

Analysis about the Text Content's Effect on the Usefulness of Online Comments based on Search Type Products

Yi Kaigang

School of Tourism and Urban-rural planning
Zhejiang Gongshang University
Hangzhou, China
yikaigang@163.com

Yu Ye*

School of Business Administration
Zhejiang Gongshang University
Hangzhou, China
422899459@qq.com

Abstract—This article aims at the Amazon comments data for the original object, the processed data as direct object, using SPSS data analysis tool to study the content of online comments about what factors will affect the online comments usefulness of search type products, and thus improve the theories about influence factors of usefulness of online comments. The study found that extreme comment star, products with high scoring and appearance of sigh emotion are significantly positively correlated with the usefulness of online comments, but large rate variance is negatively correlated with the usefulness of online comments. These conclusions can not only provide the Amazon with recommendations to improve the online comment system to improve their competitive advantage, but also provide suggestions for consumers to have a positive and effective buying behavior.

Keywords—Comments usefulness; Amazon; Empirical analysis; Search type products

I. INTRODUCTION

With the popularization of Internet, network has become an indispensable part of people's life. According to the 40th statistical report on the development of China's Internet network released by the China Internet information center (CNNIC) in August 2017, it can be seen that by June 2017, China's Internet penetration rate is 54.3%, more than the global average of 4.6%, the scale of Chinese Internet users reached 751 million, one fifth of the world's Internet users. Therefore, with the increase of Internet penetration, China's Internet population has grown steadily, which is bound to play a fundamental role in the growth of shopping users and e-commerce transactions.

The deep development of the Internet has made human interaction particularly close, and the explosion of information is not surprising, including the flood of online comments. David (2007), as the originator of the TAM model, put forward the traditional technology acceptance model, in this model, David (2007) argues that the main content of the users perception is perceived usefulness, and it influences consumer attitudes and behavior. The author points out that useful online comments can influence consumers' positive perception of usefulness and make the important basis for online purchasing decision. In practice, there are a number of websites that are beginning to use useful voting systems to assess the usefulness

of online reviews. But usefulness voting mechanism cannot solve all the problems, it is extremely difficult to get a useful vote on low-volume items and some new comments have no useful votes, the review mechanism is invalid for a recent comments—a useful comment can produce certain instruction function only after a long period of accumulation. In the case of the above, the usefulness of the voting system cannot function, and the consumers' decisions will be relatively difficult.

It is not only for consumers, but also for businesses. Cao, Duan and Gan (2011) pointed out that there were more than 20,000 comments about mp3s on foreign Amazon websites, only 38% were considered to be useful, and 81 percent of e-commerce companies were skeptical about how to improve the usability of online reviews. To manage the business effectively, Enterprise should find out the important factors that affect product online comments usefulness and its influence degree, to identify what factors in the online reviews can influence the effectiveness of the review.

II. THEORY AND HYPOTHESIS

Integrated predecessors' research, the author finds that the previous analysis of online reviews uses questionnaires or data experiment method. According to the model, various hypotheses of influencing factors on the usefulness of online comments are put forward, which plays a certain reference role for the author's own analysis.

But the author thinks it also has some shortcomings, such as parts of the important factors that affect online comments' usefulness have not been incorporated into the model, which affect the improvement of the model predictive power, or most scholars studied one online comment on the usefulness of goods, whether research results can be extended to search goods remains to be verified.

So the author wants to define search products as the object, focusing mainly on the effect of comments on the content itself, subdividing comments into several dimensions and is verified on the basis of the composition factors, on this basis, the relationship between each component factor and the usefulness of online comments and the degree of influence are verified.

According to the theoretical contents supported by Sussman and Siegal (2003), Jiang (2015) and other scholars, based on

research purposes of this paper, the author proposes the following six hypotheses:

A. Extreme commenting star

Yin, Liu and Zhu (2012) put forward that the usefulness of online comments is negatively correlated with the assessment of the extreme of stars, and consumers think that moderate comments are more useful than extreme comments.

Five stars and one star are recognized as the extreme star, therefore, the author also defines the extreme star as five stars and one star. On the one hand, the author thinks that all five-star ratings will give people a false sense, the comments are likely to have been made by businesses hiring professional network operators in order to gain more credibility and attention, it may also be the result of offering certain benefits to five-star consumers; On the other hand, a one-star rating conveys a sense of hostility and contempt, which makes the reader suspect that the writer is hostile to the water army with a commercial purpose. Therefore, the author proposes the following hypothesis:

H1: in search products, there is a significant negative correlation between the usefulness of online reviews and the extreme level of rating of a buyer commenting on a product.

