Indian Journal of Science and Technology, *Vol 9(11), DOI: 10.17485/ijst/2016/v9i11/89390, March 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Implementation of Speedy Emergency Alert using Tweet Analysis

W. Ancy Breen*, A. Merry Ida and M. Queen Mary Vidhya

Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai - 600062, Tamil Nadu, India; breen.cse@gmail.com, idamerry@gmail.com, queenmaryvidya@gmail.com

Abstract

Background/Objectives: The Traditional Medias such as television, newspapers and radios become less qualified because they are not much preferred. The traditional techniques become inexperienced as because they do not consider social relation data. While, the existing social recommendation approaches has not been fully considered to provide any useful information as they only classify the tweets and there is no proper alert to rescue people. **Methods/Statistical analysis**: This paper is to investigate the social problems on the basis of both economically and emotionally using twitter; summarizing the classified tweets into useful information used for both increasing revenue and cutting costs. Then in the proposed approach the particle filter mechanism is used to extract keywords from tweets. Further it uses stemming algorithm which is used for reducing variant forms of a word to a common form. Data are split and stored in the Data node and the index is maintained by the Name node. **Findings:** The factual results and analysis manifest that the proposed method significantly outperforms the existing approaches. **Applications/Improvements:** The importance of the social networks with reference to their effective usage especially during natural calamity has been highlighted and also its usefulness for people to get adapted to the society has been emphasized.

Keywords: Gradient Search, Increasing Revenue, Particle Filter Mechanism, Social Recommendation, Stemming Algorithm

1. Introduction

Social media allows instantaneous sharing of real-time information. The main use of social networking sites is to provide useful information to the people. It is observed that most of the knowledge is gained from social networks and many useful results are obtained. The traditional collaborative filtering does not consider social relations[1].

In Twitter, a popular micro blogging service has become a new information channel for users to receive and to exchange information[2]. It is an online social network used by millions of people around the world to stay connected to their friends, family members and co-workers through their computers and mobile phones. Every day, nearly 170 million tweets are created and redistributed by millions of active users which are not made productive to provide any useful information. Twitter has several unique advantages that distinguish it from news web sites, blogs, or other information channels. Firstly, tweets are created in real-time. The tweets are in 140-character-message limit and the popularity of Twitter is it is a mobile application, users can tweet and re-tweet instantly.

For example, the system could detect a tweet related to a fire accident 10 minutes after it was fired, while the first conventional news report will be appeared approximately 3 hours later. This shows an investigation about the real-time nature of twitter that is designed to ascertain whether any valid information can be extracted from it. In the proposed approach an event notification system that keeps monitoring the tweets posted by the registered users and delivers notification promptly. This is done

**Author for correspondence*

based on the knowledge obtained from the investigation. There are three steps to be followed: first, Crawl numerous tweets from the registered users in the created application; second, probabilistic models to extract events from those tweets and estimate locations of events; and third, developed an earthquake reporting system that extracts earthquakes from Twitter and sends an automatic alert message to registered users as well as to the rescue team. If the alerts are sent it is easier to take immediate rescue action to protect people from the disaster.

2. System Model

2.1 Application Creation

While creating an application, the design fields in our application will be assigned. That includes username, password, phone, date of birth, gender and other personal information. Once created the user is allowed to enter the tweet. Also the server will store the data and allow the user to enter into the original twitter chat application created. The user can post their tweets through this application.

2.2 Server

A Server is used to verify the user information submitted in the registration form to enter into the chat application and allow the user to Tweet with their friends. The server will analyze whether the user is an existing or new user. If the user is new user, they will be asked to Sign up and if the user is an already existing user, then they will have directed to the application once they Login. Also the server will analyze the contents posted by the user. So that the server will extract the keywords. The server will also retrieve the use information such as access time and location which is used to find each user's location and can provide any necessary help to them[3].

