

## Research on the Logistics Information System Based on Data Mining

Yonggang Zuo <sup>a</sup>, Weiping Zhu <sup>b</sup>

College of Air Transportation, Shanghai University of Engineering Science, Shanghai, China

<sup>a</sup>10203809@qq.com, <sup>b</sup>rret123@qq.com.com

---

*Abstract: The logistics information system realizes the intelligent platform for real-time monitoring of express data, using big data technology. The paper analyzes the application scenarios of data mining in the logistics system, and points out the problems that may arise during specific operations.*

*Keywords: logistics; Data Mining;*

---

### 1. INTRODUCTION

Logistics information is the display of information on various activities and characteristics of logistics. Logistics uses information to change from general activities to systematic activities. As long as there is information involved in the logistics operation process, logistics activities will not become a one-way operation activity. In the logistics process, logistics related information has feedback and logistics activities can become modern logistics systems such as output, conversion, input, and information feedback.

### 2. PROPERTIES AND PROBLEMS

(1) The data collection efficiency is low, and the accuracy is not high. The data collection of the existing logistics information system mainly comes from the manual input of nodes or barcode scanning. The data is passive input. It is not synchronized with the items in terms of timeliness, and the accuracy of the data is difficult to guarantee. . However, the information of the articles in motion is not collected and is in the empty area of information, which increases the risk of goods and cargo damage.

(2) Real-time logistics information is difficult to control, and supervision is difficult Because most logistics systems do not have an effective real-time information monitoring system, it brings great difficulties for supervision, such as the flow of goods, the recall of dangerous goods and non-conforming goods, and so on. Existing systems cannot monitor the real-time information of goods, especially dynamic information in logistics, resulting in numerous accidents. Once a real-time monitoring system with item traceability and traceability has been established, the safety and controllability of items is greatly increased.

(3) The use of information management software is low. At present, automated information management technology has been extensively applied to enterprise management, and the logistics industry is also the same. At present, a large number of automated management software for the

information management characteristics of the logistics industry have emerged. They broadly include: Logistics System Integration Software, Manufacturing Resource Planning (MRPII), Enterprise Resource Planning (ERP), Vendor Managed Inventory System (VMI), Supply Chain Management (SCM), etc. However, at present, in view of the entire logistics industry, there are very few companies that use a software system to manage the data in a rational and scientific manner. This has affected the information management of the logistics industry.

### 3. DATA MINING AND LOGISTICS

#### 3.1 Common Methods of Data Mining

Usually we convert information into value, and we need to experience four levels of information, data, knowledge, and value. Data mining is an important link in the middle, and it is a process of discovering knowledge from data.

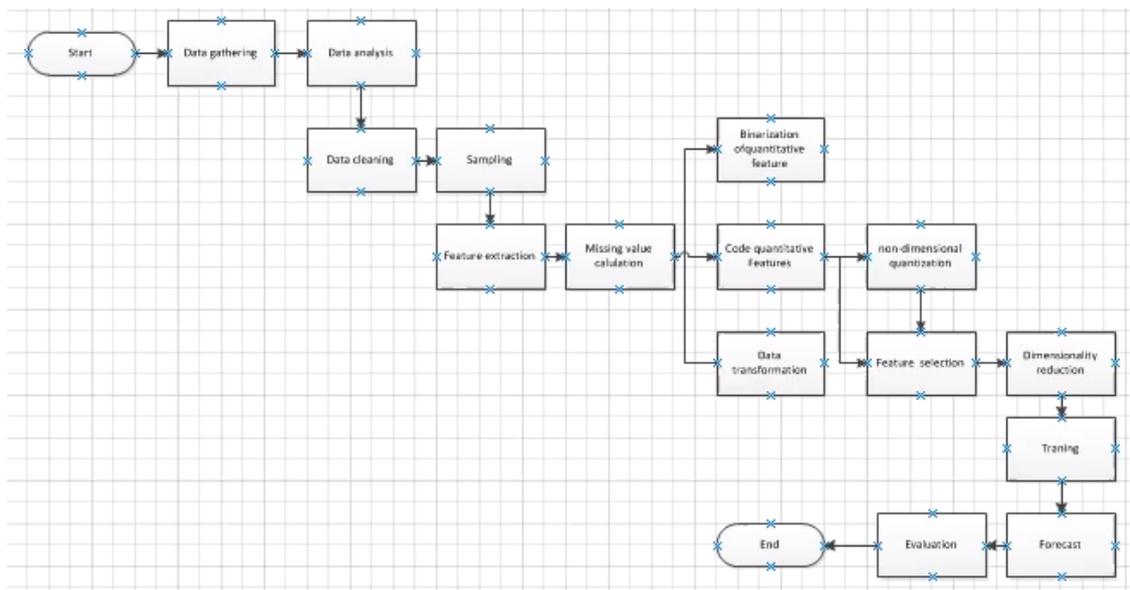


Fig 1: Data Mining Process

Cluster detection method is one of the earliest data mining techniques, also known as unguided knowledge discovery or unsupervised learning. Groups generated by clustering are called clusters, and clusters are collections of data objects. The process of clustering detection is to make any two objects in the same cluster have higher similarity, and there are higher dissimilarities between two objects in different clusters. Clustering detection methods for data mining include: partitioning methods, hierarchical methods, density-based methods, network-based methods, and model-based methods.

