
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
 Volume 25, Issues 1, pp. 137–145, DOI 10.12694/scpe.v25i1.2281

ANALYSIS AND APPLICATION OF BIG DATA FEATURE EXTRACTION BASED ON IMPROVED K-MEANS ALGORITHM

WENJUAN YANG*

Abstract. This paper addresses the challenges modelled by collecting and storing large volumes of big data, focusing on mitigating data errors. The primary goal is to propose and evaluate an enhanced K-means algorithm for big data applications. This research also aims to design an extensive energy data system to demonstrate the improved algorithm's practical utility in monitoring power equipment. The research begins with an in-depth analysis of the traditional K-means algorithm, culminating in the proposal of an improved version. Subsequently, the study outlines developing a comprehensive, extensive energy data system, encompassing architectural aspects such as data storage, mechanical layers, and data access structures. The research also involves the development of a power big data analysis platform, incorporating the improved algorithm for clustering and analyzing power equipment monitoring data. Experimental results reveal that the proposed improved K-means algorithm outperforms the traditional version, with significantly improved accuracy and reduced classification errors, achieving an error rate of less than one. The improved K-means algorithm showcased remarkable enhancements, achieving a meagre misclassification rate of just 0.08% while substantially boosting accuracy levels, consistently exceeding 95% across all datasets. Moreover, the power big data system developed in this study to meet practical requirements while enhancing storage and processing efficiency effectively.

Key words: Big data, K-means algorithm, Data errors, Power equipment monitoring, Data analysis platform, Pollution detection

1. Introduction. In today's digital age, the proliferation of data generation points, primarily through data collection terminals like sensors, has led to an unprecedented surge in the sheer volume of data. Consider the well-known online giants such as Facebook, Google, Yahoo, and Baidu to put this exponential growth into perspective. These titans of the internet realm grapple with the monumental task of processing hundreds of petabytes of data each day. Likewise, global retail giants, including Wal-Mart, Carrefour, and TESCO Group, must efficiently handle millions of user requests every hour. In particle physics, exemplified by the Large Hadron Collider (LHC) since 2008, annual data production has consistently exceeded 25 petabytes. This explosion of big data presents a dual opportunity and challenge. On the one hand, it furnishes humanity with a prosperous source of information, empowering us to comprehend and exert control over the physical world to an unprecedented extent. On the other hand, this overflow of data places ever-mounting demands on server systems' processing power and efficiency [11].

Once the torrents of big data inundate the servers, a crucial step involves streamlining their processing. To optimize the efficiency of handling vast datasets, a rational approach involves categorizing big data into meaningful groups, enabling the provisioning of similar data to processing terminals with analogous functions. This systematic classification and allocation enhance the overall efficacy of data processing, ensuring that the colossal influx of information can be harnessed and transformed into actionable insights with remarkable efficiency [16].

The K-means algorithm, a cornerstone in the field of clustering techniques, was initially conceptualized by MacQueen. This classic algorithm is celebrated for its simplicity, computational efficiency, and remarkable clustering capabilities. At its core, K-means assigns each data point to the cluster whose centroid is closest in terms of Euclidean distance. However, K-means bears a notable limitation—its categorical rigidity. It operates on a strict partitioning principle, obliging every data object to be unequivocally assigned to a single cluster. The quality of its clustering outcomes hinges heavily upon the initial placement of cluster centres and the predetermined number of clusters [4].

*Shanghai Zhongqiao Vocational and Technical University, Shanghai, 201514, China (wenjuanyang5@163.com).

The Fuzzy C-means (FCM) algorithm is a more nuanced and adaptable alternative to address some of these limitations inherent in the K-means algorithm. Building upon the foundations of K-means, FCM introduces the concept of fuzzy membership. In stark contrast to K-means' rigid assignments, FCM liberates each element from the confines of strict cluster boundaries. Instead, it allows data points to exhibit degrees of membership or confidence in multiple clusters simultaneously. This intrinsic flexibility enables FCM to capture real-world data distributions' nuanced, overlapping nature. In the FCM algorithm, each data point is not restricted to a single cluster but instead conveys its affinity or level of confidence for each cluster. It is achieved by assigning continuous numbers between 0 and 1 membership values. These membership values indicate how much a data point belongs to each cluster, reflecting the inherent uncertainty or fuzziness in many practical scenarios [5].

