

Context Based Syntactic Opinion Mining

Nandini S.

Assistant Professor

Dept. of Computer Science and Engineering

GITAM University

Bengaluru Campus, Karnataka, India

Abstract— E-commerce is evolving at such a rapid pace that new doors have been opened for users to have many opportunities to express their opinions about the product. The purpose of this project "Context Based Syntactic Opinion Mining" is to provide an effective way to view the opinions of the customers expressed in the form of customer reviews. This paper focus on aspect level opinion mining and proposes a new syntactic based approach using Natural Language Tool Kit (NLTK) and SentiWordNet. The objective of this paper is to summarize reviews of the product based on features or Aspects and classify as positive or negative opinion about a feature by assigning a score. This paper, mainly concentrates on reviews expressed about Mobile devices to extract the aspects but also applicable to other products. By analyzing the opinions of users about the features of a product, visual summary of the product can be made by plotting a graph based on score. It can also be extended to compare two mobile phone features and give an opportunity for the user to select the best among two products. This enables a user to have better understanding of the product which otherwise involves reading through long textual reviews to form a mental picture of the strengths and weaknesses of the product. This will be very useful for the customers to know about the features of product before making a buying decision. This project not only helps individuals in buying a product but also helps the organization to know how customers' perceive their product.

Keywords— *Opinion Mining, Aspect-level Opinion Mining, SentiWordNet*

I. INTRODUCTION

Opinion mining is the field of study that analyse individuals' opinions, views, sentiment, emotions, attitude towards the product, events, organizations. It is a challenging task and hence represents a large problem space. It is also known as Sentiment analysis, Opinion extraction, review mining. However, Sentiment analysis or Opinion mining is commonly known. Opinions are key influencers for both organizations and customers. In real world scenario, organizations want to find opinions about their product so that they can work on to increase the quality of product. Also, individual customers want to know about the product before purchasing by the existing users.

Opinion mining can be classified into three levels: Document level, Sentence Level and Aspect/Feature Level. In Document Level, the complete document is considered to find the opinions and then classify them as either positive or negative. For example: The document level classifies reviews into positive and negative opinions for a given set of movie reviews. This is similar to traditional topic-based classification, which classifies the document into different classes like science, sports, politics etc., In Sentence-Level, every sentence in a document is analyzed to find the opinions

and to categorize the opinion expressed in a sentence as positive or negative. This type of classification, classifies each sentence as a fact or opinion. Both Document level and Sentence level will not convey information about what the opinion holder likes or dislikes. In **Feature/Aspect Level**, the feature is identified for a product and based on these features the orientation of the feature can be made (i.e., positive or negative). Because opinion holder expresses both positive and negative views on the product. These aspects are mainly considered to as nouns or noun phrases. For example: "*The resolution of a camera is good*". Here the 'aspect' considered is 'camera' and 'good' refers to 'positive opinion' of feature. Mining opinions at document or sentence-level is not beneficiary in many cases. Because the information provided is not sufficient for decision making by a customer. Hence, the proposed project focuses on a finer level of granularity for opinion mining based on features.

A. Problem Statement

"A novel approach to context based Syntactic Opinion Mining of consumer reviews on mobile phones posted on E-commerce websites."

The problem statement deals with the reviews posted on existing users of E-commerce websites. The proposed project extracts the text reviews from the Urls and segregate the sentences based on features of Mobile phones using Natural Language Toolkit. By using SentiWordNet these sentences are assigned scores and hence positive opinion and negative opinion are analysed. Based on the these scores, Graphical representation of two products are plotting and hence consumer gets an idea of two product features without no longer reading of text reviews.

II. RELATED WORK

Opinion mining can be of manual or automation in-order to give score orientation i.e., positive or negative. Automation can be done using SentiWordNet with numerical scores indicating positive and negative sentiment information. Aspect identification in this paper is done by manual process by selecting six features of mobile phone.

In [1], Authors collected Restaurant review dataset and was tagged manually. It focused on Feature level opinion mining. This paper uses Part-of-Speech (POS) tagger to extract the features. The reviews which are collected are stored in a database, for opinion mining process these reviews are used later. Aspects or Features are identified through training process and these aspects can be a phrase or a single word. Pre-processing is made to remove unnecessary words. From

these sentences extraction of adjectives and adverbs are made in order to find the polarity of a word. By using SentiWordNet scores are assigned to the opinion words. Visualization is done based on polarity calculation of aspects by using SentiWordNet for graphical representation.

