Intensified Multidimensional Style for User Belief Mining from Social Media

Amit Singla[1] Dr. Vishal Goar[2] Prof. S.S. Sarangdevot[3]

[1] Research Scholar, J.R.N. Rajasthan Vidyapeeth, Udaipur (INDIA)
[2]Assistant Professor, Govt. Engineering College, Bikaner (INDIA)
[3]VC, J.R.N. Rajasthan Vidyapeeth, Udaipur (INDIA)

amitsingla70@yahoo.com, dr.vishalgoar@gmail.com, prof.sarangdevot@gmail.com

Abstract. Big data analytics is used to examine large sets of data which may contain diversity of different types of data, it can be used to decrypt cryptic symbiology, correlating previously not known variables, finding the trends in the market, checking preferences of customers and finding out data about various businesses and institutions. The result can be used to conduct informative market strategizing, checking out chances to generate higher income, to provide effective consumer-oriented services, to improve effectiveness of operations and to provide competition-edge over competitors and other institutional profits. The main aim of Big data Analysis is to aid in better and informative decision making for the firms by taking advantage of capable data-scientists, genius model makers as well as other trained scientists to verify chunks of information that may be unused by the traditional programs. It may include analyzing special log and internet based information, internet network data and social-analysis of reports. It can also be used to analyze network records, caller details and other information gathered and operated by IOT devices. It can be used to bind with big data and unstructured data as well as partially structured data.

Keywords: Big Data Analytics, Emotions Mining, Social Media Analytics, User Belief Mining

1 Introduction

Big data analysis is done by using specialized software's for over-the-top level analysis and specialized fields such as statistical-analysis, data-mining, text- analysis and predictive analysis. Modern tools such as Business Intelligence as well as Visualization is majorly used in big data analytics. Also for the regularly updating data as well as continuous data update poses a problem in the analysis process- i.e. data analytics in case of mobile apps, and real time data in airlines as well as petroleum pipes.

Following is the statistics of Twitter used by assorted media and global locations. The following statistical data depicts the Twitter users of USA in year 2016.

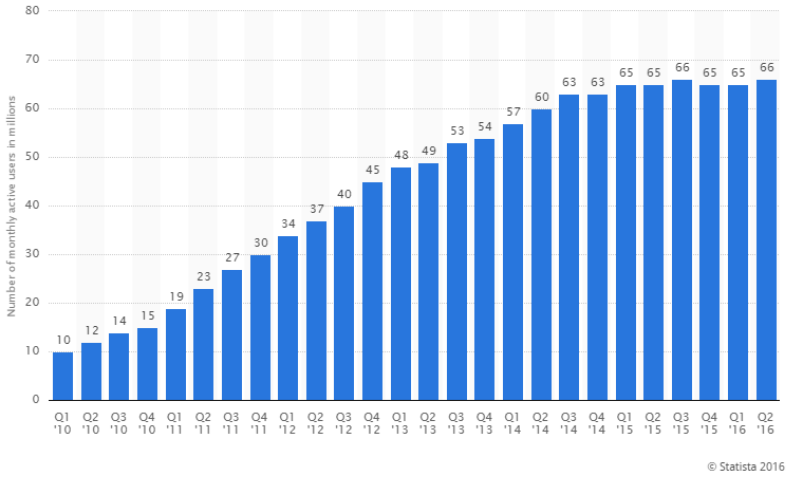

Fig. 1. Statistical Description of Twitter Users in India (2006)

The above depicted graph represents the escalating values of Twitter Users in USA and that is the key base of big data which is part of this research work in which the real time data analytics is done.The following statistical data and report depicts the forecasting and predictive analysis of Twitter users.

