Postprint: Comment Quality Detection Based on Text Content Feature Selection

Authors: Meng Yuan, Wang Hongwei

Date: 2017-10-11T00:00:00+00:00

Abstract

Purpose: On the basis of effectively extracting multi-dimensional features, this study investigates the influence of review content features on review quality detection. Method: Based on the hybrid nature of information feature measurement and sentiment orientation in review text, we quantify and extract review content features, employ the GBDT model to evaluate the classification effectiveness of feature sets, combine it with a greedy feature selection algorithm to identify effective content features, and analyze their influence on review quality detection. Results: Applying review content features to the review quality detection task achieves favorable results, significantly improving experimental accuracy and recall rates. Limitations: The experimental subjects are primarily review data for search products, and no validation or comparison has been conducted for other experiential products (such as movies, music, etc.). Conclusion: Information gain of review content, information gain of product feature words, review objective sentiment orientation, and content difference have a significant effect on review quality detection.

Full Text

Review Quality Detection Based on Text Content Feature Selection

Meng Yuan, Wang Hongwei
School of Economics and Management, Tongji University, Shanghai 210000

Abstract

[**Objective**] This study examines the impact of review content features on review quality detection through effective multi-dimensional feature extraction. [**Methods**] Based on the hybrid nature of textual information characteristics and sentiment orientation in reviews, we quantified and extracted review content features. The Gradient Boosting Decision Tree (GBDT) model was employed to evaluate the classification effectiveness of feature sets, while a greedy feature selection algorithm identified effective content features and analyzed their influence on review quality detection. [**Results**] Applying review content features to quality detection tasks achieved favorable results, significantly improving experimental accuracy and recall rates. [**Limitations**] The experiments primarily focused on review data for search-type products, without validation or comparison for other experiential products (such as movies or music). [**Conclusions**] Information gain from reviews, information gain from product feature terms, degree of objective sentiment orientation, and content differences all demonstrate significant effects on review quality detection.

Keywords: Review quality; Information feature; Sentiment orientation; Content feature; Greedy feature selection

---

With the increasing maturity of internet technology, consumers' enthusiasm for online reviews has grown, generating massive volumes of review data online. While users leverage these reviews to inform purchase decisions, they also suffer from issues such as inconsistent review quality and information overload. Manual methods alone cannot effectively identify truly valuable information from the vast sea of reviews, necessitating automated approaches to assist in screening. Consequently, detecting the quality of online reviews holds important research value.

Some e-commerce platforms rank review quality through "helpfulness voting." Based on this, scholars generally consider consumer-perceived helpfulness as a measure of review quality or utility, where higher helpfulness indicates higher quality. Thus, review quality, utility, and helpfulness are typically treated as equivalent concepts [1-2]. Existing literature on review quality detection employs two main approaches: econometric regression methods and supervised learning methods. The former typically uses metadata features (e.g., review ratings, reviewer identity) or linguistic features (e.g., word count, number of terms) as independent variables and helpfulness voting ratio as the dependent variable to examine which features significantly impact review quality. The latter treats review quality detection as a classification problem, establishing helpfulness voting thresholds or manual annotation to generate training sets for useful reviews, then employing optimal models to automatically identify high-quality reviews with relatively better results. Since review quality is influenced by multiple features, selecting effective features is key to detection. Current domestic research on helpful review feature selection has focused on metadata and linguistic features [2-6], with insufficient depth in mining textual content features and limited analysis of feature contribution and selection mechanisms.

This paper employs the Gradient Boosting Decision Tree (GBDT) model as the classification framework. Based on multi-dimensional feature extraction, we specifically examine the performance of review content information features and semantic sentiment features in classification models, further utilizing a greedy feature selection algorithm to identify effective content feature sets and reveal the impact effects of multi-dimensional review features.

2.1 Feature Analysis Affecting Review Quality

Existing research categorizes features affecting review quality into three major types: metadata features, linguistic features, and review content features. Yang et al. [2] noted that metadata features are independent of textual content and language characteristics, with review ratings, helpful vote counts, and total vote counts being important metadata features. Kim et al. [7] demonstrated that time elapsed since review publication significantly affects quality. Ghose et al. [8] considered reviewer-related information as effective metadata features, such as the number of previously published reviews, reviewer helpfulness rate, and reviewer identity.

