Positivity and Negativity Attributes of Users in Twitter

Mrs.Sapana Mahaveer Patil

Abstract— **Social media is a platform where people create content, share their opinions, vision and concepts. Examples include Facebook, MySpace, Digg, Twitter and JISC list serves on the academic side, because of its simplicity, speed and reach, social media is rapidly changing the public chat in society and setting trends and agendas in case that limits from the environment and politics to technology and the entertainment industry. Since social media can also be presumed as a form of mutual wisdom, it can be used to predict real-world outcomes. "The Wisdom of Crowds" is about the gathering of info in groups, emerging in decisions that are often better than could have been made by any single member of the group. The popular saying on social media goes as follows: "We use Facebook to record the protests, Twitter to coordinate, and YouTube to tell the world."**

Twitter is the place where we all gather to express precisely the point of view and feelings about specific topics. Opinions reveal beliefs about specific matter commonly considered to be subjective. Twitter has millions of users that spread millions of personal posts on a daily basis. And this gives us the opportunity to study social human subjectivity. Manual classification of thousands of posts for opinion mining task is unfeasible for a human being.

Keywords— **Mood, emotions, twitter, facebook, social network graph.**

I. INTRODUCTION

There has been a study since long time about how the mood and emotion affects on a person's behavior, actions and his/her interactions with other. On the basis of psychological theories on emotions, emotions are generally divided into mainly two groups, i.e. positive and negative.

Precisely, there are two main methods to categorize emotions: a discrete way considers essence of basic emotions that hold anger, disgust, fear, happiness, sadness and surprise, and a multi-dimensional way that classify different emotional states along two or more dimensions. Negative emotions can have terrible corollary leading to firm depressions, family neglect, and acts of violence, criminal activities and self-destruction. As a result, it is vital to find such people suffering from negative emotions in a timely manner, so that crucial aid and guide can be outfitted to them. However, a key difficulty in identifying people with negative emotions is that they often do not propose to share information and stay secluded. Indeed, scientists have been examining distinct ways due to which people with pessimistic emotions may be classified which include the use of a person's physiological signals, body sensing tactics, speech patterns as well as body movement. This field of study has attracted the attention of scientists from many other disciplines.ASONAM 2014, August 17-20, 2014, Beijing, China 978-1-4799-5877-1I14/$31.00 ©2014 IEEE 365 Online social media such as Twitter and Facebook are increasingly being used by people now to share thoughts and opinions and connect with other people. Indeed, there is evidence that even people leading otherwise a lonely life do yield in online social activities. A natural question that arises in this context is: Are there any differentiating aspects presented by people with positive or negative emotions in their online social activities? If so, can we build appropriate classifiers that can identify positive or negative users with high accuracy and low false positive and negative rates? This paper addresses the first question. Social network activities can be used to obtain behavioral aspects of individuals. The main goal of this paper is to investigate differentiating aspects of users with their positive emotions and users with negative emotions through user activities in online social network. In particular, we analyze user activities in a Twitter dataset released by the University of Illinois. This dataset is composed of 3 million users swapping 50 million tweets. Our study is done on the two kinds of user networks: a neutral network and a biased network. For these types of networks, we study behavioral aspects form of the patterns of tweeting, replying and following in the relation of positive and negative users. We consider the bond between positive and negative users, like how they follow and communicate with different groups and how they make and keep their friendships, and also what are the main differences in messaging activities such as to tweet and retweet, replying and mentioning. We also compare the behavior of positive, negative and neutral users through the above analysis. We also examine view point and behavioral variations across 2 groups of users (positive and negative) in both the subsequent alliance and signifying activities. The key finding of this paper is that there are some rare, varying attributes of positive and negative users. We believe that this research can be used to design classifiers to identify positive and negative users.

