขั้นตอน k-fold Cross-Validation โดยที่ k เท่ากับ 10 เพื่อวนไขว้แบ่งข้อมูลออกเป็นชุดสอนและชุดทดสอบ โดยนำชุดสอนไปผ่านขั้นตอนที่ 3.1 ถึง 3.6 เพื่อสร้างนิวรอลเน็ตเวิร์ก จากนั้นนำชุดทดสอบไปทดสอบนิวรอลเน็ตเวิร์กที่ได้จากการสร้างโดยชุดสอนและทำการบันทึกเก็บค่าความแม่นยำต่อไป

3.7.2 วัดผลกระทบของการแบ่งกลุ่ม

การวัดผลกระทบของการแบ่งกลุ่มจะทำโดยการเก็บข้อมูลของความถูกต้องและการคำนวณที่ใช้ เมื่อจำนวนโครงข่ายของนิวรอลเน็ตเวิร์กที่ใช้ในการสอนมีจำนวนต่างกัน

3.8 เปรียบเทียบผล

การเปรียบเทียบผลจะทำการเปรียบเทียบค่าความแม่นยำและเวลาที่ใช้ในการสอนระหว่างนิวรอลเน็ตเวิร์กที่ได้จากการรวมกับนิวรอลเน็ตเวิร์กที่ได้จากการสอนโดยชุดข้อมูลขนากใหญ่โดยตรง โดยจะทำการสรุปผลด้วยวิธีการทดสอบสมมติฐานทางสถิติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะเป็นการอธิบายถึงขั้นตอนการทดลอง ชุดข้อมูลที่ใช้ และเครื่องมือที่ใช้ และผลการทดลอง

4.1 ชุดข้อมูลที่ใช้

ข้อมูลที่ใช้ในงานวิจัยนี้ประกอบด้วยชุดข้อมูล 5 ชุดดังในตารางที่ 2 โดยชุดข้อมูลที่ 1 3 4 และ 5 เป็นชุดข้อมูลจาก UCI Machine Learning Repository[12] ส่วนชุดข้อมูลที่ 2 เป็นชุดข้อมูลที่ได้จากการสร้างขึ้น

ตารางที่ 2 แสดงชุดข้อมูล

Name	Sample	Feature	Class
Covertype	581012	53	7
Synthesized	200000	2	2
MiniBooNE	130065	50	2
Statlog	58000	8	7
King-Rook vs. King	28056	6	18

4.2 เครื่องมือที่ใช้

ฮาร์ดแวร์

- หน่วยประมวลผล intel(R) core(TM) i5 M460 ความเร็ว 2.53 กิกะเฮิร์ต
- หน่วยความจำ 8 กิกะไบต์

ซอฟต์แวร์

- วินโดวส์ 7 64 บิต
- โปรแกรม matlab เวอร์ชัน R2012a(7.14.0.739) 64 บิต

4.3 ขั้นตอนการทดลอง

ในงานวิจัยนี้จะเขียนโปรแกรมและทดลองบนโปรแกรม matlab และใช้นิวรอลเน็ตเวิร์กที่มีชั้นซ่อนเพียง 1 ชั้นซึ่งประกอบด้วยโหนดซ่อน 10 โหนด จะทำการสอนนิวรอลเน็ตเวิร์กด้วยวิธีการแบบ Gradient descent ที่มีการปรับอัตราการเรียนรู้อัตโนมัติและมีการใช้โมเมนตัมเข้าร่วมด้วย โดยการสอนนั้นจะใช้ validation set เข้าร่วมด้วย และกำหนดเงื่อนไขในการหยุดสอนเมื่อค่าความผิดพลาดบน validation set ไม่ลดลงเป็นเวลา 20 รอบติดต่อกัน

จำนวนกลุ่มย่อยที่จะทำการแบ่งได้แก่ 5 10 20 40 และ 80 กลุ่มย่อย การวัดผลจะใช้ค่าเฉลี่ยจากการทำ 10-flow cross validate

4.4 ผลการทดลอง

ผลการทดลองในงานวิจัยนี้จะแบ่งออกเป็น 2 ส่วนได้แก่ ผลการทดลองด้านความถูกต้องและผลการทดลองด้านเวลา

4.4.1 ความถูกต้อง

ในงานวิจัยนี้จะนำชุดข้องมูลแต่ละชุดไปทำการสอนและทดสอบด้วยวิธีการแบบ 10-flow cross validate โดยค่าความถูกต้องของชุดข้อมูลต่างๆสามารถแสดงได้ดังนี้

4.4.1.1 Covertype

ค่าความผิดพลาดในการทดลองของชุดข้อมูล Covertype สามารถแสดงได้ในตารางที่ 3-5

ตารางที่ 3 แสดงค่าความผิดพลาดบนชุดข้อมูล Covertype ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	0.2321	0.2384	0.2330	0.2365	0.2357	0.2574
2	0.2736	0.2354	0.2424	0.2347	0.2364	0.2378
3	0.2368	0.2302	0.2304	0.2392	0.2343	0.2437
4	0.2415	0.2398	0.2448	0.2305	0.2323	0.2388
5	0.2323	0.2307	0.2321	0.2309	0.2284	0.2422
6	0.2286	0.2253	0.2272	0.2370	0.2414	0.2422
7	0.2382	0.2365	0.2403	0.2336	0.2388	0.2435
8	0.2444	0.2315	0.2312	0.2380	0.2307	0.2416
9	0.2305	0.2271	0.2328	0.2332	0.2318	0.2377
10	0.2331	0.2313	0.2339	0.2327	0.2427	0.2465
เฉลี่ย	0.2391	0.2326	0.2348	0.2346	0.2352	0.2432

ตารางที่ 4 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Covertype

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.2380	0.2351	0.2375	0.2375	0.2562
2	0.2362	0.2462	0.2380	0.2379	0.2394
3	0.2298	0.2327	0.2409	0.2385	0.2443
4	0.2412	0.2473	0.2364	0.2383	0.2453
5	0.2323	0.2335	0.2373	0.2301	0.2436
6	0.2301	0.2300	0.2369	0.2448	0.2481
7	0.2365	0.2417	0.2363	0.2401	0.2460
8	0.2325	0.2374	0.2434	0.2337	0.2456
9	0.2282	0.2333	0.2367	0.2342	0.2433
10	0.2331	0.2361	0.2368	0.2453	0.2536
เฉลี่ย	0.2338	0.2373	0.2380	0.2381	0.2466

จากนั้นจะทำการคำนวณหาค่าความเชื่อมั่นด้วยการทดสอบค่าที (t-test) โดยสมการที่ 16

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D^2)}{n-1}}} \tag{16}$$

เมื่อ t คือค่าทีที่คำนวณได้ D คือความต่างระหว่างค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุด และ n คือจำนวนของ flow

