2022 3rd International Conference on Education, Knowledge and Information Management (ICEKIM) | 978-1-6654-1666-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICEKIM55072.2022.00234

An Analysis Method for Online Shopping Platform Comments Based on NLP-AHP: Taking Amazon as An Example

Junjie Tang [1,a*]
[1] School of Cultural Heritage and Information Management, Shanghai University, Baoshan District, Shanghai, China
[a,*] e-mail: keventnt@outlook.com

Jiaxin Guo [2,b]
[2] Faculty of Science, Molecular Biology & Biotechnology, the University of Hong Kong, Pokfulam, Hong Kong, China
[b]e-mail: u3577171@connect.hku.hk

***Abstract*- This research proposes a new method for analyze online shopping platform review data based on the NLP-AHP method and conducts an empirical analysis with the review data of microwave ovens on Amazon shopping sites. Through empirical analysis, we can find that the NLP-AHP method can quickly find more important comments and time nodes. It can help stakeholders improve product quality and improve sales strategies by analyzing these important comments.**

Keywords- NLP algorithm; AHP method; data mining; online shopping review data

I. Introduction

With the development of Internet technology, online shopping has become an indispensable part of people's lives around the world, so businesses need to take new measures to discover customer perceptions for products and improve the competitiveness of their products [1].

Amazon is one of the biggest online markets in the world. In the reviews of Amazon, there are many interesting things: the first one is the Individual ratings, it can also be named as star ratings, which is provided by Amazon in the online marketplace it created allow purchasers to express their level of satisfaction with a product. The second one is customers can express their further comments on the product through reviews which can be rated by others as being useful or not. Other customers can submit ratings on these reviews as being helpful or not by helpfulness rating. Using date above can improve the market's perception, the timing of that participation and the potential success of product design feature choices of companies. The third one is Amazon Vine program is an exclusive review program launched by Amazon to brand sellers. It is a shortcut for sellers to quickly obtain high-quality reviews for new products. The members who join the program are called Amazon Vine reviewers (also known as Vine Voice, or Vine spokespersons) who are responsible for free use of products provided by Amazon sellers and publishing real buyer reviews for these products. However, researchers at this stage do not understand how to use these multi-dimensional comment data to improve product quality and improve sales strategies.

Online customer reviews posted on an online market's website, provide feedbacks of various products from a large number of users. The majority of the studies in the online review literature focus on issues such as the impact of reviews on sales [2-3], customers' participation in reviews [4-5], and helpfulness of online reviews [6-7]. However, there are few literatures to study the online user reviews from the perspective of combining the two methods of Analytic Hierarchy Process (AHP) and Natural language processing (NLP). In addition, most of the studies only study the main body of the review without considering such as the title of the review, user rating, and other people's feedback on the review.

Motivated by the recent advances in deep learning and natural language processing (NLP), We propose a new method for analyze online shopping platform review data based on the NLP-AHP. In the implementation of NLP technology, we use Baidu's NLP capability engine, which is an open-source technology, it unites industry authorities and universities to jointly establish an open-source data for Chinese natural language processing. Baidu's NLP technology supports a wide range of internal and external businesses and has a deep reserve of basic natural language processing technologies such as lexical analysis and word meaning similarity. In the realization of the AHP model, we use the Python language for programming, which provides some efficient high-level data structures. The data used in this study is based on the review data on all microwave ovens on the Amazon platform from 2004 to 2015. These data are retrieved from an open access MCM database.

II. Research Method

A. NLP Algorithm

NLP stands for Natural language processing, which refers to the branch of computer science, to be more specifically, it is a branch of artificial intelligence which concerned with giving computers the ability to understand text and spoken words in much the same way human beings can [8]. NLP is the driving force behind machine intelligence in many modern real-world applications, such as spam detection, machine translation, virtual agents and chatbots, social media sentiment analysis and text summarization.

Sentiment analysis, as the most important part of NLP, refers to the process of analyzing, processing, inducing and reasoning about subjective texts with emotional colors. On the Internet (such as blogs and forums), a large number of users participate in valuable comment data, this comment data express people's various emotion and tendency, such as joy, anger and sadness. Based on this, potential users can understand the public opinion on a certain event or product by browsing

978-1-6654-1666-5/22/$31.00 ©2022 IEEE
DOI 10.1109/ICEKIM55072.2022.00234

these subjective comments. Therefore, the mining of review data in this article is based on this technology.

