A Survey on Clusturing Methods based on Fuzzy Rules

Aditi Mehto, Prof. Sandeep Kumar, Namrata Shrivastava

Abstract— Clustering is a technique of grouping the similar items and dissimilar items so that the analysis of any data can be done efficiently and effectively. Although there are various clustering techniques implemented for the analysis of data but the clustering technique used here is based on fuzzy based clusters. Here in this paper a brief survey of all the papers that are related to the clustering based on fuzzy rules are introduced here. The paper also provides the techniques used for the clustering of data using fuzzy rules.

Index Terms— Membership function, fuzzy clustering, supervised clustering, FC Means, FLC.

1 INTRODUCTION

Fuzzy clustering is used when the boundaries among the clusters are vague and confusing. Though, the main limitation of both fuzzy and crisp clustering algorithms is their sensitivity to the number of potential clusters. Moreover, the comprehensibility of obtained clusters is not expertise, whereupon in data-mining, the discovered knowledge is not understandable for human users. FRBC is a supervised classification approach to do the unsupervised clustering. Here it tries to automatically explore the potential clusters in the data patterns and identify them with some other same interpretable fuzzy set of rules. Using the Simultaneous classification of data patterns with these fuzzy rules can reveal the actual boundaries of the clusters [1].

Several methods that uses a fuzzy clustering process within a rapid-prototyping approach and attempt to generate a first approximation to a fuzzy model. Any system is to be described through the existing relations between its input variables and its output variables. To identify such relations, a functional input–output description may be available, but in the case of many complex processes, this is not feasible and we need look for alternative methods. The use of fuzzy models and, more particularly, those described through fuzzy rules has been shown to be successful.

The problem of generation of fuzzy IF-THEN rules is one of the more important problems in the development of fuzzy systems models. This first one is essentially a parameter identification approach; it is characterized by a back-propagation like learning of the antecedent and consequent membership function parameters. In this approach, the description of the antecedent and consequent fuzzy subsets are reduced to a functional form whose parameters are estimated.

Clustering:

Clustering is the unsupervised classification of patterns into groups known as clusters. Clustering is a difficult problem combinatorial, and differences in assumptions and contexts in different communities have made the transfer of useful generic concepts and methodologies slow to occur. Cluster analysis aims at identifying groups of related objects and, hence helps to discover distribution of patterns and interesting correlations in large data sets. So that it can be used in wide research since it arises in many application domains. Especially, in the last years the availability of huge transactional and experimental data sets and the arising requirements for data mining created needs for clustering algorithms that scale and can be applied in diverse domains.

Clustering is considered an interesting approach for finding similarities in data and putting similar data into dissimilar sets. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. The idea of data grouping, or clustering, is simple to use and in its nature and is very near to the human way of thinking; whenever they are presented with a large amount of data, humans are usually tend to summarize this huge number of data into a small number of groups or categories in order to further facilitate its analysis. Most of the data collected in many problems seem to have some inherent properties that lend themselves to natural grouping.

Fuzzy Rule-Based Clustering

This approach attempts to resemble the unsupervised issue of cluster analysis as a supervised classification problem by the addition of some auxiliary data patterns to the main data and usage of a fuzzy classifier to solve this new problem. In this manner, the FRBC tries to repeatedly extract all possible clusters in the problem's data patterns. To extract each cluster, the FRBC considers all unlabeled data patterns of the problem as main data and labels them as Class. It then tries to generate some uniformly distributed instances as auxiliary data. Since the generation of a specific number of uniform patterns is impossible for high-dimensional problems, they are produced randomly and added to the pattern space of the problem as Class 2 to form a two-class problem. The number of main data patterns and their distribution in the problem space would control the number of auxiliary data (i.e., added random instances).

2 RELATED WORK

In 2011 Eghbal G. Mansoori proposed fuzzy rule-based classifiers, the FRBC. It is a supervised classification scheme to do the unsupervised clustering. It will automatically explore the potential clusters in the data patterns and identify them with some interpretable fuzzy rules. The classification of data patterns with these fuzzy rules can reveal the actual boundaries of the clusters.By using an illustrating the capability of FRBC to explore the clusters in data [1].

