

การพยากรณ์ความต้องการใช้งานหน่วยจำหน่ายไฟฟ้าด้วย ซัพพอร์ตเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน

The Demand Forecasting of Electrical Power Distribution Units by Support Vector Regression with 3 Folds Cross-Validation

รณชัย ชื่นธวัช*, กิตติศักดิ์ เกิดประสพ และนิตยา เกิดประสพ

สาขาวิชาวิศวกรรมคอมพิวเตอร์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี จ.นครราชสีมา 30000

*Email: c_ronnachai@hotmail.com

บทคัดย่อ

การพยากรณ์ความต้องการใช้ไฟฟ้าได้อย่างแม่นยำ จะทำให้เกิดการวางแผนผลิตกระแสไฟฟ้า ได้อย่างมีประสิทธิภาพ ส่งผลให้ต้นทุนการผลิตกระแสไฟฟ้านลดลง และค่าใช้ไฟฟ้าของผู้บริโภคลดลงได้ ดังนั้น การวิจัยนี้ จึงมีวัตถุประสงค์เพื่อพยากรณ์หน่วยจำหน่ายไฟฟ้า ที่สะท้อนถึงความต้องการใช้ไฟฟ้า ด้วยการสร้างตัวแบบการพยากรณ์จากเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน (Support Vector Regression: SVR with 3 Folds Cross-Validation) จากนั้น เปรียบเทียบความแม่นยำของการพยากรณ์กับตัวแบบการถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression: MLR) และตัวแบบ ARIMA (Autoregressive Integrated Moving Average) ซึ่งข้อมูลที่นำมาวิเคราะห์ มีลักษณะเป็นอนุกรมเวลา โดยเป็นข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง (กฟน.) ที่เก็บรวบรวมข้อมูลจากรายงานสถานการณ์การจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง ตั้งแต่เดือนมกราคม ปีพ.ศ. 2553 ถึงเดือนพฤษภาคม ปีพ.ศ. 2558 และข้อมูลที่สนใจ มีขอบเขตเฉพาะจำนวนหน่วยจำหน่ายไฟฟ้าของผู้บริโภคประเภทบ้านอยู่อาศัย โดยนำข้อมูลตั้งแต่เดือนมกราคม ปีพ.ศ. 2553 ถึงเดือนธันวาคม ปีพ.ศ. 2557 ไปสร้างตัวแบบการพยากรณ์ด้วยภาษาอาร์ จากนั้น ใช้ข้อมูลตรวจสอบ ตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม ปีพ.ศ. 2558 เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ของตัวแบบทั้ง 3 ด้วยการวัดค่าคลาดเคลื่อน ได้แก่ ค่า Root Mean Square Error (RMSE) และค่า Mean Absolute Percentage Error (MAPE) จากผลการทดลอง พบว่า ตัวแบบ SVR เป็นตัวแบบที่ให้ค่า RMSE และค่า MAPE ต่ำที่สุด ดังนั้น ตัวแบบ SVR เป็นตัวแบบที่เหมาะสมที่สุดสำหรับนำมาใช้พยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวงกับผู้บริโภคประเภทบ้านอยู่อาศัย

คำสำคัญ: ซัพพอร์ตเวกเตอร์รีเกรสชัน การตรวจสอบสลับ การถดถอยเชิงเส้นแบบพหุ ตัวแบบอาร์มา อนุกรมเวลาการพยากรณ์ความต้องการใช้งานไฟฟ้า

Abstract

The accurate forecasting of electricity usage makes the efficiently planing about generating electricity, will reduce the electricity generating cost, moreover the electricity usage charge of customer will decrease as well. Therefore, this research applied the forecasting of electrical power distribution unit that indicated to the demand of electricity usage by creating the model of Support Vector Regression (SVR) with 3 folds cross-validation and compare the forecasting accuracy with the model of Multiple Linear Regression (MLR) and the model of Autoregressive Integrated Moving Average (ARIMA). The data set in this research is time series data in terms of units of household electricity usage were obtained from the Metropolitan Electricity Authority of Thailand. The data had been collected monthly from January 2010 to May 2015. Forecasting models with different parameters are generated from three techniques using the training data, which are

the series from January 2010 to December 2014 and each model is implemented with R language. The remaining data from January 2015 to May 2015 are employed as the validating data. Forecasting performance of each model is measured by the root mean square error (RMSE) and the mean absolute percentage error (MAPE) metrics. When we compare RMSE and MAPE of three models, the experimental result reveals that SVR model has lowest RMSE and MAPE. As a result, we can conclude that SVR model is the most powerful model to forecast this time series data.

Keywords: Support Vector Regression; Cross-validation; Multiple Linear Regression; ARIMA model; Time series; Electricity Usage Demand Forecasting

บทนำ

การพยากรณ์ปริมาณการใช้ไฟฟ้าของผู้บริโภคได้อย่างแม่นยำ จะช่วยวางแผนด้านทรัพยากร จัดสรรงบประมาณ และลดค่าใช้จ่ายในการดำเนินการลงได้ โดยสอดคล้องกับงานวิจัยของ Bunn และ Farmer [1] ที่ได้ศึกษาการพยากรณ์ความต้องการไฟฟ้า พบว่า ค่าใช้จ่ายการดำเนินการจะเพิ่มขึ้น 10 ล้านปอนด์ ทุกปี เมื่อมีการพยากรณ์ผิดพลาดไปทุก 1% นอกจากนี้ ยังมีงานวิจัยในอดีต [2] [3] [7] ที่กล่าวว่า การพยากรณ์ปริมาณการใช้ไฟฟ้ามากกว่าค่าจริง อาจทำให้เกิดพลังงานไฟฟ้ามากกว่าความต้องการจริง เป็นผลให้สูญเสียค่าใช้จ่ายโดยไม่จำเป็น แต่ถ้าพยากรณ์ปริมาณการใช้ไฟฟ้าน้อยกว่าค่าจริง อาจทำให้เกิดพลังงานไฟฟ้าไม่เพียงพอต่อความต้องการ เป็นผลให้เกิดความสูญเสียทางเศรษฐกิจ การดำเนินการธุรกิจ สูญเสียเวลาและภาพลักษณ์ ดังนั้น งานวิจัยนี้ จึงนำเสนอการพยากรณ์หน่วยจำหน่ายไฟฟ้า ซึ่งเป็นข้อมูลการจำหน่ายไฟฟ้าของการไฟฟ้านครหลวงแห่งประเทศไทย ที่จำหน่ายให้กับผู้บริโภคที่อยู่ในเขตพื้นที่จังหวัด กรุงเทพมหานคร สมุทรปราการ และนนทบุรี ประเภทบ้านอยู่อาศัย โดยเก็บรวบรวมข้อมูลจากรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือน ตั้งแต่เดือนมกราคม 2553 ถึงเดือนพฤษภาคม 2558 [4] ซึ่งหน่วยจำหน่ายไฟฟ้า 1 หน่วย หมายถึง การใช้ไฟฟ้า 1,000 วัตต์ต่อชั่วโมง ดังนั้น หน่วยจำหน่ายไฟฟ้า จึง

เป็นข้อมูลที่สะท้อนถึงความต้องการใช้ไฟฟ้าของผู้บริโภค

ข้อมูลหน่วยจำหน่ายไฟฟ้าที่ทำการศึกษามีลักษณะเป็นอนุกรมเวลา ที่บันทึกข้อมูลเป็นรายเดือนตามลำดับเวลา โดยการพยากรณ์ข้อมูลที่มีลักษณะเป็นอนุกรมเวลา จะดำเนินการสร้างตัวแบบการพยากรณ์จากนั้นวัดความแม่นยำในการพยากรณ์ของตัวแบบด้วยการวัดค่าคลาดเคลื่อน (Residual) ซึ่งการสร้างตัวแบบการพยากรณ์ มีหลากหลายเทคนิค และงานวิจัยในอดีต ที่ศึกษาเกี่ยวกับการสร้างตัวแบบการพยากรณ์ข้อมูลอนุกรมเวลา มีหลากหลายงานวิจัย ที่สามารถนำมากล่าวถึงได้ ดังนี้

Wang และคณะ [5] ศึกษาการปรับปรุงความแม่นยำของพยากรณ์ความต้องการไฟฟ้าในเขตไฟฟ้าภาคตะวันตกเฉียงเหนือของประเทศจีน ที่ได้จากตัวแบบ Seasonal ARIMA ด้วยการปรับปรุงค่าคลาดเคลื่อน ผลการศึกษา พบว่า การปรับปรุงค่าคลาดเคลื่อนให้ความแม่นยำเพิ่มขึ้น

Lee และ Tong [6] ศึกษาการพยากรณ์อนุกรมเวลา ซึ่งเป็นข้อมูลการบริโภคพลังงานไฟฟ้ารายปีในจีน โดยนำเสนอตัวแบบผสมผสานระหว่างตัวแบบ ARIMA และ Genetic Programming (GP) ผลการศึกษา พบว่า วิธีการผสมผสานระหว่างตัวแบบ ARIMA และ GP มีความแม่นยำในการพยากรณ์มากกว่าการใช้ตัวแบบ ARIMA หรือตัวแบบ GP เพียงอย่างเดียว

Kaytez และคณะ [7] ศึกษาการพยากรณ์ปริมาณการบริโภคพลังงานไฟฟ้าในตุรกี ด้วยการเปรียบเทียบเทคนิค Artificial Neural Networks

(ANN), Multiple Linear Regression (MLR) และ Least Squares Support Vector Machines (LSSVM) โดยวัดความแม่นยำของการพยากรณ์ด้วยค่า MAPE, MSE, RMSE และ SSE ผลการศึกษา พบว่า ตัวแบบที่สร้างจากเทคนิค LSSVM มีความแม่นยำในการพยากรณ์มากที่สุด

