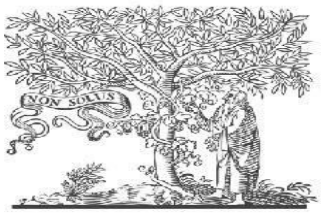
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

IJIEMR Transactions, online available on 29th Jun 2021. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-06

DOI: 10.48047/IJIEMR/V10/I06/52

Title ADVANCED DATA MINING CLUSTERING ALGORITHM BASED ON ARTIFICIAL INTELLIGENCEDECISION MECHANISM
Volume 10, Issue 06, Pages: 264-268
Paper Authors
D.ANIL KUMAR REDDY, B.SIVA KOTI REDDY, D.GANESH, D.MAHESWARA REDDY, P.MADHU SUDHAN, SHAIK MOHAMMAD AYESHA

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ADVANCED DATA MINING CLUSTERING ALGORITHM BASED ON ARTIFICIAL INTELLIGENCEDECISION MECHANISM

[1]D.ANIL KUMAR REDDY, [2]B.SIVA KOTI REDDY,[3]D.GANESH, [4]D.MAHESWARA REDDY, [5]P.MADHU SUDHAN, [6]SHAIK MOHAMMAD AYESHA

[1,2,3,4,5]B.Tech Scholar, Dept of CSE, Nalanda Inst of Engineering & Technology, Sattenapalli, Guntur, Andhra Pradesh, India

[6]Assistant Professor, Dept of CSE, Nalanda Inst of Engineering & Technology, Sattenapalli, Guntur, Andhra Pradesh, India

ABSTRACT: Today a huge amount of data is present in the internet as the network technology and the information technology has been developing rapidly and thus lack of knowledge on it is become increasingly a serious phenomenon. The Data mining in the cloud combines application of conventional data mining under the cloud computing. One of the most popular and widely used algorithms in this cloud data mining is the clustering algorithm. The clustering algorithm maps each and every data to a one group called clusters and hence form a clean partition of the specified data. In this paper data mining clustering by using Fuzzy C-Means (FCM) algorithm based on Artificial Intelligence (AI) decision making mechanism is proposed. One of the most popular unsupervised clustering algorithms is FCM algorithm. The FCM algorithm is developed with the adding of fuzzy entropy function. In this, Probability Based Matching (PBM) index as well as F-measure method is used to validate the clustering results since there is a necessary of using them because of the requirement in FCM algorithm to define the number of clusters in prior, and to define the different cluster values corresponding to different fuzzy partitions. From the results it can be shown that the proposed fuzzy c-mean algorithm with fuzzy entropy can achieve better performance compared with the traditional FCM algorithm and the optimum number of clusters can be determined automatically.

KEYWORDS: Data mining, Cloud computing, Fuzzy K-means Clustering algorithm, Artificial intelligence.

I. INTRODUCTION

At present an incredible speed of data is generated in the internet by this modern society because of its increased popularity and fast development.In addition with the ubiquitous social and community activities various types of data is generated constantly with the logicaltesting, logistic transmission, website access, mobile communication, etc., which indicates that users have entered into a new era of huge increasing of big data. Although this big data in a real time environment isn't just a "big" it has an unpredictable data and difficult problems can be solved using the various data structures for technology requirements in data analysis. Big data is simply seemed in a literal point of view as increaseddata size. An analytical technology is require for this which filter outs thelow density or low valued data and then extracts best of data into the high density or high valued data from knowledge [1].

Various types of new technologies, approaches and applications have been generated in the latest decades with the rapid development of informative industry. Those generated new technologies which contribute to the development of big data are of analytics, mass storage and Internet [2].The data mining has been developed into an interdisciplinary subjectover changing and developing from the few decades in whichvarious disciplineslike databases, statistics, pattern recognition, machine learning, parallel computing and artificial intelligence of relevant information are integrated [3].From the original normal data to today's messy and large amount of data,

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

the objects of data that are examined have been evolved because of the development in the data mining. Thus, research scope has become broadened and the technical requirements became increase.

