Abstract

We examine the sentiment analysis of social media by collecting and analyzing textual data. Sentiment analysis is an efficient and effective evaluation of people’s opinions in real time. It allows data collection and analysis from a very large sample without hindrances, obstructions and time delays. In this paper we the analyze opinion about youngsters by the public keeping in view with the recent Chennai Flood occurred, their participation and contribution towards it using Hadoop environment.

Keywords: Sentiment analysis, HDFS, Flume, R Tool.

1. Introduction

Nowadays micro blogging websites such as Twitter, Facebook,Yahoo have become so famous, and people try to express their views on a particular topic on these blogs. So here we analyze the sentiment of the people expressed in social media form to determine the opinion about youth’s participation in Chennai flood. This paper will consist of creating a sentiment analysis system using big data concepts like Hadoop, Hive, and flume. The algorithm used is k means clustering and Naive Bayes algorithm. We score the opinion as positive, negative and neutral. The Twitter stream and Facebook feeds a pass in JSON format. We accomplish this by mining tweets using Twitter’s search API, and Facebook API subsequently the data’s are stored in Hadoop Distributed File Service. Here we use Linux based environment to get connected with HDFS. The modules of the system includeretrieving the tweets, storing into HDFS analyzing scoring in a tool called R, and the graph.

2. Literature Survey

Social media are computer-mediated tools that allow people to create, share, or exchange information, career interests, ideas, and pictures/videos in virtual communities and networks. Social media defined as "a group of Internet-based applications that build on the ideological and technological foundations of Web."
Types of Analysis in Twitter.

3. Architecture of the System

![Diagram of the system architecture](image)

**Figure 3.1**

**Algorithm:**

Step 1: Start

Step 2: Read values Flume, Hadoop, Hue & Hive

Step 3: Configure Flume

Step 4: Creattwiapi

Step 5: Initilizeapi();

Step 6: Start Hadoop

Step 7: Call Hadoop();

Step 8: End Hadoop
Step 9: Start Flume
Step 10: Call Flume();
Step 11: End Flume
Step 12: Start Hue
Step 13: Function naïve bayes();
Step 14: Call Hive();
Step 15: End Hive
Step 16: Display 'graph"

Initializeapi()
{
    readAccess_Tokenacss_t;
    read Customer_Keycus_k;
    read Customer_Secretcus_s;
    InitializeAlchemyApi
    read alcapi;
}

Call Hadoop();
{
    declare namenodenn;
    declare datanodedn;
    declare yarnnodeyn;
}

Call Flume()
{
    config flume;
    declare flume_tweets;
    Add tweets to Hdfs;
    Store flume_tweets;
Function naïve bayes()
{
    mat= create_matrix(tweet_all, language="english",
    removeStopwords=FALSE, removeNumbers=TRUE,
    stemWords=FALSE, tm::weightTfIdf)
    mat = as.matrix(mat)
    classifier = naiveBayes(mat[1:160,], as.factor(sentiment_all[1:160]))
    predicted = predict(classifier, mat[161:180,]); predicted
    table(sentiment_test, predicted)
    recall_accuracy(sentiment_test, predicted)> 0.65
}

Call Hive()
{
    Query Hive
    declare keyword;
    search keyword;
    Analyze keyword;
}

4. Retrieving data from tweets:

Tweets are retrieved from Twitter using Twitter API, Download Flume and Tweets are stored in memory channel of HDFS sink. They are stored in HDFS File System.
Data Sets:

Tweets in Json Format:


#JacquelineBuilds","contributors":null,"retweeted_status":{"filter_level":"low","contributors":null,"text":'@Asli_Jacqueline will visit #Chennai today to help build homes for Tamil.'},{"filter_level":"low","contributors":null,"text":'@Asli_Jacqueline will visit #Chennai today to help build homes for Tamil.'}

Positive Tweets:

5. Dictionary development:

A dictionary is developed based on Noun, Adjective, and verb. Adverb the counts for certain words are added according to their score in Dictionary.

Verb & Adverb Strength

<table>
<thead>
<tr>
<th>Verb</th>
<th>Strength</th>
<th>Adverb</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Love</td>
<td>1</td>
<td>Complete</td>
<td>+1</td>
</tr>
<tr>
<td>Adore</td>
<td>0.9</td>
<td>most</td>
<td>0.9</td>
</tr>
<tr>
<td>Like</td>
<td>0.8</td>
<td>totally</td>
<td>0.8</td>
</tr>
<tr>
<td>Emotion</td>
<td>Strength</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>enjoy</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>smile</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>impress</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>attract</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>excite</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>relax</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reject</td>
<td>-0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disgust</td>
<td>-0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>suffer</td>
<td>-0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dislike</td>
<td>-0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>detest</td>
<td>-0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>suck</td>
<td>-0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hate</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Meaning</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>😊</td>
<td>Smile</td>
<td>1</td>
</tr>
<tr>
<td>😞</td>
<td>Sad</td>
<td>-1</td>
</tr>
<tr>
<td>:-D</td>
<td>Laugh</td>
<td>1</td>
</tr>
<tr>
<td>:-B</td>
<td>Cool</td>
<td>1</td>
</tr>
<tr>
<td>&lt;3</td>
<td>Heart</td>
<td>1</td>
</tr>
<tr>
<td>:*(</td>
<td>Crying</td>
<td>-1</td>
</tr>
<tr>
<td>X-(</td>
<td>Angry</td>
<td>-1</td>
</tr>
<tr>
<td>😐</td>
<td>Straight face</td>
<td>0</td>
</tr>
<tr>
<td>:*</td>
<td>Kiss</td>
<td>0.5</td>
</tr>
<tr>
<td>\n/</td>
<td>Hi 5</td>
<td>1</td>
</tr>
</tbody>
</table>

6. Experimental Setup:

Step 1:

Download Cloudera on CentOS and Install Hadoop.


Link For Hadoop Download [https://hadoop.apache.org/](https://hadoop.apache.org/)

For running Hadoop, it requires >8GB of RAM

Step 2:

Download and Configure Flume
Step 3:
Configure HDFS and dump Tweets using Flume.
Analyzing the tweets stored in HDFS using the dictionary developed.

Step 4: Generating R tool for Graphical representation.

7. Results

Graph Generated in R Tool
8. Conclusion and future work

In this paper, we have seen the sentiment analysis of the major disaster Chennai Flood occurred in recent times reflected in Tweets and Facebook newsfeeds using big-data concepts like Flume, Hive, and Hadoop HDFS. The result shows that the youngsters participation in rescuing people affected by flood as positive. So we can conclude that “Youngsters are not useless they are used less” by our analysis. This project can be extended by creating a mobile type app could be created as a base line with this idea. The algorithm could be made effective in the future.

9. References


2. Harvinder Jeet Kaur and Rajiv Kumar, ”Sentiment Analysis from Social Media in Crisis Situations”, IEEE 2015.


**Corresponding author:**

Prof. Priya. V*,

Email: vpriyacse@vit.ac.in