IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 01-05
www.iosrjournals.org

Category Based Sentiment Analysis in Travel And Tourism Domain

Devika M D[1], Sunitha C[2], Amal Ganesh[3]
[1(]*Department of CSE, Vidya Academy of Science and Technology,India)*
[2(]*Department of CSE, Vidya Academy of Science and Technology, India)*
[3(]*Department of CSE, Vidya Academy of Science and Technology, India)*

Abstract: *Sentiment analysis (SA) is an intellectual process of extricating user's feelings and emotions. It is one of the pursued fields of Natural language processing (NLP). The evolution of Internet based applications has steered massive amount of personalized reviews for various related information on the Web. These reviews exist in different forms like social Medias, blogs, Wiki or forum websites. Both travellers and customers find the information in these reviews to be beneficial for their understanding and planning processes. The boom of search engines like Yahoo and Google has flooded users with copious amount of relevant reviews about specific destinations, which is still beyond human comprehension. Sentiment analysis poses as a powerful tool for users to extract the needful information, as well as to aggregate the collective sentiments of the reviews. The chief objective of this paper is to implement the category based sentiment analysis in travel and tourism domain.*
Keywords: *Setiments, lexicons, semantics, polarity*

I. Introduction

Sentiment analysis refers to the use of NLP, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

Sentiment analysis is the process of detecting the contextual polarity of the text. It determines whether given text is positive, negative or neutral. The alternate word is opinion mining as it derives the opinion or attitude of the speaker. There are lots of researches going on in the field of sentiment analysis because of the marketing level competition and the change in needs of the people. Several methods are evolved for doing this task. Performing sentiment analysis by various methods will earn different results.

Sentiment analysis require the usage of the training set for its performance, the quality of the training set also plays a great role in the accurate evaluation of the text. An essential phenomenon in natural language processing is the use of discourse relations to establish a coherent relation, linking phrases and clauses in a text. The presence of linguistic constructs like connectives, modals, conditionals and negation can alter sentiment at the sentence level as well as the clausal or phrasal level. The semantic analysis of the sentence will also increase the meaning and accuracy of the result. Sometimes in the reviews POS tagging will be helpful for understanding whether the review or comment corresponds to the subject we are searching for. We can say that the result is good when we treat the sentences as n-grams instead of word by word estimation.

The paper is organized as follows: Section 1 provides an introduction. Section 2 gives related work. Section 3 provides the overview of the proposed system. Section 4 gives performance evaluation and Section 5 concludes the manuscript.

II. Related Work

Sentiment analysis played a great role in the area of researches done by many; there are many methods to carry out sentiment analysis. Still many researches are going on to find out better alternatives due to its importance in this scenario. Some of the methods are discussed in the previous related works are Machine Learning Approach, Rule Based Approach, Lexicon Based Approach.

Machine learning techniques first trains the algorithm with some particular inputs with known outputs so that later it can work with new unknown data [2].

Rule based approach is used by defining various rules for getting the opinion, created by tokenizing each sentence in every document and then testing each token, or word, for its presence. If the word is there and has with a positive sentiment, a +1 rating was applied to it. If the input sentence contains any word which is not present in the database which may help in the analysis of movie review, then such words are to be added to the database.

Lexicon Based techniques work on an assumption that the collective polarity of a sentence or documents is the sum of polarities of the individual phrases or words. In the seminar ROMIP 2012 the lexicon based method proposed in [14] was used. This method is based on emotional research for sentiment analysis dictionaries for each domain. Next, each domain dictionary was replenished with appraisal words of appropriate training collection that have the highest weight, calculated by the method of RF (Relevance Frequency) [15].

Performing sentiment analysis by various approaches will produce different results. Each approach has its own pros and cons. By considering the key factors like performance, efficiency, and accuracy, the machine learning approach yields the best result and most of the work has been done in this approach.

III. Proposed System

Travel planning and hotel booking on website has become one of an important commercial use. Sharing on web has become a major tool in expressing customer thoughts about a particular product or Service. The proposed system is to develop a Category based Sentiment analysis on travel and tourism domain.

