[1]Dr. Ankita Bansal

[2]Dr. Aruna Jain

[3]Dr. Abha Jain

Sentiment Analysis on Hotel Reviews to Quantify Amenities

Abstract: - In today's tech-savvy time, online reviews of a particular product/service play a significant role in determining the decision of a potential buyer. These reviews are crucial not only for the buyers/customers, but also for the sellers as these reviews can have a huge impact on their revenue. Thus, it becomes very essential to analyze, harness and process these online reviews for gaining useful insights. In this paper, the broad aim is to summarize the reviews of hotels using sentiment analysis but the idea can be extended to any kind of product or service. The contribution of the study is to develop a model/prototype to quantify the features/amenities (words that describe the hotel) of various hotels as a percentage of their positive and negative rating given by the user in their review. From quantified scores obtained by the prototype, further a complete summary about the product can be generated. Although, work has been done in the direction which accepts reviews from users, summarize the reviews and provide qualitative analysis of the reviews, but there had been very few works to quantify the features of a product based on their opinion words. This quantitative approach impacts the human decision quality to a very large extent. This work has manifold benefits which can be explained as follows; at the user end it provides the user a complete amenity based quantified summary of the product, i.e. the user gets a concrete idea about each and every attribute of the product under consideration. For instance, let us consider a hotel as a product. The user gets a summary incorporating amenities of the hotel like rooms, staff, location, terrain etc. On the other hand, the seller of a product or service gets the idea about what amenities he can put forth as the product's unique selling proposition and what amenities of the product/service needs improvement.

Keywords: Significant, Quantitative, Incorporating, Improvement, Proposition.

1. Introduction

Web 2.0 revolutionized the internet and took it by a data explosion [1]. Various social media platforms like Twitter, Facebook; product based sites like Amazon, Flipkart; Hotel comparing sites like Trivago, TripAdvisor allow customers to share their experiences in terms of reviews and at the same time allow them to rate their experiences in the form of quantitative ratings. Such ratings and experiences allow the customers to share their opinion about the quality, usability and various other attributes of the product [2]. The data generated from reviews is of prime importance to both the buyer and the seller of the product/service. It has been observed that in today's tech-savvy time, selection of a particular product/service primarily depends on the online reviews given by the people. In other words, people's review about a particular product/service plays a significant role in determining the decision of a potential buyer. These reviews are crucial not only for the buyers/customers, but also for the sellers as these reviews can have a huge impact on their revenue. Online reviews determine the reputation of the brand of the product which can either make or break the e-commerce business of any organization [3]. Thus, online reviews have gained a huge popularity amongst customers for making an opinion about any product/service. In addition to using the traditional process of questioning friends, relatives or acquaintances to make an opinion about any product, the customers prefer to read online reviews available on different websites. In other words, nowadays, the more popular method is to read the reviews of the product/process posted by the customers and other stakeholders on various blogs and social networking websites which are easily accessible on internet [4]. People share their opinions, suggestions and other thoughts, which may be either in favourable or unfavourable on various review sites.

However, these reviews are largely unstructured, contain sarcasm, language slangs etc. Hence, it becomes necessary to understand these reviews properly to derive meaningful insights from them, which is the basic idea behind sentiment analysis [5, 6]. Sentiment analysis is a domain which classifies reviews, comments or opinions

[1] Department of Information Technology, Netaji Subhas University of Technology, Delhi, India

ankita.bansal06@gmail.com

[2]Department of Computer Applications, Bharati College, University of Delhi

aruna.jain@bharati.du.ac.in

[3]Department of Computer Science, Deen Dayal Upadhyaya College, University of Delhi

abhajain.src@gmail.com

into two basic and intrinsic emotional indicators namely positive and negative [7]. Sentiment analysis means understanding of the emotional essence of any text and the evaluation of the nature of opinion of the text [8]. The nature can be either on the extremes of good and bad or just neutral [9].

In this paper, the broad aim is to summarize the reviews of hotels using sentiment analysis but the idea can be extended to any kind of product or service. Thus, we aim to develop a prototype model which can be used for any other product or service. The task of Sentiment Analysis of hotel reviews mainly include the following steps in sequence- preprocessing [10], amenities (words that describe the hotel) extraction which is followed by selection of the relevant amenities [11], opinion mining [12] and finally amenities quantification.

The proposed model aims to provide the customer a comprehensive rating of the hotel. One of the simple approaches to provide a comprehensive review is to compute an average quantitative rating based on individual user ratings. But this approach does not reflect the true nature of the hotel and the net hotel rating is not comprehensive in terms of the information conveyed by a single number. Thus, such average ratings are generally not a true indicator of the nature of the hotel. So to deal with the above problem and give a more comprehensive description, the authors in this paper aim at giving feature based summary of a hotel on the basis of the reviews available. The work is concerned with mining the sentiments gathered from the online reviews of the people given for a particular hotel which can greatly influence the decision of a traveler who needs to book a hotel when planning a vacation. These reviews are also highly crucial for the hotel management having a huge impact on their revenue and thus, determining their reputation which can either make or break the e-commerce business of hoteliers. Thus, access to a technological analytical tool that collects the online information and integrates it to support the decision making is highly mandatory. Strategic actions need to be defined using analytical tool which can determine the success or failure of a hotel business. Much work has been done in this direction which accepts reviews from users, summarize the reviews and provide qualitative analysis of the reviews [5-9]. But there had been very few works to quantify the features of a product based on their opinion words. Thus, more research and experimentation is needed in this field to draw useful conclusions.

Keeping the above issue in mind, the authors intend to convert the subjective review into objective with quantified values rather just the positive or negative labeling.

The broad contributions of this paper are as follows:

1. Develop a prototype which can be used to provide rating to any product or services based on the online reviews
2. Provide the overall rating (recommendation percentage) of each hotel so that the customers can draw a comparison amongst them
3. Identify the features (amenities) of each hotel and quantify each amenity based on the percentage of positive and negative rating
4. Provide feature based summary of each hotel
    - Identify all the features (amenities) of the product.
    - Detect the adjectives (opinion words) corresponding to those features using SpaCy. A sentence parse tree is generated which binds a particular feature to its various adjectives.
    - Assign orientation **(**positive, negative, neutral**)** to each adjective using SentiWordNet.
    - Quantify each amenity based on percentage of positive rating and negative rating.

This quantitative approach impacts the human decision quality to a very large extent. From quantified scores obtained by the prototype further a complete summary about the product can be generated. In this paper, the features/amenities (words that describe the hotel) are being quantified as a percentage of their positive and negative rating given by the user in their review. This work has manifold benefits which can be explained as follows; at the user end it provides the user a complete amenity based quantified summary of the product, i.e. the user gets a concrete idea about each and every attribute of the product under consideration. For instance, let us consider a hotel as a product. The user gets a summary incorporating amenities of the hotel like rooms, staff, location, terrain etc. On the other hand, the seller of a product or service gets the idea about what amenities he

can put forth as the product's unique selling proposition and what amenities of the product/service needs improvement.

For the purpose of empirical validation, the authors have used deceptive opinion spam dataset available on www.kaggle.comhttps://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus). The Deceptive opinion spam dataset is a corpus consisting of a total of 1600 truthful and deceptive hotel reviews of 20 Chicago hotels.

The paper is organized as: Next section briefly discusses the studies in the related field. Section 3 explains the dataset used for empirical validation. After section 3, section 4 focusses on the research methodology, where the framework of the proposed model is explained in detail. The proposed algorithm along with an illustration is written in section 5. Section 6 presents the results of the study. Finally, the work is concluded in section 7.

2. Related Work

Sentiment analysis has gained wide importance to handle various Natural Language Processing tasks at different levels of granularity. Much work has been done at document level [13, 8, 14, 15] and sentence level [16, 17]. More recent work involving analysis of sentiments is also performed at phrase level [18, 19, 20]. However, performing sentiment analysis and opinion mining on different attributes/features extracted from a sentence is a naïve field and requires research and experimentation for gaining useful insights. Thus, in this work, the features are being quantified as a percentage of their positive and negative rating given by the users in their reviews. This provides a complete aspect based quantified summary of the product, i.e. every attribute/aspect of the product under consideration gets quantified.

