

A Survey On OMCHAR: Opinion-mining Methods and their CHARacteristics

Prof. Sandipkumar C. Sagare¹, Prof. Vinod. G. Khetade², Prof. Mrs . Smita. S. Darbastwar³

D.K.T.E. Society's Textile and Engineering Institute, Ichalkaranji^{1,2,3}.

sagaresandip999@gmail.com¹, vgkhetade10@gmail.com², smitasangewar@gmail.com³

Abstract

Natural language processing has an important type called opinion mining that can be used to track the opinion of the public about a particular thing. Opinion mining is widely applicable to context of customer voice such as online surveys, reviews of products and social media. This paper focuses on the different methods and their characteristics for the classification of a given piece of natural language textual data based on the opinions described in it.

Keywords — *Sentiment Analysis, Opinion.*

1. INTRODUCTION

What others think has always been an important piece of information. For example, “Which smartphone should I buy?”, “Which college should I apply to?”, “Whom should I vote for?”, “Which Professor to work for?”. These examples indicate towards a natural technique of human being to take opinions of others while taking important decisions. Before introduction of WWW, we were taking these opinions manually by asking friends, relatives, experts and consumer reports.

Web 2.0 has become advanced so much that many individuals use it to express their opinion and feedback as comments, reviews or question answers on forums, blogs and social websites. This increases the amount of user generated content on the internet. For both user and an organization, this user generated content can be very useful. For example, buyers on online shopping website can check experiences and reviews written by other buyers on that website before purchasing any product. For the online shopping website, these reviews and feedbacks available on the website could be used to make the decisions or make focus groups, surveys in market research. Since very large number of opinions is published on the web, it is difficult for users to analyze all web opinions. To get valuable information from these reviews which are in the form of plain text written in any natural language, we need the help from other domains such as Data Mining and Natural Language Processing (NLP). To summarize and analyze the opinions expressed on websites manually is a difficult task. For that, we require automated sentiment analysis systems [1].

World Wide Web is becoming very popular day by day. So, the task of taking decisions can make use of WWW and analyze the opinions of users. These opinions are helpful for the stakeholders and other public in the decision making. Opinion mining is a technique for retrieving the information through Web blogs, search engines, and social networking sites. Since there is huge number of reviews and opinions in the form of unstructured text, it is not possible to manually summarize this information. So, computational methods that can be efficient are needed for extracting and summarizing the reviews from the web documents [2].

2. OMCHAR: Opinion-mining Methods and their CHARacteristics

Huge amount of research has been done on Opinion mining since an evolution in the field of NLP. Research in the opinion mining has been reviewed by K. Khan et al. [2]. Also, Tsytarau et al. [3] classifies opinion mining methods in four different approaches: Machine Learning, Dictionary, Statistical, and Semantic.

The Machine Learning approach [3] had the characteristic that it included training dataset (annotated or not) with sentiment labels. Then each review is represented as a features vector. After that a classifier is trained to distinguish between sentiment labels by analysis of the relevant features. Finally, sentiments are identified by the trained classifier for new reviews.

The Dictionary Approach [3] had the characteristic that it used a dictionary which was pre-built and defined semantic orientation of words, for example, the SentiWordNet. The task to identify semantic orientation can be performed by using dictionaries. Calculation of Semantic orientation of a single sentence or review is generally performed by making average of the semantic orientation values of individual words.

The Statistical Approach [3] had the characteristic that it is the one which obtains values of polarity through the co-occurrence of adjectives in a corpus. Here, adaptability can be achieved by creating corpus specific dictionary. This approach uses a very large corpus for solving the problem of unavailability of some of the words. Its also possible to make use of the complete set of indexed documents on the Web as the corpus for the construction of dictionary [4].

The Semantic Approach had the characteristic that it provides values of sentiments directly by calculating the similarity between words. For that, it uses various principles. The most basic of the principles of this approach is that the words which are related semantically should have the similar sentiment values [3].

Chetashri Bhadane, Hardi Dalal, Heenal Doshi [5] explained two areas of research in sentiment classification i.e. lexical and machine learning approaches. polarity values of lexicons are stored in a dictionary in case of lexical approach. If polarity score is present in the dictionary of each word of the text, it is added to get an 'overall polarity score'. A large amount of work is needed to discover which lexical information is most efficient because the polarity of the text mostly depends on the score given to each lexicon. Chetashri Bhadane, Hardi Dalal, Heenal Doshi [5] implemented a set of techniques for classification of aspects and identification of polarity of product review using combination of machine learning (SVM) and domain specific lexicons. They showed that the proposed techniques achieved about 78% accuracy and were very promising in performing their tasks.

