

# Research on Supermarket Product Layout Method Based on Comprehensive Correlation Degree

Yuqing Xia<sup>1</sup>, Chenxia Jin<sup>1</sup>, Fachao Li<sup>1</sup>

<sup>1</sup>School of Economics and Management, Hebei University of Science & Technology, Shijiazhuang, China

**Abstract**—The commodity layout of a large supermarket has an important impact on the sales and benefits of the supermarket. The quality of the commodity layout directly affects the consumer's psychological feeling and purchasing behavior in the shopping process. According to the previous research, we can find that the traditional commodity layout method has the shortcomings of lack of data support, poor stability, and does not include the actual factors such as the impulse purchase behavior of consumers and the purchase law of commodities in the research scope, resulting in the problem that the optimization effect is not obvious. In view of this deficiency, this paper proposes a commodity layout method based on comprehensive correlation degree. Firstly, the concepts of adjacency effect value and placement freedom degree are given. Based on this, transaction data is used to mine association rules of commodities in customers' shopping baskets, and the influence of the same commodity on surrounding commodities due to different placement positions is further analyzed. Based on the maximization of the comprehensive correlation value, an analytic model for the optimization of commodity layout is established, and the optimal commodity layout is obtained by using genetic algorithm. In this article, the layout of the B supermarket commodity layout is taken as an example. The overall layout is more closely related. It verifies the feasibility and superiority of the commodity layout method based on comprehensive correlation, which can better improve the market competitiveness of the supermarket.

## 1 Introduction

Consumers' impulse buying behavior is actually a kind of stimulus response, which can stimulate consumers' inner purchasing desire through external stimulation, thus making purchasing behavior. When consumers purchase commodities, in addition to the commodities they plan to buy in advance, they will also purchase some commodities because of the impulse of external stimulus. As an external stimulus, commodity layout is also an important way to promote consumption. Relevant data show that scientific, professional, commodity layout that adapts to consumer psychology and consumer demand can drive sales growth of 30%-40%, far greater than the sales increase brought about by promotion [1]. The distribution of commodities in large supermarkets has a direct impact on consumers' psychological feelings and purchasing behavior during the shopping process. Reasonable layout will leave a good impression on consumers and stimulate their purchasing desire. Therefore, studying the stimulation of external layout on consumers' purchasing behavior and optimizing the distribution of supermarket commodities is a way to attract and retain consumers to buying and improve supermarket sales.

The main purpose of supermarket commodities layout research is to improve supermarket operation efficiency, increase consumer satisfaction, and increase sales. Many scholars conduct research from different perspectives. From the perspective of improving supermarket operation efficiency, the most common method is to use SLP (System Layout Design) to plan and design the supermarket layout. For example, Zhao Feng and Wang Ze [2] aimed at reducing congestion and saving time, and based on consumer traffic as a logistics factor, a dynamic SLP method was used to lay out the interior of the supermarket to obtain a clear regional layout; Liu Zhihai and Zhang Dandan [3] aimed at more concise zoning and stronger inter-regional relationships. Based on the traditional SLP method to refine non-logistics factors, they added dynamic analysis and feedback to the supermarket layout planning process to obtain corresponding optimization solutions. SLP method can effectively avoid congestion, but the method to determine the correlation degree between regions is too subjective, and the layout scheme is diverse, so the optimal scheme can not be determined.

From the perspective of improving consumer satisfaction, in order to reduce consumer shopping time and settlement time, some scholars based on data mining technology [4-6], using service operation theory, queuing theory and other methods to layout supermarket

---

E-mail: 843139009@qq.com

commodities. For example, Li Xiao et al. [7] take the shortest path for consumers to shop as the goal, based on consumer shopping data and spatial modeling, use a heuristic search algorithm to generate an approximate optimal layout for consumers to complete shopping quickly; Wang Tingting et al. [8] take the maximum satisfaction of consumers as the main goal, based on the triple representation method to achieve the description of the problem, using the PGSA (plant growth simulated algorithm) to achieve the optimal scheduling of the supermarket cashier. This method effectively reduces the consumer's picking time and improves the consumer shopping experience, but does not consider the proportion of impulse purchases, and ignores consumers' impulse purchase behavior on external stimuli when purchasing, resulting in loss of potential sales benefits. Impulsive purchases account for a significant portion of product sales [9], and this ratio has changed from the lowest 27% to more than 80% [10], and has shown a clear upward trend in recent years.

From the perspective of increasing turnover, mainly through the magnet point design, customers can go through as many areas as possible during the in-store shopping process, see more commodities, and stimulate consumer purchases. For example, Jia Tongling [11] aims to increase the sales volume of the supermarket, and based on the magnet point theory and the dynamic line of customers, obtains a feasible layout plan for the supermarket; Gu Guangsheng [12] aimed to increase sales, based on the magnet point theory, and used the shelf display as the main factor to layout the supermarket stores. At present, the magnet point theory is widely used in supermarket, but it has some shortcomings, and the guiding effect on consumers is poor. Masao ohtah and Yoshiyuki Higuchi [13] research found that the second and third magnet point customer loss rates are 56% and 59%, respectively.