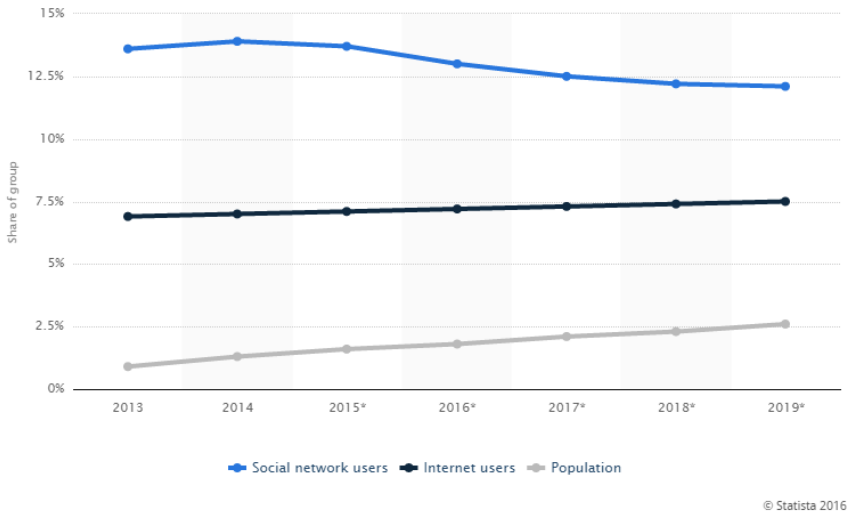

Fig. 2. Statistical Description of Twitter Users in India (2012-2019)

2 Review of Prominent Literature

Significant amount of research work is done in the segment of opinion mining and sentiment extraction from assorted social media platforms. Using this approach, the actual analytics from the live data can be done and this approach is having the effectual analytics on the social media users in terms of multiple parameters. Even the analytics from Internet of Things (IoT) based environment can be done using the sensor based technology.

As literature review is one of the key tasks that is performed for defend and justify the research work, following are the excerpts from assorted research manuscripts working on same direction as well as perspectives.

Bifet (2009) worked on the approach towards Machine Learning and Prognostic Analytics This work process the social media analytics dataset from Multiple Social Media Platforms.

Asur (2010) worked on the approach towards Sentiment Status Score. This research work is having key focus on the twitter analytics with the key aim towards user belief mining from hashtags and other attributes of Twitter.

Tan (2011)worked on the approach towards Recommender Engine. The work is having key focus on the attributes of Twitter.

Saif (2012)worked on the approach towards Popularity Score and Predictions. The work is having a unique mathematical foundation towards the feature points extraction from social media for prediction and sentiment mining.

Leong (2012) worked on the approach towards Opinion Scoring This work process the social media analytics dataset from SMS.

Hung (2013) worked on the approach towards Classification and Predictive Analysis. This work process the social media analytics dataset from Twitter

Dong (2013) worked on the approach towards Multilayered Model with Deep Learning. This work process the social media analytics dataset from Twitter

Rao (2014)worked on the approach towards Multilayered Model Based Approach. This work process the social media analytics dataset.

Lawrence (2014) worked on the approach towards Sentiment Mining This work process the social media analytics dataset from Models of Car

Poria (2015) worked on the approach towards Opinion Scoring This work process the social media analytics dataset from Twitter

Kotwal (2016) worked on the approach towards Sentiment Mining on HADOOP.. This work process the social media analytics dataset from Facebook

Davis (2016) worked on the approach towards Rule Mining This work process the social media analytics dataset from Multiple Social Media Platforms. This research work is having key focus on the twitter analytics with the key aim towards user belief mining from hashtags and other attributes of Twitter.

3 Flow of Fetching Process from Social Media

- Fetching live data feeds from social media
- Data Cleaning or Refinement
- Identification of Feature Points
- Extraction of the Mandatory Aspects
- Implementation of the Algorithm for Predictive Analytics
- Investigation of Positive and Negative Tweets
- Analysis of Popularity Score or simply Opinion Mining
- Detailed Analytic Report

This research work is focused on the predictive analysis and association of user accounts with their timeline. There are multiple phases of proposed work and implementation by which the extraction of live social media and respective attributes are interrelated for the association and relationship. Using this approach, the profile of any user can be directly fetched from multiple social media and then the probability or association aspects with other applications can be made.

In this research work, the sentiment data analysis and prediction is done on the big data fetched using Python Scripts. Using prediction tools and algorithms, the effective and accurate results shall be measured.

Proposed Flow of Work

Train Multinomial is a function that is used for training the dataset. First of all, we take the dictionary or repository of words. These words are sent to variable V. N is the total tweets or docs. Then every tweet is analyzed till its total count. In every iteration there is the concatenation of text with the score so that final cumulative score can be measured. Probability is measured so that the total occurrences can be analyzed on particular type of sentiment. In second phase, the further extraction of token is done with the calculation of overall score.

