Discriminating Topical Influencers Based on the User Relative Emotion

LANLIN GAO, YUE WU, XI XIONG, AND JINYUAN TANG

1School of Computer and Software Engineering, Xihua University, Chengdu 610039, China
2School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China
3School of Aeronautics and Astronautics, Sichuan University, Chengdu 610065, China
4Department of Information Engineering, Chengdu Industry and Trade College, Chengdu 610039, China

Corresponding author: Yue Wu (e-mail: wuyue@mail.xhu.edu.cn).

This work was supported in part by the National Natural Science Foundation of China under Grant 61602389 and 61602064, in part by the Scientific Research Project of Sichuan Education Department of China under Grant 18ZB0028, and in part by the Innovation Fund of Postgraduate of Xihua University under Grant YCJJ2018088.

ABSTRACT Topical influencers are experts on a specific topic, who always play an important role in the opinion dissemination. From the perspective of influence polarity, topical influencers can be classified into opinion leaders who gain a great deal of support, trolls who are widely condemned, and controversial figures who trigger debate. In this paper, discriminating topical influencers of social networks is addressed. Firstly, the trained BiLSTM-CRF model is constructed to extract emotional elements, and then the proposed Emotional Matching and Emotional Transforming algorithms are leveraged to obtain the relative emotion of users. Secondly, the rank of influencers is calculated by the quantified user behavior characteristics and the Multi-centrality algorithm. Finally, according to relative emotion of users, the topical influencers are divided into three categories—opinion leaders, trolls and controversial figures. Experimental results show that the user relative sentiment analysis method proposed in this paper has higher accuracy, and the empirical analysis manifests that the three kinds of influencers affect the opinion distribution in various degrees.

INDEX TERMS User relative emotion, influencers discrimination, opinion leader, troll, controversial figure.

I. INTRODUCTION

In the social networks, influencers are able to attract a lot of attention, and determine the emotional tendency of the entire network group. The polarity of influence differs among the influencers in the topics. Influential users with correct opinions can gain broad support, and are able to release positive posts by virtue of their credibility and authority [1], [2]. Some influencers can be as destructive as the opinion leaders are positive [3], [4], such users receive opposition and condemnation at large, being good at presenting false information to trigger the extreme opinion. At present, several scholars have studied the controversial influencers who often get the similar number of supporters and opponents with the ability to attract more users to discuss. Different types of influencers play different but equally important roles in a variety of networks [5]–[7]. Discrimination of influencers plays an impressive role in ad delivery [8], information dissemination [9], [10] and group opinion analysis [11]. This paper proposes a comprehensive discriminating method to divide the influencers into opinion leaders, trolls and controversial figures. Opinion leaders are “activists” who provide information to others and trusted by the majority in the social network. Opinions leaders in terms of marketing field are also defined as Key Opinion Leader (KOL) [12], that considered as an absorbent goal for marketing, advertising and brand evaluation [8], [13]. The trolls are originally proposed by Hardaker who determined the four basic characteristics of the trolls: deception, attack, destruction and success [14]. Trolls are usually defined as users who are marked by a negative behavior [14], or as individuals who initially pretend to be a legitimate participant and then attempt to disrupt the social network while affecting to attract the maximum number of responses [15]. This paper has also defined the controversial figures who are able to contribute people to talk easily, raise the controversy points in the topic and lead evenly the emotional tendency of the user to appear polarization.

The main difference between opinion leaders, trolls and controversial figures is the proportion of supporters and opponents. Therefore, identifying the emotional tendencies correctly can be necessarily discriminate these influencers. A lot
of in-depth research on the analysis of emotional tendencies have been done by scholars, on which the results are good [16], [17]. However, the relativity of emotional expression has been ignored by these research. For example, when analyzing the forum post, such as “Num89: Landlord, I do not agree with Xin Liu either!”, the result of the traditionally sentiment analysis is that the emotional holder “Num89” opposes the user object “Landlord”, but it is obvious that the emotional tendency of the user object “Landlord” is support. If the relativity of emotions is taken into account, the result is that “Num89” opposes the topic object “Xin Liu” while supports the user object “Landlord”. Accordingly, this paper proposes an emotional transforming algorithm, which transforms the emotional tendency of the topic objects into the user objects.

The structure of the topical influencer discrimination model is shown concisely in Figure 1, in which “+” represented the positive emotional tendency. The first step is to manually annotate some reply contents to train the BiLSTM-CRF model in order to automatically label the sentimental elements. Then, the emotional matching and transforming algorithms are used to accurately identify the relative emotion between users. Meanwhile, the response relationship between users is extracted to form a social network and the rank of influencers is calculated by the multi-centrality algorithm and the quantified user behavior characteristics. The final step is to divide the influencers into opinion leaders, trolls, and controversial figures based on the combination of relative emotional tendency and influencers ranking.

The main contributions of this paper are illustrated as follows.

1. Proposed the Emotional Matching and Emotional Transforming algorithms to transform the emotional tendency of the topic objects into the user objects.
2. Divided topical influencers into opinion leaders with a high degree of identification, trolls who are opposed by the majority, and controversial figures.