2.3 Extracting the Keyword using Particle Filter

After the tweets are posted, the Server will analyze them and extracts the keyword using particle filter. The particle filter will extract the keyword and filters the other words using stemming algorithm. By using stemming algorithm, the unwanted words in the chat can also be removed so that the extracted word counts can be calculated easily.

2.4 Automatic Alert

Server sends the SMS and Email alert to the registered tweet users once it attains a maximum peak of the extracted keyword. Figure 1 indicates the system architecture for automatic alert. Server is used to verify the user information submitted in the registration. For Email Alert, the email via Internet will be generated. For sending an SMS, connect PC suite configured mobile via Data cable with Server. This PC suite configured mobile will transmit the SMS to the rescue team and also to the registered tweet users.

3. System Analysis

In this paper, the proposed idea is, use a particle filter to estimate the location from the tweets posted by the registered users in our application. This can be achieved by considering each user as a sensor, and analyzing a spatial and temporal pattern of an event and applies particle filtering which is used for location estimation. By using these information, the system can analyze if the maximum peaks of the keywords like "Earthquake, Tsunami" etc. are reached. A peak will generate an automatic alert. So that the rescue team may be able to do the rescue process once the message is got in the server. This is a less time consuming process as the alert is sent immediately. By this the loss rate can be reduced by taking immediate rescue action to save the people from disaster. The system provides a way to analyze the tweets posted in the twitter using an Application Programming Interface (API)[4].

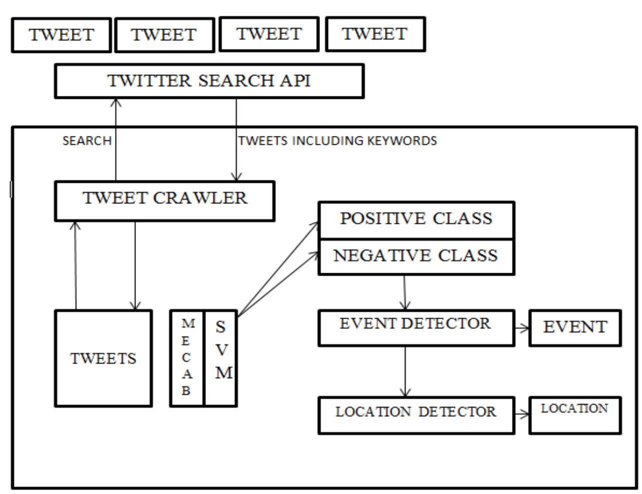

Figure 1. System Architecture Diagram for Automatic Alert.

4. Proposed Scheme

The system uses a manually given training set for the target event and location. For example, it will search for “tsunami” and “water” if our target event is tsunami. By searching every second, the system can obtain a subset of all the tweets that contain the keyword in our searching queries or training set. Then, for each tweet the system will search result and obtain the semantic features as its feature vector, then apply the classifier concept to the tweet and get a value for positive and negative. A support vector machine is used to separate positive and negative class.

4.1 Particle Filter Mechanism

This mechanism allows us to extract the important keyword from the tweets. A particle filter is a probabilistic approximation algorithm implementing a Bayes filter. It is used along with stemming algorithm to obtain the location with time.

4.2 Stemming Algorithm

It is a process of linguistic normalization, in which variant forms of a word are reduced to a common form. It improves the performance of IR systems. It is used for eliminating the stopping words (this, is was, etc.) and helps us to identify only the keywords.

4.3 SYM

Support Vector Machines is based on the concept of decision planes that define decision boundary. A decision plane is one that separates between a set of objects having different class memberships i.e., positive class and negative class[5].

5. System Model Evaluation

The proposed model reports about the disaster and it functions in the following ways:

Consider the example and its features,
I am in Japan, earthquake right now!

1. Statistical features: (7 words, 5th word). This shows the number of words in the tweet and the position of the query within the tweet.
2. Keyword features: (I, am, in, Japan, earthquake, right now). This shows that the words in the tweets are classified to identify the keyword.
3. Word context features: (Japan, right). These are the words before and after the query word.

