

Emergent System using Tweet Analyzer: Naturally Inspired Computing Approach

N.sharmila¹, B .Anantharaj², R.Hariraman³
Student¹, HOD², Faculty³ Department of CSE,
Thiruvalluvar college of Engineering and Technology
Anna University, India

Email: sharmilagunasekaran@yahoo.com¹/ ananthu_arun72@yahoo.com²/ hariraman.vs@gmail.com³

Abstract--Nowadays much system has developed to reach the people during disasters. A social interaction with the micro blogging services has vastly increased. Twitter a well popular social medium has scarcity of attention makes people interacting with each other. This paper explores the use of twitter for disaster event using text classifier. It analyses the targeted event with tweeted text and identifies the target and its location. By the use of SVMs, the text classifiers are achieved.SVM performs the methods of Bayesian filtering for the informative messages of particular event. This method of filtering works best compare with the other methods for estimating the informative messages. As an application we are sending an E-mail and SMS alert message through the twitter and the declared set of friends and followers follows the messages.

Index terms - Twitter, Informative, SVM, Bayesian Filtering.

1. INTRODUCTION

1.1 Naturally Inspired Algorithm.

Nature is the best teacher and its designs and capabilities are extremely enormous and mysterious that researchers are trying to mimic nature in technology. Bio inspired computing has come up as a new era in computing encompassing a wide range of applications, covering all most all areas including computer networks, security, robotics, bio medical engineering, control systems ,parallel processing ,data mining, power systems, production engineering and many more. New or persistent problems in computer science could have a lot in common with problems nature has encountered and resolved long ago. Thus an easy mapping is possible between nature and technology.

Towards technology through Nature

The real beauty of nature inspired algorithms lies in the fact that it receives its sole inspiration from nature. They have the ability to describe and resolve complex relationships from intrinsically very simple initial conditions and rules with little or no knowledge of the search space Nature is the perfect example for optimization, because if we closely examine each and every features or phenomenon in nature it always find the optimal strategy, still addressing complex interaction among organisms ranging from microorganism to fully fledged human beings, balancing the ecosystem, maintaining diversity, adaptation, physical phenomenon like river formation, forest fire ,cloud, rain .etc

1.2 Twitter

Social media, a very old and pervasive mechanism for communicating distal interactions among people, have become widespread in the age of Web[1]. With interfaces that allow people to follow the live of friends, social contact and family. This network is used millions of people around

the world. Twitter is an online social networking and micro blogging service that enables users to send and read "tweets", which are text messages limited to 140 characters. Registered users can read and post tweets but unregistered users can only read them. It's a platform wherein users share their thoughts, news, information and jokes. Twitter makes global communication cheap and measurable. Profiles are public anyone in the world can see what you write, unless you elect to make your profile private. Users "follow" each other in order to keep tabs on and converse with specific people

Twitter users are able to post straight and derivative updates. Straight posts are used when a user aims her update to a specific person, whereas derivative updates are used when the update is meant for anyone that cares to read it. Even though straight updates are used to communicate directly with a specific person, they are public and anyone can see them. Often times two or more users will have conversations by posting updates directed to each other. Around 25.4% of all posts are directed, which shows that this feature is widely used among Twitter users.

A predominant characteristic that is common among the micro blogging services is their real-time nature. Although blog users usually update their web logs once every days, twitter user usually write tweets many times in a day. Several important occurrences symbolize their real time nature.

Many people have made research work on the social media, Twitter most of the studies based on the structure of twitter [2],[3],[4],and some of them on the characteristic of twitter[5],[6] and some on the application usage of twitter[7][8].

This paper illustrates the detection of the earthquake disaster by monitoring the tweet messages and it helps to detect the occurrence in the real-time nature.

This paper presents the probe of the real-time nature. Twitter is sketched to establish whether we can

extract valid information from it. Proposing an event that monitors tweet messages and delivers the warning from the explore.

First, dragging numerous tweets related to target events; second, extracting events from those tweets and estimate locations of events by using SVM; finally, develop a reporting system that extracts earthquakes from Twitter and sends a message to registered users.

This paper is ordered as follows. Section 2 deals about the Emergent Alert System. Section 3 focuses the ways of Algorithm used. Section 4 discusses about the related works and discussion. Section 5 discusses about the future enhancements of NIA. Section 6 reviews the conclusion of this paper.

2. EMERGENT ALERT SYSTEM

Based on the previous research we are taking, earthquake as our target. Why we have choose earthquake as a target because of the following reasons:

1. It is very easy and accurately validates the event by event detection methodology.
2. Its gives out the meaningful value by detecting the event.

