DOCTORAL THESIS

A Study on Regional Support for Inbound Tourism Based on Review Analysis

レビュー解析に基づく地域インバウンド観光支援に関する研究

by

Zhenzhen Liu

Advised by:
Fumito Masui

KITAMI INSTITUTE OF TECHNOLOGY
GRADUATE SCHOOL OF ENGINEERING

September 2023

Acknowledgements

This thesis would not have been possible without the tremendous support I have received, and I would like to take this opportunity to express my gratitude.

First and foremost, I am deeply grateful to my supervisors, Professor Fumito Masui, Professor Michal Ptaszynski and Professor Toshio Eisaka, for their invaluable help, guidance, and insights throughout the course of my research. Their expertise and unwavering support have been instrumental in shaping my academic development, and I am indebted to them for their continued mentorship. They have provided me with valuable feedback and constructive criticism, pushing me to constantly improve and refine my work.

I would also like to express my sincere appreciation to the members of my dissertation committee, Professor Yasunari Maeda and Professor Shinichi Morita. Their expertise and critical evaluation of my dissertation was invaluable in helping me to refine and strengthen my research. Their feedback and comments have contributed significantly to the overall quality of this work.

Finally, I would like to express my deepest gratitude to my family and friends who have been by my side throughout this journey. Their unwavering support, understanding, and encouragement have been invaluable. I am especially grateful to my boyfriend, whose constant support and belief in me has been a source of motivation during challenging times.

To all those who have supported and contributed to this research, I offer my sincere thanks. Your help and encouragement have played a vital role in my academic growth and success.

Zhenzhen Liu
September 2023

ABSTRACT

The COVID-19 pandemic has significantly impacted the tourism industry, including the number of inbound tourists to Japan. While Japan had been experiencing a steady increase in inbound tourism in recent years, the pandemic caused a sharp decline in visitor numbers in 2020. This decline raises concerns, as Japan had set a target of attracting 60 million inbound tourists annually by 2030. To aid in the recovery of Japan's tourism industry post-pandemic, I propose a method for identifying the key elements that attract the attention of inbound tourists (focus points) by analyzing reviews of tourist sites.

Currently, there is a lack of a comprehensive and objective approach to identifying the focus points of tourist destinations. Previous studies have relied on subjective methods such as surveys and expert opinions, or utilized keyword extraction techniques that may not capture the full spectrum of factors influencing tourists' preferences. In this thesis, I present a novel method that combines keyword extraction, scoring based on motivational factors, and principal component analysis to pinpoint the most attention-grabbing aspects of tourist spots. By applying this method to popular tourist destinations in Hokkaido, with a specific focus on Chinese tourists, my study provides valuable insights into the specific elements that hold the greatest appeal for this particular group of visitors.

In the first phase of my research, I collected a substantial number of reviews from a prominent Chinese travel industry website that focused on popular tourist spots in Hokkaido. The goal was to extract relevant keywords from these reviews to uncover the potential points of interest for Chinese tourists. I used two different kinds of keyword extraction methods: TF-IDF (term frequency-inverse document frequency) and TextRank. Through my analysis, I found that the TF-IDF algorithm produced the most promising results in this study. By examining the extracted keywords derived from TF-IDF, I was able to gain clear insights into the specific elements that Chinese tourists prioritize when considering tourist destinations.

Building on this foundation, I proceeded to extract high-frequency n-gram patterns from the reviews written by Chinese inbound tourists, with each n-gram pattern containing the previously extracted keywords. To assess the focus of each destination, I used seven types of motivational factors commonly associated with tourist behavior and applied principal component analysis. This allowed us to effectively quantify and identify the distinctive features and focus points of each tourist destination.

Subsequently, I took my analysis a step further by clustering the n-gram patterns extracted from the tourists' reviews. This clustering process allowed us to group similar patterns together and delve deeper into the underlying themes and characteristics that emerged from the reviews. By carefully examining the clustering results, I was able to provide insightful advice and recom-

mendations to improve the overall tourist experience and develop effective industry strategies based on the identified areas of focus.

Through this comprehensive research approach, I aimed to provide valuable insights into the preferences and expectations of Chinese tourists when selecting and experiencing tourist sites in Hokkaido. The results of this study can serve as a basis for industry stakeholders to tailor their offerings and marketing strategies to better meet the interests and needs of Chinese tourists, ultimately strengthening the tourism industry in Hokkaido and beyond.

ABSTRACT IN JAPANESE (論文内容の要旨)

COVID-19のパンデミックは，日本へのインバウンド観光客数をはじめ，観光産業に大きな影響を与えた.日本は近年，インバウンド観光客が順調に増加していたが，パンデミックにより，2020年には観光客数が激減した.日本は2030年までに年間6,000万人のインバウンド観光客を誘致する目標を掲げていたため，この減少が懸念される.そこで，パンデミック後の日本の観光産業の回復を支援するため，観光地のレビューを分析することで，インバウンド観光客が注目するポイント（フォーカスポイント）を特定する手法を提案する.

現在，観光地のフォーカスポイントを特定するための包括的かつ客観的なアプローチは欠如している.これまでの研究では， アンケートや専門家の意見といった主観的な手法に頼ったり，観光客の嗜好に影響を与える要因の全容を把握できない可能性のあるキーワード抽出技術を利用したりしてきた.本論文では，キーワード抽出，動機付け要因に基づくスコアリング，主成分分析を組み合わせて，観光地の最も注目される点をピンポイントで特定する新しい手法を提案する.この手法を北海道の人気観光地に適用し，中国人観光客に焦点を当てることで，観光客グループにとって最も魅力的な要素を特定することができた.

中国の旅行業界のウェブサイトから，北海道の人気観光地に関するレビューを収集した.これらのレビューから関連キーワードを抽出し，中国人観光客の潜在的な関心を明らかにした.キーワードの抽出には，2種類の方法を用いた：TF-IDF（用語頻度-逆文書頻度）とTextRankである.分析を通じて，TF-IDFアルゴリズムが本研究で最も優れた結果をもたらすことがわかった.TF-IDFで抽出されたキーワードから，中国人観光客が観光地を訪れた際に注目した要素を明確に把握することができた.

これをもとに，中国人観光客が書いたレビューから，先に抽出したキーワードを含む高頻度のn-gramパターンを抽出した.また，各観光地の注目要素を7種類の観光動機を用いてスコアリングを行い，主成分分析を適用した.これにより，各観光地のフォーカスポイントを定量化・特定することができた.

さらに，n-gramパターンをクラスタリングすることで，フォーカスポイントを可視化した.このクラスタリングプロセスにより，類似したパターンをグループ化し，レビューから浮かび上がった根本的なテーマやトピックをより深く掘り下げることができた.クラスタリングの結果から，観光体験を向上させ， 効果的な産業戦略を開発するためのアドバイスなどを提供することができる.

本研究は，中国人観光客が北海道の観光地を選択・体験する際の好みと動機を自動抽出し観光産業へ提供することを目指した.本研究の結果は，業界関係者が中国人観光客の関心やニーズに合わせて，より良いサービスやマーケティング戦略を提供するための基礎となり，最終的に北海道とそれ以外の地域も観光産業を強化することができる.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The COVID-19 pandemic had a profound impact on the global tourism industry. According to the World Tourism Organization [1], international tourist arrivals fell by 72% in 2020 compared to the previous year. The impact of the pandemic on the tourism industry in Asia and the Pacific has been particularly severe, with international arrivals falling by 84% in 2020. The pandemic has led to the closure of many tourism-related businesses, such as hotels, restaurants and airlines, resulting in job losses and economic hardship for many people in the tourism sector. To control the spread of the virus, many countries have implemented travel restrictions and quarantine measures, further reducing international travel. The pandemic has also accelerated the trend towards domestic tourism, with many people choosing to travel within their own countries rather than abroad. However, as vaccines become more widely available and travel restrictions are gradually lifted, the tourism industry is showing signs of returning to pre-pandemic levels of activity.

The Japanese government had launched a "Visit Japan" campaign with the goal of reaching 40 million inbound tourists annually by 2020 and 60 million by 2030 [2]. However, due to the significant impact of the coronavirus pandemic, the number of tourists decreased drastically. According to the announcement by the Japan National Tourism Organization, the number of inbound tourists to Japan in April 2020 decreased by 100% compared to the same month of the previous year [3]. It has also been pointed out that the impact of the COVID-19 pandemic on the tourism industry is more severe than that of the Great East Japan Earthquake in 2011 [4].

According to the "White Paper on Tourism in Japan (2020)"[5] issued by the Japan Tourism Agency of the Ministry of Land, Infrastructure, Transport and Tourism, the plan is to "further enhance the attractiveness of Japan in order to revive inbound tourism after COVID-19, and to promote the recovery of the tourism industry, improve local tourism resources that attract domestic and foreign tourists, and promote the development of reception environments that utilize cutting-edge technology during the period before recovery". Therefore, it has become crucial to develop strategies to revive inbound tourism with a focus on the post-COVID era.

In order to rebuild inbound tourism activities after COVID-19, it is necessary to objectively understand and utilize the points that attract foreign tourists in the future. The research question of this study is: What are the underlying motivations that drive tourists to visit certain destinations, and how can I use a statistical method to identify and extract the interest of tourists from online reviews that are related to these motivations?

Previous studies have used subjective approaches, such as surveys and expert evaluations, or relied solely on keyword or topic extraction techniques, which may not account for the full range of factors that affect tourists. Therefore, I proposed a new statistical method to objectively capture the interests of inbound tourists visiting Japanese tourist destinations and identify the main tourism motivations associated with them. Collecting real-time tourism reviews and using this method to automatically extract tourists' points of interest can provide valuable information to the tourism industry on an ongoing basis. This enables the adjustment of specific travel plans according to visitors' needs, better catering to their preferences, attracting more tourists, and facilitating the recovery of the tourism sector.

Furthermore, since my proposed method is not language dependent, I can apply my method to other languages and countries, making it useful for global tourism where multilingualism is involved. Moreover, by using my method to compare samples from before and after COVID-19, I can demonstrate the difference in the focus points of tourists. It means that my approach will help to prepare for the next emergency situation affecting tourism.

In this study, I propose a method for automatically extracting focus points, which include attention-grabbing aspects and related tourism motivations, from reviews written by inbound tourists about tourist spots. First, I extracted keywords from

the reviews related to inbound tourism at each tourist spot. Then, I identified the most frequently occurring n-gram patterns that included the previously extracted keywords. Next, I assigned scores to the n-gram patterns based on the seven types of motivational factors that influence tourists. I then performed Principal Component Analysis (PCA) on the scoring results and used the obtained Principal Component Charts to determine the focus points of the tourist destinations. Finally, I applied clustering methods to the n-gram patterns to further analyze the details of the focus points. The experimental results indicate the interest of tourists. In addition, the corresponding motivations toward different tourist spots were estimated.

In this thesis, I focus on tourists from China, who accounted for the majority of inbound tourism in Japan before the outbreak of the COVID-19 pandemic. In addition, I have analyzed reviews of popular tourist spots in Hokkaido, which is one of the top destinations for inbound tourists to Japan.

1.1 Contributions

This thesis makes several contributions to the field of tourism research. The main contributions can be summarized as follows:

- A novel method for identifying the focus points of tourist sites based on the analysis of reviews of tourist sites written by inbound tourists.
- Applying the proposed methodology to analyze reviews of popular tourist destinations in Hokkaido, with a focus on Chinese tourists, to gain insights into what aspects of these destinations are most appealing.
- Providing a global approach to tourism focus point analysis and a potential solution to prepare for future tourism emergencies.

First, I propose a novel method to identify the focus points of tourist sites, which are the attention-grabbing aspects that attract inbound tourists. The method involves analyzing tourist site reviews written by inbound tourists and extracting highly frequent n-gram patterns, which are scored based on seven types of motivational factors that influence tourists. I use principal component analysis

and clustering methods to further analyze the data and identify the focus points of each tourist site.

Moreover, I apply my methodology to analyze reviews of popular tourist spots in Hokkaido, with a special focus on Chinese tourists. The clustering of n-gram patterns extracted from tourist reviews allows the identification of key features and focus points of each tourist spot, which can be used to better target tourism marketing strategies and improve the tourism industry in Japan.

Since the method is not language dependent, it can be applied to other languages and countries, allowing for a global approach to tourism. The method can also be used to compare the difference in the focus points of tourists before and after the COVID-19 pandemic, which could help prepare for future emergencies in tourism.

1.2 Organization

The remainder of this thesis is organized as follows. Chapter 2 describes the background of my research. First, I go through the research related to the impact of the COVID-19 pandemic to the tourism industry. Second, I go through research assessing travel motivation. This means the underlying reasons and factors that influence a person's decision to travel to a particular destination or engage in travel-related activities. Third, I go through research that concentrates on analyzing the focus points of tourists, meaning the analysis of the interest of tourists and the attention-grabbing features of tourism spots. In Chapter 3 I go through the fundamentals of my research and give an overview of the used methods and techniques used. In Chapter 4, I introduce the dataset used in this thesis and go through the different keyword extraction methods. Additionally I evaluate the methods in order to find most useful method to be used later in the focus point analysis. In Chapter 5 I use the extracted keywords with an n-gram based technique and principal component analysis to extract tourism focus points. Additionally, I perform clustering analysis to gain further insights into the attention-grabbing aspects of tourism spots. In Chapter 6, I discuss the results of this research and their impact to the tourism industry. In Chapter 7 I summarize and review all of the principal findings of this research and discuss ideas for future work.

Chapter 2

Related Works

2.1 Tourism and COVID

The COVID-19 pandemic has negatively impacted the tourism and hospitality industry. There have been numerous studies discussing the state of the tourism industry and how it is coping with the pandemic.

For example, Faeni et al. [6] propose a model for increasing human capital competitiveness in the tourism industry in emerging economies, using Indonesia as an example. The authors conducted a survey of 199 tourism workers in the city of Magelang and analyzed the data using a structural equation model. The study found that social and human capital influences firm performance and that innovation moderates the influence of human and social capital on firm performance. The authors highlight the potential of creating and sharing knowledge to strengthen micro, small, and medium-sized enterprises in the tourism industry in emerging economies during and after the COVID-19 pandemic.

A study by Pramana et al.[7] analyzes the impact of the COVID-19 pandemic on Indonesia's tourism industry by clustering provinces based on their room occupancy rates. Big data sources, including the Google Mobility Index, flight trackers, and reviews from Tripadvisor and Booking.com, are used to examine the impact on Bali and Yogyakarta. The study shows that the pandemic has impacted the tourism industry and its supporting sectors across Indonesia, but the patterns of impact vary across provinces. The authors suggest that big data sources can serve as a useful proxy for inferring the impact of the pandemic on tourism.

Zhao et al. [8] investigate the factors influencing online reservation intentions of tourist attractions in the COVID-19 context, based on the technology acceptance model (TAM). The study found that subjective norms had no significant effect on reservation behavior, perceived usefulness had a positive effect on tourists' reservation intention, and perceived risk had a significant negative effect on reservation intention. Government policy was the most important factor influencing tourists' reservation intentions. The findings improve the understanding of tourists' reservation intentions and extend TAM theory, and suggest that tourist attraction operators should improve tourists' experience and reduce perceived risk, and the government should promote the reservation system to create a good reservation atmosphere.

The pandemic has also highlighted the need for physical distance. One solution to alleviate this situation is the use of robots to guide and serve customers. There are several research studies that focus on the use of robots to assist the tourism industry during the COVID-19 pandemic.

For example, Zeng et al. state that [9]robotics, artificial intelligence, and human-robot interactions are increasingly being used to manage the spread of the virus in various sectors. They say the use of humanoid robots, autonomous vehicles, drones, and intelligent robots can help reduce human contact and potential transmission of the virus. While controversial in the past, the use of robotics and AI in the travel and tourism industry is likely to continue beyond the pandemic. According to the authors, tourism scholars should seize this opportunity to develop robotic applications that enhance tourist experiences, protect natural and cultural resources, involve citizens in tourism development decisions, and create new employment opportunities for industry workers.

Romero and Lado [10] found that customers believe that robots in hotels can reduce the risk of contagion during the COVID-19 pandemic. They say the use of anthropomorphic robots can increase the perceived effectiveness of COVID-19 prevention, resulting in more positive attitudes and higher booking intentions. In addition, hotels can increase demand by promoting robots as a COVID-19 prevention measure, especially in markets heavily impacted by the pandemic. They also state that robots should be used in contexts with low social presence.

