Efficient rough sets based dynamic Agglomerative clustering

Y.TEJASWINI[1]
1M.Tech student,Department of Computer Science & Engineering
Vardhaman college of Engineering,.
Kacharam, Andhra Pradesh,India
tejaswiniy9@gmail.com

H VENKATESWARA REDDY [2]
2Associate Professor, Department of Computer Science & Engineering
Vardhaman college of Engineering,.
Kacharam, Andhra Pradesh,India
h.venkateswarareddy@vardhaman.org

Abstract: There are different types of techiniques are there to extract knowledge from various sources. Critical / rough set has been applied to extract knowledge from various types of databases. Some limitations have been discovered in rough set, such as label inconsistency, the lack of flexibility and excessive dependency on discretization of the initial attributes. To overcome these limitations, a novel agglomerative clustering method using improved rough set is proposed. The idea of using equivalence class was also incorporated to merge and divide subclass. The experimental applications in data extraction and cooperative object localization showed the effectiveness of the presented improved rough set combined with agglomerative clustering.

Keywords: rough set, knowledge, agglomerative clustering.

I. INTRODUCTION

The theory or the scheme of rough sets was developed by scientist known as Zdzislaw Pawlak in the early 1980s. It is a powerful mathematical tool fitted in the area of inductive machine learning to uncover hidden patterns in data. It is also capable of assigning uncertainty to the extracted knowledge, identifying partial or total dependencies (cause–effect relationships) in databases, and eliminating redundant data. And rough set can be used in classification problems. Since it has been intensively studied by many researchers, the theory of rough set has made great advances [4]. And rough set has been applied in the area of medicine [5], engineering [6], finance [7] and many others fields. In addition, hybrid methods have been put forward between rough set and other mathematical methods that improve the quality of decision rules induced by rough set method [8]. The data reduction based on rough set can be very useful in preprocessing input data to neural networks. As a result, hybrid methods have been developed between rough set and neural networks, leading to new models of neurons [9]. Similarly, evolutionary program-based optimization can efficiently generate structures of rough set as strongholds, data patterns and decision rules [10–12]. Moreover, the hybridization of rough set with classical methods such as principal component analysis or Bayesian classifiers produces better quality. More recently, Wen and Lee [13] applied the rough set theory for the function group analysis for phenolic amide compounds. Wang et al. [14] presented a systematic study of the generalized rough sets in six coverings and pure reflexive neighborhood systems. Zhang [15] proposed a new concept of intuitionistic fuzzy soft set and investigated the relationship between intuitionistic fuzzy soft sets and intuitionistic fuzzy relations. Yanto et al. [16] proposed variable precision rough set to deal with problems in clustering categorical data. Long and Huang [17] uses the rough set theory to device an effective source camera identification method.

II. PROBLEM STATEMENT

When working with large data sets of very inconsistent data samples, the knowledge extraction based on rough set generally suffers from some drawbacks: These drawbacks include but not limited to, the lack of flexibility and excessive dependency on the intervals chosen in the discretization of the attributes. We propose a new algorithm that tries to overcome by introducing two improvements: one is the equivalence classes obtained by the method of rough set and the other is the post made from new samples reserved in the data set. The objective will primarily improve the learning of equivalence classes belonging to the boundary region of which has not been able to obtain any certain rule in the application of Variable Precision Rough Set Model. To achieve this goal, it incorporates the concept of an equivalence class. These are composed after a process of clustering of the samples, and it will be useful in the partitioning of equivalence classes.Thus, there are two possible separation of the equivalence classes (an example is shown in Fig. 1), one made from the centers obtained in the clustering and the other working with the new updated examples of knowledge. They have created‘‘subclasses of equivalence’’ which will be defined by the discredited values of condition attributes and new attributes

generated by mathematical equations involving attributes without discretization. These new subclasses may generate both positive and uncertain new rules.

III. SYSTEM DEVELOPMENT

A novel improved rough set combined with agglomerative clustering To carry out the rough set combined with agglomerative clustering, an algorithm has been developed, and consists of the following steps:

- Create the table of decision: the examples are distributed in the data set to be discussed at a table.
- Remove initial knowledge: Variable Precision Rough Set Model is used to refined the results from a clustering of equivalence classes.
- Updating of knowledge: separate the examples closed to equivalence classes other than their own by hyperplanes.
- Test: final rules are tested with new examples obtained.

