Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 154 (2019) 806–816

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology, ICICT 2019

The Impact of Topic Factors on Commodity Reviews

Yong He Lu[a,*], Wei Ting Zhang[a]

[a] *School of Information Management, Sun Yat-sen University, Guangzhou 510006, China*

Abstract

This paper predicted the usefulness and polarity of commodity reviews for clothing, shoes, and Jewelry accessories using neural network method, of which the input layer was the topic–document matrix abstracted from the corpus with LDA. After that, the parameters of neural network trained by the topic–document matrix were obtained to further calculate the relation between each topic and the predictions. The experiments show that there are mainly 19 topics in the reviews for clothing, shoes, and Jewelry accessories. However, the predictions for usefulness and polarity of reviews didn't achieve good performance, since the input layer didn't consider other factors that influence the usefulness and polarity except for topic factors. The topics that most influence the usefulness and polarity were selected respectively through the relationship strength calculated by the parameters of neural network. The beauty and design of the dress, the size of clothes, the color, the views of the reviewers and the information provided by the reviewers are the top five topics that influence the usefulness of reviews. At the same time, the family's feelings about the goods, the cost-effectiveness of goods, the expression of emotion, the theme of jewelry accessories and the comfort of clothing are the top five topics that influence the polarity of reviews. The study will help the sellers understand what customers care about when they buy a certain category of products, as well as tell the reviewers how to edit their comments in order to provide more useful information for the buyers and sellers.

Keywords: Commodity reviews; topic; LDA; neural network; relationship strength

* Corresponding author. Tel.: +86-136-1021-9952.
E-mail address: luyonghe@mail.sysu.edu.cn

10.1016/j.procs.2019.06.079

1. Introduction

As e-commerce is developing rapidly nowadays, people gradually tend to purchase their requirements through Internet. From big equipment such as furniture and electrical appliances to the small goods such as clothing and food, even the immaterial products such as online programmes and game softwares can be easily accessed with network payment, which greatly saves the time and energy of customers. Although our material and spiritual life are greatly improved, there are also some hidden problems in e-commerce which prevent the environment of e-commerce from developing actively and rapidly. For example, product quality problems and service issues during payment and transport. These problems not only damage the interests of the customer, but also do bad to the business reputation and credibility.

An effective solution to these problems is to strengthen the interaction between the sellers and buyers as well as between the buyers. The solution which inspired by Web 2.0 didn't focus too much on increasing more direct profit, since comparing with Web 1.0 that led to the popularization of information technology and Web contents, Web 2.0 enhanced the function of network as intermedium and society to create more personalized information[1].

But it doesn't mean that e-commerce reviews won't generate any benefits for e-commerce sellers. Actually, the abundant customer comments break people's psychological barrier of choosing online shopping by providing decision information for potential buyers to reference as well as by performing effective supervision and feedback on the seller's products and services. The advantages above enhanced the reliability of shopping, which also indirectly brought a considerable profit to the whole e-commerce industry. Although the rise of commodity reviews inevitably lead to some unfair competition such as employing 'Internet water army' to influence the rating of products, positive aspects that strong interaction brings for the e-commerce are still indelible.

At present, all e-commerce platforms at home and abroad have provided the customer with comment function, among which the most complete platform for academic analysis is the amazon in the United States. Compared with the review mechanism of domestic e-commerce platform, the advantages of amazon are listed as follow: ① The ranking and evaluation mechanism of customer reviews. It is possible to evaluate the content usefulness of a reviewer according to all of his or her comments and rank him or her among all the reviewers. Reviewers will feel motivated and supervised to a certain degree through their ranking. ② The useful vote. Customers can choose whether a comment on the page they read is useful, which domestic platforms have also followed suit but customers rarely press the useful buttons when they read the comments. ③ The organization of the comments. Each comment can be traced back to the product and the reviewer, and even to all the products that the reviewer has commented on. These advantages make it possible to use text processing techniques to analyze the impact factors of comments usefulness.

There has been many researches that analyzed the impact factors of comments usefulness. These papers, however, focused more on non-text or topic factors such as length and polarity. Moreover, scholars preferred to regression equation method rather than text mining algorithm. Therefore, this article mainly explored the relationship between the topic of comments and the usefulness and polarity of comments. Firstly, LDA (Latent Dirichlet Allocation) was adopted to extract topic information from commodity reviews corpus. Then, the document - topic matrix achieved from LDA was used as the input of neural network to predict the comments usefulness and polarity. Finally, the parameters of the neural network were extracted to calculate the relationship values between the inputs and outputs in order to discover the topic factors that influence comments usefulness and polarity.

