A New Way to Evaluate Products with Reviews: A Case Study of A Hair Dryer on Amazon Platform

Xiting Jin, Xiangdong Liu* and Zhan Xiangsheng Jinan
University, Guangzhou, GuangDong, China
E-mails: jxting0622@163.com, tliuxd@jnu.edu.cn

Abstract. The era of big data is coming. How to make effective use of users' feedback information data is the key to success in various online industries. In order to help target enterprise enter the market, we build a comprehensive rating evaluation model that can evaluate the sales status, future development and contain comprehensive indicators. We firstly study the one-dimensional index. We adopt the method of correlation analysis to find the ripple effect of star rating. Then we use SPSS to draw scatter plots to analysis correlation between specific quality description and star rating level. We found through the research that “star rating”, “comment validity”, “positive and negative indicator in comments” and “member effect” are all informative for target enterprise to track, and will determine the market development of the product to a certain extent. We take advantage of the machine learning to turn the information in the text into quantifiable metrics and get polarity and subjectivity. Use NLTK for sentiment analysis of each comment to get negative, positive, neutral and compound value. Then use AHP to assign weight to the indicators wisely. With the comprehensive rating evaluation model that we build, we calculate the comprehensive rating value of all products in this industry and take the median as the benchmark value, then the products with the comprehensive rating value higher than the benchmark value can be regarded as successful products. We provide an accurate assessment. In order to effectively evaluate and predict the future sales trend of new products after their launch and effectively measure the increase and decrease of product market reputation, we establish a time-based measurement and prediction model based on Grey model theory. We can use the model to make comprehensive evaluation and effective prediction of the target company's products. After inspection, our model can be effectively applied to different sales fields, which has great practical significance.

Keywords. Machine learning; Grey model; AHP; NLTK

1. Introduction
In recent years, people are more and more inclined to shop online while the online sales platform represented by amazon has become more and more popular. Therefore, making good use of users' feedback information on the platform is conducive to the company's effective evaluation of products. Meanwhile, the company can make targeted adjustments to customers' demands, so as to produce products more in line with the needs of the public and ultimately help the company earn more profits.

In this paper, a comprehensive rating evaluation model will be established and used to extend to different commodity fields. And take a hair dryer on Amazon platform as an example to illustrate and explain the model.
2. Analysis
Starting from the one-dimensional level index, we analyse the correlation between two indexes, such as peer comments, comment sentiment and star rating. Then identify data measures that are most informative for Company to track, and the metrics that best provide product and market information.

Then we build a comprehensive rating evaluation model that can evaluate the sales status, future development and contain comprehensive indicators. And based on the model, a time-based measurement and prediction model is established to effectively evaluate and predict the future sales trend of new products after their launch and effectively measure the increase and decrease of product market reputation.

Lastly we summarize and comprehensively analyse the advantages and disadvantages of some competitive products, and analyse useful conclusions for target company to help them achieve success in the new online market products. style to the title page. This paragraph follows a section title so it should not be indented.

3. Assumptions and Justifications
- Each comment reflects a true consumer assessment of the product.
- The same product has the same characteristics.
- The sentiment analysis of the library on the review text can quantitatively reflect the consumers’ emotional attitude towards the product.
- Comment validity and buyer rating are only related to the metrics mentioned in the algorithm.

4. Notations

| Symbols | Definition |
|---------|------------|
| λ       | Comment validity |
| H_r     | Helpfulness rating number for comments |
| T_r     | Total number of votes |
| φ       | Weight given to voting comments |
| L       | Emotional intensity |
| N_i     | Comment words’ number |
| N_a     | Number of words that reflect emotion |
| E       | Member effect |
| a_1     | Positive indicator |
| a_2     | Negative indicator |
| ω       | Buyer’s score |
| v       | Star rating |
| Z       | Comprehensive rating value |

5. One-dimensional index analysis
In order to build a comprehensive rating evaluation model that can evaluate the sales status, future development and contain comprehensive indicators, we firstly study the one-dimensional index.