Ogcu, Demirel และ Zaim [8] ศึกษาการพยากรณ์ปริมาณการบริโภคพลังงานไฟฟ้าในประเทศตุรกี ซึ่งเป็นอนุกรมเวลาที่มีคาบเวลาเป็นรายเดือน โดยใช้เทคนิค ANN และ SVR จากนั้น เปรียบเทียบความแม่นยำของตัวแบบ ANN และ SVR ด้วยการวัดค่าคลาดเคลื่อน MAPE ผลการศึกษา พบว่า ตัวแบบ SVR มีความแม่นยำในการพยากรณ์มากกว่าตัวแบบ ANN

Wang และคณะ [9] ศึกษาการพยากรณ์โหลดไฟฟ้าประจำปีของเมืองปักกิ่ง ประเทศจีน ซึ่งเป็นอนุกรมเวลาที่มีคาบเวลาเป็นรายปี โดยใช้เทคนิค SVR ร่วมกับการคัดเลือกพารามิเตอร์ที่เหมาะสมด้วย Differential Evolution Algorithm และเรียกว่าเทคนิค DESVR จากนั้น เปรียบเทียบความแม่นยำของการพยากรณ์กับตัวแบบที่สร้างจากเทคนิค Back Propagation Neuron Network (BPNN) และเทคนิค Regression ผลการศึกษา พบว่า ตัวแบบ DESVR มีความแม่นยำในการพยากรณ์มากที่สุด

จากการทบทวนงานวิจัยในอดีต พบว่า วิธีการสร้างตัวแบบพยากรณ์ ด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) ซึ่งเป็นวิธีการวิเคราะห์ความถดถอยระหว่างอินพุตที่อยู่ในรูปของเวกเตอร์กับตัวแปรเอาต์พุต ด้วยหลักการด้านแมชชีนเลิร์นนิง (Machine learning) ได้ถูกนำมาใช้ในงานวิจัยอย่างแพร่หลายมากขึ้น เพราะเป็นวิธีการสร้างตัวแบบการพยากรณ์ที่เป็นทางเลือกที่ดี (Alternative modeling) เนื่องจากให้ความแม่นยำสูง [7], [8], [9], [10] และเป็นวิธีการที่ประมวลผลได้รวดเร็ว เหมาะสมกับชุดข้อมูลที่มีขนาดเล็ก [10] ดังนั้น งานวิจัยนี้ จึงมีวัตถุประสงค์ เพื่อศึกษาการสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน สำหรับพยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของการไฟฟ้า

นครหลวงแห่งประเทศไทย จากนั้น เปรียบเทียบความแม่นยำของการพยากรณ์กับตัวแบบการถดถอยเชิงเส้นแบบพหุและตัวแบบ ARIMA ด้วยการวัดค่าคลาดเคลื่อน RMSE และ MAPE

วัสดุอุปกรณ์และวิธีดำเนินการวิจัย

งานวิจัยนี้ มีการดำเนินการวิจัยด้วยการสร้างตัวแบบการพยากรณ์ โดยใช้วิธีการที่แตกต่างกัน 3 วิธี ดังนั้น จึงแบ่งวิธีการดำเนินการวิจัย เป็น 3 ส่วน ดังนี้

1. ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน

ซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) เป็นเทคนิคที่ใช้วิธีการของซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ที่นำเสนอโดย Vapnik [11] มาวิเคราะห์ความถดถอยระหว่างอินพุตเวกเตอร์ (Input vector) และตัวแปรเอาต์พุต (Output variables) ซึ่งนำมาใช้กับการพยากรณ์อนุกรมเวลาได้ โดยเปลี่ยนการจำแนกคลาสด้วย SVM เป็นการทำนายค่าด้วย SVR ที่มีเป้าหมาย คือ ต้องการค้นหาความสัมพันธ์เชิงเส้นระหว่างอินพุตเวกเตอร์ในมิติขนาด n ($X \in R^n$) และตัวแปรเอาต์พุต ($y \in R$) และเนื่องจาก SVR ตัดแปลงมาจาก SVM ดังนั้น สมการความถดถอยของ SVR จึงคล้ายคลึงกับสมการไฮเปอร์เพลน (Hyperplane) ของ SVM โดยสมการความถดถอยของ SVR แสดงได้ดังสมการที่ 1 [12]

$$f(x) = w^T x + b \quad (1)$$

เมื่อ w และ b เป็นความชันและออฟเซต (Offset) ของเส้นความถดถอย โดยการกำหนดค่าของ w และ b ทำได้ด้วยการหาค่าต่ำสุดของสมการที่ 2 [12]

$$R = \frac{1}{2} \|w\|^2 + \frac{C}{l} \sum_{i=1}^l |y_i - f(x_i)|_e \quad (2)$$

การใช้เทคนิค SVR ทำนายค่าเอาต์พุตจากอินพุตเวกเตอร์ จะมีการสร้างท่อเอปซิลอน (Epsilon tube) โดยใช้ ฟังก์ชันสูญเสีย (Loss function) ที่มีหลากหลายรูปแบบ ในงานวิจัยนี้ใช้ฟังก์ชันสูญเสียแบบเอปซิลอนอินเซนซิฟ (\mathcal{E} -insensitive loss function) ที่นำเสนอโดย Vapnik [11] เพราะเป็นฟังก์ชันสูญเสีย ที่นิยมใช้ในงานวิจัยอย่างแพร่หลาย [2] [9] [10] [12] [14] โดยแสดงได้ดังสมการที่ 3

$$|y_i - f(x_i)|_\varepsilon = \begin{cases} 0 & \text{if } |y_i - f(x_i)|_\varepsilon \leq \varepsilon \\ |y_i - f(x_i)|_\varepsilon - \varepsilon & \text{Otherwise} \end{cases} \quad (3)$$

การแก้ปัญหาของสมการที่ 2 ที่มีเงื่อนไขตามสมการที่ 3 สามารถปรับให้อยู่ในรูปแบบการแก้ปัญหาแบบคู่อัล (Dual problem) ด้วยการใช้ตัวคูณลากรองจ์ (Lagrange multipliers) ดังสมการที่ 4-5 [12]

$$\begin{aligned} & \text{Maximize } L_p(\alpha_i, \alpha_i^*) \\ & = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) x_i^T x_j \\ & - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \end{aligned} \quad (4)$$

$$\text{Subject to } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \\ 0 \leq \alpha_i^* \leq C, \quad i = 1, \dots, l \end{cases} \quad (5)$$

เมื่อ α_i, α_i^* เป็นตัวคูณลากรองจ์ C เป็นจำนวนเต็มที่เป็นค่าคงที่ ซึ่งเป็นค่าใช้จ่าย (Cost) เมื่อมีข้อผิดพลาด (Error) เกิดขึ้น ε เป็นความกว้างของท่อเอปซิลอน หรือเป็นความคลาดเคลื่อนของชุดข้อมูลฝึกสอน และ l คือจำนวนของซัพพอร์ตเวกเตอร์ (Support vector) ซึ่งอินพุตเวกเตอร์ที่เป็นซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* > 0$ ส่วนอินพุตเวกเตอร์ที่ไม่ใช่ซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* = 0$ และหลังจากที่คำนวณค่า α_i และ α_i^* จากชุดข้อมูลฝึกสอน จะสามารถสร้างสมการ SVR เพื่อใช้ทำนายค่าเอาต์พุตจากอินพุตเวกเตอร์ ได้ดังสมการที่ 6 [12]

$$f(x) = w_0^T x + b = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i^T x + b \quad (6)$$

โดยที่ เวกเตอร์ถ่วงน้ำหนัก (w_0) เป็นดังสมการที่ 7

$$w_0 = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad (7)$$

สมการที่ 6 อยู่ในรูปแบบสมการถดถอยเชิงเส้น แต่ในกรณีที่เป็นกรณีถดถอยแบบไม่เป็นเชิงเส้นสามารถส่งอินพุตเวกเตอร์ไปยังมิติที่สูงขึ้น (High dimensional feature space) โดยใช้เคอร์เนลฟังก์ชัน (Kernel function) ซึ่งเคอร์เนลฟังก์ชันที่นิยมใช้ใน SVR มีดังนี้ [12]

(1) ลีเนียร์เคอร์เนล (Linear kernel): $k(x_i, x) = x_i^T x$

(2) โพลีโนเมียลเคอร์เนล (Polynomial kernel):

$$k(x_i, x) = (1 + x_i \cdot x)^d$$

(3) เกาส์เซียนเคอร์เนล (Gaussian (RBF) kernel):

$$k(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$$

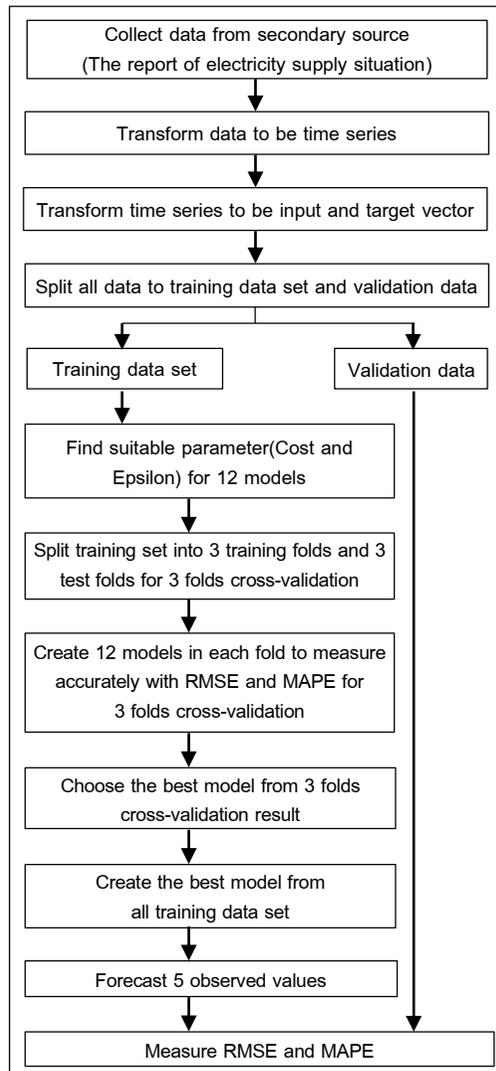
ดังนั้น สมการที่ 6 สามารถเขียนใหม่ในรูปแบบการถดถอยไม่เป็นเชิงเส้น โดยใช้เคอร์เนลฟังก์ชันได้ดังสมการที่ 8

