

# Opinion Mining: Aspect Level Sentiment Analysis using SentiWordNet and Amazon Web Services

Kajal Sarawgi  
MTech  
Computer Science & Engineering  
Global Engineering College,  
Jabalpur(M.P.)

Vandana Pathak  
Associate Professor  
Information Technology  
Global Engineering College,  
Jabalpur(M.P.)

## ABSTRACT

In today's linked world, users can purchase items at any time. However, in online shopping sites customers can locate their concerned product by visiting the site of the trader directly or by seeking among different vendors by using a shopping search engine, which demonstrate the similar product's accessibility and costing at alternative e-retailers. The active progress of the audience of shopping sites on the internet lead to the development of these resources as a new origin of the public's mood and opinion about particular product. The tracking of public's responses through reviews and feedbacks in online shopping sites has attracted a greater level of enthusiasm in the research society. Researchers notice that the millions of public opinion polls can't be processed manually. This figure out the requirement of computerized methods for intelligent analysis of text instructions, which allows to process a large amount of data in short time and to interpret customer's feedbacks. This interpretation of feedback is the most valuable and complicated element of the computerized processing. These notions provide the opportunity to perform large-scale research and to observe Online shopping sited in real-time. The main focus of this paper is to determine the aspect terms present in each sentence, searching out their polarities, discovering the polarity of sentences and the polarity of each aspect category.

## General Terms

Data-mining, Text-mining, Opinion mining

## Keywords

Sentiment analysis, Aspect based opinion mining, POS tagging, SentiWordNet.

## 1. INTRODUCTION

Sentiment analysis is grown up field in research area. It allows to mine customer suggestions, product reviews, and even article. The feedback helps the organization to improve the product and the individual to make their mind to purchase the particular product. Sentiment analysis is the analysis of public's evaluation, attitudes, sentiments, opinions, emotions and appraisals towards various kinds of entities expressed in textual information. An entity could be anything like any product, organization, service, individuals, topics, issues, events and their features. Opinion mining have open out to every practical sphere, from consumer products, health care, service, financial, e-commerce and telecommunication to political campaigns, elections and social events. From Data Mining Outlook Opinion Mining or Sentiment Analysis could be treated as multi stage classification issue. Online shopping sites like Amazon, Myntra, Flipkart, Snapdeal have become exceptionally trendy. Today millions of people share their experience about the online shopping and purchasing of

products through feedbacks. Fig 1. Showing levels of Opinion mining.

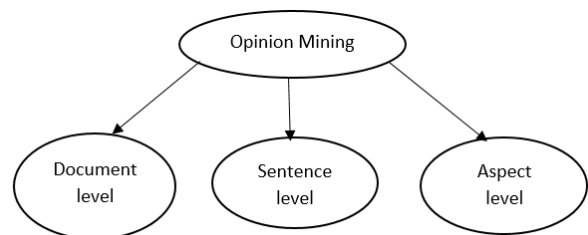


Fig 1: Levels of Opinion mining

Opinion mining can be performed at three levels-

1. Document level
2. Sentence level
3. Aspect level

### 1.1 Document level

In this method the whole document is considered as a single entity and the whole document is analyzed at a time. Sometimes the outcome given by this approach is inappropriate. A document that is positively opinioned about an entity doesn't implicate that the author has only positive reviews about all the feature of that particular entity, likewise a document that is negatively opinioned about an entity doesn't signifies that the author is wholly negative for all the features of that entity. In conventional opinioned text the author express both positive and negative opinions about the entity and its attributes.

### 1.2 Sentence level

In this method the document is break into sentences and then each sentence is treated as a single entity and a single sentence is analyzed at a time. The result generated by this approach is better than document level and are more refined. Majority of present techniques, attempt to establishing the overall polarity of document, paragraph, sentence or text regardless of the entity being expressed.

