A Multi-attribute Model to Compute Review's Value in C2C

Leiyue Yao
Information College, JiangXi University of Technology, Nanchang, China
special8212@sohu.com

Wei Yang
Information College, JiangXi University of Technology, Nanchang, China
yang.wei@163.com

***Abstract*—The valuable review can help consumers to make decisions in C2C e-commerce. Due to the subjectivity of review, instability of product features, and cheat behavior in C2C, the main work in this paper is to design a multi-attribute model for the assessment of review's value based on the review's content, reliability, timeliness. The preliminary experiments show that the proposed way can filter informative review effectively than a quality assessment approach constructed based on user-generated helpful votes, and can resist the fraud review in C2C e-commerce.**

***Index Terms*—online review, review's value, multi-attribute model, C2C e-commerce**

I. INTRODUCTION

Most e-commerce sites allow users to give reviews after purchasing. The reviews published by buyers not only promote the interaction of buyers and sellers on C2C e-commerce site, but also play a good role in word-of-mouth marketing [1]. On the one hand, the product reviews will be used to compare different goods for potential buyers, and influence purchasing decision-making; on the other hand, the reviews also can be as the data source for market feedback surveying for manufacturers [2]. With the great development of the C2C e-commerce, the number of reviews grows rapidly, and some hot commodities often have hundreds or even thousand reviews. A huge massive of review data available on the Website has created an information overload problem among potential buyers [3]. So it is difficult to rely on artificial way to discover useful review among the complicated data, and that will affect the user to build consumer confidence [4].

The online user review system can provide basis to consumers to share their opinions and experience with the products. The system offers valuable resources of information for potential buyers to make more efficient and rational purchase decisions. But the anonymity, non-contact characteristics of the C2C E-commerce, will bring unreliability to reviews. These subjective reviews will lead potential buyers have to guess the reliability of each review, and even to weigh some conflicting perspectives [5]. Comparing to other e-commerce, C2C e-commerce is a personal trading pattern. And the asymmetric information will bring instability and risk. In order to make the product more attractive or to eliminate the impact of negative reviews, sellers may take some fraudulent transactions to forge some reviews. These fraudulent reviews will increase searching costs of goods, reduce the efficiency of decision-making, and even affect the buyers to make decisions on purchasing behavior. Apparently, the potential buyers want to focus on more reliable and more influential reviews [6].

Therefore, there is an urgent need to use certain techniques to make the process convenient, allow users to find trustworthy and reliable review more quickly and accurately, and improve the efficiency of C2C e-commerce.

II. RELATED WORK

The concept of "helpfulness of online reviews" was proposed by Chatterjee in 2001 at first time [7]. In the period of Web 2.0, how to identify these helpful reviews is becoming an increasingly important subject for both academics and practitioners. The "helpfulness" of reviews can be understood as an influence degree of user on the information used. Many B2C, C2C websites, such as Amazon, Dangdang, Taobao, allow users vote the "usefulness" reviews while they read these reviews. And the system can judge the value of review according to the number of votes these reviews gained. However, the vote mechanism, which completely relies on artificial judgment, is difficult to play an important role in review's value evaluation. On one hand, the buyer can determine which review is reliable only after have received the product, but few buyers will come back to confirm and vote the value of another review after the transaction. On the other hand, newly published online reviews can reflect the current quality of goods timely. But these reviews will take some time to obtain the vote from buyers, and some truly reliable reviews may also be difficult to be presented to potential buyers [8].

In order to solve the limitations of online product review's value and utility judgment in e-commerce site, many scholars have done some research work based on evaluation of the information quality and text mining method. Ghose (2007) believed that the subjective

