International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.7 (2015), pp.201-212
http://dx.doi.org/10.14257/ijmue.2015.10.7.21

The Impact of Logistics Factors on Customer Reviews in E-Commerce

Qingtian Zhang[1], Youfang Huang[2], Wei Yan[3] and Yu Wang[4]

[1] Logistics Research Center of Shanghai Maritime University, Shanghai, 201306, P. R. China
[2] Logistics Research Center of Shanghai Maritime University, Container Supply Chain Technology Engineering Research Center of MOE, Shanghai, 201306, P. R. China.
[3] Container Supply Chain Technology Engineering Research Center of MOE, Shanghai Maritime University, Shanghai, 201306, P. R. China,
[4] Container Supply Chain Technology Engineering Research Center of MOE, Shanghai Maritime University, Shanghai, 201306, P. R. China,
15020697923@163.com; yhuang@shmtu.edu.cn; weiyan@shmtu.edu.cn; yuwang@shmtu.edu.cn

Abstract

Logistics is the next step and the server of the business, which has a crucial importance for an intensive research. In this paper, taking the amazon reviews for example, we will segment the logistics factors in customer reviews from multiple perspectives, and use the method of data mining analysis to construct a model to mine the association rules which based on APRIORI algorithm. At last, the reliability was validated by an actual example. Such study can help businesses understand the influence degree of logistics in the comments better and help them reduce bad reviews causing by logistics factors. Finally, businesses could improve their reliability and increase their efficiency.

Key words: *e-commerce, user comments, Logistics factors, correlation analysis*

1. Introduction

In recent years, as online shopping become a more and more frequent trading, people pays increasing attention to the online consumption comments, which is the main source of information [1]. For the importance of customer reviews as well as the influence factors, researchers, at home and abroad, have done much work, and both have contribute to the development of this field from different angles. Karen L. [2] identified the business value of consumer reviews and found that overall rating, attribute rating of purchase value, variation and volume of consumer reviews are significantly associated with hotel performance. Cheol Park [3] made consumer investigations in United States and South Korea, the results show that national culture has important moderating effects on the relationships among online reviews and its antecedents. Zheng Xiaolin [4] also revealed that the social features of reviewers are important in deriving better classification results. Kim and Gupta [5] focused on the effects of emotional expressions in eWOM and revealed that negative emotional expressions in a single negative review tend to decrease the reviews' informative value. Wenjing Duan [6] examined the persuasive effect and awareness effect of online user reviews on movies' daily box office performance and finally attributed the effect to the online user reviews as an indicator of the intensity of underlying word-of-mouth that plays a dominant role in driving box office revenues. Zhang Ziqiong [7] found that the volume of online consumer reviews are positively

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2015 SERSC

associated with the online popularity of restaurants. Jimenez [8] through their study indicated that more credible reviews lead to higher purchase intentions.

Although many scholars have made a lot of research, the study of user reviews mainly in the comment of two aspects: one is the effectiveness analysis and another is the influencing factors analysis. Actually, research on the influencing factors are summarized, such as service factor, quality factor, price factor, logistics factor, emotional factor and other broad categories. With the development of e-commerce, logistics seems to become more and more important. In the whole process of e-commerce transactions, logistics is actually the follow-up person and server for the business flow, and if there is no modern logistics, business flow activities will degenerate into a dead letter. Logistics plays a very important role in ensuring customer loyalty [9]. In this paper, on the basis of existing research, taking amazon reviews for example, as the commodity must be under the premise of a certain quality, we will study the effect of logistics factors on user comments and by subdividding the logistics factors into many smaller parts, the complex relationship between product categories and varieties of logistics factors could be found. The study may help the businesses improve its services specifically, and avoid the logistics factors that could affect the user satisfaction.

2. Logistics Factors Analysis

Online shopping reviews include many factors, for example, quality, service, logistics, emotional, commentators' subjective factors [10,11]. And yet logistics is one of the most critical influences that affect customer reviews.

