

Quality of classification with LERS system in the data size context

Quality of
classification
with LERS
system

29

M. Sudha
Department of Mathematics, AMET University, Chennai, India, and
A. Kumaravel
School of Soft Computing, Bharath University, Chennai, India

Received 26 November 2017
Revised 31 December 2017
Accepted 1 February 2018

Abstract

Rough set theory is a simple and potential methodology in extracting and minimizing rules from decision tables. Its concepts are core, reduct and discovering knowledge in the form of rules. The decision rules explain the decision state to predict and support the new situation. Initially it was proposed as a useful tool for analysis of decision states. This approach produces a set of decision rules involves two types namely certain and possible rules based on approximation. The prediction may highly be affected if the data size varies in larger numbers. Application of Rough set theory towards this direction has not been considered yet. Hence the main objective of this paper is to study the influence of data size and the number of rules generated by rough set methods. The performance of these methods is presented through the metric like accuracy and quality of classification. The results obtained show the range of performance and first of its kind in current research trend.

Keywords Rough set theory, Quality of classification, Rule induction, LERS, Accuracy

Paper type Original Article

1. Introduction

Mining datasets is becoming very crucial now days because gaining information from various domain datasets play a vital role in business development. All fields both commercial and official industries need these valuable information let the management to take effective decision in order to improve their sales or to achieve the target. Many of the researchers are involved in proposing algorithms which mining the data's efficiently in the way of improving the performance while generating the decision rules. For the past decade, researchers have been agreed with Rough set theory (RST) that performing effectively in mining with their work. The advantages of rough set theory make the researchers experimenting that in many ways. Zdzislaw pawlak described those advantages in his book on Rough sets [2]. Initially it was around the similarities and differences with fuzzy set theory. After that it has been continuing to evolve RST to reasonable extent. Most of the proposed hybrid methods involve RST and its applications in Data mining. RST is a valuable tool compared to other practices where it can make things easier with quantitative and qualitative attributes in similar without any preliminary information required [18]. Particularly RST was found to be very

© M. Sudha and A. Kumaravel. Published in *Applied Computing and Informatics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Publishers note: The publisher wishes to inform readers that the article "Quality of classification with LERS system in the data size context" was originally published by the previous publisher of Applied Computing and Informatics and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Sudha, M., Kumaravel, A. (2020), "Quality of classification with LERS system in the data size context", *Applied Computing and Informatics*. Vol. 16 No. 1/2, pp. 29-38. The original publication date for this paper was 03/02/2018.



useful for rule generations and attribute selection. RST requires some factors from decision table to induce rules. Rules will be generated once the data get reduced. Classification techniques and effective reduction methods were discussed as RST extension in [20].

LEERS (Learning from Examples based on Rough Sets) is a rule induction algorithm which is based on Rough set theory. We know Rough set handles inconsistent data which makes LEERS to compute lower and upper approximation results to have two types of rules: certain and possible [1]. LEERS system is declared as a successful Rough set theory application of data mining [16]. NASA's Johnson Space Center adopted LEERS as a development tool for expert system. Some institutions found benefits in their research work in Rough sets using this software [19]. In particular, LEERS was effectively applied in the medical field to diagnosis of melanoma, to prediction of behavior in mental retardation, etc [17].

In this paper, rules generated by LEM2, Modified Lem2 under Entropy measure and Laplace measure were analyzed using different kind of benchmark datasets. We know that the equivalence class is the main part of rough set theory to do approximation. Number of equivalence classes stimulated by the datasets contains the original information of datasets. It should be equal before and after reduction process. The change in the number of equivalence classes, quality of classification and number of rules are monitored by increasing the number of instances of a dataset. We outline the interpretation of rules that how they are related and dependent with rule generating algorithms.

2. Materials and methods

2.1 Basic concepts of rough set theory

RST can be defined in terms of lower and upper approximations. The set of instances is called universe U and we assume an equivalence relation R to represents the knowledge about instances in U. To characterize the set X with respect to R, we need the concept of rough set theory that is lower and upper approximations [4].

2.2 Lower and upper approximations

Rough set theory analysis is based on two approximations such as upper and lower approximations [7].

