

Research Article

Microblog Sentiment Orientation Detection Using User Interactive Relationship

Liang Wang, Mei Wang, Xinying Guo, and Xuebin Qin

Xian University of Science and Technology, Xian 710054, China

Correspondence should be addressed to Liang Wang; wangliang@xust.edu.cn

Received 26 November 2015; Accepted 16 February 2016

Academic Editor: Jiang Zhu

Copyright © 2016 Liang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The development and popularity of microblog have made sentiment analysis of tweets and Weibo an important research field. However, the characteristics of microblog message pose challenge for the sentiment analysis and mining. The existing approaches mostly focus on the message content and context information. In this paper, we propose a novel microblog sentiment analysis framework by incorporating the social interactive relationship factor in the content-based approach. By exploring the interactive relationship on social network based on posted messages, we build social interactive model to represent the opposition or acceptance behavior. Based on the interactive relationship model, the sentiment of microblog message with sparse emotion terms can be deduced and identified, and the sentiment uncertainty can be alleviated to some extent. Afterwards, we transform the classification problem into an optimization problem. Experimental results on Weibo data set indicate that the proposed method can outperform the baseline methods.

1. Introduction

With the advent of Web 2.0, users become more eager to publish and share their opinions in various domains on social networks, such as Twitter and Weibo. The opinions can reflect people's sentiments and views across areas as diverse as commercial products, services, and public events, while these opinions can influence the ultimate decisional process related to individual behavior and public policy. For example, by analyzing the sentiment of consumers for a certain brand and released product, it can help companies to improve their marketing campaign, product design, and user experience. After comprehending mass voters' opinions towards a candidate, it is conducive to making their political strategy better [1–3]. So, as it were, the microblog sentiment analysis has become a hot area which attracted attentions in both academic and industrial fields.

However, the microblog message is usually short, noisy, and ambiguous due to its informal expression style. These characteristics present a challenge for the text sentiment analysis. Recently, many works are making great efforts to deal with the microblog sentiment classification problem. Generally speaking, the previous works are mostly based on the

content-based approach, such as lexicon-based method and corpus-based method [4–6]. The content-based approach needs to establish a sentiment word dictionary beforehand. And, through the sentiment word matching calculation, the corresponding sentiment could be labeled to the microblog message to be analyzed. Actually, the previous researches mostly rely on predefined sentiment vocabularies, which are highly domain specific.

The existing research works have proved that the content-based method is effective in tackling microblog messages with strong sentiment expressions. However, sometimes the message does not contain obvious sentiment terms, but there still exists emotional tendency implicitly to a certain extent. For this situation, the pure content-based approach is inadequate to identify the sentiment orientation. In other words, the sparse sentiment terms in microblog messages make it difficult to classify the hidden emotion. Moreover, in the microblog message dissemination process, the sentiment characteristic may not be discriminative to classify its emotion polarity. In view of this situation, some scholars consider incorporating the hidden social relationship, such as friend and follower, into message content to settle the

sparse sentiment terms problem [7, 8]. The research works in [7, 8] are based on the hypothesis that people with friend relationship often keep similar sentiment towards a given topic.

Meanwhile in practice people may be in disagreement with others in their friend circle as they have different social cognition, knowledge structure, and experience. That is to say, one may oppose or accept the opinion reflected in message posted by someone else. And the opposition or acceptance behavior also reflects the emotion polarity of the participants in the microblog environment. In this paper, this opposition or acceptance behavior is regarded as a social interactive relationship. By leveraging this kind of interactive relationship, it may be feasible to infer the sentiment of message for participants and build the learning model. In this work, we exploit the social relationship between users on microblog platform and incorporate this factor into content-based analysis process. Via exploiting the social interactive relationship among microblog users, a voting-based social interactive model has been built to link the microblog participants based on the message context. Afterwards, the sentiment classification problem can be transformed into an optimization problem integrating the content-based strategy and social interactive relationship factor. Experimental results on microblog message data sets validate the effectiveness and efficiency of our method.