The remainder of the paper is organized as follows. Related work is discussed in Section II. Our approach to discriminate influence users leveraging on relative emotion between users is described in Section III. Section IV describes the experiments and the results, and Section V concludes the paper.

II. RELATED WORK

Discriminating and further dividing the influencers in the topic into opinion leaders, trolls, and controversial figures requires a variety of models and methods. Firstly, a user influence ranking model needs to be built to dig out the influencers in the topic. Then, the emotional tendency expressed by the post will be analyzed, and the emotion expressed to the topic object is required to be transformed into the emotion expressed to the user object. Finally, the influencers need to be divided into three types by combining the social networks with the relative emotion between users. At present, some researches about user influence model, text sentiment analysis model, and method of combining social networks with relative emotions have been conducted. The research progress in this respects is illustrated as follows.
A. USER INFLUENCE MODEL

Currently, the user influence models mainly studies on social structure, user interaction, and emotional polarity. [18]–[21] are based on social structure, leveraging the measurement index such as degree centrality, betweenness centrality, and closeness centrality respectively to measure user influence. However, measuring the user influence with only one measurement index not able to fully reflect the influence of the node. Li et al. [22] found influencers based on behavior characteristics between users, studied on the change in the influence of followers on their interviewees and exploited the user interaction to propose a new influencer ranking method. Cha et al. [23] compared three methods of influence measure based on followers relationship in Twitter and interactive contents of users: “Indegree”, “Retweets”, “Mentions”. They found that it is not necessary to consider the “Indegree” when discriminating influencers based on user interactive information. Lang et al. [24] analyzed the emotional polarity contained in the contents posted by users and proposed an opinion leader discovery algorithm based on the emotional tendency, but failed in considering the relative emotion of the users and further divide the influencers. Nowadays, user influence is reflected in all aspects of social networks. Influential users cannot be identified accurately if only analyzing any aspect of social structure, user interaction and emotional polarity. Therefore, this paper combines multi-dimensional features such as social topology, behavior characteristics and relative emotion of users to identify influencers, and further divides influence users into opinion leaders, trolls and controversial figures.

B. TEXT SENTIMENT ANALYSIS

At a word, the relative emotion of the user is that the sentiment analysis not only extracts the emotional words and identifies emotional tendencies in the contents, but also mines the emotional objects [25], [26], and then discriminates the emotional object corresponding to the emotional word whether a user object or a topic object. The analysis found that there are few studies on mining objects from the comments of Weibo, Forum, and Blog. Wu et al. [27] extracted objects from the comments of the Forum and analyzed the emotions of the comments for classifying the emotional tendency of objects. Although the research explored the objects in the comments, the objects are only the topic objects and the user objects should be considered. With the development of deep learning, the task of mining emotional objects has gradually achieved good results. Zhao et al. [28] presented that deep learning has emerged as an effective means for solving sentiment classification problems recently in which a neural network intrinsically learns a useful representation automatically and proposed a novel deep learning framework for the sentiment classification of product comments. According to these research, this paper proposes a method of text sentiment analysis. The BiLSTM-CRF model [29] is used to identify the sentimental elements in the comments and further work about emotional matching and transforming algorithms are calculated to determine the emotional tendencies of emotional holdings to user objects.

C. THE CORRELATION BETWEEN SOCIAL NETWORKS AND RELATIVE EMOTIONS

The emotion varies and propagates with the spatial and temporal information of individuals through social network [30]. Some research demonstrates that the friendship in the social network contribute to the classification of emotional polarity [31]. Majority of scholars use friendship as a connection between users to build social networks. However, Pozzi et al. [32] believed that the assumption about the friendship relations does not reflect the real world where two connected users could have different opinions about the same topic. In order to overcome these shortcomings, an approval network is proposed, which consists of users and approval relations that formed when responding in Twitter. As the study went on, Kaewpitakkun et al. [33] thought that the previous method of combining social network to classify emotional tendencies only focuses on explicit links, such as fans, mentions, and replies, while some networks do not include the explicit link. In their research, an implicit link was defined, in which users had a link when they expressed similar emotions on the same topic and experiments showed that this method effectively increases the accuracy of sentiment classification. The emotional matching and transforming algorithm in this paper combines a directed social network topology formed by users and reply relationships. In this topology, if two linked users have no explicit emotional tendency, the emotion of the two user objects are judged by their emotional tendency of the same topic object.

III. TOPICAL INFLUENCERS DISCRIMINATION MODEL

The design idea of the influencer discriminant model is shown in Figure 2 below. Obviously, the ultimate aim of the influencer discriminant model is to identify opinion leaders, trolls, and controversial figures in a topic. To reach this objective, our priority is to clean and pretreat the data, which has two sub-tasks. One of the sub-tasks is to manually label the emotional elements in the posts, sentiment words and topic target set to train the BiLSTM-CRF model, and the other sub-task is to extract the reply relationships of users to construct the social networks. Then, we utilize the BiLSTM-CRF model to identify the emotional elements in the posts, and take advantage of the emotional matching and emotional transforming algorithms to calculate the relative emotional tendencies between users. Meanwhile, we calculate the rank of influencers based on quantified user behavior characteristics and the multi-centrality. Finally, it is necessary to discriminate the influencers by combining the emotional tendency and influence ranking. It is noteworthy that we exploit the emotional elements to supplement reply relationship and combine the reply relationship to discriminate the relative emotion between users, by comparing the emotional tendencies expressed by the users to the topic objects. The specific methods will be elaborated in section III-A, section
III-B, and section III-C respectively.

