

การพยากรณ์ค่าความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอนด้วยเครือข่ายประสาทเทียมร่วมกับรูปแบบออโตรีเกรสซีฟ

Forecasting the PM10 Concentration by Using the Artificial Neural Network and the Autoregressive Form

รณชัย ชื่นธวัช

Ronnachai Chuentawat

หลักสูตรวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครราชสีมา

Program in Computer Science, Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University

Received : 2 March 2017

Accepted : 16 November 2017

Published online : 20 November 2017

บทคัดย่อ

งานวิจัยนี้ ศึกษาการพัฒนาตัวแบบพยากรณ์ค่าความเข้มข้นเฉลี่ยรายวันของฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ในเขตพื้นที่ภาคเหนือของประเทศไทย จำนวน 3 ชุดข้อมูล แต่ละชุดข้อมูล เป็นอนุกรมเวลาเชิงเดี่ยว ที่มีคาบเวลารายวัน ตั้งแต่วันที่ 1 มกราคม ถึงวันที่ 31 พฤษภาคม 2559 และดำเนินการสร้างตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียม ผสมผสานกับรูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA เรียกว่า ตัวแบบ AR-ANN จากนั้น เปรียบเทียบความแม่นยำของตัวแบบ AR-ANN กับตัวแบบ ARIMA ด้วยการใช้ค่าคลาดเคลื่อน RMSE และ MAPE ผลการทดลอง พบว่า ตัวแบบผสมผสาน AR-ANN ให้ค่า RMSE และค่า MAPE ต่ำกว่าตัวแบบ ARIMA ทั้ง 3 ชุดข้อมูล ดังนั้น ตัวแบบ AR-ANN สามารถนำมาใช้พยากรณ์ค่าความเข้มข้นเฉลี่ยรายวันของ PM10 ในเขตพื้นที่ภาคเหนือของประเทศไทย ได้อย่างเหมาะสม

คำสำคัญ : การพยากรณ์ ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน เครือข่ายประสาทเทียม รูปแบบออโตรีเกรสซีฟ

Abstract

This research aims to study a development of the forecasting model to predict the daily average PM10 concentration in the North area of Thailand with 3 datasets. Each of datasets is the univariate time series that is a daily data, since 1st Jan - 31th May 2016. To generate the forecasting model, we present a forecasting model by the artificial neural network technique combine with the autoregressive of ARIMA model, called AR-ANN model. We evaluated our experiment by measurement error between our AR-ANN model and the ARIMA model. The error measurement of each model is measured by the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE). From the result, we found that RMSE and MAPE of the AR-ANN model has lower than ARIMA model for whole datasets. Therefore, we concluded that our AR-ANN model can use to forecast the daily average PM10 concentration appropriately.

Keywords : forecasting, PM10, Artificial Neural Networks, autoregressive

*Corresponding author. E-mail : c_ronnachai@hotmail.com

บทนำ

ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) เป็นสารเจือปนในอากาศอย่างหนึ่งที่เกิดมลพิษทางอากาศ ถ้าในอากาศมี PM10 ในปริมาณมาก จะมีผลเสียต่อระบบทางเดินหายใจ และอาจทำให้เกิดโรคที่เกี่ยวข้องต่อระบบทางเดินหายใจ ร้ายแรงถึงขั้นเสียชีวิตได้ ซึ่งเอกสารรายงานเรื่อง “แนวทางการเฝ้าระวังพื้นที่เสี่ยงจากมลพิษทางอากาศ กรณีฝุ่นละอองขนาดเล็ก” ของกรมอนามัย (Department of Health, 2015) กล่าวว่า “ผลกระทบต่อสุขภาพจากฝุ่นละอองขนาดเล็กสามารถก่อให้เกิดผลกระทบต่อสุขภาพได้หลายระบบ เช่น ระบบทางเดินหายใจ (การไอและอาการของระบบทางเดินหายใจส่วนล่าง) ระบบหัวใจ และหลอดเลือด (กล้ามเนื้อหัวใจขาดเลือด หัวใจเต้นไม่สม่ำเสมอ หัวใจวาย) ระบบตา ระบบผิวหนัง ฝุ่นขนาดเล็กยังเพิ่มความเสี่ยงของอัตราการตายจากภาวะเส้นเลือดอุดตันในสมอง และทำให้น้ำหนักของทารกในครรภ์ลดลงอีกด้วย ทำให้้อตราป่วยและอัตราการตายด้วยโรคระบบทางเดินหายใจ และระบบหัวใจและหลอดเลือดเพิ่มขึ้น และอัตราดังกล่าวจะเพิ่มขึ้นตามปริมาณความเข้มข้นของฝุ่นในอากาศ” ดังนั้น กรมควบคุมมลพิษ (Pollution Control Department, 2016) จึงกำหนดค่ามาตรฐานของ PM10 สำหรับประเทศไทย ไว้ว่า ค่าเฉลี่ย 24 ชั่วโมงของ PM10 ถ้ามากกว่า 120 ไมโครกรัมต่อลูกบาศก์เมตร ($\mu\text{g}/\text{m}^3$) จะเริ่มมีผลเสียต่อสุขภาพของมนุษย์ และด้วยผลกระทบต่อสุขภาพ จากปริมาณความเข้มข้นของ PM10 ในอากาศ จึงมีงานวิจัยเพื่อพัฒนาตัวแบบพยากรณ์สำหรับพยากรณ์ปริมาณความเข้มข้นของ PM10 ด้วยเทคนิคต่างๆ หลากหลายงานวิจัย และสามารถนำมากล่าวถึง ได้ดังนี้

Antanasijevic และคณะ (2013) ได้ศึกษาและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียม (Artificial neural network) ร่วมกับการคัดเลือกตัวแปรนำเข้าด้วยเจเนติกอัลกอริทึม (Genetic algorithm) เพื่อพยากรณ์ปริมาณการปลดปล่อย PM10 ประจำปี ในเขตพื้นที่ 26 ประเทศของทวีปยุโรป จากนั้น เปรียบเทียบความแม่นยำของค่าพยากรณ์กับเทคนิคการวิเคราะห์ความถดถอยเชิงพหุ (Multiple linear regression) และวิธีวิเคราะห์ความถดถอยขององค์ประกอบหลัก (Principal component regression) ผลการทดลอง พบว่า ตัวแบบพยากรณ์ที่นำเสนอ มีความคลาดเคลื่อนประมาณ 10% และมีความแม่นยำมากกว่าวิธีการอื่น ที่นำมาเปรียบเทียบเกือบสามเท่า

Grivas และ Chaloulakou (2006) ได้ศึกษาการพยากรณ์ค่าความเข้มข้นรายชั่วโมงของ PM10 โดยพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียม (Artificial neural network) และดำเนินการศึกษาในพื้นที่ 4 แห่งของ Greater Athens Area ที่ประสบปัญหามลพิษจาก PM10 ข้อมูลที่ใช้ศึกษาเป็นข้อมูลระหว่างปี 2001-2002 ซึ่งพัฒนาตัวแบบพยากรณ์โดยใช้ตัวแปรนำเข้าเป็นข้อมูลทางอุตุนิยมวิทยา จากนั้น ใช้ขั้นตอนวิธีเชิงวิวัฒนาการ (Genetic Algorithms) เพื่อคัดเลือกตัวแปรนำเข้าที่เหมาะสม และประเมินผลด้วยการวัดค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ค่า Root Mean Squared Error (RMSE) และค่า Mean Absolute Error (MAE) เปรียบเทียบกับตัวแบบพยากรณ์ที่สร้างจากเทคนิคการวิเคราะห์ความถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression) ผลการทดลอง พบว่า ตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียมเพียงพอก็จะนำมาใช้พยากรณ์ค่าความเข้มข้นของ PM10 เมื่อมีความเข้มข้นอยู่ในระดับสูง และสามารถนำไปใช้สร้างตัวแบบพยากรณ์ เพื่อพยากรณ์ความเข้มข้นของ PM10 แบบทันที (Real-time) ในช่วงการแข่งขันกีฬาโอลิมปิก ปี 2004

He และคณะ (2014) ได้ศึกษาและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียม เพื่อพยากรณ์ความเข้มข้นของ PM10 ในเขตพื้นที่ Mong Kok, Hong Kong โดยสร้างตัวแบบเครือข่ายประสาทเทียมที่มีโครงสร้างแบบ Multilayer Perceptron (MLP) จากนั้น ใช้อัลกอริทึมในการฝึกสอนตัวแบบพยากรณ์ที่แตกต่างกัน 3 ชนิด คือ (1) Levenberge-Marquardt (2) Particle Swarm Optimization (PSO) (3) Chaotics Particle Swarm Optimization (CPSO) โดยประเมินผลการทดลองด้วย

การวัดค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) และค่า Mean Square Error (MSE) ผลการทดลอง พบว่า ตัวแบบ MLP ที่ใช้อัลกอริทึมฝึกสอนแบบ CPSO ให้ผลดีที่สุด โดยมีค่า MSE ต่ำที่สุด และมีค่าสัมประสิทธิ์สหสัมพันธ์สูงสุด

Hooyberghs และคณะ (2005) ได้ศึกษาและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียม เพื่อพยากรณ์ความเข้มข้นเฉลี่ยประจำวันของ PM10 ล่วงหน้า 1 วัน ในเขตพื้นที่ของประเทศเบลเยียม โดยใช้ข้อมูลสังเคราะห์ด้านอุตุนิยมวิทยาในพื้นที่ 10 จุด ระหว่างปี 1997-2001 เป็นตัวแปรนำเข้า (Input parameter) และประเมินผลด้วยการวัดค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ค่า Root Mean Squared Error (RMSE) และค่า Success Index (SI) ผลการทดลอง พบว่า ตัวแปรนำเข้าที่มีผลต่อความแม่นยำในการพยากรณ์มากที่สุดคือ ความสูงของชั้นขอบเขต (Boundary Layer Height) เมื่อเพิ่มตัวแปรนำเข้าด้านอื่น เช่น ความเร็วลม อุณหภูมิ ความหนาแน่นของเมฆ หรือทิศทางลม พบว่า สามารถเพิ่มความแม่นยำได้เพียงเล็กน้อยเท่านั้น

Perez (2012) ได้ศึกษาและพัฒนาตัวแบบพยากรณ์เพื่อนำมาใช้พยากรณ์ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมงของ PM10 ในเขตพื้นที่เมืองซานติอาโก ประเทศชิลี (Santiago, Chile) โดยการพยากรณ์ในงานวิจัยนี้ได้สร้างตัวแบบพยากรณ์ด้วยเทคนิคเครือข่ายประสาทเทียมผสมผสานกับวิธี Nearest Neighbor ซึ่งมีอินพุต คือ ค่าความเข้มข้นของ PM10 จากสถานีตรวจวัดที่กระจายอยู่โดยรอบ และข้อมูลทางอุตุนิยมวิทยา และมีเอาต์พุต คือ ค่าความเข้มข้นสูงสุดของ PM10 ล่วงหน้า 1 วัน โดยประเมินผลการทดลองด้วยการวัดค่าคลาดเคลื่อน Percentage Error (PE) ผลการทดลอง พบว่า ตัวแบบเครือข่ายประสาทเทียมผสมผสานกับวิธี Nearest Neighbor สามารถนำไปใช้เป็นเครื่องมือควบคุมมลพิษทางอากาศได้

