บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. ทฤษฎีที่เกี่ยวข้อง

1.1 มาตรฐานแหล่งน้ำผิวดิน

มาตรา 32 แห่งพระราชบัญญัติส่งเสริมและรักษาคุณภาพสิ่งแวดล้อมแห่งชาติ พ.ศ. 2535 บัญญัติให้คณะกรรมการสิ่งแวดล้อมแห่งชาติ กำหนดมาตรฐานคุณภาพสิ่งแวดล้อมเพื่อเป็นเป้าหมายในการรักษาคุณภาพสิ่งแวดล้อมให้อยู่ในเกณฑ์ที่เหมาะสม ซึ่งมาตรฐานคุณภาพสิ่งแวดล้อม นี้จะต้องอาศัยหลักวิชาการ และหลักการทางวิทยาศาสตร์เป็นพื้นฐานโดยจะต้องคำนึงถึงความเป็นไปได้ในเชิงเศรษฐกิจ สังคม และเทคโนโลยีที่ เกี่ยวข้อง [7]

มาตรฐานคุณภาพแหล่งน้ำเป็นมาตรฐานคุณภาพสิ่งแวดล้อมชนิดหนึ่ง มีวัตถุประสงค์

1. เพื่อควบคุมและรักษาคุณภาพน้ำในแหล่งน้ำให้เหมาะสมกับการใช้ประโยชน์ และมีความปลอดภัยต่อสุขภาพอนามัยของประชาชน

2. เพื่ออนุรักษ์ทรัพยากร และสภาพแวดล้อมตามธรรมชาติ

หลักการสำคัญในการกำหนดมาตรฐานคุณภาพแหล่งน้ำ ได้แก่ การกำหนดค่ามาตรฐานเพื่อรักษาคุณภาพน้ำให้เหมาะสมกับการใช้ประโยชน์การจัดแบ่งลักษณะการใช้ประโยชน์ของแหล่งน้ำ และการกำหนดหลักเกณฑ์และวิธีการตรวจสอบคุณภาพน้ำ

1.1.1 หลักเกณฑ์ในการพิจารณากำหนดมาตรฐานคุณภาพน้ำ

1. ความเหมาะสมต่อการนำมาใช้ประโยชน์ในกิจกรรมต่อละประเภทในกรณีที่แหล่งน้ำนั้นมีการใช้ประโยชน์หลายด้าน โดยคำนึงถึงการใช้ประโยชน์หลักเป็นสำคัญ ทั้งนี้ ระดับมาตรฐานจะไม่ขัดแย้งต่อการใช้ประโยชน์หลายด้านพร้อมกัน

2. สถานการณ์คุณภาพน้ำในแหล่งน้ำหลักของประเทศและแนวโน้มของคุณภาพน้ำที่อาจมีการเปลี่ยนแปลงเนื่องจาการพัฒนาด้านต่าง ๆ ในอนาคต

3. คำนึงถึงสุขภาพและความปลอดภัยของชีวิตมนุษย์และสัตว์น้ำส่วนใหญ่

4. ความรู้สึกพึงพอใจในการยอมรับระดับคุณภาพน้ำในเขตต่าง ๆ ของประชาชนในพื้นที่ลุ่มน้ำหลักและของประชาชนส่วนใหญ่

อย่างไรก็ตาม การปรับปรุงค่ามาตรฐานในอนาคต จำเป็นจะต้องพิจารณาถึงความเหมาะสมของระดับการลงทุนและภาวะทางเศรษฐกิจในพื้นที่ลุ่มน้ำ ที่อยู่ในแผนการพัฒนาตลอดจนความเป็นไปได้ในเทคโนโลยีในการบำบัดของเสียและสารพิษจากแหล่งกำเนิดของเสีย ซึ่งได้แก่

กิจกรรมที่เกิดขึ้นจากการวางแผนพัฒนาเศรษฐกิจและสังคมด้วย

1.1.2 การกำหนดประเภทแหล่งน้ำผิวดิน

ประเภทที่ 1ได้แก่ แหล่งน้ำที่คุณภาพน้ำมีสภาพตามธรรมชาติโดยปราศจากน้ำทิ้งจากกิจกรรมทุกประเภทและสามารถเป็นประโยชน์เพื่อ

1. การอุปโภคและบริโภคโดยต้องผ่านการฆ่าเชื้อโรคตามปกติก่อน
2. การขยายพันธุ์ตามธรรมชาติของสิ่งมีชีวิตระดับพื้นฐาน
3. การอนุรักษ์ระบบนิเวศน์ของแหล่งน้ำ

ประเภทที่ 2ได้แก่ แหล่งน้ำที่ได้รับน้ำทิ้งจากกิจกรรมบางประเภท และสามารถเป็นประโยชน์เพื่อ

1. การอุปโภคและบริโภคโดยต้องผ่านการฆ่าเชื้อโรคตามปกติและผ่านกระบวนการปรับปรุงคุณภาพน้ำทั่วไปก่อน
2. การอนุรักษ์สัตว์น้ำ
3. การประมง
4. การว่ายน้ำและกีฬาทางน้ำ

ประเภทที่ 3ได้แก่ แหล่งน้ำที่ได้รับน้ำทิ้งจากกิจกรรมบางประเภท และสามารถเป็นประโยชน์เพื่อ

1. การอุปโภคและบริโภคโดยต้องผ่านการฆ่าเชื้อโรคตามปกติและผ่านกระบวนการปรับปรุงคุณภาพน้ำทั่วไปก่อน
2. การเกษตร

ประเภทที่ 4ได้แก่ แหล่งน้ำที่ได้รับน้ำทิ้งจากกิจกรรมบางประเภท และสามารถเป็นประโยชน์เพื่อ

1. การอุปโภคและบริโภคโดยต้องผ่านการฆ่าเชื้อโรคตามปกติและผ่านกระบวนการปรับปรุงคุณภาพน้ำเป็นพิเศษก่อน
2. การอุตสาหกรรม

ประเภทที่ 5ได้แก่ แหล่งน้ำที่ได้รับน้ำทิ้งจากกิจกรรมบางประเภท และสามารถเป็นประโยชน์เพื่อการคมนาคม

1.2 ดัชนีคุณภาพน้ำทั่วไป (General Water Quality Index ,WQI)