In 2011 Jingjing Cao and Sam Kwong proposed scheme consist a multi-objective evolutionary hierarchical algorithm to obtain a non-dominated fuzzy rule classifier set and a reduce-error based ensemble pruning method to decrease the size and enhance the accuracy. In this each chromosome represents a fuzzy rule classifier and compose of three different types of genes: control, parameter and rule genes. Similar classifiers are removed to preserve the diversity of the fuzzy system. This approach can maintain a good trade-off among accuracy, change of interpretability and diversity among fuzzy classifiers [2].

In 2009 Jos´e M. Alonso, Manuel Oca˜na, Miguel A. Sotelo, Luis M. Bergasa, and Luis Magdalena proposed a protocol which uses robot localization inside buildings using WiFi signal strength measure. The WiFi signal strength of all visible Access Points (APs) are collected and stored in a database or Wifi map. The protocol uses of Fuzzy Rule-based Classification to obtain the robot position during the estimation Stage. This protocol

is easily adaptable to new environments where triangulation algorithms cannot be applied since the AP physical location is unknown [3].

In 2012 M. Naga Lakshmi, K Sandhya Rani proposed a privacy preserving clustering method that is based on fuzzy approach and random rotation perturbation. The method achieves privacy preservation and retains the information for clustering analysis. The proposed method satisfying the privacy constraints and retains the clustering quality. This method protects the underlying sensitive attribute values when shars the data for clustering over centralized data. The proposed method based on the concept of fuzzy logic and random rotation perturbation. This approach ensures secrecy of confidential numerical attributes without losing accuracy in results. It is effective and provides a feasible approach to balancing privacy and accuracy [4].

In 2012 Keon-Jun Park, Jong-Pil Lee and Dong-Yoon Lee introduced a new category of fuzzy neural networks with multiple-output based on fuzzy clustering algorithm, especially, fuzzy c-mean clustering algorithm (FCM-based FNNm) for pattern classification. The scatter partition of input space generated by FCM clustering algorithm. It will automatically partitioned local spaces describe the fuzzy rules and the number of the partitioned local spaces is equal to the number of clusters. The consequence part of the rules is represented by polynomial functions with multiple-output for pattern classification. Where each of the coefficients of the polynomial functions is learned by back propagation algorithm [5].

In 2010 Sankar K. and Krishnamoorthy K introduced the ant based algorithm which provides a relevant partition of data without any knowledge of the initial cluster centers. During the past researchers have used ant based algorithms that are based on stochastic principles coupled with the k-means algorithm. The new method use the Fuzzy C means algorithm as the deterministic algorithm for ant optimization. This model is used after reformulation and the partitions obtained from the ant based algorithm were better optimized than those from randomly initialized hard C Means. This technique executes the ant fuzzy in parallel for multiple clusters. This enhance the speed and accuracy of cluster formation for the required system problem [7].

In 2010 Mohamed Walid Ayech Karim El Kalti Bechir El Ayeb proposed a method is based on an adaptive distance which calculated according to the spatial position of the pixel in the image. It is novel version of FCM based on dynamic and weighted similarity Distance. The results have shown a significant improvement of this approach performance compared to the standard version of the other FCM techniques especially those regarding the robustness face to noise and the accuracy of the edges between regions [8].

In 2010 Dmitri A. Viattchenin introduced a new scheme for deriving fuzzy classification rules from the interval-valued data. This scheme is based on a heuristic method of possibilistic clustering and a special method of the interval-valued data preprocessing. The concepts of the

heuristic method of possibilistic clustering based on the allotment concept are described and the method of the intervalvalued data preprocessing is also given [9].

Uncertain Data Clustering

Most fuzzy [9] clustering methods are designed for treating crisp data. However, we often have to deal with objects that cannot be described by the quantitative, large or binary signs. In other words, there exists a sign of the object that may assume several values at the same time. Traditional fuzzy clustering methods cannot be applied directly to such types of objects. Fuzzy clustering creates a problem of uncertain data arises. Such a need occurs mostly in medicine, biology, chemistry, economy, sociology and some other domains.

In 2009 G.Sudhavani and Dr.K.Sathyaprasad proposed a modified fuzzy C-means clustering algorithm to the lip segmentation problem. The modified fuzzy C-means algorithm can take the initial membership function from the spatially connected neighboring pixels. The method can perform Successful segmentation of lip images [10].

