Semi-Supervised Aspect Extraction for Generic Aspect-based Sentiment Analysis

E M K U Ekanayake

2018

Semi-Supervised Aspect Extraction for Generic Aspect-based Sentiment Analysis

A dissertation submitted for the Degree of Master of Computer Science

E M K U Ekanayake
University of Colombo School of Computing
2018

Abstract

With the increase of popularity of web 2.0 and easy access, rapidly increasing the amount of user generated data. The tendency to rely on reviews of products and services has become more natural. But most of the contents are too much to read and unstructured to get necessary information. This research is proposing an unsupervised approach to extract sentiment for different aspects considering user reviews for hotels. The focus is to do the sentiment analysis for a collection of reviews than individual reviews.

Frequency based aspect word extraction for hotel reviews, aspect category detection and aspect sentiment classification is discussed and evaluated. Usage of general purpose corpus for aspect category detection is experimented. Aspect based sentiment classification is experimented using sentiment analysis implementation in Python. The data set employed is extracted from tripadvisor.com web site using a self-implemented python tool and preprocessed data with NLP techniques for data preprocessing. Preprocessed data was employed to find aspect words and then each review text was parsed to determine aspects which were discussed in each review. One of the major goal was to determine sentiment value for aspects in each review texts. Positive or negative sentiment was identified using sentiment classifiers.

Aspect based sentiment analysis which was used in this research was evaluated under aspect words extraction, aspect category detection and aspect sentiment detection. A manually annotated dataset was used for the evaluation. According to the evaluation results, 70% of accuracy achieved in aspect words extraction. Aspect words were identified using a frequency based approach. Different threshold values for the frequency was evaluated. An aspect words list with less synonyms were detected by specifying a high frequency threshold which was resulting 36% of words as aspect words. When detecting the correct aspect category for a review sentence, 22% of reviews were identified with correct aspect category. Both aspect category and the sentiment value identified correctly in 18% of reviews. At the end of evaluation, 0.4808 level of accuracy found on correctly classified aspect polarity occurrences. Evaluation results reveals further improvement areas which can increase the accuracy and reduce error levels.

The thesis proposes an unsupervised approach for aspect sentiment analysis problem and possible future improvement suggestions to implement an application based on the suggested process.

Acknowledgement

I was able complete this thesis as result of proper guidance and support that I got from several people who have been extremely supportive. First, I would like to thank Dr. A.R. Weerasinghe for being my supervisor, guiding me throughout the research and for the encouragement that I was given to work on this topic.

I would like to thank all the lecturers in University of Colombo School of Computing and MCS project coordinator who helped me various stages throughout the research to make it success and on time.

I always thank my friends who were being with me to encourage all the time whenever I am in trouble of completing the research.

Then my heartfelt gratitude goes to my loving family who were dedicating days and nights for me to make this research a success.

Table of Content

List of Figures and Tables

List of abbreviations

ABSA	-	Aspect Based Sentiment Analysis
ML	-	Machine Learning
SVM	-	Support Vector Machine
DP	-	Double Propagation
LDA	-	Latent Dirichlet Allocation
PDP	-	Poisson Dirichlet Process
CRF	-	Conditional Random Field
POS	-	Part Of Speech tagging
NLP	-	Natural Language Processing

Chapter 1

Introduction

People are massively sharing their thoughts and comments as reviews, about products and services. There are several sites available to publish reviews and websites that selling products and services online have facility to share user reviews and anyone can view those reviews to get an idea about the product or service. Many companies and organizations are interested in utilizing those user reviews to make improvements in their businesses. Therefore, Aspect Based Sentiment Analysis and summarizing opinions from reviews can help community in many ways. Consumers can decide what to purchase and businesses can monitor and understand the needs of the market.

When considering hotels, people tend to do online hotel booking and there are several sites which allow online hotel booking. Those sites contain user reviews collection and people read those reviews to choose a hotel that they are looking for. Reviews contain information with different aspects such as food, service atmosphere so on. When going through all the reviews, it is difficult to get the right overview about the hotel. Some reviews do not contain useful information. Some are having very general ideas. Then it is better if people can see a summary according to aspects such as price, food, service, staff, cleanliness, rooms from all individual reviews.

When number of reviews are huge, it is difficult to identify all aspects that people have mentioned in reviews. Therefore, it needs an automatic way of identifying hidden aspect words. In websites like tripadvisor.com there are some identified aspects and rating values per aspect. But those aspect ratings are entered by users and only very few aspects can be rated. This research is going to address how to find hidden aspects with an unsupervised way. Then focus on employing an aspect based sentiment analysis on identified aspects. Sometimes people may have mentioned about aspects, directly mentioning a specific word such as “service”. Sometimes there can be review texts without mentioning “service” but express implicitly about service by mentioning about staff like “Staff are very attentive”. Then there is a need of extracting hidden aspects by review text analysis.

This research will be focused on identifying aspects using an unsupervised model and enhancing accuracy level. Then aspect based sentiment analysis will be applied to identify sentiments of aspects automatically.

1.1 Aspect Based Sentiment Analysis

"In aspect-based sentiment analysis (ASBA) the aim is to identify aspects of entities and the sentiment expressed for each aspect. The end goal is to be able to generate summaries listing all the aspects and their overall polarity"

Many of existing solutions are based on text analysis in word-level and expressing explicit sentiment and not to identify opinions for any aspect. That is not sufficient to take any decision to customers or venders about their products or services. Also classifying opinions or sentiments on document level is not sufficient get to know about the overall opinion about the product or service.

Even though we assume that a single review evaluates a single entity, a document with positive opinion does not mean that author has positive opinion for all aspects of the entity. In the same way a document with negative opinions does not mean that author has negative opinions about all other aspects of the entity. For more complete analysis, we need to go through aspects and decide whether the sentiment is positive or negative.

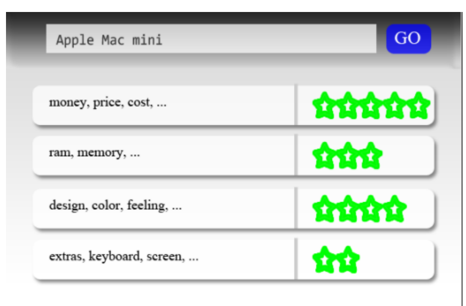

Figure 1.1 - Aspect based sentiment analysis

Figure 1.1 is an example for a product (Apple mac mini) and summarize the content of reviews in aspect-sentiment table Pontiki et al., 2014. But at the same way, hotels have aspects which can be summarized with opinions for aspects. Classifying a review as positive, negative or neutral in the document level is “sentiment polarity detection” and it is only a single dimension. But aspect or feature level sentiment analysis involves two dimensions where revealing all discussed aspects and identifying reviewer thought about aspects sentiment and polarity. Therefore, Aspect Based Sentiment Analysis provides opportunity to determine what aspects customer feel positive and negative about and goes one step further to analyze customer’s sentiment regarding different aspects of product and service.

1.2 Motivation

The amount of available information in the web make concerns about handling information overload, ensuring that user can get expected source of information with least effort. Travel and tourism domain also getting affected by amount of data produced by customers or who has visited places before. With the introduction of web 2.0, customer opinions represent valuable and unique type of information that should not be ignored by the research community. When the number of customer opinions increasing dramatically, need of having a mechanism to make use of data is more and more important. This research aims to provide a better way of utilizing customer opinions available in hotel booking sites.

From the customer perspective, before travelling somewhere else, knowing other’s opinion from those who have visited before was a common behavior even before the internet. But internet is providing services to booking places, people interested in knowing shared opinions from others. But during this time people have access to thousands of opinions which is not an easy task to make a good decision. Therefore, it is necessary to have an automated mechanism to know the summary from thousands of opinions where researches can help to the community.

From the hotel or travel site management perspective, receiving customer feedback is a key to improve their business strategies to improve the profit of their business. For an example, hotels can improve their quality of service of staff members by considering any negative feedback from customers. Large amount of opinion texts makes it harder to manage and harder to get a better idea about customer expectation and difficult to prioritize area of improvements. Then, having a mechanism to analyze those reviews automatically is important.

The major problem with existing service in booking sites, is lack of aspect based summary. Only having rating values for each review text. But there is a motivation to know about aspects such as food, room quality, service, cleanliness and so on. rather than a numeric value. This work presents a way to identify key aspects and their polarity by considering all review texts per hotel. And considering improving the mechanism which is not requiring any prior knowledge or predefined keywords in the aspect extraction and analysis.

1.3 Objective

The main objective is to do an Aspect Based Sentiment Analysis for hotel reviews representing the hotel domain. Also try to contribute by applying an unsupervised model to identify aspects that discussed in review texts. There are several researches available for sentiment analysis for given aspects. But there can be hidden aspects that are useful for hoteliers to know about, to enhance their services and facilities. Also, customers having different interests. There are different aspects that different people are considering. At the same time number of online users and their reviews are increasing. It is very difficult to go through each aspect to get an idea or decide.

There are websites which facilitate for online booking for hotels and those hotels having ratings as well as aspect based summary for each review. But there is no polarity mentioned. The objective of the research is to provide an overall aspect based sentiment analysis for the hotel considering all reviews and aspects along with reviews. One objective will be improving accuracy level of aspect identification and analysis.

1.4 Scope

The project will focus on aspect based sentiment analysis and summarization of hotel reviews. It will consider predefined set of aspects related to a hotel. Those aspects would be extracted form review texts from a site which has reviews. This project will be considering reviews only in English language.

The data set will be limited to TripAdvisor site (https://www.tripadvisor.com/). The reason to use TripAdvisor is that, it contains review along with ratings to individual reviews and aspects for individual reviews. Then research will be based on those extracted data and that data will be using for supervised learning and then to evaluate the model.

1.5 Summary of chapters

This work is divided into six chapters, including this chapter. Related works in Aspect based Sentiment Analysis is presented in Chapter 2. It includes literature survey for ASBA including opinion mining and feature extraction. This chapter will provide a basic background of related areas and will demonstrate related strategies used around feature extraction and sentiment classification. Chapter 3 is dedicated for problem analysis and design of the proposed solution. This chapter address the expectation of the aspect based sentiment analysis for hotel reviews. Further explains what kind of aspect summarization is required with existing review texts. Then present the overall design of the proposing solution in high level and later discussed in detail level about each component of the system. Experimental setup for the proposed design is described in Chapter 4. It mentions about preparing data set, implemented program for aspect term extraction, aspect category detection of review using WordNet and aspect sentiment classification method. Chapter 5 presents the evaluation results of three main modules in the experimental setup. Aspect term extraction, aspect category detection and aspect sentiment classification are three levels. Also, present used experimental measures. Chapter 6 is the final chapter which summarize and conclude the work carried out through the research mentioning future works.

Chapter 2

Related Work

With raising of web 2.0 people can express their ideas about product and service as reviews and comments over the World Wide Web. Those reviews and comments are useful for both customers and merchants to make judgments about products and services. It is not an easy task to extract useful information from huge number of review texts. Then it requires to apply some analysis techniques to make useful information. Aspect based opinion mining has been a topic which has been involved in reviewing text analysis based on different aspects. There are several approaches that can be identified in the literature which diminish this problem into certain extent.

This literature survey will investigate researches in opinion mining area and aspect based sentiment analysis. This chapter will reveal similar works done to analyze customer reviews and different approaches which has been followed. Another intention is to identify pros and cons with different methodologies that has been employed.

2.1 Binary Classification

There are several studies have been conducted to analyze overall sentiment from review texts. The focus has been products that can be purchased online. At the beginning, studies were to classify review text about a product by understanding its polarity [1]. Several machine learning techniques have been employed to categorize review texts and later identified that it is more challenging than topic categorization. Collection of movie reviews have been the dataset and they have found that in terms of performance, Naïve Bayes is the worst and SVM (Support Vector Machine) is the best. Employed methodology has not resulted well as for topic categorization [2]. "Mining *the Peanut Gallery*" is a web based opinion classifier which uses to classify a given text in to positive or negative based on features and resulting a quantitative summery. The success of the classification depends on selection of feature set [3]. Similar binary classification has been done to analyze financial news texts. The aim of that research has been to recognize positive or negative polarity of financial news texts. It had tried contributing for sentiment analysis for financial area. [4] Another research has been done to do an experimental study of sentiment analysis with large scale data set and target was to positive negative classification for online product reviews. According to the results, high order n-grams has helped to make a better judgment of article's polarity in a mixed context. [5] Binary classification further has been used to propose a system which was doing a sentiment analysis for a given topic and identify the person who holds the opinion. The system only capable of analyzing single topic and not texts that has multiple and week opinions.