Table 1. Logistics-Factor Comments

Comments	Logistics factors
A delivery that three days after the order is too slow	Speed of sending out goods
A near distance just a few hours but you need a few days	The delivery speed
Whether the goods arrived at my place is uncertain	The reminder service
I received a broken one	The commodity packaging
Size doesn't fit but returns is a hassle, so I just make do with it	Easy to return
The goods not be sent to my home, I had a long way to get it	Home delivery service
The courier came in spite of the rain	The humanized service
I could not find the express for me, maybe they lose it	The loss or lack of express items
The courier has a bad attitude	Service attitude

As can be seen from Table 1, except the quality of the goods itself, logistics is a part that cannot be ignored for merchants. First of all, taking the speed factor into account, the speed factors is subdivided into two part. On one hand is the business delivery speed, refers to a period of time that has expired between the order generation and the site shows they have sent out the goods. Another one is the transport speed, which is a period of time needed for the goods sent to the destination. The convenience factors consider more about the flexible and humanization of logistics services, such as whether the customers could get a home delivery. As we know, there may be many differences between virtual photos and real goods that is different from buyers expected, so whether you can easily return the goods is also a key part. The quality of service can be subdivided into several factors that

 Copyright ⓒ 2015 SERSC

are relatively close to the vital interests of the customers, which may involve, the commodity packaging, the loss or lack of express items and the service attitude. For goods that is fragile or need to treated with care, whether the delivery process pay attention to these details is an important manifestation of the logistics enterprise service quality. Standardized operation without a missing parts, a good attitude in the last mile delivery service and customer contact process, all have a great effect on the customer satisfaction. In addition, considering some special circumstances, such as weather condition, traffic impact. We take the delay delivery due to weather or traffic factors into account. From what has been discussed above, the logistics factors are divided into four categories, speed, convenience, service quality and service reliability. All is divided into a plurality of small parts, as is shown in Figure 1. At the same time, through the practice of investigation and research, the above factors of logistics got greatly identity.

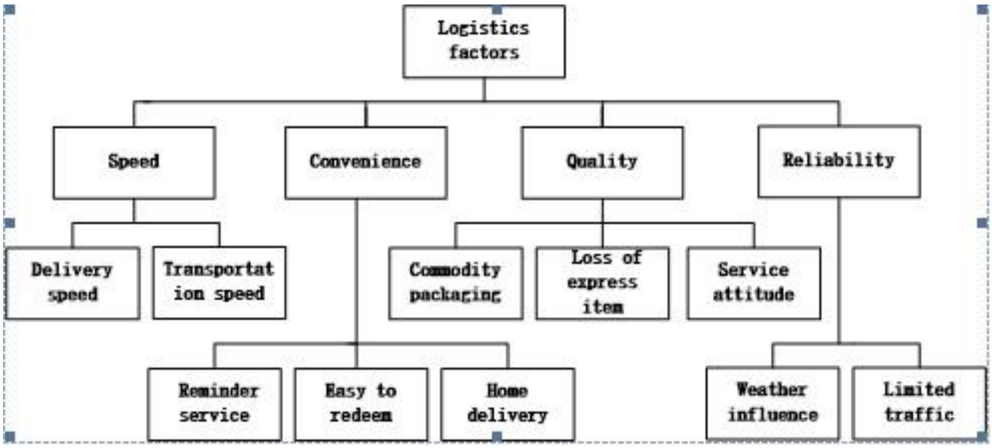

Figure 1. Logistics Factors

3. Logistics Factors Analysis Based on Association Rules

3.1. Correlation Analysis Model of Logistics Factors

Association rules analysis is an important branch of data mining, which was put forward for the first time in 1993 by Agrawal, in carries on the analysis to the market basket problem that aim to find customer buying patterns in merchandise sales [12,13]. This paper aims at mining association rules in online reviews and getting related information between the review ratings and the logistics factors given by users.