In 2009 S.Vijayachitra, A.Tamilarasi, and M. Pravin Kumar proposed a new scheme `clustering strategy is implemented in the design of a Fuzzy Logic Controller (FLC) and for the determination of the optimal values of clustering parameters such as weighting exponent and the number of clusters; Genetic Algorithm (GA) is used. Water treatment process, a MIMO process, is chosen here as an application example and GA based Minimum Cluster Volume (MCV) algorithm is proposed which minimizes the sum of the volumes of the individual clusters based on the elimination of redundant rules in the fuzzy rule base thereby reducing the rule firing and computational time and improving optimization [11].

In 2012 Priscilla A. Lopes and Heloisa A. Camargo introduced a semi-supervised clustering algorithm is applied to a partially labeled data set and the obtained results are used to automatically label the remaining data in the set. The supervised learning algorithm is used to generate fuzzy rules from the labeled data. This scheme is good for tasks that have encountered difficulties due to partially labeled data [12].

In 2012 Moez Soltani, Abdelkader Chaari, Fayçal Ben Hmida proposed a new algorithm for fuzzy c-regression model clustering. It is based on adding a second regularization term in the function of a Fuzzy C-Regression Model (FCRM) clustering algorithm which take into account noisy data. A error measure is used in the objective function of the FCRM algorithm. Then, particle swarm optimization is used to finally tune parameters of the obtained fuzzy model set. The method like orthogonal least squares method is used to identify the unknown parameters of the local linear model [13].

In 2012 Khalid Abdel Hafeez, Lian Zhao, Zaiyi Liao, Bobby Ngok-Wah Ma introduced a novel cluster head selection criteria where cluster heads are selected based on their relative speed and distance from vehicles within their neighborhood. The maintenance phase is for drivers behavior on the road and has a learning mechanism for predicting the future speed and position of all cluster members using fuzzy logic inference system. This scheme gives high average cluster head lifetime and more stable cluster topology with less communication and coordination between cluster members compared to other existing schemes [14].

In 2009 Yong Yang introduced an improved fuzzy *c*-means (IFCM) clustering algorithm for image segmentation, it is based on the fact that the conventional FCM-based algorithm considers no spatial context information, which makes it sensitive to noise. The protocol uses the spatial neighborhood information into the original FCM algorithm by *a* priori probability and initialized by a histogram based FCM algorithm. The probability indicates the spatial influence of the neighboring pixels on the centre pixel. This protocol is effective and robust to noise [16].

In 2011 Shahin Ara Begum and O. Mema Devi given the introduction of all the algorithmic methods of fuzzy pattern recognition for medical imaging. They shown several complementary and competing approaches to computer aided diagnosis, including different fuzzy logic,

neural networks and hybrid algorithms [17].

In 2012 Long Sheng a, Xiaoyu Mab proposed a compact and accurate fuzzy rule-based classification method for audio signals. In this protocol Fuzzy *c*-means clustering algorithm with Jumping-gene optimization is used. The novel audio classification scheme classifies audio data into speech, music and background sound. The *c*-means clustering algorithm is added in this method to identify the audio classes in a mixture audio database, in a first step ninitial fuzzy model is determined, in the second step Jumping-Genes Genetic algorithm is determined and at last, vector similarity measure is used to acquire the final simplified model. This protocol can produce satisfactory results, especially for music signals Keywords [18].

In 2012 Ramjeet Singh Yadav & P. Ahmed introduced a protocol which uses K-means and Fuzzy C-Means clustering algorithms to student allocation problem. this method is capable to allocate new students to homogenous groups of specified maximum capacity. This protocol automatically converts crisp sets into fuzzy sets by C-Means clustering algorithm method. The analysis shows that the student group formed by Fuzzy C-Means clustering algorithm performed better than groups formed by K-Means and Hard CMeans clustering algorithm [19].

In 2007 Paulo Salgado and Getúlio Igrejas introduced Probabilistic Fuzzy C-Means (PFCM) algorithm. It applied to fuzzy sets clustering. The protocol leads to a fuzzy partition of the fuzzy rules, one belongs to each cluster, which relates to a new set of fuzzy sub-systems. The protocol applied to the clustering of a flat fuzzy system it gives the result of set of decomposed subsystems that conveniently linked into a Parallel Collaborative Structure [20].

In 2006 Dechang Pi, Xiaolin Qin , and Qiang Wang proposed FCABTAR algorithm for association rules clustering is proposed and applied to association rules managing. The protocol is presented to demonstrate the weakness by the distance clustering, the definition of fuzzy simulation degree, simulated matrix for association rules are put forward and new algorithm based on a dynamic tree is brought forward, which can be used to implement thefuzzy clustering. This protocol can efficiently cluster the association rules for a user to understand [21].

