Opinion Mining and Sentiment Analysis - Challenges and Applications

Haseena Rahmath P

Dept. of Computer Science and Engineering, Al-Falah School of Engineering, Dhauj, Haryana, India

Abstract
People's opinions and experience are very valuable information in decision making process. Now a days several websites encourage users to express and exchange their views, suggestions and opinions related to product, services, polices, etc. publically. The increased popularity of these sites resulted huge collection of people opinion on the web in much unstructured manner. Extracting the useful content from these opinion sources becomes a challenging task. This situation created a new area of research called opinion mining and sentiment analysis. Opinion mining and sentiment analysis extract and classify the people's opinion automatically from the internet. This paper discusses various application and challenges related to the Opinion Mining and Sentiment Analysis.

Keywords – Opinion Mining, Sentiment Analysis, Applications, Challenges.

1. INTRODUCTION
The social media, blogs, forums, e-commerce web sites, etc. encourages citizens to share their opinion, emotions and feelings publically. Today these internet sites are very popular and this resulted a huge accumulation of such opinions. People's opinions and experience are very valuable information in decision making process, but to get benefits from these opinion and experience, the accumulated content should be extracted and analyzed properly. This extracted and analyzed sentiments or opinion are useful for consumer as well as manufactures as consumers can get a chance to evaluate others opinion and experience related to some product or services before purchasing them. Similarly the manufactures can take these opinions as feedback from the consumers and thereby they can improve the quality of their product or services.

It also very useful for decision makers or policy makers of the country since it is impossible for the policy makers to interact to the mass and collect their opinions effectively. It is true that via these media citizens can express their desires, problems, emotions and feelings and the experts can make use of it by properly extracting and analyzing it. But the extraction and analysis of huge unstructured internet content is beyond the human power and time. The content is mostly written in natural language. This situation necessitate an automatic natural language processing tool that extract and analyze the people sentiments from this unstructured texts. Numerous researches are undergoing in this direction. This research domain is called Opinion mining and sentiment analysis.

In this context, this paper discusses major application areas of sentiment analysis and the real challenges that are being faced by this technique. The paper concludes by discussing various areas where further researches are required.

2. OPINION MINING AND SENTIMENT ANALYSIS
It is an extension of data mining which utilizes natural language processing techniques to extract people's opinion from World Wide Web. The recent trend in internet that encourages users to contribute their opinion and suggestion created a huge collection of valuable information in the web. The Opinion mining system analyze each text and see which part contain opinionated word, which is being opinionated and who has written the opinion.

Sentiment analysis analyzes each opinionated word or phrase and determines its sentiment polarity orientation, whether it is positive or negative or neutral. It gives the summarized opinion of a writer or speaker. Sentient analysis can be done at word level, sentence level and document level.

3. APPLICATIONS AREAS OF OPINION MINING AND SENTIMENT ANALYSIS
Since the Opinion based or feedback based application are more fashionable, now a days, the natural language processing community shows much interest in Sentiment Analysis and Opinion Mining system**.** The explosion of internet has changed the people's life style, now they are more expressive on their views and opinions [1], and this tendency helped the researchers in getting user-generated content easily.

The major applications of Opinion mining and sentiment analysis are the following:

1) **Purchasing Product or Service:** While purchasing a product or service, taking right decision is no longer a difficult task. By this technique, people can easily evaluate other's opinion and experience about any product or service and also he can easily compare the competing brands. Now people don't want to rely on external

consultant. The Opinion mining and sentiment analysis extract people opinion form the huge collection of unstructured content, the internet, and analyze it and then present to them in highly structured and understantable manner.

2) **Quality Improvement in Product or service:** By Opinion mining and sentiment analysis the manufactures can collect the critic's opinion as well as the favorable opinion about their product or service and thereby they can improve the quality of their product or service. They can make use of online product reviews from websites such as Amazon and C|Net [2, 3], RottenTomatoes.com [4] and IMDb [5].

3) **Marketing research:** The result of sentiment analysis techniques can be utilized in marketing research [6]. By sentiment analysis techniques, the recent trend of consumers about some product or services can be analyzed. Similarly the recent attitude of general public towards some new government policy can also be easily analyzed. These all result can be contributed to collective intelligent research [7].

4) **Recommendation Systems:** By classifying the people's opinion into positive and negative, the system can say which one should get recommended and which one should not get recommended[8].

