

REVIEW: FEATURE SELECTION USING ROUGH SET**Pramod Mehta^{1*}**^{1*} Department of Mathematics, Mewar University, India**Pragati Jain ^{2*}**^{2*}Department of Science, St. Paul Institute of Professional Studies, India

Abstract : The aim of this paper is to give a review of literature on rough set approach for feature selection and also on adjoining key topics written from 2009 to 2014 in the research area of feature selection using rough set. The advantage of this review is to find research papers according to given key topics.

Key words : Rough Set, Feature Selection, Information System.

Introduction

To identify and remove unnecessary, irrelevant and redundant attributes from data that do not contribute to the accuracy of a predictive model or may in fact decrease that accuracy of the model, Feature selection method can be used. It is also called variable selection. It is also possible to automatic selection of those features in our data that are most useful or most relevant for the problem on which we are working.

Feature selection is the property in which focus of interest is for quite some time and more work has been done. According to M. Dash and H. Liu in 1997 feature selection can be viewed as one of the most fundamental problems in the field of machine learning [4]. Objective of feature selection is to find out a minimal feature subset from a problem domain with high accuracy.

In the year 1982 mathematician Z. Pawlak proposed Rough Set Theory which is used as a mathematical tool to treat the vague and the imprecise data. We assume in rough set theory with every object of the universe of discourse; we associate some information, data or knowledge. It is concerned with the classification and analysis of imprecise,

uncertain or incomplete information and knowledge and is also considered as one of the first non statistical approaches in data analysis. Now a day, Rough Set Theory is used in the areas of medicine, analysis of database attributes, process control, pattern recognition, security analysis and Stock market etc [18].

The fundamental concept of RST is based on approximation of lower and upper space of a set. Given two finite, non – empty sets U and A , where U is the universe of objects, and A is attributes, the pair (U, A) is called information table. Any subset B of A determines a binary relation on U , called an in discernibility relation (R) .

If $(x, y) \in B$ we denote U/B equivalence class of B . To this end we require the basic concept of RST. The lower approximation of a set B with respect to in discernibility relation (R) is the set of all objects which can be for certain classified as B with respect to R . The upper approximation of a set B with respect to R is the set of all objects which can be possibly classified as B with respect to R [13].

The boundary region of a set B with respect to R is the set of all objects, which can be classified neither as B nor as NOT- B with respect to R . If

the boundary region is empty the n set B is called crisp otherwise it is called rough. This article provides a very useful insight for rough set application in feature selection. In this article we give a review on the feature selection using rough set theory between years 2009 to 2014.

Analysis

Rough Set approach using ACO: Ant colony optimization has been successfully applied to many difficult combinatorial problems like travelling salesman, scheduling etc.

In 2009, Debahuti Mishra, Amiya Kumar Rath, Milu Acharya and Tanushree Jena proposed a novel method for dimensionality reduction of a feature set by selecting a subset of the features that contains most of the essential information [8].

In 2010, Yumin Chen, Duoqian Miao and Ruizhi Wang proposed a new rough set approach to feature selection using ACO, adopting mutual information based on feature significance as heuristic information. They also gave a novel feature selection algorithm. Their experiments are carried out on standard UCI datasets conclusion shows that the algorithm can provide noteworthy solution to find a minimal subset of the features [3].

In 2012, Tamara Qablan, Qasem A. Al-Radaideh and Sawsan Abu Sheqeir proposed an approach for reduct computation based on ACO. This approach has three features (i) The updated pheromone trails are directed to the nodes by the ants, rather than the visited edges connecting these nodes; (ii) the pheromone trails values are limited between max and min trails limits (iii) the heuristics value is evaluated dynamically during the ant search. Several experiments have been carried out to verify this approach and result have showed that this approach can produce a reduct with less number of iteration in comparison to other ACO based feature selection [14].

In 2013, Majdi Mafarja and Derar Eleyan proposed an attribute reduction method based on ACO and RST. It was tested on standard benchmark datasets. This result was that this algorithm performs well and competes with other attributes reduction approaches in terms of the number of the selected features and the running time [7].

In 2013, Hesham Arafat, Rasheed M. Elawady Sherif Barkat and Nora E. Elrashidy proposed a investigating strategy based on Rough Set Theory using ACO. Result of their experiment was that the hybrid approach can help in improving classification accuracy and also in finding more robust features to improve classifier performance [1].

Rough Set approach for feature selection without ACO:

In 2009, Yong Yang, Guoyin Wang and Hao Kong proposed a feature selection for facial emotion recognition using Rough Set Theory. Author proposed a novel attribute selection algorithm SARA. The experimental result of this approach shows a high recognition rate [21].

In 2011, Norhalina Senan, Rosziati and Nazir Mohd Nawi proposed a alternative feature selection technique using rough set theory based on Maximum degree of dependency of Attributes for Traditional Malay musicals instruments. Findings of the best 17 selected features are increased up to 99.82% and 98.03% with 1-NN and PART classifiers respectively [17].

In 2012, Hema Banati and Monika Bajaj proposed a feature selection approach which combines RST and 'firefly' algorithm. In this algorithm the author simulates the attraction system of real fireflies that guides the feature selection procedure. After the experiment the result shows that this approach is better in terms of time and optimality [2].

In 2012, Walid Moudani, Ahmad Shahin, Fadi Chakik and Felix Mora – Camino proposed

Dynamic Rough Sets Feature Reduction, they proposed a new solution approach in order to reduce the complexity of the classification problem faced by retailer business. They proposed DRSAR; their approach produces an exact solution in mathematical terms and appears to be quite adapted [9].