Satoshi Morinaga et al., [6] has proposed a framework which is based on opinion mining for product reputation from internet. It collects people's opinion about products from websites and apply opinion mining techniques to obtain recognition about products. This paper introduces further enhancement about polarity classification. It identifies opinion text from the web and then label those opinions as negative/positive and then use those labeled opinion text to reputation mining works. They have applied some text mining tasks such as extraction of characteristic words and co-occurrence words to analyze reputation of products.

Bing Lu et al., has introduced "opinion observer", which is analyzing and comparing customer opinions on competing products. It can show strengths and weaknesses of each products in a way that customer can decide by looking at the visualization. For an example it can be visualized customer's negative and positive opinions of features of a digital camera such as picture, zoom, battery and size. That study has tried to present extracted information from unstructured reviews collection and visualize in an understandable manner [7].

2.2 Opinion Mining and summarization

Li Zhuang et al. has been carried out a study on movie review mining and summarization. Websites like imbdb.com allows people to add reviews about movies that they have watched and many people used to look at rating value before selecting a movie to watch. This previous study has been to discover much more hidden information behind user entered review texts. When many other opinion mining studies focused on product reviews, author has focused on a specific domain as movie reviews. And the objective is to generate feature class based summary from review texts for a movie automatically. The summary consists of classification of review texts as negative and positive [8].

Opinion summarization was further illustrated to generate comparative summary of a product. Hyun Dik et al. has proposed a framework which is solving problem called *contrastive opinion summarization*. It identifies opinions as negative and positive and then further identify comparative words to generate comparative summary. The proposed system has used a heuristic method of removing sentiment words in computing contrastive similarity which had found more effective [9].

Similar study on identifying comparative sentences from review texts has done by Nitin Jindal et al. People always like to compare products when exploring which one is to buy. This paper proposes to study the comparative sentence identification problem. It first categorizes comparative sentences into different types, and then presents a new integrated pattern discovery and supervised learning approach to identify comparative sentences from text documents [10].

2.3 Introducing Aspect Ratings

Later, the definition of opinion mining is generalized to a multi-point (ex. One to five starts) rating scale [17, 16]. Several approaches have been introduced to solve the problem, including supervised, un-supervised, and semi-supervised approaches, but they all have attempted to predict an overall sentiment class or rating of a review, which is not much informative as revealing aspect summery as we are intended to do.

Some of the recent studies has attempted to find aspect level rating than overall rating for a review. Each product aspect gives a rating value. For an example a review text in tripadvisor.com site which has hotel reviews and aspect level ratings along with the overall rating as in Figure 2.1

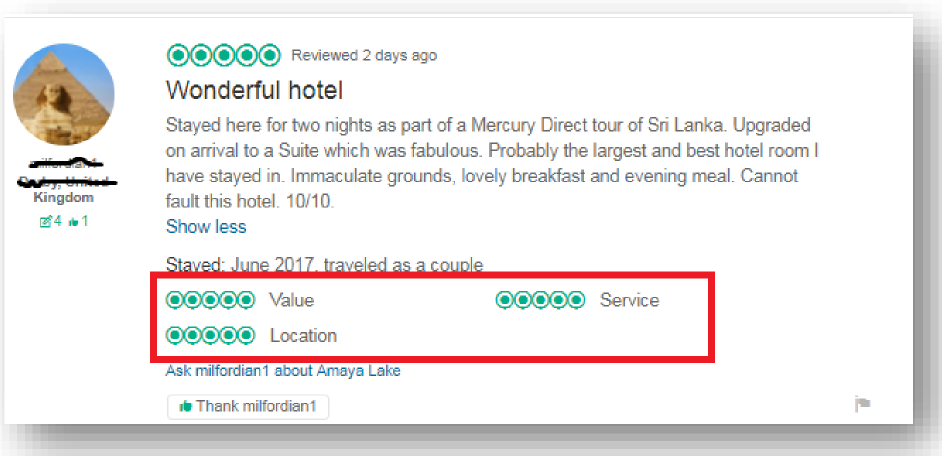

Figure 2.1- review text from tripadvisor.com

Benjamin Snyder has studied on ranking multiple aspects. The study has been focused on producing a set of numerical scores, one for each aspect. They have presented an algorithm that jointly learn ranking model for individual aspects by modeling the dependencies between ranks [11].

In [12], Titov et al. proposed a system to extract aspects and predict relevant rating value simultaneously. They have used topics that describe aspects and incorporated a regression model fed by the ground-truth ratings. How they have assumed is that training data set is provided with aspect ratings. The proposed statistical model called the Multi-Aspect Sentiment model (MAS) which can discover corresponding topics in review texts and extract textual evidence from reviews supporting each of aspect ratings.

2.4 Human preferred sentiment summarization model

There are various sentiment summarization models and [18] has performed a research on human evaluation of summarization models. They have evaluated results with different sentiment summarization algorithms. Their results indicated that humans prefer sentiment informed summaries over a simple baseline, suggesting the usefulness of modeling sentiment and aspects when summarizing opinions.

2.5 Aspect based summarization

However, existing works on aspect-based summarization [13, 14, 6, 8] only aimed at aggregating all the reviews and representing major opinions on different aspects for a given topic.

Minqing Hu et al., [13] has suggested a technique to summarize all the customer reviews of a product. It has been different from traditional text summarization because they have only mined the features of the product on which the customers have expressed their opinions and whether the opinions were positive or negative. They have not summarized the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization.

Recent work by Lu et al. [15] is the closest to ours, but their goal is still to generate an aggregated summary with aspect ratings inferred from overall ratings. Also, they have chosen product reviews. They attempted solution supporting unsupervised and minimum supervision of aspect data. As mentioned in the Figure 2.2, then input data represents what users normally can see through a community website which generally consist of large number of short comments with companion overall ratings. With such data, a user can only get an overall impression by looking at the average over all rating. It is infeasible to get over the large number of comments for more detail analysis.

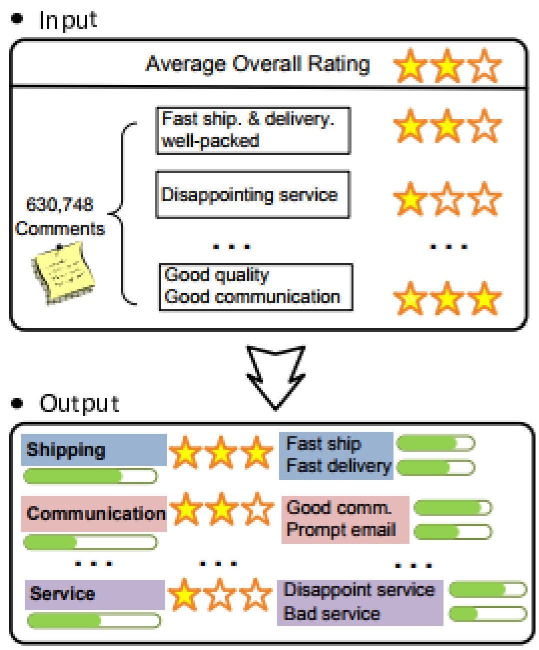

Figure 2.2 – Aspect – summary

2.6 Sentiment classification

Sentiment classification classifies opinion texts or opinion sentences as positive or negative [30] or sometimes as neutral. Classification methods can be divided in to two classes as supervised and unsupervised methods. Also, there are semi- supervised or minimum supervised approaches. Machine learning and lexicon based approaches are common in the latest state-of-art solutions.

2.6.1 Machine learning approach

Machine learning approaches are found under supervised approaches category among aspect based sentiment analysis tasks. Machine learning has been applied to identify aspect categories. Aspect categories are predefined set of classes and not explicit as aspect terms. In SemEval 2014 [1] task, had a subtask to detect aspect category.

e.g. “The hotel was expensive. But the menu was great.” Aspect categories are {price, food}.

When comparing results of machine learning approaches with unsupervised lexicon based approaches, Machine Learning has given results above the average [27]

2.6.2 Lexicon based approach

The lexicon based approach predict sentiment of review text using a database which contain opinion polarity (positivity, negativity) values. An example is SentiWordNet. SENTIWORDNET 3.0 is an enhanced version of lexical database that is supporting for sentiment classification and opinion mining applications. It has an automatic annotation process. SentiWordNet 3.0 consists of numerous number of synset terms along with polarity of each term.

Below is an equation that calculates sentiment polarity of a review texts.

$$S(D) = \frac{\sum w \in DS_w.weight(w).modified(w)}{\sum weight(w)}$$

Where S_w is a polarity score value of word w. S(D) is the sentiment polarity of the text and D is the text. This value is generated for dictionary using function weight(). Function modifier() handling , negation, word position in sentence etc.

2.7 Aspect terms Extraction

Aspect terms extraction is the fundamental task in fine-grained opinion analysis. There are many researches in aspect extraction, but still it is a challenging task. Aspect extraction is studied by many researchers in two different approaches. They are *supervised* and *unsupervised* [19].

In supervised aspect extraction approaches, predefined set of aspects word lists are used and search those aspects in review text. Ganu et al, [21] has done text review classification for

[1] http://alt.qcri.org/semeval2014/task4/

restaurant reviews and there they have identified Food, Service, Price, Ambience, Anecdotes, and Miscellaneous. Aspect terms among those six categories, first four are typical restaurant aspects and Anecdotes is to sentences which express user experience and not about quality. All other sentences that are not belong to fist five are added miscellaneous. In this study, analysis can be done only for identified features. If user's reviews contain more other features those are not included in the analysis.

Some systems have used manual feature annotations where domain specific features are annotated [22]. Those systems can perform well in different domains. But our work will be focused on hotel domain to solve its specific problems.

Later, aspect extraction is done with less human support as number of review texts are increasing. There are two techniques is finding aspects words from a collection of review texts. They are symbolic approach which rely on frequent noun phrases and statistical approach that finding terms tend to be close each other and reoccurring. Finding frequent words end up with too many terms that are not necessary and in the statistical approach tend to miss many low frequency terms. Additionally, using noun phrases is not enough to find aspects. Therefore [7] has considered other language components such as verbs and adjectives. They have employed supervised rule discovery to extract product features. They have first prepared a dataset by manually tagging large number of review text and then applied association rule mining to find all rules. Even though it is giving higher precision to the analysis results, with large number of data, manual tagging is not always feasible. Later studies have focused on algorithms that can identify aspects without manual tagging. [14]

In previous attempts, aspect identification was based on the classic information extraction approaches of using frequently occurring noun extraction [13]. When an aspect is tightly associated with a single noun, such approaches work well. But not giving better results when aspects incorporate many less frequent terms (e.g. the food aspect of a hotel can involve many dishes and meals), or when aspect is described without using any concrete noun at all. Common solution to this problem in the literature involves clustering using knowledge-rich methodologies. Employing rules that are manually created, sematic hierarchies, or both are some solutions [14, 20].

2.7.1 Frequent based approach

Hu & Liu, 2004 used frequent word extraction in their system to find frequent features. In their system, semantic orientation of identified frequent features are identified using WordNet. Hu & Liu has focused on product features and they are usually nouns or noun phrases in review texts. Therefore, they have used Part-Of-Speech tagging in natural Language processing to identify simple noun and verb groups. In this frequent feature extraction approach, it will only consider sentences which are explicitly mention feature words and not considering sentences that talk about features without mentioning feature words.

2.7.2 Syntactic Dependency

Jin & Zho has conducted a research to summarize movie reviews and they have used grammatical rules to identify feature and the corresponding opinion words as pairs. Feature-Opinion pairs are mined through some grammatical rules and the keyword list. Also, they have used statistical results on manually labeled reviews. A sentence can contain more than one feature word and opinion word. In order to find correct feature -opinion pair, they have used dependency grammar graph. This is further precise approach than frequency word extraction.

.

2.7.3 Supervised sequent labeling

Jakob and Gurevych, 2010, has evaluated feature opinion extraction through Anaphora Resolution (AR) instead of noun and noun phrases. They have tried to resolve “it” and “this” as anaphora candidates. They have implemented AR algorithm which has performed well when opinion targets are personal pronoun and not given high precision when resolving impersonal.

2.7.4 Topic Modeling

As Marrese et al. [31] expressed, that topic modeling is the simple way of analyzing large volume unlabeled data. Topic modeling effectively identifies word frequencies and concurrences. A “topic” consist of cluster of words that frequently occurs together. Topic models are able to connect words with similar meanings. Later researches applied topic modeling as an automatic way of identifying aspects.

Part-of-speech tagging and syntax tree parsing has used to find noun and noun phrases in the review text. Then they have employed frequent item set mining to extract most frequent nouns and noun phrases. Afterwards some special linguistic rules have been applied to filter nouns and noun phrases. But there can be non-frequent feature related noun and noun phrases which will not be identified in previous steps. Those non-frequent aspects are identified by finding noun or noun phrases that are appear near to opinion words with high frequency.

