

# Prediction of Polysemous Words in Sentiment Analysis: A Review

Manisha Malik<sup>1\*</sup>, Neetu Verma<sup>2</sup>

<sup>1,2</sup>Dept. of Computer Science and Engineering, Deenbandhu Chhotu Ram University of Science and Technology Murthal, Sonapat, Haryana, India

\*Corresponding Author: manishamalik1706@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v8i5.100104> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Received: 08/May/2020, Accepted: 18/May/2020, Published: 31/May/2020

**Abstract:** In the last some years, new methods of communication channels have appeared and become indifferent. These communication channels are the social networking sites which have experienced an exponential growth. During the translation or communication, the problem of polysemy may cause difficulties. Therefore, there is a dire need for sentiment analysis process which can automatically extract and detect the sentiments of data extracted from micro blogging sites. It requires efficient techniques to collect a large amount of social media data and extract meaningful information from them. This paper presents a document level lexicon-based approach to detect the sentiment polarity. So, we focused on pre-processing of data. Instead of removing all polysemy, it includes some polysemous words in the complete procedure of sentiment analysis. We use specific number of polysemy words there but in future we will focus on different words and enhance the accuracy of our documentation.

**Keywords:** Sentiment Analysis, polysemy, polarity, wordnet.

## I. INTRODUCTION

In many situation’s sentiment analysis can also be termed as opinion mining. In many research extents particularly in data mining it becomes a challenging field in social media where number of applications like product ratings, analysis of feedback given, client decision making, etc [13]. With the evolution and growing popularity of social networking sites, it has become easy to gather people opinions in comment form, appraisals, posts etc. All this promotes analysis of our sentiments by giving it rich source of information like twitter. People opinions can be treated to find sentiments using sentiment analysis process. It can be performed in any data analytics programming language with the help of data analysis tools [18].

In today’s scenario the Internet is growing rapidly that network media now has been recognized as the fourth media after newspaper, television and radio. The existence of Internet has arranged a path that connect millions of people with each other beyond space and time. Nowadays, we are living in a “data age” rather in “information age”. When we increase the amount of data that is user generated on the social networking sites like twitter, Facebook, YouTube have provided several new prospects to the organizations that keeps track on the reviews of customers and their opinions about the products.

Earlier what happen, people used to gather reviews or feedback from their friends or relative before buying any product but now the scenario is different. Now people

check reviews of individuals around the globe about any product online using micro blogging sites. To capture the public opinion about a brand, we need automatic process which can automatic and analyse the Sentiment of data extracted.

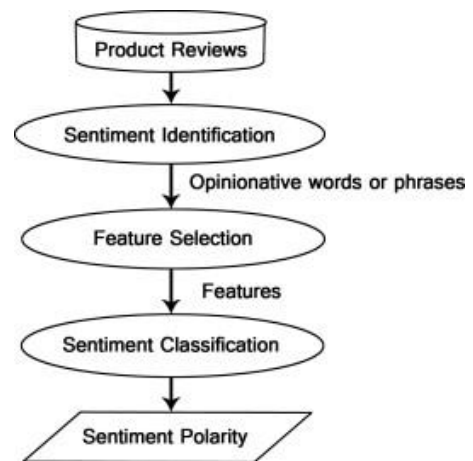


Fig 1: Structure of Sentiment Analysis

Researchers who are working in the area of NLP and text mining focusing on the sentiment analysis and their aspects. If we want to know about what people feel and think about their products then we introduce opinion mining [16]. Individuals state their opinions in every possible aspect in day to day life with the help of Social media. When we have to collect data from social media in a large amount, we must require efficient and effective techniques.

Manual extraction of facts and figures from network is very difficult work. Because these days there are billions of networking sites which have billions of active users which keep on sharing their opinions. So, analysing this huge amount of data manually is impossible. This arouses the need for a process which can automatically extract data from these sites and do analysis. This complete procedure is defined as sentiment analysis. This field is a very emerging and trending research in the era of technology. This research field comes under the domain of data mining.

## II. SENTIMENT ANALYSIS LEVELS

### 1. Document level

It is the most basic and simple level of Sentiment analysis. Entire document is occupied as input in this level and then results are produced according to it. It is generally assumed that in document level, text contains opinion about single object.

### 2. Sentence Level

Sentence level is the finest level among all three. It is the procedure the information we required is extracted from people's emotions and their opinions. Each and every sentence is measured as a separate block in this level. Sentiment polarity is calculated for each sentence word by word.

### 3. Aspect Level

Aspect level can also be defined as Feature level. In this level, there is need to identify aspect or feature which is being talked about in text extracted. For example, Beauty does not last forever. Here in this sentence, "beauty" is the feature which is being talked about and "does not last forever" is opinion.