ดัชนีคุณภาพน้ำเป็นดัชนีที่บ่งบอกสภาพของแม่น้ำโดยทั่วไป โดยมิได้ระบุโดยตรงว่าสามารถนำไปใช้ประโยชน์อะไรได้ โดยเหตุนี้จึงเรียกว่า ดัชนีคุณภาพน้ำทั่วไป (General Water Quality Index) เพื่อบ่งบอก ระดับคุณภาพน้ำว่าอยู่ในเกณฑ์ดีมาก ดีพอใช้หรือต่ำ ซึ่งจะทำให้ทราบว่าแม่น้ำดังกล่าวจะต้องดำเนินการควบคุมดูแลอย่างไรบ้าง ซึ่งจะแก้ไขมากน้อยเพียงไร ก็ต้องดูว่าอาการที่เกิดขึ้นรุนแรงมากหรือน้อยและสาเหตุเกิดเนื่องมาจาก ธรรมชาติเอง เช่น ความขุ่น หรือจากการกระทำของมนุษย์ เช่นการระบายน้ำเสีย Unweighted Multiplicative River Water Quality Index เป็นวิธีใช้ในการเผยแพร่ให้ความรู้ทางด้านคุณภาพน้ำแก่สาธารณชนทราบ ด้วยคำที่ง่าย วิธีการรวบรัด และเข้าใจโดยง่าย ไม่สลับซับซ้อน ซึ่งใช้อยู่ในสหรัฐอเมริกา และเป็นวิธีหนึ่งที่ถูกใช้ในการจัดทำรายงานเสนอต่อสภาผู้แทนราษฎรของสหรัฐอเมริกา (Brown,1970) ดัชนีคุณภาพน้ำทั่วไป (WQI) ที่กล่าวถึง มีหน่วยเป็นคะแนน เริ่มจาก 0 ถึง 100 คะแนน 91–100 คะแนน ถือว่าคุณภาพน้ำอยู่ในเกณฑ์ดีมาก 71-90 คะแนน คุณภาพน้ำอยู่ในเกณฑ์ดี 61-70 คะแนน คุณภาพน้ำอยู่ในเกณฑ์พอใช้ 31-60 คะแนน คุณภาพน้ำอยู่ในเกณฑ์เสื่อมโทรม 0-30 คะแนนคุณภาพน้ำอยู่ในเกณฑ์เสื่อมโทรมมาก

คะแนนเหล่านี้คำนวณจากการรวมคะแนน ดัชนีคุณภาพน้ำ 5 ดัชนี ได้แก่ ออกซิเจนละลายน้ำ (DO), แบคทีเรียกลุ่มฟีคัลโคลิฟอร์ม (Fecal Coliform Bacteria, FCB) , แอมโมเนีย (NH_3) , การปนเปื้อนของแบคทีเรียกลุ่มโคลิฟอร์มทั้งหมด(TCB) และความสกปรกในรูปสารอินทรีย์ (Biological Oxygen Demand, BOD) เข้าด้วยกันเป็นคะแนนรวมอย่างเดียว [7]

ค่า WQI ที่ได้จากการรวมคะแนนดัชนีคุณภาพน้ำ 5 ดัชนี สามารถนำมาวิเคราะห์ร่วมกับมาตรฐานคุณภาพน้ำผิวดินได้ วิธีการคำนวณยังให้ผลที่สอดคล้องกัน

1.3 การทำความสะอาดข้อมูล (Data Cleansing)

เป็นกระบวนการที่ตรวจจับและตรวจความถูกต้อง ข้อมูลที่ผิดปกติหรือไม่ถูกต้อง ของข้อมูลที่อยู่ในฐานข้อมูลหรือชุดข้อมูลที่กำหนด ซึ่งใช้อย่างแพร่หลายในระบบฐานข้อมูล ซึ่งวิธีการนี้จะช่วยในการจัดการข้อมูลที่ ไม่สมบูรณ์ ผิดพลาด หรือไม่มีความเกี่ยวข้องกัน เป็นต้น ซึ่งวิธีการแก้ไขมีทั้ง การลบข้อมูล ปรับค่า หรือแทนด้วยค่าใหม่ หลังจากแก้ไขแล้วชุดข้อมูลจะไม่มีความเปลี่ยนแปลงจากข้อมูลเดิมมากนัก

1.3.1 กระบวนการในการทำความสะอาดข้อมูล

Data auditing ข้อมูลจะถูกตรวจสอบโดยกระบวนการทางสถิติ เพื่อตรวจหาสิ่งผิดปกติในชุดข้อมูล กระบวนการนี้ทำให้ทราบลักษณะของสิ่งผิดปกติของข้อมูล และเพื่อให้ทราบตำแหน่งของข้อมูลเหล่านั้น

Workflow specification การลบหรือแก้ไขสิ่งผิดปกติในชุดข้อมูล จะมีการทำงานเป็นลำดับ ที่เรียกว่า work flow ซึ่งกระบวนการนี้จะต่อเนื่องจากการ Data auditing และกระบวนการ workflow ข้อมูลที่จะเข้าสู่กระบวนการนี้จะต้องถูกพิจารณาข้อผิดพลาดอย่างละเอียดเสียก่อน

Workflow execution ในกระบวนการส่วนนี้จะทำการแก้ไข เพื่อให้ได้ชุดข้อมูลตามที่ได้พิจารณาไว้ก่อนแล้ว ซึ่งการเตรียมการกระบวนการก่อนหน้านี้ ให้พร้อมและเหมาะสมจะช่วยให้ระบบทำงานได้ดีขึ้น ซึ่งจะช่วยลด cost ของการคำนวณโดยคอมพิวเตอร์ได้

Post-Processing and Controlling กระบวนการนี้คือการตรวจสอบความถูกต้อง เนื่องจากกระบวนการที่ผ่านมามีการแก้ไขแต่ไม่ได้ตรวจสอบข้อมูลไปด้วย อาจใช้วิธีการตรวจสอบด้วยผู้วิจัย ดูด้วยตาเปล่า หากผลลัพธ์ไม่ได้ดังที่ต้องการอาจมีการทำซ้ำกระบวนการทั้งหมดอีกครั้ง

1.3.2 วิธีการที่นิยมใช้ในการทำความสะอาดข้อมูล

Parsing ในการทำความสะอาดข้อมูลวิธีการนี้เป็นการตรวจจับ syntax errors กระบวนการจะตรวจสอบคำ ว่าตรงกับที่ต้องการหรือไม่ ซึ่งเป็นวิธีการที่คล้ายคลึงกับการตรวจสอบไวยากรณ์หรือภาษา

Data Transformation กระบวนการนี้เป็นการแปลงข้อมูลจากรูปแบบหนึ่งเป็นอีกรูปแบบหนึ่งที่ต้องการใช้งาน

Duplicate Elimination เป็นวิธีการตรวจสอบความซ้ำซ้อนของข้อมูล โดยการใช้อัลกอริทึม เข้ามาช่วย

