EasyChair Preprint

№ 3312

An Overview of Different techniques of Sentiment Analysis

Simranjot Kaur and Raman Maini

May 2, 2020

An Overview of Different techniques of Sentiment Analysis

[1]Simranjot Kaur,[2]Raman Maini

University College of Engineering,
Punjabi University, Patiala-147002(India)
k.simranjot@gmail.com

[2] University College of Engineering,
Punjabi University, Patiala-147002(India)
Research_raman@yahoo.com

Abstract

Sentiment analysis is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. This paper discusses about Sentiment analysis methods used.

Keywords*: sentiment analysis, opinion mining , data mining, web mining, text mining,qualitative Assurance*

1. Introduction

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. A sentiment analysis system for text analysis combines natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores to the entities, topics, themes and categories within a sentence or phrase. The best businesses understand the sentiment of their customers—what people are saying, how they're saying it, and what they mean. Customer sentiment can be found in tweets, comments, reviews, or other places where people mention your brand. Sentiment Analysis is the domain of understanding these emotions with software, and it's a must-understand for developers and business leaders in a modern workplace. As with many other fields, advances in deep learning have brought sentiment analysis into the foreground of cutting-edge algorithms. Today we use natural language processing, statistics, and text analysis to extract, and identify the sentiment of text into positive, negative, or neutral categories.

The sentiment may be a judgment, mood or evaluation of the writer [19]. A core issue in this field is an opinion classification, where a review is classified as a positive or negative evaluation of a subjected object (film, book, etc.). The assessment of sentiment can be done in two ways :

1.1 Direct opinions: It gives positive or negative sentiment about the product directly [18]. For example, "The food quality of this hotel is poor" expresses a direct opinion.

1.2 Comparison: It means to compare the subject with any other similar objects [18]. For example, "The food quality of the hotel-a is better than that of hotel-b." expresses a comparison.

2. Applications of Sentiment Analysis

Sentiment analysis systems are being applied in almost every business and social domain because opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are largely conditioned on how others see and evaluate the world.

In Marketing, to know how the public reacts on your product/service, which parts of it they like/dislike, how they want it to be improved, etc.

- In Quality Assurance, to detect errors in your products based on your actual users experience
- In Politics, to determine the views of the people regarding specific situations, what are they angry or happy for, etc

- In Risk Prevention, to detect if some people are being attacked or harassed (for example in Social Media or internal chats)
- In Finances, to try to detect the sentiment towards a brand, to anticipate their market moves

3. Sentiment Analysis techniques :

This section provides a brief description of the three sentiment analysis methods investigated in this paper. These methods are the most popular in the literature (i.e., the most cited and widely used) and they cover diverse techniques such as the use of Natural Language Processing (NLP) in assigning polarity, the use of Amazon's Mechanical Turk (AMT) to create labeled datasets, the use of psychometric scales to identify mood-based sentiments, the use of supervised and unsupervised machine learning techniques, and soon.