In this Paper, section 1 introduces the background of this research. Section 2 refers to the related works the paper may involve. Section 3 describes the process of experiments to evaluate the impact of topic factors on commodity reviews. Section 4 shows the results of the experiments and gives a detailed explanation about the similarities and differences between this work and the previous research. Finally, section 5 is the conclusion of this work.

2. Related Works

2.1. Factors that influence the usefulness of comments

According to some existed researches[2-4], the characteristics of commodity reviews include three aspects ,quantity, valence (or polarity) and quality. These characteristics will have positive or negative impacts on the sales of products and the usefulness of comments.

With respect to the quantity, many scholars found that the number of comments has a positive impact on product sales[2,3,5]. Among them, Liu not only discovered the relationship between the quantity of comments and product sales, but also proposed that the valence of comments doesn't show strong correlation with product sales as the quantity of comments do. Ludwig et al.'s further study[6] on the valence of comments concluded that the negative comments will be more valuable than the positive comments to those customers who are risk averters. This conclusion was drawn based on the comparison of conversion rate of the positive and negative comment to the product sales. The quality of comments is a more abstract concept, which mainly represents the reliability of the comments. The measurement of quality is not fixed as the actual situation always varies. Common measurement indicators are the credibility of reviewers, the length of comments and the readability of the comments[7].

The reason that the commodity reviews affect the product sales because they provide information about the purchasing decisions for the potential customers. In addition to the macro impact of the number of comments, the content of each comment will also have an impact on the buyer's decision. The function of voting the usefulness of each comment on the comments page makes it possible to quantify the impact of the comments on buyers, which makes it possible for the scholars to measure the impact of various characteristics of the commodity reviews on the usefulness of the comments.

A number of studies have drawn significant conclusions about the impact of various comment characteristics on the usefulness of comments. For instance, the larger the comment length is, the more useful the comment is [8,9]; Comments with extreme star rating tend to be more useful[10,11]; The usefulness of comments is also influenced by non-content items such as product types [11], the characteristics of reviewers[12], etc. None of the studies above explored the relationship between the topics and the usefulness of the comments, though it is obvious that the topic information contained in the comments has a significant impact on the buyer's decision.

Due to the complexity of text mining technology, few studies adopted text mining to research the relationship between the topics and the usefulness of comments. Although Cao, Duan and Gan used LSA (Latent Semantic Analysis) to extract the topic factors contained in the comments and analyzed the impact of these topics on the usefulness of comments[13], they didn't explain the meaning of each factor. On this basis Ahmad and Laroche also employed LSA, but they further explained the various topics that influence the usefulness of comments[14]. What's more, the research considered the positive and negative comments separately according to the star rating, which resulted in different conclusions for different valences. The study has shown that the topics that increase the usefulness of positive comments are commodity function and technical parameters, while the topics that increase the usefulness of negative comments are order and transport services. The commodity type that the study has selected, however, may affect the results of the analysis, thus reducing the universality of the conclusion. Therefore, this paper chose another commodity type to analyze the impact of topics so that we can perform comparisons and supplements to the existed studies.

2.2. LDA and neural networks

LDA is a kind of topic generation model for documents, which, like LSA, belongs to topic model. Both of them can decompose the document - word matrix to the document – topic matrix and topic – word matrix in order to achieve dimension reduction. Other from LSA which achieves dimension reduction by ignoring unimportant information through the calculation of matrix, LDA gets the document – topic matrix and topic – word matrix directly by training the corpus using bayes estimation and sampling method according to the topic number set in advance. Though LSA don't need to set the number of topics in advance, the number of factors to extract still have to be determined artificially based on the singular values after applying SVD (Singular Value Decomposition). If there are too many singular values or these values are very similar with each other, the model will finally ignore too

much information. Meanwhile, LDA method does not need to lose any textual information, but it is difficult to determine the appropriate number of topics.

Instead of applying LSA like existing studies[13,14], this paper chooses LDA to extract the topic information of comments. This is because in the process of SVD it is difficult to determine the number of principle factors to extract, while the topic number of LDA can be tested through experiments and finally determined by comparing the values of log likelihood of different models. Log likelihood and perplexity are both the indicators to evaluate the efficiency of LDA model. When the value of log likelihood is higher and the perplexity is lower, the model is considered to be better.