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

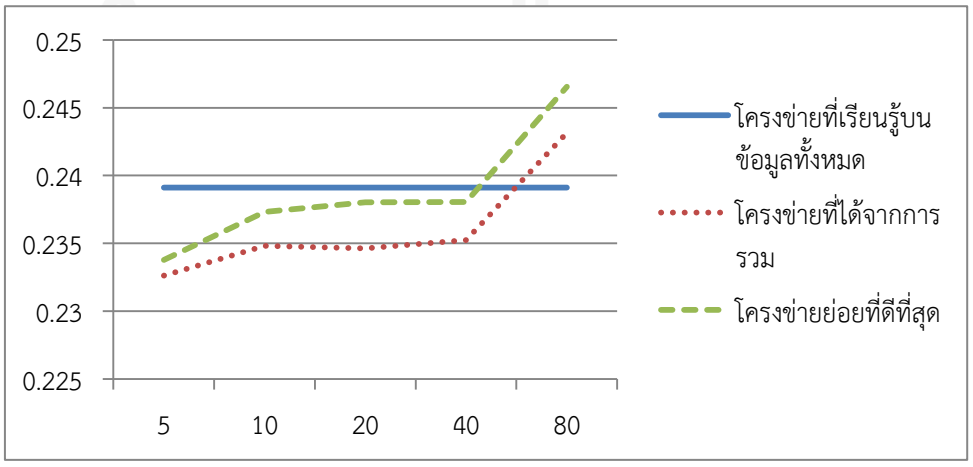
ตารางที่ 5 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Covertype

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.00034	-0.00210	-0.00107	-0.00182	0.00120
2	-0.00074	-0.00382	-0.00337	-0.00157	-0.00167
3	0.00046	-0.00227	-0.00167	-0.00422	-0.00057
4	-0.00145	-0.00250	-0.00589	-0.00601	-0.00656
5	-0.00160	-0.00143	-0.00640	-0.00170	-0.00141
6	-0.00487	-0.00279	0.00010	-0.00343	-0.00589
7	0.00002	-0.00134	-0.00270	-0.00131	-0.00246
8	-0.00093	-0.00614	-0.00540	-0.00298	-0.00401
9	-0.00115	-0.00050	-0.00344	-0.00246	-0.00558
10	-0.00172	-0.00224	-0.00413	-0.00255	-0.00707

ตารางที่ 6 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล Covertype

	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
$\sum D$	-0.01163	-0.02513	-0.03398	-0.02804	-0.03401
$\sum D^2$	0.00003	0.00009	0.00016	0.00010	0.00019
t	-2.41565	-5.09639	-5.03694	-6.13571	-3.80732
ระดับนัยสำคัญ	98.05564	99.96758	99.96486	99.99142	99.79149

โดยค่าความผิดพลาดสามารถแสดงได้ดังรูปที่ 28



รูปที่ 28 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล Covertype

4.4.1.2 Synthesized

ค่าความผิดพลาดในการทดลองของชุดข้อมูล Synthesized สามารถแสดงได้ในตารางที่ 7-19

ตารางที่ 7 แสดงค่าความผิดพลาดบนชุดข้อมูล Synthesized ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	0.0565	0.0506	0.0497	0.0576	0.0490	0.0548
2	0.0845	0.0518	0.0697	0.0564	0.0908	0.0514
3	0.0652	0.0933	0.1030	0.0617	0.0960	0.0652
4	0.0447	0.0499	0.0598	0.0484	0.0467	0.0506
5	0.0664	0.0769	0.0604	0.0620	0.0612	0.0646
6	0.0657	0.0466	0.0609	0.0710	0.0614	0.0651
7	0.0537	0.0538	0.0658	0.0759	0.0901	0.0567
8	0.0514	0.0524	0.0526	0.0523	0.0578	0.0556
9	0.0517	0.0498	0.0511	0.0507	0.0531	0.0500
10	0.0924	0.0539	0.0585	0.0597	0.0993	0.0579
เฉลี่ย	0.0632	0.0579	0.0631	0.0595	0.0705	0.0572

ตารางที่ 8 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Synthesized

Flow	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.0508	0.0518	0.0593	0.0524	0.0561
2	0.0533	0.0671	0.0581	0.0922	0.0541
3	0.0975	0.1060	0.0688	0.0983	0.0678
4	0.0512	0.0635	0.0484	0.0479	0.0521
5	0.0781	0.0620	0.0645	0.0630	0.0618
6	0.0467	0.0634	0.0797	0.0641	0.0673
7	0.0559	0.0705	0.0742	0.0964	0.0599
8	0.0525	0.0529	0.0540	0.0585	0.0576
9	0.0503	0.0518	0.0510	0.0564	0.0513
10	0.0544	0.0596	0.0622	0.0938	0.0599
เฉลี่ย	0.0591	0.0648	0.0620	0.0723	0.0588

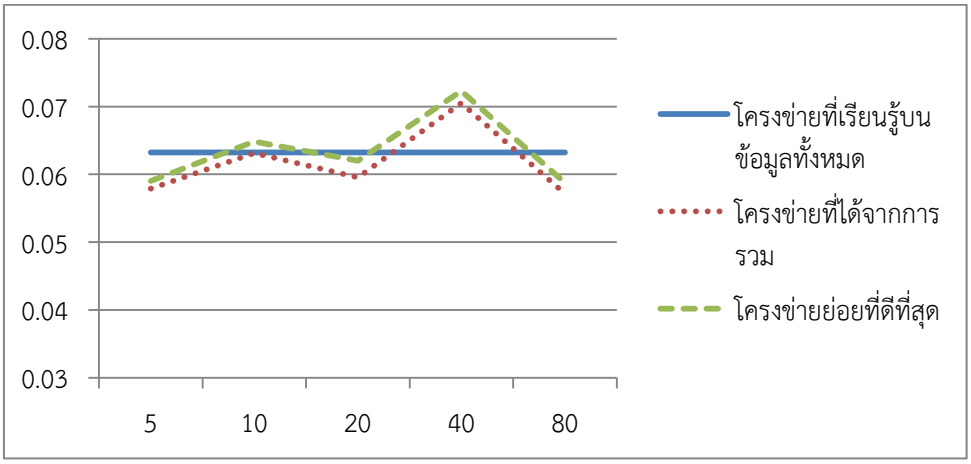
ตารางที่ 9 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Synthesized

