Accounting and Corporate Management (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/acccm.2023.050805
ISSN 2523-5788 Vol. 5 Num. 8

Research on Enterprise Intelligent Knowledge Management and Decision Making Based on Big Data Mining

Xiong Jian[1,2], Yang Feng[3]

[1]*Guizhou Industry Polytechnic College, Guiyang, 550008, China*
[2]*Jose Rizal University, Mandaluyong, Manila, Philippines, 0900, Philippines*
[3]*Lijiang Culture and Tourism College, Lijiang, Yunnan, 674100, China*

Keywords: Big data, Enterprise intelligence, Knowledge management, Decision making

Abstract: With the continuous development of the Internet, the Internet of Things and cloud computing, as well as the wide application of intelligent terminals, various forms of data explosion, the era of big data has arrived. Big data is critical to the growth of manufacturing and strategic assets and has huge potential to help revitalize the data industry and change the way people research and think. Big data mining refers to extracting valuable information for users from massive data with various data types. These data types change rapidly and the data density is low, making relevant services acceptable to users. Enterprise knowledge management uses big data mining to preserve basic knowledge. By filtering, selecting, extracting, saving, converting and using original knowledge, it can conduct intelligent processing according to the normative knowledge, experience and management intention required by the company. Providing the right knowledge to the right people at the right time in the right way can increase the feasibility of big data mining knowledge, reduce the chaos caused by information explosion, and effectively support the company to make effective management decisions and improve the level of knowledge management. Based on the above content, enterprises must introduce big data analysis technology in the process of knowledge management, which can better help enterprises to achieve the goal of sustainable development.

1. Introduction

1.1 Research Background

With the rapid development of communication and data acquisition technology and the wide application of intelligent mobile terminals, human life, organization and social activities are becoming more and more digital. For enterprises, they collect and store a large amount of data in production and business activities, such as production process monitoring data, customer order data, transaction data, market data, inventory records and so on[1]. These data reflect the enterprise's internal production and operation status and management status, called enterprise internal data. At the same time, the business activities of contemporary enterprises are increasingly affected by the

external environment, and this influence is increasingly strong with the extensive participation of social media and users. For example, users on weibo, forums and e-commerce platforms provide various comments and word-of mouth on products and enterprises, which constitute the external data of enterprises[2]. These data from various channels, ways and forms related to enterprise production, operation, sales and management are called multi-source enterprise operation data or enterprise multi-source operation data. Analysis and management knowledge mining based on enterprise multi-source operation data plays an important role in business understanding, performance measurement, customer relationship improvement and business opportunities creation. For example, the management knowledge mined from data analysis can help enterprises identify key factors affecting product quality, summarize the causal relationship between product quality and production process, and thus avoid potential risks. It can help enterprises understand customer needs, strengthen customer loyalty, maintain a good relationship with customers, and then solve problems before they turn into crises.

1.2 Problem Statement

(1) Dimensional disaster problem: But with real-time monitoring technology such as rfid and bar code in the extensive use in the process of production, acquisition of granularity of data related to the manufacturing process more meticulous, the dimension increases rapidly, and the dimension between the coupling effect, the relationship is complex, that for the data analysis and management of existing knowledge mining method bring huge difficulties, seriously affect the effectiveness and efficiency of analysis work; (2) large volume and a lot of redundant data, class imbalance, update problem not in time: At the same time, a large amount of collected data occupies a large amount of storage space and increases the storage cost of enterprises. Moreover, the update cost of large volume data is also very high, but the backward management knowledge and decision-making errors caused by slow data update will certainly greatly reduce the competitiveness of enterprises[3]. Finally, there is a serious problem of category imbalance in the data related to product quality generated in the production process of enterprises, that is, the data related to unqualified products that managers care about most account for a very small proportion, and it is difficult to analyze the causes; (3) The black-box method leads to reliable results, but the model mechanism analysis is insufficient, and it is difficult for managers to understand and use. Many existing data analysis model are black box model, the goal is accuracy, namely the high resolution of the model, the result is reliable, but some mechanism analysis model, and interpretability don't meet the needs of management, or to management problem, compared with the precision of the model, the management is more concerned about the problem of background knowledge, That is, the refinement and aggregation of management rules and knowledge[4].

1.3 Research Objectives

In the context of the network era, enterprise operation data has a variety of special properties, such as diversity of sources, high dimension, massive, not timely update, category imbalance and multi-marking, so that enterprise management workers are generally faced with the problem of “sufficient data but lack of knowledge”. Therefore, in view of the special attributes of enterprise operation data, the research on the processing and management knowledge mining method for enterprise multi-source operation data is not only a deepening of data mining technology research, but also provides a powerful information processing tool to promote the paradigm shift of enterprise management from “experience-driven” to “data-driven”[5].

1.4 Research Questions

Enterprise management knowledge mining method of multi-source operational data research, mainly for data back down times, the characteristics of the enterprise operating data and existing data analysis and the shortage of management knowledge mining method, brings to the managers of the problem of lack of sufficient data and knowledge, put forward the systematic solution, and application of the scheme, respectively, in view of the enterprise's internal and external data analysis, Extract the corresponding management rules and knowledge, and then assist managers to carry out scientific prediction, evaluation and decision-making activities[6-7].

1.5 Research Hypothesis

This section mainly uses the correlation between conditional factors to establish a dimensionality reduction algorithm. The basic idea is: when mining causality in enterprise operation data, different conditional factors have different effects on the result factors, so the influential conditional factors should be considered first[8]. At the same time, the correlation between the conditional factors should be considered, so as to make the conditional factors irrelevant. If conditional factor A determines the state of b to some extent, then conditional factor B should be deleted so that the result of deletion can ensure that the conditional factors participating in causal induction are as irrelevant as possible[9]. To solve this problem, firstly, we introduce the information of determining degree of conditional factors to result factors, and rank the importance degree of conditional factors according to this information. Second, calculating the mutual decision of condition factors between information, and according to the characteristics of the information filtering, realize the purpose of dimension reduction, said that the enterprise data dimension reduction algorithm based on factors determine the degree of conditional factors in dimension reduction algorithm (Reduction of Conditional Factors, RCF).

2. Objectives

It is generally believed that the creation, acquisition, processing, storage, dissemination and application of knowledge in the process of knowledge management are generally defined in a narrow sense, while the integration of knowledge related information is generally defined in a broad sense. Knowledge management can realize the transformation of explicit knowledge and tacit knowledge inside and outside the company, and use the team wisdom of external employees and customers to improve the innovation ability and decision-making level of the company. Comprehensive and systematic knowledge management system, on the one hand, can improve the enthusiasm for work; On the other hand, they can organize the training and education of employees to keep explicit knowledge in the company, so as to improve the retention and flexibility of the company's knowledge. Knowledge management can help companies shape internal mechanisms for continuous learning. When a company uses it, it can contain and store displayed knowledge and extract it as it extracts it. It can also reduce the economic losses caused by poor labor relations and the resulting lay-offs [10]. It is a control mechanism. It is knowledge-based and its main role is to help companies shape their core knowledge so that they can share innovation, update and business knowledge. It can improve the quality of business services by providing an environment for employees to share knowledge, improve overall performance and innovate [11]. It continues to create value throughout a company's growth, contributing to its adaptability and resilience, and driving business innovation and sustainability. Big data mining refers to mining valuable information for users from massive data with multiple data types, rapid changes and low data value density to provide and serve users. Big compared with the traditional data mining, data mining technology development background, facing the breadth of data environment and the mining depth, large data mining relies on cloud computing, Internet, mobile

Internet, mobile intelligent terminal technology of booming development, centered around the characteristics of big data, the paper analyses the problems existing data mining system, with the help of advanced technology, Real-time processing and multi-dimensional analysis of complex data, processing data in a wider range, mining and analysis more in-depth and comprehensive. Intelligent knowledge management discusses how to make use of possible technical means to build a set of systematic programs, so that the massive data can provide intelligent personalized knowledge services, improve the ability of enterprises to use information, and ultimately improve the decision-making level of decision makers. The traditional knowledge management from aspects of the algorithms, structure, process, this paper studies the object is "the knowledge management system", do not attach importance to the intelligent management of knowledge itself, without the system of knowledge management research, also does not have the method of using knowledge integration design, so that when the amount of knowledge is too large, the traditional knowledge management platform, it will be difficult to finish the task. Therefore, knowledge itself also has to have a certain degree of intelligence, self-management[12-13].

3. Methodology: is It Quantitative or Qualitative Study

3.1 Build the Big Data Mining Architecture Based on Cloud Computing

Cloud computing, as the basic support service of big data technology, is a new virtualization computing mode with high scalability and high elasticity. It is mainly applied in data storage and data processing[14]. These features provide basic storage services for big data and efficient data processing power. The architecture of big data mining combines distributed storage and distributed parallel computing of cloud computing, presenting the structure pattern of cloud computing/client. Combined with cloud computing, big data mining using cloud computing services provided by the support, and huge storage capacity with powerful computing ability as the core, the integration of a variety of large, analysis, data mining technology and mining, analysis algorithm, the cloud to form a new mining, digging the cloud to realize the real-time analysis and exploration of big data, and results of the analysis and mining in visualization technology is presented to the customer[15].

3.2 Build a Big Data Mining Platform Based on Hadoop

We can build a Hadoop big data mining platform integrating various technologies and functions, which is divided into data source, big data mining platform and user display layer. The data source can be structured data, semi-structured data, or even unstructured data; Big data mining platform is a real-time data analysis and processing platform based on Hadoop[16]. The platform can analyze and mine data content by using different computing modes and methods. The user display layer presents data results to the user or accepts a request service using a visual technique. By building a Big data mining platform based on Hadoop, big data can be collected, stored, calculated, screened, analyzed and displayed.

3.3. Study the existing knowledge management system structure, establish the theoretical model of intelligent knowledge management, design and implement the intelligent knowledge management system based on big data mining

Study the existing knowledge management architecture and find the source of knowledge. The original knowledge comes from the results of big data mining. Combining with the concept of intelligent knowledge management, the original knowledge has the characteristics of intelligence, so that it can update itself and find the deep knowledge from it, so as to gradually build the theoretical

model of intelligent knowledge management[17]. Finally, based on the theoretical model, the intelligent knowledge management system based on big data mining is designed using computer technology, so that it can really serve users.

4. Discussion

In order to fully test the effectiveness of CMBPC and UABP algorithms, nine data sets are selected from UCI database that are widely used for comparison of data volume reduction algorithms. Table 1 shows the statistical characteristics of each data set. Among them, the number of samples is at least 699 and at most 10,000, the range of dimensions is from 4 to 28, and the range of the number of categories is from 2 to 29. The selected data sets are rich in types.

Table 1: Experimental Data Set of Data Volume Reduction

The data set	Sample	Conditional factor	class
cell	699	9	2
messidor	1 151	19	2
banknote	1 372	4	2
hcv	1 385	28	4
contraceptive	1 473	9	3
segmentation	2 310	19	7
abalone	4 177	8	29
page	5 473	10	5
electrical	10 000	13	2

Table 2: Comparison of Experimental Results of Volume Reduction Algorithms

Sample set		Baseline	SRSWOR	SRSWR	SS	CS	STS	CMBPC
messidor	accuracy	0.548	0.532	0.549	0.541	0.541	0.544	0.610
	Sampling rate	1.000	0.900	0.800	0.700	0.800	0.800	0.474
segmentation	accuracy	0.829	0.818	0.780	0.799	0.809	0.799	0.857
	Sampling rate	1.000	0.900	0.700	0.900	0.900	0.900	0.771
cell	accuracy	0.950	0.940	0.945	0.940	0.925	0.940	0.970
	Sampling rate	1.000	0.800	0.800	0.900	0.800	0.700	0.574
hcv	accuracy	0.263	0.235	0.246	0.248	0.238	0.236	0.279
	Sampling rate	1.000	0.800	0.600	0.700	0.700	0.900	0.853
abalone	accuracy	0.205	0.187	0.183	0.183	0.187	0.179	0.216
	Sampling rate	1.000	0.800	0.800	0.700	0.900	0.900	0.774
banknote	accuracy	0.972	0.960	0.957	0.957	0.957	0.957	0.962
	Sampling rate	1.000	0.900	0.700	0.700	0.700	0.800	0.745
electrical	accuracy	1.000	0.942	0.939	0.950	0.947	0.952	1.000
	Sampling rate	1.000	0.800	0.700	0.900	0.800	0.700	0.897
page	accuracy	0.965	0.948	0.945	0.942	0.944	0.944	0.965
	Sampling rate	1.000	0.800	0.900	0.700	0.800	0.900	0.536
contraceptive	accuracy	0.474	0.477	0.476	0.477	0.484	0.481	0.467
	Sampling rate	1.000	0.900	0.700	0.800	0.900	0.900	0.640

The CMBPC was compared with 5 classical volume reduction algorithms, including: Simple random sampling without (SRSWOR), simple random sampling with (SRSWR), SystematicSampling (SS), ClusterSampling (CS), and STratifiedSampling (STS). In order to avoid the influence of randomness on the experimental results and to be fair, the cross-validation method was adopted in the comparative experiment. For each data set, the training set extracted each time adopts the above six algorithms for volume reduction processing, that is, the original data set used for all the volume reduction algorithms is the same. Then, the data sets after volume reduction and the

original data sets were used as new training samples to establish the classification model (in the experiment, all the data sets after volume reduction and the original data sets used the CART tree as the base classifier, and all used the same parameters). Finally, the same test set is used to test the algorithm performance, and the average of 10 classification accuracy is used as the test standard. The comparative experimental results are shown in Table 2. The Baseline is the classification result of the original data (without volume reduction) by using the base classifier, and the optimal result under each data set is represented in bold. Sampling rate refers to the ratio of the amount of data after volume reduction to the amount of original data, corresponding to the sampling rate of each algorithm under the optimal accuracy[18].

In order to measure the effect of the algorithm more intuitively, combined with the data in Table 2, the bar chart is used to compare the Sampling rate and classification accuracy of each data set before volume reduction and after volume reduction using the six algorithms. The results are shown in Figure 1: (1) On the messidor, segmentation, cell, HCV, and Abalone datasets, CMBPC not only performs effective volume reduction, but also has greater classification accuracy than the full data, indicating that the proposed algorithm effectively removes redundant data and improves the quality of the dataset[19]. (2) Accuracy of other data sets is basically the same as that of the full sample, but only 74%, 90%, 54% and 64% of the full sample are needed, indicating that CMBPC can effectively compress the data volume used for analysis and mining methods, reduce the computational complexity and improve the analysis efficiency under the principle of maintaining or approaching the full sample effect. (3) In all 9 data sets, CMBPC accuracy ranked first for 7 times and second for 1 time, indicating that the algorithm has certain robustness to data types [20].

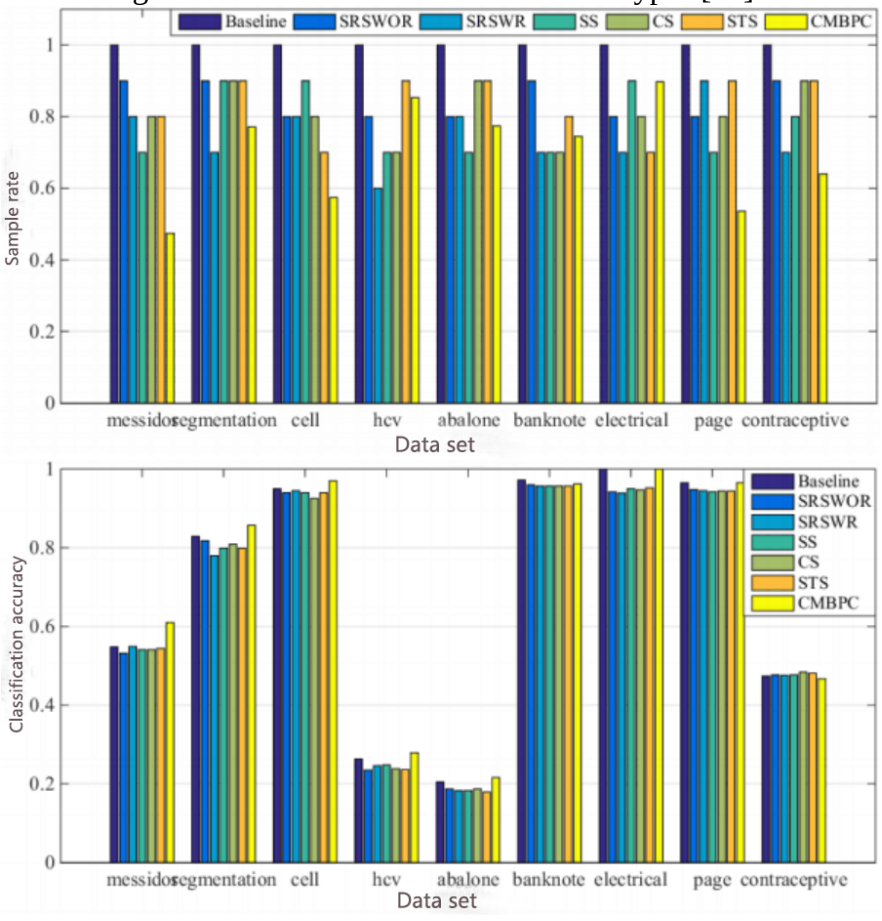

Figure 1: The results the six algorithms

5. Conclusion

To solve the problem of high-dimensional data, a dimensionality reduction algorithm of enterprise data based on factor determination is established. First of all, the model assumes that the expansion of the enterprise data dimension is interaction relationship between the dimensions of rights, led to the inevitable exist a large number of redundant factors, and determine the coupling relationship between degrees can depict factors, so it can use the redundancy factor reduction, based on the analysis of enterprise data dimension reduction algorithm based on factors determine the degree of RCF. Secondly, in order to test the effectiveness of RCF, comparative analysis is conducted on six high-dimensional data sets widely used for algorithm comparison with six classical dimensionality reduction algorithms. Finally, the experimental results show that: (1) RCF can effectively reduce redundant factors, improve enterprise data quality, and improve the efficiency and effectiveness of late classification method by considering the correlation between conditional factors and result factors as well as the coupling between conditional factors. (2) Compared with the dimensionality reduction algorithm based on feature transformation, the dimensionality reduction process of RCF can retain the actual meaning of the original dimension, which is easy for enterprise managers to understand and use; (3) Effectively reducing the number of conditional factors can enhance the availability and interpretability of management rules and knowledge in later mining, and facilitate enterprise managers to conduct causal induction analysis.

Acknowledgment

Philosophy and Social Sciences Project of Yunnan Province,Study on the path and countermeasu res of Lijiang's active integration into RCEP to suggest the core competitiveness of tourism service t rade, NO. PY202210.

References

[1] Fang J, Peng C. Research on Enterprise Intelligent Manufacturing Capability Evaluation Based on Maturity Model [J]. Francis Academic Press, 2020(9).
[2] Wang X, Xu X, Zhang J, et al. Research on Intelligent Construction Algorithm of Subject Knowledge Thesaurus Based on Literature Resources [J]. Journal of Physics: Conference Series, 2021, 1955(1):012038 (10pp).
[3] Zeng H. Scenario Design of Intelligent SPOC Knowledge Farm Based on Flexible Learning. 2020.
[4] Mahiddin N B, Othman Z A, Bakar A A, et al. An Interrelated Decision-Making Model for an Intelligent Decision Support System in Healthcare [J]. IEEE Access, 2022(10):10.
[5] Corrales-Garay D, Ortiz-De-Urbina-Criado M, Mora-Valentin E M. Understanding open data business models from innovation and knowledge management perspectives[J]. Business process management journal: Developing re-engineering towards integrated process management, 2022(2):28.
[6] Alkadhimi A, Cunningham S J, Parekh S, et al. Decision making regarding management of compromised first permanent molars in patients with molar incisor hypomineralisation: A comparison of orthodontists and paediatric dentists [J]. Journal of orthodontics, 2022(1):49.
[7] Wlabc D, Szabc D, Ywabc D, et al. Research on intelligent damage detection of far-sea cage based on machine vision and deep learning. 2021.
[8] Peng F, Xiong W, Xia D, et al. Research on Intelligent Load Forecasting Method for Electric Power Users Based on Data State Drive[J]. Electric Engineering, 2020.
[9] Liang J. Analysis of Intelligent English Learning System Based on Intelligent Data Collection Technology[C]// International Conference On Signal And Information Processing, Networking And Computers. Springer, Singapore, 2022.
[10] Tokarev K, Lebed N, Prokofiev P, et al. Monitoring and Intelligent Management of Agrophytocenosis Productivity Based on Deep Neural Network Algorithms[C]// International Conference on Intelligent Computing & Optimization. Springer, Cham, 2023.
[11] Ding P, Zheng Q, Shen Y, et al. Research on Cloud-Edge Collaboration Architecture for Intelligent Acquisition of Digital City Information Based on 5G Customized Network [J].2022.

[12] Zhang X, Li P, Wu W, et al. Research on Interactive Modeling and Intelligent Analysis Platform Based on Railway Big Data[J]. DEStech Publications, 2020.
[13] Allam S. Research on intelligent medical big data system based on Hadoop and blockchain [J]. JETIR (www.jetir.org), 2021(4).
[14] Mian T S, Ghabban F. Competitive Advantage: A Study of Saudi SMEs to Adopt Data Mining for Effective Decision Making [J]. Data analysis and information processing, 2022, 10(3):15.
[15] Salahat M A. Employee Empowerment, Knowledge Management and Decision-making Agility; Mediating Role of Extra-role Performance [J]. Human Resources Management Academic Research Society (HRMARS), 2021(7).
[16] Ms M. The Impact of Decision-Making Models and Knowledge Management Practices on Performance [J]. Allied Business Academies, 2021.
[17] Mukhtar M, Yamin M. The Influence of Academic Culture, Management Knowledge and Interpersonal Communication on Decision Making by the Head of Private Islamic Colleges in Jambi Province [J].2020(2).
[18] Xia Y. Big data based research on the management system framework of ideological and political education in colleges and universities [J]. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021(2):40.
[19] Li M. Research on Financial Management Algorithm Based on Machine Learning in Big Data Era. 2021.
[20] Li J. Research on the Construction of Intelligent Employment Platform for College Graduates Based on Big Data[J]. DEStech Publications, 2020.