Linguistic features primarily involve identifying review characteristics from word frequency statistics. Ghose et al. [8], Li et al. [9], and Liu et al. [10] indicated that key linguistic features include review length, sentence count, and counts of different parts of speech (nouns, verbs, adjectives, etc.). Chen et al. [11] emphasized that among nouns in reviews, the frequency of product attribute nouns constitutes an important linguistic feature, with high-quality reviews containing a certain number of product attribute terms. These studies reveal that metadata and linguistic features represent extrinsic review characteristics. Correspondingly, intrinsic features are based on review text content. After reading reviews, consumers can understand other users' positive or negative evaluations of products, thereby acquiring a certain amount of information to reduce uncertainty about product cognition. Wang et al. [12] noted that this information obtained from review content truly influences consumers' purchase intentions, while Nie et al. [1] validated that review sentiment features effectively assess review utility. Thus, intrinsic review features serve as important criteria for consumers' quality judgments.

2.2 Review Quality Detection Methods

Existing research primarily employs econometric and supervised learning methods for review quality detection. Econometric studies generally use helpfulness voting ratio as a proxy variable for review quality, where higher ratios indicate higher quality. For instance, Ghose et al. [8] employed multiple linear regression on DVD product review data, concluding that reviewer characteristics and linguistic features significantly and positively affect quality. Similar econometric approaches in literature [5-6] identified review length, ratings, and other features as quality influencers. Conversely, supervised methods treat review quality detection as a classification problem, establishing training sets through manual annotation or helpfulness voting thresholds, then using extracted feature sets to test and evaluate classifier performance to identify effective review features for automatic high-quality review identification. For example, Nie et al. [1] used helpfulness voting ratio as a quality proxy, setting reasonable thresholds to generate training sets and employing random forests for quality detection. Liu et al. [10] used manual annotation to obtain training sets and compared machine learning methods including support vector regression and decision trees to identify optimal classification models. Chen et al. [11] constructed a multi-layer support vector machine for review quality classification.

2.3 Literature Review and Problem Definition

Feature selection is critical for review quality detection. Existing research has primarily focused on extrinsic features like linguistic and metadata characteristics. Although few scholars have specifically validated the role of textual sentiment features, comprehensive examination of both extrinsic and intrinsic features' impact on review quality remains rare. Do review content features significantly affect quality? Does feature selection order influence classification performance? These questions require answers. Addressing these gaps, this study aims to employ GBDT supervised learning to thoroughly and comprehensively mine effective feature sets affecting review quality and examine the impact of review content features.

The research process follows three aspects: (1) Extract intrinsic review content features, including information features and semantic sentiment features; (2) Employ GBDT classification and greedy feature selection algorithms to identify effective feature sets and optimal classification models; (3) Conduct performance evaluation and comparison of classification models.

This paper models review quality detection as a binary classification problem. Based on effective extraction of multi-dimensional textual features, we employ the Gradient Boosting Decision Tree (GBDT) model and greedy feature selection algorithm for optimal model identification. The GBDT model combines "base learners" through multiple iterations, establishing decision tree models in the gradient descent direction of the loss function during each iteration to minimize the cumulative loss function. This iterative improvement yields superior classification performance compared to base learners, demonstrating excellent performance in classification and regression tasks [13-14]. The main research tasks include experimental review selection, sentence segmentation, feature extraction, feature selection, model training and identification, and experimental result analysis. The research framework is shown in Figure 1 [Figure 1: see original paper].

3.1 Text Content Feature Extraction

This study focuses on examining the effectiveness of review content features, emphasizing the extraction methods for content-related features. We extract eight features from two aspects: information characteristics and semantic characteristics embedded in review text, as summarized in Table 1 .

(1) Information Feature Extraction

Information Content of Reviews

From an information theory perspective, the greater the information content a review r contains, the more useful it is to users. Different words in a review contribute varying amounts of information to usefulness. Therefore, this paper employs information gain of words to quantify the information content of review r.