Neural network is a kind of supervised learning method that simulate the working mode of biological brain neurons. Compared with the traditional regression models, neural networks can often achieve higher accuracy. The neural network does not need to set the form of regression function such as linear and logical regression in advance to represent a certain linear or non-linear relationship. Instead, the model set simply a large number of neurons to express all the complex relationship between the independent variables and dependent variables, which greatly simplify the work. But the disadvantage is that the complex relationship is calculated automatically by the training process of model like a black box. What human can finally obtain after the training process is just a variety of parameters, which are difficult to use to explain the relationship between the independent and dependent variables like the parameters in regression models do.

But Yoon, Guimaraes and Swales[15] have proposed an indicator called strength relationship to approximately evaluate the correlation between independent and dependent variables with the parameters of the neural network which includes a three-layer architecture with a hidden layer. (1) shows the calculation formula of the strength relationship:

$$RS_{ji} = \frac{\sum_{k=0}^{n}(w_{ki} \times w_{jk})}{\sum_{i=0}^{m}\sum_{k=0}^{n}|w_{ki} \times w_{jk}|} \quad (1)$$

Where the w_{ki} represents the weight between the ith input unit and the kth unit in the hidden layer; w_{ik} represents the weight between the jth output unit and the kth unit in the hidden layer; RS_{ji} means the strength relationship between the ith input unit (independent variable) and the jth output unit (dependent variable).

According to the existed study about the prediction of the usefulness of comments, neural network model has shown better performance than the multiple linear regression method[7]. But the input variables that the experiment has selected are 20 factors related to the commodity features, the characteristics of reviewers and the depth of the comment, rather than the topic factors in the text of the comments. Therefore, this paper also adopts neural network method and mainly focuses on the relationship between the topic factors and the usefulness as well as the polarity of the comments.

3. Methodology

Fig.1 is the block diagram of the method and the experiment processes. A detailed explanation of the methodology is described in the following parts.

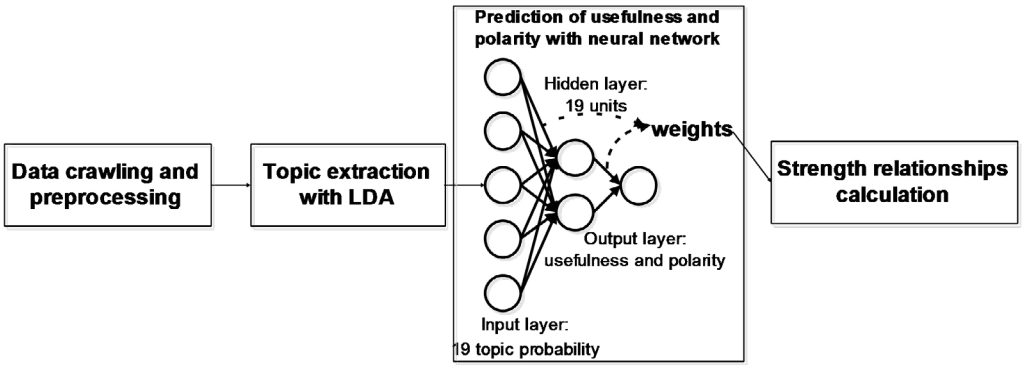

Fig. 1. The block diagram of the process of experiments.

3.1. Data source and preprocessing

In this paper, the original corpus is the commodity reviews collected from the American amazon website. The crawler is written by Python with the help of selenium package. The crawling strategy is to collect all the comments about the _Clothing, Shoes & Jewelry products, which are posted by the top 2000 reviewers in the ranking list of reviewers on the website. Conducted in July, 2017, the crawling task finally obtained 23220 comments posted by 1249 reviewers.

During the process of preprocessing, the content of the comments needs to be cleaned. At the same time, the useful score and polarity score of the comments must be discretized before using as the dependent variables.

The cleaning of data was completed by Python programs using nltk and wordnet package. At last, 23212 comments remained to enter the following analysis. Details of cleaning are as follows:

Using the relevant corpus of the NLTK package to remove the punctuations and numbers.

Using the stop word list of the NLTK package to remove the stop works in the comments.

Using the morphy function in the wordnet package to extract the stems of words in the comments .

