

# Opinion Classification from Online Reviews based on Support Vector Machine

Sangram Ashok Patil<sup>\*1</sup>, Prof. K. B. Manwade<sup>2</sup>

<sup>1</sup>ME (CSE ) Student, Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India

<sup>2</sup>Asst. Prof., Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India

## ABSTRACT

With the rapid growth of internet, a huge number of product reviews are evolved up on the Web. From these reviews, customers can get direct assessments of product information and direct supervision of their purchase actions. Meanwhile, manufacturers can obtain immediate advice and opportunities to improve the quality of their products in a timely fashion. Any customers who buy a product can place their opinion about product features. New customers who want to buy a product he reads the reviews of the previous customers. So reviews are helps to new customers to decide to buy a product or not. Product manufacturer are also use the reviews of customer for product development. Opinions of customers are sentiments about product. Opinions given by customers are not in specific format and it does not have any syntax to define it. Reviews may be in single word or may be in one sentence or it may be in multiline format. And every customer has different opinions about product so that new customer need to read all the reviews this process is difficult and time consuming process. Here we have implement a system for online reviews classification based on polarity by using support vector machine and provide a review based rating system. From these online reviews we also find opinion target and opinion word using word alignment model and shows the topical relation.

**Keywords:** Feature Extraction, Opinion target and word, Text Classification

## I. INTRODUCTION

Now a day's due to the rapid growth of ecommerce and increase in online merchants online shopping dramatically increases. Due to the number of online merchant sites large number of user's increases. This number of customers which purchase online products increases rapidly. To improve the customer fulfillment, product manufacturers and merchants allow customers to review or express their feelings or opinion about the product or services. The customers can now place a review for products at merchant sites. These online customer reviews, thereafter, become an important source of information which can be useful for both potential customers and product manufacturers. New customers will use this information to support their decision on whether to purchase the product or not. For product manufacturer customers feedback is more important and

valuable for product development and marketing. Since feedbacks of customer affects decision of other customer's decision, the review documents have become an important source of information for business organizations and useful to take development plans. Opinion Mining deals with the broad area of natural language processing and text mining involving the computational study of opinions, sentiments and emotions expressed in text. Hence, an alternate term for Opinion Mining is also called as Sentiment Analysis. Opinion mining has many application including science and technology, entertainment, education, politics, marketing, accounting, law, research and development. But with the tremendous growth of the World Wide Web, has large number of text opinions in the form of blogs, reviews, discussion groups are available for analysis making the World Wide Web the fastest, most broad and easily accessible medium for sentiment analysis.

## II. METHODS AND MATERIAL

### A. Literature Review

1. According to **Hu and Liu, Ding et al., Li et al**[1],[2], opinion target extraction can be divided into two main categories: supervised and unsupervised methods. In supervised approaches, the opinion target extraction task was usually regarded as a sequence labeling task where several classical methods are used such as CRFs.
2. According to **Li, Jin, Huang, Ma and Wan Wu et al**[3], [4], et al., the main limitation of supervised and unsupervised methods is that labeling training data for each domain is time consuming and impracticable. In unsupervised methods, approaches regarded opinion words as the important indicators for opinion targets.
3. **Hu, Liu et al** [5], exploited an association (Nearest Neighbor rule) mining rule algorithm and frequency information to extract relation among words.
4. **Qiu et al** [6], proposed a Double Propagation method to expand sentiment words and opinion targets iteratively, where they also exploited syntactic relations between words. The main limitation of Qiu's method is that the patterns based on dependency parsing tree may introduce many noises for the large corpora. Besides the patterns used in Qiu's method, they adopted some other special designed patterns to increase recall.
5. **Lui et al**[5], employed word alignment model to capture opinion relations rather than syntactic parsing. These alignment methods are more effective than syntax based approaches for online informal texts.
6. The Naive Bayes algorithm [8] is generally used algorithm for document classification. Naive Bayes is a simple and effective classification algorithm. The basic idea is to find probabilities of categories for given a test document by using the joint probabilities of words and categories. The naïve part of model is the assumption of word independence. The simplicity makes the computation of Naive Bayes classifier far more efficient.
7. The K-nearest neighbor (KNN) [9] is example based classifier that does not build an explicit, declarative representation of the category. For given a test document  $d$ , then among training

documents system finds the  $k$  nearest neighbors. The resemblance score of each nearest neighbor document to the test document is used as the weight of the classes of the neighbor document (**Songho tan**, 2008). Besides this classifier other classifiers like ID3 and C5 are also investigated (**Rudy Prabowo**, 2009).

