Automated Evaluation of Tourism Motivation from Chinese Tourists in Japan using Transformers

Zhenzhen Liu[a] , Juuso Eronen[b] , Fumito Masui[a], and Michal Ptaszynski[a]

[a]Kitami Institute of Technology, Hokkaido, Japan;
[b]Prefectural University of Kumamoto, Kumamoto, Japan

ARTICLE HISTORY
Compiled April 28, 2025

ABSTRACT
The COVID-19 pandemic caused a significant drop in tourism in 2020, threatening Japan's goal of attracting 60 million tourists annually by 2030. To aid recovery, we propose an automated method to identify tourist motivations by analyzing tourist spot reviews. While previous studies on tourism motivation have primarily relied on labor-intensive approaches like questionnaires and interviews, this research uses deep-learning to automatically assess tourism motivations, reducing time and effort. Our study focuses on Hokkaido, a popular destination for Chinese tourists, building on previous research that manually scored motivation factors in reviews and used PCA to quantify focus points (the strongest tourism motivation) of the tourist spots. We enhance this by using pre-trained transformer models, to automatically assess reviews based on seven motivation factors. The results show that RoBERTa effectively scores these factors, closely matching manual assessments while significantly reducing the required time and human effort. However, the model is less accurate in identifying motivations linked to personal experiences. This automated approach offers valuable insights into tourist preferences, with the potential to inform and optimize Japan's tourism strategies, playing a crucial role in the industry's recovery and helping to meet national tourism goals.

KEYWORDS
Inbound tourism; Tourist motivation; Transformers

1. Introduction

The COVID-19 pandemic severely impacted global tourism, with international arrivals dropping 72% in 2020 (UN Tourism, 2023). Travel restrictions and business closures led to significant job losses and financial strain. As vaccinations became widespread and restrictions eased, tourism began recovering. In early 2024, Japan recorded 8.56 million foreign visitors, largely driven by the Easter holiday and cherry blossom season (JNTO, 2024). Japan is expected to surpass its pre-pandemic goal of 32 million tourists annually by 2025.

Despite this recovery, challenges remain. Chinese tourists, who made up 30% of visitors before the pandemic, have yet to return in full (The Asahi Shimbun, 2024). Additionally, tourism remains concentrated in major cities. In 2023, 70% of visitors stayed in Tokyo, Osaka, Kyoto, and surrounding areas. Japan Tourism Agency director

CONTACT Z.Liu Email: ryushinshintust@gmail.com

Ichiro Takahashi emphasized the untapped potential of rural regions and the need to attract more visitors beyond urban centers (The Japan Times, 2024).

This study aims to support the recovery of Japan's tourism industry by developing a system to more effectively analyze the travel motivations of foreign tourists, identify their main attractions and the reasons behind them. This study, focusing on the Hokkaido region and Chinese tourists, aims to enhance the attractiveness of rural areas and increase the number of Chinese tourists by analyzing travel motivations and verifying the effectiveness of the system.

Tourism motivation analysis has traditionally relied on manual content analysis and surveys. Foundational studies (Cohen, 1972; Crompton, 1979) identified key travel drivers like cultural exploration and escapism, while Dann (1981) highlighted push-and-pull factors. Pearce (2012) introduced the "Travel Career Ladder," showing how motivations evolve with experience. While influential, these frameworks depend on small datasets and subjective interpretation.

Later studies applied data-driven methods, such as Huang and Hsu (2009)'s cluster analysis of tourist segments and Xiang et al. (2015)'s text analytics on reviews, revealing links between guest experience and satisfaction. Researchers (Alaei et al., 2019; Gregoriades et al., 2023; Liu et al., 2023; Mishra et al., 2021) further used sentiment analysis to examine tourist behavior, demonstrating the benefits of NLP over biased surveys. Unlike surveys, NLP can analyze vast real-time datasets from online sources, offering more authentic insights. However, sentiment analysis primarily classifies general sentiments (positive, neutral, negative) and requires extensive annotated data, making large-scale implementation challenging (Liu et al., 2023).