The usefulness score is the value of dividing the number of useful votes by the total number of votes. Setting 0.6 as the boundary, comments whose usefulness score are greater than or equal to 0.6 are seen as highly useful comments. Meanwhile, comments whose polarity score are greater than 3 stars are the positive comments.

3.2. Topic extraction based on LDA

Before using LDA to generate the document – topic matrix and topic – word matrix, word - document matrix of the comments corpus needs to be prepared, which is completed by the text feature extraction module of sklearn package in python. After that, the word - document matrix was input into LDA package to generate LDA model. The LDA package in Python allows the model to achieve converge by setting the number of topics and the maximum number of iterations. Each time a iteration is completed, there will be a log likelihood value of the model. When the model converges, the value of log likelihood will be stable. The maximum number of iterations is set to 1000. We found that all the experiments have achieved convergence within 1000 iterations. However, the determination of the topic number still needs to be tested through multiple experiments. In order to reduce the difficulty of the explanation of the topics, the number of topics should not be too large. At the same time, the number of topics should not be too small for the sake of ensuring the differentiation of the topics. Finally, the range of the topic number is set to 5-20, and the optimal number of topics is judged by comparing the log likelihood values of the LDA models of different topic number.

Fig. 2 is the line chart of the log likelihood value of the LDA model when the topic number ranges from 5 to 20. According to Fig. 2, if we require the number of the topics to be as large as possible, the indicator of log likelihood will achieve maximum value when the topic number is 19. Therefore, 19 is selected as the optimal topic number for LDA model, which is also the number of units in the input layer of the neural network.

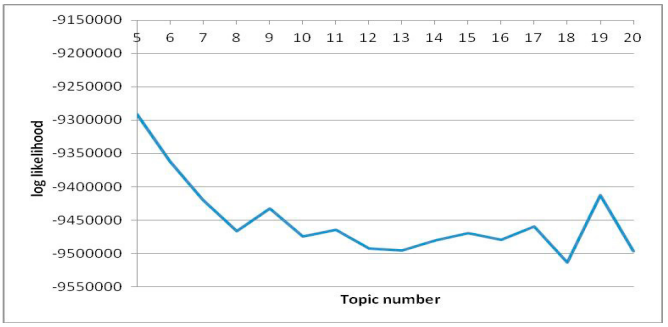

Figure. 2. The log likelihood value of the LDA model when the topic number is 5-20.

3.3. The prediction of usefulness and polarity based on neural network

In the document-topic matrix obtained by LDA, each comment contains 19 topics and the corresponding topic probability. The probabilities of the 19 topics were used as the input variables of neural network which is a three-layer architecture with one hidden layer. Similar to the input layer, the number of units in the second and the hidden layer is also set to 19. The output layer contains only one unit, which represents either the usefulness or polarity of the comments.

The MLP Classifier module in the sklearn package was applied to perform neural network calculations. In the process of model evaluation, taking f1-score as the evaluation indicator and using k fold cross-validation to eliminate the accidental cases and evaluate the average performance of the classifiers.

The ultimate goal of this paper is to find all the topics that related to the usefulness and polarity of the comments. Therefore, after the training of neural network as well as obtaining the connecting weights between the neighbor layers, the strength relationships between the 19 topics and the usefulness as well as the polarity of comments were calculated respectively using (1).

4. Results and Discussion

4.1. Explanation of the comment topics

The meaning of each topic in LDA model can be explained by the topic - word matrix. When the topic number of LDA is set to 19, the top 30 words of each topic and the corresponding explanation are shown in Table 1:

In this section, we will explain the general meanings of the topics in terms of the top 30 words of each topic in Table 1.

It is obvious that topic 1, 4, 7, 8, 10, 11, and 12 are related to the material, function, and design of a particular category of products. Topic 1 includes materials, gems and packaging of jewelry accessories; Topic 4 mainly talks about the waterproof and thermal function of the underwear, jackets and raincoats; Topic 7 is related to the fabric of the dress as well as the design of the breast and waist position. Topic 8 is related to the design of collars and sleeves of the shirts and sweaters; Topic10 and 11 is related to the material and design of shoes and bags respectively; Topic 12 relates to the function and design of the watches.

Topic 0, 2, 15 and 18 have nothing to do with a specific kind of product, but reflect certain characteristics of the products. Topic 0 mainly expresses the reviewers' feelings about the brands and styles of the products; Topic 2 is the description of color; Topic 15 is the comments and suggestions of the reviewers on the size of clothes; Topic 18 is related to wearing comfort.