Statistical Methods การวิเคราะห์ข้อมูลโดยใช้ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ขอบเขตข้อมูล หรือการจัดกลุ่ม เป็นไปได้ที่จะพบข้อมูลที่ผิดปกติ และกระบวนการทางสถิติสามารถช่วยแก้ปัญหาข้อมูลที่หายไปในชุดข้อมูล โดยแทนค่าด้วยข้อมูลที่ยอมรับได้ ซึ่งได้จากกระบวนการวิเคราะห์โดยใช้อัลกอริทึม ที่เหมาะสม

1.4 การทำเหมืองข้อมูล

เป็นวิธีการที่ทำการค้นหาองค์ความรู้จากข้อมูลที่มีอยู่จำนวนมาก โดยองค์ความรู้ที่ได้มีทั้งรูปแบบของข้อมูล แนวโน้มของข้อมูล ความสัมพันธ์กันของข้อมูล เพื่อให้ได้สิ่งที่กล่าวมาแล้ว เทคนิคการทำเหมืองข้อมูล (Data mining)ถูกนำมาใช้ โดยสามารถค้นหารูปแบบในข้อมูล วิเคราะห์ทางสถิติ รวบรวมข้อมูล การทำเหมืองข้อมูลจะมีการใช้อัลกอริทึมเฉพาะ โดยเป้าหมายหลักจะมีอยู่ 2 ส่วน คือการค้นพบ (Discovery)และการตรวจสอบความถูกต้อง(Verification) ในส่วนของการค้นพบคือค้นพบรูปแบบของข้อมูลและสามารถแบ่งย่อยเป็นการพยากรณ์(Prediction) ซึ่งเป็นผลของการค้นพบรูปแบบของข้อมูล จึงสามารถพยากรณ์ข้อมูลได้ การทำเหมืองข้อมูลมีวิธีการดังนี้

การคัดแยกข้อมูล (Classification) เป็นรูปแบบการเรียนรู้ ที่วิเคราะห์ข้อมูลโดยอ้างอิงกับรูปแบบข้อมูลที่มีการวิเคราะห์ความถูกต้อง เหมาะสมอยู่แล้ว และสามารถใช้ในการพยากรณ์ข้อมูลได้

การจัดกลุ่มข้อมูล (Clustering) เป็นการสร้างรูปแบบของเซต หรือกลุ่มที่สามารถอธิบายข้อมูล หรือแบ่งแยกข้อมูลได้

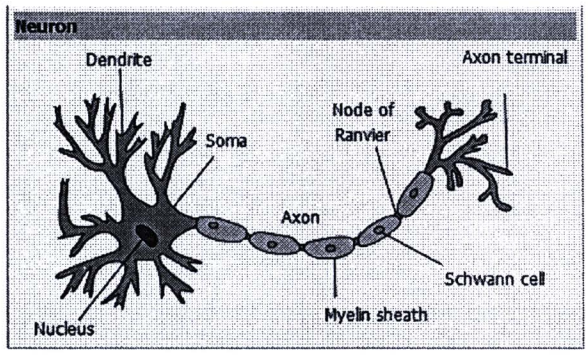
กฎความสัมพันธ์ของข้อมูล(Associate Rules) การค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ

อัลกอริทึมที่ใช้ในการทำเหมืองข้อมูล มีหลายรูปแบบแล้วแต่ผู้ใช้ที่ต้องการวิเคราะห์ข้อมูลในรูปแบบใด เช่น การคัดแยกข้อมูลสามารถใช้อัลกอริทึมต้นไม้การตัดสินใจ , C4.5 หรือโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชันและการจัดกลุ่มข้อมูล สามารถใช้อัลกอริทึมโครงข่ายประสาทเทียมแบบ Self organizing map และการหากฎความสัมพันธ์ของข้อมูลใช้ Apiori เป็นต้น

ในงานวิจัยนี้จะใช้เทคนิคการคัดแยกข้อมูล เป็นหลักซึ่งวิธีการนี้ที่สามารถบ่งบอกถึงความแตกต่างของข้อมูลและสามารถวิเคราะห์ว่าข้อมูลนั้นๆควรจะอยู่ในกลุ่มใด ซึ่ง เป็นวิธีการหนึ่งในการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยแบบจำลองที่ใช้ในการคัดแยกข้อมูลจะมีกระบวนการสร้างฟังก์ชันเชื่อมโยงกันระหว่างข้อมูลที่เรียนรู้กับกลุ่มที่ต้องการ ดังนั้นจึงมีการให้ข้อมูลเพื่อให้แบบจำลองได้เรียนรู้ ซึ่งข้อมูลเหล่านั้นจะประกอบด้วย คุณสมบัติของข้อมูลและกลุ่มที่ต้องการ เมื่อแบบจำลองได้เรียนรู้แล้ว จะมีความสามารถในการคัดแยกชุดข้อมูลอื่นๆที่ได้พบในภายหลังได้

1.4.1 โครงข่ายประสาทเทียม (Artificial neural network)

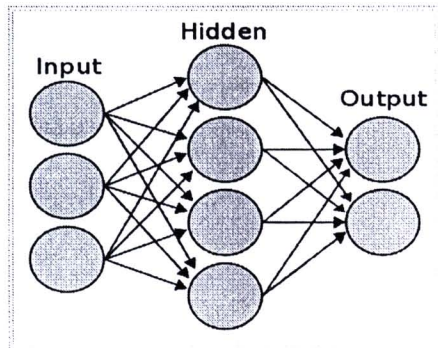
เป็นเทคโนโลยีที่มีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ (Artificial Intelligence) โดยการจำลองการทำงานของเซลล์ประสาทในสมองของมนุษย์



ภาพที่ 1 เซลล์ประสาทในสมองของมนุษย์ [8]

ในการประมวลผลโดยสิ่งที่สำคัญในการจำลองเซลล์ประสาทคือ การจำลอง synapses ทั้งหลายในโครงข่ายประสาทเทียม ซึ่งเปรียบเสมือนแหล่งสะสมความรู้ของสมอง การจำลอง synapses นั้นใช้หลักการที่ว่า แต่ละ synapses ทำหน้าที่เป็นตัวปรับเปลี่ยนสภาพสัญญาณไฟฟ้าที่ส่งมาจากเซลล์ประสาทตัวอื่นๆ ก่อนส่งสัญญาณนั้นผ่าน Dendrite เข้าสู่ตัวเซลล์ประสาท และการปรับเปลี่ยนสัญญาณดังกล่าว จะขึ้นอยู่กับความเหนียวแน่นของการเชื่อมต่อบริเวณรอยต่อ synapses โดยความแข็งแกร่งนี้จะเปลี่ยนไปตามความรู้ที่สมองได้เรียนเข้าไป และใช้ตัวแปรตัวหนึ่งเรียกว่าค่าน้ำหนัก (weight) ในการจำลองsynapses โดยวิธีการดังกล่าวสามารถประยุกต์ให้คอมพิวเตอร์สามารถทำการเรียนรู้ได้ และแก้ปัญหาที่มนุษย์กำหนดให้ได้ ดังนั้นจึงมีการประยุกต์ใช้โครงข่ายประสาทเทียมในงานเกี่ยวกับการทำเหมืองข้อมูล ซึ่งในการจำแนกข้อมูลมีการนำเทคนิคนี้ไปใช้อย่างแพร่หลาย

