

# Intelligent Analysis and Processing Technology of Big Data Based on Clustering Algorithm

Zheng Zheng<sup>1</sup>, Fukai Cao<sup>1\*</sup>, Song Gao<sup>2</sup>, Amit Sharma<sup>3</sup>

<sup>1</sup>Jitang College, North China University of Science and Technology, Tangshan, 063210, China

<sup>2</sup>Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd, Tangshan, 063000, China

<sup>3</sup>Southern Federal University, Russia

Emails: zhengzheng873@163.com, fukaicao@126.com, songgao56@163.com, amit.amitsharma90@gmail.com

**Keywords:** Clustering algorithm; Big data intelligence; Smart meter; Project cost; Genetic algorithm

**Received:** February 15, 2022

*An attribute category clustering method based on hierarchical clustering is proposed in order to study the big data intelligent analysis and processing technology. The proposed model combines the attribute categories with similar fault type distribution, reduces the data dimension, and binarizes it. To address the problem of more missing values of continuous data, a data completion method based on attribute distribution function is adopted. Through the perspective of selection and estimation of project unit price in construction enterprises, this paper summarizes the data mining process facing the characteristics of project cost data, and puts forward the method of analyzing and processing project cost data based on clustering algorithm. Finally, the processed data sets are subjected to bottom-up hierarchical clustering analysis, and finally the ideal analysis results can be obtained. The experimental results show that the preprocessing method based on attribute clustering proposed in this paper can effectively merge attributes, reduce the dimension after binary transformation and effectively reduce the amount of data under the condition of ensuring data information.*

*Povzetek: S hierarhičnim gručenjem je narejena inteligentna analiza velikih podatkov.*

## 1 Introduction

The hidden value of big data promotes the derivation of big data mining technology and methods. Big data mining is to mine valuable knowledge for data processing through massive multiple data sources. Therefore, how to quickly and accurately mine valuable knowledge through big data has attracted much attention. In fact, data mining is also a decision support process. Its common methods mainly include classification, clustering, prediction, regression analysis, association rules and so on. Clustering is the most key technology. Big data is unstructured data, which is difficult and large in processing and analysis, making the structural analysis mode too complicated, and the traditional data analysis cannot effectively process, mine and analyze as shown in Figure 1 [1]. The classical methods of cluster analysis can be summarized as: partition method, hierarchical method, density-based method, grid-based method, model-based method, neural network method based on computational intelligence, evolutionary computing method, fuzzy method and so on, as well as the semi supervised clustering method which has attracted much attention at present. Recently, the new cluster integration method has rapidly become a new research hotspot of cluster analysis. The purpose of clustering integration is to fuse the results from multiple clustering algorithms to obtain higher quality and robust clustering results. The method based on graph theory is one of the fast-

developing methods recently. It is a method to realize clustering by using the principles of graph theory and graphics. Compared with traditional algorithms, this algorithm can deal with more complex cluster structures, such as nonconvex structures, and can converge to the global optimal solution [2].

In recent years, with the rapid development of network information technology, the era of big data has come and penetrated into many fields. There are more and more big data application research for specific professional fields. However, for the field of project cost, this aspect has always been a blank. Every day, with the help of the Internet and various project cost systems, a large number of project cost data are generated, but there is no scientific and accurate processing method to process it, so that it is lost in vain. The acquisition and transmission of project cost information still rely on the traditional way, and the timeliness and accuracy cannot meet the needs of today's project management field [3]. To process and mine these huge project cost information data and provide basis and reference for the decision-making of project management process, it is not enough to rely on manual processing technology. We should innovate and apply data mining technology to make full use of the value of massive project cost data, so as to promote the rapid and healthy development of the industry.

