Enterprise Commodity Marketing Innovation Based on Big Data Mining Technology

Feng Jun

Inner Mongolia Vocational College of Chemical Engineering, Inner Mongolia,China

Abstract

The advent of the era of big data has brought many opportunities and challenges to the marketing of enterprises. Enterprises should develop marketing channels according to the requirements of the market. At the same time, enterprises further mine valuable data information, so as to improve customer satisfaction for enterprise products. This paper analyzes the opportunities and challenges brought by the era of big data to the marketing market of enterprises, and explores how to innovate the marketing strategies of enterprises. This paper describes the background of the current data mining and the main data mining technology in this field. Then, it focuses on the association rule algorithm which is widely used in knowledge data mining technology and its application in marketing strategy.

Keywords: Big data, enterprise marketing, data mining, association rules algorithm.

I. Introduction

With the development and wide application of computer network and database technology, the important role of information in the development of enterprises is more and more recognized by people [1-2]. People's ability to use information technology to produce and collect data has been greatly improved. Behind these data is extremely important business knowledge, but these business knowledge is implicit and unknown in advance. Therefore, how to avoid being submerged by the vast ocean of information, find useful knowledge in time, and improve the utilization rate of information is particularly important. In the face of this challenge, data mining technology came into being and flourished, showing its strong vitality [3].

II. Data mining technology

Data mining is a process of extracting potentially useful information and knowledge from a large number of, incomplete, noisy, fuzzy and random data that people do not know in advance. There are many similar terms, such as KDD, data analysis, data fusion and decision support. In particular, data mining technology is application-oriented from the beginning [4-5]. It is not only a simple retrieval and query call for a specific database, but also a micro, meso and even macro statistics, analysis, synthesis and reasoning for these data, so as to guide the solution of practical problems, try to find the correlation between events, and even use existing data to predict future activities [6].For example, the coach of NBA, a famous national basketball team in the United States, used the data mining technology provided by a company to decide to replace the players on the spot, which was once well-known in the database industry. In this way, people's application of data is improved from low-level terminal query operation to decision support for all levels of business decision-makers. This demand driving force is more powerful than database query. It should be pointed out that all the discovered knowledge is relative, with specific preconditions and constraints, oriented to specific fields, and easy to be understood by users [7-8].

The commonly used data mining technology can be divided into statistical analysis, knowledge discovery and other types of data mining technology.

2.1 Statistical analysis

Statistics is a branch of data collection and description mathematics. Statistical inferential analysis is usually completed by means of statistical mathematical model. He infers unknown information from his own information, such as inferring future from past data, inferring population from local data, etc. Statistical analysis tools can be used in a series of business activities, such as using statistical tools for data analysis, in order to seek the best opportunities, increase market share and profits, etc. Statistical data mining technology has become the most mature data mining technology. Among them are [9]:

(1)Data aggregation and measurement technology. This technology needs to use the commonly used aggregation functions in database, such as: cunt (), sun (), avg (), max (), etc.

(2)Histogram data mining technology. This technology can directly display some important information in the database.

(3)Linear regression data mining technology. Regression is an analytical method to study the relationship between independent variables and dependent variables. Its purpose is to estimate and predict the total mean value of dependent variables according to known independent variables. For example, the profit of an enterprise depends on the purchasing power of customers and the cost of sales. Through the analysis of this dependence, we can predict the average profit level of the enterprise under the condition that we know the number of customers, the purchasing power of customers and the cost of sales.

(4) Nonlinear regression data mining technology. When the relationship between variables is roughly a straight line to reflect its changing relationship, a straight line can be fitted to reflect its changing relationship. However, in many cases, the relationship between variables is in the form of a curve, that is, non-linear, then we should fit a curve to reflect the relationship between variables. There are many kinds of nonlinear regression models: hyperbolic model, conic model, logarithmic model, trigonometric function model and so on.

(5) Clustering data mining technology is a data mining technology that groups data objects into multiple classes or clusters.

(6) Nearest neighbor data mining technology. The basic concept of prediction with this method is that objects close to each other have similar prediction values. If you know the predicted value of one of the objects, you can predict its nearest neighbor object. For example, when predicting a person's income, it is often necessary to know what class he is in or what degree he has obtained. Because people's income is often related to the people they live with and their education level, it is necessary to detect the people closest to them.

2.2 Knowledge discovery

Knowledge discovery is a technology to extract information from data in an indirect way, which is implicit, unknown and has potential application value [10]. Knowledge discovery can be regarded as a valuable information search process, it does not have to presuppose or ask questions, but it can still find those unexpected and interesting information; It can also find valuable business rules through complete and comprehensive information discovery and data analysis. It can be divided into three parts:

(1) Association rule based knowledge mining technology. It is not only a main form of data mining, but also the most common form of knowledge discovery in non teacher learning. It is also a data mining form most similar to the data mining process imagined by most people, that is, to discover some information about the database that is not known or can not be clearly expressed in the large-scale data warehouse.