Recent NLP advances, particularly transformer-based models like BERT and RoBERTa, offer more precise and scalable analysis of tourist motivations. Despite their potential, limited research has applied these methods in tourism. This study bridges that gap by evaluating transformers' effectiveness in identifying motivation factors and comparing their performance to human-assigned scores. Automating this analysis can enhance marketing strategies and destination management by providing deeper, data-driven insights with reduced manual effort. This linkage ensures that the research questions are firmly grounded in addressing the documented limitations of prior approaches while advancing the field with innovative methods.

Effective analysis of tourist motivations is critical to address these challenges. Traditional methods, such as surveys and manual reviews, are labor-intensive, time-consuming, and difficult to scale. While these approaches provide valuable insights, their reliance on subjective human effort limits their applicability to large datasets and real-time applications.

To address these challenges, this study explores the use of transformer-based models, such as BERT and RoBERTa, for automating the analysis of tourist motivations. These models leverage advances in natural language processing (NLP) to analyze large volumes of textual data, providing insights with reduced reliance on manual effort. By automating this process, it becomes possible to assess tourist motivations more efficiently and accurately, enabling stakeholders to make data-driven decisions that enhance marketing strategies and destination management.

The research questions for this study are:

- How can automated methods, like deep learning, be utilized to accurately identify the primary motivation factors that influence tourists' preferences?
- How does the automated approach of scoring reviews compare to previous manual methods, specifically human-assigned scores, in terms of accuracy, efficiency,

and insights gained?

Previously, Liu et al. (2023) proposed a method that automatically extracts concerns from inbound tourists' reviews of tourist spots that include attention-grabbing aspects and relevant motivations for traveling. First, they extract frequent n-gram patterns from the reviews of each tourist spot. Then, they manually scored these patterns based on seven factors that influence tourist motivations. Then, they performed principal component analysis (PCA) on the scoring results to identify the attention-grabbing aspects of each tourist spot.

In this paper, we build upon this method by employing pre-trained language models to automatically score tourism motivations based on online reviews, thus reducing the need for manual evaluation and improving the method's performance. The term 'manual evaluation/methods' specifically refers to the process where human evaluators manually assign scores to reviews based on predefined tourism motivation factors. These scores are determined by a group of evaluators who analyze the content of each review and rate it on a five-point scale according to the strength of its relationship with each factor. This method, while reliable, is labor-intensive and time-consuming. By comparing these human-assigned scores with the automated predictions generated by transformer models, we aim to assess the efficiency and reliability of the automated approach in replicating or surpassing traditional manual evaluation methods.

We first collected tourist reviews for ten spots in the Hokkaido area, scoring them based on seven tourism motivation factors. We then compared the performance of three different pre-trained transformer models, namely BERT, RoBERTa, and ELECTRA. We fine-tuned the best model using the collected review data and analyzed how the artificial scores compared to those made by people.

In addition, our approach is not language-specific, so we may use it in other languages and nations as well, which expands its potential applications in international multilingual tourist research. Moreover, our approach can identify shifts in travelers' worries by comparing samples taken before and after the outbreak. This implies that our approach might aid in being ready for calamities in the future that can impact the travel and tourist sector.

The study develops and evaluates a novel automated approach to scoring tourist motivations using transformer-based models. We demonstrate the applicability of these models through a case study on Chinese tourists in Hokkaido, offering insights into how automated methods can enhance tourism motivation analysis. The findings provide a foundation for scaling this approach to other regions and contexts, supporting the development of more efficient, data-driven strategies in tourism management.

The structure of this paper is organized as follows: Section 1 outlines the main objectives and motivation behind this study. Section 2 reviews related literature and previous research in the field. Section 3 details the methodology and the research steps undertaken. Section 4 presents the experimental setup and analyzes the outcomes. Section 5 provides an in-depth discussion of the findings and validation of the proposed methods. Finally, Section 6 summarizes the conclusions drawn from this study.

2. Previous Research

2.1. Travel Motivation

Motivation and satisfaction are key concepts in tourism, influencing tourist behavior. Travel motivation refers to the underlying reasons and forces that drive individuals to travel to specific destinations or engage in travel-related activities. Understanding these motivations is crucial for the tourism industry to enhance traveler experiences and develop effective strategies for attracting visitors (Hsu and Huang, 2007). Scholars have proposed various theories to explain tourist motivations, each offering unique perspectives.

