Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2009 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

July 2009

Why some online product reviews have no usefulness rating?

Yuan Yuan Hao
Harbin Institute of Technology, hyy0112@hotmail.com

Yi Jun Li
Harbin Institute of Technology, liyijun@hit.edu.cn

Peng Zou
Harbin Institute of Technology, zoupeng@hit.edu.cn

WHY SOME ONLINE PRODUCT REVIEWS HAVE NO USEFULNESS RATING？

Hao Yuan Yuan
PH.D. Candidate of Management Science and Technology
Harbin Institute of Technology
Harbin, Heilongjiang 150001, China
hyy0112@hotmail.com

Li Yi Jun
Professor of Management Science and Technology
Harbin Institute of Technology
Harbin, Heilongjiang 150001, China
liyijun@hit.edu.cn

Zou Peng
Lecturer of Business Administration
Harbin Institute of Technology
Harbin, Heilongjiang 150001, China
zoupeng@hit.edu.cn

Abstract

Combining econometric analysis with text mining techniques, this study attempts to explore why some online product reviews have no usefulness rating through examining review posting time and text features. Later posting time may reduce the probability of some online reviews being seen and thus lead to their being not rated for usefulness. Besides, the neutral diagnosticity of reviews reflected from the text features may cause difficulty for readers to judge and evaluate the usefulness of these reviews. Our study finds that, though not being seen due to later posting time obviously explains no usefulness rating for some online reviews, the neutral diagnosticity of these reviews is also an important and non-neglectable cause for their lack of usefulness rating. Further, we identify the text features which may lead to the neutral diagnosticity of the review. Our study has implications for online product reviews website managers in identifying and dismissing the reviews with no usefulness rating to improve readers' information retrieving efficiency and also for reviewers in improving the quality of their reviews.
Keywords: Online Product Reviews, Usefulness Rating, Text Features, Posting Time, Binary Logistic Regression.

This study was partially funded by National Science Foundation of China (70771032, 70890080-70890082).

1. INTRODUCTION

With the widely developing of Internet applications, online reviews have become a very important source of word of mouth (Dellarocas, 2003). Many websites publish user generated reviews about products. Some websites also allow readers to rate usefulness of these reviews. The usefulness rating for online reviews makes the information retrieving and decision making more efficient for potential consumers. Besides, it also inspires consumer's voluntary behavior of posting reviews. On some websites, like yahoo.com, the usefulness rating for online reviews is usually shown in the form of a usefulness vote ratio (ratio of helpful votes to total helpfulness votes), which is cumulatively calculated according to opinions of previous readers. For a review, higher the vote ratio is, more useful it is. For later readers, a review may be perceived as useful and worth of reading if this indicator is above 50% or higher, and not worth reading if it is below 50%. Thus, an online review may have three states of usefulness ratings: useful, useless or no usefulness rating. Several researchers have explored to identify the useful and useless online reviews by text features (Ghose and Ipeirotis 2004/2007, Sen and Lerman 2007). However, the great amount of online reviews with no usefulness rating mixed within the useful and useless reviews have been ignored up to now. Why no reader rates these reviews? What factors can explain this phenomenon?

Clarifying the factors that may explain why some online reviews have no usefulness rating can be valuable in the following aspects: From the reviewers' view, if these reviews lack of usefulness rating are only attributed to having not been read due to the later posting time, automatically identifying and scoring the usefulness of these reviews through text mining techniques may stimulate reviewers to post more high quality reviews; Otherwise, if neutral diagnosticity of these reviews also has significant impact, finding out text features indicating their neutral diagnosticity may help reviewers to enhance the reviews usefulness. From the consumers' perspectives, if neutral diagnosticity of these reviews also has significant influences beside of posting time, automatically identifying and removing these reviews can be helpful in enhancing consumers' information filtering efficiency.

Thus, this study aims to explore the underlying characteristics of no usefulness rating reviews by mainly focusing on posting time and text features related factors. Though text mining techniques are helpful for achieving this study target by automatically identifying no usefulness rating reviews, it may be powerless in answering the following questions: What differences there are among reviews with usefulness rating and ones lack of usefulness rating? Further, how to convert potential no usefulness rating reviews to useful reviews for review posters? What factors have more weight on explaining no usefulness rating for some reviews? Thus, we try to combine the text mining techniques and econometric regression analysis to solve these questions. In our study, text mining techniques are used to automatically extract text features from review textual contents. Then we employ two binary logistic regression models to identify the differences in these factors between the reviews with no usefulness rating and ones with useful or useless rating. According to the results of difference comparisons, we further attempt to explain why some reviews are lack of usefulness rating.

