Information Credibility on Twitter in Emergency Situation

Xin Xia, Xiaohu Yang, Chao Wu, Shanping Li, and Linfeng Bao

Computer Science College, Zhejiang University
38 Zheda Road, Hangzhou, 310027, China
{xxkidd,yangxh,wuchao,shan,geniusblf}@zju.edu.cn

Abstract. Twitter has shown its greatest power of influence for its fast information diffusion. Previous research has shown that most of the tweets posted are truthful, but as some people post the rumors and spams on Twitter in emergence situation, the direction of public opinion can be misled and even the riots are caused. In this paper, we focus on the methods for the information credibility in emergency situation. More precisely, we build a novel Twitter monitor model to monitoring Twitter online. Within the novel monitor model, an unsupervised learning algorithm is proposed to detect the emergency situation. A collection of training dataset which includes the tweets of typical events is gathered through the Twitter monitor. Then we manually dispatch the dataset to experts who label each tweet into two classes: credibility or incredibility. With the classified tweets, a number of features related to the user social behavior, the tweet content, the tweet topic and the tweet diffusion are extracted. A supervised method using learning Bayesian Network is used to predict the tweets credibility in emergency situation. Experiments with the tweets of UK Riots related topics show that our procedure achieves good performance to classify the tweets compared with other state-of-art algorithms.

Keywords: Twitter, Bayesian Network, Sequential K-means, emergency situation, information credibility.

1 Introduction

Twitter, as a popular micro-blogging service, is playing a more and more important role in our social lives. It provides a fast and easy form of communication that enables the users to express their views, chat with other friends and share their status. Twitter allows the users to post and exchange 140-character-long information, which are known as tweets. Generally speaking, the tweets have the following characters: the use of "@user" syntax to remind other people to join the conversation, the use of hashtags (#) to mark the tweet's topic, and the retweets functionality to propagate the information faster and effective [1].

There are various channels for users to publish the tweets, such as sending the Email, sending the SMS and using the web-based service in PC or mobile phones. Therefore, Twitter accelerates the information propagation around the world via its rich client application and the compact 140-long-character tweets. Different from the

M. Chau et al. (Eds.): PAISI 2012, LNCS 7299, pp. 45–59, 2012.

Fig. 1. The tweets about the UK Riots

traditional media such as web portals, Twitter can disseminate the burst news directly from the news source at the first time. Figure 1 shows one example of tweets during the UK riots.

In emergency situation, Twitter has shown its power for the information diffusion [2].However, we found Twitter not only enables the effective broadcasting of valid news, but also the false rumors. For example, the Wenzhou motor car accident event in China, on July 23th, 2011. Most of the tweets are discussing the current status of the motor car accident, but one piece of tweet attracts most of interesting. The tweet is about "the reason of the motor car accident has been found. Two programmers who don't have the certificate from the government should take the great responsibility for this accident", and in a short time, this tweet is forward thousands of times, and millions of people complaint about that the government is shrinking its responsibility. Obviously, this tweet is a rumor, but in the emergency situation like the motor car accident, a lot of people choose to believe it.

The false rumor will mislead the emergency's attention, especially in the emergency situation when the public is fractious. If we just let those rumors propagate, the results are unexpected, and even cause the riots. In this paper, we mainly focus on the information credibility on Twitter in emergency situation, the main contribution of this paper is as follows:

1. We propose a novel Twitter Monitor model to monitor Twitter online based on the dynamic keywords filter and an unsupervised learning method to detect the emergency situation.
2. We propose a supervised learning method (mainly using Bayesian Network) to generate the classifier, and evaluate the proposed classifier with the tweets from Twitter.

The rest of this paper is organized as follows: In Section 2, we briefly outline the related works about the information credibility in Social Network. In Section 3, the dynamic keywords filter algorithm for the Twitter monitor, and the sequential K-means algorithm for emergency situation detection are addressed. In Section 4, we

collect the training dataset, extract the relevant features from the labeled tweets, and use supervised learning methods to generate the classifier. In Section 5, the experiment with the UK Riots topic-related tweets is provided. The final conclusions and future work appear in Section 6.

