Abstract

In general rule induction algorithms have a bias that favors the discovery of large disjuncts, rather than small disjuncts. In the context of data mining, small disjuncts are rules covering a small number of examples. Due to their nature, small disjuncts are error prone. It correctly classify individually only few examples but, collectively, cover a significant percentage of the set of examples, so that it is important to develop new approaches to cope with the problem of small disjuncts. This paper presents a classification algorithm based on Evolutionary Algorithm (EA) that discovers interesting small-disjunct rules in the form If P Then D. The proposed system specifically designed for discovering rules covering examples belonging to small disjuncts. The proposed algorithm is validated on several datasets of UCI data set repository and the experimental results are presented to demonstrate the effectiveness of the proposed scheme for automated small-disjunct rules mining.

References

- Weiss, G. M. and Hirsh, H. 2000. A quantitative study of small disjuncts. In...
An Evolutionary Algorithm for Automated Discovery of Small-Disjunct Rules

Another relevant study is the work of Holte, Acker, and Porter (1989) on the problem of small disjuncts, which they present as a challenge in the context of concept learning.

In the realm of data mining, Weiss (1995) discusses the problem of learning with rare cases and small disjuncts, particularly for the case of the 12th International Conference on Machine Learning (ICML-95). The paper by Holte, Acker, and Porter (1989) is also referenced, highlighting the concept learning and the problem of small disjuncts.

Holte, Acker, and Porter (1989) argue that concept learning involving small disjuncts is a significant challenge and propose methods to address this issue. Their work provides a foundation for subsequent research in this area.

Holte, Acker, and Porter (1989) also emphasize the importance of understanding and mitigating the impact of small disjuncts in machine learning models. Their findings underscore the need for robust algorithms capable of handling such disjuncts.

Thornton (1992) further explores techniques in computational learning, providing insights into the development of algorithms that can effectively deal with small disjuncts.

Witten and Frank (2005) discuss the practical aspects of data mining, including techniques for handling small disjuncts, which is a critical aspect of real-world applications.


Quinlan (1991) presents methods for improving the accuracy of small disjuncts, contributing to the ongoing research in this area.

Index Terms

Computer Science

Pattern Recognition

Keywords

Interestingness Small Disjunct Predictive Accuracy Genetic Algorithm