



An Algorithm to Quantify Sentiments of Product Reviews for a Given Feature using Text Mining

Smita Suresh Daniel¹, Dr. Neelam Sahu², Dr Ani Thomas³
 PhD Scholar¹, Associate Professor², Professor³
 Department of IT & Computer Applications^{1,2}, Department of IT³
 Dr C.V Raman University, Bilaspur, Kota C.G, India^{1,2}
 Bhilai Institute of Technology, Durg, C.G, India³

Abstract:

Quantifying sentiments of customer reviews of E-Commerce sites based on given features provide a convenient way to analyse large unclassified online text. Every product has many features but if reviews on specific features are to be analyzed then it becomes easier for manufacturers and producers to take a decision. In this paper we propose a system that extracts reviews based on given features and quantifies the sentiments in it.

Keywords: Quantify Sentiments – Product review, Quantification, Text Mining.

I. INTRODUCTION

Sentiment Analysis, studies people's sentiments towards certain entities. It is the process of computationally identifying, categorizing and quantifying opinion expressed in a piece of text, in order to determine the measure of the writer's attitude towards a particular product. The feedback of consumers is important for business, because it enables to plan marketing strategies based on the consumers' reception so it is necessary to understand their sentiments about its distinctive features as overall sentiment of a product is a collective one and does not provide information about the goodness or badness of a particular feature. Our model aims to quantify the sentiments of the text about a particular product feature provided and pictorially represent the overall sentiments. It mainly concentrates on finding out the sentiments of the various features of a product separately and quantifies them. For example, if the manufacturer wants to analyze the fabric feature of their product bags then all reviews related to the feature fabric is extracted into a file and the sentiments are classified as positive or negative at sentence level first and then at document level. Finally quantify the sentiments at entity level and graphically represents it. A review like "The fabric of the bag is very good.", can be classified under positive sentiments.

II. THE PROPOSED SYSTEM

Within Sentiment Analysis, quantification plays a major role, since in many applications we are interested in estimating sentiment not at the individual level, but at the aggregate level as it is the sentiment of the crowd, and how it is distributed. It attempts to make sentiments definable and measurable. Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity, positive or negative (or neutral), We calculate the sentiment score at three levels, First each sentence of a review is categorized at sentence level, then at document level and finally at the entity level. The document level concerns whether a document, as a whole, expresses negative or positive sentiment, The entity level then

targets the exact measure of people's likeness or dislikeness about a product feature.

III. THE BASIC STEPS

1. Preprocessing Database
2. Accept Feature for Analysis
3. Extract Specific Reviews
4. POS Tagging
5. Sentiment Sentence Extraction
6. Sentiment Word and Phrase Identification
7. Calculate Total Positives and Negatives in each sentence
8. Sentiment Score Computation at document level
9. Sentiment Score Computation at Entity level
10. Result Interpretation

1) PRE PROCESSING REVIEW DATABASE

The database has to be cleaned so that we get proper text review on which analysis can be performed. Here we are to remove all unwanted characters like! () * and emoticons etc which can meddle with our results. It is suggested by Pang and Lee that all objective content should be removed for sentiment analysis.

2) EXTRACT FEATURED REVIEWS ONLY

Each review is parsed for the given feature. Only those reviews are collected whose words most closely match with the features provided. This is done by finding the feature words or its synonyms in the review. The word or its synonyms can be searched for in a sentence in a review and if found the review can be extracted else rejected. Each feature review will be stored in separate database file.

3) POS TAGGING

The subjective content consists of all sentiment sentences. A sentiment sentence is the one that contains, at least, one positive or negative word. All of the sentences were firstly

tokenized into separated English words. Every word of a sentence has its syntactic role that defines how the word is used. The syntactic roles are also known as the parts of speech. There are 8 parts of speech in English: the verb, the noun, the pronoun, the adjective, the adverb, the preposition, the conjunction, and the interjection. In natural language processing, part-of-speech (POS) taggers have been developed to classify words based on their parts of speech. For sentiment analysis, a POS tagger is very useful because of the following two reasons: 1) Words like nouns and pronouns usually do not contain any sentiment. It is able to filter out such words with the help of a POS tagger; 2) A POS tagger can also be used to distinguish words that can be used in different parts of speech. For instance, as a verb, "enhanced" may conduct different amount of sentiment as being of an adjective.

4) SENTIMENT PHARSE EXTRACTION AT SENTENCE LEVEL

The sentiment word or phrase has to pass through a polarizer module which finds its polarity .It has the following procedures .

Negation Phrases identification

Words such as adjectives and verbs are able to convey opposite sentiment with the help of negative prefixes. For instance, consider the following sentence that was found in an electronic device's review: "The built in speaker also has its uses but so far nothing revolutionary." The word, "revolutionary" is a positive word according to the list. However, the phrase "nothing revolutionary" gives more or less negative feelings. Therefore, it is crucial to identify such phrases. In this work, there are two types of phrases have been identified, namely negation-of-adjective (NOA) and negation-of-verb (NOV). Most common negative prefixes such as not, no, or nothing are treated as adverbs by the POS tagger.

