International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT206384

A Literature Review on Sentiment Analysis

Vandana C P[*1], Siddharth Indoria[2], Sharan Gouda[3], Sinchana Bhaskar[4]
[*1]Assistant Professor, Information Science, New Horizon College of Engineering, Bangalore, Karnataka, India
[2,3,4]Information Science, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This paper is an intend to consolidate the review and perform the literature survey on the sentiment analysis and on opinion mining. In this paper we try to analyze people sentiments, opinions, and emotions from their text language by which we can try to understand in what mood or emotion was the person while writing the text message. There are many types of sentimental moods according to which person writes the text it can be classified like happy, sad, neutral, angry. Also there are times when the user can be sad and angry at the same time which is needed to be identified by the analysis.

Keywords : Opinion Mining, Sentiment Analysis

I. INTRODUCTION

Natural language processing has a task called sentiment analysis to understand if a text message contains some important information and what important information it expresses, whether the point of view behind this text is positive, negative or neutral. Understanding the opinions behind user-generated content automatically is of great help for commercial use, among others. The process can be conducted on different levels, classifying the polarity of words, sentences or entire documents.

In this process, we propose a method of elementary discourse unit (EDU) level sentiment analysis using discourse features. When the process starts we experiment that when we want to predict the sentiment of a certain EDU.

Rapid increase in internet users and with that the online review sites was one of the main reason of the invention of the analysis of the sentiment by which we can understand and predict what people are thinking about any particular product, issues, services. The internet users are interested to look for the positive and negative opinions which have likes and dislikes and also its is been shared by the users by which they can find out others opinion.

Sentiment analysis can solve some important issues of the company and provide answer to it, the analysis can be Automated considering the customers perspective and by that the decision can be made based on the data and is useful for quick gaining using large volumes of text data. In that by adding the customer feedback analysis use case, which we saw above and getting the output. The visualize results of sentiment analysis many people uses techniques such as graphs emoji. Because of the results when the output is un the visual form it represents in a good way.

Businesses now a days receive a large amounts of information. Daily more than 2.5 quintillion bytes of data are produced. this makes impossible for human beings to analyze the data manually and get the results accurate and efficient. Instead of this, there is a model called machine learning which can perform automatically tag your texts in just seconds called as sentiment analysis, it does not matter if you receive 10 or 1,000 reviews or social media comments.

Sentiment Analysis has three different main levels According to Bing Lui. The three levels of the Sentiment Analysis. 1. Levels of Sentiment Analysis

a. Document level Sentiment Analysis

In this Sentiment Analysis level whole document has analyzed and classify whether a whole opinion document expresses a positive or negative sentiment. In single document only reviews of single product has been reviewed. And task is to find out the opinion about that product. So this kind of task is known as document-level sentiment classification. In this level, expressed opinion is on single entity.

If a document contains many product reviews this is not applicable

b.Sentence Level Sentiment Analysis

In this level, task goes to every sentence and tries to understand whether the sentence expresses the positive, negative or neutral opinion. This level attentively suitable to Subjectivity Classification [4], which separates objective sentences and subjective sentences. Objectives sentences express factual info about sentences where Subjective sentences express the subjective info about sentences. Many objective sentence can involve opinions. This task is majorly known as Sentence Level Sentiment Analysis.

c.Aspect level Sentiment Analysis

Aspect Level sentiment Analysis was earlier called Feature level Sentiment Analysis. Document and Sentence Level Sentiment Analysis do not find out the history of what people like or did not like. It achieves finer-grained analysis. In this level mainly focuses at the opinion instead of looking to documents, paragraphs, sentences, or phrases. This level consider the entity, aspect of that particular entity, opinion of the aspect, opinion holder and time. Because of these limits this level can find what people like means which feature of product mostly likes by customers and also on which time. This task is more interesting and more difficult too.

II. LITERATURE SURVEY

When applied to a single type of text those techniques typically have a range of accuracy from 70% to 80% depending on amount of human input and type of text [7].

The users can use extractive and abstractive methods from which efficient results are achieved. For generating compressed and readable information for users, the hybridization technique proposed here proves to be highly efficient as per the test results [5]. A number of machine learning like Naïve Bayes and Random Forest models performed sentiment analysis on product review data [10].

Some work in this field includes experiments with mood classification on blog posts. One of the researches also deals with review of aspect-based opinion polling from unlabeled free-form textual customer reviews without requiring customers to answer any questions [11].

At the end, they basically measured the performance of classifier in terms of recall, precision and accuracy. Bo Pang et al used machine learning techniques to investigate the effectiveness of classification of documents by overall sentiment. Experiments demonstrated that the machine International Journal of Computer Applications (0975 – 888) Volume 47– No.11, June 2012 37 learning techniques are better than human produced baseline for sentiment analysis on movie review data. The experimental consists of movie-review corpus with randomly taken 500 positive sentiment and 500 negative sentiment reviews.

Alekh Agarwal et al proposed a machine learning method incorporating linguistic knowledge assembled through synonymy graphs, for more effective opinion classification.

