

การประสานคำตอบจากการเลือกเครื่องจักรเรียนรู้เอ็กซ์ทรีม โดยพิจารณาค่าความผิดพลาดเฉลี่ย

Ensemble of Selective Extreme Learning Machine base on Mean Error

ขจรศักดิ์ แสงโทโพธิ์ (Khachonsak Saengthopho)* สิริภัทร เชื้อชาญวัฒนา (Sirapat Chiewchanwattana)**

คำรณ สุณัติ (Khamron Sunat)***

บทคัดย่อ

บทความนี้ได้นำเสนอวิธีการเพิ่มความถูกต้องของขั้นตอนวิธีการจำแนกข้อมูลเครื่องจักรเรียนรู้เอ็กซ์ทรีม (Extreme Learning Machine : ELM) โดยใช้การประสานคำตอบจาก ELM หลายๆตัวที่ทำงานพร้อมกันเพื่อแก้ปัญหาใดปัญหาหนึ่ง ซึ่งขั้นตอนการทำงานหลักๆมี 2 ขั้นตอนคือ 1) คัดเลือก ELM ที่มีค่าผิดพลาดจากการสอนน้อยกว่าค่าเฉลี่ยความผิดพลาดของ ELM ทุกตัว 2) นำ ELM ที่ผ่านการคัดเลือกจากขั้นตอนที่ 1 มาพยากรณ์ข้อมูลทดสอบ แล้วเลือกคำตอบโดยการลงคะแนนเสียง ซึ่งเรียกขั้นตอนทั้ง 2 วิธีนี้ว่า การประสานคำตอบจากการเลือกเครื่องจักรเรียนรู้เอ็กซ์ทรีมโดยพิจารณาค่าความผิดพลาดเฉลี่ย (Ensemble of Selective Extreme Learning Machine base on Mean Error : ES-ME) โดยในการวิจัยใช้ชุดข้อมูลจริงจำนวน 14 ชุดข้อมูลจาก UC Irvine Machine Learning Repository เปรียบเทียบระหว่างขั้นตอนวิธี ES-ME และขั้นตอนวิธี ELM แบบเดี่ยว พบว่า ขั้นตอนวิธี ES-ME สามารถเพิ่มความถูกต้องและน่าเชื่อถือของการพยากรณ์ให้เพิ่มมากขึ้นกว่าการใช้ขั้นตอนวิธี ELM แบบเดี่ยวทุกชุดข้อมูล โดยให้ความถูกต้องของคำตอบเพิ่มขึ้นโดยเฉลี่ย 2.24% และให้ค่าความเบี่ยงเบนมาตรฐานลดลงโดยเฉลี่ย 0.561% รวมทั้งให้ความถูกต้องสูงที่สุดเพิ่มขึ้นโดยเฉลี่ย 1.64%

ABSTRACT

This paper introduces the method for improve accuracy of Extreme Learning Machines (ELM) by using ensemble multi ELM to solve one of problem. The main procedure will work on two steps, 1) select ELM contained lower mean learning error. 2) predict test data by ELM passed mean from 1) then select an answer by voting. This procedure is Ensemble of Selective Extreme Learning Machine base on Mean Error (ES-ME). The experimental in this paper used 14 real datasets from UC Irvine Machine Learning Repository to compared between propose algorithm and single ELM algorithm. In the results, the propose algorithm contained more accuracy and reliable. The result increased average accuracy by 2.24%, decreased standard derivation by 0.561% and increased average max accuracy by 1.64%.

คำสำคัญ: เครื่องจักรเรียนรู้เอ็กซ์ทรีม การประสานคำตอบ ความผิดพลาดเฉลี่ย

Key Words: Extreme learning machine, Ensemble, Mean error

* นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

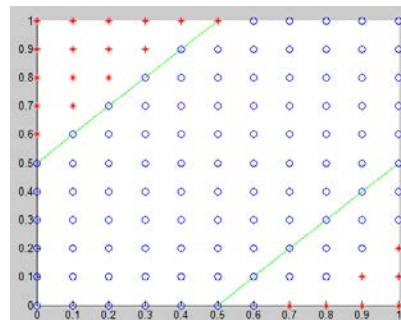
** ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

*** อาจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

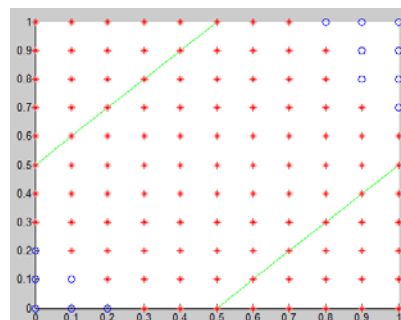
บทนำ

ปัจจุบันมีการนำข้อมูลต่างๆมาวิเคราะห์เพื่อแก้ปัญหาหรือดึงเอาองค์ความรู้ที่ได้มาใช้ประโยชน์ต่างๆอย่างมากมาย โดยเทคนิคที่นิยมนำมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลเหล่านี้คือเทคนิคโครงข่ายประสาทเทียม เนื่องจากมีความสามารถในการเรียนรู้การจดจำรูปแบบ (Pattern Recognition) (Bishop, 2006) และการอนุมานความรู้ (Knowledge deduction) (Barstow, 1979) เช่นเดียวกับที่มีในสมองมนุษย์ แต่มีข้อเสียคือมีการคำนวณที่ช้า มีความซับซ้อนมากและใช้เวลาคำนวณนาน (Huang, Wang and Lan, 2011) ดังนั้นจึงมีนักวิจัยหลายท่านได้เสนอขั้นตอนการฝึกสอนโครงข่ายประสาทเทียมโดยใช้การคำนวณเมทริกซ์ช่วยในการแก้สมการ และงานที่น่าสนใจ คือ ขั้นตอนวิธีเอ็กซ์ทรีม (Extreme Learning Machine) เรียกย่อว่า ELM (Huang, Wang and Lan, 2011) ซึ่งเป็นโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า (Single Layer Feed Forward : SLFN) ที่มีความเร็วในการคำนวณสูงกว่าขั้นตอนวิธีฝึกสอนข่ายงานประสาทเทียมแบบดั้งเดิมมาก พร้อมทั้งมีงานวิจัยอื่นๆที่ออกมาเพื่อหาขั้นตอนวิธีที่ช่วยเพิ่มทั้งความเร็วและประสิทธิภาพของ ELM อย่างต่อเนื่อง แต่ขั้นตอนวิธี ELM โดยทั่วไปยังมีปัญหาคือผลลัพธ์จากการสอนมีความแตกต่างกันมากในการพยากรณ์แต่ละครั้ง (Huang, Zhu and Siew, 2006) เป็นผลมาจากขั้นตอนวิธี ELM ได้ทำการสุ่มค่าน้ำหนักระหว่างชั้นนำเข้า (Input layer) และ ชั้นซ่อน (Hidden layer) รวมทั้งสุ่มค่าไบแอส (Bias) ของโหนดซ่อน (Hidden node) ด้วย ดังนั้นความถูกต้องของการฝึกสอนข้อมูลแต่ละครั้งจึงขึ้นอยู่กับความเหมาะสมของค่าที่สุ่มขึ้นมาได้

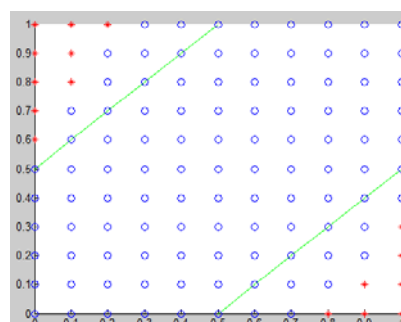
ซึ่งปัญหาของ ELM ที่มีผลการพยากรณ์ที่ไม่แน่นอนนั้นสามารถแสดงให้เห็นได้จากการทดลองเปรียบเทียบการทำงานของ ELM แบบเดี่ยว ในปัญหา XOR โดยการทดลองนี้ได้กำหนดโหนดชั้นซ่อนเป็นจำนวน 10 โหนด แล้วทำการทดลอง 3 ครั้ง



ภาพที่ 1 ผลการพยากรณ์ในปัญหา XOR ครั้งที่ 1



ภาพที่ 2 ผลการพยากรณ์ในปัญหา XOR ครั้งที่ 2



ภาพที่ 3 ผลการพยากรณ์ในปัญหา XOR ครั้งที่ 3

จากภาพที่ 1, 2 และ 3 แสดงให้เห็นถึงความแตกต่างกันมากในการพยากรณ์คำตอบของ ELM ในแต่ละครั้ง ซึ่งมีข้อมูลในการสอนเหมือนกันดังนี้คือ

$$\text{Data}_{\text{Train}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \text{Target} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

ส่วนข้อมูลที่ใช้ในการทดสอบคือค่าที่ฝึกของแกน X และแกน Y ตรงจุดที่แสดงผลการพยากรณ์ซึ่งมีข้อมูลในการทดสอบทั้งหมด 100 รายการ โดยเริ่มจาก [0,0] จนถึง [1,1] ดังนี้

[0,0], [0,0.1], [0,0.2] ,..., [0,1],
[0.1,0], [0.1,0.1], [0.1,0.2] ,..., [0.1,1],
| | | |
[1,0], [1,0.1], [1,0.2] ,..., [1,1]

จากภาพที่ 1, 2 และ 3 แสดงผลการพยากรณ์ที่ได้จากตัวแบบ ELM ของข้อมูลทดสอบแต่ละตัวโดย 0 คือผลจากการพยากรณ์เป็น 0 และ * คือผลจากการพยากรณ์เป็น 1 ส่วนเส้นคั่นจะบอกถึงขอบเขตเป้าหมายของข้อมูลว่าเป็น 0 หรือเป็น 1 โดยข้อมูลที่อยู่ระหว่างเส้นจะมีเป้าหมายเป็น 0 หรือ 0 ส่วนข้อมูลที่อยู่นอกเส้นจะมีเป้าหมายเป็น * หรือ 1

จากการทดสอบทั้ง 3 รอบ โดยมีข้อมูลสอนและข้อมูลทดสอบชุดเดียวกัน สามารถแสดงให้เห็นได้ว่าข้อมูลที่ได้จากการพยากรณ์ในภาพที่ 1 และ 3 มีความถูกต้องมากกว่าข้อมูลที่ได้จากการพยากรณ์ในภาพที่ 2 มาก ซึ่งทำให้ทราบได้ว่าการพยากรณ์แต่ละครั้งของ ELM จะให้ผลการพยากรณ์ที่แตกต่างกันมาก หากใช้ ELM เพียงตัวเดียว จะไม่สามารถทราบได้ว่าตัวแบบพยากรณ์ที่ได้เป็นตัวอย่างที่ดีแล้วหรือไม่

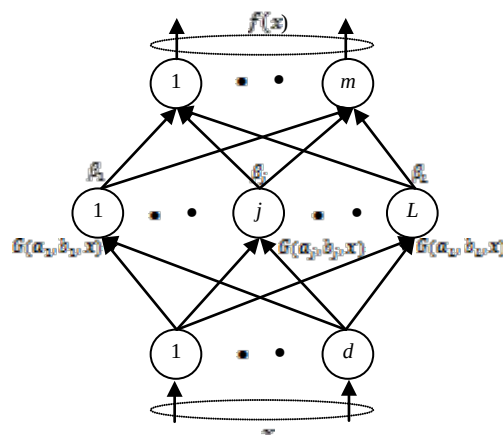
งานวิจัยนี้จึงมีแนวคิดที่จะพัฒนาขั้นตอนวิธีการประสานคำตอบจากผลการพยากรณ์ที่ได้จากตัวแบบ ELM หลายๆตัวที่ทำงานพร้อมกันเพื่อแก้ปัญหาใดปัญหาหนึ่ง โดยพิจารณาว่าควรใช้ ELM ตัวใดในการพยากรณ์คำตอบ เพื่อเพิ่มความถูกต้องของการพยากรณ์ให้มากขึ้นกว่าการใช้ ELM แบบเดี่ยว