By introducing fuzzy memberships, the FCM algorithm accommodates datasets with intricate patterns and substantial overlap and provides a more granular and nuanced representation of data relationships. This adaptability and finesse make FCM a powerful extension of the K-means algorithm, particularly well-suited for applications where data points may exhibit varying degrees of association with multiple clusters, as is often the case in complex real-world datasets [18].

The paper is organized as follows: Section 2 presents a thorough literature review, critically assessing prior work in big data and the K means algorithm. Section 3 outlines the proposed method, exploring the details of the improved K-means algorithm for feature extraction from big data. Section 4 comprehensively presents results obtained through experiments and engages in a robust discussion of these findings. Finally, Section 5 concludes the paper by summarizing key insights highlighting the contributions in big data feature extraction and analysis.

2. Literature Review. The comprehensive implementation and modernization of electricity collection systems have ushered in an era where traditional manual on-site meter readings are largely past. While this transition has undeniably boosted meter reading efficiency and dramatically curtailed labour costs, it has also brought about an unintended consequence - a reduced frequency of direct interactions between power supply authorities and consumers. Consequently, this diminished engagement has created a potential blind spot in promptly and accurately ascertaining users' actual electricity consumption behaviours, rendering them susceptible to electricity theft. Electricity theft, a clandestine practice with various modus operandi, has the unfortunate consequence of distorting real-time electricity usage data. Fortunately, the immense volume of data on residents' electricity consumption is harnessed in the age of fully integrated electricity collection systems. Leveraging advanced artificial intelligence algorithms, this wealth of big data is meticulously analyzed, thereby facilitating the effective identification of irregular electricity usage patterns among consumers who deviate from the norm [8].

In a related domain, K-means clustering and Haar wavelet transform underpin a novel optimal heart sound segmentation algorithm [17]. This innovative algorithm comprises three integral components, each contributing to a more precise and refined heart sound segmentation process. Concurrently, an advanced Orthogonal Matching Pursuit (OMP) technology significantly enhances existing methodologies. Building on this foundation, Prabhakar has replaced the K-SVD technique with K-means clustering and the Method of Optimal Direction (MOD) technology, yielding six distinctive combinations in sparse representation optimization [13].

Meanwhile, the applications of the GB-BP neural network algorithm are explored in wrestling. This research resulted in the development of a sports athlete action recognition and classification model based on the GB-BP neural network algorithm. Wang's work commenced with a comprehensive analysis of the current state of wrestling action recognition, subsequently addressing and enhancing the limitations of existing action recognition and big data analysis techniques in the domain. Through these diverse endeavours, innovative solutions and algorithmic advancements are emerging to tackle complex problems across a spectrum of domains, driven by the growing availability and utilization of big data [15].

In recent years, there has been a notable surge in research focused on big data, underscored by its profound significance in shaping the design and implementation of cutting-edge solutions across diverse applications. This surge is particularly pertinent when addressing big data's current status and challenges in various domains. In alignment with this overarching trend, the author of this study has embarked on an ambitious endeavour. The core objective of this research is to develop a robust and versatile big energy data analysis platform meticulously tailored to address the specific and evolving needs of the big energy data landscape [14].

At the heart of this initiative lies the aspiration to empower real-world analysis of big energy data, uncovering valuable insights and patterns that might otherwise remain concealed within the vast data reservoirs. By harnessing the capabilities of this platform, stakeholders can effectively pinpoint and identify crucial messages and information pertinent to energy equipment pollution. This, in turn, is a pivotal step in the larger mission to ascertain the presence and extent of equipment pollution.

Crucially, the platform leverages advanced data analysis techniques, including but not limited to the utilization of cutting-edge K-tools. These tools are instrumental in systematically collecting and analyzing data about energy equipment pollution. The platform can discern intricate patterns and anomalies within the data by employing K-means clustering and related methodologies, thereby facilitating accurate determination of equipment pollution [21].