Based on the above analysis, we can see that the traditional supermarket commodity layout method has the disadvantages of subjective correlation degree determination method, lack of data support, diversified results and poor guidance effect, and the model consistency is very strong, and some improvements have not been made according to the changes in the social environment, but there have been improvements in the detailed operation of the existing model. Therefore, the purpose of this article is to strengthen the relationship between commodities to stimulate consumer purchases, and to plan the layout of supermarket commodities. First, use data mining technology to conduct in-depth analysis of historical consumption data; secondly, use the adjacency effect value and placement freedom as the main observation indicators to establish a commodity layout model that maximizes the comprehensive correlation value, and improve the overall correlation of the layout; Then, use the optimization algorithm to determine the final supermarket layout. Finally, experimental results show that the method in this paper can successfully design a reasonable supermarket commodity layout, and the layout design algorithm is simple and efficient.

## 2 Association rule

Association rules are used to reflect the interdependence and relevance between one thing and other things. The implication expression is  $X \rightarrow Y$ , where  $X \subset I, Y \subset I$  and  $X \cap Y = \emptyset$ , that is,  $X$  and  $Y$  are two different things. Suppose  $I$  is a set of items, given a transaction database  $D$ , where each transaction  $T$  is a non-null subset of  $I$ , that is, each transaction corresponds to a unique identifier  $TID$ .  $num(\ )$  is the number of occurrences of a particular item set in  $D$ . For example,  $num(D)$  is the number of transaction data, and  $num(XY)$  is the number of transactions with  $\{XY\}$ . The strength of association rules can be measured by support degree (S) and confidence degree (C). The formulas of support degree and confidence degree are defined as follows:

$$S(X \rightarrow Y) = P(XY) \quad (1)$$

$$C(X \rightarrow Y) = P(Y|X) \quad (2)$$

Support and confidence describe the strength of association rules by probability. The support degree is the probability that two commodities  $X \cap Y$  appear in the trading set  $D$ , that is, the probability that  $X$  and  $Y$  are purchased at the same time. In formula (1),  $P(XY) = num(XY) / num(D)$ ; The confidence degree is the probability that  $Y$  also appears after  $X$  appears in data set  $D$ . In formula (2),  $P(Y|X) = num(XY) / P(X)$  represents the probability that  $X$  occurs in trading set  $D$ . However, it is found that the degree of support and confidence cannot exactly check the effectiveness of rules, and the ability to express rules is not strong, which cannot guarantee the effectiveness of decisions. In order to enhance the guiding significance, the lifting degree is put forward, whose intuitive meaning is the ratio of the probability of containing  $Y$  under the condition of containing  $X$  to the probability of containing  $Y$  under the condition of not containing  $X$ . It describes how much the use of rules can be improved compared with the use of no rules. The formula is:

$$L(X \rightarrow Y) = \frac{P(Y|X)}{P(Y)} \quad (3)$$

Where  $P(Y|X) = P(XY) / P(X)$  is the conditional probability, it represents the probability of containing  $Y$  under the condition of  $X$ ,  $P(XY) = num(XY) / num(D)$ ;  $P(X) = num(X) / num(D)$ ,  $P(X)$  indicates that  $X$  accounts for the overall occurrence probability; Similarly,  $P(Y)$  indicates that  $Y$  accounts for the overall occurrence probability, and the formula is expressed as  $P(Y) = num(Y) / num(D)$ . The value range of  $L$  is  $[0, +\infty)$ , when  $L(X \rightarrow Y) < 1$ , indicates that the appearance of  $X$  will suppress the appearance of  $Y$ , they are negatively correlated; When  $L(X \rightarrow Y) = 1$ , it

means that the appearance of  $X$  has nothing to do with  $Y$ , they are independent; When  $L(X \rightarrow Y) > 1$ , it means that the appearance of  $X$  can increase the appearance of  $Y$ , and they are positively related.

Association rule mining is an important technology for data mining. It is used to mine valuable association rules between data items from a large amount of data. Association rule mining has many applications in life, which are often used in the recommendation system of physical stores or online e-commerce. Beer and diapers are typical applications of this kind of technology. Through association rule mining, get the rules that can guide decision-making, the Apriori algorithm [11] proposed by Agrawal, Imielinski and Swami is the most commonly used in association rule mining research. The execution process of association rule mining using this algorithm is:

(1) Generation of frequent itemsets, all item sets that meet the minimum support threshold are found, which are called frequent item sets.

(2) Generation of association rules, all the rules that meet the minimum confidence threshold are extracted from the found frequent item set. These rules are called strong association rules.

As an external stimulus, supermarket commodity layout is a very important means to promote consumption. Mining association rules based on shopping data can get effective commodity association rules, provide decision support for supermarket commodity layout, and further improve sales ability and revenue. Therefore, by mining association rules on the customer's purchase record data set, the ultimate goal of this article is to discover the inherent customer purchasing habits. For example, the purchase of commodity A will increase the probability of purchasing commodity B. According to the mining results, design a reasonable commodity layout to achieve sales increase. The specific method of this article is introduced below.

