

ปรัชญา ผ่องอักษร 2552: การค้นหารูปแบบช่วงของข้อมูลโดยใช้คลาสเพื่อลดปริมาณ
ช่วงของรูปแบบที่ได้ วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์)
สาขาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ อาจารย์ที่ปรึกษา
วิทยานิพนธ์หลัก: รองศาสตราจารย์กฤษณะ ไวยมัย, Ph.D. 57 หน้า

การแบ่งช่วงข้อมูล เป็นเทคนิคหนึ่งในการเตรียมข้อมูล ใช้สำหรับแปลงข้อมูลที่เป็นค่าต่อเนื่อง ให้เป็นค่าไม่ต่อเนื่อง โดยประโยชน์จากการนำข้อมูลที่เป็นค่าไม่ต่อเนื่องไปใช้นั้น นอกจากจะช่วยให้ผู้ใช้เข้าใจข้อมูลได้ง่ายขึ้นแล้ว ยังช่วยลดเวลาในการนำไปใช้งาน อีกทั้งยังช่วยเพิ่มความแม่นยำ และประสิทธิภาพเมื่อนำไปใช้ในการจำแนกประเภทอีกด้วย งานวิจัยนี้ได้นำเสนออัลกอริทึม DCR (Discretization using Class Information to Reduce Number of Intervals) ซึ่งเป็นอัลกอริทึมในการค้นหารูปแบบช่วง ที่ได้นำประโยชน์จากความสัมพันธ์ระหว่างคลาสกับคุณลักษณะต่างๆ ของฐานข้อมูล และการจัดลำดับของการแบ่งช่วงข้อมูลตามค่าเกนสารสนเทศมาใช้ ซึ่งจำนวนช่วงของรูปแบบที่ได้นั้น จะถูกลดจำนวนลงโดยกระบวนการตัดระยะเบี่ยงของฐานข้อมูลออกในแต่ละขั้นตอนการแบ่งช่วง จากผลการทดลองจะพบว่า อัลกอริทึม DCR นี้ สามารถค้นหารูปแบบช่วงของข้อมูลที่มีจำนวนช่วงน้อย แต่ให้ความแม่นยำในการนำไปใช้จำแนกคลาสที่สูงได้ โดยใช้เวลาในการทำงานที่น้อยกว่า เมื่อเทียบกับเทคนิคอื่นๆ ที่ใกล้เคียงกัน

Prachya Pongaksorn 2009: DCR: Discretization Using Class Information to Reduce Number of Intervals. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Kitsana Waiyamai, Ph.D. 57 pages.

Discretization techniques for data set features have received increasing research attention. Results using discretized features are usually more compact, shorter, and accurate than using continuous values. In this paper, an algorithm called Discretization using Class information to Reduce number of intervals (DCR) is proposed. DCR uses both class information and order between attributes to determine the discretization scheme with minimum number of intervals. Attribute discretization order is determined based on information gain of each attribute with respect to the class attribute. The number of intervals is reduced by deleting training data at each step of attribute discretization. Experiments are performed to compare the predictive accuracy and execution time of this algorithm with several well-known algorithms. Results show that discretized features generated by the DCR algorithm contain a smaller number of intervals than other supervised algorithms using less execution time, and the predictive accuracy is as high or higher.