International Journal for Modern Trends in Science and Technology
Volume: 02, Issue No: 10, October 2016
http://www.ijmtst.com
ISSN: 2455-3778

Opinion Mining and Product Features Ranking: A Survey

Chandana Oak[1] | Prof. Manisha Patil[2]

[1, 2] Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

To Cite this Article
Chandana Oak and Prof. Manisha Patil, Opinion Mining and Product Features Ranking: A Survey, International Journal for Modern Trends in Science and Technology, Vol. 02, Issue 10, 2016, pp. 77-81.

ABSTRACT

With the ever increasing popularity of e-commerce, many customers are buying their products online. They also rate and review these products over multiple review sites, social networking sites, blogs, etc. These online reviews are helpful for potential customers in decision making to buy the products and also to manufacturers/ sellers to get immediate feedback. Since the number of reviews for a product is usually large, it is very difficult to read all the reviews and form an opinion about the product. Hence, online review mining is of immense importance. Also, if the product features are ranked as per their popularity, then it increases review usefulness.

This survey paper studies different methodologies for online review mining as well as for product feature ranking.

KEYWORDS: *Opinion mining, Product Aspect Ranking, Online review mining*

I. INTRODUCTION

Online reviews of different products are springing up with the ever increasing popularity of e-commerce. The manufacturers and sellers of products ask the purchasers to review their products online to enhance customer satisfaction and improve the overall purchase experience. This enables other potential customers to gain first hand information about the product and helps them take a decision whether to buy the product or not. It also enables the product manufacturers and sellers to get an immediate feedback about the product, the delivery, after sales service, etc. This feedback serves as an opportunity for the manufacturers and sellers to improve on their service.

A. Opinion Mining -

Opinion mining or sentiment analysis deals with processing of natural language for deciphering the mood or feedback of customers about a particular product. [15] It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features. For example, a review on a website might be overall positive about a digital camera, but can be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer. The generic architecture for opinion mining can be shown with the help of Fig. 1

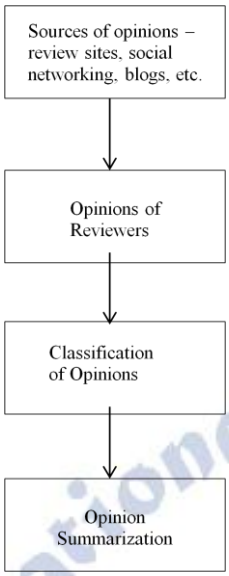

Fig. 1: Generic architecture for opinion mining

Reviewers express their opinions about a particular product or a service (or restaurants, hotels, movies, etc.) through some review sites, social networking sites like facebook, twitter, etc. or even over blogs. All these act as sources for opinions.

Opinions need to be retrieved from these sources and information retrieval techniques like web crawler can be used to fetch data from these sources and store them into a database. These opinions are analyzed, processed and then classified as positive or negative. Opinion summarization is the main part in opinion mining. Reviews need to be summarized in a format which can be easily deciphered. "Strength" (strongly or mildly positive or negative) can also be assigned to reviews.

[15] Challenges faced by opinion mining -

• A word which is considered to be positive in some reviews can be treated as negative in other reviews. For example, long battery life for a laptop is a positive indicator whereas long start up time is a negative indicator.

• Reviewers can express their opinion in a mixed fashion having a combination of both positive and negative comments. For example, the hotel was not clean but the food was excellent. This is easy to understand if someone is reading the feedback but is difficult to parse.

B. Product Aspect Ranking –

Opinion mining goes hand in hand with Product Aspect Ranking. Generally, a product may have hundreds of features, some of which holding more importance than others. Some examples of product aspects or features can be screen resolution, sound quality, battery life, etc. for a smart phone. These important features have a greater impact on decision making by potential customers as well as product development strategies by organizations. Ranking these features as per some parameters, like frequency at which the product features were commented on would help increase the usefulness of online reviews.

II. MOTIVATION

Many of the customer reviews are long with usage of multiple sentences mentioning their feedback about the product. If there are more than a hundred reviews associated with a product, it makes it difficult for a manufacturer/seller and a potential customer to go through all the reviews and build their own opinion for the product. Then he/she might go through just a few reviews making their own opinion biased. Additionally, customers can express their opinions through posts on social media or even through blogs. It makes it very difficult for the manufacturer/seller and a potential customer to browse multiple sites to get all the reviews. Thus, opinion mining has become a necessary activity.

Product aspect ranking organizes the important features of a product to improve review usability. But, manual identification of the important features of a product is very difficult due to the large number of reviews, most of them which might be verbose. Thus, a methodology to automatically extract the important aspects of a product is much necessary.

