

Opinion Ana Data Mining of Customer Feedback Data on the Web

Malarvizhi S

Research Scholar, Dept. of MCA, VELS University, Chennai.

Dr.S.Prasana

Asst. Prof, Dept. of MCA, VELS University, Chennai.

Dr.A.Muthukumaravel

Professor & Head, Department of MCA, BHARATH University, Chennai.

Abstract - The advent of Web 2.0 and social media content has stirred much excitement and created abundant opportunities for understanding the opinions of the general public and consumers toward social and political movements, strategies, campaigns, and product preferences. Many new and exciting social, business -related research questions can be answered by analyzing the evens , comments and responses expressed in various blogs , forums (such as Yahoo,Google,etc.), social media and social network sites (including YouTube, Face book.), virtual worlds (such as Second Life), and tweets (Twitter). Opinion mining, a sub discipline within data mining and computational linguistics, refers to process of computing for assessing the opinions expressed in various online services social media comments, and other user-generated content. The main goal of this project is to accessing the opinions expressed in various online services. Here opinion mining refers to computational techniques for analyzing the opinions that are extracted from all kind of sources. Current opinion research focuses on business and e-commerce such as product reviews and movie ratings.

Keywords: Opinion mining, customer feedback, sentiment classification, linguist.

I. INTRODUCTION

The World Wide Web is growing at an alarming rate not only in size but also in the types of services and contents provided. Individual users are participating more actively and are generating vast amount of new data. These new Web contents include customer reviews and blogs that express opinions on products and services – which are collectively referred to as customer feedback data on the Web. As customer feedback on the Web influences other customer’s decisions, these feedbacks have become an important source of information for businesses to take into account when developing marketing and product development plans.

Let us consider an example of customer feedback. “This camera is my first digital one and was super easy to learn to use. The picture looks great and it’s simple to get the correct exposure. The memory card that comes with the Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

camera has a very small capacity though, (it holds about 4 photos) so a separate memory card is a necessity. I’m not very happy with the memory card.”

In this example, we can extract several phrases such as ‘super easy to learn to use’, ‘the picture looks great’, ‘simple to get the correct exposure’, ‘very small capacity’, and ‘not very happy with the memory card’, which convey customer’s opinion rather than facts. In particular, subjective words such as ‘super easy’, ‘looks great’, ‘simple’, ‘very small’, and ‘not very happy’ are used to express customer’s positive/negative sentiment regarding the product features, which are referred by ‘learn to use’, ‘picture’, ‘exposure’, ‘capacity’, and ‘photo’. Although information gathered from multiple reviews are more reliable compared to information from only one review, manually sorting through large amounts of review one by one requires a lot of time and cost for both

businesses and customers. Therefore it is more efficient to automatically process the various reviews and provide the necessary information in a summarized form.

Because of the importance of automatically extracting actionable knowledge from customer feedback data on the Web, “opinion mining (OM)” has become a significant subject of research in the field of data mining. The ultimate goal of OM is to extract customer opinions (feedback) on products and present the information in the most effective way that serves the chosen objectives. This means that the necessary steps and techniques used for OM can be different depending on how the summarized information is presented.

For example, if we were to get the number of negative and positive reviews about a given product, classifying each review as positive or negative would be the most important task. On the other hand, if we want to show customer feedback on each of the different features of a product, it is necessary to extract product features and analyze the overall sentiment of each feature.

The methods that are needed for feature extraction, sentiment classification, and opinion summarization have already been targets of research in other areas such as document classification and text summarization. These can be modified and applied to OM.

However, the focus of opinion mining is on the sentiment that the customer is expressing and this is where the methods are applied differently. As can be seen from the prior example, making the linguistic distinction between objective words that express facts and subjective words that express opinions is important.

In this paper, we examine the two tasks that are specific to opinion mining: development of linguistic resources and sentiment classification. In addition, we present opinion summarization by looking into the existing opinion mining systems which extract opinion expression from large reviews and show how each system applies the methods in order to effectively summarize and present the opinions.

The remainder of the paper is organized as follows. Section 2 introduces the tasks for opinion mining and section 3 presents the methods for defining and developing linguistic resources to be used for OM.

II. TASKS FOR OPINION MINING

As mentioned in Section 1, opinion mining can be roughly divided into three major tasks of development of linguistic resources, sentiment classification, and opinion summarization.

Appraisal theory [14, 1] developed by Martin J.R., a computational linguistics researcher, concisely defines the sentiment properties of the linguistic resources that can be used for opinion mining. The techniques used for text classification and text summarization can also be applied to OM, along with linguistic resources. Although sentiment classification and opinion summarization share several steps or techniques, sentiment classification focuses on classifying each review while opinion summarization is about how to effectively extract opinion expressions and summarize them from a large number of reviews of a given product.

III. DEVELOPMENT OF LINGUISTIC RESOURCE

Sentiment related properties are well defined in appraisal theory [14] which is a framework of linguistic resources for describing how writers and speakers express inter-subjective and ideological positions. However, most researches for developing linguistic resources have focused on determining three properties: subjectivity, orientation, and strength of term attitude. For example, ‘good’, ‘excellent’, and ‘best’ are positive terms while ‘bad’, ‘wrong’, and ‘worst’ are negative terms. ‘Vertical’, ‘yellow’, and ‘liquid’ are objective terms. ‘Best’ and ‘worst’ are more intense than ‘good’ and ‘bad’

There are four major approaches in developing linguistic resources for OM: the conjunction method, the point wise mutual information (PMI) method, the WorldNet exploring method, and the gloss classification method.

