

# Development and Upgrading Plan for FMCG Industry Based on Data Mining

Feiyang Liu

Accounting, Eli Broad of Business, MSU, Michigan, United States, 48824  
Email: rainaliufeiyang@hotmail.com

## ABSTRACT

With the increasing amount of data in the enterprise, more and more managers need to understand the integrated data and historical data in the operations. Thus, many companies have the problem of how to use the data resources of each part rationally and assist the senior management of the company to complete relevant decision analysis. In this paper, the management platform system of FMCG industry will be taken as an example to discuss how to establish a data warehouse based on traditional database management, how to implement the classification of data mining objects, the preparation of data mining model and big data analysis, and the application of new data mining model.

**Keywords:** Data Mining, FMCG Industry, Development and Upgrading Plan

## 1. INTRODUCTION

FMCG, just as its name implies, is Fast Moving Consumer Goods. This shows that it is a new industry with simple, intuitive and time-saving methods to achieve marketing [1]. It has a high volume of turnover due to its rapid progress, short conversion cycle and high consumption efficiency . The sales of products such as beverages, tobacco, alcohol and daily living products can be counted in the FMCG industry. The focus of the FMCG industry is on "sales", the purpose of which is to get the product to the consumers in some way quickly, so that both parties can get what they need [2].

The production costs of FMCG industry are low, supply channels are extensive, and there is no other means of production technology. There are no industrial barriers. As long as the production basis is well established, the speed of development of enterprises can also be rapid. FMCG products are at low cost, and the marketing approach is in a form of small profits but quick turnover [3]. The positioning of the price also determines the market, the consumer coverage is also large. But make the opposite, FMCG industry has stepped into the intense stage in the current competitive market conditions. The company's position in the market differs due to the different types of products it produces, and the company's management uses the value chain of each industry to form a portfolio interest structure. However, with the gradual saturation of the

market and the Traut binary principle of the company's industrial positioning, most of the industries have been phased out, leading to a significant saturation of the number of companies in the industry and the occurrence of intra-company mergers and acquisitions [4].

## 2. OVERVIEW OF DATA MINING TECHNOLOGY IN FMCG INDUSTRY

### 2.1 Overview of Data Mining Technology in FMCG Industry

#### 2.1.1 Understand the Data Mining Technology

Data Mining (DM) refers to the acquisition of implicit, unknown, and useful messages in big data analysis. Sometimes, we take data information resource mining as Knowledge Discovery in Database (KDD). From a business perspective, data mining can be expressed as an advanced scientific and rational research approach to explore and classify a large amount of company data according to the company's established operation target, so as to discover implicit, unknown or verify existing economic laws and then make them modularized, advanced, scientific and rational research method [5].

#### 2.1.2 Analysis of Data Mining Technology in FMCG Industry

The research methods of data analysis development

mining applications integrate the research of application areas such as database, artificial intelligence, statistics, model recognition, mechanics institute, data analysis development mining.

(1) Description of concepts or classifications: characterization and classification. A general summary of a statistical collection containing a large amount of data is made to obtain a concise and precise representation.

Qualitative theoretical concept formulation, i.e. statistics characterization, is a generalization of the basic characteristics as well as characteristics of the target category of statistics. In contrast, the comparative theoretical concept formulation, i.e. statistical classification, compares the basic characteristics of the main object of the target category of statistics with the basic characteristics of a single or numerous objects of comparative classification. The target type and the comparative category are determined by the user, and their corresponding statistics are retrieved by database queries. For example, the user may want to compare product A, whose sales increased by 10% in the previous year, with product B, whose sales decreased by at least 20% in the same period. Another example: comparing customers who regularly buy a product with those who occasionally buy it. The result description provides a general outline of the customer comparison, such as comparing the age, education level and occupation of the two consumers, and possibly a deeper comparison on a particular characteristic to discover more distinguishing characteristics between the two categories. Some of the common methods used in both descriptions are: creating statistical measures, simplified computational summaries of graphs, data cubes, attribute-oriented induction, etc.

## (2) Frequent models and associations mining

High-frequency models, which are models generated on the frequency in the data. It includes item sets, subsequences and substructures.

An item set is a composed form of subitems that constantly coexist with the set of transaction statistics, such as milk and steamed stuffed buns [5]. Another example is the model of consumers buying a PC first and then selecting a digital camera before buying a memory card, which is also a (frequent) ordering model. Substructure forms usually include a variety of construction styles, such as graphs, trees, or lattices, which are composed of sets of items or suborders. A substructure is called a (frequent) substructure model if it is assumed to exist continuously. Mining the continuous model has produced some of the most interesting connections in research statistics. For example, the famous story of "diapers and beer" is a concrete application of the association rule.