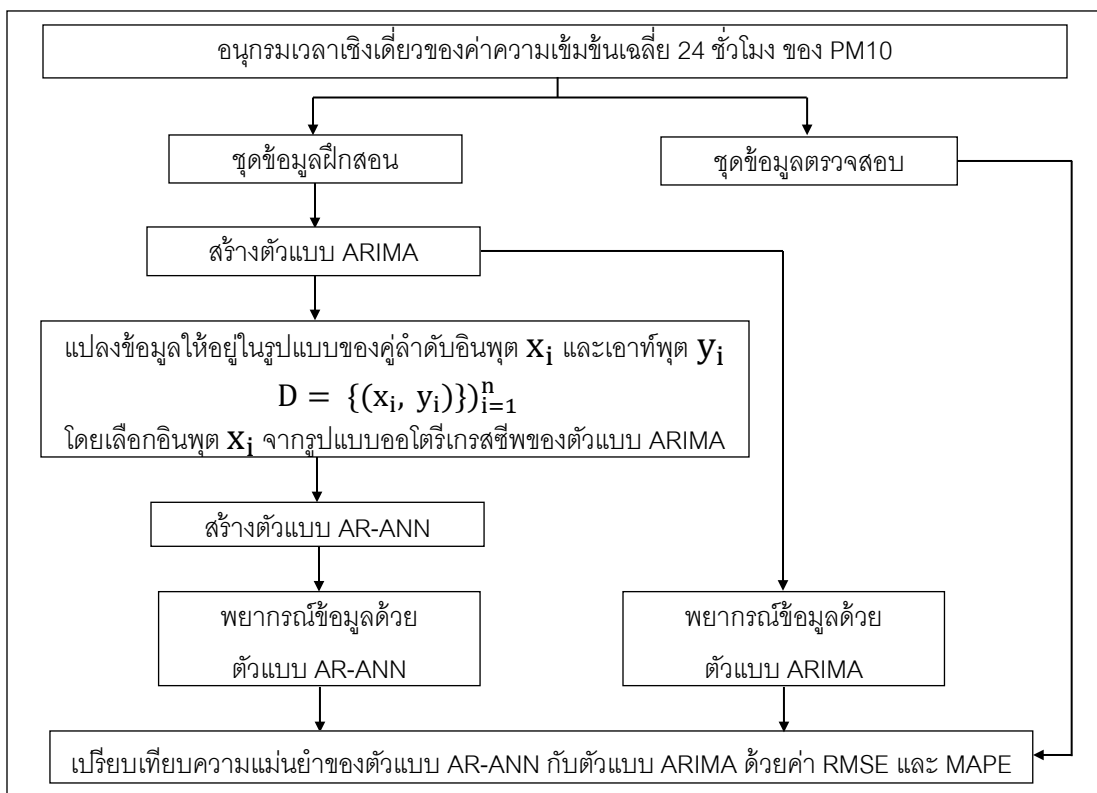
งานวิจัยนี้ สนใจศึกษาการพัฒนาตัวแบบพยากรณ์ เพื่อพยากรณ์ค่าความเข้มข้นของ PM10 ในเขตพื้นที่ภาคเหนือของประเทศไทย ซึ่งประสบปัญหามลพิษทางอากาศ เนื่องจากค่าความเข้มข้นของ PM10 สูงกว่าค่ามาตรฐานที่กรมควบคุมมลพิษกำหนดไว้ ในบางช่วงเวลาของปี โดยมีการจัดตั้งศูนย์ตรวจวัดค่าความเข้มข้นของ PM10 กระจายอยู่ตามจังหวัดต่างๆ ในเขตพื้นที่ภาคเหนือ และกรมควบคุมมลพิษได้เผยแพร่ข้อมูล ซึ่งเป็นค่าความเข้มข้นเฉลี่ย 24 ชั่วโมง (รายวัน) ของ PM10 ในแต่ละจังหวัดภาคเหนือของประเทศไทย ที่เว็บไซต์ของกรมควบคุมมลพิษ หน้ารายงานสถานการณ์คุณภาพอากาศและเสียงประเทศไทย (Pollution Control Department, 2016) โดยข้อมูลดังกล่าว มีลักษณะเป็นอนุกรมเวลาเชิงเดี่ยว (Univariate time series) ที่มีค่าสังเกต 1 ค่า คือ ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมง ของ PM10 บันทึกค่าตามลำดับเวลา ด้วยคาบเวลาเป็นรายวัน ซึ่งจากการศึกษาข้อมูลดังกล่าว พบว่า ข้อมูลมีค่าสูญหาย (Missing value) ในทุกชุดข้อมูล ดังนั้น ผู้วิจัยจึงเลือกชุดข้อมูลที่มีค่าสูญหายจำนวนน้อยที่สุด 3 ชุดข้อมูล ได้แก่ 1) ชุดข้อมูลของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่ 2) ชุดข้อมูลของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง และ 3) ชุดข้อมูลของ ต.ในเมือง อ.เมือง จ.ลำพูน โดยแก้ไขปัญหาค่าสูญหายด้วยวิธีการหาค่าเฉลี่ยระหว่างวันก่อนหน้าและวันถัดไป และข้อมูลทั้ง 3 ชุดเป็นข้อมูลรายวัน ตั้งแต่วันที่ 1 มกราคม ถึงวันที่ 31 พฤษภาคม 2559 จำนวน 152 ค่าสังเกต เพื่อดำเนินการศึกษาในงานวิจัยนี้

จากลักษณะของข้อมูลที่ใช้ศึกษา มีลักษณะเป็นอนุกรมเวลาเชิงเดี่ยว ซึ่งตัวแบบพยากรณ์ที่ได้รับความนิยมในการนำมาพยากรณ์อนุกรมเวลาเชิงเดี่ยว คือ ตัวแบบ ARIMA (Autoregressive Integrated Moving Average) ที่เป็นวิธีการของบ็อกซ์และเจนกินส์ (Box and Jenkins, 1990) เนื่องจากให้ความแม่นยำสูง (Lee and Tong, 2011; Wang et al., 2012; Wang and Meng, 2012) โดยตัวแบบ ARIMA สามารถตรวจจบบรูปแบบเชิงเส้นในอนุกรมเวลาได้ดี แต่ตรวจจบบรูปแบบไม่เป็นเชิงเส้นได้ไม่ดี ส่วนเทคนิคเครือข่ายประสาทเทียม (Artificial Neural Networks: ANN) ซึ่งเป็นเทคนิคด้านแมชชีนเลิร์นนิง (Machine learning) สามารถตรวจจบบรูปแบบไม่เป็นเชิงเส้นได้ดี แต่ตรวจจบบรูปแบบเชิงเส้นได้ไม่ดี (Wang and Meng, 2012; Wongsathan and Seedadan, 2016) ดังนั้น งานวิจัยนี้ จึงมีวัตถุประสงค์ เพื่อนำเสนอตัวแบบผสมผสานระหว่างรูปแบบออโตเรกเรสซีฟ

(Autoregressive) ของตัวแบบ ARIMA กับเทคนิค ANN เรียกว่า ตัวแบบ AR-ANN ซึ่งเป็นตัวแบบพยากรณ์ที่มีความแตกต่างจากตัวแบบพยากรณ์ของงานวิจัยที่ได้นำมาทบทวนข้างต้น เพราะมีการใช้รูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA มากำหนดจำนวนนิเวรอนในชั้นอินพุตของโครงข่ายประสาทเทียม (ANN) ทำให้ได้ตัวแบบผสมผสานที่สามารถตรวจจับรูปแบบเชิงเส้นและรูปแบบไม่เป็นเชิงเส้นได้ทั้ง 2 รูปแบบ โดยนำตัวแบบ AR-ANN มาพยากรณ์ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมงของ PM10 สำหรับชุดข้อมูลทั้ง 3 ชุด จากนั้น เปรียบเทียบความแม่นยำของค่าพยากรณ์จากตัวแบบ AR-ANN กับค่าพยากรณ์จากตัวแบบ ARIMA เพื่อประเมินประสิทธิภาพของตัวแบบ AR-ANN ว่าสามารถเพิ่มความแม่นยำของค่าพยากรณ์ได้สูงกว่าค่าพยากรณ์จากตัวแบบ ARIMA ซึ่งเป็นวิธีการทางสถิติแบบดั้งเดิม (Traditional statistic) หรือไม่

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบ AR-ANN ซึ่งเป็นตัวแบบที่ผสมผสานระหว่างรูปแบบออโตรีเกรสซีฟ (Autoregressive) ของตัวแบบ ARIMA ซึ่งสามารถตรวจจับรูปแบบเชิงเส้นได้ดี ร่วมกับเทคนิค ANN ที่สามารถตรวจจับรูปแบบไม่เป็นเชิงเส้นได้ดี ภายใต้สมมติฐานว่า ตัวแบบ AR-ANN จะสามารถตรวจจับรูปแบบเชิงเส้นและรูปแบบไม่เป็นเชิงเส้นได้ทั้ง 2 รูปแบบ และให้ความแม่นยำได้ดีกว่าตัวแบบ ARIMA ที่สามารถตรวจจับรูปแบบเชิงเส้นได้ดี เพียงรูปแบบเดียว โดยดำเนินการทดสอบสมมติฐานด้วยการวัดค่าคลาดเคลื่อน RMSE และ MAPE ของตัวแบบ AR-ANN เปรียบเทียบกับตัวแบบ ARIMA และมีกรอบแนวคิดของงานวิจัย ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดงานวิจัย

จากกรอบแนวคิดงานวิจัย งานวิจัยนี้ได้นำเสนอการสร้างตัวแบบพยากรณ์อนุกรมเวลาเชิงเดี่ยว ด้วยการนำเทคนิคที่แตกต่างกัน 2 เทคนิค มาผสมผสานกัน ได้แก่ เทคนิคการสร้างตัวแบบ ARIMA และเทคนิคเครือข่ายประสาทเทียม (ANN) โดยเทคนิคการสร้างตัวแบบ ARIMA เป็นวิธีการวิเคราะห์อนุกรมเวลาเชิงเดี่ยวแบบดั้งเดิม ที่นำเสนอโดยบ็อกซ์และเจนกินส์ (Box and Jenkins, 1990) เรียกว่า วิธีการของบ็อกซ์และเจนกินส์ ส่วนเทคนิค ANN เป็นวิธีการด้านแมชชีนเลิร์นนิง ที่มีแนวคิดจากวิธีการทำงานของเครือข่ายประสาทในสมองของมนุษย์ โดยการใช้การเรียนรู้แบบแพร่กลับ (Back propagation) ที่นำเสนอโดย Kumar (2004) ซึ่งการสร้างตัวแบบพยากรณ์จากทั้ง 2 เทคนิค มีขั้นตอนแตกต่างกัน เมื่อนำมาผสมผสานเป็นตัวแบบ AR-ANN ตามกรอบแนวคิดงานวิจัย จึงมีขั้นตอนดำเนินการวิจัยที่สามารถแบ่งส่วน ได้ดังนี้

