การเรียนรู้แบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในข่ายงานประสาทเทียม

นายอัครชัย อรรถกุลเรขา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

SIMULTANEOUS LEARNING WITH REMOVAL CRITERIA IN ARTIFICIAL NEURAL NETWORKS

Mr.Akarachai Atakulreka

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2007

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเรียนรู้แบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในข่ายงานประสาทเทียม
โดย	นายอัครชัย อรรถกุลเรขา
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.ดาริชา สุธีวงศ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์
(ผู้ช่วยศาสตราจารย์ ดร.ดาริชา สุธีวงศ์)

..... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.วีระ บุญจริง)

..... กรรมการ
(อาจารย์ ดร.วิษณุ โคตรจรัส)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

อัครชัย อรรถกุลเรขา : การเรียนรู้แบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในข่ายงานประสาทเทียม (SIMULTANEOUS LEARNING WITH REMOVAL CRITERIA IN ARTIFICIAL NEURAL NETWORKS)
อ.ที่ปรึกษา : ผศ. ดร.ดาริชา สุธีวงศ์, 61 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในการพิจารณาข่ายงานที่มีแนวโน้มของผลลัพธ์ที่ไม่ดี ซึ่งวิธีการนี้สามารถช่วยลดโอกาสการเกิดปัญหาค่าต่ำสุดเฉพาะที่ได้ พร้อมทั้งใช้ทรัพยากรในการเรียนรู้ได้อย่างมีประสิทธิภาพ โดยตัวแปรเปรียบเทียบที่ใช้ประมาณค่าผิดพลาดในอนาคตที่พิจารณาประกอบด้วย ค่าผิดพลาดต่ำสุด ค่าผิดพลาดท้ายสุด ความชันของค่าผิดพลาด และค่าผิดพลาดผสม ในงานวิจัยนี้ได้พิจารณาวิธีการกำหนดความถี่ของการพิจารณาเลือกหยุด (จำนวนระหว่างการเลือกหยุด) สองรูปแบบด้วยกัน กล่าวคือ แบบจำนวนรอบคงที่ และแบบจำนวนรอบเปลี่ยนแปลงตามความชัน โดยจำนวนข่ายงานที่พิจารณาเลือกหยุดในแต่ละครั้งสามารถกำหนดให้เป็นแบบคงที่หรือแบบเพิ่มขึ้นตามจำนวนรอบการเรียนรู้ การทดลองนี้ได้ใช้ชุดข้อมูลในการฝึกอบรมจากกลุ่มปัญหา Proben1 ซึ่งเป็นที่นิยมใช้ในการทดสอบประสิทธิภาพของวิธีการทางปัญญาประดิษฐ์ โดยผลจากการทดลองได้แสดงการเปรียบเทียบประสิทธิภาพของวิธีการที่นำเสนอกับวิธีการปกติ และการวิเคราะห์ความไวได้เปรียบเทียบและชี้ให้เห็นแนวทางของการกำหนดค่าตัวแปรต่างๆ ให้มีความเหมาะสมสำหรับการใช้งานวิธีการนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต อัครชัย อรรถกุลเรขา
สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา ดาริชา สุธีวงศ์
ปีการศึกษา 2550

4970695621 : MAJOR COMPUTER ENGINEERING
KEY WORD : ARTIFICIAL NEURAL NETWORKS / SIMULTANEOUS LEARNING / LOCAL MINIMA

AKARACHAI ATAKULREKA : SIMULTANEOUS LEARNING WITH REMOVAL CRITERIA IN ARTIFICIAL NEURAL NETWORKS.
THESIS ADVISOR : ASST. PROF. DARICHA SUTIVONG, Ph.D., 61 pp.

This research proposes a simultaneous training method with removal criteria to eliminate less promising neural networks. The algorithm helps decrease the probability of achieving a local minimum while efficiently utilizing resources. Comparative parameters proposed to estimate future error include: minimum error, last generation's error, error slope and mixed error. This research also explores two approaches in determining frequency of network removal (the number of epochs between removals), namely fixed-length interval and dynamic-length interval depending on the error slope. The number of networks to be removed may be fixed or incrementally increasing. The experiments used the dataset obtained from Proben1, which is widely used as training datasets in examining artificial intelligence methods. The experimental results demonstrate the effectiveness and efficiency of the proposed method in comparison with the conventional method. Moreover, sensitivity analysis offers comparison and guidelines for suitable assignment of several parameter values in using this algorithm.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department Computer Engineering Student's signature Akarachai Atakulreka
Field of study Computer Engineering Advisor's signature Darica Sutivong
Academic year 2007

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จออกมาได้ด้วยความกรุณาของ ผศ. ดร.ดาริชา สุธีวงศ์ อาจารย์ที่ปรึกษา ซึ่งได้ให้คำแนะนำและแนวทางที่เป็นประโยชน์ต่องานวิจัย คอยติดตามและดูแลให้งานวิจัยนี้สำเร็จออกมาด้วยดี

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ ผศ. ดร.โปรดปราน บุณยพุกกณะ อ. ดร.วิษณุ โคตรจรัส และ รศ. ดร.วีระ บุญจริง ที่กรุณาให้คำแนะนำและช่วยตรวจสอบแก้ไขวิทยานิพนธ์ฉบับนี้ และขอขอบพระคุณคณาจารย์ในภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ทุกท่านที่ประสิทธิ์ประสาทความรู้อันมีค่า

สุดท้ายนี้ ขอขอบพระคุณ คุณพ่อ คุณแม่ ญาติพี่น้องทุกคน ที่คอยดูแลห่วงใย เป็นกำลังใจ และให้การสนับสนุนในทุกๆ ด้านตลอดมา และขอขอบคุณพี่ๆ และเพื่อนๆ ทุกคนในภาควิชาที่คอยช่วยเหลือและให้คำปรึกษาในการทำงานวิทยานิพนธ์ฉบับนี้ให้สำเร็จลุล่วงในที่สุด และขอขมาในสิ่งที่ล่วงเกินไว้ ณ โอกาสนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญภาพ ..... ฌ
สารบัญตาราง ..... ญ
บทที่ 1 บทนำ 1
1.1 ที่มาและความสำคัญของปัญหา 1
1.2 วัตถุประสงค์ของการวิจัย 2
1.3 ขอบเขตของการวิจัย 2
1.4 ขั้นตอนและวิธีดำเนินการวิจัย 2
1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย 2
1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์ 3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4
2.1 ทฤษฎีที่เกี่ยวข้อง 4
2.1.1 ข่ายงานประสาทเทียม (Artificial neural networks) 4
2.2 งานวิจัยที่เกี่ยวข้อง 8
2.2.1 งานวิจัยเกี่ยวกับปัญหาค่าต่ำสุดเฉพาะที่ในข่ายงานประสาทเทียม 8
2.2.2 งานวิจัยด้านการประยุกต์ใช้แนวคิดการทำงานแบบพร้อมกัน 9
บทที่ 3 การเรียนรู้แบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในข่ายงานประสาทเทียม 11
3.1 การเรียนรู้แบบพร้อมกันในข่ายงานประสาทเทียม 11
3.2 ตัวแปรเปรียบเทียบของการพิจารณาเลือกหยุดข่ายงาน 12
3.2.1 ค่าผิดพลาดต่ำสุด 12
3.2.2 ค่าผิดพลาดท้ายสุด 12
3.2.3 ความชันของค่าผิดพลาด 12
3.2.4 ค่าผิดพลาดผสม 13
3.3 การกำหนดจำนวนรอบที่พิจารณาเลือกหยุด 14
3.3.1 การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่ 15
3.3.2 การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชัน 16
3.4 การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุด 17
3.4.1 การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่ 17
3.4.2 การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบเพิ่มขึ้น 17

หน้า

3.5 การกำหนดจำนวนรอบสูงสุด 18
บทที่ 4 การทดลอง 19
4.1 ข้อมูลที่ใช้ในการฝึกอบรมและการทดสอบ 19
4.2 การกำหนดรูปแบบของข่ายงานประสาทเทียม 21
4.3 ขั้นตอนการทดลอง 22
บทที่ 5 ผลการทดลองและวิเคราะห์ผล 23
5.1 การเปรียบเทียบคุณภาพคำตอบระหว่างวิธีการฝึกอบรมแบบต่าง ๆ 23
5.2 การเปรียบเทียบเวลาที่ใช้ระหว่างวิธีการฝึกอบรมแบบต่าง ๆ 25
5.3 ผลกระทบของการกำหนดจำนวนรอบและข่ายงานที่พิจารณาเลือกหยุดแบบคงที่ 27
5.3.1 การวิเคราะห์ความไวของรอบที่พิจารณาเลือกหยุดครั้งแรก 28
5.3.2 การวิเคราะห์ความไวของรอบที่พิจารณาเลือกหยุดครั้งสุดท้าย 30
5.3.3 การวิเคราะห์ความไวของจำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้ง 33
5.4 ผลกระทบของการกำหนดจำนวนรอบและข่ายงานที่พิจารณาเลือกหยุดแบบไม่คงที่ 35
5.4.1 การวิเคราะห์ความไวของรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชัน 35
5.4.2 การวิเคราะห์ความไวของจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบเพิ่มขึ้น 37
5.5 ผลกระทบของการกำหนดจำนวนรอบสูงสุด 39
บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ 44
6.1 สรุปผลการวิจัย 44
6.2 ข้อจำกัดและข้อเสนอแนะ 44
รายการอ้างอิง 46
ภาคผนวก 49
บทความทางวิชาการ 50
ประวัติผู้เขียนวิทยานิพนธ์ 61

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 2.1 โครงสร้างของหน่วยเซลล์ในข่ายงานประสาทเทียม 4
รูปที่ 2.2 โครงสร้างของข่ายงานประสาทเทียมป้อนไปหน้าแบบหลายชั้น 5
รูปที่ 2.3 รหัสเทียมการทำงานของอัลกอริทึมแบบการแพร่กระจายย้อนกลับ 7
รูปที่ 3.1 ลำดับขั้นตอนของการฝึกอบรมข่ายงานประสาทเทียม 11
รูปที่ 3.2 ตัวอย่างการใช้ค่าผิดพลาดผสมในการเปรียบเทียบระหว่างข่ายงาน 14
รูปที่ 3.3 ตัวอย่างการฝึกอบรมที่มีการพิจารณาเลือกหยุดแบบคงที่ 16
รูปที่ 5.1 แผนภูมิเปรียบเทียบคุณภาพของคำตอบของวิธีการฝึกอบรมแบบต่าง ๆ 24
รูปที่ 5.2 แผนภูมิเปรียบเทียบเวลาที่ใช้ของวิธีการฝึกอบรมแบบต่าง ๆ 27
รูปที่ 5.3 แผนภูมิเปรียบเทียบค่าผิดพลาดของค่ารอบพิจารณาเลือกหยุดครั้งแรก 29
รูปที่ 5.4 แผนภูมิเปรียบเทียบจำนวนรอบของค่ารอบพิจารณาเลือกหยุดครั้งแรก 30
รูปที่ 5.5 แผนภูมิเปรียบเทียบค่าผิดพลาดของค่ารอบพิจารณาเลือกหยุดครั้งสุดท้าย 31
รูปที่ 5.6 แผนภูมิเปรียบเทียบจำนวนรอบของค่ารอบพิจารณาเลือกหยุดครั้งสุดท้าย 32
รูปที่ 5.7 แผนภูมิเปรียบเทียบค่าผิดพลาดของค่าจำนวนข่ายงานเลือกหยุด 34
รูปที่ 5.8 แผนภูมิเปรียบเทียบจำนวนรอบของค่าจำนวนข่ายงานเลือกหยุด 34
รูปที่ 5.9 จำนวนรอบสูงสุดเปรียบเทียบระหว่างการกำหนดจำนวนรอบต่าง ๆ 40
รูปที่ 5.10 ความสัมพันธ์ของค่าผิดพลาดจากจำนวนรอบประมาณความชันและวัดการลู่เข้า ... 41
รูปที่ 5.11 ส่วนเบี่ยงเบนมาตรฐานของค่าผิดพลาดในการกำหนดตัวแปรจำนวนรอบทั้งสอง ... 42

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 4.1 จำนวนข้อมูลฝึกอบรมและทดสอบที่ใช้ในแต่ละปัญหา 20
ตารางที่ 4.2 รายละเอียดโครงสร้างของข่ายงานที่ใช้ในแต่ละปัญหา..... 21
ตารางที่ 5.1 ผลการเปรียบเทียบคุณภาพของคำตอบ (ค่าผิดพลาด) ของวิธีการแบบต่าง ๆ 24
ตารางที่ 5.2 ผลการเปรียบเทียบเวลาที่ใช้ (จำนวนรอบ) ของวิธีการแบบต่าง ๆ..... 26
ตารางที่ 5.3 การกำหนดค่าพื้นฐานของตัวแปร..... 28
ตารางที่ 5.4 ค่าผิดพลาดจากการกำหนดรอบที่พิจารณาเลือกหยุดครั้งแรก 28
ตารางที่ 5.5 ค่าผิดพลาดจากการกำหนดรอบที่พิจารณาเลือกหยุดครั้งสุดท้าย..... 31
ตารางที่ 5.6 ค่าผิดพลาดจากการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุด..... 33
ตารางที่ 5.7 ค่าผิดพลาดและจำนวนรอบที่ใช้ของการกำหนดรูปแบบรอบที่พิจารณา 36
ตารางที่ 5.8 ค่าผิดพลาดของจำนวนข่ายงานที่พิจารณาแบบคงที่และแบบเพิ่มขึ้น..... 38
ตารางที่ 5.9 จำนวนรอบที่ใช้ของจำนวนข่ายงานที่พิจารณาแบบคงที่และแบบเพิ่มขึ้น 38
ตารางที่ 5.10 ค่าจำนวนรอบสูงสุดที่คำนวณจากการกำหนดตัวแปรให้มีค่าต่าง ๆ..... 41

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ข่ายงานประสาทเทียม (Artificial neural networks) เป็นหนึ่งในวิธีการเรียนรู้ของเครื่อง (Machine learning) ซึ่งเป็นการจำลองการทำงานบางส่วนของสมองมนุษย์ในรูปแบบทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศ โดยได้มีงานวิจัยที่นำข่ายงานประสาทเทียมมาประยุกต์ใช้ในรูปแบบต่างๆ มากมาย ทั้งในการรู้จำ (Recognition) การพยากรณ์ (Prediction) หรือ การจัดแบ่งประเภท (Classification) ด้วยลักษณะที่เป็นเครื่องมือที่มีความสามารถในการเรียนรู้ แต่อย่างไรก็ตาม วิธีการเรียนรู้ของข่ายงานประสาทเทียมยังคงถูกปรับปรุงและพัฒนาขึ้นอย่างต่อเนื่อง

หนึ่งในปัญหาสำคัญที่มักจะพบในการเรียนรู้ของข่ายงานประสาทเทียม คือ ปัญหาค่าต่ำสุดเฉพาะที่ (Local minimum problem) ซึ่งเกิดจากการปรับค่าน้ำหนักโดยใช้หลักการของการเคลื่อนลงตามความชัน (Gradient descent) ตัวอย่างเช่น อัลกอริทึมการแพร่กระจายย้อนกลับ (Backpropagation) ด้วยเหตุนี้จึงได้มีงานวิจัยต่างๆ ที่ศึกษาการเกิดปัญหาค่าต่ำสุดเฉพาะที่ [1-6] ซึ่งวิธีการแก้ไขที่น่าสนใจและใช้กันแพร่หลาย คือ การฝึกอบรมข่ายงานประสาทเทียมมากกว่าหนึ่งข่ายงานโดยใช้ชุดค่าน้ำหนักแตกต่างกันในแต่ละข่ายงาน และเลือกใช้ข่ายงานที่ได้ผลลัพธ์ดีที่สุด [7, 8] วิธีการนี้เป็นวิธีที่นำไปใช้ได้ง่ายและสามารถนำไปประยุกต์กับอัลกอริทึมอื่นๆ ได้ แต่อย่างไรก็ตาม วิธีการนี้ยังมีข้อจำกัดในเรื่องเวลาที่ใช้ในการฝึกอบรมที่ค่อนข้างมาก ดังนั้นจึงเป็นที่มาของการเสนอวิธีการฝึกอบรมข่ายงานประสาทเทียมที่อาศัยการเรียนรู้แบบพร้อมกัน โดยใช้ร่วมกับหลักเกณฑ์การพิจารณาเลือกหยุดข่ายงานที่มีแนวโน้มที่จะให้ผลลัพธ์ที่ไม่ดี ซึ่งตัวแปรเปรียบเทียบที่นำมาใช้ในการพิจารณาเลือกหยุดนี้จะพิจารณาจากค่าผิดพลาดต่างๆ ของข้อมูลทดสอบ โดยนำค่าที่ได้จากข่ายงานต่างๆมาเปรียบเทียบกัน ข้อดีของวิธีการนี้ คือ การได้รับประโยชน์จากการฝึกอบรมข่ายงานประสาทเทียมมากกว่าหนึ่งข่ายงาน ซึ่งช่วยลดโอกาสการเกิดปัญหาค่าต่ำสุดเฉพาะที่ได้ แต่ยังใช้ทรัพยากรในการฝึกอบรมที่ไม่มากเกินไป

ในการทดสอบประสิทธิภาพของวิธีการฝึกอบรมข่ายงานประสาทเทียมนี้จะถูกนำมาเปรียบเทียบกับวิธีการฝึกอบรมแบบปกติเพื่อเปรียบเทียบคุณภาพของคำตอบและเวลาที่ใช้ในการฝึกอบรม โดยใช้หลักเกณฑ์ในการพิจารณาเลือกหยุดที่ได้นำเสนอแบบต่างๆ การวิเคราะห์แบบนี้จะทำให้เห็นข้อดีข้อเสียในการใช้หลักเกณฑ์แต่ละแบบและแสดงให้เห็นผลกระทบของแต่ละตัวแปรที่กำหนด นอกจากนี้งานวิจัยนี้ได้มีการศึกษาและวิเคราะห์การกำหนดค่าที่ใช้ในการเรียนรู้ต่างๆ เพื่อนำเสนอแนวทางการใช้งานที่ทำให้เกิดประสิทธิภาพสูงสุดด้วย

1.2 วัตถุประสงค์ของการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการฝึกอบรมข่ายงานประสาทเทียมแบบเรียนรู้พร้อมกันหลายข่ายงานและเกณฑ์การพิจารณาเลือกหยุดข่ายงาน เพื่อให้การฝึกอบรมมีโอกาสการเกิดปัญหาค่าต่ำสุดเฉพาะที่ลดลงพร้อมกับใช้ทรัพยากรในการเรียนรู้อย่างมีประสิทธิภาพ

1.3 ขอบเขตของการวิจัย

1. นำเสนอวิธีการฝึกอบรมข่ายงานประสาทเทียมชนิดป้อนไปหน้าแบบหลายชั้นที่เรียนรู้พร้อมกันหลายข่ายงาน พร้อมทั้งเสนอเกณฑ์การพิจารณาเลือกหยุดข่ายงาน
2. เปรียบเทียบประสิทธิภาพของวิธีที่นำเสนอกับการฝึกอบรมข่ายงานประสาทเทียมแบบปกติ โดยพิจารณาจากค่าผิดพลาดของชุดข้อมูลทดสอบ
3. วิเคราะห์ความไว (Sensitivity analysis) ของจำนวนข่ายงานที่เรียนรู้พร้อมกันและระยะจำนวนรอบที่พิจารณาเลือกหยุดข่ายงาน

1.4 ขั้นตอนและวิธีดำเนินการวิจัย

1. ศึกษาทฤษฎีพื้นฐานของข่ายงานประสาทเทียม
2. ศึกษาวิธีการแก้ไขปัญหาค่าต่ำสุดเฉพาะที่ของข่ายงานประสาทเทียม
3. ออกแบบวิธีการฝึกอบรมข่ายงานประสาทเทียมแบบเรียนรู้พร้อมกัน
4. ออกแบบหลักเกณฑ์การพิจารณาเลือกหยุดข่ายงานประสาทเทียม
5. ทดสอบวิธีการที่นำเสนอ
6. วิเคราะห์และสรุปผลการทดลองโดยการวิเคราะห์ความไว
7. จัดทำวิทยานิพนธ์ พร้อมทั้งนำเสนองานวิจัยทั้งหมด

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

ได้วิธีการฝึกอบรมข่ายงานประสาทเทียมแบบเรียนรู้พร้อมกันหลายข่ายงาน พร้อมทั้งหลักเกณฑ์การพิจารณาเลือกหยุดข่ายงานที่เหมาะสม ที่แสดงให้เห็นประสิทธิภาพในการแก้ปัญหาค่าต่ำสุดเฉพาะที่โดยใช้ทรัพยากรในการเรียนรู้อย่างมีประสิทธิภาพ โดยสามารถใช้แนวทางในการกำหนดการฝึกอบรมเพื่อให้ผลลัพธ์ที่ได้มีความเหมาะสมตามความต้องการ ซึ่งแนวคิดนี้ยังสามารถใช้ในวิธีการคำนวณแบบขนานเพื่อลดเวลาในการเรียนรู้และสามารถประยุกต์ใช้กับวิธีการฝึกอบรมแบบอื่นต่อไปได้

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของงานวิทยานิพนธ์ได้รับการตีพิมพ์เป็นบทความวิชาการในหัวเรื่อง “Avoiding Local Minima in Feedforward Neural Networks by Simultaneous Learning” โดย อัครชัย อรรถกุลเรขา และ ดาริชา สุธีวงศ์ ในงานประชุมวิชาการ “20th Australian Joint Conference on Artificial Intelligence” ซึ่งจัดขึ้น ณ เมืองโกลด์โคสต์ รัฐควีนส์แลนด์ ประเทศออสเตรเลีย ระหว่างวันที่ 4-6 ธันวาคม 2550 ดังรายละเอียดในภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

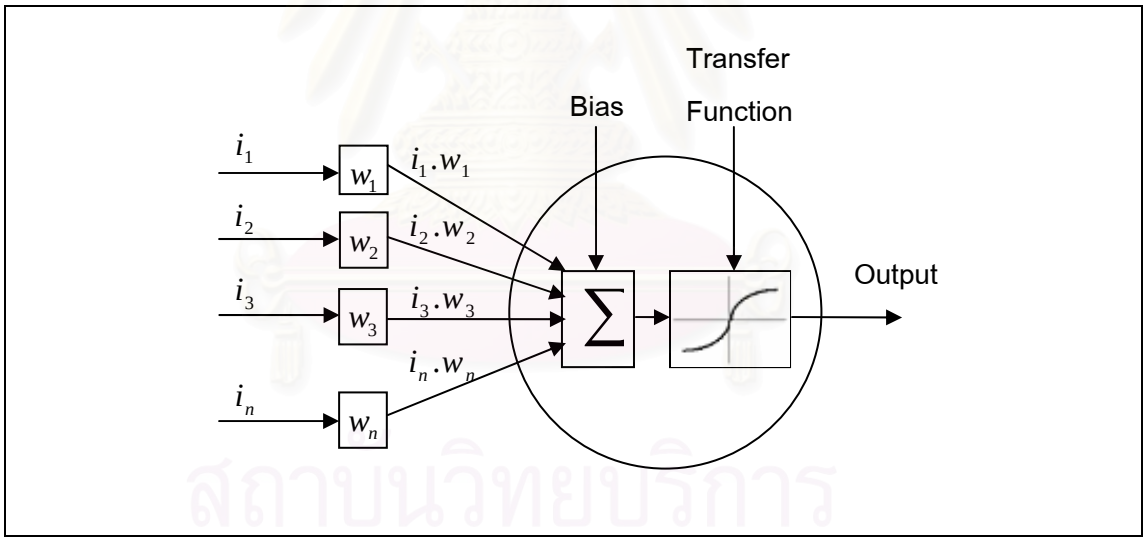
บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ข่ายงานประสาทเทียม (Artificial neural networks)

ข่ายงานประสาทเทียมเป็นหนึ่งในวิธีการเรียนรู้ของเครื่อง (Machine learning) ที่จำลองการทำงานบางส่วนของสมองมนุษย์ ซึ่งประกอบด้วยเซลล์ประสาทที่เชื่อมโยงกันหลายๆ เซลล์ เมื่อเซลล์ได้รับสัญญาณไฟฟ้าเคมี เซลล์ก็จะถูกกระตุ้นและส่งสัญญาณไปยังเซลล์อื่นๆ ข่ายงานประสาทเทียมได้จำลองการทำงานเหล่านี้ โดยแต่ละเซลล์จะมีส่วนที่เป็นข้อมูลขาเข้า (Input) และมีการรวมผล (Summation) ของข้อมูลขาเข้าที่เข้ามาปรับด้วยค่าน้ำหนัก (Weight) ซึ่งจะนำผลรวมนี้ไปเข้าฟังก์ชันเชื่อมโยง (Transfer function) หรือฟังก์ชันกระตุ้น (Activation function) เพื่อนำผลที่ได้ออกมาเป็นข้อมูลขาออก (Output) โดยลักษณะของหน่วยเซลล์ในข่ายงานประสาทเทียมสามารถแสดงได้ดังรูปที่ 2.1

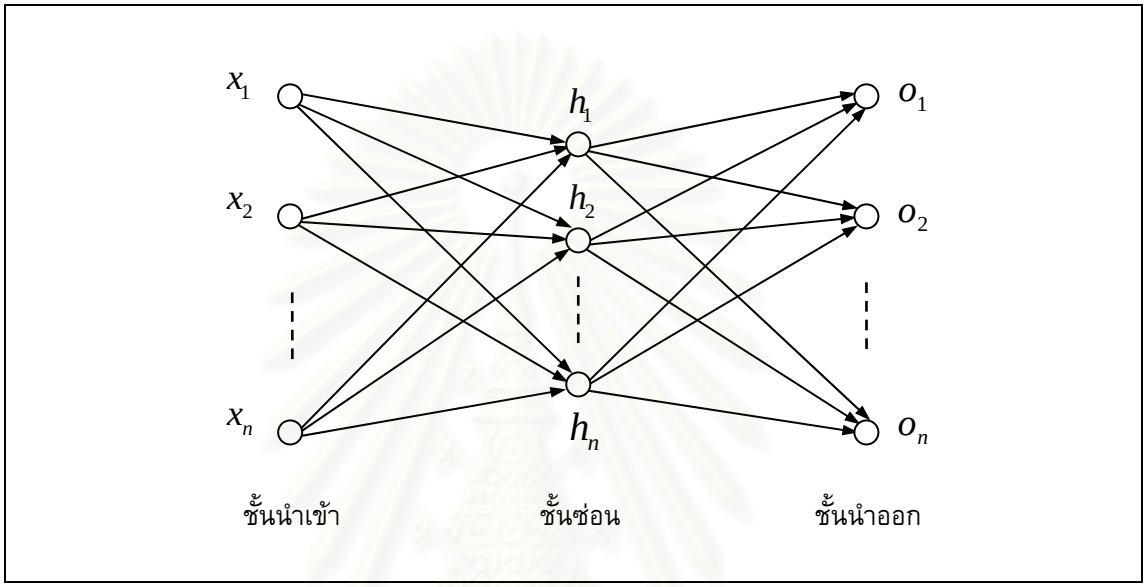


รูปที่ 2.1 โครงสร้างของหน่วยเซลล์ในข่ายงานประสาทเทียม

1) ข่ายงานป้อนไปหน้าแบบหลายชั้น (Multilayer feedforward networks)