Decision trees are mainly used for classification and prediction, and provide a way to show rules similar to what will be obtained under what conditions. A decision tree represents a series of questions, each of which determines what the continuation problem will be. The basic components of a decision tree include decision nodes, branches, and leaves. The top node is called the "root" and the distal node is called the "leaf." Decision tree in data mining is a frequently used technology. Commonly used algorithms include CHAID, CART, Quest, ID3, and C5.0. The decision tree is suitable for dealing with non-numeric data, but if the generated decision tree is too large, it will bring difficulties to the analysis of the results. Therefore, the decision tree needs to be pruned after the

decision tree is generated, and finally the decision tree is transformed into Rules for classifying new cases. The

The neural network method is getting more and more attention, mainly because it provides a relatively effective and simple method for solving large-scale complexity problems. Artificial neural network methods are mainly used for classification, clustering, feature mining, and prediction. It learns and applies learned knowledge to a training data set to generate patterns for classification and prediction. For situations where the data is indefinite and there are no obvious patterns, applying artificial neural network algorithms is more effective. The artificial neural network method simulates a biological neural network whose basic unit imitates the neurons of the human brain and is called a node. Simultaneously, a connection link node is used, which is similar to the connection between neurons in the human brain. Artificial neural network methods mainly include: feedforward network, feedback network and self-organizing network.

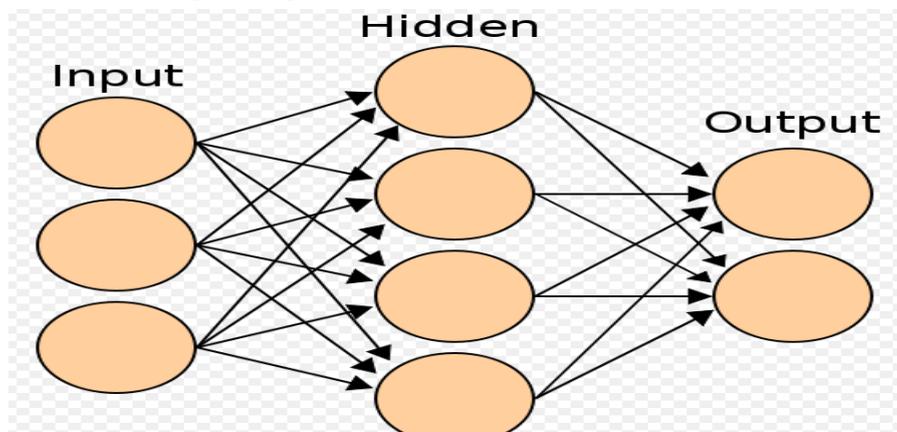


Fig 2:An artificial neural network.

### 3.2 Application of Data Mining Technology in Logistics

The location of the logistics center is an extremely important part of the process of constructing the logistics system. It is mainly to solve the problem of minimizing the sum of transportation costs, variable processing costs, and fixed costs. Site selection needs to consider how the center points are distributed and the number of center points, especially the problem of multi-center site selection. The multi-center site selection refers to selecting a certain number of locations among a number of known alternative locations to set up a logistics center so that the total cost (mainly the cost) of the formed logistics network is the lowest. In practice, mathematical planning has some difficulties when the scale of the problem becomes large or some market factors (such as customer demand) are taken into consideration. To solve this problem, we can use the method of classification tree in data mining to solve.

The rapid development of electronic commerce has made modern logistics management increasingly demanding warehousing. Reasonably arranging the storage and placing of goods, improving the efficiency of picking goods, reducing the storage cost of goods, providing more customized products and services for customers, and providing more value-added services are issues that logistics managers must consider. Using correlation analysis methods in data mining techniques can help

optimize storage in warehouses. The main purpose of the association analysis method is to find out the hidden relationship between data.

#### **4. CONCLUSION**

The application of data mining technology in the logistics industry is becoming more and more popular. More logistics companies are aware of the importance of data mining for decision support, especially in the areas of market forecasting, solution selection, warehousing, distribution, and customer analysis. Great effect. China's logistics companies are still at the initial stage of data mining applications. However, these enterprises can combine their own actual conditions, starting with the most basic data mining technology applications. With the continuous development of the logistics industry, data mining technology will be for management decisions. Provide more powerful support functions for the development of logistics companies.

#### **REFERENCES**

- [1] HUANG F, "Research on logistics management system of tropical agricultural products based on SOA", Northeast Agricultural University, 2011 Vol. 2 (10), p212-227
- [2] MA XT , "Heilongjiang saves produce content to shed development problem and politic research",Jilin University, 2013.vol.6 (2), p17-22
- [3] CHEN C, LI B. "Present situation problems and countermeasures of farm produce logistics in China under the background of urbanization",Research of Agricultural Modernization, 2013 vol(3), p:328 -332.
- [4] DING H,"On the application of supply chain theory in enterprises distributing farm produce",China Business and Market, 2014 vol(1),p:21-25.
- [5] WANG YQ,"Agricultural products logistics in China: current problems and policy alternatives", GrainDistribution Technology, 2009 vol(4) p:40 -43.