1. ขั้นตอนการสร้างตัวแบบ ARIMA

ตัวแบบ ARIMA เป็นตัวแบบที่สร้างจากวิธีการของบ็อกซ์และเจนกินส์ ซึ่งจะพยากรณ์ข้อมูลจากรูปแบบหลัก 2 รูปแบบ ได้แก่ รูปแบบออโตรีเกรสซีฟ (Autoregressive: AR) คือ รูปแบบที่กำหนดว่า ค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าสังเกตก่อนหน้า และรูปแบบมูฟวิงเอเวอร์เรจ (Moving Average: MA) คือ รูปแบบที่กำหนดว่า ค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าคลาดเคลื่อน (Errors) ก่อนหน้า ซึ่งอนุกรมเวลาที่นำมาใช้ในการสร้างตัวแบบ ARIMA ต้องเป็นอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary) ถ้าอนุกรมเวลามีลักษณะไม่นิ่ง (Non-stationary) จะต้องแปลงอนุกรมเวลาให้มีลักษณะนิ่ง ด้วยการหาผลต่างระหว่างค่าสังเกตในอนุกรมเวลาเดิม เมื่อนำทุกส่วนรวมกัน จะเรียกว่า รูปแบบ ARIMA (Autoregressive Integrated Moving Average) โดยมีรูปแบบทั่วไปที่เขียนในรูปสมการของตัวดำเนินการย้อนกลับ (Backward shift operator: B) ได้ดังสมการที่ 1 (Wang *et al.*, 2012)

$$\theta_p(B)\Theta_p(B^S)(1-B)^d(1-B^S)^D Y_t = w_q(B)W_q(B^S)\alpha_t \quad \text{----- (1)}$$

เมื่อ

Y_t = ค่าสังเกต ณ เวลา t ใดๆ

S = จำนวนคาบเวลาใน 1 ฤดูกาล

α_t = ค่าคลาดเคลื่อน (Error) ณ เวลา t ใดๆ

โดยที่

$$BY_t = Y_{t-1} \text{ และ } B^k Y_t = Y_{t-k}$$

$$w_q(B) = 1 - w_1 B - w_2 B^2 - \dots - w_q B^q$$

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p$$

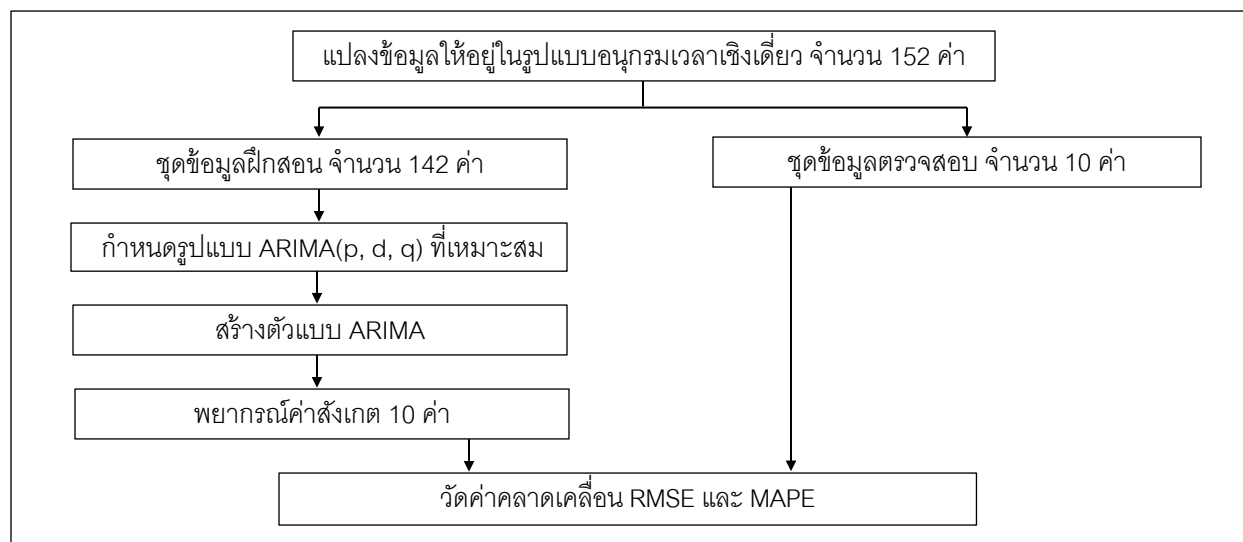
$$\Theta_p(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_p B^{pS}$$

$$W_q(B^S) = 1 - W_1 B^S - W_2 B^{2S} - \dots - W_q B^{qS}$$

การสร้างตัวแบบ ARIMA จะต้องวิเคราะห์อนุกรมเวลา เพื่อกำหนดรูปแบบ ARIMA(p, d, q)x(P, D, Q)_S ที่เหมาะสมกับอนุกรมเวลา โดยอันดับ p, d และ q เป็นตัวเลขจำนวนเต็ม หมายถึง อันดับจากส่วนประกอบแนวโน้มของอนุกรมเวลา เมื่ออันดับ p หมายถึง อันดับของรูปแบบออโตรีเกรสซีฟ อันดับ d หมายถึง ระยะห่างของคาบเวลาที่นำมาหาผลต่างระหว่างค่าสังเกตในการปรับอนุกรมเวลาให้มีลักษณะนิ่ง และอันดับ q หมายถึง อันดับของรูปแบบมูฟวิงเอเวอร์เรจ ส่วนอันดับ

P, D, Q หมายถึง อันดับจากส่วนประกอบฤดูกาลของอนุกรมเวลา แต่ละอันดับ มีความหมายเช่นเดียวกับอันดับจากส่วนประกอบแนวโน้ม

ข้อมูลที่ใช้ศึกษาในงานวิจัยนี้ เป็นอนุกรมเวลาเชิงเดี่ยว ที่มีค่าสังเกต คือ ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมงของ PM10 ดังนั้น ข้อมูลมีคาบเวลาเป็นรายวัน ทำให้ไม่มีรอบฤดูกาล จึงทำให้อันดับจากส่วนประกอบฤดูกาล และจำนวนคาบเวลาใน 1 ฤดูกาล มีค่าเท่ากับศูนย์ ดังนั้น จึงใช้รูปแบบ ARIMA(p, d, q) สำหรับสร้างตัวแบบ ARIMA และจากการศึกษางานวิจัยของ Thanapala, Charoensiri และ Soponpimol (2016) ที่ได้สร้างตัวแบบ ARIMA เพื่อพยากรณ์ราคาสับปะรดที่ส่งเข้าโรงงานในประเทศไทย โดยใช้ข้อมูลจำนวน 106 ค่า ซึ่งแบ่งเป็นชุดข้อมูลฝึกสอน (Training dataset) จำนวน 100 ค่า และชุดข้อมูลตรวจสอบ (Validating dataset) จำนวน 6 ค่า หรืองานวิจัยของ Khanthongkham, Charoensiri และ Soponpimol (2016) ที่ได้สร้างตัวแบบ ARIMA เพื่อพยากรณ์ราคาไก่พันธุ์เนื้อในประเทศไทย โดยใช้ข้อมูลรายเดือน จำนวน 103 ค่า ในการสร้างตัวแบบพยากรณ์ แล้วนำตัวแบบพยากรณ์มาใช้ในการพยากรณ์ราคาไก่พันธุ์เนื้อล่วงหน้า 2 ค่า ดังนั้น งานวิจัยนี้ จึงได้แบ่งข้อมูลจำนวน 152 ค่า เป็นชุดข้อมูลฝึกสอนในการสร้างตัวแบบพยากรณ์ จำนวน 142 ค่า และชุดข้อมูลตรวจสอบในการวัดความแม่นยำของตัวแบบพยากรณ์ จำนวน 10 ค่า เพื่อให้มีสัดส่วนใกล้เคียงกับงานวิจัยที่ได้กล่าวมาข้างต้น อีกประการหนึ่งงานวิจัยนี้ คือ การสร้างตัวแบบพยากรณ์ สำหรับพยากรณ์ค่าความเข้มข้นของ PM10 ในอากาศ จึงเลือกกำหนดข้อมูลตรวจสอบที่จะนำมาเปรียบเทียบกับค่าพยากรณ์ล่วงหน้าไม่มากเกินไป เพื่อไม่ให้เกิดค่าคลาดเคลื่อนสะสมสูงเกินไป ดังนั้น ขั้นตอนการสร้างตัวแบบ ARIMA ของงานวิจัยนี้ จึงกำหนดได้ดังภาพที่ 2



ภาพที่ 2 ขั้นตอนการสร้างตัวแบบ ARIMA

การสร้างตัวแบบ ARIMA จะดำเนินการเขียนโปรแกรมด้วยภาษาอาร์ (R language) ซึ่งมีแพ็คเกจชื่อ forecast ที่เผยแพร่โดย Hyndman (2015) สำหรับใช้สร้างอัลกอริทึม ตามขั้นตอนดังภาพที่ 2 แต่ละขั้นตอน อธิบายได้ดังนี้

1. แปลงข้อมูลให้อยู่ในรูปแบบของอนุกรมเวลาเชิงเดี่ยว เป็นขั้นตอนการแปลงข้อมูลที่เป็นค่าความเข้มข้นเฉลี่ย 24 ชั่วโมงของ PM10 ตั้งแต่วันที่ 1 ม.ค. ถึงวันที่ 31 พ.ค. 2559 จำนวน 152 ค่า ให้มีโครงสร้างในรูปแบบของอนุกรมเวลาเชิงเดี่ยวซึ่งภาษาอาร์ มีฟังก์ชันชื่อ ts() ทำหน้าที่แปลงข้อมูลให้อยู่ในรูปแบบของอนุกรมเวลา ที่มีค่าสังเกตเรียงตามลำดับเวลา
2. แบ่งข้อมูลเป็น 2 ส่วน ส่วนที่ 1 เป็นชุดข้อมูลฝึกสอน โดยใช้ข้อมูลตั้งแต่วันที่ 1 ม.ค. ถึงวันที่ 21 พ.ค. 2559 จำนวน 142 ค่า สำหรับสร้างตัวแบบ ARIMA และส่วนที่ 2 เป็นชุดข้อมูลตรวจ แบ่งแยกออกมา เพื่อใช้ประเมินความแม่นยำของตัวแบบพยากรณ์ โดยใช้ข้อมูลตั้งแต่วันที่ 22 พ.ค. ถึงวันที่ 31 พ.ค. 2559 จำนวน 10 ค่า สำหรับวัดความแม่นยำของตัวแบบพยากรณ์
3. กำหนดรูปแบบ ARIMA(p, d, q) ที่เหมาะสม เป็นขั้นตอนการระบุอันดับ p, d, q ของตัวแบบ ARIMA ที่เหมาะสมที่สุด ซึ่งภาษาอาร์ มีฟังก์ชันชื่อ auto.arima() เพื่อใช้กำหนดอันดับ p, d, q ให้กับตัวแบบ ARIMA
4. สร้างตัวแบบ ARIMA ด้วยฟังก์ชันชื่อ arima()
5. พยากรณ์ค่าสังเกต 10 ค่า ซึ่งเป็นค่าความเข้มข้นเฉลี่ย 24 ชั่วโมงของ PM10 ในวันที่ 22 พ.ค. ถึงวันที่ 31 พ.ค. 2559 เพื่อนำไปวัดค่าคลาดเคลื่อนกับค่าสังเกตจริงของชุดข้อมูลตรวจสอบ ที่เป็นข้อมูลในช่วงเวลาเดียวกันกับค่าพยากรณ์
6. วัดความแม่นยำของค่าพยากรณ์จากตัวแบบ ARIMA ด้วยการคำนวณค่าคลาดเคลื่อน RMSE ที่เป็นการวัดขนาดของความคลาดเคลื่อนระหว่างค่าพยากรณ์กับค่าจริง โดยมีหน่วยวัดเป็นหน่วยเดียวกับหน่วยข้อมูล และค่า MAPE ที่เป็นการวัดร้อยละของความคลาดเคลื่อน โดยวัดค่าพยากรณ์คลาดเคลื่อนจากค่าจริงในรูปของร้อยละ ซึ่งสมการคำนวณค่า RMSE และ MAPE แสดงได้ดังสมการที่ 2 และ 3 ตามลำดับ (Wang *et al.*, 2012)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad \text{----- (2)}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \frac{y_t - \hat{y}_t}{y_t} \right| \quad \text{----- (3)}$$