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (8)$$

สำหรับงานวิจัยนี้ ดำเนินการเขียนโปรแกรมทุกขั้นตอน ด้วยการใช้ภาษา R โดยการสร้างตัวแบบพยากรณ์ด้วยเทคนิค SVR จะใช้ฟังก์ชัน svm() ในแพ็คเกจชื่อ e1071 ที่เผยแพร่โดย David Meyer และคณะ [13] ในส่วนเก็บข้อมูลของภาษา R (CRAN repository of R language) ซึ่งการนำ SVR ไปใช้กับการพยากรณ์อนุกรมเวลา จะต้องปรับชุดข้อมูลให้อยู่ในรูปแบบของอินพุตเวกเตอร์และทาร์เก็ต ($D = \{(x_i, y_i)\}_{i=1}^n$) จากนั้นสร้างตัวแบบพยากรณ์จำนวน 12 ตัวแบบ โดยเหตุผลที่เลือกสร้างตัวแบบ จำนวน 12 ตัวแบบ เนื่องจาก อนุกรมเวลาที่นำมาศึกษา มีคาบเวลาเป็นรายเดือน และมีจำนวนคาบเวลาใน 1 ฤดูกาล เท่ากับ 12 คาบเวลา ตามรอบปีปฏิทิน ซึ่งแต่ละตัวแบบ มีความแตกต่างที่จำนวนของอินพุตเวกเตอร์ นั่นคือ ตัวแบบที่ 1 มีอินพุตเวกเตอร์ 1 ตัว เป็นค่าสังเกตล่าสุด 1 คาบเวลา ไล่ไปตามลำดับจนถึงตัวแบบที่ 12 มีอินพุตเวกเตอร์ 12 ตัว เป็นค่าสังเกตล่าสุดตั้งแต่ 1 ถึง 12 คาบเวลา

การสร้างตัวแบบ SVR จะต้องกำหนดค่าพารามิเตอร์ C และ ε โดยแพ็คเกจ e1071 มีฟังก์ชัน tune() [13] ที่ใช้สำหรับการกำหนดค่า C และ ε ที่เหมาะสมที่สุด ดังนั้น จึงใช้ฟังก์ชัน tune() ในการกำหนดค่า C และ ε ให้กับตัวแบบ SVR นอกจากนี้ต้องระบุเคอร์เนลฟังก์ชันด้วย ซึ่งจากการสุ่มทดสอบสร้างตัวแบบ พบว่า ลีเนียร์เคอร์เนลให้ผลลัพธ์ที่ดีที่สุดกับข้อมูลชุดนี้ ดังนั้น งานวิจัยนี้ จึงใช้ลีเนียร์เคอร์เนลเป็นเคอร์เนลฟังก์ชัน และเพื่อกำจัดการวัดความแม่นยำของการวัดความแม่นยำของตัวแบบ จึงนำเสนอวิธีการตรวจสอบสลับ 3 ส่วน (3 folds cross-validation) สำหรับเลือกตัวแบบที่เหมาะสมที่สุด ในการนำมาใช้

พยากรณ์อนุกรมเวลา โดยสรุปขั้นตอนการสร้างตัวแบบ SVR สำหรับพยากรณ์ข้อมูลอนุกรมเวลา เพื่อ นำไปเปรียบเทียบกับตัวแบบ MLR และตัวแบบ ARIMA ได้ดังรูปที่ 1



รูปที่ 1 ขั้นตอนการสร้างตัวแบบ SVR ที่เหมาะสมที่สุด

จากรูปที่ 1. อธิบายแต่ละขั้นตอนได้ดังนี้

Step 1: Collect data from secondary source คือ ขั้นตอนรวบรวมข้อมูล จากข้อมูลทุติยภูมิที่ได้จากรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือนของการไฟฟ้านครหลวง ตั้งแต่เดือนมกราคม 2553

ถึงเดือนพฤษภาคม 2558 ของผู้บริโภคประเภทบ้านอยู่อาศัย

Step 2: Transform data to be time series คือ ขั้นตอนที่น่าข้อมูลทุติยภูมิที่รวบรวมได้ มาสร้างเป็นข้อมูลอนุกรมเวลา เรียงตามลำดับเวลา โดยมีคาบเวลาเป็นรายเดือน แสดงได้ดังรูปที่ 2

	A	B	C	D	E
1	Years	Months	Electric Units	No_Consumers	Mean_Units
2	2010	1	776.32	2420767	320.69
3	2010	2	836.16	2423046	345.09
4	2010	3	925.55	2433022	380.41
5	2010	4	1026.5	2439689	420.75
6	2010	5	1082.18	2448223	442.03
7	2010	6	1003.41	2456309	408.5
8	2010	7	930.91	2463398	377.9
9	2010	8	876.83	2473742	354.45

รูปที่ 2 ข้อมูลอนุกรมเวลาที่สร้างจากข้อมูลทุติยภูมิ

Step 3: Transform data to be input and target vector คือ ขั้นตอนการแปลงข้อมูลตามรูปที่ 2 ให้เป็นข้อมูลตั้งรูปที่ 3 โดยแต่ละคอลัมน์หมายถึงค่าสังเกตที่ล่าช้าไปตั้งแต่ 1 ถึง 12 คาบเวลา ($y_{t-1}, y_{t-2}, \dots, y_{t-12}$) ตามลำดับ และคอลัมน์สุดท้ายคือ ค่าของทาร์เก็ต ซึ่ง

เป็นค่าสังเกต ณ เวลา t (y_t) โดยค่าของทาร์เก็ตในแถวแรก คือ ค่าสังเกตที่ล่าช้าเวลา $t=13$ เมื่อแปลงข้อมูลแล้ว จะได้ชุดข้อมูลจำนวน 53 ชุด จากอนุกรมเวลาเดิมที่มีค่าสังเกต 65 ค่า ตามจำนวนคาบเวลาตั้งแต่เดือนมกราคม 2553 ถึงเดือนพฤษภาคม 2558

A	B	C	D	E	F	G	H	I	J	K	L	M
ValT1	ValT2	ValT3	ValT4	ValT5	ValT6	ValT7	ValT8	ValT9	ValT10	ValT11	ValT12	Target
773.99	787.33	820.6	873.47	876.83	930.91	1003.4	1082.2	1026.5	925.55	836.16	776.32	707.16
707.16	773.99	787.33	820.6	873.47	876.83	930.91	1003.4	1082.2	1026.5	925.55	836.16	787.29
787.29	707.16	773.99	787.33	820.6	873.47	876.83	930.91	1003.4	1082.2	1026.5	925.55	806.64
806.64	787.29	707.16	773.99	787.33	820.6	873.47	876.83	930.91	1003.4	1082.2	1026.5	864.58
864.58	806.64	787.29	707.16	773.99	787.33	820.6	873.47	876.83	930.91	1003.4	1082.2	968.86
968.86	864.58	806.64	787.29	707.16	773.99	787.33	820.6	873.47	876.83	930.91	1003.4	933.59
933.59	968.86	864.58	806.64	787.29	707.16	773.99	787.33	820.6	873.47	876.83	930.91	912.8
912.8	933.59	968.86	864.58	806.64	787.29	707.16	773.99	787.33	820.6	873.47	876.83	902.59
902.59	912.8	933.59	968.86	864.58	806.64	787.29	707.16	773.99	787.33	820.6	873.47	886.08

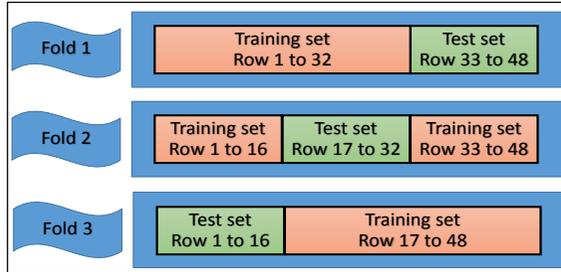
รูปที่ 3 ลักษณะของชุดข้อมูลที่ใช้สร้างตัวแบบ SVR

Step 4: Split all data to training data set and validation data set คือ ขั้นตอนการแบ่งชุดข้อมูลที่ได้จากขั้นตอนที่ 3 ออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลฝึกสอน (Training data set) จำนวน 48 ชุด และชุดข้อมูลตรวจสอบ (Validation data set) จำนวน 5 ชุด