This paper is structured as follows. In Section 2, we review related work on sentiment classification. After that, we describe and formulate the social interactive relationship in microblog message and propose the classification model in detail in Section 3. In Section 4, we then evaluate our models and present experimental results. Finally, in Section 5, we summarize our findings and contributions.

2. Background and Related Works

In recent years, as an opinion-rich resource, microblog messages provide the possibility of mining and analyzing the emotion of people in different scales. And microblog sentiment orientation identification has gained huge popularity and attracted researchers' attention across many fields. In this section, an overview of works related to microblog sentiment analysis and classification is presented.

Previous works usually compare the text content of a given message with a lexicon or a dictionary to classify its sentiment and calculate its strength. For instance, SentiWordNet contains as many as 200,000 entries to match each word with positive, negative, or objective scores. Mostafa et al. propose a lexicon-based method to analyzing consumers' sentiment towards some commercial brands [1, 2]. Go et al. [2] study the problem of sentiment classification for Twitter messages by the means of machine learning algorithms. Based on distant supervised learning, the work compares the sentiment classification performance for three algorithms, including Naïve Bayes, Maximum Entropy, and SVM. Ghiassi et al. [3] introduce an approach to supervised feature reduction using n -grams and statistical analysis to develop a Twitter-specific lexicon for sentiment analysis. Kumar and Sebastian [4]

expound a hybrid approach to analyze the sentiment of Twitter. The approach employs a corpus-based method to identify the semantic orientation of adjectives and a dictionary-based method to find the semantic orientation of verbs and adverbs. Saif et al. [5] add semantics information as additional features in the process of tweets sentiment analysis. Moreover, the correlation of the representative concept is calculated with positive and negative sentiment. Bollen et al. [6] extract six dimensions of emotion using an established psychometric instrument POMS and perform tweet sentiment analysis in the field of stock market, crude oil price indices, and major public events.

As tweets are usually short and more ambiguous, Jiang et al. [9] study the target-dependent Twitter sentiment classification problem. Through incorporating target-dependent features and context information related to the given target tweet, the sentiment of query tweet can be classified as positive, negative, or neutral. Pang and Lee [10] review the related approaches for opinion mining and sentiment analysis. Besides, some other issues, including privacy protection, manipulation, and economic impact, are also discussed in [10]. Although many researchers point out that the properties of short and noisy present new challenges to the sentiment analysis for microblog message, Birmingham and Smeaton [11] find classifying sentiment in microblogs easier than in blogs. Davis and O'Flaherty [12] evaluate the automated sentiment coding accuracy and misclassification errors of six leading third-party companies for a broad range of comment types and forms.

Actually, beside the content-based information, the hidden social relationship in microblog message can also be reflected and may be used in the sentiment analysis process. Recently, there are many works that focus on the social relationship in microblog space and leveraged the relationship to improve the classification accuracy [7, 8]. Hu et al. [7] exploit social relationship and introduce two social theory hypotheses into social network, which are sentiment consistency and emotional contagion. Pang and Lee [10] expand the hypothesis verified by Hu et al. in [7] and proposed a microblog sentiment classification framework. Inspired by the work in [7], we consider combining the implicit social relationship behind the microblog message to classify its sentiment polarity.

3. Problem Definition

3.1. Motivation. The purpose of sentiment orientation identification for microblogs is to build a program which can automatically identify whether a given microblog message is expressing positive, negative, or no sentiment. In other words, this problem can be defined as follows: given a collection of microblog message set D and a set of classes $P_G = \{\text{positive, neutral, negative}\}$, the goal is to construct a mapping function $\Phi_G : D \rightarrow P_G$, which describes the sentiment orientation of messages according to an integrated model.

However, the character of microblog message, such as noisy, short, and ambiguous, poses a challenge for sentiment analysis and mining. In addition, the informal language

expression and the use of emoticon make the sentiment analysis problem more difficult. If the message to be analyzed contains distinct sentiment vocabulary, it is not hard to classify the sentiment of this message by the means of established semantic lexicon and dictionary. However, when there is no obvious sentiment vocabulary in the message but, with sentiment orientation, it is impossible to obtain correct sentiment estimation, sometimes it is difficult to determine the implicit emotion via the content-based approach. Especially in the process of information dissemination, such as forward, comment, and @ operations, the original microblog message may change its sentiment to some extent and even generate polarity reversal. Let us consider two examples as follows.