A. Creating the decision table

Our aim is to express the data set, in which knowledge is extracted, so that it can be treated in the following steps. To do this, select the attribute of decision, which classifies the examples, and the condition attributes, which are the factors able to perform this classification. The ultimate goal is to determine the decision attribute value from the information provided by the condition attributes, such as to get knowledge of underlying rules governing the relationship between these attributes. In the method of rough set, the examples are provided to the algorithm in a decision table in which rows are distributed by the examples available for training and in which each column corresponds to one of the attributes considered. Each cell of the table shows the value of an item in one of these attributes. The value will be expressed both in discrete form and standard form.

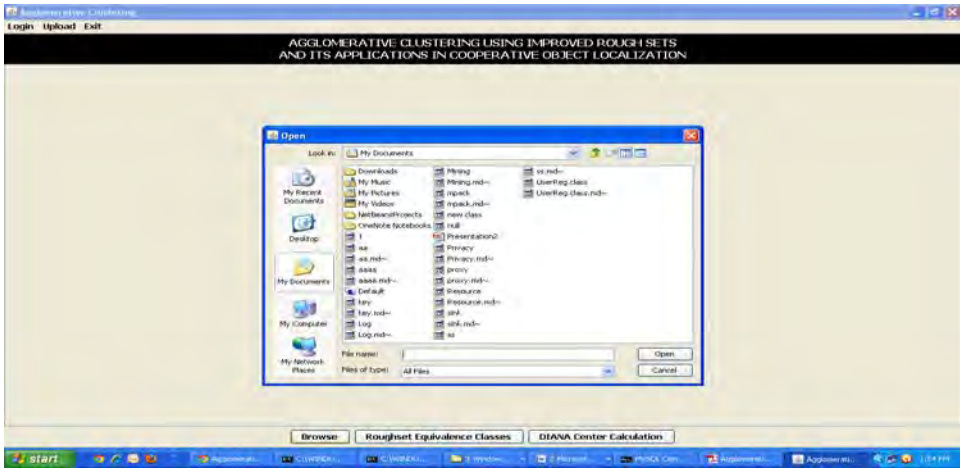

Figure 1: upload dataset

B. Initial extraction:

The aim is to discover rules hidden in the data set. Variable Precision Rough Set Model is used, and then there will be a grouping or clustering process by DIANA method to obtain new knowledge with a greater number of certain rules. Thus, this step is composed of two phases:

- The first applies the method of rough set with an acceptable error level classification, as proposed by the Variable Precision Rough Set Model model.
- The second will be held on clustering in each of the resulting equivalence classes not included in the positive region and have generated certain rules.

Application in knowledge and data extraction:

At this point, there will be a comparison between the methods in knowledge extraction. This will work with real data sets, very different from the UCI repository of databases for machine learning [18]. These sets are widely known and used in this area of knowledge, so that will deepen their description: ‘‘Forest’’ is the most simple, ‘‘Connect-4’’ are also simple sets but not as linearly separable ‘‘Poker’’ are less linear and more complex. Each date set is divided into training set (2/3) and test set (1/3). In rough set divisible, 2/3 of the first form is used for initial training (step 2) and the rest is updated (step 3). In each set, parameters are chosen by trial and error, showed in Table 1.

For analysis rough set, it is necessary for discretization of continuous variables in first step. The Modified Chi2 [28] with a variation, will lower the initial level of consistency L, and more examples are needed to obtain more representative equivalence classes.

As shown in Table 1, the proposed improved rough set combined with agglomerative clustering provide a progressive increase in the hit rate. This is mainly due to:

- The examples of tests that do not belong to any equivalence class of forms in the training are awarded the group's decision included in the positive region with the nearest center. This can be done from the Phase 2 of the initial knowledge, when the centers are calculated.
- A greater number of examples fall into certain rules (many in the following generated equivalence classes), which can be assigned with greater certainty than a decision as examples as true.