The rest of this study is arranged as follows: First, we identify three situations which can lead to no usefulness rating for some online reviews; Based on these situations, we identify the factors (especially text features) causing these situations to happen. Then, we establish two binary logistic regressions to compare the differences in these factors between reviews with no usefulness rating and ones with useful rating or useless rating. Finally, we provide conclusions and managerial implications.

2. FACTORS EXPLAINING WHY SOME ONLINE REVIEWS HAVE NO USEFULNESS RATING

According to Ahluwalia et, al. (2000), the information processing procedure of a review may be broken into two steps: (i) decision to pay attention to and read the review and (ii) the actual processing of the information and the decision to use it because it is helpful (relevance or diagnosticity of the information). Extending on this frame, we develop this process into the three stages (see Figure 1 below): See an online review, decide to read it, and decide to use it. Every stage impacts the latter stage. No see, no read. No read, no use and rate. Based on these three stages, we identify three corresponding situations in which consumers may not rate usefulness for some reviews: (1) These reviews have not been seen. No see, no impact at all. (2) Though the titles of these reviews have been scanned, consumers may not decide to further read the full text because of the titles diagnosticity (or unattractiveness). (3) Though consumers have read the full texts of these reviews, they may think them neutral in diagnosticity and hard to give definite usefulness rating.

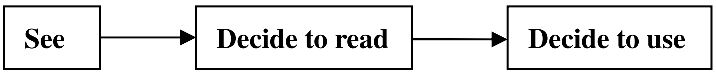

Figure 1. Three stages for information processing procedure of an online review

For the first situation, the time length from product release to review posting may impact the probability of the review being seen. On one hand, the later/delayed posted reviews might have less chance to be seen because of weakened attentions and decreased information retrieving behaviors on the product as time goes by after the product release. Besides, some earlier posted reviews are less likely to be seen for a reader due to information overload (Later reviews may provide enough information). However, every review posted in different time always has chance to be seen and even rated by some readers. Thus, we expect one reason that no one rates the usefulness for some reviews is: These reviews were not seen due to delayed posting time rather than information overload.

For the second and third situation, whether people decide to read and rate the review separately depends on the diagnosticity (or attractiveness) of its title and full text. Though the grade rating (valence) shown beside a certain review may indicate its general opinion on some product, people still need to depend on review textual contents (including review title and full text) for further product evaluation (Ghose and Ipeirotis 2008). It can be justified by the phenomenon that reviews with the same grade rating have different usefulness ratings. In the texts of a review, an attractive title may drive consumers to read its full text further. However, whether consumers decide to use and rate a review depends on the diagnosticity of its full text. Thus, we expect diagnosticity (or attractiveness) of both review title and full text may partly account for why some reviews are lack of usefulness rating.

From the view of the text features, several factors may indicate review diagnosticity, such as positive vs. negative orientation and deviation, subjective rating vs. objective description orientation and deviation, subjective expression vs. objective expression orientation and deviation, average sentence length of reviews. (Ghose and Ipeirotis 2004/2007, Sen and Lerman 2007) As for the word-of-mouth communication direction, **negative or positive**, some researches indicate that the direction of WOM impacts the consumers' perception of WOM value. Negative WOM has greater influence on consumers' purchase decision than positive one. In psychology field, Skowronskij and Carlston (1989)'s study finds that, negative ratings are usually granted with more weights than positive ratings in the process of some object evaluation. It is because that the psychological responses triggered by negative rating, such as arouse, awareness, emotion, attribution, are stronger than that by positive one. And in marketing field, Ahluwalia et, al. (2000) also finds that people depend more on negative rating information than positive one because negative rating information is more diagnostic. Therefore, we expect that positive orientation online reviews are more likely to have no usefulness rating due to their weaker diagnosticity, compared with the negative ones. The direction of online reviews may partly explain why they are lack of usefulness rating. Besides, according to Ghose and Ipeirotis (2004/2007), an online review is usually a mixture of positive and negative sentences and the mixture degree between these two orientations sentences may influence the consumers' perception of reviews diagnosticity. Thus, we also examine the possible effect of the two orientations sentences deviation (namely the positive deviation below) for each review on its usefulness.

