



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 9, September 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

A Technical Survey on Data Mining Clustering Approach in the Personalized E-Learning System

S.Keerthana¹, P.Vanitha², Dr.S.K.Mahalingam³

PG Scholar, Department of Computer Science and Engineering, Nandha Engineering College, Erode,
Tamil Nadu, India¹

Assistant Professor, Department of Computer Science and Engineering, Nandha Engineering College, Erode,
Tamil Nadu, India²

Professor, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India³

ABSTRACT: Educational data mining techniques aimed toward increasing effectiveness of E-learning process also because the idea of adaptive feedback, individual assessment and more personalized attention to students profile. The personalized e-learning system architecture is presented, which detects and responds to teaching contents consistent with the students' learning capabilities. The primary objective includes the discovery of optimal settings, in which the learners can improve their learning capabilities. The clustering methods K-Means, K-Medoids, Density-based Spatial Clustering of Applications with Noise, Agglomerative Hierarchical Cluster Tree and Clustering by Fast Search and Finding of Density Peaks via Heat Diffusion (CFSFDP-HD) are analyzed using educational data mining. The proposed educational data mining techniques allow assessing student's behavior within the E-learning system for understanding student's interest in studying the training materials and assessing the standard of educational content

KEYWORDS: Educational data mining, Clustering, Profile learning, E- learning.

I. INTRODUCTION

Currently, there is a rapid development of e-education, during which any activity of the student is tracked and recorded in numerous databases, log files, personal profiles, etc. E-learning constantly describing the interaction between training and teaching, electronic systems and students. The analysis of the studies showed that data mining techniques, actively utilized in the E-learning system, are aimed toward different characteristics of the student's profile. In general, there is a wide variety of data mining methods that can be applied in the field of education. These methods are often categorized into classification, clustering, neural networks, and relationship mining. Clustering may be a primary unsupervised method to partition datasets into separate groups (clusters) based on the estimated intrinsic characteristics or similarities.

1.1 CLUSTERING MODELS:

Clustering is that the grouping of specific objects supported their characteristics and their similarities. Clustering helps to splits data into several subsets. Each of those subsets contains data almost like one another, and these subsets are called clusters. There are different types of clustering which are given as follows:

K-Means Clustering

K-Means is one among the foremost widely used and maybe the only unsupervised algorithms to solve the clustering problems. Using this algorithm, classify a given data set through a particular number of predetermined clusters or "k" clusters.

K-medoids

The k-medoids drawback could be a virtually just like the k-means. k-medoids may even be a classical partitioning technique of clump that splits the info set of n objects into k clusters, wherever the number k of clusters assumed known a priori.

DBSCAN

DBSCAN is thought as density-based spatial applications with noise. It finds arbitrary formed clusters and clusters with noise (i.e. outliers).

Agglomerative Hierarchical Clustering

AHC is a reiterative classification technique whose principle is simple. The method starts by conniving the difference between the N objects.

1.2 DATA MINING

The process of extracting information to identify patterns, trends, and useful data that may alter the business to want the data-driven decision from massive sets of knowledge is called data mining. process involves effective knowledge assortment and reposition further as pc processing data processing is additionally referred to as data Discovery in data (KDD).

- Data mining is that the method of analyzing an oversized batch of knowledge to tell apart trends and patterns.
- Data mining programs break down patterns and connections in knowledge supported what data users request or offer.

II. LITERATURE SURVEY

Jingchao Ni , Wei Cheng , Wei Fan, and Xiang Zhang[1],COMCLUS: A Self-Grouping Framework for Multi-Network Clustering,Joint cluster of multiple networks has been shown to be additional correct than activity cluster on individual networks separately. This is often as a result of multi-network cluster algorithms generally assume there's a standard cluster structure shared by all networks, and completely different networks will offer compatible and complementary info for uncovering this underlying cluster structure.

Avory Bryant and Krzysztof Cios[2],RNN-DBSCAN: A Density-Based Clustering Algorithm Using Reverse Nearest Neighbor Density Estimates, Clustering is performed employing a DBSCAN-like approach supported k nearest neighbor graph traversals through dense observations. Every of those agglomeration approaches is delineate by a standard graph-based interpretation whereby clusters of dense observations ar outlined as connected elements, along side a discussion on their computational complexity.

Pradeep Kumar and Anita Kanavalli[3],A Similarity based K-Means Clustering Technique for Categorical Data in Data Mining Application, the analysis study enforced cluster techniques supported the similarity of categorical information. At the same time, the attributes of inter and intra-clusters' similarities area unit known, and so the performance of planned methodology is improved by group action those similarities. The overhead is reduced by developing the Similarity-based K-means cluster (SKC) approach for cluster the attributes that depends on divergence distance.

