International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Sentiment Analysis of Restaurant Reviews

Pooja Tyagi

Student, Dept. of Computer Science Engineering, Galgotias University, UP, India

---

Abstract - *In the past for decision making people ask about the product/service from their relatives, friends, neighbors etc., and then they take decision whether to go for purchase or discard the idea. In today's time, the internet is our best advisor, because lots of reviews are available that can be considered as word of mouth that will help us to take decision. The positive and negative both experiences are shared by the actual user of the product/service. But it is impossible to reads all the reviews. We don't even know how many reviews can affect our decisions, so as we know that this era belongs to Machine Learning (ML). A machine can be learned in such a manner, so that it will read all the reviews for us and give us on how many reviews are related to you, there are some good unsupervised and supervised methods are defined in the field of Natural Language Processing (NPL). In this paper we are going to introduce an unsupervised method that will help to take decision. When we do not have enough data to make a supervised machine, unsupervised methods are coming into the picture to do the same work but without labeled data. Labeled data is nothing but pre-classified reviews, Classes are major aspects, and the reviews will be classified by the method that will be categorized into five aspects food, services, price, ambience and anecdotes / miscellaneous. Food aspects are used to assign by the method if any review will have any food related information, and the same condition are applicable for other aspects categories. If we have multiple aspects in a review, then it will be added into the respective aspect. The opinion in the form of electronic word of mouth, reviews of the user has great impact on decision making. This analysis can help a business or organization to find out best and worst about them. That will help to correct worst cases on respective aspects.*

***Key Words*: Sentiment analysis, aspect category, restaurant reviews, root words, fire word, unsupervised method.**

1. INTRODUCTION

This technology processes the logical structure to identify the most relevant elements in text and then understand the topic discussed. The sentiment analysis is also known as opinion mining, in which the opinions, appraisals, emotions or attitude towards a topic, person or entity are analyzed. The expressions can be classified as positive, negative or neutral. For example: "I really liked the garlic noodles of your restaurant "- this is a positive expression. The overall sentiment polarity shows a preference on service in the reviews, which might hint the customers to "self-select" the food they like. Natural language processing in artificial intelligence applications makes it easy to gather product reviews from a website and understand what consumers are actually saying as well as their sentiment in reference to a specific product. Companies with a large volume of reviews can actually understand them and use the data collected to recommend new products or services based on customer preferences.

1.1 RELATED WORK

[1.1.1] Kim Schouten, Onne van der Weijde, Flavius Frasincar, and Rommert Dekker, "Supervised aspect category detection for sentiment analysis with co-occurrence data" IEEE transactions on Cybernetics, Volume: PP, issue: 99, page 1-13, April 2017.Many researchers introduced supervised and unsupervised learning methods for detecting aspect category for sentiment analysis in this; they used both supervised and unsupervised methods for aspect category detection for sentiment analysis with co-occurrence data.

[1.1.2] S. Kiritchenko, X. Zhu, C. Cherry, and S. M. Mohammad, "NRC Canada-29014: Detecting aspects and sentiment in customer reviews," in Proc. 8th Int. Workshop Eval. (SemiEval), Dublin, Ireland, 2014, pp.437-442. In 2014, M. Pontiki et al's paper introduced 4 subtasks for restaurant and laptop review data. The aspect category for restaurant data is given but for the laptop data, there are not such categories as given. As for restaurant data, all the tasks are applicable.

[1.1.3] M. Pontiki et al., "SemEval-2014 Task 4: Aspect based sentiment analysis," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), Dublin, Ireland,2014, pp. 27-35.The paper introduced 4 subtasks for restaurant and laptop review data. The aspect category for restaurant data is given but for the laptop data, there are no such categories are given. As for restaurant data [service, food, price, ambience and miscellaneous], all the four tasks are applicable for the restaurant data. But for laptop data only two tasks are open.

[1.1.4] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE transactions knowledge and data engineering, vol. 28, no. 3, pp. 813-830, Mar.2016. The paper explained about Joint aspect detection and sentiment analysis methods, there are syntax-based method, supervised machine learning and hybrid machine learning. It also explains problems in all methods for aspect detection and sentiment analysis. the major problems are occurred by comparative opinions, conditional sentenced, negations modifiers and also explained aggregation methods in brief.

