Classification of Business Reviews using Sentiment Analysis

Shilpa Shendre[1], Prof. Pramila Chawan[2]

[1]*Mtech Student, Dept of Computer Engineering and IT, VJTI College Mumbai, Maharashtra, India*
[2]*Associate Professor, Dept of Computer and IT Engineering, VJTI College Mumbai, Maharashtra, India*

---***---

Abstract - *A The rapid increase in mountains of unstructured textual data accompanied by the proliferation of tools to analyze them has opened up great opportunities and challenges for text research. The research area of sentiment analysis has gained popularity in the last years. Business developers not only want to know about there product marketing and profit based on the number of sales been done but also want to know about the reviews and thoughts of people using these products. The feedback they receive via social media and other internet services becomes very important to measure the quality of a product they are serving. Sentiment analysis is a domain where the analysis is focused on the extraction of feedback and opinions of the users towards a particular topic from a structured, or unstructured textual data. In this paper, we try to focus our effort on sentiment analysis on restaurant review database. We examine the sentiment expression to classify the reviews of the restaurant business whether it is positive or negative and perform the feature extraction and use these features for updating and maintenance of the business.*

***Key Words*: sentiment analysis; opinion mining; classification; text reviews, Machine learning**

1. INTRODUCTION

Sentiment analysis has become an important research area for understanding people's opinion on a matter by differentiating a huge amount of information. The present era of the Internet has become a huge Cyber Database which hosts the gigantic amount of data which is created and consumed by the users. People across the world share their views about various services or products using social networking sites, blogs or popular reviews sites. The Internet is been growing at an exponential rate giving rise to communicate across the globe in which people express their views on social media such as Facebook, Twitter, Rotten Tomatoes and Foursquare. Opinions which are being expressed in the form of reviews provide a platform for new explorations to find collective reviews of people. One such domain of reviews is the domain of business reviews which affects business people. The feedback from the customer is valuable for companies to analyze their customer's satisfaction and survey the competitors. This is also useful for other people or consumers who want to buy a product or a service prior to making a purchase.

In this paper, we are going to present the results of machine algorithms for classifying reviews using semantic analysis. A large number of customer-generated reviews for businesses and service providers are classified as either positive or negative. We propose a method to automatically classify customer sentiments using only business text review. This helps us to generate the result using feedback without manual intervention. By studying only rating, it is very difficult to judge why the user has rated the product as 1 or 5 stars. However, the text content contains a more quantitative value for analyzing more than rating itself.

In this paper, we are going to mention the preprocessing steps require in order to achieve accuracy in the classification task. There is no previous research available on classifying sentiment of business review using the latest reviews forms restaurant dataset. Determining the underlying sentiment of restaurant business review is a difficult task taking into account several factors such as the connotation of a word depending on the context, language used, words ambiguity when using words that don't express a particular sentiment or when using sarcasm. We show that a sentiment analysis algorithm built on top of machine learning algorithms such as Naïve Bayes and Linear Support Vector Classification (SVC) has accuracy above 90% business reviews.

2. WORK REALTED

Hu et al. perform the classification of a document at the sentence level. Instead of the whole document and feature extract on which views have been expressed, identifying comments words by proposing a technique that uses the WordNet lexical database. For each feature extracted, the related reviews sentence is stored in positive or negative categories and computes a total count. The features are ranked on the bases of there frequency of the appearance in the reviews. The feature-based summary of the reviews of the product sold online was provided by the authors.

Usually work related to sentiment analysis using machine learning techniques in determining if the overall review is positive or negative movie reviews as data. The writer's used unigram model and Navie Bayes, entropy classification, and SVM to perform the classification and achieve accuracy upto 80%. They finally concluded that their results outperform the method based on human tagged features.

A system was built by Blair-Goldensohn et al. which automatically summarize sentiment from a set of reviews for a local service such as restaurant or hotel and combine the review sentiment per aspect such as food, service, decor, value etc., Basically they have implemented a custom built lexicon based on WordNet and used a classifier at the sentence level.

3. THE PROPOSED METHOD

The basic methodology to determine polarity is the one with a lexical approach, where we look at the words comprising the document and apply some algorithms to quantify words with some sentiment score and determine the collective polarity. We have based our computational method on the publically available library SentiWordNet

In this work for determining the polarity of the reviews , we have focused on two areas: 1) Feature Selection and Ranking 2) Classification using Machine Learning techniques. We use the restaurant review dataset comprising 51Mb reviews. We tend to label the polarity as follows : 0- Strong Negative, 1- Weak Negative, 2-Neutral, 3-week, Positive, 4-Strong Positive. The proposed methodology can be well explained from the below figure.