doi:10.4304/jcp.9.11.2648-2655

opinion of product reviews can be used to observe its value, and review's content is an important aspect of the performance of its value [8]. Li Chen (2012) defined perceived usefulness of online reviews, and pointed out that the online reviews content mining should include product attributes, and consumer attitudes [9]. Kim (2007) proposed an automatic detection method for low-quality reviews by information quality, subjectivity, readability aspects [10]. The information quality reveals product features; subjectivity reflects the emotional characteristics; readability reflects text feature. These studies only consider the review's property in isolation, and ignore the impact from the relation of the review's semantic. If only take the review's content into account, some conflicting conclusions may be deduced from different data mining methods, which will make e-commerce site managers confused. Liu (2008) supported that the helpfulness of review depends on the reviewer's expertise, writing style of review, timeliness of review [11]. And get a better result to predict the movie review data by use of the nonlinear regression model. In addition to considering the factors of the review content, the time of review has also been analyzed, which can reflect the product attribute timely. Liu (2007) realized the automatic evaluation of review's helpfulness from structural, lexical, syntactic, semantic, meta-data by using support vector machine regression method. The experimental results show that review's length, unigram, and product ratings are the key elements in the assessment of the usefulness [12]. Danescu (2009) proposed the revised argument by analyzing the statistic characteristics of the product rating [13]. He believed that the usefulness of review deduced with the deviation from mean rating, the positive rating higher than the mean is more useful than negative reviews lower than the mean. It is in line with filtering untrue feedback by statistical properties. The longer review can reflect more product features in a certain extent, but that may also doesn't contain product features, and users have no interest in it.

In this paper, we approach this problem from a different perspective in order to synthesize a variety of review's characteristics. Our paper differs significantly from previous studies in that we investigate the factors that determine the value of review in according to considering the subjectivity of review, instability of product features in C2C, and cheat behavior in C2C. Another key aspect of our study that differs significantly from prior research is that in order to find the concerned characteristics, we use the semantic relation among reviews to build review's networks, and compute the information quality based on importance of review nodes. Finally the approach import trust technology to compute the reliability of each review to inhibit fraud feedback. In addition to quality of review' content, the behavioral characteristics of the reviewers, and the time of review published are also important to assess the value of review. We design a multi-attribute model to measure the value of review, which including information quality of review's content, the reliability and timeliness of feedback. However, our approach and prior studies are not mutually exclusive, and in fact, our approach complements previous studies.

III. The Proposed Multi-attribute Review's Value Calculation Method

A. An Overview of Multi-attribute Method

The review's value calculation in our study is a consumer-oriented mechanism which can present the helpful and valuable reviews to potential buyers and improve their efficiency of decision making. Fig.1 displays the main components and procedures to calculate value of product review in C2C website.

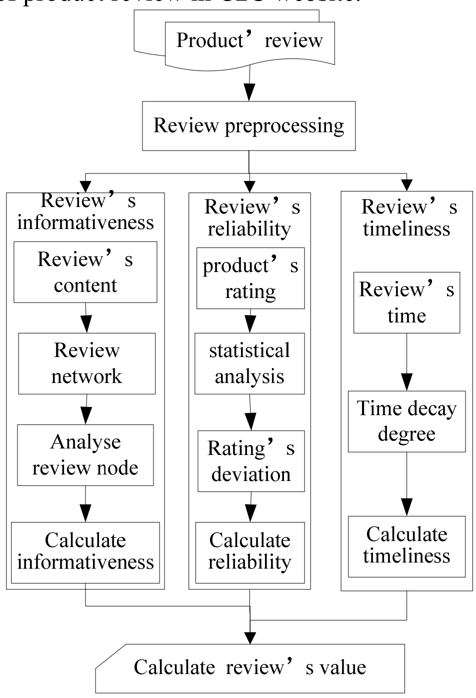


Figure 1. Review's value calculation architecture

In the work, we developed a multi-attribute calculation method by considering the following three main factors in unison. The review's information quality, which is the most important index of review's value, will provide users with product's information and help users to understand product's quality. The review's reliability, which can reflect whether the review is fair and objective, will help users to filter the unfair or fraud feedback. The reviewer's timeliness is a consensus that time decay should be considered to maintain a more accurate product's features in time for online C2C. Based on the three proposed factors described above, the global review's value can be synthesized as follows.

$$RV_i = RI_i \times (w_1 \cdot RR_i + w_2 \cdot RT_i) \tag{1}$$

RI is the review's information quality, *RR* is the review's reliability, *RT* is the review's timeliness. From the aspect of user reading reviews, review's value *RV* is directed to *RI*, and adjust by *RR* and *RT*, so with respect to review *i*, the *RV* is defined as (1), Where w_1, w_2 is the weight of the corresponding factor, and $w_1 + w_2 = 1$.

B. Review Preprocessing

The development of e-commerce brings rapid growth of the online review. Some hot commodities usually have hundreds or even thousands of reviews. However, a lot of reviews may have no value, such as the default review, expired review and so on. In order to improve the efficiency of value computing, we set a review preprocessing mechanism refer to other research on feature selection as [14]. The preprocessing work will filter worthless reviews to reduce the amount of reviews. The pretreatment can be divided into two categories.

