

Study of Sentiment Analysis on Reviews

Bhupendra Singh Shaktawat¹, Sharad Sarangkar², Shivansh Sharma³, Vikrant Dubole⁴, Varisha Khan⁵,
Prof. Jayashree Chaudhari⁶

^{1,2,3,4} Student, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegoan, Savitribai Phule Pune University, Pune, Maharashtra, India

⁵Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegoan, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT

Sentiment analysis and opinion mining is the area of study that analyses people's opinions, sentiments, attitudes, and emotions from written language. Sentiment analysis systems are being applied in almost every business and social domain which helps them to analyse customer satisfaction of their product because opinions are central to almost all human activities and are key influences of our behaviours. The suggested framework for Movie Review consists mainly of data collection and pre-processing, and measurement of customer satisfaction. In Data collection and pre-processing stage, text mining is utilized to compile customer-review-based dictionaries of attributes and sentiment words. Then, using sentiment analysis, sentiment scores for attributes are calculated for each Movie Review. An empirical case study will be conducted on customer reviews on movies. We believe that the our proposed customer review based approach not only saves time and effort in measuring customer satisfaction, but it also captures the real voices of customers.

Keywords : Customer Satisfaction, NLP, HSWN, Customer Review, Sentiment Analysis, ML.

I. INTRODUCTION

Sentiment analysis is computational study of people's opinions, sentiments, emotions, and attitudes.

Sentiment Analysis finds orientation of a person opinion or feelings over an entity. It is a task under natural language processing. It deals with analysing personal emotions, feelings, attitude and opinion of a speaker or a writer over an object. The primary target of sentimental analysis is to find the sentiments expressed by person over an information or object.

Hindi is the 4th highest speaking language in the world. The web compared to previous years is currently enriched with non-English languages too.

There exist very small no of systems which calculate sentiment associated with Hindi text as Sentiment Analysis is highly difficult for Hindi language due different complexity associated with Hindi text. Well annotated standard linguistic data are still not available for Hindi language. Hindi language lacks availability of efficient resources like parser and tagger which are essential for extracting sentiment. HindiSentiWordNet (HSWN) like well know English SentiWordNet is available but consists of limited numbers of adjectives and adverbs, which still needs to improvement to achieve higher accuracy.

There are many situations where same words may be used in multiple contexts. Context dependent word mapping is still a difficult task, error prone and it

requires manual efforts to find accurate polarity of word.

A framework for performing sentiment analysis on Reviews of Hindi and English language is presented in this paper. Section 2 shows related work done in this field. In section 3 shows our proposed system and section 4 gives the conclusion about the work.

II. LITERATURE SURVEY

In this section we will see the relevant past literature of research work done in the field of sentiment analysis for Hindi language.

Opinion Mining System is proposed by authors named as “Hindi Sentiment Orientation System” which is based on Hindi language. Unsupervised approach which based on using dictionary is used to determine the polarity of reviews written by users in Hindi language. Many challenges like negation associated in text which reverse the sentiment are also handled. The accuracy of system is evaluated by using 50 sentences of movie reviews and their result showed the accuracy of 65% in finding sentiment associated with text.

This fascinating problem is increasingly important in businesses and society. Sentiment analysis is a kind of Natural Language Processing (NLP) for tracing the mood of the public about a particular product or topic. The application which considers every individual as a separate class and predict the next work on the basis of previous content available is known as language modelling.

For classification applied in machine learning mainly two steps are followed. Utilising the training data set for learning the model is the first step. In the second step the trained model is applied to the test data set.

From paper[11] presented „Pulse“ a prototype system for mining topics and sentiment orientation from free text customer feedback. Blogs, newsgroups, feedback email from customers, and web sites that collect

product reviews these all are source of free text customer feedback. The proposed system is designed to handle the free form information of the customer feedbacks as the sources of information are less structured than traditional surveys. A machine learned sentiment classifiers and clustering technique were used in the proposed method. Sentiment and topic detections are not performed at the document level it is done at sentence level. The data set used for the survey contained almost 900,000 sentences in total. Sentiment analysis was performed using 3000 randomly selected sentences from data set. Each sentence is classified as positive, negative and others, where other category contained both positive and negative sentiment and sentences with no complex sentiments. Training of the sentiment classifier was done using 2500 sentences and the remaining 500 sentences were reserved for test set. Results reflect the efficiency of the proposed system.

Miniqing Hu et al., [12] performed mining and summarization process to all the customer reviews of a product. The proposed process was carried out in three steps: 1. The features of product commented by the customer in the review are mined. Natural language processing(NLP) and Data mining techniques are used for mining. 2. The opinions in the review are identified and the opinions are classified as +ve or -ve. Set of adjectives words are identified and semantic orientation of the opinion words is determined. SentiWordNet can be used to identify the semantic orientation and the opinion orientation of each sentence. 3. Summarize the results. The objective of the study is to perform summary of a large number of customer reviews of a product sold online.

Qui et al., [13] analyzed the problems related to opinion mining such as opinion lexicon expansion and opinion target extraction. Opinion targets are entities and there attributes on which opinions have been expressed. The list of opinion words such as good, bad, lovely, poor used to indicate positive, negative sentiments is Opinion lexicon. The link between the opinion words and targets Syntactic relations are

identified using dependency parser which is based on bootstrapping. The process uses supervised methods, opinion word seeds are used in the initial opinion lexicon. Bootstrapping process is started using the initial opinion lexicon. Double propagation method is used as information are propagated back and forth between opinion words and target.

