

A BRIEF REVIEW OF SENTIMENT ANALYSIS METHODS

Shreya Banker¹ and Rupal Patel²

¹Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, CHARUSAT, CHANGA

²Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, CHARUSAT, CHANGA

ABSTRACT

On today's fast spreading use of social media, many websites have offer reviews of items like books, cars, mobiles, movies etc. They describe the product in some detail and evaluate them as good/bad, preferred/not preferred, so it is necessary to categorize these reviews in an automated way. Sentiment analysis is one kind of computational technique of Artificial Intelligence. It is the task of identifying positive and negative opinions, emotions, and evaluations. This article represents Sentiment analysis, it's issues, applications and some of the methods used to evaluate the review using sentiment analysis.

KEYWORDS

Sentiment analysis, Review Mining, Opinion Mining, Data mining, Machine learning

1. INTRODUCTION

Human opinion is important for making decision or choosing from multiple options. Before the advent of the Internet, people were generally asking to friends or relatives for the product recommendations. Nowadays, with the growing use of social media, people have access to refer reviews on a different websites to take decision.

People's opinion and experience are very valuable in decision-making process, but to get profits from these opinion and experience, the accumulated content should be analysed properly. These analysed opinions are useful for consumer to get a chance to evaluate others view and involvement related to some products or services before purchasing them. Likewise, the tradesperson can take this review as feedback from the consumers, and they can improve the quality of their product or services.

Sentiment analysis also called "Opinion Mining." It is useful in several ways, such as social media monitoring. There are also many names and slightly different tasks, e.g., sentiment analysis [1], opinion mining [2], opinion extraction [3], sentiment mining [4], subjectivity analysis [5], affect analysis [6], emotion analysis [7], review mining [8], etc. Sentiment analysis involves classifying opinions into categories like "positive" or "negative" view. Approaches to this task mainly drawn from the areas of natural language processing [9], data mining [10], and machine learning [11].

2. LITERATURE REVIEW

There are different methods to perform sentiment analysis, such as keyword spotting, lexical affinity and statistical methods.

Sentiment analysis includes a process called keyword spotting that includes developing a list of keywords that relate to a certain sentiment. These words, known as affect words, are usually positive or negative adjectives because such words can be strong indicators of sentiment. Keyword spotting classifies text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Keyword spotting method used to detect fraud. Fraud examiners search for affect words in employee emails or other communications as part of their fraud detection procedures. For example, a fraud examiner might receive a clue that a manager is manipulating his division's reported sales to meet company quotas. In evaluating the merit of the clue, the examiner might run a list of keywords — flexible, unreasonable, temporary and worried — against the alleged party's emails and, if there are positive results investigate the issues further. The limitation of this approach lies in two areas: poor recognition of affects when negation is involved and reliance on surface features. About its first weakness, while the approach can correctly classify the sentence “today was a happy day” as being happy, it is likely to fail on a sentence like “today wasn’t a happy day at all.” About its second weakness, the approach relies on the presence of obvious affect words that are only surface features of the prose.

Lexical affinity is slightly more sophisticated than keyword spotting as, rather than simply detecting obvious affect words, it assigns arbitrary words a probabilistic ‘affinity’ for a particular emotion. For example, ‘accident’ might assign a 75% probability of being indicating a negative effect, as in ‘car accident’ or ‘hurt by accident’. Though often outperforming pure keyword spotting, there are some problems with the approach. Lexical affinity, operating solely on the word-level, can easily be tricked by sentences like “I avoided an accident” (negation) and “I met my girlfriend by accident” (other word senses). [23].

