

## **Feature transformation: A genetic-based feature construction method for data summarization**

### **Abstract**

The importance of input representation has been recognized already in machine learning. This article discusses the application of genetic-based feature construction methods to generate input data for the data summarization method called Dynamic Aggregation of Relational Attributes (DARA). Here, feature construction methods are applied to improve the descriptive accuracy of the DARA algorithm. The DARA algorithm is designed to summarize data stored in the nontarget tables by clustering them into groups, where multiple records stored in nontarget tables correspond to a single record stored in a target table. This article addresses the question whether or not the descriptive accuracy of the DARA algorithm benefits from the feature construction process. This involves solving the problem of constructing a relevant set of features for the DARA algorithm by using a genetic-based algorithm. This work also evaluates several scoring measures used as fitness functions to find the best set of constructed features. © 2010 Wiley Periodicals, Inc.