ZKG INTERNATIONAL ISSN: 2366-1313

HADOOP FRAMEWORK TO PROGRESS THE PROFILE

[1]Mrs.C.ARCHANA, [2]KATCHA PRABATH RAJ, [3]YASHWASRI MOTE, [4]KODARI ANIL, [5]KAKARLA PAVANSAI

(Assistant Professor) ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

B,tech scholar ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

ABSTRACT

Social media is a great platform in our modern society to share our views as we feel to the situation. Here, it is the platform turned to like, dislike and comments to every individual posts of even stranger. Comment section is a crucial one which is playing a role of express the feel. Based on this comments, we are introducing a website where the who are using informal language towards a post leads to the low profile rate and who are not they get high profile rate and others remains neutral rate. This project examines libraries developed on the basis of machine learning algorithms, as well as algorithms that use sentimental analysis. The project compares the probability with which the described algorithms determine the tone of texts. Most sentiment analysis systems use bag-of-words approach for mining sentiments from the online reviews and social media data. Rather considering the whole sentence/paragraph for analysis, the bag-ofwords approach considers only individual words and their count as the feature vectors. This may mislead the classification algorithm especially when used for problems like sentiment classification. Traditional ma- chine learning algorithms like Naive Bayes, Maximum Entropy, SVM etc. are widely used to solve the classification problems. These machine learning algorithms often suffer from biasness towards a particular class. In this paper, we propose Natural Language (NLP) based approach to enhance the sentiment classification by adding semantics in feature vectors and thereby using ensemble methods for classification. Adding semantically similar words and context-sense identities to

ISSN: 2366-1313

the feature vectors will increase the accuracy of prediction. Experiments conducted demonstrate that the semantics based feature vector with ensemble classifier outperforms the traditional bag-of-words approach with single machine learning classifier by 3-5%

1.INTRODUCTION

Mining opinions and analyzing sentiments from social network data help in various fields such as event prediction, analyzing overall mood of public on a particular social issue and so on. This helps in exploring potential customers, business demands, and predicting events. To understand the view or opinion implied in the text, the study of sentiment analysis is needed. Text sentiment analysis involves the application of computational powers in understanding sentiment implied in the text. Natural Language Processing involves the task of finding the semantic meaning of the text content and thereby analyzing the text information. There are lot of NLP techniques available at present to aid the sentiment classification problem. Stemming, Stop Word removing, Parts of Speech Tagging (POS), Named Entity Recognition (NER) followed by bag-of-words are some of the preprocessing techniques used for feature vector formation. The objective of the project is to successfully implement a sentiment analysis tool that retrieves tweets and is able to infer the sentiment of these tweets. Sentiment analysis can be classified into 3 types: 1. Sentence Level: Sentence Level sentiment analysis focuses on an entire sentence and calculates the sentiment of the sentence as a whole. 2. Aspect Level: Aspect Level sentiment analysis focuses on certain aspects/characteristics of an entity. 3. Document Level: In a document there may be many opinions expressed throughout the length of the document.

1.2 PROBLEM STATEMENT

The main problem with bag-of-words approach is that, it considers only individual words and their frequencies during feature vector creation. Rather considering the whole sentence/paragraph for analysis, the bag-of-words approach considers only individual words and their count as the feature vectors. This may mislead the classification algorithm especially when used for problems like sentiment classification. In supervised learning systems when the feature vectors are selected based on keywords the

ISSN: 2366-1313

classification will happen based on the keywords only. The relationship between the keywords will be missed out.

1.2 DESCRIPTION

To overcome the above said problem, in this paper we propose adding semantics while extracting feature vectors. Exploring semantic relationships between words can be best done with WordNet's synsets. Synsets are sets of cognitive synonyms expressing a particular concept. So using synsets along with the original keyword will cover the whole domain or concept. This helps in increasing the predicting power of a classification algorithm. To solve the problem of identifying the sense of a polysemic word based on the context of its occurrence, Word Sense Disambiguation (WSD) is used . To some extent, WSD solves the problem of plain bag-of- words approach by considering whole sentence while sense. Various Machine Learning algorithms say Support Vec- tor Machines (SVM), Maximum Entropy, Navies Bayes are widely used to solve the classification problems. Ensemble methods of machine learning, combine the effect of multiple machine learning algorithms to obtain better predictive power than their constituent algorithms in solitude.

There are many types of ensemble classifiers depending on how the training datasets are chosen and how the results are combined. Our method differs from the existing frequency based bag- of-words approach by adding sentence level meaning to feature vectors with the help of word senses. Also it uses the ensemble framework to overcome the class imbalance problem. This paper involves analyzing the mood of the society on a particular news or topic from Twitter posts. The key idea pro- posed in the paper is to increase the accuracy of classification by including Natural Language Processing Techniques (NLP) especially Synsets and Word Sense Disambiguation with ensemble methods for classification. The key contributions in the proposed work are: 1. To the best of our knowledge, we are the first to propose the use of Synsets and Word Senses as features in the sentiment analysis systems. 2. To the best of our knowledge, this is the first work on the combined framework of semantics based ensemble classification system.

