Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 3, 4601– 4612

Multi-pipeline Processing Algorithm for Intensive Data Stream Processing of educational data using Soft Computing Techniques

Muzammil H Mohammed

Associate Professor

Department of Information Technology

College of Computers and Information Technology

Taif University, Taif, Saudi Arabia

Abstract

Information mining of monetary administration upheld by soft computing techniques can offer some hypothetical help for the supportable improvement of certain undertakings in monitoring smart girds. To consummate this innovation of our country, in this paper, the important speculations were perceived. Further, the framework was built and the information base was developed. Then, at that point, the development undertakings were taken as instances of the examination, and the information mining was completed for financial administration. The examination results give the premise to advancement of the undertaking. The reason for this investigation is to offer specialized help and reference for follow-up research. The information stream in the time of huge information is expanding, straightforwardly prompting the expanded trouble of the clustering examination. Considering this, the grouping calculation of information stream mining dependent on multi-pipeline processing was contemplated and investigated in this paper. Above all else, the clustering cycle was investigated and the multi-pipeline processing has been presented with the qualities of the average grouping calculations by extended particle swarm optimization (PSO). Finally, the regular grouping calculation and the clustering calculation for information stream mining dependent on the pipeline advancement also been addressed. The outcomes show that the grouping calculation dependent on the multi-pipeline is useful to further develop the clustering virtue in control system.

Keywords: Multi-pipeline process; data stream mining; clustering algorithm; particle swarm optimization, unsupervised learning

1. Introduction

With the rapid development of the times, the world economy is beginning to tend towards globalization and internationalization of information, the demand for basic information about the development of the industry in various industries and fields of the world is increasing under the new development concept, and sharing and mining of information gradually become the basic link of the development of modern industry. As a relatively new technology industry in the development of the times, the development of computer technology provides a new environment for further promotion of various industries. Under the influence of this new development technology, the more perfect development concept and the information dissemination pattern are permeating unceasingly in a variety of fields and industries. As a result, information storage, delivery, communication and processing among different industries have become more complex. As an important link in the development of all industries or fields, the economic management has provided a certain positive influence to its information excavation and the better management of a variety of operating resources in the operation

of the whole enterprise. The role of this link in the economic development of the times has gradually increased, and how to better grasp the opportunities provided by the management of data mining technology under the background of the new era has attracted more and more attention of experts. In recent years, with the popularity of the Internet, the cloud computing, the networking technology and the rapid development of intelligent terminal equipment, all fields of society are producing a large number of data, and the whole world is already in the super data ocean [1]. Not just the size of the information is quickly extending, yet in addition the measure of information is becoming quicker than at any other time. Specifically, the broad utilization of new data advances as of late has brought about remarkable development of information. The quick difference in information climate has become the focal point of consideration and examination in different fields. Moreover, the appearance of the period of huge information has influenced all degrees of society, and it is changing our perspectives, life, work and correspondence, etc [2]. The appearance of the period of enormous information significantly affects the improvement of the PC information handling. As a significant information asset stockpiling focus and the data administration focus, the inward and outside information preparing of PC information are in the sea [3]. PC information handling information assets have been portrayed by the period of large information. Simultaneously, with the quick advancement of data innovation and the nonstop difference in the data climate, the interest of the clients to the data has shown the qualities of the broadening, personalization, information and elements [4].

At present, in the field of the educational management, the data mining, the association analysis, the clustering, the outlier analysis, the concept description and the deviation detection have been applied and popularized. The use of data mining technology can discover interesting knowledge from the data, so as to make decisions can further strengthen the teaching management and improve the level of teaching by using these knowledge [5]. Clustering analysis is an effective way to deal with a large number of complex and large data, and it has been widely used in many areas of life.

In addition, applying the particle swarm optimization algorithm to the clustering algorithm can make clustering algorithm group learning and self-learning in the search process, and this unique memory function can improve the function of the traditional static search. Therefore, realizing the dynamic adjustment of particle swarm algorithm is very useful for the application of clustering algorithm based on particle swarm optimization in the data processing, and we can analyze the correlation between various factors in the flow of information. With an ever increasing number of information data, the bunching calculation of information stream mining dependent on the particle swarm enhancement can additionally improve, change and foster the exactness of data expectation [6].