ทฤษฎีที่เกี่ยวข้อง

1. Extreme Learning Machine (ELM)

Extreme learning machine (ELM) (Huang, Wang and Lan, 2011) เป็นขั้นตอนวิธีที่ได้รับความนิยมจากนักวิจัย ซึ่ง ELM นี้จะใช้งานได้กับสถาปัตยกรรมโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าที่มีโหนดชั้นซ่อนเพียงชั้นเดียว (Single-hidden layer feedforward networks : SLFNs) และที่สำคัญ

ELM ไม่ต้องการปรับค่าน้ำหนักในโหนดชั้นซ่อนของ SLFNs และเมื่อเปรียบเทียบกับเทคนิคการคำนวณแบบดั้งเดิมแล้ว ELM จะมีความเร็วสูงกว่าหลายเท่า (ปัญญาพล, 2008)



ภาพที่ 4 โครงสร้างของ โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าที่มีโหนดชั้นซ่อนเพียงชั้นเดียว (Single-hidden layer feedforward networks)

จากภาพที่ 4 สำหรับตัวอย่างที่ใช้สอนจำนวน N รายการ $\{(x_i, t_i)\}_{i=1}^N$ เมื่อ $x_i \in \mathbb{R}^d$ และ $t_i \in \mathbb{R}^m$ แล้ว เอาท์พุทที่ได้จาก SLFNs ที่มีจำนวนโหนดในชั้นซ่อนเท่ากับ L โหนด สามารถเขียนเป็นสมการได้คือ

$$o_i = \sum_{j=1}^L \beta_j G(a_j, b_j, x_i), \quad i = 1, 2, \dots, N \quad (1)$$

เมื่อ $a_j \in \mathbb{R}^d$ และ $b_j \in \mathbb{R}^m (j = 1, 2, \dots, L)$

คือพารามิเตอร์เรียนรู้ ของโหนดซ่อนลำดับที่ j และ $\beta_j \in \mathbb{R}^m$ คือ เส้นเชื่อมระหว่างโหนดซ่อนลำดับที่ j กับโหนดผลลัพธ์ โดย $G(a_j, b_j, x_i)$ คือผลลัพธ์ที่ได้จากโหนดซ่อนลำดับที่ j ที่เกิดจากการป้อนข้อมูลตัวอย่าง x_i

จากตัวอย่างข้อมูล N รายการ สามารถเทียบเคียงแทนสมการที่ (1) ได้ว่า

$$O = H\beta \quad (2)$$

เมื่อ $H_{ij} = G(a_j, b_j, x_i)$ ถูกแทนด้วยข้อมูลจำนวน i แถว และ j คอลัมน์ ของผลลัพธ์จากเมทริกซ์

H ในโหนดชั้นซ่อน และ $\beta = (\beta_1, \beta_2, \dots, \beta_L)$ และ

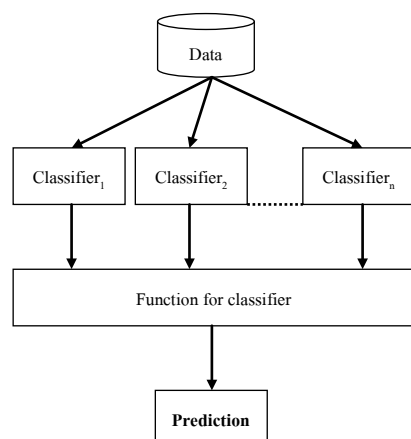
$$O = (o_1, o_2, \dots, o_M)$$

เพื่อให้ค่าเครือข่ายของฟังก์ชันต้นทุน (Network cost function) $\|O - T\|$ น้อยที่สุด เมื่อ $T = (t_1, t_2, \dots, t_M)$ คือ เมทริกซ์เป้าหมายของเอาต์พุต โดยทฤษฎีของ ELM ได้กล่าวไว้ว่า พารามิเตอร์ของ โหนดชั้นซ่อน a_j และ b_j สามารถกำหนดค่าได้จากการสุ่มโดยไม่ต้องพิจารณาข้อมูลนำเข้า ดังนั้น ระบบสมการที่ (1) จะกลายเป็นโมเดลเชิงเส้น (Linear model) และ คำนวณน้ำหนักของเอาต์พุตสามารถวิเคราะห์โดยหาค่าจากผลลัพธ์ของกำลังสองน้อยสุดของระบบเชิงเส้นดังสมการต่อไปนี้

$$\beta = H^+ T \quad (3)$$

เมื่อ H^+ คือ การหาเมทริกซ์ผกผันแบบมอร์เพนโรส (Moore-Penrose)

ซึ่งขั้นตอนวิธีของ ELM นี้จะเป็นการหาเมทริกซ์ผกผันแบบมอร์เพนโรส เพื่อแก้ไขสมการที่มี m ตัวแปรและมี n สมการ (ปัญญาพล, 2008) ซึ่งช่วยลดเวลาในการฝึกสอนข่ายงานประสาทเทียมได้เป็นอย่างมาก ปัจจัยที่ช่วยลดเวลาฝึกสอนของเมทริกซ์ผกผันคือการแก้สมการ $Ax = b$ โดยในขั้นตอนวิธี ELM นี้ได้มีการนำการคำนวณเมทริกซ์ผกผันมอร์เพนโรสมาประยุกต์ใช้งาน



ภาพที่ 5 โครงสร้างการทำงานของ การประสม คำตอบ

2. การประสมคำตอบ (Ensemble)

