การหาค่าเหมาะสมของสารสร้างตะกอนโดยแบบจำลองปัญญาประดิษฐ์

A COAGULANT DOSAGE OPTIMIZATION:

ARTIFICIAL INTELLIGENCE MODELLING APPROACH

นิลวัฒน์ เฟื่องโชค

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2563

การหาค่าเหมาะสมของสารสร้างตะกอนโดยแบบจำลองปัญญาประดิษฐ์

นิลวัฒน์ เฟื่องโชค

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2563

A COAGULANT DOSAGE OPTIMIZATION:
ARTIFICIAL INTELLIGENCE MODELLING APPROACH

NINLAWAT PHUANGCHOKE

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2020

ปริญญานิพนธ์

เรื่อง

การหาค่าเหมาะสมของสารสร้างตะกอนโดยแบบจำลองปัญญาประดิษฐ์

ของ

นิลวัฒน์ เฟื่องโชค

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าปริญญานิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.วราภรณ์ วิยานนท์)

..... ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.อัครา ประโยชน์)

..... ที่ปรึกษาร่วม

(อาจารย์ ดร.เสฏฐา ศาสนนันทน์)

..... กรรมการ

(อาจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	การหาค่าเหมาะสมของสารสร้างตะกอนโดยแบบจำลองปัญญาประดิษฐ์
ผู้วิจัย	นิลวัฒน์ เฟื่องโชค
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2563
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. วราภรณ์ วิยานนท์
อาจารย์ที่ปรึกษาร่วม	อาจารย์ ดร. เสฏฐา ศาสนนันทน์

กระบวนการที่มีความสำคัญที่สุดของกระบวนการผลิตน้ำประปา (Water treatment plant process) คือกระบวนการสร้างตะกอนโดยปัจจุบันโรงงานผลิตน้ำบางเขนใช้สารส้มและโพลิอลูมิเนียมคลอไรด์ ซึ่งมูลค่าการใช้ต่อวันคิดเป็นเงินหลายแสนบาท เพราะฉะนั้นการกำหนดปริมาณสารเคมีทั้งสองชนิดเป็นปัจจัยสำคัญที่สุดที่จะกำหนดให้การผลิตน้ำนั้นประหยัดและมีคุณค่า งานวิจัยนี้ประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียม (Artificial neural network, ANN) โดยใช้การฝึกสอนโครงข่ายประสาทเทียมแบบ Levenberg–Marquardt algorithm สร้างแบบจำลองทางคณิตศาสตร์ (Soft jar test) สำหรับการทำนายค่าเฉลี่ยปริมาณสารเคมีที่ใช้ในการสร้างตะกอนได้แก่ สารส้ม (Alum) และ โพลิอลูมิเนียมคลอไรด์ (Polyaluminum chloride, PACL) ซึ่งข้อมูลอินพุตประกอบด้วย ค่าความขุ่นน้ำดิบ (turbidity) ค่าความเป็นกรด-เบสของน้ำดิบ (pH) ค่าความเป็นด่างของน้ำดิบ (alkalinity) ค่าความนำไฟฟ้าของน้ำดิบ (conductivity) และค่าการบริโภคออกซิเจน (oxygen consumed, OC) ของโรงงานผลิตน้ำบางเขน การประปานครหลวง เป็นข้อมูลที่เก็บรวบรวมตั้งแต่วันที่ 1 มกราคม พ.ศ. 2562 ถึง วันที่ 31 ธันวาคม พ.ศ 2562 โดยครอบคลุมการเปลี่ยนแปลงฤดูกาลของประเทศไทย ข้อมูลอินพุตแบ่งออกเป็น 3 กลุ่มคือ กลุ่มข้อมูลฝึกสอน (training set) กลุ่มข้อมูลตรวจสอบ (test set) และกลุ่มข้อมูลทดสอบ (validation set) วัดประสิทธิภาพแบบจำลองด้วยค่าสัมประสิทธิ์แสดงการตัดสินใจ (Coefficient of determination, R^2) และ ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error, MAE) โดยประสิทธิภาพแบบจำลอง ANN ที่ดีที่สุดทำนาย Alum เท่ากับ 0.73, 3.18 และแบบจำลองทำนาย PACL เท่ากับ 0.59 , 3.21 ตามลำดับและเพิ่มประสิทธิภาพ Soft jat test โดยการประยุกต์ใช้เทคนิคการแบ่งกลุ่มข้อมูลที่เรียกว่า Self organizing map ซึ่งทำให้แบบจำลองมีประสิทธิภาพการทำนายปริมาณสารส้มและโพลิอลูมิเนียมคลอไรด์แม่นยำและถูกต้องกว่าเดิม

คำสำคัญ : Soft jar test, Jar test, กระบวนการผลิตน้ำประปา, โครงข่ายประสาทเทียม, Self organizing map

Title	A COAGULANT DOSAGE OPTIMIZATION: ARTIFICIAL INTELLIGENCE MODELLING APPROACH
Author	NINLAWAT PHUANGCHOKE
Degree	MASTER OF SCIENCE
Academic Year	2020
Thesis Advisor	Asst. Prof. Dr. Waraporn Viyanon
Co Advisor	Prof. Dr. Setta Sasananand

The most important process for water treatment plants is coagulation using alum and polyaluminum chloride (PACL), and the value of usage per day was one hundred thousand Baht. Therefore, determining the dosage of alum and PACL is the most important factor to be prescribed. Water production is economical and valuable. This research applies Artificial Neural Network (ANN), which uses the Levenberg–Marquardt algorithm to create a mathematical model (Soft Jar Test) for a prediction chemical dose used to coagulation, such as alum and PACL input data consists of turbidity, pH, alkalinity, conductivity, and, oxygen consumption (OC) of Bangkhen water treatment plant (BKWTP) of the Metropolitan Waterworks Authority. The data collected from 1 January 2019 to 31 December 2019 covered the changing seasons of Thailand. The input data of ANN was divided into a three-group training set, test set and validation set which the best model performance with a coefficient of determination and mean absolute error of alum are 0.73, 3.18 and PACL is 0.59, 3.21 respectively. The soft jar test was optimized by applying a data clustering technique called a self-organizing map, which made the model more efficient and gave more accurate predictions of alum and polyaluminum chloride dosage.

Keyword : Soft jar test, Jar test, Water treatment plant process, Artificial neural network, Self organizing map

กิตติกรรมประกาศ

ปริญญานิพนธ์นี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอกราบขอบพระคุณ ผศ.ดร.วราภรณ์ วิยานนท์ อาจารย์ที่ปรึกษา อ. ดร. เสฏฐา ศาสนนันทน์ อาจารย์ที่ปรึกษาร่วม และคณาจารย์ในหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการข้อมูลทุกท่านที่ให้คำปรึกษา คำแนะนำในการทำปริญญานิพนธ์ตลอดจนสนับสนุนข้อมูลทางวิชาการ และข้อมูลที่เกี่ยวข้องสำหรับทำปริญญานิพนธ์นี้ นอกจากนี้ขอขอบคุณเพื่อน ๆ พี่ ๆ ทุกคนที่เป็นกำลังใจ และให้ความช่วยเหลือในการทำปริญญานิพนธ์นี้

ขอกราบขอบพระคุณ บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ สำหรับทุนสนับสนุนการนำเสนอผลงานวิจัยของนิสิตบัณฑิตศึกษา สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณบิดามารดา และครอบครัว ที่เปิดโอกาศให้ได้รับเล่าเรียนศึกษาและช่วยเหลือให้กำลังใจผู้วิจัยตลอดเสมอมา

นิลวัฒน์ เฟื่องโชค

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญรูปภาพ

หน้า

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของงานวิจัย

น้ำเป็นปัจจัยสำคัญในการดำเนินชีวิตทั้งการบริโภคและอุปโภค การผลิตน้ำประปาต้องได้คุณภาพผ่านเกณฑ์มาตรฐานองค์การอนามัยโลก (World health organization) ที่กำหนด น้ำในกรุงเทพมหานครใช้กำลังผลิตจาก 3 โรงงานผลิตน้ำ (Water treatment plant, WTP) ได้แก่ บางเขน (Bangkhen water treatment plant, BKWTP) มหาสวัสดิ์ (Mahasawat water treatment plant, MSWTP) และสามเสน (Samsen water treatment plant, SSWTP) ซึ่งโรงผลิตน้ำบางเขนเป็นโรงงานผลิตน้ำขนาดใหญ่ที่มีกำลังการผลิตสูงที่สุดในบรรดากลุ่มโรงงานผลิตน้ำของการประปานครหลวง สามารถผลิตน้ำได้วันละประมาณ 4,400,000 ลูกบาศก์เมตร โดยรับน้ำดิบจากแม่น้ำเจ้าพระยาซึ่งมีสถานีสูบน้ำดิบสำแลสูบน้ำเคลื่อนที่ผ่านคลองส่งน้ำดิบฝั่งตะวันออกจนถึงโรงงานผลิตน้ำ เป็นระยะทางโดยประมาณ 18 กิโลเมตร ครอบคลุมพื้นที่ 30 เขต

น้ำดิบที่ใช้ในการผลิตน้ำประปาอาจมีสิ่งเจือปนอยู่มาก ทำให้น้ำมีความขุ่น อนุภาคคอลลอยด์ที่อยู่ในน้ำทั้งขนาดใหญ่และขนาดเล็กมีแรงกระทำอนุภาคดึงดูดและผลักกันซึ่งส่วนใหญ่แรงผลักกันมากกว่าแรงดึงดูดทำให้น้ำอนุภาคเหล่านี้ไม่รวมตัวกันและลอยอยู่ในน้ำ กระบวนการที่สำคัญในการผลิตน้ำประปา คือ ระบบจ่ายสารเคมีเพื่อปรับปรุงคุณภาพน้ำโดยพิจารณาข้อมูลหลายองค์ประกอบด้วยกันก่อนการเติมสารเคมี ไม่สามารถเติมสารเคมีได้อย่างอิสระ แต่มีการวัดค่าพารามิเตอร์คุณภาพน้ำในสายการผลิตหนึ่งในวิธีการตัดสินใจเติมสารเคมีคือวิธีการทำจาร์เทส (jar test) ซึ่งดำเนินการวันละ 2 ครั้ง ได้แก่ ตอนเช้า และตอนเย็น เป็นวิธีการคำนวณปริมาณสารเคมีที่ใช้การสร้างตะกอนโดยที่โรงงานผลิตน้ำบางเขนใช้สารเคมี 2 ชนิดได้แก่ สารส้ม (alum) และ โพลิอลูมิเนียมคลอไรด์ (polyaluminum chloride, PACL) สารเคมีทั้ง 2 ชนิดนี้เมื่อเติมลงไปในน้ำจะไปลดแรงผลักกันของอนุภาคเมื่ออนุภาคสัมผัสกันทำให้แรงดึงดูดดูดอนุภาคติดกันเกิดการตกตะกอน (coagulation) และรวมตัวกันจนได้ตะกอนที่ใหญ่ขึ้น ซึ่งต้องใช้สารเคมีและเจ้าหน้าที่ในการทดสอบ เป็นงานประจำเจ้าหน้าที่นักวิทยาศาตร์ที่โรงงานผลิตน้ำบางเขนทดสอบจาร์เทสทุกวันวันละ 2 ครั้ง (เช้าและเย็น) การทดสอบแต่ละครั้งต้องใช้เวลาในการรอผล 16 นาที การกำหนดค่าเหมาะสมของ Alum และ PACL โดยการใช้จาร์เทสแบบดังเดิมในที่นี้ขอเรียกว่าการทำจาร์เทสแบบออฟไลน์ (offline jar test) ทำให้เป็นปัญหาคอขวดของการผลิตน้ำประปา ซึ่งการทดสอบลักษณะเช่นนี้ไม่สามารถใช้เป็นตัวแทนของการเปลี่ยนแปลงน้ำประปาได้เนื่องจากน้ำมีการเคลื่อนที่อยู่ตลอดเวลา ถ้าเปรียบเทียบกับพารามิเตอร์อื่น เช่น ค่าความเป็นกรด-

ด่าง (pH) หรือ ค่าความขุ่น (turbidity) ที่การประปานครหลวงนั้นมีเซนเซอร์วัดที่สามารถรู้ค่าได้ทันที แต่ส่วนใหญ่ยังไม่ถูกนำมาใช้บันทึกจริงเนื่องจากมีเหตุการณ์กรณีที่เซิร์ฟเวอร์และเซนเซอร์เสียบ้าง การเติมสารเคมีทั้งสารส้มและโพลิอลูมิเนียมคลอไรด์จะถูกเติมในถังตกตะกอน (clarifier) น้ำที่ออกจากถังตกตะกอนต้องมีความขุ่นระหว่าง 3 NTU ถึง 5 NTU เมื่อน้ำเข้าสู่ขั้นตอนการกรองน้ำจะถูกกรองได้อย่างมีประสิทธิภาพและล้างย้อน (backwash) ทุก ๆ 48 ชั่วโมงโดยน้ำที่ผ่านกระบวนการกรองแล้วจะมีความขุ่นไม่เกิน 1 NTU ซึ่งเป็นน้ำที่ใสมาก

เพื่อทดแทนการทดสอบจาร์เทสแบบออฟไลน์ผู้วิจัยจึงนำเสนอวิธีการจาร์เทสแบบออนไลน์โดยใช้การสร้างแบบจำลองเทคนิคโครงข่ายประสาทเทียม (Artificial neural network, ANN) เพื่อทำนายปริมาณสารเคมีสารส้มและโพลิอลูมิเนียมคลอไรด์ซึ่งขอเรียกกระบวนการนี้ว่าซอฟต์จาร์เทส (Soft Jar test)

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยนี้ผู้วิจัยได้ตั้งวัตถุประสงค์ไว้ดังนี้

1.2.1 เพื่อสร้างแบบจำลองปัญญาประดิษฐ์ โดยใช้เทคนิค ANN เพื่อทำนายปริมาณสารส้ม (alum) และ โพลิอลูมิเนียมคลอไรด์ (PACL)

1.2.2 พัฒนาและปรับปรุงประสิทธิภาพการทำงานของแบบจำลองซอฟต์จาร์เทสโดยใช้เทคนิคการแบ่งกลุ่มข้อมูลด้วยวิธี Self organizing map

1.3 ขอบเขตการวิจัย

1.3.1 ข้อมูลที่ใช้ในการวิจัย

ข้อมูลคุณภาพน้ำดิบและข้อมูลการทำจาร์เทส ตลอดปี พ.ศ. 2562 ของโรงงานผลิตน้ำบางเขน

1.3.2 ตัวแปรที่ศึกษา

1.3.1.1 ตัวแปรต้น แบ่งเป็นดังนี้

ข้อมูลคุณภาพน้ำดิบ

- ความขุ่น (turbidity)
- สภาพความเป็นด่าง (alkalinity)
- ความเป็นกรด-ด่าง (pH)
- ความนำไฟฟ้า (conductivity)
- การบริโภคออกซิเจน (oxygen consumption)

1.3.1.2 ตัวแปรตาม แบ่งเป็นดังนี้

ข้อมูลการทดสอบจาร์เทส

- ปริมาณสารส้ม (Alum)
- ปริมาณโพลิอลูมิเนียมคลอไรด์ (PACL)

1.3.3 เทคนิคที่ใช้ในการวิจัย

การวิจัยนี้ใช้เทคนิคโครงข่ายประสาทเทียม (ANN) 2 ลักษณะได้แก่ การเรียนรู้แบบมีการสอน (supervised learning) และการเรียนรู้แบบไม่มีการสอน (unsupervised learning) ด้วยวิธี Self organizing map มาใช้ในทำนายและวัดประสิทธิภาพแต่ละแบบจำลองด้วย (Mean absolute error, MAE) และ R^2

1.4 วิธีการดำเนินการวิจัย

1.4.1 ศึกษาค้นคว้าทฤษฎีและงานวิจัย (Literature review) ที่เกี่ยวข้อง

1.4.2 กำหนดแหล่งข้อมูลและดำเนินการขอข้อมูลการเติมสารเคมีเพื่อวัดคุณภาพน้ำ

1.4.3 การสำรวจและวิเคราะห์ข้อมูล (Exploratory data analysis)

1.4.4 เตรียมข้อมูลเพื่อให้สามารถนำไปสำหรับสร้างแบบจำลอง

1.4.5 สร้างแบบจำลองในการทำนายข้อมูล

1.4.6 วัดประสิทธิภาพของแต่ละแบบจำลอง

1.4.7 ประเมินและสรุปงานวิจัย

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

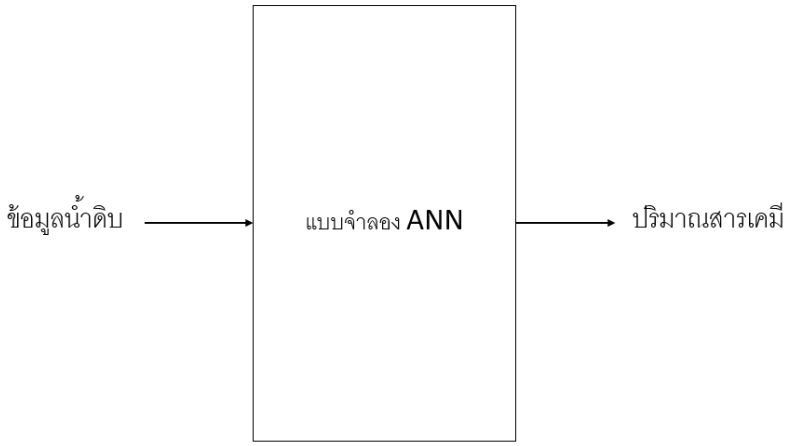
1.5.1 ได้แบบจำลองการหาปริมาณที่เหมาะสมของสารส้ม

1.5.2 ได้แบบจำลองการหาปริมาณที่เหมาะสมของ PACL ทั้ง 2 แบบจำลองสามารถนำมาทดแทนการทำจาร์เทสแบบเดิม

1.6 กรอบแนวคิดในงานวิจัย

แบบจำลองซอฟต์จาร์เทสสร้างจากข้อมูลคุณภาพน้ำดิบ และ การกำหนดใช้สารเคมี Alum และ PACL ในช่วงเดือน มกราคม ถึง ธันวาคม ปี 2562 (โดยได้รับอนุญาติจาก การประปานครหลวง ตามหนังสือ เลขที่ อว 8712.1/1596)

แผนผังแนวคิดในการวิจัย ดังภาพประกอบ 1 แบบจำลอง ANN มีอินพุตเป็นข้อมูลคุณภาพน้ำดิบและเอาต์พุตทำนายปริมาณสารเคมีสารส้มและโพลิอลูมิเนียมคลอไรด์



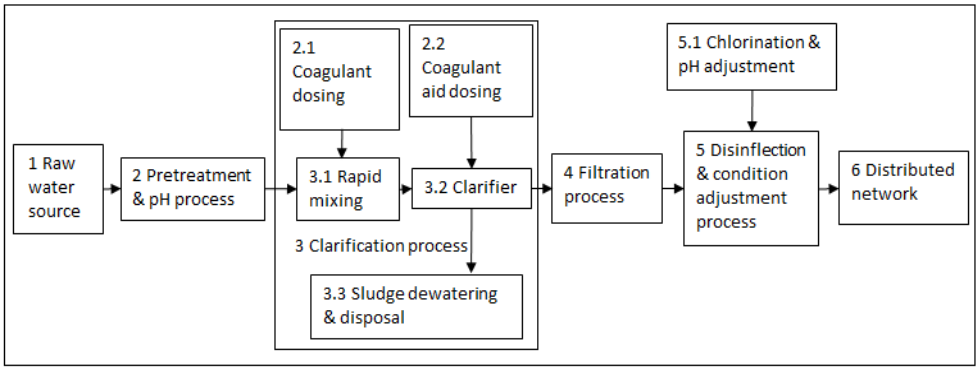
ภาพประกอบ 1 แผนผังแนวคิดการงานวิจัย

บทที่ 2
หลักการ ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้ศึกษากระบวนการที่เกี่ยวข้องกับการผลิตน้ำประปาการสร้างตะกอนซึ่งเป็นขั้นตอนหนึ่งในการผลิตน้ำประปา และ หลักการสร้างแบบจำลองโครงข่ายประสาทเทียม โดยมีรายละเอียดดังนี้

2.1 กระบวนการผลิตน้ำประปา

เพื่อให้น้ำประปามีคุณภาพตามเกณฑ์มาตรฐานองค์การอนามัยโลก (World Health Organization, WHO) ในแต่ละขั้นตอนการผลิตจึงต้องพิถีพิถันซึ่งกระบวนการผลิตน้ำประปาประกอบไปด้วย 6 ขั้นตอนหลัก ดังภาพประกอบ 2 ประกอบไปด้วย 1) ระบบสูบน้ำดิบ 2) ระบบจ่ายสารเคมี 3) ระบบตกตะกอน 4) ระบบการกรองน้ำ 5) ระบบฆ่าเชื้อโรค และ 6) ระบบสูบน้ำ ซึ่งขั้นตอนการผลิตน้ำทั้งหมดตั้งแต่เริ่มต้นจนได้น้ำที่มีคุณภาพใช้เวลาทั้งหมดประมาณ 4 ชั่วโมง



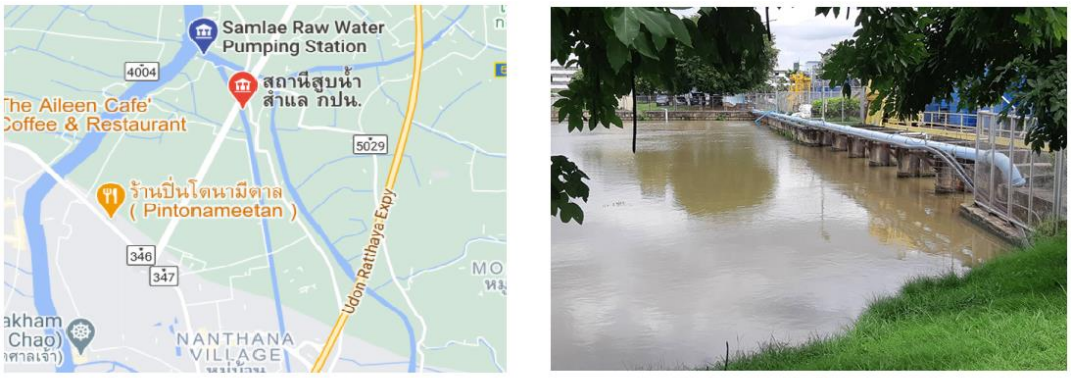
ภาพประกอบ 2 กระบวนการผลิตน้ำประปา

ที่มา : S. Sasananan, "Water Treatment Plant Clarifier Control: An Artificial Intelligence Approach," Doctoral dissertation, University of Tasmania, Australia, 2009.