Clustering seeks to locate a collection of comparable object groupings inside a dataset yet maintaining dissimilar items segregated into various groups or the noisy points category. It is a fundamental aspect of data mining which has been widely employed in a variety of sectors. Clusters are defined as places with high intensity than the rest of the dataset in density-based grouping, that is a popular type of clustering technique. Noise and boundary points are commonly thought to be objects in such scarce regions that are necessary to divide groupings. Density-based clustering algorithms don't need the number of cluster nodes as an argument, and clusters don't have to be convex; they may be formed anyway they want in the database.

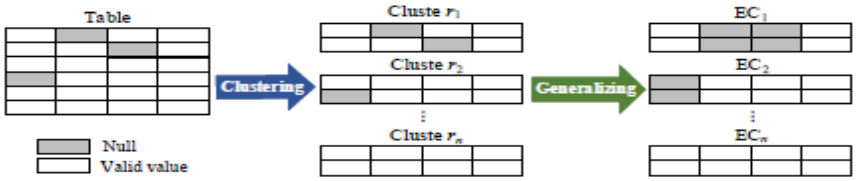

Figure 1 Generalization Procedure in Clustering

The preceding anonymization approaches, unlike those used to anonymize static datasets, confront unique obstacles when dealing with data streams. A data flow, for starters, has an endless capacity and flows indefinitely. Stationary data is frequently retrieved to prevent systematic error. Several scanners need a considerable amount of time and area. For streaming data, such a condition is unacceptable. As a result, anonymizing data streams should entail only one scan followed by ongoing analysis. When it comes to abstracted information availability, large tuple delays should be avoided while processing. Furthermore, due of inconsistencies in acquisition technology, equipment malfunctions, or data loss, missing values are typical in actual data sources. As per a review of the UCI supervised classification repository, 11 of the 21 health-related datasets include missing values, with the proportion of incomplete data exceeding 15%. Statistics reveal that discrepancies make up a significant fraction of datasets. Traditional privacy preservation methods, on the other hand, are predicated on the premise that data does not include missing values. As a result, prior to processing, the treatment is to eliminate incomplete records. Immediate removal of insufficient data, on the other hand, might result in significant data loss. As a result, we'll need to devise a strategy for dealing with missing data during confidentiality. K-anonymity, is perhaps the inherent widely used approach, has difficulty anonymizing partial data streams. K-anonymity is commonly translated into a clustering issue , in which the goal is to locate a collection of clusters, i.e. Equivocation Categories (EC), each of which comprises at least k entries. The euclidean distance used to quantify tuple similarity in the clustering job demands itemsets with the same properties and correct values for each parameter. In partial streaming data, such a criterion is impossible to meet. As a result, an actual challenge in incomplete data stream anonymization is the lack of computing the similarity for partial data. Furthermore, several studies have been undertaken to include incomplete information in the obfuscation process, however this strategy results in attribute data pollution. Generalization is indeed the final step in the asymmetric encryption process. All legitimate morals of other attribute values in same intermediate node are supplemented with more general assumptions when a missing value exists, as seen in Figure 1. Incomplete data pollution that is caused in this situation, which reduces the functionality of anonymous data and may decrease the precision of successive analysis of the data gathered.

2. Research Methodology

With the development of the new era, network information technology is penetrated into various fields of industry, at the same time, the process of information with rapid development provides a new environment and opportunities for the development and progress of various industries [7]. In the process of combining network information computer technology and traditional operation mode of economic industry, compared with the traditional model of enterprise development, this kind of more innovative industrial development model has more complex industrial institutions, and the basic information contained in it is more mass, these advantages have a positive impact on the sustainable development of enterprises and have a higher demand for theory and technology of economic management [8]. the management, audit and publication of economic basic information is taken as an example, in the operation of today's enterprise, only a better grasp of the data in this series of links can timely control the actual situation of the operation of the enterprise, so that the development of enterprises has data support with higher credibility [9]. However, the strong dependence on manual processing methods in the use of data in traditional economic management of enterprise may lead to the release of false information due to some subjective factors, and the use of some bad means makes the financial scandal phenomenon continues to increase, which may cause financial losses, indirectly affect the normal operation and progress of enterprises [10]. Thus, more economists suggest that computer technology should be introduced into practical data mining of economic management to provide technical support to ensure rational mining of economic data of the entire enterprise[11].