เมื่อ y_t = ค่าสังเกตที่เวลา t , \hat{y}_t = ค่าพยากรณ์ที่เวลา t , n = จำนวนช่วงเวลาที่ยากรณ์

2. ขั้นตอนการสร้างตัวแบบ AR-ANN

เครือข่ายประสาทเทียม (ANN) คือ การจำลองวิธีการทำงานของเครือข่ายประสาทในสมองของมนุษย์ เมื่อนำมาใช้ในการพยากรณ์อนุกรมเวลาเชิงเดี่ยว จะนำเครือข่ายประสาทเทียมที่มีโครงสร้างไปข้างหน้าแบบหลายชั้นด้วยการเรียนรู้แบบแพร่กลับ มาใช้พยากรณ์ค่าสังเกตล่วงหน้า โดยมีเอาต์พุตของเครือข่ายเป็นค่าสังเกต ณ เวลา t (y_t) และอินพุตเวกเตอร์ (Input vector) เป็นค่าสังเกต ณ เวลาที่ผ่านมาไป 1 ถึง p คาบเวลา ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) ดังนั้น ชุดข้อมูลฝึกสอน (Training dataset) ที่ใช้สร้างเครือข่ายจะประกอบด้วยข้อมูล 2 ชุด ได้แก่ ชุดที่ 1 คือ อินพุตเวกเตอร์ เป็นค่าสังเกต ณ เวลาที่ผ่านมาไป 1 ถึง p คาบเวลา และชุดที่ 2 คือ ทาร์เก็ต (Target) หรือเอาต์พุต เป็นค่าสังเกต ณ เวลา t ซึ่งเป็นค่าตัวเลข 1 ค่า ทำให้มีนิวรอน (Neuron)

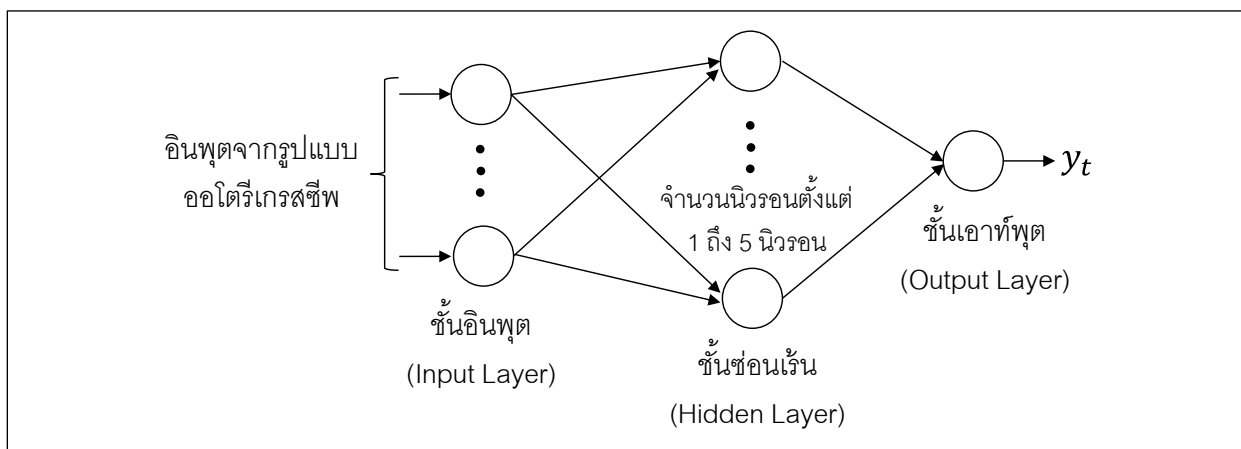
ในชั้นเอาต์พุต (Output layer) 1 นิวรอน ซึ่งสามารถแสดงความสัมพันธ์ระหว่างเอาต์พุตและอินพุต ได้ดังสมการที่ 4 (Wang and Meng, 2012)

$$y_t = w_0 + \sum_{j=1}^q w_j g(w_{0j} + \sum_{i=1}^p w_{ij} y_{t-i}) + e_t \quad \text{----- (4)}$$

เมื่อ w_j ($j = 1, 2, \dots, q$) เป็นน้ำหนักซินแนปส์ (Synapse weight) ระหว่างชั้นซ่อนเร้น (Hidden layer) และชั้นเอาต์พุต และ w_{ij} ($i = 0, 1, 2, \dots, p; j = 1, 2, \dots, q$) เป็นน้ำหนักซินแนปส์ระหว่างชั้นอินพุต (Input layer) และชั้นซ่อนเร้น โดยที่ p เป็นจำนวนโหนดของชั้นอินพุต และ q เป็นจำนวนโหนดของชั้นซ่อนเร้น ส่วนฟังก์ชัน $g()$ เป็นแอคติเวชันฟังก์ชัน (Activation function) ของชั้นซ่อนเร้น ซึ่งเป็นฟังก์ชันแบบซิกมอยด์ (Sigmoid) ดังสมการที่ 5. (Wang and Meng, 2012)

$$\text{sig}(x) = \frac{1}{(1+\exp(-x))} \quad \text{----- (5)}$$

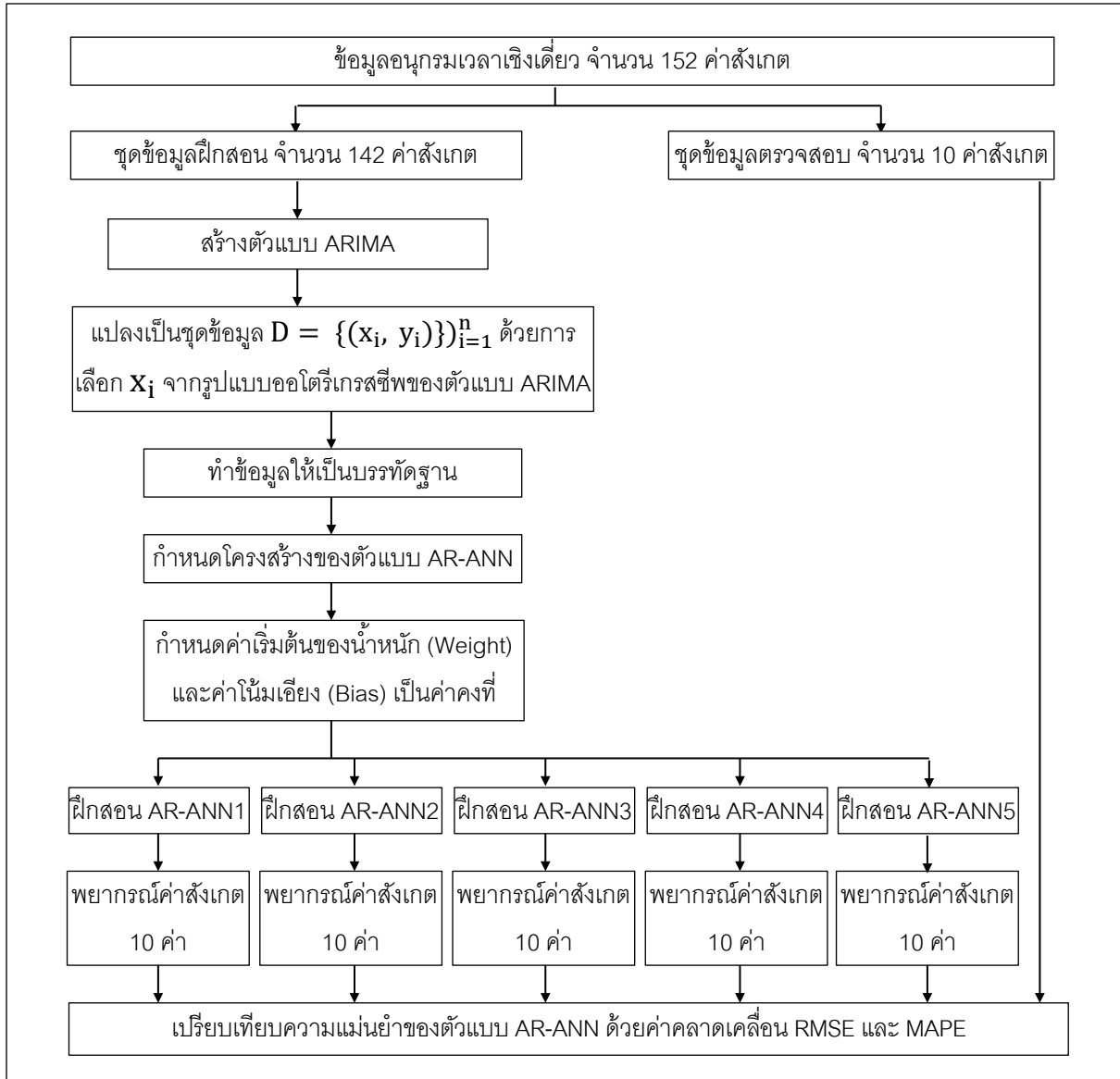
งานวิจัยนี้ นำเสนอเทคนิคใหม่ในการสร้างตัวแบบพยากรณ์อนุกรมเวลาเชิงเดี่ยว เรียกว่า ตัวแบบ AR-ANN โดยนำรูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA มาใช้ร่วมกับเทคนิคเครือข่ายประสาทเทียม ซึ่งรูปแบบออโตรีเกรสซีฟจะเป็นตัวกำหนดอินพุตของเครือข่ายประสาทเทียม และจากงานวิจัยของ Wang และ Meng (2012) กล่าวว่า เมื่อนำเครือข่ายประสาทเทียมมาใช้ในการพยากรณ์อนุกรมเวลา ควรมีชั้นซ่อนเร้นเพียงชั้นเดียว ประกอบกับงานวิจัยของ Cadenas และ Rivera (2010) ที่ได้พัฒนาตัวแบบพยากรณ์แบบผสมผสานระหว่าง ARIMA และ ANN โดยกำหนดโครงสร้างของ ANN ให้เรียบง่าย ด้วยการกำหนดนิวรอนในชั้นซ่อนเร้นเพียง 2 ถึง 3 นิวรอน ทั้งนี้ เพื่อป้องกันการเกิดปัญหาโอเวอร์ฟิตติง (Over-fitting problem) ดังนั้น งานวิจัยนี้จึงกำหนดโครงสร้างของตัวแบบ AR-ANN ที่กำหนดจำนวนนิวรอนในชั้นอินพุตจากรูปแบบออโตรีเกรสซีฟของ ARIMA และกำหนดจำนวนนิวรอนในชั้นซ่อนเร้นตั้งแต่ 1 ถึง 5 นิวรอน เพื่อหลีกเลี่ยงปัญหาโอเวอร์ฟิตติง จากการอ้างอิงงานวิจัยที่ได้กล่าวไว้ข้างต้นจำนวน 5 ตัวแบบ แล้วนำมาทดสอบกับชุดข้อมูลทั้ง 3 ชุด โดยมีโครงสร้างของตัวแบบ AR-ANN ที่ใช้ในการทดลอง ดังภาพที่ 3



