

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

A Survey on Aspect Based Opinion Mining from Product Reviews

A G Dongre¹, Sushmit Dharurkar², Swanand Nagarkar³, Rahul Shukla⁴, Vivek Pandita⁵

Professor, Department of Computer Engineering, Pune Vidyarthi Griha's College of Engineering and Technology,
Pune, Maharashtra, India¹

U.G Students, Department of Computer Engineering, Pune Vidyarthi Griha's College of Engineering and Technology,
Pune, Maharashtra, India^{2, 3, 4, 5}

ABSTRACT: Opinion mining and sentiment analysis is the process of analysing the text about a topic written in a natural language and classify them as positive, negative or neutral based on the humans sentiments, emotions, opinions expressed in it. Most of the previous work is in the field of document or sentence level opinion mining. Currently the work is going on to implement this work using aspect level opinion mining. Aspect-level opinion mining is also well-known as phrase-level (or feature level or word level) opinion mining. This paper proposes an alternative model to implement aspect-level opinion mining. The model proposed is a syntactic unsupervised approach for explicit aspects for achieving aspect level opinion mining. Using this approach, we can extract most important aspects of the product along with the opinion words describing that aspect. The sheer volume of online reviews makes it difficult for a human to process and extract all meaningful information in order to make an educated and informed decision. This approach does not require any seed word or domain-specific knowledge of the product. The result of the proposed model will show the extracted aspects from the reviews, various opinion words describing that aspect and its rating. The rating of aspects will be done with the help of sentiment scores from the lexical resource, SentiWordNet.

KEYWORDS: Opinion Mining, Natural Language Processing.

I. INTRODUCTION

Opinion mining or sentiment analysis is the field of study that analyses peoples opinions, sentiments, attitudes, and emotions towards entities such as products, organizations, services, individuals, issues, events and topics and which comes under the area of natural language processing. It is basically the generation of subjective knowledge from source materials. We can say that opinion mining is a sub problem of traditional text classification problem because in traditional text classification the documents are classified based on the topics, e. g. , politics, sciences, and sports, but in opinion mining texts are classified based on the human sentiments, emotions or opinions. Opinion mining represents a large problem space. There are various tasks which comes under the category of opinion mining, e. g. , sentiment analysis, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. The term opinion mining is commonly used in academia and sentiment analysis is in industry. Due to the explosive growth of social media (e. g., reviews for various products on e-commerce websites and mobile applications, forum discussions on various issues or topics, blog's reviews, Twitter, comments, and postings in social network sites) on the Web, users now have many opportunities to express their opinions about a product or topic. Users express their opinion through the reviews. These reviews are used by the individuals and organizations for decision making. This trend makes attention of organizations businesses and researches around the world towards opinion mining area. This new trend provides a strong motivation to the researchers and also offers many challenging research problems. Opinion mining is a hard problem to be solved due to the highly unstructured nature of natural language and the difficulty of a machine to interpret the meaning of a sentence. But the reviews and usefulness of the opinion from the reviews is increasing day by day. For solving this problem a system must be made to understand and interpret the human emotions and feelings. Opinion mining and sentiment analysis are approaches used for implementing the same. Opinion mining can be done at three different levels, which are Document Level, Sentence Level and Aspect Feature Level. In

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

document level, the overall opinion about the document is finding out and classifies them as positive or negative. In sentence level, each sentence in the document is analysed for finding the fine grained opinions about different topics in a document. Finally classify the opinion expressed in a sentence as positive, negative, or neutral. The product and restaurant reviews are a mixture of positive and negative opinion about different aspects. It needs more fine- grained analysis of reviews to mine these mixed opinions, aspect level perform this task. Hence aspect based opinion mining is preferred in this work. This survey paper mainly focuses on the aspect based opinion mining. The core tasks in aspect based opinion mining is aspect identification, aspect based opinion word identification and its orientation detection. For example, consider a review of a mobile phone, "The performance of mobile is good but battery backup is low". First step is to identify the aspects, which are performance and battery, then find aspect related opinion word, which are good and low. Then detect its orientation, i.e. whether that opinion word expresses positive or negative opinion. By analysing the above example we get that performance has positive opinion and the battery has negative opinion. Current approaches for opinion mining, attempt to detect the overall polarity of a sentence, paragraph or text span regardless of the aspects mentioned in it. The work proposes a new syntactic based approach for aspect based opinion mining which uses syntactic dependency, aggregate score of opinion words, SentiWordNet and aspect table together for opinion mining process.