Example 1. Well, Jane's opinion towards genetically modified food is so. What do you think?

Example 2. @Jack, I do not agree with your views about genetically modified food.

As described above, in Example 1, the poster of this message does not express his opinion for the topic discussed by Jane. But it is undeniable that the message's holder also delivers subtle emotion in the published message. For Example 2, the user claims that he disapproves the opinion of Jack's viewpoint, but the explicit sentiment towards the given topic remains unknown for analysts. That is to say, the user's emotion is opposite to the sentiment of Jack's. For these two cases, it is quite clear that context-based approach cannot obtain sentiment analysis results.

3.2. Social Interactive Relationship Model. Obviously, microblog message possesses not only explicit text content information but also implicit social interactive relationship due to its social network nature. In microblog website, a message is published by a user to express his/her opinion towards event, public figure, and commercial products, while this message may be forwarded and commented on by other people who are interested in it. Via the social interactive operation, the opinion implied in a message diffuses in virtual cyber space and influences many participants. It is quite clear that the social interactive operation plays a major role in the information dissemination and can reflect the emotion among the participants. Moreover, in the process of microblog message diffusion, the social interactive operation may enhance or weaken or even reverse the emotion hidden in the original message. In other words, social interaction is a key issue for dissemination and evolution of microblog message.

In practice, some people usually hold an objective and comprehensive opinion in their published messages, while some people may keep a subjective or extreme opinion towards the evaluated objects, such as events or other people. As a matter of common sense, the posted messages of the former microblog users should obtain more support or acceptance from other people. And the latter ones will receive criticisms for their one-sided opinions. The support or opposition can be seen as a vote from other people to

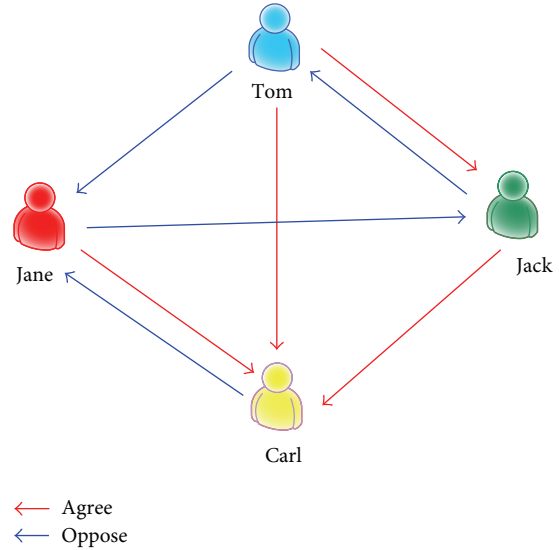


FIGURE 1: A toy example for social interactive relationship.

the publisher of certain message. If the pattern of social interactive relationship can be obtained, it is possible to infer and predict the sentiment of microblog messages. Based on the observation, we consider leveraging this kind of social interactive relationship in the process of microblog sentiment classification.

Let us demonstrate the principle by a toy example in Figure 1. In the toy example, there are four users Tom, Jack, Jane, and Carl. The social interactive relationship is represented by the directed line, where the red one denotes agreement relation and the blue one is opposition relation. And the direction of line is a forward, comment, or @ operation to diffuse microblog messages. As in Figure 1, Tom has opposed message published by Jane once and accepted the view of message posted by Carl once. Thus, assuming that Carl has posted a microblog message related to a certain topic, if there exists social interactive behavior between Carl and Jack on this published message, it is probably that the emotion of Jack's opinion is similar to Carl's.