The information size of the information stream is enormous. With the fast improvement of the distributed computing, the

systems administration, the interpersonal interaction and other new advances, the fame of the Internet permits clients to rapidly access and offer data, and the data can be shared to deliver a lot of information. The constructions of the information are assorted, and the information of the enormous information is exceptionally mind boggling. The consistent refreshing of the interpersonal organization and the expansion of the method of the client's data trades will cause the client to want to get the information. In any case, the thickness of most information is extremely low. The information design of large information is semi-organized and unstructured, yet the worth significance between unstructured information isn't high. Through the computation of data information, information mining can get valuable and likely data from it. Information mining innovation covers a wide scope of advancements, including the insights, the AI, the example acknowledgment, etc. Through the exhaustive utilization of these innovations, the capacity of dissecting and extricating helpful data can be acknowledged [12].

Information mining innovation has no prerequisite for the information design, the information stockpiling mode, etc. The info information mining innovation can be information put away in various data sets, and can likewise be the website pages on the Internet; information can be a straightforward book, and can likewise be an intricate sight and sound data framework; data set can be a progressive data set, and can likewise be a social data set [13]. Information mining innovation can take various information objects and various information mining techniques as indicated by various prerequisites, so it is more adaptable. The innovation of information mining has some exceptional qualities: the information size of the information mining measure is huge, so in the reason of a lot of information, the mining results will have the down to earth importance and general worth; the arbitrary interest of mining is huge, and it as a rule has no uncommon necessities, so the yield of information mining might be covered up and applicable data; mining data set data has the estimate capacities, and it can anticipate the mishap; it reacts rapidly to changes in the information; the exactness relies upon the measure of information, so the outcomes must be logical and the law can be valid with the examination of an adequate measure of information got after evaluation [14].

Computational ways to coping with insufficiency have a simple foundation. heuristics knowledge is utilised to decrease attribute data, and then current clustering techniques are used to the simulated data matrices . The removal of data samples with missing items is a logical way to deal with incomplete data. To put it another way, this approach creates a completely viewable new information set from an incomplete information matrices. If the asymmetric information rate is low, its grouping effectiveness may be adequate (for example, less than 10 percent). This technique is commonly utilised in the healthcare profession and is quite simple to implement. Nevertheless, they neglect incomplete data information in favour of looking for underlying patterns, which might result in a significant decrease of the initial data matrices when the absence rate is substantial. Furthermore, earlier technologies were incapable of handling learning challenges with partial sample matrices. Heuristic techniques are usually unsatisfying since it is incredibly difficult to anticipate their success for a learning job, despite how easy and simple they are to execute. Statistical approaches, unlike the preceding metaheuristic algorithms, search for even more pertinent data from incomplete information. The bulk of them uses performance parameters to restore null values and therefore do not reject flawed information. They use a constant value to fill in the blanks, result in the complete sample data that may be utilized to cognitive difficulties. In that dimension, the simplest and most generally used filler values are zero, conditional mean number, median, and modal number. Furthermore, by doing extrapolation on the entire characteristics, the items missing might be reproduced.

3. Multi-pipelining Algorithm for Data Stream Mining

Particle swarm optimization algorithm is a kind of swarm intelligence algorithm, which is based on the research of the intelligent behavior of birds and fish groups in the foraging activities. The theory comes from the theory of bionics and the evolutionary

computation [10]. By simulating the information exchange and cooperation among birds in foraging behavior, the group behavior can be optimized [11].

The position change of the particle in the search space is based on the success of the individual to follow other social psychological tendency, so the position change of particles will be affected by the experience or knowledge of other particle swarms [12]. When flying in the search space, the particle position will be affected by the individual best particle position (i.e. individual experience optimal), the best position of the swarm (optimal social experience) and the current velocity of the particle.

With the multi-pipelining process for the data stream mining based on the multi-pipelining process, we can deal with the information under the conditions of the data era. The main functions are processing and storing the data. By mastering the rules between the data, the user data can be used to optimize the analysis and calculation to enhance their abilities to deal with and spread information [13]. The key of information clustering is to compute and store the data. On the one hand, multi-pipelining process for data stream mining based on the particle swarm optimization includes the information virtualization technology, the cloud technology, the automated resource mobilization technology and the information resources protection technology. Among them, the development and application of the cloud technology has attracted many attentions [14].

