INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

Surveying Fuzzy Classification and Clustering Techniques

Komal[1], Navdeep Kumar[2]

[1]Haryana Engineering college
Old Chhachhrauli Road ,Jagadhari, Haryana - 135003
Komalrelan.18@gmail.com

[2]Haryana Engineering college
Old Chhachhrauli Road ,Jagadhari, Haryana - 135003
Navdeepkumar17@gmail.com

Abstract: *Data acquisition and maintenance to any industry is multiplying at an enormous speed. In any industry the produced data is not only massive but also quite complex. As a result, proper handling of data is of prime importance in order to convert the available data into useful information that leads to knowledge and apposite decision making. Use of data mining in the healthcare industry is proving to be a boon for attaining speedy, accurate and futuristic results. This paper presents a review of Fuzzy approaches of data mining techniques in and a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use today.*

Keywords: *Data Mining, Clustering, Fuzzy, Classification*

1. INTRODUCTION

Fuzzy logic [1] is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values) fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.

Fuzzy Logic is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PC or workstation-based data acquisition and control systems, MRI Scans[2][8]. It can be implemented in hardware[3], software, or a combination of both. FL provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information.

Compared with Conventional Control Systems

Fuzzy Logic incorporates a simple, rule-based if X and Y then Z approach to a solving control problem rather than attempting to model a system mathematically. The Fuzzy Logic model is empirically-based, relying on an operator's experience rather than their technical under-standing of the system. For example, rather than dealing with temperature control in terms such as "SP =500°F", "T <1000°F", or "210°C <TEMP <220°C", terms like "IF (process is too cool) AND (process is getting colder) THEN (add heat to the process)" or "IF (process is too hot) AND (process is heating rapidly) THEN (cool the process quickly)" are used shown in fig 1.

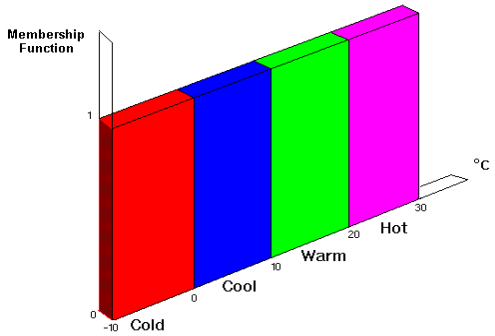

Figure 1: Fuzzy systems in contrast with Non-Fuzzy System

2. HOW IS FL USED[1][2]

- Fuzzy logic defines the control objectives and criteria: What am I trying to control? What do I have to do to control the system? What kind of response do I need? What are the possible (probable) system failure modes?
- Determine the input and output relationships and choose a minimum number of variables for input to the FL engine (typically error and rate-of-change-of-error).
- Using the rule-based structure of FL, break the control problem down into a series of IF X AND Y THEN Z rules that define the desired system

output response for given system input conditions. The number and complexity of rules depends on the number of input parameters that are to be processed and the number fuzzy variables associated with each parameter. If possible, use at least one variable and its time derivative. Although it is possible to use a single, instantaneous error parameter without knowing its rate of change, this cripples the system's ability to minimize overshoot for a step inputs.

- Create FL membership functions that define the meaning (values) of Input/output terms used in the rules.
- Test the system, evaluate the results, tune the rules and membership functions, and retest until satisfactory results are obtained.

3. LITERATURE REVIEW

Puri, Shalini, et al[4] in ""A Fuzzy Similarity Based Concept Mining Model for Text Classification " presents new Fuzzy Similarity Based Concept Mining Model (FSCMM) is proposed to classify a set of text documents into predefined Category Groups (CG) by providing them training and preparing on the sentence, document and integrated corpora levels along with feature reduction, ambiguity removal on each level to achieve high system performance. Fuzzy Feature Category Similarity Analyzer (FFCSA) is used to analyze each extracted feature of Integrated Corpora Feature Vector (ICFV) with the corresponding categories or classes. This model uses Support Vector Machine Classifier (SVMC) to classify correctly the training data patterns into two groups; i.e., +1 and –1, thereby producing accurate and correct results. The proposed model works efficiently and effectively with great performance and high accuracy results.

Jiang, Jung-Yi, , et al[5] in "A fuzzy self-constructing feature clustering algorithm for text classification " presents a a fuzzy similarity-based self-constructing algorithm for feature clustering. The words in the feature vector of a document set are grouped into clusters, based on similarity test. Words that are similar to each other are grouped into the same cluster. Each cluster is characterized by a membership function with statistical mean and deviation. When all the words have been fed in, a desired number of clusters are formed automatically. The authors have presented a fuzzy self-constructing feature clustering (FFC) algorithm, which is an incremental clustering approach to reduce the dimensionality of the features in text classification. Features that are similar to each other are grouped into the same cluster. Each cluster is characterized by a membership function with statistical mean and deviation. If a word is not similar to any existing cluster, a new cluster is created for this word.

Au, W-H., and Keith CC Chan.[6] in "Mining fuzzy rules for time series classification "presents a Time series classification is concerned about discovering classification models in a database of pre classified time series and using them to classify unseen time series. To better handle the noises and fuzziness in time series data, the authors propose a new data mining technique to mine fuzzy rules in the data.