Seyitolu et al. [11] state that service robots can be useful in maintaining physical

distance, but they can also create a technological barrier between tourists and employees. They suggest that tourism and hospitality companies need to use additional technologies to promote social connection and mitigate the negative effects of physical distance.

2.2 Travel Motivation

Travel motivation refers to the underlying reasons and factors that influence a person's decision to travel to a particular destination or engage in travel-related activities. Understanding travel motivation is important for the tourism industry to provide a better travel experience and attract more tourists[12].

There are many studies that analyze and discuss tourists' travel motivation. For example, Hayashi et al. [13] surveyed Japanese overseas travelers to determine their tourism motivation and identified the factors that contribute to their motivation. They performed factor analysis on the questionnaire results and confirmed seven types of motivational factors for tourists, including stimulation, cultural observation, local communication, health recovery, experiencing nature, unexpectedness, and educating oneself.

As for the motivation of Chinese tourists, Wen et al. [14] explore the relationship between Chinese cultural values and tourist motivations in Israel. Through surveys and interviews, they found seven main reasons for Chinese tourists to visit Israel, including knowledge enhancement/learning, business development, sightseeing, self-fulfillment, escape/relaxation, destination uniqueness, and adventure. In particular, business development was a significant factor. Chen et al. [15] examine the role of "face" in Chinese tourism, which has received limited research attention. Through qualitative interviews with 20 Chinese tourists, the study found that "face" influences travel behavior and destination choice, leading to the selection of high-value and prestigious destinations, luxury tourism products, and social media sharing to "gain face".

A study by Simeon et al. [16] analyzes online reviews of cultural attractions in Naples, Italy posted on TripAdvisor to explore tourists' experiences and identify their preferences. Content analysis and principal component analysis are used to identify five critical components of tourists' experiences: wonder, authenticity,

relaxation, discovery, and knowledge. The study provides practical implications for destination managers and policymakers to enhance the attractiveness of cultural attractions and provide more satisfying cultural experiences.

As this study focuses on tourist destinations in Japan, we choose to use the seven types of motivational factors proposed by Hayashi et al. [13], which also helps to compare and discuss the differences between Chinese and Japanese tourists.

2.3 Focus Points of Tourists

There are many studies that analyze the focus or image of various destinations in Japan as perceived by inbound tourists. These studies mainly use reviews posted on social media and travel websites as data sources. For example, Okubo et al. [17] analyzed the gap between the expectations and evaluations of tourist destinations from travel guidebooks and review data, and revealed the differences in the image of tourist destinations among different countries. Specifically, they extracted and compared nouns and adjectives with higher Jaccard indexes from text data of travel guidebooks and reviewed data of tourist destinations for five areas in Tokyo. They then performed PCA on the extracted business nouns and adjectives to analyze the image of tourist destinations by nationality. Ohkawa et al. [18] visualized the similarities and differences in the image of tourist destinations from multilingual reviews posted on the Internet, and conducted a comparative analysis. They extracted frequently occurring nouns from review texts in Japanese, English, and Chinese for 10 tourist destinations in Japan, and performed correspondence analysis of the extracted words using the k-means method. Then, they analyzed the similarities and differences in the images of tourist spots based on different languages.

There are also many studies that compare the elements of attention between domestic tourists and inbound tourists for the same tourist resources. For example, Hoshino et al. [19] collected Twitter posts related to a fireworks event in central Tokyo and conducted topic extraction using LDA for Japanese and English tweets, respectively. Then, using the obtained results, they compared and analyzed the tourism information that Japanese and foreign tourists are interested in. In addition, they examined the problems of providing tourism information to foreign tourists

and made suggestions on how to disseminate information on websites. Yasuhara et al. [20] collected Japanese and English reviews of Japanese gardens from travel websites and discussed the similarities and differences between them. They ranked the words contained in the reviews in both languages by frequency of occurrence and created a collocation network using the top 100 words. They then compared the visit experiences of Japanese and foreign visitors using the obtained collocation network and revealed their similarities, differences, and evaluation tendencies.

Currently, however, there is a lack of a comprehensive and objective method for identifying the focus points of tourist destinations. Previous studies have relied on subjective approaches such as surveys and expert opinions, or on keyword extraction methods that may not capture the full range of factors that influence tourists. This paper proposes a novel method that combines keyword extraction, scoring based on motivational factors, and principal component analysis to identify the most attention-grabbing aspects of tourist spots as indicated by reviews written by inbound tourists. By applying this method to popular tourist spots in Hokkaido, with a special focus on Chinese tourists, the paper provides insights into what aspects of these destinations are most appealing to them.

In this study, we extracted frequent n-gram patterns for each tourist destination from online reviews written by Chinese tourists. These patterns indicate the most frequently mentioned aspects of each destination by Chinese tourists. We then quantified the tourism motivation of each spot by evaluating the extracted n-gram patterns against tourism motivation factors. We defined the combination of strong tourism motivation factors for a tourist spot by statistical threshold as focus points. Our analysis not only focused on the most frequently mentioned keywords or topics by tourists, but also included an examination of the tourism motivation factors related to these spots. This approach helped us to better understand the needs and preferences of inbound tourists.

Chapter 3

Basic Idea and Methodology

This study introduces a novel method that uses reviews of tourist sites as a valuable source of information to identify and analyze the focus of inbound tourists. My approach goes beyond simply examining the frequently mentioned content in these reviews; I delve deeper into the underlying tourism motivations that drive tourists' preferences and choices.

My methodology includes Keyword Extraction and Focus Point Extraction, which are the extraction of feature patterns, quantifying the motivation factors using scoring and PCA analysis, and focus point analysis. The research procedure is shown in Figure 3.1.

To understand Chinese tourists' interests toward Hokkaido tourist spots, I first collected online reviews written by Chinese tourists and extracted keywords from these reviews using various keyword extraction techniques, including TF-IDF (term frequency-inverse document frequency) and TextRank, to identify the frequently mentioned contents in the reviews. I compared different keyword extraction algorithms to determine the most appropriate one for my research. The keywords extracted from each tourist spot revealed different interests and preferences associated with different destinations.

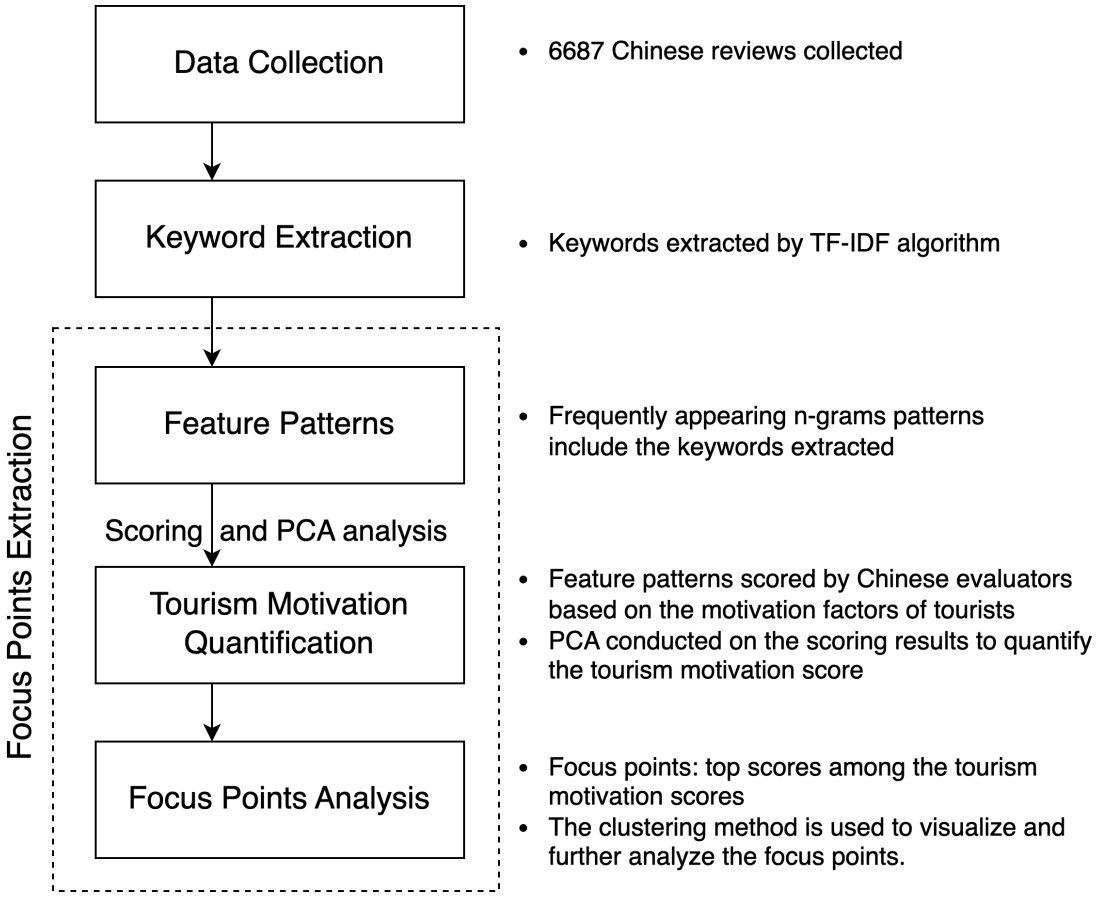

Figure 3.1: Flowchart of the Proposed Method

To delve deeper into tourists' motivations and focus points, I employ a comprehensive framework that quantifies motivation factors through scoring and PCA analysis.

Specifically, I extract high-frequency n-gram patterns from the reviews written by Chinese inbound tourists as feature patterns. These n-gram patterns capture the combinations of keywords and phrases that are commonly used together in the context of reviewing tourist sites. By identifying these patterns, I can uncover specific clusters of information that are of particular interest to Chinese tourists.

Considering various tourism motivation factors such as cultural exploration, natural beauty, and more, I assign scores to each factor based on the presence and relevance of n-gram patterns. The scores reflect the relative importance of each motivation factor in influencing tourists' preferences and choices. I then use PCA

analysis to further analyze and identify the key components that contribute most to tourists' motivations.

Finally, I conduct focus point analysis to identify the main attractions and features of each destination. By clustering the feature patterns, it is possible to gain a comprehensive understanding of the focus points that resonate with Chinese tourists. These focus points provide valuable insights for tourism stakeholders to enhance marketing strategies, improve visitor experiences, and attract more inbound tourists to Hokkaido.

By employing this multi-step methodology, it is not only possible to uncover the specific elements that attract the attention of Chinese tourists, but also provide a deeper understanding of their motivations and preferences. The insights gained from this analysis have implications for marketing strategies, product development, and destination management not only in Hokkaido, but also in other regions and countries. The comprehensive approach presented in this study contributes to a better understanding of inbound tourists' priorities and facilitates the development of tailored tourism experiences worldwide.

Chapter 4

Keyword Extraction

In this chapter, I focus on analyzing the information provided by Chinese tourists and extracting keywords to identify their points of interest. As the largest group of inbound tourists to Japan, the insights of Chinese tourists can provide valuable information for improving tourism experiences and promoting destinations.

To collect relevant data, I collected reviews from a popular travel industry website that caters specifically to Chinese travelers. These reviews were from popular tourist destinations in Hokkaido, a picturesque region in Japan known for its natural beauty and unique attractions. My goal was to extract keywords from these reviews to uncover the key elements and attractions that Chinese tourists prioritize when selecting tourist spots.

To analyze the reviews and identify the most important keywords, I used two different methods: TF-IDF (term frequency-inverse document frequency) and TextRank. The TF-IDF method calculates the importance of a term within a document by considering its frequency in the document and its rarity in the entire collection of documents. TextRank, on the other hand, uses graph-based algorithms to identify important keywords based on the co-occurrence and relationships between words in the text.

By applying keyword extraction methods and analyzing the collected data, my goal is to provide valuable insights to tourism stakeholders, enabling the development of more targeted marketing strategies and improving the overall tourism experience for Chinese visitors.

4.1 Data Set

I selected spots that are popular and have a high number of reviews in Hokkaido from Ctrip (https://www.ctrip.com/), a well-known Chinese travel industry website. The Ctrip website provides a list of recommended tourist spots in Hokkaido, sorted by spot ratings, number of reviews, and other factors using its own algorithm. From this list, I selected ten locations with high rankings and a significant number of reviews. I collected a total of 6687 reviews from these spots by scraping the site and extracting useful information, such as content, rating scores, and dates.

As a data pre-processing step, I first checked the list of reviews for duplicates and removed them. Since the Chinese written language does not use spaces between words, text segmentation was necessary [21]. To segment the words, I used a popular Chinese segmentation tool called Jieba [22]. Since Jieba's dictionary did not include words specific to Hokkaido spots, I manually updated the list to better suit my needs. The added words included the names of tourist spots such as '洞爺湖 (Lake Toya)', '五稜郭 (Goryokaku)', '富良野 (Furano)'. In addition, I removed the stop words [23] from the segmented texts to reduce the amount of redundant words and punctuation.

4.2 Keyword Extraction Method

In order to find the focus points of Chinese tourists, I extracted keywords from the collected reviews using TF-IDF and TextRank.

4.2.1 TF-IDF

TF-IDF [24] or term frequency with inverse document frequency $tf * idf$ is a numerical weighing factor that can be used to extract keywords from texts. In TF-IDF, term frequency $tf(t, d)$ refers to the number of times a term t (word, token) appears in a document d, while inverse document frequency $idf(t, D)$ is the logarithm of the total number of documents $|D|$ in the corpus divided by the number of documents containing the term n_t. Lastly, $tf * idf$ refers to the multiplication of these two as shown in equation (4.1).

$$idf(t, D) = log(\frac{|D|}{n_t}) \tag{4.1}$$

Compared to raw frequencies, TF-IDF helps adjust for the fact that some words appear more often in general as it is offset by the number of documents in the corpus that contain the word, while still increasing proportionally to the number of times a word appears in the document.

4.2.2 TextRank

TextRank algorithm is derived from the classical PageRank algorithm[25]. PageRank is a famous algorithm by Google, which used to measure the importance value of particular web pages. The algorithm works by checking if there is a large number of web pages linking to a certain site, or if some important pages themselves link to the certain sites. These pages will result in a high value. In other words, the score of a page comes from the importance scores of all the pages that are linked to it through iteration calculation.

The idea behind the PageRank algorithm treats each web page as a node in a graph, and the links between web pages represent the edges of the graph. Each web page has an outlink set and an inlink set. In addition, each web page is assigned a PR (PageRank) value, which indicates its importance or authority. The PR value is calculated as the weighted sum of the PR values of all the web pages that link to it.

In essence, the links between web pages can be thought of as votes. To calculate the PR value of a website A, we need to know how many votes other websites have given to website A, which means we need to know its inlinks. However, the weight of each vote depends on the authority of the source site. The higher the PR value of the source site, the more authoritative it is, and site A may receive more votes from it. Another factor to consider is how many sites the source site has voted for. If a site casts too many votes, it dilutes its own authority. Therefore, the number of inbound votes is generally equal to the average of the source site's PR divided by the total number of outbound links [26].

With these ideas in mind and through optimization, the formula for the PageRank algorithm is derived as follow (4.2):

$$PR(V_i) = (1-d) + d * \sum_{j \in ln(V_i)} \frac{1}{|Out(V_j)|} PR(V_j) \tag{4.2}$$

In this context, d represents the damping factor, typically in the range of 0.60 to 0.85. V_i represents a specific web page, and V_j represents web pages that link to V_i (the inlinks of V_i), $PR(V_i)$ represents the PR value of webpage V_i, and $In(V_i)$ represents the set of all inlinks for webpage V_i. $|Out(V_j)|$ represents the total number of outlinks from the source webpage V_j, and $1/|Out(V_j)|$ represents that the PR value of V_j should be equally distributed among all webpages linked by the outlinks.

TextRank is the application of PageRank to the field of natural language processing and is widely used in keywords extraction [27]. It establishes a graph-based relationship between a word and its preceding N words, as well as the following N words (similar to an N-gram grammatical model). The implementation involves setting up a sliding window of length N, where all words within this window are considered adjacent nodes to the word node. Co-occurrence relationships are then used to create edges between any two points. The word graph created by TextRank is undirected. The diagram 4.1 shows a word graph constructed from a document (stop words have been removed, and filtering based on word types has been applied).

The formula for the TextRank algorithm is derived as follow (4.3):

$$WS(V_i) = (1-d) + d * \sum_{V_j \in ln(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \tag{4.3}$$

As can be seen, this formula is only slightly different from PageRank, with the addition of a weight term w_{ji}. Considering that different word pairs may have different co-occurrences, TextRank assigns co-occurrence as the weight of undirected graph edges. Words that share co-occurrence relationships will mutually support each other to become keywords [26].