Once the user register, they can enter into the twitter application (Table 1). In the application, user can tweet with their friends. Tweets which are posted are stored in the server and analyzed. These posted tweets are classified based on above features and particular keywords in tweets are obtained. After identifying the keywords and reaching maximum peak of a particular keyword, an email alert will be sent (Figure 2).

The registered users in the tweet application will also receive a SMS alert (Figure 3) once the threshold for the keyword reaches a maximum limit as email alert. This helps the users to know about what kind of natural disaster is occurring and at which place it is affecting. Thereby the information reaches to many users as the information is broadcasted[6].

6. Proposed Model Evaluation

The performance report of the proposed model with the existing model is described in this section. Table 2 indicates the classification of performance-earthquake query and Table 3 represents the classification performance-shaking query.

The conditional distribution over the observed entries in R is defined as:

$$P(R\,|\,S,U,V,\sigma^2 R) = \prod_{i=1}^{M} N \prod_{j=1}^{N} N(R_{ij}\,|\,S_i G_j{}^T \bigcirc U_i{}^T V_{j,} \sigma^2 R)$$

By incorporating the social contextual factors, the posterior distribution is defined as

$$P(S,U,V|R,G,W,C,F,\Omega)$$

$$=(P(R,W,C,F,G|S,U,V,\Omega)P(S,U,V|\Omega))/(P(R,G,W,C,F,\Omega))$$

$$\alpha\ P(R|S,U,V,\Omega)P(W|U,\Omega)P(C|V,\Omega)P(F|S,\Omega)$$

$$P(S|\Omega)P(U|\Omega)P(V|\Omega)$$

Table 1. Tweets

Feature Name	Features
Features A	7 Words, The Fifth Word
Features B	I, Am, In, Japan, Earthquake, Right Now
Features C	Japan, Right

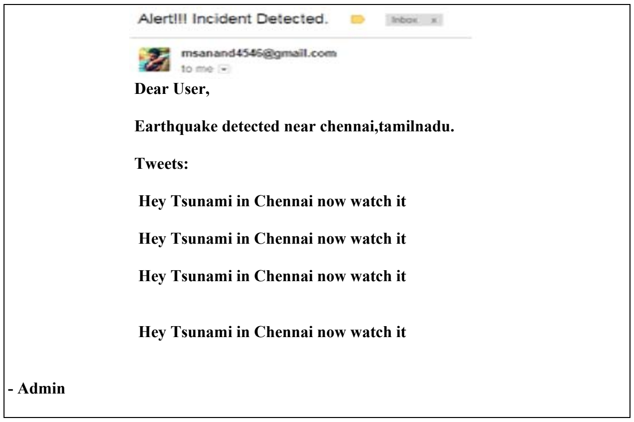

Figure 2. Sample Email alert.

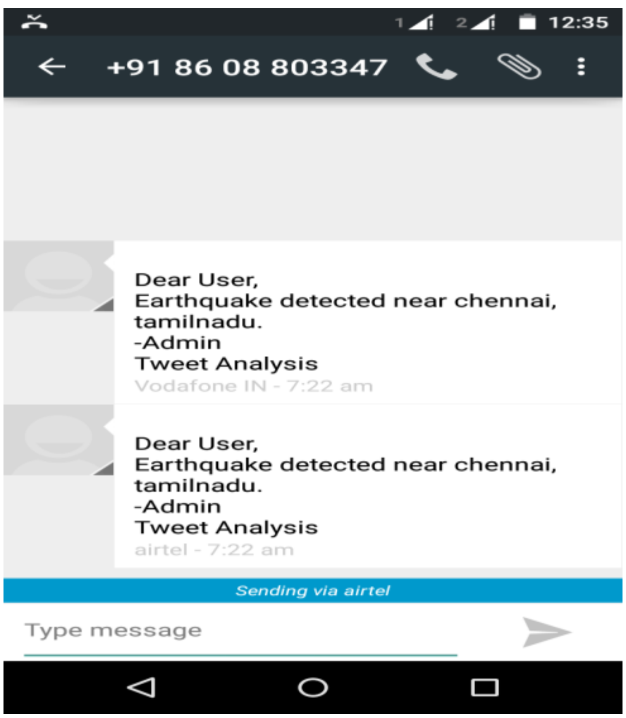

Figure 3. Sample SMS alert.