8. **Rui Xia, Ziqiong, Songho tan, Rudy Prabowo**, [7] [8] et.al. proposed a Support vector machines (SVM). SVM is a discriminative classifier is considered the best text classification method. The support vector machine is a statistical classification method. SVM seeks a decision surface to separate the training data points into two classes and makes decisions based on the support vectors that are selected as the only effective elements in the training set. Multiple variants of SVM have been developed in which Multi class SVM is used for Sentiment.

### B. Proposed Work

Opinion mining is viewpoint of text mining. To conclude some decision from the reviews or the feedbacks retrieved from the user customers opinions are generally taken to collect the thoughts and various aspects of the product that also include the technical details. Once the opinions are collected the filtration over reviews are performed to identify most relevant information in the form of opinion class or sentiment class.

The proposed work has following objectives

- ✓ To design a system to perform feature extraction from user reviews and relevant feature selection from multi sentence reviews.
- ✓ To implement opinion classification based on polarity.
- ✓ To implement rating system based on data which is classified in different polarities.
- ✓ To design and implement a system which construct topical relation between sentences of reviews

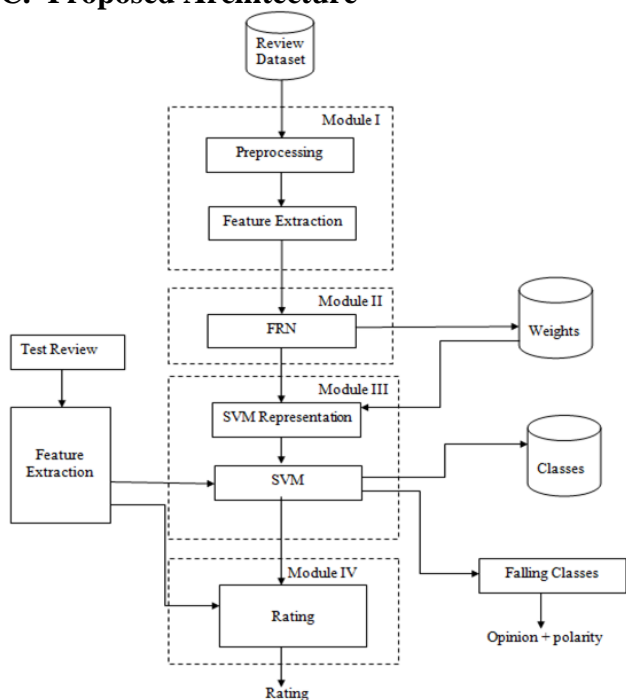
Following techniques are used for proposed work

- ✓ Feature Relational Network (FRN) algorithm can be used to for feature extraction. FRN is based on N-gram technique which is used for separating

text in different grams such as N-char or N-words. These N-grams are used for feature extraction and to find association that will be used in classification further.

- ✓ The Support Vector Machine is discriminative (SVM) classifier used for text classification. SVM is based on decision planes that define decision boundaries. Decision plane separates set of objects having different class membership. When SVM is used with FRN it gives good performance
- ✓ Calculating the rating for product based on the reviews.
- ✓ The Word alignment model used to find the opinion target and words in the reviews. We use here phrase based technique which use noun verb phrases in the review.

### C. Proposed Architecture



**Figure 1:** System Architecture

The proposed architecture consists of four modules.

#### 1. Feature Extraction Module

Feature Extraction Module take review dataset as input. The dataset contains the product reviews which are given by customer on merchant’s site. This review dataset is used for preprocessing to remove noisy, erroneous, and inconsistent data. This preprocessed data is then used for feature extraction. All text features are extracted using n-gram feature set. The n-gram feature set using character level n-gram, word level n-gram, and part of speech tag n-grams and so

on. Character N-gram based on set of characters. Word level N-gram is based on group of words. It extracts all the feature text from reviews. Suppose the sentence is like “I love this smart phone”