### 3 Supermarket product layout method based on maximization of comprehensive correlation value

#### 3.1 Formal description of supermarket space

In order to facilitate the theoretical analysis, this article analyzes the three-dimensional supermarket space according to the supermarket space environment, and gives a more intuitive formal description of the supermarket space. Assume that the distance between adjacent shelves is one unit. Consumers can reach in one step between adjacent goods shelves. For non-adjacent shelves, consumers cannot reach directly and must pass through other areas. The commodity shelf is represented by rectangle, and the distance between adjacent shelves is standardized as above. The formal description of the supermarket's plane topology is shown in Figure 1:

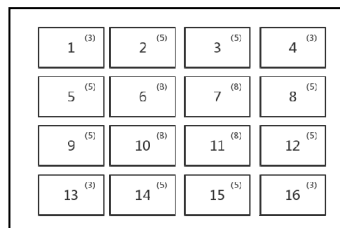


Figure 1. Flat topology of the supermarket

#### 3.2 Mathematical Model of Supermarket Product Layout Based on Adjacent Effect Value

Suppose there are  $n$  shelves. According to the properties of the commodities, the goods are also divided into  $n$  categories, numbered  $1, 2, \dots, n$ . As the lift is used to describe the influence degree of one commodity on another commodity, firstly, the degree of lift of each commodity to other  $n-1$  kinds of commodities except itself. The lift degree matrix is as follows:

$$\begin{matrix} & 1 & 2 & \dots & n \\ \begin{matrix} 1 \\ 2 \\ \dots \\ n \end{matrix} & \left( \begin{matrix} L(1 \rightarrow 1) & L(2 \rightarrow 1) & \dots & L(n \rightarrow 1) \\ L(1 \rightarrow 2) & L(2 \rightarrow 2) & \dots & L(n \rightarrow 2) \\ \dots & \dots & \dots & \dots \\ L(1 \rightarrow n) & L(2 \rightarrow n) & \dots & L(n \rightarrow n) \end{matrix} \right) & (4)
 \end{matrix}$$

When the current and subsequent items are the same, that is, when  $x=y$ ,  $L(x \rightarrow y)=1$ , the product has no lifting effect on itself. The lift matrix reflects the mutual relationship of commodities and provides basic data for subsequent calculations. For the description of commodity relationship, the commonly used method is based on association rules, using implication expression to describe the association relationship between two kinds of commodities, but this method can not fully reflect the impact of commodities on multiple commodities around. In order to better describe the relationship, this paper give the concepts of the degree of freedom of placement and the value of adjacency effect according to the position on the shelf space and the number of shelves.

Definition 1: select a shelf, the number of shelves that can be placed adjacent to it, which is called the degree of freedom of this shelf. This article uses  $k$  to express it.

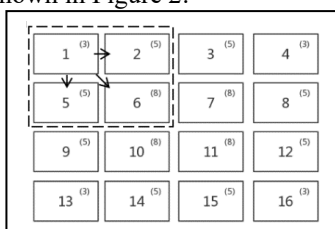
Definition 2: the adjacency effect value is the average promotion effect value of the commodities on the location to the surrounding commodities. Suppose  $a$  represents a layout and  $i$  represents the shelf number, then  $f(i|a)$  is the adjacency effect value of shelf  $i$  under the  $a$  layout. The quantization formula of  $f(i|a)$  is expressed as follows:

$$f(i|a) = f(a_{i_1}, a_{i_2}, a_{i_3}, \dots, a_{i_k} | a_i) = \frac{1}{k} \sum_{j=1}^k L(a_i \rightarrow a_{i_j}) \quad (5)$$

Among them,  $a_i$  indicates the products placed on the shelf  $i$  under the  $a$  layout,  $k$  indicates the degree of freedom of placement, and  $a_{ik}$  indicates the products placed on the  $k$ -th shelf around the shelf  $i$ .

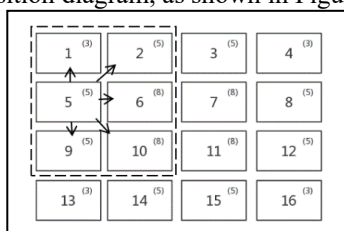
In the actual supermarket product layout, the space has a variety of shapes, such as fan-shaped, circular, rectangular, irregular, etc., but each shelf has adjacent shelves, which are placed according to the different positions of the shelves. There are many cases of  $k$  values, and three cases in rectangular space. These three cases are explained below:

Case 1: It is placed in the corner, and placing the degree of freedom is  $k=3$ . Taking shelf 1 as an example, it is relatively close to shelves 2, 5, and 6 in position, which will have a correlation and interaction. Therefore, when  $k=3$ , it is necessary to consider the influence of the surrounding three shelves on it. This situation is visually represented by a shelf position diagram as shown in Figure 2:



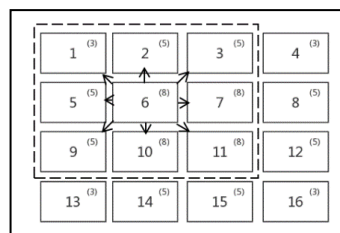
**Figure 2.** Shelf relationship diagram in the corner area

Case 2: Place it against the wall, placing the degree of freedom is  $k=5$ . Taking shelf 5 as an example, it is in a relatively surrounding state with the surrounding shelves 1, 2, 6, 9, and 10 in position, and has a correlation and interaction with each other, when  $k=5$ , it is necessary to consider the influence of the surrounding 5 shelves on it. This situation is shown by the shelf position diagram, as shown in Figure 3:



**Figure 3.** Shelf relationship in the wall area

Case 3: Place it in the middle area, placing the degree of freedom  $k=8$ . Taking shelf 6 as an example, it is surrounded by shelves 1, 2, 3, 5, 7, 9, 10, and 11, which are interrelated and affect each other. Therefore, when  $k=8$ , we need to consider the impact of the surrounding 8 shelves on it. This situation is shown by the shelf position diagram, as shown in Figure 4:



**Figure 4.** Shelf relationship diagram in the middle area

Because the commodity layout involves all shelves, a comprehensive correlation value is proposed. The comprehensive correlation value represents the sum of the adjacency effect values of all shelves under the current layout. Assuming  $a$  is a layout and  $n$  shelves are, then  $F(a)$  is called the comprehensive correlation value of layout  $a$ , the formula is expressed as:

$$F(a) = \sum_{i=1}^n f(i | a) \quad (6)$$

The comprehensive correlation value can reflect the lifting effect of the entire layout. Therefore, based on the adjacency effect value, this paper establishes a mathematical model for the comprehensive correlation value of the comprehensive product layout. According to equation (4) lifting degree matrix and equation (5) adjacency effect value expression, a quantitative expression is given as follows:

$$\begin{aligned} \max F(a) &= \sum_{i=1}^n \left[ \frac{1}{k} \sum_{j=1}^k L(a_i \rightarrow a_j) \right] \\ \text{s.t. } &\begin{cases} a \in M \\ i \in 1, 2, \dots, n \end{cases} \end{aligned} \quad (7)$$

Among them,  $n$  indicates the number of shelves,  $k$  indicates the degree of freedom for placing shelves  $i$ , and  $M$  indicates all possible layout situations.

This article studies the distribution of supermarket commodity with the purpose of maximizing the comprehensive correlation value. This commodity layout model has the following characteristics in practical applications: (1) The model uses the degree of freedom of placement and the commodity correlation degree as observation index to calculate the value of adjacency effects when the goods are placed at different positions. Obtain quantifiable specific results, avoid the problem of bias in evaluation parameters due to strong subjectivity, and improve rigor. (2) This model describes the relationship between locations from four directions, and the adjacency effect value is proposed. Using the association rules of data mining to select the goods with the maximum value of location adjacency effect, and then match the commodities with the positions to obtain the optimal commodity layout. So that the relationship between the goods is closer, as long as commodities are adjacent, they must be related, and the whole layout is more systematic. However, the layout method in this paper involves a large number of shelves, each shelf has a different  $k$  value, has a large amount of calculation,



and is difficult to complete by manual calculation. Therefore, the genetic algorithm is used to solve the model.

#### 4 Solving the model of product layout by maximizing comprehensive correlation values on genetic algorithm

Genetic Algorithm is a computational model that simulates the natural selection and genetic mechanism of Darwin's biological evolution. It is a method to search for the optimal solution by simulating natural evolution. The genetic algorithm can directly operate on the structural objects, and adopts a probabilistic optimization method, which can adaptively adjust the search direction, and has parallelism and better global optimization capabilities. Because of the large number of parameters of the commodity layout model with the maximum adjacency effect, the complexity of the algorithm is high, and the difficulty of manual calculation is large, the genetic algorithm is used to solve it. The solution process is as follows:

##### 4.1 Encoding

It adopts integer permutation coding. For the combinatorial optimization of  $n$  commodities, the chromosomes are divided into  $n$  segments, each of which is the corresponding commodity number. For example, the combination optimization problem of 16 kinds of commodities is  $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16\}$ , then  $|1|3|2|4|16|6|9|11|7|10|8|12|13|14|15|5|$  is a valid chromosome.

##### 4.2 Population initialization

After completing the chromosome encoding, an initial population must be generated as the starting solution, so first of all, we need to determine the number of initialization population. The number of initialized populations is generally based on experience. In general, the number of populations depends on the size of the commodity type, and the value of case in this paper is 30.

##### 4.3 Fitness Function

Let  $|1|2|\dots|i|\dots|n|$  be a chromosome with integer coding,  $L_{ij}$  is the correlation value of commodity  $i$  to commodity  $j$ , then the fitness formula of the individual is  $fitness = \sum_{i=1}^n f(i|a)$ , it means that the fitness function is the sum of the adjacent effect values of each commodity, and the goal of the optimization is to select the chromosome with the largest fitness function value.

