

Sentiment Analysis using Opinion Mining

Benito Alvares

Student

Xavier Institute of Engineering
Mahim, Mumbai, India

Nishant Thakur

Student

Xavier Institute of Engineering
Mahim, Mumbai, India

Siddhi Patil

Student

Xavier Institute of Engineering
Mahim, Mumbai, India

Daniel Fernandes

Student

Xavier Institute of Engineering
Mahim, Mumbai, India

Kavita Jain

Associate Professor

Dept. of Computer Engineering
Xavier Institute of Engineering
Mahim, Mumbai, India

Abstract— Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. The idea of this paper is to “process a set of user reviews for a given product, generating a summarization (quality, features) and aggregating of user opinion”. The existing systems have greater emphasis on the product in particular rather than what the user is saying about it. Our research aims to shift focus on the user’s opinion after semantically analyzing and mining the data to find the hidden sentiment it.

Keywords—Sentiment Analysis, Opinion Mining, Review Crawling, Text Analysis, POS Tagging, Classification, Summarization.

I. INTRODUCTION

In order for the consumers to make a better choice, they ought to have access to reviews & experiences from other consumers who have made similar choices. This will help them not only to avoid mistakes that other consumers made but also help clear confusion, if any, about the product or service.

In the present system, most of the leading e-commerce websites, tend to focus on the product or its features to quite a large extent. While it’s not a crime to do so, they give very little emphasis to what other people are saying about the product. In this paper we are aiming to perform opinion mining using sentiment analysis.

Our paper aims to help consumers in this decision making process during a purchase. This paper focuses on semantically analyzing & evaluating product reviews as they are, minimizing human bias or preference. This paper does not compare products feature-wise, rather it tries to detect hidden sentiments in a review and along with the product’s features, and it gives an overall rating.

II. RELATED WORK

Our work is closely related to Hu & Liu’s work in [2] on text mining & summarization. Their research focused on studying the problem of generating feature-based summaries of customer reviews of products sold online. The summarization here is different from traditional text summarization tasks because our focus here to classify opinions & features in each sentence in a review and not just the entire review as a whole. The extraction of the feelings from the review is done by Part-of-Speech Tagging. And then the system would comprehend the sentiments by recognizing the data sets from the database. For the prototype, we have used SentiWordNet[1] for the opinion mining.

III. IMPLEMENTATION METHODOLOGY

There are basically four major components in the implementation of the proposed theory.

First is the user and review databases which will store all the reviews from users and web crawler used on e-commerce sites.

Second part is the POS tagging and feature pruning. Here all the words are tagged into various part of speech. The part of reviews containing insignificant features are removed by feature pruning. Now what remains are the sentiments along with frequent features.

Next, we extract the opinion from the given review using the Opinion Word Extraction. Then the orientation of the tagged opinion is found by Opinion Sentence Orientation Identification.

Finally, we summarize the result which provides a non-biased overall rating. This summary is generated using clustering algorithms.

