การพัฒนาแบบจำลองเพื่อพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมาด้วยเทคนิคเหมืองข้อมูล

Development of the Average of Water Stream Flow Prediction Models for Nakhonratsima Province using Data mining Techniques

มังกร เจนชัย[1*], จารี ทองคำ[2]

Mungkorn Janechai[1*], JareeThongkam[2]

Received: 22 August 2016 ; Accepted: 23 December 2016

บทคัดย่อ

น้ำเป็นทรัพยากรธรรมชาติที่มีค่า และมีความสำคัญมากกับสิ่งมีชีวิตไม่ว่าจะเป็นพืชหรือสัตว์ แต่ละปีน้ำมีผลกระทบต่อชาวนาเป็นอย่างมากโดยเฉพาะจังหวัดนครราชสีมาซึ่งเป็นจังหวัดที่มีแม่น้ำหลายสาย มีการทำเกษตรกรรมจำนวนมาก การทราบถึงปริมาณน้ำได้ล่วงหน้าสามารถเป็นแนวทาง ช่วยในการบริหารจัดการปริมาณน้ำให้เพียงพอต่อการทำเกษตรกรรมในจังหวัดนครราชสีมา ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาและพัฒนาแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมาในช่วงฤดูฝนของเดือน กรกฎาคม ถึง ตุลาคม ของปี ค.ศ. 2012 ถึง 2015 ของจังหวัดนครราชสีมา โดยใช้ 4 เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลองคือ Artificial Neural Networks (ANN), linear regression (LR), Sequential Minimal Optimization for Regression (SMOreg) และ Radial basis function networks (RBFN)มาใช้ คณะผู้วิจัยได้ใช้หลักการ Sliding Window ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ เพื่อวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่า mean absolute error (MAE) และ root mean square error (RMSE) จากการศึกษาพบว่า แบบจำลอง SMOreg มีประสิทธิภาพในการพยากรณ์ปริมาณน้ำท่าเฉลี่ยในช่วงฤดูฝนของจังหวัดนครราชสีมามากที่สุด

คำสำคัญ น้ำท่าเฉลี่ย อนุกรมเวลา เทคนิคเหมืองข้อมูล

Abstract

Water is a valuable natural resource and critical to all living things. Each year, water affects famers especially in Nakhon Ratchasima province. Nakhon Ratchasima has many rivers used for planting agriculture. To know the amount of water in the rivers (called stream flow) in advance can assist Aquatic Toxicology officers manage supply and the amount of water stream flow. The objective of this research is to study and develop a stream flow prediction model for Nakhon Ratchasima during the rainy season between July and October 2012 – 2015. Four techniques including Artificial Neural Networks (ANN), linear regression (LR), Sequential Minimal Optimization for Regression (SMOreg) and Radial basis function networks (RBFN) were utilized to build the models. In order to evaluate the techniques, Sliding Window was employed to divide information into training and testing sets. The training set was used to build the modes while the testing set was used to evaluate the effectiveness of the models. The average mean absolute error (MAE) and root mean square error (RMSE) is the criteria to select the best technique. The experimental result showed that SMOreg model is superior to ANN LR and RBFN.

Keywords: Water Stream Flow, Time series, Model of Data Mining

---

[1] นิสิต, สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม 44150.

[2] อาจารย์ที่ปรึกษา, สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม 44150

[1] Student Faculty of Informatics, Mahasarakham University, MahaSarakham, Thailand, 44150.

[2] Advisors Faculty of Informatics, Mahasarakham University, MahaSarakham, Thailand, 44150.

** Corresponding author: Tel: +66 088 1010036 Email address: kwangaew72@gmail.com*

บทนำ

น้ำเป็นทรัพยกรธรณีรมชาติที่มีค่า และมีความสำคัญมากกับสิ่งมีชีวิตไม่ว่าจะเป็นพืช หรือ สัตว์ จังหวัดนครราชสีมาเป็นจังหวัดที่มีพื้นที่ใหญ่ เป็นอันดับสองของประเทศไทย เนื่องจากจังหวัดนครราชสีมาเป็นจังหวัดที่ขนาดใหญ่ ทำให้ปริมาณน้ำส่งผลกระทบต่อคนในจังหวัดนครราชสีมา มากไม่ว่าจะเป็นน้ำท่วม หรือน้ำแล้ง เช่นในช่วงน้ำท่วม ส่งผลกระทบทางด้านน้ำท่วมบนถนนทำให้รถติด พืชผลทางการเกษตรได้รับน้ำมากเกินไปทำให้เกิดความเสียหาย สัตว์ที่เลียงเป็นโรคที่เกิดจากน้ำ ในช่วงที่มีน้ำในจังหวัดมีปริมาณที่น้อย ก็ส่งผลกระทบมากเช่นกัน ตัวอย่างน้ำในจังหวัดไม่พอใช้ ทำให้สัตว์ที่เลี้ยงไว้ขาดน้ำ พืชที่ปลูกขาดน้ำ ทำให้จังหวัดนครราชสีมามีผลกระทบมากในเรื่องปริมาณน้ำเนื่องจากจังหวัดนครราชสีมาเป็นจังหวัดขนาดใหญ่ มีพื้นที่มากถึง 20,493.964 ตารางกิโลเมตรหรือประมาณ 12,808,728 ไร่คิดเป็นร้อยละ 12.12 ของพื้นที่ภาคตะวันออกเฉียงเหนือ มีอำเภอทั้งหมด 32 อำเภอ มีแหล่งน้ำธรรมชาติที่สำคัญของจังหวัดนครราชสีมา 9 ลุ่มน้ำโดยมีพื้นที่ลุ่มน้ำประมาณ 20,905 ตารางกิโลเมตรอำเภอที่ได้รับน้ำจากลุ่มน้ำได้แก่อำเภอชุมพวง พิมาย ห้วยแถลง จักราช โนนสูง เมือง เสิงสาง ครบุรี โชคชัย ปักธงชัย สูงเนิน ปากช่อง ขามทะเลสอ สีคิ้ว โนนไทย ประทาย โนนแดง บัวใหญ่ ขามสะแกแสง คง แก้งสนามนาง บ้านเหลื่อม หนองบุญมาก รวม 23 อำเภอ มีแหล่งเก็บน้ำขนาดใหญ่ 5 แห่ง อย่างไรก็ตามผลกระทบจากน้ำก็ยังมีมาโดยตลอดในช่วงสามปีที่ผ่านมา

ในปี พ.ศ. 2556 จังหวัดนครราชสีมา[1] ประสบภัยแล้งรุนแรง จำนวน 26 อำเภอ 191 ตำบล 1,905 หมู่บ้าน ราษฎรประสบภัย 100,197 ครัวเรือน พื้นที่การเกษตรเสียหาย 863,433 ไร่ โดยภาพรวมอำเภอที่มีตำบลตั้งแต่ 10 ตำบลขึ้นไปประสบภัยแล้งรุนแรง เช่น อ.ประทาย 13 ตำบล 147 หมู่บ้าน, อ.บัวใหญ่ 10 ตำบล 146 หมู่บ้าน, อ.คง 10 ตำบล 156 หมู่บ้าน, อ.ปักธงชัย 16 ตำบล 148 หมู่บ้าน, อ.ด่านขุนทด 14 ตำบล 108 หมู่บ้าน, อ.เมือง 12 ตำบล 66 หมู่บ้าน, อ.สีคิ้ว 11 ตำบล 65 หมู่บ้าน, อ.โนนไทย 10 ตำบล 131 หมู่บ้าน เป็นต้น เป็นพื้นที่ประสบภัยแล้งรุนแรงและอยู่ในโซนที่หนักที่สุด ส่วนอีก 6 อำเภอยังไม่มีการประกาศแต่อย่างใด เนื่องจากอยู่ระหว่างการสำรวจ ประกอบด้วย อ.ปากช่อง, พิมาย, ครบุรี, ขามทะเลสอ, ขามสะแกแสงและ เสิงสาง