## III. APPROACHES FOR SENTIMENT ANALYSIS

**1. Lexicon based approach:** It is a simple, feasible and real-world approach to analyse data and training is not necessary. It is also termed as knowledge-based approach. Aman Mehto [17] proposed an approach based on lexicons for performing analysis. The proposed methodology takes into account an extra feature of aspect directory. Santanu Mandal [15] future a lexicon-based approach that helps to analyse polarity of user's sentiments i.e. Positive, negative or neutral from reviews based on text classification. The lexicon-based technique determines the use of glossaries of words marked up with their Sentiment alignment.

Lexicon approach is further of two types.

**1.1 Dictionary based approach:** This approach is simple yet elegant approach used for sentiment analysis. The sentiment classification using this approach therefore depends on size of lexicons [21] as we increase the size lexicons this approach becomes more inaccurate and takes lots of time. If we have to understand that the lexicon-based technique makes use of dictionaries of

words marked up with their semantic alignment read paper [14].

Dictionary approach is based on collection of small set of opinions. This set of opinion contains small number of words. This set is then grown by finding words synonyms, antonyms using WordNet or thesaurus and then these words are added to set. This process goes on as long as no new word is found.

**1.2 Corpus based approach:** This approach is totally depending on large quantities for semantic outlines of opinion mining. The words we produced may require large amount of categorized dataset.

**2. Machine learning based approach:** This approach is based on well-known machine learning algorithms. It uses supervised and unsupervised learning to classify data. Machine learning approach is also known as statistical based approach.

Most commonly used algorithm to perform sentiment analysis are mentioned below.

**2.1 Naïve Bayes Algorithm:** Naïve Bayes algorithm comes under supervised learning. It is named after the famous Bayes theorem that was used to find the probability. Naïve Bayes probabilistic classifier is one of the renowned algorithms used for sentiment analysis. Its accuracy varies from 70- 85%.

It is based on calculation of subsequent and past probability calculations.

Po-Wei Liang et al. [8] presented a new system architecture which was able to process short and colloquial micro-blog tweets. K S Kushwanth Ram [12] proposed a sentiment analyser engine which used twitter as data source. Retrieved tweets are then classified as different polarities like positive, negative or neutral.

Purtata Bhoir [22] at aspect level proposed system aimed to recapitulate the reviews of a movie. Proposed system performed subjectivity analysis before finding the particular opinions about movie. There are two set to implement in supervised learning – training set and test set. It was observed that Naive Bayes classifier gave more accuracy as compared to SentiWordNet.

**2.2 Support Vector Machine (SVM):** It is also defined as supervised learning which is suitable for both classification and regression challenge. The main target of SVM is to find hyperplane which can best divide it in different areas. SVM is quite used on text data. Jyoti Ranteke et al. [20] proposed an approach to predict election results. Two classifiers are used in this approach. First, the results are predicted using multinomial Naïve Bayes and after that, results are predicted using Support Vector Machine (SVM).

Supervised learning's main problem is that it was not able to perform well in different fields and that leads to in

accuracy and insufficient performance in sentiment analysis and inability to deal with multifaceted sentences which doesn't provide simple analysing [10].

**3. Hybrid approach:** To increase the precision of Sentiment analysis some researchers combine the both machine learning and lexicon-based which is known as Hybrid approach. It combines the prevailing features of both approaches and then performs analysis. The collaboration of the approach actually improved the performance and accuracy. Douglas R. Rice [2] built a semi-supervised where a text is drawn from a specific dictionary in such a way that they need small quantity of coding that are basically unsupervised by nature.

#### IV. CHALLENGES AND ISSUES

Nowadays social media become very important part where people share their opinions on day to day life. It requires well-organized systems that extract meaningful information and collect the data from the social media and Sentiment Analysis takes care of this. Many researchers have worked on sentiment analysis using different techniques.

The presence of polysemous word in text document is the major factor that affect the quality of cataloguing. The problem of polysemy is also defined as the problem of interrelation that have various meaning of the same word. "Polysemy is the phenomenon of a single word having two or more meanings, no matter how meaning is defined in a given approach".

"A polyseme is a word or phrase with different, but related senses. Since the test for polysemy is the vague concept of the relatedness, judgments of polysemy can be difficult to make". It is the capacity for a phrase that have multiple senses within a semantic area. In English language most of the words are poly-sentiment in nature.

#### V. RELATED TASKS

Lesk Michael [5] proposed an algorithm to identify senses of the polysemous words known as Lesk algorithm. He used the overlay of word classification from the "Oxford Advanced Learner's Dictionary of Current English (OALD)" to disambiguate the senses of words.

