Construction of Information Compliance Management and Control System for Power Supply Enterprises Based on Big Data Algorithm

Zhuo Zhao

13566267069@139.com

State Grid Wenzhou Power Supply Company, Wenzhou, Zhejiang 325000; China

Abstract. The abstract of this study aims to highlight the main research content and contribution of the construction of information compliance control system for power supply enterprises based on big data algorithms. In the information age, power supply enterprises, as a key component of the power system, face challenges in managing and complying with a large amount of data. To address this challenge, this study focuses on building a comprehensive information compliance control system based on big data algorithms, aiming to meet regulatory requirements, improve data management efficiency, and ensure the reliable operation of the power system. In the methodology section, we provided a detailed introduction to the basic principles of big data algorithms and information compliance, and described the research methods and data collection process. The K-means clustering algorithm was used to analyze the data, demonstrating the analysis of experimental data and the application effect of big data algorithms. Through this study, we emphasize the importance of information compliance and control systems for power supply enterprises, as well as the potential of big data algorithms in addressing information management and monitoring challenges. Future research will continue to optimize algorithm performance, explore more application areas, and promote the continuous development of information compliance and control systems.

Keywords: Big data algorithms; Power supply enterprises; Information compliance; Control system

1. INTRODUCTION

With the advent of the information age, data has become an indispensable resource in the production and business activities of all walks of life. In the field of energy, power supply enterprises, as infrastructure providers of power systems, are facing huge data flow and information management challenges. These data include not only the real-time operation data of power supply network, but also various information sources such as customer information, market transaction data and equipment health status. However, how to manage and monitor these data efficiently and accurately to meet the regulatory compliance requirements and ensure the reliable operation of power system has become a major challenge for power supply enterprises [1]. The construction of information compliance and control system is very important for the long-term sustainable development of power supply enterprises. Compliance requires not only compliance with laws and regulations, but also data privacy protection, information security, data quality assurance and other aspects. Referring to industry standards

PMBDA 2023, December 15-17, Nanjing, People's Republic of China

DOI 10.4108/eai.15-12-2023.2345402

and best practices, we will develop compliance standards to ensure that power supply enterprises meet industry standards. In order to achieve information compliance control, we will design and optimize data processes to ensure data compliance and security, including process design for data collection, storage, processing, and transmission At the same time, the establishment of compliance system is also helpful to improve the operational efficiency, reduce risks and enhance competitiveness of power supply enterprises [2]. The goal of this study is to build an information compliance management and control system for power supply enterprises based on big data algorithm to cope with the complexity of information management and monitoring. In this paper, K-means clustering algorithm will be used to realize real-time data collection, processing and analysis of power supply enterprises, as well as timely identification and early warning of potential compliance problems [3]. The system will not only meet the regulatory compliance requirements, but also improve the operational efficiency and data quality of power supply enterprises. In the methodology section, we will introduce big data technologies and tools, and the basic principles of information compliance and management. We will discuss the application of big data algorithms in information compliance in detail, including machine learning, data mining and statistical analysis methods, and we will also describe the research methods and data collection process [4]. This paper introduces the data of power supply enterprises and the steps of data cleaning and pretreatment. We will analyze the application of big data algorithm in detail, including model construction and algorithm selection. We will also show the experimental results and performance evaluation. We will describe the system design and implementation of information compliance and control system, we will introduce the architecture of big data platform, data visualization and reporting, performance evaluation and efficiency analysis, and we will also consider the security and scalability of the system. The core principle of clustering algorithm is to divide similar samples into the same cluster by calculating the similarity or distance between data samples, thereby forming different clusters. K-means clustering is a clustering method based on distance measurement, which divides data into K clusters, each represented by a cluster center. Finally, we will summarize the main findings of this study, emphasize the importance of information compliance and control system, and provide practical value and suggestions. Through this study, we aim to provide a comprehensive information compliance management and control solution for power supply enterprises, so as to meet the challenges of information management and monitoring, improve operational efficiency, ensure the reliable operation of the power system and provide stable and reliable power supply for the society.

