ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

A Survey: Rough Set Theory in Incomplete Information Systems

M. Pushpalatha[1], Dr. V. Anuratha[2]

M.Phil Scholar, PG, Department of Computer Applications, Sree Saraswathi Thyagaraja College, Pollachi, India[1]

Assistant Professor of PG, Department of Computer Applications, Sree Saraswathi Thyagaraja College, Pollachi India[2]

ABSTRACT: Rough Set theory has been conceived as a tool to conceptualize, organize and analyze various types of data, in particular, to deal with inexact, uncertain or vague knowledge in applications related to Artificial Intelligence. In the case of classification, this theory implicitly calculates reducts of the full set of attributes, eliminating those that are redundant or meaningless. Such reducts may even serve as input to other classifiers other than Rough Sets. The typical high dimensionality of current databases precludes the use of greedy methods to find optimal or suboptimal reducts in the search space and requires the use of stochastic methods. Rough set theory, which has been used successfully in solving problems in pattern recognition, machine learning, and data mining, centers around the idea that a set of distinct objects may be approximated via a lower and upper bound. In order to obtain the benefits that rough sets can provide for data mining and related tasks, efficient computation of these approximations is vital.

KEYWORDS: Rough set, Artificial Intelligence, greedy method, data mining, and approximations.

I. INTRODUCTION

Large amounts of data are generated every day and the ability to analyse them is normally a challenge. Experts need efficient data mining methods to extract useful information and to perform the analysis of the data. This is the case of the Rough Sets Theory (RST);Pawlak [1] introduced mathematical rough set theory in the early 1980's. The theory was based on the discernibility of objects. Rough set theory provides systems designers with the ability to handle uncertainty. If a concept is 'not definable' in a given knowledge base, rough sets can 'approximate' with respect to that knowledge. From a medical point of view, the attribute-value boundaries are usually vague.

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). For example, if objects are patients suffering from a certain disease, symptoms of the disease form information about patients. Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. This understanding of indiscernibility is related to the idea of Gottfried Wilhelm Leibniz that objects are indiscernible if and only if all available functional take on them identical values (Leibniz's Law of Indiscernibility: The Identity of Indiscernible). However, in the rough set approach indiscernibility is defined relative to a given set of functional (attributes).

A weak aspect of RST is the non-availability of free RST software, except for limited implementations. On the other hand, there is RST proprietary software. RST is an extension of the set theory and has the implicit feature of compressing the dataset. Such compression is due to the definition of equivalence classes based on indiscernibility relations and to the elimination of redundant or meaningless attributes. A central concept in RST is attribute reduction, which generates reducts.

II. RELATED WORK

Marzena Kryszkiewicz[1] presented Rough Set approach to incomplete information systems, i.e. to systems in which attribute values for objects may be unknown (missing, null). His main concern was devoted to find rules from such

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

systems. He proposed reduction of knowledge that eliminates only that information, which is not essential from the point of view of classification or decision making.

PuntipPattaraintakorna, Nick Cerconeb [2], described Medical science is not an exact science in which processes can be easily analyzed and modeled. Rough set theory has proven well suited for accommodating such inexactness of the medical profession. As rough set theory matures and its theoretical perspective is extended, the theory has been also followed by development of innovative rough sets systems as a result of this maturation. Unique concerns in medical sciences as well as the need of integrated rough sets systems are discussed. We present a short survey of ongoing research and a case study on integrating rough set theory and medical application. Issues in the current state of rough sets in advancing medical technology and some of its challenges are also highlighted.

Alex SandroAguiar Pessoa, Stephan Stephany[3], proposed the innovative use of two known met heuristics for the calculation, the Variable Neighborhood Search, the Variable Neighborhood Descent, besides a third heuristic called Decrescent Cardinality Search. The last one is a new heuristic specifically proposed for reduct calculation. Considering some database commonly found in the literature of the area, the reducts that have been obtained present lower cardinality, i.e., a lower number of attributes.