Step 5: Find suitable parameter (Cost and Epsilon) for 12 models คือ ขั้นตอนการกำหนดค่า C และ \mathcal{E} ที่เหมาะสมที่สุดของแต่ละตัวแบบ ซึ่งกำหนดตัวแบบไว้ 12 ตัวแบบ โดยตัวแบบที่ 1 มีอินพุตเวกเตอร์เป็นค่าสังเกตที่ล่าช้าไป 1 คาบเวลา (y_{t-1}) , ตัวแบบที่ 2 มีอินพุตเวกเตอร์เป็นค่าสังเกตที่ล่าช้าไป 1 และ 2 คาบเวลา (y_{t-1}, y_{t-2}) ไปไล่ไปตามลำดับจนถึงตัวแบบที่ 12 มีอินพุตเวกเตอร์เป็นค่าสังเกตที่ล่าช้าไป 1, 2, ..., 12 คาบเวลา ($y_{t-1}, y_{t-2}, \dots, y_{t-12}$) โดยการหาค่า C และ \mathcal{E} ที่เหมาะสมที่สุดของแต่ละตัวแบบ จะใช้ฟังก์ชัน

tune() ซึ่งจากการทดลองสุ่มกับชุดข้อมูลฝึกสอนพบว่า ค่า C มีค่าอยู่ระหว่าง 1 ถึง 16 ส่วนค่า \mathcal{E} มีค่าอยู่ระหว่าง 0 ถึง 1 ดังนั้น จึงกำหนดช่วงการค้นหาค่า C ตั้งแต่ 1 ถึง 16 และช่วงการค้นหาค่า \mathcal{E} ตั้งแต่ 0 ถึง 1 เพิ่มขึ้นครั้งละ 0.001 จากนั้นเก็บค่า C และ \mathcal{E} ที่เหมาะสมที่สุดของแต่ละตัวแบบไว้ในไฟล์ เพื่อให้สามารถนำไปใช้ต่อ ในขั้นตอนการเขียนคำสั่งสำหรับสร้างตัวแบบ SVR

Step 6: Split training set into 3 training folds and 3 test folds for 3 folds คือ ขั้นตอนการแบ่งชุดข้อมูลฝึกสอนจำนวน 48 ชุด ออกเป็น 3 ส่วน (Fold) ดังนั้น แต่ละส่วน จะมีจำนวนชุดข้อมูลเท่ากับ 16 ชุด สำหรับนำไปใช้เลือกโมเดลที่เหมาะสมที่สุดด้วยการตรวจสอบสลับแบบ 3 ส่วน และวิธีการแบ่งข้อมูลเป็น 3 ส่วน แสดงได้ดังรูปที่ 4



รูปที่ 4 การแบ่งชุดข้อมูลสำหรับการตรวจสอบสลับ 3 ส่วน

Step 7: Create 12 models in each fold to measure accurately with RMSE and MAPE for 3 folds cross-validation คือ ขั้นตอนการสร้างตัวแบบพยากรณ์ 12 ตัวแบบ โดยใช้ชุดข้อมูลที่แบ่งออกเป็น 3 ส่วน จากขั้นตอนที่ 6 จากนั้นวัดค่า RMSE และ MAPE กับชุดข้อมูลทดสอบ (Test data set) ของแต่ละตัวแบบ ในแต่ละส่วน เพื่อนำมาหาค่าเฉลี่ย และใช้เป็นเกณฑ์ในการเลือกตัวแบบที่เหมาะสมที่สุดสำหรับการนำมาพยากรณ์ค่าสังเกตล่วงหน้า 5 ช่วงเวลา

Step 8: Choose the best model from 3 folds cross-validation result คือ ขั้นตอนการเลือกตัวแบบที่เหมาะสมที่สุดด้วยการพิจารณาจากค่าเฉลี่ย RMSE และ MAPE ของทั้ง 3 ส่วน

Step 9: Create the best model from all training data set คือ ขั้นตอนการสร้างตัวแบบที่ดีที่สุดจากชุดข้อมูลฝึกสอนจำนวน 48 ชุด เพื่อนำไปใช้ใน

การพยากรณ์ค่าสังเกตล่วงหน้า 5 ช่วงเวลา ด้วยอินพุตเวกเตอร์ของชุดข้อมูลตรวจสอบที่แบ่งไว้จำนวน 5 ชุด

Step 10: Forecast 5 observed values คือ ขั้นตอนการใช้ตัวแบบที่ดีที่สุดพยากรณ์ค่าสังเกต 5 ค่า ซึ่งเป็นค่าสังเกตของเดือนมกราคมถึงเดือนพฤษภาคม 2558

Step 11: Measure RMSE and MAPE คือ ขั้นตอนการนำค่าพยากรณ์จากขั้นตอนที่ 10 ไปเปรียบเทียบกับชุดข้อมูลตรวจสอบ ที่แบ่งไว้ในขั้นตอนที่ 4 โดยเป็นค่าสังเกตในช่วงเวลาเดียวกับค่าพยากรณ์ เพื่อวัดค่า RMSE และ MAPE ของตัวแบบ SVR ที่เหมาะสมที่สุดสำหรับบางส่วนของชุดค่าสังเกตที่สำคัญ ตามขั้นตอนการสร้างตัวแบบ SVR ที่เหมาะสมที่สุด (รูปที่ 1) แสดงได้ดังรูปที่ 5

```
AllData<-read.csv(file="d:/CodeSVR/Data/EuseTran.csv" , header=TRUE, sep=",")
RowTest=5
RowTrain=dim(AllData)[1] - RowTest
TrainSet<-matrix(nrow=RowTrain,ncol=13)
k=1
for(i in 1:RowTrain) {
  for(j in 1:13) {
    TrainSet[k,j]=AllData[i,j] }
  k=k+1 }
TrainSet<-data.frame(TrainSet)
names(TrainSet)<-c("ValT1","ValT2","ValT3","ValT4","ValT5","ValT6","ValT7","ValT8","ValT9",
,"ValT10","ValT11","ValT12","Target")
```

(a) Command set for transform data to training data set

```
InV2<-data.frame(Trainload$ValT1,Trainload$ValT2)
tune2 <- tune(svm,InV2, Trainload$Target, data = data,ranges = list(epsilon = seq(0,1,0.001), cost = 1:16))
```

(b) Command set with using tune() function for model2

```
InVM2F1 <- data.frame(TrainLoadF1$ValT1,TrainLoadF1$ValT2)
model2F1 <-
svm(x=InVM2F1,y=TrainLoadF1$Target,kernel="linear",cost=TuneLoad[2,3],epsilon=TuneLoad[2,4])
newdataM2F1 = data.frame(Val1=TestLoadF1$ValT1,Val2=TestLoadF1$ValT2)
```

(c) Command set for create model2 in first fold by using svm() function

รูปที่ 5 บางส่วนของชุดคำสั่งสำหรับสร้างตัวแบบ SVR

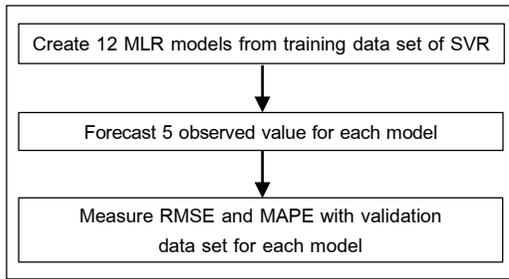
2. ตัวแบบการถดถอยเชิงเส้นแบบพหุ

การวิเคราะห์การถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression: MLR) เป็นวิธีการวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มของตัวแปรอิสระที่มีมากกว่า 1 ตัวแปรและตัวแปรตาม 1 ตัวแปร โดยตัวแบบพยากรณ์ที่สร้างจากวิธีการ MLR เป็นตัวแบบที่อยู่ในรูปแบบสมการเชิงเส้น ที่แสดงความสัมพันธ์ระหว่างกลุ่มของตัวแปรอิสระและตัวแปรตาม ดังสมการที่ 9 [14]

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon \quad (9)$$

การนำวิธี MLR มาวิเคราะห์หอนุกรมเวลาเชิงเดี่ยวที่มีค่าสังเกตค่าเดียว ในแต่ละคาบเวลา ทำให้ตัวแปรตาม y หมายถึง ค่าสังเกต ณ.เวลา t ส่วนกลุ่มของตัวแปรอิสระ x_i ($i=1, 2, \dots, k$) หมายถึง ค่าสังเกตที่ล่าช้า

ไป $1, 2, \dots, k$ คาบเวลา ซึ่ง β_i ($i=1, 2, \dots, k$) คือ ค่าสัมประสิทธิ์ความถดถอย (Regression Coefficient) ของ x_i และ ε คือค่าคลาดเคลื่อน (Residual) เมื่อ x_i หมายถึง ค่าสังเกตที่ล่าช้าไป $1, 2, \dots, k$ คาบเวลา และ y หมายถึง ค่าสังเกต ณ.เวลา t ดังนั้น ข้อมูลที่ใช้ในการสร้างตัวแบบ MLR จึงอยู่ในรูปแบบ $D = \{(x_i, y_i)\}_{i=1}^n$ เช่นเดียวกับข้อมูลที่ใช้สร้างตัวแบบ SVR จึงใช้ชุดข้อมูลฝึกสอนของตัวแบบ SVR มาสร้างตัวแบบ MLR และใช้ชุดข้อมูลตรวจสอบของตัวแบบ SVR มาใช้วัดค่าคลาดเคลื่อน RMSE และ MAPE ของตัวแบบ MLR เพื่อนำไปเปรียบเทียบกับตัวแบบ SVR และตัวแบบ ARIMA โดยสรุปขั้นตอนการสร้างตัวแบบ MLR สำหรับพยากรณ์ข้อมูลอนุกรมเวลา 5 คาบเวลาเป็นดังรูปที่ 6



รูปที่ 6 ขั้นตอนการสร้างตัวแบบ MLR

จากรูปที่ 6 อธิบายแต่ละขั้นตอนได้ดังนี้

Step 1: Create 12 MLR models from training data set of SVR คือ ขั้นตอนการสร้างตัวแบบ MLR 12 ตัวแบบ โดยตัวแบบที่ 1 มีตัวแปรอิสระ 1 ตัวแปร และเป็นค่าสังเกตที่ล่าช้าไป 1 คาบเวลา (y_{t-1}) , ตัวแบบที่ 2 มีตัวแปรอิสระ 2 ตัวแปร และเป็นค่าสังเกตที่ล่าช้าไป 1 และ 2 คาบเวลา (y_{t-1}, y_{t-2}) ไปไปตามลำดับจนถึงตัวแบบที่ 12 มีตัวแปรอิสระ 12 ตัวแปร และเป็นค่า

