

# **Unsupervised Learning of Mutagenesis Molecules Structure Based on an Evolutionary-Based Features Selection in DARA**

## **Abstrak**

The importance of selecting relevant features for data modeling has been recognized already in machine learning. This paper discusses the application of an evolutionary-based feature selection method in order to generate input data for unsupervised learning in DARA (Dynamic Aggregation of Relational Attributes). The feature selection process which is based on the evolutionary algorithm is applied in order to improve the descriptive accuracy of the DARA (Dynamic Aggregation of Relational Attributes) algorithm. The DARA algorithm is designed to summarize 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 issue of optimizing the feature selection process to select relevant set of features for the DARA algorithm by using an evolutionary algorithm, which includes the evaluation of several scoring measures used as fitness functions to find the best set of relevant features. The results show the unsupervised learning in DARA can be improved by selecting a set of relevant features based on the specified fitness function which includes the measures of the dispersion and purity of the clusters produced.