Detecting False Information of Social Network in Big Data

Yi Xu[1(✉)], Furong Li[1], Jianyi Liu[1], Ru Zhang[1], Yuangang Yao[2], and Dongfang Zhang[3]

[1] Information Security Center, Beijing University of Posts and Telecommunications, Beijing, China
{xuyi0511, ronger1993071l, liujy, zhangru}@bupt.edu.cn
[2] China Information Technology Security Evaluation Center, Beijing, China
yaoyg@itsec.gov.cn
[3] First Research Institute of the Ministry of Public Security of PRC, Beijing, China
40319005@qq.com

Abstract. With the rapid development of social network, the information announced by this platform attracts more and more attention, because of the great harm brought by the false information, researching the false information detection of social network has great significance. This paper presents a model of social network false information detection, which firstly converting the information announced by social network into a three-dimensional vector, then comparing this vector with the three-dimensional vector converted by Internet events and calculating the similarity between social network and Internet, detecting the consistency of social network event and Internet event afterwards, finally gathering statistics and analyzing then we can get the similarity between social network event and Internet event, according to this, we can judge that the social network information is false or not.

Keywords: Social network · Information · Similarity · False · Detection

1 Introduction

At present, with the rapid growth of data volume, we enter the era of Big Data. While Big Data brings us extremely rich information, followed by this is a large number of false or outdated data, greatly reducing the application value of Big Data. The so-called "false information" is information that it is not authentic and maybe cause negative influence. Especially social network in Big Data, because announcing information by social network is open, anonymous, convenient and the spread of information is extensive and rapid, the problem of false information become more and more serious. For example, on April 24, 2013, hackers stole the Twitter of AP announced that the White House has been attacked by bomb attack, result in the Dow Jones Industrial Average Index plunged in a short time [1]. On March 2016, in microblog some say that somebody infect H7N9 because eating chicken, this led to people's panic for pandemic virus. This information may fool the cyber citizens, more seriously maybe cause social

S. Wang and A. Zhou (Eds.): CollaborateCom 2016, LNICST 201, pp. 642–651, 2017.
DOI: 10.1007/978-3-319-59288-6_65

unrest [2]. It follows that it is important that detecting false information to prevent the spread of false information.

At the moment, the research of social network false information detection has been attracting more and more attention, domestic and international scholars have achieved some research result in this aspect. Akritidis et al. [3] presents two mechanisms: BP-Index Mechanism and BI-Index Mechanism. BP-Index Mechanism in charge of assessing amount of blogs that users announced; BI-Index Mechanism judges bloggers' influence according to the number of links and comment in a certain time. Combining the two mechanisms to evaluate whether the bloggers are recently influential or recently productive. Zolfaghar and Aghaie [4] presents a method to forecast users' trust issues with Machine Learning, this paper considers that social credibility can divided into five aspects: relationship credibility, honor credibility, knowledge credibility, similarity credibility and individuation credibility, maps these five aspects into the characteristic sets consist of context, behavior and feature information of credibility network topological structure. However, this paper doesn't analyze forecast of users' trust issues in dynamic state. Calais et al. [5] proposes a real-time emotion analysis method based on transfer learning strategy, this method acquires users' prejudice to information in social network and sets this prejudice as the essential attribute of users' behaviors to translate into textual features, so that structure emotion classification model to realize emotion analysis. Castillo et al. [6] proposes a method of automatically assessing Twitter information credibility aim at social network typical representative Twitter. The paper analyzes text content to judge information credibility by means of users' emotion and opinion on the information. However, this method depends on manual work, its efficiency is low in the practical application. Qiao et al. [7] presents a trust calculating algorithm based on the social network users' context. This method divides the user trust into two parts: generated by familiarity and similarity among the different social network users. This paper also provides the specific computing method. In the research of information credibility, there are a few typical system can help user judge network information credibility from several angles, mainly include: WISDOM, reframeit.com, Honto Search, Blekko.com, etc. [8].

