HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

GRADUATION THESIS

Aspect Based Sentiment Analysis for Tourism

TRẦN QUỐC KHÁNH
khanh.tq204881@sis.hust.edu.vn

Major: Data Science
Specialization: Data Science and Artificial Intelligence

Supervisor: Ph.D. Kiem-Hieu Nguyen ______________

Signature

Department: Data Science

School: School of Information and Communications Technology

HANOI, 06/2022

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my mentor, Mr. Nguyen Kiem Hieu, whose guidance and meticulous supervision have been instrumental in shaping this thesis. His insightful feedback and expertise have not only enriched my understanding of the subject matter but have also inspired me to strive for excellence.

A special note of appreciation goes to my family, especially my mother, whose unwavering support and encouragement have been the cornerstone of my journey. Her boundless love and belief in my abilities have been a constant source of strength throughout this endeavor.

I am also indebted to my friends for their continuous support and understanding during the challenging phases of this project. Their encouragement and camaraderie have provided me with the motivation to persevere.

I am grateful to all those who have contributed, directly or indirectly, to the completion of this thesis. Your support has been invaluable, and I am truly blessed to have such wonderful individuals in my life.

PLEDGE

Student's full Name: Trần Quốc Khánh
Student ID: 20204881
Contact phone number: 0941039169
Email: khanh.tq204881@sis.hust.edu.vn
Class: IT-E10-01
Program: DSAI

I, **Trần Quốc Khánh**, commit that the Graduation Thesis (GT) is my own research work under the guidance of Ph.D. Kiem-Hieu Nguyen. The results presented in the GT are truthful and are my own achievements, not copied from any other works. All references in the GT, including images, tables, data, and quotations, are clearly and fully cited in the bibliography. I take full responsibility for any violations of the school's regulations concerning plagiarism.

Hanoi, 06/2024

Student

ABSTRACT

In the business and service sectors, user feedback is crucial, and this is especially true for the tourism industry. Natural Language Processing (NLP), particularly Aspect-Based Sentiment Analysis (ABSA), has proven to be effective in understanding and analyzing user feedback. However, both in Vietnam and globally, there is a lack of comprehensive and accurate datasets specifically tailored to tourism. In this thesis, I identify key aspects relevant to tourism as the initial step of ABSA. I then conduct sentiment analysis with four polarities: none/neutral, positive, negative, and conflict. Finally, I introduce a novel dataset along with a fine-tuned Roberta-base model trained on this dataset. Experimental results show that the model have a well generalization for a part of tourism.

Student

(Signature and full name)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
ABSA	Aspect-based Sentiment Analysis
ACD	Aspect Category Detection
ACSA	Aspect Category Sentiment Analysis
ACSD	Aspect-Category-Sentiment Detection
amn	Amenities
AOPE	Aspect-Opinion Pair Extraction
ASC	Aspect Sentiment Classification
ASQP	Aspect Sentiment Quad Prediction
ASTE	Aspect Sentiment Triplet Extraction
ATE	Aspect Term Extraction
BERT	Bidirectional Encoder Representations from Transformers
BT	Back Translation
ch	Cultural Heritage
E2E-ABSA	End-to-End Aspect-based Sentiment Analysis
LDA	Latent Dirichlet Allocation
mgt	Management
MLM	Masked Language Modeling
nat	Nature
NLP	Natural Language Processing
OTE	Opinion Term Extraction
ppl	People
RoBERTa	Robustly Optimized BERT Approach
SA	Sentiment Analysis
SB	Sentence Combining

CHAPTER 1. INTRODUCTION

1.1 Problem Statement: ABSA

Entering the 4.0 era, the need to share and collect data has significantly increased. As a result, most current websites feature options for users to share experiences, reviews, and feedback about products and services, including travel services. These reviews greatly influence tourists' decision-making processes. Concurrently, businesses and organizations collect this feedback to make informed decisions. Despite the abundance of feedback available on various platforms, extracting meaningful insights from this data remains a challenge.

To address this issue, there is a need for a system that can automatically analyze and summarize all feedback for customers, allowing them to make more informed choices. Additionally, such a system would enable businesses to make faster and more accurate decisions. Aspect-Based Sentiment Analysis [1] presents a viable solution for this system. ABSA provides a nuanced understanding of user sentiments by identifying specific aspects of services (such as accommodation, food, and transportation) and determining the sentiment expressed towards each aspect.

1.1.1 Definition of ABSA Problem

The ABSA problem can be formally defined as follows:

1. **Input**: A set of textual reviews or feedback from users. Each review may contain multiple sentences discussing various aspects of a service or product.
2. **Output**: For each sentence or segment of the review, the system identifies the relevant aspect categories and assigns a sentiment polarity to each aspect. For example, given a sentence like "The hotel's location was perfect, but the service was terrible," the system should output (Location-Positive, Service-Negative)

1.1.2 Common Approaches to ABSA

There are several prevalent approaches to ABSA, including:

1. **Rule-based Methods**: These rely on predefined rules and lexicons to identify aspects and sentiments. Although straightforward, they often lack adaptability to diverse and evolving language usage.
2. **Machine Learning-based Methods**: These involve training classifiers using labeled datasets. While more adaptable than rule-based methods, they require substantial annotated data and may struggle with unseen terms.
3. **Deep Learning-based Methods**: Leveraging neural networks, particularly

models like BERT (Bidirectional Encoder Representations from Transformers), these methods excel in capturing contextual nuances and have shown state-of-the-art performance in various sentiment analysis tasks.

1.1.3 Approach in This Research

This research adopts a deep learning-based approach, specifically targeting the tourism domain and focusing on user reviews in the English language. Key elements of the methodology include:

1. **Aspect Category Identification**: Emphasis is placed on identifying aspect categories pertinent to tourism services.
2. **Selection of Sentiment Polarities**: Unlike the conventional triad of sentiment polarities (positive, neutral, negative), this project incorporates four sentiment polarities: None/Neutral, Positive, Negative, and Conflict. This expanded polarity set aims to capture a broader range of sentiments expressed in user reviews.
3. **Sentence Embedding with Aspect Information**: Each sentence is embedded with its corresponding aspect category, allowing the model to recognize and process the sentiment polarity in the context of that specific aspect.

By implementing ABSA with these enhancements, the proposed system aims to effectively categorize and analyze user feedback in the tourism industry. This advanced sentiment analysis will enable more nuanced and actionable interpretations of user reviews, thereby enhancing decision-making processes for both consumers and businesses.

1.2 Research problem

Aspect-Based Sentiment Analysis (ABSA) has been widely researched and applied in various fields, with significant advances made in recent years. ABSA focuses on identifying and classifying sentiments expressed towards specific aspects of entities, providing a detailed understanding of user feedback. Several notable studies and datasets have contributed to progress in this field:

First, a standard dataset comprising reviews from the restaurant and laptop domains, along with four sub-tasks, was introduced by (Pontiki et al., 2014)[2]. This dataset has become a benchmark for developing and evaluating ABSA models.

Then, (Jiang et al., 2019)[3] proposed a new dataset called MAMS (Multi-Aspect Multi-Sentiment), which includes sentences with at least two aspects and potentially different sentiment polarities for each aspect. This dataset is designed to handle more complex scenarios and provides a richer source of information for training and testing ABSA models.

Additionally, (Peng et al., 2019)[4] introduced a new subtask in ABSA, called Aspect Sentiment Triplet Extraction (ASTE). This approach extracts triplets (aspect, sentiment, reason) from the inputs, providing more detailed information, improving accuracy, and aiding in the interpretation of results.

Despite these advancements, there are notable limitations in current ABSA research and datasets. First, most existing datasets are tailored to **specific domains** such as restaurants and laptops, limiting their generalizability and effectiveness when applied to the tourism sector, which has its own unique set of aspects and user concerns. Second, advanced models like ASTE require detailed annotations which are labor-intensive and costly to produce. This limits the availability of comprehensive datasets, particularly in less-studied domains.

1.3 Research Objectives

In this thesis, I aim to achieve the following main objectives:

1. Identify aspects and polarities to create a new dataset tailored to the ABSA problem.
2. Train and fine-tune the RoBERTa-base model. Concurrently, apply data augmentation techniques such as fill-mask BERT, back translation, and sentence combining, which have been proven effective in various studies.
3. Evaluate the improvement in the model's performance when using data enrichment techniques. Conduct a comprehensive analysis to evaluate how well the model performs on different sentiment polarities and aspect categories. This will help in understanding the strengths and limitations of the model in various contexts.

1.4 Conceptual Framework

The conceptual framework for this research is designed to systematically address the problem of Aspect-Based Sentiment Analysis in the tourism domain by leveraging advanced deep learning techniques. The framework encompasses the following components:

Dataset Construction Methodology:

1. **Selection of Sentiment Polarities**: Define a comprehensive set of sentiment polarities, including None/Neutral, Positive, Negative, and Conflict, to capture the full spectrum of user sentiments.
2. **Selection of Aspect Categories**: Identify relevant aspect categories specific to

the tourism industry, such as amenities, cultural heritage, management, nature, and people.