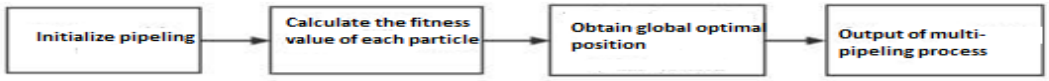

Figure 2 multi-pipelining Algorithm on the data stream mining

Computer technology is an important data analysis technology in the development of the times, which can collect the basic data of some operations in real time, make a more objective analysis and evaluation, and further provide a more accurate development plan for the sustainable development of enterprises in the future on the basis of understanding the current situation of enterprise development. As an information technology in computer data processing technology, cloud computing technology has been applied and expanded in different industries in the country and has achieved great results. The application of cloud computing technology in the process of data mining of economic management has also effectively reduced the cost of human resources, so that the production efficiency of enterprises has been greatly improved [15]. However, different from the use of many western developed countries, our country only uses the computing function of this technology, but ignores the mining of the connection of the technology when introducing the cloud computing technology in the process of data mining for economic management, this results in a large amount of waste of data resources. Therefore, cloud computing technology should be faster and better integrated into data mining of economic management in today's enterprise to reduce the uncoordinated situation of the lack of accurate data in the process of enterprise development. In addition, a large amount of data and information resources have been utilized to study the future development trend of enterprises, this has gradually become the main research topic of related industry development in society. Therefore, cloud computing technology was taken as the object of study. The application of data mining in economic management in related industries or fields was studied to determine the data mining model of economic management which was more suitable for the development of enterprise. Then, dependent on the examination of a lot of basic information assets, the hypothetical premise and specialized help were accommodated, the practical turn of events and progress of the business straightforwardly or in a roundabout way using related models.

Therefore, the advantages of the application of the technology were explained and discussed based on the understanding of the importance and related research background of cloud computing technology and data mining of economic management, the detailed methods of study as follows:

First of all, based on the query of related data and the cognition of related theories, the existing cloud computing technology platform was taken as an example, the main service performance and features of this technology platform were analyzed and summarized, and then the future direction of the platform was discussed and analyzed.

On the basis of understanding the construction process of platform, the related algorithms of data mining of economic management in cloud computing were analyzed and studied. The relevant building steps of model are as follows:

$$Gain\ ratio(Xi)=\frac{Gain(Xi)}{Split\ (Xi)} \quad (1)$$

Among them, Gainratio (Xi) indicates the gain rate of the information mining of the economic management data, Gain (Xi) and Split (Xi) are the gain and the subsequent splitting information of a certain data information, the relevant formulas are as follows (2) and (3).

$$Gain(Xi) = I(p,n) - E(Xi) \quad (2)$$

When the particle swarm optimization algorithm is used to solve the clustering problem, each particle is a feasible solution to the problem. According to the different understandings of the clustering results, the description of the solution set can have two forms: the first one is the clustering result; the second one is the clustering result. In this study, second forms are adopted, and the corresponding solution is based on the clustering center set. In other words, the solution set is composed of a plurality of cluster centers.

Secondly, we should determine the cluster partition: for each sample, we should calculate the distance from the cluster center. According to the nearest neighbor rule, we should determine the clustering of the sample, and then update the particle fitness value. According to the corresponding clustering, we should calculate the new clustering center. If we reach the end condition (to get a good clustering effect or to achieve the maximum number of iterations), then we will end the algorithm and output the global optimal solution.

Finally, we define a generalized notion of distance to handle time-series, and in particular multi-dimensional time-series. Given a pair of multi-dimensional time-series XN(z) and YN(z), where z represents the sample in Z samples, and N the dimensions, we define our generalized distance as follow:

$$d(X_N, Y_N) = \sqrt[L]{\sum_{n=0}^{N}\sum_{z=0}^{Z} |X_n(z) - Y_n(z)|^L} \quad (3)$$

where L represents the metric distance. For our implementation, we rely on L = 2, i.e., the Euclidean distance. We use the distance d(.) to find the closest centroid in the K-means algorithm. Notice that our generalized distance assumes that XN(z) and YN(z) are synchronous multi-dimensional time-series. In case the multi-dimensional time-series may present different phases but the same shape, i.e., synchronous but not aligned time-series.

