Deep Learning Based User Belief Extraction from Social Media and Predictive Analytics

Yatin Kumar Manocha
M.Tech. Research Scholar
Department of Computer Science and Engineering
Universal Group of Institutes
Lalru, Punjab, India

Priyanka
Head of Department, Assistant Professor
Department of Computer Science and Engineering
Universal Group of Institutes
Lalru, Punjab, India

Abstract

Social Media Analytics is quite prominent now days in assorted factors with the association of user belief mining in which the real time evaluation of popularity of any movie, celebrity, person or any other real world object can identified. This research work is having key focus on the dynamic analytics of the multiple keywords from Twitter Social Media using Java Platform. With this particular platform, the real time analytics from social media and related popularity with the negative and positive sentiments can be extracted and by this way, the overall performance and emotions about that person or object can be identified. This research work extracts the real time tweets from Twitter Social Medial using Twitter4J that is Java based API and further analytics can be done effectually. In association with these technologies, the Stanford Libraries for natural language processing is integrated so that the sentiments can be extracted with accurate performance.

Keywords: User Belief Mining, User Sentiment Mining, Sentiment Extraction

INTRODUCTION

Bigdata 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 bigdata 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. Thus, organizations and firms famous for their data-analytics proficiency have switched to modern and more innovative technologies in the filed for e.g., Hive, NoSQL, Hadoop Tools, YARN etc. which can be used to process clustered information which are of big and sundry nature with the help of their open-source nature [2].

Kalev Leetaru of Wired recently wrote, "For its flagship new reality show Opposite Worlds the Syfy channel wanted to let the audience 'remote control' the show via social media. The task worked with Syfy to create what ultimately became its real-time 'Twitter Popularity Index.' The Index combines the intensity of conversation around each character, the number of unique discussants, and the emotion of that discussion using a new sentiment engine powered by over 1.6 million words, phrases and common misspellings and colloquial expressions. Using our Index, Opposite Worlds records across the board in Twitter engagement for a cable television series."

Leetaru continues, "Sentiment mining is a hot emerging field, yet the underlying technology has changed little from the first computerized sentiment mining system created in 1961, the General Inquirer. It still treats emotion measurement as merely a technical problem. This has yielded a stream of pioneering technical achievements that have focused on algorithms rather than the actual outcome of how to better measure tone online. To build the first sentiment engine that could actually understand real-time tweets, we had to start from scratch, asking the question: how can big data combine with human insight to change the way we interact with our world? In the process, we identified 16 limitations to current sentiment mining approaches."

Big data analytics is used to examine huge chunks of data and can be used to decrypt cipher texts, correlating previously not known variables, finding the trends in the market, checking preferences of customers and finding out data about various businesses and institutions. Data professionals and perform analytic operations on the large amount of data which is unconquerable by conventional operations and methodology. Using conventional methodology here is a futile effort and thus Big-Data Analytics proves essential as well as effective solution. Big-Data Analytics can reduce the wait-time of results which comprises of GBs and TBs of information.

Modern tools such as mining, prediction, text-analytics and relevant data analysis result in conclusive results and valid and informative decision making process. It also aids in suiting modern requirements in innovative fashion [3].

LITERATURE SURVEY

The extraction, analysis and learning of earlier work done on the related domain are very important aspect for any research work. To propose and defend the new and effective approach, there is need to present to the excerpts from assorted research papers with the deep learning of journals, conference proceedings and seminars. In this section, the extracts of earlier research work from assorted sources are mentioned so that the proposed work can be under investigation effectively.

Bifet (2009) [1] – In this research work, the deep learning algorithms and machine intelligence approaches are presented with the statistical analysis at the base level. Using the implementation of assorted statistical methods, there is specific predictive analysis in association with assorted data mining approaches. The work in this paper is presenting the effectual results with the analytical results with integration of assorted datasets in investigation.

Bollen (2009) [2] - In this article, we play out an assumption investigation of every single open tweet communicated by Twitter clients between August 1 and December 20, 08'. This work finds that occasions in the social, political, social and financial circle do have a significant, prompt on the different measurements of open disposition. The

creators estimate that vast scale investigations of disposition can give a strong stage to model aggregate emotive patterns as far as their prescient esteem with respect to existing social and also financial indicators.