โครงสร้างของโครงข่ายประสาทเทียมจะประกอบด้วยนิวรอล สำหรับ Input , Output และการประมวลผล กระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ชั้น Input, ชั้น Output และชั้น Hidden การประมวลผลของโครงข่ายประสาเทียมจะอาศัยการส่งการทำงานผ่านนิวรอล ต่าง ๆ ในชั้นเหล่านี้



ภาพที่ 2 แบบจำลองโครงข่ายประสาทเทียม [9]

ในแต่ละชั้นของโครงข่ายประสาทเทียมจะจำลองให้คล้ายคลึงกับการทำงานของเซลล์ประสาท ลักษณะการทำงานจะเริ่มจากส่งข้อมูลเข้าที่ชั้น Input และผ่านไปยังชั้น Hidden เพื่อประมวลผล และได้ผลลัพธ์ออกมาที่ชั้น Output

เพื่อให้โครงข่ายประสาทเทียมสามารถทำงานได้ตามปัญหาที่ต้องการแก้ไขได้ จำเป็นต้องมีการสอนโครงข่ายให้เรียนรู้ข้อมูลที่ใช้ โดยลักษณะการเรียนรู้ของโครงข่ายประสาทเทียมเป็นดังนี้

การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ การให้โครงข่ายได้เรียนรู้ข้อมูลที่ถูกต้องก่อนและทำการปรับค่าน้ำหนักในโครงข่าย และสามารถสร้างผลลัพธ์ออกมาตามที่ได้เรียนรู้ หรือกล่าวได้ว่ากำหนด คู่ของข้อมูล $(x,y),x\in X,y\in Y$ และมีความต้องการที่จะค้นหาฟังก์ชัน $f:X\rightarrow Y$ ซึ่งผลตอบจะเป็นอย่างไร ก็ขึ้นอยู่กับสภาวะในตอนที่เริ่มเรียนรู้ของโครงข่ายประสาทเทียม การเรียนรู้รูปแบบนี้มักใช้ในงานจำพวก การรู้จำรูปแบบ (pattern recognition) และการวิเคราะห์ถดถอย (regression)

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) วิธีการนี้ไม่จำเป็นต้องมีค่าเป้าหมายของแต่ละข้อมูลตัวอย่าง ในระหว่างการเรียนรู้ โครงข่ายประสาทเทียมค้นหาความสัมพันธ์ของข้อมูลด้วยตนอง โดยทำการค้นหารูปแบบเฉพาะของข้อมูลนั้นๆ และจะทำการจัดกลุ่มรูปแบบต่างๆ เหล่านั้นเองตามต้องการ การเรียนรู้รูปแบบนี้มักใช้ในงานจำพวก clustering

1.4.2 กลไกเรเดียลเบสิสฟังก์ชัน (Radial basis function Neural Network)

ในช่วงท้ายยุคปี 80 โครงข่ายประสาทเทียมรูปแบบหนึ่งมีการผนวกใช้เรเดียลเบสิสฟังก์ชัน ซึ่งเรเดียลเบสิสฟังก์ชัน ถูกใช้ในส่วนชั้น Hidden และชั้นOutput โดยที่ชั้นHidden จะใช้ ฟังก์ชันกระตุ้นแบบเรเดียลเบสิสฟังก์ชัน ในส่วนของ input ที่เข้าสู่โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน จะเป็นแบบไม่เชิงเส้นแต่ในส่วนของ Output จะเป็นเชิงเส้น เมื่อต้องการที่จะใช้งานโครงข่ายต้องมีการกำหนดฟังก์ชันกระตุ้นที่ใช้ และกำหนดจำนวนของนิวรอลที่จะประมวลผล และกำหนดอัลกอริทึมที่ใช้ในการเรียนรู้ (Training Algorithm) ซึ่งโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน นั้นมีลักษณะการเรียนรู้แบบมีผู้สอน (Supervised training) จึงต้องมีข้อมูล input – output ชุดหนึ่งที่ต้องให้โครงข่ายเรียนรู้และการเรียนรู้ของโครงข่ายก็คือการหาค่าน้ำหนัก (weight)นั่นเอง

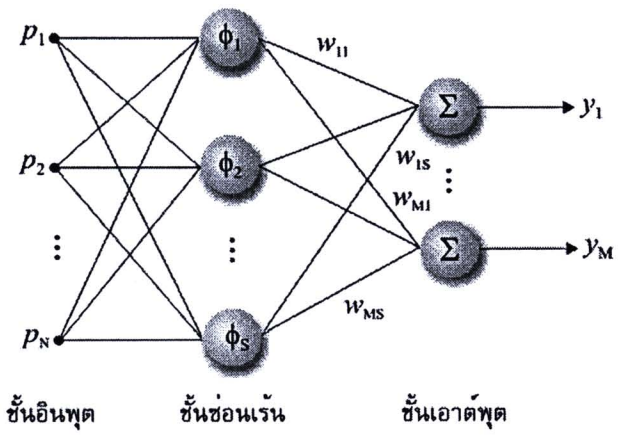
โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชันมีลักษณะเป็นโครงข่ายประสาทเทียมแบบไปข้างหน้า (feed forward networks) โครงสร้างของโครงข่ายเรเดียลเบสิสฟังก์ชัน โดยทั่วไปมีลักษณะดังนี้

ชั้น Input ข้อมูล input ที่ชั้นนี้จะมีลักษณะเป็น Input Vector เหมือนในโครงข่ายแบบ

Multilayer perceptron ทั่วๆไป

ชั้น Hidden ในชั้น Hidden จะมีฟังก์ชันที่มีลักษณะการทำงานที่พิเศษคือ ผลตอบสนองของฟังก์ชันจะขึ้นอยู่กับระยะทางของ Input กับจุดศูนย์กลางของฟังก์ชัน ถ้าหาก Input Vector อยู่ใกล้จุดศูนย์กลางของฟังก์ชันมากจะได้ Output ที่มากด้วย

