

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

SURVEY OF DATA CLASSIFICATION USING FUZZY RULE BASE SYSTEM

¹Md Mudassir Reyaz, ²Dipak Kulhare, ³Lokendra Songare.

^{1,2,3}Computer Science Engineering Department, Central India Institute of Technology, Indore
(M.P.) India

ABSTRACT

Data mining is individual of the mainly significant possessions of any organization and especially in healthcare environment. Healthcare milieu is affluent of information though generate familiarity absent of this information is still a grave challenge. Realistic use of healthcare database systems and information discovery and supervision equipment like data mining can exceptionally provide to improvement resolution creation in healthcare. Renovate massive, complex and various healthcare data into information can assist in overprotective cost and preserve high quality of patient care. A variety of data mining technique have gradually more functional to attempt a variety of problems and confront of information detection in administrative and clinical aspect of healthcare. We have pre-process the data so that values that are deliberate in dissimilar scale can be adjusted to notionally common scale. To Perform feature selection for dimensionality reduction. Perform clustering by implementing Fuzzy c-means clustering algorithm. generate Fuzzy rules sousing generated clusters. generate a Fuzzy Rule Base System using these rules for classification

Keywords: Data mining, Fuzzy c-means clustering , feature selection.

INTRODUCTION

Clustering technique is an exploratory data [1, 2,11] analysis technique that can be used to structure the data into groups based on similarities of objects in data. Clustering is mostly unsupervised methods [9] that are k-means [8] and fuzzy c-means [5]. There are many different clustering techniques available to get an appropriate structure into the data, but the obtained results of the techniques and time period for completion of clustering work are highly important. This paper is strongly feel that fuzzy clustering technique is suitable one to have appropriate groups into given data, since fuzzy clustering does not require any prior information about the objects of the data and also fuzzy clustering assigns every object to every cluster with meaningful membership weights between 0 and 1, but crisp or deterministic clustering assigns each object to a unique cluster[10]. The main heart of fuzzy clustering is an objective function which controls all computation parts of clustering and strength of clusters. The shortcoming is that the objective function of fuzzy clustering has to improve the effect of Fuzzy C-Means algorithms Data mining uses information from past data to analyze the outcome of a particular problem or situation that may arise. Data mining works to analyze 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 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 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) 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. 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. The main objective of the project is to use Rough Set Theory for dimensionality reduction and then perform fuzzy rules generation.

RELETED WORK

Sivasankar in at al [1] This paper thus assess the purpose of the data mining method namely Fuzzy Logic rule based classifier in the diagnosis of severity of appendicitis in patients present with right iliac fossa (RIF) pain. It is based on the information already composed about the occurrence of appendicitis from patients data set of around 2230 data sets composed from BHEL Hospital, Tiruchirappall The termination is that Fuzzy logic rule based classifiers can be used an effectual tool for precisely diagnosing the severity of appendicitis. E.P.Ephzibah in at al [2] in this paper a system that merge genetic algorithms and fuzzy expert system is proposed. Genetic algorithm is used to resolve the attributes which have a say more towards the diagnosis of heart ailments which ultimately reduce the number of tests which are essential to be taken by a patient. scheming of this system with fuzzy in assessment with other process improves outcome. The experts' knowledge and well as the support sets have been used in framing the fuzzy rules. Richard Jensen in at al [3] this paper present a original loom based on an included use of fuzzy and rough set theories, to gravely lessen this data dismissal. Formal perception of fuzzy-rough attribute decrease are bring in and illustrated with a simple illustration. The work is functional to the crisis of web classification, significantly reducing dimensionality with minimal loss of information. tentative consequence illustrate that fuzzy-rough reduction is more powerful than the conventional rough set-based approach. Richard Jensen in at al [4] Discovery of hidden patterns and relationships often goes available. Advanced data mining method can help remedy this circumstances. This research has developed a prototype clever Heart Disease

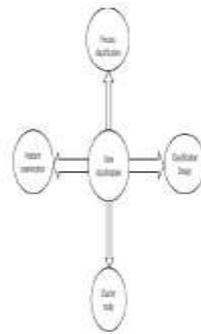
Prediction System (IHDPS) using data mining method, namely, Decision Trees, Naive Bayes and Neural Network. Cordon, O in at al [5] a scheme is future to routinely learn the facts base by decision an suitable data base by means of a hereditary algorithm while using a easy production technique to derive the rule base. Our genetic process learns the number of linguistic terms per variable and the membership function parameters that describe their semantics Latha Parthiban in at al [6] Coactive Neuro-fuzzy replica was proposed as a dependable and strong method developed to classify a nonlinear relationship and mapping between the unlike attributes. It has been shown that of GA is a extremely useful method for auto-tuning of the CANFIS parameters and selection of optimal feature set. C.K. Kwong in at al [7] . In this paper, a genetic algorithm (GA) based knowledge discovery system is proposed to discover information about the fluid provision process. This knowledge is expressed in the form of rules derived from tentative data sets. Rules generate by the GA based knowledge detection system have been validated using a computational system for process optimization of fluid provision. Novruz ALLAHVERDI in at al [8] The aim of this study is to intend a Fuzzy Expert System to decide coronary heart disease (CHD) risk of patient for the next ten-years. The designed system give the user the ratio of the risk and might recommend using one of three results; (1) normal live; (2) diet;(3) drug action.

