

การเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดอุบัติเหตุทางถนน โดยใช้ เทคนิคอนุกรมเวลา

Performance Comparison of the Road Occurrence Accidents Prediction Models using Time Series Techniques

ปัทมวิภา บัญรักษา^{1*} และ จาริ ทองคำ²

นิสิตสาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม¹

สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ ศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์จำนวนผู้ประสบอุบัติเหตุบนท้องถนน ตั้งแต่ เดือนมกราคมพ.ศ.2554 ถึง เดือนธันวาคม 2559 ของจังหวัดขอนแก่น โดยใช้ 5 เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลองคือ Linear regression (LR), Artificial Neural Network (ANN), Sequential Minimal Optimization for Regression (SMOreg), Support Vector Machine Regression (SVR) และ Gussian Process (GP) มาใช้ คณะผู้วิจัยได้ใช้หลักการ Sliding Window ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ เพื่อวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่า เพื่อวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่า mean absolute error (MAE) และ root mean square error (RMSE)

ผลการวิจัยพบว่า เทคนิค SVM มีประสิทธิภาพในการสร้างแบบจำลองในการพยากรณ์จำนวนผู้ประสบอุบัติเหตุบนท้องถนนที่มีค่า ซึ่งเป็นค่าความผิดพลาดต่ำสุดเมื่อเปรียบเทียบกับแบบจำลองที่สร้างจากเทคนิค LR, ANN, SMOreg และ Gussian Process

คำสำคัญ: การพยากรณ์, อุบัติเหตุ, อุบัติเหตุบนท้องถนน, อนุกรมเวลา,เหมืองข้อมูล

ABSTRACT

The purposes of the research were to study and comparison of effective models for forecasting number of road accidents. The data were collected from January 2010 to December 2016 at Khon Kaen province. Five techniques were used including Linear regression (LR), Artificial Neural Network (ANN), Sequential Minimal Optimization for Regression (SMOreg), Support Vector Machine Regression (SVR) and Gussian Process (GP). The Sliding Window method was utilized to segment data into learning and testing sets. The predictive efficiency of the models were measured by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

The research finding showed that SVM techniques is effective in predicting the number of road accident victims. This is the lowest error value when compared with LR, ANN, SMOreg and Gussian Process models.

Keywords: Accident, Road Accidents, Time Series, Data Mining

บทนำ

การเกิดอุบัติเหตุจราจรทางบกเป็นปัญหาที่สำคัญปัญหาหนึ่งของโลกและมีแนวโน้มความรุนแรงของการเกิดอุบัติเหตุจราจรทางบกเพิ่มมากขึ้น ซึ่งไม่เพียงจะก่อให้เกิดผลเสียต่อร่างกาย ชีวิต และทรัพย์สินแต่ยังสร้างความเสียหายทางเศรษฐกิจของประเทศ สำนักงานอำนวยการความปลอดภัย กรมทางหลวง รายงานว่ามีผู้ประสบอุบัติเหตุ 211,499 ครั้ง ทั่วทั้งประเทศ ในปี พ.ศ. 2560 ซึ่งมีจำนวนผู้บาดเจ็บ 240,899 ราย ทุพพลภาพ 711 และเสียชีวิต 7,074 ราย โดยเฉพาะจังหวัดขอนแก่นเป็นเมืองที่ขนส่งสินค้าเส้นใหญ่ขึ้นเพิ่มความถี่ของการเกิดอุบัติเหตุบนท้องถนน ในปี พ.ศ. 2560 มีผู้ประสบอุบัติเหตุเกิดขึ้น 3,364 ครั้ง เป็นจำนวน 4,811 ราย มีจำนวนผู้บาดเจ็บ 3,667 ราย ทุพพลภาพ 12 ราย และเสียชีวิต 152 ราย [1]

ข้อมูลอุบัติเหตุจราจรทางบก เป็นข้อมูลอนุกรมเวลา ซึ่งเป็นชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลา รายวันต่อเนื่องกัน ข้อมูลดังกล่าวสามารถวิเคราะห์ได้ด้วยการวิเคราะห์อนุกรมเวลา (Time Series Analysis) [2] ซึ่งเป็นการพยากรณ์ที่อาศัยข้อมูลในอดีตมาพยากรณ์ การเปลี่ยนแปลงของข้อมูล เมื่อเวลาเปลี่ยนการเกิดอุบัติเหตุจะมีแนวโน้มสูงขึ้นหรือลดลง เพื่อให้ทางกรมทางบก โรงพยาบาล ได้นำไปช่วยในการจัดเตรียมบุคลากรทางการแพทย์ เครื่องมือแพทย์ และอื่นๆ ซึ่งการวิเคราะห์อนุกรมเวลาโดยทั่วไปใช้การวิเคราะห์ถดถอย (Linear Regression)