ภาพที่ 3 โครงสร้างของตัวแบบ AR-ANN ที่ใช้ดำเนินการทดลอง

การสร้างตัวแบบ AR-ANN จะดำเนินการทดลองกับชุดข้อมูลที่ใช้ศึกษาจำนวน 3 ชุดข้อมูล ในลักษณะเดียวกัน โดยใช้โปรแกรม MatLab เป็นเครื่องมือ ซึ่งดำเนินการสร้างตัวแบบ AR-ANN จำนวน 5 ตัวแบบ แต่ละตัวแบบ มีความแตกต่างที่จำนวนนิเวรอนในชั้นซ่อนเร้น ตั้งแต่ 1 ถึง 5 นิเวรอน เรียกว่า ตัวแบบ AR-ANN1 คือ มีจำนวนนิเวรอนในชั้นซ่อนเร้น 1 นิเวรอน เรียงตามลำดับถึงตัวแบบ AR-ANN5 คือ มีจำนวนนิเวรอนในชั้นซ่อนเร้น 5 นิเวรอน โดยวัดค่าคลาดเคลื่อน RMSE และ MAPE ของแต่ละตัวแบบ เพื่อกำหนดโครงสร้างที่เหมาะสมที่สุดของตัวแบบ AR-ANN สำหรับนำไปเปรียบเทียบความแม่นยำกับตัวแบบ ARIMA ดังนั้น ขั้นตอนการสร้างตัวแบบ AR-ANN ที่มีโครงสร้างที่เหมาะสมที่สุดของงานวิจัยนี้ เป็นดังภาพที่ 4 ซึ่งสามารถอธิบายแต่ละขั้นตอนได้ ดังนี้

1. จากข้อมูลอนุกรมเวลา แบ่งข้อมูลเป็นชุดข้อมูลฝึกสอน และชุดข้อมูลตรวจสอบเช่นเดียวกับการสร้างตัวแบบพยากรณ์ด้วยวิธีการวิเคราะห์อนุกรมเวลาเชิงเดี่ยวแบบดั้งเดิม
2. สร้างตัวแบบ ARIMA จากชุดข้อมูลฝึกสอน
3. แปลงชุดข้อมูลฝึกสอนให้อยู่ในรูปแบบของข้อมูล $D = \{(x_i, y_i)\}_{i=1}^n$ ด้วยการใช้การวิเคราะห์รูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA
4. ทำข้อมูลให้เป็นบรรทัดฐาน (Data normalization) มีค่าระหว่าง 0 ถึง 1
5. กำหนดโครงสร้างของตัวแบบ AR-ANN จำนวน 5 ตัวแบบ ด้วยฟังก์ชันชื่อ "newff()" ที่อยู่ในส่วน Library ของโปรแกรม Matlab
6. กำหนดค่าเริ่มต้นของน้ำหนัก (Weight) และค่าโน้มเอียง (Bias) เป็นค่าคงที่ แทนการสุ่มค่า เพื่อให้เอาท์พุทของตัวแบบ AR-ANN ลู่เข้าสู่ค่าตอบได้ดีขึ้น โดยค่าที่กำหนด ได้จากการทดลองสุ่ม ที่ทำให้ค่าคลาดเคลื่อนมีค่าน้อยที่สุด
7. ฝึกสอนตัวแบบ AR-ANN จำนวน 5 ตัวแบบ ตามโครงสร้างที่กำหนดไว้ในข้อ 5 โดยใช้ข้อมูลที่ผ่านมาทำให้เป็นบรรทัดฐานของข้อ 4 สำหรับฝึกสอนตัวแบบ AR-ANN ทั้ง 5 ตัวแบบ
8. นำตัวแบบ AR-ANN ทั้ง 5 ตัวแบบ พยากรณ์ค่าสังเกตล่วงหน้า 10 คาบเวลา จากนั้น ทำการดีนอร์มอลไลซ์ (Denormalize) แปลงข้อมูลกลับเป็นค่าปกติ ซึ่งเป็นค่าสังเกตของวันที่ 22 ถึงวันที่ 31 พฤษภาคม 2559 เพื่อนำไปวัดค่าคลาดเคลื่อน RMSE และ MAPE
9. เลือกตัวแบบ AR-ANN ที่เหมาะสมที่สุด โดยพิจารณาจากค่าคลาดเคลื่อน RMSE และ MAPE สำหรับนำไปพยากรณ์ค่าสังเกตล่วงหน้า เพื่อเปรียบเทียบความคลาดเคลื่อนในการพยากรณ์กับตัวแบบ ARIMA



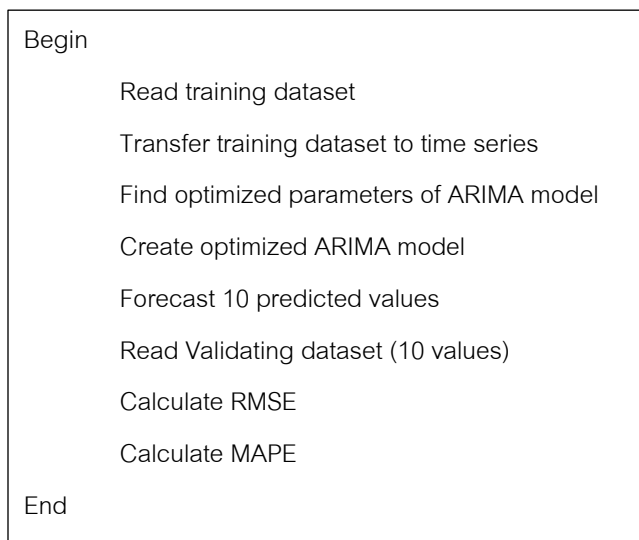
ภาพที่ 4 ขั้นตอนการสร้างตัวแบบ AR-ANN ที่มีโครงสร้างเหมาะสมที่สุด

ผลการวิจัยและวิจารณ์ผล

งานวิจัยนี้ ดำเนินการศึกษากับชุดข้อมูลจำนวน 3 ชุด ได้แก่ 1) ชุดข้อมูลของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่ 2) ชุดข้อมูลของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง และ 3) ชุดข้อมูลของ ต.ในเมือง อ.เมือง จ.ลำพูน ซึ่งข้อมูลทั้ง 3 ชุด มีลักษณะเป็นอนุกรมเวลาเชิงเดี่ยว ที่มีค่าสังเกตเป็นค่าเฉลี่ยความเข้มข้น 24 ชั่วโมง ของปริมาณ PM10 ในอากาศ และมีคาบเวลาเป็นรายวัน โดยมีจำนวนค่าสังเกตของแต่ละชุดข้อมูล เท่ากับ 152 ค่าสังเกต ตั้งแต่วันที่ 1 ม.ค. ถึง 31 พ.ค. 2559 ซึ่งแบ่งข้อมูลจำนวน 142 ค่าแรก เป็นชุดข้อมูลฝึกสอน สำหรับสร้างตัวแบบพยากรณ์ และใช้ข้อมูล 10 ค่าหลัง ของวันที่ 22 ถึงวันที่ 31 พ.ค. 2559 เป็นชุดข้อมูลตรวจสอบ สำหรับวัดความแม่นยำของค่าพยากรณ์ ในขั้นตอนประเมินผลการทดลอง โดยดำเนินการวิจัยกับชุดข้อมูลทั้ง 3 ชุด ตามกรอบแนวคิดงานวิจัย ด้วยขั้นตอนเดียวกันทั้ง 3 ชุด และได้ผลการทดลอง สำหรับข้อมูลทั้ง 3 ชุด ดังนี้

1. ผลการทดลองของตัวแบบ ARIMA

เมื่อดำเนินการทดลองสร้างตัวแบบ ARIMA กับข้อมูลทั้ง 3 ชุด ตามขั้นตอนการสร้างตัวแบบ ARIMA โดยใช้รหัสเทียม (Psudo code) ดังภาพที่ 5 ผลการทดลอง พบว่า รูปแบบ ARIMA ที่เหมาะสมของแต่ละชุดข้อมูล ซึ่งได้จากการใช้คำสั่ง auto.arima() ที่มีการวัดค่า AIC (Akaikie Information Criterion) โดยประมาณความแปรปรวนของข้อสนเทศคูลส์แบ็ค-ไลท์เบอร์ (Kullback Leibler Information) ระหว่างรูปแบบของตัวแบบที่แท้จริงกับรูปแบบของตัวแบบที่เหมาะสม ซึ่งรูปแบบที่เหมาะสมที่สุด คือ รูปแบบที่ให้ค่า AIC ต่ำที่สุด โดยรูปแบบ ARIMA ที่เหมาะสมของข้อมูลทั้ง 3 ชุด สรุปได้ดังตารางที่ 1



ภาพที่ 5 รหัสเทียมในการสร้างและวัดค่าคลาดเคลื่อนของตัวแบบ ARIMA

ตารางที่ 1 รูปแบบ ARIMA ที่เหมาะสมของแต่ละชุดข้อมูล

ชุดข้อมูล	รูปแบบ ARIMA ที่เหมาะสม
ชุดข้อมูลที่ 1 ของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่	ARIMA(2, 1, 2)
ชุดข้อมูลที่ 2 ของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง	ARIMA(0, 1, 2)
ชุดข้อมูลที่ 3 ของ ต.ในเมือง อ.เมือง จ.ลำพูน	ARIMA(1, 0, 2)

เมื่อได้รูปแบบ ARIMA ที่เหมาะสมของแต่ละชุดข้อมูล จึงนำมาสร้างตัวแบบ ARIMA สำหรับพยากรณ์ค่าสังเกต 10 ค่า ตั้งแต่วันที่ 22 ถึงวันที่ 31 พ.ค. 2559 จากนั้น นำไปเปรียบกับค่าสังเกตจริงในช่วงเวลาเดียวกันของชุดข้อมูลตรวจสอบ เพื่อวัดค่า RMSE และ MAPE ได้ค่าคลาดเคลื่อน RMSE และ MAPE ของแต่ละชุดข้อมูล ดังตารางที่ 2

ตารางที่ 2 ค่าคลาดเคลื่อน RMSE และ MAPE จากตัวแบบ ARIMA ของแต่ละชุดข้อมูล

ชุดข้อมูล	RMSE	MAPE
ชุดข้อมูลที่ 1 ของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่	7.8841	22.3924
ชุดข้อมูลที่ 2 ของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง	20.5247	27.9511
ชุดข้อมูลที่ 3 ของ ต.ในเมือง อ.เมือง จ.ลำพูน	18.9031	67.5820

2. ผลการทดลองของตัวแบบ AR-ANN

เนื่องจากขั้นตอนการสร้างตัวแบบ AR-ANN ได้นำรูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA มากำหนดเป็นอินพุตให้กับตัวแบบ AR-ANN และจากผลการทดลองของตัวแบบ ARIMA ทำให้ทราบรูปแบบ ARIMA ที่เหมาะสมของแต่ละชุดข้อมูล และสามารถวิเคราะห์รูปแบบออโตรีเกรสซีฟของแต่ละชุดข้อมูล ด้วยสมการที่ 1 ได้ดังนี้

1. ชุดข้อมูลที่ 1 ของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่ มีรูปแบบ ARIMA(2, 1, 2) ดังนั้น แทนค่าอันดับ $p=2$ และ $d=2$ ลงในสมการที่ 1 โดยไม่สนใจอันดับ q เนื่องจากเป็นส่วนของรูปแบบมูฟวิงเอเวอเรจ (Moving average) ได้ว่า

$$\begin{aligned} (1 - \theta_1 B - \theta_2 B^2)(1 - B)Y_t &= \text{Moving average term} \\ (1 - \theta_1 B - \theta_2 B^2 - B + \theta_1 B^2 + \theta_2 B^3)Y_t &= \text{Moving average term} \\ Y_t &= (1 + \theta_1)Y_{t-1} + (\theta_2 - \theta_1)Y_{t-2} - \theta_2 Y_{t-3} + \text{Moving average term} \end{aligned} \quad \text{----- (6)}$$