### 1.3 Aspect level

In this method fine grained analysis is performed. The main focus is on aspect and feature terms of the product. Outcome of this approach is to discover sentiments on feature of items. Mostly of the current approaches tries to determine overall sentiment of the sentence regardless the attributes and features of target entity (e.g. mobile phones, laptops) and their attributes (size, memory, screen, camera). In aspect based sentiment analysis polarity for each aspect is determined. This innovates entirely new way to analyze the data. Aspect Based

Sentiment Analysis determines the aspects of each entity which is used in indicating the sentiments. Aspect Based Sentiment Analysis have wide varieties of applications in different fields like travels, services, movie reviews, restaurants and computers. It uses text that is expressed in various ways like comments, feedbacks, forum reviews, messages and discussions. The Process of Sentiment Analysis is shown in Fig 2.

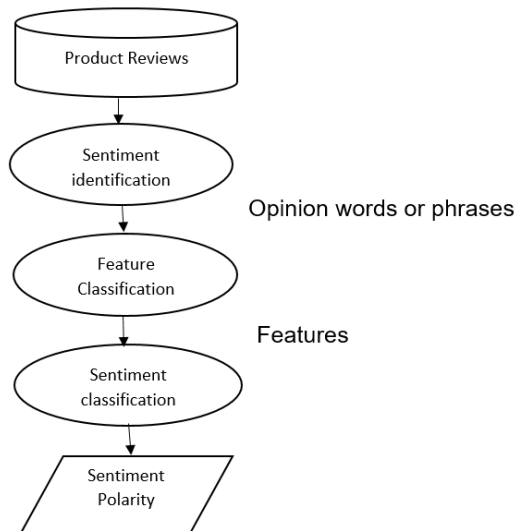


Fig 2: Process of Sentiment Analysis

Data Mining- Data Mining is the computational process of determining patterns in large database (also known as knowledge discovery). Data mining tools predict behavior and future trends. The main process of data mining involves data preprocessing, data analysis, data interpretation and data classification. Data classification is the process of categorizing and sorting data into various types or any other class. Classifier algorithm is used for data classification. This algorithm maps the input to the data category, it is basically a mathematical function.

Text Mining- The process of obtaining high quality information from text is known as text mining. Text mining is useful for particular purposes and it is the process of analyzing text. Different kinds of data can be stored in database, text is unstructured, vague and it is difficult to deal it algorithmically. A typical application of text mining is a set of documents to be scanned which is written in natural language and it is used for predictive classification purpose.

Here in this paper SentiWordNet is used for orientation of words and for extracting features of product Amazon web services are used. NLProcessor is used for sentence segmentation and parts of speech tagging.

The rest of the paper is organized in the following manner. Section 2 is dedicated for related work. It covers aspect based sentiment analysis and the previous work in opinion mining field. Section 3 represent the aspect based sentiment analysis task. Section 4 representing the database collection and annotation of corpus. Section 5 represent methodology and working of system. Section 6 is dedicated for conclusion and future possibilities and Section 7 represent reference of this research work.

## 2. RELATED WORK

The last two decades have seen improvement in the field of opinion mining or sentiment analysis. A number of experimentation papers have also been published and issued presenting new techniques and novel ideas to perform sentiment analysis. Still there required many ideas for the field of corpus creation and data extraction. Most of the work in this area focuses on finding the sentiment orientation of the data at various levels but for the accuracy improvement data pre-processing and feature selection is performed. Here the research work is influenced by Bo Wang and Min Liu [17] deep learning of neural nets to solve the aspect level task. For finding the aspects and their polarities syntactic structure and convolution neural nets are used. SentiWordNet for Indian languages paper by Amitava Das [5], various techniques are used like WordNet based, dictionary based and Corpus based. Thellaamudhan C [15] Various approaches like Frequency based, Syntax based, Supervised and unsupervised machine learning approach and hybrid approach are discussed. K.S Ilakiya [8] comparison between various algorithm like Naïve Bayes algorithm, Support Vector Machine and Maximum Entropy their benefits and limitation are compared. A book by Bing Liu and Lei Zhang Survey of opinion mining and sentiment analysis helped in this paper writing process. In the work by Pooja A Rangari [20] Machine learning approach and lexical based approach are described. Here the research work is different from above in many aspects. 1) data mining association rules are used for mining association rules. 2) For finding orientation of words SentiWordNet is used. And 3) Reviews are extracted from Amazon web services.