(1) Filter expired reviews: review has timeliness, as the shelf life of the product. The distant reviews can prove the current status of the goods poorly, and will lack of value. When the review time is beyond the default effective date, the reviews will be filtered by the system.

(2) Filter no informative reviews. Many users do not take the initiative to give reviews in feedback system of e-commerce, and the system will create several default reviews. Some users do not responsible to write reviews, and not involving any product feature. These reviews can be filtered for their poor value.

C. Review's Information Quality

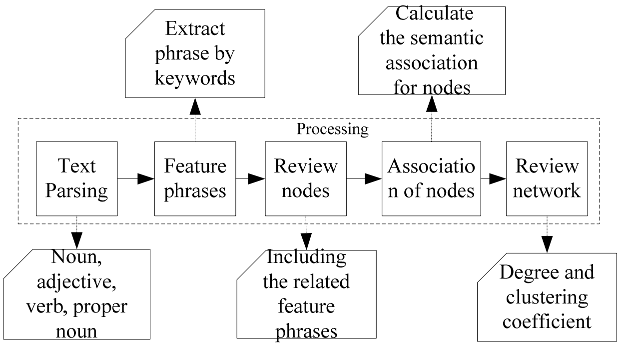


Figure 2. Procedure of calculation of review content information

Review's information quality is used to evaluate the quality of review content, which based on review's network. The product's features which user concern in common will extract from massive, unstructured, noise, uncertainties reviews. In this process, the information quality of review's content determines its value. All the reviews of one product will present a feature of "universal contact and high cohesion". Even the content is random to an individual review. If each review look as a node and the semantic association between any two reviews look as edge, all reviews data will perform the network topology for content interconnection. The information quality of review's content is depending on the possession degree of the product characteristics other people focus on. That is to say, the more valuable of the reviews, the more features it mentioned will be displayed frequently in another review. From the view of network, because there is an edge between the same or similar product reviews, more edges connect to other review indicated more nodes are affected by the node in the network. The node will have a relatively important role in the network, which will make it more informative.

The review's mining methodology to compute information quality is shown in Fig. 2. The right side of the figure illustrates the outcome of each step, and the left side shows how to mine the review text.

(1) Text parsing.

In this paper, we use the Chinese word segmentation tool ICTCLAS which developed by Chinese Academy of Sciences to comment text word and filter out duplicate words. Similarly, the ICTCLAS is used to comment after the speech tagging. Tagging method can require one label or two. The difference is one marked only nouns, verbs, etc.; two can mark a more specific situation, including those with functional nouns or adjectives verbs, nouns, etc. In order to improve the mining precision, we use two dimensions. Then we compiled a list of the meaningless words (generally called "stop word" like "a, an, the") and eliminated them from the term-by-frequency Matrix, which reduced the number of columns, but keep the punctuation marks to form some natural keywords group.

(2) Feature phrases.

The feature phrases can be constructed by these segmented keywords. Feature phrases are used to describe the main features of the product based on experience. These features include descriptive characteristics, just like color, style; and behavioral characteristics, just like wear, touch feeling. The combination rules can be defined as noun or verb +adjective, and some independent adjective. So the keywords group of the reviews will be listed according to the order of reviews; the keywords that are marked as nouns or verbs will connect with the neighbor adjectives automatically in one keywords group. Some adjectives can't match with the verbs and nouns can be used alone to form a characteristic phrase.

(3) Review nodes.

The most intuitive review network construction method is to use each text of review as node, use semantic association between the reviews as edge. However, the reviews are subjective, personalized and unstructured that bring difficulties to measure the information quality. Through the pretreatment of review text, the structured review consisted of key feature phrases can be used as a review node. Supposed that the review r consists of k feature phrases, it can be defined as $r = \{p_1, p_2, \ldots pk\}$. p_i is a feature phrases of review, which can be as an attribute or item in the review node.

(4) Review semantic association.

In the process of building review network, we use the ESA to calculate the semantic association between key features phrases in each review. In natural language processing and information retrieval, explicit semantic analysis (ESA) is represented by vectorial text based on Wikipedia. Specifically, in ESA, a word is represented as a column vector in the tf–idf matrix of Wikipedia's article text and a document is represented as the centroid of the vectors.

(5) Review network.