จากสมการที่ 6 สรุปได้ว่า อินพุตของตัวแบบ AR-ANN สำหรับชุดข้อมูลที่ 1 คือ ค่าล่าหลังที่เวลา $t-1$, $t-2$ และ $t-3$ (Y_{t-1} , Y_{t-2} , Y_{t-3})

2. ชุดข้อมูลที่ 2 ของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง มีรูปแบบ ARIMA(0, 1, 2) ดังนั้น แทนค่าอันดับ $p=0$ และ $d=1$ ลงในสมการที่ 1 โดยไม่สนใจอันดับ q เนื่องจากเป็นส่วนของรูปแบบมูฟวิงเอเวอเรจ ได้ว่า

$$\begin{aligned} (1 - B)Y_t &= \text{Moving average term} \\ Y_t &= Y_{t-1} + \text{Moving average term} \end{aligned} \quad \text{----- (7)}$$

จากสมการที่ 7 สรุปได้ว่า อินพุตของตัวแบบ AR-ANN สำหรับชุดข้อมูลที่ 2 คือ ค่าล่าหลังที่เวลา $t-1$ (Y_{t-1})

3. ชุดข้อมูลที่ 3 ของ ต.ในเมือง อ.เมือง จ.ลำพูน มีรูปแบบ ARIMA(1, 0, 2) ดังนั้น แทนค่าอันดับ $p=1$ และ $d=0$ ลงในสมการที่ 1 โดยไม่สนใจอันดับ q เนื่องจากเป็นส่วนของรูปแบบมูฟวิงเอเวอเรจ ได้ว่า

$$\begin{aligned} (1 - \theta_1 B)Y_t &= \text{Moving average term} \\ Y_t &= \theta_1 Y_{t-1} + \text{Moving average term} \end{aligned} \quad \text{----- (8)}$$

จากสมการที่ 8 สรุปได้ว่า อินพุตของตัวแบบ AR-ANN สำหรับชุดข้อมูลที่ 3 คือ ค่าล่าหลังที่เวลา $t-1$ (Y_{t-1})

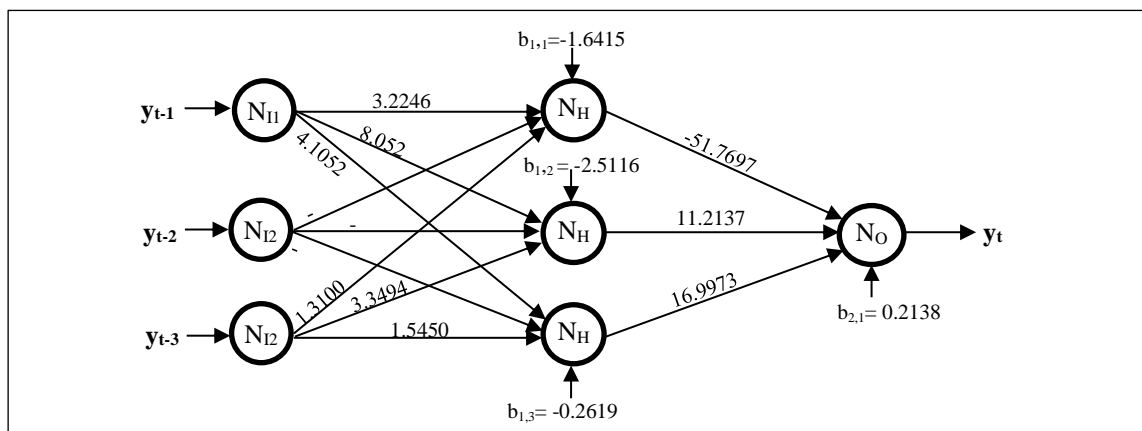
หลังจากทราบอินพุตของตัวแบบ AR-ANN ของแต่ละชุดข้อมูล จึงนำไปสร้างตัวแบบ AR-ANN จำนวน 5 ตัวแบบ จากนั้น วัดค่าคลาดเคลื่อน RMSE และ MAPE เพื่อกำหนดโครงสร้างของตัวแบบ AR-ANN ที่เหมาะสมที่สุด ตามขั้นตอนการสร้างตัวแบบ AR-ANN โดยใช้โปรแกรม Matlab สามารถสรุปผลได้ดังตารางที่ 3

ตารางที่ 3 ค่าคลาดเคลื่อน RMSE และ MAPE จากตัวแบบ AR-ANN ของ 3 ชุดข้อมูล

ชุดข้อมูล	ตัวแบบพยากรณ์	RMSE	MAPE
ชุดข้อมูลที่ 1 ของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่	AR-ANN5	8.4004	21.2296
	AR-ANN4	7.5277	18.5945
	AR-ANN3	4.7727	11.8080
	AR-ANN2	5.8424	17.1497
	AR-ANN1	5.7013	17.7842
ชุดข้อมูลที่ 2 ของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง	AR-ANN5	13.8751	20.1108
	AR-ANN4	12.6597	19.2211
	AR-ANN3	12.7777	19.2477
	AR-ANN2	13.0932	19.6222
	AR-ANN1	12.9322	19.4397
ชุดข้อมูลที่ 3 ของ ต.ในเมือง อ.เมือง จ.ลำพูน	AR-ANN5	7.7698	24.7266
	AR-ANN4	7.8670	30.1251
	AR-ANN3	9.3014	35.2303
	AR-ANN2	9.3051	35.2514
	AR-ANN1	9.2787	35.1152

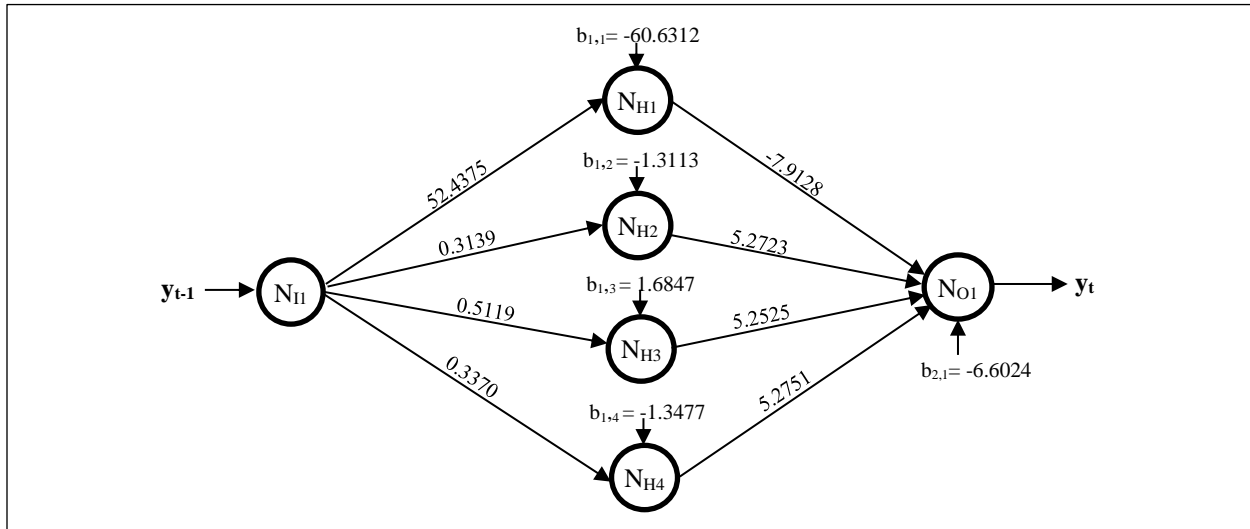
จากตารางที่ 3 โครงสร้างที่เหมาะสมของตัวแบบ AR-ANN สำหรับแต่ละชุดข้อมูล สรุปได้ดังนี้

- ชุดข้อมูลที่ 1 ของจังหวัดเชียงใหม่ ได้ตัวแบบ AR-ANN3 ที่มีจำนวนนิวรอนในชั้นซ่อนเร้นเท่ากับ 3 นิวรอน เป็นตัวแบบพยากรณ์ที่ให้ค่าคลาดเคลื่อน RMSE และ MAPE ต่ำที่สุด ดังนั้น ตัวแบบ AR-ANN สำหรับชุดข้อมูลที่ 1 เป็นดังภาพที่ 6



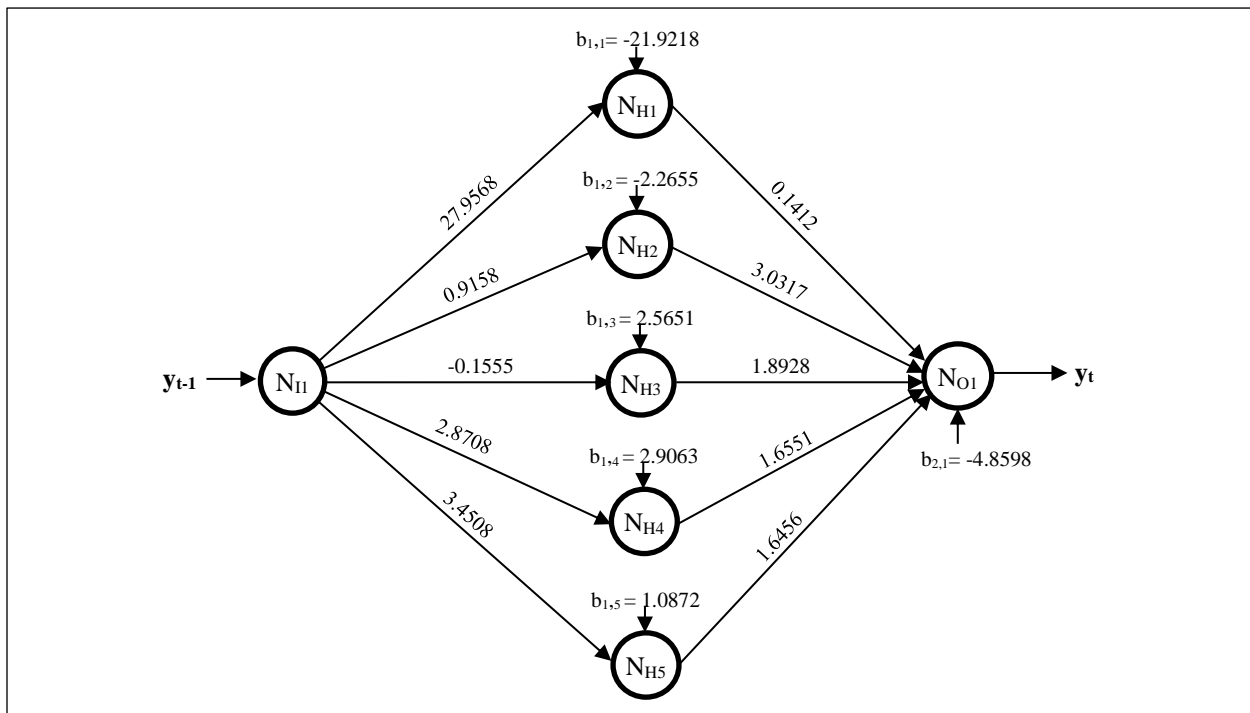
ภาพที่ 6 ตัวแบบ AR-ANN ที่มีโครงสร้างเหมาะสมที่สุดของชุดข้อมูลที่ 1 จังหวัดเชียงใหม่

2. ชุดข้อมูลที่ 2 ของจังหวัดลำปาง ได้ตัวแบบ AR-ANN4 ที่มีจำนวนนิวรอนในชั้นซ่อนเร้นเท่ากับ 4 นิวรอน เป็นตัวแบบพยากรณ์ที่ให้ค่าคลาดเคลื่อน RMSE และ MAPE ต่ำที่สุด ดังนั้น ตัวแบบ AR-ANN สำหรับชุดข้อมูลที่ 2 เป็นดังภาพที่ 7