Arian Dhini and Dhea Indriyanti [4], Clustering High-Dimensional Stock Data using Data Mining Approach, Clustering is helpful to pick out the acceptable stock for investors. Sadly, stock costs keep varied from time to time. Paper presents High Dimensional data cluster (HDDC), a model based mostly agglomeration supported Gaussian Mixture Model, with the Expectation-Maximization (EM) algorithmic rule.

IrfanKamil, BambangPharmasetiawan [5], Fingerprint Presence Fraud Detection Using Tight Clustering on Employee's Presence and Activity Data, analyzing using a supervised algorithm cannot handle unlabeled data that generated uniquely for this case. Tight clustering method to detect fraud in fingerprint data using DBSCAN (Density-based spatial clustering of applications with noise) algorithm, as tight distance calculation removes non-fraud data because non-fraud data is generated to be unique naturally.

Qibing Zhu[6], Improvement of Spatial Data Clustering Algorithm in City Location, has become additional and additional mature, wide utilized in numerous fields. spatial bunch analysis formula will /deeply discovers the information that hidden within the geospatial data, determine the representative node of 1 or variety of spatial information assortment, discovery the law of the spatial distribution.

Bens Pardamean, Join W. C. Sigalingging, Kartika Purwandari, uhammad Fhadli,Shinta Nur Arizky[7], Data Mining for Predicting Customer Satisfaction Using Clustering Techniques,This study aims to work out the applying of the K-means, Spectral cluster (SC), and agglomerated cluster (AC) technique for activity client satisfaction. The cluster analysis supported agglomerated cluster approach performs moreover because the K-means approach to cluster an equivalent characteristics of the customer.

Saryu Chugh and Vanshita R Baweja[8], Data Mining Application in Segmenting Customers with Clustering, This paper describes concerning competition level that's raised between the organizations to retain the customers. Two-phase cluster technique is applied for customers' retention. initial stage is employed to alter the k-means algorithmic program by utilizing a heuristic approach. collective cluster is employed to notice outliers.

Kais Allab, Lazhar Labiod, and Mohamed Nadif[9], A Semi-NMF-PCA Unified Framework for Data Clustering, his paper propose a completely unique to contemplate the cluster and therefore the reduction of the dimension at the same time. Indeed, our approach takes advantage of the mutual reinforcement between knowledge reduction and cluster tasks. the utilization of a low-dimensional representation is of facilitate in providing less complicated and additional explicable solutions.

Qadeem Khan, Saad Rehman[10], Dataset Designing of Software Architectures Styles for Analysis through Data Mining Clustering Algorithms, There are multiple designs in software, like Object orientated, Client-Server, CORBA, Repository, Event-based, Interpreter. The analysis additionally performs agglomeration algorithms on the projected information set with Rapid Miner Studio tool for analysis.

Krishna Rao, Mangathayaru and Sreenivasa Rao[11], Evolution and Prediction of Radical Multi-Dimensional E-Learning System with Cluster based Data Mining Techniques, In the present education system performance evaluation is a big task, because of its multi-dimensional data base. A clustering data mining technique is used to analyses large student data set. A clustering technique will boost the searching process speed and analysis.

Guoqiang Han, Hau-San Wong, and Jane You[12], Adaptive Ensembling of Semi-Supervised Clustering Solutions, In this paper, we have a tendency to propose the transitive closure primarily based constraint propagation approach, that makes use of the transitive closure operator and therefore the affinity propagation to address the primary limitation. Then, the random mathematical space primarily based semi-supervised cluster ensemble framework with a collection of projected confidence factors is meant to deal with the second limitation and supply a lot of stable, robust, and correct results.

Gang Li, Wenqian Jiang, Xiqiao Lin and Zhou Yang[13], An Electricity Data Cluster Analysis Method based on SAGA-FCM Algorithm, The key technology to analyzing electricity knowledge is cluster strategies, of that the standard method has already lost its agility and quality due to the increasing knowledge volume. This paper planned SAGA-FCM algorithm to enhance the information process results, which is a combination of Simulated hardening, Generic algorithmic rule and FCM (Fuzzy C Mean) algorithmic rule.

Rafal A. Angryk and Ruizhe Ma[14], Distance and Density Clustering for Time Series Data, we propose a Distance Density cluster methodology that's a medoid-based cluster with statistic knowledge density thought that provides cluster results in a hierarchy fashion. the space Density cluster technique on the UCR dataset demonstrates that cluster initialization is crucial in obtaining stable and higher results than random initialization on the average, and is additionally additional correct than traditional distance cluster.

