



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Survey of Various Opinion Mining Approaches

Gayathri R Krishna¹, Jothi S², Minojini N³, Sowmiyaa P⁴

PG Students, Department of Computer Science and Engineering, Dr.NGP Institute of Technology, Anna University, Coimbatore, India^{1, 2, 3, 4}

ABSTRACT: Opinion mining or sentiment analysis extract specified information from a large amount of text or reviews given by the internet users. Opinion mining classifies the large text of opinions as positive (good), negative (bad) or neutral. According to the number of positive, negative and neutral reviews, the product or service will be rated. Sometimes an overall rating for a review cannot be helpful to identify various features of a product or service. For example, a camera may come with excellent battery life but poor image quality. Hence more sophisticated aspect level opinion mining approaches have been proposed to extract information from online reviews. In this paper, we are discussing various approaches used for opinion mining. They are frequency-based approach, relation-based approach, supervised learning and topic modelling.

KEYWORDS: Aspect mining, opinion mining, supervised learning, text mining.

I. INTRODUCTION

Most of us always give importance to what other people think. It is an important piece of information during the decision making process. Many of us asked our friends or relatives to recommend a good car or to say who they are planning to vote in various elections, requested reference letters for job application, or consulted *Consumer Reports* to decide what washing machine to buy. Now World Wide Web become widespread, it is possible to find out opinions and experiences from a vast pool of people that are neither personal acquaintances nor well-known professional critics - that is, people we have never heard of. More and more people are making their opinions available to strangers via the Internet.

With the advent of Web 2.0 [1], [2], people are encouraged to contribute their own contents to the web. Now many user-centered platforms are available for information sharing and user interaction. Some of them are Epinion, Amazon, Facebook and Twitter. When people are interested in buying a product or a service, they usually not only look for official information from product manufacturers or service providers, experienced and practical opinions from the customers' and users' points of view are also influential. Hence, online reviews, blogs and forums dedicated for different kinds of products are pervasive, and how to effectively analyse and exploit such immense online information source is a challenge.

Opinion mining or sentiment analysis [3]–[5] involves the computational study of opinions. It extracts information from a large amount of text opinions or reviews given by Internet users. The information is positive or negative sentiments of a product. Based on the positive and negative aspects of a product, the product or service can be rated. Most of the time overall rating for a review cannot correctly reflect the various characteristics of a product or service. Hence, more effective opinion mining approaches have been proposed to extract and groups aspects of a product or service and predict their sentiments or ratings [3], [6]–[9]. In this paper, we are going to discuss some of the approaches used for opinion mining. They are frequency-based approach [10], relation-based approach [11], [12], supervised learning [13] and topic modelling [14], [15]. Frequency based approach extract information from reviews based on the frequency or strength of opinions. Relation-based approach extracts information from reviews based on the relation between aspect and sentiment. Supervised learning approach used the correctly defined labels to classify the reviews. Topic modelling approach is mostly used in probabilistic models.

The paper is organized as follows. Various approaches used for opinion mining are described in section 2. Section 3 describes the conclusion of the paper.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

II. DIFFERENT APPROACHES USED FOR OPINION MINING

What other people think has always been an important part of our information-gathering behaviour. Availability and popularity of opinion-rich resources such as online review sites and personal blogs are growing. Hence new opportunities and challenges arise as people can actively use information technologies to find out and understand the opinions of others. Opinion mining or sentiment analysis helps us to process the opinion from others. In this section, we are analysing some of the approaches used for opinion mining.

A. Frequency-based approach:

Product reviews on Internet sites such as amazon.com and elsewhere often associate meta-data with each review indicating how positive (or negative) it is using a 5-star scale, and also rank products by how many positive reviews at the site. However, the reader's preference may differ from the reviewers'. For instance, the reader may want to know about the quality of the gym in a hotel, but reviewers may focus on other aspects of the hotel, such as the decor or the location. Hence, reader is forced to wade through a large number of reviews looking for information about particular features of interest.

