IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 2, February 2016

Data Analysis of Twitter feeds on Google Maps

Apoorva Ramkumar[1], Krishna Khamankar[2], Trishala Ghone[3], Prof. Sneha Annappanavar[4]

Department of Computer Engineering, Vidyalankar Institute of Technology, Mumbai, India[1,2,3,4]

Abstract: With increasing popularity in microblogging sites, we are in the era of information explosion. As of June 2011, about 200 million tweets are being generated every day. Spurred by that growth, companies and media organizations are increasingly seeking ways to mine Twitter for information about what people think and feel about their products and services. One advantage of this data, over previously used data-sets, is that the tweets are collected in a streaming fashion and therefore represent a true sample of actual tweets in terms of language use and content. Although Twitter provides a list of most popular topics people tweet about known as trending topics in real time, it is often hard to understand what these trending topics are about and it becomes difficult to classify them based on the emotions that each tweet is trying to portray. Therefore a question arises as to what happens when you combine the social networking power of Twitter with the usefulness of Google Maps? The result can contain some rather innovative and intriguing applications. With Twitter and Google Maps mashup's here, it's all about taking advantage of both the Google Maps and Twitter API to create a mashup that's unique, practical, and highly efficient. The aim of our project is to display real time globally trending tweets on google maps and perform sentiment analysis on the tweets related to the topic.This paper primarily focuses on the trend detection and sentiment analysis techniques and the applications of using a Twitter and Google maps mashup. Further, we analyze the sentiments expressed within a particular sentence, paragraph or document etc. The analysis based on sentiments can pave way for automatic trend analysis, topic recognition and opinion mining etc. Furthermore, we can fairly estimate the degree of positivity and negativity of the opinions and sentiments based on the content obtained from a particular social media.

Index Terms: Trend Analysis, Sentiment Analysis, Data analysis, Geocoding, microblogs, trend detection, Twitter, Google maps.

I. INTRODUCTION

Social media has evolved into a vibrant platform where people communicate freely with each other, share ideas and comment on various events and issues. Twitter is one of the evolving social media which, on average, hosts around 200 million tweets per day. Tweets generally correspond to infinite number of products, services, social issues, news, incidents and reviews etc. Further, people also comment and share views about tweets pertaining to various topics/issues. Twitter shares these tweets in a unique way, which cannot be out rightly used to judge the essence of the tweets and topics being discussed on social media. Planned events include general elections in a country, music concert, educational or employment workshops, sports tournament etc. and unplanned events pertain to unanticipated, out of the blue and sudden incident such as earthquakes, hurricanes, bomb blasts, spot-fixing etc. People express their views/comments about any event of interest and share the latest information about the particular event/incident. This can be quite useful for creating awareness and getting solution to the problem as well as ascertaining general public‘s trend on that issue.

Tweets can correspond to innumerable products, services, social issues, news, incidents and reviews etc. Further, people also comment and share views about tweets pertaining to various topics/issues. Different organizations rely on such information to analyze and evaluate the customers‘and consumers‘views about their products or services. Twitter share tweets in a unique way, which cannot out rightly be used to judge the essence of the tweets and topics being discussed on social media. For example, TV channels use tweets as a major source of feedback about their programs and talk shows.

Social media such as Facebook, Twitter etc. is one of the fast evolving phenomena for sentiment analysis to know how people think about a particular event. In the present day technology driven culture, we can get opinions from different polls and advertisements placed on blogs and social media sources.

In general, human beings have natural instinct to share information or give feedback about the product or items they purchase in their daily lives. And this very trait of sharing information has now moved on to social media sites like Twitter, Facebook, Linkedin and other microblogging sites. By this means, these sources are becoming very useful in identifying and analyzing diverse opinions on different topics and areas. Twitter is one of the important sources for getting opinion from microblogging data available in different languages. Such data can be obtained through the Twitter‘s ―Tweet Entities‖ using various applications.