The authors proposed a new data mining technique to discover fuzzy rules in time series data. The fuzzy rules employ fuzzy sets to represent the revealed regularities and exceptions hidden in the data. The use of fuzzy set allows the proposed approach to be resilient the noises hidden in the time series data. To distinguish interesting association relationship from uninteresting ones, their approach utilizes the residual analysis, which has an advantage that it does not require any user-specified thresholds.

A Padmapriya, KSC Maragatham.[7] in **"**Priority Based Apriority Algorithm For Cancer Prediction Using Fuzzy Classification" presents a association rule mining to support the prediction of cancer. In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. They reduce the features using FUZZY based rough set theory and then apply priority based approach. They propose priority based apriority for rule generation. Finally they apply the FUZZY classification approach to classify the dataset as normal or abnormal prediction of cancer. FUZZY based roughest theory is applied for reducing attributes and then increasing accuracy is described. It also proposed a priority based apriori for frequent item generation to reduce time complexity and improving accuracy. Then classification for prediction of class labels is applied. This show proposed work is better than existing one.

OA Tamayo, M Zuluaga, S Ourselin,[8] in – "Fuzzy classification of brain MRI using a priori knowledge: weighted fuzzy C-means" presents a new formulation for the cost function of the well-known fuzzy C-means classification technique whereby they introduce weights. They derive the equations of this new weighted fuzzy C-means algorithm (WFCM) in the presence of additive and multiplicative bias field. They show that the weights can be designed in the same manner as prior probabilities commonly used in maximum a posteriori classifier (MAP) to introduce prior

knowledge (e.g. using atlas), and increase robustness to noise (e.g. using Markov random field). Using prior probabilities of three popular MAP algorithms, they compare the performances of their proposed WFCM scheme using the simulated MRI T1W Brain Web datasets, as well as five T1W MR patient scans. A new weighted fuzzy C-means algorithm has been described which allows to include prior probability atlases and Markov random field techniques developed for maximum a posteriori methods. The new framework allows comparing more fairly non parametric and probabilistic classifiers. Experiments with the Brain Web datasets showed overall best performances using SPM5 with our new proposed WFCM modification. Segmentation of patients T1W images with a signal to noise ratio of about 20 , showed substantial differences between Gaussian mixture model and WFCM.

4. CONCLUSION

The aim of clustering or classification is to minimize a set of data points into self-similar groups such that the points that belong to the same group are more similar than the points belonging to different groups. Each group is called a cluster. This work will present a new algorithm Non-Assisted Fuzzy Based Model for Text Classification for data mining based on Fuzzy Similarity to mine association Rules. The algorithm considers not only exact matches between items, but also the fuzzy similarity between them. The will work unassisted i.e. there should not be a requirement to have an expert for finding similarity between items. It summarizes the techniques to find all association rules that satisfy user-specified minimum support and minimum confidence constraints.

5. FUTURE SCOPE

The aim of clustering or classification is to minimize a set of data points into self-similar groups such that the points that belong to the same group are more similar than the points belonging to different groups. Each group is called a cluster. In our future work we will present a new algorithm Non-Assisted Fuzzy Based Model for Text Classification for data mining based on Fuzzy Similarity to mine association Rules. The algorithm considers not only exact matches between items, but also the fuzzy similarity between them. The algorithm will work unassisted i.e. there should not be a requirement to have an expert for finding similarity between items. It summarizes the techniques to find all association rules that satisfy user-specified minimum support and minimum confidence constraints.

REFERENCES

[1] Novák, V., Perfilieva, I. and Močkoř, J. (1999) Mathematical principles of fuzzy logic Dodrecht: Kluwer Academic. ISBN 0-7923-8595-0

[2] Rajasekaran, S., and GA Vijayslakshmi Pai. Neural networks, Fuzzy logic and Genetic algorithms. PHI Learning Private Limited, 2011.

[3] Kandel, Abraham, and Gideon Langholz, eds. Fuzzy hardware: architectures and applications. Springer, 1998.

[4] Puri, Shalini. "A Fuzzy Similarity Based Concept Mining Model for Text Classification." International Jtheirnal of Advanced Computer Science & Applications 2, no. 11 (2011).

[5] Jiang, Jung-Yi, Ren-Jia Liou, and Shie-Jue Lee. "A fuzzy self-constructing feature clustering algorithm for text classification." Knowledge and Data Engineering, IEEE Transactions on 23, no. 3 (2011).

[6] Au, W-H., and Keith CC Chan. "Mining fuzzy rules for time series classification." In Fuzzy Systems, 2013. Proceedings. 2013 IEEE International Conference on, vol. 1, pp. 239-244. IEEE, 2013.

[7] A Padmapriya, KSC Maragatham –" Priority Based Apriori Algorithm For Cancer Prediction Using Fuzzy Classification" International Journal of Engineering …, 2013

[8] OA Tamayo, M Zuluaga, S Ourselin–"Fuzzy classification of brain MRI using a priori knowledge: weighted fuzzy C-means" 2007 – ieeexplore