Research on Sentiment Orientation of Hot Topics Based on Topic Relevance (Postprint)

Authors: He Yue, Xiao Min, Zhang Yue

Date: 2017-11-08T00:00:00+00:00

Abstract

【Objective】Hot topics exert substantial influence. This study investigates the sentiment orientation toward hot topics and their associated sentiment objects.

【Method】We propose a subjectivity classification model incorporating topic relevance to extract subjective microblogs related to hot topics; employ an improved sentiment classification method based on machine learning to analyze the sentiment polarity of extracted posts; and comprehensively evaluate sentiment classification performance using recall, precision, and F-value.

【Results】Empirical analysis demonstrates that incorporating topic relevance effectively enhances both subjectivity classification and sentiment polarity classification for hot-topic microblogs, achieving F-value improvements of 7.4% and 2.2%, respectively.

【Limitations】Further research should thoroughly consider data distribution characteristics, refinement of sentiment classification granularity, and temporal changes in sentiment trends of sentiment objects.

【Conclusion】By incorporating topic relevance to improve microblog sentiment classification effectiveness and extracting sentiment orientations of key sentiment objects within hot topics, this study provides relevant intelligence for precision marketing on microblogging platforms.

Full Text

Sentiment Analysis of Trending Topics Based on Relevance

He Yue, Xiao Min, Zhang Yue
Business School, Sichuan University, Chengdu 610064, China

Abstract

[**Objective**] Trending topics exert significant influence, and this study investigates the sentiment orientation of trending topics and their emotional objects. [**Methods**] We propose a subjective-objective classification model that incorporates topic relevance to extract subjective microblog posts related to trending topics. We then analyze the sentiment polarity of extracted posts using an improved sentiment classification method based on machine learning. The classification performance is evaluated in detail using recall, precision, and F-measure. [**Results**] Empirical analysis demonstrates that incorporating topic relevance effectively improves both the subjective-objective classification and sentiment polarity classification of trending topic microblogs, with F-measures increasing by 7.4% and 2.2% respectively. [**Limitations**] Further research is needed to consider data distribution states, refine sentiment classification granularity, and analyze emotional trend changes of sentiment objects. [**Conclusions**] Considering topic relevance enhances microblog sentiment classification effectiveness. By extracting sentiment tendencies of key emotional objects within trending topics, this approach provides relevant intelligence for precision microblog marketing.

Keywords: Trending Topic; Subjective-Objective Classification; Emotion Orientation Classification; TF-IDF-SIM; Machine Learning

1. Introduction

With the development of network technology, the Internet has become a crucial carrier for information dissemination. Social networks such as microblogs and blogs have rapidly integrated into people' s lives due to their rich service content and convenient operation features. The emergence of Web 2.0 has made microblogging a powerful product of this era, enabling rapid information dissemination and exchange and exerting significant influence on various aspects of society including politics, culture, and economy. An increasing number of enterprises, businesses, and celebrities use microblogs to expand their visibility and enhance public image. Among these, trending topics possess tremendous influence, affecting not only the formation and development of various events in the virtual network society but also influencing people' s perceptions and judgments of events in the real world, and even impacting government and judicial decisions regarding events. Therefore, facing massive amounts of textual data on microblogs, quickly and accurately identifying hot topics and focal points and extracting and analyzing user opinions and emotional information is highly meaningful for both enterprises and governments. This paper studies the sentiment orientation of trending topics and their emotional objects, providing emotional intelligence for precision microblog marketing.

As microblogging' s influence and user base continue to expand, related research has gradually increased, with growing literature on microblog sentiment orientation analysis. Existing methods can be divided into two categories: dictionary-based approaches and machine learning approaches [1].

