International Journal of Advanced Technology in Engineering and Science
Vol. No.5, Issue No. 07, July 2017
www.ijates.com
ijates
ISSN 2348 - 7550

Opinion Mining using RSS Feeds and Social Media News Streams

Miss.Kalyani D.Gaikwad[1], Prof.P.P.Rokade[2]

[1,2]SND COE & RC,Yeola

ABSTRACT

Analysis of the contents which are generated onlineis useful for analysis of social media tasks. A lot of work hasbeen carried out for extracting people sentiments from textualdata. The textual sentiment analysis is needed by researchersto develop systems for predicting political elections, measureeconomic indicators, and so on. Although, social media is sourceof most recent information, it cannot be trustworthy as it iscomposed of several aspects generated by different peoples.In this work we are proposing hybrid approach of sentimentanalysis for area of interest. The hybrid approach consists ofaggregating sentiments from both social media and news feeds.After extracting sentiments from both approaches, they are thenclustered and will be made available for analysis.

Keywords: RSS feeds, Twitter, Sentiment analysis, opinionmining, emotion.

I. INTRODUCTION

Opinion mining is art and science of identifying views, thoughts and opinions of people. Today, due to high availability and usage of internet, people are interested in expressing their views on social media like Twitter. Socialmedia sites, blogs and forums are playing important role in collecting feedbacks of people which are too important to improving the performance of certain product. People also express their feelings about any news, movie on social media platform publically. These opinions can be categorized into positive, negative or neutral classers feeds are xml data files that are used as source of information like news's feed are useful in giving time to time updates to users from their favorite websites.RSS is XML formatted plain text. Format of RSS can be read easily by both automated process and human.

In research field contribution of Opinion mining is in large amount. Sentiment analysis classifies the collection of data into positive, negative and neutral emotions. Two underlying approaches for sentiment analysis are dictionary based and machine learning. The former is popular for public sentiment analysis, and the latter has found limited use for aggregating public sentiment from Twitter data. The research presented in this project is aimed to widen the machine learning approach for aggregating publicemotion. A lot of work has been proposed and models are implemented to compute twitter sentiment analysis.

However the research has proven that real time publicopinions are not always be correct because of inclusionof human emotions. This emotion minimizes degree ofaccuracy and hence cannot beconsidered in applicationswhere accuracy is crucial. E.g. efficient market hypothesis (EMH).The presented research explains how combination of RSS feeds and twitter improves the accuracy of sentiment analysis. Opinions are classified into positive or negativeclasses for analysis of public mood.

In this paper explanation is given how machine learning methodology excels the degree of sentiment accuracy when news data feeds are used in combination with twitter as a more accurate data source. This paper is classified into following sections: Section II contains the previous work done in this field. Our proposed work is described in Section III. Section IV consists of the results generated. Conclusions and future scope are discussed in Section V.

II. REVIEW OF LITERATURE

Many researchers use the sentiment analysis by choosing the words which are used to formulate the view or opinion on specific subjectFor sentiment analysis, combining common-sense knowledge with sentiment analysis can be done through sentic computing [2]. In other researches like [10], clause level sentiment analysis was done in the researchers. They extracted independent clause from the statement.SentiWordNet was used for opinion mining. [6]In this work, a novel approach is proposed based on SentiWordNet, which generates count of score words into seven categories such as strong-positive, positive, weak-positive, neutral, weak-negative, negative and strong-negative words for the opinion mining task and evaluated using machine learning algorithms like Nave Bayes, SVM and Multilayer Perception (MLP) [7].

The domain of analysis of news articles has been traversed before also like [2] but most research uses machine learning techniques to extract sentiments. Researches like [11] have described the way for opinion mining but by analyzing complete articles. There are a few researches like [12] which have analyzed the sentiments by using news headlines only, but by using nave Bayes classifier technique. In our current paper weare aggregating RSS with tweets. In [7] authors have presented a conceptual emotion detection and analysis system for eLearningusing opinion mining techniques.