Statistical methods, such as Bayesian inference and support vector machines, have been popular for affect classification of texts. By feeding a machine learning algorithm a large training corpus of affectively annotated texts, it is possible for the system to not only learn the affective valence of affect keywords (as in the keyword spotting approach), but also to take into account the valence of other arbitrary keywords (like lexical affinity), punctuation, and word co-occurrence frequencies. The polarity of a word can be identified by studying the occurrence frequency of the word in a large annotated corpus of texts [27]. If the word occurs more frequently among positive texts, then its polarity is positive. If it occurs more frequently among negative texts, then its polarity is negative. If it has equal frequencies, then it is a neutral word. The similar opinion words frequently appear together in a corpus. This is the main observation that the state of the art methods are based on. Therefore, if two words appear together frequently within the same context, they are likely to have the same polarity. Therefore, the polarity of an unknown word can be determined by calculating the relative frequency of co-occurrence with another word. However, traditional statistical methods are generally semantically weak, meaning that, with the exception, of obvious affect keywords, other lexical or co-occurrence elements in a statistical model have little predictive value individually. As a result, statistical text classifiers only work with acceptable accuracy when given a sufficiently large text input. Therefore, while these methods may be able to affectively classify user’s text on the page- or paragraph- level, they do not work well on smaller text units such as sentences or clauses [23].

3. PRELIMINARIES OF SENTIMENT ANALYSIS

Following are the preliminaries required for the sentiment analysis as:

3.1. Natural Language Processing

With the amount of information that is shared on social media, forums, blogs, etc., it is easy to see why we need to automate sentiment analysis: there is simply too much information to manually process. The problem is that machine-learning approaches are typically not that accurate. For a simple task, separating ‘positive’ from ‘negative’ sentiment on social media, many automated solutions only perform with around 80% accuracy. This can still be useful for tracking broad trends over time, but it limits fine-grained analysis.

We need the natural language processing techniques to have a deeper understanding of the context. Compare these examples:

1. It was a great restaurant.
2. It should have been a great restaurant.
3. The restaurant was great in that it will make all future meals seem more delicious.
4. Despite a pleasant experience, I cannot support the many reviews that it was a great restaurant.

The first sentence is positive, but the rest are all negative sentiment, despite having the first sentence embedded within them. The second sentence “should have been” indicates the desired outcome, leaving the actual sentiment implied (it is possible that this sentence could follow by “... and it was!”), but the implication is that it was not). The third is the more complicated, bordering on the kinds of sarcasm that are very hard for machines to identify: even a person might misread this as positive when skimming quickly. The fourth statement is even more complicated because the overall sentiment is negative, but it begins with the (weakly) positive “pleasant experience” and also finishes by reporting that many other people expressed positive sentiment.

3.2. Opinion Mining

We use the following review about a Canon camera to introduce the problems: Posted by: John Smith Date: September 10, 2011

1. I bought a Canon G12 camera six months ago.
2. I simply love it.
3. The picture quality is amazing.
4. The Battery life is also long.
5. However, my wife thinks it is too heavy for her.

From this review, we identify some important issues as:

The review has a number of opinions, both positive and negative, about Canon G12 camera. Sentence (2) expresses a positive opinion about the Canon camera as a whole. Sentence (3) expresses a positive opinion about its picture equality. Sentence (4) expresses a positive opinion about its battery life. Sentence (5) expresses a negative opinion about the weight of the camera. From these opinions, we can make the following important observations:

An opinion consists of two key components: a target g and a sentiment s on the target. i.e. (g, s) where g can be any entity or aspect of the entity about which an opinion has been expressed, and s is a positive, negative, or neutral sentiment, or a numeric rating score expressing the strength/intensity sentiment analysis and opinion mining 18 of the sentiment (e.g., 1 to 5 stars). Positive, negative, and neutral are called sentiment (or opinion) orientations (or polarities). For example, the target of the opinion in sentence (2) is Canon G12, and the target of the opinion in sentence (3) is the picture quality of Canon G12. Target also called topic in the literature.

This review has opinions from two persons, which are called opinion sources or opinion holders. The holder of the opinions in sentences (2), (3), and (4) is the author of the review (“John Smith”), but for sentence (5), it is the wife of the author. 3. The date of the review is

September 10, 2011. This date is important in practice because one often wants to know how opinions change with time and opinion trends.

There are other factors to currently stop us from relying blindly on tools for sentiment analysis:

Context: A positive or negative sentiment word can have the opposite connotation depending on context.

Sentiment Ambiguity: A sentence with a positive or negative word does not necessarily express any sentiment (e.g., “can you recommend a good tool I could use?” does not express any sentiment, although it uses the positive sentiment word “good”). Likewise, sentences without sentiment words can express sentiment too (e.g., “This browser uses a lot of memory” does not contain any sentiment words, although it’s clearly negative at a document level.).