N-gram Category	Examples
1-char	I, l, o, v, e, t, h, i, s, s, m, a, r, t, p, h, o, n, e
2-char	Il, ov, et, ,hi, ss ma, rt, ph
1-word	I, love, this, smart, phone
2-word	I love, this smart

**Table 1 :** N-gram Feature set

#### 2. Feature Selection Module

Text features selected by Feature Extraction Module are used by Feature Relational Network (FRN). FRN algorithm is used to select text feature considering relevance between features and redundancy factors. It uses a rule based multivariate text feature selection method that considers semantic information and syntactic relationship between n-grams features in order to efficiently remove redundant and irrelevant features. FRN utilizes subsumption relation which enables intelligent comparison between features to facilitate removal of redundant and irrelevant features. FRN also assigns numeric values called as weight to the features.

Feature group	Subsumption Relation
N-Char	1-char -> 2-char, 1-char -> 3-char, 2-char -> 3-char,
N-Word	1-word->2-word, 1-word->3-word, 2-word->3-word

**Table 2 :** Subsumption relation

#### 3. Feature Classification Module

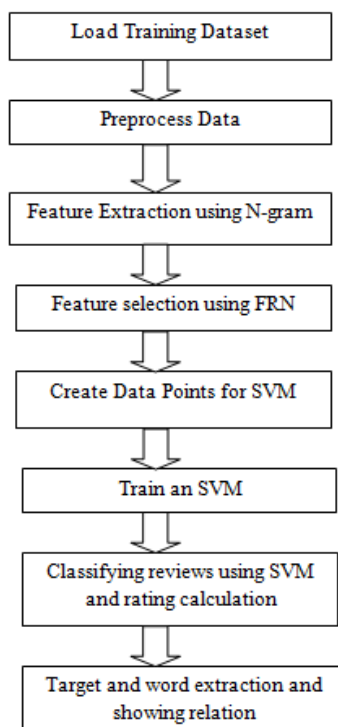
Features are classified by using Support Vector Machine (SVM) classifier SVM is discriminative classifier and it is the best text classifier. It is a statistical classification method. SVM separates decisions surface into separate training data points into classes and perform classification based on Support vector. Reviews are in the form of text so

reviews need to be converting in SVM representation to perform classification. Each review is then converted into numeric value names as data points. Each review is considered as four dimensional points. Using numeric representation of review text SVM perform classification on input data. Features are placed in different classes depends on polarity.

#### 4. Rating Module

Rating module used to calculate the rate for test review and its following class of review. This module also used to calculate the overall rating of the product from all reviews.

#### D. Implementaion Steps



**Figure 2:** Implementation Steps

The implementation steps are as follows:

Step 1: In first step we load the dataset which is used for training. The data set is in the json format we first extract and then transform it mysql database for easy access and then extract only reviews in single file.

Step 2: In step 2 preprocessing is done on the reviews from file to remove stopwords.

Step 3: Feature extraction is done using n-gram in step 3.

Step 4: Feature Selection is performed by using feature relational network.

Step 5: In step 5 SVM data points are calculated which are used for SVM training. These data points are numeric values on which SVM is works.

Step 6: By using SVM data points SVM is trained in step 6.

Step 7: Test on single and multiple reviews are performed and ratings are calculated.

Step 8: In last step opinion words and targets are calculated and we show the relation between them graphically.

#### E. Scope

Proposed work is based on opinion mining and text classification. Classification is done on based on polarity.

Using N-gram feature set we find the features from text review. N is any positive number used in feature extraction. Here we use value of N is 3 because to find feature relational network it requires the same value.

Feature relational network (FRN) finds the relation between features and association between them.

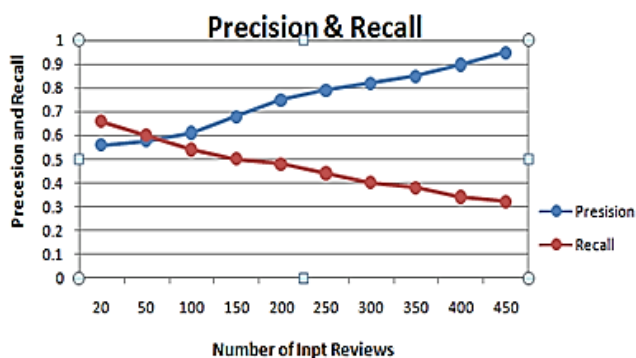
Support Vector Machine is classifier which classifies text in classes. Classes need to be predefined. It works on numeric data so here needs to convert text into numeric value and based on these values SVM is trained.

