

Harnessing Multi-Source Data about Public Sentiments

M Ranganathappa

Assistant Professor

School of Computer Science and Applications

REVA University

ranganathappam@reva.edu.in

Syed Ameen

PG student

School of Computer Science and Applications

REVA University

ameensa546@gmail.com

ABSTRACT

As net is growing bigger, its horizons are becoming wider. Social Media and Micro blogging platforms like Facebook, Twitter, Instagram dominate in spreading encapsulated information and trending subjects across the globe at a rapid pace. A theme will become trending if greater and greater customers are contributing their opinion and judgements, thereby making it a treasured source of on-line perception. These subjects commonly supposed to unfold attention or to promote public figures, political campaigns for the duration of elections, product endorsements and enjoyment like movies, shows. Thus, there is a huge practicability of discovering and analysing interesting patterns from the countless social media statistics for business-driven applications.

Index Terms- **Opinion Mining, Sentiment Analysis, Sentic Computing.**

Introduction

The word social media refers to the use of web based and cellular applied sciences to flip verbal exchange into an interactive dialogue. Social media takes on many distinct forms inclusive of magazines, Internet forums, weblogs, social blogs, micro blogging, wikis, podcasts, pictures or pictures, video, ranking and social bookmarking. With the world during a social media revolution, it is more than apparent that social media like Facebook, Twitter, Instagram, skype etc., are used significantly for the motive of communication. This structure of communication can be with a man or woman or a group of persons. Today most of the people especially the children are hooked on to the one of a kind social media for retaining in contact with their peers.

In present scenario an increasing number of population are using social media to express their opinions on various social media application using sentiments, as a result of which a vast amount of unstructured opinionated data has become available. By analysing this data for sentiments, we can extract the public's opinion on several classes and use the conclusions derived from this to make informed choices and predictions concerning those subjects [1]. However, due to the volume of data generated, manual sentiment analysis is not feasible. Thus, automated sentiment analysis is becoming exceedingly popular. 5.

Methodology

First off, let's talk about what social media is. Social media represents low priced equipment that are used to combine technological know-how and social interaction with the use of words. These tools are usually net or mobile based. A few that you have probably heard of consist of Twitter, Facebook, and YouTube. Sources of data: Data have been

Polarity classification is a sub-task of sentiment analysis that focusses on classifying text into effective and negative, or positive, negative and neutral. Document-level polarity classification involves figuring out the polarity of opinions expressed in an whole document, whereas sentence-level polarity classification involves determining the polarity of opinions expressed in a single sentence. Another stage of polarity classification is aspect-based polarity classification, which involves extracting opinion ambitions from text and then determining the polarity of the text in the direction of that specific target. Surveys of strategies used for a number of levels of sentiment analysis.

Allocate resource

To analyse the have an impact on Social Media as a capability of Communication tools.

To check the usefulness of social media as a tool of communication

The social media activity worldwide is mind blowing:

1 billion Facebook users, 42 million pages and 9 million apps

4 billion views per day on YouTube

400 million registered users on Google+

175 million LinkedIn users

140 million Twitter users

collected from each primary and secondary source. Primary sources consist of survey work accomplished among the residents of Guwahati city. Document evaluation of the secondary sources includes internet, magazines, books and journals has been carried out to study the affect and the sue of social media in the social context and its conceivable for future growth.

Twitter

Twitter is an online networking site driven by tweets which are 140-character limited messages. Thus, the character limit enforces the use of hashtags for text classification. Currently around 6500 tweets are published per second, which results in approximately 561.6 million tweets per day [1]. These streams of tweets are generally noisy reflecting multi topic, changing attitudes information in unfiltered and unstructured format. Twitter sentiment analysis involves the use of natural language processing to extract, identify to characterize the sentiment content.

Sentiment Analysis

Sentiment analysis is the prediction of emotions in a word, sentence or corpus of documents. It is intended to serve as an application to understand the attitudes, opinions and emotions expressed within an online mention. The intention is to gain an overview of the wider public opinion behind certain topics. Precisely, it is a paradigm of categorizing conversations into positive, negative or neutral labels. Many people use social media sites for networking with other people and to stay up-to-date with news and current events. These sites (Twitter, Facebook, Instagram, google+) offer a platform to people to voice their opinions. These information forms a basis for people to evaluate, rate about the performance of not only any movie but about other products and to know about whether it will be a success or not. This type of vast information on these sites can used for marketing and social studies [1]. Therefore, sentiment analysis has wide applications and include emotion mining, polarity, classification and influence analysis.

