Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews

Tahura Shaikh[#1], Dr. Deepa Deshpande [#2]

[#1] *Department of Computer Science, JNEC College of Engineering, Aurangabad – 431001, India*

[#2] *Assistant Professor, Department of Computer Science, JNEC College of Engineering, Aurangabad – 431001, India*

Abstract — *Sentiment Analysis or Opinion Mining is a nascent field of data mining, which is expanding and much research work is being done in this field. Opinion Mining mines people's opinion towards a topic. Opinion mining's main objective is to extract opinion or views of a person for a particular topic or subject. Mainly Opinion Mining classifies the given review as positive, neutral or negative. Opinion Mining has accomplished much focus nowadays due to availability of vast amount of opinion rich web resources such as online product reviews, blogs, social networking sites etc. As the use of e-commerce websites are increasing profusely and people are opting for online shopping there is vast amount of data generated which can be useful for Opinion Mining.*

In this paper, different feature extraction or selection techniques for opinion mining are performed. Work is carried out in different steps. First step is the data collection step in which amazon dataset is used. Second is the pre-processing step which is used for the removal of stop words and special characters. In the third step, feature selection or extraction techniques like phrase level, single word and multiword are applied over the amazon dataset. The fourth step is used to generate the vector of the extracted features. In the final step, Naïve Bayes classifier is applied to classify the reviews. Step one to four is used for training the system and last step is used for testing. In the paper Supervised learning method is used for classification of reviews.

Keywords — *Opinion Mining, Sentiment Analysis, User Reviews, Feature Extraction, Classification, Naïve Bayes*

I. INTRODUCTION

With the explosive growth of social networking sites, blogs, forums etc., lot of useful information is been generated which forms a source for opinion mining. Nowadays many e-commerce websites have advanced and people are opting for online shopping due to high discounts, vast choice of products etc. Buyer can compare the product with other products beforehand; they can look into opinion provided by users who already bought the item. Thus the user reviews give important information of the product quality, price etc. which can be useful for the buyer to take ethical decision. The main aim of sentiment analysis is polarity classification. Also the reviews give insight to the manufacturers about the flaws in their product which can be helpful for the product preferment. Sentiment classification is a sub discipline of text classification which is concerned with the opinion a topic expresses [11]. Sentiment analysis also has different names, among which are opinion mining, sentiment analysis, sentiment extraction, or affective rating. Sentiment analysis is performed to find the semantic orientation of the given review or comment [3]. Sentiment analyses consider two types of information i.e. facts and opinions. Facts are objective in nature describing the nature of a product or event. The majority of researches done on objective nature of the product but recent trend are to focus on opinions. There are a number of challenges in opinion mining. Word based challenge is the first kind of a challenge, as sometimes it is difficult to understand the emotion of a given word, same word can have positive meaning in one statement but negative meaning in other statement.

Proposed method is evaluated using freely available amazon dataset for feature selection and sentiment classification. Different steps are performed which are data collection were the available amazon dataset is being used, next pre-processing step is applied which filters the reviews by removing the stop words and special characters which are not required for further processing. Feature selection for phrase level, single word and multi word is performed, vector generation is the next step which generates the vector against the positive and negative class labels and finally sentiment classification is executed using Naïve Bayes algorithm. Performance analysis is done for classification. In section 2 the related work is given. Section 3 gives detail about the different steps which are performed or the proposed framework. Section 4 gives the experimental result and section 5 gives the conclusion and further work which can be performed in this field.

II. Related Work

J. Ashok Kumar and S. Abirami [1] implemented the OMSA approach and analysed the results by using a single dataset for different feature extraction or selection techniques namely single word, Multiword, Document Level, Phrase Level, Tf-idf single word and Tf-idf Multiword. Experimental procedure has been carried out with an extension of the OMSA approach [2]. In this approach, the Polarity Classification Algorithm (PCA) and evaluation procedure is applied to verify the accuracy. The evaluation procedure is tested with four different datasets.

Gurneet Kaur, Abhinash Sinha [2] studied efficacy of classifying product reviews by semantic meaning. They also presented the fundamentals of opinion mining and also pros and cons of past opinion mining systems. The authors proposed different approaches including spelling correction in review text and classification of comments. They employed hybrid algorithm combining Naïve Bayes and Decision Tree. For their work they made use of Flipkart comments for MOTO X play phones. Naive Byes classifier is used by them to classify positive and negative words and decision tree used to calculate overall polarity.

Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande [4] proposed a system that performs the classification of customer reviews followed by finding sentiment of the reviews. Rule based extraction of product feature sentiment is also done in their work. They tested their work for iphone 5 reviews from amazon dataset.

Jeevanandam Jotheeswaran and S. Koteeswaran[5] used the method which used twitter dataset for evaluation. Their system is applicable to product reviews, emotion detection, knowledge transformation and predictive analysis. It gives the result as decision forest based feature extraction improves precision of classifier when compared with decision tree based feature selection. They studied the movie review features obtained from Twitter using inverse document frequency and the importance of the word found. Principal component analysis was used for feature selection based on the importance of the work with respect to the entire document. They concluded from the experimental results that the LVQ classifier performs better than the CART and Naïve Bayes classifiers. And the proposed decision forest based feature selection improves the efficiency of the classifiers.

Hu and Liu [6] made use of distance based approach for extraction of opinion words and phrases after extracting aspects. They used WordNet for calculating the polarity of each extracted opinion word .Set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods was proposed by them . Their objective was to provide a feature-based summary of a large number of customer reviews of a product sold online. Summarizing the reviews is not only useful to common shoppers, but also crucial to product manufacturers.

J. Ashok Kumar, S. Abirami and S. Murugappan [7] presented the OMSA approach with different frameworks and algorithms as a review and their results were compared and analysed for readily available datasets. In this paper, they presented the recent role of OMSA in Social Networks with different frameworks such as data collection process, text pre-processing, classification algorithms, and performance evaluation results. The achieved accuracy level is compared and shown for different frameworks in the paper. A brief description about different developments in OMSA is given in this paper.

Su Su Htay and Khin Thidar Lynn [8] proposed a novel idea to find opinion words or phrases for each feature from customer reviews in an efficient way. Focus in this paper is to get the patterns of opinion words/phrases about the feature of product from the review text through adjective, adverb, verb, and noun. The extracted features and opinions are useful for generating a meaningful summary that can provide significant informative resource to help the user as well as merchants to track the most suitable choice of product. We use a part-of-speech tagger to identify phrases in the input text that contains adjective or adverb or verb or nouns as opinion phrases.

Gayathri R Krishna, Kavitha S, Yamini S, Rekha A[11]discussed and analysed different opinion mining algorithms like LDA, sLDA, NMF, SSNMF, DiscLDA and PAAM.

Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori [9] Addresses the aspect-based summarization task by introducing a novel methodology for retrieving product features from a collection of free-text customer reviews about a product or service. The proposal relies on a language modeling framework that combines a probabilistic model of opinion words and a stochastic mapping model between words to approximate a language model of products. Their work extends a preliminary approach introduced that addresses the modeling of a language of product features from customer reviews.They propose a more general methodology that effectively allows, for example, the use of grammatical dependency relations between words in modeling the language of features. They also provide a more formalized methodology for the retrieval of (multiword) product features from the estimated language model of features, along with a more comprehensive evaluation.

III. METHODOLOGY

Work consists of first collecting the reviews from Amazon dataset. The dataset which we have used is divided into different categories. Each category consists of Positive, Negative and Unlabelled reviews. For training; the system is given known positive and negative reviews. Pre-processing method is done which consist of removing the stop words. Then the features are extracted using phrase level and POS (Parts Of Speech) i.e. Single word and Multiword methods. After feature selection/ extraction is completed vector is generated. The vector is then used for training the system. In our model we have not used any dictionary instead we have generated the vector of extracted features and used it as a dictionary to classify the unlabelled reviews. Then Naïve Bayes algorithm is used for classification. Feature selection techniques which are used are Phrase level, Single word and multiword. Workings of the different steps are divided as A) Feature Selection/Extraction and B) Sentiment Classification.

A. Feature Selection/Extraction

I) Phrase Level:

The Phrase level feature selection/extraction technique is used to extract the expression which is useful or important from the given review. This method makes use of suffix array. Array of all suffixes for a given string is the Suffix Array. Suffix array comes handy if there is a large collection of documents and we need to find whether collection contains documents with any related phrase despite the phrase size or when we need to compute ngram frequencies for large phrases. Suffix array works like, whenever a linear scan is done for collection of words it gives number of times the required word appears in the collection, despite the size of the word. Our work gives the score of the unique phrase which appears in all the reviews we have selected. Feature selection gives the result as the important features from the given phrases which are required to complete the meaning of the phrase. The phrase level method in our work makes use of suffix array, LCP, index, rank etc.