TRAINMULTINOMIALNB($\mathbb{C}, \mathbb{D}$)
1 $V \leftarrow$ EXTRACTVOCABULARY($\mathbb{D}$)
2 $N \leftarrow$ COUNTDOCS($\mathbb{D}$)
3 **for each** $c \in \mathbb{C}$
4 **do** $N_c \leftarrow$ COUNTDOCSINCLASS($\mathbb{D}, c$)
5 $prior[c] \leftarrow N_c / N$
6 $text_c \leftarrow$ CONCATENATETEXTOFALLDOCSINCLASS($\mathbb{D}, c$)
7 **for each** $t \in V$
8 **do** $T_{ct} \leftarrow$ COUNTTOKENSOFTERM($text_c, t$)
9 **for each** $t \in V$
10 **do** $condprob[t][c] \leftarrow \frac{T_{ct}+1}{\sum_{t'}(T_{ct'}+1)}$
11 **return** $V, prior, condprob$

APPLYMULTINOMIALNB($\mathbb{C}, V, prior, condprob, d$)
1 $W \leftarrow$ EXTRACTTOKENSFROMDOC(V, d)
2 **for each** $c \in \mathbb{C}$
3 **do** $score[c] \leftarrow \log prior[c]$
4 **for each** $t \in W$
5 **do** $score[c] \mathrel{+}= \log condprob[t][c]$
6 **return** $\arg\max_{c \in \mathbb{C}} score[c]$

Results and Discussion

Table 1: Database structure for Implementation

Column	**Type**
Id	int(11)
Username	varchar(255)
Twittertimestamp	varchar(255)
Twittertext	Text
Screenname	varchar(255)
Followers	varchar(255)
Friends	varchar(255)
Listed	varchar(255)

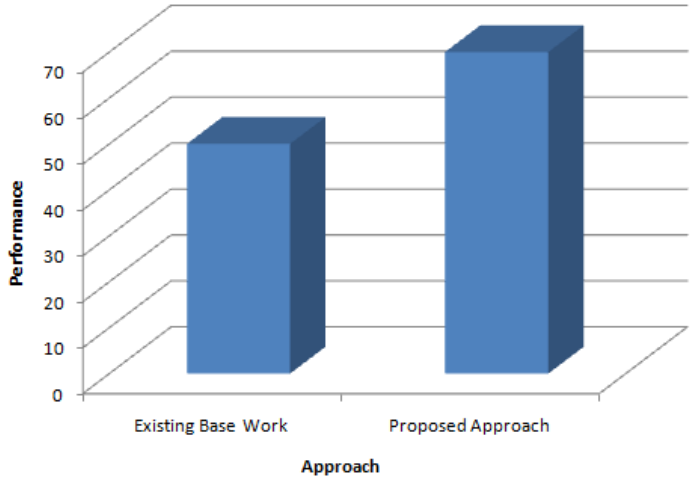

Figure 3 – Classical and Proposed approach (Performance in Percentage)

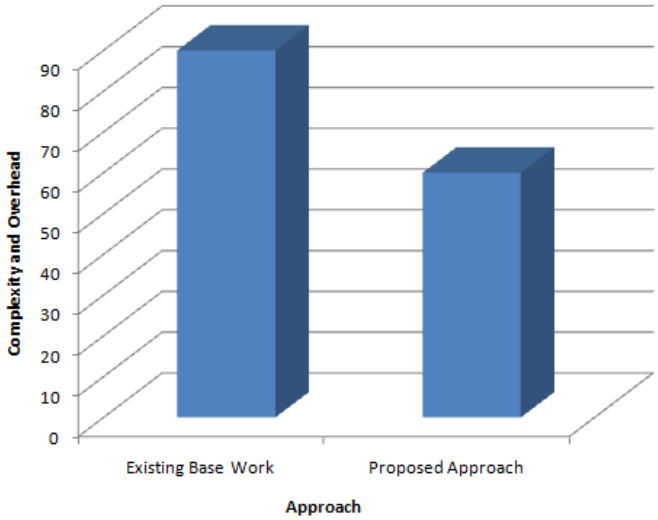

Figure 4 – Comparison of Classical and Proposed approach (Complexity in Points)

Table 2: Comparison of Existing Base Work and Proposed Approach

Existing Base Work (Overall Effectiveness)	Proposed Approach (Overall Effectiveness)

50	60
60	88
70	89
50	69

Table 2 underlines the comparative results in terms of effectiveness. The effectiveness in taken and logged in terms of percentage and its evident from the results that the results of proposed approach is better than the classical approach.