Topic 5 and 6 are similar to each other in meaning. The words in topic 5 are mostly related to the payment, packaging and transporting services. Although topic 6 also contains words related to transporting, it includes more words about price and quality. So topic 6 represents cost performance of the products.

The contents expressed in topics 3, 9, 13 and 14 have nothing to do with products. Topic 3 contains many adjectives, which are the subjective emotion expression of the reviewers; Topic 9 contains a lot of possible synonyms of hope, seem, think and expect, which express the anticipation and opinions of the reviewers on the products; Topic 13 contains two kinds of speech. One kind expresses one's different feelings about some products from the other reviewers. The other one is to provide information for the other buyers; Topic 14 appears a lot of words about families. Reviewers usually express their feelings about the products worn by their families, or whether their families like the products they receive.

It is difficult to judge what topic 16 and 17 talk about, since the words in these two topics rarely express specific meanings. But topic 16 contains some words related to time, while topic 17 contains some words related to quantity.

Table 1. The top 30 words of 19 topics.

No.	The top 30 words	Explanation
0	like brand fit wear make shorts look great good one different buy well size get pair much line really also feel go tag say even style love underwear want bit	Feelings about the brand and style.
1	beautiful gift necklace ring piece stone box jewelry bracelet chain silver gold pendant face lovely gorgeous elegant diamond clear glasses clasp charm crystal sparkle sunglasses shiny come metal set finish	Jewelry accessories.
2	color colors belt blue bra black hat picture tie red dark pink head bright white outfit match girl different scarf adorable brown baby buckle green adjustable child hair gray photo	Color.
3	nice really well make look design nicely style fun add different together amaze seem detail fan unique five yet wonderful chart close cool expensive super cat chic touch giving classic	Emotion expression.
4	pants warm keep jacket like make well cold winter lightweight weight water style heavy coat weather thick men cool hot glove fall layer robe live pajama rain inside collar days	The function of keeping warm.
5	say cant know buying sure people tell wife let arrive real saw money ask must huge seller spend deal like far nothing course call day package mean friend idea bad	Transport and payment services.
6	great good quality price look recommend looking highly cheap excellent definitely worth costume expect high exactly fast shipping absolutely surprise awesome pay decent value durable impress deal party arrive halloween	The feeling of cost performance of the products.
7	dress top waist length stretch look fabric skirt show body tall medium flatter hip sexy suit area bust inch material shape cut beautiful style back skin chest knee sew short	The fabric and design of each part of the dress.
8	shirt soft material long wash like top fabric sleeve loose short dry cotton medium button legging length thin sweater hang casual washing print neck low black shrink colors tank summer	The fabric and design of each part of the shirts.
9	like bit pretty think much seem though little want expect quite might better actually hope wish way thin sure maybe id thought almost slightly end side ill bigger kind probably	The expectations and views of the reviewers.
10	shoes foot boot pair shoe wide heel toe support sandal walking walk leather ankle narrow sole arch slipper running comfort width lace good slip half looking sneaker flip try strap	The material and design of each part of the shoes.
11	bag pocket zipper strap leather inside wallet hold use side carry open purse front small pack shoulder keep two stuff room enough handle cards travel compartment backpack phone item case	The fabric and function design of wallets and bags.
12	watch earring time band light set easy looking want look much back case wrist keep give however simple hand second black may works easily come remove way quite show cut	The function and design of watches.
13	review purchase product receive amazon see please give way star always happy help item know provide found picture free return photo honest opinion helpful disappoint update however full base yes	Stressing they have bought products according to others' comments, or providing comments to help the others.
14	love buy perfect little cute super big old daughter want year husband happy month son also kid christmas still glad already fall school compliments worry play perfectly excite wait age	Referring to families' opinion on the products or stressing the products were purchased for their families.
15	size fit order small large tight run jean perfectly larger sizing true smaller woman normally return xl extra bottom snug us ordering skinny usually baggy however suggest denim half least	Comments and suggestions on the size of clothing.
16	go even still first back get time try last put another take years going thing best right	Something about time.

	never issue problem wont able couple times week away hole several place notice	
17	one get use need also come lot many two find see better favorite easy new hard take store three things either getting hold reason easily making start stand especially clean	Something about quantity.
18	wear comfortable feel wearing around work sock enough day right thick pair fine comfy usually every high extremely bottom house stylish stay need summer true find often comfortably problem dressy	Wear comfort.