## (3) Classification knowledge discovery

Classification means that the given data is classified in the appropriate category. The key to analysis is to classify which norms, or what principles, have been followed for the data.

There are some general approaches to mining classification laws as follows: decision trees, plain Bayesian methods, k-nearest neighbor analysis, artificial neural networks, rough set methods, and genetic computation. All types of computations are suitable for various characteristics in data sets. The most classical way of classification is based on decision trees.

## (4) Predictive knowledge

Predictive knowledge is mainly based on event sequence type statistical methods to infer future historical data from historical events or current historical data, and it can also be regarded as relevant knowledge with time as an important attribute. For example, an example of (numerical) estimation in the FMCG industry is to predict the profit of each product in future sales based on previous sales statistics.

At present, regression analysis is one of the most commonly used statistical approaches to numerical prediction, along with various other approaches such as neural networks and machine learning.

## (5) Cluster analysis

Cluster analysis processes data results in no predetermined classification attribution, but rather divides the overall database into different types of clusters. It aims to have significant differences between groups and clusters, with the number within a group being as close as possible.

The use of cluster in the FMCG industry can assist market analysts in finding different customers in a basic pool of customers, as well as using purchasing models to portray the characteristics of different customers. In addition, systematic cluster analysis algorithms can be used as a pre-processing process for certain calculations (e.g. characteristics and types, etc.). There are generally two main categories of systematic clustering approaches, generally the statistics and the neural.

## (6) Outlier analysis

In data analysis set, the data analysis objects that do not follow the rules (models) composed of these data analysis objects at all are called outliers or anomalies. Because the vast majority of data mining techniques easily discard outliers as noise or anomalies. But in some special application scenarios (e.g., automatic detection of commercial fraud), things that occur with small probability (data analysis) are more valuable for discovery research than things that usually occur (data analysis). Anomaly detection is often used to find

outliers, and anomaly detection should be performed in a statistical-based, distance-based and deviation point-based manner.

## 2.2 Characteristics of Data Analysis and Mining Technology in FMCG Industry

### 2.2.1 Characteristics of FMCG Industry

#### (1) Small amount per item

Even in developed countries like China, the main expenses of daily life are spent on consumer durables and other investment expenses, which is especially common in developed countries. Chinese consumers rarely spend too much on the financial burden that FMCG products impose on their daily lives.

(2) The market influence of the brand is strong, but the maximum degree of differentiation of each brand is not large

Many FMCG products are low-tech. Although there are small barriers to the introduction of the product market, they usually have great competitiveness in the international financial market. Because each uses the same manufacturing equipment or raw materials, making the difference between the various competitors in terms of the characteristics of the goods themselves is not large.

(3) The production method is mostly based on mass production

This is partly because the consumer group is generally ordinary people, it is difficult to do some personalization for products, but at the same time, the consumer group pays less attention to these goods or is less picky, there is no need to emphasize the personalized characteristics of goods. The manufacturing equipment and process technology of many FMCG products meet the requirements of high-volume operations, and unit costs can only be reduced in high-volume production.

(4) Many products have obvious seasonal and fashionable characteristics

Because there are many consumer behavior is closely related to the season or fashion, one of the most classic cases is cold drinks. It can also be said that all trendy goods are fast-moving consumer goods, because the trend is originally oriented to the rapid change of consumers.

(5) Wide range of customers covered and a large number of customers

In the case of low commodity dependence and low market attention, the consumer usually only chooses the commodity that he can easily obtain. And if the location of sale is a little far from the consumers, they generally

cannot choose the commodity. In addition, by expanding the range of distribution, the consumers can be inspired to try new products when they often see them, thus achieving the purpose of developing new users. This is why all the major mature FMCG companies consider the range of distribution as an important indicator to examine the strength of the company.

### 2.2.2 Characteristics of Data Mining Technology in FMCG Industry

#### (1) Bases on large amount of data

There is no claim that data mining cannot be achieved on small data, because most of the calculations can be performed on small data and conclusions can be drawn. Only that small data volume is fully capable of using manual data analysis to summarize regularity. Moreover, although small data volume cannot show universality in general situations, even for the special situation of FMCG industry with excessive product coverage and large number of target customers, small data volume cannot solve the actual needs of operators, so it is critical that whether the small data volume is sufficient in data mining in FMCG industry.