Lower approximation is the union of elements possibly belonging to a concept (set) with respect to R. They definitely belong to the set.

$$R_*(x) = \bigcup_{x \in U} \{R(x) : R(x) \subseteq X\} \quad (1)$$

Upper approximation is the union of elements possibly and partially belonging to a concept (set) with respect to R. i.e., they roughly are in the set.

$$R^*(x) = \bigcup_{x \in U} \{R(x) : R(x) \cap X \neq \emptyset\} \quad (2)$$

Boundary set is the set of all objects that can be neither classified as X and nor X compliment with respect to R. that is the boundary region of the set is the difference between the lower and upper approximations

$$RN_R(x) = R^*(x) - R_*(x) \quad (3)$$

2.3 Quality and accuracy of approximations

Using lower and upper approximations one can calculate the quality and accuracy of approximation [6]. The values will be the numbers between [0,1] and this will describe the instances using the information prescribed in the original data.

The accuracy of approximation is defined as

Accuracy related to negative or positive class

$$= \frac{\text{No. of objects belongs to lower approximation of neg or pos}}{\text{No. of objects belongs to upper approximation of neg or pos}} \quad (4)$$

The quality of classification is defined as

Quality related to decision class

$$= \frac{\text{No. of objects correctly classified as both classes by the attributes}}{\text{No. of objects in the universal set}} \quad (5)$$

The accuracy of approximation decides whether the set is a rough set or a crisp set with respect to set of attributes [5]. If accuracy is not equal to 1, then the set is Rough otherwise it is a Crisp set.

2.4 Core and reduct

The concept of reduction is to find the redundant attributes in the information system and at the same time the irrelevant attributes will be removed. The attributes hold the original information of the data remains as redundant that is they fully describe the knowledge in the database. Those sets of attributes are derived using rough set theory and an attribute set is called a reduct [7]. The set of attributes which is intersection of all reducts called core. An attribute of core cannot be removed from the system unless it fails the equivalence class formation.

Then it is important to analyze the dependencies of attributes with other attributes excluding core. The dependency measure has to be investigated according to the quality of classification.

2.5 Rule induction from decision table

Now the decision table is ready to give perfect pattern in the form of if and then rules. This pattern is the knowledge computed by all instances with respect to the set of attributes and it can identify or test any instance belonging to that knowledge or not. It induces rules in three forms; Minimum set, exhaustive set and satisfactory set. Minimum set induces minimum number of rules enough to express all examples. Exhaustive set induces all rules from examples. Satisfactory set induces rules that convince requirements defined by the user. In rule induction, process of converting numerical attribute to nominal attribute is important. This discretization process may vary according to algorithms. Decision rules and their algorithms were discussed in [8]. The algorithms based on local covering introduced by Grzymala induce all three above mentioned types of decision rule sets [9], ROSE2 [10] is a modular software system implementing basic concepts of the rough set theory and the rule finding methods are used for this study.

2.5.1 LERS system. This represents a system for learning LERS from examples based on rough sets (Figure 1). It is for rule induction that handles missing attributes, numerical attributes and inconsistencies using rough set theory. It will try to form rules mainly from attributes with highest priorities by computing lower and upper approximations for each concept when it finds the input data as inconsistent. Then it induces certain and possible rules from the approximations. We have LEM2 and MODLEM algorithms for rule induction from decision table. LERS system induces a set of rules from a decision table and classifies new examples using the set of rules induced previously by the system [11].

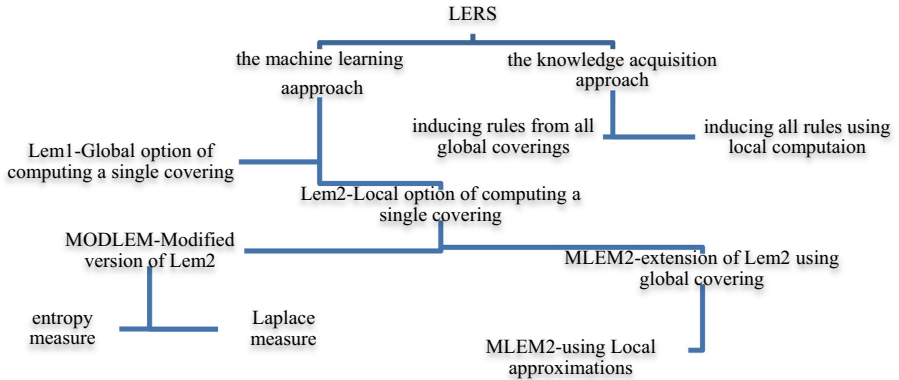


Figure 1. Taxonomy of algorithms available for users in LERS system.