It is a network constructed by the semantic association among reviews. But time-dependent is not taken into consideration. The network can be presented as a directed weighted graph, $G = (V, E)$, where $V = \{v_1, v_2 \dots v_k\}$, V is a non-empty finite set of the n nodes; E is the set of edges between nodes. Only if the semantic association is greater than a certain threshold value, then the semantic relations is established.

1) Degree analysis: The degree of node i is the number of other nodes connected to it. The degree of the node is referring to the number of nodes which is semantically associated with the current node in review network. Some nodes with high degree will contain most of the product's features, which is conducive to quickly identify user needs. The statistical characteristics also suggest that a bigger size of the review data is not the key to find valuable reviews, but is how to find useful reviews.

2) Clustering coefficient: The clustering coefficient reflects the close links between adjacent nodes. The nodes in review network are gathered together to describe specific product features. The node with high degree indicated the common features are concerned by users.

In the review network, the degree of review i reflects the amount of other review related to the review i. And clustering coefficient reflects a close association. The key words are extracted from the review by characterizing processing. Supposed that the number of key words of review i is n_i, the number of key words of review j have an association with key words in the review i is n_{ij} . The associated value must greater than the threshold. Comparing with the review j, the information of review i is n_i / n_{ij}. In the review network, all the nodes related to review i are defined as a review list $\{ r_1, r_2 \dots r_m \}$, and the information quality of review i can be expressed as the sum of information in the review list. In order to avoid the denominator is zero that makes computing meaningless, the denominator and numerator will plus 1 when compute the review information quality RI as follows:

$$RI_i = \sum_{j=1}^{m} \frac{n_i + 1}{n_{ij} + 1} \tag{2}$$

D. Review's Reliability

Review's reliability depends on whether it is a fair, objective feedback. Theoretically, a high reliable feedback indicated it has a lot of influence, and should be displayed first. In C2C e-commerce, buyers can rate the products they purchased. It is reasonable to give equal importance to all reviews if all the reviewers are honest. Some reviewers will give unfair feedback to achieve its own purposes, such as exaggeration and slander. If it is supposed that most of the reviewers are honest in a valid C2C community, the reliability of the review will decrease with the rating deviation from the perspective of the statistical distribution of the feedback. If the product's rating is a ten-point scale in C2C business, just like {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. The higher scores show the better quality. f_i is the rating corresponding to review i, and the rating set is $\{f_1, f_2 \dots f_i\}$. The process to compute review's reliability is as follows.

Step 1: Calculate average rating f_a of all reviews corresponding to a product.

$$f_a = \frac{1}{n} \sum_{i=1}^{n} f_i \tag{3}$$

Step 2: Calculate the deviation of each rating related to review i. Because of the ten-point taken in the work, the normalized deviation degree p_i is as eq.5.

$$p_i = \frac{1}{10} \sqrt{(f_i - f_a)^2} \tag{4}$$

Step 3: Calculate the reliability of each review i. Because of value of normalized deviation is between 0 and 1, the value of reliability is also in [0,1].

$$RR_i = 1 - p_i \tag{5}$$

Theoretically, a higher value of the reliability means that the review will have an important influence and should be presented preferentially in the system. If a reliable customer gives the positive review, the potential buyers will likely to believe that the goods or services provided by the seller are with high quality. Otherwise, if a reliable customer gives the negative review, it is easy to imply that the goods or services provided by the seller are with lower quality [15].

D. Review's Timeliness

Reviewer's timeliness implies that the authorized reputation value of product will diminish over time. Very old transactions and reviews are difficult to prove the latest status of goods, and reduce the effect to influence user's purchasing decisions. In view of this, we have defined the timeliness as follow:

$$RT_i = \begin{cases} 1 - d \cdot (n/u), & if \quad n \le M \\ 0, & if \quad n > M \end{cases} \tag{6}$$

Where n is the number of days between now and the review's date. If select a period of time u for the evaluation of the time unit, all the evaluations in the time unit will have same attenuation effect. d is the rate of decay for time unit u, M represents the longest number of days that review remains effective. Comparing with the more commonly used exponential time decay equation, such as $e^{-n/M}$, the characteristics of Eq.6 are as follows:

1) Because trading records are sparse for lots of users, the unit of time decay is an adjusted time unit u. 2) RT_i can be adjusted according to the preferences of the user by changing u, d and M, which will help managers make adjustments based on actual transactions.