Bin et al. [23] has investigated about efficacy of topic model based approaches in multi-aspect sentiment analysis tasks. They have evaluated variations of topic modeling approaches for multi-aspect labeling and multi-aspect prediction. They have been proposing a weekly supervised approach that utilize minimal prior knowledge in the form of seed words. Later researches have been based on LDA (Latent Dirichlet Allocation) [24]. LDA has been applied with aim of uncovering latent topics in document collection and found that performing quite well. Earlier researches have used seed words to bootstrap aspect terms. [23] Has considered review texts which aspect ratings are available and aspect ratings are not available. Web sites like tripadvisor.com has review texts with aspect ratings but texts can contain more additional aspects and implicit ratings. Therefore, LDA variations can be considers in this research as well.

Further illustrating four types of LDA topic models which were evaluated in Bin's research.

LDA and local LDA

LDA is a probabilistic generative model in which documents are represented as mixture of over latent topics. LDA can effectively model word co-occurrence at the document level. But [25] has shown that aspects in a review texts are more likely to be discovered at the sentence level. Therefore, they have proposed local LDA at sentence level.

Multi-grain LDA

This is designed to overcome limitations in standard LDA for multi-aspect works. MG-LDA jointly model document specific global themes. Global themes are common throughout the corpus and correlates with ratable aspect words called 'local topics'. Local topic proportions are varied across the document.

Segmented topic models

STM [9] which jointly models document-level and sentence level topic propositions using two parameter Poisson Dirichlet Process (PDP). STM is an extension of local LDA and additionally consider document level topic distribution.

Inference

While exact inference for the models just presented is largely intractable [24], approximate techniques such as variation inference or Gibbs sampling can be used instead. They have used collapsed Gibbs sampling approach for inference.

Finally, as the evaluation results, they have found that there is a higher precision for local LDA than other variations. Also implies that finding sentence level co-occurrence in more suitable for studying aspects from review texts. Because most sentences usually focus on just one or two aspects and whole review text is not about one aspect.

2.7.5 Unsupervised approaches

Unsupervised approaches of aspect extraction have been studied by many researches. Key advantage of unsupervised dependency base methods is that it is not require any human labeled data. They are based on opinion words and aspect words (e.g. friendly staff). By considering such relationships DP (Double Propagation) is introduced as an unsupervised dependency based method by qui et al. [35]. Also, they have showed that DP can perform better than CRF (Conditional Random Fields) [35].

Further illustrated aspect extraction in a different approach [23]. They have used set of opinion words to extract set of aspects words. Then those extracted aspect set has been used to extract new aspect words. This approach is much like the DP.

Below Figure 2.3 further illustrate the concept.

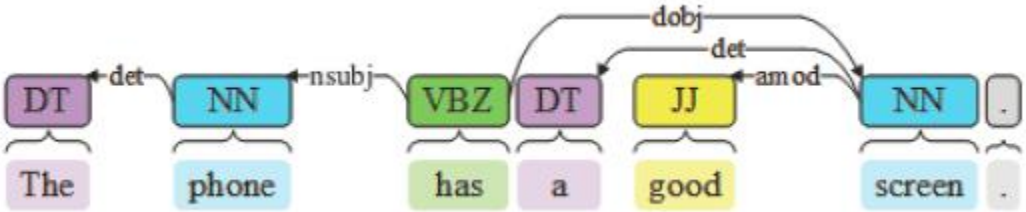

Figure 2.3 – DP Concept

"If a word A, whose part-of-speech (POS) is a singular noun (nn), has the dependency relation amod with (i.e., modified by) an opinion word O, then A is an aspect."

The research shows that DP can further improve if a system can learn from past and then use that knowledge for extraction. Then the results are much better. They have been used recommendation for aspect extraction. They have used large number of reviews of other products to extract aspect for the current product. So, their base id DP and then further improved with recommendation.

When extracting aspects, they have used a word vector trained from large corpus of 5.8 million reviews for similarity comparison. For an example extracted word "picture" has used to recommend "photo' as an aspect word. To identify correlations of aspects, association rule mining is employed on reviews of other products in the web [36]. Aim of applying association rule mining is to generate association rules and use them to recommend specific aspects rather than general aspects to all products.

Word vector and association rules learnt from corpus Hu and Liu [13] were two forms of recommendation approaches that they have used and outperformed with state-of-art dependency rule based methods of those times.

Later unsupervised aspects extraction approaches are mainly based on topic modeling [34]. Then syntactic rules are defined using dependency relations. Topic modeling only gives some rough topics rather than precise aspect or topical term.

Summary

Many researches are found in the literature in sentiment analysis area with user generated data in the web. Approaches can be identified as supervised and unsupervised approaches. Aspect based sentiment analysis is targeting, summarizing aspect ratings. Machine learning and lexicon based approaches are used for sentiment classification. Aspect terms extraction is done using frequent based approaches, supervised sequent labeling and various topic modeling techniques. Double propagation is introduced as an unsupervised dependency based method for aspect term extraction.

Chapter 3

Analysis and Design

This chapter includes a critical analysis of the requirement and will present the design of proposed solution to the ABSA in the hotel domain. This design highlights the aspect term extraction as an unsupervised approach and natural language processing combined with review text preprocessing task.

This chapter is divided into multiple subsections. Section 3.1 will be an analysis of the problem and explain the requirement. Section 3.2 is the overall architecture of the system. The main objective is to identify major components of the proposed solution. Section 3.3 explains each component in detail. Section 3.4 will be about creating the data set.

3.1 Problem Analysis

The major problem being addressed is an aspect based sentiment analysis for hotel reviews. A review text contains information about more than one aspect. Then, to a person who tries to get an overview of the hotel, based on an aspect, unable to decide by reading several reviews. The proposing system tries to summarize the actual sentiment value of an aspect by analyzing all review texts for a hotel. As in Figure 3.1, a review has three major components.

1) Overall rating which a user can give when writing the review.

2) Review text which user explain the thoughts and experience during the stay. It is the key in our analysis.

3) Aspect rating, is a predefined set of aspects that user can rate when entering the review. It is provided by tripadvisor.com.

This research focuses on the text where it contains addition aspect terms and sentiments about those aspects. In this example user has not rated for “Foods”. But in the review text, user has mentioned “Breakfast”. And, “pool is good”. Therefore, it is important and reasonable to do a deep analysis of review text where can reveal many interesting information for both customers and hotel management of the hotel than relying on predefined set of aspect terms provided by the website.

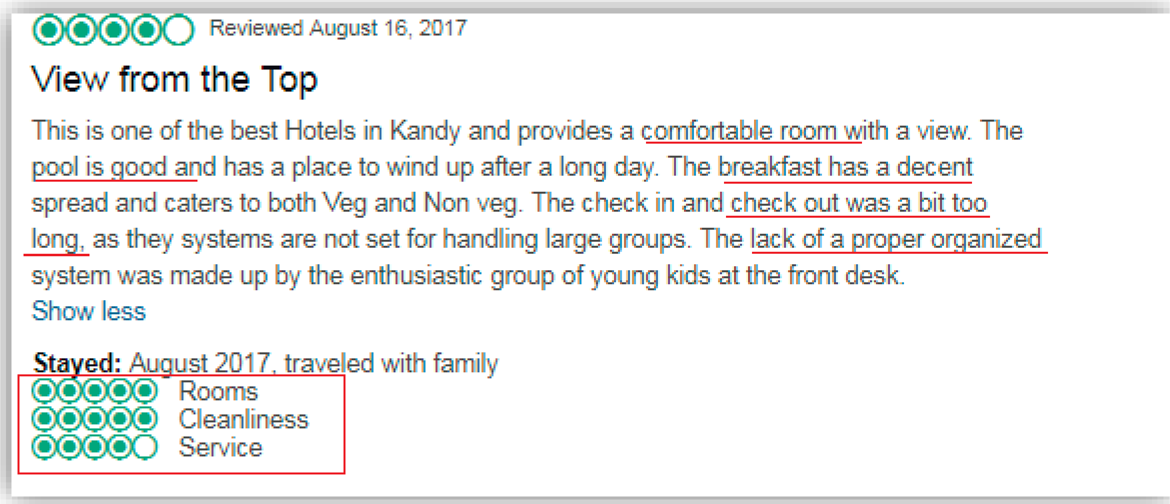

Figure 3.1 – Example of a hotel review from tripadvisor.com

3.2 Research Method

The main objective of this research is to develop a methodology to facilitate aspect based sentiment analysis for hotel reviews using unsupervised approach. This research was conducted investigating several related works and techniques used for unsupervised approaches. A literature survey was used to identify requirements and propose a design. Came across this approach by analyzing several approaches in the literature.

3.3 Design Approach

According to previous works discussed in the previous chapter, there are several improved versions of unsupervised approaches which can be employed across different domains Hu and Liu [13]. Many researches done with supervised and semi-supervised based approaches. And many of them were requiring domain knowledge and predefined aspect words to get results with higher precision. That is why we wanted to focus on unsupervised design which require less domain knowledge and less human intervention. Then, wanted to identify possible ways of improving the accuracy of sentiment analysis.

The following design selected after analyzing several methods discussed in the literature. Liu and Hu [7] had pointed three types of review formats and suitable approaches. Hotel reviews are collection of full sentences and no predefined format for text. Reviewers can freely express their ideas. For those kinds of reviews can be analyzed through unsupervised approaches more effectively. Also in many cases, full sentences are presented and sentiment analysis can be performed at sentence level.

3.4 Design Assumptions

Before arriving to the actual design, there are few limitations and constraints to be concerned as we are dealing with natural language. As mentioned in previous studies, below are some of the limitations to concern when designing a model for sentiment analysis task.

This research is dealing with review text which are only in English language. Most of the online user generated content containing many misspelled words, grammatical errors, incomplete words, abbreviations, playful words known as internet slangs, syntactically incorrect sentences. But this study assumes that sentences in reviews are syntactically correct. Reviewer's location or country is not counted and not considered about different expectation levels. Another design assumption is that all reviews are independent of observations.

Below Figure 3.2 is the proposed design based on assumptions. There is a separate module to collect data from the website. Even though there are texts in different languages, only the English language is considered. When extracting the text, language is not much a concern. But in data preprocessing steps, preprocessing techniques will have a concern with language. Techniques can be different from language to language.

Another assumption is that reviewers are in the same level of expectation. There is a high chance to having different expectation levels based on county, religion and nation. But, this research would not count those factors.

3.5 Proposed design

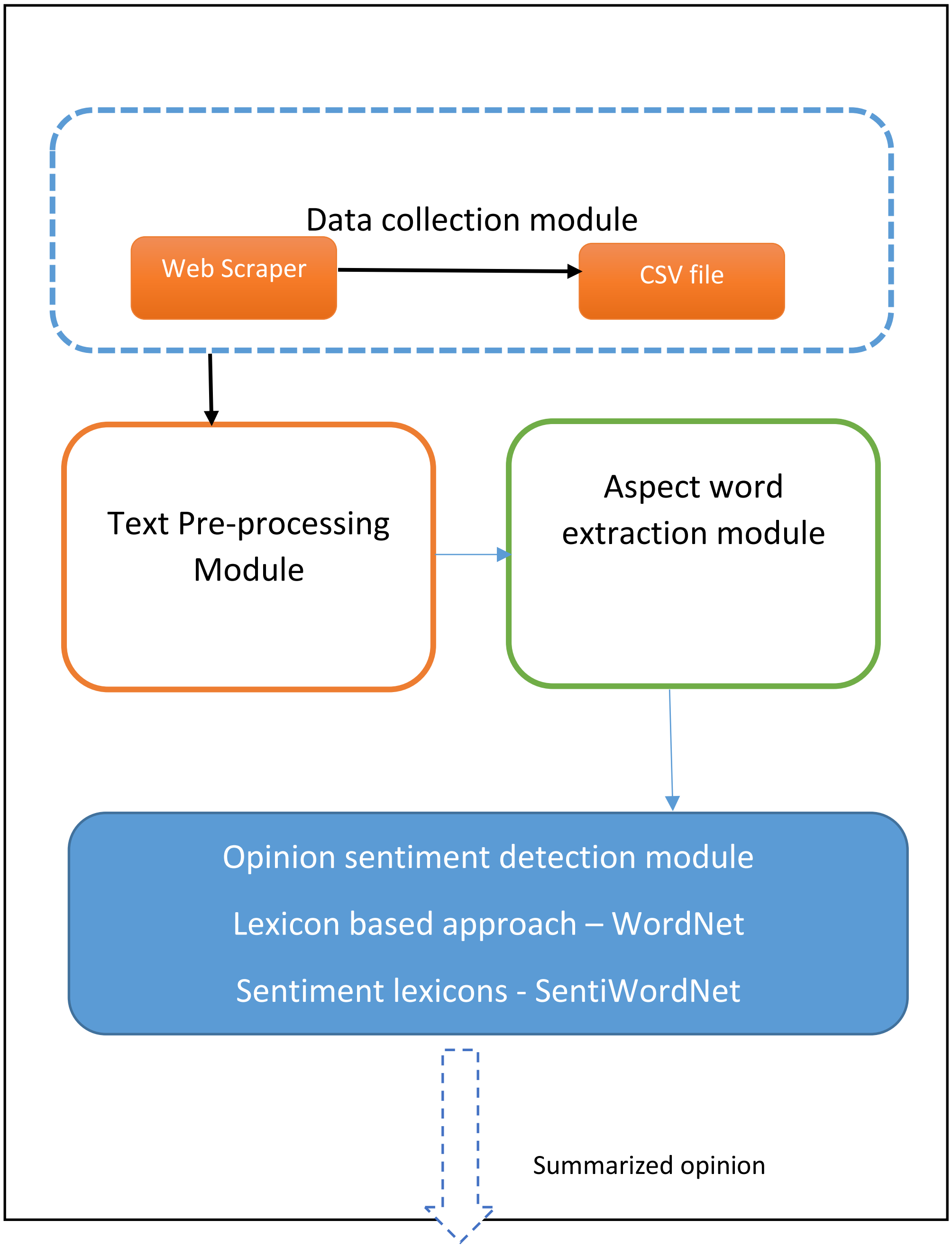

Figure 3.2 – General system design

As shown in the Figure 3.2, overall system contains four major components. They are data collection module, Text preprocessing module, Aspect word extraction module and Opinion sentiment classification module.