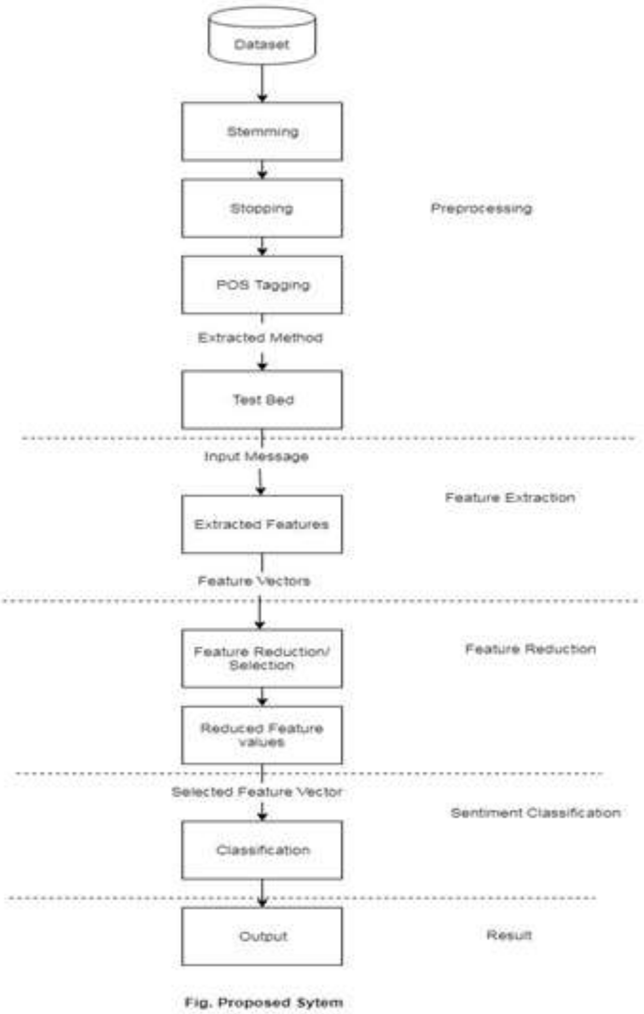

Fig1. Proposed System

Since SemEval 2016 has independent tasks for aspect extraction and aspect categorization, we need to add sentence preprocess before applying aspect categorization. We split a sentence into two sentences if it has contrary conjunction, such as "but", "however", etc. This preprocess solves problem when a sentence has two or more aspects with same category but different sentiment polarity. We define two rules i.e.: If both clauses before and after contrary conjunction contain extracted aspects, we split the sentence into two sentences by using contrary conjunction as delimiter. Example such as I like the dessert but I don't like the meal is splitted into first: I like the dessert but and second: I don't like the meal. If clause before contrary conjunction contains aspect but clause after contrary conjunction does not contain aspect, we also split it into two sentences by using contrary conjunction as delimiter and inserting aspect from the first sentence into the second sentence. For example, "food price is expensive but comparable" is splitted into "food price is expensive" and "food price is comparable".

We build multilabel binary relevance classifier with MaxEnt algorithm for aspect categorization. The illustration of binary relevance classifier. We define four categories: food, place, price, and service. Each category has its binary classifier so the total classifier is same as the total of category. Each sentence in the corpus is labeled with Boolean value for each category: true if a sentence has certain category and false otherwise. The classifier for each category classify the Boolean value. After that, we collect categories with true values as multilabel output.

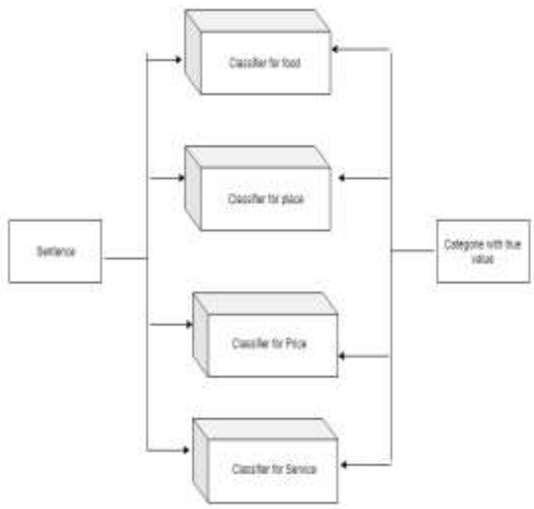

Fig 2.Relevance Classifier for aspect categorization

We apply the same algorithm as in aspect categorization for sentiment classification. We define two labels of sentiment for each category in a sentence: positive and negative. For example, a sentence has food and place category. Each category has its own classifier to classify sentiment of categories in the sentence. To classify the sentiment polarity, we use classifier for food and place category. As in aspect categorization, we use general features from English. Bag of clusters is obtained from CBOW, LDA, and GloVe. We use 2 to 5 words skips for Skip-bigram. N value of 1 and 2 is also used for bag of N-gram feature.