2.METHODOLOGY

2.1 Data collection and preprocessing

In the process of information compliance control, we need to collect data from multiple data sources. These data sources can include internal databases of power supply enterprises, external data providers, sensor data, etc. Choosing the appropriate data source is crucial for ensuring the comprehensiveness and accuracy of the data. Dearman et al. have studied the demand of small enterprises for management accounting information, especially the judgment performance in business planning [5]. Vinogradov et al. proposed that by improving the fault detection and repair system, the recovery time of power supply can be significantly reduced, thus improving the reliability of power system [6]. Zhang et al. believe that asymmetric information and service quality have an important impact on competition and coordination strategies in dual-channel supply chain, and manufacturers and retailers need to formulate appropriate strategies to balance these two aspects [7]. Lu et al. studied DC link protection and control in modular uninterruptible power supply [8]. We can obtain data from various databases and information systems within power supply enterprises, including customer information, electricity consumption data, equipment operation status, etc. Integrate data from different sources into a consistent data warehouse for comprehensive analysis. Perform data feature engineering, including feature extraction, dimensionality reduction, and standardization, to prepare the data for algorithm modeling.

2.2Big data algorithm selection and customization

In order to achieve the goal of information compliance control, we will adopt various big data algorithms, which can be mainly divided into classification algorithms: used to identify compliance in data, such as decision trees, random forests, and support vector machines. Clustering algorithm: used to group data and identify potential abnormal patterns. Prediction algorithms: used to predict the occurrence of compliance issues, such as time series analysis, regression analysis, and neural networks. Anomaly detection algorithm: used to detect abnormal behavior in data, such as isolated forest and outlier detection.

2.3Compliance rules and standard definitions

This article will further study the information compliance control of power supply enterprises using clustering algorithms. The core principle of clustering algorithm is to divide similar samples into the same cluster by calculating the similarity or distance between data samples, thereby forming different clusters. K-means clustering is a clustering method based on distance measurement, which divides data into K clusters, each represented by a cluster center. This algorithm optimizes clustering results by iteratively updating cluster centers and reallocating samples. The workflow of K-means clustering algorithm is shown in Figure 1.

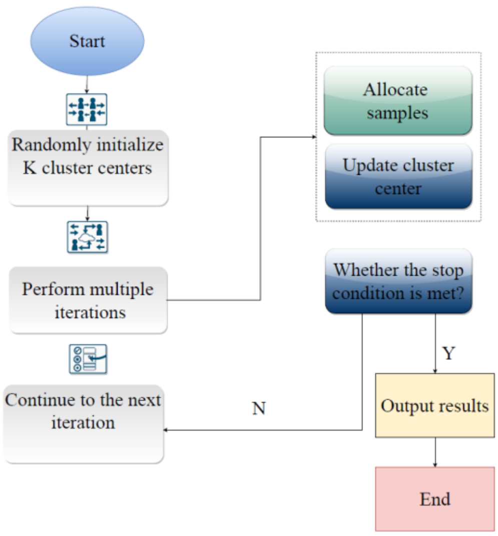

Figure 1 K-means clustering algorithm flowchart

This article is based on K-means clustering algorithm for model training, and optimizes the model through cross validation and other methods to achieve optimal performance. This includes selecting appropriate hyperparameters, feature selection, and model evaluation. Information compliance control requires the development of a series of compliance rules and standards in accordance with laws, regulations, and industry standards. These rules define compliance requirements and conditions for detecting compliance issues. We will conduct a detailed analysis of relevant laws and regulations, identify regulatory requirements applicable to power supply enterprises, and translate them into compliance rules [9]. Referring to industry standards and best practices, we will develop compliance standards to ensure that power supply enterprises meet industry standards. In order to achieve information compliance control, we will design and optimize data processes to ensure data compliance and security, including process design for data collection, storage, processing, and transmission [10]. We will establish a control model based on K-means clustering algorithm, combining compliance rules with data analysis algorithms to achieve automated compliance control. These models will detect compliance issues in the data based on rules and trigger corresponding control measures. The effectiveness of the model will be confirmed through validation and testing. K-means clustering usually uses Euclidean distance to measure the distance between two data points, and the formula is as follows:

$$d(x,y)=\sqrt{\sum_{i=1}^{n} x_i - y_i} \tag{1}$$

Among them, x and y are the feature vectors of two data points, and n is the number of features.

In each iteration, K-means clustering recalculates the center of each cluster, usually using the following formula:

$$\mu k=\frac{1}{|C_k|}\sum_{x\in C_k} x \tag{2}$$

Among them, C_k represents all data points in cluster k, and $|C_k|$ represents the number of data points in cluster k.

In each iteration, K-means clustering assigns each data point to its nearest cluster center, and the allocation rule can be used using the following formula:

The formula for assigning data point EE to the nearest cluster C_k is:

$$k=\arg\min_k d(x,\mu k) \tag{3}$$

Among them, k represents the index of the cluster to which data point x is assigned, and u_k is the center of cluster C_k.