2.5.2 LEM2 algorithm. LEM2 algorithm is a single local covering approach based on LERS system produces minimal discriminated description and gives better result in most cases. Discriminant is a consistent and complete rule set of rule induction. When inducing rules LEM2 is based on the idea blocks of attribute–value pairs by exploring its search space. Rough set theory effectively deals with inconsistent and creates consistent data using definition of lower and upper approximation of a concept. The transformation converts local covering into a rule set in the core of the algorithm.

2.5.3 MODLEM algorithm. MLEM2 (Modified Learning from Examples Module, version 2) is an extension of Lem2 algorithm. In 2003 MLEM2 global approach was introduced that needs preprocessing even though it can induce rules from numerical and missing attribute values. But in 2006 MLEM2 was combined with local approximations that use blocks of attribute-value pairs provide without preliminary preprocessing. So we can input the data with numerical attributes directly where discretization and rule induction is performed at the same time like C4.5 (J48). MODLEM algorithm has two different measures called MODLEM-Entropy and MODLEM –Laplace to evaluate conditions. Entropy measure induces rules covering only small number of examples [12]. As another option, Laplace measure which preferring higher values of Laplace accuracy to induce rules than entropy. For this Laplace measure was used to reduce bias of entropy to get good results. MODLEM-Laplace method is better in inducing the number of rules than entropy method, at the same time entropy is better than Laplace measure in generating rule strength [3].

Data set	Number of instances	Number of attributes	Number of classes
Glass	214	9	7
Iris	150	4	3
Soybean	542	36	14
Cars	159	44	6
Yeast	1484	9	4
Segment	2310	20	7
Abalone	4174	9	28
Thyroid	7200	22	3

Table 1. Data sets.

3. Experiments

To analyze the kinds of rules while varying the size of datasets, we used benchmark datasets of ROSE tool and some from UCI repository. Eight datasets namely glass; iris, cars, soybean, yeast, segment, abalone and thyroid were used for this study. Initially the datasets were preprocessed and minimum number of instances was taken for mining using rough set

