

# Aspect Based Sentiment Analysis

1.Namaswi Chandarana, University of Illinois at Chicago, [nchand22@uic.edu](mailto:nchand22@uic.edu)

2.Vanisre Jaiswal, University of Illinois at Chicago, [vjaisw2@uic.edu](mailto:vjaisw2@uic.edu)

## I. ABSTRACT

Aspect Based Sentiment Analysis receive as input a set of texts discussing a entity. The systems attempt to detect the main aspects of the entity and to estimate the average sentiment of the texts per aspect i.e. how positive, negative or neutral the opinions are on average for each aspect. Although several ABSA systems have been proposed, mostly research prototypes, there is no established task decomposition for this analysis, more are there any established evaluation measures for the subtasks are required to perform.

## II.INTRODUCTION

Sentiment Analysis (SA), so called Opinion Mining (OM) is the study through which analysis of people's opinions, sentiments, appraisals, evaluations, attitudes and emotions towards entities of various kinds expressed in Text. An entity can be a product of various types, services, organizations, individuals, issues, events, topics and their aspects. Opinion mining or sentiment analysis have spread to every possible domain, from health care, financials, consumer products, service, telecommunications and e-commerce to social events, political campaigns and elections. The two expressions SA and OM are interchangeable, they express the mutual meaning. However, in view of some researchers SA and OM have different notions. Opinion Mining extracts and analyses people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyses it. From Data Mining Standpoint Sentiment Analysis or Opinion Mining can be

considered as multi-step classification problem. In the first step, there are three main classification types like Document Level, Sentence Level and Aspect Level. In second step related to identification of polarity of the document or sentence or aspect as positive or negative or neutral. Document level sentiment analysis aims to classify an opinion document as positive or negative by considering the whole document as basic information unit. Sentence level sentiment analysis aims to determine whether a sentence expresses positive or negative opinion Classifying opinionated texts at document level or at sentence level is useful in many applications; but it does not provide necessary details needed for many applications. A positive opinionated document about an entity does not mean that the author has complete positive opinion about all the features of the entity. Likewise, a negative opinionated document does not mean that the author does not at all like all the features of the entity. In a typical opinionated text, the authors write both positive and negative opinions with respect to entities and their attributes. Majority of current approaches, however attempt to identify the overall polarity of a document, sentence, paragraph or text irrespective of the entities involved. To obtain these hidden details, we need to go to Aspect Based Sentiment Analysis. Aspect Based Sentiment Analysis aims to identify the aspects of entities being used in expressing sentiments. It is also used to determine the sentiment that expressed by author towards each aspect of the entity. Aspect Based Sentiment Analysis is much critical in mining and summarizing opinions from any kinds of datasets. In the last decade several Aspect Based Sentiment Analysis systems have been developed for wide variety of entities like Movie reviews, Travels, Digital cameras, Services, Computers and Restaurants. In more

detail, Aspect Based Sentiment Analysis systems receive a set of texts (product reviews, comments, forum discussions and messages from web 2.0 sources) discussing about a entity (Ex: Phone). The system attempts to find the main (frequently discussed features) Aspects („Screen“, “Size“, “Price“) of the entity and to find the sentiment expressed towards each aspect and their summary of polarity. The rest of the paper is organized as follows. Section 2 provides the details of Data sets and required pre-processing to be done for Aspect Based Sentiment Analysis. Section 3 provides the details of Aspect Based sentiment Analysis Tasks, Section 4 gives the comparison of results from various approaches used by researchers and section 5 provides conclusion and future scope of the ABSA.

### III. TECHNIQUES

The techniques used for the classification are discussed below but before let's discuss the features of the available datasets.

The two-dataset available (restaurant reviews and laptop reviews) for aspect-based sentiment analysis had the following columns:

- **Example\_id:** Unique Id for each review.
- **Text:** Reviews mainly focusing on restaurants and laptops. The reviews had “[comma]” instead on the symbol “,” to separate from the csv file delimiter.
- **Aspect Term:** the term that has to be classified as positive, negative or neutral from the review.
- **Aspect location:** represented start and end location of the aspect term in the review in the following format” start-end”  
Example:””
- **Sentiment Label:** -1(negative), 0(neutral) and 1(positive)

#### Pre-processing

Preprocessing is the most important step for any data analysis. Preprocessing is the most important

step for any data analysis. Data preprocessing used for extracting interesting and non-trivial and knowledge from unstructured text data.

The preprocessing steps were mainly on the columns Text, Aspect Term and Aspect Term Location.

#### Basic Preprocessing Steps:

Preprocessing for Text and Aspect term column

- Removal of uppercase letter and replacing them with lowercase for better training the classifier model
- Replacing “[comma]” with “,” (only for Text column) to add value to the sentence

Preprocessing for Aspect Term Location only involved splitting the start and end location values into different columns.

#### POS Tagging

POS tagging is the process of assigning a ‘tag/category’ (in the form of an abbreviated code) to each word (token) in a given sentence.

In the English language for example, common POS categories are:

- nouns
- verbs
- adjectives
- adverbs
- pronouns
- prepositions
- conjunctions
- interjections

Other categories can be derived from different forms of the above, for example a verb can be in its base form or in past tense.

