



Journal Homepage: -www.journalijar.com
**INTERNATIONAL JOURNAL OF
 ADVANCED RESEARCH (IJAR)**

Article DOI:10.21474/IJAR01/6834
 DOI URL: <http://dx.doi.org/10.21474/IJAR01/6834>



RESEARCH ARTICLE

A STUDY OF DM TECHNIQUES IN SOFT COMPUTING FRAMEWORK.

Prof. R. K. Dhuware¹, Dr S.R. Pande² and Dr. S. J. Sharma³.

1. Research Scholar, Department of Electronics and Computer Science, RTM Nagpur University, Nagpur.
2. Asso. Prof. and Head Department of Computer Science, S.S.E.S Science College., Nagpur.
3. Prof. and Head, Department of Electronics and Computer Science, RTM Nagpur University, Nagpur.

Manuscript Info

Manuscript History

Received: 03 February 2018
 Final Accepted: 05 March 2018
 Published: April 2018

Keywords:-

Knowledge discovery in data bases (KDD), Data mining, Soft Computing tools, Fuzzy logic, Neural networks, Genetic Algorithms and Rough sets.

Abstract

Data Mining is one of the fundamental steps in KDD process and is concerned with the algorithmic means by which patterns or structures are enumerated from the data under acceptable computational efficiency. Soft computing tools are individually or in integrated manner, are turning out to be strong candidates for performing Data mining Tasks efficiently.

The main constituents of soft computing indicates Fuzzy logic, Neural networks, Genetic Algorithms and Rough sets. Each of them contributes a distinct methodology for addressing problems in its domain. This is done in a co-operative, rather than a competitive manner. The result is a more intelligent and robust system providing a human – interpretable, low cost, approximate solution, as compared to traditional techniques. The present article provides an overview of the available literature on use of data mining in the soft computing framework.

Copy Right, IJAR, 2018., All rights reserved.

Introduction:-

With the development of computer hardware and software and rapid computerization of business, huge amount of data have been collected and stored in database. As a result, traditional statistical techniques and data management tools are no longer adequate for analyzing this vast collection of data. With the explosive growth in stored and transient data, hence to transfer the vast amount of data in to useful information and knowledge remain the challenges. Data mining refers to the extraction of useful information from large set of data. It is a techniques for the discovery of patterns hidden in large data sets, focusing on issue relating to their feasibility, usefulness, effectiveness and scalability. On the other hand, soft computing deals with information processing. If these two key properties can be combined in a constructive way, then this formation can effectively be used for knowledge discovery in large databases.

Referring to this synergistic combination, the basic merits of DM and soft computing Paradigms are pointed out and novel data mining implementation coupled to a soft computing approach for knowledge discovery is presented.

Overview of Knowledge Discovery and Data Mining:-

KDD focuses on the overall process of knowledge discovery from large volumes of data, including the storage and accessing of such data, scaling of algorithms to massive data sets, interpretation and visualization of results, and the modeling and support of the overall human machine interaction.

Corresponding Author:-R. K. Dhuware.

Address:-Research Scholar, Department of Electronics and Computer Science, RTM Nagpur University, Nagpur.

An important step in the KDD process is data mining. Data mining involves fitting models to or determining patterns from observed data. The fitted models play the role of inferred knowledge. Deciding whether the model reflects useful knowledge or not is basically a part of the overall KDD process for which subjective human judgment is usually required.

Development of new generation algorithms is expected to encompass more diverse sources and types of data that will support mixed-initiative data mining, where human experts collaborate with the computer to form hypotheses and test them. The main challenges to the data mining procedure involve the following : Massive data sets and high dimensionality, User interaction and prior knowledge, Overfitting and assessing the statistical significance, Understandability of patterns, Nonstandard and incomplete data, Mixed media data, Management of changing data and knowledge and Integration

Role and Significance of Soft computing Methods in Data Mining:-

Soft computing is a consortium of methodologies (like Fuzzy Logic, Neural Networks, Genetic Algorithms, Rough Sets), that works Synergistically and provides, in one form or another, flexible information processing capabilities for handling real life problems.

Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness and low cost solution.

Recently various soft computing methodologies have been applied to handle the different challenges posed by data mining. The main constituents of soft computing, at this juncture, include fuzzy logic, neural networks, genetic algorithms, and rough sets. Each of them contributes a distinct methodology for addressing problems in its domain. This is done in a cooperative, rather than a competitive, manner. The result is a more intelligent and robust system providing a human-interpretable, low cost, approximate solution, as compared to traditional techniques. It may be mentioned that there is no universally best data mining method; choosing particular soft computing tool(s) or some combination with traditional methods is entirely dependent on the particular application and requires human interaction to decide on the suitability of an approach.