For a given online review set R, let $C = (c_1, c_2)$ represent the review helpfulness categories, where c_1 denotes the helpful category and c_2 denotes the unhelpful category. The total information entropy $H(C)$ required for the binary classification system is:

$$H(C) = -\sum_{i=1}^{2} P(c_i) \log P(c_i)$$

where $Pr(c_i)$ is the occurrence probability of category c_i in the system.

Review r consists of multiple different words. Considering a word t in the review with two possible values—presence or absence, represented by w and $\bar{w}$ respectively—the system entropy $H(C|w)$ under the condition that t appears (i.e., t takes value w) is:

$$H(C|w) = -\sum_{i=1}^{2} P(c_i|w) \log P(c_i|w)$$

where $P(c_i|w)$ is the probability of category c_i occurring in reviews where t appears.

Similarly, we can obtain the system entropy $H(C|\bar{w})$ under the condition that t does not appear (i.e., t takes value $\bar{w}$). Considering the information increment brought to the system under these two different value conditions of t, the information gain $G(t)$ of t is:

$$G(t) = H(C) - P(w)H(C|w) - P(\bar{w})H(C|\bar{w})$$

where $P(w)$ represents the appearance probability of t, and $P(\bar{w})$ represents the non-appearance probability of t.

Since $G(t)$ considers the sum of word t' s contribution across both categories (c_1 and c_2), and given that helpful reviews assist users in reducing product uncertainty while unhelpful reviews may hinder correct product judgment, the information gain direction of word t differs across helpful categories. To better reflect the difference in information gain of t between the two categories, we adapt the information gain of word t from literature [15] by comparing the occurrence probabilities of helpful and unhelpful categories among all reviews containing t, i.e., $P(c_1|w)$ and $P(c_2|w)$. If the former is larger, word t represents positive information gain; otherwise, it represents negative information gain. The improved information gain $IG(t)$ of word t is expressed as:

$$IG(t) = \begin{cases} P(c_1|w) - P(c_2|w) & \text{if } P(c_1|w) > P(c_2|w) \\ P(c_2|w) - P(c_1|w) & \text{otherwise} \end{cases}$$

Thus, the information content $IGain(r)$ of review r is represented as the sum of information gains of all words in r, as shown in the following formula:

$$IGain(r) = \sum_{t \in r} IG(t)$$

Literature [12] indicates that product feature words in reviews contribute more significantly to users' quality judgment than other words. Therefore, we examine the information gain $IG(f)$ provided by each product feature word f and extract the information content contributed by all feature words in each review $IGain_f$, calculated as:

$$IGain_f(r) = \sum_{f \in r} IG(f)$$

Content Difference of Reviews

Literature [16] notes that reviews with more similar content are more likely to be fake reviews, reflecting that content differences among reviews affect users' quality perception. Based on text information entropy, Jelinek defined the concept of perplexity, which together can measure the content difference between a review and other reviews [17]. For review set R, if review r differs more from other reviews in content, its information entropy and perplexity values will be larger.

Assuming review r consists of a sequence of words $w_1, w_2, \ldots, w_n$ in a specific order, where $p(w_i)$ is the occurrence probability of word w_i in r, the information entropy $Entropy(r)$ and perplexity $Perplexity(r)$ of the review are expressed as:

$$Entropy(r) = -\sum_i p(w_i) \log p(w_i)$$

$$Perplexity(r) = 2^{Entropy(r)}$$

This paper uses each product model' s corresponding review subset as training corpus to build a unigram statistical language model, then employs the trained model to calculate the information entropy and perplexity of each review within the subset.

(2) Semantic Sentiment Feature Extraction
Reviews often present mixed opinions, containing both positive/negative sentiments and subjective/objective expressions. Typically, positive/negative sentiment orientation is determined by the polarity of opinion words, while subjective/objective orientation is measured by the consistency between the reviewer' s product comments and merchant descriptions [18]—the more similar the review text is to merchant descriptions, the more formal the language and the more objective the review tends to be.