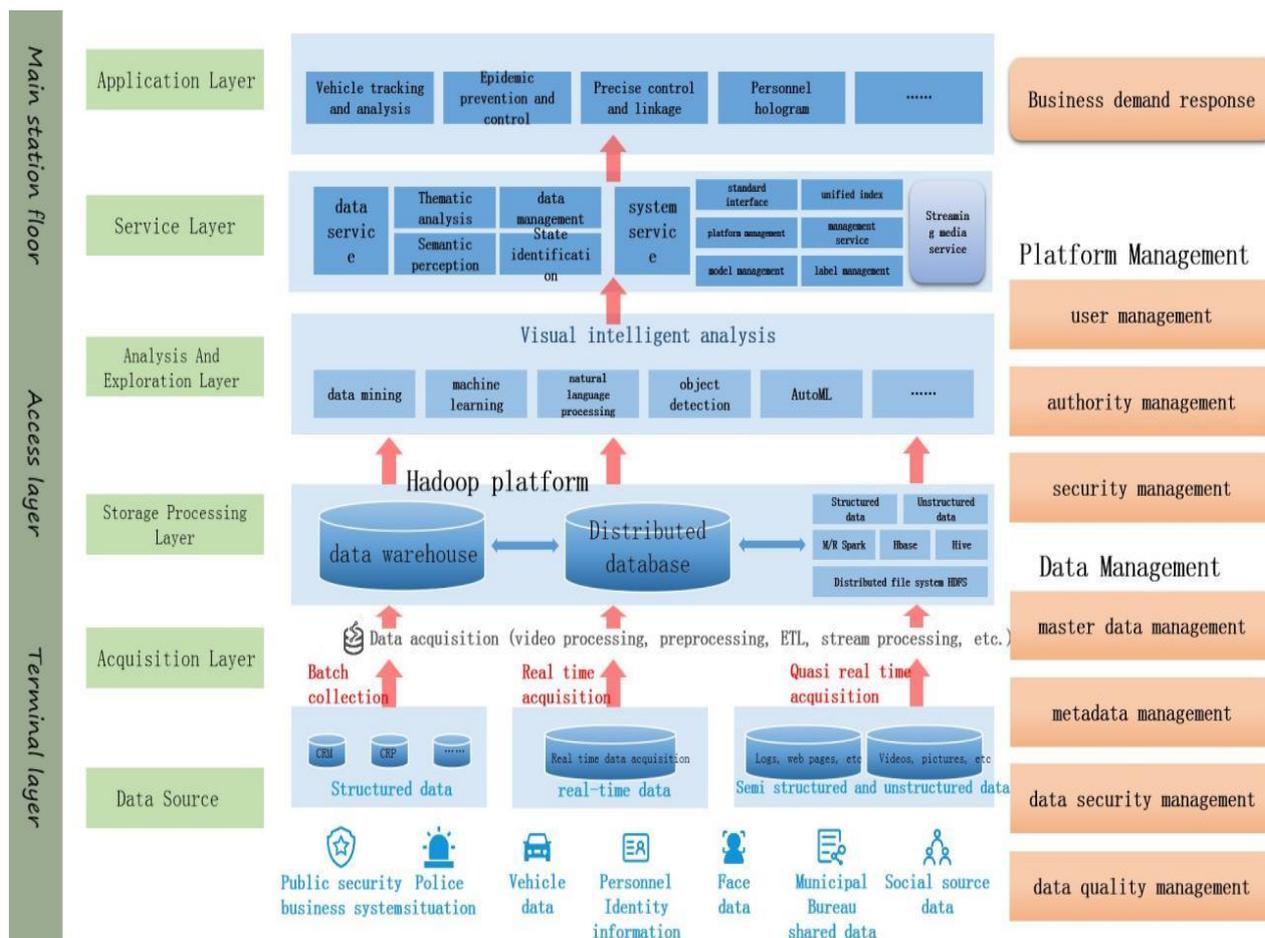


Figure 1: Big data intelligent analysis and processing technology

The rest of manuscript is organized as the most recent work done is discussed in Section 2. The research methodology, optimization of clustering algorithm, complexity and project acquisition is presented in Section 3. Results and analysis of the proposed model is discussed in Section 4 which is followed by the conclusion in Section 5.

## 2 Related work

In this section various state-of-the-art work in the field of big data processing based on clustering algorithm is presented.

Zhu *et al.* proposed an initial clustering center selection method based on point density, and processed outliers specially [4]. Ser *et al.* proposed an improved algorithm to determine the optimal cluster number  $k$  by calculating the contour coefficient of each object in the cluster under different  $K$  values, and determine the initial cluster center by hierarchical aggregation method [5]. Wu proposed a clustering method based on patent technology efficacy matrix. This method uses  $K$ -means to cluster by calculating the similarity of technology, and achieves good results.  $K$ -medoids and PAM algorithms are very effective for small data sets, but they do not have good scalability for large data sets [6]. Duan and

Wang proposed a new heuristic search algorithm clarans algorithm based on PAM [7]. The algorithm finds the center point of the representative cluster by random search of the graph. Clarans algorithm is the first clustering algorithm successfully applied in the field of spatial data mining. It overcomes the shortcomings that other classical clustering algorithms cannot deal with large-scale data sets, but it still fails to solve the problem of low execution efficiency. Its time complexity is  $2O(KN)$ . In order to speed up the execution speed of the algorithm, the parallel clarans algorithm based on PVM mechanism proposed by Xing and Li effectively improves the speed of the algorithm [8]. In the artificial neural network, Cai applied the classical hierarchical clustering algorithm and partition algorithm to cluster SOM, which aims to reduce the computational complexity of the classical clustering method [9]. In addition, in terms of network application: Xu *et al.* proposed a three-dimensional facial expression clustering method based on network, which overcomes the shortcomings of limited information contained in data and sharp decline in recognition performance in the case of two-dimensional facial expression recognition [10]. In terms of project cost, Li *et al.* others established the power grid cost management method system and the construction framework of cost analysis information platform under the big data environment [11]. Shi and

Zhu designed the cost management system of mine engineering construction project based on cost data [12]. Wendong *et al.* put forward the statistics and analysis method of project cost information data under the background of big data, and constructed the statistical calculation model of project cost information data [13]. The evolution of artificial intelligence and Internet of Things is considered for several industrial applications and contributing towards social life [14-17].