Significant research has been done towards sentiment analysis on the data generated from social media like Twitter [21, 22, 23, 24]. Go et al. (2009) [21] analyzed distant learning data and used the tweets ending with positive and negative emoticons. The authors constructed Naïve Bayes, MaxEnt and SVM models. In terms of feature space, they tried a Unigram, Bigram model in conjunction with parts-of-speech (POS) features. They found SVM and unigram models outperformed other models. Pak and Paroubek (2010) [23] collected the similar distant learning data as was done by Go et al. (2009) [21]. They performed subjective versus objective classification task. For subjective classification, they analysed the tweets ending with positive and negative emoticons as was done by [21]. For objective data, they crawled twitter sections of popular newspapers. Their results were contradictory to the results of Go et al. (2009) [21] as they found bigrams to be useful.

Another popular field of research is analysing the product reviews using sentiment analysis. Huettner and Subasic (2001) [25] manually constructed a discriminant-word lexicon and used fuzzy logic to classify sentiments. Tong (2001) [18] generated sentiment timelines. The study tracked online discussions about movies and displayed a plot of the number of positive and negative sentiment messages over time. Messages were classified by looking for specific phrases that indicate the author's sentiment towards the movie (e.g., "great acting", "wonderful visuals", "uneven editing"). Each phrase was required to be manually added to a special lexicon and manually tagged as positive or negative sentiment. The lexicon is domain dependent (e.g., movies) and must be rebuilt for each new domain. In contrast, in this work, we have only manually created a small list of seed adjectives tagged with positive or negative labels. This list of seed adjective is also domain independent. We have also proposed an effective technique to grow this list using WordNet. Turney et al. (2003) [15] applied a specific unsupervised learning technique based on the mutual information between document phrases and the words "excellent" and "poor", where the mutual information is computed using statistics gathered by a search engine. Wawre1 and Deshmukh (2016) [26] examined several supervised machine learning methods for sentiment classification of movie reviews and concluded that machine learning techniques outperform the method based on human-tagged features.

However, these works determine the sentiments of each document, i.e. they do not consider the domain specific information, nor did they take into account the users' views on different aspects of a given product. Moreover, they also do not find the features/aspects on which opinions have been expressed. Unlike document-level or sentence-level sentiment analysis, aspect-based sentiment analysis determines the polarity in each aspect being examined [27]. There are works in literature which have incorporated the user sentiments towards different aspects of a product [28, 29, 30, 31]. However, these works did not propose any approaches on how to combine feature specific sentiments to obtain aggregate review polarity of a given user towards the product.

In this paper, we aim to develop a prototype to obtain the feature based scores. From feature based scores obtained by the prototype, further a complete summary about the product can be generated. The work revolves around summarizing the reviews of different hotels and quantifying different features as a percentage of their positive and negative rating given by the user in their reviews. This quantitative approach impacts the human decision quality in interpreting the reviews to a large extent.

3. Data Collection

This section provides a detailed view of the structural and pictorial representation of the dataset used in the study. For the purpose of empirical validation, the deceptive opinion spam dataset is used which is an open source dataset available on www.kaggle.com https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus). The deceptive opinion spam dataset is shown in figure 1. It consists of five columns which are as follows:

- Deceptive: It gives the nature of the review whether it is true or misleading
- Hotel: It specifies the name of hotel for which the review is given
- Polarity: It specifies the nature of the review whether is positive or negative
- Source: It specifies the name of the site from where the review is taken
- Text: It specifies the actual review in text format

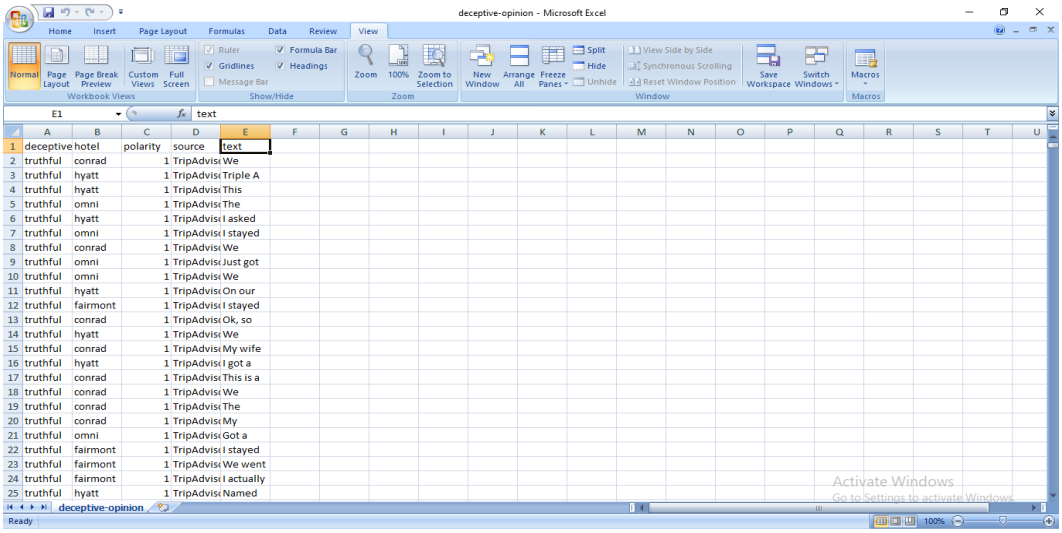


Figure 1: The Deceptive Opinion Spam Dataset

The corpus contains a total of 1600 reviews corresponding to 20 Chicago hotels taken from different sources like TripAdvisor, Expedia, Orbitz etc. The distribution of 1600 reviews as given in the corpus is as follows:

- 400 truthful, positive reviews from TripAdvisor
- 400 deceptive positive reviews from Mechanical Turk
- 400 truthful, negative reviews from Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor, and Yelp
- 400 deceptive negative reviews from Mechanical Turk

Figure 2 depicts the percentage of reviews which are pertaining to each of the 20 hotels. For instance, the percentage of reviews which are pertaining to 'Hyatt' hotel is 5% (figure 2). This implies that 5% of total number of reviews (1600) is that of 'Hyatt' hotel i.e. out of 1600 reviews, there are 80 reviews pertaining to 'Hyatt' hotel. Similarly, it can be observed from the figure that the percentage of reviews pertaining to each of the 20 hotels is 5% which evaluates to 80 reviews corresponding to each hotel.

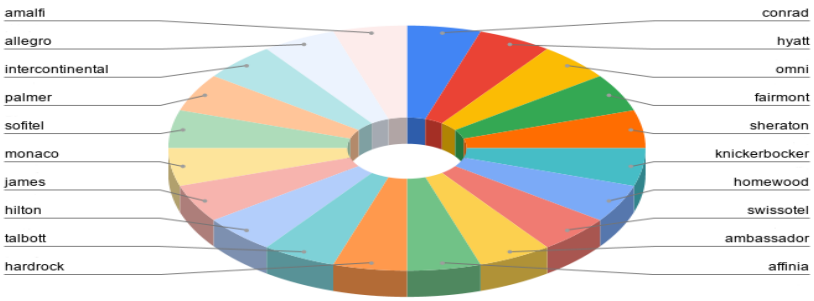


Figure 2: Percentage of reviews corresponding to each of the 20 hotels

The textual reviews are extracted from the dataset and are cleaned and re-organized. The data is cleaned as the collected data is not well structured and needs some processing before it becomes ready to use. In this study, cleaning of the data is done by: (1) Removing of HTML tags and URLs, (2) Correcting spelling errors. Reviews may contain bold /underlined/ italic words to emphasize the meaning of some word or sentence. Different tags in HTML are used for this purpose, <b> for bold, <u> for underline and <i> for italics. However, while analyzing the reviews, such emphasizing is of no use as they do not provide any useful information towards the sentiment, thus, they are removed. Similarly, punctuation marks, special characters, white spaces, stop words etc. are also removed during cleaning. In addition to this, there may be some spelling errors in the review which can result in deviation from correct analysis. TextBlob is a Python library for processing textual data. We used the correct() method in TextBlob library to attempt spelling correction.