Each of the components are described in detail in next subsection.

3.6 system components

3.6.1 Data Collection module

This module is to manage review text extraction from the website. In this research it is tripadvisor.com and extracting review texts for a given hotel and store in a csv file. The sub component web scraper is to retrieve data from tripadvisor.com. It automatically browses through each pages of the review section and read the complete review text for each review. And then, store each text in a CSV file.

3.6.2 Text preprocessing Module

CSV file contains very noisy review texts. Therefore, preprocessing techniques will be employed to reduce noise and it will make faster in further steps of the analysis. First, each review text should be **split in to sentences** as topic modeling would be applied to sentence level rather than considering whole review text. Because a review text can contain information about more than one aspect (topic).

Preprocessing techniques and the order of applying techniques depend on the application and the modeling method Mathew and Arthur [29]. That research was about selecting a matching preprocessing specification to unsupervised learning approach. **Removing punctuation** is one preprocessing task which remove all non-letter characters from the document. Special characters, any markup (html tags), extra white spaces, and punctuations are things that would be removed. As our modeling method is interested only in texts, we can remove punctuations at the preprocessing step. Also, numbers in the text will not provide much information in our domain. Therefore, we can **remove numbers** as well. Another preprocessing step that would be better to be applied is **lowercasing** letters. **Stemming** is another preprocessing step that can be employed. Stemming comes under standard text preprocessing pipeline which is often employed as a vocabulary reduction technique. For an example the words “party”, “partying” and “parties” all share a common stem “parti”. So, in our context also there can be such words. Therefore, stemming would be useful to apply to our review text collection as a preprocessing step. Stop words removal is the next preprocessing task. Some words in a sentence referred to as “stop words” and do not having much information. They are consisting of function words such as “the”, “it”, “and” and “she”. Also include some domain-specific examples such as “hotel”. There is no standard or guideline to list stop words. Then an existing NLP technique will be applied to identify and remove stop words of English language texts.

Most commonly used preprocessing procedure in unsupervised approaches was, remove punctuation, numbers, do stemming, lowercasing and remove stop words.

3.6.3 Aspect word Extraction module

This is the step which unsupervised modeling going to be employed. This module is going to identify aspect terms discussed in review texts. Feature extraction is based on word frequency. Because this is focusing on unsupervised approaches. In this step, frequency of nouns is calculated along with the noun and create a list of word, frequency pairs. This the list is sorted from high number of frequency to low number of frequency. Then the candidate aspect word list is generated by defining a threshold value for the frequency. There can be synonyms of words in the list. For an example word “food” is extracted and another word “breakfast” also listed. Both are related to “food” aspect. Therefore, a manual inspection needed to be carried out to make the candidate aspect terms list. Hu and Liu [13] has firstly proposed a similar approach based on association rule mining. The occurrence of nouns and noun phrases are calculated and only the frequent nouns are kept as aspects. The same approach is used here with minimal manual interaction.

3.6.4 Opinion sentiment classification module

opinion words and polarity of the opinion should be identified. Sapna and Paul [33] fist identify opinion words. Adjectives in a sentence can be an opinion word. Then SentiWordNet or WordNet will be used to identify polarity of opinion words. For each sentence aspect term opinion word and the polarity will be stored in database. Then those automatically annotated data will be used in summary generation for a given hotel.

Summary

Aspect based Sentiment Analysis is focusing on generation an aspect level summary by considering all reviews for a hotel. Research aims to follow an unsupervised approach with a narrowed down scope. A design is proposed by assuming that reviews are syntactically correct. The design contains three main modules including a module for data collection, text preprocessing, aspect word extraction and sentiment detection for aspects.

Chapter 4

Implementation

This chapter discuss how the implementation works carried out to experiment the concept discussed in chapter 3. This will discuss about the dataset that was used to do the experiment, and how data set was prepared, implementing web scraper for collecting data, how the aspect terms extracted. We performed this experiment using publicly available tools as well as self-implemented tools. This chapter will highlight about tools, algorithms used, technologies and evaluation matrices used.

4.1 Data Sets

Dataset consists of 5864 review texts extracted form tripadvisor.com which is a very famous hotel booking web site all over the world. Dataset is a collection of reviews from seven different hotels in Sri Lanka. Review texts are extracted using a web scraper implemented using python and selenium web driver. Dataset contains reviews for seven hotels listed in tripadvisor.com until September 2017.

A review in tripadvisor.com contains many information. This research interested only the review texts. The web scraper capable of extracting only the text part for all reviews for a given hotel.

Figure 4.1 explains a structure of a review in tripadvisor.com.

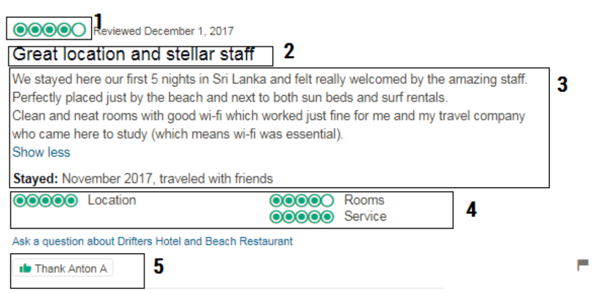

Figure 4.1 – Example of a tripadvisor.com hotel review

1. Overall rating

2. Title for a review entered by user

3. Review text where people have expressed their idea

4. Aspect based rating. These aspects are predefined

5. reply from hotel

This research was targeting to extract only the review text (3) through the scraper. About 5800 texts were extracted and stored in CSV format. Each raw contains the raw number and the review text.

The whole corpus was used when extracting aspect terms. But, there were limited number of reviews are used for evaluation.

100 of review texts which selected randomly from large data set were manually annotated to test the model. Each sentence of review texts is annotated with aspect term and positive or negative value of the aspect term.

According to the annotator agreement, either sentence or part of sentence in the review would be positive or negative. And if there are sentences that would describe anything else that would not have related to an aspect term listed, would be ignored. Sentences with multiple aspect terms also annotated. For an example, “The restaurant was expensive, but menu was good” would be annotated for ‘price’ and ‘food’ separately. Annotators were asked to annotate every explicit aspect as well as implicit aspects. If there are multiple sentences per review with same aspect were clustered.

The annotation procedure which was applied had better results proven and had used in a previous work to annotate hotel reviews [31]. Each review text was annotated following a set of guidelines and by a person who has computer science background. When there is a sentence found in a review which seems ambiguous and hard to find the aspect term, then had a discussion with linguistic expert and come to an agreement. This procedure has made a difference in studies with other common annotating procedures. Other approach is to do the annotation to the same set of reviews by two annotators and at the end, compare annotated reviews and define a final choice. This different approach is used here due to time constraints.

Table 4.4 is an example for annotated texts. All review texts were annotated using the extracted aspect term lists. Section 4.3 in the chapter 4 will describe, how aspect terms were extracted. Annotated 100 review texts will be used as test data set for the evaluation process.

Review text	Aspect, sentiment, sentences
"Lovely hotel that overlooks the beautiful Indian Ocean. Excellent location and ambience (the xmas deco is up and it is very cheery). Hotel provides free wifi too, enabling to connect w friends and family."	[[**view**, positive,[Lovely hotel that overlooks the beautiful Indian Ocean.]], [**place**, positive,[Excellent location and ambience (the xmas deco is up and it is very cheery).]], [**wifi**, positive, [Hotel provides free wifi too, enabling to connect w friends and family.]]]
"I loved this hotel. Cheerful decor, an extremely comfortable bed, and a beautiful view of the lotus pond, pool and sea right in front of my room. Staff are really lovely and polite. Wasn't too impressed that you could only have the buffet for breakfast, a la carte would have been good for small breakfast eaters. This place is ideal for all travelers, lots of room for kids to run around as well. The beach is clean and not at all crowded. Overall, I loved my stay there, will definitely be back."	[[**place**, *positive*, [I loved this hotel, This place is ideal for all travellers, lots of room for kids to run around as well.]], [**view** , *positive* , [Cheerful decor,and a eautiful view of the lotus pond pool and sea right in front of my room, The beach is clean and not at all crowded]], [**room**, *positive*, [an extremely comfortable bed.]], [**staff**, *positive*, [Staff are really lovely and polite.]], [**food**, *negative*, [Wasn't too impressed that you could only have the buffet for breakfast a la carte would have been good for small breakfast eaters.]], [**stay**, *positive*, [Overall I loved my stay there, will definitely be back.]]]

Table 4.4 - Example of annotated reviews

4.2 Tools and programming languages

This section provides and overview of tools and programming language used. In this research, mainly used programming language is python and many python libraries that are available to natural language processing were used. Tools needed to collect data and aspect extraction was self-implemented tool.

4.2.1 Python

Python is a general-purpose programming language, which is very power full for text analysis tasks. There are many in built libraries available in python for natural language processing. I've used NLTK for python, genism, TextBlob for text processing. Also used python to implement web scraper to collect data. Aspect extracting also done by a tool implemented by python. Also, the training model also implemented by using python.

4.3 Language resources

Apart from the data set that is collected and annotated, some other linguistic resources were used. These resources include synonyms for words and sentiment lexicons.

4.3.1 Word Net

WordNet is a lexical database for English language. It groups words in to set of synonyms called synsets. I've used WordNet to identify synsets and then match synset with words in review test to identify aspect words. WordNet is not domain specific and may not contain all synonyms words in hotel domain. A similarity can be measured between two synsets. Similarity is a value indicating how much similar two words. That functionality available for the python implementation for WordNet.

4.4 Sentiment lexicons

A sentiment lexicon is a mapping from words to an association score corresponding to positive or negative sentiment. Such list can be created from annually and annotated data or automatically annotated data.

4.4.1 SentiWordNet

SentiWordNet is a lexical resource for opinion mining. Each sysnset of WordNet are assigned with positivity, negativity and objectivity scores. I’ve used SentiWordNet in this research to get the polarity of sentences in reviews. I considered positivity and negativity but not the objectivity.

4.5 Experimental Flow

This section explains steps involved in experiment process. Throughout the process data collected from tripadvisor.com and then select a random dataset to annotate manually and then same dataset is used as an input to the program which is implemented to generate summary of aspects. All the steps that program follows are explained below.

4.5.1 Data collection

This study requires a collection of review texts from the hotel domain. I have selected **tripadvisor.com** web site which is very popular when searching for places to visit. And, especially for hotels all over the world. It has information about places, hotels all over the world and people can enter reviews on those places and hotels. Therefore, people those who are planning a travel can visit this web site to see what other’s opinions. A review text in tripadvisor.com contains an overall star rating for the hotel, and star ratings for seven aspects which have already defined. Every review does not contain that aspect based rating but contain a text expressing people’s experience at the hotel and their feedbacks.

As this research is interested only the text part or user’s opinion in a review and a program required to extract the text from reviews of multiple hotels. I selected five different hotels in Sri Lanka located in different areas in the country.

The implemented program is capable of extracting review texts data from web sites. Web scraper implemented using selenium library for python and web driver. Program can extract all review texts for a give URL of hotel page in tripadvisor.com. For an example, if a hotel has 100 user reviews, they are included in 50 pages where each page contains 5 reviews. Then scraper go through all pages and extract review texts. Then extracted information are written into a csv file. Review data collected from 7 different hotels in Sri Lanka was stored in 7 different .CSV files. Later all files were combined and added to a single file which contains 5864 review texts.

