Dynamic Data Mining Method of Cold Chain Logistics in Drug Distribution Under the Background of Cloud Computing

Meng-li Ruan(✉)

Shandong Management University, Jinan 250357, China
ruanmengli988@163.com

Abstract. Because of the huge volume of cold chain logistics data, the traditional data dynamic mining method can not mine the whole local drug circulation data, resulting in the lack of a large number of data mining results, reducing the integrity of the data. Therefore, in the context of cloud computing, a new dynamic mining method is proposed for the cold chain logistics data of drug circulation. Under the cold chain logistics model, the method is further developed by defining the drug circulation mode. The data mining uses cloud computing technology to extract the target data; uses data cleaning, data elimination, data supplement and data conversion to preprocess the target data; according to the association rules between the acquired data, realizes the dynamic mining of cold chain logistics data information. Experiments show that, compared with the traditional methods, the proposed mining method can find the target data in the huge cold chain logistics data, and achieve all the data mining. It can be seen that the data mining method proposed in this paper has higher data integrity.

Keywords: Cloud computing · Drug circulation · Cold chain logistics data · Dynamic mining

1 Introduction

For a long time, high-performance computing (HPC), as an important topic in the field of computer research, has been concerned. Since the 1990 s, HPC has grown from 1 billion floating-point operations per second to 10 billion floating-point operations per second. However, with the increase of computing speed, the cost of supercomputers has also increased, and its high cost makes people rarely able to afford [1]. However, cloud computing technology has changed this situation, people began to use parallel and distributed computing technology to build high-throughput cloud computing system. The development style of high-end computing system is changing, and the computing strategy has changed from HPC to HTC, that is, to high-throughput computing. Generally, in order to make the results of data mining more accurate, people tend to choose larger data sets. Considering the problems of running efficiency and computing speed, the distributed computing and storage of cloud computing system have more advantages than single processing. Therefore, more and more technologies are put into the use of cloud computing technology [2].

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 347, pp. 239–250, 2021.
https://doi.org/10.1007/978-3-030-67871-5_22

Data mining is a process of building models and discovering the relationships among data in large-scale data of database by using various analysis means and tools, and obtaining valuable information that has not been known before. It belongs to an interdisciplinary field, including statistics, database, pattern recognition, machine learning, artificial intelligence, high-performance computing, data visualization and other fields. Figuratively speaking, data mining looks for small valuable parts from large raw materials, just like extracting gold from soil or sand. The main task of data mining is to extract patterns from data sets. According to function classification, data mining can be roughly divided into two general characteristics: description and prediction. Description is mainly to find the data stored in the database through data mining, while prediction is the basic data knowledge of data mining, which can be divided into feature analysis, association analysis, classification prediction, clustering analysis, inference and prediction [3].

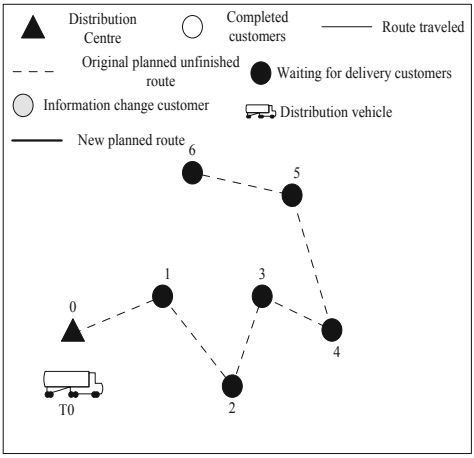
Cold chain logistics generally refers to the system engineering to reduce the loss of products in the production, storage, transportation, sales and all links before consumption, which is always in the specified low temperature environment to ensure product quality. Cold chain logistics is established with the progress of science and technology and the development of refrigeration technology. It is a low-temperature logistics process based on refrigeration technology and means of refrigeration technology [4]. Therefore, at this stage, some drugs circulation, as long as the cold chain logistics mode is adopted, while ensuring the quality of drugs, the information in each link is clear. In order to grasp the status of drugs in circulation in time, a dynamic mining method of cold chain logistics data is proposed under the background of cloud computing.

