



## Analysing Customer Reviews Using Opinion Mining

Dr. S. Sagar Imambi, Y. Chandana, G. Raphi

Department of CSE, NEC, AP,  
India

**Abstract**— This paper shows how opinion mining is applied to customer reviews. We applied a feature weighting technique, which gives weights not only for the features but for the adjectives, used in opinion sentence. We used Sentence type Opinion Mining. Based on the adjective database, our technique produced 88% of precision in analysing the customer reviews or opinions.

**Keywords**— Opinion mining-Customer reviews-Positive feedback on products- Negative feed back.

### I. INTRODUCTION

In recent years, we found that opinionated postings in social media have helped remodel e-businesses. Public sentiments and feelings always impact social and political systems. Opinion mining relates to computational techniques for analyzing the opinions that are extracted from various sources [1][9].

Opinion mining is essential for both individuals and companies. Individuals may want to see the opinion of other customers about a product to analyze it before buying it. Companies want to analyze the feedback of customers about their products to make future decisions. Therefore companies can take decisions about their products based on customers' opinion. Thus companies can modify their products according to customers' opinions in a better and faster way so that they can establish better customer relationship by giving them exactly what they need. By acquiring the knowledge of customer demands, the companies can easily attract and retain customers. Existing research work on Opinion is based upon business and e-commerce such as product reviews and music albums ratings. Reviews are an increasingly important for online consumers. The ease of posting content to the Web creates opportunities to improve the businesses. A recent survey found that 64% of consumers would read online reviews when purchasing technology items such as MP3 players and cameras[10].

Opinion mining can be defined as a sub-discipline of computational linguistics that focuses on extracting people's opinion from the web. It is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents.

A number of opinion mining tools are available in the market which are suitable for extracting general opinions. Some of them are Meltwater, Google Alerts, People Browser, Google Analytics, Facebook insights, Marketing Grader e.tc.

#### 1.1. CHALLENGES OF OPINION MINING:

- Identifying synonyms:

Different words or phrases can be used to refer to the same quality of the object. So, such words should be identified and grouped together. It is a difficult task to identify these words. A lot of research is required to be done on this issue as it has not been much addressed in the past.

- Identifying positive or negative opinion:

Similarly detecting an opinion as positive, negative or neutral can be a difficult task in opinion mining. A word could be measured positive in one situation and negative in another situation. This can be difficult to calculate as a sentence can be considered negative because of the use of negative words in it.

#### 1.2 DIFFERENT LEVELS OF OPINION MINING:

##### 1.2.1 Document level Opinion Mining:

Generally Opinions are not stored in full document, unless it is some organizational feedback report. But Customer share their opinions in blogs, forums which are not in document form. Hence for customer opinion mining document level mining is not suitable. It is much useful for high level or formal feedback or sentiment mining.

##### 1.2.2 Sentence level Opinion Mining:

In this method, individual sentences, bearing sentiments in the text are considered for classification. In sentence level Opinion Mining, the relations between sentences are calculated. The same document level classification methods can be applied to the sentence level classification. The subjective sentences contain opinion words which help in determining the sentiment about the entity. After that sentence classification is done into positive and negative classes.

##### 1.2.3 Phrase/feature level Opinion Mining

Opinion Mining can be based on some specific phrases. The phrase level sentiment classification is a much more precise in identifying opinions. In this method the phrases that contain opinion words are found out and a phrase level classification is done. But in some other cases, where contextual relations matters, the result may not be accurate.

**1.3 Research Areas in Opinion Mining:**

- Customer feedback for Individual Product
- Identification of highly rated experts
- Overall positive and negative relations at paragraph level .
- improving the accuracy of algorithm for opinion detection
- Ranking of best paragraph or best sentence based on best feature
- Continuous Improvement of the algorithms for opinion detection Decrease the human effort needed to analyze contents
- Sentiment classification
- Reduction of human effort needed to analyze content

**II. SURVEY OF LITERATURE**

Morinaga et al. compare reviews of different products in one category to find the reputation of the target product. This cannot mine product features and summarize reviews. They find some repeated features like “doesn’t work”, “benchmark result” and “no problem(s)”[3]. Cardie et al discuss opinion-oriented information extraction. They aim to create summary representations of opinions to perform question answering. They propose to use opinion-oriented “scenario templates” to act as summary representations of the opinions expressed in a document, or a set of documents. They also generated summary automatically [4].

