Aspect Based Sentiment Analysis on Travel Review Rating Prediction

S.N. De Silva

2021

Aspect Based Sentiment Analysis on Travel Review Rating Prediction

A dissertation submitted for the Degree of Master of Business Analytics

S.N. De Silva
University of Colombo School of Computing
2021

Abstract

Online reviews play an integral role in the process of decision making among the consumers of a product or service. In the travel and tourism industry traveller reviews and ratings make a significant impact on the choices made by travellers. Existing rating systems are designed to collect overall rating and plain-text review. A rating of a particular aspect is a good metric for the service provider to improve the service and consumer to find the perfect fit without reading through reviews. Therefore a systematic approach for predicting a rating, for different aspect categories in reviews have been identified as a need. This study aims to introduce a systematic approach to calculate the aspect category based rating of customer reviews in travel tourism domain. This study adapts a novel approach Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction, with the aim to extract all aspect-category-opinion-sentiment quadruples in a review sentence while considering implicit aspects and opinions and finally calculate the aspect based rating in the travel reviews. The experiment results demonstrate that this method is effective in generating the aspect based rating even though a small number(782) of review sentences have been used.

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: *Suhashi Nihara De Silva*

Registration Number: *2019ba005*

Index Number: *19880057*

Signature: Date: 19/11/2022

This is to certify that this thesis is based on the

work of ~~Mr~~./Ms. Suhashi Nihara De Silva

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Viraj Welgama

Signature: Date: 25 / 02 / 2023

Acknowledgements

I would like to express my gratitude and appreciation to the Postgraduate division of University of Colombo School of Computing for giving the opportunity to conduct an individual project which involved in identifying a contemporary research problem in business analytics and finding a solution via a scientific method. Special thanks to my project supervisor, Mr. Viraj Welgama for the guidance and continuous support given in carrying out this project.

I would also like to sincerely thank all my family members, relatives and friends who contributed and supported immensely throughout this period of one year.

Finally i express my profound thanks to all who have indirectly guided and supported to conduct this project successfully.

Table of Contents

List of Figures

List of Tables

Nomenclature

Acronyms / Abbreviations

ABSA Aspect Based Sentiment Analysis

ACOS Aspect-Category-Opinion-Sentiment

BERT Bidirectional Encoder Representation from Transformers

BiLSTM Bidirectional Long Short Term Memory

BIO Begin-Inside-Outside

CNN Convolutional Neural Network

LSTM Long Short Term Memory

MNBC Multinomial Naive Bayes Classifier

PLSI Probabilistic Latent Semantic Indexing

RNN Recurrent Neural Network

SVM Support Vector Machines

Chapter 1

Introduction

According to reports of the World Tourist Organization there has been an up-trend in tourism over the last few decades. People have a wide range of budgets and variety of tastes. On the other hand there are wide variety of resorts, hotels and other attractions which provide distinct experience. Before travelling most people will look on reviews and ratings on rating systems like Google, Trip Advisor or even on Facebook, when deciding where to go and what kind of experience they would prefer. Online reviews have become very common to be found in many websites across the internet and have become the biggest source of social proof. Many people are reluctant to trust businesses that have low ratings. Online reviews expand the conversation about the products and services while very good (or bad) reviews have a way of quickly spreading. Studies show that almost two-thirds of people think that online reviews are an essential part of the decision-making process in purchasing products as well as when traveling

Among the most popular global travel sites that provide travel related reviews are Google, TripAdvisor, Booking.com, Expedia and Yelp. Out of them Trip Advisor is one of the largest travel community, reaching 390 million unique visitors each month and listing 465 million reviews and opinions about more than 7 million accommodations, restaurants, and attractions in 49 markets worldwide (Valdivia, Luzón, and Herrera 2017). When posting reviews on Trip advisor a reviewer could comment on the different rides available and experience of a leisure park in a positive manner, however the service of the staff negatively and overall rating could be given considering both aspects, summarizing their feedback in a scale of 1 to 5. Further people who are reading the comments could have different preferences about the place they would be willing to visit. One would prefer visiting a place with exciting rides and attractions while another would be more interested about the delicious meals and service offered. However, the readers will have to struggle finding the appropriate information since there are a large number of reviews. Readers will find it easier to know the rating based on different aspect categories. On the other hand the business itself would also have the benefit to use these reviews in order to improve their service and customer experience if a rating can be estimated based on aspect. Even so, most commercial off-the-shelf tools are limited to the extraction of a polarity value associated to the whole document, rather than determining the values related to the different aspects(Poria et al. 2016).

A rating of a particular aspect is a good metric for the service provider to improve the service and consumer to find the perfect fit without reading through reviews. This study focuses on predicting a aspect-based rating on a scale of 1-5, for the different aspect categories identified in reviews. In order to represent the reviewers preferences based on different aspect categories, an aspect based sentiment analysis needs to be conducted. Therefore this study will focus on sentiment analysis on aspect level. Aspect-based Sentiment Analysis (ABSA)is a specific Sentiment Analysis that aims to extract most important aspects of an entity and predict the polarity of each aspect from the text. A review of the recent state-of-the-art in ABSA, shows the remarkable growing in finding both aspect, and the corresponding sentiment (Madhoushi, Hamdan, and Zainudin 2019). There are many studies conducted on sentiment analysis at aspect level.

1.1 Motivation

Online reviews have become very popular and plays an integral role in the process of decision making among the public for many businesses. Especially in the travel and tourism industry traveller reviews and ratings have a significant impact for sales. However when ratings are given in websites an overall rating is given by people in most cases and ratings based on different aspects can not be extracted directly. Therefore the aspect based rating is given by the readers and could be subjective for each individual. As a result a systematic rating prediction, based on different aspects in reviews have been identified as a need. Further the identified ratings could be used to get useful insights to improve customer experience.

1.2 Objectives

The objectives of this analysis will be to:

- Identify the different aspect categories people in general comment related to travel experiences
- Identify a systematic approach to calculate aspect based ratings for travel reviews
- Estimate the aspect category based review ratings for the travel reviews

1.3 Project Scope

The proposed study will be based on a dataset extracted from Trip Advisor platform on customer Reviews and ratings of a Universal studio branch located in Orlando United States. Universal studio is one of the most popular theme park chain owned and operated by NBC Universal. There are number of Universal studio theme parks operated in different countries across continents.

Reviews from Trip Advisor consist of title, text, rating, and reviewer's profile. For this study, only the text part of the reviews will be used. The dataset considered for this study consists of 200 reviews in number. The scope of this project will include identifying aspects, aspect categories, opinions and related sentiments of the customer reviews and finally estimate ratings for the 7 different aspect categories identified as Rides, Food, Service, Price, Attractions, Experience and Miscellaneous in a scale of 1-5, for the Universal studio branch in Orlando in the year 2019. Further when identifying aspects and opinions even the implicit aspects and opinions terms will be taken into consideration.

1.3.1 Out of Scope Items

Following will be considered as out of scope when conducting this analysis.

- Reviews given in English language will be considered in this analysis and reviews given in other languages will not be considered
- Sentiments only available in the form of text will be considered in the analysis and reviews given in other forms such as symbols, images will not be considered.

1.4 Research Contributions

In this paper, a method for systematically calculating aspect based rating for travel reviews is being evaluated. Specifically, this paper makes the following major contributions.