2 Related Work

The literatures on information credibility and the media's effect in emergency situation have long histories, various works focus on this subject. In this section, we provide an outline of researches that are most related to our works. This section is divided into two parts, one is the information credibility, and the other is about the Twitter's effect in emergency situation.

2.1 Information Credibility

Information Credibility in Traditional Media on the Internet. The traditional media on the Internet includes the various web portals, blogs and forums. Researches show that people trust the news on the online news portals as well as the other media, for example, the TV and newspaper [3]. Nowadays, Internet has become the most important channel for the young people in US to gain the current news, according to a survey in 2008.

Besides the online news portals, blogs and forums are considered as the less trustworthy. Similar to our work on information credibility on Twitter, the information in blogs and forums are faced the same credibility problem. A lot of researches pay attentions to this problem. [4] ranks the blogs credibility by exploit the verified content, it uses a variant of PageRank algorithm. [5] does some research on blog's credibility among the people with politically-interested, and it found that the people who rate the high credibility of blogs are the heavy blog users.

Spam Detection in Social Network. Spam is the use of electronic messaging systems to send unsolicited bulk messages indiscriminately[1] [6-8] .[6] describes a method to detecting the spams on Twitter. It extract the obvious features from the original tweets, such as the URL fraction of the tweets, the fraction of spam word of tweets, the number of hashtags of the tweets…etc, and analyses the significance for those features for spam detection. Lastly, the SVM classifier is used to detect whether a new arrived tweet is a spam. [7] detects the spams from an opinion mining view, which provides three types of spams for product reviews: the untruthful opinions, the reviews on brands only, and the non-reviews, and proposes a supervised learning method to spam detection.

Information Credibility on Twitter. [9] provides a whole framework to analyses the information credibility on Twitter. It uses TwitterMonitor [10] to monitor twitter and analyses the trend of Twitter. The crawled tweets are dispatched to the workers in Mechanical Turk[2] , to label the credibility of the tweets. Lastly, a J48 decision tree algorithm is used to make the final classification.

[1] http://en.wikipedia.org/wiki/Spam_(electronic)
[2] http://www.mturk.com

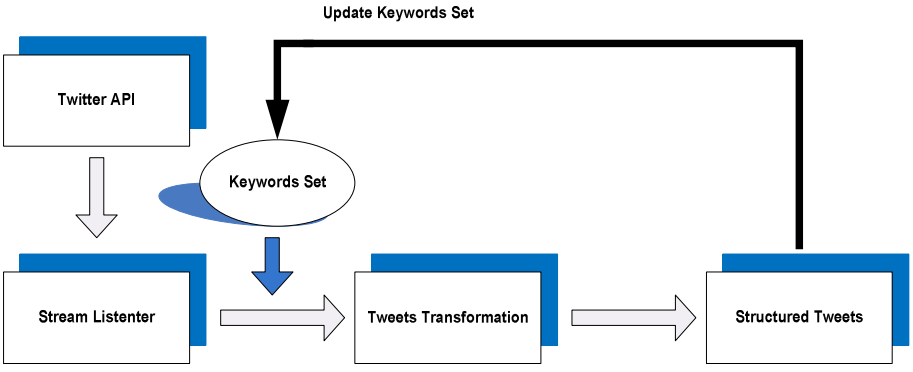

Fig. 2. The Architecture of Tweets Monitor Model

2.2 Twitter's Effect in Emergency Situation

Although most of the messages on Twitter are the conversation and chatter, people also use it to share some emergency information and to report the current news. Many researches show that Twitter has shown great power of influence during emergency situation, such as Earthquake [11], hurricanes [12], floods [2]…etc. [12] analyses the number of tweets ,the percentage of tweets contain the URL and the percentage of new users who have become low-active/inactive and active users in emergency situation comparing with the ordinary situation. [11] analyses the tweets during the Japanese Earthquake to produce the trace of earthquake, and compare with the real earthquake trace from the government. [13] analyses the whole Twitter community's activity during the period of Chilean earthquake in 2010.

3 Emergency Situation Monitor

We focus on the time-sensitive tweets, especially the tweets about current emergency event (such as terrorist attacking, earthquake, riots, etc). In this section, we describe the way we collect the emergency event related tweets and detect the emergency situation.