#### (2) Non-trivial

Non-trivial means that the information mining is not that simple, it can never be as a well-known sports commentator believes that "through my statistics, I saw a very interesting phenomenon, before the end of the game, the number of goals scored and goals conceded in this World Cup is the same. What a special coincidence!" In this case, the data mined must have its own characteristics to be meaningful.

#### (3) Implicit

The main purpose of data mining is to explore more deeply the information implied in the big data, not only the information revealed on the surface of the big data.

#### (4) Novelty

The knowledge that comes out of data mining should be previously unknown, because only new knowledge can help companies gain further insights.

#### (5) Value

The conclusion drawn from data mining should be able to bring direct benefits or indirect economic benefits to the company. For the FMCG industry, which has a large variety of goods and a wide range of products, only by ensuring the value of the data can provide strong data support for the management. Thus, in the process of data mining, it is necessary to ensure the accuracy of the mining methods used.

### 3. DATA WAREHOUSE CONSTRUCTION PROCESS OF FMCG INDUSTRY

#### 3.1 Demand Study

For a data warehouse, each data has a fixed subject. The data information is based on the individual subject domains defined in the enterprise data model, which are relatively independent from the use. The range of subjects can include both subjects describing the data of various basic entities in the enterprise, as well as subject purposes describing the liaison relationships between entity data. Precisely because the data management work environment in a data warehouse is subject-oriented, it is also determined that its product design is data-driven. But traditional operational conditions are application-oriented and product design is demand-driven. In view of the special characteristics of FMCG industry, this paper sets five major subjects of sales, inventory, order, consumer and employee respectively based on product, personnel and consumer relationships, and each subject corresponds to an fact table and a dimension table, and the fact table has a metric value in addition to the key value information connected to each dimension table. For example, the merchandising fact table has a database of key values such as item number, time code, and consumer code, as well as metric values such as store sales, branch sales, and store cost of sales; the employee fact table has a database of key values such as employee code, department code, and position code, as well as metric values such as monthly salary. The Sales subject, mainly oriented to the analysis of product sales. By setting up this subject, we can analyze the overall sales status of a product in a certain period of time, the overall sales status of a market or store, etc. The Order subject allows us to view the status of orders, deliveries, returns and payments for ordered products from various perspectives, such as date, product and supplier. The Employee subject enables us to view the number of employees, total payments, and average age of employees in a department in a given year; we can also ask about the number and age of managers in a given period, etc., to get a broader picture of the current status of the company's personnel.

#### 3.2 Warehouse Design

##### 3.2.1 Design and Construction of Data Warehouse Based on Data Mining

Because the FMCG industry covers a wide range of big information data (including marketing, customers, employees, goods delivery, product sales and after-sales service), there are many ways to design a big information data warehouse. The level of detail information it involves may vary greatly. Since the important role of data and information warehouse is to

support analysis and statistical discovery, the conclusion of an example of data and information mining in advance can be used as a reference base for product design and research on data warehouse architecture. This product design determines the level at which those are maintained and any pre-processing required to ensure quality and efficient data information mining.

##### 3.2.2 Multi-dimensional Design for Sales, Customer, Product, Time and Region

Considering consumer needs, sales of goods, market trends and fashions, as well as the quality, value, profit and service of daily products, and the FMCG industry needs real-time information. Thus, a powerful multidimensional analysis and visualization tool is also critical, which also contains a need to build a complex big data analysis cube based on big data analysis (as figure below). The basic cube covers all the city, item and year, and is able to return to any composition of these three dimensions. The vertex cube represents the case where the grouping is empty and also covers the sum of all sales. The characteristic data cube is a practical data analysis architecture in retail data analysis, as it facilitates data analysis on collections with more complex requirements.

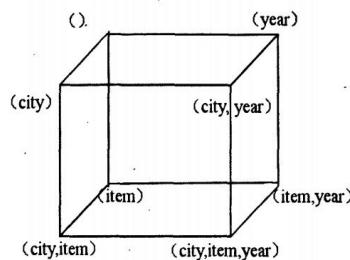


Figure1 vertex data cube

##### 3.2.3 Effective Design of Promotional Activities

The FMCG industry often uses advertising campaigns, coupons, discounts and concessions to promote their products and retain consumers. Careful analysis of the effectiveness of sales promotion activities can increase the company's revenue. Multidimensional analysis can also meet the requirements of big data analysis in this area by comparing the sales of products during the sales period with the actual number of transactions and the situation before and after the sales activity. In addition, correlation analysis can also help to identify which products may be acquired together with the fake products, especially in comparison with the direction of sales before the actual sales activity, applying it to the analysis methods of shopping cart or object data.