For example, the review sentence: "This product's performance is quite superior, and its appearance is also very compact and beautiful. The cost-performance ratio is average because the price is a bit high. Overall, I really like it!" expresses both positive and negative sentiment orientations, but the overall sentiment is positive. Analyzing subjectivity, the first two sentences are more objective compared to the last one. To comprehensively measure the impact of sentiment mixing on review quality, we define four feature items: objective sentiment orientation degree $ObjDegree$ and its deviation $DevObj$, positive sentiment orientation degree $PosDegree$ and its deviation $DevPos$.

Objective Sentiment Orientation Degree and Deviation
Using review clauses as units, we examine the cosine similarity between review content and product description text to determine objectivity. Representing review clause s and product description d as vectors using tf-idf weights, we calculate their cosine similarity $sim(s,d)$ and set threshold λ to judge clause objectivity. Let s^+ denote objective clauses in r, and $total(r)$ be the total number of clauses in review r. The objective sentiment orientation degree is calculated as:

$$ObjDegree(r) = \frac{count(s^+)}{total(r)}$$

For all reviews of product p, the average objective sentiment orientation degree uniformly reflects the stable proportion of overall subjective/objective clauses. Comparing review r' s objective sentiment orientation degree with the overall mean, larger deviation indicates more consistent viewpoints in r (all objective or all subjective), while smaller deviation suggests mixed subjective/objective viewpoints. Therefore, based on the average objective sentiment orientation degree of all reviews for product p, we define the objective sentiment orientation deviation $DevObj(r)$ to represent the mixing degree of subjective/objective sentiment:

$$DevObj(r) = |ObjDegree(r) - Avg(ObjDegree(r))|$$

Positive Sentiment Orientation Degree and Deviation
Reviews generally consist of multiple opinion clauses. The proportion of positive

clauses represents the review’ s positive sentiment orientation degree—larger proportion indicates more positive overall sentiment, and vice versa. Thus, positive sentiment orientation degree expresses the sentiment polarity feature of reviews. Using clauses as units, we determine sentiment polarity. This paper employs machine learning methods for clause sentiment classification, selecting 1,000 reviews each with 5-star and 1-star ratings to build a sentiment classifier. Based on chi-square statistics, we select the top 1,500 unigrams and bigrams as text sentiment polarity classification features [1], using BernoulliNB (which achieved the best classification performance in Python) to discriminate positive/negative sentiment polarity of review clauses. Let r^+ denote positive clauses in r, and $total(r)$ be the total number of clauses. The positive sentiment orientation degree is calculated as:

$$PosDegree(r) = \frac{count(r^+)}{total(r)}$$

Similarly, the positive sentiment orientation deviation measures the mixing degree of positive/negative sentiment in review r:

$$DevPos(r) = |PosDegree(r) - Avg(PosDegree(r))|$$

3.2 Feature Selection

(1) Basic Feature Template

Since review quality is closely related to metadata features (Meta) and linguistic features (Lan), we construct a classification model feature template using these as the basic feature set. Based on conclusions from literature [3,5,7-8,10-11], we extract six effective metadata features (M1-M6) and three linguistic features (L1-L3), as shown in Table 2 :

Table 2. Basic Feature Set

Feature	Description
M1	Review helpfulness voting ratio: helpful votes divided by total votes
M2	Number of helpful votes received by review r
M3	User rating corresponding to review r
M4	Time elapsed since review r publication
M5	Reviewer rank corresponding to review r
M6	Average helpfulness rate of reviewer’ s previous reviews
L1	Number of characters in review r
L2	Number of words in review r
L3	Number of product feature words in review r

(2) Greedy Feature Selection Algorithm
To select feature sets beneficial for review quality detection from the extracted text content features, we use metadata and linguistic features as the basic feature set and extracted content features as candidate features. We employ a greedy feature selection algorithm [19] with the GBDT classification model for feature selection. The main idea is: based on each candidate content feature' s contribution to classification performance on the development set DevData, we iteratively select the feature with maximum contribution to add to the basic feature set. The algorithm terminates when adding any remaining candidate feature causes classification evaluation metrics on DevData to decline or when the candidate feature set becomes empty.