### 3 Research methods

This section includes the project design process, structural seismic analysis and detailed modeling steps of proposed design.

As unstructured data, big data is difficult to be characterized by two-dimensional logic table of database. The multi-dimensional de clustering analysis algorithm shows the hidden structure of observation variables through the Bayesian network model structure, and constructs the logical correlation between leaf nodes and other nodes. In this model, multiple hidden variables are allowed to exist, corresponding to the corresponding data clustering methods. Based on the probability dependence between random variables, the multi-dimensional de reunion class analysis algorithm analyzes unstructured data, and quantitatively describes the reasonable distribution with the conditional concept as the carrier. The specific flow of data processing is as follows:

Data preprocessing, that is, data cleaning, avoiding noise and solving the problem of data loss. During data processing, discrete the continuous values in the attribute and convert the data. The data result set and test training set are studied, and the data set is divided into two parts: data result set and test training set. The classifier is constructed by classification algorithm. Through the test set, the accuracy evaluation mode is selected to evaluate the classifier. The classifier that meets the accuracy standard is applied in practice, otherwise it will be modified.

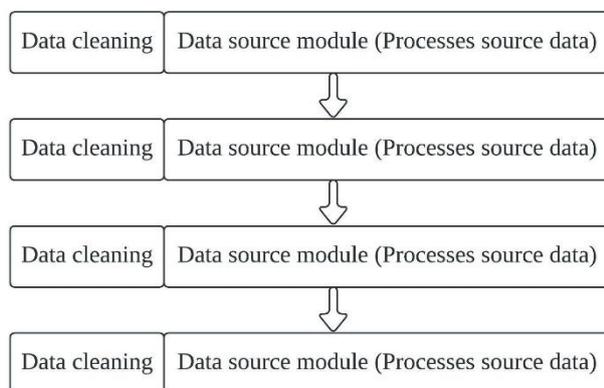


Figure 2: Data processing flow of cluster analysis model

Word segmentation and document vectorization processing, reorganize the continuous word sequence according to the established norms to form the word sequence. In order to transform the document after word

segmentation into a pattern that can be recognized and processed by computer, it is necessary to quantify the word features as the feature vector, which is currently processed by vector space model. Feature selection and multi-dimensional cluster analysis, word features will lead to a certain sparsity and high dimensionality in the document vector feature space, so an effective feature selection method is selected to reduce the dimension of the feature space and further improve the classification efficiency and accuracy [18]. The detailed data processing steps of the analysis model are shown in Figure 2.

Clearly build a functional model for the classification process of big data and unstructured data. The problem can be described as a given data set and category set which is evaluated using Equation 1 and 2.

$$F = \{F_1, F_2, F_3, F_4, \dots, F_m\} \tag{1}$$

$$G = \{G_1, G_2, G_3, \dots, G_m\} \tag{2}$$

The classification problem is to clarify the function mapping to make the data items of the data set map to the corresponding categories. Given the big data variable set, the variable takes the parent node set as the carrier, the carrier correlation between nodes can be characterized by a directed graph, that is, for each variable, it can be characterized as a node, and each node guides a directed edge from each directed node of the parent node set to enter the variable. Suppose that the variables of Bayesian network are a and b respectively, and X is the node set without a and b. once z separates a and b, the conditions remain independent based on a given z. The so-called isolation and conditional independence show the close relationship between the graph theory side and probability theory side of Bayesian network. Set to classify objects based on the evidence provided by the feature vector, then:

$$e\left(\frac{vj}{x}\right) > e\left(\frac{vi}{x}\right) i \neq j \tag{3}$$

$$e\left(\frac{x}{vj}\right)e(vj) > e\left(\frac{x}{vi}\right)e(vi) \tag{4}$$

Decision rules are likelihood test rules which are evaluated using Equation 3 and 4. Bayesian network reasoning, through probability decomposition, reduces the reasoning complexity to localize the operation. Through the edge processing and analysis of the elimination process, the decision rules can be tested by likelihood rate for all given large data sets to obtain the minimum error probability calculation samples [19].

### 3.1 Optimization of clustering algorithm

Based on the function model, an optimized clustering algorithm is constructed to divide the overall big data into multiple data intervals, which are stored through multiple files, and each file represents the corresponding interval. After scanning and comparing all the data, divide them into multiple sections, and sort and remove multiple files. The data quantity of each file is  $1M$  and  $2M$  respectively. After the data is de duplicated, cluster analysis is carried out and Bayesian formula is used to calculate, that is calculated in Equation 5.

$$BIC\left(\frac{N}{F}\right) = \text{Max} \alpha \log e\left(\frac{N}{F}, \alpha\right) - f(n) \log \frac{M}{2} \quad (5)$$

$\text{Max} \alpha \log e\left(\frac{N}{F}, \alpha\right)$  represents the effect of data and model integration;  $f(n) \log$  represents that when the data is closely integrated with the model, it should be taken as the negative amount of difference, while when combined with sparse, it should be taken as the compensation amount. Based on the specific specification of Bayesian formula and the organic combination of model and data, on the basis of meeting the clustering characteristics, it is necessary to calculate and analyze the model through multi-dimensional clustering algorithm. The input of this algorithm contains  $m$  objects. The objects in the same cluster have high similarity, on the contrary, the similarity is small. The algorithm description process is shown in Figure 3.

