Summer 7-26-2022

Information Cascade and Spam E-commerce reviews filtering

Haoran Zhang
School of Inforamtion, Central University of Finance and Economics, China, haoran_hr_zhang@126.com

Yin Chen
School of Applied Economics, Renmin University of China, China

Short Research Paper

Information Cascade and Spam E-commerce reviews filtering

Haoran Zhang[1*], Yin *Chen*[2]

[1]School of Inforamtion, Central University of Finance and Economics, China

[2]School of Applied Economics, Renmin University of China, China

Abstract: Rating and reviews provided by customers can send many signals to the market, especially in online marketplace. How to formulate an outstanding sales strategy and show users the most useful information of the product is particularly important. In this paper, we study the review information cascade and build a MLP model to distinguish the spam and informative reviews from the noise reviews. The information cascade phenomenon in E-commerce reviews is identified by Multinomial Logistic Regression. After seeing a series of one/five star ratings, consumers are more likely to write low/high score reviews and the effect caused by low star rating is more significant than that of high star rating. Therefore, we use the helprate of the review as the label to distinguish the informative reviews and spam reviews. The one-hot review vectors is reduce to 136 principal components components as the input variables to train the MLP Model. The accuracy of the model on the test set is 78.5%, and AUC is 0.713. The E-commerce companies can evaluate the informative review, improve online sales strategies, enhance the product desirability and identify customers' preferences through the model proposed in this paper.

Keywords: E-commerce reviews, Information Cascade, MLP, spam detection

1. INTRODUCTION

Rating and reviews provided by customers can send many signals to the market. In E-commerce, the rating and reviews system consists of three components: individual ratings called "star ratings", text-based messages called "reviews" and ratings on these reviews as being helpful or not called "helpfulness rating". In the online marketplace, all the rating and reviews are available to the public to help customers have a clear image about the product, which may influence the product's reputation to a large extent. On the other hand, companies can understand consumers' demands better and adapt their design features and sales strategy to enhance product desirability. Therefore, understanding and analyzing the patterns of rating and reviews has high research value for both purchaser and producer. However, information cascade phenomenon in E-commerce reviews will influence the above comment system. Information cascade is also called the bandwagon effect, which refers to the phenomenon that readers are more likely to give an unfair review after reading an extreme star rating. For example, when "One-star" rating appears, the following users's probability to give an extremely negative emotion will significantly increase.

In this paper, we will establish an excellent e-commerce review identification model based on open source dataset which contains the ratings and reviews for some products sold in Amazon marketplace during 2002-2015. We make the following assumptions to simplify the problem and ensure its rationality: 1) Whether the review is informative has nothing to do with the product category; 2) We do not take some special logic (like double negation) of language into account; 3) The sample distribution of each industry is the same as the overall distribution; 4) A product's reputation consists of customers' attention and praise (high-level star rating rate). We use the helprate of the review as the label to distinguish the informative reviews and spam reviews. The review vectors is reduce to 136 principal components. The principal components and the text features selected by LASSO is the input variables to train the MLP Model. The accuracy of the model on the test set is 78.5%, and

* Corresponding author. Email: haoran_hr_zhang@126.com

AUC is 0.713.

To sum up, our contributions are threefold:

- We identified the information cascade effect in e-commerce comments through dynamic multinomial logistic regression, which proves that users will be affected when reading the comments on e-commerce platforms and causes bias.
- This paper propose a spam comment detection model combines text features and semantic information, which can effectively filter spam comments in the real world e-commerce scenario.
- Our approach can help the E-commerce companies to evaluate the informative review, improve online sales strategies, enhance the product desirability and identify customers' preferences through the model proposed in this paper.

2. RELATED WORKS

Spam comments mainly spread on three platforms: blog, social software and e-commerce system. In this paper, we mainly focus on the junk comments on e-commerce platform, which is also called the spam review.The kind of spam we want to detect in this paper is usually referred to as in e-commerce platform. Our study purpose is to detect the commodity comments on e-commerce platform are spam comments or not. Spam reviews are highly sequential related and different from the other spams categories from various perspectives. One of the most prominent differences is the fuzziness because it is hard to measure whether the subjective comments given by users are objective or not. However, spam reviews have a great negative impact on both the purchaser and producer. It not only give the consumers unreasonable reference, but also bring bad reputation to a brand.