3.1 Automatic Related Tweets Collection

Currently, there are some tools for Twitter monitor, such as TwitterMonitor [10], which detects bursts and analyses the trends from the tweets. But our Tweets Monitor Model is a bit different from TwitterMonitor. Figure 2 shows the architecture of Tweets Monitor Model.

The Stream Listener module receives the data stream from Twitter stream, via Twitter API[3]. Stream Listener collects the original tweets information, and then those tweets are filtered by the Keywords Set.

[3] http://apiwiki.twitter.com/Twitter-API-Documentation

The Keywords Set contains the keywords related to the emergency event area we are interested, for example, if we are interested in the topic of terrorism event, we can set the Keywords Set with "terrorist, terrorism…etc"; if we are interested in the topic of natural disaster, we can set the Keywords Set with "earthquake, tsunami... etc".

Table 1. The meanings of the attributes from the twitter monitor component

Feature	Meanings
time	The time when the author sent the tweet, it is in the datatime format.
author	The author who sent the tweet.
citation	Identify whether this tweet cite another tweet (re-tweet). If this attribute is the same as author, then it is an original tweet; else it cites another guy's tweet.
content	The main text of the tweet.

The Keywords Set will change according to the word distance between different words. The procedure of updating dynamically updating keywords set is discussed later.

Tweets Transformation module makes the unstructured tweets information into structure information. The structured format of the tweets is:

$$\text{tweets} = (\text{time}, \text{author}, \text{citation}, \text{content}) \tag{1}$$

The meaning of those features is in table 1. For example, a tweet from John at 8:00:15 am, 8/2/2011, in the airport can be like the following:

$$\text{tweet}_i = (\text{"8:00:15 am, 8/2/2011 "}, \text{"John"}, \text{" John"}, \text{" What a beautiful day! @Lee "})$$

The tweets in the Structured Tweets are well-structured, and the dynamic keywords filter algorithm is used to update the Keywords Set. Figure 3 shows a heuristic dynamic keywords filter algorithm. The Word Distance is defined as for each two different words w_i, w_j, the distance is:

$$\text{distance}(w_i, w_j) = \frac{\text{The number of tweets } w_i, w_j \text{ both appear}}{\text{The number of total tweets}} \tag{2}$$

The whole Keywords Set contain two parts: the initial constant keywords set D and the dynamic keywords det D'. The initial constant keywords set D contain all the original words, for example, the set can be a collection of noun words such as {"terrorist", "terrorism"…etc}. For each noun (especially the name of places, such as

Algorithm 1. a heuristic dynamic keywords filter algorithm

Input : The initial Constant Keywords set D ,Dynamic keywords Set D', a user-specify threshold μ, the total number of tweets t

Output: The updated Keywords Set DUD'

For Each new tweets T from the Tweets Transformation

t=t+1

Words set $W = \text{Fetch Noun from } T.\text{Content}$

For Each Word w in W

For i=1 to |D|

If distance $(w, D_i) > \mu$

Add w to D'

For Each Word w in D'

If the distance of w and all the words in D are less than μ

Remove w to D'

Return DUD',t

Fig. 3. A heuristic dynamic keywords filter algorithm

Beijing, New York …etc) in the new arrived tweet, the distance between the noun and the words in D is computed. Then if the distance more than the user-specify threshold, those words add to the dynamic keywords set D'. And for each word w in D', if the distance between w and all the words in D are less than a user-specific threshold μ, then w is removed from D'.

3.2 Semi-automatic Emergency Situation Detection

How to detect the emergency situation from Twitter is one of the difficulties in this paper. One feature of the emergency situation is a sudden burst of tweets from Twitter. From this point, we propose an emergency situation detection algorithm. Figure 4 shows one example of emergency situation detection in 2-dimensionality feature space. The small circle represents the tweets posted in about 15 minutes. There are 3 clusters in Figure 4 using k-means algorithm. Notice that the left up cluster has the highest density and shortest diameter. After extracting the common element in this cluster, and analyzing by the experts, we can finally decide whether an emergency situation really happens.

