ISSN 2723-6471

Survey Opinion using Sentiment Analysis

Taqwa Hariguna [1,*], Husni Teja Sukmana [2], Jong Il Kim [3]

[1] Information System Program, Amikom Purwokerto University, Indonesia
[2] Informatics Program, Syarif Hidayatullah University, Indonesia
[3] Wookyung Information Technology, South Korea
[1] taqwa@amikompurwokerto.ac.id; [2] husniteja@uinjkt.ac.id; jikim2@wkit.co.kr [3]
* corresponding author

(Received: July 15, 2020; Revised: July 21 2020; Accepted: August 29, 2020; Available online: September 1, 2020)

Abstract

Sentiment analysis or opinion mining is a computational study of the opinions, judgments, attitudes, and emotions of a person towards an entity, individual, issue, event, topic, and attributes. This task is very challenging technically but very useful in practice. For example, a business always wants to seek opinion about its products and services from the public or the consumers. Additionally, potential consumers want to learn what users think they have when using a service or purchasing a product. To get public opinion on food habits, ad strategies, political trends, social issues and business policy, this is a very critical factor. This paper will explain a survey of key sentiment-extraction approaches.

Keywords: Opinion mining, Business, Machine learning, Sentiment Analysis, Survey

1. Introduction

Sentiment analysis or opinion mining, is a field of research that analyzes the thoughts , feelings, perceptions, decisions, attitudes and emotions of individuals towards entities such as goods, services, organisations, persons, issues, events, topics and their attributes. It provides a very wide space for issues. However there are a range of slightly different titles and functions, including sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, review mining, etc. They 're all under study of sentiment or opinion mining now, though. The term sentiment analysis is more widely used in the manufacturing sector but both sentiment analysis and opinion mining are often used in the education field. They serve essentially the same area of analysis.

Various forums, blogs , social networks, e-commerce portals and news stories serve as a medium for sharing views that can be used to identify people's and customers ' perceptions about current issues , political trends, business tactics, marketing promotions, user tastes, and credibility monitoring[1]. In the past decade and a half, researchers and academics have worked hard on an analysis of sentiments to complete these tasks. Sentiment analysis is a computational study of opinions , feelings, emotions , and attitudes expressed towards entities in the text[2]. Sentiment analysis or mining is the task of detecting, extracting and classifying opinions, sentiments and attitudes on different topics, as stated in text input [3].

Given the increasing growth in e-commerce, which is also the key outlet for sharing and evaluating thoughts , feeling, appraisal and analysis become very important. Customers on e-commerce sites currently rely on reviews posted by existing clients, manufacturers and service providers in order to analyze customer opinions to increase the quality and standards of their services and products. For instance, opinions given on e-commerce sites like Amazon can influence customer decisions in purchasing certain products[4].

2. Literature review

Opinion mining or sentiment analysis is a method of subjectivity analysis that attempts to identify points of view, feelings and judgments expressed in natural language. This has been studied in recent years by a lot of researchers, in order to predict the sentiment orientation (negative or positive) they analyzing sentiment or opinion word and expressions in sentences and documents. Factor behind this study is:

- Improvement on machine learning techniques in the analysis of natural language and information extraction.
- The accessibility of data sets for the training of machine learning algorithms due to the growth of the World Wide Web and, in particular, the creation of websites for review aggregation.
- Awareness of the interesting theoretical obstacles caused by the environment and business and technology applications.

Sentiment analysis has been used for such a many purpose like tracking the popularity for some brand[5], Analyzing product potency[6], predicting market value, improving product quality, etc. Based on Pirolli[7] research sentiment analysis has a role as a helper for a researcher or data finder to get and identify relevant information from someone (customer, user, participant). In general, the sentiments analysis was studied mainly on three levels:

Document Level, At this point the challenge is to determine how all opinion papers convey negative or positive sentiment[8]. For example, when you get a review of the product, the system determines whether the review expresses an overall positive or negative view of the product. This task is commonly referred to as classification of the sentiment at document level. This analytical level assumes that each text is voicing an view about a particular person. That does not apply, however, to documents which assess other entities.

