

Product Promotion Prediction Model Based on Evaluation Information

Qixiu Kang* Jing Tang Yuming Wang

College of Artificial Intelligence, North China University of Science and Technology, Tangshan, Hebei, 063000, China

ARTICLE INFO

Article history

Received: 3 March 2021

Revised: 10 March 2021

Accepted: 9 April 2021

Published Online: 16 April 2021

Keywords:

NLP

Markov chains

TOPSIS

Product promotion

ABSTRACT

This paper mainly studies the impact of evaluation information on e-commerce platform on the future of products. Through natural language processing and rating, an evaluation model based on user rating and evaluation is defined to measure product quality. Among them, evaluations are differentiated: review sentiment coefficient (R) and review length (L). The evaluation model is: In order to predict the future reputation of products, based on the above evaluation model, time series is used to rank the products studied. Each customer purchases the product through Markov chain model, so as to predict the probability of future word-of-mouth spread of the product. Use TOPSIS method to select monthly sales, stars and comment sentiment coefficient as indicators. The comprehensive measurement method based on text and score is determined to predict whether the product is successfully promoted.

1. Rating Evaluation Model

Based on the interactivity, anonymity and convenience of the Internet, more and more people are willing to share their review information on various websites. Compared with traditional offline reviews, online reviews have the characteristics of wide range, large amount of information, anonymity, and storage.^[1] Online reviews can express personal opinions more conveniently. Taking Amazon as an example, it provides customers with an opportunity to rate and review purchases. Customers can use “star ratings” to indicate their level of satisfaction with a product. In addition, customers can submit “reviews” to express further opinions and information about the product. At the same time, other customers can submit “helpfulness ratings” based on these reviews to assist their own product purchasing decisions. Companies can analyze these data to gain insights into the markets operation and potential success of product design feature choice.

The important factors that affect the product are rating and comment, so these two factors should be considered when building the model. The higher the star rating, the better the product; the longer the comment, the higher the review sentiment coefficient, and the higher the comment helpfulness.^[2] It should be noted that reviews and star ratings should correspond to each other. But in real life, “five-star bad reviews” and “one-star black powder” and other situations appear frequently. So the data needs to be processed first, remove data that does not match star ratings with reviews.

The process of natural language processing is as follows:

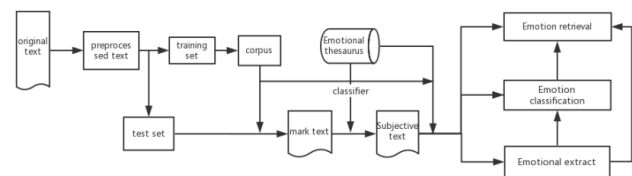


Figure 1. NLP flowchart

*Corresponding Author:

Qixiu Kang,

College of Artificial Intelligence, North China University of Science and Technology, Tangshan, Hebei, 063000, China;

E-mail: 1239562479@qq.com.

Using the normalized star rating and review sentiment coefficient, the star rating and review rating are standardized as weights, and the normalized star ratings and review rating data were used to normalize star ratings and the weight is used to measure the quality of a certain product.

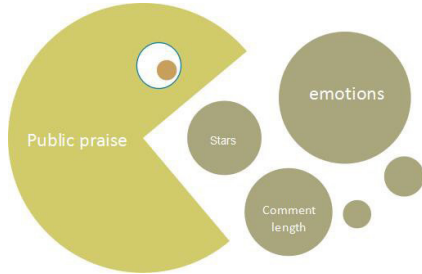


Figure 2. NLP flowchart

A research shows that customers usually tend to look at reviews rather than star ratings when purchasing products. So this article considers customer purchase psychology when setting weights. The product star weight is set to 30%, and the comment weight is set to 70%.^[3-4]

Among them, the comment length weight is set to 30%, and the review sentiment coefficient weight is set to 70%. That is:

$$D=0.3 \times S+0.7 \times (0.3 \times L+0.7 \times R)$$

2. Product Reputation Development Trend Model

Every sales and review process of the product is an event. The comprehensive prediction of the event can not only obtain various possible results of the event, but also give the probability of each result. Therefore, Markov prediction method is selected to predict the future time changes according to the current time situation and explore the development trend of product reputation in online market.^[5]

2.1 State Division

According to the data measurement method determined above, the final score of the selected product can be obtained.^[6] Since the final score of the selected product is between 0 and 1, the final score of the product is divided into five states. The criteria are as follows:

Table 1. Reputation status classification criteria

Status	Level	Condition value interval
S_1	C-	0-0.2
S_2	C	0.2-0.4
S_3	B	0.4-0.6
S_4	A-	0.6-0.8
S_5	A	0.8-1

2.2 State Transition Probability and Transition Probability Matrix

There are five types of product reputation states: S_1, S_2, S_3, S_4, S_5 , but only one state can be realized at a time. So each state will have n-1 different turns and one self-turn.