## 3. ASPECT BASED SENTIMENT ANALYSIS TASK

This scenario is classified into following categories:

1. Aspect term extraction
2. Aspect term polarity
3. Aspect category determination
4. Aspect category polarity

Fig 3. Representing Aspect based Sentiment analysis process.

### 3.1 Aspect Term Extraction

First step of aspect level sentiment analysis is data preprocessing. This is a significant step towards information extraction and to identify each aspect term present in review sentence. There can be multiple aspects in a single sentence and the aspect could be single word or multi word. Every aspect need to be extracted. The aspects can be represented by noun, verb, adjective and adverb. There are the following ways for data preprocessing-

- a) Stemming
- b) (b)POS tagging
- c) Tokenization
- d) Stemming of 3 words
- e) POS tagging of 3 words
- f) Feature indication

#### 3.1.1 Stemming

In this process suffix is removed from the actual word (for ex. ing, tion, sion, ed etc.) e.g. after removing suffix Buying will become Buy.

### 3.1.2 POS tagging

This is the parts of speech tagging step in the sentence. So that all the aspects can be identified based on noun and noun phrases and also get the importance of other words.

### 3.1.3 Tokenization

In this process each sentence is tokenized by spaces that means space removal, stop word removal like a, an, the etc.

### 3.1.4 Stemming of 3 words

Here stemming of three words that means before and after the aspect term is performed.

### 3.1.5 POS tagging of 3 words

Here POS tagging of three words that means before and after the aspect term is performed.

### 3.1.6 Feature indication

An aspect term is indicated by capital letter. For example, Image quality of this phone is awesome. Here Image quality is the aspect term which is represented by this phone.

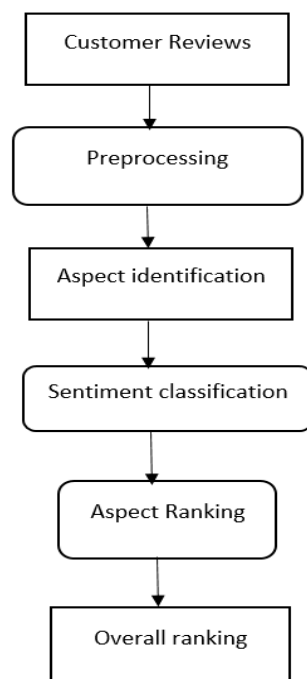


Fig 3: Aspect Based Sentiment Analysis

## 3.2 Aspect term polarity

For every given set of aspect term, the polarity is determined as positive, negative, neutral or bipolar (i.e. both negative or positive). For example, Battery backup of Mi4i is good. Here Battery backup is aspect term which is positive as the sentiment term good is positive.

## 3.3 Aspect category determination

Given a predefined set of aspect categories (e.g. cost, memory), here the task is to identify the aspect category in the given sentence. Aspect category are coarser than aspect terms and they occur very few in the sentence.

## 3.4 Aspect category polarity

Given a predefined set of aspect categories (e.g. cost, memory), here the task is to identify the aspect category polarity in the given sentence.