การประสมคำตอบเป็นการเอาคำตอบจากตัวจำแนกหลายๆตัวมารวมกัน เพื่อให้ได้การพยากรณ์ที่มีประสิทธิภาพ ซึ่งตัวจำแนกข้อมูลแต่ละตัวต้องมีการทำงานเป็นของตัวเอง และเมื่อได้ผลจากการทำงานข้างต้นแล้วก็จะนำผลจากการจำแนกทั้งหมดมาประสมกัน โดยการรวมคำตอบที่ได้จากแต่ละตัวจำแนกข้อมูล เพื่อหาตัวตัดสินใจที่สามารถจำแนกข้อมูลได้มีประสิทธิภาพมากที่สุดดังภาพที่ 5

ในการรวมคำตอบที่ได้นั้นจะต้องประกอบด้วย คำตอบจากตัวแยกแยะข้อมูลหลายๆตัว และแต่ละตัวต้องเป็นอิสระต่อกัน (Zhou and Jiang, 2003) ซึ่งถ้าแต่ละตัวมีการแยกแยะข้อมูลที่ดีและมีจำนวนมากเท่าใดก็จะมีผลต่อความถูกต้องที่มากขึ้นเท่านั้น ในทางตรงกันข้าม ถ้าตัวแยกแยะข้อมูลมีการทำงานผิดพลาดซ้ำๆกัน เมื่อนำมารวมคำตอบด้วยตัวประสมคำตอบแล้ว ผลที่ได้ก็จะมีความผิดพลาดมากด้วยเช่นกัน

เนื่องจากตัวจำแนกข้อมูลจะมีประสิทธิภาพมากน้อยนั้นขึ้นอยู่กับความยากง่ายของข้อมูลที่นำมาจำแนก ดังนั้นจึงมีการหาขั้นตอนวิธีต่างๆมาช่วยในการจำแนกข้อมูล เพื่อให้การจำแนกข้อมูลมีประสิทธิภาพมากขึ้น ซึ่งตัวจำแนกข้อมูลที่ทำาร่วมกันหรือที่เรียกว่า การประสมคำตอบของการจำแนกข้อมูล เป็นอีกทางเลือกหนึ่งที่มีประสิทธิภาพดีกว่าการใช้ตัวจำแนกข้อมูลเพียงตัวเดียว

ชุดข้อมูลและการจัดการชุดข้อมูลในการวิจัย

ในการวิจัยครั้งนี้ใช้ข้อมูลทั้งหมด 14 ข้อมูลดังตารางที่ 1 ซึ่งสามารถดาวน์โหลดได้จากเว็บไซต์ <http://archive.ics.uci.edu/ml/> โดยข้อมูล 9 ข้อมูลแรกประกอบด้วยข้อมูล Glass, Breast Cancer, Balance, Diabetes, Statlog, Contraceptive Method Choice, Yeast, Wine Quality และ Waveform จะใช้การแบ่งข้อมูลในการสอนและทดสอบโดยการเลือกสุ่มแบบความเที่ยงตรง K กลุ่ม (K-Fold Cross Validation) ซึ่ง

ในการทดลองได้กำหนดการเลือกสุ่มเป็น 10 Fold ส่วนข้อมูล 5 ข้อมูลหลังประกอบด้วยข้อมูล Hayes, Monk1, Monk2, Monk3 และ Digits จะเป็นข้อมูลแบบกำหนดส่วนที่ใช้สอนและทดสอบตายตัว โดยตัวเลขชุดแรกในช่อง #Instances คือจำนวนข้อมูลที่ใช้สอนและตัวเลขชุดหลังในช่อง #Instances คือจำนวนข้อมูลที่ใช้ทดสอบ

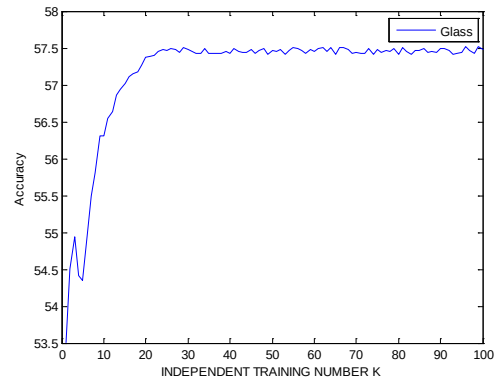
ตารางที่ 1 ข้อมูลและลักษณะข้อมูลที่ใช้วิจัย

Datasets	#Attributes	#Classes	#Instances
Glass	10	6	214
Breast Cancer	30	2	569
Balance	4	3	625
Diabetes	8	2	768
Statlog	20	2	1000
Contraceptive Method Choice	9	3	1473
Yeast	8	10	1484
Wine Quality	12	7	4898
Waveform	21	3	5000
Hayes	4	3	132/28
Monk1	6	2	124/432
Monk2	6	2	169/432
Monk3	6	2	122/432
Digits	16	10	7494/3498

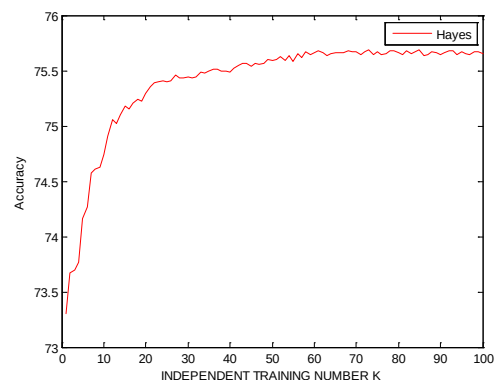
วิธีการวิจัย

ในการวิจัยได้นำข้อมูลแต่ละชุดมาเข้าตัวแบบการพยากรณ์ทั้ง 2 แบบคือ ELM แบบเดี่ยว และ ES-ME โดยผลการทดลองได้จากการหาค่าเฉลี่ยจากการทดลองจำนวน 30 ครั้ง ในแต่ละโหนดชั้นซ่อน จากโหนดชั้นซ่อนจำนวน 1 โหนด จนถึงโหนดชั้นซ่อนจำนวน 200 โหนด โดยใช้ฟังก์ชันกระตุ้นของ ELM แบบเดี่ยว และ ES-ME เป็นฟังก์ชันซิกมอยด์ (sigmoid function) ดังสมการที่ 4