III. RELATED WORK

The popularity and necessity for online review mining are increasing day by day. [1], [2] and [3] showcase methods to collectively extract opinion targets and opinion words in a bootstrapping manner. Opinion target is the product feature which is reviewed while opinion word is the word which expresses the reviewer's opinion about the product feature. For example, for the review "The laptop battery lasts long", laptop battery is the opinion target and long indicates the user opinion about laptop battery, so it is the opinion word. [1] classifies each sentence in the review and not the complete review as a whole. It uses association mining technique to identify the frequently occurring nouns and noun phrases in a review sentence. The most frequent ones are then

extracted as opinion targets. It then picks up the positive and negative review and generates a summary of the product feature and the positive and negative reviews associated with it. Pronoun resolution and investigating opinions associated with adverbs and verbs can be worked upon further. Also, this method does not take into consideration the strength of opinions i.e. whether an opinion is strongly or mildly positive/negative.

[2] proposes a graph based algorithm, Word-based Translation Model (WTM) for extracting opinion target or for product feature extraction. WTM overcomes the drawbacks of traditional opinion target extraction methods viz. adjacent methods and syntax based methods. Thus, this method is independent of parsing performance which plays a major role in syntax based methods and of window size which is used in adjacent methods to find opinion relations with the surrounding opinion words. It can also capture one to many (in a sentence one opinion word modifies many opinion targets) and many to one (many opinion words modify one opinion target) relations. This WTM algorithm mines all possible options for opinion targets in a sentence, uses graphical algorithm to assign a confidence to each option, ranks them as per confidence level and selects the best possible option as opinion target. It can be also improved upon by adding some syntactic information for higher precision to identify opinion relations between words. Also, this model needs to be enhanced to consider verbs as opinion words for opinion target extraction.

[3] proposed a method called Double Propagation which extracts opinion words (or targets) iteratively from the words and targets already mined during the previous iteration using syntactic relations. This method defines certain rules to find the relationship between opinion targets and words and works iteratively until no more new opinion words or targets can be identified. The main drawback for this method is that syntactic rules cannot cover all opinion relations.

Bootstrapping methods in [1], [2], [3] have a problem of error propagation. If some errors are extracted in an iteration, they will not be eliminated in subsequent iterations.

[1] and [4] use POS tagging technique. This Part of Speech (POS) tagging technique is used to identify nouns or noun phrases in the reviews. Thus it can be used to extract product features from reviews. [4] uses nearest neighbor rule which regards nearest adjective/verb to a noun as its opinion word. Thus, it works in a limited span and cannot obtain accurate results.

[5] and [6] focus on sentence level extraction wherein the task of opinion target or word extraction is from sentences in reviews. In [5], Conditional Random Fields (CRFs) are used to jointly extract object features and positive/negative opinions from review sentences. In [6], a concept of phrase dependency parsing is used by the observation that many product features such as image quality, battery life, etc. are noun phrases rather than just nouns.

An OPINE algorithm is proposed in [7] which is built on top of the Know-ItAll Web information-extraction system (Etzioni et al., 2005). It uses syntactic parsing of reviews to derive opinion relations among words. But it is heavily dependent on parsing performance which is not good in case of informal writing style of online reviews.

A partially supervised word alignment model (PSWAM) is used in [8] which extracts opinion targets and opinion words based on a graph-based co-ranking algorithm which estimates the confidence of each candidate. The candidates with higher confidence are co-extracted as opinion targets and opinion words. This method extracts opinion words and targets more precisely than nearest neighbor rules or syntax based methods. PSWAM uses syntactic parsing to obtain the primary partial alignments and then uses a hill climbing algorithm to determine all alignments in sentences. Opinion Relation Graph is constructed to model all opinion words/targets and relations among them. A co-ranking algorithm finds out candidate's confidence. Candidates with higher confidence than a threshold are extracted as opinion targets and words.

The references covered so far in this section focus on extracting opinion targets or opinion words from online reviews. None of these references cover product aspect ranking. [9], [10], [11], [13] and [14] focus on product aspect ranking and its different applications. [9] uses double propagation method combining it with part-whole relationship pattern and "no" pattern for feature extraction. This will improve recall, as double propagation alone cannot find all features for large or small corpora. The reference paper also proposes ranking of products based on feature importance i.e. a combination of feature relevance and feature frequency. It states that if an extracted feature is correct and occurs frequently in reviews, it should be ranked high, else low. This paper does not take into consideration the influence of

customers' opinion on these aspects, on the overall opinion though.

[10] uses an aspect ranking algorithm which takes into consideration both the aspect frequency as well as their influence on overall opinion. For Pros and Cons reviews, this algorithm identifies the aspects as frequent nouns in the reviews and for free text reviews, a Stanford parser is used to parse each review and extract noun phrases as aspects. Aspect Sentiment classification is carried out to classify aspects as positive or negative based on the reviews. This uses an SVM classifier. The ranking algorithm is then applied on top of sentiment classification results.

[11] was studied for product aspect ranking. Normally, a product has many features, some being more important than others. These important features have a greater impact while potential customers buy products online. Thus, identifying important features will improve the usability of the reviews for both potential customers and manufacturers/sellers. The parameter considered for ranking is the frequency with which the product feature is commented upon and their influence on overall opinion.