5.1. The ripple effect of star rating
Here we first investigate the ripple reaction effect of star rating, and propose the following questions to be studied: the impact response among star reviews and the influence reaction between one-star reviews and five-star reviews.
5.1.1. Methods
The key and starting point of systematic analysis are how to correlate factors and how to quantify the degree of correlation. We know that the basic method of factor analysis used to be regression analysis. In addition, the method of regression analysis is computationally heavy and abnormal. In order to overcome the above disadvantages, we adopt the method of correlation analysis to do system analysis. As a changing system, correlation analysis is actually a quantitative comparative analysis of the development trend of dynamic process. The so-called development trend comparison is the comparison of geometric relations of relevant statistical data in each period of the system.

5.1.2. Sample selection
We counted and sorted the total number of comments of all the same hair dryer products, and took the top five comments as statistical samples. For each case, the number of one-star evaluation and the number of five-star evaluation in each quarter from 2013 to 2015 was calculated, and the trend analysis was made by using line graph.

5.1.3. Chart analysis and conclusion
we analysis the impact response among one-star reviews and five-star reviews:

![Figure 1. Impact response among one-stars and among five-stars](image)

From the change trend of the broken lines in the chart, we can see that the change trend of the number of one-star reviews in the next quarter is basically the same as that of the previous quarter. From the position of the broken lines in the chart, we can see that when the number of one-star reviews in the previous quarter increases, the number in the next quarter also increases.

![Figure 2. The influence reaction between stars](image)

Therefore, we came to a preliminary conclusion that a given star rating, such as a one-star rating, would lead to more one-star ratings, with a certain chain reaction. And that’s the same can be inferred in one-star reviews:

Then we analysis the influence reaction between one-star reviews and five-star reviews:
According to the change trend of the broken lines in the chart, the change trend of the number of five-star reviews in the next quarter has no obvious trend with that of the number of one-star reviews in the previous quarter.

Generally, the conclusion of our investigation on the chain reaction effect of star rating is as follows: The evaluation of the same emotional trend has a great influence on each other, but has little influence on the opposite trend. It can be considered that people are more inclined to refer to the same opinions, so it can be seen that fixed praise group customers are of certain importance.

5.2. Correlation analysis between specific quality description and star rating level
We now analysis that are specific quality descriptors of text-based reviews such as “enthusiastic” and “disappointed”, and others, strongly associated with rating levels.

5.2.1. Algorithm discussion
Considering the large number of comments, it is unrealistic to analyse the information contained in these comments only by relying on the text[1]. It is necessary to convert the text into several quantifiable indicators for data analysis.

We decided to take advantage of the Python language to turn the information in the text into quantifiable metrics. Meanwhile, through the processing of comment text by the TextBlob library, the values of emotional polarity in the range of (-1, 1) are obtained. When P is negative, the emotion is negative; When P is positive, it indicates positive emotion, and the value reflects the intensity[2-3].

5.2.2. Sample selection
We processed all comment data of the hair dryer, removed invalid data combination, and took 1/5 of the whole population as sample data.

5.2.3. Data analysis
We used SPSS to analyse the data and make a scatter chart of star rating-emotional polarity values.

![Figure 3. scatter chart of star rating-emotional polarity values](image)

5.2.4. Conclusion
From the scatter diagram, it can be seen clearly that the emotional polarity values of higher scores (3-5) are generally concentrated above the red line, that is, greater than 0. However, the emotional polarity value of the lower score (1~2) is generally concentrated below the red line, that is, less than 0.
At the same time, the higher the star rating, the scatter distribution shows a trend of concentration upward, the lower the star rating, the scatter distribution shows a trend of concentration downward.