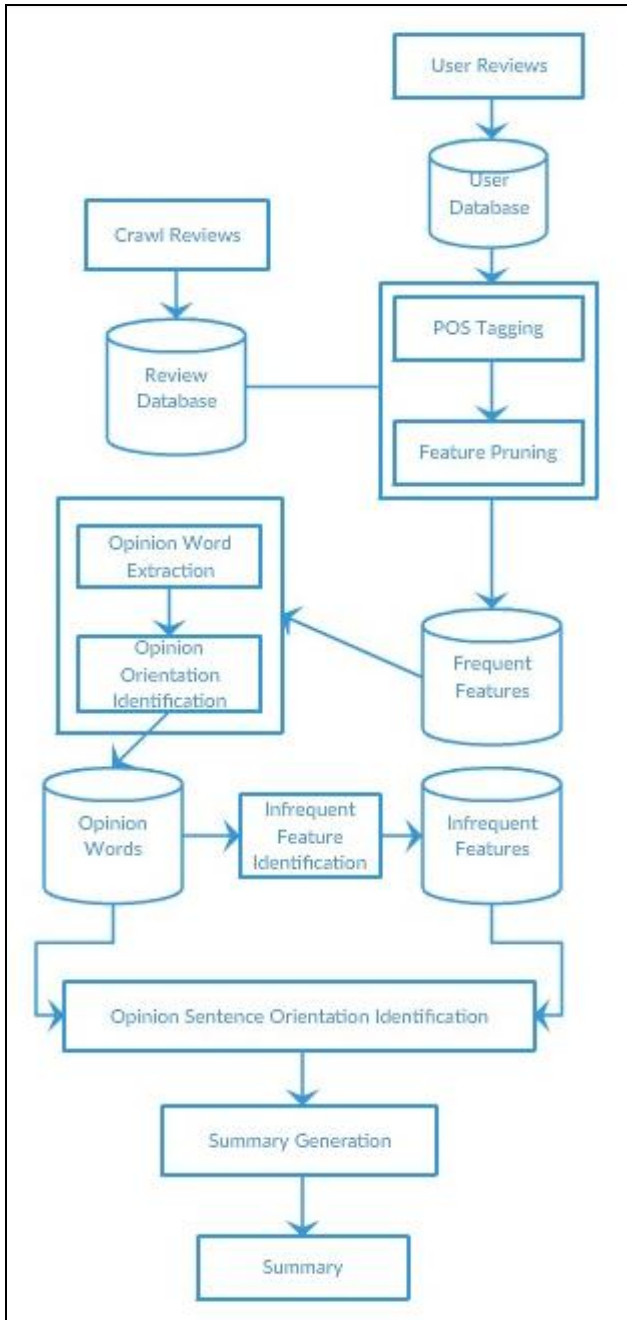


Figure 1: Flow Chart of Proposed System.

IV. WORKING OF THE PROPOSED SYSTEM

Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. Our proposed system focuses singularly on reviewing movies and is mainly performed in four steps:

A. Review Crawling

We need to have a database or dataset of reviews for the process of opinion mining. This dataset can be used as raw material to get set of features. We use a simple web crawler to gather reviews from the web from popular ecommerce sites and store them into a database.

B. Feature Extraction

The distinction between the terms *aspects* and *features* is an important concept that needs to be understood. *Features* are basically the characteristics that a product or service possesses or does. Whereas *Aspects* on the other hand are the important features rated by reviewers. A product may have many features but not all aspects always appeal to the user. This task of extraction of important features are given below.

- First we must split each review into individual sentences and then analyze each sentence.
- Part-Of-Speech tagging (POS tagging or POST) is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. It does the basic identification of words as nouns, verbs, adjectives, adverbs, etc.
- Using Part-of-Speech Tagger, tag each part of the sentence. We only need to extract the sentences with tagged words.
- Identify the most frequently extracted words and create a table along with its synonyms.
- This table is the aspect table for that particular product. It is the table containing important aspects and words that have same meaning as that of the aspects included in it.
- We have used SentiWordNet datasets along with the crawled datasets to create aspect tables for smartphone review purpose.

C. Opinion Word Extraction.

For the purpose of identifying the orientation of word opinion and ultimately the sentence opinion, we build on the algorithm proposed by Hu and Liu in [2]. We take the adjective seed list and a set of extracted opinion words whose orientations need to be determined.

Algorithm 1: SentenceOrientation ()

1. Procedure SentenceOrientation()
2. begin
3. for each opinion sentence s_i
4. begin
5. orientation = 0;
6. for each opinion word op in s_i
7. orientation += wordOrientation(op, s_i);
8. /*Positive = 1, Negative = -1, Neutral = 0*/
9. if (orientation > 0) s_i 's orientation = Positive;
10. else if (orientation < 0) s_i 's orientation = Negative;
11. else {
12. for each feature f in s_i
13. orientation +=
14. wordOrientation(f 's effective opinion, s_i);
15. if (orientation > 0)
16. s_i 's orientation = Positive;
17. else if (orientation < 0)
18. s_i 's orientation = Negative;
19. else s_i 's orientation = s_{i-1} 's orientation;
20. }

21. endfor;
22. end

Algorithm 2: wordOrientation ()

1. Procedure wordOrientation(word, sentence)
2. begin
3. orientation = orientation of word in seed_list;
4. If (there is NEGATION_WORD appears closely around word in sentence)
5. orientation = Opposite(orientation);
6. end

Every time an adjective with its orientation is added to the seed list, the seed list is updated. Along with the word orientation, we have also included the sentence orientation. Here the orientation of a particular sentence is determined on the amount of positivity or negativity contained in it. Since the user opinion words are mostly positive or negative, we can use variants of these algorithms to predict the semantic orientation of the polarity of a sentence.