4. Research Methodology

In this section, research methodology of the work is presented. Figure 3 depicts the steps involved in carrying out the proposed work viz. pre-processing, amenities extraction and finally opinion mining. Each of the steps has been explained in the subsequent sub-section.

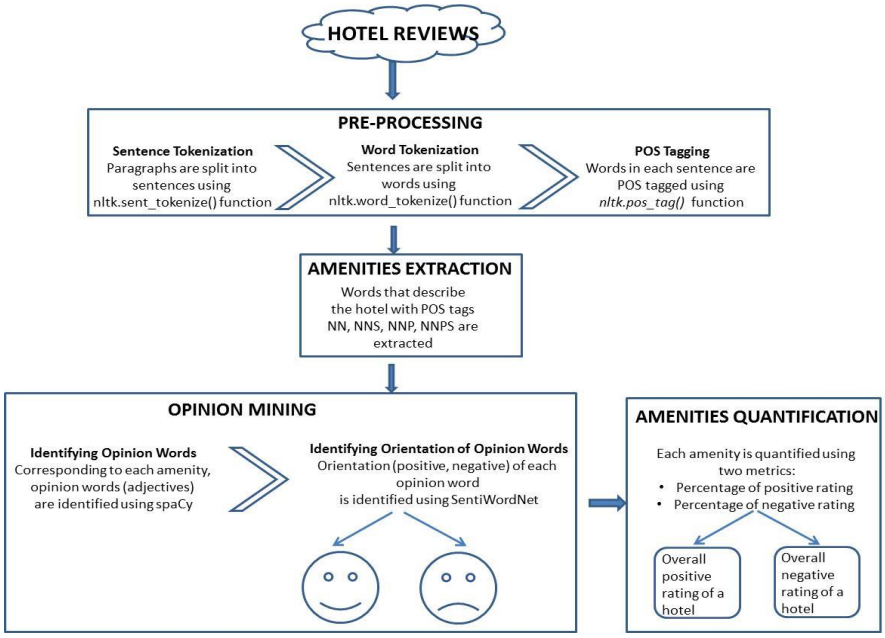


Figure 3: Framework of the Proposed Methodology

4.1 Pre-processing

Firstly, pre-processing of all 1600 reviews which are in the textual format is carried out using two main steps viz. sentence and word tokenization and Part-of–Speech (POS) tagging.

4.1.1 Sentence and word tokenization

A token is a block of text which could be a word, a paragraph, a sentence, a syllable or a phoneme depending upon the need of the application in which tokenization is to be performed. Tokenization represents the process of converting a stream of characters into a sequence of tokens. In this work, the authors had firstly performed sentence tokenization, followed by word tokenization.

In sentence tokenization, a token is considered as a sentence and the process is to split the given text (hotel reviews) which is in the form of a paragraph into different sentences. The authors have used *nltk.sent_tokenize()* function available in NLTK library of python language to perform sentence tokenization. This function uses an instance of PunktSentenceTokenizer from the nltk.tokenize.punkt module, which has already been trained and knows what punctuations and characters mark the beginning of a new sentence and end of a sentence. Figure 4 depicts a snippet in python language which has been used to perform sentence tokenization.

Once a set of sentences have been retrieved from a given text, then each sentence is further tokenized into a set of words (word tokenization) using *nltk.word_tokenize()* function.

Figure 4: Python Snippet to Perform Sentence Tokenization

4.1.2 POS Tagging

We need to differentiate subjective reviews from objective reviews. Subjective reviews are those reviews which contain the reviewer's opinion about the hotel. These reviews must be containing some adjectives describing various features of the hotel like rooms, food, service etc. On the other hand, there are some reviews which do not contain the reviewer's opinion about the hotel and therefore will not contain any of such features. Such reviews are called objective reviews. For our purpose, only the subjective reviews are useful and the remaining reviews can be ignored. In order to filter the reviews (subjective review or objective review), we first need to identify in each sentence the Part-of-Speech (POS) of each word. For this, the authors have performed POS tagging. It is the process of tagging different parts of speech in a sentence like tagging of nouns, adjectives, verbs etc. POS tagging is a supervised learning solution that uses features like the previous word, next word, is first letter capitalized etc. The authors have used *nltk.pos_tag()* function available in NLTK library of python language which returns the list words along with their POS tags. Figure 5 depicts a snippet in python language which has been used to perform POS tagging. Each sentence is first word tokenized using *nltk.word_tokenize()* function and then passed to the *nltk.pos_tag()* function.

Figure 5: Python Snippet to Perform POS Tagging of Sentence

Some of the popularly used POS tags are: CC Coordinating Conjunction (CC), Cardinal Digit (CD), Determiner (DT), EX Existential there (EX) e.g. there is, there exists, Foreign Word (FW), Preposition/subordinating conjunction (IN), adjective (JJ) e.g. big, Comparative Adjective (JJR) e.g. bigger, Superlative Adjective (JJS) e.g. biggest, list marker (LS), Modal (MD) e.g. could, will, Singular Noun (NN) e.g. desk, Plural Noun (NNS) e.g. desks, Singular Proper Noun (NNP) e.g. Harrison, Plural Proper Noun (NNPS) e.g. Americans, Pre-Determiner (PDT), Possessive Ending (POS) e.g. parent's, Personal Pronoun (PRP) e.g. I, he, she, Possessive Pronoun (PRP) e.g. my, his, hers, Adverb (RB) e.g. very, silently, Comparative Adverb (RBR) e.g. better, Superlative Adverb (RBS) e.g. best, To (TO) e.g. go to the store., UH Interjection (UH), Verb in base form

(VB) e.g. take, Verb in past tense (VBD) e.g. took, Verb in gerund/present participle (VBG) e.g. taking, Verb in past participle (VBN) e.g. taken, wh-determiner (WDT) e.g. which, wh-adverb (WRB) e.g. where, when.

4.2 Amenities Extraction

This step deals with the extraction of words from the textual reviews that tell us about the hotel like rooms, service, staff, location, food etc. In this work, these words are referred to as amenities. As we know, these words will fall into the category of nouns i.e. common nouns and proper nouns both singular as well as plural. From POS tagging performed on each word of the sentence, we can easily identify the words which are nouns. These words will have POS tags as: NN, NNS, NNP and NNPS**.** These amenities are extracted from the reviews.

Let us understand the process of amenities extraction with the help of a suitable example. Let us consider a sentence as: 'The rooms were big'. POS tagging of the given sentence will be: The (DET) rooms (NNS) were (VBD) big (JJ). (.). From POS tagging, it can be seen only the word 'rooms' has a NNS (Plural Noun) POS tag. Therefore, only rooms will be selected as an amenity.

We have also considered noun phrases in the reviews for summarization. These are identified by the presence of two or more feature words i.e. words with tags NN, NNS, NNP, NNPS consecutively. For instance, consider a sentence as: 'The Taj hotel rooms were big'. POS tagging of the given sentence will be: The (DET) Taj (NNP) hotel (NN) rooms (NNS) were (VBD) big (JJ) . (.). From the above POS tagging, Taj hotel rooms will be selected as an amenity because the POS tag of Taj is NNP, hotel is NN and rooms NNS and they are consecutive.

To avoid using the same amenities but in different forms such as room and rooms, the authors have done lemmatization. We lemmatized each word before finalizing the list of selected features. Lemmatization is different from stemming as it takes into consideration the morphological analysis of the word which is not considered in stemming. Also lemma of a given word will always be an actual language word whereas stem may or may not be an actual language word. The authors have used lmtz.lemmatize function available in NLTK library of python language which returns the base form of the given word. Figure 6 depicts a snippet in python language which has been used to perform lemmatization.

```
In [1]: import nltk
        from nltk.stem.wordnet import WordNetLemmatizer

In [2]: word = "cars"
        lmtz = WordNetLemmatizer()
        base_word = lmtz.lemmatize(word)
        base_word

Out[2]: u'car'
```

Figure 6: Python Snippet to Perform Lemmatization

4.3 Opinion Mining

Opinion mining consists of two steps viz. identifying opinion words and identifying orientation of opinion words.