From the stylistic point of view, contents of online reviews can be classified into two types: **subjective ratings** and **objective information descriptions**. According to Pang&Lee (2004) and Ghose&Ipeirotis (2007), objective information descriptions are defined as follows: review contents that list "objective" information, listing the characteristics of the product and giving an alternate product description that confirms (or rejects) the description given by the merchant; Subjective ratings are defined as follows: review contents with "subjective", sentimental information, in which the reviewers give very personal evaluations of the product, and give information that typically does not appear in the official description of the product.. Ghose and Ipeirotis (2007)'s study on audio/video players and digital cameras finds that a review with high subjective rating orientation and low deviation between these two orientations contents may have a low usefulness. Therefore, we also expect that reviews with high subjective rating orientation and low deviation between these two categories of sentences (namely subjective rating deviation below) are more likely to have neutral diagnosticity and thus lack of usefulness rating.

From the aspect of text expression manners, Duhan et, al. (1997) classified review contents into two types: affective cues evaluations and instrumental cues evaluations. They defined the differences between these two contents: Affective cues evaluations are generally based on reviewers' intrinsic subjective criteria (such as aesthetic, art effects) (e.g. "My wife hates blue cars."). Whereas instrumental cues evaluations are generally based on characteristics of the product (such as technical or performance oriented aspects) (e.g. "This car gets good gas mileage."). Affective cues evaluations are more related to the reviewer, whereas instrumental cues evaluations are more independent of the reviewers. According to these comparisons, we expect instrumental cues evaluations are more objective than affective cues evaluations. In addition of the evaluative cues, we find different expression tones and wordings can also impact the objectivity of expression by observing several online reviews. For example, "a must see film." may be more objective and acceptable than "you must see it!". So based on Duhan et, al. (1997)'s taxonomy, we classify and define two types of review contents according to expression manners (including evaluative cues, tones and wording): **Objective expression ratings** and **subjective expression ratings**. Both objective expression ratings and subjective expression ratings belong to subjective ratings. But they are distinguished in expression manners. In details, "Objective expression ratings" are defined as ratings that objectively evaluate total or particular features of the product, generally taking the third person tones in form of statement sentences, which is similar with critics' ratings in expression. (e. g. "This movie is in a good quality.

The animation has improved quite a lot and is spectacularly realistic to watch. The music and sound effects were also buzzing with delightful energy.") Totally different from objective expression ratings, "subjective expression ratings" are defined as ratings that are opinioned or attitudinal ratings with strong personal and subjective colors. (e. g. "For us, ICE AGE2 is wonderful. We love Sid with big eyes!", "Terrible. even small children with no understanding of what poor acting is will think this movie if poorly acted.") Objective expression ratings usually evaluate the quality of product in more objective expression form and may be more persuadable than subjective expression ratings. Therefore, we expect that reviews with high subjective expression orientation are more likely to have neutral diagnosticity and thus lack of usefulness rating. Besides, we also examine the possible effect of deviation between these two categories of sentences (namely the subjective expression deviation below) for each review on its usefulness.

Online reviews usually consist of two parts: **title** and **full text**. Different from offline WOM which is usually passively received, online reviews have to be achieved through consumers' actively searching. Online surfing consumers have more freedom to selectively read useful online reviews they are interested in. Usually, potential consumers decide whether to read the full text of one review firstly according to its title. As the summary and advertisement of the whole review, the more attractive the title is, the more likely it is for consumers to read the full text content. Since the diagnosticity (or attractiveness) of title impacts the probability of reviews full texts being read and further impacts the later usefulness rating, we distinguish title from full text and separately examine their effects of the above mentioned text features on reviews usefulness. Because title of an online review is usually short and has only one sentence, the deviation indicator is not examined.

Beside of the above text features, we also measure the average number of words per sentence in the review, defined as the ratio of the length of the review in words to the number of sentences. Ghose and Ipeirotis (2007) think this indicator can reflect the cognitive cost or readability (how easy it is for a user to read a review). In our study, we find this factor can also indicate the information enrichment of reviews which may impact reviews usefulness rating. Reviews with greater average sentence length are likely to have richer product evaluation information and higher usefulness rating. Thus, we expect the average number of words per sentence in the review has an impact on review usefulness.

3. RESEARCH METHOD AND DATA

In our study, we select online movie reviews for further empirical analysis mainly out of the following considerations: First, as a typical experience goods industry, movie industry is significantly influenced by the WOM (especially e-WOM), which has become the important reference sources for consumers. Second, usefulness rating mechanism of online movie reviews is completely developed and movie reviews data are detailed and publicly available.