On the problem of social network false information detection, scholars have obtained some achievement from different view, however, the achievement is little and scattered so that there is no systematic theory, we still have lots of problems to solve. This paper presents a model of social network false information detection, comparing the social network event with Internet event, calculating the similarity between the two, then detecting their consistency, according to this, we can judge that the social network information is false or not. This paper also provides specific computational formula about these process. Comparing with the common false information detection methods, this paper follows the point of view in information itself instead of users, and compares social network information with Internet information, it can guarantee the information detection is authentic and authoritative. In this paper, the Sect. 2 presents the model of social network false information detection and explains every module of this model; the Sect. 3 designs contrast experiments to prove that the model in this paper is feasible; at last, makes summary of work and outlooks the future work.

2 Detecting False Information of Social Network

This paper firstly extracts social network information keywords to be query items put into Google, screens the Internet information from web pages return by Google; then based on webpages screen, extracting the webpages event, this paper converts the information into a three-dimensional vector $\vec{E}(e,\ t,\ p)$, in this vector, E represents event vector, e represents event name, t represents time, p represents place, so that we can get the three-dimensional vector $\vec{E}_i(e_i, t_i, p_i)$ converted by social network events and $\vec{E}_j(e_j, t_j, p_j)$ converted by Internet events; calculating the similarity between this two vector; detecting the emotional tendency consistency of social network event and Internet event; after gathering statistics and analyzing we can get the similarity between social network event and Internet event, according to this, we can get the result that the social network information is false or not. The model as shown in Fig. 1.

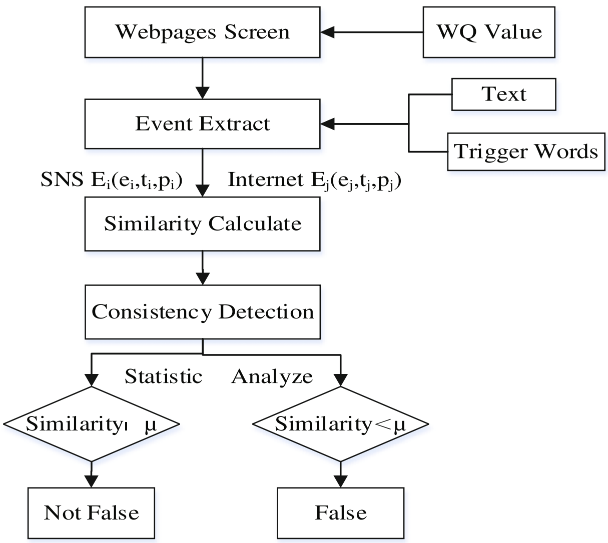

Fig. 1. The model of social network false information detection

2.1 Internet Information Screen

In order to screening the Internet information, this paper firstly extracts social network information keywords to be query items put into Google, aim at web pages return by Google, we screen the webpages by website quality value.

At present, the methods describing website quality are those used often: PageRank and Alexa. PageRank describes importance degree of website [9]. Alexa embodies popularity degree of website [10]. This paper comprehends PageRank and Alexa to rank the webpages, the definition of Website Quality Value as follows:

$$WQ(A) = \alpha \cdot PageRank(A) + \beta \cdot [1 - \frac{Alexa(A)}{10000}] \tag{1}$$

In this formula, A represents the website we need to calculate, 10000 is the quantity of Chinese website list which Alexa published, α, β ($0 \leq \alpha, \beta \leq 1 \&\& \alpha + \beta = 1$) is the weight of PageRank and Alexa.

2.2 Event Extract

In the social network information and Internet information, all of the events can be constituted of three elements, event name, time and place, this paper converts the information into a three-dimensional vector $\vec{E}(\mathrm{e}, t, p)$, in this vector, E represents event vector, e represents event name, t represents time, p represents place.

