ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2019.0266

SENTIMENT ANALYSIS FOR PRODUCT REVIEW

Najma Sultana[1], Pintu Kumar[2], Monika Rani Patra[3], Sourabh Chandra[4] and S.K. Safikul Alam[5]
[1,2,3,4]Department of Computer Science and Engineering, Calcutta Institute of Technology, India
[5]Department of Computer Science and Technology, Raja Ranajit Kishore Government Polytechnic, India

Abstract

Sentiment analysis is defined as the process of mining of data, view, review or sentence to predict the emotion of the sentence through natural language processing (NLP). The sentiment analysis involve classification of text into three phase "Positive", "Negative" or "Neutral". It analyzes the data and labels the 'better' and 'worse' sentiment as positive and negative respectively. Thus, in the past years, the World Wide Web (WWW) has become a huge source of raw data generated custom or user. Using social media, e-commerce website, movies reviews such as Facebook, twitter, Amazon, Flipkart etc. user share their views, feelings in a convenient way. In WWW, where millions of people express their views in their daily interaction, either in the social media or in e-commence which can be their sentiments and opinions about particular thing. These growing raw data are an extremely high source of information for any kind of decision making process either positive or negative. To analysis of such huge data automatically, the field of sentiment analysis has turn up. The main aim of sentiment analysis is to identifying polarity of the data in the Web and classifying them. Sentiment analysis is text based analysis, but there are certain challenges to find the accurate polarity of the sentence. This states that there is need to find the better solution to get much better results than the previous approach or technique used to find polarity of sentence. Therefore, to find polarity or sentiment of, user or customer there is a demand for automated data analysis techniques. In this paper, a detailed survey of different techniques or approach is used in sentiment analysis and a new technique which is proposed in this paper.

Keywords:

Sentiment Analysis, Naïve Bayes, Mining, Support Vector Machine, Polarity, Semantic

1. INTRODUCTION

Every single day huge amount of information, reviews or opinions are getting stored in the websites of social media or e-services in the form of raw data. To work with those raw data proper methods required. Most of the methods either focus on verbs, nouns, adverbs or adjectives. Although a recent study has shown that combination of adverbs and adjectives in sentiment analysis is better than adjectives alone [8]. But no work has focused on all the possible combinations of adverbs, adjectives and verbs. This paper presents the theoretical analysis of some well-known methods or proposal of Sentiment Analysis. Both the advantages and disadvantages of the discussed methods are considered to add new features in the proposed approach. The new approach follows machine learning technique at document level with combination of adjectives, adverbs, and verbs. The following combinations are taken into for analysis, adverbs-adjectives, adverbs-verbs, adjectives-verbs and adverbs-adjectives-verbs along with adverbs, adjectives and verbs. The Standard classifier like Naive Bayes (NB), Linear Model and Decision Tree are used to deduct result and for analysis. This section presents the classification of Sentiment Analysis followed by detailed revision of the existing methods related to sentiment analysis.

1.1 CLASSIFICATION OF SENTIMENT ANALYSIS

The approaches made in sentiment analysis can be categorized based on techniques used, structure of dataset and level of rating, etc. These categorization are again sub-categorized as below:

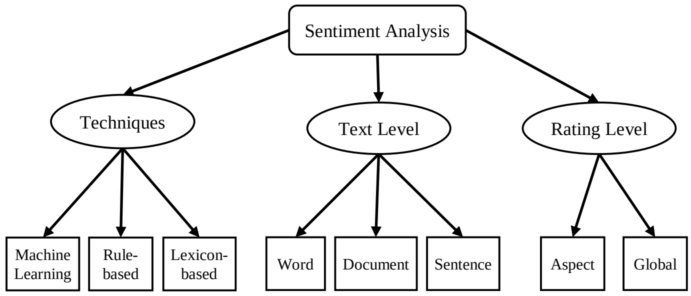

Fig.1. Categorization of Sentiment Analysis [6]

Technically Sentiment Analysis can be done by either,

1. *Machine Learning*: Dataset are to be trained beforehand. Using standard machine algorithms polarities are detected [6].
2. *Rule based*: Extracts information from dataset and try to asses them according to the polarity of words. There are different rules such as negation words, idioms, dictionary polarity, emoticons etc. [13].
3. *Lexicon-based*: Using Semantic orientation i.e. measurement of opinion and subjectivity of a review or comment it generates sentiment polarity (either positive or negative) [7].