(1) Dictionary-based methods typically utilize the sentiment polarity and intensity of words in dictionaries to weight given texts and determine overall sentiment orientation. Common Chinese sentiment dictionaries include HowNet, NTUSD Sentiment Dictionary, Student Commendatory and Derogatory Dictionary, and Tsinghua Commendatory and Derogatory Dictionary. Dictionary-based methods do not require training data and can be applied to many domains, but they have several drawbacks in microblog sentiment analysis:

First, the cost of obtaining and updating sentiment dictionaries is high. Gui et al. [2] proposed an automatic method for building sentiment dictionaries based on microblog emoticons, but this approach is only suitable for applications where requirements for microblog sentiment analysis are not particularly stringent. Bravo-Marquez et al. [3] proposed expanding dictionaries in a supervised manner from automatically annotated Twitter data with emoticons and existing dictionaries, using pointwise mutual information and stochastic gradient descent to establish connections between words and sentiments. Experimental results showed improved performance for SentiWordNet (an English sentiment analysis dictionary). While automatically annotated data using emoticons can reduce costs, it also decreases method reliability.

Second, word coverage in sentiment dictionaries is low, making it difficult to cover emerging vocabulary, misspelled words, abbreviations, and informal expressions in microblogs. Ning et al. [4] constructed a new dictionary by merging and deduplicating multiple dictionaries, but microblog vocabulary evolves daily, with new strongly emotional words constantly emerging. Zhou et al. [5] expanded sentiment dictionaries by adding domain-specific opinion words, and experiments on 56 topics showed improved accuracy for dictionary-based classifiers.

Third, the fixed sentiment polarity and intensity of words make these methods domain-independent, while sentiment expression typically involves specific objects or domains and carries different emotional intensities in different contexts. This domain-independence of sentiment dictionaries significantly impacts sentiment classification. Increasingly, scholars have focused on solving this problem. For example, Saif et al. [6-8] conducted a series of studies proposing the SentiCircles method, which dynamically updates sentiment scores of words in sentiment dictionaries when applied to specific datasets through word co-occurrence patterns. This method fully considers the context in which words appear and can improve sentiment classification results in specific domains, but shows no significant improvement in cross-domain scenarios. Additionally, they proposed extracting semantic relationships from DBpedia to improve dictionary adaptability [9], with results showing effective improvements in sentiment classification accuracy and F-measure. Zhao et al. [10] also achieved good performance by combining semantics and prior sentiment.

(2) Machine learning methods require training data for sentiment classification learning, typically manually annotated microblog sentiment orientations (positive, negative, neutral, etc.). Commonly used methods include Support

Vector Machines, Naive Bayes, Neural Networks, and Maximum Entropy [11-12], among which Support Vector Machines have been proven to produce good classification results in many studies. Machine learning methods rely on training data and thus often have advantages when analyzing sentiment for specific microblog topics, but they also have limitations:

First, training data acquisition costs are high. Automatically annotating data using emoticons in text is a common method [3-4], but annotation results still require improvement. Second, training data selection significantly impacts sentiment classification results. Palguna et al. [13] analyzed Twitter sampling algorithms and proposed new statistical metrics to quantify sample representativeness. Song et al. [14] argued that emotional expression in microblogs reflects user personality, and user representativeness in training data is also a concern. Additionally, training dataset size affects results [15]. Third, training data balance impacts classifier performance, with current research mostly addressing this through improved sampling methods [16-17].

When applying machine learning methods to microblog topic sentiment analysis, many previous studies have often overlooked the relevance between microblog content and topics due to the domain-specific nature of supervised learning methods themselves, resulting in noise data reducing training effectiveness. This represents the main entry point for our research.

These two categories of methods each have their strengths, and many studies currently combine their advantages to investigate microblog sentiment orientation [1,18]. We believe that machine learning methods have greater advantages when studying sentiment orientation of specific trending topics. Therefore, this paper improves research on microblog trending topic sentiment orientation by fully considering the impact of noise data, specifically the relevance strength between post content and topics. This study of microblog topic sentiment orientation has three main tasks: text preprocessing, sentiment information extraction, and sentiment classification.

