International Journal of Engineering & Technology, 7 (4.5) (2018) 143-146

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Semi-supervised Approach for Opinion Mining using Online Product Review

Bhagyashree G. Bhongade[1]*, Ashwini V. Zadgaonkar[2]

[1, 2]Shri Ramdeobaba College of Engineering and Management,
Nagpur, Maharashtra, India,
**Corresponding author E-mail [1]bhongadebg@rknec.edu . [2]zadgaonkarav1@rknec.edu*

Abstract

The growth of the internet as a secure online shopping channel has developed since 1994. With the Increasing number of e-commerce portal, we are now heavily inclined to online shopping. One of the benefits of online shopping is the ability to read reviews about the product purchased. This paper presents a semi-supervised approach for opinion mining using online product reviews obtained from Amazon website. A semi-supervised model regards identifying opinion relation as an alignment process and gives more precision in comparison to unsupervised model. Opinion mining of online reviews is needed for first-hand assessments of product information and direct supervision of their purchase actions. Manufacturers can obtain immediate feedback and opportunities to improve the quality of their products in a timely fashion.

Keywords: *Online product reviews, Opinion Mining, Opinion target, Opinion words, Semi supervised model.*

1. Introduction

In the modern era, online shopping has become a new trend. Online shopping also provides the customer to express their views on the product they have purchased. These reviews discover the opinions, experiences and feelings of customers. Opinion mining is a type of natural language processing that analyzes the mood of the customer about a particular product. An opinion of the customers can be categorized as positive, negative or neutral. This type of classification is known as polarity classification.

In many cases, overall sentiment of the product is unsatisfactory for the readers. Many times the reviewers expect fine-grained opinion about the features of the product. For example “This phone’s camera is excellent but battery-life is very poor”.

From this reviews, readers can understand that the reviewer has expressed the positive opinion about the camera but the negative opinion about the battery life. To fulfil, this goal, list of opinion targets and opinion words must be extracted. Opinion targets are the words about which the users express their opinion like camera, battery-life and opinion words are the words which are used to express the opinion like very poor and excellent. Opinion words are important for determining the expression of the user. Opinion targets and opinion words usually co-occur in a sentence. Our task is to determine the relationship between these opinion words and opinion targets. We have used a semi-supervised method for this relation extraction. Semi-supervised learning is a class of supervised learning tasks and techniques that use unlabelled data for training purpose.

Previous methods have usually generated an opinion target list from online product reviews. As a result, opinion targets usually are product features or attributes. Accordingly, this subtask is also called as product feature extraction. In addition, opinion words are the words that are used to express users’ opinions. In the above example, “colourful”, “big” and “disappointing” are three opinion words. Constructing an opinion words lexicon is Also important because the lexicon is beneficial for Identifying opinion expressions. For these two subtasks, previous work generally adopted a collective extraction strategy. The intuition represent-ed by this strategy was that in sentences, opinion words usually co-occur with opinion targets, and there are strong modification relations and associations among them(which in this paper are called opinion relations or opinion associations). Therefore, many methods jointly extracted opinion targets and opinion words in a bootstrapping manner. For example, “colourful” and “big” are usually used to modify “screen” in the cell-phone domain, and there are remarkable opinion relations among them. If we know “big” to be an opinion word, then “screen” is very likely to be an opinion target in this domain. Next, the extracted opinion target “screen” can be used to deduce that “colourful” is most likely an opinion word. Thus, the extraction is alternatively performed between opinion targets and opinion words until there is no item left to extract.

2. Related Work

This section presents various work related to the opinion min-ing or sentiment analysis. There are basically two methods used for the sentiment classification first method is supervised method and the second method is an unsupervised method. The supervised methods like naive Bayesian, SVM, and maximum entropy classify the reviews based on machine learning. The second method is an unsupervised method where the classification is based on certain syntactic patterns that are used express opinions. In the paper [6], [21] presents a novel approach to identify feature specific expressions of opinion in product reviews with different features and mixed emotions. Here the author has used Direct Neighbor

