

# A Trace Oriented Analysis of Social Network Application Based Sentiments Based on Polarity Consistency

Mrs.P.Bhavani <sup>1</sup>, J.Anandhabalaji <sup>2</sup>, M.Arjun <sup>2</sup>, V.Harish <sup>2</sup>

<sup>1</sup>Assistant Professor, Computer Science and Engineering, Manakula Vinayagar Institute Of Technology, Puducherry

<sup>2</sup>Student, Computer Science and Engineering, Manakula Vinayagar Institute Of Technology, Puducherry

\*\*\*\*\*

**Abstract:** Nowadays, online discourse group community social media is online conversation where people enhance create content, share it, save it, and network in a distributed file system for Hadoop technologies at an amazing rate. The faster idea and ease of use in team group community social media today is Tweets content posts details material. The important points on Tweets content posts details content include opinions and opinions on certain topics such as sports, film, book, product, state suggestions, and so on. Based on this condition, this studies to use the details of twitter to review a world cup cricket series by using perspective discovering or feeling analysis. Viewpoint discovering symbolizes the application of natural language managing, computational linguistics, and written text discovering to recognize or classify whether the who will win or lost the arrange based on idea perspective. Trace Oriented Analysis (TOA) is supervised learning methods that assess data and recognize the styles that are used for classification. This analysis issues on binary classification which is classified into two classes. Those classes are positive and negative with some emotions too. The beneficial classification shows excellent idea opinion; otherwise the negative classification shows the bad idea perspective of certain matches. This approval is based on the level of high top, high quality of TOA with the approval process uses Fold cross validation approval and confusion matrix. The Polarity and the HDFS balance is used to enhance the selection of best parameter in order to fix the dual marketing problem.

**Keywords:** Trace oriented analysis, Polarity, HDFS balance, Hadoop technologies, Supervised learning method

## INTRODUCTION

Nowadays social network play an important role in most of the people's daily life. We have taken twitter as an example for social network. In this people try to express

their feelings like happy, sad, sorrow etc. in the social network by updating their status, posting their comments in the internet. In this we found that nearly 8 tera bytes of data has been stored by twitter every day. In near the nearby future it may increase even more. So handling of that much data in the database is not easy thing. And the retrieving of data from its huge set of database is difficult and it may take even more time. And there is some failure, that is the data which we retrieved is correct or not.

To avoid this problem we have implemented the trace oriented analysis (TOA) algorithm. In this the polarity consistency method which means finding whether the sentence is positive or negative. This method helps the user to identify the sentence easily. And we also define the neutral comments that are intermediate between positive and negative comments. By splitting the sentence in TOA, the mining of particular database can be done easily. During these we have analyzed that the people were expressing their sorrow and happiness by using some abused words in their comments. This may affect the feeling of other users who views this comments. To avoid this, we are having a set of abused words that the people used normally in the social network. By analyzing that particular word will be removed. The next thing we have concentrated is analyzing the stopping and stemming words. The stopping words will be like is, was, were, then etc. and stemming words will be like ing, ies, ied, ent etc. By removing these words we can reduce some memory space and also accessing of data is faster in this process.

The objectives of this venture to research the people's reviews in a large amount of information for verifying

and estimate the results under the sentimental analysis for beneficial, adverse and other psychological emotions by decoding natural terminology with the help of TOA Procedure in Big Data.

## 2. EXISTING SYSTEM

The existing system of this project is polarity classification of terms is essential for programs such as Viewpoint Exploration and Feeling Research. A variety of sentiment word/sense dictionaries have been personally or (semi)automatically designed. We observe that these sentiment dictionaries have several discrepancies. Besides apparent circumstances, where the same term seems to be with different polarities in different dictionaries, the dictionaries display complicated situations of polarity inconsistency, which cannot be recognized by simple guide examination. We present the idea of polarity reliability of words/senses in sentiment dictionaries in this document. We display that the reliability issue is NP-complete. We decrease the polarity reliability issue to the satisfiability issue and implement two quick SAT solvers to identify variance in a sentiment vocabulary. We execute tests on five sentiment dictionaries and WordNet to demonstrate inter- and intra-dictionaries variance.

The problem of these applications are geared towards analyzing pupils feedback about the consistency for domain checking independent and therefore skewed towards sentiment analysis that detects positive/negative feelings by interpreting natural language.

### 2.1 DISADVANTAGES

Sentimental Emotions however comes at a significant performance cost. On smart gadgets where data, like the disgust, bad, sad etc are very limited, it is important to keep a low footprint on such solutions.

## 3. PROPOSED SYSTEM

The proposed system of this project is the file system for Hadoop technology at an efficient manner. The quicker idea and ease of use in team group community social media today is Twitter posts content details. Everything

on Twitter posts content information include opinions and opinions on certain topics such as sports, movie, book, product, state suggestions, and so on. Based on this condition, this studies to use everything of twitter to review a world cup cricket series by using perspective discovering or feeling analysis. Viewpoint discovering symbolizes the application of natural language handling, computational linguistics, and published written text discovering to recognize or classify that who will win or lost the arrange based on idea perspective. Tracking oriented Analysis (TOA) is supervised learning methods that evaluate data and recognize the styles that are used for classification. This analysis issues on binary classification which is classified into two classes. Those classes are good and bad with some feelings too. The beneficial classification shows outstanding idea opinion; otherwise the negative classification shows the bad idea perspective of certain suits. This acceptance is based on the level of outstanding top, top outstanding top outstanding top great top high quality of TOA with the acceptance process uses 5-Fold cross validation approval and confusion matrix. The Polarity and the HDFS balance is used to enhance the selection of best parameter in order to fix the double marketing problem.