LCP stands for Longest Common Prefix and is an auxiliary or supporting data structure to the Suffix Array. The elements on the stack are lcp-intervals represented by tuples where lcp is the lcp-value of the interval [3].Calculation of LCP is useful when there is a need to speed up the search.

It is also useful when we want to know which the longest string which is repeated is or when there is a need to find ngram counts. After the suffix array and the LCP (longest common prefix) array are computed LCP array is scanned and the LCP array are given Rank based on the number of prefixes.

The Review with rank one is the most distinctive. The phrases which are extracted are arranged ascendingly based on the given ranks. In sorted suffix array the length of prefix common between any two consecutive suffixes is given by LCP.

Figure 1 Phrase Level Feature Extraction

II) Single Word:

Single Word is a unigram model. It uses POS (Parts Of Speech) tagging [1]. Single words or unigram from the given reviews are extracted. We have used wsj left3words model of Stanford Tagger. The code is restricted to extract unigram or single words which are Nouns from the given set of reviews. The WSJ corpus contains a million words published in the Wall Street Journal in 1989[4]. Positive and Negative reviews are given as input to the system, it selects the required features and gives the result. Features selected are the common nouns from the given reviews. For common nouns 'N' is the simplified noun tag i.e.: monograms are extracted from given reviews.

It then constructs feature vector. Vector is a matrix consisting of the features extracted and it holds value 1 for positive and 0 for negative. We have trained the system by providing the positive and negative reviews.

III) Multi Word:

Multi Word is a bigram model. It uses POS (Parts Of Speech) tagging [1]. Multi words or bigrams from the given reviews are extracted. We have used wsj left3words model of Stanford Tagger, same as single word but multigram extracts binary words which are Nouns from the given set of reviews. Negative and Positive reviews are given as input to the system, it selects the required features and gives the result. Features selected are the common nouns from the given reviews. For common nouns 'N' is the simplified noun tag i.e.: bigrams are extracted from reviews which are the features. It then constructs feature vector. Vector is a matrix consisting of the features extracted and it has value 1 for positive and 0 for negative. Our system is trained by providing positive and negative reviews.

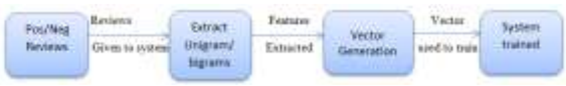

Figure 2 Unigram/Bigram Feature Extraction

B. Feature Selection/Extraction

After vector generation phase is completed, sentiment classification can be performed. The vector generated consists of all the extracted features with their respective values 0 (negative) and 1(positive), which is useful to train the system. We have considered two Machine Learning Classification algorithms for sentiment classification which are Naïve Bayes.

Naïve Bayes algorithm is a text classification approach which assigns class to the given documents [2]. We have classified the reviews using the vector which has been generated after selection of the features using phrase level, single word and multiword. Thus training data is generated where data which is trained is the probability of occurrence of train data files in positive and negative class. Naïve Bayes classifier does simple classification based on Bayes approach. It is Bag of Words (BOW) which is collection of words. Naïve Bayes Classifier makes use of Bayes Theorem for class 'c' and document'd':

$$P(c/d) = P(d/c)\,P(c) / P(d)$$

Class c is assigned to the document d using Naïve Bayes algorithm: c*= argmaxc P (c/d). All the documents are represented as feature vector , in our work the documents are the reviews .When we select the reviews for classification it will probabilistically check all the reviews for all the features present in vector against their class and will assign the review to that class in which its feature match the most. If more matches are found for positive class then that review will be assigned to positive class and will be classified as positive review or else it will be classified as negative review. The working is as first we calculate the probability of label P (label= X) then calculation of probability for each feature is done and result is given. The label is the class; here it is positive and negative which has the value either 1 or 0.We have list of features against their label.

C. Algorithm for feature extraction and sentiment classification

1. Procedure Review check()
2. begin
3. for each positive and Negative reviews remove stopwords;
4. for each review r;
5. begin
6. check c=0.5;
7. extract features f;
8. for each feature f value is (0,1);
9. value 0= Negative, Value 1= Positve;
10. vector generated for all f's against(0,1);
11. system trained;
12. if c<0.5, check=negative;
13. else if (c>0.5) check=positive;
14 end for;
15. end;

1. Procedure check(word w, review r)
2. begin
3. check= orientation of review;
4. select classification by Naive Bayes
5. select method for training system p,s or m;
6. if method selected p
7. begin
8. check= all review word w against all the features f against classes positive and negative;
9. If (positive word w appears to close for certain features f from vector)
10. check= assign review to class