Relation and Dependency Relation for Relation extraction and Classified the extracted opinion words as positive or negative using Rule-Based Classification and Supervised Classification. In [13], [19], the authors have focused on aspect extraction at a sentence level using different NLP techniques. The first task is to prepare the dataset by employing NLP techniques. The second task is to find Opinion words and map them to the product aspect. Supervised Methods for Aspect-Based Sentiment Analysis is used by [11], [18]. This paper has focused on contribution in SemEval2014. The author used CRF model with different features for aspect extraction. Z-score model for category detection Multinomial Naive-Bayes for polarity detection. [5], [9], [20] has used a two-fold rule-based method. In the first fold, SPRs are been used for the extraction of nouns/noun phrases as aspects and associated opinions using opinion lexicon. In the second fold, the proposed model searches aspects associated with the domain dependent opinions using the concept lexicon. In [3], [16], sentimental analysis of tweets from twitter microblogging site are classified and various classification algorithm is analyzed based on their precision and recall. [9], [17] uses WorldNet for statistical analysis and to gain some information about the movies. In this paper, WordNet identifies features and opinion from the blogs. [7], [8], [18] focuses on the sentimental analysis in three non-English languages and analyses the efficiency of the different algorithm for the purpose. [15], [21] has used an unsupervised approach named SentiWordNet lexicon. The author of this paper has done aspect level sentiment analysis and for polarity classification of sentiment calculation of sentiment, the polarity is done using SentiWordNet which is a lexical resource for opinion mining. [10], [5] has surveyed various supervised method for opinion mining, such as Naive Bayes Classification, Maximum Entropy, and SVM. [4], [14], [2] used a supervised method called LCCT (Lexicon-based and Corpus-based, Co-Training). The dataset used here is both English and Chinese.

In this paper, we are using a semi-supervised approach to build an opinion mining system. Our proposed system will be using real-time product reviews from Amazon websites. A JSON file of reviews on the mobile phone is taken from Amazon website. After Preprocessing like stopword removal, tokenization on this file Frequent Features are chosen like phone battery, screen sim card etc.

3. Proposed System

In this proposed system, initially, mobile phone reviews are collected from Amazon site. The collected reviews are in the form of JSON file format. Data pre-processing like stopword removal is performed on this. In Feature selection step we will take 14 frequent features from the review. Opinion words for this features will be extracted.The architecture of proposed method is depicted from the above diagram. The proposed method can be divided into two main tasks. First is extracting opinion target and opinion words from the reviews and the second is finding the polarity of the reviews. In the proposed method we are using a semi-supervised method for classification of online reviews. First, reviews are collected from Amazon site. Frequently occurring Opinion targets are selected from these reviews and opinion words for these opinion targets are extracted. Bag of words model is used for feature extraction in opinion target. In this model, we consider different reviews of the same user as one document and use 8 different patterns to extract opinion words for opinion target. And finally, the polarity of each opinion words is used to calculate the score. For example, the opinion targets in the sentence “This phone’s camera is excellent but battery-life is very poor” is camera and battery life. We can consider different re-view of the same person as a single document and do the part of speech tagging and figure out important patterns in it This pattern provide opinion target and opinion words. For example, in above example, excellent and very poor (phrase) are opinion words of opinion target camera. We use WordNet to extract synonyms of both selected predefined opinion target and we find out where these opinion targets are present in our dataset and extract opinion words from those reviews. The flowchart of the proposed system is shown in figure 1.

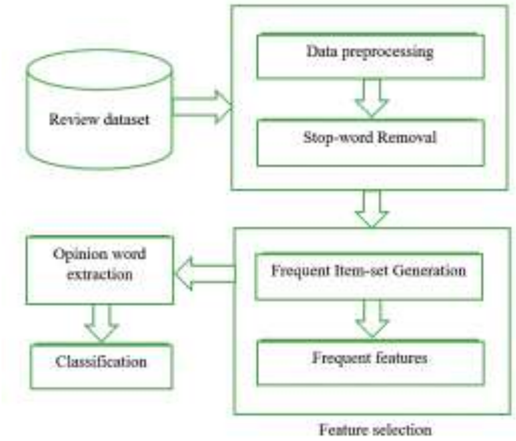


Fig 1: Flowchart of the proposed system

3.1. Stopword Removal

A stopword is a commonly used word such as “a”, “the”, “an”, “in” in a sentence. Such words are not so helpful in opinion mining. These words rarely indicate any meaning and thus needs to be removed. The reviews are given to stopword removal algorithm to remove the redundant words. Hence the review “The memory is sufficient” becomes “memory, sufficient” after stopword removal.