First a dataset is taken in this and select a set of documents (X). Each of those selected documents(X_i)have 'm' number of features or elements with a m-dimensional vector. Every one of these features are generally normalized to a uniform scaling before clustering as those m number of features regarding to the each document may have different units. In an m-dimensional elements space every X_i document is considered as a point and a set of points with 'n' number of elements is considered as a set of documents(X) from the geometric point of view.The Number of clusters that are specified prior to the clustering process is used by the FCM algorithm in the analysis of fuzzy clustering method to calculate the coordinates of the cluster centers and the partition matrix. Then for several number of clusters values FCM algorithm is run and PBM index as well as F-measure methods are used to calculate the results of clustering validity function for both internal and external performance measurements respectively.

II. DATA MINING AND CLOUD COMPUTING

2.1 Data Mining

The process of extracting inherent, possibly useful and previously unknown information from data is defined as Data mining. The information which is useful to the humans from a huge amount of data is discovered and presented in an understandable format by using the visual, statistical and machine learning methods [4].The semi automated and automated ways are used in the data miming to determine the significant rules and patterns by examining and analyzing the huge quantity of data. Extracting the huge quantity of data is impossible without automation way [5].

The problem of uncovering the hidden as well as useful information of the data in large databases which helps to the help governments and companies for making decisions to get the most out of it can be solved by using this data mining.The Data mining is also called as the KDD (Knowledge Discovery Databases). Prediction and description are the two main goals of this data mining. The future and unknown values of some variables or fields of interest are predicted in the prediction by using the other than those variables of interests in the database. Whereas the finding the human interpretable patterns which describe the datais focused on thedescription [6].

Classification, regression, clustering, transform and deviation detections are some of the methods that achieve the prediction and description objectives.A model is included in the dependency modeling that describes the existence of dependency models on two levels such as structural level and quantitative level. The variables which were depending locally on each other is generally specified in a graphical form in a structural level while the strengths of the dependencies are specified by the some numerical scaling inthe quantitative level of the modeling [7].

2.2 Cloud Computing

Cloud computing is an umbrella term for anything that includes the provision of Internet-hosted services. These services are roughly divided into three types as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). The cloud computing name

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

was inspired by the symbol of cloud that's frequently used in flowcharts and diagramsto represent the Internet. Cloud computing is fast becoming one of the next industry buzzwords. It brings together the following terms: Computing Grid, Utility Computing, Virtualization, Clustering, etc. The Cloud Computing overlaps to some distribution concepts, grid and utility computing, but its meaning when used properly in the context. The conceptual overlap is due in part to changes in technology, usage and implementation over the years. The cloud is a self-managed and self-managed resource virtualization. Of course, there is staff to keep the hardware, operating systems and networks in order. From a user or app developer's perspective, however, this only refers to the cloud. The required services resources for performing the functions with dynamically changing requirements can effectively accessed by the cloud computing. Instead of access requirement from the named resource or specific endpoint, an access from the cloud is required for an application or service developer.

III. ADVANCED DATAMINING CLUSTERING BASED ON AI DECISION MECHANISM

The flow diagram of proposed data mining clustering by fuzzy c-means algorithm is shown in Figure (1). Documents selection then classifying the selected documents, selecting the appropriate keywords, calculating FCM algorithm and selecting the best scored cluster result with the use of PBM index as well as F-measure are the four major stepspresented in the proposed frame work which implement data mining clustering by using Fuzzy c-means algorithm.

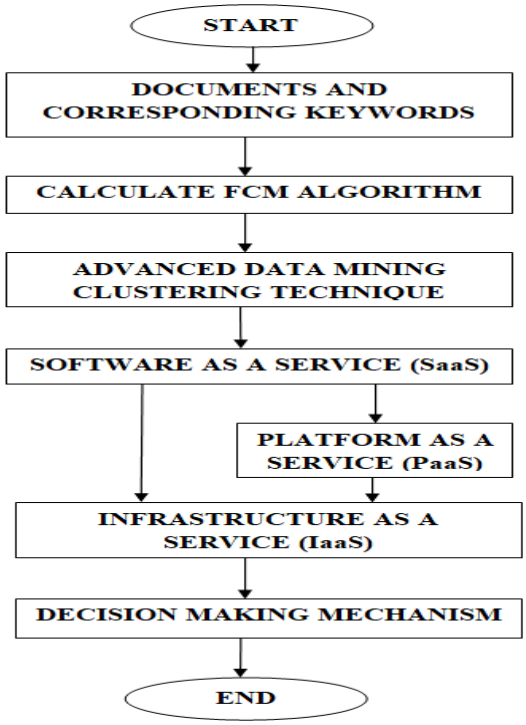


Fig. 1: PROPOSED ALGORITHM

Fuzzy clustering is a form of clustering in which each data point can belong to more than one cluster. Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible.