Application in cooperative object localization:

The data with localization information provided by wireless multimedia sensor networks (WMSNs) has important application values. WMSNs can monitor environment data, track objects, communicate urgently, and provide intuitive spot audio–visual data in complex environment [23–25]. The data always mapped to a certain coordinates [23–27]. So, the data plays an important role in the application of WMSNs. Localization is the necessary supporting technology that WMSNs must have. At present, the method of locating nodes of WMSNs and that of wireless sensor networks (WSNs) are primarily similar. They mostly use the communication signals between the inner-nodes to locate the unknown nodes. There is a method employed by WMSNs to locate the position of objects with images captured by WMSNs cameras, but the cameras need to be calibrated firstly [25].This paper puts forwards WMSNs objects localization algorithm based on Calibration-Free & Tri-Camera Cooperative method, namely CFTCC object localization algorithm. This method does not use communication signals nor need camera calibration, instead, it works for target segmentation of images got by the cooperative of the beacon cameras, and the scale of the objects can be obtained, so the distance between the targets and the beacon camera can be inversed. Then, we use trigonometric orientation methods to estimate the localization of the object. The results of the experiments show that, the algorithm has good performance of localization in the region constructed by WMSNs cameras.

The main idea for CFTCC localization algorithm can be described as following, the scale of the targets are estimated in the images sampled by the WMSNs cameras, so as to inverse the relatively value of distances from the cameras to the targets.

According to the positions of three or more cooperative WMSNs nodes, the positions of objects are estimated using trilateral measuring method.

Let us consider two ball of color red and black. A red ball with perimeter length of 160 cm is selected to be located. The experiment area is 11 times 8 squares, and the size of each square is 0.6 m times 0.6 m. Take the square floor as the coordinate, and take (0, 0), (8,0) and (5,6) as the cameras' position. Using fixed-focus cameras, the images are sampled in three orientations. The height of the camera from the ground is 105 cm.

- The sampled images and the segmented ones are shown in Fig. 2a and b.
- The estimated positions of the objects and localization errors are shown.
- The experiment results show that the average localization error is 0.1342. The accuracy of objects localization in center area is higher than that in edge. The objects localization performance is good, and it will be a good method to locate and track object for WMSNs. However, in practice, we hope that the locating objects fell in the triangle regions constructed by three WMSN nodes. We hope that this method has broad application prospects in object tracking based on WMSNs and communication.

