International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 05, May 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

VISUAL DATA MINING IN SOCIAL MEDIA

Shilpy Gandharv[1], Mr. Vivek Richhariya[2], Dr. Vineet Richhariya[3]

[1]M.Tech(CSE), [2]Prof. Dept. of CSE, [3]Prof. & Head, Dept. of CSE

LNCT, Bhopal(M.P.)(India)

ABSTRACT

With the rapid growth of social media, the amount of customer feedback available to corporations, business owners, and IT services managers interested in obtaining customer input is greater than ever. The total volume of twitter comments was much larger than that of normal reviews on the web site. The tweets came in by the minute in contrast to normal web reviews that come in on a daily basis. The enormous size of the data stream, the diversity of the comments, and the uneven distribution of opinions over time make the analysis of twitter data very challenging. One of the main reason of using visualizing techniques is it is easily understandable the analysis result. Text analysis is still somewhat in its infancy, but is very promising.

Keywords—Social data, text mining, data mining, visualization.

I. INTRODUCTION

Over past 10 years, industries and organizations doesn't have requirement to store and perform operations and analytics on information of the shoppers. However around from 2005, the requirement to rework everything into information is way amused to satisfy the wants of the individuals. therefore huge information came into image within the real time business analysis of process information. From twentieth century ahead this World Wide Web has modified the means of expressing their views. gift scenario is totally they're expressing their thoughts through on-line blogs, discussion forms and additionally some on-line applications like Facebook, Twitter, etc [3]. If we have a tendency to take Twitter as our example nearly 1TB of text information is generating inside per week within the sort of tweets. So, by this it's perceive clearly however this web is ever-changing the means of living and elegance of individuals. Among these tweets will be classified by the hash worth tags that they're commenting and posting their tweets. So, currently several corporations and additionally the survey corporations area unit mistreatment this for performing some analytics[5] specified they will predict the success rate of their product or additionally they will show the various read from the information that they need collected for analysis. But, to calculate their views is extremely troublesome during a traditional means by taking these serious information that area unit about to generate day by day.

Text mining has become a preferred approach to analyzing and understanding massive datasets not done by the traditional analysis techniques. These tools are applied to a spread of data issues, like understanding themes in social media or facilitating info retrieval in unstructured information. Text mining may be a extremely great tool within the beginnings of analysis exploration, permitting the matter information to counsel themes and ideas to the research worker throughout analysis. this will give a helpful start line for framing additional analysis queries and analysis approaches, notably if hypotheses and further queries are not known (as is typical with an inductive analysis approach). moreover, these tools also can assist in improvement and structuring text-based information

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

for future analysis in mental image or different graphical tools. And, additionally to the tangible analysis advantages, text mining may be a fun and fruitful method of discovery!. Text Mining [4], is one amongst the foremost frequent nevertheless difficult exercise sweet-faced by beginners in information science / analytics consultants. the most important challenge is one has to completely assess the underlying patterns in text, that too manually. For example: it's pretty common to delete numbers from the text before we have a tendency to do any reasonably text mining. however what if we would like to extract one thing like "24/7". Hence, the text cleansing exercise is very customized as per the target of the exercise and therefore the kind of text patterns.

II. LITERATURE REVIEW

In [1] ,they present a system for the acquisition, analysis and visualisation of Twitter data. Twitter messages are harvested and stored in a distributed cluster, and the data is processed using algorithms implemented in a MapReduce framework. We present a clustering algorithm capable of identifying the main topics of interest in a tweet data set. Also, we designed a visualization method which allows to follow the intensity of twitter activity at a given geographical location. In this paper we have presented a system for the acquisition, analysis and visualisation of Twitter data. Twitter messages are harvested and stored in a distributed cluster, and he data is processed using algorithms implemented in a MapReduce framework. We presented a clustering algorithm capable of identifying hot topics of interest in a tweet data set. Also, we designed a visualization method which allows to follow the density of twitter activity in a given geographical location. The system is a prototype and was meant to present the potential use of a social media platform as source of large scale spatio-temporal information. It represents the building ground for future social media related applications targeting a multitude of possible applications with high social impact such as emergency situation management, risk and damage assessment and even social unrest.

In this paper they can visualize the twitter data using matlab and matlab is a traditional technique which can not handle bigdata, and twitter data generates huge amount of data per data which is not able to process by traditional tools and technique, due to which we need a powerful visualizing techniques which can work on bigdata directly.

In [2],Twitter, as a social media could be a very fashionable manner of expressing opinions and interacting with others within the on-line world. Once taken in aggregation tweets will give a mirrored image of public sentiment towards events. During this paper, we offer a positive or negative sentiment on Twitter posts employing a well-known machine learning methodology for text categorization. Additionally, we tend to use manually labeled (positive/negative) tweets to make a trained methodology to accomplish a task. The task is probing for a correlation between twitter sentiment and events that have occurred. The trained model relies on the Bayesian supply Regression (BLR) classification methodology. We tend to used external lexicons to notice subjective or objective tweets, other Unigram and written word options and used TF-IDF (Term Frequency-Inverse Document Frequency) to strain the options. Exploitation the FIFA journey 2014 as our case study, we tend to used Twitter Streaming API and a few of the official tourney hashtags to mine, filter and method tweets, so as to research

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

the reflection of public sentiment towards sudden events. An equivalent approach will be used as a basis for predicting future events. Twitter, one in all the foremost common on-line social media and micro-blogging services, could be a very fashionable methodology for expressing opinions and interacting with others within the on-line world. Twitter messages give real data within the format of short texts that categorical opinions, ideas and events captured within the moment. Tweets (Twitter posts) are well-suited sources of streaming information for opinion mining and sentiment polarity detection Opinions, evaluations, emotions and speculations usually mirror the states of individuals; they contains narrow-minded information expressed during a language composed of subjective expressions .During this paper, we tend to examine the effectiveness of a usually used text categorization methodology known as Bayesian supply Regression (BLR) Classification for providing positive or negative sentiment on tweets. We tend to use extracted Twitter sentiment to seem for correlations between this sentiment and major FIFA journey 2014 events as our case study.