E. A Complete Example for Review Value Calculation

In this section, we use an example to explain how the proposed method calculates the value for a given review. It is supposed that there are ten reviews as Tab. I.

In the review preprocessing, some expired reviews should be filtered first. It is supposed that the current time is April 1, 2013, the longest number of days M is 720 days, the time unit U is 30 days, and the decay factor d is 0.05. The timeliness of No.9 and No.10 reviews are expired and should be filtered. Then some meaningless reviews should be filtered too. Some reviews only have a little text, including meaningless text and can not reflect the product's features. These reviews have no value to users, like No.7 and No.8 reviews. So after the preprocessing of review set, six reviews No.1 to No.6 will be obtained.

TABLE I
EXAMPLE OF REVIEWS

No.	Review' content	Time	Rating
1	It is suitable, have nice color, fast shipping, I like it very much.	2013.4.1	10
2	It is quick shipping, and it is very cheap, have a great quality	2013.1.30	8
3	The seller is great, it is a very nice camisole, have pretty color, shipping was fast too!	2013.1.1	9
4	It's very suitable, the color is beautiful, fast shipping,	2012.11.1	7
5	It is very expensive and style is bad too	2012.4.1	1
6	Excellent seller, lighting fast shipping, perfect quality, thanks!	2011.10.1	9
7	A	2011.9.1	2
8	!	2011.8.1	9
9	Size is ok, color is beautiful	2011.1.1	7
10	Size is a little small, price is very cheap	2010.9.1	8

To calculate the timeliness, we have defined the value of d, M, U above, use the No.2 review's timeliness as an example. The data is on January 30:

RT_2=1-0.05*(60/30)=0.90

To calculate the reliability, we should calculate the average score f_a of all score first, and obtain deviation p_i for review i, use the No.2 review's reliability as an example:

f_a=(10+8+9+7+1+9+2+9+7+8)/10=7.0

p_2=(8-7.0)/10=0.1 then RR_2=1- p_2=0.9

Thus, the feature phrases of review, the timeliness and reliability of six valid reviews can be calculated, the result is as Tab.II.

TABLE II
REVIEWS AFTER FILTERING

No	Feature phrases	Timeliness	Reliability
1	suitable, nice color, fast shipping, like	1	0.7
2	quick shipping, cheap, great quality	0.9	0.9
3	seller great, nice camisoles, pretty color, shipping fast	0.85	0.8
4	suitable, color beautiful, fast shipping	0.75	1
5	expensive, style bad	0.4	0.4
6	excellent seller, fast shipping, perfect quality	0.1	0.8

We need to calculate the correlation between the reviews to construct the review network. Semantic association of keywords in the reviews is summarized to calculate information quality in accordance with the proposed method. Use No.1 review as an example, it has six keywords, and keyword list is {Suitable, nice, color, fast, shipping, like}. They will form four feature phrases {suitable, nice color, fast shipping, like}. The No.2 review has three feature phrases, and feature phrase of No.2 review is {quick shipping, cheap, great quality}. There is only one feature phrase in No.2 review can create semantic associations with No.1 review, which is {quick shipping}. In a similar way, according to the infomativeness measurement, the relative information of No.1 review compared to No.2 review is (4+1) / (1+1) =5/2. The relative information of No.1 review compared to reviews No.3, No.4, No.5, No.6 is {5/3, 5/4, 5/1, 5/2} as the Tab.III. So the information quality of No.1 can be summed: (5/2+5/3+5/4+5/1+5/2=12.92).

TABLE III
RELATIVE INFORMATION OF EACH REVIEW

Review	No.1	No.2	No.3	No.4	No.5	No.6
No.1	NULL	4/2	5/3	4/4	3/1	4/2
No.2	5/2	NULL	5/2	4/2	3/1	4/3
No.3	5/3	4/2	NULL	4/3	3/1	4/3
No.4	5/4	4/2	5/3	NULL	3/1	4/2
No.5	5/1	4/1	5/1	4/1	NULL	4/1
No.6	5/2	4/3	5/3	4/2	3/1	NULL
Sum	12.92	11.33	12.50	10.33	15.00	10.67

The review network as the Fig. 3 can be constructed by the compared information as the Tab.III.