2 Drug Circulation Under Cold Chain Logistics Model

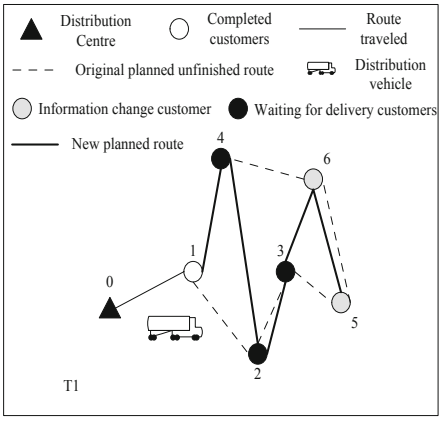
In a certain period of time, according to the change of drug information, drug information is divided into fixed drug information and changed drug information. Because the traffic network information changes all the time, in order to increase the anti-interference of the preset model, a coefficient of variation can be set. When the coefficient of variation is greater than a threshold, it is understood that the traffic network information has changed; when the coefficient of variation is less than the threshold, the traffic network has not changed. The changed drug information and the changed traffic network information can be collectively referred to as the changed information. Therefore, when the changed information is not generated, the cold chain logistics distribution network is static. According to the existing information, the distribution cloud platform plans the distribution path, sends the planned route map to the terminal of the cloud distribution system, and guides the driver to drive. With the implementation of refrigerated vehicle distribution task, once the distribution cloud platform receives the change information request, the distribution cloud platform will adjust and plan the original driving route according to the new data to respond to the change information.

Since the change information is generated based on the concept of time, the strategy of “time axis recording + cloud service platform delivering distribution information” is

adopted. When the change information is generated, record the time t, analyze the unfinished, ongoing and newly generated drug information at the time t, so as to divide the dynamic distribution problem into static problems by using the time axis, and the cloud service platform will The distribution information is known and transferred to the intelligent vehicle scheduling system for global optimization [5], as shown in Fig. 1.



(a) At t0



(b) At T1

Fig. 1. Dynamic distribution path of cold chain under time axis

(a) It means that when the refrigerated vehicle has not started in the distribution center, it plans the distribution path according to the known customer information and real-time road condition information. It is known that the acceptable service time of

customer 5 and customer 6 is adjusted at time t, and the vehicle just serves customer 1, so a new path planning is carried out for the remaining customers.

3 Dynamic Data Mining Method of Cold Chain Logistics

3.1 Cloud Computing Technology Extracts Target Data

According to the cold chain logistics model, the cloud computing technology is used to extract the target data. Cloud computing technology is widely used in the logistics industry. Through the collection of historical data, current data and follow-up new data in the cold chain information source, the historical data and current data are analyzed to explore the general law of follow-up new data. According to the new law, the status of drugs in the circulation link is analyzed to facilitate the cold chain logistics enterprises to formulate targeted development plans and lines Industry planning [6].

In the aspect of drug safety, the data collection and visual analysis under the background of cloud computing can obtain the logistics data in the circulation link; the distributed computing results and database query results of drug safety can be obtained by integrating the distributed processing and data warehouse technology; the sensor inspection information processing and fusion technology can be used to analyze the drug in each link by building a cloud platform Stream information and real-time environment data. Figure 2 below is a schematic diagram of the cold chain logistics data collection model under cloud computing technology.