The algorithm execution process is as follows:

Input: All feature set $F_{all} = \{M1 \sim M6, L1 \sim L3, I1 \sim I4, S1 \sim S4\}$
Output: Effective feature set $F_{select} = \{\text{set of selected features}\}$, $M_{select} = \{\text{selected model}\}$

1. Initialize basic feature set and candidate content feature set: $F_{select} = \{M1 \sim M6, L1 \sim L3\}$, $F_{can} = F_{all} - F_{select}$
2. Train model to obtain initial classification performance: $M_{select} = GBDT\_Train(F_{select})$, $E_{select} = Evaluate(M_{select}, DevData)$
3. Perform text content feature selection
4. **loop**
5. **for each** feature f_i in F_{can} **do**
6. `$F_i = F_{select} \cup f_i$`
7. `$M_i = GBDT\_Train(F_i)$`
8. `$E_i = Evaluate(M_i, DevData)$`
9. **end for**
10. $E_{max} = Max(E_i)$
11. **if** $E_{max} > E_{select}$ **then**
12. `$f_{max} = \arg\max_{f_i}(E_i)$`
13. `$F_{select} = F_{select} \cup f_{max}$`
14. `$M_{select} = M_{max}$`
15. `$F_{can} = F_{can} - f_{max}$`
16. **end if**
17. **if** $F_{can} = \emptyset$ or $E_{max} \leq E_{select}$ **then**
18. `**return** $F_{select}, M_{select}$`

19. **end if**
20. **end loop**

In the algorithm, lines 5-9 select one feature f_i from candidate set F_{can} each time, add it to effective set F_{select}, execute the classification model, and record its corresponding classification metric E_i. Lines 10-16 compare the classification metrics corresponding to each feature f_i to determine the feature with maximum contribution and its addition order to the effective feature set.

4.1 Experimental Data and Annotation Standards

We used a crawler program to collect relevant review information and product data for digital cameras from Chinese Amazon, including review text, metadata, and product descriptions. Data collection ended on September 2, 2013, covering review publication dates from January 7, 2009, to September 1, 2013, totaling 15,327 reviews. We selected products with more than 50 reviews as experimental objects. After preprocessing to remove duplicates and advertisements, we obtained 10,568 valid reviews covering 10 camera brands and 67 product models. Detailed statistics are shown in Table 3 .

Table 3. Review Data Characteristics Statistics

Review Attribute	Minimum	Maximum	Average
Number of words			
Number of product feature words			
Number of helpful votes			
Helpfulness ratio			
Review publication time (log)			
Reviewer rank (log)			
Average helpfulness rate of reviewer' s previous reviews			

Given the large deviation in helpfulness voting ratios, we adopted manual annotation following literature [20]. Two experienced digital product users were invited to independently annotate the experimental data. Annotators read each review and answered the question: "Is this review content useful for understanding or purchasing the product?" In addition to review text, brief product descriptions were provided. Cohen-Kappa testing of annotation results yielded a Kappa value of 83.45%, indicating high consistency in quality evaluation standards between annotators. Annotator 1' s results were used to train and test the model, ultimately obtaining 5,307 high-quality reviews and 5,261 low-quality reviews.

4.2 Model and Evaluation Metrics

We employed the GBDT model for review quality classification. Testing revealed that building 25 tree models achieved optimal classification performance.

Experimental data were split 4:1 into training and test sets for feature selection. Average precision, average recall, and average F1-score were used as evaluation metrics, denoted as P, R, and $F1$ respectively.

Experiments were conducted on a Win7 32-bit operating system with 4GB RAM. All programs were written in Python 2.7.3. Below are partial text content feature extraction results.

(1) Word Information Gain
Using the review quality annotation results from experimental data, we performed word segmentation and stop-word removal on reviews, then calculated information gain for words in the review set (considering only words with parts of speech n, v, a, d, vn). A total of 11,729 effective words were obtained. Partial calculation results are shown in Table 4 .

Table 4. Partial Word Information Gain Calculation Results

WordID	Word	Positive Information Gain	WordID	Word	Negative Information Gain
20	帮别人				

(2) Partial Feature Extraction Results
After word segmentation and sentence segmentation of experimental corpus, features were extracted according to the method described in Section 3.1. Partial review content feature extraction results are shown in Table 5 .