4.5.2 Data preprocessing

All texts are in English language and required some cleanups for the text before extracting aspect terms and sentiment values. It is very important to apply some preprocessing steps to reduce size of corpus. When identifying aspects in review text, program is parsing through each word. Therefore, preprocessing can reduce time taken by the program to aspect identification.

Simple pre-processing steps were performed to tokenize in to sentences, lowercase, remove stop words and remove punctuation marks.

4.5.2.1 Tokenizing

Python Implementation for nltk is used. In this Step, each review was tokenized in to sentences. I’ve used nltk *sen_tokaniz (text)* function. All the unnecessary spaces are removed and review text is tokenized in to a list of sentences.

4.5.2.2 Lowercasing

Another simple preprocessing step taken to is Lowercasing. Each word in the review text is converted in to lowercase. The reason for doing this is that the first letter of the sentence starts with an uppercase but not always. And converting a word in to lower case doesn’t has impact to the meaning. “Elephant” and “elephant” is same meaning but it can impact on counting in the analysis if it would be counting as two separate words.

4.5.2.3 Remove stop words

In text Processing, there are some words referred as “stop words” which are unlikely convey any information, for example “the”, “it”, “he” can be considered as stop words. I’ve used nltk stop word corpus and all the words in that nltk stop world list are searched in the corpus and removed. Here considered stop word “English” for each tokenized sentence in the review, stop words are removed. Now, analysis task would be much efficient as it reduced the numbers of words it is parsing through.

4.5.2.4 Remove punctuation marks

A library called “genism” for python is used to remove punctuation maker in each review sentences. A method called *strip_punctuation (sentence)* is used with a sentence as punctuator. It is very first task in research to filter out valid text and remove special characters ($, #, *, &, %, etc…) and punctuation marks. Then none letter characters are not informative for this research.

4.5.3 Aspect words extraction

Aspect word extraction is key in aspect based statement analysis task. Unsupervised based approach was applied to identify aspect terms from text. Whole coypus with 5864 review texts was used for this task. A frequency based approach is selected to be employed to this corpus to find most frequent words as aspect terms. This is an adoption of HU and Liu [13]. According to Hu and Liu frequent nouns and noun phrases can be identified as aspect terms. Anyway, using noun phrases tend to be produce lots of none aspect terms. But in this research, a threshold value is set for frequency value and words that are having more than 150 occurrences are considered as candidate aspect terms.

Nouns and noun phrases in the review are identified using part-of-speech tagging techniques in NLP. This research has used nltk part of speech tagger to tag words in review text with noun and noun phrases.

4.5.3.1 Frequent word identification

The output of the part of speech tagger is used to get frequencies of tagged words and phrases. Similar words are counted and added to a list as a key value pair. Key is the word or phrase and the value is number of occurrences. A program implemented using python, able to do the task. The output list is saved in a text file to be used as the input of candidate aspect list identification. Figure 4.2 is screen capture of output word list.

4.5.3.2 Preparing candidate aspect word list

The outcome from the frequent words identification process is very big list of words along with frequency. There was a need of filter aspects, as all the words are not meaning full as aspect terms and there are semantically similar words with multiple frequencies.

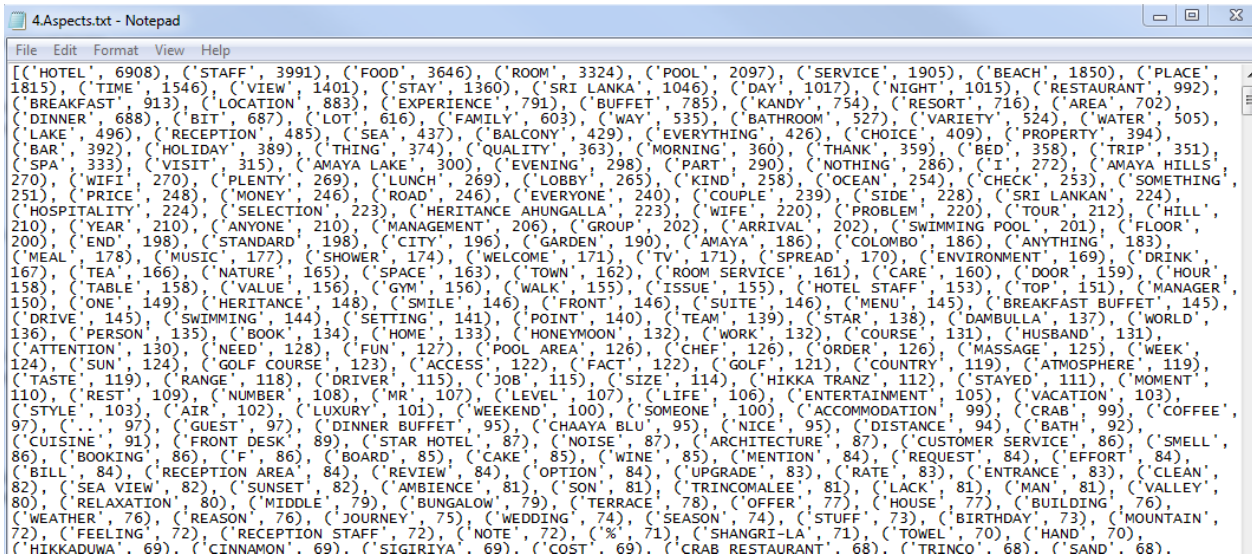

Figure 4.2 – screen capture of aspect word list

Figure 4.2 is part of extracted aspect terms before applying any filtering. 5304 words listed. It has reported all nouns and noun phrases. Words like 'room', 'bed', 'bathroom', 'balcony' are related to 'room'. Those words are mentioned in reviews when mentioning about room. But in the aspect list, each of related words have separate frequencies but can be cluster into one group. There is an ability to do clustering to the words list and identify different clusters. On cluster can represent one aspect. But in this method, not applied clustering and selected words by defining a threshold value for the frequency. The threshold vale was set to 100 for this instance. When analyzing the aspect list, words with low frequency found as names of people, names of places, or a word that cannot be interpreted as an aspect.

For an example, below is an example for words with less frequency,

('CINNAMON', 69), ('SIGIRIYA', 69), ('COST', 69), ('CRAB RESTAURANT', 68), ('TRINCO', 68), ('SAND', 68), ('TOUCH', 68), ('LOOK', 68)

In this list, 'COST' can be taken as an aspect terms, but other words are does not mean any aspect when it appears as a single word. 'CRAB RESTAURANT' can be considered as a word mentioned about 'FOOD'. But 'FOOD' has higher frequency and get as an aspect term.

Below are the words with more than or equal 100 times occurrences in the whole corpus,

('HOTEL', 6908), ('STAFF', 3991), ('FOOD', 3646), ('ROOM', 3324), ('POOL', 2097), ('SERVICE', 1905), ('BEACH', 1850), ('PLACE', 1815), ('TIME', 1546), ('VIEW', 1401), ('STAY', 1360), ('SRI LANKA', 1046), ('DAY', 1017), ('NIGHT', 1015), ('RESTAURANT', 992), ('BREAKFAST', 913), ('LOCATION', 883), ('EXPERIENCE', 791), ('BUFFET', 785), ('KANDY', 754), ('RESORT', 716), ('AREA', 702), ('DINNER', 688), ('BIT', 687), ('LOT', 616), ('FAMILY', 603), ('WAY', 535), ('BATHROOM', 527), ('VARIETY', 524), ('WATER', 505), ('LAKE', 496), ('RECEPTION', 485), ('SEA', 437), ('BALCONY', 429), ('EVERYTHING', 426), ('CHOICE', 409), ('PROPERTY', 394), ('BAR', 392), ('HOLIDAY', 389), ('THING', 374), ('QUALITY', 363), ('MORNING', 360), ('THANK', 359), ('BED', 358), ('TRIP', 351), ('SPA', 333), ('VISIT', 315), ('AMAYA LAKE', 300), ('EVENING', 298), ('PART', 290), ('NOTHING', 286), ('I', 272), ('AMAYA HILLS', 270), ('WIFI', 270), ('PLENTY', 269), ('LUNCH', 269), ('LOBBY', 265), ('KIND', 258), ('OCEAN', 254), ('CHECK', 253), ('SOMETHING', 251), ('PRICE', 248), ('MONEY', 246), ('ROAD', 246), ('EVERYONE', 240), ('COUPLE', 239), ('SIDE', 228), ('SRI LANKAN', 224), ('HOSPITALITY', 224), ('SELECTION', 223), ('HERITANCE AHUNGALLA', 223), ('WIFE', 220), ('PROBLEM', 220), ('TOUR', 212), ('HILL', 210), ('YEAR', 210), ('ANYONE', 210), ('MANAGEMENT', 206), ('GROUP', 202), ('ARRIVAL', 202), ('SWIMMING POOL', 201), ('FLOOR', 200), ('END', 198), ('STANDARD', 198), ('CITY', 196), ('GARDEN', 190), ('AMAYA', 186), ('COLOMBO', 186), ('ANYTHING', 183), ('MEAL', 178), ('MUSIC', 177), ('SHOWER', 174), ('WELCOME', 171), ('TV', 171), ('SPREAD', 170), ('ENVIRONMENT', 169), ('DRINK', 167), ('TEA', 166), ('NATURE', 165), ('SPACE', 163), ('TOWN', 162), ('ROOM SERVICE', 161), ('CARE', 160), ('DOOR', 159), ('HOUR', 158), ('TABLE', 158), ('VALUE', 156), ('GYM', 156), ('WALK', 155), ('ISSUE', 155), ('HOTEL STAFF', 153), ('TOP', 151), ('MANAGER', 150), ('ONE', 149), ('HERITANCE', 148), ('SMILE', 146), ('FRONT', 146), ('SUITE', 146), ('MENU', 145), ('BREAKFAST BUFFET', 145), ('DRIVE', 145), ('SWIMMING', 144), ('SETTING', 141), ('POINT', 140), ('TEAM', 139), ('STAR', 138), ('DAMBULLA', 137), ('WORLD', 136), ('PERSON', 135), ('BOOK', 134), ('HOME', 133), ('HONEYMOON', 132), ('WORK', 132), ('COURSE', 131), ('HUSBAND', 131), ('ATTENTION', 130), ('NEED', 128), ('FUN', 127), ('POOL AREA', 126), ('CHEF', 126), ('ORDER', 126), ('MASSAGE', 125), ('WEEK', 124), ('SUN', 124), ('GOLF COURSE', 123), ('ACCESS', 122), ('FACT', 122), ('GOLF', 121), ('COUNTRY', 119), ('ATMOSPHERE', 119), ('TASTE', 119), ('RANGE', 118), ('DRIVER', 115), ('JOB', 115), ('SIZE', 114), ('HIKKA TRANZ', 112), ('STAYED', 111), ('MOMENT', 110), ('REST', 109), ('NUMBER', 108), ('MR', 107), ('LEVEL', 107), ('LIFE', 106), ('ENTERTAINMENT', 105), ('VACATION', 103), ('STYLE', 103), ('AIR', 102), ('LUXURY', 101), ('WEEKEND', 100), ('SOMEONE', 100)

The word 'STAFF' has resulted with high frequency as it appears in many reviews. 'MANAGEMENT', 'HOTEL STAFF' and 'MANGER' are domain synonyms and focusing on 'STAFF'. Therefore, domain synonyms can be grouped together. Then I have applied a semi-supervised based approach to group nouns and noun phrases. Domain synonyms are grouped together and reduced the list into smaller count as it easy to be used by further analysis works. The grouping was done manually by using the domain knowledge. But grouping can be automated by applying a topic modeling algorithm to the word list to get dominant topics as aspects.

At the end, an aspect list was produced with minimal words after grouping. Table shows final aspect list

1	('STAFF', 3991),('KIND', 258),('HOSPITALITY', 224),('MANAGEMENT', 206),('CARE', 160),('MANAGER', 150),('SMILE', 146),('ATTENTION', 130),('CHEF', 126), ('HOTEL STAFF', 153)
2	('FOOD', 3646),('LUNCH', 269),('MEAL', 178),('DRINK', 167),('TEA', 166),('SUITE', 146),('MENU', 145),('BREAKFAST BUFFET', 145),('VARIETY', 524),('BAR', 392),('RESTAURANT', 992)
3	('ROOM', 3324),('BED', 358),('BALCONY', 429),('FLOOR', 200),('TV', 171),('SPACE', 163),('TABLE', 158),('SIZE', 114),('SHOWER', 174),('BATHROOM', 527)
4	('POOL', 2097),('SWIMMING POOL', 201),('SWIMMING', 144),('POOL AREA', 126)
5	('SERVICE', 1905),('ROOM SERVICE', 161),('ORDER', 126)
6	('PLACE', 1815),('ENVIRONMENT', 169),('GARDEN', 190),('NATURE', 165),('ATMOSPHERE', 119),('STYLE', 103),('SETTING', 141)
7	('VIEW', 1401),('LAKE', 496),('BEACH', 1850),('SEA', 437),('OCEAN', 254),('SUN', 124)
8	('STAY', 1360), ('HONEYMOON', 132),('FUN', 127),('LIFE', 106),('REST', 109),('HOLIDAY', 389),('VISIT', 315),('TOUR', 212),('VACATION', 103),('WEEKEND', 100),('WEEK', 124)
9	('WATER', 505)
10	('RECEPTION', 485),('LOBBY', 265),('CHECK', 253),('ARRIVAL', 202),('WELCOME', 171),('BOOK', 134)
11	('SPA', 333),('MASSAGE', 125)
12	('WIFI', 270)
13	('PRICE', 248),('MONEY', 246),('VALUE', 156)
14	('GYM', 156),('GOLF COURSE', 123),('GOLF', 121)

Table 4.2 - Aspect categories

In this aspect identification approach, it requires a minimal prior knowledge at the beginning. When extracting words, it does not require any domain knowledge and can applied this approach to any domain. Only grouping involves some supervision but it can be reduced by using a topic modeling approach.