ปัจจุบันเทคนิคในเหมืองข้อมูลเข้ามามีบทบาทในการวิเคราะห์ข้อมูลอนุกรมเวลา เช่น เทคนิค Linear regression (LR), Artificial Neural Network (ANN), Sequential Minimal Optimization for Regression (SMOreg), Support Vector Machine Regression (SVR) และ Gaussian Process (GP) ผู้วิจัยสนใจศึกษาการประยุกต์เทคนิคการพยากรณ์ที่เหมาะสมการวิเคราะห์อนุกรมเวลาในการพยากรณ์พื้นที่จังหวัดขอนแก่นเพื่อเปรียบเทียบสมรรถนะของแบบจำลองการเกิดอุบัติเหตุจากอุบัติเหตุทางถนนในพื้นที่จังหวัดขอนแก่น คณะผู้วิจัยได้ใช้หลักการ Sliding Window ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ และวัดประสิทธิภาพการพยากรณ์ของแบบจำลอง เพื่อหาเทคนิคการพยากรณ์ที่มีความแม่นยำมากที่สุดมาใช้ในการพัฒนาแบบจำลอง

1. วัตถุประสงค์การวิจัย

- 1.1 เพื่อศึกษาเทคนิคอนุกรมเวลาที่เหมาะสมในการเปรียบเทียบสมรรถนะของแบบจำลองการเกิดอุบัติเหตุทางถนน
- 1.2 เพื่อเปรียบเทียบประสิทธิภาพผลการพยากรณ์การเกิดอุบัติเหตุบนท้องถนน

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 เอกสารที่เกี่ยวข้อง

ข้อมูลอนุกรมเวลา (Time Series data) คือ ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วง ๆ อย่างต่อเนื่องกัน เช่น ข้อมูลยอดขายสินค้าที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลาย ๆ เดือน ข้อมูลรายได้ ประชาชาติปีต่าง ๆ ที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลาย ๆ ปี เป็นต้น ข้อมูลอนุกรมเวลาอาจอยู่ในลักษณะที่เป็นข้อมูลรายปี รายไตรมาส หรือรายเดือนก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์ เนื่องจากข้อมูลทางธุรกิจมีการเปลี่ยนแปลงตลอดเวลา ผู้นำทางธุรกิจหรือองค์กรต้องหาวิธีพัฒนาต่าง ๆ ที่สามารถนำไปใช้ประกอบการตัดสินใจวางแผน เกี่ยวกับผลที่เกิดจากความเปลี่ยนแปลงในการดำเนินการอย่างใดอย่างหนึ่ง ดังนั้นการวิเคราะห์อนุกรมเวลาจึงเข้ามามีบทบาทช่วยในการตัดสินใจ เทคนิคอย่างหนึ่งที่ใช้ช่วยในการควบคุมการดำเนินการในปัจจุบันและในการวางแผนความต้องการในอนาคต คือ การพยากรณ์ (forecasting) ซึ่งการพยากรณ์นั้นทำได้หลายวิธี แต่ละวิธีต่างมีเป้าหมายเดียวกัน คือ ทำนายเหตุการณ์ในอนาคต

เทคนิคการวิเคราะห์การถดถอย (Linear Regression) เป็นเทคนิคที่ใช้ในการวิเคราะห์การถดถอย 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) เป็นวิธีการทางสถิติอย่างหนึ่ง ที่ใช้ในการตรวจสอบลักษณะของความสัมพันธ์ระหว่างตัวแปร ตั้งแต่ 2 ตัวขึ้นไป โดยแบ่งเป็นตัวแปรอิสระและตัวแปรตาม ซึ่งวรารพร งามสุข [3] ได้ทำการศึกษาเพื่อเปรียบเทียบการพยากรณ์อนุกรมเวลาด้วยวิธีการถดถอย โดยใช้ข้อมูลสถิติจำนวนการเกิดอุบัติเหตุในประเทศไทย ระหว่างเดือนมกราคม พ.ศ. 2545 ถึง เดือนธันวาคม พ.ศ. 2554 พบว่าค่าประสิทธิภาพที่ได้ RMSE เท่ากับ 51.59% ในงานวิจัยนี้ได้ใช้ Linear Regression (LR)

