Effective Co-extraction of Online Opinion Reviews and Product Aspect Ranking

Chandana Oak
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Manisha Patil
Professor
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

ABSTRACT
As the popularity of e-commerce is increasing day by day, many customers have started buying their products online through e-commerce sites. Customers also normally review and rate the products they have bought over multiple review sites, social networking sites, blogs, etc. Such online reviews are of great help to people who are going for that product in decision making and also to manufacturers/sellers to get immediate feedback about the product quality or after sales service, etc. 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. Also, there are multiple sources of these online reviews. Hence, online review mining is gaining importance.

Another feature that can add to review usefulness is ranking the product aspects as per their importance and popularity. Ranking product aspects manually is very difficult since a product may have hundreds of features. So, an automated method to do this is needed.

This paper presents a methodology for online review mining as well as for product feature ranking. The paper proposes use of Naïve Bayes Classifier for product review category classification, partial parsing of classified reviews and an algorithm to co-extract opinion target and corresponding opinion words. It also proposes an algorithm for product aspect ranking using the extracted opinion target and word pairs.

General Terms
Online opinion mining

Keywords
Opinion mining, Sentiment Analysis, Product Aspect Ranking, Online review mining

1. INTRODUCTION

1.1 Opinion Mining
Opinion mining or sentiment analysis deals with mining and analysis of natural language for tracking the mood or feedback of people about a particular product. It can be treated in short as a system to gather and classify different opinions about a particular product or service. [1] It can help the marketing teams evaluate the success of an ad campaign, determine which versions of a product are popular in which areas, whether customers are satisfied about the after sale service provided for a particular product, etc. For example, a laptop review can have very high ratings on the overall functionality, but the reviews can specify that the laptop is pretty heavy and thus not easy to carry. Since this feedback comes directly from the customer, it is essential to decipher it in a systematic way. Such a feedback can provide data points about the product development strategy for future versions as well.

[2] gives the generic architecture for opinion mining.

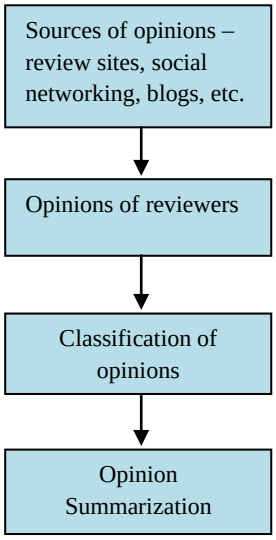


Fig 1: Generic architecture for opinion mining

People log their opinions on dedicated online review sites, social networking sites, blogs, etc. So, there can be multiple sources which a potential customer can refer to for gaining first hand information of the product which he/she is interested in buying. Information retrieval techniques like web crawling, etc. can be used to extract opinions from all these sources. These extracted opinions are stored into a database for further analysis. These extracted opinions are analyzed and classified using different sentiment classifiers.

1.2 Product Aspect Ranking
A potential customer is mainly interested in some important features or aspects of a product. E.g. while buying a smart phone, a customer will look for screen resolution rather than middle button. Thus, some product features hold more importance than others and they can also influence the overall rating for the product. Ranking product aspects as per some parameters would help increase the usefulness of online reviews.

2. MOTIVATION
Many online reviews are quite verbose. Also there might be a lot many reviews associated with the product of interest. This makes it difficult for a person to read the reviews and thus reduces the interest of any person reading those. Then he/she

might go through just a few reviews making their own opinion biased. Additionally, customers can express their opinions either on e-commerce sites, or social networking sites or blogs, etc. It makes it very difficult for a person to browse multiple sites to get all the reviews. Thus, opinion mining has become a necessary activity.

Product aspect ranking organizes the extracted features of a product as per their importance to improve review usability. But, manual identification of features of a product is very difficult due to the large number of reviews. Thus, an automated method to extract and rank the important aspects of a product is much necessary.

3. RELATED WORK

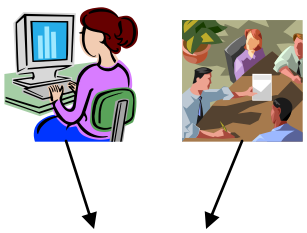
The popularity and necessity for online review mining are increasing day by day. [3] uses association mining technique to identify the frequently occurring nouns and noun phrases in a review sentence. The most frequent ones are then extracted as opinion targets. This technique is not suitable for opinion word extraction. [4] proposes Word-based Translation Model (WTM) which is a graph based algorithm for extracting opinion target. WTM does not involve syntactic parsing and is thus independent of parsing performance and also of window size which is used in adjacent methods to find opinion relations with the surrounding opinion words. [5] 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. [6] uses nearest neighbor rule which regards nearest adjective/verb to a noun as its opinion word. It can thus lead to incorrect results since it focuses on only the adjacent words. It cannot treat a distant opinion word as pairing with an opinion target.