Data Augmentation Techniques:

1. **Masked Language Modelling (MLM)**: Employ the MLM technique, where certain tokens in a sentence are masked and the model predicts these masked tokens. This method generates diverse syntactic variations of sentences while maintaining their semantic integrity, enriching the dataset with varied training examples.

2. **Back Translation**: Translate sentences into another language and then back into the original language. This introduces natural variations in sentence structure and wording, thereby enhancing the robustness and diversity of the training data.

3. **Sentence Combining**: Merge multiple sentences to create more complex and informative training examples. This technique helps the model handle varied and intricate inputs, improving its generalization capabilities and overall performance.

Model Training and Fine-tuning:

1. **RoBERTa-base Model**: Utilize the RoBERTa-base model, known for its contextual understanding and efficiency in handling large datasets, as the foundation for sentiment analysis.

2. **Fine-tuning Process**: Adapt the pre-trained RoBERTa-base model to the specific nuances of the newly created dataset, optimizing it for improved performance in ABSA tasks.

Performance Evaluation:

1. **Metrics**: Use key performance metrics, such as macro-recall, macro-precision, and macro-F1 scores, to quantitatively assess the model's effectiveness.

2. **Comparative Analysis**: Conduct a thorough comparison of the model's performance with each data augmentation techniques to evaluate the impact of dataset enrichment.

By integrating these components, the conceptual framework provides a structured approach to enhance the effectiveness of Aspect-Based Sentiment Analysis in the tourism domain. This framework not only addresses the immediate research objectives but also sets the stage for future advancements in sentiment analysis methodologies.

1.5 Contributions

This thesis makes several key contributions to the field of Aspect-Based Sentiment Analysis (ABSA) in the context of the tourism industry:

1. Development of a Specialized Dataset: I have created a novel dataset specifically tailored to the tourism sector, focusing on reviews from two prominent tourist destinations: Hue Imperial City in Vietnam and Changdeokgung Palace in South Korea. This dataset includes meticulously annotated aspects and sentiment polarities, providing a valuable resource for future research in tourism-related sentiment analysis.

2. Model Fine-tuning and Data Augmentation: By training and fine-tuning the RoBERTa-base model on newly developed dataset, I have enhanced its ability to accurately analyze sentiments in tourist reviews. I also applied advanced data augmentation techniques, including synonym replacement, fill-mask BERT, back translation, and sentence combining. These techniques have significantly enriched the dataset and improved the model's performance.

3. Evaluation of Data Enrichment Techniques: I conducted a thorough evaluation of the impact of data enrichment techniques on the model's performance. These findings demonstrate the effectiveness of these techniques in enhancing the accuracy and robustness of sentiment analysis models.

These contributions collectively advance the state of ABSA research, particularly in the underexplored domain of tourism, and provide practical tools and insights for better understanding and leveraging user feedback in this industry.

1.6 Organization of Thesis

The thesis is structured as follows:

Chapter 2 provides a comprehensive Literature Review encompassing related works, foundational concepts of Sentiment Analysis (SA), and Aspect-Based Sentiment Analysis (ABSA). This chapter will offer an overview of previous research conducted in ABSA, elucidating the theoretical background and contextual framework for the study.

Chapter 3 introduces the Methodology, detailing the steps involved in aspect identification for ABSA, rationale behind selecting four sentiment polarities, and methods for data augmentation and balancing. This chapter will delineate the data collection and preprocessing procedures, algorithms and techniques utilized for model development and training, as well as justification for the chosen methodologies.

Chapter 4 presents Numerical Results, encompassing model evaluation metrics, fine-tuning strategies. This chapter will expound upon the performance evaluation metrics employed, empirical findings, and analysis of model improvements following the application of data augmentation techniques.

Chapter 5 concludes by summarizing the addressed issues, lingering challenges, and proposes future directions for research. This final chapter synthesizes the achieved results, highlights key contributions of the thesis, and suggests further research avenues to overcome current limitations and expand the application scope of ABSA within the tourism domain.

CHAPTER 2. LITERATURE REVIEW

2.1 Overview

This chapter provides a comprehensive review of the relevant literature, organized into five main sections, building upon the foundational concepts introduced in Chapter 1. The section 2.2 outlines the focus of the study, particularly on ABSA applied to tourism-related user feedback. The section 2.3 includes an examination of state-of-the-art models such as BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining Approach), which have significantly advanced the field of natural language processing by providing powerful contextual embeddings and enhancing performance in various sentiment analysis tasks. The section 2.4 and 2.5 presents the essential background knowledge, covering key principles and techniques in Sentiment Analysis and Aspect-Based Sentiment Analysis. Finally, the section 2.6 examines existing research on ABSA in the tourism industry, identifying key studies and methodologies.

2.2 Scope of Research

The research focuses on tourists' reviews extracted from feedback on `tripadvisor.com` about two tourist destinations: Hue Imperial City (Vietnam) and Changdeokgung Palace (South Korea).

2.3 BERT and RoBERTa

2.3.1 BERT

In 2018, Google researchers including Jacob Devlin et al. [5] introduced BERT, a pre-training technique for NLP that leverages transformers, a powerful machine learning architecture.

The original BERT model in English is available in two pre-trained versions: the BERTbase and BERTlarge models. The BERTbase model features a neural network architecture with 12 layers, 768 hidden units, 12 attention heads, and a total of 110 million parameters. In contrast, the BERTlarge model comprises a more complex architecture with 24 layers, 1024 hidden units, 16 attention heads, and 340 million parameters. Both models were trained on the BooksCorpus[6], containing 800 million words, and an English Wikipedia dataset with 2.5 billion words.

BERT is based on pre-trained contextual representations, utilizing semi-supervised sequence learning techniques. Unlike its predecessors, BERT is an unsupervised, bidirectional language representation model.

Context-free models such as Word2Vec[7] or GloVe[8] generate a single embedding representation for each word in the vocabulary, regardless of context. In contrast, BERT considers the context for each occurrence of a given word. For instance, the word "bank" would have the same Word2Vec vector representation in the sentences "She sat by the bank of the river" and "He went to the bank to deposit money." However, BERT provides contextual word embeddings that differ depending on the sentence, capturing the different meanings of "bank" in each context.

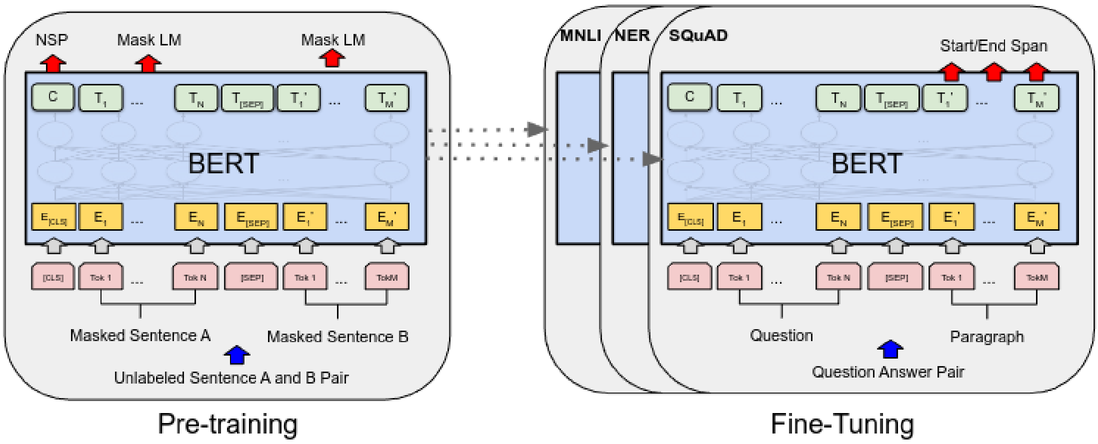

Figure 2.1: The architecture of the BERT model is presented in [5]

2.3.2 RoBERTa

RoBERTa (Robustly optimized BERT approach)[9] is an enhanced version of BERT, trained on a larger dataset with several optimizations. These improvements include training on longer sequences, removing the next sentence prediction objective, and dynamically changing the masking pattern applied to the training data.

2.4 Fundamental Concept of Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a subfield of Natural Language Processing (NLP) that focuses on identifying and extracting subjective information from text. It aims to determine the sentiment expressed in a piece of text, which is typically categorized as positive, negative, or neutral. Sentiment analysis is widely used in various applications, including market research, social media monitoring, customer feedback analysis, and more.

2.4.1 Sentiment Classification

1. Lexicon-Based Methods: Using predefined lists of words (sentiment lexicons) with associated sentiment scores to determine sentiment.
2. Machine Learning-Based Methods: Training algorithms on labeled datasets to classify sentiment. Common algorithms include Naive Bayes, Support Vector

Machines, and Decision Trees.

3. Deep Learning-Based Methods: Utilizing neural networks, such as Convolutional Neural Networks and Recurrent Neural Networks, for more advanced and accurate sentiment analysis.