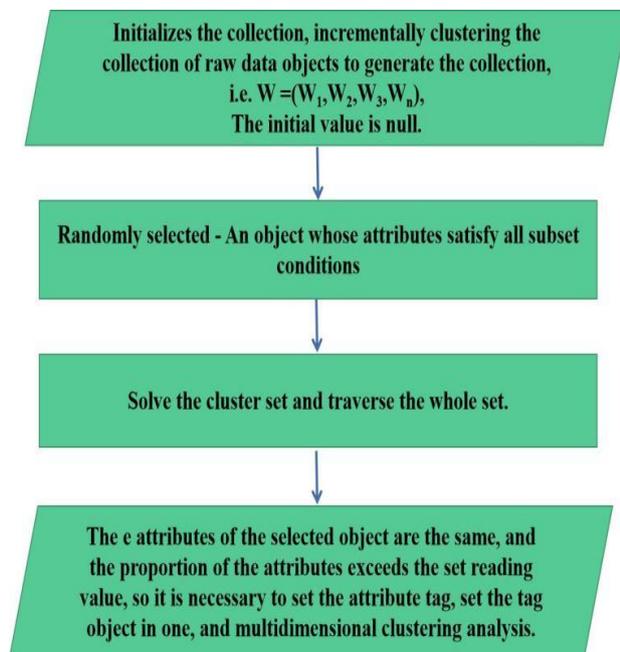


Figure 3: Algorithm description process

### 3.2 Complexity

The space cost generated by the new algorithm needs to fully consider the characteristic samples of big data. If hierarchical clustering is used to optimize the

clustering algorithm, all clusters to be clustered need to be reasonably set according to the serial mode, the total clustering time ( $R$ ) and the cost ( $n$ ). Then the space complexity ( $W$ ) is expressed in Equation 6.

$$W = n + R(m^2 \log m) \quad (6)$$

In terms of optimization rules, when the model and data fusion are sparse, set  $x$  and  $y$  as the dimension of the data set. When dividing attributes, only scan the data set at one time, in which  $z$  identifies clustering data, and the results will not be affected by factors such as multidimensional space and input order [20, 21]. Then the multi-dimensional spatial clustering can be found in time through the evaluation of weight and threshold, and the amount of calculation can be simplified. The total clustering time  $u \times n$  is divided by the linear arrangement of the consumption time ( $n$ ) and the de duplication time ( $m$ ); Total weight removal time  $u \times n$ ; Time complexity  $u \times m^2 \log m$ , then the total time complexity of the algorithm is calculated by Equation 7.

$$R(m) = u \times m + u \times m + u \times m^2 \log m \quad (7)$$

### 3.3 Acquisition of project cost data

There are two ways to obtain project cost data based on big data.

- i. There are generally two methods of internal collection in the platform. First method is to build a unified project cost information data collection template, collect and import the relevant data in the platform according to the user-defined unified specifications, so as to directly convert the target cost data information and store it in the local database for backup. The second method is to set up fields conforming to certain specifications on the relevant cost information platform, collect the information of the same field and store it in the local database [22].
- ii. The specific methods and principles are as follows: create a unified data exchange format through the corresponding platform interface, and realize the information exchange of relevant businesses inside and outside the platform.

According to the collection method and the form of price change trend, we generally use the box method to process the project cost data studied in this paper. Before processing, we must first solve the problem of detection. For the detection of noise data, the change of cost data is mainly based on the overall change of market economy [23]. From the perspective of time series, it changes continuously, and is largely affected by the overall economic development. Generally, there will be no major

fluctuations and changes. We set the percentage of the annual change threshold range of cost data to 19%. Within the sampling range, the data points exceeding 20% of the average value are regarded as noise, the regression curve is calculated, and its value is re solved and corrected [24-26]. Handling method of inconsistent data format: to deal with the problem of inconsistent data format, the common method is to establish a general data acquisition template and collect according to the general data template to ensure the consistency of data acquisition format. According to the requirements and characteristics of data analysis in this paper, the data acquisition template is established, as shown in Table 1 and Table 2.

Number	Region	Unit	Time	Source
1	Jiangsu	yuan	January	Data survey
2	Shanghai	yuan	February	Data survey
3	Beijing	yuan	March	Data survey

Table 1: Data collection template - labor unit price expense template

Listing	Type	Accuracy	Format	Explain
Region	text	--	--	Area code
Number	double	1	XXX	Sample number
Company	text	--	--	Collection unit
Unit Price	double	0.02	XX.XX	Unit Price
Single time	Date	s	--	Acquisition time
Source	Date	...	--	Data sources

Table 2: Template description - labor unit price expense template

As the material cost accounts for a large proportion of the project cost, usually about 0% ~ 70%, the material price has a great impact on the specific final settlement results and decisions [27]. Therefore, this paper selects the material price as the research object, and focuses on the specific application of material price data in the fields of relevant project cost index prediction, project price

information analysis and investment estimation. Due to the dynamic, massive, multi-source and heterogeneous characteristics of project cost big data, we choose K-means clustering algorithm for specific solution [28].

