

Research on the Price Prediction in Supply Chain based on Data Mining Technology

Yang LanQin¹, Xu Xin²

¹Hebei College of Industry and Technology ,

²Hebei Education Examination Authority

Email: xugemyson@163.com

Abstract—Through using data mining methods, we can find useful hidden trends and relationships in the mass data. This can help supply chain companies to improve the quality of decision-making on supply chain management with the gained knowledge. Take the supply chain product polyester filament as an example, through the influence factor analysis of polyester filament price; this paper uses data mining methods to predict the prices of polyester filament. The established predictive models and analytical results can be used in the supply chain enterprises and as the basis of macro-control on the chemical fiber industry of and relevant departments.

Keywords- Price Prediction; Supply Chain; Data Mining

I. INTRODUCTION

Supply chain management is a modern enterprise management mode, which organize, plan, coordinate the information flow, capital flow and trade partnerships in the supply chain, and it needs a full business and market information. However, the cost of getting the supply chain enterprises and product information is very high by traditional methods. Information technology provides a force support for enterprises to implement supply chain management, making the supply chain information sharing easily, and all enterprises in the supply chain can create value through information management. Data mining as a novel information processing technology is increasingly used to solve problems of corporate management. Supply chain enterprises exchange data and information frequently, and this provides the necessary foundation for data mining, and the scope of data mining expanded from the internal to outside of the enterprise. With these information resources, suppliers, manufacturers and vendors can develop a practical demand, production and supply planning, so that information flows smoothly along the supply chain to help run and coordinate the whole industry.

There are a large number of synthetic intermediate products from the chain of chemical fiber raw materials to chemical fiber production, and these intermediate products are both as finished products of upstream chain and raw material products of downstream chain. Thus, there is a close interdependence between fiber products companies. Chemical fiber supply chain is formed based on the closely linked characteristics and inter-enterprise information sharing demand. This paper has used data mining methods to predict the product prices, and the results is satisfactory. This can help synthetic supply chain companies'

decision-making and the relevant departments' macro-control in chemical fiber industry.

II. DATA MINING

Data mining is to extract people interested knowledge from large data sources, and this knowledge is implicit, previously unknown and potentially useful information. It brings together the achievement from the various disciplines of machine learning, pattern recognition, databases, statistics, artificial intelligence and management decision-making systems, etc., and is still in a rapid change and development.

Data mining capability can generally be divided into two categories: predictive and descriptive. Descriptive mining function describes the general characteristics of the data in the database, and predictive mining function uses the current data to extrapolate and predict. Data mining is to extract knowledge from large amounts of data, because of the nature of the data itself, those algorithms in line with the reasoning and statistical theory, information theory, genetic algorithms and neural networks knowledge are first applied to the design of the knowledge discovery algorithms.

The statistical analysis is currently the earliest application, the most mature and effective form of data mining methods. The key of the method is to construct appropriate statistical and mathematical models to explain the data analysis mode. This method requires that the user has a rich professional of knowledge. Statistical analysis are generally divided into two major steps: First, users extracts the appropriate data from the database; Second, users perform visualization and analysis functions provided by statistical analysis tool to find the relationships between data, and construct statistical models and mathematical models to explain the data. The second step is repeated and continuous refinement. If analysis step is further broken down, it includes problem definition, data extraction, data pre-processing, knowledge extraction and knowledge assessment process.

III. AN EXAMPLE OF THE TECHNOLOGY

A. The definition of the instance

In 2008, the economic benefits of China's chemical fiber supply chain enterprises are severe downturn, mainly for the following two aspects: First, chemical fiber supply chain enterprise's profit level dropped significantly, which is 2.56 billion yuan in 2008, loss 3.44 billion yuan

compared to 2007. Second, the operation quality of chemical fiber supply chain enterprise is decreased. The debt ratio continues to rise, and the rate of return on total assets, profit on sales and other operating indicators decreased rapidly.

There are a number of factors leading to chemical fiber supply chain enterprises' economic benefits decline, mainly because the low price of chemical fiber products. Thus predicting the price of chemical fiber products has a great significance. There is a wide range of products in the chemical fiber industry. In this paper, we choose polyester filament in chemical fiber supply chain for example. In the next step before the data extraction, we not only need to clarify the relationship between the on polyester filament yarn in which the product supply chain, and also need to understand the background of the entire chemical fiber industry even related industries.