4. EXPERIMENT, EVALUATION, AND DISCUSSION

In order to use supervised learning and train a classifier, we usually require a predefined training data, but taking into account the large range of restaurant businesses and the large number of reviews, it would be very difficult to manually annotate the data to train a sentiment classifier for reviews.

4.1 FEATURE EXTRACTION

Mostly the researchers apply standard feature selection in there approach to improve performance with few using more practical approaches. We are focusing completely on feature election to improve sentiment analysis are few. One of them is the famous Pang & Lee, who removed objective sentences on a testbed consisting of objective and subjective text trained on SVM. Initially, they found that sentiment classification result is actually slow and moderate. They then concluded it was more likely that sentences adjacent to discarded sentences improved classification result over their baseline.

For opinion structure generation, we employ CBOW model to find similarity between extracted aspect with seed words for each category. Category that has maximum similarity score will be paired with the extracted aspect. For example, a sentence has food and place category and we want to pair "cake" with a category from the sentence. First, we find similarity score for each seed words for food and place category. The maximum similarity score for food and place category are 0.2 and 0.1 respectively. After that, we can pair "cake" with food category because similarity score for food category is higher than place category. The generated opinion structures are used to calculate rating for each category with equation as in.

Feature	Definition
Bag of N-grams	The occurrence of a N-gram in the context window
Bag of Head Words	Bag of word that determines the syntactic category of that word from the dependency tree
Bag of Clusters	The occurrence of word's cluster in the context window
Bag of k-skip-bigram	Bag of N-gram which has skipped over gaps.

Table 1. Features For Sentiment Classification

TABLE 1 show token distribution and example of training data for aspect extraction step. For clusters and clusters bigram feature, there are four scenarios to experiment with the number of clusters. The best combination for this step is bag of N-gram, bag of POS N-gram, clusters with 5000 clusters, and clusters bigram with 100 clusters CBOW.

4.2 FEATURE CATEGORIZATION

The label distribution in training data and example of training data for aspect categorization . For bag of clusters feature, we also use four scenarios and additional six scenarios to experiment with number of topic for LDA. The best feature for this step is bag of clusters using CBOW model with 1000 clusters.

This is the highest accuracy obtained using this method. Also it's worth noting that giving equal importance to all factors i.e. giving each a value of 0.165 has resulted in a lower accuracy of 78.268% than the highest accuracy obtained by unequal distribution of factors. Thus by changing the importance of that aspect, we can see its effect in the accuracy of the overall classification of the review.

Thus we can interpret from the results that in the reviews used from the dataset, the user has given more importance to these factors while writing the review. It also means that if the user tends to give a positive review towards these aspects then, due to their increased importance, the overall review will tend to be positive even if the user gives a negative feedback towards the other aspects. Giving more importance to certain factors also has an added advantage, it tends to suppress the users opinion about other factors. Suppose we have a reviewed 'X' and it contains user's opinion about two factors F1 and F2. Also the overall orientation of the review is positive in nature. The user has given a positive review about F1 and a negative about F2. Also the amount of text in the review for F1 aspect is less as compared to the F2 aspect. If we use any non-aspect based sentiment analysis method then since text size of F2 is greater than text size of F1 and also since F2 is negative in orientation, the overall review score will tend to reduce and skew towards.

The various performance measures used were:

Accuracy = (Total correctly classified word / Total number of words)

Precision = tp / (tp +fp)

Specificity = (tn / Total number of negatively oriented review in the dataset)

Recall = (tp / Total number of positively oriented review in the dataset)

Where tp, fp and tn are the true positives, false positives and true negatives obtained during the classification. On the other hand if driving factors are used and F1 is given more importance the review score will better reflect the positivity of the review. Since each aspect of a restaurant is analysed separately in this method, we can track the effect each aspect has towards the overall score of the review. This individual aspect based tracking can be used in a fined grained aspect based recommendation system, which recommends restaurants based on its various aspects instead of the overall rating of the restaurant. Also this method can be applied on a product review dataset thus enabling us to see what opinion each user has on the various aspects of the product, thus helping in the development of proper product placement strategy. It is very difficult to acquire such in-depth knowledge from the dataset using non-aspect based methods.