Furthermore, the insights from this comprehensive analysis pave the way for the platform to offer tailored and customized advice. This advice is indispensable in ensuring the safety and stability of electricity usage, a paramount concern in modern energy management. In essence, this research endeavour underscores the critical role that big data plays in our contemporary world. By developing a purpose-built platform, the author contributes to advancing big energy data analytics and equips stakeholders with the tools and insights needed to navigate the complex landscape of energy equipment pollution. Ultimately, this work aligns with the broader trajectory of harnessing the power of big data to inform and optimize decision-making across many domains.

3. Proposed Improved K-means Algorithm.

3.1. Principle of improved K-means algorithm. The author introduces an upgraded version of the K-means algorithm to enhance the analysis of monitored power big data. This algorithm enhancement's essence lies in altering the conventional K-means clustering rules. Expressly, during the computation of the distance from the centroid, a novel component, represented as the particle weight proportion 'w', is incorporated. This addition enables a data point's category assignment to be determined based on the magnitude of the distance, effectively refining the clustering process.

Choose 'k' centre points from the dataset 'm', determine the cluster to which the remaining points belong, compute the mean for each cluster as the new centre point, and iterate this process until convergence. The algorithm's procedural steps can be summarized as follows:

Training sample $\{x_1, \ldots, x_m\}, x_i \in R^n$, divide it into k categories:

Step 1: Randomly select k out of m sample data: $\mu_1, \mu_2, \ldots, \mu_k \in R^n$;

Step 2: Calculate the distance between the remaining data and these k data separately;

$$C_i = \arg\min_j \|x_i - \mu_j\|^2 \tag{3.1}$$

Step 3: Redetermine the centre point of each class and recalculate the average value;

$$\mu_j = \frac{\sum_{i=1}^{m} \{C_i = j\} x_i}{\sum_{i=1}^{m} \{C_i = j\}} \tag{3.2}$$

Step 4: If the measurement function converges, terminate the program; Otherwise, continue with Step 2. The improved K-means method minimizes the evaluation function fitness $(A[1], A[2], \cdots, A[n])$.

$$\text{fitness}(A[1], A[2], \ldots, A[n]) = \sum_{i}^{k} \sum_{i}^{n} \text{Dist}\,(x_i, C_k) \tag{3.3}$$

$$\sum_{i}^{k} \sum_{i}^{n} \text{Dist}\,(x_i, C_k) = \cos\,(x_i, C_k) = \frac{\sum_1^m x_{ij} c_{kj}}{\sqrt{\sum_1^m x^2 ij \sum_1^m c^2{}_{kj}}} \tag{3.4}$$

In the formula: n is the number of data; $\text{Dist}\,(x_i, C_k)$ is the distance between x_i and the centre point C_k. The process of implementation is to generate C_k and continuously improve C_k based on the value of x_i, so that

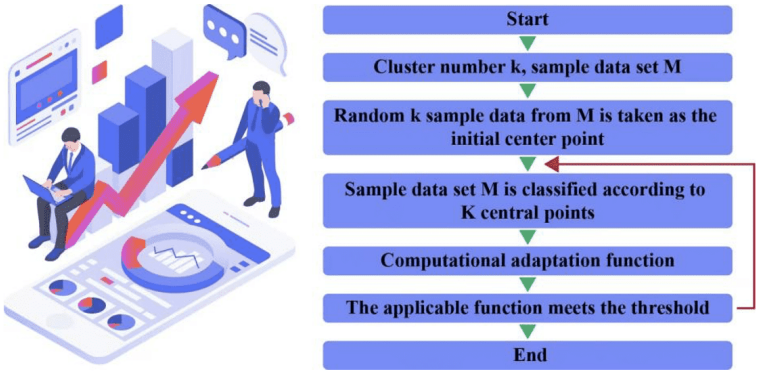

Fig. 3.1: Process flow of the improved K-means algorithm

Table 3.1: Experimental data attributes

Data set	Centre point	Number	Radius	Covariance
1	[0,0,0]	100	2	[0.300;00.350;000.3]
2	[1.25,1.25,1.25]	100	2	[0.300;00.350;000.3]
3	[-1.25,-1.25,-1.25]	100	2	[0.300;00.350;000.3]

data of the same class is more clustered and finally reaches the convergence condition.