Two types of analyses, trend analysis and sentimental analysis, can be highly beneficial to determine how people think and get emotional on certain social, religious or political issues. One reason for trend analysis can be to detect an emergent or suspicious behaviour happening on the social media platform. For example, trend analysis can be used to see how certain groups of people are using it to launch their propaganda or forging facts about certain political or religious issues. Corporate sector can also use it to get feedback about their products.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 2, February 2016

Social media trend can be broadly categorized into four types namely positive, negative, neutral and uninterested. The level of trend can further be classified into low, medium, high which will mainly be linked to the critical comments/tweets. Sentiments are evaluated and extracted from the social media content, which can either be in positive or negative attitude. The positive attitude of a person can be conceived as being happy or pleased with the content expressed by someone on certain issue. On the other hand, negative expressions can pertain to being unhappy or angry with the content posted by someone on certain issue.

Trend analysis in a traditional sense can be defined as the frequently mentioned topics throughout the stream of user activity [1]. Hence, for generating an effective trend out of the social media content, the need for an automated classifier becomes necessary to reduce the time for analyzing the large amount of data and improving efficiency of the analysis process.

Sentiment analysis basically tries to judge different aspects of natural language which help people to find valuable information from large amount of unstructured data [2]. It is an emerging concept in which different human emotions are determined from textual content. It enables us to extract opinions and sentimental feelings of the people. To know people‘s opinion about a particular event and its future impact (commonly termed as social media trends), there is a need for an automated system that can analyze such a huge amount of data and produce desired results with certain level of accuracy so that such results can be made acceptable by the masses. On Internet, people use blog posts and forums for promoting products or services as well as discussing any topic and expressing their views. The sentiment analysis on this platform possesses very important information for security analysts to keep an eye on the activities of miscreants and terrorists etc.

However, it becomes a serious challenge to perform such types of analysis on a big data. Numerous events take place regularly in our daily life, therefore, it is not possible to manually analyze every event and predict its future impact. It is really hard for the computing machines to automatically extract the meaning and tone of content as people express so many things in many different ways and styles etc. Sentiment analyses can prove very useful when we analyze search engine results, different blogs, social networks, web forums, different review of people on books, movies, sport and products etc [3]. This can help reduce the efforts required to go through large amount of documents to generate an opinion about the nature of the contents.

II. LITERATURE SURVEY

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Starting from being a document level classification task

Some of the early and recent results on sentiment analysis of Twitter data are by Go et al.[4], Bermingham and Smeaton[5] and Pak and Paroubek [7]. Go et al.[4] used distant learning to acquire sentiment data. They use tweets ending in positive emoticons like ":)" ":-)" as positive and negative emoticons like ":(" ":-(" as negative. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM), and they report SVM outperforms other classifiers.

Since Twitter data is one of the important sources of microblogging platform, so it has been used for sentiment analysis and Pak et al. [7] used this corpus for sentiment analysis and opinion mining. The corpus containing emotions like happy smiley ―:)‖ or sad smiley ―:(‖ are readily evaluated as positive or negative sentiments respectively. After corpus collection, it is analyzed to check how data has been distributed into subjective corpus (containing positive or negative set) and objective corpus (containing neutral set). The authors calculated the presence of n-gram for extracting binary feature and keyword frequency was used to obtain rest of the general information. The analytics reported by the authors showed that objective sets contained more common and proper nouns which in turn have often used personal pronouns. Similarly, the objective sets bloggers addressed themselves as third person while the subjective sets bloggers described themselves as first person or second person.

Lima et al. [8] propose a classifier for Twitter messages that comprises three modules: Support Counting, Database Selection and Classification modules. Support counting module counts percentage of the tweets that contain at least one word or emotion in the tweet. Database Selection module divides data into two sets: training set and testing set. Classification module classifies data using Naïve Bayes algorithm.

Another significant effort for sentiment classification on Twitter data is by Barbosa and Feng [9]. They use polarity predictions from three websites as noisy labels to train a model and use 1000 manually labelled tweets for tuning and another 1000 manually labelled tweets for testing. They however do not mention how they collect their test data. They propose the use of syntax features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words.

Gamon [10] perform sentiment analysis on feedback data from Global Support Services survey. One aim of their paper is to analyze the role of linguistic features like POS tags. They perform extensive feature analysis and feature selection and demonstrate that abstract linguistic analysis features contributes to the classifier accuracy.