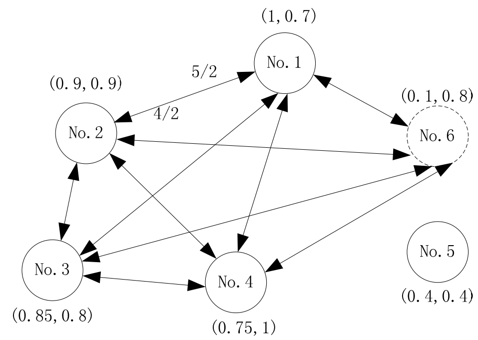


Figure 3. Review network

We can get the Information quality, timeliness, reliability of each review from the Tab. II and Tab. III. The weight of timeliness and reliability are set {w_1=0.5, w_2=0.5}. Like the Tab. IV, if we compute the value of No.1 review, the *RI*, *RT*, *RR* are {12.92, 1, 0.7}, and w_1, w_2 are {0.5, 0.5}, so the value is 12.92*(0.5*1+0.5*0.7)=10.98. Thus, the results of all reviews are as Tab. IV.

TABLE IV
THE RESULT OF REVIEW'S VALUE

No	Information quality	Timeliness (w_1=0.5)	Reliability (w_2=0.5)	Review's Value
1	12.92	1	0.7	10.98
2	11.33	0.9	0.9	10.20
3	12.50	0.85	0.8	10.31
4	10.33	0.75	1	9.04
5	15.00	0.4	0.4	6.00
6	10.67	0.1	0.8	4.80

IV. EXPERIMENTAL EVALUATIONS

A. *Description of Experiment*

(1) The basis of experiment

The experiments use the customer reviews about woman's shoe. The data are collected from the C2C website which is famous in China. Products on this website have a large number of reviews. Each of the review includes text content, and additional information available used in this project, and includes product's rating, review's time. We set the weight of timeliness is 0.5 in the experiments. In order to examine the efficiency of the method, we collect the reviews of woman's shoe from two sellers. By analyzing the properties and reviews related to product features, we summarize the main features *S* that buyers interested in as follows:

{material, quality, sole, heel-height, durable, size, style, color, decoration, Softness, Comfort, anti-skidding, workmanship, smell, brand, price, packaging, shipping, Service}

In order to prove the validity of the model, we select different sizes of theses reviews as a valid set *H*. The coverage *r* can reflect how many features of the valid review are included in main features set *S*. If supposing that the features in *H* defined as *F*, all features defined as *S*, so the *r* can be defined as:

$$r = \frac{|F \cap S|}{S} \tag{7}$$

We define the anti-fraud rate af to verify the ability to inhibit false feedback of the method,. If add some untrue reviews with number *T* and their related ratings, *N* reviews are identified by review's value, then the definition of af is as follow:

$$af = \frac{T - N}{T} \tag{8}$$

(2) Related method

Two compared models are selected in the experiment. The first is time-based model that sort the review by published time, and use TSM for short. The second is user votes based evaluation model, and use VSM for short. VSM model also reflect the common perception of the public, which it is a more accurate and fair evaluation. VSM consists by helpful votes and the total votes, so the review's value is defined as:

$$v_i = \begin{cases} \dfrac{HV_i}{TV} & HV_i > 0 \\ 0 & HV_i = 0 \end{cases} \tag{9}$$

B. Simulation Analysis

We collected two review sets. The first set has 153 reviews, and the second set has 135 reviews. The validity of review's life is one year. We get 102 reviews and 90 reviews after filtering out expired reviews and meaningless reviews. The coverage rate *r* and anti-fraud rate af are computed.

(1) The coverage rate

RVM is the method presented in this paper. Reviews are sorted by TSM, VSM and RVM in the experiment. The previous *N* reviews are selected to compute the coverage rate. From the Fig.4, the overall trend of the two sets of reviews, the coverage rate is increasing with the review amount. The phenomenon as we expect appears that more reviews will reflect more concerned features. TSM that is based on time can reflect most timely product characteristics, and with high timeliness. But the TSM doesn't consider the review content, and may not include product features, which will induce TSM with lower coverage.