End

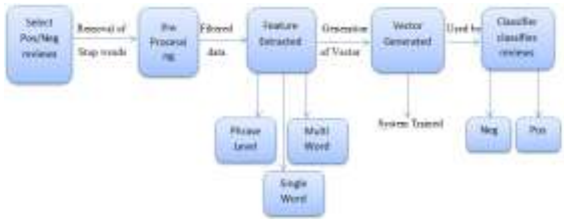

Figure 3 System Flow

IV. EXPERIMENT AND RESULT ANALYSIS

A. Dataset and evaluation measure

We have performed experiments using reviews from different categories of amazon dataset. The dataset consisted of different categories like books, electronics, music, camera and magazines.

Table 1 Size of dataset for different categories

Categories	Sentiment	Number Of Reviews
Books	Positive	130
	Negative	130
Camera	Positive	130
	Negative	130
Magazines	Positive	130
	Negative	130
Electronics	Positive	130
	Negative	130
Video	Positive	130
	Negative	130

Each category consisted of positive, negative and unlabelled reviews. The unlabelled reviews where classified using Naïve Bayes algorithm. 130 Positive and Negative Reviews were used to train the system and 25 known positive and Negative reviews were used for classification from which evaluation measures were calculated.

For classification, True Positives (TP), True Negatives (TN), False Negatives (FN) and False

Positives (FP) are used to compare the class labels (Positive and Negative) assigned to documents by a classifier with the classes in which the items truly belongs . True positive are terms truly classified as the positive. True Negative is the ones truly classified as Negative. Evaluation measures like precision, recall, F-measure, specificity and accuracy are easily calculated once we get the confusion matrix consisting of TP, TN, FP, and FN. Different evaluation measures we used for study are:

1. Confusion Matrix: Is the matrix which describes the performance of a classifier on test data for which the true values or actual values are known. Confusion matrix is helpful in interpreting the accuracy of the result for the given classification problem.

Table 2 Confusion Matrix

		Actual Value	
	Classes	Positive	Negative
	Positive	TP	FP
Result	Negative	TN	FN

2. Precision: It gives the fraction of retrieved instances that are relevant. [8].

$$\text{Precision} = \frac{\text{Number of correct predictions}}{\text{Number of predictions}}$$

$$\text{Precision} = TP / (TP + FP)$$

3. Recall: Fraction of relevant instances retrieved.
Recall=Number of correct predictions / Number of examples

$$\text{Recall} = TP / (TP + FN)$$

4. Accuracy: Accuracy is one of the evaluation measures which is helpful in evaluating the performance of the classifier or to find error rate. Accuracy is the fraction of correctly classified examples to total number of examples [2]. Error rate considers the incorrectly classified examples.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

5. F-Measure: Harmonic mean of precision and recall is calculated using F-Measure. The result is a score which is balanced between recall and precision [2]. The formula used to calculate F-Measure is as :

$$\text{F-Measure} = \frac{2(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

B. Experimental Results

We have trained the system using positive and negative reviews. To test the system for Naïve Bayes classifiers the next 25 reviews are used. Different reviews are used for training and testing. By selecting the classifier we can test the system for phrase level, single word and multi word feature selection methods. Results are given for evaluation measures Precision, Recall, Accuracy and F-Measure for Naïve Bayes algorithm against feature selection method. We have performed the experiment using the categories books, music and camera. Below is category wise results calculated for Naïve Bayes algorithm.

Table 3 Books Category

	Precision	Recall	Accuracy	F-Measure
Phrase Level	0.692	0.72	0.7	0.353
Single Word	0.666	0.64	0.7	0.652
Multi-word	1	0.6	0.8	0.75

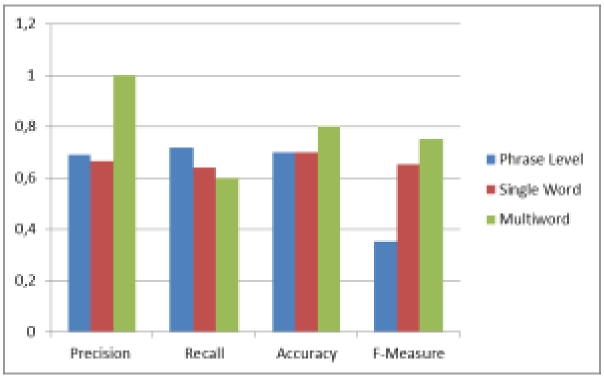

Figure 4 Results for Books Category

Table 4 Music Category

	Precision	Recall	Accuracy	F-Measure
Phrase Level	0.607	0.68	0.62	0.641
Single Word	0.77	0.84	0.8	0.803
Multi-word	0.714	0.6	0.68	0.651