3.2. Frequent Itemset Selection

In Frequent itemset we select most common features in the reviews like screen, sim-card, power supply, application, display etc. These are the features on which mostly reviews are given by the reviewers. These are the features on which mostly reviews are given by the reviewers. In the proposed work we have initially selected 13 most common feature of the reviews such as:

Table 1: 13 most common features from review.

1.	Software
2.	Application
3.	Services
4.	Power supply
5.	Sim card
6.	Display
7.	Storage space
8.	Sensors
9.	Wireless charging
10.	Design
11.	CPU
12.	Accessories
13.	Camera

3.3. Opinion Word Extraction

In opinion extraction, we extract the opinions related to the features. For example, “The screen is big and colourful.” The opinion words related to the feature screen is big and colourful. Hence the feature screen is aligned with opinion word “big” and “colourful”. In this, we have used naïve based classifier word alignment. Here, the target words are related to their opinion words.

Table 2: Examples of opinion targets and their corresponding opinion words

Opinion targets	**Opinion words**
Display	Big, colourful,
Application	Slow, fast

Camera	Tiny,sensor,pixel etc.
Power supply	Poor, long-lasting
Storage space	Small, huge, less
Services	Excellent, bad, rich
Accessories	Cool, reliable
Simcard	Small, dual
Software	Excellent, useful, worthless
CPU	Speed, processing power, high performance

3.4. Classification

In classification, we classify the opinion words as positive or negative by assigning weight to each opinion words using.

Algorithm:

The steps of opinion mining using a semi-supervised method:

Step 1: Reviews are collected about mobile phone from the Amazon site

Step 2: Pre-processing like stopword removal is performed on this reviews. Here the common occurring word like “a”, “an”, “the”, “is”, “in”, “on” which gives little information about the sentence is removed.

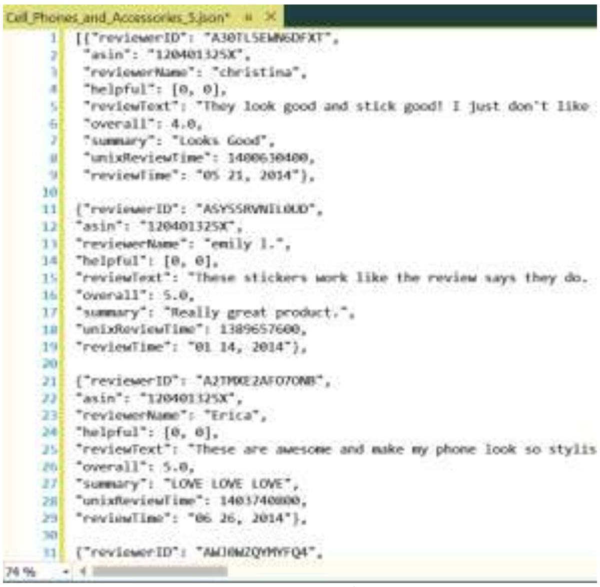
Step 3: Most frequent features of mobile like “screen”, “display”, ”battery”, “sim card”, “application”,”software” are selected man ally. This are called the target words about which mostly an opinion is given.

Step 4: After collecting the target words the next step is to collect all the opinion words related to this target words. We used 8 patterns to extract opinion target and its related opinion words using part of speech tagging. These patterns are taken from [1]

Table 3: Eight Patterns taken from paper [1]

Pattern	The first word	The second word	The third word
Pattern 1	JJ	NN/NNS	-
Pattern 2	JJ	NN/NNS	NN/NNS
Pattern 3	RB/RBR/RBS	JJ	-
Pattern 4	RB/RBR/RBS	JJ/RB/RBR/RBS	NN/NNS
Pattern 5	RB/RBR/RBS	VBN/VBD	-
Pattern 6	RB/RBR/RBS	RB/RBR/RBS	JJ
Pattern 7	VBN/VBD	NN/NNS	-
Pattern 8	VBN/VBD	RB/RBR/RBS	-