2.2.1 ระบบสูบน้ำดิบ

โรงงานผลิตน้ำบางเขนรับน้ำดิบมาจากแม่น้ำเจ้าพระยามีโรงสูบน้ำดิบสำแลทำหน้าที่สูบน้ำเป็นระยะทาง 18 กิโลเมตร ผ่านเส้นทางธรรมชาติที่ไม่มีผู้อยู่อาศัยดังภาพประกอบ 3 (ก) โดยทางเข้าน้ำดิบที่โรงงานผลิตน้ำบางเขนเป็นดังรูปภาพประกอบ 3 (ข) น้ำเข้าสู่โรงงานผลิตน้ำ

บางเขนภายในจะมีกำแพงหยาบ และ กำแพงละเอียดทำหน้าที่กรองสิ่งวัสดุที่ติดมากับน้ำดัง รูปภาพประกอบ 4



ภาพประกอบ 3 (ก) แผนที่เส้นทางแหล่งน้ำดิบ (ข) ทางเข้าน้ำดิบโรงงานผลิตน้ำบางเขน

ที่มา : https://www.google.com/maps/search/โรงงานสูบน้ำสำแล/@14.0139746,100.55031,14z, [5 ตุลาคม 2563]

ภาพประกอบ 4 กำแพงหยาบสำหรับกรองวัสดุขนาดใหญ่

2.2.2 ระบบจ่ายสารเคมี

ระบบจ่ายสารเคมีมีวัตถุประสงค์ [1] เพื่อปรับสภาพน้ำให้ได้คุณภาพตามมาตรฐานองค์การอนามัยโลกโดยมีขั้นตอน 1) ใช้ปูนไลม (lime) หรือ กรดกำมะถัน (sulfuric acid) ปรับความเป็นกรด-ด่าง (pH) ของน้ำ 2) เพื่อเพิ่มประสิทธิภาพในการตกตะกอน ซึ่งต้องอาศัยสาร 2 ประเภทคือ 2.1) สารตกตะกอน (coagulant) คือ สารเคมีชนิดสารอนินทรีย์ เช่น สารส้ม (alum)

โพลิอลูมิเนียมคลอไรด์ (polyaluminum chloride) เป็นต้น 2.2) สารช่วยสร้างตะกอน (coagulant aid) คือ สารเคมีชนิดสารอินทรีย์เพื่อเพิ่มประสิทธิภาพในการตกตะกอนแต่ต้องใช้ร่วมกับสารตกตะกอนในการทำปฏิกิริยา เช่น โพลีเมอร์ (anionic polymer) เป็นต้น 3) การตกตะกอนผลึก (precipitation) เช่น การกำจัดความกระด้างโดยใช้สารเคมี ปูนขาว (calcium oxide) โซดาแอช (Na_2Co_3) เป็นต้น และ 4) เพื่อฆ่าเชื้อโรค (disinfection) เช่น การเติมคลอรีน (chlorine) เป็นต้น ในขั้นตอนนี้นับเป็นขั้นตอนสำคัญที่ต้องเลือกชนิดและปริมาณที่เหมาะสมกับสภาพของน้ำดิบอย่างระมัดระวังซึ่งอาศัยความเชี่ยวชาญของเจ้าหน้าที่ปฏิบัติงานที่ต้องคำนวณปริมาณสารเคมีแต่ละชนิดให้เหมาะสมกับคุณภาพน้ำดิบ ณ เวลาขณะนั้นโดยทดลอง เพื่อเก็บข้อมูลคุณภาพน้ำดิบประกอบการตัดสินใจว่าควรใช้ปริมาณสารเคมีมากน้อยเท่าใดด้วยการทำจาร์เทส

2.2.3 ระบบตกตะกอน

หลังจากที่ใส่สารสร้างตะกอนในปริมาณที่เหมาะสมกับความขุ่นลงในน้ำดิบแล้ว น้ำดิบจะไหลสู่ถังตกตะกอนผสมกัน เพื่อแยกตะกอนกับน้ำออกจากกัน ในกรณีที่น้ำดิบยังคงมีความขุ่นสูงมากนั้นจะมีการใส่สารช่วยในการตกตะกอนคือ โพลีอิเล็กโทรไลต์ (polyelectrolyte) เพื่อเพิ่มประสิทธิภาพของการตกตะกอน ที่ปลายของถังตกตะกอนจะมีการกวนน้ำทำให้สารแขวนลอยสัมผัสกันและจับตัวรวมกัน เป็นเม็ดตะกอนที่มีขนาดใหญ่ขึ้นและน้ำหนักมากขึ้นซึ่งจะตกลงสู่ก้นถัง แล้วไหลสู่ด้านล่างของถังโดยที่พื้นที่ด้านล่างจะมีเครื่องกวาดช่วยกวาดตะกอนเหล่านี้ลงสู่ก้นถังเพื่อควบคุมประสิทธิภาพในถังตกตะกอนโดยต้องคำนึงถึงปัจจัยเหล่านี้เช่น ความเร็วกังหัน (turbine speed) และเวลาในการระบายตะกอน (sludge drain) เป็นต้น ความขุ่นของน้ำที่ตกตะกอนแล้วถูกควบคุมไม่เกิน 4 NTU ภาพประกอบ 5 แสดงถังตกตะกอนชนิดหมุนวน (Solid contact recirculation) ซึ่งตะกอนที่ได้จะนำไปบำบัดและกำจัดต่อไปโดยในกระบวนการตกตะกอนตามที่กล่าวใช้เวลาทั้งหมดประมาณ 2 ชั่วโมง

ภาพประกอบ 5 ถังตกตะกอนชนิดหมุนวน

2.2.4 ระบบการกรองน้ำ

น้ำที่ได้จากกระบวนการตกตะกอนจะไหลเข้าสู่กระบวนการกรองผ่านสารกรองน้ำ 2 ชั้น ชั้นที่ 1 แอนทราไซต์เป็นชั้นถ่านหินชนิดหนึ่งและไหลสู่ชั้นที่ 2 เป็นทรายโดยมีหัวกรองน้ำหรือหัวนอซเซิล (Nozzle) ที่พื้นบ่อกรอง ดังภาพประกอบ 6 เพื่อป้องกันไม่ให้สารกรองหลุดปนออกไปกับน้ำ โดยน้ำที่ผ่านกระบวนการกรองน้ำจะมีความขุ่นไม่เกิน 1 NTU ซึ่งน้ำจะมีความใสมาก เมื่อการใช้งานผ่านไป สารกรองจะมีความสกปรกมากขึ้นหรือน้ำที่ผ่านจากถังตะตอนยังมีความขุ่นมากส่งผลให้การกรองมีความสกปรกยิ่งขึ้น จึงต้องมีล้างทำความสะอาดเพื่อให้หัวกรองมีความสะอาดโดยเรียกวิธีการนี้ว่า การล้างย้อน (backwash) [1] โดยการพ่นลมดันย้อนขึ้นทำให้สารกรองเคลื่อนตัวเสียดสีกันตะกอนและสิ่งสกปรกที่ติดอยู่จะหลุดออก จากนั้นจึงใช้น้ำดันย้อนเพื่อพัดพาตะกอนและสิ่งสกปรกออกไปสู่ระบบกำจัดตะกอนต่อไปแต่การล้างย้อน (backwash) บ่อยๆ นั้นทำให้สูญเสียน้ำและค่าใช้จ่ายซึ่งโดยปกติจะทำ ทุก ๆ 48 ชั่วโมง ซึ่งจะใช้เวลาประมาณ 15 นาที ต่อการล้าง 1 ครั้ง

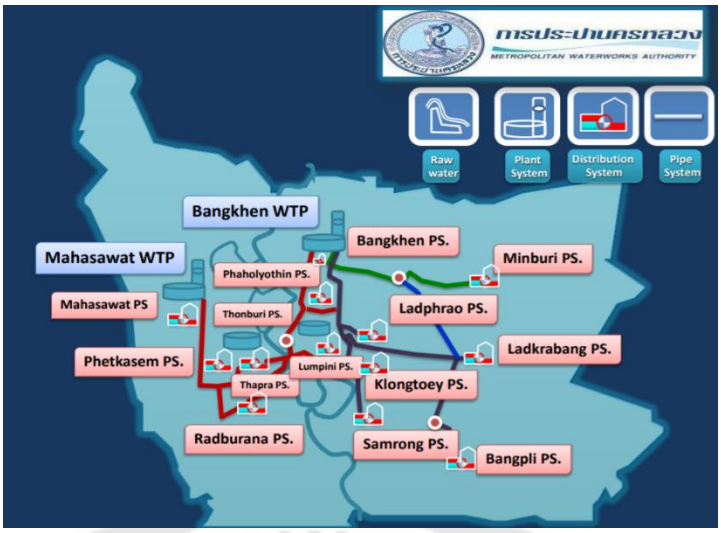
ภาพประกอบ 6 พื้นที่บริเวณบ่อกรองน้ำของโรงงานผลิตน้ำบางเขน

2.2.5 ระบบฆ่าเชื้อโรค

ระบบฆ่าเชื้อโรคเป็นขั้นตอนที่สำคัญมาก น้ำที่ผ่านกระบวนต่าง ๆ มา จำเป็นต้องถูกค่าเชื้อโรค เนื่องจากในน้ำธรรมชาติมีจุลินทรีย์ปนเปื้อนอยู่มาก ทั้งชนิดและปริมาณขั้นตอนในการบำบัดน้ำโดยทั่วไปเพราะการตกตะกอนและการกรองไม่สามารถกำจัดเชื้อโรคได้จึงจำเป็นต้องมีระบบกำจัดเชื้อโรคด้วย เพื่อให้ได้น้ำประปาที่สะอาดปราศจากเชื้อโรค เหมาะแก่การอุปโภคบริโภคตามมาตราฐานองค์การอนามัยโลก (World health organization) บางครั้งอาจมีการเติมปูนขาวอีกเล็กน้อยเพื่อปรับความเป็นกรดเป็นด่าง เพื่อป้องกันการกัดกร่อนในระบบเส้นท่อที่ใช้ในการลำเลียงน้ำประปาและคลอรีนเพื่อฆ่าเชื้อโรค เมื่อน้ำถูกฆ่าเชื้อโรคเรียบร้อยจะถูกส่งไปยังถังเก็บน้ำใส (water tank)

2.2.6 ระบบสูบน้ำ

เมื่อน้ำผ่านระบบฆ่าเชื้อโรคเรียบร้อยแล้วน้ำจะมีความสมบูรณ์พร้อมที่จะอุปโภคและบริโภค กระบวนการสุดท้ายคือ การสูบน้ำลำเลียงน้ำประปาจากถังเก็บน้ำใสไปยังโรงสูบน้ำเพื่อส่งน้ำไปยังสถานที่ต่าง ๆ ดังภาพประกอบ 7 ให้กับประชาชน



ภาพประกอบ 7 สถานีสูบจ่ายน้ำ ณ ปี พ.ศ.2557

ที่มา : S. Deesawasmongkol, "An introduction to Metropolitan Waterworks Authority (MWA)." สืบค้นจาก https://www.unescap.org/sites/default/files/Day%201_10.30SD_Water%20Supply.pdf, [5 ตุลาคม 2563]

2.2 การสร้างตะกอนด้วยวิธีจาร์เทส

การเลือกใช้ชนิดสารเคมี สารตกตะกอน (coagulant) หรือ สารช่วยสร้างตะกอน (coagulant aid) และปริมาณที่ต้องใช้จะต้องทำการทดลองให้ได้ค่าที่เหมาะสม (optimum dose) โดยการทดลองที่นิยมใช้คือจาร์เทส (jar test) เป็นวิธีการทดลองเพื่อหาปริมาณของสารสร้างตะกอน (coagulant) สำหรับน้ำดิบ (raw water) ประกอบด้วยขั้นตอน ดังภาพประกอบ 8

ขั้นตอนการหาค่าความขุ่น (turbidity) พีเอช (pH) และค่าความเป็นด่างของน้ำดิบ (Alkalinity)

2.2.1 ตวงน้ำดิบใส่ในบีกเกอร์ ทั้ง 6 ใบ ๆ ละ 1 ลิตร วางในเครื่องทดสอบการตกตะกอน

2.2.2 ปีเปตสารละลายสารส้ม 1 % ลงในบีกเกอร์แต่ละใบในปริมาณที่เพิ่มขึ้นตามลำดับ (เติมสารส้มเมื่อมีการกวนเร็วเกิดขึ้นแล้ว)

2.2.3 กวนน้ำและสารส้มอย่างรวดเร็ว (Rapid mixing) โดยใช้ความเร็ว 100 รอบ/นาที เป็นเวลา 1 นาที

2.2.4 กวนอย่างช้า (Slow mixing) โดยใช้ความเร็ว 50 รอบ/นาที เป็นเวลา 5 นาที หลังจากนั้นลดความเร็วรอบให้เป็น 20 รอบ/ นาที และกวนต่ออีก 5 นาที

2.2.5 ปล่อยให้ตกตะกอนเป็นเวลา 5 นาทีแล้วนำน้ำส่วนใสข้างบนของบีเกอร์แต่ละใบ (supernatant) มาวิเคราะห์หาค่าความขุ่น (turbidity) พีเอช (pH) และค่าความเป็นด่าง (alkalinity) ตามตาราง 1

ตาราง 1 พารามิเตอร์ที่วิเคราะห์ในการทำจาร์เทส (Jar test)

พารามิเตอร์	หน่วย	คำอธิบาย
ความเป็นกรด-ด่าง (pH)	pH	น้ำที่มีค่า pH น้อยกว่า 7 เป็นน้ำที่มีความเป็นกรดมีฤทธิ์กัดกรอน แต่มากกว่า 7 น้ำนั่นจะมีฤทธิ์เป็นเบสหรือด่าง
สภาพความเป็นด่า (alkalinity)	mg/L	ความสามารถหรือคุณสมบัติของน้ำที่ทำให้กรดเป็นกลาง
ความขุ่นของน้ำ(turbidity)	NTU	น้ำที่มีความขุ่นหรือความมั่วเกิดจากการปนเปื้อนของอนุภาคต่าง ๆ ทำให้แสงทำไม่สามารถส่องผ่านได้

ภาพประกอบ 8 แสดงเครื่องที่ใช้ในการทดสอบจาร์เทส (jar test)

โดยเฉลี่ยในขั้นตอนดังกล่าวใช้เวลาดำเนินการ 16 นาทีและน้ำดิบ (Raw water) เดินทางจากปากปล่องมาห้องปฏิบัติการผ่านท่อใช้เวลาประมาณ 20 นาที

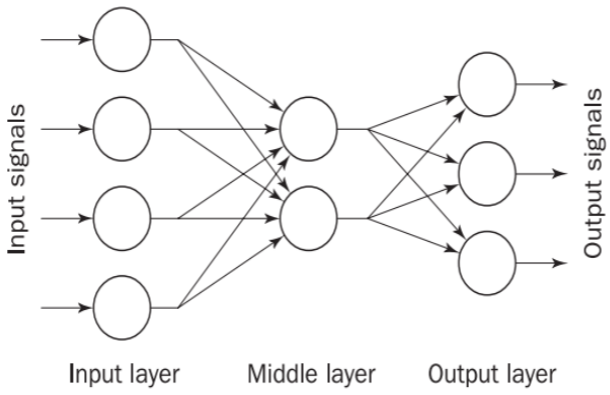
2.3 โครงข่ายประสาทเทียม (Artificial neuron network)

2.3.1 พื้นฐานการทำงานโครงข่ายประสาทเทียม

เทคโนโลยีที่นำมาช่วยในการทำนายปริมาณสารส้ม และ โพลิอลูมิเนียมคลอไรด์ที่ใช้ในการจ่ายสารเคมีของผู้ปฏิบัติงานและปริมาณแนะนำการจากทำจาร์เทส (Jar test) คือเทคนิคโครงข่ายประสาทเทียม (Artificial neuron network) [2] พื้นฐานและหลักการทำงานเบื้องต้นที่ถูกคิดค้นโดย Frank Rosenblatt ในปี 1957 โครงข่ายประสาทเทียมเหมาะสำหรับนำมาแก้ไขปัญหาที่เป็นความสัมพันธ์ที่ไม่เป็นเชิงเส้น (Non-Linear) ไม่มีความสัมพันธ์ที่ชัดเจนระหว่างตัวแปรโดยจะนำเทคนิคการส่งค่าย้อนกลับ (Back-propagation network) มาใช้ในงานวิจัย

2.3.1.1 โครงสร้างโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมสามารถกำหนดรูปแบบของการใช้เหตุผลตามมนุษย์ สมองประกอบด้วยชุดเซลล์ประสาทที่เชื่อมต่อกันอย่างหนาแน่นหรือหน่วยประมวลผลข้อมูลพื้นฐาน ซึ่งเรียกว่า เซลล์ประสาท [2] สมองของมนุษย์เราประกอบด้วยเซลล์ประสาทเกือบ 10,000 ล้าน และการเชื่อมต่อ 60 ล้านไซแนปส์ (Synapse) ด้วยการใช้เซลล์ประสาทหลาย ๆ ตัวพร้อมกัน ทำให้มนุษย์มีความฉลาดมากกว่าคอมพิวเตอร์ แต่ในบางเรื่องคอมพิวเตอร์ก็เก่งกว่ามนุษย์ โครงสร้างของโครงข่ายประสาทเทียมดังภาพประกอบ 9 ประกอบด้วย 3 องค์ประกอบ ดังนี้



ภาพประกอบ 9 ชั้นของโครงข่ายประสาทเทียม

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

1.ชั้นอินพุต (input layer) เป็นชั้นแรกที่รับข้อมูลจะมีจำนวนนิวรอนเท่ากับจำนวนอินพุต (input signals) ที่รับเข้ามา

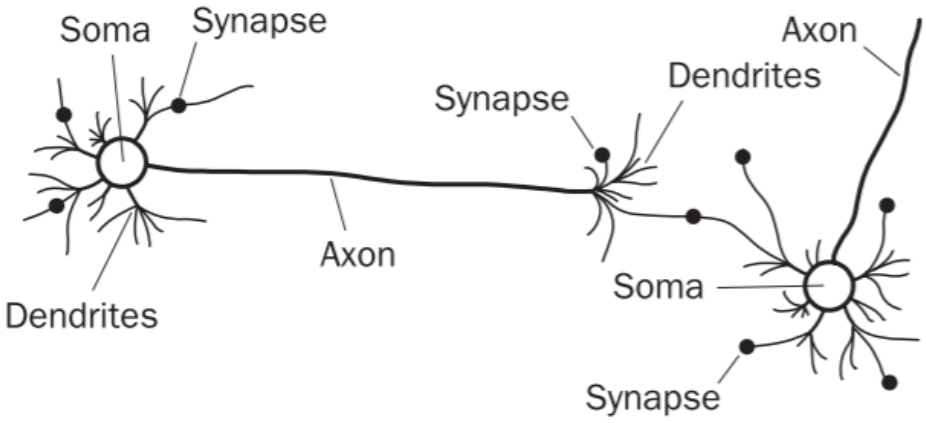
2.ชั้นกลาง (middle layer หรือ hidden layer) เป็นชั้นที่อยู่ระหว่างชั้นอินพุตและชั้นเอาต์พุตทำหน้าที่รวมอินพุตที่คูณด้วยน้ำหนัก (Weight) แล้วคำนวณแปลงให้เป็นผลลัพธ์ (output) แล้วส่งไปชั้นเอาต์พุต

3.ชั้นเอาต์พุต (output layer) เป็นชั้นสุดท้ายในโครงสร้างของโครงข่ายประสาทเทียมในการแสดงผลข้อมูลมีจำนวนนิวรอนเท่ากับ จำนวนเอาต์พุต (output signals) ขึ้นอยู่กับนักวิจัยว่าต้องการเท่าไรเพื่อให้เหมาะสมกับงานวิจัยนั้น ๆ

วิธีการเรียนรู้ของโครงข่ายประสาทเทียมมี 2 ลักษณะคือ การเรียนรู้แบบมีผู้ช่วยสอน (supervised learning) และ การเรียนรู้แบบมีไม่มีผู้ช่วยสอน (unsupervised learning) ซึ่งขึ้นอยู่กับข้อมูลที่นำมาวิเคราะห์ วิธีที่ใช้ในการแก้ไขปัญหา และผลลัพธ์ที่ต้องการ หลักการทำงานของโครงข่ายประสาทเทียม (Neuron network) พยายามเลียนแบบการทำงานอวัยวะภายในของสมองได้แก่ ใยประสาท (dendrite) เซลล์ (soma) แกนประสาท (axon) และจุดประสานประสาท (synapse) ดังภาพประกอบ 10 ซึ่งเมื่อเปรียบเทียบอวัยวะสมองกับโครงสร้าง ANN เป็นดังตาราง 2

ตาราง 2 การเปรียบเทียบกับระหว่างโครงข่ายทางชีววิทยาของสมองกับโครงข่ายประสาทเทียม

Biological neural network	Artificial neural network
Soma	Neuron
Dendrite	Input
Axon	Output
Synapse	Weight



ภาพประกอบ 10 โครงข่ายประสาทชีวภาพ

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

2.3.1.1 เพอร์เซปตรอน

เพอร์เซปตรอนเป็นโครงข่ายประสาทอย่างง่ายประกอบไปด้วยหนึ่งเซลล์ประสาทโดยการปรับค่า Weight และ Activation function เพื่อสร้างขอบเขตการตัดสินใจได้จากสมการ X ซึ่ง Single Layer Perceptron สามารถเรียนรู้แก้ไขปัญหาเฉพาะรูปแบบเชิงเส้นเท่านั้น

$$\sum_{i=1}^{n} X_i W_i - \Theta = 0 \qquad (1)$$

โดยที่ X_i คือค่าของอินพุต i

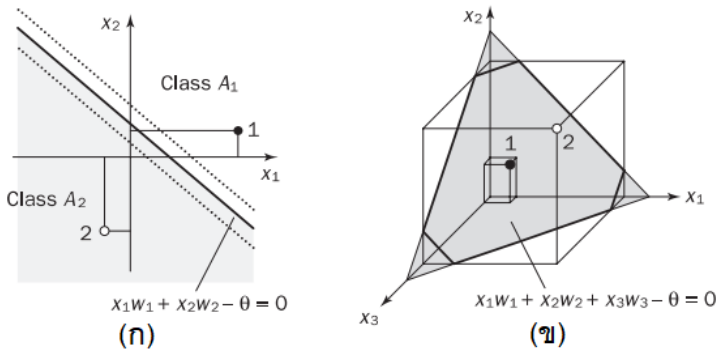
W_i คือน้ำหนักของอินพุต i

Θ คือ Threshold สามารถใช้เปลี่ยนขอบเขตการตัดสินใจ (Decision boundary)