In [2], focused on aspect-based opinion mining. Here, the authors named as phrase-level opinion mining. It focuses on number of positive and negative opinions of each aspect in an online review. It extracts the online review and splits into individual sentences. For data Pre-processing, each review sentence is given as an input. Next, the aspects are extracted and the Pre-processing step includes Stop Word removal, Stemming and POS-tagging.

The main task is to identify the number of positive and negative opinions of an extracted aspects. Both sentence and aspect orientations are implemented by using Naive Bayesian algorithm which considers the Supervised term counting. By using Naive Bayesian Classifier, the probabilities of positive and negative counts are found and from this the summary of the product can be made which helps in decision making before purchasing a product. There is no visual summary of extracted aspects by plotting graphs. Hence the entire review has to be studied in order to review the product summary.

In [3], the authors focused as feature based level mainly on Chinese reviews. The authors extracts features and their corresponding opinions by considering the adjectives. Because these adjectives can modify whole sentence or feature of a sentence. Next, Pre-processing is made by eliminating noise. It is an important step. The noise is removed by considering the low frequency words. These low frequency words are removed. Later, by choosing high confidence of opinions, filter the opinions and the corresponding feature with low frequency. The low confidence opinion is then checked with the corresponding feature, if it is low confidence then remove and again re-calculate the co-occurrence matrix. There is no visual representation for reviews based on features.

In [4], the authors proposed based on phases. Phase 1 is a pre-processing phase which is necessary for detection of implicit feature and questioned based aspects. *Wordtag* tool is used for tokenising and for POS tagging. In Phase 2 reviews are identified and orientation of reviews are made by using SentiWordNet tool. In Phase 3 summarization of reviews are made. There is no graphical representation of reviews and the dataset considered is very small in size with only 40 reviews.

In [5], Proposed paper is about extracting the text data and to overcome the errors usually made by Humans. The errors can be corrected by the methodology using Artificial Neural Network (ANN). This neural network plays a important role in correcting the errors and developing a training data for further analysis on this unstructured text data. Natural Language Processing (NLP) helps in eliminating the majority of these errors made by humans. As because NLP enables computers to derive the meaning from humans. This system is well suited for social Networking trends. The un-structured data such as images, videos, audios are also challenging tasks which are not considered by the authors.

III. DESIGN OF PROPOSED TECHNIQUE

The proposed for opinion mining process in given the Figure 1. The following section describes the proposed opinion mining model with each subsequent step in detail.

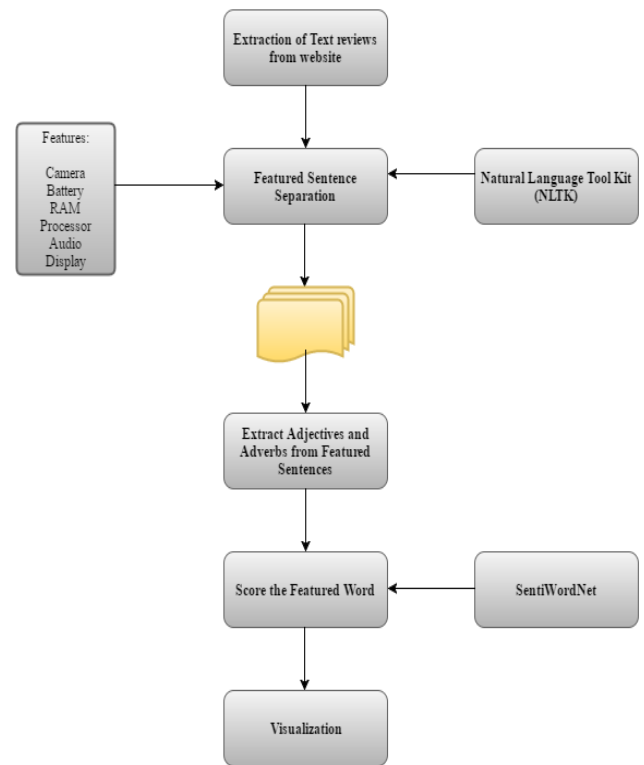


Figure 1: Proposed design for Opinion Mining

The proposed work is based on the mobile reviews which are extracted from Amazon website. The six main features of mobile is considered: Camera, Battery, RAM, Processor, Display, Audio.