Each of the soft computing methods have their own characteristic ,based upon which they can be suitably used in data mining process .Encapsulation of each of these methods in the data mining process has brought about a significant difference in the approach of information extraction and processing.

Fuzzy Logic in Data Mining:-

Since fuzzy sets allow partial membership, Fuzzy logic is basically multi valued logic that allows intermediate values to be defined between conventional evaluations such as yes/no, true/false etc. This represents a more human like thinking approach in the programming of computer. Since in data mining with large data bases the most common challenges are the noisy, imprecise, vague data, therefore by suitable extracting the relevant characteristics of fuzzy sets the data mining techniques can be made more efficient. The use of fuzzy techniques has been considered to be one of the key components of data mining systems because of the affinity with human knowledge representation. Wei and Chen have mined generalized association rules with fuzzy taxonomic structures. Fuzzy logic is useful for data mining systems performing classification such as CRM, Health care and finance. At this junction fuzzy data mining comes as great help to data miners.

Neural Networks in Data Mining:-

Neural Network based data mining approach consists of three major phases:

Network Construction and Training:-

This phase constructs and trains a three layer neural network based on the number of attributes and number of classes and chosen input coding method.

Network Pruning:-

The pruning phase aims at removing redundant links and units without increasing the classification error rate of the network. A small number of units and links left in the network after pruning enable us to extract concise and comprehensible rules.

Rule Extraction:-

This phase extracts the classification rules from the pruned network. The rules generated are in the form of "if (a, Bv,) and (x, Bv,) and ... and (x, Bv,) then Cy where a.s., are the attributes of an input tuple, v ,are constants,& are relational operators (=, <, >, <=, <>), and Ci is one of the class labels.

The nns exhibit mapping capabilities that is they can map input patterns to their associated output patterns. The nns learn by examples, thus nn architecture can be trained with known examples of a problems before they are tested for their inference capabilities on unknown instances of the problem, they can therefore identify new objects previously untrained. The nns possess the capability to generalize, thus they can predict new outcomes from past trends. The nns are robust systems and are fault tolerant. They can therefore recall full patterns from incomplete partial or noisy patterns. Based upon the above characteristics of nn which are very closely associated with the functionality what is required by the data mining application. They can be efficiently embodied with data mining methods for increasing efficiency of the outcome of different data mining techniques. Data mining, cleaning and validation could be achieved by determining which records suspiciously diverge from the patterns of their peers. Hence for this proper approaches for combining the ANN and datamining technologies should be found to improve and optimize data mining technology. Fuzzy neural networks and self organizing neural networks are gaining fast importance in the field of data mining. Neural networks have found wide application in areas such as pattern recognition, image processing, optimization, fore casting.

Genetic Algorithm in Data Mining:-

Genetic algorithm plays an important role in data mining technology, which is decided by its own characteristics and advantages. To sum up, mainly in the following aspects:

1. Genetic algorithm processing object not parameters itself, but the encoded individuals of parameters set, which directly operate to set, queue, matrices, charts, and other structure.
2. Possess better global overall search performance; reduce the risk of partial optimal solution. At the same time, genetic algorithm itself is also very easy to parallel.
3. In standard genetic algorithm, basically not use the knowledge of search space or other supporting information, but use fitness function to evaluate individuals, and do genetic Operation on the following basis.
4. Genetic algorithm doesn't adopt deterministic rules, but adopts the rules of probability changing to guide search direction. Genetic algorithm has been efficiently used in multimedia databases.
5. Knowledge discovery systems have been developed using genetic programming concepts.

Rough Sets in Data Mining:-

The main goal of rough sets is induction of approximation of concepts rough sets constitutes a sound basis for KDD. It offers mathematical tools to discover patterns hidden in data and hence used in the field of mining. Rough sets does not require any preliminary information as fuzzy sets require membership values or probability is required in statistics hence this is special. Hence rough sets can be used as a framework for data mining especially in the areas of soft computing where exact data is not required and in some areas where approximation data can be of great help. Rough set theory can be used in different steps in data processing such as computing lower and upper approximation.

