ISSN: 2688-8653

Analysis of Commodity Evaluation Mechanism based on Regression Analysis

Qianqian Chen

Faculty of science, Hangzhou Normal University, Hangzhou 311121, China

1475874017@qq.com

Abstract

Online shopping has become the trend of today's society, shopping platform is constantly improving, personal rating, customer evaluation, helpful votes...They are important factors in the sale of goods, which is the focus of attention of the merchants. Therefore, star rating, reviews, helpful votes and other factors are crucial to the success of a product. In this article, we aim to determine the relationship and impact between star rating, reviews and helpful votes. In the meanwhile, we hope that we could use our analysis results to provide Sunshine company with our recommendations. Firstly, We clean the data and make a preliminary processing of the review text by referring to the idea of NLP. Next, we analyze the relationship among star rating, reviews and helpful votes by using the analytic hierarchy process(AHP) to analyze their influence on customers' purchase of products. After that, based on the previous preliminary analysis and a large number of data, we use the regression analysis model to estimate the linear relationship between comment rating and star rating. It is concluded that the key factors influencing the quantity of product sales are comment rating and star rating. Secondly, we use the time series model to analyze the change of rating in time dimension. Based on the trend of favorable rating in time, it is further predicted that the product's reputation in the online market is rising or falling.

Keywords

AHP; Regression Analysis; Reviews and Star Rating; Time Sequence.

1. Main Body

With the coming of the information age, great changes have taken place in our lifestyles, such as shopping mode. On the one hand, most young people now prefer to shop online through some platforms. They can shop around when they buy, and get what they want without being restricted by geography with the help of developed and convenient logistics. On the other hand, businesses need to face a bigger test. Because there are many factors would affect consumers shopping online, it is important to recommend appropriate products according to customers' needs accurately. So, it is necessary to rely on the mining of big data to obtain information, to put forward shopping Suggestions that meet customers' requirements.

Amazon's online marketplace offers customers the ability to rate and rate their purchases, and other users can submit helpful or unhelpful ratings in those reviews to help them make their own product purchase decisions. By analyzing the data of this evaluation mechanism, merchants and producers can have an in-depth understanding of customers' shopping needs, as well as the product's design function selection and market prospect.

ISSN: 2688-8653

2. Model of Qualitative Analysis Data

2.1. Data Cleaning and Filtering

2.1.1. Analysis of Star Rating and Review's Helpful Votes

Due to the large amount of data, we initially screened the data when processing. In order to consider the helpfulness rating, we define the helpful review which meet the requirement $V > 0.6$ and $V_{helpful} \geq 10$. Avoid brushing, malicious comments and other phenomena, we use the helpful review as our standard and require the verified purchase line is Y, and then pick out initial effective data. After that, we could get the following conclusions:

(1) It is observed that the star rating of most customers on the products more than three star from the figure 1. this shows that most customers are satisfied with the product.

(2) From the figure 2, there are a certain number of reviewers support the review, which shows the reliability and rationality of data screening.

(3) It can be seen from figure 2 that the higher the star rating of the product is, the more the corresponding number of helpful reviews will be. On this basis, we can guess that there is a certain relationship between the star rating and the number of helpful reviews.