Compared with TF-IDF which only considers word frequency itself, TextRank also considers the semantic relations between words in the document.

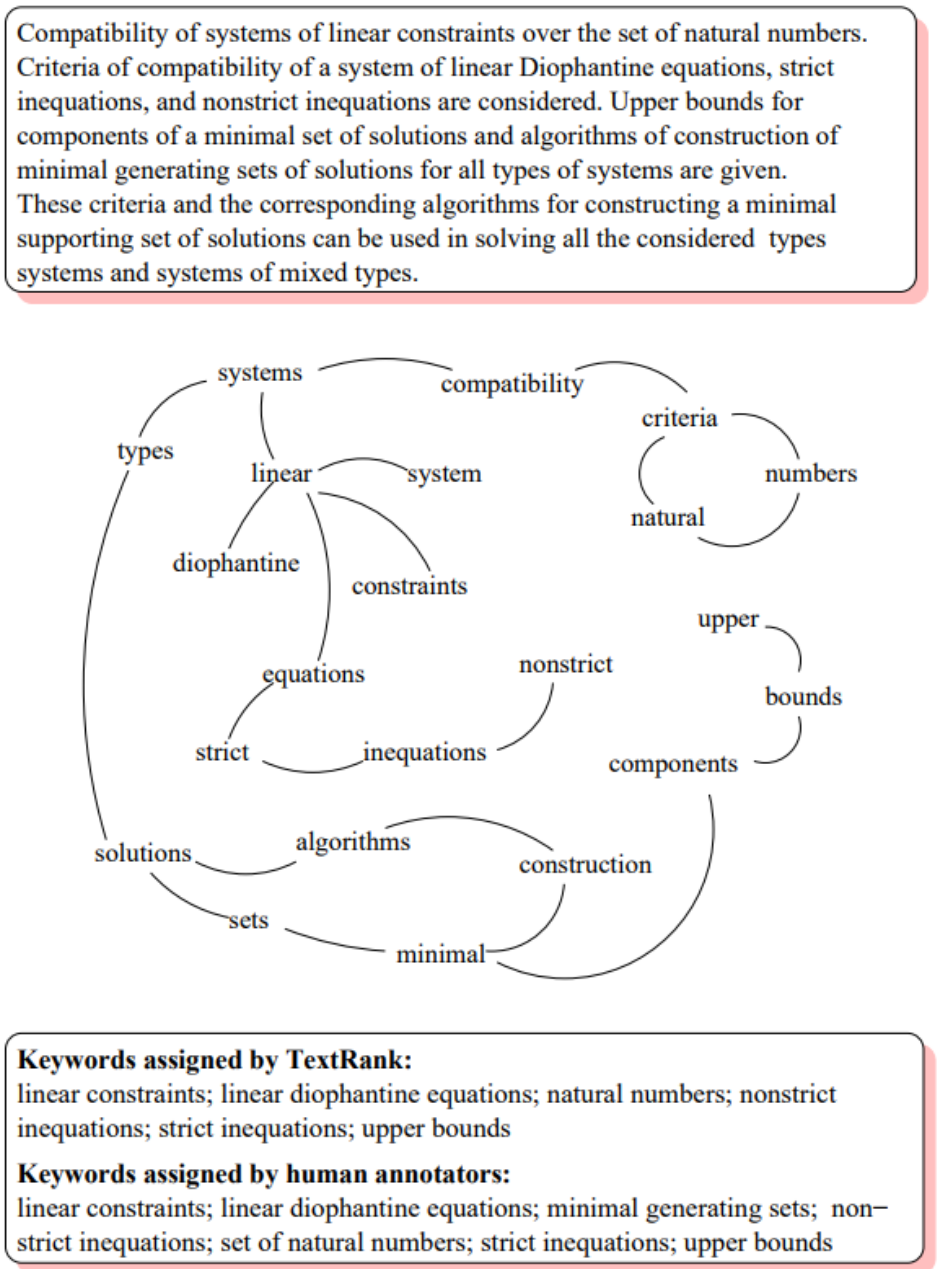

Figure 4.1: Sample graph build for key phrase extraction from an *Inspec* abstract

4.3 Experiments

I used TF-IDF and TextRank methods to extract keywords from reviews. And I used Jieba [22] and Scikit-learn [28]. "Jieba" (Chinese for "to stutter") is a Chinese text segmentation and keyword extraction tool, one of the popular Chinese text data processing tool. Scikit-learn (sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python [29].

I first used Jieba's TF-IDF implementation with its own dictionary to get top 50 keywords from all the collected Hokkaido reviews, which are from 18 most popular spots. The results are sorted by the TF-IDF weight and shown in Table 4.1. In order to delete the keywords that are not specific to a certain spot, I then extract top 50 keywords from reviews of each spot and count how many times the keywords of Hokkaido (Table 4.1) appeared in all of the spots. This is also shown in Table 4.1, sorted by the frequency from least to most. For example, the frequency of the keyword '地獄谷(Hell Valley)' is 1, which means this keyword only appeared in one spot. On the other hand, the frequency of the keyword '日本(Japan)' is 18, which means it appeared in all of the spots.

In Table 4.1, The keywords sorted by frequency clearly show more specific features of each spot on the top, such as '巧克力(chocolate)', '企鵝(penguin)' and '白色恋人(Shiroi Koibito)'. The more general words like '美麗(beautiful)', '特別(special)' and '日本(Japan)', mainly go to the bottom. I highlight the words in the bottom, which appeared in over half of the spots, and removed them from the keyword list.

Some spots like '大通公園(Odori Park)', '函館(Hakodate)' and '美瑛(Biei)' show lots of the distinct features. For example in Table 4.2, spot '大通公園' shows many keywords, such as '電視塔(TV tower)', '冰雪節(ice festival)', '噴泉(fountain)', which described the features well. While spot '小樽音楽盒堂(Otaru Music Box Museum)', '登別地獄谷(Noboribetsu Hell Valley)' and '北海道旧道庁(Former Hokkaido Govt. Office)' show more common words such as '喜歡(like)', '味道(smell)', '大楼(buildings)', etc.

Similarly, I used Scikit-learn's TF-IDF function and Jieba's TextRank function to extract keywords from the collected reviews. In Scikit-learn's case however, I

Table 4.1: Top 50 keywords sorted by TF-IDF (left) and Bottom 50 keywords sorted by frequency (right)

Rank	**Keyword**		**Freq**	**TF-IDF**	**Keyword**		**Freq**
1	北海道	Hokkaido	18	0.1444	地獄谷	Hell Valley	1
2	札幌	Sapporo	16	0.1083	巧克力	chocolate	1
3	公園	park	8	0.0693	洞爺湖	Lake Toya	1
4	音楽盒	music box	3	0.0682	企鵝	penguin	1
5	函館	Hakodate	3	0.0672	神宮	shrine	1
6	夜景	night view	6	0.0576	大学	university	1
7	白色恋人	Shiroi Koibito	2	0.0552	工厂	factory	1
8	運河	canal	3	0.0547	餅乾	biscuits	1
9	地獄谷	Hell Valley	1	0.0406	白色恋人	Shiroi Koibito	2
10	日本	Japan	18	0.0405	動物園	zoo	2
11	温泉	spa	4	0.0398	函館山	Mt. Hakodate	2
12	動物園	zoo	2	0.0381	電視塔	TV Tower	2
13	景点	attractions	17	0.0370	美瑛	Biei	2
14	巧克力	chocolate	1	0.0334	富良野	Furano	2
15	洞爺湖	Lake Toya	1	0.0324	纜車	cable car	2
16	登別	Noboribetsu	3	0.0300	朝市	morning market	2
17	企鵝	penguin	1	0.0292	音楽盒	music box	3
18	地方	local	17	0.0290	函館	Hakodate	3
19	大通	Odori	3	0.0282	運河	canal	3
20	JR	JR	12	0.0279	登別	Noboribetsu	3
21	函館山	Mt. Hakodate	2	0.0276	大通	Odori	3
22	狸小路	Tanukikoji	3	0.0270	狸小路	Tanukikoji	3
23	電視塔	TV Tower	2	0.0267	時間	time	3
24	不錯	not bad	17	0.0261	温泉	spa	4
25	遊客	tourist	14	0.0260	海鮮	seafood	4
26	海鮮	seafood	4	0.0252	旭川	Asahikawa	4
27	景色	view	13	0.0249	浪漫	romantic	5
28	旭川	Asahikawa	4	0.0243	好吃	delicious	5
29	冬天	winter	13	0.0239	夜景	night view	6
30	美瑛	Biei	2	0.0239	参観	visit	6
31	富良野	Furano	2	0.0234	札幌市	Sapporo City	7
32	浪漫	romantic	5	0.0215	酒店	hotel	7
33	札幌市	Sapporo City	7	0.0214	公園	park	8
34	酒店	hotel	7	0.0212	建築	building	8
35	値得	worth	12	0.0209	拍照	take pictures	8
36	神宮	shrine	1	0.0199	晚上	night	8
37	喜歓	like	9	0.0197	喜歓	like	9
38	建築	building	8	0.0197	美麗	**beautiful**	**10**
39	特別	especially	13	0.0195	**JR**	**JR**	**12**
40	参観	visit	6	0.0189	値得	**worth**	**12**
41	大学	university	1	0.0180	景色	**view**	**13**
42	工厂	factory	1	0.0171	冬天	**winter**	**13**
43	拍照	take pictures	8	0.0170	特別	**especially**	**13**
44	餅乾	biscuits	1	0.0166	遊客	**tourist**	**14**
45	美麗	beautiful	10	0.0166	札幌	**Sapporo**	**16**
46	好吃	delicious	5	0.0164	景点	**Attractions**	**17**
47	纜車	cable car	2	0.0162	地方	**local**	**17**
48	朝市	morning market	2	0.0162	不錯	**not bad**	**17**
49	時間	time	3	0.0161	北海道	**Hokkaido**	**18**
50	晚上	night	8	0.0159	日本	**Japan**	**18**

Table 4.2: Examples of the top 10 keywords of some individual spots

Rank	大通公園	**Odori Park**	函館	**Hakodate**	美瑛	**Biei**
1	公園	park	函館	Hakodate	美瑛	Biei
2	大通	Odori	夜景	night view	富良野	Furano
3	電視塔	TV Tower	函館山	Mt. Hakodate	之樹	tree
4	札幌市	Sapporo City	海鮮	seafood	四季	four seasons
5	冰雪節	ice festival	朝市	morning market	自行車	bicycle
6	噴泉	fountain	三大	three biggest	接布	patchwork
7	街心公園	city center park	温泉	spa	花田	flower field
8	冰雕	ice sculpture	倉庫	warehouse	薰衣草	lavender
9	雪雕	snow sculpture	海胆	sea urchin	丘陵	hills
10	雪祭	snow festival	金森	Kanamori	騎行	cycling
Rank	小樽音楽盒堂 **Otaru Music Box Museum**		登別地獄谷 **Noboribetsu Hell Valley**		北海道旧道庁 **Former Hokkaido Government Office**	
1	音楽盒	music box	地獄谷	Hell Valley	紅磚	red brick
2	音楽	music	登別	Noboribetsu	旧道	old road
3	各式各様	various	温泉	spa	建築	building
4	博物館	museum	硫黄	sulfur	巴洛克	Baroque
5	精緻	exquisite	硫黄味	sulfur smell	免費参観	free visit
6	喜歓	like	火山	volcano	歴史	history
7	蒸汽	steam	地獄	hell	札幌市	Sapporo City
8	琳瑯滿目	dazzling	味道	smell	大樓	building
9	建築	building	酒店	hotel	風格	style
10	童話	fairy tale	噴出	erupt	参観	visit

created the IDF dictionary directly from my review dataset.

4.4 Keyword Evaluation and Discussion

In order to compare those different methods of extracting keywords, I evaluate the top 10 keywords of each spot manually as shown in Table 4.3:

Table 4.3: Evaluation score of the top 10 keywords of the "Asahiyama Zoo" spot

Rank	**Keyword**		**Evaluation**
1	動物園	zoo	good
2	企鵝	penguin	good
3	旭川	Asahikawa	good
4	動物	animal	good
5	旭山	Asahiyama	good
6	北極熊	polar bear	good
7	散步	take a walk	good
8	可愛	lovely	good
9	時間	time	bad
10	近距離	close range	acceptable
Evaluation Score			80%

The keywords were labeled by three Chinese native speakers using the following criteria.
(1)If the word shows the distinct features of the spot, or it can otherwise describe the spot very well, it is labeled as ‘good’.
(2)If the word does not show any features of the spot, or has no relation to the spot, it is labeled as ‘bad’.
(3)If there is a possibility that the word could be a feature of the spot, or the annotators are not sure, it is labeled as ‘acceptable’.

And I use the percentage of ‘good’ as the evaluation score, which represents how well they describe the features of each spot. The evaluation scores of TF-IDF (Jieba, Scikit-learn) and TextRank are compared in Table 4.4.

From Table 4.4, I can find out that TF-IDF (Jieba) shows the best result with an average score of 85%, while TextRank (Jieba) shows the worst result with a

Table 4.4: Comparison of evaluation results

Spot name		TF-IDF (Jieba)	TF-IDF (sklearn)	TextRank (Jieba)
旭山動物園	Asahiyama Zoo	80%	80%	60%
札幌電視塔	Sapporo TV Tower	80%	60%	70%
小樽	Otaru	90%	80%	70%
小樽音楽盒堂	Otaru Music Box Museum	70%	70%	50%
小樽運河	Otaru Canal	90%	80%	70%
大通公園	Odori Park	100%	90%	70%
狸小路商店街	Tanukikoji Shopping Street	80%	80%	60%
登別地獄谷	Noboribetsu Hell Valley	70%	70%	70%
洞爺湖	Lake Toya	90%	80%	60%
白色恋人公園	Shiroi Koibito Park	90%	90%	80%
函館	Hakodate	100%	90%	80%
函館山	Mt. Hakodate	80%	80%	50%
函館朝市	Hakodate's morning market	80%	70%	80%
美瑛	Biei	100%	100%	60%
富良野	Furano	90%	80%	60%
北海道旧道庁	Former Hokkaido Govt. Office	70%	70%	50%
北海道神宮	Hokkaido Shrine	90%	80%	70%
北海道大学	Hokkaido University	80%	80%	60%
Average		85%	79%	65%

score of 65%. The differences in the results of Jieba and Scikit are most likely due to the fact that Jieba uses a prebuilt IDF dictionary trained on a huge corpus, while Scikit's dictionary was directly trained on the reviews themselves. With a larger review dataset, the evaluation ranking of the two TF-IDF methods could be the opposite. In the future I am planning to utilize a larger review dataset.

Table 4.5: Top 10 keywords of the "Asahiyama Zoo" spot

Rank	TF-IDF (Jieba)		TF-IDF (sklearn)		TextRank (Jieba)	
1	動物園	zoo	動物園	zoo	動物園	zoo
2	企鵝	penguin	企鵝	penguin	企鵝	penguin
3	旭川	Asahikawa	動物	animal	動物	animal
4	動物	animal	旭川	Asahikawa	散步	take a walk
5	旭山	Asahiyama	旭山	Asahiyama	可愛	lovely
6	北極熊	polar bear	散步	take a walk	北極熊	polar bear
7	散步	take a walk	可愛	lovely	展示	display
8	可愛	lovely	北極熊	polar bear	設計	design
9	海豹	seal	海豹	seal	生活	life
10	近距離	close range	展示	display	日元	JPY

Table 4.5 shows the top 10 keywords of 'Asahiyama Zoo' spot, which are extracted by TF-IDF(Jieba, sklearn) and TextRank(Jieba). One can see that TF-IDF(Jieba) and TF-IDF(sklearn) show more words related to the features of the spot, like '旭川(Asahikawa)', '海豹(seal)' and '近距離(close range)', which contains the information of location, popular animals and the design of zoo. While TextRank shows more general words, like '設計(design)', '生活(life)' and '日元(JPY)', which are not so strong features associated with the spot. Comparing to TextRank, TF-IDF performed better in my data set. It not only considered the frequency of words, but also the unique factor of the spot.

And from the extracted keywords, I can get an idea what Chinese tourists pay attention to. For example in Table 4.5, I can see that '企鵝(penguin)', '散步(take a walk)', '動物(animal)', '可愛(lovely)', '北極熊(polar bear)' and '海豹(seal)' could be the things that attract Chinese tourists to Asahiyama zoo the most. In the next step, I will extract the sentences, which contain the keywords and use them to analyze the topics in order to extract the main focus points.

