



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Innovative Sentiment Analysis for Smart Devices and Digital Technologies

Dr.P.B.Edwin Prabhakar¹, Dr.M.Sindhuja², G.Saritha³, Sneha V⁴, Basheer Ahmed⁵

Department of Computer Science and Engineering, New Prince Shri Bhavani College of Engineering and Technology,
Chennai, India¹

Department of Information Technology, New Prince Shri Bhavani College of Engineering and Technology,
Chennai, India^{2,4}

Department of Electronics and Communication Engineering, Sri Sairam Institute of Technology, Chennai, India³
Rotek Electronics, Chennai, India⁵

ABSTRACT: As years are passing individuals are getting increasingly more pulled in towards Social Media Sites and they had begun joining there existence with these electronic devices. As this things are going on the data being made and shared by a large number of individuals that is expanding each second and these data are difficult to store and make due. As we realize that everything in this world makes them mean full presence in this world in that comparable manner this big data can likewise be utilized for some application and for that the analysis done is called Sentiment Analysis. In any case, investigating this huge dataset isn't so natural and same way Sentiment Analysis is missing for Negation Statement case. This issue is lessening the exactness of Sentiment Analysis which can be confirmed effectively with the proposed philosophy in this paper.

KEYWORDS: Big Data, Sentiment Analysis, Digital Technologies and Negation.

I. INTRODUCTION

Big Data is the most arising word during this timeframe and it is the most significant theme to be worried for. Web-based Media has assumed control over an exceptionally huge space in human way of life to that end Big data is additionally turned out to be extremely big issue to deal with and use it simply [4]. What's more, this data from online media is vital and can be utilized for different investigations like Sentiment analysis. With the development of computerized Technologies and shrewd gadgets, an enormous measure of advanced data is being created consistently. This data is exceptionally difficult to oversee utilizing customary procedures like Relational Database Management System [7], these framework can't ready to store this colossal data and difficult to investigate this data fastly and precisely. Because of the disappointment of traditional procedure of capacity of data for big Data another framework is designed named as Hadoop, Hadoop is an open source system which has capacity to store the data and furthermore to dissect the data precisely, exactly and fastly moreover. The Apache Hadoop programming library is a structure that considers the circulated processing of enormous data sets across groups of PCs utilizing straightforward programming models. It is essentially intended to keep huge number of machine through a solitary server. Hadoop has different apparatuses which upholds various languages for its processing, model, HDFS, Hive, Map-Reduce, Pig, Oozie, flume and so on [13]. This paper will primarily have to reach out to Flume, Hive, HDFS, Pig. Flume upholds for the assortment of dataset from the engineers site of Twitter and hive principally utilizes HSQL language which is for the data dealing with as in tables, sees and so on HDFS is basically for the capacity reason. What's more, in this task Pig is utilized for making a client characterized work for various method. As different investigations have been led in this space however the precision of sentiment analysis isn't sufficient till now. Also, if there should arise an occurrence of Big Data the exactness isn't great in light of the fact that to deal with enormous measure of data is definitely not a simple undertaking. Our principle design isn't just to store huge measure of data yet additionally to break down it to serve associations and furthermore for individuals' government assistance.

II. OVERVIEW ON CLOUD COMPUTING

The crude Data that should be handled for assessment mining or sentiment analysis can be of two kinds:

1. Structure Data - Structured data is data, generally text documents, showed in named segments and columns which can without much of a stretch be arranged and handled by data mining apparatuses. This could be envisioned as an impeccably coordinated file organizer where everything is recognized, named and simple to get to.

2. Unstructured Data -Unstructured data is what has no recognizable inside structure. It very well may be imagined as a room with chaotic articles .Some instances of unstructured data are E-sends, PDF documents, Digital Images, Video, Audio and so on For This sort of Unstructured Data [5] the Traditional Database Management System won't work, RDBMS can't ready to deal with huge and unstructured Data for that reason it need to have other Big Data taking care of framework which store and interaction it in quicker speed.