The original tweet format in (1) need to be extended. The features include the Author-based features, Content-based features and Diffusion-based features.

The Author-based features include the author's name, the author profile's location.

The Content-based features include the tweets posting time, the address information (if not, just set the features to be null), the words which match the Keywords Set.

The Diffusion-based features include the re-tweet information (if it is a re-tweet, this feature is equal to the original author of the tweet, else just the same of the author's name).

For example, an author “Leon” who is in London, UK sent a tweet “I saw some young guys rob the Thinkpad retail shop in the London Street. So crazy they were! London is in a Riot! ” at 9:00:15 am, 8/23/2011, Then this tweet is transformed into the structured format:

$$\text{tweet} = (\text{“Leon”}, \text{”London, UK”}, \text{"9:00:15 am, 8/23/2011"}, \text{“London Street, London”}, \text{“Riot”}, \text{“Leon”})$$

With the above notation, the distance of two tweets is computed by the “XOR” operation for all the feature except the time feature is computed by the ordinary subtract operation. For the features with multiple-values (such as the words which match the Keywords Set), the minimum “XOR” operation is used. The definition of distance between two tweets is ($\otimes$ denotes the XOR operation):

$$\text{distance}(\text{tweet}_m, \text{tweet}_n) = \sqrt[2]{\begin{array}{c}(\text{name}_m \otimes \text{name}_n)^2 + (\text{loc}_m \otimes \text{loc}_n)^2 + \\ (\text{time}_m - \text{time}_n)^2 + (\text{add}_m \otimes \text{add}_n)^2 + \\ (\text{word}_m \otimes \text{word}_n)^2 \\ +(\text{re}_m \otimes \text{re}_n)^2\end{array}} \quad (3)$$

The emergency situation detection algorithm is working under two assumptions: 1. the tweets will burst in emergency situation; 2. the diffusion procedure of the emergency event is from the event source place and then the other places, and there are more people in the source place posting the tweets at the beginning. Figure 5 shows the algorithm for emergency situation detection. The algorithm is a modification of sequential K-means algorithm [14], which monitor the data stream in a short time interval, such as 15 minutes, and find the cluster with highest density. If the density is bigger than a user-specific threshold, the common feature of the tweets

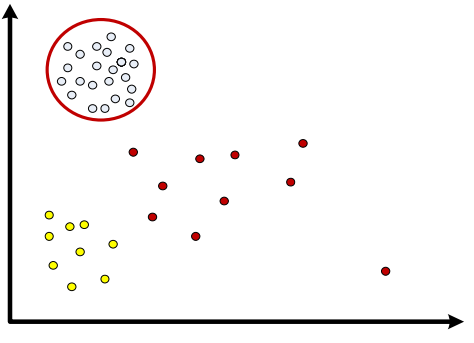

Fig. 4. Emergency situation detecting in 2-dimensionality feature space

Algorithm 2. Sequential K-means algorithm for emergency situation detection

Input: the data stream in the time interval t
Output: the cluster center ρ_i and diameter d_i
Make initial guesses for the means $m_1, m_2 \ldots m_k$
Set the counts $n_1, n_2, \ldots n_k$ to zero
Set the diameter for K clustering $\boldsymbol{d_1, d_2, \ldots, d_k}$ to zero
Set the density for K clustering $\boldsymbol{\rho_1, \rho_2, \ldots \rho_k}$ to zero
While LOOP
    Acquire the next example, X
    If m_i is closest to X
        Increment n_i
        Replace m_i by $m_i = \left(\frac{1}{n_i}\right) * (X - m_i)$
        If $\mathbf{distance}(X, m_i) > \boldsymbol{d_i}$
            Set $\boldsymbol{d_i} = \mathbf{distance}(X, m_i)$
            Set $\boldsymbol{\rho_i} = \boldsymbol{n_i}/\boldsymbol{d_i^2}$
    Choose the minima $\boldsymbol{d_i}$ and the highest $\boldsymbol{\rho_j}$ and sent back

Fig. 5. Sequential K-means algorithm for emergency situation detection

in the clusters is extracted, and experts are required to judge the emergency level. By this way, the emergency situation is detected semi-automatically.