In this paper the author calculate the polarity of the tweets with the help of Bayesian supply Regression (BLR) classification methodology and predict some events based on correlation between the events. But these methodology fails when data is very huge in terms of petabyte and also it cannot do real time analysis, for this we need a new tool and technique which can handle such huge and large datasets.

III. PROBLEM DEFINITION

Text mining [7] help an organization derive potentially valuable business insights from text-based content such as word documents, email and postings on social media streams like Facebook, Twitter and LinkedIn. Data mining or Text mining plays a important role in decision making because through these mining techniques we can analyse the data and on the basis of result we can take a decision. Now a days social media sites like twitter are widely used to share user opinions on various topics, twitter gives a platform to user to share their views and thoughts on various field like political, industrial, education and there is a petabytes of data generated by twitter in a day.

So the mining techniques are used to analysis the social twitter data thorough we get large amount of datasets to analysis, so the analysis of twitter data provides a better way for making decision.

IV. PROPOSED WORK

The work involved in this usually requires several computational techniques (such as data and text mining, natural language processing, etc.) and complex analytical processes required to manipulate varied data sources. Besides that, reach a point of balance between the computational side of the process and the aesthetic side using tables, charts, colours and other visual features, could favour a good analysis and quicker understanding of such data. Many researchers have been involved with the study of the evolution of these techniques. In several situations, e.g., a simple line or bar chart were not good enough to translate the complexity of the data to a general audience and, therefore, our primary goal is to understand how visualization techniques can help media

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

studies and journalism professionals and students to better understand user behaviour (and also user sentiment) in social networks.

V. PROPOSED METHODOLOGY

Our Steps or Algorithm Steps will follow:

1. First we get a complex social data and stored.
2. After retrieving we transformed the text, tweets are first converted to a data frame and then to a corpus. After that, the corpus needs a couple of transformations, including changing letters to lower case, removing punctuations/numbers and removing stop words.
3. In many cases, words need to be stemmed to retrieve their radicals. For instance, "example" and "examples" are both stemmed to "exampl". However, after that, one may want to complete the stems [6] to their original forms, so that the words would look "normal".
4. After transforming and stemming process is done then we build a document term matrix. Based on the matrix, many data mining tasks can be done, for example, clustering, classification and association analysis.
5. With the help of matrix we can identify the frequent words and their association between words.
6. After building a document-term matrix, we can now visualize the outputs.

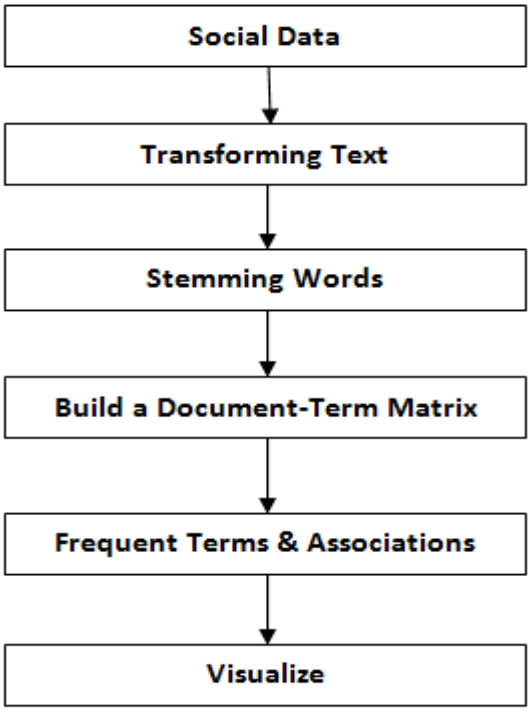

Figure 2. Analysis Steps

VI. CONCLUSION

Twitter data is very useful in decision making because its provide a variety of opinions on various topics. so the texting mining will perform on twitter data and we are using a visualizing techniques. Through which we are

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

perform preprocess on these data, and analyse the unstructured data comes from twitter sites and than we can visualize the analysis result because visualization result is more easy to understand.

REFERENCES

[1] Andrei Sechelea , Tien Do Huu, Evangelos Zimos, and Nikos Deligiannis, "Twitter Data Clustering and Visualization", in 2016 23rd International Conference on Telecommunications (ICT), 2016 IEEE.

[2] Mr. Peiman Barnaghi and John G. Breslin , ", Opinion Mining and sentiment polarity on Twitter and correlation between Events & Sentiment", International Conference on Big Data Computing and Application , IEEE 2016.

[3] Judith Sherin Tilsha S, Shobha M.S.," A Survey on Twitter Data Analysis Techniques to Extract Public Opinion.", IJARCSE , Vol. 5 , Issue 11 , Nov 2015 , 2277128X.

[4] Lokmanyathilak Govindan Sankar Selvan," A Framework for Fast-Feedback Opinion Mining on Twitter Data Streams", IEEE 2015.

[5] T. K. Das , D.P. Acharjya & M. R. Patra, " Opinion Mining about a product by Analyzing Public Tweets in Twitter ", ICCCI- 2014, Jan 03-05, 2014.

[6] Porter M.F, Snowball: A language for stemming algorithms. 2001.

[7] Ning Zhong, Yuefeng Li, and Sheng-Tang Wu, "Effective Pattern Discovery for Text Mining", *IEEE Transactions on Knowledge And Data Engineering*, Vol. 24, No.1, January 2012.