In this scenario there are 14 number of different groups were identified. The number 14 is not a fixed value for the number of groups. That was gained by reducing the word list as much as possible and grouping into groups as much as distinct. If the number of distinct groups are high it is difficult to evaluate. Because evaluation is done by using a manually annotated dataset.

Figure 4.5 summarize the process of aspect word extraction.

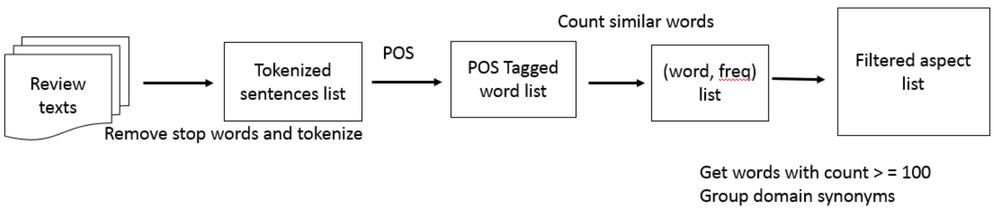

Figure 4.5 - process of preparing candidate aspect list

4.5.4 Aspect category detection in reviews

The problem of aspect category detection can me modeled with multi – label classification methods. Because each review belongs zero or more aspect categories. In this case rather than classifying using a classification model it assigns labels based on review sentences. Multi-label classification problem can be solved using two approaches, as: i). binary relevance approach, ii). Label power set approach [32] In binary relevance approach, it first building n distinct model for each n unique labels. The final prediction of n models. In this research also, individual sentences are classified in to aspect categories and the label of the review is decided by combining aspect categories of sentences. Which is a binary relevance approach. But in, label power set approach each label combination is treated as unique label.

The implemented model is targeting to identify aspect words in reviews and the sentiment for identified aspects. First all aspects of review are extracted. When a review text is given as an input, it is tokenized in to sentences and then each then stop words and punctuation marks are removed. Now the preprocessed sentences are further tokenized in to words and each word is compared with

WordNet synsets (WordNet organize words in synonym sets, called synsets) of aspect words to identify aspects. If this process further illustrated below in steps:

1. Tokenize reviews into sentences
2. Preprocess each sentence to remove stop words and punctuation
3. For each aspect word in the aspect list, get path similarity for a sentence.
4. Word with highest similarity is identified as the dominant aspect word for that sentence.
5. Path similarity is calculated using python implementation for WordNet.
6. *path_similarity (synset_1, synset_2)* was the method used.
7. *synset_1, synset_2* are synsets for aspect words and a word in a sentence

```
#get the maximum path similarity between an aspect word
sentence words
for sentence in doc_set:
            line=[]
            line=str(sentence)
            sim=[]
            for word in line.split():
                w={}
                for aspect in aspect_list:
                    #print(get_best_synset_pair(word,aspect))
                    w[a]=get_best_synset_pair(word,aspect)
                max_sim=max(w.values())

def get_best_synset_pair(word_1, word_2):
        max_sim = -1.0
#get synset from WordNet
    synsets_1 = wn.synsets(word_1)
    synsets_2 = wn.synsets(word_2)
    if len(synsets_1) == 0 or len(synsets_2) == 0:
        return 0
    else:
        max_sim = -1.0
        best_pair = None, None
        for synset_1 in synsets_1:
            for synset_2 in synsets_2:
               sim = wn.path_similarity(synset_1, synset_2)
               if sim != None:
                   if sim > max_sim:
                       max_sim = sim
                       best_pair = synset_1, synset_2
        return max_sim
```

4.5.4.1 Similarity distance

Similarity distance is a measure of how similar two-word senses are. In WordNet, similarity score can be determined based on shortest path that connects the sense in is-a (hypernym/hyponym) taxonomy. The score is in the range 0 to 1. *Path_similarity ( )* is the method that returns the similarity score of two words in python WordNet. A score of 1 represents that two words are identical.

4.5.5 Aspect level sentiment classification

When an aspect word is found, a sentiment lexica (SentiWordNet) is used to identify the polarity value to detect sentiment of the sentence. When more than one sentence is found in the review for the same aspect, the polarity value of each sentence in aggregated and consider aggregated polarity as the overall polarity value for that aspect word. Then the sentiment is decided based on the polarity value. If polarity value is positive, then sentiment is positive. If the polarity value is negative then the sentiment is negative. Output of a review is annotated sentences with aspect term and polarity. Final summary is generated aggregating all aspect polarities.

Input:

Friendly multilingual staff. Excellent pool and amazing views. Food is simply fantastic. Definitely worth the price. We had a very relaxing time. I would recommend the full package (meals)). Unluckily the beach was too rough for us to swim in but we still took awesome pictures. But room had uncleaned dust and quality was not as expected.

Output:

[[**'staff'**, 'positive', 0.38, ['friendly multilingual staff ']], [**'place'**, 'positive', 0.45, ['unluckily beach rough us swim still took awesome pictures ']], [**'food'**, 'positive', 0.275, ['food simply fantastic ', 'would recommend full package meals   ']], [**'price'**, 'positive', 0.3, ['definitely worth price ']], [**'room'**, negative, -0.357, [But room had uncleaned dust and quality was not as expected ']], [**'pool'**, 'positive', 0.8, ['excellent pool amazing views ']]]

4.5.5.1 Get polarity value using TextBlob

In this research I have used a python library call “TextBlob” for sentiment analysis task. TextBlob is a python library for text processing. It provides simple API s for NLP tasks such as part of speech tagging, sentiment analysis, noun phrase extraction, classification and more. I have used an API for sentiment analysis in this research to simplify the sentiment extractions in sentences. This research has used textblob.sentiments module for sentiment analysis implementation. textblob.sentiments is based on “pattern” library and NaiveBayesAnalyzer (NLTK classifier trained on movie reviews corpus).

This step can be further improved using a different sentiment lexicon like SentWordNet for better results. TexBlob has been employed as it was easy to use.

4.5.6 Aspect based summary generation

When each review is annotated through the program, the program itself generate the over al summary by aggregating each review summary. It is list with aspects and polarity values. For an example ‘staff’:[44,4] means that aspect ‘staff’ has 44 positive reviews and 4 negative reviews.

```
{
    'staff': [44, 4],
    'place': [48, 6],
    'pool': [38, 0],
    'wifi': [6, 2],
    'water': [18, 4],
    'stay': [38, 10],
    'gym': [3, 1],
    'room': [45, 6],
    'food': [40, 7],
    'reception': [18, 1],
    'spa': [36, 1],
    'service': [28, 4],
    'price': [11, 3],
    'view': [45, 3] }
```

Summary

Data setup is created by extracting reviews from tripadvisor.com website. 5864 of reviews are used for the aspect word extraction. But only 100 of reviews are used to aspect category detection and aspect sentiment identification task which is comparatively less number of data from the whole corpus. But the challenge is to manually annotate data for the test setup. Python is used as the language NLTK is heavily used for text preprocessing tasks. Data preprocessing is a key part of the implementation.

Dynamic extraction of aspect words is considered during the implementation process. Frequency of nouns and noun phrases are identified by considering whole corpus. Words with high frequency are added to the candidate aspect terms list. Then the list is reviewed to select words that have sense of an aspect term. Because every word in the list cannot be considered as a meaningful aspect word.

Then the challenge is to detect right aspect category for review texts. Each sentence is parsed and detected the mentioned aspect word. It the closest aspect word. Then grouped sentences according to the aspect category. Aspect sentiment is calculated for that group of sentences. Then overall sentiment value is assigned to each aspect in a review. At the end, similar aspect terms are grouped and their overall sentiment is calculated. Labeled data is used in different ways in the evaluation process.

Chapter 5

Evaluation

During the evaluation process, three major things will be evaluated. As the research focusing on unsupervised approach for aspect base sentiment analysis, three main tasks are going to be evaluated.

1) Aspect term extraction.

2) Aspect category detection

3) Sentiment classification

The evaluation approach is to compare results obtained for the same dataset two different approaches. One is the unsupervised approach discussed in this research which is implemented and another one from manually annotated data which is supervised approach. Evaluation is done based on the comparison of two result sets. 100 of review texts have been randomly selected from the large dataset (reviews scraped from hotel review web site) and those will be manually annotated. Annotation will include aspect term, opinion word and opinion polarity for each sentence in all 100 review texts. Then same data set used as an input to the implemented program and generated results. This chapter describes on how dataset was prepared and there levels of evaluation process.

5.1 Preparing test dataset

We have review texts already extracted using the web scraper in the previous chapter. 100 of review texts were randomly selected. Target of preparing the data set is to have test which is annotated manually with the aspect list generated from aspect words extraction module.

5.1.1 Data annotation

We manually read each review text and identify if the sentence has mentioned about an aspect in the aspect list. Then tag the sentence with aspect term and the polarity. If the sentence is not about an aspect or user has not mentioned a sentiment, those sentences are not tagged. Because this research is not interested in sentences without polarity and aspects. Section 4.4 has explained about the dataset and annotation in with more detail information. Aspect term and polarity tagged in the sentence level. But in the next step grouped sentences based on aspects and then aggregated polarity is assigned. At the end a review text has aspect term and their overall polarities.

Below is an example of a manually annotated review text. It has annotated with three aspect terms. ‘view’, ‘food’ and ‘room’ are three aspects.

[[**view**, positive, [I would say that this hotel has the most beautiful view in Kandy, the hotel lies up in the hill just above Kandy]],[**food**, positive, [The hotel was organised and their buffet was delicious]],[**room**, positive, [However, I had a deluxe room with a mountain view that were booked through my travel agency, When I arrived they tried to charge me an extra to have the mountain view room]]]

5.2 Evaluation measures

There are multiple evaluation metrics available in the literature to do qualitative evaluations

1) Confusion matrix

Confusion matrix also known as error matrix is table that visualize performance of an algorithm and it is being used in supervised learning and unsupervised learning. In unsupervised learning it is usually called matching matrix. This table is with two dimensions “predicted” and “actual”. If a classification system has trained to distinguish between multiple classes, a confusion matrix can summarize the result of algorithm for further inspection.

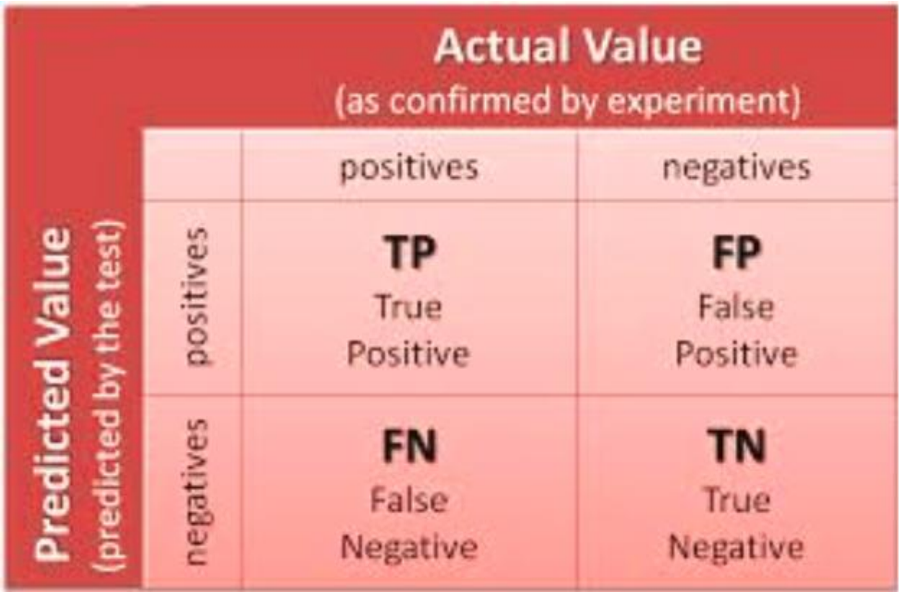

Figure 5.2.1 - Confusion matrix

2) Precision

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

3) Recall

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

4) Accuracy

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

5.3 Evaluation results

5.3.1 Aspect extraction

Aspect words are extracted using a word frequency based approach as described in previous chapter. And later more sensible aspect words are extracted based on threshold value. Final aspect list is created after grouping aspect words based on topics and domain synonyms.

