Learning Relational Data Based on Multiple Instances of Summarized Data Using DARA

Abstrak

DARA (Dynamic Aggregation of Relational Attributes) algorithm is designed to summarize non-target records stored in a non-target table. These records have many-to-one relationships with records stored in the target table. The records stored in the non-target table are summarized and the summarized data is then appended to the target table. With these summarized data appended into the target table, a classifier will be applied to learn this data in order to perform the classification task. However, the predictive accuracy of the classification task is highly influenced by the representation of the summarized data. In our previous works, several types of feature construction methods have been introduced especially for the DARA algorithm in order to improve the descriptive accuracy of the summarized data and indirectly improve the predictive accuracy of the target data. This paper proposes a method that learns relational data based on multiple instances of summarized data that are obtained using different types of feature construction methods. This involves investigating the effect of selecting several sets of summarized data which have been summarized using the feature construction methods and appending these summarized data into the target table before the classification task can be performed. The predictive accuracy of the classification task is expected to be improved when multiple instances of summarized data appended into the target table. The experiment results show that there are some improvements in the predictive accuracy of the classification by selecting multiple instances of summarized data and appending them into the target table.