A. IDENTIFY RELATIVE EMOTIONAL TENDENCY OF USERS

The opinion leader is an influencer supported by the majority. While, the troll is the influencer who is opposed by the public. The controversial figure is the influencer who gets half supporters and half opponents. Therefore, accurately identifying the emotional tendency of the user is the basic condition for influencer discrimination. Firstly, the trained BiLSTM-CRF model is used to label the sentimental elements such as emotional words, user objects and topic objects in the content of posts. Then the above sentimental elements are combined and matched, and there are two types of matching results: < emotional words, user objects >, < emotional words, topic objects >. Finally, the sentimental elements matching results < emotional words, topic objects > are transformed to obtain emotional tendencies between user objects and emotional holders.

1) Label sentimental elements automatically

The study found that the use of sentimental words to classify emotional tendency is hard to fulfill the needs of the relative emotion discrimination task. Therefore, scholars have proposed the task of sentimental element extraction [34], [35]. The emotional element extraction task is also called fine-grained opinion mining, and most of its related work is to do sequence labeling. At present, the combination of deep learning and machine learning models such as BiLSTM-CRF has achieved good results in sequence labeling tasks [32]. The sentimental elements in this paper are labeled by the IOB2 labeling system which was used in the Bakeoff-3 [33]. B-SUB and I-SUB represent emotional holders (subjects), B-EMO and I-EMO indicate emotional words, B-UOBJ and I-UOBJ denote user objects, B-TOBJ and I-TOBJ mean topic objects, and O stands for other words that are not belonging to emotional elements, as shown in Figure 3.

First, vectorize the words and features in the post text and denote the sentence $S = (w_1, w_2, ..., w_n)$ containing $n$ words as $x = (x_1, x_2, ..., x_n)$, where $x_i$ represents the id of the $i$-th word in the dictionary. Then, input $x = (x_1, x_2, ..., x_n)$ into the BiLSTM-CRF model, and the architecture of the BiLSTM-CRF algorithm model is shown in Figure 4.

The first layer of the model is the look-up layer, which utilizes one-hot encoding to get the corresponding word vector for each word in the sentence. The second layer is the BiLSTM layer, in which word vectors $H_t = [F_t; B_t]$ are used as the input of BiLSTM, and the forward sequence of hidden states $(F_1, F_2, ..., F_t)$ and the backward sequence of hidden states $(B_1, B_2, ..., B_t)$ are calculated respectively. Then splice the sequences of hidden states to obtain a complete sequence of hidden states $(H_1, H_2, ..., H_t)$. The third layer is the CRF layer. For a sequence of labels $y = (y_1, y_2, ..., y_n)$ whose length is equal to the length of the sentence, the score of $y$ as
the label of the word $x$ is shown as follows.

$$\text{score}(x, y) = \sum_{i=1}^{n} P_{i,y_i} + \sum_{i=0}^{n} T_{y_i,y_{i+1}}$$  \hspace{1cm} (1)

Where $P$ is the score matrix output by BiLSTM, and $P_{i,y_i}$ is the score of the $i$-th word in the sentence classified to the $j$-th label. $T$ is the transfer matrix of the CRF layer, and $T_{y_i,y_{i+1}}$ represents the score of the $i$-th label transferring to the $j$-th label. $y_0$ and $y_{n+1}$ are the initial and terminate states added to the header and the tail of sentence respectively, so $T$ is a $k + 2$ matrix. When predicting all possible label sequences $y$ of the sentence, the results are normalized by using Softmax, as shown in (2).

$$P(y|x) = \frac{e^{\text{score}(x,y)}}{\sum_{y \in Y} e^{\text{score}(x,y)}}$$  \hspace{1cm} (2)

Maximize the log-likelihood ratio of the correct label sequence during model training. In decoding, the Viterbi algorithm for dynamic programming shown in (3) is leveraged to obtain the final output sequence.

$$\hat{y}^* = \arg \max_y \text{score}(x, \hat{y})$$  \hspace{1cm} (3)

2) Emotional matching algorithm (EMA)

Due to randomness of the user post, the sentimental elements (words, topic objects, user objects) in the post may be null, this paper classifies the matching case of the emotional words and the emotional objects: 1. There are only sentimental words in the text, such as “Support”, “Oppose”, “Like” etc. These words directly express the emotional tendency between objects with a reply relationship. 2. The posts contain both sentimental words and user objects, such as “Landlord is right”!. The sentimental elements matching result is $< \text{emotional object: landlord}, \text{emotional word: right} >$. If there are sentimental words and emotional objects in the text and the emotional objects have both user objects and topic objects, the relative distance minimization algorithm is used to match the emotional objects with the emotional words.