Yih-Jeng LIN, Ming-Shing YU and Chin-Yu LIN in 2010 [7] projected a sophisticated approach that determine pronunciation of the polysemous words in a C2T TTS system and use Academic Sinica Balanced Corpus 3.0, a recognized Chinese corpus with segmentation information, as data that they used for their experiment. They performed their experiment by using 6 words that have polysemy. As a result, the given approach shows that the models they use attain higher accuracies than the previous model used like decision list classifier and language models.

M. Sinha [3] developed an automatic Word Sense Disambiguation for Hindi language using Hindi WordNet. For determining the senses, they used statistical technique. They made a comparison between context of polysemous word in the sentence and context raised from WordNet. The range of their accuracy was found 40-70%.

To disambiguate the polysemy word in Nepali language U. R. Dhungana and S. Shakya [4] used the adapted Lesk algorithm. They performed the experiments that contains 348 words (which includes the 59 polysemy words with different senses and context words) with test data that contains 201 Nepali sentences. The experiment shows that the accuracy of their system is to be 88.05%.

Udaya Raj Dhungana, Subarna Shakya, Kabita Baral and Bharat Sharma [1] developed a new model of WordNet. According to this model we can systematizes the polysemous word with single sense as well as different different senses with the help of clue words. With the help of new model of WordNet, the accuracy of the system increased by 3.48% in comparison with the accuracy attained in [4]. So, after that experiment they concluded that new model of WordNet gives the higher accuracy for WSD algorithms instead of the conventional WordNet.

In 2019 Andrey Kapitanov, Ilona Kapitanova, Vladimir Troyanovskiy, Vladimir Ilyushechkin and Ekaterina Dorogova [16] proposed a method of context clustering to eliminate the polysemy. By performing experiment of context clustering they stated that by selecting parameters, window width increasing and the size of clusters diminishes significantly, and the speech context becomes too narrow.

Table 1 Content of Previous Work

S.No.	Author's Name	Tool Used	Advantages	Disadvantages
1	Lesk Michael	WordNet	To disambiguate the senses of word he uses "Oxford Advanced Learner's Dictionary of Current English" (OALD)	This approach does not provide sufficient vocabulary to relate fine-grained sense dissimilarities.
2	Yih-Jeng LIN, Ming-Shing YU and Chin-Yu LIN	Chinese to Taiwanese text-to-speech (TTS) system.	The model they used to distinguish the senses of word attain higher accuracies than the other models like language model and decision list classifier	Use only 6 words so we can't determine that it will also works if try with more words.
3	M. Sinha, M. K. Reddy, P. Pandey and L. Kashyap	Hindi WordNet	They made a comparison between context of polysemous word in the	They use modest overlap method to determine simple

			sentence and context raised from WordNet. The range of their accuracy was found 40-70%.	words and fails to detect sense of complex words.
4	N. Shrestha	WordNet	Use Modified Lesk Algorithm where every single word has comparison with every word in the collection of words that formed by the gloss and synset of every possible senses of the mark word	For the target word the number of examples given was only one.
5	Udaya Raj Dhungana, Subarna Shakya, Kabita Baral and Bharat Sharma	Nepali WordNet	Use adapted Lesk algorithm to perform the experiments that contains 348 words with test data which contains 201 Nepali sentences which gives the accuracy of 88.05% to system	Accuracy level should be increased at some extent where we can detect correct sense of the word.
6	Chandrika Prasad and Jagadish S Kallimani	Kannada dictionary and shallow parser	For identifying the polysemy word, they use string-matching algorithm. The naïve-Bayesian classifier identifies the context	Text summarization left for future work.
7	Rahul Rao and Jagadish S Kallimani	Shallow parser	With the approach of POS tagging polysemy word that was collected from the large document in Kannada Language can be summarized easily	If in a sentence there are more than one polysemous word than uncertain outputs are obtained and we have to resolve these.
8	Kana Oomoto, Haruka Oikawa and Eiko Yamamoto	Word2Vec	They performed a statistical test for detecting polysemy by calculating the neighbouring uniformity of around 100 most frequently used words	To improve the accuracy of the test performed there is a need to find a way that select the number of neighbours
9	Andrey Kapitanov, Ilona Kapitanova, Vladimir Troyanovskiy, Vladimir Ilyushechkin	Context Clustering Algorithm	By eliminating polysemy from the text document, the quality of automatic classification improves	Elimination of polysemy in words sometimes loses sense of the text document.

## VI. CONCLUSION

As we know that the social media is the biggest podiums where enormous posts posted on a daily basis and seizing various opinions such as products, things, etc. But sometimes polysemous words presence may affect sense of the sentence and change the meaning of the expression. Sometimes we just remove the polysemy of words but some are worth to be enlightened as it adds value to the sentiments expressed in sentence. In this paper, we explain that how polysemy will affect our sentiment analysis and change the meaning of the expression. In future we will focus on different words so that our work should be more polished and we can easily identify the sense related to that word.