2.4.2 Levels of Sentiment Analysis

Sentiment analysis can be performed at different levels of granularity:

1. Document-Level Sentiment Analysis: Determines the overall sentiment of an entire document or piece of text. This approach assumes that the text expresses a single sentiment towards the subject matter.

2. Sentence-Level Sentiment Analysis: Analyzes the sentiment of individual sentences. This is useful when different sentences in a document express different sentiments.

3. Aspect-Level Sentiment Analysis (ABSA): Identifies sentiments towards specific aspects or features of the subject. For example, in a restaurant review, aspect-level analysis might separately assess sentiments towards food, service, ambiance, etc.

2.4.3 Challenges in Sentiment Analysis

Despite its widespread application, sentiment analysis faces several challenges:

1. Sarcasm and Irony: Detecting sarcasm and irony, where the sentiment expressed is the opposite of the literal meaning, is difficult for automated systems.

2. Context and Ambiguity: Understanding the context in which a word or phrase is used is crucial. Words can have different sentiments depending on the context (e.g., "mad" can mean angry or enthusiastic).

3. Domain-Specific Sentiments: Words may have different sentiment orientations in different domains. For instance, "unpredictable" may be negative for a car review but positive for a movie review.

4. Multilingual Sentiment Analysis: Analyzing sentiment across different languages requires robust models that can handle linguistic nuances and cultural differences.

2.5 Fundamental Concept of Aspect-based Sentiment Analysis

This section provides an overview of the fundamental concepts of Aspect-Based Sentiment Analysis (ABSA), primarily drawing on the paper "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges" by (Zhang et al., 2022)[10]. ABSA delves into the nuanced understanding of sentiments expressed towards specific aspects or features within textual data.

2.5.1 Four Sentiment Elements of ABSA

ABSA involves the interaction of four critical sentiment elements:

1. Aspect Category: Broad domains or categories under which specific aspects fall. For example, in a restaurant review, aspect categories might include "food," "service," or "ambiance."
2. Aspect Term: Specific words or phrases in the text that indicate the precise aspect being discussed. For instance, in the sentence "The pizza was delicious," the aspect term is "pizza."
3. Opinion Term: Words or phrases expressing sentiment towards an aspect term. In the earlier example, "delicious" is the opinion term associated with the aspect term "pizza."
4. Sentiment Polarity: The sentiment expressed towards an aspect term, typically categorized as positive, negative, or neutral.

Additionally, an ontology tree is often used to organize aspect categories and terms hierarchically. This helps in understanding the relationships and dependencies between different aspects and their subcategories, improving the accuracy and depth of ABSA.

2.5.2 Single ABSA Subtasks

ABSA can be broken down into several single subtasks, each focusing on a specific component of the analysis:

1. Aspect Term Extraction (ATE): The process of identifying and extracting aspect terms from the text. This task aims to pinpoint specific features or components being discussed.
2. Aspect Category Detection (ACD): Classifying text segments into predefined aspect categories. This involves mapping textual mentions to broader categories, such as grouping "pizza" under the category "food."
3. Opinion Term Extraction (OTE): Identifying words or phrases that express opinions or sentiments. This task focuses on extracting terms that convey the writer's sentiment.
4. Aspect Sentiment Classification (ASC): Determining the sentiment polarity (positive, negative, or neutral) associated with each aspect term or category. This subtask assigns a sentiment value to the identified aspects.

2.5.3 Compound ABSA Tasks

Compound ABSA tasks target multiple sentiment elements to provide a more comprehensive analysis:

1. Aspect-Opinion Pair Extraction (AOPE): Extracting pairs of aspect terms and their corresponding opinion terms. This task identifies which opinion is associated with which aspect.
2. End-to-End ABSA (E2E-ABSA): Integrating all single ABSA subtasks into a unified process. This involves extracting aspects, opinion terms, and determining sentiment polarity in one go.
3. Aspect Category Sentiment Analysis (ACSA): Analyzing sentiment at the category level rather than individual terms. This task evaluates the overall sentiment towards broader aspect categories.
4. Aspect Sentiment Triplet Extraction (ASTE): Identifying triplets consisting of an aspect term, its corresponding opinion term, and the sentiment polarity. This task provides a detailed sentiment analysis for specific aspects.
5. Aspect-Category-Sentiment Detection (ACSD): Detecting aspect categories and their associated sentiments within a text, often focusing on the relationship between categories and sentiments.
6. Aspect Sentiment Quad Prediction (ASQP): Predicting quads that include an aspect term, opinion term, aspect category, and sentiment polarity. This comprehensive task combines all elements for a detailed analysis.

2.5.4 Challenges

ABSA faces several challenges that impact its effectiveness and accuracy:

1. Data Scarcity: The limited availability of labeled datasets for training and testing ABSA models poses a significant challenge.
2. Domain Adaptability: ABSA models often struggle to generalize across different domains, requiring extensive retraining for each new domain.
3. Aspect-Sentiment Coherence: Ensuring that identified aspects and their corresponding sentiments are coherent and accurately matched.
4. Complex Sentences: Handling sentences with multiple aspects and mixed sentiments, which require sophisticated models to disentangle.
5. Multilingual ABSA: Extending ABSA methods to multiple languages adds complexity due to linguistic diversity and the need for multilingual resources

and models.

Understanding these fundamental concepts, tasks, and challenges is crucial for advancing research and development in ABSA, aiming to create more accurate and adaptable sentiment analysis systems.

2.6 ABSA in the tourism domain

(Arianto et al., 2020)[11] proposed a dataset comprising mixed Indonesian and English reviews from Google Maps, focusing on two temples in Indonesia. The authors identified relevant aspects for the problem, namely Attractions, Amenities, Accessibility, Image, Price, and Human Resources, as recommended by the World Tourism Organization. They employed machine learning methods to classify sentiment polarities into Positive, Negative, Neutral, and None. Extensive text preprocessing steps were applied to ensure model performance with mixed-language reviews. However, the dataset exhibited errors in the Image aspect, which led to reduced model performance. Additionally, the study primarily evaluated traditional machine learning models, suggesting the potential for further research utilizing more advanced deep learning models.

(Af'idah et al., 2023)[12] proposed a two-step approach for Aspect-Based Sentiment Analysis (ABSA). The first step involves using four binary classification models to determine the presence or absence of aspects such as attractiveness, accessibility, facilities, and accommodation. In the second step, detected aspects are analyzed using a model to identify sentiment polarities, specifically positive and negative. Both steps utilize Long Short-Term Memory or Bidirectional LSTM models. Although this method achieves good results, it requires up to eight models for analysis, which can result in the loss of relationships between aspects, increased system complexity, high computational resource consumption, and difficulties in optimization. Furthermore, this approach lacks flexibility when applied to different datasets.

In addition to identifying aspects, (Ali et al., 2022)[13] employed the Latent Dirichlet Allocation (LDA) algorithm to automatically discover topics (aspects) such as atmosphere, shopping experience, citizen behavior, and overall touristic experience. However, a limitation of this approach is the inherent assumption of LDA that reviews are a mixture of topics, which is not entirely accurate for tourism reviews, as some reviews focus exclusively on a single topic. Moreover, aspect-terms across different topics can be interrelated; for instance, terms like "buy" and "price" in the Shopping Experience topic may relate to "pay" and "money" in the Citizen Behavior topic.

This chapter has provided a comprehensive review of the literature relevant to Aspect-Based Sentiment Analysis (ABSA), particularly in the context of tourism-related user feedback. It covered essential principles and techniques of ABSA, examined state-of-the-art models like BERT and RoBERTa, and discussed their contributions to enhancing sentiment analysis. This review sets the stage for Chapter 3, which will detail the methodology, including the construction of the dataset and the application of data augmentation techniques, to develop a robust ABSA system for the tourism domain.

CHAPTER 3. METHODOLOGY

3.1 Overview

Building upon the theoretical foundations established in Chapter 2, this chapter outlines the practical methodology employed to tackle the challenges of aspect-based sentiment analysis (ABSA). The methodology encompasses several crucial steps, including the construction of a tailored dataset designed specifically for the tourism domain. It further details the implementation of the RoBERTa model. Additionally, the chapter explores various strategies for data augmentation and balancing, aiming to enhance the robustness and effectiveness of the ABSA system in capturing nuanced sentiment insights from user feedback.

3.2 Dataset Construction Methodology

3.2.1 Data Source

Tripadvisor is the primary data source for this study. As a widely-used website, Tripadvisor offers extensive and high-quality reviews from travelers, providing comprehensive evaluations of various travel-related services. In tourism research, many studies have utilized Tripadvisor data to analyze trends, customer satisfaction, and factors influencing travel experiences. For this study, data were collected from two prominent tourist locations: Hue Imperial City (Vietnam) and Changdeokgung Palace (South Korea).

Rationale for Selection

1. **UNESCO World Heritage Sites:** The designation of both Hue Imperial City and Changdeokgung Palace as UNESCO World Heritage sites underscores their global cultural and historical significance, making them exemplary subjects for tourism research.