Opinion mining refers to the use of natural language processing, text analysis and computational linguistics to find and elicitation relevant data in source materials. It aims to determine the attitude of a speaker or a writer against some case or the thorough provisional

Manuscript received November 23,2015
Sapana Mahaveer patil, Information Technology, Savitribai Phule university Pune,Sidhhant College Of Engineering , Pune, India, Mobile No. +918793595443, (e-mail: sapnapatil6@rediffmail.com/ itsapna6@gamil.com).

polarity of a document. The attitude may be his or her judgment or evaluation of affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication.

Business intelligence (BI) systems comprises of operational contents with analytical tools to present complex and ambitious data to architect and choice makers, in order to improve the timeliness and quality of the decision-making course. BI systems contribute actionable information delivered at the right time when decisions need to be made.

Mining the existing contents within social media applications is bright to produce the required information to fit the suited objective, but social media data are mostly vast, unstructured, and dynamic in nature, and thus mining the data is not easy to do.

In recent years, social media has become very famous and important for data sharing, social networking, etc. The data generated from these websites remains largely unused. Social media data largely contains unstructured text. Foremost thing is to retrieve the data in the unstructured text. This paper presents the influence of social media data for analysis and how the content can be used to anticipate real-world decisions that enhance business intelligence, by applying the text mining methods.

This paper presents the influence of social media data for research and how the content can be used to anticipate real-world decisions that enhance business intelligence, by applying the text mining methods. This paper presents a case study where we analyzed publicly available tweets that were posted when a product is launched. We use standard text-mining process to answer questions. This kind of business intelligence can be used by companies in improving product, in targeted marketing to audience, in gathering competitive intelligence and many other ways.

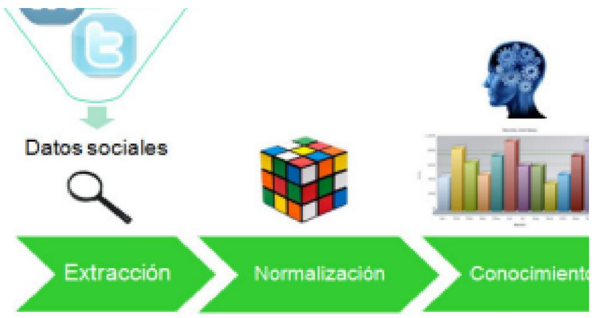


Image: Social media data analysis proces

As we know Twitter is an online social networking and micro blogging service that enables users to publish and read text-based messages upto 140 characters, which are known as "tweets". Over 350 millions tweets are generated daily and a not inconsiderable count of them are akin to your business or industry, don't underestimate this knowledge.

Twitter contains the following items to be studied:

Tweets". A tweet is a post or status update on this social network; they are little pieces of data containing news, conversations, and opinions.

II. LITERATURE SURVEY

We have sought related work on twitter data-mining in three areas: trend detection, sentiment mining and user characterization which are the areas that we cover in the case study.

M. Cataldi, L. Di Caro, and C. Schifanella. Emerging topic detection on twitter based on temporal and social terms evaluation. 2010.

Prior research efforts on topic or trend detection have directed on abrupt activity discovery on the general Twitter platform. Use of aging theory to illustrate term lifecycle of arising topics in real-time has been studied in this paper.

B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11):2169–2188, 2009.

A system that identifies trends by identifying sets of words that come out at a very high recurrence and analyzing them is discussed in this paper. We follow the same concept but define a similarity measure differently. Moreover we are not as interested in real-time topic detection but are biased in 'outlining' all the tweets that appear in a given context (in our case the Jay-Leno Show). On Sentiment mining, uses managed learning to classify twitter posts on brands by using n-gram features. They also conclude that automated sentiment analysis accuracy is equal to manual classification. Use of emoticons in tweets for distant supervised learning in sentiment analysis is explored in this paper.

U. Waltinger. Polarity reinforcement: Sentiment polarity identification by means of social semantics. 2009.

Uses set of twitter-specific words to improve sentiment classification. We have not found any significant reference on mining user information in Twitter to develop profiles (though many have tried to identify influential users).