We decompose opinion mining problem into the following main subtasks:

- Identify features associated with the product
- Identify opinions regarding product features.
- Determine the polarity of opinions as positive and negative.
- Rank opinions based on their strength.

OPINE [10], an unsupervised information extraction system can be used to solve all these tasks. OPINE uses the frequency based approach for opinion mining. It mines reviews to build a model of important product aspects. Given a particular product and a corresponding set of reviews, OPINE solves the opinion mining tasks outlined above and outputs a set of *product features*, each accompanied by a list of *associated opinions* which are ranked based on strength (e.g., "abominable" is stronger than "bad). This output information can then be used to generate various types of opinion summaries.

OPINE uses association rule mining to extract *frequent* review noun phrases as features. Frequent features are used to find *potential opinion* words (only adjectives) and the system uses Word-Net synonyms/antonyms in conjunction with a set of seed words in order to find actual *opinion* words. Finally, opinion words are used to extract associated *infrequent* features. The system only extracts *explicit* features.

B. Relation-based approach:

- *Opinion Observer*

A prototype system called *Opinion Observer* [11] uses the relation-based approach for opinion mining. In this, with a single glance of its visualization, the user can identify the strengths and weaknesses of each product in the minds of consumers in terms of various product aspects. Both potential customers and product manufacturers can benefit from this comparison. For a potential customer, although he/she can read all reviews of different products at merchant sites to mentally compare and assess the strengths and weaknesses of each product in order to decide which one to buy, it is much more convenient and less time consuming to see a visual feature-by-feature comparison of customer opinions in the reviews. A system like ours can be installed at a merchant site that has reviews so that potential buyers can compare not only prices and product specifications (which can already be done at some sites), but also opinions from existing customers. For a product manufacturer, finding the strengths and weaknesses of their product is very crucial. Because market research can be done using this information. Product benchmarking can be also done using this information. *Opinion Observer* is helpful for product manufacturer also.

- *Multi-facet rating*

Software tools to organize product reviews and to make them easily accessible to prospective customers are going to be more and more popular. Some of the issues that the designers of these software tools need to address are pulling together reviews from various resources, filtering our fake reviews given by authors with vested interests and ranking products automatically products in terms of the satisfaction of consumers that have purchased the product before.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Multi-facet rating [12] address a problem related to automatically rating (i.e., attributing a numerical score of satisfaction to) consumer reviews based on their textual content. This problem arises when some online product reviews consist of a textual evaluation of the product *and* a score expressed on some ordered scale of values, while other reviews contain a textual evaluation only. These latter reviews are difficult to manage automatically, especially when a qualitative comparison among them is needed in order to determine whether product x is better than product y , or to identify the best product in the lot.

Tools capable of interpreting a text-only product review and scoring it according to how positive the review is, are thus of the utmost importance. In particular, our work addresses the problem of rating a review when the value to be attached to it must range on an *ordinal* (i.e., discrete) scale. This scale may be in the form either of an ordered set of *numerical* values (e.g., one to five “stars”), or of an ordered set of *non-numerical* labels (e.g., Poor, Good, Very good, Excellent); the only difference between these two cases is that, while in the former case the distances between consecutive scores are known, this is not true in the latter case.

In *multi-facet* rating of product reviews, the review of a product (e.g., a hotel) must be rated several times, according to several features of the product (for a hotel: cleanliness, centrality of location, etc.). The system realized could work as a building block for other larger systems that implement more complex functionality. Multi-facet rating uses the relation-based approach for opinion mining.

C. Supervised learning:

The OpinionMiner system used the supervised approach for opinion mining. Some merchants who sell their products on the Web ask their customers to share their opinions and hands-on experiences on products they have purchased. But, reading through all customer reviews is difficult, especially when the number of reviews can be up to hundreds or even thousands. Hence it is difficult for a potential customer to read them to make an informed decision. The OpinionMiner [13] system is designed with a aim to mine customer reviews of a product and extract high detailed product entities on which reviewers express their opinions. This system first identifies opinion expressions and then opinion orientations for each recognized product entity are classified as positive or negative.