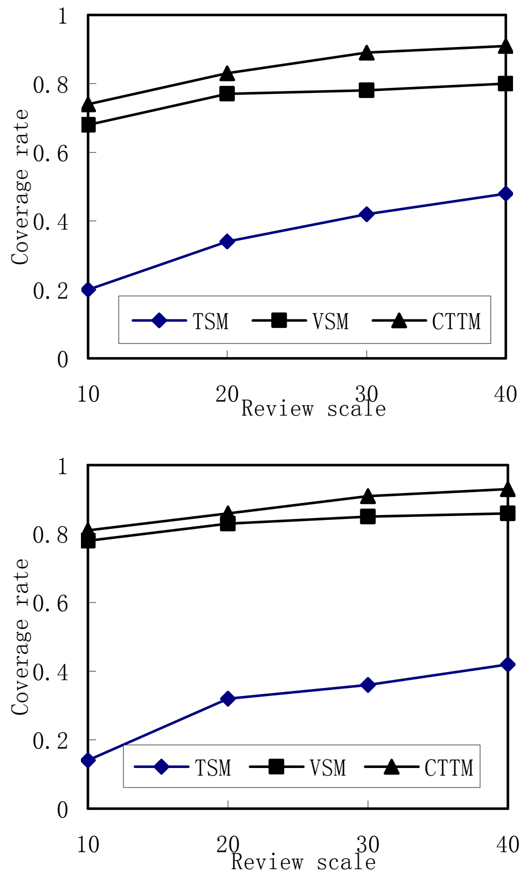
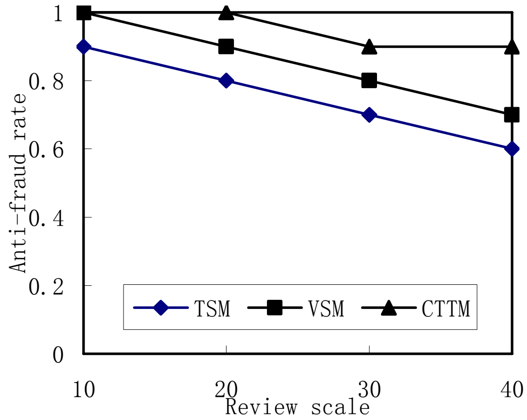


Figure 4. Coverage Rate of each group

From the results of two group review set, none of the rate beyond 50%. It also brings more contingency due to the subjectivity of the reviewer. VSM sort the reviews by voting score. It can largely reflect the product's features. The review with lots votes should include more information. So some reviews selected by VSM can achieve higher coverage rate. However, it has two disadvantages of computing the review's value. The first is some new reviews which may contain more valuable information will not get voting in time. Because few of people would vote these new reviews, and it must take some time to accumulate votes. Secondly, because the buyers may have not enough enthusiasm in voting, and it induces only a few reviews will get votes. Some reviews with none voting will be sorted by time in the experiment. The coverage rate of VSM isn't significantly improved with increasing the amount of review. In the RVM model, the value score is computed by product information, timeliness and reliability. These reviews with high scores in RVM will have a lot of information compared to other reviews, and include more features. The RVM will compute the value of every review. With the amount of review is increasing, more valuable reviews will be selected by RVM, so the coverage rate will be improved significantly.

(2) Anti-fraud rate

We add 10 fraud reviews into the two sets. The ratings of fraud reviews largely deviate from the average score, and have a uniform time span. Because these fraud reviews may hope to attract more attention to achieve their personal goals, the coverage of these reviews is set randomly between 50% and 80%.