Supervised Learning for Sentiment Analysis

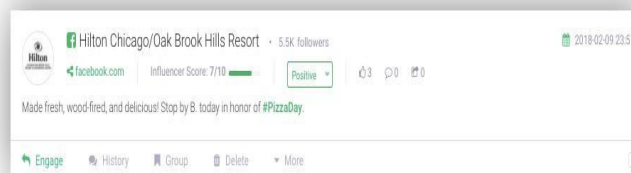
A text pattern is converted to a function vector that represents its most essential characteristics. Given the feature vectors X and class labels Y for N number of training tweets, the supervised learning algorithm approximates a feature F such that $F(X) = Y$. Now, in the testing phase, given feature vectors $X_{_}$ for T number of unlabelled tweets, the feature F predicts labels $Y_{_}$ using $F(X_{_}) = Y_{_}$ for every of the unlabelled tweets. The most commonly used points for sentiment analysis are time period presence and term frequency of single tokens or unigrams. The use of higher order n-grams (presence or frequency of 2,3,...n contiguous tokens in a text) such as bigrams and trigrams is also prevalent, and allows for encoding of the tokens' positional information in the feature vector. Parts-of-speech and negation based aspects are also commonly used in sentiment analysis. exploitmicroblogging text or Twitter-specific elements such as emoticons, hashtags, URLs, @symbols, capitalisations, and elongations to enhance sentiment evaluation of tweets.

Unsupervised Learning and Linguistic Rules for Sentiment Analysis

Usually, unsupervised methods for sentiment analysis such as [21] contain first creating a sentiment lexicon in an unsupervised manner, and then determining the polarity Modeling Public Sentiment in Twitter 53 of a textual content using some characteristic dependent on the number or measure of positive and poor words and/or phrases present in the text. A comparison of supervised methods and other unsupervised techniques can be found in [22].In [9], the authors define dependency-based linguistic rules for sentiment analysis and merge those rules with frequent sense knowledge, and machine learning to enhance sentiment analysis. Our proposed method is based totally on the concept illustrated in [9], alternatively the linguistic rules we define are restrained and now not dependency based, because most dependency parsers do now not perform nicely for micro blogging texts such as tweets. Moreover, it is appropriate to function sentiment analysis of social media texts in real-time, and dependency parsers can't be used in real-time due to the large time complexity of their algorithms. In this paper, our purpose is to create a Twitter sentiment analysis classifier that classifies tweets in real-time while countering the two challenges postulated in 2.

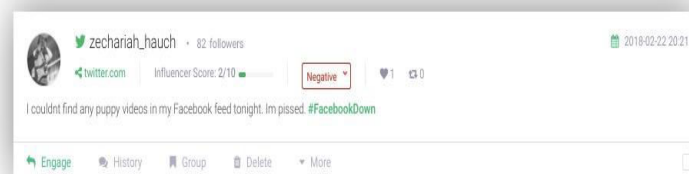
Positive tweets

In the twitter people will see the posts and give there feed back. If the post is good user will use #tag and write his comment on that post. User will write what he/she will feel about post and give there opinion. Here is the example of positive tweets.



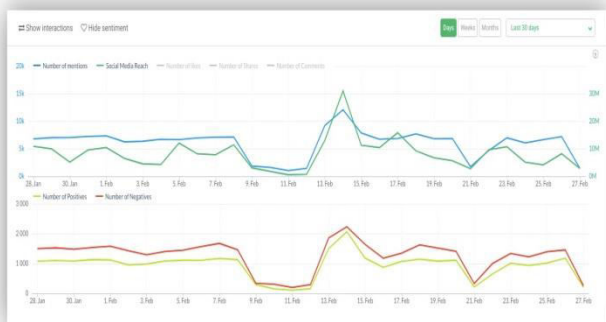
Negative tweets

In the twitter people will see the posts and give there feed back. If the post is bad user will use #tag and write his comment on that post. User will write what he/she will feel about post and give there opinion. Here is the example of negative tweets.



Sentiment analysis chart

This chart shows how it works. About positive and negative tweets about the post.