IV. PROPOSED METHOD

A. Extraction of Text Reviews from Website

The reviews are extracted from a website by assigning URLs as an input. The reviews contain only textual sentences and eliminates the images which are found in the website. These text reviews are redirected to separate text document.

B. Featured Sentence Separation

From the text file extract the sentences which contain only features using Natural Language Toolkit (NLTK). This step eliminates the unwanted sentences from the review. This step is also known as Pre-Processing which improves the accuracy. These featured sentences are stored in a separate text file. For each feature separate text file is maintained.

Extraction of Adjectives and Adverbs from Featured Sentences

In Opinion Mining, opinion are mainly considered. Opinion words are the words which expresses the opinions towards the aspect/feature. Hence, these Opinion words has to be identified. The Opinion words considered in this paper are: Adjectives and Adverbs. In English POS, adjectives are considered as the describing words. Adverbs are the words which modify the adjectives.

The Adjectives and Adverbs are extracted from the reviews using NLTK. In NLTK JJ, JJS, JJR, RB, RBR, RBS represents adjectives, adjective superlative, adjective comparative, adverb, adverb comparative, adverb superlative respectively. These words are extracted.

D. Score the Featured Word

The positive and negative opinion words are scored using SentiWordNet. SentiWordNet is a dictionary of words which consists of words with their scores.

The dataset is maintained for the words with their scores. The words from reviews are compared with this dataset. Later, the count of these words are considered which is necessary for plotting a graph.

E. Visualization

By using the count, graph is plotted with Features on X-axis and scores on Y-axis. By analysing this graph, the user can make a firm decision before purchasing a product. It is also possible to compare two mobile products so that users can compare two phones features.

Below Figure 2 represents the Algorithm of the proposed system

```

Step 1 : Retrieve all text reviews from the given Urls and redirect to text file
Step 2 : Extract the featured sentences from the text and redirect to another text file
Step 3 : Score Calculation-
    i. Extraction of adjectives and adverbs from the above text document
    ii. rating = 0
        with open ('result.txt', 'rt') as f:
            reader = csv.reader(f, delimiter = ", ")
            for row in reader :
                if i == row [0]:
                    print i , row [1]
                if row [1] == 'pos' :
                    rating = rating + 1
                elif row [1] == 'neg' :
                    rating = rating - 1
            print rating
Step 4 : Graphical Representation
    
```

- Import the score calculation python files
- Plot bar graph of two phones with X-axis consisting of features and Y-axis consisting of scores.

Figure 2: Proposed Algorithm

V. EXPERIMENTAL RESULTS

The visualisation of large number of reviews by graphical representation makes the one simpler to analyse the reviews given by existing users. Hence, the user no need to depend on relatives or family members opinions. He/she can directly make a decision by analysing a graph w.r.t features and scores.

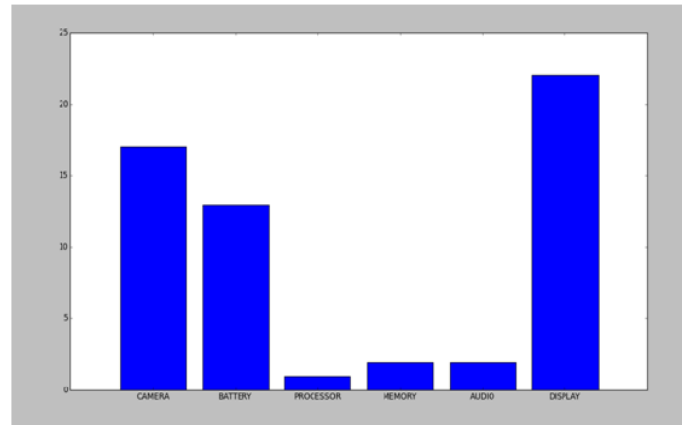


Figure 3: Graphical Representation of a Single Phone

Below Figure 3 depicts the graphical representation of a single phone. The proposed work can also be extended to compare two products based on features.

