

Comparison of Different Hybrid Approaches Used for Sentiment Analysis: Survey

Research Scholar Mansi Chauhan, Research Scholar Devangi Paneri

Department of Computer Engineering,
V.V.P Engineering College, Rajkot, India
mansih1998@gmail.com, 15cse101@gardividypath.ac.in

Abstract- In Today's Technological Life Social media quirts a specious amount of Information. Social media has become a tremendous source of acquiring Users Opinions. It also helps to analyze how people, particularly consumers, feel about a particular topic, product or Idea. Among such opinions plays an important role in analyzing different business aspects. Sentiment analysis therefore becomes an effective way of Understanding public Opinions. Business Organizations can predict best Decision with Using Sentiment analysis. A lot of Research work has been done on Sentiment analysis in order to classify the opinions. Researchers have tested a variety of methods for automating the sentiment analysis process but very few Researchers are using Hybrid Approaches. This research paper shows the advantage of hybrid approaches to improve classification accuracy Compare to individuals. In proposed work a comparative study of the effectiveness of hybrid approaches was used for Sentiment analysis. Empirical results indicate that the hybrid approaches outperform compare to this individuals Classifiers.

Keywords- Machine learning, Data mining, Hybrid approach, Sentiment analysis.

I. INTRODUCTION

Recently, the use of Internet all Over the World is getting more and more increasing day by day. Internet based applications or websites that allow users to create and share information, ideas, Opinions, etc. Through virtual communities and social networking. With the growth of the internet based social media and also increase the social media users. So, social media is playing a vital role in modern life without any restrictions and hesitation, users are allowed to post their view and thoughts on social media [4].

The Users are giving their opinions not Only on product but also giving opinions on politics, terrorism, economy, movies, fashion, justice, humanity, etc. these opinions are very much useful to make knowledgeable decision. and these opinions are not only useful for customers but also useful for business organizations and sellers to know about good or bad features related to them [7].

So, A huge amount of data is being Created on a daily basis. This data is in the form of text and also in an unstructured way. Making sense of this data is an important task. This is where data mining comes in to picture [8]. For making knowledgeable decision we need to be classify properly this data in order to provide any useful information to customers, Business organization and seller. This is why Sentiment analysis needed.

Sentiment analysis also known as Opinion mining, Sentiment mining, opinion extraction, emotion mining,

attitude mining or subjective mining. Sentiment analysis refers to the classification of a text based on sentiment polarities present in it. Sentiment polarities are positive, negative, neutral. Sentiment analysis is very much beneficial in predicting future scope.

However, manually analyzing a large amount of data in limited time can be wearisome and challenging task.

For Sentiment Classification of this huge amount of data, researchers have tested variety of methods for automating the sentiment analysis process. Methods are from machine learning and deep learning techniques like Support Vector Machine, Naive Bases, Decision tree, Random forest, Genetic algorithm, Long-short term memory, Convolution Neutral network etc. this technique giving best accuracy in sentiment analysis.

Also, Hybrid approaches of these techniques are available and most important thing is hybrid approaches giving higher accuracy compare to individuals. also, some drawbacks are overcome in hybrid approaches. But Very few researchers are using this hybrid approaches.

So, this paper gives Comparison study of different hybrid approaches with Individuals to show the advantages of hybrid approaches.

This Paper has been Structured in four Sections. Section 2 briefs about the related work done in this area. Section 3 discusses Comparison of hybrid approaches with individuals. In Section 4 Conclusion has been deducted.

II. RELATED WORK

Konstantinas korovkinas and Paulius Danenas[1] Worked on hybrid method of Support Vector Machine and K-means clustering algorithm for Sentiment Classification on textual data. Also, hybrid technique to improve support vector machines classification accuracy using training data Sampling and hyper parameter turning. Training data for Subset30k is selected randomly, which might negatively affect accuracy in different runs. Therefore, multiple runs are required for more objective results. In this paper advocate the use of clustering base instance selection method (K-means). So, the use of hybrid method considered as a positive and significant step towards more efficient sentiment analysis.