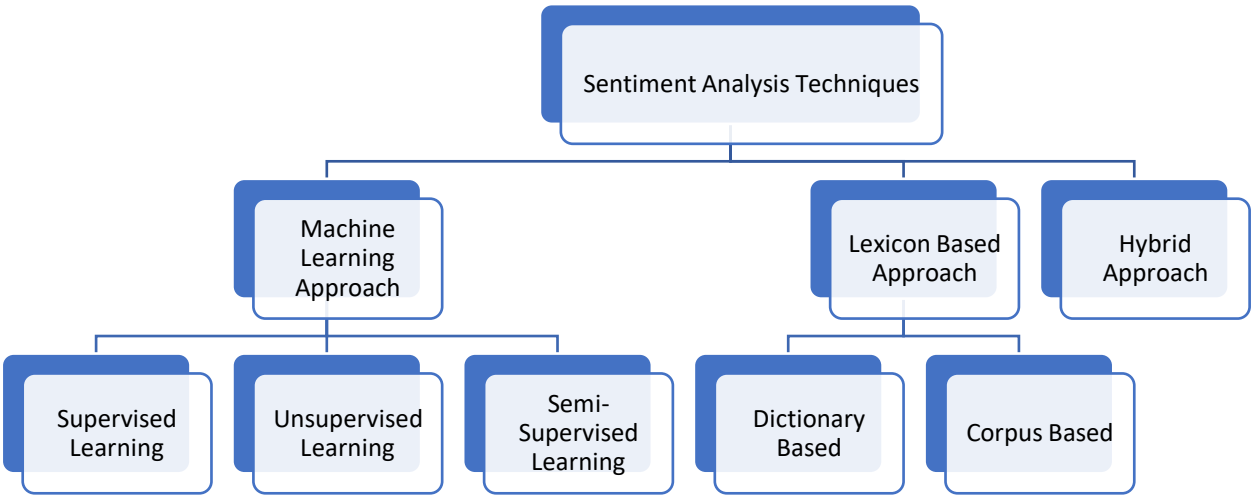

Fig 1 : Different Techniques of Sentiment Analysis

A. Lexicon based Approach :

lexicon-based approaches for sentiment classification are based on the insight that the polarity of a piece of text can be obtained on the ground of the polarity of the words which compose it. It deals with counting the number of positive and negative words in the text. If the text consists of more positive words, the text is assigned a positive score. If there are more number of negative words the text is assigned a negative score. If the text contains equal number of positive and negative words then it is assigned a neutral score. There are several approaches to compile and build an opinion lexicon .

- Dictionary based approach: A small set of opinion words is collected manually with known orientations. Then, synonyms and antonyms of these words are searched in corpora like WordNet or thesaurus and added to the set. The set gradually grows until no new words are found.
- Corpus based approach: They depend on large corpora for syntactic and semantic patterns of opinion words. The words that are generated are context specific and may require a huge labelled dataset.

The most common lexicon resources are SentiWordNet, WordNet, and ConceptNet, and among these resources, SentiWordNet is the most widely used[13]

B. Machine Learning Approach :

In such approaches, the analysis of customers' emotional inclinations is considered to be a problem of polarity classification. It starts with collection of dataset containing labelled tweets. This dataset might be boisterous and subsequently should be pre handled utilizing various Natural Language processing (NLP) techniques. Then

features that are relevant for sentiment analysis need to be extracted and finally the classifier is trained and tested on unseen data.

Machine Learning-Based Sentiment classification methods can be categorized into three types: supervised, unsupervised, and semi-supervised learning methods[8]

1) Supervised learning
   Supervised learning is a mature and successful solution in traditional topical classification and has been adopted and investigated for opinion detection with satisfactory results.
   The biggest limitation associated with supervised learning is that it is sensitive to the quantity and quality of the training data and may fail when training data are biased or insufficient.
2) Unsupervised learning
   In text classification, it is sometimes difficult to create labeled training documents, but it is easy to collect the unlabeled documents. The unsupervised learning methods overcome these difficulties. Traditional topic models such as LDA and pLSA are unsupervised methods for extracting latent topics in text documents.
   The limitation of unsupervised approaches is that they normally need a large volume of data to be trained accurately.
3) Semi-Supervised learning (SSL)
   SSL models drive from either supervised or unsupervised methods. In contrast with supervised learning, which learns from labeled data only, SSL learns from both labeled and unlabeled data. SSL is a relatively new machine learning approach to opinion mining, motivated by the lack of labeled data in real world applications.

*C.*Hybrid Approach :

In order to improve sentiment classification performance few research techniques suggest using a combination of both lexicon based and machine learning techniques. The main advantage of this hybrid approach is that we can attain best of both world.

This method uses natural language processing (NLP) essential techniques, a sentiment lexicon enhanced with the assistance of SentiWordNet, and fuzzy sets to estimate the semantic orientation polarity and its intensity for sentences, which provides a foundation for computing with sentiments.

Mudinas et al. [20] combined lexicon based and learning-based approaches to develop a concept-level sentiment analysis system, pSenti. It utilized advantages of both the approaches and attained stability and readability from semantic lexicon and high accuracy from a powerful supervised learning algorithm. They extracted sentiment words and considered it as features in machine learning

4. Comparison

The performance of various sentiment analysis techniques was measured on the basis of accuracy. A brief comparison of different techniques used in Sentiment Analysis is shown in table 1. Performing sentiment analysis by various approaches will produce different results. Each approach has its own pros
and cons .