4 Credibility Analysis

With the emergency situation detected, we need to identity the credibility of the tweets. In this section, we propose a supervised learning method to analyses the tweets credibility.

4.1 Labeling the Related Tweets

Generally speaking, the tweets from Twitter can be divided into two types: the news and the chat. After we match the tweets with keyword set, the most of the unrelated chat tweets are eliminated. Those collected tweets mainly include the news and the news related conversation. Then we need to identity the credibility of those tweets. To make the problems easy, those collected tweets are labeled into two classes: the credible tweets and the non-credible tweets. The non-credible tweets contain the tweets which are not the credible news, such as spam, rumors, and the unrelated conversations…etc.

Table 2. Four-types of Features: Author-based, Content-based, Topic-based and Diffusion-based

Type	Feature
Author-based	Time interval of last 2 tweets; Total number of tweets per day; Registration age of the author; Total usage times of Twitter; Number of Followers; Number of followees; Whether a verified user; Whether has description
Content-based	Length of the tweet; Number of the reply comments; Number of words match the keyword set; Whether a re-tweet; Whether contains address information; Number of "!" character; Number of "?" character; Number of "@" character; Number of the emotion smile; Number of the emotion frown
Topic-based	the URL fraction of the tweets; the hashtags (#) fraction of the tweets; whether the address information match the poster's personal profile address information; Number of Positive words; Number of negative words
Diffusion-based	Time of the tweet been cited; Time of the original tweets been cited if it is a re-tweet

Those collected tweets are dispatched into about 5 or more experts, each expert reads a part of the tweets, and labels the tweets into credible or non-credible. In this paper, we collect 350 tweets with the topic of "UK Riots", after labeled by 5 experts, there are about 30.3% of the tweets are non-credible, while 52.3% of the tweets are credible, and the remaining 17.4% of the tweets can't decide their credibility. In the remaining paper, we ignore the 17.4% of the tweets, and using the 82.6% of the tweets as the training set, which is total of 289 tweets are considered.

4.2 Features Extraction

The original tweets format in (1) is insufficient for classification. We need to extend the original feature space. Those features are selected from some of previous researches, such as opinion mining, spam detection and information credibility. The feature set is listed in Table 2. We definite four types of features: Author-based features, Content-based features, Topic-based features and Diffusion-based features.

Author-based features mainly describe the personal and statistics information, such as : the time interval of last 2 tweets, total number of tweets per day, the registration age of the author, number of followers, number of followees, etc.

Content-based features mainly identify the characteristics of the tweet, such as: the tweet's length, the number of words match the keyword set, whether it is a re-tweet, the number of special characters (!, @,?) in the tweet, etc.

Topic-based features are aggregates computed from the tweets content, such as: the URL fraction of the tweets, the hashtags (#) fraction of the tweets, the number of sentiment words, etc.

Diffusion-based features mainly consider the diffusion attributes of the tweet, such as: the times of the tweet has been cited by other authors, the time of the original tweets has been cited if it is a re-tweet.

4.3 Supervised Learning Algorithm

The analysis of the tweets credibility can be modeled as: Given a new arrived tweet, judge it whether a credible or non-credible tweets. It is a traditional 2-class classification problem. The formal notation is: denote the tweets feature space χ, and the class set $Y = \{-1, 1\}$, where Y=1 denotes that the tweet is a credible tweet, Y=-1 denotes the tweet is a non-credible tweet. Given the training dataset $(X_i, Y_i)_{i=1}^{m}$ where $X_i \in \chi$, and $Y_i \in Y$, learn the classifier: $h: \chi \to Y$ which is used to predict the class of new arrived tweet X_{new}.

The technologies for solving the classification are mature, such as SVM [15], Decision Trees [16], Bayesian Network [17], and etc. In this paper, we use Bayesian Network Classifier to partition the terrorism event. Generally speaking, Bayesian Network [17] is a directed acyclic graph (DAG) over a set of variables U, which represent the probability distributions $P(U) = \prod_{u \in U} P(u|parents(u))$. There are various algorithms for Bayesian Network Classification, such as K2 [18],Hill Climbing [19] , TAN [20], etc.