### 4 Results and Analysis

This section illustrates the analysis of results obtained by comparing the seismic forces and finally presents its discussion and summary. In this proposed model, cluster analyze the quotations of 20 local suppliers for composite Portland cement.

Number	Region	Specifications	Unit Price	Source
1	SSX	PC32.1	452	Merchant A
2	SSX	PC32.1	326	Merchant A
3	SSX	PC32.1	419	Merchant A
4	SSX	PC32.1	385	Merchant A
5	SSX	PC32.1	453	Merchant A
6	SSX	PC32.1	376	Merchant A
7	SSX	PC32.1	413	Merchant A
8	SSX	PC32.1	306	Merchant A
9	SSX	PC32.1	378	Merchant A
10	SSX	PC32.1	403	Merchant A
11	SSX	PC32.1	487	Merchant A
...	SSX	PC32.1	...	Merchant A
20	SSX	PC32.1	346	Merchant A

Table 3: Data acquisition results

The 20 data listed in Table 3 are combined according to the price and serial number to obtain the initial data set  $A$ ,  $A$  is  $\{x_1, x_2, x_3, \dots, x_{20}\}$ . Before calculation, it should be noted that K-Means algorithm must give K value before solution, which directly

determines the accuracy and efficiency of the algorithm. This paper determines the  $K$  value according to the following methods: firstly, compare the distance between each sample in the sample data set, select the point furthest from other points as the initial center point of the calculation according to the calculation results, and then determine the value of  $K$  through the newly generated classification [29, 30].

- i. Select the two data with the smallest distance in the data sequence. In this example, the distance between the two points  $x_9$  and  $x_{12}$  is the largest. Take these two points as the center of the cluster for cluster calculation to obtain two cluster sets.

They are:  $S_{21} = \{x_9, x_2, x_4, x_8, x_{10}, x_{13}, x_{14}, x_{18}\}$  and  $S_{22} = \{x_{12}, x_3, x_5, x_7, x_{11}, x_{12}, x_{16}, x_{17}\}$ .

- ii. Combined with the above clustering calculation results, for the two clustering sets, first solve the first type of data and cluster center  $x_9$  respectively, for example, to obtain the farthest distance of 83, the second type of data and cluster center  $x_{12}$  respectively, with the maximum distance of 85, and then select the point  $x_{11}$  with the maximum distance as the third cluster point.

- iii. Recalculate, select  $x_9, x_{12}$  and  $x_{11}$  as three cluster centers, and calculate three cluster sets as follows:

$$S_{31} = \{x_9, x_2, x_{10}, x_{20}\}, S_{32} = \{x_{12}, x_1, x_5\} \text{ and } S_{33} = \{x_{11}, x_3, x_4, x_6, x_7, x_8, x_{10}, x_{13}, x_{14}, x_{15}\}$$

- iv. Calculate the distance between the data elements in the three set classes and each cluster center, continue the cluster analysis, and then obtain four cluster sets [31].
- v. Based on the above calculation results, the cluster numbers of different cluster centers are listed, as shown in Table 4.

Serial number	Center point	Numerical value	Number of clusters
1	$x_9$	315	5
2	$x_{11}$	406	4
3	$x_{12}$	475	4
4	$x_{18}$	413	9

Table 4: Cluster analysis results

According to the analysis of the results of the clustering algorithm in Table 4 and Figure 4, point  $x_{18}$  is

the center with the largest number of clustering samples in all clustering centers, so it can better reflect the real price of the market compared with other centers [32]. Taking this as an example, in the practical application of project cost budget and final accounts, we can analyze the market price of materials through the data mining algorithm proposed in this paper. By analyzing the solution results, we can assist relevant personnel to accurately grasp the market price information and help auditors judge the authenticity of price information in time.