M. Govindarajan [2] proposed a new hybrid technique Naive Bayes-Genetic Algorithm method. hybrid technique is investigated and evaluated their performance based on the movie review data. A hybrid Naive Bayes and Genetic Algorithm model and Naive Bayes, Genetic Algorithm models as base classifiers are designed. A hybrid system was proposed to make optimum use of the best performances delivered by the individual base classifiers and the hybrid approach. Genetic algorithm exhibits better performance than Naive Bayes in the important respects of accuracy. The hybrid classifier shows the Significant improvement over the single classifier.

H. A. Shehu and S. Toka [3] proposed hybrid approach of Support Vector Machine and Random forest algorithm for Sentiment Classification on Turkish Twitter data. In this work 13k Turkish tweets has been collected from the twitter. the datasets are classified to be either positive, Negative and Neutral based on their contents. The main dataset was converted to a stemmed dataset. In this paper Experimental results have shown that while support vector machine performs better when it comes to classifying negative and neutral stemmed data, random forest algorithm performs better in classifying positive stemmed data. Hence, developed hybrid approach achieves higher accuracy compare to individual classifiers.

Priya Kumari and Md. Tanwir Uddin Haider [8] proposed a Hybrid approach foe sentiment analysis on Aadhaar. The Government of India is linking the Aadhaar card with many government schemes, but there are many issues with respect to security and privacy of the aadhaar database that needs to be addressed. Main scope of this paper is to find the sentiment of people on Aadhaar limited to the aspects “Enrollment process”, “mAdhaar”, “Customer care”, and “Security” using hybrid approach. This dataset is further labeled for polarity to define positive, negative and neutral values for training dataset using SentiWordNet. Hybrid approach is combination of machine learning classifiers with Long Short-Term memory (LSTM) network to maximize classification accuracy. Also in hybrid approach, Wrapper method

(recursive elimination cross-validation) is used to select the best performing feature subset that depends on the performance of machine learning classification algorithm. In hybrid approach we have used base classifiers such as naive bayes (NB), Logistic regression (LR) and Support vector machines (SVM) to do comparative analysis of result. the experimental result of classifiers confirms that hybrid approach performs better than base classifier for Aadhaar domain dataset.

Nisha Yadav and Rajeev Kumar [14] proposed a new hybrid classification method for Sentiment Analysis on online reviews of Restaurants and the analysis of their quality on the basis of the review data for restaurants. hybrid classification method id combination of three machine learning classifiers which are Naive bayes, Support Vector machine and Genetic algorithm. First, the hybrid approach and the Naive bayes, Support Vector machine and Genetic algorithm models are built as base classifiers. then, a hybrid approach is proposed to make optimum use of the best performance of the individual base classifiers and of the hybrid approach. The hybrid model shows a higher percentage of classification accuracy than the individual classifier and decreases the test time due to the data dimensions reduction.

Yassine Al Armani and Mahamed Lazzar [5] proposed A novel hybrid approach to identify product reviews offered by Amazon. In this paper approach focuses on the analysis of feeling resulting from reviews of products using original text search techniques. These reviews could be classified as having a positive or negative feeling based on certain aspects in relation to a query based on terms. A novel hybrid approach was combination of two machine learning classifiers which are Support Vector machine and Random forest. This is useful for consumers who want to research the sentiment of products before purchase or also help to Business Organization that want to monitor the public sentiment of their brands. The result summarize that the proposed method gives higher accuracy compare to individual classifiers in this Amazon dataset.

Abinash Tripathy and Abhishek Anand [6] proposed hybrid machine learning approach from two machine learning algorithm which are Support Vector machine and Artificial Neural Network In order to classify the sentiment of Review data associated with movies. The Reviews framed are often made in text format. while preprocessing the text reviews, each word of the reviews is considered as a feature. Thus, selection of right kind of features needs to be carried out to select the best features from the set of all features. As the Support Vector machine method analyzes most of the features like words present in the review, it provides a sentiment value to it, then using the sentiment value criteria, it selects the best features from the list of features. By varying the no. of neurons in the hidden layer of Artificial Neural Network, the value of accuracy is found out. The final output indicates whether

the review is of positive or negative. The accuracy obtained in hybrid approach is found to be comparably better than the results obtained by individual approach.