B. Grade of products

Liao (2013) believed that in the process of information searching, consumers believe that favorable comments tend to have subjective emotional tendencies of purchasers, more of its information is non-objective, medium and poor evaluations often objectively reveal the shortcomings of commodities. Therefore, in order to obtain the real information about the goods that are useful to consumers, consumers will browse the content of the comments and bad comments of reviewers.

The author thinks that, generally speaking, a low grade will arouse the readers' disgust, while the high score will attract the readers. Low grade product reviews often shows the lack of this product with defect, after readers seeing such comments, they generally more tend to view the details of the goods in order to study its defects, rather than think the comments are useful, and high grade tend to cause the curiosity of readers, thus increase the likelihood of readers thinking the comments are useful. Therefore, the author proposes a second hypothesis:

H2: in search products, the usefulness of online reviews is positively correlated with the rating of products by reviewers.

C. Comment on emotions

Yan (2012) pointed out that many of the factors that motivate consumers to make purchase decisions come from evaluating emotions, for example, the appearance of "very" and other emotive strong words can effectively promote the purchase.

The author thinks that the tone of overreaction, anger and even provocation, or excessive enthusiasm and affirmation, all show the color of hypocrisy to some extent, which will affect the usefulness of the comments, all tones are expressed through punctuation, especially exclamation marks.

Therefore, the author mainly studies the exclamation mark. The author believes that the use of exclamation marks by commentators is for two purposes only: first, internet troopers accentuate the comments in order to impress others, and it does attract consumers to view the comments and to think they are effective in making purchases; Second, real buyers express uncontrollable excitement or anger when they receive large surprises or extreme disappointments, in this case, the review readers are more likely to trust the comments and find they are useful. Therefore, the author proposes a third hypothesis:

H3: in search products, the appearance of exclamation emotion in comment text is significantly positively correlated with the usefulness of online comments.

D. Text typo

The author thinks that the text factor will also have some influence on the usefulness of online comments, such as misspellings, wrong sentences and grammatical comments often give people a bad impression, so it is difficult to get a high degree of recognition. For mobile power products, common mistake is miswriting "filling" to "blunt", so the author uses the word error as dummy variable to judge and proposes a third hypothesis:

H4: in search products, the usefulness of online comments is negatively correlated with the occurrence of text typo.

E. Representativeness of the scores

Guo and Xu (2006) pointed out that the psychological and behavioral process of changing one's attitude or behavior is conformity with the principle of the majority opinion. Judging from this angle, the closer the score of the goods to the average score of the commodity, the more the advantages and disadvantages of the product will be reflected, and the more valuable it will be for the later purchasers. Therefore, the author thinks that the mean square statistic score can be representative, the greater the mean square of the score, the less representative the score is, and the less useful the review is. Accordingly, the author proposes the fifth hypothesis:

H5: in search products, the usefulness of online reviews is positively correlated with the representativeness of the scores.

F. Depth of comments

Mudambi and Schuff (2010) believed that the depth of comments can be measured using the length of words. The more words there are, the less certain the reader will feel about the product or service, and the more useful it will be. Similarly, Chinese scholars Wang and Dai (2012) argued that the length of the review content can reflect the comment on the amount of information, especially for search products, consumers are more likely to query objective descriptions of products, and it's natural to pay more attention to longer reviews when viewing product reviews to help them better understand the product. To sum up, the author proposes the sixth hypothesis:

H6: in search products, the depth of online comments is positively correlated with the usefulness of online comments.