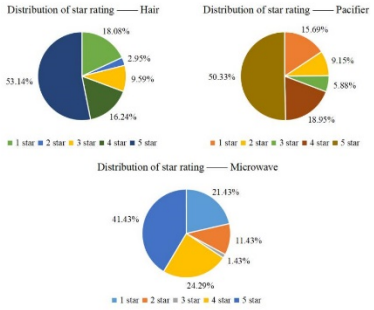

Figure 1. Distributions of star rating

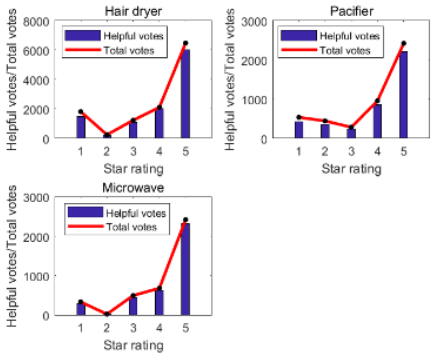

Figure 2. Relationship of star rating and helpful votes/total votes

2.1.2. Extract Comment Text Information with Keywords

Natural language processing (NLP) is a theory or method that can realize the effective communication between human and computer in natural language. Natural language processing is divided into two stages: natural language understanding and natural language generation. However, due to the complexity of word formation and the influence of different situations in linguistics, it is easy to cause ambiguity during the transformation.

In order to simplify our data analysis, we decided to analyze the review data by extracting keywords based on the idea of natural language processing. And we also consider the star rating, helpful votes and the review that from the members of Amazon Vine program. We filter the text

ISSN: 2688-8653

content of review_headline, and put the key words with "five stars", "fine", "great", "Nice" as a good comment. In the same way, it is considered bad to include "awesome", "one start" and other keywords. For those that don't include good or bad keywords, we can think of this as a general comment. From this, we can replace those with good comment with "1", and the bad comment with "0", and the general comment with "0.5".

2.1.3. Conclusion and Analysis

After cleaning and screening of the data, as all types of data show the same trend of change, we can get product stars, helpful reviews and review text content have a relatively close relationship with product sales.

In order to better explain the effect of these factors on customers' purchase of products, we use analytic hierarchy process (AHP) to establish a product purchase analysis model. We hope to get the influence weight of each factor on the result and put forward the sales suggestions for sunshine company.

3. Product Selection Model based on AHP

3.1. Introduction of the Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) is a decision-making method that decomposes the elements related to decision making into levels such as objectives, criteria and plans, and conducts qualitative and quantitative analysis on this basis. It is characterized by the essence of complex decision problems, and the use of less quantitative information to make the thinking process of decision mathematical, to provide a simple decision method for complex decision problems.

It is especially suitable for situations which it is difficult to measure the result of decision directly and accurately, and its basic steps are as follows:

Step 1 Build a hierarchical model.

Step 2 Construct a pairwise comparison matrix.

Step 3 Structure judgment matrix.

Step 4 Calculate weight vector.

Step 5 Test the consistency.

3.2. Determine the Weight based on the Analytic Hierarchy Process (AHP)

(1)Build a hierarchical model

By referring to the paper and analyzing the data, we determined four factors of the criterion layer, which are star rating, suitable review, product price, review's rating.

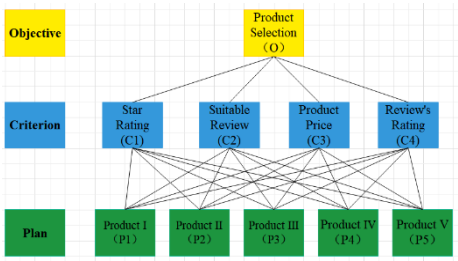

Figure 3. The hierarchical model of product option

(2) Determine the weight of each index

In order to determine the importance of each indicator to the system evaluation process , we assign values of importance between each index element. We can get the objective layer - the criterion layer discriminant matrix.

ISSN: 2688-8653

Table 1. Discriminant matrix O-C

O	C1	C2	C3	C4
C1	1	2	4	2
C2	1/2	1	4	3
C3	1/4	1/4	1	2
C4	1/2	1/3	1/2	1

Average random consistency indicator as shown in Table 2 :

Table 2. Average random consistency indicator RI value table

Order	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49

(3) Comprehensive evaluation

According to the analytic hierarchy process (AHP), We can get the weights of the factors from the discriminant matrix which is: ω= 0.3773, 0.3610, 0.1393, 0.1224.

After calculating that we can get the consistency test index CI is 0.0276, and the consistency ratio of evaluation matrix O-C is 0.0311. Both of them are less than 0.1, this result complies with the corresponding consistency test standards.

