International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 9, Number 1 (2019), pp. 5-10
© International Research Publications House
http://www. irphouse.com

Automatic Sentiment Analysis of User Reviews

Prof.(Mr.) Prashant Kanade
Department of Computer Engineering,
Vivekanand Education Society's
Institute of Technology, Mumbai, India

Lakhan Rangwani
Computer Engineering,
Vivekanand Education Society's
Institute of Technology, Mumbai, India

Pritee Wadhwa
Computer Engineering,
Vivekanand Education Society's
Institute of Technology, Mumbai, India

Jitesh Watwani
Computer Engineering,
Vivekanand Education Society's
Institute of Technology, Mumbai, India

Nitin Hazarani
Computer Engineering,
Vivekanand Education Society's
Institute of Technology, Mumbai, India

Abstract

As there are large number of unstructured reviews available online for various services which can be useful for new users to make the correct choice. To organize these reviews, sentiment analysis is proposed. The sentiment analysis system can be developed using machine learning approaches as well as lexicon based approaches. The proposed system will take into consideration all the reviews and then it will classify them using the SVM(support vector machine) classifier. After which, the system will generate the generalized rating for that service as an output.

Keywords: Sentiment analysis, machine learning, lexicon based, unstructured reviews, SVM classifier

I. INTRODUCTION

In recent years, various online services have gained popularity all over the world. One can say that travel and tourism is one of those services. Most of the tourists like to make their stay arrangements prior to the trip to reduce the overhead. The large number of reviews available online for various hotels plays an important role in helping the tourists to make the correct choice of hotel. However it is not possible for the user to read every review. So sentiment analysis of reviews is done.

Moreover, people can easily read through the reviews if they speak that particular language and from that they can deduce whether the writer of the review had a positive or negative opinion for the service. However, it is not possible for the machine to read the language and deduce the opinion of the writer. This problem can be solved with the help of sentiment analysis by using natural language processing which will identify the keywords indicating the opinion of the writer.

The sentiment analysis can be defined as the measure of satisfaction of the customers from the services provided to them. The sentiment analysis extracts the opinion of the customers from the reviews given by them which are useful for the new customers. The opinions will then be classified as- positive, neutral and negative. Furthermore, the scope of sentiment analysis can be at document level, sentence level or sub-sentence level.

II. LITERATURE REVIEW

A. Automatic Sentiment Analysis of User Reviews

Abinaya R, Aishwaryaa P, Baavana S, Thamarai Selvi N.D. in [1] have performed the sentiment analysis of the reviews for the videos present in their database. The entire process is divided into three steps- review extraction, review evaluation and sentiment graph visualization. Word clustering is performed to process the stored reviews and then the relevant phrases, words are classified as positive, neutral or negative using the SVM(support vector machine) algorithm and accordingly, the sentiment score is calculated. The words without any meaning are treated as hypothetical words. The supervised learning algorithm classifies these words iteratively in order to obtain a proper score. By taking into consideration these scores, the sentiment bar graph is generated.

B. Challenges and Techniques for Sentiment Analysis: A Survey

K.S. Ilakiya, M. Lovelin Ponn Felciah in [2] have described various challenges for sentiment analysis which includes domain dependency, detection of spam and fake reviews, classification filtering. The process of sentiment analysis have been divided into six tasks which involve entity, aspect, opinion holder extraction and their categorization, time extraction which includes times when opinions are given and then different time formats are standardize, aspect sentiment classification to classify an opinion as positive, neutral or negative and then the generation of opinion quintuple. There is a comparative study of various supervised and unsupervised learning methods for document level sentiment classification. Naive bayes classification, maximum entropy rule and SVM are tested and compared amongst each other, of which SVM classifier achieved the best performance. These same techniques can also be applied to sentence level sentiment analysis. The sentences are identified as objective or subjective, from which the opinion words are extracted and then their polarity is calculated.

C. Sentiment Analysis based on Support Vector Machine and Big Data

The paper in [4] determines the valency of text. As the existing methods for analysing the text requires complex text preprocessing and are relatively expensive. Moreover, the optimization methods also create problems as they are often language dependent. The main aim is to find the language independent method of classification for sentiment analysis. Due to this, large datasets having texts with positive and negative valence were used for the purpose of testing and training on four different languages which include english, german, czech, spanish. For training, the first half of the dataset was used and the remaining half was used for testing purpose. In order to reduce the overhead of labeling the large datasets, they were downloaded from the web pages along with the product rating. The english language gave the best accuracy of classification. Moreover, the SVM classifier achieved better accuracy than k-NN(k-nearest neighbour) classifier when they were experimented on the dataset with 7000 samples.

D. Sentiment Analysis using Support Vector Machine

This paper implies the different levels on which the sentiment analysis can be done-sentence level, document level, phrase level, word level. In order to identify multiple sentiments within a sentence, we can combine phase level categorization with the sentence level classification as in [7]. Document level sentiment analysis is challenging and it can be accomplished using two approaches- term counting and machine learning approach. A sentiment measure is generated in the term counting approach by calculating the positive and negative terms. However, machine learning approaches have proven to provide better performance than the term counting approach. The process of sentiment classification in [7] is divided into various steps. Initially, preprocessing of text is performed to reduce the noise in the text in order to enhance the performance of the classifier. Further, the transformation is done which involves calculation of weight of each word in the corpus using TF-IDF(term frequency-inverse document frequency). There is feature selection step which reduces the amount of data under consideration to make classifiers more effective. Then the text classification process is carried out using SVM algorithm by classifying the text into positive and negative classes. In order to achieve this, the input document is first converted into the format suitable for the machine.