Y.H.Qian, J.Y.Liang, W.Pedrycz, and C.Y.Dang[4], Feature selection (attribute reduction) from large-scale incomplete data is a challenging problem in areas such as pattern recognition, machine learning and data mining. In rough set theory, feature selection from incomplete data aims to retain the discriminatory power of original features. To address this issue, many feature selection algorithms have been proposed, however, these algorithms are often computationally time-consuming. To overcome this shortcoming, they introduced a theoretic framework based on rough set theory, which is called positive approximation and can be used to accelerate a heuristic process for feature selection from incomplete data. As an application of the proposed accelerator, a general feature selection algorithm was designed. By integrating the accelerator into a heuristic algorithm, they obtain several modified representative heuristic feature selection algorithms in rough set theory.

J.B.Zhang,T.R.Li,D.Ruan,Z.Z.GaoandC.B.Zhao [5], The effective computation of approximations is vital for improving the performance of data mining or other related tasks. The recently introduced MapReduce technique has gained a lot of attention from the scientific community for its applicability in massive data analysis. The authors proposed a parallel method for computing rough set approximations. Consequently, algorithms corresponding to the parallel method based on the MapReduce technique are put forward to deal with the massive data. An extensive experimental evaluation on different large data sets shows that the proposed parallel method is effective for data mining.

G.L.Liu[6] proposed a new matrix view of the theory of rough sets, they started with a binary relation and redefine a pair of lower and upper approximation operators using the matrix representation. Different classes of rough set algebras are obtained from different types of binary relations. Various classes of rough set algebras are characterized by different sets of axioms. Axioms of upper approximation operations guarantee the existence of certain types of binary relations (or matrices) producing the same operators. The upper approximation of the Pawlak rough sets, rough fuzzy sets and rough sets of vectors over an arbitrary fuzzy lattice are characterized by the same independent axiomatic system.

G.L.Liu[7] presented outer product method which was used in rough set study for the first time. By that approach, they propose a unified lower approximation axiomatic system for Pawlak's rough sets and fuzzy rough sets. As the dual of axiomatic systems for lower approximation, a unified upper approximation axiomatic characterization of rough sets and fuzzy rough sets without any restriction on the cardinality of universe was also given. These rough set axiomatic systems will help to understand the structural feature of various approximate operators.

J.W.Grzymala-Busse,andW.Ziarko,[8] gave basic ideas of rough set theory - a new approach to vague data analysis. The lower and the upper approximation of a set the basic operations of the theory, are intuitively explained and formally defined. Some applications of rough set theory are briefly outline and some future problems pointed out.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Y.H.Qian,C.Y.Dang,J.Y.Liang, and D.Tang [9] introduced the conjunctive/disjunctive set-valued ordered information systems, and developed an approach to queuing problems for objects in presence of multiple attributes and criteria. Then, they presented a dominance-based rough set approach for these two types of set-valued ordered information systems, which is mainly based on substitution of the indiscernibility relation by a dominance relation. Through the lower/ upper approximation of a decision, some certain/possible decision rules from a so-called set-valued ordered decision table can be extracted. Finally, they presented attribute reduction (also called criteria reduction in ordered information systems) approaches to these two types of ordered information systems and ordered decision tables, which can be used to simplify a set-valued ordered information system and find decision rules directly from a set-valued ordered decision table.

T.R.Li,D.Ruan,W.Geert,J.Song,and Y.Xu, [10]proposed a new approach for incrementally updating approximations of a concept was presented under the characteristic relation-based rough sets. Finally, the approach of direct computation of rough set approximations and the proposed approach of dynamic maintenance of rough set approximations are employed for performance comparison. An extensive experimental evaluation on a large soybean database from MLC showed that the proposed approach effectively handles a dynamic attribute generalization in data mining.

III. PROPOSED SYSTEM

The Proposed system presents an application of rough sets and statistical methods to feature reduction and pattern recognition.

- The presented description of rough sets theory emphasizes the role of rough sets reducts in feature selection and data reduction in pattern recognition.
- Applying kernel induced fuzz logic-rough concept to overcome the keys and values Rough set limitations. Therefore, to build an efficient classifier Genetic Algorithm (GA) has been applied to obtain optimal subset of attributes, sufficient to classify the objects.
- The proposed algorithm reduces dimensionality to a great extent without degrading the accuracy of classification and avoid of being trapped at local minima.
- Results are compared with the existing algorithms, demonstrate compatible outcome.

