Dynamic Aggregation of Relational Attributes Based on Feature Construction.

Abstrak

The importance of input representation has been recognised already in machine learning. This paper discusses the application of genetic-based feature construction methods to generate input data for the data summarisation method called Dynamic Aggregation of Relational Attributes (DARA). Here, feature construction methods are applied in order to improve the descriptive accuracy of the DARA algorithm. The DARA algorithm is designed to summarise data stored in the non-target tables by clustering them into groups, where multiple records stored in non-target tables correspond to a single record stored in a target table. This paper 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.