$$C_k = \frac{\sum_{i=1}^{N} C_{ki} x_{ki}}{\sum_{i=1}^{N} C_{ki}} \tag{3.5}$$

The improved K-means algorithm introduces ω_k, and the implementation process of the same is expressed as:

$$\omega_k = \frac{1}{\sqrt{\omega_k}} \tag{3.6}$$

ω_k is the standard deviation, and when the target is C_k, the function increases by ${\Delta\varepsilon_k}^2$ after k times.

$$\Delta\varepsilon_k = \frac{1}{2m} \left(\omega_k \cdot \operatorname{dist}\left(c_k, x\right)\right)^2 \tag{3.7}$$

To assess the precision of the enhanced K-means classification method, we selected three distinct datasets for a comparative analysis of their clustering outcomes. The process flow of the improved K-means algorithm is shown in Figure 3.1.

The data attributes of the improved K-means algorithm [12] are detailed in Table 3.1, and the comparative evaluation of classification accuracy is presented in Table 3.2.

Table 3.2 reveals when applied to the classification of identical datasets, the enhanced K-means method consistently demonstrates significantly lower error rates than the traditional K-means approach, concurrent with a substantial increase in accuracy [10]. This compelling evidence underscores the efficacy of employing the improved K-means method in analyzing power monitoring equipment pollution big data, as it enables swift and precise clustering.

3.2. Design of power big data system. The following are the steps involved in data collection and processing.

Table 3.2: Comparison of data classification errors

Algorithm	Misclassification rate	Accuracy (Data 1)	Accuracy (Data 2)	Accuracy (Data 3)
K-means	13.21%	90.26%	91.94%	91.53%
Improve K-means	0.08%	95.21%	96.17%	95.82%

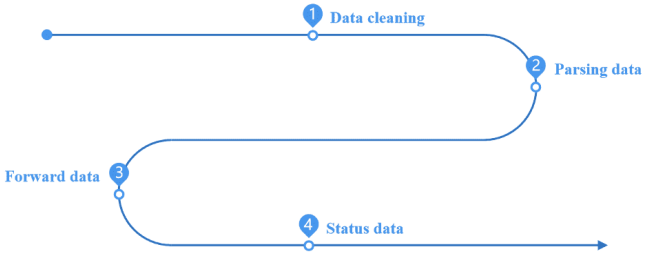

Fig. 3.2: Process flow in power big data system

(1) System architecture

The architecture of the power big data system encompasses various components, including data access, the mechanical floor, data storage, and data calculation. The mechanical floor comprises essential elements such as switching equipment and servers. Data access, on the other hand, encompasses multiple modules, including gateways, load balancing, and message middleware. The web management component enables diverse functionalities, including gateway management, terminal management, user management, and the parsing of original terminal messages [2]. During the data access, users can subscribe to data via message middleware, facilitating data analysis and storage.

The data import service also empowers users to import data of interest into storage, supporting various storage methods such as Hadoop, Redis, and Rdbms [20]. The platform monitoring component plays a pivotal role in overseeing the operational status of nodes and offers comprehensive monitoring across various aspects, including business, software, and systems. It also supports alarm notifications, including SMS and email.

(2) System data flow

The step-by-step procedure of the process of system data flow is represented in Figure 3.2 and explained as follows:

1) The terminal creates a long connection through LVS load balancing and gateway;

2) The gateway uses data packet decoding to encapsulate platform general data and writes it into Kafka;

3) Realize subscription of Kafka raw message data through real-time computing module and achieve data parsing; the parsed data is written in Kafka;

4) Subsequent modules can subscribe to the raw data of the terminal through Kafka or analyze the data. Thus, data can be stored and analyzed offline [9];

5) The forwarding service subscribes to data through Kafka and forwards it to other platforms;

6) The business management platform utilizes a data exchange interface to access the big data collection access module.

(3) Processing of streaming data

In the power big data system context, flow data pertains to the continuous stream of real-time data generated throughout power production, monitoring, and operational processes. The computing infrastructure employed for this purpose is the Storm system, which enhances real-time data analysis by persistently storing the computation outcomes in HBase. Stream data analysis, as a critical component, encompasses the processing, acquisition, and storage of streaming data within the framework of the power big data platform, delineating the interplay among data sources, processing stages, and computing platforms. The details of this intricate

Fig. 3.3: Module structure of power big data system

process are elucidated in Figure 3.3 to provide a comprehensive understanding of the processing workflow.

