TREND ANALYSIS ON SOCIAL MEDIA USING NLP

Mohit Khatwani[1]

[1]*Student, Computer Science, MITCOE, Maharashtra, India*

Abstract

As we are heading towards Internet age resource demanding services are getting more and more popular. However size of these resources are growing rapidly, storage is becoming a problem. Natural Language Processing refers to ability of computer to process natural languages such as English, Marathi, Hindi, etc. Natural Language Processing is one of the most important stream in Artificial Intelligence,which helps computer understand an important part in which humans communicate. This paper focuses on trend analysis algorithms using Natural Language Processing. The result from this assessment shows how people think and are opinionated about different issues. For our result to be efficient these analysis techniques are applied among large masses. Nowadays we are surrounded by various platforms to give our opinion on something so it has become much easier to get the data.

Keywords: *NLP, Twitter, Trend Analysis, and Social Media.*

----------***----------

1. INTRODUCTION

Natural Language Processing is a domain in computer science which comes under Artificial Intelligence. Natural Language Processing is primarily focused on interaction between computer and human being. NLP involves Natural Language Understanding as well as Natural Language Processing.

Earlier it involved algorithms based on decision trees nowadays it is focused on statistical models which gives a weight to each input. Machine Learning algorithms are better than old algorithms in the sense that they can learn from the previous output produced and automatically focus on highly occurring cases. Algorithms can be made more efficient just by providing them more data.

Social media has evolved a lot in the past years. People can share ideas among each other on a click of a button. Twitter is one of the elite social media trends among people. Some of the topics get a lot of attention. Some of these information spread among a crowd of people and this phenomenon of spreading information is known as going 'viral'. It is necessary for a business to analyze these trends and further take necessary steps in improving the business. Trends in these social media is also very important among research community

2. BUILDING SOCIAL GRAPH

Social Graph consists for nodes which are nothing but users. Information about these users can be collected from Twitter with the help of their API (Application Programming Interface). Only restriction about creating this social graph is number of calls that can be made through this application programming interface. These calls return the necessary data about the node that is the users. One of the important parameter to consider for understanding social graph is a clique. A Clique in layman terms is group of people which belong to a specific community. In graph theory clique is minimum number of nodes which are connected to each other which form a complete graph.

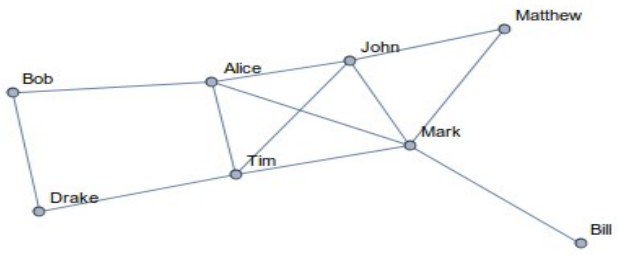

Fig -1: Social graph

3. CLASSIFYING TWEETS

Mainly tweets are classified into two types 1.) Informative tweets 2.) Chat tweets. Informative tweets consists of information which can be a promotional tweet. These type of tweets are important as they provide us with necessary information. They should be given higher priority among the other tweets. Chat tweets are tweets which are focused on a single person and they should not be considered while performing analysis as they are from a person with a single point of view and it may be wrong.

4. NLP TOOLS FOR TWITTER

There are a lot of NLP tools available but they are not applicable while doing trend analysis on twitter as they are specifically trained for performing analysis on news article like text. Twitter gave us a different type of challenge due its informal language used while tweeting and restriction of number of letters in a specific tweet.

It is also to be noted that this tools gives NER tags along with class of entities tagged such as location, person, date etc. Using number of distinct classes of NER in a tweet gives better result than just using number of NER's in a tweet.

Table shows how tweets are classified on the basis of their word count, language structure. Language structure helps in categorizing tweets. If the tweet consists of a image along with a 2-3 line text then it will be safe to assume that it can be a tweet from a company or it is pre-planned tweet as live tweet is something which cannot be sent with a picture.

Number of re-tweet count helps us identify how famous is the tweet and a lot of people support the tweet as their own point of view. Number of followers of the source of the tweet can help us identify whether the twitter user is a celebrity or not.

Nowadays most of the live events are announced on twitter. Event organizer gives a hash tag so that he can segregate all the tweets. This makes information gathering much easier. Location of the tweet can help a business man identify where he should think to increase or focus his business more.

Table -1: Classification for tweets

Tweet text based	Tweet meta data	NLP tags
Word count Hash Tags	Retweet count Number of Followers	NER (Name Entity Relationship) Location

5. NLP BEYOND TWITTER

Twitter, as general-purpose platform with an enormous user base, is a great source for tasks that require general population monitoring, such as influenza and behavioral risk factor monitoring. However, many health-focused on-line communities provide more detailed health information. With expansive texts, analyzing these communities requires NLP algorithms.

Facebook is one of the world's top most social networking websites. Facebook data can be used by analytics for searching trending topics among a specific target audience. Facebook is used by almost everyone. It is can also be used to gather day to day information in the form of news feed. We are going to use a web application known as "Quintly"[6] which is mostly used for analyzing Facebook data.

Quinlty gives the result in the form of different graphs which are helpful for analyzing the result accurately. In this report we have compared two celebrities and their on-line presence and fan base on Facebook.

Fig -2 shows total fans on Facebook , as we can see that growth of both celebrities is constant but total number of fan of Amitabh Bachchan are much more than that of ShahRukh Khan. From this result we can infer that Amitabh is much more active on Facebook than ShahRukh Khan.