Table 5. Partial Content Feature Extraction Results

ReviewID	IGain	IGainf	Entropy	Perplexity	ObjDegree	DevObj	PosDegree	DevPos

5 Classification Experimental Results

(1) Effectiveness of Text Content Features
Using metadata and linguistic features as the basic feature set, we then added text content features sequentially. Table 6 shows the effects of adding individual text content features.

Table 6. Effects of Adding Individual Text Content Features

Feature Set	F1(%)
Meta+Lan	
Meta+Lan+I1	
Meta+Lan+I2	
Meta+Lan+S1	

Feature Set	F1(%)
Meta+Lan+I3	
Meta+Lan+I4	

The results demonstrate that adding individual content features improves review quality classification accuracy and recall to some extent, with F1 increasing by nearly 3 percentage points, validating the effectiveness of content features for quality detection. However, classification metrics show an initial rise followed by decline when features are added sequentially, indicating that some features have limited effect. Therefore, feature selection is necessary to remove unhelpful features. According to the feature selection algorithm in Section 3.2, the optimal classification performance is achieved by sequentially combining features in the order {I1, I2, S1, I3, I4}.

Table 7 shows the classification performance of feature combinations selected by the greedy feature selection algorithm. Therefore, the feature combination {I1, I2, S1, I3, I4} and its order constitute the final effective review content feature set.

Table 7. Greedy Feature Selection Results

Feature Set	F1(%)
Meta+Lan	
Meta+Lan+I1	
Meta+Lan+I1+I2	
Meta+Lan+I1+I2+S1	
Meta+Lan+I1+I2+S1+I3	
Meta+Lan+I1+I2+S1+I3+I4	

The results show that among the filtered features, only objective sentiment orientation degree (S1) is an effective sentiment feature, indicating that formal, objective review content affects quality. All information features demonstrate significant effects on quality, with the largest contributions from overall review information content (I1) and information content from product feature words (I2), followed by information entropy (I3) and perplexity (I4) measuring content differences. The information content of entire reviews (I1) helps users understand product information, while product feature words provide more valuable information for quality judgment. Content differences play a crucial role in evaluating review quality, indirectly validating literature [16]' s conclusion that more similar reviews are more likely to be spam.

(2) Model Comparison Based on Effective Feature Set

To examine the classification performance of the GBDT model, we compared it with the Random Forest (RF) model used by Ghose et al. [8] and the basic

Decision Tree (DT) model to evaluate gradient boosting optimization effects. Experimental comparison results are shown in Figure 2 [Figure 2: see original paper]. The results indicate that the GBDT method significantly improves both accuracy and recall compared to RF and DT models. Overall, F1 improves by approximately 9 percentage points compared to DT, demonstrating good optimization effects, and by about 2.3 percentage points compared to RF, indicating strong model performance.

This paper introduces the application effects of textual information features and semantic sentiment features in review quality detection. The results show that after greedy feature selection with a specific feature order, GBDT achieves optimal classification performance on the selected feature set, outperforming decision tree and random forest models. This validates the effectiveness of feature extraction and selection, thereby more effectively assisting merchants in automatically identifying high-quality reviews.

Future work will continue to explore other effective content features to further improve and perfect the application of textual features in review quality monitoring.

References

[1] Nie Hui, Rong Zhe. Review Helpfulness Prediction Research Based on Review Sentiment Feature Sets [J]. New Technology of Library and Information Service, 2015(7-8): 104-112.

[2] Yang Ming, Qi Wei, Yan Xiangbin, et al. Utility Analysis for Online Product Review [J]. Journal of Management Science in China, 2012, 15(5): 65-75.

[3] Gao Ya, Li Hong, Shi Huibin. The Research of the Impact Factors of the Online Review Votes [J]. China Management Informationization, 2012, 15(17): 88-91.

[4] Yan Jianyuan, Zhang Li, Zhang Lei. An Empirical Study of the Impact of Review Content on Online Reviews Helpfulness in E-Commerce [J]. Information Science, 2012, 30(5): 713-719.