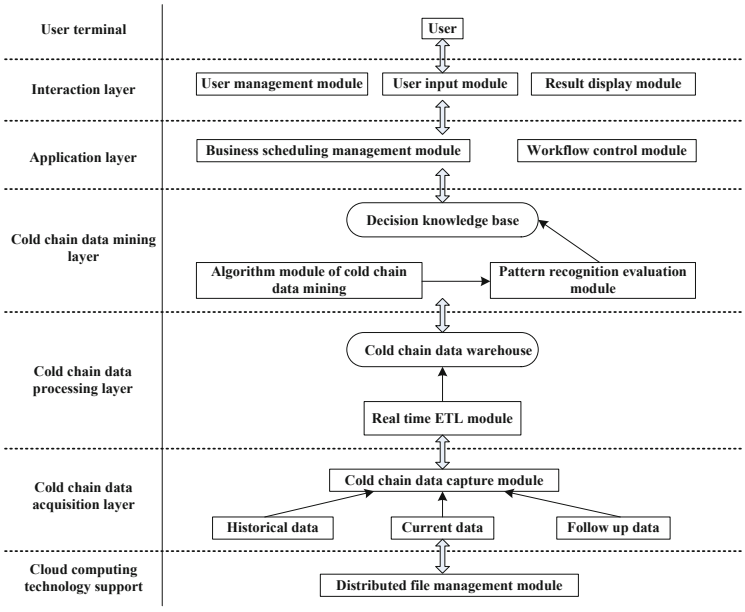


Fig. 2. Cold chain logistics data collection model under cloud computing

One of the key points of cold chain logistics data collection is temperature information, because the temperature is too high or too low, which has an impact on the quality of drugs. According to this characteristic, temperature and humidity recorder and acquisition standard are constructed. In order to ensure the temperature monitoring in cold chain logistics and provide the most timely data, cold chain logistics cloud computing center provides a non-contact temperature and humidity recorder that can realize the automatic collection and storage of temperature data.

The logic components of the temperature and humidity recorder for cold chain logistics include temperature data acquisition, data processing and data storage. The corresponding hardware is integrated digital temperature and humidity sensor, low-power microcontroller and data memory. According to the length of refrigerated transport car, install appropriate number of temperature and humidity probes, 3 below 10 m, 4 above 10 m. The single-chip controller suitable for industrial environment is selected, which has the design of anti disassembly and anti destruction. The internal integrated GPS module, six axis gyroscope, TF memory card interface, SIM card and standby power supply. After the temperature and humidity recorder is installed in the cold chain transport car, it can collect data in real time, record vehicle turning, braking, tilting and position information, and send it to the cloud computing center through 3G/4G mobile data network. Table 1 is the main collection index of temperature and humidity recorder [7].

Table 1. Main collection indexes

Index item	Parameter	Value
Temperature	Scope of work	–40 °C–80 °C
	Temperature error range	±0.5 °C
	Waterproof grade	IP65
	Working voltage	3–36 V
Humidity	Scope of work	0–100%
	Humidity error range	3%
	Waterproof grade	IP65
	Working voltage	3–36 V

The cloud computing technology is used to count these data and classify them according to the data sources, so as to provide reliable data sources for further data mining.

3.2 Target Data Preprocessing

The data extracted from cloud computing is often incomplete and noisy. The quality of data often determines the quality of data mining. In a mess of disordered data, it is impossible to carry out effective data mining and data analysis. The four processes of data cleaning, data integration, data conversion and data reduction are collectively referred to as data preprocessing. Before data mining, data preprocessing can improve

data integrity, accuracy and availability, remove redundant data, reduce the amount of computation, and improve the quality of data mining.

Data cleaning is the processing of data noise and incomplete data cleaning; data integration is the unified storage and management of data of different data formats and different data sources; data transformation is to transform data of different structures into data input structures with applicable data mining algorithms on the premise of modifying the data structure input by data mining algorithms; data reduction In the face of massive data, data reduction can reduce the amount of data while keeping the integrity of data unchanged [8].

The main content of data cleaning includes two parts: noise removal and incomplete data processing. Noise removal refers to the treatment of random errors, deviations and redundancy in data, mainly including smoothing, clustering and de duplication. Smoothing is to smooth data values by deleting data that does not conform to the general data distribution curve. Clustering is a group of similar data, and outliers may be noise. The reason of de duplication is that there is no need to repeat data in many analyses, so de duplication is used to remove noise. Processing incomplete data refers to the processing method of treating incomplete data, including two processing methods: deleting data and supplementing complete data. Specifically, it includes data deletion, mean value supplement and neighbor supplement. Among them, deleting data is to delete all data with null attribute, and the number of data applicable to null attribute is far less than the total number of non null data; mean value supplement is to use the mean value of all non null values of attribute instead of null value, and the change degree between attribute values is small; nearest neighbor supplement is to make up for the missing data through proximity data.