This paper based on website screen, extracts the webpages return by Google. The format of the web information in webpages mostly is HTML. So we present a method based on Chinese text density to extract main body. Firstly, taking out HTML tags from webpages and reserving blank position, the left is text denoted by Ltext totals L lines. Dividing pieces turn down with k spacing from the first line denoted by $Block_i$, counting the total number of characters $SChar_i$ and the number of Chinese characters $CChar_i$ in the $Block_i$, the density of Chinese characters denoted by Den_i as shown in formula 2.

$$Den_i = \frac{CChar_i}{SChar_i} \tag{2}$$

The text is divided into L-k pieces, we draw the distribution image using [1, L-k] as horizontal axis and Den_i as vertical axis. There inevitably are a large number of Chinese characters in the main body of webpages, it lead to sudden rise of Chinese characters density. What we need to do is confirm the point of sudden rise and sudden fall to delimit a region with high Chinese characters density, the text in this region is the main body of webpages.

After obtaining the main body of webpages, this paper search the sentences which contain trigger words. The so-called "trigger word" is a word that describes the status of one event, it represents the event occurred so that it can commendably deciding the type of event, such as "happen" and "outbreak" [11]. In the natural language processing, the context field is from −8 to +9 apart from Core Words can contain more than 85% amount of information [12], we call this field effective range of event sentences. Because the sentence between two nearest periods from trigger word mostly in this effective range, so this paper put the sentence between two nearest periods from trigger word as the Internet event sentence. Aim at event sentences, using the method of matching rule to extract event information, based on the ICTCLAS, this paper defines the matching rule "\{ }/t" to extract time information. Similarly, defining the matching rule "\{}/ns" or "\{}/nsf" to extract place information. The process of extracting the event three-dimensional vector as follows:

Step 1: Segmenting the text and traversing the words, then searching the words whether belong to trigger words dictionary or not. If not, searching go on. If yes, putting the sentence between two nearest periods from trigger word as the event sentence.

Step 2: Using the method of matching rule to extract time information denoted by t in the event sentence; extract place information denoted by p; taking out t and p, the rest of the sentence is so called as event name denoted by e.

Step 3: Using the three extract elements to constituting the event three-dimensional vector $\vec{E}(e, t, p)$.

2.3 Similarity Calculate

In accordance with the event three-dimensional vector $\vec{E}(e, t, p)$, we compare social network event vector $\vec{E}_i(e_i, t_i, p_i)$ with webpages extract event vector $\vec{E}_j(e_j, t_j, p_j)$. The formula that how to calculate similarity value as follows:

$$Sim(\vec{E}_i, \vec{E}_j) = \cos(\vec{E}_i, \vec{E}_j) = \frac{\vec{E}_i \cdot \vec{E}_j}{|\vec{E}_i| \cdot |\vec{E}_j|} \tag{3}$$

Taking into consideration that there are lots of Chinese words in the similarity calculating proposed by this paper and the semantic information similarity these two key factors, this paper uses the method of calculating similarity based on HowNet Semantic Information. Using this method to obtain the event name similarity, time similarity and place similarity. The sentence similarity can translate into word similarity. In the HowNet, the word is expressed by one or more concepts and the concept is explained by primitive. So calculating word similarity can translate into calculating primitive similarity. In this paper, supposing two sentences are expressed as s_1 and s_2, they respectively include two words w_1 and w_2, the two words respectively include two concepts c_1 and c_2, this two concepts respectively include two primitive p_1 and p_2. Therefore, calculating similarity between s_1 and s_2 can translate into calculating similarity between p_1 and p_2. In the HowNet, the formula that how to calculate similarity between two primitive as follows:

$$Sim(p_1, p_2) = \frac{\delta}{d + \delta} \tag{4}$$

In this formula,p_1 and p_2 represent two primitive, d represents the distance between p_1 and p_2 in primitive gradation system, δ is an adjustable parameter [13].

In the HowNet, words are divided into notional words and function words, the similarity between notional words and function words is 0. Calculating function words similarity just need to calculate the primitive similarity. The function words concepts are mainly represented by four parts of primitive: $Sim_1(p_1, p_2)$, $Sim_2(p_1, p_2)$,

$Sim_3(p_1, p_2)$, $Sim_4(p_1, p_2)$, they represent four kinds of primitive, so the formula of calculating similarity between notional words as follows:

$$Sim(\mathrm{c}_1, c_2) = \sum_{i=1}^{4} \lambda_i \prod_{j=1}^{i} Sim_j(p_1, p_2) \tag{5}$$

In this formula, $\lambda_i(1 \leq \mathrm{i} \leq 4)$ is an adjustable parameter and $\lambda_1 + \lambda_2 + \lambda_3 +$ $= 1$, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$, it reflects the impact to ensemble similarity of sim_1 to sim_4 is decreasing.