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	-0.0002	-0.0022	-0.0018	-0.0034	-0.0014
2	-0.0015	0.0027	-0.0017	-0.0014	-0.0028
3	-0.0042	-0.0030	-0.0071	-0.0023	-0.0027
4	-0.0013	-0.0037	-0.0001	-0.0012	-0.0015
5	-0.0012	-0.0017	-0.0025	-0.0018	0.0029
6	-0.0001	-0.0025	-0.0087	-0.0027	-0.0022
7	-0.0021	-0.0047	0.0017	-0.0063	-0.0032
8	0.0000	-0.0004	-0.0017	-0.0007	-0.0020
9	-0.0005	-0.0006	-0.0003	-0.0033	-0.0012
10	-0.0004	-0.0011	-0.0025	0.0055	-0.0020

ตารางที่ 10 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล Synthesized

	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
$\sum D$	-0.01165	-0.01715	-0.02460	-0.01760	-0.01605
$\sum D^2$	0.00003	0.00007	0.00015	0.00011	0.00005
t	-2.91096	-2.64840	-2.45165	-1.85562	-3.00572
ระดับนัยสำคัญ	99.13579	98.67271	98.16714	95.17576	99.25909

โดยค่าความผิดพลาดสามารถแสดงได้ดังรูปที่ 29



รูปที่ 29 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล Synthesized

4.4.1.3 MiniBooNE

ค่าความผิดพลาดในการทดลองของชุดข้อมูล MiniBooNE สามารถแสดงได้ในตารางที่ 11-13

ตารางที่ 11 แสดงค่าความผิดพลาดบนชุดข้อมูล MiniBooNE ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	0.0854	0.0827	0.0804	0.0970	0.0887	0.0961
2	0.0948	0.0939	0.0971	0.0910	0.0965	0.1032
3	0.0961	0.0923	0.0946	0.0936	0.1031	0.1056
4	0.0867	0.0865	0.0879	0.0925	0.1004	0.1005
5	0.0948	0.0957	0.0889	0.0984	0.1003	0.1007
6	0.0965	0.0910	0.0996	0.0914	0.1010	0.1005
7	0.0937	0.0876	0.0926	0.0974	0.0993	0.1016
8	0.0992	0.0843	0.0954	0.0907	0.0985	0.1080
9	0.0806	0.0876	0.0881	0.0911	0.0999	0.1016
10	0.0823	0.0889	0.0861	0.0900	0.0996	0.1000
เฉลี่ย	0.0910	0.0891	0.0911	0.0933	0.0987	0.1018

ตารางที่ 12 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล MiniBooNE

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.0841	0.0819	0.0975	0.0923	0.1011
2	0.0933	0.0978	0.0898	0.0985	0.1045
3	0.0942	0.0954	0.0942	0.1043	0.1086
4	0.0886	0.0919	0.0936	0.1021	0.1035
5	0.0957	0.0895	0.0978	0.1023	0.1030
6	0.0904	0.1013	0.0925	0.1026	0.1018
7	0.0877	0.0920	0.0979	0.1011	0.1023
8	0.0853	0.0982	0.0915	0.0966	0.1086
9	0.0870	0.0879	0.0926	0.1024	0.1056
10	0.0907	0.0866	0.0928	0.0990	0.1048
เฉลี่ย	0.0897	0.0922	0.0940	0.1001	0.1044

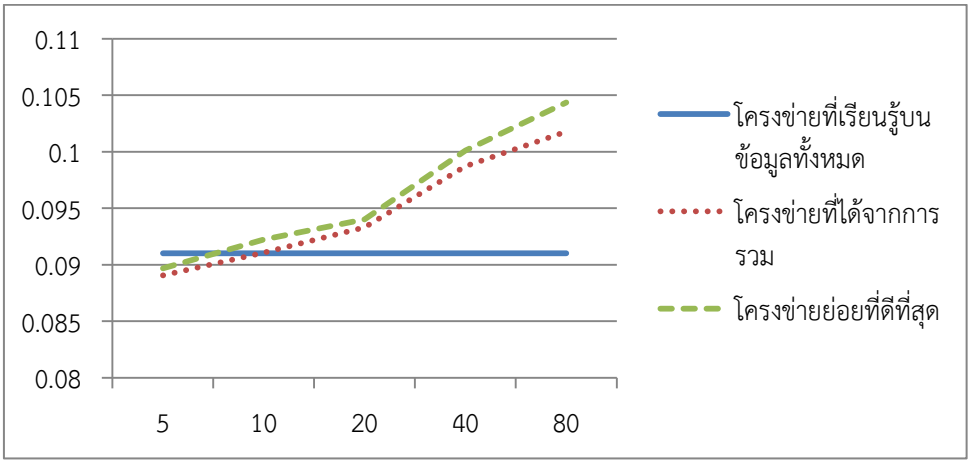
ตารางที่ 13 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล MiniBooNE

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	-0.0014	-0.0015	-0.0005	-0.0036	-0.0050
2	0.0007	-0.0007	0.0012	-0.0020	-0.0013
3	-0.0019	-0.0008	-0.0005	-0.0012	-0.0031
4	-0.0022	-0.0040	-0.0012	-0.0017	-0.0030
5	0.0000	-0.0006	0.0006	-0.0020	-0.0022
6	0.0005	-0.0018	-0.0011	-0.0016	-0.0013
7	-0.0001	0.0006	-0.0005	-0.0018	-0.0007
8	-0.0009	-0.0028	-0.0008	0.0018	-0.0005
9	0.0007	0.0002	-0.0015	-0.0025	-0.0041
10	-0.0018	-0.0005	-0.0028	0.0007	-0.0048

ตารางที่ 14 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล MiniBooNE

	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
$\sum D$	-0.00630	-0.01169	-0.00707	-0.01392	-0.02599
$\sum D^2$	0.00002	0.00003	0.00002	0.00004	0.00009
t	-1.76367	-2.66242	-2.01316	-2.81415	-4.99337
ระดับนัยสำคัญ	94.41930	98.70287	96.25278	98.98804	99.96272

โดยค่าความผิดพลาดสามารถแสดงได้ดังรูปที่ 30



รูปที่ 30 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล MiniBooNE

4.4.1.4 Statlog

ค่าความผิดพลาดในการทดลองของชุดข้อมูล Statlog สามารถแสดงได้ในตารางที่ 15-17

ตารางที่ 15 แสดงค่าความผิดพลาดบนชุดข้อมูล Statlog ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	0.0046	0.0046	0.0018	0.0046	0.0044	0.0046
2	0.0037	0.0034	0.0018	0.0039	0.0034	0.0037
3	0.0051	0.0051	0.0053	0.0053	0.0034	0.0051
4	0.0041	0.0039	0.0037	0.0044	0.0037	0.0048
5	0.0039	0.0032	0.0039	0.0034	0.0034	0.0025
6	0.0041	0.0041	0.0037	0.0044	0.0046	0.0046
7	0.0060	0.0044	0.0051	0.0053	0.0055	0.0041
8	0.0025	0.0023	0.0018	0.0044	0.0039	0.0046
9	0.0046	0.0030	0.0032	0.0053	0.0051	0.0055
10	0.0034	0.0028	0.0034	0.0030	0.0034	0.0030
เฉลี่ย	0.0042	0.0037	0.0034	0.0044	0.0041	0.0043