สังเกตที่ล่าช้าไป 1, 2, ..., 12 คาบเวลา ($y_{t-1}, y_{t-2}, \dots, y_{t-12}$) ซึ่งใช้ชุดข้อมูลฝึกสอนของตัวแบบ SVR ในการสร้างตัวแบบ MLR

Step 2: Forecast 5 observed values คือ ขั้นตอนการนำแต่ละตัวแบบ พยากรณ์ค่าสังเกตล่วงหน้า 5 คาบเวลา ซึ่งเป็นค่าสังเกตของเดือนมกราคมถึงเดือนพฤษภาคม 2558 สำหรับนำไปเปรียบเทียบกับชุดข้อมูลตรวจสอบ ซึ่งเป็นค่าสังเกตในช่วงเวลาเดียวกัน

Step 3: Measure RMSE and MAPE with validation data set for each model คือ ขั้นตอนการวัดค่า RMSE และ MAPE ของแต่ละตัวแบบกับชุดข้อมูลตรวจสอบ เพื่อเลือกตัวแบบ MLR ที่เหมาะสมที่สุด จากนั้น จะได้นำไปเปรียบเทียบความแม่นยำของการพยากรณ์กับตัวแบบ SVR และตัวแบบ ARIMA

ตัวอย่างชุดคำสั่งในการสร้างตัวแบบ MLR ตัวแบบที่ 2 ซึ่งมีตัวแปรอิสระ 2 ตัว คือ y_{t-1} และ y_{t-2} แสดงได้ดังรูปที่ 7

```

formularMLR<-Trainload$Target ~Trainload$ValT1+Trainload$ValT2
modelMLR <- glm(formularMLR,family = gaussian())
Testload<-read.csv(file="TestData2.csv")
newdata = data.frame(Val1=Testload$ValT1,Val2=Testload$ValT2)
ForeMLR<-modelMLR$coefficients[1]+modelMLR$coefficients[2]*newdata[1]+modelMLR$coefficients[3]*newdata[2]
err1 <- Testload$ActualValue - ForeMLR
err2 <- err1/Testload$ActualValue
rmseMLRtest <- sqrt(mean(err1[,1]^2))
mapeMLRtest<-mean(abs(100*err2[,1]))
rmseMLRtest # show RMSE
mapeMLRtest # show MAPE
    
```

รูปที่ 7 ชุดคำสั่งสร้างตัวแบบ MLR ตัวแบบที่

3. ตัวแบบ ARIMA

ตัวแบบ ARIMA เป็นตัวแบบที่สร้างจากวิธีของ บอกซ์และเจนกินส์ (Box and Jenkins method) [15] โดยพยากรณ์ข้อมูลจากรูปแบบหลัก 2 รูปแบบ ได้แก่ รูปแบบ Autoregressive (AR) คือ รูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใดๆ ขึ้นอยู่กับค่าสังเกตก่อนหน้า และรูปแบบ Moving average (MA) คือ รูปแบบที่

กำหนดว่า ค่าพยากรณ์ที่เวลาใดๆ ขึ้นอยู่กับค่าคลาดเคลื่อนก่อนหน้า และเรียกรูปแบบ AR รวมกับรูปแบบ MA ว่า รูปแบบ ARMA ซึ่งเป็นรูปแบบการวิเคราะห์อนุกรมเวลาที่เป็นสเตชันนารี ในกรณีที่อนุกรมเวลาไม่เป็นสเตชันนารี จะต้องแปลงอนุกรมเวลาให้เป็นสเตชันนารี และเรียกรูปแบบที่ใช้วิเคราะห์อนุกรมเวลาที่แปลงให้เป็นสเตชันนารี ได้ว่า รูปแบบ

ARIMA โดยมีรูปแบบทั่วไปที่เขียนในรูปของตัวดำเนินการย้อนกลับ (Backward shift operator: B) ในรูปสมการ [5] ดังนี้

$$\theta_p(B)\Theta_p(B^S)(1-B)^d(1-B^S)^D Y_t = w_q(B)W_Q(B^S)\alpha_t$$

เมื่อ $BY_t = Y_{t-1}$ และ $B^k Y_t = Y_{t-k}$ โดยที่

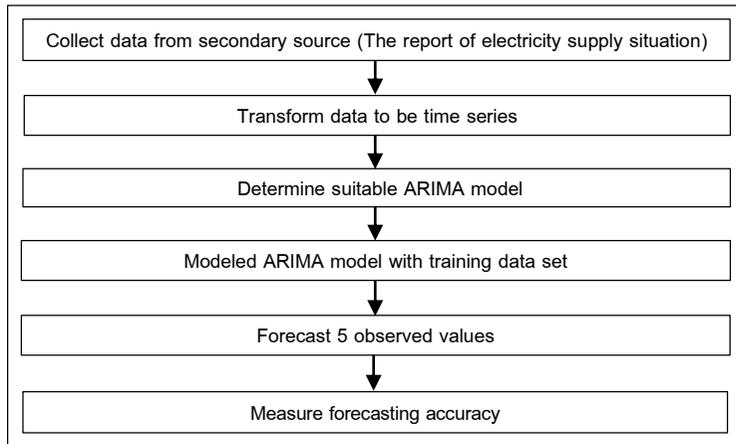
$$w_q(B) = 1 - w_1 B - w_2 B^2 - \dots - w_q B^q$$

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

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

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

สำหรับการสร้างตัวแบบ ARIMA ต้องวิเคราะห์อนุกรมเวลา เพื่อกำหนดรูปแบบ ARIMA(p, d, q)x(P, D, Q)_S ที่เหมาะสมกับอนุกรมเวลา โดยเป็นขั้นตอนในการเลือกอันดับ p, d, q เนื่องจากแนวโน้ม และอันดับ P, D, Q เนื่องจากฤดูกาล ซึ่งภาษา R มีฟังก์ชัน auto.arima() ที่ใช้ในการระบุรูปแบบ ARIMA(p, d, q)x(P, D, Q)_S ที่เหมาะสม ดังนั้น การพยากรณ์อนุกรมเวลาด้วยตัวแบบ ARIMA จึงมีขั้นตอนดังรูปที่ 8 และอธิบายแต่ละขั้นตอนได้ ดังนี้



รูปที่ 8 ขั้นตอนการสร้างตัวแบบ ARIMA

Step 1: Collect data from secondary source คือ ขั้นตอน รวบรวมข้อมูล จากข้อมูลทุติยภูมิที่ได้จากรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือนของการไฟฟ้านครหลวง เช่นเดียวกับขั้นตอนที่ 1 ของการสร้างตัวแบบ SVR

Step 2: Transform data to be time series คือ ขั้นตอนที่น่าข้อมูลทุติยภูมิที่รวบรวมได้ มาสร้างเป็นข้อมูลอนุกรมเวลา เรียงตามลำดับเวลา โดยมีคาบเวลาเป็นรายเดือน เช่นเดียวกับขั้นตอนที่ 2 ของการสร้างตัวแบบ SVR ดังรูปที่ 2 จากนั้น แบ่งข้อมูลเป็น 2 ชุด ได้แก่ 1) ชุดข้อมูลฝึกสอน (Training data set) ซึ่งนำมาแปลงเป็นอนุกรมเวลาด้วยฟังก์ชัน ts() ในภาษา R เพื่อนำไปสร้างตัวแบบ ARIMA โดยใช้ข้อมูลตั้งแต่เดือนมกราคม 2553 ถึงเดือนธันวาคม 2557 สำหรับสร้างตัวแบบพยากรณ์ 2) ชุดข้อมูลตรวจสอบ (Validation data set) ซึ่งใช้ข้อมูลตั้งแต่

เดือนมกราคมถึงเดือนพฤษภาคม 2558 สำหรับวัดความถูกต้องของการพยากรณ์

Step 3: Determine suitable ARIMA คือ ขั้นตอนการกำหนดรูปแบบของตัวแบบ ARIMA ที่เหมาะสมกับชุดข้อมูลฝึกสอน โดยเป็นการกำหนดอันดับ p, d, q และอันดับ P, D, Q ของตัวแบบ ARIMA ด้วยฟังก์ชัน auto.arima()

Step 4: Modeled ARIMA model with training data set คือ ขั้นตอนการสร้างตัวแบบ ARIMA จากชุดข้อมูลฝึกสอน โดยนำค่าอันดับ p, d, q และอันดับ P, D, Q ที่ได้จากขั้นตอนที่ 3 มาใช้สร้างตัวแบบ ARIMA ด้วยฟังก์ชัน arima()

Step 5: Forecast 5 observed values คือ ขั้นตอนการนำตัวแบบ ARIMA พยากรณ์ค่าสังเกตล่วงหน้า 5 คาบเวลา ด้วยฟังก์ชัน predict() ในภาษา R โดยค่าพยากรณ์เป็นค่าสังเกตของเดือนมกราคมถึง

เดือนพฤษภาคม 2558 สำหรับนำไปเปรียบเทียบกับชุดข้อมูลตรวจสอบ ซึ่งเป็นค่าสังเกตในช่วงเวลาเดียวกัน

Step 6: Measure forecasting accuracy คือขั้นตอนการวัดความแม่นยำของการพยากรณ์ โดยนำค่าพยากรณ์ที่ได้จากตัวแบบมาเปรียบเทียบกับค่าจริงของชุดข้อมูลตรวจสอบ ด้วยการวัดค่าคลาดเคลื่อนทางสถิติ ได้แก่ ค่า RMSE (Root Mean Square Error) และค่า MAPE (Mean Absolute Percentage Error) [16] ซึ่งมีสมการดังนี้