(4) Calculation of Product Selection Considerations

Combining with the weight of the four factors of product selection rating by the analytic hierarchy process (AHP), we can obtain the product selection rating table of microwave, hair and pacifier. we take microwave for an example as shown in Table 3.

Table 3. Product Selection Rating--microwave

Criterion	ω	Danby (P1)	Samsung (P2)	Sharp (P3)	Whirlpool (P4)	Others (P5)
Star Rating (C1)	0.3773	0.1974	0.2265	0.1893	0.1850	0.2017
Suitable Review (C2)	0.3610	0.2252	0.2523	0.1502	0.1682	0.2042
Product Price (C3)	0.1393	0.1765	0.1765	0.2353	0.2353	0.1765
Review's Rating (C4)	0.1224	0.1944	0.2222	0.1944	0.1944	0.1944
Total	1	0.2042	0.2283	0.1822	0.1871	0.1982

(5) Result and the analysis

According to the weight of each factor, we can get that the star rating and the suitable review of product evaluation have a greater impact on customers' purchase, and the product price and review's rating also have a certain impact.

From the above product selection evaluation form, we can get the relative burning brands according to the total score, such as Samsung microwave, Remington hair and Fisher pacifier.

ISSN: 2688-8653

But the scores show a similar level of competition among brands, with some growing brands among others.

During the process of data processing, it is not difficult to find that the greater the influence of a brand is, the better its product rating, price and review rating are. Therefore, we would like to suggest sunshine company to pay attention to the choice of product brand when selling its products, so as to expand its impact and attract customers by increasing advertising. Nevertheless, at the same time also should pay attention to the reasonable price, product quality control.

4. Regression Model based on Time Sequence

4.1. Linear Regression Equation

In order to give you a better reasonable suggestions on the number of basis. After grasping the general sales data of the product, we selected several factors which may be related to the above data for SPSS regression analysis. After comprehensive consideration of VIF, R2 and the significance level of each factor, we selected the following four influencing factors and calculated their coefficients as follows:

$$y = \beta_0 + \beta_1 \times S + \beta_2 \times G + \beta_3 \times P + \beta_4 \times V$$

$$\beta_1 = 0.327, \beta_2 = 0.534, \beta_3 = 0.186, \beta_4 = 0.096$$

Here, S stands for star rating, G for praise rate, P for price, V for helpful votes.

From this formula, we can make it clear that praise rate and star rating are the two decisive factors for the sales volume of products. We estimate that there is a strong correlation between reviews and star rating (which can also be obtained from the correlation analysis below), so we believe that the word of mouth and reputation of products are the decisive factors to describe the success and failure of products.

Given the number of good reviews and the total number of merchandises sold that year, we get an annual report of good reviews. The following is a specific analysis of praise rate in time dimension to see the change of its sales volume. It is generally believed that the topics related to the microwave, hair dryer, and pacifier do not have quarterly changes, and we can take a year as an interval for statistics. Therefore, taking the hair dryer as an example, 8 brands were randomly selected from the brands that users bought more from 2009 to 2015, and the relative purchase quantity of these brands was calculated. The following distribution is obtained.

The figure shows that can be concluded that every popular brand has a lot of volatility. Taking the Mangroomer as an example, we can only observe its listing activity in 2012-2013, and as time goes on, more and more new products will occupy the market, such as Panasonic.