Based on structure of dataset or text level, SA can be further classified as at aspect level, document level or sentence level [14]. Document-level sentiment analysis aims to classify an opinion or view in terms of "positive" or "negative" sentiment. It considered the whole document as a bunch of information unit. In sentence level sentiment analysis aims to classify sentiment expression or opinion in each sentence. However, there is no fundamental difference between document level and sentence level classification because sentences are just short then document. Thus we have use document level Approach to find the polarity of the sentence or document in terms of "positive" or "negative" sentiment. Rating of a product can be done at both aspect level and global level. This is another classification of SA. Although most of the e-shopping portals, movie review sites determine strength of sentiment at global level [13].

2. EXISTING WORKS

In this 21st century, people are more social in social media, internet, online shopping etc. Thus directly or indirectly online judgments, opinions are eventually gaining great attention. But the real deal is analysis or mining of opinions. Below is the review of some existing solutions available for SA. These methods are also briefly tabulated in Table.1.

OPINE, an unsupervised, web-based information extraction system proposed by Propescu et al. [5] extracted product feature and opinions from reviews. It identifies product feature, opinion regarding product feature, determines polarity of opinions and then ranks product accordingly [9]. In feature identification, nouns from dataset or reviews are extracted. Frequencies higher than the threshold frequency are kept else discarded. OPINE's feature assessor is used to extract explicit features (occurrence of frequent features) [4]. Researchers have used manual extraction rule to extract data [4]. Advancement of OPINE is its domain independency. But fails to find its real life uses as OPINE system is not easily available.

Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone, is a linguistic approach of sentiment analysis at document level, proposed by Benamara et al. [8] in the year 2006. This research work began with measuring the intensity of degree of adverbs (using Linguistic Classifiers) and adverb-adjective combinations (using Scoring Methods). Variable Priority Scoring, Adjective Priority Scoring and Adverb First Scoring are the said Scoring methods used herein [8]. The goal of all these methods are nothing but to add a relative weight (in a variable, on a scale of 0 to 1) of score of adverb relative to the score of adjective. This paper aim to determine which weight most closely matches human assignments of opinions. Experimenting on about 200 documents of news resources it shows that analysis that best matches the human sentiments must comprise of 35% of adverbs along with adjectives. Produces Pearson correlation (correlation between human sentiment and Sentiment Analysis Algorithms) and of about 0.47 (ranging in between -1 and 1) [8]. Though this approach shows higher Pearson correlation but considered very few dataset.

One of the solutions to Sentiment Analysis namely Opinion Digger was introduced by Moghaddam and Ester [1]. This unsupervised Machine Learning methodology works at Sentence level. Correlates and compares product aspect and standard rating guidelines (used in Amazon, Snapdeal, flipkart1 etc). This proposed work is divided into two sub methods. At first, input information is fragmented into sentences. Repeated nouns in the sentences are coined as aspects. Aspect (repeated nouns) if forms any pattern, are stored. Secondly, aspects are compared to the rating guideline (like 4 means "Good", 3 means "Average", etc) and accordingly labeled as "Good" ,"Average" and "Bad" [1]. Major advantage is its high performance in product rating at aspect level with a loss of 0.49 only. Demanding guidelines and known data to rate are its major drawbacks and it was compared with very few methodologies. Therefore lacks more number of performance comparisons.

Table.1. Comparison Table of Existing Techniques

Method	Year of proposal	Classification	Text Level	Prediction Accuracy	Pros	Cons
OPINE	2005	Unsupervised rule-based approach	Word	87%	Domain independent	Difficulty in availing OPINE system, thus rare to get applied in real life.
Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone	2006	Linguistic approach	Document	Pearson correlation of 0.47	Adjectives are given more priority(adjectives expresses human sentiments better than adverbs alone)	None
Opinion Digger	2010	Unsupervised machine learning method	Sentence	51%	Rates product at aspect level	Requires rating guidelines to rate. Works only on known data.
Sentiment Classification Using Lexical Contextual Sentence Structure	2011	Rule based approach	Sentence	86%	Said to be domain independent [6]	Depends solely on wordNet
Interdependent Latent Dirichlet Allocation	2011	Probabilistic graphical model	Document	73%	Faster in comparing and correlating sentiment and rating	Correlation between identified clusters and feature or ratings are not explicit always[6]
A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating	2011	Machine Learning	Document	71% (in 3 categories) 46.9% (in 5 categories)	Automatic calculation of feature vector	Use of WordNet

Sentiment Classification from Online Customer reviews Using Lexical Contextual Sentence Structure was proposed by Khan et al. [2] is a semantic or Rule-based (Dictionary Polarity) approach of analyzing customer reviews [2]. Firstly, input is fragmented into sentences and using method "POS" each word is stored. Secondly, based on the context and structure of the sentence polarity of the given sentence is calculated. Nouns are coined as "aspects". Concept of semantic score of words available in SentiWordNet are used to label the sentence as either positive or negative [6]. Accuracy of 86% is produced. Said to be domain independent (subject of review), advantage but the author collected few data (about 3600). Major drawback is it full dependency on WordNet [2].