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n [(y_t - \hat{y}_t)]^2} \quad (10)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (11)$$

เมื่อ y_t = ค่าสังเกตที่เวลา t , \hat{y}_t = ค่าพยากรณ์ที่เวลา t และ n = จำนวนช่วงเวลาที่ยพยากรณ์ สำหรับชุดคำสั่งทั้งหมดของภาษา R ตามขั้นตอนในรูปที่ 8 แสดงได้ดังรูปที่ 9.

```
library("forecast", lib.loc="--/R/win-library/3.2")
arimaDF <- read.csv(file="d:/R/Euse2010_2014.csv", header=TRUE, sep=",")
arimaTS <- ts(arimaDF[,3], start=c(2010,1), end=c(2014,12), frequency=12)
auto.arima(arimaTS)
arimaM <- arima(arimaTS, order=c(p,d,q), list(order=c(P,D,Q), period=12))
# p,d,q and P,D,Q is an integer number, it's can be obtain by using auto.arima()
PredictARIMA <- predict(arimaM, n.ahead=5)
PredictARIMA # show forecasting value
TestSet<-read.csv(file="d:/R/Etest1_5_2015.csv")
ActualARIMA<-c(TestSet$Electric_Units)
ForecastARIMA<-c(PredictARIMA$pred[1:5])
err1 <- ActualARIMA-ForecastARIMA
err2 <- (ActualARIMA-ForecastARIMA)/ActualARIMA
rmseARIMA <- sqrt(mean(err1^2))
mapeARIMA<-mean(abs(100*err2))
rmseARIMA # show RMSE
mapeARIMA # show MAPE
```

รูปที่ 9 ชุดคำสั่งของภาษา R ตามขั้นตอนการสร้างตัวแบบ ARIM

ผลการวิจัย

งานวิจัยนี้ได้ศึกษาการวิเคราะห์หอนุกรมเวลาเพื่อพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัย ของการไฟฟ้านครหลวงแห่งประเทศไทย ซึ่งได้ดำเนินการทดลอง ด้วยวิธีการสร้างตัวแบบพยากรณ์ที่แตกต่างกัน 3 วิธี ได้แก่ วิธีการสร้างตัวแบบ SVR วิธีการสร้างตัวแบบ MLR และวิธีการสร้างตัวแบบ ARIMA โดยใช้การเขียนโปรแกรมด้วยภาษา R เป็นเครื่องมือดำเนินการวิจัย ดังนั้น จึงแบ่งผลการวิจัยตาม

วิธีดำเนินการวิจัย จากนั้น เปรียบเทียบผลการวิจัยที่ได้จากวิธีดำเนินการวิจัยทั้ง 3 วิธี ซึ่งผลการวิจัยแต่ละส่วนเป็นดังนี้

1. ผลการวิจัยจากตัวแบบ SVR

เทคนิค SVR เป็นวิธีการที่มีพารามิเตอร์ ซึ่งพารามิเตอร์ที่สำคัญคือ C และ \mathcal{E} โดย C คือ ค่าคงที่ที่ใช้กำหนดความเหมาะสมระหว่างค่าคลาดเคลื่อนที่ยอมรับได้ และขนาดของเวกเตอร์ถ่วงน้ำหนัก ($\|w\|$) ส่วน \mathcal{E} คือ

การกำหนดค่าความกว้างของท่อเปิดซิลอน หรือค่าคลาดเคลื่อนที่ยอมรับได้ของชุดข้อมูลฝึกสอน ซึ่งงานวิจัยนี้ใช้ฟังก์ชัน tune() ค้นหาค่า C และ \mathcal{E} ที่เหมาะสม โดย

ฟังก์ชัน tune() จะค้นหาค่าที่เหมาะสมแบบกริดเสิร์จ (Grid search) ตลอดช่วงของค่าพารามิเตอร์ที่กำหนดไว้ และประเมินความเหมาะสมด้วยการวัดค่า MSE (Mean Square Error) [13] ได้ผลลัพธ์ดังตารางที่ 1

ตารางที่ 1 ค่า C และ \mathcal{E} ที่เหมาะสมจากการค้นหาด้วยฟังก์ชัน tune() ของตัวแบบ SVR

Model	C	\mathcal{E}	Model	C	\mathcal{E}
1	1	0.433	7	2	0.001
2	1	0	8	2	0.024
3	5	0.567	9	2	0.081
4	2	0.686	10	3	0.193
5	2	0.263	11	2	0.011
6	2	0.325	12	3	0.135

ตารางที่ 2 ผลการตรวจสอบสลับ 3 ส่วนของตัวแบบ SVR

Model	RMSE	RMSE	RMSE	RMSE	MAPE	MAPE	MAPE	MAPE
	fold 1	fold 2	fold 3	Average	fold 1	fold 2	fold 3	Average
1	90.13	66.06	74.43	76.88	7.17	5.28	6.97	6.47
2	80.02	56.49	89.28	75.26	6.55	4.03	8.12	6.24
3	75.45	65.62	88.10	76.39	6.45	5.50	8.20	6.72
4	77.03	70.52	80.04	75.86	6.70	5.99	7.29	6.66
5	75.15	63.70	103.87	80.91	6.74	5.09	9.34	7.06
6	80.58	65.95	104.83	83.79	7.02	5.38	9.75	7.38
7	64.73	54.29	110.28	76.43	5.85	4.07	11.05	6.99
8	62.73	56.17	119.19	79.36	5.63	4.80	12.41	7.61
9	56.61	57.00	106.40	73.34	5.22	4.81	10.84	6.96
10	49.99	55.80	99.17	68.32	4.70	4.61	9.91	6.41
11	49.19	49.22	96.53	64.98	4.65	4.14	9.69	6.16
12	54.20	54.89	93.60	67.56	5.01	4.38	9.26	6.22

เมื่อได้ค่า C และ \mathcal{E} ที่เหมาะสมสำหรับแต่ละตัวแบบ จะนำไปสร้างตัวแบบ SVR จำนวน 12 ตัวแบบ และเลือกตัวแบบที่ดีที่สุดด้วยการวัดค่า RMSE และ MAPE โดยวิธีตรวจสอบสลับ 3 ส่วน ซึ่งผลการวัดค่า RMSE และ MAPE ด้วยวิธีตรวจสอบสลับ 3 ส่วนของแต่ละตัวแบบเป็นดังตารางที่ 2

จากตารางที่ 2 ตัวแบบที่ 11 เป็นตัวแบบที่ให้ค่าเฉลี่ย RMSE และ MAPE ต่ำที่สุด จึงเป็นตัวแบบ

SVR ที่ดีที่สุดสำหรับนำมาใช้ในการพยากรณ์อนุกรมเวลาชุดนี้ ดังนั้น ทำการสร้างตัวแบบ SVR ด้วยชุดข้อมูลฝึกสอน ที่มีอินพุต 11 ตัว คือ ค่าสังเกต้าหลังตั้งแต่ 1 ถึง 11 คาบเวลา จากนั้นนำไปเปรียบกับค่าสังเกตจริงในช่วงเวลาเดียวกันของชุดข้อมูลตรวจสอบเพื่อวัดค่า RMSE และ MAPE ได้ผลลัพธ์เป็นดังตารางที่ 3

ตารางที่ 3 ค่าพยากรณ์และค่าคลาดเคลื่อนของตัวแบบ SVR ที่เหมาะสมที่สุด

Value	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	RMSE	MAPE
Actual Value	745.82	853.60	1017.89	1100.93	1216.77	-	-
SVR model 11	749.60	795.30	960.09	1144.16	1197.00	42.46	3.71

2. ผลการวิจัยจากตัวแบบ MLR

สำหรับการสร้างตัวแบบพยากรณ์ด้วยเทคนิค MLR จะใช้ชุดข้อมูลฝึกสอนที่มีลักษณะเช่นเดียวกับชุดข้อมูลฝึกสอนที่ใช้สร้างตัวแบบ SVR โดยทำการสร้างตัวแบบ MLR จำนวน 12 ตัวแบบ

จากชุดข้อมูลฝึกสอน ในลักษณะเช่นเดียวกับตัวแบบ SVR จากนั้น นำตัวแบบ MLR ทั้ง 12 ตัวแบบไปพยากรณ์ค่าสังเกต 5 คาบเวลา แล้วนำไปเปรียบเทียบกับค่าสังเกตจริงในช่วงเวลาเดียวกันของชุดข้อมูลทดสอบ เพื่อวัดค่า RMSE และ MAPE ได้ผลลัพธ์เป็นดังตารางที่ 4

ตารางที่ 4 ค่าพยากรณ์และค่าคลาดเคลื่อนของตัวแบบ MLR

Value	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	RMSE	MAPE
Actual Value	745.82	853.60	1017.89	1100.93	1216.77	-	-
MLR model 1	865.83	795.18	874.89	996.39	1057.80	122.04	11.91
MLR model 2	813.97	768.57	929.83	1060.45	1076.56	90.47	8.59
MLR model 3	821.41	759.93	917.04	1068.54	1090.44	91.29	8.87
MLR model 4	821.31	760.18	916.76	1068.13	1090.64	91.26	8.87
MLR model 5	824.85	762.87	911.22	1074.87	1100.06	89.62	8.73
MLR model 6	814.68	755.67	908.80	1085.66	1089.13	92.47	8.66
MLR model 7	754.15	726.63	881.39	1070.78	1119.97	94.98	8.02
MLR model 8	772.42	744.48	901.54	1091.95	1127.63	82.68	7.18
MLR model 9	785.11	792.74	935.99	1129.05	1174.92	53.84	5.29
MLR model 10	782.15	797.80	943.23	1131.58	1184.00	49.03	4.84
MLR model 11	760.70	789.74	950.92	1143.49	1190.37	47.52	4.42
MLR model 12	762.18	792.08	952.35	1143.28	1188.78	46.74	4.40