กรณีที่มี 2 อินพุต X_1 และ X_2 ขอบเขตการตัดสินใจอยู่ในรูปแบบของเส้นตรงหนา ดังภาพประกอบ 11 (ก) โดยจุดที่ 1 อยู่เหนือเส้นแบ่งเขตเป็นคลาส A_1 และจุดที่ 2 อยู่ใต้เส้นแบ่งเขตเป็นคลาส A_2

กรณีที่มี 3 อินพุต X_1, X_2 และ X_3 ขอบเขตการตัดสินใจยังคงสามารถมองเห็นได้ ดังภาพประกอบ 11 (ข) สำหรับ Perceptron 3 อินพุต สามารถเขียนอยู่ในสมการคณิตศาสตร์ได้ดังนี้

$$X_1 W_1 + X_2 W_2 + X_3 W_3 = 0 \qquad (2)$$

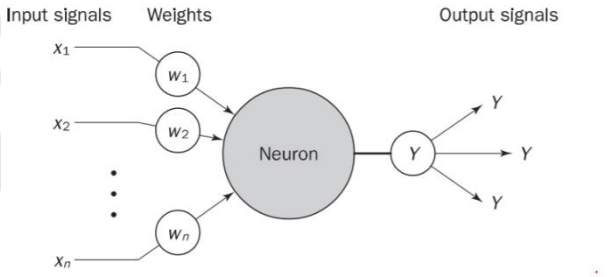


ภาพประกอบ 11 (ก) ความสามารถการแยกเชิงเส้นของ Perceptron 2 อินพุต (ข) Perceptron 3 อินพุต

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

2.3.1.3 เซลล์ประสาทเทียม

การทำงานของเซลล์ประสาท (Neuron) ดังภาพประกอบ 12 สามารถอธิบายด้วยสมการคณิตศาสตร์ได้ดังนี้



ภาพประกอบ 12 แผนภาพของเซลล์ประสาท

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

$$X = \sum_{i=1}^{n} X_i W_i \quad (3)$$

$$Y = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (4)$$

เมื่อ X คืออินพุตถ่วงน้ำหนักสุทธิของเซลล์ประสาท (Neuron)

X_i คือค่าของอินพุต i

W_i คือน้ำหนักของอินพุต i

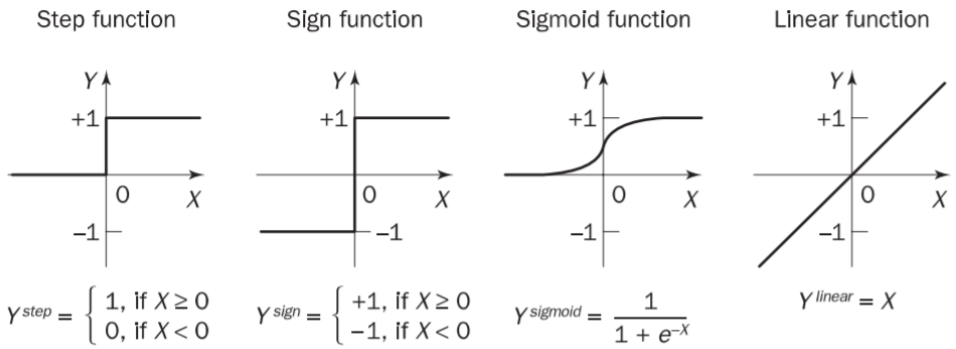
n คือจำนวนอินพุตของเซลล์ประสาท

Y คือเอาต์พุตของเซลล์ประสาท

ประเภทของ Activation function ของ Y เรียกว่า Sign function ดังนั้น ผลลัพธ์ที่แท้จริงของเซลล์ประสาทสามารถแสดงเป็น

$$Y = sign[\sum_{i=1}^{n} X_i W_i - \theta] \quad (5)$$

Activation function ทำหน้าที่รับข้อมูลอินพุตมาจากหลายเซลล์ประสาท (Neuron) แล้วนำมาประมวลผลว่าควรจะส่งต่อผลลัพธ์ไปเป็นเท่าไร ซึ่ง มีหลากหลายประเภทมาก ดังภาพประกอบ 13 เมื่อสังเกต Step function และ Sign function ผลลัพธ์ที่ได้จะมีเพียง 2 ค่าคือ 1 หรือ 0 กับ +1 หรือ -1 ซึ่งเหมาะกับการนำไปแก้ปัญหาการจำแนกประเภทเช่น การแยกประเภทเหรียญ เป็นต้น ในทางกลับกัน Sigmoid function และ Linear function ผลลัพธ์จะตอบสนองกับข้อมูลทุกช่วงเหมาะกับปัญหาประเภทเชิงเส้นและไม่เป็นเชิงเส้นเช่น การทำนายราคาหุ้น เป็นต้น

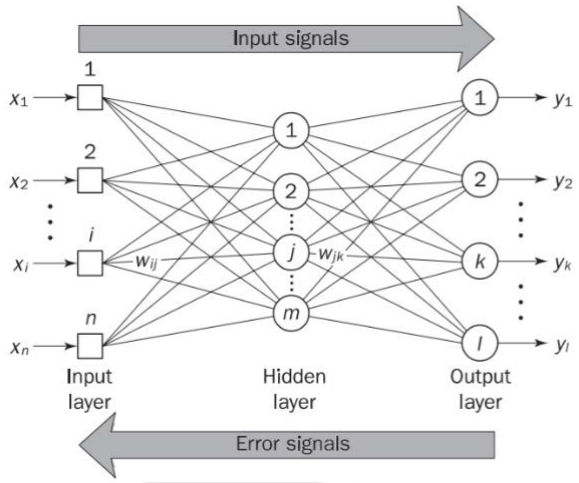


ภาพประกอบ 13 Activation functions ของเซลล์ประสาท

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

2.3.1.4 การส่งถ่ายย้อนกลับ

ภาพประกอบ 14 แสดงโครงข่ายประสาทเทียมมีทั้งหมด 3 ชั้น โดยมีตัวแปร i j และ k อ้างถึง จำนวนเซลล์ในชั้นอินพุต จำนวนเซลล์ในชั้นกลาง และ จำนวนเซลล์ในชั้นเอาต์พุต ตามลำดับ



ภาพประกอบ 14 โครงข่ายประสาท Back-propagation 3 ชั้น

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

โดยมีสัญญาณอินพุตคือ $X_1, X_2, X_3, X_4, \dots, X_n$ ถ่ายทอดจากด้านซ้ายไปด้านขวา สัญญาณข้อผิดพลาดคือ $e_1, e_2, e_3, e_4, \dots, e_n$ ถ่ายทอดจากด้านขวาไปด้านซ้าย มี W_{ij} เป็นน้ำหนักที่เชื่อมต่อระหว่างเซลล์ประสาทในชั้นอินพุตและชั้นกลาง และ W_{jk} เป็นน้ำหนักที่เชื่อมต่อระหว่างเซลล์ประสาทในชั้นอินพุตและชั้นเอาต์พุต

การถ่ายทอดสัญญาณข้อผิดพลาดเริ่มต้นที่ชั้นเอาต์พุตและย้อนกลับไปที่ชั้นกลางสัญญาณข้อผิดพลาดที่เอาต์พุตของเซลล์ประสาท k ที่ iteration เท่ากับ p อธิบายด้วยสัญลักษณ์ทางคณิตศาสตร์โดย

$$e_k(p) = y_{d,k}(p) - y_k(p) \qquad (6)$$

เมื่อ $y_{d,k}(p)$ คือเอาต์พุตที่ต้องการของเซลล์ประสาท k ที่ Iteration เท่ากับ p

เนื่องจาก k ถูกกำหนดในชั้นเอาต์พุตดังนั้นเราสามารถอัพเดทน้ำหนัก w_{jk} ได้โดยตรงและสามารถเขียนสมการคณิตศาสตร์การปรับปรุงน้ำหนักได้เป็น

$$w_{ik}(p+1) = w_{jk}(p) - \Delta w_{jk}(p) \qquad (7)$$

เมื่อ $\Delta w_{jk}(p)$ คือ การแก้ไขน้ำหนัก (Weight corrections)

การแก้ไขน้ำหนักสำหรับ Perceptron สามารถใช้สัญญาณอินพุต X_i แต่ถ้าโครงข่ายประสาทเทียมหลายชั้น อินพุตของเซลล์ประสาทในชั้นเอาต์พุตมีความแตกต่างจาก

อินพุตของเซลล์ประสาทในชั้นอินพุต การปรับปรุงน้ำหนักในโครงข่ายหลายชั้น ถูกคำนวนโดย [3] ดังสมการ

$$\Delta w_{jk}(p) = \alpha * y_j(p) * \delta_k(p) \quad (8)$$

เมื่อ $\delta_k(p)$ คือ Error gradient ที่เซลล์ประสาท k ในชั้นเอาต์พุต ที่ Iteration เท่ากับ p เขียนเป็นสมการคณิตศาสตร์ได้เป็น

$$\delta_k(p) = y_k(p) * [1 - y_k(\mathrm{p})] * e_k(p) \quad (9)$$

เมื่อ $y_k(\mathrm{p})$ คือ ผลลัพธ์ของเซลล์ประสาท k ที่ Iteration เท่ากับ p

เราสามารถอัพเดทน้ำหนัก w_{ij} ได้โดยตรงและสามารถเขียนสมการคณิตศาสตร์การปรับปรุงน้ำหนักได้เป็น

$$\Delta w_{ij}(p) = \alpha * x_j(p) * \delta_j(p) \quad (10)$$

เมื่อ $\delta_j(p)$ คือ Error gradient ที่เซลล์ประสาท j ในชั้นเอาต์พุต ที่ Iteration เท่ากับ p เขียนเป็นสมการคณิตศาสตร์ได้เป็น

$$\delta_j(p) = y_i(p) * \left[1 - y_j(\mathrm{p})\right] * \sum_{k=1}^{l} \delta_k(p) W_{ik}(p) \quad (11)$$

เมื่อ l คือจำนวนของเซลล์ประสาทในชั้นเอาต์พุต

$$y_i(p) = \frac{1}{1 + e^{-X_j(p)}}$$

$$X_j(p) = \sum_{i=1}^{n} x_i(p) * W_{ij}(p) - \Theta_j$$

และ n คือจำนวนเซลล์ประสาทในชั้นอินพุต

สรุปหลักการทำงานของ Back-propagation ได้ดังนี้

1) กำหนดค่าเริ่มต้น ได้แก่ น้ำหนัก และ Threshold ของโครงข่ายประสาทเทียม

2) กำหนด Activation function ที่ต้องการจากนั้นคำนวณผลลัพธ์ที่แท้จริงของเซลล์ประสาทชั้นกลาง และผลลัพธ์ที่แท้จริงของเซลล์ประสาทในชั้นเอาต์พุต

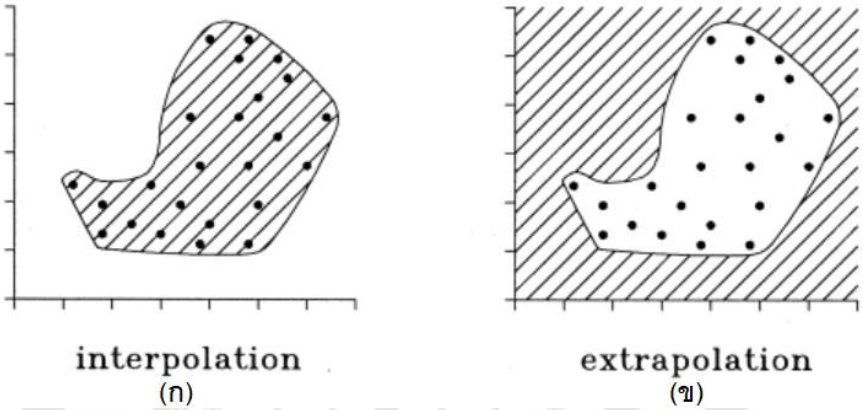
3) อัพเดตน้ำหนักในการถ่ายถอดย้อนกลับ (Back-propagation) ข้อผิดพลาดเซลล์ประสาทที่อยู่ในชั้นเอาต์พุตหลังจากนั้นคำนวน Error gradient คำนวณการแก้ไขน้ำหนัก (Weight corrections) และอัพเดตน้ำหนักที่เซลล์ประสาท โดยการคำนวนต่าง ๆ นั้นต้องคำนวนในชั้นอินพุตและชั้นเอาต์พุต

4) เพิ่ม Iteration ทีละหนึ่ง และกลับไปทำงานในขั้นตอน (2) และทำซ้ำจนกว่าข้อผิดพลาดของผลลัพธ์เป็นที่พึงพอใจ

2.3.2 เทคนิคการสร้างแบบจำลอง

2.3.2.1 ข้อจำกัดของ ANN

โครงข่ายประสาทเทียมสามารถสร้างแบบจำลองที่ไม่เป็นเชิงเส้นและถูกนำไปช่วยแก้ไขปัญหาที่ซับซ้อนได้ แต่ก็มีข้อจำกัดคือโครงข่ายประสาทสามารถเรียนรู้สร้างขอบเขตพื้นที่เกือบทุกฟังก์ชันกันการทำงานได้โดยการปรับค่าพารามิเตอร์ต่าง ๆ ตามข้อมูลชุดฝึกสอนแต่สำหรับขอบเขตพื้นที่ข้อมูลที่ไม่ได้ฝึกสอน ผลลัพธ์โครงข่ายประสาทที่ทำนายได้นั้นอาจไม่ถูกต้องแม่นยำประสิทธิภาพแย่ลง สำหรับข้อมูลที่อยู่ในกรอบพื้นที่เรียกว่า Interpolation ดังภาพประกอบ 15 (ก) และจุดอื่น ๆ ที่อยู่นอกพื้นที่เรียกว่า Extrapolation ดังภาพประกอบ 15 (ข)



ภาพประกอบ 15 (ก) จุดข้อมูลในพื้นที่ Interpolation และ (ข) จุดข้อมูลในพื้นที่ Extrapolation

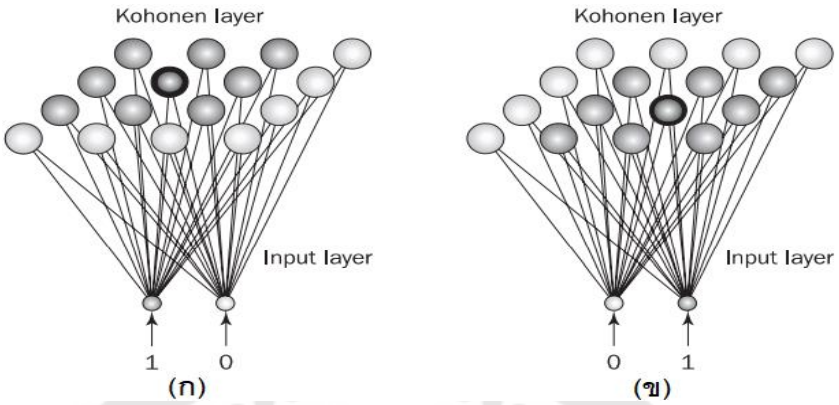
ที่มา: H. Lohninger, 2012, "Fundamentals of Statistics." สืบค้นจาก http://www.statistics4u.info/fundstat_eng/cc_ann_extrapolation.html, [5 ตุลาคม 2563]

2.3.2.2 Self-organizing map

ข้อจำกัดของโครงข่ายประสาทเทียม (Artificial neuron network) ตามที่กล่าวในข้อ 2.3.2.1 จึงนำเทคนิค Self-organizing map มาช่วยในการจัดกลุ่มของข้อมูลเพื่อทำให้ข้อมูลชุดฝึกสอน ข้อมูลชุดทดสอบ และ ข้อมูลชุดตรวจสอบ อยู่ในกลุ่มพื้นที่กลุ่มเดียวกัน

Self-organizing map เป็นโครงข่ายประสาทเทียมประเภทหนึ่ง จัดอยู่ในกลุ่มการเรียนรู้แบบมีไม่มีผู้ช่วยสอน (unsupervised learning) ถูกคิดค้นโดย Kohonen เมื่อปี 1990 สามารถจัดชุดข้อมูลที่มีลักษณะคล้ายกันคุณลักษณะเฉพาะรูปแบบอยู่ในกลุ่มเดียวกันหรืออยู่ในกลุ่มที่ใกล้เคียงกัน ในภาพประกอบ 16 โครงสร้างถูกออกแบบให้มี 2 ชั้นคือชั้นอินพุตและชั้นโครงข่าย Kohonen ประกอบด้วยเซลล์ประสาท 4 * 4 เซลล์ประสาทแต่ละตัวจะมีอินพุต 2

แบบ โดยเซลล์ประสาทที่ชนะจะมีลักษณะออกเป็นสีดำและเซลล์ประสาทที่อยู่ในเคียงกับเซลล์ประสาทที่ชนะจะจัดอยู่ในกลุ่มเดียวกัน

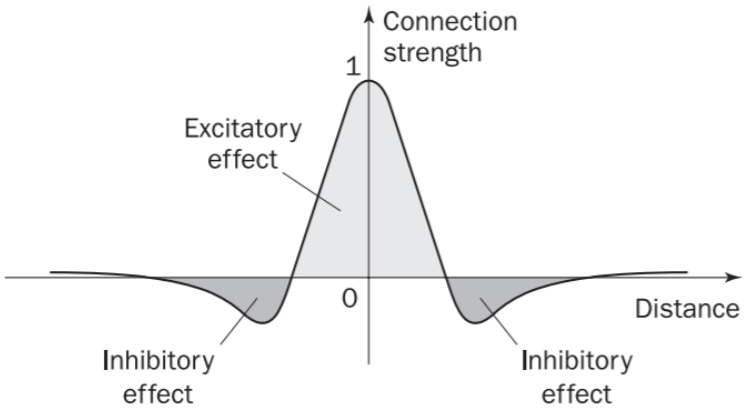


ภาพประกอบ 16 แผนที่คุณลักษณะแบบจำลอง Kohonen (ก) อินพุตตัวแรกมีค่าเท่ากับ 1 อินพุตตัวที่สองมีค่าเท่ากับ 0 (ข) อินพุตตัวแรกมีค่าเท่ากับ 0 อินพุตตัวที่สองมีค่าเท่ากับ 1

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

โครงข่าย Kohonen มีการเชื่อมต่อไปข้างหน้า (Forward connections) จากเซลล์ประสาทในชั้นไปยังชั้นเอาต์พุตและยังมีการเชื่อมต่อด้านข้าง (Lateral connections) ระหว่างเซลล์ประสาทในชั้นเอาต์พุต การเชื่อมต่อด้านข้างใช้เพื่อสร้างการแข่งขันระหว่างเซลล์ประสาทในชั้นเอาต์พุต เซลล์ประสาทตัวไหนที่มีระกับการกระตุ้นมากที่สุดจากเซลล์ประสาททั้งหมดจะเป็นผู้ชนะ

โดยการเชื่อมต่อระหว่างโหนดก่อให้เกิดการกระตุ้น (Excitatory effect) หรือการยับยั้ง (Inhibitory effect) ขึ้นอยู่กับระยะทางจากเซลล์ประสาทที่ชนะทำได้โดยการใช้ Mexican hat function ซึ่งอธิบายน้ำหนักกระหว่างเซลล์ประสาทในชั้น Kohonen



ภาพประกอบ 17 Mexican hat function ของการเชื่อมต่อระหว่างโหนด

ที่มา : M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.