ปี 2557 นครราชสีมาประกาศพื้นที่ภัยแล้ง 11 อำเภอ[2] จังหวัดนครราชสีมาประกาศภัยแล้งใน 11 อำเภอ จากทั้งหมด 32 อำเภอ 81 ตำบล 1 เทศบาล 884 หมู่บ้าน 16 ชุมชนราษฎรได้รับความเดือนร้อน 61,603 ครัวเรือนความเสียหายด้านการเกษตร 936,679 ไร่ ประกอบด้วยอำเภอบัวใหญ่ สูงเนิน บัวลาย โนนแดง บ้านเหลื่อม ขามสะแกงแสง สีดา ห้วยแถลง โนนสูงประทาย และอำเภอด่านขุนทด รวม 11 อำเภอซึ่งอยู่ในเขตลุ่มน้ำลำสะแทด เนื้อที่ 2,457 ตร.กม. ข้อมูลน้ำท่าเฉลี่ยมีการเก็บในรูปแบบของอนุกรมเวลา โดย ชนิดของข้อมูลที่ได้จากการเก็บค่าจุดข้อมูลอย่างต่อเนื่องเป็นลำดับของเวลาก่อนหลัง ในการวิเคราะห์ข้อมูลอนุกรมเวลาได้มีนักวิจัยหลายท่านได้นำเอาเทคนิคในเหมืองข้อมูลมาใช้ในการวิเคราะห์เช่น

Soomlek, Kaewchainam, Simano และ So-In[3] ได้นำเอาเทคนิค back propagation neural networks มาทำการประยุกต์ใช้ในการการพยากรณ์ระดับน้ำ ที่สถานีวัด C.35 ตั้งอยู่ที่แม่น้ำเจ้าพระยา พระนครศรีอยุธยา ประเทศไทย โดยมีช่วงการพยากรณ์ระดับน้ำ 1-3 วัน ล่วงหน้าในแม่น้ำเจ้าพระยา ข้อมูลที่ใช้ในการวิเคราะห์เก็บรวบรวมใน ปี พ.ศ. 2551-2553 เป็นข้อมูลแบบอนุกรมเวลา จากสถานีวัดระดับน้ำ C.2 , C.13, C.35, C.36 และ C.37 เพื่อใช้ในการวิเคราะห์และพัฒนาแบบจำลองเพื่อพยากรณ์ปริมาณระดับน้ำในแม่น้ำเพื่อเตือนภัยน้ำท่วม และทดสอบประสิทธิภาพของแบบจำลองโดย Mean Squared Error (MSE) และผลการการคาดการณ์ระดับน้ำความถูกต้องของแบบจำลองที่ได้ถึง 90.1218 %

Mustafa และ Isa[4] ประยุกต์ใช้ Radial Basis Function Neural Network เพื่อสร้าง Model ในการพยากรณ์ตะกอนแขวนลอย เพื่อระบายตะกอนแขวนลอยออกจากแม่น้ำ Pari รัฐเประ ประเทศมาเลเซียเป็นข้อมูลรายวันย้อนหลัง 5 ปี จัดเก็บอยู่ในรูปแบบข้อมูลอนุกรมเวลาของตะกอนแขวนลอยออกจากแม่น้ำ Pari รัฐเประ ประเทศมาเลเซีย ซึ่งข้อมูลได้จากกรมชลประทาน กรุงกัวลาลัมเปอร์ประเทศมาเลเซีย โดยแบ่งข้อมูล 70:30 ในการทดสอบ 6 มกราคม ค.ศ. 1993 ถึง 3 มีนาคม ค.ศ. 1997 ประมาณ 70 % เป็นข้อมูล training และข้อมูล 4 มีนาคม ค.ศ. 1997 ถึงเดือน ตุลาคม 24 ค.ศ.1998) สำหรับการ testing ในการทดสอบประสิทธิภาพการทำงานของ Model ใช้ root mean square error (RMSE) and coefficient of efficiency (CE) เป็นเครื่องในการทดสอบประสิทธิภาพการทำงานของของ Model RMSE Training 24 Testing 31, CE Training 0.99 Testing 0.99, R2 Training 0.99 Testing 0.99 ผลที่ได้ Model RBF เหมาะสำหรับในการพยากรณ์ตะกอนแขวนลอย เพื่อระบายตะกอนแขวนลอยออกจากแม่น้ำ Pari รัฐเประ ประเทศมาเลเซีย

Wanatada และ Sittidech[5] ได้พัฒนาแบบจำลองที่เหมาะสมกับ การพยากรณ์ระดับน้ำท่าในเขตเทศบาลเมืองชัยภูมิได้ล่วงหน้า โดยวิเคราะห์จากข้อมูลที่ได้จากสถานีวัดน้ำท่าและน้ำฝน ที่ส่งผลกระทบให้เกิดน้ำท่วมในเขตเทศบาลเมืองชัยภูมิ ใช้ข้อมูล เป็นรายวันระหว่างปี พ.ศ. 2550 – พ.ศ

2554 จำนวนทั้งสิ้น 1,824 ระเบียน โดยคัดเลือกสถานี คัดเลือกตัวแปร รวมทั้ง ระยะเวลาเดินทางของน้ำ สร้างตัวแบบโดยใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบแพร่กลับ และ ประเมินผลประสิทธิภาพของตัวแบบด้วยค่าเฉลี่ยของความ คลาดเคลื่อนสมบูรณ์ ผลการทดลองได้ตัวแบบจำลองที่ใช้ ข้อมูลนำเข้า 15 ตัวแปรได้ค่าเฉลี่ยของความคลาดเคลื่อน สมบูรณ์เป็น 1.008

Choubey, Mishra and Pandey[6] ได้ศึกษาแบบจำลอง SMOreg ซึ่งเป็นเทคนิคหนึ่งในตระกูลเทคนิคเครื่องเวกเตอร์การสนับสนุน (Support Vector Machine: SVM) เพื่อพยากรณ์ปริมาณน้ำที่ไหลบ่า และปริมาณน้ำฝน ซึ่งเป็นพื้นที่ลุ่มน้ำ 39,372 ตารางกิโลเมตร โดยใช้หลักการ Sliding Window ช่วงละสี่เดือนของแม่น้ำ Narmada จังหวัด Madhya Pradesh ประเทศอินเดีย โดยมีข้อมูลปริมาณน้ำฝน อุณหภูมิ ตะกอน ปริมาณน้ำฝน ระดับน้ำจากการปล่อยน้ำ เป็นค่าพารามิเตอร์ เป็นตัวแปรเพื่อใช้ในการจำแนก (Classification) ข้อมูลที่นำมาใช้จากปี ค.ศ. 1975 – 2010 ด้วยเทคนิค SMOreg ในการทดลองคณะวิจัยได้แบ่งข้อมูลออกเป็นสองชุด คือ ชุดสอน (Training) และชุดทดสอบ (Testing) โดย ชุดสอนเป็นข้อมูลจากปีค.ศ.1975 - 2000 ส่วนชุดทดสอบใช้ข้อมูลจากปี ค.ศ. 2001 – 2010 ในการวัดประสิทธิภาพการทำงานของแบบจำลองเขาได้ใช้ Mean Absolute error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) และ Root Relative Square Error (RRSE) ผลการทดสอบพบว่า แบบจำลอง SMOreg มีค่า RMSE เท่ากับ 2.3731 ค่า RAE เท่ากับ 65.28% และค่า RRSE เท่ากับ 62.491% ซึ่งเป็นค่าที่เปรียบเทียบกับค่าจริง จะเห็นว่าความแตกต่างระหว่างค่าที่พยากรณ์และค่าจริงมีความแตกต่างกันมากโดยมีนัยสำคัญ

วัตถุประสงค์เพื่อศึกษาและพัฒนาแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมาในช่วงหน้าฝน ช่วงเดือน กรกฎาคม ถึง ตุลาคม ของปี ค.ศ. 2012 ถึง 2015 ของจังหวัดนครราชสีมา เพื่อช่วยในการพยากรณ์น้ำท่าเฉลี่ยในการสนับสนุนการบริหารจัดการน้ำในจังหวัดนครราชสีมา โดยนำข้อมูลจากสถานีวัดปริมานน้ำท่าเฉลี่ยในเขตพื้นที่จังหวัดนครราชสีมา ในงานวิจัยนี้ได้โดยใช้ 4 เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลอง คือ Artificial Neural Networks (ANN), Linear regression (LR), Sequential Minimal Optimization for Regression (SMOreg) และ Radial basis function networks (RBFN) คณะผู้วิจัยได้ใช้หลักการ Sliding Window ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่า mean absolute error (MAE) และ root mean square error (RMSE)