Anwar Ur Rehman and Ahmad Kamran Malik [13] proposed a hybrid model using Long Short-term Memory (LSTM) model and Convolution Neural Network (CNN) model named as hybrid CNN-LSTM model to accuracy classify the sentiment polarities. The CNN model efficiently extracts higher level features using convolution layers and max – pooling layers. The LSTM model is capable to capture long-term dependencies between words sequences. The proposed model combines set of features that are extracted by convolution and global max – polling layers with long-term dependencies. The proposed model also uses dropout technology, normalization and a rectified linear unit for accuracy Improvement. Here, two movie review datasets were used which are IMDB movie reviews and Amazon movie reviews datasets. The proposed hybrid CNN-LSTM model performs very well on two movie reviews datasets as compared to single CNN and LSTM models in terms of accuracy.

Yassine Al Armani and Mahamed Lazzar [9] proposed hybrid approach of Support vector machine and decision tree. In this paper shows comparison of popular classifier was performed to classify SMS text either positive and negative. Compared Several methods with proposed approach, which are very suitable for generating rules in classification technique. It was concluded that the proposed approach has high accuracy and low CPU time than the other algorithm.

M. Govindrajan [4] proposed new hybrid classification method based on coupling classification methods using arcing classifier and their performances are analyzed in terms of accuracy. Hybrid method was designed using Naive Bayes (NB) and Support Vector Machine (SVM). New hybrid method is investigated and evaluated their performance based on the movie review dataset. hybrid method is proposed to make optimum use of the best performances delivered by the individual classifiers and the hybrid method. the hybrid NB-SVM shows higher percentage of classification accuracy than the individual classifiers.

III. COMPARISON BETWEEN HYBRID APPROACHES AND INDIVIDUAL CLASSIFIER

Dataset	Classifiers	Accuracy
Amazon reviews	Support vector machine (SVM)	87.70 %
	K-means	87.59 %

	Hybrid approach of SVM - Kmeans	88.32 %
Movie review data	Naive – Bayes (NB)	91.15 %
	Genetic algorithm (GA)	91.25 %
	hybrid NB-GA method	93.80 %
Small dataset of Turkish Twitter data	Random forest (RM)	75.9%
	Support vector machine (SVM)	76.4%
	hybrid RF-SVM approach	86.4%
large dataset of Turkish Twitter data	Random forest (RM)	71.2%
	Support vector machine (SVM)	67.3%
	hybrid RF-SVM approach	82.8%
Aadhaar Twitter Data	Naive – Bayes (NB)	71.4%
	Support vector machine (SVM)	80.63 %
	Logistic regression	77.7%
	hybrid approach of NB-LSTM	85.3%
Restaurant Review data	Naive – Bayes (NB)	85%
	Support vector machine (SVM)	85.20 %
	Genetic algorithm (GA)	85.30 %
	Hybrid method of NB-SVM-GA	93%
Amazon product Review	Random forest (RM)	82%
	Support vector machine (SVM)	82.4%
	hybrid RF-SVM approach	84.7%
IMDB Dataset	Support vector machine (SVM)	86%
	Hybrid SVM and ANN method	95%

polarity dataset	Support vector machine (SVM)	93%
	Hybrid SVM and ANN method	96%
SMS-Text Dataset	Support vector machine (SVM)	98.95 %
	Decision Tree (DT)	96.60 %
	Hybrid approach of SVM - DT	98.98 %
Movie review data	Naive – Bayes (NB)	91.15 %
	Support vector machine (SVM)	91.35 %
	hybrid NB-SVM method	94.15 %
IMDB Dataset	Convolutional neural network (CNN)	86%
	Long short-term memory (LSTM)	89%
	Hybrid CNN-LSTM model	91%

IV. CONCLUSION

Now a days very few researchers are using hybrid approaches. In this paper, we presented comparative study on the different Hybrid approaches with its individual classifiers for sentiment analysis.

Classifiers used in this paper from machine learning and deep learning models. Comparison between Hybrid approaches with its individual classifiers shows that Hybrid approaches giving higher accuracy compare to individual classifiers. So, we recommend and urge researchers to use these hybrid approaches for their research work.