The data collection and preprocessing functions serve the crucial role of facilitating the precise, comprehensive, and instantaneous acquisition of data. These functions merge disparate datasets, encompassing incomplete data, various formats, and noisy inputs, to yield standardized data through noise reduction techniques. Subsequently, data processing applies specialized business logic to analyze standardized data thoroughly. Users have the flexibility to craft custom implementations through dedicated interfaces. Ultimately, the outcomes of these processes find their repository in HBase for storage and retrieval [19].

(4) Data collection and preprocessing

The power big data platform is a comprehensive data fusion platform, aggregating data from electricity generation and consumption domains, including sources like SCADA and energy metering systems. Notable examples of flow data encompass power equipment status monitoring data. Figure 3.4 provides a visual representation of the data collection and processing workflow. Upon gathering data from diverse facets of the power system, it is initially stored on an FTP server. Storm does not impose rigid constraints on data sources and formats, accommodating input types such as message queues, databases, and log files [6]. All that is required is implementing the corresponding interface in Spout.

(5) Result storage

Diverse data types exist within the power big data platform, each characterized by intricate structures and substantial volume. Consequently, adopting a multi-tiered storage system is imperative to cater to the varied demands of different business operations. This approach allows the platform to store its extensive big data reserves in alignment with performance and analytical requisites. For instance, data with substantial volumes and unstructured attributes, such as monitoring video data, finds its storage solution through the HDFS file system. In contrast, real-time processed data is efficiently stored in HBase, with its storage structure meticulously designed, as illustrated in Table 3.3. This strategic approach to data storage optimizes the platform's ability to handle and retrieve data following specific operational needs.

4. Experimental Results and Analysis.

4.1. System development. The power big data platform's requisites and constituent modules were meticulously examined to develop a robust power big data analysis platform. To execute system analysis and calculation tasks, a B/S architecture was employed. This framework guides users through a user-friendly interface to make crucial selections. These selections encompass picking data files, opting for intelligent algorithms, and configuring pertinent parameters. The system then seamlessly executes the chosen processes, automatically analyzing the data and presenting the results through intuitive icons for user comprehension and interpretation.

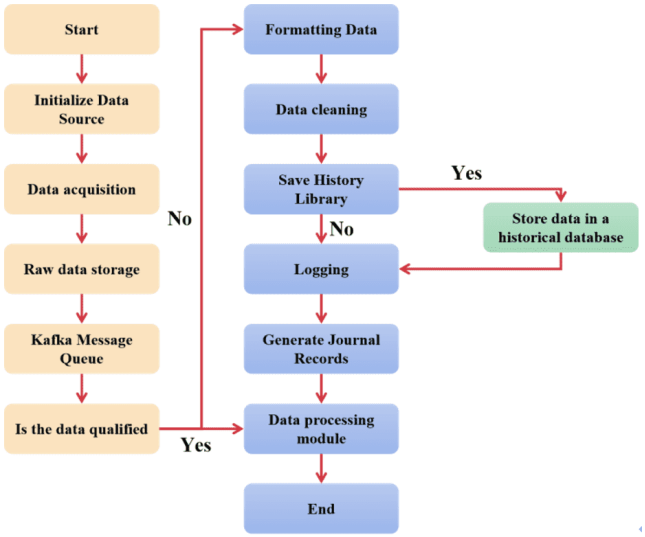

Fig. 3.4: Flow of data collection and processing

Table 3.3: HBase table storage data

Line keywords	Column cluster		Specific values of the N sampling points		
	Temperature	Dampness	V1	V2	Vn
Mac1+id	20	55%	13	12	16
Mac2+id	23	58%	14	12	15
Mac3+id	22	54%	16	14	18

4.2. System based on improved algorithms. The author monitors a specific power system daily, gathering data regularly. Whenever the environmental monitoring value surpasses the threshold of 0.1, it signifies a potential safety issue within the power grid during operation. The author harnesses an enhanced K-means algorithm to gauge the pollution status of various locations and ascertain equipment-related pollution situations.