Therefore, we can infer that specific emotional description is exactly related to star rating. The more positive emotion description, the stronger positive emotionality and the higher score. The more positive emotions it describes, the stronger its positive emotionality and the higher its score. The more negative emotions it describes, the stronger its negative emotionality and the lower its score.

6. A comprehensive rating evaluation model

6.1. Selection of evaluation indicators

Combined with the analysis of one-dimensional indicators, we find that the star rating of products and the sentiment description attributes of comments all play an important role in providing product and market information.

At the same time, in order to build a comprehensive rating evaluation model with comprehensive coverage of evaluable sales status and future development, we added the following two indicators: comment effectiveness and membership effect by combining the given data and considering the enhancement of comment credibility by helpfulness rating and Amazon Vine Voices.

6.2. Model building

6.2.1. Comment validity λ

We define comment effect as affected by three indicators: "helpfulness rating, the emotional intensity of comment, membership effect".

• Helpfulness Rating:

A helpfulness rating is the degree to which a comment is acceptable to other users. Considering that users of amazon generally evaluate useful information through voting and thumb up, therefore, generally speaking, the higher the number of comments with helpful votes, the closer the product description is to the actual using situation of users, and the higher the authenticity and helpfulness. This paper deals with the relationship between the number of helpful votes and the total number of votes through ratio method. The more votes there are, the more the result of the vote will reflect the value of the comment[4-5].

Therefore, we set the quantization attribute of no useful vote to 0, and the higher the number of comments of the higher the number of helpfulness rating the higher the index weight. The quantization method of the value of useful index is as follows:

\[
H_r = \begin{cases} 
0 & T_r = 0 \\
\frac{H_r}{T_r} & 1 \leq T_r \leq 10 \\
\frac{H_r}{T_r} \cdot \varphi_1 & 10 < T_r \leq 50 \\
\frac{H_r}{T_r} \cdot \varphi_2 & 50 < T_r \leq 100 \\
\frac{H_r}{T_r} \cdot \varphi_3 & T_r > 100 
\end{cases}
\]  

(1)

Where \( r \) is the number of comments on a product, \( H_r \) is the number of helpfulness rating for comment \( r \), \( T_r \) is the total number of votes, \( \varphi (\varphi > 0) \) is the increased weight given to voting comments. Different weights are set for different threshold intervals. In order to ensure the rationality of weight setting, this paper appropriately takes: \( \varphi_1 = 1.5, \varphi_2 = 2, \varphi_3 = 2.5 \)

• The emotional intensity of comment:
The emotional intensity of comment refers to the degree to which online comments describe emotional attributes in detail. After referring to numerous literatures, we believe that the longer reviews are more detailed in the description of products and services, which usually contain more comprehensive and detailed information. But at the same time, the literature believes that too long comments are scattered and tend to deviate from the theme, which is not easy for consumers to understand. Therefore, the emotional intensity of comment intensity on credibility is within a certain threshold[6-7].

This paper analyzes the information content of online comments from the perspective of the effective length of comments, and measures the effective length by measuring the ratio between the total number of emotional characteristic words and the total length of comments, and adopts the logarithmic method to make the depth value of comments tend to be smooth, which is conducive to weakening the value difference of the denominator in the measurement. The calculation formula is as follows:

$$L = \frac{\ln(N_a)}{\ln(N_t)} \#(2)$$

Where $L$ is the emotional intensity of comment, $N_t$ is the number of words in the comment, $a_1$ is "positive" indicator, $a_2$ is "negative" indicator. $N_a$ is the number of words that reflect emotion $= N_t(a_1 + a_2)$. To avoid the situation that the emotional intensity of comment is negative, if $L$ is negative, we make $L = 0$.

