CLASSIFICATION OF DATA BY USING ROUGHF SET THEORY AND FUZZY RULE BASE SYSTEM FOR DATA MINING.

" MD MUDASSIR REYAZ","GOPAL PRAJAPATI","AARIF KHAN

1.Student(M.tech),CIIT,indore,India,,2.Asst prof. CIIT,indore,India,3. .Asst prof CIIT,indore,India

ABSTRACT

The paper gives basic ideas of classification of data by using 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 prob-lems pointed out. Rough set theory (RST) is one of the techniques used for feature selection. The rough set theory is a mathematical approach to data analysis, based on classification. One of the main objectives of RST is to reduce data size. RST can solve many problems occurred in data reduction, feature selection and pattern extraction so that we can get rid of redundant data even in the information system with null values or missing data.And a rule base system consists of if-then rules, a bunch of facts, and an interpreter controlling the application of the rules. Fuzzy rule base System extracts rules for the datasets for classification.The other family uses clustering. Fuzzy c-means (FCM) [6], [12], [14] is a data clustering technique in which a data set is grouped into n clusters with every data point in the dataset belonging to every cluster will have a high degree of belonging or membership to that cluster and another data point that lies far away from the center of a cluster will have a low degree of belonging or membership to that cluster.

***Index Terms*—Dimensionality reduction, feature selection, fuzzy rules.clustr**

1.Introduction:

In real world, there are many fields in which huge amount of data is stored and increasing day by day. If we want some information or any kind of fact from that data then that vast amount of data cannot be processed manually by individual or group of persons. So here we need Data Mining to solve these problems.Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives

Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.Data mining uses information from past data to analyse the outcome of a particular problem or situation that may arise. Data mining works to analyse data stored in data warehouses that are used to store that data that is being analyzed. That particular data may come from all parts of business, from the production to the management. Managers also use data mining to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company.There are many Applications of data mining that can be divided into four main types, Classification, Numerical prediction, Association and Clustering. Classification [2,3,4]is a data mining (machine learning) technique used to predict group membership for data instances. The goal of classification is to accurately predict the target class for each case in the data.A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time. In addition to the historical credit rating, the data might track employment history, home ownership or rental, years of residence, number and type of investments, and so on. Credit rating would be the target, the other attributes would be the predictors, and the data for each customer would constitute a case. In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown.Feature selection is the process of selecting a subset of relevant features

One

for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context. For supervised learning, feature selection algorithms maximize some function of predictive accuracy. Because we are given class labels, it is natural that we want to keep only the features that are related to or lead to these classes.

Rough set theory (RST)[9,16,17] is one of the techniques used for feature selection[5,6]. The rough set theory is a mathematical approach to data analysis, based on classification. One of the main objectives of RST is to reduce data size. RST can solve many problems occurred in data reduction, feature selection and pattern extraction so that we can get rid of redundant data even in the information system with null values or missing data.A rule base system consists of if-then rules, a bunch of facts, and an interpreter controlling the application of the rules. Fuzzy rule base System extracts rules for the datasets for classification. There are many ways to extract useful fuzzy rules from the dataset. There are two main approaches to fuzzy rule extraction. One family of approaches uses a fixed partition of the input space to generate fuzzy rules, while the other family uses clustering[12,15,17]

2.BRIEF SURVEY OF SOME EXISTING METHODS:

of the main attraction of a fuzzy rule-based system is its interpretability which is hindered severely with an increase in the dimensionality of the data. For high-dimensional data, the identification of fuzzy rules is also a big challenge. Feature selection methods often ignore the subtle nonlinear interaction that the features and the learning system can have. Most methods of fuzzy rule-based system identification (SI) either ignore feature analysis or do it in a separate phase. To solve this problem Nikhil R. Pal [11] gave a novel neuro-fuzzy system that can simultaneously do feature analysis and SI in an integrated manner. It has a five-layered feed-forward network for realizing a fuzzy rule-based system which can realize a fuzzy rule-based inferencing system and at the same time can find out the features which are not important. The description and architecture of five layered network is shown below:

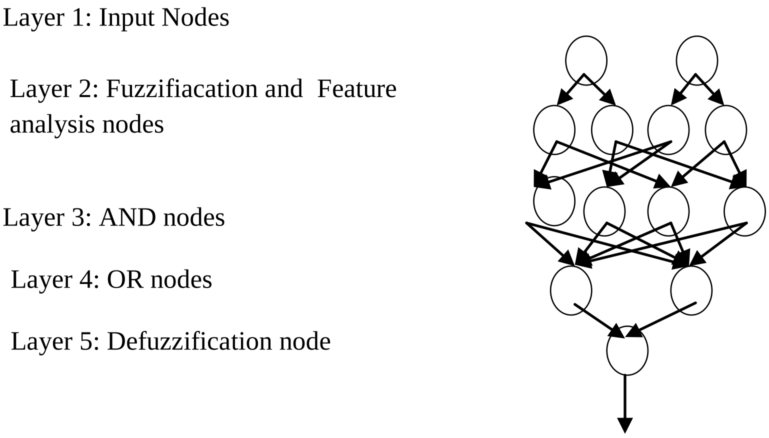


Fig. 3.1 : Network Structure

The neural fuzzy system is realized using a five-layered network, as shown in Fig. 3.1. The node functions with its inputs and outputs are discussed layer by layer. Suffixes p, n, m, l and k is used to denote, respectively, the suffixes of the nodes in layers 1 through 5 in order. The output of each node is denoted by z.

Layer 1: Each node in layer 1 represents an input linguistic variable of the network and is used as a buffer to transmit the input to the next layer, that is to the membership function nodes representing its linguistic values. Thus, the number of nodes in this layer is equal to the number of input features in the data. If x_p denotes the input to any node in layer 1 then the output of the node will be

$$x_p = z_p$$

Layer 2: Each node in layer 2 represents the membership functions of a linguistic value associated with an input linguistic variable. Moreover, this layer also does the feature analysis. The output of these nodes lies in the interval [0,1] and represents the membership grades of the input with respect to different linguistic values. Therefore, the nodes in this layer acts as fuzzifiers. The most commonly used membership functions are triangular, trapezoidal and bell shaped. Although any one of these choices may be used, we consider bell shaped membership functions. All connection weights between the nodes in layer 1 and layer 2 are unity. If there are N_i fuzzy sets associated with the ith feature and if there s are input features then the number of nodes in this layer would be $N^2 = \sum_{i=1}^{s} N_i$. The output of a node in layer 2 is denoted by

$$\bar{z}_n = exp\left\{-\frac{\left(z_p - \mu_n\right)^2}{\sigma_n^2}\right\}$$

Layer 3: This layer is called the AND layer. Each node in this layer represents an IF part of a fuzzy rule. There are many operators for fuzzy intersection. Product is chosen as the operator for intersection. The number of nodes in this layer is $N^3 = \prod_{i=1}^{s} N_i$. The output of the m th node in the layer is

$$z_m = \prod_{n \in P_m} z_n$$

Layer 4: This is the OR layer and it represents the THEN part (i.e., the consequent) of the fuzzy rules. The operation performed by the nodes in this layer is to combine the fuzzy rules with the same consequent. The nodes in layers 3 and 4 are fully connected.

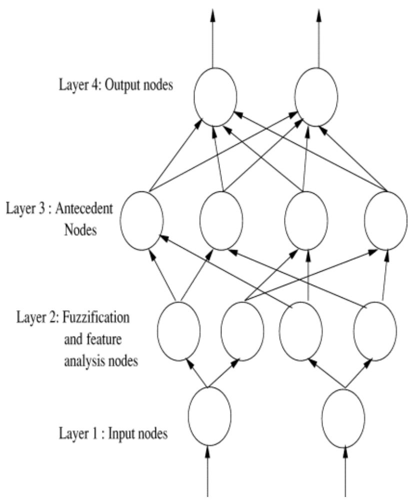
Layer 5: This layer is the defuzzification layer. Each node of layer 5 represents an output linguistic variable and performs defuzzification, taking into consideration the effects of all membership functions of the associated output linguistic variable. The number of nodes in this layer is equal to the number of output features.