If the attribute values of the known data a and b are a_i and b_i respectively, there is a formula:

$$f(a_i, b_i) = \begin{cases} 1 & if \quad a_i = b_i \\ 0 & else \end{cases} \tag{1}$$

According to the above formula, the nearest neighbor measures of a and b are calculated as follows:

$$\eta(a, b) = \sum f(a_i, b_i) \tag{2}$$

According to the above calculation results, find out the nearest m neighbor with the largest measure, that is, the most similar neighbor, use the attribute value of neighbor instead of null value, this step can deal with binary Boolean data. The threshold supplement first calculates the distance C between the incomplete data and all other data. If $C < Threshold$ is less than the threshold, the corresponding attribute value of the data is used instead of the null value, which C can be Euclidean distance, and its calculation equation is:

$$C = \left(\sum_i \left| a_i - b_j \right|^2 \right)^{1/2} \tag{3}$$

You can also default the above formula to Manhattan distance:

$$C = \sum_i \left| x_i - x_j \right| \tag{4}$$

The formula can deal with discrete and continuous data. Experience supplement is to supplement the vacancy value manually according to experience, which requires the processing personnel to have rich experience and low data complexity.

According to the above process of cleaning data, according to the results of cleaning data conversion processing. Data conversion includes four cases: unit conversion, data generalization, normalization and attribute construction: unit conversion is the unit of original data, which may need unit conversion. For mobile cold chain data, common unit conversions include: time and normal time conversion; time measurement second, minute and hour conversion. Data generalization is to layer the concept and replace the low-level or "original" data with the high-level concept. For mobile cold chain data, common data generalization includes: attributes of mobile user location can be generalized to higher-level concepts; time of a week can be generalized to workdays and weekends; time of a day can be generalized to morning, noon and evening. Normalization is to give equal weight to attributes of different dimensions so that they have the same influence on the analysis results. The normalization methods for data include attribute normalization and attribute standardization.

Data specification includes four main methods: dimension reduction, numerical reduction, data sampling and discretization: dimension reduction uses coding mechanism to reduce the size of data set and delete irrelevant or weakly related attributes in data. Numerical reduction is to replace a large amount of data with a small amount of data in an average sense. Data sampling is faced with a large number of data samples and can not be processed. It can sample the data and analyze the subset. Discretization is the discretization of continuous attribute values. The continuous attribute values in the original data are marked with discrete intervals to simplify the number of attribute values and simplify the original data. According to the above data preprocessing method, the cold chain logistics data in the drug circulation link is preprocessed to ensure the authenticity and integrity of drug related logistics data. Through dynamic mining method, the required logistics data is obtained.

3.3 Dynamic Mining of Logistics Data

The dynamic data mining task is implemented, which is based on the preprocessed data and finds out the same type of data characteristics according to the association 2^m rules between the data. And processed data sets, generally, there are m data clusters of different phases. There may be frequent s data clusters, and there will be a rule.

Therefore, in these complex data sets, all frequent data clusters that meet the minimum support threshold are found, and then association rules with high confidence

are mined from these clusters. The possible rules between frequent data clusters can be expressed by the following formula (5):

$$\sigma = \sum_{\varphi=1}^{m} \left[EDI \begin{pmatrix} m \\ \varphi \end{pmatrix} \times \sum_{Z=1}^{m-\varphi} (m-\varphi)/Z \right] \tag{5}$$

In the formula: σ possible association rules obtained from cluster data cluster analysis; φ is number of candidate clusters; Z support count of each candidate cluster. According to these possible association rules, the association degree between these rules is measured [9–11].

According to these possible association rules, the association degree between these rules is measured [12]. Set $B = \{w_1, w_2, \ldots, w_m\}$ to represent the set of all items. If $X \subseteq B$ is a drug logistics data X mode, X is called the item set when there are h data information in the mode h.

It is assumed that data P is a set in drug circulation, in which each circulation data H is a set of items, and each data type is represented by HP. U and V are two cluster item sets, which exist in $U \subseteq B$, $V \subseteq B$ and $U \cap V\lambda$ are included U in the drug logistics data H. If and only then $U \subseteq H$, the association rule is $U \Rightarrow V[u, v]$. Therefore, the description equation of support and confidence is:

$$\begin{cases} support(U \Rightarrow V) = P(U \cup V) \\ confidence(U \Rightarrow V) = P(V/U) = P(U \cup V)/P(U) \end{cases} \tag{6}$$

$$\frac{|W - K_i|}{|W|} \leq \beta, 1 \leq i \leq m, \beta \in [0, 1] \tag{7}$$

In the formula: β indicates the specified parameter, which W is used to describe the division precision Y_i. It will be divided into data blocks. If the above formula $K_1, K_2, \ldots, K_n$ and W does not hold for all sums, it needs W to be divided into data blocks Y_n'. So far, under the background of cloud computing, the dynamic mining method of cold chain logistics data of drug circulation link can be realized.