The larger the fitness function value, the better the chromosome, and vice versa.

##### 4.4 Select operation

The selection operation is to select individuals to a new population with a certain probability from the existing population. The probability that an individual is selected is related to the fitness value. The greater the individual fitness value, the greater the probability of being selected. This article uses the roulette selection method.

##### 4.5 Cross operation

Using partial mapping hybridization, determine the parent of the cross operation, and divide the parent sample into two groups, and repeat the following process for each group (assuming the number of products is 16):

1) Generate two random integers  $q_1$  and  $q_2$  in the interval  $[1,16]$ , determine two positions, and cross the middle data of the two positions, such as  $q_1=5, q_2=10$

1 3 7 2 |11 15 10 16 9 5| 6 12 8 14 13 4  
 3 4 7 8 |3 9 10 11 12 5| 6 13 1 14 15 2

After crossing it becomes:

1 \* 7 2 |3 9 10 11 12 5| 6 12 8 14 13 4  
 3 4 7 8 |11 15 10 16 9 5| 6 13 1 14 \* 2

2) After crossing, in the same individual, there are duplicate product numbers, and the non-duplicate numbers are reserved. The conflicting numbers (with \* position) are eliminated by partial mapping, that is, mapping is based on the corresponding relationship of the middle segment. The result is:

1 15 7 2 |3 9 10 11 12 5| 6 12 8 14 13 4  
 3 4 7 8 |11 15 10 16 9 5| 6 13 1 14 12 2

##### 4.6 Mutation operation

The mutation strategy takes two points randomly and swaps them. Generate two random integers  $q_1$  and  $q_2$  in the range  $[1,16]$ , determine the two positions, and swap them, such as  $q_1=6, q_2=13$ :

1 15 7 2 3 |9| 10 11 12 5 6 12 |8| 14 13 4

After mutation it becomes:

1 15 7 2 3 |8| 10 11 12 5 6 12 |9| 14 13 4

##### 4.7 Reverse operation

In order to improve the local search ability of genetic algorithms, continuous multiple evolution reversal operations are introduced after selection, crossover, this operation added to accelerate evolution. Mutation. "Evolution" here refers to the unidirectionality of the reversal operator, it means that only after the reversal, the fitness value will be accepted, otherwise the reversal will be invalid. Generate two random integers  $q_1$  and  $q_2$  in the interval  $[1,16]$ , determine the two positions, and swap them, such as  $q_1=4, q_2=9$ :

1 15 7 |2 3 9 10 11| 12 5 6 12 8 14 13 4

After the reversal it becomes:

1 15 7 |11 10 9 3 2| 12 5 6 12 8 14 13 4

Perform cross mutation for each individual, and then substitute it into the fitness function for evaluation. Select individuals with large fitness values to perform the next generation of crossover, mutation, and evolution reversal operations. Perform a loop operation. Determine whether the set maximum generation number MAXGEN is met, if not, then enter the calculation of fitness value, otherwise, the genetic operation is ended.

## 5 Case analysis

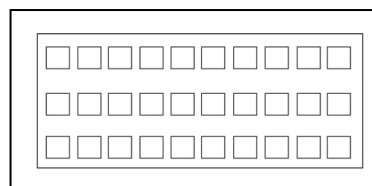
In order to meet people's growing demand, the scale and number of supermarkets have increased significantly in recent years, and industry competition has become increasingly fierce. However, there is not a big gap between the operating models and promotional methods of supermarkets, which has caused supermarket sales to be tested. In order to develop well, B supermarket urgently needs to change its business strategy and improve its competitiveness and turnover. Scientific and reasonable product layout can affect consumers' consumption behavior, thereby increasing supermarket sales. Therefore, in the layout of products, this article considers the impact on customer consumption psychology, changes the traditional product layout, and proposes a new product layout method to stimulate consumers to change their initial consumption behavior, thereby helping supermarkets increase consumption. In addition, the cost of optimizing the product layout is relatively small, the implementation difficulty is relatively low, and it can produce the effect relatively quickly, which has certain feasibility.

B Supermarket focusing on food and daily necessities. It is located at 188 Zhongshan East Road, northeast of the intersection of Ping'an North Street and Yuhua East Road. It belongs to the core business district. It is adjacent to the civic center and the first block of finance and commerce, and the transportation is convenient. There are many high-end communities such as Huicui Garden, Chang'an Flower Garden, Zhongji Liyu, Yinhong Flower Garden and so on, with a large and very stable consumer group, the purchasing power and purchasing demand in the region are guaranteed. In this paper, we randomly collected part of the shopping data of B medium-sized chain supermarket in Shijiazhuang City, Hebei Province in the past month, including 58 attributes (including customer number, shopping time, shopping amount and 55 kinds of commodities), with a total of 2877 transaction records. The attribute value "0" indicates that the customer did not purchase the product in this transaction, and the value "1" indicates that the product was purchased. After necessary processing of the original data, the product categories are integrated into 30 types, and each piece of data is a record of different customer purchases, that is, a shopping basket, and the number of product types purchased by each customer at a time becomes 1-30. In the mining software Clementine12.0, the Apriori

algorithm was used to mine association rules. The rule support degree was set to be greater than 10%, the confidence degree was greater than 20%, and 988 association rules were run.