A given transaction set D is a user reviews data sets. Comment T for any affairs and T represents one of the user comments. A data set is the set of all data in T and it is a set of review ratings and logistics factors. Both A and B are sets, as a combination of several comments categories and logistics factors. Association rules is the implication of the form $A \Rightarrow B$, $A \neq \phi, B \neq \phi$ and $A \cap B \neq \phi$ [14]. If the rule was established in D, it needed to meet certain support and confidence. Support is the percentage for D contains $A \cap B$, that is a probability of $p(A \cup B)$, Confidence is the percentage for D contains A that also contains B, that is a conditional probability of $p(B \mid A)$. If the support and the confidence both meet their respective threshold, then the rule $A \Rightarrow B$ can be regarded as an interesting association rule. Discovering association rules between review categories and logistics factors is to find association from a large number of reviews data and fet the ones whose support and confidence are greater than the minimum support (min-sup) and the minimum confidence(min-conf).

$$\sup port(AB) = p(AB)$$

Copyright ⓒ 2015 SERSC

$confidence(AB) = p(B \mid A)$

where support and confidence have,

$confidence(A \Rightarrow B) = p(B \mid A) = \mathrm{support}(A \cup B) / \mathrm{support}(A)$

$confidence(A \Rightarrow B) = p(B \mid A) = \sup port\_count(A \cup B) / \sup port\_count(A)$

Such as a rule, { good review, delivery speed} $\Rightarrow$ { service attitude}, We have,

That is, the confidence here is the ratio of the probability value included {good review, delivery speed, service attitude} items and the probability value that contains {good review, delivery speed} items.

Only the support and confidence are not enough to filter out uninteresting association rules [15], so we can use the correlation measure degree to expand the support_confidence framework. That is,

$A \Rightarrow B[\sup port, confidence, correlation]$

A more effective measure of correlation namely, lift. If $p(A \cup B) = p(A)p(B)$, then, set A independent set B, or A and B set are related.

$$Lift(A,B) = \frac{p(A \cup B)}{p(A)p(B)}$$

by the probability knowledge,

$$Lift(A,B) = \frac{p(A \cup B)}{p(A)p(B)} = p(B \mid A) / P(B) = confidence(A \Rightarrow B) / \sup port(B)$$

As for a rule {good review, delivery speed } $\Rightarrow$ {service attitude},

If this value is less than 1, it means that the emergence of A and B is negative correlation. that is, if {good review, delivery speed} appears in a comment ,it is unlikely to have the {service attitude}; If the value is greater than 1, then A and B are positive correlation, namely, if {good review, delivery speed} appears in a comment, it is most likely to have the {service attitude};If the resulting value is equal to 1, then A and B are independent. there is no correlation between them.

3.2 Apriori Method

Apriori algorithm is a basic algorithm for frequent pattern discovery by limiting the candidates. The following Figure 2. is a flow chart of the Apriori algorithm. At the same time, in order to improve the efficiency of frequent item sets generated by layer, it can also be based on the Apriori property that is an important property for compressing the search space. That is, an item set must be frequent if it was generated by the non-empty frequent item set.

