A Survey on Sentiment Analysis of (Product) Reviews

A. Nisha Jebaseeli
Assistant Professor
Dept. of Computer Science
BDU Constituent College Lalgudi, Trichy

E. Kirubakaran, PhD.
Additional General Manager
Outsourcing Department
Bharat Heavy Electricals Ltd. Trichy, India

ABSTRACT
With the help of wireless technology, the internet becomes a valuable place for online learning, exchanging ideas, reviews for a product or service. Reviews in the internet could be in millions for a product or services which make it difficult to track and understand customer opinions. Sentiment analysis is an emerging area of research to extract the subjective information in source materials by applying Natural Language processing, Computational Linguistics and text analytics and classify the polarity of the opinion stated. This paper provides an overall survey about sentiment analysis or opinion mining related to product reviews.

Keywords
Opinion Mining, Product Reviews, Sentiment Analysis.

1. INTRODUCTION

OPINION mining (often referred as Sentiment Analysis) refers to identification and classification of the viewpoint or opinion expressed in the text span; using information retrieval and computational linguistics. The opinion expressed on the topic is given significance rather than the topic itself [1]. Sentiment analysis or opinion mining extracts the subjective information from the source materials such as reviews using techniques such as natural language processing, and text analytics. Opinion plays essential part in our information-gathering behavior before taking a decision. Online review sites, and personal blogs facilitate gathering of sentiments of products or object using information technologies. The main objective of Opinion mining is to determine the polarity of comments (positive, negative or neutral) by extracting features and components of the object that have been commented on in each document [2, 3]. Studies related to opinion mining, on the implication of economic impact due to the reviews, issues about breach of privacy are given attention.

Generally, the opinion expressed in a review document could either be a direct opinion or comparative opinion. Direct sentiment expressions on some target objects such as products, events, topics, persons. E.g.: "The picture quality of this camera is great." Comparison opinion expresses the similarities or differences of more than one object usually stating an ordering or preference. E.g.: "car x is cheaper than car y." Different types of comparatives are Non equal Gradable (less than), Equative (same), Superlative (longest).

Opinion mining is carried at either document level or sentence level as follows:

- Sentence level opinion mining is performed by two tasks subjective or Objective.

   Objective: I bought an iPhone a few days ago.

   Subjective: It is such a nice phone.

- For subjective sentences or clauses, classify positive or negative.

   Positive: It is such a nice phone.
   Negative: The phone has poor reception.

In document level, a document (e.g., a review) is classified based on the overall sentiment expressed by opinion holder.

Classes: Positive or negative

Assumption: each document focuses on a single object and contains opinions from a single opinion holder.
E.g., thumbs-up or thumbs-down, star ratings (2 stars, 3 stars…)

Opinions can also be made based on features as shown in example. "I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop. …"

Each feature of the product is classified and overall sentiment is judged. This paper presents a survey on different methods of sentiment analysis available in literature related to product reviews.

2. LITERATURE SURVEY

Lina Zhou et al., [4] investigated movie review mining using machine learning and semantic orientation. Supervised classification and text classification techniques are used in the proposed machine learning approach to classify the movie review. A corpus is formed to represent the data in the documents and all the classifiers are trained using this corpus. Thus, the proposed technique is more efficient. Though, the machine learning approach uses supervised learning, the proposed semantic orientation approach uses "unsupervised learning" because it does not require prior training in order to mine the data. Experimental results showed that the supervised approach achieved 84.49% accuracy in three-fold cross validation and 66.27% accuracy on hold-out samples. The proposed semantic orientation approach achieved 77% accuracy of movie reviews. Thus, the study concludes that the supervised machine learning is more efficient but requires a considerable amount of time to train the model. On the other hand, the semantic orientation approach is slightly less accurate but is more efficient to use in real time applications. The results confirm that it is practicable to automatically mine opinions from unstructured data.

Bo Pang et al., [5] used machine learning techniques to investigate the effectiveness of classification of documents by overall sentiment. Experiments demonstrated that the machine

learning techniques are better than human produced baseline for sentiment analysis on movie review data. The experimental setup consists of movie-review corpus with randomly selected 700 positive sentiment and 700 negative sentiment reviews. Features based on unigrams and bigrams are used for classification. Learning methods Naïve Bayes, maximum entropy classification and support vector machines were employed. Inferences made by Pang et al., is that machine learning techniques are better than human baselines for sentiment classification. Whereas the accuracy achieved in sentiment classification is much lower when compared to topic based categorization.