In 2012, You – Shyang Chen proposed approach for credit ratings for banks across Asia using integrating feature selection and the CPDA – based rough sets approach. They proposed an integrated feature selection approach for selecting 16 condition attributes of financial ratios. Their experimental focus was the Asian banking industry and experimental result demonstrates that the procedure is an effective method removing irrelevant attributes and achieving increased accuracy [22].

In 2012, sombut Foithong, Ouen Pinngern and Boonwat Attachoo proposed a novel feature selection method based on the hybrid model. The authors develop a feature selection method using the information criterion without requiring a user – defined parameter for the selection of the candidate feature set. The authors proposed the superreduct-wrapper approach for finding a proper reduct. The efficiency and effectiveness of their technique is demonstrated through extensive comparison with other methods and the result shows an excellent performance [6].

In 2012, Usman Qamar and John A. Keane proposed clustering using Rough – Set feature selection. The authors analyze the effects of rough sets on clustering using 10 datasets. Each dataset include a decision attribute. This generates a framework for applying rough sets for clustering. This paper shows the effectiveness of the rough set feature selection for clustering data and it provides a framework for applying the rough set for classification [15].

In 2012, Jue Wang, Abdel- Rahman Hedar, Shouyang Wang and Jian Ma Proposed RSFS approach using rough set and scatter search. The experimental result of this approach shows that RSFS has a superior performance in saving the computational costs and improve classification accuracy [20].

In 2013, Jianhua Dia and Qing Xu proposed attribute selection approach based on information gain ratio with application to tumor classification. They used an attribute selection method based on fuzzy gain ratio under the framework of fuzzy rough set theory. This approach is compared to several other approaches on three real world tumor data sets in gene expression and the result shows that this method is more effective [5].

In 2014, Sabu M. K. proposed a rough Set based feature Selection approach for the predication of the learning disabilities (LD). His approach mainly focuses on the effect of feature selection in predication of LD in school aged children. The experiments conducted on the LD datasets demonstrate the effectiveness of RST in selecting significant symptoms of LD [16].

Conculsion

Rough Set Theory is used for imprecise and incomplete information. This paper presents a review of different techniques for feature selection using Rough Set Theory. Also paper consider various techniques using ACO and without ACO for solving many problems by various researchers. A number of simplification and evaluation technique of feature selection using rough set theory have been proposed by various researcher in many aspect of real life. The paper will contribute in the field of feature selection in rough set theory.

References

1. Arafat H., Elawady R., Barakat S., Elrashidy N., "Using Rough Set and Ant Colony optimization in feature Selection", Vol. - 2, 148-155, 2013.
2. Banati H., Bajaj M., "Fire Fly Based Feature Selection Approach", Vol.-8, 473-480, 2011.
3. Chen Y., Miao D., Wang R., "A rough set approach to feature selection based on ant colony optimization", Vol. - 31, 226-233, 2010.
4. Dash M., Liu H., "Feature selection for classification. *Intell. Data Analysis*", Vol.-1, 131 -156. 1997.
5. Dia J., Xu Q., "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification", Vol.-13, 211- 221, 2013.
6. Foithong S., Pinnern Q., Attachoo B., "Feature subset selection wrapper based on mutual information and rough sets", Vol.- 39, 574 – 584 , 2012.
7. Mafarja M., Eleyan D., "Ant Colony Optimization based Feature Selection in Rough Set Theory", Vol.-1, 244 -247, 2013.
8. Mishra D., Rath A., Acharya M., and Jena T., "Rough ACO: A Hybridized Model for Feature Selection in Gene Expression Data", Vol. -1, 85 -98, 2009.
9. Moudani W., Shahin A., Chakik F., Mora- Camino F., "Dynamic Rough Sets Features Reduction" , Vol.-9, 1 -10, 2011.
10. Muthurajkumar S., Kulothungan K., Vijayalakshmi M., Jaisankar N., Kannan A. , "A Rough Set based feature Selection Algorithm for Effective Intrusion Detection in Cloud Mode", 2013.
11. Pawlak Z., Marek W., "Information storage and retrieval: Mathematical foundation", *Theoretical Computer Science*, 331-354, 1976.
12. Pawlak Z., "Rough sets: "Theoretical aspects of reasoning about data", Kluwer, Netherland, 1991.
13. Pawlak Z., "Rough Set, Institute of Theoretical and Applied Information, Polish academy of Sciences".
14. Qablan T., Qasem A., Radaideh A., Shugeir S., "A Reduct Computation Based on Ant Colony Optimization" ,Vol. -21, 29-40, 2012.
15. Qamar U., Keane J., "Clustering Using Rough – Set Feature Selection", Vol.-21, 5915-5920, 2012.
16. Sabu M.K, "A Rough Set Based Feature Selection Approach for the Predication of Learning Disabilities", Vol.-2, issue-12, 43-48, 2014.
17. Senan N., Ibrahim R., Mohd Nawi N., "Rough Set Approach for Attributes Selection of Traditional Malay Musical Instruments Sounds Classification", Vol.-4, issue-3, 59-78, 2011.
18. Silvia R. and Germano L.-T., "Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications".
19. Walczak B., Massart D. L., "Rough Set Theory", Vol.-47, 1 -16, 1999.
20. Wang J., Hedar A., Wang S., Ma J., "Rough Set and scatter search met heuristic based feature selection for credit scoring", Vol.- 39, 6123-6128, 2012.
21. Yang Y., Wang G., Kong H., "Self-Learning Facial Emotional Feature Selection Based on Rough Set Theory", Vol.-2009, Article ID-802932, 2009.
22. You – Chen S. , "Classifying credit ratings for Asian banks using integrating feature selection and the CPDA – based rough sets approach" , Vol-26, 259 -270, 2014.
