Various Techniques for Sentiment Analysis of Products Review – A Review

Anjani Chanji
Punjabi University Regional Centre for
Information Technology and Management,
Mohali, Punjab, India

Rekha Bhatia
Punjabi University Regional Centre for
Information Technology and Management,
Mohali, Punjab, India

ABSTRACT

The sentiment analysis is the analysis of the sentiment in the numeric form from the given text input. The text data obtained from the social networks primarily undergoes the emotion mining method to analyze the sentiment of the user messages posted online. The best known sentiment analysis approaches are supervised approaches and utilizes the dictionary with pre-enlisted sentiment weights to evaluate the overall weight of the sentence sentiment. The pre-enlisted sentiment weights are computed in the linguistic dictionaries such as WordNet by Princeton University, SenticNet by MIT andHindi SentiWordnet by IIT Bombay. The appropriate dictionaries areutilized to evaluate the sentiment of the input text data for the social threads, product review, opinion mining, etc.This work focuses upon the detailed literature study over the sentiment analysis and concludes the shortcomings of the existing models to improve the design of the sentiment analysis models

Keywords
Sentiment analysis, Product review classification, Porter stemming, Quick response, Opinion mining.

1. INTRODUCTION
With the flourishing growth of social media everyone expresses their views on the social media. So keeping on an eye on social media is vital to get the real scenario among the general public; how the general public is responding to the event or the brand.

1.1 Sentiment Analysis
Sentiment analysis is a type of opinion mining or it can said that the both terms are interchangeable; it determines the point of view or perspective of the writer. It analyses positive,negative, neutral or no sentiment texts. In decision making process opinion are the key influencers; while making decisions wouldprefer to know what the other people opinions are. Sentimentanalysis determines the mood of people about various entity and their attributes.With the rise of social mediadecision making get influenced with the content over the social media. For example before voting a candidate during any election everyone reads about other people opinion before making their decision. During election various surveys, opinion polls are conducted to know the estimated opinion of the public which is beneficial to the political campaign companies.

1.2 Product Reviews
Product reviews are mostly used on ecommerce websites and are considered as sales drivers. It gives an opportunity for the customers to review/comment or rate the product. Other people can take benefit from it; by reading the comments they can came to know about pros and cons of the product that would help them to make up their decision in purchase of the product. Even organization can take benefit from reviews;they came to know about what the general public opinions are about their product and the services. The organization comes to know more about the pros and cons of their product. Let's take an example of a mobile phone of company XYZ model P. A customer buys the same model of company XYZ and found that the sound quality for the phone is excellent while the camera quality of the phone is low as per the expectation. So a person making up mind to buy the same model reads review and if the person gets satisfied after reading the comments of the customer he will stick to its decision otherwise he will drop the idea of buying the phone. So products reviews are helpful in making up decision in purchase of the product.

1.3 Review Categorization
Review categorization is a process of categorization of review in various classes. Each class determines the property of the product. For example A mobile phone will have following classes:-Camera, RAM, Battery, Screen, Sound, Storage etc. Reviews will be categorized into various classes and then sentiment analysis will be done on the properties of the product that is each review class will be classified into positive negative or no sentiment texts. Review categorization describes the sentiments over various classes of the product. Review categorization ease the public to make the decision as it is easily understandable. Public can distinguish between sentiments over different classes.

1.4 Review Summarization
Review Summarization is a process of summarizing the reviews or opinions of the product. In decision making process only one review or opinion is not sufficient; various reviews for en efficient decision making process are needed. So review summarization is used solution for that problem. Analyzing various reviews to take and finalize decision.So the need ofthe summary to consolidate the reviews and brief about the opinion of the reviews. Review Summarization can be indicative, informative or critical. Generally thesummarization of reviews is based on feature also called as aspect based summarization.