Spam reviews can be further divided into four aspects: advertising comments, repeated comments, irrelevant comments and fake comments. The first two aspect can be filtered based on simple rules; The latter two need to be recognized in combination with text content and text features which are also the main research aspects of this paper. In the field of e-commerce, identifying spam comments often starts from two aspect: text features and reviewer characteristics [1]. The spam review detection model on e-commerce platforms often use a two-class classifier from two aspects[2]: the user's background features, and the features extracted from the comment content and the data labels are labeled manually. One of the earliest researches on spam review is Jindal and Liu[1], in which they attempt to detect fake product reviews on Amazon. Since then, this topic has been brought into highlight, and there are many following researches like Lim et al.[3] and Xie et al.[4].

However, the above method has great drawbacks, because the manually constructed features have bias and require a large number of manual rules. In addition, these manual rules need to update frequently from the candidate feature set. In the e-commerce spam reviews detect tasks, even annotators can't tell whether a comment is spam or not, which will cause the unfairness of the labels [5]. Myle et al. reported a very low annotator agreement score when annotating the spams from a review corpus[6]. In contrast, most of the email spams, web spams, or social network spams are fairly easy to spot by an experienced user. At the same time, manual annotation consumes a large labor cost. To address the above problems, this paper uses automatic feature extraction method LASSO and the review's help rate from the users as label to filter spam comments, in order to make the model more practical and unbiased.

3. INFORMATION CASCADE IN E-COMMERCE REVIEWS

In this section, we identify information cascade in e-commerce reviews. The result find that the influence of the existing ratings and reviews to potential consumers is significant. It is important to discuss how the existing ratings and reviews influence the attitude of potential consumers who want to write reviews. Given that

reviews are strongly associated with rating levels, we use star ratings as a proxy variable. This problem can be solved by performing a Dynamic Multinomial Logistic Regression(M-logit). A dynamic penal data is established according to the time-based patterns, using "day" as the unit of time.

We first establish a Utility Function of a potential consumer: $u_{in} = v_{in} + \varepsilon_{in}$, where u_{in} represents the utility of individual i when choosing product n, ε_{in} is the error term, and v_{in} represents the influence of the existing ratings and reviews, which can be further described as: $v_{in} = x_{in}' \cdot \beta$. Let P_{in} denote the probability that the customer can maximize his utility, we have: $P_{in} = P\left(u = \max_k u_{kn}\right)$ For each customer, their objection is to maximize their utility under any circumstances. Suppose the error term $\{\varepsilon_{in}\}$ is i.i.d and it follows Type I extreme value distribution, then we have:

$$P\left(y_i = j \mid x_i\right) = \frac{\exp\left(x_{in}' \cdot \beta_j\right)}{\sum_{k=1}^{J} \exp\left(x_{in}' \cdot \beta_j\right)} \tag{1}$$

Table 1 shows the M-logit regression results. In the current period, we use Three-Star as the benchmark and eliminate the collinearity. It can be seen that there is a bandwagon effect of ratings and reviews. When specific star rating appears, it is likely that some type of reviews may appear. For example, as is shown in Table 1, when the appearance of "One-star" rating increase one unit, there is significant increase in the probability of more reviews with extremely negative emotion (L. One-Star). This bandwagon effect may last for more than one phase. However, as for positive ratings, the bandwagon effect of "Five-star" rating is much weaker if we focus on the reviews in lag 1 phase. What deserves noticing is that in lag 2 phase, the influence of "Five-Star" rating become stronger in margin. Although a "Five-star" rating cannot bring more positive reviews with Five-Star rating, it can significantly reduce the amount of negative ratings and reviews in lag 2 phase.