4.3.1 Identifying opinion words

Once the amenities are extracted using POS tagging, then the opinion words corresponding to each amenity are identified. Opinion words are words that people use to express a positive or negative opinion about a particular amenity. In other words, we can say opinion words are the adjectives of that amenity. To identify the opinion words corresponding to a particular amenity, we go through all the reviews in the database to find out the reviews containing that amenity. A review that contains the said amenity is then used to find the corresponding adjectives (opinion words) related to that amenity. For instance, consider a review as: 'The food was delicious'. In this review, food is the amenity extracted using POS tagging and delicious is the opinion word which is expressing a positive opinion about food.

In this work, authors have used spaCy to find out the opinion words from the reviews. SpaCy has been used because of its superior precision over NLTK. It generates a parse tree which connects all the relations in the sentence. It is an open-source software library for advanced Natural Language Processing (NLP), written in the

programming languages (Python and Cython). It offers the fastest syntactic parser in the world. It features convolutional neural network models for part-of-speech tagging, dependency parsing and named entity recognition, as well as API improvements around training and updating models and constructing custom processing pipelines. Figure 7 depicts a snippet in python language which has been used to generate spaCy parse tree using spaCy library.

```
In [1]: import spacy
        from nltk import Tree
        en_nlp = spacy.load('en')

In [2]: doc = en_nlp(unicode("The food was delicious."))
        def to_nltk_tree(node):
            if node.n_lefts + node.n_rights > 0:
                return Tree(node.orth_, [to_nltk_tree(child) for child in node.children])
            else:
                return node.orth_
        [to_nltk_tree(sent.root).pretty_print() for sent in doc.sents]
```

Figure 7: Python Snippet to Generate spaCy Parse Tree for Identifying Opinion Words

4.3.2 Identifying orientation of opinion words

Once we have identified the opinion words corresponding to each amenity, we find the orientation of each opinion word/adjective using SentiWordNet [32]. SentiWordNet is a lexical resource for opinion mining in which each synset of WordNet is assigned to three numerical scores: positive Pos(s), negative Neg(s), objective Obj(s) indicating how neutral (the value of Ob(s)) or how affective (the values of Pos(s) and Neg(s)) the terms contained in the synset are (SentiWordNet = Wordnet + Sentiment Information). Each of the three scores ranges from 0.0 to 1.0 and their sum is 1.0 for each synset. A synset that has value 1.0 for the objectivity scores indicates that the term is objective, while 0.0 means that the term conveys some strong sentimental (positive or negative) meaning [33]. A subjective term is considered a term with important positive or negative polarity (PN-polarity), while an objective term can be defined as a term that does not have either positive or negative characteristics [34]. The positive sentiment represents the amount of good rating given by the user in their review and the negative sentiment represents the amount of poor rating given by the user in their review. Figure 8 represent the amount of positive and negative polarity of a given adjective. The green color represents the amount of positive value of the adjective and red color represents the amount of negative value of the adjective. Figure 9 depicts a snippet in python language which has been used to find the orientation of each opinion word using SentiWordNet.

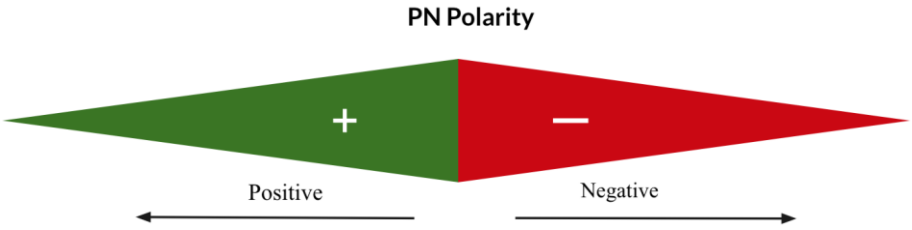


Figure 8: Graphical Representation of SentiWordNet

```
In [1]: import nltk
        from nltk.corpus import sentiwordnet
        from nltk.corpus import wordnet

In [2]: def orientation(inputWord):
            wordSynset=wordnet.synsets(inputWord)
            if(len(wordSynset) != 0):
                word=wordSynset[0].name()
                orientation=sentiwordnet.senti_synset(word)
                if(orientation.pos_score()>orientation.neg_score()):
                    print("Positive")
                elif(orientation.pos_score()<orientation.neg_score()):
                    print("Negative")

        orientation("good")
        Positive
```

Figure 9: Python Snippet to Find Orientation of Opinion Word using SentiWordNet

Most of the negative reviews and negative subjective sentences consist of a negative opinion word before a positive opinion word which completely inverts the meaning of the whole sentence. In order to handle such cases, the authors have maintained a list of negative words commonly used to negate the meaning of a word or sentence. When we find any of the words from the list in a subjective sentence with an opinion word, we invert the orientation of the opinion word found from the SentiWordNet. The list of negative words is don't, never, nothing, nowhere, none, not, hasn't, hadn't, can't, couldn't, shouldn't, won't, wouldn't, don't, doesn't, didn't, aren't, ain't.

4.4 Amenities Quantification

Once the orientation of each opinion word corresponding to a particular amenity is determined using SentiWordNet, we quantify each amenity using two metrics viz. percentage of positive rating and percentage of negative rating. Thereafter, the overall positive rating and overall negative rating of that hotel is determined based on the quantified value of all amenities associated with the hotel. The formulae have been given below:

$$\text{Percentage of positive rating of an amenity} = \frac{\text{Total number of positively oriented opinion words associated with an amenity}}{\text{Total number of opinion words associated with that amenity}} * 100$$

$$\text{Percentage of negative rating of an amenity} = \frac{\text{Total number of negatively oriented opinion words associated with an amenity}}{\text{Total number of opinion words associated with that amenity}} * 100$$

$$\text{Overall positive rating of a hotel} = \frac{\text{Sum of all the positive rating percentages of each amenity}}{\text{Sum of both positive and negative rating percentages of that amenity}}$$

$$\text{Overall negative rating of a hotel} = \frac{\text{Sum of all the negative rating percentages of each amenity}}{\text{Sum of both positive and negative rating percentages of that amenity}}$$

5. Proposed Algorithm

The algorithm proposed by the authors is explained in this section with the help of an example. Each step of algorithm is written, followed by an illustration for better understanding of the algorithm. In the presented algorithm, there are few shortcomings which authors came across while analysing the results. These shortcomings are also listed in this section. The authors plan to work on these and resolve them in their future work.

Step 1: Apply the SentenceTokeniser to split each review (which is in the form of paragraph) in separate sentences.

```
tokenizer = nltk.tokenize.punkt.PunktSentenceTokenizer()
```

Step 2: Then, to each sentence apply the Post Tagging which finds the part of speech Tagger to each Word in a sentence using NLTK POS TAG.

```
tagged_text_list=[]
tagged_text_list.append(nltk.pos_tag(word_tokenize(text)))
```

Step 3: Attach the PosTag along with the word in the sentence.

Step 4: Find the most occurring noun words in all the tokenized sentences corresponding to specific hotel reviews. The top most occurring words will become the features of our hotel reviews.

Step 5: Repeat step 5-8 until each sentence of the review is iterated.

Step 6: Now iterate through each sentence in the reviews and check whether the feature is present in the sentence or not. If the feature is present in the tokenize sentence, then find the adjectives associated with that feature in the sentence.

Step 7: Maintain two counters associated with each identified feature. First counter stores the positive count associated with the feature and second counter stores the negative count associated with the feature.