- Constructing a new travel review dataset, manually annotated by a domain expert that provides full support for the ACOS quadruple extraction task.
- Experimenting Naive Bayes and Support Vector Machine Learning models adapting new features, comparing and presenting the results for sentence subjectivity classification.
- Demonstrating that the proposed method of ACOS extraction task for travel reviews provides adequate performance compared to other baseline models which support ABSA considering implicit aspects and opinions.

The remainder of this paper is organized as follows. Chapter 2 will discuss related work, followed by the approach taken or the methodology of this project being described under Chapter 3. Chapter 4 will present the technical details of this project and the experimental results and evaluation strategy. Finally the Conclusion and Future work will be presented under the last section, chapter 5.

Chapter 2

Literature Review

Aspect Based Sentiment Analysis[ABSA] has been a popular research area during last few decades. Researches on Aspect Based Sentiment Analysis has been carried out targeting several domains such as electronic product reviews(laptop, camera, and phone) and hospitality reviews (restaurant, hotels). According to Zhang and B. Liu (2014) aspect extraction methods are classified into 3 categories of language rule, sequential and topic model. Following this Madhoushi, Razak, and Zainudin (2019) classified ABSA methods with 2 added additional categories as Deep Learning and hybrid Models as shown in (Fig.2.1) This section summarizes and critically evaluates recent literature in ABSA methods.

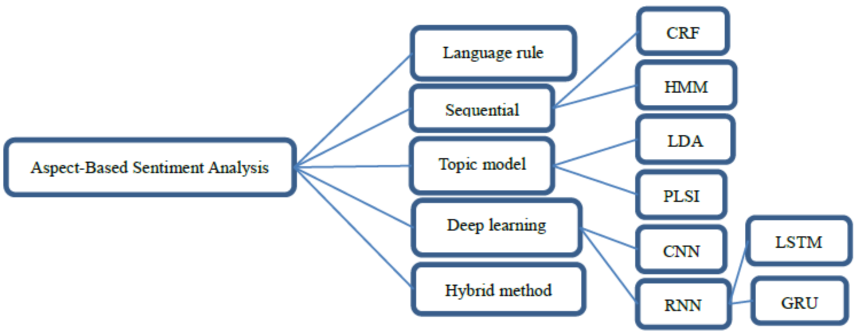


Fig. 2.1 ABSA Methods

2.1 Language Rule Models

In reviews people discuss about several aspects more commonly which gives the idea that aspects should be frequent nouns. However not all frequent nouns are aspects and aspects can be expressed by a noun, adjective, verb or adverb. Therefore, different filtering techniques are usually applied on frequent nouns to filter out nonaspects. Most of these models try to find the most frequent nouns and noun phrases of the reviews in dataset, ordered by decreasing sentence frequency in the first step.

In work presented by Hu and B. Liu (2004) first identifies all frequent noun phrases from full text reviews as candidate aspects. Then two pruning methods are applied to remove those candidate aspects with meaningless string, based on association rule mining, and those which are subsets of others (B. Liu 2012). Marrese-Taylor, Velásquez, and Bravo-Marquez (2014) improves the algorithm to estimate the orientation of sentence for compound aspects. Zhang and B. Liu (2014) presents three methods, noun phrase extraction, Named Entity Recognition and a combination of both for aspect extraction. The semantic-based approach in P. Liu, Joty, and Meng (2015) is similar to Hu and B. Liu (2004). Moreover, they estimate personalized aspect polarity estimation for each individual user from his/her review. Their model removes irrelevant pairs of aspect-sentiment if they are not similar to any of the pre-defined aspects and then group relevant pairs into their corresponding aspect. A pair is grouped into an aspect if the semantic similarity between the noun of the pair and the pre-defined aspect word is above the specific threshold. WordNet Similarity is used to compute the semantic similarity between words.Double propogation was a typical rule-based method proposed for aspect-opinion-sentiment triple extraction by Qiu et al. (2011). The aspect-opinion-sentiment triples are extracted, by making use of the syntactic relations between aspects and opinions to extract them iterative in each review, and rely on the sentiment lexicon to assign sentiments (i.e., Positive, Negative, and Neutral) to aspects and opinions in a bootstrapping manner. Lal and Asnani (2014) can be seen as an advanced extension of method in Hu and B. Liu (2004). It is designed specifically to identify aspects that mention implicitly in review sentences. Secondly, the approach distinguished sentiment words and aspect words with predefined rules. For example, sentiment words can only occur in the rule antecedents, while rule consequents must be aspects. Thirdly association rules are made directly from the Latent Semantic Analysis (LSA) matrix of sentiments and aspects. Wan et al. (2020) introduced one of the state-of-the-art method for aspect-category-sentiment triple extraction, that integrates aspect category-based sentiment classification and aspect extraction in a unified framework by adding the aspect category and the sentiment polarity to the review sentence and using it as the input for BERT (Bidirectional Encoder Transfer) model.

Author	Data Set	Domain	Result (%)
Hu and B. Liu 2004	FBS	Electronic products	Precision: 69.3 Recall: 64.2, Accuracy: 84.2
B. Liu, Hu, and Cheng 2005	-	Electronic products	Precision: 88.9/79.1 Recall: 90.2/82.4
Blair-Goldensohn et al. 2008	Tripadvisor google maps	Restaurant Hotel	Precision: 85.2 Recall: 66 F-score: 74
W. Du and Tan 2009	ctrip	Hotel (in Chinese)	Precision: 78.90
Albornoz et al. 2011	bookings.com	Hotel	F-score : Logistic: 70 LibSVM: 69 FT: 67
Qiu et al. 2011	-	Electronic products	Average : Precision: 88 Recall: 83 F-Score: 86
Wan et al. 2020	SemEval-2015 SemEval-2016	Restaurant	SemEval-2015 : F-Score: 58.09 SemEval-2016 F-Score: 65.44

Table 2.1 Language Rule Models

2.2 Sequential Models

Sequential learning is one of the most prominent information extraction where the current state-of-arts methods have been recognized as Hidden Markov Model(HMM) and Conditional Random Field(CRF). Since they are supervised Learning models, they need to annotate the aspect and non-aspect manually. These methods deduce a function from labeled training data to apply for unlabeled data (Madhoushi, Razak, and Zainudin 2019). A hybrid approach has been presented in Jin, Ho, and Srihari (2009), integrating POS information with the lexicalization technique under the HMM framework. In this model the current tag is related with the previous tag and also corresponds with previous observations (word token and part of speech). Meanwhile a CRF model trained on multi-domain review sentences has been proposed in Jakob and Gurevych (2010). A list of domain independent features such as tokens, POS tags, etc have being used in this work. Another CRF-based model has being discussed in Choi and Cardie (2010) which proposes a set of sequential rules which are extracted using a sequential rule mining technique considering class labels, dependency and word distance. In Kiritchenko et al. (2014) literature, presents a new sequence tagger for aspect terms extraction and supervised classifiers for aspect category detection. A major limitation of the Language rule method is that it requires tuning various parameters manually. Thus, the models cannot be generalized for unseen data set. An end-to-end framework, one of the state-of-the-art approaches for aspect-opinion-sentiment triple extraction, was introduced by L. Xu et al. (2020), by combining the identification of aspects, their corresponding opinions, and their sentiment polarities with a position-aware tagging scheme.