Information Cascade theory which was originally studied by Lisa and Charles [7-8] can perfectly explain this phenomenon. In certain environments the decisions of individuals tended to reflect only the decisions of those who went before them and did not reflect the information they held privately. May's Theorem also shown that the only group decision function that satisfies decisiveness, anonymity, neutrality and positive responsiveness is simple majority rule. Another research result shows that the number of negative reviews and the quality of the comment review text content have a negative impact on consumer purchasing behavior. More importantly, there is no interaction between the number of negative reviews and the quality of the interactive review text-based content. It will inevitably have a negative impact on consumers' buying behavior. Only when high-quality bad review appears will the number of bad reviews reduce the possibility of consumers buying [9].

Table 1. Dynamic Multinomial Logistic Regression

	One-Star	One-Star	Two-Star	Two-Star	Four-Star	Four-Star	Five-Star	Five-Star
L.One-Star	0.575***	0.570***	0.174**	0.134*	0.0925	0.130	-0.362***	-0.278***
	(4.18)	(4.07)	(2.17)	(1.83)	(0.79)	(1.09)	(-3.43)	(-2.60)
L.Two-Star	0.341**	0.0331	-0.152	-0.0824	-0.1000	-0.0814	-0.336***	-0.277**
	(2.30)	(0.21)	(-0.84)	(-0.45)	(-0.75)	(-0.60)	(-2.89)	(-2.35)
L.Three-Star	-0.213	-0.231	-0.159	-0.0823	0.124	0.136	-0.197*	-0.142
	(-0.13)	(-1.54)	(-0.99)	(-0.50)	(1.08)	(1.17)	(-1.94)	(-1.37)
L.Four-Star	-0.508***	-0.253**	-0.462***	-0.157*	0.237**	0.264***	-0.0840	-0.0411
	(-4.02)	(-2.01)	(-3.24)	(-1.75)	(2.52)	(2.76)	(-0.99)	(-0.47)
L.Five-Star	-0.424***	-0.302***	-0.266**	-0.183	0.137*	0.140*	0.193**	0.111*
	(-3.95)	(-2.77)	(-2.28)	(-1.54)	(1.80)	(1.79)	(1.97)	(1.76)

L2.One-Star		0.276*		0.137*		-0.256**		-0.566***
		(1.71)		(1.66)		(-2.26)		(-5.70)
L2.Two-star		-0.186		-0.276		-0.0954		-0.433***
		(-1.14)		(-1.49)		(-0.73)		(-3.73)
L2.Three-Star		-0.355**		-0.273*		-0.0333		-0.333***
		(-2.41)		(-1.66)		(-0.30)		(-3.35)
L2.Four-Star		-0.160		0.00592		-0.110		-0.288***
		(-1.25)		(0.04)		(-1.23)		(-3.65)
L2.Five-Star		-0.562***		-0.511***		0.0197		0.0722*
		(-5.18)		(-4.15)		(0.27)		(1.73)
_cons	0.154	0.504***	-0.321**	0.0795	0.532***	0.558***	1.962***	2.126***
	(1.31)	(3.25)	(-2.49)	(0.46)	(6.14)	(5.14)	(25.17)	(21.80)

Note: (1) t statistics in parentheses (* p < 0.10, ** p < 0.05, *** p < 0.01); (2) "L." represents lag one, and "L2." represents lag two.

In conclusion for this part, bandwagon effect of ratings and reviews does exist. Furthermore, bandwagon effect of negative ratings and reviews pass faster than that of positive ones. The online e-commerce platform should take care of some ratings and reviews with extreme value closely to adjust their sales strategy and product design features.

4. OUR APPROACH

4.1 DATASET

In this paper, we use the Amazon product data to evaluate our method. This dataset contains product reviews and metadata from Amazon, including 142.8 million reviews spanning May 1996 - July 2014. This dataset includes reviews (ratings, text, helpfulness votes), product metadata (descriptions, category information, price, brand, and image features), and links.