2.LITERATURE SURVEY

In literature there are two main approaches followed in mining sentiment from the SNS: Dictionary Based (DB) and Machine Learning Based (MLB). In general all the

ISSN: 2366-1313

DB techniques refer to a pre-built dictionary for classifying the sentiment. The potential limitation with DB systems is that, the strength of classification depends on the reference dictionary used. Also, most of these systems use Bag-of-Words concept which lacks domain/context based semantics. On the other hand MLB systems achieve domain based sentiment classification due to the presence of domain specific training data and the learning efficiency of classification algorithms. Class imbalance problem and linguistic variations in text can be overcome by the bootstrapping ensemble framework. Coletta et al. demonstrated the performance of SVM combined cluster ensemble classification on Twitter data. The objective of the project is to successfully implement a sentiment analysis tool that retrieves tweets and is able to infer the sentiment of these tweets. Sentiment analysis can be classified into 3 types:

1. Sentence Level: Sentence Level sentiment analysis focuses on an entire sentence and calculates the sentiment of the sentence as a whole.

2. Aspect Level: Aspect Level sentiment analysis focuses on certain aspects/characteristics of an entity.

3. Document Level: In a document there may be many opinions expressed throughout the length of the document.

3.SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

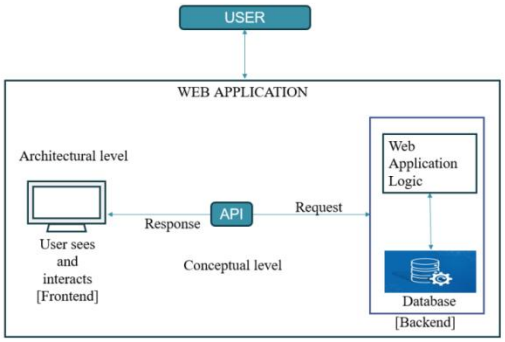

Fig.3.1 : System Architecture

3.2 ACTIVITY DIAGRAM:

The process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions.

ISSN: 2366-1313

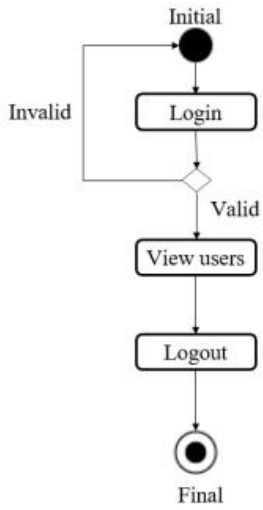

Fig.3.2 a Activity diagram for admin

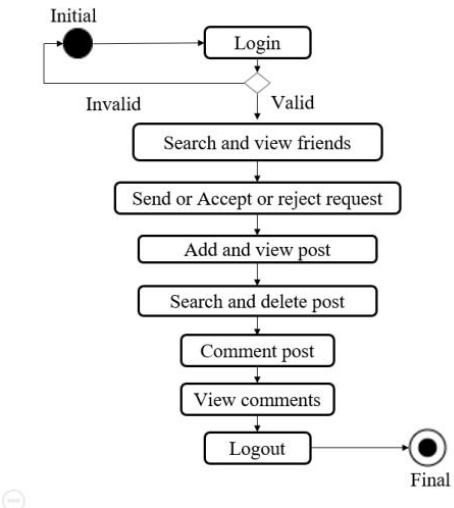

Fig.3.2 b Activity diagram for user

4. OUTPUT SCREENS

ADMIN SESSION

Admin login

- This page helps admin to login into the account using his credentials.

Fig.4.1 Admin login

Admin dashboard

- This page guides to check the activities of user.

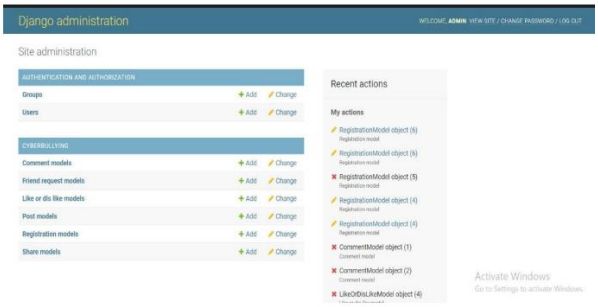

Fig.4.2 Admin dashboard

User profile

- This page shows the complete details of the users, which can be edit or delete by admin.

ISSN: 2366-1313

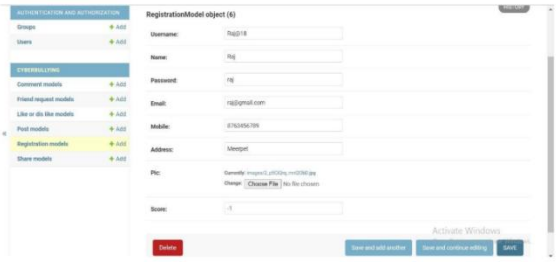

Fig.4.3 User profile

USER SESSION

User login

• This page helps users to login into their account using their credentials. It also leads to registration page for new users.