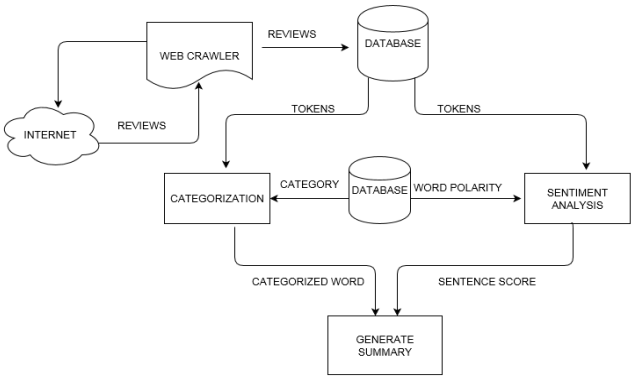

Fig 1: System architecture

First major process is to collect the corpus mainly the travel reviews from the major site. Next step is sentence splitting and tokenization. Tokenization mainly deals with the splitting of sentence into various tokens. After tokenization, stop words are removed to generate keywords. Word polarity of each keyword is check with the pre-stored positive and negative word. If the word does not match with either positive or negative then it can be consider as neutral word. If the word matches to positive then that word is marked as positive polarity. In similar fashion the negative words also checked. For example (good) is marked as positive and (bad) is marked as negative. Like this, positive and negative values are assigned to various words which convey either positive or negative sentiment. Each word is marked its polarity until the last word is encountered. Thus the sentiments of each word are obtained for this input sentence. After assigning the sentiments to each word, negation rule for finding the overall Sentiment of the given text is applied. If a negative word is encountered, then the count of negative value is incremented by one. This same process is carried out in the case of positive word also. The word which does not hold either negative or positive value, then count of neutral value is incremented. The next process is to find the overall sentiment of the text, for this process, the count of negative and positive word is taken. After assigning the sentiments to each sentence, next step is to find out the category from the generated keywords. Each keyword is check with the pre-defined category words. For example keyword weather, rain, sunny etc. mapped to the category climate. In similar fashion each sentence is categorized. Finally with the generated sentence polarity and category, the sentiment is generated for the whole collected review as a summary using pre-defined sentences.

The program has been done using the python script. The input given to the analysers as text and the output obtained are in form of summary as shown below. The Graphical User Interface (GUI) of the program is designed in python.

Fig 2: System Output

IV. Performance Measure And Evalution

This chapter describes about the experimental set up created for developing and testing the system.

IV.1 Experimental setup

The program has been done using the python script with Djnago web frame. The name of the place where the user wishes to visit is the input given to the analyser and the output obtained is in a categorized form. The Graphical User Interface (GUI) of the program is designed in python.

IV.2 Experimental Analysis and Results

There are two basic functions being performed in our project which are sentiment analysis and categorization. They have their own different methods hence they are to be evaluated individually. The first step in sentiment analysis where the reviews are obtained from the Trip advisor website is classified into positive, negative and neutral using a classifier. A Sample of 60 reviews is used which is then tested manually for accuracy. A threshold is set and the actual classification into positive, negative and neutral are done for sentiment analysis. Similarly, this approach is used for categorization.

- ***Recall***

Recall measures the completeness, or sensitivity, of a classifier. Higher recall means less false negatives, while lower recall means more false negatives. Improving recall can often decrease precision because it gets increasingly harder to be precise as the sample space increases.