In this five-layered feed-forward network for realizing a fuzzy rule-based system, he has used a neural fuzzy system for the purpose of SI. He has not given any guidelines to decide on the number of input and output fuzzy sets and their definitions which are important for designing a good system. A novel scheme for simultaneous feature selection and SI in a neuro-fuzzy framework has been given in paper [11].

In further work N. R. Pal and D. B. [12] gave a neuro-fuzzy scheme for designing a classifier along with feature selection.It is a four-layered feed-forward network for realizing a fuzzy rule-based classifier. The network is trained by error backpropagation in three phases. In the first phase, the network learns the important features and the classification rules. In the subsequent phase, the network is pruned to an “optimal” architecture that represents an “optimal” set of rules. Pruning is found to drastically reduce the size of the network without degrading the performance. The pruned network is further tuned to improve performance. The network can select good features along with the relevant rules in an integrated manner. The network starts with all possible rules and the training process retains only the rules required for classification, thus resulting in a smaller architecture of the final network. The final network has a lower running time than the initial network. This is a new approach to previous one which is given in [11] for designing fuzzy rule based Classifier in a neuro-fuzzy framework. The network architecture and description of network is shown below:

Layer 1: Each node in layer 1 represents an input linguistic variable of the network and is used as a buffer to transmit the input to the next layer, that is to the membership function nodes of its linguistic values.

Layer 2: This is the fuzzification and feature analysis layer, which is similar to the layer 2 of the network described in previous approach [4]. Each node in this layer represents the membership function of a linguistic value associated with an input linguistic variable. The output of a layer 2 node represents the membership grade of the input with respect to a linguistic value



Layer 3: This layer is called the antecedent layer. Each node in this layer represents the IF part of a fuzzy rule.

Layer 4: This is the output layer and each node in this layer represents a class. Thus, if there are c classes then there will be c nodes in layer 4. The nodes in this layer perform an OR operation, which combine the antecedents of layer 3 with the consequents.In further work, to address the same problem of structure identification and feature selection, N. R. Pal and S. Saha [13] propose an integrated method that can find the bad features simultaneously when finding the rules from data for Takagi–Sugeno-type fuzzy systems. It is an integrated learning mechanism that can take into account the nonlinear interactions that may be present between features and fuzzy rule-based systems. Hence, it can pick up a small set of useful features and generate useful rules for the problem at hand. Such an approach is computationally very attr active because it is not iterative in nature like the forward or backward selection approaches. Yet, there is no universally acceptable solution to the structure identification problem, particularly the problem of selecting useful features that are adequate for the task at hand. This issue becomes more serio

when the original dimension of the input is very high. N. R. Pal and S. Saha have proposed an innovative way of handling this problem. Although it is not sure that they have provided a universal solution to the problem, this proposed philosophy is a departure from the usual approaches, and it enjoys some advantages. For the structure identification of fuzzy systems, if features are selected offline (i.e., the importance of features is evaluated in a separate phase) or it is done in an iterative (incremental) manner, then it would be difficult to account for the nonlinear interaction between features and that between the features and the tool being used. There is another family of approaches where different features are combined to form a lower dimensional representation of the input. However, this takes away a very attractive attribute of fuzzy systems, i.e., its readability, because combined features are usually not interpretable. Here, they have proposed a system in which the feature analysis step is integrated into the rule extraction process in such a manner that the system can pick up the required features while solving the task at hand. Thus, such an approach can take into account nonlinear interactions that may be present between features and that between the features and the tool being used. Consequently, it can find a smaller but adequate set of features. Moreover, it is also very attractive computationally because in this neitherthere is need to evaluate different subsets of features northe need to go through an iterative process like forward selection or backward selection approaches. The effectiveness of the approach is demonstrated using several function-approximation/prediction-type problems. In this investigation, they did not deal with very high dimensional data, which needs to be done. In this context, it is worth mentioning that classifier systems designed based on such a philosophy can easily select useful features from data in 20–30 dimensions. If the dimensionality of the data is very high, for example, a few thousands, most clustering algorithms may not find useful clusters, and the proposed method may not be very effective. To deal with such data sets, divide-and-conquer-type approaches, although they may not be optimal, can be tried. The set of features selected by the method may depend on the choice of the initial rule base.