(2) Neural network knowledge mining technology. It is mainly composed of the interconnection of "neurons", or

nodes organized by layers. Generally, neural network model consists of three layers: input layer, middle layer and output layer. After each neuron gets the input value, the total input value is calculated; The input value is compared by the filtering mechanism to determine the output value of the network.

(3) Genetic algorithm based knowledge mining technology. Genetic algorithm is a computational model simulating the process of biological evolution. It is a computational method combining natural genetics and computer science. In genetic analysis, search technology is applied to find out two suitable parent samples, generate the next generation of samples through "crossover", "mutation" and other operations with biological genetic characteristics, repeat the "crossover" and "mutation" operation on the sub samples until the sub samples converge, and then find another two suitable parent samples to repeat the above process, and the next generation of sample set can be obtained. The possible development direction of the current sample set is obtained.

(4) Rough set knowledge mining technology. Rough set was put forward by Pawlak Z in 1982, which is a mathematical tool to study uncertainty problems. As an extension of set theory, it is mainly used to study incomplete and incomplete data mining technology. It can solve the problem of fuzzy or uncertain data analysis and processing based on the ability of data classification without prior knowledge of data.

2.3 O2O E-commerce User Data Mining Method in Big Data Environment

Data mining makes proactive and knowledge-based decisions by predicting future trends and behaviors. The commonly used methods for data analysis by data mining mainly include classification, clustering, association rules, etc. They mine data from different angles. O2O e-commerce user data mining methods mainly include association rule analysis, classification and clustering analysis, and social network analysis. Classification and cluster analysis. Classification is to find out the common characteristics of data objects and divide them into different classes according to classification mode. Its purpose is to map data items to a given class through classification model. It can be applied to O2O user classification, user attribute and feature analysis, user satisfaction analysis, user purchase trend prediction and so on. In O2O user data mining, cluster analysis is an effective tool for market segmentation, which is used to discover different customer groups, study consumer behavior, and characterize different customer groups through purchase patterns. It can be applied to O2O user individual classification, user background and interest analysis, user purchase trend prediction and so on.

2.4 Other data mining technologies

In addition to the data mining technology described above, in recent years, with the rapid development of various data processing tools, advanced database technology and Internet technology, there are also some other data mining technologies. For example: text mining technology, web mining technology, classification analysis technology, geographic information system and spatial data mining technology and distributed data mining technology.

III. Implementation of mining tools

Software vendors quickly joined the data mining competition. This kind of competition not only makes the existing tool mining ability further enhanced, development and use more convenient, but also makes its price threshold to be reduced rapidly, which brings the possibility for the popularization of application. Especially with the company headed by Microsoft putting data mining on PC platform, data mining will become a popular technology. Here are the contributions of several software vendors in this area.

(1) Data Mining Technology launched by IBM: IBM has developed a set of data mining products called intelligent mine, which is used to help business analysis planners discover various hypotheses from data. IBM's intelligent miner is one of the few tools that can build six patterns: classification, regression, time series, clustering, association and sequence. At the same time, IBM's intelligent mine: can also connect these six modes, greatly improving the efficiency of data mining. Because olai and data mining can complement each other, IBM integrates

intelligent decision server and intelligent miner, so that users can drive intelligent miner by IDS graphical interface and analyze the results it finds. As an extension of DB2, the newly released IBM DB2 intelligent miner scoring service directly integrates data mining technology into related database management system, which meets the rapid growth of application development and configuration, and makes the application faster.

(2)Microsoft's data mining technology: Microsoft's solution is to provide a data analysis platform-—SQL Server 2000, instead of a single DM algorithm. Such a platform can provide a basic structure for application problems, so that it can be applied to many different application backgrounds. This analysis platform also makes it possible to integrate various technologies and tools. In addition, Mi-crasof t also provides an application programming interface (OLE DB FOR DM), which makes data mining possible to become a popular technology. Microsoft's new research goal is to make the database system achieve self-adjustment and self-management. The self-adjusting component added in Microsoft SQL.Server can make the database actively adapt to the application requirements, instead of the application adapting to the database as before. Microsoft's scheme enables developers who are not familiar with data mining algorithms to use data mining to achieve the purpose of analysis and prediction.