4.2. Prediction results of comment usefulness and polarity

Table 2 shows the evaluation of the prediction of the comments usefulness and polarity. 20-fold cross-validation and the indicator of f1-score is adopted to evaluate the average performance of two models.

Table 2. The performance of the two models which predict the comment usefulness and polarity.

No.	f1-score(usefulness)	f1-score(polarity)
1	0.61	0.78
2	0.61	0.77
3	0.62	0.8
4	0.62	0.8
5	0.6	0.81
6	0.62	0.82
7	0.65	0.82
8	0.6	0.8
9	0.61	0.79
10	0.6	0.82
11	0.62	0.82
12	0.59	0.79
13	0.59	0.82
14	0.63	0.8
15	0.63	0.81
16	0.6	0.79
17	0.61	0.79
18	0.59	0.78
19	0.6	0.79
20	0.6	0.81

As seen from Table 2, it is difficult to predict the usefulness and polarity of comments with topic factors, since the f1-score of the two models are less than 0.65 and 0.82 respectively. This may be because topic is not the only factor that determines the usefulness and polarity of comments, other factors not considered in the paper such as the depth of comments and the characteristics of products would also influence the usefulness and polarity. For higher performance of prediction, it's better to fully consider both topic and non-topic factors in the following studies.

4.3. Strength relationship between the topics and the comments usefulness and polarity

Although the performance of neural network is not good enough, we can still calculate the strength relationship using parameters of the model to discover the contribution of different topics to the dependent variables.

As Table 3 shows, if the topics are ranked according to their strength relationship with the dependent variables, then the topics most related to the usefulness of comments are: topic 7, 15, 13, 2 and 9. The reason that the topic about dress is most related to the usefulness may be because the customers who enjoy buying dresses especially

attach great importance to the beauty and design, thus the comments are often well edited and become important information for other girls to reference; Meanwhile, the size of the clothes must be carefully verified, since the buyers will face the return of products if the clothes are not suitable for them, which is a troublesome business; In addition, the comments about the colors also have influence on the usefulness, probably because chromatic difference may be existed between the pictures on the website and the real products; The reviewers' opinions of the products and the information provided actively are also important for others.

The topics that are the most related to the polarity of comments are: topic 14, 6, 3, 1 and 18, which indicates that the main factors that influence the reviewers' satisfaction with the products they have bought include the family's feelings on products, the cost performance, the expression of emotion, jewelry accessories and the comfort level of clothing.

Compared with the study of Lee and Choeh[7], the results are slightly different in this paper. Lee and Choeh proposed that the positive comments about technical aspects, core function and aesthetic mainly influence the usefulness, while this paper concluded that the design of dress, the size of clothes and the colors are associated with comments usefulness. In addition, this paper also believes that the information provided by the reviewers is important, while the comparing study didn't consider this aspect. Finally, Lee and Choeh believes that service in the negative comments is also the main factor of usefulness, while the study of this paper found that service is not the top factors that influence the usefulness of comments.

Table 3. Strength relationship (RS) between the topics and the comments usefulness and polarity.

Topic class	Topic No. (explanation)	RS (usefulness)	RS (polarity)
Material, function, and design of a particular category of products.	1 (Jewelry accessories.)	-0.01699749	**0.018141**
	4 (The function of keeping warm.)	-0.00340885	0.013067
	7 (The fabric and design of each part of the dress.)	**0.02580175**	0.006789
	8 (The fabric and design of each part of the shirts.)	-0.00362676	0.001918
	10 (The material and design of each part of the shoes.)	-0.02197463	-0.01583
	11 (The fabric and function design of wallets and bags.)	-0.00154039	-0.00672
	12 (The function and design of watches.)	-0.00416032	-0.00535
Having nothing to do with a specific kind of product, but reflecting certain characteristics of the products.	0 (Feelings about the brand and style.)	0.00157286	0.00709
	2 (Color.)	**0.00852707**	0.015055
	15 (Comments and suggestions on the size of clothing.)	**0.01162244**	-0.00033
	18 (Wear comfort.)	0.00687508	**0.017607**
Payment, packaging, transporting services and cost performance.	5 (Transport and payment services.)	0.00761404	-0.01959
	6 (The feeling of cost performance of the products.)	-0.0046399	**0.02282**
Emotions and opinions of selves and others	3 (Emotion expression.)	-0.00258311	**0.021947**
	9 (The expectations and views of the reviewers.)	**0.00833111**	-0.02632
	13 (Stressing they have bought products according to others' comments, or providing comments to help the others.)	**0.0091873**	-0.01774
	14 (Referring to families'	0.00378097	**0.033359**

	opinion on the products or stressing the products were purchased for their families.)		
Related to time and quantity	16 (Something about time.)	-0.00153423	-0.00329
	17 (Something about quantity.)	0.00448099	0.008822