From the social interactive relationship depicted in Figure 1, we conceive that the support and opposition relation among microblog participants can be exploited to deduce the sentiment of messages under the condition of poor emotion expressions. Firstly, it is necessary to build the social interactive relationship with mathematical formula. Suppose a message posted by user i is supported by user j ; that is to say, user j keeps similar sentiment to user i , while if user i 's message is opposed by user j , it is regarded that user j holds opposite sentiment to user i . Via the interactive history, we can build an interactive relationship matrix R to represent the attribution based on the microblog messages.

More specifically, element $r_{i,j}$ in matrix R denotes the interactive relationship from user i to user j . $r_{i,j} = 1$ means user i has supported user j 's opinion in message, while $r_{i,j} = -1$ means user i has opposed user j 's opinion hidden in the message published by user j . If there exists no kind of this interactive relationship, element $r_{i,j}$ is assigned to

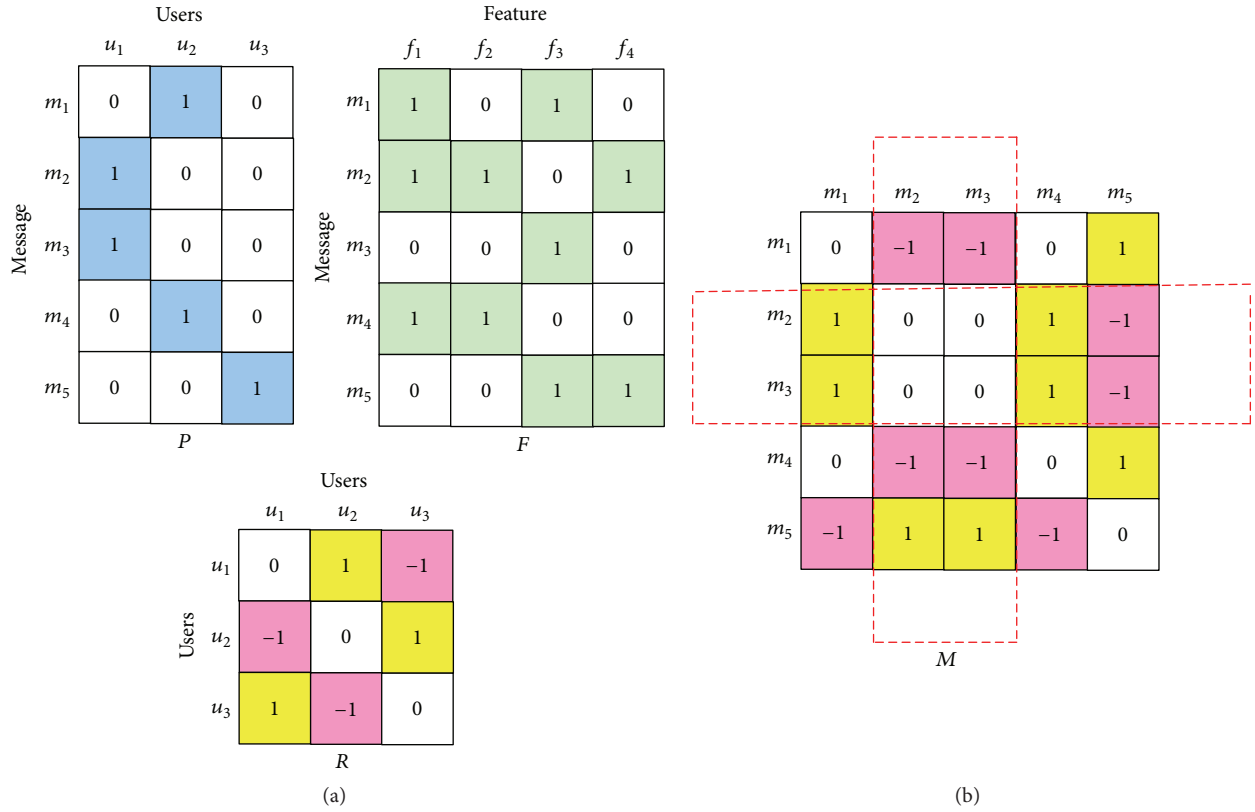


FIGURE 2: An example of microblog message matrix representation.