Interdependent Latent Dirichlet Allocation presented by Moghaddam and Ester [1] is a probabilistic graphical model of rating product at aspect level [6]. Majority of the review sites considers number of stars as the tool to rate a product. This proposed work also does the same assuming interdependency between aspect (feature) and its matching rating. This model tries to generate and showcase cluster head terms into aspects and reviews into ratings in the form of multinomial distributions [10]. Each item in the pool of discrete data is represented as a finite mixture over some latent variables. Found to gain a rating accuracy of about 73%. Since graphical representation suffers from chances of having errors and mistakes in representation of data, this technique might not produce expected output always.

A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating was presented by de Albornoz et al. [11]. This machine learning method rates product at global level considering whole opinion at once. This approach is basically carried out in four steps. At first important features in the document or review are marked. Secondly, sentences containing features (aspects) are identified. Very next, polarity and strength of those sentences are calculated. At last, products are rated globally at aspect level. Feature weights are calculated automatically. Researchers have used the concept of Vector Feature Intensity Graph (VFIG) to represent the reviews [6]. Though use of WordNet is the major disadvantage of this work, it produces an average prediction accuracy of 71% (3 categories) and 46.9% (5 categories) [11].

3. PROPOSED ALGORITHM FOR SENTIMENT ANALYSIS

This section illustrates the proposed algorithm for sentiment analysis. This proposed algorithm is divided into three phases as shown in Fig.2.

- Data Filtration
- Training model
- Testing model

The detailed algorithms of all phases are discussed below. The data filtration flow diagram is given in Fig.3, for training model flow diagram is given in Fig.4 and for testing model flow diagram is given in Fig.5.

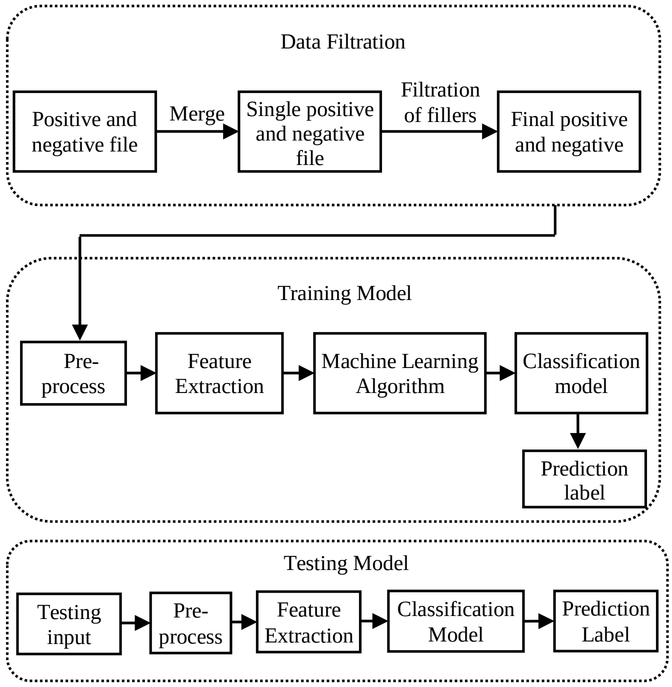

Fig.2. Illustration of the proposed model for sentiment analysis

Data Filtration: Data Filtration importing all positive and negative datasets from file and combining them into a single file. The data sets may contain lots of unwanted symbols, and number. These factors need to be corrected or solved to increase the efficiency. Therefore, in this process the unwanted symbols and number are removed.