(3)MLC++ is a machine learning library developed by AbTech with C ++. MLC ++ is free and open source code. It includes common inductive algorithms, such as ID3, nearest neighbor algorithm, Naive Bayes, oneR(Holte), Winnow and decision table. It also includes interfaces of common algorithms such as C4. 5pf rlsbi-4 OCl CN2. MLC++also includes evaluation method (crates-validation bootstrap holdout), feature selection, discrete filtering, automatic parameter selection of C4.5, combined classifier and so on. Efficient predictive data mining software includes expert mining strategy, new proprietary modeling technology, compatibility with client/server structure and windows, and a variety of useful reporting and analysis functions.

(4) In addition, there are data mining software snob of LEVEL5 Quest, MineSet (SGI) , Partek,SE - Learn, SPSS, Snob, Ashraf Azmy's SuperQuery, WINROSA, XmdvTool and so on.

III. Association rule algorithm and its application in marketing strategy

Association rule is the core technology of data mining, which was first proposed by R. Agrawal and others. Association rules are given a set of items and a record set, and the correlation between items is deduced by analyzing the record set. Association rules are widely used in business, medical insurance, financial industry, judicial departments, etc., so it is of great significance to study them.

In the association rule system, the rule itself is a simple form of "if the condition is what, then the result or situation is what". The association rule can be expressed as "A→B", which includes two parts: the left part a is called the condition, and the right part b is called the afterpart. The antecedent can include one or more conditions. In a given accuracy, all the conditions in the antecedent must be true at the same time. The latter part generally contains only one case, not multiple cases.

Association rules aiming at accuracy (also called confidence) mainly express the possibility that the former part is true and the latter part is true. For the "A→B" association rule, its confidence can be defined as:

Confidence (A→B) 2. The number of tuples containing a and b/the number of tuples containing A.

The most important thing for users is the accuracy of rules. Rules with an accuracy rate of more than 80% show that the relationship found is very strong. Even if their coverage of database is low, they do not appear frequently.

Association rules aiming at coverage (also called support) represent the number of records applicable to the rules in the database. Can be defined as:

Coverage (A→B)= number of tuples containing A and B/total number of tuples

High coverage indicates that rules are often used.

Association analysis can be mathematically and formally described as:

Definition 1: let I = {i1, i2,..., im} be a set of m different attributes (predicates or items) (conventionally, we also call I an item set, but the elements in it are different from those defined by R. Agrawal et al. Where the elements in the item set may be predicates or items, while the item set defined by R. Agrawal et al. Only contains items). Given a database D, each record T is a set of attributes in I, that is, t is contained in I. If set X is contained in I and X is contained in T, then record T contains set X. An association rule is an implication in the form of X → Y, where x is contained in I, Y is contained in I, XY= Φ。 The conditions of association rule X → Y are as follows:

(1) It has support s. That is, at least s% of the records in database d contain XUY;
(2) It has confidence C. That is, at least C% of X records in database D also contain Y. Traditionally, association rules are expressed as X → Y (S%, C%).

Among them, support defines the proportion of the project in the whole database; Confidence defines the strength of discovery rules. Next, we propose a feasible correlation analysis method based on the example of customer purchase.

A department store operates cosmetics and bath products. The company has many sales outlets in the city. The company has collected the detailed purchase information of customers within a certain period of time, as shown in Table 1 (limited to space, only 6 customers and 5 products are listed as examples).

Table 1 Customer purchase information

Customer	Buy products
A	Day cream, facial cleanser, evening cream
B	Shampoo, night cream, shower gel
C	Facial cleanser, evening cream
D	Shampoo, shower gel, facial cleanser, day cream
E	Shampoo
F	Shampoo, shower gel

For the association analysis of Table 1, we first construct the association table between two kinds of goods, as shown in Table 2. Each value in the table represents the number of times that two kinds of goods represented by rows and columns are purchased by a user at the same time.

Table 2 Association table between two commodities

X (Longitudinal) Y (Horizontal)	Facial Cleanser	Sun frost	Night Cream	Shampoo	Shower Gel
Facial Cleanser	3	2	2	1	1
Sun frost	2	2	1	1	1
Night Cream	2	1	3	1	1
Shampoo	1	1	1	4	3

| Shower Gel | 1 | 1 | 1 | 3 | 3 |

According to the set minimum support threshold, calculate the minimum support of each X (in this example, set the minimum support threshold as 0.3); Support = 0.33; Support = 0.33; Support = 0.5. Others are not listed.

According to the set minimum confidence threshold, the minimum confidence is calculated, as shown in Table 3.