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. พื้นที่ศึกษา

จังหวัดนครราชสีมา[7] ตั้งอยู่ในภาคตะวันออกเฉียงเหนือบนที่ราบสูงโคราช ระหว่างเส้นรุ้ง ที่ 14 องศา 17 ลิปดาเหนือกับ 15 องศา 46 ลิปดาเหนือ เส้นแวงที่ 101 องศา 11 ลิปดาตะวันออก กับ 102 องศา 53 ลิปดาตะวันออก อยู่ห่างจากกรุงเทพฯโดยทางรถยนต์ 255 กิโลเมตร ทางรถไฟประมาณ 264 กิโลเมตร มีพื้นที่รวมทั้งสิ้น 20,493.964 ตารางกิโลเมตร หรือ 12,808,728 ไร่ที่ตั้งกลุ่มจังหวัดนครชัยบุรินทร์ แหล่งน้ำธรรมชาติของจังหวัดนครราชสีมา มีแหล่งน้ำธรรมชาติที่สำคัญ 9 ลุ่มน้ำ ได้แก่ ลุ่มน้ำมูลตอนบน, ลุ่มน้ำลำพระเพลิง, ลุ่มน้ำลำมูลตอนล่าง, ลุ่มน้ำลำปลายมาศ, ลุ่มน้ำลำตะคอง, ลุ่มน้ำลำเชียงไกร, ลุ่มน้ำลำสะแทด, ลุ่มน้ำลำชี และ ลุ่มน้ำลำจักราช รวมพื้นที่ลุ่มน้ำ 20,905 ตร.กม. ปริมาณน้ำฝนเฉลี่ย 1,078.6 มม./ปี ปริมาณน้ำท่าเฉลี่ย 3,512 ล้าน ลบ.ม. / ปี มีพื้นที่ชลประทาน 702,458 ไร่ คิดเป็นร้อยละ 8.57 ของพื้นที่จังหวัดทั้งหมดและ มีสถานีวัดน้ำท่าในจังหวัดนครราชสีมาจำนวน 16 สถานีดัง (Figure 1) และ (Figure 2)

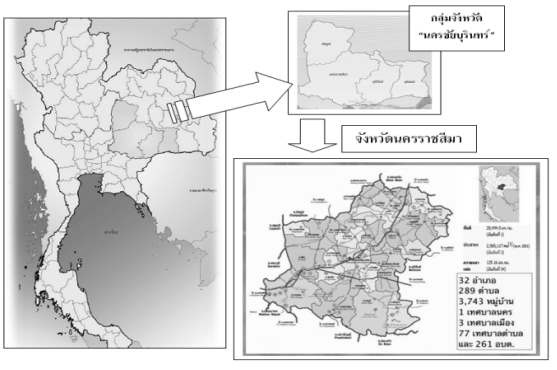

Figure 1 Nakhonratsima Province

Figure 2 Station of Average of Water Stream Flow

2. ข้อมูลอนุกรมเวลา

ข้อมูลอนุกรมเวลา[8] หมายถึง ข้อมูลที่ได้จากการเก็บค่าจุดข้อมูลอย่างต่อเนื่องเป็นลำดับของเวลาก่อนหลัง เช่น ข้อมูลตลาดหุ้น ข้อมูลอุณหภูมิรายวัน ข้อมูลคลื่นหัวใจ อัตราการเติบโตทางการตลาด ในปัจจุบันข้อมูลอนุกรมเวลาได้เข้ามามีบทบาทที่สำคัญกับงานวิจัยในศาสตร์แขนงต่าง ๆ เช่น ชีวสนเทศศาสตร์ (Bioinformatics) วิทยาการหุ่นยนต์ (Robotics) การแพทย์ (Medicine) เคมี (Chemistry) การรู้จำอากัปกิริยาท่าทาง (Gesture Recognition) การรู้จำคำพูด (Speech Recognition) การสะกดรอย (Tracking) การเงิน (Finance) ชีวมาตร (Biometrics) ดาราศาสตร์ (Astronomy) อุตสาหกรรม (Manufacturing) เป็นต้น

3. เทคนิคในเหมืองข้อมูล

เทคนิคในเหมืองข้อมูลได้ถูกนำมาใช้ในการสร้างแบบจำลองกันอย่างแพร่หลายและมีประสิทธิภาพ เช่น โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN), การถดถอยเชิงเส้น (Linear regression: LR), Sequential Minimal Optimization for Regression (SMOreg) และ Radial basis function networks (RBFN) ซึ่งสามารถอธิบายได้ดังต่อไปนี้

1. โครงข่ายประสาทเทียม (Artificial Neural Networks)หรือ (Neural Networks หรือ Neural Net) เป็นเทคนิคของการทำเหมืองข้อมูลใช้โมเดล ทางคณิตศาสตร์ในการคำนวณแบบ (Connectionist) มีการนำเอาการทำงานของโครงข่ายประสาทเทียม (Artificial Neural Networks) ซึ่งเหมือนกับการทำงานของเครือข่ายประสาทในสมองมนุษย์ ซึ่งมีความสามารถในการเรียนรู้ การจดจำ รูปแบบ (Pattern Recognition) และการสร้างความรู้ใหม่ (Knowledge Extraction) ดัง Figure 3

จากแนวคิดนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วยเซลล์ประสาท หรือ “นิวรอน” (Neurons) และ “จุดประสานประสาท” (Synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรท์” (Dendrite) ซึ่งเป็น input และปลายในการส่งกระแสประสาทเรียกว่า “แอคซอน” (Axon) ซึ่งเป็นเหมือน output ของเซลล์การทำงานของ Neural Networks คือการนำ Input เข้าไปใน Network โดยเอา Input มาคูณกับ Weight ของแต่ละขาผลที่ได้จาก Input ทุกขาของ Neuron จะเอามารวมกันต่อมา เอาค่าที่ได้มาเทียบกับ Threshold ที่กำหนดไว้ถ้าผลรวมมีค่ามากกว่า Threshold แล้ว Neuron ก็จะส่ง Output ออกไป Output นี้จะถูกส่งไปยัง Input ของ Neuron ที่เชื่อมกันใน Network ถ้าค่าน้อยกว่า Threshold ก็จะไม่เกิด Output สมการทางคณิตศาสตร์ที่ได้ดัง (Figure 4)

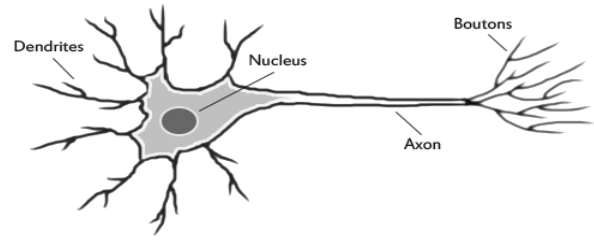

Figure 3 Basic stru cture of a brain cell

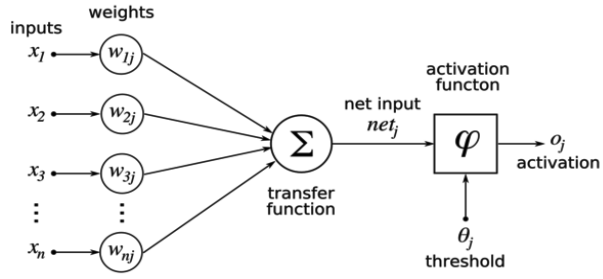

Figure 4 Common activation functions in use with neural networks

มีนักวิจัยหลายท่านได้นำเอาเทคนิคนี้มาใช้ในการพยากรณ์ เช่น Soomlek, Kaewchainam, Simano และ So-In[3] ปี พ.ศ. 2558 ทำการประยุกต์ใช้ Neural Network สำหรับการคาดการณ์ระดับน้ำ ของสถานีวัดระดับน้ำ C.35 ตั้งอยู่ที่แม่น้ำเจ้าพระยา พระนครศรีอยุธยาประเทศไทย โดยใช้ back propagation neural networks สมามรถนำมาพยากรณ์ระดับน้ำ 1-3 วันล่วงหน้าในแม่น้ำเจ้าพระยา ข้อมูลที่ใช้ในการวิเคราะห์เก็บรวบรวมใน ปี พ.ศ. 2551-2553 เป็นข้อมูลแบบอนุกรมเวลา จากสถานีวัดระดับน้ำ C.2, C.13, C.35, C.36 และ C.37 เพื่อใช้ในการวิเคราะห์และพัฒนาแบบจำลองเพื่อพยากรณ์ปริมาณระดับน้ำในแม่น้ำเพื่อเตือนภัยน้ำท่วม และทดสอบประสิทธิภาพของแบบจำลองโดย Mean Squared Error (MSE) และผลการคาดการณ์ระดับน้ำความถูกต้องของแบบจำลองที่ได้ถึง 90.1218 %

2. การถดถอยเชิงเส้น (Linear regression : LR) ถดถอยเชิงเส้นผ่านเส้น Regression การวิเคราะห์การถดถอย Regression Analysis แบ่งออกเป็น สองชนิดคือ Linear Regression, Non-linear Regression การวิเคราะห์ Linear Regression แบ่งออกเป็น Simple Linear Regression (SLR), Multiple Linear Regression (MLR) การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว (ในที่นี้คือตัวแปร X และ Y) ที่มีความสัมพันธ์กันในลักษณะเชิงเส้น (Linear) โดยมีสมการ

