กีระชาติ สุขสุทธิ์ : การจำแนกข้อมูลไม่สมดุลโดยใช้การปรับปรุงข้อมูลร่วมกับการหาค่าพารามิเตอร์ที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรมที่มีการเริ่มต้นใหม่ (IMBALANCED DATA CLASSIFICATION USING DATA IMPROVEMENT AND PARAMETER OPTIMIZATION WITH RESTARTING GENETIC ALGORITHM) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.กิตติศักดิ์ เกิดประสพ, 152 หน้า.

การทำเหมืองข้อมูล เป็นกระบวนการหาองค์ความรู้จากข้อมูลที่มีขนาดใหญ่ เพื่อค้นหารูปแบบ หรือหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ภายในข้อมูลเหล่านั้น เพื่อที่จะนำความรู้ที่ได้รับนั้นมาสร้างโมเดลเพื่อใช้ในการทำนายข้อมูลในอนาคตหรือนำไปใช้จำแนกประเภทข้อมูลที่ยังไม่ทราบกลุ่ม แต่โดยทั่วไปแล้วอัลกอริทึมสำหรับการจำแนกประเภทข้อมูลจะมีประสิทธิภาพและมีความแม่นยำในการจำแนกประเภทข้อมูลสูงถ้าหากข้อมูลที่นำมาใช้ในการเรียนรู้มีจำนวนข้อมูลในแต่ละคลาสสมดุลกัน หรือใกล้เคียงกัน แต่เทคนิคในการจำแนกจะมีประสิทธิภาพในการจำแนกด้อยลงเมื่อข้อมูลที่นำมาใช้ในการเรียนรู้มีความไม่สมดุลเกิดขึ้น เนื่องจากอัลกอริทึมในการจำแนกอาจจะมีความเอนเอียงไปทางกลุ่มของข้อมูลที่มีจำนวนสมาชิกในคลาสนั้นมากกว่าอีกกลุ่ม

ดังนั้นงานวิจัยนี้จึงได้เสนอการปรับปรุงสมดุลข้อมูลร่วมกับการหาค่าพารามิเตอร์ที่เหมาะสมด้วยวิธีการเชิงพันธุกรรมที่มีการเริ่มต้นใหม่โดยได้นำ SMOTE Technique ร่วมกับวิธีการสุ่มลดข้อมูลมาใช้ในการปรับข้อมูลให้เกิดความสมดุล และใช้ขั้นตอนวิธีเชิงพันธุกรรมในการหาค่าพารามิเตอร์ที่เหมาะสมสำหรับสร้างโมเดลจำแนกประเภทข้อมูลด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน และได้ใช้เทคนิคการเริ่มต้นขั้นตอนวิธีเชิงพันธุกรรมใหม่เพื่อปรับปรุงประสิทธิภาพในการค้นหาค่าพารามิเตอร์ที่เหมาะสม การทดสอบประสิทธิภาพของวิธีการที่เสนอขึ้นใหม่นี้ ใช้การทดสอบเปรียบเทียบประสิทธิภาพในการจำแนกประเภทข้อมูลด้วยค่าความแม่นยำในการจำแนก (Accuracy) ค่าความเที่ยง (Precision) ค่าความไว (Sensitivity) และค่าการวัดเอฟ (F-measure) โดยผลการทดสอบพบว่าเทคนิคที่นำเสนอมีประสิทธิภาพในการจำแนกข้อมูลจากคลาสส่วนน้อยได้ดีกว่าเทคนิคอื่น ๆ

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา กีระชาติ

ปีการศึกษา 2559 ลายมือชื่ออาจารย์ที่ปรึกษา

KEERACHART SUKSUT : IMBALANCED DATA CLASSIFICATION USING DATA IMPROVEMENT AND PARAMETER OPTIMIZATION WITH RESTARTING GENETIC ALGORITHM. THESIS ADVISOR : ASSOC. PROF. KITTISAK KERDPRASOP, Ph.D., 152 PP.

IMBALANCED DATA CLASSIFICATION/SUPPORT VECTOR MACHINE/ RESTARTING GENETIC ALGORITHM/DATA IMPROVEMENT

Data mining is the process to find knowledge from the huge amount of stored information and use the discovered knowledge to predict or classify the new data item that its class label is unknown. Classification algorithms are normally efficient on classifying balanced data but show poor performance with imbalanced data. Most learning algorithms cannot classify the data within minority class correctly because the model that is learned from imbalanced data tends to bias toward the majority class. In this research we thus propose a hybrid technique for data balancing with random under sampling the data from majority class and synthetically generate the new data from minority class with SMOTE technique, and also finding the optimal parameter with restarting genetic algorithm for model learning with support vector machine algorithm. We test the performance of the proposed method for minority data classification with four measurements: accuracy, precision, recall, and f-measure. The results show that the proposed technique has high performance on classifying the data from minority class and outperforms other techniques.

School of Computer Engineering Student's Signature ____________

Academic Year 2016 Advisor's Signature ____________