ชั้น Output ทำหน้าที่รวบรวมผลลัพธ์ที่ได้จากชั้น Hidden



ภาพที่ 3 กลไกเรเดียลเบสิสฟังก์ชัน [10]

จากโครงข่ายดังภาพที่ 3 จะได้ว่า Output ของกลไกเรเดียลเบสิสฟังก์ชัน มีค่าดังสมการ (1)

$$\begin{aligned} y_i &= \sum_{k=1}^{S} w_{ik}\phi_k(\mathbf{p}, \mathbf{c}_k) \\ &= \sum_{k=1}^{S} w_{ik}\phi_k(\|\mathbf{p} - \mathbf{c}_k\|_2) \end{aligned} \tag{1}$$

โดยที่ $\phi_k(\cdot)$ คือ ฟังก์ชันส่งค่า(Transfer function)

$\|\cdot\|_2$ คือ ฟังก์ชันระยะทางแบบยูคลิด

w_{ik} คือ ค่าน้ำหนักประสาทในชั้น Hidden

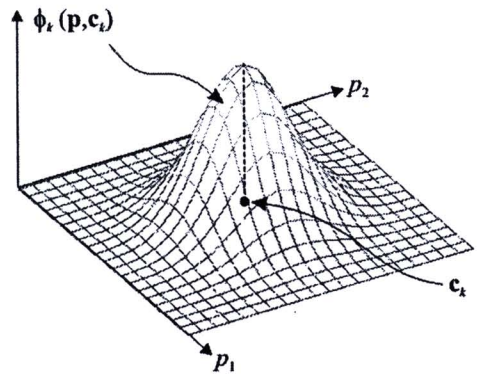
S คือ จำนวนนิวรอลในชั้น Hidden

$\mathbf{c}_k$ คือ เวคเตอร์จุดศูนย์กลางในชั้น Hidden

ในแต่ละนิวรอลของชั้น Hidden ระยะทางของจุดศูนย์กลางและ Input Vector จะถูกคำนวณ และ Output ของแต่ละนิวรอลจะได้จากฟังก์ชัน $\phi_k(\cdot)$ และ Output สุดท้ายของโครงข่ายจะได้จาก

สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุดงานวิจัย
วันที่ 10 ก.ย. 2555
เลขทะเบียน 248351
เลขเรียกหนังสือ

ผลรวมของค่าน้ำหนักและ Output ของแต่ละนิวรอล ในการคัดแยกข้อมูลนิยมเลือกใช้ Gaussian function เพื่อใช้เป็นฟังก์ชันการส่งค่า



ภาพที่ 4 Gaussian function [10]

โดยรวมแล้วโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน คือ การปรับเส้นโค้งระหว่างข้อมูล Input กับ Output นั่นเอง โดยมีลักษณะเป็นการ Mapping ระหว่าง Input และ Output และเป็นที่ยอมรับกันอย่างแพร่หลายว่าเป็นโครงข่ายที่มีประสิทธิภาพสูง

1.4.3 กลไกแบบคณะกรรมการ (Committee machine)

กลไกแบบคณะกรรมการ [3] เป็นรูปแบบหนึ่งของโครงข่ายประสาทเทียมที่ใช้เทคนิค divided and conquer ที่จะทำการรวมผลลัพธ์ของโครงข่ายประสาทเทียมหลายๆอันให้เป็นผลลัพธ์เดียว ซึ่งทำให้ประสิทธิภาพของการการคำนวณสูงกว่าการคำนวณเพียงโครงข่ายประสาทเทียมอันเดียว และยังสามารถแยก input ให้เป็นหลายๆกลุ่มเพื่อคำนวณได้ อีกทั้งยังช่วยลดความซับซ้อนในการคำนวณ

ในงานวิจัยนี้เลือกใช้โครงข่ายคณะกรรมการแบบ Boosting ซึ่งมีลักษณะการทำงานคือในสมาชิกของกลไกแบบคณะกรรมการจะทำการเรียนรู้ตามลำดับ โดยการเรียนรู้และประสิทธิภาพของสมาชิกของกลไกแบบคณะกรรมการจะขึ้นอยู่กับสมาชิกตัวที่อยู่ลำดับก่อนหน้า โดยสมาชิกในลำดับถัดมาจะทำการวิเคราะห์ข้อมูลที่วิเคราะห์ผิดพลาดโดยสมาชิกตัวก่อนหน้านี้ Boosting จะสามารถลดการกระจายและการลำเอียงของการพยากรณ์ข้อมูลได้

Boosting ถูกคิดค้นโดย Freund ในปี ค.ศ.1996 ซึ่งเป็นวิธีการที่ช่วยลดข้อผิดพลาดในการวิเคราะห์ข้อมูลของอัลกอริทึมที่มีความสามารถต่ำ โดยวิธีการคือ ทำการวิเคราะห์ข้อมูลโดย

อัลกอริทึมใดๆ หลายๆครั้งและรวมผลนั้นให้เป็นค่าเดียว [11] หลักการทำงานของ Boosting เป็นดังนี้

Input: ข้อมูลทั้งหมด M ชุด $\{(x_i, y_i),\ldots,\{x_m, y_m)\}$

กำหนดให้ WeakLearn คือ อัลกอริทึมใดๆที่เลือกใช้

ค่า T กำหนดจำนวนรอบในการเรียนรู้

ให้ $B=\{(i,y) : i$ เป็นสมาชิกของ $\{1,..,M\}, y \neq y_i\}$

ค่าเริ่มต้นกำหนดให้ $D_1(i,y) = 1 / |B|$ สำหรับ (i , y) เป็นสมาชิกของ B

สำหรับค่า $t = 1,2,\ldots,T$

1.เรียกใช้ WeakLearn ส่งค่าให้อัลกอริทึม ซึ่งเป็นข้อมูลที่วิเคราะห์ผิดพลาด D_t

2.ตรวจสอบสมมติฐาน h_t X x Y → [0,1]

3.คำนวณค่า pseudo-loss ของ h_t

$$\epsilon_t = \frac{1}{2} \sum_{(i,y)\in B} D_t(i,y)(1 - h_t(x_i,y_i) + h_t(x_i,y)) \quad (2)$$

4.กำหนดค่า

$$\beta_t = \epsilon_t / (1 - \epsilon_t) \quad (3)$$

5.ปรับปรุงค่า D_t :

$$D_{t+1}(i,y) = \frac{D_t(i,y)}{Z_t} \cdot \beta_t^{(1/2)(1+h_t(x_i,y_i)-h_t(x_i,y))} \quad (4)$$