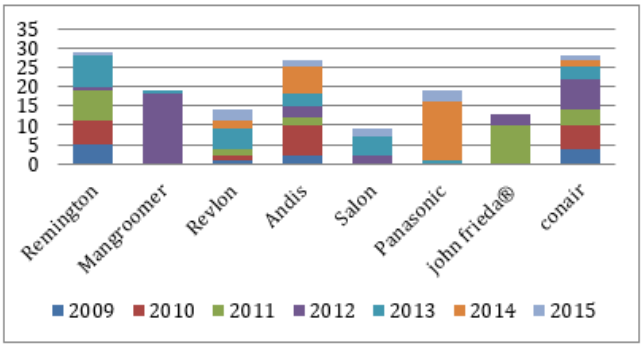

Figure 4. Relative brand purchases ,2009-2015

ISSN: 2688-8653

4.2. Time Series Analysis

In order to analyze the change of praise rate based on time dimension, we adopted time series analysis. Since the information contained in the data varies from year to year, the Figure 4 shows that the recent data contains more information about the future. Therefore, in order to both consider the data of each period and highlight the importance of the recent data, we give greater weight to the recent data, which is the weighted moving average method. Define the following formula:

$$N_{t\omega} = \frac{\omega_1 y_1 + \omega_2 y_t + \cdots + \omega_M y_{t-M+1}}{\omega_1 + \omega_2 + \cdots + \omega_M}, t \geq M$$

$$\hat{y}_{t+1} = N_{t\omega}$$

Where $N_{t\omega}$ is the t period weighted moving average;ω_i is the weight of y_{t-i+1},which reflects the importance of the corresponding y_t in the weighted average. Take Revlon for example:

Table 4. Annual report of praise rate of Revlon ,2006-2015

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Revlon y_t	0.534	0.763	0.286	0.923	0.899	0.534	0.038	0.669	0.582	0.637
Praise rate $\hat{y}_{t+1}$				0.4863	0.6480	0.8048	0.7205	0.3468	0.4362	0.5203
$\Delta y_i(\%)$				0.4731	0.2392	0.5072	17.9605	0.4816	0.2506	0.1832

After substituting the data into the formula:

$$\hat{y}_{t+1} = \frac{3y_1 + 2y_{t-1} + y_{t-2}}{3 + 2 + 1}$$

$$\hat{y}_{2016} = \frac{3 \times 0.637 + 2 \times 0.582 + 0.669}{3 + 2 + 1} = 0.624$$

For example, the relative error in 2009 is 0.4731. By analogy, we can use the formula to obtain the relative error from 2009 to 2015 and calculate the average.

$$\left(1 - \frac{\sum \hat{y}_t}{\sum y_t}\right) \times 100\% = 0.0661$$

It can be estimated that the average of the total predicted value is 6.61% lower than the actual value, so the predicted value of 2016 can be revised to 0.6682.

By analogy, we can get the forecast value of other commodities' praise rate respectively in 2016.

Table 5. The praise rate of prediction

	Remington1	Mangroomer	Andis	Panasonic	john frieda®	Conair
2016	0.5648	0.0952	0.9060	0.8366	0.0401	0.7333

From the distribution of the data, we can predict that Andis is the brand with the highest praise rate in the next year. Combining with the past favorable rating, we can conclude that its reputation is on the rise, and Remington predicts that its reputation is on the decline. Results

ISSN: 2688-8653

we can draw from the data is under one year of the goods is in nearly three years relative purchases a large number of data base, so regardless of whether or not active for more than ten years before, we think that have good credit in the next year.

References

[1] Xu yingnan, ganli ren. Analysis of knowledge support in online commodity selection decision of consumers based on recommendation service [J]. Intelligence theory and practice, 2013,36 (03) : 107-111 + 116.

[2] Zhao jingsheng, song mengxue, gao xiang. Development and application of natural language processing [J]. Information technology and informatization, 2019 (07) : 142-145.

[3] Xu Jun, Liu Na. Basic thought and practical application of analytic hierarchy process [J]. Information exploration, 2008 (12) : 113-115.

[4] Yang luming, li qinyun. Research on influencing factors of mobile consumer shopping [J]. Academic exploration, 2016 (05) : 97-103.

[5] Li jianchao. Research on the usefulness of online commodity evaluation [D]. Jinan university, 2018.17-18.

[6] Si shoukui. Mathematical modeling algorithm and program [M]. Naval aeronautical engineering college: 475-478.