

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

งานวิจัยฉบับนี้นำเสนอวิธีการคำนวณน้ำหนักของคุณลักษณะข้อมูล คือ วิธีการแบ่งกลุ่มข้อมูล (Clustering) และต้นไม้ตัดสินใจ (Decision Tree) ซึ่งปรับปรุงจากงานวิจัยของ Hall เนื่องจากค่าน้ำหนักที่ใช้ในวิธีของ Hall เป็นค่าน้ำหนักแบบสากล (Global) จึงทำให้บางกรณีมีประสิทธิภาพการทำนายไม่ดีพอ จึงมีแนวคิดในการใช้ค่าน้ำหนักแบบท้องถิ่น (Local) ซึ่งพิจารณาจากค่าน้ำหนักคุณลักษณะที่เหมาะสมกับข้อมูลทดสอบ โดยการแบ่งกลุ่มข้อมูลตามคุณลักษณะข้อมูล สำหรับวิธีการแบ่งกลุ่มที่นำเสนอในการศึกษาคั้งนี้จะใช้ 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