First, we use text preprocessing techniques to extract sentiment information (such as feature words and sentiment objects) from trending topic microblogs. Then we propose a subjective-objective classification model incorporating topic relevance to help extract subjective microblog text collections related to trending topics. We use an improved subjective microblog sentiment classification method to analyze sentiment orientation and evaluate classification performance in detail through recall, precision, and F-measure. In empirical analysis, we study the sentiment orientation of relevant emotional objects for the trending topic #Feng Xiaogang Criticizes Film Critics# and propose recommendations for precision microblog marketing based on the results.

3. Methodology

3.1 Data Acquisition and Text Preprocessing To study microblog trending topic sentiment orientation, we first acquire relevant data through web crawler software and perform appropriate preprocessing. Preprocessing procedures include: extracting emoticons from microblogs; cleaning meaningless microblog text including pure reposts, images, videos, URLs, emoticons, and URL addresses; word segmentation and part-of-speech tagging; and filtering stop words.

3.2 Sentiment Information Extraction (1) Feature Word Extraction Feature word extraction primarily involves extracting words from text that can represent content and play a decisive role in classification, and calculating their feature weights. Common feature extraction methods include Document Frequency (DF), Information Gain (IG), Mutual Information (MI), and Chi-square test (CHI). Different classification tasks use different feature extraction methods. This paper involves three types of classification: topic relevance classification, subjective-objective text classification, and sentiment polarity classification. Topic relevance classification features include words, part-of-speech, and similarity values with topics, using the TF-IDF-SIM algorithm (an improvement based on TF-IDF) to determine feature word weights. Zhang [19] added three new features to the five-dimensional non-text features commonly used in subjective-objective classification, comparing the effects of five-dimensional and eight-dimensional features (as shown in Table 1) using four common classifiers (SVM, ANN, NB, and LR), proving that eight-dimensional features perform better. Further investigation of emoticon features' impact on subjective-objective classification showed that emoticon features effectively improve classification performance. Wu et al. [20] constructed a relatively complete sentiment polarity dictionary including basic sentiment dictionaries, polarity adverb dictionaries, emoticon dictionaries, microblog neologism dictionaries, and domain dictionaries. Based on literature review, we believe that the six feature types shown in Table 2 are the most common emotional features in microblogs, extracted using Mutual Information (MI). Emoticons are stored in the form "[text]" and are extracted and analyzed according to their sentiment in the HowNet sentiment dictionary. Network expressions are manually collected from websites. For sentiment polarity classification, based on Table 2, we divide emoticons, sentiment words, and network expressions into positive and negative categories and add transition words. Note that the sentiment polarity of network expressions is manually annotated, ultimately yielding ten feature types as shown in Table 3 , extracted using Mutual Information (MI).

(2) Sentiment Object Extraction and Merging

Sentiment objects, also called evaluation objects, refer to words or phrases modified by sentiment words in subjective sentences, which can be individuals, organizations, events, products, etc. Sentiment object extraction and sentiment orientation judgment facilitate precision microblog marketing. Current microblog sentiment object extraction methods include rule-based methods, syntactic analysis-based methods, and sequence labeling model methods. Rule-based methods are relatively simple and efficient, so we adopt this approach for evaluation object extraction. In evaluation objects, nouns, noun phrases, and hashtags constitute the main components. If specific nouns or noun phrases exist, they are used as evaluation objects; otherwise, hashtags serve as evaluation objects. However, extracted evaluation objects often contain large numbers of similar terms, such as "冯小刚" (Feng Xiaogang) and "冯导" (Director Feng), which express the same meaning. Therefore, we use the K-means clustering algorithm based on similarity calculation and word coverage to merge evaluation objects.