Label	Total Sentence
Food	503
Service	97
Price	125
Place	440

TABLE 2. DISTRIBUTION OF ASPECT CATEGORIZATION IN DATA

Some misclassifications occur when the sentence has word “restaurant” or restaurant name. For example, sentence “we come to Atmosphere cafe for celebrating our anniversary with high expectation” is classified as place category while the sentence actually does not have any category. The sentence that have word “atmosphere” only appear once in the training data and it is labeled as place category so the sentence misclassified as place category. Another misclassification happens in sentence “this is an old restaurant that still exists until now”. The sentence has word “restaurant” so it is classified as place category by the model. But word “restaurant” does not have any sentiment so the sentence is not labeled as place category even though it has word “restaurant".

Misclassification can happen if words never co-occur in the training data as in aspect extraction. For example, sentence “but the speciality is the environment” is classified as food category while the sentence actually has place category. Most sentences have word “special” labeled as food category in the training data because it co-occurs with word related to food and it never cooccurs with word “environment”. Besides that, the word "environment" never appears in training data. Because of that, the sentence is misclassified as food category.

Aspect Category	Example Seeds
Food	Food, beverage, dessert, meal, taste
Service	Service, waiter, waitress
Price	Price
Place	place, atmosphere,

TABLE 2. ASCPECT CATEGORIES

After we have all aspect categories and its aspect, we will calculate the rating for each aspect categories. The rating calculation will follow the equation 1.

Rating = (p / (p+n) *4) + 1

Variable P/N is the total of positive/negative opinion in the aspect category. The rating is scaled in 1 – 5.

Example: *The place was comfortable, the view was nice, and the price was affordable. In my opinion, the food was good, but the cocktail was not too good, the bartender still has a lot to learn.*

Experiments are conducted for various aspect level sentiment classification with feature selection methods and different feature set size. The aim of this analysis is to see 1) if machine learning algorithms for aspect level sentiment classification work 2) if the size of the feature set influence the performance of classification.

Category	Sentiment	Aspect	Rating
Food	Positive	food	5.00
Negative	-		
Place	Positive	-	0.00
Negative	-		
Price	Positive	price	5.00
Negative	-		
Service	Positive	place view	5.00
Negative	-		

TABLE 4. EXAMPLE OF GENERATED OPINION STRUCTURES AND RATING

4.3 Aspect and Sentiment Extraction
For token classification, the accuracy of the label is quite high with 88.48. The results also show that F1-Measure for OP_NEG_I is quite low while for OTHER is high. Many misclassifications occurred and the tokens are mostly classified as OTHER class. This is because of the use of infrequent words to describe the aspects and the opinions. Those infrequent words are then classified as OTHER.

TABLE 4. EVALUATION RESULT ON TOKEN CLASIFICATION

Label	Precision	Recall	F1
ASPECT-B	0.7104	0.7455	0.7275
ASPECT-I	0.4929	0.5475	0.5188
OP_POS-B	0.7524	0.8505	0.7985
OP_POS-I	0.6885	0.8235	0.75
OP_NEG-B	0.6923	0.5373	0.605
OP_NEG-I	0.5926	0.4444	0.5079
OTHER	0.943	0.9243	0.9336
Accuracy	0.8848		

One of the example of class ASPECT classified as OTHER is in the sentence "order grilled carp, the crap slightly over burn". In that sentence, the aspect is "the crap" however the model cannot detect the aspect. It is caused by that word never appeared in training data. The opinion in that sentence is "slightly over burn". Those words also never appear in the data and also classified as OTHER.

In contrast, all tokens in the sentence "the place is nice, the atmosphere is also comfortable, and a lot of variation in the menu" are correctly classified. The reason behind this is all of the tokens have occurred in the training data and the model can easily recognize the pattern in the sentence.

5. CONCLUSION

Sentiment analysis is vast research area and it has wide variety of issues to be discussed with several challenges. This paper explains the aspect based feature selection methods in combination with other machine learning algorithms. The results of experiment explain that features or aspects are selected and iterative classifier using machine learning technique is proposed. Aspect-based sentiment analysis has six steps i.e. preprocess, aspect extraction, aspect categorization, sentiment classification, opinion structure generation, and rating calculation. The experiment includes review data sets which includes positive and negative aspects. Our method gives best results for precision, recall and accuracy compared to SVM and naives bayes method. The proposed method iteratively runs while processing the data and analyses based on previous experience. The accuracy increases up to 83.5% the accuracy scaled up to a great extent. The naïves method scaled up to 78.44% and SVM scaled up to 80.34 percent. We identify and calculate the precision, recall and accuracy for the models .It shows that the proposed method gives better results. Future work would be to combine different feature selection schemes for analyzing the accuracy of the review data sets.

