

LITERATURE REVIEW ON SENTIMENT ANALYSIS

Hemamalini, Dr.S. Perumal

Abstract— Opinion Mining a sub field of Natural Language Processing (NLP) in analysing text data and determining the polarity, which also uses Sentiment Analysis as a part of Opinion Mining. It is extremely useful for monitoring the people feelings about a particular topic, product or idea. Document analysers adapted the technique of Sentiment Analysis for summarizing or customizing particular knowledge text. Analysis of sentiment is widely used in mining of subjective information from internet content using various techniques including NLP, Statistical techniques and Machine Learning methods. The application of Sentiment Analysis is very powerful and broad field which mainly focuses on huge text analysing process rather analysing manually. The Opinion Mining technology is widely used for extracting social data and makes the business process profitably. This survey describes about the research accomplished in Sentiment Analysis in last decade to show how the problems have solved until now and the various algorithms used in Sentiment Analysis.

Index Terms— Opinion Mining, Sentiment analysis, Natural Language processing, Machine learning, Accuracy, Polarity, Time, Reviews.

1 INTRODUCTION

The emotions in textual data is analysed and processed in Sentiment Analysis. In other words the Sentiment Analysis specifies that the given information is positive, negative or neutral about a specific topic or product. Due to this scenario it is widely termed as Opinion Mining. For processing the textual information Sentiment Analysis adapts the approaches of NLP (Natural language processing), AI (Artificial Intelligence) and ML (Machine Learning). Conclusively analysis of sentiment deals with the people opinion about a particular topic or product. Sentiment Analysis is competent in understanding the people's opinion and to provide a basis of evidence and reasoning on particular product. Sentiment Analysis analyses abundant textual information presented in internet to provide future insight for the organization and aid the public to take decision on their purchase. For instance, in an e-commerce website, people purchase product and give their reviews with or without using the product. These reviews must help the customers who desire to purchase the product or item. But the issue is, as the number of reviews is more the customers may find it difficult to read all reviews. So, there is a need of automated process to give an appropriate conclusion for particular product or topic and this task is known as Sentiment Analysis. The Sentiment Analysis classification framework (fig .1) is illustrated as follows,

- **Case Normalisation:** The entire documents are changed to lower case or upper case.

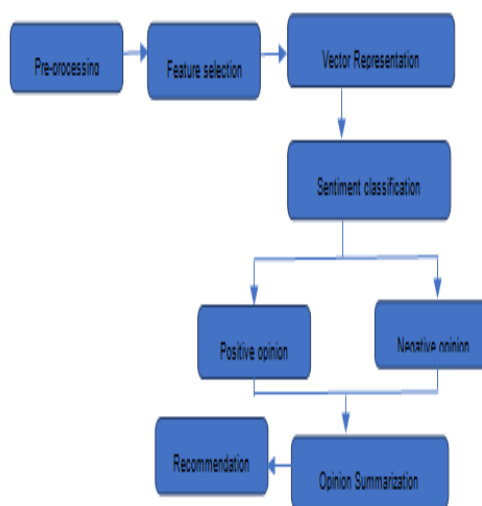


Fig 1. Framework for Sentiment Classification

a. Preprocessing: In Preprocessing the unanalyzed data is handled for feature extraction. It is further divided into below step:

- **Tokenisation:** White spaces, symbols and special characters are removed and a sentence is divided into words.
- **Stop Word Removal:** Articles are removed.
- **Stemming:** Token or words are reduced for root forms.

b. Feature extraction: Feature extraction handles the following task:

- **Feature Type:** In this step features are identified like the term frequencies, term co-occurrences, Opinion word, OS information, Negation Syntactic Dependencies.
- **Selection of Feature:** Good features are selected for classification using the following ways like Information gain, Document frequency, Odd ratio and Mutual Information.
- **Feature Weighting Mechanism:** The features are ranked by computing the weight using term presence, term frequency and Inverse document frequencies.
- **Reduction of Feature:** To optimize the classifiers performance the vector size is reduced.

c. Sentiment Analysis: Polarity of text is classified by Sentiment Analysis. This process is done in 3 different levels.

- Mrs.U.Hemamalini , Research Scholar, Department of Computer Science, VISTAS Pallavaram, Chennai, Assistant Professor, Prince Shri Venkateshwara Arts and Science College, Chennai.hemababu2501@gmail.com
- Dr.S.Perumal, Associate Professor, Department of Computer Science, VISTAS, Pallavaram, Chennai.