4 Experimental Study

This paper proposes a comparative experiment, compares the data mining method of this study with the traditional data mining method, analyzes the differences between the two mining methods, and draws specific experimental conclusions according to the experimental test results.

4.1 Experiment Preparation Process

Set up the experimental test platform, select the experimental test software, and the following Fig. 3 is the operation page of the experimental test software.

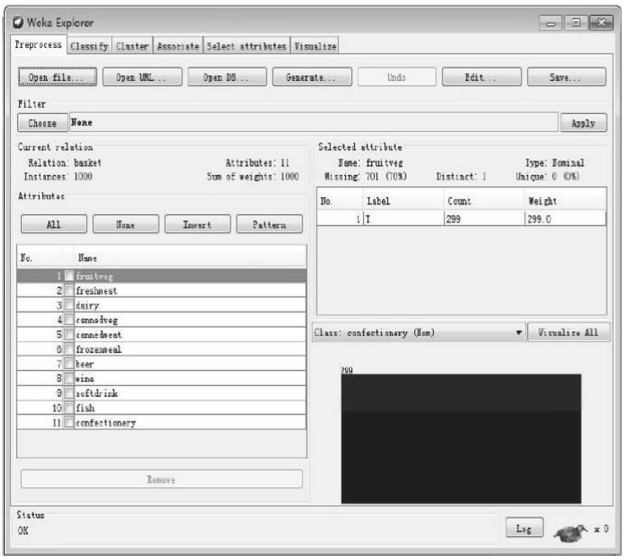


Fig. 3. Test software

Randomly select a pharmaceutical group in a certain city, take the cold chain logistics selected by its drug circulation as the experimental test object, randomly select the drug circulation data in three directions as the object to be mined, respectively recorded as N1, N2, N3. Table 2 below is the basic information of the object to be mined.

Table 2. Basic information of objects to be mined

	Data size	Data type
N1	45.5 GB	Distribution link
N2	28.63 GB	Transportation link
N3	104.27 GB	Transportation link

Figure 4 below is the distribution diagram of these drug circulation data in the cold chain logistics data storage database.

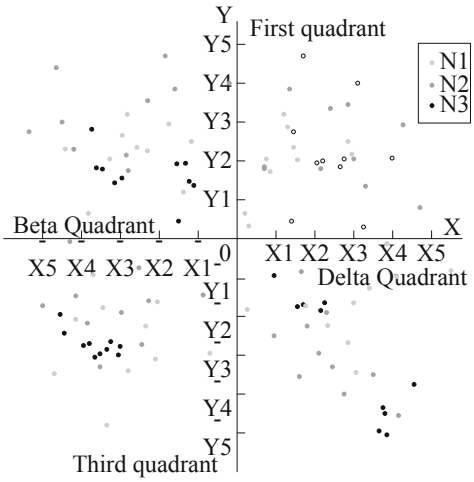


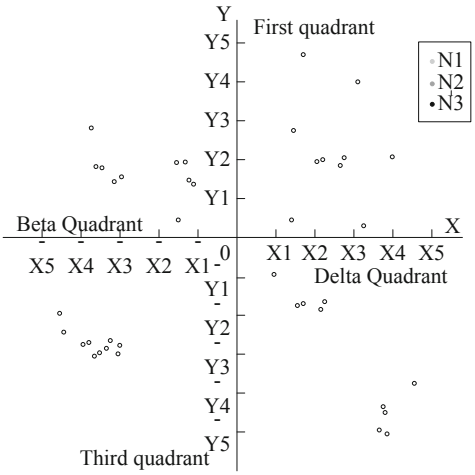
Fig. 4. Distribution of experimental objects

According to the above figure, the data in N1-N3 group are randomly distributed in four quadrants and do not have specific laws, which meet the requirements of this experiment. Therefore, two data mining methods are used to mine the data in Fig. 4 and analyze the mining effect of the two methods.