Copyright ⓒ 2015 SERSC

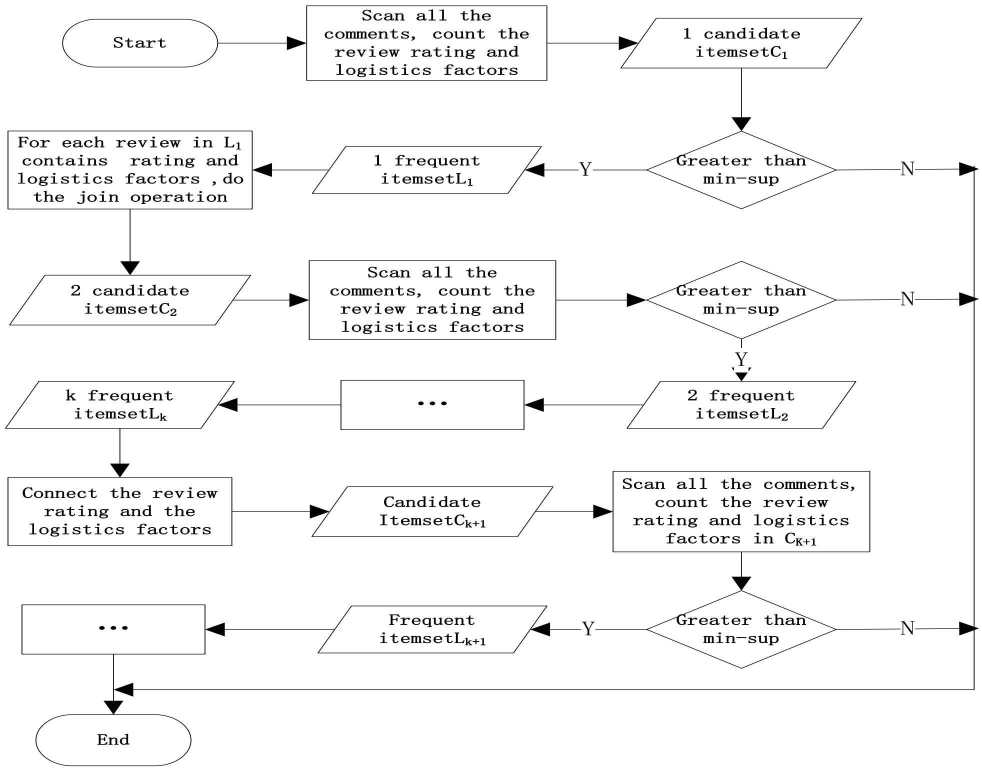

Figure 2. Flow Chart of Apriori Algorithm

(1)In the first iteration algorithm, each item was took as a candidate for the 1 item sets, that is denoted as C_1. It may involves items such as {delivery speed}, {transportation speed}, {home delivery}, {reminder service}, {easy to redeem}, {commodity packaging}, {loss of express item}, {service attitude}, {weather influence}, {limited traffic}.We also calculate the supports of them.

(2) Setting minimum support min_sup. Then, find the 1 frequent item sets L_1, they are all these C_1 whose support is greater than min_sup.

(3) As is shown in Figure 3, the following will be how to get the 2 candidate item sets C_2 and the frequent item sets L_2. One link operation on any two of frequent item sets in L_1, we will get the candidate item sets C_2, such as {good review, medium reviews}, {good review, bad review}, {good review, delivery speed}, {good review, transportation speed}, {good review, home delivery}, {good review, reminder service}, {good review, easy to redeem}, {good review, commodity packaging}, {good review, loss of express item}, {good review, service attitude}, {good review, weather influence}, {good review, limited traffic}, knowing as the permutation and combination of knowledge, C_2 will be composed of C^2_{L1} items. Scanning the comments to find the support of each C_2 items. By determining whether the support of each C_2 is greater than the min_sup, we will get the 2 frequent item sets $L_{2.}$

(4) Similarly, using the join operations to connect the two item sets that has a common item of L_2. we will get the candidate item sets C_3. for example, assume that L_2 has frequent item sets {good review, delivery speed}, {good review, commodity packaging}. So we can obtain a candidate item set {good review, delivery speed, commodity packaging}.

(5) In item sets C_3, also, we will calculate support for each candidate item sets, and compare the candidate's support with min_sup, and choose those support great than the min_sup as the frequent item sets L_3. Doing the join operations to connect the two item sets that has a common item of L_3. We will get the candidate item sets C_4. For example, there are frequent item sets {good review, delivery speed, commodity packaging}, {good

review, delivery speed, service attitude} in L_3, So we can obtain a 4 candidate item set {good review, delivery speed, commodity packaging, service attitude}.