Table 2. Classification Performance-Earthquake query

Features	Recall	Percision	F-value
A	87.50%	63.64%	73.69%
B	87.50%	38.89%	53.85%
C	50.00%	66.67%	57.14%
All	87.50%	63.64%	73.69%

Table 3. Classification Performance-shaking query

Features	Recall	Percision	F-value
A	66.67%	68.57%	67.61%
B	86.11%	57.41%	68.89%
C	52.78%	86.36%	68.20%
All	80.56%	65.91%	72.50%

$$= \prod i, jN(R_{i,j} \mid S_i G_j^T \circ U_i^T V_{j}, \sigma^2 R)$$

$$\prod p,q\, N(W_{pq} \mid U_p^T U_q, \sigma^2 w) \prod m,n\, N(C_{mn} \mid V_m^T V_n, \sigma^2 c)$$

$$\prod s,t\, N(F_{st} \mid S_{st}\, \sigma^2 F) \prod x\, N(S_z \mid o, \sigma^2 s)$$

$$\prod y\, N(U_y \mid o, \sigma^2 u) \prod z\, N(V_z \mid o, \sigma^2 v)$$

Where denotes that zero-mean spherical Gaussian priors are placed on latent feature vectors and observed matrices. Then

$$\ln P(S,U,V|R,G,M,C,F,\Omega)$$

$$\alpha - \frac{1}{2\sigma^2 R} \sum_{i,j} (R_{ij} - S_i G_j^T \circ U_i^T V_j)^2$$

$$-\frac{1}{2\sigma^2 W} \sum_{p,q} (W_{pq} - U_p^T U_q)^2 - \frac{1}{2\sigma^2 C} \sum_{m,n} (C_{mn} - V_m^T V_n)^2$$

$$-\frac{1}{2\sigma^2 F} \sum_{s,t} (F_{st} - S_{st})^2 - \frac{1}{2\sigma^2 S} \sum_x (S_x^T S_x)$$

$$-\frac{1}{2\sigma^2 U} \sum_y (U_y^T U_y - \frac{1}{2\sigma^2 V} \sum_z (V_z^T V_z)$$

Maximizing the posterior distribution is equivalent to minimizing the sum-of-squared errors function with hybrid quadratic regularization terms:

$$J = ||R - SG^T O\, U^T V||^2_F + \alpha||W-U^T U||^2_F + \beta||C-V^T V||^2_F + \gamma||F - S||^2_F + \delta||S||^2_F + \eta||U||^2_F + \lambda||V||^2_F$$

where

$$\alpha = \frac{\sigma^2 R}{\sigma^2 W}, \beta = \frac{\sigma^2 R}{\sigma^2 C}, \gamma = \frac{\sigma^2 R}{\sigma^2 F}, \delta = \frac{\sigma^2 R}{\sigma^2 S}, \eta = \frac{\sigma^2 R}{\sigma^2 U}, \lambda = \frac{\sigma^2 R}{\sigma^2 V},$$

and $||.||_F$ is the Frobenius Norm

Each equation above is solved alternatively with the other two matrices fixed and proceed step by step until convergence. As the objective is obviously lower bounded by 0 and the alternating gradient search procedure will reduce it monotonically, the algorithm is guaranteed to be convergent. In this paper, the gradient search method to solve the problem has been used. Specifically, the gradients of the objective with respect to the variables are

$$\frac{\partial J}{\partial S} = -2(R - SG^T \bigcirc U^T V)G - 2\gamma(F - S) + 2\delta s$$

$$\frac{\partial J}{\partial U} = -2V(R - SG^T \bigcirc U^T V)^T - 4\alpha U(W - U^T U) + 2\eta u$$

$$\frac{\partial J}{\partial V} = -2U(R - SG^T \bigcirc U^T V) - 4\beta V(C - V^T V) + 2\lambda V$$

Thus, the following gradient-based approach to our social contextual model is applied in Algorithm 1. J decreases the fastest in the direction of gradients during each iteration and the sequence ($J^{(t)}$) converges to the desired minimum.