S. Asur and B.A. Huberman. Predicting the future with social media,2010.

Regarding applications of Twitter data-mining in media, we found rather fewer articles. There have been a few on trends and sentiment analysis on tweets during political debates on TV. Recently [5] talked about predicting box-office sales from mining twitter data. However, we have not found any current article on business intelligence using Twitter data in media. We explore this space through a case-study in our current paper.

Mining Social Media Data for Understanding Students' Learning Experiences Xin Chen, Student Member, IEEE, Mihaela Vorvoreanu, and Krishna Madhavan

This study explores the previously uninstrumented space on Twitter in order to understand engineering students' experiences, integrating both qualitative methods and large-scale data mining techniques. As an initial attempt to instrument the uncontrolled social media space, we propose many possible directions for future work for researchers who are interested in this area. We hope to see a proliferation of work in this particular area in the near upcoming future. We support that great consideration needs to be paid to protect students' confidentiality when trying to give good education and services to them.

Predicting User-Topic Opinions in Twitter with Social and Topical Context Fuji Ren, Senior Member, IEEE, and Ye Wu ,Ieee transactions on affective computing, vol. 4, no. 4, october-december 2013
With popular micro blogging services like Twitter, users are able to online share their real-time feelings in an easier way. The user set up information in Twitter is thus regarded as a resource providing individuals' spontaneous emotional info, and thus has attracted much consideration of researchers. Prior work has measured the emotional expressions in users' tweets and then did various analysis and learning. However, how to utilize those learned knowledge from the noticed tweets and the context info to predict users' opinions toward specific topics they had indirectly given is an unusual problem presenting both challenges and opportunities. In this paper, we mainly focus on resolving this issue with a Social context and Topical context incorporated Matrix Factorization (ScTcMF) framework. The preliminary conclusion on a actual Twitter data set show that this framework outperforms the state-of-the-art collective filtering process, and show that both social context and topical context are effective in improving the user-topic opinion prediction performance.
Social media is stage where people build content, share their opinions, views and ideas. Examples include Facebook, MySpace, Digg, Twitter and JISC list deliver on the academic side, because of its ease of use, speed and reach, social media is rapidly changing the public chat in community and setting trends and agendas in topics that range from the habitat and politics to technology and the entertainment corporation. Since social media can also be understood as a mode of unified foresight, it can be used to foresee real-world outcomes. "The Wisdom of Crowds" is about the aggregation of info in groups; appear in choice that is often better than could have been made by any single member of the group. There is a popular saying on social media that goes as follows: "We use Facebook to schedule the protests, Twitter to counterpart, and the YouTube to tell the world."
Twitter is the place where we all gather to express particular points of view and feelings about specific topics. Opinions reveal beliefs about specific matter commonly considered to be subjective. Twitter has millions of users that span millions of personal posts on a daily basis. And this gives us the opportunity to study social human subjectivity. Manual classification of thousands of posts for opinion mining task is unfeasible for a human being.

III. OBJECTIVE AND AIM

Develop system which uses, Social Media data using Text Mining methods for Decision making. Through a combination of automated and manual text-mining process we found out the opinion about product, the customers response to the product as well as some interesting profiles of the users. We believe that such information can aid companies in understanding its customer and better planning of future products.

IV. PROBLEM STATEMENT

The problem in sentiment analysis is segregating the conflict of a given text at the document, sentence, or feature/aspect level. Whether the expressed opinion in a doc, a sentence or a matter feature/aspect is positive, negative, or neutral.

V. SCOPE

Classify users within positive and negative association depends upon network density and degree of social activity either in information sharing or emotional synergy and social alertness. We consider that our findings will be useful in developing tools for forecasting positive and negative users and aid give the best recommendation towards helping negative users through online social media. We analyzed publicly available tweets that were posted when a popular TV show was aired for the first time. We use basic text-mining process to answer questions such as, what were the main themes during the show, what flock the demand of the show and what was the kind of viewers who were tweeting on the show. This type of BI can be used by content producers in fixing future shows, in targeted marketing to audience, in collecting competitive intelligence and many such other ways.

