ORIGINAL RESEARCH PAPER

Computer Science

STUDY OF FUZZY METHODS IN DATA MINING

|	
Chirag I. jagani	Assistant Professor, Department of Computer Science & I. T. Shree M. & N. Virani Science College, Rajkot.
Dr. Kishor H. Atkotiya*	Professor & Head, Department of Statistics Saurashtra University, Rajkot. *Corresponding Author

ABSTRACT

"*Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner*" [2]. This is to eliminate the randomness and discover the hidden pattern. As these data processing methods are nearly always computationally intensive. We use data mining tools, methodologies, and theories for revealing patterns in data [1]. Fuzzy logic has been applied to various fields, from control theory to AI. It is designed to allow the computer to determine the distinctions among data which is neither true nor false. In this paper our aim to provide how fuzzy logic and fuzzy set can be used in data mining instead of classical crisp theory.

INTRODUCTION

As we know in our society the amount of data doubles every year. Hence we need some computational techniques which can classify these data and extract the useful information out of it. Data mining is of a more explanatory in nature, and patterns discovered in a data set are usually of a descriptive rather than of a predictive nature. This provides the scope for applications of fuzzy logic in the mining system. The main characteristic of fuzzy logic is that it allows us to define values without specifying a particular value, something which isn't possible with classical logic, upon which computer development has been based so far[3]. In classical logic, the membership to one class or set is binary, i.e., one is either member or not, so that only two precise values are worked with (1 and 0, yes or no). But Fuzzy logic is efficient technique for logic having more than one value. The aim of this paper is to provide the information about the contributions of Fuzzy logic in data mining.

The computational strategy of Fuzzy Inference System as follows:
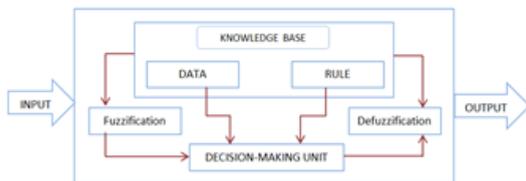
Figure 1: Fuzzy Logic Architecture:

Stage 1: The fresh info is changed over into fuzzy utilizing the fuzzification technique.

Stage 2: After fuzzification, the standard base is shaped.

Step3: The standard base and database mutually allude to the information base.

Stage 4: defuzzification is utilized to change over a fuzzy

incentive to this present reality esteem which is the yield.

Fuzzification assists with changing over fresh numbers into fuzzy sets. Intelligence assists with deciding the level of match between fuzzy information and the standards [6][10]. In view of the rate coordinate, it figures out which rules should be executed as indicated by the given info field. After that, the applied guidelines are joined to build up control activities.

As the title of this section appears, its point is to offer a rundown about fuzzy information handling techniques. Fuzzy information mining strategies can mean information mining techniques that are fuzzy techniques too; then again it can likewise mean ways to deal with dissecting fuzzy information. Fuzzy information means loose, unclear, unsure, questionable, conflicting and additionally fragmented information [14]. In this manner, the wellspring of vulnerability is simply the information. It is essential to create strategies that can deal with this sort of information since the information from a few data sources may be fuzzy. There are calculations to the bunch, to a group or to see fuzzy information. In any case, the methodologies the information is examined with handle the vulnerability on the possibility of emblematic rationale, for example by a bunching issue an example can be the individual from each group all the while with various degrees somewhere in the range of 0 and 1. In the accompanying, a few fuzzy techniques will be displayed for a few sorts of issues.

APPLICATIONS OF FUZZY SET IN DATA MINING
1) Clustering using Fuzzy

In standard clustering, each object is assigned to one cluster in an unequivocal way. Consequently, the individual clusters are separated by sharp boundaries. In practice, such boundaries are often not very natural or even counterintuitive. In fact, the boundary of single clusters and the transition between different clusters are usually "smooth" rather than abrupt. This is the most motivation underlying fuzzy extensions to clustering algorithms. In fuzzy clustering an object may belong to different clusters at the same time, at least to some extent, and the degree to which it belongs to a particular cluster is expressed in terms of a membership degree. The membership functions of the different clusters are usually assumed to form a partition of unity. This version, often called probabilistic clustering, can be generalized further by weakening this constraint: In possibilistic clustering, the sum of membership degrees is constrained to be at least 1. Fuzzy clustering has proved to be extremely useful in practice and is now routinely applied also outside the fuzzy community (e.g. in recent bioinformatics applications). A fuzzy clustering method assigns degrees of membership in several clusters to each input pattern. A fuzzy clustering can be converted to hard clustering by assigning each pattern to the cluster with the largest measure of membership.

2) Association rule mining using fuzzy

An example of simple association rule would be {bread} {butter}. This rule says that if bread was in transaction, butter was in most cases in that transaction too. Such a rule is based on the observations of the customer behavior and is a result from the data stored in transaction databases. The most

complex task of the whole association rule mining process is the generation of frequent item sets. As most of the business databases are very large, the need for efficient algorithms that can extract item sets in a reasonable amount of time is high. Often, a compromise has to be made between discovering all item sets and computation time. Generally, only those item sets that fulfill a certain support requirement are taken into consideration. Support and confidence are the two most important quality measures for evaluating the interestingness of a rule [10]. Based on classical association rule mining, a new approach has been developed expanding it by using fuzzy sets. The new fuzzy association rule mining approach emerged out of the necessity to mine quantitative data frequently present in databases efficiently. Algorithms for mining quantitative association rules have already been proposed [4].When dividing an attribute in the data into sets covering certain ranges of values, we are confronted with the sharp boundary problem. Sometimes, it is not ultimately possible to assign an item to a category. As an example, one can say that a tomato is a vegetable but also, in a way, a fruit. Crisp sets would only allow assigning the item to one single category; fuzzy sets allow different grades of membership to more than one set. Three different approaches to fuzzy association rules can be found in literature which is the quantitative approach, fuzzy taxonomic structures and the approximate item set approach.