[7] and [8] focus on sentence level extraction. In [7], Conditional Random Fields (CRFs) are used to jointly extract object features and positive/negative opinions from review sentences. In [8], phrase dependency parsing is used as many product features such as image quality, battery life, etc. are noun phrases rather than just nouns.

An OPINE algorithm is proposed in [9] which uses syntactic parsing of reviews to find opinion relations among words. Since it is very much dependent on parsing performance, its performance is not good for online reviews which are mostly written in an informal manner.

[10] uses a partially supervised word alignment model (PSWAM) which co-extracts opinion targets and opinion words based on a graph-based co-ranking algorithm. This method extracts opinion words and targets more precisely than nearest neighbor rules or syntax based methods.

[11], [12], [13], [14], [15] and [16] focus on product aspect ranking and its different applications. [11] uses double propagation for product feature extraction along with algorithm for ranking of products based on feature importance. [12] and [13] use an aspect ranking algorithm which takes into consideration both the aspect frequency as well as their influence on overall opinion.



[12] gives an overview of different algorithms for opinion target and opinion word extraction as well as methodologies for product aspect ranking.

[17] gives different opinion mining techniques.

Table 1: Summary of related work

Reference	Technique used for feature extraction	Is product feature ranking used?
Mining and summarizing customer reviews [3]	Association mining	No
Opinion target extraction using word based translation model [4]	Word Translation Model	No
Opinion word expansion and target extraction through double propagation [5]	Double Propagation	No
Mining opinion features in customer reviews [6]	Nearest neighbor rules	No
Structure-aware review mining and summarization [7]	CRF	No
Phrase dependency parsing for opinion mining [8]	Phrase dependency parsing	No
Extracting product features and opinions from reviews [9]	OPINE	No
Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model [10]	PSWAM	No
Extracting and ranking product features in opinion documents [11]	Double propagation with part-whole relationship	Yes
Aspect ranking: Identifying important product aspects from online consumer reviews [12]	Stanford parser and SVM classifier	Yes
Product Aspect Ranking and Its Applications [13]	Stanford parser and SVM classifier	Yes

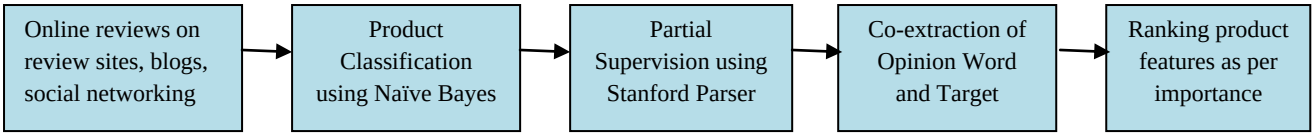


Fig 2: Architecture for Proposed System

The proposed work uses a Naïve Bayes Classifier for product category classification. The classified product reviews are partially parsed using a Stanford parser.

An Opinion TW Co-extraction Algorithm is proposed which co-extracts opinion targets and opinion words from classified and semi-supervised product reviews. In addition to this, product aspect ranking is also included as part of proposed work.

4. PROPOSED WORK

Fig. 2 shows the overall architecture for the proposed system.

4.1 Opinion Mining:

Millions of reviews of different products are available online on multiple sites such as different e-commerce sites like Flipkart, Amazon, etc. Users of the product can also express their feedback through 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. All these sources will act as the source for co-extraction algorithm.

An additional step for classifying the product reviews will be added before going for partial supervision. This product review classification will be done using Naïve Bayes classifier which segregates reviews into different categories such as mobile phones, laptops, cameras, etc. As the product reviews will be classified before feeding them to the Stanford parser for partial supervision, a more accurate training data will be generated as compared to feeding a combined set of all product reviews together. This is because opinion targets (i.e. nouns and noun phrases) and opinion words (i.e. adjectives) extracted from online reviews for different product categories will be different. E.g. – a review for mobile phone will have opinion targets i.e. product features as battery life, screen resolution, etc. whereas an LCD TV will have opinion targets as no. of modes supported, HDMI cable supported, etc. In a different domain, restaurant reviews can have opinion targets like quality of service, ambience, etc. Thus, training data generated for a mobile phone review will be irrelevant for a restaurant review. Classifying and categorizing product reviews will improve the reliability of the co-extracted opinion target and opinion word pairs in turn increasing the precision and recall values and reducing error rate or incorrect opinion targets and words generated.

Partial supervision technique will be used and parsing would be done using Stanford parser. Of the classified online product reviews, 50% of the review data would be considered for training the parser and remaining 50% data would be test data. This will be more accurate as compared to completely unsupervised model and would not be heavily dependent on parsing performance as in completely supervised model.

Co-extraction of opinion target and opinion words would be done using "Opinion TW Co-extraction Algorithm". This algorithm takes the partially supervised set of classified product reviews as input.