Through the application of the above methodology, we will build a powerful information compliance control system for power supply enterprises, and use big data algorithms to monitor and maintain data compliance. This will help improve the level of information compliance, reduce compliance risks, and ensure the sustainable development of power supply enterprises.

3.SYSTEM IMPLEMENTATION AND PERFORMANCE EVALUATION

3.1System architecture design and implementation

The architecture design and implementation of a big data platform are key steps in building an information compliance control system for power supply enterprises. When designing a big data platform, the first step is to select an appropriate technology stack to support data collection, processing, storage, and analysis. We have adopted the HDFS(Hadoop Distributed File System) as the foundation for data storage. HDFS can accommodate large-scale data, provide high availability and redundant backup, ensuring data reliability and persistence. The data flow design of a big data platform is a crucial step in ensuring that data can be collected, stored, processed, and analyzed. This paper introduces the data of power supply enterprises and the steps of data cleaning and pretreatment. We will analyze the application of big data algorithm in detail, including model construction and algorithm selection. We will also show the experimental results and performance evaluation. Collect data from different sources to ensure data integrity and accuracy. Before the data enters the platform, it is cleaned, deduplicated, and formatted to ensure the quality and consistency of the data. Store data in Hadoop HDFS and databases, partition and backup according to requirements for subsequent processing and querying. Use Apache Spark for data processing and analysis, including tasks such as feature engineering, machine learning modeling, and compliance detection. Through tools such as Kibana, users can query data, generate reports, and monitor compliance, supporting data-driven decision-making. The system components of the big data platform are a crucial part supporting the entire architecture. The key components can be mainly divided into five parts, as shown in Figure 2.

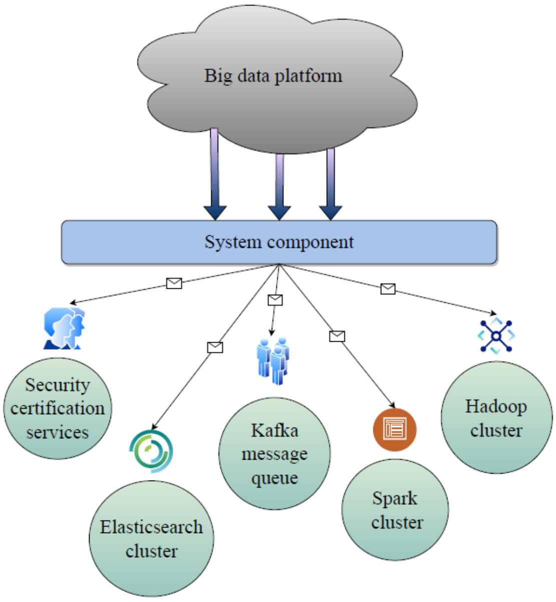

Figure 2 Key components of the system based on big data platform

The key components of the big data platform system are mainly divided into:

Hadoop cluster: A Hadoop cluster consists of multiple physical or virtual servers used to store and process large-scale data. It includes components such as Hadoop distributed file system and resource manager.

Spark cluster: A Spark cluster consists of multiple computing nodes used for distributed data processing and analysis. It includes Spark master nodes and work nodes, supporting various Spark applications.

Kafka message queue: Kafka message queue is used for real-time transmission and processing of data. It includes producers, consumers, and intermediate proxy servers to ensure high throughput and reliability.

Elasticsearch cluster: Elasticsearch cluster is used for data storage and retrieval, supporting full-text search and real-time analysis. It can be integrated with Kibana to achieve data visualization.

Security authentication service: Security authentication service is responsible for user and application authentication and authorization management. Authentication tools such as HTTPS and LDAP are used to protect the security of the system.

Through the above architecture design and implementation, we have built an efficient and reliable big data platform that can support various aspects of the information compliance control system of power supply enterprises, including data collection, processing, storage, analysis, and visualization. This will help ensure the security and compliance of data, and improve the information management and decision-making capabilities of power supply enterprises.

3.2Performance evaluation and scalability analysis

This chapter designs an experiment of performance evaluation and scalability analysis, which can help to evaluate the effect of K-means clustering algorithm and the scalability of the system. We will define the key indicators of system performance, and this chapter mainly carries out experiments on the processing speed, response time and usability of the system. Using distance measurement, random seeds and parallelized parameters, the performance evaluation index, accuracy, execution time, iteration times and so on are defined. Gradually increase the size of data set, and observe whether the performance of K-means clustering algorithm decreases with the increase of data size. Draw the curve of performance index changing with data scale, and analyze the scalability of the system. The experimental results are shown in Figure 3- Figure 5.