จากตารางที่ 4 ตัวแบบ MLR ที่ดีที่สุด คือ ตัวแบบที่ 12 เพราะให้ค่า RMSE และ MAPE ต่ำที่สุด โดยมีค่า RMSE = 46.74 และค่า MAPE = 4.40 ดังนั้นตัวแบบ MLR ที่ดีที่สุด คือ ตัวแบบที่มีตัวแปรอิสระ 12 ตัวแปร ซึ่งเป็นค่าสังเกตล่าหลังตั้งแต่ 1 ถึง 12 คาบเวลา ($y_{t-1}, y_{t-2}, \dots, y_{t-12}$) ส่วนตัวแปรตาม คือ ค่าสังเกต ณ เวลา t (y_t) นอกจากนี้ พิจารณาผลการทดลองตามตารางที่ 4. พบว่า การเพิ่มตัวแปรอิสระ จะทำให้ค่าคลาดเคลื่อนลดลง ซึ่งสอดคล้องกับหลักการ

วิเคราะห์ความถดถอยเชิงเส้นแบบพหุ แต่การเพิ่มตัวแปรอิสระเข้าไป เพื่อให้ค่าคลาดเคลื่อนลดลง จะต้องใช้ข้อมูลจำนวนมาก และเป็นข้อด้อยของหลักการวิเคราะห์ความถดถอยเชิงเส้นแบบพหุ

3. ผลการวิจัยจากตัวแบบ ARIMA

การสร้างตัวแบบ ARIMA จะต้องกำหนดรูปแบบของตัวแบบ ARIMA ด้วยการระบุอันดับ (p, d, q) และอันดับ (P, D, Q)_s ซึ่งภาษา R มีฟังก์ชัน `auto.arima()` สำหรับกำหนดอันดับ (p, d, q) และ

อันดับ (P, D, Q)_s ที่เหมาะสม โดยผลจากการใช้ฟังก์ชัน auto.arima() เพื่อกำหนดรูปแบบของตัวแบบ ARIMA เป็นดังรูปที่ 10 ซึ่งทำให้ทราบว่า ตัวแบบ ARIMA มีรูปแบบเป็น ARIMA(1,0,0)x(1,0,0)₁₂ โดยมีคาบเวลา 1 ฤดูกาลเท่ากับ 12 (S=12) และสรุปได้ว่าตัวแบบ ARIMA ประกอบด้วย Autoregressive (AR) อันดับ p=1 และ Seasonal autoregressive (SAR) อันดับ P=1 โดยประมาณค่าสัมประสิทธิ์ AR₁ (θ₁) และ SAR₁ (Θ₁) ได้เท่ากับ 0.7301 และ 0.6403 ตามลำดับ โดยที่อันดับ d=D=0 และอันดับ q=Q=0 ดังนั้นสามารถสร้างสมการพยากรณ์ได้ดังนี้

$$(1 - \theta_1 B)(1 - \Theta_1 B^{12})Y_t = \theta_0 \text{ เมื่อ } \theta_0 = \text{ค่าคงที่}$$

$$(1 - \theta_1 B - \Theta_1 B^{12} + \theta_1 B \Theta_1 B^{12})Y_t = \theta_0$$

$$Y_t - \theta_1 Y_{t-1} - \Theta_1 Y_{t-12} + \theta_1 \Theta_1 Y_{t-13} = \theta_0$$

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \Theta_1 Y_{t-12} - \theta_1 \Theta_1 Y_{t-13}$$

$$Y_t = \theta_0 + 0.73Y_{t-1} + 0.64Y_{t-12} - 0.47Y_{t-13}$$

จากค่าสังเกตของอนุกรมเวลา ทำให้ทราบว่าค่าคงที่ $\theta_0 = 92.42$ ดังนั้น ได้สมการพยากรณ์จากตัวแบบ ARIMA เป็นดังนี้

$$Y_t = 92.42 + 0.73Y_{t-1} + 0.64Y_{t-12} - 0.47Y_{t-13}$$

เมื่อนำตัวแบบ ARIMA พยากรณ์ค่าสังเกต 5 คาบเวลา ตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม 2558 จากนั้น นำไปเปรียบเทียบกับค่าสังเกตจริงในช่วงเวลาเดียวกันของชุดข้อมูลตรวจสอบ เพื่อวัดค่า RMSE และ MAPE จะได้ผลลัพธ์เป็นดังตารางที่ 5

```

Console D:/R/code/
>
> auto.arima(arimaTS)
Series: arimaTS
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
      ar1      sar1  intercept
  0.7314  0.6425  929.9975
s.e.  0.0833  0.0980   57.5709

sigma^2 estimated as 3300:  log likelihood=-331.78
AIC=671.56  AICC=672.29  BIC=679.94
> |
    
```

รูปที่ 10 ผลลัพธ์ของการใช้ฟังก์ชัน auto.arima()

ตารางที่ 5 ค่าพยากรณ์และค่าคลาดเคลื่อนของตัวแบบ ARIMA

Value	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	RMSE	MAPE
Actual Value	745.82	853.60	1017.89	1100.93	1216.77	-	-
ARIMA	799.24	859.33	969.40	1065.12	1117.64	57.18	4.80

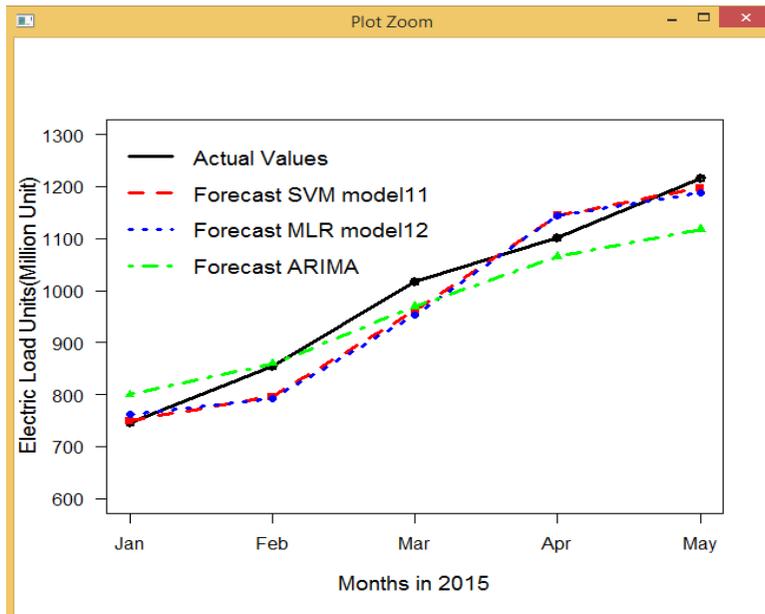
4. เปรียบเทียบผลการวิจัยของตัวแบบ SVR ตัวแบบ MLR และตัวแบบ ARIMA

จากการทดลองสร้างตัวแบบพยากรณ์อนุกรมเวลาของหน่วยจำหน่ายไฟฟ้าในเขตพื้นที่กรุงเทพมหานคร ของประเทศไทย พบว่า ตัวแบบพยากรณ์ที่มีความแม่นยำมากที่สุด คือ ตัวแบบที่สร้างจากเทคนิค SVR ซึ่งวัดความแม่นยำของการพยากรณ์

ด้วยการวัดค่า RMSE และ MAPE โดยตัวแบบ SVR ที่มีความแม่นยำมากที่สุด คือ ตัวแบบที่ 11 ซึ่งเป็นตัวแบบที่มีอินพุตเป็นค่าสังเกตล้าหลังตั้งแต่ 1 ถึง 11 คาบเวลา โดยวัดค่า RMSE = 42.46 และค่า MAPE = 3.71 ส่วนค่าพยากรณ์และค่าคลาดเคลื่อนของ 3 ตัวแบบ สรุปได้ดังตารางที่ 6 จากนั้น นำมาสร้างกราฟเส้นเปรียบเทียบกับค่าสังเกตจริงในช่วงเวลาเดียวกัน ได้ผลดังรูปที่ 11

ตารางที่ 6 ค่าพยากรณ์และค่าคลาดเคลื่อนของ 3 ตัวแบบ

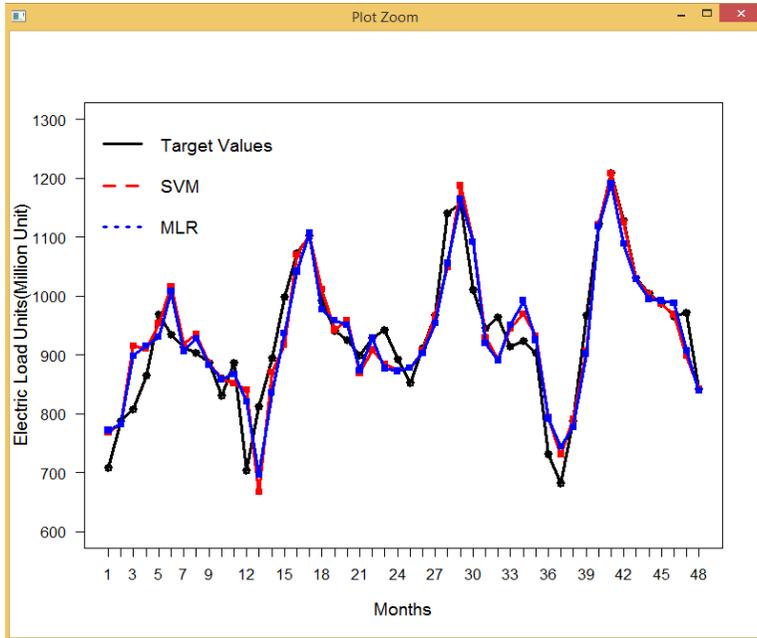
Value	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	RMSE	MAPE
Actual Value	745.82	853.60	1017.89	1100.93	1216.77	-	-
SVR model 11	749.60	795.30	960.09	1144.16	1197.00	42.46	3.71
MLR model 12	762.18	792.08	952.35	1143.28	1188.78	46.74	4.40
ARIMA	799.24	859.33	969.40	1065.12	1117.64	57.18	4.80