Step 8: Now check the orientation of the adjectives associated with the feature using the SentiWordNet. It tells whether the adjective is positive or negative.

```
orientation=sentiwordnet.senti_synset(word)
if(orientation.pos_score()>orientation.neg_score()):
   return True
elif(orientation.pos_score()<orientation.neg_score()):
   return False
```

Here, True represents that an adjective is positive and false represents that an adjective is negative. If orientation is positive, then increase the positive counter else increase the negative counter.

```
if wordOrien is True:
   opinionList[feature][0] += 1
elif wordOrien is False:
   opinionList[feature][1] += 1
```

Step 9: Create a pretty table with three columns (Features, Positive Rating (%), Negative Rating (%)) and for each identified feature calculate the percentage of negative and percentage of positive rating using the positive and negative counter associated to that feature and add them to the pretty table. The percentage of positive rating of a particular feature is calculated as the ratio of total the number of positively oriented adjectives associated with that feature to the total number of adjectives associated with that feature. The percentage of negative rating of a particular feature is calculated as the ratio of total the number of negatively oriented adjectives associated with that feature to the total number of adjectives associated with that feature.

Opinion list is the 2D array whose 0th index stores the positive rating percentage of the feature and the 1st index stores the negative rating percentage of the feature.

```
t = PrettyTable(['Features', 'Positive Rating (%)', 'Negative Rating (%)'])
 opinionList[feature][0] = round(100*opinionList[feature][0]/count,2)
opinionList[feature][1] = round(100*opinionList[feature][1]/count,2)
 t.add_row([feature,opinionList[feature][0],opinionList[feature][1]])
```

Step10: Calculate the overall positive and negative rating of the hotel based on the percentage of positive rating and negative rating of all the features corresponding to that hotel. The overall positive rating of a particular hotel is calculated by adding positive rating percentage of each feature divided by the sum of positive and negative rating percentages of each feature. Similarly, the overall negative rating of a particular hotel is calculated by adding negative rating percentage of each feature divided by the sum of positive and negative rating percentages of each feature.

An Illustration: Let us understand the algorithm with the help of an example. Consider the following review:

"It was a lovely hotel. Large room with a comfortable bed."

Step 1: Firstly, the given review consisting of multiple sentences would be sentence tokenized as shown below:

```
Out[4]: ['It was a lovely hotel.', 'Large room with a comfortable bed.']
```

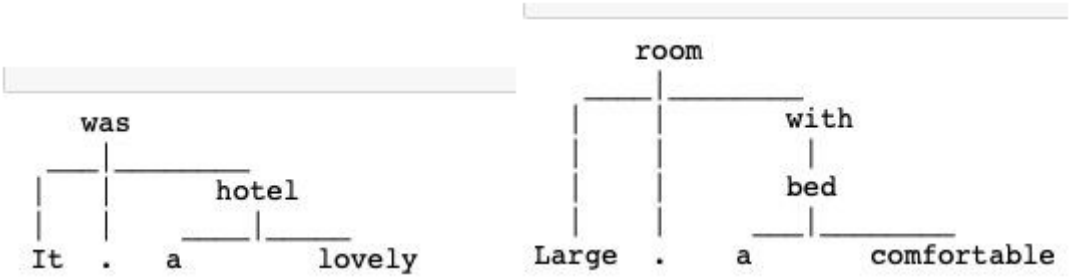
Step 2: Thereafter, the two tokenized sentences will be word tokenized and POS tagged shown below:

```
Out[5]: [('It', 'PRP'),
         ('was', 'VBD'),
         ('a', 'DT'),
         ('lovely', 'JJ'),
         ('hotel', 'NN'),
         ('.', '.')]
```

```
Out[6]: [('Large', 'JJ'),
         ('room', 'NN'),
         ('with', 'IN'),
         ('a', 'DT'),
         ('comfortable', 'JJ'),
         ('bed', 'NN'),
         ('.', '.')]
```

The features selected from the POS tagged sentences are: ['hotel', 'room', 'bed'].

Step 3: Once, features are extracted using POS tagging, then the opinion words corresponding to each feature are identified using spaCy parse tree as shown below:



The opinion words found using spaCy parse tree for each feature are: ‘lovely’ corresponding to hotel, ‘large’ corresponding to room and ‘comfortable’ corresponding to bed.

Step 4: Finally, we identify the orientation of each opinion word (lovely, large and comfortable) using SentiWordNet. The orientation of these three opinion words is positive.

Step 5: Once the orientation of each opinion word corresponding to a particular feature is determined using SentiWordNet, we quantify each feature using two metrics viz. percentage of positive rating and percentage of negative rating. Thereafter, the overall positive rating and overall negative rating of that hotel is determined based on the quantified value of all features associated with the hotel.

5.1 Shortcomings of the Algorithm

Following are few shortcomings on which we intend to work on in our future study:

● Certain useful features might not get selected from the reviews while some irrelevant/inappropriate features may be selected.

● Certain adjectives are ambiguous due to which right polarity of sentence gets difficult to be determined.

● Algorithm is unable to detect sarcasm in the reviews.

● Proper negation handling is not implemented.

● Algorithm takes a simple average to find score but this is not recommended as some adjectives are more prominent than others.

6. Experimental Results

In this section, the experimental results are explained in detailed with the emphasis on the quantification of the amenities for different hotels.

The output has been represented in the form of a table using the prettyTable which is a Python library used for generating simple ASCII tables. The results are provided in table 1 (a-g) which consists of three columns viz. features, positive rating (%) and negative rating (%). Features are the amenities of a hotel like room, bed, location etc. that have been extracted and quantified using two metrics i.e. percentage of positive rating and percentage of negative rating. Percentage of positive rating corresponding to a particular feature/amenity represents the amount of good rating given by the user in his/her review for that feature/amenity. Similarly,

percentage of negative rating corresponding to a particular feature/ amenity represents the amount of poor rating given by the user in his/her review for that feature/amenity. Amenities quantification will help people in making an appropriate selection of the hotel on the basis of the quantification values of their desired amenities across all the available hotels. In other words, these ratings will help the customers to compare different types of hotels according to their needs and then select the hotel which fits best according to their requirements. For example, if a person wants a very comfortable room, then the person will analyze the quantification values of the amenity 'room' across all the hotels. Since, the quantification value of amenity 'room' is maximum corresponding to hotel Hilton (57.14%), therefore the person will select hotel Hilton for his stay. Quantification of amenities will not only be useful to common customers, but will also be crucial to hotel owners. Owners will try to improve the services of their hotels after analyzing the quantified amenities generated by this project. Improving services will in turn increase the business sales. Once each feature has been quantified using the percentage of positive and negative rating, then the overall positive rating and overall negative rating of that hotel is determined based on the rating of all features/amenities associated with the hotel.

Since, there are only 80 reviews (limited number of reviews) pertaining to a hotel, therefore the number of features retrieved using POS tagging is not the same for all the hotels and may vary from one hotel to another i.e. one hotel may have more number of features than the other hotel. There are very few common features across all the hotels is very less. In this work, the authors have considered top-100 most commonly used features. Table 1 (a-g) represent the quantified values (percentage of positive and negative rating) of all the features corresponding to each of the 20 hotels along with the overall rating of these hotels.