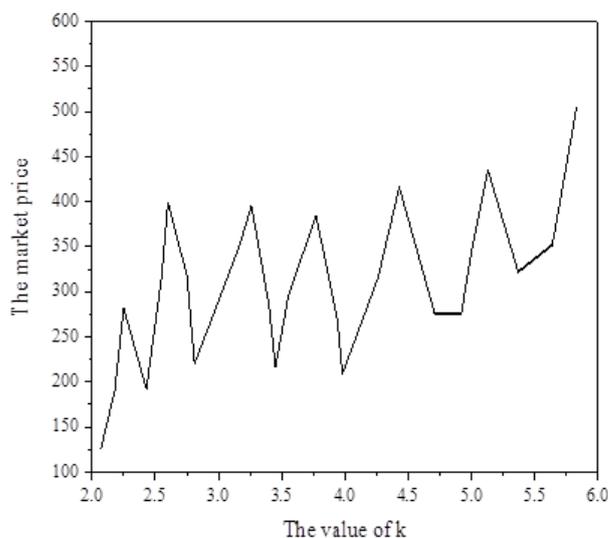


Figure 4: Results of clustering algorithm

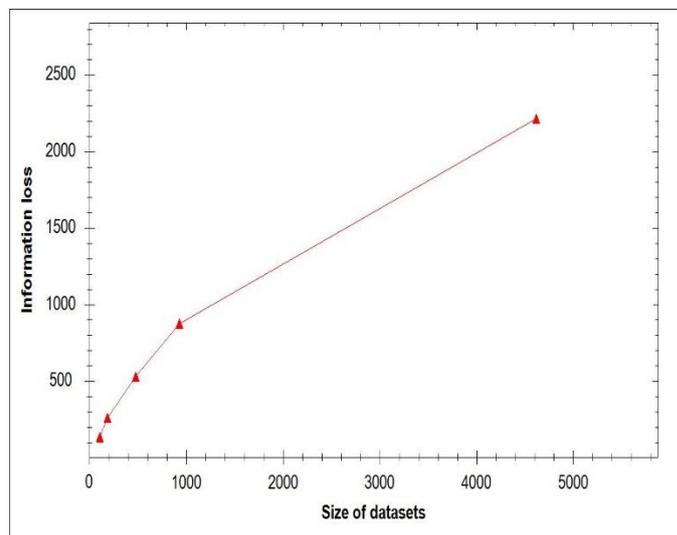


Figure 5(a): Result for different size of datasets for information loss

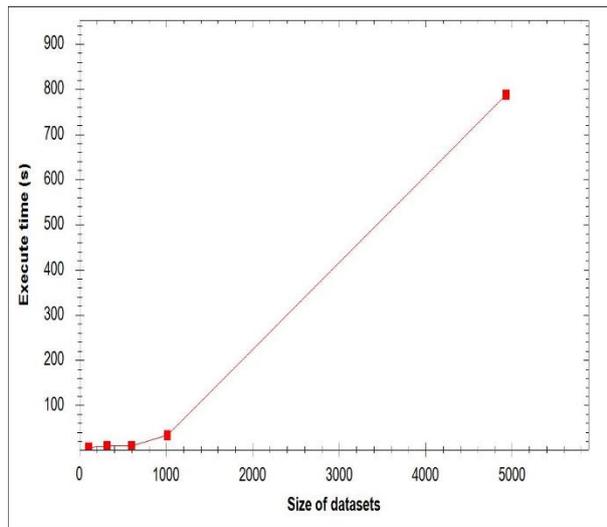


Figure 5(b): Result for different size of datasets for execute time

Different number of records are separated from the grown-up dataset and assess the exhibition of further developed anonymity model on various size of datasets, as depicted in Figure 5 (a and b). As shown in this figure, execute time increment and information loss with the increasing size of datasets. Execute time rises quickly, yet the incensement of information loss reportedly slows progressively. Clearly, the rising size of datasets genuinely affects execution time on the grounds that the grouping system of finding proportionality classes is perplexing and time consuming.