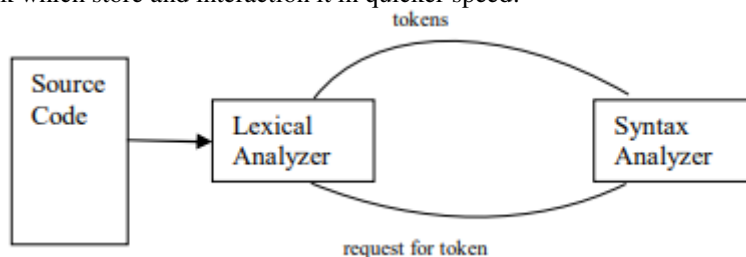


Figure.1 Lexical Analysis

2.1 Lexical Analysis

Sentiment Analysis can be broadly partitioned into different techniques like Lexical analysis Lexical analysis essentially utilize the analysis from the reference of any distinct corpus. This corpus contains English words with its significance and the grammatical features given for individual words. On account of sentiment analysis, a basic dictionary based approach is order the sentiments of tweets in view of number of 'positive' and 'negative' terms contained in the tweets and pick the name with the most contained terms.[12]. The flowchart for this is displayed in figure 1. However, lexical Analysis likewise comes up short now and again to that end we can't take just lexical analysis for our review, for further developed exactness and accuracy it ought to have followed another strategy or way to conquer the disservice of the lexical Analysis.

2.2 Machine Learning Algorithm

Machine learning algorithm principally partitioned into two sections for example Directed Approach and Unsupervised Approach. There are different Machine Learning algorithm till date which can be utilized and looked at which can give the better exactness. Yet, as nothing is amazing Machine Learning algorithm likewise not give the ideal exactness and accuracy for instance this machine learning algorithm [12] needs preparing and test data set in enormous sum to get wonderful precision. That is the reason by reading up both the viewpoint for this paper the blend of Lexical Analysis and Machine Learning Algorithm both have been considered to beat the impediment of nullification issue from lexical analysis Machine learning Algorithm suits the best and for giving the enormous measure of preparing dataset Lexical Analysis give it.

2.3 Natural Language Processing

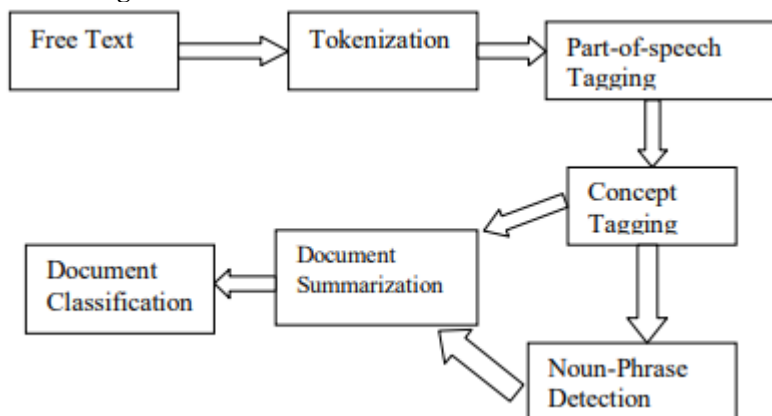


Figure.2 Process of NLP

Natural language processing (NLP) is the capacity of a PC program to comprehend human discourse as it is spoken. NLP [9] is a part of man-made consciousness (AI). Today NLP is an essential apparatus since human needs to impart to the PCs and PC see just machine language for that Naturally changing over processing is required. Current methodologies of Natural Language Processing is using Machine Learning Algorithm. NLP can mostly fills different roles (displayed in Figure 2) like:

- Tokenization
- Part of Speech Tagging
- Parsing
- Named Entity Recognition
- Chunking

There are various NLP toolkits available like:

- Stanford NLP
- Natural Language Toolkit (NLTK)
- Apache OpenNLP
- Apache Lucene & Solr

For this paper the Apache OpenNLP have been viewed as better which is open source with the utilization of Maximum Entropy Algorithm as the Machine Learning Algorithm.