3.3 Subjective-Objective Classification Model Incorporating Topic Relevance Before studying topic sentiment orientation, we need to extract subjective posts related to the topic, as subjective text content is based on assertions or comments with personal emotions and intentions, while objective text content is based on factual descriptions without personal preferences or biases. This paper proposes a subjective-objective classification model incorporating topic relevance, decomposing the problem into two parallel sub-problems: relevance and subjectivity, then using Logistic regression for integration to obtain subjective posts related to trending topics. The model is illustrated in Figure 1 [Figure 1: see original paper].

As shown in Figure 1, the topic relevance-based subjective-objective classification model contains two sub-models: a topic relevance classification sub-model and a subjective-objective classification sub-model. Both sub-models follow the same main process including feature extraction, feature matrix construction, sample sequence establishment, and model learning. The sample sequence establishment phase uses manual annotation, and the model learning phase uses SVM algorithms. In the feature matrix construction phase, the topic relevance sub-model uses the TF-IDF-SIM method, while the subjective-objective classification sub-model uses the Mutual Information (MI) method.

(1) TF-IDF-SIM Method is an improved algorithm based on TF-IDF that comprehensively considers both the importance of a term to a specific topic corpus and its relevance to the specific topic, thereby assigning a comprehensive weight to the term. TF (Term Frequency) represents term frequency, IDF (Inverse Document Frequency) represents inverse document frequency, and SIM represents the maximum similarity value between the term and topic words. In text feature representation, each microblog post D_j can be represented by word features in the post, with these word features and their weights forming a vector in "space" :

$W_{i,j}$ is the weight of term i in microblog post D_j, expressed as:

$$W_{i,j} = TF_{i,j} \times IDF_i \times SIM_i$$

where $TF_{i,j}$ represents the occurrence count of term i in microblog post D_j; IDF_i is the inverse document frequency coefficient, where N represents the total number of microblog posts in the corpus and n_i represents the number of posts containing term i; SIM_i is the similarity value between term i and current topic words. The similarity algorithm is as follows:

Input: Current word w and current trending topic word $hotTopic$
Output: Maximum similarity $sim(w, hotTopic)$ between word w and current trending topic words

1. Set $sim(w, hotTopic) = 0$;
2. Segment $hotTopic$ to obtain $hotTopicSet = \{H_1, H_2, ..., H_n\}$;
3. For each H_i, if $w = H_i$, then $sim(w, hotTopic) = 1$ and proceed to step 5; otherwise proceed to step 4;
4. Calculate $sim(w, H_i)$ using the word similarity calculation algorithm based on Tongyici Cilin [21]. If $sim(w, H_i) > sim(w, hotTopic)$, update $sim(w, hotTopic) = sim(w, H_i)$;
5. Algorithm ends.

(2) Mutual Information (MI) Method extracts the six types of emotional features shown in Table 2 (emoticons, sentiment words, etc.) and calculates their mutual information with subjective and objective text classes. Mutual Information is a useful information measure in information theory that refers to the correlation between two event sets, calculated through formula (4). Its meaning is the amount of information provided by the occurrence of event A associated with the occurrence of event B:

$$I(A, B) = \log_2 \frac{P(AB)}{P(A)P(B)}$$

where $P(AB)$ represents the probability of events A and B occurring simultaneously, $P(A)$ represents the probability of event A occurring, and $P(B)$ represents the probability of event B occurring. When extracting features for classification problems, mutual information measures the correlation between a feature term and a specific category. Larger information volume indicates greater correlation between the feature and category, and vice versa. The basic process of using mutual information for feature extraction is: assuming feature term t, subjective class c_1, and objective class c_2, the calculation result is $MI(t, c)$. Define a threshold θ; if the feature term satisfies formula (5), extract the feature term:

$$MI(t, c_1) > MI(t, c_2) > \theta \quad (5)$$

Finally, Logistic regression combines the two sub-models to construct a topic relevance-based subjective-objective classification model. From the above analysis, the essence of trending topic subjective-objective classification is finding a functional relationship between random variable Y and random vector (x_1, x_2), where Y represents whether current discourse is topic-relevant and subjective, x_1 is the independent variable of whether the post is topic-relevant, and x_2 is the independent variable of whether the post is subjective text. Since labels (x_1, x_2) and classification results Y are discrete data, we use Logistic regression analysis to solve this problem.