To test extreme commenting star, grade of products, emotional depth, text typo, representative of the score and depth of comments whether has a significant influence on online comments usefulness and its influence degree, this article will make a multiple linear regression to these factors:

$$Y=H1*extreme\ commenting\ star+H2*grade\ of\ products+H3*comment\ on\ emotions+H4*text\ typo+H5*representativeness\ of\ the\ scores+H6*\ depth\ of\ comments$$

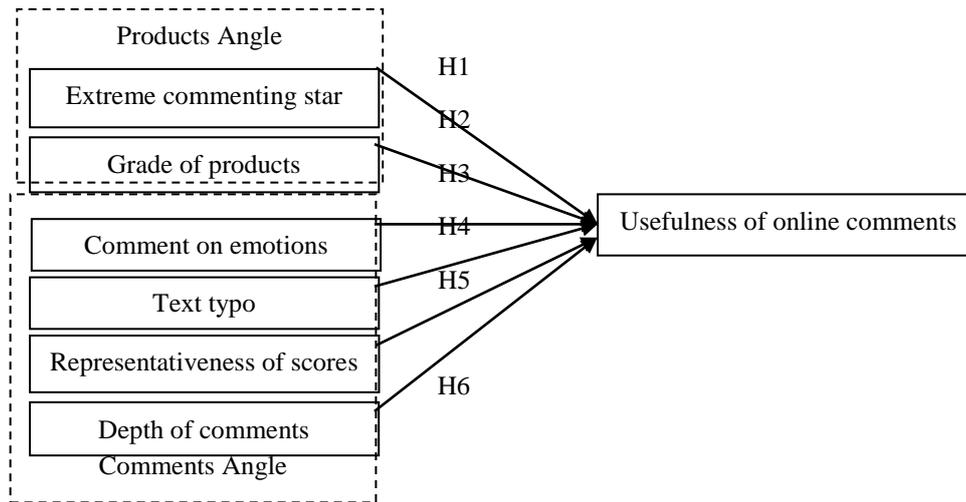


Fig. 1 Theoretical model based on hypothesis

The specific variables and descriptive statistics of the regression equation are shown in table 1 below:

TABLE I VARIABLE INTERPRETATION AND DESCRIPTIVE STATISTICS

variable	name	interpretation	mean	standard deviation
Y	Usefulness of online comments	The number of people considered useful for evaluation/total comments	0.61	0.23
X1	Extreme commenting star	Dummy variable, five - or one - star comments recorded as 1, not recorded as 0	0.72	0.45
X2	Grade of products	Reviewers rate the product on a scale of 1-5	3.07	1.76
X3	Comment on emotions	Dummy variable, exclamation point that appears is written as 1, and the absence of exclamation point is written as 0	0.24	0.43
X4	Text typo	Dummy variable, "charge" is recorded as 1	0.10	0.31
X5	Representativeness of the scores	The variance between the reviewers' ratings and the average product scores	4.30	4.24
X6	Depth of comments	Length of comment	51.42	54.04

III. SAMPLES AND DATA

The rich variety of products on Amazon and a large number of user reviews are in line with the data requirements of this study, which not only avoids the uniformity of samples, but also guarantees the representativeness of the research data. The author based on the study of the original data from Amazon reviews data sets, the original 359800 raw data, these data began on October 23, 2015, ended on April 9, 2018, the main contents were the Product ID, Product name, reviewers, commenter rating, headline, body content, products, etc. The products mainly cover the products such as Samsung, Apple and other mobile phone brands, mobile phones, mobile power sources and earphones. The price is RMB 10RMB-1000+RMB.

In view of the strong representativeness and feasibility of the data required for research and analysis, the author finally sorts the sample data into 1594 items after the 5 steps. Processed data generally include the product ID, product name, people who feel useful, comments, title, text, timing, price, the number of product reviews, product average score and so on.

When processing data, the author considers a unified product is beneficial to the research structure comparison and verification, so the author chooses move power supply as the research object, and calculates the mean and standard, finds that the processed data have good fitting degree from descriptive statistic results.

A. Correlation analysis

Comment on the significance of usefulness and product grade is 0.000, r is within 0.05, which means it has a significant meaning, the Pearson coefficient of usefulness and product rate is 0.619, thus related. According to it, the usefulness of online comments can be found at 5% level, and the usefulness of comments is strongly correlated with product ratings and extreme ratings, which is negatively correlated with the score level.

B. Regression analysis

- Explanatory power of independent variables

For the explanatory power of the independent variable, the model fit coefficient R squared of SPSS is 0.392, and the adjusted R2 reaches 0.390, indicating that the explained variable could be interpreted by the model as 39.0%. Moreover, the smaller the standard error is, the better the model will be, and the standard error is 0.18 for the analysis. In conclusion, the independent variable can explain the dependent variable better.

- Colinear diagnosis

The eigenvalues of the common linear diagnosis of independent variables does not appear to be very close to 0. And the author finds that assumption of six independent variables of tolerance is greater than 0.15, is not extremely close to 0, accordingly, VIF is less than 6.5, not more than 10, so it can be seen that the six variables' collinearity is not strong, all from the collinearity diagnosis can be retained.