4.2 FEATURE SELECTION

In our paper, the review data is vectorized by the TF-IDF Model, and the high-dimensional vectors for each review are obtained. For the traditional statistics model, the high-dimensional data will cause the overfitting problem and dimension disaster. Therefore, we use the PCA and feature selection method to reduce the input dimension. The classic solution of the feature selection is using regularization. For linear regression models, the classic penalty regression is LASSO Regression (Least Absolute Shrinkage and Selection Operator). However, the Lasso regression is biased. In order to get the unbiased estimation of the feature of words, Belloni et al. proposed the Post Double Lasso Regression[10]. The method is to delete the redundancy variables that are under the significant level in the first regression stage then use the selected variables to carry out the OLS regression.

Table 2. Lasso regression

Positive	Lasso	PD Lasso	Negative	Lasso	PD Lasso
adorable	0.4148102	0.6984165	annoy	-0.7837718	-1.8394925
amazing	0.4173960	1.0487653	awful	-2.0708660	-2.9537236
awesome	0.7473301	1.0796962	bad	-2.0043258	-2.1390828
comfortable	0.0018916	0.6662282	burnt	-0.4057664	-1.7981232
compliment	0.0604489	1.1652974	choke	-0.3809821	-1.6375437
comfortable	0.0018916	0.6662282	dangerous	-2.7818355	-3.0493823
daughter	0.2484637	0.4398958	disappoint	-1.9507684	-2.5260236
enjoy	0.0652949	0.6999416	fake	-0.6825718	-2.2437396

exact	0.1441077	0.6264478	frustrating	-0.5152176	-2.4439440
excellent	0.6508276	0.9310531	heavy	-0.3185593	-0.7726373
fantastic	0.6232899	1.3951805	junk	-3.7279031	-3.8595142
favorite	0.7950107	1.0285492	leak	-0.2242432	-0.8863464
five	1.8256588	3.0494801	loud	-0.6290767	-1.1889072
friend	0.3755859	0.7863640	melt	-0.5488885	-1.5694183
gift	0.3420769	0.4272349	overprice	-0.1376485	-1.6820070
good	0.2112128	0.5556372	plastic	-0.4323263	-0.5954471
grandson	0.0322434	0.4156137	plug	-0.0692597	-0.3577292
happy	0.6948543	1.0246901	red	-0.2101598	-1.1239613
love	1.6620482	1.5169349	sad	-1.1457216	-2.0824643
Mom	0.2568337	0.7824537	smoke	-1.5854177	-2.1438746
nice	0.4253121	0.7073229	spark	-2.5221133	-2.8322112
recommend	0.2184024	0.1403427	terrible	-2.7162331	-3.1586878
satisfied	0.1264550	0.9114049	uncomfortable	-0.2758357	-1.5649763
shower	0.0715228	0.4098420	unsafe	-2.6324747	-4.2564315
wonderful	0.6871538	1.1169403	useless	-3.4313305	-4.0348976

Table 2 show the Lasso regression result of 25 main words. The left column reports the Lasso Estimators and the right column reports the Post Double Lasso Estimators. As is shown in the Table 2, those representative words can be divided in to two parts: positive words and negative words. Positive words are strongly correlated with high helpful rate reviews with high star level ratings, while negative words are correlated with high helpful rate's reviews with low star level ratings. When customer give reviews with negative emotion, they are intended to describe the quality problem they met in details. For example, they use "overprice" to complain the high price when the product is far from their satisfaction.As for positive reviews, it is interesting that words like "daughter", "friend", "grandson" and "mom" appear in the reviews. Customers are likely to recommend the products they like to others. This may increase the number of potential consumers. The selected text feature will concatenate with the tf-idf principal component after PCA as the input of the spam review filtering model.

4.3 EXPERIMENT DESIGN

Ratings and reviews may influence the reputation of products so that consumers may change their desire to buy this product. The problem is how to choose useful information. Since ratings are strongly correlated with the type of reviews[11], we can use both text-based and rating-based measures to choose informative ratings and reviews.

In this paper, we define what is an informative rating and review. In order to identify informative ratings and reviews, we form a classifier based on TF-IDF and PCA. After the data preprocessing, we filter out the first 1000 words according to their importance based on the term frequency using the TF-IDF model. However, the dimensionality of the filtered words is still too high for us to apply to practice. We use PCA to reduce the dimensionality of the filtered words and a small set of unrelated variables were obtained to reflect the content information of the text. The scatter plot based on the eigenvalues of each principal component as shown in Figure 1. We choose 136 PCs with the cumulative contribution rate of 50.3527%.