- Membership effect
Considering that amazon invites customers with accurate and insightful reviews to be its members and provides free copies of products, we have incorporated the membership effect into the review influence system in consideration of the members' trustworthiness:

$$E = \begin{cases} 1 & \text{Comments filled in by members} \\ 0 & \text{Comments filled in by non-members} \end{cases} \#(3)$$

We use the analytic hierarchy process (AHP) to assign weight to the indicators, and finally we got:

$$[0.6442 \quad 0.0852 \quad 0.2706]$$

And the the consistency is acceptable, So the comment validity:

$$\lambda = 0.6442H_r + 0.0852L + 0.2706E\#(4)$$

6.2.2. Comprehensive rating evaluation model
In order to help sunshine company to make guidance opinions on the overall decision of the products to be launched, we establish a comprehensive rating evaluation model to calculate the comprehensive evaluation value of each product and reflect whether the product is successful or not.

For each comment text, according to the judge of NLTK, we have positive indicator $a_1$, negative indicator $a_2$. Set the number of stars for each review as $v$. The above five indicators compose buyer's score $\omega$.

Be similar to the previous discussion, through the analytic hierarchy process to assign weight to the three indicators, we get the compose of buyer's score $\omega$. Then we got this:

$$\omega = 0.1047a_1 + 0.6370a_2 + 0.2583v\#(5)$$

For each product, its comprehensive evaluation of product value $Z$ is the sum of the buyer's score for each comment. Based on the concept of comment validity $\lambda$, a comment’s contribution to $Z$ can transfer to $\lambda$ comments’ contribution. We assume this product have n comments,” the comment number n’s validity is $\lambda_n$, the buyer’s score of “the comment number n” is $\omega_n$, then the comprehensive evaluation of product value is:

$$Z = \frac{\sum \omega_n \cdot \lambda_n}{\sum \lambda_n} \#(6)$$

Finally, we calculate the comprehensive rating value of all products in this industry and take the median as the benchmark value, then the products with the comprehensive rating value higher than the
benchmark value can be regarded as successful products, and the products with the comprehensive rating value lower than the benchmark value can be regarded as failed products.

7. Comprehensive rating value calculation
Take hair dryer products for example. To avoid accidental results, we screened out brands with less than 50 comments, and calculated the index item by item according to the data to obtain the comprehensive evaluation value. The result are as follows:

![Figure 4. Comprehensive rating value](image)

| Distribution     | Value   |
|------------------|---------|
| Median value     | 106.477 |
| Mean value       | 103.335 |
| Minimum value    | 62.058  |
| Maximum value    | 123.517 |
| 25%              | 97.517  |
| 75%              | 112.427 |

8. A time-based measurement and prediction model
Take hair dryer products for example. To avoid accidental results, we screened out brands with less than 50 comments, and calculated the index item by item according to the data to obtain the comprehensive evaluation value. The result are as follows:

8.1. Model establishment
Assume that the known reference data is listed as:

\[ x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)) \] (#7)

Do one accumulation to generate the sequence:

\[ x^{(1)} = (x^{(1)}(1), x^{(1)}(1) + x^{(0)}(2), \ldots, x^{(1)}(n - 1) + x^{(0)}(n)) \] (#8)

Among them:

\[ x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \text{ for } k = 1, 2, \ldots, n \] (#9)

Calculate the average sequence:

\[ z^{(1)}(k) = 0.5x^{(1)}(k), 0.5x^{(1)}(k - 1) \] (#10)

So:
Thus, the grey differential equation is established as follows:
\[ x^{(0)}(k) + az^{(1)}(k) = b, k = 2,3,\ldots,n \] # (12)
The corresponding albinism equation is:
\[ \frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b \] # (13)
Therefore, the albinism equation can be solved as follows:
\[ x^{(1)}(k + 1) = (x^{(0)}(1) - \frac{b}{a}) e^{-ak} + \frac{b}{a} k = 1,2,\ldots,n - 1 \] # (14)
By establishing GM(1,1), the predicted value can be obtained:
\[ \hat{x}^{(1)}(k + 1) = (x^{(0)}(1) - \frac{b}{a}) e^{-ak} + \frac{b}{a} k = 1,2,\ldots,n - 1 \] # (15)