- Significance diagnosis

In addition to the two independent variables of 'text error' and 'comment depth', the absolute value of the four independent variables corresponding to t is above 1.96, and the strength sig is below 0.05. In other words, the coefficient of only 4 independent variables in this model is not 0 in the 5% significance, which is statistically significant. Therefore, the 'text error' and 'comment depth' should be removed in the regression model.

Standardized regression coefficient is used to display the relative importance of variables, when you need to compare multiple independent variables on the dependent variable

relative size, using standardized regression coefficient. We can see in the table from big to small in turn are grade of products, representativeness of the score, the extreme comment star, emotion, it also explains the relative effects on various independent variables on the size of the online comments usefulness.

- Standardized residual

The multivariate linear regression is the dependent variable that conforms to the normal distribution, and the dependent variable 'online comment usefulness' basically conforms to the normal distribution, so the author supports the above model analysis.

- Heteroscedasticity test

The results of the regression analysis of requirements for unexplained residual X axis, with the dependent variable as the Y-axis scatterplot distribution shape does not show obvious trumpet-shaped, such ability is given to illustrate the effectiveness of this model. The analysis of scatter plot does not present a trumpet-shaped, but close to a ribbon, there is obvious linear relationship, so the author thinks that the analysis is reasonable and effective, assuming four independent variables can be verified.

- AIC test

Based on the equation of the front, representativeness of the scores is the smallest independent variable in the influence coefficient score, the second is emotion, so the author keeps the extreme evaluation star and product ratings, removing representativeness of the scores and emotion, establishing regression analysis.

To sum up, the author believes that the established model can be established, and the regression equation is as follows:

$$Y=0.060*extreme\ commenting\ star+0.061*grade\ of\ products+0.022*comment\ on\ emotions+(-0.010)*representativeness\ of\ the\ scores$$

IV. CONCLUSION

A. Result discussion

Based on previous data analysis, the verification results of previous assumptions can be summarized as shown in table 2:

TABLE II SUMMARY OF RESEARCH CONCLUSIONS

	Research hypothesis	Results
H1	in search products, there is a significant negative correlation between the usefulness of online reviews and the extreme level of rating of a buyer commenting on a product	unsupported
H2	in search products, the usefulness of online reviews is positively correlated with the rating of products by reviewers	supported
H3	in search products, the appearance of exclamation emotion in comment text is significantly positively correlated with the usefulness of online comments	supported
H4	in search products, the usefulness of online comments is negatively correlated with the occurrence of text typo	unsupported
H5	in search products, the usefulness of online reviews is positively correlated with the representativeness of the scores	unsupported
H6	in search products, the depth of online comments is positively correlated with the usefulness of online comments	unsupported

This article obtains data from Amazon website as the research object, in text content as a starting point, from the product reviews and comments on the previous comments to study the influence factors of online comments usefulness, which using the SPSS linear regression processing tools, finally draws the following conclusions: online comment usefulness is significantly positively related to extreme star, product grade and comment on emotions, comments usefulness is significantly negative correlation with representativeness of the scores.

Specifically, the author makes the following detailed interpretation of the research conclusion:

First, the hypothesis 'in search products, there is a significant negative correlation between the usefulness of online reviews and the extreme level of rating of a buyer commenting on a product' is unsupported. The coefficient of this independent variable is 0.060, which is a positive number. It indicates that the usefulness of online comments is positively correlated with the severity of the rating. According to the author's re-thought, the result shows that people are more concerned about the extreme scores, and the scores of 5 and 1 star will give the later purchasers a sense of reality. This may be due to water army has been removed when the writer sorts data, and it is also due to the high score can effectively explain the advantage of this product, to give those potential buyers psychological support; A low score can reveal the weakness of the product and the dissatisfaction of the reviewers, which will give more objective reference to later purchasers, so both can make consumers think those comments are useful.

Second, the hypothesis 'in search products, the usefulness of online reviews is positively correlated with the rating of products by reviewers' is supported. According to the data analysis, the coefficient of the analysis is 0.061, which is also the most significant factor, which is in line with the author's expectation. It indicates that a relatively high score can indeed win favor and approval from buyers. This is exactly what online sellers want now.