Output :

$$h_{fin}(x) = \arg\max_{y\in Y} \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t}\right) h_t(x,y) \quad (5)$$

ผลลัพธ์สุดท้ายคือ การโหวตค่าน้ำหนักของ สมมติฐานที่ใช้ในการวิเคราะห์ โดยค่าน้ำหนักที่มากที่สุดจะถูกเลือกเพราะให้ค่าความผิดพลาดที่ต่ำที่สุด

2. งานวิจัยที่เกี่ยวข้อง

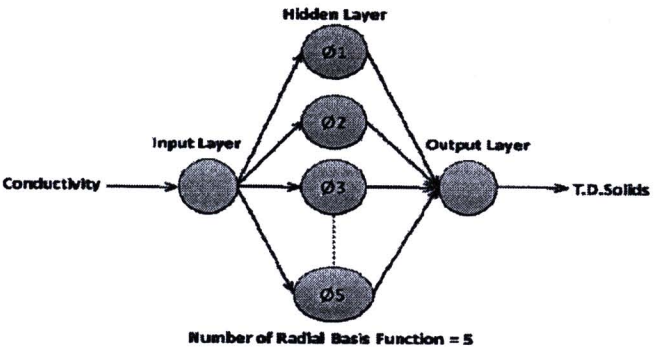
งานวิจัยที่มีการศึกษาโดยการใช้โครงข่ายประสาทเทียมในการแก้ปัญหานั้นมีอยู่มากมาย มีการประยุกต์ใช้ในกลุ่มงานต่างๆ ในด้านการวิเคราะห์ข้อมูลคุณภาพน้ำ โครงข่ายประสาทเทียมถูกประยุกต์ใช้ในการพยากรณ์ค่าพารามิเตอร์ของคุณภาพน้ำ โดยอัลกอริทึมที่ใช้นั้นมีทั้งโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน, Multilayer perceptron ซึ่งให้ผลแตกต่างกันทั้งความแม่นยำในการพยากรณ์ค่า และความเร็วในการเรียนรู้ของโครงข่าย ส่วนในด้านการเพิ่มประสิทธิภาพโดยใช้กลไกแบบคณะกรรมการนั้นได้มีการทดสอบเปรียบเทียบกับการใช้เพียงโครงข่ายประสาทเทียมหรืออัลกอริทึมเพียงอย่างเดียวเพื่อให้เห็นประสิทธิภาพที่เพิ่มขึ้นหรือลดลงอย่างชัดเจนและในบางงานวิจัย เทคนิคอื่นๆยังถูกใช้ร่วมด้วยเพื่อให้ได้ผลการวิเคราะห์ข้อมูลที่แม่นยำขึ้น เช่น สมการถดถอย (Regression)

2.1 On investigation of radial basis function neural network for prediction of water quality parameters

งานวิจัยนี้ผู้วิจัย [12] ใช้โครงข่ายประสาทเทียม ในการพยากรณ์ค่าพารามิเตอร์ของคุณภาพน้ำ โดยเลือกใช้ อัลกอริทึม Radial basis function Neural network(RBF-NN) ข้อมูลที่ใช้ศึกษานั้นได้จากการวัดค่าที่แม่น้ำ Johor ประเทศมาเลเซีย มีทั้งหมด 30 ค่าตัวอย่าง ตัวอย่างที่ถูกวัดดำเนินการในช่วงปี ค.ศ.1998-2002

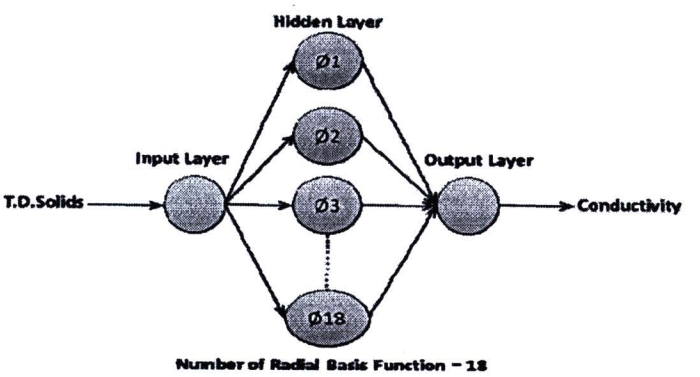
เพื่อวิเคราะห์ความสัมพันธ์เชิงปริมาณของพารามิเตอร์คุณภาพน้ำผู้วิจัยได้เลือกใช้ สมการถดถอย (regression model) จากการวิเคราะห์ค่าพารามิเตอร์คุณภาพน้ำผู้วิจัยพบว่า ความนำไฟฟ้าของน้ำ มีความสัมพันธ์กับความเข้มข้นของสารละลายของแข็งในน้ำ ซึ่งส่วนมากคือ เกลือ และพบว่าการเพิ่มขึ้นของสารแขวนลอยในน้ำส่งผลกระทบโดยตรงกับ ความขุ่นของน้ำ และได้สร้างแบบจำลองการทดลองเพื่อพยากรณ์ค่าดังกล่าว

แบบจำลองสำหรับพยากรณ์ค่าความเข้มข้นของสารละลายของแข็งในน้ำประกอบด้วยนิวรอลจำนวน 5 ตัว และข้อมูล Input คือ ค่าความนำไฟฟ้า



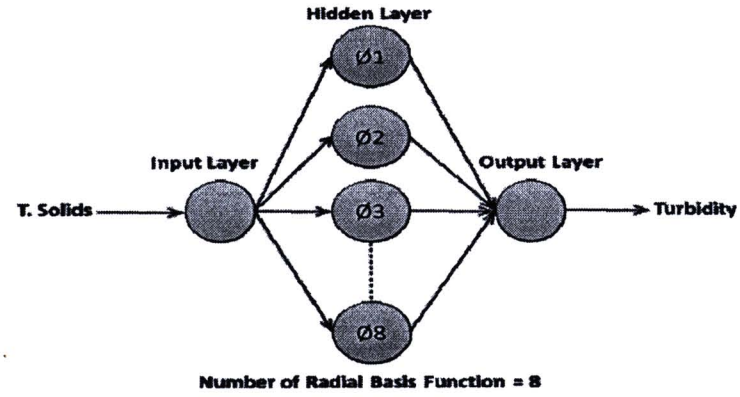
ภาพที่ 5 แบบจำลอง RBF-NN เพื่อพยากรณ์ค่าความเข้มข้นของสารละลายของแข็งในน้ำ [12]