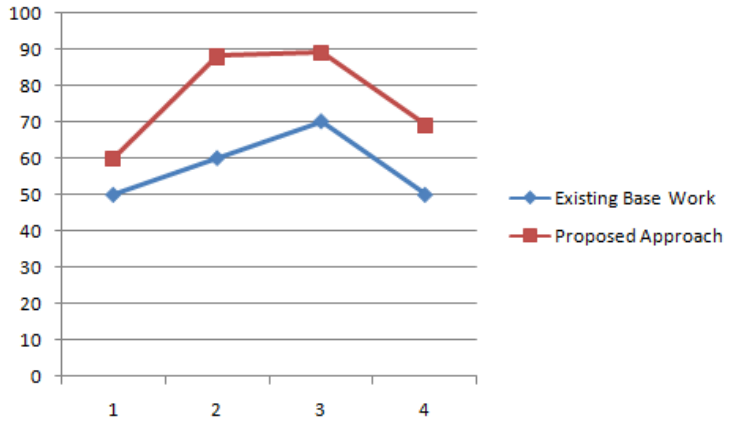

Figure 5 – Effective Comparison of Classical and Proposed Algorithm

It is evident from the above mentioned figures and graphical results that the proposed approach is hugely effectual in terms of assorted parameters including performance, effectiveness and other related dimensions.

Conclusion

Fetching the live social media or related dimension sentiment analysis is under research from a long time for detailed analysis and prediction of the events with respect to the social cause. In this research work, the live extraction of time-line from social media platforms are implemented so that a common as well shared dataset can be prepared for future login and predictive analysis of the user behavior. In this work the key focus rely on the fetching of Twitter Timelines with the usage of SDK and API for research and development and real time dataset can be evaluated for predictive analysis. There are number of optimization approaches using which the efficiency, accuracy and performance factors can be improved. The integration of soft computing approaches are prevalent in the research community which provides fuzzy based execution and global optimization from existing results.

References

1. Awrahman, B., & Alatas, B. (2017). Sentiment Analysis and Opinion Mining within Social Networks using Konstanz Information Miner. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(1), 15-22.
2. Hürlimann, M., Davis, B., Cortis, K., Freitas, A., Handschuh, S., & Fernández, S. (2016, September). A Twitter Sentiment Gold Standard for the Brexit Referendum. In Proceedings of the 12th International Conference on Semantic Systems (pp. 193-196). ACM.
3. Kotwal, A., Fulari, P., Jadhav, D., & Kad, R. (2016). Improvement in Sentiment Analysis of Twitter Data Using Hadoop. Imperial Journal of Interdisciplinary Research, 2(7).
4. Cambria, E., Fu, J., Bisio, F., & Poria, S. (2015, January). AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In AAAI (pp. 508-514).
5. Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in Twitter. Natural Language Engineering, 20(01), 1-28.
6. Abdul-Mageed, M., Diab, M., & Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. Computer Speech & Language, 28(1), 20-37.
7. Cambria, E., Schuller, B., Liu, B., Wang, H., & Havasi, C. (2013). Knowledge-based approaches to concept-level sentiment analysis. IEEE Intelligent Systems, 28(2), 12-14.
8. Shahheidari, S., Dong, H., & Daud, M. N. R. B. (2013, July). Twitter sentiment mining: A multi domain analysis. In Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on (pp. 144-149). IEEE.
9. Saif, H., He, Y., & Alani, H. (2012, November). Semantic sentiment analysis of twitter. In International Semantic Web Conference (pp. 508-524). Springer Berlin Heidelberg.
10. Leong, C. K., Lee, Y. H., & Mak, W. K. (2012). Mining sentiments in SMS texts for teaching evaluation. Expert Systems with Applications, 39(3), 2584-2589.
11. Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011, August). User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1397-1405). ACM.
12. Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492-499). IEEE.
13. Bifet, A., & Frank, E. (2010, October). Sentiment knowledge discovery in twitter streaming data. In International Conference on Discovery Science (pp. 1-15). Springer Berlin Heidelberg.
14. Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, 11, 450-453.
15. Awrahman, B., & Alatas, B. (2017). Sentiment Analysis and Opinion Mining within Social Networks using Konstanz Information Miner. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(1), 15-22.
16. Zhao, J., & Gui, X. (2017). Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis. IEEE