Neuro-Fuzzy Computing:-

Neuro-fuzzy computation is one of the most popular hybridizations widely reported in literature. It comprises a judicious integration of the merits of neural and fuzzy approaches, enabling one to build more intelligent decision-making systems. This incorporates the generic advantages of artificial neural networks like massive parallelism, robustness, and learning in *data-rich* environments into the system. The modelings of imprecise and qualitative knowledge in natural/linguistic terms as well as the transmission of uncertainty are possible through the use of fuzzy logic. Besides these generic advantages, the neuro-fuzzy approach also provides the corresponding application specific merits as highlighted earlier.

The rule generation aspect of neural networks is utilized to extract more *natural* rules from fuzzy neural networks. The fuzzy MLP and fuzzy Kohonen network have been used for linguistic rule generation and inferencing. Here the input, besides being in quantitative, linguistic, or set forms, or a combination of these, can also be incomplete. The components of the input vector consist of membership values to the overlapping partitions of linguistic properties *low*, *medium*, and *high* corresponding to each input feature. Output decision is provided in terms of class membership values.

The models are capable of

1. Inference based on complete and/or partial information;
2. Querying the user for unknown input variables that are key to reaching a decision;
3. Producing justification for inferences in the form of IF THEN rules.

Conclusion:-

The Synergistic combination of data mining methods and soft computing tools like Fuzzy logic, Genetic Algorithms, Neural Networks, Rough sets and their hybridizations can greatly improve the efficiency of data mining methods. The soft computing tools are suitable for solving the problem of data mining because of its characteristics of good robustness, self-organizing, adaptive, parallel processing, distributive storage and high degree of fault tolerance. Fuzzy sets provide a natural framework for the process in dealing with uncertainty. Neural networks and Rough sets are widely used for classification and rule generation. Genetic Algorithms are involved in various optimization and search processes, like query optimization and template selections. Other approaches like case-based reasoning and decision trees are also widely used to solve data mining problems.

Hence it may be concluded that both paradigms have their own merits and by observing these merits synergistically, these paradigms can be used in a complimentary way for knowledge discovery in databases.

References:-

1. A. B. Tickle, R. Andrews, M. Golea, and J. Diederich, "The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks," *IEEE Trans. Neural Networks*, vol. 9, pp. 1057–1068, 1998.
2. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds., *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA:AAAI/MIT Press, 1996.
3. "Special issue on knowledge discovery in data- and knowledge bases," *Int. J. Intell. Syst.*, vol. 7, 1992.
4. L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Commun. ACM*, vol. 37, pp. 77–84, 1994.
5. S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*. New York: Wiley, 1999.
6. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
7. *Data Mining Concepts and Techniques*- Jiawei Han, Micheline Kamber.
8. *Effective Data Mining Using Neural Networks* Hongjun Lu, Rudy Setiono, and Huan Liu, *IEEE Transactions on knowledge and data engineering*, vol. 8, No. 6 December 1996.
9. S. Mitra and S. K. Pal, "Fuzzy multi-layer perceptron, inferencing and rule generation," *IEEE Trans. Neural Networks*, vol. 6, pp. 51–63, 1995.
10. J. F. Baldwin, "Knowledge from data using fuzzy methods," *Pattern Recognition Lett.*, vol. 17, pp. 593–600, 1996.
11. *Principles of Soft Computing*-S. N. Sivanandam and S. N. Deepa.
12. J. Grzymala-Busse, R. Swiniarski, N. Zhong and Z. Zizko *International Journal of Applied mathematics and Computer science*, Special Issue on Rough Sets, and Its Applications.
13. Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning about Data*. Dordrecht: Kluwer, 1991. A. Skowron and C. Rauszer, "The Discernibility Matrices and functions in Information Systems," *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, R. Slowinski, ed. pp. 331- 362, Dordrecht: Kluwer Academic, 1992.
14. Q. Wei and G. Chen, "Mining generalized association rules with fuzzy taxonomic structures," in *Proc. NAFIPS 99*, New York, June 1999, pp. 477–481.
15. I. W. Flockhart and N. J. Radcliffe, "A genetic algorithm-based approach to data mining," in *Proc. 2nd Int. Conf. Knowledge Discovery Data Mining (KDD-96)*. Portland, OR, Aug. 2–4, 1996, p. 299.
17. M. L. Raymer, W. F. Punch, E. D. Goodman, and L. A. Kuhn, "Genetic programming for improved data mining: An application to the biochemistry of protein interactions," in *Proc. 1st Annu. Conf. Genetic Programming 1996*, Stanford Univ., CA, July 28–31, 1996, pp. 375–380.
18. S. Mitra and Y. Hayashi, "Neuro-fuzzy rule generation: Survey in soft computing framework," *IEEE Trans. Neural Networks*, vol. 11, pp. 748–768, 2000.