D. Topic modelling:

Topic-Sentiment Mixture (TSM) [14] is an opinion mining model which uses the topic modelling approach. It is a probabilistic model. This model addresses the Topic-Sentiment Analysis (TSA) problem and extracts the multiple subtopics and sentiments in a collection of blog articles. A blog article is considered to be “generated” by sampling words from a mixture model. The mixture model contains a background language model, topic language models, and two (positive and negative) sentiment language models. By using this model, we can extract the topic/subtopics from blog articles, reveal the correlation of these topics and different sentiments, and further model the dynamics of each topic and its associated sentiments.

Aspect and Sentiment Unification Model (ASUM) also uses the topic modelling approach for opinion mining. It is also a probabilistic generative model which automatically discover aspects people evaluate and different sentiments toward these aspects. ASUM incorporates sentiment and aspects together to discover from reviews the aspects that are evaluated positively and the ones evaluated negatively. The model reviews electronic devices, restaurants and photo critiques. The results show that the aspects discovered by ASUM match evaluative details of the reviews and capture important aspects that are closely coupled with a sentiment.

III. CONCLUSION

Various approaches have been proposed for doing opinion mining or sentiment analysis. Some of them are frequency-based approach, relation-based approach, supervised learning and topic modelling. Frequency based approach mines information based on the strength of opinions. Relation-based approach mines information based on the relation between product features and people sentiments. Supervised learning approach defines the data correctly in terms of labels. Topic modelling approach is implemented using probabilistic models.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

REFERENCES

1. T. O'Reilly, "What is web2.0: Design patterns and business models for the next generation of software," Univ. Munich, Germany, Tech. Rep. 4578, 2007.
2. D. Giustini, "How web 2.0 is changing medicine," *BMJ*, vol. 333, no. 7582, pp. 1283–1284, 2006.
3. B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Ret.*, vol. 2, no. 1–2, pp. 1–135, Jan. 2008.
4. L. Zhuang, F. Jing, and X. Zhu, "Movie review mining and summarization," in *Proc. 15th ACM CIKM*, New York, NY, USA, pp. 43–50, 2006.
5. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. KDD*, Washington, DC, USA, pp. 168–177, 2004.
6. C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proc. 18th ACM CIKM*, New York, NY, USA, pp. 375–384, 2009.
7. I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," in *Proc. 46th Annu. Meeting ACL*, pp. 308–316, 2008.
8. Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: Modeling facets and opinions in weblogs," in *Proc. 16th Int. Conf. WWW*, New York, NY, USA, pp. 171–180, 2007.
9. S. Moghaddam and M. Ester, "Aspect-based opinion mining from online reviews," in *Proc. Tutorial 35th Int. ACM SIGIR Conf.*, New York, NY, USA, 2012.
10. A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proc. Conf. Human Lang. Technol. Emp. Meth. NLP*, Stroudsburg, PA, USA, pp. 339–346, 2005.
11. B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. 14th Int. Conf. WWW*, New York, NY, USA, pp. 342–351, 2005.
12. S. Baccianella, A. Esuli, and F. Sebastiani, "Multi-facet rating of product reviews," in *Proc. 31st ECIR*, Berlin, Germany, pp. 461–472, 2009.
13. W. Jin, H. Ho, and R. Srihari, "Opinionminer: A novel machine learning system for web opinion mining and extraction," in *Proc. 15th ACM SIGKDD Int. Conf. KDD*, New York, NY, USA, pp. 1195–1204, 2009.
14. Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: Modeling facets and opinions in weblogs," in *Proc. 16th Int. Conf. WWW*, New York, NY, USA, pp. 171–180, 2007.
15. Y. Jo and A. Oh, "Aspect and sentiment unification model for online review analysis," in *Proc. 4th ACM Int. Conf. WSDM*, New York, NY, USA, pp. 815–824, 2011.