Table 1 (a-g): Quantified Values of all the Features and Overall Rating of all 20 Hotels

Affinia hotel			**Allegro hotel**			**Amalfi hotel**		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	room	60	40	hotel	70.59	29.41
room	71.43	28.57	hotel	60.87	39.13	room	70.59	29.41
staff	76.92	23	time	25	75	night	50	50
night	33.33	66.67	staff	80	20	floor	50	50
stay	66.67	33.33	bed	28.57	71.43	view	88.89	11.11
bed	25	75	service	80	20	bed	10	90
service	75	25	i	51.75	48.25	review	66.67	33.33
elevator	33.33	66.67	door	50	50	place	50	50
review	28.57	71.43	thing	75	25	i	65.79	34.21
pillow	50	50	city	75	25	manager	66.67	33.33
place	66.67	33.33	friend	42.86	57.14	experience	81.82	18.18
experience	15.38	84.62	bar	50	50	bathroom	33.33	66.67
way	50	50	experience	42.86	57.14	service	66.67	33.33
view	41.67	58.33	price	50	50	choice	66.67	33.33
food	75	25	end	25	75	city	50	50
price	50	50				thing	75	25
hallway	66.67	33.33				suite	50	50
i	54.22	45.78						
rate	66.67	33.33						
end	90.91	9.09						
group	50	50						
shopping	50	50						

city	50	50						
Overall Rating of the Hotel Positive Rating = 54.24 Negative Rating = 45.76			**Overall Rating of the Hotel Positive Rating = 53.13 Negative Rating = 46.87**			**Overall Rating of the Hotel Positive Rating = 59.57 Negative Rating = 40.43**		

(a)

Conrad hotel			**Fairmont hotel**			**Hard Rock hotel**		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	hotel	60	40	hotel	60	40
room	71.43	28.57	room	71.43	28.57	room	71.43	28.57
staff	76.92	23	staff	76.92	23	staff	76.92	23
night	33.33	66.67	night	33.33	66.67	night	33.33	66.67
stay	66.67	33.33	stay	66.67	33.33	stay	66.67	33.33
bed	25	75	bed	25	75	bed	25	75
service	75	25	service	75	25	service	75	25
elevator	33.33	66.67	elevator	33.33	66.67	elevator	33.33	66.67
review	28.57	71.43	review	28.57	71.43	review	28.57	71.43
pillow	50	50	pillow	50	50	pillow	50	50
place	66.67	33.33	place	66.67	33.33	place	66.67	33.33
experience	15.38	84.62	experience	15.38	84.62	experience	15.38	84.62
way	50	50	way	50	50	way	50	50
view	41.67	58.33	view	41.67	58.33	view	41.67	58.33
food	75	25	food	75	25	food	75	25
price	50	50	price	50	50	price	50	50
hallway	66.67	33.33	hallway	66.67	33.33	hallway	66.67	33.33
i	54.22	45.78	i	54.22	45.78	i	54.22	45.78
rate	66.67	33.33	rate	66.67	33.33	rate	66.67	33.33
end	90.91	9.09	end	90.91	9.09	end	90.91	9.09
group	50	50	group	50	50	group	50	50
shopping	50	50	shopping	50	50	shopping	50	50
city	50	50	city	50	50	city	50	50
Overall Rating of the Hotel Positive Rating = 59.03 Negative Rating = 40.98			**Overall Rating of the Hotel Positive Rating = 56.41 Negative Rating = 43.59**			**Overall Rating of the Hotel Positive Rating = 50.56 Negative Rating = 49.24**		

(b)

Hilton hotel			**Homewood hotel**			**Hyatt hotel**		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	hotel	60	40	hotel	60	40
room	71.43	28.57	room	71.43	28.57	room	71.43	28.57
staff	76.92	23	staff	76.92	23	staff	76.92	23
night	33.33	66.67	night	33.33	66.67	night	33.33	66.67

stay	66.67	33.33	stay	66.67	33.33	stay	66.67	33.33
bed	25	75	bed	25	75	bed	25	75
service	75	25	service	75	25	service	75	25
elevator	33.33	66.67	elevator	33.33	66.67	elevator	33.33	66.67
review	28.57	71.43	review	28.57	71.43	review	28.57	71.43
pillow	50	50	pillow	50	50	pillow	50	50
place	66.67	33.33	place	66.67	33.33	place	66.67	33.33
experience	15.38	84.62	experience	15.38	84.62	experience	15.38	84.62
way	50	50	way	50	50	way	50	50
view	41.67	58.33	view	41.67	58.33	view	41.67	58.33
food	75	25	food	75	25	food	75	25
price	50	50	price	50	50	price	50	50
hallway	66.67	33.33	hallway	66.67	33.33	hallway	66.67	33.33
i	54.22	45.78	i	54.22	45.78	i	54.22	45.78
rate	66.67	33.33	rate	66.67	33.33	rate	66.67	33.33
end	90.91	9.09	end	90.91	9.09	end	90.91	9.09
group	50	50	group	50	50	group	50	50
shopping	50	50	shopping	50	50	shopping	50	50
city	50	50	city	50	50	city	50	50
Overall Rating of the Hotel Positive Rating = 48.44 Negative Rating = 51.56			**Overall Rating of the Hotel Positive Rating = 51.43 Negative Rating = 48.57**			**Overall Rating of the Hotel Positive Rating = 57.57 Negative Rating = 42.43**		

(c)

Intercontinental hotel			**James hotel**			**Knickerbocker hotel**		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	hotel	60	40	hotel	60	40
room	71.43	28.57	room	71.43	28.57	room	71.43	28.57
staff	76.92	23	staff	76.92	23	staff	76.92	23
night	33.33	66.67	night	33.33	66.67	night	33.33	66.67
stay	66.67	33.33	stay	66.67	33.33	stay	66.67	33.33
bed	25	75	bed	25	75	bed	25	75
service	75	25	service	75	25	service	75	25
elevator	33.33	66.67	elevator	33.33	66.67	elevator	33.33	66.67
review	28.57	71.43	review	28.57	71.43	review	28.57	71.43
pillow	50	50	pillow	50	50	pillow	50	50
place	66.67	33.33	place	66.67	33.33	place	66.67	33.33
experience	15.38	84.62	experience	15.38	84.62	experience	15.38	84.62
way	50	50	way	50	50	way	50	50
view	41.67	58.33	view	41.67	58.33	view	41.67	58.33
food	75	25	food	75	25	food	75	25
price	50	50	price	50	50	price	50	50
hallway	66.67	33.33	hallway	66.67	33.33	hallway	66.67	33.33

i	54.22	45.78	i	54.22	45.78	i	54.22	45.78
rate	66.67	33.33	rate	66.67	33.33	rate	66.67	33.33
end	90.91	9.09	end	90.91	9.09	end	90.91	9.09
group	50	50	group	50	50	group	50	50
shopping	50	50	shopping	50	50	shopping	50	50
city	50	50	city	50	50	city	50	50
Overall Rating of the Hotel Positive Rating = 57.85 Negative Rating = 42.15			**Overall Rating of the Hotel Positive Rating = 57.85 Negative Rating = 42.15**			**Overall Rating of the Hotel Positive Rating = 56.56 Negative Rating = 43.44**		

(d)

Omni hotel			**Monaco hotel**			**Palmer hotel**		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	hotel	60	40	hotel	60	40
room	71.43	28.57	room	71.43	28.57	room	71.43	28.57
staff	76.92	23	staff	76.92	23	staff	76.92	23
night	33.33	66.67	night	33.33	66.67	night	33.33	66.67
stay	66.67	33.33	stay	66.67	33.33	stay	66.67	33.33
bed	25	75	bed	25	75	bed	25	75
service	75	25	service	75	25	service	75	25
elevator	33.33	66.67	elevator	33.33	66.67	elevator	33.33	66.67
review	28.57	71.43	review	28.57	71.43	review	28.57	71.43
pillow	50	50	pillow	50	50	pillow	50	50
place	66.67	33.33	place	66.67	33.33	place	66.67	33.33
experience	15.38	84.62	experience	15.38	84.62	experience	15.38	84.62
way	50	50	way	50	50	way	50	50
view	41.67	58.33	view	41.67	58.33	view	41.67	58.33
food	75	25	food	75	25	food	75	25
price	50	50	price	50	50	price	50	50
hallway	66.67	33.33	hallway	66.67	33.33	hallway	66.67	33.33
i	54.22	45.78	i	54.22	45.78	i	54.22	45.78
rate	66.67	33.33	rate	66.67	33.33	rate	66.67	33.33
end	90.91	9.09	end	90.91	9.09	end	90.91	9.09
group	50	50	group	50	50	group	50	50
shopping	50	50	shopping	50	50	shopping	50	50
city	50	50	city	50	50	city	50	50
Overall Rating of the Hotel Positive Rating = 49.60 Negative Rating = 50.40			**Overall Rating of the Hotel Positive Rating = 67.13 Negative Rating = 31.87**			**Overall Rating of the Hotel Positive Rating = 61.08 Negative Rating = 38.92**		

(e)