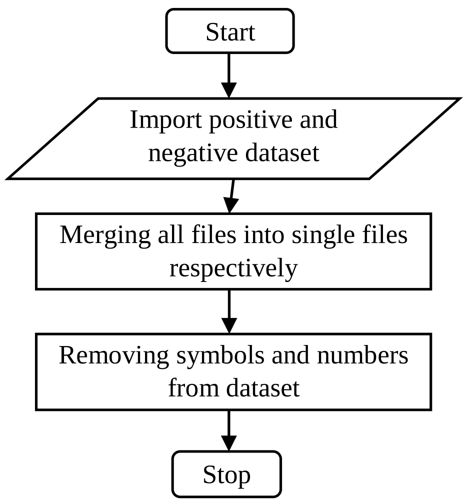

Fig.3. Flow diagram for performing Data Filtration Algorithm

Training Model: Fetching the datasets from the file and extracting all the corresponding words (feature words) like adjective, adverb and verb. Then datasets are labelled a respectively as "pos" for positive and "neg" for negative. Then performing frequency distribution over collected words and selecting 5000 words for training. Again, the shuffling of data is performed using random seed for better training. Here, the labeled datasets are divided into the percentile of 70-30% for training and testing, respectively. Training dataset to classification algorithms like Naïve Bayes classification algorithm [5], Linear Model algorithm [16], SVM algorithm and Decision tree.

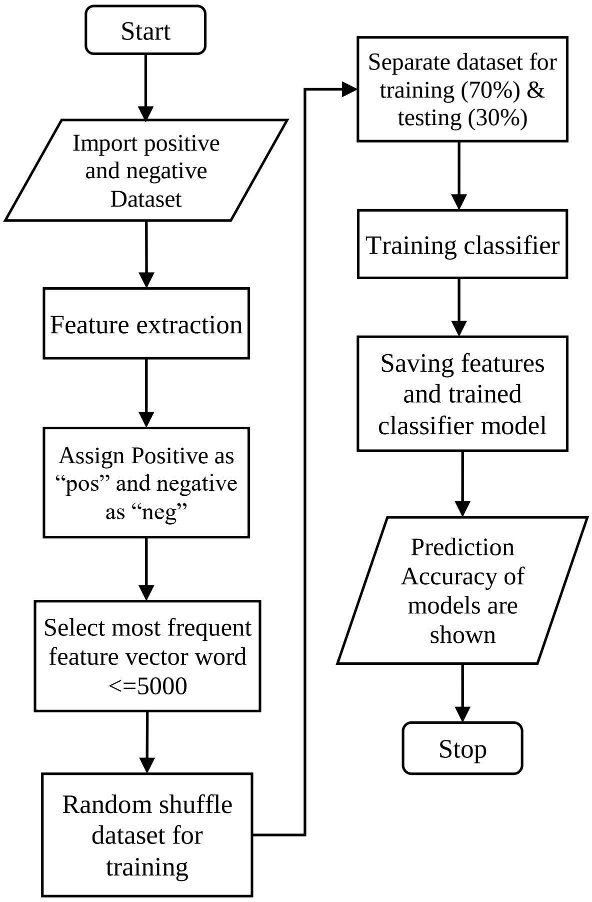

Fig.4. Schematic diagram for implementing machine learning algorithms

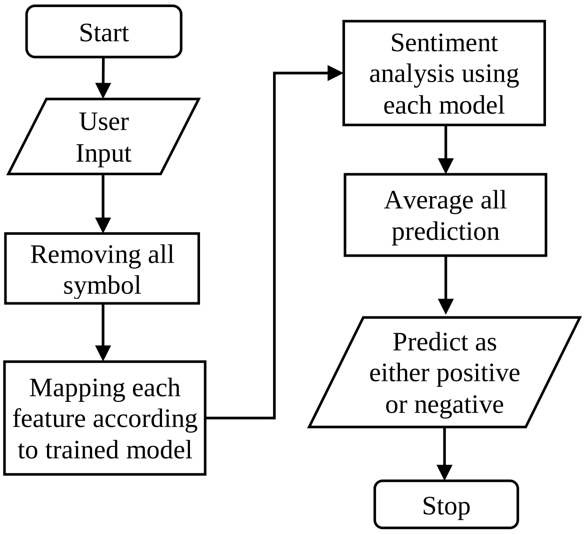

Fig.5. Diagram for testing the proposed model on datasets

Testing Model: Here user can test and analysis the respective model by performing preprocessing over the input data. The preprocessing contains the removal of the symbol and number. Mapping to user input using saved featured (based on training dataset). Then feed to saved model for prediction.

Algorithm 1: Data Filtration Algorithm

Step 1: Importing both positive and negative files and combining them into single file

Step 2: Removal of punctuations and numbers from the dataset

Step 3: Output (Filtered data)

Algorithm 2: Algorithm for Machine Learning Implementation

Step 1: Fetching text paragraph from dataset

Step 2: Feature Extraction phase: Extracting words corresponds to adjective, adverb and verb.

Step 3: All the positive sentences are labeled as "pos" and all the negative ones are labeled as "neg".

Step 4: Most frequent feature vector word is set to 5000 words.