8.2. Accuracy test – posterior variance test
We assume that, in accordance with the GM(1,1) has been calculated \( \hat{x}^{(1)} \), then accumulated by subtracting we got \( \hat{x}^{(0)} \):
\[ \hat{x}^{(0)} = [x^{(0)}(1)\hat{x}^{(0)}(2),n\hat{x}^{(0)}(n)] \] # (16)
We calculate the residuals:
\[ e(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1,2,\ldots,n \] # (17)
The original sequence \( \hat{x}^{(0)} \) and residual error sequence E variance is as follows:
\[ S_1^2 = \frac{1}{n} \sum_{k=1}^{n} [x^{(0)}(k) - \bar{x}]^2 \] # (18)
\[ S_2^2 = \frac{1}{n} \sum_{k=1}^{n} [e(k) - \bar{e}]^2 \] # (19)
Among them:
\[ \bar{x} = \frac{1}{n} \sum_{k=1}^{n} x^{(0)}(k), \bar{e} = \frac{1}{n} \sum_{k=1}^{n} e(k) \] # (20)
We now use MATLAB programming to achieve the process of modeling.

8.3. Examples of prediction process
The comprehensive evaluation scores of each year were imported into the grey prediction system ordinarily, and the prediction results of effective measurement based on time were obtained. Sunshine Company can compare the predicted value with the base value to effectively know the future sales trend and effectively measure the increase and decrease of the product's market reputation.
Taking the product with the highest comprehensive rating value among the hair dryer products as an example, we substituted the average score of each quarter into the gray prediction model. The prediction results and charts are as follows:
The input data: [90.571, 93.328, 97.284, 100.1, 103.46875, 108.8875]
The next fitting value is 112.3183
Again the next fitting value is 116.5739
After testing, the posterior difference ratio is 0.094418, the system prediction accuracy is good. And the predicted value is a perfect prediction of the future development trend of the product. So this model can be applied to other product analysis.

9. Strengths and Weaknesses

9.1. Strengths

- In the grey prediction model, we use the posterior variance test to make the model more accurate and get more accurate results.
- Although we substituted all the information and data into the model, we still believed that the reference value of products with more comments was higher and more concerned, and did not cause the products with less information and less reference value to affect our judgment of product standards.
- After testing, the model used is applicable to both large projects and small projects. The model is effective.

9.2. Weaknesses

- In the use of analytic hierarchy process, even though the construction of judgment matrix is reasonable, there are subjective factors.
- Though the weight is small, the calculation of the emotional intensity of comment is a bit rough.
- The calculation method will have errors, so the calculated value may be not so accurate.

References

[1] Liu j et al. Low-quality product review detection in [C] //Proceedings of the 2007 joint confer Ence on empirical methods in natural language processing and com-Putational natural language learning. The incapable: Assaciation for Com -Putational Linguistics. 2007:334-342.

[2] Li x, Ding sc. Research on the identification of junk review information [J]. Modern book Journal of technology, 2013(1): 63-68.

[3] Yun t. Feature extraction of spam comment detection in ecommerce Law [J]. Modern library and information technology, 2014(10): 93-100.

[4] Huang tingting, Zeng guo-sun, Xiong huan-liang. Shopping customer reviews based on the correlation degree of product characteristics Trusted sort method [J]. Computer applications, 2014,34 (8): 2322-2327,2341.

[5] Wang qianqian. A ranking method for the credibility of online commodity review information [J]. Intelligence miscellaneous Chi, 2015,34 (3): 181-185.
[6] Cai xiaozhen. Research on user comment filtering model for emotion analysis [J]. Modern library and information technology, 2014(4): 58-64.

[7] Lu shaohua, Yang liang, Lin hongfei. Ranking of restaurant reviews based on LDA model [J]. meter Computer engineering, 2011(10): 62-64, 67.