- **Document Level:** The entire document is taken and is labeled as either positive or negative.
- **Sentence Level:** The entire document is parsed into sentence and the polarity is classified as positive, neutral or negative.
- **Word or Phrase Level:** Product attributes or components are analyzed.

d. Sentiment Classification: Sentiment classification uses two approaches to classify the nature of documents/sentence.

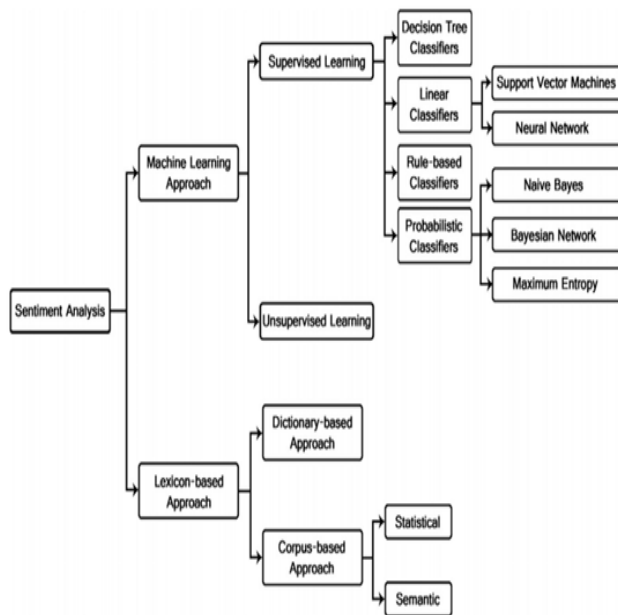


Fig 2. Sentiment classification techniques.

They are Machine Learning Approaches and Lexicon Based Approaches. The Machine Learning belongs to supervised learning and classification of text in particular. So, it is called as "Supervised Learning". It is composed of several methodologies like Naïve Bayes (NB), Maximum Entropy (ME), Support Vector Machine (SVM), K-Nearest Neighborhood, and Neural Networks. Lexicon Based approach consist of Dictionary based and corpus based. From different point of view the recent work on Sentiment Analysis can be characterized by the technique used, rating level, view of text and level of detailed text analysis. Machine learning, lexicon-based, statistical based and rule-based methods are identified on technical point of view. By training the known data set the sentiments are analyzed by several learning algorithms in machine learning methods. The quality of review and opinion in text is used to regulate the polarity for a review in lexicon-based approaches. The rule-based approach makes use of different classification method such as negation words, booster words, dictionary polarity, idioms, mixed opinions to analyze the opinion terms in a text and to regulate the polarity of opinion words. The latent aspect and rating are represented in statistical models. The structure of text is very crucial for classification. The construction of the text can be analyzed in document level, sentence level, word or feature level. The past literature shows that the document

level classification is focused by many methods. In global level classification, each aspect has strength and attempts to determine a rate for a review. Methods used for global review classification depend on machine learning technique and consider only the polarity of the review. For a detailed classification of review more linguistic feature has to be included like negation, intensification, discourse structure and modality.

2 PARAMETERS AND MEASUREMENTS IN SENTIMENT ANALYSIS

Table 1: Parameters and Measurements in Sentiment Analysis.

AUTHOR	TOPIC	MEASUREMENTS	DESCRIPTION
Lik Mui, Arihalberstadt, Mojdeh Mohtashemi.	A Computational model of Trust and Reputation	Accuracy (Size)	This paper presented a model to show the difference between the trust and reputation in terms of the size of the network.
A. Arenas, L. Donon, P.M. Gleser and R. Guimera, A. Diaz-Guilera	Community analysis in social network	Accuracy (Size)	Presented a model to scale the characteristic of the community size distribution of different social network with two different exponents.
BoPang, Lillian Lie	A Sentimental Education: Sentimental Analysis using subjectivity summarization based on Minimum Cut.	Accuracy (Performance)	This paper proposed that SVM and NB are better technique for improving the performance of a model up to 86.4 %.
Erik.Boiy, Pieter Hens, Koen Dschacht, Marie Francine Moens	Automatic Sentiment Analysis in Online Text	Accuracy (speed and size)	This paper shows the varying level of accuracy when symbolic and machine learning methods were applied to different social network dataset.
Doreen Hii	Using Meaning specificity to aid negation handling in	Accuracy (Performance)	This paper compared the accuracy of 1-,2-,3-,4-, gram and

