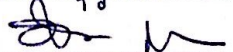


ศุภจิตร์ บุญอำนวยการ : การปรับปรุงประสิทธิภาพต้นไม้ตัดสินใจแบบจำแนกและแบบ
ถดถอยด้วยเทคนิคการสุ่มซ้ำ (IMPROVING THE CLASSIFICATION AND
REGRESSION TREE PERFORMANCE WITH RESAMPLING TECHNIQUE)
อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. นิตยา เกศประสพ, 63 หน้า.

ปัจจุบันการทำเหมืองข้อมูลได้รับความนิยมอย่างมากในด้านการวิเคราะห์ข้อมูล โดยการนำโมเดลที่ได้จากกระบวนการเรียนรู้มาใช้ในการจำแนกข้อมูล แบ่งกลุ่มข้อมูล หรือทำนายข้อมูลในอนาคต ในอัลกอริทึมเหมืองข้อมูลที่มีหลากหลายอัลกอริทึมต้นไม้ตัดสินใจแบบจำแนก และแบบถดถอย เป็นอัลกอริทึมที่มีจุดเด่นคือสามารถนำมาวิเคราะห์ข้อมูลได้ทั้งตัวเลข และข้อความ ทั้งยังมีประสิทธิภาพค่อนข้างสูงเมื่อนำมาวิเคราะห์ข้อมูลที่มีลักษณะไม่สมดุล ซึ่งเป็นลักษณะของข้อมูลส่วนใหญ่ในชีวิตจริง งานวิจัยนี้จึงเสนอวิธีการเพิ่มประสิทธิภาพอัลกอริทึมต้นไม้แบบจำแนก และแบบถดถอยด้วยการนำเทคนิคการสุ่มซ้ำเข้ามาช่วยในขั้นตอนการเตรียมข้อมูลด้วยการหาอัตราส่วนระหว่างจำนวนข้อมูลคลาสส่วนมากต่อจำนวนข้อมูลคลาสน้อย ซึ่งจากการทดลองปรากฏว่าเมื่ออัตราส่วนของสองคลาสเริ่มใกล้เคียงกัน ค่าประสิทธิภาพของการจำแนกข้อมูลยิ่งสูงขึ้น โดยเฉพาะเมื่ออัตราส่วนเป็น 50 : 50 และเนื่องด้วยข้อมูลของการวิจัยนี้ค่าของแอททริบิวต์ต่าง ๆ นอกเหนือจากคลาสเป้าหมายเป็นข้อมูลตัวเลขทั้งหมด จึงได้ทดลองนำเทคนิคการจัดกลุ่มข้อมูลเข้ามาช่วย โดยจัดกลุ่มข้อมูลคลาสส่วนมากให้เท่ากับจำนวนข้อมูลคลาสน้อย แล้วใช้ค่าเฉลี่ยของข้อมูลแต่ละแอททริบิวต์ในแต่ละกลุ่มเป็นข้อมูลตัวแทนของข้อมูลชุดนั้น ๆ ซึ่งจากวิธีการดังกล่าว สามารถสร้างแบบจำลองที่จำแนกคลาสน้อยได้ดียิ่งขึ้น

สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2562

ลายมือชื่อนักศึกษา ศุภจิตร์ บุญอำนวยการ
ลายมือชื่ออาจารย์ที่ปรึกษา 

SUPAJITTREE BOONAMNUAY : IMPROVING THE CLASSIFICATION
AND REGRESSION TREE PERFORMANCE WITH RESAMPLING
TECHNIQUE. THESIS ADVISOR : ASSOC. PROF. NITTAYA
KERDPRASOP, Ph.D., 63 PP.

CLASSIFICATION AND REGRESSION TREE/RESAMPLING TECHNIQUE

Data mining is a very popular method for data analysis by introducing models derived from the learning process for classification, clustering or predicting. From numerous data mining algorithms, Classification and Regression Tree (CART) algorithm is the prominent one with its advantage of being able to analyze data in both numeric and categorical forms. CART performance is also highly effective when analyze imbalanced data that is the normally found in real life. This research proposes to improve CART performance by using resampling technique to find the appropriate ratio between the number of majority class and the number of minority class that can best improve CART performance. When we try to randomly reduce the number of majority class to a close proportion of minority class, CART shows better discriminative performance, and the best ratio is 50:50. Since the value of all attributes other than the target class in this research is numeric, clustering technique is also used to help grouping majority class to be equal to minority class. The mean value of each attribute in each group is the representative data. From the applied technique, models can detect more of minority class and have good performance. Finally hopefully this research will be useful in the future.

School of Computer Engineering

Academic Year 2019

Student's Signature Supajittree Boonamnuy

Advisor's Signature Nittaya