ภาพที่ 7 ตัวแบบ AR-ANN ที่มีโครงสร้างเหมาะสมที่สุดของชุดข้อมูลที่ 2 จังหวัดลำปาง

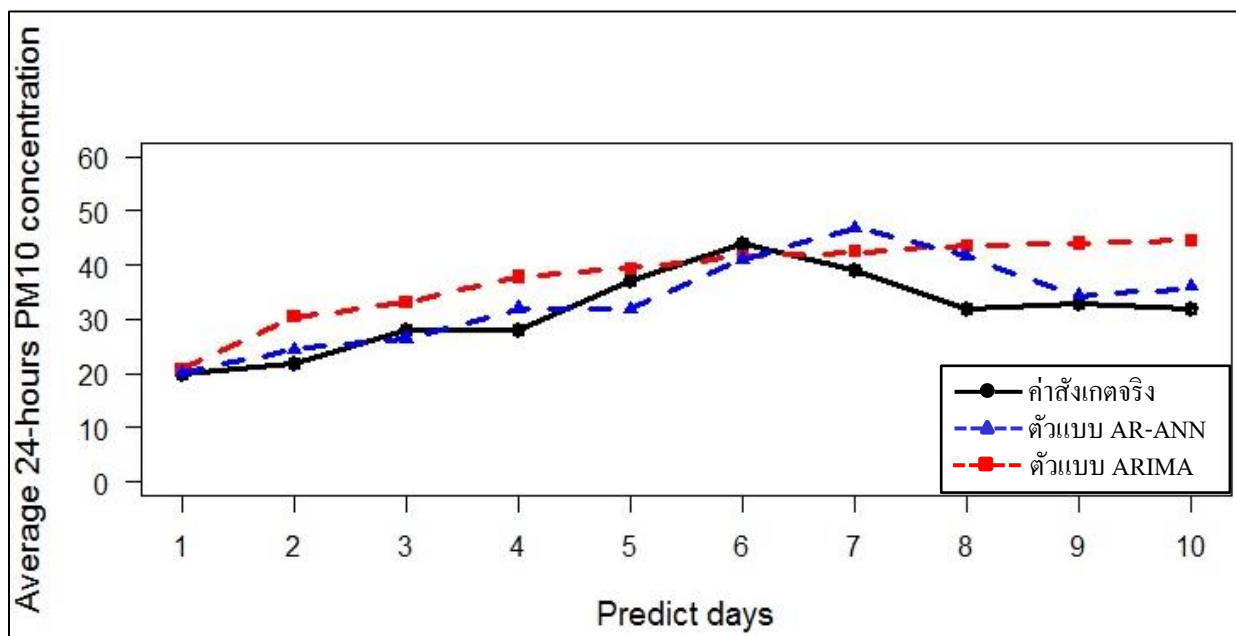
3. ชุดข้อมูลที่ 3 ของจังหวัดลำพูน ได้ตัวแบบ AR-ANN5 ที่มีจำนวนนิวรอนในชั้นซ่อนเร้นเท่ากับ 5 นิวรอน เป็นตัวแบบพยากรณ์ที่ให้ค่าคลาดเคลื่อน RMSE และ MAPE ต่ำที่สุด ดังนั้น ตัวแบบ AR-ANN สำหรับชุดข้อมูลที่ 3 เป็นดังภาพที่ 8



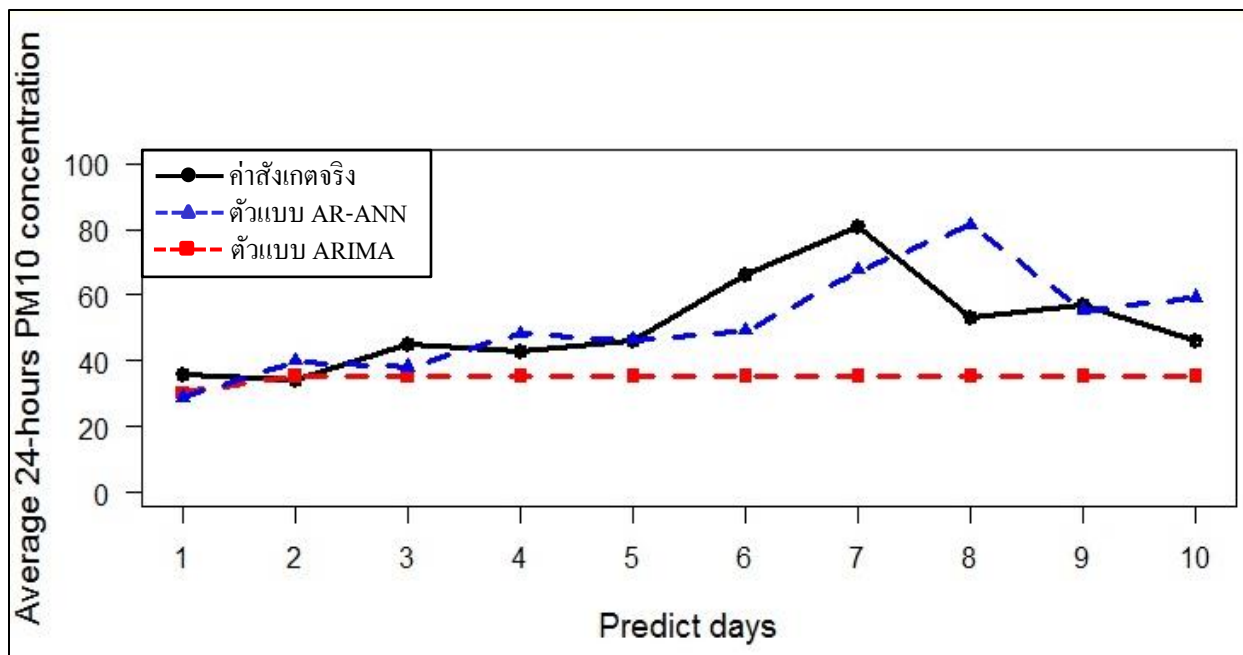
ภาพที่ 8 ตัวแบบ AR-ANN ที่มีโครงสร้างเหมาะสมที่สุดของชุดข้อมูลที่ 3 จังหวัดลำพูน

3. การเปรียบเทียบความคลาดเคลื่อนของค่าพยากรณ์ ระหว่างตัวแบบ AR-ANN และตัวแบบ ARIMA

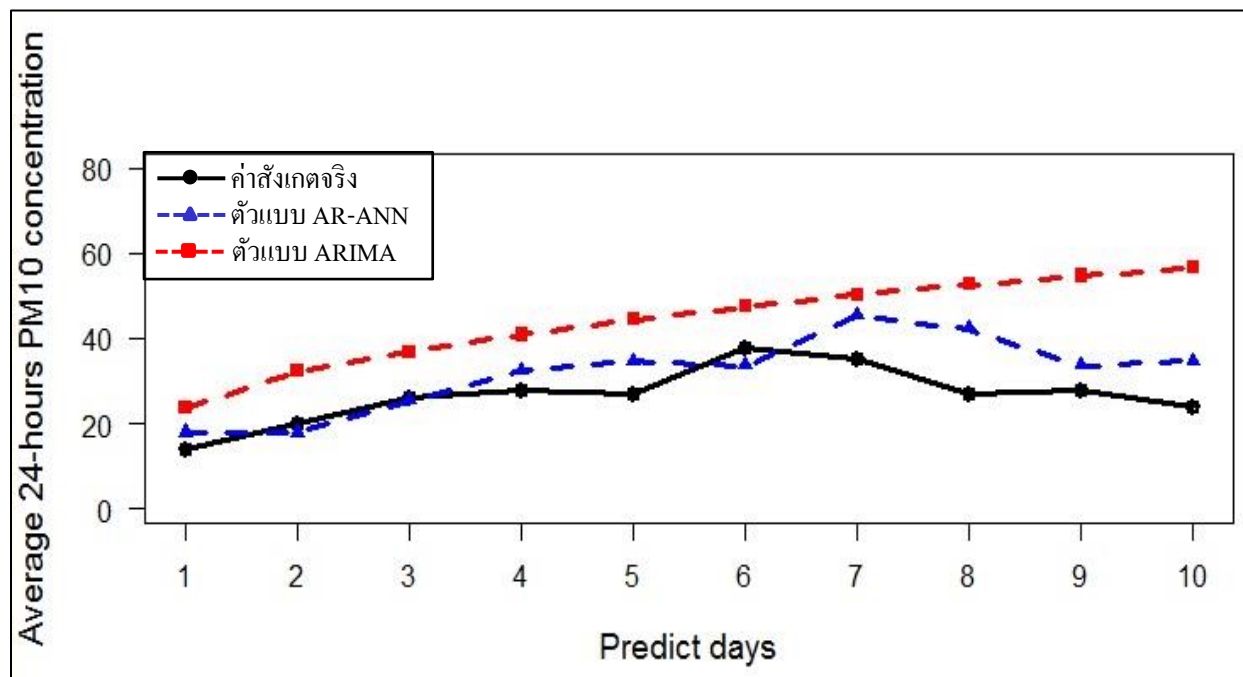
จากการนำเสนอวิธีการสร้างตัวแบบ AR-ANN เพื่อลดความคลาดเคลื่อนของค่าพยากรณ์ เมื่อเปรียบเทียบกับตัวแบบ ARIMA ภายใต้สมมติฐานว่า ตัวแบบ AR-ANN จะให้ความคลาดเคลื่อนของค่าพยากรณ์ต่ำกว่าตัวแบบ ARIMA เนื่องจากสามารถตรวจจับรูปแบบเชิงเส้นและรูปแบบไม่เป็นเชิงเส้น ได้ทั้ง 2 รูปแบบ โดยดำเนินการทดลองกับชุดข้อมูล 3 ชุด แต่ละชุดมีลักษณะเป็นอนุกรมเวลาเชิงเดี่ยว ที่มีค่าสังเกต คือ ค่าเฉลี่ย 24 ชั่วโมงของความเข้มข้น PM10 ในอากาศ ผลการทดลองพบว่า ตัวแบบ AR-ANN ที่มีโครงสร้างเหมาะสมที่สุด ให้ค่าคลาดเคลื่อนต่ำกว่าตัวแบบ ARIMA ทั้ง 3 ชุดข้อมูล ซึ่งแสดงได้ดังตารางที่ 4 เมื่อเปรียบเทียบค่า MAPE ของตัวแบบ AR-ANN และตัวแบบ ARIMA พบว่า ตัวแบบ AR-ANN มีค่าความคลาดเคลื่อน RMSE และ MAPE ต่ำกว่าตัวแบบ ARIMA ทั้ง 3 ชุดข้อมูล โดยเฉพาะ ชุดข้อมูลที่ 3 ของ จ. ลำพูน ที่ตัวแบบ AR-ANN สามารถลดค่าคลาดเคลื่อนได้ 63.41% นอกจากนี้ เมื่อนำค่าพยากรณ์ของตัวแบบ AR-ANN มาสร้างกราฟเส้นเปรียบเทียบกับกราฟค่าพยากรณ์ของตัวแบบ ARIMA และกราฟของค่าสังเกตจริง พบว่า กราฟค่าพยากรณ์ของตัวแบบ AR-ANN มีการเคลื่อนไหวของเส้นกราฟใกล้เคียงกับกราฟของค่าสังเกตจริง มากกว่ากราฟของตัวแบบ ARIMA ทั้ง 3 ชุดข้อมูล ดังภาพที่ 9-11 ดังนั้น จากผลการทดลองที่ได้ สรุปได้ว่า ชุดข้อมูลที่ศึกษามีส่วนผสมของรูปแบบที่เป็นเชิงเส้นและรูปแบบที่ไม่เป็นเชิงเส้น เมื่อสร้างตัวแบบ ARIMA ที่สามารถตรวจจับรูปแบบเชิงเส้นได้ดี แต่ตรวจจับรูปแบบไม่เป็นเชิงเส้นได้ไม่ดี จึงให้ค่าคลาดเคลื่อนสูงกว่า เมื่อเปรียบเทียบกับตัวแบบ AR-ANN ที่สามารถตรวจจับรูปแบบเชิงเส้นและรูปแบบไม่เป็นเชิงเส้นได้ทั้ง 2 รูปแบบ ซึ่งผลการทดลองจากงานวิจัยนี้ สอดคล้องกับงานวิจัยในอดีต (Wang and Meng, 2012) ดังนั้น ตัวแบบ AR-ANN สามารถลดความคลาดเคลื่อนของค่าพยากรณ์เมื่อเปรียบเทียบกับตัวแบบ ARIMA และสามารถนำไปใช้พยากรณ์ค่าเฉลี่ย 24 ชั่วโมงของความเข้มข้น PM10 ในอากาศ สำหรับเขตพื้นที่ภาคเหนือของประเทศไทย