เมื่อเซลล์หลายๆ เซลล์ประกอบกันจะเกิดเป็นข่ายงานประสาทเทียมขึ้นมา ซึ่งรูปแบบหนึ่งที่ใช้กันอย่างแพร่หลาย คือ ข่ายงานประสาทเทียมแบบหลายชั้น (Multilayer neural networks) ดังแสดงในรูปที่ 2.2 ข่ายงานนี้เป็นข่ายงานประสาทเทียมแบบส่งผ่านข้อมูลไปข้างหน้า (Feedforward networks) โดยภายในข่ายงานชนิดนี้จะมีการแบ่งออกเป็นชั้นๆ

ประกอบด้วย ชั้นนำเข้า (Input layer) ชั้นซ่อน (Hidden layer) และชั้นนำออก (Output layer) ในแต่ละชั้นจะมีบัพซึ่งเป็นเหมือนเซลล์ที่มีการเชื่อมต่อกันระหว่างชั้น โดยเส้นเชื่อมจะเชื่อมต่อในทิศทางเดียวระหว่างเซลล์เฉพาะในชั้นที่ติดกันตั้งแต่ชั้นนำเข้าไปจนกระทั่งชั้นนำออก เมื่อมีการส่งผ่านข้อมูลของแต่ละเซลล์จะส่งผ่านไปยังเซลล์ที่อยู่ในชั้นถัดไป ไม่มีการส่งไปยังเซลล์ที่อยู่ในชั้นเดียวกันหรือชั้นก่อนหน้า สำหรับชั้นซ่อนอาจมีจำนวนชั้นได้มากกว่าหนึ่งชั้นก็ได้



รูปที่ 2.2 โครงสร้างของข่ายงานประสาทเทียมป้อนไปหน้าแบบหลายชั้น

2) อัลกอริทึมการแพร่กระจายย้อนกลับ (Backpropagation algorithm)

อัลกอริทึมการแพร่กระจายย้อนกลับ (Backpropagation algorithm) เป็นวิธีการเรียนรู้หนึ่งของข่ายงานประสาทเทียม อัลกอริทึมนี้จะใช้การเคลื่อนลงตามความชัน (Gradient descent) ในการปรับเวกเตอร์น้ำหนักเพื่อให้ข่ายงานมีค่าผิดพลาดต่ำที่สุด ซึ่งกระบวนการเรียนรู้และปรับปรุงแก้ไขจะเป็นไปโดยอัตโนมัติ ถ้าข่ายงานให้คำตอบที่ผิดในข้อมูลสอน (Training data) ค่าน้ำหนักก็จะถูกปรับจนกว่าค่าความผิดพลาดจะลดลงหรืออยู่ในเกณฑ์ที่ยอมรับได้ จากโครงสร้างของข่ายงานข้างต้น เมื่อข่ายงานได้รับข้อมูลนำเข้าก็จะคำนวณค่าน้ำหนักจากบัพข้อมูลเข้าไปยังแต่ละบัพของชั้นซ่อน และคำนวณน้ำหนักจากบัพในชั้นซ่อนไปยังบัพที่เป็นผลลัพธ์ เมื่อคำนวณถึงบัพที่เป็นชั้นนำออกแล้วก็จะได้ผลลัพธ์ ซึ่งเมื่อทำการเปรียบเทียบผลลัพธ์ที่ได้กับค่าจริงจะได้ค่าความต่างระหว่างผลลัพธ์ จากนั้นข่ายงานจะพยายามปรับค่าความผิดพลาดจากชั้นนำออกโดยการปรับค่าเวกเตอร์น้ำหนักและจะปรับค่าแพร่ย้อนกลับไปยังชั้นซ่อนแต่ละชั้นต่อไป

ในการปรับค่าเวกเตอร์น้ำหนักโดยขั้นตอนวิธีแบบการแพร่กระจายย้อนกลับนั้น มีการนิยามค่าผิดพลาดการฝึกอบรม ($E(\vec{w})$) เพื่อนำไปใช้ในการหาค่าเวกเตอร์น้ำหนักที่ทำให้ค่าผิดพลาดนี้มีค่าต่ำที่สุด ซึ่งนิยามค่าผิดพลาดการฝึกอบรมแสดงดังสมการ 2.1

$$E(\vec{w}) = \frac{1}{2}\sum_{d \in D}\sum_{k \in outputs}(t_{kd} - o_{kd})^2 \quad (2.1)$$

เมื่อ $\vec{w}$ คือ เวกเตอร์น้ำหนักของข่ายงาน

$outputs$ คือ เซตของบัพนำออกในข่ายงาน

t_{kd} คือ ค่าผลลัพธ์เป้าหมาย (Target output)

o_{kd} คือ ค่าผลลัพธ์ที่ได้จากข่ายงาน (Actual output)

D คือ จำนวนตัวอย่าง

ค่าผิดพลาดที่ได้นี้นอกจากจะนำมาใช้ในการปรับค่าเวกเตอร์น้ำหนักแล้ว ยังบ่งบอกถึงความพอดีกับตัวอย่างทั่วไป (Generalization) ซึ่งเป็นความสามารถในการให้คำตอบที่ถูกต้องของข่ายงานประสาทเทียมที่ผ่านการเรียนรู้ไม่เฉพาะเพียงกับชุดข้อมูลสอนเท่านั้น แต่สามารถใช้เพื่อหาคำตอบกับชุดข้อมูลอื่นๆ ได้เป็นอย่างดี ดังนั้น เพื่อเป็นการตรวจสอบความสามารถนี้ขณะกำลังฝึกอบรมข่ายงานประสาทเทียม จึงต้องมีการตรวจสอบกับข้อมูลประเมิน (Validation data) โดยชุดข้อมูลนี้แบ่งมาจากข้อมูลทั้งหมดที่มีอยู่ก่อนการฝึกอบรม ส่วนชุดข้อมูลอีกหนึ่งชุดที่นำมาใช้ในทดสอบประสิทธิภาพของข่ายงาน คือ ข้อมูลทดสอบ (Test data) ชุดข้อมูลนี้จะถูกใช้ในการทดสอบกับข่ายงานภายหลังการฝึกอบรมเสร็จสิ้น เพื่อประเมินการทำงานของข่ายงานกับข้อมูลรูปแบบใหม่ๆ ซึ่งเปรียบเสมือนข้อมูลที่เกิดขึ้นจากการใช้งานจริง

ขั้นตอนวิธีการปรับเวกเตอร์น้ำหนักสำหรับข่ายงานประสาทเทียมที่ใช้อัลกอริทึมแบบการแพร่กระจายย้อนกลับสามารถแสดงเป็นรหัสเทียม (Pseudo code) ได้ดังรูปที่ 2.3

For each training example, do

1. Input the training example to the network and compute the network outputs
2. For each output unit k

$$\delta_k \leftarrow o_k(1-o_k)(t_k-o_k)$$

3. For each hidden unit h

$$\delta_h \leftarrow o_h(1-o_h)\sum_{k\in outputs} w_{h,k}\delta_k$$

4. Update each network weight $w_{i,j}$

$$w_{i,j} \leftarrow w_{i,j} + \Delta w_{i,j}$$

where

$$\Delta w_{i,j} = \eta \delta_j x_{i,j}$$

รูปที่ 2.3 รหัสเทียมการทำงานของอัลกอริทึมแบบการแพร่กระจายย้อนกลับ

เมื่อ w คือ ค่าน้ำหนักของข่ายงาน

δ คือ ค่าผลลัพธ์ที่ได้จากข่ายงาน (Error term)

η คือ อัตราการเรียนรู้ (Learning rate)

$x_{i,j}$ คือ ข้อมูลนำเข้า (Input)

จากการใช้อัลกอริทึมการแพร่กระจายย้อนกลับในการเรียนรู้จำเป็นต้องมีการกำหนดค่าอัตราการเรียนรู้และโมเมนตัม ซึ่งค่าเหล่านี้มีความไวต่อการเปลี่ยนแปลงมาก จึงได้มีการพัฒนาอัลกอริทึมขึ้นมาใหม่ คือ อัลกอริทึมการแพร่กระจายแบบยืดหยุ่น (Resilient propagation: RPROP) [9] ซึ่งเป็นอัลกอริทึมที่มีการประยุกต์มาจากอัลกอริทึมการแพร่กระจายย้อนกลับเดิม แต่อัลกอริทึมนี้จะมีการปรับเพิ่ม-ลดค่าน้ำหนักตามการเปลี่ยนแปลงของค่าผิดพลาด ทำให้ไม่ต้องคำนึงถึงการกำหนดค่าอัตราการเรียนรู้และโมเมนตัม นอกจากนั้นค่าผิดพลาดที่ได้จากการเรียนรู้ยังสามารถลู่เข้าได้เร็วและมีการผันผวนน้อย ดังนั้นในงานวิจัยนี้จึงได้เลือกใช้อัลกอริทึมนี้ในการทดลอง

การทำงานของอัลกอริทึมที่ใช้แนวคิดแบบเคลื่อนลงตามความชันมีหลักการทำงานโดยการค้นหาเวกเตอร์น้ำหนักที่ทำให้ค่าผิดพลาดมีค่าต่ำที่สุด แต่ในกรณีการใช้ข่ายงานป้อนไปหน้าแบบหลายชั้น จุดที่ให้ค่าต่ำสุดมักมีมากกว่าหนึ่งจุด ดังนั้นคำตอบที่ได้จากการใช้อัลกอริทึม

นี้จึงอาจเป็นค่าต่ำสุดเฉพาะที่ (Local minima) จากปัญหานี้เองจึงเป็นที่มาของการวิจัยเพื่อหาวิธีการแก้ไขเพื่อให้ได้คำตอบของการเรียนรู้ที่น่าพอใจ

2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ศึกษาเกี่ยวกับการแก้ไขปัญหาค่าต่ำสุดเฉพาะที่ เพื่อให้ข่ายงานประสาทเทียมที่ผ่านการฝึกอบรมเป็นข่ายงานที่สามารถให้คำตอบที่ดี ซึ่งงานวิจัยนี้จะเสนอวิธีการฝึกอบรมข่ายงานประสาทเทียมที่อาศัยการเรียนรู้แบบพร้อมกัน ดังนั้นจึงได้ศึกษาถึงปัญหาและวิธีการแก้ไขปัญหาค่าต่ำสุดเฉพาะที่ในงานวิจัยต่างๆ เพื่อแสดงให้เห็นลักษณะของปัญหาพร้อมทั้งแนวคิดในการจัดการกับปัญหานี้ และได้ศึกษาเกี่ยวกับแนวคิดการทำงานแบบพร้อมกันที่นำไปใช้ในงานอื่นๆ ซึ่งงานวิจัยที่เกี่ยวข้องนี้จะเป็นองค์ความรู้ที่จะนำไปใช้ในการกำหนดวิธีการฝึกอบรมเพื่อแก้ปัญหาต่อไป

2.2.1 งานวิจัยเกี่ยวกับปัญหาค่าต่ำสุดเฉพาะที่ในข่ายงานประสาทเทียม

ในการฝึกอบรมข่ายงานประสาทเทียมโดยทั่วไปแล้วมักจะประสบกับปัญหาค่าต่ำสุดเฉพาะที่ เนื่องจากการเรียนรู้ส่วนใหญ่จะอาศัยการปรับค่าน้ำหนักที่ใช้หลักการของการเคลื่อนลงตามความชัน ซึ่งหลักการนี้จะมีการปรับค่าน้ำหนักของเส้นเชื่อมต่อภายในข่ายงานจากค่าผิดพลาดที่ได้จากการเรียนรู้ชุดข้อมูลสอนเพื่อให้ได้ค่าน้ำหนักที่เหมาะสม แต่บ่อยครั้งที่การปรับค่าน้ำหนักปรับเข้าสู่ค่าต่ำสุดเฉพาะที่ของพื้นผิวค่าผิดพลาด (Error surface) ด้วยเหตุนี้จึงได้มีงานวิจัยต่างๆ ที่ศึกษาและวิเคราะห์ถึงการเกิดของปัญหานี้ [1-6] โดยการพิจารณาลักษณะของโครงสร้างของข่ายงานประสาทเทียมและสภาวะการเรียนรู้ที่ทำให้ไม่เกิดปัญหาค่าต่ำสุดเฉพาะที่ และในงานของ Sprinkhuizen-Kuyper และ Boers [10] ได้จำแนกและอธิบายประเภทของค่าต่ำสุดเฉพาะที่เพื่อทำให้เกิดความเข้าใจในพฤติกรรมของพื้นผิวค่าผิดพลาดบริเวณใกล้เคียงค่าต่ำสุดเฉพาะที่

โดยแท้จริงแล้วค่าต่ำสุดเฉพาะที่เกี่ยวข้องกับปัจจัยหลักๆ 2 ปัจจัยด้วยกัน คือ รูปแบบการเรียนรู้ และโครงสร้างของข่ายงาน ดังนั้นวิธีการต่างๆ ที่ถูกเสนอเพื่อแก้ไขปัญหานี้จึงมุ่งเน้นที่การปรับปรุงสองปัจจัยนี้ โดยวิธีที่นำมาใช้ในการแก้ไขสามารถแบ่งออกได้เป็น 2 แนวทาง คือ แนวทางเชิงกำหนด (Deterministic approach) และแนวทางเชิงน่าจะเป็น (Probabilistic approach) ซึ่งในแต่ละแนวทางมีงานวิจัยที่เกี่ยวข้องดังนี้

สำหรับแนวทางเชิงกำหนด Cetin และคณะ [11] ได้นำเสนอวิธีการเรียนรู้โกลเบิลเดสเซนท์ (Global descent) เพื่อมาแทนที่หลักการของการเคลื่อนลงตามความชันเดิม โดยวิธีการนี้ได้นำอัลกอริทึมการหาค่าเหมาะสมที่สุด (Global optimization) มาประยุกต์ใช้ ทำให้การเรียนรู้สามารถออกจากค่าต่ำสุดเฉพาะที่ได้ อย่างไรก็ตามยังคงมีข้อจำกัดเกี่ยวกับการไม่มีข้อกำหนดการหยุดที่แน่นอน จึงไม่เหมาะกับการใช้ในปัญหาที่ไม่รู้แนวโน้มของคำตอบ ซึ่ง

ต่อมางานวิจัยของ Toh [12] ได้ศึกษาวิธีการเรียนรู้โกลเบิลเดสเซนท์นี้ถึงผลกระทบของการใช้ทรัพยากรในการฝึกอบรมที่เพิ่มขึ้นมากเมื่อเทียบกับวิธีปกติ เช่นเดียวกับงานวิจัยอื่นที่คล้ายกันของ Jordanov และ Rafik [13] ที่ยังคงมีปัญหาในเรื่องนี้

ส่วนในแนวทางเชิงน่าจะเป็น โดยส่วนใหญ่มักจะให้ความสนใจในการปรับปรุงเกี่ยวกับชุดค่าน้ำหนักของข่ายงาน เช่น ในงานของ Wessels และ Barnard [14] ได้เสนอวิธีการกำหนดค่าเริ่มต้นของค่าน้ำหนักเพื่อลดโอกาสการเกิดปัญหาค่าต่ำสุดเฉพาะที่ อีกวิธีหนึ่งที่น่าสนใจ คือ การฝึกอบรมข่ายงานประสาทเทียมมากกว่าหนึ่งข่ายงานโดยใช้ชุดค่าน้ำหนักแตกต่างกันในแต่ละข่ายงานและเลือกใช้ข่ายงานที่ดีที่สุด โดยมักพิจารณาจากค่าผิดพลาดที่ต่ำที่สุดจากการทดสอบ ตัวอย่างในงานของ Park และคณะ [7] ได้นำเสนอวิธีการในการกำหนดรูปแบบของข่ายงานในการพยากรณ์จุดดับบนดวงอาทิตย์ ในงานวิจัยนี้ได้ใช้วิธีการเลือกข่ายงานที่ดีที่สุดจากข่ายงานที่ฝึกอบรมทั้งหมด 10 ข่ายงาน ซึ่งแต่ละข่ายงานจะกำหนดชุดค่าน้ำหนักเริ่มต้นที่แตกต่างกันโดยการสุ่ม อย่างไรก็ตาม การใช้วิธีการนี้ยังไม่มีการกำหนดจำนวนครั้งที่ฝึกอบรมข่ายงานที่แน่นอน ต่อมา Iyer และ Rhinehart [8] ได้ใช้วิธีเชิงน่าจะเป็นในการกำหนดจำนวนครั้งของการเริ่มต้นฝึกอบรมข่ายงานด้วยชุดค่าน้ำหนักที่สุ่มขึ้นมาใหม่ เพื่อให้ข่ายงานที่ได้มีความน่าจะเป็นต่ำที่จะเกิดปัญหาค่าต่ำสุดเฉพาะที่ แม้ว่าวิธีการนี้จะเป็นวิธีที่นำมาใช้ได้ง่าย แต่ยังจำเป็นต้องอาศัยเวลาค่อนข้างมากในการฝึกอบรม

ข่ายงานอีกประเภทหนึ่งที่กำลังถูกศึกษาอย่างกว้างขวาง คือ ข่ายงานประสาทเทียมเชิงวิวัฒน์ ข่ายงานประเภทนี้เป็นการนำแนวคิดอัลกอริทึมเชิงวิวัฒน์ (Evolutionary algorithm) มาประยุกต์ใช้กับข่ายงานประสาทเทียม โดยได้นำมาใช้กับข่ายงานในระดับหลักๆ 3 ระดับด้วยกัน คือ ค่าน้ำหนักเชื่อมโยง โครงสร้าง และ กฎการเรียนรู้ของข่ายงาน ตัวอย่างเช่น EPNet [15] ที่เป็นหนึ่งในระบบที่มีการนำวิธีข่ายงานประสาทเทียมเชิงวิวัฒน์มาใช้ ซึ่งระบบนี้จะใช้ตัวดำเนินการการกลายพันธุ์ (Mutation) ในการปรับปรุงค่าน้ำหนักและโครงสร้างของข่ายงาน โดยงานวิจัยของ Yao [16] ได้ทำการสำรวจเกี่ยวกับข่ายงานประสาทเทียมเชิงวิวัฒน์ไว้ และในงาน [17, 18] ได้แสดงผลการทดสอบเปรียบเทียบข่ายงานประเภทนี้ไว้ ซึ่งแสดงให้เห็นถึงประสิทธิภาพในการค้นหาค่าต่ำสุดที่แท้จริง แต่จำเป็นต้องแลกกับทรัพยากรในการคำนวณเป็นอย่างมาก

2.2.2 งานวิจัยด้านการประยุกต์ใช้แนวคิดการทำงานแบบพร้อมกัน

แนวคิดการทำงานแบบพร้อมกันได้ถูกนำไปประยุกต์ใช้ในงานวิจัยต่างๆ เพื่อจุดประสงค์ในการพัฒนาการทำงานให้มีประสิทธิภาพมากยิ่งขึ้น ทั้งทางการเพิ่มความสามารถในการทำงานจากการนำข้อมูลของแต่ละงานมาใช้ร่วมกัน การลดเวลาการทำงานรวมให้น้อยลงเนื่องจากสามารถทำงานแบบขนานได้ และการนำแนวคิดนี้มาใช้ช่วยแก้ไขปัญหาบางประการในงานนั้นๆ อีกด้วย

ตัวอย่างงานวิจัยที่ได้นำแนวคิดการทำงานแบบพร้อมกันมาประยุกต์ใช้ เช่น การนำไปใช้ในสภาพแวดล้อมที่มีหลายผู้กระทำ (Multi-agent environment) ดังในงานของ Ikenoue และคณะ [19] และงานของ Takahashi และคณะ [20] ซึ่งได้นำเสนอวิธีการเรียนรู้แบบพร้อมกันไปใช้ในงานที่อาศัยความร่วมมือกันทำงาน เช่น การแข่งขันฟุตบอลหุ่นยนต์ โดยผลที่ได้จากการเรียนรู้นี้ทำให้ทราบข้อมูลภาพรวมของการทำงาน ซึ่งช่วยลดขอบเขตการทำงานต่อไปให้น้อยลง และทำให้การทำงานของหุ่นยนต์แต่ละตัวสามารถปรับตัวให้สอดคล้องกับตัวอื่นๆ ได้ดียิ่งขึ้น ส่วนการนำวิธีการเรียนรู้แบบพร้อมกันไปใช้เพื่อเพิ่มประสิทธิภาพการเรียนรู้ในงานอื่นๆ ให้สามารถทำงานได้มากกว่าหนึ่งงานในเวลาใดๆ เช่น ในงานที่ศึกษาการเรียนรู้เกี่ยวกับภาพ งานวิจัยของ Artac และคณะ [21] ได้แสดงวิธีการเรียนรู้และจดจำภาพไปในเวลาเดียวกัน เพื่อลดปัญหาการจัดการกับภาพที่มีขนาดใหญ่ โดยวิธีการที่ใช้นี้จะปรับปรุงการทำงานให้เป็นแบบเพิ่มขึ้น (Increment) และในงานของ Declercq และ Piater [22] ได้เสนอแนวทางการเรียนรู้กราฟแสดงลักษณะของภาพ (Visual feature graphs) ซึ่งในการเรียนรู้จะทำแบบเพิ่มขึ้นโดยอาศัยข้อมูลที่ได้จากที่ผ่านมา นำมาช่วยในการติดตาม (Tracking) เพื่อเรียนรู้ต่อไป สำหรับในการนำแนวคิดนี้มาใช้ในการเรียนรู้ของข่ายงานประสาทเทียม งานวิจัยของ Liu และ Yao [23] ได้เสนอระบบการเรียนรู้ของข่ายงานประสาทเทียมที่ทำงานร่วมกัน (Neural network ensemble) โดยวิธีการนี้จะทำให้ข่ายงานแต่ละข่ายงานมีความแตกต่างกันเพื่อให้ข่ายงานทั้งหมดสามารถเรียนรู้ชุดข้อมูลได้อย่างครอบคลุม และการฝึกอบรมจะทำแบบพร้อมกันเพื่อให้มีการสร้างความสัมพันธ์ระหว่างข่ายงานจากข้อมูลที่แลกเปลี่ยนกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

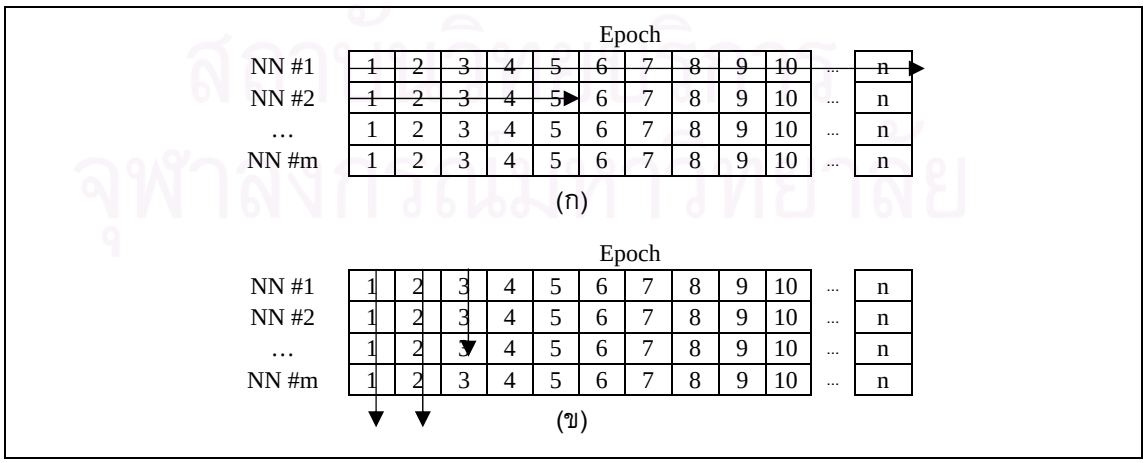
บทที่ 3

การเรียนรู้แบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในข่ายงานประสาทเทียม

งานวิจัยนี้นำเสนอวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในการพิจารณาข่ายงานที่มีแนวโน้มของผลลัพธ์ที่ไม่ดี ทั้งนี้ได้นำแนวคิดการแก้ไขปัญหาค่าต่ำสุดเฉพาะที่โดยการฝึกอบรมข่ายงานประสาทเทียมแบบหลายข่ายงานโดยใช้ชุดค่าน้ำหนักที่แตกต่างกันมาประยุกต์ใช้ เพื่อให้การฝึกอบรมมีโอกาสการเกิดปัญหาค่าต่ำสุดเฉพาะที่ลดลงพร้อมทั้งใช้ทรัพยากรในการเรียนรู้อย่างมีประสิทธิภาพ โดยวิธีการเรียนรู้ที่นำเสนอนี้ประกอบด้วยแนวคิดที่แบ่งออกเป็นส่วนต่างๆ คือ การเรียนรู้แบบพร้อมกัน ตัวแปรเปรียบเทียบของการพิจารณาเลือกหยุดข่ายงาน การกำหนดจำนวนรอบที่พิจารณาเลือกหยุด การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุด และการกำหนดจำนวนรอบสูงสุด โดยในแต่ละส่วนมีรายละเอียดดังต่อไปนี้

3.1 การเรียนรู้แบบพร้อมกันในข่ายงานประสาทเทียม

แนวทางการเรียนรู้แบบพร้อมกันในข่ายงานประสาทเทียมนี้ได้นำแนวคิดการทำงานแบบขนานมาใช้ในการฝึกอบรมข่ายงาน ซึ่งวิธีการนี้จะมีลำดับขั้นตอนในการทำงานที่แตกต่างจากการฝึกอบรมข่ายงานประสาทเทียมแบบปกติ เนื่องจากการฝึกอบรมแบบปกตินั้นจะเริ่มต้นจากการสร้างข่ายงานขึ้นมาหนึ่งข่ายงานด้วยการสุ่มค่าน้ำหนักเริ่มต้นและฝึกอบรมข่ายงานนั้นโดยการปรับค่าน้ำหนักจนกระทั่งครบจำนวนรอบสูงสุดที่ได้กำหนดไว้ ในกรณีที่ต้องการฝึกอบรมซ้ำจะสร้างข่ายงานขึ้นมาใหม่ด้วยการสุ่มค่าน้ำหนักเริ่มต้นอีกชุดหนึ่งและทำตามขั้นตอนเช่นเดิม ซึ่งตัวอย่างขั้นตอนจะแสดงในรูปที่ 3.1(ก) โดยช่องแต่ละช่องแสดงรอบที่ฝึกอบรม (Epoch) ของข่ายงานแต่ละข่ายงาน และลูกศรแสดงลำดับของการเรียนรู้



รูปที่ 3.1 ลำดับขั้นตอนของการฝึกอบรมข่ายงานประสาทเทียม
(ก) แบบปกติ (ข) แบบพร้อมกัน