The posts published by users with emotions are mostly like “Landlord is correct!””, “Support the landlord” and “Landlord, I do not agree with Xin Liu either.”. After analyzing the contents of posts, it has been found that the relative distance between sentimental words and emotional objects (user objects, topic objects) are as close as possible. So, the relative distance between the emotional words and the emotional objects can be minimized to obtain the result of emotional matching. Get the relative distances $D = \{<w_{u, obj}, w_{emo}>, ..., <w_{t, obj}, w_{emo}>\}$ between emotional words and emotional objects based on the context. If the minimum $D$ is $<w_{u, obj}, w_{emo}>$, the corresponding matching result is $<\text{user object, sentimental word}>$. On the contrary, the matching result is $<\text{topic object, sentimental word}>$. If the matching result obtained is $<\text{user object, sentimental word}>$, the relative emotional tendency between the users

<table>
<thead>
<tr>
<th>Algorithm 1: Emotional matching algorithm (EMA)</th>
</tr>
</thead>
</table>
| **Input:** 1. Sequence label set $L = \{<\text{SUB, User, Emo, O,Other, Topic, EMOP, EMON, O}>, <\text{I-" or "B-"}>\}$
| 2. Sequence labeling results for each text obtained by the BiLSTM-CRF model: $\text{Seqlabel} = \{<\text{char}, \text{tag}>,...,<\text{char}, \text{tag}>\}, \text{tag} \in L$
| 3. Punctuation set $P$ |
| **Output:** Results of emotional matching $R = \{<\text{Sub, Emo, User, Tend(Emo)}>\}$ |

1. **#Text segmentation**
2. For $<\text{char}, \text{tag}>$ in $\text{Seqlabel}$ do
3.   Cut text to $\text{Clauses}();$
5. **#Label of sentimental elements**
6. Perform a chunk operation on the IOB2 tag in the $\text{Seqlabel}$ to get the label block set $\text{ChunkL} = \{\text{SUB, User, TOBJ, EMO}, \text{O}, \text{EMO} \in \{\text{EMOP, EMON}\}$, and the label result $\text{Chunk label} = \{<\text{word}, \text{label}_i>, ..., <\text{word}, \text{label}_n>\}$ where $\text{word}_i \in \{\text{Sub, User, Topic, Emo, Other}\}$ and $\text{label}_i \in \text{ChunkL};$
7. **#Identification of emotional tendency**
8. if $\text{EMO} \in \text{EMOP}$ then
9.   $\text{Tend(Emo)} \leftarrow 1;$
10. else
11.  $\text{Tend(Emo)} \leftarrow -1;$
12. **#Analysis of sentimental elements**
13. For clause which have labeled $\text{EMO}$ in $\text{Clauses}$ do
14.   if $\text{Chunk label} = \{<\text{Sub, SUB}>$, $<\text{Emo, EMO}>$,
15.      $<\text{Other, O}>\}$ then
16.     find $\text{User of Sub based on the reply relationship};$
17.     $\text{ User} \leftarrow \text{Tend(Emo)}$ of Sub;
18.   else if $\text{Chunk label} = \{<\text{Sub, SUB}>$, $<\text{User, UOBJ}>$, $<\text{Topic, TOBJ}>$, $<\text{Emo, EMO}>$,
19.      $<\text{Other, O}>\}$ then
20.     $\text{ Topic} \leftarrow \text{Tend(Emo)}$ of Sub;
21.     $\text{EmoTopic}() \leftarrow \{<\text{Sub, Tend(Emo), Topic}>;$
22.   else if $\text{Chunk label} = \{<\text{Sub, SUB}>$, $<\text{User, UOBJ}>$, $<\text{Topic, TOBJ}>$, $<\text{Emo, EMO}>$,
23.      $<\text{Other, O}>\}$ then
24.     Compare the minimum distance of $\text{UBOBJ}$ and $\text{TOBJ}$ to
25.     $\text{EMO}$ if $\text{OBJ}$ on the left side, the distance is represented by $D_{UB}$ otherwise represented by $D_{UB}$;
26.     if $D_{UB} \leq D_{UB}$ then
27.       $\text{-OBJ} \text{which is on the left side of EMO} \leftarrow \text{Tend(Emo)}$;
28.     else
29.       $\text{-OBJ} \text{which is on the right side of EMO} \leftarrow \text{Tend(Emo)}$;
30. if $\text{OBJ} = \text{TOBJ}$ then
31.   $\text{EmoTopic}() \leftarrow \{<\text{Sub, Tend(Emo), Topic}>;$
32.   find $\text{User of Sub based on the reply relationship};$
33.   Using $\text{Emotional Transforming Algorithm}$ to get $\text{User} \leftarrow \text{Tend(Emo)}$;
34. else
35.   $\text{User} \leftarrow \text{Tend(Emo)}$ of Sub;
36. add $\{<\text{Sub, Emo, User, Tend(Emo)}>\}$ into $R$. 

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.
can be directly discriminated. Otherwise, need to transform the emotion of the topic into the emotion of the user by the emotional transforming algorithm. The details are as shown in the emotional matching algorithm (EMA).

3) Emotional transforming algorithm (ETA)

Some user posts put emotion on the topic objects, which need to be transformed into the emotions on the user objects and topic objects may not be unique. This paper studies the case of (1) **one class topic object**, (2) **two opposite class topic objects**, and (3) **multiple opposite class topic objects**, and proposes corresponding transforming algorithm.