II. RELATED WORK

One of the earliest works to extract the aspects was done in [3], which was based on frequent nouns and noun phrases. They used association rule mining combined with pruning strategies to find the candidates of features. They assumed that the product features are nouns or noun phrases. They first performed Part- of- Speech (POS) parsing. A tag of POS was then given to each word. A transaction was built from the noun words of each sentence. All transactions were fed into the association rule mining algorithm to find the frequent item sets. The returned frequent item sets were used to identify product features. In [1], authors proposed aspect based sentiment analysis using support vector machine classifier. They propose a different approach which combines the use of dependency parsing, co-reference resolution and SentiWordNet together for the sentiment analysis. The training of the system is done using the support vector machine. They did this work in a single domain and tests are done only for reviews about digital cameras. They consider only explicit aspects. The average accuracy is 77. 98%. More training data is needed for this approach and it may fail when training data are insufficient. The proposed method does not use any training data. The use of emoticons in sentiment analysis is explained in [4], the results show that accuracy is improved. The emoticons are less in online reviews like restaurant reviews; hence it is not included in this study. In [2] a new feature based heuristic was used for aspect level sentiment classification of movie reviews. The authors proposed AAAC algorithm based on SentiWordNet to find the sentence level aspect score. This scheme calculates (Adverb + Adjective) and (Adverb + Verb) combined score using SentiWordNet. They used SentiWordNet scheme to compute the document-level sentiment for each movie reviewed and compared the results with results obtained using Alchemy API. Accuracy of 78.7% obtained over movie review dataset. In this method, features are extracted by using POS Tagger. Firstly, aspect indicating term in a sentence is located, and then searches up to 5-gram forward or backward for the occurrence of features in it. Directly modified feature term of an aspect cannot be determined in it. For resolving it, the proposed method use dependency grammar for feature extraction.

III. DESIGN OF PROPOSED WORK

The proposed design for opinion mining process can be understood from the given proposed architecture in Fig.1. The following section describes the proposed method for the proposed opinion mining architecture.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