2. **Comprehensive Tourism Characteristics:** These locations exhibit a wide range of tourism attributes, including historical significance, architectural beauty, and cultural heritage. This diversity ensures that the collected data encapsulates various aspects of the tourism industry.

3. **Generalizability of Data:** By selecting two geographically and culturally distinct sites, the dataset becomes more generalized, enhancing its applicability across different contexts within the tourism sector.

The inclusion of data from Hue Imperial City and Changdeokgung Palace contributes to a richer, more comprehensive dataset, supporting a robust aspect-based sentiment analysis (ABSA) framework. This approach ensures that the research findings are not only specific to individual sites but also extendable to a broader range of tourism destinations.

3.2.2 Sentence-Level Analysis

Rationale for Choosing Sentence-Level Analysis

1. **Compatibility with Machine Learning Methods:** Sentence-level processing aligns well with various machine learning techniques, facilitating more efficient and effective model training.
2. **Simplicity and Efficiency:** Sentences are shorter and simpler compared to paragraphs. This reduces the labeling effort and increases accuracy, as annotators can focus on more manageable chunks of text.
3. **Detailed Analysis:** Processing at the sentence level allows for more granular sentiment analysis, providing a clearer understanding of specific opinions and sentiments expressed in individual sentences.
4. **Reduced Contextual Complexity:** Sentences, being more concise, often contain more focused and explicit expressions of sentiment, making it easier to identify and classify sentiment accurately.

While sentence-level processing offers several advantages, it also has some limitations:

1. **Loss of Context:** Sentences often rely on surrounding sentences to convey complete meaning. Isolating sentences may lead to a loss of contextual information, which can affect the accuracy of ABSA.
2. **Fragmented Opinions:** Opinions that span multiple sentences may be fragmented, leading to incomplete sentiment extraction.

Despite these limitations, sentence-level processing is chosen for its overall benefits in terms of simplicity, efficiency, and compatibility with machine learning methods. Addressing the challenges associated with context loss can be a focus for future enhancements in the model.

3.2.3 Labeling Approach

During the annotation process, I exclusively label the Aspect Category without specifying the location of the Aspect Term.

1. **Generalization and Sufficiency:** Although aspect terms provide detailed insights, aspect categories offer a more generalized view that is often sufficient for analysis purposes. This abstraction helps in capturing broader trends and patterns without the need for excessive detail.

2. **Reduced Labeling Effort:** Labeling aspect categories requires significantly less effort compared to aspect terms. This reduction in labeling complexity makes the annotation process more efficient and less time-consuming.
3. **Handling Missing Aspect Terms:** In some instances, sentences may not explicitly mention aspect terms but still imply an aspect category. For example, the sentence “Free for children under 7” clearly indicates the aspect category "service" even though no specific aspect term is mentioned.

Similarly, I label only the Sentiment Polarity instead of the Opinion Term.

1. **Simplification:** Labeling sentiment polarity is simpler and more straightforward than labeling opinion terms. This simplicity reduces the cognitive load on annotators and enhances the consistency and accuracy of the labels.
2. **Focused Analysis:** Sentiment polarity labels provide direct insights into the sentiment expressed, making it easier to analyze and interpret the data. This directness is particularly useful in applications where understanding the sentiment towards a category is more critical than identifying the specific opinion terms.

To summarize, the sentence "Free for children under 7" will be annotated as a pair (Aspect Category-Sentiment Polarity): service-positive

By focusing on aspect categories and sentiment polarity, I streamline the annotation process while still capturing essential information for effective aspect-based sentiment analysis. This approach balances the need for detailed insights with practical considerations of labeling efficiency.

3.2.4 Selection of Sentiment Polarities

In sentiment analysis (SA) and aspect-based sentiment analysis (ABSA), the primary polarities considered are **positive** and **negative**.

Annotations in this study include **"none"** for an aspect category, indicating that the aspect category is not referenced within the text. Similarly, **"neutral"** for an aspect category when it is mentioned in the text but without clear indication of being positive or negative.

This study employs four sentiment polarities: none/neutral, positive, negative, and conflict. "none" and "neutral" are combined into a single polarity due to their similar characteristics based on the polarity combining rule (outlined below), while **"conflict"** results from positive and negative according to the polarity aggregation rules. For brevity, I denote none/neutral, positive, negative, and conflict as 0, 1, 2, and 3, respectively.

Polarity Combining Rule: Two annotated sentences can be combined to form a new sentence with the polarity label defined as follows (sentiment polarities refer to the same aspect category, where the The left side represents the emotional polarities in the original two sentences and the right side represents the emotional polarity obtained in the new sentence, (+) represents combination and the commutative property allows the exchange of emotional polarities feelings of sentences 1 and 2 without changing the result):

Table 3.1: Polarity Combing Rule

	0	1	2	3
0	$0+0 \to 0$	$0+1 \to 1$	$0+2 \to 2$	$0+3 \to 3$
1	$1+0 \to 1$	$1+1 \to 1$	$1+2 \to 3$	$1+3 \to 3$
2	$2+0 \to 2$	$2+1 \to 3$	$2+2 \to 2$	$2+3 \to 3$
3	$3+0 \to 3$	$3+1 \to 3$	$3+2 \to 3$	$3+3 \to 3$

3.2.5 Selection of Aspect Categories

Selecting appropriate aspect categories is a crucial step that must align with the specific characteristics of Hue Imperial City and Changdeokgung Palace while maintaining general applicability to the tourism domain. Two approaches were considered:

Approach 1: Topic Modeling

Topic modeling techniques such as Latent Dirichlet Allocation (LDA) were utilized to identify relevant aspect categories. However because of its mentioned limitations I use the Approach 2

Approach 2: Synthesis from Previous Studies

(Af'idah et al., 2023)[12] identified: Attractions, Accessibility, Facility, Accommodation.

(Arianto and Budi, 2020)[11] utilized: Attractions, Amenities, Accessibility, Image, Price, Human Resources.

(Nguyen Thi Le Huong and Phan Thanh Hoan, 2017)[14] identified aspects such as: Natural resources, Cultural resources (tangible, intangible), Accommodation facilities, Dining facilities, Shopping centers, Entertainment venues, Information service centers, Infrastructure, Accessibility, Scenery, Culture-cuisine, Environment, Safety levels, Facilities, Local community friendliness, Human resources.

Additionally, referencing 4640/QD-BVHTTDL[15], a comprehensive 6 categories was considered:

1. Tourism resources: Diversity and uniqueness of resources; Capacity of resource points; Resource protection and embellishment.
2. Products and services: Providing information to customers; Information guidance throughout the tourist area; Interpretation; Tourist information centers; Technical infrastructure serving tourist accommodation; Services provided to guests in accommodation areas; Restaurant system serving tourists; Dining services; Entertainment and recreation facilities; Entertainment services; Performance activities, artistic performances; Tourist services, resort, exploration, understanding natural and cultural values; Event organization services, conferences, seminars; Shopping services.
3. Destination management: General management; Environmental hygiene and general sanitation; Waste treatment; Public restroom system; Social environment; Security forces organization, order; Plans ensuring security, safety for tourists; Technical infrastructure.
4. Infrastructure: Transportation system; Signage instructions for accessing tourist areas by road, waterway; Internal transportation system; Electricity system; Water supply and drainage system; Mobility capabilities.
5. Participation of local community: Proportion of local labor in tourist areas.
6. Visitor satisfaction: Tourist satisfaction through surveys.

Based on these references, I derived a preliminary set of aspect categories:

1. **Scenery, natural environment**
2. **Weather**
3. **Tangible cultural heritage**
4. **Intangible cultural heritage**
5. **History**
6. **Resource protection and embellishment**
7. **Destination management:** General management, regulations, laws; Environmental hygiene, waste management; Safety levels, organization of security forces, order, plans ensuring security, safety for tourists;
8. **Infrastructure:** Physical facilities, public restroom system, transportation system, signage instructions for accessing tourist areas by road, waterway, internal transportation system, electricity system, water supply and drainage system, mobility capabilities.

9. **Amenities:** Accommodation services, services provided to guests in accommodation areas; Dining facilities, dining services; Shopping centers, shopping services; Tourist services, exploration, understanding natural and cultural values, tourist information services, information guidance throughout the tourist area, interpretation; Entertainment venues, entertainment services; Performance activities, artistic performances; Event organization services, conferences, seminars.
10. **Pricing**
11. **People:** Friendliness, expertise, image of tourist workers, locals, tourists.

These categories provide a foundational framework for capturing diverse aspects pertinent to tourism evaluation and aspect-based sentiment analysis in the context of Hue Imperial City and Changdeokgung Palace.

3.3 Dataset Construction Process

Labeling the dataset construction process is a challenging task that demands considerable time and effort. To streamline this process and enhance the accuracy of new dataset, I have employed several methodologies. These methods are designed to optimize efficiency while ensuring the dataset captures the nuances necessary for robust analysis in aspect-based sentiment analysis (ABSA).