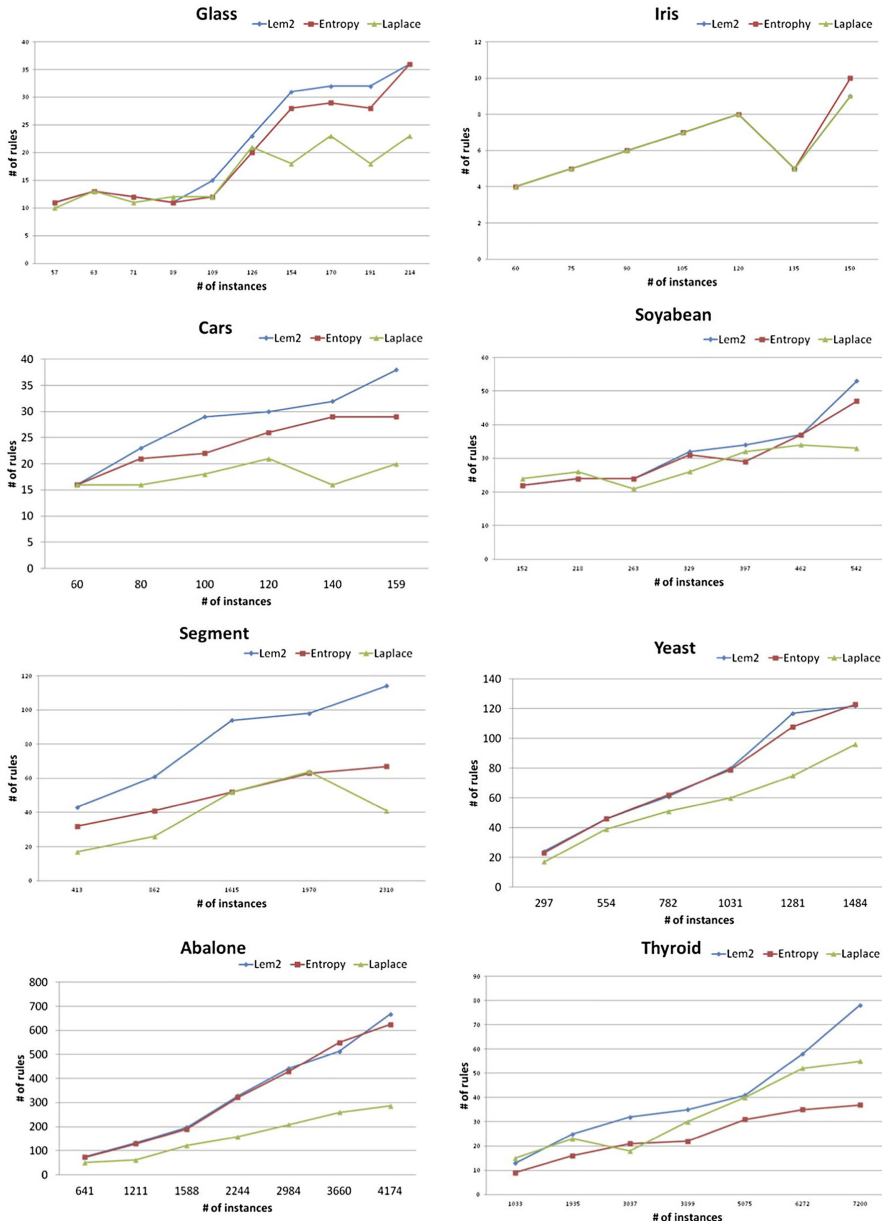


Figure 2.
Number of rules
induced in each data.

theory. Then the number of instances was increased gradually and performances of analysis were monitored and investigated to capture the variations of decision rules in each stage.

Most of the studies on rough set theory applications in data mining are around attribute reduction. This paper tried to test the quality of rule induction algorithms based on rough set theory and also we examined the variations among different size of datasets. Percentages of decision classes of datasets should be stratified in each stage of size. For example the distribution of classes in each set of 'glass dataset' has to be stratified. This dataset is about the study of classification of types of glass that was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence. This dataset has 11 attribute and 7 type of class attributes. More number of instances is labeled as 1 and 2. If majority of a set has one type of class and also no number of instances is labeled by some type of classes are not good for classification and validation. Table 1 summarizes the datasets used for experiments.

The same way the other datasets are also experimented. For each size of dataset, number of equivalence classes, quality of classification, number of rules generated (Lem2, Modlem2 entropy and Laplace methods) and prediction accuracy are calculated. The results are analyzed in order to find the similarities and differences among rules induced methods with respect to prediction accuracy as well as quality of classification.

The strength and specificity of rules among three discretization methods for rule induction such as MODELM-Entropy, MODLEM-Laplace and LEM2 (entropy based discretization) were compared and found that modified version of LEM2 is better than entropy based LEM2 [3]. Also MLEM2 algorithm [11], an extension of LEM2, can induce rules without any prior discretization and can handle missing attributes [12]. The viewpoint of rule induction from examples is double folded. Firstly it is classification-oriented induction whose objective is to build a classifier, and secondly it is discovery-oriented induction whose objective is to extract interesting rules [15]. Explore is a discovery oriented algorithm which extracts all decision rules that satisfy users requirements. The rough set based rule induct algorithms were compared with that explore algorithms in [13,14].