Further this problem has been solved by Ishibuchi et al.[20]. They proposed a hybrid algorithm of two fuzzy genetics-based machine learning approaches (i.e., Michigan and Pittsburgh) for designing fuzzy rule-based classification systems. First, they examine the search ability of each approach to efficiently find fuzzy rule-based systems with high classification accuracy. Next, they combine these two approaches into a single hybrid algorithm. This hybrid algorithm is based on the Pittsburgh approach where a set of fuzzy rules is handled as an individual. Genetic operations for generating new fuzzy rules in the Michigan approach are utilized as a kind of heuristic mutation for partially modifying each rule set. Then, they compare their hybrid algorithm with the Michigan and Pittsburgh approaches. Finally, they examine the generalization ability of fuzzy rule-based classification systems designed by a hybrid algorithm.In paper [20], the authors first examined the search ability of two fuzzy GBML algorithms through computational experiments on commonly used data sets. These two algorithms were based on the Michigan approach and the Pittsburgh approach, respectively. From experimental results, they had the following observations: the Michigan-style fuzzy GBML algorithm had high search ability to efficiently find good fuzzy rules. Because the evolution of fuzzy rule-based systems in the Michigan-style algorithm was driven only by the performance of each fuzzy rule, it did not have high search ability to find a good combination of fuzzy rules. That is, the execution of the

Michigan-style algorithm was not directly related to the optimization of fuzzy rule-based systems. On the other hand, the Pittsburgh-style algorithm could directly optimize fuzzy rule-based systems. Thus, it could find a good combination of fuzzy rules. The Pittsburgh-style algorithm, however, did not have high search ability to efficiently find good fuzzy rules because the performance of each fuzzy rule was not taken into account in the evolution of fuzzy rule-based systems. Next, theycombined the two fuzzy GBML algorithms into a single hybrid algorithm based on these observations. In this hybrid algorithm, the Michigan approach was used for generating good fuzzy rules while the Pittsburgh approach was used for finding good combinations of generated fuzzy rules. In this manner, advantages of these two approaches were utilized in their hybrid algorithm. It was shown by computational experiments that this hybrid algorithm outperformed its non-hybrid versions.

We also studied some methods for only dimensionality reduction. Lot of work has been done in this area. Feature selection techniques aim at reducing the number of unnecessary features in classification rules. Rough set theory represents an objective approach to imperfections in data, all computations are performed directly on data sets, i.e., no feedback from additional experts is necessary. One of the feature of rough set theory is to find a minimal subset of the attribute set that may be used to identify all concepts. That is called 'reduct'. In paper [12,16,18,] the author has given a method to compute reduct.This method of computing all reducts is of exponential worst time complexity with respect to the number of attributes. There are some other methods of feature selection in which, rough set theory has been used to define the necessity of features. In paper [19]a rough set based feature selection approach called Parameterized Average Support Heuristic has been given.