Mexican hat function จากภาพประกอบ 17 พื้นที่จะแบ่งออกเป็น 3 ส่วนหลัก ๆ ประกอบด้วย 1) พื้นที่ใกล้เคียงคือบริเวณด้านข้าง มีแรงกระตุ้นผลกระทบมาก (Excitatory effect) 2) พื้นที่ใกล้เคียงระยะไกล คือพื้นที่มีผลยับยั้ง (Inhibitory effect) เล็กน้อย 3) พื้นที่ใกล้เคียงที่ห่างไกลมาก คือพื้นที่รอบ ๆ ตัวยับยั้ง ในโครงข่าย Kohonen น้ำหนักจะถูกเปลี่ยนให้เหมาะสมกับเซลล์ประสาทที่ชนะ สัญญาณเอาต์พุต $\boldsymbol{X}$ ที่ชนะ ถูกกำหนดให้มีค่าเท่ากับ 1 และสัญญาณเอาต์พุตของเซลล์ประสาทอื่น ๆ ทั้งหมดถูกมีค่าเป็น 0 ตามหลัก Standard competitive learning rule [4] กำหนดให้เป็นดังนี้

$$\Delta w_{ij} = \begin{cases} \alpha(x_j - w_{ij}), & if\ neuron\ j\ wins\ the\ competition \\ 0, & if\ neuron\ j\ loses\ the\ competition \end{cases} \quad (12)$$

เมื่อ x_j คือสัญญาณอินพุต

α คืออัตราการเรียนรู้ (Learning rate)

ภาพรวมกฎการเรียนรู้ของการแข่งขันขั้นอยู่กับการเปลี่ยนแปลงเวกเตอร์น้ำหนัก $\boldsymbol{W}_j$ ของเซลล์ประสาทชนะที่ j ต่อรูปแบบอินพุต $\boldsymbol{X}$ โดยเกณฑ์การจับคู่คือเทียบระยะทางต่ำสุดของยุคลิด (Euclidean distance) ระหว่างเวกเตอร์

$$d = \|\boldsymbol{X} - \boldsymbol{W}_j\| = [\sum_{i=1}^{n} (x_i - w_{ij})^2]^{1/2} \quad (13)$$

เมื่อ x_i และ w_{ij} เป็นองค์ประกอบของเวกเตอร์ $\boldsymbol{X}$ และ $\boldsymbol{W}_j$ ตามลำดับ

สรุปหลักการทำงานของ Self-organizing map network ได้ดังนี้

1) กำหนดค่าเริ่มต้น ได้แก่ กำหนดค่า น้ำหนัก เริ่มต้น และกำหนดค่าอัตราการเรียนรู้ (Learning rate)

2) เริ่มต้นการทำงานโครงข่าย Kohonen โดยใช้เวกเตอร์อินพุต X และค้นหาเซลล์ประสาทผู้ชนะโดยใช้วิธีหาระยะทางต่ำสุดของยุคลิด (Euclidean distance)

3) อัปเดตน้ำหนัก

4) เพิ่ม Iteration ทีละหนึ่ง และกลับไปทำงานในขั้นตอน

5) จนผลลัพธ์ระยะทางยูคลิด (Euclidean distance) เป็นที่พึงพอใจ

2.3.2.3 T-test

T-test เป็นเทคนิคทางสถิติที่ใช้ในการทดสอบสมมุติฐานเพื่อเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างกับประชากร หรือเปรียบเทียบระหว่างกลุ่มประชากร 2 กลุ่ม เช่น การเปรียบเทียบคะแนนสอบของนิสิตจำแนกตามเพศ เป็นต้น

2.3.2.4 F-test

F-test เป็นเทคนิคทางสถิติที่ใช้ในการทดสอบสมมุติฐานเพื่อเปรียบเทียบการกระจายข้อมูลของกลุ่มตัวอย่างกับประชากร หรือเปรียบเทียบระหว่างกลุ่มประชากร 2 กลุ่ม เช่น การเปรียบเทียบรายได้ของอาจารย์ในมหาวิทยาลัยกับอาจารย์ในโรงเรียน เป็นต้น

2.3.2.5 Range-test

Range-test เป็นเทคนิคการทดสอบกลุ่มข้อมูลที่พิจารณามีค่าสูงสุดและค่าต่ำสุดครอบคลุมกับอีกกลุ่มข้อมูลหนึ่งหรือไม่

2.3.2.6 เทคนิคในการเลือกจำนวนโหนดและจำนวนเลเยอร์

จำนวนชั้นกลาง (hidden layer) ของโครงข่ายประสาทเทียมไม่มีกฎเกณฑ์ที่แน่นอน มีหลากหลายทฤษฎี [5] โดย สามารถสรุปความสามารถของโครงข่ายประสาทเทียมที่มีจำนวนชั้นกลางต่าง ๆ กันได้ ดังตาราง 3

ตาราง 3 การกำหนดจำนวนชั้นกลาง (Middle layer)

จำนวนชั้นกลาง (hidden layer)	ผลลัพธ์
ไม่มี	สามารถทำนายผลหรือตัดสินใจผลลัพธ์ความสัมพันธ์เป็นแบบเชิงเส้นเท่านั้น
1	สามารถเริ่มทำนายผลหรือเริ่มตัดสินใจผลลัพธ์ที่มีความซับซ้อน มนุษย์เริ่มยากที่จะตัดสินใจได้
2	สามารถทำนายผลหรือตัดสินใจผลลัพธ์ความสัมพันธ์ที่มีความซับซ้อนและมีความถูกต้องแม่นยำมากขึ้น

ที่มา : J. Heaton, Introduction to Neural Networks for Java, 2nd ed. St. Louis: Heaton Research, Inc., 2008.

ถัดมาคือการตัดสินใจเลือกจำนวนโหนดในชั้นกลาง ซึ่งเป็นองค์ประกอบสำคัญอีกอย่างหนึ่งที่ต้องคำนึงถึงซึ่งมีผลต่อประสิทธิภาพการทำนายหรือการตัดสินใจ

การเลือกจำนวนโหนดน้อยเกินไปในชั้นกลางก็อาจทำให้เกิด Underfitting เนื่องจากมีจำนวนโหนดไม่เพียงพอในการประมวลผลข้อมูลที่มีความซับซ้อน [4]

ตรงกันข้ามกับการเลือกจำนวนโหนดมากเกินไปในชั้นกลางก็อาจทำให้เกิด Overfitting เนื่องจากมีจำนวนโหนดมากเมื่อนำโมเดลไปสอนด้วยชุดข้อมูลฝึกสอนแบบจำลองสามารถจดจำรายละเอียดต่าง ๆ ได้เกือบทั้งหมดได้เมื่อตรวจสอบผลลัพธ์ได้ความแม่นยำสูงมาก แต่เมื่อนำโมเดลนี้ทดสอบด้วยข้อมูลชุดตรวจสอบได้ความแม่นยำต่ำ

ปัจจุบันมีกฎจำนวนมากที่ใช้เป็นเกณฑ์ในการตัดสินใจเลือกจำนวนโหนดอย่างเช่น วิธี Rule-of-thumb [4] ที่กำหนดไว้ จำนวนโหนดที่ชั้นกลาง (hidden layer) ควรอยู่ระหว่างขนาดของชั้นอินพุต (input layer) และชั้นเอาต์พุต (output layer) จำนวนโหนดที่ชั้นกลาง (hidden layer) ควรมีขนาด 2/3 ของชั้นอินพุต (input Layer) รวมกับขนาดของชั้นเอาต์พุต (output layer) เป็นต้น

2.3.2.7 วิธีการทำ Normalization

การ Normalization คือการปรับข้อมูลให้อยู่สเกลเดียวกันเพื่อนำข้อมูลไปใช้สร้างแบบจำลองไม่ให้เกิดไบแอสกับอินพุตพารามิเตอร์ตัวใดตัวหนึ่ง และช่วยลดเวลาในในการสร้างโมเดลเพราะข้อมูลมีค่าน้อยลงซึ่งวิธีการทำ Normalization มีหลายวิธีดังนี้

1) Min-max normalization ข้อมูลถูกสเกลให้มีค่าอยู่ในช่วงระหว่าง 0 และ 1 โดยคำนวณจากค่าสูงสุดและต่ำสุดจากข้อมูลทั้งหมดกำหนดให้ x แทนข้อมูลที่สนใจและ x' แทนข้อมูลที่ถูกสเกลซึ่งเขียนเป็นสมการได้ดังนี้

$$x' = \frac{x - Min(x)}{Max(x) - Min(x)} \quad (14)$$

2) Mean normalization ข้อมูลถูกสเกลคล้ายกับวิธีก่อนหน้านี้แต่เปลี่ยนจากการคำนวณด้วยค่าต่ำสุดเป็นค่าเฉลี่ยแทนเขียนเป็นสมการได้ดังนี้

$$x' = \frac{x - Mean(x)}{Max(x) - Min(x)} \quad (15)$$

3) Z-Score normalization หรือ Standardization เป็นวิธีการสเกลข้อมูลให้มีค่าเฉลี่ยเท่ากับ 0 และ ค่าความแปรปรวนเท่ากับ 1 กำหนดให้ σ แทนค่าเบี่ยงเบนมาตราฐานและ z แทน Standard score เขียนเป็นสมการได้ดังนี้

$$z = \frac{x - Mean(x)}{\sigma} \quad (16)$$

2.3.2.8 การเลือก Activation function

Activation function เป็นส่วนหนึ่งของโครงข่ายประสาทเทียมทำหน้าที่ประมวลผลโดยรับผลรวมจากทุกอินพุตใน 1 โหนด แล้วพิจารณาว่าควรส่งต่อเอาต์พุตมีค่าเท่าใดซึ่งมีหลายประเภท โดยสามารถสรุปคุณสมบัติของ Activation function ได้ดังนี้

1) Monotonic คือพิจารณากราฟ Activation function ใด ๆ มีแนวโน้มไปในทิศทางเดียวกล่าวคือมีแนวโน้มสูงขึ้นหรือต่ำลงอย่างใดอย่างหนึ่ง

2) One to one corresponding mapping คือพิจารณาจากกราฟ Activation function ใด ๆ ค่า x หนึ่งค่าให้จะให้ค่า y เพียงหนึ่งค่าเท่านั้น

3) High sensitivity คือพิจารณาจากกราฟ Activation function ใด ๆ เปรียบเทียบค่า x_1 ให้ค่า y_1 และเมื่อพิจารณาที่ x_2 ให้ค่า y_2 เมื่อนำค่า $y_1 - y_2$ จะให้ผลลัพธ์ที่แตกต่างกันมาก

2.4 งานวิจัยที่เกี่ยวข้อง

ในปัจจุบันมีงานวิจัยที่แสดงการทำนายปริมาณสารสร้างตะกอนที่เหมาะสมที่ช่วยให้น้ำดิบตกตะกอนอย่างมีประสิทธิภาพด้วยวิธีการทดสอบจาร์เทส (Jar test) หรืองานที่มีลักษณะเกี่ยวข้อง ตัวอย่างเช่น

2.4.1 Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters

งานวิจัยนี้ใช้โครงข่ายประสาทเทียม (Artificial neuron network) ในการทำนายปริมาณสารส้มที่เหมาะสมและพารามิเตอร์คุณภาพน้ำที่ได้รับการบำบัดแล้ว [6] โดยรวบรวมข้อมูลมาจากสถานที่หลายแห่งในรัฐวิคตอเรียและรัฐเซาท์ออสเตรเลีย ทั้งหมด 202 แถว แบ่งเป็นชุดข้อมูลฝึกอบรบ 126 แถว ชุดข้อมูทลทดสอบ 44 แถว และ ชุดข้อมูลตรวจสอบ 32 แถว

จุดประสงค์ของงานวิจัยนี้คือ 1) ทำนายคุณภาพน้ำที่ได้รับการบำบัดแล้วโดยใช้ข้อมูลน้ำดิบ และข้อมูลสารส้ม 2) ทำนายปริมาณสารส้มโดยใช้ข้อมูลของน้ำดิบและข้อมูลน้ำที่ได้รับการบำบัดแล้ว และ 3) สร้างโมเดลที่ใช้งานง่ายและควบคุมปริมาณสารส้มโดยอัตโนมัติเพื่อที่จะทำนายคุณภาพน้ำที่ได้รับการบำบัดแล้ว งานวิจัยนี้ได้สร้างโมเดลขึ้น มา 2 โมเดล โมเดลที่ 1 ทำนายค่าพารามิเตอร์ ความขุ่นของน้ำ,สีของน้ำและ UVA-254 มีค่า R^2 เท่ากับ 0.90,0.92 และ 0.98 ตามลำดับ และมีค่า MAE เท่ากับ 0.12 NTU,1.45 HU และ 0.01 cm-1 โมเดลที่ 2 ทำนายค่าพารามิเตอร์ อลูมิเนียมที่เหลือและ ค่าความเป็นกรด-เบส มีค่า R^2 เท่ากับ 0.96,0.85 ตามลำดับและมี ค่า MAE เท่ากับ 0.02 mg/l และ 0.11 ตามลำดับ และโมเดลที่ 3 เพื่อทำนายปริมาณสารส้ม มีค่า R^2 เท่ากับ 0.94 และ ค่า MAE 3.2 mg/l นอกจากนี้มีการปรับใช้โมเดลและสร้าง GUI (Graphical user interface) โดยใช้โปรแกรม LabVIEW เครื่องมือ Sim TTP และ Sim WT

2.4.2 Clarifier Models using Artificial Neural Networks— Case Study: Bangkhen Water Treatment Plant

งานวิจัยนี้เป็นการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียม สำหรับการทำนายค่าความขุ่น (Turbidity) ที่ออกจากแคริไฟเออร์ [7] ด้วยข้อมูลการปฏิบัติงานในรูปแบบของอนุกรมเวลา Clarifier เป็นกระบวนการกำจัดความขุ่นของน้ำซึ่งส่งผลต่อประสิทธิภาพผลิตน้ำประปา ในงานวิจัยนี้แบ่งข้อมูลให้มีอัตราส่วนของจำนวณข้อมูล Training set: Test set: Validation set เท่ากับ 4:1:1 แบบจำลอง ANN ที่มีชั้นซ้อนจำนวนหนึ่งชั้นมีค่า R^2 เท่ากับ 0.89 และค่า MAE 0.018 NTU สำหรับแบบจำลองโมเดลที่มีชั้นซ้อนจำนวนสองชั้น R^2 เท่ากับ 0.92 และค่า Mean

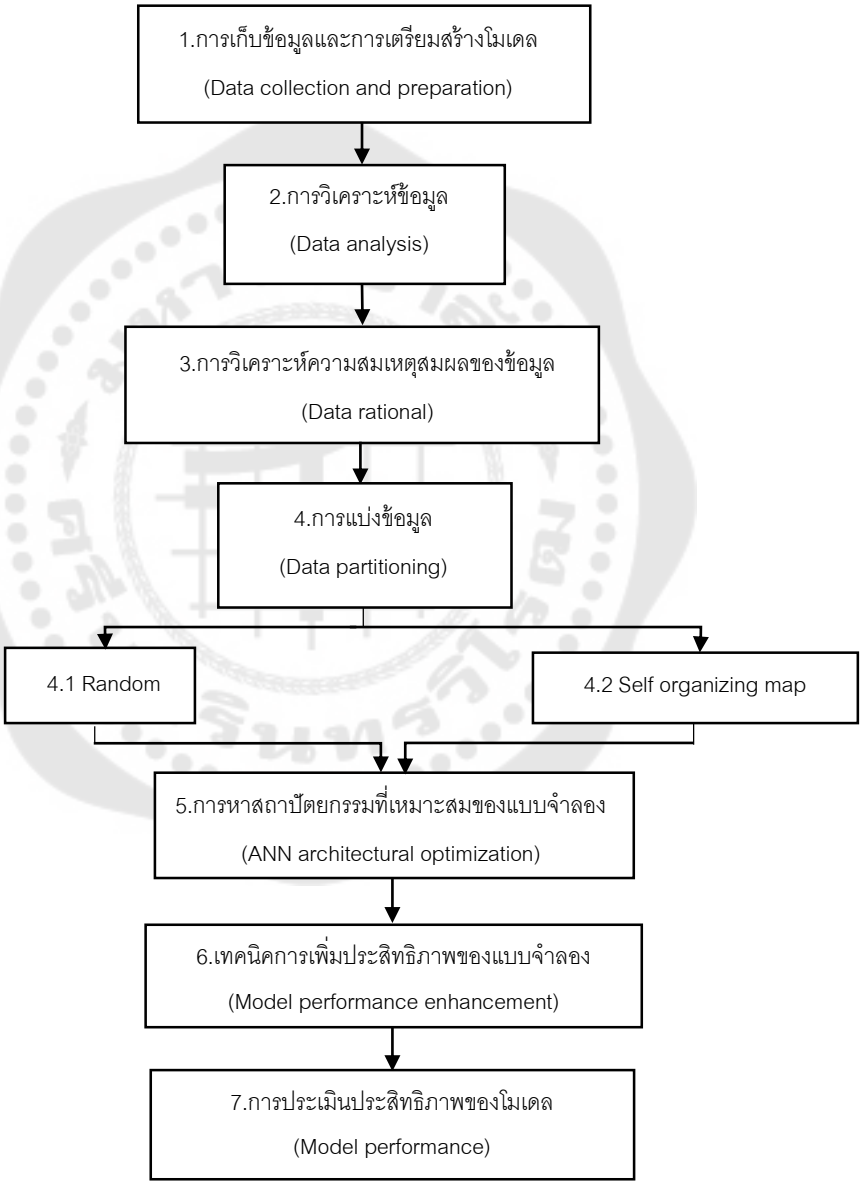
Absolute Error 0.016 NTU ซึ่งจะเห็นว่า แบบจำลอง ANN ที่มีสองชั้นให้ผลดีกว่าที่เวลาย้อนหลัง 8 ชั่วโมง ให้ประสิทธิภาพในการทำนายดีที่สุด

2.4.3 Prediction of Optimal Coagulant Dosage in Drinking Water Treatment by Artificial Neural Network

งานวิจัยนี้ใช้โครงข่ายประสาทเทียม (Artificial neuron network) สร้างแบบจำลองเพื่อทำนายพารามิเตอร์ของน้ำที่ผ่านการบำบัดแล้วและปริมาณการตกตะกอนที่เหมาะสม [8] ข้อมูลที่นำมาใช้ในการสร้างแบบจำลองรวบรวมมาจากข้อมูลการปฏิบัติงานจริงจากน้ำดื่มโรงงานบำบัดน้ำในอิสตันบูล (Istanbul) โดยใช้โปรแกรม NeuroShell® Predictor ซึ่งอัลกอริทึมเรียนรู้เฉพาะชื่อว่า TurboProp2 โดยไม่จำเป็นต้องกำหนดพารามิเตอร์ที่ใช้ในโครงข่ายประสาทเทียมเช่น Learning rate, Activation function เป็นต้นและอีกทั้งไม่ต้องปรับสเกลอินพุตพารามิเตอร์ให้อยู่ในช่วงเดียวกันโดยโปรแกรม โปรแกรม NeuroShell® Predictor จะทำการประมวลผลล่วงหน้า (preprocessing) โดยในงานวิจัยสร้างแบบจำลองทั้งหมด 2 แบบจำลอง แบบจำลองแรกใช้ในการทำนายพารามิเตอร์น้ำที่ผ่านการบำบัดมีอินพุต ได้แก่ ความเป็นกรด-ด่าง (pH), ความขุ่น (turbidity), การนำไฟฟ้า (conductivity) , สี (color), UV_{254}, ฟลูออไรด์ (fluoride) และปริมาณสารส้ม (alum dosage) และแบบจำลองที่สอง ทำนายปริมาณสารส้มที่เหมาะสมด้วยอินพุตลักษณะเดียวกับแบบจำลองแรก แต่ไม่รวมปริมาณสารส้ม

บทที่ 3
วิธีดำเนินการวิจัย

ในการวิจัยนี้ ผู้วิจัยได้ดำเนินการขั้นตอน ดังภาพประกอบ 18 มีทั้งหมด 7 ขั้นตอนซึ่งในส่วนของรายละเอียดจะกล่าวในหัวข้อถัดไป



ภาพประกอบ 18 แผนผังวิธีดำเนินการวิจัย

3.1 **การเก็บข้อมูลและการเตรียมสร้างโมเดล** (Data collection and preparation)

ข้อมูลได้รับความอนุเคราะห์จากผู้อำนวยการโรงงานผลิตน้ำบางเขนซึ่งเป็นข้อมูลคุณภาพน้ำดิบตั้งแต่วันที่ 1 มกราคม 2019 ถึงวันที่ 31 ธันวาคม 2019 ข้อมูลช่วงหนึ่งปีที่นำมาใช้เพื่อให้มั่นใจว่าข้อมูลนี้ครอบคลุมฤดูกาลทั้งหมดในประเทศไทยได้แก่ ฤดูร้อน ฤดูฝนและฤดูหนาว ข้อมูลทั้งหมดที่เกี่ยวข้องกับการทดสอบจาร์เทสจัดเก็บรวมกันอยู่ในไฟล์ Excel เป็น 2 กลุ่ม ได้แก่ 1) ปริมาณสารเคมีประกอบไปด้วยสารส้มและโพลีอลูมิเนียมคลอไรด์ถูกบันทึก 2 ส่วนคือ Recommended dosage ซึ่งได้จากการทดสอบจาร์เทสเก็บในรูปแบบ XX-YY โดย XX เป็นปริมาณสารเคมีสูงสุดและ YY คือปริมาณสารเคมีต่ำสุด และส่วนที่สอง Used เป็นปริมาณสารเคมีจริงที่เติมในกระบวนการผลิตน้ำ 2) คุณสมบัติของน้ำดิบประกอบไปด้วย Turbidity, pH, Alkalinity, Conductivity และ Oxygen consumption ข้อมูลบางอย่างสามารถได้มาจากเซนเซอร์ที่ติดตั้งไว้อยู่หลายตัวโดยบางข้อมูลต้องรวบรวมด้วยตนเองโดยอธิบายการบันทึกข้อมูลและเวลาการเก็บรวบรวมข้อมูลไว้ในตาราง 4 และพารามิเตอร์ทั้งหมดอธิบายไว้ในตาราง 5 โดยมีจำนวนข้อมูลทั้งหมด 730 แถว

เนื่องจาก BKWTP มีการเก็บข้อมูลอยู่ในรูปแบบ 2 ประเภทได้แก่ 1) ข้อมูลแบบออนไลน์ในแต่ละวันจะทำการวัดทุก 4 ชั่วโมงโดยวิธีการ Lookup table คือมองด้วยตาแล้วจดบันทึกพารามิเตอร์คุณภาพน้ำดิบ 2) ข้อมูลแบบออฟไลน์เป็นค่าที่จำเป็นต้องทำการทดลองผ่านอุปกรณ์วิทยาศาสตร์ เช่น การทดสอบจาร์เทสเพื่อหาปริมาณสารส้มหรือปริมาณโพลิอลูมิเนียมคลอไรด์ที่เหมาะสม

ตาราง 4 การบันทึกข้อมูลและเวลาการเก็บรวบรวมข้อมูล

Parameter	Recommended dosage	Used	Time					
			00:00	04:00	08:00	12:00	16:00	20:00
Turbidity	-	-	x	x	x	X	X	x
pH	-	-	x	x	x	x	X	x
Conductivity	-	-	x	x	x	x	x	x
Alkalinity	-	-	X	x	x	x	x	x
OC	-	-	x	x	x	x	x	x
Alum	20-25	25	-	x	-	-	x	-
PACL	10-12	11	-	x	-	-	x	-

โดยเซนเซอร์ออนไลน์เหล่านั้นมักจะอุดตันและหยุดชะงักจากการผลิตน้ำจำนวนมากต่อวัน ดังนั้นจึงต้องมีการสอบเทียบเซนเซอร์เป็นประจำ ข้อผิดพลาดของข้อมูลส่วนใหญ่มาจากข้อมูลที่รวบรวมในช่วงเวลาการสอบเทียบหรือความผิดพลาดของมนุษย์ เพื่อแก้ไขข้อผิดพลาดของข้อมูล เกณฑ์การกรองข้อมูลจะถูกกำหนดมาจากการสัมภาษณ์ของผู้ปฏิบัติงานในงาน [9] และได้แสดงไว้ในตาราง 6 ถ้าค่าหากข้อมูลอินพุตไม่ได้อยู่ตามเงื่อนไขผู้วิจัยก็จะทำการตัดข้อมูล Row นั้นออกทั้ง Row ไม่มีการแก้ไขค่า

