Abstract

The quality of the data being analyzed is a critical factor that affects the accuracy of data mining algorithms. There are two important aspects of the data quality, one is relevance and the other is data redundancy. The inclusion of irrelevant and redundant features in the data mining model results in poor predictions and high computational overhead. Feature extraction aims to reduce the computational cost of feature measurement, increase classifier efficiency, and allow greater classification accuracy based on the process of deriving new features from the original features. This paper represents an approach for classifying students in order to predict their final grades based on features extracted from logged data in an educational web-based system. A combination of multiple classifiers leads to a significant improvement in classification performance. By weighing feature vectors representing feature importance using a Genetic Algorithm (GA), we can optimize the prediction accuracy and obtain a marked improvement
over raw classification. We further show that when the number of features is few, feature weighting and transformation into a new space works efficiently compared to the feature subset selection. This approach is easily adaptable to different types of courses, different population sizes, and allows for different features to be analyzed.

Reference


**Index Terms**

Computer Science  
Data Mining

**Key words**

Feature Subset Selection  
Student repository

Classification

Rule Generations