Sentence Level, At this point, the assignment is a phrase and decides each sentence to convey a positive, negative , or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to the classification of subjectivity[9], which differentiates sentences (called objective sentences) expressing factual information from sentences (called subjective sentences) expressing subjective views and opinions. We must note, however, that subjectivity is not equivalent to sentiment, because most objective phrases can express an opinion.

Entity and Aspect Level, If the analysis of document level and sentence level does not find what people really like and dislike. Aspect level does a narrower analysis. The previous aspect level is referred to as Feature Level[10]. The level of aspects directly see the opinion itself compared to seeing language construction. This level is based on the idea that opinion is based on feeling (positive or negative) and goal (opinion). Opinions are of limited use without identified targets.

Being aware of the importance of targets for opinion can also help us better understand the problem of analyzing sentiments. For example, although there is clearly a positive tone in the phrase "even though the service is not very good, I still like this cafe," we can't say that this phrase is entirely positive. The sentence itself was actually positive (emphasized) about cafe, but negative about service (not emphasized). In most applications , different entities and/or aspects explain opinion targets. The purpose of this level of analysis is, therefore, to find sentiment about the entity and/or its aspects.

In order to decide which expressions are positive or negative sentiment analysis also requires certain resources like sentiment lexicons. There are various methods have been developed for understanding the polarity of sentiment analysis. Most of those recent studies have been focused on document level analysis for determining the sentiment orientation of a document. However, when in-depth sentiment analysis of review texts is requested these document level sentiment analysis approaches are less efficient. Lately researchers have tried out sentence level sentiment

ISSN 2723-6471

analysis to analyze and obtain opinions about different aspects of a examined text. For most instances, a sentence level sentiment analysis seems to be more advanced than a document level one.

3. Research Method

3.1. Emoticons

The simplest way we can detect the polarity of a message is based on the emoticons it contains. Emoticons are really popular these days, especially for teenagers. Emoticons shortened for "emotion icon" or emote, is a textual depiction of a facial expression using characters normally punctuating symbols, numbers, and symbols to illustrate a person's emotions or mood, or as a time-saving tool.

Table. 1. Emoticons and their variations

<table>
<tr><th>Polarity</th><th>Symbols</th></tr>
<tr><td>Neutral</td><td>: | =| :-| >-< >.< >0< :o :x -_-</td></tr>
<tr><td>Positive</td><td>:) :] : } :o) :o] :o} :-] :=) :B :-D :-B ^.^ ^-^ ^_^ ^^ :* <3 =b ;) ;]</td></tr>
<tr><td>Negative</td><td>D: D= : (:[:'(o_0</td></tr>
</table>

3.2. SentiStrength

SentiStrength is software for sentiment analysis. it's indicates the positive and negative strength sentiment in short texts. Except for political texts, It has high accuracy for short social web texts in English. This software feature is a list of positive and negative keywords and a database of booster words to strengthen (very, so much, highly, etc) or weaken (awful, bad, ugly, etc) sentiments. SentiStrength, utilizes a combination of learning techniques which give the best results and empirically obtained best training models.

3.3. SenticNet

SenticNet[11] is an opinion mining and sentiment analysis method that explores the techniques of artificial intelligence and semantic Web. The aim of this method is to assume the concepts polarity of common sense from the natural language text on a semantic rather than syntactic level. To Create polarity this method using NLP or Natural Language Processing techniques. For example, SenticNet will translate a message "I hate monday", then SenticNet will identify the concept, which are "Hate" and "Monday". After that it will give a polarity score to each concept, -0.478 for "hate" and +0.119 for "Monday". The result of this example is -0.359 which is pretty bad.

