

# Sentiment Analysis and Text Summarization of Online Reviews: A Survey

Pankaj Gupta, Ritu Tiwari and Nirmal Robert

**Abstract**—Sentiment analysis and text summarization has evoke the interest of many scientists and researchers in last few years, since the textual data has become useful for many real world applications and problems. Sentiment analysis is a machine learning approach in which machine learns and analyze the sentiments, emotions etc about some text data like reviews about movies or products. These reviews are increasing day by day, due to which summarization of reviews comes in role where summarized form of text in needed, which provides useful information from the large number of reviews. It is very difficult for a human being to extract useful data or summarize it from the very large document. In Text summarization, importance of sentences is decided based on linguistic features of sentences. This paper provides the comprehensive overview of recent and past research on sentiment analysis and text summarization and provides excellent research queries and approaches for future aspects.

**Index Terms**—Text summarization, Sentiment analysis, Information extraction, Natural language processing, Opinion mining.

## I. INTRODUCTION

SENTIMENT ANALYSIS AND TEXT SUMMARIZATION uses natural language processing, machine learning, text analysis, statistical and linguistics knowledge to analyze, identify and extract information from documents[1]. It is generally used to determine the emotions, sentiments and summarization from large data and that information can be used to make some predictions.

There are a lot of sources from where one can extract reviews of different movies or products. Data can be movie reviews which we can get from IMDB(Internet Movie Database), product reviews provided by Amazon or Flipkart etc, social comments or tweets from Facebook or Twitter. Sentiment analysis helps to analyze the behavior of reviews whether it is positive, negative or neutral. Text summarization helps to extract useful information from these large data which can be used to make the summary of public opinions. A lot of research work is done since late 90s on sentiment analysis and text summarization.

In this survey, we aim to review how different methods have been used to build summarization systems and perform reviews analysis. The rest of the paper is organized as follows: Section II is to discuss about past work done on the sentiment analysis and text summarization. Section III elaborates the techniques used for the reviews analysis and summarization. and also the discussion of research gaps and future work of past work is done. Finally Section IV concludes the survey.

## II. PAST RESEARCH WORK

### A. Sentiment analysis using machine learning technique[1]

The author elaborates the studied the measurement of effect of applying machine learning techniques on sentiment analysis problems.

The main challenge of this work is to distinguish sentiment analysis from traditional topic-based classification, i.e. identifying by keywords alone, sentiment can be expressed in more precise manner. Sentiment require more understanding than normal topic based classification, it requires more understanding of sentence sense. This work basically consist of two machine learning methods Naive Bayes Classifier and Support machines Machines(SVM).

The results generated by techniques of machine learning are more improved and precise in comparison to earlier techniques like human generated baselines. As comparing in between both the techniques of machine learning SVMs results are better than Naive Bayes Classifier, although difference weren't very large.

### B. Determining the sentiment of Opinions[2]

The work proposed a method to find the sentiments about the topic in the textual data, and also identify the people who hold each sentiment.

The system proposed in the paper operates in four steps. First it selects the sentence which contains the topic and the holder persons. Then holder-based regions of opinions are delineated. Then further polarity of each sentiment holding words is calculated using sentiment classifier individually. Finally, system combines the sentiments to generate the holder's sentiment for whole sentence.

In the Fig. 1 is the system architecture which represents the flow of methodology used in the sentiment analysis by determining the opinion using WordNet.

The author concluded that the results are encouraging and also be generated even with simple architecture, model and also with small amount of seeding efforts.

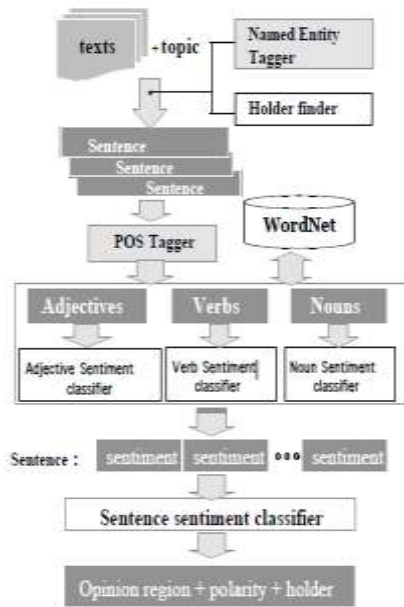


Fig. 1. System Architecture represent methodology used in opinion mining using WordNet[2]

### C. Sentiment analysis of Movie Reviews[3]

In this paper author proposed two methods for identifying sentiment(positive, negative and neutral) of movie reviews. They basically examined the effect of polarity and valence shifter in classifying procedure of reviews.

They examined three types of shifters: negation, intensifiers and diminishes. To reverse the polarity of sentence negation is used and to increase or decrease the degree of positive and negative of individual term the intensifiers and diminishes had been used.