Sequential method overcomes the limitations of language rule methods by learning the parameters from the data automatically. Even though supervised methods achieve reasonable effectiveness, the disadvantage of these models is that they require labeled data for training. Labeled data are not usually available and constructing enough labeled data will be expensive and time consuming. Therefore, it is desired to develop a model that works with unlabeled data or less labeled data (Madhoushi, Razak, and Zainudin 2019)

Author	Data Set	Domain	Result (%)
Jin, Ho, and Srihari 2009	FBS	Electronics	F1: 78.8-82.7 Recall: 64.2, Accuracy: 84.2
Choi and Cardie 2010	The MPQA corpus	-	Precision: 48.0 Recall: 87.8 F-Score: 62.0
Kiritchenko et al. 2014	SamEval 2014	Laptop	Precision: 78.77 Recall: 60.70 F-Score: 68.57
Zhang and B. Liu 2014	SamEval 2014	Laptop Restaurant	Laptop F-score: 65.88 Restaurant F-score: 78.24
L. Xu et al. 2020	14Lap 16Rest	Laptop Restaurant	Laptop F-score: 51.33 Restaurant F-score: 62.86

Table 2.2 Sequential Models

2.3 Topic Models

Topic Modelling is an unsupervised model where topics are being extracted from texts, assuming that text does have a combination of topics and every topic has a probability distribution over it. Probabilistic Latent Semantic Indexing (PLSI) and Latent Dirichlet Allocation (LDA) are two common methods under Topic modelling.

According to Lu, Zhai, and Sundaresan (2009)literature, first the model creates sentiment phrases(head term, modifier), and then models are learned to generate sentiment phrases only and not all the words of a review. They cluster the head terms using PLSI to extract aspects. The polarity of a head term is considered as the polarity of the corresponding short comment. In the model presented in Moghaddam and Ester (2011), it introduces an Interdependent Latent Dirichlet Allocation (ILDA) model that utilize the assumption of inter dependency between aspects and sentiments. In Z. Chen, Mukherjee, and B. Liu (2014), prior knowledge is learned automatically from a large amount of review data available on the review websites. Then this prior knowledge is used by a topic model to find more relevant aspects. Some recent works try to jointly extract aspects and their polarity in a single phase (Moghaddam and Ester 2011). According to a model proposed in Bagheri, Saraee, and De Jong (2014), it introduces a model that can extract aspects automatically using the structure of reviewed sentences. On the contrary with sequential models, there is no need for manually labeled data. In addition, topic models perform both aspect extraction and aggregation at the same time in an unsupervised manner. A limitation of topic model is that it requires a large volume of data to be trained accurately. Further with Topic modelling these models are only able to find some general aspects, and has difficulty in finding detailed aspects (Zhang and B. Liu 2014).

Author	Data Set	Domain	Result(%)
Z. Chen, Mukherjee, and B. Liu 2014	Amazon	Electronic products	Average Precision: 5: 90 Average Precision: 10: 85
Bagheri, Saraee, and De Jong 2014	FBS	Electronic products	Rand Index: 85.18
Lu, Zhai, and Sundaresan 2009	eBay	-	Aspect rating prediction: Correlation: 11-49 Ranking Loss: 15-63 Representative phrases: Precision: 26-59 Recall: 29-63
Moghaddam and Ester 2011	Epinions	Electronic products	Aspect extraction: 83 Polarity prediction: 73

Table 2.3 Topic Models

2.4 Deep Learning Methods

Deep learning automatically learns latent features as distributed vectors and has recently shown better results than many machine learning methods on similar tasks. When applying Aspect Based Sentiment Analysis, in each review there could be more than one aspect and accordingly more than one class for each review. Hence a simple supervised model cannot classify each review to different classes. To overcome this challenge many studies have been conducted and developed models using deep learning techniques (Madhoushi, Razak, and Zainudin 2019). A number of studies have used attention mechanism to let the model learn representation with attention on a specific part of text.

The model proposed in Baziotis, Pelekis, and Doulkeridis (2017) use Bidirectional Long Short Term Memory (BiLSTM) for both aspect and sentence. Then, concatenates the hidden layers and adds attention on top of it. As a work that combine classical models with deep learning in P. Liu, Joty, and Meng (2015) considers the task as BIO sequence labelling problem. A general class of discriminative models based on Recurrent neural networks (RNNs) and word embedding has being proposed in this study. In Dhanush, Thakur, and Diwakar (2016) presents a model made of separate models for aspect extraction and sentiment classification. The first model extracts aspects by tagging aspects in a sentence using RNN and the second model classify sentences using Convolutional Neural Network(CNN). A coupled multi-layer attention framework was proposed by Wang et al. (2017) which performs aspect-opinion co-extraction and has also been adapted in our system. This model is a multi-layer attention network, where each layer consists of a couple of attentions with tensor operators. One attention is used for extracting aspect terms, while other is for extracting opinion terms.

2.5 Hybrid Methods

Considering both pros and cons discussed in previous models, several studies attempted to apply hybrid solutions in customer review domains. According to Xue et al. (2017))literature, the aspect terms and aspect categories are closely related, so a multi-task framework of Bidirectional LSTM for Opinion Target Extraction and CNN for Aspect Category Detection was proposed. The main benefits of this framework is the mutual information sharing of two tasks, in which the CNN can also utilize extra information learned in the BiLSTM to improve its informative features.

Author	Data Set	Domain	Result
P. Liu, Joty, and Meng 2015	SemEval-2014 Task 4	Laptop Restaurant	Best F1-score: Restaurant: 78.00 Laptop: 81.56
Dhanush, Thakur, and Diwakar 2016	SemEval-2014 Task 4	Laptop Restaurant	Precision: 88.6 Recall: 82.4 F1 score: 85.4
J. Xu et al. 2016	Yelp Dataset	Restaurant Computer	Restaurant Accuracy: 68.34 Computer Accuracy: 76.90
Baziotis, Pelekis, and Doulkeridis 2017	SemEval-2017 Task 4	Twitter	F1-score: 82
Wang et al. 2017	SemEval-2014 Task 4 SemEval-2015 Task 12	Laptop Restaurant	F-Score - Laptop Aspect Extraction: 77.80 Opinion Extraction: 80.17 F-Score - Restaurant Aspect Extraction: 85.29 Opinion Extraction: 83.18

Table 2.4 Deep Learning Methods

Similarly, P. Chen et al. (2016) also combined Long Short Term Memory (LSTM) and CNN together for sentiment classification but used LST for generating context embedding and CNN for detecting features. In literature Ye et al. (2017), proposed a dependency-tree based convolutional stacked neural network (DTBCSNN) for aspect term extraction, in which the convolution is included in the sentence's dependency parse trees to capture syntactic and semantic features. This can overcome the practical limitations of sequential models which cannot capture the tree-based dependency information

Author	Data Set	Domain	Result
Wang et al. 2016	SemEval 2014	Laptop Restaurants	Restaurant: Aspect F1-score: 84.93 Sentiment F1-score: 84.14 Laptop: Aspect F- score: 78.42 Sentiment F- score: 79.44
H. Du et al. 2016	-	Electronic products	Accuracy: Electronics: 92.08 Movies and TV: 92.05 CDs and Vinyl: 94.38
Popescu and Etzioni 2007	FBS	Electronic products	Precision: 94 Recall: 76
Blair-Goldensohn et al. 2008	Tripadvisor	Restaurant Hotel	Aspect classification: Precision: 85.2 Recall: 66 F-score: 74
Sauper and Barzilay 2013	Yelp	Restaurant Medical	Aspect cluster prediction: Precision: 74.3 Recall: 86.3 Sentiment classification: Accuracy: 82.5

Table 2.5 Hybrid Methods

Chapter 3

Methodology

This chapter outlines the methodology used for the proposed model in this study along with the data preprocessing steps followed and machine learning models used in different phases. The proposed system consist of three main phases and has been depicted in Figure 3.4.