The different categories of products selected by Lee and Choeh's study and this paper probably caused the divergence between the two studies. Lee and Choeh chose the comments of kitchen appliances as corpus. As the products always have large volume and high price, the customers need to consider about the warranty issues and pay more attention to the transporting and payment services. However, the products of clothing, shoes and jewelry accessories selected by this paper are always small and not expensive most of the time. So the customers usually don't need to worry about the transportation damage and warranty. At the same time, the size of clothing and shoes is the most important feature for this kind of products, just as the technical aspects for kitchen appliances. So it is reasonable for clothes size to have a great influence on the usefulness of comments.

5. Conclusion

This paper employed LDA to extract the topic message from the corpus of commodity reviews. 19 topics were explained and input into neural networks in order to predict the usefulness and polarity of comments. Finally the parameters of neural networks were used to calculate the correlation between the topics and the usefulness and polarity of comments. In this process, five topics that mostly influence the comments usefulness and polarity respectively were generated. The study will help the sellers understand what customers care about when they buy a certain category of products, as well as tell the reviewers how to edit their comments in order to provide more useful information for the buyers and sellers.

When analyzing the impact factors of comments usefulness, this paper chose a new category of products different from the existed study. Through comparison of two studies, some relations and distinctions are found, which enrich the existing conclusion. However, the models of neural networks for the predictions of comments usefulness and polarity didn't achieve good performance, which may be because the topic factor is not the only factor affecting the usefulness and polarity of the comments. At present many successful studies about comments usefulness consider the non-topic factors as main objects to research, which implies that the future study may focus on how to combine the topic factors with the non-topic factors to advance the performance of classifiers. furthermore, there are still not enough studies about the analysis of comments usefulness that applies text mining methods. Therefore more effective methods need to be find in the future to reveal the correlation between the topic and usefulness of comments.

Acknowledgement

This work was supported by Science and Technology Planning Project of Guangdong Province, China (No. 2015A030401037) .

References

1. Jun CN, Chung CJ. Big data analysis of local government 3.0: Focusing on Gyeongsangbuk-do in Kore. *Technological Forecasting & Social Change*; 2016, 110:3-12.
2. Chevalier JA, Mayzlin D. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*; 2003,43(3): 345-354.
3. Goldsmith RE, Horowitz D. Measuring Motivations for Online Opinion Seeking. *Journal of Interactive Advertising*; 2006, 6(2):2-14.
4. Tsao WC, Hsieh MT. eWOM persuasiveness: do eWOM platforms and product type matter?. *Electronic Commerce Research*; 2015, 15(4):1-33.
5. Liu Y. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*; 2006, 70(3):74-89.
6. Ludwig, Stephan, de Ruyter, et al. More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*; 2013, 77(1): 87-103.

7. Lee S, Choeh JY. Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*; 2014, 41(6):3041-3046.
8. Kim SM, Pantel P, Chklovski T, et al. Automatically assessing review helpfulness. *Conference on Empirical Methods in Natural Language Processing*; 2006:423-430.
9. Huang AH, Chen K, Yen DC, et al. A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*; 2015, 48(C):17-27.
10. Mudambi SM, Schuff D. What makes a helpful online review? a study of customer reviews on amazon.com. *Society for Information Management and The Management Information Systems Research Center*; 2010.
11. Pan Y, Zhang JQ. Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*; 2011, 87(4):598-612.
12. Baek H, Ahn J, Choi Y. Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*; 2012, 17(2):99-126.
13. Cao Q, Duan W, Gan Q. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*; 2011, 50(2):511-521.
14. Ahmad SN, Laroche M. Analyzing electronic word of mouth: A social commerce construct. *International Journal of Information Management*; 2017, 37(3):202-213.
15. Yoon Y, Guimaraes T, Swales G. Integrating artificial neural networks with rule-based expert systems. *Elsevier Science Publishers B. V.*; 1994.