(1) **One class topic object**:

When users express emotional tendencies to the same type of topic object in the topic, only the emotional tendencies of $U_A$ (the user A) and $U_B$ (the user B) to Objj (the topic object) need to be identified. Then compare whether the two emotional tendencies are the same. If the same, the user relationship is support, and vice versa. (4) is an algorithm used to determine whether two users have the same emotional tendency toward the same type of topic object. (5) is a transforming algorithm that obtains a relative emotional tendency between users based on the emotional tendency of users toward the same type of topic object.

$$Sim(U_A, U_B, Obj) = \begin{cases} 
   1 & \text{if } e(U_A, Obj) = e(U_B, Obj) \\
   -1 & \text{else}
\end{cases}$$ (4)

$$e(U_B, U_A) = Sim(U_A, U_B, Obj)$$ (5)

Where $Sim()$ indicates the similarity of the emotional tendencies of the user $U_A$ and the user $U_B$ to the topic object Obj, with 1 representing similarity and −1 representing dissimilarity. $e()$ denotes the emotional tendency of the user $U_B$ to the user $U_A$, with 1 representing support and −1 representing opposition. As shown in Figure 5 (a), the user $U_A$ and the user $U_B$ respectively oppose and support the topic object Obj, and the above algorithms are used to derive the user $U_B$ against the user $U_A$.

(2) **Two opposite class topic objects**:

When users express the emotional tendency to the topic objects of the two opposite classes in the topic respectively, it is necessary to discriminate the emotional tendency of the user $U_A$ to the topic object Obj and the emotional tendency of the user $U_B$ to the topic object −Obj. Then compare whether the two emotional tendencies are the same. If the same, the user relationship is opposition and vice versa. (6) is an algorithm used to determine whether two users have the same emotional tendency toward the two opposite types of topic objects. (7) is a transforming algorithm, which is applied to obtain the relative emotional tendency between users.

$$Sim(U_A, U_B, Obj, −Obj) = \begin{cases} 
   1 & \text{if } e(U_A, Obj) = e(U_B, −Obj) \\
   -1 & \text{else}
\end{cases}$$ (6)

$$e(U_B, U_A) = −Sim(U_A, U_B, Obj, −Obj)$$ (7)

Where $Sim()$ indicates the similarity of the emotional tendencies of the user $U_A$ to the topic object Obj and the emotional tendencies of the user $U_B$ to the topic object −Obj, with 1 representing similarity and −1 representing dissimilarity. $e()$ denotes the emotional tendency of the user $U_B$ to the user $U_A$, with 1 representing support and −1 representing opposition. As shown in Figure 5 (b), a topic has topic objects Obj and −Obj in opposite states. That is, support for the topic object Obj is against −Obj. The user $U_B$ does not directly evaluate the user $U_A$, but expresses a positive emotional tendency to the topic object −Obj. At the same time, the user $U_A$ also expresses a positive emotional tendency to the topic object Obj. Using the traditional sentiment analysis method to determine that the user $U_B$ supports the user $U_A$, which is the opposite of the real situation. The result obtained by the transforming algorithm is that the user $U_B$ against the user $U_A$.

(3) **Multiple opposite class topic objects**:

The sentiment analysis of multiple topic objects is shown in Figure 5 (c). Since the opposite relationship between the topic objects is unclear when users express emotional tendencies to multiple opposite topic objects in the topic, it is impossible to accurately transform the emotional tendency of the emotional holder to the topic object into the user object. Therefore, this paper prunes the situation of multiple opposite class topic objects.

After the analysis of the above three situations, this paper proposes a user relative emotion transforming algorithm (ETA).

### Algorithm 2: Emotional transforming algorithm (ETA)

**Input:** 1. Topic objects $\{Objj, n \in [1, 2, 3, ...]\}$ and the relationship (opposite or consensus) between them 2. Emotional sets of topics: $\text{Emo2Topic}() = \{< \text{Sub}, \text{Tend}(\text{Emo}), \text{Topic}> \}$

**Output:** Results of emotional transforming $\text{User } \leftarrow \text{Tend}(\text{Emo})$

1. Get $e(U_A, Obj_j)$ (the emotional tendency of Sub to Obj) and $e(U_B, Obj_j)$ (the emotional tendency of User to Obj) by matching $U_A$ and $U_B$ with Topic in Emo2Topic() respectively, $\{Obj_j, Obj_j \in \text{Topic}\}$

2. **If in the case of a single class topic object**
   3. if $Obj_j == Obj_j == Obj_j$ then
      4. if $e(U_A, Obj) == e(U_B, Obj)$ then
         5. $\text{Tend}(\text{Emo}) \leftarrow 1$
         6. else
         7. $\text{Tend}(\text{Emo}) \leftarrow -1$
   8. **If in the case of two opposite class topic objects**
   9. else
      10. $Obj_j \leftarrow Obj_j$
      11. $Obj_j \leftarrow -Obj_j$
      12. if $e(U_A, Obj) == e(U_B, -Obj)$ then
         13. $\text{Tend}(\text{Emo}) \leftarrow 1$
         14. else
         15. $\text{Tend}(\text{Emo}) \leftarrow -1$
   16. return $\text{User } \leftarrow \text{Tend}(\text{Emo})$
and the edge with the arrow points to the parent node.

users is represented by the connected side between the posts, the post published by the user. The reply relationship between users is constructed based on the user posts in each topic, thus it is vital to analyze the post characteristics of the user, including the post structure and other users mentioned in the posts. At present, most web portals such as Weibo, BBS and post bars adopt a tree structure in the process of organizing user posts. Each post can be replied by many posts, but one user can only reply to one post. The tree structure contains the post contents and the reply relationships between users, which is of significance to the construction of user topology network in social networks. The user reply relationship topology can be constructed by mapping the user reply relationship tree structure to the social network. Part A of Figure 6 shows the user reply relationship tree structure in the forum, and Part B of Figure 6 shows the user reply relationship topology.

