Abstract

Real world big data are uncertain and imprecise in nature. Receiving higher accuracy in data analysis from those data is an important and challenging task. For the analysis task significant and informative set of data is more required. Feature selection (FS) refers to a problem for selecting maximum relevant and less redundant data, which can provide a superior precision for data analysis purpose. Rough set theory (RST), one of the successful compelling components of soft computing is used for analyzing different data. It has been widely used as a mathematical tool for FS. This paper starts with an outline of the basic concepts of FS and RST. A theoretical comparative review on some existing RST methods are discussed with their pros and cons. Three RST based FS algorithms like Quickreduct (QR), Relative Reduct (RR) and Entropy based Reduct (EBR) are presented briefly. An experimental study on these three algorithms is carried out. Six public domain datasets available in UCI machine learning repository is analyzed for their performance.

References

- J. Han and M. Kanber, "Data Mining concepts and techniques 2nd
Comparative Review on Classical Rough Set Theory based Feature Selection Methods


Index Terms

Computer Science

Information Science
Keywords
Feature selection  Rough set  Soft computing  Supervised  Filter  Dependency