In 2007 Yong Yang and Shuying Huang introduced FCM algorithm for image segmentation which overcomes the noise sensitiveness of conventional fuzzy c-means (FCM). It modifies the objective function of the standard FCM algorithm and takes the influence of the neighboring pixels on the centre pixels. The penalty term acts as a regularizer in this algorithm, which is inspired from the neighborhood expectation maximization algorithm. The analysis on segmentation of synthetic and real images shows that the this algorithm is effective and robust [22].

In 2005 Xing Zong-yi, Zhang Yong, Jia Li-min, Hu Wei-li proposed a new scheme to construct interpretable fuzzy classification system which is based on fuzzy clustering initialization . In this scheme at first the precision index is defined, and the necessary conditions of interpretability are analyzed and at second the initial fuzzy classification system is identified using a fuzzy clustering algorithm, and the number of fuzzy rules is determined by cluster validity measure. In this the genetic algorithm is used to optimize the model in order to improve its precision. [23]

In 2005 Nikhil R. Pal, Kuhu Pal, James M. Keller, And James C. Bezdek introduced possibilistic fuzzy c-means (PFCM) model. It produces memberships and possibilities with the usual point prototypes for each cluster. PFCM is combination of of possibilistic c-means (PCM) and fuzzy c-means (FCM). It solves the problems of PCM, FCM and FPCM overcomes. It solves coincident clusters problem of PCM and eliminates the row sum constraints of FPCM. PFCM prototypes are less sensitive to outliers and it avoids coincident clusters, it is a strong candidate for fuzzy rule-based system identification. [24]

In 2005 Agus Priyono, Muhammad Ridwan, Ahmad Jais Alias, Riza Atiq O. K. Rahmat, Azmi Hassan & Mohd. Alauddin Mohd. Ali proposed a technique expert system part of the urban traffic control system (UTCS) that is developed and implemented to control multi-junctions. The parallel hybrid genetic algorithm optimizes the phase timing. The Two-stage neural network model recognize the traffic pattern and then decide the traffic control strategies. A fuzzy-genetic model estimates the objective values in the optimization process with iterative adjustment of signal timings and offset. This method is applicable to an on-line system because it is trained for extensive traffic condition. It can contribute to an improvement in traffic performance, reliability, and human expert satisfaction.[25]

REFERENCES

[1] Eghbal G. Mansoori "FRBC: A Fuzzy Rule-Based Clustering Algorithm" 2011 IEEE transactions on fuzzy systems, vol. 19, no. 5, october 2011.

[2]Jingjing Cao and Sam Kwong proposed "Combining Interpretable Fuzzy Rule-Based Classifiers via Multi-Objective Hierarchical Evolutionary Algorithm" 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1771 – 1776, 2011.

[3]Jos´e M. Alonso, Manuel Oca˜na, Miguel A. Sotelo, Luis M. Bergasa, and Luis Magdalena "WiFi Localization System Using Fuzzy Rule-Based Classification" 2009 Book Computer Aided Systems Theory - EUROCAST, pp. 383-390, 2009.

[4] M. Naga Lakshmi, K Sandhya Rani "A Privacy Preserving Clustering Method Based On Fuzzy Approach and Random Rotation Perturbation" 2013 Publications Of Problems & Application In Engineering Research – Paper, Vol 04, Special Issue01, Issn: 2230-854, 2012.

[5] Keon-Jun Park, Jong-Pil Lee and Dong-Yoon Lee "Optimal Design of Fuzzy Clustering-based Fuzzy Neural Networks for Pattern Classification" 2012 International Journal of Grid and Distributed Computing Vol. 5, No. 3, September, 2012.

[6] Hamed Malek · Mohammad Mehdi Ebadzadeh Mohammad Rahmati "Three new fuzzy neural networks learning algorithms based on clustering, training error and genetic algorithm" 2011Springer Science+Business Media, LLC 2011, Dec 2011.

[7] Sankar K. and Krishnamoorthy K. "Ant based rule mining with parallel fuzzy cluster" 2010 Bioinfo Publications Advances in Information Mining, ISSN: 0975–3265, Volume 2, Issue 1, 2010, pp.13-17, 2010.

