

จักรกฤษณ์ พลรบ : ขั้นตอนวิธีการหาค่าเหมาะที่สุดแบบวาทแปลงสำหรับปรับปรุงสมดุล  
ข้อมูลอิงกลวิธีการเลือกตัวอย่างลด (A MODIFIED WHALE OPTIMIZATION ALGORITHM  
FOR IMPROVING DATA BALANCE BASED ON UNDERSAMPLING TECHNIQUES)  
อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.เบญจวรรณ โรจนดิษฐ์, 112 หน้า.

คำสำคัญ: ข้อมูลไม่สมดุล/การเลือกตัวอย่างลด/ขั้นตอนวิธีการหาค่าเหมาะที่สุดแบบวาท/ขั้นตอน  
วิธีการหาค่าเหมาะที่สุดแบบไบนารีวาท

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างขั้นตอนวิธีการเลือกตัวอย่างลดแบบใหม่ โดยประยุกต์ใช้  
ขั้นตอนวิธีการหาค่าเหมาะที่สุดแบบวาทและไบนารีวาทร่วมกับเทคนิคเพื่อนบ้านใกล้ที่สุด เพื่อ  
แก้ปัญหาข้อมูลไม่สมดุล ซึ่งมี 2 รูปแบบ คือ ตรึงค่าพารามิเตอร์เคเท่ากับหนึ่งและปรับ  
ค่าพารามิเตอร์เคได้ ทั้งนี้ได้เลือกชุดข้อมูลทดสอบจำนวน 10 ชุดจาก KEEL และ imbalanced-learn  
ซึ่งชุดข้อมูลจะมีอัตราส่วนความไม่สมดุลอยู่ในช่วง 1.82 ถึง 42.01 ขั้นตอนการทำงานเริ่มจากนำชุด  
ข้อมูลมาแบ่งเป็นสองชุด คือ ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบสำหรับข้อมูลชุดฝึกสอน ชั้นข้อมูล  
กลุ่มน้อยยังคงเก็บไว้เหมือนเดิม ในขณะที่ชั้นข้อมูลกลุ่มมากจะถูกวิเคราะห์เพื่อดึงชุดข้อมูลย่อยที่เป็น  
ตัวแทนที่ดีที่สุดด้วยขั้นตอนวิธีที่นำเสนอ โดยประเมินประสิทธิภาพด้วยเทคนิคดังนี้ เทคนิคซัพพอร์ต  
เวกเตอร์แมชชีน เทคนิคป่าสุ่ม และ เทคนิคต้นไม้ตัดสินใจ ผลการวิจัยพบว่าขั้นตอนวิธีที่นำเสนอแบบ  
ปรับค่าพารามิเตอร์เคได้ร่วมกับแบบจำลองเทคนิคป่าสุ่มมีประสิทธิภาพการทำงานโดยรวมสูงที่สุด  
โดยมีผลการวัดประสิทธิภาพเฉลี่ยดังนี้ Accuracy = 0.8387, F1 score = 0.5783, G-mean =  
0.8794, AUROC = 0.9212, AUPRC = 0.6457, Sensitivity = 0.9399, Precision = 0.4703,  
MCC = 0.5705, และ Kappa = 0.5123 โดยรวมแล้วขั้นตอนวิธีที่นำเสนอมีประสิทธิภาพดีกว่าสาม  
ขั้นตอนวิธีการเลือกตัวอย่างลดดังต่อไปนี้ คือ การลดจำนวนตัวอย่างข้อมูลแบบสุ่ม คลัสเตอร์เซน  
ทรอยด์ และ เนียร์มิส

สาขาวิชาคณิตศาสตร์  
ปีการศึกษา 2564

ลายมือชื่อนักศึกษา จักรกฤษณ์ พลรบ  
ลายมือชื่ออาจารย์ที่ปรึกษา Benjawan Rojanadit  
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม J. Tanthanud  
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม Cu

JAKKRIT POLROB : A MODIFIED WHALE OPTIMIZATION ALGORITHM FOR IMPROVING DATA BALANCE BASED ON UNDERSAMPLING TECHNIQUES. THESIS ADVISOR : ASST. PROF. BENJAWAN RODJANADID, Ph.D. 112 PP.

Keyword: IMBALANCED DATA, UNDERSAMPLING TECHNIQUE, WHALE OPTIMIZATION ALGORITHM, BINARY WHALE OPTIMIZATION ALGORITHM

This thesis is aimed at developing a novel undersampling algorithm, by combining the idea of whale and binary whale optimization algorithms with K-nearest neighbor classification for solving imbalanced data problems. There are two versions, one with fixed-parameter  $K = 1$ , and the other with adjustable parameter  $K$ . Ten datasets of varying imbalance ratios ranging from 1.82 to 42.01 were selected from the Knowledge Extraction based on Evolutionary Learning (KEEL) and imbalanced-learn repositories to be used in the evaluation of the novel algorithms. The work started by splitting each dataset into two parts, the training set and the testing set. Whereas the minority class of each training set remained untouched, the majority class was analyzed by the proposed algorithms to extract an optimal representative subset. Then, the support vector machine, random forest, and decision tree classifiers were trained with the new training set for assessing performance. It was found that the proposed algorithms applied to the random forest model had the highest overall performance, with average efficiency measurement results as follows: Accuracy = 0.8387, F1 score = 0.5783, G-mean = 0.8794, AUROC = 0.9212, AUPRC = 0.6457, Sensitivity = 0.9399, Precision = 0.4703, MCC = 0.5705, and Kappa = 0.5123. In particular, the proposed algorithm performed better overall than three common undersampling algorithms, namely random undersampling, cluster centroid, and near-miss algorithms.

School of Mathematics  
Academic Year 2021

Student's Signature Benjamin Rodjanadid  
Advisor's Signature Benjamin Rodjanadid  
Co-advisor's Signature J. Tautandi  
Co-advisor's Signature C. S.