ถดถอย คือสมการเส้นตรง$Y = \alpha + \beta X$ ซึ่ง α และ β เป็นพารามิเตอร์ที่ไม่ทราบค่าจึงจะต้องประมาณค่าโดยใช้ข้อมูลจากตัวอย่างโดยวิธีที่นิยมใช้ในการประมาณค่าของ α และ β ก็คือวิธี กำลังสองน้อยที่สุด (The Least Squares Method) ซึ่งจะแทนค่าของ α และ β ด้วยค่า a และ b โดยที่ a ก็คือค่าคงที่ (Constant) เป็นค่าที่เส้นกราฟถดถอยตัดกับแกน Y ส่วน b เป็นความชัน (Slope) ของเส้นกราฟซึ่งแสดงถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงเรียกส่วนนี้ว่า สัมประสิทธิ์การถดถอย (Regression Coefficient) หรือสัมประสิทธิ์การพยากรณ์สมการที่ได้คือ Y = a + bX ดัง (Figure 5)

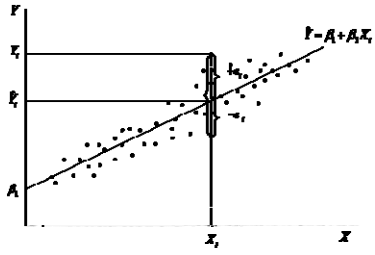

Figure 5 Regression line

GAO Xiang, BAI Lina[9] ใช้ Multivariate Linear Regression (MLR) ปรับปรุงประสิทธิภาพการทำงาน refreshed by improved RFF-LS ของ Partial Regression Coefficients (PRC) ในการประเมินค่าการไหลมารวมกันของปริมาณน้ำเสีย ให้ดีขึ้นของโรงบำบัดน้ำเสียเนื่องจาก Partial Regression Coefficients (PRC) มีรูปแบบข้อมูลและกระบวนการบางอย่างที่ไม่เหมาะสมจากการทดสอบแสดงให้เห็นประสิทธิภาพการพยากรณ์ที่ดีขึ้น ของดัชนีตัวชี้วัดของข้อมูลสองค่า Chemical Oxygen Demand (COD) ในน้ำที่มีคุณภาพดี ในการทดสอบได้แบ่งข้อมูลเป็นสองส่วน ส่วนที่ 1 มีข้อมูลจำนวน 165 Shifts จากฐานข้อมูลเดือน มกราคม เมษายน และ กรกฎาคม ในปี 2007 เป็นข้อมูล training ส่วนที่ 2 มีจำนวนข้อมูล 45 Shifts ในเดือนสิงหาคม ปี 2008 เป็นข้อมูลในการทดสอบ

3. (Radial basis function networks: RBFN) ประกอบด้วย 3 Layer คือ หนึ่ง Input Lavers สอง Hidden Layer และสาม Output Layer ในส่วน Hidden Layer จะเป็นส่วนกำหนดการทำงานและรูปแบบในการป้อนข้อมูล คำตอบของค่าที่เป็นบวก เท่ากับค่า y

คำตอบที่ลดลงเท่ากับ 0 คือ |y| “0 สมการที่ได้คือ $f(y) = e^{-y^2}$ พัฒนาได้เป็น $f'(y) = -2ye^{-y^2} = -2yf(y)$

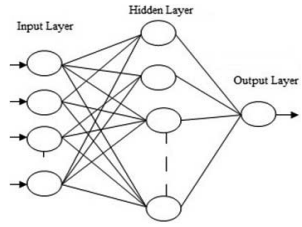

Figure 6 Model Radial Basis Function Neural Network

Mustafa และ Isa[4] ประยุคใช้ Radial Basis Function Neural Network เพื่อสร้าง Model ในการพยากรณ์ตะกอนแขวนลอย เพื่อระบายตะกอนแขวนลอยออกจากแม่น้ำ Pari รัฐเประ ประเทศมาเลเซียเป็นข้อมูลรายวันยอนหลัง 5 ปี จัดเก็บอยู่ในรูปแบบข้อมูลอนุกรมเวลาของตะกอนแขวนลอยออกจากแม่น้ำ Pari รัฐเประ ประเทศมาเลเซีย ซึ่งข้อมูลได้จากกรมชลประทาน กรุงกัวลาลัมเปอร์ ประเทศมาเลเซียโดยแบ่งข้อมูล 70:30 ในการทดสอบ 6 มกราคม ค.ศ. 1993 ถึง 3 มีนาคม ค.ศ. 1997 โดยประมาณ 70 % เป็นข้อมูล training และข้อมูล 4 มีนาคม ค.ศ. 1997 ถึง 24 ตุลาคม ค.ศ.1998) สำหรับการ testing ในการทดสอบประสิทธิภาพการทำงานของ Model ใช้ root mean square error (RMSE) and coefficient of efficiency (CE) เป็นเครื่องในการทดสอบประสิทธิภาพการทำงานของของ Model RMSE Training 24 Testing 31 , CE Training 0.99 Testing 0.99, R2 Training 0.99 Testing 0.99 ผลที่ได้ Model RBF เหมาะสำหรับในการพยากรณ์ตะกอนแขวนลอย เพื่อระบายตะกอนแขวนลอยออกจากแม่น้ำ Pari รัฐเประ ประเทศมาเลเซีย

4. Sequential Minimal Optimization for Regression (SMOreg) ถูกพัฒนาจาก SMO อัลกอลิทึม บนพื้นฐานของ SVM (Support Vector Machine) วิธีนี้ใช้การแทนค่าที่ข้อมูลสูญหายและแปลงข้อมูลคุณลักษณะเชิงกลุ่ม (nominal) ให้เป็นข้อมูลไบนารี่ (Binary)นอกจากนี้ยังทำให้ข้อมูลคุณลักษณะทุกค่าอยู่ในรูปแบบมาตรฐาน (Normalized) SMOreg ยังมีคุณสมบัติใช้งานกับ non-linear ได้อย่างมีประสิทธิภาพ SMOreg ยังช่วยในการจัดการโครงสร้างของ Model และลดความเสี่ยงของข้อมูลทำให้ข้อมูลมีความน่าเชื่อถือในการพยากรณ์

Choubey, Mishra and Pandey [6] ได้ศึกษาแบบจำลอง SMOreg ซึ่งเป็นเทคนิคหนึ่งในตระกูลเทคนิคเครื่องเวกเตอร์การสนับสนุน (Support Vector Machine: SVM) เพื่อพยากรณ์ปริมาณน้ำที่ไหลบ่า และปริมาณน้ำฝน ซึ่งเป็นพื้นที่ลุ่มน้ำ 39,372 ตารางกิโลเมตร โดยใช้หลักการ Sliding Window ช่วงละสี่เดือนของแม่น้ำ Narmada จังหวัด Madhya Pradesh ประเทศอินเดีย โดยมีข้อมูลปริมาณ

น้ำฝน อุณหภูมิ ตะกอน ปริมาณน้ำฝน ระดับน้ำจากการปล่อยน้ำ เป็นค่าพารามิเตอร์ เป็นตัวแปรเพื่อใช้ในการจำแนก (Classification) ข้อมูลที่นำมาใช้จากปี ค.ศ. 1975 – 2010 ด้วยเทคนิค SMOreg ในการทดลองคณะวิจัยได้ แบ่งข้อมูลออกเป็นสองชุด คือ ชุดสอน (Training) และชุดทดสอบ (Testing) โดย ชุดสอนเป็นข้อมูลจากปีค.ศ.1975 - 2000 ส่วนชุดทดสอบใช้ข้อมูลจากปี ค.ศ. 2001 – 2010 ในการวัดประสิทธิภาพการทำงานของแบบจำลองเขาได้ ใช้ Mean Absolute error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) และ Root Relative Square Error (RRSE) ผลการการทดสอบพบว่า แบบจำลอง SMOreg มีค่า RMSE เท่ากับ 2.3731 ค่า RAE เท่ากับ 65.28% และค่า RRSE เท่ากับ 62.491% ซึ่งเป็นค่าที่เปรียบเทียบกับค่าจริง จะเห็นว่าความแตกต่างระหว่างค่าที่พยากรณ์และค่าจริงมีความแตกต่างกันมากโดยมีนัยสำคัญ