$$G(a, b, x) = \frac{1}{(1 + \exp(-(a \cdot x + b)))} \quad (4)$$



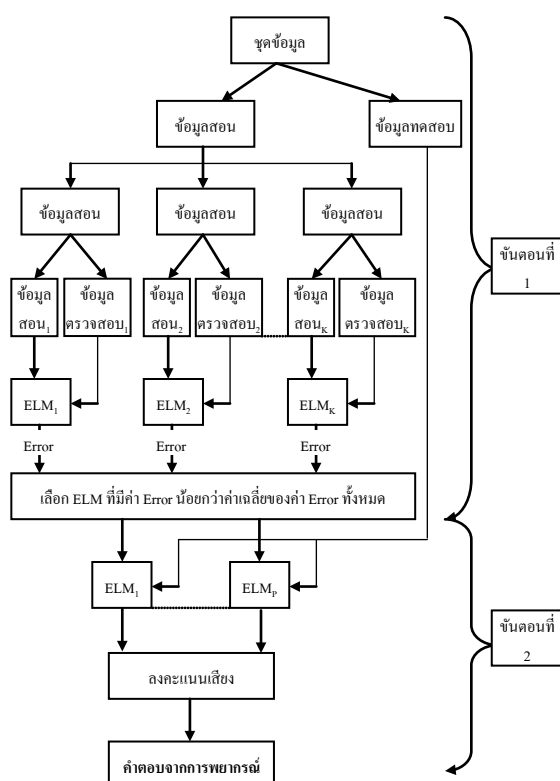
ภาพที่ 6 กราฟค่าเฉลี่ยของความถูกต้องของข้อมูล Glass ในแต่ละจำนวน ELM ที่ใช้ในขั้นตอนวิธี ES-ME



ภาพที่ 7 กราฟค่าเฉลี่ยของความถูกต้องของข้อมูล Hayes ในแต่ละจำนวน ELM ที่ใช้ในขั้นตอนวิธี ES-ME

ซึ่งในการทดลองได้กำหนดจำนวน ELM ที่ใช้ในขั้นตอนวิธี ES-ME เป็น 21 ตัว เนื่องจากการทดลองกำหนดจำนวน ELM ที่ใช้ในขั้นตอนวิธี ES-ME เป็น 1 - 100 ตัว ทดลองกับข้อมูล Glass และ ข้อมูล Hayes ดังภาพที่ 6 และ 7 แสดงให้เห็นว่าจำนวน ELM ในขั้นตอนวิธี ES-ME ที่เหมาะสมกับข้อมูลทั้งสองชุดอยู่ระหว่าง 18-24 ตัว โดยเมื่อกำหนดจำนวน ELM ในขั้นตอนวิธี ES-ME ให้ต่ำกว่า 18 ตัว จะส่งผลให้ค่า

ความถูกต้องในการพยากรณ์มีแนวโน้มลดลงมาก ซึ่งจะทำให้การพยากรณ์มีความถูกต้องต่ำ และเมื่อกำหนดจำนวน ELM ในขั้นตอนวิธี ES-ME ให้มากกว่า 24 ตัว จะส่งผลให้ค่าความถูกต้องในการพยากรณ์มีแนวโน้มเพิ่มขึ้นน้อย ซึ่งจะทำให้เสียเวลามากขึ้นโดยได้ค่าความถูกต้องเพิ่มขึ้นเพียงเล็กน้อย จึงเลือกใช้จำนวน ELM ในขั้นตอนวิธี ES-ME เป็นจำนวน 21 ตัว



ภาพที่ 8 แผนภาพการทำงานของขั้นตอนวิธี ES-ME

จากภาพที่ 8 สามารถอธิบายขั้นตอนการทำงานของขั้นตอนวิธี ES-ME ได้ดังนี้

ขั้นตอนที่ 1 คัดเลือก ELM ที่มีค่าความผิดพลาด (Error) จากการสอนน้อยกว่าค่าความผิดพลาดเฉลี่ยโดย

1.1 นำชุดข้อมูลที่ต้องการทดลองมาแบ่งข้อมูลออกเป็น 2 ชุด คือ ข้อมูลสอนและข้อมูลทดสอบเมื่อได้ข้อมูลสอนแล้วจะนำส่วนหนึ่งของข้อมูลสอนมาเป็นข้อมูลตรวจสอบแล้วจึงส่งทั้งข้อมูลสอนและข้อมูลตรวจสอบให้ ELM แต่ละตัว

1.2 เมื่อได้ข้อมูลสอน ELM แต่ละตัวจะเริ่มขั้นตอนการเรียนรู้ของตัวเอง แล้วจึงส่งข้อมูลตรวจสอบให้ ELM แต่ละตัวเพื่อพยากรณ์คำตอบพร้อมทั้งหาค่าความผิดพลาดโดยเปรียบเทียบจากค่าที่ได้จากการพยากรณ์กับเป้าหมายของข้อมูลตรวจสอบแล้วส่งออกมาเป็นค่า Error

1.3 นำค่า Error ของ ELM แต่ละตัวมาหาค่าเฉลี่ยแล้วคัดเลือกเฉพาะ ELM ที่มีค่าผิดพลาดน้อยกว่าค่าผิดพลาดเฉลี่ยไปใช้ในขั้นตอนที่ 2

ขั้นตอนที่ 2 นำโครงสร้างของ ELM ที่ได้จากขั้นตอนที่ 1 ทุกตัวมาพยากรณ์ข้อมูลทดสอบและเลือกคำตอบโดยการลงคะแนนเสียง

