DARA: Data Summarisation with Feature Construction

Abstrak

This paper addresses the question whether or not the descriptive accuracy of the DARA (Dynamic Aggregation of Relational Attributes) algorithm benefits from the feature construction process. This involves solving the problem of constructing a set of relevant features used to generate patterns representing records in the TF-IDF weighted frequency matrix in order to cluster these records. In this paper, feature construction will be applied to enhance the results of the data summarisation approach in learning data stored in multiple tables with high cardinality of one-to-many relations. It is expected that the predictive accuracy of a classfication problem can be improved by improving the descriptive accuracy of the data summarisation approach, provided that the summarised data is fed into the target table as one of the features considered in the classification task.