{tag} {/tag} International Journal of Computer Applications

Number 5 - Article 9

Year of Publication: 2010

Authors:

Ayesha butalia

Divya Shah

Dr. R.V Dharaskar

10.5120/123-239

{bibtex}pxc387239.bib{/bibtex}

Abstract

The issues of Real World are Very large data sets, Mixed types of data (continuous valued, symbolic data), Uncertainty (noisy data), Incompleteness (missing, incomplete data), Data change, Use of background knowledge etc. Lot of knowledge related to the application can be generated through these large data sets.

Rough set is the methodology which can be used to deduce rules from these data sets.

The main goal of the rough set analysis is induction of approximations of concepts [4]. Rough sets constitute a sound basis for KDD. It offers mathematical tools to discover patterns hidden in data [4] and hence used in the field of data mining.

Rough Sets does not require any preliminary information as Fuzzy sets require membership values or probability is required in statistics. Hence this is its specialty.

Two novel algorithms to find optimal Reducts of condition attributes based on the relative attribute dependency, out of which the first algorithms gives simple Reduct whereas the second one gives the Reduct with minimum attributes,

This project highlights on the case study of mushroom which consists of twenty two attributes depending on which the decision is taken whether the mushroom plant is edible or poisonous. The technique of Reduct is very useful as when tested, through the algorithms, the twenty one attributes, excluding the decision attribute gets reduced to two to three attributes.

Reference

[1] Z.Pawlak, "Rough Sets", International Journal of Computer and Information Sciences, Vol.11, 341-356(1982).

[2] Z.Pawlak, Rough Sets - Theoretical Aspect of Reasoning about Data, Kluwer Academic Publishers (1991).

[3] A.Skowron, N. Zhong, and N. Cercone Computational Intelligence, An International Journal, Special Issue on Rough Sets, Data Mining, and Granular Computing.

[4] J. Grzymala-Busse, R. Swiniarski, N. Zhong, and Z. Ziarko International Journal of Applied Mathematics and Computer Science, Special Issue on Rough Sets and Its Applications.

[5] H. S Nguyen and S.H Nguyen, "Discretization Methods in Data Mining", Vol.1, 451-482, Physica-Verlag (1998).

[6] T.Y Lin, Journal of Intelligent Automation and Soft Computing, Vol.2, No. 2, Special Issue on Rough Sets (1996).

[7] T.Y Lin, International Journal of Approximate Reasoning, Vol.15, No. 4, Special Issue n Rough Sets (1996).

[8] Z. Ziarko Computational Intelligence, An International Journal, Vol.11, No.2, Special Issue on Rough Sets (1995).

[9] Z. Ziarko Fundamental Informaticae, An International Journal, Vol.27, No. 2-3, Special Issue on Rough Sets (1996).

[10] MushroomUCI Machine Learning Repository Mushroom Data Set.htm

Index Terms

Computer Science

Pattern Recognition

Key words

Reduct Technique

Mixed types of data

Uncertainty

Data change