ส่วนวิธีการฝึกอบรมที่นำเสนอจะฝึกอบรมข่ายงานประสาทเทียมหลายๆ ข่ายงานไปพร้อมกัน เริ่มต้นจากการสร้างข่ายงานด้วยการสุ่มค่าน้ำหนักเริ่มต้นที่แตกต่างกัน แล้วทำการฝึกอบรมแต่ละข่ายงานไปทีละรอบ กล่าวคือ เริ่มต้นจากรอบแรกของข่ายงานแต่ละข่ายงาน และทำเช่นเดียวกันในรอบต่อๆ ไป ซึ่งเมื่อนำวิธีการนี้ไปใช้กับหน่วยประมวลผลแบบหลายแกน (Multiple core processors) ข่ายงานทั้งหมดจะสามารถถูกฝึกอบรมได้ในเวลาเดียวกัน โดยตัวอย่างขั้นตอนการเรียนรู้แสดงดังรูปที่ 3.1(ข) ลำดับการฝึกอบรมที่ได้เสนอนี้ทำให้สามารถทราบถึงแนวโน้มของค่าผิดพลาดได้ เนื่องจากข่ายงานที่เรียนรู้พร้อมกันหลายข่ายงานจะแสดงข้อมูลของค่าผิดพลาดได้ดียิ่งขึ้น ซึ่งข้อมูลเหล่านี้จะช่วยในการกำหนดจำนวนรอบสูงสุดของการฝึกอบรมได้เหมาะสมมากขึ้น สำหรับแนวโน้มของค่าผิดพลาดจะถูกนำไปใช้ในการเปรียบเทียบระหว่างข่ายงานเพื่อพิจารณาเลือกหยุดฝึกอบรมข่ายงานที่ไม่ดีโดยใช้ตัวแปรเปรียบเทียบของการพิจารณาเลือกหยุดข่ายงานซึ่งจะกล่าวถึงในหัวข้อถัดไป

3.2 ตัวแปรเปรียบเทียบของการพิจารณาเลือกหยุดข่ายงาน

การฝึกอบรมข่ายงานประสาทเทียมหลายๆ ข่ายงาน และเลือกข่ายงานที่ให้ค่าผิดพลาดน้อยที่สุด สามารถช่วยให้หลีกเลี่ยงการเกิดปัญหาค่าต่ำสุดเฉพาะที่ได้ แต่อย่างไรก็ตามการใช้วิธีการนี้มีข้อเสียที่ใช้ทรัพยากรในการเรียนรู้ที่ค่อนข้างมาก ดังนั้นเพื่อลดข้อบกพร่องนี้ งานวิจัยนี้จึงเสนอตัวแปรเปรียบเทียบของการพิจารณาเลือกหยุดข่ายงานซึ่งจะนำมาใช้ร่วมกันในการฝึกอบรมแบบพร้อมกัน เพื่อเลือกหยุดข่ายงานที่มีแนวโน้มที่จะให้ผลลัพธ์ที่ไม่ดี โดยค่าที่ใช้ในการเปรียบเทียบนี้ได้กำหนดขึ้นมาเป็นเหมือนกับตัวพยากรณ์ลักษณะของค่าผิดพลาดในอนาคต ซึ่งข่ายงานที่มีค่าที่แย่ที่สุดจะถูกหยุดการฝึกอบรม โดยค่าผิดพลาดที่นำมาพิจารณานี้จะใช้ค่าผิดพลาดจากการทดสอบข่ายงานด้วยชุดข้อมูลประเมิน ซึ่งค่าที่ใช้ในการเปรียบเทียบทั้งหมดประกอบด้วย 4 ค่า ดังนี้

3.2.1 ค่าผิดพลาดต่ำสุด

ค่าผิดพลาดต่ำสุดเป็นค่าผิดพลาดที่มีค่าต่ำที่สุดจากรอบฝึกอบรมทั้งหมดที่ผ่านมา

3.2.2 ค่าผิดพลาดท้ายสุด

ค่าผิดพลาดท้ายสุดเป็นค่าผิดพลาดที่ได้จากการฝึกอบรมรอบท้ายสุดของข่ายงานประสาทเทียม

3.2.3 ความชันของค่าผิดพลาด

ความชันของค่าผิดพลาดเป็นการประมาณค่าความชันของเส้นตรงที่มาจากการถดถอยเชิงเส้น (Linear regression) ของค่าผิดพลาดในรอบท้ายสุด 50 รอบ

3.2.4 ค่าผิดพลาดผสม

ค่าผิดพลาดผสมเป็นค่าผิดพลาดที่ประมาณด้วยค่าผิดพลาดท้ายสุดและนำมาลดลงด้วยแนวโน้มของค่าผิดพลาด ดังแสดงในสมการที่ 3.1

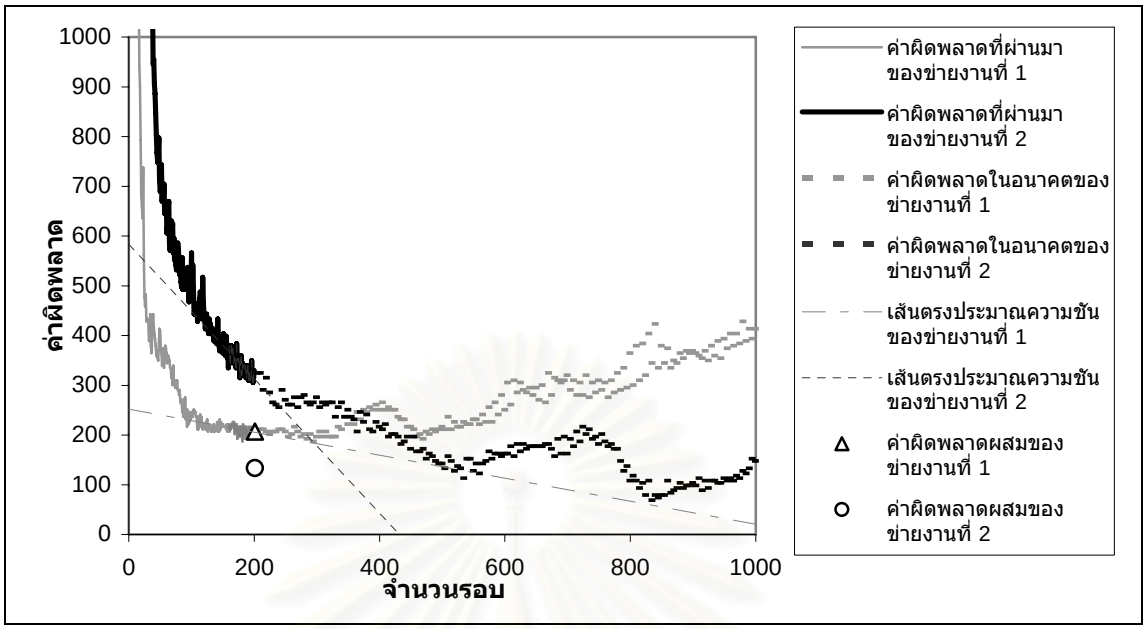
$$mixedError = lastGenError \times (1 - (\arctan(errorSlope) / (\pi / 2))) \quad (3.1)$$

เมื่อ *mixError* คือ ค่าผิดพลาดผสม

lastGenError คือ ค่าผิดพลาดท้ายสุด

errorSlope คือ ค่าความชันของค่าผิดพลาด

ในการกำหนดค่าที่ใช้ในการเปรียบเทียบเพื่อเลือกข่ายงานที่มีแนวโน้มที่ไม่ดีให้หยุดการเรียนรู้เพื่อลดเวลาและทรัพยากรที่ใช้ให้น้อยลงนั้น ได้พิจารณาจากค่าที่สามารถเป็นตัวชี้วัดว่าข่ายงานไหนดีหรือไม่ดี ตัวอย่างเช่น ค่าผิดพลาดต่ำสุด เป็นค่าที่ใช้บ่งบอกได้ว่าข่ายงานนั้นเคยให้ผลลัพธ์ที่ดีสุดเท่าใด ในกรณีที่ปัญหาที่กำลังฝึกอบรมอยู่เกิดความพอดีกับตัวอย่างมากเกินไป (Overfitting) ค่าผิดพลาดต่ำสุดนี้ยังคงสามารถใช้เปรียบเทียบได้อย่างถูกต้อง แม้ว่าค่าผิดพลาดจะมีแนวโน้มที่สูงขึ้นแล้วก็ตาม ส่วนความชันของค่าผิดพลาด สามารถใช้บอกได้ว่าที่รอบการฝึกอบรมเดียวกัน ข่ายงานไหนมีแนวโน้มการลดลงของค่าผิดพลาดน้อยหรือมาก เป็นต้น แต่สำหรับกรณีการเปรียบเทียบข่ายงานในรูปที่ 3.2 ซึ่งเป็นตัวอย่างการใช้ค่าผิดพลาดผสมในการเปรียบเทียบระหว่างข่ายงานในรอบการสอนที่ 200 ข่ายงานหนึ่งให้ค่าผิดพลาดที่ลดลงเร็วในช่วงแรกแต่หลังจากนั้นจะเริ่มมีค่าคงที่ (ข่ายงานที่ 1) กับอีกข่ายงานหนึ่งมีค่าผิดพลาดลดลงทีละน้อยอย่างต่อเนื่อง (ข่ายงานที่ 2) การพิจารณาเลือกหยุดข่ายงานโดยใช้ค่าผิดพลาดต่ำสุดหรือค่าผิดพลาดท้ายสุดอาจทำให้การเลือกหยุดข่ายงานเกิดการผิดพลาดได้ เนื่องจากข่ายงานที่ 2 มีโอกาสที่จะให้ค่าผิดพลาดที่ลดลงได้อีกมากในอนาคต ซึ่งอาจทำให้ได้ค่าที่ต่ำกว่าข่ายงานที่ 1 ก็ได้ ด้วยเหตุนี้จึงเป็นที่มาของการกำหนดค่าผิดพลาดผสมขึ้นมา ค่าผิดพลาดผสมนี้เป็นค่าที่คำนวณมาจากค่าผิดพลาดท้ายสุดที่ปรับลดค่าลงเป็นสัดส่วนตามค่าความชันของค่าผิดพลาดจากการฝึกอบรมข่ายงานนั้น ดังเช่นตัวอย่างในรูปที่ 3.2 เมื่อพิจารณาหยุดข่ายงานในรอบการสอนที่ 200 แม้ว่าทั้งค่าผิดพลาดต่ำสุดและค่าผิดพลาดท้ายสุดของข่ายงานที่ 1 จะน้อยกว่าข่ายงานที่ 2 แต่ที่จริงแล้วค่าผิดพลาดในอนาคตของข่ายงานที่ 2 มีค่าน้อยกว่า ซึ่งจากการเปรียบเทียบโดยการใช้ค่าผิดพลาดผสมของข่ายงานที่ 1 (จุดสามเหลี่ยมสีเทา) มีค่ามากกว่าข่ายงานที่ 2 (จุดวงกลมสีดำ) ทำให้การพิจารณาเลือกหยุดข่ายงานสามารถทำได้อย่างถูกต้อง



รูปที่ 3.2 ตัวอย่างการใช้ค่าผิดพลาดผสมในการเปรียบเทียบระหว่างข่ายงาน

3.3 การกำหนดจำนวนรอบที่พิจารณาเลือกหยุด

แนวคิดในการเรียนรู้แบบพร้อมกันของข่ายงานประสาทเทียมที่ใช้เกณฑ์การเลือกหยุดที่ได้นำเสนอนี้ สิ่งสำคัญของการเรียนรู้แบบพร้อมกัน คือ การกำหนดค่าตัวแปรต่างๆ สำหรับเป็นเกณฑ์ในการพิจารณาเลือกหยุด ตัวอย่างเช่น รอบที่พิจารณาเลือกหยุดครั้งแรก (First elimination epoch) รอบที่พิจารณาเลือกหยุดครั้งสุดท้าย (Last elimination epoch) และจำนวนข่ายงานที่เลือกหยุดในแต่ละครั้ง (No. of networks eliminated) ซึ่งค่าตัวแปรเหล่านี้มีความสำคัญอย่างมากต่อประสิทธิภาพของคำตอบที่ได้และทรัพยากรที่ใช้ของการฝึกอบรมข่ายงานประสาทเทียม ดังนั้น การกำหนดค่าจำนวนรอบและจำนวนข่ายงานที่พิจารณาเลือกหยุดนี้จึงจำเป็นต้องกำหนดให้มีค่าที่เหมาะสม โดยในหัวข้อนี้จึงได้นำเสนอถึงแนวคิดและวิธีการทำงานของค่าตัวแปรจำนวนรอบที่พิจารณาเลือกหยุด พร้อมทั้งยกตัวอย่างการฝึกอบรมที่มีการพิจารณาเลือกหยุดตามค่าตัวแปรต่างๆ ตามที่กำหนด

วิธีการกำหนดจำนวนรอบที่มีการพิจารณาเลือกหยุดนี้ทำได้หลายวิธี ซึ่งวิธีที่นำเสนอแบ่งได้เป็นสองวิธีด้วยกัน คือ การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบมีค่าคงที่ และการกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบที่มีค่าขึ้นกับความชันของค่าผิดพลาด ซึ่งแต่ละวิธีมีรายละเอียดดังต่อไปนี้

3.3.1 การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่

การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่เป็นการกำหนดรอบแต่ละครั้งแบบคงที่ ซึ่งจำนวนรอบที่ได้จะเป็นค่าเปอร์เซ็นต์ของจำนวนรอบสูงสุดของการฝึกอบรม โดยสามารถระบุได้จากรอบที่พิจารณาเลือกหยุดครั้งแรก รอบที่พิจารณาเลือกหยุดครั้งสุดท้าย จำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้ง และจำนวนรอบสูงสุดของการฝึกอบรม

วิธีการกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่นี้เริ่มจากการระบุรอบที่พิจารณาหยุดครั้งแรกและครั้งสุดท้าย จากนั้นพิจารณาค่าจำนวนข่ายงานที่เลือกหยุดแต่ละครั้งแล้วนำมาหารจากจำนวนข่ายงานทั้งหมดที่ฝึกอบรม ซึ่งจะได้จำนวนครั้งที่ต้องพิจารณาเลือกหยุด (No. of eliminations) ดังสมการที่ 3.2

$$noOfElim = noOfAllNW \ / \ noOfElimNWPerTime \tag{3.2}$$

เมื่อ *noOfAllNW* คือ จำนวนข่ายงานทั้งหมด

noOfElimNWPerTime คือ จำนวนข่ายงานที่เลือกหยุดแต่ละครั้ง

จากนั้นนำจำนวนครั้งที่ได้มาแบ่งช่วงจำนวนรอบการเลือกหยุด (Elimination period) ตั้งแต่รอบที่พิจารณาหยุดครั้งแรกถึงครั้งสุดท้าย ได้ดังสมการที่ 3.3

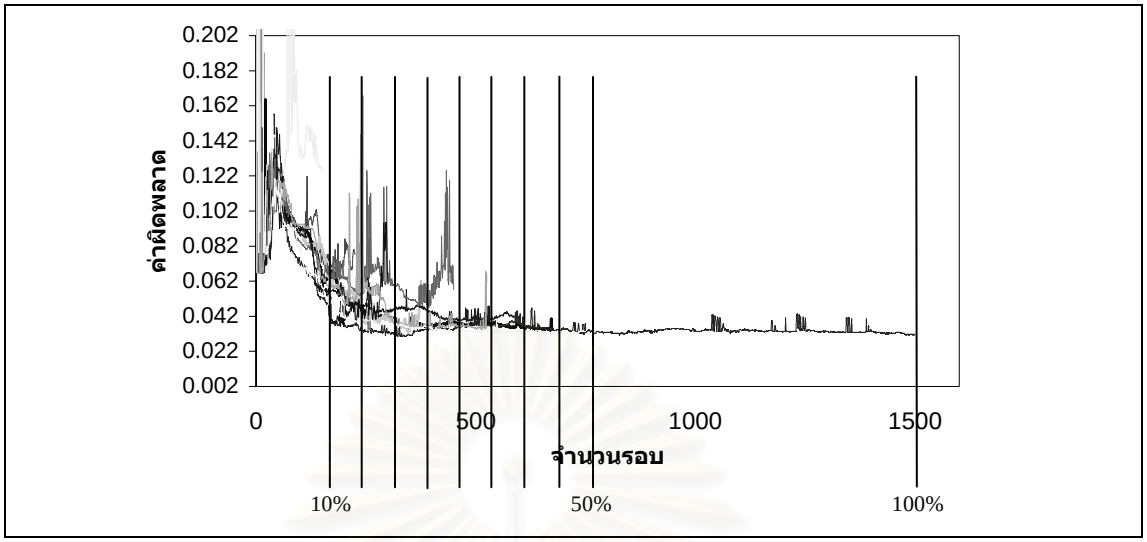
$$eliminationPeriod = (lastElimEpoch - firstElimEpoch) / noOfElim \tag{3.3}$$

เมื่อ *firstElimEpoch* คือ รอบที่พิจารณาหยุดครั้งแรก

lastElimEpoch คือ รอบที่พิจารณาหยุดครั้งสุดท้าย

ดังตัวอย่างในรูปที่ 3.3 ที่กำหนดให้รอบที่พิจารณาหยุดครั้งแรกเท่ากับรอบที่ 10% ของจำนวนรอบสูงสุด รอบที่พิจารณาหยุดครั้งสุดท้ายเท่ากับรอบที่ 50% ของจำนวนรอบสูงสุด และจำนวนรอบสูงสุดของการฝึกอบรมเท่ากับ 1500 รอบ โดยจำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้งเท่ากับ 1 ข่ายงาน จากจำนวนข่ายงานที่ฝึกอบรมทั้งหมด 10 ข่ายงาน ซึ่งผลการแบ่งจำนวนรอบที่ได้แสดงได้จากเส้นตรงแนวตั้งที่แบ่งรอบต่างๆ ในรูปดังกล่าว

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.3 ตัวอย่างการฝึกอบรมที่มีการพิจารณาเลือกหยุดแบบคงที่

ในการทดลองเรื่องการกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่ จะพิจารณาตัวแปรต่างๆ ที่กำหนดการแบ่งจำนวนรอบนี้ โดยจะทำการวิเคราะห์ความไวเพื่อศึกษาถึงความเหมาะสมในการกำหนดค่าตัวแปรแต่ละตัว

3.3.2 การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชัน

การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชันเป็นการกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแต่ละครั้งแบบไม่คงที่ โดยจะพิจารณาเลือกหยุดตามระดับความชันที่กำหนด เพื่อให้การเลือกหยุดสามารถทำได้อย่างเหมาะสมในกรณีที่แนวโน้มการลดลงของค่าผิดพลาดจากการฝึกอบรมมีลักษณะที่แตกต่างกันในแต่ละปัญหา ซึ่งวิธีการนี้จะยึดความชันของค่าผิดพลาดที่ได้เป็นหลักในการระบุรอบที่พิจารณา

ที่มาของการกำหนดจำนวนรอบที่พิจารณาเลือกหยุดด้วยวิธีการนี้มาจากการที่การฝึกอบรมข่ายงานประสาทเทียมในแต่ละปัญหาอาจได้แนวโน้มการลดลงของค่าผิดพลาดที่แตกต่างกัน เช่น ค่าผิดพลาดที่มีอัตราการลดลงที่เร็วในช่วงแรก แต่ช่วงหลังมีอัตราการลดลงที่ช้า หรือค่าผิดพลาดที่มีอัตราการลดลงที่ค่อนข้างคงที่ เป็นต้น ทำให้การกำหนดจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่นั้นอาจให้ผลลัพธ์ที่ไม่ดีได้ เนื่องจากหยุดเร็วหรือช้าเกินไป ดังนั้นการกำหนดด้วยวิธีการนี้น่าจะทำให้จำนวนรอบที่ใช้สอดคล้องกับการลดลงของค่าผิดพลาดในแต่ละปัญหาได้มากกว่า

งานวิจัยนี้จะศึกษาการพิจารณาเลือกหยุดโดยการกำหนดรอบแบบขึ้นกับความชันเปรียบเทียบกับวิธีการกำหนดจำนวนรอบแบบคงที่ โดยจำนวนข่ายงานที่พิจารณาเลือกหยุดในแต่ละครั้งจะกำหนดให้เป็นแบบคงที่ ซึ่งรายละเอียดและผลการทดลองจะนำเสนอต่อไปในหัวข้อที่ 5.4.1

3.4 การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุด

นอกเหนือจากค่าจำนวนรอบที่พิจารณาเลือกหยุดแล้ว ค่าตัวแปรอีกตัวหนึ่งที่มีความสำคัญต่อประสิทธิภาพในการฝึกอบรมข่ายงานประสาทเทียม คือ จำนวนข่ายงานที่พิจารณาเลือกหยุดในแต่ละครั้ง เนื่องจากการเลือกหยุดข่ายงานครั้งละมากๆ จะทำให้การเรียนรู้มีการใช้จำนวนรอบที่น้อยลงและใช้ทรัพยากรน้อยลง แต่ก็อาจส่งผลกระทบต่อคุณภาพของคำตอบที่ได้ ดังนั้นจึงต้องมีการวิเคราะห์หาค่าที่เหมาะสม หัวข้อนี้จะอธิบายเกี่ยวกับรูปแบบและวิธีการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุด

วิธีการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดที่นำเสนอนี้แบ่งได้เป็นสองวิธีด้วยกัน คือ การกำหนดจำนวนข่ายงานแบบมีค่าคงที่ และการกำหนดจำนวนข่ายงานแบบมีค่าเพิ่มขึ้นในแต่ละครั้งที่พิจารณาเลือกหยุด ซึ่งแต่ละวิธีมีรายละเอียดดังต่อไปนี้

3.4.1 การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่

การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่จะเลือกหยุดข่ายงานจำนวนคงที่เท่ากันทุกๆ ครั้ง ตัวอย่างเช่น เมื่อกำหนดให้จำนวนข่ายงานที่พิจารณาเลือกหยุดมีจำนวน 2 ข่ายงาน ในการฝึกอบรมนี้ เมื่อถึงรอบที่ต้องทำการพิจารณาเลือกหยุดข่ายงานแต่ละครั้ง จะเลือกหยุดข่ายงาน 2 ข่ายงาน ที่ให้ผลลัพธ์ที่แย่ที่สุดจากการเปรียบเทียบโดยอาศัยเกณฑ์ที่กำหนด จนกระทั่งเหลือข่ายงานสุดท้ายที่มีแนวโน้มที่ดีที่สุดข่ายงานเดียวที่จะฝึกอบรมจนครบจำนวนรอบสูงสุด โดยการทดลองในงานวิจัยนี้ได้พิจารณาจำนวนข่ายงานที่พิจารณาเลือกหยุด 5 จำนวนด้วยกัน คือ 1 2 3 4 และ 5 ข่ายงาน

จะสังเกตได้ว่า การกำหนดจำนวนข่ายงานจะมีความสอดคล้องกับจำนวนรอบที่พิจารณาเลือกหยุด โดยมีวิธีการในการกำหนดดังที่กล่าวมาแล้วในหัวข้อที่ 3.3.1 โดยจะได้วิเคราะห์ค่าผลลัพธ์ที่ได้จากการกำหนดค่าจำนวนข่ายงานต่างๆ และแสดงผลการทดลองต่อไปในบทที่ 5

3.4.2 การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบเพิ่มขึ้น

การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบเพิ่มขึ้นเลือกหยุดข่ายงานให้มีจำนวนเพิ่มขึ้นในแต่ละครั้งที่พิจารณา ซึ่งมาจากแนวคิดที่ว่า เมื่อจำนวนรอบการเรียนรู้ผ่านไป

มากขึ้นจะทำให้แนวโน้มของการค่าผิดพลาดมีความชัดเจนมากขึ้นด้วย ทำให้เกณฑ์ที่ใช้พิจารณาเลือกหยุดสามารถประเมินได้อย่างถูกต้อง ดังนั้นเพื่อลดเวลาในการฝึกอบรมให้มากยิ่งขึ้น การกำหนดให้พิจารณาเลือกหยุดจำนวนข่ายงานให้เพิ่มมากขึ้นน่าจะให้ผลลัพธ์ที่ดี

งานวิจัยนี้จะศึกษาการเลือกหยุดโดยกำหนดจำนวนข่ายงานแบบเพิ่มขึ้น เปรียบเทียบกับวิธีการกำหนดจำนวนข่ายงานแบบคงที่ โดยจำนวนรอบที่พิจารณาเลือกหยุดจะกำหนดให้เป็นแบบคงที่ ซึ่งรายละเอียดและผลการทดลองจะนำเสนอต่อไปในหัวข้อที่ 5.4.2

3.5 การกำหนดจำนวนรอบสูงสุด

ในการศึกษาลักษณะการเรียนรู้ของข่ายงานประสาทเทียมด้วยการกำหนดค่าตัวแปรต่างๆ เพื่อเปรียบเทียบประสิทธิภาพในการให้คำตอบของข่ายงานจากค่าผิดพลาดที่เกิดขึ้น สิ่งหนึ่งที่เป็นตัวแปรสำคัญในการฝึกอบรม คือ จำนวนรอบสูงสุดที่ใช้เรียนรู้ เนื่องจากค่าจำนวนรอบสูงสุดนี้ จะส่งผลต่อค่าผิดพลาดต่ำสุดที่ได้และทรัพยากรที่ใช้ ตัวอย่างเช่น เมื่อมีการกำหนดจำนวนรอบสูงสุดที่ใช้เรียนรู้น้อยเกินไปจะส่งผลให้ค่าผิดพลาดที่ได้มีโอกาสได้ค่าที่สูง และเมื่อมีการกำหนดจำนวนรอบที่มากเกินไปจะส่งผลให้ใช้เวลาและทรัพยากรในการคำนวณที่เกินความจำเป็น ดังนั้นจึงควรกำหนดค่านี้ให้เหมาะสมกับความต้องการ

การกำหนดจำนวนรอบสูงสุดโดยทั่วไปแล้วจะกำหนดให้มีค่าเท่ากับจำนวนรอบที่ทำให้ค่าผิดพลาดที่ได้เกิดการลู่เข้า (Converge) (ค่าผิดพลาดมีการเปลี่ยนแปลงน้อยกว่า ε) หรือค่าผิดพลาดเริ่มมีค่าสูงขึ้นในกรณีปัญหาที่เกิดความพอดีกับตัวอย่างมากเกินไป (Overfitting) ซึ่งการกำหนดจำนวนรอบสูงสุดในงานวิจัยนี้ได้กำหนดขึ้นโดยการวิเคราะห์ค่าความชันและจำนวนรอบที่ใช้วัดการลู่เข้าของค่าผิดพลาดที่เกิดขึ้น กล่าวคือ รอบสูงสุดสำหรับการฝึกอบรมข่ายงานประสาทเทียมเท่ากับรอบที่ค่าผิดพลาดมีค่าประมาณความชันน้อยกว่า ε มีจำนวนครั้งติดต่อกันเท่ากับจำนวนรอบที่ใช้วัดการลู่เข้า ดังนั้น ค่าทั้งสองที่ใช้ คือ ค่าความชันและจำนวนที่ใช้วัดการลู่เข้ามีส่วนสำคัญอย่างยิ่งที่ส่งผลต่อจำนวนรอบสูงสุดที่ได้ ซึ่งการทดลองในงานวิจัยนี้จะได้ศึกษาถึงผลกระทบจากการกำหนดค่าตัวแปรทั้งสองนี้พร้อมทั้งให้หลักเกณฑ์ในการกำหนดค่าตัวแปรเพื่อประโยชน์ในการนำไปใช้งานต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