Table 1. Comparison of Three Approaches

Approaches	Classification	Advantages	Disadvantages
Machine Learning Approach	• Supervised and Unsupervised learning.	• Dictionary is not necessary. • Demonstrate the high accuracy of classification.	• Classifier trained on the texts in one domain in most cases does not work with other domains.
Hybrid Based Approach	• Supervised and Unsupervised learning.	• Performance accuracy of 91% at the review level and 86% at the sentence level. • Sentence level sentiment classification performs better than the word level.	• Efficiency and accuracy depend the defining rules.
Lexicon Based Approach	Unsupervised learning. procedure	Labelled data and the of learning is not required.	• Requires powerful linguistic resources which is not always available

5. Conclusion

Various sentiment analysis methods have been studied in this paper such as machine learning, Lexicon based and Hybrid based approach . Many different methods have been developed and tested. However, alot of work is yet to be done. The most common approach is machine learning, a method that needs a significant data set for training and learning the aspects and sentiments associated. Also, models tend to target a simple global classification of reviews, rather than rating individual aspects of the reviewed product. Only a few of the methods are able to reach a somewhat high level of accuracy. Thus, the solutions for sentiment analysis still have a long way to go before reaching the confidence level demanded by practical applications.

6. References

[1] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, pp. 1-167, 2012.

[2] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," Computers in Human Behavior, vol. 31, pp. 527-541, 2014.

[3] A. Nisha, Jebaseeli, E. Kirubakaran, PhD., “A Survey on Sentiment Analysis of (Product) Reviews”, International Journal of Computer Applications (0975 – 888) Volume 47– No.11, June 2012

[4] Siddhi Patni, Avinash Wadhe, “Review Paper on Sentiment Analysis is – Big Challenge”, International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 2, ISSN: 2321-7782 (Online), February 2014 .

[5] G.Vinodhini, RM.Chandrasekaran, “Sentiment Analysis and Opinion Mining: A Survey”, International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 6, ISSN: 2277 128X, June 2012 .
[6] Abdullah Dar*, Anurag Jain, “Survey paper on Sentiment Analysis: In General Terms”, International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-3, Issue-11).
[7] Jiawen Liu, Mantosh Kumar Sarkar and GoutamChakraborty, “Feature-based Sentiment Analysis on Android App Reviews Using SAS® Text Miner and SAS® Sentiment Analysis Studio”, SAS Global Forum 2013.
[8] Bing Liu, “Sentiment Analysis and Opinion Mining”, Morgan and Claypool Publishers, p.18-19, 27-28, 44-45, 47, 90-101, May 2012.
[9] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.
[10] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", Expert Systems with Applications, vol. 36, pp. 6527-6535, 2009.
[11] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach" , Journal of Informetrics, vol. 3, pp.143-157, 2009
[12] A. Khan, B. Baharudin, K. Khan; “Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure” ICSECS 2011: 2nd International Conference on Software Engineering and Computer Systems, Springer, pp.317-331, 2011.
[13] A. Mudinas, D. Zhang, M. Levene, “Combining lexicon and learning based approaches for conceptlevel sentiment analysis”, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York,NY, USA, Article 5, pp. 1-8, 2012.
[14] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. 5 (2014) 1093–1113.
[15] A. Montejo–Ráez, E. Martínez-Cámara, M.T. Martín-Valdivia, L.A. Urena-López, Ranked WordNet graph for sentiment polarity classification in Twitter, Comput. Speech Lang. 28 (1) (2014) 93–107.
[16] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowl–Based Syst. (2015), Available Online 29June 2015.

[17] Schouten, K., Frasincar, F.: Ontology-driven sentiment analysis of product and service aspects. In: Proceedings of the 15th Extended Semantic Web Conference (ESWC 2018). vol. 10843, pp. 608–623. Springer (2018)
[18] Nidhi Mishra et al, “Classification of Opinion Mining Techniques”, International Journal of Computer Applications, Vol 56, No 13, Oct 2012Pg No 1-6.
[19] David Osimo and Francesco Mureddu, “Research Challenge on Opinion Mining and Sentiment Analysis”.
[20] A. Mudinas, D. Zhang, M. Levene, “Combining lexicon and learning based approaches for conceptlevel sentiment analysis”, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York,NY, USA, Article 5, pp. 1-8, 2012.