Step 5: The last step is to perform the classification using weights assigned to opinion words. The weights below 0.2 are for negative polarity, weights between 0.2 and 0.8 are for neutral polarity and above o.8 are positive polarity.



```
[{"reviewerID": "A3OTL5EWN6DFXT",
 "asin": "120401325X",
 "reviewerName": "christina",
 "helpful": [0, 0],
 "reviewText": "They look good and stick good! I just don't like
 "overall": 4.0,
 "summary": "Looks Good",
 "unixReviewTime": 1400630400,
 "reviewTime": "05 21, 2014"},

{"reviewerID": "ASYSSRVNI1OUD",
"asin": "120401325X",
"reviewerName": "emily l.",
"helpful": [0, 0],
"reviewText": "These stickers work like the review says they do.
"overall": 5.0,
"summary": "Really great product.",
"unixReviewTime": 1389657600,
"reviewTime": "01 14, 2014"},

{"reviewerID": "A2TMXE2AFO7ONB",
"asin": "120401325X",
"reviewerName": "Erica",
"helpful": [0, 0],
"reviewText": "These are awesome and make my phone look so stylis
"overall": 5.0,
"summary": "LOVE LOVE LOVE",
"unixReviewTime": 1403740800,
"reviewTime": "06 26, 2014"},

{"reviewerID": "AWJ0WZQYMYFQ4",
```

Fig 2: Screenshot of dataset

Some of the reviews from the datasets are:

1. *“It worked great for the first couple of weeks then it just stopped completely.. so basically a small waste of money.”*

   Opinion targets for this review are: {'waste', 'waste money', 'weeks'}

   Opinion words for this review are: {'basically small', 'completely', 'first couple', 'small'}

   Polarity of above Opinion Words : {-0.25, 0.1, 0.25, -0.25}

   Sum Of above polarity : 0.15

   Normalized polarity : 0.067

   Conclusion : Negative semantic

2. *“This is a fantastic case. Very stylish and protects my phone. Easy access to all buttons and features, without any loss of phone reception. But most importantly, it double power, just as promised. Great buy”*

   Opinion targets for this review are: {'access', 'access buttons', 'buy', 'case', 'power', 'protects', 'protects phone'}

   Opinion words for this review are: {'double', 'easy', 'fantastic', 'great', 'importantly double', 'stylish'}

   Polarity of above Opinion Words: {0.0, 0.43, 0.4, 0.8, 0.0, 0.5}

   Sum Of above polarity : 2.13

   Normalized polarity : 0.960

   Conclusion : Positive semantic

3. *“These make using the home button easy. My daughter and I both like them. I would purchase them again. Well worth the price.”*
   Opinion target for this review is: { *price}*

   Opinion word for this review is: {'well worth', 'worth'}

Polarity of above Opinion Words: { 0.3, 0.3}

Sum Of above polarity : 0.6

Normalized polarity : 0.270

Conclusion: Neutral semantic

4. Conclusion

The architecture of proposed method is depicted from the Fig 1. The proposed method can be divided into two main tasks. First is to define some important opinion target which have potential to get relevant opinion words and then extracting opinion target and their corresponding opinion words from the reviews and the second task is finding the polarity of the reviews. The former task uses part of speech tagging and 8 predefined patterns. WordNet can be used to figure out similar opinion target and corresponding opinion words from the review text. Additionally, we can increase opinion word size by adding similar words from WordNet. This is useful when we have small size of predefined opinion targets extracted from corpus or the opinion target chosen are not relevant to domain (i.e. when domain information is hidden). Finally, the classification of the review is given using weights assigned to opinion words. Our algorithm classifies the product based on features and helps the customer as well as manufacturer to gain knowledge about individual feature of the mobile phone. The customer as well as manufacturer can benefit from this. The customer gets to know the efficiency of the product and the manufacturer can further improve their product.

References

[1] Htay, s. S., & lynn, k. T. (2013). “Extracting product features and opinion words using pattern knowledge in customer reviews.” *The scientific world journal*, 2013.

[2] Anisha P Rodrigues, D. N. (2016). "Mining Online Product Reviews And Extracting Product Features Using Unsupervised Method”. IEEE.