The methods used in the proposed system are the step by step process. First the preprocessing of data is performed. Data pre-processing is an important step in the data mining process. The data gathering methods are often loosely controlled and analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis. Then we classify the data based on the feature. They are stopping words, stemming words. The elimination of these sopping and stemming words can reduce the more memory space occupied by it in the database. This method also increases the performance by increasing the speed of accessing the data and also the accuracy is maintained high by this process.

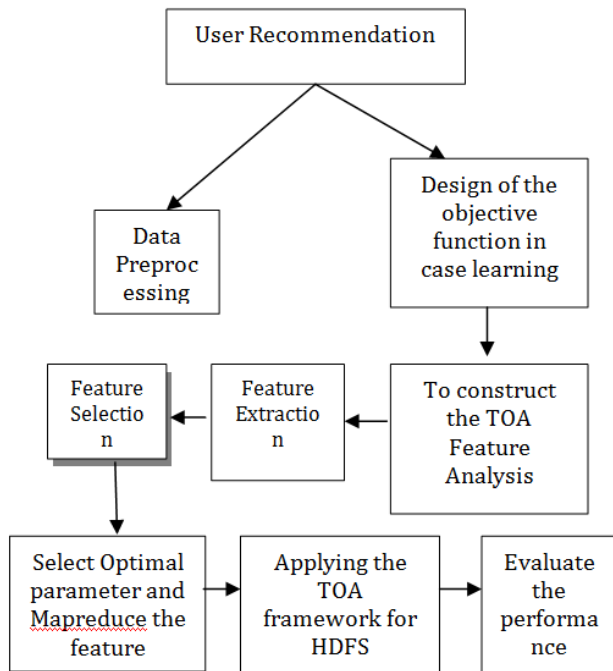


Figure.2 Architecture for proposed system

Stemming is the term used for information retrieval to describe the process for reducing inflected (or sometimes derived) words to their word stem, base or root form generally a written word form. The stem need not be identical to the morphological root of the word, it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

Stop words are words which are filtered out before or after processing of natural language data. Though the stop words usually refer to the most common words in a language, there is no single universal list of stop words used by all processing of natural language big data tools (Hadoop), which indeed not all tools even use such a list. Some tools specifically avoid removing these stop words to support phrase search. Any group of words can be chosen as the stop words for a given purpose.

Then the data cleaning is performed. Data cleaning is a technique that is applied to remove the noisy data and correct the inconsistencies in data. Data cleaning involves transformations to correct the wrong data. Data cleaning is performed as a data preprocessing step while preparing the data for a data warehouse. Sentimental analysis is the process to frame by uploading the positive and negative words in the sentimental dictionary with the help of word net dictionary. In detail, it is a data preprocessing technique that merges the data from multiple

heterogeneous data sources into a coherent data store. Data integration may involve inconsistent data and therefore needs data cleaning.

Polarity based sentimental analysis is an automated sentiment analysis in which the application of text analytics techniques for the identification of subjective opinions in text data. It normally involves the classification of text into categories such as Positive and Negative feedbacks. We tremendously increase in demand for polarity consistent for sentimental tools of hadoop network will monitor the feedbacks of the people's opinions of the world cup scenario to fulfill their demands increasing more and more research contents to perform sentiment analysis for claiming to perform any domain for big data tools working is so difficult in different meanings of words are associated with distinct emotions depending on the domain in which they are being used.

### 3.1 ADVANTAGES

- To achieve better performance, this project proposes to optimize the Trace oriented Algorithm which is easily recognize their users feedback with a strong validation.
- On analyzing the performance for persistent storage protection using TOA on smart gadget devices.

### 4. CONCLUSION

The proposed system recommends that a new technique for sentiment analysis using real twitter data. We created three classification frameworks. With the advents of polarity of words and its consistency were maintained that depicts the sentimental classification in a streamlined manner. Thus the analysis of big data is enhanced. This large volume of data is processed efficiently in Hadoop environment. The analyzing of the data in huge volume with less time complexity proves its efficiency. The Polarity and the HDFS balance are used to enhance the selection of best parameter in order to fix the dual marketing problem. The result shows the improvement of level of top high quality from 71.87% to 77%.

### 5. ACKNOWLEDGEMENT

We would like to express our gratitude to all teaching and non-teaching staff members of our department. And above all, the blessings of the Almighty has kept up our spirit and enabled us to complete our study successfully.

## REFERENCES

- [1] A. Cui, M. Zhang, Y. Liu, S. Ma, Emotion Tokens: Bridging the Gap among Multilingual Twitter Sentiment Analysis, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 238–249.
- [2] A. Bifet, E. Frank, Sentiment Knowledge Discovery in Twitter Streaming Data, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 1–15.
- [3] A. Bifet, G. Holmes, B. Pfahringer, MOA-Tweet Reader: real-time analysis in twitter streaming data, in: T. Elomaa, J. Hollm'en, H. Mannila (Eds.), DS 2011, LNCS 6926, Springer-Verlag, Berlin Heidelberg, 2011, pp. 46–60.
- [4] <http://www.edupristine.com/courses/big-data-hadoop-program/?jscfct=1>
- [5]<http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>
- [6] <http://efytimes.com/e1/fullnews.asp?edid=123367> Tom White, Hadoop The Definitive Guide, 1st edition, 2009.
- [7] Pete Warden, Big Data Glossary, International edition, 2011.
- [8] Eric Siegel and Thomas H. Davenport, Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Pearson edition, 2011.
- [9] Willyan D. Abilhoa, Leandro N. de Castro, A keyword extraction method from twitter messages represented as graphs, 2014.
- [10] Luca Cagliero, Tania Cerquitelli, Paolo Garza, Luigi Grimaudo, Twitter data analysis by means of Strong Flipping Generalized Itemsets, 2014.