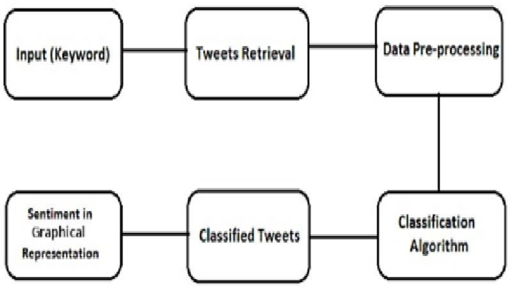
VI. ARCHITECTURE

A) Existing System
Social media is a platform where people create content, share their opinions, vision and concepts. Examples include Facebook, MySpace, Digg, Twitter and JISC list serves on the academic side, because of its simplicity, speed and reach, social media is rapidly changing the public chat in society and setting trends and agendas in case that limits from the environment and politics to technology and the entertainment industry. Since social media can also be presumed as a form of mutual wisdom, it can be used to predict real-world outcomes. "The Wisdom of Crowds" is about the gathering of info in groups, emerging in decisions that are often better than could have been made by any single member of the group. The popular saying on social media goes as follows: "We use Facebook to record the protests, Twitter to coordinate, and YouTube to tell the world."
Twitter is the place where we all gather to express precisely the point of view and feelings about specific topics. Opinions reveal beliefs about specific matter commonly considered to be subjective. Twitter has millions of users that spread millions of personal posts on a daily basis. And this gives us the opportunity to study social human subjectivity. Manual classification of thousands of posts for opinion mining task is unfeasible for a human being.

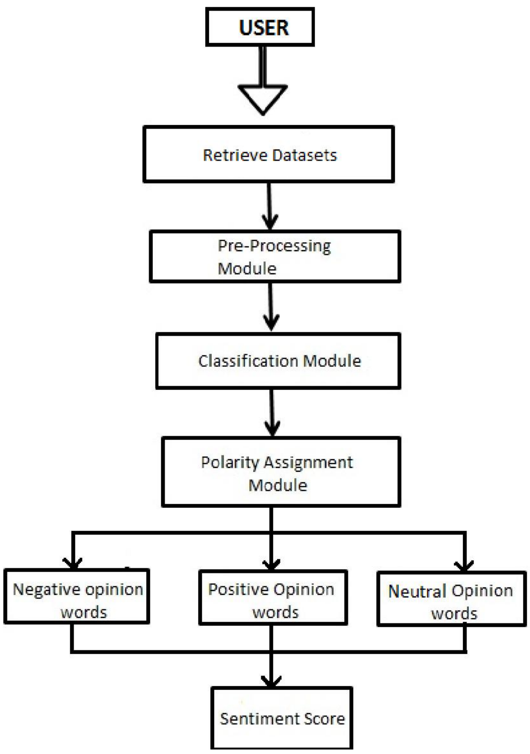
B) Proposed System
This paper reviews some basic terminologies, definitions and applications of Social Media data using Text Mining methods and Business Intelligence. Some of the applications of Text Mining methods in Social Media which influences Decision making is also collected and discussed in this paper. Case study can be considered from the twitter data and an analysis.

Through a combination of automated and manual text-mining techniques we found out the opinion about product, the customers response to the product as well as some interesting profiles of the users. We believe that such information can aid companies in understanding its customer and better planning of future products.