### 5.1 Formal description of supermarket space

Supermarket B with a total area of 3255 square meters and a business area of 1955 square meters. Protective products, cosmetics and other products, about 15,000 varieties of products. Although the supermarket is a three-dimensional structure, in fact, the behavior of shoppers is mainly on a two-dimensional plane. The standardized supermarket form description is shown in Figure 5:



**Figure 5.** Standardization plan of B supermarket

### 5.2 Goods location optimization process

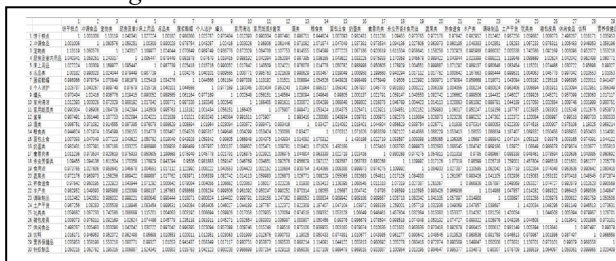
1) *Classification of goods:* Based on the analysis of the historical consumption data of the supermarket, this chapter proposes to carry out calculation analysis on the major commodity categories. According to the product function and use attributes, the categories are classified to solve the problems of small analysis granularity and high calculation complexity caused by analysis on specific products. All commodities are divided into 30 categories. Integrate the initial data according to the category and number the products. The product classification is shown in Table 1:

**TABLE I.** CLASSIFICATION OF OPERATING COMMODITIES

Classification of operating commodities			
Numbering	Product Types	Numbering	Product Types
1	Biscuits and cakes	16	Egg and dairy products
2	Prepare food	17	Avian meat
3	Pet Supplies	18	Insecticidal aromatics
4	Kitchenware and housewares	19	Cooking oil
5	bed linings	20	Fruits and vegetables
6	Frozen products	21	Instant cooked food
7	Clothing and shoes	22	Aquatic products
8	Personal cleaning	23	Seasonings

Classification of operating commodities			
Numbering	Product Types	Numbering	Product Types
9	Canned food	24	Local specialty dry goods
10	Household cleaning	25	Toys
11	Household paper	26	Luggage leather goods
12	Pickles	27	Snack foods
13	Alcohol	28	Drink
14	Foodstuff	29	Nutrition and health products
15	Bread staple food	30	Knitwear

2)Organize the association matrix: Based on historical consumption data, perform data mining operations to analyze the correlation between products. Clementine 12.0 is a general-purpose data mining software, which can help users build a complete data mining process, and provides a series of functions so that users can perform any of these mining steps. For the processed shopping data, use Clementine 12.0 to establish the Apriori algorithm model to find the association rules between products. The process mainly includes establishing data, establishing a process, importing data, setting parameters, exporting results, and analyzing. According to the promotion degree of each commodity, the association matrix is sorted out, as shown in Figure 6:

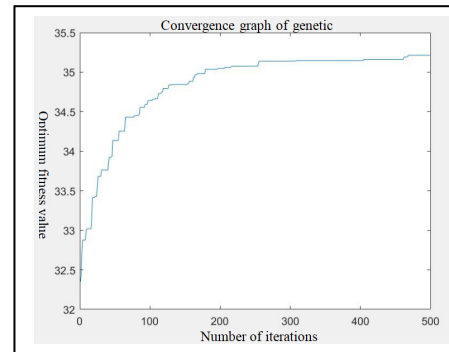


**Figure 6.** Association matrix

3)Experimental results: The algorithm design and coding involved in this article are written by MATLAB. The data input of the algorithm is: (1) the two-dimensional correlation matrix of goods (30 \* 30 matrix); (2) the layout of goods shelves; Set the layout of the shelf space in the program, substitute the correlation matrix arranged in the number sequence into the program, and run the program. First generate a random set of initial solutions, so we get a set of unoptimized random genes:  
 13→8→18→25→28→21→4→5→2→  
 >12→9→17→30→15→26→7→24→29→  
 >1→10→22→23→3→6→27→16→20→  
 >11→19→14

Calculate the comprehensive correlation function value of the unoptimized genes, that is, the sum of all adjacent effect values, unoptimized comprehensive

correlation value function value is 30.9144. In the experiment of supermarket product layout, the specific supermarket product layout is randomly generated and is not repeated. Continue to run the program to get the convergence graph of genetic algorithm, as shown in Figure7:

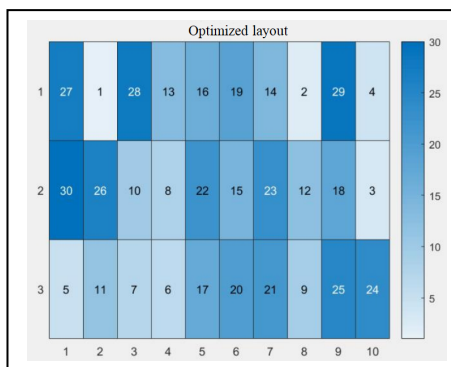


**Figure 7.** Convergence diagram of genetic algorithm optimization

In different supermarket product layout, the value of adjacency effect is calculated based on the historical consumption data of the supermarket, and the comprehensive correlation value under each supermarket product layout is obtained. By comparing and analyzing the layout with the largest comprehensive correlation value, we consider this layout to be the optimal convergence process of the model. It can be seen that when the number of iterations is 256, the optimal fitness value tends to be stable, and the comprehensive correlation function value is 35.14. When the number of iterations reaches 470, the optimal fitness value no longer increases and reaches the maximum, the comprehensive correlation function value is 35.21, which is the optimal solution to the problem. At this time, a group of genes optimized by genetic algorithm can be obtained:

27→30→5→1→26→11→28→10→7→  
 >13→8→6→16→22→17→19→15→20→  
 >14→23→21→2→12→9→29→18→25→  
 >4→3→24

The value of the comprehensive correlation value function of the gene after genetic algorithm optimization is 35.2138, which is the global optimum. At the end of the program, the optimized product layout is obtained. The darker the color is, the greater the value of adjacency effect is, and vice versa. The optimal product layout of the supermarket can be obtained by matching the products represented by the numbers with those in the result chart one by one. The optimal product layout is shown in Figure 8:



**Figure 8.** Optimal layout results of genetic algorithm

4) *Conclusion:* As can be seen from Figure 9, it is relatively successful to optimize the layout of supermarket products based on the comprehensive correlation value of the products. Under the initial random production layout, the comprehensive correlation value of the product layout is 30.9144, while the relatively optimal product layout has a comprehensive correlation value of 35.2138. the difference between the two is 4.2994, and the difference ratio between the two before and after optimization is about 14%, which shows that the optimization effect is very good. It can effectively optimize the supermarket layout, and get the relatively best supermarket product layout.

## 6 Conclusions and research prospects

The purpose of this article is to provide a constructive plan for the layout of supermarket products. By designing a reasonable product layout, it will affect consumers' consumption behavior, thereby increasing supermarket sales.

This article first analyzes the impact of goods on the surrounding goods due to different cargo locations, and put forward the concept of placing degrees of freedom, considering the impact of goods on other goods in multiple directions, the overall layout is more closely related. A commodity network with adjacency relation is constructed to make suggestions for commodity layout. By understanding the purchasing habits of customers, the mutual influence of product sales is expanded, and customers are guided to continue to purchase products, thereby driving the sales of products with less sales. Then, according to the different degrees of freedom of placement, the model of maximizing the comprehensive correlation value is established, calculate the adjacency effect value of each product and use genetic algorithm to solve it, and get the comprehensive set of related values under each supermarket product layout. Through comparative analysis, the layout with the largest comprehensive correlation value is obtained. After determining the layout of the product, it is mapped to the corresponding space area to form the final supermarket

product layout. In this way, when customers choose products by themselves in the supermarket, they will go to the products they most want to buy, and the related products are around them, which gives customers a kind of implicit guidance, arouses customers' interest in purchasing other products, thus promoting the sales of products and playing a role of commodity recommendation. At the same time, the quantitative method is adopted to avoid the deviation of evaluation parameters due to the strong subjectivity, which is more rigorous. The experimental results show that the method in this paper can successfully design a reasonable supermarket layout, and the layout design algorithm is simple and efficient. At the same time, related products drive sales to each other, and increasing sales brings economic benefits to the enterprise. The data analyzed in this article comes from consumer shopping information of B supermarkets. Due to the small amount of data extracted and a certain degree of randomness, the results obtained may not be accurate and perfect. However, the mining principle is basically the same, and the results have certain reference and guidance significance.

## Acknowledgment

This work is supported by the National Natural Science Foundation of China (71771078, 71371064)

## References

1. Kacen J J, Hess J D, Walker D. "Spontaneous selection: The influence of product and retailing factors on consumer impulse purchases." *Journal of retailing and consumer services*, 2012, 19(6), pp. 578-588.
2. Zhao Feng, Wang Ze. "Research on The Application of Improved SLP and AHP Methods in Supermarket layout." *Logistics Science and Technology*, 2018(9), pp. 13-17.
3. Liu Zhihai, Zhang Dandan. "Supermarket Layout Optimization Based on Improved SLP." *Logistics engineering and management*, 2017(3), pp. 117-120
4. FAYYAD U. "From Data Mining to Knowledge Discovery in Databases." *AI Magazine*, 1996, 17(3), pp. 37-54.
5. HE X, MIN F, ZHU W. "Parametric Rough Sets with Application to Granular Association Rule Mining." *Mathematical Problems in Engineering*, 2013, 2013(2), pp. 1-13.
6. Francisco Azuaje, Witten IH, Frank E. "Data Mining: Practical Machine Learning Tools and Techniques 2nd edition." *Biomedical Engineering Online*, 2006, 5(1), pp. 51.
7. Li Xiao, Chen Wenbin, Cheng min. "Supermarket Layout Design Algorithm Based on Shopping Data." *Digital design*, 2016, 10(5), pp. 36-40.
8. Wang Tingting, Yang Qin. "Optimal Scheduling of Supermarket Cash Registers Based on Simulated