	Sentiment analysis.		suggested 4-gram is the best performing window size in Lexicon method.
Junyi Jessy Li, Ani Nenkova	Fast and Accurate predictions of sentence specificity	Accuracy (Performance)	Suggested a training algorithm using Shallow and Brown cluster for predicting sentence specificity with highest accuracy of 71%.
Matthew Richardson, Rakesh Ajarwal, and Pedro Domingos.	Trust Management for the semantic web	Accuracy (Time, Noise)	Suggested that Semi Naïve is the simplest one which runs in (ON) ⁴ times.
AK. Varadharajulu, Y Ma	Data Mining algorithms for a feature-based customer review process model with engineering informatics approach	Accuracy (Time)	Indicates that J48 is the best classifier with highest accuracy and Simple K- mean is used for a smaller number of cluster when compared with other techniques.
Diamh Alahmadi, Xiaojian Zeng.	Improving Recommendations using Trust and Sentiment Inference from OSN.	Accuracy (Noise)	Indicates that decision tree achieves the highest accuracy level among the other methods.
Zhenzhen Xu, Jialiang Kang, Huizhen Jiang, Xiangjie Kong, and Wie Wang, Feng Xia.	Cross Domain item Recommendation based on user similarity.	Accuracy (Time, Quality)	Suggested that Random walk algorithm works better for time and quality measurements and also elevates the problem of Cold start and Sparsity.
Hao Tian, and Peifeng Liang	Improved recommendations based on trust relationships in social networks	Accuracy (Performance, Noise)	This paper proposes IRATR as a best approach in prediction the performance and is dealing with cold start and sparsity

			when compared with NCut and Trust walker list method.
Sophie de kok, Linda Punt, Rosita van den Puttelaar, Karoluna Ranta, Kim Schouten, and Flavius Frascinear.	Review aggregated aspect-based sentiment analysis with ontology feature	Accuracy (Noise)	This paper shows that pure review level algorithms outperforms the sentence aggregation method in terms of sensitivity.
Xiaomie Zou, Jing Yang, and Jianpei Zhang.	Microblog sentiment analysis using social and topic context.	Accuracy (Performance)	This paper proposes a new method to identify the polarity of the sentiment and shows the structure similarity has a better accuracy than user direct relations.
A K Varadharajulu and Y Mu	Datamining algorithm for a feature-based customer review model with engineering informatics approach	Accuracy (Time, Noise)	This paper investigates that J48 classification methods, and simple K – Mean clustering algorithm are best in terms of accuracy
Dr. Sefer Kurnaz, and Mustafa Ahmed Mahmood.	Sentiment Analysis in data of Twitter using Machine learning algorithm	Accuracy (Time)	This paper proposes a new technique which offers an accuracy of 98 % when compared with Deep learning method, SVM and Maximum Entropy method.

The past literature and various sentiment analysis techniques are discussed in this section. The sentiment of the reviews is identified by approaches [1],[2],[3],[4],[5],[6],[7],[8]. The accuracy of the classifier is increased by reducing the noise in textual data by pre-processing it. Bo Pang and Lillian Lee [9] proposed that Naïve Bayes and SVM is the efficient method for providing highest accuracy. Xiangjie Kong, Huizhen Jiang, Zhuo Yang, Zhenzhen Xu, feng Xia, Amr Tobla [10] embraced Random Walk model for giving highest accuracy in academic domain. Huakang Li, Yixiong Bian, Xiuying Xu, Guozi Sun [14] suggested Monte Carlo decision tree algorithm for mining interest similarity with highest performance when compared

with other techniques. G. Vinodhini [11] proposed NB is the best technique for estimating the quality of the document. Jayashri Khairnar and Mayura Kinikar [12] put forward that SVM excelled when comparing with other techniques in sentiment classification. Faruk and Arnab [13] initiated a model for trust management with highest accuracy. Rudy Prabowo [16] explored a new approach for improving the performance of the classifier. Dongjoo Lee et al [15] says the when dealing with huge volume of data, PMI give better accuracy. Soudamini Hota, Sudhir Pathak [19] recommended SVM and KNN as a best method for handling noisy data in textual information. Table 1 describes the list of measurements used in different application in sentiment analysis. Fig-2. shows the parameters to determine the efficiency of classifier in sentiment analysis. Table 2- describes various algorithms in sentiment classification.