Algorithm 3. Find the skeleton of the Bayesian Network

Input: set of variables $U = \{u_1, u_2, \ldots u_n\}$, Training dataset D

Output: undirected graph $\mathcal{H}$, separation variables sets Sep_{u_i,u_j} for each pair of variables

Initial c to be a complete undirected graph over U

For each u_i, u_j in U

$Sep_{u_i,u_j} = \emptyset$

Traverse the training dataset D, if find $(u_i \perp u_j | U')$ //means that u_i, u_j is conditioned independent given some subset $U' \subseteq U$

$Sep_{u_i,u_j} = U'$

Remove $u_i - u_j$ in $\mathcal{H}$

Return $\mathcal{H}$ and Sep_{u_i,u_j}

Fig. 6. The algorithm of finding the skeleton of the Bayesian Network

Algorithm 4. Find the potential immorality

Input: set of variables $U = \{u_1, u_2, \ldots u_n\}$, undirected graph $\mathcal{H}$, separation variables sets Sep_{u_i,u_j} for each pair of variables

Output: a partially DAG S

Initial S to be the same of $\mathcal{H}$

For each u_i, u_j, u_k in U with $u_i - u_j - u_k \in S$ and $u_i - u_k \notin S$

If $u_j \notin Sep_{u_i,u_k}$

Update S, set $u_i \rightarrow u_j$ and $u_k \rightarrow u_j$

Return S

Fig. 7. The algorithm of Finding the potential immorality

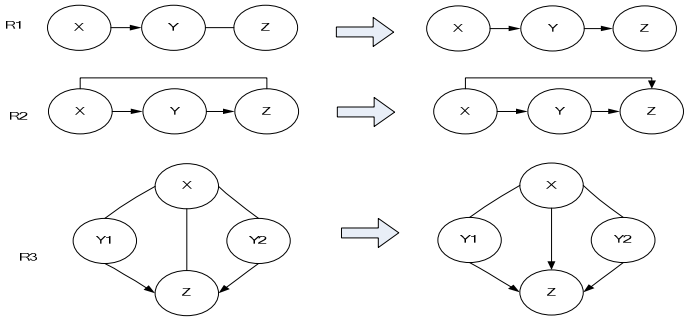

Fig. 8. Rules for orienting edges in the partially DAG

Algorithm 5. Conditional independence test based structure learning in Bayesian Network classification

Input: set of variables $U = \{u_1, u_2, \ldots u_n\}$, Training dataset D

Output: a supposed Bayesian Network B_S

$\left(\mathcal{H}, Sep_{u_i,u_j}\right) =$ Find the skeleton of the Bayesian Network

S= Find the potential immorality

While not convergent

Find the subgraph of S that satisfies left hand of R1-R3

Transform subgraph of S into the right hand of R1-R3

For the remaining undirected edges, choose either orientation.

Return S.

Fig. 9. Conditional independence test based structure learning in Bayesian Network classification

To make the classification simple, we use the conditional independence test based structure learning algorithm (CIT) [17] to get the final results. The conditional independence test based structure learning algorithm mainly includes 3 steps:

1. Find the skeleton of the Bayesian Network. Figure 6 shows the detail algorithm for finding the skeleton.
2. Find the potential immorality. A triplet of variables X, Y and Z is a potential immorality if the skeleton X—Y—Z but doesn't contain and edge between X and Y [17]. Figure 7 shows the algorithm of finding the potential immorality.