2.3 Maximum Entropy Algorithm

Maximum Entropy Algorithm is a Machine Learning Algorithm .The Max Entropy classifier is a probabilistic classifier which has a place with the class of remarkable models. Our objective is to utilize the logical data of the report (unigrams, bigrams, different attributes inside the text) to sort it to a given class (positive/impartial/negative, objective/emotional and so forth) Maximum Entropy primary Principle is Higher the entropy higher is the consistency. The maximum entropy guideline depends on choosing the most uniform dispersion which is to be known by the one having maximum entropy.

III. METHODOLOGY

For working on the exactness of Sentiment analysis the framework in which these model must be carried out ought to have RAM of limit least of 8GB and for Big Data we require a structure called Hadoop in the System. With the utilization of Flume we can gather the dataset which is to be handled. As the technique followed is Lexical Analysis first it need some corpus of English words which will analyze the sentence and check with the corpus to be made. After the assortment of dataset from twitter it can additionally continue to finish the analysis and the result will be as Positive, Negative and Neutral cases .In this paper three cases have been accepted that is of Positive, Negative and Neutral as it were. In this Model every one of the means of data pre processing is done in Hive instrument and after that it should make a User Defined Function which is written in Java utilizing Pig device of hadoop. As displayed in Figure 4.1, it addresses the general techniques to be followed during this course of Sentiment Analysis. In above system for tackling the nullification expulsion issue it is not difficult to compose a User Defined Function in Java Language which joins the negative intensifier word with its adjoining descriptor positive word and make its extremity to - 1 for example negative. Furthermore, as examined prior about Machine Learning Algorithm and out of different machine learning Techniques Maximum Entropy Algorithm is taken up, and applied with the foundation of OpenNLP tool compartment. As Natural Language Processing toolbox is the instrument which is equipped for Understanding the Naturally utilizing Human language to machine till some degree.

IV. RESULTS AND DISCUSSION

Through this work, a productive algorithm for Sentiment Analysis has been checked and ended up being proficient. This study demonstrates that Sentiment Analysis improves exactness and Precision by utilizing the Lexical and Machine Learning Algorithm together. Figure 5,1 displayed underneath give the result of the Maximum Entropy Algorithm utilizing OpenNLP device. This Maximum Entropy Algorithm Log-probability data predicts the graphical result based on its cycles which is displayed in beneath figure 3.

```

.. done.
Computing model parameters...
Performing 100 iterations.
1: .. loglikelihood=-2169688.899571239      0.0
2: .. loglikelihood=-2040859.7804328476     0.7494770108406036
3: .. loglikelihood=-1940176.6629931435     0.7561765115568784
4: .. loglikelihood=-1859007.6408229321     0.7617041587382319
5: .. loglikelihood=-1791755.667663587     0.7662062743364075
6: .. loglikelihood=-1734750.9520797003     0.7697997881207022
7: .. loglikelihood=-1685528.3119008588     0.7728501854902297
8: .. loglikelihood=-1642380.4030299599     0.7754188066518343
9: .. loglikelihood=-1604088.1371733875     0.7774660360552624
10: .. loglikelihood=-1569756.204198629     0.7790966632205397
11: .. loglikelihood=-1538710.337909146     0.7805062100569825
12: .. loglikelihood=-1510431.375628575     0.7817988271241283
13: .. loglikelihood=-1484511.7540273038     0.7827936886038841
14: .. loglikelihood=-1460626.042972584     0.7836703423933671
15: .. loglikelihood=-1438510.4936582488     0.7844818222130779
16: .. loglikelihood=-1417948.5217154683     0.7852549644035112
17: .. loglikelihood=-1398760.1910357862     0.7858926469704967
18: .. loglikelihood=-1380794.4540989576     0.7865443866682172
19: .. loglikelihood=-1363923.3302048233     0.7870613057029779
20: .. loglikelihood=-1348037.4714930023     0.7874759910596648
21: .. loglikelihood=-1333042.7396964652     0.7878683127992732
22: .. loglikelihood=-1318857.5304275514     0.7882452994871703
23: .. loglikelihood=-1305410.6581469618     0.7886529562784896
24: .. loglikelihood=-1292639.6670965056     0.789060613069809

```

Figure.3 Output of Maximum Entropy Algorithm

V. CONCLUSIONS

This proposed work is appropriate for English Language as it were. In future it tends to be tried for other Language data Set likewise and it very well may be more exact and exact by seeing each line and sentence in any language like as a human being understanding. And furthermore tackle various issues like Increment - decrement explanations and furthermore address the emoticon cases to be broke down. One further more work is Possible to speed up the working of entire analysis. In this paper data is taken from Twitter just yet we can take data from any of the online media destinations. This task is fundamentally taking care of the big Data which should be taken care of better with association of equal circulated frameworks and associated with a solitary server this should be possible in future work.