3.3.1 Handling Aspect Categories

During analysis, We observed several similar categories, such as tangible and intangible cultural resources. Additionally, certain categories, like amenities and pricing, frequently co-occur. These similar or co-occurring categories can be merged into broader categories during the annotation process.

Merging aspect categories offers several benefits: it reduces complexity, increases accuracy (by minimizing instances of missed aspect terms or difficulty in classifying aspect terms into specific categories), alleviates data imbalance among aspects, and simplifies the training model, thereby enhancing its accuracy.

During the annotation process, I combined these categories, utilizing polarity aggregation rules to facilitate the process easily:

1. **Nature**: Combines 'Scenery, Natural Environment' and 'Weather'.
2. **Cultural Heritage**: Combines 'Tangible Cultural Resources', 'Intangible Cultural Resources', and 'History'.
3. **Management**: Combines 'Resource Protection and Embellishment', 'Destination Management', and 'Infrastructure'.

4. **Amenities**: Combines ’Amenities’ and ’Pricing’.

To address exceptional cases (where there is an opinion term without an aspect term, or the aspect term does not fit into the defined categories), I propose two special aspect categories: “Other” and “All-Inclusive”.

1. **Other Aspects**: This category includes all aspect terms that do not belong to any of the predefined categories. The polarity of "Other Aspects" is calculated using the polarity aggregation rules with all corresponding polarities of its aspect terms.
2. **All-Inclusive Aspects**: This category includes all aspect terms (similar to a general sentiment analysis problem, but the polarity assignment method differs). Its polarity is also calculated using the polarity aggregation rules across all aspect categories, including "Other Aspects".

For simplicity, the polarity of "Other Aspects" serves as an intermediate step to calculate the polarity for "All-Inclusive" then will not be annotated. Ultimately, We have five aspect categories: Nature (nat), Cultural Heritage (ch), Management (mgt), Amenities (amn), People (ppl), and two special categories: Other (other) and All-Inclusive (all).

By consolidating these categories, I ensure a more streamlined and accurate annotation process, which is crucial for effective aspect-based sentiment analysis.

3.3.2 An example of Annotation

In this section I go over how a sentence is annotated in dataset. **Note:**

1. In this context, the operator + is understood as combining.
2. An opinion term can be associated with multiple aspect terms or aspect categories.
3. An aspect category can be determined using three pieces of information: the opinion term, the aspect term, and the context of the sentence. However, all three are not always necessary. For example, in the example below, the opinion term "dirty" alone is sufficient to identify the aspect category "management" (mgt) without needing "Hue" to provide additional information. Conversely, in some cases, even with both the opinion term and the aspect term present, we might still be unable to determine the relevant aspect category due to a lack of contextual information in the sentence.

Consider sentence: “Hue is the ugliest city I’ve been to, it is dirty and the people here are so rude and pushy and annoying they will force you to sell their goods.”

Step 1: Initialization

$$\text{nat\_polarity} = 0$$
$$\text{ch\_polarity} = 0$$
$$\text{mgt\_polarity} = 0$$
$$\text{amn\_polarity} = 0$$
$$\text{ppl\_polarity} = 0$$
$$\text{other\_polarity} = 0$$

Step 2: For each opinion in sentence

Opinion “ugliest” with Term “Hue”

$\Rightarrow$ Category “other”

$\Rightarrow$ Polarity “2”

$\text{other\_polarity} = \text{other\_polarity} + 2$

Opinion “dirty” with Term “Hue”

$\Rightarrow$ Category “mgt”

$\Rightarrow$ Polarity “2”

$\text{mgt\_polarity} = \text{mgt\_polarity} + 2$

Opinion “rude and pushy and annoying” with Term “people”

$\Rightarrow$ Category “ppl”

$\Rightarrow$ Polarity “2”

$\text{ppl\_polarity} = \text{ppl\_polarity} + 2$

Opinion “force you to sell their goods” with Term “people”

$\Rightarrow$ Category “ppl”

$\Rightarrow$ Polarity “2”

$\text{ppl\_polarity} = \text{ppl\_polarity} + 2$

Step 3: Get Sentiment Polarity of All-Inclusive

$$\begin{aligned}\text{all\_polarity} &= \text{nat\_polarity} + \text{ch\_polarity} + \text{mgt\_polarity} \\ &\quad + \text{amn\_polarity} + \text{ppl\_polarity} + \text{other\_polarity} \\ &= 0 + 0 + 2 + 0 + 2 + 2 \rightarrow 2\end{aligned}$$

Final Label is “all-2 nat-0 ch-0 mgt-2 amn-0 ppl-2” or short “all-2 mgt-2 ppl-2”

3.3.3 Handling human error

With the aforementioned method for labeling sentences, We can apply labeling technique to complex sentences, considering both semantic and syntactic elements. This process is prone to high human error, largely dependent on the annotator’s experience. Additionally, during the dataset construction, there are the following challenges:

1. **Lack of consensus**: The assigned labels heavily depend on the subjective viewpoint of the annotator.
2. **Ambiguous boundaries**: In some instances, the distinction between sentiment polarities is unclear.

To address these challenges:

1. Sentences with unresolved annotation disagreements are discarded.
2. Train the model concurrently with the annotation process. Even though the model is not fully trained, it serves as an effective tool for identifying and correcting mislabeled sentences.

Specifically, the figure 3.1 below describes the dataset construction process.

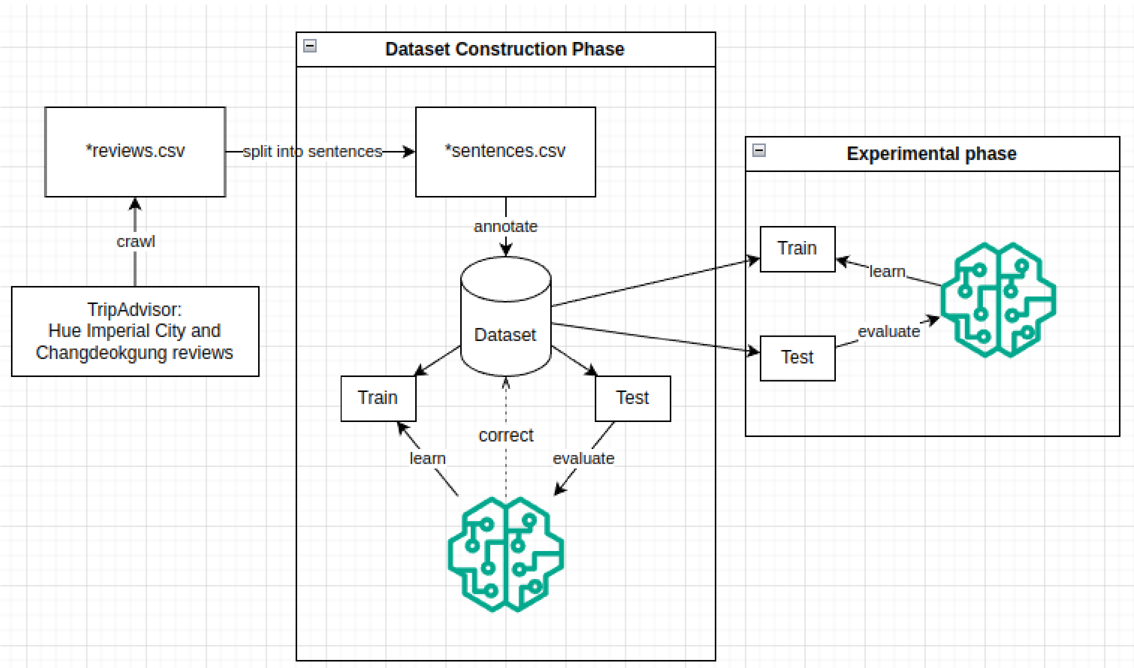

Figure 3.1: Dataset Construction Process

Through the above steps, I have created the following dataset: **Hue Imperial City** - 2621 instances; **Changdeokgung** - 995 instances.

3.4 Model Fine-Tuning

In this study, I fine-tuned the cardiffnlp/twitter-roberta-base-sentiment-latest model, which was pre-trained on the TweetEval dataset specifically for sentiment analysis tasks. This pre-trained model provides a robust starting point, leveraging extensive sentiment data from Twitter.

To adapt the model to new dataset, I modified the number of output classes from 3 to 4 to accommodate specific sentiment polarities. Additionally, to ensure the model understands which aspect category's sentiment it needs to classify, I appended the name of the aspect category to the beginning of each text instance. This method offers several advantages over modifying the model's head for Aspect Category Sentiment Analysis (ACSA) in compound ABSA tasks, which handle all aspect categories simultaneously:

1. **Improved Data Balancing:** By processing aspect categories individually, it becomes easier to balance the dataset for each category.
2. **Focused Learning:** The model can more directly associate sentiment polarities with specific aspect categories.
3. **Flexible Architecture:** There is no need to alter the model's architecture when merging or adding new aspect categories.