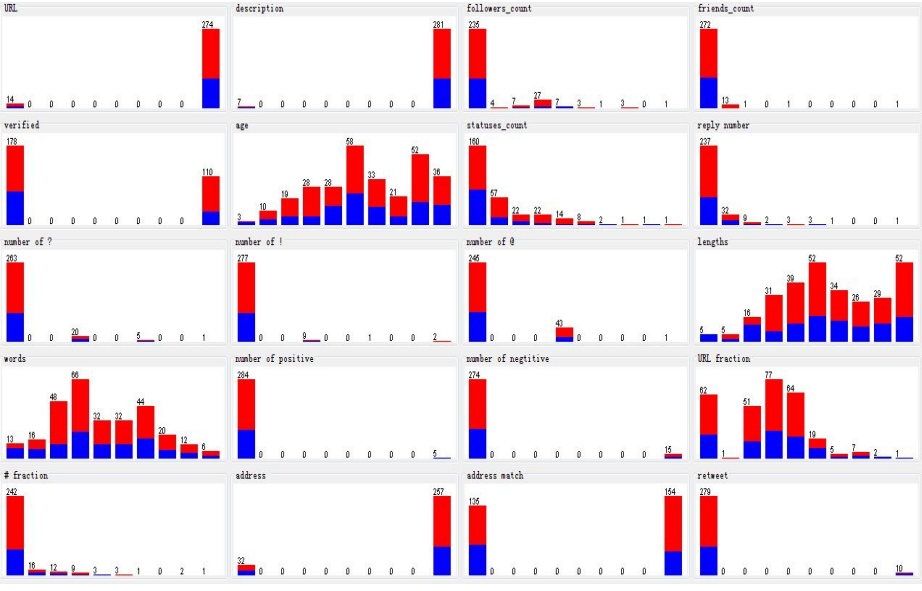

Fig. 10. The experiments features with the UK Riots Topic, the red represents the tweets with credibility, and the blue represents the tweets with incredibility

3. Based on the step 1 and 2, generate the final Bayesian Network. The main job for step 3 is to check whether the partially DAG S in step 2 satisfies left hand of Rule 1 to Rule 3 in Figure 8, and transform it into the right hand of Figure 8.

Above all, the whole algorithm for conditional independence test based structure learning in Bayesian Network classification is summarized in Figure 9.

5 Experiments and Results

In this section, we mainly show our experiments on Twitter in emergency situation. The experiments contain two parts: the data collection and the data analysis. The experiment is built with the help of Weka [21].

5.1 Data Collection

The data is collected with our previous proposed Twitter monitor model, and the tweets are generated from August 6 to August 8, 2010 with the keywords initially

with “Riots”. As time went on, we found that the distance between the words “UK” and “Riots” is closer and closer, so the “UK” related words are added to the Keyword Set in our Twitter monitor model.

Totally, more than 5,000 tweets are collected during August 6 to August 8. We collected the most typical 350 tweets to form the training dataset. After dispatched those tweets to 5 experts, 61 tweets are considered topic-unrelated. Finally, the training dataset contains 289 tweets, with 183 tweets considered with credibility and 106 tweets considered with incredibility 20 features are extracted from the original 289 tweets, as discussed before, with the four type of features: Author-based, Content-based, Topic-based and Diffusion-based features. Figure 10 shows the distribution of the features. Those features are discretized and transform from the number type to the nominal type.

Table 3. Summary for the credibility classification with different algorithms

	CIT	J48	SVM	K2	Hill Climbing
Correctly Classified Instances	63.6678 %	61.2457 %	66.782 %	61.5917 %	62.2837 %
Kappa statistic	0.1259	0.0399	0.2614	0.1263	-0.0206
Mean absolute error	0.3897	0.4315	0.3322	0.3979	0.458
Root mean squared error	0.4865	0.5041	0.5764	0.4967	0.4784
Relative absolute error	83.8411 %	92.8249 %	71.4626 %	85.6077 %	98.5356 %
Root relative squared error	100.9339 %	104.5962 %	119.5779 %	103.0438 %	99.2633 %

Table 4. Detail results for credibility with CIT

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
credibility	0.858	0.745	0.665	0.858	0.749	0.655
incredibility	0.255	0.142	0.509	0.255	- 0.34	0.655
AVERAGE	0.637	0.524	0.608	0.637	0.599	0.655

5.2 Data Analysis

We compare our proposed conditional independence test based structure learning algorithm (CIT) with other state-of-are algorithms, such as J48 decision trees, SVM, Bayesian Network with Hill Climbing algorithm…etc.

Table 3 shows the summary information for the credibility classification with different algorithms. From the table, we can conclude that our proposed CIT algorithm achieves almost the same performance with the SVM algorithm, which is famous for its high precision. And the other algorithm, such as J48, K2, CIT shows better results. The detail of evaluation for each class with CIT is in Table 4. The detail of evaluation shows that our CIT algorithm achieves a bitter better performance in predicting credible tweets than the non-credible tweets.