The state transition probability matrix obtained according to the final score of the product status and the status level division criteria is as follows:

$$P = \begin{bmatrix} 1/34 & 4/17 & 4/17 & 1/34 \\ 5/69 & 17/69 & 19/69 & 1/69 \\ 13/103 & 16/103 & 24/103 & 1/103 \\ 16/151 & 24/151 & 51/151 & 2/151 \\ 0 & 4/5 & 0 & 0 \end{bmatrix}$$

And P_{ij} satisfies the condition:

$$\begin{cases} 0 < P_{ij} < 1, (i, j = 1, 2, \dots, 5) \\ \sum_{j=1}^5 P_{ij} = 1, (i = 1, 2, \dots, 5) \end{cases}$$

2.3 Markov Forecast

Introduce state probability $\pi_j(k)$ here. State probability represents the probability of being in the state S_j at k times after k state transitions under the condition that the state is known at the initial time ($k = 0$). According to the nature of probability:

$$\sum_{j=1}^N \pi_{ij}(k) = 1$$

The state transition process from the initial state transition to S_j can be regarded as first reaching state S_i ($i = 1, 2, \dots, 5$) after (k-1) state transitions, and then reaching state S_j after another state transition. According to the aftereffects of the Markov process and the Bayes conditional probability formula we can get:

$$\pi_j(k) = \sum_{i=1}^5 \pi_i(k-1) P_{ij}, (j = 1, 2, \dots, 5)$$

Record line vector . From formula 11, a recursive formula for successively calculating the state probability can be obtained.

$$\begin{cases} \pi(1) = \pi(0)P \\ \pi(2) = \pi(1)P = \pi(0)P^2 \\ \dots \\ \pi(k) = \pi(k-1)P = \pi(0)P^k \end{cases}$$

In the above formula, $\pi(0) = [0, 0, 1, 0, 0]$ is the initial

state probability vector.

It can be known from the analysis that if the initial state of an event at the 0-th moment is known, we can get the probability of being in various possible states at the k-th moment by using recursive formula. Thus, the state probability prediction result of the event at the k-th moment is obtained.

2.4 Ultimate State Probability Predictio

The state probability obtained after infinite state transitions is called the ultimate state probability, or the equilibrium state probability. If we record the ultimate state probability vector as $\pi = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5]$:

$$\pi_i = \lim_{k \rightarrow \infty} \pi_i(k), (i = 1, 2, \dots, 5)$$

That is:

$$\lim_{k \rightarrow \infty} \pi_j(k) = \lim_{k \rightarrow \infty} \pi_j(k + 1) = \pi$$

The above formula is substituted into the recurrence formula of Markov prediction model and given:

$$\pi = \pi P$$

In this way, we get the conditions that the ultimate state probability should satisfy:

$$\begin{cases} \pi = \pi P \\ 0 \leq \pi_i \leq 1, (i = 1, 2, \dots, 5) \\ \sum_{i=1}^5 \pi_i = 1 \end{cases}$$

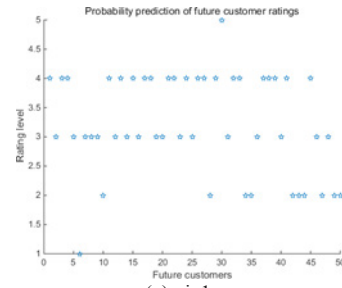
Let the probability of the ultimate state be $\pi = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5]$ to get:

$$(\pi_1 \pi_2 \pi_3 \pi_4 \pi_5 = \pi_1 \pi_2 \pi_3 \pi_4 \pi_5)$$

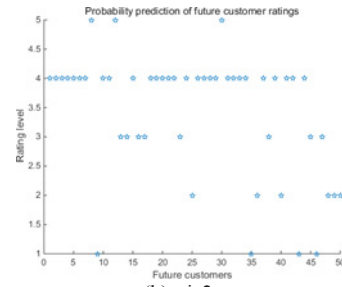
$$\begin{bmatrix} 1/34 & 4/17 & 4/17 & 8/17 & 1/34 \\ 5/69 & 17/69 & 19/69 & 9/23 & 1/69 \\ 13/103 & 16/103 & 24/103 & 49/103 & 1/103 \\ 16/151 & 24/151 & 51/151 & 58/151 & 2/151 \\ 0 & 4/5 & 0 & 1/5 & 0 \end{bmatrix}$$

2.5 Result Analysis

Solving formula 16 gives that $\pi_1 = 364/3775, \pi_2 = 188/985, \pi_3 = 51/181, \pi_4 = 835/2002, \pi_5 = 43/3102$



(a) pic1.



(b) pic2.



(c) pic3.



(d) pic4.

Figure 4. Probability prediction of future customer ratings

It can be seen from the figure 4 that the product reputation is roughly in S4. According to the established model, the probability of each future product reputation in different states can be calculated.

3. TOPSIS model establishment

In order to better identify a potential successful or failed product, we have determined a comprehensive measurement method based on text and rating. C.L.Hwang and K.Yoon first proposed the TOPSIS in 1981. [7] The TOPSIS sorts according to how close a limited number

of evaluation objects are to idealized targets, which is a relatively good evaluation of existing objects. The basic principle is to sort by detecting the distance between the evaluation object and the optimal solution and the worst solution. If the evaluation object is closest to the optimal solution and farthest from the worst solution, it is the best; otherwise it is not optimal.