III. EXISTING SYSTEMS

Sentiment analysis can be performed on three different levels:

- review-level
- sentence-level
- Phrase-level

Sentence-level analysis and review -level analysis tries to classify the sentiment of a whole review to one of the predefined sentiment which, including positive, negative and neutral.

While phrase-level analysis tries to extract sentiment polarity of each feature that a user expresses his/her attitude to the specific feature of a specific product.
Zhang et al. propose a self-supervised and lexicon-based sentiment classification approach to determine sentiment polarity of a review that has both emoticons and textual words. And they use sentiment for recommendation.

Lee et al. propose a recommended system using the concept of Experts to find both novel and relevant recommendations.

IV. Proposed Model

We propose a sentiment based grade prediction method in the structure of matrix. In this process we use the social media users sentiment to denote ratings accordingly.

In starting, we take texts which can be a sentence or a paragraph from user reviews. Then, we find out the sentiment words, which are used to describe the mood of the user. By sentiment dictionaries to calculate sentiment of a specific user.

In sentiment analysis polarity of many of the words are domain and context specific. For example when we consider the word "long" it is positive in "long battery life", but somehow negative in "long shutter lag". Excluding such type of expressions lead to poor coverage while tagging them with overall polarity tendency may lead to poor precision.

The main contributions of our approach are as follows: We propose a user sentimental measurement approach, which is based on the mind sentiment words in user reviews.

For prediction rating We make use of sentiment. User sentiment focuses on the user different words. User sentiment influence reflects the rating scale and the results.

We fuse the two factors: user sentiment depending if the text is positive or negative and then the process of the sentiment analysis is done by gathering the data, preparing the data for the process of creating sentiment model and in the visualization of results.

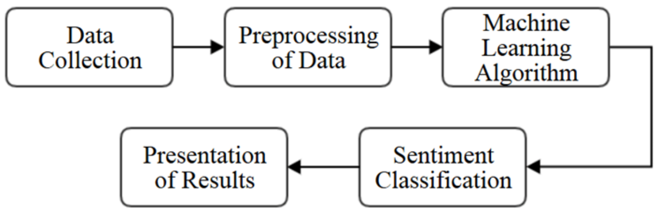


Figure 1

V. Proposed Approach

In the field of machine learning , classification methods have been developed, which the unlabeled are classified by different strategies. Classifiers could possibly require training data. Examples of machine learning classifiers are Naive Bayes.

Naïve Bayes This is a classification method that relies on Bayes' Theorem with strong (naive) independence assumptions between the features. A Naive Bayes classifier awaits that the closeness of a specified quality (element) in a class is detached to the closeness of some other elements. For example, an organic fruit might be considered to be an apple if its color is red, its shape is round and it measures approximately three inches in breadth. Regardless of whether these features are dependent upon one another or upon the presence of other features, a Naïve Bayes classifier would consider these properties independent due to the likelihood that this natural fruit is an apple. Alongside effortlessness, the Naive Bayes is known to out-perform even modern order strategies. As in, it is made to simplify the computation and in this sense considered as "Naive".

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Where p (a/b) is the posterior probability of class a given predictor b and p (b/a) is the likelihood that is the probability of predictor b given class a. The prior probability of class a is denoted as p (a), and the prior probability of predictor p is denoted as p (b).

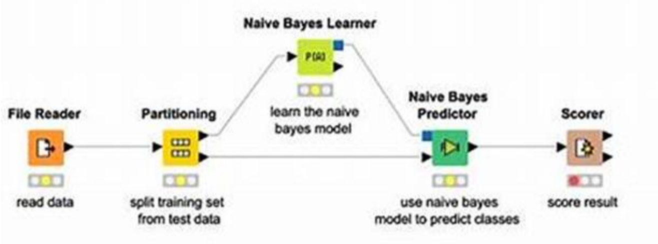


Figure 2

Applications to Review-Related Websites

a) Applications as a Sub-Component Technology Detecting antagonistic, heated language in mails, spam detection, and many more.
b) All type Movie Reviews, Product based Reviews etc.
c) Applications in Business and Government Intelligence Knowing Consumer attitudes and trends
d) Applications across Different Domains

Sentiment analysis cycle

Sometimes authors go even further and present methods for specific text format, for instance reviews where positive and negative features are explicitly separated is different areas. These above approach is shown by Hu and Liu in their work of customer reviews analysis

Figure 3

Sentiment analysis cycle

Pick up any data As you read, you form opinions about the character and prospects of the myriad companies featured in the daily news. Your brain arrives at a "sentiment" score based on a rubric of positive, negative, or neutral emotions stimulated by the text. In the computer science identical of reading the news, sentiment analysis is the system which processes the attributes from words taken from text mining.

What is clear from looking at a page in the newspaper, text heavily outnumbers numerical information. Charts and graphs are matched by recollections, and quotes. Financial analysis, last constrained to price tag ratios and differences, is currently undergoing a sentiment revolution.