Sakaki et al. [11] propose earthquake detection and reporting system which sends email alerts to the registered users when it gets any tweet originated from Japan only about an earthquake. In the proposed system, all the Twitter users are considered as a sensor because they send sensory information. After every second, the system searches tweets which match the given keywords and applies semantic analysis to get accurate results. To know the location or area where that event has occurred, the system uses event detection algorithm. The proposed algorithm uses search API to get the time and location of a

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 2, February 2016

tweet and the same is automatically attached with the tweet when it is posted via iPhone or phone that has GPS system. The other alternate it uses for finding event location is to get the registered location of the user through Kalman filter or particle filtering algorithms.

A proximity-based sentiment analysis is proposed by Hasan et al. [12] that uses features based on word proximities within a sentence. The authors used three proximity-based features which are called proximity distribution, mutual information between proximity types, and proximity patterns. The dataset was divided into number of segments in which each segment contained over 100 words. The distance between positive and negative pair of word was calculated. Three proximity based features were used. In Proximity Distributions, different numbers of bin were considered which returned the distribution of pair-wise distances from the proximity models. In Mutual Information between Proximity Types, the relationship between the proximity types was used to determine the polarity of the document. Then, the theoretic quantities of entropy for each sequence, was used to get mutual information between pairs of proximity types. In Proximity Patterns, it described polarity of words which were used in the document.

Mizumoto et al. [13] proposed the system in which the author created polarity dictionary to determine the sentimental polarities of stock market. While constructing the dictionary, the author has used the semi-supervised learning approach from which small polarity dictionary is made. Using the co-occurrence frequency with words in polarity dictionary, those new words are added to the dictionary whose polarities are unknown. For estimating the polarity of the text, the author has used sentiment analysis method. The polarity of article is determined according to the frequency of words in the polarity diction; hence, the articles are determined as positive, negative or neutral.

In [14], the authors introduced text sentiment classification for the contextual information. For this purpose, the flow of the sentiments and keywords in the paragraph were taken out from the contextual information. Finally, by computing the contextual information degree (linearly combined weighted sum of contextual information), the overall sentiment of sentence was classified.

Davies et al [15] proposed a new language-independent model for sentiment analysis of short statuses. They demonstrated this on data from Twitter, modelling happy vs sad sentiment, and show that in some circumstances this outperforms similar Naive Bayes models by more than 10%.

III. RELATED WORK

Today, different services are offered by Google Maps such as detailed information about geographical regions and sites around the world. In addition to conventional road maps, Google Maps offers aerial and satellite views of many places. In some cities, Google Maps offers street views comprising of photographs taken from vehicles. The purpose of this project is to create a web-based application that uses Twitter APIs to access localization services and social network information.

Using the Twitter API the user's location will be retrieved along with the depiction of the display picture on the current location on the map. An API is a set of functions and procedures that allow the creation of applications which access the features or data of an operating system, application, or other service. Addition of the click action will enable the user to retrieve the latest Tweets. The users can access their previously visited locations from which they had tweeted. The user will also have access to all the current topics of discussion popular in their area.

A global overview of the hashtags will also be provided. 'Trendsmaps' [20] is one of the services which offers real-time mapping of Twitter trends across the world. They are displayed as hashtags, @mentions or keywords super-imposed over a world map. The User can click on any word to see a real-time stream of relevant tweets to, in the words of the site, "see what the global, collective mass of humanity are discussing right now."

One of the main focuses of the application is usability therefore the display of information will be user-friendly by using a clear and simple design. The main user interface consists of the overview of the Google Map all over the world.

Another angle to the Trendsmap [20] is the sentimental analysis feature that will be incorporated onto the Google Map. Here, we aim to examine the sentimental analysis of Twitter data. In this project, we investigate the utility of linguistic features for detecting the sentiment of Twitter messages. We evaluate the usefulness of existing lexical resources as well as features that capture information about the informal and creative language used in microblogging. We take a supervised approach to the problem, but leverage existing hashtags in the Twitter data for building training data.

Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life.

In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many times these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment. Our project shows the features that have to deal with Twitter-specific features such as emoticons, hashtags and various words with a positive or negative score.

IV. MARKET ANALYSIS AND BENEFITS

This project is a great tool for anyone who needs to or wants to stay on top of popular topics. The application utilizes Tweets, which might reveal some interesting trends throughout the world. The application is very easy to use which makes it appealing to almost anyone.

Expected Outcome

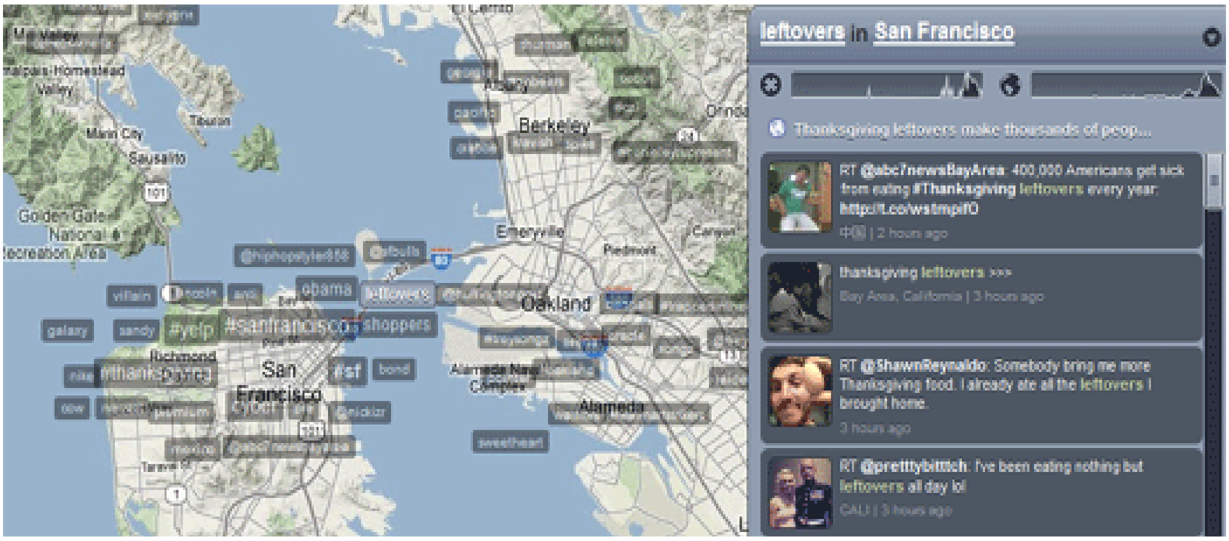

Fig 1: references: Google

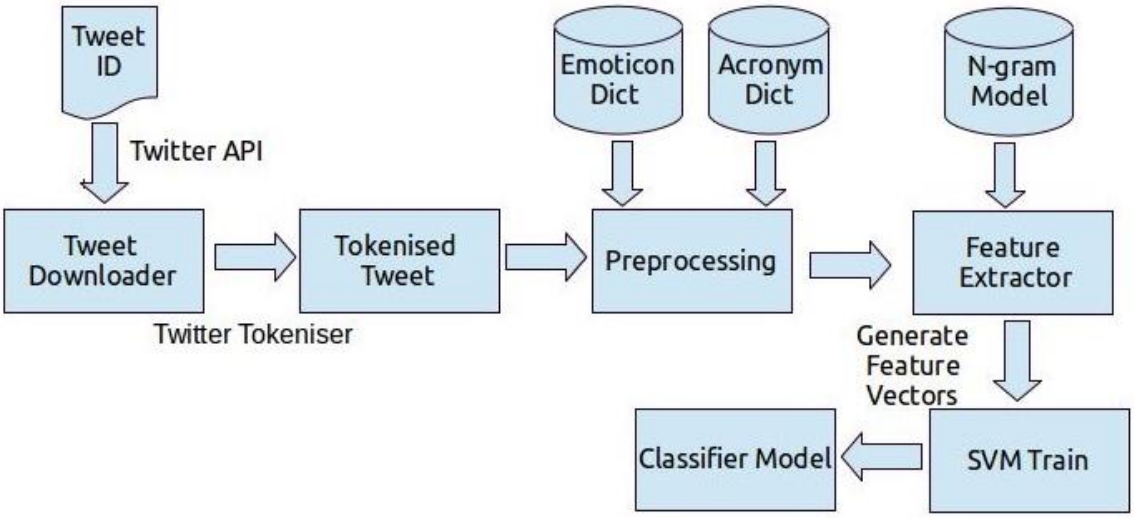

Fig 2: https://researchweb.iiit.ac.in/~ayushi.dalmia/projects/sentimentanalysisintwitter.html

It will be a great journalism tool as it's really good for doing research, for tracking tweets and for getting that one sentence in the story that speaks to the broader social media perspective in a story.

It will be of great help for advertising related fields as it's not uncommon for clients to ask about the current hot topics. More than ever, it's important for planners to stay on top of the trends that can impact their clients' business. The tool will not just display the trends but also the tweets related to it too. This kind of information can give us more insight into the kinds of topics that are important to people as it can help us stay aware of the kinds of topics people and communities are talking about.

V. CONCLUSION

In this study, we looked into different techniques used for trend analysis on social media such as Twitter. We also studied that how data which is available on social media can be used in different ways to analyze and predict future trends. We observed that there is not any flexible system that has data dictionary with more appropriate keywords for predicting trends over Facebook. As most of the work is done using Twitter as a source, we have plan to design and develop a system to display twitter trends on Google map interface by collecting all type of public posts and related user comments and applying sentiment analysis. We also intend to propose a model that is capable to analyze different events being discussed on social media. Since, most of the content available on Twitter is unstructured text, therefore, we have planned to develop the system which can automatically analyze sentiments from the available content and verify the opinions that are expressed in the contextual information.

REFERENCES

[1] Budak, C., Agrawal, D., & El Abbadi, A. (2011). Structural trend analysis for online social networks. Proceedings of the VLDB Endowment, 4(10), 646-656

[2] Bloom, K., Garg, N., & Argamon, S. (2007). Extracting appraisal expressions. HLT-NAACL 2007, 308-315.

[3] Hasan, S. S., & Adjeroh, D. A. (2011). Proximity-based sentiment analysis. In Applications of Digital Information and Web

Technologies (ICADIWT), 2011 Fourth International Conference on the (pp. 106-111). IEEE.
[4] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.
[5] Adam Bermingham and Alan Smeaton. 2010. Classifying sentiment in microblogs: is brevity an advantage is brevity an advantage? ACM, pages 1833–1836
[6] Apoorv Agarwal, Fadi Biadsy, and Kathleen Mckeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 24–32, March.
[7] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC
[8] Lima, A. C., & de Castro, L. N. (2012, November). Automatic sentiment analysis of Twitter messages. In Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on (pp. 52-57). IEEE.
[9] Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.
[10] Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. Proceedings of the 20th international conference on Computational Linguistics
[11] Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World Wide Web (pp. 851-860). ACM
[12] Hasan, S. S., & Adjeroh, D. A. (2011). Proximity-based sentiment analysis. In Applications of Digital Information and Web Technologies (ICADIWT), 2011 Fourth International Conference on the (pp. 106-111). IEEE.
[13] Mizumoto, K., Yanagimoto, H., & Yoshioka, M. (2012, May). Sentiment Analysis of Stock Market News with Semi-supervised Learning. In Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on (pp. 325-328). IEEE.
[14] Colbaugh, R., & Glass, K. (2011, September). Agile Sentiment Analysis of Social Media Content for Security Informatics Applications. In Intelligence and Security Informatics Conference (EISIC), 2011 European (pp. 327- 331). IEEE.
[15] Davies et al, Language-independent Bayesian sentiment mining of Twitter
[16] James Benhardus(2010), Streaming Trend detection. 2010, UCCS REU for Artificial Intelligence, Natural Language Processing and Information retrieval
[17] Vineeth Rakesh, Chandan K. Reddy, Dilpreet Singh (2013), Location-Specific Tweet Detection and Topic Summarization in Twitter, 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Department of Computer Science, Wayne State University, Detroit, MI, USA.
[18] https://engineering.twitter.com/opensource
[19] http://developer.twitter.com
[20] http://www.trendsmap.com