Supposing two words w_1 and w_2, w_1 has m concepts: c_{11}, c_{12} … $\mathrm{c}_{1\mathrm{m}}$, w_2 has n concepts: c_{21}, c_{22} … $\mathrm{c}_{2\mathrm{n}}$, defining the similarity between w_1 and w_2 is:

$$Sim(w_1, w_2) = \max_{\substack{i = 1, 2...m \\ j = 1, 2...n}} Sim(c_{1i}, c_{2j}) \tag{6}$$

Supposing two sentences $s_1(w_{11}, w_{12}...w_{1m})$, $s_2(w_{21}, w_{22}...w_{2n})$, w_{1i} and w_{2j} respectively represent the word contained within s_1 and s_2, so the formula that how to calculate similarity between sentences s_1 and s_2 as follows:

$$Sim(\mathrm{s}_1, \mathrm{s}_2) = \frac{\sum_{i=1}^{m} \max[\sum_{j=1}^{n} sim(w_{1i}, w_{2j})]}{m} \tag{7}$$

The process of calculating similarity between social network event and Internet event as follows:

Step 1: After extracting the social network event vector $\vec{E}_i(e_i, t_i, p_i)$ and Internet event vector $\vec{E}_j(e_j, t_j, p_j)$, aim at the e_i and e_j, t_i and t_j, p_i and p_j, calculating event name similarity $SimE(e)$, time similarity $SimT(t)$, place similarity $SimP(p)$ by formula 7.

Step 2: Setting μ as the threshold value of judging the information is false or not, if $Sim(\vec{E}_i, \vec{E}_j)$ less than μ, it can be say that the event is false. Taking into consideration that if any one of event name, time and place differ greatly, it will heavily affect the authenticity of event. So that this paper stipulates that whichever the three values less than μ, we say that $Sim(\vec{E}_i, \vec{E}_j) = 0$. If $SimE(e)$, $SimT(t)$ and $SimP(p)$ are all not less than μ, commanding $\vec{E}_i(e_i, t_i, p_i)$ is the unit vector (1, 1, 1) and $\vec{E}_j(e_j, t_j, p_j)$ is $\vec{E}_i(SimE(e_i, e_j), SimT(t_i, t_j), SimP(p_i, p_j))$. Using formula 3 can get the similarity between these two events; on the contrary, $Sim(i, j) = 0$.

2.4 Consistency Detection

After receiving the result of the similarity between social network information and Internet information, detecting the consistency of social network information and Internet information, detecting whether these two event information express same semantic information or not, mean to detect the emotional tendency in the two events is consistent or not.

At present, the method of calculating similarity not consider that oriented words may affect the sentences information, for example, "I very like Chinese football team" and "I very dislike Chinese football team", the similarity between two sentences will cause error comparing with the fact using the present similarity algorithm. Therefore, this paper detects the consistency of emotional tendency in the event information.

Acquiring the emotional tendency of sentence mainly through analyzing the commendation or derogation of the words, firstly segmenting the sentences and traversing the words, then extracting the adversative and appraisable words refer to the Antisense Primitive Dictionary and Adversative Words Dictionary so that judge the emotional tendency of the sentences. The specific computational process as follows:

Step 1: Segmenting sentences and traversing the words. Searching for the adversative words, if find that mark the sentences before the adversative words as 1 and mark the sentences after the adversative words as 2, giving up the sentence 1 and analyzing the sentence 2; if not find that analyze the whole sentences.

