



## INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

### A Survey Based on Opinion Classification Techniques

Poongodi S<sup>\*1</sup>, Radha N<sup>2</sup>

<sup>\*1</sup> Research scholar, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India

<sup>2</sup> Assistant Professor, GRG School of Applied Computer Technology, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India

Poongodisri23@gmail.com

#### Abstract

The essential one is to collect the information as what the people think forever. Nowadays people can share everything in online Social network such as twitter, facebook. People articulate their views and opinions regarding products and services. People can easily access and understand the opinions of others via web resources like as blog, review, and forum. These opinions are individual information which represents user's opinions. The precise way for predicting opinions are to extract sentiments from the web, which could be valuable for marketing research. Opinions are so vital that whenever need to make a decision before want to know others' opinions. Opinion is not only important for a user but is also useful for an organization. This survey is about various methods and techniques used to classify the user opinions like as positive, negative.

**Keywords:** Opinion Mining, Machine Learning Algorithm, Negation Handling.

#### Introduction

Opinion mining is computational study about opinion, sentiment and expressions in text on online Social Network. Online Service is used to create the short text by the user, who updates on their activities, observations and interesting content. User opinion provides interesting as well as valuable data for making decision in business. Gathering feedback from user opinions about a given product (e.g., car, mobile phone, etc.), which may give companies valuable information as satisfaction or dissatisfaction of their customers. And this information is also useful for customers in their decisions to buy a particular product. Basic task of opinion mining are holder detection (who expressed the sentiments), target recognition (what expressed towards) and polarity classification (nature of sentiment). For example "Samsung galaxy s4 camera is taking too good quality picture". Here Samsung galaxy s4 is target; camera is features and too good is polarity. Opinion mining could be performed at several levels of context. Opinion mining task is classifying the polarity of text at document level, sentence and term /phrase level, whether it is expressed the opinion at document, sentence and features as positive, negative and neutral.

The Opinion Mining is to recognize the subjectivity and objectivity of a text and further classify the opinion orientation of subjective text. It is an automated extraction of subjective content from text and identifying the orientation such as positive or negative in

that text which aims to explore feelings of a person who write the text. It used Natural Language Processing and Machine Learning ethics to determine opinion in the text. In this opinion is classified in various approaches on text using some machine learning algorithm like Naïve Bayes, Maximum Entropy classification, and Support Vector Machines. In this paper describe various techniques used to classify the opinion.

#### Data Sources

User's opinion expressed their activities and interest through blogs, reviews and micro blog such as twitter face book.

#### Micro blogs:

Users may update their status through micro blog. And also their opinions expressed on micro blogs. For example twitter, face book. In twitter message may contains texts, emoticons and hashtag. Emoticons are the combination of symbols, which represent facial expressions. Hash tag (#) followed by one word or code to group the related messages. People use emoticons and hash tag to describe their emotions or attitudes, as to show intended humor and thoughts. Alec go et al.(2009) collected twitter data set from this URL <http://twittersentiment.appspot.com/>.

#### Reviews:

User posted reviews about the product commands and services on review site, which may

opinion expressed. Review data also used to classify the sentiments. Pang et al. (2002) collected movie review data set from <http://reviews.imdb.com/Reviews/>. Yulan He (2010) collected dataset from <http://www.cs.cornell.edu/People/pabo/movie-review-data> and <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

#### Flow of opinion mining:

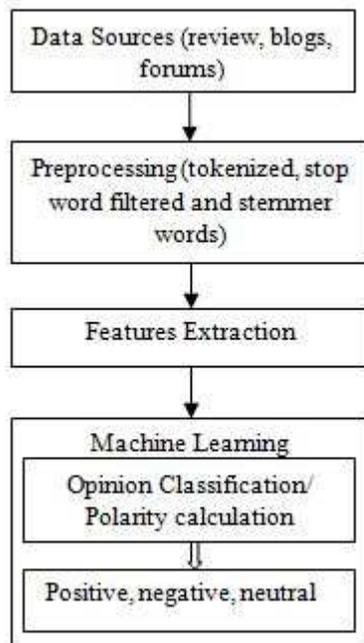


Fig 1: flow chart for opinion classification

### Opinion Classification

The sentence level opinion mining is associated with two tasks. First one is to identify whether the given sentence is subjective or objective. The second one is to find opinion of an opinionated sentence as positive, negative or neutral. The assumption is taken at sentence level is that a sentence contain only one opinion for e.g., “The picture quality of Samsung Galaxys4 is good”. Riloff and Wiebe (2004) use a method known as bootstrap approach to recognize the subjective sentences then attained the result 90% accuracy during their tests. Wilson et al. proposed that not only a single sentence may contain multiple opinions, but both subjective and factual clauses have also multiple opinions. It is also important to identify the strength of opinions.

Pang et al. (2002) proposed machine learning method, namely Naïve Bayes, Maximum Entropy and Support Vector Machine to classify the movie review data set. The experimental guides to a both classes as positive and negative classification problem and it can be performed after determining subjective text. Movie

Review data set achieved 81.6% accuracy on support vector machine.

Go et al. (2009) proposed for classifying the sentiment of Twitter messages, which are classified as either positive or negative with used to a query term automatically. The result of machine learning algorithms for classifying the sentiment of Twitter messages using distant supervision. In training data consists of Twitter messages with emoticons, which are used as noisy labels. The result showed that accuracy above 80% when trained with emoticon data.