III. SYSTEM ARCHITECTURE

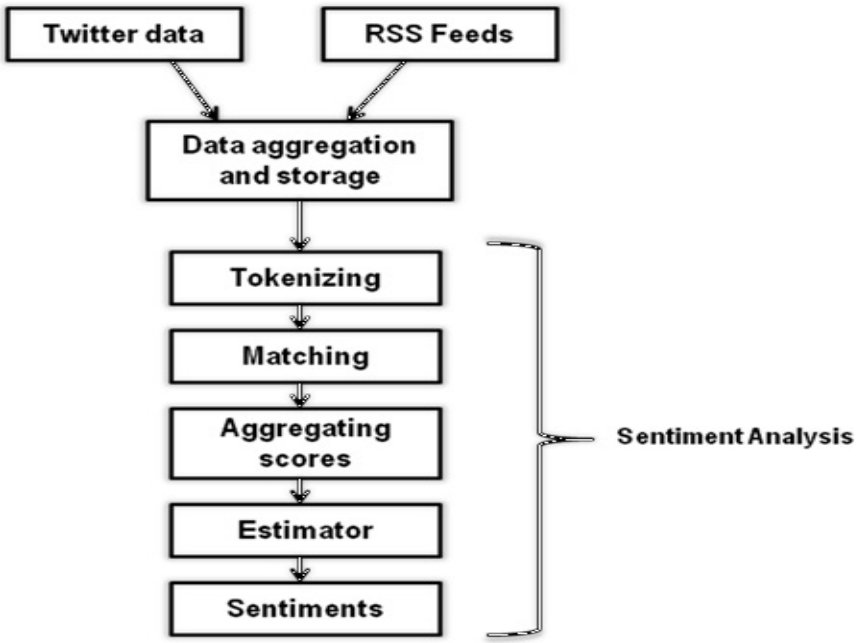

Fig 1: working of proposed system

A.Data aggregation and storage

In this paper,we are first collecting the real time news data stream from Twitter using twitter streaming API.After thatwe are grabbing real time news from RSS feeds. The datafrom RSS feeds is collected, processed and stored using stepsprovided in fig.2

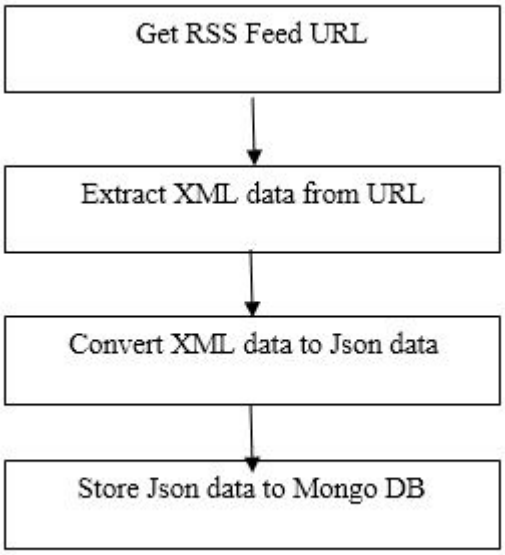

Fig 2: RSS feeds collection and processing

B. Sentiment analysis

Sentiment analysis classifies the collection of data into positive, negative and neutral emotions. Two underlying approaches for sentiment analysis are dictionary based and machine learning. The former is popular for public sentiment analysis, and the latter has found limited use for aggregating public sentiment from Twitter data. The research presented in this project is aimed to widen the machine learning approach for aggregating public emotion. The analyzer which brings out sentence-scores follows an algorithm which calculates cumulative sentiment scores of words present in each line of the parsed sentences. The algorithm which runs for every sentence to calculate its score is comprised of 4 steps tokenizing, matching, aggregatingand estimating.

1) Tokenizing: The news headlines are tokenized using a lexical analyzer using R. The algorithm to tokenize each sentence is as followed:

a.Cleaning up the sentence: The punctuations, control characters, and digits of the parsed sentence were cleaned up and stemmed.

b.Change sentence to lower case: The punctuations, control characters, and digits of the parsed sentence were cleaned up and stemmed.

c.Split the elements into character vector.

d.Flatten lists: To produce a vector which contains all the atomic components which occur in the list, it gets flattened or unlisted.

2) Matching: Since here in this paper, we aim at analyzing it using the machine learning based approach; the next step is to match the words of each sentence with the dictionary vectors in R. The dictionary vector in R has two sub dictionaries of positive and negative words. For each sentence, each word gets matched with the dictionary.If theres a match between the word of the sentence and the word of the positive or negative dictionary

it returns TRUE, otherwise FALSE.The total sentiment score is calculated and the sentence score is stored into an array:

3) Aggregating sentence scores: Once each sentence gets a score, the number of sentences with same score gets counted and put along with the particular sentence score. It gets stored into a data frame with Number of sentences with a unique score and the unique score.

Number of sentences with a unique score = ∑sentences with the unique score.

4) Estimator: The estimator estimates the sentence wise scores and brings out the degree of positivity, negativity and neutrality in percentage.

IV. SYSTEM ANALYSIS

A. Mathematical model

Input Set= I1, I2

Where,

I1= Tweets

I2= RSS feeds.