Step 2: Searching for the appraisable words and commanding the emotional tendency value of the sentences is 0, if find the commendatory words, mark the value +1; if find the derogatory words, mark the value −1;

Step 3: Working out the emotional tendency value of the social network sentences and Internet sentences denoted by Jud, comparing the social network sentences Jud with the Internet sentences Jud, if properties of positive and negative in social network sentences and Internet sentences are consistent, the emotional tendency of two sentences can be considered is consistent, whereas is not consistent;

Step 4: The Internet sentences can be considered as passing the consistency detection which emotional tendency is consistent with social network sentences, retaining the similarity between two events; if the sentences which emotional tendency is not consistent with social network sentences, they can be thought that not pass the consistency detection and the social network information is considered as false information, the similarity between two events is 0.

2.5 Statistics and Conclusion

This paper chooses the top n webpages return by Google ranked by WQ Value to analyze, it means that choose the top n webpages respectively denoted by $\vec{E}_1, \vec{E}_2, \vec{E}_3 \ldots \vec{E}_j \ldots \vec{E}_n$. Comparing the similarity between social network event vector $\vec{E}_i(e_i, t_i, p_i)$ and Internet event vector $\vec{E}_j(e_j, t_j, p_j)$ according to Sects. 2.3 and 2.4 can obtain the similarity value $Sim(\vec{E}_i, \overrightarrow{\vec{E}_j})$ denoted by $Sim(i,j)$, sum to $Sim(i,1), Sim(i,2) \cdots Sim(i,j) \cdots Sim(i,n)$ n similarity data. The formula that how to calculate similarity between social network event i and Internet events as follows:

$$Sim(i) = \sum_{j=1}^{n} \frac{WQ(j)}{\sum_{j=1}^{n} WQ(j)} Sim(i,j) \tag{8}$$

If $Sim(i) < \mu$, the social network event i is false;

If $Sim(i) \geq \mu$, the social network event i is not false.

3 Experiment and Analysis

In order to verify the effect of the social network false information detection method propose by this paper, we design contrast experiments to compare the method propose by this paper with pre-existing methods and use the events announced by social network which have already taken place. This experiment mainly adopts the following performance indexes to measure system performance so as to objectively measure the performance of false information detection method:

$$\text{Precision:} \qquad P = \frac{\textit{The number of correctly detect false information}}{\textit{The number of the test information}} \times 100\% \tag{9}$$

$$\text{Recall:} \qquad R = \frac{\textit{The number of correctly detect false information}}{\textit{The number of the actual false information}} \times 100\% \tag{10}$$

$$\text{F - Measure:} \qquad F = \frac{2PR}{P+R} \times 100\% \tag{11}$$

This paper selects the Sina microblog hot events data set as the experimental data set, on the basis of refuting rumors by Sina official, chooses top 100 web pages return by Google as the Internet comparative information. Outside the false information detection method (denoted by Method 1) that this paper propose, at present, the common false information detection methods mainly include: Judging the information authenticity by calculating credibility of users (denoted by Method 2); Carlos Castillo proposes the method of automatically assessing social network popular theme information credibility through users' emotion and opinion on the information (denoted by Method 3). Comparing with these two methods, the experiment result as shown in Table 1.

Table 1. The experiment result of different false information detection methods

Method	Precision	Recall	F-Measure
1	88.33%	93.33%	90.76%
2	69.01%	70.04%	69.52%
3	86.10%	86.00%	86.05%

From the contrastive experiment data as shown in Table 1, the precision and recall of the method this paper proposed has a certain degree of increase comparing with pre-existing methods. Comparing with Method 2, this paper follows the point of view in information itself instead of users, this mainly considers about that the high-credibility users are hard to avoid announcing some unsupported information follow the trend, the credibility of users can not strictly correspond to the authenticity of information they announced. Comparing with Method 3, this paper compares social network information with Internet information to guarantee the contrastive information is authentic and authoritative, it makes the method this paper proposed has more practicability.

In order to verify significance of consistency detection to increase false information detection property propose by this paper, we design contrast experiments to compare conducting consistency detection with not conducting consistency detection, the comparison result as shown in Fig. 2.