REFERENCES

- [1] Konstantinas KOROVKINAS, Paulius DAN'ENAS, Gintautas GAR'SVA "SVM and k- Means Hybrid Method for Textual Data Sentiment Analysis", International Journal of Advanced Computer Research 2019, vol.7, pp.47-60.
- [2] M. Govindarajan, "Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm", International Journal of Advanced Computer Research, December-2013, vol.7, pp.139-145.
- [3] H. A. Shehu, S. Tokat, "A Hybrid Approach for the Sentiment Analysis of Turkish Twitter Data", Springer-2020, pp. 182–190.
- [4] M. Govindarajan, "Sentiment Analysis Of Restaurant Reviews Using Hybrid Classification Method", International Journal of Soft Computing and Artificial Intelligence, May-2014, vol.2, pp.17-23.
- [5] Yassine Al Amrani, Mohamed Lazaar, Kamal Eddine El Kadiri, "A Novel Hybrid Classification Approach for Sentiment Analysis of Text Document", International Journal of Electrical and Computer Engineering (IJECE), December 2018, Vol. 8, pp. 4554-4567.
- [6] Abinash Tripathy, Abhishek Anand, Santanu Kumar Rath, "Document-level sentiment classification using hybrid machine learning approach", Springer-2017, vol.53, pp.805–831.
- [7] Ashwini Save, Narendra Shekokar, "A Brief Survey of Sentiment Analysis", Springer 2020, pp.353-368
- [8] Priya Kumari and Md. Tanwir Uddin Haider, "Sentiment Analysis on Aadhaar for Twitter Data—A Hybrid Classification Approach", springer-2020, pp-309-318.
- [9] Yassine Al Amrani, Mohamed Lazaar, Kamal Eddine El Kadiri, "Sentiment Analysis Using Hybrid Method Of Support Vector Machine And Decision Tree", Journal of Theoretical and Applied Information Technology, April 2018, Vol.96, pp. 1886-1895.
- [10] Sisi Liu and Ickjai Lee, "A Hybrid Sentiment Analysis Framework for Large Email Data", International Conference on Intelligent Systems and Knowledge Engineering, November 2015, pp. 324-330.
- [11] Addlight Mukwazvure, K.P Supreethi, "A Hybrid Approach to Sentiment Analysis of News Comments", IEEE-2015.
- [12] Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, "Sentiment Analysis of Tweets using Machine Learning Approach", International Conference on Contemporary Computing, August 2018.
- [13] Anwar Ur Rehman, Ahmad Kamran Malik, Basit Raza, Waqar Ali, "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis", springer 2019, vol.78, pp.26597–26613.
- [14] Nisha Yadav, Rajeev Kumar, Dr. Bhupesh Gour, Dr. Asif Ullah Khan, "Extraction-Based Text Summarization and Sentiment Analysis of Online Reviews Using Hybrid Classification Method", International Conference on Wireless and Optical Communications Networks, February 2020.
- [15] Balage Filho PP, Avanço L, Pardo TA, Nunes MG, NILC USP: "an improved hybrid system for sentiment analysis in Twitter messages", Sem Eval 2014.
- [16] Jagtap B, Dhotre V, "SVM and HMM based hybrid approach of sentiment analysis for teacher feedback assessment", International journal Emerging Trends

- Technology Computer Science (IJETCS) 2014, pp.229–232.
- [17] S Chandrakala and C Sindhu, (2012), “Opinion Mining and sentiment classification a survey”, ICTACT journal on soft computing, vol.3, pp. 420-427.
- [18] Rui Xia, Chengqing Zong, Shoushan Li, (2011), “Ensemble of feature sets and classification algorithms for sentiment classification”, Information Sciences, 181, pp. 1138–1152.
- [19] Haque, T.U., Saber, N.N., Shah, F.M. (2018). Sentiment analysis on large scale Amazon product reviews. In Innovative Research and Development (ICIRD), 2018 (pp. 1–6).
- [20] Shivani Bahri, Pranav Bahri, Sangeeta Lal,” A Novel approach of Sentiment Classification using Emoticons”, International Conference on Computational Intelligence and Data Science (ICCIDS 2018), vol.132, pp.669-678.