In 2009 ,Qiu et al. proposed Double Propagation which is a progressive unsupervised technique for clustering features. It mainly extracts noun features, and fit for medium-size corpora. But for large corpora, this method can introduce low precision, and for small corpora, it can miss important features. To deal with these two problems Zhang et al proposed a new method which enhances double propagation. It first uses part-whole and “no” patterns to increase recall. It then ranks the extracted feature candidates by feature importance, which is determined by two factors: feature relevance and feature frequency. The Web page ranking algorithm HITS was applying to compute feature relevance. Experimental results using diverse real-life datasets show promising results[7].

Movie review mining is different from other topic-based classifications. Movie review mining classifies movie reviews into two polarities: positive and negative. Pimwadee Chaovali et.al investigates movie review mining using two approaches: machine learning and semantic orientation. The approaches are adapted to movie review domain for comparison.[6]. Morinaga, Satoshi, et al .[3]proposed a new framework for mining product reputations on the Internet. It automatically collects people's opinions about target products from Web pages, and it uses text mining techniques to obtain the reputations of those products. On the basis of human-test samples, they generate in advance syntactic and linguistic rules to determine whether any given statement is an opinion or not, as well as whether such any opinion is positive or negative in nature.

**III. CLASSIFICATION OF OPINIONS FROM SOCIAL NETWORKS.**

The opinions are gathered from social networks. They are in the form of sentences. We gathered data from various social networking sites like, Blogs, Twitter, and Facebook. Classification follows these steps:

- Step1 Preprocessing of opinions
- Step2: Extraction of feature-phrases of products
- Step 3: Classification of opinions.

**3.1 Preprocessing of opinion:**

In this process noise or common words are removed by comparing with list of common words. Product features are usually noun or noun phrases. We used NLPProcessor to add part of speech tags for each word. For example <NN> indicates noun.

**3.2 Extraction of feature-phrases of products:**

In this process we extract feature phrase pattern. The pattern looks like this NN, NNNN jj NN where NN represents noun and jj represents adjective. These nouns are verified with product feature ontology. Only matched nouns are used in classification. The below table shows the some of the positive and negative phrases used in the customer opinions.

Table 1: positive and negative adjectives

Positive	Negative
dazzling	useless
brilliant	slow
phenomenal	bad
excellent	stupid
fantastic	clichéd
gripping	boring
mesmerizing,	hideous

mesmerizing	Awful
riveting,	unwatchable
spectacular	terrible
awesome,	Suck

The sentences are first divided in to groups based on the product by identifying <NN> tags. Each product features are also identified with the feature phrase like <NN>jj where NN is feature and jj is adjective. Each adjective is given weightage in the sentence. The positive adjective has positive weight, negative adjectives have negative weight. Each feature weights are arranged in vector form. One sentence became one vector. Features may differ for each product.

### 3.3 Classification of opinions:

Main Significant challenges of opinion mining are difficulty to capture opinions from natural language sentences/documents, identify positive or negative opinions.

We used Bayes classification model to predict whether the opinion is positive or negative feedback. On many real-world example datasets naive Bayesian learning gives better test set accuracy than any other method, including backpropagation and C4.5 decision trees.

The probability of a opinion (O) being in class 'C'(positive,negative) is computed as

$$P(c | s) = P(c) \prod_{1 \leq k \leq n_s} P(f_k | c) \quad \text{Where } P(f_k | c) \text{ is the conditional probability of feature } f_k \text{ occurring}$$

in a sentence of class .  $P(f_k | c)$  shows how much this feature contributes for predicting correct class. P(c) is the prior probability of a sentence occurring in class 'C'. This Probability is calculated for each class i.e., positive and negative. In text classification, our goal is to find the best class for the sentence. The best class in NB classification is the class which has maximum posteriori probability .