1.2 MOTIVATION

As Natural language processing is used in many artificial intelligence applications, and it is an area where lots of work already done and lots of word need to be done in future. Natural language processing can be beneficial for both consumer and business. The work given by SemEval2014, it is similar kind of work where customers and owners of the restaurant can use our application to find their beneficial information. Earlier applications using unsupervised method are able to give a usable system, which cannot be used to find right information. I found myself that I can contribute to this method in positive manner. Unsupervised method is usable, where the labeled data is not available to use supervised learning methods for classifying raw data.

2. PROPOSED APPROACH

The proposed method is an unsupervised method and co-occurrence relation and spreading activation algorithm, it will totally depend on root words, which can be called from Wordnet for each aspect category and some special {for example: food- dosa, sandwich, lime soda, etc} words that are not available in the Wordnet, that are closely related to the food, and likewise for other aspect categories. The root words the major part of this approach, that can be easily extract from various source for example: menu of the restaurant these words are helpful to find food related reviews. After collection of all root words for each category. The data will be pre-processing an all pre-processing steps need to perform, determined steps for Unsupervised method, to build occurrence matrix(X) need to find total distinct words are in the data, after that co-occurrence matrix(Y) is designed using all distinct words available in the entire reviews. The co-occurrence directed graph G (V, E) will be designed based on co-occurrence matrix of words, if Yi and Yj are occurred in the same review and the occurrence of Yi is before Yj then the Yj will have an edge directed form Yi. The same method will be applicable to entire data. The weight of each edge between two vertices will be decided by (Wi, j) = Yij / Xj, generating the activation value to each V will be calculated by applying spreading activation algorithm. After getting activation value it will be decided that which word is close word to the root words and of the which aspect category, those words activation value will be used to sum up and if the value crosses the threshold value, then the aspect category will be assigned for that sentence by implicitly. After assigning the aspect category, the plain text and pre-processed data that is called comes under that particular aspect will be used to calculate for sentiment analysis for that particular aspect category. The sentiment score will be called from sentiwordnet3.0 and assigned to the sentence that is already have aspect. Spreading activation algorithm, in this algorithm we have four input- root words, G (V, E), decay factor, threshold value and we will get two output- lists of fire-words and activation value for each word of graph. For each vertices of graph V(G) 0 will be assigned. After that for each G(Vi) that belong to root words (Rc) 1 will be assigned. Rest of the vertices will get the activation value Avi Min(Avj+ Avi* Wij* dc,1), if the value of Avj is greater than threshold value i.e. 0.1, the word will be added into the fire word list for the respective aspect category. This process will be done for each aspect category. Association rule algorithm will be used to get aspect category of each sentence and for each aspect category. In this we have four inputs for each aspect category (Reviews, Graph (V, E), threshold value atc and fireWordList Fc and one output aspect category for the textT. (1). For each word in sentence Ti. (2). If Ti is available in the graph and its value is 1 than 3. The sentence will get the aspect category for respective input aspect category, if not 4. Find the co-occurrence words in the graph and sum up the activation values and compare with threshold value aspect category (Atc 1.0) if the sum up value is greater than or equal to Atc, the text will get the respect category if less than it will be assigned default aspect category. A limitation of the proposed approach is range of root words. Root words are basically, synonyms and some extra words, that are used in general English language in any restaurant for target aspect, if we are able to list more words under root words, it would lead to get more accurate fire words for the same aspect, then it is very easy to find aspect of the data, fire words are nothing but the words, which were frequently occurred with the root words. It will help the machine, when any root word is not there but fire words are there than the sum up the value of each fire word available in the review, will assign the aspect category as implicitly.