แบบจำลองสำหรับพยากรณ์ค่าความนำไฟฟ้าประกอบด้วยนิวรอลจำนวน 18 ตัว และข้อมูล Input คือ ความเข้มข้นของสารละลายของแข็งในน้ำ



ภาพที่ 6 แบบจำลอง RBF-NN เพื่อพยากรณ์ค่าความนำไฟฟ้า [12]

แบบจำลองสำหรับพยากรณ์ค่าความขุ่นประกอบด้วยนิวรอลจำนวน 8 ตัว และข้อมูล Input คือ ความเข้มข้นของสารละลายของแข็งในน้ำ



ภาพที่ 7 แบบจำลอง RBF-NN เพื่อพยากรณ์ค่าความขุ่น [12]

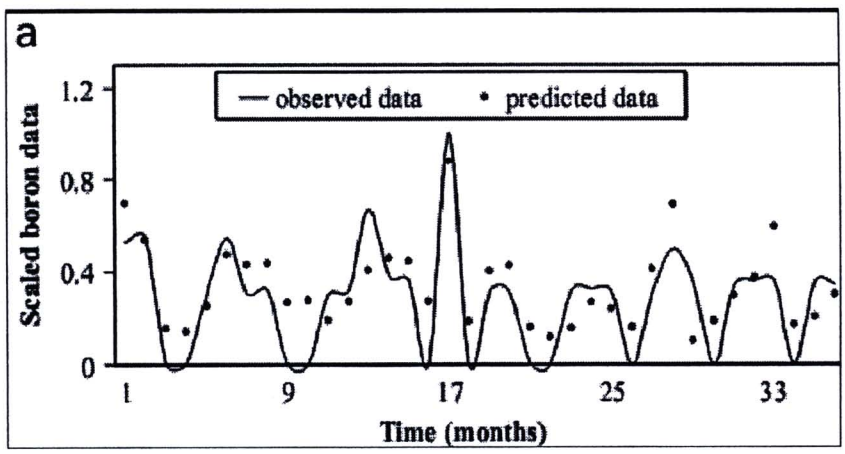
ผลการทดลองพบว่าการพยากรณ์โดยใช้ RBF-NN ให้ผลที่แม่นยำมากว่าการพยากรณ์โดยใช้สมการถดถอย โดยเฉพาะในสถานการณ์ที่ข้อมูลที่ใช้วิเคราะห์มีจำนวนน้อยและไม่จำเป็นต้องรู้ลำดับความสำคัญของข้อมูลและสมการถดถอยสามารถใช้ศึกษาความสัมพันธ์ของตัวแปรที่ศึกษาได้ โครงข่ายประสาทเทียมที่มีจำนวนนิวรอลต่างกันให้ผลการพยากรณ์ที่ต่างกัน ดังนั้นในงานวิจัยได้แนวความคิดว่าการสร้างกลไกแบบคณะกรรมการ ในแต่ละสมาชิกของโครงข่ายที่เป็น โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน สามารถกำหนดจำนวนนิวรอลที่ต่างกันเพื่อให้ได้ผลการพยากรณ์ที่ต่างกันได้

2.2 A hybrid neural network and ARIMA model for water quality time series prediction

งานวิจัยนี้ [2] มีเป้าหมายเพื่อพัฒนาแบบจำลองผสม ซึ่งประกอบด้วยโครงข่ายประสาทเทียมและAuto regressive integrated moving average (ARIMA) เพื่อทำการพยากรณ์ค่าคุณภาพน้ำและเพื่อเปรียบเทียบค่าที่ได้กับค่าที่วัดไว้แล้วและเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองผสมกับโครงข่ายประสาทเทียมและARIMA โดยชุดข้อมูลที่ใช้ในการทดลอง คือค่าพารามิเตอร์คุณภาพน้ำที่วัดในปี ค.ศ.1996-2004 ซึ่งประกอบด้วย ค่าออกซิเจนละลาย(DO),อุณหภูมิน้ำ และโบรอน (Br)

อัลกอริทึมที่ใช้ในโครงข่ายประสาทเทียมคือ Back propagation ซึ่งเป็นแบบป้อนกลับ และ ARIMA โดยวิธีการทั้งสองได้แยกกันทดลองเพื่อวัดประสิทธิภาพของแบบจำลอง ชุดข้อมูลที่ใช้ในการวิจัยมีการบันทึกค่าทั้งหมด 9 ปี หรือ 108 เดือน ซึ่งทำการแบ่งข้อมูลสำหรับเรียนรู้โดยแบบจำลองจำนวน 72 เดือนและแบ่งเพื่อตรวจสอบความแม่นยำของแบบจำลองอีก 36 เดือน ในการวัดประสิทธิภาพของแบบจำลองในงานวิจัยนี้ได้ทำการวัดค่า root mean square error

(RMSE)และmean absolute percentage error (MAPE) ซึ่งเป็นการวิเคราะห์ค่าผลลัพธ์ที่ได้จากการพยากรณ์โดยแบบจำลอง เปรียบเทียบกับค่าที่วัดจริง

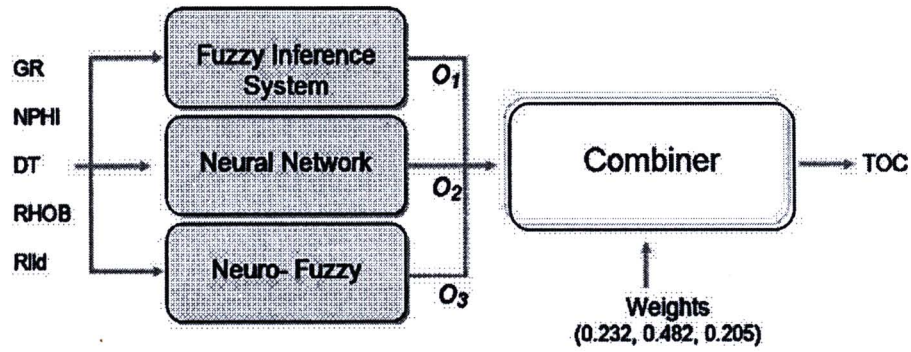


ภาพที่ 8 ผลการพยากรณ์ค่าโบรอนโดยแบบจำลอง ARIMA [2]