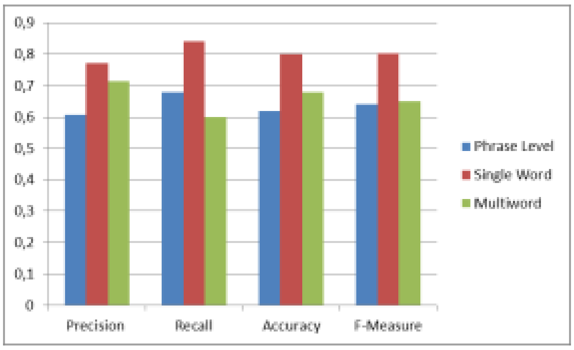

Figure 5 Results for Music Category

Table 5 Camera Category

	Precision	Recall	Accuracy	F-Measure
Phrase Level	0.607	0.68	0.62	0.641
Single Word	0.77	0.84	0.8	0.803
Multi-word	0.714	0.6	0.68	0.651

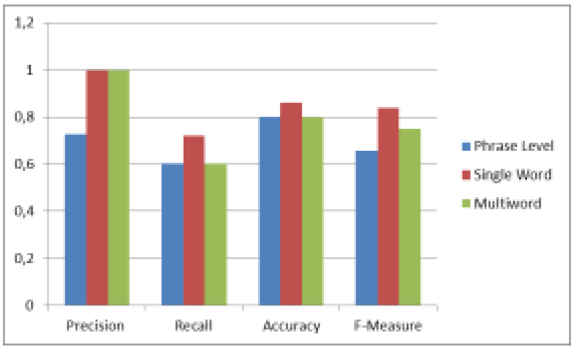

Figure 6 Results for Camera Category

V. CONCLUSIONS

From above experimental work we conclude that Naïve Bayes algorithm gives better result for phrase level feature extraction method compared to other two. This accuracy in phrase level is due to the implementation of suffix array method. Naïve Bayes classifier works well even if the dataset used for training consists of less data and tested on large dataset. The main advantage of using Naïve Bayes is that it is easy to implement. In this paper we have extracted the features using phrase level, single word and multi word methods and classification of reviews is exhibited using Naïve Bayes algorithm and results are analyzed.

REFERENCES

[1] J. Ashok Kumar and S. Abirami, "An Experimental Study Of Feature Extraction Techniques In Opinion Mining," International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.4, No.1, February 2015.

[2] GurneetKaur and AbhinashSingla, “Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm,” International Journal Of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 5, ISSN: 2278- 1323, January 2016.

[3] Mohamed Ibrahim Abouelhoda , Stefan Kurta and Enno Ohlebusch,”Replacing suffix trees with enhanced suffix arrays”, 1570-8667, 2004.

[4] Aashutosh Bhatt, Ankit Patel, Harsh Chheda and Kiran Gawande, "Amazon Review Classification and Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 6, ISSN:0975-9646, 2015.

[5] Jeevanandam Jotheeswaran and S. Koteeswaran,“Feature Selection using Random Forest method for Sentiment Analysis,” Indian Journal of Science and Technology, vol.9. ISSN : 0974-5645, Jan. 2016.

[6] SuSuHtay and KhinThidarLynn , “Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Review,” Hindawi Publishing Corporation TheScientific World Journal, Vol.2013.

[7] Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori, “Retrieving Product Features and Opinions from Customer Reviews”, IEEE Intelligent Systems, vol. 28, pp. 1541-1672, May/Jun 2013.

[8] Tina R. Patil and Mrs. S. S. Sherekar, “Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification,” International Journal Of Computer Science And Applications, vol. 6, ISSN: 0974-1011, Apr 2013.

[9] Dim En Nyaung, Thin Lai Lai Thein, "Feature Based Summarizing From Customer Reviews", International Journal Of Scientific Engineering and Technology Research, vol.03, ISSN 2319-8885, Dec-2014.

[10] Bing Liu, Sentiment analysis and opinion mining, Morgan & Claypool Publishers, 2012

[11] R Krishna, Kavitha S, Yamini S,Rekha A,“Analysis of Various Opinion Mining Algorithms Gayathri,” International Journal of Computer Trends and Technology, vol. 22 , Number 2, ISSN: 2231-2803 ,April 2015.