Genetic algorithm (GA) is a search heuristic, used to generate useful solutions to optimization and search problems. In the proposed optimal rough set based genetic algorithm, the dependency of the decision feature attribute on different set of conditional variables is calculated and attributes with highest dependency value is selected as optimum reduct by applying genetic algorithm.

Step 1:

Take the roughest feature input into the information system along with feature values of each data objects in every class.

Step 2:

Initialize $\gamma_{prev} = 0.0$, $\gamma_{best} = 0.0$, flag = 0, count-of-generation = 0.

Step 3:

do repeat until (flag == 1) :

(a) Count-of-generation ++1.

(b) Select randomly number of features to be taken.

number-features ←− rand()

number-features = number-features +1.

(c) Generate the combination set containing all combinations of number − features number of attributes.

(d) Select a combination from the combination set.

comb-num ←− rand() % total-number-of-elements in combination set.

(e) Take the reduced information system

number−features number of attributes for *comb−num*th combination of the combination set.

(f) Find out the crossing over probability.

(g) Modify the roughest information system (x) as required after mutation.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

(h) Call function variation (*x*, no-of-objects, no-of attributes) for generating different combination of attributes.

IV. CONCLUSION

In this survey paper, the several real word datasets, the information collected to represent various decisions along with variables contains ambiguity. For few identical attributes, decisions made or the classes labeled are different for different patterns. The improved Rough set theory has emerged as an essential tool to handle such ambiguity. This survey paper presents an overview of the rough set theory, terms used in the rough sets. Rough sets can be applied to the important process of feature selection and learning. The investigations and developments made in these areas are tabled and discussed in the paper. Rough sets can be combined with other techniques when they alone are not able to produce better results. Further, applications of rough sets are numerous; some of the applications is summarized in the paper with references. The available literature in rough sets opens a promising gate towards future research directions in many other complex areas including big data, communications, computational intelligence etc.

REFERENCES

1. M. Kryszkiewicz, "Rough set approach to incomplete information systems," Inf. Sci., vol. 112, no. 1–4, pp. 39–49, Dec. 1998.
2. PuntipPattaraintakorna, Nick Cerconeb "Integrating rough set theory and medical applications", Volume 21, Issue 4, Pages 400–403, April 2008
3. Pessoa, A.S.A. and Stephany, S. "An Innovative Approach for Attribute Reduction in Rough Set Theory". Intelligent Information Management, 6, 223-239. (2014)
4. Y. H. Qian, J. Y. Liang, W. Pedrycz, and C. Y. Dang, "An efficient accelerator for attribute reduction from incomplete data in rough set framework," Pattern Recognit., vol. 44, no. 8, pp. 1658–1670, Aug. 2011.
5. J. B. Zhang, T. R. Li, D. Ruan, Z. Z. Gao, and C. B. Zhao, "A parallel method for computing rough set approximations," Inf. Sci., vol. 194, pp. 209–223, Jul. 2012.
6. G. L. Liu, "The axiomatization of the rough set upper approximation operations," FundamentaInformat., vol. 69, no. 3, pp. 331–342, 2006.
7. G. L. Liu, "Axiomatic systems for rough sets and fuzzy rough sets," Int. J. Approximate Reasoning, vol. 48, no. 3, pp. 857–867, Aug. 2008.
8. J. W. Grzymala-Busse, and W. Ziarko, "Data mining and rough set theory," Commun. ACM, vol. 43, no. 4, pp. 108–109, Apr. 2000.
9. Y. H. Qian, C. Y. Dang, J. Y. Liang, and D. Tang, "Set-valued ordered information systems," Inf. Sci., vol. 179, no. 16, pp. 2809– 2832, Jul. 2009.
10. T. R. Li, D. Ruan, W. Geert, J. Song, and Y. Xu, "A rough sets based characteristic relation approach for dynamic attribute generalization in data mining," Knowl.-Based Syst., vol. 20, no. 5, pp. 485–494, Jun. 2007.