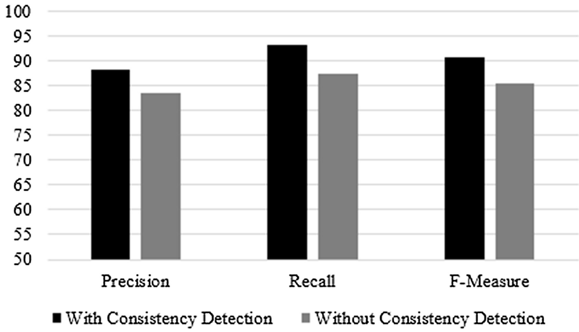

Fig. 2. Comparison of with or without consistency detection

From the contrastive experiment data as shown in Fig. 2, judging the emotional tendency of sentences and detecting the emotional tendency in the events is consistent or not can increase properties such as precision of false information detection. Thus it can be seen that the consistency detection module in the model proposed by this paper has great significance. This paper synthesizes advantage of pre-existing methods and presents a model of social network false information detection so that improves validity of false information detection, as a consequence, this paper has a certain feasibility.

4 Conclusions

With the rapid development of social network, announcing information by social network is open, anonymous, convenient and the spread of information is extensive and rapid, the information announced by this platform attracts more and more attention, because of the great harm brought by the false information, researching the false information detection of social network has great significance. This paper presents a model of social network false information detection, converting the information announced by social network into a three-dimensional vector, comparing this vector with the three-dimensional vector converted by Internet events, calculating the similarity between social network and Internet, detecting the consistency of social network event and Internet event, gathering statistics and analyzing then we can get the similarity between social network event and Internet event, according to this, we can judge that the social network information is false or not. This paper describes the essential flow and particular process of the model in detail, and conducts experiments with authentic social network data to verify whether the method this paper proposed is reasonable and effective or not. However, the method how to detect the information that can't searched out by Google is lacking, it need to be resolved in the later studies.

Acknowledgments. The author thanks the editor and reviewers for their suggestions to improve the quality of paper. This work was supported by the National Key Research and Development Program of China under Grant 2016YFB0800404 and NSF of China (U1536118) and NSF of China (U1433105).

References

1. Sina Technology. The Twitter of AP is attacked: announcing the false information that Obama has been attacked, 24 April 2013. http://www.ahwang.cn/zbah/20160319/1503617.shtml
2. Anhui Network: Somebody infect H7N9 because eating chicken in Huainan? Official: It is false information, 19 March 2016. http://tech.sina.com.cn/i/2013-04-24/08198273255.shtml
3. Akritidis, L., Katsaros, D., Bozanis, P.: Identifying the productive and influential bloggers in a community. IEEE Trans. Syst. Man Cybern. **41**(5), 759–764 (2011)
4. Zolfaghar, K., Aghaie, A.: A syntactical approach for interpersonal trust prediction in social web applications: combining contextual and structural data. Knowl.-Based Syst. **26**, 93–102 (2012)
5. Calais, G.P.H., Adriano, V., Wagner, M., Virgílio, A.: From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 150–158. ACM, CA (2011)
6. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the International Conference on World Wide Web (WWW), pp. 675–684 (2011)
7. Qiao, X.Q., Yang, C., Li, X.F., et al.: A trust calculating algorithm based on social networking service users' context. Jisuanji Xuebao Chin. J. Comput. **34**(12), 2403–2413 (2011)
8. Liu, T.: The problem of Internet information reliability, 01 December 2012. http://caai.cn/contents/49/102.html
9. Page, L., Brin, S., Motwani, R., et al.: The PageRank citation ranking: bringing order to the web (1999)
10. Sites, T.: Alexa Rank. The top, 500
11. Chen, Z., Ji, H.: Language specific issue and feature exploration in Chinese event extraction. In: Proceedings of the HLT-NAACL 2009, pp. 209–212. Omnipress, Madison (2009)
12. Lu, S., Bai, S.: Quantitative analysis of context field in natural language processing. Chin. J. Comput. Chin. Ed. **24**(7), 742–747 (2001)
13. Liu, Q., Li, S.: Word similarity computing based on How-Net. Comput. Linguist. Chin. Lang. Process. **7**(2), 59–76 (2002)