Fig.2 Factors in Sentiment Analysis.

Table 2: Algorithms used in Existing research work

ALGORITHM	OPERATION	STATUS
Rule Based	Mine product feature or polarity of sentence for a product	Decrease in Recall rate, difficult to list the rules.
Naïve Bayes	It is carried out to identify the polarity of the textual information	Assumes only unconventional attribute.
Support Vector Machine(SVM)	It is used to split the data points of the classes from one another.	Transparency of the result is lacking.
Multilayer perception	Handles the data which flows in one direction alone.	Requires additional time for execution and less flexible.
Maximum Entropy	Used to estimate the data from previous test and is better for probability distribution.	Effort of human is required.
Decision Tree	It has the potential of describing the decision-making knowledge from the given data.	Not efficient for regression test and for predicting continuous values.
Convolutional Neural Network (CNN)	Classifies unorganised data in textual information.	Loses Phrase level labels.
Bayesian Network	Presents the dependencies of variables	Hypothesis of independent features.
Statistical	Discovers the receptacle of the sentiment and target	Noisy and the sentiment is often classified as neutral.
Semanti	Recognise the patterns in the unorganised collection of information.	Dependencies between the entities are not considered.

3 CONCLUSION

Opinion mining is a field where a large volume of data is being generated through person-to-person communication. Opinion Mining in market analysis and teaches what changes is necessary for forecoming product generation with the help of historical market data. The product designing and development may further construct in an efficient way using Opinion mining. Usually customers use Opinion Mining platform for purchasing a product which they never used before. The customization and review process reveal the product in an efficient way and makes the customer satisfaction level high through Opinion Mining. This paper presented the various parameters and measurements for determining the efficiency of the classifier and concludes classifier alone cannot give complete efficiency in accuracy since the result is based on number of factors.

REFERENCES

- [1] E. Kouloumpis, T. Wilson, J. Moore, Twitter sentiment analysis: The good the bad and the omg!, Proc. 5th Int. AAAI Conf. Weblogs Social Media, pp. 538-541, 2011.
- [2] D. Terrana, A. Augello, G. Pilato, Automatic unsupervised polarity detection on a Twitter data stream, Proc. IEEE Int. Conf. Semantic Comput., pp. 128-134, Sep. 2014.
- [3] H. Saif, Y. He, M. Fernandez, H. Alani, Semantic patterns for sentiment analysis of Twitter, Proc. 13th Int. Semantic Web Conf., pp. 324-340, Apr. 2014.
- [4] H. Saif, Y. He, H. Alani, Alleviating data sparsity for Twitter sentiment analysis, Proc. CEUR Workshop, pp. 2-9, Sep. 2012.
- [5] H. G. Yoon, H. Kim, C. O. Kim, M. Song, Opinion polarity detection in Twitter data combining shrinkage regression and topic modeling, J. Informetrics, vol. 10, pp. 634-644, 2016.
- [6] F. H. Khan, U. Qamar, S. Bashir, SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection, Appl. Soft Comput., vol. 39, pp. 140-153, Apr. 2016.
- [7] A. Agarwal, B. Xie, I. Vovsha, Sentiment analysis of Twitter data, Proc. Workshop Lang. Social Media Assoc. Comput. Linguistics, pp. 30-38, 2011.
- [8] Bo Pang, Lillian Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Article No.271, 2004.
- [9] Kong X, Jiang H, Yang Z, Xu Z, Xia F, Tolba A (2016) Exploiting Publication Contents and Collaboration Networks for Collaborator Recommendation. PLoS ONE

