ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Survey on Sentiment Analysis

Shabnam, Neha Gupta

Assistant Professor, Dept. of C.S, JMIT Radaur, India

M. Tech Student, Dept. of C.S, JMIT Radaur, India

ABSTRACT: Internet has become an important part of in each and everyone life. It is mainly useful for user to share their opinion and views in a very short time .Sentiment analysis means extracting important information from user's views. The extracted information is helpful in decision making process. It aims to analyze user's sentiments, opinions, attitudes, emotions, etc. around elements like individuals, topics, services and products. Sentiment analysis is useful in social media observing to necessarily characterize the overall mood or feeling of consumers as reflected in social media regarding a specific brand or company and decide even if they are viewed positive or negative on the web. Sentiment classification of comments and reviews are the most useful execution in the area of sentiment analysis. Many software tools and different techniques are being developed to move out sentiment analysis. Feature based sentiment and bag of words are the popular approaches to deal with sentiment analysis used by researchers of opinions about product such as movies, cars, electronics etc.

KEYWORDS: text mining, sentiment analysis, sentiment classification, bag of words, feature based sentiment.

I. INTRODUCTION

Sentiment analysis is an information gathering task to get user's feelings. By analyzing a huge number of documents, these feelings can be manifest in positive or negative ways in the form of questions, request and comments. Generally, sentiment analysis is used to find the opinion of a writer about any topic or the overall sentiment of a text and document. Sentiment analysis crosses natural language processing and text analytics to find and extract essential information in source materials.

In recent years, the content growth in internet has made a huge volume of information available. This information is available in different formats namely comments, reviews, posts, news articles. From the sentiment aspect, we have two types of textual information such as facts and opinions. Facts means objective statement that tells the nature of product, opinions set out emotions, attitudes about the product. Sentiment analysis is an interdisciplinary field which is the application of natural language processing, artificial intelligence, and text analysis. Since many opinions are accessible in the text format and its processing is straightforward than other formats, sentiment analysis has appeared as a subfield of text mining [14]. Sentiment analysis is used in many domains such as shopping, education, entertainment, marketing, politics, development and research.

From the technical aspect, sentiment analysis has two approaches that are Bag of Words (BOW) and Feature Based Sentiment (FBS) [13]. In the BOW approach, every document is seen as a set of words. As a result, the semantic and syntactic information between words are lost. The BOW approach is not useable when estimation about products and their features have to be examined. In such type of cases, it is need to extract features. FBS has appeared as an approach for determining the sentiments of the products and their features.

Sentiment analysis can be performed at three distinct levels, viz. Document Level, Sentence Level and Feature or Aspect Level. Document level sentiment analysis means classifying the overall sentiments expressed by the author in the whole document text in positive, negative or neutral classes. The sentence level sentiment analysis is used to identify whether the sentence is subjective or objective and then only subjective sentences are determined to be positive, negative or neutral. The objective types of sentences are ignored, as there is no sentiment bearing words in objective type of sentences. The entity or aspect level of sentiment analysis is also called as fine grained sentiment analysis, where analysis is done by deep study of the text. The goal is to not only decide text subjectivity and polarity, but also, what the text reviewer liked or disliked about a particular aspect.

Negation problems are not solvable at document level or sentence level. Handling of negation is very important task in sentiment analysis. For example "I do not like the way he talks." "The battery is not superior, but I like the screen of

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

mobile phones." In first example, 'like' word has a positive meaning, but overall sentence has a negative meaning. In second example, the negation word has an impact on first half of the sentence, but in second part of sentence has positive sentiment towards the screen of mobile phones. This has to be handled at entity level. To take care of negation, we use bi-gram as a feature.

The two main approaches of sentiment analysis are machine learning based and lexicon based. Machine learning approach is used to classify the text whereas lexicon based approach is used for sentiment dictionary with opinion words and find the polarity by matching them with words.

In sentiment analysis, the main fields of research are sentiment classification, opinion summarization and feature based sentiment classification. Sentiment classification is used to classify the whole document according to the opinion around certain objects. In opinion summarization, the features of the product are mined on which the customers have expressed their opinions. Feature based sentiment classification considers the opinion on features of certain objects.

For instance, user-generated reviews of products not only assess the overall product but also reveal some sentiments on product specific aspects, such as performance, price etc.