Step 5: Random shuffling the dataset for training

Step 6: Dividing dataset into 70% training and 30% testing dataset

Step 7: Training dataset to classification algorithms like Naïve Bayes classification algorithm [5], Linear Model algorithm [16], SVM algorithm [17]

Step 8: Save the outputs of step 2, and step 7

Step 9: Output (Representation of Accuracy of each model)

Algorithm 3: Proposed algorithm to perform Sentiment Analysis

Step 1: User Input

Step 2: Preprocessing:

- **a.** Removal of " ' " symbol from the text
- **b.** Mapping to user input using saved featured (based on training dataset)

Step 3: Feeding Mapped data to different model for sentiment analysis

Step 4: Output (Averaging all the models)

4. FEATURE EXTRACTION

The dimensionally reduction process of extracting informative and non-redundant values from a given dataset is called Feature Extraction. The Bag of Words model is used for creation of vocabulary after the cleaning up of the 50,000 reviews from the trained set and the frequency of occurrence of each of these words is calculated. The features obtained in this process are used to train the classifier. This action is performed by using sci-kit learns feature extraction module. This module extracts numerical features from the given movie or product reviews which are in text format in the following way:

1) Each string is converted into a unique 'token'.
2) Frequency of occurrence of each of these tokens is calculated.
3) Tokens are organized based on the frequency of occurrences.

With the possibility of obtaining a very large number of features while dealing with 50,000 reviews, one cannot use all the features that are extracted. A certain number of feature vectors need to be selected. Upon testing and experimenting, it was observed that selecting << 5000 or >> 5000 features was resulting in poor prediction accuracy. Therefore, a final array of 50,000 reviews in rows and 5000 features was created.

5. EXPERIMENTAL SETUP

The proposed sentiment analysis algorithm is tested on freely available Stanford dataset for 50,000 movie reviews [15]. The given dataset which comprises of labeled 50,000 movie reviews with half of them being positive and half of them negative. From the given dataset, 70% of the dataset is used for training and other 30% for testing. This process can be easily demonstrated with the help of basic design models given above in Fig.3, Fig.4 and Fig.5 respectively.

Dataset: In order to make the classifier learn and predict it deals with two types of datasets: Training dataset and Test dataset.

Training Dataset**:** It obtains features from the training dataset and forms a classification logic based on the extracted features in order to classify a given test review as a positive or negative review.

Test Dataset**:** The set of data that is used for test of our algorithm is test dataset. The main objective is to feed this test set to our classifier, which then can label accurately given reviews as positive ('pos') or negative ('neg').

Format of the Training Dataset: There are 50,000 reviews in the training dataset out of which 25,000 are positive and 25,000 are negative. The training dataset is available in the format illustrated in Fig.6.

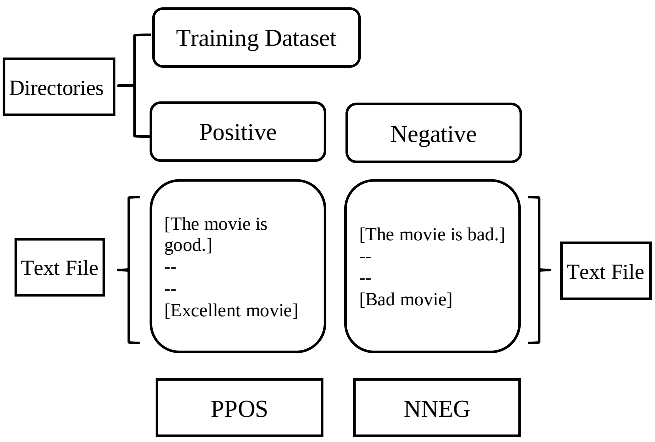

Fig.6. Demonstration of Dataset format used in the proposed algorithm

6. EXPERIMENTAL RESULTS

As it is well known that sentence is the combination of different combination of parts of speech, the different combination produces different accuracy rate. The accuracy of these parts of speech is shown in Table.2 and Fig.7 below with respect to the different types of classifiers. The Figure on various combinations on adverb, adjective and verb and various combinations on datasets has been showcased in Fig.7, Fig.8, Fig.9 and Fig.10. Along with the accuracy rates of different POS the execution time of training and testing dataset has been showcased in Table.3, Table.4 and Fig.7-Fig.10, respectively.