ตารางที่ 16 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล Statlog

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.0046	0.0039	0.0046	0.0044	0.0046
2	0.0034	0.0032	0.0039	0.0039	0.0037
3	0.0051	0.0053	0.0053	0.0034	0.0041
4	0.0039	0.0037	0.0044	0.0037	0.0048
5	0.0037	0.0039	0.0039	0.0034	0.0037
6	0.0041	0.0037	0.0044	0.0046	0.0046
7	0.0062	0.0051	0.0055	0.0053	0.0053
8	0.0021	0.0021	0.0055	0.0046	0.0053
9	0.0048	0.0048	0.0053	0.0051	0.0055
10	0.0030	0.0037	0.0037	0.0034	0.0034
เฉลี่ย	0.0041	0.0039	0.0046	0.0042	0.0045

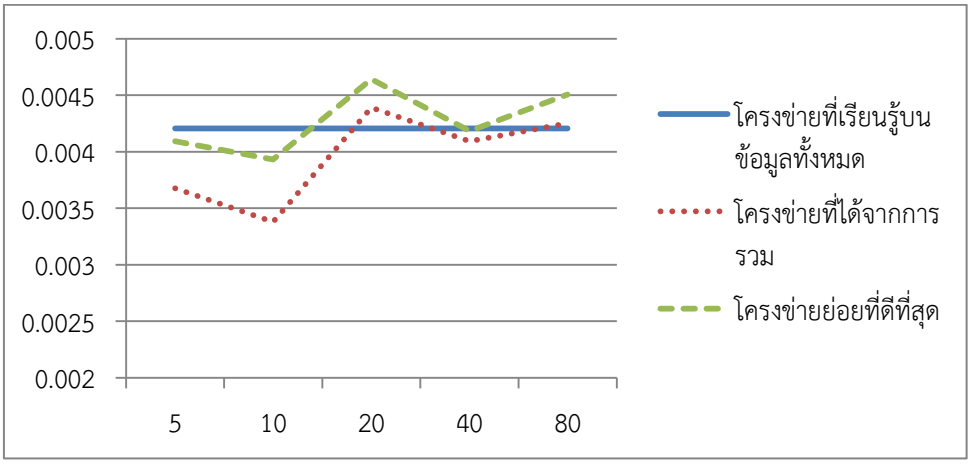
ตารางที่ 17 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล MiniBooNE

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.00000	-0.00207	0.00000	0.00000	0.00000
2	0.00000	-0.00138	0.00000	-0.00046	0.00000
3	0.00000	0.00000	0.00000	0.00000	0.00092
4	0.00000	0.00000	0.00000	0.00000	0.00000
5	-0.00046	0.00000	-0.00046	0.00000	-0.00115
6	0.00000	0.00000	0.00000	0.00000	0.00000
7	-0.00184	0.00000	-0.00023	0.00023	-0.00115
8	0.00023	-0.00023	-0.00115	-0.00069	-0.00069
9	-0.00184	-0.00161	0.00000	0.00000	0.00000
10	-0.00023	-0.00023	-0.00069	0.00000	-0.00046

ตารางที่ 18 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล Statlog

	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
$\sum D$	-0.004138	-0.005517	-0.002529	-0.000920	-0.002529
$\sum D^2$	0.000007	0.000009	0.000002	0.000001	0.000004
t	-1.694131	-2.166945	-2.012046	-1.077632	-1.275854
ระดับนัยสำคัญ	93.77579	97.07984	96.24601	84.53834	88.30251

โดยค่าความผิดพลาดสามารถแสดงได้ดังรูปที่ 31



รูปที่ 31 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล Statlog

4.4.1.5 King-Rook vs. King

ค่าความผิดพลาดในการทดลองของชุดข้อมูล King-Rook vs. King สามารถแสดงได้ในตารางที่ 19-21

ตารางที่ 19 แสดงค่าความผิดพลาดบนชุดข้อมูล King-Rook vs. King ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	0.6240	0.5681	0.5634	0.5862	0.6123	0.6750
2	0.7066	0.6396	0.6253	0.6816	0.7130	0.7123
3	0.6707	0.5649	0.5702	0.5966	0.6333	0.7028
4	0.6706	0.5608	0.5697	0.6357	0.6770	0.7419
5	0.6818	0.5916	0.6009	0.6793	0.6782	0.6939
6	0.5998	0.5770	0.5841	0.6066	0.6889	0.6429
7	0.6813	0.5686	0.6756	0.6681	0.7012	0.6902
8	0.6810	0.6461	0.6237	0.6689	0.6828	0.6903
9	0.6827	0.6253	0.6046	0.6677	0.6841	0.7027
10	0.6639	0.5485	0.5845	0.5727	0.7038	0.7092
เฉลี่ย	0.6662	0.5890	0.6002	0.6363	0.6775	0.6961

ตารางที่ 20 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล King-Rook vs. King

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.5780	0.5902	0.6301	0.6693	0.7345
2	0.6674	0.6367	0.7062	0.7658	0.7451
3	0.6037	0.5841	0.6568	0.6939	0.7480
4	0.5661	0.5900	0.6970	0.7255	0.7679
5	0.6115	0.6087	0.7167	0.7484	0.7644
6	0.5848	0.6101	0.6436	0.7124	0.6960
7	0.5704	0.7187	0.7248	0.7633	0.7369
8	0.6550	0.6572	0.7274	0.7327	0.7423
9	0.6424	0.6221	0.7102	0.7451	0.7729
10	0.5752	0.5937	0.5944	0.7666	0.7655
เฉลี่ย	0.6055	0.6212	0.6807	0.7323	0.7474