In the initial step, the power big data analysis platform is employed to amass data about site pollution. This data is subjected to mean and variance analyses to derive valuable insights. Subsequently, the improved K-means method is applied to conduct a clustering analysis of site pollution levels. The outcomes of this analysis are vividly depicted in Figure 4.1, showcasing the clustering analysis results for power environmental monitoring stations [7].

An evaluation of the data representation within the system reveals that user identity verification is a fundamental security measure. Access to the system is granted solely upon entering the correct account and password [3, 1]. Once inside the system, users can select data files and intelligent machine learning algorithms, fine-tune parameters, and initiate data analysis. The analysis results are then elegantly presented through charts and tables, offering decision-makers precise information. This capability greatly facilitates the assessment of the power system's performance.

These findings in Table 4.1, which presents the outcomes of site classification. Following the research detailed in this article, it becomes evident that data values indicate pollution levels, with heavily polluted sites yielding the highest numerical values, moderately polluted data falling in the middle range, and lightly polluted data exhibiting the smallest numerical values. This distinct differentiation between the three data categories is accompanied by pronounced periodicity in heavily polluted data, in contrast to minimal fluctuations in lightly

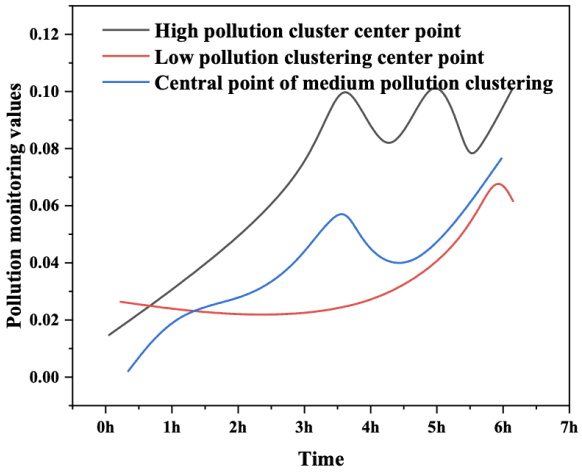

Fig. 4.1: Cluster analysis of power environmental monitoring stations

Table 4.1: Results of site classification

Classification	Pollution level	Number of sites	Processing strategy
1	Mild	353	Heavy
2	Degrees	83	Observation
3	Heavy	15	Regular cleaning

polluted sites. The extensive data mining efforts to monitor power station equipment pollution status enable timely equipment updates.

5. Conclusion. A comprehensive analysis system with both front-end and back-end components has been designed and successfully developed in response to power big data analysis requirements. The proposed system automatically performs data analysis and presents results in graphical form by guiding users through a structured process, including data file selection, intelligent algorithm choice, and parameter configuration. The improved K-means method has led to quantifiable advancements, including enhanced accuracy and a substantial reduction in misclassification rates compared to traditional K-means algorithms. When applied to classifying pollution levels in power equipment, this method significantly improves determining the true state of power equipment, thereby furnishing actionable insights for power equipment management. This comparative analysis shows that the traditional K-means algorithm yielded a relatively higher misclassification rate of 13.21% and slightly lower accuracy rates across all three datasets. These outcomes undeniably highlight the superior performance and efficacy of the improved K-means algorithm in the precise clustering and classification of data.

REFERENCES

[1] M. Aljasem, A. Irtaza, H. Malik, N. Saba, A. Javed, K. M. Malik, and M. Meharmohammadi, *Secure automatic speaker verification (sasv) system through sm-altp features and asymmetric bagging*, IEEE Transactions on Information Forensics and Security, 16 (2021), pp. 3524–3537.

[2] W. Bai and J. Liu, *Analysis of test scores of insurance salesman based on improved k-means algorithm*, in Proceedings of the International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, vol. 153, Springer, 2022, pp. 1192–1201.

[3] W. Bao, K. Wang, X. Gao, J. Song, P. Zeng, D. Zhou, H. Zhu, and Q. Geng, *Verification of security measures for smart substations based on visualized simulation*, in IOP Conference Series: Earth and Environmental Science, vol. 461, IOP Publishing, 2020, p. 012002.