The research procedure is conducted as follows: First, according to the usefulness vote ratio which has been accumulatively calculated since movie release, we classify online sample movie reviews into three categories: useful reviews, useless reviews and no usefulness rating reviews. Useful reviews are defined as those whose usefulness vote ratio is above 50% with the total usefulness vote above 1; and useless reviews are defined as those with usefulness vote ratio not more than 50%. Then, the probability values of above mentioned text features orientations or deviations are assessed through text mining techniques. Next, we separately establish two binary logistic regressions with all independent variables including probability values of text features and review posting time length since movie release. In these two regressions, we further identify the differences of reviews with no usefulness rating from ones with useless rating and useful rating in text features and posting time.

3.1 Data Collection

Data in this study are collected by automatically crawling and parsing online movie reviews web pages through Java programs. Online movie reviews are extracted from Yahoo! Movies (movies.yahoo.com), an American popular movie reviews website. The usefulness vote ratio of a movie is usually changing and not steady within its life cycle. Considering the usefulness vote ratio after the movie's life cycle closes is ready and more representative of review usefulness, we sample movies published in the previous years. Firstly, in the 500 top box-office movies list of Year 2006, two movies are randomly sampled for each of seven types of movies, Animation, Comedy, Horror, Thriller, War, Adventure, Fiction. Next, on Yahoo! Movies, online movie reviews data for these 14 sample movies are crawled, including consumers' numerical rating for the movie, review posting time, usefulness rating, review title, and review full text. (See Figure 2 for a sample of the information about these data shown on Yahoo! Movies). In Figure 2, the usefulness rating of this review is shown as "6 of 11 people found this review helpful", indicating the helpful votes are 6 and the total helpfulness votes are 11 for this review. Thus, the usefulness vote ratio for this review is about 55%. And the overall positive/negative orientation grade rating of this review for the related movie is "B-" (Yahoo! Movies totally assign 13 rating grades), which equals to 9 in the 1-13 numerical grades if converted into the numerical rating.

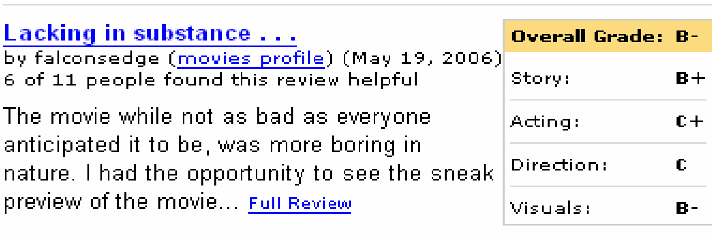

Figure 2. Information presentation of reviews on Yahoo! Movies website

Since the reasons related with review diagnosticity are practically more meaningful, we mainly analyze the reviews within the first week for these sample movies. Usefulness rating in the form of ratio of helpful votes to total votes is classified into useful rating and useless rating by the criteria of 50%. In the original data, there are some reviews only having one total vote. Considering the usefulness rating of this kind of reviews is poor in representation, so we remove these reviews with only one total usefulness vote. Finally, 3332 sample reviews data are achieved for empirical analysis.

3.2 Text Feature Identification Method

We calculate the text feature, the average number of words per sentence in the review, through dividing the total numbers of words by numbers of sentences in the review. Besides, for positive/negative orientation of the online review full text for the related movie, we select the reviewer's overall rating grade shown beside it as the indicator, which is original and objective compared with the probability values calculated through text mining method.

For the other text features, referring to Ghose and Ipeirotis (2007)'s study, instead of classifying each review, we first classify each sentence in each review and assess the probabilities of text features each sentence, then calculate the average value and standard deviation of these probabilities of each sentence as scores of the whole review text features orientations and deviations. E.g., subjective expression orientation of one review is the average value of probability being subjective expression of each sentence; deviation between subjective expression and objective expression sentences of one review is the standard deviation of probability being subjective expression of each sentence.

The probability scores of these text features, including subjective ratings orientation and deviation of full text, subjective expression orientation and deviation of full text, positive orientation, subjective ratings orientation, subjective expression orientation of title, are all identified and assessed by machine text mining tool LingPipe (available at www.alias-i.com/lingpipe). The procedure of mining and evaluating these text features probabilities is that: Create corresponding classification machines using Dynamic Language Model by Java program, and train these classification machines with already manually classified training data collections, then evaluate the classification outcomes with additional testing data. Finally, calculate the text features scores of 14 sample movies' reviews with well trained classification machines.