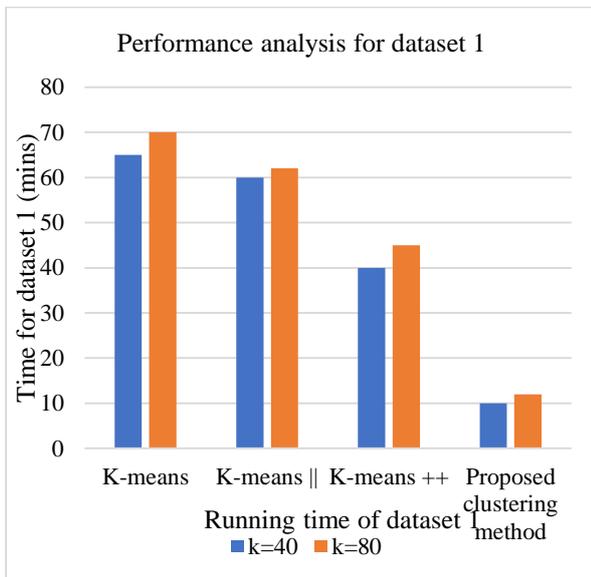


Figure 6: Performance comparison of time measured for dataset 1

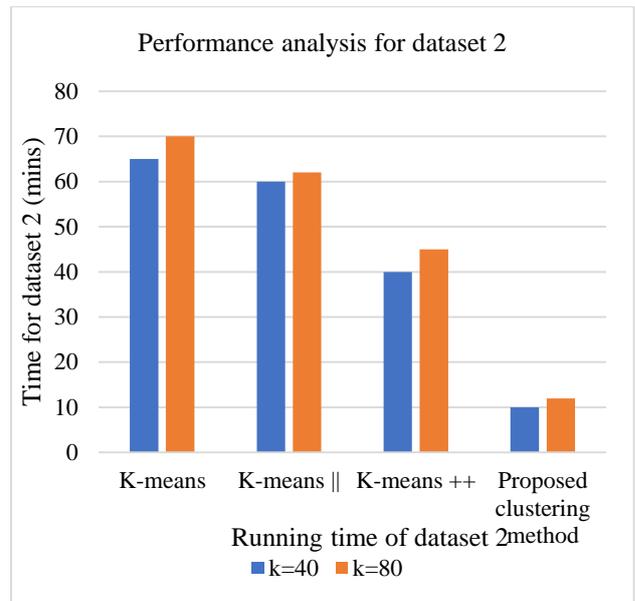


Figure 7: Performance comparison of time measured for dataset 2

The performance of the proposed clustering scheme is measured on two different datasets, dataset 1 i.e., BoW (Bag of words) dataset and dataset 2 i.e., HOUSE (household electric power consumption) dataset. To analyze the performance of clustering cost of proposed algorithm we have compared it with existing baseline models. The value of  $k$  is considered as 40 and 80 for BoW and HOUSE datasets. Figure 6 and 7 illustrates the experimental analysis of HOUSE and Bag of words (BoW) datasets and the total running time of proposed model is observed. It is observed from the experimentation that the proposed model achieves higher performance in comparison with K-means ++, K-means and K-means || when implanted to execute in parallel.

## 5 Conclusions

Different data analysis and mining methods are required for different purposes of project cost data mining under the background of big data. From the perspective of the selection and estimation of engineering unit price in construction enterprises, this paper summarizes the data mining process facing the characteristics of engineering cost data, and puts forward the method of analyzing and processing engineering cost data based on clustering algorithm. The proposed model provides a meaningful exploration for the research of massive engineering cost data mining. From the experimentation it is analyzed that the proposed clustering model achieves better time measurement when compared with existing baseline models. The clustering models based on computational intelligence are proposed. However, these intelligent technologies are not organically integrated. Machine learning and data mining technology have made great breakthroughs in today's academic and industrial circles. Therefore, how to integrate various intelligent technologies to give full play

to the functional characteristics of this kind of algorithm applied to cluster analysis is also one of the future research directions.

## References

- [1] Li, W., & Huang, Q. (2017). Research on intelligent avoidance method of shipwreck based on bigdata analysis. *Polish Maritime Research*, 10.1515/pomr-2017-0125
- [2] Li, L., Wang, J., & Li, X. (2020). Efficiency analysis of machine learning intelligent investment based on K-means algorithm. *Ieee Access*, 8, 147463-147470. 10.1109/ACCESS.2020.3011366
- [3] Dong-rui, L. (2017). Cluster analysis algorithm based on key data integration for cloud computing. *International Journal of Reasoning-based Intelligent Systems*, 9(3-4), 123-129. 10.1504/IJRIS.2017.090041
- [4] Zhu, K., Joshi, S., Wang, Q. G., & Hsi, J. F. Y. (2019). Guest editorial special section on big data analytics in intelligent manufacturing. *IEEE Transactions on Industrial Informatics*, 15(4), 2382-2385. 10.1109/TII.2019.2900726
- [5] Del Ser, J., Sanchez-Medina, J. J., & Vlahogianni, E. I. (2019). Introduction to the special issue on online learning for big-data driven transportation and mobility. *IEEE Transactions on Intelligent Transportation Systems*, 20(12), 4621-4623. 10.1109/TITS.2019.2955548
- [6] Wu, C. (2019, June). Research on Clustering Algorithm Based on Big Data Background. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022131). IOP Publishing. 10.1088/1742-6596/1237/2/022131
- [7] Duan, S., & Wang, Z. (2021). Research on the service mode of the university library based on data mining. *Scientific Programming*, 2021. <https://doi.org/10.1155/2021/5564326>
- [8] Xing, Z., & Li, G. (2019). Intelligent classification method of remote sensing image based on big data in spark environment. *International Journal of Wireless Information Networks*, 26(3), 183-192. <https://doi.org/10.1007/s10776-019-00440-z>
- [9] Cai, Z. M. (2020). Network community partition based on intelligent clustering algorithm. *Компьютерная оптика*, 44(6), 985-989. 10.18287/2412-6179-CO-724
- [10] Xu, Z., Shi, D., & Tu, Z. (2021). Research on diagnostic information of smart medical care based on big data. *Journal of Healthcare Engineering*, 2021. <https://doi.org/10.1155/2021/9977358>
- [11] Li, W., Luo, Y., Tang, C., Zhang, K., & Ma, X. (2021). Boosted Fuzzy Granular Regression Trees. *Mathematical Problems in Engineering*, 2021. <https://doi.org/10.1155/2021/9958427>
- [12] Shi, F., & Zhu, L. (2019). Analysis of trip generation rates in residential commuting based on mobile phone signaling data. *Journal of Transport and Land Use*, 12(1), 201-220. <http://dx.doi.org/10.5198/jtlu.2019.1431>
- [13] Wendong, X., Yuanfeng, L., & Deli, C. (2017). Algorithm of key data ensemble clustering and approximate analysis in cloud computing. *International Journal of Reasoning-based Intelligent Systems*, 9(3-4), 177-184. 10.1504/IJRIS.2017.090038
- [14] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, 100, 107912. <https://doi.org/10.1016/j.compeleceng.2022.107912>
- [15] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892. <https://doi.org/10.1111/exsy.12892>
- [16] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826. <https://doi.org/10.1002/dac.4826>
- [17] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, 61, 102332. <https://doi.org/10.1016/j.scs.2020.102332>
- [18] Tseng, F. H., Cho, H. H., & Wu, H. T. (2019). Applying big data for intelligent agriculture-based crop selection analysis. *IEEE Access*, 7, 116965-116974. 10.1109/ACCESS.2019.2935564
- [19] Zhao, Y., Ding, F., Li, J., Guo, L., & Qi, W. (2019). The intelligent obstacle sensing and recognizing method based on D-S evidence theory for UGV. *Future Generation Computer Systems*, 97, 21-29. <https://doi.org/10.1016/j.future.2019.02.003>
- [20] Yuan, W., Deng, P., Taleb, T., Wan, J., & Bi, C. (2015). An unlicensed taxi identification model based on big data analysis. *IEEE Transactions on Intelligent Transportation Systems*, 17(6), 1703-1713. 10.1109/TITS.2015.2498180