## 4. INPUT REQUIRES (DATABASE COLLECTION)

For accomplishing Aspect based sentiment analysis there are many datasets are available for different entities. Datasets can be gathered from websites like amazon.com, imdb.com, tripadvisor.com and many others resources for analysis. Before the implementation of Aspect based sentiment analysis, data is preprocessed to get the understanding of data. There are various techniques which researchers have used like remove the punctuation marks, stop words, tokenization, normalization and parts of speech tagging. Large number of mobile reviews are available on internet so they are collected from different-different websites. Mobile reviews contain users and critic reviews. There are various websites presented on the web which contain mobile reviews like www.digit.in, m.gadgets.ndtv.com, www.hemobileindian.com, amazon customer reviews etc.

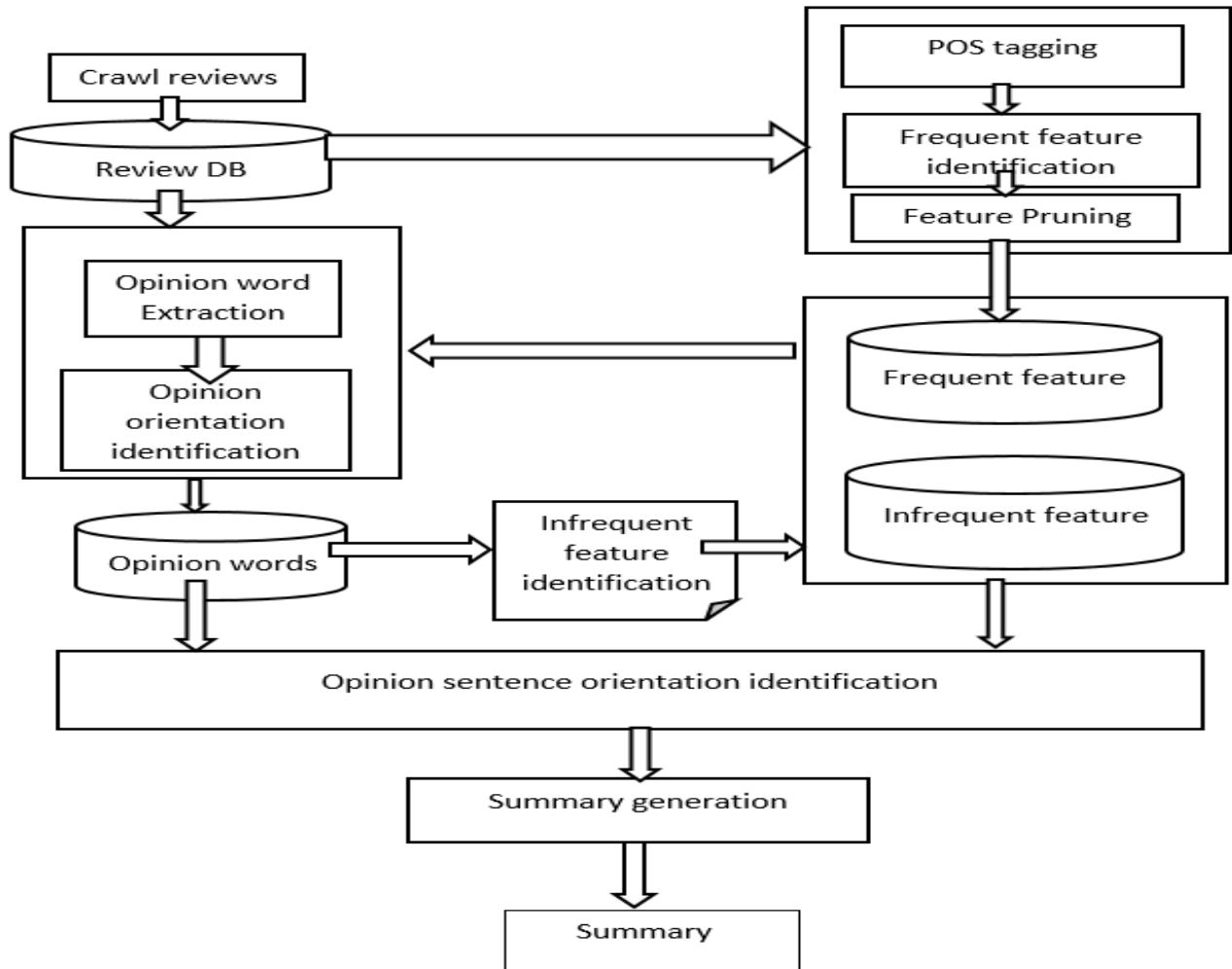


Fig 4: Step by step working of the system

## 5. WORKING OF SYSTEM

- Gather review of product (ex. mobile) from amazon.com
- Select the common features of mobile phone from the description of the product.
- Select a set of feature to rate on.
- Determine the ratings for the selected features based on the sentiments of the features.
- Now summarize the ratings for the features by calculating the total number of positive and negative reviews.

The working block diagram is shown in Fig 4.

### 5.1 Step by step implementation details

#### 5.1.1 Review Collection

There are many resources available on WWW that provides reviews about products. Here Amazon is used for collecting reviews because a large number of domain is covered by It as well as It provides large number of choices to the customer. It is also covered considerable different reviews for an item. The reviews are acquired from Amazon's Web Services API; from there an XML file is obtained. And then this XML file is parsed for getting reviews.

#### 5.1.2 Sentence segmentation and POS tagging

Here NLProcessor program is used for accomplishing this step. This program facilitates both Unix and Windows. After getting the product reviews, these are run through this NLProcessor software. This gives output in the NLProcessor format.