These strategies collectively enhance the model's performance and adaptability, ensuring that it can effectively classify sentiment polarities across various aspect categories with improved accuracy and efficiency.

3.5 Data augmentation and Balancing Techniques

Data augmentation has become increasingly significant in recent years, allowing for the generation of additional training data from the existing dataset. This process helps reduce overfitting, especially when data is limited. Although these techniques are straightforward to implement and do not rely on external factors, they substantially improve model performance. In this study, I employ the following techniques: Masked Language Modeling, Back Translation, and Sentence Combining.

3.5.1 Masked Language Modeling

The Masked Language Modeling (MLM) is utilized during the training of models such as BERT and RoBERTa, where the model predicts masked words based on the context. Leveraging this characteristic, we can augment our dataset. In this study, after tokenizing the input, a certain proportion of tokens are masked, except for the initial tokens related to the aspect category.

3.5.2 Back Translation

Back Translation has been employed to generate new data and improve model performance. This approach targets low-frequency words to create new sentences containing rare words and diverse contexts. The technique involves translating the original data into a specific language and then using an independent translator to translate it back to the original language.

For example, the English sentence: “Get to the site early as the bus loads of tourists seems to mount up after lunch and the weather seemed a lot cooler in the morning.”

Is translated into French: “Rendez-vous sur le site tôt que les charges de bus des touristes semble monter après le déjeuner et le temps semblait beaucoup plus frais le matin.”

Then back to English: “Visit the site early that the bus loads of tourists seem to go up after lunch and the weather seemed much cooler in the morning.”

In this study, I use four intermediary languages: **German, Spanish, Russian, and French**. The translation models used are “Helsinki-NLP/opus-mt-*”.

3.5.3 Sentence Combining

As previously mentioned, we can combine two sentences and compute the sentiment polarity using the Polarity Combining Rule. Data belonging to pairs of (aspect category - sentiment polarity) are generated randomly until the quantities in these pairs are balanced. Note that the two sentences can be swapped.

To generate none/neutral sentiment polarity: $0 + 0 \rightarrow 0$

To generate positive sentiment polarity: $1 + 0 \rightarrow 1$ (50%); $1 + 1 \rightarrow 1$ (50%)

To generate negative sentiment polarity: $2 + 0 \rightarrow 2$ (50%); $2 + 2 \rightarrow 2$ (50%)

To generate conflict sentiment polarity: $1+2 \rightarrow 3$ (50%); $3+0/1/2/3 \rightarrow 3$ (50%);

This chapter emphasized the meticulous construction of a relevant and comprehensive dataset, ensuring the inclusion of key tourism-related aspects and sentiments. Fine-tuning the RoBERTa model was pivotal in optimizing performance for sentiment analysis tasks. Additionally, advanced data augmentation and balancing techniques were integrated to bolster model robustness and generalizability. These foundational steps lay the groundwork for the numerical results discussed in Chapter 4, demonstrating the efficacy of our ABSA framework in extracting and analyzing sentiment from tourism-related user feedback.

4.1 Overview

In this chapter, I delve into the detailed analysis of numerical results derived from extensive experiments conducted to evaluate our aspect-based sentiment analysis (ABSA) framework. Building upon the methodologies outlined in Chapter 3, the experiments focus on assessing the effectiveness of data augmentation techniques and the performance of the RoBERTa-base model across various aspect categories and sentiment polarities. The computational experiments were executed using a single NVIDIA RTX 3060 GPU, necessitating fine-tuning of the RoBERTa-base model. Specifically, attention was given to optimizing the final encoding layer (layer 11) and the classifier layer, consisting of one dense layer and an output layer. These experiments provide critical insights into the robustness and applicability of our ABSA framework in the context of tourism-related user feedback analysis.

4.2 Dataset

The dataset used for these experiments was split using a hold-out method with a training-to-testing ratio of 4:1. The figure 4.1 below depicts the distribution of polarities with respect to aspect categories.

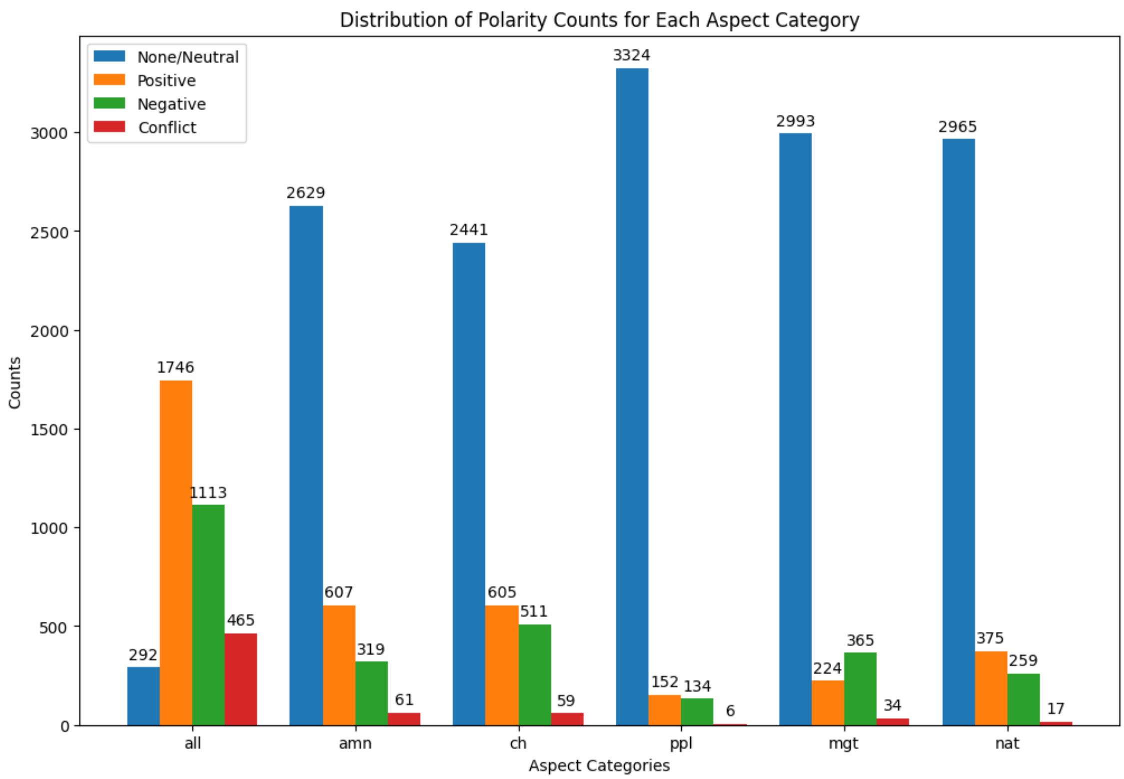

Figure 4.1: Sentiment Polarity Distribution in Aspect Categories

Among the sentiment polarities, excluding the all-inclusive category, None/Neutral polarity has the highest count across all aspect categories. In contrast, Conflict polarity has the lowest count, indicating that sentences classified under this category are rare. This scarcity poses challenges for training and evaluating models, as the limited data can lead to underrepresentation and difficulty in accurate prediction.

Positive sentences generally outnumber Negative ones, except in the Management (mgt) category, where Negative sentences prevail. This imbalance may influence the sentiment analysis models' performance, making it easier for the models to learn and predict positive sentiments compared to negative ones.

Excluding the all-inclusive category and None/Neutral polarity, the Culture Heritage (ch) category has the largest number of sentences, followed by Amenities (amn), Nature (nat), and Management (mgt). The People (ppl) category has the fewest sentences. This distribution is understandable given that the remaining categories have been aggregated from smaller sub-categories, whereas People is a more specific aspect.

4.3 Effectiveness of Data Augmentation Techniques

This section compares the model's performance when utilizing different data augmentation techniques. The results are consolidated in the table below for clarity.

Table 4.1: Comparison of Data Augmentation Techniques

Technique	Accuracy	Macro-Recall	Macro-Precision	Macro-F1
SC	0.8690	**0.7937**	0.6583	0.6937
SC + MLM	0.8801	0.7764	0.6870	0.7087
SC + MLM + BT	**0.9040**	0.7600	**0.7304**	**0.7321**

As illustrated in the table above, by trade-off between precision and recall, the combination of Sentence Combining, Masked Language Model, and Back Translation yielded the highest accuracy and Macro-F1 score. This suggests that leveraging multiple data augmentation techniques can significantly enhance the model's performance by providing a richer and more varied training dataset.

4.4 Results of Aspect Categories and Polarities

This section details the model's performance when applying all three data augmentation techniques. Table 4.2 and Figure 4.2 illustrate the F1-scores across different polarities and aspect categories. The experimental results align consistently with the data distribution analyzed.