[5] Yang Shuang. The Impact Mechanism of Information Quality and Community Status on Perceived Usefulness for User-Generated Product Reviews——Tobit Regression Analysis [J]. Management Review, 2013, 25(5): 136-143, 154.

[6] Yin Guopeng. What is the Kind of Online Reviews that Consumer Think are More Useful? The Effect of Social Factors Influence [J]. Management World, 2012(12): 115-124.

[7] Kim S M, Pantel P, Chklovski T, et al. Automatically Assessing Review Helpfulness [C]. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), Sydney, Australia. Stroudsburg, PA, USA: ACL, 2006: 423-430.

[8] Ghose A, Ipeirotis P G. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(10): 1498-1512.

[9] Li F, Zhang Y L, Dang Y, et al. Analyzing Sentiments in Web2.0 Social Media Data in Chinese: Experiments on Business and Marketing Related Chinese Web Forums [J]. Information Technology Management, 2013(14): 231-242.

[10] Liu Y, Jin J, Ji P, et al. Identifying Helpful Online Reviews: A Product Designer' s Perspective [J]. Computer-Aided Design, 2013, 45(2): 180-194.

[11] Chen C C, Tseng Y-D. Quality Evaluation of Product Reviews Using an Information Quality Framework [J]. Decision Support Systems, 2011, 50(4): 755-768.

[12] Wang Wei, Wang Hongwei. The Influence of Aspect-based Opinions on User' s Purchase Intention Using Sentiment Analysis of Online Reviews [J]. Systems Engineering—Theory & Practice, 2016, 36(1): 63-76.

[13] Ayaru L, Ypsilantis P-P, Nanapragasam A, et al. Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting [J]. PLoS ONE, 2015, 10(7). DOI: 10.1371/journal.pone.0132485.

[14] Semanjski I, Gautama S. Smart City Mobility Application-Gradient Boosting Trees for Mobility Prediction and Analysis Based on Crowd Sourced Data [J]. Sensors, 2015, 15(7): 16658-16683.

[15] Zhang R, Tran T. An Information Gain-based Approach for Recommending Useful Product Reviews [J]. Knowledge and Information Systems, 2011, 26(3): 419-434.

[16] Jindal N, Liu B. Review Spam Detection [C]. In: Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada. New York, NY, USA: ACM, 2007: 1189-1190.

[17] Wu Jun. The Beauty of Mathematics [M]. Beijing: Posts and Telecom Press, 2012: 60-64.

[18] Pang B, Lee L. A Sentiment Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts [C]. In: Proceeding of the 42nd Annual Meeting of the Association for Computational Linguistic (ACL). Morristown, NJ, USA: ACL, 2004: 271-278.

[19] Jiang Z P, Ng H T. Semantic Role Labeling of NomBank: A Maximum Entropy Approach [C]. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA, USA: ACL, 2006: 419-426.

[20] Liu J, Cao Y, Lin C-Y, et al. Low-Quality Product Review Detection in Opinion Summarization [C]. In: Proceedings of the 2007 Joint Conference on EMNLP-CoNLL. ACL Press, 2007: 334-342.

Author Contributions

Meng Yuan: Data collection, cleaning, and analysis; manuscript drafting and final revision.
Wang Hongwei: Topic formulation, research idea proposal, study design, and paper refinement.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data is available in the journal's online version at http://www.infotech.ac.cn.

[1] Meng Yuan, Wang Hongwei. quality_classify.py. Greedy feature selection and model comparison algorithm for review quality.
[2] Meng Yuan, Wang Hongwei. exp_data_initial.xlsx. Raw Amazon crawled data.
[3] Meng Yuan, Wang Hongwei. classifier.pkl. Text sentiment polarity classifier.
[4] Meng Yuan, Wang Hongwei. segged_pos_words.txt. Review word segmentation data.
[5] Meng Yuan, Wang Hongwei. sub_sentence.txt. Review sentence segmentation data.
[6] Meng Yuan, Wang Hongwei. Refine_Igain_word.txt. Improved review word information gain.
[7] Meng Yuan, Wang Hongwei. all_feature_value.txt. Review feature extraction data.

Received: December 9, 2015
Revised: February 7, 2016

Note: Figure translations are in progress. See original paper for figures.