Bollen (2010)[3] - It is observed on analysis of Twitter handles that it may sometimes be linked to the DJIA value and shifts accordingly with it. In this study, feed-analysis (Twitter) is done by two effective software-solutions-GPOMS (Google Profile of mood states)- It checks and analyzes the mood of collective opinion by its six dimensional analyses (Surety,Kindness, Vitalness, Calmness, Alertness and Happiness). Opinion Finder - It analyzes whether the mood is negative or positive. In this work, a Mood-time variable is cross validated and analyzed against a sample case of prediction of a president-level election on Thanksgiving in 08'. Fuzzy networks and a granger causality analytics is done for investigation of collective lever mood indication, i.e. by GPOMS and Opinion-finder tool which is the collective result of prediction by dow jones industrial average. In this study, we also signify that with the inclusion of collective emotional spectrum variables the verity of results can be increased by quite a margin. It is also concluded that by significant margin (6 percent) MAP error is decreased and the results were accurate by 87.6 percent in the prediction of regular up-down shifts in closing variables in dow jones average.

Asur (2010) [4] - The domain of sentiment scoring, recommendations and prediction is touched in this research manuscript by forms the foundations for development and deployment of a recommender system. In this research manuscript, the dataset from Twitter is fetched and then the recommendations about opinions or popularity of movies are implemented. The research manuscript presents the results in effectual aspects with the cavernous graphical representations having recommendations and predictions.

Tan (2011) [5] – The research work by authors in this manuscript is focus towards the extraction and predictive analysis of hashtags and profile tags of Twitter so that the effective recommender engine can be proposed. In this work, the authors focus on the assorted aspects by which the Twitter profiles, timelines and related information can be fetched

out from social media platforms. The results in this paper depict the prognostic investigation of user profiles and timelines using assorted programming models.

Saif (2012) [6] – This research manuscript is focused and diverted towards the market predictions and forecasting with the extraction of live dataset from social media. Using the live dataset fetching from Twitter and other social media profiles, the research work can be used to present the effective market survey and then product based prediction in terms of their related popularity, scoring and user sentiments.

Leong (2012) [7] - As smartphone dataset can be used for the extrapolative investigation on assorted aspects, this research work is having the SMS dataset and further analyzing the sentiments from these extracted messages. Using this approach, the classification and opinion scoring can be presented with effectual aspects.

Wang (2013) [8] –This research work is having key focus on the development and deployment of a new and effective classification model having analysis of popularity dataset simply known as word of mouth dataset. Using this approach, the words associated with positive and negative sentiments can be classified and further predicted

Dong (2013) [9] – The work in this paper is implemented in three different but integrated layers which performs the predictive analysis from user tweets fetched from Twitter social media platform. The first phase in this work fetches the real time streaming dataset from Twitter for training of the upcoming layers. The second layer in this work develops and implements the effective model with Naïve Bayes for classification of sentiments. The third layer or phase of this work is having focus on the testing and predictive presentation of results.

Cambria (2014) [10] – In this research work, the authors present their work with two effective and novel models. The initial model of this research work is more diverted towards the extraction of live microblogs and tweets then training in association with the supervised model so that predictions can be done. This work is done with the approach similar to artificial neural networks. The second model of this research work present the predictive

results in forms of classification of emotions, tweet sentiments and microblog opinions.

The work begins with the experimentation done on the fetched tweets according to the categories. In addition thorough analysis of different tweets is done using deep mining and association of tweets and tokens. Further the proposed technique is being compared with the existing techniques.

Google Refine or Open Refine is the free and open source tool used for parsing or transformation of JSON to the understandable and formatted aspects so that the data mining and machine learning algorithm can be implemented. Earlier version was Google Refine which was further transformed to Open Refine under the URL openrefine.org so that the research community including scientists, academics and research scholars can use it.
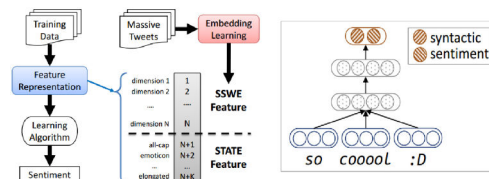

Figure 1. Sentiment Extraction [22]