C.K. Jha and Seema Maitrey[15], Handling Structured Data Using Data Mining Clustering Technique, Various application areas needed this method, thus, resulted into associate evolution of the many data processing ways. Though many data processing ways get evolved not all of them were capable to manage high voluminous knowledge. This paper place concentrate on CURE cluster technique that found appropriate for operating with massive databases.

Subhash Chandra and Vadlana Baby[16], Distributed threshold k-means clustering for privacy preserving data mining, Privacy protective is vital in whereby data processing turns into a cooperative assignment among members. This paper, propose distributed threshold privacy-preserving kmeans cluster algorithmic program that use the code primarily based threshold secret sharing as a privacy-preserving mechanism.

Bala Mikat Tyoden, Hamza Erol and Recep Erol[17], Classification Performances Of Data Mining Clustering Algorithms For Remotely Sensed Multispectral Image Data, This study compares classification algorithm performances of data mining clustering algorithms for remotely sensed multispectral image data using WEKA data mining software. Clustering algorithm selection is very important for data mining classification method based clustering. This study determines data free clustering algorithms for classification.



Adel Ali Alkhaibari[18], Cluster Analysis for Reducing City Crime Rates, Data analysis plays an important role within the data discovery method of extracting of attention-grabbing patterns or data for understanding varied phenomena or wide applications. Analyzing the situation of the crime and stopped individuals exploitation the explanation of stopped so as to cut back town crime rates. Our analytic and visual results disclosed that the most effective cluster algorithmic program is K-Means algorithmic program, and its smart options making certain that the models area unit useful.

FeipingNie, Joshua Zhexue Huang and Zhou Zhao[19], PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data, Clustering of customer group action information is a crucial procedure to investigate client behaviors in retail and e-commerce companies. “personalized product tree”, named purchase tree, to represent a customer’s group action records that the customers’ group action digital audiotapea set are often compressed into a group of purchase trees. we have a tendency to propose a partitioned agglomeration algorithmic rule, named PurTreeClust, for quick agglomeration of purchase trees. Finally, the agglomeration results area unit obtained by assignment every client to the closest representative. we have a tendency to conjointly propose a spot data point primarily based technique to guage the amount of clusters.

Chaojie Zheng, David Feng and Hui Cui[20], A Unified Collaborative Multikernel Fuzzy Clustering for Multiview, Clustering is progressively necessary for multiview knowledge analytics and current algorithms area unit either supported the cooperative learning of native partitions or directly derived world cluster from multikernel learning. The collaborative learning strategy enables the mutual and interactive clustering from local partitions and global clustering the proposed algorithm outperformed the related state-of-the-art algorithms in comparison, which included multitask, multikernel, and multiview clustering approaches.

III. COMPARATIVE ANALYSIS

Title	Techniques & Mechanisms	Parameter Analysis	Future Work
ComClus : A self grouping framework for multi network clustering	COMCLUS is novel in combining the clustering approach of non-negative matrix factorization (NMF)	Similar clustering structures, Better clustering performance	Network clustering accuracy can be further boosted
RNN-DBSCAN: A Density-Based Clustering Algorithm Using Reverse Nearest Neighbor Density Estimates	. Reverse nearest neighbor based clustering approaches (RECORD, IS-DBSCAN, and ISB-DBSCAN) along with DBSCAN and OPTICS	Problem complexity, improved ability, heterogeneous density, choice of k nearest neighbors.	RNN-DBSCAN is presented leveraging an existing approximate k nearest neighbor technique
A Similarity based K-Means Clustering Technique for Categorical Data in Data Mining Application	Similarity-based K-means Clustering (SKC)	Memory utilization, time consumption, overhead, computation complexity.	The results state that the developed study achieved 98.45% accuracy for the publicly available dataset when comparing with the existing Techniques.