3.1 Conjunction Method

The work presented in [7] is the first attempt to automatically develop linguistic resources for opinion mining. The approach Relies on an analysis of textual corpora that correlates linguistic features or indicators with semantic orientation. The authors Demonstrated that conjunctions between adjectives provide indirect information about orientation, based on the hypothesis that “The conjoined adjectives and conjunctions usually have Similar orientation, though ‘but’ is used with opposite orientation.”

Their system identifies and uses this indirect information in the following steps: First, all conjunctions of adjectives are extracted from the corpus along with relevant morphological relations. And then, a log-linear regression model combines information from different conjunctions to determine if each of the two conjoined adjectives is of the same or different orientation. The result is a graph with hypothesized same- or different-orientation links between

adjectives. Here, clustering algorithm that separates the adjectives into two subjects of different orientation is applied. It places as many words of the same orientation as possible into the same subset.

Finally, the average frequencies in each group are compared and the group with the higher frequency is labeled as positive. Through this approach, decisions on individual words are aggregated to provide decisions on how to group words into a class and whether to label the class as positive or negative. Thus the overall result can be much more accurate than the individual indicators.

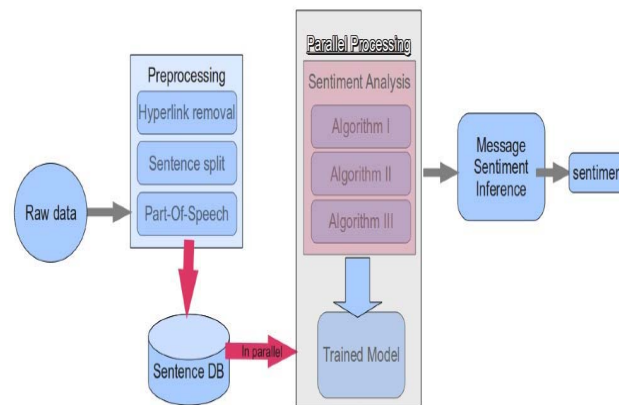
IV. SENTIMENT ANALYSIS OR OPINION MINING –ALGORITHMAM

Refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation

4.1 Sentiment Analysis

The goal is to mine opinions from social media data at sentence and document level.

Design a framework that can determine the sentiment (positive, negative, or neutral) overcoming social slang and lingos, non-textual expressions and language.



SES Architecture

4.2 Text classification

The goal is to group similar text or messages that can make information more manageable. Employ framework that can cluster similar text belong to the same topic and several similar topics that can belong to a broad category.

V. CONCLUSION

In this paper, we proposed a number of techniques for mining opinion features from product reviews based on data mining and natural language processing methods. The objective is to produce a feature-based summary of a large number of customer reviews of a product sold online. We believe that this problem will become increasingly important as more people are buying and expressing their opinions on the Web. Our experimental results indicate that the proposed techniques are effective in performing their tasks. In our future work, we plan to further improve these techniques. We also plan to group features according to the strength of the opinions that have been expressed on them, e.g., to determine which features customers strongly like and dislike. This will further improve the feature extraction and the subsequent summarization.

REFERENCES

- [1] Appraisal Analysis. <http://www.grammatics.com/appraisal/index.html>.
- [2] Argamon, S, Bloom, K., Esuli, A. and Sebastiani, F. 2007. Automatically determining attitude type and force for sentiment analysis. In *Proceedings of the 3rd Language and Technology Conference* (Poznań, Poland, October 05-07, 2007). 369-373.
- [3] Dave, K., Lawrence, S., and 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* (Budapest, Hungary, May 20 - 24, 2003). WWW '03. ACM, New York, NY, 519-528.

- [4] Esuli, A. and Sebastiani, F. 2005. Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international Conference information and Knowledge Management* (Bremen, Germany, October 31 - November 05, 2005). CIKM '05. ACM, New York, NY, 617-624.
- [5] Esuli, A. and Sebastiani, F. 2006. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of the Eleventh Conference on European Chapter of the Association for Computational Linguistics* (Trento, Italy, April 03 - 07, 2006). European Chapter Meeting of the ACL. Association for Computational Linguistics. 193-200.
- [6] Esuli, A. and Sebastiani, F. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of Fifth Conference on Language Resources and Evaluation* (Genova, Italy, May 22-28, 2006). 417-422.
- [7] Hatzivassiloglou, V. and McKeown, K. R. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics* (Madrid, Spain, July 07 - 12, 1997). European Chapter Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ, 174-181.
- [8] Hiroshi, K., Tetsuya, N., and Hideo, W. 2004. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th international Conference on Computational Linguistics* (Geneva, Switzerland, August 23- 27, 2004). International Conference on Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 494.
- [9] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining* (Seattle, WA, USA, August 22 - 25, 2004). KDD '04. ACM, New York, NY, 168-177.
- [10] Hu, M., and Liu, B. 2004. Mining Opinion Features in Customer Reviews. In *Proceedings of Nineteenth National Conference on Artificial Intelligence* (San Jose, California, USA, July 2-29, 2004). The AAAI Press, Menlo Park.