III. PROPOSED SYSTEM

The system will deliver variety of features which includes- sentiment analysis of the hotel reviews for the hotel selected by the user, recommendation of the tourist places nearby selected hotel, recommendation of similar categories of hotels.

The first step involves creation of web based UI which will take a particular hotel name as the input from the user and then by using web crawling technique like BeautifulSoup in python, the reviews about that hotel will be searched on different web sources and then extracted in a .csv file.

Further, the next step will be preprocessing of reviews which involve following operations- tokenization, stop words removal, stemming and POS(part-of-speech) tagging. Tokenization is splitting the text into words and then discarding the non-relevant words like pronouns, prepositions and articles. There are certain words in the english language such as "I", "it", "the", "of", which do not carry any meaning. Stop words removal is the process of removing these words. Stemming is the process in which the slang words and the words which are synonyms will be replaced by their root meaning words. After performing POS tagging, feature selection is done to reduce the amount of data to be investigated and also to identify relevant features for the consideration in the classification process. This data will be fed to the SVM classifier. The SVM classifier will classify the given set of words into three different categories- positive, neutral or negative and also, it will find out their corresponding polarity. The overview of the system is as shown in Fig. 1.

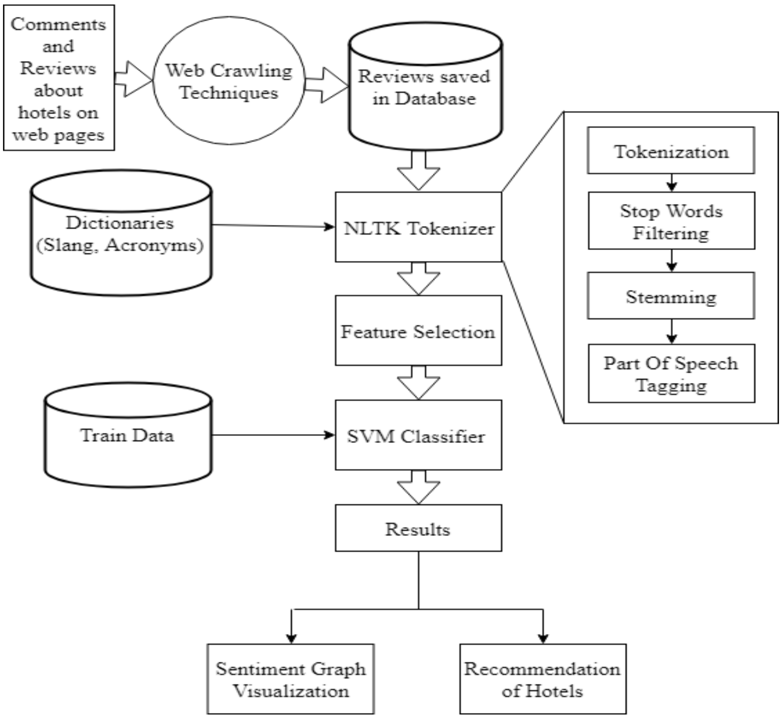

Fig. 1 Proposed System Architecture

SVM algorithm is supervised learning technique which first constructs the hyperplane and then maps the data points also known as vectors, on either side of the hyperplane and then collectively predicts the final output. Along with the linear classification, it can efficiently perform non-linear classification as well by using kernel trick which is implicitly mapping inputs into high-dimensional feature spaces.

After performing the sentiment analysis, a visual sentiment graph will be generated as the final output and also the polarity will be displayed in the output after considering the text valence in the SVM classifier. By taking into consideration the hotel name and its location as entered by the user, the recommendation engine will recommend the other similar nearby hotels. This will help the user to make an informed decision.

IV. CONCLUSION

The proposed sentiment analysis system in this paper aims to help the users to make the correct choice of hotel according to their needs by considering the reviews of previous users. However, sentiment analysis is also very significant for an organization because this will help them to know about their areas of improvement. There are multiple approaches available for sentiment analysis such lexicon based, machine learning based and the hybrid(combination of both) approach. From the comparative study of various approaches, it can be concluded that SVM algorithm is the most widely used and gives better performance. Moreover, in the taxonomy of sentiment analysis methods, SVM comes under linear classification which is a supervised learning technique in machine learning approach.

REFERENCES

[1] Abinaya R, Aishwaryaa P, Baavana S, 2016, “Automatic Sentiment Analysis of User Reviews”, IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development.

[2] K.S. Ilakiya, Mrs. M.Lovelin Ponn Felciah, 2015, “Challenges and Techniques for Sentiment Analysis: A Survey”, International Journal of Computer Science and Mobile Computing, Vol. 4, Issue. 3, pg. 301-307.

[3] Hailong Zhang, Wenyan Gan, Bo Jiang, 2014, “Machine Learning and Lexicon based Methods for Sentiment Classification: A Survey”, 11th Web Information System and Application Conference.

[4] Lukas Povoda, Radim Burget, Malay Kishore Dutta, 2016, “Sentiment Analysis based on Support Vector Machine and Big Data”, IEEE.

[5] Esha Tyagi, Dr. Arvind Kumar Sharma, 2017, “An Intelligent Framework for Sentiment Analysis of Text and Emotion”, International Conference on Energy, Communication, Data Analytics and Soft Computing.

[6] Munir Ahmad et al, 2018, “Sentiment Analysis using SVM: A Systematic Literature Review”, International Journal of Advanced Computer Science and

Applications, Vol. 9, No. 2.

[7] Aamera Z. H. Khan et al, 2015, '"Sentiment Analysis using Support Vector Machine", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, Issue 4.

[8] Devendra Kamalapurkar et al, 2017, "Sentiment Analysis of Product Reviews", International Journal of Engineering Sciences and Research Technology.