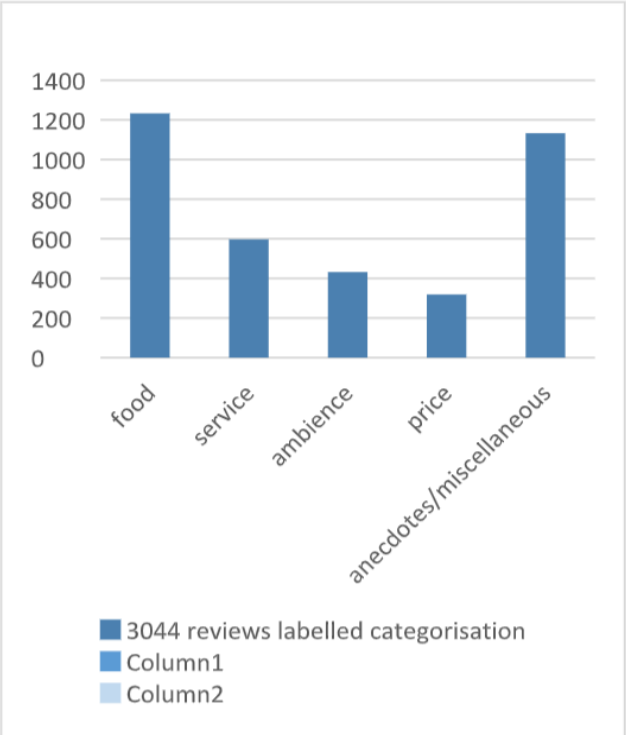

Chart -1: 3044 reviews labeled categorization

3. EVALUATION

We have total 3044 word, but after applying the proposed method it become 3566, the reason of such a major changes is a review can have multiple aspect; the result is in table 1, in the second column labeled data is given by the SemEval 2014 and next in column the result of proposed approach for each aspect category. TRUE and FALSE columns are representing the truly and falsely identified by the proposed approach is performing better than many unsupervised methods that are used to do the same task. In Fig.2 in the blue lines are showing correctly identified reviews and green line represents for proposed approach data volume. In proposed data volume we have truly and falsely identified both.

Table -1: Total Reviews: 3044

Categories	Labeled data	Proposed approach	TRUE	FALSE	% Accuracy
Food	1233	1097	991	106	90.34
Price	319	244	218	26	89.34
Ambience	432	414	389	25	93.96
Service	597	418	398	20	95.22
Anecdotes/ miscellaneous	1133	1393	1067	326	76.6
Total	3714	3566	3063	503	85.89

3. CONCLUSION

In this paper we introduced and unsupervised method for aspect category detection from the reviews of restaurant. It would be hard to get aspect category if the target aspects are not given. The main drawback of proposed approach is fire word threshold value, and aspect threshold value, two different values are used to maximize the performance. It needs to be carefully set before applying the proposed approach.

The orange line is representing the data volume in each aspect category, and blue line is representing actual labeled data, here we can see food, price, ambience and services have a smaller number of reviews but in anecdotes/miscellaneous we have a greater number of reviews than labeled data. It is because we have a smaller number of reviews in other category, all those reviews, which are not identified by the proposed approach, will transferred into default category. And default category is anecdotes/miscellaneous.

The result by the proposed approach for correct identified in:

Food= 398,

Price= 218,

Service= 398,

Ambience= 389,

Anecdotes= 1067.

The total aspect was *3566* created by the proposed approach and out of *3566* the correctly identified aspects are *3063.* It gives a good accuracy of *85.89%.*

Chart -2: data volume of proposed approach

4. Future work

While calculating the sentiment, the sentiment is taken form SentiWordNet3.0, it is for English language only, and we will use different techniques and different library like TextBlob to get the sentiment. This work is done only in English language, the work can be extended for other Indian languages, we have to study more research paper on sentiment analysis on Indian languages, and aspect category detection for Indian languages, as far as concerned about my knowledge, there are not any SentiWordNet3.0 and WordNet for Indian languages, so we need to implement the work from the scratch. It would be a great experience of learning and implementing.

REFERENCES

[1] Kim Schouten, Onne van der Weijde, Flavius Frasincar, and Rommert Dekker, "Supervised aspect category detection for sentiment analysis with co-occurrence data" IEEE transactions on Cybernetics, Volume: PP, issue: 99, page 1-13, April 2017.

https://www.researchgate.net/publication/316144039_Supervised_and_Unsupervised_Aspect_Category_Detection_for_Sentiment_Analysis_With_Co-Occurrence_Data

[2] S. Kiritchenko, X. Zhu, C. Cherry, and S. M. Mohammad, "NRC Canada-29014: Detecting aspects and sentiment in customer reviews," in Proc. 8th Int. Workshop Eval. (SemiEval), Dublin, Ireland, 2014, pp.437-442.

https://www.aclweb.org/anthology/S14-2076.pdf

[3] M. Pontiki et al., "SemEval-2014 Task 4: Aspect based sentiment analysis," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), Dublin, Ireland, 2014, pp. 27-35.

https://www.aclweb.org/anthology/S14-2004.pdf

[4] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE transactions knowledge and data engineering, vol. 28, no. 3, pp. 813-830, Mar.2016.

https://ieeexplore.ieee.org/document/7286808