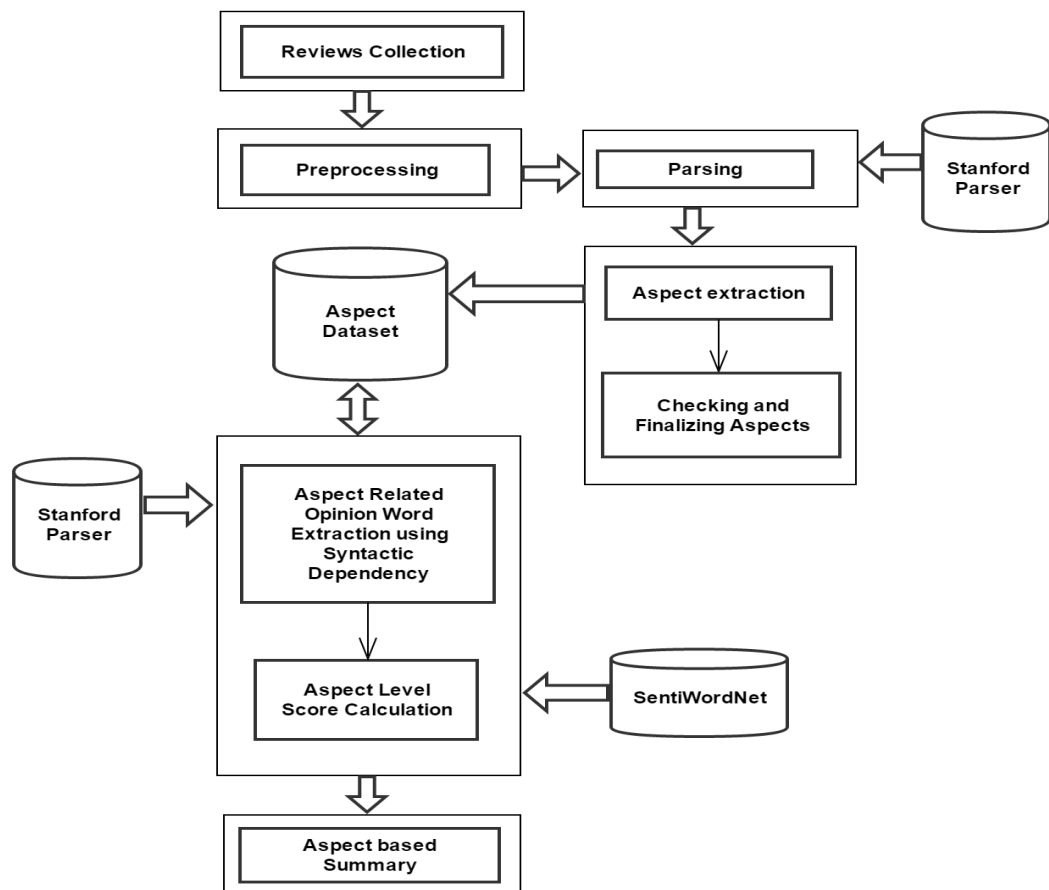


Fig.1. Proposed Architecture for Opinion Mining

IV. PROPOSED METHOD FOR IMPLEMENTATION

WORKING:

Through our model, we propose the following tasks which includes aspect identification (explicit aspects), aspect based opinion word identification and its orientation detection. The proposed model should work irrespective of product. We do not provide any explicit data or domain specific information to the model related to the product. This will involve, gathering of the opinion corpus or reviews from various available research datasets or e-commerce website manually. The next step will involve pre-processing the data to make it suitable for aspect extraction. Pre-processing is done in order to improve the accuracy and also avoid the unnecessary processing overhead of opinion mining process, hence the collected reviews should be pre-processed. It mostly involves stop words removal and eliminating unnecessary non alphabetic characters and smiley. Aspect or feature extraction involves identification of important aspects or feature of the particular product from the user's reviews [frequent item set mining]. A count is set for the extraction of various aspects, i.e., the various aspects generated can only be considered for the further processing if it is fulfilling the criteria of minimum count of occurrence of that aspect in the data set. After generating the final aspect list, we will cluster the similar aspects (based on word-net similarity) and maintain a list of the aspects. The second half of work will focus on rating the product aspects. This will be done by identification of opinion words (adjectives and verbs) from the subjective sentences. The subjective sentence will comprise of at least one of the aspects/feature. The identification of opinion words for a particular aspect is a very difficult task and will involve either

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

of syntactic-dependency paths (unsupervised learning) or machine learning algorithms (semi-supervised learning). Finally each aspect is given a polarity score. This is done by aggregating the sentiment rating of the opinion words of a particular aspect. In order to obtain the sentiment score we will use the external sentiment lexicon SentiWordNet, an extension of WordNet. Finally we will output the various aspects with their sentiment scores. However the result will not be completely correct due to ambiguity of word polarity and inclusion of objective statement. There are limitations to the state of the art dependency parser and used algorithm. Most of the errors will occur due to inability to deal with the colloquial language used in the review.

DATASET:

The review dataset is collected from various e-commerce websites like Amazon, Flipkart, etc. The review dataset is passed as a input in the form of a text file. The text file will comprise of all the reviews in unstructured form.

PERFORMANCE EVALUATION:

The performance measures used for the evaluation purpose were Accuracy, Precision and Recall. The formulas are given below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

In the above formula, TP is the true positive instances. TN is the true negative instances. FP is the false positive instances and FN is the false negative instances. These formulas will help in getting the performance evaluation of our implementation.

V. CONCLUSION

In this model, we have presented a system for summarizing sentiment from online reviews about aspects of product. This model actually represents a generic tool which does not require any seed word or pre requisite knowledge about the product or service. We can say that using this proposed model of aspect level opinion mining, it can help people in taking the right decision about the quality of product or service. The resulting system is highly precise for frequently queried product, yet also sufficiently general to produce quality summaries for all product types. This model uses a new different syntactic approach to aspect level opinion mining, which use aspect dictionary, SentiWordNet, Dependency parsing, for opinion mining process with automatic acquisition of aspects. It is a syntactic based approach hence there is no need of any training data. In the proposed system aspect matched opinion words are extracted using dependency parsing on subjective sentences. Polarity of opinions i.e. positivity or negativity of an aspect is find out using SentiWordNet, based on which a rating can be produced, which shows how good or bad each aspect of the product is from total reviews.

REFERENCES

- [1] R. Varghese and M. Jayasree, "Aspect based sentiment analysis using support vector machine classifier," in Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on. IEEE, 2013, pp. 1581-1586.
- [2] V. Singh, R. Piriyani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," in Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on. IEEE, 2013, pp. 712-717.
- [3] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004, pp. 168-177.
- [4] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de long, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in Proceedings of the 28th Annual ACM Symposium on Applied Computing. ACM, 2013, pp. 703-710.