The training and testing data for text feature identification are mainly derived from two sources: one is from the existing text materials base provided by Pang and Lee (available at http://www.cs.cornell.edu/people/pabo/movie-review-data/), and another is from reviews of other movies except of 14 sample movies on Yahoo! Movies. In details, (1) To classify subjective ratings and objective information descriptions, according to the above definitions, we select 800 texts of movie plots in the movie plots introduction materials base provided by Pang and Lee as training and testing data for movie objective information descriptions. (Of this data collection, 2/3 are used for training, and the other 1/3 are used for testing, the ratio arrangement of training and testing data is the same below); On the other hand, we sample 800 subjective rating sentences from other movies except of 14 sample movies on Yahoo! Movies, as training and testing data collections for movie subjective ratings. (2) To classify subjective expression ratings and objective expression ratings, according to the above definitions, 500 critic review texts are randomly extracted from large-scale famous movie reviews organizations listed on Yahoo! Movies, as training and testing data for objective expression ratings; On the other hand, 500 subjective expression sentences with strong personal colors are extracted from online reviews of other movies except of 14 sample movies as training and testing data for subjective expression ratings. (3) Considering positive deviation of review full text still needs to be evaluated by text mining, we separately extract 400 positive and 400 negative sentences of other reviews out of 14 sample movies as training and testing materials.

The three classification machines for classifying subjective ratings vs. objective information descriptions, positivity vs. negativity, and subjective expression ratings vs. objective expression ratings, are trained and then tested separately using the above mentioned corresponding training and testing materials. The classification correctness ratios for these well-trained classification machines are separately 85.15%, 83.34% and 78.71%, which show that classification effects are all acceptable.

3.3 Binary Logistic Regression Models

To compare and identify the differences between no usefulness rating reviews and useful or useless ones, we set the dependent variable, usefulness rating, as a categorical variable with two levels, useful vs. no usefulness rating, or useless vs. no usefulness rating. Independent variables are continuous ones, including review text features and posting time length mentioned above. Binary logistic regression is suitable for this analysis task. We establish two original binary logistic regression models: one for no usefulness rating reviews vs. useless ones, and another for no usefulness rating reviews vs. useful ones. (See Equation 1 and 2, and see Table 1 for specifications of variables in these equations) In both of the equations, “no usefulness rating” is coded as the reference category of the dependent variable.

$$\log it(P_{usefulness=useless}) = \alpha_0 + \alpha_1 \cdot Rating_{kr} + \alpha_2 \cdot (DevPos)_{kr} + \alpha_3 \cdot (AvgSub)_{kr} + \alpha_4 \cdot (DevSub)_{kr} + \alpha_5 \cdot (AvgSubExp)_{kr} + \alpha_6 \cdot (DevSubExp)_{kr} + \alpha_7 \cdot (title\_AvgPos)_{kr} + \alpha_8 \cdot (title\_AvgSub)_{kr} + \alpha_9 \cdot (title\_AvgSubExp)_{kr} + \alpha_{10} \cdot Elapsedays_{kr} + \alpha_{11} \cdot Read_{kr} + \varepsilon_{kr} \quad (1)$$

$$\log it(P_{usefulness=useful}) = \beta_0 + \beta_1 \cdot Rating_{kr} + \beta_2 \cdot (DevPos)_{kr} + \beta_3 \cdot (AvgSub)_{kr} + \beta_4 \cdot (DevSub)_{kr} + \beta_5 \cdot (AvgSubExp)_{kr} + \beta_6 \cdot (DevSubExp)_{kr} + \beta_7 \cdot (title\_AvgPos)_{kr} + \beta_8 \cdot (title\_AvgSub)_{kr} + \beta_9 \cdot (title\_AvgSubExp)_{kr} + \beta_{10} \cdot Elapsedays_{kr} + \beta_{11} \cdot Read_{kr} + \varepsilon_{kr} \quad (2)$$

4. ANALYSIS RESULTS

4.1 Descriptive Statistics

The specifications and descriptive statistics of the independent variables are listed in Table 1.