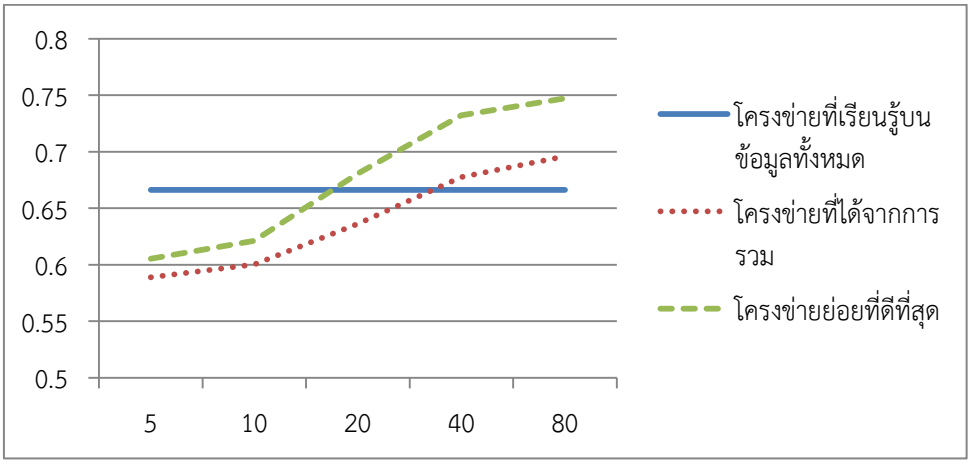
ตารางที่ 21 แสดงผลต่างของของความความผิดพลาดระหว่างนิวรอลเน็ตเวิร์กโครงข่ายที่ได้จากการรวมและโครงข่ายย่อยที่ดีที่สุดบนชุดข้อมูล King-Rook vs. King

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	-0.00998	-0.02673	-0.04383	-0.05702	-0.05952
2	-0.02781	-0.01141	-0.02460	-0.05276	-0.03280
3	-0.03885	-0.01390	-0.06023	-0.06058	-0.04526
4	-0.00535	-0.02032	-0.06132	-0.04848	-0.02602
5	-0.01996	-0.00784	-0.03742	-0.07021	-0.07056
6	-0.00784	-0.02602	-0.03706	-0.02352	-0.05310
7	-0.00178	-0.04314	-0.05668	-0.06203	-0.04670
8	-0.00891	-0.03350	-0.05845	-0.04989	-0.05203
9	-0.01711	-0.01747	-0.04242	-0.06096	-0.07023
10	-0.02673	-0.00927	-0.02174	-0.06272	-0.05631

ตารางที่ 22 แสดงการคำนวณค่านัยยะสำคัญทางสถิติบนชุดข้อมูล King-Rook vs. King

	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
$\sum D$	-0.16431	-0.20958	-0.44376	-0.54819	-0.51254
$\sum D^2$	0.00396	0.00556	0.02159	0.03153	0.02813
t	-4.39102	-5.81162	-9.66482	-13.53989	-11.26052
ระดับนัยสำคัญ	99.91285	99.98721	99.99976	99.99999	99.99993

โดยค่าความผิดพลาดสามารถแสดงได้ดังรูปที่ 32



รูปที่ 32 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กบนชุดข้อมูล King-Rook vs. King

4.4.2 เวลาที่ใช้ในการสอน

การเปรียบเทียบเวลาที่ใช้ในงานวิจัยนี้จะเป็นเวลาที่ใช้ในการสอน โดยนิวรอลเน็ตเวิร์กโครงข่ายที่สอนโดยชุดข้อมูลทั้งหมดจะเริ่มนับเวลาตั้งแต่เริ่มสอนจนถึงสิ้นสุดการสอน นิวรอลเน็ตเวิร์กที่สอนโดยเทคนิคในงานวิจัยนี้จะทำการเริ่มจับเวลาตั้งแต่แบ่งกลุ่มข้อมูลจนกระทั้งสิ้นสุดกระบวณการทั้งหมด โดยเวลาที่ใช้ในการสอนชุดข้อมูลต่างๆสามารถแสดงได้ดังต่อไปนี้

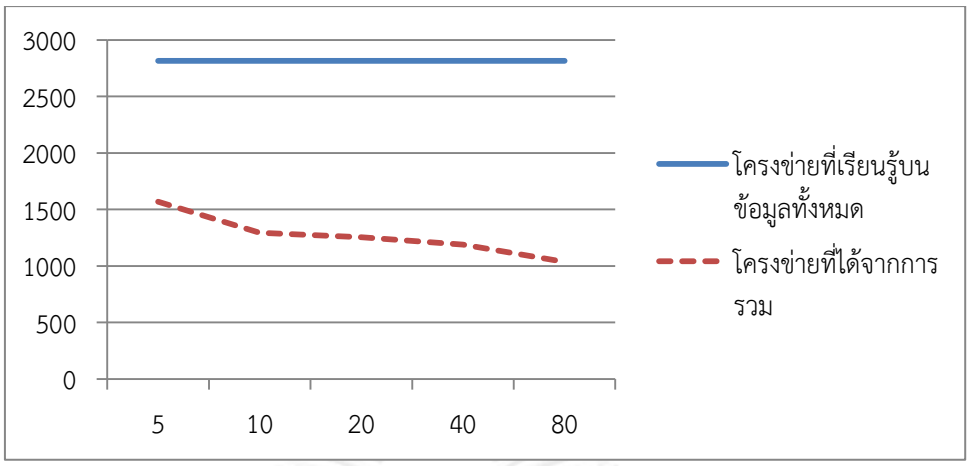
4.4.2.1 Covertype

เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กบนข้อมูล Covertype สามารถแสดงได้ในตารางที่ 23

ตารางที่ 23 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Covertype ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	2842	1339	1008	2174	917	1135
2	599	1535	1069	984	976	1274
3	3395	1538	1390	1049	1334	844
4	3402	1425	1208	1401	1156	754
5	2437	1768	1899	1669	1171	1077
6	4443	1595	1356	902	1219	994
7	1810	2097	1320	1189	1693	1229
8	2085	1195	899	885	1474	779
9	3357	1632	1282	1077	1080	864
10	3777	1572	1508	1204	888	1408
เฉลี่ย	2815	1570	1294	1253	1191	1036

โดยเวลาที่ใช้สอนสามารถแสดงได้ดังรูปที่ 33



รูปที่ 33 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Covertype

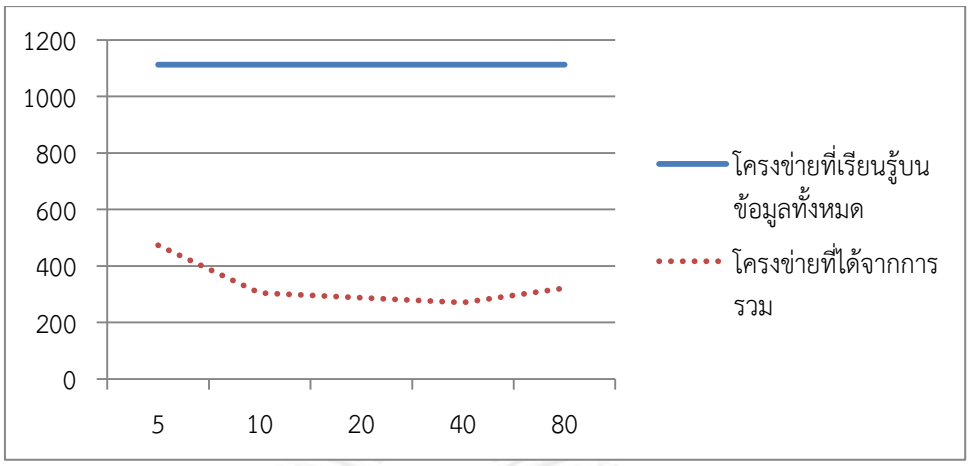
4.4.2.2 Synthesized

เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กบนข้อมูล Synthesized สามารถแสดงได้ในตารางที่ 24