The push-pull model, proposed by Dann (1981) categorizes motivations into two types. Push factors are internal drivers such as relaxation, adventure, or escape from routine, while pull factors represent external destination attributes like beaches, cultural attractions, or entertainment facilities. Push factors like relaxation and adventure interact with pull factors such as destination attributes to guide tourist decisions. This model is widely applied in tourism research to segment tourist markets and design destination-specific strategies.

The hierarchy of needs theory, based on Maslow's hierarchy, was adapted by Pearce Pearce (2012) to tourism in the form of the "Travel Career Ladder." This model suggests that tourist motivations evolve as individuals gain more travel experience, emphasizing a progression from basic needs, such as safety, to higher-level aspirations, like self-actualization. Once tourists arrive at their destination, their motivations influence how they allocate time and resources. Motivations tied to cultural exploration often lead tourists to prioritize visits to museums, historical landmarks, and local festivals. In contrast, tourists motivated by relaxation may prefer activities like spa treatments or leisure time at resorts. Understanding these on-site behaviors helps destination managers tailor experiences to match tourist expectations.

Socio-psychological motivations, identified by Crompton Crompton (1979), outline nine key motives for travel, including novelty, education, and enhancing relationships. These motivations highlight the diverse personal and social drivers influencing tourist behavior. For example, they found that socio-psychological motives, such as the desire for escape or novelty, are key drivers of destination choice.

Devesa et al. (2010) explored the relationship between motivation and visitor satisfaction through a survey conducted at a rural destination in Spain. Using ANOVA, factor, and cluster analyses, the results confirmed that motivation significantly impacts visit assessment criteria and specific satisfaction factors. The study provides insights for management and marketing strategies.

Dunn Ross and Iso-Ahola (1991) examined motivation and satisfaction among 225 sightseeing tourists, assessed before and after their tours. Data from 10 different tour buses revealed a strong alignment between motivation and satisfaction, with knowledge seeking, social interaction, and escape being key factors. Tourists who joined by chance showed higher scores in knowledge-seeking and five satisfaction dimensions compared to regular and convention groups, leading to high overall satisfaction.

The study by Park and Yoon (2009) focuses on segmenting and profiling tourist motivations in rural tourism in Korea through a survey of 252 tourists. It identifies four distinct segments: Family Togetherness Seekers, Passive Tourists, Want-it-All Seekers, and Learning and Excitement Seekers. The findings reveal that relaxation is the primary motivation for these tourists, highlighting the importance of leisure in rural travel. Additionally, there are significant differences in socio-economic characteristics among the identified segments. These insights can inform targeted marketing strategies for the development of rural tourism.

Also, Su et al. (2020) addressed limitations in analyzing tourist satisfaction solely through travel motivation by incorporating experience-related factors and destination image. A survey of 352 visitors at Hoi An UNESCO World Heritage Site, Vietnam, conducted during their on-site experience, used Partial Least Square-Structural Equation Modeling. The results showed that motivation significantly influenced visitor engagement, experience, and destination image, leading to satisfaction. An indirect link between motivation and satisfaction was also confirmed. Practical implications for Destination Management Organizations (DMOs) are provided.

A study by Chi and Phuong (2022) explored how travel motivations influence tourists' intention to visit city destinations, alongside the roles of time perspective and city image. It finds that travel motivations are significantly positively related to the intention to visit city tourism sites. Furthermore, travel motivations mediate the relationship between time perspective, city image, and travel intentions. Tourists are driven by factors such as knowledge enhancement, self-fulfillment, socializing, and escape. The study suggests that city tourism providers should focus on these motivations and develop a strong city image to attract visitors. The findings emphasize the importance of understanding travel motivations in shaping effective tourism strategies and branding.

Additionally, a study by Annika Aebli and Taplin (2022) explored travel motivations and demotivators during the COVID-19 pandemic. Through interviewing potential tourists and tourism destination managers, they find that people are motivated by the need for mental wellbeing and social connections, but are discouraged by health and safety concerns. The research adds to our understanding of how travel preferences are affected during a global health crisis.

Many studies analyze tourist motivations for a specific traveler or destination by using surveys or interview data combined with statistical methods. For instance, Hayashi and Fujihara (2008) conducted surveys with 1,014 Japanese travelers to identify the key factors influencing their travel decisions. Seven main motivation variables were identified by their analysis: self-expansion, excitement, unexpectedness, natural experience, local engagement, cultural observation, and health recovery.