การทดลอง

งานวิจัยนี้ได้ทดลองนำวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดดังที่กล่าวมาแล้ว มาทดสอบกับชุดข้อมูลทดสอบเพื่อประเมินประสิทธิภาพของวิธีการที่นำเสนอ ซึ่งในบทนี้จะกล่าวถึง ข้อมูลที่ใช้ในการฝึกอบรมและการทดสอบ การกำหนดรูปแบบของข่ายงานประสาทเทียมที่ใช้ในแต่ละปัญหา และขั้นตอนวิธีการทดลอง ส่วนรายละเอียดและการวิเคราะห์ผลการทดลองจะนำเสนอต่อไปในบทที่ 5

4.1 ข้อมูลที่ใช้ในการฝึกอบรมและการทดสอบ

ข้อมูลที่ใช้ในการทดสอบวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดเป็นชุดข้อมูลในกลุ่มปัญหา Proben1 [24] ซึ่งชุดข้อมูลที่นำมาใช้แบ่งออกเป็น 2 ประเภท คือ ชุดข้อมูลปัญหาการประมาณค่า (Approximation) และชุดข้อมูลปัญหาการจัดแบ่งประเภท (Classification) โดยมีทั้งสิ้น 8 ปัญหาด้วยกัน ประกอบด้วย

1. ปัญหา Building เป็นปัญหาการประมาณค่าหรือการพยากรณ์การใช้พลังงานในอาคารสิ่งปลูกสร้าง โดยพยายามที่จะคาดคะเนการใช้ไฟฟ้า น้ำร้อน และน้ำเย็น จากข้อมูลวัน เวลา อุณหภูมิ ความชื้น แสงแดด และความเร็วลม

2. ปัญหา Flare เป็นปัญหาการประมาณค่าจำนวนการระเบิดที่เกิดขึ้นบนพื้นผิวดวงอาทิตย์ในระดับขนาดต่างๆ ที่เกิดขึ้นใน 24 ชั่วโมงถัดไป จากข้อมูลที่ผ่านมาของการเกิดและประวัติพื้นผิวที่เกิดนั้น

3. ปัญหา Heart เป็นปัญหาการประมาณค่าของการวินิจฉัยโอกาสการเกิดโรคหัวใจ โดยประเมินจากข้อมูลพื้นฐานของแต่ละบุคคล อาทิ อายุ เพศ นิสัยการสูบบุหรี่ รวมทั้งผลการทดสอบทางการแพทย์ต่างๆ

4. ปัญหา Card เป็นปัญหาการจัดแบ่งประเภทเพื่อคาดคะเนการอนุมัติเครดิตการ์ดให้กับลูกค้าของธนาคารว่าสมควรได้รับหรือไม่ ซึ่งข้อมูลขาเข้าในการฝึกอบรมของข่ายงาน คือ คุณลักษณะของลูกค้าแต่ละคนที่น่าจะเป็นปัจจัยในการอนุมัตินี้

5. ปัญหา Mushroom เป็นปัญหาการจัดแบ่งประเภทเพื่อแยกแยะประเภทของเห็ดมีพิษว่ารับประทานได้หรือไม่ ซึ่งลักษณะที่นำมาพิจารณาจะดูจากรูปร่าง สี กลิ่น และถิ่นที่อยู่ของเห็ดนั้นๆ

6. ปัญหา Gene เป็นปัญหาการจัดแบ่งประเภทของการตรวจสอบโครงสร้างของยีนในสายลำดับนิวคลีโอไทด์ (Nucleotide) ถึงคุณสมบัติการเป็นตัวรับ ตัวให้ หรือไม่ได้เป็นทั้งตัวรับและตัวให้ เพื่อใช้ในการศึกษาหารูปแบบในการนำไปใช้งานต่อไป

7. ปัญหา Soybean เป็นปัญหาการจัดแบ่งประเภทที่แยกความแตกต่างของโรคที่เกิดขึ้นกับถั่วเหลืองทั้ง 19 โรค โดยข้อมูลลักษณะที่นำมาใช้ในการศึกษาประกอบด้วย ลักษณะของถั่ว ลักษณะของต้นพืช และข้อมูลในอดีต

8. ปัญหา Thyroid ที่เป็นปัญหาการจัดแบ่งประเภทของการวินิจฉัยรูปแบบโรคไทรอยด์ที่เกิดขึ้นกับผู้ป่วย โดยพิจารณาข้อมูลจากทั้งการสอบถามและการตรวจคนไข้ เพื่อที่จะแยกรูปแบบของการทำงานของต่อมไทรอยด์ว่าเป็นแบบทำงานมากกว่าปกติ แบบทำงานปกติ หรือแบบทำงานต่ำกว่าปกติ

ในการทดสอบวิธีการเรียนรู้จะทำการทดลองกับชุดข้อมูลของปัญหาต่างๆ เหล่านี้ ซึ่งได้แบ่งข้อมูลที่ใช้เป็น 2 ส่วน คือ ข้อมูลสอนและข้อมูลประเมิน โดยจำนวนข้อมูลที่ใช้สอนจะมีจำนวนเป็นสองเท่าของข้อมูลประเมิน และจำนวนข้อมูลที่ใช้ในทั้งสองส่วนนี้จะมีจำนวนแตกต่างกันในแต่ละปัญหา ซึ่งมีรายละเอียดดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 จำนวนข้อมูลฝึกอบรมและทดสอบที่ใช้ในแต่ละปัญหา

ปัญหา	จำนวนข้อมูลสอน	จำนวนข้อมูลประเมิน
ปัญหาการประมาณค่า		
Building	2104	1052
Flare	533	267
Heart	460	230
ปัญหาการจัดแบ่งประเภท		
Card	345	173
Mushroom	4062	2031
Gene	1588	794
Soybean	342	171
Thyroid	3600	1800

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.2 การกำหนดรูปแบบของข่ายงานประสาทเทียม

ข่ายงานประสาทเทียมที่ใช้ในการทดลองนี้กำหนดให้เป็นข่ายงานประสาทเทียมป้อนไปหน้าแบบหลายชั้น ซึ่งรูปแบบนี้เป็นรูปแบบที่นิยมใช้ในการแก้ไขปัญหาทั่วไป โดยการใช้งานข่ายงานประสาทเทียมในปัญหาใดปัญหาหนึ่งจำเป็นต้องมีการกำหนดตัวแปรต่างๆ ที่เป็นการกำหนดรูปแบบโครงสร้างของข่ายงานให้เหมาะสมกับปัญหานั้นเพื่อให้การเรียนรู้มีประสิทธิภาพสูงสุด ซึ่งในการทดลองนี้ได้กำหนดค่าต่างๆ ให้เหมาะสมตามค่าที่แนะนำในงานวิจัย [24] ดังตารางที่ 4.2

ตารางที่ 4.2 รายละเอียดโครงสร้างของข่ายงานที่ใช้ในแต่ละปัญหา

ปัญหา	จำนวนเซลล์ประสาทในข่ายงาน			จำนวนรอบสูงสุด
	นำเข้า	ชั้นซ่อน	นำออก	
ปัญหาการประมาณค่า				
Building	14	16	3	411
Flare	24	32	3	253
Heart	35	32	1	244
ปัญหาการจัดแบ่งประเภท				
Card	51	32	2	239
Mushroom	125	32	2	1619
Gene	120	4+2	3	711
Soybean	82	16+8	19	926
Thyroid	21	16+8	3	842

โดยจำนวนเซลล์ประสาทหรือจำนวนบัพที่ใช้ในแต่ละชั้นของข่ายงานประสาทเทียมที่ประกอบไปด้วย ชั้นนำเข้า ชั้นซ่อน และชั้นนำออก ซึ่งในชั้นซ่อนสามารถมีจำนวนชั้นได้มากกว่าหนึ่งชั้น ซึ่งในกรณีที่มีตั้งแต่สองชั้นขึ้นไป ค่าจำนวนบัพที่แสดงแต่ละชั้นจะคั่นด้วยเครื่องหมายบวก (+) โดยเรียงจากชั้นที่อยู่ติดกับชั้นนำเข้าไปยังชั้นนำออก และการกำหนดจำนวนรอบสูงสุดของการฝึกอบรมที่ได้แสดงไว้ ค่าจำนวนรอบสูงสุดนี้มาจากการทดลองหาค่าที่เหมาะสมกับปัญหานั้นๆ ซึ่งสามารถศึกษารายละเอียดได้จากการศึกษาผลกระทบของการกำหนดจำนวนรอบสูงสุดในหัวข้อที่ 5.4

4.3 ขั้นตอนการทดลอง

1. นำข่ายงานประสาทเทียมที่ใช้วิธีการฝึกอบรมแบบปกติมาทดสอบกับชุดข้อมูลปัญหาต่างๆ ทั้ง 8 ปัญหา เพื่อศึกษาคุณภาพของคำตอบที่ได้และเวลาที่ใช้ในการเรียนรู้

2. นำข่ายงานประสาทเทียมที่ใช้วิธีการฝึกอบรมแบบที่ได้นำเสนอมาทดสอบกับชุดข้อมูลปัญหา โดยกำหนดค่าจำนวนรอบและจำนวนข่ายงานที่พิจารณาหยุดแบบคงที่ที่ค่าต่างๆ เพื่อทำการวิเคราะห์ความไวของตัวแปรแต่ละตัว

3. นำข่ายงานประสาทเทียมที่ใช้วิธีการฝึกอบรมแบบที่ได้นำเสนอมาทดสอบกับชุดข้อมูลปัญหา โดยใช้วิธีการกำหนดค่าจำนวนรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชันและวิธีการกำหนดค่าจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบเพิ่มขึ้น

4. เปรียบเทียบคุณภาพของคำตอบระหว่างวิธีการฝึกอบรบแบบปกติและแบบที่ได้นำเสนอ จากผลลัพธ์ของการทดสอบด้วยวิธีต่างๆ

5. เปรียบเทียบเวลาที่ใช้ระหว่างวิธีการฝึกอบรบแบบปกติและแบบที่ได้นำเสนอ จากผลลัพธ์ของการทดสอบด้วยวิธีต่างๆ

6. เปรียบเทียบประสิทธิภาพของวิธีการฝึกอบรบแบบที่ได้นำเสนอระหว่างวิธีการกำหนดค่าจำนวนรอบที่พิจารณาเลือกหยุดแบบคงที่และแบบขึ้นกับความชัน

7. เปรียบเทียบประสิทธิภาพของวิธีการฝึกอบรบแบบที่ได้นำเสนอระหว่างวิธีการกำหนดค่าจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่และแบบเพิ่มขึ้น

8. ศึกษาถึงผลกระทบของการกำหนดจำนวนรอบสูงสุด โดยการพิจารณาจำนวนรอบที่ใช้ในการประมาณค่าความชันและจำนวนรอบที่ใช้วัดการลู่เข้าของค่าผิดพลาดที่เกิดขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

ผลการทดลองและวิเคราะห์ผล

งานวิจัยนี้ได้ทดลองนำวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดข่ายงานที่มีแนวโน้มของผลลัพธ์ไม่ดีมาทดสอบกับชุดข้อมูลในกลุ่มปัญหา Proben1 [24] ผลการทดลองที่ได้แสดงการเปรียบเทียบระหว่างวิธีการฝึกอบรมแบบต่างๆ ทั้งในแง่ของคุณภาพของคำตอบและเวลาที่ใช้ในการฝึกอบรม โดยได้วิเคราะห์เปรียบเทียบถึงผลกระทบของการกำหนดจำนวนรอบที่พิจารณาเลือกหยุด จำนวนข่ายงานที่พิจารณาเลือกหยุด และจำนวนรอบสูงสุด ซึ่งผลการทดลองทั้งหมดมีรายละเอียดดังต่อไปนี้

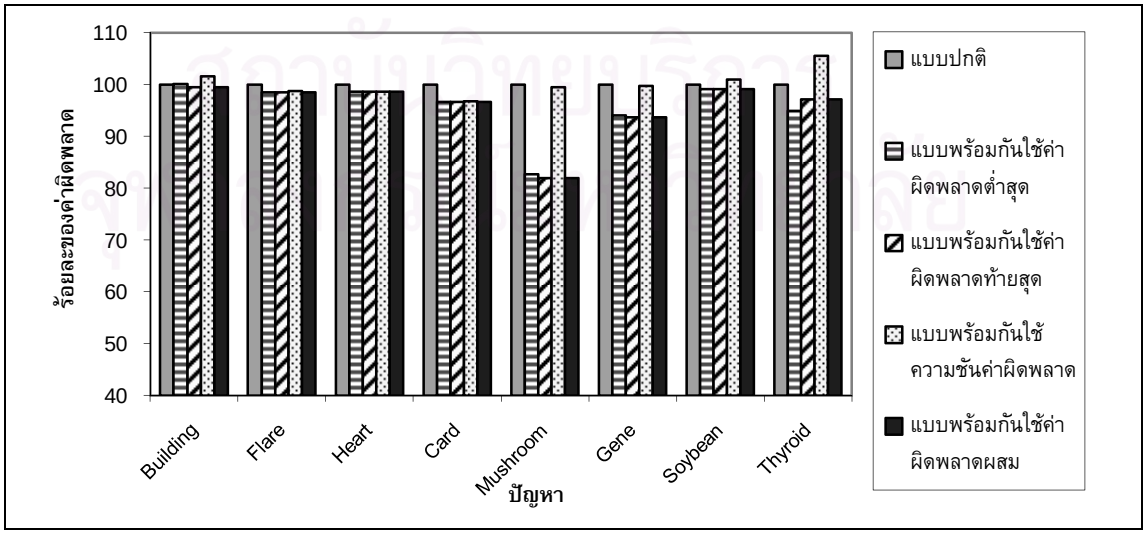
5.1 การเปรียบเทียบคุณภาพคำตอบระหว่างวิธีการฝึกอบรมแบบต่าง ๆ

ในการเปรียบเทียบคุณภาพของคำตอบที่ได้จากการฝึกอบรมข่ายงานประสาทเทียมด้วยวิธีต่างๆ นี้ ได้ใช้วิธีการกำหนดจำนวนรอบทั้งหมดที่ใช้เรียนรู้ของแต่ละวิธีให้มีค่าเท่ากัน ซึ่งค่าจำนวนรอบรวมทั้งหมดนี้จะมีค่าเท่ากับจำนวนรอบทั้งหมดที่ใช้ในวิธีการฝึกอบรมแบบพร้อมกัน ดังนั้น วิธีการฝึกอบรมแบบปกติจะฝึกอบรมข่ายงานประสาทเทียมทีละข่ายงานจนครบจำนวนรอบสูงสุด และจะเริ่มต้นฝึกอบรมข่ายงานใหม่จนกระทั่งจำนวนรอบรวมทั้งหมดเท่ากับค่าจำนวนรอบนี้ ตัวอย่างเช่น เมื่อจำนวนรอบสูงสุดมีค่าเท่ากับ 100 รอบ จำนวนรอบรวมทั้งหมดที่ใช้ในวิธีการฝึกอบรมแบบพร้อมกันจะเท่ากับ 460 รอบ วิธีการฝึกอบรมแบบปกติจึงสามารถเริ่มต้นฝึกอบรมข่ายงานใหม่ได้ทั้งสิ้น 5 ข่ายงาน โดยข่ายงานสุดท้ายจะฝึกอบรมได้เพียง 60 รอบเท่านั้น เป็นต้น ซึ่งค่าคำตอบที่จะนำมาเปรียบเทียบกัน คือ ค่าผิดพลาดที่มีค่าต่ำที่สุดจากการทดลองในครั้งนั้น ดังแสดงในตารางที่ 5.1 โดยค่าผิดพลาดที่นำมาเปรียบเทียบในตารางนี้มาจากการเฉลี่ยค่าที่ได้จากการทดลองทั้งสิ้น 20 ครั้ง และภายในวงเล็บได้แสดงค่าเปอร์เซ็นต์ที่ดีขึ้นของคำตอบจากวิธีฝึกอบรมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดเมื่อเทียบกับวิธีปกติ ซึ่งค่าคำตอบของวิธีการที่ให้ผลลัพธ์ดีที่สุดในแต่ละปัญหาจะถูกแสดงในตารางด้วยค่าที่ขีดเส้นใต้

ในรูปที่ 5.1 ได้นำค่าผิดพลาดที่ใช้เปรียบเทียบคุณภาพของคำตอบในตารางที่ 5.1 มาแสดงในรูปแบบแผนภูมิเพื่อวัดประสิทธิภาพของการฝึกอบรมแบบพร้อมกันที่ใช้ค่าเปรียบเทียบแบบต่างๆ โดยนำมาเทียบกับการฝึกอบรมแบบปกติที่มีการปรับค่าผิดพลาดให้เป็นเกณฑ์ที่ 100 เปอร์เซ็นต์ ดังนั้น ในกรณีที่วิธีการฝึกอบรมแบบพร้อมกันแบบหนึ่งให้ค่าผิดพลาดน้อยกว่า 100 เปอร์เซ็นต์ แสดงว่าวิธีการฝึกอบรมแบบนั้นสามารถให้คำตอบที่ดีกว่าวิธีการฝึกอบรมแบบปกติ เช่นเดียวกันในทางตรงข้าม ถ้าให้ค่าผิดพลาดมากกว่า 100 เปอร์เซ็นต์ แสดงว่าวิธีการฝึกอบรมแบบปกติจะให้คำตอบที่ดีกว่าวิธีการฝึกอบรมแบบนั้น

ตารางที่ 5.1 ผลการเปรียบเทียบคุณภาพของคำตอบ (ค่าผิดพลาด) ของวิธีการแบบต่าง ๆ

ปัญหา	การฝึกอบรมแบบปกติ	การฝึกอบรมแบบพร้อมกัน			
		ค่าต่ำสุด	ค่าท้ายสุด	ความชัน	ค่าผสม
ปัญหาการประมาณค่า					
Building	0.008050	0.008058	0.008013	0.008180	0.008013
		(-0.097%)	(0.463%)	(-1.614%)	(0.463%)
Flare	0.003841	0.003784	0.003784	0.003793	0.003784
		(1.478%)	(1.478%)	(1.232%)	(1.478%)
Heart	0.042731	0.042163	0.042163	0.042163	0.042163
		(1.329%)	(1.329%)	(1.329%)	(1.329%)
ปัญหาการจัดแบ่งประเภท					
Card	0.395954	0.382659	0.382659	0.383237	0.382659
		(3.358%)	(3.358%)	(3.212%)	(3.358%)
Mushroom	0.010364	0.008567	0.008493	0.010315	0.008493
		(17.339%)	(18.052%)	(0.475%)	(18.052%)
Gene	0.180416	0.169647	0.169018	0.179975	0.169018
		(5.969%)	(6.318%)	(0.244%)	(6.318%)
Soybean	0.438304	0.434210	0.434503	0.442690	0.434211
		(0.934%)	(0.867%)	(-1.001%)	(0.934%)
Thyroid	0.028389	0.026944	0.027583	0.029972	0.027583
		(5.089%)	(2.838%)	(-5.577%)	(2.838%)



รูปที่ 5.1 แผนภูมิเปรียบเทียบคุณภาพของคำตอบของวิธีการฝึกอบรมแบบต่าง ๆ

ผลที่ได้จากการเปรียบเทียบคุณภาพของคำตอบนี้แสดงให้เห็นว่า วิธีการฝึกอบรมแบบพร้อมกันให้ผลคำตอบที่ดีกว่าวิธีการฝึกอบรมแบบปกติในทุกชุดข้อมูลปัญหา แต่ไม่มีหลักเกณฑ์การเลือกหยุดใดที่สามารถให้คำตอบที่ดีที่สุดได้สำหรับทุกปัญหา ค่าผิดพลาดต่ำสุด ค่าผิดพลาดท้ายสุด และค่าผิดพลาดผสม เป็นค่าที่ใช้เปรียบเทียบได้อย่างมีประสิทธิภาพในปัญหาต่าง ๆ โดยเฉพาะอย่างยิ่งในการใช้หลักเกณฑ์ค่าผิดพลาดผสมที่สามารถให้ผลลัพธ์ที่ดีที่สุด 7 ใน 8 ปัญหาที่ทดสอบ ในขณะที่ค่าความชันของค่าผิดพลาดเป็นหลักเกณฑ์ที่ให้คำตอบที่ดีที่สุดได้เพียงปัญหาเดียวเท่านั้น อย่างไรก็ตามค่าเฉลี่ยของค่าผิดพลาดที่ได้จากการใช้วิธีการที่นำเสนอนี้สามารถช่วยให้ได้คำตอบที่ดีขึ้นจากวิธีการแบบปกติสูงสุดถึง 18.05% ในปัญหา Mushroom ซึ่งเป็นปัญหาที่มีการลู่เข้าของค่าผิดพลาดสุดท้ายที่แตกต่างกันมากระหว่างแต่ละข่ายงานประสาทเทียม ซึ่งแปลว่าอาจเกิดปัญหาค่าต่ำสุดเฉพาะที่ ทำให้การใช้วิธีการฝึกอบรมนี้ให้ผลลัพธ์ที่ดีขึ้นกว่าวิธีการปกติมากยิ่งขึ้น

5.2 การเปรียบเทียบเวลาที่ใช้ระหว่างวิธีการฝึกอบรมแบบต่าง ๆ

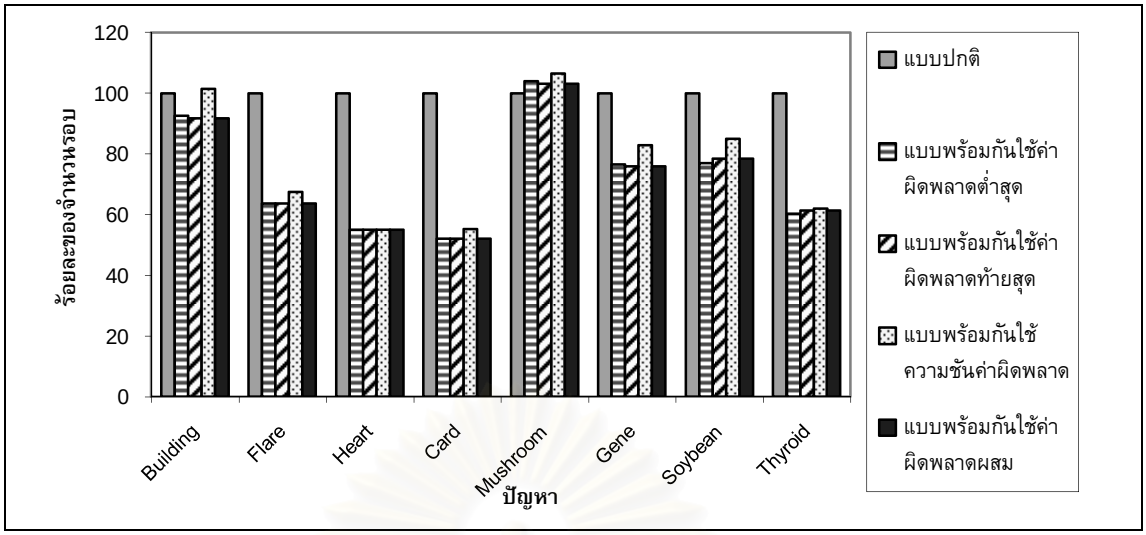
ในการเปรียบเทียบเวลาหรือจำนวนรอบที่ใช้ในการฝึกอบรมข่ายงานประสาทเทียมด้วยวิธีต่าง ๆ นี้ ได้ใช้วิธีการนับจำนวนรอบทั้งหมดในการเรียนรู้ของแต่ละวิธีที่ใช้ฝึกอบรมจนกระทั่งได้คำตอบที่มีค่าต่ำกว่าค่าที่กำหนด โดยในการฝึกอบรมแบบปกติจะนับจำนวนรอบเริ่มจากฝึกอบรมข่ายงานที่หนึ่งไปจนกระทั่งได้คำตอบที่ต่ำกว่าค่าผิดพลาดที่กำหนด ซึ่งเมื่อครบจำนวนสูงสุดแล้วยังไม่ได้ค่าที่ต่ำกว่า จะนับจำนวนรอบต่อจากการฝึกอบรมข่ายงานใหม่จนกระทั่งได้คำตอบที่ต่ำกว่าค่าที่กำหนด และนำจำนวนรอบรวมที่นับได้มาพิจารณาเปรียบเทียบ โดยกำหนดให้การทดลองแต่ละครั้งฝึกอบรมไม่เกิน 10 ข่ายงาน สำหรับวิธีการฝึกอบรมแบบพร้อมกันจะนับจำนวนรอบทั้งหมดจากทั้ง 10 ข่ายงานที่เรียนรู้พร้อมกันที่ใช้หลักเกณฑ์การพิจารณาเลือกหยุดข่ายงานแบบต่าง ๆ ในการฝึกอบรมจนได้คำตอบที่ต่ำกว่าค่าผิดพลาดที่กำหนด และนำค่าจำนวนรอบที่ได้มาเปรียบเทียบดังแสดงในตารางที่ 5.2 โดยค่าที่แสดงนี้มาจากการเฉลี่ยค่าที่ได้จากการทดลองทั้งสิ้น 20 ครั้ง และภายในวงเล็บได้แสดงค่าเปอร์เซ็นต์ที่ลดลงของจำนวนรอบเมื่อเทียบกับวิธีปกติ ซึ่งจำนวนรอบของวิธีการที่ให้ค่าผิดพลาดน้อยที่สุดในแต่ละปัญหาจะถูกแสดงในตารางด้วยค่าที่ขีดเส้นใต้

สำหรับค่าผิดพลาดที่ใช้เป็นเกณฑ์ในการวัดจำนวนรอบที่ใช้เรียนรู้ เพื่อให้ได้ผลลัพธ์ที่มีค่าต่ำกว่าค่าผิดพลาดนี้ของแต่ละปัญหา จะกำหนดโดยการฝึกอบรมจนครบจำนวนรอบสูงสุด และหาค่าผิดพลาดที่ต่ำที่สุดในแต่ละครั้งมาเรียงลำดับ โดยจะเลือกค่าที่ต่ำที่สุดอันดับที่ 20 จาก 100 หรือเกณฑ์ที่ใช้ค่านี้จะช่วยจำแนกการฝึกอบรมข่ายงานที่ให้ค่าผิดพลาดที่ดีที่อยู่ในระดับร้อยละ 80 ขึ้นไป โดยค่าที่ได้นี้จากทดลองในแต่ละปัญหาจะถูกนำมาใช้เป็นเกณฑ์ในการทดลอง