Self-organizing feature map (SOFM), learning vector quantization (LVQ), and adaptive resonance theory are commonly used in neural network techniques (ART). For supervised learning, the LVQ approach is employed, whereas neural network-based clustering methods are dominated by SOFM and ART [16]. SOFM's major goal is to express high-dimensional input patterns using prototype vectors that may be observed in a two-dimensional space. SOFM has the benefit of being independent of input pattern ordering, although it, like the K-means method, requires a predetermined lattice size.

Because of the particularity of data stream, there is a big difference between the data stream and traditional static data. Stream algorithm involves the batch knowledge. The algorithm deals with P data points in each data bucket, and the bucket size is related to the memory. Firstly, the data points in the bucket are divided into several clusters and information are used to represent the centroid and the weight of each cluster. The center of mass is the center of the cluster, and the weight mainly represents the number of data points in the cluster. With the continuous inflow of data points, the number of data points will be an operation. The CluStream algorithm uses a dual layer structure- online layer and offline layer. The online layer uses the cluster to store the summary information, uses the tower time frame, and follows the latest data priority principle. The offline layer is used to deal with the micro cluster information online, and the results are fed back to the user. When a new data point arrives, DenStream first determines the feasibility of the data points received by the nearest potential core. If the merged cluster fails, it will again merge it to the nearest outlier cluster. D-Stream algorithm does not need to calculate the distance or weight, so when the amount of data increases, the efficiency of the algorithm is higher than the D-Stream algorithm. By using the density grid structure, we can deal with the data points. Clustering algorithm is summarized and analyzed as follows:

Table 1 Comparison and analysis of clustering algorithm

algorithm	Characteristic	Main disadvantage
STREAM	Partition based algorithm Clustering using centroid and weight representation	Not applicable to non-globular clusters
CluStream	Hierarchical algorithm Double layer structure Introduce the concept of micro cluster	Not applicable to non-globular clusters
DenStream	Density based algorithm Double layer structure; potential Core micro cluster and outlier clustering	When clustering a large amount of data, the clustering efficiency is low
D-Stream	Based on density and grid algorithm Using double layer structure data structure	Cannot reflect the clustering results at any time

In conclusion, most current data stream clustering algorithms[17][18][19][20] use the double layer structure: the online layer and the offline layer, and can obtain good effects, so it can be used in the double framework and it can use the grid technology to solve the cluster spherical clusters problem. Although the algorithm has done the similar treatment, most current clustering algorithms only rely on the traditional algorithm to get improvement, which is easy to make it difficult for the data stream clustering algorithm limited to expand.

In this experiment, the network data stream was considered as the research object and was divided into two kinds of data stream mining methods: the typical clustering algorithm and the clustering algorithm based on the particle swarm optimization. In light of the bunching investigation of various information streams, the information preparing

after effects of two diverse mining strategies were thought about and dissected. Among them, the preparing time, the bunch investigation and the information grouping precision were determined by the PC framework as per the real circumstance. Among them, the score of the bunch investigation and the information grouping exactness were 40 and 20 individually. The results are shown in the following table:

Table 2 Comparison of clustering results of data mining in the two models

	Item	Standard (unit)	Evaluation
Typical clustering algorithm	processing time	40	34.7
	cluster analysis	30	24.9
	Data clustering accuracy	30	23.6
	Overall evaluation	100	83.8
Clustering Algorithm based on particle swarm optimization	processing time	40	31.3
	cluster analysis	30	24.6
	Data clustering accuracy	30	22.5
	Overall evaluation	100	78.4

The comparison of the clustering results of the two patterns of data stream mining is shown below:

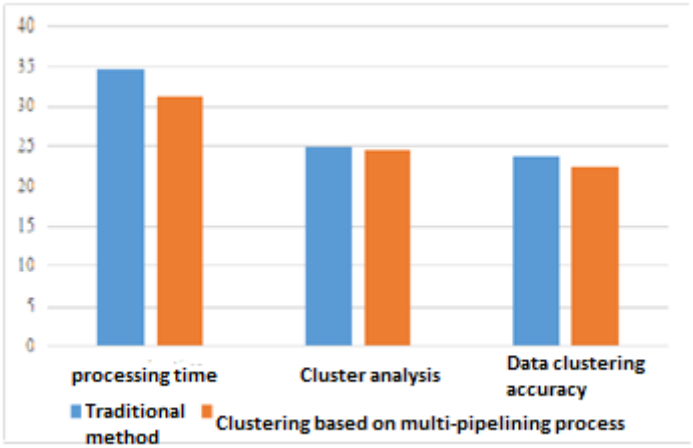

Figure 3 Comparison of clustering results

Based on Figure 2, the treatment results of the typical clustering algorithm and the clustering algorithm based on the particle swarm optimization have significant differences. It shows that the clustering algorithm based on the particle swarm optimization (PSO) brings new opportunities for the development of the data mining and data clustering. Under the new environment, the problems of the data stream, the data storage and other computer information processing are increasing day by day. The network has a large number of rich data, so by mining and clustering the analysis of these

data, we can grasp the direction of the development of the times, the needs of social groups, the information behaviour and other characteristics.

Based on the understanding the theory of cloud computer technology and its advantages in data mining of economic management, furthermore, in the current economic management, the system structure and function of cloud computing platform based on main applications of data mining were summarized. First of all, the functions of the related systems are mainly to store, process and display the relevant data information. The selection and annotation of data are implemented mainly through the logical definition layer. As can be seen from the table, the primary function of the logical definition layer is to explain the function of the economic link, can pass on the relevant data information, further analyze all the data, and make reasonable scheduling of the resource data acquired.

Table 3 Functions of the logical definition layer

Control system layer	Definition of visual mining task of data reduction	Display of results mining
Intermediate layer	The user's request for a mining task Mining task of business logic analysis Transfer resolution	Results of excavation The representation and interpretation of the results of excavation
	Unified logical form and the decomposition of subtask	Subtask after the binding
	Binding of scheduling resources of data resources	The execution of monitoringresources
	Calls to algorithm and business models	

After the logical definition layer of cloud computing's data mining technology of economic management was understood, information tables for the database of the technology needed to be designed. The function of the designed information table of database is mainly to query and collect the data information, and add annotation information that can be used for subsequent management comprehension into data information.

Table 4 Design of information table

Name of field	Description of Chinese characters	Type	Length	Description
IDBNO	Number of digital library	Int		Automatically increase 1 to ensure record unique
strDBName_Y	The name of the digital library (English)	Varchar	15	Internal use of the program
strDBName_C	The name of the digital library (Chinese)	Varchar	15	User's external display usage
strDBMemo	Description of the	Varchar	60	Some information, comments, and so on of the related business,

	function of the database			usually added by users
strDBOther				
strDBMachine	Server of database	Varchar	30	Name of storage server
strConnectionString	A contiguous string in a database	Varchar	150	
iTableNums	The number of tables in the entire database	Int		
iFieldNums	The number of fields in the database	Int		

The clustering algorithm analysis of data stream mining in the era of big data is not limited to a fixed clustering anymore. The evaluations index of the particle swarm optimization algorithm for the clustering data stream mining are divided into the following sections: data stream storage, data stream transmission, data stream mining efficiency, data stream clustering analysis, and data stream mining clustering.

Table 5 Score table of clustering algorithm based on multi-pipelining algorithm

Evaluation factors	Evaluation grade				
	1	2	3	4	5
Data stream storage	-	-	-	√	-
Data stream transmission	-	-	-	-	√
Data stream mining efficiency	-	-	-	-	√
Data stream clustering analysis	-	-	-	-	√
Data stream mining clustering	-	-	-	√	-

According to the above table, in the comprehensive evaluation results of the clustering algorithm based on the data stream mining of particle swarm optimization, the highest proportion of the index reached 60%. It is proved that the data transmission based on the particle swarm optimization in data stream mining is stable and fast, and the data stream mining efficiency is high and the data stream clustering analysis is more accurate.