Recall = tp/(tp + fn) (1)

Where tp and fN are the numbers of true positive and false negative predictions for the considered class. The result is always between 0 and 1. The recall for a class is shown as:

• Hospitality= 5/(5+3)=0.63
• Climate= 7/(7+4)= 0.64
• Transportation=10/(10+6) = 0.62
• Food and Accommodation= 11/(11+5) = 0.69
• Entertainment= 5/(5+2) = 0.71

Average Recall =0.658

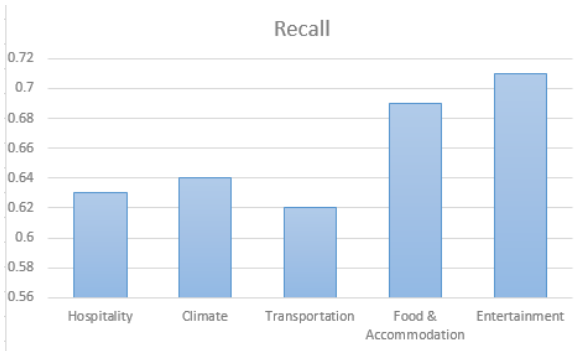

Fig 3: Recall

- **Precision:** Precision measures the exactness of a classifier. A higher precision means less false positives, while a lower precision means more false positives. This is often at odds with recall, as an easy way to improve precision is to decrease recall.

Precision = tp(tp + fp) (2)

Where tp and fp are the numbers of true positive and false positive predictions for the considered class. The result is always between 0 and 1. The precision for a class is shown as:

• Hospitality= 5/(5+2)=0.71
• Climate= 7/(7+3)= 0.70
• Transportation=10/(10+2) = 0.83
• Food and Accommodation= 11/(11+4) = 0.73
• Entertainment= 5/(5+1) = 0.83

Average Precision =0.76

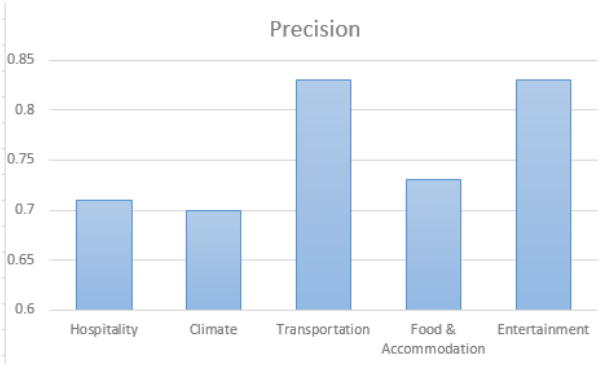

Fig 4: Precision

- *F-Measure Metric*

Precision and recall can be combined to produce a single metric known as F-measure, which is the weighted harmonic mean of precision and recall

F-Measure = 2 *(precision* recall)/(precision + recall) (3)

F-Measure is given as = 0.70

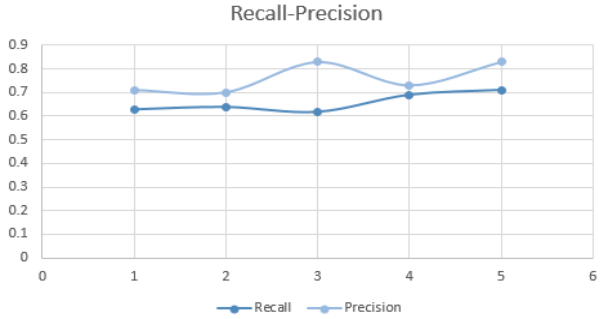

Fig 4: Recall-Precision

- *Accuracy*

Accuracy is the overall correctness of the model and is calculated as the sum of true positive and true negative divided by the total number of classifications.

Accuracy of system = (38+14)/60= 0.86

Sentiment Analysis done for the above data is depicted in figure below

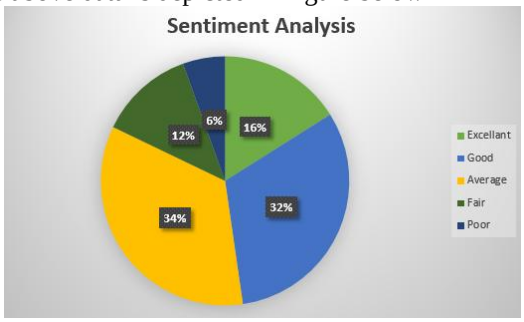

Fig 4: Sentiment Analysis

In future this project can be extended by including the popular tourist destinations in the continent. Here accuracy depends on Sentiment analyser, as well as the size of corpus. So there is a great chance for future works on enriching the analyser and the corpus. By empowering these better performance can be achieved. Since it is time consuming it is considered as tedious task

V. Conclusion

The aim of this report is focused mainly to develop a system that performs category based sentimental analysis on travel and tourism domain. In literature survey many machine learning methods like SVM, NB, Maximum Entropy methods are reviewed. Semantic analysis of the text is of great consideration and researches are going on for better analysis method.. In the world of Internet all information can be obtained from the sites.