3.4 Improved Sentiment Classification for Subjective Microblogs After obtaining topic-relevant subjective texts, we need to classify sentiment orientation of subjective texts and determine sentiment orientation for each evaluation object based on extracted objects. Current Chinese microblog sentiment classification methods can be divided into two categories: semantic dictionary-based sentiment calculation and machine learning-based sentiment classification. There is not yet a universal and complete sentiment dictionary in the microblog text sentiment analysis field, and due to context shift, most existing sentiment dictionaries suffer from insufficient emotional coverage and poor classification effects in microblog sentiment analysis [14-15]. Therefore, this paper adopts a machine learning approach using sentiment words, emoticons, network expressions, etc. as classification features. Through classification algorithm training, we build a classifier that categorizes sentiment orientation into positive and negative tendencies.

4. Empirical Analysis

4.1 Microblog Data Acquisition and Preprocessing We used a web crawler to acquire 91,361 posts from the Sina Weibo trending topic #FengXiaogangCriticizesFilmCritics#. After text preprocessing, 88,571 posts remained, which were randomly split into a training set (68,889 posts) and test set (19,682 posts), with the training set being 3.5 times the test set. Three experts manually annotated the microblog texts as relevant/irrelevant in terms of relevance, and as objective, positive, or negative in terms of sentiment polarity. Before annotation, each expert received a booklet explaining classification concepts for reference. For example, positive texts typically contain user support or optimistic attitudes toward events. Inter-expert agreement exceeded 80%, passing reliability tests. Final classification results were determined by majority rule (frequency-based method). Annotation results are shown in Table 4 .

4.2 Subjective-Objective Classification of Trending Topics According to the research methods in Section 2.3, we extracted feature words and matrices. Using the WEKA platform, we employed SVM classifiers to separately annotate the topic relevance classification sub-model and subjective-objective text classification sub-model, then used Logistic algorithms to unify the dual annotations in one model. Results are shown in Table 5 .

We also investigated the impact of adding the topic relevance classification sub-model on trending topic subjective-objective text classification, with results shown in Table 6 . The results show that incorporating the topic relevance classification sub-model yields better performance for trending topic subjective-objective text classification, with overall F-measure improving by 7.4%. This indicates that discourse unrelated to trending topics in the dataset affects subjective-objective text classification performance, and incorporating topic relevance significantly improves subjective-objective classification effectiveness.

4.3 Sentiment Orientation Classification of Subjective Microblogs According to the improved subjective microblog sentiment classification method in Section 3.4, we judged the sentiment polarity of extracted topic-relevant subjective microblogs and compared results before and after improvement, as shown in Table 7 .

We also studied the impact of adding the topic relevance-based subjective-objective classification model on sentiment orientation classification, with results shown in Table 8 . Adding the topic relevance-based subjective-objective text classification model improved sentiment orientation classification performance, with overall F-measure increasing by 2.2%. This shows that screening relevant and subjective texts for sentiment classification can significantly reduce classifier burden and provide more accurate classification results.

4.4 Sentiment Object Extraction and Sentiment Orientation Judgment According to the method in Section 3.2, we extracted and merged sentiment objects, and judged sentiment orientation for the top 5 most discussed evaluation objects, as shown in Table 9 .

From Table 9, we can see that the hashtag (#FengXiaogangCriticizesFilmCritics) has approximately 1.7 times more positive sentiment than negative sentiment, indicating that Sina Weibo users' support for this event far exceeds opposition. The evaluation object "Feng Xiaogang" has about 1.6 times more positive than negative sentiment, while "Small Stories" and "Ge You" are plot elements and actors in the movie "Personal Tailor," which received more negative sentiment from users. This shows that users' negative feelings about the movie did not affect their fondness for Feng Xiaogang or their support for this event. Overall, user discussions of the topic focused more on the trending topic itself and Feng Xiaogang.