ตาราง 5 พารามิเตอร์ทั้งหมดของงานวิจัยนี้

ลำดับ	พารามิเตอร์	หน่วย	คำอธิบาย [10]	ตัวอย่าง	Online/Offline
1	pH	pH	เป็นตัววัดค่าความเป็นกรด/ด่างของน้ำ	7.9	Online
2	Alkalinity	mg/L	เป็นตัววัดความต้านทานการเปลี่ยนแปลงความเป็นกรดของน้ำ	101	Online
3	Turbidity	NTU	เป็นตัววัดความบริสุทธิ์ของน้ำ ว่ามีอนุภาคอื่นปนเปื้อนมากน้อยแค่ไหนอาจจะมองเห็นหรือไม่เห็นด้วยตาเปล่า ซึ่งเป็นตัววัดคุณภาพน้ำที่สำคัญ	19	Online/Offline
4	Conductivity	µs/cm	เป็นตัวที่ใช้วัดความนำไฟฟ้าของน้ำ	284	Online
5	Oxygen consumption	mg/L	(ออกซิเจนคอนซูม) เป็นตัววัดค่าปริมาณการปนเปื้อนของสารอินทรีย์ในน้ำ	3.27	Online
6	Alum dosage	mg/L	ค่าเฉลี่ยของ Alum ที่ใช้มากที่สุดและน้อยที่สุดแนะนำจากการทดสอบ Jar test หารด้วย 2	20	Offline

ตาราง 5 (ต่อ)

ลำดับ	พารามิเตอร์	หน่วย	คำอธิบาย [10]	ตัวอย่าง	Online/Offline
7	polyaluminum chloride dosage	mg/L	ค่าเฉลี่ยของ PACL ที่ใช้มากที่สุดและน้อยที่สุดแนะนำจากการทดสอบ Jar test หารด้วย 2	25	Offline

ตาราง 6 เกณฑ์กรองข้อมูลโรงผลิตน้ำบางเขน

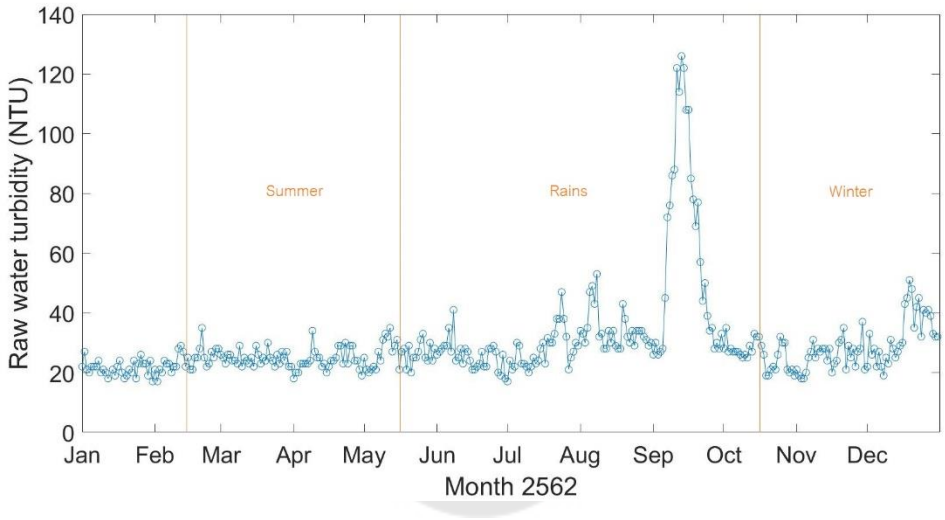
พารามิเตอร์	เกณฑ์
Turbidity (NTU)	> 15
pH	6.8 - 8.5
Conductivity (μS/cm)	> 90
Alkalinity (mg/L)	35 – 125
Oxygen consumption (OC) (mg/L)	< = 5.5

3.2 **การวิเคราะห์ข้อมูล** (Data analysis)

ข้อมูลที่ผ่านการกรองแล้วเหลือจำนวน 717 แถว ซึ่งถูกนำมาวิเคราะห์ค่าสถิติดังแสดงในตาราง 7 และแสดงแนวโน้มของความขุ่นของน้ำดิบดังภาพประกอบ 20 เพื่อแสดงการเปลี่ยนแปลงของน้ำดิบตลอดทั้งปี พบว่าความขุ่นสูงสุด 126 NTU ในช่วงฤดูฝนในเดือนกันยายน เห็นได้ชัดว่าค่าเบี่ยงเบนมาตรฐานเท่ากับ 14.86 ในขณะที่ค่าเฉลี่ย 27.42 แสดงให้เห็นว่ามีการเปลี่ยนแปลงอย่างมากในความขุ่นของน้ำดิบ จากสถานการณ์เหล่านี้การคาดการณ์สารเคมีที่ใช้ในโรงงานอย่างแม่นยำจึงจำเป็นต้องมีการติดตามบ่อยขึ้นซึ่งเป็นขีดจำกัดของการใช้จาร์เทสแบบเดิม

ตาราง 7 ค่าสถิติของพารามิเตอร์น้ำดิบ

	Raw water					Alum	PACL
	Turbidity	Alkalinity	pH	Conductivity	OC		
Max	126	115	8.2	3500	5.42	52.5	29
Min	15	69	7.25	201	1.62	12.5	0
Mean	27.42	97.14	7.7	361.16	3.67	26.45	11.23
Median	24	97	7.7	322	3.62	27.5	13
Range	111	46	0.95	3299	3.8	40	29
STD	14.86	7.19	0.16	237.10	0.74	7.11	6.61

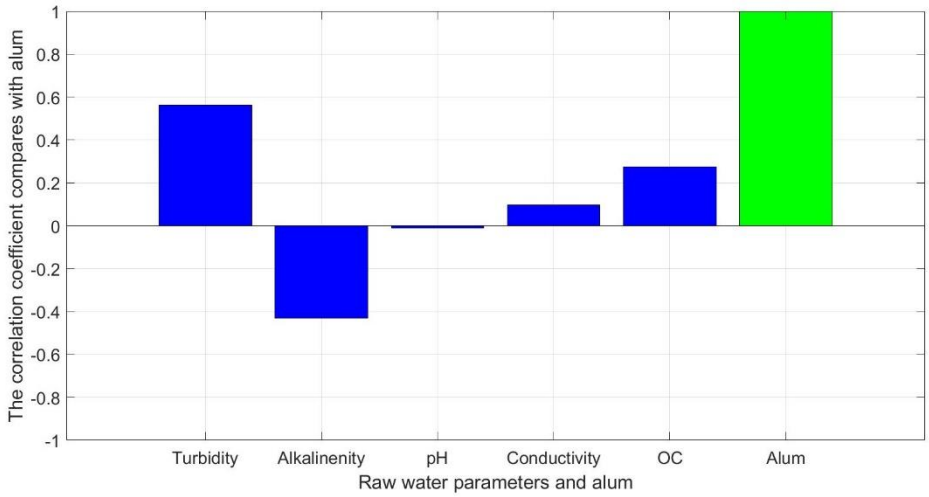


ภาพประกอบ 19 แนวโน้มความขุ่นของน้ำดิบสูงสุดรายวัน

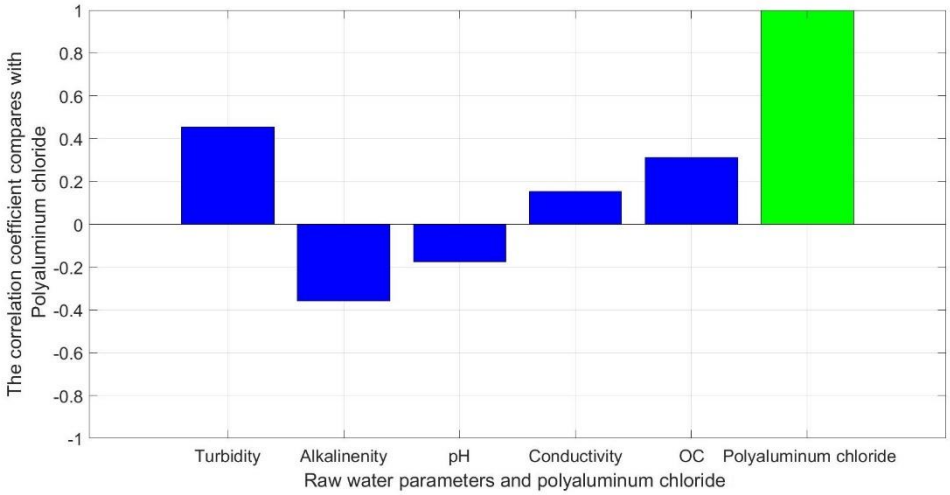
3.3 **การวิเคราะห์ความสมเหตุสมผลของข้อมูล** (Data rational)

ภาพประกอบ 21 และ 22 แสดงค่าสัมประสิทธิ์สหสัมพันธ์หรือค่า Correlation coefficient (R) ของพารามิเตอร์คุณภาพน้ำกับค่าปริมาณสารส้ม (alum) และโพลีอะลูมิเนียมคลอไรด์ (PACL) ตามลำดับ ค่า R ของสารส้มมีค่าน้อยมากอยู่ในช่วง -0.43 ถึง 0.56 กรณีที่ค่า R เข้าใกล้ศูนย์สะท้อนพฤติกรรมไม่เป็นเชิงเส้น (non-linear) อย่างไรก็ตามข้อมูลที่เป็นเหตุเป็นผลสามารถตรวจสอบได้จากค่า R เช่น ถ้าน้ำดิบมีความขุ่นสูง ปริมาณของสารส้มและโพลิอลูมิเนียม

คลอไรด์ ที่ต้องการปรับคุณภาพน้ำก็สูงเช่นกันความสัมพันธ์มีแนวโน้มไปทิศทางเดียวกัน ในทางกลับกันความเป็นด่าง (alkalinity) และ pH จะแปรผกผันกับสารส้มเนื่องจากการเติมสารส้มจะทำให้ pH และความเป็นด่าง (alkalinity) ลดลง



ภาพประกอบ 20 ค่าสัมประสิทธิ์สหสัมพันธ์ของพารามิเตอร์คุณภาพน้ำดิบเทียบกับสารส้ม



ภาพประกอบ 21 ค่าสัมประสิทธิ์สหสัมพันธ์ของพารามิเตอร์คุณภาพน้ำดิบเทียบกับโพลิอลูมิเนียมคลอไรด์

3.4 **การแบ่งข้อมูล** (Data partitioning)

ในขั้นตอนนี้ข้อมูลจะถูกแบ่ง 2 วิธีได้แก่ 1) การแบ่งข้อมูลโดยใช้ Self organizing map (SOM) เป็น ANN ประเภทการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) สำหรับการจัดกลุ่มข้อมูลซึ่งจะได้กล่าวรายละเอียดในหัวข้อ 3.6 2) สุ่มแบบ Random แบ่งข้อมูลออกเป็น 3 กลุ่ม คือกลุ่มข้อมูลฝึกสอน (training set) กลุ่มข้อมูลทดสอบ (test set) และกลุ่มข้อมูลตรวจสอบ (validation set) โดยมีอัตราส่วน 1/3 , 1/6 และ 1/6 ตามลำดับจึงได้แบ่งกลุ่มข้อมูลฝึกสอนจำนวน 431 รายการ กลุ่มข้อมูลทดสอบจำนวน 143 รายการและกลุ่มข้อมูลตรวจสอบจำนวน 143 รายการ โดยวัดประสิทธิภาพแบบจำลองด้วยกลุ่มข้อมูลทดสอบและเพื่อป้องกัน Overfitting จะใช้ชุดข้อมูลตรวจสอบสำหรับ Early stopping criteria

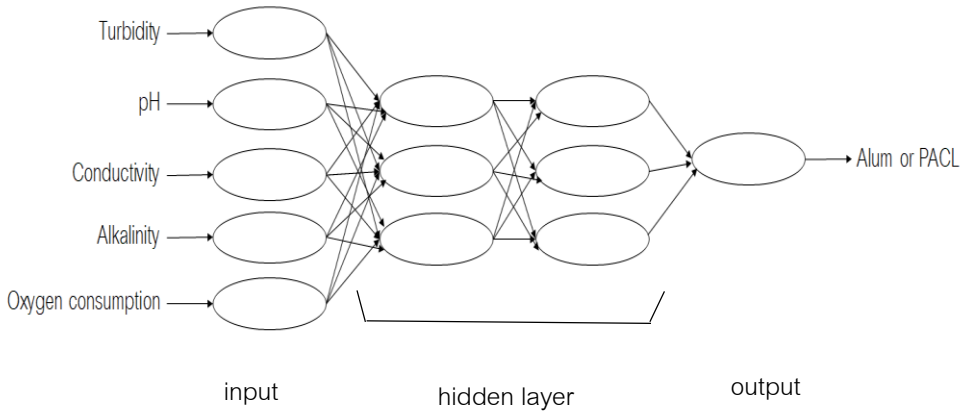
3.5 **การหาสถาปัตยกรรมที่เหมาะสมของแบบจำลอง** (ANN architectural optimization)

สถาปัตยกรรมแบบจำลองที่เหมาะสมที่สุด (จำนวนชั้นซ้อนและจำนวน Node ใน Hidden layer) สามารถพบได้โดยใช้วิธีการทดลองลองถูกลองผิด (trial and error) อย่างเป็นระบบ [7] แบบจำลองที่ดีที่สุดคือมีค่าสัมประสิทธิ์แสดงการตัดสินใจ (Coefficient of determination, R^2) สูงที่สุดและค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean absolute error, MAE) น้อย ๆ ในงานวิจัยนี้มีการทดสอบจำนวนชั้นซ้อนไม่เกิน 2 ชั้น ในทางกลับกัน กระบวนการ Trial – error ถูกดำเนินการอย่างเป็นระบบโดยการเพิ่มจำนวนโหนดในแต่ละชั้นซ้อนเริ่มต้นด้วย 5 โหนดและค่อย ๆ เพิ่มทีละ 5 โหนด รวมเป็น 20 ค่า ได้แก่ 5,10,15, …,100

ตัวอย่างของสถาปัตยกรรมเครือข่ายประสาทเทียมแบบ 2 ชั้นซ้อน แสดงในภาพประกอบ 23 ในงานวิจัยนี้มี ANN model 2 แบบคือ 1) แบบจำลอง ANN สำหรับทำนายปริมาณสารส้มและ 2) แบบจำลอง ANN สำหรับทำนายปริมาณโพลิอลูมิเนียมคลอไรด์ อินพุตและเอาต์พุตของแต่ละแบบจำลองแสดงในตาราง 8

ตาราง 8 อินพุต (I) และเอาต์พุต (O) สำหรับพัฒนาแบบจำลอง ANN

ANN model	Raw water					Alum dosage	polyaluminum chloride dosage
	pH	Alkalinity	Turbidity	Conductivity	OC		
1	I	I	I	I	I	O	-
2	I	I	I	I	I	-	O

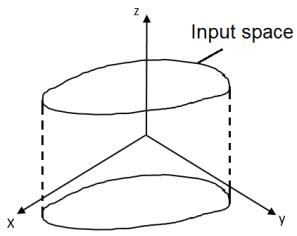


ภาพประกอบ 22 สถาปัตยกรรมตัวอย่างชั้นซ้อน 2 ชั้น

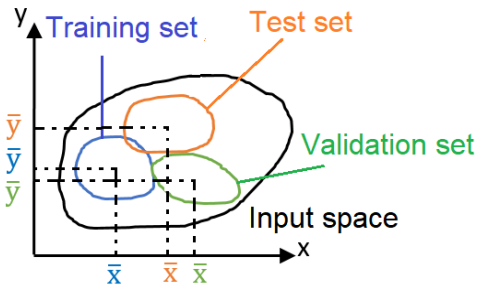
3.6 **เทคนิคการเพิ่มประสิทธิภาพของแบบจำลอง** (Model performance enhancement)

เทคนิคการจัดกลุ่มข้อมูล (clustering) ที่ช่วยให้กลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบ และกลุ่มข้อมูลตรวจสอบ เป็นตัวแทนซึ่งกันและกันโดยตรวจสอบจากการทดสอบสมมุติฐาน F-test และ T-test และทำให้ช่วงกลุ่มข้อมูลฝึกสอนครอบคลุมกลุ่มข้อมูลทดสอบและกลุ่มข้อมูลตรวจสอบ

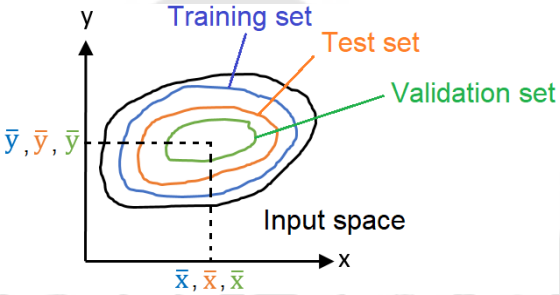
เมื่อเปรียบเทียบวิธีการแบ่งข้อมูลระหว่างวิธี Random กับ SOM โดยมี Input space ดังภาพประกอบ 23 ในระนาบ xyz แล้ว Projection ลงแกน x และ แกน y พบว่าการแบ่งข้อมูลแบบ Random ไม่การันตีว่ากลุ่มข้อมูลแต่ละกลุ่มเป็นตัวเดียวกันจะเห็นได้ชัดเจนว่าค่า $\bar{x}$ และ $\bar{y}$ ไม่เท่ากันดังภาพประกอบ 24 และภาพประกอบ 25 เป็นการแบ่งข้อมูลแบบ SOM ซึ่งทำให้ได้ ค่า $\bar{x}$ และ $\bar{y}$ ที่ใกล้เคียงกันหรือเท่ากันอีกทั้งกลุ่มข้อมูลแต่ละกลุ่มจะเป็นตัวแทนซึ่งกันและกัน ตรวจสอบได้จากการทดสอบ F-test null hypothesis เป็นการทดสอบสมมุติฐานเพื่อเปรียบเทียบค่าความแปรปรวนของกลุ่มข้อมูล 2 กลุ่มและ T-test null hypothesis เป็นการทดสอบสมมุติฐานเพื่อเปรียบเทียบค่าเฉลี่ยของกลุ่มข้อมูล 2 กลุ่ม



ภาพประกอบ 23 พื้นที่อินพุตข้อมูลในระนาบ xyz

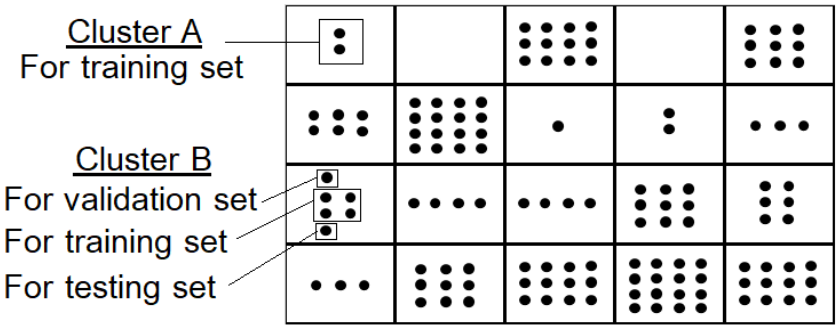


ภาพประกอบ 24 ค่าเฉลี่ย $\bar{x}$ และ $\bar{y}$ การแบ่งข้อมูลแบบ Random



ภาพประกอบ 25 ค่าเฉลี่ย $\bar{x}$ และ $\bar{y}$ การแบ่งข้อมูลแบบ SOM

วิธีการแบ่งข้อมูลแบบ SOM เป็นไปดังภาพประกอบ 26 ตารางสี่เหลี่ยมแทนเซลล์ประสาทในชั้น Kohonen จุดสีดำแทนจำนวนข้อมูลที่มีอยู่ในแต่ละ Cluster ในตัวอย่างนี้ Cluster A มีข้อมูลเพียง 2 จำนวนถูกกำหนดให้เป็นกลุ่มข้อมูลฝึกสอนและสำหรับ Cluster B มีจำนวนข้อมูลทั้งหมด 6 จะสุ่มเลือกอย่างละ 1 ให้เป็นกลุ่มข้อมูลทดสอบและกลุ่มข้อมูลตรวจสอบส่วนที่เหลือกำหนดเป็นกลุ่มข้อมูลฝึกสอน โดยทำกระบวนการแบบนี้ซ้ำจนครบทุก Cluster



ภาพประกอบ 26 กระบวนการสุ่มตัวอย่างที่ใช้ในการแบ่งข้อมูล SOM

3.7 การประเมินประสิทธิภาพของโมเดล (Model performance)

แบบจำลอง ANN ทั้งหมดถูกประเมินโดยใช้ค่าสัมประสิทธิ์แสดงการตัดสินใจ (Coefficient of determination, R^2) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean absolute error, MAE)

R^2 คือการวัดที่ใช้อธิบายความแปรปรวนของปัจจัยหนึ่งที่อาจเกิดความสัมพันธ์กับปัจจัยอื่นที่เกี่ยวข้อง ความสัมพันธ์นี้เรียกว่า "Goodness of fit" มีค่าอยู่ระหว่าง 0.0 – 1.0

MAE เป็นค่าเฉลี่ยกลุ่มข้อมูลทดสอบของความแตกต่างสัมบูรณ์ระหว่างค่าจาการทำนายและค่าจริง