Sofitel hotel			Sheraton hotel			Talbott hotel		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	hotel	60	40	hotel	60	40
room	71.43	28.57	room	71.43	28.57	room	71.43	28.57
staff	76.92	23	staff	76.92	23	staff	76.92	23
night	33.33	66.67	night	33.33	66.67	night	33.33	66.67
stay	66.67	33.33	stay	66.67	33.33	stay	66.67	33.33
bed	25	75	bed	25	75	bed	25	75
service	75	25	service	75	25	service	75	25
elevator	33.33	66.67	elevator	33.33	66.67	elevator	33.33	66.67
review	28.57	71.43	review	28.57	71.43	review	28.57	71.43
pillow	50	50	pillow	50	50	pillow	50	50
place	66.67	33.33	place	66.67	33.33	place	66.67	33.33
experience	15.38	84.62	experience	15.38	84.62	experience	15.38	84.62
way	50	50	way	50	50	way	50	50
view	41.67	58.33	view	41.67	58.33	view	41.67	58.33
food	75	25	food	75	25	food	75	25
price	50	50	price	50	50	price	50	50
hallway	66.67	33.33	hallway	66.67	33.33	hallway	66.67	33.33
i	54.22	45.78	i	54.22	45.78	i	54.22	45.78
rate	66.67	33.33	rate	66.67	33.33	rate	66.67	33.33
end	90.91	9.09	end	90.91	9.09	end	90.91	9.09
group	50	50	group	50	50	group	50	50
shopping	50	50	shopping	50	50	shopping	50	50
city	50	50	city	50	50	city	50	50
Overall Rating of the Hotel Positive Rating = 55.0 Negative Rating = 45.0			**Overall Rating of the Hotel Positive Rating = 56.28 Negative Rating = 43.72**			**Overall Rating of the Hotel Positive Rating = 55.72 Negative Rating = 44.28**		

(f)

Swissotel hotel			Ambassador hotel		
Features	**Positive Rating (%)**	**Negative Rating (%)**	**Features**	**Positive Rating (%)**	**Negative Rating (%)**
hotel	60	40	hotel	60	40
room	71.43	28.57	room	71.43	28.57
staff	76.92	23	staff	76.92	23
night	33.33	66.67	night	33.33	66.67
stay	66.67	33.33	stay	66.67	33.33
bed	25	75	bed	25	75
service	75	25	service	75	25
elevator	33.33	66.67	elevator	33.33	66.67
review	28.57	71.43	review	28.57	71.43

pillow	50	50	pillow	50	50
place	66.67	33.33	place	66.67	33.33
experience	15.38	84.62	experience	15.38	84.62
way	50	50	way	50	50
view	41.67	58.33	view	41.67	58.33
food	75	25	food	75	25
price	50	50	price	50	50
hallway	66.67	33.33	hallway	66.67	33.33
i	54.22	45.78	i	54.22	45.78
rate	66.67	33.33	rate	66.67	33.33
end	90.91	9.09	end	90.91	9.09
group	50	50	group	50	50
shopping	50	50	shopping	50	50
city	50	50	city	50	50
Overall Rating of the Hotel Positive Rating = 55.0 Negative Rating = 45.0			**Overall Rating of the Hotel Positive Rating = 64.0 Negative Rating = 36.0**		

(g)

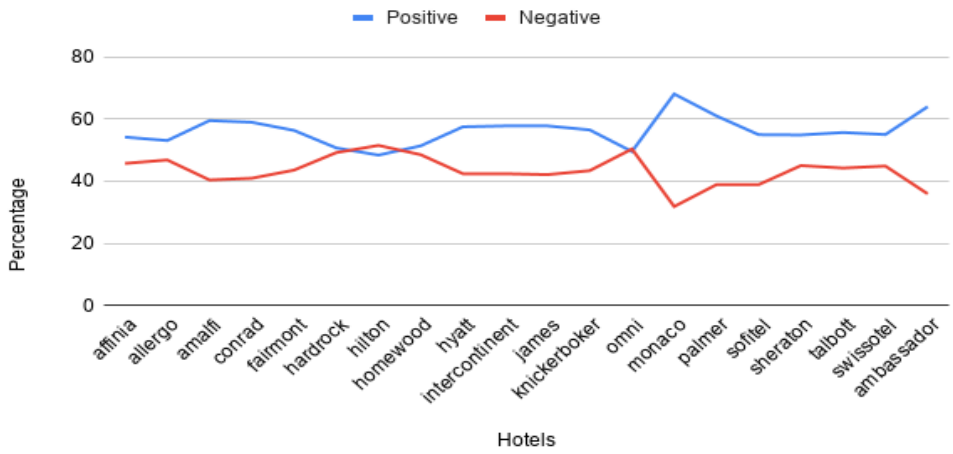


Figure 8: Overall Rating of Hotels

Figure 8 shows the overall rating of each hotel. The blue line depicts the overall positive rating and the red line depicts the overall negative rating of a particular hotel. These ratings will help the users to compare different hotels and then select a particular hotel according to his/her need. However, comparison of hotels based on overall rating (positive and negative) will be more meaningful if comparison of different amenities of these hotels is also provided. For instance, as can be seen from figure 8, the overall rating of hotel 'Hilton' is more inclined toward the negative rating than that of positive rating. But the figure does not provide a clear distinction between the good amenities and bad amenities of this hotel. Thus, this ambiguity is resolved in figure 9 which shows the comparison of most commonly referred amenities with respect to all the 20 hotels.

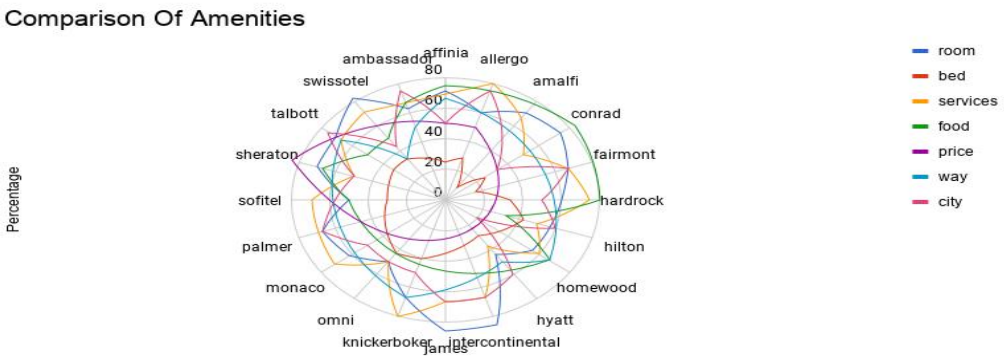


Figure 9: Comparison of Amenities

Figure 9 gives a detailed analysis of each of the amenities/features of different hotels. For this detailed analysis, the most common features among all the hotels are extracted manually. That is, all these hotels have been compared for these common features for a better comprehension. The most common features picked are room, bed, services, food, price, way and city. All other remaining features are not that much effective for a comparison among all hotels as they would be present in one hotel's review and absent in another hotel's quantified values.

Figure 9 helps the customer to check the rating of specific features of the hotels and help them to compare the specific feature of one hotel with other hotels. This chart is based on the radar chart. The innermost circle or point represents 0 value and this value increases with the outward concentric circles. Thus, the outermost circle represents maximum value. Using this specific feature rating users can select the hotels with the amenities e.g. the amenity room can be compared for each hotel. For 'Swissotel' hotel the value for room amenity lies between concentric circles of 20 and 40 and has a minimum value of 33.33. On the other hand, for hotels Intercontinental and James, value for room amenity is outside the concentric circle of value 80 and has values around 85. Thus, 'James' hotel and 'Intercontinental' hotel would be very beneficial for the customers looking for good room facilities. In this way, customers will be at an ease to take decision of choosing required hotels based on the values of different amenities. This rating will not only be beneficial to the customers, but also the employees of the hotel. They can get an idea about what amenities they can put forth as the hotel's unique features and what amenities of the hotel need improvement. For instance, employees of 'Swissotel' hotel can improve the rooms of the hotel as the value of this amenity is the lowest. This will enable them to uplift their business by improving upon the weakness of the hotel. Similarly, other hotels can take advantage of this quantification and can work on their specific target of action for improvements.