Indepent Variable	Variable Specification	Valid Obs.	Mean	Std.	Min.	Max.
$Rating_{kr}$	numerical rating of Review r for Movie k (indicating Review r full text positive orientation)	3332	10.050	3.660	1	13
$(DevPos)_{kr}$	positive deviation of Review r full text for Movie k	3332	0.301	0.181	0	0.999
$(AvgSub)_{kr}$	subjective rating orientation of Review r full text for Movie k	3332	0.884	0.177	0	1
$(DevSub)_{kr}$	subjective rating deviation of Review r full text for Movie k	3332	0.146	0.188	0	0.567
$(AvgSubExp)_{kr}$	subjective expression orientation of Review r full text for Movie k	3332	0.474	0.317	0	1
$(DevSubExp)_{kr}$	subjective expression deviation of Review r full text for Movie k	3332	0.294	0.197	0	0.500
$(title\_AvgSub)_{kr}$	subjective rating orientation of Review r title for Movie k	3332	0.930	0.238	0	1
$(title\_AvgSubExp)_{kr}$	subjective expression orientation of Review r title for Movie k	3332	0.455	0.465	0	1
$(title\_AvgPos)_{kr}$	positive orientation of Review r title for Movie k	3332	0.585	0.442	0	1
$Read_{kr}$	average number of words each sentence of Review r full text for Movie k	3332	14.800	9.437	2	120
$Elapseddays_{kr}$	the days differences from Movie k release to Review r posting	3332	3.360	1.835	1	7

Table 1. Specifications and descriptive statistics of the independent variables

Encoding method and frequency statistics for the categories of the dependent variable in model 1 and model 2 are listed in Table 2.

Model	Category of Dependent Variable	Frequency	Encoding Value
Model 1	review has no usefulness rating	2123	0
	review is useless	573	1
Model 2	review has no usefulness rating	2123	0
	review is useful	636	1

Table 2. Encoding method and frequency statistics for the dependent variable categories

We also examine the average percentage statistics of no usefulness rating reviews every week in the first month for 14 sample movies (see Table 3). From Table 3, after the first week of movie release, average over 80% of all reviews every sample movie are ones with no usefulness rating. It shows the trend of later posted reviews being less likely to be seen and rated as time goes after the first week. Thus, based on our study target of finding out the underlying characteristics of no usefulness rating reviews except of review posting time, it is suitable to analyze the reviews of these samples within the first week.

	First Week	Second Week	Third Week	Fourth Week
Mean	52.85%	83.16%	80.78%	89.89%
Std.	0.27	0.2	0.25	0.14

(Note: the percentage values in the first row of the table are calculated by averaging the ratios of no usefulness rating reviews to all reviews every sample movie in the corresponding week)

Table 3. Average percentage statistics of no usefulness rating reviews every week in the first month for 14 sample movies

4.2 Results of Binary Logistic Regression Models

To avoid the multicollinearity effects, we adopt Forward Stepwise Likelihood Ratio method for screening independent variables into the regression model. To demonstrate whether review text features obviously contribute for model variances, we compare the overall fitness of each model with all variables (full model) and its incomplete model with only the variable *Elapeddays* for the two models. The model fitness indices for full models and their corresponding incomplete models are listed in Table 4. From Table 4, the two pseudo R-Squared indices (Cox & Snell R Square and Nagelkerke R Square) for Model 1 and Model 2 are separately 0.129 and 0.229. For large sample data, these values, usually not more than 2, indicate acceptable model fitness. Besides, compared with that of the models with only *Elapseddays*, the overall fitnesses of full models for Model 1 and Model 2 are both obviously improved, including the decreased -2log likelihood values and two increased pseudo R-Squared indices. It shows that these review text features significantly explain the partial variances of the dependent variable.

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Model 1 with only *Elapseddays*	2638.722	0.062	0.097
full Model of Model 1	2359.156	0.147	0.229
Model 2 with only *Elapseddays*	2858.372	0.051	0.077
full Model of Model 2	2732.570	0.086	0.129

Table 4. Model fitness indices of the four models

The results of coefficients estimations for these two regressions are shown in Table 5. For the independent variables, standardized coefficients should be used for comparing the effects of them due to their difference in dimensions. The results of model 1 (See the former four columns of Table 5) show that no usefulness rating reviews differ from useless ones mainly in two aspects: one is posting time and another is text features of review full text. Compared with useless reviews, reviews with no usefulness rating are more likely to be those with later posting time (indicated by variable $Elapseddays_{\mathrm{kr}}$, standardized coefficient of which is -0.434), or higher positive orientation ($Rating_{kr}$, -0.400), or less subjective rating deviation of full text ($(DevSub)_{kr}$, 0.06). According to the effects of these factors, the main causes for no usefulness rating reviews being not rated to be useless are twofold: On one hand, having not been seen due to later posting time accounts for it; Besides, when posting time and other text features are controlled to be the same, high positive orientation reviews, which may be regarded as the low diagnosticity ones, are less likely to be rated for usefulness.