- Plant Growth Algorithm.” Computer application, 2014,34(5), pp. 1516-1520.
9. Andrade E B. “Behavioral Consequences of Affect: Combining Evaluative and Regulatory Mechanisms.” *Journal of consumer research*, 2005, 32(3), pp. 355-362.
  10. Zhang Yunlai. “Research on Integration of Impulsive Purchase Influence Mechanism Based on Emotional Perspective.” *Journal of Harbin Business University (SOCIAL SCIENCE EDITION)*, 2009(04);, pp. 3-7.
  11. Jia Tongling. “Store Line and "Magnetic Point" Design.” *China Cooperation Times*, 2014-06-06 (B03).
  12. Gu Guangsheng. “"Magnetic Point" Theory and Display Skills of Small and Medium-sized Supermarkets: A Case Study of Meizhou Pacific Supermarket.” *Commercial Research*, 2007 (06);, pp. 144-147.
  13. Masao Ohta, Yoshiyuki Higuchi. “Study on the Design of Supermarket Store Layouts: The Principle of “Sales Magnet”.” *World Academy of Science, Engineering and Technology Vol:7* 2013-01-26.
  14. R. Agrawal, T. Imielinski, A. Swami. “Mining Association Rules between Sets of Items in Large Databases”. Y.L.Diao. *Acm Sigmod Record*, Washington DC, May 26-28, 1993, New York: ACM Press, 1003, pp. 207-216.
  15. Hui, S.K, Fader, P.S. Bradlow. “Research note-The traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality.” *Marketing Science*. 2009, pp. 566-572.
  16. X. Wenjie, L. Senbiao, H. Lihong.” Simulation model of supermarket queuing system.” *Chinese Control Conference (CCC)*. Hangzhou, 2015, pp. 8819-8823.
  17. C. Yanfei, Z. Yafeng. SWOT Analysis on the Supermarkets in China[A]. *International Conference on Business Computing and Global Informatization[C]*. Shanghai, 2012, pp. 206-209.
  18. Huang Xigeng. The Role of Magnetic Dots in The Display Design of Drugstores [N]. *21st Century Pharmacy*. 2017-11-20 (B02).
  19. Yu Liqiong. “Exploring The Combination of Regional Shopping Mall Formats Based on Consumer Preferences”. *Commercial Economy Research*, 2018 (1): 47-49.
  20. Péter Boros, “Orsolya Fehér, Zoltán Lakner, e.l. Modeling supermarket re-layout from the owner’s perspective.” *Annals of Operations Research*, 2015, 238(1), pp. 60-63.
  21. Xiong Ping. “Data Mining Algorithm and Clementine Practice.” Beijing: Tsinghua University Press, 2011.
  22. Kacen J, Hess J D, Walker D. “Spontaneous selection: The influence of product and retailing factors on consumer impulse purchases.” *Journal of retailing and consumer services*, 2012, 19(6), pp. 578-588.
  23. Iris I M. “In-store impact on impulse shoppers.” *Marketing*, 2002,34(2), pp. 357-364.
  24. Neha P. Mehta, Pawan K. Chugan. “The Impact of Visual Merchandising on Impulse Buying Behavior of Consumer: A Case from Central Mall of Ahmedabad India.” *Social Science Electronic Publishing*., 2011, 57 (3), pp. 599-610.
  25. Zhang Shenghua, Li Jiahui, Li Na. “Application of Association Rule Analysis in the Commodity Layout of Supermarkets.” *SME Management and Science and Technology*, 2019 (05), pp. 61-63.
  26. Mowrey C H, Parikh P J, Gue K R. “A Model to Optimize Rack Layout in a Retail Store.” *European Journal of Operational Research*, 2018, 377(2), pp. 71-83.
  27. Volpe R, Cho C. “Market structure determinants of performance for independent supermarkets.” *Managerial and Decision Economics*, 2019, 40(8), pp. 871-881.
  28. Bill Page, Giang Trinh, Svetlana Bogomolova. “Comparing two supermarket layouts: The effect of a middle aisle on basket size, spend, trip duration and endcap use.” *Journal of Retailing and Consumer Services*, 2019, 47(1), pp. 49-56.
  29. Elif Ozgormus, Alice E. Smith. “A data-driven approach to grocery store block layout.” *Computers & Industrial Engineering*, 2020, 139(10), pp. 55-62.
  30. Jean-Paul Arnaout, Caline ElKhoury, Gamze Karayaz. “Solving the multiple level warehouse layout problem using ant colony optimization.” *Operational Research*, 2020, 20(1), pp. 473-490.