Software-as-a-Service (SaaS) is a software licensing model, which allows access to software a subscription basis using external servers. SaaS allows each user to access programs via the Internet, instead of having to install the software on the user's computer.

Infrastructure as a Service (IaaS) is a cloud computing service where enterprises rent or lease servers for compute and storage in the cloud. Users can run any operating system or applications on the rented servers without the maintenance and operating costs of those servers.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Platform as a service (PaaS) is a cloud computing model where a third-party provider delivers hardware and software tools to users over the internet.

IV. RESULTS

The experiment results for the proposed FCM added with the fuzzy entropy frame work on clustering the document data is presented in this. 6 number of documents are selected from the dataset with the 19 number of keywords.

Table 1: PERFORMANCE COMPARISON

SI.NO	Parameter	Traditional FCM Clustering	Proposed FCM with Fuzzy Entropy
1	Accuracy (%)	90.05	98.93
2	Precision (%)	88.06	98.14
3	Sensitivity (%)	97.65	98.35
4	Specificity (%)	53.84	98.65

V. CONCLUSION

Clustering analysis has been most commonly used as a significant part of data mining in different fields. There are variety of algorithms for different clustering techniques, each one have their own characteristics and have been utilized in various application areas. This paper proposed data mining clustering using a fuzzy c-means algorithm based on an AI decision mechanism. This Clustering Fuzzy C-means algorithm is a dominant unsupervised technique which constructed and analyzed the data. As initially number of clusters is unknown, usefulness of estimating the optimum number of clusters is proved with the measurements of cluster validity. From the results it can be shown that the better results of clustering calculated using the method of PBM index is differed from the better results calculated with the F-measure due to their individual nature. Then the best clustering results were obtained for proposed Fuzzy c-means algorithm using fuzzy entropy compared to the tradition Fuzzy c-means algorithm by the evaluation of clustering efficacy function, accuracy, precision, sensitivity and specificity.

VI. REFERENCES

[1]Yuan Huang, Qianyu Zhou, Ruixiao Zhao, Yuxing Xiang andZhe Cheng, “Data Mining Algorithm for Cloud Network Information Based on Artificial Intelligence Decision Mechanism”, IEEE Access, Volume: 8, 2020.

[2]M. Shengdong, X. Zhengxian and T. Yixiang, "Intelligent traffic control system based on cloud computing and big data mining", *IEEE Trans Ind. Informat.*, vol. 15, no. 12, pp. 6583-6592, Dec. 2019.

[3]Fabrizio Marozzo, Domenico Talia and Paolo Trunfio, “A Workflow Management System for Scalable Data Mining on Cloud”, IEEE Transactions on Services Computing, Volume: 11, Issue: 3, 2018.

[4]Z. Xia, X., Wang, X., Sun, Q., et al., "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data", *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 340-352, Feb. 2016.

[5]L. Jiang, L. D., Xu, H., Cai, Z., et al., "An IoT-oriented data storage framework in cloud computing platform", *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1443-1451, May 2014.

[6]O. Rubel, C. Geddes, M. Chen, E. Cormier-Michel and E. Bethel, "Feature-based analysis of plasma-based particle acceleration data", *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 2, pp. 196-210, Feb. 2014.

[7]Temporal Yun Yang and Ke Chen, "Clustering via Weighted Clustering Ensemble with Different Representations",IEEE Transactions on Knowledge and Data Engineering, Vol: 23, Issue: 2, 2011

[8], Kun-Ta Chuang, Hung-Leng Chen and Ming-Syan Chen, "On Data Labeling for Clustering Categorical Data", IEEE Trans., Knowledge and Data Engineering, Vol: 20, Issue: 11, 2008.

[9] P.P. Rodrigues, J. Gama and J.P. Pedroso, "Hierarchical Clustering of Time-Series Data Streams", *IEEE Trans. Knowledge and Data Eng.*, vol. 20, no. 5, pp. 615-627, May 2008.

[10]Hichem Frigui and Cheul Hwang, "Fuzzy Clustering and Aggregation of Relational Data With Instance-Level Constraints", IEEE Transactions on Fuzzy Systems, Volume: 16, Issue: 6, 2008.