ตารางที่ 24 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Synthesized ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	1060	541	328	302	308	280
2	1014	1023	478	415	520	643
3	1548	405	569	383	157	349
4	3154	887	292	226	438	225
5	570	224	242	274	145	252
6	513	302	197	145	165	281
7	1097	417	201	300	360	315
8	1024	263	318	508	153	193
9	742	273	191	171	181	409
10	407	400	226	150	282	274
เฉลี่ย	1113	474	304	287	271	322

โดยเวลาที่ใช้สอนสามารถแสดงได้ดังรูปที่ 34



รูปที่ 34 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Synthesized

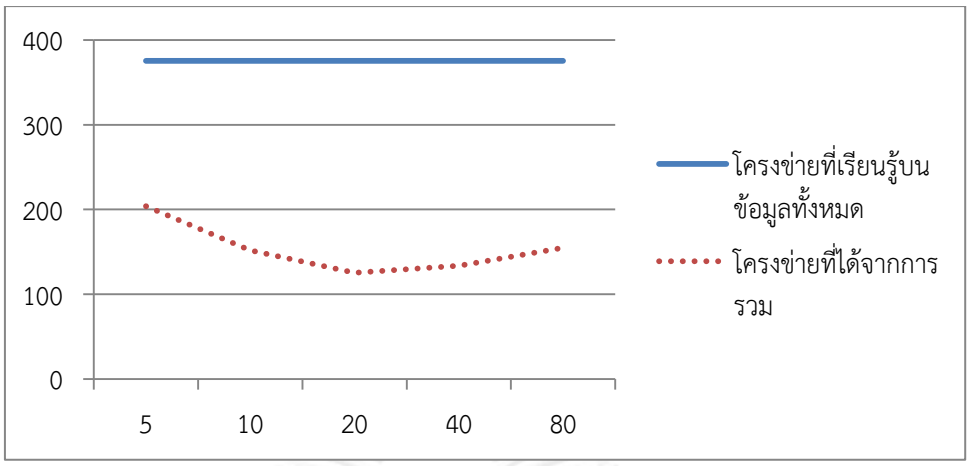
4.4.2.3 MiniBooNE

เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กบนข้อมูล MiniBooNE สามารถแสดงได้ในตารางที่ 25

ตารางที่ 25 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล MiniBooNE ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	420	194	147	131	119	123
2	210	140	131	104	136	131
3	194	151	149	104	112	158
4	521	241	173	137	149	127
5	219	177	158	179	132	153
6	203	364	157	130	137	180
7	209	186	155	142	113	157
8	207	305	140	114	190	152
9	909	154	152	118	113	169
10	662	127	156	96	137	196
เฉลี่ย	375	204	152	125	134	155

โดยเวลาที่ใช้สอนสามารถแสดงได้ดังรูปที่ 35



รูปที่ 35 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล MiniBooNE

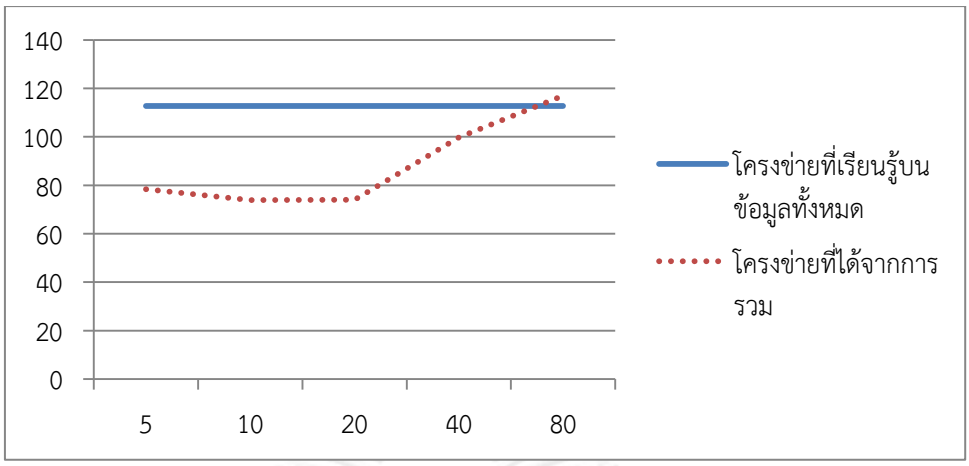
4.4.2.4 Statlog

เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กบนข้อมูล Statlog สามารถแสดงได้ในตารางที่ 26

ตารางที่ 26 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Statlog ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	166	64	100	151	98	122
2	159	117	42	118	102	146
3	122	47	81	47	75	147
4	144	111	78	105	64	98
5	130	145	79	56	137	137
6	84	101	91	64	107	141
7	81	53	57	54	157	85
8	87	42	44	41	113	110
9	75	52	73	51	64	118
10	81	52	93	51	81	66
เฉลี่ย	113	78	74	74	100	117

โดยเวลาที่ใช้สอนสามารถแสดงได้ดังรูปที่ 36



รูปที่ 36 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล Statlog

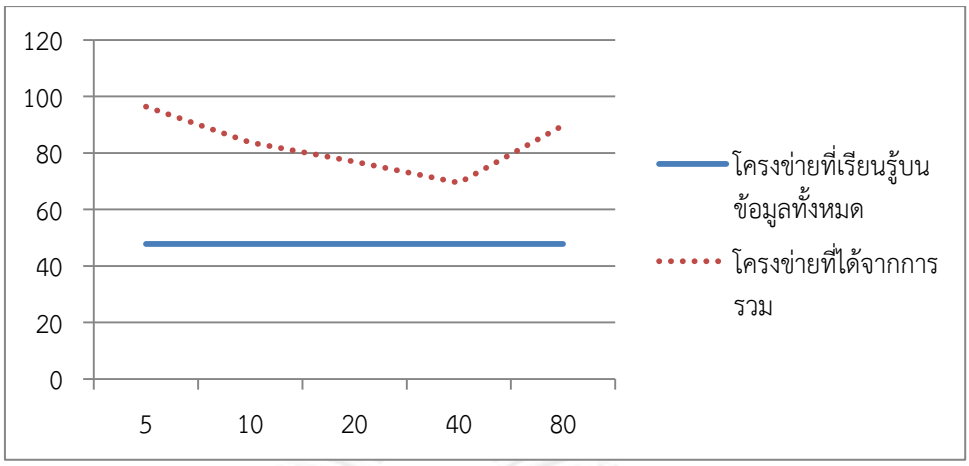
4.4.2.5 King-Rook vs. King

เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กบนข้อมูล King-Rook vs. King สามารถแสดงได้ในตารางที่ 26