Results of using this shifters(polarity and valence) increase the accuracy of classification very high than the earlier proposed methods.

### D. Sentiment classification considering Negation[4]

An approach to classify the information considering the negation and transition of sentence into sentiment analysis has been proposed in this paper.

The approach classify the sentences in reversed sentences and non-reversed sentences parts, which represent them as two different bag-of-words(BOW) model. Then for classification of these two bag-of-words three general machine learning strategies are presented.

Above Fig. 2 shows the proportion of negation sentences in all positive and negative reviews examined in each domain which conclude that negation sentences occurs more often in reviews the positive ones. The results obtained considering the negation sentences into account are very good and performs robustly better than traditional machine learning approaches which consider bag-of-words.

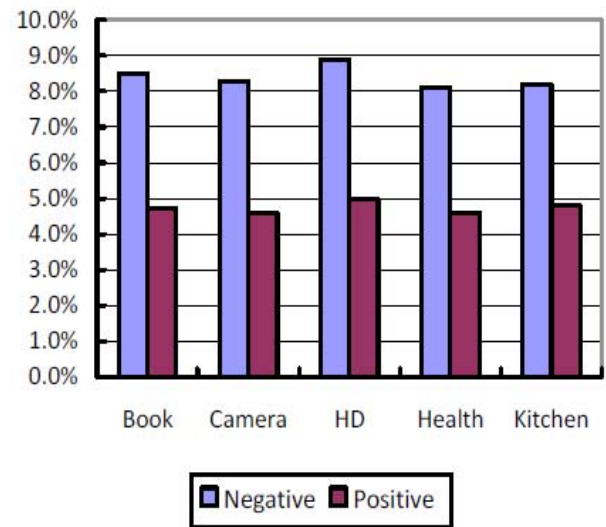


Fig. 2. The proportion of negation sentences in negative and positive reviews[4]

The future works of this paper emphasis the more effective approach to recognize sentiment-reversed sentences.

### E. Dual Sentiment Analysis[5]

The author proposed a model for sentiment analysis which is Dual sentiment analysis over the earlier methods like Bag-of-words, traditional approach for sentiment analysis.

They first propose the novel-based approach for the sentiment reversed reviews for each training and test reviews. On this basis they proposed the dual training of original and reversed for each review using a sentiment classifier(logistic regression classifier).

	Review Text	Class
Original review	<i>I don't like this book. It is boring.</i>	Negative
Reversed review	<i>I like this book. It is interesting.</i>	Positive

Fig. 3. Reversed review generated using original review which changes its polarity[5]

In Fig. 3 the reversed review is generated using original review. Polarity has been changed from Negative to positive.

In Fig. 4 the methodology used for dual sentiment analysis is shown where dual training is done using both original and reversed reviews then run for the test data.

They also used another classifier that are Naive Bayes Classifier and Support Vector Machine(SVM) for comparing the results with their model and classifier. The Results produced by their model are more efficient and more accurate than earlier approaches.

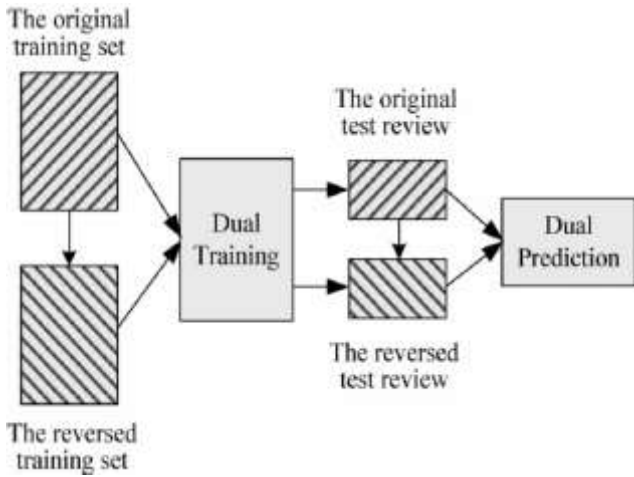


Fig. 4. The process of dual sentiment analysis[5]

#### F. Automatic Text Summarization[6]

In this paper author conferred a content-based and graph-based approach for text summarization. Automatic Text summarization helps in extraction of summary from a large document quickly. They introduce Hopfield Network algorithm for ranking text segments. In addition Hopfield Network algorithm on undirected graphs obtained best text segment ranking algorithm according to the study.