The Proposed Method

in this process before analysing the raw tweets for sentiment analysis we pre-process them. during pre-processing all the @ <username> converted into @USER and all the URLs are changes to http://URL.com. we use the CMU Twitter Tokeniser and Parts-of-Speech Tagger to tokenise the tweets and assign a parts-of-speech tag to each token. Apart from nouns, verbs, adjectives and adverbs, this tagger is also able to tag injunctions, and micro blogging-specific tokens such as emoticons, hashtags, and URLs.

this will not understand the the nouns, verbs, adjectives and adverbs to give the best opinion. so it will convert into zeros and one's then it will give the results of the tweet.

Emoticon Rules

An emoticon is short form for emotion icon also known as an emote, is a pictorial representation of a facial representation using character. Usually punctuation marks, numbers and letters to express a person's feelings or mood or as a time-saving method. It is an easy method; it will save time. In the instance of typing people's feelings and mood, we can send emojis. A manually created list of usually polar emoticons along with their semantic orientation (*positive or negative*). We match emoticons in the bag-of-emoticons of the tweet to the list of positive or negative emoticons, and count the number of positive and the number of negative emoticons present in the tweet.

Manually Created List of Positive Emoticons

(-: , (: , =) , :) , :-) , =') , :') , :'-) , =-d , =d , ;d , :d , :-d , ^-^ , ^_^ , :] , ^_^ , ^_* , ^^

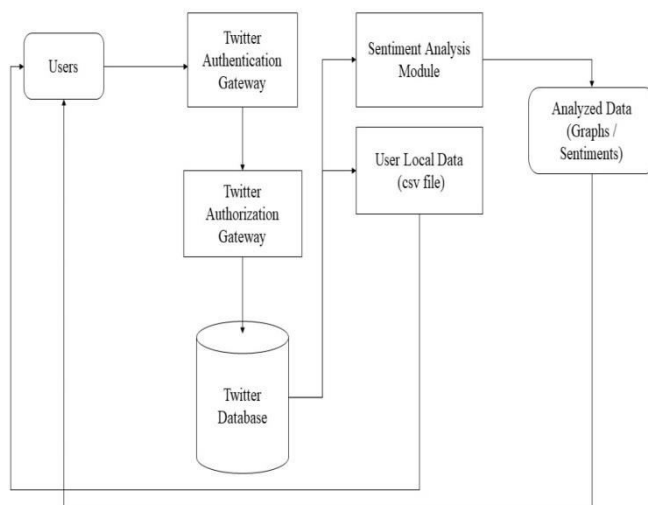
Manually Created List of Negative Emoticons

)-: ,): , =(,]: , :[, :(, :- (, >:(, >:(, :_(, d'x , :(, :)'(, ='[, :'(, :'- (, \: , :/ , (~~) , >__> , <('')> , </3

We apply the following rules for tweet.

- If tweet contains one or more positive tweets it will consider as a positive tweet.
- If tweet contains one or more negative tweets it will consider as a negative tweet.
- If the tweet contains both positive and negative same it will consider as neutral

System architecture



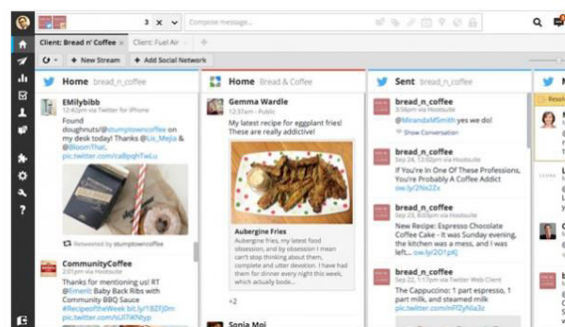
DISCUSSION AND FINDINGS

A survey was conducted among the residents of Guwahati city in the urban areas. Guwahati is the gateway of North-East India. The spurt in the growth of industries has truly made it a world-class city. The total sample size was 200, which consisted of students, teachers, engineers, marketing professionals, businessmen, etc. A careful analysis of the data reveals the following results-

Some tools used in trending are:

Hootsuite: Hootsuite is one of the best free social media listening tools available and covers multiple social networks, including Twitter, Instagram, Facebook, LinkedIn, WordPress, Foursquare, and Google+. It is well known for its social media management functions.

You can also monitor specific search terms in real time. This can be really handy for tracking mentions of our brand, products, or relevant keywords you're interested in.



TweetReach: TweetReach is a great monitoring tool for your business if you're interested in checking how far your tweets travel. TweetReach measures the actual impact and implications of social media discussions.