Example:-Feature/Aspect based opinion summary→

Phone XYZ:

Aspect	Positive	Negative	Summarized Review Sentences
Sound Quality	202	4	\<Sound Quality Summarized Review>
Picture Quality	45	1	\<Picture Quality

			Summarized Review>
Battery Life	150	19	<Battery Life Summarized Review>
Camera	111	23	<Camera Summarized Review>

1.5 Advantages of product review

- Public can know the real quality of the product; the real experience of using the product is wordish by the customers.

- Public can compare various products on the basis of quality that the customers had reviewed.

- Companies can improve the product by working on problems of the product that the customers are facing written in the reviews.

In the introduction (section 1) the overview of the sentiment analysis methods has been discussed in detail. The introduction section describes the various aspects of the sentiment analytical models. Section 2 posts the study of the various existing systems based on the sentiment analysis in the form of literature survey.Section 3 describes the gaps in the sentiment analysis model.The final section (Section 4) concludes the overview about the sentiment analysis method and highlights the important aspects from the existing study and future projections.

2. LITERATUTE SURVEY

Che, Wanxianget al. [1]have developed sentence compression model for aspect based sentiment analysis. In this model sentence is compressed before the sentiment analysis process. Sentence compression 'SENTCOMP' model removes the unnecessary words for sentiment analysis process and therefore compacting sentence that is easier to parse and high performance for sentiment analysis. 90% accuracy is achieved using this model. 'Sent comp' model removes the repeated, unwanted words but it preserves polarity based information that is necessary for sentiment analysis.

Liu, Siyuanet. Al. [2]have developed SVM model 'TASC' that is topic adaptive sentiment classification. The author has developed a classifier that works on dynamic tweets. The classifier works for general features and blended labeled data for diversified topics. It increases the performance by selecting more reliable tweets after selecting unlabeled data collaboratively. It fits the unlabeled data and the features without any hinge loss for all sentiment words and connection of sentiments acquired from '@'tweets and it is designated topic adaptive features. They had even designed an TASC-t that is a timeline based model on dynamic tweets that has achieved notable accuracy and F-score. Finallyan visualizing and color gradation based graph that shows sentiment trends.

Andrea Esuli and FabrizioSebastiani [3]have developedSentiWordNetthat is an lexical resource for mining opinion.WordNetsynset have scores 'obj(s)','Neg(s)','Pos(s)' having some numerical values that outlines how negative,positive,objective is accommodated in synset. It comes with GUI. Eight Ternary classifiers are trained to get three scores for synset. Training set and learning device for each ternary classifier is distinct and hence distinct classification outcome of the like synset is produced. Scores for opinions are calculated by normalization on the basis of assigned labels.If every ternary classifier assigns same label to synset than synset will have maximal score otherwise the score will be proportional to ternary classifiers.

Soo-Min Kim and Eduard Hovy [4]have developed a model that automatically identifies pros and cons reasons in the review that is the reasons behind the liking or disliking of the product. As reviewed opinion have some reasons behind recommendation or non-recommendation of the product. Sodeveloped a technique where it automatically labels pros and cons using clue phrases in online reviews. Reason identification and reason classification were the two key areas they have worked upon. They have achieved 71% F-score in reason identification whereas 61% F-Score in reason classification.

Minqing Hu and Bing Liu [5]have developed a model that summarizes opinion based on features. Authors worked on the features that the customers have commented on. Firstly what are the features of the product that the customers have reviewed on is identified than scores are assigned on the basis of the number of times recurrence in the review. For every feature commented positive and negative opinions are identified. Than opinion summarization is performed by summarizing the opinions for extracted features in two classes positive and negative.

Jiaming Zhan a, Han Tong Loh a, Ying Liu [6]have developed approach for summarization of text by gathering concerns of customers from reviews of online product. It isbased on topical structure. Firstly the pre-processing step requires the removal of stop words and word stemming is done to bring down noisy information. Main idea here is to determine the important topics in the review and develop subjective topical design. The topic identification is done by frequent words recurring and correlated classes. Previous approaches were based on ranking of sentences whereas here ranking of topics are done. The summary is created based on the ranked topics. This algorithm has capability to explore the most important topic.