3.1 Standardization of indicators

Select monthly sales (M), star rating (S), and review sentiment coefficient (R) as indicators.

Table 2. Indicator type

	Monthly sales	Star rating	Review Sentiment Coefficient
Indicator type	Very large	Very large	Very large

The monthly sales volume is standardized as the formula: $\frac{M}{max}$. The star rating is standardized an the formula: $\frac{S-1}{max}$. Generate a normalized matrix.

3.2 Calculate the Normalization Matrix

Assume the generated normalization matrix is:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{bmatrix}$$

3.3 Determine the Ideal Solution

Determine the positive ideal solution:

$$Z^+ = (Z_1^+, Z_2^+, \dots, Z_m^+)$$

$$= (\max\{z_{11}, z_{21}, \dots, z_{n1}\}, \max\{z_{12}, z_{22}, \dots, z_{n2}\}, \dots, \max\{z_{1m}, z_{2m}, \dots, z_{nm}\})$$

Determine the negative ideal solution:

$$Z^- = (Z_1^-, Z_2^-, \dots, Z_m^-)$$

$$= (\min\{z_{11}, z_{21}, \dots, z_{n1}\}, \min\{z_{12}, z_{22}, \dots, z_{n2}\}, \dots, \min\{z_{1m}, z_{2m}, \dots, z_{nm}\})$$

3.4 Calculate the Distance Scale

Define the distance between the i-th evaluation object and the minimum value as:

$$D_i^+ = \sqrt{\sum_{j=1}^m (Z_j^+ - z_{ij})^2}$$

Define the distance between the i-th evaluation object and the minimum value as:

$$D_i^- = \sqrt{\sum_{j=1}^m (Z_j^- - z_{ij})^2}$$

Thus, the unnormalized score of the i-th evaluation object is:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

It is clear that $S_i = \frac{D_i^-}{D_i^+ + D_i^-}$. The larger S_i is, the smaller $S_i = \frac{D_i^-}{D_i^+ + D_i^-}$ is, the closer to the maximum.

The TOPSIS evaluation method is authentic, intuitive, and reliable which is a very classic and effective method in multi-objective decision analysis.^[8] This paper combines the evaluation indicators and uses the TOPSIS method to sort the evaluation targets.

4. Conclusion

In terms of marketing strategy, this article starts with the data provided. Analyze past customer reviews through NLP algorithm to obtain quantitative indicators of product preference. At the same time, it is found that the above three are related to the practicality of customer reviews. The quality of reviews will affect future customers' purchase intentions. Therefore, in order to promote product launches and successful promotion, it is necessary for companies to fully pay attention to customers' star ratings, emotional index and product word length.

In order to promote the company's product marketing, the concept of time is incorporated into the model. First, this article arranges the data of selected brands of hair dryers in chronological order. Then, use Markov to predict the word-of-mouth evaluation of future customer purchases, and obtain a dispersion distribution map of product word-of-mouth evaluation. Through analysis, we come to the following conclusion: In order to ensure long-term product reputation and promote product marketing, we need to consider the product score under time causality.

4.1 Strengths

The established model is closely related to the actual situation. And the proposed problem can be solved in combination with the actual situation, so that the model is closer to the reality which has strong promotion and generality.

TOPSIS-based multi-objective decision analysis im-

proves the scientificity, accuracy, and operability of the analysis.

4.2 Weaknesses

Markov prediction models require that all events be independent. But in fact, customers may write the same type of review after seeing a series of star ratings. As a result, the accuracy of prediction is reduced.

5. References

- [1] Li Zongwei. Research on the Model Construction of Online Reviews Affecting Consumers' Purchase Decisions [J]. *Innovation*, 2013, 5 (7): 55-58.
- [2] Adam Lingreen, Lin Zhang, Baolong Ma. The impact of online user reviews on cameras sales[J]. *European Journal of Marketing*, 47(7):1115-1128.
- [3] Lingreen A , Zhang L , Ma B , et al. The impact of online user reviews on cameras sales[J]. *European Journal of Marketing*, 2013, 47(7):1115-1128.
- [4] Chrysanthos Dellarocas. The Digitization of Word of Mouth: Promise and Challenges of On-line Feedback Mechanisms[J]. *Management Science*, 2003,49(10):1407-1424.
- [5] Wang Junkui. Research on the usefulness of online reviews for e-commerce websites [D]. Xi'an: Xidian University, 2014.
- [6] Chen Geheng .Text sentiment analysis based on polarity transfer and bidirectional LSTM [J] .*Information Technology*, 2018, (2): 149-152.
- [7] Tian Wuzi, Sun Ze, Sun Run, et al. Analysis of Urban Water Consumption Based on SPSS Multiple Linear Regression [J]. *Journal of North China Institute of Science and Technology*, 2019, 16 (02): 114-117.
- [8] Gu Xiaohan. Research on the Performance Evaluation of Precision Poverty Alleviation Based on TOPSIS [D]. Jiangxi: Jiangxi University of Finance and Economics, 2019.