B. BUILD A USER RESPONSE NETWORK

The topology of the user reply relationship network is constructed based on the user posts in each topic, thus it is vital to analyze the post characteristics of the user, including the post structure and other users mentioned in the posts. At present, most web portals such as Weibo, BBS and post bars adopt a tree structure in the process of organizing user posts. Each post can be replied by many posts, but one user can only reply to one post. The tree structure contains the post contents and the reply relationships between users, which is of significance to the construction of user topology network in social networks. The user reply relationship topology can be constructed by mapping the user reply relationship tree structure to the social network. Part A of Figure 6 shows the user reply relationship tree structure in the forum, and Part B of Figure 6 shows the user reply relationship topology.

C. INDEX OF USER INFLUENCE RANKING

1) The multi-centrality

Users in online networks exert different influence during the process of information propagation. Analyzing the topology of social networks can investigate the influence strength of user on their neighbors [36]. Centrality is mainly used to measure the importance of nodes in social networks. Classical centrality algorithms have Degree centrality, Betweenness centrality, and Closeness centrality. Degree centrality defines the centrality of the node as the node degree, which is the degree of a node connected to all other nodes [18], [37]. A node with a greater node degree means that this node is connected to more nodes and has more influence in the network. Betweenness centrality [19], [20], [38] can identify the node that plays an indispensable role in the process of information dissemination. Closeness centrality [21], [39], [40] defines the centrality of the node as the proximity, and the smaller the value of the average distance, the more autonomous the node is in the process of information dissemination. This paper proposes a Multi-centrality algorithm to identify users with a greater influence in the social network topology. The Multi-centrality algorithm calculates the influence of the user by assigning different weight values to Degree centrality, Betweenness centrality, and Closeness centrality. The expression of the Multi-centrality is shown as (8).

\[
C_M(N_i) = C_D(N_i) \ast \omega_D + C_B(N_i) \ast \omega_B + C_C(N_i) \ast \omega_C \tag{8}
\]

To compare the influence of users in the same topic, the maximum normalization is performed as shown in (9).

\[
C_M(N_i)^* = \frac{C_M(N_i)}{\max(C_M)} \tag{9}
\]

Where \( N_i \) represents a set of nodes except node \( i \), \( C_D(N_i) \) denotes the Degree centrality of the node \( i \), \( C_B(N_i) \) denotes the Betweenness centrality, \( C_C(N_i) \) denotes the Closeness centrality of the node \( i \), and \( \omega_D, \omega_B, \omega_C \) are the weight values of each centrality in the Multi-centrality. The weight value of each centrality \([\omega_D, \omega_B, \omega_C] = [0.5882, 0.2942, 0.1176]\) is obtained by using Saaty [41], which gains the relative importance level between the measured indicators according to the cognitive habits and judgment ability of humans.

FIGURE 5: Emotional transforming (a) in the case of single topic object, (b) in the case of two topic objects, and (c) in the case of multiple topic objects.

FIGURE 6: The process of the user reply relationship tree structure mapping to reply relationship topology.

The node in the user reply relationship tree represents the post published by the user. The reply relationship between users is represented by the connected side between the posts, and the edge with the arrow points to the parent node.

VOLUME 6, 2019

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.
2) The quantified user behavior characteristics

Today, social networks are essentially enriched so that users can behave more diversity on social networks. Therefore, to analyze the users in current complex social networks by using traditional topology cannot achieve the desired results, user behavior characteristics needs to be considered, such as forward rate, reply rate, and like rate. The mining of user personalized information such as social content and user social behavior can not only describe the influence of users more specifically but also measure the intensity of influence between users in more detail. The way to quantify the behavioral characteristics of other users to the user $u_i$ is as shown in (10).

$$f_{ui} = \frac{n_{i+} + n_{ic} + n_{ic}}{\max(n_{i+} + n_{ic} + n_{ic})} \quad (10)$$

Where $n_{i+}$ indicates the number of forwarded posts of user $i$, $n_{ic}$ indicates the number of replied posts of user $i$, and $n_{ic}$ indicates the number of posts liked by other user of user $i$. $\max(n_{i+} + n_{ic} + n_{ic})$ is for normalization of the maximum value.

IV. EXPERIMENT AND ANALYSIS

A. EMOTIONAL TENDENCY IDENTIFICATION ANALYSIS

We train the BiLSTM-CRF model on two data sets to identify emotional elements and the matching algorithm and transforming algorithm are leveraged to achieve the final classification of emotional tendency. The first data set is the published Weibo data, including 100K data, which is mainly used to identify emotional elements such as emotional objects and emotional words in user post. The second data set is the data crawled from Tianya BBS, including the 5850 data about a topic, which is mainly used to mine the emotional holders in the user posts. The emotional labels in this data sets are annotated manually.