## IV. EXPERIMENTAL RESULT

Customers can post reviews on web communities, blogs, discussion forums, twitters, product's web site these comments are called user generated contents. SO there is lot of data available freely thru various websites. In our experiment we used data set from <https://archive.ics.uci.edu/ml/datasets>. This data set contains reviews obtained from amazon.com Edmunds.com e.t.c, It includes 51 products such as Toyota car, Ipod etc. There are approximately 100 sentences on an average per product. Several others also used this data set to summarize the opinions.[8].

Table2: Sample opinions of Toyota car

<i>Opinion 1</i>	
I'm enjoying my new car everyday,	I'm really happy with our choice.
Our experience at the dealership was very good and easy, I've not been so relax before doing this kind of transaction. I felt the team was great and Carmack took care of us and made sure we were taken care of.	
<i>Opinion 2</i>	
Everything is perfect	no frustration. This is wonderful.
I will recommend for my friends.	

The dataset file also comes with gold standard summaries used for the Opinion summarization. After applying the classification algorithm we measured the effectiveness in terms of precision. Precision is calculated using the formula TP/(TP+FP). Precision is the fraction of the opinions retrieved that are relevant to the positive or negative feed back

Table3.. Confusion matrix

Classified	Relevant /True	Not Relevant /False
TRUE	2350	655
FALSE	275	1750

$$\text{Precision} = 2650 / (2650 + 355) = 0.8818$$

The precision and recall for some of the products is tabulated in the below table

Table 4. Precision and recall for some of the product

Product	Precision	Recall
Toyata car	0.8123	0.7852
IPOD	0.7834	0.7232
Digital camera	0.6781	0.5895
MP3 player	0.8832	0.8002

## V. CONCLUSION

In this paper we applied a technique to weigh the features and adjectives used in the opinion sentences. This technique resulted overall 88% of precision for predicting opinion class which is either positive or negative. Individual categories also showed nearly the same precision and recall. But our procedure depends upon the common adjectives list and the phrases used in opinions. In future we would like improve adjective phrases database, so that we can get better result.

## REFERENCES

- [1] Pang, Bo., Lillian Lee. "Opinion Mining and Sentiment Analysis-Foundations and Trends in Information Retrieval" pp. 1-135 2008.
- [2] Qiu, Guang., Bing, Liu., Jiajun Bu and Chun Chen "Expanding Domain Sentiment Lexicon through Double Propagation", in Proceedings of IJCAI 2009.
- [3] Morinaga, S., YaYamanishi, K., Tateishi, K, and Fukushima, T" Mining Product Reputations on the Web". KDD'02,2002.
- [4] Cardie, C., Wiebe, J., Wilson, T. and Litman, D. "Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering". AAAI Spring Symposium on New Directions in Question Answering,2003.
- [5] Chieu, Hai Leong and Hwee-Tou Ng. "Name Entity Recognition: A Maximum Entropy Approach Using Global Information", in Proceedings of the 6th Workshop on Very Large Corpora, 2002.
- [6] PimwadeeChaovalit, Lina Zhou, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches",Hawaii International Conference on System Sciences(HICSS)2005, pp. 112 .2005.
- [7] Zhang, L., B. Liu, S. Lim, and E. O'Brien-Strain. " Extracting and Ranking Product Features in Opinion Documents" in Proceedings of International Conference on Computational Linguistics (COLING-2010), 2010.
- [8] KavitaGanesan, ChengXiangZhai, Jiawei Han. Opinions: "A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions". In Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). Beijing, China.
- [9] Liu, Bing. 2006. "Web Data Mining: Exploring Hyperlinks, contents and usage data". Springer,2006.
- [10] <https://econsultancy.com/blog/9366-ecommerce-consumer-reviews-why-you-need-them-and-how-to-use-them/>.