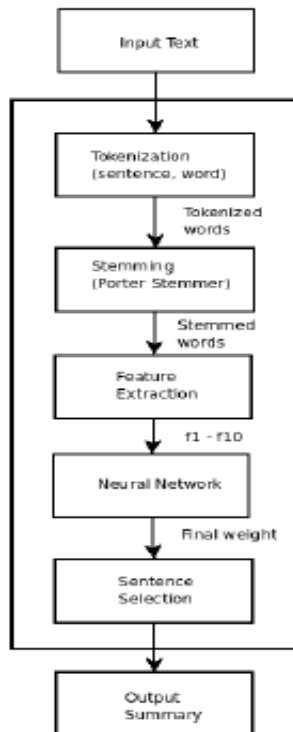


Fig. 5. Text Summarization Methodology[6]

The Fig. 5 shows the traditional methodology used by the Text summarization systems.

The Results are generated against the baseline and peer systems, performance of proposed algorithm was measured against these systems and the outcome were good and efficient.

#### G. Text Summarization based on Natural Language Processing[7]

The author conferred an automatic text summarization technique based on natural language understanding by using RST(Rhetorical Structure Theory) and CIT(Comprehensive Information Theory). The system extracts the rhetorical structure of text and also the rhetorical relations between sentences and removes the less import words from the sentence keeping only keywords and important words. At last, author studied sentences in extracted structure to develop text summarization by using CIT.

#### H. Content Based Text Summarization Systems[8]

This work emphasis that the quality of the summary achieved when one consider the combinations of sentences scoring methods which are based on textual data. Such hypothesis is evaluated by use a good advantage of three different context which are news, blogs and articles.

The results generated after the work shows the validation of hypothesis proposed and developed which techniques are more impressive in each of these contexts considered.

The future work proposed is to develop a system which discovers which combination of sentences of sentence scoring methods best summarizes it.

#### I. Automatic Text Summarization of Wikipedia Articles [9]

In this paper, the author used the extractive text summarization with Wikipedia articles as input to the system by analyzing text features and scoring the sentences.

The work contain tokenizing the sentences and perform stemming of the words, initially the text is pre-processed. Then scoring of sentences has been done using various text features. Traditional methods are used to score the sentences. This score helps in classifying the sentences that if they will in summary text.

It was realized that scoring the sentences using citations gave the excellent results.

#### J. Other Related Work

Other Related works in the field of Sentiment Analysis: Other work on sentiment analysis contain opinion extraction and semantic classification on product reviews[10] which identifies the important aspects of the reviews on a particular product and the method proposed classify the reviews in positive and negative reviews. The sentences are scored based on the importance of the sentence or review. Methods used gives better results than the traditional machine learning methods.

Effect of simple linguistic processing in sentiment analysis[11-12] is also observed on product reviews. SVM(Support Vector Machines) are used for categorizing in different features which gave better results. And also it was hypothesized that the syntactic and semantic processing plays an important role in sentiment analysis. Polarity shifting had been also implemented for sentiment classification[13-15] which depict that polarity of the words are not always same as of the sentence. In this paper they proposed the polarity model which can be trained in two different ways- word-wise and sentence-wise learning. They also combine the features used in past work

such as bag-of-words and n-grams. For small amount of training data their result and performance of system had been improved.

Other work is related to phrase-level sentiment analysis[15], in this paper they model the classifier for phrases and shown that the results will improve if one combines two method to classify the sentence. Some work also include improvement of scope of negation to classify[16]. In this paper author presents the negation detection system using features from English parser.

Other Related works in the field of Text summarization: Other related approaches for text summarization contain Hybrid approach for text summarization[17-18], in this paper a hybrid approach KCS is used to enhance the quality of summarization. The capability of summary is tested whether it increasing the text classification accuracy or not. As the result KCS perform best among all approaches.

Another work include Text summarization using Rule reduction[19]. In this paper, for analyzing the text the rule reduction techniques had been used in three stages, Token creation, Feature Identification and categorization and summarization. It produces noteworthy results. Experiment validate the selection of parameter and efficiency of approach. Another work is text summarization using WordNet[20] which includes unsupervised learning for summarizing the text. WordNet is a online semantic library. Simplified Lesk algorithm is used to evaluates the weights of all the sentences and arranges them in decreasing order of weights. Then according to the percentage of summarization the sentences are selected from the ordered list. The proposed approach gives best results up to 50% summarization of original text.

### III. DISCUSSION

We have seen that a lot of work is done in sentiment analysis and text summarization field using any textual data from social website like twitter, facebook, product reviews from Flipkart, Amazon and also reviews of movies, news, articles etc. Many techniques are devised to improve accuracy of classification of sentences and summarization of large data. The future work gives us idea about improving the classification and summarization of text and also efficiency of model and accuracy of data manipulation. Recognizing sentiment-reversed sentence can be more effective if it is done with advance machine learning approaches rather than traditional approaches. Area of sentiment classification of sentence where more improvements can be done is to classify neutral reviews and also one can increase the accuracy of positive and negative reviews detection. Negation conversion of sentence is also a field in sentiment analysis where improvement is possible. In Text summarization improvement is done by improving the accuracy of summarization of text from original input data and also a system can be modeled which discovers which combination of sentences scoring methods best summarizes it. Apart from this summary an intelligent system can also be prepared that will be dynamic, more robust in nature that help in extracting useful data from large data in faster and efficient way.