Whether the information is useful depends on ratings and reviews. In addition to word frequency, the length of the review body and the star rating are also taken into account. Because the ultimate goal of screening useful review information is to understand the customer's demands, we uses the proportion of customers who consider the review to be useful (helprate) to measure the usefulness of the information, with 0.5 as the cut-off

point, divided into useful or useless.

$$helprate = \frac{helpful\_votes}{total\_votes} \tag{2}$$

We take review body, length of review body, and star rating as input variables. Then we use helprate to complete machine supervised learning. The classifier helps us get a binary variable (whether the information is useful) as output variables. We divide the data set into two part: training set and test set with the ratio 8: 2.

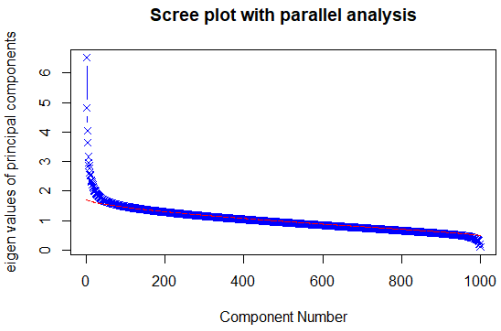

Figure 1. Eigenvalues of the term frequency

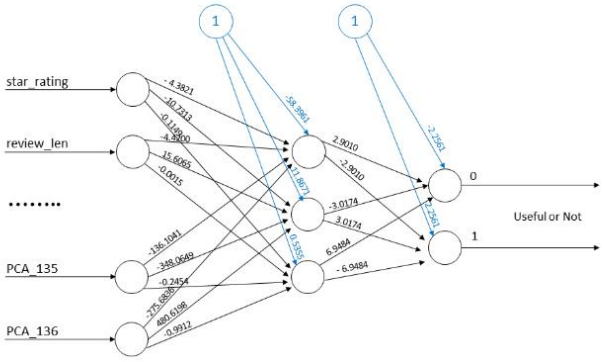

Figure 2. BP-neural network weights

We use BP Neural Network and determine the hidden layers and the number of nodes of the neural network. The activation function we use in this case is Sigmoid function. The Sigmoid function can reflect the gradual transition process of the model from approximately linear to non-linear during the continuous weight correction process well. It is not only non-linear and monotonic, but also infinitely differentiable, which enables BP to use gradient descent to adjust weights continuously. The principle is minimizing the omission of any useful information. Since it is a binary classification task, we assume the value of positive class is 1 and that of negative class is 0. The BP neural network weights are shown in Figure 2.

5. RESULT AND ROBUSTNESS TEST

In this section, we use the left 20% of sample as test set to test the BPNN model. The text-based information is retained if they are the same as that of training set. The PCs of the test set is obtained by multiplying the term frequency of test set by the eigenvalue matrix given by training set. We will compare the results of test set with training set.

We define the confusion matrix as:

Table 3. Confusion Matrix

Confusion Matrix		Predicted value	
		Positive	Negative
Truth Value	Positive	TP	FP
	Negative	FN	TN

Furthermore, Sensitivity and Specificity are defined as:

$$TPR = Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$TNR = Specificity = \frac{TN}{TN + FP} \tag{4}$$

We use a classifier to classify the sample data. The classifier will give the probability of each data being a

positive example. We can set a threshold for this. When the probability that a sample is judged as positive is greater than the threshold, the sample is considered to be a positive example, otherwise it is considered to be negative. Then we can get a (Specificity, Sensitivity) pair, which is a point shown in figure. After continuously adjusting this threshold, we get several points to draw ROC (Receiver Operating Characteristic) curve. AUC (Area Under Curve) of the ROC curve represent the area under it. The ROC curve of BP-neural network is shown in Figure 3.