Table 3 Minimum confidence table of X → Y

X (longitudinal) Y (horizontal)	Facial Cleanser	Sun frost	Night Cream	Shampoo	Shower Gel
Facial Cleanser	/	0.667	0.667	0.333	0.333
Sun frost	1.0	/	0.5	0.5	0.5
Night Cream	0.667	0.333	/	0.333	0.333
Shampoo	0.25	0.25	0.25	/	0.75
Shower Gel	0.333	0.333	0.333	1.0	/

The rules that are greater than the minimum confidence threshold are listed (in this case, the minimum confidence threshold is set to 0.7):

Rule: Facial Cleanser → daily cream, support = 0.33, confidence = 0.667
Rule 2: Facial Cleanser → night cream, support = 0.33, confidence = 0.667
Rule 3: Daily Cream → facial cleanser, support = 0.33, confidence = 1.0
Rule 4: Night Cream → facial cleanser, support = 0.33, confidence = 0.667
Rule 5: Shampoo → shower gel, support = 0.5, confidence = 0.75
Rule 6: Shower Gel → shampoo, support = 0.5, confidence = 1

The conclusion can be drawn from the above rules:
(1) A considerable proportion of the customers who buy our products almost certainly want to buy facial cleanser when they buy daily cream. This shows that most customers have to use facial cleanser to wash their face before maintaining their skin during the day (It is estimated that a certain proportion of customers are white-collar beauties!) And the proportion of customers who buy facial cleanser to buy day cream or night cream is the same, depending on their hobbies.

(2) Most customers who buy shampoo will buy Shower Gel at the same time, while customers who buy Shower Gel will almost certainly buy shampoo (because most people wash their hair at the same time and wash it more often than bathe).

According to the above rules, the company has taken the following measures in marketing:
(1) The day cream and facial cleanser, shampoo and shower gel are placed together to facilitate customers to buy.

(2) After the customer has bought one kind of goods, the salesperson recommends another kind of goods appropriately.

(3) Related products will be arranged in production, delivery and transportation. After taking these measures, the

cross consumption of customers is greatly improved, and the satisfaction of shopping malls and customers is also improved.

IV. Conclusion

At present, many large foreign retail enterprises and financial insurance industry have established marketing database, and use data mining technology to find customers interested in a certain product, help make plans to know and retain the best customers, enhance the relationship with customers, and identify and track the market with development prospects. According to customer feedback to determine the product development plan, improve the effectiveness of sales personnel deployment, these have brought a unique competitive advantage for enterprises.

At present, the research of data mining technology is in the ascendant, and it is expected to form a greater climax in the 21st century. The research focus may focus on the following aspects: the research of data mining language specialized in knowledge discovery may move towards formalization and standardization like SQL language. The visualization method in the process of data mining is sought to make the process of knowledge discovery understood by users and facilitate the human-computer interaction in the process of knowledge discovery. This paper studies the data mining technology in the network environment, especially the establishment of knowledge discovery (data mining) server on the Internet, and the cooperation with the database server to realize data mining. Strengthen the mining of all kinds of unstructured data, such as text data, graphic image data, multimedia data. However, in any case, demand traction and market driving are eternal. DMKD will first meet the urgent needs of users in the information age, and a large number of decision support software tools based on data mining will come out.

References

[1] Cao Shuyan, Li Zhenxin. Research on the Third Party Logistics Mode of Cross Border E-commerce. E-commerce, 2013 (03): 23-25

[2] Hu Zhiying. New Ways of International Trade: the Latest Research on Cross Border E-commerce. Journal of Sichuan Cadre Correspondence College, 2017 (03): 138-140

[3] Jin Hong, Lin Xiaowei. Development Mode and Strategy Suggestions of Cross Border E-commerce in China. Macroeconomic Research, 2015 (9): 40-49

[4] Ji Fang, Zhang Xiaheng. Cross Border E-commerce Logistics Model Innovation and Development Trend. China's Circulation Economy, 2015 (06): 20-26

[5] Wang Wailian, Wang Mingyu, Liu Shuzhen. Analysis and Suggestions on the Current Situation of Cross Border E-commerce in China. E-commerce, 2013, 000 (009): 23-24

[6] Yang Jianzheng, Yu Lu. Analysis on the Application of Cross Border E-commerce in China's Foreign Trade Enterprises. Contemporary Economic Management, 2014, 36 (006): 58-63

[7] Manual Book for Expert Pure Data for Analytics. https://www.ibm.com/cn-zh/cloud/support?lnk=hpmls_busu_cnzh&lnk2=learn

[8] He Jia. the Dilemma and Countermeasures of Cross Border E-commerce and Logistics Integration. Modern Marketing (next Issue), 2018, 000 (007): 147

[9] Wen Jing. Difficulties and Countermeasures of Cross Border E-commerce and Logistics Integration. Global Market Information Guide, 2017 (49): 14-14

[10] Jiang Youlin. Analysis on the Current Development Mode and Strategy of Cross Border E-commerce. Business Manager, 2016 (32): 302