ตารางที่ 5.2 ผลการเปรียบเทียบเวลาที่ใช้ (จำนวนรอบ) ของวิธีการแบบต่างๆ

ปัญหา	การฝึกอบรมแบบปกติ	การฝึกอบรมแบบพร้อมกัน			
		ค่าต่ำสุด	ค่าท้ายสุด	ความชัน	ค่าผสม
ปัญหาการประมาณค่า					
Building	1707.50	1580.05 (7.46%)	1566.45 (8.26%)	1730.90 (-1.37%)	1566.45 (8.26%)
Flare	670.40	426.90 (36.32%)	426.90 (36.32%)	452.45 (32.51%)	426.90 (36.32%)
Heart	553.20	304.15 (45.02%)	304.15 (45.02%)	304.15 (45.02%)	304.15 (45.02%)
ปัญหาการจัดแบ่งประเภท					
Card	926.35	482.30 (47.94%)	482.30 (47.94%)	511.15 (44.82%)	482.30 (47.94%)
Mushroom	4533.25	4713.45 (-3.98%)	4677.75 (-3.19%)	4827.00 (-6.48%)	4677.75 (-3.19%)
Gene	2819.65	2158.75 (23.44%)	2142.50 (24.02%)	2339.20 (17.04%)	2142.50 (24.02%)
Soybean	4176.50	3211.25 (23.11%)	3278.55 (21.50%)	3554.25 (14.90%)	3278.60 (21.50%)
Thyroid	6179.20	3728.95 (39.65%)	3792.15 (38.63%)	3828.70 (38.04%)	3792.15 (38.63%)

เมื่อนำค่าผิดพลาดที่ใช้เปรียบเทียบเวลาในการฝึกอบรมในตารางที่ 5.2 มาแสดงในรูปแบบแผนภูมิเพื่อวัดประสิทธิภาพของการใช้วิธีการฝึกอบรมแบบพร้อมกันเมื่อเปรียบเทียบกับวิธีการฝึกอบรมแบบปกติจะแสดงได้ดังรูปที่ 5.2 ในกรณีที่วิธีการฝึกอบรมแบบพร้อมกันแบบหนึ่งให้เปอร์เซ็นต์ของจำนวนรอบการฝึกอบรมน้อยกว่า 100 เปอร์เซ็นต์ แสดงว่า วิธีการฝึกอบรมแบบนั้นใช้เวลาในการฝึกอบรมที่น้อยกว่าวิธีการฝึกอบรมแบบปกติเพื่อให้ได้คำตอบที่ดีเท่ากัน



รูปที่ 5.2 แผนภูมิเปรียบเทียบเวลาที่ใช้ของวิธีการฝึกอบรมแบบต่างๆ

ผลที่ได้จากการเปรียบเทียบเวลาที่ใช้ในการฝึกอบรมนี้แสดงให้เห็นว่า ในปัญหาส่วนใหญ่วิธีการฝึกอบรมแบบพร้อมกันใช้เวลาน้อยกว่าวิธีการฝึกอบรมแบบปกติ ซึ่งอาจช่วยลดเวลาการคำนวณได้สูงสุดถึง 47.94% โดยค่าผิดพลาดต่ำสุด ค่าผิดพลาดท้ายสุด และค่าผิดพลาดผสม ทั้งสามค่าที่ใช้เปรียบเทียบนี้สามารถเพิ่มประสิทธิภาพได้ใกล้เคียงกัน

5.3 ผลกระทบของการกำหนดจำนวนรอบและข่ายงานที่พิจารณาเลือกหยุดแบบคงที่

การกำหนดจำนวนรอบและข่ายงานที่พิจารณาเลือกหยุดนั้น ประกอบด้วยตัวแปรที่เกี่ยวข้องหลายตัวด้วยกัน คือ รอบที่พิจารณาเลือกหยุดครั้งแรก รอบที่พิจารณาเลือกหยุดครั้งสุดท้าย และจำนวนข่ายงานที่เลือกหยุดในแต่ละครั้ง ซึ่งค่าของตัวแปรแต่ละตัวนี้ได้ส่งผลกระทบต่อประสิทธิภาพของการเรียนรู้ ทำให้ในการกำหนดค่าตัวแปรเหล่านี้จำเป็นต้องเลือกใช้ค่าที่เหมาะสมเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด งานวิจัยนี้จึงได้ศึกษาถึงผลกระทบของตัวแปรแต่ละตัวเพื่อให้เกิดแนวทางการกำหนดค่าของตัวแปรในการนำไปใช้งานต่อไป

การกำหนดจำนวนรอบและจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบมีค่าคงที่นั้นทำได้โดยปรับค่าตัวแปรรอบที่พิจารณาเลือกหยุดครั้งแรกและครั้งสุดท้าย รวมถึงจำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้งให้มีค่าต่างๆ ซึ่งตัวแปรแต่ละตัวจะมีความเกี่ยวข้องในการฝึกอบรมที่แตกต่างกัน โดยค่าจำนวนรอบทั้งสองจะบอกตำแหน่งรอบแรกและรอบสุดท้ายที่พิจารณาเลือกหยุดตามลำดับ ตัวอย่างเช่น ค่าที่กำหนดเป็น 10-50 แสดงว่า รอบแรกของการพิจารณาเลือกหยุด คือ รอบที่ 10% ของจำนวนรอบสูงสุด และรอบสุดท้าย คือ รอบที่ 50% ของจำนวนรอบสูงสุด และมีจำนวนข่ายงานที่เลือกหยุดแต่ละครั้งเป็นค่าที่ใช้แบ่งช่วงการพิจารณาเลือกหยุด ซึ่งในการแบ่งจะให้มีระยะห่างจำนวนรอบในการพิจารณาเลือกหยุดแต่ละครั้งเท่าๆ

กัน โดยในการทดลองนี้กำหนดให้มีข่ายงานที่เรียนรู้พร้อมกันทั้งหมด 10 ข่ายงาน และมีการนำค่าผิดพลาดต่ำสุดที่ได้ในแต่ละครั้งจากการใช้ค่าผิดพลาดผสมเป็นตัวแปรเปรียบเทียบ

ในการวิเคราะห์ความไวของตัวแปรต่างๆ ที่กำหนด จะทำการทดลองโดยกำหนดให้เฉพาะค่าตัวแปรใดค่าหนึ่งที่มีค่าเปลี่ยนแปลง ส่วนค่าตัวแปรที่เหลือมีค่าคงที่ ซึ่งในที่นี้ค่าคงที่ที่กำหนดให้เป็นค่าพื้นฐานของแต่ละตัวแปรมีค่าแสดงดังตารางที่ 5.3

ตารางที่ 5.3 การกำหนดค่าพื้นฐานของตัวแปร

ตัวแปร	ค่าพื้นฐาน
รอบที่พิจารณาเลือกหยุดครั้งแรก	20
รอบที่พิจารณาเลือกหยุดครั้งสุดท้าย	60
จำนวนข่ายงานที่เลือกหยุดแต่ละครั้ง	1

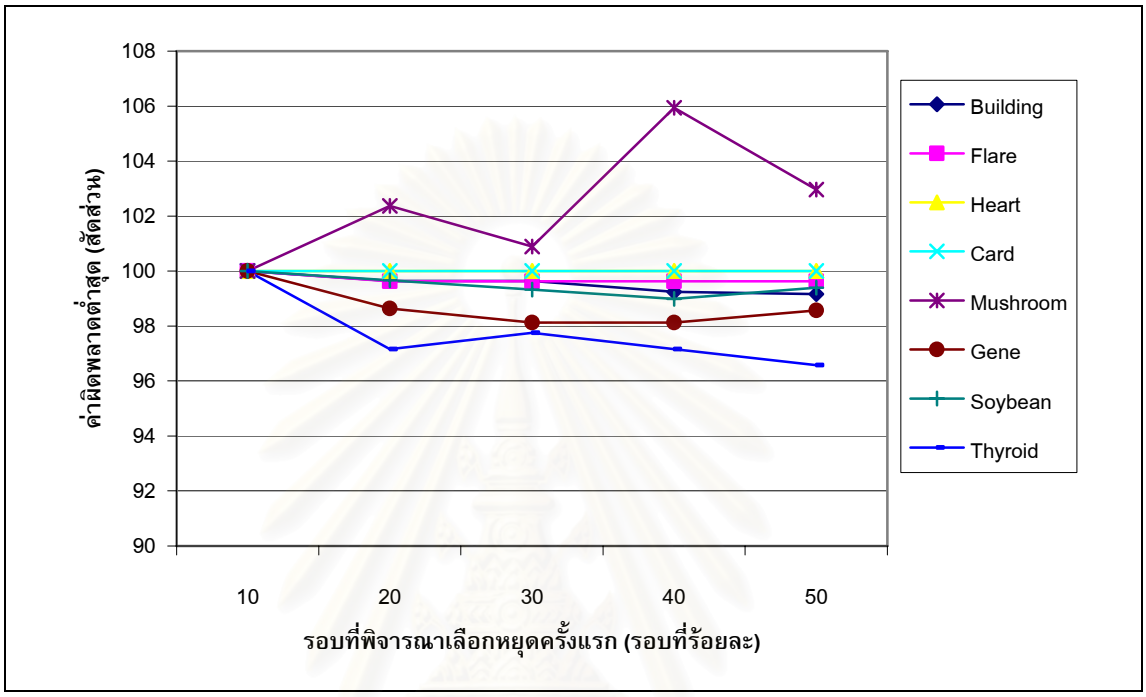
5.3.1 การวิเคราะห์ความไวของรอบที่พิจารณาเลือกหยุดครั้งแรก

เมื่อกำหนดรอบที่พิจารณาเลือกหยุดครั้งแรกให้มีค่าต่างๆ ผลลัพธ์ค่าผิดพลาดต่ำสุดจากแต่ละปัญหามีค่าดังตารางที่ 5.4

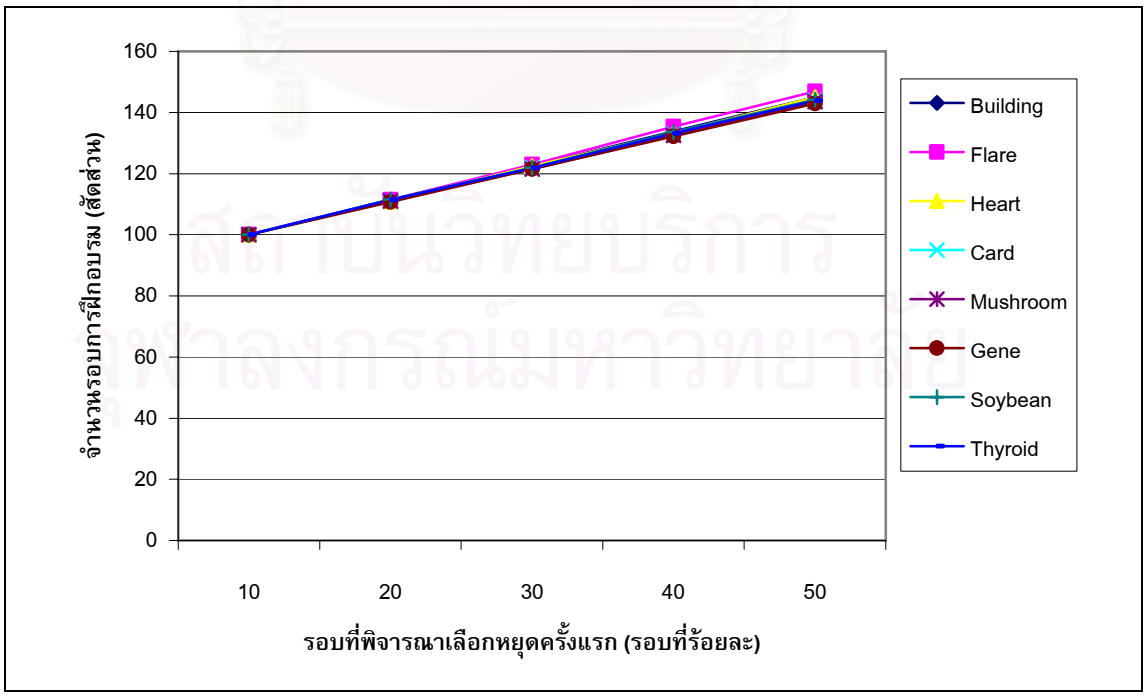
ตารางที่ 5.4 ค่าผิดพลาดจากการกำหนดรอบที่พิจารณาเลือกหยุดครั้งแรก

ปัญหา	ค่าผิดพลาดต่ำสุดของการเรียนรู้				
	รอบที่ ร้อยละ 10	รอบที่ ร้อยละ 20	รอบที่ ร้อยละ 30	รอบที่ ร้อยละ 40	รอบที่ ร้อยละ 50
ปัญหาประมาณค่า					
Building	0.008087	0.008058	0.008058	0.008026	0.008019
Flare	0.003798	0.003784	0.003784	0.003784	0.003784
Heart	0.042163	0.042163	0.042163	0.042163	0.042163
ปัญหาจัดแบ่งประเภท					
Card	0.382659	0.382659	0.382659	0.382659	0.382659
Mushroom	0.008296	0.008493	0.008370	0.008789	0.008543
Gene	0.171347	0.169018	0.168136	0.168136	0.168892
Soybean	0.435672	0.434211	0.432748	0.431286	0.433041
Thyroid	0.028389	0.027583	0.027750	0.027583	0.027417

จากตารางที่ 5.4 การกำหนดรอบที่พิจารณาเลือกหยุดครั้งแรกมีค่าตั้งแต่ร้อยละ 10 ของจำนวนรอบสูงสุด ถึงร้อยละ 50 ของจำนวนรอบสูงสุด โดยเพิ่มขึ้นครั้งละ 10 ส่วนค่ารอบที่พิจารณาเลือกหยุดครั้งสุดท้ายและจำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้งกำหนดให้มีค่าคงที่ คือ รอบที่ร้อยละ 60 และ 1 ข่ายงาน ตามลำดับ ดังตารางที่ 5.3



รูปที่ 5.3 แผนภูมิเปรียบเทียบค่าผิดพลาดของค่ารอบพิจารณาเลือกหยุดครั้งแรก



รูปที่ 5.4 แผนภูมิเปรียบเทียบจำนวนรอบของค่ารอบพิจารณาเลือกหยุดครั้งแรก

เมื่อนำข้อมูลค่าผิดพลาดของแต่ละปัญหาจากตารางที่ 5.4 มาแสดงในรูปแบบกราฟเพื่อเปรียบเทียบระหว่างรอบที่พิจารณาเลือกหยุดครั้งแรกที่มีค่าต่างๆ ซึ่งแสดงดังรูปที่ 5.3 โดยนำค่าผิดพลาดแต่ละค่ามาหารด้วยค่าผิดพลาดค่าแรกในปัญหานั้นๆ เพื่อปรับค่าให้อยู่ในสัดส่วนเดียวกัน เช่นเดียวกับ ในรูปที่ 5.4 ที่แสดงการเปรียบเทียบจำนวนรอบที่ใช้ของแต่ละปัญหาระหว่างรอบที่พิจารณาเลือกหยุดครั้งแรกที่มีค่าต่างๆ

จากรูปที่ 5.3 แสดงให้เห็นว่า เมื่อค่าของรอบที่พิจารณาเลือกหยุดครั้งแรกมีค่าน้อยกว่าร้อยละ 20 ผลลัพธ์ของค่าผิดพลาดที่ได้ในปัญหาส่วนใหญ่จะมีค่าผิดพลาดที่สูง แต่จะมีค่าต่ำลงเมื่อค่าของรอบที่พิจารณาเลือกหยุดครั้งแรกมีค่าตั้งแต่ร้อยละ 20 ขึ้นไป และแนวโน้มการเปลี่ยนแปลงในช่วงหลังนี้จะเริ่มมีค่าน้อยลงด้วย ยกเว้นในกรณีของปัญหา Mushroom ที่ค่าผิดพลาดจากการฝึกอบรมมีค่าน้อยมากหรือมีค่าเข้าใกล้ศูนย์ ซึ่งเมื่อผลลัพธ์ของการเรียนรู้มีการเปลี่ยนแปลงเพียงเล็กน้อยจะส่งผลให้ค่าผิดพลาดที่นำมาเปรียบเทียบกันมีแตกต่างกันมาก อันเป็นสาเหตุให้กราฟที่แสดงของปัญหานี้มีความแปรปรวนมาก เมื่อพิจารณาจำนวนรอบที่ใช้ทั้งหมดดังแสดงในรูปที่ 5.4 ค่าของจำนวนรอบรวมมีการเพิ่มขึ้นแบบคงที่ตามการเพิ่มขึ้นของรอบที่พิจารณาเลือกหยุดครั้งแรก ซึ่งแนวโน้มการเพิ่มขึ้นจะมีลักษณะเช่นเดียวกันในทุกปัญหา

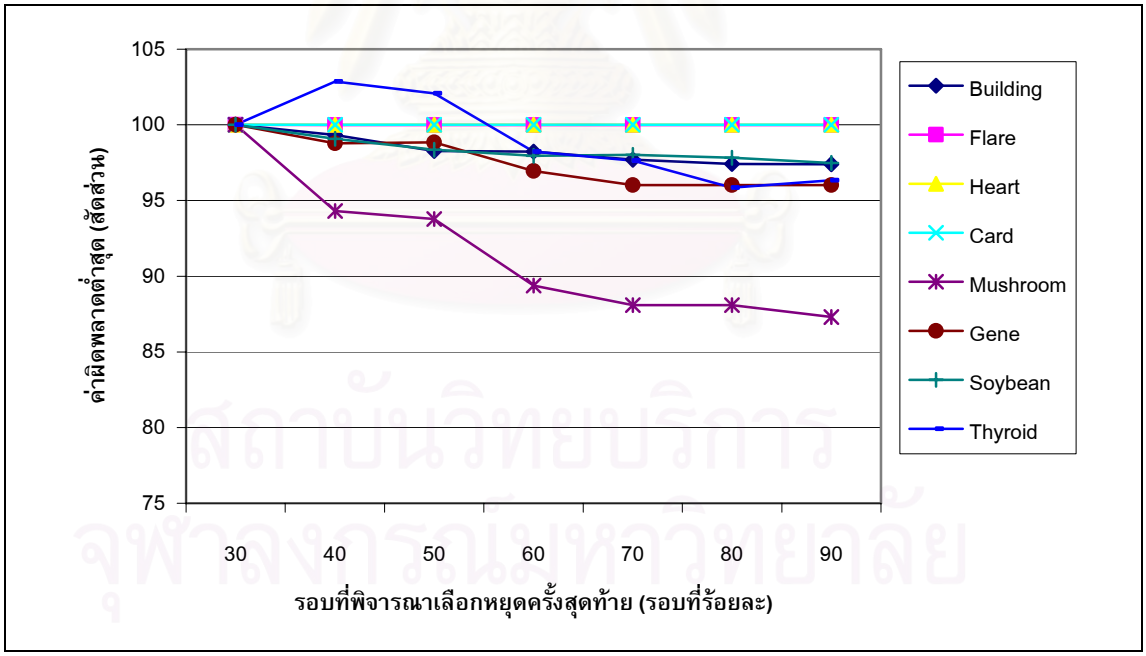
การกำหนดรอบที่พิจารณาเลือกหยุดครั้งแรกให้มีค่าต่ำหรือน้อยกว่าร้อยละ 20 ในการทดลองนั้น ส่งผลให้ได้ค่าผิดพลาดจากการเรียนรู้ที่สูง สาเหตุน่าจะมาจากการเลือกพิจารณาหยุดข่ายงานที่เร็วตั้งแต่ในช่วงแรกของการฝึกอบรม แนวโน้มค่าผิดพลาดของข่ายงานอาจยังมีค่าที่ไม่ชัดเจน ส่งผลให้การพิจารณาเลือกหยุดข่ายงานกระทำได้ไม่ถูกต้อง ดังนั้น ในการกำหนดรอบที่พิจารณาเลือกหยุดครั้งแรกที่เหมาะสมในการฝึกอบรมจึงควรกำหนดให้มีค่าตั้งแต่ร้อยละ 20 ของจำนวนรอบสูงสุด

5.3.2 การวิเคราะห์ความไวของรอบที่พิจารณาเลือกหยุดครั้งสุดท้าย

การเปรียบเทียบค่าผิดพลาดเมื่อกำหนดรอบที่พิจารณาเลือกหยุดครั้งสุดท้ายให้มีค่าต่างๆ คือ มีค่าตั้งแต่ร้อยละ 30 ถึงร้อยละ 90 ของจำนวนรอบสูงสุด และกำหนดให้ตัวแปรอื่นมีค่าคงที่ตามค่าพื้นฐานในตารางที่ 5.3 ซึ่งในการทดสอบนี้ให้ผลลัพธ์ค่าผิดพลาดต่ำสุดจากแต่ละปัญหามีค่าดังตารางที่ 5.5

ตารางที่ 5.5 ค่าผิดพลาดจากการกำหนดรอบที่พิจารณาเลือกหยุดครั้งสุดท้าย

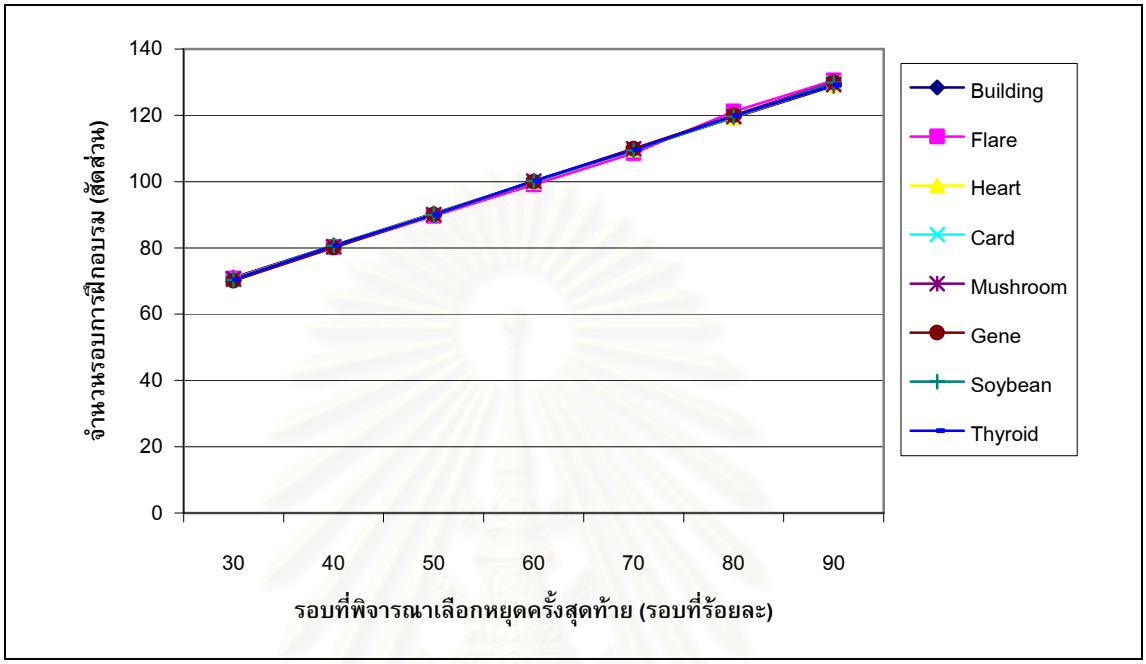
ปัญหา	ค่าผิดพลาดต่ำสุดของการเรียนรู้						
	รอบที่ ร้อยละ 30	รอบที่ ร้อยละ 40	รอบที่ ร้อยละ 50	รอบที่ ร้อยละ 60	รอบที่ ร้อยละ 70	รอบที่ ร้อยละ 80	รอบที่ ร้อยละ 90
ปัญหาประมาณค่า							
Building	0.008203	0.008148	0.008062	0.008058	0.008014	0.007992	0.007991
Flare	0.003785	0.003784	0.003784	0.003784	0.003784	0.003784	0.003784
Heart	0.042163	0.042163	0.042163	0.042163	0.042163	0.042163	0.042163
ปัญหาจัดแบ่งประเภท							
Card	0.382659	0.382659	0.382659	0.382659	0.382659	0.382659	0.382659
Mushroom	0.009503	0.008961	0.008912	0.008493	0.008370	0.008370	0.008296
Gene	0.174307	0.172166	0.172292	0.169018	0.167380	0.167380	0.167380
Soybean	0.443275	0.439182	0.435965	0.434211	0.434503	0.433625	0.432163
Thyroid	0.028083	0.028889	0.028667	0.027583	0.027417	0.026917	0.027056



รูปที่ 5.5 แผนภูมิเปรียบเทียบค่าผิดพลาดของค่ารอบพิจารณาเลือกหยุดครั้งสุดท้าย

เมื่อนำข้อมูลค่าผิดพลาดของแต่ละปัญหาจากตารางที่ 5.5 มาแสดงในรูปแบบกราฟ เพื่อเปรียบเทียบระหว่างรอบที่พิจารณาเลือกหยุดครั้งสุดท้ายที่มีค่าต่าง ๆ ซึ่งแสดงดังรูปที่ 5.5

โดยนำค่าผิดพลาดแต่ละค่ามาหารด้วยค่าผิดพลาดค่าแรกในปัญหานั้นๆ เพื่อปรับค่าให้อยู่ในสัดส่วนเดียวกัน เช่นเดียวกับ ในรูปที่ 5.6 ที่แสดงการเปรียบเทียบจำนวนรอบที่ใช้ของแต่ละปัญหาระหว่างรอบที่พิจารณาเลือกหยุดครั้งสุดท้ายที่มีค่าต่างๆ



รูปที่ 5.6 แผนภูมิเปรียบเทียบจำนวนรอบของค่ารอบพิจารณาเลือกหยุดครั้งสุดท้าย

จากรูปที่ 5.5 เมื่อกำหนดรอบที่พิจารณาเลือกหยุดครั้งสุดท้ายในมีค่าต่างๆ จะเห็นได้ว่า ค่าของรอบที่พิจารณาเลือกหยุดครั้งสุดท้ายที่มีค่าน้อยในช่วงแรก ค่าผิดพลาดต่ำสุดที่ได้จะมีค่าสูง แต่จะเริ่มลดลงเรื่อยๆ ตามจำนวนรอบที่เพิ่มขึ้น เมื่อพิจารณาจากผลการทดลองในปัญหาต่างๆ จะพบว่าที่รอบพิจารณาเลือกหยุดครั้งสุดท้ายรอบที่ร้อยละ 60 ของจำนวนรอบสูงสุด ค่าผิดพลาดที่ได้จะเริ่มมีแนวโน้มการเปลี่ยนแปลงลดลง ดังเช่นในรูปที่ 5.5 เมื่อพิจารณาค่าผิดพลาดที่ลดลงระหว่างรอบที่ร้อยละ 50 และ 60 ค่าผิดพลาดมีการเปลี่ยนแปลงลดลงสูงสุดถึง 4.73 เปอร์เซ็นต์ แต่ค่าผิดพลาดที่ลดลงระหว่างรอบที่ร้อยละ 60 และ 70 ค่าผิดพลาดมีการลดลงสูงสุดเพียง 1.39 เปอร์เซ็นต์ แสดงให้เห็นว่าเป็นช่วงที่ค่าผิดพลาดมีการเปลี่ยนแปลงน้อยลง และเมื่อพิจารณาจำนวนรอบที่ใช้ทั้งหมดของการกำหนดจำนวนรอบที่ค่าต่างๆ ดังแสดงในรูปที่ 5.6 เมื่อเพิ่มค่าของรอบที่พิจารณาเลือกหยุดครั้งสุดท้าย ค่าของจำนวนรอบรวมจะมีค่าสูงขึ้นตาม โดยค่าที่เพิ่มขึ้นนี้มีค่าเป็นแบบคงที่