Sarcasm: A positive or negative sentiment word can switch sentiment if there is sarcasm in the sentence (e.g. “Sure, I’m happy for my browser to crash right in the middle of my coursework”).

Language: A word can change the sentiment and meaning depending on the language used.

This is often seen in slang, dialects, and language variations. An example is the word “sick,” which can change meaning based on context, tone, and language, although clear to the target audience.

4. LEVELS OF SENTIMENT ANALYSIS

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Depending on whether the target of study as a whole text or document, one or several linked sentences, or one or several entities or aspects of those entities, different NLP, and sentiment analysis tasks can be performed. Hence, it is necessary to distinguish three levels of analysis that will clearly determine the different tasks of sentiment analysis:

Document Level: Analysing the overall sentiment expressed in the text. This works best on the assumption that the whole comment only discusses one topic. Document level considers that a document is an opinion on an entity or aspect of it. This level is associated with the task called document-level sentiment classification. The task is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is known as document-level sentiment classification.

Sentence Level: The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification that distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions.

Aspect Level: Both the document level and the sentence level analysis do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Aspect level was called feature level (feature-based opinion mining and summarization). Aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion). It is closely related to tasks like Feature-based Opinion Mining and Opinion Summarization.

5. APPLICATIONS OF SENTIMENT ANALYSIS

As the online review sites, blogs, social networking sites provide huge amount of opinions, this helps in easier decision-making process. Some of the applications of sentiment analysis are discussed as below:

Most of the e-commerce activities use sentiment analysis. Most of the websites allow users to give their feedback about shopping and quality of products. They give the information about different features and summary of product by giving ratings to products. Then, the users can view reviews, read opinions about products and their special features. Moreover, summary of product and its features is present to users in graphical form. Most of the popular commercial websites like amazon.com give the information about reviews from customers with rating information. By analysing these large volumes of opinions, Sentiment Analysis helps the websites by changing dissatisfied customers into promoters [24].

Twitter and Facebook are a focal point of many sentiment analysis applications. The most common application is monitoring the reputation of a specific brand on Twitter and/or Facebook. One application that performs real-time analysis of tweets that contain a given term is tweetfeel (<http://www.tweetfeel.com>) [25].

Voice of the Customer helps in determining what the individual customer is saying about products or services. It means examining the reviews, opinions, and feedback of the customers. Voice of the Customer is an important part of Customer Experience Management. To invent new products, VOC helps in finding such new opportunities. Extracting customer opinions also helps identify functional requirements of the products and some non-functional requirements like performance and cost [24].

Sentiment analysis helps government in assessing their strength and weaknesses by analyzing opinions from public. For example, “If there is situation like, The MP who is investigating 2G scam himself is deeply corrupt then how do you expect that truth will 4 come out?.” This example shows the negative sentiment about government. Sentiment analysis also helps in identifying strengths and weaknesses in a recruitment campaign in government job, assessing success of electronic submission of tax returns, and many other areas [24].

Sentiment analysis can provide substantial value to candidates running for various positions. It enables campaign managers to track how voters feel about different issues and how they relate to the speeches and actions of the candidates.[25]

Another important domain for sentiment analysis is the financial markets. There are numerous news items, articles, blogs, and tweets about each public company. A sentiment analysis system can use these various sources to find articles that discuss the companies and aggregate the sentiment about them as a single score that can be used by an automated trading system. One such system is The Stock Sonar.[25]

6. CONCLUSIONS

Sentiment analysis has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. Sentiment classifiers are dependent on topics. It has found that different types of features and classification algorithms combined in an efficient way in order to overcome their individual drawbacks, benefit, and finally enhance the sentiment classification performance. In future, more work is needed on further improving the performance. The techniques and algorithms used for sentiment analysis are advancing fast; however, many problems in this field of study remain unsolved. More future research could be dedicated to these challenges.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the many people who saw us through this article; to all those who provided support, allowed us to quote their remarks and assisted in the editing, proofreading and design.

We would like to thank Dr. Atul Patel for enabling us to publish this paper and who encouraged us lots throughout the entire survey.