วิธีการดำเนินการวิจัย

ในการทำวิจัยคณะผู้วิจัยได้มีแบ่งวิธีการดำเนินการวิจัยออกเป็น 4 ขั้นตอนหลัก 1) ขั้นตอนการเก็บรวบรวมข้อมูล 2) ขั้นตอนก่อนกระบวนการสร้างแบบจำลอง 3) ขั้นตอนการสร้างแบบจำลอง และ 4) ขั้นตอนการวัดประสิทธิภาพแบบจำลอง ดัง (Figure 7)

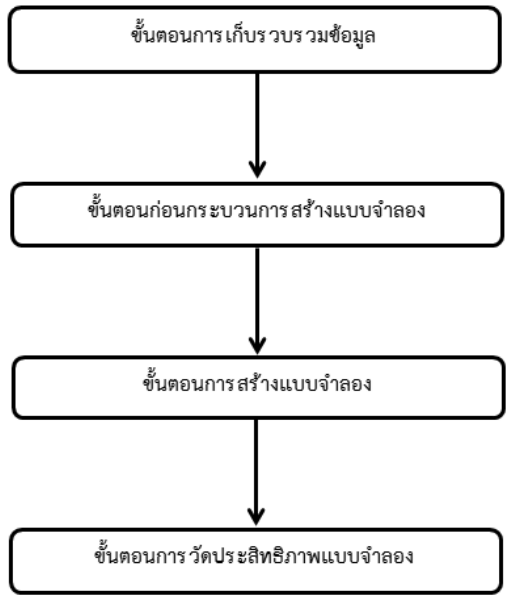

Figure 7 The process of data mining

1. ขั้นตอนการเก็บรวบรวมข้อมูล

งานวิจัยนี้ข้อมูลได้จากเว็บไซต์ http://hydro-4.com/ ของศูนย์อุทกวิทยาชลประทานภาคตะวันออกเฉียงเหนือตอนล่างสำนักบริหารจัดการน้ำและอุทกวิทยา กรมชลประทาน ซึ่งมีการจัดเก็บข้อมูลปริมาณน้ำท่าเฉลี่ยของสถานีวัดปริมาณน้ำท่าเฉลี่ย ผู้วิจัยใช้ข้อมูลปริมาณน้ำท่าเฉลี่ยรายวันจาก 16 สถานีในเขตลุ่มน้ำของจังหวัดนครราชสีมา โดยข้อมูลทั้งหมดของสถานีวัดปริมาณน้ำท่าเฉลี่ยจัดเก็บข้อมูลในรูปแบบ ระดับน้ำเทียบเฉลี่ยรายวัน 16 สถานี ในเขตลุ่มน้ำจังหวัดนครราชสีมา ผู้วิจัยได้จัดทำการแบ่งค่าข้อมูลเป็น Attribute โดยแต่ละ Attribute จะมีค่าข้อมูลของแต่ละสถานีเป็นข้อมูลสรุปเป็นรายวันในแต่ละสถานี และทำการเลือกสถานีที่มี ปริมาณน้ำท่าเฉลี่ยมากที่สุด 4 สถานี เนื่องจากสถานีทั้ง 4 มีปริมาณน้ำที่มากและส่งผลกระทบกับประชาชนมากกว่า สถานีวัดน้ำท่าที่มีปริมาณน้ำน้อยได้แก่ สถานี M38C, M145, M177 และ M184 ข้อมูลปริมาณน้ำท่าเฉลี่ยรายวัน ช่วงปี 2012 ถึง 2015 เดือนกรกฎาคม ถึง ตุลาคม ข้อมูลที่ได้จำนวน 492 ระเบียน

จาก (Figure 8) แสดงให้เห็นถึงปริมาณน้ำของสถานีวัดปริมาณน้ำท่าเฉลี่ย M.38C, M.145, M.177 และ M.184 ในจังหวัดนครราชสีมาซึ่งมีปริมาณน้ำท่าในลำน้ำลุ่มน้ำในจังหวัดนครราชสีมา ปี ค.ศ. 2012 – 2015 จากรูป สังเกตได้ว่า สถานีวัดมีปริมาณน้ำมากที่ในช่วงเดือน กันยายน – พฤศจิกายน ในช่วงหน้าฝนของจังหวัดนครราชสีมา และเป็นช่วงที่ต้องเฝ้าระวังปริมาณน้ำเพื่อไม่ให้ส่งผลกระทบแก่ประชาชน ผู้วิจัยจึงสนใจที่จะพยากรณ์ปริมาณน้ำท่าในช่วงฤดูหน้าฝนของจังหวัดนครราชสีมา

2. ขั้นตอนก่อนกระบวนการสร้างแบบจำลอง

ขั้นตอนกระบวนการก่อนการสร้างแบบจำลองโดยการแปลงข้อมูล (Data transformation) จากข้อมูลที่ได้จากเว็บไซต์ http://hydro-4.com/ จากศูนย์อุทกวิทยาชลประทานภาคตะวันออกเฉียงเหนือตอนล่างสำนักบริหารจัดการน้ำและอุทกวิทยา กรมชลประทานได้จัดเก็บข้อมูลปริมาณน้ำท่าเฉลี่ยเป็นรายสถานี ผู้วิจัยจึงได้รวบรวมข้อมูลแต่ละสถานีวัดปริมาณน้ำท่าเฉลี่ย เข้าด้วยกัน โดยแบ่งข้อมูลแต่ละสถานีตามชื่อของสถานีวัดปริมาณน้ำท่าเฉลี่ย ดัง (Table 1) จากข้อมูลที่ได้มาแต่ละสถานีวัดปริมาณน้ำท่าเฉลี่ย ผู้วิจัยได้แบ่งข้อมูลเป็นตัวแปรตามชื่อของแต่ละสถานีวัดปริมาณน้ำท่าเฉลี่ยแสดงตาม (Table 2) ข้อมูลหลังจากการแปลงข้อมูลได้เรียบร้อย

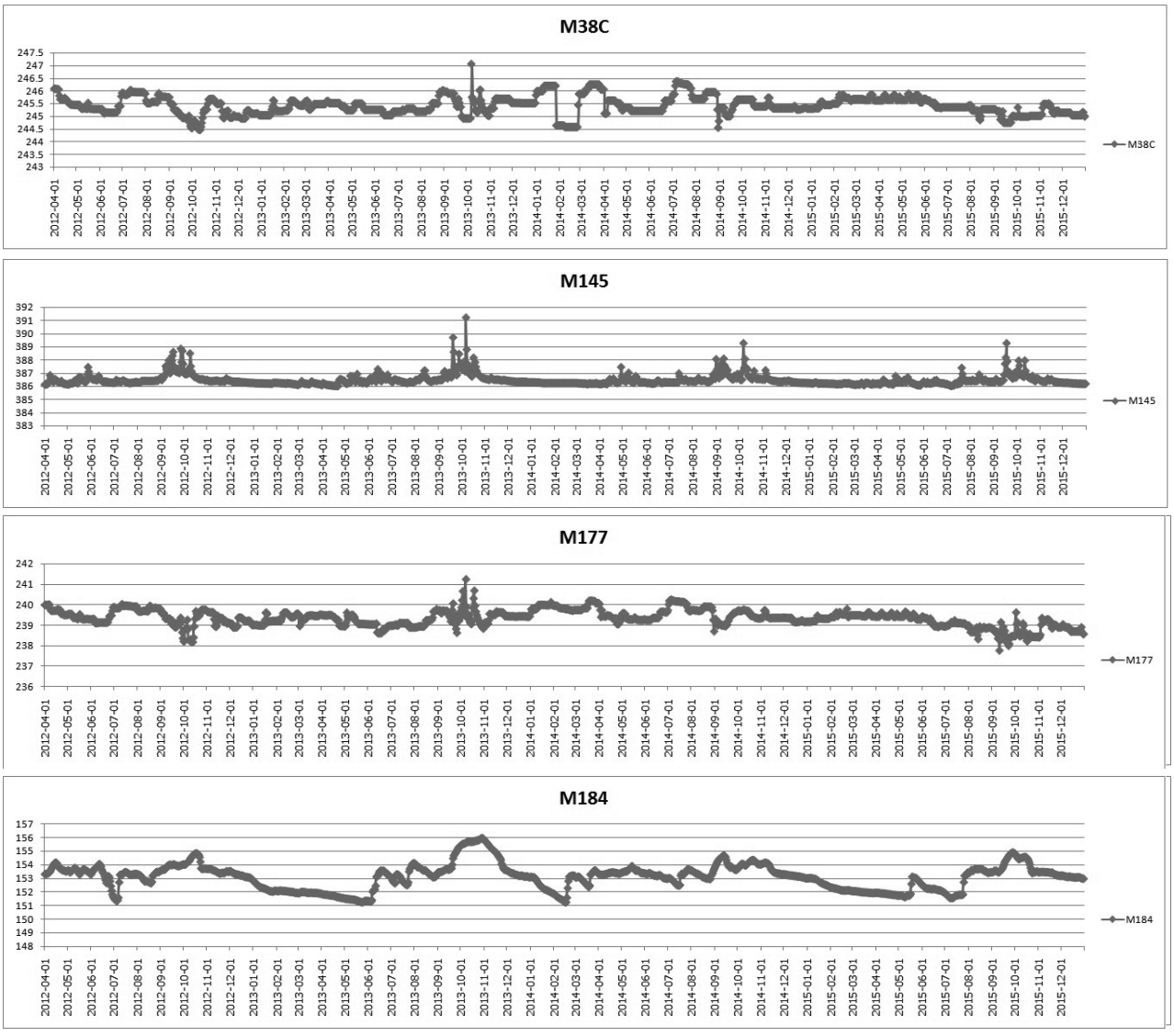

Figure 8 Level of Water Stream Flow Station M.38C, M.145, M.177 and M.184 year 2012 - 2015