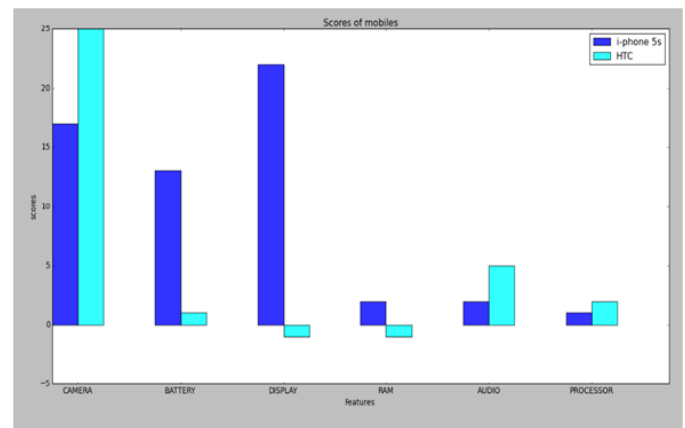


Figure 4: Comparison of two phones

From the above two figures, it is easy for one to analyse the product by comparing two phones and can make a firm decision

CONCLUSION AND FUTURE WORK

Due to explosive usage of internet and social media, one can feel free to express their opinions and views towards the product. This helps to know the product with its merits and demerits posted on website.

The proposed work, "Context Based Syntactic Opinion Mining" concentrates on Feature level of Opinion Mining. It is more useful, since it helps the users in making firm decision before purchasing the product. It also provide the graphical representation which is a vital element of the proposed work based on features and scores for the features. It also helps the users to compare two products of phones and make a best decision out of it. The proposed work also helps the organizations and business as they can retrieve the opinions for their product or service from the users. This helps to improve the quality of product or service to meet customer requirement and also to satisfy the customer. But the application must be coded in a language which supports Natural language Processing (NLP). Extraction of reviews for different products simultaneously is an overhead associated with the proposed work. The featured-based proposed work

concentrates on only six features of mobile phone with 200 reviews and comparison of products are made through these dataset.

Since in this fast growing world, there is a need of such visualized opinion mining which helps to know about the strength and weakness of a product rather than reading whole textual reviews. It helps the users in saving time and energy to make a decision before purchasing.

In the proposed work, we can retrieve text reviews by the given number of Urls, it can be extended to retrieve all the text reviews by giving the base Url loop until the last Url of the website. It can also be enhanced by taking the consideration of spelling mistakes made by users while writing the reviews. The proposed work only concentrates on the explicit sentences of reviews, it can be further enhanced for implicit sentences i.e., the sentence which contains 'not bad', 'not good', 'camera is good but phone is expensive' etc.

REFERENCES

- [1] Shibily Joseph, Chinsha T C, "A Syntactic Approach for Aspect Based Opinion Mining" Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)
- [2] C.S.Kanimozhi Selvi, A.Jeyapriya, "Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm" IEEE Sponsored 2nd International Conference On Electronics And Communication Systems(ICECS 2015)
- [3] Lizhen Liu, Zhixin Lv, Hanshi Wang, "Opinion Mining Based on Feature-Level" 2012 5th International Congress on Image and Signal Processing (CISP 2012)
- [4] Jibrán Mir and Muhammad Usman, "An Effective Model for Aspect Based Opinion Mining for Social Reviews" Published in The Tenth International Conference on Digital Information Management (ICDIM 2015)
- [5] ShivKumar Goel and Anil Kumar Bhand, "Text Analytics on Un-structured Text Data using Artificial Neural Network (ANN)" published in International Journal of Science and Research (IJSR) ISSN, Volume 4 Issue 6, June 2015
- [6] Haseena Rahmath P, "Opinion Mining and Sentiment Analysis-Challenges and Applications" published in International Journal of Application or Innovation in Engineering & Management (IJAEM), Volume 3, Issue 5, May 2014
- [7] Asmitha Dhokrat, Sunil Khillare, C. Namrata Mahender, "Review on Techniques and Tools used for Opinion Mining" published in International Journal of Computer Applications Technology and Research Volume 4– Issue 6, 419 - 424, 2015
- [8] Sneha, B. Akshatha Bhat and Preetham Kumar, "Weighted summarization of Student Feedback using Sentiment Analysis" published in International Journal of Computer Applications (0975 8887) Volume 97 - No. 3, July 2014
- [9] Vignesh S, Srinath N K and Sandeep R V, "Data Analytics on E-Commerce Transaction Logs for Payment Management" published in International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 10, October-2015
- [10] Bing Liu white paper entitled as "Opinion Mining"
- [11] Felipe Jordao Almeida Prado Mattosinho, "Mining Product Opinions and Reviews on the Web" published in 2010