1. Sentence Subjectivity Classification Model
2. Aspect Category Opinion Sentiment (ACOS) Detection Model
3. Aspect based Review Rating Prediction

Before feeding the data to the main processing pipeline of the proposed methodology preprocessing steps are followed.

3.1 Preprocessing of Data

First the review texts are splitted into sentences and selected preprocessing techniques are followed in order to remove special characters and normalize the accented characters. Next we further process the sentences by removing stop words, tokenizing and lemmatizing the text based on part of speech tags before feeding into the sentence subjectivity classification models. However for the second phase (ACOS detection model), the review sentences are not subjected to removing stop words, tokenizing and lemmatizing. In other words, we only remove special characters and normalize the sentences when feeding to the ACOS model.

3.2 Sentence Subjectivity Classification Model

In the context of this study sentence Subjectivity Classification refers to identifying sentences which has a sentiment value related to the predefined aspects in a review text. There can be two types of sentences found in a review text, subjective sentences and Objective sentences. Note that the terminology used in this study related to sentence being subjective and objective is not from the usual language context and terms used for ABSA tasks. Objective sentences are sentences which are not opinionated or does not contain a sentiment value. For example:

- Objective Sentence: “Went to Universal studios on a Wednesday hoping for smaller crowds”
- Subjective Sentence: "The staff was friendly and helpful"

In the Objective sentence given above the reviewer does not comment or provide any opinion on any of the aspects identified. Therefore there is no use of further processing this type of sentences. In Contrary, in the subjective sentence given above the word ’staff’ can be identified as the aspect of the sentence and the opinions are ’friendly’ and ’helpful’ which has a positive sentiment value. Therefore subjective sentences are the sentences which will be used for further processing under this project in order to find the aspects, categories, opinions and there related sentiment values to predict the aspect based ratings in reviews.

In order to identify the subjective sentences from a given review following machine learning models are being used:

- Multinomial Naive Bayes Classifier(MNBC)
- Support Vector Machines(SVM)

MNBC achieved good results in (Wiebe, Bruce, and O’Hara 1999) work by training model with the binary feature of adjective, adverb, pronoun, modal, number, punctuation, and the position of the sentence. Therefore in this study in order to classify subjective sentences MNBC and SVM is trained using TF-IDF values, Bag of Features and binary feature of ngrams and finally results are being compared.

3.3 Aspect Category Opinion Sentiment Quadruple Detection Model

This is the most important and crucial phase in this system which aims to identify the aspect term, the category related to the aspect term, opinion term and the sentiment polarity of a given sentence. In short form this has being referred to as ACOS quadruple extraction (Cai, Xia, and Yu 2021). In majority of the work carried out related to ABSA identifying explicit aspect terms and explicit opinion terms are only taken into consideration and there is no support for implicit aspects and opinions. Among the baseline models which support implicit aspects and opinions in review sentences, ACOS model provides better performance (ibid.).

Further in order to find a rating related to the predefined aspect categories it is required to first identify the aspect terms and then the related aspect category for which ACOS model provides full support by extracting the aspect, related aspect category opinion and related sentiment quadruples in a review sentence. These entities can be defined as follows.

- Aspect - refers to an entity and its aspect specifying the opinion target, which is normally a word or phrase in the text
- Category - represents a unique predefined category for the aspect in a particular domain
- Opinion - refers to the subjective statement on an aspect, which is normally a subjective word or phrase in the text
- Sentiment - refers to the predefined semantic orientation (e.g. Positive, Negative, or Neutral) toward the aspect.

In a review sentence usually there could be multiple aspects and opinions. Therefore the ACOS Quadruple Extraction task does not only identify four elements, but also merge them into a set of valid quadruples, also considering implicit aspects & opinions. As the implicit aspect & opinion is not explicitly indicated by a word or phrase, in the event of implicit aspects 'a' is being set as null and category c is used to describe the opinion target, and in the event of implicit opinion 'o' is set as null and sentiment 's' is used to describe the semantic orientation. Examples:

1. It was good for all ages -> *Sentence with implicit aspect*
2. It was basically a bunch of teenagers that serve the bare minimum -> *Sentence with implicit opinion*

When analyzing the data set we can identify that around 0.19 of the sentences contain either implicit aspects or implicit opinions which is a considerable amount and therefore considering the implicit aspects and opinions for ABSA tasks is crucial.

- No of Quadruples - 947
- No of Quadruples with implicit aspects/opinions - 169

A quadruple is a collection of the aspect, aspect category, opinion and related sentiment. There are two main steps under this phase where the first step performs aspect-opinion co-extraction, and the second step predicts category-sentiment from the extracted aspect-opinion pairs.

Before training the model data needs to be pre processed into the expected models input format. A pretrained BERT model $bert_{u}ncased_{L} - 6_{H} - 768_{A} - 12$ specification has being used for this system and it is then being fine tuned to learn the specific text classification tasks.

3.3.1 BERT Model

BERT encoder which refers to Bidirectional Encoder Representation from Transformers comprises of stacked encoders. Bert uses transformers to pretrain model for common Natural Language Processing tasks. Transformers use encoders and decoders to solve language translations. In transformer model encoding refers to taking english words simultaneously and generate embeddings simultaneously. Embeddings are transformed into vectors and similar words have closer numbers in the vectors. Encoders are best suited for tasks requiring an understanding of the full sentence.

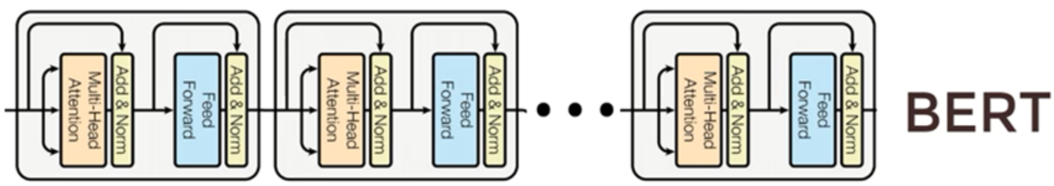


Fig. 3.1 Formation of BERT encoder stacking multiple encoders

Transformers provide a set of preprocessing classes to prepare a dataset for the model. For text classification tasks a tokenizer is used to convert text into a sequence of tokens, which then creates a numeric representation of the tokens and assemble them into tensors. Therefore the text is passed through a tokenizer to convert the words into a sequence of numeric representation of the tokens. Wordpiece Tokenizer which is a tokenization algorithm developed by Google to train pretrained BERT has been used for this task. Wordpiece tokenizer splits on spaces and punctuation. Figure 3.2 depicts the highlevel functioning of a tokenizer providing example of a wordpiece tokenizer.