4.5 Conclusions

In this study, I illustrated the research plan that realizes a new method to extract focus points to attract inbound tourists. I want to build an automatic system to extract the focus points of inbound tourists from online reviews and analyze what attracts inbound tourists, then provide useful tourism information for supporting the post-pandemic tourism industry in Hokkaido.

For the first step, I collected Hokkaido's tourism spot reviews and used TF-IDF and TextRank to extract keywords from the collected reviews. For TF-IDF, I used two different implementations, Jieba and Scikit-learn. To compare the extraction methods, I evaluated the top 10 keywords from each spot by checking how the keywords show the distinct features of each spot. The evaluation results indicate that TF-IDF(Jieba) shows the best result compared to the other two methods. Also, the keywords show the main topics Chinese tourists discuss most often in the reviews, which helps in finding the focus points of Chinese tourists.

For the next step, I plan to extract n-gram patterns from the reviews which contains the keywords, aiming to clarify the focus points of Chinese tourists towards Hokkaido travel spots.

Chapter 5

Focus Point Extraction

In the previous chapter, I presented the process of extracting keywords from Chinese reviews to identify the most frequently mentioned content for each tourist destination. These keywords serve as valuable indicators of Chinese tourists' interests and preferences.

In this chapter, I delve deeper into the analysis by examining the focus points that are closely related to these keywords. First, I extract the frequently occurring n-gram patterns from the reviews, focusing on those that contain the identified keywords. These n-gram patterns provide a more comprehensive understanding of the topics and themes that resonate with Chinese tourists. By scoring these patterns against seven different motivation factors, I quantitatively measure the tourism motivation values associated with each spot.

In addition, I use clustering methods to further visualize and analyze the extracted patterns and the major motivation factors associated with them. Clustering allows for a comprehensive examination of the similarities and differences between the feature patterns, providing valuable insights into the unique characteristics of each spot.

To validate and illustrate the effectiveness of the proposed methodology, I conduct experiments using reviews from Hokkaido spots. And I zoom in on specific spots to provide in-depth discussions on their unique features. Through this detailed analysis, I aim to shed light on the differentiating factors that make each spot stand out and attract the attention of Chinese tourists.

By thoroughly examining the focus points and their relationship with the

extracted keywords and n-gram patterns, this chapter aims to provide valuable insights into Chinese tourists' preferences and interests when selecting tourist spots. These findings can enable tourism stakeholders to make informed decisions, improve marketing strategies, and ultimately enhance the overall tourism experience for Chinese visitors.

5.1 Feature Patterns

In previous research [30], I compared TF-IDF and TextRank [26] for keyword extraction. To compare the extraction methods, I evaluated the top 10 keywords from each spot by checking how the keywords showed the distinctive features of each spot. The evaluation results showed that TF-IDF better captured the keywords of each spot. In addition, I considered a topic modeling approach using LDA [31]. However, while LDA is designed to identify underlying topics in a corpus of text, it can be difficult to interpret the topics that are generated. The topics may be highly abstract or difficult to label, making it difficult to understand the underlying themes. Therefore, TF-IDF was chosen as the keyword extraction method.

I used Jieba's TF-IDF implementation with its own dictionary to obtain the top n keywords from the segmented reviews of each tourist spot. I removed some common words from the keyword list, such as '美麗 (beautiful)', '特別 (special)', and '日本 (Japan)', which appeared in more than half of the spots. The keywords for each spot were sorted in descending order based on their TF-IDF values. As an example, Table 5.1 shows the top 10 keywords ($n = 10$) extracted from the reviews of Asahiyama Zoo. For the experiments, I chose top 5 keywords ($n = 5$) based on the pilot survey.

Table 5.2: n-gram examples.

n Size	n-gram Pattern List	English Translation
1	動物園, 有, 企鵝, 游行, 活動	zoo, has, penguin, walking, event
2	動物園 有, 有 企鵝, 企鵝 游行, 游行 活動	zoo has, has penguin, penguin walking, walking event
3	動物園 有 企鵝, 有 企鵝 游行, 企鵝 游行 活動	zoo has penguin, has penguin walking, penguin walking event

Table 5.1: Top 10 keywords used to describe Asahiyama Zoo.

No.	Keyword	English	TF-IDF Value
1	動物園	zoo	0.5025
2	企鵝	penguin	0.3924
3	旭川	Asahikawa	0.2666
4	動物	animal	0.1827
5	旭山	Asahiyama	0.1719
6	北極熊	polar bear	0.1090
7	散步	walk	0.0986
8	可愛	cute	0.0900
9	北海道	Hokkaido	0.0773
10	遊客	tourist	0.0550

To check what tourists wrote about in relation to the keywords and obtain more complete information about them, I extracted n-gram patterns from the reviews.

n-gram [32] is a method used in natural language processing to represent a sequence of words in a text. It involves breaking down the text into smaller units of n consecutive words called "n-grams". For example, in the sentence "動物園(zoo) 有(has) 企鵝(penguin) 游行(walking) 活動(event)", the n-gram examples are as shown in Table 5.2. The n-gram method is widely used in various applications, such as language modeling, text classification, and information retrieval. It can help to capture important contextual information and improve the performance of these applications.

After generating the n-gram patterns from all reviews of each tourist spot and

filtering for those containing each keyword about the tourist spot, I calculated the frequency of occurrence for each pattern and selected the most highly occurring n-gram patterns as the feature patterns. For example, Table 5.3 shows the feature patterns containing the keyword "zoo" extracted from the reviews of the Asahiyama Zoo. Here, the 3-gram patterns with a frequency of occurrence of 5 or more were identified as feature patterns based on a pilot survey for the experiments. All feature patterns containing any of the top n keywords for a tourist site were extracted to create a feature pattern set for that site.

Table 5.3: Feature patterns containing the keyword "zoo".

No.	**Feature Patterns**	**English**	**Frequency**
1	日本 北端 動物園	Japan north zoo	15
2	旭山 動物園 日本	Asahiyama zoo Japan	12
3	北海道 旭山 動物園	Hokkaido Asahiyama zoo	10
4	東京都 上野 動物園	Tokyo Ueno zoo	10
5	旭川 動物園 企鵝	Asahikawa zoo penguin	8
6	動物園 里 動物	animals in the zoo	7
7	到達 旭山 動物園	arrive at Asahiyama zoo	7
8	日本 有名 動物園	famous zoo in Japan	7
9	参観 人数最多 動物園	the most visited zoo	6
10	前往 旭山 動物園	go to Asahiyama zoo	6

5.2 Quantification of Tourism Motivation

Next, I used the feature pattern sets to quantify tourism motivation. The tourism motivation factors proposed by Hayashi et al. [13] (stimulation, cultural observation, local communication, health recovery, experiencing nature, unexpectedness, and educating oneself) were used for this purpose. The impression of the feature patterns was manually scored for each tourism motivational factor, and the results were quantified as tourism motivational scores of the tourism spots by using PCA.

Six Chinese-speaking evaluators were tasked with scoring the feature patterns on a five-point scale, where 5 indicated the strongest relationship and 1 indicated

no relationship between the pattern and the motivation factor. Tables 5.4 and 5.5 provide the descriptions of the Tourism Motivation Scale and the scoring criteria, respectively. The detailed definition of the motivational factors is listed in the Appendix A, Table A.1.

For example, for the feature pattern "zoo, penguin, walking" of the Asahiyama Zoo spot, the scores could be assigned as shown in Table 5.6. The evaluator determined that the pattern "zoo, penguin, walking" was closely related to health recovery; moderately related to stimulation and experiencing nature; somewhat related to cultural observation, local communication, and educating oneself; and unrelated to unexpectedness.

Table 5.4: Explanation of Tourist Motivation Scale.

The Tourism Motivation Scale	**Explanation**
Stimulation (Stimul.)	Experiencing novelty and change
Cultural observation (Culture)	Interest in the culture of the visited area
Local communication (Local)	Communication with local people
Health recovery (Health)	Recovery from daily fatigue and stress
Experiencing nature (Nature)	Getting into direct contact with nature
Unexpectedness (Unexpect.)	Surprising, unexpected experiences
Educating oneself (Self-exp.)	Improvements/changes in your inner self

Table 5.5: The scoring criteria.

Degree of Judgement	**Score**
Strongly related	5
Closely related	4
Moderately related	3
Somewhat related	2
Unrelated	1

Table 5.6: Examples of scoring.

Feature Pattern	**Stimul.**	**Culture**	**Local**	**Health**	**Nature**	**Unexpect.**	**Self-Exp.**
Zoo Penguin Walking	3	2	2	4	3	1	2

After scoring, I performed PCA on the scoring results. In statistics and data analysis, principal component analysis (PCA) is a technique used to reduce the complexity of high-dimensional data by finding the most important variables, or principal components, that explain most of the variation in the data [33].

The first principal component is the linear combination of variables that explains the most variance in the data. It is the direction in the feature space that captures most of the variation in the data. The importance of the first component is that it represents the dominant source of variation in the data and provides a way to summarize the data along a single dimension. The weights of the first principal component represent the weights or coefficients applied to each variable in the linear combination to obtain the score of the first principal component. In other words, the weights indicate the relative importance of each variable in contributing to the first principal component [34].

The values of the first principal component weights were used as the tourism motivation scores of each site. For example, Figure 5.1 illustrates the tourism motivation scores of Asahiyama Zoo. The horizontal axis represents the seven tourism motivation factors, while the vertical axis represents the values of the first principal component weight. The highest tourism motivation score for Asahiyama Zoo is related to health recovery. This may indicate that visitors feel relaxed when surrounded by animals and natural scenery at the zoo.

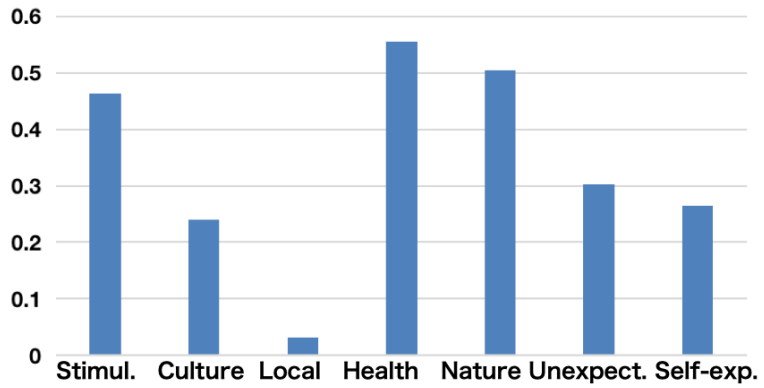

Figure 5.1: Example of tourism motivation score (Asahiyama Zoo).

5.3 Definition of Focus Point

I define the focus points of the destinations as the highest scores among the tourist motivation factors. For example, in Figure 5.1, the focus points of Asahiyama Zoo are health recovery, experiencing nature, and stimulation, which have higher scores compared to other tourist motivation factors. This indicates that when considering this site, tourists focus on aspects of the zoo that allow them to recover from daily stress, engage with nature, and experience novelty and change.

In addition, the term "top scores" here refers to those with large absolute values and is not related to whether they are negative or positive. The sign of the weight value (whether it is positive or negative) only indicates the direction of the relationship between the original variable and the principal component, but the magnitude of the weight value is what determines the strength of this relationship. When a PCA weight is negative, it indicates an inverse relationship between the original variable and the principal component or other variables that have a positive weight [35].

5.4 Clustering Analysis

The PCA results show the main tourist motivational factors of the site. To clarify the specific content related to these motivational factors, I used and compared

the following clustering methods to analyze the feature patterns by clustering them based on their first and second principal component scores.

- K-Means

  K-means clustering is a popular unsupervised machine learning algorithm used to partition a given data set into K clusters, where K is a user-specified number. The algorithm works by iteratively assigning each data point to the nearest centroid, and then recalculating the centroid of each cluster based on the new assignments. The process is repeated until convergence, i.e., when the assignment of data points to clusters no longer changes [36].

- Gaussian Mixture Model (GMM)

  GMM clustering is a probabilistic clustering method that models the distribution of the data as a mixture of Gaussian distributions. In GMM clustering, each data point is assumed to be generated from one of K Gaussian distributions, where K is the number of clusters. The parameters of the Gaussians (mean and covariance) are estimated using the Expectation–Maximization (EM) algorithm [37].

- MeanShift

  MeanShift clustering is a non-parametric clustering method that does not assume an underlying distribution of the data. It works by iteratively shifting the data points toward the mode of their density estimate until convergence. The mode can be interpreted as the center of a cluster, and the final clusters are obtained by assigning data points to the nearest mode [38].

- Spectral Clustering

  Spectral clustering is a powerful clustering method that uses the spectral properties of a similarity matrix to group data points into clusters. It is based on the idea that the eigenvectors and eigenvalues of a similarity matrix contain useful information about the structure of the data and can be used to transform the data into a lower-dimensional space where clusters are more easily separable [39].

The cluster results show the classified groups of feature patterns, which helps in understanding the themes of the focus points. They also reveal the relationships between feature patterns and tourist motivation factors, which helps in understanding the focus points of the spot.

In addition, I can also extract reviews that include the feature patterns representing the focus point as recommended reviews, making it easier for users to view relevant reviews more efficiently.

5.5 Experiments and Results

In this section, I describe the experiments conducted to evaluate the basic performance of the proposed method.

5.5.1 Focus Point Extraction

5.5.1.1 Data Set

In this experiment, I focused on 10 popular spots in Hokkaido. The data set that I collected for this experiment is summarized in Table 5.7.

- Collected reviews

  I collected reviews for each spot from tourism websites, and then removed duplicate reviews, leaving us with 6687 reviews.

- Feature Patterns

  I extracted the top 5 keywords of each spot from the reviews. Then, I extracted frequently occurring 3-gram patterns containing the keywords from the reviews as the feature patterns of each spot. The total number of feature patterns of 10 spots was 395.

- Customer reviews

  I randomly chose 50 reviews from the review list of each spot, for a total of 500 reviews.

Table 5.7: Tourism spots and the numbers of reviews.

Spot Name	Collected Reviews	Feature Patterns	Customer Reviews
旭山動物園 (Asahiyama Zoo)	599	51	50
北海道庁旧本庁舎 (Former Hokkaido Govt. Office)	402	39	50
北海道神宮 (Hokkaido Shrine)	361	34	50
登別地獄谷 (Noboribetsu Hell Valley)	662	45	50
大通公園 (Odori Park)	789	35	50
小樽運河 (Otaru Canal)	1060	58	50
小樽音楽盒堂 (Otaru Music Box Hall)	473	33	50
札幌電視塔 (Sapporo TV Tower)	433	24	50
白色恋人公園 (Shiroi Koibito Park)	1012	33	50
狸小路商店街 (Tanukikoji Shopping Street)	896	43	50
Total	6687	395	500

5.5.1.2 The Focus Points of Hokkaido

First, I applied the proposed method to all the data on the Hokkaido spots that I collected. I extracted feature patterns from all the spots and evaluated each pattern using the motivational factors of tourists by employing six Chinese-speaking university students. Then, I calculated the average scores of all the raters and performed PCA on the scores.

In addition, as a control experiment, I scored all customer reviews using the same evaluators and motivation factors, and then performed PCA.

The PCA results are shown in Figure 5.2. The height of the bars on the graph represents the principal component weight of each tourist motivation factor. Factors with high scores indicate the focus points, which are the main points of interest for the site. PC1, PC2, and PC3 denote the first, second, and third components of the PCA, respectively.

