

หทัยชนก กรชี 2551: การคำนวณน้ำหนักของคุณลักษณะข้อมูลสำหรับตัวจำแนกในอีฟเบย์ โดยวิธีการจัดกลุ่มข้อมูลและต้นไม้ตัดสินใจ ปริญญาวิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์) สาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ ปรธานกรรมการที่ปรึกษา: รองศาสตราจารย์อนงค์นาฏ ศรีวิหค, Ph.D. 92 หน้า

งานวิจัยฉบับนี้นำเสนอวิธีการคำนวณค่าน้ำหนักของคุณลักษณะข้อมูล คือ วิธีการแบ่งกลุ่มข้อมูล (Clustering) และต้นไม้ตัดสินใจ (Decision Tree) ซึ่งปรับปรุงจากงานวิจัยของ Hall เนื่องจากค่าน้ำหนักที่ใช้ในวิธีของ Hall เป็นค่าน้ำหนักแบบสากล (Global) จึงทำให้บางกรณีมีประสิทธิภาพการทำนายไม่ดีพอ จึงมีแนวคิดในการใช้ค่าน้ำหนักแบบท้องถิ่น (Local) ซึ่งพิจารณาจากค่าน้ำหนักคุณลักษณะที่เหมาะสมกับข้อมูลทดสอบ โดยการแบ่งกลุ่มข้อมูลตามคุณลักษณะข้อมูล สำหรับวิธีการแบ่งกลุ่มที่นำเสนอในการศึกษารั้งนี้จะใช้ 2 วิธี คือ (1) อัลกอริทึมเค-มีน (K-Means) และ (2) อัลกอริทึมสองขั้นตอนโดยใช้ต้นไม้การจัดกลุ่มตามคุณลักษณะ (Cluster Feature Tree) และวิธีการลำดับชั้นแบบ Agglomerative พร้อมทั้งเปรียบเทียบประสิทธิภาพและความแม่นยำของตัวจำแนกในอีฟเบย์แบบน้ำหนักจากวิธีการแบ่งกลุ่มข้อมูลและต้นไม้ตัดสินใจกับวิธีการต้นไม้ตัดสินใจของ Hall โดยวัดค่าความถูกต้องของการทำนายและค่า Root Relative Square Error (RRSE) งานวิจัยนี้ใช้ข้อมูลจากมหาวิทยาลัยแคลิฟอร์เนีย (University of California, Irvin (UCI)) จำนวน 4 ชุดข้อมูล ได้แก่ (1) Chess End-Game (KR) (2) Balance Scale Weight และ Distance Database (Bal_sc) (3) Wave Form และ (4) German Credit dataset (Crd)

จากผลการวิจัยพบว่า ตัวแปรที่มีผลต่อการเพิ่มประสิทธิภาพและความแม่นยำให้แก่วิธีการคำนวณน้ำหนักคุณลักษณะจากการจัดกลุ่มและต้นไม้ตัดสินใจคือ จำนวนข้อมูลที่มีขนาดใหญ่ จำนวนคุณลักษณะของข้อมูลที่มีจำนวนมากและ จำนวนคลาสของข้อมูลที่มีจำนวนน้อย สำหรับเวลาที่ใช้ในการสร้าง โมเดลของงานวิจัยนี้มีค่ามากเมื่อเปรียบเทียบกับงานวิจัยของ Hall แต่มีประสิทธิภาพการทำนายที่เพิ่มขึ้น

หทัยชนก กรชี

ลายมือชื่อนิสิต

ลายมือชื่อประธานกรรมการ

26 / 5 / 2551

Hathaichanok Kornchee 2008: Calculation of Attribute Weights for Naive Bayes Classifier by Using Data Clustering and Decision Tree. Master of Science (Computer Science), Major Field: Computer Science, Department of Computer Science. Thesis Advisor: Associate Professor Anongnart Srivihok, Ph.D. 92 pages.

The objective of this study is to improve Hall's algorithm by applying a local weighting scheme for optimizing each test data and solving problems of attribute weights which equal to zero. To improve the prediction performance, we propose a data clustering and decision tree algorithms to calculate weights for each attribute and applied to each node of Naïve Bayes classifier. In this study two clustering methods and decision tree were used to calculate weights of Naïve Bayes. The clustering methods include (1) K-Means algorithm and (2) two step clustering by Cluster Feature Tree and Agglomerative Clustering. Then the performance of the two proposed methods were compared with Hall's method. The measures included percent of correct data prediction and Root Relative Square Error. Four Data sets used in the experiment were obtained from the University of California, Irvin (UCI) machine learning repository. They included (1) Chess End-Game (KR) (2) Balance Scale Weight and Distance Database (Bal_sc) (3) Wave Form and (4) German Credit dataset (Crd).

Result showed that data clustering and decision tree algorithms outperformed Hall's algorithm upon 3 parameters which included number of data (large), number of attribute (large) and number of class (small). However the time complexity of this approach is more than Hall's because it uses more algorithms in calculating attribute weights, clustering algorithm.

Hathaichanok Kornchee

Student's signature

Anongnart Srivihok

Thesis Advisor's signature

26 / 5 / 2008