Algorithm 1 Social Contextual Model Context

Require: $0 < \alpha_S^{(t)}, \alpha_U^{(t)}, \alpha_V^{(t)} < 1, t = 0$.Initialization

$J^{(0)} = J(S^{(0)}, U^{(0)}, V^{(0)})$.

Ensure : $J^{(0)} \geq 0$, $J^{(t+1)} < J^{(t)}$

For t = 1,2,3,… do

Calculate $\frac{\partial J^{(t-1)}}{\partial S}, \frac{\partial J^{(t-1)}}{\partial U}, \frac{\partial J^{(t-1)}}{\partial V}$

$S^{(t)} \leftarrow S^{(t-1)} - \alpha_S^{(t-1)} \cdot \frac{\partial J^{(t-1)}}{\partial S}$

$U^{(t)} \leftarrow U^{(t-1)} - \alpha_U^{(t-1)} \cdot \frac{\partial J^{(t-1)}}{\partial U}$

$V^{(t)} \leftarrow V^{(t-1)} - \alpha_V^{(t-1)} \cdot \frac{\partial J^{(t-1)}}{\partial V}$

$J^{(t)} \leftarrow J(S^{(t)}, U^{(t)}, V^{(t)})$.

end for.

7. Conclusion

In this paper, the message posted in the twitter is used efficiently. This make the social networking sites to get more reach by the people. The real-time nature of twitter is investigated, thereby devoting particular attention to an event. Semantic analyses were applied to tweets to classify them into a positive and negative class. Each twitter user is regarded as a sensor, and also set the disaster as a detection of event based on sensory observations. Location estimation methods such as particle filtering are used to estimate the locations of the event. As an application, is developed as a disaster reporting system which identifies the disaster from the tweets and sends an automatic alert. This is a novel approach to notify people promptly of an earthquake event. This is a less time consuming process, because the automatic alert is immediately sent to the nearest users' location. By this approach the loss rate can be reduced, due to the immediate action taken by rescue team to save the people from disaster[7]. Using twitter, it can also able to reduce the loss rate, by implementing this methodology for various events and due to the immediate action taken by rescue team and the people can also be saved. This helps people to get updated with the events occurring at a particular location. They can prevent themselves and can protect the people present in the disaster location.

8. References

1. Ramesh N, Andrews J. Personalized search engine using social networking activity. Indian Journal of Science and Technology. 2015 Feb; 8(4):301–306. doi: 10.17485/ijst/2015/v8i4/60376.
2. Sarah M, Abdur C, Gregor H, Ben L, Roger M. Twitter and the micro-messaging revolution; O'Reilly Media Radar: USA; 2008.
3. Java A, Song X, Finin T, Tseng B. Why we twitter: understanding microblogging usage and communities. Proceeding of Ninth Web KDD and First SNA-KDD Workshop Web Mining and Social Network Analysis, (WebKDD/SNA-KDD '07); 2007. p. 56–65.
4. Huberman BA, Romero DM, Wu F. Social networks that matter: Twitter under the microscope. First Monday. 2008; 14(1):1–9.
5. Kwak H, Lee C, Park H, Moon S. What is twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web (WWW '10); 2010. p. 591–600.
6. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: conversational aspects of retweeting on twitter. 2010 43rd Hawaii Internatioanl Conference System Sciences (HICSS-43); Honolulu: HI; 2010. p. 1–10.
7. Tumasjan A, Sprenger TO, Sandner PG, Welpe IM. predicting elections with twitter: what 140 characters reveal about political sentiment. Proceedings of the 4th Internatioanl AAAI Conference on Weblogs and Social Media. WWW'10; 2010. p. 178–85.