1. The interactive relationship matrix R for the above toy example is listed as follows. It is not hard to obtain that Carl's influence is powerful among the four people. For other people, including Tom, Jack, and Jane, all vote in favor of him. That is to say other people's sentiment polarity for certain topic is homogenous to Carl's. It is noted that matrix R is a nonsymmetric matrix as the relationship between any two users is bidirectional:

$$R = \begin{bmatrix} 0 & 0 & -1 & 1 \\ -1 & 0 & -1 & 1 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{bmatrix}. \quad (1)$$

3.3. Sentiment Classification Model. In the following, we should extract the interactive relation from posted messages in data set. Given a microblog message data set, $P \in \mathfrak{R}^{l \times n}$ can be used to represent the content matrix, where l is the number of messages and n is the number of participant users, as shown in the left of Figure 2. $P_{i,j} = 1$ denotes that message m_i is published by user u_j . For each message in the data set, we can build the message-feature vector. And the corresponding message-feature matrix $F \in \mathfrak{R}^{l \times q}$, where q is the number of extracted feature, can be shown in Figure 2. And the user-user matrix which is obtained from history messages is $R \in \mathfrak{R}^{n \times n}$, as shown in the right of Figure 2.

Next we can define the problem of sentiment orientation identification in this work. Consider a microblog message data set $P \in \mathfrak{R}^{l \times n}$, the message-feature representation for the given message data set $F \in \mathfrak{R}^{l \times q}$, and a user-user social interactive relationship matrix $R \in \mathfrak{R}^{n \times n}$, and the goal of sentiment analysis is to obtain classifier W to label sentiment automatically.

Inspired by the work [7], we propose a hybrid approach incorporating microblog social interactive relationship into traditional content-based sentiment analysis process. Specifically, via modeling the social interactive relationship history, we establish a matrix to represent the voting relationship among many microblog users. Afterwards, we construct the relationship for messages based on the users' social interaction. The relationship can be calculated by the formula as follows:

$$M = P \times R \times P^T. \quad (2)$$

The elements in matrix M represent whether two different messages are published by two participants in microblog with social interactive relationship. If $m_{i,j} = 1$, that means the author of message i has a support interactive relationship with the author of message j and sentiments of these two microblog messages are identical, while $m_{i,j} = -1$ indicates that message i 's holder has an opposed relationship with message j 's and the sentiments of the related two messages are inverse. As message 2 and 3 are all posted by user 1 in the

above example, the values in the integrated matrix M have similar polarity, which are circled by red dotted lines.

In this paper, we employ the n -gram model to extract the feature vector $f_i \in \mathcal{R}^{Q \times 1}$ of textual messages, where Q is the dimension of feature vector. By leveraging the Least Squares method, it is not hard to generate the content-based sentiment classifier score. More generally, the classification problem can be transformed into an optimization problem as

$$\min \frac{1}{2} \|X^T W - Y\|_F^2, \quad (3)$$

where X is the message to be analyzed, W denotes the classifiers, and Y is the sentiment label vector from context-based classifiers. The goal is to find a vector of W to make the Frobenius norm of matrix reaches minimum value.

Moreover, it is necessary to integrate the aforementioned social interactive relationship factor into the learning process depicted in formula (2). The basic idea is to establish a correlation to make two messages as similar as possible if their holders have a support interactive relationship and make two messages different in emotion if their authors have an opposed relationship. So the social interactive relationship between messages can be formulated as follows:

$$\sum_{i=1}^N M_{i,j} |L_i - L_j|, \quad (4)$$

where $L = X^T W$, $L = \{L_1, L_2, L_3, \dots, L_N\}$.