References

[1] Kudakwashe Z, Oludayo O., et al. "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews.2018". ICTAS

[2] Kulvinder S.,& Sanjeev D., &Pratibha "Real-time Data Elicitation from Twitter: Evaluation and Depiction Strategies of Tweets Concerned to the Blazing Issues Through Twitter Application " LREC. Vol. 10. 2010.

[3] Rincy J.,& Varghese S. "Prediction of Election Result by Enhanced Sentiment Analysis on Data using Classifier Approach " Computational linguistics 267-307.

[4] Xiaobo Z., &Qingsong Y., "Hotel Reviews Sentiment Analysis Based on Word Vector Clustering " 2017 2nd IEEE International Conference on Computational Intelligence and Applications.

[5] Wang Z., Qin S., A Sentiment Analysis Method of Chinese Specialized Field Short Commentary" 2017 3rd IEEE International Conference on Computer and Communications

[6] Harpreet K., Veenu M., Nidhi., "A Survey of Sentiment Analysis techniques International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017) Mining Workshops. IEEE, 2011.

[7] Jie Li; Lirong Qiu, "A Sentiment Analysis Method of Short Texts in Microblog" in 2017 IEEE International Conference on Computational Science and Engineering (CSE)

[8] Kudakwashe Z., Oludayo O., "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews" in 2018 (ICTAS)

[9] V. S. Jagtap, Karishma Pawar, "Analysis of different approaches to Sentence-Level Sentiment", International Journal of Scientific Engineering and Technology Volume 2 Issue 3, PP (ISSN : 2277) (2013).

[10] Nidhi Mishra, "Classification of Opinion Mining Techniques",IEEE Trans.Syst., International Journal of Computer Applications (0975 – 8887) Volume 56– No.13, October 2012.

[11] Aansi A. Kothari and Warish D. Patel," A Novel Approach towards Context Sensitive Recommendations Based On Machine Learning Methodology" 2015 Fifth International Conference on Communication Systems and Network Technologies.

[12] A. Mudinas et al, 2012, Combining lexicon and learning based approaches for concept-level sentiment analysis, Proceedings of the First International Workshop on Issues of Sentiment
[13] Discovery and Opinion Mining, ACM, New York, NY, USA, Article 5, 1-8.

[14] M. RushdiSaleh et al, 2011, Experiments with SVM to classifyopinions in different domains, Expert Systems with Applications 38.

[15] Emma Haddi, Xiaohui Liu, Yong Shi, "The Role of Text PreProcessing In Sentiment Analysis", Information Technology and Quantitative Management (Itqm2013), 1877-0509, Elsevier © 2013.

[16] T. K. Das, D. P. Acharjya and M. R. Patra: "Business Intelligence from Online Product Review-A Rough Set Based Rule Induction Approach", International Conference on Contemporary Computing and Informatics (IC3I-2014), November 03 – 05, 2014, Mysore, India, IEEE Xplore, pp. 800 – 803, 2014.

[17] T. K. Das, D. P. Acharjya and M. R. Patra: "Opinion Mining about a Product by Analyzing Public Tweets in Twitter", International Conference on Computer Communication and Informatics (ICCCI - 2014), January 03 – 05, 2014, Coimbatore, India, IEEE Xplore, pp. 1- 4, 2014.

[18] D. P. Acharjya and L. Ezhilarasi: "A Knowledge Mining Model for Ranking Institutions using Rough Computing with Ordering Rules and Formal Concept Analysis"; International Journal of Computer Science Issues; Vol. 8 (2), pp. 417 – 425, 2011.

[19] T. K. Das and D. P. Acharjya: "A Decision Making Model using Soft Set and Rough Set on Fuzzy Approximation Spaces", International Journal of Intelligent Systems Technologies and Applications, Vol. 13 (3), pp. 170-186, 2014.

[20] T. K. Das, D. P. Acharjya and M. R. Patra: "Multi Criterion Decision Making using Intuitionistic Fuzzy Rough Set on Two Universal Sets", International Journal of Intelligent Systems and Applications, Vol. 7 (4), pp. 26-33, 2015.