Since PC1 shows the total influence of the tourism motivational factors, it becomes clear that the focus points of Hokkaido for the results of the feature pattern are experiencing nature and stimulation. From PC2, I conclude that experiencing nature has the highest absolute value. Local communication has the

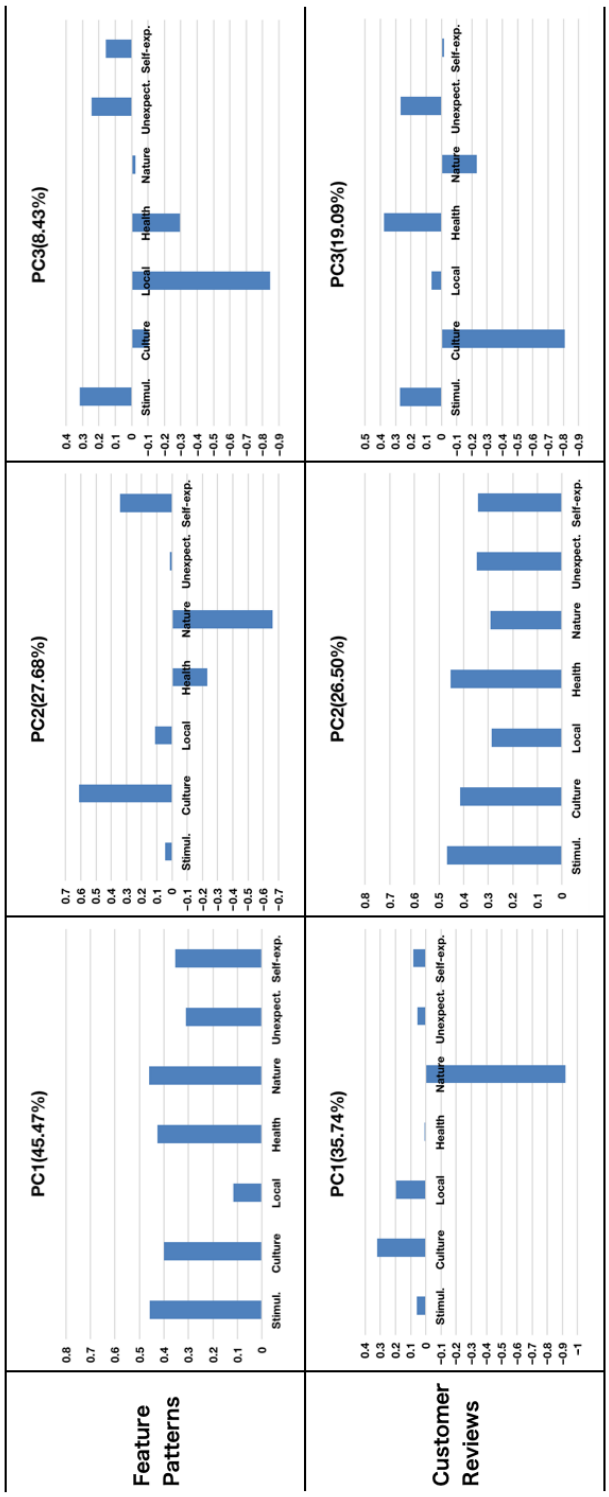

Figure 5.2: The tourism motivation scores of Hokkaido

highest absolute value on PC3, for which the contribution rate is quite low (8.43%).

For customer reviews, the focus is on experiencing nature, with stimulation being the highest (absolute value) on PC2, and cultural observation the highest (absolute value) on PC3. In this experiment, I examine the motivational factors with the highest absolute values. When the PCA weight is negative, it indicates an inverse relationship between the original variable and the principal component or other variables with a positive weight. For example, experiencing nature has a negative relationship with other motivators and principal components on PC1, suggesting that customer reviews that emphasize a connection with nature may place less emphasis on culture and other motivators.

The extracted focus point from both feature patterns and customer reviews is the same, which is experiencing nature. However, the customer reviews have a much higher score for experiencing nature compared to other tourist motivations. This is because there are many descriptions related to nature in the reviews, but, when extracting keywords and feature patterns, the context may not have been effectively considered. In addition, analyzing Hokkaido as a whole makes it difficult to see the unique features of each site. To address this, I also extracted the focus points for each tourist spot individually.

In addition, I created a word cloud using the reviews of ten Hokkaido tourist spots, as shown in Figure 5.3. In the word cloud, one can observe many frequently appearing words related to nature, such as “park”, “snow”, “canal”, and so on, which is consistent with our analysis.

Figure 5.3: Word cloud of the reviews of Hokkaido.

5.5.1.3 Spot-Specific Focus Points

In order to find the spot-specific focus points, I evaluated the feature patterns and customer reviews of each tourist spot and conducted PCA. The PCA results of 10 spots are shown in Appendix A, Table A.2.

I took three different types of spots as an example. Figure 5.4 shows the first component weights of the PCA results, which represent the tourism motivation scores of the spots.

Figure 5.4 (feature patterns) shows that the focus points of Asahiyama Zoo are health recovery, experiencing nature, and stimulation. The focus points of Tanukikoji Shopping Street are cultural observation, local communication, and unexpectedness. The Former Hokkaido Govt. Office has focus points centered on cultural observation and educating oneself.

The results of customer reviews are similar to those of the feature patterns, but there is a significant difference in the case of the Former Hokkaido Government Office location. The results of customer reviews show a very high score for experiencing nature, which could be due to the large amount of content related to the natural scenery around the attraction. The keyword and feature pattern extraction method are mainly based on the frequency of words, so it may not be able to extract the context or various expressions concerning nature effectively.

I also created a word cloud using the reviews from each spot—for example,

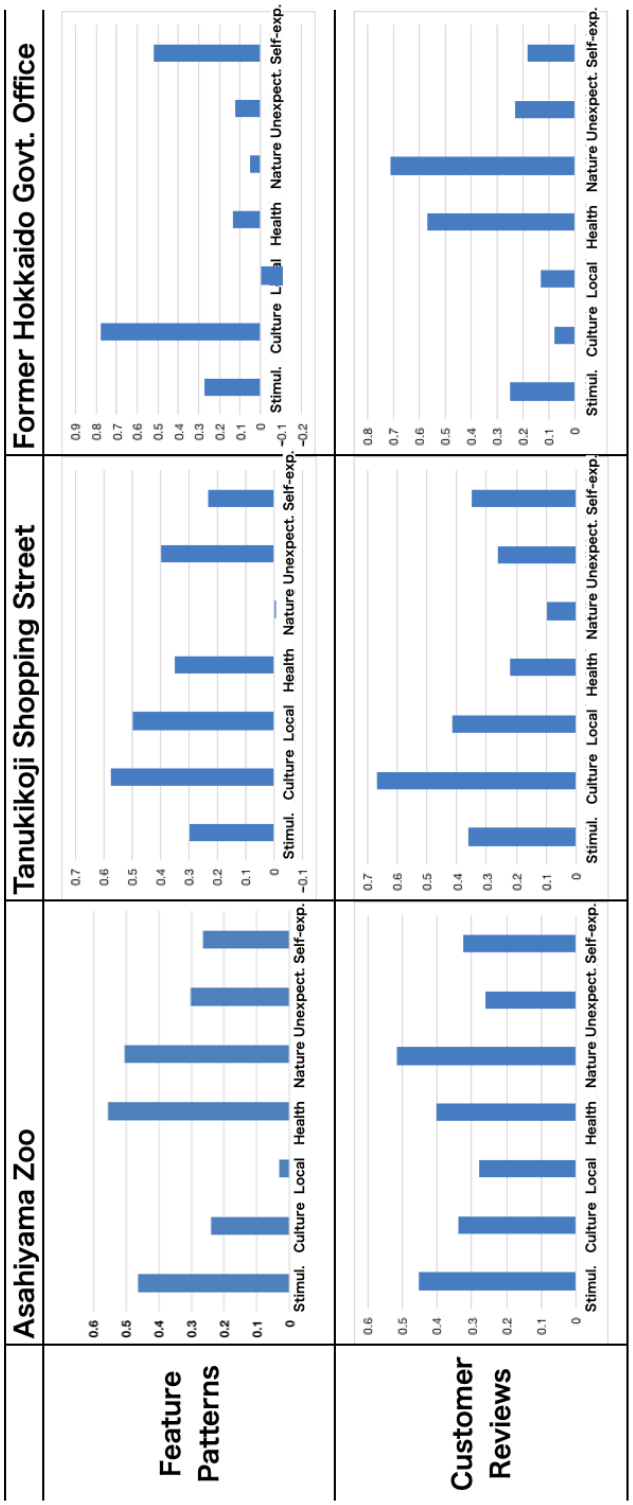

Figure 5.4: The tourism motivation scores of the spots

Asahiyama Zoo, which is shown in Figure 5.5.

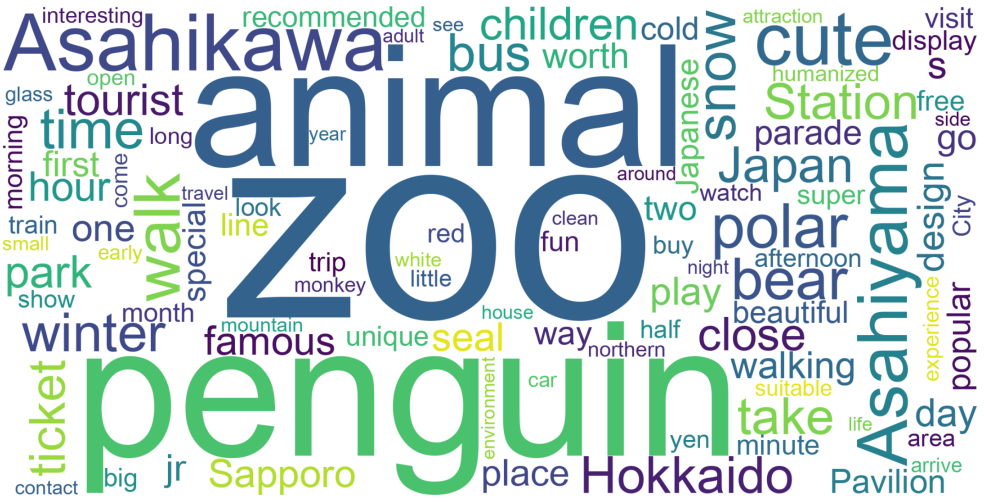

Figure 5.5: Word cloud of the reviews of Asahiyama Zoo.

The word cloud shows many important words related to animals and nature, which is consistent with my focus point analysis results.

5.5.1.4 Verification

To verify how well the feature patterns covered the features of the spot, I compared the contribution rates of the first principal component.

The contribution rate of the principal component indicates how much the principal component covers the variance of the entire data. In this study, it is assumed that a higher contribution rate explains more of the data, thus more clearly representing the characteristics of the tourist spots.

Table 5.8 shows the contribution rates of the first principal component for the tourist motivation scores of each spot and the average for all spots. It shows that the feature pattern has a higher contribution rate than the customer reviews. This indicates that the feature pattern represents more features of the tourist spots and, by consequence, that the extracted feature patterns are considered to be effective.

Table 5.8: The contribution rates of the first principal component.

Spot Name	Feature Patterns	Customer Review
旭山動物園 (Asahiyama Zoo)	0.79	0.69
北海道庁旧本庁舎 (Former Hokkaido Govt. Office)	0.73	0.46
北海道神宮 (Hokkaido Shrine)	0.55	0.45
登別地獄谷 (Noboribetsu Hell Valley)	0.62	0.32
大通公園 (Odori Park)	0.75	0.44
小樽運河 (Otaru Canal)	0.49	0.59
小樽音楽盒堂 (Otaru Music Box Hall)	0.56	0.43
札幌電視塔 (Sapporo TV Tower)	0.60	0.41
白色恋人公園 (Shiroi Koibito Park)	0.47	0.39
狸小路商店街 (Tanukikoji Shopping Street)	0.62	0.60
Average	0.62	0.48

5.5.2 Feature Pattern Clustering

To identify specific content related to the focus points, I clustered the feature patterns based on their first and second principal component scores using clustering methods. Specifically, I used the first and second principal components as the features on which to base the clustering algorithm. I applied several clustering methods to the feature patterns of five spots and calculated the accuracy based on the rate of correctly clustered feature patterns. I did this by manually evaluating each feature pattern and considering whether it matched the assigned cluster or not. I then calculated the average accuracy over five tourism spots. Table 5.9 shows the results, which indicate that spectral clustering is the best method, while GMM gives the lowest accuracy in this experiment.

The discussion of the focus points and related feature patterns obtained from the spectral clustering results can be found in Section 5.6.2, while the extracted recommended reviews for the focus points are discussed in Section 5.6.3.

Table 5.9: The accuracy of different clustering methods.

Spot Name	GMM	MeanShift	K-Means	Spectral Clustering
旭山動物園 (Asahiyama Zoo)	0.94	0.92	0.94	0.96
北海道庁旧本庁舎 (Former Hokkaido Govt. Office)	0.61	0.71	0.84	0.74
北海道神宮 (Hokkaido Shrine)	0.55	0.85	0.97	1.00
白色恋人公園 (Shiroi Koibito Park)	0.48	0.66	0.87	0.93
狸小路商店街 (Tanukikoji Shopping Street)	0.41	-	0.74	0.97
Average	0.59	0.78	0.87	0.92

5.6 Discussion

5.6.1 Feature Patterns

The tourist motivation scores of the feature patterns represent the characteristics and focus points of each tourist spot. I analyzed the focus points from the PCA results (Figure 5.4) and customer reviews (Table 5.10; the original Chinese text of these reviews is presented in Appendix A, Table A.1). For example, health restoration, experiencing nature, and stimulation are the focus points in the case of Asahiyama Zoo. This is interpreted as tourists recognizing the healing effect on the body and mind achieved through engaging with nature and animals in the zoo. Cultural observation, local communication, and unexpectedness are the focus points of Tanukikoji Shopping Street. Tanukikoji Shopping Street is lined with a variety of shops, offering Japanese cosmetics as well as souvenirs and Japanese restaurants, while also staging live performances that sometimes take place on the street. Tourists can experience Japanese culture and meet locals and tourists from other regions.

The focus points of the Former Hokkaido Govt. Office are cultural observation and educating oneself. The Former Hokkaido Govt. Office houses archives, a museum, and an exhibition room for local specialties, where visitors can learn about Hokkaido's history, nature, and culture. In addition, the building itself is a beautiful red brick American Neo-Baroque-style building (https://www.japan.travel/en/spot/1938/) and it is an attractive spot to take pictures by and share them on SNS.

The knowledge of these focus points can help tourism businesses and organizations to improve their marketing strategies and develop better products and services that align with tourists' interests. By understanding which aspects of destinations are most appealing to visitors, tourism businesses can create targeted marketing campaigns that showcase these features and attract more visitors. In addition, tourism organizations can use this information to improve the visitor experience by optimizing destinations based on the identified focus points. For example, if the focus points include food, cultural experiences, and outdoor activities, businesses can work to enhance these aspects by offering more food options, cultural events, and adventure activities. Ultimately, improving the visitor experience can lead to higher visitor satisfaction, more repeated visits, and positive word-of-mouth recommendations, which can attract yet more tourists to the destination.

Table 5.10: Examples of the customer reviews.

Spot	Customer Reviews
Asahiyama Zoo	1. The most popular activity at Asahiyama Zoo in winter is the penguin walk. Chubby king penguins wiggle and sway on the snow, and you get to observe the baby penguins, known as "kiwis", which adds even more fun to the experience. The penguin walk takes place at 11am and 2pm, so I recommend going to see the penguins first thing in the morning, after the zoo opens. Another popular attraction at Asahiyama Zoo is the red panda enclosure. They are so cute and cuddly. Every day there are different animal feeding events at Asahiyama Zoo, where you can see polar bears, red pandas, seals and more animals eating. **Every time you see the cute little animals, you feel that time goes by especially fast and you forget about the things that bother you in life. That's why visiting the zoo is so healing.** 2. I love to go to Asahiyama Zoo to see the penguins walking in winter. They are so cute and wiggly! **Watching them walk and wiggle in the snow really makes you happy and you can forget all your worries**. Remember to go to Hokkaido's Asahiyama Zoo if you're traveling in winter. There's a penguin walk every day at 11am and 2:30pm! It's quite crowded, so remember to go early and take your place!
Tanukikoji Shopping Street	1. If you come to Japan to buy **Japanese specialties, such as cosmetics or Japanese souvenirs**, you can spend at least a day shopping on Tanukikoji Shopping Street. 2. This is not the first time I have seen a scene like this. These days as long as walking through the path always see **a group of such young people, boys and girls, Japanese and Europeans and Americans, playing various musical instruments, singing in harmony or beautiful songs**. Not peddling and singing, just out of love for music and youthful energy. **Their companions and friends are directly in front of them sitting on the ground quietly listening, or on the side of the pace of the beat to hum along. Passers-by, whether they love music or not, will be infected by such a youthful picture, and stop to watch.**
Former Hokkaido Govt. Office	1. It is a government agency, but the red brick exterior looks quite beautiful, and I personally still like it very much, like some of the buildings inside the Sherlock Holmes movies, and **it is very nice to take pictures, especially when it is snowing.** 2. The former site of the Hokkaido Office is a European-style building, and admission is free. Inside, **you can learn about the history of Hokkaido's development**, and there is an introduction booklet in Chinese. There is a commemorative stamp at the entrance.