6 Conclusions

In this paper, we propound a new and interesting problem, which is around the information credibility on Twitter in emergency situation.

To solve this problem, we first propose a novel Twitter Monitor model which is based on the dynamic keywords set. With the monitor model, a modified sequential K-means algorithm is address to semi-automatic detect the emergency situation.

With the emergency situation detected, we proposed a CIT Bayesian Network structure learning algorithm to judge the information credibility. Lastly, experiment with the UK riots related tweets shows that our whole procedure of information credibility on Twitter in emergency situation display good performance compared with other state-of-art algorithm.

As this research is going on, in the future, we will combine the opinion mining algorithm into the information credibility area, since currently we consider little for natural language process in this paper. The emergency situation detection will be modified to be automatic with the ontology technologies. What's more, the training dataset will be changed, as currently we just ask experts for help, and only 10 experts are available. In the future, we will put the job of labeling the tweets into the Mechanical Turk, to make the dataset more precise.

Acknowledgement. This work was supported by the Ministry of Industry and Information Technology of China (No. 2010ZX01042-002-003-001).

References

[1] Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: HICSS-43, pp. 1–10. IEEE, Kauai (2010)

[2] Vieweg, S.: Microblogged contributions to the emergency arena: Discovery, interpretation and implications. Computer Supported Collaborative Work (2010)

[3] Flanagin, A.J., Metzger, M.J.: Perceptions of Internet information credibility. Journalism and Mass Communication Quarterly 77, 515–540 (2000)

[4] Juffinger, A., Granitzer, M., Lex, E.: Blog credibility ranking by exploiting verified content. In: WICOW, Madrid, Spain, pp. 51–58 (2009)

[5] Johnson, T.J., Kaye, B.K., Bichard, S.L., Wong, W.J.: Every Blog Has Its Day: Politically-interested Internet Users' Perceptions of Blog Credibility. Journal of Computer-Mediated Communication 13, 100–122 (2008)

[6] Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V.: Detecting spammers on twitter. In: Seventh Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, Redmond, Washington, US (2010)

[7] Jindal, N., Liu, B.: Opinion spam and analysis. In: WSDM 2008, Palo Alto, California, USA, pp. 219–230 (2008)

[8] Stringhini, G., Kruegel, C., Vigna, G.: Detecting spammers on social networks. In: ACSAC 2010, Austin, Texas, USA, pp. 1–9 (2010)

[9] Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: WWW 2011, Hyderabad, India, pp. 675–684 (2011)
[10] Mathioudakis, M., Koudas, N.: TwitterMonitor: trend detection over the twitter stream. In: Proceedings of the 2010 International Conference on Management of Data, pp. 1155–1158 (2010)
[11] Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: WWW 2010, Raleigh, North Carolina, pp. 851–860 (2010)
[12] Hughes, A.L., Palen, L.: Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management 6, 248–260 (2009)
[13] Mendoza, M., Poblete, B., Castillo, C.: Twitter Under Crisis: Can we trust what we RT? In: 1st Workshop on Social Media Analytics (SOMA 2010), Washington, DC, USA, pp. 71–79 (2010)
[14] MacQueen, J.: Some methods for classification and analysis of multivariate observations. Presented at the Proceedings of the Fifth Berkeley Symposium (1967)
[15] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998)
[16] Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S.: Top 10 algorithms in data mining. Knowledge and Information Systems 14, 1–37 (2008)
[17] Koller, D., Friedman, N.: Probabilistic graphical models. MIT Press (2009)
[18] Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9, 309–347 (1992)
[19] Buntine, W.: A guide to the literature on learning probabilistic networks from data. IEEE Transactions on Knowledge and Data Engineering 8, 195–210 (1996)
[20] Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning 29, 131–163 (1997)
[21] Bouckaert, R.R., Frank, E., Hall, M.A., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: WEKA"Cexperiences with a java opensource project. Journal of Machine Learning Research 11, 2533–2541 (2010)