Abstract

A majority of contribution in the domain of rule mining overemphasize on maximizing the predictive accuracy of the discovered patterns. The user-oriented criteria such as comprehensibility and interestingness are have been given secondary importance. Recently, it has been widely acknowledged that even highly accurate discovered knowledge might be
worthless if it scores low on the qualitative parameters of comprehensibility and interestingness. This paper presents a classification algorithm based on evolutionary approach that discovers comprehensible and interesting in CNF form in which along with conjunction in between various attributes there is disjunction among the values of an attribute. A flexible encoding scheme, genetic operators with appropriate syntactic constraints and a suitable fitness function to measure the goodness of rules are proposed for effective evolution of rule sets. The proposed genetic algorithm is validated on several datasets of UCI data set repository and experimental results are presented which clearly indicate lower error rates and more comprehensibility across a range of datasets. Some of the rules show the interesting and valuable nuggets of knowledge discovered from small disjuncts of high accuracy and low support which are very difficult to capture otherwise.

Reference

- Greene, D.P. and Smith, S.F. 1993. Competition-based induction of decision models from

- Fadl, M., Ba-Alwi, and Bharadwaj, K.K. 2005. Automated discovery of hierarchical ripple down rules (HRDRs). In proceedings of the 23rd IASTED International Multi-Conference Artificial Intelligence and Applications, Innsbruck, Austria.

Index Terms

Computer Science

Data Mining

Key terms

Comprehensibility interestingness classification rules genetic algorithm