We compare the performance of our approach with the following methods.

Naive Bayes (NB). [42] implies that a highly accurate and fast sentiment classifier can be built using a simple Naive Bayes model. NB also met a good result in finding the polarity of the ambiguous tweets [43].

Support Vector Machine (SVM). Another common sentimental classification technique is based on the Support Vector Machine. For unlabeled data, the SVM model achieves promising increases in accuracy [44].

Convolutional Neural Network (CNN). Recently, the classifiers based on convolutional neural network (CNN) achieve good performances in sentiment classification tasks. [45] propose the architecture of a Convolutional Neural Network that takes into account not only the text but also user behavior, which outperforms current baseline models.

Long Short-Term Memory (LSTM). The LSTM, proposed in [46], has become the state-of-the-art models for a variety of machine learning problems. After research, the LSTM models have achieved impressive performance in the sentiment classification task [47].

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>For the Negative Label</th>
<th>For the Positive Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>NB</td>
<td>0.681</td>
<td>0.6147</td>
<td>0.97</td>
</tr>
<tr>
<td>SVM</td>
<td>0.685</td>
<td>0.6209</td>
<td>0.95</td>
</tr>
<tr>
<td>CNN</td>
<td>0.609</td>
<td>0.5935</td>
<td>0.692</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.611</td>
<td>0.5952</td>
<td>0.694</td>
</tr>
<tr>
<td>Bi-CEE</td>
<td>0.849</td>
<td>0.8262</td>
<td>0.884</td>
</tr>
</tbody>
</table>

Here, we processed a performance evaluation of each model in terms of correct prediction for the positive and negative labels, including Precision, Recall, F1 and Accuracy. Due to the purpose of this paper is to mine the relative emotion between user objects, the test data is focused and relevant. 80% of the test data is to discriminate the relative emotional tendency between users, and 20% of the test data is to identify the emotional tendency of the topic. The results are showed in Table 1. Firstly, the accuracy of all sentimental classification methods demonstrates that the best classification performance accuracy is achieved by our model Bi-CEE([Bi-LSTM + CRF + EMA + ETA]). Then, NB has the highest recall for negative label and the lowest recall for negative label, which indicate that NB is unstable. Although the recall for negative label and the precision for positive label are lower than NB and SVM, the model proposed by this paper is able to give us a stable performance. Finally, the poor performance of CNN and LSTM are not surprised since they are not take the relative emotion of users into account. These results confirm that when the analysis of the text needs to consider the relative emotions between users, the model proposed in this paper can improve the classification accuracy.

B. EMPIRICAL RESEARCH ON INFLUENCERS

In order to verify the necessity of the influencer discrimination model, we conducted an empirical analysis. From the Tianya BBS, in which including a large number of topics, we have crawled the experimental data about society, film, and entertainment. Firstly, we analyzed and processed all the posts in topics, and utilized the Multi-centrality and the quantified user behavior characteristics to identify the influencers. Then, combined with the relative emotions of users, opinion leaders, trolls and controversial figures in each topic were explored. Finally, we analyzed the ranks, levels, post numbers, and reply-post numbers of influencers in each topic to discover the different characteristics of opinion leaders, trolls and controversial figures.

Due to the disorder of data, this paper filtered out user comments data with unclear reply objects. The final data totally had four topics, consisted of 5967 comments and 3662 reply relationships. And the basic information of the...
TABLE 2: The basic information of Tianya BBS data sets

<table>
<thead>
<tr>
<th>TopicId</th>
<th>Topic1</th>
<th>Topic2</th>
<th>Topic3</th>
<th>Topic4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes(us)</td>
<td>987</td>
<td>434</td>
<td>651</td>
<td>692</td>
</tr>
<tr>
<td>Edges(reply relationships)</td>
<td>1300</td>
<td>716</td>
<td>844</td>
<td>802</td>
</tr>
<tr>
<td>Comment</td>
<td>2071</td>
<td>2115</td>
<td>1121</td>
<td>927</td>
</tr>
<tr>
<td>Emotional proportion(Pos/Neg)</td>
<td>0.7087</td>
<td>0.2907</td>
<td>0.5901</td>
<td>1.1000</td>
</tr>
</tbody>
</table>

1 The post corresponding to Topic1 are “Chongqing Bus Crash”.
2 The post corresponding to Topic2 are “GeJiang Murder Case”.
3 The post corresponding to Topic3 are “The Wandering Earth”.
4 The post corresponding to Topic4 are “MenygyaoXi Wrestling Event”.