Zhu et al., [6] proposed aspect based opinion polling from free form textual customers reviews. The aspect related terms used for aspect identification was learnt using a multi-aspect bootstrapping method. A proposed aspect-based segmentation model, segments the multi aspect sentence into single aspect units which was used for opinion polling. Using a opinion polling algorithm, they tested on real Chinese restaurant reviews achieving 75.5 percent accuracy in aspect-based opinion polling tasks. This method is easy to implement and are applicable to other domains like product or movie reviews.

Jeonghee Yi et al., [7] proposed a Sentiment Analyzer to extract opinions about a subject from online data documents. Sentiment analyzer uses natural language processing techniques. The Sentiment analyzer finds out all the references on the subject and sentiment polarity of each reference is determined. The sentiment analysis conducted by the researchers utilized the sentiment lexicon and sentiment pattern database for extraction and association purposes. Online product review articles for digital camera and music were analyzed using the system with good results.

Alekh Agarwal et al., [8] proposed a machine learning method incorporating linguistic knowledge gathered through synonymy graphs, for effective opinion classification. This approach shows the degree of influence among relationships of documents have on their sentiment analysis. This is brought about by the use of graph-cut technique and opinion words got through synonymy graphs of Wordnet. The proposed approach also improves the accuracy of predictions in classification task. Experiments using the system have given results with an accuracy of over 90%, with an added advantage of reduction in processing time, with minimal difference in final accuracies.

The proposed methodology from the authors resulted in the following conclusions:

1. Automated mining of linguistic information is possible, so demonstrated with the structure of links in Wordnet.
2. Generic method of using graph-cut technique for efficient opinion classification.

Ahmed Abbasi et al., [9] proposed novel sentiment analysis methods to classify web forum opinions in multiple languages. The proposed sentiment analysis method utilized the function of stylistic and syntactic features to evaluate the sentiment in English and Arabic content. The Entropy weighted Genetic Algorithm is incorporated to enhance the performance of the classifier and achieve the true assessment of the key features. Experiments were conducted using movie review data set and the results demonstrated that the proposed techniques are efficient.

Anidya et al., [10] ranked the product reviews based on customer-oriented and manufacturer ranking mechanism. The expected helpfulness of the review is used for the ranking and also ranking is based on the expected effect on sale. The proposed methods identify the reviews which have the most impact. For feature based products, he reviews that confirm the information contained in the product description are used, and reviews with subjective point of view are useful for experience goods. Econometric analysis with text mining techniques and with subjectivity analysis is used in the proposed method. Product prices and sales ranking publicly available on amazon.com were used to compile the data set. The product and sales data are the two sets of information collected for each product. Products such as audio and video players, digital cameras were used to form the data set. The empirical analysis is performed using the compiled data set.

Michael et al., [11] presented 'Pulse' a prototype system for mining topics and sentiment orientation from free text customer feedback. Blogs, newsgroups, feedback email from customers, and web sites that collect product reviews are all source of free text customer feedback. The proposed system is designed to handle the free form information of the customer feedbacks as the sources of information are less structured than traditional surveys. A clustering technique and machine learned sentiment classifiers were used in the proposed method. Sentiment and topic detections are performed at the sentence level not at the document level. The Pulse was evaluated using car reviews database, and the sample data contains 4, 06,818 customer car reviews written over a four year period. The data set contained almost 900,000 sentences in total. Sentiment analysis was performed using 3000 randomly selected sentences. Each sentence is classified as positive, negative and others. The other category contained both positive and negative sentiment and sentences with no complex sentiments. Training of the sentiment classifier was done using 2500 sentences and the remaining 500 sentences are reserved for test set. Results reflect the efficiency of the proposed system.

Miniqing Hu et al., [12] performed mining and summarization process to all the customer reviews of a product. The proposed process is carried out in three steps:

1. The product features commented by the customer in the review are mined. Natural language processing and Data mining techniques are used for mining.

2. The opinions in the review are identified and the opinions are classified as positive or negative. Set of adjectives words called opinion words are identified and semantic orientation of the opinion words is determined. WordNet can be used to identify the semantic orientation and the opinion orientation of each sentence is decided.

3. Summarize the results.

The objective of the study is to perform feature based summary of a large number of customer reviews of a product sold online.

Qui et al., [13] analyzed the problems related to opinion mining such as opinion lexicon expansion and opinion target extraction. Opinion targets are entities and their attributes on which opinions have been expressed. The list of opinion words such as good, bad, excellent, poor used to indicate positive and negative sentiments is Opinion lexicon. The links between the opinion words and targets Syntactic relations are identified using dependency parser based on bootstrapping. The process uses semi-supervised methods, opinion word seeds are used in the initial opinion lexicon. Bootstrapping

process is started using the initial opinion lexicon. Double propagation method is used as information is propagated back and forth between opinion words and targets.