(6) Calculate the support of each candidate item sets in C_4 ,and compared them with min_sup to get the 4 frequent item sets L_4.

(7) Continue to generate candidate item sets by connection operation and compare each support with min_sup to get frequent item sets.

(8) Until the candidate item sets generated by connection operation do not satisfy the support conditions, that is do not greater than min_sup, the algorithm terminates.

(9) During the process of generating the candidate item sets and then generating the frequent item sets, we can use the Apriori property for a Pruning operation in order to improve the efficiency of frequent item sets generation. The Apriori property is that, if there is a subset of candidate item sets, it is not frequent, then, the candidate item sets is not a frequent item sets. for example, assuming the candidate item set {good views, delivery speed, commodity packaging}, then its 2 subset {good review, delivery speed}, {good review, commodity packaging}, {delivery speed, commodity packaging}. But in all of the 2 item sets of frequent item sets in L_2, they are not including the elements {delivery speed, commodity packaging}, Then, based on the Apriori property, {good review, delivery speed, commodity packaging} can be removed from the C_3. This is definitely not the frequent item set, there is no need to scan all the comments to calculate the support to determine whether it is the frequent item set.

Generate C_2 by L_1 →

Itemsets	Support
{good review, delivery speed}	I_{11}
{good review, packaging}	I_{12}
{good review, service attitude}	I_{13}
{delivery speed, packaging}	I_{14}
{delivery speed, service attitude}	I_{15}
{ packaging, service attitude}	I_{16}

Greater than min-sup →

Itemsets	Support
{good review, delivery speed}	I_{21}
{good review, packaging}	I_{22}
{good review, service attitude}	I_{23}
{delivery speed, service attitude}	I_{25}

Generate C_3 by L_2 →

Itemsets	Support
{good review, delivery speed, packaging}	I_{31}
{good review, delivery speed, service attitude}	I_{32}
{good review, packaging, service attitude}	I_{33}

Figure 3. Candidate Item-Sets and Frequent Item-Sets Generated Schematic Diagram

(10)To the finally get frequent item sets, lists all nonempty proper subset of each item sets. Such as, if, finally received only 3 frequent item sets {good review, delivery speed, service attitude}, the non empty subset are {good review, delivery speed}, {good review, service attitude}, {delivery speed, service attitude}, {good review}, {delivery speed}, {service attitude}. For each non empty subset, calculate the confidence as,

Also, we can calculate the confidence of {good review, service attitude } $\Rightarrow$ {delivery speed}, {delivery speed, service attitude} $\Rightarrow$ {good review}, {good review} $\Rightarrow$ {delivery speed, service attitude}, {service attitude} $\Rightarrow$ {good review, delivery speed}, {delivery speed, service attitude} $\Rightarrow$ {good review, service attitude}. If the confidence of a rule is greater than the minimum confidence threshold (min_conf), then output of the rules.

 Copyright ⓒ 2015 SERSC

4. Case Study

4.1 Data Preparation

Choose the data from amazon sales reviews. As many researchers used Amazon acquisition for experimental data analysis, so the data got certain reliability and persuasive [16,17]. The reviews data includes data categories, the evaluation level and whether it involves logistics factors. Here, considering a variety of logistics factors, including speed (delivery speed, transportation speed), convenience (home delivery, reminder service, easy to redeem), quality of service (packaging, loss of express item, service attitude), reliability (weather factor, traffic factor).

Some data interception involves the ID number of commodities, the commodity categories and the logistics factors are given as follows (see Table 2).