So, the sentiment classification can be solved by the synthetical formula:

$$\frac{1}{2} \sum_{i=1}^N (L_i - Y_i)^2 + \gamma \sum_{i=1}^N M_{i,j} |L_i - L_j|. \quad (5)$$

4. Experiment and Result Analysis

Each experiment is repeated ten times independently and the average results are reported. The experiments are conducted on microblog message data set which is crawled from Sina Weibo. The data set consists of 3.7 million messages. The corresponding social network is also included in the data set. Firstly, we conduct experiments to verify the performance of our proposed SSTI (Sparse Sentiment Terms Identification) approach by comparing it with existing common sentiment classification approaches including LS (Least Squares) and SVM (Support Vector Machine). The experimental results are shown in Figure 3. From there, it is not hard to see that the performance of proposed method SSTI is better than that of the baseline methods on different data size. The reason is that the existing methods are purely content-based approach and ignore the social relationship. Moreover, the sentiment identification accuracy for microblog messages increases as the size of training data set increases.

We also investigate parameter γ in our proposed sentiment classification learned model. The role of γ is to control the contribution of social interactive relationship factor in the optimization formula. Experiments are conducted to study

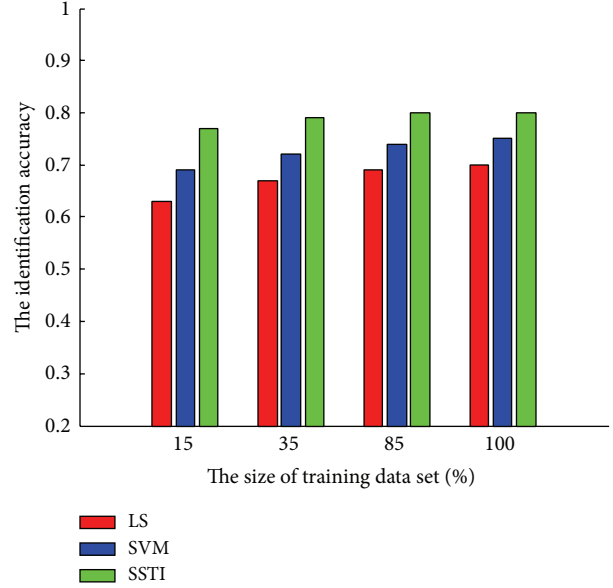


FIGURE 3: Sentiment identification on microblog messages.

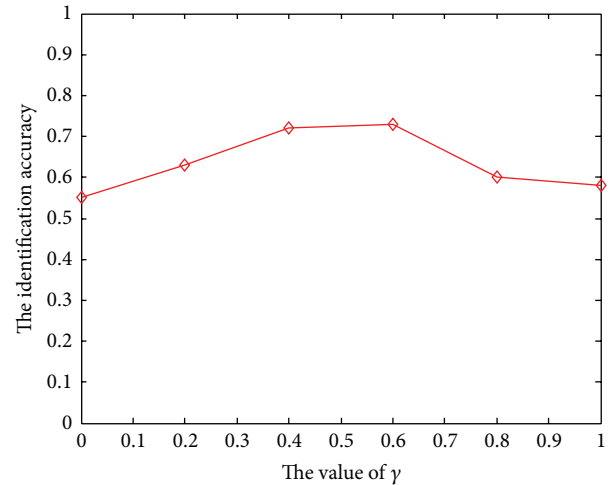


FIGURE 4: The influence of parameter γ in sentiment identification.

the influence of γ in the sentiment identification process. The corresponding results are shown in Figure 4 as follows. As shown in Figure 4, when the value of parameter γ is not large, the performance of sentiment identification for microblog messages could be improved as γ increases, while if the parameter increases to some extent, the performance will drop obviously. The reason for this phenomenon is that when the parameter achieves a large value, the influence of social interactive relationship factor is overstressed. So, it is necessary to choose a suitable value for the parameter.