5) **Detection of "flame" :** The monitoring of newsgroup and forums, blogs and social media is easily possible by sentiment analysis. Opinion mining and sentiment analysis can automatically detect arrogant words [6], over heated words or hatred language used in emails or forum entries or tweets on various internet sources.

6) **Opinion spam detection:** Since internet is available to all, anyone can put anything on internet, this increased the possibility of spam content on the web. People may write spam content to mislead the people. Opinion mining and sentiment analysis can classify the internet content into' spam' content and 'not spam' content [1].

7) **Policy Making:** Through Sentiment analysis, policy makers can take citizen's point of view towards some policy and they can utilize this information in creating new citizen friendly policy.

8) **Decision Making:** People's opinion and experience are very useful element in decision making process. Opinion mining and Sentiment analysis gives analyzed people's opinion that can be effectively used for decision making.

4. RESEARCH CHALLENGES IN OPINION MINING AND SENTIMENT ANALYSIS
The main challenges that are faced by Opinion mining and sentiment analysis are the following:

1) **Detection of spam and fake reviews:** The web contains both authentic and spam contents. For effective Sentiment classification, this spam content should be eliminated before processing. This can be done by identifying duplicates, by detecting outliers and by considering reputation of reviewer [1].

2) **Limitation of classification filtering:** There is a limitation in classification filtering while determining most popular thought or concept. For better sentiment classification result this limitation should be reduced. The risk of filter bubble [11] gives irrelevant opinion sets and it results false summarization of sentiment.

3) **Asymmetry in availability of opinion mining software:** The opinion mining software is very expensive and currently affordable only to big organizations and government. It is beyond the common citizen's expectation. This should be available to all people, so that everyone gets benefit from it.

4) **Incorporation of opinion with implicit and behavior data:** For successful analysis of sentiment, the opinion words should integrate with implicit data. The implicit data determine the actual behavior of sentiment words.

5) **Domain-independence:** The biggest challenge faced by opinion mining and sentiment analysis is the domain dependent nature of sentiment words. One features set may give very good performance in one domain, at the same time it perform very poor in some other domain.

6) **Natural language processing overheads:** The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. created hindrance in sentiment analysis too.

5. RESEARCH SCOPE IN OPINION MINING AND SENTIMENT ANALYSIS
The major research scope areas in sentiment analysis are:
1) Spam Detection Sentiment Analysis;
2) Sentiment Analysis on short Sentence like abbreviations;
3) Improving sentiment word identification algorithm;
4) Developing fully automatic analyzing tool;
5) Effective Analysis of policy opinionated content;
6) Successful handling of bi polar sentiments;
7) Generation of highly content lexicon database.

6. CONCLUSION
Thus, Opinion Mining and Sentiment analysis has wide area of applications and it also facing many research challenges. Since the fast growth of internet and internet related applications, the Opinion Mining and Sentiment Analysis become a most interesting research area among natural language processing community. A more innovative and effective

techniques required to be invented which should overcome the current challenges faced by Opinion Mining and Sentiment Analysis.

References
[1] Liu, B. (2010), "Sentiment Analysis and Subjectivity". Appeared in Handbook of Natural LanguageProcessing, Indurkhya, N. & Damerau, F.J. [Eds.].
[2] Dave K., Lawrence, S. & Pennock, D.M. (2003), "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews". In Proceedings of the 12th International Conference on World Wide Web, p. 519-528.
[3] Hu, M. & Lui, B. (2004), "Mining and Summarizing Customer Reviews". In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004), p. 168–177.
[4] Pang, B. & Lee, L. (2004), "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts". In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, p. 271-278.
[5] Pang, B., Lee, L. & Vaithyanathan, S. (2002), "Thumbs Up? Sentiment Classification Using Machine Learning Techniques". In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2002), p. 79-86.
[6] Boiy, E., Hens, P., Deschacht, K. & Moens, M.-F. (2007), "Automatic Sentiment Analysis in OnLine Text". In Proceedings of the Conference on Electronic Publishing (ELPUB-2007), p. 349-360.
[7] Segaran, T. (2007), Programming Collective Intelligence. Sebastopol: O'Reilly Media, Inc.
[8] Pang, B. & Lee, L. (2008), "Opinion Mining and Sentiment Analysis". In Foundations and Trends in Information Retrieval 2 (1-2), p. 1–135.
[9] Gamon, M. (2004), "Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis". In Proceedings of the International Conference on Computational Linguistics (COLING 2004), p. 841-847.