Table 2. Data Collection Style

Id	Items	Description
1	Food	<good review, shops were very good and very fast delivery speed>
2	Clothes	<bad review， slow delivery, return postage is too expensive>
3	Digital products	<good review, a fast delivery speed, I got the SMS notification, courier had a good attitude>
4	Cosmetics	<medium review, just from Yantai to Jinan shipping for 4 days>

Table 3. Review Data Includes Rating

Id	*Food*	*Clothes*	*Digital*	*Cosmetics*	*Good review*	*Medium review*	*Bad review*
1	yes	no	no	No	yes	no	no
2	no	yes	no	No	no	no	yes
3	no	no	yes	No	yes	no	no
4	no	no	no	Yes	no	yes	no
5	yes	no	no	No	no	no	yes

Table 4. Review Data Includes Logistics Factors

Id	*Delivery speed*	*Transportation speed*	*Home delivery*	*Reminder service*
1	yes	No	no	no
2	yes	No	no	no
3	no	Yes	no	yes
4	no	Yes	no	no
5	yes	No	no	yes

Copyright ⓒ 2015 SERSC

In the data acquisition and processing process, data were randomly selected to ensure the consistency. Randomly selected a certain category of goods in different classes, and the data of each category were random sampling.

The data sets get categories as clothing, shoes and hats, food, digital products, home appliances, cosmetics *e.g*. As well as the corresponding commodity reviews which may involves good review, medium review and bad review .

For the logistics factors affecting comments, we considered about 10 factors. that are included in speed, convenience, service and quality.

The logistics factors involved in the review are in a variety of ways and will be classified into different logistics factors categories according to semantic expression. Just connect Table 3 and Table 4 according to the same Id, then we can get the data processing style.

If the use of 0 represents not relates to this property, 1 represents relates to, the data can be represented as is shown in Figure 4.

$$\begin{bmatrix} 10001001000 \\ 01000011000 \\ 00101000101 \\ 00010100100 \\ 10000011001 \end{bmatrix}$$

Figure 4. 0-1 Data Processing Style

4.2. Association Rules Analysis

Clementine is used for data mining and Apriori algorithm for constructing the rule model.

Table 5. Correlation Analysis of the Preceding and Consequent Paragraph

Preceding	*Consequent*	*Support*	*Confidence*
Digital product	Express speed	14.286%	96.0%
Cosmetic	Delivery speed	12.571%	92.0%
Shoes	Delivery speed	28.571%	86.0%
Home applicant&Good view	Home delivery	13.429%	95.745%
Cosmetic&Good view	Packaging	12.857%	88.889%
Food&Medium view	Packaging	11.143%	92.308%
Clothes&Good view	Return convenience	12.571%	86.364%
Shoes&Good view	Express speed	22.286%	83.333%
Digital product&Good view	Delivery speed	12.286%	83.721%

Intercept part experimental results as shown in Table 5. Overall, for e-commerce shopping transactions, goods delivery speed and the transportation speed are the main parts of the impact of user comments, and for different commodity classes:

 Copyright ⓒ 2015 SERSC

(1) Among customers who bought food commodities and gave medium review, 92.3% of people will be involved in the breakage of the packages problems in their comments.

(2) Among customers who bought clothing products and gave good review, 86.36% of people will be involved in returned goods convenience problems in their comments.

(3) Among customers who bought cosmetics products and gave good review, 88.89% of people relates to the packing problems in their comment.

(4) Among customers who bought home appliances products and gave good review, 95.7% of people involved door-to-door problem in their comment.

Table 6. Model Assessment

$A-9 field-1	*$AC-9 field-1*	*$A-Rule_ID-1*	*$A-9 field-2*	*$AC-9 field-2*	*$A-Rule_ID-2*	*$A-9 field-3*	*$AC-9field-3*	*$A-Rule_ID-3*
factor_1	0.910	27	Factor_2	0.867	15	factor_5	0.86	11
factor_1	0.910	27	Factor_2	0.867	15	factor_5	0.86	11
factor_1	0.910	27	Factor_2	0.867	15	factor_5	0.86	11
factor_1	0.910	27	Factor_2	0.867	15	factor_5	0.86	11
factor_1	0.910	27	Factor_2	0.867	15	factor_5	0.86	11
factor_1	0.910	27	Factor_2	0.867	15	factor_5	0.86	11
factor_2	0.865	14	Factor_1	0.85	28	factor_5	0.84	12
factor_2	0.865	14	Factor_1	0.85	28	factor_5	0.84	12
factor_3	0.96	8	Factor_2	0.96	10	factor_1	0.9	9
factor_3	0.96	8	Factor_2	0.96	10	factor_1	0.9	9
factor_3	0.96	8	Factor_2	0.96	10	factor_1	0.9	9

