

เทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative Classification) เป็นเทคนิคหนึ่งในสาขาวิชาทางด้านคานาไมนิ่ง ซึ่งได้รวมเทคนิคการสืบค้นกฎความสัมพันธ์เข้าไว้ด้วยกันกับเทคนิคการจำแนกประเภทข้อมูล โดยที่เทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์สามารถที่จะให้ผลลัพธ์ในการทำนายข้อมูลที่มีประสิทธิภาพและความแม่นยำสูงกว่าเทคนิคการจำแนกประเภทข้อมูลที่ผ่านมาก่อนหน้านี้ ซึ่งไปกว่านั้น การที่นำกฎความสัมพันธ์แบบมีคลาสมาใช้ในการสร้างโมเดลในการทำนายข้อมูล จะให้ความสมบูรณ์ของข้อมูลมากกว่าเทคนิคการจำแนกประเภทข้อมูลที่ใช้หลักทางสถิติหรือความน่าจะเป็น โดยในการเพิ่มประสิทธิภาพและความแม่นยำสำหรับเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ ก็ได้มีผู้นำเสนออัลกอริทึมต่างๆ ไม่ว่าจะเป็น CBA, CMAR และ CPAR ซึ่งอัลกอริทึมเหล่านี้ เน้นปรับปรุงในส่วนของการสร้างโมเดลในการทำนายเป็นหลัก ซึ่งส่งผลให้ในส่วนของการสร้างกฎความสัมพันธ์มีการสร้างกฎที่มีขนาดใหญ่และมีความซับซ้อนออกมาเป็นจำนวนมาก ถึงแม้ว่าจะมีวิธีการในการจัดเรียงกฎ รวมไปถึงการกำจัดกฎที่ไม่มีประโยชน์ออกไป แต่กฎจำนวนมากที่มีความซ้ำซ้อนกันก็ยังคงมีอยู่

ดังนั้นในวิทยานิพนธ์เล่มนี้จึงได้เสนอวิธีการในการกำจัดกฎที่ซ้ำซ้อน เพื่อทำให้จำนวนกฎความสัมพันธ์แบบมีคลาสมีจำนวนลดลง และเสนอวิธีการจัดเรียงกฎเพื่อเพิ่มประสิทธิภาพความแม่นยำให้กับโมเดลที่ใช้ในการทำนาย โดยได้เสนออัลกอริทึมใหม่สำหรับเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ เรียกว่า CBEAR (Classification Based on Essential Class-Association Rules) แทนที่จะใช้กฎความสัมพันธ์แบบมีคลาสทั้งหมด แต่ CBEAR ใช้เฉพาะกฎความสัมพันธ์ที่เรียกว่า ECAR (Essential Class-Association Rules) สำหรับการสร้างโมเดลในการทำนาย และในขั้นตอนของการสร้างโมเดลในการทำนายนั้น CBEAR จะมีวิธีการจัดเรียงกฎ รวมถึงวิธีการเลือกกฎที่จะนำไปใช้ในการทำนาย โดยจะพิจารณาเฉพาะกฎที่ยาวที่สุดก่อน ในการเปรียบเทียบประสิทธิภาพความแม่นยำกับอัลกอริทึม C4.5, CBA และ CMAR นั้น ได้ใช้ฐานข้อมูลมาตรฐานจาก UCI machine learning database repository ซึ่งจากผลการทดลองพบว่า อัลกอริทึม CBEAR ให้ประสิทธิภาพในการจำแนกประเภทข้อมูล ได้แม่นยำมากกว่าอัลกอริทึมตัวอื่นๆ ซึ่งเห็นเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ที่ได้รับความนิยม

Associative classification is a data mining technique that integrates classification with association rule mining. Associative classification can produce more efficient and accurate classifiers than traditional classification techniques. Moreover, generated classifiers in the form of class-association rules (CARs) are more comprehensive than statistical classifiers. A few accurate and effective classifiers based on associative classification have been presented recently, such as CBA, CMAR and CPAR. With the focus on classifier phase, these algorithms generate very large and complex rule sets during the rule generator phase. Despite strategies for sorting and pruning unuseful rules, large number of redundant rules still exists.

The objective of this research is to propose pruning methods for minimizing and reducing number of class-association rules, and sorting strategies for improving accuracy of the output classifier. We propose a new algorithm, CBEAR (Classification Based on Essential Class-Association Rules) for associative classification. Instead of using a complete rule sets, CBEAR uses only ECAR (Essential Class-Association Rules) for building prediction model. At the classifier phase, CBEAR considers only maximal frequent itemsets for sorting and selecting rules. To compare its accuracy with C4.5, CBA, and CMAR, we use standard datasets from UCI machine learning database repository. Experimental results show that CBEAR yields better classification accuracy compared with other associative classification techniques.