4. Results and discussion

The number of rules induced by lem2 and Modlem entropy is similar to some extent. In overall, Lem2 algorithm induced more number of rules than other methods and the pattern of rules is coinciding with entropy method (Figure 2). We know that Modlem is analogous to Lem2. But here we are analyzing not only the number of rules, also the number of certain and

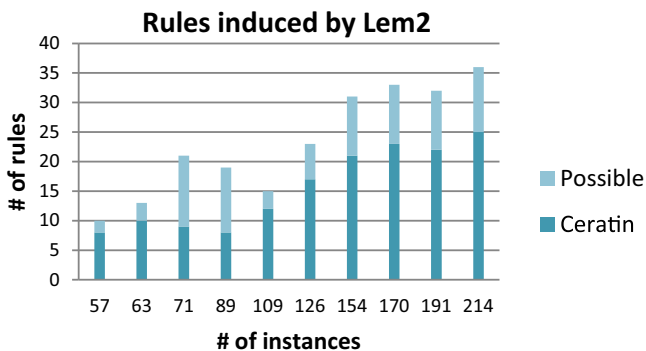


Chart 1.
Lem2 algorithm's rule induction – glass dataset.

possible rules generated by them. Laplace method generates less number of rules with certainty. All algorithms get trained after four or five iterations that is after several stages only, we are able to see some up and downs between them. When the set is in small size, algorithms generates similar type of certainty and when they are in large size, the percentage of certain rules are varying in large number. The pattern obtained from the number of certain rules is similar to the pattern of quality of classification (see Chart 1).

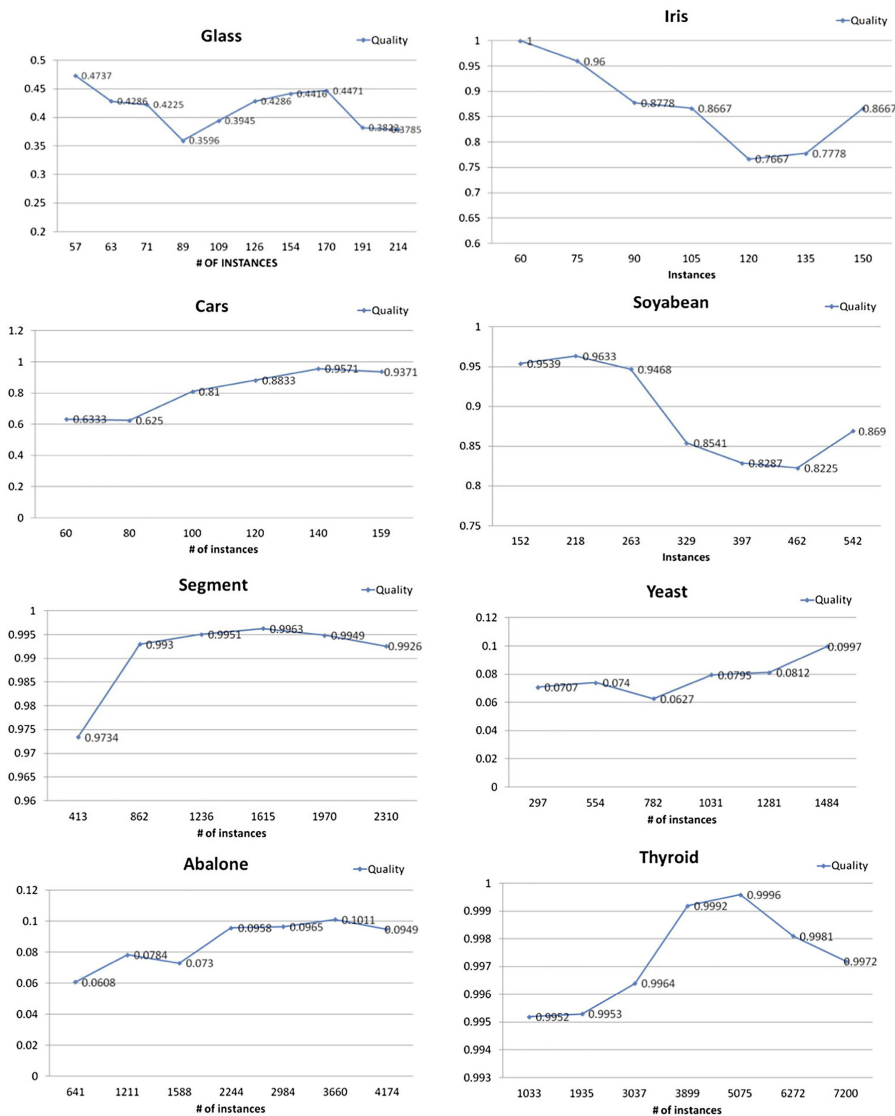


Figure 3.
Quality of
classification
of datasets.