Copyright ⓒ 2015 SERSC

factor _3	0.96	8	Factor _2	0.96	10	factor _1	0.9	9

The model is applied to the data prediction. We just intercept some of the results data as is showed in Table 6, with 9 rows of data. The data shows the predicted results using 3 rules with the highest confidence of the model. The recorded prediction results:

$A-9field-1=factor_1 $AC-9field-1=0.910 $A-Rule_ID-2=27

$A-9field-2=factor_2 AAC-9field-2=0.867 $A-Rule_ID-3=15

On the analysis of logistics factors of all categories in the comments, probability of the factors that related to factor_1 (delivery speed) is 91%, and the probability of the factors related to the factor_2 (logistics delivery speed) is 86.7%. and the rules ID are 27 and 15. The model predicted results are in conformity with our test results.

Thus, for customers, whatever commodity they bought, speed is more important logistics factor than others. For food commodities, as is the consumption of goods, so customers tend to care more about their packaging. For apparel items, sometimes due to the color and the size are not fit, and thus for the customer, whether it can be easily returned appears more important. For cosmetics that women commonly used, packaging has a direct impact on the customer’s impress whether the goods is true and whether they can use at ease. For the home appliance products, generally because they are the heavier, larger ones, customers may concern more about the door-to-door service.

For businessman engaged in different categories, they could pay attention to logistics factors according to their different goods categories. For businessman engaged in food business, a complete and tight packing, make sure that the buyer receives a not crush deformation package, will indirectly increase the buyer's satisfaction For clothing sellers, if you showed in the web page, such as, no reason to return, or the seven day package returns, it will definitely eliminate the buyer’s hesitation and concerns when they placing a order. For cosmetics sellers, delicate and tight packing will be impressing, and buyers may believe that the merchants and their goods are trustworthy. For the business of home appliance, do a good job at the after sale service is very important, and a door to door delivery, free installation and other services will be a plus for businesses in virtually.

5. Conclusions

In this paper, based on the association rule in data mining method, we analyzed the impact of logistics factors in electronic commerce transactions on user comments. Through the analysis of the logistics factors in different categories of goods, we explored the logistics factors impact on user reviews of different categories of goods. So from a merchant’s perspective, they could ameliorate many issues according to their own businesses and so that they could improve their services and customer satisfaction.

In the logistics factors classification, we divide them into different categories of factors according to the user comment semantics [18]. But how to borrow other technology method, so as to avoid the potential man-made factors will be the direction of further improvement.

Acknowledgements

This work is sponsored by National Natural Science Foundation project(71101090), Shanghai Top Academic Discipline Project- management science & engineering, Doctoral Fund of the Ministry of Education (20133121110001), Shanghai Yangfan Program (14YF1411200), Shanghai Municipal Education Commission Project (12ZZ148, 13YZ080, 14YZ112), and Shanghai Maritime University Research Project (20120102). We also thank anonymous referees.