Third, the hypothesis 'in search products, the appearance of exclamation emotion in comment text is significantly positively correlated with the usefulness of online comments' is supported. Analyzing its coefficient is 0.022, is also a positive number, the author understands it as a comment presents exclamation mark not only won't affect the authenticity of the comment, it may show the full expression of the relationship and enhance its authenticity and credibility.

Fourth, the hypothesis 'in search products, the usefulness of online comments is negatively correlated with the occurrence of text typo' is unsupported. I think this is probably although there is text typo, it is useful for subsequent buyers, especially the development of the Internet age, making the most of people show lower attention on text form and correction, and show more attention on text's information, which means online comments are not perceived to be significantly less useful because of errors in the comments.

Fifth, $B=0.010$ available through the data analysis, the hypothesis 'in search products, the usefulness of online reviews is positively correlated with the representativeness of the

scores' is unsupported. The coefficient is -0.010, which means that the usefulness of online comments in search products is significantly negatively correlated with the rating representativeness of comments. I understand it as when the comments' readers find rate close to the average scores of other consumers, they will think that these comments lack credibility, and thus they more tend to think that comments with higher variance are more useful.

Based on the conclusions of the first and second study, the author speculates that, in terms of rating, the usefulness of comments and the status of rating are generally in line with the u-shaped distribution, which is partially corrected.

Sixth, after analysis, the hypothesis 'in search products, the depth of online comments is positively correlated with the usefulness of online comments' does not hold. Writer will interpret it as the length of the text is not a factor in whether online comments are useful or not, consumers may pay more attention to whether keywords they care about appear in online comments, so some very long online comments probably are useless for consumers, the usefulness of extremely short online comments is not necessarily low.

B. Theoretical contribution and management implications

• Theoretical contribution

In theory, in this paper, the research of the predecessors' theory is obtained in the text content, and search products online reviews research field is improved, the influence of the usefulness of the information acceptance model in the field of online comments usefulness can have some new interpretation, and the paper puts forward with product reviews and comments on the review of the new model for data demonstration and analysis, the final research conclusion and constructive suggestions are put forward smoothly.

• Management implications

Variable quality and comment information are overload, these seriously interferes with the readers' effective judgment on the quality of good, thus increasing the cost of information search, reducing the decision-making efficiency, therefore, consumers need to evaluate usefulness to identify goods; In order to manage users' comments quickly and effectively, e-commerce enterprises also need to find out the important factors and influence factors that affect the usefulness of online comments. The practical significance of this paper is to study the influence of text content on the usefulness of online comments on search products, which can play an auxiliary role in enterprises' marketing activities and consumer decision-making.

Amazon has yet to be perfect, but given the comment on the reader's actual need, there is still more worthy of improvement, such as by adding text prompt and encouraging consumers to comment more objectively. For example, comment content can be subdivided into physical objects and product description, website description and so on. Of course, the above research also has certain enlightenment to the consumer.

In view of the first study, the author suggests for consumers that consumers are watching comments should pay attention to the extreme of the usefulness of marking, but at the same time, consumers should also prevent water army, not only improve their reasoning abilities, but also pay attention to product coefficient and quality of the product itself the objective description. For merchants, the conclusion is sure of their current practice, it is useful to promote the brand, but they need to pay attention to product quality as the first principle, only in this way the consumers will give product due scores, which makes readers think the comments are effective.

For the second study, the author suggests consumers can pay attention to how high or low the rating is when they watch the comments, as the basis of their own whether to buy or not. With regard to business, they can encourage buyers comment for the goods with high scores, which can influence more potential buyers think these comments are useful to produce purchase behavior, in the midst of this, they can adopt certain incentives, such as high score for coupons, of course the most important is to promote the quality of the product.

The third study finds that consumers should pay attention to the sentiment of the phrasing in the comments, but at the same time, they should prevent the flooding of the Internet. For businesses, they can make the conclusion that there is no doubt that they carry out the following measures, encouraging buyers comment more earnestly, and that can influence more potential buyers think these comments are useful to produce purchase behavior.

For the fourth study, the author suggests that consumers should be aware of the motives of extreme ratings when viewing comments, in case they are deceived. For merchants, the conclusion encourages them to invite the appropriate buyer to rate the content satisfied with the product in the comments with high marks, give low marks for content that they are not satisfied with but it is best to include a compliment for certain qualities of the product, which can influence more potential buyers think these comments are useful to produce purchase behavior.