Clustering High-Dimensional Stock Data using Data Mining Approach	This paper Presents High Dimensional Data Clustering (HDDC), a model based clustering based on Gaussian Mixture Model, using the Expectation-Maximization (EM) algorithm.	.Dimension reduction, better quality, improve the performance	EM algorithm enables to handle the high-dimensional data better.
Fingerprint Presence Fraud Detection Using Tight Clustering on Employee's Presence and Activity Data	we propose a tight clustering method to detect fraud in fingerprint data using DBSCAN	Detect fraud, distance calculation	Distance calculation removes non-fraud data because non-fraud data is generated to be unique naturally.
Improvement of Spatial Data Clustering Algorithm in City Location	Spatial clustering analysis algorithm can deeply discover the knowledge which hidden in the geospatial information.	Construct the spatial knowledge base, optimize the query, analysis and optimize the spatial data.	This paper is based on the spatial data mining method, analysis and optimizes the spatial data clustering algorithm in the Location Problem in the city, providing scientific location decisions.
Data Mining for Predicting Customer Satisfaction Using Clustering Techniques.	This study aims to determine the application of the K-means Spectral Clustering and Agglomerative Clustering method for measuring customer satisfaction.	customer satisfaction, cluster analysis	Result of customer satisfaction and provides improvement suggestion to the restaurant concerned.
A Semi-NMF-PCA Unified Framework for Data Clustering	NMF, PNMf and Semi-NMF	Data reduction and clustering tasks	Good performances in terms of reparability between clusters, hence they can also be beneficial for visualization.
Dataset Designing of Software Architectures Styles for Analysis through Data Mining Clustering Algorithms	Software architecture is an important part of the software systems which states that how multiple components of the system interact with each other.	Software reusability, testing and maintenance	Researchers and software industrialists to define their own data set of their organizations for analysing the projects through data mining approaches.



<p>Evolution and Prediction of Radical Multi- Dimensional E-Learning System with Cluster based Data Mining Techniques</p>	<p>In the present education system performance evaluation is a big task, because of its multi-dimensional data base.</p>	<p>Performance evaluation, achievement and future performance.</p>	<p>Evaluating student performance would enhance the student skills, and help the educators and institutions to adopt new learning methods or modifications in the present techniques.</p>
<p>Adaptive Ensembling of Semi-Supervised Clustering Solutions</p>	<p>adaptive semi-supervised clustering ensemble framework is proposed to address the third limitation, which adopts a newly designed adaptive process to search for the optimal subspace</p>	<p>Effectiveness , remove redundant constraints</p>	<p>The proposed approaches work well on most of the real-world datasets.</p>
<p>An Electricity Data Cluster Analysis Method based on SAGA-FCM Algorithm</p>	<p>The key technology to analyzing electricity data is cluster methods, of which the traditional way has already lost its agility and quality due to the increasing data volume. To</p>	<p>Reduced – dimensional, improves the speed, slow convergence and low accuracy.</p>	<p>Prove its availability and the other is to compare its efficiency to conventional algorithm.</p>
<p>Distance and Density Clustering for Time Series Data</p>	<p>Clustering is an important branch in the field of data mining as well as statistical analysis and is widely used in exploratory analysis.</p>	<p>Good and stable clustering, improve distance-based clustering, high time complexity</p>	<p>It would be very useful to develop an effective method to cluster multivariate time series data.</p>



Handling Structured Data using data mining Clustering Technique	In the new era, every organization has the capability to store the extremely large amount of data. The continuous rise in the capturing of data is turning it into a huge tomb of data.	Achieving scalability efficiency, and improving concurrency.	Quality of output clusters produced by it is much improved with respect to those clusters that are resulted from the prior algorithms.
Distributed Threshold K-Means Clustering For PrivacyPreserving Data Mining	.Privacy preserving is important in wherein data mining turns into cooperative assignment among members	Secret sharing scheme and secure addition and comparison protocols	Secret sharing scheme and secure addition and comparison protocols
Classification Performances Of Data Mining Clustering Algorithms For Remotely Sensed Multispectral Image Data	K-means algorithm, EM clustering, Density Based clustering	Algorithm performance comparisons, correctly classified instances.	This study determines data free clustering algorithms for classification.
Cluster Analysis For Reducing City Crime Rates	K-Means clustering, agglomerative clustering, Density-based spatial clustering.	Measures of cluster goodness or quality	We realized that good features play an important role in ensuring that the models are helpful
Text Documents Clustering Using Data Mining Techniques	Rapid automatic keyword extraction (RAKE) algorithm, documents clustering approaches.	Increase the effectiveness, identifying the needs, accurate and reliable results.	Mainly focuses on developing and analyzing the classification of research papers based on clusters topics.
Unified Collaborative Multikernel Fuzzy Clustering For Multiview Data	k-means, fuzzy c-means (FCM), spectral clustering , entropy clustering and possibility fuzzy c-means	Robustness and accuracy, effective description, clustering accuracy	We will extend our method to multimodality biomedical IS and will take the domain knowledge into consideration when choosing kernels.

IV. CONCLUSION

The use of kernel fuzzy cluster analysis to study the educational behavior of students in the E-learning system allowed to identify the main groups of students depending on the time of implementation of educational tasks. With that suitable choice of clustering algorithm that justifies the research questions on student data can be more

effectively used for performance analysis from the large set. The results of the study showed that different students with different levels of academic performance differently distribute their activities in the online environment that is, they have different educational behavior.