Table.2. Performance Results of different classification models corresponding to different parts of speech (The best case considering each of the speech is styled in bold)

Parts of Speech Considered	Classifier			
	Naïve Bayes	Logistic Regression	Linear SVC	Decision Tree
Adjective	83.81764	**84.29752**	81.92482	82.15142
Verb	80.96507	**81.28499**	78.65902	78.04585
Adverb	79.49880	**81.43161**	79.89869	78.55238
Adjective + Verb	**89.85500**	88.60000	88.75000	87.87500
Adjective + Adverb	**89.85500**	**88.470000**	**88.66000**	83.82500
Verb + Adverb	**89.85500**	**87.255000**	86.95000	86.40500
Adjective + Adverb + Adverb	**89.85500**	**89.575000**	**89.36000**	87.78500

NB → Naïve Bayes, LRC → Logistic Regression Classifier, LSVC → Linear SVC Classifier, D TREE → Decision Tree, Ad → Adverb, A → Adjective and V → Verb.

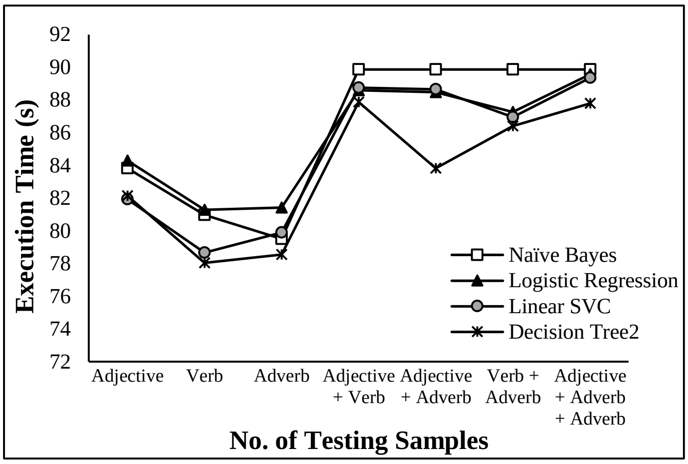

Fig.7. Graphical representation of classification model vs. parts of speech tested on Stanford Dataset [15]

From the above Table.2 and Fig.6 it is clear that prediction accuracy is different for different POS over same datasets and classifiers. Naïve Bayes give an accuracy of 89.855% for POS combination adjective-verb, adjective-adverb, verb-adverb, adjective-verb-adverb which defeats other classifiers thus produced liner line when permutation is performed. LRC works on some logistic functions produces prediction accuracy of 84.29752%, 81.28499%, 81.43161%, 88.47%, 87.255% and 89.575% for reviews that contains POS combination of adjective, verb, adverb, adjective-adverb, verb-adverb, adjective-verb-adverb respectively which is more impressive than other classifiers considered in. LSVC produces prediction accuracy of 88.66% and 88.36% for review containing adjective-adverb and adjective-verb-adverb respectively. Adjective-verb-adverb gives better result than any other combination of adjective adverb verb. Thus the new approach is not only competent enough but also promises to be more efficient than the existing methods [2] (86%, on an average in Table.1). The Fig.7-Fig.10 on various combination on adverb adjective and verb and various combination on datasets.