#### 5.1.3 frequent features identification

Here candidate features are obtained from all the noun and noun phrases occurred in each sentence and then they are aggregated into a transaction file. And then Apriori algorithm is applied to get the features which are frequently commented, this conclude that these are the features which are really matters for the product.

#### 5.1.4 Feature pruning

In this step irrelevant features are removed from the set of candidate features by using heuristics.

#### 5.1.5 Opinion words extraction

In this step opinion words are identified from the set of product features. For this; set of adjectives is extracted which are describing product features.

#### 5.1.6 Opinion orientation identification

After getting opinion words, its orientation is calculated i.e. opinion expressed is either positive or negative. SentiWordNet is used here for calculating the score of

positivity, negativity and neutrality.

### 5.1.7 Opinion sentence orientation identification

After calculating the orientation of opinion words, orientation of sentence is estimated. For this only those sentences are selected which are consisting at least one opinion word.

### 5.1.8 Opinion summarization

By using orientation of each of the feature, total number of positive and negative sentence are calculated. Here the tuple is formed <feature, positive count, negative count>.

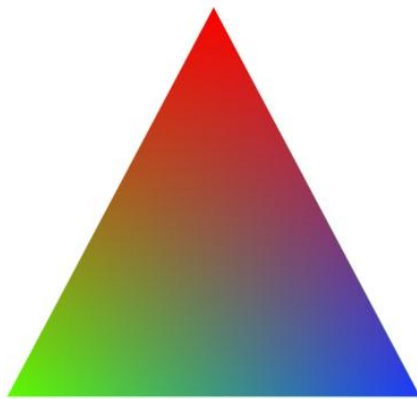
## 5.2 Tools and data

SentiWordNet is publically available tool, it is lexical resource for opinion mining task. Sentiment resource as positivity, negativity and objectivity. Current version SentiWordNet 3.0 is used here. SentiWordNet is made up with tens of thousands of words with their meanings, synonyms, parts of speech tags and their degree of polarity ranging from 0 to 1. The user can also visualize these scores, here each category is linked to the color like positivity is represented by green color, negativity is represented by red color and objectivity through blue color. This is shown in

Fig. 5 and table 1 represents tools and data used in this research work.