Algorithm 1: Opinion TW Co-extraction Algorithm

Input: Partially supervised set of classified product reviews I = {p1, p2, …, pn}

Output: The probability of co-alignment for sentences based on optimal association score

1. Initialization:
2. Initialize R as review data of products
3. Initialize review = i[data]
4. Initialize o_word [ ] = review [opinion word]
5. Initialize o_target [ ] = review [opinion target]
6. for each review[]
7. for each o_target[]
8. Nt := Count (opinion targets)
9. end for
10. for each o_word[]
11. Nw := Count (opinion words)
12. end for
13. Ntw := Count (collocated opinion target and opinion word)
14. P (o_target | o_word) = Ntw / Nw
15. P (o_word | o_target) = Ntw / Nt
16. Optimal Association Score OAS = [P (o_target | o_word) + P (o_word | o_target)] / 2
17. end for
18. Co-extract the opinion target and opinion word with higher OAS as being aligned with each other

---

The input to Opinion TW Co-extraction Algorithm is the set of product reviews which is already classified and segregated on a per product basis. Consider n to be the total number of products. R denotes the review collection per product. For each category of product review, assign "review" collection with the partially parsed review data for that product. Fetch the opinion words from review into o_word based on training

data. Similarly, fetch the opinion targets from review into o_target based on training data.

From the entire corpus, find Nt to be the total number of opinion targets and Nw to be the total number of opinion words. Also find the count of collocated opinion target and opinion word as Ntw. Find the estimated alignment probability for a potential opinion target and word pair as -

P (o_target | o_word) and P (o_word | o_target)

To find the optimal association score, take a mean of the 2 probability values fetched earlier. Co-extract the opinion target and opinion word with higher OAS as being aligned with each other.

To find one-many relation between opinion targets and opinion words (E.g. – The ambience and food in this restaurant are very good – here, good indicates both ambience and food), find 2 opinion targets separated by a conjunction and aligned to one opinion word. Similarly, find 2 opinion words separated by a conjunction and aligned to one opinion target. Co-extract these as opinion word and target having one-many relation.

4.2 Product Aspect Ranking:

Opinion target and word co-extraction can go hand in hand with product aspect ranking.

A product has many features. E.g. – considering mobile phone as a product, it has features like screen resolution, battery life, wifi connectivity, etc. As per the terminology used so far, they are also termed as opinion targets. Of all the features that a product has, some can be more important than others. A potential customer gives more importance to these features rather than giving much weightage to the less important ones while buying any product. Also, manufacturers can focus mainly on the more important features while developing their product development strategies. 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.

The proposed work will build a ranking framework on top of co-extraction framework 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. The same set of opinion targets extracted by the Opinion TW Co-extraction Algorithm will be used as product aspects or features and Naïve Bayes classifier for identifying the polarity of the feature expressed in the review based on opinion word extracted for the feature. The polarity will determine if the opinion associated with that review is positive or negative.

The Naïve Bayes classifier will classify each opinion word a positive or negative. Calculate the opinion of various aspects and also the importance weight for aspects. The ranking would be done based on these 2 parameters.

Since this work is currently in proposed stage, implementation is not done and results are not available to compare against the base system.

Algorithm 2: Product Aspect Ranking Algorithm

Input: Product name and opinion target-word pairs

Output: Ranked product aspects

```
1   Initialization:
2   Initialize set of product reviews as R
3   Initialize Negative sentimental word set N[]
4   Initialize Positive sentimental word set P[]
5   Initialize set of product aspects as aspect[]
6   Initialize set of overall ranking of aspect Or[]
7   for each aspect[]
8       Initialize array for sentiments as S[]
9       Int i = 0
10      For each i
11          If S[i] is negative then
12              N[i] = S[i]
13          Else if S[i] is positive then
14              P[i] = S[i]
15      End if
16    end for
17  end for
18  Overall rating of product as o[]
19  Calculate Importance weight for aspect wt[] as frequency of the aspect commented on
20  Or[] = o[] * wt[]
21  Rank the product aspects as per Or[] value
```

4.3 System Requirements:

The system should be able to classify product reviews into different categories and perform semi-parsing on those classified reviews. The system should be able to identify opinion target-word pairs and also rank product aspects as per their importance. The system should authenticate users and only authorized users should be provided access. The system should exhibit good performance.

5. CONCLUSION AND FUTURE WORK

This paper presents an overview of the proposed architecture and algorithms which will be used for opinion target and corresponding opinion word co-extraction from online reviews. It also proposes ranking of different product features as per the frequency they are commented on and the impact the product feature makes on the overall product.

As future work, the aim is to implement the application on the basis of these algorithms and compare the results achieved against state of the art methods. Co-extracting and ranking can also be integrated with recommendation system to provide recommendations to the user with similar features and their importance as the product that is searched for.

6. ACKNOWLEDGMENTS

This work would not have been possible without the support and help of many individuals; a sincere thanks to all of them. A special thanks 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 research.

7. REFERENCES

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

[2] C. Oak, M. Patil, "Opinion Mining and Product Features Ranking: A Survey", International Journal for Modern Trends in Science and Technology, Volume: 02, Issue No: 10, October 2016

[3] 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.

[4] 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.

[5] 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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

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

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