ตารางที่ 27 เวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล King-Rook vs. King ของนิวรอลเน็ตเวิร์กโครงที่เรียนรู้ชุดข้อมูลทั้งหมด และโครงข่ายที่ได้จากการรวม

Fold	เรียนรู้บนข้อมูลทั้งหมด	จำนวนชุดข้อมูลย่อย				
		5	10	20	40	80
1	43	90	75	66	73	77
2	60	91	98	81	77	92
3	47	121	82	132	85	94
4	46	93	87	66	69	116
5	45	91	80	76	65	66
6	61	93	85	71	64	122
7	40	105	84	75	68	69
8	53	88	65	69	62	71
9	45	95	76	68	66	127
10	39	98	104	66	65	63
เฉลี่ย	48	96	84	77	69	90

โดยเวลาที่ใช้สอนสามารถแสดงได้ดังรูปที่ 37



รูปที่ 37 แสดงเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์ก(เป็นวินาที)บนชุดข้อมูล King-Rook vs. King

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุป

งานวิจัยนี้ได้นำเสนอวิธีการในการสอนนิวรอลเน็ตเวิร์กบนชุดข้อมูลขนาดใหญ่โดยการแบ่งชุดข้อมูลขนาดใหญ่ออกเป็นชุดข้อมูลย่อยๆจำนวน 2 กลุ่ม จากนั้นนำชุดข้อมูลย่อยภายในแต่ละกลุ่มไปทำการสอนนิวรอลเน็ตเวิร์ก ทำการหานิวรอลเน็ตเวิร์กโครงข่ายที่ดีที่สุดจากทั้ง 2 กลุ่ม จากนั้นทำการรวมนิวรอลเน็ตเวิร์กที่ดีที่สุด 2 โครงข่ายนั้นเข้าด้วยกัน

จากการผลการทดลองในบทที่ 4 จะเห็นได้ว่านิวรอลเน็ตเวิร์กที่ได้จากการรวมนั้น สามารถจำแนกตัวอย่างได้แม่นยำมากกว่านิวรอลเน็ตเวิร์กตั้งต้น และในกรณีที่ทำการแบ่งชุดข้อมูลใหญ่ออกเป็นชุดข้อมูลย่อยจำนวนไม่มาก ความแม่นยำของนิวรอลเน็ตเวิร์กที่ได้จากการรวมนั้น สามารถทำได้ใกล้เคียงกับนิวรอลเน็ตเวิร์กที่เรียนรู้ชุดข้อมูลขนาดใหญ่โดยตรง

การสอนนิวรอลเน็ตเวิร์กด้วยวิธีการที่งานวิจัยนี้นำเสนอนั้นสามารถลดเวลาที่ใช้ในการสอนนิวรอลเน็ตเวิร์กลงจากการสอนด้วยข้อมูลทั้งหมดประมาณ 50 เปอร์เซ็นต์บนชุดข้อมูลขนาดใหญ่ แต่หากชุดข้อมูลมีขนาดเล็กลง ประสิทธิภาพในการลดเวลาสอนนั้นอาจลดลง และหากชุดข้อมูลมีขนาดเล็ก อาจทำให้ต้องใช้เวลาในการสอนมากการการสอนนิวรอลเน็ตเวิร์กด้วยชุดข้อมูลขนาดใหญ่โดยตรง เช่นในชุดข้อมูลที่ 5

จำนวนของการแบ่งกลุ่มชุดข้อมูลนั้นส่งผลต่อทั้งความแม่นยำและเวลาที่ใช้ในการสอน การเพิ่มจำนวนการแบ่งกลุ่มนั้น จะทำให้ในการสอนนิวรอลเน็ตเวิร์กโครงข่ายย่อย โครงข่ายย่อยแต่ละโครงข่ายพบข้อมูลจำนวนไม่มาก ทำให้ใช้เวลาในการสอนไม่มากเช่นกัน แต่การที่นิวรอลเน็ตเวิร์กย่อยแต่ละโครงข่ายพบข้อมูลไม่มากจะมีข้อเสียคือนิวรอลเน็ตเวิร์กย่อยๆนั้นอาจไม่สามารถเรียนรู้รูปแบบของข้อมูลทั้งหมดได้ ทำให้ประสิทธิภาพของโครงข่ายลดลง ดังที่แสดงในผลการทดลอง เมื่อทำการแบ่งกลุ่มข้อมูลเพิ่มขึ้น ประสิทธิภาพของโครงข่ายย่อยที่ดีที่สุดจะลดลง และเมื่อโครงข่ายย่อยไม่สามารถเรียนรู้ชุดข้อมูลทั้งหมดได้ดี จึงจำเป็นต้องใช้เวลาในการปรับค่าถ่วงน้ำหนักในขั้นตอนสุดท้ายเพิ่มขึ้น ดังที่แสดงในผลการทดลองด้านเวลา เมื่อเพิ่มจำนวนการแบ่งกลุ่มจะทำให้เวลาที่ใช้ในการสอนลดลง แต่เมื่อเพิ่มจำนวนการแบ่งกลุ่มขึ้นเรื่องๆ กลับทำให้ต้องใช้เวลาการสอนเพิ่มขึ้น

5.2 แนวทางการพัฒนา

ในงานวิจัยนี้เป็นการนำเสนอการลดเวลาในการสอนนิวรอลเน็ตเวิร์กบนชุดข้อมูลขนาดใหญ่ โดยการแบ่งชุดข้อมูลขนาดใหญ่ออกเป็นชุดข้อมูลขนาดย่อยหลายชุด ซึ่งจำนวนในการแบ่งชุดข้อมูลนั้น ส่งผลต่อทั้งเวลาและความแม่นยำ

ในการพัฒนาต่อนั้นอาจทำการวิเคราะห์หาจำนวนที่เหมาะสมในการแบ่งชุดข้อมูล โดยอาจคำนวณจากจำนวนข้องข้อมูลทั้งหมดเพื่อจำนวนการแบ่งที่เหมาะสมซึ่งทำให้เวลาที่ใช้สอนและค่าความผิดพลาดในการจำแนกข้อมูลมีค่าน้อยที่สุด