Based on the study of different research papers it can be concluded that many approaches have been given to solve the problem of fuzzy rule extraction, but they cannot perform well on high dimension data. As we can see in all these given approaches, feature selection is performed for dimensionality reduction, but that feature selection technique is integrated in to fuzzy rule generation, that's why fuzzy rule generation has some problem with these approaches. So we can reduce the dimensions explicitly. Rough Set theory is used to reduce the dimension of the data.So that, the proposed system can perform well on high dimensional data.

3.Testing and Results

The proposed system is tested on various datasets to show itseffectiveness.The accuracies obtained with the proposed methodare compared with accuracies obtained with existing methods.

Experimentation Performed with Various Datasets

For experimentation some data sets like Diabetes, BCW, Wine, SPECT heart, Sonar dataset are used. All these datasets are taken from UCI Machine Learning Repository [18].Brief description of all the datasets used, is given in Table 7.1. The details include the information about no. of instances, no. of classes, and no. of features.

Table 7.1: Details of Datasets used for Experimentation

Dataset	No. of Instances	No. of Classes	No. of Features
Diabetes	768	2	8
BCW	683	2	9
Wine	178	3	13

Testing Result with Diabetes Dataset

Diabetes data set is pima indian diabetes data set which is taken from UCI machine learning repository. The data set contains 2 classes of 768 instances each where each class refers to a type of class. When we perform attribute reduction on this dataset then there is no reduction in attribute based on reduct of rough set theory. This data has only 8 attributes and 768 instances and indiscernible classes of 8 attributes is not equal to any reduct of these 8 attributes, so there is no reduction in the attributes in this dataset.

Reduced attributes = 0

Accuracy obtained by existing approach = 49.86%

Accuracy obtained by proposed approach = 78.12%

Testing Result with BCW Dataset

BCW stands for Breast Cancer Wisconsin, this dataset is binary class dataset which means 2 classes are there in BCW dataset. It has 683 instances and 9 attributes, as we can see 683 is big enough for instances. As like diabetes dataset, there is no reduction in attributes when we perform attribute reduction in this dataset. This data has 683 instances and only 9 attributes and indiscernible classes of 9 attributes is not equal to any reduct of these 9 attributes, so there is no reduction in attributes in BCW dataset.

Reduced attributes = 0

Accuracy obtained by existing approach = 32.50%

Accuracy obtained by proposed approach = 96.04%

Testing Result withWine Dataset

This dataset contains 178 instances and 13 attributes. This dataset is 3 class dataset. When attribute reduction is performed on this dataset then we get 10 attributes out of 13 attributes, it means the set of indiscernible classes of 13 attributes is equal to set of indiscernible classes of 10 attributes. Therefore 10 attributes is a reduct of 13 attributes in this dataset.

Reduced attributes = 3

Accuracy obtained by existing approach = 51.12%

Accuracy obtained by proposed approach = 93.25%

Result of Comparison of Classification Accuracy based on Various Datasets:

Various datasets are taken for experimentation of the proposed system and results corresponding to these datasets are presented in Table 7.2. The system is run for some number of iterations and the average classification accuracy obtained for various datasets is as presented in the table. Itshows the comparison of accuracy between the existing method and the proposed method.

Table 7.2 :Comparison of Average Accuracy withVarious Datasets

Dataset	**Accuracy with existing method (%)**	**Accuracy with Proposed method (%)**
Diabetes	49.86	78.12
BCW	32.50	96.04
Wine	51.12	93.25

As we can see, this proposed approach performs better than the existing approach in the case of high dimensional data.

4. CONCLUSION:

In this project, a system for feature selection and extraction of Fuzzy rules is implemented successfully. We have demonstrated that unlike other feature selection methods used in connection with fuzzy rules . The effectiveness of the method is demonstrated using several datasets from the UCI machine learning repository as well as using a synthetic dataset. Our rule generation scheme is not a partition-based scheme. We cluster the data from a class into a number of clusters and each cluster is converted into a rule. Thus, the increase in the dimensionality of the feature space does not have a direct impact on the number of rules. Hence, the rule generation is scalable..In the proposed system, the problem of high dimension data is solved by reducing the dimension through Rough Set Theory. Here dimension reduction using Rough Set Theory gives very useful contribution because one of the main attractions of a fuzzy rule-based system is its interpretability which is hindered severely with an increase in the dimensionality of the data.The proposed system is implemented using Matlab 7.11 (R2010b).

