Abstract

Data mining has emerged as most significant and continuously evolving field of research because of its ever growing and far reaching applications into various areas such as medical, military, financial markets, banking etc. One of the most useful applications of data mining is extracting significant and earlier unknown knowledge from real-world databases. This knowledge may be in the form of rules. Rule generation from raw data is a very effective and most widely used tool of data mining. Real life data are frequently imperfect, erroneous, incomplete, uncertain and vague. There are so many approaches for handling missing attribute values. In this paper we use the most common attribute value approach i.e. replacing all the missing attribute values by most frequently occurring attribute value and thereby completing the information table. Subsequently, we find the reduct and core of the complete decision table and verify that the reduct and core find by our method is same as the reduct and core find by ROSE2 software. Thereafter we generate the rules based on reduct. Our results are validated by conducting the same rough set analysis on the incomplete information system using the software ROSE2.

References

- Z. Pawlak Rough Sets and Intelligent Data Analysis[J]. Information Sciences, 2002,
A Rough Set Approach for Generation and Validation of Rules for Missing Attribute Values of a Data Set

- Jiawei Han, and Micheline Kamber, Data Mining: Concepts and Techniques. California??Morgan Kaufmann Publishers, 2000.
- Communications of The ACM, special issue on Data Mining, vol. 39, no. 11.

Index Terms

Computer Science
Artificial Intelligence

Keywords
Data Mining  Knowledge Discovery From Database  Machine Learning  Reduct  Core
Missing Attribute Values
Rule Generation