Table 1 Location stations of Water Stream Flow in Nakhon Ratchasima province

ชื่อสถานี		ลุ่มน้ำ	อำเภอ	จังหวัด
สถานี	M.38C	ลำตะคอง	อ.สีคิ้ว	จ.นครราชสีมา
สถานี	M.145	ลำพระเพลิง	อ.ปากช่อง	จ.นครราชสีมา
สถานี	M.177	ลำตะคอง	อ.สีคิ้ว	จ.นครราชสีมา
สถานี	M.184	แม่น้ำมูล	อ.พิมาย	จ.นครราชสีมา

Table 2 Description of parameters and variables.

ชื่อตัวแปร	คำอธิบาย	ชนิดของตัวแปร
Date Water Level	วัน / เดือน / ปี ในการบันทึกน้ำท่าเฉลี่ย	Date
M38C	จำนวนข้อมูลปริมาณน้ำท่าเฉลี่ยของสถานี M38C	Number
M145	จำนวนข้อมูลปริมาณน้ำท่าเฉลี่ยของสถานี M145	Number
M177	จำนวนข้อมูลปริมาณน้ำท่าเฉลี่ยของสถานี M177	Number
M184	จำนวนข้อมูลปริมาณน้ำท่าเฉลี่ยของสถานี M184	Number

3. การวิเคราะห์ประสิทธิภาพของแบบจำลอง

การวิเคราะห์ประสิทธิภาพของแบบจำลองในอนุกรมเวลา โดยใช้หลักการหน้าต่างเคลื่อนที่ (Sliding Window) และมีการวัดค่าความผิดพลาดโดยการเปรียบเทียบกับค่าจริง

1. หลักการ Sliding Window

Sliding Window คือ หลักการที่เป็นที่นิยมนำมาใช้ในการทดลองในการทำเหมืองข้อมูลของข้อมูลอนุกรมเวลา โดยการแบ่งข้อมูลออกเป็น 2 ชุด โดยคำนึงถึงเวลาและทำการเคลื่อนที่หน้าต่าง ไปตามเวลา ดังนี้

ชุดที่ 1 คือ ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data)

ชุดที่ 2 คือ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data)

ในงานวิจัยนี้ผู้วิจัยใช้ข้อมูล ปี ค.ศ. 2012 - 2015 ช่วงเดือน กรกฎาคม ถึง สิงหาคม สามารถแบ่งข้อมูลออกเป็น 3 ปีแรก เพื่อเป็นชุดข้อมูลฝึกสอน และให้หนึ่งเดือนเป็นชุดข้อมูลทดสอบ แล้วจึงทำการเคลื่อนหน้าต่างข้อมูลไปรอบละหนึ่งเดือน โดยจะทำการเคลื่อนหน้าต่างทั้งหมด 4 รอบ ดัง (Table 9)

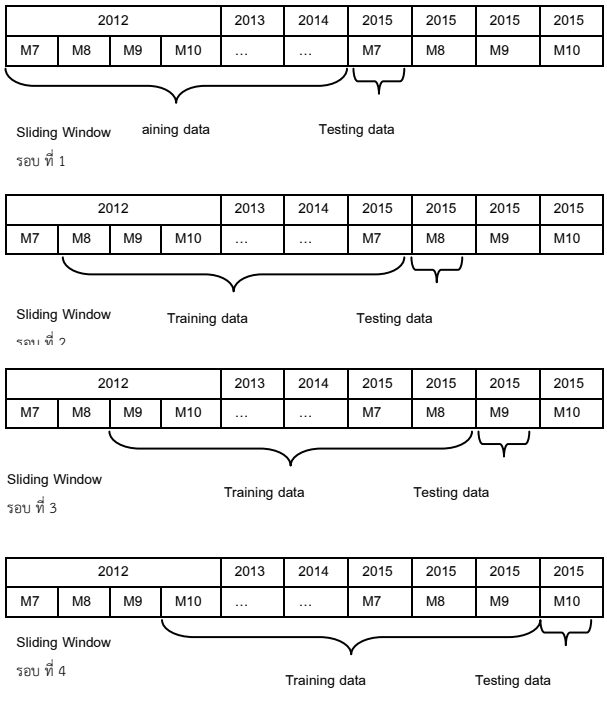

Figure 9 The sliding window method

โอพาร ธรรมานนท์, มณฑลลีลา จินดาไกรฤกษ์และ สมชาติ จิริวิภากร[10] ได้ใช้หลักการหน้าต่างเคลื่อนที่ในการทดลองประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมเพื่อพยากรณ์ความต้องการไฟฟ้า 13 ปีข้างหน้า

4. การวัดค่าประสิทธิภาพของแบบจำลอง

การวัดค่าประสิทธิภาพของแบบจำลองเพื่อวัดประสิทธิภาพของแบบจำลองที่ดีที่สุดสำหรับการพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมา ในงานวิจัยนี้ผู้วิจัยใช้ค่าสำหรับประสิทธิภาพของแบบจำลอง ดังนี้

1) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) มีสูตรในการคำนวณดังนี้

$$MSE = \left(\frac{\Sigma e_t^2}{n}\right)$$

การวัดความถูกต้องของการพยากรณ์ที่นิยมใช้กันอย่างแพร่หลายอีกวิธีหนึ่งคือการหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยหรือ MSE ซึ่งค่านี้ใช้หลักการเดียวกันกับการหาค่าความแปรปรวนในทางสถิติการวัดค่าความคลาดเคลื่อนด้วยวิธีนี้จะได้ค่าความคลาดเคลื่อนที่สูงเนื่องจากเป็นการนำความคลาดเคลื่อนเวลาใด ๆมายกกำลังสองก่อนที่จะหาผลรวมแล้วจึงนำมาหาค่าเฉลี่ยอีกครั้งหนึ่ง

2) ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ความคลาดเคลื่อนเฉลี่ยกำลังสอง หรือค่ารากที่สอง ของค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐาน (RMSD) เป็น วิธีการวัดความคลาดเคลื่อนจากค่าที่พยากรณ์จากแบบจำลองกับค่าจริงที่เกิดขึ้น หากค่า RMSE มีค่าน้อยแสดงว่าแบบจำลองสามารถประมาณค่าได้ใกล้เคียงกับ ค่าจริง ดังนั้นหากค่านี้มีค่าเท่ากับศูนย์แล้ว จะหมายความว่า ไม่เกิดความ คลาดเคลื่อนในแบบจำลอง ค่า RMSE สามารถคำนวณได้ดังสมการ

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

ผลการศึกษา

ในการศึกษาประสิทธิภาพของแบบจำลองนี้คณะผู้วิจัยได้นำเอาโปรแกรม WEKA เวอร์ชัน 3.7.10 มาเป็นเครื่องมือมาใช้ในการสร้างแบบลองด้วยเทคนิค LR, ANN, RBFN และ SMOreg วัดประสิทธิภาพของแบบจำลองด้วย MAE และ RMSE ผลการทดลองสามารถแสดงได้ดังหัวข้อ 4.1 หัวข้อ 4.2 และในหัวข้อ 4.3 แสดงผลการพยากรณ์น้ำในปี 2016 และ 2017

1. การเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Absolute error: MAE) เป็นค่าที่ใช้วัดความถูกต้องของการพยากรณ์ที่นิยมใช้กันอย่างแพร่หลาย จากการทดลองสามารถแสดงผลการเปรียบเทียบของค่า MAE ของ 4 แบบจำลอง LR, ANN, RBFN และ SMOreg ในการสร้างแบบจำลองปริมาณน้ำท่าเฉลี่ยของ 4 สถานีได้แก่ สถานี M38C, M145, M177 และ M184 ได้ดัง (Table 3)