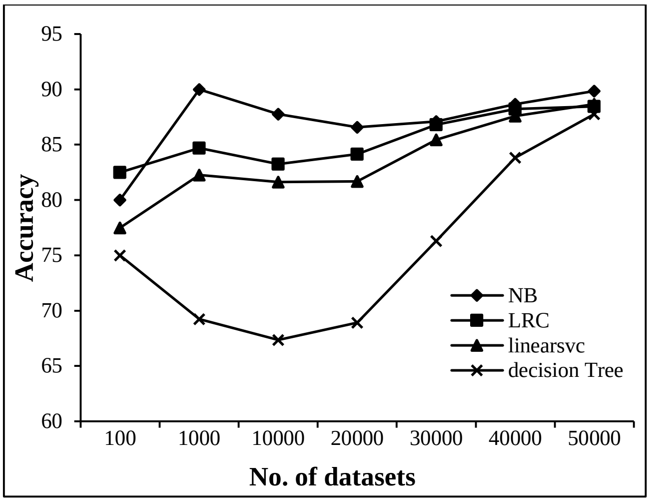

Fig.8. Adjective adverb

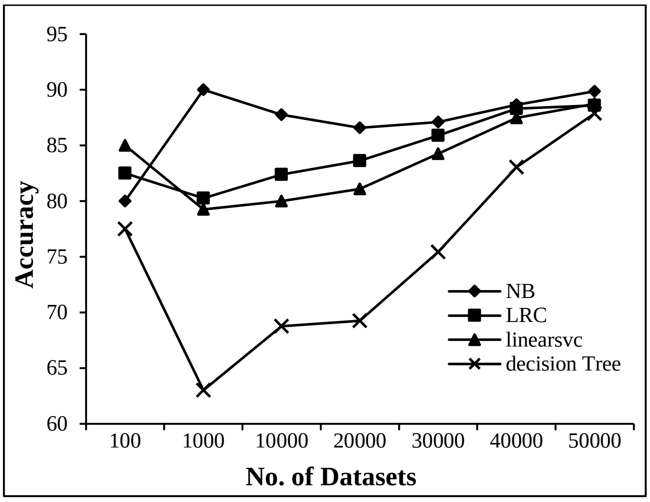

Fig.9. Adjective verb

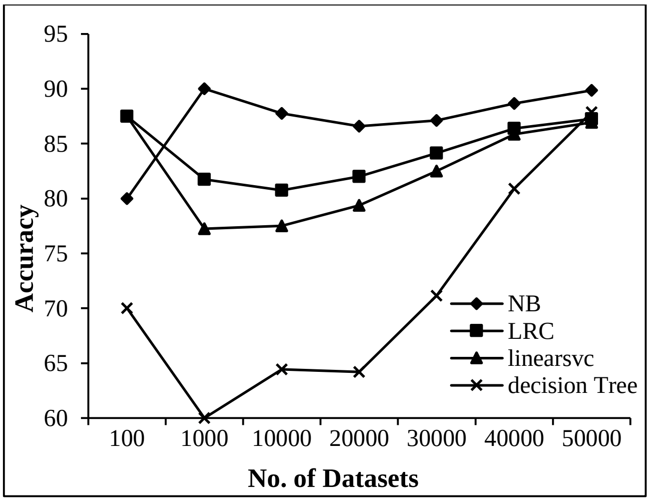

Fig.10. Verb adverb

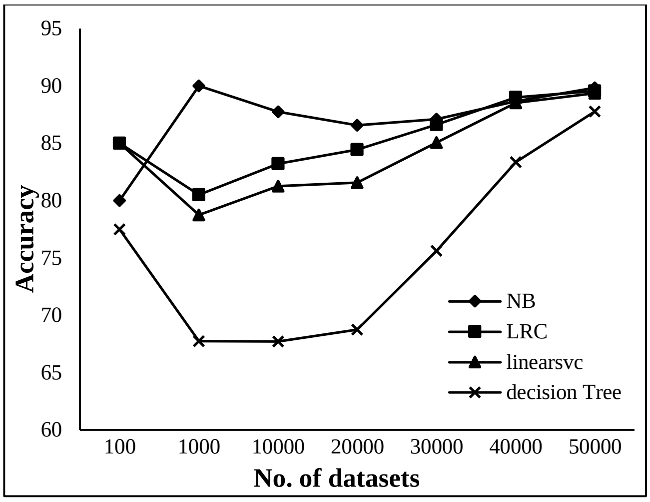

Fig.11. Adjective verb adverb

The Fig.8-Fig.11 mention above represents the accuracy of the various numbers of datasets and also represent how the accuracy fluctuates over the number of datasets.

Table.3. Tabular representation of execution time (in seconds) of training dataset corresponding to each classifier

Representation of Execution Time				
Datasets	Naïve Bayes	Logistic Regression	Linear SVC	D-Tree
50000	26.7274	13.7237	17.4956	50.64934

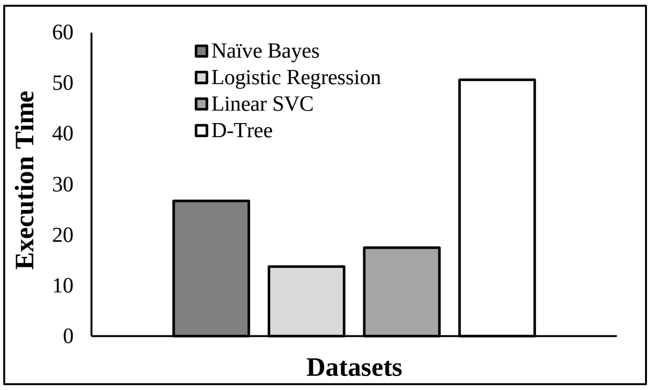

Fig.12. Graphical representation of execution time for training the dataset w.r.t different classifiers

From the above Fig.12, graphical representation of execution time for training the dataset with respect to different classifiers over the same datasets.