Lei Zhang et al., [14] identified domain dependent opinion words. Noun and noun phrases that indicate the product feature which implies opinions are found using a feature based opinion mining model. Two steps are used to identify the noun product feature which means the positive or negative opinion. Sentiment context of each noun feature is determined in the Candidate identification step. And also a list of candidate features with positive opinions and list of candidate features negative opinions is produced. Noun product feature is directly modified into positive and negative opinion words in pruning step. Opinion lexicon complied by Ding et al. (2008) was used to identify the opinion polarity on each product feature in a sentence. For a sentence s which contains a product feature f, opinion words in the sentence are first identified by matching with the words in the opinion lexicon. An orientation score for f is computed and the semantic orientation of the positive word is assigned the score of +1, and a negative word is assigned the score of -1. On summing up of all the scores, if the final score is positive, then the opinion on the feature in s is positive. If the score is negative, then the opinion on the feature in s is negative.

Xiaowen Ding et al., [15] proposed a holistic lexicon-based approach which uses external indications and linguistic conventions of natural language expressions to determine the semantic orientations of opinions. Advantage of this approach is that opinion words which are context dependent are easily handled. The algorithm used uses linguistic patterns to deal with special words, phrases. Researchers built a system called Opinion Observer based on this technique. Experiments using product review dataset was highly effective. It was shown that multiple conflicting opinion words in sentence are also dealt with efficiently. This system shows better performance when compared to existing methods.

3. DISCUSSION

Opinion mining plays vital role in Businesses and organizations.

- ❖ Product and service bench marking.
- ❖ Market intelligence.

People get the other's opinion to make some decision about product or services.

- ❖ Finding opinions while purchasing a new product.
- ❖ Finding opinions on political topics.
- ❖ In Advertisement (ads) opinion mining helps to display the product based on the stake holders view.
- ❖ Placing ads in the user-generated content.
- ❖ Place ads when one praises a product.
- ❖ Place ads from a competitor if one criticizes a product.

Finally opinion mining can be served in the field of Information search & Retrieval. In opinion mining Determining sentiments seems to be easier, determining objects and their corresponding features is harder. Combining both the task is very tedious and also accuracy is the problem.

4. CONCLUSION

In this literature survey paper it is seen that sentiment analysis/opinion mining play vital role to make decision about

product /services. Opinion mining not only encompasses concepts of text mining but also the concepts of information retrieval. Major challenges in opinion mining includes feature weighting which plays a crucial role for good classification. Also it is seen that soft computing techniques have not been extensively used in the literature. Without opinion life is like an empty vessel. The work can be further extended to emerging areas like Mobile learning and investigation with soft computing techniques like neural network.

5. REFERENCES

[1] Bing Liu,"Exploring User Opinions in Recommender Systems", Proceeding of the second KDD workshop on Large Scale Recommender Systems and the Netflix Prize Competition", Aug 24, 2008, Las Vegas, Nevada, USA.

[2] Dave.D, Lawrence.A, Pennock.D,"Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews", Proceedings of International World Wide Web Conference (WWW'03), 2003.

[3] Turney, P, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", ACL'02, 2002.

[4] Lina Zhou, Pimwadee Chaovalit, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches", Proceedings of the 38th Hawaii International Conference on system sciences, 2005.

[5] 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.

[6] Zhu, Jingbo Wang, Huizhen Zhu, Muhua Tsou, Benjamin K. Ma, Matthew, "Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on Affective Computing, Volume: 2,Issue:1 On page(s): 37. Jan-June 2011.

[7] Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack: 2003, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques", In: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003). Melbourne, Florida.

[8] 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.

[9] Ahmed Abbasi, Hsinchun Chen, And Arab Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums", ACM Trans. Inf. Syst., Vol. 26, No. 3. (June 2008), pp. 1-34.

[10] Anindya Ghose, Panagiotis G. Ipeirotis, "Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews", Proceedings of the Ninth International conference on Electronic commerce ICEC07 (2007), pp: 303-310.

[11] Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger, "Pulse: Mining Customer Opinions from Free Text Natural Language Processing", Microsoft Research, Redmond, WA 98052, USA.

[12] Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews", Proceedings of the tenth ACM SIGKDD International conference on knowledge discovery in data mining (KDD-2004), August 22-25.

[13] Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen. "Opinion Word Expansion and Target Extraction through Double Propagation." Computational Linguistics, March 2011, Vol. 37, No. 1: 9.27.

[14] Lei Zhang and Bing Liu. "Identifying Noun Product Features that Imply Opinions." *ACL-2011* (short paper), Portland, Oregon, USA, June 19-24, 2011.

[15] Xiaowen Ding. A Holistic Lexicon-Based Approach to Opinion Mining- *WSDM'08*, February 11-12, 2008, Palo Alto, California, USA.2008 ACM 978-1-59593-927-9/08/0002.