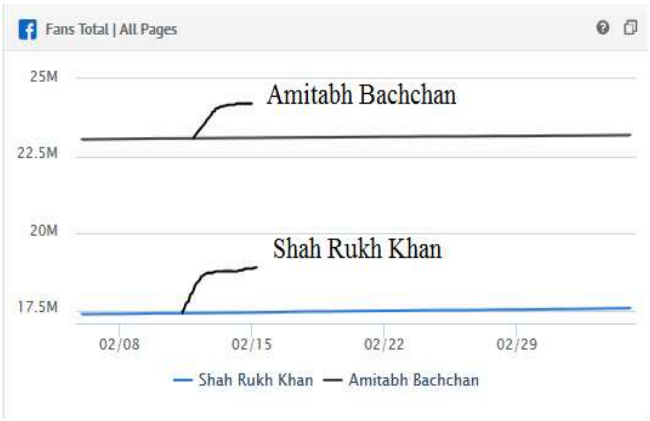

Fig -2: Total number of fans

It also provides with change of rate of fans following or disliking their profile,which is shown in Fig-3.

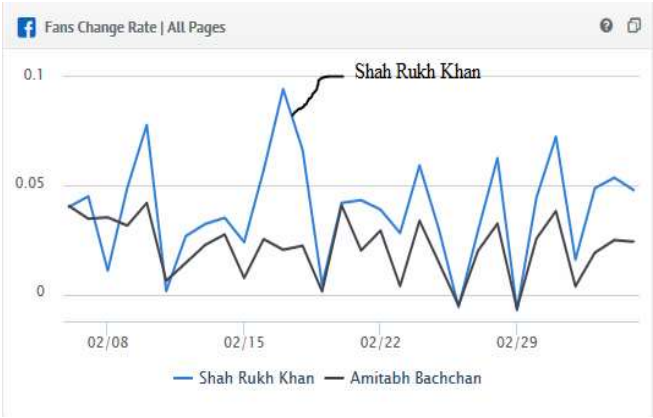

Fig -3: Rate of change of number of fans

From the analysis we can also know when the account holder was active so that we can have a rough idea about his /her preferences or timing he/she prefers for interacting with his fan base. If at a specific time account holder was not active then we can infer many things depending on the situation.

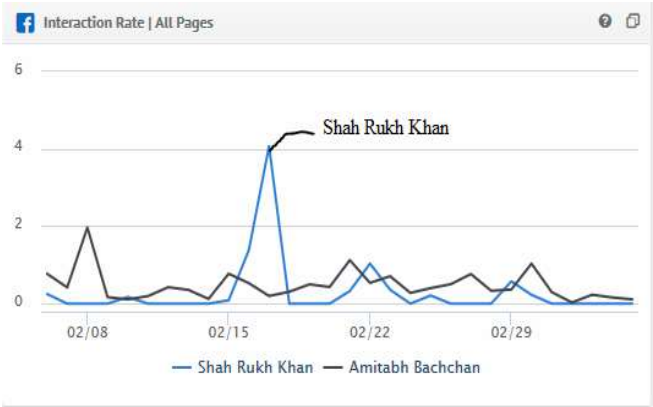

Fig -4: Interaction rate of account holder

There are various metrics related to this Trend analysis using Quinlty. These metrics can be used in various different ways. Diagram below shows graph of various metrics which are discussed further.

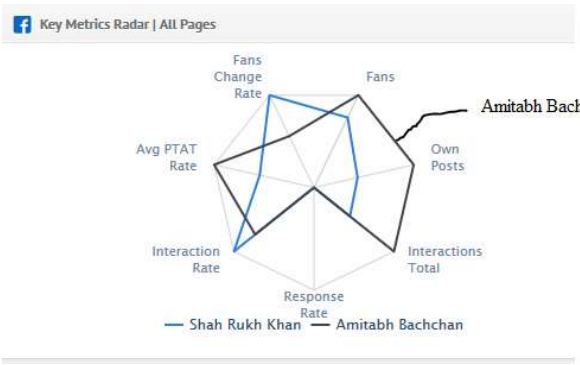

Fig -5: Metric Radar

In Fig - 5 there are various metrics discussed and their relationship among each other.

Interaction rate states that how the account holder often interacts with his fans. Own posts states whether his/her posts are original or not. Whether he/she often shares the posts from some other site or writes them themselves. Fans metric just compares the total value of their followers.

Apart from social media Trend analysis techniques can also be applied on Health related sectors which can be more useful for our society. Next chapter focuses on Discovering Health issues and trying to rectify them by using the knowledge of Natural Language Processing.

6. CONCLUSION

The research involved use of Natural Language Processing tools for an important aspect of Social Media.The paper involves use of websites like quintly.com for analyzing data and producing in a more graphical format. Social Media giants like Twitter and Facebook are more focused because they have most impact on an individual's views.

REFERENCES

[1]. Asad Mehmood “A Study of Sentiment and Trend Analysis Techniques for Social Media Content”, J.J. Modern Education and Computer Science ,February 2014.
[2]. Takeshi Sakaki “Tweet Trend Analysis in an Emergency Situation”, Tokyo Journal ,December 2014
[3]. Jinjian (James) Zhai, ”Sentiment analysis of news articles for financial signal prediction”,CS224N Final Report,2009
[4]. Ricardo Riberio,” Natural Language Processing for the Social Science”, ISCTE, January 2010.
[5]. Basics of developing android application on http://developer.android.com/training/basics/
[6]. Facebook data analysis on http://quintly.com/.

BIOGRAPHIES

Mohit Khatwani, Student of BE in Computer Science at MITCOE, Pune