การวัดประสิทธิภาพ

กำหนดให้

L คือ ขอบเขตของจำนวนโหนดชั้นซ่อนที่กำหนด โดยในการทดลองนี้กำหนดเป็น 200

C คือ จำนวนรอบในการทดลองแต่ละโหนดชั้นซ่อน โดยในการทดลองนี้กำหนดเป็น 30

ในการวัดประสิทธิภาพเปรียบเทียบระหว่างขั้นตอนวิธี ES-ME กับขั้นตอนวิธี ELM แบบเดียวจะใช้ค่าดังต่อไปนี้

1. ค่าเฉลี่ยของความถูกต้อง (Average accuracy) คือ ค่าที่ได้จากการนำค่าเฉลี่ยของความถูกต้องในโหนดชั้นซ่อนจำนวน 1 โหนด จนถึงโหนดชั้นซ่อนจำนวน 200 โหนด มาหาค่าเฉลี่ยดังสมการต่อไปนี้

$$\text{Average accuracy} = \frac{1}{L} \sum_{i=1}^L \left(\frac{1}{C} \sum_{j=1}^C \text{Accuracy}_{ij} \right) \quad (5)$$

โดยค่าเฉลี่ยของความถูกต้องบอกถึงความสามารถในการจำแนกข้อมูลของตัวจำแนกข้อมูล ยิ่งค่าเฉลี่ยของความถูกต้องสูง แสดงว่าตัวจำแนกข้อมูลสามารถจำแนกข้อมูลได้ดี

2. ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐาน (Average SD) คือ ค่าที่ได้จากการนำค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานในโหนดชั้นซ่อนจำนวน 1 โหนด

จนถึงโหนดชั้นซ่อนจำนวน 200 โหนด มาหาค่าเฉลี่ย
ดังสมการต่อไปนี้

$$\text{Average SD} = \frac{1}{L} \sum_{i=1}^L SD_i \quad (6)$$

เมื่อ

$$SD = \left(\frac{1}{C} \sum_{j=1}^C (Accuracy_j - \overline{Accuracy})^2 \right)^{\frac{1}{2}} \quad (7)$$

และ

$$\overline{Accuracy} = \frac{1}{C} \sum_{j=1}^C Accuracy_j \quad (8)$$

โดยค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานบอก
ถึงการกระจายตัวของกลุ่มข้อมูล ยิ่งค่าเฉลี่ยของความ
เบี่ยงเบนมาตรฐานน้อย แสดงว่าข้อมูลนั้นมีความ
น่าเชื่อถือสูง

3. ค่าความถูกต้องสูงสุด (Max accuracy) คือ
ค่าที่มากที่สุดจากค่าความถูกต้องในโหนดชั้นซ่อน
ตั้งแต่โหนดที่ 1 จนถึงโหนดที่ 200

โดยค่าความถูกต้องสูงสุดบอกถึงความ
สามารถในการจำแนกข้อมูลที่มีความถูกต้องสูงสุด
ของตัวจำแนกข้อมูลในโหนดใดโหนดหนึ่ง ตั้งแต่

โหนดที่ 1 จนถึงโหนดที่ 200 ยิ่งค่าความถูกต้องสูงสุด
มาก แสดงว่าตัวจำแนกข้อมูลสามารถจำแนกข้อมูลได้ดี
ในโหนดใดโหนดหนึ่งตั้งแต่โหนดที่ 1 จนถึงโหนดที่
200

4. ค่าความแตกต่าง (Difference : Dif) คือการ
เปรียบเทียบความแตกต่างระหว่างค่าของข้อมูลโดยนำ
ค่าที่ได้จาก ES-ME ลบด้วยค่าที่ได้จาก ELM ซึ่งได้
เปรียบเทียบค่าต่างๆดังต่อไปนี้

4.1 ความแตกต่างของค่าเฉลี่ยความถูกต้อง
ถ้าขั้นตอนวิธี ES-ME มีค่าเฉลี่ยของความถูกต้องดีกว่า
ขั้นตอนวิธี ELM แบบเดี่ยว ค่าที่ได้จะเป็น “+”

4.2 ความแตกต่างของค่าเฉลี่ยความเบี่ยงเบน
มาตรฐาน ถ้าขั้นตอนวิธี ES-ME มีค่าเฉลี่ยความ
เบี่ยงเบนมาตรฐานดีกว่าขั้นตอนวิธี ELM แบบเดี่ยว
ค่าที่ได้จะเป็น “-”

4.3 ความแตกต่างของค่าความถูกต้องสูง
ที่สุด ถ้าขั้นตอนวิธี ES-ME มีค่าความถูกต้องสูงที่สุด
ดีกว่าขั้นตอนวิธี ELM แบบเดี่ยว ค่าที่ได้จะเป็น “+”

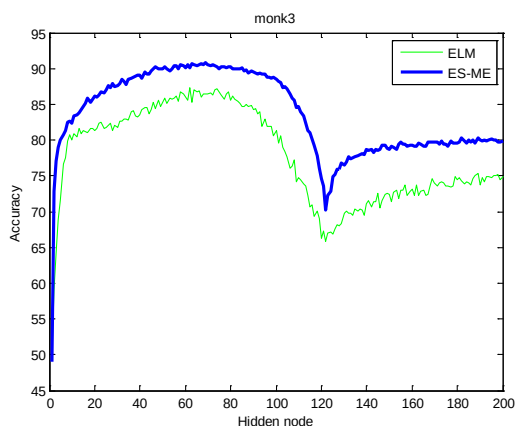
ตารางที่ 2 ผลการทดลองเปรียบเทียบ ค่าเฉลี่ยของความถูกต้อง (Average accuracy) ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐาน
(Average SD) และ ค่าความถูกต้องสูงสุด (Max accuracy) ระหว่าง ELM และ ES-ME