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (17)$$

โดยที่ SS_{res} คือค่าผลรวมกำลังสองของ Residuals

SS_{tot} คือค่าผลรวมกำลังสองทั้งหมด

$$SS_{res} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (18)$$

$$SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \quad (19)$$

โดยที่ y_i คือค่าเอาต์พุตแท้จริง

$\hat{y}_i$ คือค่าจากการทำนายของแบบจำลอง

$\bar{y}_i$ คือค่าเฉลี่ยเอาต์พุตแท้จริง

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (20)$$

โดยที่ y_i คือค่าเอาต์พุตแท้จริง

$\hat{y}_i$ คือค่าจากการทำนายของแบบจำลอง

บทที่ 4
ผลลัพธ์การวิจัย

4.1 การหาสถาปัตยกรรมที่เหมาะสม

ค่าสัมประสิทธิ์แสดงการตัดสินใจ (Coefficient of determination, R^2) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean absolute error, MAE) ของแบบจำลองที่แบ่งข้อมูลเป็นกลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบและกลุ่มข้อมุลตรวจสอบด้วยวิธีการสุ่ม (Random) ผลลัพธ์ของ (a) MAE และ (b) R^2 ของแบบจำลองแสดงในภาพประกอบ 27 และ 28 จะเห็นได้ว่าแบบจำลองทำนายปริมาณสารส้มที่มีค่า R^2 มากที่สุดคือ 0.73 และ MAE ต่ำสุดที่ 3.18 มิลลิกรัม/ลิตร ที่แบบจำลอง (25 โหนดที่มีเลเยอร์ชั้นซ่อนหนึ่งชั้น) ดังภาพประกอบที่ 27 ในทางกลับกันภาพประกอบที่ 28 แสดงให้เห็นว่าแบบจำลองทำนายปริมาณ PACL โดยมี (a) MAE ที่ 3.21 มิลลิกรัม/ลิตร (b) R^2 เท่ากับ 0.59 และที่โมเดล (5 โหนดที่มีเลเยอร์ชั้นซ่อนหนึ่งชั้น)

สำหรับโมเดลประเภทเลเยอร์ชั้นซ้อนสองชั้น ภาพประกอบ 29 แสดงให้เห็นถึงการประเมินประสิทธิภาพแบบจำลองสำหรับการทำนายปริมาณสารส้มซึ่งแสดงให้เห็นว่า (a) MAE ต่ำสุดคือ 2.94 มิลลิกรัม/ลิตร (b) R^2 สูงสุด 0.66 (AA-BB คือจำนวนโหนดในชั้นแรกและชั้นที่สองตามลำดับ) ภาพประกอบ 30 แสดงการประเมินแบบจำลองการทำนาย PACL (a) MAE ต่ำสุด 3.25 มิลลิกรัม/ลิตร (b) R^2 สูงสุด 0.57 ที่แบบจำลอง (15-5) ภาพประกอบ 31 และ 32 แสดงให้เห็นถึงปริมาณสารส้มและ PACL ที่ทำนายและจริงในชุดทดสอบตามลำดับ

4.2 ผลการแบ่งข้อมูลแบบสุ่ม

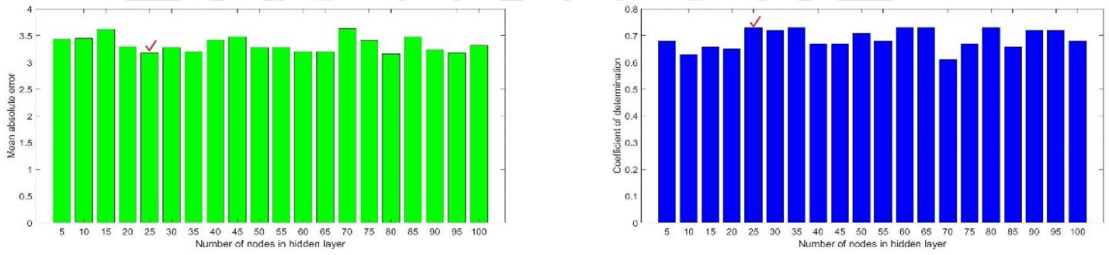
ดังนั้นสามารถสรุปรูปแบบของแบบจำลองที่ดีที่สุดสำหรับการทำนายสารส้มจึงมาจากเลเยอร์ที่ซ่อนสองชั้น (30-85) โดย R^2 คือ 0.66 MAE คือ 2.94 มิลลิกรัม/ลิตร ในทางกลับกันรูปแบบที่ดีที่สุดสำหรับ PACL คือรูปแบบ 5 โหนดโดยมีชั้นซ่อนหนึ่งชั้นมีค่า R^2 เท่ากับ 0.59 และ R^2 เท่ากับ 3.21 มิลลิกรัม/ลิตร

จากการประเมินประสิทธิภาพพบว่า ANN ทั้งสองแบบจำลองสามารถจำลองการทดสอบ Jar test แม้ว่าแบบจำลองสารส้มจะทำงานได้ดีกว่าแบบ PACL ก็ตามแบบจำลองเหล่านี้ยังคงต้องได้รับการปรับปรุงเนื่องจาก R^2 ที่ค่อนข้างน้อยและ MAE ที่เยอะ จะเห็นได้ว่า MAE ของสารส้มมีค่าประมาณ 12 เปอร์เซ็นต์เมื่อเทียบกับค่าเฉลี่ยจริง 26.45 mg / L (แสดงในตาราง 7) และสำหรับ PACL ประมาณ 28 เปอร์เซ็นต์เมื่อเทียบกับค่าเฉลี่ยจริง 11.45 mg / L (แสดงในตาราง 7)

หากตั้งค่าข้อผิดพลาด 10 เปอร์เซ็นต์เป็นเกณฑ์จะพบว่าแบบจำลองสารส้มเกินมาเล็กน้อยและแบบจำลอง PACL เกินเกณฑ์ที่กำหนด

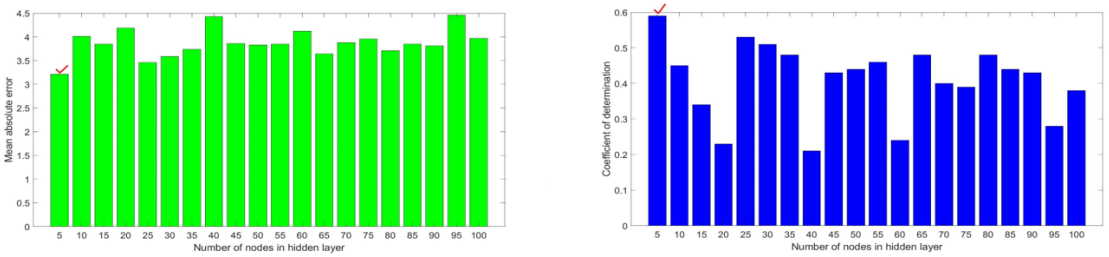
เนื่องจากรับข้อมูลไปยังแบบจำลองถูกแบ่งแบบสุ่ม (Random) ออกเป็น 3 กลุ่ม ได้แก่ ชุดฝึกชุดทดสอบและชุดตรวจสอบความถูกต้องโดยมีอัตราส่วน 1/3, 1/6 และ 1/6 ตามลำดับ แม้ว่ากลุ่มข้อมูลฝึกสอนจะเป็นกลุ่มที่ใหญ่ที่สุด แต่ก็มีการเลือกแบบสุ่มไม่มีการรับประกันว่ากลุ่มข้อมูลทดสอบและตรวจสอบทั้งหมดเป็นชุดย่อยของกลุ่มข้อมูลฝึกสอน (กล่าวคือกลุ่มข้อมูลฝึกสอนหลากหลายครอบคลุมกลุ่มข้อมูลทดสอบและตรวจสอบ) โดยทั่วไปโมเดล ANN จะให้ประสิทธิภาพในการแก้ไขที่ดีในช่วง (Interpolation range) ซึ่งเป็นที่นิยมมากกว่าช่วงการประมาณค่านอกช่วง (Extrapolation range) [1] ผลลัพธ์เหล่านี้ R^2 ขนาดใหญ่และ MAE ขนาดเล็กของการทำนายทั้งสองแบบจำลอง ANN และที่แย่กว่านั้นคืออินพุตช่วงแคบ ๆ ของแบบจำลอง PACL

ดังนั้นการใช้แบบจำลอง ANN เพื่อแทนที่การทดสอบ Jar จึงจำเป็นต้องใช้ชุดข้อมูลการฝึกอบรมเพื่อให้ครอบคลุมผลลัพธ์ที่เป็นไปได้เพื่อที่จะสอนแบบจำลอง เทคนิคกระบวนการนี้สามารถปรับปรุงและประยุกต์ใช้เพื่อทดแทนการทดสอบ Jar test แบบออฟไลน์ในห้องปฏิบัติการ



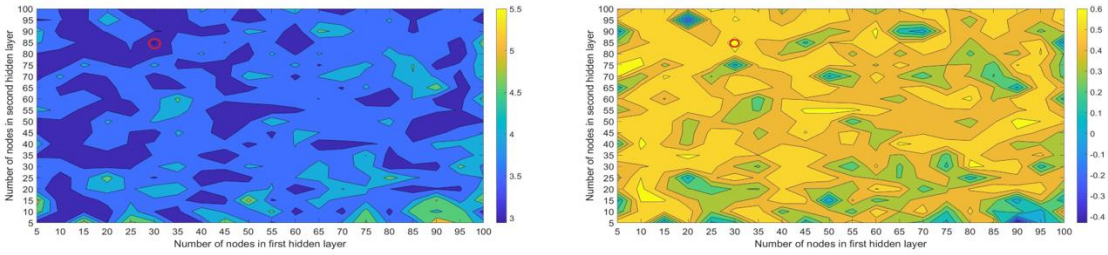
(a) MAE (b) R^2

ภาพประกอบ 27 แบบจำลองการทำนายสารส้มชั้นซ่อนหนึ่งชั้น



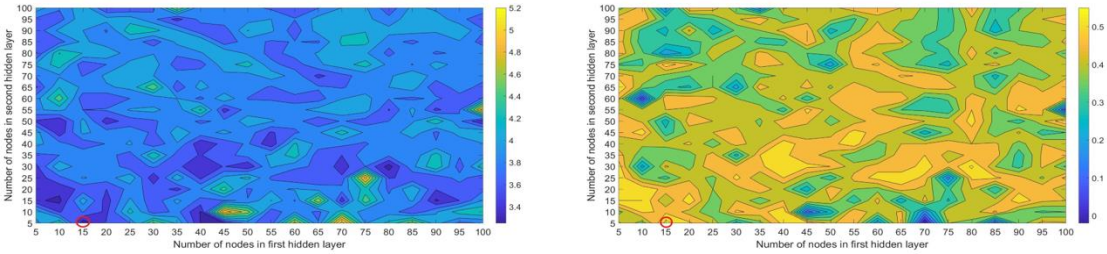
(a) MAE (b) R^2

ภาพประกอบ 28 แบบจำลองการทำนายโพลิอลูมิเนียมคลอไรด์ชั้นซ่อนหนึ่งชั้น



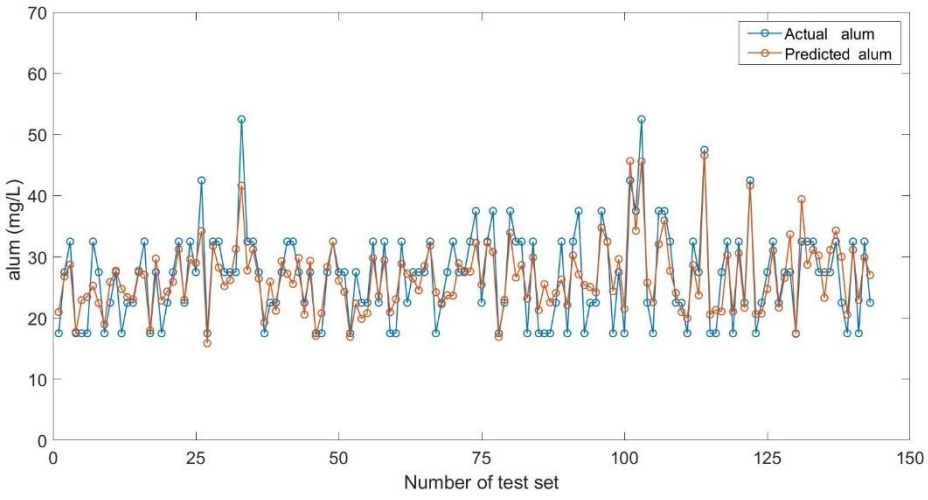
(a) *MAE* (b) R^2

ภาพประกอบ 29 แบบจำลองการทำนายสารส้มชั้นซ่อนสองชั้น

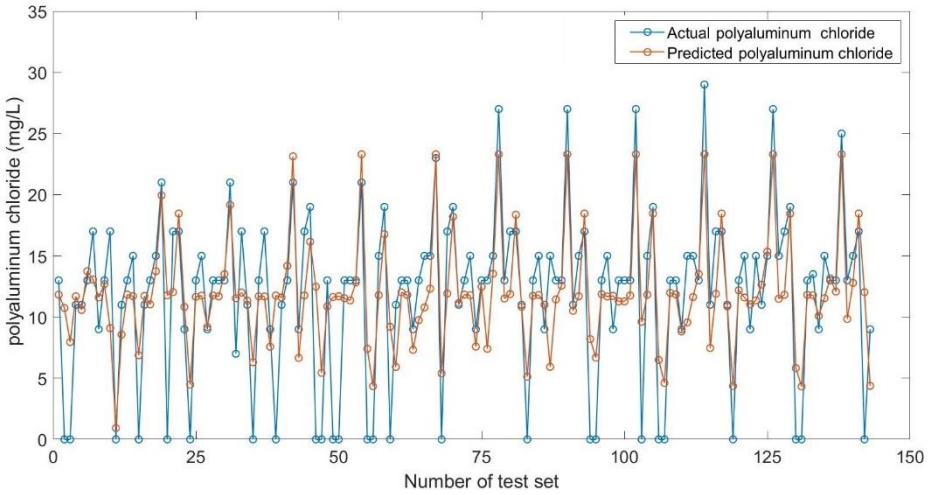


(a) *MAE* (b) R^2

ภาพประกอบ 30 แบบจำลองการทำนายโพลิอลูมิเนียมคลอไรด์ชั้นซ่อนสองชั้น



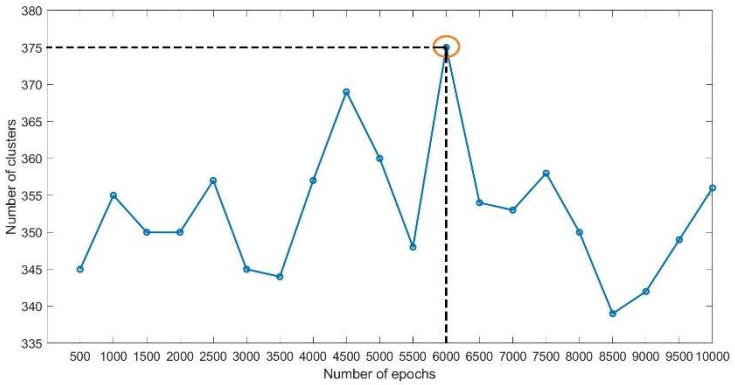
ภาพประกอบ 31 การทำนายสารส้มในชุดทดสอบ



ภาพประกอบ 32 การทำนายโพลิอลูมิเนียมคลอไรด์ในชุดทดสอบ

4.3 ผลการแบ่งข้อมูลด้วยวิธี SOM

ผู้วิจัยเห็นข้อจำกัดจากการแบ่งข้อมูลแบบสุ่ม จึงนำวิธีการแบ่งข้อมูลแบบ SOM มาประยุกต์ใช้ เริ่มต้นจากการนำ ข้อมูลในงานวิจัยนี้รวมทั้งหมดมี 717 รายการ และ 5 พารามิเตอร์อินพุต โดยกำหนดโครงสร้างของ SOM มีขนาด 25 x 25 ชั้น Kohonen มีลักษณะแบบโครงสร้างหกเหลี่ยม สอนแบบจำลองจนได้จำนวน Clusters สูงสุดโดยจำนวน Epoch เพิ่มขึ้นทีละ 500 จนถึง 10,000 Epochs เพื่อแสดงให้เห็นว่าจำนวน Cluster มีแนวโน้มไม่เพิ่มขึ้นอีกต่อไป ภาพประกอบ 33 นอกจากนี้ดูจากกราฟมีลักษณะเป็น Monotonic trend ดังนั้นจึงสันนิษฐานได้ว่า 375 Cluster เป็นจำนวนสูงสุดของข้อมูลและผลของการรัน SOM algorithm



ภาพประกอบ 33 แสดงจำนวน Cluster ของแต่ละ Epochs

ข้อมูล 717 รายการถูกจัดเป็นกลุ่มได้ 375 กลุ่มมี 291 กลุ่มประกอบด้วยข้อมูลน้อยกว่า 3 รายการกำหนดให้เป็น Training set มี 84 กลุ่มมีข้อมูลอย่างน้อย 3 รายการโดยกลุ่มเหล่านี้เราสุ่มสำหรับเลือกเป็น Test set และ Validation set อย่างละ 1 รายการส่วนที่เหลือถูกกำหนดให้เป็น Training set เมื่อทำการแบ่งข้อมูลลักษณะนี้จึงมี Training set 549 รายการและ Test set กับ Validation set มีอย่างละ 84 รายการ

ผลลัพธ์ค่าสถิติ ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation), ค่าเฉลี่ย (Mean), ค่าสูงสุด (Maximum), ค่าต่ำสุด (Minimum) และ ช่วงข้อมูล (Range of value) มีความใกล้เคียงกันดังตาราง 9 และการทดสอบสมมุติฐานหลัก Test set กับ Training set และ Validation set กับ Training set ของทั้งหมด 5 พารามิเตอร์อินพุตระดับนัยสำคัญ = 0.05 พบว่า T-test null hypothesis ถูกปฏิเสธมากที่สุด 2 อินพุตพารามิเตอร์คือ Turbidity และ pH ของกลุ่มข้อมูลทดสอบกับกลุ่มข้อมูลฝึกสอนและกลุ่มข้อมูลตรวจสอบกับกลุ่มข้อมูลฝึกสอนคือ Turbidity และ Oxygen consumption สำหรับ F-test null hypothesis ไม่มีพารามิเตอร์อินพุตปฏิเสธสมมุติฐานการทดสอบ Range test โดยผลปรากฏว่าทุกตัวแปรของ Test set และ Validation set อยู่ในภายช่วง Training set การทดสอบสมมติฐานได้สรุปในภาพประกอบ 34, 35 และ 36

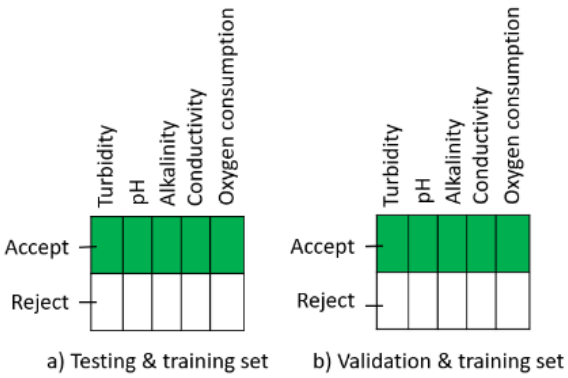
ตาราง 9 ค่าสถิติของกลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบ และกลุ่มข้อมูลตรวจสอบ แบ่งข้อมูลโดยใช้วิธีการ SOM

Parameter and data set	Standard deviation	Mean	Maximum	Minimum	Range of value
Input : Turbidity					
Training set	15.14	27.65	126	15	111
Test set	13.56	26.5	107	16	91
Validation set	14.34	26.78	108	15	93
Input : pH					
Training set	0.16	7.69	8.2	7.25	0.95
Test set	0.19	7.73	8.18	7.41	0.77
Validation set	0.17	7.7	8.12	7.31	0.81
Input : Alkalinity					
Training set	7.07	97.37	115	69	46
Test set	7.44	96.34	113	70	43
Validation set	7.7	96.44	113	70	43

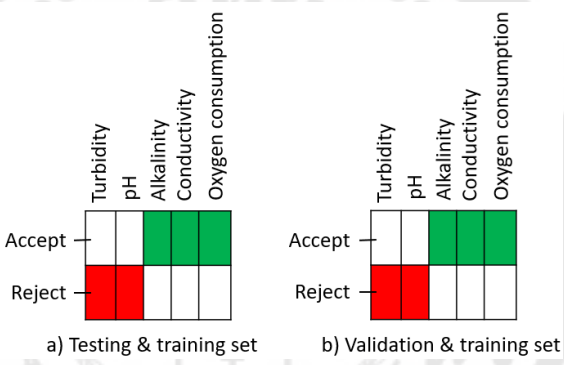
ตาราง 9 (ต่อ)

Parameter and data set	Standard deviation	Mean	Maximum	Minimum	Range of value
Input : Conductivity					
Training set	239.26	587.56	3500	201	3299
Test set	253.46	353.25	2464	213	2251
Validation set	205.99	347.17	2008	209	1799
Input : Oxygen consumption					
Training set	0.73	3.68	5.42	1.62	3.8
Test set	0.79	3.59	5.41	1.91	3.5
Validation set	0.71	3.68	5.32	2.24	3.08

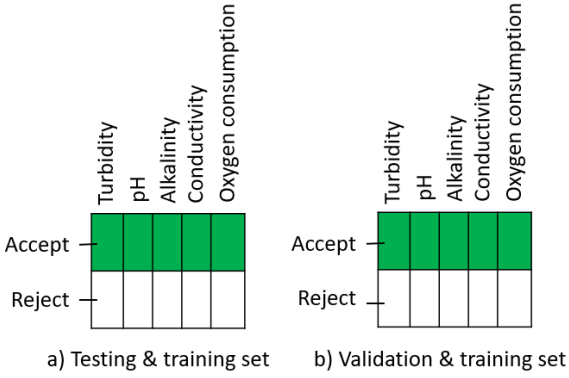
เพื่อเปรียบเทียบประสิทธิภาพกับการแบ่งข้อมูลแบบ SOM จึงได้แบ่งข้อมูลโดยใช้อัตราส่วนของ SOM มาทำการแบ่งแบบ Random ข้อมูลทั้งหมด 717 รายการ (100 %) แบ่งเป็นกลุ่มข้อมูลฝึกสอน 431 รายการ (76.57 %) กลุ่มข้อมูลทดสอบ 143 รายการ (11.715 %) และกลุ่มตรวจสอบ 143 รายการ (11.715 %) ตาราง 10 ค่าสถิติของพารามิเตอร์น้ำดิบจะเห็นได้ว่าค่าทางสถิติของอินพุตบางตัวมีความแตกต่างกันมากเมื่อดูการทดสอบสมมุติฐานระดับนัยสำคัญ = 0.05 Range test โดยกลุ่มข้อมูลทดสอบกับกลุ่มข้อมูลฝึกสอนไม่ครอบคลุม 1 พารามิเตอร์อินพุตคือ pH และกลุ่มข้อมูลตวจสอบกับกลุ่มข้อมูลฝึกสอนไม่ครอบคลุม 2 พารามิเตอร์อินพุตคือ Turbidity กับ pH ส่วน T-test ซึ่งกลุ่มข้อมูลทดสอบกับกลุ่มข้อมูลฝึกสอนมี Oxygen consumption ที่ยอมรับสมมุติฐานหลักที่และกลุ่มข้อมูลตรวจสอบกับกลุ่มข้อมูลฝึกสอนปฏิเสฎสมมุติฐานหลักทุกพารามิเตอร์อินพุตและท้ายสุดการทดสอบ F-test ซึ่งกลุ่มข้อมูลตรวจสอบกับกลุ่มข้อมูลฝึกสอนมี Conductivity เท่านั้นที่ปฏิเสธสมมติฐานหลักสำหรับกลุ่มข้อมูลทดสอบกับกลุ่มข้อมูลฝึกสอนทุกพารามิเตอร์อินพุตยอมรับสมมุติฐานหลักสามารถสรุปได้ในภาพประกอบ 37, 38 และ 39