RESEACH METHODOLOGY

For numerous function district, given a quantity of input–output data, we require to discover a technique of the fundamental system that convert the input to the output. Such request areas include Prediction of medical data classification based on past observations, deciding on organize action based on the present state of the plant, diagnosing diseases based data, and many more. For each of these problems, we have $\{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$, where an unknown system S transforms x to y ; $y = S(x)$. Here, $x = (x_1, x_2, \dots, x_p) \in R_p$ is an input vector and

$y = (y_1, y_2, \dots, y_r) \in R_r$ is the matching output vector of a multiple-input multiple-output system. Our purpose is to discover an S' that approximates S to explain the given input–output data (X, Y) . The popular tools that are used for such a problem are neural networks, fuzzy rule-based systems (FRBS), regression, support vector machines, etc. The achievement of such a system classification task depends strongly on the set of features that is used as input. This is true irrespective of the computational model tool that is used to recognize the comparative between the input and output. Contrary to the customary confidence, more features are not essentially good for system identification. Many features might lead to improved data achievement time and cost, additional design time, more decision making time, more hazards, further degrees of freedom and additional complexity in recognize the system (localminima).Hence, reducingthe dimensionality, if probable, is forever attractive. Classification is used to classify the objects into dissimilar classes. a lot of technique be present for classification, Fuzzy Rule Base system is one of them. Fuzzy Rule Base system cannot execute well in case of High dimensional data; therefore feature selection is necessary for the data set. Feature selection converts high dimensional data into low dimensions, i.e. it take away inappropriate features and selects constructive features. Clustering is required so that the similar type of data can be assigned into the same group. After clustering fuzzy rules are generated, fuzzy rules are generated for classification of dataset.

The Fuzzy C-mean algorithm and Subtractive clustering algorithm are realized to discover the number & the position of clusters for a set of highly non-linear data.

**Figure No. 1: Data Classification**

The clustering consequence find are tested using legality quantity indices that calculate the largely integrity of the clustering effect. The preeminent clustering consequence obtained is used to build input/output fuzzy technique. These fuzzy techniques are used to replica the imaginative data entered to the algorithm. The consequence acquire from the subtractive clustering algorithm are used straight to construct the system, whereas the FCM productivity enter to a Surgeon-type training practice. The find least square modeling error consequence are compared beside comparable current researches in this field. When prospect reliable, the alike functions and settings have been used for each case after that two extremely nonlinear functions are modeled using both algorithms. Ultimately investigational data are representation.

CONCLUSION

This paper has initiated some effective fuzzy c- means technique. For the function of compare the outcomes and identify best scheme the planned technique were introduced to discover the structure in a random data which is produce by this paper. In direct to estimate the effect of technique the random data was separated into two clusters using the proposed technique, and the outcome clusters have converted into data for performance visually the difference in performance of each technique. Available dataset of Heart disease from Repository has been deliberate and pre-processed and clean out to organize it for classification process.

REFERENCES

1. Sivasankar, *EKnowledge discovery in medical datasets using a Fuzzy Logic rule based classifier Electronic Computer Technology (ICECT), 2010 International Conference.*
2. E.P.Ephzibah, *Framing Fuzzy Rules using Support Sets for Effective Heart Disease Diagnosis International Journal of Fuzzy Logic Systems (IJFLS) Vol.2 No.1, February 2012.*
3. Richard Jensen, *Fuzzy-rough attribute reduction with application to web categorization, Copyright © 2003 Elsevier B.V. All rights reserved.*
4. Palaniappan, S. *Intelligent heart disease prediction system using data mining techniques Computer Systems and Applications, 2008. AICCSA 2008.IEEE/ACS International Conference.*
5. Cordon, O. *Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base Fuzzy Systems, IEEE Transactionson (Volume:9, Issue: 4)*
6. Latha Parthiban and R.Subramanian *Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm International Journal of Biological and Life Sciences 3:3 2007.*

7. C.K. Kwong A genetic algorithm base knowledge discovery system for the design of fluid dispensing processes for electronic packaging, *Expert Systems with Applications*, 36(2), 3829–3838. (2008).
8. Novruz allahverdi, design of a fuzzy expert system for determination of coronary heart disease risk, *international conference on computer systems and technologies -compsystech'07*.
9. J.Shapiro: *Genetic algorithms in machine learning. Lecture notes in computer science (Vol.2049/2001, pp. 146–168) (2001)*.
10. Zhu, F., & Guan, S.: *Featur selection for modular GA-based classification. Applied Soft Computing pp.381–393.(2004)*
11. Soler, V., Roig, J., Prim, M.: *Finding Exceptions to Rules in Fuzzy Rule Extraction, KES 2002, Knowledge-based Intel. Information Engineering Systems, Part 2,1115-1119.(2002)*.
12. Allahverdi N., *Expert Systems. An Artificial Intelligence Application Atlas, Istanbul, 248 s., 2002*.
13. Dubitzky W., Schuster A.Hughes J. and Bell D. *An Advanced Case-Knowledge Architecture Based on FuzzyObjects, Applied Intelligence, Volume 7,Number 3 / July, 1997, pp. 187-204*.
14. Kemal Polata, Salih Gune,sa &Sulayman Tosunb: *Diagnosis of heart diseaseusingartificialimmun recognition system and fuzzy weighted pre-processing, Elsevier,Pattern recognition,(2007)*.
15. JJ. C. Bezdek, J. Keller, R. Krisnapuram, N. R. Pal, “*Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*”, Springer, 2005