เทคนิค Artificial Neural Networks [4] เป็นศาสตร์แขนงหนึ่งทางด้านปัญญาประดิษฐ์ (Artificial Intelligence : AI) ที่สามารถนำไปประยุกต์ใช้กับงานหลายด้านได้อย่างมีประสิทธิภาพ หลักการสำคัญของโครงข่ายประสาทเทียม คือ ความพยายามที่จะลอกเลียนแบบการทำงานของเซลล์ประสาทในสมองมนุษย์เพื่อทำงานได้อย่างมีประสิทธิภาพ ลักษณะทั่วไปของโครงข่ายประสาทเทียม คือ การที่โหนด (node) ต่าง ๆ จำลองมาจากไซแนป (synapse) ของเซลล์ประสาทระหว่าง เดนไดรต์ (dendrite) และแอกซอน (axon) โดยมีฟังก์ชันเป็นตัวกำหนดสัญญาณส่งออก (activation function or transfer function) นั่นเอง ลักษณะของโครงข่ายประสาทเทียมสามารถแบ่งได้ 2 แบบ คือ 1) โครงข่ายประสาทเทียมแบบ ชั้นเดียว (single layer) ซึ่งจะมีเพียงชั้นสัญญาณประสาทขาเข้า และชั้นสัญญาณประสาทขาออก เท่านั้น เช่น โครงข่ายเพอเซปตรอนอย่างง่าย (simple perceptron) และโครงข่ายโฮปฟิลด์ (hopfield networks) เป็นต้น และ 2) โครงข่ายประสาทเทียมแบบหลายชั้น (multilayer) ซึ่งมีลักษณะเช่นเดียวกับโครงข่ายประสาทเทียมแบบชั้นเดียว แต่จะมีชั้นแอบแฝง (hidden) เพิ่มขึ้น โดยอยู่ส่วนกลางระหว่างชั้นนำข้อมูลป้อนเข้าและชั้นส่งข้อมูลออก ทั้งนี้ชั้นแอบแฝงอาจมีมากกว่า 1 ชั้น ในงานวิจัยนี้ได้ใช้ multilayer ในการสร้างแบบจำลอง

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

เทคนิค Support Vector Machine (SVM) เป็นเทคนิคการจำแนกประเภทของข้อมูลที่มีพื้นฐานมาจากทฤษฎีการเรียนรู้จากสถิติ คล้ายเทคนิคโครงข่ายประสาทเทียม แต่ SVM ใช้หลักการลดค่าความเสี่ยงเชิงโครงสร้างให้ต่ำสุด (Structural Risk Minimization) เพื่อลดค่าความผิดพลาดของการทำนาย (Minimized Error) พร้อมกับเพิ่มระยะการแบ่งแยกให้มากที่สุด (Maximized Margin) หลักการของ SVM คือการหาสมมติของสมการเพื่อสร้างเส้นแบ่งแยกประเภทข้อมูล ทำโดยการเลือกเส้นหรือระนาบแบ่งแยกประเภทข้อมูลที่เหมาะสมที่สุด A Library for Support Vector Machines (LibSVM) [5] เป็นไลบรารีสำหรับ SVM ที่พัฒนาโดย Chih-Chung Chang และ Chih-Jen Lin ซึ่ง LibSVM นี้จะช่วยให้การวิเคราะห์ SVM นั้นง่ายขึ้น โดย LibSVM ได้รองรับซอฟต์แวร์เวกเตอร์แบบการแบ่งกลุ่ม (Support Vector Classification: SVC) แบบการวิเคราะห์แบบรีเกรสชัน (Support Vector Regression: SVR) Sven และ Rohit [6] ได้สร้างแบบจำลองเพื่อการพยากรณ์ข้อมูลอนุกรมเวลาในรูปแบบฤดูกาลจำนวน 8 ชุดข้อมูลด้วยโครงข่ายประสาทเทียม โดยทดลองกับโมเดลโครงข่ายประสาทเทียมแบบต่างๆ แสดงให้เห็นว่าโครงข่ายประสาทเทียมมีประสิทธิภาพในการแก้ปัญหการพยากรณ์อนุกรมเวลาได้เป็นอย่างดี

เทคนิค Gaussian Processes (GP) [7] เป็นเทคนิคที่นำเอาระบบ Stochastic Process ซึ่ง ปัจจุบันถูกยอมรับแล้วว่าเป็นเครื่องมือที่มีประสิทธิภาพในการแก้ปัญหการพยากรณ์ Regression, Classification, และ Decision ใน Machine learning สามารถทำงานได้ดีถึงแม้ว่ามี Training Data น้อยและมีประสิทธิภาพและ Convergence rate ดีกว่า ARMA, NN และ SVR ตามลำดับ[8] ซึ่งมีงานวิจัยของทัศนีย์ พลอยสุวรรณและคณะ[9] นำเสนอการ

พยากรณ์หาค่าความต้องการไฟฟ้าสูงสุด (Peak load) ระยะยาวระหว่างปี 2011-2012โดยใช้ Gaussian Process ผลการพยากรณ์พบว่าให้ค่าความผิดพลาดน้อยและมีประสิทธิภาพดีกว่าที่ศึกษาเมื่อเปรียบเทียบกับกับการใช้โครงข่ายประสาทเทียม (Neural Networks)