The tourist motivation scores of the feature patterns and customer reviews are similar to some extent. For example, the graphs of Asahiyama Zoo for the feature pattern and customer review show that the scores for health recovery, experiencing nature, and stimulation are high, while the score for local communication is low.

Customer reviews reflect more comprehensive and authentic content. The goal of this study is to reduce the amount of data processing by extracting keywords

and feature phrases from the reviews, while closely approaching the true travel motivations expressed in customer reviews. Therefore, the expected outcome is that for the results from the feature phrases will closely resemble those from the customer reviews.

However, in the case of some tourist spots, I can observe large differences. One such example is the Former Hokkaido Govt. Office, for which the graph of the feature pattern shows higher scores for cultural observation and educating oneself, while the customer review shows higher scores for health recovery and experiencing nature. Table 5.11 shows that the reviews describe not only the building itself and its history, but also the Odori Park next to it, the snowy landscape, and the nearby river. These contents are related to health recovery and experiencing nature, while the feature patterns were related only to buildings and history. The descriptions of nature are in the context of reviews, and the keywords related to experiencing nature (e.g., park, white snow, landscape) were not included in the top 10 keywords list that I extracted for each tourist spot. Thus, there is a possibility that the feature patterns could not demonstrate the focus points similarly to the reviews.

Table 5.11: Examples of customer reviews and feature patterns for the Former Hokkaido Govt. Office.

Customer Review	**Feature Pattern**
...I went to Odori Park and played in the snow, then I walked to the Former Hokkaido Govt. Office... I was taking photos in front of the red office building, and together with the snowy scenery, it gives a very exotic feeling. There was a river nearby, but it was too cold to go around it...	Hokkaido, documents, museums documents, museum, Hokkaido introduction, Hokkaido, history Hokkaido, development, history red bricks, green tiles, baroque Hokkaido, government, building Hokkaido, reclamation, period style, architecture, Hokkaido This, red brick, green tile Hokkaido, building, former office

5.6.2 Clustering Result

In a comparison performed with several clustering methods on the five tourist spots, the spectral clustering method gave the best results. Here, I discuss the clustering results obtained using the spectral clustering method on the example of two tourist spots, the "Asahiyama Zoo" and the "Former Hokkaido Govt. Office".

5.6.2.1 Asahiyama Zoo

The clustering result for the "Asahiyama Zoo" spot is shown in Figure 5.6. The horizontal and vertical axes represent the values of the first and second principal components of the feature patterns, respectively. The direction of the weight vectors (tourist motivations) indicates how much of the information is carried by the first principal component (x-axis) and the second principal component (y-axis). The length of the vector indicates the strength of the relationship. The clustering method formed three distinct clusters from the feature patterns, which could be interpreted as "location", "zoo", and "animal". The "location" cluster contains patterns related to the location of the zoo, such as "Hokkaido, Asahikawa, city". The "zoo" cluster mainly shows descriptive information about the zoo, such as "Japan, famous, zoo". The "animal" cluster contains patterns about the most famous animals in the zoo, or some special activities and events related to them, such as "penguin, walk, activity".

Moreover, the feature patterns near the tourist motivation factors point to additional factors related to the tourist motivation. In particular, the feature patterns near the focus point indicate which aspects tourists mentioned the most regarding the focus point. In Figure 5.6, stimulation, health recovery, and experiencing nature have a higher score compared to other motivational factors on PC1 (the x-component is longer), which can then be considered as the focus points of the spot. In addition, many feature patterns in the animal cluster are close to these focus points, as they are placed further along the x-axis. This means that the penguin walking activity, close contact with animals, and the free and energetic nature of the animals make Chinese tourists feel excited, relaxed, and connected to nature.

With this approach, I can analyze the needs of visitors based on the feature patterns around the focus points and improve the visitor experience. For Asahiyama

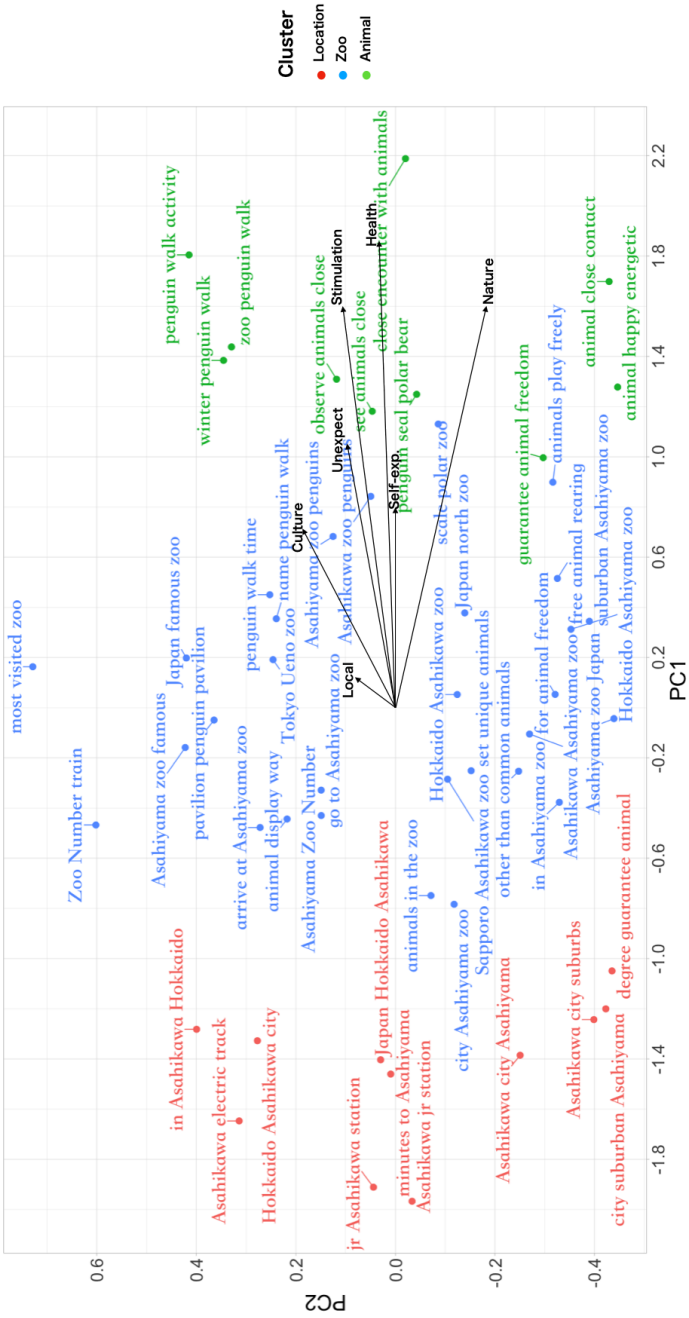

Figure 5.6: Spectral clustering result for “Asahiyama Zoo”.

Zoo, in addition to the penguin walk activity, the close contact with animals and the interesting design of the zoo might also be very attractive to Chinese tourists. The reason may be that zoos in China are designed differently from those in Japan.

5.6.2.2 Former Hokkaido Govt. Office

The clustering result of the spot "Former Hokkaido Govt. Office" is shown in Figure 5.7. The clusters are "history", "location", "building", and "gov. office".

The focus points are cultural observation and educating oneself. The feature patterns from the "history" cluster have a higher score on PC1, which is strongly influenced by these focus points. This may indicate that Chinese tourists who visit this spot are interested in culture and history and are looking to expand their knowledge.

Additionally, the unexpectedness and stimulation factors have a higher score on PC2. Many feature patterns in the "building" cluster also have a high score on PC2, which suggests that the baroque-style building itself is so beautiful that it surprises many Chinese tourists. This is likely due to the fact that this style of building is not very common in China.

By using this method of analysis, it is possible to determine the possibility to provide Chinese tourists with more historical materials in Chinese to help them to better understand the history of Hokkaido and gain more cultural knowledge. Additionally, if the information about baroque-style buildings is emphasized more on websites or other media, it could encourage Chinese tourists to visit and take pictures.

5.6.3 Recommended Reviews

Based on the feature pattern clustering results, one can recommend reviews to users that contain feature patterns that are highly related to the focus points of tourist spots. This approach is more efficient than simply reading all customer reviews to find the focus points of tourist spots.

For example, in Figure 5.6, the focus points of Asahiyama Zoo are stimulation, health recovery, and experiencing nature, and the feature patterns that have high scores on these factors are "penguin walking activity", "observe animal close", "animal

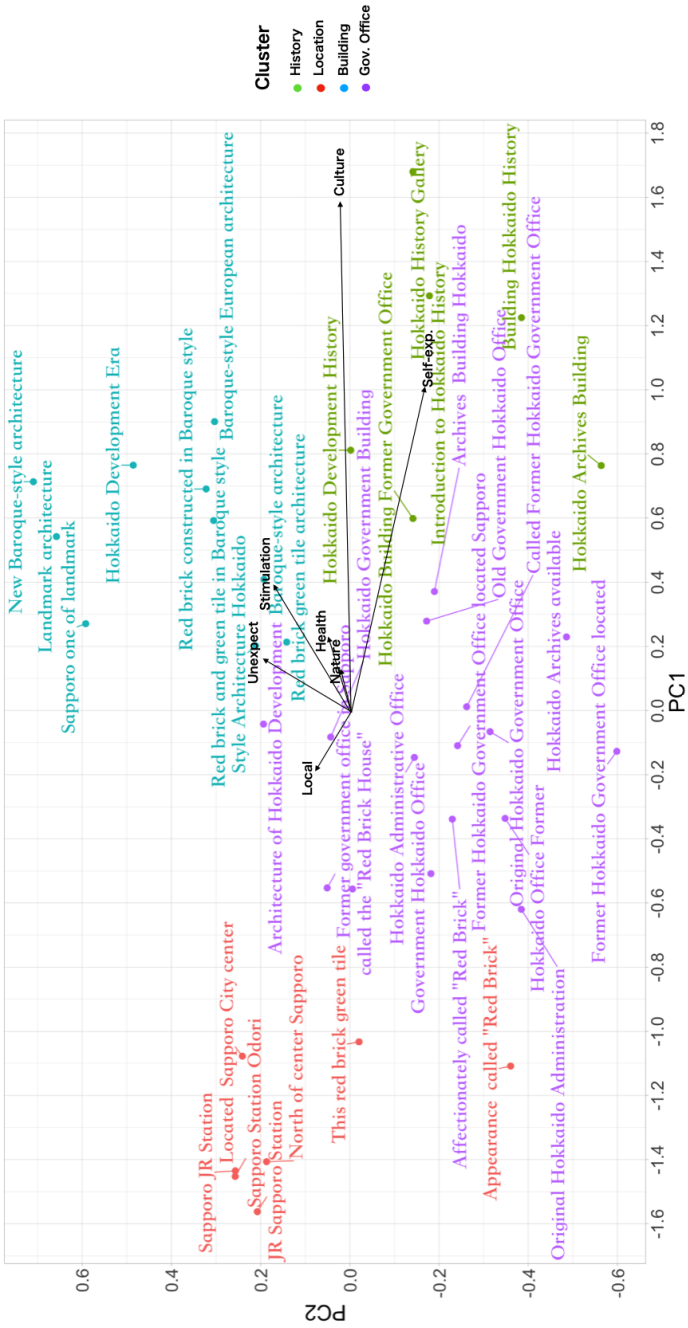

Figure 5.7: Spectral clustering result for "Former Hokkaido Govt. Office".

happy energetic", and so on. In this way, one can automatically extract the reviews that contain these feature patterns as recommended reviews. Table 5.12 shows some examples of feature patterns, recommended reviews, and customer reviews. One can see that compared to the customer reviews, the recommended reviews are more related to the feature patterns and contain the focus points of this tourist spot.

5.7 Conclusions

In this study, I propose a method for extracting and quantifying tourists' points of interest based on the analysis of tourism reviews for the purpose of capturing and evaluating the attractiveness of tourist spots as seen by inbound tourists.

An evaluation experiment was conducted on tourist sites in Hokkaido, Japan. I confirmed that the proposed method can sufficiently capture the characteristics of tourist spots. The results of the customer reviews are different from those of the feature patterns because they include contexts other than the words contained in the patterns. The focus points can be seen most clearly when the feature patterns are used instead of a random set of customer reviews. Moreover, the recommended reviews seem to be a better choice than customer reviews if one is looking for information about a location and wishes to focus on the most important information.

I clustered the feature patterns from each spot to further clarify the specific content of the focus points. The clustered groups of feature patterns clearly sum up the information related to the focus points, which can help to improve the travel experience and attract more inbound tourists.

In the future, I plan to calculate the inter-rater agreement to validate the scoring process. I will also try to reduce the bias in the scoring process by comparing the obtained scores with those assigned by expert raters.

In addition, I will analyze the second and third principal components of focus points to gain a more comprehensive understanding of the data.

It is also necessary to improve the performance of feature pattern extraction. The feature patterns extracted in this experiment represented the characteristics of each tourist site to some extent. However, there were some feature patterns that

Table 5.12: Examples of feature patterns, recommended reviews, and customer reviews.

Feature Patterns	**Recommended Reviews**	**Customer Reviews**
penguin walk activity	I can't remember how long it's been since I've been to the zoo, but once in a while, I get as excited as a little kid to see these little cuties. Apart from the signature penguins, the zoo also has a number of animals not to be missed, including polar bears, mooses, tigers, wolves, and a snow owl that blends in with the snow. Not to be missed are the two daily **penguin walk activities**, which are super cute!	1. The small animals are so cute and adorable, I would love to come back again if I have the chance. 2. Asahiyama Zoo's best feature is its humane design, which maximizes the animals' ability to be kept in a relatively free-range environment. Through clever displays, visitors can observe the animals up close. In the winter, the zoo features a special event called the Penguin Parade, where visitors have two opportunities each day to follow the cute penguins as they walk around. 3. A must-visit spot in Asahikawa, Hokkaido for families with children. The animal zoo is very user-friendly and has a wide variety of animals. My child had a great time playing there.
animal happy energetic	Seals, polar bears, penguins, and other animals from colder regions can be seen here. But what is really impressive is that the zoo promotes presenting the animals as they are, allowing visitors to see the **animals as happy and energetic** as they are.	
observe animal close	The animals can be viewed directly, from the top, the bottom and directly in front of you, so you can get to enjoy their appearance from all sides. It's really close up, except for the monkeys and tigers, which have enclosures, so you can basically **observe the animals up close**. It's worth a visit!	

could not be obtained only by computing the occurrence frequency of keywords and n-gram patterns. It is necessary to extract more comprehensive content by using other information extraction methods.

To extract information from other sources and platforms, I plan to apply image recognition and sentiment analysis. Analyzing images of tourist sites could provide valuable insights into the focus of the destination. For example, identifying the most frequently photographed landmarks or attractions can give an indication of what tourists find most interesting. Sentiment analysis involves using natural language processing techniques to determine the sentiment expressed in a text, whether positive, negative, or neutral. This can help to identify aspects of a tourist destination that tourists like or dislike. This method could be applied to social media posts and travel blogs, for example. In addition, analyzing social media platforms such as Instagram or Twitter can provide a wealth of information about tourist behavior and preferences. For example, analyzing hashtags related to a particular destination can help to identify the most popular tourist spots or activities.

The next step is to expand the range of tourist destinations to be evaluated. The proposed method of identifying the focus points of tourist spots has the potential to be applied to a wider range of languages and tourist destinations beyond the scope of the current study. While this study focused on reviews written in Chinese and popular tourist spots in Hokkaido, the methodology can be adapted to analyze reviews written in other languages and for tourist spots in other regions or countries.

In the sample obtained in the course of this research, there were multiple tourist spots that focused on experiencing nature, which is a common denominator of tourist attractions in Hokkaido. The method can be extended to analyze different types of tourist spots beyond the scope of this study, such as cultural sites or historical landmarks. The method could also be applied to different types of reviews, such as social media posts, travel blogs, or online forums.

In addition, I plan to compare the elements of interest and the image of destinations for each tourist site based on the reviews of Japanese and Chinese tourists. Extending the application of the proposed method to a wider range of languages, destinations, and types of reviews and tourist spots could provide a more comprehensive understanding of which aspects of tourist destinations are

most appealing to inbound tourists around the world.

Chapter 6

Discussions

6.1 Focus Points

The keywords and n-gram patterns extracted from customer reviews provide valuable insights into the interests and preferences associated with each tourist destination. These extracted patterns serve as indicators of the specific topics and aspects that attract the attention of visitors. By analyzing these patterns, I gain a deeper understanding of the focal points that shape tourists' perceptions and experiences.