ภาพประกอบ 34 Range test null hypothesis แบ่งกลุ่มข้อมูลแบ่งวิธี SOM

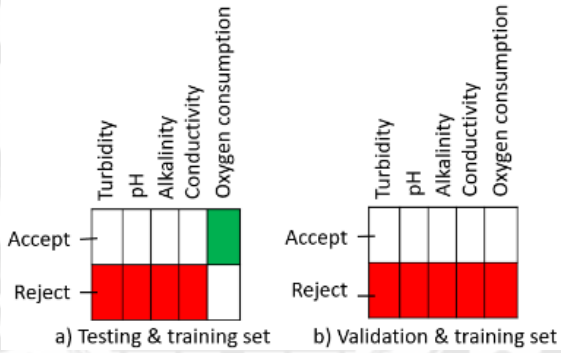


ภาพประกอบ 35 T-test null hypothesis แบ่งกลุ่มข้อมูลแบ่งวิธี SOM

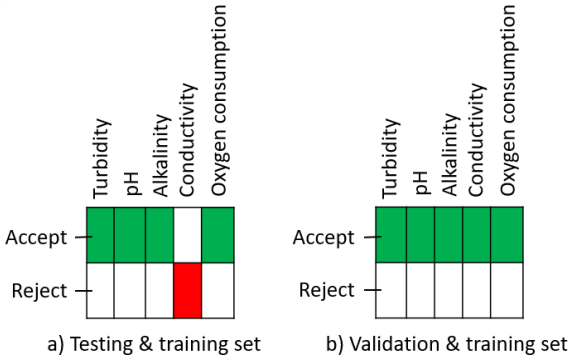


ภาพประกอบ 36 F-test null hypothesis แบ่งกลุ่มข้อมูลแบ่งวิธี SOM

ภาพประกอบ 37 Range test null hypothesis แบ่งกลุ่มข้อมูลด้วยวิธี Random



ภาพประกอบ 38 T-test null hypothesis แบ่งกลุ่มข้อมูลแบ่งวิธี Random



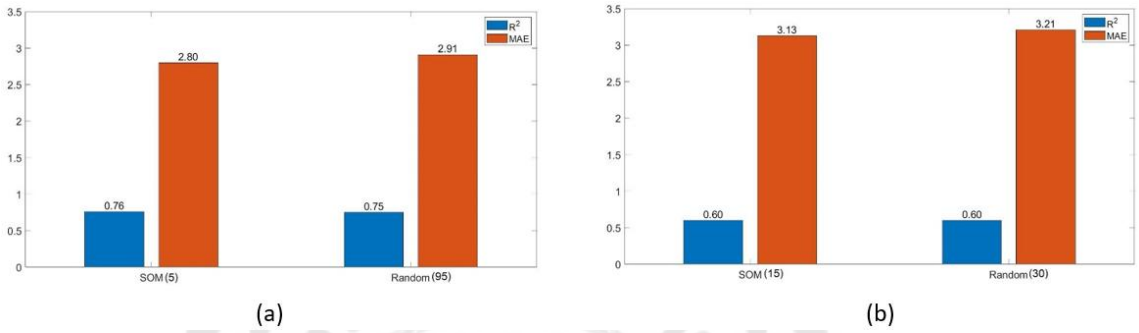
ภาพประกอบ 39 F-test null hypothesis แบ่งกลุ่มข้อมูลแบ่งวิธี Random

ดำเนินการทดลองที่จำนวนชั้นซ้อน 1 ชั้น โดยเริ่มต้น 5 โหนด และทำการเพิ่มจำนวนโหนดทีละ 5 โหนด จนถึง 100 โหนด (5,10,15, … , 100) การแบ่งกลุ่มข้อมูลฝึกสอน (training set) กลุ่มข้อมูลทดสอบ (test set) และกลุ่มข้อมูลตรวจสอบ (validation set) ด้วยเทคนิคการแบ่งกลุ่ม SOM แบบจำลองทำนายปริมาณ Alum ที่มีประสิทธิภาพการทำนายดีที่สุดที่แบบจำลอง (5) R^2 และ MAE เท่ากับ 0.76, 2.80 มิลลิกรัม/ลิตร ตามลำดับ

ตาราง 10 ค่าสถิติของกลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบ และกลุ่มข้อมูลตรวจสอบ แบ่งข้อมูลโดยใช้วิธีการ Random

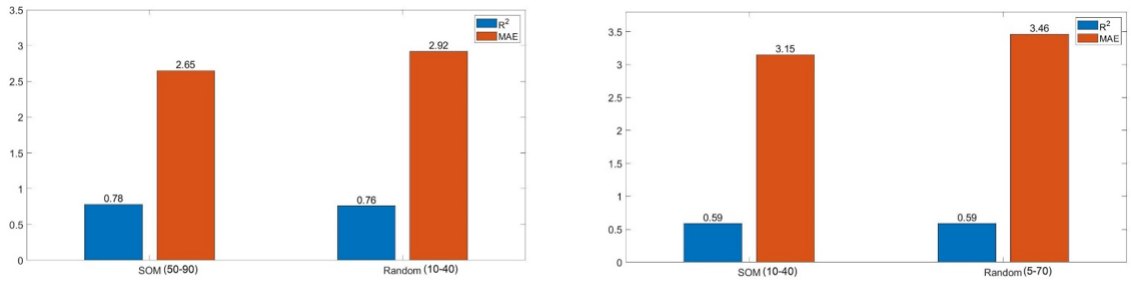
	Standard deviation	Mean	Maximum	Minimum	Range of value
Input : Turbidity					
Training set	13.88	27.03	122	15	107
Test set	12.58	26.8	108	15	93
Validation set	21.53	30.51	126	15	111
Input : pH					
Training set	0.16	7.7	8.18	7.25	0.93
Test set	0.17	7.68	8.2	7.31	0.89
Validation set	0.16	7.68	8.18	7.35	0.83
Input : Alkalinity					
Training set	6.62	97.4	115	69	46
Test set	7.97	96.53	113	70	43
Validation set	9.54	96.1	113	70	43
Input : Conductivity					
Training set	257.73	367.77	3500	201	3299
Test set	72.31	325.99	799	209	590
Validation set	198.84	353.09	1920	208	1712
Input : Oxygen consumption					
Training set	0.74	3.66	5.42	1.62	3.8
Test set	0.74	3.74	5.31	2.22	3.09
Validation set	0.71	3.7	5.1	2.24	2.86

การแบ่งข้อมูลแบบ Random ด้วยอัตราส่วน Training set: Test set: Validation set เป็น 0.77: 0.115: 0.115 จากเทคนิค SOM ที่แบบจำลอง (95) ได้ R^2 เท่ากับ 0.75 และ MAE เท่ากับ 2.91 มิลลิกรัม/ลิตร ในส่วนแบบจำลองทำนายปริมาณ PACL เทคนิค SOM (15) ได้ค่า R^2 เท่ากับ 0.6, MAE เท่ากับ 3.13 มิลลิกรัม/ลิตร และด้วยวิธี Random (30) ได้ค่า R^2 เท่ากับ 0.6, MAE เท่ากับ 3.21 มิลลิกรัม/ลิตร ตามลำดับดังภาพที่ 40



ภาพประกอบ 40 ประสิทธิภาพการทำนายปริมาณของ Alum (a) และ PACL(b) 1 ชั้นซ้อน

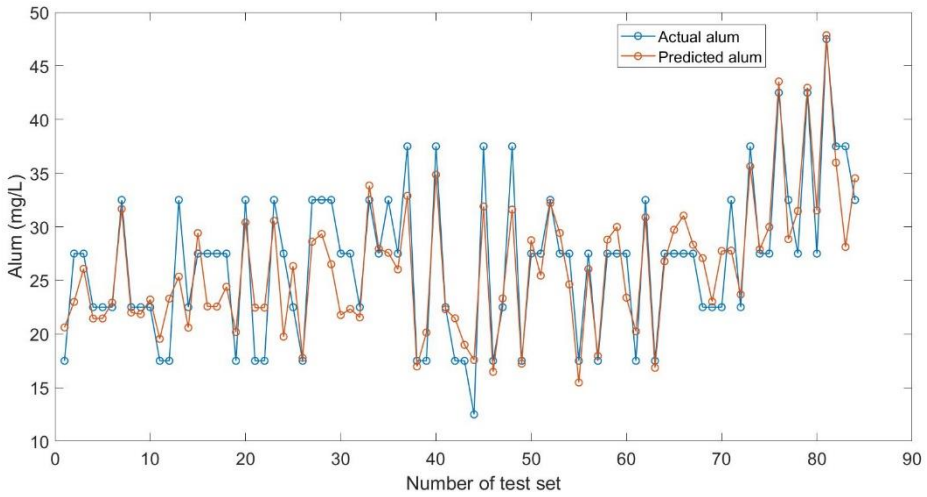
นอกจากนี้ผู้วิจัยได้ทำการเพิ่มจำนวนชั้นซ้อนเป็น 2 ชั้นซ้อน และเพิ่มโหนดด้วยวิธีการเดียวกับแบบ 1 ชั้นซ้อน การแบ่งกลุ่มข้อมูลด้วย SOM และ การแบ่งแบบ Random ด้วยอัตราส่วนเดียวกันกับ SOM แบบจำลองทำนายปริมาณ Alum ที่มีประสิทธิภาพการทำนายดีที่สุดที่ด้วย SOM แบบจำลอง 50-90 ได้ค่า R^2 เท่ากับ 0.78, MAE เท่ากับ 2.65 มิลลิกรัม/ลิตร และด้วย Random แบบจำลอง 10-40 ได้ค่า R^2 เท่ากับ 0.76, MAE เท่ากับ 2.92 มิลลิกรัม/ลิตร สำหรับแบบจำลองทำนายปริมาณ PACL จากเทคนิค SOM ได้ค่า R^2 เท่ากับ 0.59 และ MAE เท่ากับ 3.15 มิลลิกรัม/ลิตร ที่แบบจำลอง 10-40 และด้วย Random กับ 5-70 ได้ค่า R^2 เท่ากับ 0.59, MAE เท่ากับ 3.46 มิลลิกรัม/ลิตร ดังรูปภาพ 41 โดยกราฟให้เห็นถึงปริมาณสารส้มและโพลิอลูมิเนียมคลอไรด์ที่คาดการณ์และค่าจริงในชุดทดสอบถูกแสดงในภาพประกอบ 42, 43, 44 และ 45



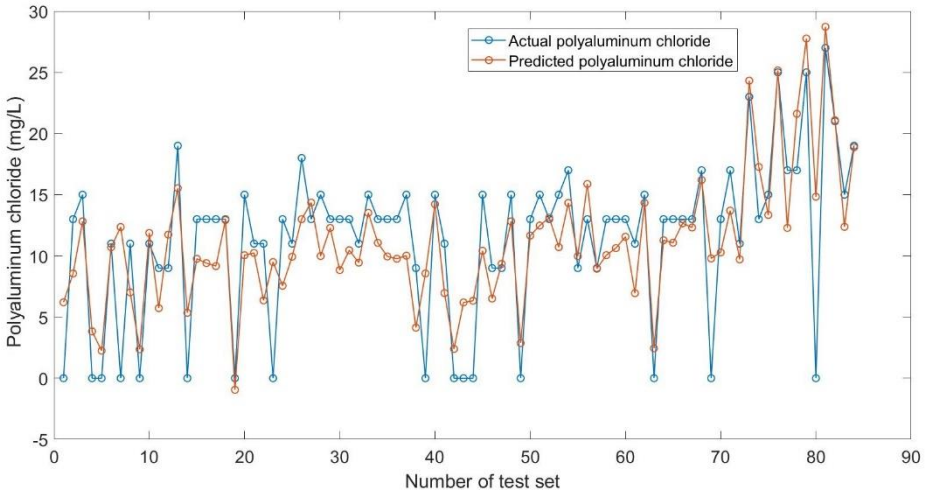
(a) (b)

ภาพประกอบ 41 ประสิทธิภาพการทำนายปริมาณของ Alum (a) และ PACL (b) 2 ชั้นซ้อน

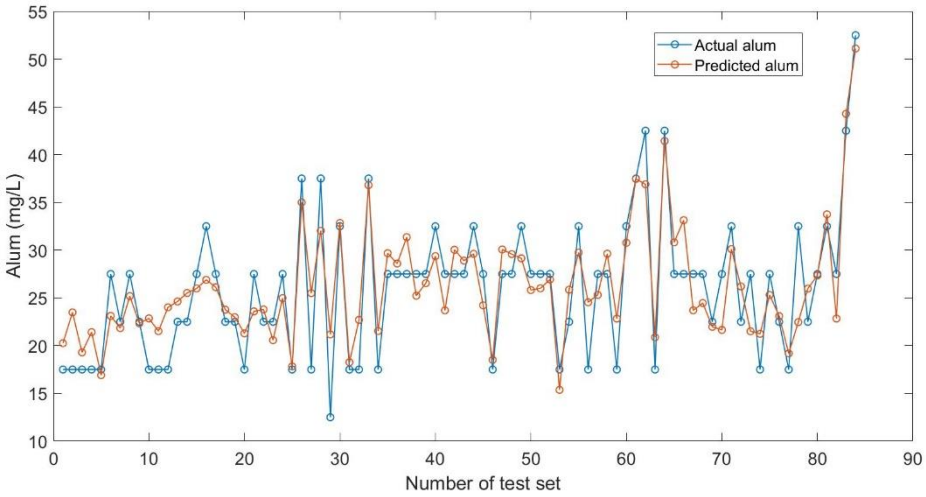
แผนภูมิแท่งในภาพประกอบ 41 แสดงการเปรียบเทียบแบบจำลองที่แบ่งกลุ่มโดยใช้เทคนิค SOM จะให้ R^2 ที่สูงกว่าหรือเท่ากับการแบ่งข้อมูลแบบ Random เป็นเพราะว่าเมื่อเราได้การจัดกลุ่มจาก SOM ผู้วิจัยทำการหยิบสุ่มข้อมูลแบ่งเป็นกลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบและกลุ่มข้อมูลตรวจสอบดังที่กล่าวก่อนหน้านี้ซึ่งการสุ่มหยิบ Random แบบนี้ก็อาจจะทำให้ผลลัพธ์ที่ได้อาจจะออกมาดีหรือไม่ดีก็เป็นไปได้และเมื่อพิจารณาแบบจำลองที่มี R^2 เท่ากันแต่แบบจำลองที่แบ่งกลุ่มเทคนิค SOM ให้ค่า MAE ที่น้อยกว่าเนื่องจากว่าผลการทดสอบค่าทางสถิติ Range test, F-test และ T-test ระหว่างกลุ่มข้อมูลทดสอบกับกลุ่มข้อมูลฝึกสอนและกลุ่มข้อมูลตรวจสอบกับกลุ่มข้อมูลฝึกสอนยอมรับสมมุติฐานหลักมากกว่าการแบ่งกลุ่มแบบ Random.



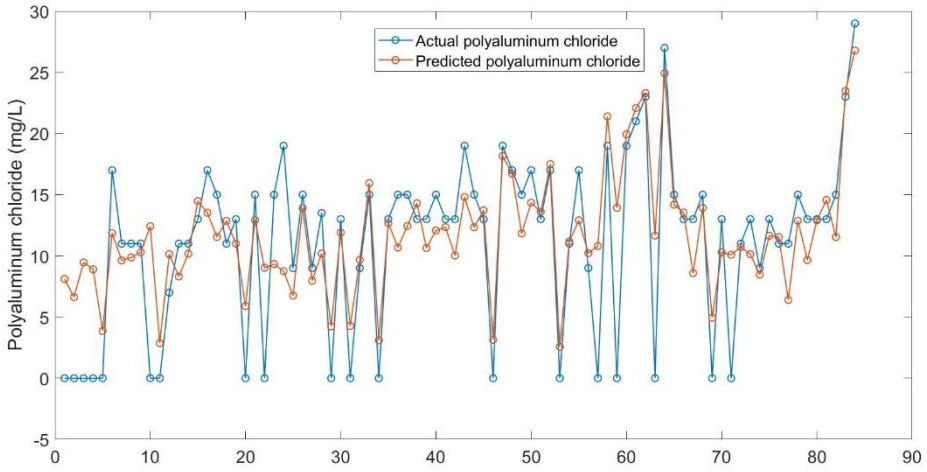
ภาพประกอบ 42 การทำนายสารส้มในชุดทดสอบวิธีการแบ่งข้อมูล SOM



ภาพประกอบ 43 การทำนายโพลิอลูมิเนียมคลอไรด์ในชุดทดสอบวิธีการแบ่งข้อมูล SOM



ภาพประกอบ 44 การทำนายสารส้มในชุดทดสอบวิธีการแบ่งข้อมูล Random



ภาพประกอบ 45 การทำนายโพลิอลูมิเนียมคลอไรด์ในชุดทดสอบวิธีการแบ่งข้อมูล Random

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้สร้างแบบจำลองซอฟต์จาร์เทสโดยการประยุกต์ใช้โครงข่ายประสาทเทียมเพื่อทำนายปริมาณสารส้มและโพลิอลูมิเนียมคลอไรด์โดยนำข้อมูลการทดสอบจาร์เทสจากโรงงานผลิตน้ำบางเขน การประปานครหลวงตั้งแต่วันที่ 1 มกราคม พ.ศ 2562 ถึง 31 ธันวาคม พ.ศ 2562 มีพารามิเตอร์อินพุตน้ำดิบได้แก่ ความขุ่น สภาพความเป็นด่าง ความเป็นกรด-ด่าง ความนำไฟฟ้า การบริโภคออกซิเจน และพารามิเตอร์เอาต์พุตได้แก่สารส้มและโพลิอลูมิเนียมคลอไรด์ และทำการแบ่งข้อมูลแบบสุ่มออกเป็น 3 กลุ่มได้แก่ กลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบ และกลุ่มข้อมูลตรวจสอบด้วยอัตราส่วน 1/3, 1/6, 1/6 ตามลำดับ โดยแบบจำลองที่ดีที่สุดได้มาจากการเพิ่มจำนวนชั้นซ้อนไม่เกิน 2 ชั้นซ้อนและจำนวนเลเยอร์ในชั้นซ้อนเริ่มต้นที่ 5 โหนดโดยเพิ่มขึ้นทีละ 5 จนถึง 100 โหนดของโครงข่ายประสาทเทียม สำหรับแบบจำลองทำนายสารส้มดีที่สุดที่แบบจำลอง 25 โหนด 1 ชั้นซ้อนได้ R^2 เท่ากับ 0.73, MAE เท่ากับ 3.18 มิลลิกรัม/ลิตร และโพลิอลูมิเนียมคลอไรด์ที่แบบจำลอง 5 โหนด 1 ชั้นซ้อนได้ R^2 เท่ากับ 0.59, MAE เท่ากับ 3.21 มิลลิกรัม/ลิตร

การแบ่งข้อมูลแบบสุ่มไม่สามารถยืนยันว่าแบบจำลองถูกสอนด้วยกลุ่มข้อมูลฝึกสอนที่ครอบคลุมกลุ่มข้อมูลทดสอบรวมถึงโครงข่ายประสาทเทียมนั้นมีข้อจำกัดคือการทำนายข้อมูลนอกช่วงกลุ่มข้อมูลฝึกสอนทำให้ประสิทธิภาพของผลการทำนายที่ได้ไม่ดี ผู้วิจัยจึงดำเนินการเพิ่มประสิทธิภาพแบบจำลองโดยใช้เทคนิคการแบ่งกลุ่มข้อมูลแบบ SOM เพื่อให้กลุ่มข้อมูลฝึกสอนกลุ่มข้อมูลทดสอบและกลุ่มข้อมูลตรวจสอบเป็นตัวแทนซึ่งกันและกันโดยตรวจสอบได้จากการทดสอบสมมุติฐานสถิติได้แก่ F-test, T-test และ Range test และเปรียบเทียบการแบ่งข้อมูลแบบสุ่มโดยใช้อัตราส่วนที่ได้จากการแบ่งกลุ่มข้อมูลแบ่งแบบ SOM

แบบจำลองที่ดีที่สุดสำหรับการทำนายปริมาณสารส้มด้วยวิธีการแบ่งกลุ่ม SOM ที่แบบจำลอง 50-90 โหนด 2 ชั้นซ้อนได้ R^2 เท่ากับ 0.78, MAE เท่ากับ 2.8 มิลลิกรัม/ลิตร และการแบ่งกลุ่มข้อมูลแบบสุ่มที่แบบจำลอง 10-40 โหนด 2 ชั้นซ้อนได้ R^2 เท่ากับ 0.76, MAE เท่ากับ 2.92 มิลลิกรัม/ลิตร และแบบจำลองที่ดีที่สุดสำหรับการทำนายปริมาณโพลิอลูมิเนียมคลอไรด์ด้วยวิธีการแบ่งกลุ่ม SOM ที่แบบจำลอง 15 โหนด 1 ชั้นซ้อนได้ R^2 เท่ากับ 0.6, MAE เท่ากับ 3.13 มิลลิกรัม/ลิตร ตามลำดับ การแบ่งกลุ่มข้อมูลแบบสุ่มที่แบบจำลอง 30 โหนด 1 ชั้นซ้อนได้ R^2 เท่ากับ 0.60, MAE เท่ากับ 3.21 มิลลิกรัม/ลิตร