[4] S. Bo, *Application of k-means clustering algorithm in evaluation and statistical analysis of internet financial transaction data*, arXiv preprint arXiv:2202.03146, (2022).
[5] Z. Dong, Y. Men, Z. Li, Z. Liu, and J. Ji, *Chilling injury segmentation of tomato leaves based on fluorescence images and improved k-means++ clustering*, Transactions of the ASABE, 64 (2021), pp. 13–22.
[6] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*, vol. 72, Springer, 2015.
[7] J.-C. Huang, P.-C. Ko, C.-M. Fong, S.-M. Lai, H.-H. Chen, and C.-T. Hsieh, *Statistical modeling and simulation of online shopping customer loyalty based on machine learning and big data analysis*, Security and Communication Networks, 2021 (2021), pp. 1–12.
[8] W. A. Ibrahim and M. M. Morcos, *Artificial intelligence and advanced mathematical tools for power quality applications: a survey*, IEEE Transactions on Power Delivery, 17 (2002), pp. 668–673.
[9] X. Li, *Research on english teaching ability evaluation algorithm based on big data fuzzy k-means clustering*, in Proceedings of the EAI International Conference, BigIoT-EDU, vol. 467, Springer, 2022, pp. 36–46.
[10] Y. Li, R. Liu, Y. Bo, and H. Wei, *Analysis and research of students' mental health status based on k-means clustering under the background of big data*, in Proceedings of the International Conference on Cognitive based Information Processing and Applications, vol. 85, Singapore, 2022, Springer, pp. 437–444.
[11] W. Lv, W. Tang, H. Huang, and T. Chen, *Research and application of intersection clustering algorithm based on pca feature extraction and k-means*, in Proceedings of the 5th International Workshop on Advanced Algorithms and Control Engineering, vol. 1861, Zhuhai, China, 2021, IOP Publishing, p. 012001.
[12] X. Meng, J. Lv, and S. Ma, *Applying improved k-means algorithm into official service vehicle networking environment and research*, Soft Computing, 24 (2020), pp. 8355–8363.
[13] S. K. Prabhakar and S.-W. Lee, *Improved sparse representation based robust hybrid feature extraction models with transfer and deep learning for eeg classification*, Expert Systems with Applications, 78 (2022), pp. 18023–18050.
[14] F. Terroso-Saenz, A. González-Vidal, A. P. Ramallo-González, and A. F. Skarmeta, *An open iot platform for the management and analysis of energy data*, Future Generation Computer Systems, 92 (2019), pp. 1066–1079.
[15] L. Wang, K. Qiu, and W. Li, *Sports action recognition based on gb-bp neural network and big data analysis*, Computational Intelligence and Neuroscience, 15 (2021), pp. 795–798.
[16] X. Wang, C. Song, and M. Yu, *Research on power security early warning system based on improved k-means algorithm*, in Proceedings of the Big Data and Security - Third International Conference, vol. 1563 of Communications in Computer and Information Science, Shenzhen, China, 2021, Springer, pp. 73–89.
[17] X. Xu, X. Geng, Z. Gao, H. Yang, Z. Dai, and H. Zhang, *Optimal heart sound segmentation algorithm based on k-mean clustering and wavelet transform*, Applied Sciences, 13 (2023), pp. 547–552.
[18] J. Ye, J. Zou, J. Gao, G. Zhang, M. Kong, Z. Pei, and K. Cui, *A new frequency hopping signal detection of civil uav based on improved k-means clustering algorithm*, IEEE Access, 9 (2021), pp. 53190–53204.
[19] C. Zhu, Z. Liu, B. Zou, Y. Xiao, M. Zeng, H. Wang, and Z. Fan, *An hbase-based optimization model for distributed medical data storage and retrieval*, Electronics, 12 (2023), p. 987.
[20] N. Zhu and Q. Dai, *Basketball data analysis based on spark framework and k-means algorithm*, in Proceedings of the International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy:, vol. 98, Springer, 2022, pp. 853–857.
[21] J. Zhuang, C. Ren, D. Ren, Y. Li, D. Liu, L. Cui, G. Tian, J. Yang, and J. Liu, *A novel single-cell rna sequencing data feature extraction method based on gene function analysis and its applications in glioma study*, Frontiers in Oncology, 11 (2021), p. 797057.

Edited by: Venkatesan C
Special issue on: Next Generation Pervasive Reconfigurable Computing for High Performance Real Time Applications
Received: May 13, 2023
Accepted: Sep 16, 2023