II. LEVELS OF SENTIMENT ANALYSIS

There are three levels of sentiment analysis

- Document level
- Sentence level
- Phrase level

A. Document level sentiment analysis

In document level classification, single review is considered about the single topic. The information unit is a single document of opinionated text. In case of forum and blogs, document level analysis is not desirable when customer compare one product with another that has the same characteristics. Subjectivity/objectivity classification is very much important in this classification because document may not be relevant in expressing the opinion about an entity. Document level classification used supervised or unsupervised learning methods. Supervised learning algorithm such as naïve base, support vector machine can be used to train the system. The unsupervised learning can be done by extracting the opinion words inside a document.

B. Sentence level sentiment analysis

The polarity of each and every sentence is calculated by the sentence level classification. Subjective and objective sentences must be found out. The subjective sentences may contain opinion words which help to verify the sentiment about an entity whereas objective sentences are ignored, there is no sentiment bearing words.

C. Phrase level sentiment analysis

In phrase level classification, opinions words are find out that phrase level contain and the classification is done. It is much more identifying approach to opinion mining. In some cases, the correct opinion about an entity can be correctly taken out. But in some cases negation of words can occur locally. In this type of cases, the level of sentiment analysis is satisfied.

III. LITERATURE SURVEY

There is a large body of work concerning sentiment analysis of customer reviews. Most of these studies treat sentiment analysis as a classification problem and execute supervised learning methods where the positive and negative classes are determined by reviewer ratings. In [1] Mostafa Karamibekr focuses on the sentiment analysis of social issues. They initially conduct a statistical investigation on the differences between sentiment analysis of products and social issues and proposed a method for sentiment analysis of social issues. It takes out the opinions from each and every sentence and constructs correspondence opinion structures, and then determines their orientations regarding the social issue. This method has focused on verbs as the core element of opinion structures. They implemented and evaluated method as well as the BOW approach to classify the sentiment orientation of comments about a social issue. This method accomplishes the classification 10% more than the BOW approach. The average accuracy of this approach is 65%,

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

which is better than some previous work. In [2] N. D. Valakunde computes the document level sentiment analysis by calculating the aspect level sentiment score based upon underlying entity. When performing the aspect based document level sentiment analysis accuracy is high as correlate to sentiment analysis at direct document level. It offers the multiclass sentiment analysis providing fine-grained view of sentiments instead of giving only positive or negative class. It also shows that SVM has better accuracy over NB. In [3] Pavitra.R proposes a document level sentiment classification in co-occurrence with topic detection and topic sentiment analysis of bigrams at the same time. This model is based on the weakly supervised Joint Sentiment- Topic model. The usage of bigrams for classification instead of unigrams improves the efficiency of sentiment classification. In [4] Deepali Virmani proposed algorithm which collaborates aspect factor with sentiment value. Aspect factor helps evaluating the preciseness of the aspect. The higher the aspect factor, more specific is the review about the aspect. Sentiment value is calculated taking into account the polarity of the review. In [5] Omar Qawasmeh provides a dataset for aspect-based sentiment analysis of Arabic text. Moreover, fosters the domain of Arabic ABSA and provides a benchmark dataset. The human annotated Arabic dataset (HAAD) consists of 1513 books review sentences in Arabic which has been annotated by humans with aspect terms and their polarities. In [6] Ms.K.Mouthami uses a document level sentiment classification with bag of words in existing system to determine the opinion document whether positive or negative. To accurate the classification proposed a new sentiment fuzzy classification algorithm with parts of speech tag to improve classification accuracy on the benchmark dataset of Movies reviews. In [7] Siti Rohaidah Ahmad proposed a metaheuristic algorithm which can be used as feature selection in sentiment analysis. Study of type of feature selection which is based on natural language processing and modern method such as Rough set theory and genetic algorithm. This algorithm is used for selecting the optimum features from the customer reviews. In [8] Roliana Ibrahim proposed a novel learning model which is based on combination of semi-supervised self-training and uncertainty based active learning approaches to increase the performance in cross-lingual sentiment classification. To unlabelled sentiment documents were translated from the target language in source language by using self-training and active-learning and density is also measured from unlabelled data to avoid the selection of outlier examples. In [9] B.Baharudin proposed the rule based domain independent sentiment analysis method. It classifies the objective and subjective sentences from comments and reviews. The subjective sentence is take out from SentiWordNet to determine their polarity as positive, negative or neutral. This method achieves the accuracy of 83% at sentence level and 87% at feedback level. In [10] Abdullah