- 11(2): e0148492.
- [10] G. Vinodhini et al, Sentiment Analysis and Opinion Mining: A Survey, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Vol 2, Issue 6, June 2012.
- [11] Jayashri Khairnar and Mayura Kinikar, Machine Learning Algorithms for Opinion Mining and Sentiment Classification, International Journal of Scientific and Research Publications (IJSRP), Volume 3, Issue 6, ISSN 2250-3153, June 2013.
- [12] Faruk Alam; Arnab Paul, A computational model for trust and reputation relationship in social network, Published in: International Conference on Recent Trends in Information Technology (ICRTIT), ISBN: 978-1-4673-9802-2, 2016
- [13] Li, H.; Bian, Y.; Xu, X.; Sun, G. SRRTC: Social Recommendation Based on Relationship Transmission with Convergence Algorithm. Preprints 2017, 2017110033 (doi: 10.20944/preprints201711.0033.v1).
- [14] Dongjoo Lee et al, Opinion Mining of Customer Feedback Data on the Web. Seoul National University
- [15] Rudy Prabowo, Mike Thelwell, Sentiment Analysis: A Combined Approach.
- [16] Arti Buche, Dr.M.B. Chandak, Akshay Zadgoanakar, Opinion Mining and Analysis: A Survey, International Journal on Natural Language Computing (IJNLC) Vol 2 No 3 June 2013Pg No 39-48.
- [17] Nilesh M. Shrike et al, Survey of Techniques for Opinion Mining, International Journal of Computer Applications, Vol 57 No 13. Nov 2012Pg No 30-35.
- [18] Hota, Soudamini, Pathak, Sudhir, KNN classifier-based approach for multi-class sentiment analysis of twitter data, Volume - 7, 10.14419/ijet. v7i3.12656, International Journal of Engineering and Technology (UAE)
- [19] Bodapati J.D., Veeranjanyulu N., Shaik S. (2019). Sentiment analysis from movie reviews using LSTMs, Vol. 24 No. 1, pp. 125-129. <https://doi.org/10.18280/isi.240119>.
- [20] Mui, L. & Mohtashemi, Mojdeh & Halberstadt, A. (2002). A computational model of trust and reputation. 2431 - 2439. 10.1109/HICSS.2002.994181.
- [21] Arenas, Alex et al., Community analysis in social networks, The European Physical Journal B 38 (2004): 373-380.
- [22] Bo Pang, Lillian Lee. (2004), A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts.
- [23] Boiy, & Erik, & Hens, & Pieter, & Deschacht, & Koen, & Moens, Marie-Francine & Marie-Francine,. (2007). Automatic Sentiment Analysis in On-line Text. ELPUB2007.Proceedings of the 11th International Conference on Electronic Publishing held in Vienna, Austria 13-15 June 2007, ISBN 978-3-85437-292-9, 2007, pp. 349-360.
- [24] DeWitt, T., Nguyen, D. T., & Marshall, R. (2008). Exploring Customer Loyalty Following Service Recovery: The Mediating Effects of Trust and Emotions. Journal of Service Research, 10(3), 269-281. <https://doi.org/10.1177/1094670507310767>.
- [25] Doreen Hii (2019), Using Meaning specificity to aid negation handling in Sentiment analysis.
- [26] Junyi Jessy Li, and Ani Nenkova (2015), Fast and Accurate prediction of sentence specificity.
- [27] Matthew Richardson, Rakesh Agrawal, Pedro Domingos, Trust management for the semantic web, Proceedings of the Second International Conference on Semantic Web Conference, October 20-23, 2003, Sanibel Island, FL
- [28] AK. Varadharajulu, Y Ma (2013), Data Mining algorithms for a feature-based customer review process model with engineering informatics approach. Diah Alahmadi, Xaio - Jan Zeng. (2015), Improving Recommendation using Trust and Sentiment Inference from OSN.
- [29] Xu, Z., Jiang, H., Kong, X., Kang, J., Wang, W., Xia, F.: Cross-Domain Item Recommendation Based on User Similarity. Computer Science and Information Systems, Vol. 13, No. 2, 359-373. (2016), <https://doi.org/10.2298/CSIS150730007Z>
- [30] Hao Tian, Peifeng Liang (2017), Improved recommendations based on trust relationships in social networks.
- [31] Hongyu Han, Yongshi Zhang, Jianpei Zhang, Jing Yang, Xiaomie Zou. (2018), Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias.
- [32] Zou X, Yang J, Zhang J (2018) Microblog sentiment analysis using social and topic context. PLoS ONE 13(2): e0191163. <https://doi.org/10.1371/journal.pone.0191163>
- [33] Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). Using Twitter trust network for stock market analysis. Knowledge-Based Systems, 145, 207-218. <https://doi.org/10.1016/j.knosys.2018.01.016>
- [34] A K Varadharajulu and Y Mu (2019), Datamining algorithms for a feature-based customer review model with engineering informatics approach.
- [35] Dr. Sefer Kurnoz, Mustafa Ahmed Mahmood (2019), Sentiment Analysis in data of Twitter using Machine learning algorithms.
- [36] Nupur Kalra, Deepak Yadav, Gourav Bathla. (2019), SynRec: A prediction technique using collaborative filtering and Synergy Score.