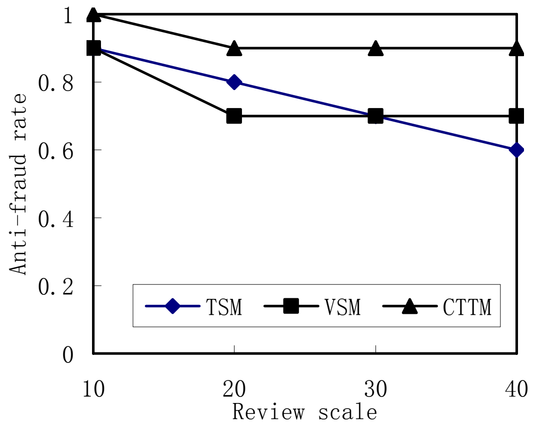


Figure 5. Anti-fraud rate of each group

As shown in the Fig.5, these fraud reviews have an impact on all models with the increasing of review scale. In the TSM model, the anti-fraud rate will decrease while increasing the reviews scale. It is because these fraud reviews are uniformly distributed in the time span. If the reviews are sorted by time, more fraud reviews will be selected while increasing the review scale and will deduce the anti-fraud rate of TSM. In the VSM model, ten buyers will vote these reviews to update voting scores. The buyers generally pay more attention to reviews' content, and have a big possibility to vote these fraud reviews for the fraud review often have rich content to attract more attention. So the anti-fraud rate of VSM is decreased with review scale increasing. In the RVM model, the statistical regularities are used to analyze the differences of feedback scores. The review has a big deviation from the average feedback score will indicate that it is deviated from majority of buyers and is likely to be a fraud feedback. The RVM is possible to get a high anti-fraud rate in theoretically. But taking into account the time and the content in comprehensively, some reviews with high timeliness and containing a lot of information will also get high review value. By increasing the review scale, the anti-fraud rate may decline in a certain way.

V. Conclusion

The proposed method in our work pays more attention to the semantic association among reviews, and uses the network view to describe the complex relationship among reviews. The information quality problem is changed into the importance of a node in the network. The work also takes the timeliness, reliability into considering. The reasonable and effective method is designed to measure the two factors in review's value computation. Future work involves testing the proposed mechanisms based on a larger set of reviews to improve the efficiency of the method. In addition, the factors which have influence on review's value should be refined to find the optimal set of features to predict the value of the review.

Acknowledgment

This work was supported by Youth fund projects and National Science Foundation, from Science and Technology Agency of Jiangxi Province, NO. 2012ZBAB201003, NO. 20132BAB201055.

References

[1] Chevalier, J. A., & Mayzlin, D. The effect of word of mouth on sales: Online book reviews. Journal of marketing research, vol.43, No.3, pp.345-354, 2006.

[2] Liu, L., Lv, Z., & Wang, H. Extract Product Features in Chinese Web for Opinion Mining. Journal of Software, vol.8, no.3, pp. 627-632, 2013.

[3] Brynjolfsson, E., & Smith, M. D. Frictionless commerce? A comparison of Internet and conventional retailers. Management Science, vol.46 , no.4, 563-585, 2000.

[4] Balahur, A., & Montoyo, A. Multilingual feature-driven opinion extraction and summarization from customer reviews. In Natural Language and Information Systems (pp. 345-346). Springer Berlin Heidelberg, 2008.

[5] Zhang, Z. Weighing stars: Aggregating online product reviews for intelligent e-commerce applications. Intelligent Systems, IEEE, vol.23, no.5, pp.42-49, 2008.

[6] Duan, W., Gu, B., & Whinston, A. B. Do online reviews matter?—An empirical investigation of panel data. Decision Support Systems, vol.45, no.4, pp.1007-1016, 2008.

[7] Chatterjee, P. Online reviews: do consumers use them?. Advances in consumer research, vol.28, no.1, pp.129-133, 2001.

[8] Ghose, A., & Ipeirotis, P. G. Designing novel review ranking systems: predicting the usefulness and impact of reviews. In Proceedings of the ninth international conference on Electronic commerce (pp. 303-310), ACM, August, 2007.

[9] Chen, L., Qi, L., & Wang, F. Comparison of feature-level learning methods for mining online consumer reviews. Expert Systems with Applications, vol.39, no.10, pp.9588-9601, 2012.

[10] Kim, S. M., Pantel, P., Chklovski, T., & Pennacchiotti, M. Automatically assessing review helpfulness. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 423-430). Association for Computational Linguistics, July, 2006.

[11] Liu, Y., Huang, X., An, A., & Yu, X. Modeling and predicting the helpfulness of online reviews. In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on (pp. 443-452). IEEE, December, 2008.
[12] Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. Low-Quality Product Review Detection in Opinion Summarization. In EMNLP-CoNLL, pp. 334-342, 2007.
[13] Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. How opinions are received by online communities: a case study on amazon.com helpfulness votes. In Proceedings of the 18th international conference on World wide web, pp. 141-150, ACM, April , 2009.
[14] Xu, Y. A Data-drive Feature Selection Method in Text Categorization. Journal of Software, vol.6, no.4, pp. 620-627, 2011.
[15] Turney, P. D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics, 2002, July.

Leiyue Yao, is a senior engineer in information system development. He was born in July 1982 in Jiangxi in China. His major is computer application technology. He received his master’s degree in NanChang University in 2006. His research interests are in the pervasive computing, Mobile Commerce, system optimization.

Wei Yang is a Chief software Architect. He received his bachelor's degree in 2010 and major in network technology. His research interests are in information management system, Service Computing.