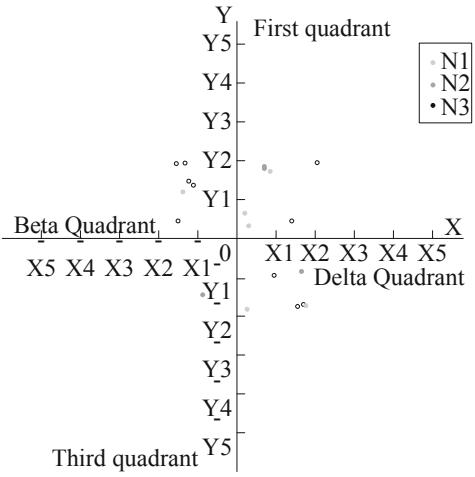
4.2 Experimental Results and Analysis

Taking the proposed data mining method as the experimental group and the latest data mining method proposed by other scholars as the control group, all data in N3 group are mined, as shown in Fig. 5.

After analyzing the two groups of test results, it is known that under the same experimental conditions, the mining method proposed in this paper is to mine all N3 data in four quadrants, and there is no problem of missing and mining errors. However, under the latest mining methods, for the remote N3 data, it is not mined at all. At the same time, there are some mining results of N2 and n3ex data in a small range. Based on the above experimental results, we can see that the proposed data mining method has already cleaned and compensated the data in the preprocessing stage, so as to get the complete mining results.



(a) Test results of experimental group



(b) Control group test results

Fig. 5. Experimental test results

5 Concluding Remarks

On the basis of combing the relevant research results at home and abroad, this paper improves the original traditional data mining methods by using the technology of cloud computing, data mining and logistics management. By means of cleaning and compensation, the repeated data and missing data in the circulation of drugs are preprocessed to achieve a higher degree of data mining integrity. However, the proposed mining method does not consider the mining efficiency, so there may be a certain

degree of backwardness in efficiency, so in the future research and innovation tasks, we can also improve the data mining technology from the perspective of efficiency.

Acknowledgements. This work is supported by the Key R & D plan of Shandong Province (public science and Technology) with No.2019GGX105013, and the Social science planning research project of Shandong Province with No.17CQXJ11.

References

1. Shi, W., Liu, Z.: Virtual machine migration algorithm to reduce energy consumption in datacenter. Comput. Digital Eng. **46**(01), 39–41+ 203 (2018)
2. Zhigang, W., Wenbin, P.U., Pengguo, T.: Data security storage technology in cloud computing. Commun. Technol. **52**(02), 471–475 (2019)
3. Weijiong, C., Wen, F., Xiaolin, Z., et al.: DBN method for risk assessment of dairy products cold chain logistics. J. Syst. Simul. **31**(05), 936–945 (2019)
4. Zhou, L., Shang, R., Zhan, C.: Research on the optimization of drug safety supervision system under the perspective of social co-governance. Health Econ. Res. **36**(06), 54–57.52 (2019)
5. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
6. Yan, Y., et al.: Analysis on mechanisms and medication rules of herbal prescriptions for gout caused by heat-damp accumulation syndrome based on data mining and network pharmacology. China J. Chin. Mater. Med. **43**(13), 2824–2830 (2018)
7. Fuwen, X.I.E.: Optimization Analysis of Discontinuous Data Path Mining in Cloud Computing Environment[J]. Computer Simulation **35**(06), 432–435 (2018)
8. Liu, M., Lü, D., An, Y.: Applications research of data mining technology in big data era. Sci. Technol. Rev. **36**(09), 73–83 (2018)
9. Wang, L.: Detection of offensive information on Internet via data mining. J. Hunan Ind. Polytech. **19**(01), 25–27 + 46 (2019)
10. Huang, Q.: Analysis on application of data mining in cloud computing. wuxian hulian keji **16**(14), 145–146 (2019)
11. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
12. Gui, G., Yun, L. (eds.): ADHIP 2019. LNICST, vol. 301. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-36402-1