B. AHP Method

The Analytic Hierarchy Process (AHP) is a kind of multi-criteria decision (evaluation) analysis method that combines quantitative and qualitative analysis proposed by the American professor Saaty at the University of Pittsburgh in the early 1970s [9]. This method is widely used in engineering, economics, military, politics, diplomacy and other fields, it solves many important issues such as system evaluation, resource allocation, price forecasting item selection. The process of the AHP method is as follows [10]:

(1) Constructing the judgment matrix. The specific method is to compare the relative importance of each element at each level and give a judgment. These judgments are expressed in numerical values and written in the form of a matrix, the so-called judgment matrix.

Assuming that the element C_k of the previous level is dominated by the element $A_1, A_2, \ldots, A_n$ in the next level, a pairwise comparison judgment matrix between the elements $A_1, A_2, \ldots, A_n$ with C_k as the judgment criterion can be established. The matrix is denoted as A, and the form is as follows:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

The scaling method of $a_{ij}(i = 1,2, \ldots, n; j = 1,2, \ldots, n)$ is shown in Table 1.

Table 1 scaling method of AHP

Scaling	Meaning
1	Compared with the two factors, they have the same importance
3	Compared with the two factors, the factor *i* is slightly more important than the factor *j*
5	Compared with the two factors, the factor *i* is obviously more important than the factor *j*
7	Compared with the two factors, the factor *i* is strongly more important than the factor *j*
9	Compared with the two factors, the i factor is extremely important than the j factor
2,4,6,8	The median of the above two adjacent judgments
reciprocals	If activity *i* has one of the above numbers assigned to it when compared with activity *j,* then j has the reciprocal value when compared with *i*

(2) Single hierarchical arrangement. The single hierarchical arrangement refers to finding out the non-sequential weights of the relative importance of corresponding elements at the same level to an element at the previous level. In fact, it is to determine the relative priority. The method is to calculate the largest eigenvalue λ_{max} of the judgment matrix *A* that satisfies the equation $AW = \lambda_{max}W$ and the corresponding eigenvector *W*. This eigenvector is the sorting weight under the single criterion.

By solving the eigenvalue of the matrix *A*, the corresponding eigenvector can be obtained, and the weight vector obtained after normalization is $\omega = (\omega_1, \omega_2, \omega_3, \ldots, \omega_n)^{\mathrm{T}}$, Where $\omega_i(i = 1,2, \ldots, n)$ is the relative weight of different indicators.

(3) Consistency check of single hierarchical arrangement. In order to judge the degree to which the matrix deviates from the consistency, the consistency index *CI* can be calculated. The calculation of *CI* is shown in equation (2), where *n* represents the number of items being compared.

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{2}$$

Cl represents the degree to which the judgment matrix deviates from the consistency. The closer it is to 0, the better the consistency of the matrix; when $CI \leq 0.1$, the judgment matrix is considered to have satisfactory or acceptable consistency. According to the law of people's understanding of things, when we construct the judgment matrix, the fewer the factors of the pairwise comparison, the higher the accuracy of the judgment result. In other words, the more dimensionality of the judgment short matrix, the easier it is to deviate from the consistency, and the greater the *CI*.

Therefore, in order to measure the size of *CI,* the random consistency index *RI* is introduced to modify *CI.*

Table 2 the random consistency index RI

n	1	2	3	4	5
RI	0	0	0.58	0.9	1.12
n	6	7	8	9	10
RI	1.24	1.32	1.41	1.45	1.49

(4) Consistency ratio test. Based on the index *RI* and *CI*, a consistency test method commonly used in AHP is proposed: the consistency ratio *CR*, which is defined as the ratio of the consistency index *CI* of the judgment matrix to the average random consistency index *RI* of the same order:

$$CR = \frac{CI}{RI} \tag{3}$$

When $CR \leq 0.1$, it can be considered that the judgment matrix has satisfactory consistency.

III. Data Processing

A. General Description of Data

The empirical data used this time comes from Amazon's microwave oven review data from 2004 to 2015, with a total of 1,615 items.

The number of comments changes over time as shown in Figure 1, we found that the number of comments before 2007 was basically zero, the sum of the number of comments from 2002 to 2006 was only 0.16% of the total. From 2007 to 2011

In the year, comments gradually increased. Explosive growth appeared after 2012, the number of reviews from 2012 to 2015 accounted for 90.56% of the total.