LaddaSuanmali, NaomieSalim and Mohammed Salem Binwahlan [7]have developed a method based on fuzzy logic that improves the text summarization. This method works on text extraction approach. The main step includes sentence selection. The Sentence is weighted and then best ones out of the sentences are selected.The stop words are removed and word stemming is applied. Important features are selected and Scores are calculated for each sentence.The top leading scores sentences set are extracted as summary for the document.

Seyed Hamid Ghorashi, Roliana Ibrahim, ShirinNoekhah and NiloufarSalehiDastjerdi [8]have developed an algorithm for frequent pattern mining for extraction of reviews of customer. Here some group of reviewed are mined to extract the product features.The algorithm they have used is H-mine. They have only worked on the frequent recurring features in the review. By using this algorithm they are easily able to deal with two major issues that are taking numerous scans of large database to generate frequent features of items and to another one is while the new item sets are generated the transposition of words are also considered. The main steps include pre-processing for removal of stop words and stemming of words. POS tagging for identification of frequent features. Mining of frequent patterns gives potential features. Frequentfeatures are extracted using pruning. Finally summary with potential features is generated

AUTHORS	PAPER TITLE	PROPOSED WORK	MERITS	DEMERITS
Che, Wanxiang et al. [1]	Sentence Compression for Aspect-Based Sentiment Analysis.	Sentence compression (SentComp) is performed to eliminate needless word prior to aspect based sentiment analysis.	The accuracy of nearly 90% has been achieved.	Performance can be improved as the aspect-based model needs critical requirements.
Liu, Siyuan, Xiaoyin Cheng, and Fan Li[2]	TASC: Topic-Adaptive Sentiment Classification on Dynamic Tweets.	A semi supervised TASC model,a classifier that works on dynamic tweets. The classifier works for general features and blended labeled data for diversified topics.	It adjusts to untagged data and features. Visualizing and color gradation based graph that shows sentiment trends	Needs to improvework on the performance on more source of the social data.
Esuli, Andrea, and Fabrizio Sebastiani.[3]	Sentiwordnet: A publicly available lexical resource for opinion mining.	It is a lexical resource in which every WordNetsynset have scores 'obj(s)','Neg(s)','Pos(s)' having some numerical values that outlines how negative,positive,objective is accommodated in synset.	Robust data classification. Utilizes the large word database for the sentiment analysis.	Require word compression method to compress the dictionary size to decrease the elapsed time.
Kim, Soo-Min, and Eduard Hovy. [4]	Automatic identification of pro and con reasons in online reviews.	Automatically identifies pros and cons reasons in the review that is the reasons behind the liking or disliking of the product Reason identification and reason classification were the two key areas they have worked upon.	Achieved 71% F-score in reason identification whereas 61% F-Score in reason classification.	Based on observation in online reviews, most reviews have both pros and cons even if sometimes one of them dominates.
Hu, Minqing, and Bing Liu. [5]	Mining and summarizing customer reviews.	A model that summarizes opinion based on features. Opinion summarization is performed by summarizing the opinions for extracted features in two classes positive and negative.	Text segmentation is used to segment the text based on similarity of adjacent passages and detect the boundary of topics.	Feature identification is based upon the frequent word sequence which carries the probability of the high polarization and can cause false classification results.
Suanmali, Ladda, Naomie Salim, and Mohammed Salem Binwahlan.[7]	Fuzzy logic based method for improving text summarization.	Works on text extraction approach. The main step includes sentence selection.	Uses the good tokenization process. The feature extraction technique is producing better results.	Not suitable for categorization. Uses the small dictionary.

This section has described the literature survey over the sentiment analysis methods. The next section 3 describes the gaps

3. RESEARCH GAPS

- The existing model requires high computational power and slower the process of sentiment analysis. The model can be extended to increase the process execution speed of the process.