**Table 1. Tools and data**

S. No.	Data	Tool
1.	Mining association rule	Data mining association rules
2.	Sentence segmentation and POS tagging	NLProcessor
3.	Orientation of words	SentiWordNet
4.	Extraction of reviews	Amazon web services



**Fig 5: Visualization of scored synset in SentiWordNet 3.0**

## 6. CONCLUSION

In this paper, Aspect based sentiment analysis on mobile phone have been performed. For analyzing textual content i.e. reviews and feedback of consumers through the process of text mining. Text mining is performed by the tool called SentiWordNet, NLProcessor and amazon web services. We believed this research paper is progressively will be utilized by more individuals give their sentiment in e-business. As well as looking on reviews from site on various different brands is more valuable for consumer and for manufacturer.

In future, many issues regarding aspect based sentiment analysis would be performed like determination of aspect terms in interrogative sentence. The effect of negation is not considered in this thesis work. Negation reverse the opinion so considering it, will vary the result. Few sentence may be written in sarcastic way or in the form of jokes and they destroy the value of whole document and violet the sense of particular sentence. Comparison between products of different manufacturer have to be handled. Sometimes it happens spam reviews are posted by the opposition to degrade the brand publicity. These aspects would be handled in future.

## 7. REFERENCES

- [1] Kim Schouten and Flavius Frasinca, Survey on aspect level sentiment analysis, DOI 10.1109/TKDE.2015.2485209, IEEE Transactions on Knowledge and Data Engineering
- [2] Deepak Singh Tomar, Pankaj Sharma, A text polarity analysis using SentiWordNet based algorithm, IJCSIT 2016 ISSN:0975-9646
- [3] Haseena Rahmath P, Opinion mining and Sentiment Analysis- Challenges and applications, IJAIEM 2014 ISSN: 2319-4847
- [4] Samaneh Abbasi Moghaddam, Aspect based opinion mining on online reviews, Spring 2013
- [5] Amitava Das, SentiWordNet for Indian languages, 8th workshop on Asian Language Resources, Beijing, China, August 2010.
- [6] Arti Buche, Opinion mining and analysis: A survey, IJNL 2013
- [7] Mrs. A. Nirmala, An enhanced sentence level sentiment classification opinion mining system with POS tagging, IJETCSE 2016, ISSN: 0976-1353
- [8] K.S Ilakiya, Challenges and techniques for Sentiment Analysis: a survey, IJCSMC March 2015, ISSN 2320-088X
- [9] Vivek Kumar Singh, Computing sentiment polarity o texts at document and aspect levels. ECTI 2014
- [10] Erik Cambria, Knowledge based approaches to concept level sentiment analysis, IEEE 1541-1672, 2013
- [11] Josef Steinberger, Aspect level sentiment analysis in Czech, 2014
- [12] Aishwarya Mohan, An approach to perform aspect level sentiment analysis on customer reviews using SentiScore algorithm and priority based classification, IJCSIT 2014, ISSN: 4145-4148
- [13] Chetan Mate, Product aspect ranking using Sentiment analysis: A Survey, IRJET 2015, ISSN:2395-0056
- [14] IEEE Computer society, New avenues in Opinion mining and sentiment analysis, IEEE 2013, ISSN :1541-1672
- [15] Vidisha M. Pradhan, A survey on sentiment analysis algorithm or opinion mining, IJCA 2016, ISSN: 0975-8887
- [16] Thellaamudhan C, A comprehensive Survey on aspect based sentiment analysis, IJARCSSE 2016, ISSN: 2277-128X

- [17] Aurangzeb khan, Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs, *Int. J Comp Sci. Emerging Tech*
- [18] Mining Hu and Bing Liu Department of Computer Science from University of Illinois at Chicago 851 South Morgan Street Chicago, IL 60607-7053 “Mining and Summarizing Customer Reviews.”
- [19] Bo Wang and Min Liu, Deep learning for aspect based sentiment analysis
- [20] Pooja A Rangari, A survey on Aspect based opinion mining, *IJSETR*, March 2015, ISSN: 2278 – 7798