INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 2)

Available online at: www.ijariit.com

Google maps based trending decision

Anuj Upadhyay
anujupadhyay151296@gmail.com
IMS Engineering College, Ghaziabad, Uttar Pradesh

Ashu Dikshit
ashudikshit356@gmail.com
IMS Engineering College, Ghaziabad, Uttar Pradesh

Anurag Mishra
anurag.mishra@imsec.ac.in
IMS Engineering College, Ghaziabad, Uttar Pradesh

ABSTRACT

In the last decade, social network has gained remarkable attention. People access social network sites such as Twitter, Facebook LinkedIn, and Google+. People are becoming more interested in and depend on the social network for information, news, and opinion of other users on different subject matters. The heavy dependency on social network sites causes them to generate massive data. The main contribution of this project is the integration of sentimental analysis on location-based Twitter data. Another contribution of this project is the development of a novel user interface, which allows the user to search on a map interactively on Google map. This integrated sentiment analysis works efficiently and performs location-based popularity on persons, products, event or any given topic.

Keywords: *Google Maps API, Support Vector Machine (SVM), Twitter Data (TWD).*

1. INTRODUCTION

The recent explosion of social networks is so vast that it has become a very important platform for communication on the web. It is getting popular every day. Every user posts their daily moments and their thoughts and opinions regarding different topics in the social media, users post messages about their daily life and initiate discussions by sharing personal opinions and emotions on different topics. According to C. Aggarwal [1] "The richness of this network provides unprecedented opportunities for data analytics in the context of social networks [1]". Data mining techniques may detect implicit or hidden patterns within a social networking site. This technique provides feedback to sense user sentiments for purchasing behavior, identification of social groups and understanding the hidden trends in network evolution.

2. PROBLEM STATEMENT

Analysis of the trends or the popularity of any topic (which may be related to any product, person, events etc.) by analyzing the different public statements gathered from different social media sources and to find whether the tweet or the response of the person is positive, negative or neutral towards the particular topic.

3. WORKING TECHNOLOGY

A. DATA MINING

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. It is an essential process where intelligent methods are applied to extract data patterns. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD.

B. GOOGLE MAPS API

Google Maps is a web mapping service that offers satellite images, street view and maps through Google Maps API, released in June 2005. This API allows users to embed robust functionality to explore the world. It also allows users to identify map locations with custom markers. Google Maps API is a free service for private or commercial use. The proposed system has integrated Google

API for location selection, data interaction, and visualization. Once the application is run, the very first interface will be with Google Map. If we know the location we can search tweets from the region. Even if we do not know the location we can simply search tweets and identify the targeted location by density.

C. TWITTER API

Twitter allows users to interact with its data using Twitter APIs. Twitter has three API (Rest, Search and Streaming). "The Rest API provides simple interfaces for most Twitter functionality. The Twitter Search API is part of Twitter's v1.1 REST API. It allows queries against the indices of recent or popular Tweets and behaves similarity to, but not exactly like the Search feature available in Twitter mobile or web clients, such as Twitter.com search. Before getting involved, it's important to know that the Search API is focused on relevance and not completeness. This means that some Tweets and users may be missing from search results. If you want to match for completeness you should consider using a Streaming API instead. The Streaming API is a family of powerful real-time APIs for Tweets and other social events" [10]. For efficient data extraction server-side programming/scripting language such as PHP, RUBY or PYTHON is needed. An Oath-authentication is required to access data from Twitter. Four authentication parameters (consumer key, consumer secret, Oauth_token, Oauth_secret) are needed to enable API call. These keys look like following:

Consumer key xvz1evFS4wEEPTGEFPHBog

Consumer secret L8qq9PZyRg6ieKGEKhZolGC0vJWLw8iEJ88DRdyOg

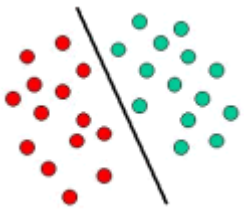
OAUTH_TOKEN 5501815334-YG8YkYD8tjFLrT3zw61GmhT1uS8iqvzQxojsu49

OAUTH_SECRET bfK5P3kZVsfOwgNrM0Gjab56BhLDRiYgMhGRAzxIRry5b

Each application has different Consumer-key, Consumer-secret, Oauth_token, and Oauth_secret. Once the request is sent to twitter API with proper authentication it will deliver results in JSON (JavaScript Object Notation) format. JSON format is very popular and can easily be read by any program.

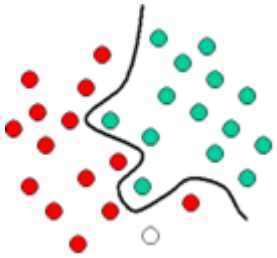
D. SUPPORT VECTOR MACHINE

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the illustration below. In this example, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object (white circle) falling to the right is labeled, i.e., classified, as GREEN (or classified as RED should it fall to the left of the separating line).



Linear classification by hyperplane

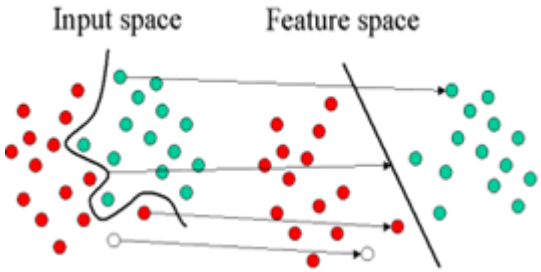
The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (train cases). This situation is depicted in the illustration below. Compared to the previous schematic, it is clear that a full separation of the GREEN and RED objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyper plane classifiers. Support Vector Machines are particularly suited to handle such tasks.



Nonlinear classification by hyper plane

The illustration below shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly

separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the GREEN and the RED objects.



Transformation classification

4. BACKGROUND LITERATURE SURVEY

A. TWITTER

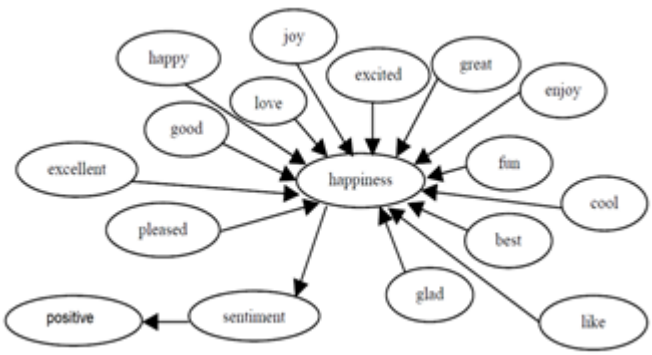
Twitter is an online social media website. Twitter allows its users to communicate through micro-blogging services and opens a way to share their feelings, opinions or messages to other users over the internet. Twitter was developed in 2006. Currently, Twitter is one of the top three popular online social media websites [2]. It provides micro-blog services to write messages up to 140 characters at one time, which are typically not more than 30 words. As of September 2013, Twitter has 645 million users and who produced about 300 billion tweets. More than 143,199 Tweets are tweeted (i.e. transmitted or delivered) per second [3].Twitter uses several character symbols for performing operations. For instance, '@USER_ID' transmits direct messages, 'RT' for performing re-tweet. To enforce a category or a topic discussion '#' (hash tag) is widely used. Twitter itself also suggests its users use hash tag for putting extra wastage of the words. Approximately 98% users do not have enough followers [4]. Some of the tweeter users have millions of followers but they are very few in number (in the year 2010) [5]. These users are a different kind of media-stars, politicians or popular news web sites (CNN, newspaper websites etc.). The follower is very important to keep the flow of tweeting. A tweet user must have at least 10 followers for a regular flow, otherwise, they will lose the interest of tweeting [6]. Twitter has a very high number of users and many new users are joining every day. Location-enabled tweet can play an important role in business [7]. Benjamin and Krzyszt of mentioned that the geographic-location is a key component for information retrieval on the web for recommendation systems especially in mobile computing, social networks, and place-based integration.

B. LOCATION BASED ANALYSIS WITH SEMANTIC WEB

Location-based system and services have a great influence on our daily life. People experience these influences through the usage of smartphone and location enabled electronic gadgets. Widely used services through these devices include weather forecasts checking, traffic updates, locations of nearby gas stations, restaurants etc. and sometimes as a GPS device. Celino et al. [8] have developed an android application BOTTARI that processes social media content to identify people's point of interest (POI) and visualize it on Google map. The system has integrated semantic analysis for identifying people's influence through location-based microblog postings. They have used machine learning algorithm SVM, rule base approach for language processing and a kernel syllable for sentiment elicitation. This system can process only single location based objects but our system can process multiple user-defined objects on targeted locations.

C. ONTOLOGY-BASED TEXT MINING

Ontology is a metaphysical nature of being, existence, becoming or reality, along with the basic categories of being relations. In the wide sense human ontology is to investigate the relation relying on institutional, social, and technological conventions representing an intellectual activity. Sam et al. have shown that the emotional-ontology from social networking sites combined with product-ontology can analyze the customer behavior [9]. For Ontology-based Text-Mining, unstructured data inside the tweet are analyzed for a particular topic and extracted keywords. Then the ontology of the particular topic and emotions from text mining are matched to discover behavior of the consumer in the market. For example let us consider the class sentiment of "happiness". The ontology subclass of "happiness" has been shown



Ontology sample

5. LITERATURE SURVEY

We have to survey from the other author research paper and they have to measure some related work of our research project. They are:

1) This research aims to analyze location-based social media data to measure the trends of the products or persons or any given user parameter. For this purpose this work has integrated sentimental analysis, location-based system. An application with a novel user interface has been developed to search and visualize the data on Google map. This research work uses publicly available and location-enabled social media data. This work also has the capability to process tweet data without user's locations.

2) Data mining is the extraction of projecting information from large data sets, is a great innovative technology which helps corporations focus on the most important information in their data stockrooms. Data mining makes use of various statistical, machine learning and graphical methods and separate the knowledge in to a form which is very much useful for many real-world applications. Social network analysis has become a very popular field of research as it is useful for many applications. In this paper we have overviewed various data mining techniques used for social network analysis. Keywords: Data Mining, Graph Mining, Social Networks.

3) Social network has gained remarkable attention in the last decade. Accessing social network sites such as Twitter, Facebook LinkedIn and Google+ through the internet and the web 2.0 technologies has become more affordable. People are becoming more interested in and relying on social network for information, news and opinion of other users on diverse subject matters. The heavy reliance on social network sites causes them to generate massive data characterised by three computational issues namely; size, noise and dynamism. These issues often make social network data very complex to analyse manually, resulting in the pertinent use of computational means of analysing them. Data mining provides a wide range of techniques for detecting useful knowledge from massive datasets like trends, patterns. Data mining techniques are used for information retrieval, statistical modelling and machine learning. These techniques employ data pre-processing, data analysis, and data interpretation processes in the course of data analysis. This survey discusses different data mining techniques used in mining diverse aspects of the social network over decades going from the historical techniques to the up-to-date models.

4) This paper provides a methodology and a critique of social media tools. Analyzing social media, in particular Twitter feeds for sentiment analysis, has become a major research and business activity due to the availability of web-based application programming interfaces (APIs) provided by Twitter, Facebook and News services. This has led to an 'explosion' of data services, software tools for scraping and analysis and social media analytics platforms. It is also a research area undergoing rapid change and evolution due to commercial pressures and the potential for using social media data for computational (social science) research. Using a simple taxonomy, this paper provides a review of leading software tools and how to use them to scrape, cleanse and analyze the spectrum of social media. In addition, it discussed the requirement of an experimental computational environment for social media research and presents as an illustration the system architecture of a social media (analytics) platform built by University College London.

5) Research on social media has intensified in the past few years as it is seen as a means of garnering insight into human behaviors. The unstructured nature of social media data also provides unique challenges and opportunities for researchers across a variety of disciplines. Businesses are tapping into social media as a rich source of information for product design, relations management and marketing. Scientists utilize social media data as a platform for developing new algorithms for text mining and sentiment analysis and focus on social media as a sensor network for natural experimentation for exploring social interactions and their implications. Currently, the visual analytics community has begun focusing on social media analytics with respect to developing tools and frameworks to collect, monitor, analyze and visualize social media data. Studies have ranged from geo-temporal anomaly detection to topic extraction to customer sentiment analysis. Such work focuses on capturing the incoming streams and enables the analysts to perform exploratory data analysis. However, little work has been done on developing tools for predictive analytics using social media. In 2013, the Visual Analytics Science and Technology (VAST) conference ran the VAST Box Office challenge using social media data to predict the opening weekend gross of movies. This particular contest served as an entry point to explore how users interact with visualization tools to develop predictions. Continuing from this contest, our work has focused on utilizing movie data from social media to explore the promises and pitfalls of visualization for predictive analytics. In this paper, we present a framework for social media integration, analysis and prediction. This framework consists of tools for extracting, analyzing and modeling trends across various social media platforms

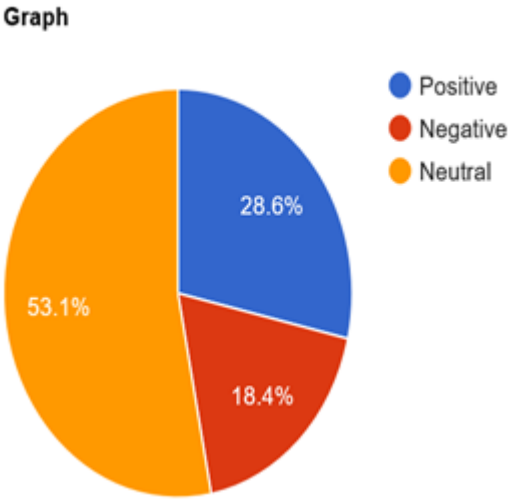
6. METHODOLOGY

1) EXPERIMENT

This research is performing the experiment to measure the popularity of our proposed system and the data collection spanned over couple of months. The whole system was divided in multiple modules and was implemented in different time frames. Twitter does not allow accessing its database directly; we had to collect the data with a data crawler and stored it in a local database. So the beginning of our research implementation, the very first module coded was a data crawler which was scheduled to run to collect the data set on given individuals. Since data collection was a time consuming process other modules were developed as data crawler was working on the background. Once we had adequate data, the system was online to produce final result.

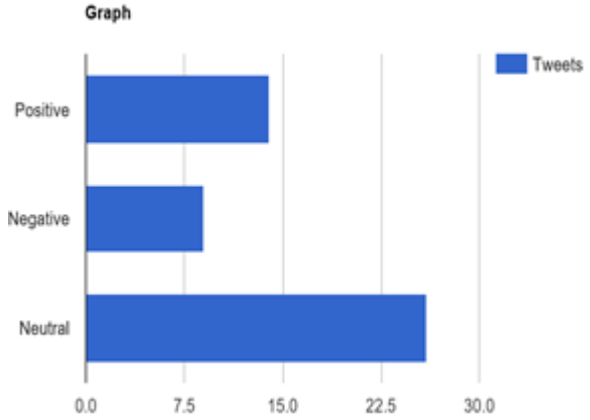
2) DATA VALIDATION

The below figure shows the percentage of positive, negative and neutral tweets about cricket in different locations. In this figure the percentage of positive tweets is 28.6%, negative tweets is 18.4% and neutral tweets is 53.1%.



3) Pie chart for input cricket

After processing the tweets we are also showing the results on a bar chart. The percentage of positive tweets is 28.6%, negative tweets are 18.4% and neutral tweets is 53.1% for cricket.



Bbar chart for input cricket

7. CHALLENGES AND LIMITATIONS

- Location-based and Location independent data extraction.
- Data extraction from social media sources.
- Sentiment analysis of the extracted data.
- Trend scaling.
- Deception detection: Information intended to deceive can spread through social media the same as the valid information in different social channels and formats (for example text, link, audio, video, and photo, multimedia).

8. CONCLUSION

This research has developed a user-friendly interface with Google Map. Users can draw single /multiple locations on Google map to perform location-based search. The application can extract location independent (the tweet has no location) twitter data. It allows performing a semantic search for tweets either by keyword. The application sends a request to Twitter API through Data Crawler. The Crawler collects related tweet data and generates data visualization instantly on the map with Google markers. The Text processor examines the data to decide what actions are needed to perform, store and manage the data. Data cleaning process cleans the no-structured format tweet content written in the informal language. Word Parsing process splits the cleaned tweet string into parts and develops an array of words. Word-Matching function compares each word with lexicon dictionary which returns numerical values. Score Calculation process adds up the returned values by word-matching function and generates the final score for a tweet string.

9. REFERENCES

[1] C. Aggarwal, "An Introduction to Social Network Data Analytics in Social Network Data Analytics", New York: Kluwer Academic, January 2011.
[2] L. Barbosa and J. Feng, "Robust sentiment detection on Twitter from biased and noisy data," 23rd International Conference on Computational Linguistics, Beijing, China, 2010.
[3] Digital Marketing Ramblings http://expandedramblings.com/

[4] B. Huber, D. M. Romero and F. Wu, "Social networks that matter: Twitter under the microscope," website, 2008.
[5] H. Kwak , C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in Proc. Int. World Wide Web Conf. , San Francisco,2010.
[6] N. Jamil and A. Alhadi, "A Collaborative Names Recommendation in the Twitter Environment based on Location," in Semantic Technology and Information Retrieval (STAIR), International Conference, Putrajaya, Malaysia, 2011.
[7] A. Benjamin and J. Krzysztof, "On the Geo-Indicativeness of Non-Georeferenced Text," in 6th International AAAI Conference on Weblogs and Social Media, Dublin, Ireland 2012.
[8] I. Celino, D. Daniele, E. D. Valle, B. Marco, Y. Huang, L. Tony, K. Seon-Ho, and T.Volker, "Bottari: Location based social media analysis with semantic web," International Semantic Web Conference (ISWC), Bonn, Germany 2011.
[9] K. Sam and R. Chatwin , "Ontology-based text-mining model for social network analysis," in 2012 IEEE International Conference on Management of Innovation &Technology, Bali, Indonesia 2012.
[10] Twitter Documentation https://dev.twitter.com/docs.