From a microblog marketing perspective, film producers and other enterprises can obtain user attitude and preference feedback, and can actively promote the movie' s concept and underlying meaning for objects receiving negative evaluations like "Personal Tailor," helping audiences understand more positive aspects of the film to guide public opinion. Government departments can also focus on monitoring and eliminating sensitive information after large amounts of negative information appear, formulating relevant strategies for public opinion guidance and control by targeting the most discussed evaluation objects.

5. Conclusion

This study investigates sentiment orientation of trending topics and their emotional objects. We propose a subjective-objective classification model incorporating topic relevance to help extract subjective microblogs related to trending topics, use improved sentiment classification methods to analyze extracted posts' sentiment orientation, and evaluate classification performance through recall, precision, and F-measure.

Empirical results show that the topic relevance-based subjective-objective classification model facilitates trending topic subjective-objective classification and improves microblog sentiment classification effectiveness. By extracting sentiment orientation of key emotional objects within trending topics, this approach provides intelligence for precision microblog marketing.

Key innovations include: extracting and judging sentiment orientation of evaluation objects in topics, which holds more value than pure theoretical and technical discussions of sentiment classification; proposing a topic relevance-based subjective-objective classification model that improves subjective-objective text classification and consequently sentiment polarity classification, likely because considering relevance reduces noise data impact; and proposing improved sentiment classification methods that consider non-standard text like emoticons, with results proving the method and model' s effectiveness and practicality.

Several aspects warrant improvement: sample distribution imbalance may affect classification results, but since this study examines all microblog data within a topic, we did not adjust sample distribution; sentiment orientation classification only divides into positive and negative, while practical applications require finer-grained sentiment types; and the study lacks emotional trend analysis for sentiment objects. Future research can address these issues.

References

[1] Chen G. Microblog Sentiment Analysis Basing on Emotion Dictionary and Semantic Rule[J]. Information Research, 2016(2): 1-6.

[2] Gui B, Yang X, Zhang Z, et al. Research on Building Lexicon for Sentiment Analysis Based on the Chinese Microblogging Smiley[J]. Transactions of Beijing Institute of Technology, 2014, 34(5): 537-541.

[3] Bravo-Marquez F, Frank E, Pfahringer B. Building a Twitter Opinion Lexicon from Automatically-annotated Tweets[J]. Knowledge-Based Systems, 2016, 108(SI). DOI: 10.1016/j.knosys.2016.05.018.

[4] Ning H, Yang S, Zhao Y, et al. Study of Microblog Sentiment Analysis Based on Semantic Feature[J]. Applied Science and Technology, 2016, 43(3): 70-74.

[5] Zhou Z, Zhang X, Sanderson M. Sentiment Analysis on Twitter Through Topic-Based Lexicon Expansion[A]// Databases Theory and Applications[M]. Springer International Publishing, 2014: 98-109.

[6] Saif H, Fernandez M, He Y, et al. SentiCircles for Contextual and Conceptual Semantic Sentiment Analysis of Twitter[A]// The Semantic Web: Trends and Challenges[M]. Springer, Cham, 2014: 83-98.

[7] Saif H, He Y, Fernandez M, et al. Adapting Sentiment Lexicons Using Contextual Semantics for Sentiment Analysis of Twitter[A]// The Semantic Web: ESWC 2014 Satellite Events[M]. Springer, Cham, 2014: 54-63.

[8] Saif H, He Y, Fernandez M, et al. Contextual Semantics for Sentiment Analysis of Twitter[J]. Information Processing & Management, 2015, 52(1): 5-19.