- The compression method used in the existing model can lower the performance of the sentiment analysis by removing the necessary bias and affecting the total emotion of the text data. The real-time aspect and position and grammar-based dictionary database can be used for the purpose of the highly accurate and human-like compression to understand the real meaning of the sentence in order to improve the accuracy of the sentiment analysis.

- The existing model can be made bilingual, which must include the English and Hindi, which are mostly used languages in the Indian region for the social network discussion.

- The existing model offers the accuracy of the nearly 90%, which can be improved up to the higher level. The accuracy of the system can be improved by using the various improvements in the existing model.

This section has described the gaps in sentiment analysis methods. The next section describes the conclusion.

4. CONCLUSION

The model can been designed with the stemming porter based dictionary reduction with N-gram analysis for the keyword feature extraction. This model can be entirely based upon product review classification, which has been divided into the two major layers. The first layer of the model can be designed to the auto-categorize the reviews according to the product features, whereas the second module calculates the sentiment score for the categorized message data. The model will be described in the form of bilingual sentiment analytical models. The proposed model will be capable of work on the reviews written inHindi written in English (Hinglish) and English.The overall system will be utilized as the complete automatic product review classification and analytical engine to facilitate the online purchases over the shopping portals where the personal touch and feel is never available. The model can be even aimed at solving the problem with higher accuracy and precision.

5. REFERENCES

[1] Che, Wanxiang, Yanyan Zhao, Honglei Guo, Zhong Su, and Ting Liu. "Sentence Compression for Aspect-Based Sentiment Analysis." Audio, Speech, and Language Processing, IEEE/ACM Transactions on 23, no. 12 (2015): 2111-2124.

[2] Liu, Siyuan, Xiaoyin Cheng, and Fan Li. "TASC: Topic-Adaptive Sentiment Classification on Dynamic Tweets." (2015).

[3] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." In Proceedings of LREC, vol. 6, pp. 417-422. 2006.

[4] Kim, Soo-Min, and Eduard Hovy. "Automatic identification of pro and con reasons in online reviews." In Proceedings of the COLING/ACL on Main conference poster sessions, pp. 483-490. Association for Computational Linguistics, 2006.

[5] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." InProceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177. ACM, 2004.

[6] Jin, Jian, Ping Ji, and Ying Liu. "Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach." Engineering Applications of Artificial Intelligence 41 (2015): 115-127

[7] Suanmali, Ladda, Naomie Salim, and Mohammed Salem Binwahlan. "Fuzzy logic based method for improving text summarization." arXiv preprint arXiv:0906.4690 (2009).

[8] Ghorashi, Seyed Hamid, Roliana Ibrahim, Shirin Noekhah, and Niloufar Salehi Dastjerdi. "A frequent pattern mining algorithm for feature extraction of customer reviews." In IJCSI International Journal of Computer Science Issues. 2012.

[9] Yazhini, R., and Raja P. Vishnu. "Automatic summarizer for mobile devices using sentence ranking measure." In Recent Trends in Information Technology (ICRTIT), 2014 International Conference on, pp. 1-6. IEEE, 2014.

[10] Zha, Zheng-Jun, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua. "Product aspect ranking and its applications." Knowledge and Data Engineering, IEEE Transactions on 26, no. 5 (2014): 1211-1224.

[11] Jin, Jian, Ping Ji, and Ying Liu. "Prioritising engineering characteristics based on customer online reviews for quality function deployment." Journal of Engineering Design 25, no. 7-9 (2014): 303-324.

[12] Kherwa, Pooja, Anish Sachdeva, Dhruv Mahajan, Nishtha Pande, and Praveen Kumar Singh. "An approach towards comprehensive sentimental data analysis and opinion mining." In Advance Computing Conference (IACC), 2014 IEEE International, pp. 606-612. IEEE, 2014.