สำหรับในปัญหาที่มีการเกิดความพอดีกับตัวอย่างมากเกินไป (Overfitting) เช่น ในปัญหา Flare ปัญหา Heart และปัญหา Card ลักษณะของค่าผิดพลาดต่ำสุดที่ได้จะต่างจากปัญหาอื่น คือ ค่าผิดพลาดที่ได้จะมีลักษณะคงที่ ที่เป็นเช่นนี้เนื่องจากในปัญหาที่เกิดความพอดี

กับตัวอย่างมากเกินไป ค่าผิดพลาดเมื่อเริ่มฝึกอบรมจะมีค่าลดลงจนถึงค่าหนึ่ง และจะมีแนวโน้มเพิ่มมากขึ้นหลังจากนั้น ทำให้ค่าผิดพลาดต่ำสุดที่ได้จากการทดลองในปัญหาลักษณะนี้ เมื่อได้ค่าที่ต่ำที่สุดแล้วจะไม่มีการเปลี่ยนแปลงอีก

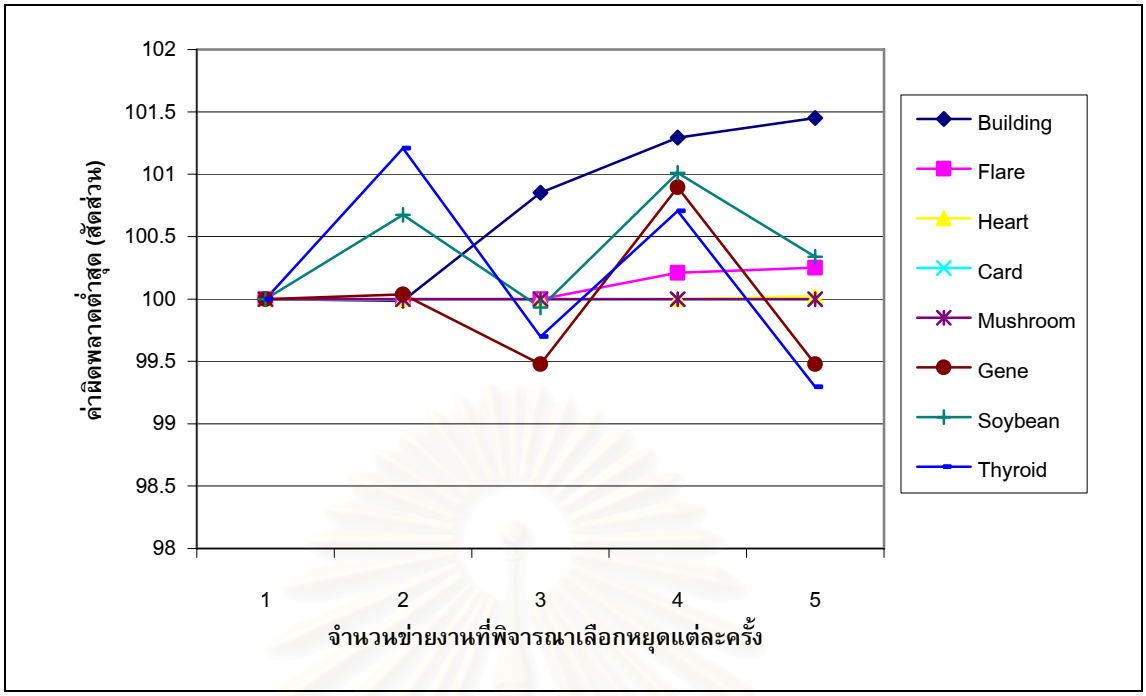
5.3.3 การวิเคราะห์ความไวของจำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้ง

การทดลองนี้ได้ทำการเปรียบเทียบค่าผิดพลาดเมื่อกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแต่ละครั้งให้มีค่าต่างๆ โดยในการทดลองนี้กำหนดให้มีจำนวนข่ายงานเริ่มต้นที่เรียนรู้พร้อมกันเท่ากับ 10 ข่ายงาน และพิจารณาเลือกหยุดข่ายงานแต่ละครั้งตั้งแต่ 1 ถึง 5 ข่ายงาน ซึ่งกำหนดให้ตัวแปรอื่นมีค่าคงที่ตามค่าพื้นฐานในตารางที่ 5.3 ในการทดสอบนี้ให้ผลลัพธ์ค่าผิดพลาดต่ำสุดจากแต่ละปัญหามีค่าดังตารางที่ 5.6

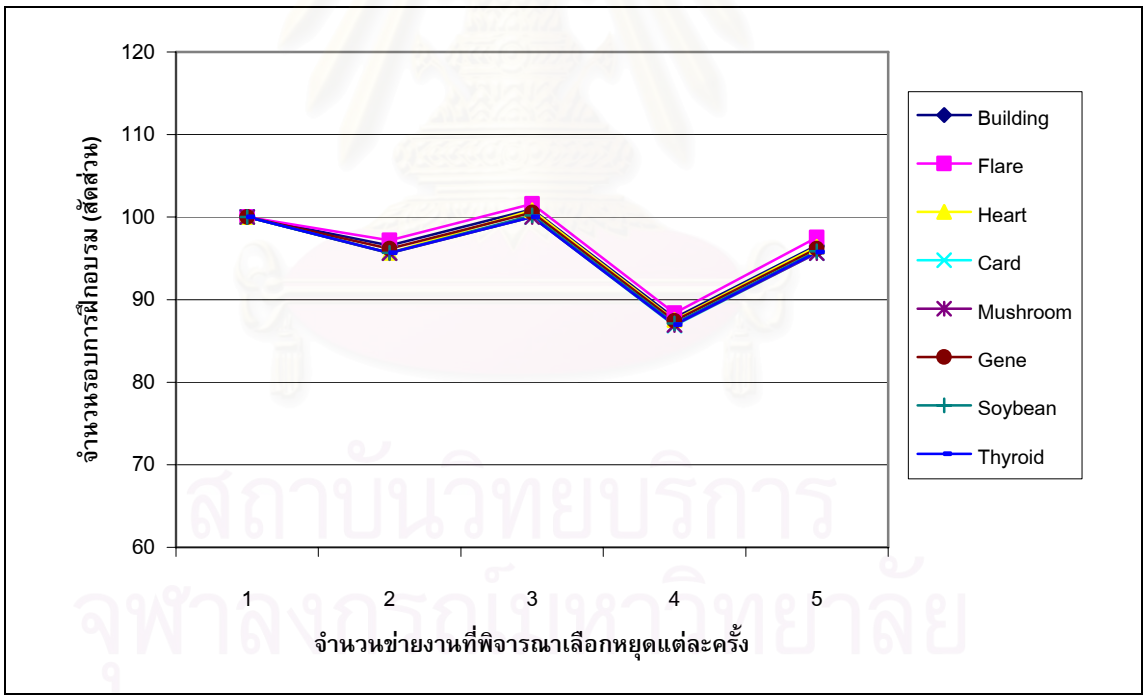
ตารางที่ 5.6 ค่าผิดพลาดจากการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุด

ปัญหา	ค่าผิดพลาดของการเรียนรู้				
	1 ข่ายงาน	2 ข่ายงาน	3 ข่ายงาน	4 ข่ายงาน	5 ข่ายงาน
ปัญหาประมาณค่า					
Building	0.008058	0.008058	0.008142	0.008171	0.008143
Flare	0.003784	0.003784	0.003784	0.003792	0.003793
Heart	0.042163	0.042163	0.042163	0.042163	0.042173
ปัญหาจัดแบ่งประเภท					
Card	0.382659	0.382659	0.382659	0.382659	0.382659
Mushroom	0.008493	0.008493	0.008493	0.008493	0.008493
Gene	0.169018	0.169081	0.168136	0.170529	0.168136
Soybean	0.434211	0.437134	0.433918	0.438596	0.435672
Thyroid	0.027583	0.027917	0.027500	0.027778	0.027389

เมื่อนำข้อมูลค่าผิดพลาดของแต่ละปัญหาจากตารางที่ 5.6 มาแสดงในรูปแบบกราฟเพื่อเปรียบเทียบระหว่างการใช้จำนวนข่ายงานที่เลือกหยุดแต่ละครั้งที่ค่าต่างๆ สามารถแสดงได้ดังรูปที่ 5.7 ซึ่งค่าที่เปรียบเทียบนี้เป็นการนำค่าผิดพลาดแต่ละค่ามาหารด้วยค่าผิดพลาดค่าแรกในปัญหานั้นๆ เพื่อปรับค่าให้อยู่ในสัดส่วนเดียวกัน เช่นเดียวกับจำนวนรอบที่ใช้ของแต่ละปัญหาสามารถแสดงการเปรียบเทียบได้ดังรูปที่ 5.8



รูปที่ 5.7 แผนภูมิเปรียบเทียบค่าผิดพลาดของค่าจำนวนข่ายงานเลือกหยุด



รูปที่ 5.8 แผนภูมิเปรียบเทียบจำนวนรอบของค่าจำนวนข่ายงานเลือกหยุด

จากการเปรียบเทียบค่าผิดพลาดของการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดที่ใช้ค่าต่างๆ จะเห็นได้ว่า การเลือกหยุดครั้งละ 1 และ 3 ข่ายงานให้ผลของค่าผิดพลาดที่น้อยกว่าในปัญหาส่วนใหญ่ ซึ่งเมื่อคำนวณค่าเฉลี่ยจากสัดส่วนค่าผิดพลาดแล้ว พบว่าการเลือก

หยุดครั้งละ 1 ข่ายงาน ให้ค่าผิดพลาดที่ต่ำที่สุด อย่างไรก็ตามจำนวนรอบที่ใช้นั้นมีค่าที่สูงกว่าการเลือกหยุดด้วยจำนวนข่ายงานอื่นๆ ส่วนใหญ่

เมื่อพิจารณาการเลือกหยุดครั้งละ 2 ข่ายงานอย่างละเอียดจะพบว่า การเลือกหยุดครั้งละ 2 ข่ายงานที่ใช้จำนวนรอบน้อยกว่าการเลือกหยุดครั้งละ 1 ข่ายงาน โดยเฉลี่ยถึง 3.98 เปอร์เซ็นต์ สามารถให้ค่าผิดพลาดที่ดีใกล้เคียงกับการเลือกหยุดครั้งละ 1 ข่ายงาน ยกเว้นในบางปัญหา คือ ปัญหา Soybean และปัญหา Thyroid ซึ่งเมื่อพิจารณาลักษณะของค่าผิดพลาดที่เกิดขึ้นจากการฝึกอบรมใน 2 ปัญหานี้จะเห็นได้ว่า ค่าผิดพลาดของการเรียนรู้แต่ละข่ายงานมีความแปรปรวน ซึ่งต่างจากปัญหาอื่นที่มีแนวโน้มของค่าผิดพลาดที่ต่อเนื่อง จึงน่าจะเป็นข้อสังเกตหนึ่งที่ใช้ในการพิจารณาเลือกจำนวนข่ายงานที่ใช้หยุดแต่ละครั้ง

ดังนั้น ในการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดในแต่ละครั้ง สำหรับปัญหาทั่วไปสามารถใช้การเลือกหยุดครั้งละ 1 ข่ายงาน ได้โดยไม่จำเป็นต้องคำนึงถึงลักษณะของค่าผิดพลาดที่เกิดขึ้นจากปัญหานั้นๆ แต่ในกรณีที่ทราบว่า ค่าผิดพลาดจากปัญหาที่ต้องทดสอบมีลักษณะที่ต่อเนื่อง หรือมีความแปรปรวนน้อย การใช้การเลือกหยุดครั้งละ 2 ข่ายงานสามารถช่วยลดเวลาการเรียนรู้ลงได้

จากการกำหนดจำนวนข่ายงานเริ่มต้นในการทดลองนี้ที่กำหนดให้มีจำนวน 10 ข่ายงาน ซึ่งมีค่าค่อนข้างน้อยเมื่อเทียบกับการพิจารณาเลือกหยุดครั้งละ 1 ถึง 5 ข่ายงาน ส่งผลให้ค่าจำนวนรอบที่ใช้ทั้งหมดของการเรียนรู้แต่ละแบบมีจำนวนที่แตกต่างกันอย่างเห็นได้ชัด ดังรูปที่ 5.8 แต่ถ้าจำนวนข่ายงานเริ่มต้นมีค่าเพิ่มมากขึ้น จำนวนรอบที่ใช้ทั้งหมดของการเลือกหยุดที่มีจำนวนต่างกันจะมีค่าใกล้เคียงกันมากยิ่งขึ้น

5.4 ผลกระทบของการกำหนดจำนวนรอบและข่ายงานที่พิจารณาเลือกหยุดแบบไม่คงที่

การกำหนดจำนวนรอบและข่ายงานที่พิจารณาเลือกหยุดนั้น ประกอบด้วยตัวแปรที่เกี่ยวข้องหลายตัวด้วยกัน คือ รอบที่พิจารณาเลือกหยุดครั้งแรก รอบที่พิจารณาเลือกหยุดครั้งสุดท้าย และจำนวนข่ายงานที่เลือกหยุดในแต่ละครั้ง ซึ่งค่าของตัวแปรแต่ละตัวนี้ส่งผลกระทบต่อประสิทธิภาพของการเรียนรู้ ทำให้ในการกำหนดค่าจำเป็นต้องเลือกใช้ค่าที่เหมาะสมเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด การทดลองนี้จึงได้ศึกษาถึงผลกระทบของตัวแปรแต่ละตัว เพื่อให้เกิดแนวทางในการกำหนดค่าของตัวแปรเหล่านี้

5.4.1 การวิเคราะห์ความไวของรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชัน

การใช้วิธีกำหนดรอบที่พิจารณาแบบขึ้นกับความชันในการทดลองนี้จะกำหนดให้มีการเลือกหยุดข่ายงานเมื่อความชันเฉลี่ยของค่าผิดพลาดทุกข่ายงานในรอบนั้นๆ มีค่าตรงตามที่กำหนด โดยจะเปรียบเทียบกับวิธีการกำหนดรอบที่พิจารณาเลือกหยุดแบบคงที่

การกำหนดค่าตัวแปรที่ใช้ในการทดลองนี้เพื่อให้สอดคล้องกับลักษณะของแนวโน้มค่าผิดพลาดที่เกิดขึ้น ได้กำหนดระดับความชันของค่าผิดพลาดที่จะพิจารณาเลือกหยุดให้มีค่าตั้งแต่ -10^{-2} ถึง -10^{-6} ตามจำนวนครั้งในการพิจารณา โดยให้การเลือกหยุดครั้งแรกมีความชันของค่าผิดพลาดเท่ากับ -10^{-2} ครั้งสุดท้ายเท่ากัน -10^{-6} และระยะห่างในแต่ละครั้งจะแบ่งเท่าๆ กันตามระดับในฟังก์ชันลอการิทึม (Logarithm) สาเหตุจากที่แนวโน้มของค่าผิดพลาดมีลักษณะการลดลงมีรูปแบบคล้ายกับฟังก์ชันลอการิทึม และในการพิจารณาเลือกหยุดครั้งแรกจะกำหนดให้มีการพิจารณาเมื่อจำนวนรอบของการเรียนรู้มีค่ามากกว่าร้อยละ 5 ของจำนวนรอบสูงสุด เนื่องจากค่าผิดพลาดที่เกิดขึ้นของการฝึกอบรมในช่วงแรกของข่ายงานประสาทเทียมจะยังคงมีค่าที่ผันผวนอยู่ ทำให้ความชันของค่าผิดพลาดที่ประมาณได้อาจมีค่าที่ไม่ถูกต้อง และในที่นี้ได้กำหนดให้ใช้ข่ายงานในการทดลองทั้งหมด 10 ข่ายงาน และจำนวนข่ายงานที่เลือกหยุดในแต่ละครั้งเท่ากับ 1 ข่ายงาน ทำให้ในการทดลองนี้จะพิจารณาเลือกหยุดทั้งสิ้น 9 ครั้ง โดยแต่ละครั้งจะพิจารณาระดับความชันมากกว่า -1×10^{-2} -0.3×10^{-2} -1×10^{-3} -0.3×10^{-3} -1×10^{-4} -0.3×10^{-4} -1×10^{-5} -0.3×10^{-5} และ -1×10^{-6} ตามลำดับ

ตารางที่ 5.7 ค่าผิดพลาดและจำนวนรอบที่ใช้ของการกำหนดรูปแบบรอบที่พิจารณา

ปัญหา	รอบที่เลือกหยุดแบบคงที่		รอบที่เลือกหยุดแบบขึ้นกับความชัน		% การเปลี่ยนแปลงแบบขึ้นกับความชันเทียบกับแบบคงที่	
	ค่าผิดพลาด	จำนวนรอบ	ค่าผิดพลาด	จำนวนรอบ	ค่าผิดพลาด	จำนวนรอบ
ปัญหาประมาณค่า						
Building	0.008013	1859.00	0.008150	1367.95	1.71%	-26.41%
Flare	0.003784	1125.00	0.003784	878.65	0.00%	-21.90%
Heart	0.042163	1098.00	0.042163	885.30	0.00%	-19.37%
ปัญหาจัดแบ่งประเภท						
Card	0.382659	1084.00	0.382659	941.60	0.00%	-13.14%
Mushroom	0.008493	7432.00	0.009010	4399.55	6.09%	-40.80%
Gene	0.169018	3239.00	0.172733	2637.60	2.20%	-18.57%
Soybean	0.434211	4237.00	0.440352	2981.90	1.41%	-29.62%
Thyroid	0.027583	3856.00	0.030611	1393.45	10.98%	-63.86%

การกำหนดรอบที่พิจารณาเลือกหยุดแบบคงที่และแบบขึ้นกับความชันของค่าผิดพลาดนี้มีความแตกต่างกันของจำนวนรอบที่ใช้ เนื่องจากจำนวนรอบที่ใช้วิธีการกำหนดแบบขึ้นกับความชันนั้นจะมีจำนวนรอบที่ไม่แน่นอน ทำให้ในการเปรียบเทียบระหว่างทั้งสองวิธีการไม่สามารถทำได้โดยตรง อย่างไรก็ตามเมื่อนำค่าผิดพลาดต่ำสุดที่ได้และจำนวนรอบที่ใช้ของทั้งสองวิธีมาพิจารณา ทำให้สามารถเห็นลักษณะการใช้งานของวิธีการทั้งสองได้ โดยผลลัพธ์ค่าผิดพลาดและจำนวนรอบที่ใช้ในปัญหาที่ทดสอบต่างๆ ได้นำมาเปรียบเทียบดังตารางที่ 5.7

จากตารางที่ 5.7 ผลการเปรียบเทียบค่าผิดพลาดและจำนวนรอบที่ใช้ระหว่างวิธีการกำหนดรอบที่พิจารณาเลือกหยุดแบบคงที่และแบบขึ้นกับความชันแสดงให้เห็นว่า ค่าผิดพลาดที่ได้จากการใช้วิธีการแบบขึ้นกับความชันให้ค่าที่มากกว่าหรือเท่ากันกับวิธีการแบบคงที่ในทุกปัญหา ซึ่งมีถึง 5 ปัญหาที่มีค่ามากกว่าจากปัญหาทั้งหมดที่ทำการทดลอง 8 ปัญหา แต่อย่างไรก็ตามเมื่อพิจารณาจำนวนรอบที่ใช้จะพบว่า วิธีการแบบขึ้นกับความชันได้ใช้จำนวนรอบในการเรียนรู้ทั้งหมดน้อยกว่าวิธีการแบบคงที่ ตัวอย่างเช่น ในปัญหา Thyroid ที่ค่าผิดพลาดจากวิธีการแบบขึ้นกับความชันมีค่ามากกว่าวิธีการแบบคงที่ 10.98 เปอร์เซ็นต์ แต่ใช้จำนวนรอบที่ลดลงถึง 63.86 เปอร์เซ็นต์ และสำหรับในปัญหา Flare ปัญหา Heart และปัญหา Card ซึ่งเป็นปัญหาที่เกิดความพอดีกับตัวอย่างมากเกินไป จะสังเกตได้ว่า ค่าผิดพลาดที่ได้จากทั้งสองวิธีมีค่าเท่ากัน แต่จำนวนรอบที่ใช้ของวิธีการแบบขึ้นกับความชันมีค่าน้อยกว่า 13.14 ถึง 21.90 เปอร์เซ็นต์

จากผลการทดลองในการทดสอบวิธีการกำหนดรอบที่พิจารณาเลือกหยุดแบบขึ้นกับความชันแสดงให้เห็นได้ว่า การฝึกอบรมข่ายงานประสาทเทียมโดยใช้วิธีการกำหนดจำนวนรอบแบบขึ้นกับความชันแม้ว่าจะได้ค่าที่ไม่ดีเท่าวิธีการกำหนดแบบคงที่เล็กน้อย แต่ใช้จำนวนรอบที่น้อยกว่าอย่างมีนัยสำคัญ ซึ่งในปัญหาที่เกิดความพอดีกับตัวอย่างมากเกินไปวิธีการนี้ให้ค่าผิดพลาดที่เท่ากัน ในขณะที่ใช้จำนวนรอบน้อยกว่า ดังนั้น ในการฝึกอบรมปัญหาลักษณะนี้ วิธีการกำหนดจำนวนรอบแบบขึ้นกับความชันจึงมีความเหมาะสมกว่า และข้อดีอีกประการหนึ่งของวิธีการนี้ คือ การนำไปใช้ที่ไม่จำเป็นต้องคำนึงจำนวนรอบสูงสุดล่วงหน้า เนื่องจากรอบที่พิจารณาเลือกหยุดจะขึ้นกับความชันของค่าผิดพลาดที่เกิดขึ้นในขณะเรียนรู้

5.4.2 การวิเคราะห์ความไวของจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบเพิ่มขึ้น

ในการทดลองการพิจารณาเลือกหยุดโดยกำหนดจำนวนข่ายงานให้เป็นแบบเพิ่มขึ้นในแต่ละครั้งนี้ ได้ทำการทดสอบเปรียบเทียบกับวิธีการกำหนดจำนวนข่ายงานแบบคงที่ โดยกำหนดให้รอบที่พิจารณาเลือกหยุดนั้นมีค่าเป็นแบบคงที่ ซึ่งการใช้วิธีกำหนดจำนวนข่ายงานให้เป็นแบบเพิ่มขึ้นในการทดลองนี้จะกำหนดให้มีค่าจำนวนข่ายงานที่เลือกหยุดเพิ่มขึ้นครั้งละ 1 ข่ายงาน เนื่องจากได้ใช้ข่ายงานในการทดลองทั้งหมด 10 ข่ายงาน การพิจารณาหยุดจึงได้กำหนดค่าครั้งละ 1 2 3 และ 3 ข่ายงาน ตามลำดับในแต่ละครั้ง

ตารางที่ 5.8 ค่าผิดพลาดของจำนวนข่ายงานที่พิจารณาแบบคงที่และแบบเพิ่มขึ้น

ปัญหา	ค่าผิดพลาดของการเรียนรู้					
	แบบ 20-60			แบบ 30-60		
	คงที่	เพิ่มขึ้น	%Δ	คงที่	เพิ่มขึ้น	%Δ
ปัญหาประมาณค่า						
Building	0.008058	0.008041	-0.21%	0.008058	0.008007	-0.63%
Flare	0.003784	0.003784	0.00%	0.003784	0.003784	0.00%
Heart	0.042163	0.042163	0.00%	0.042163	0.042163	0.00%
ปัญหาจัดแบ่งประเภท						
Card	0.382659	0.382659	0.00%	0.382659	0.382659	0.00%
Mushroom	0.008493	0.008493	0.00%	0.00837	0.008493	1.47%
Gene	0.169018	0.168073	-0.56%	0.168136	0.168136	0.00%
Soybean	0.434211	0.432163	-0.47%	0.432748	0.432455	-0.07%
Thyroid	0.027583	0.027556	-0.10%	0.02775	0.026833	-3.30%

ตารางที่ 5.9 จำนวนรอบที่ใช้ของจำนวนข่ายงานที่พิจารณาแบบคงที่และแบบเพิ่มขึ้น

ปัญหา	จำนวนรอบที่ใช้ทั้งหมดของการเรียนรู้					
	แบบ 20-60			แบบ 30-60		
	คงที่	เพิ่มขึ้น	%Δ	คงที่	เพิ่มขึ้น	%Δ
ปัญหาประมาณค่า						
Building	1869	2067	10.59%	2058	2215	7.63%
Flare	1135	1264	11.37%	1252	1353	8.07%
Heart	1108	1220	10.11%	1225	1309	6.86%
ปัญหาจัดแบ่งประเภท						
Card	1094	1206	10.24%	1202	1286	6.99%
Mushroom	7442	8198	10.16%	8144	8738	7.29%
Gene	3249	3587	10.40%	3564	3835	7.60%
Soybean	4247	4682	10.24%	4643	4983	7.32%
Thyroid	3866	4258	10.14%	4226	4538	7.38%

การกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่และแบบเพิ่มขึ้นนี้มีความแตกต่างในการกำหนดค่าตัวแปรที่ใช้เป็นหลักเกณฑ์การพิจารณา ทำให้ในการเปรียบเทียบระหว่างทั้งสองวิธีการไม่สามารถทำได้โดยตรง ดังนั้น เพื่อให้เห็นประสิทธิภาพของการเรียนรู้ทั้งสองรูปแบบนี้ การทดลองนี้ได้เลือกรูปแบบการฝึกอบรมที่กำหนดให้รอบที่พิจารณาเลือกหยุดครั้งแรกและครั้งสุดท้ายมีค่าเป็นแบบ 20-60 และแบบ 30-60 ในการพิจารณาเปรียบเทียบทั้งสองวิธีการเพื่อศึกษาค่าผิดพลาดและจำนวนรอบที่ใช้ ซึ่งค่าผิดพลาดต่ำสุดของรูปแบบที่เลือกนำมาเปรียบเทียบได้แสดงในตารางที่ 5.8 และจำนวนรอบที่ใช้แสดงในตารางที่ 5.9

จากข้อมูลค่าผิดพลาดในตารางที่ 5.8 ที่เปรียบเทียบระหว่างการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่และแบบเพิ่มขึ้นจะเห็นได้ว่า โดยส่วนใหญ่แล้ววิธีการแบบเพิ่มขึ้นจะให้ค่าผิดพลาดที่ต่ำกว่าเพียงเล็กน้อยหรือเท่ากับแบบคงที่ อย่างไรก็ตาม ข้อมูลจำนวนรอบที่ใช้ในตารางที่ 5.9 ได้แสดงจำนวนรอบที่ใช้ของวิธีการแบบเพิ่มขึ้นที่มากกว่าแบบคงที่ในทุกปัญหา ซึ่งมีค่าเฉลี่ยสูงถึง 10.41 เปอร์เซ็นต์ ในการกำหนดแบบ 20-60 และ 7.39 เปอร์เซ็นต์ ในการกำหนดแบบ 30-60