7. Conclusion

In today's time, users are getting more and more involved on social media platforms and discussion forums which allow audiences to share their perspective or give ratings and reviews about a particular item/product or service. Hence, a large amount of data containing reviews and opinions are generated. Such data possess a great value and worth if harnessed properly as people's review about a particular product/service plays a significant role in determining the decision of a potential buyer. These reviews are crucial not only for the buyers/customers, but also for the sellers as these reviews can have a huge impact on their revenue. Online reviews determine the reputation of the brand of the product which can either make or break the e-commerce careers of the employees.

The work focused on developing a prototype which could be used to provide rating to any product or services based on the online reviews. In our work, online reviews given by the users for the hotels were considered. The core idea of the paper was to extract important and useful features (amenities) from the reviews of the hotel i.e. amenities extraction and based on that perform a semantic analysis providing a comprehensive quantification about the hotel. The intent was to make the task of hotel selection at the user end much easier and efficient by providing the users a more comprehensive quantification of amenities as compared to those available on the basis of hotel's average ratings.

In order to conduct empirical validation, the Deceptive opinion spam dataset was used which is a corpus consisting of truthful and deceptive hotel reviews of 20 Chicago hotels. The corpus contains a total of 1600 reviews taken from different sources like TripAdvisor, Expedia, Orbitz etc. Firstly, pre-processing of all 1600 reviews which were in the textual format was carried out using sentence and word tokenization and Part-of–Speech (POS) tagging. Thereafter, amenities extraction was carried out which dealt with extraction of words from the textual reviews that tell us about the hotel like rooms, service, staff, location, food etc. Once, amenities corresponding to each hotel were extracted, then identification of opinion words (adjectives) corresponding to each of these amenities was carried out. This was followed by identifying the orientation of the opinion words (positive, negative). Once the orientation of each opinion word corresponding to a particular amenity was determined, then each amenity was quantified using two metrics viz. percentage of positive rating and percentage of negative rating. Thereafter, the overall positive rating and overall negative rating of that hotel was determined based on the quantified value of all amenities associated with the hotel.

This kind of quantitative approach has a huge impact on the quality of the human decision. From quantified scores obtained by the prototype further a complete summary about the product could be generated which in turn helps the user in getting a concrete idea about each and every attribute of the product under consideration. On the other hand, the seller of a product or service gets the idea about what amenities he can put forth as the product's unique selling proposition and what amenities of the product/service needs improvement.

References

[1] O'Reilly, T. (2005). What is Web 2.0, O'Reilly Network.

[2] Sasikala, P., Sheela, Immaculate. (2018). Sentiment Analysis and Prediction of Online Reviews with Empty Ratings. International Journal of Applied Engineering Research. 13. 11525-11531

[3] Chen, C. C., Tseng, Y. (2011). Quality evaluation of product reviews using an information quality framework. Decision Support Systems, 50(4), pp. 755-768.doi:10.1016/j.dss.2010.08.023

[4] Kumar,H., Harish, B., Darshan,H. (2018). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method. International Journal of Interactive Multimedia and Artificial Intelligence

[5] Ortigosa, A., Martín, J., Carro, R. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, pp. 527-541

[6] Xiaomei, Z., Jing,Y., Jianpei, Z., Hongyu, H. (2018). Microblog Sentiment Analysis with Weak Dependency Connections. Knowledge-Based Systems, 142, pp. 170–180, doi:10.1016/j.knosys.2017.11.035

[7] Bhatia, P., Nath, R. (2020). Using Sentiment Analysis In Patient Satisfaction: A Survey. Advances in Mathematics: Scientific Journal, 9(6), pp. 3803–3812, ISSN: 1857-8365 (printed), 1857-8438 (electronic)

[8] Pang, B., Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proceedings of the 42nd annual meeting on Association for Computational Linguistics, pp. 271

[9] Tripathy, A., Agarwal, A., Rath,S. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. Procedia Computer Science, 57, pp. 821–829

[10] Latha, I., Varma, G., Govardhan, A. (2012). Preprocessing the informal text for efficient sentiment analysis. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 1(2), pp. 58-61

[11] Manek, A., Shenoy, P., Mohan, M., Venugopal, K. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World Wide Web, 20(2), pp.135-154

[12] Norambuena, B. K., Lettura, E.F., Villegas, C.M. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. Intelligent Data Analysis. 23(1), pp. 191-214, 10.3233/IDA-173807.

[13] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, pp. 417-424.

[14] Pang, B., Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity analysis using subjectivity summarization based on minimum cuts. ACL, pp. 271-278.

[15] Turney, P. D., Littman, M. L. (2003). Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus. Technical report, ERB-1094 NRC-44929

[16] Hu, M., Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, pp. 168-177

[17] Kim, S.M., Hovy, E. (2004). Determining the sentiment of opinions. COLING, Proceedings of the 20th international conference on Computational Linguistics

[18] Tong, R.M. (2001). An operational system for detecting and tracking opinions in on-line discussion. Working Notes of the SIGIR Workshop on Operational Text Classification, New Orleans, Louisianna, pp. 1-6

[19] Wilson, T., Wiebe, J., Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), Vancouver, pp. 347–354

[20] Agarwal, A., Biadsy, F., McKeown, K. (2009). Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-Grams. Proceedings of 12th Conference of the European Chapter of the Association for Computational Linguistics, pp. 24-32, 10.3115/1609067.1609069

[21] Go, A., Bhayani, R., Huan, L. (2009). Twitter sentiment classification using distant supervision. Technical report, Stanford

[22] Bermingham, A., Smeaton, A. (2010). Classifying sentiment in microblogs: is brevity an advantage is brevity an advantage? ACM, pp. 1833–1836

[23] Pak, A., Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC, pp. 1320-1326.

[24] Zimbra, D., Ghiassi, M., Lee, S. (2016). Brand related twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks. 49th Hawaii International Conference on System Sciences (HICSS), IEEE, pp. 1930–1938.

[25] Subasic, P., Huettner, A. (2001). Affect analysis of text using fuzzy semantic typing. Fuzzy Systems, IEEE Transactions, pp. 483 - 496

[26] Wawre1, S.V., Deshmukh, S.N. (2016). Sentiment Classification using Machine Learning Techniques. International Journal of Science and Research (IJSR), 5(4), pp. 819-821

[27] Zhou, J., Huang, J.X., Chen, Q., Hu, Q.V., Wang, T., He, L. (2019), Deep learning for aspect-level sentiment classification: Survey, vision, and challenges. *IEEE Access*, 7, pp. 78454-78483, 2019

[28] Wu, Y., Zhang, Q., Huang, X., Wu, L. (2009). Phrase dependency parsing for opinion mining. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 3, pp. 1533–1541

[29] Chen, M., Yao, T. (2010). Combining dependency parsing with shallow semantic analysis for chinese opinion-element relation identification. 4th International Universal Communication Symposium, Beijing, China, pp. 299–305

[30] Mukherjee, S., Bhattacharyya, P. (2012). Feature specific sentiment analysis for product reviews. Computational Linguistics and Intelligent Text Processing, pp. 475–487

[31] Lin, C., He, Y. (2009). Joint sentiment/topic model for sentiment analysis. Proceedings of the 18th ACM conference on Informationand knowledge management, ACM, pp. 375–384

[32] Cernian, A., Sgarciu, V., Martin, B. (2015). Sentiment analysis from product reviews using SentiWordNet as lexical resource. 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). Bucharest, Romania doi:10.1109/ecai.2015.7301224

[33] Ohana, B., Tierney, B. (2009). Sentiment classification of reviews using SentiWordNet. 9th IT&T Conference, Technological University Dublin, Dublin, Ireland, doi:10.21427/D77S56

[34] Colhon, M., Vladutescu, S., Negrea, X. (2017). How Objective a NeutralWord Is? A Neutrosophic Approach for the Objectivity Degrees of Neutral Words. Symmetry, 9, 280, doi:10.3390/sym9110280