3) Data mining with fuzzy methods

The exploration in information disclosure in databases and information mining has lead to an enormous number of proposals for a general model of the information revelation process. An ongoing proposal for such a model, which can be relied upon to have impressive effect, since it is upheld by a few huge organizations like NCR and DaimlerChrysler, is the CRISP-DM model (Cross Industry Standard Process for Data Mining) [5]. The circle shows that information mining is basically a round procedure, where the assessment of the outcomes can trigger a re-execution of the information readiness and model age steps. In this procedure, fuzzy set strategies can gainfully be applied in a few stages: The business comprehension and information understanding stages are normally emphatically human focused and just little computerization can be accomplished here. These stages serve primarily to characterize the objectives of the information revelation venture, to gauge its latent capacity advantage, and to distinguish and gather the fundamental information. Furthermore, foundation space information and Meta information about the information is accumulated. In these stages, fuzzy set techniques can be utilized to define, for example, the foundation space information in dubious terms, yet at the same time in a structure that can be utilized in an ensuing demonstrating stage.fuzzy database questions are valuable to discover the information required and to check whether it might be helpful to take extra, related information into account. In the information arrangement step, the assembled information is cleaned, changed and possibly appropriately scaled to deliver the contribution for the demonstrating procedures. The displaying stage, in which models are built from the information all together, for example, to anticipate future advancements or to assemble classifiers, can, obviously, advantage most from fuzzy information examination draws near. These methodologies can be separated into two classes. The top of the line, fuzzy information examination, comprises of approaches that break down fuzzy information got from uncertain estimation instruments or from the depictions of human space specialists. The second class, fuzzy information investigation, comprises of strategies that utilization fuzzy procedures to structure and break down fresh information, for example, fuzzy bunching for information division and rule age and neuro-fuzzy frameworks for rule age. In the assessment stage, in which the outcomes are tried and their quality surveyed, the handiness of fuzzy displaying strategies turns out to be generally self-evident. Since they yield interpretable

frameworks, they can without much of a stretch be checked for credibility against the instinct and desires for human specialists. Furthermore, the outcomes can give new bits of knowledge into the space viable, rather than, e.g., unadulterated neural systems, which are secret elements.

CONCLUSION

As described in the paper, data mining is the way toward removing nontrivial connections from information. No single procedure is regularly characterized in view of the ideal method from information handling. Fuzzy logic modeling is a probability-based method. It has many advantages as compare to traditional rule induction algorithm; this is discussed in this paper. The selection of technique depends on the matter and therefore the data set. There are great potentials for exploiting these techniques on different data sets to extract information.

REFERENCES:
[1] Data Mining with Fuzzy Methods:Status and Perspectives Rudolf Kruse, DetlefNauck, and Christian Borgelt
[2] Madigan, David. (2002). Principles of Data Mining by David Hand; Heikki Mannila; Padhraic Smyth. SIAM Review. 44. 501-502. 10.2307/4148400.
[3] J.C. Bezdek, J.M. Keller, R. Krishnapuram, and N. Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. The Handbooks on Fuzzy Sets. Kluwer, Dordrecht, Netherlands 1999
[4] Z. Chen and G. Chen, An approach to classification based on fuzzy association rules, in T. Li, Y. Xu, D. Ruan (Eds.), Proc. Int. Conf. Intelligent Systems and Knowledge Engineering, 2007.
[5] C. Borgelt, J. Gebhardt, and R. Kruse. Chapter F1.2: Inference Methods. In: E. Ruspini, P. Bonissone, and W. Pedrycz, eds. Handbook of Fuzzy Computation. Institute of Physics Publishing Ltd., Philadelphia, PA, USA 1998
[6] C. Borgelt, J. Gebhardt, and R. Kruse. Fuzzy Methoden in der Datenanalyse. In: R. Seising, ed. Fuzzy Theorie und Stochastik | Model le und Anwendungen in der Diskussion, 370{386. Vieweg, Braunschweig, Germany 1999
[7] Mandel, J.M. Mar 1995, Fuzzy logic systems for Engineering, Proceedings of the IEEE.
[8] G. Cooper and E. Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning 9:309{347. Kluwer, Dordrecht, Netherlands 1992
[9] P. Chapman, J. Clinton, T. Khabaza, T. Reinartz, and R. Wirth. The CRISP-DM Process Model, 1999 Available from http://www.ncr.dk/CRISP/.
[10] D. Dubois, H. Prade, and R.R. Yager. Merging Fuzzy Information. In: J.C. Bezdek, D. Dudois, and H. Prade, eds. Approximate Reasoning and Fuzzy Information Systems, The Handbooks of Fuzzy Sets, chapter 6, 335{402. Kluwer, Dordrecht, Netherlands 1999 (to appear)
[11] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds. Advances in Know ledge Discovery and Data Mining. MIT Press, Menlo Park, CA, USA 1996
[12] The Prediction Algorithm Based on FuzzyLogic Using Time Series Data Mining Method I. Aydin, M. Karakose, and E. Akin