	Model 1 (useless rating vs. no usefulness rating)				**Model 2 (useful rating vs. no usefulness rating)**			
	B	**sig.**	**exp(B)**	**standar-dized B**	**B**	**sig.**	**exp(B)**	**standar-dized B**
$\boldsymbol{Rating_{kr}}$	-0.198	0.013	0.820	-0.400	-	-	-	-
$\boldsymbol{(DevPos)_{kr}}$	-	-	-	-	0.944	0.001	2.570	0.094
$\boldsymbol{(AvgSub)_{kr}}$	-	-	-	-	3.140	0.000	23.094	0.306
$\boldsymbol{(DevSub)_{kr}}$	0.574	0.035	1.776	0.060	2.942	0.000	18.956	0.305
$\boldsymbol{(AvgSubExp)_{kr}}$	-	-	-	-	-0.486	0.003	0.615	-0.085
$\boldsymbol{(DevSubExp)_{kr}}$	-	-	-	-	0.710	0.008	2.034	0.077
$\boldsymbol{(title\_AvgSub)_{kr}}$	-	-	-	-	-	-	-	-
$\boldsymbol{(title\_AvgSubExp)_{kr}}$	-	-	-	-	0.281	0.007	1.324	0.072
$\boldsymbol{(title\_AvgPos)_{kr}}$	-	-	-	-	-	-	-	-
$\boldsymbol{Read_{kr}}$	-	-	-	-	0.035	0.000	1.036	0.182
$\boldsymbol{Elapsedday_{kr}}$	-0.429	0.000	0.651	-0.434	-0.324	0.000	0.723	-0.328
Constant	1.784	0.000	5.952	-	-4.339	0.000	0.013	-

(Note: "-" denotes the coefficients of the corresponding variables are not significant at the level of 0.05.)

Table 5. Coefficient estimation results for the two models

According to the results of model 2 (the later four columns of Table 5), reviews with no usefulness rating differ from useful reviews in three aspects: posting time, text features of review full text, and text features of review tile. Compared with useful reviews, reviews with no usefulness rating are more likely to be those with later posting time (indicated by variable $Elapseddays_{kr}$, standardized coefficient of which is -0.328), or lower subjective rating orientation of full text ($(AvgSub)_{kr}$, 0.306)，or less subjective rating deviation of full text ($(DevSub)_{kr}$, 0.305), or less average number of words per sentence ($Read_{kr}$, 0.182)，or less positive deviation of full text ($(DevPos)_{kr}$, 0.094), or higher subjective expression orientation of full text ($(AvgSubExp)_{kr}$, -0.085), or less subjective expression deviation of full text ($(DevSubExp)_{kr}$, 0.077), or lower subjective rating orientation of title ($(title\_AvgSub)_{kr}$, 0.072). One main cause for some reviews being lack of usefulness rating is the later posting time. However, when posting time factor is controlled to be the same, the neutral diagnosticity of review full text is also an important cause. The neutral diagnosticity is reflected by some text features including: Less subjective rating orientation and deviation of full text, less average number of words per sentence, less positive deviation of full text, higher subjective expression orientation and less subjective expression deviation of full text. In addition, subjective rating orientation of title also has significantly positive impact, indicating poor attractiveness of the title may hamper readers from further reading full text for the review with no usefulness rating.

5. DISCUSSIONS

In this study, we classify the usefulness rating of reviews into three categories, in which useful rating and useless rating are distinguished by the criteria of 50% usefulness vote ratio. Considering the results may be impacted by the different classification criteria values, we also distinguish the useful rating and useless rating by 60% and re-conduct the binary logistic regressions. We find the similar results with those under the criteria of 50%. To test the reliability and independence of the analysis results in our study, we also analyze the two models with first two weeks, first three weeks and the first month reviews data for 14 sample movies. As a result, we find similar results for review text features differences among three types of reviews, which show the results in this study are reliable.