Lie Zhang et al., [14] identified domain dependent opinion words. Noun,noun phrases that indicate the product feature which implies opinions are found using a feature based opinion mining model. Two steps are used for identifying the noun product feature which means the positive or negative opinion. In the Candidate identification step sentiment context of each noun feature is determined. And also a list of candidate features with positive opinions and list of candidate features with negative opinions is produced. In pruning step noun product feature is directly modified into positive and negative opinion words. Opinion lexicon compiled by Ding et al. It was used to identify the opinion polarity on each product feature in a sentence. For a sentence *s* which contains a product feature *f*, opinion words in the sentence are first identified by matching with the words in the opinion lexicon. An orientation score for *f* is computed and the orientation of the positive word is assigned the score of +1, and a negative word is assigned the score of -1. On summing up of all the scores, if the final score is positive(+ve), then the opinion on the feature in *s* is positive(+ve). If the score is negative(-ve), then the opinion on the feature in *s* is negative(-ve).

Xiaowen Ding et al., [15] proposed a holistic lexicon-based approach which uses external indications and linguistic conventions of natural language expressions to determine the semantic orientations of opinions. Opinion words which are context dependent are easily handled which is an advantage. The algorithm used linguistic patterns to deal with special words, phrases. Based on this technique researchers built a system called Opinion Observer. Experiments using product review dataset was highly effective. Multiple

conflicting opinion words in sentence were also dealt with efficiently.

Table 1

Sr. No.	Paper Name	Method Proposed	Limitations
1.	Sentiment analysis of mobile network [6]	Supervised learning, naïve Bayes, bag of words	Execution time is more
2.	Sentiment Analysis for twitter data [2]	Lexicon approach	Less accuracy
3.	Sentiment analysis experiment [3]	Natural language processing	Large data set not accepted.
4.	Sentiment analysis for movie review [10]	OPEN NLP library	No hindi reviews are considered
5.	Pulse: Mining Customer Opinions from Free Text Natural Language Processing [11]	Component based mining	Limited to certain products only
6.	Mining and Summarizing Customer Reviews [12]	NLP and search methodology	Based on only English reviews

III.CONCLUSION

Based on above study and survey on sentiment analysis about different languages and using different

methodologies we proposed a system in which the system allows finding sentiment associated with review where overall polarity of the review is classified as positive, negative or neutral using HindiSentiWordNet and RNN. It also includes Hindi reviews.

IV. ACKNOWLEDGEMENT

It gives us a great pleasure in presenting the paper on "SENTIMENT ANALYSIS ON REVIEWS". We would like to take this opportunity to thank Dr. Pankaj Agarkar, Head of Computer Engineering Department, DYPSOE, Pune for giving us all the help and support we need during course of the Paper writing work. We are really grateful to him. Our special thanks to Dr M. Z. Shaikh, Principal DYPSOE who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

V. REFERENCES

- [1]. Bing Liu, "Exploring User Opinions in Recommender Systems", Proceeding of the second KDD workshop on Large Scale Recommender Systems and the Netflix Prize Competition", Aug 24, 2008, Las Vegas, Nevada, USA.
- [2]. Xiaowen Ding. A Holistic Lexicon-Based Approach to Opinion Mining- WSDM'08, February 11-12, 2008, Palo Alto, California, USA. 2008 ACM 978-1-59593-927- 9/08/0002.
- [3]. R. B. W. N. Jeonghee Yi, T Nasukawa, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," ICDM03, IEEE, 2003
- [4]. D. A. A. S.M. Shamimul Hasan, "Proximity-based sentiment analysis," IEEE, 2011
- [5]. S. Mohammad, B. Dorr, and C. Dunne, "Generating HighCoverage Semantic Orientation Lexicons fom Overly Marked Words and a Thesaurus", In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp. 599-608, 2009.
- [6]. D. Rao and D. Ravichandan, " Semi-supervised polarity lexicon induction", In Proceeding of 12 conference of the European Chapter of the Association for Computational Linguistics, pp. 675-682, USA, 2009.
- [7]. P.J. Stone, "The General Inquirer: A Computer Approach to Content Analysis", The MIT Press, 1966.
- [8]. K. Dave, S. Lawrence and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews", In Proceedings of 12th International Conference on World Wide Web, pp. 519- 528, Hungary, 2003.
- [9]. S. M. Kim and E. Hovy, "Determining the sentiment of opinions", In Proceeding of COLING, pp. 1367-1373, 2004.
- [10]. A. Das and S. Bandyopadhyay, "SentiWordNet for Indian languages," Asian Federation for Natural Language Processing, China, pp. 56-63, August 2010.
- [11]. Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger, "Pulse: Mining Customer Opinions from Free Text Natural Language Processing", Microsoft Research, Redmond, WA 98052, USA.
- [12]. Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews", Proceedings of the tenth ACM SIGKDD International conference on knowledge discovery in data mining (KDD-2004), August 22-25.
- [13]. Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen. "Opinion Word Expansion and Target Extraction through Double Propagation." Computational Linguistics, March 2011, Vol. 37, No. 1: 9.27.
- [14]. Lie Zhang and Bing Liu. "Identifying Noun Product Features that Imply Opinions." ACL-2011 (short paper), Portland, Oregon, USA, June 19-24, 2011.
- [15]. Xiaowen Ding. A Holistic Lexicon-Based Approach to Opinion Mining- WSDM'08,

February 11-12, 2008, Palo Alto, California,
USA.2008 ACM 978-1-59593-927- 9/08/0002.

Cite this article as :

Bhupendra Singh Shaktawat, Sharad Sarangkar, Shivansh Sharma, Vikrant Dubole, Varisha Khan, Prof. Jayashree Chaudhari, "Study of Sentiment Analysis on Reviews", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4 Issue 8, pp. 55-59, September-October 2019.
Journal URL : <http://ijsrcseit.com/CSEIT194814>