As we know that the identification of a polysemous word is based on the context and it is very imperative in abstracting the summarization of text document. So, we have to propose such method that will help to make our text document very feasible and context-based denotation of the polysemy word should be extracted and we have to present it along with their other related senses.

## REFERENCES

- [1] Udaya Raj Dhungana<sup>1</sup>, Subarna Shakya, Kabita Baral and Bharat Sharma (2014), "Word Sense Dissambiguation using WSD specific WordNet of Polysemy Words," in *International Journal on Natural Language Computing* (IJNLC).
- [2] Rice, D. R., & Zorn, C. (2013). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Proceedings of NDATAD*, 98-115.
- [3] M. Sinha, M. K. Reddy, P. Bhattacharyya, P. Pandey, and L. Kashyap, "Hindi word sense disambiguation," Master's thesis, Indian Institute of Technology Bombay, Mumbai, India, 2004.
- [4] U. R. Dhungana and S. Shakya, "Word sense disambiguation in nepali language," in *The Fourth International Conference on Digital Information and Communication Technology and Its Application (DICTAP2014)*, Bangkok, Thailand, 2014, pp. 46-50.
- [5] M. E. Lesk, "Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone," in *SIGDOC Conference*, Toronto, Ontario, Canada, 1986.
- [6] Kaur, J. (2016). A Review Paper on Twitter Sentiment Analysis Techniques. In *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* (Vol. 4, No. 2016).
- [7] Yih-Jeng LIN, Ming-Shing YU and Chin-Yu LIN (2010), A Combined Approach to the Polysemy Problems in a Chinese to Taiwanese TTS System, *National Chung-Hsing University, Taichung*. IEEE.
- [9] Chauhan, C., & Sehgal, S. (2017, May). Sentiment analysis on product reviews. In *Computing, Communication and Automation (ICCCA), 2017 International Conference on* (pp. 26-31). IEEE.
- [10] Astya, P. (2017, May). Sentiment analysis: Approaches and open issues. In *Computing, Communication and Automation (ICCCA), 2017 International Conference on* (pp. 154-158). IEEE.
- [11] Kolkur, S., Dantal, G., & Mahe, R. (2015). Study of different levels for sentiment analysis. *International Journal of Current Engineering and Technology*, 5(2), 768-770.
- [12] Ram, K.S.K., Araballi, S., BR, S., Shobha, G. (2014, December). Sentiment Analysis of Twitter Data. In *International Journal of Advanced Research in Computer*

- Engineering & Technology (IJARCET)* (Vol. 3 Issue 12, December 2014)
- [13] Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang, and Aijun An, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain", *IEEE Trans. On knowledge and data engineering*, vol. 24, no. 4, April 2012.
- [14] Bhuta, S., Doshi, A., Doshi, U., & Narvekar, M. (2014, February). A review of techniques for sentiment analysis Of Twitter data. In *Issues and challenges in intelligent computing techniques (ICICT), 2014 international conference on* (pp. 583-591). IEEE.
- [15] Mandal, S., & Gupta, S. (2016, December). A Lexicon-based text classification model to analyse and predict sentiments from online reviews. In *Computer, Electrical & Communication Engineering (ICCECE), 2016 International Conference on* (pp. 1-7). IEEE.
- [16] Andrey Kapitanov, Ilona Kapitanova, Vladimir Troyanovskiy, Vladimir Ilyushechkin, Ekaterina Dorogova (2019), Clustering of Word Contexts as a Method of Eliminating Polysemy of Words, *National Research University of Electronic Technology, Zelenograd, Moscow, Russia*, IEEE.
- [17] Mehto, A., & Indras, K. (2016, March). Data mining through sentiment analysis: Lexicon based sentiment analysis model using aspect catalogue. In *Colossal Data Analysis and Networking (CDAN), Symposium on* (pp. 1-7). IEEE.
- [18] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1-2), 1-135.
- [19] Shahare, F. F. (2017, June). Sentiment analysis for the news data based on the social media. In *Intelligent Computing and Control Systems (ICICCS), 2017 International Conference on* (pp. 1365-1370). IEEE.
- [20] Ramteke, J., Shah, S., Godhia, D., & Shaikh, A. (2016, August). Election result prediction using Twitter sentiment analysis. In *Inventive Computation Technologies (ICICT), International Conference on* (Vol. 1, pp. 1-5). IEEE.
- [21] Jain, A. P., & Dandannavar, P. (2016, July). Application of machine learning techniques to sentiment analysis. In *Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 2nd International Conference on* (pp. 628-632). IEEE.
- [22] Bhoir, P., & Kolte, S. (2015, December). Sentiment analysis of movie reviews using lexicon approach. In *Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on* (pp. 1-6). IEEE.