Block Diagram



C) Algorithm/Procedure/Process/Flowchart



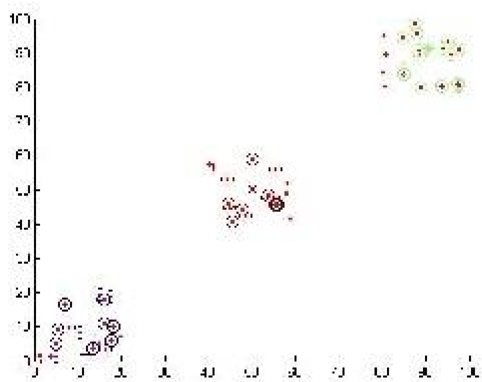
D) Methodology

- Tweet Downloader
  - Download the tweets using Twitter API
- Tokenisation
  - Twitter specific POS Tagger and tokenizer developed by ARK Social Media Search
- Pre-processing
  - Replacing Emoticons by their polarity, assign scores
  - Remove URL, Target Mentions
  - Replace #text -> text, since hash tags may contribute to the sentiment
  - Replace Sequence of Repeated Characters eg. 'cooooool' by 'cool' and assign higher score
  - Twitter specific stop word removal
  - Acronym expansion
- Feature Extractor

  Unigrams and Bigrams
  Polarity Score of the Tweet (f1)
  Count of Positive/Negative Words
  Maximum Positive/Negative Score for Words
  Clustering Positive/Negative words
  Count of Positive/Negative Emoticons and assign scores Positive/Negative special POS Tags Polarity Score

For Clustering we use QT clustering Algorithm
Quality Threshold (QT) clustering algorithm
This algorithm requires the apriori specification of the threshold distance within the cluster and the minimum number of elements in each cluster.
Now from each data point we find all its candidate data points. Candidate data points are those which are within the range of the threshold distance from the given data point. This way we find the candidate data points for all data point and choose the one with large number of candidate data points
to form cluster. Now data points which belongs to this cluster is removed and the same procedure is repeated with the reduced set of data points until no more cluster can be formed satisfying the minimum size criteria.

Algorithmic steps for QT clustering :
1) Initialize the threshold distance allowed for clusters and the minimum cluster size.
2) Build a candidate cluster for each data point by including the closest point, the next closest, and so on, until the distance of the cluster surpasses the threshold.
3) Save the candidate cluster with the most points as the first true cluster, and remove all points in the cluster from further consideration.
4) Repeat with the reduced set of points until no more cluster can be formed having the minimum cluster size.



Result of QT Clustering with threshold distance =10 & minimum number of element required within a cluster = 5

Advantages
1)Quality Guaranteed - Only clusters that pass a User-defined quality threshold will be returned.
2) Number of clusters is not specified apriori.
3) All possible clusters are considered - Candidate cluster is generated with respect to every data points and tested in order of size against quality criteria.

Disadvantages
1) Computationally Intensive and Time Consuming - Increasing the minimum cluster size or increasing the number of data points can greatly increase the computational time.
2) Threshold distance and minimum number of element in the cluster has to be defined apriori.

VII. MATHEMATICAL MODEL

Set theory:
Business intelligence (BI) systems combine operational data with analytical tools to present complex and competitive information to planners and decision makers, in order to improve the timeliness and quality of the decision-making process. Business intelligence systems provide actionable information delivered at the right time when decisions need to be made. Mining the existing data within social media applications is promising to produce the required
Information to meet the relevant objective. However, social media data are vast, noisy, unstructured, and dynamic in nature, and thus mining the data is not easy to do.
In this we are going through following phases.

- Tweet Downloader
- Tokenisation
- Pre-processing
- Feature Extractor

Above are the variables of the system According with it we are going to express dependencies among them and try to find out the scope and solution of our system
Mathematical model :
Here given is:
S=(T,P,Q,O)
A=T (t1,t2,t3…….tn)
Where t =tweets
Here we are download tweets of TV shows from twitter.
B=P (a1,a2,a3…..an)
Where P=Preprocessing
Here remove URL of tweet,#tag,Stopwords
C=Q (b1,b2,b3…..bn)
Apply QT clustering Algorithm
Mathematical Representation:
Tweet=t
Token = tk
Emotions= e
Positive emotions=pe
Negative emotions=ne
Score=sc
Neutral = nl
Neutral emotions = nle
Positive score=p
Negative score=n
User=u
Count =c
p= p +pe
n=ne + n
nl =nl + nle
as the user varying with time score of TV show varies
so
dsc=du/dt
dp=du/dt
dn=du/dt
dnl=du/dt
as per calculation we are convert it into graph for further use.

NP hard and NP Complete problems in business intelligence on tv media.
NP-hard (non-deterministic polynomial-time hard), in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP". A problem *H* is NP-hard if and only if there is an NP-complete problem L that is polynomial time Turing-reducible
to H (i.e., $L \leq_T H$). In other words, *L* can be solved in polynomial time by an oracle machine with an oracle for *H*. Informally, we can think of an algorithm that can call such an oracle machine as
a subroutine for solving *H*, and solves *L* in polynomial time, if the subroutine call takes only one step to compute. NP-hard problems may be of any type: decision problems, search problems, or optimization problems.

VIII. COMPARISON OF RESULTS LIMITATIONS

While the above analysis helped in getting information that is not readily available out of the Twitter world, there are restrictions to this intelligence.

A) Generalizing to TV audience from Tweeting audience:
All the study that has been given is for those people who were watching the show and tweeting. This part of people, probably, will not be model of the entire TV watching audience. The people who tweet are more likely to be more social network savvy and have access to the Internet while watching the show. Generalizing their comments/sentiments and profiles to the entire TV audience requires additional data and validation. This is a vital section were we need to work on. We were not able to do so due to the lack of time. Also we are unaware of any study that addresses this problem.

B) Data quality:
Few parts of Twitter have false data. Bios of users might be falsified. Names may be deliberately changed. Spam robots might create tweets on popular trending topics. All this hinders the data. While our assumption is that at a corporate level we can still trust the data we have not come across such event to prove the fallacy.

C) Automation:
Many components in our analysis have been manual (especially the study of sentiments) and is densely rest on ability of the key elements in the show and the context of the show. Evolving an automated or even a semi-automated accomplice to perform these tasks would be a more desirable out come. We are in the process of achieving this goal.

IX. CONCLUSION

This paper reviews some basic terminologies, definitions and applications of Social Media data using Text Mining methods and Business Intelligence. Some of the applications of Text Mining methods in Social Media which influences Decision making is also collected and discussed in this paper.

ACKNOWLEDGMENT

Authors are thankful to the Department of Information Technology, Sidhhant College Of Engineering, Pune, Maharashtra, India, for providing the necessary facilities for carrying out this work. Authors gratefully acknowledge the support given by Savitribai Phule university Pune, Maharashtra, India, for carrying out this research work. Authors also grateful to **Prof Nilesh Mali**, Sidhhant College Of Engineering, Pune, Maharashtra for their time to time guidance and support towards carrying out the research work. For influencing us to take forth this research concept and work towards its implementaion and also the anonymous reviewers for their valuable suggestions towards improving the paper.

REFERENCES

[1] M. Cataldi, L. Di Caro, and C. Schifanella. Emerging topic detection on twitter based on temporal and social terms evaluation. 2010.

[2] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11):2169–2188, 2009.

[3]U. Waltinger. Polarity reinforcement: Sentiment polarity identification by means of social semantics. 2009.

[4] S. Asur and B.A. Huberman. Predicting the future with social media,2010.

[5] Mining Social Media Data for Understanding Students' Learning Experiences Xin Chen, Student Member, IEEE, Mihaela Vorvoreanu, and Krishna Madhavan.

[6] Predicting User-Topic Opinions in Twitter with Social and Topical Context Fuji Ren, Senior Member, IEEE, and Ye Wu ,Ieee transactions on affective computing, vol. 4, no. 4, october-december 201

Mrs. Sapana Mahaveer Patil, ME- Information Technology, Savitribai Phule university Pune,Sidhhant College Of Engineering , Pune**.**