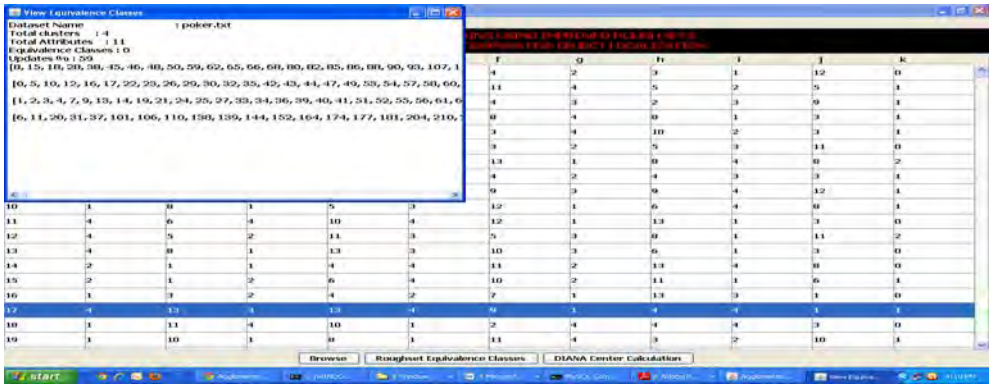

Figure 2 : Diana Center Calculation

IV. CONCLUSION AND FUTURE EHANCEMENT

This paper proposes a novel improved rough set agglomerative clustering approach. It seeks to obtain certain rules from the uncertain rules made by the method of rough set. In this way, the number of new examples increases to be assigned to a considered decision. It also allowed the assignment of a class to the examples that fall in uncertain rules or do not fall into any rules, because any such training was similar to them. The proposed approach permits the choice of several adjustable parameters such as the level of consistency achieved in the discretization, the final number of clusters up to each equivalence class or the minimum percentage of renovation examples required to proceed to the division of an equivalence class. The values assigned to these parameters should be suitable for the instant scenario, considering the fact that, poor selection can cause the generation of rules less significant. As future extension, one may try to use neural networks as the perceptron to substitute the role of the RBFN played in the phase of knowledge updating. Wireless multimedia sensor networks (WMSNs) can locate the positions of objects by images. However, it must be based on the camera calibration. In this paper, equipped with the new clustering technology, a calibration-free and tri-camera cooperative localization method for WMSNs is proposed. Firstly, the method analyzed the scale of images collected by the camera, and anti-performance the distance ratio from object to camera. Secondly, the methods use triangulation localization method to estimate the positions of objects. The results of experiments of locating thirty-two coordinate positions show that the localization algorithm of objects has good performance.

ACKNOWLEDGMENT

I would like to thank my supervisor Associate Professor Mr H Venkateswara Reddy for his guidance, encouragement and assistance throughout the preparation of this study. I would also like to extend thanks and gratitude to all teaching and administrative staff in the vardhaman engineering college, and all those who lent me a helping hand and assistance. Finally, special thanks are due to members of my family for their patience, sacrifice and encouragement.

REFERENCES

[1] Pawlak Z. Rough sets. Int J Comput Inform Sci 1982;11(5):341–56.
[2] Pawlak Z. Rough sets: theoretical aspects of reasoning about data. Dordrecht (Netherlands): Kluwer Academic; 1991.
[3] Pawlak Z, Grzymala-Busse J, Slowinski R, Ziarko W. Rough sets. Commun ACM 1995;38(11):89–95.
[4] Slowinski R. Intelligent decision support. Handbook of applications and advances of the rough sets theory. Boston: Kluwer Academic Publishers; 1992.
[5] Tsumoto S, Tanaka H. Automated discovery of medical expert system rules from clinical databases based on rough sets. In: Proceedings of the second international conference on knowledge discovery and data mining (KDD' 96). Menlo Park: AAAI Press; 1996. p. 63–9.
[6] Nowicki R, Slowinski R, Stefanowski J. Evaluation of vibroacoustic diagnostic symptoms by means of the rough sets theory. Comput Ind1992;20:141–52.
[7] Ziarko W, Golan R, Edwards D. An application of datalogic/R knowledge discovery tool to identify strong predictive rules in stock market data. In: Proceedings of AAAI workshop on knowledge discovery in databases, Washington, DC; 1993. p. 93–101.
[8] Browne C, Düntsch I, Gediga G. IRIS revisited: a comparison of discriminant and enhanced rough set data analysis. En: [15]; 1998. p. 345–68.
[9] Szczuka M. Rough sets and artificial neural networks. En: [15]; 1998. p. 451–71.
[10] Wróblewski J. Genetic algorithms in decomposition and classification problems. En: [15]; 1998. p. 472–92.
[11] Ziarko W. Variable precision rough set model. J Comput Syst Sci 2003;46(1):39–59.
[12] Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: John Wiley & Sons; 1990. p. 253–79.
[13] Wen K-L, Lee YT. Applying rough set theory in the function group analysis for phenolic amide compounds. Comput Elect Eng 2012;38:11–8.
[14] Wang L, Yang X, Yang J, Wu C. Relationships among generalized rough sets in six coverings and pure reflexive neighbourhood systems. Inform Sci 2012;207:66–78.
[15] Zhang Z. A rough set approach to intuitionistic fuzzy soft set based decision making. Appl Math Model 2012;36:

AUTHORS PROFILE

Y.Tejaswini is pursuing M.tech in computer science from Vardhaman college of Engineering, Kacharam village, Shamshabad Mandal, A.P, India. Affiliated to Jawaharlal Nehru Technological University, Hyderabad. Approved by: AICTE, NEW DELHI

Mr H Venkateswara Reddy is pursuing Ph. D in Computer Science and Engineering JNTUH,currently working as Associate Professor in Vardhaman college of Engineering, Kacharam village, Shamshabad Mandal, A.P, India Affiliated to Jawaharlal Nehru Technological University, Hyderabad. Approved by: AICTE, NEW DELHI