5. REFERENCES:

[1] N. R. Pal and K. K. Chintalapudi, “A connectionist system for feature selection,” *Neural, Parallel Sci. Comput.*, vol. 5, pp. 359–382, 1997.

[2] D. Chakraborty and N. R. Pal, “A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification,” *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 110–123, Jan. 2004.

[3] D. Chakraborty and N. R. Pal, “Integrated feature analysis and fuzzy rule based system identification in a neuro-fuzzy paradigm,” *IEEE Trans.Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 3, pp. 391–400, Jun. 2001.

[4] N. R. Pal, E. V. Kumar, and G. Mandal, “Fuzzy logic approaches to structure preserving dimensionality reduction,” *IEEE Trans. Fuzzy Syst.*,

vol. 10, no. 3, pp. 277–286, Jun. 2002.

[5] N. R. Pal and E. V. Kumar, “Two efficient connectionist schemes for structure preserving dimensionality reduction,” *IEEE Trans. Neural Netw.*,vol. 9, no. 6, pp. 1142–1153, Nov. 1998.

[6] F. Kohavi and G. John, “Wrappers for feature subset selection,” *Artif.Intell.*, vol. 97, no. 1, pp. 273–342, 1997.

[7] D. Chakraborty and N. R. Pal, “Selecting useful groups of features in a connectionist framework,” *IEEE Trans. Neural Netw.*, vol. 19, no. 3,pp. 381–396, Mar. 2008.

[8] D. Muni, N. R. Pal, and J. Das, “Genetic programming for simultaneous

feature selection and classifier design,” *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 36, no. 1, pp. 106–117, Feb. 2006.

[9] I. Guyon, J.Weston, S. Barnhill, and V. Vapnil, “Gene selection for cancer

classification using support vector machines,” *Mach. Learn.*, vol. 46,

pp. 389–422, 2002.

[10] S. Niijima and S. Kuhara, “Recursive gene selection based on maximum

[11] N. R. Pal and S. Saha, “Simultaneous structure identification and fuzzy

rule generation for Takagi–Sugeno models,” *IEEE Trans. Syst., Man,Cybern. B, Cybern.*, vol. 38, no. 6, pp. 1626–1638, Dec. 2008.

[12] N. R. Pal, “A fuzzy rule based approach to identify biomarkers for diagnostic classification of cancers,” in *Proc. IEEE Int. Conf. Fuzzy Syst.*,

2007, pp. 1–6.

[13] G. V. S. Raju, J. Zhou, and R. A. Kisner, “Hierarchical fuzzy control,”*Int. J. Control*, vol. 54, pp. 1201–1216, 1991.

[15] Y. Yoshinari, W. Pedrycz, and K. Hirota, “Construction of fuzzy models

through clustering techniques,” *Fuzzy Sets Syst.*, vol. 54, no. 2, pp. 157–

166, 1993.

[16]W. Ziarko, Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada, Octo-ber 12--15, Springer-Verlag, Berlin, 1993, P.476.

[17] Pawlak, Z.: Data Mining - a Rough Set Perspective In PAKDD’99. 1999.

[18] H. Pomares, I. Rojas, J. Ortega, J. Gonzalez, and A. Prieto, “A systematic

approach to a self-generating fuzzy rule-table for function approximation,”

IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 30, no. 3, pp. 431–447,Jun. 2000.

[19] R. Babuska and U. Kaymak, “Application of compatible cluster merging

to fuzzy modeling of multivariable systems,” in *Proc. Eur. Congr. Intell.Tech. Soft Comput.*, 1995, pp. 565–569.