### IV. CONCLUSION

As the data is increasing day by day, due to which getting all information is almost infeasible but possible steps can be taken to get most of the useful information from it. This can be achieved through sentiment analysis which provides the emotions, sense of the sentence or the reviews of different movies or product. There are various areas in sentiment analysis field which are still untouched and lot of improvement in existing techniques can be done with correct knowledge. Similarly due to the vast textual data, it is impossible to read all the data to extract useful information for a human being. For such kind of problem, text summarization is one of the vital solution. Improvement in conversion of data to its summary can be possible with efficient models, accuracy can be achieved up to greater extent in this field too. An intelligent model can also be proposed for extraction and sentiment analysis of data using machine learning techniques. There are chances that we can get a lot of new outcomes, techniques and methods if we explore these areas. Some real world application based on sentiment analysis and text summarization can be social media monitoring, customer reviews record, survey responses, document summarization, recommender systems etc. which help the users in easy featuring of sentiments and textual data.

In this paper, survey has been done of previous work related to text summarization and reviews analysis, so that new research area can be explored by looking into the merits and demerits of the current techniques and strategies.

### REFERENCES

- [1] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*(pp. 79-86). Association for Computational Linguistics.
- [2] Kim, S. M., & Hovy, E. (2004, August). Determining the sentiment of opinions. In *Proceedings of the international conference on Computational Linguistics* (p. 1367). Association for Computational Linguistics.
- [3] Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational intelligence*, 22(2), 110-125.
- [4] Li, S., & Huang, C. R. (2009). Sentiment Classification Considering Negation and Contrast Transition. In *PACLIC* (pp. 307-316).
- [5] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi & Tao Li, "Dual Sentiment Analysis: Considering Two Sides of One Review", *IEEE Transactions. Knowledge and Data Engineering*, Vol.: 27, pp: 2120 - 2133, Aug. 2015.
- [6] Sornil, O., & Gree-Ut, K. (2006, June). An automatic text summarization approach using content-based and graph-based characteristics. In *Cybernetics and Intelligent Systems, 2006 IEEE Conference on* (pp. 1-6). IEEE.
- [7] Li, C., & Wang, S. (2006, August). Study of Automatic Text Summarization Based on Natural Language Understanding. In *Industrial Informatics, 2006 IEEE International Conference on* (pp. 712-714). IEEE.
- [8] Ferreira, R., Freitas, F., de Souza Cabral, L., Dueire Lins, R., Lima, R., Franca, G., ... & Favaro, L. (2014, April). A Context Based Text Summarization System. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on* (pp. 66-70). IEEE.
- [9] Hingu, D., Shah, D., & Udmale, S. S. (2015, January). Automatic text summarization of Wikipedia articles. In *Communication, Information Computing Technology (ICCICT), 2015 International Conference on* (pp. 1-4). IEEE.
- [10] Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528). ACM.

- [11] Na, J. C., Sui, H., Khoo, C. S., Chan, S., Zhou, Y. (2004). Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. In I.C. McIlwaine (Ed.), *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference* (pp. 49-54).
- [12] Choi, Y., & Cardie, C. (2008, October). Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 793-801). Association for Computational Linguistics.
- [13] Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35(3), 399-433.
- [14] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10<sup>th</sup> ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2004, pp. 168-177.
- [15] Li, S., Lee, S. Y. M., Chen, Y., Huang, C. R., & Zhou, G. (2010, August). Sentiment classification and polarity shifting. In *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 635-643). Association for Computational Linguistics.
- [16] Council, I. G., McDonald, R., & Velikovich, L. (2010, July). What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing* (pp. 51-59). Association for Computational Linguistics.
- [17] Abbasi, A., France, S., Zhang, Z., & Chen, H. (2011). Selecting attributes for sentiment classification using feature relation networks. *Knowledge and Data Engineering, IEEE Transactions on*, 23(3), 447-462.
- [18] Chang, T. M., & Hsiao, W. F. (2008, July). A hybrid approach to automatic text summarization. In *Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on* (pp. 65-70). IEEE.
- [19] Devasena, C. L., & Hemalatha, M. (2012, March). Automatic text categorization and summarization using rule reduction. In *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on* (pp. 594-598). IEEE.
- [20] Pal, A. R., & Saha, D. (2014, February). An approach to automatic text summarization using WordNet. In *Advance Computing Conference (IACC), 2014 IEEE International* (pp. 1169-1173). IEEE.