To evaluate the accuracy of aspect words extraction process, manually created an aspect word list, by reading each review and tagging each review independently with aspect words by identifying aspects that are discussed in the review. As an Example, below Figure 5.3.1 is a review and aspect words are highlighted. Then extracted those words and counted word frequencies.

Went on a holiday with my family to Sir Lanka. This hotel was part of my self made itinerary and set up by my travel agent. This hotel is situated in an excellent location. Our kids enjoyed the chalet rooms, the pool was nice and clean. A good spread at breakfast. Staff were very helpful and kind. Overall lovely setup, location, food, hospitality. Loved every bit of it.

Figure 5.3.1 – Example of manually annotated review text to extract aspect words

The reviews are same collection of 100 reviews that used as the test set. Table 5.3 is the manually created aspect list. It shows words with count greater than 7. Rest of the words are not frequent and those can be ignored.

Aspect word	count
food	40
stay	29
pool	21
room	21
ambiance	16
view	15
staff	11
service	11
place	11
location	9

Table 5.3-Manually created aspect list

Now compared the above list with the word list that created automatically through the aspect extraction algorithm.

Below table 5.3.1 is the comparison with top 10 aspect categories detected through aspect extraction process. 70% of aspect words are same in both list. Therefore, extracting most frequent nouns and noun phrases is effective in aspect word extraction. When comparing two columns in the table, there additional aspect words that have detected in the automated process than the ones identified manually. It is an indication that aspect extraction method can reveal some hidden aspects. Aspect word categorization can be further improved to extract most relevant and meaning full categories.

Manually extracted Aspect word	**Automatically extracted aspect categories**
food	staff
stay	food
pool	room
room	pool
ambiance	service
view	place
staff	view
service	stay
place	water
location	reception

Table 5.3.1. - comparison of two aspect word lists

5.3.1.1 Effectiveness of automatic aspect extraction

The graph of Figure 5.4 shows that higher threshold is more effective when selecting set of real and sensible aspect words out of the noun and noun phrases with word frequency based list. Less number of aspect words are identified as infrequent aspect terms are ignored. Most frequent words are aspect words and other words with less frequency are synonyms or in same domain synonym category. This graph shows clearly that minimum threshold parameter impacted on accuracy and number of detected aspect words.

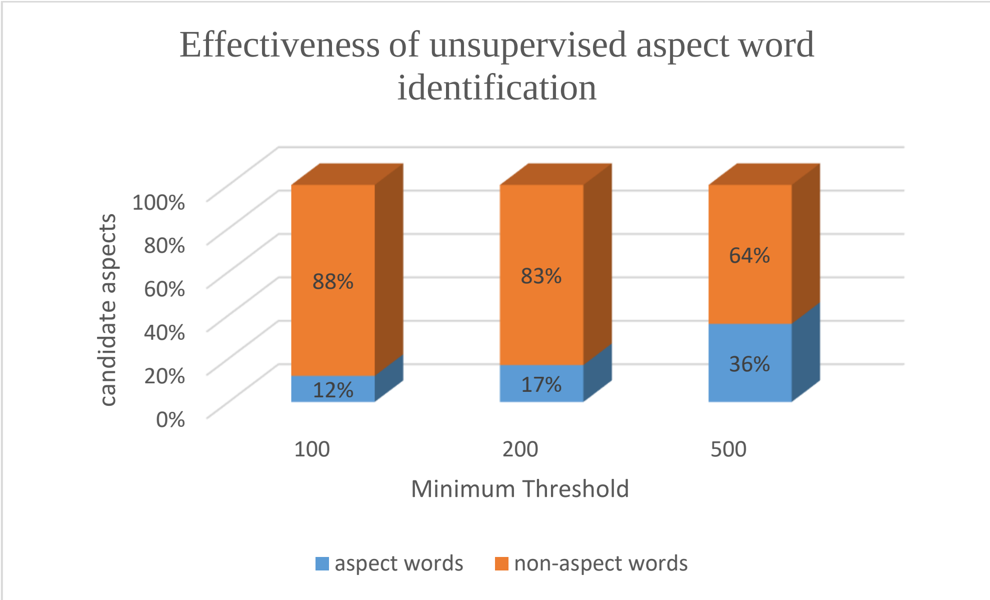

Figure 5.3 Effectiveness of unsupervised method of aspect identification

5.3.2 Aspect category detection

Evaluated the aspect terms identification in reviews by comparing two results set. Dataset annotated by the implemented program and the manually annotated dataset was compared. As the pie chart in the Figure 5.5 depicts, for 22% of reviews detected with aspect words. There are 83% of reviews have less than or equal 3 errors. Only 6% are detected with 4 or 5 errors. Here the number of errors describe number of aspects that program unable to identify or identified incorrectly. In this comparison if manually annotated reviews have right aspect value and considering as a base or 'ground-truth'.

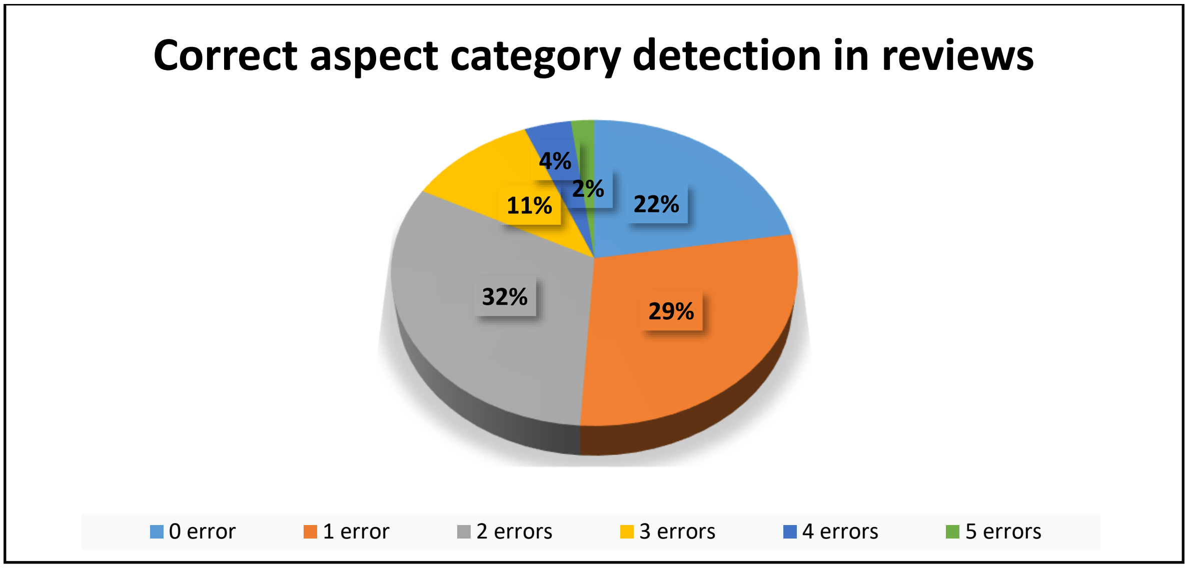

Figure 5.3.2 - Aspect category identification in reviews

Table 5.5 and table 5.6 are results output from implemented aspect category detection program and from manually annotated dataset respectively. When comparing two tables, identification of aspect like 'staff', 'service', 'stay' and 'wi fi' have closer values detection of those categories are more accurate. But it had performed badly for detecting, 'spa' and 'water'. Also, another indication is that 'staff', 'food', 'room', 'pool', 'service', 'place' are the most dominant aspect categories. Total number of 433 sentences were annotated with aspect in the manual annotation. And the program had 470 tokenized sentences. The reason to this difference is difference in splitting sentences by the tokenizer and annotators. And during the annotation, some of the sentences were ignored as they are meaningless. But the program has annotated all sentences in the review.

	Aspect Category													
polarity	staff	food	room	pool	service	place	view	stay	water	reception	spa	wifi	price	gym
Positive	44	40	45	38	28	48	45	38	18	18	36	6	11	3
negative	4	7	6	0	4	6	3	10	4	1	1	2	3	1

Table 5.5 - aspect category and polarity gives as output from the implemented program

	Aspect Category													
polarity	staff	food	room	pool	service	place	view	stay	water	reception	spa	wifi	price	gym
Positive	49	59	34	26	26	67	30	36	0	10	5	4	8	0
negative	4	13	9	1	6	9	1	2	0	5	1	5	3	1

Table 5.6 - aspect and polarity from manually annotated dataset

5.3.2.1 Confusion matrix for aspect category detection

Table 5.3.2.2 is a confusion matrix created based on results from manually annotated data set (test dataset) and results from the same data set annotated through program (training dataset). In both approaches 100 reviews are tokenized into sentences and labeled under aspect category labels. The confusion matrix is created to revel more about the accuracy level and identify and do an error analysis in the process. Columns in the confusion matrix are number of reviews labeled training data. Rows represent outcome of test dataset. “staff”, and “food” categories are classified with more accuracy than other categories. Because they to different classes with high contrast.

Table 5.3.1.2 is presenting more evaluation measures calculated based on confusion metrics results. ‘staff’ aspect has the highest accuracy.

5.3.2.1 Error analysis

As per overall results from confusion matrix, there are many misclassification of aspect categories. ‘place’ has classified as ’view’. In some reviews view is mentioned with location. Then there is a tendency to classify ‘place’ to ‘view’.   Another problem is ‘water’ is classified as ‘stay’. ‘water’ as aspect is about drinking water, clean water or hot water for bathing. But in this case sentence mentioning about ‘lake’ also labeled with ‘water’.” Celebrations near the lake were awesome” is one example which manually annotated as ‘stay’. The reason might be using general purpose word corpus for similarity detection. Usage of domain specific word corpus would reduce such misclassifications. Some category labels can be combined to one class such as ‘staff’, ‘service’. In the same way ‘place’ and ‘view’ can be combined in to one category. Misclassification may be causing due to classifier. This classifier has not trained with any domain specific data. This is only labeling sentences based on how its aspect category is detected. Then there is possibility to not giving accurate results by the confusion matrix which supposed to evaluate real classification problems.

	food	gym	noClass	place	pool	price	reception	room	service	spa	staff	stay	view	water	wifi
Accuracy	0.92	0.98	0.92	0.83	0.87	0.94	0.99	0.87	0.87	0.93	0.99	0.87	0.83	0.95	0.96
F1 score	0.71	0	0	0.37	0.21	0	0.88	0.41	0.06	0.23	0.95	0.30	0	0	0
Recall	0.58	0	0	0.32	0.21	0	1	0.41	0.06	0.83	0.90	0.34	0	0	0
Precision	0.89	0	0	0.44	0.21	0	0.78	0.41	0.06	0.13	1	0.27	0	0	0

Table 5.3.1.2 - evaluation measures based on confusion matrix

		Automatically detected aspect category														
		food	gym	noClass	place	pool	price	reception	room	service	spa	staff	stay	view	water	wifi
Manually annotated aspect category	food	**42**	0	0	0	0	0	0	**30**	0	0	0	0	0	0	0
	gym	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
	noClass	0	**4**	0	0	0	**14**	0	0	0	**11**	0	0	0	0	**8**
	place	0	0	0	**24**	0	0	0	0	0	0	0	**4**	**48**	0	0
	pool	0	0	0	0	**8**	0	0	0	**30**	0	0	0	0	0	0
	price	0	0	0	0	0	0	0	0	0	**11**	0	0	0	0	0
	reception	0	0	0	0	0	0	**15**	0	0	0	0	0	0	0	0
	room	0	0	0	0	**30**	0	0	**21**	0	0	0	0	0	0	0
	service	0	0	0	**30**	0	0	0	0	**2**	**5**	0	0	0	0	0
	spa	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0
	staff	**5**	0	0	0	0	0	0	0	0	0	**48**	0	0	0	0
	stay	0	0	0	0	0	0	**3**	0	0	0	0	**13**	0	**22**	0
	view	0	0	0	0	0	0	0	0	0	0	0	**31**	0	0	0
	water	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	wifi	0	0	0	0	0	0	0	0	0	**9**	0	0	0	0	0
	__all__	**47**	**4**	**0**	**54**	**38**	**14**	**19**	**51**	**32**	**37**	**48**	**48**	**48**	**22**	**8**

Table 5.3.1.3 - Confusion matrix for aspect category detection

5.3.3 Sentiment Classification

Evaluation of sentiment classification is to determine how accurately the positive or negative value is detected in reviews with respect to each aspect. As there were no existing bench mark dataset for hotel reviews where aspect and sentiment annotated, I have used the manually annotated data set with sentiment classified data through the proposed algorithm to experiment the results. According to the Table 5.3.3, 18% of reviews are identified with correct aspect, sentiment pair. Further analysis showed that there 81% of reviews identified with less than 3 wrong aspect categories. Comparatively 76% of them has detected with correct sentiment.