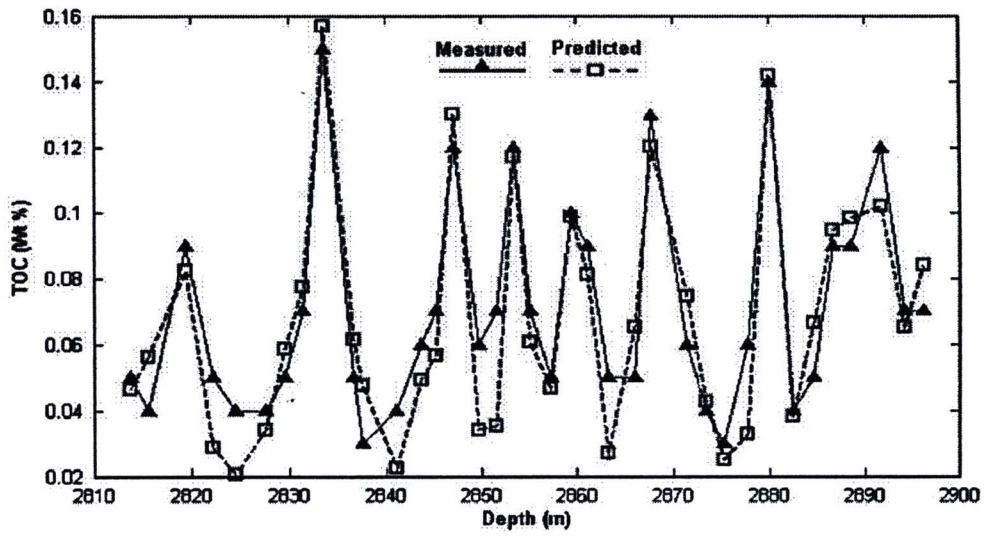
จากผลการทดลองพบว่า แบบจำลองผสมให้ผลลัพธ์ที่ไม่ดีนักในการพยากรณ์ค่าโบรอนและค่าออกซิเจนละลายแต่ค่าอุณหภูมิน้ำให้ผลที่ดี และเมื่อเปรียบเทียบกับผลการทดลองที่ใช้เพียงโครงข่ายประสาทเทียมหรือ ARIMA เพียงอย่างเดียวพบว่าผลการทดลองจากแบบจำลองผสมให้ผลการทดสองที่ดีกว่า จากงานวิจัยนี้พบว่าการวิเคราะห์ข้อมูลคุณภาพน้ำรูปแบบที่น่าสนใจอย่างหนึ่งคือ การแบ่งข้อมูลส่วนหนึ่งเพื่อใช้ในการเรียนรู้ของโครงข่ายประสาทเทียมและอีกส่วนหนึ่งใช้ในการตรวจสอบความถูกต้องของแบบจำลองการพยากรณ์ข้อมูล

2.3 A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: An example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran

งานวิจัยนี้ [13] ได้ใช้กลไกแบบคณะกรรมการในการพยากรณ์ค่าสารประกอบอินทรีย์ (TOC) โดยกลไกแบบคณะกรรมการจะทำการรวมผลการพยากรณ์ของสมาชิกทุกตัว ซึ่งสมาชิกเหล่านั้นประกอบด้วย ตรรกศาสตร์คลุมเครือ (fuzzy logic),Neuro-fuzzy และโครงข่ายประสาทเทียม โดยค่ารวมของค่าน้ำหนักจะใช้ Genetic algorithm ในการวิเคราะห์ ชุดข้อมูลที่ใช้มีทั้งหมด 137 ค่าแบ่งเป็นข้อมูลสำหรับเรียนรู้ 87 ค่า และใช้ในการทดสอบอีก 37 ค่า



ภาพที่ 9 แผนภาพของระบบกลไกแบบคณะกรรมการที่ใช้ในการทดลอง [13]



ภาพที่ 10 การเปรียบเทียบระหว่างค่าที่วัดและค่าที่ได้จากการพยากรณ์โดยกลไกแบบคณะกรรมการ [13]

ผลทดลองพบว่ากลไกแบบคณะกรรมการสามารถหาความสัมพันธ์เชิงปริมาณของพารามิเตอร์ได้ดี แต่ก็พบปัญหาในการทดลองบ้างเนื่องจากข้อมูลยังมีปริมาณน้อยไม่เพียงพอให้แบบจำลองทั้งสามชนิดประมวลผลและผลการวิเคราะห์ข้อมูลที่ได้พบว่าความแม่นยำเพิ่มมากขึ้น

2.4 Combining MLP and RBF Neural Networks for Novelty Detection in Short Time Series

งานวิจัยนี้ [14] นำเสนอการตรวจจับสิ่งแปลกปลอมในข้อมูล ซึ่งใช้ในหลากหลายงาน เช่น การผลิต ระบบคอมพิวเตอร์ การทำงานของเครื่องจักร ซึ่งส่วนมากแล้วมีข้อมูลแบบอนุกรมเวลา

ผู้วิจัยทำการพยากรณ์ข้อมูลโดยใช้โครงข่ายประสาทเทียมแบบ MLP และโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน ซึ่งมีการรวมผลลัพธ์ของทั้งสองโครงข่าย จากการทดลองพบว่าโครงข่ายแบบคณะการการของ RBF และ MLP มีประสิทธิภาพที่ดีให้ผลความผิดพลาดที่ต่ำกว่า การใช้โครงข่ายแบบ RBF หรือ MLP เพียงอย่างเดียว

เมื่อดูภาพรวมของงานวิจัยที่เกี่ยวข้องจะเห็นว่ามีการวิจัยที่ใช้โครงข่ายประสาทเทียมในด้านการพยากรณ์ข้อมูลค่อนข้างหลากหลายและให้ผลการวิจัยที่แม่นยำ โดยเฉพาะในการศึกษาคุณภาพน้ำนั้น การใช้โครงข่ายประสาทเทียมได้เลือกใช้อัลกอริทึมอื่นๆนอกเหนือจาก RBF [12] เช่น Back propagation [2] ซึ่งแสดงให้เห็นถึงความแพร่หลายและน่าเชื่อถือในความสามารถของโครงข่ายประสาทเทียม และการใช้กลไกแบบคณะกรรมการก็แสดงให้เห็นว่าสามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ดี [13] จึงสรุปได้ว่าในการวิจัยเพื่อศึกษาความสัมพันธ์และพยากรณ์พารามิเตอร์คุณภาพน้ำ โดยใช้กลไกแบบคณะกรรมการ สามารถดำเนินการศึกษาได้และคาดว่าจะได้รับผลการวิจัยที่น่าพอใจ และจากการที่ยังไม่มีการใช้กลไกแบบคณะกรรมการในการศึกษาเกี่ยวกับคุณภาพน้ำในประเทศไทยจึงเป็นเรื่องที่น่าสนใจอย่างยิ่งที่ จะศึกษาเกี่ยวกับหัวข้องานวิจัยนี้