Intermediate Output Set.

E=E1, E2

Where,

E1=Positive score

E2= Negative score

Final Output Set

D= D1, D2, D3

Where,

D1=Degree of positivity D2= Degree of Negativity D3=Degree of Neutrality Following figure shows functional dependency of system:

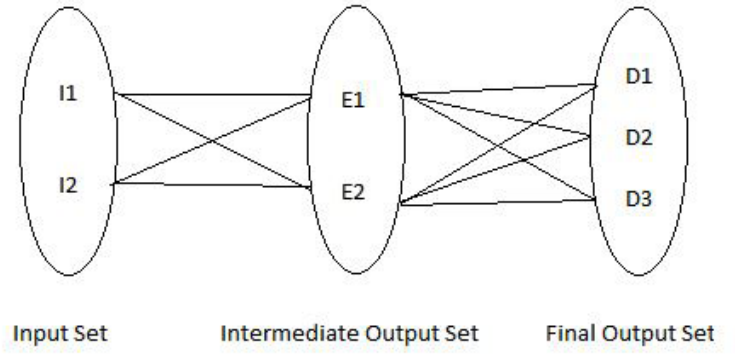

Fig. 3. Functional dependency of system

B. Results and Discussion

The comparative results for Twitter only and Twitter along with RSS feeds are provided in following table.

Table 1. Comparative Results Table

Metric / Classifier	TP	TN	FP	FN	Accuracy	Precision	Recall	F1 Score
Twitter Sentiment Analysis	10	07	09	02	0.7	0.83	0.66	0.47
Twitter + RSS Feeds Sentiment Analysis	14	12	01	01	0.92	90	0.85	0.58

V. CONCLUSION

The sentiment analysis plays a vital role in many applications including Natural language processing, Artificial Intelligence, etc. In this paper we have performed the sentiment analysis on RSS feeds along with the tweets. We can classify these News and tweets according to area which will help indecision making. It will also help to overcome the weaknesses in particular area. The opinion mining done with RSS feeds and tweets can help a lot to predict the needs of people as well as their views about particular topic. For twitter sentiment analysis lot of research has been done and many models are implemented. Our research explains that due to inclusion of emotions, real time public opinions are not always accurate. So that we have combined the twitter data with RSS feeds to achieve the accuracy.

REFERENCES

[1] Apoorv Agarwal,Vivek Sharma,Geeta Sikka and Renu Dhir,Opinion Mining of News Headlines using SentiWordNet,Inside IEEE, Symposium on Colossal Data Analysis and Networking (CDAN) ,2016.

[2] Prashant Raina,Sentiment Analysis in News Articles Using Sentic Computing, Inside IEEE 13th International Conference on Data Mining Workshops,2013.

[3] Daniel Dor,On newspaper headlines as relevance optimizers, Elsevier Journal of Pragmatics 35 (2003) 695721.

[4] Ang Yang,Jun Zhang,Lei Pan and Yang Xiang,Enhanced Twitter Sentiment Analysis by Using Feature Selection and Combination, 2015 International Symposium on Security and Privacy in Social Networks and Big Data.

[5] D. Zhogliang, Y. Yanpei, Y. Xie, W. Neng and Y. Lei,Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques, Third IEEE International Conference on Data Mining (ICDM03).

[6] Shoiab Ahmed andAjit Danti ,A Novel Approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using Web Data , IEEETrends in Automation, Communications and Computing Technology (I-TACT-15),2015.

[7] Haji H. BINALI, Chen WU and Vidyasagar POTDAR, A New Significant\ Area: Emotion Detection in E-learning Using Opinion Mining Techniques, 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009.

[8] Raj Parkhe and Bhaskar Biswas,Sentiment analysis of movie reviews: finding most important movie aspects using driving factors, Published in Journal Soft Computing - A Fusion of Foundations, Methodologies and Applications.

[9] Khairullah Khan, Baharum B.Baharudin, Aurangzeb Khan and Fazal-e- Malik Niemegeers,A Mining Opinion from Text Documents: A Survey, 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009.

[10] T. Thet, J. Na, C. Khoo and S. Shakthikumar, Sentiment analysis of movie reviews on discussion boards using a linguistic approach, Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion - TSA '09, 2009.

[11] A. Balahur and R. Steinberger, Rethinking Sentiment Analysis in the News: from Theory to Practice and back, JOINT RESEARCH CENTRE The European Commission's in-house science service, 2015.

[12] H. kaur and D. Chopra, Sentiment Analysis of News Headlines using Nave Bayes Classifier, Council For Research And Development Enterprise,2015.