รูปที่ 11 กราฟเปรียบเทียบค่าพยากรณ์ของ 3 ตัวแบบ

จากรูปที่ 11 พบว่า กราฟค่าพยากรณ์ของตัวแบบ SVR และตัวแบบ MLR มีลักษณะใกล้เคียงกันมาก นอกจากนี้ เมื่อนำค่าสังเกตจริงของชุดข้อมูลฝึกสอน และค่าเออร์ทพุต ที่ได้จากขั้นตอนการฝึกสอนของตัวแบบ SVR และ MLR มาสร้างกราฟตามลำดับเวลา ดังรูปที่ 12 พบว่า กราฟเออร์ทพุตของตัวแบบ SVR และ MLR มีลักษณะใกล้เคียงกันด้วย ดังนั้น จากผลการทดลอง แสดงให้เห็นว่า เทคนิคการพยากรณ์

ด้วยตัวแบบ SVR ที่มีเคอร์เนลฟังก์ชันเป็นลิเนียร์เคอร์เนล มีความคล้ายคลึงกับเทคนิค MLR แต่ตัวแบบ SVR มีค่าคลาดเคลื่อนต่ำกว่า และใช้จำนวนตัวแปรที่ใช้เป็นอินพุตน้อยกว่า ซึ่งแสดงให้เห็นว่า เทคนิค SVR เป็นเทคนิคที่เป็นทางเลือกที่ดี สำหรับนำมาพยากรณ์อนุกรมเวลา โดยให้ความแม่นยำของการพยากรณ์ในระดับสูง



รูปที่ 12 กราฟเปรียบเทียบเอาท์พุทของตัวแบบ SVR และ MLR ของชุดข้อมูลฝึกสอน

สรุปและเสนอแนะ

จากการศึกษาการวิเคราะห์อนุกรมเวลา เพื่อพยากรณ์ข้อมูลอนุกรมเวลา ซึ่งเป็นข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง (กฟน.) แห่งประเทศไทย โดยเก็บรวบรวมข้อมูลจากรายงานสถานการณ์การจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง ตั้งแต่เดือนมกราคม 2553 ถึงเดือนพฤษภาคม 2558 และข้อมูลที่สนใจ มีขอบเขตเฉพาะจำนวนหน่วยจำหน่ายไฟฟ้าให้กับผู้บริโภคประเภทบ้านอยู่อาศัย โดยสร้างตัวแบบการพยากรณ์ด้วยเทคนิคที่แตกต่างกัน 3 วิธี ได้แก่ ตัวแบบ SVR ตัวแบบ MLR และตัวแบบ ARIMA ซึ่งตัวแบบ ARIMA เป็นตัวแบบพยากรณ์อนุกรมเวลาเชิงเดี่ยว ส่วนตัวแบบ SVR และตัวแบบ MLR เป็นตัวแบบที่มีชุดข้อมูลในรูปแบบของอินพุตเวกเตอร์ และทาร์เก็ต ($D = \{(x_i, y_i)\}_{i=1}^n$) โดยอินพุตเวกเตอร์ คือ ค่าสังเกตล้าหลังตั้งแต่ 1-12 คาบเวลา ($y_{t-1}, y_{t-2}, \dots, y_{t-12}$) ส่วนทาร์เก็ต คือ ค่าสังเกต (y_t) ณ เวลา t ในส่วนของการสร้างตัวแบบ SVR ได้สร้างตัวแบบขึ้นมา 12 ตัวแบบ จากนั้นคัดเลือกตัวแบบที่ดีที่สุดด้วยวิธีการตรวจสอบสลับ 3 ส่วน ผลการ

ทดลอง พบว่า ตัวแบบ SVR ที่ดีที่สุด คือตัวแบบที่ 11 ซึ่งมีอินพุต 11 ตัว คือ ค่าสังเกตล้าหลังตั้งแต่ 1 ถึง 11 คาบเวลา และในส่วนของกรการสร้างตัวแบบ MLR ได้สร้างตัวแบบขึ้นมา 12 ตัวแบบ จากนั้นเลือกตัวแบบที่ดีที่สุดด้วยการวัดค่าคลาดเคลื่อน RMSE และ MAPE กับชุดข้อมูลตรวจสอบ ผลการทดลอง พบว่า ถ้ามีจำนวนตัวแปรอิสระหรือตัวแปรอินพุตมากขึ้น จะทำให้ค่าคลาดเคลื่อนลดลง ดังนั้น ตัวแบบ MLR ที่ดีที่สุด คือตัวแบบที่ 12 มีตัวแปรอิสระ 12 ตัวแปร คือ ค่าสังเกตล้าหลังตั้งแต่ 1 ถึง 12 คาบเวลา ในส่วนสุดท้าย เมื่อนำค่าพยากรณ์ 5 ช่วงเวลา ของเดือนมกราคมถึงเดือนพฤษภาคม 2558 มาวัดค่าคลาดเคลื่อน RMSE และ MAPE ผลการทดลอง พบว่า ตัวแบบ SVR เป็นตัวแบบที่ให้ค่าคลาดเคลื่อนต่ำที่สุด คือ ค่า RMSE เท่ากับ 42.46 และค่า MAPE = 3.71 ดังนั้น ตัวแบบ SVR เป็นตัวแบบที่มีความแม่นยำในการพยากรณ์อนุกรมเวลาชุดนี้มากที่สุด นอกจากนี้ ผลการทดลอง พบว่า ตัวแบบ SVR ที่มีเคอร์เนลฟังก์ชันเป็นลิเนียร์เคอร์เนล และตัวแบบ MLR มีความคล้ายคลึงกันมาก แต่ตัวแบบ SVR ให้ความแม่นยำในการพยากรณ์สูง

กว่า ด้วยจำนวนตัวแปรอินพุตน้อยกว่า แสดงให้ทราบ ว่า ตัวแบบ SVR เป็นตัวแบบที่สามารถนำมาใช้ในการพยากรณ์อนุกรมเวลาได้ดี แต่ถ้าเปรียบเทียบในด้านความซับซ้อนของตัวแบบ SVR กับ ตัวแบบ ARIMA พบว่า ตัวแบบ SVR เป็นตัวแบบที่มีความซับซ้อนมากกว่าตัวแบบ ARIMA เพราะใช้อินพุต 11 ตัว ในขณะที่ ตัวแบบ ARIMA ใช้อินพุต 3 ตัว สำหรับการพยากรณ์ ดังนั้น ข้อเสนอแนะจากงานวิจัยนี้ คือค้นหาวิธีการลดจำนวนอินพุตของ SVR เพื่อลดความซับซ้อนของตัวแบบ SVR และให้ความแม่นยำของการพยากรณ์ในระดับสูงเช่นเดิม

เอกสารอ้างอิง

- [1] Bunn, D.W. and Farmer, E.D. 1985. **Comparative Models for Electrical Load Forecasting**. New York: John Wiley and Sons.
- [2] Hong, W. C. 2009. "Electric load forecasting by support vector model". **Applied Mathematical Modelling**. 33: 2444–2454.
- [3] Ismail, Z., Yahya, A. and Mahpol, K. A. 2009. "Forecasting Peak Load Electricity Demand Using Statistics and Rule Based Approach". **American Journal of Applied Sciences**. 6(8): 1618-1625.
- [4] การไฟฟ้านครหลวงแห่งประเทศไทย, 2558. รายงานสถานการณ์การจำหน่ายไฟฟ้า. <http://www.meo.or.th/download/index.php#.20 สิงหาคม>.
- [5] 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.
- [6] 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.
- [7] Kaytez, F., Taplamacioglu, M. C., Cam, E. and Hardalac, F. 2015. "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines". **Electrical Power and Energy Systems**. 67: 431–438.
- [8] Ogcı, G., Demirel, O.F. and Zaim, S. 2012. "Forecasting electricity consumption with neural networks and support vector regression" **Procedia-Social and Behavioral Sciences**. 58: 1576-1585.
- [9] Wang, J., Li, L., Niu, D. and Tan, Z. 2012. "An annual load forecasting model based on support vector regression with differential evolution algorithm". **Applied Energy**. 94: 65-70.
- [10] Fan, G. F., Peng, L. L., Hong, W. C. and Fan, S. 2016. "Electric load forecasting by the SVR model with differential empirical mode decomposition and autoregression" **Neurocomputing**. 173: 958-970.
- [11] Vapnik, V. 1995. **The Nature of Statistical Learning Theory**. Springer-Verlag.
- [12] Bagheripour, P., Gholami, A., Asoodeh, M. and Asadi, M. V. 2015. "Support vector regression based determination of shear wave velocity". **Journal of Petroleum Science and Engineering**. 125: 95-99.
- [13] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C., and Lin, C.C. 2015. **Misc Functions of the Department of Statistics, Probability Theory Group**. <http://cran.r-project.org/web/packages/e1071/index.html>. 5 December.

- [14] Sajan, K.S., Kumar, V. and Tyagi, B. 2015. "Genetic algorithm based support vector machine for on-line voltage stability monitoring" **Electrical Power and Energy Systems**. 73: 200-208.
- [15] Box, G.E.P. and Jenkins, G. 1976. **Times series analysis forecasting and control**. San Francisco, CA: Holden-Day.
- [16] Bergmeir, C. and Benítez, J. M. 2012. "On the use of cross-validation for time series predictor evaluation". **Information Sciences**. 191: 192-213.