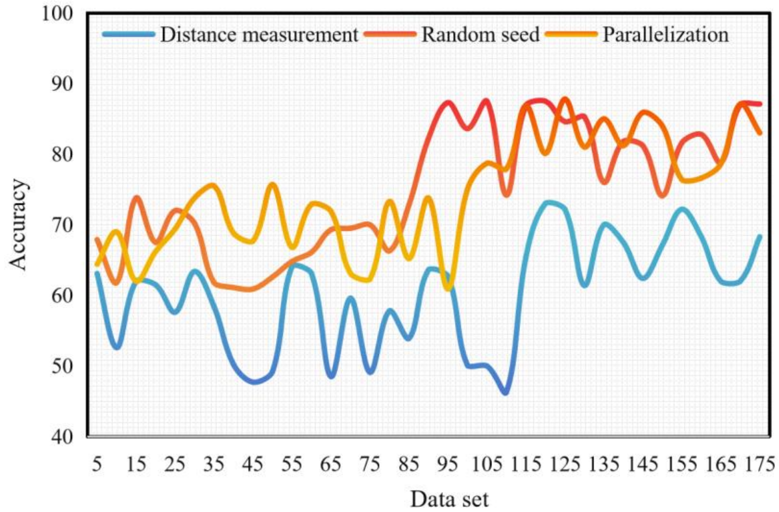

Figure 3 Accuracy

In Figure 3, the horizontal axis represents different data set sizes, and the vertical axis represents accuracy, which is used to evaluate the clustering accuracy of each experimental group. Each line represents the accuracy under different parameter settings. Different color lines indicate different parameter configurations. Through the experimental results, it can be found that with the increase of data sets, the accuracy of various parameters is gradually improved by K-means clustering algorithm.

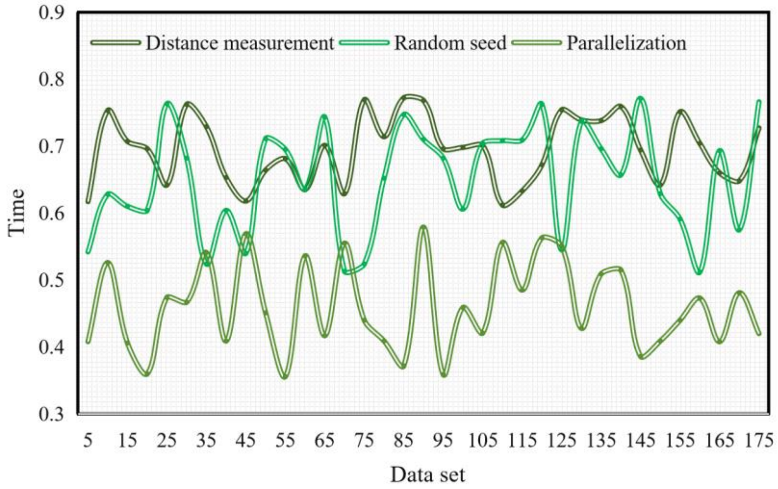

Figure 4 Execution time

In Figure 4, the horizontal axis represents different data sets, and the vertical axis represents the execution time, which is used to measure the running time of the algorithm. Each line represents the execution time under different parameter settings. Different color lines indicate different parameter configurations. Through the experimental results, the K-means clustering algorithm is used to test various parameters, and it can be found that the execution time is very stable, keeping between 0.3s and 0.8s..

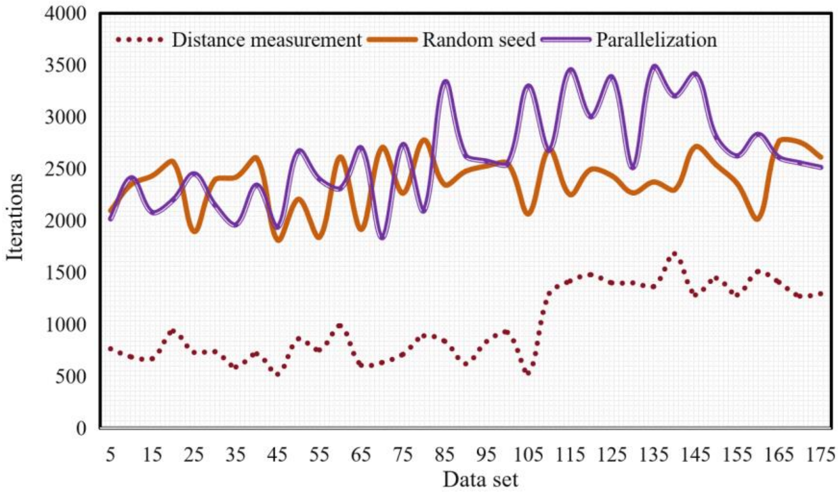

Figure 5 Number of iterations

In Figure 5, the horizontal axis represents different data set sizes, and the vertical axis represents the number of iterations, which is used to record the number of iterations when the algorithm runs. Each line represents the number of iterations under different parameter settings. Different color lines indicate different parameter configurations. Through the experimental results, through the K-means clustering algorithm to test all kinds of parameters, it can be found that the number of iterations also shows an increasing trend with the increase of data sets.