Table.4. Tabular representation of execution time (s) for testing dataset by different classifiers

Datasets	Naïve Bayes	Logistic Regression	Decision Tree	Linear SVC
1	0.0263	0.0698	0.0368	0.0254
2	0.0662	0.0551	0.0556	0.0543
3	0.0807	0.07022	0.1129	0.07233
4	0.1913	0.1188	0.1917	0.0972

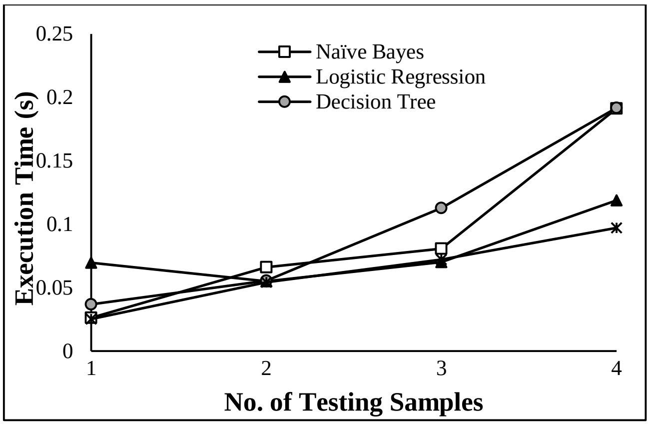

Fig.13. Graphical representation of execution time of testing dataset w.r.t different classifiers

The Table.3 and Table.4 followed by Fig.11 and Fig.12 of execution time of training and testing the dataset signifies the efficiency of the proposed approach. LRC takes about 13.7237s to train 50000 dataset that is the least time taken among all other classifiers considered here while D-Tree takes 50.64934s of time and is maximum among all. Linear SVC takes about 0.0972s to test the four samples at once and provide output which is much lesser and thus better than other classifiers considered herein. This new approach not only improvises the way of analyzing sentiments with better accuracy rate but also promises to take lesser time to train as well as to test the opinion than other existing methods.

7. CONCLUSION

Key consideration of this newly proposed technique is part of speech and tested on benchmark Standford Dataset [15] using six well-known supervised classifiers. It is noticed that the combination of adjective, adverb and verb turned out to be the best combination among various combinations of the parts of speech.

REFERENCES

[1] Samaneh Moghaddam and Martin Ester, "Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews", *Proceedings of 19th ACM International Conference on Information and Knowledge Management*, pp. 1825-1828, 2010.

[2] Aurangzeb Khan, Baharum Baharudin and Khairullah Khan, "Sentiment Classification from Online Customer Reviews using Lexical Contextual Sentence Structure", *Proceedings of International Conference on Software Engineering and Computer Systems*, pp. 317-331, 2011.

[3] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", *Proceedings of 10th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 166-177, 2005.

[4] A.M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews", *Proceedings of International Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 339-346, 2005.

[5] G. Vinodhini and R.M. Chandrasekaram, "Sentiment Analysis and opinion Mining: A Survey", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2, No. 6, pp. 28-35, 2012.

[6] A. Collomb, C. Costea, D. Joyeux, O. Hasan and L. Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation", Available at: https://liris.cnrs.fr/Documents/Liris-6508.pdf.

[7] Mika V. Mantyla, Daniel Graziotin and Miikka Kuutila, "The Evolution of Sentiment Analysis-A Review of Research Topics", *Computer Science Review*, Vol. 27, No. 1, pp. 16-32, 2018.

[8] F. Benamara, C. Cesarano and D. Reforgiato, "Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone", *Proceedings of International Conference on Weblogs and Social Media*, pp. 1-7, 2006.

[9] R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes", *Proceedings of International Conference on Computer Vision: Issues, Problems, Principles, and Paradigms*, pp. 585-605, 1987.

[10] Samaneh Moghaddam and Martin Ester, "ILDA: Interdependent LDA Model for Learning Latent Aspects and their Ratings from Online Product Reviews", *Proceedings of 34th International ACM Conference on Research and Development in Information Retrieval*, pp. 665-674, 2011.

[11] Jorge Carrillo De Albornoz, Laura Plaza, Pablo Gervas and Alberto Diaz, "A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating", *Proceedings of International Conference on Advances in Information Retrieval*, pp. 55-66, 2011.

[12] Sentiment Analysis, Available at: https://insightsatlas.com/sentiment-analysis/

[13] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts, "Learning Word Vectors for Sentiment Analysis", *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics*, pp. 1-7, 2011.

[14] Understanding Sentiment Analysis: What It Is and Why It's Used Understanding Sentiment Analysis: What It Is and Why It's Used, Available at: https://www.brandwatch.com/blog/understanding-sentiment-analysis/

[15] Sentiment Analysis Explained, Available at: https://www.lexalytics.com/technology/sentiment-analysis