Datasets	Average accuracy			Average SD			Max accuracy		
	ELM	ES-ME	Dif	ELM	ES-ME	Dif	ELM	ES-ME	Dif
Glass	54.52	57.33	+2.81	2.224	1.850	-0.374	66.76	68.84	+2.08
BreastCancer	94.78	96.40	+1.62	0.689	0.377	-0.312	96.37	97.48	+1.11
Balance	88.87	89.80	+0.93	0.601	0.415	-0.186	91.03	91.56	+0.53
diabetes	74.51	75.48	+0.97	0.837	0.637	-0.200	77.14	77.51	+0.37
Statlog	72.31	74.73	+2.42	0.992	0.649	-0.343	73.85	76.03	+2.18
Contraceptive Method Choice	52.17	53.35	+1.18	0.757	0.561	-0.196	53.80	54.75	+0.95
Yeast	58.11	58.73	+0.62	0.526	0.402	-0.124	59.80	60.05	+0.25
Wine Quality	54.53	55.16	+0.63	0.323	0.228	-0.095	55.64	56.28	+0.64
Waveform	83.98	86.02	+2.04	0.349	0.193	-0.156	86.11	86.79	+0.68
Hayes	73.05	75.27	+2.22	4.975	2.895	-2.080	78.33	83.21	+4.88
Monk1	72.01	76.41	+4.40	2.827	1.435	-1.392	77.51	80.53	+3.02
Monk2	66.54	70.20	+3.66	2.055	1.185	-0.870	73.47	76.29	+2.82
Monk3	77.74	83.60	+5.86	2.889	1.368	-1.521	87.29	90.76	+3.47
Digits	90.52	92.47	+1.95	0.875	0.356	-0.519	97.04	97.23	+0.19
Average of Dif			+2.24			-0.561			+1.64

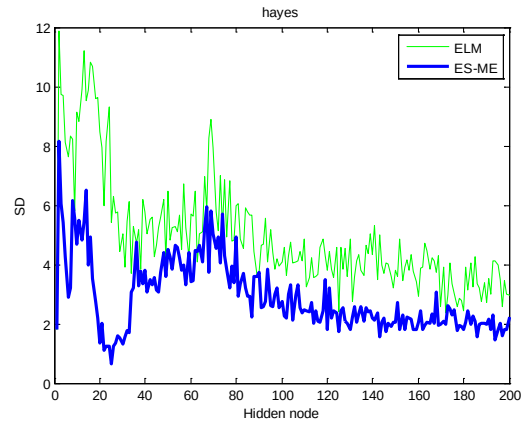
ผลการวิจัยและการอภิปรายผล

จากการทดลองเปรียบเทียบระหว่างขั้นตอนวิธี ELM แบบเดี่ยว และ ขั้นตอนวิธี ES-ME โดยกำหนดขอบเขตจำนวนโหนดชั้นซ่อนตั้งแต่ 1 - 200 โหนด และกำหนดให้จำนวน ELM ในขั้นตอนวิธี ES-ME เป็น 21 ตัว จะได้ผลการทดลองดังตารางที่ 2 ซึ่งแสดงให้เห็นว่า ขั้นตอนวิธี ES-ME ให้ค่าเฉลี่ยความถูกต้องดีขึ้นกว่าขั้นตอนวิธี ELM แบบเดี่ยวทุกชุดข้อมูล ดูได้จากช่อง Dif ของ Average accuracy ให้ผลเป็น “+” ทั้งหมด และแสดงให้เห็นว่า ขั้นตอนวิธี ES-ME ให้ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานดีขึ้นกว่าขั้นตอนวิธี ELM แบบเดี่ยวทุกชุดข้อมูลดูได้จากช่อง Dif ของ Average SD ให้ผลเป็น “-” ทั้งหมด รวมทั้งแสดงให้เห็นว่าขั้นตอนวิธี ES-ME ให้ค่าความถูกต้องสูงที่สุดดีกว่าขั้นตอนวิธี ELM แบบเดี่ยวทุกชุดข้อมูล ดูได้จากช่อง Dif ของ Max accuracy ให้ผลเป็น “+” ทั้งหมด

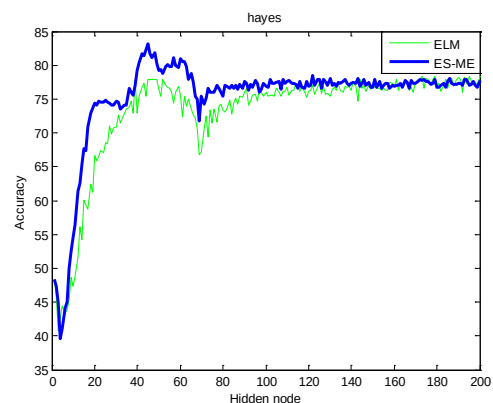
โดยข้อมูลที่มีค่าเฉลี่ยของความถูกต้องเพิ่มขึ้นมากที่สุดคือข้อมูล Monk3 มีค่าเฉลี่ยความถูกต้องเพิ่มขึ้น 5.86% และข้อมูลที่มีค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานลดลงมากที่สุดคือข้อมูล Hayes มีค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานลดลง 2.080% รวมทั้งข้อมูลที่มีค่าความถูกต้องสูงที่สุดเพิ่มขึ้นมากที่สุดคือข้อมูล Hayes มีค่าความถูกต้องสูงที่สุดเพิ่มขึ้น 4.88%



ภาพที่ 9 กราฟเปรียบเทียบค่าความถูกต้องระหว่าง ELM และ ES-ME ของชุดข้อมูล Monk3 ในแต่ละจำนวนโหนดชั้นซ่อน



ภาพที่ 10 กราฟเปรียบเทียบค่าความเบี่ยงเบนมาตรฐานระหว่าง ELM และ ES-ME ของชุดข้อมูล Hayes ในแต่ละจำนวนโหนดชั้นซ่อน