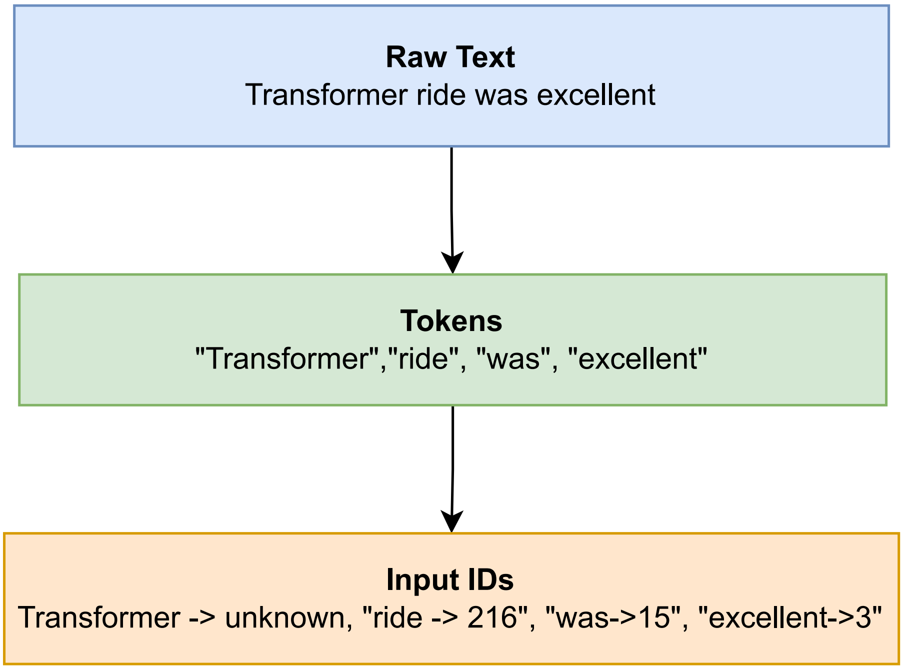


Fig. 3.2 Functioning of a Word Piece Tokenizer: word "transformer is not identified and is considered as a special token

The tokenized text is then passed through BERT encoder. Before passing first insert two [CLS] tokens at the beginning and the end of the review sentence r, and then feed the transformed input to BERT to obtain the context-aware token representations. CLS stands for classification and its added in order to represent sentence-level classification. last hidden state of BERT corresponding to this token (h[CLS]) is used for classification tasks. When solving problems using BERT, 2 main steps are being followed.

1. Pretraining BERT to understand language

2. Fine Tune BERT to learn a specific task

The text passed through the BERT encoder outputs a numerical feature vector. The values of the words are being contextualized into a numerical form, taking into consideration the word and the context of the word in the text(words around it). Once text is passed through the encoder the output feature vector is then being sent for aspect-opinion co-extraction. Figure 3.3 depicts the functioning of BERT model.

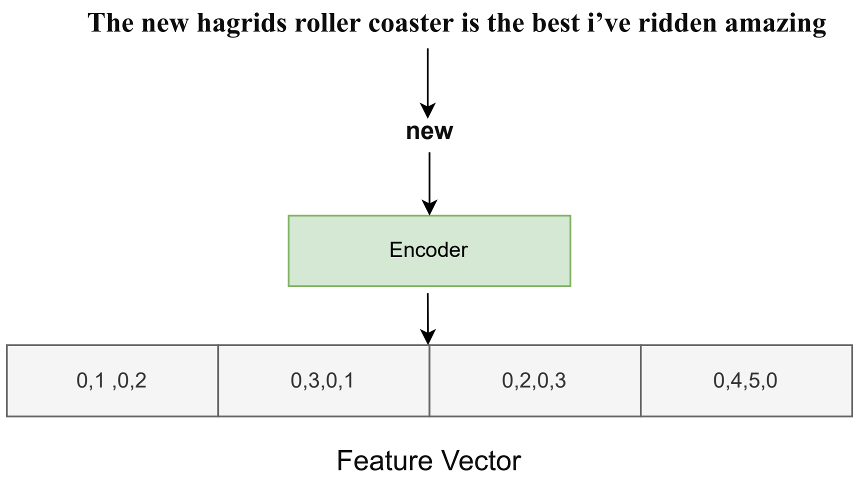


Fig. 3.3 Working of an Encoder: Value of the word and context contextualized in a numerical representation

3.3.2 Aspect-Opinion co-extraction

The explicit aspect-opinion co-extraction is based on a CRF layer with the modified BIO tagging scheme (Cai, Xia, and Yu 2021). Further two binary classification tasks are being applied on the [CLS] tokens to predict whether there are implicit aspects or implicit opinions. Next, we obtain the potential aspect set A and opinion set O, and then carry out Cartesian Product on A and O and produce a set of candidate aspect-opinion pairs (ibid.). Given a review sentence r the input is constructed as follows:

$$x = [[CLS]; r; [CLS]] \tag{3.1}$$

Then feed x to BERT to get the context-aware token representation H:

$$H = [h[CLS], hr, h[CLS]] \tag{3.2}$$

where hr = [h1; : : : ;hn] is the output representation for r and h[CLS] is used for category-sentiment verification. Perform aspect-opinion co-extraction over H by designing it as a single sequence labeling task. Apply modified Begin-Inside-Outside (BIO) tagging scheme, BA; IA; BO; IO; Og, Then feed hr to a CRF layer to extract the aspects and opinions in r. Also apply two binary classification tasks on the [CLS] tokens to predict whether there are implicit aspects or implicit opinions. Then get the potential aspect set A, opinion set O, and perform Cartesian Product on A and O and produce a set of candidate aspect-opinion pairs.

3.3.3 Category-Sentiment Classification

The category-sentiment classification is designed as a multiple multi-class classification problem. For each category c, we merge the average vectors of each aspect-opinion pair a-o, and feed them to a fully-connected layer with the following Softmax function:

$$s^{aoc} = Softmax(W_{aoc}^{T}[ua; uo] + b^{aoc}) \tag{3.3}$$

where

$$s^{aoc} \in Positive, Negative, Neutral, Invalid$$

denotes its sentiment given current a-o and c, or indicates an invalid quadruple.

3.4 Review Rating Prediction

Final step of this study is to estimate a Rating in a scale of 1-5 for the pre defined set of aspect categories. Based on the analysis done on the Universal studio Florida branch, 7 aspect categories were identified as areas which many reviewers comment. Following table 3.1 depicts the aspect categories and few sample aspect words that would fall under each of the category.