The majority depend on the social networking sites to get their valued information. So by analysing the reviews on these blogs will yield a better understanding of products and services.

References

Journal Papers:

[1] Neha S. Joshi, Suhasini A. Itkat, " A Survey on Feature Level Sentiment Analysis" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5422-5425.

[2] He Y., "Incorporating sentiment prior knowledge for weakly supervised sentiment analysis ", ACM Transactions on Asian Language Information Processing, Vol. 11(2).

[3] N. Veeranjaneyulu, Akkineni Raghunath, B, Jyostna Devi, Venkata Naresh Mandhala, "Scene Classification Using Support Vector Machines With Lda " journal of theoretical and applied information technology 31 may 2014. Vol. 63 No.3

[4] Ankush Sharma, Aakanksha, "A Comparative Study of Sentiments Analysis Using Rule Based and Support Vector Machine", IJRCCE Vol. 3, Issue 3, March 2014.

[5] A. Tamilselvi, M. ParveenTaj, "Sentiment Analysis of Micro blogs using Opinion Mining Classification Algorithm " International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Volume 2 Issue 10, October 2013.

[6] Swati A. Kawathekar1, Dr. Manali M. Kshirsagar, "Movie Review analysis using Rule-Based & Support Vector Machines methods", IOSR Journal of Engineering Mar. 2012, Vol. 2(3), March. 2012, pp: 389-391.

Proceedings Papers:

[7] Devika M D, Amal Ganesh, Sunitha C ,Sentiment Analysis: A Comparative Study on Different Approaches. Procedia Computer Science87 2016, Elsevier 44-49

[8] P. Saloun, M. Hruzk and I. Zelinka, "Sentiment Analysis e-Bussines an e- Learning Common Issue," ICETA 2013 ,11th IEEE International Conference on Emerging eLearning Technologies and Applications, Stary Smokovec, The High Tatras, Slovakia, October 24-25, 2013.

[9] Pablo Gamallo, Marcos Garcia, "Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets" Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 171–175, Dublin, Ireland, August 23-24 2014.

[10] Ankitha Srivastava, Dr.M.P. Singh, "Supervised SA of product reviews usin Weighted k-NN Algorithm," 2014 11th International Conference on Information Technology.

[11] Kerstin Denecke, "Using SentiWordNet for Multilingual Sentiment Analysis," ICDE Workshop 2008, 978-1- 4244-2162- 6/08/ 2008 IEEE

[12] Brett W. Bader, W. Philip Kegelmeyer, and Peter A. Chew "Multilingual Sentiment Analysis Using Latent Semantic Indexing and Machine Learning," 2011 11th IEEE International Conference on Data Mining Workshops.

[13] Lizhen Liu, Xinhui Nie, Hanshi Wang, "Toward a Fuzzy Domain Sentiment Ontology Tree for Sentiment Analysis," 5th Internationa Congress on Image and Signal Processing (CISP 2012).) 2012.

[14] Blinov P. D. ,Klekovkina M. V. , Kotelnikov E. V. , Pestov O. A." Research of lexical approach and machine learning methods for sentiment analysis", Vyatka State Humanities University, Kirov, Russia.

[15] Klekovkina M. V., Kotelnikov E. V., "The automatic sentiment text classification method based on emotional vocabulary" , Digital libraries: advanced methods and technologies, digital collections (RCDL-2012) , pp. 118–123

[16] Lan M., Tan C. L., Su J., Lu Y. (2009), "Supervised and traditional term weighting methods for automatic text categorization", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31(4), pp. 721–735.