4. Comparative opinion analysis

Besides expressing positive or negative opinions directly over an entity and its aspects, people can also express opinions through comparison of similar entities. Such views are referred to as contrasting views[12][13]. It is related to comparative opinion, but also different from ordinary opinion. Not only do they have different semantic significances, they also have different syntactic forms. For illustration , a standard opinion phrase is "the screen quality of this cell phone is great," and a traditional comparative opinion phrase is "the quality of the iPhone screen is better than Nokia's." This comparison phrase does not claim whether any cell phone 's screen quality is good or poor, but is enough to compare. Because of all these dissimilarities, comparative opinions require different techniques for analysis. Comparative sentences can also be criticized or not criticized, as in ordinary sentences. The above comparative statement is an opinion as it expresses its maker's comparative feelings directly, whereas the expression "iPhone 1 inch wider than a standard Nokia smartphone" communicates no emotion.

4.1. Problem Definitions

Because of all these dissimilarities, comparative opinions require different techniques for analysis. Comparative sentences can also be criticized or not criticized, as in ordinary sentences. The above comparative statement is an opinion as it expresses its maker's comparative feelings directly, whereas the expression "iPhone 1 inch wider than a standard Nokia smartphone" communicates no emotion.

A. **Gradable comparison:** These comparisons show the ordering relationships of the comparing entities. There are three subtypes:

- *Non-equal gradable comparison:* this reflects the interaction of a superior or lesser sort, varying from one group of entities to a separate set of entities depending on any of their common features. "KFC tastes better than McDonald's," for instance. This form also contains a choice, "I prefer KFC over McDonald's."
- *Equative comparison*: It states the relationship of the same type by stating 2 or more entities are the same based on several aspects of them together, for example, "KFC and McDonald's taste the same."
- *Superlative comparison*: This expresses the relationship of a type that is greater or not compared to all the others, which ranks one entity over another, for example, "Among all fast foods KFC tastes the best."

B. **Non-gradable comparison:** This kind of comparison expresses, but does not value, the relationship between two or more entities. Three sub-types are:

- Entity A is similar or different from entity B based on several aspects shared, for example, "KFC tastes different from McDonald's."
- Entity A has aspects x1, and entity B has aspects x2 (x1 as well as x2 can be substituted), "Computers use external speakers but laptops use internal speakers" for example.
- Entity A has an aspect of x, but it was not for entity B. For instance, "Nokia mobile phones come with earphones, but iPhone didn't."

In this article, we concentrate only on the gradual comparison. Non-gradual comparisons can express opinions as well but are often more subtle and hard to recognize.

4.2. Identify Comparative Sentences

While there are comparative and superlative keywords in most of the comparative sentences. For example, many sentences that contain these words, but not comparative sentences, are great, superior, and best. For example, "I can’t be more agree with you."

It is shown, according to [12], that nearly all comparative sentences possess keywords (words or phrases) which show comparison. Using a sequence of keywords, 98% (recall = 98%) of the comparative sentences were described with 32% precision based on their data collection. Keywords for this are:

1. Comparative adjective (JJR) and comparative adverb (RBR). For example, higher, taller, better, and words that end with -er. This counts as only two keywords.
2. Superlative adverbs (RBS) and superlative adjectives (JJS). Examples, Most, few, and best. that's often counted as only two keywords.

ISSN 2723-6471

3. Other words and phrases not standard indicative like help, defeat, win, exceed, surpass, prefer, ahead, than, superior, lower, fight, number one, etc. Individually measured in the number of keywords.

Since only keywords can reach high withdrawal rates, they can also be used to filter out certain sentences that may not be comparative. We also need to boost the exactness of the remaining words.

It was also observed that[12] comparative sentences have a consistent pattern containing comparative keywords which isn't really shocking. This pattern can be utilized as a learning function. Class Sequential rule (CSR) is used to identify this pattern. A special type of sequential pattern mining is the Class Sequential Mining (CSR) rules. Each example of training is a pair (x1 , y1), in which x1 is the sequence and y1 is the class label. y1 $\in$ {comparison, non-comparison}. Sequences are formed from words. CSR will be generated by using the training data.