In addition to the extracted patterns, the application of Principal Component Analysis allows to quantify the scores for different motivation factors associated with each site. The higher the scores, the greater the influence of these factors on the overall perception of the spot. These factors can be considered as the focus points that contribute significantly to the appeal and attractiveness of the target destination.

By combining keyword and n-gram pattern extraction with PCA results, I can identify the key elements that shape tourists' interests and motivations. These insights can be used by tourism businesses and organizations to refine their marketing strategies, develop targeted promotional materials, and tailor their offerings to align with the focus points of each spot. By aligning their efforts with the identified points of interest, businesses can improve the overall visitor experience and attract a wider audience.

It is important to note that the extracted feature patterns and resulting mo-

tivation factor scores are derived from the analysis of customer reviews. While these patterns and scores provide valuable information, it is critical to consider the context and nuances of the reviews themselves. Customer reviews provide a more complete understanding of tourists' perceptions, experiences, and feedback, and should be considered alongside the extracted patterns and scores for a holistic evaluation of tourism destinations.

In conclusion, the analysis of keywords, n-gram patterns, and the application of PCA contribute to a deeper understanding of the interests, motivations, and focus points associated with each tourist destination. This knowledge can guide tourism companies and organizations in developing targeted strategies, refining their offerings, and delivering exceptional experiences that resonate with the preferences of their target audiences. By incorporating these insights into their decision-making processes, tourism stakeholders can foster greater visitor satisfaction and create more memorable and rewarding experiences for tourists.

6.2 Clustering Analysis

6.2.1 Clustering Result

The clustering results of the feature patterns provide valuable insights into the different themes and characteristics associated with each place. For example, in Figure 5.6, which shows the "Asahiyama Zoo", the animal cluster highlights activities such as the penguin walk, the presence of seals and polar bears, and the opportunity to observe animals up close. The zoo cluster suggests that it is a well-known and frequently visited zoo, with penguins being a popular attraction. In addition, the display and arrangement of the animals is unique and well designed. The location cluster indicates that the zoo is located in Asahikawa, Hokkaido, possibly near the JR Asahikawa station or requiring transportation from that station.

In addition, the feature patterns near the tourist motivational factors reveal aspects relevant to tourist motivations. In particular, the feature patterns near the focus point, which are characterized by higher scores of tourist motivation factors, highlight the aspects that are frequently mentioned in relation to the spot.

Referring to Figure 5.6, the motivation factors of stimulation, health, and nature have higher scores compared to other factors, indicating that they serve as focus points of the place. And most of the feature patterns in the Animal cluster are closely aligned with the focus points. This suggests that activities such as the penguin walk, close observation of animals, and the vibrant and natural behavior of animals evoke feelings of excitement, relaxation, and a strong connection to nature in Chinese tourists.

By analyzing the feature patterns near the focus points, one can improve the visitor experience and gain insights into visitor needs. In the case of Asahiyama Zoo, in addition to the popular penguin walk activity, the close interaction with animals and the intriguing zoo design are likely to be highly attractive to Chinese tourists. This may be due to differences in zoo design and animal viewing experiences in China, where close interaction with animals may be less common.

By understanding these aspects, tourism stakeholders can improve the overall visitor experience, tailor their offerings, and cater to the specific preferences and desires of Chinese tourists.

6.2.2 Recommended Reviews

The clustering results of the feature patterns provide a practical approach for recommending reviews to users that are closely aligned with the focus of tourist spots. This method proves to be more efficient than manually reading through all customer reviews to identify the specific aspects that make a spot stand out.

This approach has several benefits. First, it saves users time and effort by directing them to reviews that specifically address the aspects they are interested in. By focusing on the feature patterns associated with the focus points, users can gain a better understanding of what makes a particular spot appealing. Second, it improves the quality of the review recommendation system because the recommended reviews are more likely to contain relevant and useful information. This improves the overall user experience and helps potential visitors make more informed decisions about where to visit.

In addition, this method has implications for businesses and tourism stakeholders. By analyzing feature patterns and identifying the focus of tourist destinations, businesses can gain insight into what aspects of their offerings are most appealing

to visitors. This information can be used to refine marketing strategies, develop targeted advertising campaigns, and improve the overall visitor experience. For example, in the case of Asahiyama Zoo, the focus on activities such as the penguin walking experience and up-close animal viewing can be highlighted in promotional materials to attract tourists who are specifically interested in these features.

In conclusion, the use of feature pattern clustering enables the recommendation of reviews that are highly relevant to the focus points of tourist spots. This approach provides a more efficient way to extract valuable information from customer reviews and offers users a targeted selection of reviews that address their specific interests. Integrating this method into review recommendation systems can improve the user experience, facilitate decision making for potential visitors, and provide valuable insights for businesses in the tourism industry.

6.3 Implications

The implications of this research are significant for the tourism industry in Japan and beyond. By identifying the focus points of tourist destinations, tourism stakeholders can gain valuable insights into which aspects of their destinations are most appealing to inbound tourists. This knowledge can be used to improve marketing strategies by focusing promotional efforts on specific attractions, activities, or cultural experiences that resonate with tourists. By effectively highlighting and promoting these unique features, tourism stakeholders can enhance the overall visitor experience and differentiate their destination from others. This, in turn, can help to attract more tourists and drive economic growth in the tourism sector.

Furthermore, the methodology proposed in this study has the potential for broad application across different languages and countries, providing a global approach to tourism analysis. By replicating the research methodology in other contexts, researchers and tourism professionals can gain a better understanding of what motivates tourists from different regions and cultures to visit specific tourist attractions. This cross-cultural analysis can lead to the development of more tailored tourism products and services that satisfy the specific preferences and interests of different target markets. For example, by identifying the focus points of Chinese tourists visiting Hokkaido, the research results can be used to

create customized tour packages, cultural experiences, or specialized services that cater specifically to this market segment.

In the context of the COVID-19 pandemic, this research also has implications regarding rebuilding and preparing for future emergencies in the tourism industry. By analyzing the differences in emphasis before and after the pandemic, tourism stakeholders can identify changes in tourist preferences and adjust their strategies accordingly. For example, if there is a shift in focus points from crowded attractions to outdoor or nature-based experiences, destinations can prioritize the development and promotion of such offerings. This adaptability and responsiveness to changing tourist preferences could be crucial in effectively managing crises and ensuring the resilience of the tourism industry in the face of unexpected challenges.

In addition, this research has implications for sustainable development in the tourism industry. By identifying the specific elements that are most attractive to Chinese tourists visiting Hokkaido, tourism stakeholders can align their efforts with sustainable tourism principles. For example, if natural landscapes and wildlife encounters are found to be major focus points, destinations can prioritize conservation efforts, implement responsible tourism practices, and engage in community-based initiatives that preserve the region's natural beauty and biodiversity. In addition, by promoting cultural heritage and local traditions that appeal to Chinese tourists, tourism stakeholders can support the social well-being of local communities, foster cultural exchange, and generate economic benefits that contribute to the sustainable development of the region.

The social implications of my research are as follows:

- Improved tourist experiences

  By understanding the focus points and preferences of tourists through online reviews, businesses can improve their offerings and provide more tailored and satisfying experiences for visitors. This can lead to increased satisfaction and positive word-of-mouth, benefiting both tourists and the local community.

- Cultural exchange and understanding

  Analyzing reviews from tourists from different regions and cultures can provide a better understanding of their motivations and preferences. This can promote cultural exchange and mutual understanding between tourists

and local communities, enriching the travel experience and fostering an intercultural dialogue.

- Destination development

  Identifying areas for improvement based on online reviews can help destination managers and policymakers to develop strategies to improve the tourism infrastructure, services, and attractions. This can contribute to the sustainable development of tourist destinations and improve the overall socio-economic well-being of local communities.

The industrial implications of my research are as follows:

- Marketing and advertising strategies

  By analyzing online reviews and understanding the focus points and preferences of tourists, businesses can tailor their marketing and advertising strategies to effectively communicate the unique features and appeal of their offerings. This can help them to attract more visitors and differentiate themselves from their competitors in the tourism industry. In addition, by analyzing online reviews and comparing the topics and themes, businesses can gain insights into areas that may need improvement. For the same type of destination, comparing the focus points of popular and unpopular spots can help to identify differences and opportunities for improvement.

- Customer relationship management

  Understanding the topics discussed in online reviews can provide valuable insights into customer perceptions, preferences, and satisfaction levels. This information can be used to improve customer relationship management strategies, personalize interactions, and increase customer satisfaction and loyalty.

- Industry collaboration and partnerships

  Research on online reviews and focus points can foster collaboration among industry stakeholders. Businesses, tourism boards, and other relevant organizations can work together to address common challenges, share best practices, and develop initiatives that improve the overall tourism experience in a destination or region.

6.4 Limitations of Proposed Method

Despite the contributions of the proposed method, there are some limitations of this study that should be acknowledged. The proposed method relies on the scoring of n-gram patterns based on the seven types of motivational factors that influence tourists. The scoring process may be subjective and biased, and different raters may interpret and score the same patterns differently. This limitation can be mitigated by including multiple raters and assessing the inter-rater reliability, and by validating the scoring process with scores assigned by experts or another group of raters.

The experiment extracted feature patterns that represented the characteristics of each tourist site. However, some feature patterns were not captured solely based on the frequency of occurrence of keywords and n-gram patterns. To obtain a more complete picture, additional information extraction methods should be used to extract more comprehensive content.

I also analyzed the focus points based on the results of the first principal component and found that the average contribution rate of the first principal component for ten tourist spots was 0.62. However, this rate did not fully represent the data. In future research, I aim to analyze the second and third principal components as well in order to obtain a more comprehensive understanding of the data.

In addition, the proposed method focuses only on Chinese tourists' evaluations of popular tourist spots in Hokkaido. Therefore, the results may not be possible to apply to other languages or tourist destinations. Future studies should replicate the proposed method with different languages and tourist destinations to provide more comprehensive insights into the foci of tourist spots.

In addition, the proposed method only considers written reviews, while some tourists may express their opinions about tourist spots through other means, such as photos or videos. Therefore, the results may not capture the full range of attention-grabbing aspects of the tourist spots.

Chapter 7

Conclusions

7.1 Conclusions

In this study, we have presented a method for extracting and quantifying tourists' points of interest by analyzing tourism reviews. Our goal was to capture and evaluate the attractiveness of tourist spots from the perspective of inbound tourists.

To achieve this, we collected a dataset of Chinese reviews of tourist spots in Hokkaido and extracted keywords from these customer reviews. By focusing on the topics most frequently mentioned by tourists, we gained valuable insights into their points of interest. To gain a more comprehensive understanding of these keywords, we further extracted 3-gram patterns that included the identified keywords.

To quantify the importance of these patterns, we assigned scores based on seven motivation factors and used PCA analysis to identify the most highly influenced factors as focus points for each site. This approach allowed us to capture the distinctive and appealing features of the tourist sites.

Our results indicate that the proposed method successfully captures the unique features of tourist destinations. In particular, the results obtained from customer reviews may differ from those derived from the feature patterns. This discrepancy is due to the additional contextual information present in the reviews, beyond the specific words contained in the patterns. However, in terms of focus, the feature patterns provide a clearer representation than a random selection of customer reviews. Moreover, the recommended reviews derived from the feature patterns

prove to be a more informative choice for individuals seeking location-specific information and focusing on the most relevant details.

In addition, we performed clustering on the feature patterns for each spot to gain a deeper understanding of the specific content related to the focus points. The clustered groups of feature patterns provide concise summaries of the information associated with these focus points. This knowledge can be used to enhance the travel experience and attract more inbound tourists to the destinations.

In conclusion, our proposed method for extracting and quantifying tourist focus points from tourism reviews provides valuable insights into the characteristics and attractiveness of tourist destinations. By using this information, tourism businesses and organizations can refine their marketing strategies, improve the visitor experience, and attract a wider audience. This research contributes to the advancement of tourism analysis and provides a framework for using customer reviews to uncover the key factors that shape tourist perceptions.

7.2 Future Work

There are several avenues for future research and development to improve and extend the findings of this study. The following areas represent potential directions for future work:

- Validation of the scoring process

  To ensure the reliability and validity of the scoring process, it is important to calculate inter-rater agreement. This will involve comparing the scores obtained by the proposed method with those assigned by expert raters. By assessing the agreement between different raters, the scoring process can be further validated and any biases can be identified and addressed.

- Principal component analysis

  In this study, we focused on the first principal component of the Focus Points. However, analysis of the second and third principal components can provide additional insights and a more complete understanding of the data. Exploring these additional components may reveal other underlying factors that contribute to the attractiveness of tourist destinations.

- Improve the feature pattern extraction

  Although the feature patterns extracted in this study captured the characteristics of each tourist site to some extent, there is room for improvement. It is necessary to use other information extraction methods to obtain more comprehensive content. Techniques such as natural language processing and machine learning can be explored to improve the feature pattern extraction process.

- Incorporate image recognition and sentiment analysis

  To expand the sources and platforms from which information is extracted, image recognition and sentiment analysis can be applied. Analyzing images of tourist sites can provide valuable insights into the visual aspects and attractions of a destination. In addition, sentiment analysis can uncover the emotional sentiment expressed in reviews, providing a deeper understanding of tourists' perceptions and experiences.

- Broadening the scope of evaluation

  The proposed method has the potential to be applied to a wider range of languages and tourist destinations beyond the current study's focus on Chinese reviews and tourist spots in Hokkaido. Evaluating reviews written in different languages and analyzing tourist spots in different regions or countries can provide a more comprehensive understanding of the focuses and attractions that resonate with inbound tourists worldwide.

- Analyze different types of destinations and reviews

  Expanding the analysis beyond natural attractions, such as cultural sites or historical landmarks, can provide insights into different types of tourist destinations. In addition, exploring different types of reviews, such as social media posts, travel blogs, or online forums, can provide a more diverse range of perspectives and opinions for analysis.

- Cross-cultural comparison

  Comparing the elements of interest and image of destinations based on reviews from Japanese and Chinese tourists can provide valuable insights into the

similarities and differences in their perceptions. Understanding the cultural nuances and preferences of different tourist groups can inform destination marketing strategies and enhance the overall visitor experience.

By addressing these future research directions, we can further develop the methodology and its applicability to different languages, destinations, and types of reviews. This will contribute to a more comprehensive understanding of what attracts inbound tourists to different destinations worldwide.

Appendix A

Additional Data and Translations

Table A.1: Examples of the customer reviews for specific tourist spots, the original Chinese review text of Table 5.10.