Table 4.2: F1 Scores for Different Aspect Categories and Polarities

Aspect Category	None/Neutral	Positive	Negative	Conflict
all	0.6780	**0.9001**	**0.8989**	**0.6907**
amn	0.9374	0.7911	0.7087	0.3200
ch	0.9331	0.7928	0.8571	0.6087
mgt	0.9539	0.6813	0.7738	0.3529
nat	0.9640	0.8187	0.8000	0.5455
ppl	**0.9746**	0.8000	0.6222	0.1667

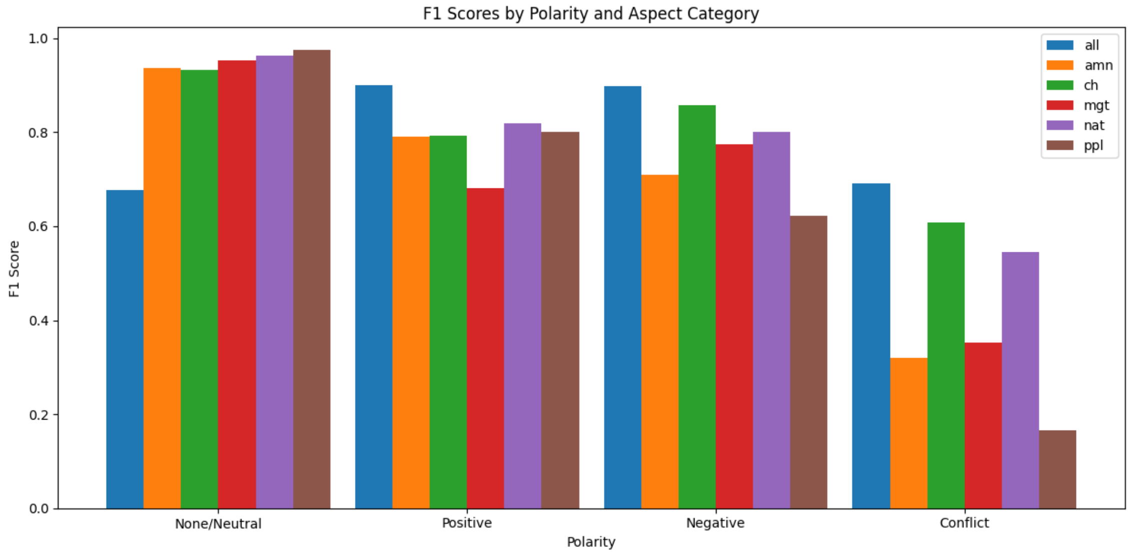

Figure 4.2: F1 Scores by Polarity and Aspect Category

For less complex tasks, such as predicting the polarity of the all-inclusive category, the model performs exceptionally well, achieving an F1-score of 0.9001 for Positive and 0.8989 for Negative polarities.

Furthermore, the performance improves with sentiment analysis and aspect category pairs that have larger data. For instance, the F1-score for the (negative-ch) pair, with 511 instances, is substantially higher than that for the (negative-amn) pair, which has only 319 instances. Similarly, the (negative-mgt) pair, with 365 instances, outperforms the (positive-mgt) pair, which contains 224 instances. This indicates that a larger dataset contributes to better model performance and more reliable predictions.

4.4.1 Comparison of Sentiment Polarities

In a similar manner, to evaluate the model's performance across different sentiment polarities, I calculated the macro-recall, macro-precision, and macro-F1 scores. These metrics represent the average performance across all aspect categories, namely: all-inclusive (all), amenities (amn), cultural heritage (ch), management (mgt), nature (nat), and people (ppl). This comprehensive approach ensures a balanced assessment of the model's effectiveness in handling various sentiment polarities across diverse categories.

Polarity	Marco-Recall	Marco-Precision	Marco-F1
None/Neutral	**0.9174**	**0.8982**	**0.9068**
Positive	0.7989	0.8015	0.7973
Negative	0.7799	0.7811	0.7768
Conflict	0.5439	0.4408	0.4474

Table 4.3: Comparison of Sentiment Polarities based on Macro-F1 Scores

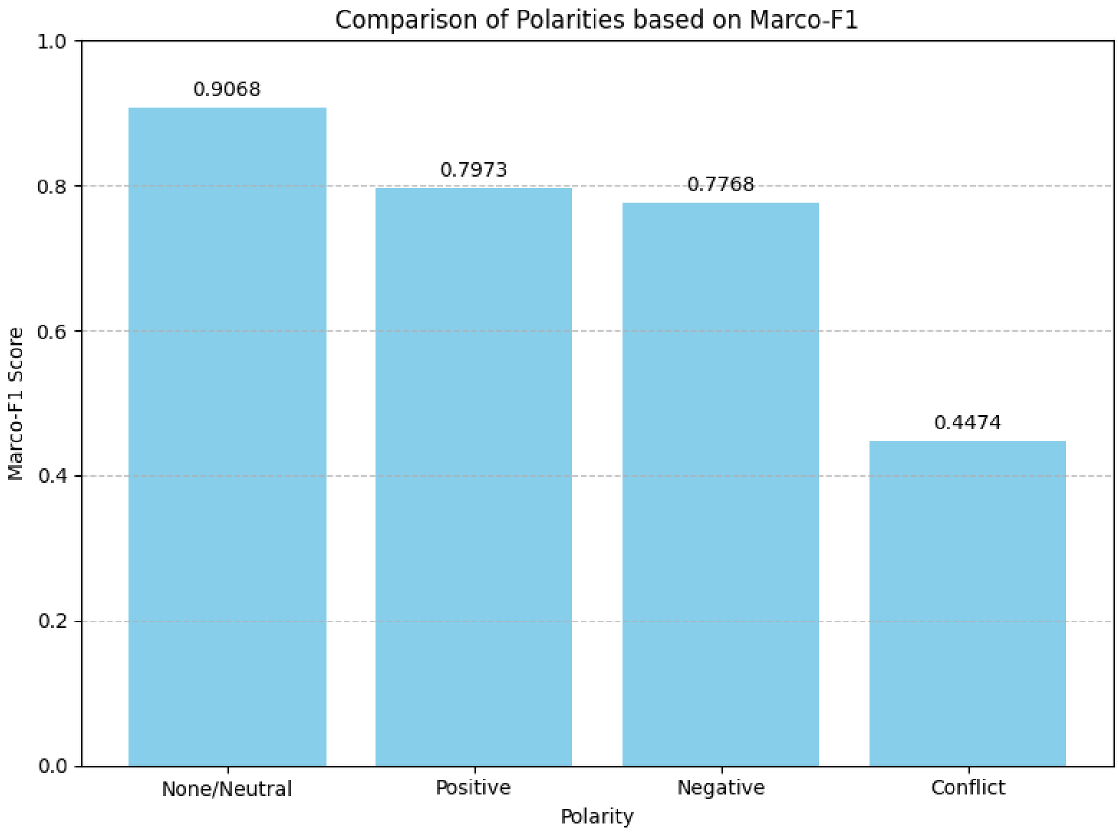

Figure 4.3: Comparison of Sentiment Polarities based on Marco-F1 Scores

Analysis of Figure 4.3 the None/Neutral polarity achieves the highest macro-F1 score, primarily due to the abundance of data. This polarity benefits from the fact that most sentences contain fewer than two aspect categories, resulting in a more straightforward analysis. In contrast, the Conflict polarity exhibits the lowest macro-F1 score. There are two primary reasons for this:

1. **Lack of Data**: The occurrence of sentences containing both positive and negative opinion terms within the same aspect category is exceedingly rare. Consequently, the dataset has a limited number of these sentences, making it challenging to train the model effectively.

2. **Sentence Complexity**: Sentences labeled as Conflict tend to be long and

complex, presenting significant difficulties during the labeling process. The intricate nature of these sentences hampers the model's ability to accurately capture and analyze the sentiments expressed.

Furthermore, the performance of the positive and negative categories is quite similar. Despite the positive category having more data (3709 instances labeled as *-1) compared to the negative category (2701 instances labeled as *-2), their macro-F1 scores are comparable. This outcome suggests that the use of Sentence Combining to balance the data has proven to be highly effective, ensuring that the model can handle both positive and negative sentiments with similar proficiency.

4.4.2 Comparison of Aspect Categories

To evaluate the quality of the model with aspect categories, I calculated marco-recall, marco-precision, marco-f1 (average of None/Neutral, Positive, Negative polarity) for each aspect category Table 4.4. Results of Conflict polarity is not reliable, so the conflict parameter is ignored so as not to affect the results.

Aspect Category	Marco-Recall	Marco-Precision	Marco-F1
all	0.8506	0.8043	0.8256
amn	0.7858	0.8455	0.8124
ch	0.8724	0.7074	0.7178
mgt	0.8379	0.7747	0.8030
nat	**0.8559**	**0.8689**	**0.8609**
ppl	0.7900	0.8100	0.7990

Table 4.4: Comparison of Aspect Categories based on Marco-F1 Scores

Figure 4.4 reveals that the Nature category achieves the highest Marco-F1-score among the aspect categories. This superior performance can be attributed to the narrower and more specific domain of the Nature category, which facilitates more precise sentiment analysis despite not having the largest amount of data. In contrast, the Cultural Heritage category, although it encompasses more data than other categories, exhibits lower F1-scores. This can be ascribed to the complexity and inconsistency inherent within this category, which adversely affects the model's performance.