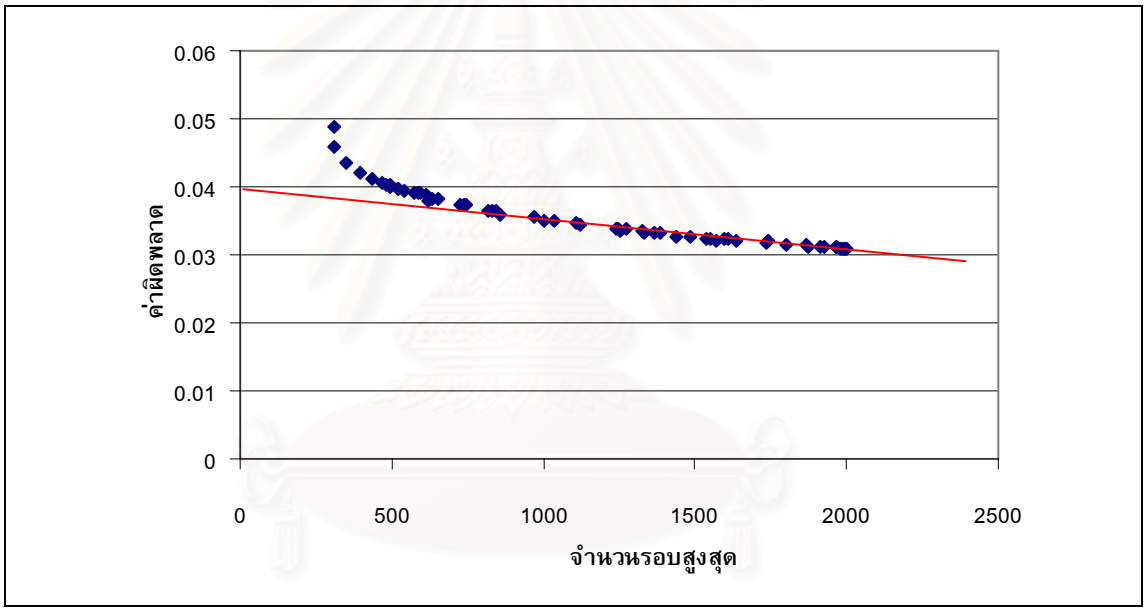
จากการเปรียบเทียบระหว่างข่ายงานที่พิจารณาเลือกหยุดแบบคงที่และแบบเพิ่มขึ้นของทั้งสองรูปแบบสามารถสรุปได้ว่า วิธีการกำหนดจำนวนข่ายงานที่พิจารณาเลือกหยุดแบบคงที่และแบบเพิ่มขึ้นให้ผลลัพธ์ของค่าผิดพลาดที่ไม่แตกต่างกันมาก แต่วิธีการแบบเพิ่มขึ้นจะใช้จำนวนรอบที่มากกว่า สาเหตุเนื่องจากการใช้วิธีการแบบเพิ่มขึ้นจะเลือกหยุดจำนวนข่ายงานที่น้อยในช่วงแรก ทำให้ต้องฝึกอบรมข่ายงานจำนวนมากไปจนกระทั่งถึงในช่วงหลัง จึงส่งผลให้จำนวนรอบรวมของวิธีการนี้มีค่ามาก ดังนั้น ในการฝึกอบรมกับปัญหาทั่วไปจึงควรใช้วิธีการกำหนดข่ายงานที่พิจารณาเลือกหยุดแบบคงที่

5.5 ผลกระทบของการกำหนดจำนวนรอบสูงสุด

ในการศึกษาลักษณะการเรียนรู้ของข่ายงานประสาทเทียมด้วยวิธีการต่าง ๆ เพื่อเปรียบเทียบประสิทธิภาพในการให้คำตอบของข่ายงานจากค่าผิดพลาดที่เกิดขึ้น สิ่งหนึ่งที่เป็นตัวแปรสำคัญในการฝึกอบรม คือ จำนวนรอบสูงสุดที่ใช้เรียนรู้ เนื่องจากค่าจำนวนรอบสูงสุดนี้จะส่งผลต่อค่าผิดพลาดต่ำสุดที่ได้และทรัพยากรที่ใช้ ตัวอย่างเช่น เมื่อมีการกำหนดจำนวนรอบสูงสุดที่ใช้เรียนรู้น้อยเกินไปจะส่งผลให้คำตอบของค่าผิดพลาดที่ได้มีโอกาสได้ค่าที่สูง หรือเมื่อมีการกำหนดจำนวนรอบที่มากเกินไปจะส่งผลให้ใช้เวลาและทรัพยากรในการคำนวณที่เกินความจำเป็น ดังนั้นจึงควรกำหนดค่านี้ให้เหมาะสมกับความต้องการ ผลการทดลองในส่วนนี้จึงนำเสนอถึงผลกระทบของค่าตัวแปรที่ใช้ในการกำหนดจำนวนรอบสูงสุด ดังต่อไปนี้

การกำหนดจำนวนรอบสูงสุดในงานวิจัยนี้ได้กำหนดขึ้นจาก 2 ตัวแปรด้วยกัน คือ จำนวนรอบที่ใช้ประมาณค่าความชัน ซึ่งเป็นค่าของจำนวนรอบท้ายสุดที่นำมาใช้ในการประมาณความชันของค่าผิดพลาดในขณะนั้น และจำนวนรอบที่ใช้วัดการลู่เข้า ซึ่งเป็นค่าจำนวนครั้งที่ใช้เป็นเกณฑ์ตัดสินว่าเกิดการลู่เข้าหรือไม่ โดยใช้ค่าที่กำหนดนี้เปรียบเทียบกับจำนวนครั้งที่ค่าประมาณความชันมีค่าน้อยกว่า ε ติดต่อกัน ดังนั้น การวิเคราะห์ความไวของการกำหนดจำนวนรอบสูงสุดจึงทำได้โดยศึกษาการผลที่เกิดขึ้นจากการเปลี่ยนแปลงค่าตัวแปรทั้งสองค่านี้

การหาจำนวนรอบสูงสุดจากการกำหนดจำนวนรอบที่ใช้ประมาณค่าความชันและจำนวนรอบที่ใช้วัดการลู่เข้าที่ค่าต่างๆ เมื่อนำค่าผิดพลาดต่ำสุดที่ได้พร้อมทั้งตำแหน่งรอบสุดท้ายมาแสดงรวมกัน จะได้กราฟที่มีลักษณะดังรูปที่ 5.9

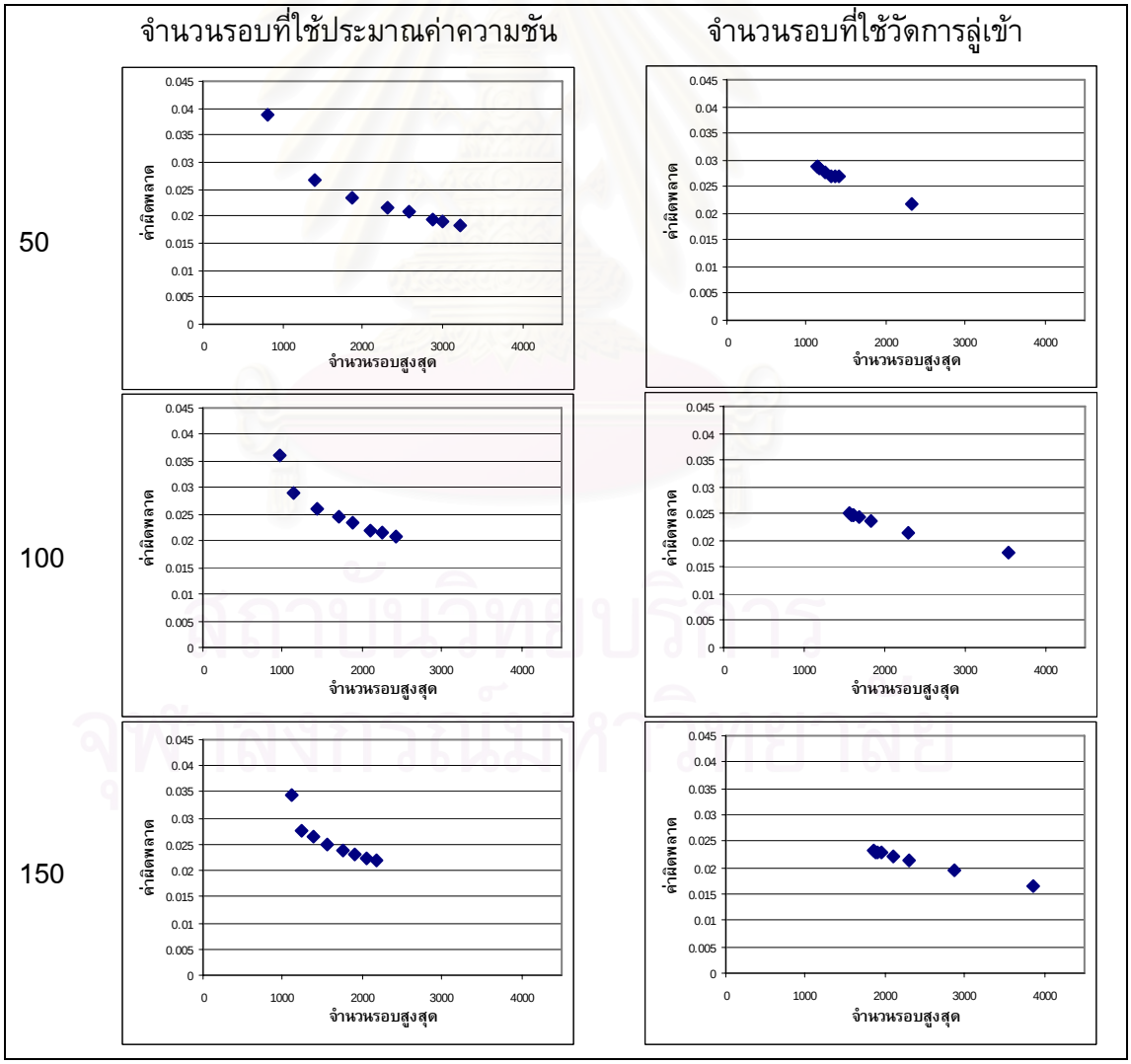


รูปที่ 5.9 จำนวนรอบสูงสุดเปรียบเทียบระหว่างการกำหนดจำนวนรอบต่างๆ

จากรูปที่ 5.9 ในกราฟนี้เป็นการแสดงค่าผิดพลาดที่ได้จากปัญหา Thyroid ซึ่งจะเห็นได้ว่า ในช่วงแรกที่จำนวนรอบสูงสุดมีค่าต่ำๆ ค่าผิดพลาดจะมีแนวโน้มการลดลงที่เร็ว แต่เมื่อผ่านไประยะหนึ่ง จำนวนรอบสูงสุดที่มีค่าเพิ่มขึ้นกับค่าผิดพลาดที่ลดลงจะมีแนวโน้มการเปลี่ยนแปลงด้วยอัตราคงที่ แสดงได้โดยเส้นตรงในรูปดังกล่าว ดังนั้น ในการกำหนดจำนวนรอบที่ใช้ประมาณค่าความชันและจำนวนรอบที่ใช้วัดการลู่เข้าให้มีค่าเหมาะสม จึงควรกำหนดให้จำนวนรอบสูงสุดมีค่าอยู่ในช่วงหลังที่มีอัตราการลดลงคงที่ และเมื่อพิจารณาค่าจำนวนรอบทั้งสองที่เป็นตัวกำหนดในปัญหานี้สามารถแบ่งจำนวนรอบที่ลดลงช่วงแรกและช่วงหลังได้จากค่าที่ขีดเส้นใต้ในตารางที่ 5.10

ตารางที่ 5.10 ค่าจำนวนรอบสูงสุดที่คำนวณจากการกำหนดตัวแปรให้มีค่าต่าง ๆ

จำนวนรอบที่ใช้วัดการลู่เข้า \ จำนวนรอบที่ใช้ประมาณค่าความชัน	25	50	75	100	125	150	175	200
25	308	309	350	395	437	467	498	522
50	1327	620	492	484	497	521	544	572
75	1910	1250	830	633	590	592	616	650
100	1980	1634	1275	1037	842	748	727	736
125	1992	1866	1571	1333	1109	971	859	815
150	1999	1966	1804	1547	1366	1249	1121	1000
175	1999	1986	1928	1738	1596	1481	1387	1241
200	1999	1994	1966	1875	1742	1607	1534	1437

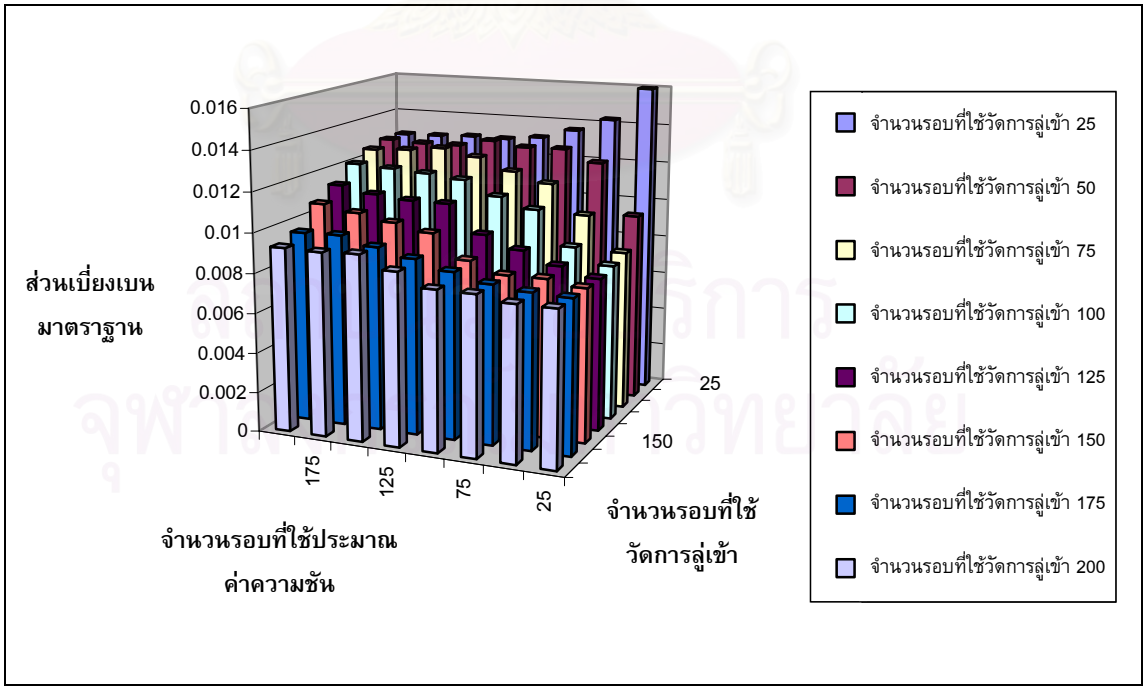


รูปที่ 5.10 ความสัมพันธ์ของค่าผิดพลาดจากจำนวนรอบประมาณความชันและวัดการลู่เข้า

จากตารางที่ 5.10 ค่าจำนวนรอบสูงสุดที่ได้จากการกำหนดจำนวนรอบทั้งสองที่ขีดเส้นใต้เป็นจำนวนรอบที่อยู่ในช่วงแรกที่มีอัตราการลดลงของค่าผิดพลาดที่เร็ว จึงไม่ควรกำหนดค่าให้อยู่ในส่วนนี้ แต่ควรกำหนดให้เป็นค่าในส่วนที่เหลือแทน โดยค่าผิดพลาดต่ำสุดที่ได้จะแปรผันตามค่าจำนวนรอบทั้งสองที่กำหนด ซึ่งความสัมพันธ์ของจำนวนรอบที่ใช้ประมาณค่าความชันและจำนวนรอบที่ใช้วัดการลู่เข้าทั้งสองสามารถแสดงในตัวอย่างดังรูปที่ 5.10

จากรูปที่ 5.10 ข้อมูลที่ถูกแสดงแบ่งเป็นสองส่วนด้วยกัน คือ ส่วนทางด้านซ้ายแสดงกราฟที่กำหนดจำนวนรอบที่ใช้ประมาณค่าความชันให้มีค่าคงที่ คือ 50 100 และ 150 รอบ ตามลำดับ และส่วนทางด้านขวาแสดงกราฟที่กำหนดจำนวนรอบที่ใช้วัดการลู่เข้าให้มีค่าคงที่ โดยมีค่าเช่นเดียวกันกับส่วนทางด้านขวา ภายในแต่ละกราฟได้แสดงค่าผิดพลาดต่ำสุดที่ตำแหน่งรอบสุดท้าย ซึ่งภายในกราฟส่วนทางด้านซ้ายจะแสดงค่าผิดพลาดและจำนวนรอบสูงสุดที่ได้จากการกำหนดจำนวนรอบที่ใช้วัดการลู่เข้าต่างๆ ที่มีค่าตั้งแต่ 25 ถึง 200 รอบ และภายในกราฟส่วนทางด้านขวาจะแสดงค่าผิดพลาดและจำนวนรอบสูงสุดที่ได้จากการกำหนดจำนวนรอบที่ใช้ประมาณค่าความชัน ซึ่งมีค่าเช่นเดียวกัน

ผลจากการเปรียบเทียบนี้จะเห็นได้ว่า การเพิ่มจำนวนรอบที่ใช้วัดการลู่เข้าส่งผลให้ค่าผิดพลาดที่ได้มีการลดลงมากกว่าการเพิ่มจำนวนรอบที่ใช้ประมาณค่าความชัน และเมื่อพิจารณาการเพิ่มจำนวนรอบที่ใช้วัดการลู่เข้าที่ส่งผลต่อค่าเบี่ยงเบนมาตรฐานของค่าผิดพลาดต่ำสุดให้มีค่าลดลงมากขึ้น ดังจะเห็นได้จากรูปที่ 5.11



รูปที่ 5.11 ส่วนเบี่ยงเบนมาตรฐานของค่าผิดพลาดในการกำหนดตัวแปรจำนวนรอบทั้งสอง

ดังนั้น การกำหนดจำนวนรอบทั้งสองที่เหมาะสม ซึ่งประกอบด้วย จำนวนรอบที่ใช้วัดการลู่เข้า และจำนวนรอบที่ใช้ประมาณค่าความชัน สำหรับการระบุค่าจำนวนรอบสูงสุด ควรกำหนดให้มีค่าที่ส่งผลให้ค่าผิดพลาดต่ำสุดที่ได้อยู่ในช่วงที่มีค่าลดลงด้วยความชันคงที่ หรือดังเช่นในตารางที่ 5.10 ที่ควรหลีกเลี่ยงการกำหนดด้วยค่าทั้งสองที่ทำให้ผลของค่าจำนวนรอบสูงสุดอยู่ในค่าที่ขีดเส้นใต้ โดยถ้าต้องการให้ค่าผิดพลาดต่ำสุดที่ได้มีค่าลดลงมากขึ้นหรือมีค่าเบี่ยงเบนมาตรฐานน้อยลงสามารถกำหนดได้จากการเพิ่มจำนวนรอบที่ใช้วัดการลู่เข้า

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดข่ายงานที่มีแนวโน้มของผลลัพธ์ที่ไม่ดี เพื่อให้การฝึกอบรมมีโอกาสการเกิดปัญหาค่าต่ำสุดเฉพาะที่ลดลงพร้อมทั้งใช้ทรัพยากรในการเรียนรู้อย่างมีประสิทธิภาพ และวิเคราะห์ความไวของตัวแปรที่ใช้ในการกำหนดวิธีการฝึกอบรมเพื่อแนวทางในการเลือกค่าที่เหมาะสมในการนำไปใช้งาน ซึ่งจากผลการวิจัยสามารถสรุปได้ดังนี้

6.1 สรุปผลการวิจัย

1. วิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การพิจารณาเลือกหยุดที่นำเสนอนี้ได้ประยุกต์ใช้แนวคิดการฝึกอบรมข่ายงานแบบหลายครั้งที่กำหนดค่าน้ำหนักเริ่มที่แตกต่างกันและเลือกข่ายงานที่ดีที่สุด ซึ่งวิธีการนี้ได้ปรับให้มีการฝึกอบรมเป็นแบบพร้อมกัน และนำเกณฑ์การพิจารณาเลือกหยุดมาใช้เพื่อลดเวลาที่ใช้หรือจำนวนรอบในการเรียนรู้ให้น้อยลง ซึ่งจากการทดลองเบื้องต้นพบว่า วิธีการที่นำเสนอนี้สามารถให้คุณภาพของคำตอบที่ดีขึ้นและใช้ทรัพยากรน้อยลงได้ในปัญหาส่วนใหญ่

2. หลักเกณฑ์ที่ใช้ในการพิจารณาเลือกหยุดที่กำหนดให้ใช้จำนวนรอบและจำนวนข่ายงานแบบคงที่นั้น เพื่อให้การฝึกอบรมได้คำตอบที่มีความเหมาะสมตามที่ต้องการในการนำไปใช้ในแต่ละงาน ซึ่งสามารถปรับเปลี่ยนค่าของตัวแปรแต่ละตัวให้สอดคล้องได้ โดยพิจารณาได้จากความสัมพันธ์ของตัวแปรต่างๆ ต่อลักษณะของคำตอบที่ได้

3. การกำหนดจำนวนรอบสูงสุดของการฝึกอบรมจากที่ได้นำเสนอ เมื่อมีความต้องการระดับค่าผิดพลาดที่แตกต่างกัน สามารถปรับค่าตัวแปรที่ใช้ในการกำหนดจำนวนรอบสูงสุดให้มีความเหมาะสมเพื่อให้ได้ระดับของคำตอบตามความต้องการ

6.2 ข้อจำกัดและข้อเสนอแนะ

1. การเรียนรู้ของข่ายงานประสาทเทียมจากวิธีการที่ได้นำเสนอนี้ยังคงมีข้อจำกัดในการฝึกอบรมกับบางปัญหาที่ค่าผิดพลาดจากการทดสอบมีลักษณะไม่คงที่หรือมีการเปลี่ยนแปลงค่อนข้างมาก จึงส่งผลให้การพิจารณาเลือกหยุดเกิดการผิดพลาดได้ ดังนั้น การนำไปใช้งานกับปัญหาทั่วไปจึงควรพิจารณาลักษณะของค่าผิดพลาดที่เกิดขึ้นจากการเรียนรู้ควบคู่ไปด้วย

2. การนำวิธีการเรียนรู้ของข่ายงานประสาทเทียมแบบพร้อมกันที่ใช้เกณฑ์การพิจารณาเลือกหยุดไปใช้กับปัญหาอื่นๆ ให้ทำงานได้อย่างมีประสิทธิภาพ ควรกำหนดจำนวนรอบสูงสุดให้มีความเหมาะสม เพื่อให้เวลาหรือจำนวนรอบที่ใช้ในการฝึกอบรมมีค่าต่ำที่สุด

3. ในการกำหนดค่าตัวแปรต่างๆ ที่ใช้ในการเรียนรู้ทั้งจำนวนรอบและจำนวนข่ายงานที่พิจารณาเลือกหยุดจะส่งผลต่อคำตอบที่ได้ ดังนั้น เพื่อให้ได้ผลลัพธ์ของการฝึกอบรมมีความเหมาะสมตามความต้องการในการนำไปใช้งาน ผู้ใช้สามารถกำหนดฟังก์ชันอรรถประโยชน์ขึ้นมาสำหรับการหาค่าตัวแปรที่กำหนดให้ตรงตามวัตถุประสงค์การใช้งาน

4. การนำแนวคิดนี้ไปใช้ต่อไปในการเรียนรู้ของข่ายงานประสาทเทียมหรือวิธีการเรียนรู้ของเครื่องอื่นๆ สามารถประยุกต์ใช้ได้โดยไม่จำเป็นต้องมีการเปลี่ยนแปลงขั้นตอนหลักดังที่เสนอ เพียงแต่เมื่อมีการเปลี่ยนอัลกอริทึมหรือวิธีการที่เรียนรู้อาจต้องมีการศึกษาลักษณะของค่าผิดพลาดที่เกิดขึ้น เพื่อนำมาปรับเปลี่ยนค่าของตัวแปรต่างๆ ให้มีความเหมาะสม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] Poston, T., Lee, C.N., Choie, Y., and Kwon, Y. (1991). Local minima and back propagation. Proceedings of the IEEE International Joint Conference on Neural Networks, pp. 173-176.

[2] Yu, X.-H. (1992). Can backpropagation error surface not have local minima. IEEE Transactions on Neural Networks 3: pp. 1019-1021.

[3] Gori, M., and Tesi, A. (1992). On the problem of local minima in backpropagation. IEEE Transactions on Pattern Analysis and Machine Intelligence 14: pp. 76-86.

[4] Fukumizu, K., and Amari, S. (1999). Local minima and plateaus in multilayer neural networks. Proceedings of the Ninth International Conference on Artificial Neural Networks, pp. 597-602.

[5] Huang, D.-S. (1998). The local minima-free condition of feedforward neural networks for outer-supervised learning. IEEE Transactions on Systems, Man and Cybernetics, Part B 28: pp. 477-480.

[6] Yu, X.-H., and Chen, G.-A. (1995). On the local minima free condition of backpropagation learning. IEEE Transactions on Neural Networks 6: pp. 1300-1303.

[7] Park, Y.R., Murray, T.J., and Chen, C. (1996). Predicting sun spots using a layered perceptron neural network. IEEE Transactions on Neural Networks 7: pp. 501-505.

[8] Iyer, M.S., and Rhinehart, R.R. (1999). A method to determine the required number of neural-network training repetitions. IEEE Transactions on Neural Networks 10: pp. 427-432.

[9] Riedmiller, M., and Braun, H. (1993). A Direct Adaptive Method for Faster Backpropagation Learning: The {RPROP} algorithm. IEEE International Conference on Neural Networks, pp. 586-591. San Francisco, CA.

[10] Sprinkhuizen-Kuyper, I.G., and Boers, E.J.W. (1999). A local minimum for the 2-3-1 XOR network. IEEE Transactions on Neural Networks 10: pp. 968-971.

[11] Cetin, B.C., Burdick, J.W., and Barhen, J. (1993). Global descent replaces gradient descent to avoid local minima problem in learning with artificial neural networks. Proceedings of the IEEE International Conference on Neural Networks, pp. 836-842.

[12] Toh, K.A. (2003). Deterministic global optimization for FNN training. IEEE Transactions on Systems, Man and Cybernetics, Part B 33: pp. 977-983.

[13] Jordanov, I.N., and Rafik, T.A. (2004). Local minima free neural network learning. Proceedings of the 2nd International IEEE Conference Intelligent Systems, pp. 34-39.

[14] Wessels, L.F.A., and Barnard, E. (1992). Avoiding false local minima by proper initialization of connections. IEEE Transactions on Neural Networks 3: pp. 899-905.

[15] Yao, X., and Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks 8: pp. 694-713.

[16] Yao, X. (1999). Evolving artificial neural networks. Proceedings of the IEEE 87: pp. 1423-1447.

[17] Sexton, R.S., and Gupta, J.N.D. (2000). Comparative evaluation of genetic algorithm and backpropagation for training neural networks. Information sciences 129: pp. 45-59.