The above mentioned figure represents the flow of sentiment classification. The process of sentiment classification solely depends on the bag of words in which positive and negative words are processed under weight and then finally checked so that the overall scoring of tweet regarding particular celebrity, person, entity or organization can be done.
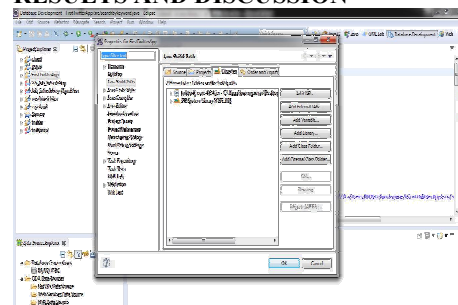
RESULTS AND DISCUSSION

Figure 2. Eclipse IDE Setup and Configuration with Twitter4J

☐ Uncompress and copy to any location
☐ In Eclipse, Create a New Project
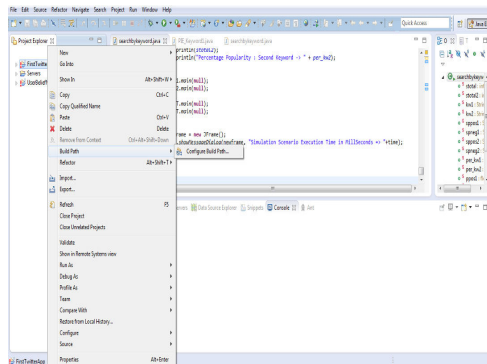☐ Right Click the Project Name, Select Build Path -> Configure Build Path

Figure 3. Build Path Settings for JAR files

GoTo Libraries, click Add External JAR and give the path of all JAR files in lib

Figure 4. Libraries in Twitter4J

Figure 5. WAMP Server for MySQL Database
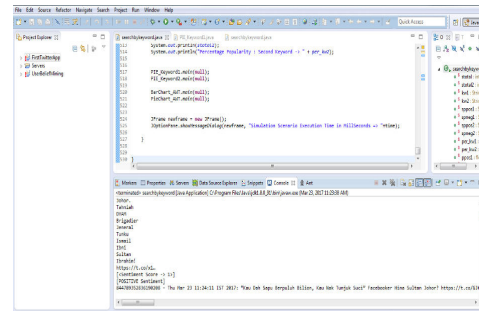
JDBC Twitter Integration Results

Figure 6. Setup of Live Tweets

SQL Result
Host: localhost
Database: twitter
Generated by: phpMyAdmin 3.4.5 / MySQL 5.5.16-log
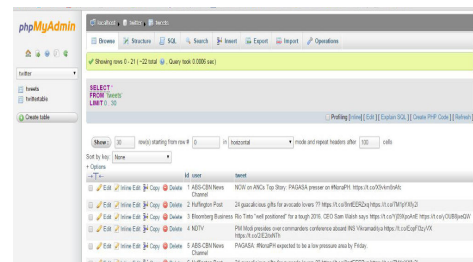**SQL query:** SELECT * FROM `tweets` LIMIT 0, 30 ;
Rows: 22

Figure 7. MySQL database with the fetched tweet
In Figure, there is phpMyAdmin which is a free and open source tool written intended to handle the administration of MySQL or MariaDB with the use of a web browser. It can perform various tasks such as creating, modifying or deleting databases, tables, fields or rows; executing SQL statements; or managing users and permissions. In our work, the live tweets are getting stored in phpMyAdmin.
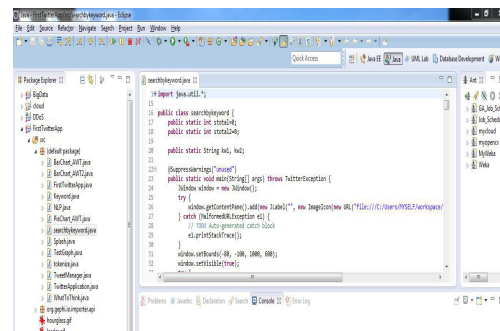
Figure 8. Searching by Keywords from Twitter and Java

This figure describes how the Java Based GUI can search Tweets from Twitter by using different keywords. It can search for Movies, Politics, Education, Celebrities, World and Others. It search for the selected keyword.
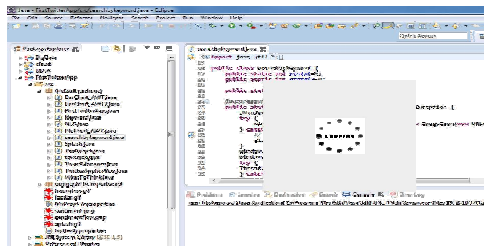

Figure 9. Splash Loader in Java

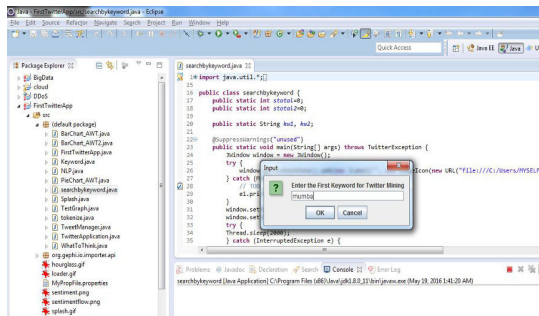It defines the Splash loader that moves during processing.