4. Conclusion

With the development of the times, the basic data in the process of the development of the industry has a very important impact on the development of enterprises. Because the massive data acquisition and data processing technologies of computer technology are used, the application of cloud computing technology can be used to mine more data for economic management, and provide a reasonable development plan for the further development of the enterprise. The related theories and technologies of cloud computing technology aren't perfect enough in the development of some industries in our country. To more readily examine and dissect the connected advancements, the primary connections of the framework development were investigated

based on understanding the pertinent hypotheses. From that point onward, the development undertaking was presented to act as an illustration of the investigation. Information on the financial administration of business activities was unearthed, and the mining result gave the improvement premise to the reasonable activity of the undertaking. Albeit the information is not difficult to be gotten, it is hard to be abused. This expects us to break down the bunching calculation of information stream mining dependent on the particle swarm enhancement in countless information streams. The information investigation and handling in the information stream climate has its own improvement issues, which expect us to group and dissect data. The grouping calculation of information stream mining dependent on particle swarm streamlining was proposed in this paper, which can unbiasedly mirror this present reality. The research and analysis of the clustering algorithm based on the particle swarm optimization for the data stream mining was done in three steps. The results show that using the particle swarm optimization algorithm to detect the redundant features can optimize the clustering effect and improve the clustering purity.

References

1. Pahl C, Donnellan D. (2002), Data mining technology for the evaluation of web-based teaching and learningsystems.
2. Romero C, Ventura S, García E. (2008), Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1).368-384.
3. Kumar V, Chadha A. (2011), An empirical study of the applications of data mining techniques in higher education. International Journal of Advanced Computer Science and Applications,2(3).
4. Sun H. (2010), Research on Student Learning Result System based on Data Mining. IJCSNS, 10(4). 203.
5. Peña-Ayala A. (2014), Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications, 41(4).1432-1462.
6. West D M. (2012), Big data for education: Data mining, data analytics, and web dashboards. Governance Studies at Brookings,1-10.
7. Papamitsiou Z K, Economides A A. (2014), Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. Educational Technology & Society, 17(4).49-64.
8. Esmin A A A, Coelho R A, Matwin S. (2015), A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data. Artificial Intelligence Review, 44(1). 23-45.
9. Alam S, Dobbie G, Riddle P. (2008), Particle swarm optimization based clustering of web usage data. Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 03. IEEE Computer Society,451-454.
10. Alam S, Dobbie G, Koh Y S, et al. (2014), Research on particle swarm optimization based clustering: a systematic review of literature and techniques. Swarm and Evolutionary Computation, 17.1-13.
11. Chen D, Liu Z, Wang Z. (2012), A novel fuzzy clustering algorithm based on Kernel method and Particle Swarm Optimization. Journal of Convergence Information Technology,7(3).
12. Hatamlou A. (2013), Black hole: A new heuristic optimization approach for data clustering. Information sciences, 222. 175-184.
13. PSarkar S, Roy A, Purkayastha B S. (2013), Application of particle swarm optimization in data clustering: a survey. International Journal of Computer Applications,65(25).
14. Eusuff M M, Lansey K E. (2003), Optimization of water distribution network design using the shuffled frog leaping algorithm. Journal of Water Resources Planning and Management, 129(3). 210-225.
15. Poli R, Kennedy J, Blackwell T. (2007), Particle swarm optimization. Swarm intelligence, 1(1). 33-57.
16. Rui Xu, and Donald Wunsch II, “Survey of Clustering Algorithms,”

IEEE transactions on neural networks, vol. 16, no. 3, May 2005, pp. 645-678.
17. K. Chitra, and Dr. D.Maheswari, " A Comparative Study of Various Clustering Algorithms in Data Mining," IJCSMC, Vol. 6, Issue. 8, August 2017, pp.109 – 115.
18. Abla Chaouni Benabdellah, Asmaa Benghabrit, and Imane Bouhaddou, "A survey of clustering algorithms for an industrial context," Procedia Computer Science 148 (2019), pp. 291–302.
19. KM Archana Patel, and Prateek Thakral, "The Best Clustering Algorithms in Data Mining," International Conference on Communication and Signal Processing, April 6-8, 2016, pp. 2042-2046.
20. Peng Zhou, Fan Ye, and Liang Du, "Unsupervised Robust Multiple Kernel Learning via Extracting Local and Global Noises," IEEE Access, 1–1 Volume 4, 2016.