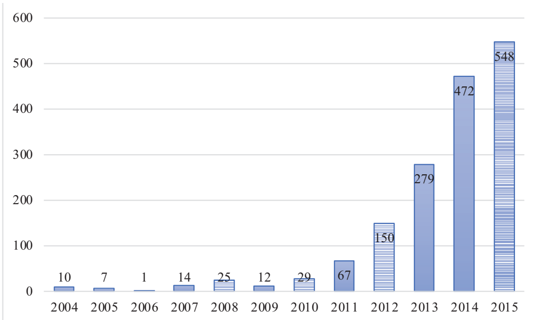


Figure 1. the number of comments in microwave oven with years

B. Data Cleaning

Because of the openness of Amazon and the complexity of reviewer, there are countless invalid reviews. Some companies bribe people to make positive or negative review deliberately, whether it is used to promote their own products or to downgrade others. We assume that the reviews made by people who are neither the vine of Amazon nor have verified purchase are invalid so that we choose to delete them (501 reviews about microwave oven were deleted).

We modify digital gibberish in text. Due to the phenomenon of the rise of digital gibberish in texts, we can convert some part of the symbols and emoticons to match with the text. Thus, we must amend them when cleaning data. There are some "
" that we transform them into blank.

IV. NLP-AHP Modeling

A. Determine the Evaluation Index

Because we need to find indicators that can quickly analyze customers' attitudes and views on products purchased online, so we choose to use the post-modernism [11]. Before the start of our team generally have no theoretical assumptions. We believe that all indicators can have an impact, so we start directly with the actual observation, concludes the experience summarized from the raw data, and then systematize them.

At the first level, we discussed all the items and targeted at nine items: star rating, helpful votes, total votes, vine voice, verified purchase, review headline, review body and review date, product title and product category to judge the attitudes and opinions of customers; At the next level, Through the analysis of causal relationship, time sequence, semantic relationship, etc., our team finally determined to evaluate customers attitudes and opinions on products according to the seven indicators of *star rating, helpful votes, useless votes, vine voice, review headline, review body and review date*; The third level refers to: choosing a core category of this study as "evaluation data", that is, "text analysis" and "scoring analysis".

B. Quantitative Processing Indicators

After determining the seven indicators (star rating, helpful votes, useless votes, vine voice, review headline, review body and review date), we conducted a quantitative analysis on them.

1) Quantitative processing of review headline and body

Our team tried to use the most active research area in NLP (Natural Language Processing), Sentiment Analysis, to quantify the review title text and review content. Our team use python language to access Baidu AIPNLP port, which is powered by Baidu Cloud Natural Language API, then we get the emotional values of review headline and review body, which is between 0 and 1, finally we can make a number to symbol the sentiment.

After using machine learning technology to automatically score the review headline sentiment and the review body sentiment, due to insufficient machine processing, some data scoring cannot be achieved, so we combined the manual evaluation mode to fill the missing data with some other data in the same group.

2) Quantitative processing of review date

According to the information lifecycle theory, the value of information has a half-life, will gradually decline with the passing of time [12]. The later the comment is posted, the more valuable it becomes. Then, we converted the time of comments into a time-effectiveness value Y (by year). The time-effectiveness Y is calculated by equation (4), where t is the deadline year, and t_0 is the year of publication.

$$D = \frac{1}{2}^{(t-t_0)} \tag{4}$$

3) Quantitative processing of vine voice

Amazon Vine Voices means that the customers who write accurate and insightful reviews. So, we use V to identify the Amazon Vine Voices customers, if he is, $V = 1$, if not, $V = 0$.

4) Quantitative processing of useless votes

The usefulness and uselessness of reviews is very important. We subtract numbers of helpful votes from numbers of total votes the review received to get numbers of useless votes.

C. Calculate Weights

After selecting the evaluation indices, we decided to use the AHP (Analytic Hierarchy Process) to calculate weights.

Our team used the AHP to evaluate the above seven indicators. Then average the weight obtained by three people to get the final indicator weights. The weights and CR values obtained are shown in Table 3, both CR are less than 0.01 and passed the consistency test.

Table 3 the values of weights

Indicators	Review Heading	Star Rating	Review Date	Review Body
M1	0.0560	0.2745	0.0211	0.0557
M2	0.0857	0.3415	0.0506	0.2801
M3	0.0368	0.2669	0.0829	0.2705
AVE	0.0595	0.2943	0.0515	0.2021
Indicators	Vine Voice	Useless Votes	Helpful Votes	CR
M1	0.2604	0.0867	0.2455	0.0941
M2	0.0938	0.0413	0.107	0.0828
M3	0.0785	0.0283	0.2361	0.0878
AVE	0.1442	0.1442	0.1962	0.0882

Finally, we found the weight of each indicator which showed as the figure 2.