Table 3 A comparison of MAE of the four models for forecasting daily Water Stream Flow Prediction

Model	LR		ANN		RBFN		SMOreg	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
M38C	0.50	0.26	0.45	0.94	0.19	0.34	0.13	0.25
M145	0.62	0.34	4.46	14.93	0.20	0.26	0.13	0.23
M177	0.59	0.18	0.50	0.86	0.26	0.63	0.14	0.42
M184	0.99	1.34	0.24	0.65	0.74	2.19	0.63	1.00
Average	0.67	0.53	1.41	4.34	0.35	0.85	0.26	0.48
S.D.	0.22	0.54	2.04	7.06	0.26	0.90	0.25	0.36

(Table 3) แสดงให้เห็นถึงประสิทธิภาพการทำงานทั้ง 4 แบบจำลองคือ LR, ANN, RBFN และ SMOreg เมื่อเปรียบเทียบการพยากรณ์น้ำท่าเฉลี่ยในฤดูฝนของจังหวัดนครราชสีมา ของสถานีวัดปริมาณน้ำท่าที่มีปริมาณน้ำท่าเฉลี่ยมากที่สุดจำนวน 4 สถานีได้ แก่ สถานี M38C, M145, M177 และ สถานี M184 โดย SMOreg เป็นแบบจำลองที่มีประสิทธิภาพเหนือกว่า แบบจำลอง LR, ANNและ RBFN ด้วยค่า MAE ระหว่าง Training และ Testing แบบจำลอง SMOreg มีค่าเฉลี่ย (Average) ของค่า MAE อยู่ที่ 0.26 และ 0.48 และค่า S.D. ของค่า MAE อยู่ที่ 0.25 และ 0.36 ตามลำดับ และแบบจำลอง LR, ANN และ RBFN มีค่า MAE ระหว่าง Training และ Testing มีค่าเฉลี่ย (Average) 0.67 0.53, 1.14 4.34, และ 0.35 0.85 ตามลำดับ มีค่า S.D. ของค่า MAE ของแบบจำลอง LR, ANN และ RBFN 0.22 0.54, 2.04 7.06 และ 0.25 0.36 ตามลำดับ แสดงให้เป็นว่าค่า MAE ของแบบจำลอง SMOreg มีประสิทธิภาพที่ดีที่สุดในการพยากรณ์น้ำท่าเฉลี่ยในฤดูฝนของจังหวัดนครราชสีมา

2. การเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง

ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) เป็นค่าที่ใช้วัดความถูกต้องโดยเฉลี่ยของการพยากรณ์ จากการทดลองสามารถแสดงผลการเปรียบเทียบของค่า RMSE ของ 4 แบบจำลอง LR, ANN, RBFN และ SMOreg ในการสร้างแบบจำลองปริมาณน้ำท่าเฉลี่ยของ 4 สถานีได้แก่ สถานี M38C, M145, M177 และ M184 ได้ดัง (Table 4)

Table 4 A comparison of RMSE of the four models for forecasting daily Water Stream Flow Prediction

Model	LR		ANN		RBFN		SMOreg	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
M38C	0.82	0.27	0.81	1.09	0.26	0.35	0.24	0.27
M145	1.29	0.42	9.18	16.28	0.38	0.34	0.37	0.32
M177	0.94	0.23	0.88	1.07	0.33	0.64	0.27	0.47
M184	1.44	1.43	1.41	2.45	0.78	1.06	0.42	0.72
Average	1.12	0.59	3.07	5.22	0.44	0.60	0.32	0.44
S.D.	0.29	0.57	4.08	7.40	0.23	0.34	0.08	0.20

(Table 4) แสดงให้เป็นถึงประสิทธิภาพการทำงานทั้ง 4 แบบจำลองคือ LR, ANN, RBFN และ SMOreg เมื่อเปรียบเทียบการพยากรณ์น้ำท่าเฉลี่ยในฤดูฝนของจังหวัดนครราชสีมา ของสถานีวัดปริมาณน้ำท่าที่มีปริมาณน้ำท่าเฉลี่ยมากที่สุดจำนวน 4 สถานีได้ แก่ สถานี M38C, M145, M177 และ สถานี M184 โดย SMOreg เป็นแบบจำลองที่มีประสิทธิมากเหนือกว่า แบบจำลอง LR, ANN และ RBFN ด้วยค่า RMSE ระหว่าง Training และ Testing แบบจำลอง SMOreg มีค่าเฉลี่ย (Average) ของค่า RMSE อยู่ที่ 0.32 และ 0.44 และค่า S.D. ของค่า RMSE อยู่ที่ 0.08 และ 0.20 ตามลำดับ และแบบจำลอง LR, ANN และ RBFN มีค่า RMSE ระหว่าง Training และ Testing มีค่าเฉลี่ย (Average) 1.12 0.59, 3.07 5.22,

และ 0.44 0.60 ตามลำดับ มีค่า S.D. ของค่า RMSE ของแบบจำลอง LR, ANN และ RBFN 0.29 0.57, 4.08 7.40, และ 0.23 0.34 ตามลำดับ แสดงให้เป็นว่าค่า RMSE ของแบบจำลอง SMOreg มีประสิทธิภาพที่ดีที่สุดในการพยากรณ์น้ำท่าเฉลี่ยในฤดูฝนของจังหวัดนครราชสีมา

3. ผลการพยาการณ์น้ำท่าเฉลี่ยของจังหวัดนครราชสีมา

ในงานวิจัยนี้ เราสามารถนำแบบจำลอง SMOreg มาใช้ในการวิเคราะห์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมาของสถานีวัดปริมาณน้ำท่าที่มีปริมาณน้ำท่าเฉลี่ยมากที่สุดจำนวน 4 สถานีได้ แก่ สถานี M38C, M145, M177 โดยใช้ข้อมูลในการ (วันที่ 1 กรกฎาคม ค.ศ. 2012 – วันที่ 31 ตุลาคม ค.ศ. 2014) เป็นชุดข้อมูลฝึกสอน (Training) ด้วยหลักการ Sliding Window และใช้ข้อมูล (วันที่ 1 กรกฎาคม ค.ศ.2015 – วันที่ 31 ตุลาคม ค.ศ.2015) ที่ใช้ในการทดสอบ (Testing)

จากนั้นผู้วิจัยได้ทำการพยากรณ์ปริมาณน้ำท่าเฉลี่ย ด้วยแบบจำลอง SMOreg ของสถานีวัด M38C, M145, M177 และ M184 ในปี ค.ศ. 2016 และ ปี ค.ศ. 2017 ดัง (Figure 10) และ (Figure 11)

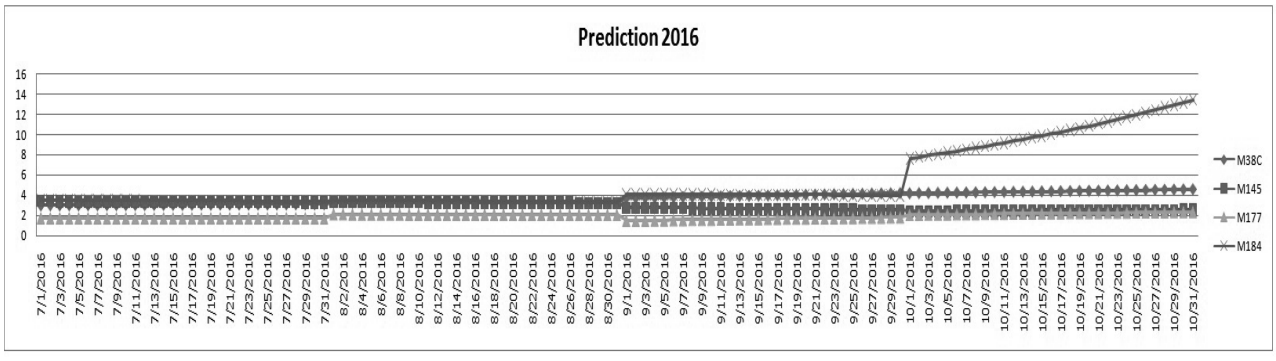

Figure 10 Result of Water Stream Flow Prediction Models for Nakhonratsima Province in year 2016