Dar deals with the study of different methodology which is used for opinion mining and sentiment analysis. This is the ontology based supervised learning technique deals with arrangement of words. It is used to calculate about the movie review, product review etc. Sentiment analysis is helpful in different fields such as expressing, identifying and calculating the sentiment. In [11] Soujanya Poria proposed a novel technique for multimodal sentiment analysis which contains the suitable features of visual data and text and uses both feature and decision level for fusing the features extracted from different modalities. This technology achieves the accuracy of 80%. In [12] Duyu Tang proposed a learning sentiment-specific word embedding. Existing studies of this algorithm only encode the word context not the sentiment text. Introducing of neural network to encode context and sentiment level information simultaneously in unified way into word embedding .The effectiveness of sentiment embedding are verified on three sentiment analysis tasks that are word-level sentiment analysis, sentence level sentiment classification and building sentiment lexicons.

IV. CONCLUSION

In sentiment analysis, it is very difficult to predict the positive or negative sentiment about any element such as topics, services and products .To overcome this problem this paper defines the various levels and techniques of sentiment classification. Researchers' uses popular approaches such as bag of words and feature based sentiment to deal with sentiment analysis about product such as movies, cars, electronics etc. Sentiment analysis task is very challenging as it is natural language processing task, which has no easy problems.

REFERENCES

[1] Mostafa Karamibekr and Ali A.Ghorbani,"sentiment analysis of social issues", International Conference on social informatics, pp.215-221, 2012.
[2] N. D. Valakunde and Dr. M. S. Patwardhan," Multi-Aspect and Multi-Class Based Document Sentiment Analysis of Educational Data Catering Accreditation Process", International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, pp.188-192, 2013.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

[3] Pavitra.R and PCD.Kalaivaani," Weakly supervised sentiment analysis using joint sentiment topic detection with bigrams", IEEE sponsored 2nd International Conference on electronics and communication system (ICECS '2015).
[4] Deepali Virmani, Shweta Taneja and Pooja Bhatia," Aspect Level Sentiment Analysis to Distil Scrupulous Opinionated Result", International Conference on Computing, Communication and Automation, pp. 59-65, 2015.
[5] Mohammad AL-Smadi, Omar Qawasmeh, Bashar Talafha and Muhannad Quwaider," Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis", 3rd International Conference on Future Internet of Things and Cloud, pp.726-730,2015.
[6] Ms.K.Mouthami and Ms.K.Nirmala Devi, "Sentiment Analysis and Classification Based On Textual Reviews", International Conference on Information Communication and Embedded Systems (ICICES), pp.271-276, 2013.
[7] Siti Rohaidah Ahmad, Azuraliza Abu Bakar and Mohd Ridzwan Yaakub," Metaheuristic Algorithms for Feature Selection in Sentiment Analysis", Science and Information Conference 2015 July 28-30, 2015 | London, UK..
[8] Mohammad Sadegh Hajmohammadi, Roliana Ibrahim, Ali Selamat and Hamido Fujita," Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples", Information Sciences 317, pp. 67–77,2015.
[9] A. Khan and B.Baharudin," Sentiment classification using sentence-level semantic orientation of opinion terms from blogs", IEEE National Postgraduate Conference, 2011.
[10] Abdullah Dar and Anurag Jain," Survey paper on Sentiment Analysis: In General Terms", International Journal of Emerging Research in Management & Technology ISSN: 2278-9359, Vol-3, Issue-11.
[11] SoujanyaPoria, ErikCambria, Newton Howard, Guang-BinHuang and AmirHussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content", Neurocomputing 174, pp. 50–59, 2016.
[12] Duyu Tang, Furu Wei, Bing Qin, Nan Yang, Ting Liu and Ming Zhou," Sentiment Embeddings with Applications to Sentiment Analysis", Transactions on knowledge and data engineering, 2015
[13] B. Liu," Sentiment analysis and subjectivity", Handbook of Natural Language Processing, 2010.
[14] V. P. H. Binali and W. Chen, "A state of the art opinion mining and its application domains", International Conference on Industrial Technology, pp.1–6, February 2009.