Here, explaining our methods using an earthquake as a target event. First, to obtain tweets on the target event, we apply semantic analysis of a tweet. For example, users might make tweets such as “Earthquake!” or “Now it is shaking,” in which earthquake or shaking could be keywords, but users might also make tweets such as “I am attending an Earthquake Conference,” or “Someone is shaking hands with my boss.” We prepare the training data and devise a classifier using a Support Vector Machine (SVM) based on features such as keywords in a tweet, the number of words, and the context of target-event words. After this classifier we find the content is informative or non informative by using Naïve-Bayesian filtering method. Thus event-detection problem can be reduced to one of object detection and location estimation. A user has a mobile device. Through infrared communication or a Wi-Fi signal of a mobile, the user location is estimated as providing location-based services such as navigation and museum guides [9], [10]. Finally an emergent message has been sent to the registered user as well as the followers.

3. ALGORITHMS

3.1 Stemming:

Stemming is used to improve retrieval effectiveness and to reduce the size of indexing files. Several approaches to stemming are described--table lookup, affix removal, successor variety, and n-gram.

To find the useful tweets from the messages or content, the tweet might include the target event. Here the target event is earthquake. So we have to extract the content or messages with the keyword 'earthquake'. For example users might make tweets like “”. The keywords mentioned are relates to the targeted keyword are confirmed by the matching algorithms of stemming. Such algorithms use a stem database, for example a set of documents that contain stem words. These stems, as mentioned above, are not

necessarily valid words themselves (but rather common sub-strings, as the "earth" in "earthquk" and in "earthquake". In order to stem a word the algorithm tries to match it with stems from the database, applying various constraints, such as on the relative length of the candidate stem within the word, for example, the short prefix "eq", which is the stem of such words as "EartQuk", "Quake" and "earth", would not be considered as the stem of the word "earthquarter".

3.2 Support Vector Machines

Support Vector Machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns. It used for classification and regression analysis[9]. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

The tweets are truly explains about the target. But they are not the original report of the event. It must confirm whether a given tweet is truly mentioning about the real target. For this confirmation we are using machine-learning algorithms. To classify tweet content into positive or negative examples, we produce a model to classify the examples.

“Here earthquake, be safe”

TABLE 1.SVM Example

<i>Classification name</i>	<i>classification</i>
1	4 words,2 nd word
2	Here,earthquake,be,safe
3	Here(finding of location)

We create three groups of classification for the above example.

1-the number of words in a tweet messages ant the position of keyword in a tweet.

2-the words in a tweet.

3-the words before and after the keyword.

The analysis of tweet is done using stop word elimination and stemming which separates the sentences into a single set of words.

3.3 SVM is good for Classifiers

High dimensional input space: Learning text classifiers, it deals with many features. Since SVMs good for protection, which does not necessarily depend on the number of features, they have the potential to handle these large feature spaces.

Few irrelevant features: in text categorization there are only very few irrelevant features. All features are ranked according to their information gain. Then a naive Bayes

classifier [2] is trained using only those features ranked 1-200, 201-500, 501 -1000, 1001-2000, 2001-4000, 4001-9962. A classifier using only that \worst" features have a performance much better to the conjecture that a good classifier should combine many features (learn a \dense" concept) and that aggressive feature selection may result in a loss of information.

Contains only few entries which are not zero. Kivinen et al. [4] give both theoretical and empirical evidence for the mistake bound model that \additive" algorithms, which have a similar inductive bias like SVMs, are well suited for problems with dense concepts and sparse instances.

Most text categorization problems are linearly separable: All Ohsumed categories are linearly separable and so are many of the Reuters tasks. The idea of SVMs is to and such linear Separators.

3.4 Naïve-Bayesian Filtering

We use probabilistic models for the detection of the target. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

A naive bayes classifier is a simple probabilistic classifier based on applying bayes theorem with strong independence assumptions. Probability model would be independent model.

$P(C|F_1, F_2, F_3, \dots, F_N)$ over a dependent class variable C with small number of outcomes or classes, conditional on several feature variables F_1 through F_N . The problem is that if the number of features N is large or when a when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. Therefore reformulate the model to make it more tractable,

$$P(C|F) = P(C) P(F|C) / P(F)$$

First we examine the actual data. It represents the quality of target event, an earthquake. It is evident that projection occurs in the number of tweets. Each corresponds to the target occurrence concerning earthquake more earthquakes are occurred.

Chi-square is useful technique for finding whether the given data is representative of particular distribution.

4. RELATED WORKS

Best of the recent social media is Twitter. Plenty of researches are done in twitter. Regarding similar research are presented in this paper. Bernardo A. Huberman [3] proposed a method to describe massive online social networks as a representation of social interactions that can be used to study the propagation of ideas, social bond dynamics and viral marketing, among others. Social

interactions within Twitter reveal that the driver of usage is a sparse and hidden network of connections underlying the declared set of friends and followers. Social networking of micro blogging services of twitter is used here. Sending of messages to the followers either in a short way of communication.

Joachims,[9] proposes the use of Support Vector Machines for learning text classifiers. It analyzes the particular properties of learning with text data and identities. SVMs achieve substantial improvements over the currently best performing methods and behave robustly over a variety of different learning tasks. Explores key techniques for handling and organizing text data. It identifies the benefits of Support Vector Machines for text categorization. SVMs are a new learning method and very open to theoretical understanding and analysis. The properties of text related to SVMs are very well suited for learning in this setting. SVMs show substantial performance gains. the conventional text classification methods SVMs will prove to be very robust, eliminating the need for expensive parameter tuning

Kerstin Borau[10] establish the chance of actively produce language and the chance to use English as tool of communication. It describes how to use the twitter with student and analyze the student message. Explains how Twitter was used to learn specific aspects of language. Our main finding is that it is suitable to train communicative and culture competence anytime anywhere with-out face-to-face interaction. Provides how to use twitter for the communication and the analyze of the text entry based on the semantic usage.

Stefania Montani[11] proposes to describe a novel framework, in which time-series features are summarized by means of Abstractions, and then retrieved resorting. Our approach grants for interpretability of the output results, and understandability of retrieval process. It is domain-independent. These of TA provide an easily interpretable output, also for end users. Expected time of completion, considering all types of inconsistency should be included to prevent starvation for some tasks. Framework of time series features are retrieved summarized and sorted for the earthquake reporting system. Time of completion of the expected event notification.

Various researches have been studied using the informative messages from the twitter and presented as probabilistic model. The motivation of our study is that the twitter data is more time-sensitive and examines out the time events such as earthquakes.

5. FUTURE ENHANCEMENTS

Many studies are undertaken to monitor the social media, twitter. Most are aimed at the changes in the social state. Our study is to use the twitter for real-time event detection. It is possible that many events can be finding by noticing twitter.

Expanding the system to detect various kinds of events using twitter. The system includes the assumption for a single instance. We can detect the event like typhoon, rainfall and even traffic jams. To realize more events at a time advanced probabilistic models are produced. That helps to detect multiple events in a time.

A search query is important for seeking tweet .we can monitor tweet to detect about the diseases or any other medical related work. Advance algorithms can be useful for our future work.

6. CONCLUSION

As described in the paper .event detection of earthquake is examined. Probabilistic models of SVMs are applied to find the informative and non-informative classes. Based on this we find an event. As an application the event is send as a message to the registered users in a micro blogging services.

Twitter is used here as a micro blogging social media and used in solving the real time problem of disasters. It is hope that it provides some future integration in the naturally inspired events.

REFERENCES

- [1] M. Sarah, C. Abdur, H. Gregor, L. Ben, and M. Roger, "Twitter and the Micro-Messaging Revolution," technical report, O'Reilly Radar, 2008.
- [2] A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," Proc. Ninth WebKDD and First SNA-KDD Workshop Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07), pp. 56-65, 2007.
- [3] B. Huberman, D. Romero, and F. Wu, "Social Networks that Matter: Twitter Under the Microscope," ArXiv E-Prints, <http://arxiv.org/abs/0812.1045>, Dec. 2008.
- [4] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, A Social Network or A News Media?" Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 591-600, 2010.
- [5] G.L. Danah Boyd and S. Golder, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," Proc. 43rd Hawaii Int'l Conf. System Sciences (HICSS-43), 2010.
- [6] A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpel, "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment," Proc. Fourth Int'l AAAI Conf. Weblogs and Social Media (ICWSM), 2010.
- [7] P. Galagan, "Twitter as a Learning Tool. Really," ASTD Learning Circuits, p. 13, 2009.
- [8] K. Borau, C. Ullrich, J. Feng, and R. Shen, "Microblogging for Language Learning: Using Twitter to Train Communicative and Cultural Competence," Proc. Eighth Int'l Conf. Advances in Web Based Learning (ICWL '09), pp. 78-87, 2009.
- [9] Joachims, Baroper Str. Dortmund "Text Categorization with Support Vector Machines: Learning with Many Relevant Features" Thorsten University at Dortmund Informatik LS8, Germany
- [10] Kerstin Borau, Carsten Ullrich, Jinjin Feng, and Ruimin Shen "Microblogging for language learning :Using twitter to train communicative and cultural competence" Distance Education College of Shanghai Jiao Tong University.
- [11] Stefania Montani, Giorgio Leonardi, Alessio Bottrighi, Luigi Portinale, and Paolo Terenziani ,(2013)"Supporting Flexible, Efficient, and User-Interpretable Retrieval of Similar Time Series", IEEE transactions
- [12] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo.(2013) "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development",IEEE



She completed B.TECH –Information Technology from Arulmigu Meenakshi Amman College of Engineering, University of Madras in the year 2014 and also She obtained her PG Degree in Computer Science from Thiruvalluvar College of Engineering and Technology, Anna University in the year 2014.