According to study conducted the ratings are calculated considering only the number of positive and negative sentences for each aspect category and neutral sen-

Aspect Category	Aspect Words/Phrases
RIDES	Roller coasters, 3D Rides, Harry Potter Ride, Fast and Furious, Men in Black
ATTRACTIONS	Diagon Alley, Volcano Bay, Islands of Adventure
FOOD	Food, Butter Beer, Pizza,
SERVICE	Staff, Tour guide
PRICE	Price, Pricey
EXPERIENCE	Overall experience, Universal experience, One time experience
MISCELLANEOUS	Universal App, Character Encounters, Shows, Parades

Table 3.1 Aspect Categories and Sample Aspect words

tences will be ignored as it does not provide insightful information (Ganu, Elhadad, and Marian 2009). Therefore in this study instead of positive and negative sentences positive and negative quadruples will be taken into consideration in order to calculate aspect based ratings and comment on the quality of the theme park. A quadruple refers to the combination of aspect, related aspect category, opinion and sentiment.

$$Rating = [P/(P+N)*4]+1 \tag{3.4}$$

where P is the number of Positive quadruples in the review, and N is the number of Negative quadruples. The rating is scaled in the [1:5] This rating indicates the overall sentiment for each aspect category expressed in the customer reviews as a whole.

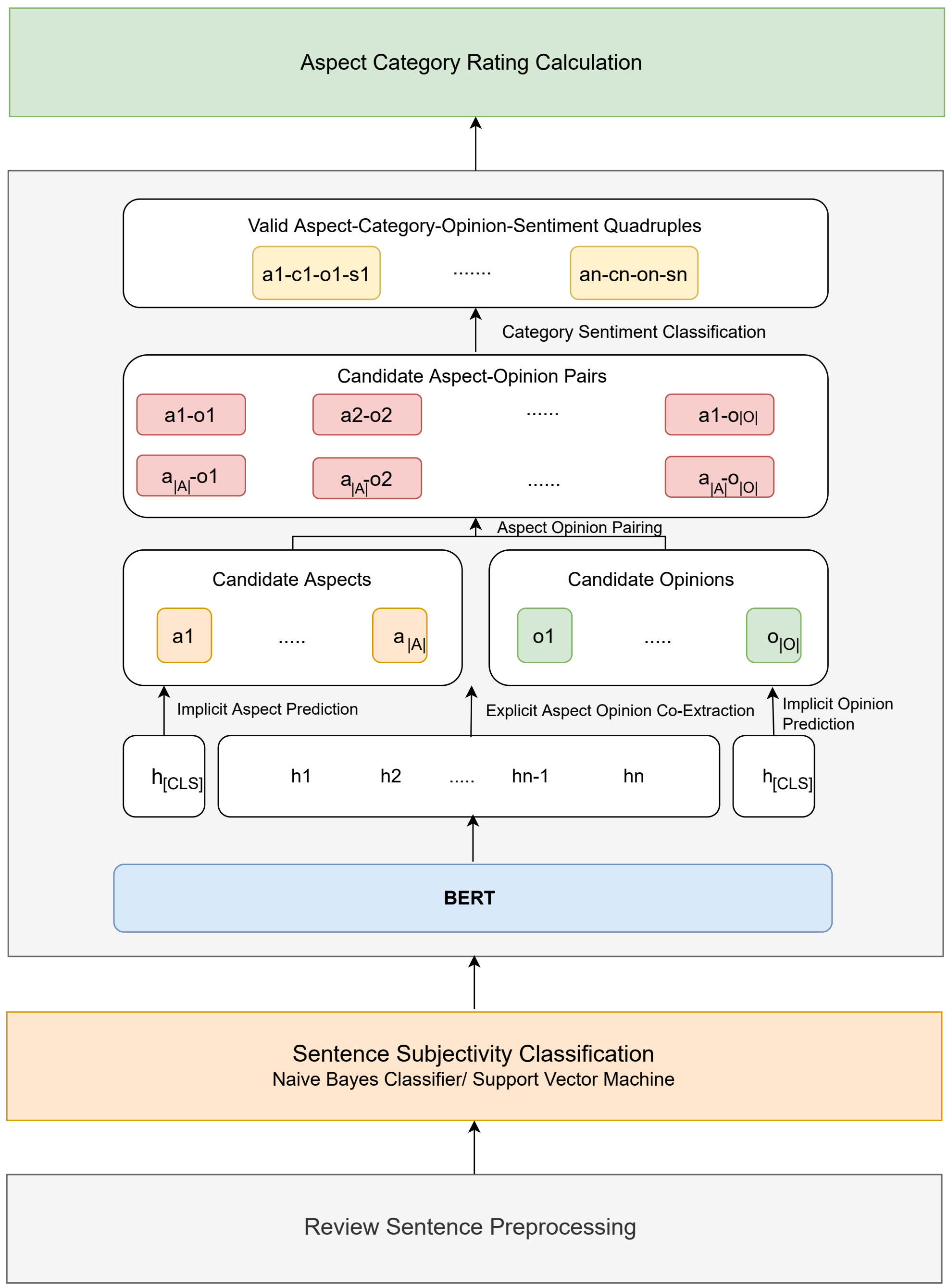


Fig. 3.4 Overview of the main phases in the proposed system

Chapter 4

Experimental Evaluation and Discussion

This chapter focuses on describing the dataset used for building classification models, experiments conducted while implementing the final system and results of different approaches and evaluation of those results.

4.1 Dataset

The dataset containing 200 customer reviews of Universal studio Orlando is manually annotated using a extended Begin-Inside-Outside(BIO) tagging scheme by a domain expert in travel and tour domain. BIO tagging scheme is a common tagging format which is used for tagging tokens in chunking tasks such as Named Entity Recognition in computational linguistics. However due to its limitations in nesting and even representing simple sentence boundaries and grammatical structures this tagging scheme is extended and a modified version has been used in this study as the data annotation strategy which is explained under section 4.1.1.

4.1.1 Annotation Strategy

Review sentences have been manually annotated using the modified BIO tagging scheme as follows.

- we enjoyed the macy balloons 3,5 MISCELLANEOUS 2 1,2
- we thoroughly enjoyed it -1,-1 EXPERIENCE 2 2,3

When annotating the first pair will denote the orientation of the aspect terms or phrase, followed by the aspect category and then the sentiment polarity (positive will be represented by 2, neutral by 1 and negative by 0) and finally the opinion orientation within a sentence. If there are implicit aspects or implicit opinions such pairs will be denoted by -1,-1.

This annotated dataset is then divided into a training set, a validation set and a testing set in the following percentages. Training data = 0.72, Validation data = 0.08, Testing data = 0.20

- No of Reviews considered : 200
- No of Review Sentences: 1398
- No of Subjective Sentences: 782
- No of Objective Sentences: 616

Both models will be evaluated using a validation and Test dataset by Precision, Recall and F1 - Measure.

4.2 Experiments Conducted

Experiments were carried out using the dataset containing 200 customer reviews about the Universal Studio in Orlando taken from year 2019. Following 2 experiments were conducted in this study.

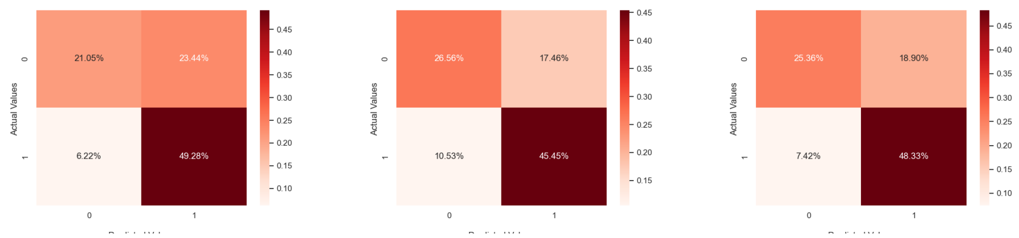
1. Subjectivity Classification
2. Aspect Category Opinion and Sentiment Detection

4.2.1 Sentence Subjectivity Classification

In order to classify subjective sentences in review texts experiments have being conducted using two machine learning models Naive Bayes and Support Vector Machine based on the good results it has achieved in (Wiebe, Bruce, and O'Hara 1999) work.