D. Summarization

Once the reviews are passed through the previous steps, the process of summarization is consisting of following steps:

- Once we have the polarity of individual sentences, we can summarize the review based on feature ranking.
- We can rank all the features based on the frequency of occurrence in the reviews written by the users.
- We can compute a count to show how many users have given positive or negative reviews to a particular feature.
- Based on weightage given to each feature along with their orientation score, we can plot the overall score for each review.

V. RESULT & ANALYSIS

We have tested our algorithm with different opinion mining algorithms and the results are given below. For the purpose of testing we had taken multiple fixed samples as inputs to each of the classification algorithms.

A. Sample Input Data

Sample A: "I like this phone, especially the camera and user interface. There is a small lag because of the bad processor, also the cost of the phone is too high which does not make it value for money. If you can afford the phone you will not be disappointed."

Sample B: "This phone is the best at this cost, though it lags because of low RAM space. I really enjoyed using the front camera as well as the superb headset. The AMOLED display was also great."

Sample C: "I like the phone, but it has bad camera, good battery life, complicated UI and bad looks, also it has bad processor."

B. Comparative Study / Result:

Table: Sentiment Accuracy Score

Name of Algorithm	Accuracy score of sentiment			Average
	Sample	Positive Score	Negative Score	
Naïve Bayesian Classifier	Sample A	0.125	1.75	-1.625
	Sample B	2.75	0.75	1.00
	Sample C	1.00	2.375	-0.6875
Maxent Classifier	Sample A	0.125	2.125	-2.00
	Sample B	2.50	0.25	1.125
	Sample C	0.75	2.00	-0.625

We analyzed & collected results from Naïve Bayesian based Sentiment Classifier and Maximum entropy Text classifier. We have found that Naive Bayes classifier is much more efficient than Maxent since it is less computationally intensive (in both CPU and memory) and it requires a small amount of training data. However it appears that Maxent provides a better accuracy result even though it took longer to compute. We used the same POS tagger & provided the same sample inputs to both the classifiers and the results show that the Maxent classifier provides the better way to ensure sentiment extraction accuracy.

CONCLUSION

Our paper proposes a different approach on Sentiment Analysis and Opinion Mining where we use web crawling, aspect tables, data mining techniques, SentiWordNet, parsing, POS tagging for opinion mining process. Mobile phone reviews were collected as test dataset from Amazon using web crawler written in python. In our paper, we consider aspects and features that are explicitly mentioned by user.

FUTURE SCOPE

In the proposed paper, there is facility of providing the review only by text. Hence, there is a future scope wherein the input to be taken can be in other methods like Speech-to-Text and hand gestures which can be further used for understanding human emotions by machines. There is also the future aim to avoid the SentiWordNet approach and find an unsupervised training approach & to implement clustering based summarization.

ACKNOWLEDGMENT

We are highly indebted to our guide, Prof. Kavita Jain, for her exemplary guidance and monitoring and for sharing her views about the paper. Also, we like to thank all our faculties from the computer engineering department at Xavier Institute of Engineering for their continued and relentless support. We would like to thank our parents and friends for their kind co-operation and encouragement.

REFERENCES

- [1] "SentiWordNet: lexical resource for opinion mining", [Online]. Available: <http://sentiwordnet.isti.cnr.it/>
- [2] Hu, M., and Liu, B., "Mining and Summarizing Customer Reviews", 2004.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.2378&rep=rep1&type=pdf> I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [3] Hu, M., and Liu, B., "Mining Opinion Features in Customer Reviews", 2004. <http://www.aaai.org/Papers/AAAI/2004/AAAI04-119.pdf>
- [4] <http://blog.datumbox.com/machine-learning-tutorial-the-naive-bayes-text-classifier/>