C. Research innovation and prospect

The author thinks that, in this paper, the main innovation is reflected in the following aspects:(1) at present, most researches on the usefulness of online comments in China have made suggestions to merchants, in addition to, the paper also gives the reference to the consumer. (2) the existing literature in the process of building models and assumptions lack reference theory framework, the author read through the literature, and the paper proposed to study the factors influencing the usefulness of online comments from two perspectives of product reviews and previous reviews, and regard the information dissemination and reception theory as an important basis of online review research under the Internet environment. (3) The deletion process of the data reflects the author's own thinking and certain rationality.

In this paper, the main disadvantages are as follows: the research of this paper focuses on Amazon website, and avoids

the uniformity of samples through different sources of data. (2) some online comment content needs further verification, and finally in this paper the fit of the R square is less than 0.5, shows that some more important variables of online comments' usefulness are not into the model, such as commentators' ranking, commentators' properties such as buyers' experience. It means some limitations. (3) It does not include the usefulness of commodity brands in online reviews. Although the filtered data show that the products are basically brand products, it is more reasonable to propose it as a hypothesis.

ACKNOWLEDGMENT

Firstly, thank teacher Yi for his tireless teaching. My teacher always looks up at the stars and keeps his feet on the ground. He treats students as if we were his own children, in addition to teaching us how to do academic work, and let us harvest and grow in various projects. During my study, I was lucky to investigate many places, meet many different people and learn a lot. He is sincere, active and responsible. I will always remember these qualities. Secondly, thank brother-sister-school students for your kindness. We always help each other. I am very happy to learn and grow with you. Thirdly, thanks to my parents' selfless dedication, my parents have always been my strong support. They are my relatives and friends. I am very happy to meet them in this life. And I will always be myself, never let myself be the child of others, what is more, I will strive to be their pride, I will try to have a happy, healthy and happy life. Fourthly, thank classmates for helping me unselfishly when I am helpless. We study together and exercise together, so that I am not lonely on the academic road. Fifthly, thanks to the dormitory administrator aunts, who treat themselves like my parents and take care of me when I am sick. Without them, I can not recover from illness so quickly. Sixthly, thank my boyfriend for not coming, let me concentrate on my academic study, and keep improving myself. In my spare time, I read, swim and watch movies instead of spending time thinking of each other. Finally, thank myself for not abandoning or giving up, although I was ill, I didn't stop learning, and I has been working hard, although tired but has been happy. In addition to learning how to write a regular article, I also learn how to be a human being, how to help others and how to accept others' help, learn to be gentle and learn to be strong, this will be the treasure of my life.

REFERENCES

- [1] David F D.A technology acceptance model for empirically testing in to negative consumer comments on the Web [J].*Journal of Interactive Marketing*, 2007, 21(4):76-94.
- [2] Cao Q, Duan W, Gan Q. Exploring determinants of voting for the "helpfulness" of online user comments: A text mining approach [J].*Decision Support Systems*, 2011, 50(3):511-521.
- [3] Sussman S W, Siegal W S. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption [J].*Information Systems Research*, 2003, 14(1):47-65.
- [4] Jiang xiaodong. What kind of product reviews are most useful? - a study on the influence of quantitative characteristics and text characteristics of online comments on their usefulness [J]. *Foreign economics and management*, 2015, 37 (4):41-54. (In Chinese)

- [5] Yin guopeng, liu wenwen, zhu shan. Research on the usability impact model of online community online review - based on information adoption and social network perspective [J]. *E-commerce*, 2012, 56(16):140-147. (In Chinese).
- [6] Liu chenglin, CAI chunjiang, li yi. An empirical study on the factors influencing the usefulness of online comments in e-commerce [J], *soft science*, 2013, 27(5):46-50. (In Chinese).
- [7] Yan jianyuan, zhang li, zhang lei. Empirical research on the impact of online comment content on the usefulness of online comment in e-commerce [J]. *Intelligence science*, 2012, 30(5):714-719. (In Chinese).
- [8] Mudambi S M, Schuff D. What makes a helpful online comment? a study of customer comments on amazon.com [J]. *MIS Quarterly*, 2010, 34(1):185-200.
- [9] Wang ping, dai bao. Empirical research on the impact of consumer online comments on usability-related insofar [J]. *Statistics and decision-making*, 2012, 2:118-120. (In Chinese)