[8] Mohamed Walid Ayech Karim El Kalti Bechir El Ayeb "Image Segmentation Based on Adaptive Fuzzy-C-Means Clustering" 2010 International Conference on Pattern Recognition. 2010.

[9] Dmitri A. Viattchenin "Derivation of Fuzzy Rules from Interval-Valued Data" 2010 International Journal of Computer Applications (0975 – 8887), Volume 7, No.3, September 2010.

[10] G.Sudhavani and Dr.K.Sathyaprasad "Segmentation of Lip Images by Modified Fuzzy C-means Clustering Algorithm" IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.

[11] S.Vijayachitra, A.Tamilarasi, and M. Pravin Kumar "Multiple Input Multiple Output (MIMO) Proces Optimization using Fuzzy GA Clustering" International Journal of Recent Trends in Engineering, Vol 2, No. 2, November 2009.

[12] Priscilla A. Lopes and Heloisa A. Camargo "Semi-supervised Clustering in Fuzzy Rule Generation" 2012 IEEE 13th International Conference on Information Reuse and Integration (IRI), pp. 279 – 286, Aug. 2012.

[13] Moez Soltani, Abdelkader Chaari, Fayçal Ben Hmida "A Novel Fuzzy C–Regression Model Algorithm Using A New Error Measure And Particle Swarm Optimization" 2012 Int. J. Appl. Math. Comput. Sci., Vol. 22, No. 3, pp. 617–628, 2012.

[14] Khalid Abdel Hafeez, Lian Zhao, Zaiyi Liao, Bobby Ngok-Wah Ma "A Fuzzy-Logic-Based Cluster Head Selection Algorithm In Vanets" 2012 Connference on Communications IEEE International (ICC), pp. 203 – 207, June 2012.

[15] Tatiane Marques Nogueira and Heloisa de Arruda Camargo "Fuzzy Rule Base Generation through Conditional Clustering" 2013 IEEE (Revista IEEE America Latina) Latin America Transactions, Volume. 11, Issue 1, ISSN No. 1548-0992, 2013.

[16] Yong Yang "Image segmentation based on fuzzy clustering with neighborhood information" 2009 Optica Applicata, Vol. XXXIX, No. 1, 2009.

[17] Shahin Ara Begum and O. Mema Devi "Fuzzy Algorithms for Pattern Recognition in Medical Diagnosis" 2011 Assam University Journal of Science & Technology on Physical Sciences and Technology, ISSN 0975-2773, Vol. 7, Number II, Dec 2011.

[18] Long Sheng a, Xiaoyu Mab "A Novel GA-fuzzy Classification Method for Audio Signals" 2012 Journal of Information & Computational Science, 2012.

[19] Ramjeet Singh Yadav & P. Ahmed "Academic Performance Evaluation Using Fuzzy C-Means" 2012 International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), ISSN 2249-6831 Vol.2, Issue 4, Dec 2012.

[20] Paulo Salgado and Getúlio Igrejas "Probabilistic Fuzzy Clustering Algorithm for Fuzzy Rules Decomposition" 2007 Advanced Fuzzy and Neural Control, Volume No. 3, pp: 115-120. 2007.

[21] Dechang Pi, Xiaolin Qin , and Qiang Wang "Fuzzy Clustering Algorithm Based on Tree for Association Rules" International Journal of Information Technology, Vol.12, No.3, 2006.

[22] Yong Yang Shuying Huang "Image Segmentation By Fuzzy C-Means Clustering Algorithm With A Novel Penalty Term" 2007 Computing And Informatics, Vol. 26, 2007.

[23] Xing Zong-yi1, Zhang Yong1, Jia Li-min2, Hu Wei-li1 "Construct Interpretable Fuzzy Classification System based on Fuzzy Clustering Initialization" 2005 International Journal of Information Technology, Vol.11 No.6 2005, 2005.

[24] Nikhil R. Pal, Kuhu Pal, James M. Keller, And James C. Bezdek "A Possibilistic Fuzzy C-Means Clustering Algorithm" 2005 IEEE Transactions On Fuzzy Systems, Vol. 13, No. 4, August 2005.

[25] Agus Priyono, Muhammad Ridwan, Ahmad Jais Alias, Riza Atiq O. K. Rahmat, Azmi Hassan & Mohd. Alauddin Mohd. Ali "Generation Of Fuzzy Rules With Subtractive Clustering" 2005 Jurnal Teknologi, pp.143–153, 2005.