We would like to thank Dr. Pritpal Singh for helping us in the process of selection and editing. Special thanks to Mr. Kanu Patel our publisher who encouraged and guided us during the whole process.

REFERENCES

- [1] Wikipedia Retrieved from https://en.wikipedia.org/wiki/Sentiment_analysis
- [2] Sentiment Mining Retrived from <http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining>
- [3] Retrived from <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
- [4] Retrived from <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
- [5] Retrived from <http://www.aclweb.org/anthology/H05-2018.pdf>
- [6] Retrived from <http://dl.acm.org/citation.cfm?id=1972851>

- [7] Retrived from <https://www.kairos.com/emotion-analysis-api>
- [8] Retrived from <https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf>
- [9] Retrived from https://en.wikipedia.org/wiki/Natural_language_processing
- [10] Retrived from https://en.wikipedia.org/wiki/Data_mining
- [11] Retrived from https://en.wikipedia.org/wiki/Machine_learning
- [12] Li Zhuang, FengJing,Feng Jing. "Movie Review Mining and Summarization" (2006): 43-50.
- [13] BakhtawarSeerat, FarouqueAzam. "Opinion Mining: Issues and Challenges (A survey)." International Journal of Computer Applications (0975 – 8887) 49 (2012): 1-10.
- [14] Vivek Narayanan, IshanArora, Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model" : 1-8.
- [15] AnasCollomb, CrinaCostea,Damien Joyeux,Omar Hasan,Lionel Brunie. "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation."
- [16] Bongirwar, Vrushali K. "A Survey on Sentence Level Sentiment Analysis." International Journal of Computer Science Trends and Technology (IJCTST) 3.3 (2015): 110-113.
- [17] G.Vinodhini, RM.Chandrasekaran. "Sentiment Analysis and Opinion Mining: A Survey." International Journal of Advanced Research in Computer Science and Software Engineering 2.6 (2002).
- [18] Mr. Saifee Vohra, Prof. Jay Teraiya. "Applications and Challenges for Sentiment Analysis : A Survey." International Journal of Engineering Research & Technology (IJERT) (2013): 1-5.
- [19] Noura Farra, Elie Challita, Rawad Abou Assi, Hazem Hajj. "Sentence-level and Document-level Sentiment Mining for Arabic Texts." 2010 IEEE International Conference on Data Mining Workshops (n.d.).
- [20] Pang, Bo. "AUTOMATIC ANALYSIS OF DOCUMENT SENTIMENT." (2006).
- [21] V.K. Singh, R. Piryani, A. Uddin,P. Waila. "Sentiment Analysis of Movie Reviews." (n.d.): 712-716.
- [22] Wouter Bancken, Daniele Alfarone and Jesse Davis. "Automatically Detecting and Rating Product Aspects from Textual Customer Reviews."1-16.
- [23] Erik Cambria. "An Introductionto Concept-Level Sentiment Analysis."478-483.
- [24] Sujata Rani "Rule Based Sentiment Analysis System".
- [25] Ronen Feldman "TechniquesandApplicationsfor SentimentAnalysis" (2013).
- [26] Aurangzeb Khan,Baharum Baharudin,and Khairullah Khan,Sentiment classification from online customer reviews using lexical contextual sentence structure.pages 317-331.Springer,2011
- [27] Read J, Carroll J. Weakly supervised techniques for domainindependent sentiment classification. In: Proceeding of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion; 2009. p. 45–52.

Authors

Shreya Banker received her M.C.A. degree from G H Patel Post Graduate Department of Computer Science and Technology, Sardar Patel University, V.V.Nagar, Gujarat in 2010 & B.C.A. degree from SEMCOM College, Sardar Patel University, V.V.Nagar, Gujarat in 2007. She is presently working as an Assistant Professor in Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, Changa. Her research area includes Machine Learning and Data Mining



Rupal Patel received her M.C.A. degree from ISTAR, Gujarat Technological University, V.V.Nagar, Gujarat, India in 2012, & B.C.A. degree from DDIT, Dharmsinh Desai University, Nadiad, Gujrat, India in 2009. She is presently working as an Assistant Professor in Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, Changa. Her research area includes Mobile Computing and Database System.