 Copyright ⓒ 2015 SERSC

References

[1] N. Hu, L. Liu and J. J. Zhang, "Do online reviews affect product sales? The role of reviewer characteristics and temporal effects," Information Technology and Management, vol. 9 no. 3 **(2008)**, pp. 201-214.
[2] K. L. Xie, Z. L. Zhang and Z. Q. Zhang, "The business value of online consumer reviews and management response to hotel performance," International Journal of Hospitality Management, vol. 10 no.43 **(2014)**, pp. 1-12.
[3] C. Park and T. M. Lee, "Antecedents of Online Reviews' Usage and Purchase Influence: An Empirical Comparison of U.S. And Korean Consumers," Journal of Interactive Marketing, vol. 23 no.4 **(2009)**, pp. 332–340.
[4] X. L. Zheng, "Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach," Decision Support Systems, vol. 56, **(2013)**, pp. 211–222.
[5] J. Y. Kim and P. Gupta, "Emotional expressions in online user reviews: How they influence consumers' product evaluations," Journal of Business Research, vol. 7 no. 65 **(2012)**, pp. 985-992.
[6] W. J. Duan, B. Gu and A. B. Whinston, "Do online reviews matter? — An empirical investigation of panel data," Decision Support Systems, vol. 45 no. 4 **(2008)**, pp.1007-1016.
[7] Z. Q. Zhang, Q. Ye and R. Law, "The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews," International Journal of Hospitality Management, vol. 29 no. 4 **(2010)**, pp. 694-700.
[8] F. R. Jiménez and N. A. Mendoza, "Too Popular to Ignore: The Influence of Online Reviews on Purchase Intentions of Search and Experience Products," Journal of Interactive Marketing, vol. 27 no. 3 **(2013)**, pp. 226-235.
[9] R. Ramanathan, "The moderating roles of risk and efficiency on the relationship between logistics performance and customer loyalty in e-commerce," Transportation Research Part E: Logistics and Transportation Review, vol. 46 no. 6 **(2010)**, pp. 950-962.
[10] L. M. Willemsen, P. C. Neijens and F. Bronner, "Highly Recommended! The Content Characteristics and Perceived Usefulness of Online Consumer Reviews," Journal of Computer-Mediated Communication, vol. 17 no. 1 **(2011)**, pp. 19-38.
[11] C. M. Y. Cheung and C. L. Sia, "Is This Review Believable? A Study of Factors Affecting the Credibility of Online Consumer Reviews from an ELM Perspective," Journal of the Association for Information Systems, vol. 13 no. 8 **(2012)**, pp. 618-635.
[12] R. Agrawal, "Mining association rules between sets of items in large databases," Proceedings of 1993 ACM SIGMOD. International Conference of Management of Data, Washington D C: ACM, **(1993)**, pp. 207-216.
[13] A. Verma and S. D. Khan, "Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports," Safety Science, vol.70, **(2014)**, pp. 89-98.
[14] Z. Z. Shi, "Knowledge discovery," Beijing: Tsinghua University Press, **(2002)**.
[15] J. W. Han and M. Kamber, "Data Mining: Concepts and Techniques," 3rd edition, Morgan Kaufmann, **(2011)**.
[16] S. Skalicky, "Was this analysis helpful? A genre analysis of the Amazon.com discourse community and its most helpful product reviews," Discourse, Context & Media, vol. 2 no.2 **(2013)**, pp. 84–93.
[17] P. Ritala, A. Golnam and A. Wegmann, "Competition-based business models: The case of Amazon.com," Industrial Marketing Management, vol. 43 no.2 **(2014)**, pp. 236-249.
[18] L. Q. Zhou, S. Ye and P. L. Pearce, "Refreshing hotel satisfaction studies by reconfiguring customer review data," International Journal of Hospitality Management, vol. 38, **(2014)**, pp. 1-10.

Authors

Qingtian Zhang, (1991-), female, born in Shandong, master graduate student, specialized in e-business and data mining.

Youfang Huang, **(**1956-), male, born in Zhejiang, doctor of engineering, professor, tutor of doctoral students, president of Shanghai Maritime university, specialized in supply chain and logistics.

Wei Yan, (1968-), male, born in Zhejiang, doctor, professor, tutor of doctoral students, specialized in knowledge engineering, logistics engineering and management.

Copyright ⓒ 2015 SERSC

Yu Wang, (1986-), male, born in Shanghai, doctor, Engineer, specialized in knowledge engineering, logistics engineering and supply chain management.

 Copyright ⓒ 2015 SERSC