##### 3.2.4 Customer Loyalty Design

With the customer honor card information, the

purchase sequence of the next consumer can be recorded. Consumer loyalty and shopping trends can be analyzed in a systematic way. The products that a consumer shops for at various times can be divided into sequences. Sequence model mining can be used to analyze changes in their spending habits or loyalty and adjust prices and product models accordingly, thus retaining old customers and acquiring new customers.

### 3.2.5 Purchase Recommendation and Product Reference Design

By finding relevant information in sales records, it can be seen that consumers who purchase a certain brand of perfume are likely to be able to purchase another product. This type of information can be used to create certain shopping recommendations, which can be promoted on web pages, weekly flyers or receipts to improve service, help consumers select products and increase sales. Also, news like "Hot Item of the Week" or an attractive deal may be introduced along with the news for promotional purposes.

## 4. APPLICATION AND DEVELOPMENT OF DATA MINING IN FMCG INDUSTRY

### 4.1 Understanding the Overall Sales Situation

Through the classified information, understand the daily operation and financial situation by product type, sales quantity, store address, price and time, and know every increase in sales volume, change in stock quantity, and increase in sales volume through sales. When selling products, it is important for distributors to check whether the product components is reasonable at all times, i.e., whether the proportion of each type of product in the business is generally equal. Adjusting the product components requires consideration of changes in supply and demand due to seasonal changes, product structure adjustments by competitors in the same industry and so on.

### 4.2 Reduce Inventory Costs

With data mining system, the number of goods sold and inventory data are brought together and statistically analyzed to determine the increase or decrease for different products and various goods to achieve reasonable inventory. The data mining system can also send the inventory information and sales forecast information directly to the supplier through electronic data exchange, which reduces the middleman in the FMCG industry, while the supplier is responsible for replenishing the inventory on a regular basis, and the retailer can reduce the pressure on the enterprise.

### 4.3 Effectiveness Analysis of Promotional Activities

The FMCG industry often takes advertising campaigns, coupons, various discount promotions and concessions to do marketing activities in order to achieve the purpose of promoting goods and retaining consumers. However, only by truly and comprehensively understanding customers, they can correctly position the marketing activities in a way that will increase customer response rates and reduce the cost of marketing activities. With data mining technology, when, where, how and to whom the company should do promotional activities can be analyzed, so that effectively achieve the purpose of publicity, thus reducing the unnecessary consumption of corporate resources? At the same time, data mining technology can also be used to find the most profitable users of the company's future investment with historical data about the company's past publicity.

### 4.4 Market and Trend Analysis

Data mining tools and statistical analysis models are used to carefully investigate the data in the database, and analyze the purchasing and consumption habits of consumer groups, the success rate of advertising campaigns, and other strategic signals. With the database system, by retrieving the merchandise sales statistics in the database in recent years, and then doing reclassification and data mining, seasonal and monthly sales changes can be predicted, and make a comprehensive analysis of the trends in product classifications and total inventory. It is also able to determine whether or not to reduce the price, thus providing decisions on product quantities and operations.

### 4.5 Customer Segmentation

Customer segmentation is the division of a person's primary consumer base into several smaller segments, with consumers belonging to the same segment in close proximity to each other. Customer segmentation enables merchants to differentiate customers located in each type with a differentiated approach.

## 5. CONCLUSION

Data warehouse technology has now turned out to be a key technology of great practical value, which can convert the existing huge amount of data information into more valuable commercial messages that can be used as decision support. The only way to make resource storage play a significant role in decision support is to organically combine resource mining technology with the huge data analysis in resource warehouses, and with the modern corporate

management decision making approach of large enterprises.

## REFERENCES

- [1] Design and Implementation of Data Warehouse System for Retail Enterprises [D]. Fudan University, 2008.
- [2] Yang Genyi. A study on Marketing Strategy of Market Conditions in Chinese FMCG Industry [J]. Guangdong Economy, 2017
- [3] Gu Hailan. Design and Application of Data Nining Model in the Development of Retail Data Warehouse [J]. Hebei Journal of Industrial Science and Technology, 2008, 025(004):241-243
- [4] Ma Linyi, Wang Yan. Design and Data Mining of Data Warehouse in Retail Industry [J]. Market Modernization, 2008,00 (013):54-55
- [5] Oraman Y, Azabagaoglu M O, Inan I H. The Firms' Survival and Competition through Global Expansion: A Case Study from Food Industry in FMCG Sector[J]. Procedia - Social and Behavioral Sciences, 2011, 24:188-197.