2.2 งานวิจัยที่เกี่ยวข้อง

พงษ์เทพ วิวรรณธนะเดช และ คณะ [10] ได้ให้ความหมายของคำว่า อนุกรมเวลา ไว้ว่า เซตของข้อมูลเชิงปริมาณที่จัดเก็บในช่วงเวลาหนึ่ง ข้อมูลอนุกรมเวลาอาจอยู่ในลักษณะที่เป็นรายปี รายไตรมาส หรือรายเดือนก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์ การวิเคราะห์อนุกรมเวลา[11]เป็นกระบวนการของการใช้เทคนิคทางสถิติเพื่อสร้างแบบจำลองและอธิบายชุดข้อมูลจุดขึ้นอยู่กับเวลา การคาดการณ์ชุดอนุกรมเวลาเป็นกระบวนการของการใช้แบบจำลองเพื่อสร้างการคาดการณ์ สำหรับเหตุการณ์ในอนาคตจากเหตุการณ์ในอดีตที่ผ่านมา ข้อมูลชุดเวลามีลำดับเวลาตามธรรมชาติซึ่งแต่ละข้อมูลเป็นตัวอย่างที่มีแนวคิดที่จะเรียนรู้และลำดับข้อมูล

Zainal และ Mustaffa [12] ได้นำเอาเทคนิค Artificial Neural Network มาใช้ในสร้างแบบจำลองเพื่อการพยากรณ์ราคาทองคำ โดยใช้ให้ค่า RMSE และ MAE ในการแสดงประสิทธิภาพของแบบจำลอง ผลปรากฏว่าเทคนิค Artificial Neural Network สามารถสร้างแบบจำลองที่มีประสิทธิภาพต่ำสุดแต่ไม่มีนัยสำคัญของความแตกต่างจากเทคนิคอื่น

Yang, Zhai, Xu และ Han [13] ได้ใช้ SMO ในสร้างแบบจำลองเพื่อการพยากรณ์จำนวนของจุดยอดบนดวงอาทิตย์ ของ Wolfer ผลปรากฏว่า เทคนิค SMO ดีกว่า SVM และ Neural Network

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

- 1.1 โปรแกรมที่ใช้ในการวิเคราะห์ค่าสถิติการพยากรณ์ ใช้โปรแกรม WEKA 3.9.1
- 1.2 เทคนิคที่นำมาพยากรณ์ จำนวน 5 เทคนิค ได้แก่ ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย (Regression Analysis) วิธีและโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เทคนิค Sequential Minimal Optimization for Regression (SMOreg) เทคนิคซัพพอร์ตเวกเตอร์แมชชีนส์ (SVM) และเทคนิค Gaussian Process (GP)

2. ขั้นตอนการดำเนินการวิจัยประกอบด้วย 4 ขั้นตอน ดังนี้

- 2.1 เก็บรวบรวมข้อมูลการเกิดของอุบัติเหตุระหว่าง เดือนมกราคมพ.ศ.2554 ถึง เดือนมิถุนายน 2560 ในพื้นที่จังหวัดขอนแก่น
- 2.2 จัดเตรียมข้อมูล โดยจัดเรียงข้อมูลตามช่วงเวลาเดือนตามปัจจัยที่เกี่ยวข้องโดยข้อมูลได้ถูกเตรียมข้อมูลใน Excel
- 2.3 สร้างแบบจำลอง โดยการทำการทดสอบด้วยเทคนิค Linear regression (LR), Artificial Neural Network (ANN), Sequential Minimal Optimization for Regression (SMOreg), Support Vector Machine Regression (SVR) และ Gaussian Process (GP)
- 2.4 วัดประสิทธิภาพ ด้วยการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง ของทั้ง 5 เทคนิค

3. ข้อมูลที่ใช้ในการวิจัย

ในงานวิจัยฉบับนี้ใช้ข้อมูลการเกิดของอุบัติเหตุรายเดือนในพื้นที่จังหวัดขอนแก่น โดยใช้ข้อมูล ย้อนหลัง ตั้งแต่ เดือนมกราคม 2554 - มิถุนายน 2560 จำนวน 7 ปี รวมทั้งหมด 78 เดือน

4. สถิติที่ใช้ในการวิจัย

การวิเคราะห์ประสิทธิภาพของแบบจำลองในงานวิจัยนี้ผู้วิจัยได้ใช้หลักการหน้าต่างเคลื่อนที่ (Sliding Window) คือ หลักการแบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอน หรือการเรียนรู้ (Training data) และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) รอบแรกข้อมูล 3 ปีแรก (พ.ศ. 2554 - พ.ศ. 2556) เป็นชุดข้อมูลฝึกสอน และให้หกเดือนเป็นชุดข้อมูลทดสอบ (เดือนมกราคม-มิถุนายน พ.ศ. 2555) แล้วจึงทำการเคลื่อนหน้าต่างข้อมูลไปรอบละ 6 เดือนรวมทั้งสิ้น 7 รอบ แล้วจึงมีการวัดค่าประสิทธิภาพของแบบจำลอง ด้วยค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute error: MAE) และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)