•Challenges in sentiment analysis

Implicit Sentiment and Sarcasm

A sentence may have an implicit sentiment even without the presence of any sentiment bearing words.

Pragmatics

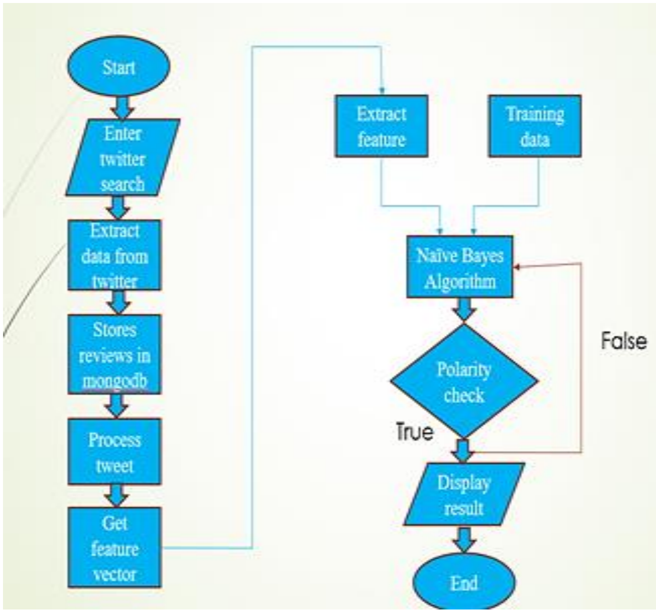
It is important to determine the pragmatics of end user opinion which may change the sentiment.

Subjectivity Detection

This is to differentiate between opinionated and non-opinionated text.

This is mostly used to improve the performance of the system by adding a subjectivity detection module to filter out objective reality.

Flowchart



Pang-Lee et al. (2002)[13] broadly classifies the applications into the following categories. a. Applications to Movie Reviews on websites, Product based Reviews and many more. b. Applications as a Sub-Component Technology Detecting antagonistic, abusive language in mails, spam detection, context sensitive information detection etc. c. Applications in Business and Government Intelligence Knowing Consumer point of view and trends in the market d. Applications across Different Domains Knowing public opinions for their political parties or party leaders or about new or old rules and regulations in a certain place etc

VI. CONCLUSION

First We provide a survey and then comparative study of existing techniques for opinion mining including machine learning and lexicon-based approaches, together with cross domain and cross-lingual methods and some evaluation metrics.

Research results show that machine learning methods, such as SVM and naive Bayes have the most high accuracy when compared to other and can be considered as the baseline learning technique, while lexicon-based system are very effective in some cases, which require few effort in human-labeled document We can focus on the study of merging machine learning method into opinion lexicon method to improve the accuracy and perfomance of sentiment classification and adaptive capacity to variety of domains and different languages.

Finally, it can be mentioned that our analysis doesn't take many factors into account. when talking First, our dataset doesn't really map the real public sentiment, it can only consider the twitter using, english speaking people. As studied it is highly possible to analyze mostly a perfect kind of mood. It may be considered that people's mood indeed affect their investment decisions, hence the correlation. But in that case, there's no direct correlation between the people who invest in stocks and the one who use twitter more frequently, though there is an indirect correlation - investment decisions of people may be affected by the moods of people around them, the general public sentiment. All these remain as areas of future research.

VII. REFERENCES

[1]. Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.

[2]. Ji fang, Bi Chen. Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classification. Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP 2011, pages 94–100, Chiang Mai, Thailand, November 13, 2011

[3]. Turney, Peter (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417–424.

[4]. 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 (ACL1999). 1999.

[5]. Mihai Dascălu, et.al, (2011)

[6]. Hu, M., Liu, B., “Mining and Summarizing Customer Reviews”, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004

[7]. Liu, B., Opinion Mining and Summarization, World Wide Web Conference, Beijing, China, 2008

[8]. Pak and P. Paroubek. “Twitter as a Corpus for Sentiment Analysis and Opinion Mining”, In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010.

[9]. 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.Yang Long, Yiping Gong,

[10]. Xing Fang, Justin Zhan. Sentiment analysis using product review data. Journal of Big Data 2015.

[11]. S. A Kanade, S. Shibu and Abhishek Chauhan. Review of Aspect Based Opinion Polling. IJREST 2014.

[12]. Neethu M S, Rajasree R. Sentiment Analysis in Twitter using Machine Learning Techniques. IEEE 2013.

[13]. Pang, Bo and Lee, Lillian and Vaithyanathan, Shivakumar, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of EMNLP 2002)

[14]. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques”, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002

[15]. Alekh Agarwal and Pushpak Bhattacharyya, “Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified”, In Proceedings of the International Conference on Natural Language Processing (ICON), 2005.

Cite this article as :

Vandana C P, Siddharth Indoria, Sharan Gouda, Sinchana Bhaskar, "A Literature Review on Sentiment Analysis", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 3, pp.357-362, May-June-2020. Available at doi : https://doi.org/10.32628/CSEIT206384

Journal URL : http://ijsrcseit.com/CSEIT206384