Abstract

During past few decades, researchers worked on data preprocessing techniques for the datasets. Data preprocessing techniques are needed, where the data are prepared for mining. The performance of data mining algorithms in most cases depends on dataset quality, since low-quality training data may lead to the construction of overfitting or fragile classifiers. Also, scientists worked on data mining areas in both algorithms section and conceptions practice section. But for better results they always used the combined or embedded or hybrid approaches. Scientists used different classifiers in different ways and also got their smoother results by arranging some modification in the algorithms. In this paper we shall describe all possible areas of attribute selection and reduction techniques. Feature selection algorithms broadly fall into three categories: filter models, wrapper models and hybrid models. Practically, scientists do the tasks in two stages for obtaining accuracy and that is, they firstly select the features and then reduce the dimensionality of feature vectors with classifiers through learning. Some promising approaches are indicated here and particular concentration is dedicated to describe different methods from raw level to experts, so that in future one can get significant instruction for further analysis.
References

- Hailiang Chen, Hongyan Liu, Jiawei Han, Xiaojin Yin, Jun He, "Exploring optimization of semantic relationship graph for multi-relational Bayesian classification", Decision Support Systems, Vol. 49, 2009, pp. 112–121
- Brian Quanz, Meenakshi Mishra, "Knowledge Transfer with Low-Quality Data: A

Index Terms

Computer Science

Information Sciences

Keywords

Embedded; hybrid; \textit{filter}; wrapper; classifiers;