Table 5.3.3.1 is describing the positive and negative distribution of two datasets. Sentiment identification is performing quite well with proposed approach. But in this case, it is difficult to prove sentiment analysis more accurately with aspect level. This gives overall sentiment classification. Table 5.6 and table 5.5 are clearer about aspect level sentiments. But still it is better to evaluate sentiment classification results.

	Sentiment classified through the algorithm
# correct aspect category detected with 0 errors	22%
# correct aspect , sentiment pairs	18%

Table 5.3.3 - correct aspect sentiment pairs

	Positive	**Negative**
Manually annotated sentiments	87%	13%
Automatically detected sentiments	89%	11 %

Table 5.3.3.2 - sentiment classification in two datasets

Aspect sentiment can be evaluated using accuracy (Acc). Acc defines the number of correctly classified (positive or negative) aspect term occurrences divided by the total number of aspect term occurrences.

$$\text{Acc} = \frac{\text{Total number of crrectly classified aspect polarity occurrences}}{Total\ number\ of\ aspect\ polarity\ occurrences}$$

Acc = 0.4808

According to the interpretation of Acc, the accuracy of the sentiment classification is 0.4808 which is not good enough. Although Acc is a well-established evaluation measure for classifications, when it applying to skewed class distributions, it is giving misleading results. Therefore, further computed some other evaluation as precision, recall and F1-score per class. Similar evaluation measures have been used in ABSA task of SEMEVAL 2014.

$$Positve\ precision = \frac{Number\ of\ aspect\ terms\ correctly\ classfied\ as\ positive}{\text{Total number of aspect terms classified as positive}}$$

Positive precision= 0.5311

Positive precision is indicating that classifying in positive class is more than 50% which is much better result. When considering ABSA systems, average polarity scores for aspect terms are more sensible. Therefore, classification errors do not matter for average polarity scores. And in this scenario, negative sentiments are very few and do not need to evaluate further. But it is more important to compute average polarity scores of most frequent aspect terms as it gives better insights.

5.4 Discussion

This chapter described the evaluation process involving all modules that is implemented for aspect word extraction, aspect category detection and aspect sentiment classification. A manually annotated data set from hotel reviews was used as the test dataset. Evaluation was carried out comparing results generated by the defined unsupervised methods. Evaluation results of aspect word extraction indicate that extraction method has in effective of extracting correct aspect terms with 70% of similarity to the supervised aspect extraction approach. Also, could reveal the impact of threshold value to identifying more sensible and distinct aspect categories. But still room for improvement in aspect categorization to make it totally unsupervised. Also, the evaluation revealed that aspect category detection has problem in the classification. Number of misclassified classes are found and the reason could be number of aspect words are as listed as distinct words but still can be merged to one category. Confusion matrix computed for the multi-class classification has pointed errors in the classification which needs to be addressed as future improvements. Aspect sentiment classification performed well in the average sentiment classification. But accuracy is lower and the reason could be skewed distribution of classes. According to table 5.3.3.2 that overall sentiment classification comparison has closer results to the results from test data. Evaluation can be further enhanced with more accurate bench mark data set and more refined classification approach.

Chapter 6

Conclusion

Aspect based sentiment analysis for hotel reviews is discussing an unsupervised approach where it is possible to see the summary of hotel reviews based on different aspects such as ‘price’, ‘service’ and ‘location’. Although several approaches are proposed in the literature, they are supervised or semi supervised as mentioned in chapter 2. This research was carried out to explore the possibility of employing unsupervised approach for hotel review. Even though the hotel domain is selected with purpose of narrowing the scope, proposed solution is domain independent.

This research investigates the role of unsupervised approach of aspect term extraction and multi-aspect sentiment analysis. We show that aspect extraction is performing quite well. But aspect category detection is not well performed. But still there is a room for improvements. If the module of aspect category detection performing well, it is possible to support for an application which summarizing hotel reviews in the aspect level. Aspect category detection having many misclassification and not performing quite well and there is a room for refinements as a multi-label classification problem. Aspect sentiment classification has shown better results with overall average of polarity but with lower accuracy computed from confusion matrix.

This research has gone through all the steps involved in unsupervised aspect based sentiment analysis process from extracting data from review site to generating aspect based sentiment summary. The web scraper developed to extract review data can be reused for data extraction. The experimental setup can accept a set of review texts and generate aspect based summary. But extracted aspect terms also should be given as an input. This can be further improved as an application that will able to summarize dominant aspects considering all reviews rather that individual review level summery that available now.

6.1 Future works

The objective of this research is to explore the effectiveness of using an unsupervised approach to aspect based sentiment analysis task. Hotel reviews are selected as the domain, as it is very common that people need to review before visiting to a place that they haven't traveled before. Possibilities and ideas to extending the experiments will be discussed here.

Further improvements can be done on aspect word extraction.

Aspect word frequency is obtained in this research bases on noun and noun phrases and that resulted many unnecessary words. This can be further improved by using some other lexicons like adjectives, neighboring words to sentiment words. Aspect word extraction can be done totally unsupervised in the process of candidate aspect category creation using a topic modeling such as LDA as a post processing step to the frequent word list, it will create word list, it will create proper categorization in totally unsupervised approach.

When considering aspect category detection, we used WordNet and it is general purpose corpus with synsets. Instead of that, if we use a domain specific words to aspect category extraction, it will give better results. Also applying multi-class aspect classification model trained with domain specific data will reduce misclassifications as we discussed in chapter 05.

In the aspect sentiment analysis task still, room for improvements. When one sentence is expressing about more than one aspects and different sentiment for each aspect in the proposed algorithm unable to determine two sentiments separately as it give overall sentiment for the whole sentence. We ignored that problem in this research as there are only few of such sentences included in reviews. But addressing that problem may increase the accuracy of sentiments which may be more important to a different domain that hotel reviews.

Most of the online user generated contents contain many incomplete sentences, incomplete words, abbreviations, grammatical errors. This study has assumed that no such problems and all review texts are grammatically correct. But this poorly formatted texts can cause extraction of wrong classification could be a possible future improvement.

This research has outlined a flow of unsupervised based approach for aspect sentiment analysis. Also, the proposed algorithm performs quite well for 100 of review texts. This approach can be adapted to produce and application to summarize set of reviews.

All the beginning of the research, we web scraper was implemented to extract review texts from tripasvisor.com site. This can be reused for future researches and there are around 5000 review text for hotel review are available to be used for future researches in this domain.

References

[1] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques,” In *EMNLP*, 2002, pp. 79–86

[2] K. Dave, S. Lawrence, and D. M. Pennock,” Mining the peanut gallery: opinion extraction and semantic classification of product reviews,” In *WWW ’03*,2003, pp. 519–528.

[3] A. Devitt and K. Ahmad, “ Sentiment polarity identification in financial news: A cohesion-based approach,” In *Proceedings of ACL’07*, 2007, pp. 984–991.

[4] H. Cui, V. Mittal, and M. Datar, “Comparative experiments on sentiment classification for online product reviews,” In *Twenty-First National Conference on Artificial Intelligence*, 2006, volume 21, page 1265.

[5] S. Kim and E. Hovy. “Determining the sentiment of opinions,” In *Proceedings of COLING*, 2004, volume 4, pages 1367–1373.

[6] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the web. In *KDD ’02*, pages 341–349, 2002

[7] B. Liu, M. Hu, and J. Cheng, ” Opinion observer: Analyzing and comparing opinions on the web,” In *WWW ’05*, 2005, pages 342–351.

[8] L. Zhuang, F. Jing, and X. Zhu, “ Movie review mining and summarization,” In *Proceedings of CIKM*, 2006, page 50..

[9] H. Kim and C. Zhai,” Generating Comparative Summaries of Contradictory Opinions in Text, “ In *Proceedings of CIKM’09*, 2009, pages 385–394.

[10] N. Jindal and B. Liu, “ Identifying comparative sentences in text documents*,”* In *Proceedings of SIGIR ’06* New York, NY, USA, 2006, pages 244–251.

[11] B. Snyder and R. Barzilay, “ Multiple aspect ranking using the good grief algorithm,” In *Proceedings of NAACL HLT*, 2007, pages 300–307.

[12] K. Lerman, S. Blair-Goldensohn, and R. T. McDonald, “ Sentiment summarization: Evaluating and learning user preferences,” In *EACL*, 2009, pages 514–522.

[13] M. Hu and B. Liu. Mining and summarizing customer reviews. In W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel, editors, KDD, pages 168–177. ACM, 2004

[14] A.-M. Popescu and O. Etzioni. "Extracting product features and opinions from reviews," In *Proceedings of HLT '05*, Association for Computational Linguistics, Morristown, NJ, USA , 2005, pages 339–346.

[15] Y. Lu, C. Zhai, and N. Sundaresan, "Rated aspect summarization of short comments," In *Proceedings of WWW'09*, 2009, pages 131–140.

[16] A. Goldberg and X. Zhu, " Seeing stars when there aren , a´rt many stars: Graph-based semi-supervised learning for sentiment categorization," In *HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*, 2006

[17] B. Pang and L. Lee, " Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," In *Proceedings of the ACL*, 2005, pages 115–124.

[18] K. Lerman, S. Blair-Goldensohn, and R. T. McDonald, "Sentiment summarization: Evaluating and learning user preferences," In *EACL*, 2009, pages 514–522.

[19] Brody, Samuel, and Noemie Elhadad, "An unsupervised aspect-sentiment model for online reviews," Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.

[20] Fahrni, Angela, and M. Klenner, "Old wine or warm beer: Target-specific sentiment analysis of adjectives," Proc. of the Symposium on Affective Language in Human and Machine, AISB. 2008.

[21] G. Ganu, N. Elhadad, and A. Marian, " Beyond the Stars: Improving Rating Predictions using Review Text Content," WebDB. Vol. 9. 2009.

[22] Marrese-Taylor, Edison, D. Juan, Velásquez, and Felipe Bravo-Marquez, "A novel deterministic approach for aspect-based opinion mining in tourism products reviews," Expert Systems with Applications 41.17 (2014): 7764-7775, 2014.

[23] B. Lu, M. Ott, C. Cardie,and B. k. Tsou, 2011, December, " Multi-aspect sentiment analysis with topic models," In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference , IEEE*, 2011, (pp. 81-88).

[24] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.

[25] S. Brody and N. Elhadad, "An unsupervised aspect-sentiment model for online reviews," in *Proceedings of ACL:HLT*, 2010, pp. 804–812.

[26] L. Du, W. Buntine, and H. Jin, "A segmented topic model based on the two-parameter poisson-dirichlet process," Machine learning, vol. 81, no. 1, pp. 5–19, 2010.

[27] Brychcín, Tomáš, Michal Konkol, and Josef Steinberger, "Uwb: Machine learning approach to aspect-based sentiment analysis," *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014.

[28] Taboada, Maite, et al, "Lexicon-based methods for sentiment analysis," Computational linguistics 37.2 (2011): 267-307, 2011

[29] Denny, J. Matthew., and Arthur Spirling, "Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it," Unpublished manuscript, Dep. Polit. Sci, Stanford Univ and Inst. Quant. Soc. Sci., Harvard Univ. https://ssrn. com/abstract 2849145 , 2017.

[30] S. Das. and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," *APFA'01*, 2001.

[31] Marrese-Taylor, Edison Velásquez, Juan D.Bravo-Marquez, Felipe., "A novel deterministic approach for aspect-based opinion mining in tourism products reviews. "Expert Systems with Applications 7764-7775 , 2014.

[32] M. S. Akhtar, A. Ekbal and P. Bhattacharyya, " Aspect Based Sentiment Analysis: Category Detection and Sentiment Classification for Hindi".

[33] S.Negi and P. Buitelaar, "INSIGHT Galway: Syntactic and lexical features for aspect based sentiment analysis," *In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 346-350.

[34] I.Titov, R. McDonald," Modeling online reviews with multi-grain topic models," *In Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 111-120.

[35] Q. Liu, B. Liu, Y. Zhang, Kim, D.S. and Z. Gao, "Improving Opinion Aspect Extraction Using Semantic Similarity and Aspect Associations," *In AAAI*, 2016, pp. 2986-2992.

[36] Z.Chen, A. Mukherjee, and B. Liu, "Aspect extraction with automated prior knowledge learning," *In ACL* , 2014, 14, 347–358.