ภาพที่ 9 กราฟค่าพยากรณ์ของชุดข้อมูลที่ 1 จังหวัดเชียงใหม่



ภาพที่ 10 กราฟค่าพยากรณ์ของชุดข้อมูลที่ 2 จังหวัดลำปาง



ภาพที่ 11 กราฟค่าพยากรณ์ของชุดข้อมูลที่ 3 จังหวัดลำพูน

ตารางที่ 4 การเปรียบเทียบค่าคลาดเคลื่อน RMSE และ MAPE ของตัวแบบ AR-ANN และตัวแบบ ARIMA

ชุดข้อมูล	ตัวแบบ ARIMA		ตัวแบบ AR-ANN		%เปรียบเทียบ ความคลาดเคลื่อน
	RMSE	MAPE	RMSE	MAPE	
ชุดข้อมูลที่ 1 ของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่	7.8841	22.3924	4.7727	11.8080	-47.26%
ชุดข้อมูลที่ 2 ของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง	20.5247	27.9511	12.6597	19.2211	-31.23%
ชุดข้อมูลที่ 3 ของ ต.ในเมือง อ.เมือง จ.ลำพูน	18.9031	67.5820	7.7698	24.7266	-63.41%

* หมายเหตุ: %เปรียบเทียบความคลาดเคลื่อน = $100\% \times (\text{ค่า MAPE ของ AR-ANN} - \text{ค่า MAPE ของ ARIMA}) / \text{ค่า MAPE ของ ARIMA}$

สรุปผลการวิจัย

งานวิจัยนี้ ศึกษาการพัฒนาตัวแบบผสมผสานระหว่างรูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA กับเทคนิคเครือข่ายประสาทเทียม เรียกว่า ตัวแบบ AR-ANN เพื่อพยากรณ์ค่าความเข้มข้นของ PM10 ในเขตพื้นที่ภาคเหนือของประเทศไทย โดยดำเนินการวิจัยกับชุดข้อมูล ที่มีลักษณะเป็นอนุกรมเวลาเชิงเดี่ยว ซึ่งมีค่าสังเกต 1 ค่า คือ ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมง ของ PM10 บันทึกค่าตามลำดับเวลา ด้วยคาบเวลาเป็นรายวัน ตั้งแต่วันที่ 1 มกราคม ถึงวันที่ 31 พฤษภาคม 2559 จำนวน 152 ค่าสังเกต จำนวน 3 ชุดข้อมูล ได้แก่ 1) ชุดข้อมูลของ ต.ศรีภูมิ อ.เมือง จ.เชียงใหม่ 2) ชุดข้อมูลของ ต.บ้านดง อ.แม่เมาะ จ.ลำปาง และ 3) ชุดข้อมูลของ ต.ในเมือง อ.เมือง จ.ลำพูน ซึ่งการสร้างตัวแบบพยากรณ์ จะดำเนินการสร้างตัวแบบ ARIMA และตัวแบบ AR-ANN กับชุดข้อมูลทั้งหมด ด้วยขั้นตอนเดียวกัน ตามกรอบแนวคิดงานวิจัย โดยแบ่งข้อมูลเป็นชุดข้อมูลฝึกสอน จำนวน 142 ค่า สำหรับสร้างตัวแบบพยากรณ์ทั้ง 2 ตัวแบบ และแบ่งเป็นชุดข้อมูลตรวจสอบ จำนวน 10 ค่า สำหรับประเมินผลการทดลอง ด้วยการวัดค่าคลาดเคลื่อน RMSE และ MAPE ผลการทดลอง พบว่า รูปแบบ ARIMA ที่เหมาะสมของชุดข้อมูลที่ 1 คือ ARIMA(2, 1, 2) รูปแบบ ARIMA ที่เหมาะสมของชุดข้อมูลที่ 2 คือ ARIMA(0, 1, 2) และรูปแบบ ARIMA ที่เหมาะสมของชุดข้อมูลที่ 3 คือ ARIMA(1, 0, 2) เมื่อนำรูปแบบออโตรีเกรสซีฟของตัวแบบ ARIMA ไปสร้างเป็นตัวแบบผสมผสาน คือ ตัวแบบ AR-ANN สามารถลดค่าคลาดเคลื่อนของค่าพยากรณ์ได้ทั้ง 3 ชุดข้อมูล โดยเปรียบเทียบกับค่าคลาดเคลื่อน MAPE ของตัวแบบ AR-ANN กับตัวแบบ ARIMA ซึ่งชุดข้อมูลที่ 1 ตัวแบบ AR-ANN ลดค่าคลาดเคลื่อนได้ 47.26% ชุดข้อมูลที่ 2 ลดค่าคลาดเคลื่อนได้ 31.23% และชุดข้อมูลที่ 3 ลดค่าคลาดเคลื่อนได้ 63.41% ผลการทดลองที่ได้รับ สอดคล้องกับงานวิจัยในอดีต จึงสรุปได้ว่า ตัวแบบ AR-ANN ที่งานวิจัยนี้ได้นำเสนอ สามารถลดค่าคลาดเคลื่อนของค่าพยากรณ์ เมื่อเปรียบเทียบกับตัวแบบ ARIMA ได้อย่างน่าพอใจ และสามารถนำไปใช้พยากรณ์ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมง ของ PM10 ในเขตพื้นที่ภาคเหนือของประเทศไทย

ข้อเสนอแนะ

งานวิจัยนี้ นำเสนอวิธีการพัฒนาตัวแบบพยากรณ์ด้วยการนำวิธีการทางสถิติแบบดั้งเดิม คือ ตัวแบบ ARIMA ผสมผสานกับวิธีการด้านแมชชีนเลิร์นนิง คือ ตัวแบบโครงข่ายประสาทเทียม (ANN) เรียกว่า ตัวแบบ AR-ANN จากนั้นเปรียบเทียบความแม่นยำกับตัวแบบ ARIMA เพื่อแสดงให้เห็นว่า การนำวิธีการพัฒนาตัวแบบพยากรณ์ 2 วิธี มาผสมผสานกัน สามารถให้ความแม่นยำในการพยากรณ์ได้มากกว่าตัวแบบพยากรณ์ที่พัฒนาจากวิธีการทางสถิติแบบดั้งเดิมเพียงวิธีเดียว อย่างไรก็ตาม ตัวแบบพยากรณ์ AR-ANN ของงานวิจัยนี้ อาจไม่ใช่ตัวแบบพยากรณ์ที่สามารถพยากรณ์ค่าความเข้มข้นเฉลี่ย 24 ชั่วโมง ของ PM10 ได้แม่นยำมากที่สุด เนื่องจากมีเทคนิคอื่น ๆ ที่ใช้พัฒนาตัวแบบพยากรณ์ได้ เช่น เครือข่ายประสาทเทียม

(Support Vector Machine: SVM) ซึ่งเป็นวิธีการด้านแมชชีนเลิร์นนิง ดังนั้น งานวิจัยในอนาคต อาจนำเทคนิคอื่น ๆ นอกเหนือจากที่ใช้ในงานวิจัยนี้ มาใช้ในการดำเนินงานวิจัยต่อไป

กิตติกรรมประกาศ

ขอขอบคุณกรมควบคุมมลพิษ ที่เผยแพร่ข้อมูลค่าความเข้มข้นเฉลี่ย 24 ชั่วโมงของ PM10 ในเขตพื้นที่ภาคเหนือของประเทศไทย โดยเป็นข้อมูลที่งานวิจัยนี้ได้นำมาศึกษาและดำเนินงานวิจัย

เอกสารอ้างอิง

- Antanasijevic, D.Z., Pocajt, V.V., Povrenovic, D.S., Ristic, M.D., and Peric-Grujic, A.A. (2013). PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization. *Science of the Total Environment*, 443, 511–519.
- Box, G.E.P., and Jenkins, G. (1990). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- Cadenas, E., and Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renewable Energy*, 35, 2732-2738.
- Department of Health. (2015). The surveillance areas at risk from air pollution in case of PM10. Retrieved October 4, 2016, from http://hia.anamai.moph.go.th/more_news.php?cid=317&filename=hia_book_2. (in Thai)
- Grivas, G., and Chaloulakou, A. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmospheric Environment*, 40, 1216–1229.
- He, H.D., Lu, W.Z., and Xue, Y. (2014). Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm. *Building and Environment*, 78, 111-117.
- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., and Brasseur, O. (2005). A neural network forecast for daily average PM10 concentrations in Belgium. *Atmospheric Environment*, 39, 3279–3289.
- Khanthongkham, W., Charoensiri, T., and Sophonpimol, C. (2016). Forecasting of Broiler Price. *Burapha Science Journal*, 21(1), 100-109.
- Kumar, S. (2004). *Neural Networks: A Classroom Approach*. New Delhi: Tata McGraw-Hill Education.
- Lee, Y.S., and Tong, L.I. (2011). Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems*, 24,66-72.
- Hyndman, R.J. (2015). Forecasting Functions for Time Series and Linear Models. Retrieved November 22, 2015, from <http://github.com/robjhyndman/forecast>.
- Perez, P. (2012). Combined model for PM10 forecasting in a large city. *Atmospheric Environment*, 60, 271-276.
- Pollution Control Department. (2016). Air Quality Index. Retrieved October 4, 2016, from http://aqmthai.com/aqi_info.php. (in Thai)

- Pollution Control Department. (2016). Data Archives for Air and Noise Pollution. Retrieved October 4, 2016, from <http://aqnis.pcd.go.th/>. (in Thai)
- Thanapala, D., Charoensiri, T., and Sophonpimol, C. (2016). Forecasting of Factory Pineapple Prices with Box-Jenkins Method. *Burapha Science Journal*, 21(1), 110-118.
- Wang, X., and Meng, M. (2012). A Hybrid Neural Network and ARIMA Model for Energy Consumption Forecasting. *JOURNAL OF COMPUTERS*, 7(5), 1184-1190.
- Wang, Y., Wang, J., Zhao, G., and Dong, Y. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy*, 48, 284-294.
- Wongsathan, R., and Seedadan, I. (2016). A hybrid ARIMA and Neural Networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area. *Procedia Computer Science*, 86, 273 – 276.