อีกส่วนหนึ่งในการพัฒนาต่อได้แก่การวิเคราะห์ผลกระทบจากขั้นตอนต่างๆในงานวิจัยนี้ที่ส่งผลกระทบต่อความถูกต้องและเวลาที่ใช้ในการสอน หากสามารถทราบได้ว่าขั้นตอนใดส่งผลกระทบต่อประสิทธิภาพโดยรวมมากน้อยเพียงใดจะทำให้สามารถปรับขั้นตอนย่อยต่างๆเพื่อให้ได้ประสิทธิภาพที่สูงสุดได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

1. กิจศิริกุล, บ., ข่ายงาน ประสาทเทียม, in ปัญญาประดิษฐ์ *Artificial Intelligence*. 2548. p. 169-185.
2. Haykin, S., *Neural Networks and Learning Machines*. 3 ed. 2009, USA: Pearson Education.
3. แววเจริญ, ช.บ.ศ., ยูคลีเดียนในปริภูมิ *n* มิติ, in เมตริกซ์ พีชคณิตเชิงเส้น และการประยุกต์ *Matrix Linear Algebra and Their Applications*. 2546. p. 230-269.
4. Lagarias, J.C., J. A. Reeds, M. H. Wright, and P. E. Wright, *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions.* SIAM Journal of Optimization, 1998. **9**: p. 112-147.
5. วงศ์รัตนะ, ช., เทคนิคการใช้สถิติเพื่อการวิจัย. 12 ed. 2553.
6. Owens, A.J. *Empirical modeling of very large data sets using neural networks*. in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*. 2000.
7. Kala, R., A. Shulkla, and R. Tiwari. *Fuzzy Neuro Systems for Machine Learning for Large Data Sets*. in *Advance Computing Conference, 2009. IACC 2009. IEEE International*. 2009.
8. Badawy, O. and A. Almotwaly. *Combining neural network knowledge in a mobile collaborating multi-agent system*. in *Electrical, Electronic and Computer Engineering, 2004. ICEEC '04. 2004 International Conference on*. 2004.
9. Hashem, S. and B. Schmeiser, *Improving model accuracy using optimal linear combinations of trained neural networks.* Neural Networks, IEEE Transactions on, 1995. **6**(3): p. 792-794.
10. Qazi, N. and Y. Hoi. *Estimation of weights to combine trained neural networks using linear estimation techniques*. in *Multitopic Conference (INMIC), 2011 IEEE 14th International*. 2011.
11. Bahrami, M. *Integration of knowledge acquired by different neural networks*. in *Artificial Neural Networks and Expert Systems, 1993. Proceedings., First New Zealand International Two-Stream Conference on*. 1993.
12. Bache, K. and M. Lichman. *UCI Machine Learning Repository*. 2013 [cited 2014 25 May]; Available from: http://archive.ics.uci.edu/ml.

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

ตารางที่ 28 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล Covertype

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	5	6	4	7	5
2	2	6	5	5	4
3	3	4	5	7	7
4	2	3	7	6	8
5	1	1	6	3	5
6	5	5	5	4	8
7	1	3	5	2	5
8	5	8	8	6	6
9	3	3	5	4	7
10	2	4	6	6	7
เฉลี่ย	2.9	4.3	5.6	5	6.2

ตารางที่ 29 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล Synthesized

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0	0	0	0	0
2	3	1	2	1	2
3	2	1	0	0	1
4	1	1	1	3	2
5	0	0	0	0	2
6	0	0	0	0	1
7	1	0	1	2	1
8	0	1	1	0	0
9	0	0	0	0	1
10	1	1	0	1	1
เฉลี่ย	0.8	0.5	0.5	0.7	1.1

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 30 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล MiniBooNE

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	3	4	3	7	6
2	2	4	4	2	2
3	2	2	0	3	3
4	2	4	3	5	6
5	0	3	3	4	3
6	1	1	4	3	1
7	3	3	5	3	1
8	1	3	4	4	0
9	1	6	4	5	5
10	5	4	3	4	3
เฉลี่ย	2	3.4	3.3	4	3

ตารางที่ 31 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล Statlog

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	5	2	4	3	6
2	5	4	6	6	4
3	2	3	3	5	2
4	1	4	4	3	4
5	1	3	3	6	6
6	2	1	5	2	2
7	2	3	4	2	4
8	0	2	3	7	5
9	2	3	3	4	3
10	0	6	4	4	5
เฉลี่ย	2	3.1	3.9	4.2	4.1

ตารางที่ 32 จำนวนโหนดซ่อนที่ถูกเพิ่มเข้าไปในนิวรอลเน็ตเวิร์กโครงข่ายหลักบนชุดข้อมูล King-Rook vs. King

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	3	5	4	5	4
2	1	3	4	5	5
3	7	5	5	6	5
4	3	4	6	4	2
5	4	4	4	5	7
6	1	2	3	7	3
7	3	1	2	4	6
8	3	7	5	5	5
9	3	4	5	4	3
10	5	3	7	4	8
เฉลี่ย	3.3	3.8	4.5	4.9	4.8

ตารางที่ 33 แสดงค่าความผิดพลาดของนิวรอลเน็ตเวิร์กที่มีโหนดซ่อนเท่ากับนิวรอลเน็ตเวิร์กที่ได้จากการรวมและสอนบนชุดข้อมูล King-Rook vs. King โดยตรง

Fold	จำนวนชุดข้อมูลย่อย				
	5	10	20	40	80
1	0.6907	0.6746	0.6875	0.6782	0.6796
2	0.6128	0.6934	0.6738	0.6988	0.6492
3	0.6700	0.6746	0.6297	0.6964	0.6746
4	0.6503	0.6809	0.6446	0.6863	0.6845
5	0.6703	0.6636	0.6850	0.6447	0.6853
6	0.6885	0.5777	0.5966	0.6565	0.5859
7	0.6752	0.6652	0.6970	0.6699	0.5651
8	0.6739	0.6155	0.6440	0.6835	0.6230
9	0.6863	0.6888	0.6777	0.7009	0.5540
10	0.7017	0.6821	0.7085	0.7031	0.6468
เฉลี่ย	0.6720	0.6616	0.6644	0.6818	0.6348

ประวัติผู้เขียนวิทยานิพนธ์

นายต้า เกียรติไกรวัลศิริ เกิดเมื่อวันที่ 30 มีนาคม พ.ศ. 2533 ที่จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาบัณฑิต หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ จากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เมื่อปี พ.ศ. 2555 และได้เข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University