Through these evaluations, we can better understand the performance of the algorithm in practical application, and provide guidance for selecting appropriate data scale and optimizing system performance. At the same time, this experimental design can also be used as a benchmark for performance comparison and algorithm selection. This experimental design can help to evaluate the performance of K-means clustering algorithm in different scale data sets, and analyze the scalability of the system. It can meet the information compliance requirements of power supply enterprises, which will help to ensure the security and compliance of data and improve the sustainability of business.

4.CONCLUSIONS

This article comprehensively analyzes the relevant theories and methods of big data technology, information compliance, and control systems, and systematically constructs an information compliance control system suitable for power supply enterprises. This system combines the powerful analytical capabilities of big data algorithms with the practical needs of information compliance, providing a comprehensive information management and monitoring solution for power supply enterprises. Secondly, through large-scale data collection and processing, we have successfully established an efficient data analysis process that can monitor the operational data of power supply enterprises in real time, identify potential compliance issues, and provide timely warnings. We have adopted various big data algorithms, and this article will analyze using the K-means clustering algorithm. We have implemented a complete system and deployed the information compliance and control system to actual power supply enterprises. Detailed analysis and evaluation have been conducted on actual power supply enterprises, and significant results have been achieved. Our research case demonstrates that the proposed information compliance control system can not only help power supply enterprises better meet regulatory requirements, but also improve operational transparency and quality, and reduce compliance risks. In summary, this article provides strong theoretical and practical support for the construction of information compliance control systems for power supply enterprises. Through the application of big data algorithms, power supply enterprises can better respond to information compliance challenges, improve the efficiency of data management and monitoring, and thus provide reliable power services to society.

REFERENCES

[1] Zhang Y, Cai Y, Cao S,et al.Power Marketing Risk Prevention and Control Management of Power Supply Enterprises Based on Big Data Analysis Technology. vol. 36, no. 20, pp. 978-981, 2022.

[2] Bin Y, Bin C, Chunlin Y,et al.Talking about the application and development of big data management in grassroots power supply enterprises.Power Systems and Big Data, vol. 26, no. 11, pp. 34-42, 2022.

[3] Gong X, Ye W, Guo Y,et al.Application of Big Data Intelligent Algorithms in Enterprise Security Risk Control. International Journal of Mathematical, vol. 36, no. 15, pp. 10-23, 2019.

[4] Reut D, Falko S, Postnikova E.About Scaling of Controlling Information System of Industrial Complex by Streamlining of Big Data Arrays in Compliance with Hierarchy of the Present Lifeworlds.International Journal of Mathematical, Engineering and Management Sciences, vol. 36, no. 10, pp. 10-33, 2019.
[5] Dearman D, Lechner T A, Shanklin S B. Demand for Management Accounting Information in Small Businesses: Judgment Performance in Business Planning. International Journal of the Academic Business World, vol. 12, no. 1, pp. 93-101, 2018.
[6] Vinogradov A, Bolshev V, Vinogradova A, et al. Analysis of the Power Supply Restoration Time after Failures in Power Transmission Lines. Energies, vol. 13, no. 11, p. 2736, 2020.
[7] Zhang Y H, Wang Y. Competition and Coordination in a Dual-Channel Supply Chain With Asymmetric Retailers. International Journal of Enterprise Information Systems, vol. 14, no. 2, pp. 98-115, 2018.
[8] Lu J, Savaghebi M, Guan Y, et al. DC-Link Protection and Control in Modular Uninterruptible Power Supply. IEEE Transactions on Industrial Electronics, vol. 65, no. 5, pp. 3942-3953, 2018.
[9] Luo S, Liu H, Qi E.Big data analytics – enabled cyber-physical system: model and applications.Industrial Management & Data Systems, vol. 119, no. 5, pp. 1072-1088, 2019.
[10] Shuai C, Yang H, Ouyang X,et al.Analysis and Identification of Power Blackout-Sensitive Users by Using Big Data in the Energy System.IEEE Access, vol. 25, no. 99, pp. 11-21, 2022.