B. Data extraction

There is a large amount of data about the above supply chain products in the database, and extracting different data will directly affect the model results. In this paper, how to extract data will involve three aspects: data of what products; the extracted product data should include what attributes; data on product attributes belong to what period of time.

Supply chain company's input is output of the upstream business, and its output is the input of its downstream enterprises. Thus, in analysis of the supply chain, we should consider the relevant attributes of its upstream and downstream products in addition to other attributes. The choice of product attributes can be used visualization operations, such as scatter plots, graphs, etc., also can be applied to clustering, classification, factor analysis and other methods to filter. This is a repeated test and requires rich knowledge and keen powers of observation.

Through the analysis of prices factors of polyester filament, we extracted data from the database related to the cost of polyester filament and supply and demand data, specific product attributes are as follows:

The price of Polyethylene terephthalate(PET), which reflected the raw material cost of the filament

Filament production, imports, exports, due to the small volume of the export, we reflect filament domestic supply as production + imports;

Output value, imports and exports of textile, we use textile output value of + textile imports-textile exports to reflect the domestic demand of the filament;

We use the price of DTY 150 which is the representative product of filament instead of the price of the filament.

The above data is from January to December in 2008. The data is shown in Table1.

In above, the unit of polyester filament's production and imports is Ten thousand tons; the unit of textile's output value, imports and exports is One hundred million yuan.

C. Data preprocessing

Data cleaning. Used to fill vacancies value, smoothed data, identify isolated points and correct data inconsistencies;

Data integration. Integrate data from different data sources into a consistent data and store them;

Data transformation. Transform the data into the form suitable for mining. For example, the data's attribute can be standardized, allowing them to fall into the smaller range;

Data reduction. Use to generate a much smaller data set, but still close to original data in integrity. In this way, the data set after reduction will be more effective and produce the same (or almost the same) analytical results.

D. Knowledge Extraction

The knowledge extraction is to use knowledge discovery algorithms to extract the required knowledge by the user from the pretreatment data. This is the central link of data mining. According to the characteristics of data set and objectives to be achieved, we can select one or several knowledge discovery algorithm such as mathematical statistics, artificial neural networks, decision trees and so on, which will be quantitative and modeling the implicit relationship in the data. This step needs Continuous experimentation in order to get the best fit model.

1) *The Forecasting Model.* Based on the above preprocessed data and a repeated comparison of variety methods, a multiple linear regression model established by SAS software is more suitable:

$$150\text{DTY's price} = -2090.47 + 1.6506\text{PET's Price} - 18.9241(\text{Polyester filament's production per month} + \text{imports per month}) + 2.2259(\text{Textile's output value per month} + \text{imports per month} - \text{exports per month})$$

In above, the unit of polyester filament's production and imports is Ten thousand tons; the unit of textile's output value, imports and exports is One hundred million yuan. The sample determination coefficient of the model is $R^2=0.9834$, and the adjusted sample determination coefficient is $\text{Adj } R^2=0.9808$.

2) *Solving the forecasting model.* According to the minimum prediction error of the historical data, we choose different forecasting method, the main methods is single-index moving method, double exponential smoothing method, method of moving average and gray forecasting model.

Due to the lack of some data on PET prices forecast, it's difficult to establish multivariate regression model, so that we take the double exponential smoothing method for prediction.

The forecasting method on polyester filament's production per month + polyester filament's imports per month(domestic supply) and textile's output value per month+ textile's imports per month - textile's exports per month (domestic demand) is the gray forecasting model.

Forecasting results of variable values from January to March 2010 are shown in Table 3.

Table 3. Forecasting results of variable values from January to March 2010

Time	PET's price(yuan)	domestic supply (Ten thousand	domestic demand (One hundred million yuan)
------	-------------------	-------------------------------	--

		tons)	
Jan, 2010	5820	33.74	750
Feb, 2010	6269	37	750
Mar, 2010	6800	41.03	750

Take the predictive value of the independent variables into the filament price forecasting model, we can get the forecasting filament price is 8550 tons/yuan, 9200 tons/yuan and 10,050 tons/yuan from January to March 2010.

The real filament price is 8990 tons/yuan, 9330 tons/yuan and 9800 tons/yuan from January to March 2010.

The predicted values and real values of filament price have shown in Table 4.