[3] Balakrishnan Gokulakrishnan, P. P. (2012). "Opinion mining And Sentiment Analysis On A Twitter Data Stream*". The International Conference On Advances In Ict For Emerging Region,* IEEE.

[4] Cho, M. Y.-P. (May 31 – June 5, 2015). "Lcct: A Semi-Supervised Model For Sentiment Classification". *Association For Computational Linguistics.*

[5] . Hussam Hamdan, P. B. (August 23-24 2014). Supervised Methods For Aspect-Based Sentiment Analysis". *International Workshop On Semantic Evaluation*, Dublin, Ireland.

[6] Kang Liu, L. X. (March 2015). “Co-Extracting Opinion Targets And Opinion Words From Online Reviews Based On The Word Alignment Models". Ieee *Transactions On Knowledge And Data Engineering*, Vol. 27, No. 3.

[7] Nipuna Upeka Pannala, C. P. (2016). "Supervised Learning Based Approach To Aspect Based Sentiment Analysis”. IEEE *International Conference On Computer And Information Technology*.

[8] Nishantha Medagoda, S. S. (2013, December).” A Comparative Analysis Of Opinion Mining And Sentiment Classification In Non-English Languages”. In *Advances In Ict For Emerging Regions (Icter), 2013 International Conference* On Ieee, 144-148.

[9] Pranali Yenkar, D. S. (January 2013). "Opinion Mining Of The Movie Blogs Based On Supervised Learning Approach*". International Journal Of Advanced Research In Computer Science* Volume 4, No.1.

[10] Richa Sharma, S. N. (2013). "Supervised Opinion Mining Techniques: A Survey". *International Journal Of Information Andcomputation Technology.* Issn 0974-2239 Volume 3, Number 8, 737-742.

[11] Samha, A. K. (Jan - Feb 2016). "Aspect-Based Opinion Mining Using Dependency Relations" . *International Journal Of Computer Science Trends And Technology (Ijcst)* – Volume 4 Issue 1.

[12] Shitanshu Verma, P. B. (N.D.). "Incorporating Semantic Knowledge For Sentiment Analysis". Proceedings Of Icon-2008: *6th International Conference On Natural Language Processing.*

[13] . Subhabrata Mukherjee, P. B. (2012). "Feature Specific Sentiment Analysis For Product Reviews". *Computational Linguistics And Intelligent Text Processing*, 475-487.

[14] Toqir A. Rana, Y.-N. C. (2017.). "A Two-Fold Rule-Based Model For Aspect Extraction". Universiti Sains, Malaysia.

[15] Vibha Soni, M. R. (5-May 2014). "Unsupervised Opinion Mining From Text Reviews Using Sentiwordnet". *International Journal Of Computer Trends And Technology (Ijctt),* Volume 11 Number.

[16] Wang, S. Z. (August 2010). "Active Deep Networksfor Semi-Supervised Sentiment Classification". Coling 2010: Poster Volume, Pages 1515–1523, Beijing.

[17] A. V. Zadgaonkar , “Knowledge Base Population: A vey“ *,International Journal of Advanced Research in Computer Science and Software Engineering* , Volume 5, Issue 2, February 2015, pp. 732-735.

[18] A. V. Zadgaonkar , “An Overview of Entity Relation Extraction Techniques “ *,International Journal of Advanced Research in Computer Science and Software Engineering* , Volume 5, Issue 11, November 2015, pp. 266-269.

[19] Ashwini A. Shende, Avinash J. Agrawal& Dr. O. G. Kakde, “Domain Specific Named Entity Recognition Using Supervised Approach *“, International Journal of Computational Linguistics* (IJCL), Volume (3) : Issue (1) : 2012, pp. 66-78.

[20] A. V. Zadgaonkar , A. J. Agrawal , S. Aote, “Facets extraction-based approach for query recommendation using data mining approach”, *International Journal of Engineering & Technology*, 7 (1) (2018) 121-125, pp.121-125.

[21] Ashwini V. Zadgaonkar, "Natural Language Understanding Using Open Information Extraction Technique*", International Journal of Computer Sciences and Engineering,* Vol.6, Issue.1, pp.347-350, 2018.