

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.

Student's signature

Thesis Advisor's signature

____ / ____ / ____