ผลการวิจัย

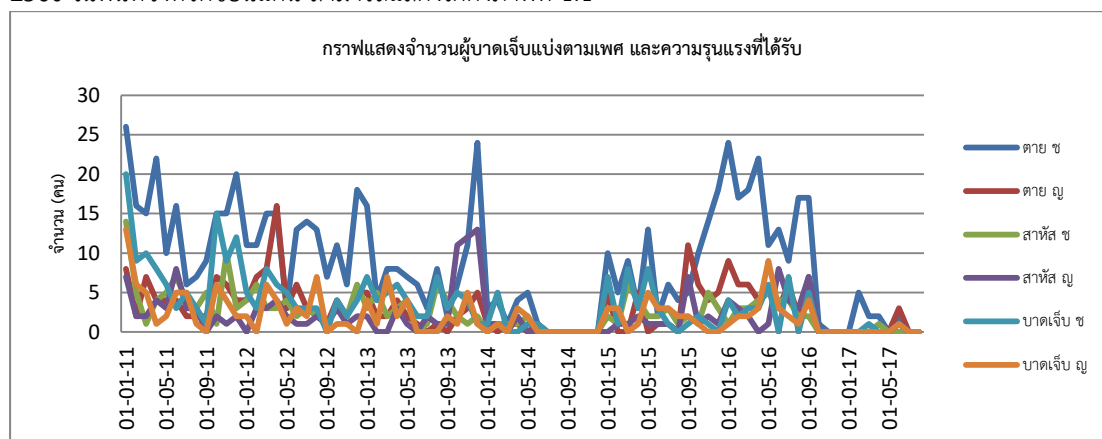
ผลการวิจัยในฉบับนี้ ผู้วิจัยได้ทำการศึกษาข้อมูล แล้วจึงใช้โปรแกรม WEKA 3.9.1 ในการทำการทดลอง และวัดค่าประสิทธิภาพของแบบจำลองการเกิดอุบัติเหตุทางถนน ด้วยเทคนิคอนุกรมเวลา 5 เทคนิคได้แก่

- 1.1 เทคนิคการถดถอยเชิงเส้น (Linear Regression: LR)
- 1.2 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)
- 1.3 เทคนิค Sequential Minimal Optimization for Regression (SMOreg)
- 1.4 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนส์ (SVR)
- 1.5 เทคนิค GaussianProcess (GP)

โดยการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute error: MAE) และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)

1. ผลการศึกษาเทคนิคอนุกรมเวลาที่เหมาะสมในการเปรียบเทียบสมรรถนะของแบบจำลองการเกิดอุบัติเหตุทางถนน

จากการศึกษาการเกิดอุบัติเหตุการเกิดของอุบัติเหตุระหว่างเดือน มกราคม พ.ศ. 2554 ถึง เดือนมิถุนายน 2560 ในพื้นที่จังหวัดขอนแก่น สามารถแสดงได้ดังภาพที่ 1.1

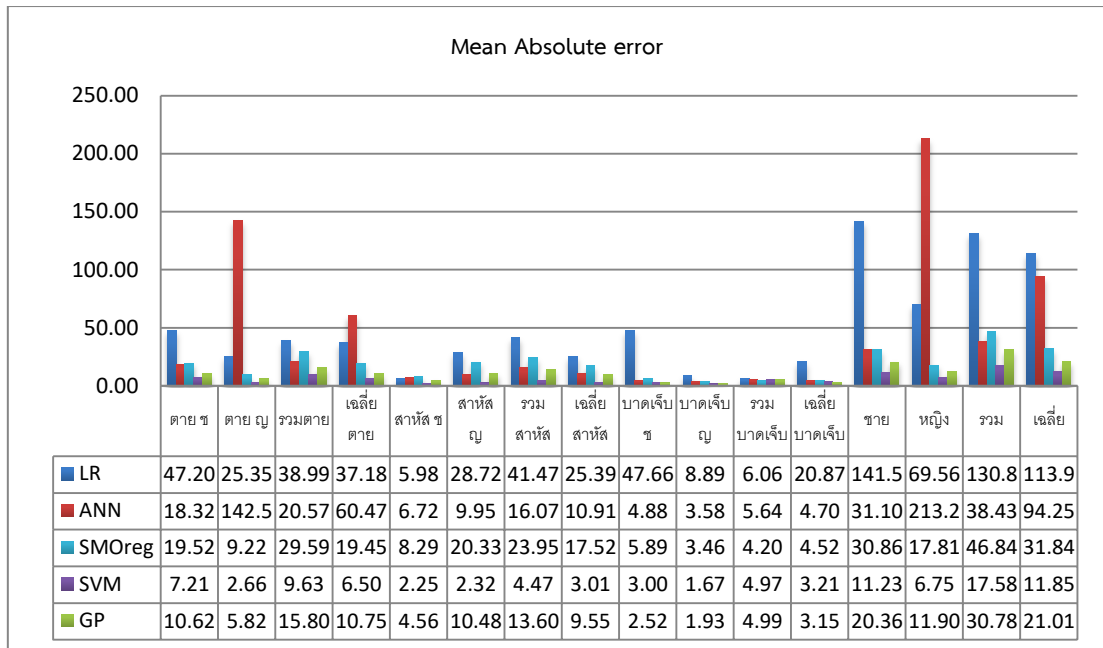


ภาพที่ 1 ข้อมูลผู้ประสบเหตุ แบ่งเพศ และระดับความรุนแรงที่เกิดขึ้น

จากภาพที่ 1 พบว่าข้อมูลอยู่ในลักษณะไม่เส้นตรง (Non-Linear) จากการเกิดอุบัติเหตุในแต่ละเดือนโดยแยกตามเพศและความรุนแรง