REFERENCES

1. Ramesh R, Divya G , Divya D , Merin K Kurian , Vishnuprabha V, " Big Data Sentiment Analysis using Hadoop" in Volume I April 2015, IJRST, ISSN(online):2349-6010.
2. Andrea Esuli and Fabrizio Sebastiani , "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining
3. Apoorv Agrawal , Jasneet Singh Sabarwal , "End to End Sentiment Analysis of Twitter data " Proceeding of the workshop on Information Extraction & Entity Analysis on Social media data, COLING 2012, Mumbai , December 2012.
4. Efhymois Kouloumpis , Theresa Wilson , Johanan Moore , "Twitter Sentiment Analysis: The Good the Bad and the OMG! "fifth International AAAI Conference on Weblogs and Social media.
5. Milagros Fernandez-Gavilanes , Tavera Alvarez-Lopez , Jonathan Juncal-Martinez , Enrique Costa-Montenegro , Francisco Javier Gonzalez Castano , "An Unsupervised Approach for Sentiment Analysis in Twitter", Proceedings of the International Workshop on Semantic Evaluation ,june 4-5 2015.
6. Manisha Shinde-Pawar , "Formation of smart Sentiment analysis Technique for Big Data "Vol.2 ,Issue 12 ,December 2014IJIRCE.
7. M.Vasuki , J.Arthi, K. Kayalvizhi , "Decision Making using Sentiment Analysis from twitter" Vol.2 ,Issue 12 ,December 2014 , IJIRCE(an ISO 3297:2007 Certified organization).
8. Chetan Kaushik and Atul Mishra , "A Scalable , Lexicon Based Technique for Sentiment Analysis" Vol.4 ,No. 5 September 2014, IJFCST.
9. Sunil B.Mane , yashwant Sawant , Saif Kazi , Vaibhav Shinde , "Real Time Sentiment Analysis of Twitter Data Using Hadoop " Vol 5(3) , 2014 , IJCSIT.
10. Mahalaxhmi R. , Suseela S, "Big-SoSA: Social sentiment Analysis and Data Visulaization on Big Data " Vol 4, Issue 4, April 2015 , IJARCCE
11. Penchalaiah.C , Murali.G , Suresh Babu.A , "Effective Sentiment Analysis on Twitter Data Using: Apache Flume and Hive" IJISET , Vol. I Issue 8, October 2014.



12. Christopher Johnson, Parul Shukla, Shipla Shukla, "On classifying the Political Sentiment of Tweets"
13. Tom White, "Hadoop: The Definitive Guide " Text Book.
14. Shlomo Argamon-Engelson and Moshe Koppel and Galit Avneri , " Style-based Text Categorization: What Newspaper Am I Reading? " AAAI Technical Report WS-98-05. Compilation copyright © 1998, AAAI (www.aaai.org).
15. Adam L.Berger, Stephen A.Della Pietra and Vincent J.Deila Pietra 1996. "A maximum Entropy approach to natural Language processing " Computational Linguistic.
16. Deepali Arora, Kin Fun Li and Stephen W. Neville, " Deepali Arora, Kin Fun Li and Stephen W. Neville" , 2015 IEEE 29th International Conference on Advanced Information Networking and Applications.
17. Vinh Ngoc Khuc, Chaitanya Shivade, Rajiv Ramnath, Jay Ramanathan," Towards Building Large-Scale Distributed Systems for Twitter Sentiment Analysis".
18. John Dodd, "Twitter Sentiment Analysis",National College of Ireland, Final Thesis Report.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details