Yulan He (2010) attempted to create a structure for sentiment classifier learning from unlabeled documents. Dataset are collected about multi domain review and movie review. The method initiates with a collection of un-annotated text and a sentiment lexicon. An initial classifier is trained by integrating prior information from the sentiment lexicon which consists of a list of words marked with their respective polarity. The labeled features use directly to constrain model’s predictions on unlabeled instances using generalized expectation criteria. The initially-trained classifier using generalized expectation is applied on the un-annotated text and the documents labeled with high confidence are fed into the self-learned features extractor to acquire domain-dependent features automatically. Self-learned features are subsequently used to train another classifier which is then applied on the test set to obtain 75.30 %.

Jiang et al. (2011) performs a target dependent twitter sentiment analysis. Given a query term, tweets are collected which are talking about the query subject. Here three approaches are implemented namely Subjectivity, Polarity and graph based optimization classification. Subjectivity classification is to decide if the tweet is subjective or neutral about the target. Polarity classification is to decide if the tweet is positive or negative about the target if it is classified as subjective. Graph-based optimization is to further boost the performance by taking the related tweets into consideration. Graph-based optimization approach improves the performance to classify the opinion. SVM classifier is achieved 68.3% to classified three classes as positive, negative and neutral.

Agarwal et al. (2011) performed three classes such as positive, negative and neutral classification of tweets. The dataset are collected using Twitter stream API and asked human judges to annotate the data into three classes. Each class had 1709 tweets, making a total of 5127 in all. POS-specific prior polarity features along with twitter specific features are introduced. Uni-gram + senti features are achieved the accuracy of 75.39%.

A Bakliwal et al. (2011) proposed to classify the sentiment using machine learning methods on movie and product review data set. Sentiment classified using basic

NLP Techniques such as NGram, POS-Tagged Ngram. A model proposed for address the problem of determining whether a review is positive or negative for use several machine learning algorithms Naive Bayes (NB), Multi-Layer Perception (MLP), Support Vector Machine (SVM). Here negation also handled and observed the improvements in classification. They achieved the accuracy of 78.32% on movie and 70.06% on multi-category dataset.

Summary of the survey:

S.No	Studies	Data Source	Performance	Mining
1	Jieng et al	Tweet message	68.3%	SVM
2	Yulan He (2010)	Movie review	74.6%	Opinion words
3	Alec Go (2009)	Twitter message	81.9%	SVM
4	Akshat Bakliwal (2011)	Movie and Product Review	78.32%, 76.06%	MLP
5	Bo Pang and Lillian Lee (2002)	Movie review	86.4%	SVM
6	Agarwal et al. (2011)	Twitter stream API	75.39%.	Uni-gram + senti features

### Conclusion

This paper represents various techniques available for classification of opinion mining on text. There are different ways to implement machine-learning algorithm using features. More work is needed on further improving the performance measures. The techniques and algorithms used for polarity calculation and a lot of problems in this field of study reside unclear. The main challenging is dealing with produce a summary of opinions based on product features/attributes, complexity of sentence/ document, etc. Many research works could be committed to these challenges in future.

### References

[1] Abbasi, A., Chen, H. and Salem, A. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Trans. Inf. Syst.*, Vol. 26, 3, pp. 1-34.2008.

[2] Akshat Bakliwal, Piyush Arora, Ankit Patil, Vasudeva Varma, "Towards Enhanced Opinion Classification using NLP Techniques", 2009.

[3] Alec Go, Richa Bhayani, Lei Huang, "Twitter Sentiment Classification using Distant Supervision", 2009.

[4] Amna Asmi, Tanko Ishaya. "Negation Identification and Calculation in Sentiment Analysis". The Second International Conference on Advances in Information Mining and Management, 2012.

[5] B. Liu 2011. Opinion Mining and Sentiment Analysis, AACL, San Francisco, USA.

[6] Bo Pang and Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", 2002

[7] Council, I. G., Mcdonald, R. and Velikovich, L. (2010) What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In: *The Workshop on Negation and Speculation in Natural Language Processing*. Uppsala, Sweden. Association for Computational Linguistics, 51-59.

[8] E. Riloff, and J. Wiebe, " Learning Extraction Patterns for Subjective Expressions, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)", Japan, Sapporo. 2003

[9] H. Yu, and V. Hatzivassiloglou, 2003. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Japan, Sapporo.

[10] Jia, C. Yu, and W. Meng, "The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness", In Proceedings of CIKM, 2009.

[11] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11, pages 151–160. Association for Computational Linguistics, 2011.

[12] Rudy Prabowo, Mike Thelwall, "Sentiment analysis: A combined approach .", *Journal of Informetrics* 3 (2009) 143–157.

[13] Rui Xia, Chengqing Zong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", *Information Sciences* 181 (2011) 1138–1152.

- [14] T. Wilson, J. Wiebe, R. Hwa, "Just how mad are you? Finding strong and weak opinion clauses". In: the Association for the Advancement of Artificial Intelligence. 2004.
- [15] Wiebe, W.-H. Lin, T. Wilson and A. Hauptmann, "Which side are you on? Identifying perspectives at the document and sentence levels," in Proceedings of the Conference on Natural Language Learning (CoNLL), 2006.
- [16] Yulan He, Deyu Zhou, "Self-training from labeled features for sentiment analysis", Information Processing and Management, 2010.