Figure 10. Searching Panel

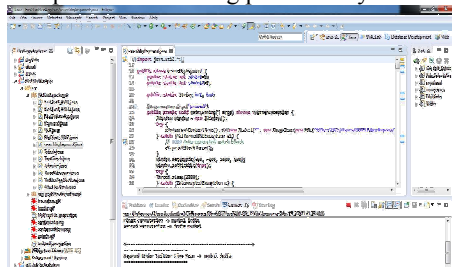It depicts the searching panel of keywords.

Figure 11. Generation of tweets of Search Strings

Abovementioned figure describes the generation of various tweets on console window search by different strings. The sentiments of tweets also found during fetching tweets on this window.
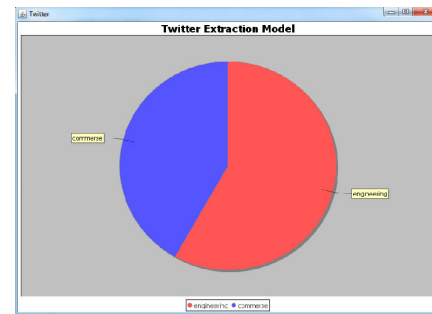

Figure 12. Result Analysis

Conclusion

Sentiment analysis, also referred to as Opinion Mining, implies extracting opinions, emotions and sentiments in text. One of the most common applications of sentiment analysis is to track attitudes and feelings on the web, especially for tacking products, services, brands or even people. The main idea is to determine whether they are viewed positively or negatively by the viewers or users on the social media. Twitter is a popular micro blogging service where users create status messages (called "tweets"). These tweets sometimes express opinions about different topics. The work build an automatic sentiment (positive or neutral or negative) extractor from a tweet. This is very useful because it allows feedback to be aggregated without manual intervention. Using this analyzer, Consumers can use sentiment analysis to research products or services before making a purchase. Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. Organizations can also use this to gather critical feedback about problems in newly released products. 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 timeline 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.

References

[1] 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.

[2] Bollen, J., Mao, H., & Pepe, A. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, 11, 450-453.

[3] Bollen, J., Mao, H., & Pepe, A. (2010). Determining the Public Mood State by Analysis of Microblogging Posts. In ALIFE (pp. 667-668).

[4] 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.

[5] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011). 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.

[6] Saif, H., He, Y., & Alani, H. (2012, November). Semantic sentiment analysis of twitter. In International Semantic Web Conference (pp. 508-524). Springer Berlin Heidelberg.

[7] 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.

[8] Wang, H., Cambria, E., Schuller, B., Liu, B., & Havasi, C. (2013). Knowledge-based approaches to concept-level sentiment analysis. IEEE Intelligent Systems, 28(2), 12-14.

[9] Dong, H., Shahheidari, S., & 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.

[10] Cambria, E., Fu, J., Bisio, F., & Poria, S. (2014). AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In AAAI (pp. 508-514).

[11] Kotwal, A., Fulari, P., Jadhav, D., & Kad, R. (2014). Improvement in Sentiment Analysis of Twitter Data Using Hadoop. Imperial Journal of Interdisciplinary Research, 2(7).

[12] Poria, S. Cambria, E., Fu, J., Bisio, F., & (2015). AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In AAAI (pp. 508-514).

[13] Guo, Y., Rao, J., Cheng, D., & Zhou, X. (2016). ishuffle: Improving hadoop performance with shuffle-on-write. IEEE Transactions on Parallel and Distributed Systems.

[14] Davis, B., Hürlimann, M., 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.

[15] 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.

[16] Abdul-Mageed, M., Diab, M., & Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. Computer Speech & Language, 28(1), 20-37.

[17] Hassan, S., Yulan, H., Miriam, F., & Harith A.(2015, June). Contextual Semantics for Sentiment Analysis of Twitter. Elsevier.