4.2.1.1 Multinomial Naive Bayes Classifier Model

In previous work (ibid.) MNBC model is trained with the binary feature of adjective, adverb, pronoun, modal, number, punctuation, and the position of the sentence.



(a) trained with TF-IDF (b) trained with BOW (c) trained with ngrams

Fig. 4.1 Confusion Matrix Visualization of MNBC Results

Therefore in this study in order to train the MNBC model we experiment by using Term-Frequency Inverse Document Frequency(TF-IDF), Bag of Words(BOW) features and ngrams and results obtained are presented in table 4.1

	Accuracy	Precision	Recall	F1-Score
TF-IDF	0.7	0.68	0.89	0.77
BOW	0.72	0.72	0.81	0.76
Ngrams(Bigrams,Trigrams)	0.74	0.72	0.87	0.79

Table 4.1 Results of MNBC for sentence subjectivity classification

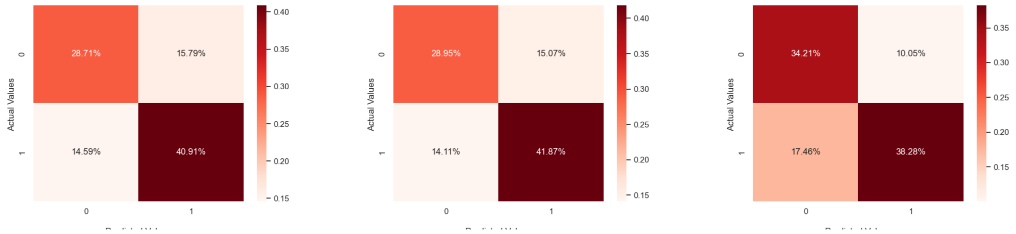
4.2.1.2 Support Vector Machine Model

When training the SVM model we use the same features used for NBC training Term-Frequency Inverse Document Frequency(TF-IDF), Bag of Words(BOW) features and ngrams and compare with MNB results. Results obtained by the SVM classifier have being presented in table 4.2

	Accuracy	Precision	Recall	F1-Score
TF-IDF	0.7	0.72	0.74	0.73
BOW	0.71	0.74	0.75	0.74
Ngrams(Bigrams,Trigrams)	0.72	0.79	0.69	0.74

Table 4.2 Results of MNBC for sentence subjectivity classifcation

According to 4.1 and 4.2 in all three instances MNB has obtained better results than SVM classifier. The highest F1-score of 0.79 has been obtained by Multinomial Classifier when trained with ngrams which include bigrams and trigrams. Therefore we will use the Multinomial Naive Bayes classifier with ngrams in the pipeline for the subjectivity classification task.



(a) trained with TF-IDF (b) trained with BOW (c) trained with ngrams

Fig. 4.2 Confusion Matrix Visualization of SVM Results

4.2.2 Evaluating results of ACOS Detection

4.2.2.1 Implementation and Experimental Setup

In order to identify aspect, aspect category, opinion and related sentiment ACOS model (Cai, Xia, and Yu 2021) has been adapted and for the implementation of the ACOS model, pyTorch an open source machine learning framework has being used. There are two sub experiments conducted in ACOS detection model as Aspect-Opinion co-extractions which aims at detecting the matching aspect and opinion pairs in a given review sentence and second step which detects the Aspect category from the aspect and sentiment from the opinion.

Experimental Setup Configuration for the step 1 in Machine Learning Model

- Train batch size = 24
- Learning rate = 2e - 5
- Number of training epochs = 30

Experimental Setup Configuration for the step 2 in Machine Learning Model

- Train batch size = 16
- Learning rate = 5e - 5
- Number of training epochs = 30

When obtaining the Final score, a quadruple(aspect, aspect category, opinion and related sentiment) is considered as correct if and only if the four elements as well as their combination are exactly the same as those in the gold quadruple. On this basis, the Precision, Recall, and F1 score is calculated.

	Precision	Recall	F1-Score
Aspect-Opinion Co-extraction	0.61	0.63	0.62
Final Score	0.23	0.26	0.24

Table 4.3 Results of Aspect-Category-Opinion-Sentiment Detection

It is evident that the final F1 score, Recall and Precision of Extract-Classify-ACOS on the dataset is very low. One reason being that the number of reviews considered in the dataset is considerably low. On the other hand it is fair since the evaluation metric is based on exact matching of all 4 elements and the ACOS Quadruple Extraction is more complex than the conventional ABSA task. When comparing the performance of the baseline models which provide support for implicit aspects and implicit opinions the results obtained for a laptop dataset (Cai, Xia, and Yu 2021) can be presented as follows:

	Precision	Recall	F1-Score
Double Propogation ACOS	0.13	0.005	0.08
JET-ACOS	0.45	0.16	0.24
Extract Classify ACOS	0.45	0.29	0.36

Table 4.4 Results of three Baseline models for ACOS Quadruple Extraction task

Based on table 4.4 it is visible that the performance of all 3 models are comparatively low. Further when comparing the F1-scores of Double Propogation ACOS and JET ACOS they are much lower than the F1-score of the Extract Classify ACOS model for the laptop dataset. Note that this dataset consists of 4076 sentences and 5758 quadruples. Therefore when evaluating the performance obtained by Extract Classify ACOS model for our travel domain dataset which consists of only 782 sentences and 947 quadruples, it is evident that the F1-score of the proposed model is adequate and there is potential for improvement. Therefore this system can be further improved by using a larger dataset.

4.3 Results of Aspect Category Rating Calculation

Finally we calculate and obtain the ratings based on the aspect categories from the valid quadruples classified based on the calculation method proposed in previous work (Ganu, Elhadad, and Marian 2009). The ratings obtained for each aspect category has been presented in table 4.5. Based on the results obtained it is evident that the Experience, Service, Attractions and Rides have received a customer rating above 4. Miscellaneous and Food aspect categories have received a rating above 3 and for Price a rating below 2 which is reasonable.

Aspect Category	Final Rating
MISCELLANEOUS	3.77
EXPERIENCE	4.22
PRICE	1.27
SERVICE	4.17
FOOD	3.29
ATTRACTIONS	4.69
RIDES	4.20

Table 4.5 Final results of Aspect Category Rating Prediction

There is no mechanism to evaluate the aspect based ratings obtained, since currently there is no aspect based ratings given in online platforms. However based on the comments received by travellers who actually visited the park, also with the experience gained while annotating the review texts it was identified that Universal Studio rides, attractions, the general Experience and service are in a good state, while the ticket prices are quite expensive which is also demonstrated through our system. Therefore, it can be concluded that the aspect-based ratings obtained are reasonable.