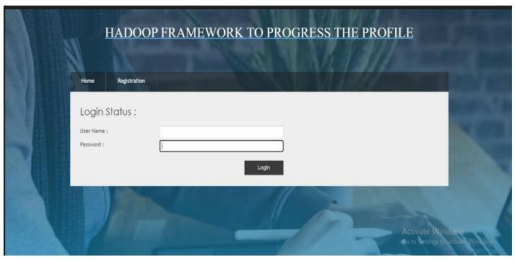

Fig.4.4 User login

Home page

• This page shows posts or information posted by other users .It also helps in finding friends or post.

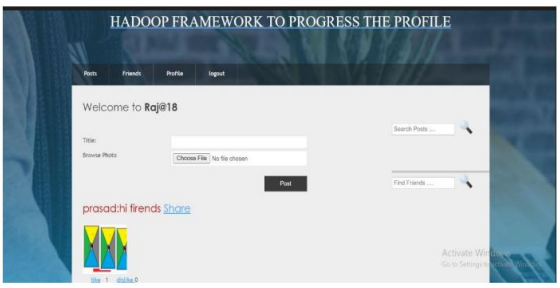

Fig.4.5 Home page

Profile update

• This page shows the user profile helps in changing their details.

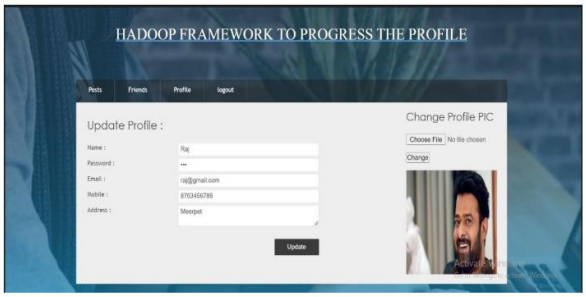

Fig.4.6 Profile update

5.CONCLUSION

The proposed system collects data from Twitter social net- work site and does NLP techniques to extract features out from the tweets. Word Sense Disambiguation and WordNet synsets are also added to the feature vector to increase the accuracy of prediction. Then various Ensemble methods of classification are applied to classify the sentiment of the data as Positive, Negative and Neutral. It is observed that the ensemble method outperforms the traditional classification methods by about 3- 5%. Among the ensemble methods Extremely Randomized Trees classification performs better than others. Though the addition of Synsets and Word Senses increase the level

of accuracy in classification, they add to the increase in size of the feature vector which in turn leads to increase in computation overhead.

6.FUTURE ENHANCEMENT

Future work will be focused on the computational cost by optimizing the feature vector size. More enhanced methods to consider the paragraph level context while classifying the sentiment will also be concentrated in the future work. The future work of our project enhance the awareness of foul language to the positive way for the content post. By the score points of the profiles they aware for their scores to be in positive because nobody want to be in negative in other perspective. In future project can be developed as it can give rating to graphic content post.

7. REFERENCES

[1] C. Fellbaum, WordNet. Wiley Online Library, 1998.

[2] M. Stevenson and Y. Wilks, "Word-sense disambiguation," The Oxford Handbook of Comp. Linguistics, pp. 249–265, 2003.

[3] A. Hassan, A. Abbasi, and D. Zeng, "Twitter sentiment analysis: A bootstrap ensemble framework," in Social Computing (SocialCom), 2013 International Conference on. IEEE, 2013, pp. 357–364.

[4] L. F. Coletta, N. F. F. d. Sommaggio Silva, E. R. Hruschka, and E. R. Hruschka, "Combining classification and clustering for tweet sentiment analysis," in Intelligent Systems, 2014 Brazilian Conference on. IEEE, 2014, pp. 210–215.

[5] J. Bollen, A. Pepe, and H. Mao, "Modeling public mood and emo- tion: Twitter sentiment and socio-economic phenomena," arXiv preprint arXiv:0911.1583, 2009.

[6] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the omg!" ICWSM, vol. 11, pp. 538–541, 2011.

[7] P. T. Ngoc and M. Yoo, "The lexicon-based sentiment analysis for fan page ranking in facebook," in Information Networking (ICOIN), 2014 International Conference on. IEEE, 2014, pp. 444–448.

[8] A. Minanovic, H. Gabelica, and Z. Krstic, "Big data and sentiment anal- ysis using knime: Online reviews vs. social media," in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th

ISSN: 2366-1313

International Convention on. IEEE, 2014, pp. 1464– 1468.

[9] A. Porshnev, I. Redkin, and A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis," in Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, 2013, pp. 440– 444.

[10] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of facebook statuses using naive bayes classifier for language learning," in Information, Intelligence, Systems and Appli- cations (IISA), 2013 Fourth International Conference on. IEEE, 2013, pp. 1–6.