REFERENCES

1. Adel Ali Alkhaibari “Cluster Analysis for Reducing City Crime Rates”, computer science and engineering, 2018.
2. Arian Dhini and DheaIndriyanti “Clustering High-Dimensional Stock Data using Data Mining Approach”, 978-1-7281-1941-0/19/\$31.00 ©2019 IEEE.
3. Avory Bryant and Krzysztof Cios “RNN-DBSCAN: A Density-Based Clustering Algorithm Using Reverse Nearest Neighbor Density Estimates”, Ieee Transactions On Knowledge And Data Engineering, Vol. 30, No. 6, June 2018.
4. . Bala Mikat Tyoden, Hamza Erol and Recep Erol “Classification Performances Of Data Mining Clustering Algorithms For Remotely Sensed Multispectral Image Data”, 978-1-5386-5150-6/18/\$31.00 ©2018 IEEE.
5. Bens Pardamean, Join W. C. Sigalingging, Kartika Purwandari, uhammad Fhadli, Shinta Nur Arizky “Data Mining for Predicting Customer Satisfaction Using Clustering Techniques”, 978-1-7281-7071-8/20/\$31.00 ©2020 IEEE 2020 International Conference on Information Management and Technology (ICIMTech).
6. ChaojieZheng, David Feng and Hui Cui “A Unified Collaborative Multikernel Fuzzy Clustering for Multiview”, Ieee Transactions On Fuzzy Systems, Vol. 26, No. 3, June 2018.
7. C.K. Jha and Seema Maitrey “Handling Structured Data Using Data Mining Clustering Technique”, 2019 2nd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) 978-1-7281-1772-0 ©2019 IEEE.
8. Feiping Nie, Joshua Zhexue Huang and Zhou Zhao “PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data”, Ieee Transactions On Knowledge And Data Engineering, Vol. 30, No. 3, March 2018.
9. Gang Li, Wenqian Jiang, Xiqiao Lin and Zhou Yang “An Electricity Data Cluster Analysis Method based on SAGA-FCM Algorithm”, 978-1-5090-4429-0/17/\$31.00 ©2017 IEEE.
10. Guoqiang Han, Hau-San Wong, and Jane You “Adaptive Ensembling of Semi-Supervised Clustering Solutions” Ieee Transactions On Knowledge And Data Engineering, Vol. 29, No. 8, August 2017 1577.
11. Irfan Kamil and Bambang Pharmasetiawan “Fingerprint Presence Fraud Detection Using Tight Clustering on Employee’s Presence and Activity Data”, 4th International Conference 2019.
12. JingchaoNi , Wei Cheng , Wei Fan, and Xiang Zhang “COMCLUS: A Self-Grouping Framework For Multi-Network Clustering”, Ieee Transactions On Knowledge And Data Engineering, Vol. 30, No. 3, March 2018.
13. Kais Allab, LazharLabiod, and Mohamed Nadif “A Semi-NMF-PCA Unified Framework for Data Clustering”, Ieee Transactions On Knowledge And Data Engineering, Vol. 29, No. 1, January 2017.
14. Krishna Rao, Mangathayaru and Sreenivasa Rao “Evolution and Prediction of Radical Multi-Dimensional E-Learning System with Cluster based Data Mining Techniques”, International Conference on Trends in Electronics and Informatics ICEI 2017.
15. Pradeep Kumar and Anita Kanavalli “A Similarity based K-Means Clustering Technique for Categorical Data in Data Mining Application”, International Journal of Intelligent Engineering and Systems, Vol.14, No.2, 2021 .
16. Qadeem Khan, SaadRehman “Dataset Designing of Software Architectures Styles for Analysis through Data Mining Clustering Algorithms”, Intelligent Systems Conference 20177-8 September 2017/London,UK.
17. Qibing Zhu “Improvement of Spatial Data Clustering Algorithm in City Location”, 8th International Conference on Advanced Computational Intelligence Chiang Mai, Thailand; February 14-16, 2016.
18. Rafal A. Angryk and Ruizhe Ma “Distance and Density Clustering for Time Series Data”, 2017 IEEE International Conference on Data Mining Workshops.
19. SaryuChugh and Vanshita R Baweja “Data Mining Application in Segmenting Customers with Clustering”, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) 978-1-7281-4142-8/20/\$31.00 ©2020 .
20. Subhash Chandra and Vadlana Baby “Distributed threshold k-means clustering for privacy preserving data mining”, 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details