[9] Saif H, Fernandez M, Kastler L, et al. A Linked Open Data Approach for Sentiment Lexicon Adaptation[C]// Proceedings of the 15th International Semantic Web Conference. 2016.

[10] Zhao J, Cao X. Combining Semantic and Prior Polarity for Boosting Twitter Sentiment Analysis[C]//Proceedings of the International Conference on Smart City/Socialcom/Sustaincom. IEEE, 2015: 832-837.

[11] Le B, Nguyen H. Twitter Sentiment Analysis Using Machine Learning Techniques[A]// Advanced Computational Methods for Knowledge Engineering[M]. Springer International Publishing, 2015: 279-289.

[12] Qasem M, Thulasiram R, Thulasiram P. Twitter Sentiment Classification Using Machine Learning Techniques for Stock Markets[C]//Proceedings International Conference on Advances in Computing, Communications and Informatics. IEEE, 2015.

[13] Palguna D, Joshi V, Chakaravarthy V, et al. Analysis of Sampling Algorithms for Twitter[C]// Proceedings of the 24th International Joint Conference on Artificial Intelligence. AAAI Press, 2015.

[14] Song K, Feng S, Gao W, et al. Personalized Sentiment Classification Based on Latent Individuality of Microblog Users[C]// Proceedings of the 24th International Joint Conference on Artificial Intelligence. AAAI Press, 2015.

[15] Abdelwahab O, Bahgat M, Lowrance C J, et al. Effect of Training Set Size on SVM and Naive Bayes for Twitter Sentiment Analysis[C]// Proceedings of the IEEE International Symposium on Signal Processing and Information Technology. 2015: 46-51.

[16] Saif H, He Y, Alani H, et al. On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter[C]// Proceedings of the 9th International Conference on Language Resources and Evaluation. 2014.

[17] Ah-Pine J, Morales E P S. A Study of Synthetic Oversampling for Twitter Imbalanced Sentiment Analysis[C]// Proceedings of the Workshop on Interactions Between Data Mining and Natural Language Processing. 2016.

[18] Sabariah M K, Effendy V. Sentiment Analysis on Twitter Using the Combination of Lexicon-based and Support Vector Machine for Assessing the Performance of a Television Program[C]//Proceedings of the International Conference on Information and Communication Technology. 2015.

[19] Zhang X. Research on Sentiment Analysis for Hot Topic Microblog[D]. Harbin: Harbin Institute of Technology, 2013.

[20] Wu Q, Wang Y. Research on the Emotional Feature Selection Method in the Chinese Microblog[J]. Journal of Inner Mongolia Normal University: Natural Science Edition, 2016, 45(1): 84-88.

[21] Tian J, Zhao W. Words Similarity Algorithm Based on Tongyici Cilin in Semantic Web Adaptive Learning System[J]. Journal of Jilin University: Information Science Edition, 2010, 28(6): 602-608.

Author Contributions

He Yue: Proposed research ideas and designed research methodology;
Xiao Min, Zhang Yue: Conducted experiments, collected, cleaned, and analyzed data;
He Yue, Xiao Min, Zhang Yue: Drafted the manuscript;
He Yue, Xiao Min: Revised the final version.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data is self-archived by the authors, E-mail: 2279332915@qq.com.
[1] He Y, Xiao M, Zhang Y. Data.xlsx. Raw data and manual classification annotation results.
[2] He Y, Xiao M, Zhang Y. Data.xlsx. Subjective-objective classification results table.
[3] He Y, Xiao M, Zhang Y. Data.xlsx. Subjective-objective text classification results with/without topic relevance classification sub-model.
[4] He Y, Xiao M, Zhang Y. Data.xlsx. Sentiment orientation classification results before and after improvement.
[5] He Y, Xiao M, Zhang Y. Data.xlsx. Sentiment orientation classification results with/without topic relevance-based subjective-objective classification model.

Received: 2016-10-17
Revised: 2017-01-25

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