According to Bei Yu, Stefan Kaufmann, Daniel Diermeier, there are two mainly important approaches for classification of sentiments. The knowledge based and supervised machine learning [7]. Knowledge-based approach has the characteristic that predefined affect dictionaries of opinion words are used to search the input words and find its effects. While in supervised machine learning has the characteristic that a trained statistical classifier is used for sentiment classification. Sentiment orientation of input documents is predicted by the trained classifier. These two approaches rely heavily on affective vocabulary although its use is different [8, 9, 10, and 11].

3. CHALLENGES AND FUTURE SCOPE

There are several different challenges through which opinion mining needs to suffer from. It includes determination of which segment of text is to be opinionated, identification of the opinion holder, determination of the positive or negative strength of an opinion [21]. Opinion mining is based on the human reviews, sentimental discussion and emotions. Every person has their own view and concern about a given topic, problem or issue. A text that is Opinionated may be fake or not relevant and or ambiguous information. Opinions are very much hard to describe than facts. Sources of opinion are not formally written and can be highly diverse.

The Authority, Credibility, Spam, Non Expert opinion, Domain Dependent, Language differences, Effects of syntax on semantics are the general challenges that are pointed out so far by the different authors [12,13,14,15,16,17,18,19 and 20]. Future work needs to be done in the direction of these challenges to optimize the opinion mining methods..

4. CONCLUSION

This paper has presented brief survey of opinion mining methods and their characteristics. The future study will focus on efficiently analyzing the opinions using opinion mining methods and to provide the decision making capability to the stakeholders of the information present on the web in the form of reviews, surveys, feedbacks etc.

REFERENCES

- [1] Solanki Yogesh Ganeshbhai, Bhumika K. Shah "Feature Based Opinion Mining: A Survey", 2015 IEEE International Advance Computing Conference (IACC), 919-923.
- [2] Khairullah Khan *, Baharum Baharudin , Aurnagzeb Khan , Ashraf Ullah, "Mining opinion components from unstructured reviews: A review",Journal of King Saud University – Computer and Information Sciences (2014) 26, 258–275. doi: <http://dx.doi.org/10.1016/j.jksuci.2014.03.009>.
- [3] Tsytsarau, Mikalai, and Themis Palpanas. "Survey on mining subjective data on the web." Data Mining and Knowledge Discovery vol. 24, pp. 478-514, 2012.
- [4] Turney, Peter D. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews." In Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, pp. 417-424, 2002.
- [5] Chetashri Bhadane,Hardi Dalal, Heenal Doshi, " Sentiment analysis: Measuring opinions ", International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015).
- [6] https://en.wikipedia.org/wiki/Sentiment_analysis.
- [7] Bei Yu, Stefan Kaufmann,Daniel Diermeier, "Exploring the Characteristics of Opinion Expressions for Political Opinion Classification", Proceeding of the 9th Annual International Digital Government Research Conference, 2007.
- [8] Hatzivassiloglou, V. & McKeown, K, "Predicting the semantic orientation of adjectives", Proceedings of the 35th ACL Conference, 174-181, 1997.
- [9] Dave, K., Lawrence, S., & Pennock, " Mining the peanut gallery: opinion extraction and semantic classification of product reviews", Proceedings of the 12th international conference on World Wide Web (WWW2003), 519-528,2003.
- [10] Hu, M. & Liu, B., "Mining and summarizing customer reviews", Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD2004), 168-177, 2004.
- [11] Pang, B., Lee, L., & Vaithyanathan, S, "Thumps up?, Sentiment classification using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP2002), 79-86, 2002.
- [12] Jack G. Cornrad et al. 'Professional Credibility: Authority on The web', WICOW08, 2008.
- [13] Nitin Jindal and Bing Liu, "Opinion Spam and Analysis", WSDM08, 2008.
- [14] Nitin Jindal and Bing Liu, "Review Spam Detection", WWW2007, ACM 978-1-59593-654-7/07/00005, 2007.
- [15] G. Salton, "Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer", Addison-Wesley, 1989.
- [16] Bo Pang and Lillian Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval Vol. 2, Nos. 1–2 (2008) 1–135 2008.
- [17] Carlo Strapparava, Rada Mihalcea, "Learning to Identify Emotions in Text", SAC'08 Fortaleza, Brazil 2008.

- [18] Michael Gamon and Anthony Aue, "Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms", In Proceedings Of the Workshop on Feature Engineering for Machine Learning in Natural Language Processing pages 57–64.ACL 2005.
- [19] Antonio Lieto, "Manually vs semiautomatic domain specific ontology building", Thesis dissertation, Chapter 1, 2008.
- [20] Farah Benamara, Carmine Cesarano, Diego Reforgiato, "Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone", ICWSM '2007 Boulder, CO USA.
- [21] Khairullah Khan, Baharum B.Baharudin, Aurangzeb Khan, Fazal-e-Malik, "Mining Opinion from Text Documents: A Survey", 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, 217-222.