Word alignment model finds the opinion words and opinion targets from reviews. Target is entity on which opinion is given and opinion word is opinion about the product or target.

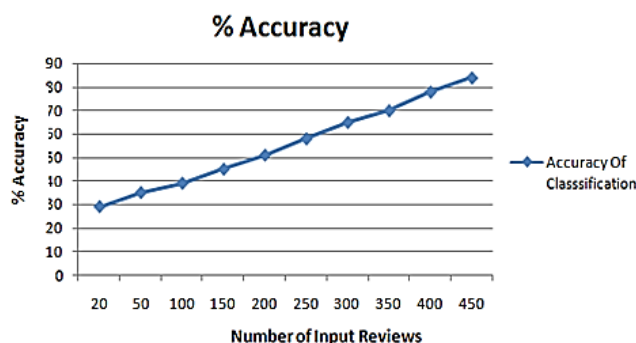
### III. RESULTS AND DISCUSSION

The experiment for the classification on data is performed by SVM algorithm carried out on review dataset. The approach proposed for text review data and classifies the review in positive, negative and neutral classes. The result analysis is done on the

classification for accuracy factor. The accuracy is calculated by calculating precision and recall for the input data. In graph following graphs horizontal axis represent number of input reviews to system and vertical axis represent precision and recall of review classification for first graph and percentage accuracy of classification for second graph. The graph 1 shows that for less number of reviews precision is less than the recall. After gradually increasing the number of input reviews precision gets increases and recall decreases. Graph 2 is for accuracy which is based on precision. Accuracy increases as the number of records increases.



**Graph 1:** Precision and recall for proposed system (Classification)



**Graph 1:** Accuracy of Classification

#### IV.CONCLUSION

In the proposed work classification of online reviews are and rating calculation is performed. In proposed framework is combination of feature extraction from reviews by feature relational network and classification of these reviews based on support vector machine. Feature relational network allows finding relevant feature by using N-gram technique and selecting the features by using FRN. Support vector machine takes review text as input and classify them to the different classes like positive, negative or neutral depending on the training pattern. Here we

also perform a rating of the product based on the review. The proposed system gives classes of reviews with rating which is useful for the user for to take decision about purchasing product and they can easily supervise the product.

#### V.REFERENCES

- [1] Kang Lui, Liheng Xu, and Jun Zhao, "CoExtracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model", IEEE Transactions on Knowledge and Data Engineering, Vol.27 No.3 March 2015
- [2] Ding et al., Xiaowen Ding, Bing Liu, and Philip S. Yu, "A holistic lexicon-based approach to opinion mining", In Proceedings of the Conference on Web Search and Web Data Mining (WSDM), 2008.
- [3] Fangtao Li, Chao Han, Minlie Huang, Xi-aoyan Zhu, Yingju Xia, Shu Zhang, and HaoYu,"Structure-aware review mining and summarization", In Chu-Ren Huang and Dan Jurafsky, editors, COLING, pages 653–661. Tsinghua University Press, 2010.
- [4] Wei Jin and Hay Ho Huang, "A novel lexicalized hmm-based learning framework for web opinion mining", In Proceedings of International Conference on Machine Learning (ICML), 2009.
- [5] Mingqin Hu and Bing Liu, "Mining opinion features in customer reviews", In Proceedings of Conference on Artificial Intelligence (AAAI), 2004.
- [6] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Che, "Expanding domain sentiment lexicon through double propagation", 2009.
- [7] Ahmad Abbasi, Stephen France, zhu zhang, and Hisinchun Chen, "Selecting Attributes for Sentiment Classification using Feature Relational Network", IEEE Transactions on Knowledge and Data Engineering. Vol 23. No 3. Page 447-462. 2011
- [8] Loubes, J. M. and van de Geer, S "Support vector machines and the Bayes rule in classification", Data mining knowledge and discovery 6 259-275.2002
- [9] Gongde Guo, Hui Wang, David Bell, Yaxin Bi and Kieran Greer, "KNN Model-Based Approach in Classification", Proc. ODBASE pp- 986 – 996, 2003