Table 1 provides a summary of related work and proposed work.

Reference	Technique used for feature extraction	Is product feature ranking used?
Mining and summarizing customer reviews [1]	Association mining	No
Opinion target extraction using word based translation model [2]	Word Translation Model	No
Opinion word expansion and target extraction through double propagation [3]	Double Propagation	No
Mining opinion features in customer reviews [4]	Nearest neighbor rules	No
Structure-aware review mining and summarization [5]	CRF	No
Phrase dependency parsing for opinion mining [6]	Phrase dependency parsing	No
Extracting product features and opinions from reviews [7]	OPINE	No
Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model [8]	PSWAM	No
Extracting and ranking product features in opinion documents [9]	Double propagation with part-whole relationship pattern and "no" pattern	Yes
Aspect ranking: Identifying important product aspects from online consumer reviews [10]	Stanford parser and SVM classifier	Yes
Product Aspect Ranking and Its Applications [11]	Stanford parser and SVM classifier	Yes
Proposed work	PSWAM	Yes

Table 1: Summary of related and proposed work

IV. PROPOSED WORK

After the survey conducted, a need was observed to apply a ranking algorithm after co-extracting opinion words and opinion targets from online reviews. The word alignment model is used for co-extracting opinion targets and opinion words from reviews. Ranking is not used after this co-extraction. The proposed work builds a ranking framework over the existing co-extraction framework (PSWAM) which will organize all the product features as per their popularity i.e. as per the number of reviews a product feature receives, and also as per the influence a feature has on the overall opinion of the product. This will improve the usefulness of the review summarization.

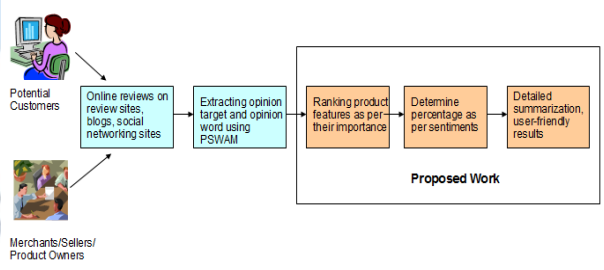

Fig. 2: Proposed work

The block diagram in fig. 2 shows the proposed work. Millions of reviews of different products are available online on different review sites, social networking sites, blogs, etc. These reviews are referred to by people wishing to buy a particular

product in order to get an opinion on the product. Also, merchants/sellers refer to these reviews in order to get a first hand feedback about their product/service. Opinion target and opinion word will be co-extracted using PSWAM from online reviews. The extracted features (or aspects) will then be ranked as per their importance. The extracted opinion words will be used for aspect sentiment classification. Depending on the positivity or negativity of the sentiment (i.e. how much positive or negative the opinion word weighs), the corresponding opinion target would be associated with the percentage as per aspect sentiment classification. The results would be presented in a detailed, user friendly format.

V. CONCLUSION

This survey paper presented an overview of different techniques for mining online customer reviews to extract opinion targets and opinion words. Review mining is useful for both potential buyers of the product in decision making as well as for manufacturers/sellers as they get quick feedback about the product.

The survey paper also provided a brief review of product aspect ranking which can be used to improve the usability of reviews.

ACKNOWLEDGEMENT

This survey would not have been possible without the support and help of many individuals. I would like to extent my sincere thanks to all of them. I am indebted to Prof. Manisha Patil, Department of Computer Engineering of Smt. Kashibai Navale College of Engineering affiliated to Savitribai Phule Pune University for her guidance and constant supervision as well as all the other staff members for providing important information regarding the survey.

REFERENCES

[1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, 2004, pp. 168–177.

[2] K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using word based translation model," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.

[3] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Comput. Linguistics, vol. 37, no. 1, pp. 9–27, 2011

[4] M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, 2004, pp. 755–760.

[5] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 653–661.

[6] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.

[7] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 339–346.

[8] Kang Liu, Liheng Xu, and Jun Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model", IEEE transactions on Knowledge and Data Engineering, Vol. 27, No.3, March 2015

[9] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 1462–1470.

[10] J. Yu, Z.-J. Zha, M. Wang, and T. S. Chua, "Aspect ranking: Identifying important product aspects from online consumer reviews," in Proc. ACL, Portland, OR, USA, 2011, pp. 1496–1505.

[11] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua, "Product Aspect Ranking and Its Applications", IEEE transactions on Knowledge and Data Engineering, Vol. 26, No.5, May 2014.

[12] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. Artificial Intelligence, 165(1):91–134.

[13] K. Zhang, R. Narayanan, and A. Choudhary. Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. WOSN, 2010.

[14] K. Zhang, R. Narayanan, and A. Choudhary, Mining Online Customer Reviews for Ranking Products, Technical Report, EECS department, Northwestern University, 2009.

[15] http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining

[16] http://ijiset.com/vol2/v2s6/IJISET_V2_I6_106.pdf