(Figure 10) แสดงให้เห็นถึงปริมาณน้ำท่าเฉลี่ยของสถานีวัดปริมาณน้ำท่าเฉลี่ย ของจังหวัดนครราชสีมา ในปี ค.ศ. 2016 ค่าที่พยากรณ์ได้จากแบบจำลอง SMOreg ในเดือนกรกฎาคม – สิงหาคม มีปริมาณน้ำท่าเฉลี่ยคงที่ ในเดือนกันยายน – ตุลาคม มีค่าปริมาณน้ำท่าจากการพยากรณ์เพิ่มมากขึ้น สถานีที่มีปริมาณน้ำท่าเฉลี่ยที่มากที่สุด คือ สถานี M184 และ สถานีที่มีปริมาณน้ำท่าเฉลี่ยที่น้อยที่สุด คือ สถานี M177

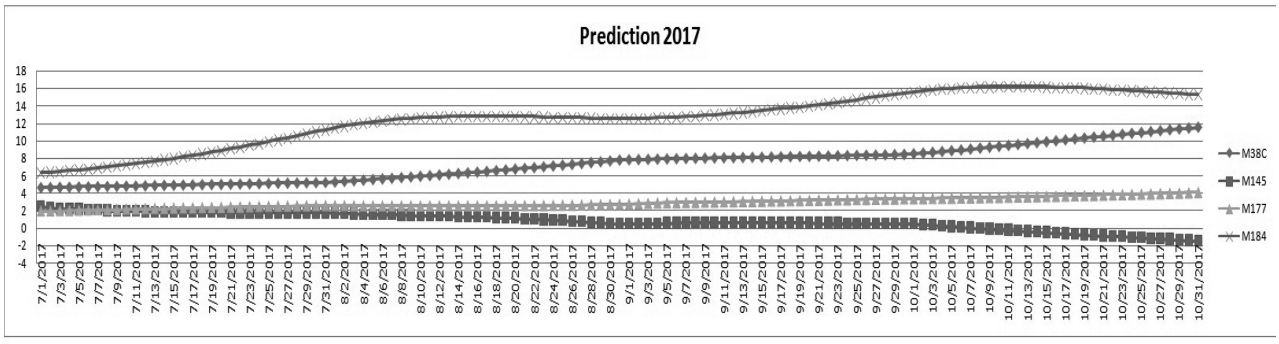

Figure 11 Result of Water Stream Flow Prediction Models for Nakhonratsima Province in year 2017

(Figure 11) แสดงให้เห็นถึงปริมาณน้ำท่าเฉลี่ยของสถานีวัดปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมาปี ค.ศ. 2017 ค่าทีพยากรณ์ได้จากแบบจำลอง SMOreg ในเดือนกรกฎาคม – ตุลาคม มีปริมาณน้ำท่าเฉลี่ย ของสถานี M38C, M177, และM184 มีปริมาณน้ำเพิ่มขึ้นอย่างต่อเนื่อง ส่วนสถานี M145 มีปริมาณน้ำท่าเฉลี่ยที่ลดลงอย่างต่อเนื่อง สถานีที่มีปริมาณน้ำท่าเฉลี่ยที่มากที่สุด คือ สถานี M184 และ สถานีที่มีปริมาณน้ำท่าเฉลี่ยที่น้อยที่สุด คือ สถานี M145

วิจารณ์และสรุป

ในการศึกษาและพัฒนาแบบจำลองที่เหมาะสมสำหรับการพยากรณ์ ปริมาณน้ำท่าเฉลี่ยในจังหวัดนครราชสีมาในช่วงฤดู

ฝน โดยแบบจำลองที่ใช้ในการเปรียบเทียบ 4 แบบจำลองได้แก่ แบบจำลอง LR, ANN RBFN และ SMOreg ผลการทดลองพบว่าแบบจำลอง SMOreg มีความเหมาะสมมากที่สุดในการพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมาในช่วงฤดูฝน โดยการวัดประสิทธิภาพการทำงานของแบบจำลองด้วยค่า MAE และ RMSE ด้วยค่า RMSE ระหว่าง Training และ Testing แบบจำลอง SMOreg มีค่าเฉลี่ย (Average) ของค่า RMSE อยู่ที่ 0.32 และ 0.44 ซึ่งความแตกต่าง จากผลการวิจัยของ Choubey, Mishra and Pandey[6] ศึกษาแบบจำลอง SMOreg ซึ่งเป็นเทคนิคหนึ่งในตระกูลเทคนิคเครื่องเวกเตอร์การสนับสนุน (Support Vector Machine: SVM) เพื่อพยากรณ์ปริมาณน้ำที่ไหลบ่า และปริมาณน้ำฝน ผลการทดรองพบว่าแบบจำลอง SMOreg มีค่า RMSE จากการวิจัยเท่ากับ 2.3731 ค่า RAE เท่ากับ 65.28% และค่า RRSE เท่ากับ 62.491% ซึ่งเป็นค่าที่เปรียบเทียบกับค่าจริง จะเห็นว่าความแตกต่างระหว่างค่าที่พยากรณ์และค่าจริงมีความแตกต่างกันมากโดยมีนัยสำคัญ

จากผลการทดลอง สรุปได้ว่า SMOreg มีความเหมาะสมในการพยากรณ์ปริมาณน้ำท่าเฉลี่ยในช่วงฤดูฝนของจังหวัดนครราชสีมา หากผู้ที่สนใจศึกษาหรือสนใจที่จะพัฒนางานวิจัยนี้ ควรเพิ่มจำนวนข้อมูลชุดฝึกสอนให้มากขึ้น และเพิ่มตัวแปรในการพยากรณ์เพิ่มขึ้นตัวอย่างเช่น อุณหภูมิ ความชื้นสัมพัทธ์ เป็นต้นเพื่อให้แบบจำลองมีประสิทธิภาพมากขึ้นในการ พยากรณ์ และในไปใช้งานทางด้านอื่นต่อไป

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณ ศูนย์อุทกวิทยาชลประทาน ภาคตะวันออกเฉียงเหนือตอนล่าง สำนักบริหารจัดการน้ำและอุทกวิทยา กรมชลประทาน และคณาจารย์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ที่ให้ความอนุเคราะห์ข้อมูลและคำปรึกษาในการดำเนินงานวิจัยครั้งนี้

เอกสารอ้างอิง

1. ไทยรัฐออนไลน์. (2013). *โคราชแล้งหนัก ประกาศพื้นที่ภัยพิบัติฉุกเฉิน 26 อำเภอ*. Available: http://www.thairath.co.th/content/318446
2. dailynews. (2014). *โคราชประกาศพื้นที่ภัยแล้ง 11 อำเภอ*. Available: http://www.dailynews.co.th/regional/288517
3. N. K. Chitsutha Soomlek, Thawat Simano, and Chakchai So-In “Using backpropagation neural networks for flood forecasting in PhraNakhon Si Ayutthaya, Thailand,” *2015 International Computer Science and Engineering Conference (ICSEC),* pp. 1 - 6, 23-26 Nov 2015.
4. M. H. I. MRUI.Mustafa “Prediction of River Suspended Sediment Load Using Radial Basis Function Neural Network-A Case Study in Malaysia,” *National Postgraduate Conference (NPC), 2011,* pp. 1 - 4, 19-20 Sept 2011.
5. P. S. Munin Wanatada, “Runoff Forecasting Using Back-propagation Neural Network Technique: Case Study of Municipality of Chaiyaphum,” *The 9th National Conference on Computing and Information Technology,* pp. 179-184, 2013.
6. S. M. V.Choubey, S.K.Pandey, “Time Series Data Mining in Real Time Surface Runoff Forecasting through Support Vector Machine,” *International Journal of Computer Applications (0975 – 8887),* vol. Volume 98– No.3,, July 2014 2014.
7. สำนักงานจังหวัดนครราชสีมา, *แผนพัฒนาจังหวัดนครราชสีมา 4 ปี พ.ศ. 2558-2561 (ฉบับทบทวน) โครราชเมืองหลวงของภาคอีสาน*, 2558.
8. ผศ. ดร.โชติรัตน์ รัตนามหัทธนะ (2011). *เหมืองข้อมูลอนุกรมเวลา*. Available: http://www.chulapedia.chula.ac.th/index.php/เหมืองข้อมูลอนุกรมเวลา
9. B. L. GAO Xiang, “The Prediction of Indices at Infall of Confluent Flow Network of Wastewater with Multivariate Linear Regression*,” *Proceedings of the 29th Chinese Control Conference,* pp. 5125-5129, July 29-31, 2010.
10. ม. ล. โอพาร ธรรมานนท์, สมชาติ จิริวิภากร, “Long-Term Peak Load Forecasting for EGAT System Using Artificial Neural Networks,” *Ladkrabang Engineering Journal,* vol. 21, 2,June 2004.