The quality of classification for each stage of datasets was calculated using (5). While increasing the size of dataset, the quality of classification attains noticeable changes which should be analyzed in order to interpret the dependency between quality and induced rules. The results show that percentage of certainty is increasing or decreasing according to the quality of classification (Figure 3). Here the input data of much knowledge induced more number of rules as well as hold précised classification. Therefore the quality is varied in each stage. Also we noticed that the dataset which instances having equal number of some decision classes in large size attains poor quality. If we need to have more number of certain rules, then it's better to select a dataset of size which attains more quality.

The below Table 2 explicate the change of quality and accuracy of an each dataset which shows several changes in data size context. Average change of quality and accuracy were calculated using percentage change formula. We observed that the overall improvement achieved in the cases where less number of rules induced. For instance in Yeast dataset, the algorithms produced more number of attributes with respect to their number of instances compared with other datasets (see Chart 2).

Data set	Average change of quality	Average change of accuracy
Glass	-1.6899	-0.4702
Iris	-1.7839	0.0174
Soybean	-1.2267	-0.5752
Cars	7.2674	3.8014
Yeast	6.8522	-3.1367
Segment	0.327	0.6867
Abalone	7.5225	3.2733
Thyroid	0.0288	0.07948

Table 2.
Observation of quality
and accuracy rate.

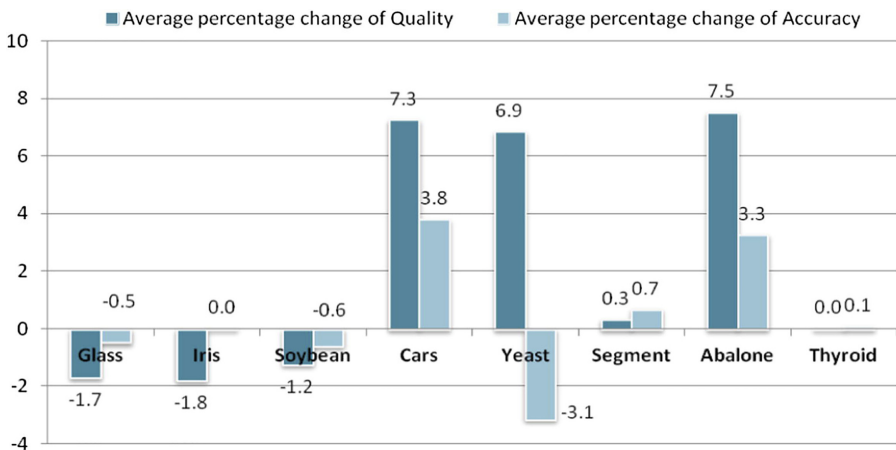


Chart 2.
Average percentage
change of Accuracy per
quality.

5. Conclusion

The paper analyzed and observed the similarities between rule induction algorithms based on Rough set theory at the same time dependency between quality of classification and percentage of certain rules were examined. LERS system based algorithms Lem2 and Modlem (entropy, Laplace method) were used. The results provide evidence that to strengthen certainty of rules requires good quality of classification. To have better knowledge prediction, size of the dataset is not the matter instead the quality and knowledge of the dataset is important. Rules induced by Modlem entropy method are effective in validation than others. The performance of LEM2 algorithm is active on classifying new examples. In case of inconsistent data, rough set based rule induction performs effectively on the sets to induce rules which help to classify new examples.

Increasing the size of instances more or less doesn't bother the quality and accuracy of classification. But all three algorithms are not similar in rule induction, because they show significant differences while generating rules and classifying new examples. In case of increasing examples in large size, the algorithms shows wide range of difference in number of rules at the same time the average percentage change of quality does not affect the prediction accuracy. When dealing with large size of datasets, LERS based algorithms gives better accuracy even if the quality of classification decreases, which minimizes the total error rate. The accuracy is affected in the case more number of rules induced evidence Yeast dataset. We hope this work will contribute to the further research while approaching the process of mining in the way to get better prediction based on rough set theory and we believe that the next step in such analysis should be investigate advance this with rule based classifiers for the development.