#### Dependency Parsing

Dependency Parsing is the task of recognizing a sentence and assigning a syntactic structure to it. The most widely used syntactic structure is the parse tree which can be generated using some parsing algorithms.

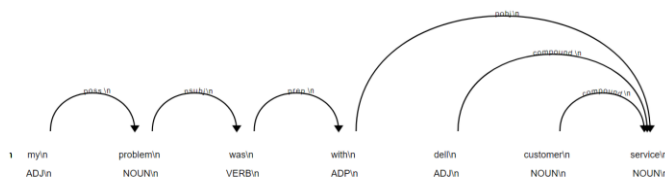
Technique of Dependency Parsing was used to figured out the terms in the sentences which helped in the classifying the sentiment of the

aspect term in the review. Words such as adjectives, adverbs, negations, etc. which were related to the aspect term were extracted from the reviews. Some of the relations that were figured out from the dependency parsing were:

An example is as follows:

“The staff was horrible” ->” horrible”

Visualization of the dependency parsing on one of the reviews is as follows:



The dependency parsing and visualization were done using the Spacy, which a very powerful tool for performing NLP operations.

### Removal of Stop-words

The stop-words (not including “not”) were removed, if any, from the extracted information after dependency parsing. The stop-words do not contribute in any way to decide the sentiment of the aspect term. Whereas, “not” in the statement can act as the opinion shifter and therefore, we have considered “not”. For example:

Text: “The staff was not good”

After Dependency Parsing: “good”” not”

Stop words removal (including “not”): “good”

Stop words removal (excluding “not”): “good” “not”

If we would have removed “not” then the meaning of the sentence would have changed. Hence, “not” is never removed from the list if present and if it is directly connected to the aspect term.

### Bag of Words Models

A bag-of-words is a representation of text that describes the occurrence of words within a document. It involves two things:

- A vocabulary of known words.

- A measure of the presence of known words.

The bag of model’s model was created using the terms which were resulted after dependency parsing and stop words removal.

## IV.MODEL TRAINING

We used Logistic Regression to train our model. Let us discuss about this model in detail.

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

In logistic regression, the dependent variable is binary or dichotomous, i.e. it only contains data coded as 1 (TRUE, success, pregnant, etc.) or 0 (FALSE, failure, non-pregnant, etc.).

The goal of logistic regression is to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest:

### Logistic regression equation

where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

- Odds= $p/(1-p)$
- Logit( $p$ )= $\ln(p/(1-p))$

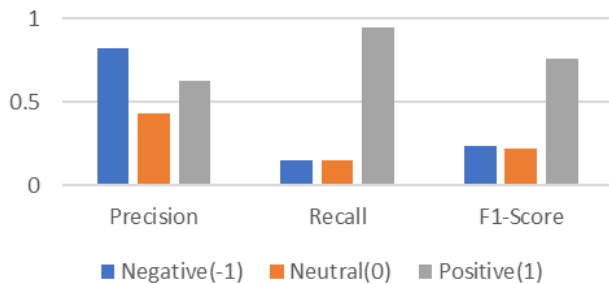
Rather than choosing parameters that minimize

the sum of squared errors (like in ordinary regression), estimation in logistic regression chooses parameters that maximize the likelihood of observing the sample values.

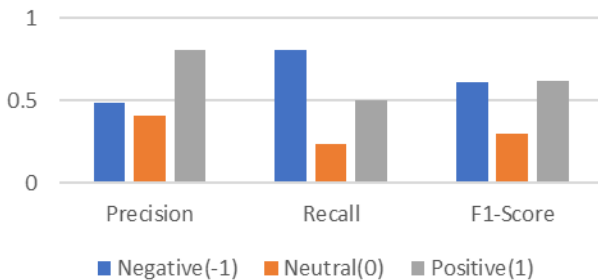
## V.EVALUATION

The evaluation parameters that are considered for the classifiers are F1 score of the three classes (i.e positive, negative and neutral).

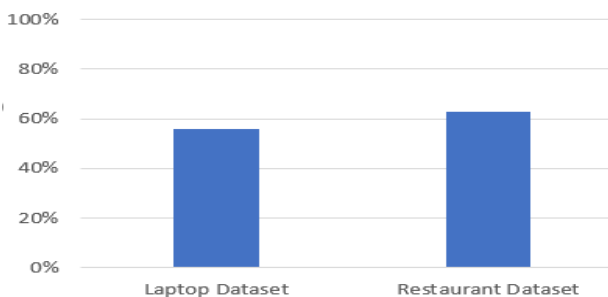
Following are the evaluations for Laptop Reviews using Logistic Regression



Following are the evaluations for Restaurant reviews using Logistic Regression



The accuracy is also considered while evaluating the model.



## References

- 1] Bo Wang, Min Liu. Deep Learning for Aspect term sentiment analysis
- 2] Naveen Kumar Laskari, Suresh Kumar Sanampudi. Aspect based sentiment analysis survey
- 3] Soufian Jebbara, Philipp Cimiano. Aspect based relational sentiment analysis using stacked neural network architecture.
- 4] Raisa Varghese, M Jayasree. Aspect based sentiment analysis using support vector machine classifier.