- [21] Wang, L. (2021, December). Intelligent analysis of accounting information processing under the background of big data. In *2021 2nd International Conference on Big Data Economy and Information Management (BDEIM)* (pp. 461-464). IEEE. [10.1109/BDEIM55082.2021.00100](https://doi.org/10.1109/BDEIM55082.2021.00100)
- [22] Ma, X., Wang, Z., Zhou, S., Wen, H., & Zhang, Y. (2018, June). Intelligent healthcare systems assisted by data analytics and mobile computing. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 1317-1322). IEEE. [10.1109/IWCMC.2018.8450377](https://doi.org/10.1109/IWCMC.2018.8450377)
- [23] Hu, H., Tang, B., Gong, X., Wei, W., & Wang, H. (2017). Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks. *IEEE Transactions on Industrial Informatics*, 13(4), 2106-2116. [10.1109/TII.2017.2683528](https://doi.org/10.1109/TII.2017.2683528)
- [24] Vedavathi, N., Dharmamah, Ghuram, Venkatadri, Kothuru and Gaffar, Shaik Abdul. Numerical study of radiative non-Darcy nanofluid flow over a stretching sheet with a convective Nield conditions and energy activation. *Nonlinear Engineering*, 10(1), 159-176, 2021. <https://doi.org/10.1515/nleng-2021-0012>
- [25] Hayat, Tasawar, Ullah, Inayat, Muhammad, Khurshed and Alsaedi, Ahmed. Gyrotactic microorganism and bio-convection during flow of Prandtl-Eyring nanomaterial. *Nonlinear Engineering*, 10(1), 201-212, 2021. <https://doi.org/10.1515/nleng-2021-0015>
- [26] Li, Zhenfang, Gao, Dong, Wu, Chuanji, Lv, Guoqing, Liu, Xin, Zhai, Haoran and Huang, Zhanfang. Mechanical performance of aerated concrete and its bonding performance with glass fiber grille. *Nonlinear Engineering*, 10(1), 240-244, 2021. <https://doi.org/10.1515/nleng-2021-0018>
- [27] Liang, H., Yun, C., Kan, M. J., & Gao, J. (2019). Research and application of element logging intelligent identification model based on data mining. *IEEE Access*, 7, 94415-94423. [10.1109/ACCESS.2019.2928001](https://doi.org/10.1109/ACCESS.2019.2928001)
- [28] He, Z., He, Y., Liu, F., & Zhao, Y. (2019). Big data-oriented product infant failure intelligent root cause identification using associated tree and fuzzy DEA. *IEEE Access*, 7, 34687-34698. [10.1109/ACCESS.2019.2904759](https://doi.org/10.1109/ACCESS.2019.2904759)
- [29] He, X., Wang, K., Lu, H., Xu, W., & Guo, S. (2020). Edge qoe: Intelligent big data caching via deep reinforcement learning. *IEEE Network*, 34(4), 8-13. [10.1109/MNET.011.1900393](https://doi.org/10.1109/MNET.011.1900393)
- [30] Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5), 3137-3147. [10.1109/TIE.2016.2519325](https://doi.org/10.1109/TIE.2016.2519325)
- [31] Srivani, B., Sandhya, N., & Padmaja Rani, B. (2020). Literature review and analysis on big data stream classification techniques. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 24(3), 205-215. [10.3233/KES-200042](https://doi.org/10.3233/KES-200042)
- [32] Liu, X., Sun, Q., Lu, W., Wu, C., & Ding, H. (2020). Big-data-based intelligent spectrum sensing for heterogeneous spectrum communications in 5G. *IEEE Wireless Communications*, 27(5), 67-73. [10.1109/MWC.001.1900493](https://doi.org/10.1109/MWC.001.1900493)