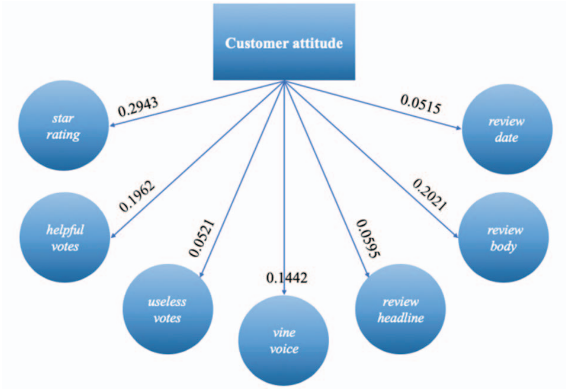


Figure 2. the weight of each indicator

We considered that there may be inconsistent manifestations between the review headline and body, which is related to whether they need to be weighted again, so we firstly analyzed the correlation between the sentiment values of title and the sentiment values of the subject by Charles Spearman correlation coefficient, the results showed as below tables.

Table 4 correlation between the review headline and body

Review title	Correlation Coefficient	1.000	0.524**
	Sig. (2-tailed)	0.000	0.000
	N	1114	1114
Review body	Correlation Coefficient	0.524**	1.000
	Sig. (2-tailed)	0.000	0.000
	N	1114	1114
Review body	Correlation Coefficient	0.524**	1.000
	Sig. (2-tailed)	0.000	0.000
	N	1114	1114
** Correlation is significant at the 0.01 level (2-tailed).			

It can be obtained from Tables 4 that the review headline sentiment and the review body sentiment have a strong correlation, so no additional weight is needed.

D. Model Implementation

1) Raw data normalization

Because the dimensions of the original indicators are different, they cannot be compared directly. So, we must standardize data at fist.

$$r_{ij} = \begin{cases} \dfrac{a_{ij} - mina_{ij}}{max\ a_{ij} - mina_{ij}} & (When\ the\ value\ is\ positive) \\ \dfrac{max\ a_{ij} - a_{ij}}{max\ a_{ij} - mina_{ij}} & (When\ the\ value\ is\ negative) \end{cases} \quad (5)$$

2) Model calculation

Based on the above analysis, we can find that all 7 indicators have an important impact on customer evaluation, so we establish the following linear model as equation (6)

$$F(X) = T \times \theta_t + R \times \theta_r + Y \times \theta_y + B \times \theta_b + V \times \theta_v + U \times \theta_u + H \times \theta_h \quad (6)$$

Based on the above analysis, we can find that all 7 indicators have an important impact on customer evaluation, so we establish the following linear model as equation (6)

Table 5 Definition of each characters

F(X)	Overall rating
T	the value of review headline sentiment
R	the value of star rating
Y	the year of review date
B	the year of review body sentiment
V	vine voice customer is 1, otherwise 0
U	the number of useless votes
H	the number of helpful votes
$\boldsymbol{\theta_t, \theta_r, \theta_y, \theta_b, \theta_v, \theta_u, \theta_h}$	the weight of each item

Model the microwave oven data after the raw data normalization, some data are shown in the appendix, the line chart of the F(X) data after the model calculation is shown in in the appendix. The monthly averages of the values after modeling are shown in Figures 3.

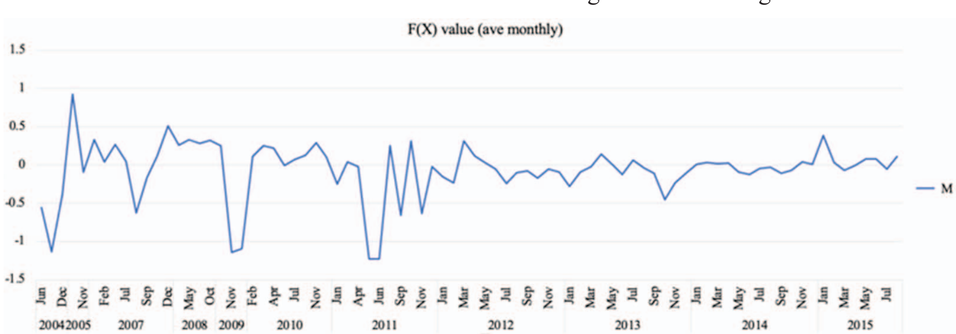


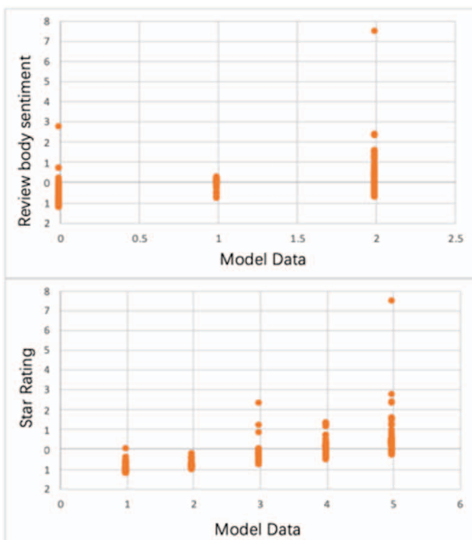
Figure 3. the monthly average of overall rating about microwave oven

With the figure 3, we can find that the reviews of microwave ovens in Amazon have changed greatly in 2004, 2009 and 2011, and the performance increased in 2004, and the performance decreased in 2009 and 2011. Therefore, relevant businesses can focus on the analysis of these three years. The quality changes in the.

