Intelligent Mining Method of Enterprise Management Information Based on ID3 Decision Tree Algorithm

Ying Liu[1,2], Zulkefly Abdul Karim[2,*], Norlin Khalid[2], Fathin Faizah Said[2]

[1]*School of Business, Jinggangshan University*
[2]*Center for Sustainable and Inclusive Development Studies (SID), Faculty of Economics and Management, Universiti Kebangsaan Malaysia(UKM), 43600 Bangi, Selangor, Malaysia*
[*]*Corresponding author. Email: zak1972@ukm.edu.my*

ABSTRACT
The intelligent mining method of enterprise management information is time-consuming in the application process. This paper designs an intelligent mining method of enterprise management information based on ID3 decision tree algorithm. According to the definition features of normalized feature vectors, the vector length is calculated. Under the condition of known candidate attributes, the information entropy of child nodes is calculated. The ID3 decision tree algorithm is used to optimize the intelligent mining mode. Experimental results: The mining time of the designed enterprise management information intelligent mining method and the other two intelligent mining methods are 38.814ms, 44.868ms, 45.032ms respectively, which proves that the time of the enterprise management information intelligent mining method is reduced after the integration of ID3 decision tree algorithm.

Keywords: *ID3 decision tree algorithm, Enterprise management information, Intelligent mining, Business processing, Sample data, Market management.*

1. INTRODUCTION

ID3 decision tree algorithm is a common method in information intelligent mining. It is a process algorithm that classifies data sets by tree rules. Generally speaking, it is an information intelligent mining algorithm that generates tree decision model from sample data set and uses it for classification. In a certain range, decision tree algorithm can be divided into "prediction" range. The innovation of enterprise management, the most important is to make full use of computer information technology, to realize the construction of enterprise management information. Information technology has become a resource to change the development of enterprise market economy and technological innovation. The use of advanced computer management information technology to create enterprise information management system is the necessary means for enterprises to obtain strong market competitive advantages and good development. Decision tree can simply and directly describe the relationship between description attributes of multiple objects and the final classification of objects. Each node in the decision tree represents an object attribute, and the branching path of each node represents different attribute values of the object and the division of data sets [1-2]. Management information intelligent mining must continually advance and development, to make an advantage in the fierce market competition environment, information resources is one of the important constitute of enterprise management information intelligent mining, selection of components to complete the enterprise management information system analysis and design of information resource, greatly raise the reuse of enterprise management information system, replaceability, Can spend low cost and rapid development and integration and expansion of enterprise management information system. The leaf node is the final classification corresponding to the path. In general, the decision tree only has a single decision result output. If multiple outputs are needed, the processing capability of multiple outputs can be obtained by establishing a series of independent but nested decision trees. Each decision tree is a tree-like decision model built from sample data, and its branches classify data tuples or objects in the actual data according to their attributes. The development trend of intelligent mining of management information is reflected in the trend of network, value, intelligence, integration and humanism

[3-4]. Management information system plays an important role in the management of modern enterprises. It is not only an effective way to complete the modernization of management, but also stimulates the process of enterprise management toward modernization. Decision tree can rely on the division of sample data set, node selection and tree structure construction, and in the construction process can be recursively "pruning" the tree, until the class can no longer be divided or there is only a separate class. At the same time, decision tree can also be constructed by mathematical calculation method, which can achieve more ideal results in some aspects. Management information intelligent mining is the most basic work, such as printing statements, personnel management, wage management, and the further development of enterprise financial management, inventory management, market management, a single business management, it belongs to the data processing system.

2. INTELLIGENT MINING METHOD OF ENTERPRISE MANAGEMENT INFORMATION BASED ON ID3 DECISION TREE ALGORITHM

2.1. Identifying the Characteristics of Enterprise Management Information

The enterprise management information system has its uniqueness, and the construction of enterprise management information system also has its laws that should be followed. The top level of enterprise (decision level) needs decision support system, the middle level of enterprise (control level) needs knowledge work system (information transmission, inquiry), and the basic level of enterprise needs business processing system. Enterprise management information should combine the characteristics of enterprise's own industry development, enhance the investment in enterprise management information and the study of leading the development of information technology in the world, not blindly trust big, step by step forward development. The establishment of the system framework of the management information system, the design of the detailed information system architecture for the mutual operation within the enterprise, and the setting, selection and implementation optimization of the organizational framework of information [5-7]. Different information service objects are different. In the manufacturing industry, some people think that the enterprise management information system contains most of the information of the enterprise, but some enterprises have few senior leaders directly enter the system to inquire information. The reason is that, on the one hand, there are many levels of query interface, and the operation is relatively complex, which is not as convenient as the paper file used by leaders. On the other hand, it is a business control system, which focuses on process control, and the information in it can only be used by management and decision-making layer after secondary development. Thus, the characteristics of enterprise management information are obtained, as shown in Figure 1:

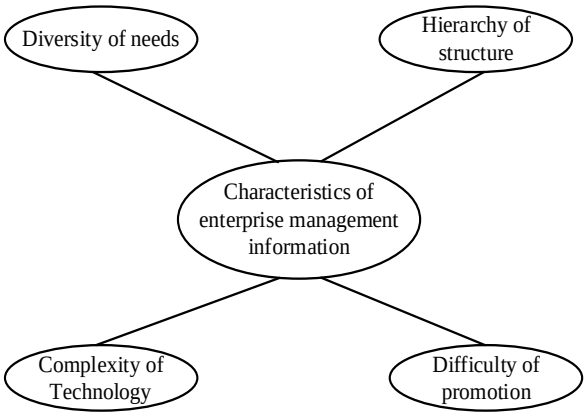

Figure 1 Characteristics of enterprise management information

As can be seen from Figure 1, the characteristics of enterprise management information include diversity of needs, hierarchy of structure, complexity of technology and arduousness of promotion. Strive for the enterprise information technology technology level to the world enterprise information enterprise first-class level. The overall design objectives and strategic planning of the system should be consistent with the development strategy of enterprise informatization to ensure that the investment decision of technology information conforms to the prospect of enterprise. Under the guidance of enterprise development strategy, management information system should analyze, reason, diagnose, evaluate and summarize enterprise development strategy planning on the basis of understanding enterprise's own development strategy planning and market planning. In fact, we can make an information platform to process various types of information in different business systems of enterprises and convert the extracted format of information to facilitate middle and high-level inquiries. In this way, the middle and senior leaders of enterprises can read various kinds of information without having to go to different systems, and can reduce the use of paper and documents that are inconvenient to keep. In view of the diversity of such demands, enterprises should deal with the contradiction between the richness and repeatability of management information system, reasonably plan the management information system of enterprises, reduce unnecessary input, clarify the division of labor of management information system, and deal with the interface between a variety of different types of management information system. Through the construction of excellent science and technology support system, lead its healthy and rapid development. Overall frame design and construction of management information, must be

enterprise market management platform, internal and external supervision platform, customer service platform, information center, internal construction from four aspects of organic integration, set up take the customer as the center of the electronic sales service channels, form efficient market operation of advanced enterprise information management system standard. The enterprise management information system is as hierarchical as the enterprise itself. First of all, an enterprise group is usually composed of many subsidiaries (divisions). In the subsidiary, all kinds of management information and systems are hierarchical, that is, mountain decision support system, knowledge work system and business processing system constitute a pyramidal hierarchy. But between groups and subsidiaries, their management information systems form another pyramidal hierarchy. Marketing Department is the external window of enterprises, related to the rise and fall of enterprises. To make an enterprise in the increasingly fierce market competition in an invincible position, it is necessary to improve the enterprise's marketing management mechanism, the introduction of computers into the enterprise's marketing management, is to improve the management level, an important measure to update the management means. In this case, there is a contradiction between the unity and uniqueness of the system. Groups seek uniformity for the sake of information exchange between systems or because of an inevitable centralization tendency. The subsidiary pursues system uniqueness in order to satisfy its unique needs for information system or because of the same unavoidable tendency of decentralization. In view of the hierarchical nature of this structure, enterprises should formulate information system and information technology standards for the whole group, and solve the contradiction between the unity and uniqueness of the system through open and unified standards. The purchasing and supply module is responsible for the recording and tracking of data and information in the whole process of purchasing materials from application to contract signing, receiving management, acceptance and warehousing, including planning maintenance, material requisition, purchase order maintenance, supplier management, contract maintenance, code maintenance and receiving management. The purpose is to reduce the stock capital as much as possible and improve the economic benefits of enterprises under the premise of ensuring the production and economic activities of enterprises. The technology involved in the construction of enterprise management information system is complicated. From the point of view of hardware, different grades of computers, network communication equipment, input/output equipment and multimedia equipment may be used according to different levels of application. Based on the above description, complete the steps to identify the characteristics of enterprise management information.

2.2. Filtering Data Access Objects

The JDBC API provides a unified programming interface for Java developers to use databases, but for many applications, persistence storage is implemented using different mechanisms, and the apis used to access these different persistence storage mechanisms vary greatly. In general, data in a data warehouse comes from heterogeneous operational databases. The data in these heterogeneous operation databases are not all correct, and there are often incomplete, inconsistent, inaccurate and repeated data inevitably, which are collectively called dirty data. Whenever data is loaded into the data warehouse, dirty data is cleaned. Data cleaning can occur either before or after the data is loaded into the data warehouse. If the data source is different, the data access is different. Access to persistent storage (such as databases) can also vary greatly if the type of database storage (relational, object-oriented, data files, and so on) and vendor implementation are different. Therefore, in addition to using JDBC, you also need to use design patterns to achieve transparent application access to the database. The Data Access Object pattern uses data access objects to encapsulate and abstract access to all data sources, manages connections to data sources for easy retrieval and storage of data, and provides a transparent underlying data access implementation for business objects. Data selection actually takes place on two dimensions. The first is the selection of column or parameter dimension, which is a part of the process of intelligent information mining [8-10]. The second choice is the row or record dimension, which is based on the value of each field. In a relational database, the selection of columns and rows can be done either by SQL language or by database front-end tools. Data selection requires a detailed and in-depth understanding of the problem domain and the underlying data. The DAO pattern completely wraps the reading and manipulation of data and wraps the data access API that interacts with the database, wrapping data read from and sent to the database in numeric objects that communicate with the business layer. The flexibility provided by the DAO pattern is due to the fact that applications do not directly access data sources, but instead create DAO objects and use them to access data sources. According to the definition features of normalized feature vectors, the calculation formula of vector length is as follows:

$$L = \sqrt{\sum \frac{\|\beta - n\|}{m^2 - 1}} \qquad (1)$$

In Formula (1), β represents the number of entries in the document, and m, n represent two associated documents respectively. On the basis of Formula (1), calculate the inner product between vectors, which is expressed as follows:

$$Sim(m,n)=\sum_{\varepsilon=1}\frac{|L+\beta|}{2}\times\varepsilon^2 \qquad (2)$$

In Formula (2), ε represents the similarity between documents, and the remaining variables are the same as in Formula (1). After selecting the data, it is necessary to preprocess the data before mining. A common requirement of information intelligent mining is to generate a new field or attribute based on two or more fields. It usually takes the form of the ratio of two data points, as well as their sum, product and difference. Other transformations can be to convert a date to a day of the week or a day of the year. When reading data, you can use numeric objects to hold the obtained data. Because the framework can reuse code, it is easy to build applications from existing component libraries because the components use interfaces defined by the framework, making communication between components easy. Computing attributes is often necessary because transaction processing applications are focused on processing as little data as possible for recording transactions, requiring only minimal storage requirements and reduced processing time, rather than collecting more transaction information. In the process of intelligent information mining, numerical data represented by vectors or arrays can sometimes be processed in groups, that is, vectors as a whole are regularized [11-12]. These standard interfaces make it possible to build a wide variety of systems by assembling existing components. New components can be inserted into the framework as long as they conform to the interface definition, and component designers can reuse the architecture's design. The framework can also reuse analysis, so that if all people analyze transactions according to the framework's ideas, they can divide it into the same artifacts and adopt similar solutions, thus enabling communication between analysts using the same framework. One of the core problems with common frameworks for reusability is the use of common interfaces. The second way is to take the sum of all the elements and divide it by each number. In this case, the sum of the regularized elements is 1.0, and the value of each element represents how much they contribute to the grouping. The third way is to take the maximum value in the vector and divide it by each of the entries. The management of raw data is an important aspect of information intelligent mining and application development by using computational intelligence method. When multiple documents are used as training samples, necessary adjustments should be made to its fitness function to meet the optimization of intelligent information mining under the condition of multiple documents [13-14]. The average or maximum relevance of keyword vector to multiple training samples can be used as fitness function. Then the average value and maximum value between vectors can be calculated as follows:

$$G_A=\frac{1}{\eta}\sum Sim(m,n) \qquad (3)$$

$$G_{\max}=\max\frac{\left\|\frac{1}{\eta}\sum Sim(m,n)\right\|}{D} \qquad (4)$$

In Formulas (3) and (4), η represents the number of keywords in the population, and D represents the individual fitness in the training document. With a common interface, developers can build new component implementations and easily integrate them into the architecture. Therefore, common interfaces are used in the concrete implementation of both the business object architecture and the service architecture described below. In the business object architecture, business objects are configured through metadata. In service architecture, dynamic invocation of business services is realized through common interfaces. For example, the method used in neural network mining is to randomly divide the raw data into two or more data sets. One data set is used to train the neural network and the other is used to test the network. You must ensure that the neural network is in training mode without seeing the test data. That is, it cannot learn from test data or adjust its weights. On the basis of the above description, complete the steps of filtering data access objects.

2.3. ID3 Decision Tree Algorithm to Optimize Intelligent Mining Mode

ID3 algorithm is a decision tree learning method based on information gain attribute selection measure method proposed by Quinlan. It is the classic algorithm of decision tree algorithm, and many subsequent algorithms proposed by scholars are improved on its basis [15]. The core idea of ID3 algorithm is to select classification attributes on all nodes of the decision tree by calculating the information gain of attributes, so that the maximum category information about the tested samples can be obtained when each non-leaf node is tested. Its basic method is: do all its properties, choose the maximum attribute of information gain splitting decision tree node, based on the attribute set up branches, the different attribute value to each branch a subset of the recursive calls to the method to set up the branch of child nodes, until all the subsets of attributes of the same category or not can be split, resulting from a decision tree. Under the condition of known candidate attributes, the information entropy of child nodes is calculated as follows:

$$H_\mu=\sum_{p=1}\frac{\mu\times\eta}{D_p-\sqrt{\mu^2}} \qquad (5)$$

In Formula (5), μ represents the number of data subsets after splitting, p represents the splitting

probability, and D represents the splitting node. There are many types of enterprise processes, including the following aspects: order processing process: the input is the customer's order (or some demand intention), and the output is the goods sent.The payment order and customer satisfaction. Product development process: input is the customer's consumption idea, concept and concept, output is the new product sample. Service process: the input is the problem that the customer needs to understand and deal with, and the output is the solution and solution of the problem. Sales process: input is potential customers, and output is the payment order. Strategy development process: input is a variety of variables in the internal and external environment of the company. And output is a variety of strategies for enterprise development. Management process: input is the enterprise internal and external environment of all kinds of relationship elements and problems, and output is the enterprise operation of all kinds of relationship rules and methods. According to formula (5), the data subset with the minimum splitting probability can be obtained, which is expressed as follows:

$$F = \left(\frac{T}{\varphi} + 1\right)\sum T^2 - L \qquad (6)$$

In Formula (6), T represents the set of categories to which the sample belongs, and φ represents the number of samples. In the informatization practice of many enterprises, we can see that the original management mode and service mode of enterprises have undergone fundamental changes through the implementation of management information system. The core of the management reform strategy of information system operation is the optimization process of enterprise business process under the environment of management information system and the reform of enterprise organization. Business object architecture is the foundation of business application development. Most of the application's business logic can be found in this layer. The business object component is the foundation on which business transactions and business logic processes are established. After the preparation of information intelligent mining data, such as cleaning, extraction, selection and arrangement, it can enter the stage of information intelligent mining (narrow concept of information intelligent mining). Information intelligent mining is an important step of knowledge discovery, which is formed by establishing mining model and implementing corresponding mining algorithm. We know that the MVC architecture design pattern also includes service-based architecture, in which business service components usually just provide transactional or procedural wrapping for these business object components, while business object components actually do most of the business logic. Therefore, business objects are a core element of the business object architecture. The scope of the process refers to the number of organizational units such as business departments or functional departments involved. Information intelligent mining is an iterative process. Only through repeated interactive implementation and verification can we find the best way to solve problems, and only through continuous generation, screening and verification can we find valuable and potential knowledge. The business object architecture includes the components in the system that implement the business objects. Each component manages the data and business logic associated with specific business objects, including the persistence of business object data, particularly relational database data. Depending on how the business object is implemented, object persistence can be implemented by the corresponding container through CMP or BMP, or by a separate data access layer through regular Java class objects. After adjusting the output node vector, the weight calculation formula of the vector is obtained, which is expressed as follows:

$$\Delta R_{ij} = \frac{\varphi_{ij}}{2} \times (\sigma - 1) \qquad (7)$$

In Formula (7), i, j represent input and output sample data respectively, and σ represent variable learning speed. In the stage of information intelligent mining, the task or purpose of mining should be determined firstly, such as classification, clustering, association rule discovery, etc. Once the task is determined, it is time to decide what kind of mining algorithm to use. The same task can be achieved by different algorithms, and the selection of model algorithm is mainly based on two factors: first, different data have different characteristics, so it is necessary to use relevant algorithms to mine. The other is the requirements of users or practical applications. Some users may want to acquire descriptive and easily understood relational knowledge, while some users or systems aim to acquire predictive knowledge with the highest accuracy of prediction. Therefore, the best way to implement data persistence is to separate access to the database into a separate data access layer DAO. The business object component structure is based on business object entities. A narrow-scope process may occur within one operating department or functional unit, and a broad-scope process may cross several operating departments or functional units. The size of a process depends on its business content. Some processes consist of just a few very simple tasks, while others may include many highly complex and interrelated tasks. Business objects are actually the foundation of the functionality of the business logic, and they often determine some of the characteristics of the application. The common features of business objects are derived from the object-oriented development concept that a

business object encapsulates both entity data and behavior. After completing the above work, we can implement the operation of information intelligent mining, and use the selected model and algorithm to extract the knowledge needed by users from the data. Based on this, the steps of optimizing intelligent mining mode are completed.

3. EXPERIMENTAL ANALYSIS

3.1. Setting up the Experimental Environment

According to the test needs of this experiment, set up the experimental environment. On the server, the operating system is Windows 2020 Server, the database system is SQL Server2019 Chinese enterprise edition or later, and the browser is Microsoft Internet Explorer. Network requirements: Supports TCP/IP, the server and client must be on the same physical network segment, and the server must have a fixed IP address. Hardware requirements: minimum server configuration requirements: Intel P 4/2.6GB/256MB memory /40GB hard disk. There needs to be a common way to include the data and other related content required by the service when using standard service interfaces. In the implementation, we use the argument list class ArgumentLis to implement the ValueObject interface, which contains a member variable attributes that holds a collection of name-value pairs for input parameters. Since it does not know the specific data type, it needs to be stored as a collection of strings that can be retrieved based on different data types using the GET method. At the same time, configure the data source for the data management analyzer, using the system DSN to carry out, in the configuration of our database type to choose Microsoft OLE DB Provider for SQL Server, and then we set up the data source, data source name and user name and other information, complete the construction of the experimental environment.

3.2. Experimental Result

This experimental test is carried out in the form of experimental comparison. The enterprise management information intelligent mining method based on cloud computing and the enterprise management information intelligent mining method based on neural network are selected for experimental comparison with the enterprise management information intelligent mining method designed this time. The mining time of the three methods is tested under different population sizes. The experimental results are shown in Figure 2-3:

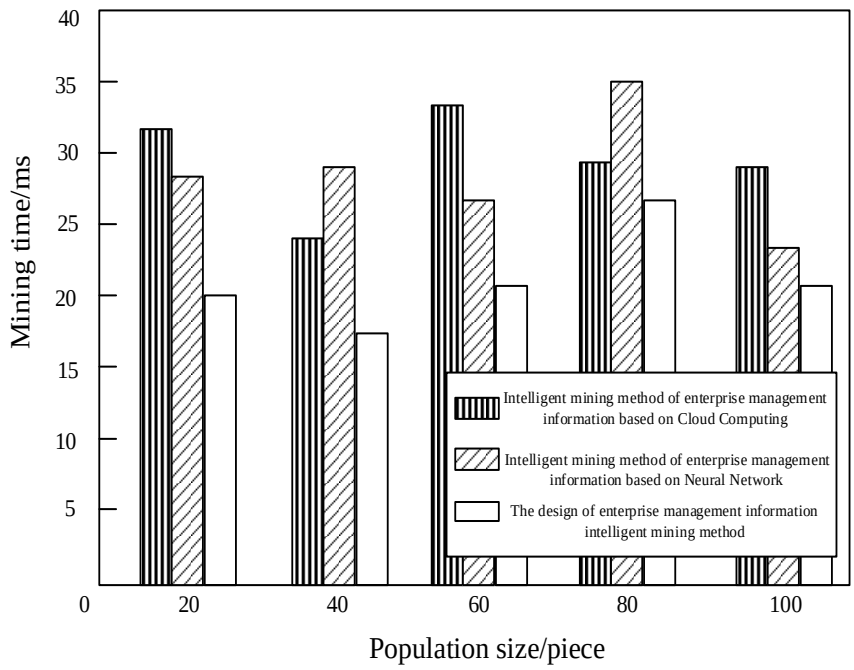

Figure 2 Mining time of population size 100 (ms)

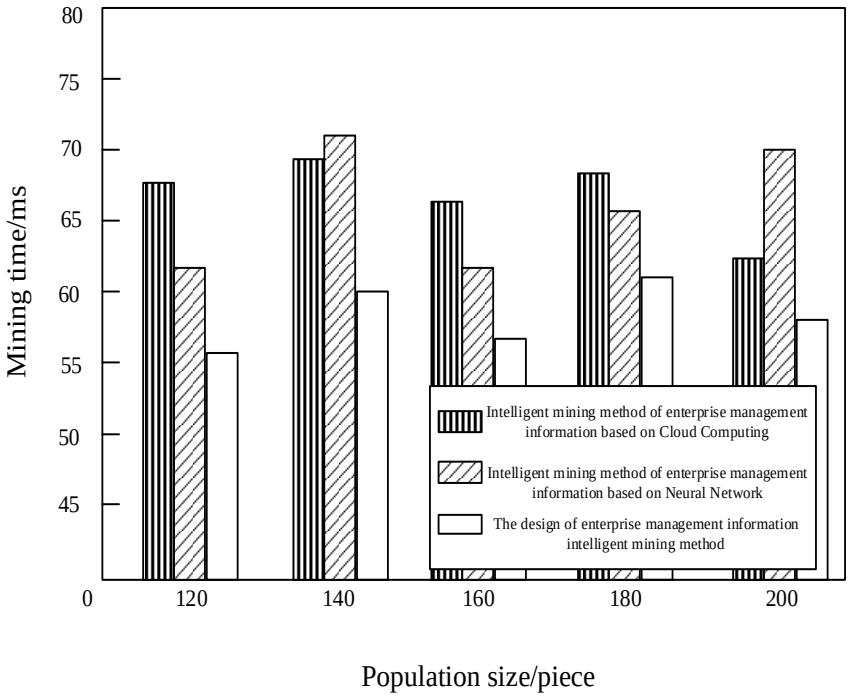

Figure 3 Mining time of population size 200 (ms)

It can be seen from Figure 2 that when the population size is 100, the mining time of the enterprise management information intelligent mining method designed this time and the other two intelligent mining methods are 21.306ms, 24.518ms and 26.559ms respectively; It can be seen from Figure 3 that when the population size is 200, the mining time of the enterprise management information intelligent mining method designed this time and the other two intelligent mining methods are 56.322ms, 65.218ms and 63.505ms respectively.

4. CONCLUSION

In this paper, the ID3 decision tree algorithm into the enterprise management information in intelligent mining method, according to different features of enterprise information in bom system divides the multiple themes, changed the way the original data management, can make the enterprise and material production, costs, sales and product related information integration, and for enterprises to accumulate and to more useful potential information. Due to the diversity of information intelligent mining tasks, there are still many information intelligent mining tasks that may not be fully competent. How to provide more intelligent mining models and make them more flexible and easy to use also needs further research.

REFERENCES

[1]YU Jian-Jun, Zhang Qiong-Zhi. Decision Tree ID3 Algorithm Based on Rough Set[J]. Computer Systems & Applications, 2020, 29(4): 156-162.

[2]Wang Li-jun. Optimization of Decision Tree ID3 Algorithm[J]. Journal of Heze University, 2020, 42(5): 15-19,30.

[3] Yang Y, Mei Z, Zheng B, et al. Design of Enterprise Management System Based on Edge Computing Architecture[J]. Mobile Information Systems, 2021, 2021(10):1-12.

[4] Rla B, Zxa B, Yh C, et al. Improving high-tech enterprise innovation in big data environment: A combinative view of internal and external governance[J]. International Journal of Information Management, 2020, 50:575-585.

[5] Evans N, Price J . Development of a holistic model for the management of an enterprise's information assets[J]. International Journal of Information Management, 2020, 54:102193.

[6] Almigheerbi T S, Ramsey D, Lamek A . Information Management in a Collaboratively- Developed Approach to Enterprise Resource Planning-A Higher Education Perspective[J]. Information (Switzerland), 2020, 11(3):146.

[7] Vasiliev V A, Velmakina Y V, Mayborodin A B, et al. Use of Information Technologies for the Integration of an Enterprise Quality Management System with the Requirements of the Related

Standards[J]. Russian Metallurgy (Metally), 2020, 2020(13):1644-1648.

[8] Luo J P, Yuan L L, Han L Q, et al. Research on Intelligent Mining Algorithm for Distribution Network Transformer Fault Early Warning[J]. IOP Conference Series: Earth and Environmental Science, 2021, 701(1):012014 (8pp).

[9] Si W, Y Men, Ge Z. Intelligent Mining and Parallel Processing Technology of Tunnel Construction Data under Complex Geological Conditions[J]. IOP Conference Series: Earth and Environmental Science, 2020, 580(1):012022 (5pp).

[10] J W at ada, Roy A, Vasant P. Preference Identification Based on Big Data Mining for Customer Responsibility Management[J]. International Journal of Intelligent Technologies and Applied Statistics, 2020, 13(1):1-24.

[11] D Zhao, Li Y, Dong Y, et al. Research on Experiential Solidification Learning Technology for Intelligent Mining of Power System Vulnerabilities[J]. Journal of Physics Conference Series, 2020, 1550:032060.

[12] Li W, Zhu J, Zhang Y, et al. Design and implementation of intelligent traffic and big data mining system based on internet of things[J]. Journal of Intelligent and Fuzzy Systems, 2020, 38(2):1-9.

[13] Mengash H A, Mahmoud H. Methodology for Detecting Strabismus through Video Analysis and Intelligent Mining Techniques[J]. Computers, Materials and Continua, 2021, 67(1):1013-1032.

[14] Shi bing. Complex Network Data Flow Frequent Itemset Artificial Intelligence Mining Simulation[J]. Computer Simulation, 2020, 37(4): 330-334.

[15] Li Yadong. Research on the Application of Decision Tree ID3 Algorithm in Employment Forecast of Vocational Graduates[J]. China Computer & Communication, 2020, 32(17): 54-56.