เทคนิคการแบ่งข้อมูล SOM มีข้อได้เปรียบกว่าวิธีการแบบสุ่มตรงที่กลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบและกลุ่มข้อมูลตรวจสอบมีความคล้ายคลึงกันสามารถแก้ปัญหาข้อจำกัดของโครงข่ายประสาทเทียมได้ นั่นคือการประมาณค่านอกช่วงข้อมูลการฝึกสอน (Extrapolation) และป้องการแบบจำลอง Overfitting พบว่าแบบจำลองที่ประยุกต์ใช้โครงข่ายประสาทเทียมโดยใช้เทคนิคการแบ่งข้อมูลแบบ SOM มีประสิทธิภาพดีกว่าแบบสุ่ม

5.2 อภิปรายผล

จากตาราง 11 สรุปผลประสิทธิภาพแบบจำลองเห็นได้ชัดว่าการแบ่งข้อมูลแบบสุ่มออกเป็นกลุ่มข้อมูลฝึกสอน กลุ่มข้อมูลทดสอบและกลุ่มข้อมูลตรวจสอบ ด้วยอัตราส่วน 1/3, 1/6, 1/6 ตามลำดับ แบบจำลองที่ดีที่สุดของสารส้มมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 3.18 มิลลิกรัม/ลิตร และโพลิอลูมิเนียมคลอไรด์เท่ากับ 3.21 มิลลิกรัม/ลิตร เมื่อประยุกต์ใช้เทคนิคการแบ่งกลุ่มข้อมูลแบบ SOM ทำให้ประสิทธิภาพของแบบจำลองได้ดีขึ้นโดยแบบจำลองทำนายสารส้มมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของสารส้มเท่ากับ 2.8 มิลลิกรัม/ลิตร และโพลิอลูมิเนียมคลอไรด์เท่ากับ 3.13 มิลลิกรัม/ลิตร และการแบ่งข้อมูลแบบสุ่มด้วยอัตราส่วนเดียวกับ SOM ได้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของแบบจำลองทำนายสารส้มเท่ากับ 2.92 มิลลิกรัม/ลิตรและโพลิอลูมิเนียมคลอไรด์เท่ากับ 3.21 มิลลิกรัม/ลิตร เมื่อเปรียบเทียบวิธีการแบ่งข้อมูลทั้ง 3 แบบพบว่าการแบ่งกลุ่มข้อมูลแบบ SOM ให้ค่าความคลาดเคลื่อนสัมบูรณ์น้อยที่สุด ด้วยการประยุกต์ใช้โครงข่ายประสาทเทียมสำหรับการสร้างแบบจำลองสามารถรับยอมรับได้เนื่องจากมีค่าความคลาดเคลื่อนน้อยกว่าค่าความคลาดเคลื่อนของเครื่องมือวัด 15 เปอร์เซ็นต์

ตาราง 11 สรุปผลประสิทธิภาพของแบบจำลอง

วิธีการแบ่งข้อมูล	แบบจำลองที่ดีที่สุดทำนาย			
	สารส้ม		โพลิอลูมิเนียมคลอไรด์	
	R^2	MAE	R^2	MAE
การแบ่งข้อมูลแบบสุ่ม	0.73	3.18	0.59	3.21
การแบ่งข้อมูลแบบ SOM	0.78	2.80	0.60	3.13
การแบ่งข้อมูลแบบสุ่มด้วยอัตราส่วนเดียวกับ SOM	0.76	2.92	0.60	3.21

5.3 ข้อเสนอแนะ

1. การปรับปรุงประสิทธิภาพของแบบจำลองด้วยวิธีอื่น เช่น ใช้เทคนิคการเรียนรู้ของเครื่องประยุกต์สร้างแบบจำลองอาจทำให้ประสิทธิภาพดียิ่งขึ้น การเรียนรู้ของเครื่องแบ่งออกเป็น 2 ประเภทได้แก่ แบบมีผู้ช่วยสอนและแบบไม่มีผู้ช่วยสอน สำหรับงานวิจัยนี้เป็นแบบมีผู้ช่วยสอน คือแบบจำลองรู้คำตอบผลลัพธ์การทำนายซึ่งอัลกอริทึมที่เลือกใช้ควรเป็นแบบการถดถอย (Regression) ยกตัวอย่างเช่น RandomForestRegressor, LinearRegression, Support Vector Regression, GradientBoostingRegressor, PolynomialFeatures และ XGBRegressor เป็นต้น

2. การหาฟีเจอร์เพิ่มเติมเนื่องจากการทดสอบจาร์เทสเป็นสภาพแวดล้อมเป็นแบบปิด กล่าวคืออยู่ในห้องปฏิบัติการแต่การเติมสารเคมี ได้แก่ สารส้มและโพลิอลูมิเนียมคลอไรด์ที่ได้จากการทดสอบจาร์เทสถูกเติมในถังตกตะกอนซึ่งเป็นสภาพแวดล้อมแบบเปิดซึ่งอาจจะมีพารามิเตอร์ที่สำคัญที่ทำให้แบบจำลองมีประสิทธิภาพมากขึ้นเช่น ความขุ่นของน้ำดิบในถังตกตะกอน อุณหภูมิของน้ำดิบ เวลา อัตราการไหลเข้าของน้ำดิบในถังตกตะกอน อัตราความเร็วของใบกวน เป็นต้น

3. การสร้างเป็นแอฟพลิเคชันหลังจากที่ได้แบบจำลองสามารถนำไปปรับใช้จริงโดยพัฒนาด้วยโปรแกรมต่าง ๆ เช่น SimWT หรือ SimTTP โดยทั้ง 2 โปรแกรมนี้เป็นเครื่องมือที่อยู่ในโปรแกรม Matlab สำหรับนำแบบจำลองที่ได้มาสร้างการติดต่อกับผู้ใช้งาน (Graphical user interface) โดยผู้ใช้งานสามารถปรับพารามิเตอร์อินพุตต่าง ๆ ของแบบจำลองได้แล้วผลลัพธ์พารามิเตอร์เอาต์พุตของแบบจำลองก็จะแสดงหน้าโปรแกรมซึ่งสะดวกต่อการใช้งานและง่ายต่อการทำความเข้าใจ

4. การสร้างแบบจำลองทำนายสารส้มและโพลิอลูมิเนียมคลอไรด์ด้วย 1 แบบจำลองอาจจะทำให้แบบจำลองมีความซับซ้อนมากเกินไปเนื่องจากในแต่ละช่วงเวลามีหลากหลายปัจจัยที่มีผลทำให้คุณภาพน้ำดิบเปลี่ยนไปซึ่งอาจจะลองวิธีการทำหลายแบบจำลองด้วยการแบ่งข้อมูลออกเป็นหลาย ๆ ช่วง เช่น ใช้ค่าพารามิเตอร์ค่าสัมประสิทธิ์สหสัมพันธ์สูงที่สุดในการแบ่งข้อมูลหรือใช้พารามิเตอร์ที่มีความไวสูง (Hight sensitivity) ของแบบจำลอง เป็นต้น ซึ่งอาจทำให้แบบจำลองมีประสิทธิภาพมากขึ้น

5. ข้อมูลที่ใช้เป็นช่วงในระยะเวลา 1 ปี ดังนั้นถ้าจะนำแบบจำลองไปประยุกต์ใช้จริงอาจต้องมีการสอนแบบจำลองเพิ่มเติมโดยนำข้อมูลในปีอื่น ๆ หรือในช่วงเหตุการณ์สำคัญต่าง ๆ เพื่อให้ครอบคลุมเหตุการณ์ที่อาจจะเกิดขึ้นได้เพราะว่าในบางปีก็มีเหตุการณ์ทางธรรมชาติที่แตกต่างกัน

บรรณานุกรม

[1] คณะทำงานย่อยที่ 3 กระบวนการผลิตน้ำประปาและควบคุมคุณภาพน้ำ. "คู่มือกระบวนการทำงาน : กระบวนการผลิตน้ำประปาและควบคุมคุณภาพน้ำ." สืบค้นจาก https://reg4.pwa.co.th/km/sites/default/files/km/upload/news1-281062.pdf (accessed 2561).

[2] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, 3rd ed. New York: Addison Wesley, 2011.

[3] L. Fu, *Neural networks in computer intelligence*. Singapore: McGraw-Hill, 1994.

[4] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Englewood Cliffs: Prentice Hall, 2008.

[5] J. Heaton, *Introduction to Neural Networks for Java*, 2nd ed. St. Louis: Heaton Research, Inc., 2008.

[6] H. Maier, N. Morgan, and C. Chow, "Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters," *Environmental Modelling and Software,* vol. 19, pp. 485-494, 05/01 2004, doi: 10.1016/S1364-8152(03)00163-4.

[7] W. Kuanthong, W. Liamlaem, and S. Sasananan, "Clarifier Models using Artificial Neural Networks - Case Study: Bangkhen Water Treatment Plant," *SWU Engineering Journal,* vol. 10, no. 1, pp. 32-44, 2015.

[8] Z. Gormez and A. Sengul, *Prediction of Optimal Coagulant Dosage in Drinking Water Treatment by Artificial Neural Network*. 2013.

[9] S. Sasananan, "Water Treatment Plant Clarifier Control: An Artificial Intelligence Approach," Doctoral dissertation, University of Tasmania, Australia, 2009.

[10] O. Degremont, *Water Treatment Handbook*, 6th ed. France, 1991.

ภาคผนวก

รายละเอียดโค้ดที่ใช้ในการวิจัย

รายละเอียดโค้ดการจัดกลุ่มข้อมูลด้วยเทคนิค SOM

```
1.% Use for SOM clustering FOR model of BKWTP ONLY
2.% change num1,num2 for maximum number of clustering,Kohonan layer
3.% change 'i' and 'net.trainParam.epochs' for max number of epoachs,
4.% function net=newsom([pr],[num1,num2]);
5.% outputs are separately saved sub1,sub2,sub3...,sub'i'
6.% in each sub'..'.mat contain saveNET,numberof cluster,tr,ac
7.% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
8.% Clear all variable
9.clc
10.clear all
11.% Define Kohonan layer
12.num1=25; % CHANGE HERE
13.num2=25; % CHANGE HERE
14.% Load input data
15.load data2.mat ;
16.%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
17.final_turbidity =input;
18.pr=minmax(final_turbidity); % Normalize input data.
19.net=newsom([pr],[num1,num2]);
20.%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
21.for i=1:20 % change here
22.                net.trainParam.epochs = 500; %change here
23.                [net tr]=train(net,final_turbidity);
24. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
25.                if i==1
26.                                saveNET1=net;
```

```
27.                         a=sim(net,final_turbidity);
28.                         ac=vec2ind(a);
29.                         main=1:max(ac);
30.                         [r1 index]=ismember(main,ac);
31.                         num_of_cluster1=max(cumsum(r1));
32.                         'num_of_cluster1'
33.                         num_of_cluster1
34.                         save sub1 saveNET1 tr a ac num_of_cluster1;
35.                         i
36. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
37.                 elseif i ==2
38.                         saveNET2=net;
39.                         a=sim(net,final_turbidity);
40.                         ac=vec2ind(a);
41.                         main=1:max(ac);
42.                         [r1 index]=ismember(main,ac);
43.                         num_of_cluster2=max(cumsum(r1));
44.                         'num_of_cluster2'
45.                         num_of_cluster2
46.                         save sub2 saveNET2 tr a ac num_of_cluster2;
47.                         i
48. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
49.                 elseif i==3
50.                         saveNET3=net;
51.                         a=sim(net,final_turbidity);
52.                         ac=vec2ind(a);
53.                         main=1:max(ac);
54.                         [r1 index]=ismember(main,ac);
55.                         num_of_cluster3=max(cumsum(r1));
```

```
56.                         'num_of_cluster3'
57.                         num_of_cluster3
58.                         save sub3 saveNET3 tr a ac num_of_cluster3;
59.                         i
60.   %%%%%%%%%%%%%%%%%%%%%%%%%%%%
61.               elseif i==4
62.                         saveNET4=net;
63.                         a=sim(net,final_turbidity);
64.                         ac=vec2ind(a);
65.                         main=1:max(ac);
66.                         [r1 index]=ismember(main,ac);
67.                         num_of_cluster4=max(cumsum(r1));
68.                         'num_of_cluster4'
69.                         num_of_cluster4
70.                         save sub4 saveNET4 tr a ac num_of_cluster4;
71.                         i
72.   %%%%%%%%%%%%%%%%%%%%%%%%%%%%
73.               elseif i==5
74.                         saveNET5=net;
75.                         a=sim(net,final_turbidity);
76.                         ac=vec2ind(a);
77.                         main=1:max(ac);
78.                         [r1 index]=ismember(main,ac);
79.                         num_of_cluster5=max(cumsum(r1));
80.                         'num_of_cluster5'
81.                         num_of_cluster5
82.                         save sub5 saveNET5 tr a ac num_of_cluster5;
83.                         i
84.   %%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
85.             elseif i==6
86.                     saveNET6=net;
87.                     a=sim(net,final_turbidity);
88.                     ac=vec2ind(a);
89.                     main=1:max(ac);
90.                     [r1 index]=ismember(main,ac);
91.                     num_of_cluster6=max(cumsum(r1));
92.                     'num_of_cluster6'
93.                     num_of_cluster6
94.                     save sub6 saveNET6 tr a ac num_of_cluster6;
95.                     i
96. %%%%%%%%%%%%%%%%%%%%%%%%%%%%
97.             elseif i==7
98.             saveNET7=net;
99.             a=sim(net,final_turbidity);
100.            ac=vec2ind(a);
101.            main=1:max(ac);
102.            [r1 index]=ismember(main,ac);
103.            num_of_cluster7=max(cumsum(r1));
104.            'num_of_cluster7'
105.            num_of_cluster7
106.            save sub7 saveNET7 tr a ac num_of_cluster7;
107.            i
108. %%%%%%%%%%%%%%%%%%%%%%%%%%%%
109.   elseif i==8
110.            saveNET8=net;
111.            a=sim(net,final_turbidity);
112.            ac=vec2ind(a);
113.            main=1:max(ac);
```

```
114.              [r1 index]=ismember(main,ac);
115.              num_of_cluster8=max(cumsum(r1));
116.              'num_of_cluster8'
117.              num_of_cluster8
118.              save sub8 saveNET8 tr a ac num_of_cluster8;
119.              i
120.  %%%%%%%%%%%%%%%%%%%%%%%%%%%
121.    elseif i==9
122.              saveNET9=net;
123.              a=sim(net,final_turbidity);
124.              ac=vec2ind(a);
125.              main=1:max(ac);
126.              [r1 index]=ismember(main,ac);
127.              num_of_cluster9=max(cumsum(r1));
128.              'num_of_cluster9'
129.              num_of_cluster9
130.              save sub9 saveNET9 tr a ac num_of_cluster9;
131.              i
132.  %%%%%%%%%%%%%%%%%%%%%%%%%%%
133.    elseif i==10
134.              saveNET10=net;
135.              a=sim(net,final_turbidity);
136.              ac=vec2ind(a);
137.              main=1:max(ac);
138.              [r1 index]=ismember(main,ac);
139.              num_of_cluster10=max(cumsum(r1));
140.              'num_of_cluster10'
141.              num_of_cluster10
142.              save sub10 saveNET10 tr a ac num_of_cluster10;
```

```
143.              i
144. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
145.     elseif i==11
146.              saveNET11=net;
147.              a=sim(net,final_turbidity);
148.              ac=vec2ind(a);
149.              main=1:max(ac);
150.              [r1 index]=ismember(main,ac);
151.              num_of_cluster11=max(cumsum(r1));
152.              'num_of_cluster11'
153.              num_of_cluster11
154.              save sub11 saveNET11 tr a ac num_of_cluster11;
155.              i
156. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
157.              elseif i==12
158.              saveNET12=net;
159.              a=sim(net,final_turbidity);
160.              ac=vec2ind(a);
161.              main=1:max(ac);
162.              [r1 index]=ismember(main,ac);
163.              num_of_cluster12=max(cumsum(r1));
164.              'num_of_cluster12'
165.              num_of_cluster12
166.              save sub12 saveNET12 tr a ac num_of_cluster12;
167.              i
168. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
169.              elseif i==13
170.              saveNET13=net;
171.              a=sim(net,final_turbidity);
```

```
172.            ac=vec2ind(a);
173.            main=1:max(ac);
174.            [r1 index]=ismember(main,ac);
175.            num_of_cluster13=max(cumsum(r1));
176.            'num_of_cluster13'
177.            num_of_cluster13
178.            save sub13 saveNET13 tr a ac num_of_cluster13;
179.            i
180. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
181.            elseif i==14
182.            saveNET14=net;
183.            a=sim(net,final_turbidity);
184.            ac=vec2ind(a);
185.            main=1:max(ac);
186.            [r1 index]=ismember(main,ac);
187.            num_of_cluster14=max(cumsum(r1));
188.            'num_of_cluster14'
189.            num_of_cluster14
190.            save sub14 saveNET14 tr a ac num_of_cluster14;
191.            i
192. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
193.            elseif i==15
194.            saveNET15=net;
195.            a=sim(net,final_turbidity);
196.            ac=vec2ind(a);
197.            main=1:max(ac);
198.            [r1 index]=ismember(main,ac);
199.            num_of_cluster15=max(cumsum(r1));
200.            'num_of_cluster15'
```

```
201.            num_of_cluster15
202.            save sub15 saveNET15 tr a ac num_of_cluster15;
203.            i
204. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
205.            elseif i==16
206.            saveNET16=net;
207.            a=sim(net,final_turbidity);
208.            ac=vec2ind(a);
209.            main=1:max(ac);
210.            [r1 index]=ismember(main,ac);
211.            num_of_cluster16=max(cumsum(r1));
212.            'num_of_cluster16'
213.            num_of_cluster16
214.            save sub16 saveNET16 tr a ac num_of_cluster16;
215.            i
216. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
217.            elseif i==17
218.            saveNET17=net;
219.            a=sim(net,final_turbidity);
220.            ac=vec2ind(a);
221.            main=1:max(ac);
222.            [r1 index]=ismember(main,ac);
223.            num_of_cluster17=max(cumsum(r1));
224.            'num_of_cluster17'
225.            num_of_cluster17
226.            save sub17 saveNET17 tr a ac num_of_cluster17;
227.            i
228. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
229.            elseif i==18
```

```
230.            saveNET18=net;
231.            a=sim(net,final_turbidity);
232.            ac=vec2ind(a);
233.            main=1:max(ac);
234.            [r1 index]=ismember(main,ac);
235.            num_of_cluster18=max(cumsum(r1));
236.            'num_of_cluster18'
237.            num_of_cluster18
238.            save sub18 saveNET18 tr a ac num_of_cluster18;
239.            i
240. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
241.            elseif i==19
242.            saveNET19=net;
243.            a=sim(net,final_turbidity);
244.            ac=vec2ind(a);
245.            main=1:max(ac);
246.            [r1 index]=ismember(main,ac);
247.            num_of_cluster19=max(cumsum(r1));
248.            'num_of_cluster19'
249.            num_of_cluster19
250.            save sub19 saveNET19 tr a ac num_of_cluster19;
251.            i
252. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
253.            elseif i==20
254.            saveNET20=net;
255.            a=sim(net,final_turbidity);
256.            ac=vec2ind(a);
257.            main=1:max(ac);
258.            [r1 index]=ismember(main,ac);
```

```
259.            num_of_cluster20=max(cumsum(r1));
260.            'num_of_cluster20'
261.            num_of_cluster20
262.            save sub20 saveNET20 tr a ac num_of_cluster20;
263.            i
264. %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
265.            end
266.end
```

รายละเอียดโค้ดสร้างแบบจำลอง ANN

```
1.% Clear all variable
2.clear all
3.close all
4.clc
5.% Load input/output data
6.load input_test.mat
7.load output_alum.mat
8.% Initialize for test data
9.test_r_sqr = [];
10.test_mse = [];
11.test_mae = [];
12.test_R = [];
13.Rq = [];
14.%Preparation data for model
15.% Transpose input data and normalize
16.k = normalize(input_test.', 'range', [-0.75 0.75]);
17.% Transpose input data and assign new variable
18.x =k.';
19.% For loop to run ANN 2 layer.
20.for i=5:5:100
```

```
21.     for j=5:5:100
22.% Delete data some variable.
23.             close all
24.             clear net,clear tr,clear St1,clear Sr1,clear r21,clear R,clear error1,clear
MSEperf1,clear MAEperf1,
25.             clear y
26.             rng(6); % Defined random state.
27.             %Create neural network 2 layer, defined activation function and algorithm
for tranning network.
28.             net=  newff(minmax(x),[i,j,1],{'tansig','tansig','purelin'},'trainlm');
29.             input = x(:,1:717); %Defined range of input
30.             target = output_alum(:,1:717); %Defined range of output
31.             net.divideFcn='divideind'; %Defined type data division
32.             %Assgin traning set,validation set and test set
33.             [trainInd,valInd,testInd] = divideind(717,1:431,432:574,575:717);
34.             net.divideParam.trainInd = trainInd;
35.             net.divideParam.valInd = valInd;
36.             net.divideParam.testInd = testInd;
37.             % Train and validation neural network
38.             [net,tr]=train(net,input,target);
39.             y = net(input);
40.             % Calculation R_square,MSE,MAE and R
41.             St1 = sum(( target(tr.testInd).' - mean(target(tr.testInd).') ).^2);
42.             Sr1 = sum(( target(tr.testInd).' - round(y(tr.testInd).',3) ).^2);
43.             r21 = (St1 - Sr1) / St1;
44.             R = corr2(target(tr.testInd),round(y(tr.testInd),3));
45.             error1=target(tr.testInd)-round(y(tr.testInd),3);
46.             MSEperf1=mse(error1);
47.             MAEperf1=mae(error1);
48.             test_r_sqr = round([test_r_sqr,r21],2);
```

```
49.                 test_mse   = round([test_mse,MSEperf1],2);
50.                 test_mae   = round([test_mae,MAEperf1],2);
51.                 test_R     = round([test_R,R],2);
52.                 Rq         = [Rq tr.trainInd];
53.         end
55.end
```

ประวัติผู้เขียน

ชื่อ-สกุล	นิลวัฒน์ เฟื่องโชค
วัน เดือน ปี เกิด	21 ธันวาคม 2537
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	ปริญญาตรี หลักสูตรวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์และระบบคอมพิวเตอร์ มหาวิทยาลัยศิลปากร
ที่อยู่ปัจจุบัน	104 ซอยเอกชัย 83 เขตบางบอน แขวงบางบอน กรุงเทพมหานคร 10150