References

- [1] J.W. Grzymala-Busse, LERS A system for learning from examples based on rough sets, in: R. Slowinski (Ed.), *Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, 1992, P3_18.
- [2] Z. Pawlak, Rough sets, *Int. J. Computer and Information Sci.* 11 (1982) 341–356.
- [3] J.W. Grzymala-Busse, J. Stefanowski, Three discretization methods for rule induction, *Int. J. Intell. Syst.* 16 (2001) 29–38.
- [4] Z. Pawlak, Rough sets and intelligent data analysis, *Inf. Sci.* 147 (1) (2002) 1–12.
- [5] M. Sudha, A. Kumaravel, Comparative Analysis between Rough set theory and Data mining algorithms on their prediction, *Global J. Pure Appl. Math.* 13 (7) (2017) 3249–3260.
- [6] A. Maciocha, J. Kisielnicki, Intellectual capital and corporate performance, *Electr. J. Knowled. Manage.* 9 (3) (2008) 271–283.
- [7] Z. Pawlak, Rough Set Theory and Its Applications To Data Analysis, *Cyber. Syst. Int. J.* 29 (7) (1998) 661–688.
- [8] Zdzislaw Pawlak, Rough Sets and Decision Algorithms. Ziarko, Y. Yao (Eds.), *RSCTC 2000, LNAI 2005, 2001*, pp. 30–45.
- [9] J. Stefanowski, On rough set based approaches to induction of decision rules, in: L. Polkowski, A. Skowron (Eds.), *Rough Sets in Data Mining and Knowledge Discovery*, Physica Verlag, 1998, pp. 500–529.
- [10] ROSE Software <<http://idss.cs.put.poznan.pl/site/rose.html>>.
- [11] W. Jerzy, Grzymala-busse, a local version of the MLEM2 algorithm for rule induction, *Fundam. Informat.* 100 (2010) 1–18.
- [12] Jerzy W. Grzymala-Busse, A comparison of three strategies to rule induction from data with numerical attributes, in: *Proceedings of the International Workshop on Rough Sets in Knowledge*

-
- Discovery, associated with the European Joint Conferences on Theory and Practice of Software, Warsaw, Poland, April 5–13, 2003, pp. 132–140.
- [13] Mert Bal, Rough sets theory as symbolic data mining method: an application on complete decision table, *Inf. Sci. Lett.* 2 (1) (2013) 35–47.
- [14] J. Stefanowski, D. Vanderpooten, Induction of decision rules in classification and discovery-oriented perspectives, *Int. J. Intell. Syst.* 16 (2001) 13–27.
- [15] R. Mienko, J. Stefanowski, K. Taumi, D. Vanderpooten, Discovery-Oriented Induction of Decision Rules. Cahier du Lamsade, No.141, Université Paris Dauphine, 1996.
- [16] T. Slimani, Application of rough set theory in data mining, *Int. J. Comput. Sci. Network Solut.* 1 (3) (2013) 1–10.
- [17] J.W. Grzymala-Busse, *Rough Set Theory with Applications to Data Mining. Real World Applications of Computational Intelligence.* Springer, Heidelberg, 2004.
- [18] M. Mohamad, A. Selamat, O. Krejcar, K. Kuca, A recent study on the rough set theory in multi-criteria decision analysis problems, in: M. Núñez, N. Nguyen, D. Camacho, B. Trawinski (Eds.), *Computational Collective Intelligence. Lecture Notes in Computer Science*, Springer, Cham, 2015.
- [19] Q. Zhang, Q. Xie, G. Wang, A survey on rough set theory and its applications, *CAAI Trans. Intell. Technol.* 11 (2016) 1–11.
- [20] Yee Leung, Manfred M. Fischer, Wu. Wei-Zhi, Ju-Sheng Mi, A rough set approach for the discovery of classification rules in interval-valued information systems, *Int. J. Approx. Reason.* 47 (2008) 233–246.

Corresponding author

M. Sudha can be contacted at: sudha.m@ametuniv.ac.in