ภาพที่ 11 กราฟเปรียบเทียบค่าความถูกต้องระหว่าง ELM และ ES-ME ของชุดข้อมูล Hayes ในแต่ละจำนวนโหนดชั้นซ่อน

จากภาพที่ 9, 10 และ 11 แสดงการเปรียบเทียบระหว่างขั้นตอนวิธี ES-ME กับขั้นตอนวิธี ELM แบบเดี่ยวในแต่ละจำนวนโหนดชั้นซ่อน จากโหนดชั้นซ่อนจำนวน 1 โหนดจนถึงโหนดชั้นซ่อนจำนวน 200 โหนด โดย

ภาพที่ 9 แสดงกราฟในข้อมูล ES-ME ให้ค่าเฉลี่ยของความถูกต้องเพิ่มขึ้นดีกว่า ELM แบบเดี่ยวมากที่สุด ซึ่งจากกราฟแสดงให้เห็นว่าในแต่ละโหนดชั้นซ่อนของขั้นตอนวิธี ES-ME สามารถให้ความถูกต้องมากกว่าขั้นตอนวิธี ELM แบบเดี่ยวมาก

ภาพที่ 10 แสดงกราฟในข้อมูลที่ ES-ME ให้ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานต่ำลงดีกว่า ELM แบบเดี่ยวมากที่สุด ซึ่งจากกราฟแสดงให้เห็นว่าในแต่ละโหนดชั้นซ่อนของขั้นตอนวิธี ES-ME สามารถให้ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานต่ำลงกว่าขั้นตอนวิธี ELM แบบเดี่ยวมาก

ภาพที่ 11 แสดงกราฟในข้อมูลที่ ES-ME ให้ค่าความถูกต้องสูงที่สุดเพิ่มขึ้นดีกว่า ELM แบบเดี่ยวมากที่สุด ซึ่งจากกราฟแสดงให้เห็นว่าในโหนดชั้นซ่อนที่ 45 ขั้นตอนวิธี ES-ME สามารถให้ความถูกต้องมากกว่าขั้นตอนวิธี ELM แบบเดี่ยวมาก

จากการทดลองเปรียบเทียบระหว่างขั้นตอนวิธี ELM แบบเดี่ยว กับขั้นตอนวิธีที่นำเสนอ แสดงให้เห็นว่าขั้นตอนวิธีที่นำเสนอให้ค่าเฉลี่ยของความถูกต้องโดยรวมเพิ่มขึ้น 2.24% และให้ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐานโดยรวมลดลง 0.561% รวมทั้งให้ค่าเฉลี่ยของความถูกต้องสูงที่สุดโดยรวมเพิ่มขึ้น 1.64%

สรุปผลการวิจัย

การทำงานของขั้นตอนวิธี ELM แม้ว่าจะมีความเร็วสูงแต่ก็มีข้อเสียคือ ผลลัพธ์ที่ได้จากการพยากรณ์มีความไม่แน่นอนเนื่องจากค่าน้ำหนักระหว่างชั้นนำเข้า และ ชั้นซ่อน รวมทั้งค่าของไบแอสภายใน ELM ได้มาจากการสุ่ม ดังนั้นงานวิจัยนี้จึงได้เสนอขั้นตอนวิธีที่ช่วยเพิ่มความน่าเชื่อถือให้กับขั้นตอนวิธี ELM โดยใช้การประสานคำตอบจากการเลือกเครื่องจักรเรียนรู้เอ็กซ์ทรีมโดยพิจารณาค่าความผิดพลาดเฉลี่ย (ES-ME) เพื่อคัดเลือกเฉพาะ ELM ที่มีความผิดพลาดในการพยากรณ์ในชุดข้อมูลสอนต่ำกว่าค่าความผิดพลาดเฉลี่ย มารวมกันพยากรณ์ข้อมูลทดสอบ แล้วเลือกคำตอบโดยใช้วิธีการลงคะแนนเสียง เพื่อเพิ่มความถูกต้องของการพยากรณ์

จากการวิจัยพบว่าชุดข้อมูลที่ใช้วิจัยมีค่าเฉลี่ยของความถูกต้อง (Average accuracy) และ ค่าเฉลี่ยของความเบี่ยงเบนมาตรฐาน (Average SD) รวมทั้งค่าความถูกต้องสูงที่สุด (Max accuracy) ดีขึ้นทุกชุดข้อมูล จึงสามารถ

สรุปได้ว่าขั้นตอนวิธีที่นำเสนอสามารถเพิ่มความถูกต้องและน่าเชื่อถือของการพยากรณ์ให้เพิ่มมากขึ้นกว่าการใช้ขั้นตอนวิธี ELM แบบเดี่ยว

เอกสารอ้างอิง

ปัญญาพล หอระตะ. 2551. การเร่งความเร็วของการ

สอนข่ายงานประสาทเทียมสำหรับปัญหาการจำแนกประเภท. การประชุมวิชาการเทคโนโลยี และนวัตกรรมสำหรับการพัฒนาอย่างยั่งยืน. 2(1): 374-379.

Barstow, DR. 1979. The roles of knowledge and deduction in program synthesis. IJCAI'79 Proceedings of the 6th international joint conference on Artificial intelligence. 1(1): 37-43.

Bishop, CM. 2006. Pattern Recognition and Machine Learning. UK: Springer.

Huang, GB., Wang, DH., and Lan Y. 2011. Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics. 2(2): 107-122.

Huang, GB., Zhu, QY., and Siew CK. 2006. Extream Learning Machine: Theory and applications. Neurcomputing. 70(1-3): 489-501.

UC Irvine Machine Learning Repository. Retrieved September 15, 2011, from <http://archive.ics.uci.edu/ml>.

Zhou, ZH., and Jiang Y. 2003. Medical Diagnosis With C4.5 Rule Preceded by Artificial Neural Network Ensemble. Information Technology in Biomedicine, IEEE Transactions on. 7(1): 37-42.