Abstract

Medical data mining has enormous potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized by the physicians to improve clinical diagnosis. Feature subset selection is one of data preprocessing step, which is of immense importance in the field of data mining. As a part of feature subset selection step of data preprocessing, a filter approach with genetic algorithm (GA) and Correlation based feature
Feature Subset Selection using Cascaded GA & CFS: A Filter Approach in Supervised Learning

selection has been used in a cascaded fashion. GA rendered global search of attributes with fitness evaluation effected by CFS. Experimental results signify that the feature subset recognized by the proposed filter GA+CFS, when given as input to five classifiers, namely decision tree, Naïve Bayes, Bayesian, Radial basis function and k-nearest neighbor classifiers showed enhanced classification accuracy. Experiments have been carried out on four medical data sets publicly available at UCI.

Reference

- Antonio Arauzo, Jose Manuel Benitez, and Juan Luis Castro, " C-focus: A continuous extension of focus", In proceedings of the 7th online World Conference on Soft Computing in Industrial Applications, 2002.
- M.A. Jayaram, Asha Gowda Karegowda, "Integrating Decision Tree and ANN for

- Asha Gowda Karegowda and M.A. Jayaram,” Cascading GA & CFS for Feature Subset Selection in Medical Data Mining”, International Conference on IEEE International Advance Computing Conference (IACC’09), Thapar University, Patiala, Punjab India, March 6-7, 2009

Index Terms

Computer Science Data Mining

Key words

Feature selection filters Genetic Algorithm
Correlation based

feature selection

Decision tree

Naïve Bayes

Bayesian Classifier

Radial Basis Function

K-Nearest Neighbor