Next, we will discuss the guidelines for how to choose an appropriate value of parameter γ . As explained above, the value of parameter γ is to determine the weight of social interactive relationship factor. Although the social interaction can improve the sentiment classification accuracy

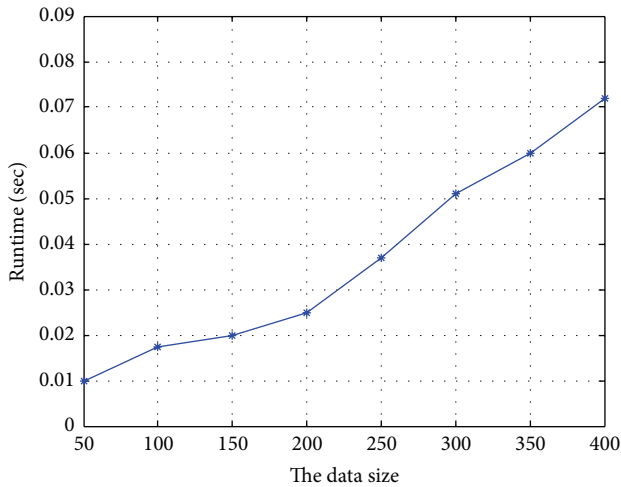


FIGURE 5: The runtime efficiency of our approach.

of microblog message, the information derived from content itself is also important. In other words, the social interactive relationship and content semantic information should be leveraged synchronously. From Figure 4, the sentiment identification accuracy can achieve best performance when the value of parameter γ varies from 0.4 to 0.6. That is to say, the parameter should be chosen from 0.4 to 0.6. For the exact value of parameter γ , it may change slightly for different training data set, social culture, and so on.

Finally, we explore the runtime efficiency of our proposed approach. The experiments are implemented with Interl(R) Core i3-3110M CPU, 4.00 GB RAM, in Matlab R2010b environment. The runtime efficiency results are shown in Figure 5. From the experiment results in Figure 5, it is not hard to obtain that the runtime of our proposed approach is approximately linear with the size of data. It takes 0.07 seconds to finish the classification process when the data size is equal to 400. The results validate the efficiency of our approach.

5. Conclusions and Future Work

Sentiment classification of microblog message is an important research area. Through classification and analysis of sentiments on microblog, one can get an understanding of people's attitudes about particular topics. However, sometimes there are not enough emotion terms in the messages to be analyzed. The sparse sentiment terms in microblog message pose a challenge to the content-based sentiment classification methods. Towards this problem, we propose a novel notion of social interactive relationship based on microblog messages and propose incorporating it in the existing content-based approach. After modeling the social interactive relationship matrix from user-message matrix, we build a sentiment identification learned model to analyze the emotion of a given message. Experiments demonstrate that our proposed approach can improve the sentiment classification performance significantly.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant nos. 61402360 and 61401015).

References

- [1] M. M. Mostafa, "More than words: social networks' text mining for consumer brand sentiments," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241–4251, 2013.
- [2] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Project Report CS224N, Stanford University, 2009.
- [3] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Systems with Applications*, vol. 40, no. 16, pp. 6266–6282, 2013.
- [4] A. Kumar and T. M. Sebastian, "Sentiment analysis on twitter," *International Journal of Computer Science Issues*, vol. 9, no. 3, pp. 372–378, 2012.
- [5] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in *The Semantic Web—ISWC 2012*, vol. 7649 of *Lecture Notes in Computer Science*, pp. 508–524, Springer, Berlin, Germany, 2012.
- [6] J. Bollen, A. Pepe, and H. Mao, "Modeling public mood and emotion: twitter sentiment and socio-economic phenomena," <http://arxiv.org/abs/0911.1583>.
- [7] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM '13)*, pp. 537–546, Rome, Italy, February 2013.
- [8] F. Wu, Y. Huang, and Y. Song, "Structured microblog sentiment classification via social context regularization," *Neurocomputing*, vol. 175, pp. 599–609, 2016.
- [9] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies—Volume 1*, pp. 151–160, Association for Computational Linguistics, June 2011.
- [10] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.
- [11] A. Bermingham and A. F. Smeaton, "Classifying sentiment in microblogs: is brevity an advantage?" in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*, pp. 1833–1836, ACM, Toronto, Canada, October 2010.
- [12] J. Davis and S. O'Flaherty, "Assessing the accuracy of automated twitter sentiment coding," *Academy of Marketing Studies Journal*, vol. 16, supplement 1, p. 35, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