2. ผลการเปรียบเทียบประสิทธิภาพผลการพยากรณ์การเกิดอุบัติเหตุบนท้องถนน

การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAE) ซึ่งเป็นค่าความแตกต่างโดยเฉลี่ยระหว่างการเกิดอุบัติเหตุทางถนนจากการพยากรณ์ และเกิดอุบัติเหตุทางถนนจริง สามารถแสดงได้ภาพที่ 2

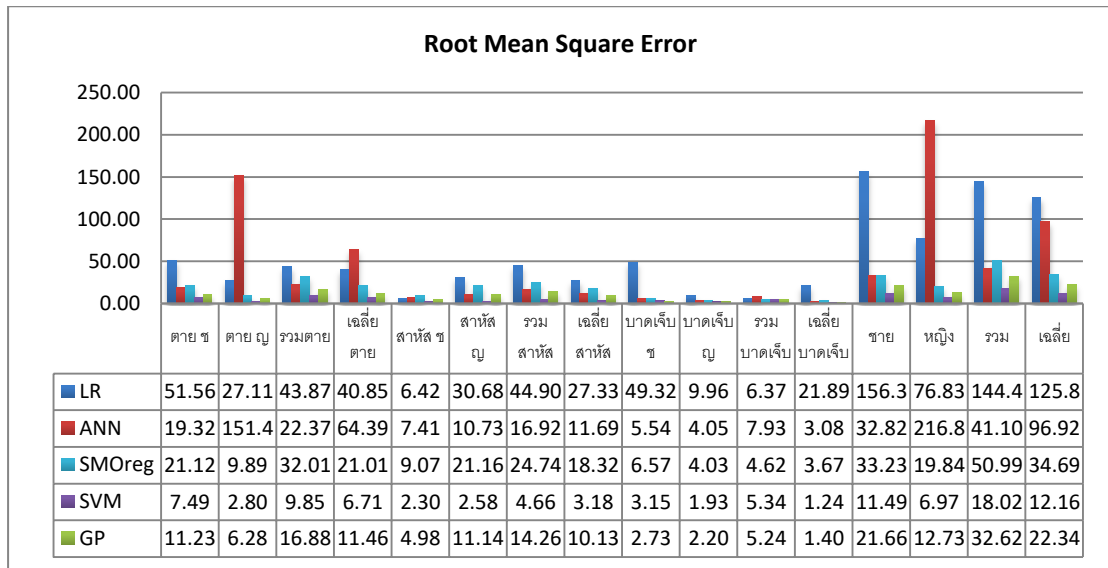


ภาพที่ 2 แสดงค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAE) ของแต่ละเทคนิค

จากภาพที่ 2 จะพบว่าค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute error: MAE) เป็นค่าที่ใช้วัดความถูกต้องของการพยากรณ์ที่นิยมใช้กันอย่างแพร่หลาย จากการทดลองสามารถแสดงผลการเปรียบเทียบของค่า MAE ของ 5 แบบจำลอง LR, ANN, SMOreg, SVM และ GP ในหัวข้อนี้ได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองเพื่อพยากรณ์จำนวนผู้บาดเจ็บ บาดเจ็บสาหัส และเสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนน ของเพศชาย (M) เพศหญิง (F) และรวมทั้งชายและหญิง ดัง ภาพที่ 1 ในการพยากรณ์จำนวนผู้บาดเจ็บจากการเกิดอุบัติเหตุบนท้องถนนของเพศชาย (M) เพศหญิง (F) และรวมทั้งชายและหญิง พบว่า SVM สามารถพยากรณ์แล้วให้ค่า MAE ต่ำสุดเพียง 3.00 ในการพยากรณ์การบาดเจ็บของชาย 1.67 การบาดเจ็บของเพศหญิง และ 4.97 การบาดเจ็บโดยรวม โดยมีค่าเฉลี่ยรวมเป็น 3.21 มีค่า 2.25 ในการพยากรณ์การบาดเจ็บสาหัสเพศชาย 2.32บาดเจ็บสาหัสเพศหญิง 4.47 บาดเจ็บสาหัสรวม โดยมีค่าเฉลี่ยรวมเป็น 3.01 และ 7.21 การพยากรณ์การเสียชีวิตเพศชาย 2.66 ในการเสียชีวิตเพศหญิง 9.62 การเสียชีวิตรวมทั้งชายและหญิง โดยมีค่าเฉลี่ยรวมเป็น 6.50

2. ผลการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง ของแบบจำลองการเกิดอุบัติเหตุทางถนน ของจังหวัดขอนแก่นโดยใช้เทคนิคอนุกรมเวลา

การเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง ซึ่งเป็นการเปรียบเทียบความแตกต่างระหว่างค่าพยากรณ์และค่าจริง ของการเกิดอุบัติเหตุทางถนนเฉลี่ยกำลังสอง ซึ่งถ้าค่าที่ได้เท่ากับ 0 แสดงว่าค่าพยากรณ์ตรงกับค่าความเป็นจริงที่สุดซึ่งหมายถึงดีที่สุด จากการทดลองสามารถแสดงผลการทดลองได้ดังภาพที่ 3



ภาพที่ 3 ความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) ของแต่ละเทคนิค

ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เป็นวิธีการวัดความคลาดเคลื่อนจากค่าที่พยากรณ์จากแบบจำลองกับค่าจริงที่เกิดขึ้น ในหัวข้อนี้ได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองเพื่อพยากรณ์จำนวนผู้บาดเจ็บ บาดเจ็บสาหัส และเสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนนโดยใช้ 5 เทคนิค แสดงค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) ซึ่งเป็นค่าที่ใช้วัดประสิทธิภาพของ 5 แบบจำลองรวมถึง LR, ANN, SMOreg, SVM และ GP ในหัวข้อนี้ได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองเพื่อพยากรณ์จำนวนผู้บาดเจ็บ บาดเจ็บสาหัส การเสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนน ของเพศชาย (M) เพศหญิง (F) และรวมทั้งชายและหญิง ดังภาพที่ 3 ในการพยากรณ์จำนวนผู้บาดเจ็บจากการเกิดอุบัติเหตุบนท้องถนนของเพศชาย (M) เพศหญิง (F) และรวมทั้งชายและหญิง พบว่า SVM สามารถพยากรณ์แล้วให้ค่า RMSE ต่ำสุดเพียง 3.15 ในการพยากรณ์การบาดเจ็บของชาย 1.93 การบาดเจ็บของเพศหญิง และ 5.34 การบาดเจ็บโดยรวม โดยมีค่าเฉลี่ยรวมเป็น 1.24 มีค่า 2.30 ในการพยากรณ์การบาดเจ็บสาหัสเพศชาย 2.58 บาดเจ็บสาหัสเพศหญิง 4.66 บาดเจ็บสาหัสรวม โดยมีค่าเฉลี่ยบาดเจ็บรวมเป็น 3.18 และ 7.49 ในการพยากรณ์การเสียชีวิตเพศชาย 2.80 เสียชีวิตเพศหญิง 9.85 การเสียชีวิตรวมทั้งชายและหญิง โดยมีค่าเฉลี่ยรวมเป็น 6.71

อภิปรายผลการวิจัย

ผู้วิจัยได้ศึกษาแบบจำลองที่เหมาะสมสำหรับการพยากรณ์การเกิดอุบัติเหตุทางถนน โดยใช้เทคนิคอนุกรมเวลา โดยใช้ 5 เทคนิค ผลการเปรียบเทียบเทคนิคการพยากรณ์การเกิดอุบัติเหตุโดยใช้ด้วยค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute error: MAE) และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) พบว่า ส่วนเทคนิค LR มีค่าของการบาดเจ็บเฉลี่ย 20.87, 21.89 บาดเจ็บสาหัสเฉลี่ย 25.39, 27.33 และการเสียชีวิตเฉลี่ยอยู่ที่ 37.18, 40.85, เทคนิค ANN ค่าของการบาดเจ็บเฉลี่ย 4.70, 3.08 บาดเจ็บสาหัสเฉลี่ย 10.91, 11.69 และการเสียชีวิตเฉลี่ยอยู่ที่ 60.47, 64.39, เทคนิค SMOreg ค่าของการบาดเจ็บเฉลี่ย 4.52, 3.67 บาดเจ็บสาหัสเฉลี่ย 17.52, 18.32 และการเสียชีวิตเฉลี่ยอยู่ที่ 19.45, 21.01, เทคนิค SVM มีค่าของการบาดเจ็บเฉลี่ย 3.21, 1.24 บาดเจ็บสาหัสเฉลี่ย 3.01, 3.18 และการเสียชีวิตเฉลี่ยอยู่ที่ 6.50, 6.71 และเทคนิค GP ค่าของการบาดเจ็บเฉลี่ย 3.15, 1.40 บาดเจ็บสาหัสเฉลี่ย 9.55, 10.13 และการเสียชีวิตเฉลี่ยอยู่ที่ 10.75, 11.46, วิธีการสรุปผลการวิเคราะห์พบว่า เทคนิค SVM ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean

Absolute error: MAE) และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง เป็นวิธีที่มีค่าความคลาดเคลื่อนต่ำที่สุด มีความเหมาะสมในการพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดอุบัติเหตุบนท้องถนนโดยใช้เทคนิคอนุกรมเวลา และ เพื่อใช้ในการวางแผนในการเดินทางและได้หลีกเลี่ยงการเกิดอุบัติเหตุ

ข้อเสนอแนะ

งานวิจัยนี้ได้มีการศึกษาเทคนิคการพยากรณ์จำนวน 5 เทคนิค โดยเทคนิคที่ทดสอบแล้วได้ค่าการพยากรณ์ที่คลาดเคลื่อนน้อยที่สุดในเอกสารฉบับนี้ หากมีการศึกษาเทคนิควิธีอื่นเข้ามาเปรียบเทียบเพิ่มเติม ทำให้ได้ค่าการพยากรณ์ที่ต่างกันในแต่ละเทคนิค โดยการนำข้อมูลการเกิดอุบัติเหตุโดยใช้ข้อมูลย้อนหลัง 7 ปี โดยแยกข้อมูลทดสอบเป็นรายเดือน ในงานวิจัยที่มีข้อมูลใกล้เคียงกันก็สามารถนำไปปรับใช้ให้เหมาะสมกับการพยากรณ์

เอกสารอ้างอิง

- [1] ศูนย์ข้อมูลอุบัติเหตุ. (2560). รายงานสถิติการใช้สิทธิ พรบ. รายจังหวัด. สืบค้นเมื่อ 23 กันยายน 2560, จาก <http://rvpreport.rvpeservice.com/viewrsc.aspx?report=0486&session=16>
- [2] ศูนย์วิจัยอุบัติเหตุแห่งประเทศไทย. (2552). การพัฒนาแบบจำลองการเกิดอุบัติเหตุ. สืบค้นเมื่อ 26 กันยายน 2560, จาก http://bhs.doh.go.th/files/Project/TACR/Accident%20Prediction%20Model_th.pdf
- [3] วราพร งามสุข. (2555). การพยากรณ์อนุกรมเวลาโดยการเปรียบเทียบวิธีแบบฉบับและวิธีบอกซ์-เจนกินส์ กรณีศึกษาจำนวนการเกิดอุบัติเหตุในประเทศไทย (ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์), มหาวิทยาลัยบูรพา.
- [4] ธนาวุฒิ ประกอบผล. (2552). โครงข่ายประสาทเทียม *Artificial Neural Networks* (Vol. 12).
- [5] Chih-Chung Chang และand Chih-Jen Lin. (April 2011). A Library for Support Vector Machines (LibSVM). *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3).
- [6] Sven F. Crone และRohit Dhawan. (2007). Forecasting Seasonal Time Series with Neural Networks: A Sensitivity Analysis of Architecture Parameters. *IEEE Trans.Neural Networks(IJCNN)*, 2099-2104.
- [7] Amine Bermark Sofiane Brahim-Belhouari. (2004). Gaussian process for nonstationary time series prediction. *Computational Statics & Data Analysis* 47(47), 705-712.
- [8] C. E. Rasmussen and C. K. I. Williams. (2006). Gaussian processes for machine learning. *Cambridge, Mass.: MIT Press*.
- [9] ทศนัย พลอยสุวรรณ และธนวันต์ ยลพันธ์. (2556). การพยากรณ์ค่าความต้องการใช้ไฟฟ้าสูงสุดของการไฟฟ้าฝ่ายผลิตแห่งประเทศไทยโดยกระบวนการเกาส์เซียน. *การประชุมและนำเสนอผลงานวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 36*.
- [10] พงษ์เทพ วิจารณ์ชนะเดช และคณะแพทย์ศาสตร์. (2550). การวิเคราะห์อนุกรมเวลา(*Time Series Analysis*)และระบบสารสนเทศทางภูมิศาสตร์(*GIS*)ในการพัฒนาองค์ความรู้ด้านสิ่งแวดล้อมและสุขภาพ. ศูนย์ประสานข้อมูลปัญหาหมอกควันในพื้นที่ภาคเหนือภาควิชาเวชศาสตร์ชุมชน คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่, pp.1-35.
- [11] Adaptavist Theme. (2560). *Time Series Analysis and Forecasting with Weka*. Retrieved 23 September 2017, from <http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka>
- [12] Mustaffa Z Zainal NA. (2016). Developing a gold price predictive analysis using Grey Wolf Optimizer. *IEEE Student Conference on Research and Development 2016*.
- [13] Jin-Fang Yang; Yong-Jie Zhai; Da-Ping Xu. (2007). SMO Algorithm Applied in Time Series Model Building and Forecast. *Pu Han 2007 International Conference on Machine Learning and Cybernetics*, 4, 2395 – 2400.