Chapter 5

Conclusion and Future Work

The focus of this study, is to systematically address and calculate Aspect Based rating for the different aspects identified in customer reviews in also considering the implicit aspect/opinions in customer reviews. The Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction taskCai, Xia, and Yu 2021 is adapted to a manually annotated new dataset with 200 customer reviews for this task. The ACOS annotations include implicit aspects and implicit opinions and finally calculate a rating based on the different aspects identified.

The proposed system is relatively simple and there is much potential for further improvements. This system can be further improved by using a larger dataset and by using a BERT model trained for this task from scratch. Therefore in order to increase the annotated data size data augmentation techniques could be applied. Even if augmentation of text data in NLP is infrequent there are a number of approaches that are currently being practised.

Bibliography

Albornoz, Jorge Carrillo de et al. (2011). "A joint model of feature mining and sentiment analysis for product review rating". In: *European conference on information retrieval*. Springer, pp. 55–66.

Bagheri, Ayoub, Mohamad Saraee, and Franciska De Jong (2014). "ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences". In: *Journal of Information Science* 40.5, pp. 621–636.

Baziotis, Christos, Nikos Pelekis, and Christos Doulkeridis (2017). "Datastories at semeval-2017 task 6: Siamese LSTM with attention for humorous text comparison". In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 390–395.

Blair-Goldensohn, Sasha et al. (2008). "Building a sentiment summarizer for local service reviews". In.

Cai, Hongjie, Rui Xia, and Jianfei Yu (2021). "Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 340–350.

Chen, Peng et al. (2016). "Clause sentiment identification based on convolutional neural network with context embedding". In: *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, pp. 1532–1538.

Chen, Zhiyuan, Arjun Mukherjee, and Bing Liu (2014). "Aspect extraction with automated prior knowledge learning". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 347–358.

Choi, Yejin and Claire Cardie (2010). "Hierarchical sequential learning for extracting opinions and their attributes". In: *Proceedings of the ACL 2010 conference short papers*, pp. 269–274.

Dhanush, D, Abhinav Kumar Thakur, and Narasimha Prasad Diwakar (2016). "Aspect-based sentiment summarization with deep neural networks". In: *International Journal of Engineering Research and Technology* 5.

Du, Hui et al. (2016). "Aspect-specific sentimental word embedding for sentiment analysis of online reviews". In: *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 29–30.

Du, Weifu and Songbo Tan (2009). "An iterative reinforcement approach for fine-grained opinion mining". In: *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 486–493.

Ganu, Gayatree, Noemie Elhadad, and Amélie Marian (2009). “Beyond the stars: improving rating predictions using review text content.” In: *WebDB*. Vol. 9. Citeseer, pp. 1–6.
Hu, Minqing and Bing Liu (2004). “Mining and summarizing customer reviews”. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177.
Jakob, Niklas and Iryna Gurevych (2010). “Using anaphora resolution to improve opinion target identification in movie reviews”. In: *Proceedings of the ACL 2010 Conference Short Papers*, pp. 263–268.
Jin, Wei, Hung Hay Ho, and Rohini K Srihari (2009). “A novel lexicalized HMM-based learning framework for web opinion mining”. In: *Proceedings of the 26th annual international conference on machine learning*. Vol. 10. 1553374.1553435. Citeseer.
Kiritchenko, Svetlana et al. (2014). “Nrc-canada-2014: Detecting aspects and sentiment in customer reviews”. In: *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pp. 437–442.
Lal, Mily and Kavita Asnani (2014). “Aspect extraction and segmentation in opinion mining”. In: *International Journal Of Engineering And Computer Science* 3.05.
Liu, Bing (2012). “Sentiment analysis and opinion mining”. In: *Synthesis lectures on human language technologies* 5.1, pp. 1–167.
Liu, Bing, Minqing Hu, and Junsheng Cheng (2005). “Opinion observer: analyzing and comparing opinions on the web”. In: *Proceedings of the 14th international conference on World Wide Web*, pp. 342–351.
Liu, Pengfei, Shafiq Joty, and Helen Meng (2015). “Fine-grained opinion mining with recurrent neural networks and word embeddings”. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1433–1443.
Lu, Yue, ChengXiang Zhai, and Neel Sundaresan (2009). “Rated aspect summarization of short comments”. In: *Proceedings of the 18th international conference on World wide web*, pp. 131–140.
Madhoushi, Zohreh, Abdul Razak Hamdan, and Suhaila Zainudin (2019). “Aspect-based sentiment analysis methods in recent years”. In: *Asia-Pacific Journal Of Information Technology And Multimedia* 7.2, pp. 79–96.
Madhoushi, Zohreh, Abdul Razak, and Suhaila Zainudin (June 2019). “Aspect-Based Sentiment Analysis Methods in Recent Years”. In: *Asia-Pacific Journal of Information Technology Multimedia* 08. DOI: 10.17576/apjitm-2019-0801-07.
Marrese-Taylor, Edison, Juan D Velásquez, and Felipe Bravo-Marquez (2014). “A novel deterministic approach for aspect-based opinion mining in tourism products reviews”. In: *Expert Systems with Applications* 41.17, pp. 7764–7775.
Moghaddam, Samaneh and Martin Ester (2011). “ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews”. In: *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 665–674.
Popescu, Ana-Maria and Orena Etzioni (2007). “Extracting product features and opinions from reviews”. In: *Natural language processing and text mining*. Springer, pp. 9–28.
Poria, Soujanya et al. (2016). “Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis”. In: *2016 international joint conference on neural networks (IJCNN)*. IEEE, pp. 4465–4473.

Qiu, Guang et al. (2011). "Opinion word expansion and target extraction through double propagation". In: *Computational linguistics* 37.1, pp. 9–27.
Sauper, Christina and Regina Barzilay (2013). "Automatic aggregation by joint modeling of aspects and values". In: *Journal of Artificial Intelligence Research* 46, pp. 89–127.
Valdivia, Ana, M Victoria Luzón, and Francisco Herrera (2017). "Sentiment analysis in tripadvisor". In: *IEEE Intelligent Systems* 32.4, pp. 72–77.
Wan, Hai et al. (2020). "Target-aspect-sentiment joint detection for aspect-based sentiment analysis". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. 05, pp. 9122–9129.
Wang, Wenya et al. (2016). "Recursive neural conditional random fields for aspect-based sentiment analysis". In: *arXiv preprint arXiv:1603.06679*.
— (2017). "Coupled multi-layer attentions for co-extraction of aspect and opinion terms". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1.
Wiebe, Janyce, Rebecca Bruce, and Thomas P O'Hara (1999). "Development and use of a gold-standard data set for subjectivity classifications". In: *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, pp. 246–253.
Xu, Jiacheng et al. (2016). "Cached long short-term memory neural networks for document-level sentiment classification". In: *arXiv preprint arXiv:1610.04989*.
Xu, Lu et al. (2020). "Position-aware tagging for aspect sentiment triplet extraction". In: *arXiv preprint arXiv:2010.02609*.
Xue, Wei et al. (2017). "MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews". In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 151–156.
Ye, Hai et al. (2017). "Dependency-tree based convolutional neural networks for aspect term extraction". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 350–362.
Zhang, Lei and Bing Liu (2014). "Aspect and entity extraction for opinion mining". In: *Data mining and knowledge discovery for big data*. Springer, pp. 1–40.