It is a good way of finding out who are your most influential followers, implicitly guiding you towards the right people you should be targeting when aiming to share and promote online content.



Advantages and Dis-Advantages of Social Media

Advantages

1. Connects with the People

The great element about social media is that it connects to the humans all over the world. Facebook and Google+ is the great channel for connecting with the pals whereas Twitter and LinkedIn are the nice for making the expert contacts. It is well worth checking

2. Easy Communication

It has made the verbal exchange nearly instant. People now no longer have to wait for the replies. There are many social messaging apps such as Facebook messenger apps, Twitter Direct Messages for the professional contacts.

3. Real Time News and Updates

It helps in discovering the updates and information almost instantly. If you desire to know some thing new, then you can without delay check the news feed and update your information with the modern happenings in the world.

4. Great Opportunities

Business owners can join their respectable pages on the social media platforms. This will assist them to sell the products the usage of social media with the aid of accomplishing to the extra customers. Two

5. Fun and Enjoy

People like sharing their posts, photos or videos on social networks. They fulfill if they see the feedback and likes on their very own posts. It is the medium to check what the friends are up to barring asking or contacting them directly.

Disadvantages

1. Information is Devastating

As there are lot of people now on social media sharing, tweeting, and posting videos on daily basis thus information is overwhelming on the internet. Over the time there are lot of followers and friends and thus lot of updates keep popping up on the feed.

2. Privacy Questions

Sharing too much information can sometimes let you in trouble and there are privacy concerns also. There is a location sharing or even sometimes sharing some information can't be undone so you have to think twice before sharing on social media.

3. Social Peer Heaviness or Cyber Harassment

People can take psychological advantage and create pressure on the teens for money or other offline settings. Social media can also lead to cyber bullying and lead to serious stress, anxiety and depression.

4. Substitution for offline Interaction

Due to online connectivity all the time and availability, people are now connected more on social media and stopped meeting offline. This has stopped the physical interactions and group discussions.

5. Affects Lifestyle

It has affected the lifestyle of people as they are now more engaged on social media. They stare at artificial light from the phone screen at night can affect their sleeping patterns as well.

Conclusion

In conclusion, social media is, and will continue to be, harmful, except something is carried out about it. The strength it has over human beings is hazardous and frequently goes unnoticed. People care due to the fact if now not addressed, or taken care, of social media could reason countrywide and global problems. Social media stunts interpersonal increase and increases narcissism in many internet users. Even if the person is now not psychologically affected via social media, they are nonetheless in danger. Nobody is immune to the unsafe side outcomes social media such as identification theft, stalkers, or hackers. Social networking lets in hoaxes and false statistics to purpose sizable chaos. "The darkish facet

of social media is that, inside seconds, whatever can be blown out of percentage and taken out of context. And it is very difficult no longer to get swept up in it all “. If there are no regulations or barriers on social media, the repercussions will accumulate. People want to exchange the damaging patterns of social media use earlier than it destroys society. The fast, effortless get entry to to data that is supposed to be advisable to society is truly dangerous to society.

In this paper, we describe the Twitter sentiment analysis system that enhances supervised learning, by using modified features for supervised learning as well as applying rules based on linguistics and sentic computing. Based on our results, we can conclude that unsupervised emoticon rules and modified n-grams for supervised learning help improve sentiment analysis. They do so by handling peculiar linguistic characteristics introduced by special parts-of-speech such as emoticons, conjunctions and conditionals. Moreover, we have shown that verifying or changing the low-confidence predictions of a supervised classifier using a secondary rule-based (high-confidence, unsupervised) classifier is also immensely beneficial.

Reference

- [1]Li, H., Liu, B.,Mukherjee, A., Shao, J.: Spotting fake reviews using positive-unlabeled learning. *Computación y Sistemas* 18, 467–475 (2014).
- [2] Alonso-Rorís, V.M., Santos Gago, J.M., Pérez Rodríguez, R., Rivas Costa, C., Gómez Carballa, M.A., AnidoRifón, L.: Information extraction in semantic, highly-structured, andsemi-structured web sources. *Polibits* 49, 69–75 (2014)
- [3] Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems* 28, 15–21 (2013)
- [4]Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Foundations and Trends in InformationRetrieval* 2, 1–135 (2008)
- [5] Liu, B.: Sentiment analysis and opinion mining. *Synthesis Lectures on Human LanguageTechnologies*5,1–167(2012)