Spot	**Customer Reviews**
旭山動物園	1. 冬天的旭山動物園最人氣的活動應該就是企鵝散步了。胖胖的國王企鵝一扭一搖的走在雪地上，還可以看的到企鵝寶寶"獼猴桃"，更好玩兒。企鵝散步的時間是上午11點和下午2點，個人建議是上午動物園開門之後，就先去看企鵝散步比較好。除此以外，旭山動物園的一個人氣王就要數小熊貓了。萌萌的樣子，實在是可愛。旭山動物園每天都會有不同的動物食的活動，可以看到北極熊、小熊貓、海豹等等吃飯的樣子。每一次看到小動物們可愛的樣子，就覺得時間過的特別快，也會忘記生活中煩心的事情。所以，動物園是一種特別治愈心情的事情。 2. 很喜歡去旭山動物園看冬天的企鵝散步。搖搖擺擺的，憨態可掬！看他們走路，在雪地裡扭扭擺擺，真的是開心到什麼煩惱都可以忘記。大家冬天去旅遊的話記得去北海道的旭山動物園呀，每天上午11點和下午2點半各有一次企鵝的散步活動，記得去看哦！人還蠻多的，記得早點兒去佔位置啊！
狸小路商店街	1. 來日本如果想買日本特色商品，比如藥妝或日本土特，來狸小路就對了，可以至少上一天 2. 我並非第一次見到類似的場景了。這幾天只要走過狸小路總會看到一組這樣的年輕人，有男孩有女孩，有日本人有歐美人，演奏著各種樂器，唱著動感或是優美的歌。不是乞討賣唱，只是出於對音樂的喜愛和年輕的活力。他們的同伴和朋友或直接在他們面前席地而坐安靜聆聽，或在邊上著步伐打著節拍跟著唱。走過的路人看到了，不論愛不愛音樂，都會被這樣一幅年輕的畫面所感染，駐足觀看。
北海道庁旧本庁舍	1. 是一個政府機構，但是紅色的外表看著還挺漂亮的，個人還是非常的喜歡，特別像福爾摩斯電影裡面的一些建築物，拍照特別好看，特別是下雪的時候。 2. 北海道廳的舊址，歐式建築，入場免費，裡面可以了解北海道開發的歷史，裡面有中文的介紹冊。門口有紀念章蓋。

Tourism motivation factors	Definition
Stimulation Experience novelty and change	Want to try new experiences in an environment different from where I live. Want to feel the excitement of my heart pounding during the journey. Travel to bring changes to life. Want to break out of a static life through travel. It's boring to stay in the same environment all the time, so I want to travel. During the journey, I want to try something new and varied. Want to get in touch with the culture and customs not available in my place of residence.
Cultural observation Interested in local culture	Want to visit a famous monument or building. Want to visit art galleries, museum artworks. Want to learn more about local history and traditions. Want to appreciate local arts (music, drama, dance).
Local communication Communicate with people on the journey/ local	Want to build friendly relations with the locals. Want to chat with locals. Want to establish friendly relations with tourists from other regions. Want to know the living conditions/lifestyles of locals.
Health recovery Recovery from daily fatigue or stress	Want to eliminate the pressure accumulated on weekdays. Want to heal my tired body and mind. Want to forget my daily life and enjoy the journey at ease.
Experiencing nature Direct contact with nature	Want to feel the vast nature. Want to experience nature, walking in the mountains, etc. Want to feel the fresh air and water, and feel the beauty of nature. Want to see local animals or plants.
Unexpectedness A trip without knowing what will happen	During the journey, the destination is not clearly determined, and everything will take its course. Want to go on an unplanned trip. When traveling, I want to make a good schedule and plan. Want to travel as I want.
Educating oneself: The growth of the self	Want to use it as an opportunity to change our values and outlook on life. Want to re-examine myself/self-reflection. Want to discover who I am not the same as usual. Want to accumulate growth experience.

Figure A.1: The definition of each motivation factor

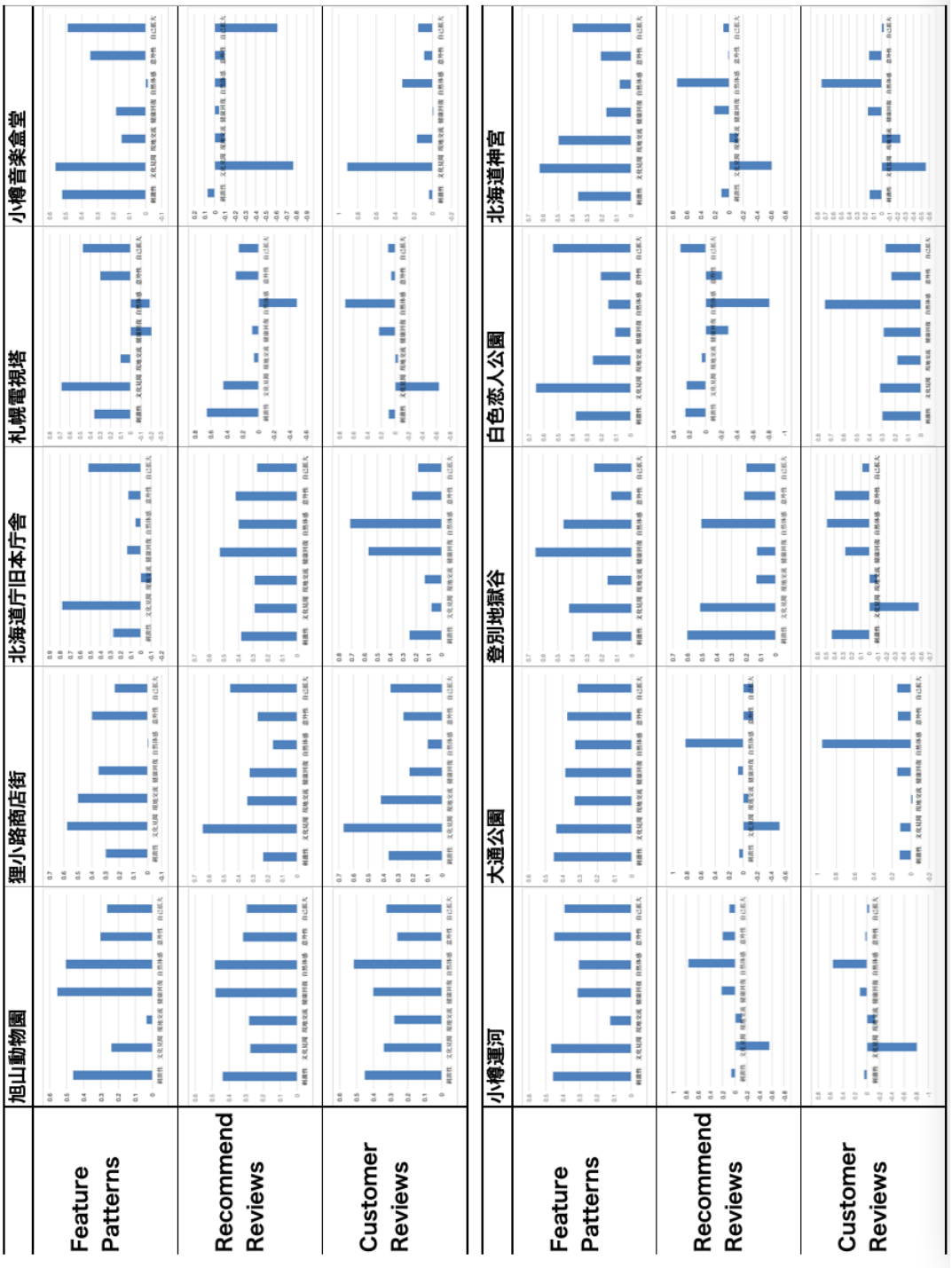

Figure A.2: The PCA results of 10 spots

Table A.2: Examples of feature patterns, recommended reviews, and customer reviews. The original Chinese review text of Table 5.12.

Feature Patterns	**Recommended Reviews**	**Customer Reviews**
企鵝 散步 活動	記不清有多久沒有去過動物園了，偶爾去一回，看到這些小可愛也會如小盆友一樣興奮。動物園裡除了招牌企鵝們，也有不少動物不可錯過，有北極熊、麋鹿、老虎，狼，還有這個與雪融為一体的雪梟。最不能錯過的就是每日2場的企鵝散步活動，超萌	1. 小動物們太可愛了，超萌，有機會還來。 2. 這裡最大的特點就是人性化的設計，最大程度保證了動物能處在相對自由的飼養狀態中，並通過巧妙的展示讓遊客零距離觀察動物，冬天動物園會推出企鵝大遊行的特色活動，遊客每天有兩次機會跟隨在萌萌的企鵝隊伍身後四處行走。 3. 帶兒童去北海道旭川必打卡景點，很人性化的動物園，各類動物很多，孩子玩了很開心。
動物 快樂 充滿 活力	在這裡可以看到海豹、北極熊、企鵝等寒冷地區的動物。不過真正讓人印象深刻的是動物園主張呈現動物的生活原貌，讓遊客看到動物們快樂、充滿活力的模樣。	
近 距 離 觀察 動物	動物都可以直接圍觀的，從上層，底層，直接直視都是可以全方位去了解動物的屬性。真的很近距離，除了猴子以及老虎類的有圍欄，基本都是可以近距離觀察動物。值得一來！	

Bibliography

[1] International tourism and covid-19. https://www.unwto.org/tourism-data/global-and-regional-tourism-performance. Accessed on: May 1, 2023.

[2] White paper on land, infrastructure, transport and tourism in japan, 2020. https://www.mlit.go.jp/hakusyo/mlit/r01/hakusho/r02/pdf/np101200.pdf.

[3] Honichi gaikyaku su 2020-nen 12-gatsu suikei-chi oyobi nenkan suikei-chi [Number of foreign visitors to japan, estimated values for December 2020 and whole year 2020] (in Japanese). https://www.jnto.go.jp/jpn/news/press_releases/pdf/210120_monthly.pdf. Accessed: 2021-01-20.

[4] The COVID-19 pandemic impact on Japanese inbound tourism. https://www.wakayama-u.ac.jp/ctr/news/2020101400019/. Accessed: 2021-03-28.

[5] White paper on tourism in japan, 2020. https://www.mlit.go.jp/kankocho/en/siryou/content/001375676.pdf.

[6] D. P. Faeni, R. Puspitaningtyas Faeni, H. Alden Riyadh, and Y. Yuliansyah. The covid-19 pandemic impact on the global tourism industry smes: a human capital development perspective. *Review of International Business and Strategy*, 33(2):317–327, 2023.

[7] S. Pramana, D. Y. Paramartha, G. Y. Ermawan, N. F. Deli, and W. Srimulyani. Impact of covid-19 pandemic on tourism in indonesia. *Current Issues in Tourism*, 25(15):2422–2442, 2022.

[8] Y. Zhao, H. Wang, Z. Guo, M. Huang, Y. Pan, and Y. Guo. Online reservation intention of tourist attractions in the covid-19 context: an extended technology acceptance model. *Sustainability*, 14(16):10395, 2022.

[9] Z. Zeng, P.-J. Chen, and A. A. Lew. From high-touch to high-tech: covid-19 drives robotics adoption. *Tourism Geographies*, 22:724–734, 2020.

[10] J. Romero and N. Lado. Service robots and covid-19: exploring perceptions of prevention efficacy at hotels in generation z. *International Journal of Contemporary Hospitality Management*, 2021.

[11] F. Seyitoğlu and S. Ivanov. Service robots as a tool for physical distancing in tourism. *Current Issues in Tourism*, 24:1631–1634, 2020.

[12] C. Hsu and S. Huang. Travel motivation: a critical review of the concept's development. *Tourism management: Analysis, behaviour and strategy*:14–27, Dec. 2007. DOI: 10.1079/9781845933234.0014.

[13] Y. Hayashi and T. Fujihara. Sightseeing motives of japanese overseas tourists as a function of destination, tour type and age. *The Japanese Journal of Experimental Social Psychology*, 48(1):17–31, 2008. DOI: 10.2130/jjesp.48.17.

[14] J. Wen, S. Huang, and T. Ying. Relationships between chinese cultural values and tourist motivations: a study of chinese tourists visiting israel. *Journal of Destination Marketing & Management*, 14:100367, 2019. ISSN: 2212-571X. DOI: https://doi.org/10.1016/j.jdmm.2019.100367. URL: https://www.sciencedirect.com/science/article/pii/S2212571X19300186.

[15] L.-H. Chen, J. P. Loverio, W. Mei-jung, B. Naipeng, and C.-C. Shen. The role of face (mien-tzu) in chinese tourists' destination choice and behaviors. *Journal of Hospitality and Tourism Management*, 48:500–508, 2021. ISSN: 1447-6770. DOI: https://doi.org/10.1016/j.jhtm.2021.08.009. URL: https://www.sciencedirect.com/science/article/pii/S1447677021001303.

[16] M. I. Simeon, P. Buonincontri, F. Cinquegrani, and A. Martone. Exploring tourists' cultural experiences in naples through online reviews. *Journal of Hospitality and Tourism Technology*, 2017.

[17] R. Ohkubo and Y. Muromachi. A study of destination image of foreign tourists to japan by analyzing travel guidebook and review site. *Journal of the City Planning Institute of Japan*, 49(3):573–578, 2014. DOI: 10.11361/journalcpij.49.573.

[18] S. Ohkawa. A cross-language comparison of tourist image using text mining. *Japan Society for Information and Management*, 39(2):85–94, 2019. DOI: 10.20627/jsim.39.2_85.

[19] Y. Hoshino, K. Shibata, E. Ishii, T. Otomo, and K. Yamada. Extracting information from twitter to create tourist information for foreigners [Gaikoku-jinmuke kanko joho sakusei no tame no Twitter kara no joho chushutsu] (in Japanese). In *The 16th National Conference of Tourism Information Society (2019)*, pages 7–8. Tourism Information Society, 2019.

[20] A. YASUHARA and M. LIU. A study of the differences between japanese and foreigners' tour experiences of hamarikyu gardens by analyzing review site. *Papers on Environmental Information Science*, 35:227–232, 2021.

[21] T. Emerson. Segmenting chinese in unicode, 2000.

[22] "Jieba" Chinese text segmentation. https://github.com/fxsjy/jieba. Accessed: 2023-03-28.

[23] Chinese stopword list. https://blog.csdn.net/shijiebei2009/article/details/39696571.

[24] J. E. Ramos. Using tf-idf to determine word relevance in document queries. In 2003. URL: https://api.semanticscholar.org/CorpusID:14638345.

[25] A. N. Langville and C. D. Meyer. *Google's PageRank and beyond: the science of search engine rankings*. In *Mathematics and Technology*. Princeton: Princeton University Press, 2006.

[26] R. Mihalcea and P. Tarau. TextRank: bringing order into texts. In *Proceedings of EMNLP-04and the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelonaand Spain, July 2004.

[27] Yujun Wen, Hui Yuan, and Pengzhou Zhang. Research on keyword extraction based on Word2Vec weighted TextRank. In *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, pages 2109–2113, 2016. DOI: 10.1109/CompComm.2016.7925072.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, G. Louppe, P. Prettenhofer, R. Weiss, R. J. Weiss, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: machine learning in python. *ArXiv*, abs/1201.0490, 2011. URL: https://api.semanticscholar.org/CorpusID:10659969.

[30] Z. Liu, F. Masui, and M. Ptaszynski. Supporting inbound tourism in hokkaido:keyword extraction and focus point analysis from spot reviews. In *Proceedings of the 2021 International Workshop on Modern Science and Technology*, pages 151–156. Kitami Institute of Technology, 2021.

[31] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[32] D. Jurafsky and J. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, volume 2. Feb. 2008. Chapter 4.

[33] H. Abdi and L. J. Williams. Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459, 2010. DOI: https://doi.org/10.1002/wics.101.

[34] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning*. Jan. 2013. Chapter 12. ISBN: 1461471389.

[35] I. Jolliffe. *Principal component analysis*. In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2014. ISBN: 9781118445112. DOI: https://doi.org/10.1002/9781118445112.stat06472. eprint: https://

onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat06472. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06472.

[36] A. Chakraborty, N. Faujdar, A. Punhani, and S. Saraswat. Comparative study of k-means clustering using iris data set for various distances. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 332–335, 2020. DOI: 10.1109/Confluence47617.2020.9058328.

[37] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st edition, 2007.

[38] D. Comaniciu and P. Meer. A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, June 2002. DOI: 10.1109/34.1000236.

[39] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pages 849–856. MIT Press, 2001. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.8100.

Research Achievements

First Author Publications

1. Z. Liu, F. Masui, J. Eronen, S. Terashita, and M. Ptaszynski. A new approach to extracting tourism focus points from chinese inbound tourist reviews after covid-19. *Sustainability*, 2023. Accepted for publication: 2023-05-24.
2. Z. Liu, F. Masui, and M. Ptaszynski. Examination of focus point extraction method from inbound tourism reviews using n-gram patterns and motivation factors of tourists. In *Information Processing & Management Conference*, Oct. 2022. Online poster.
3. Z. Liu, F. Masui, and M. Ptaszynski. Supporting inbound tourism in hokkaido: keyword extraction and focus point analysis from spot reviews. In *Proceedings of the 2021 International Workshop on Modern Science and Technology*. The International Center of National University Corporation Kitami Institute of Technology, Oct. 2021.
4. S. Ryu, F. Masui, T. Shun, and M. Ptaszynski. N-gram patan to kanko doki o riyo shita inbaundo kanko rebyu kara no fokasupointo chushutsu shuho no kento [a study on focus point extraction method from inbound tourism reviews using n-gram patterns and tourism motives (in japanese)]. In *Proceedings of the 17th National Conference of the Japanese Association of Tourism Informatics*, Nov. 2021.
5. S. Ryu, F. Masui, and M. Ptaszynski. Chubun supamurebyu kenshutsu no tame no shoppingusaitorebyu bunseki [shopping site review analysis for chinese spam review detection (in japanese)]. In *Proceedings of the 30th National*

Conference of The Japanese Society for Artificial Intelligence, 2P112in1–2P112in1. The Japanese Society for Artificial Intelligence, 2016.

6. S. Ryu, F. Masui, and M. Ptaszynski. Detecting spam reviews on the chinese online shopping site taobao. In *Proceedings of the International Workshop on Modern Science and Technology*, Nov. 2016.