5) SENTIMENT QUANTIFICATION

A Sentiment word or a Phrase consists of a positive (negative) word and its part-of-speech tag. All such words and phrases are collected as many times it occurs in a review. As in a Bag of words model a list of positive words and a list of negative words, respectively, based on customer reviews are created and both lists included some misspelled words that are frequently present in review content. Based on this Sentiment categorization is done. Now each words and phrases that contain sentiment information should be identified in the bag of words before the classification. If it is found in the list of positive word a score of +1 is provided else if found in the list of negative word a score of -1 is provided thus categorizing the text into one specific sentiment polarity, positive or negative (or neutral). Finally, the sentiment score information for positive and negative words can be calculated statistically for a review at the sentence level and then at document level. Similarly each review has to be polarized resulting in entity level polarization. The entity level concerns whether a document, as a whole, expresses negative or positive sentiment. Finally the quantification targets to find the measure of how much people like or dislike the particular feature from their opinions. All processed output is stored in a file for further use. The file is then loaded into the **database** for Summarization.

Summarization of sentiments is generated as charts finally.

IV. THE SYSTEM FLOW CHART

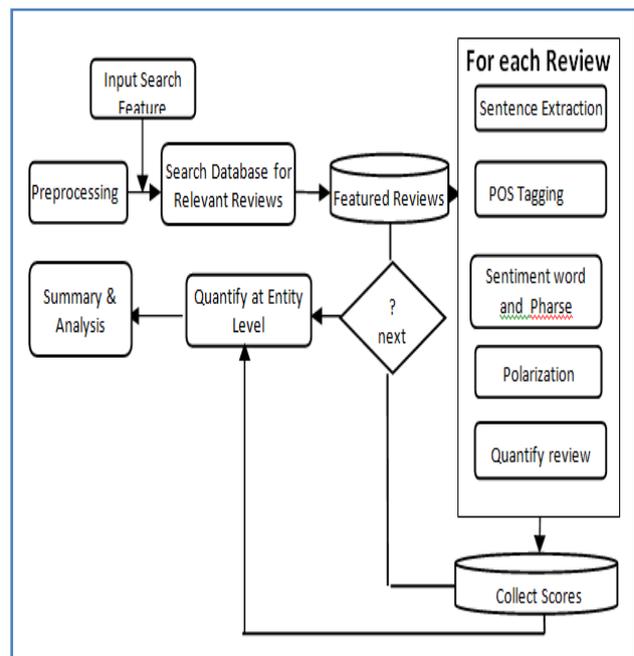


Figure.1. System flow chart

V. THE PROPOSED ALGORITHM

1. Product Reviews will be collected from commercial websites. These online reviews posted by the reviewers includes the following information: 1) reviewer ID; 2) product ID; 3) rating; 4) time of the review; 5) helpfulness 6) review text. Extract all required details from it.
2. **Cleaning the data** All special characters (such as: “./, ’#\$%^&-), binary, emoticon must be removed in order to retrieve best results. This also saves our review processing time.
3. Read the review file for **Processing**, for each review and the following is done.
 - i. A review is tested for occurrence of feature words or its synonyms , i.e. if the review words are found in the review, then it is classified and stored in a file. .If the review fails for this test then it is rejected.
 - ii. For each sentence of the review. We extract its sentiment words and phrases using POS tagging and rule based extraction (using regular expressions).
 - iii. These are then sent to a polarizer module that return 1 if the sentiment is positive else -1 which means the sentiment is negative.
4. Sum of positives and negatives are calculated for all sentences in a given review and its mean is taken.
 - 4 The sum of the sentiments of all the reviews are collected .A new final file is generated with the classification and sentiment of the feature phrases. This file is then loaded into the database for creating the visualizations by querying data from the database.
- 5 Summaries can be textual or non-textual and they can make use of graphs and/or charts. Obviously, when we have to summarize a lot of opinions, we have to represent only the information which is relevant for our goal.

VI. QUANTIFICATION FOR SENTIMENT ANALYSIS

Before Quantifying Sentiments there are many problems which have to be handled ,some of which are stated below ,

- **Sentiment Detection**, In the Bag of Words model, the list of positive and negative words should be collected carefully after manually going through some of the reviews for opinion-bearing words in the sentences.
- **Score computation on a 5 -point scale**, We can classify the sentiment words in a multipoint scale according to the intensity of the sentiment word. The list of words in the Bag of Model can be categorized as positive, very positive, neutral, negative and very negative.

VII. REFERENCES

[1]. Aashutosh Bhatt and, Ankit Patel. and Harsh Cheda ., "Amazon Review Classification and Sentiment Analysis ", IJCSIT , 2015, Vol 6(6).

[2]. Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04.. Association for Computational Linguistics, Stroudsburg, PA, USA.

[3]. Minqing Hu and Bing Liu, "Mining and summarizing customer reviews", KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004

[4].Bo Pang¹ and Lillian Lee² "Opinion Mining and Sentiment Analysis