6. CONCLUSION

Based on the three stages of information processing procedure, we attempt to explore why some online product reviews are lack of usefulness rating mainly by the posting time and review text features related factors. We identify text features of these reviews and assess their probability scores through text mining techniques. And then, through two logistic regression models, we compared differences in these factors between reviews with no usefulness rating and those with useful or useless rating. According to these comparisons, we tried to disclose why some reviews have no usefulness rating. Beside of posting time, our study finds that the diagnosticity of these reviews reflected from their text features is also a non-neglectable cause. Another important finding is that the reviews with higher positive orientation are less likely to be rated for usefulness, probably due to the weaker perceived diagnosticity than negative ones. In addition, compared with useful and useless reviews, no usefulness rating reviews have lower subjective rating deviation of full text. This text feature may well demonstrate the neutral diagnosticity of no usefulness rating reviews, which makes it hard for consumers to evaluate the usefulness. Besides, less subjective rating orientation of full text, less average number of words per sentence, less positive deviation of full text, higher subjective expression orientation, less subjective expression deviation of full text and subjective rating orientation of title also partly explain why reviews with no usefulness rating are not rated to be useful.

From the theoretical aspect, this study adds to the existing literature as follows: Based on the three stages of information processing procedure, see->decide to read->decide to use, we attempt to analyze why some online product reviews have no usefulness rating. Through combining econometric analysis and text mining techniques, we explore the path of improving potential no usefulness rating reviews with neutral diagnosticity into useful reviews and what text features account more for some reviews being not rated for usefulness. Solving of these questions in our study may help extend the existing theoretical findings on online reviews usefulness rating behaviors. Besides, we attempt to study review readers' complex behaviors through online reviews text features using text mining techniques, which may offer new methodology for other similar consumer online behavior researches.

On the other hand, this study has several managerial implications: First, to improve the efficiency of consumers information retrieving and decision, the online reviews website manager can timely eliminate the reviews probably with no usefulness rating according to their unique text features by text mining analysis, including less subjective rating orientation and deviation of full text, less average number of words per sentence, less positive deviation of full text, higher subjective expression orientation, less subjective expression deviation of full text and subjective rating orientation of title, etc. Second, identifying differences of text features between three categories of reviews can help reviewers to improve their review quality. On one hand, our findings suggest reviewers may enhance the subjective rating deviation of review full text when posting reviews, which can increase the diagnosticity of reviews. Furthermore, to turn a review probably with no usefulness rating into a useful one, reviewers need to further delicate on improving subjective rating orientation and deviation of full text and average number of words per sentence, as well as enhancing positive deviation of full text, subjective expression deviation of full text, objective expression orientation of full text and subjective rating orientation of title.

This research has several limitations: The classification correctness ratios for both binary logistic regressions are not improved greatly after introducing review posting time and text features mentioned above. Thus, other important factors properly explaining why some online reviews have no usefulness rating may not have been included in our study model. This study only use the online movie reviews data for empirical analysis. It is yet to be justified whether our findings are also reasonable for online reviews of other products.

REFERENCES

Ahluwalia, R., R. E. Burnkrant, and H. R. Unnava. (2000). "Consumer Response to Negative Publicity: The Moderating Role of Commitment." *Journal of Marketing Research,* 37: 203-214.

Dellarocas, C. (2003). "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms" *Management Science*, 49(10): 1407-1424.

Duhan, D. F., S. D. Johnson, J. B. Wilcox, and G. D. Harrell. (1997). "Influences on consumer use of word-of-mouth recommendation sources." *Journal of the Academy of Marketing Science*, 25:283-295.

Ghose, A., and P. G. Ipeirotis. (2007). "Designing novel review ranking systems: predicting the usefulness and impact of reviews." *Proceedings of the ninth international conference on Electronic commerce*, pp. 303-310.

Ghose, A., and P. G. Ipeirotis. (2004). "Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality." *Proceedings of the International Converence on Decision Support System*.

Ghose, A. and P. G. Ipeirotis. (2008). "Estimating the Socio-Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics." *Working Paper*.SSRN: http://ssrn.com/abstract=1261751.

Pang, B., and L. Lee. (2004). "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts." *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistic*, pp. 271-278.

Sen, Shahana, and Dawn Lerman. (2007). "Why are you telling me this? An examination into negative consumer reviews on the Web." *Journal of Interactive Marketing*, 21: 76-94.

Skowronski, J.J., and D.E. Carlston. (1989). "Negativity and extremity biases in impression formation: a review of explanations." *Psychological bulletin*, 105: 131-142.