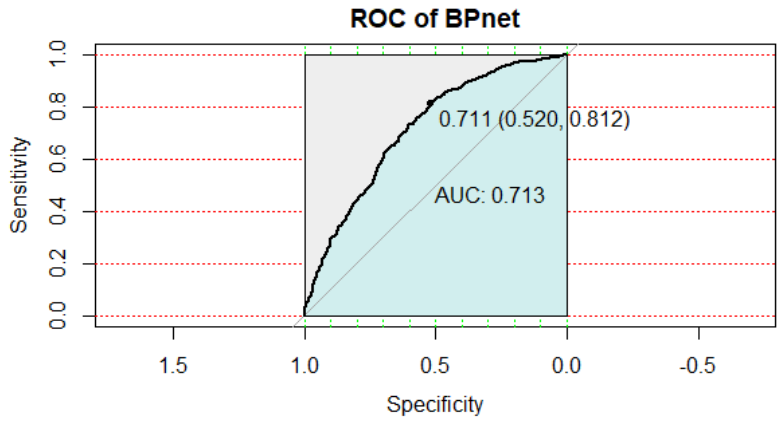

Figure 3. The ROC of BP-neural network

As is shown in Figure 3, the value of AUC equals 0.713, meaning the classifier performs relatively satisfactorily. Given the threshold as 0.5, the confusion matrix of the test set and the training set is shown in Table 4. The accuracy rate of the training set is 81.56%, and the accuracy rate of the test set is 78.53%. Informative ratings and reviews are distinguished well by this classifier.

Table 4. The confusion matrix of the test set and the training set of BPNN

Train Data

		Predicted value	
		0	1
Truth Value	0	591	1395
	1	185	6401

Test Data

		Predicted value	
		0	1
Truth Value	0	122	372
	1	88	1561

6. CONCLUSIONS

With the rapidly development of mobile phones, online e-commerce has become an indispensable part of people's life, and the rating and reviews provided by customers can send many signals to the market, especially in online marketplace. How to formulate an outstanding sales strategy and show users the most useful information of the product is particularly important. In this paper, we study the review information cascade and build a MLP model to distinguish the spam and informative reviews from the noise reviews. The information cascade phenomenon in E-commerce reviews is identified by Multinomial Logistic Regression. After seeing a series of one/five star ratings, consumers are more likely to write low/high score reviews and the effect caused by low star rating is more significant than that of high star rating. Therefore, we use the helprate of the review as the label to distinguish the informative reviews and spam reviews. The accuracy of the model on the test set is 78.5%, and AUC is 0.713. The E-commerce companies can evaluate the informative review, improve online sales strategies, enhance the product desirability and identify customers' preferences through the model proposed in this paper.

REFERENCES

[1] Jindal, N., & Liu, B. (2008, February). Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining (pp. 219-230).

[2] Harris, C. G. (2012, July). Detecting deceptive opinion spam using human computation. In Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.

[3] Lim, E. P., Nguyen, V. A., Jindal, N., Liu, B., & Lauw, H. W. (2010, October). Detecting product review spammers using rating behaviors. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 939-948).

[4] Xie, S., Wang, G., Lin, S., & Yu, P. S. (2012, August). Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 823-831).

[5] Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. Journal of Investigative Psychology and Offender Profiling, 5(1‐2), 39-43.

[6] Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. arXiv preprint arXiv:1107.4557.

[7] Anderson, L. R. , & Holt, C. A. . (2008). Information cascade experiments. Handbook of Experimental Economics Results.

[8] ANGELA, A., HUNG, CHARRLES, R., & PLOTT. (2001). Information cascades: replication and an extension to majority rule and conformity-rewarding institutions. The American economic review.

[9] Lu H.X., Wu X.D., Su L.X., (2014). Is a negative online review really that scary? A research on the effect of negative online customer review on purchase behavior. Social Science of Beijing: 000(005), 102-109.

[10] Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies, 81(2), 608-650.

[11] Mcauley, J. , & Leskovec, J. . (2013). Hidden factors and hidden topics: Understanding rating dimensions with review text. Proceedings of the 7th ACM conference on Recommender systems. ACM.