TABLE 3: Top-6 influencers and their corresponding relative emotional tendencies

<table>
<thead>
<tr>
<th>Topic1</th>
<th>Topic2</th>
<th>Topic3</th>
<th>Topic4</th>
</tr>
</thead>
<tbody>
<tr>
<td>UserN</td>
<td>InfR</td>
<td>EmoT</td>
<td>UserN</td>
</tr>
<tr>
<td>1.5647</td>
<td>1</td>
<td>-1</td>
<td>蓝蓝的明天</td>
</tr>
<tr>
<td>1.5589</td>
<td>-1</td>
<td>RHB之再战江湖</td>
<td>1.7545</td>
</tr>
<tr>
<td>1.3165</td>
<td>-1</td>
<td>公民xh</td>
<td>1.2059</td>
</tr>
<tr>
<td>1.1274</td>
<td>-1</td>
<td>烈火山</td>
<td>0.7132</td>
</tr>
<tr>
<td>0.757</td>
<td>-1</td>
<td>天边的鸟1972</td>
<td>0.3816</td>
</tr>
<tr>
<td>0.5242</td>
<td>-1</td>
<td>口袋里的风</td>
<td>0.3594</td>
</tr>
</tbody>
</table>

Note: UserN(User Name), InfR(Influence Ranking), EmoT(Emotional Trendy).

We analyzed the data from the four topics. The influencers of the four topics are identified based on multi-centrality and user behavior characteristics, and the opinion leaders, trolls and controversial figures were further discriminated by analyzing the emotional tendencies of the other users to influencers. Table 3 illustrated the top-6 influencers in each topic. InfR represents the influence value of users, and EmoT denotes the overall relative emotion between users. For EmoT, 1 indicates that the emotional tendency is positive and the influencer is the opinion leader, 0 indicates that the positive and negative emotional tendency is the same and the influence user is the controversial figure, −1 indicates that the emotional tendency is negative and the influencer is the troll.

Then, we statistically analyzed the ranks, levels, posts and reply-posts of Top-20 influencers as opinion leaders, trolls and controversial figures in each topic to explore the characteristics of each influencer. Figure 7 shows the ranks of opinion leaders, trolls and controversial figures in each topic. Found that the ranks of opinion leaders are generally in the middle of the upper position, the ranks of the trolls are more evenly distributed in a topic, and the influential ranks of the controversial figures are mostly lower.

FIGURE 7: The ranks of opinion leaders, trolls and controversial figures in each topic.

FIGURE 8: The levels of opinion leaders, trolls and controversial figures in each topic.

Figure 8 demonstrates the levels of opinion leaders, trolls and controversial figures in the portal on each topic. It can be seen that the levels of opinion leaders and controversial figures are generally concentrated at level5-level10, while the levels of trolls are usually concentrated at level5-level15 and the troll in the experimental data has the highest level of 28.

The Figure 9 demonstrated the features of the posts and reply-posts of influencers. It can be found that the number of posts of opinion leaders is basically proportional to reply-posts. In general, the posts and reply-posts of trolls are inversely proportional, which means that the amount of reply-posts from users with relatively large posts is small. Controversial figures have a large number of posts and a small number of reply-posts because this type of influencers are willing to initiate a discussion.

An interesting phenomenon was found in the analysis of Table 2 and Table 3. Table 2 shows that the number of users participating in the discussion in Topic2 is about 0.6 times the number of users in Topic4, while the number of comments in Topic2 is 2.3 times that of Topic4. As seen in...
whether the emotional word and the emotional object can be matched in the emotional matching algorithm according to the distance minimization remains open to question. For that, future studies will focus on the following two directions. One of which is to improve the accuracy of the BiLSTM-CRF model by obtaining optimized training parameters, another is to match the emotion of the emotional word and the emotional object with the semantic dependency tree.

REFERENCES


[8] K. Grissa, “What makes opinion leaders share brand content on professional networking sites (e.g. LinkedIn, Viadeo, Xing, SkilledAfricans⋅")”, in Proc. IEEE Int. Conf. Digital Economy (ICDEc)., Apr. 2016, pp. 8-15.


Table 3, Top-6 influencers in Topic2 are mostly trolls, while the most influencer in Topic4 is a controversial figure and Top-6 influencers have the same number of opinion leaders and trolls. Furthermore, it is not hard to find that the greater influence of the opinion leaders or the more opinion leaders in the topic, the more posts with positive emotions in combination with Table 2 and Table 3. Research has concluded that opinion leaders can trigger users to post comments with positive emotion, trolls can improve the participation of users and controversial figures can increase the number of users participating in topic discussions.

V. CONCLUSION

In this paper, the BiLSTM-CRF model trained with the annotated manually data is leveraged to identify the sentimental elements, which are included in a specific topic. The matching algorithm and transforming algorithm are proposed to obtain the relative emotions of users. Influencers are discriminated as opinion leaders, trolls, and controversial figures respectively with three combination of methods include the relative emotions of users, the multi-centrality algorithm and the quantitative behavioral characteristics.

The results of contrast experiment suggest that when the analysis of the text needs to consider the relative emotions between users, the model proposed in this paper can improve the classification accuracy. Hence, we find that extracting the emotional elements of posts and performing emotional matching and transforming algorithms can improve performance of the emotional tendency discrimination. The empirical analysis demonstrates that diverse influencers play a different role in the topic. Though all of opinion leaders, trolls, and controversial figures contribute to increase the number of participants, the rate of emotional tendencies is widely divergent. Obviously, the more trolls there are, the more users with negative emotional tendencies in a topic.

Currently, the accuracy of the BiLSTM-CRF model labeling emotional elements needs to be improved. Moreover,