From these observations, we infer that the quality of the model for different aspect categories is predominantly influenced by three key factors, in descending order of importance: the cleanliness of the data, the complexity and breadth of the aspect category domain, and the quantity of data available.

This chapter provides a comprehensive analysis of our aspect-based sentiment analysis (ABSA) framework, focusing on data analysis, comparison of data augmen-

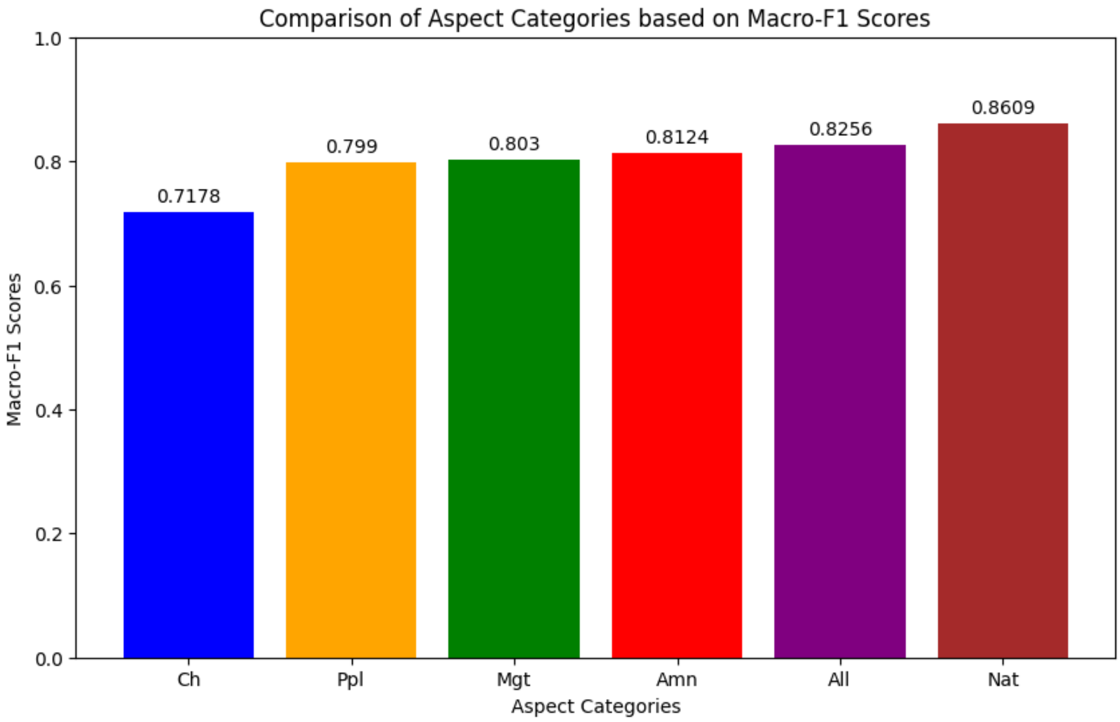

Figure 4.4: Comparison of Aspect Categories based on Marco-F1 Scores

tation techniques, and evaluation of model performance across sentiment polarities and aspect categories. The results demonstrate the effectiveness of data augmentation methods in improving model robustness. These impressive metrics across different aspect categories and sentiment polarities highlight the model's ability to identify categories while also capturing emotional expressions in travel-related responses. These insights contribute to advancing ABSA applications in the tourism sector, setting the stage for detailed conclusions in Chapter 5.

CHAPTER 5. CONCLUSIONS

5.1 Summary

This thesis addressed several key issues in aspect-based sentiment analysis (ABSA), focusing on enhancing model performance through data augmentation techniques. By employing methods like Sentence Combining (SC), Masked Language Model (MLM), and Back Translation (BT), we observed significant improvements in accuracy, macro-precision, and macro-F1 scores. These techniques effectively enriched the training data, contributing to better model performance.

A critical component of this research was the **Dataset Construction Methodology**. This involved meticulous selection of sentiment polarities and aspect categories to ensure a comprehensive and balanced dataset. The chosen sentiment polarities—None/Neutral, Positive, Negative, and Conflict—covered a wide range of sentiments, providing a robust foundation for analysis. Similarly, the aspect categories were selected to represent diversse facets of the target domains, including amenities, management, culture heritage, nature, and people. This careful selection process was crucial in creating a dataset that could accurately reflect the complexities of real-world sentiment and aspect interactions.

The project also analyzed the distribution of sentiment polarities across various aspect categories, highlighting the challenges posed by the uneven distribution of data, especially for the Conflict polarity, which had the lowest representation. The results demonstrated a consistent relationship between data quality, data quantity and model performance, with more clean and abundant data leading to higher F1 scores.

However, some issues remain unresolved. The Conflict polarity's low representation poses challenges for model training and evaluation.

5.2 Suggestion for Future Works

Future research can build on this work by exploring the following directions:

1. **Enhanced Data Augmentation**: Investigate more sophisticated data augmentation methods to address the recall decrease observed in some techniques. This could include adversarial training or synthetic data generation to create more diverse and challenging datasets.

2. **Balanced Dataset Creation**: Develop strategies to balance the dataset, especially for underrepresented polarities like Conflict. This might involve targeted data collection or synthetic data generation techniques tailored to specific

polarities.

3. **Aspect and Polarity Interaction Analysis**: Conduct a deeper analysis of the interactions between different aspect categories and sentiment polarities. Understanding these interactions could lead to more nuanced models that can better capture the complexity of real-world sentiment.

4. **Cross-Domain ABSA**: Extend the current ABSA models to other domains beyond the initial dataset, such as cultural heritage sites like Hue Imperial City in Vietnam and Changdeokgung Palace in South Korea. This could involve fine-tuning models on domain-specific datasets to improve their applicability and robustness. By addressing these areas, future research can build on the foundation laid by this thesis to create even more effective and versatile aspect-based sentiment analysis models.

REFERENCE

[1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[2] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 27–35.

[3] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, "A challenge dataset and effective models for aspect-based sentiment analysis," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 6280–6285.

[4] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," *arXiv preprint arXiv:1911.01616*, 2019, Version 4.

[5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018. arXiv: `1810.04805`. [Online]. Available: `http://arxiv.org/abs/1810.04805`.

[6] Y. Zhu, R. Kiros, R. S. Zemel, *et al.*, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," *CoRR*, vol. abs/1506.06724, 2015. arXiv: `1506.06724`. [Online]. Available: `http://arxiv.org/abs/1506.06724`.

[7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient estimation of word representations in vector space*, 2013. arXiv: `1301.3781 [cs.CL]`. [Online]. Available: `https://arxiv.org/abs/1301.3781`.

[8] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds., Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. DOI: `10.3115/v1/D14-1162`. [Online]. Available: `https://aclanthology.org/D14-1162`.

[9] Y. Liu, M. Ott, N. Goyal, *et al.*, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019. arXiv: `1907.11692`. [Online]. Available: `http://arxiv.org/abs/1907.11692`.

[10] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," *arXiv preprint arXiv:2203.01054*, 2022, Version 2.

[11] D. Arianto and I. Budi, "Aspect-based sentiment analysis on indonesia's tourism destinations based on google maps user code-mixed reviews (study case: Borobudur and prambanan temples)," in *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, Hanoi, Vietnam: Association for Computational Linguistics, 2020, pp. 359–367.

[12] D. I. Af'idah, P. D. Anggraeni, M. Rizki, A. B. Setiawan, and S. F. Handayani, "Aspect-based sentiment analysis for indonesian tourist attraction reviews using bidirectional long short-term memory," *Juita : Jurnal Informatika/Jurnal Informatika*, vol. 11, no. 1, p. 27, 2023. DOI: `10.30595/juita.v11i1.15341`. [Online]. Available: `https://doi.org/10.30595/juita.v11i1.15341`.

[13] T. Ali, B. Omar, and K. Soulaimane, "Analyzing tourism reviews using an lda topic-based sentiment analysis approach," *MethodsX*, vol. 9, p. 101 894, 2022. DOI: `10.1016/j.mex.2022.101894`. [Online]. Available: `https://doi.org/10.1016/j.mex.2022.101894`.

[14] N. T. L. Huong and P. T. Hoan, "Factors constituting tourist destination at phu loc – thua thien hue," *Khoa học - Đại học Huế*, vol. 126, no. 5A, pp. 101–113, 2017, ISSN: 2588-1205.

[15] T. t. v. D. l. Bộ Văn hóa, *Quyết định 4640/qđ-bvhttdl về việc ban hành bộ tiêu chí đánh giá điểm đến du lịch*, 2016. [Online]. Available: `https://thuvienphapluat.vn/van-ban/van-hoa-xa-hoi/quyet-dinh-4640-qd-bvhttdl-2016-bo-tieu-chi-danh-gia-diem-den-du-lich-340594.aspx`.