Table 4. Filament price's predicted values and real values

Time	real values(ton/yuan)	predicted values (ton/yuan)	Prediction error %
Jan, 2010	8990	8550	4.89
Feb, 2010	9330	9200	1.39
Mar, 2010	9800	10050	2.55
average value			2.94

E. Knowledge Assessment

Knowledge assessment is a further explanation of knowledge(results, pattern or model) found in data mining. Besides, according to certain methods, evaluate the applicability of decision support. If the scope of the sample can be expanded, it should be tested whether the model still meets; If we found the first sample data is not fully representative, or the model itself is not perfect, we need re-process the knowledge mining.

PET's price increases, which means raw material costs of polyester filament arise, leading to its price arise, its impact amount equals to 1.6506 times of PET's price;

(Polyester filament production + imports) polyester filament domestic supply, the greater its value, the lower prices of filament, and affect the prices of polyester filament with 18.9241 times;

(Output value of textiles + textile imports per month-exports per month) the domestic demand of polyester filament, the greater its value, the higher the filament prices, and affect the prices of polyester filament with 2.2259 times.

Based on the above analysis, the state authorities can make the macro-control, such as cuts in polyester filament

production or stimulate the demand of textile can contribute to the prices of polyester filament yarn.

IV. CONCLUSION

Data mining is a multidisciplinary areas, it absorbs nutrients from database technology, artificial intelligence, machine learning, neural networks, statistics, pattern recognition, knowledge-based systems, knowledge acquisition, information extraction, high-performance computing and data visualization and other disciplines, and its unfolded prospects have attract a large number of research institutions and commercial companies, in recent years, it is a booming field. This paper uses data mining to predict the price of the supply chain products of polyester filament and establish predictive models, and analytical results can be used as the basis for supply chain companies and national authorities in decision-making.

Nowadays, with the continuous development of supply chain management, finding the trends and relationships hidden in massive data from supply chain companies through data mining can help supply chain companies improve the quality of decision-making for supply chain management and build competitive advantage.

REFERENCES

[1] Michael T. Rosenstein, Paul R. Cohen. Concepts from time series, Proc.of the fifteenth national/tenthc on ferencon Artificialin telligence/Innovativea plications of artificialin telligence, M adison,Wisconsin, United States, 2005, p.739~745.

[2] Bruck J. A. Study on Neural Networks. International Journal of Intelligence Systems, Vol.3(2006), p.122~127.

[3] Sprague R H. A Framework for the Development of Decision Support Systems. MIS Quarterly, Vol.4 (2010), p.1~26.

[4] Calvanese D, Dragone L, Nardi D, et al. Enterprise modeling and Data Warehousing in Telecom Italia. Information System, Vol.31 (2006), p.1~32.

[5] Trisolini S M, Lenzerini M, Nardi D. Data integration and warehousing in Telecom Italia. In proceedings of the 1999 ACM SIGMOD international conference on Management of data, Philadelphia, Pennsylvania, USA, 1999, p.538~539.

[6] Zhai LY, Khoo L P, Fok S C.Feature extraction using rough set theory and genetic algorithms-an application for the simplification of product quality evaluation. Computers& Industrial Engineering, Vol.43 (2002), p. 661~676.

[7] Forgy E. Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications. Biometrics, Vol.21 (2007), p. 68~781.

Table1. Original data in 2008

2008	January	February	March	April	May
PET's price	5190.3	5237.6	5284.8	5332.1	5379.3
Polyester filament's production	19.25	20.56	19.87	20.19	20.51
Polyester filament's imports	1.29	1.29	1.37	1.26	1.34
Textile's output value	3087.2	3205.0	3322.8	3450.6	3558.4
Textile's imports	83.88	86.75	88.83	91.30	97.78
Textile's exports	978.3	988.7	999.0	1009.4	1019.8
150 DTY's price	12500.0	12547.5	12595.1	12642.6	12690.2

June	July	August	September	October	November	December
5426.5	5473.8	5521.0	5568.3	5615.5	5662.8	5710.0
20.72	20.35	20.53	20.72	20.71	21.09	21.28
1.33	1.31	1.20	1.28	1.27	1.35	1.24
3576.2	3794.0	3911.8	4059.6	4147.4	4265.2	4353.0
96.25	98.72	101.20	107.67	103.29	102.91	102.53
1030.2	1040.6	1050.9	1061.3	1071.7	1082.1	1092.5
12737.7	12785.3	12832.8	12880.4	12927.9	12975.5	13023.0