The left-hand sequence pattern of CSR rules with high conditional probabilities is used as a feature for classification building models. To build that model, Naïve Bayes was hired. The problem is also studied in[14] but in the Korean language context. The algorithm used for learning is transformation-based learning, which produces rules.

Classification of comparative sentences into four categories: The algorithm also classifies them into four types after comparative sentences are defined, gradable not-equal, equative, superlative, and non-gradable. [12] shows that unique keywords and phrases as features are enough for this mission.

5. Discussion

The performance of various methods used for opinion mining is evaluated by measuring different parameters such as precision, recall and F-measure. Precision is the fraction of relevant instances retrieved, although recall is the fraction of relevant instances that are retrieved.
Table 2 presents an average of the results obtained for each labeled dataset to compare the results of prediction output for each method. A score of 1, and 0 is the worst possible, is ideal for the F-measure. The best F-measure method was Emoticons(0.846) that had the least coverage. The 2nd greatest F-measure method is SentiStrength, which obtained a much higher coverage compared to Emoticons. It's important to remember that version of SentiStrength we 're using is already qualified, potentially with the entire dataset. Thus, conducting SentiStrength experiments using this dataset could theoretically be biased, because training and research would be performed using the same dataset. Instead we calculate SentiStrength 's predictive performance metrics based on the numbers identified in their experiments.

Table. 2. Average prediction performance for all labeled dataset.

Metric	Emoticons	SenticNet	SentiStrength
Recall	0.856	0.562	0.767
Precision	0.867	0.934	0.780
Accuracy	0.817	0.590	0.815
F-measure	0.846	0.658	0.765

6. Conclusion

This paper provides an examination of the mining of feelings or opinion. It has become very active research in recent years thanks to many challenging research problems and various practical applications. It has also spread from the computer science to the science of management. Although there have been several works, the comparative sentences from sentiment analysis have not been studied as extensively as other topics. this is still a need for more research.

One difficult question is how many types of non-standard or implied comparative sentences can be found. For instance, "I'm really happy my iPhone isn't like my old ugly smartphone." Further processing of feelings is difficult without defining. In order to identify comparative sentences and their types, researchers are also studying the extraction of comparable entities, comparative aspects and comparative words. Jindal and Liu[13] use mining sequential labeling rules in the context of sequential pattern-based learning methods.

References

[1] M.R. Saleh, M.T. Martín-Valdivia, A. Montejo-Ráez, L.A. Ureña-López, Experiments with SVM to classify opinions in different domains, Expert Systems with Applications 38 (2011) 14799–14804

[2] W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http://dx.doi.org/10.1016/j.asej.2014.04.011

[3] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679.

[4] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2 (2012) 1-8

[5] Greco, Francesca, and Alessandro Polli. "Emotional Text Mining: Customer profiling in brand management." *International Journal of Information Management* 51 (2020): 101934.

[6] Jeong, Junhyung, et al. "Harmonics suppressed band-pass matching network for high efficiency power amplifier." *2017 47th European Microwave Conference (EuMC)*. IEEE, 2017.

[7] Pirolli, Peter. *Information foraging theory: Adaptive interaction with information*. Oxford University Press, 2007.

[8] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002). 2002.

[9] Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. in Proceedings of the Association for Computational Linguistics (ACL-1999). 1999.

[10] Hu, Minqing and Bing Liu. Mining and summarizing customer reviews. in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004). 2004.

[11] Cambria, Erik, et al. "Senticnet: A publicly available semantic resource for opinion mining." *2010 AAAI fall symposium series*. 2010.

[12] Jindal, Nitin and Bing Liu. Identifying comparative sentences in text documents. in Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-2006). 2006.

[13] Jindal, Nitin and Bing Liu. Mining comparative sentences and relations. in Proceedings of National Conf. on Artificial Intelligence (AAAI-2006). 2006.

[14] Yang, Seon and Youngjoong Ko. Extracting comparative entities and predicates from texts using comparative type classification. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011). 2011