E. Model Checking

Owing to the model may have errors, model test is required. We decided test this model by goodness of fit of important factors. From the AHP, we know that the relationship between star rating and review body sentiment is the most important, accounting for 49.64%, so we will perform a goodness-of-fit analysis with Linear Regression on the model data and the above two data to ensure that the model accuracy. The statistic that measures the goodness of fit is the determination coefficient R^2.

We create two linear regressions for each research object. The first one with the value of review body sentiment as the dependent variable(a), with model data as the dependent variable, the second one with the value of star rating as the dependent variable(a), with model data as the dependent variable, then calculate the adjusted R square. As we see all the R^2 are greater than 0.5, so we think the fit between the two is good, our model is considered reliable. data are showed as figure 4.



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.723ª	.523	.523	.367212136

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.740ª	.548	.548	.389769855

Figure 4. Scatter plot for microwave oven data and the model summary

V. Conclusion

This research proposes the NLP-AHP online shopping platform comment mining method and introduces the process of using NLP-AHP to analyze related comments in detail. Based on the review data of Amazon platform microwave ovens, an empirical analysis was carried out to prove the validity of the model. The empirical results show that the model can quickly find out the time nodes of important reviews. At the merchant's level, based on the results of this model, the shortcomings of the product can be quickly improved, and public opinion can be quickly controlled, so that the reputation of their own products can rise rapidly. At the platform's level, online shopping platforms can understand the quality and sales effects of a certain product, then they can better introduce some new manufacturers' products to increase the sales volume and sales speed of the platform, and adjust sales strategies in a targeted manner, including time, economic cost, which can ensure the sustainable development of online shopping platforms.

Even though this research makes valuable practical contributions, it faces several limitations. Firstly, the data collection process only depends on the microwave review data in the Amazon platform, more samples might be more convincing. Nevertheless, future research should consider determining methods to obtain data from multiple different platforms and different fields to minimize the possibility of response bias. Secondly, due to the development of information technology, it is possible to mine comment data from other different levels, including by using AI technology where more comprehensive findings can be generated after changes are accomplished.

Acknowledgments

This work benefited from the Educational Science Project of Changsha City (CJK2021052), and special thanks to Prof. Dr. Xiaoping Sheng for some advice on this paper.

References

[1] Tate, D. F., Wing, R. R., Winett, R. A. (2001). Using Internet technology to deliver a behavioral weight loss program. Jama,285:1172-1177.

[2] Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., Yan, X. (2018). Measuring customer agility from online reviews using big data text analytics. Journal of Management Information Systems,35:510-539.

[3] Qi, J., Zhang, Z., Jeon, S., Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. Information & Management, 53:951-963.

[4] Sonnier, G. P., McAlister, L., Rutz, O. J. (2011). A dynamic model of the

effect of online communications on firm sales. Marketing Science,30:702-716.

[5] Forman, C., Ghose, A., Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information systems research,19:291-313.

[6] Chen, Y., Fay, S., Wang, Q. (2011). The role of marketing in social media: How online consumer reviews evolve. Journal of interactive marketing,25:85-94.

[7] Sonnier, G. P., McAlister, L., Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. Marketing Science,30:702-716.

[8] Cambria, E., White, B. (2014). Jumping NLP curves: A review of natural language processing research. IEEE Computational intelligence magazine,9:48-57.

[9] Saaty, T. L. (1988). What is the analytic hierarchy process? In Mathematical models for decision support. Springer Publishing, Berlin, Heidelberg.

[10] Ricardo, J. E., Rosado, Z. M. M., Pataron, E. K. C., Vargas, V. Y. V. (2021). Measuring legal and socioeconomic effect of the declared debtors usign the AHP technique in a neutrosophic framework. Neutrosophic Sets and Systems, 44:357-366.

[11] Hassan, I. (1985). The culture of postmodernism. Theory, Culture & Society, 2:119-131.

[12] Mika, M. (2020). The Half-Life of Radiotherapy and Other Transferred Technologies. Technology and Culture, 61: S135-S157.