[18] Cantu-Paz, E., and Kamath, C. (2005). An empirical comparison of combinations of evolutionary algorithms and neural networks for classification problems. IEEE Transactions on Systems, Man and Cybernetics, Part B 35: pp. 915-927.

[19] Ikenoue, S., Asada, M., and Hosoda, K. (2002). Cooperative behavior acquisition by asynchronous policy renewal that enables simultaneous learning in multiagent environment. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and System, pp. 2728-2734.

[20] Takahashi, Y., Edazawa, K., Noma, K., and Asada, M. (2005). Simultaneous learning to acquire competitive behaviors in multi-agent system based on a modular learning system. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2016-2022.

[21] Artac, M., Jogan, M., and Leonardis, A. (2002). Incremental PCA for on-line visual learning and recognition. Proceedings of the 16th International Conference on Pattern Recognition, pp. 781-784.

[22] Declercq, A., and Piater, J.H. (2007). On-line Simultaneous Learning and Tracking of Visual Feature Graphs. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-6.

[23] Liu, Y., and Yao, X. (1999). Simultaneous training of negatively correlated neural networks in an ensemble. IEEE Transactions on Systems, Man and Cybernetics, Part B 29: pp. 716-725.

[24] Prechelt, L. (1994). Proben1: A set of Neural Network Benchmark Problems and Benchmarking Rules. Technical Report 21/94: Fakultat fur Informatik, University of Karlsruhe.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทความทางวิชาการ

ส่วนหนึ่งของงานวิทยานิพนธ์เรื่องการเรียนรู้แบบพร้อมกันที่ใช้เกณฑ์การเลือกหยุดในข่ายงานประสาทเทียม ได้รับการตีพิมพ์เป็นบทความวิชาการในหัวเรื่อง “Avoiding Local Minima in Feedforward Neural Networks by Simultaneous Learning” โดย อัครชัย อรรถกุลเรขา และ ดาริชา สุธีวงศ์ ในงานประชุมวิชาการ “20th Australian Joint Conference on Artificial Intelligence” ซึ่งจัดขึ้นระหว่างวันที่ 4-6 ธันวาคม 2550 ณ เมืองโกลด์โคสต์ รัฐควีนส์แลนด์ ประเทศออสเตรเลีย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Avoiding Local Minima in Feedforward Neural Networks by Simultaneous Learning

Akarachai Atakulreka[1] and Daricha Sutivong[2]

[1] Department of Computer Engineering,
[2] Department of Industrial Engineering,
Faculty of Engineering, Chulalongkorn University,
Pathumwan Road, Bangkok 10330 Thailand
akarachai.a@hotmail.com, daricha.s@chula.ac.th

Abstract. Feedforward neural networks are particularly useful in learning a training dataset without prior knowledge. However, weight adjusting with a gradient descent may result in the local minimum problem. Repeated training with random starting weights is among the popular methods to avoid this problem but it requires extensive computational time. This paper proposes a simultaneous training method with removal criteria to eliminate less promising neural networks, which can decrease the probability of achieving a local minimum while efficiently utilizing resources. The experimental results demonstrate the effectiveness and efficiency of the proposed training method in comparison with conventional training.

Keywords: Local Minima, Simultaneous Learning, Removal Criteria, Feedforward Neural Networks.

1 Introduction

Artificial neural networks (ANNs) are widely used in various applications such as classification, approximation [1], signal processing, and pattern recognition [2]. One of the interesting properties of a neural network is the ability to learn from its environment in order to improve its performance. When learning takes place, the connection weights of a network are adjusted using a training dataset without any prior knowledge. The most common learning method used for supervised learning with feedforward neural networks (FNNs) is backpropagation (BP) algorithm. The BP algorithm calculates the gradient of the network's error with respect to the network's modifiable weights. However, the BP algorithm can result in movement toward the local minimum.

To overcome the local minimum problems, many methods have been proposed. One of them is to train a neural network more than once, starting with a random set of weights [3,4]. An advantage of this approach lies in the simplicity of using and applying to other learning algorithms. Nevertheless, this approach requires more time to train the networks. Therefore we propose simultaneous training along with certain removal criteria. By randomizing starting weights of multiple neural networks, a local minimum is probabilistically avoided. Elimination of less promising neural networks then help save on computational resources. In the experiments, the proposed method

is applied with Resilient Backpropagation (RPROP) [5], which is a fast and computationally economical variant of BP algorithm, for comparison of effectiveness and efficiency with conventional neural networks.

This paper is organized as follows: Section 2 describes the local minimum problem in neural networks. Section 3 explains our proposed algorithm which consists of the training method and removal criteria. Section 4 discusses the experimental procedure. Section 5 compares and analyzes the results. Section 6 shows the comparison with evolving artificial neural networks (EANNs). Finally, section 7 concludes the paper with a summary.

2 Local Minima in Neural Networks

The supervised learning of multilayered neural networks with conventional learning algorithms faces the local minimum problems. Gradient descent type learning algorithms include BP which is a learning algorithm that changes the connection weights of a network to learn via a training set of input-output pairs without any prior knowledge. Using gradient descent to adjust the weights involves following local slope of the error surface which may lead towards some undesirable points, or the local minima.

In this situation, conventional training of neural networks often gets stuck in the local minima. There are several studies [6]-[11] that investigate this problem, by exploring the appearance on the architecture and the learning environment for the local minima-free condition. Different types of local minima are described in [12] in order to understand the behavior of an error surface in the neighborhood of a local minimum and to explain the global behavior of the error surface. In fact, the local minima are mainly associated with two factors: the learning style and the network structure. The methods handling the problem can be based on a deterministic approach or a probabilistic approach.

In a deterministic approach, a new learning algorithm, Global Descent [13], was proposed in place of primary gradient descent rules. Optimization algorithms are applied in the learning process to avoid local minima in [14] and [15]. These methods can assist in finding the global optimum; however, they are rather time-consuming.

Another alternative, a probabilistic approach, often focuses on the set of weights such as the weight initialization method [16], which can decrease the probability of achieving local minima. Neural networks learning, in which training occurs more than once by starting with a random set of weights is another interesting method. The best neural network is often selected as the one with the lowest error. For example, Park et al. [3] proposed a procedure to determine a structure of the neural network model for predicting sun spots. Their best network was selected from among ten networks of the same architecture, each of which was initiated with a different random set of weights. But the number of random starts can not be specified for certain. To avoid restarting too many times, Iyer and Rhinehart [4] took a probabilistic approach to determine the number of random starts needed in the neural network training. Although this approach is easy to implement, it requires extensive time for training.

Another interesting learning paradigm of ANNs which have been widely studied is the evolving artificial neural networks (EANNs). This model is the combination of ANNs and evolutionary algorithms (EAs), which have been applied in three major levels: connection weights, architecture, and learning rules. For instance, EPNet [17] is an automatic system for EANNs that uses mutation operators to modify architectures and weights. A survey of EANNs can be found in Yao [18] and the comparison results are demonstrated in [19] and [20]. This method is effective in discovering global minima; however, it still requires substantial resources.

3 The Proposed Simultaneous Learning Method

The proposed method involves learning of multiple neural networks which is analogous to the concept of repeated training with a random set of weights that help avoiding local minima. However, in this approach, the neural networks learn simultaneously in parallel using multiple initial weights. The method also incorporates removal criteria in order to gradually stop running some neural networks to achieve efficient resource utilization. The following subsections describe the method in more details in two parts: simultaneous learning and removal criteria.

3.1 Simultaneous Learning

The proposed approach applies parallel computing to neural network training, which offers a different training order than the conventional way. The conventional training starts from creating a neural network with a set of initial weights and then train by adjusting the weights. When the number of epochs reaches the max epoch size, in the case of repeated training, a new network will be recreated with a different set of initial weights and trained as before (Fig. 1(a) shows an example procedure).

In the proposed training method, neural networks are trained simultaneously. In order to train a number of neural networks in parallel, all networks will be created each with a random set of weights and trained simultaneously, all starting from the first epoch continuing to the max epoch. When training with multiple core processors, all networks may be trained at the same time. Fig. 1(b) shows an example procedure. The proposed training order allows us to perceive the error trend, as more networks provide better information of the error. Observing a set of errors can assist in determining the max epoch more appropriately. The error trend is then used for comparison among the networks in order to select a poor network to stop training. The removal criteria are introduced in the next subsection.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

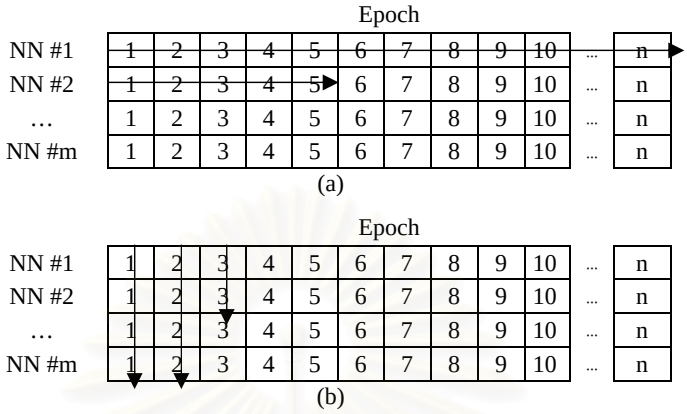


Fig. 1. (a) The conventional training procedure. (b) The simultaneous training procedure. Each slot represents a training epoch. The arrows show the order of the training.

3.2 Removal Criteria

In order to reduce the resource consumption in learning, evaluation criteria for neural network removal are incorporated in the simultaneous training in order to stop the neural networks with poor tendency. Specifically, we utilize certain comparative parameter as a predictor of future error behavior. The network with the worst value will be stopped. The comparative parameters under consideration include the followings.

MinimumError	The smallest error from all previous training epochs
LastGenError	The error of the last training epoch
ErrorSlope	The slope computed by linear regression on the last 50 errors
MixedError	The approximate error estimated from the last generation error scaled down by its error tendency, as shown in equation (1) and Fig. 2

$$MixedError = LastGenError\,(1 - (\arctan(ErrorSlope)\,/(\pi / 2))) \qquad \textbf{(1)}$$

In the experiment, we set the maximum epoch to be the number of the epochs that the error begins to converge, i.e., the error decreases by less than ε or the error starts to increase indicating that overfitting occurs in the problem. We simultaneously train ten networks and consider choosing the network to stop for the first time when the number of the training epoch reaches 10% of the max epoch and subsequently at the interval of 5% of the max epoch. Therefore, nine networks will be eliminated after 50% of the max epoch, and the last network is trained until the max epoch. Fig. 3 shows an example of applying the approach with ten networks.

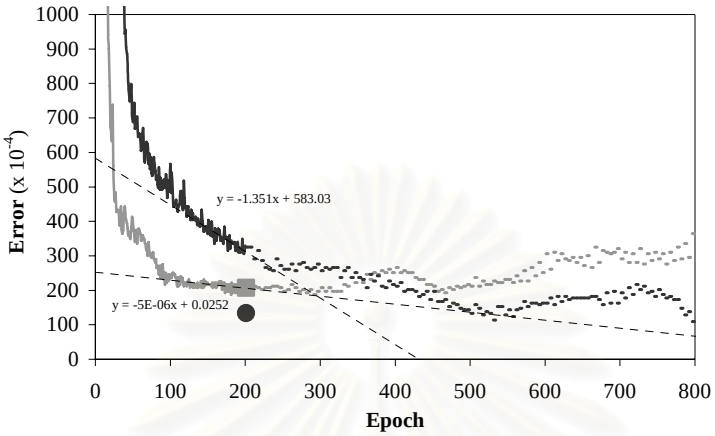


Fig. 2. An example of using the mixed error criterion to evaluate between two networks at 200th epoch. Black line and gray line show the error of each network. Solid lines represent the passed error, and dotted lines represent the error in the future. Each network is computed its error slope by linear regression (dashed line) in order to calculate equation (1). The approximate error of black line and gray line are plotted by using circle dot and square dot, respectively. Even if the error of black line's network at the present epoch is higher than the error of gray line's network, this mixed error criterion can predict the future error behavior correctly.

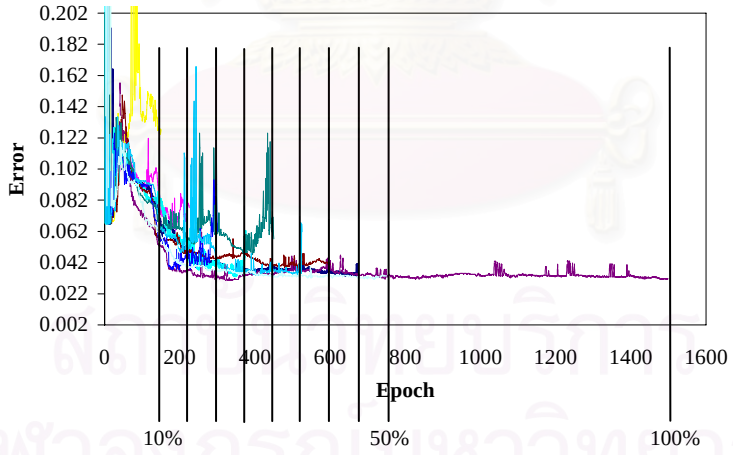


Fig. 3. An example of applying the method with network removal to ten neural networks that are trained simultaneously. The vertical lines indicate the epochs at which one network is stopped. The max epoch (100%) is set where error seems to converge.

4 Experimental details

In our experiment, we apply the proposed method to eight problems. The datasets were obtained from Proben1 [21], which can be categorized into two types; classification and approximation problem. Approximation problems include three testing problems: Building, Flare, and Heart. Classification problems include five testing problems: Card, Mushroom, Gene, Soybean, and Thyroid. Numeric data used in simulations are shown Table 1.

Table 1. Neural network architecture and dataset for testing problems, which include data size, network architecture (input, neurons in hidden layers, output), and the max epoch size.

Problem	Dataset size	Network Architecture			Max epoch size
		Input	Hidden	Output	
Approximation Problem					
Building	4208	14	16	3	800
Flare	1066	24	32	3	200
Heart	920	35	32	1	160
Classification Problem					
Card	690	51	32	2	200
Mushroom	8124	125	32	2	4000
Gene	3175	120	4+2	3	500
Soybean	683	82	16+8	19	1000
Thyroid	7200	21	16+8	3	1500

5 Results and Discussions

The results from all testing problems are used for comparison between the conventional method and the proposed method using various removal criteria. The network architecture is adjusted for each problem. The performances of the two methods are compared using two measures: effectiveness and efficiency.

5.1 Effectiveness

The effectiveness measure compares the solution quality between the two methods using the validation errors from training after the same number of epochs. The number of epochs for both methods is equal to the sum of epochs used by all networks in the simultaneous training. This amount limits the number of epochs used in the conventional training, which repeatedly initializes weights when each cycle reaches the max epoch. The experimental results show the average minimum error after 20 simulations of each method as depicted in Table 2.

Table 2. The effectiveness (solution quality) comparison. The table compares conventional training with simultaneous training using various removal criteria. The numbers show a validation error (x 10^{-3}) followed by a percentage improvement from conventional training as shown in parentheses. The best result for each test problem is highlighted in bold.

Problem	Conventional training	Simultaneous training			
		Minimum	Last Gen.	Slope	Mix
Approximation					
Building	8.154	7.809(4.2%)	**7.796(4.4%)**	8.094(0.7%)	**7.796(4.4%)**
Flare	3.897	**3.801(2.5%)**	3.816(2.1%)	3.807(2.3%)	3.821(2.0%)
Heart	43.47	**42.21(2.9%)**	42.24(2.8%)	42.33(2.6%)	42.44(2.4%)
Classification					
Card	410.1	**382.7(6.7%)**	**382.7(6.7%)**	404.6 (1.3%)	**382.7(6.7%)**
Mushroom	9.872	6.204(37.2%)	6.179(37.4%)	8.198(17.0%)	**6.155(37.7%)**
Gene	201.1	181.2 (9.9%)	**180.3(10.3%)**	197.1(2.0%)	181.3(9.8%)
Soybean	450.3	**434.5(3.5%)**	435.4(3.3%)	445.6(1.0%)	436.0(3.2%)
Thyroid	27.36	**25.75(5.9%)**	26.14(4.5%)	27.47 (-0.4%)	26.31(3.9%)

From Table 2, simultaneous training outperforms conventional training for all test problems. There is no single removal criterion that is best for all problems. Minimum error, last error, and mixed error are effective error predictors in different case, while error slope is never the best choice. An average error over all removal criteria shows an improvement of the proposed approach over conventional training by 2.76-32.30%.

5.2 Efficiency

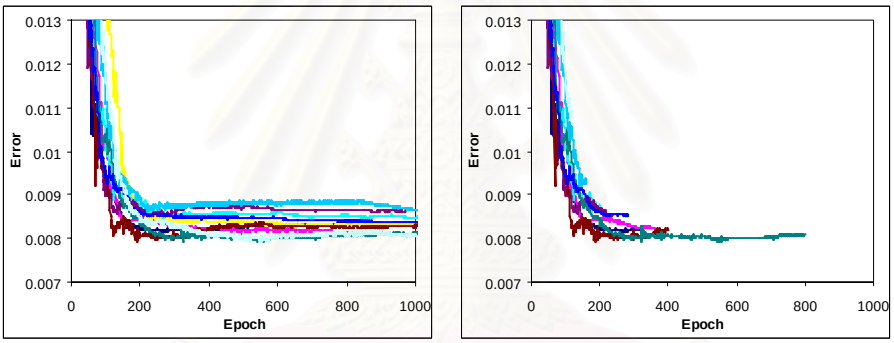
The efficiency measure compares the training time or the number of epochs used in order to obtain a specified error. Table 3 shows the results, which are the sum of the number of training epochs for each problem. For conventional training, the number of epochs is counted from the start of the first network until an error of the training is less than a specified error for a particular problem, but not more than ten networks. Each neural network training has a limitation on the number of epochs at the max epoch size. For simultaneous training, the number of epochs is calculated from the sum of ten networks' training using various removal criteria. The results are an average over 20 simulations.

From Table 3, the proposed approach outperforms the conventional approach in most cases and could lead to a decrease in computational time of up to 53.87%.

Fig. 4 and 5 show the error behaviors of the two methods on Building and Card problems respectively. In Fig. 4(a), errors from different initial weights converge to different values in conventional training. In simultaneous training, a network with poor error tendency will be cut off at each epoch as shown in Fig. 4(b). Fig. 5 displays error behavior when overfitting problem occurs but a good solution is finally found.

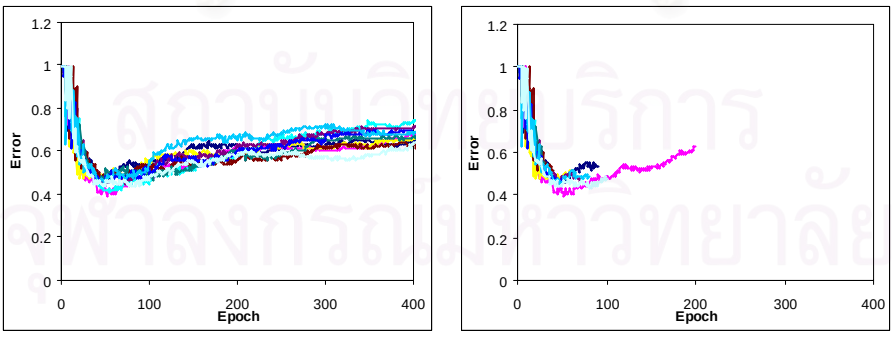
Table 3. The efficiency (computation time) comparison. The table presents an average number of epochs use followed by a percentage improvement from conventional training as shown in parentheses. The best result for each test problem is highlighted in bold.

Problem	Conventional training	Simultaneous training			
		Minimum	Last Gen.	Slope	Mix
Approximation					
Building	3695.85	2386.3(35.4%)	**2361.6(36.1%)**	2723.6(26.3%)	**2361.6(36.1%)**
Flare	540.55	**402.5(25.6%)**	427.3(21.0%)	406.6(24.8%)	436.2(19.3%)
Heart	784.00	**361.7(53.9%)**	376.1(52.0%)	370.6(52.7%)	387.0(50.6%)
Classification					
Card	782.05	**417.1(46.7%)**	417.1(46.7%)	612.3(21.7%)	417.2(46.7%)
Mushroom	**2632.70**	3034.0(-15.2%)	3034.0(-15.2%)	3034.0(-15.2%)	3034.0(-15.2%)
Gene	2862.90	**1627.7(43.2%)**	**1627.7(43.2%)**	1682.2(41.2%)	1635.6(42.9%)
Soybean	4465.25	**2825.5(36.7%)**	2922.0(34.6%)	3176.6(28.9%)	2922.0(34.6%)
Thyroid	4677.00	**4578.3(2.1%)**	4598.0(1.7%)	5026.9(-7.5%)	4598.0(1.7%)



(a) (b)

Fig. 4. Comparison results of the learning process on the Building dataset. (a) The conventional training procedure. (b) The simultaneous training procedure with mixed error removal criteria.



(a) (b)

Fig. 5. Comparison results of the learning process on the Card dataset. (a) The conventional training procedure. (b) The simultaneous training procedure with minimum error removal criteria.

Note that among all removal criteria, a minimum error seems to be the best indicator, as it performs best 5 out of 8 in the effectiveness comparison and 6 out of 8 in the efficiency comparison. In most problems, the proposed method can improve resource utilization and performance in neural network training.

6 Comparison with EANNs

In [17], EPNet is an automatic system for EANNs that uses mutation operators to modify architectures and weights. The experimental results showed that this method could improve the generalization ability. The numerical result for the thyroid problem is shown in the last row of Table 4. We apply the conventional approach and our proposed approach to thyroid problem for comparison with the EANNs approach, and the results are shown in Table 4.

Table 4. Comparison results of the number of epochs, error and standard deviation of training methods trained with the thyroid problem. The total time, the number of epochs, used by each training method is the maximal number of epochs. The average testing error and standard deviation also present in the table.

Method	Number of epochs	Error	SD
Conventional training	2000	0.03033	0.00213
Simultaneous training	7400	0.02575	0.00312
EPNet	109000	0.02115	0.00220

From comparisons in Table 4, EPNet outperforms in achieving the lowest error and reduces the conventional training's error by nearly one third. However, this approach and its accuracy come at the cost of additional computation time in order to perform search. Our proposed simultaneous training uses about the number of epochs less than four times more than the conventional training but much fewer than EPNet to achieve reasoned improvement in accuracy. However, this is one of the comparisons between the proposed method and the EANN. In the future, we will investigate comparison on more problems.

7 Conclusions

In this paper, we propose a simultaneous training method with removal criteria for neural network training to avoid local minima common in many problems. We experiment on multiple testing problems. Results demonstrate that the proposed method can decrease the probability of achieving local minima while utilizing resources efficiently.

For time-critical application, the proposed method can help achieve better results using limited computation times. Moreover, this method can be practicable in the parallel computing environment.

References

1. Ferrari, S., Stengel, R.F.: Smooth function approximation using neural networks. In: IEEE Transactions on Neural Networks, vol. 16, pp. 24--38 (2005)
2. Bishop, C.M.: Neural Networks for Pattern Recognition. Clarendon Press, Oxford (1995)
3. Park, Y.R., Murray, T.J., Chung, C.: Predicting sun spots using a layered perceptron neural network. In: IEEE Transactions on Neural Networks, vol. 7, pp. 501--505 (1996)
4. Iyer, M.S., Rhinehart, R.R.: A method to determine the required number of neural-network training repetitions. In: IEEE Transactions on Neural Networks, vol. 10, pp. 427--432 (1999)
5. Riedmiller, M., Braun, H.: A Direct Adaptive Method for Faster Backpropagation Learning: The {RPROP} algorithm. In: Proceedings of the IEEE International Conference on Neural Networks, pp. 586--591 (1993)
6. Poston, T., Lee, C.N., Choie, Y., Kwon, Y.: Local minima and back propagation. In: IJCNN-91-Seattle International Joint Conference on Neural Networks, vol. 2, pp. 173--176 (1991)
7. Yu, X.H.: Can backpropagation error surface not have local minima. In: IEEE Transactions on Neural Networks, vol. 3, pp. 1019--1021 (1992)
8. Gori, M., Tesi, A.: On the problem of local minima in backpropagation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, pp. 76--86 (1992)
9. Xiao-Hu, Y., Guo-An, C.: On the local minima free condition of backpropagation learning. In: IEEE Transactions on Neural Networks, vol. 6, pp. 1300--1303 (1995)
10. Fukumizu, K., and Amari, S.: Local minima and plateaus in multilayer neural networks. In: 9th International Conference on Artificial Neural Networks, vol. 2, pp. 597--602 (1999)
11. De-Shuang, H.: The local minima-free condition of feedforward neural networks for outer-supervised learning. In: IEEE Transactions on Systems, Man and Cybernetics, part B, vol. 28, pp. 477--480 (1998)
12. Sprinkhuizen-Kuyper, I.G., Boers, E.J.W.: A local minimum for the 2-3-1 XOR network. In: IEEE Transactions on Neural Networks, vol. 10, pp. 968--971 (1999)
13. Cetin, B.C., Burdick, J.W., Barhen, J.: Global descent replaces gradient descent to avoid local minima problem in learning with artificial neural networks. In: IEEE International Conference on Neural Networks, vol. 2, pp. 836--842 (1993)
14. Toh, K.A.: Deterministic global optimization for FNN training. In: IEEE Transactions on Systems, Man and Cybernetics, part B, vol. 33, pp. 977--983 (2003)
15. Jordanov, I.N., Rafik, T.A.: Local minima free neural network learning. In: 2nd International IEEE Conference on Intelligent Systems, vol.1, pp. 34--39 (2004)
16. Wessels, L.F.A., Barnard, E.: Avoiding false local minima by proper initialization of connections. In: IEEE Transactions on Neural Networks, vol. 3, pp. 899--905 (1992)
17. Yao, X., Liu, Yong.: A new evolutionary system for evolving artificial neural networks. In: IEEE Transactions on Neural Networks, vol. 8, pp. 694--713 (1997)
18. Yao, X.: Evolving artificial neural networks. In: Proceedings of the IEEE, vol. 87, pp. 1423--1447 (1999)
19. Sexton, R.S., Gupta, J.N.D.: Comparative evaluation of genetic algorithm and backpropagation for training neural networks. In: Information sciences, vol. 129, pp. 45--59 (2000)
20. Cantu-Paz, E., Kamath, C.: An empirical comparison of combinations of evolutionary algorithms and neural networks for classification problems. In: IEEE Transactions on Systems, Man and Cybernetics, part B, vol. 35, pp. 915--927 (2005)
21. Prechelt, L.: Proben1: A set of Neural Network Benchmark Problems and Benchmarking Rules. Technical report 21/94, Global Grid Forum (1994)

ประวัติผู้เขียนวิทยานิพนธ์

นายอัครชัย อรรถกุลเรขา เกิดเมื่อวันที่ 3 พฤษภาคม พ.ศ. 2527 ที่จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาระดับมัธยมศึกษาจากโรงเรียนสวนกุหลาบวิทยาลัย สำเร็จการศึกษาปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2548 และเข้าศึกษาในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2549

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย