

# Application of Big Data in Industry and Commerce Based on Data Mining Algorithm

Haifeng Liao

{lhd8898@yeah.net}

Dean's Office

Guangdong University of Science and Technology, Dongguan, Guangdong, China

**Abstract.** Data collection is the main content industrial and commercial big data, but in the process of big data processing, the quality of data and the source of data will be processed. The result is an impact and reduces the accuracy of big data for business and industry. To improve the effectiveness of data, The data mining algorithm to mine and analyze industrial and commercial big data, improve the development level of industrial and commercial big data, and shorten it data decision making time. Then, a comprehensive evaluation of the big data of industry and commerce is carried out. The continuous monitoring can analyze big data processing results. The data mining algorithm can accurately mine industrial and commercial big data, take the level of industrial and commercial big data, and the rationality is greater, and is better than traditional data processing methods. Therefore, the data mining algorithm can meet the requirements of big data collation, the big data algorithm will optimize the level and management indicators of business administration, and is suitable for the continuous analysis of the development.

**Keywords:** Data Collection, Big Data for Industry and Commerce, Data Mining Algorithms, Traditional Data Processing, Data Decisions

## 1. Introduction

The data analysis is a comprehensive evaluation method of industrial and commercial big data [1], and to choice a comprehensive evaluation of market and production data [2], and unreasonable analysis [3]. At present, industrial and commercial big data has the low accuracy [4], long actual deviation time[5].To apply intelligent algorithms in industrial and commercial big data and identify main factors for better decision-making analysis [6]. The data acquisition and data accuracy is still low. The data mining algorithms to analyze data accuracy, data collection and data diversity, and make decisions on data processing develop and conduct time series analysis to improve the accuracy of business big data [7-9]. the key values in industrial and commercial big data and verifies the decision-making results. In view of the above problems, this paper studies the following contents, mainly in three aspects. On the one hand, it collects the data of business administration and processes the data standard. Secondly, comparing different intelligent algorithms, we find the advantages of big data algorithms, and describe business administration problems mathematically. Finally, the simulation

method proves that the big data algorithm is reasonable and the evaluation index of business administration is effective.

## 2. Big data Analysis for Business

Industrial and commercial big data analysis makes reasonable decisions on data collection, data diversity and data accuracy [10], and detects the data, and can effectively process data [11]. Big data analysis of industry and commerce the data indicators [12-14]. The data mining algorithm uses integrated information theory to mine and analyze industrial and commercial big data where the direction of change, which are as follows.

Definition 1: the business data is  $x_i$ , the mining formulation function is  $w(y_i)$ , the data set is  $z_i$ , and the duration is  $l_i$ . Well, the process is  $w(y_i)$  shown in Equation (1).

$$w(y_i) = z_i \cdot l_i \cdot x_i \quad \square\square\square\square(1)$$

Definition 2: The mining function is  $h(x)$  decision-making rationality is  $P$ , data mining rationality, is  $Y_z$  and data mining is  $Y_t$  accuracy. the calculation process is  $h(x)$  as in Equation (2).

$$h(x \cdot P) = x \xrightarrow{y} \sqrt{Y_z \cup Y_t} \quad \square\square\square(2)$$

Definition 3: the result function is  $v(x_i)$ , mining change to is  $\bar{x}_i$ , mining set is  $ste_i$ , mining number is  $h_i$ . Well, the reasult is  $v(x_i)$  shown in Equation (3).

$$v(x_i) = \sum ste_i \div h_i \cdot x_i \quad (3)$$

Definition 4: The data mining function is  $f(x, b)$ , is  $w$  the threshold of mining and is  $\xi$  the decision rate. Well, the calculation process is  $f(x)$  shown in Equation (4).

$$f(x) = \sum_{i=1} x_i \times w \div \zeta_i \quad (4)$$

Among them,  $\zeta_i$  is the external factor of business administration, which mainly adjusts the error of calculation results.

### 3. Decision Making by Data Mining Algorithms

In the process, the result of big data information should to reduce the error rate [16]. By the theory of integrated information, and identify active and conservative decisions with differences and calculate the rationality of decision-making. So to conduct random analysis of data is collection, data accuracy and data diversity under different information conditions.

Definition 5: A data mining function is  $m(x_i \cdot y_i)$  a reasonable decision when a spike occurs is  $k(x_i, y_i) \in [0, 1]$ . The calculation is  $m(x_i \cdot y_i)$  shown in Equation (5).

$$m(x_i \cdot y_i) = \frac{k(x_i^2 \cdot y_i^2) \cdot (1 - P_i)}{Ya \cdot \sqrt{1 - k^2}} \quad (5)$$

Formula: If  $m(x_i) \mapsto A$  the data mining results are the data content and data accuracy should be computer. If  $m(x_i) \not\mapsto A$  the data mining results low, and not meet the market demand, adjust the data content.

Definition 6: The data mining judgment function is  $y(e)$  calculated as shown in Equation (6).

$$y(e) = y(x_i, y_i) \cdot e_i \cdot f(x_i) \quad (6)$$

### 4. Steps of Data Mining Techniques

the invalid information meet the data mining requires, which is sampling analysis of data collection, data accuracy and data diversity, including data cleaning and data transformation, as well as analysis of information extraction. In addition, the decision-making scheme is analyzed by the data mining algorithm, and the different analysis and different information content is calculated. The interference analysis is to eliminate the influence of different information on data mining results [17-18].

The first step is to collect information data, evaluate the information data, and then determine the conditions.

Step 2: Mining and calculation of data collection, data accuracy and data diversity, and continuous analysis of industrial and commercial big data.

Step 3: in different decisions, verify the rationality, and accuracy of information processing, and take the mining results in set.

Step 4: Mining calculations are performed on decision-making, and if constraints are exceeded, or information data changes, the analysis is terminated, otherwise mining calculations.

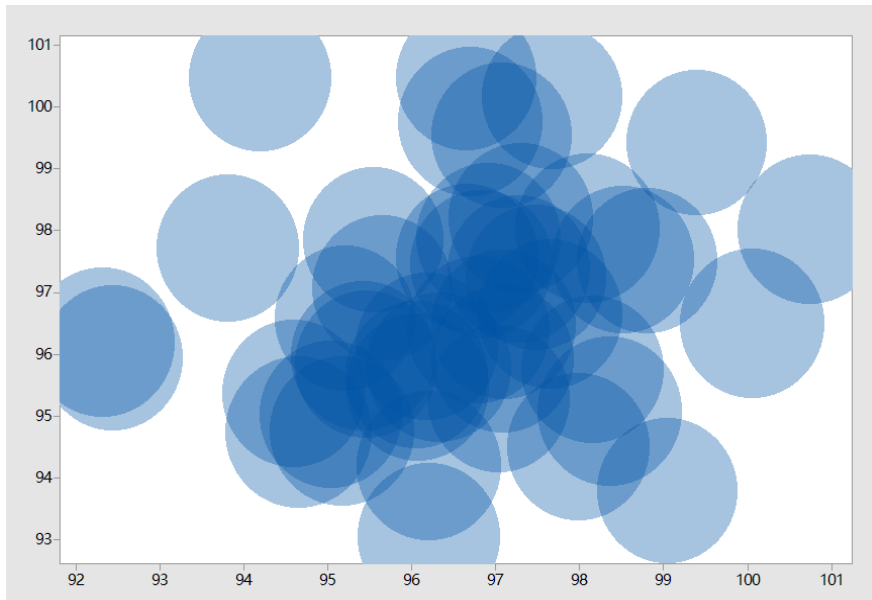
## 5. Practical Cases of Big Data in Industry and Commerce

### 5.1 Data Collection Briefing

the rationality of data mining is analyzed is shown in Table 1.

**Table 1.** Status of Mining Calculation Data (unit: %)

Computer parameter	Comprehensive	rationality
comput acquisition	95.21	96.30
Data preprocessing	95.82	97.10
Feature extraction	96.68	95.74
Data cleansing	95.20	95.41
Data integration	97.74	95.72
Data transformation	96.64	96.07



**Fig.1** Distribution status of mining calculation data (The data comes from questionnaires and practical investigations)

In the data analysis in Table 1, there is no difference in data range, iteration relationship, coincidence degree and distribution direction, which shows that the collected data meets the requirements of mining analysis as a whole and can provide support for data mining. The overall data mining process is shown in Figure 1.

The data changes in Figure 1 show that the data dispersion, data independence, data acquisition accuracy and calculation process can meet the actual requirements, indicating that the data integrity is good.

### 5.2 Comprehensiveness and Rationality of Mining Calculations

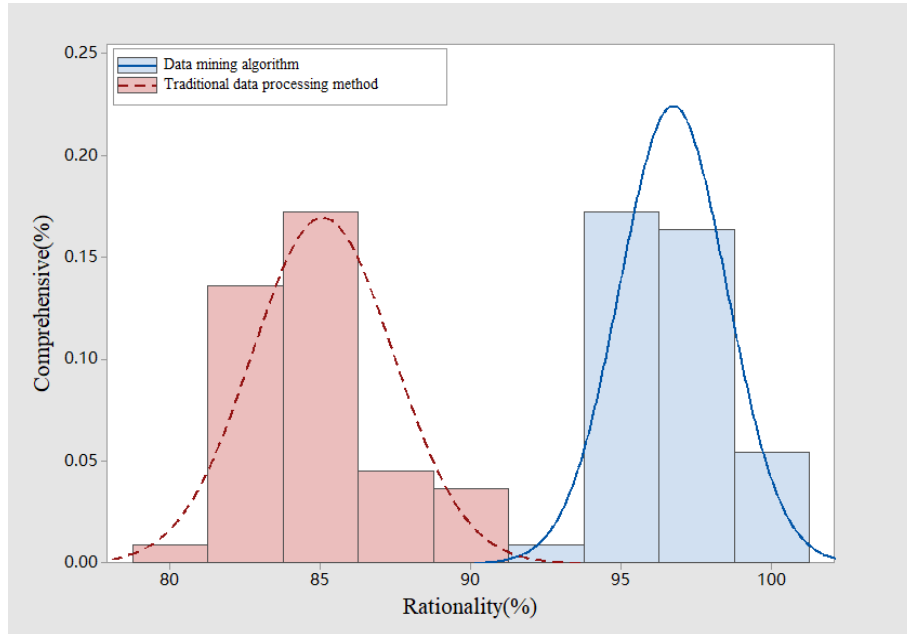
The mining calculation should be done complete, and it will reduce the mining calculation results, and the comprehensiveness and rationality of the mining calculation results will be tested, and the specific decisions are shown in Table 2.

**Table 2.** Comparative Decision-making on Comprehensiveness and Rationality (unit: %)

algorithm	Decision-making direction	parameter	rationality	Comprehensive	Average magnitude of change
Data mining algorithms	Proactive decision-making	Association analysis	97.42	97.01	0.41
		Sequence analysis	97.35	97.66	0.31
		Categorical analysis	96.55	98.13	2.32
	Conservative decision-making	Association analysis	95.70	96.18	1.75
		Sequence analysis	98.74	98.69	1.25
		Categorical analysis	99.73	99.11	0.72
Traditional data processing methods	Proactive decision-making	Association analysis	81.12	88.31	6.36
		Sequence analysis	83.32	86.19	3.53
		Categorical analysis	82.42	84.01	2.15
	Conservative decision-making	Association analysis	87.35	82.66	5.09
		Sequence analysis	86.55	88.13	2.32
		Categorical analysis	85.70	82.18	3.75
Comparison of decisions		X <sup>2</sup> =9.211, P<0.03			

In the data comparison results in Table 2, the data mining algorithm has been fully confirmed, the overall rationality is 98%, the change range is 2.4%, and the difference between different methods is 5.6%. This result shows that the decision content, decision result and decision change are reasonable and meet the preliminary requirements of data mining. However, there are some differences between data mining algorithm and traditional algorithm in the results, mainly because the difference of comprehensive results is more than 5%, which shows that the research results of this paper are ideal.

The reasonableness has not reached 90%. In the case of complete market, the calculation results of data mining algorithm are more comprehensive and reasonable, and the overall change trend is shown in Figure 2.



**Fig.2** Comparison of the comprehensiveness and rationality of different algorithms

the rationality and comprehensiveness of the data mining algorithm is more concentrated to reach 97%, while the traditional data processing method is the same. The rationality and comprehensiveness can only reach 85%, which is consistent with the results of the study in Table 2. The reason is that data mining algorithms mine data collection, data accuracy, data diversity making, decision-making, calculate the convex function values of different values, simplify the decision-making.

Through the water comparison in Figure 2, we can see that the data mining algorithm can realize continuous analysis, and the analysis results are in various ups and downs. However, the fluctuation of calculation deviation is equal and stable, and there is no abnormal mutation value, which shows that data mining algorithm can deal with data more ideally and make data change meet the requirements.

### 5.3 Mining and Decision-making Time of Industrial and Commercial Big Data

Mining calculation time is decision-making effect, including: association analysis, classification analysis, etc., the specific decision-making is shown in Table 2.

**Table 3.** Data Mining Calculation Time (unit: seconds).

method	Calculation time in seconds	Proactive decision-making			Conservative decision-making		
		Association analysis	Categorical analysis	Sequence analysis	Association analysis	Categorical analysis	Sequence analysis
Data mining algorithm	10~20 20~60 60~100	96.11 96.36 95.91	96.63 97.85 96.59	95.20 96.74 97.01	97.86 97.56 96.88	97.11 96.36 95.91	96.63 96.85 96.59

The degree of clustering	1.00~3.15						
Traditional data processing methods	10~20	87.52	83.08	87.72	84.17	87.52	83.08
The degree of clustering	20~60	89.44	89.28	80.82	87.04	89.44	89.28
The degree of clustering	60~100	89.91	88.65	81.02	85.32	89.91	88.65
The degree of clustering	4.33~7.41						

From the result in Table 2, the data mining algorithm can short the time of mining calculation, the time is relatively stable. Among them, the information content mining calculation of association analysis, classification analysis, and sequence analysis in mining calculation. The final time of the proposed method is shorter than that of traditional data processing, and the average time is less than 4 seconds, so it can be accepted. So the data mining algorithm is based on industrial and commercial big data, and iteratively analyzes and determines the association analysis, classification analysis and sequence analysis mining calculation time. The redundant market variation of traditional data processing methods has increased. The mining calculation time is shown in Figure 3.

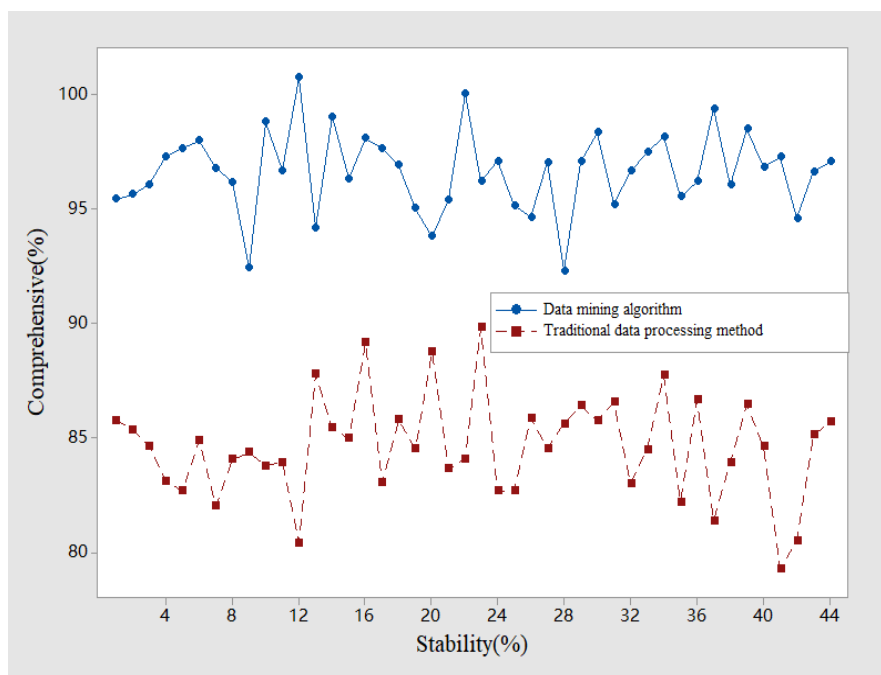


Fig.3 Comparison of decision clustering degree of different methods

the data analysis in Figure 3, we can be seen that the decision clustering degree of the data mining algorithm is high and the data change is stable, while the decision-making of the traditional data processing method is poor and shows large fluctuations.

## 6. Conclusion

In the data collection process, the former intelligent methods have some problems, such as unsatisfactory decision-making effect and deviation of decision-making results. In order to improve efficiency, the data mining algorithm to evaluate association analysis and sequence analysis to determine the final data processing decision. The comprehensive calculation of decision expectations is carried out through data mining algorithms to reduce the data of the decision-making. Mining calculations are carried out with constraints, and data collection, data accuracy and data diversity are formulated to improve reasonableness of decisions. The data mining algorithm has a high comprehensiveness, the rationality of the overall calculation is more than 89%, and the change range of some data is less than 4%, which is in sharp contrast with the previous calculation results, the changes in data accuracy, data collection and data diversity are relatively small, so the total decision of the data mining algorithm is better. The rationality and comprehensiveness in the traditional data processing method is low, the result is between 4~7, and the data accuracy data collection and data diversity are less than 90%, which is same. Among them, the information content mining calculation of association analysis, classification analysis, the calculation time of sampling samples is 1 ~ 2 seconds, and the calculation time of continuous sampling is 3 ~ 5 seconds, and the overall calculation time meets the requirements. There are some shortcomings in this research, which are mainly reflected in data collection. Because the data of business administration is less open and some data are qualitative data, the quantitative analysis effect of data is poor. In the future, quantitative data will be collected or qualitative data will be verified twice to improve the completeness of data.

## References

- [1] Alharbi, R., et al. (2022). "Prediction of Oil Production through Linear Regression Model and Big Data Tools." *International Journal of Advanced Computer Science and Applications* 13(12): 380-387.
- [2] Berawi, M. A., et al. (2020). "Impact of rail transit station proximity to commercial property prices: utilizing big data in urban real estate." *Journal of Big Data* 7(1).
- [3] Bifulco, L. and S. Bory (2020). "The redefinition of soccer plots. Big data application in football's narratives, performance, prediction, and evaluation." *H-Ermes-Journal of Communication* 17: 73-95.
- [4] Biggio, L., et al. (2021). "Uncertainty-Aware Prognosis via Deep Gaussian Process." *Ieee Access* 9: 123517-123527.
- [5] Caltanissetta, F., et al. (2022). "In-situ monitoring of Material Extrusion processes via thermal videoimaging with application to Big Area Additive Manufacturing (BAAM)." *Additive Manufacturing* 58.
- [6] Conde, J., et al. (2022). "Applying digital twins for the management of information in turnaround event operations in commercial airports." *Advanced Engineering Informatics* 54.
- [7] Deppner, J., et al. (2023). "Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach." *Journal of Real Estate Finance and Economics*.
- [8] Dragone, G., et al. (2020). "Innovation and strategic orientations for the development of advanced biorefineries." *Bioresource Technology* 302.



- [9] Holmi, J. T. and H. Lipsanen (2022). "WITio: A MATLAB data evaluation toolbox to script broader insights into big data from WITec microscopes." *Software* 18.
- [10] Joo, C., et al. (2022). "Development of physical property prediction models for polypropylene composites with optimizing random forest hyperparameters." *International Journal of Intelligent Systems* 37(6): 3625-3653.
- [11] Roberto Diaz-Reza, J., et al. (2020). "Design, process and commercial benefits gained from AMT." *Journal of Manufacturing Technology Management* 31(2): 330-352.
- [12] Santana, K., et al. (2021). "Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products." *Frontiers in Chemistry* 9.
- [13] Senova, A. and L. Sebescakova (2020). "EUROINTEGRATION TENDENCIES IN IMPLEMENTATION OF 'INDUSTRY 4.0' IN EUROPEAN COMPANIES." *Management-Journal of Contemporary Management Issues* 25(2): 251-261.
- [14] Truby, J. and R. Brown (2021). "Human digital thought clones: the Holy Grail of artificial intelligence for big data." *Information & Communications Technology Law* 30(2): 140-168.
- [15] Uz, D. and S. Buck (2020). "Comparing Water Use Forecasting Model Selection Criteria: The Case of Commercial, Institutional, and Industrial Sector in Southern California." *Sustainability* 12(10).
- [16] Vallabh, C. K. P. and X. Zhao (2022). "Melt pool temperature measurement and monitoring during laser powder bed fusion based additive manufacturing via single-camera two-wavelength imaging pyrometry (STWIP)." *Journal of Manufacturing Processes* 79: 486-500.
- [17] Vasina, M., et al. (2023). "In-depth analysis of biocatalysts by microfluidics: An emerging source of data for machine learning." *Biotechnology Advances* 66.
- [18] Verma, C. and P. K. Suri (2021). "'Big data' patentometrics for R&D decision-making." *Digital Policy Regulation and Governance* 23(4): 317-336.