M.Carvache-Franco et al. (2020) aimed to segment tourist demand at a coastal destination based on motivations, sociodemographic factors, and trip characteristics. A survey of 390 visitors in Manta, Ecuador, was analyzed using multivariate statistical techniques. Results identified three motivational dimensions—ecotourism/gastronomy, sun/beach/entertainment, and relaxation—leading to three tourist clusters: 'beach lovers,' 'eco-coastal,' and 'multiple motives.' Older tourists showed stronger motivations for sun, beach, and gastronomy, while frequent visitors reported higher satisfaction and motivation. These insights can help tourism providers tailor products to specific tourist demands.

A study by Valverde-Roda et al. (2022) analyzed the gastronomic motivations of tourists visiting Granada, Spain, home to two UNESCO World Heritage Sites. A total of 1,612 surveys were collected at culinary establishments and historical sites, assessing tourists' opinions on gastronomy and travel motivations. The findings segmented tourists into three groups—survivors, enjoyers, and experiencers—based on their interest in gastronomy. The study confirmed that gastronomy influences tourist satisfaction and enhances the destination's competitiveness.

Recent studies highlight Last Chance Tourism (LCT) as a growing motivation for visiting natural sites before they disappear. Montenvers-Mer-de-Glace, France's most visited glacier, has become an LCT destination, with visitors drawn primarily by its environmental features and LCT purposes. Exploratory Factor Analysis (EFA) reveals that motivations include LCT and five other factors, while qualitative data identifies 'Fame' as an additional motivational category. LCT motivations are categorized into four dimensions: observing, understanding, urgency, and witnessing environmental change (Salim and Ravanel, 2023).

Considering the motivations of Chinese tourists, Wen et al. (2019) examined the connection between Chinese cultural values and the driving force for Chinese tourists' journey to Israel. Through surveys and interviews, analyzed using Canonical Correlation Analysis, they discovered that the primary motivations of Chinese tourists were related to self-actualization, escapism/relaxation, adventure, destination uniqueness, business development, sightseeing, and knowledge enhancement/study. Business development emerged as a significant factor.

Additionally, Jiang et al. (2020) investigated the motivations of Chinese outbound leisure tourists and their relationship to cultural values. Through 60 in-depth laddering interviews, researchers identified 15 key motivations categorized into three main themes: self-enhancement, nurturing the soul, and harmonious relationships. Self-enhancement focuses on gaining experiences and knowledge, while nurturing the soul emphasizes the pursuit of pleasure and appreciation of nature's beauty. Harmonious relationships highlight the importance of family and friendships in travel experiences. The research utilized content analysis and developed a Hierarchical Value Map to visualize the connections between attributes, consequences, and values.

Chinese tourist motivations and behaviors have also been studied in the context of Japan. One of the primary motivations for Chinese tourists visiting Japan is the appeal of its cultural experiences, natural beauty, shopping opportunities, and historical attractions. Tourists often seek leisure, novelty, and self-fulfillment, with many drawn to seasonal phenomena like cherry blossoms or cultural artifacts such as traditional architecture and cuisine. Repeat visitors tend to focus on fewer destinations with deeper engagement, while first-time travelers explore widely, reflecting different motivations and travel strategies. Social Network Analysis (SNA) of tourist flows in Japan highlighted that destinations can be categorized into core, secondary, and peripheral nodes based on their roles in the network, revealing distinct patterns such as multi-center agglomeration and single-center dispersion (Zeng and He, 2019), (Zeng, 2021).

Several factors influence Chinese tourist flows at the inter-destination level. Tourist-specific conditions, such as travel purpose, budget, and prior experiences, are critical in determining travel patterns. The role of online reviews and opinions from others is particularly influential among younger tourists. Destination-specific factors, such as accessibility, uniqueness of attractions, and seasonal features, further shape decisions. Transportation considerations, including the cost and complexity of networks within Japan, also play a crucial role. Additionally, unforeseen circumstances like adverse weather or flight delays affect travel itineraries and experiences (Zeng and He, 2019).

Constraints faced by Chinese tourists also play a significant role in shaping their travel behavior. Structural barriers, such as high costs, visa requirements, and limited time availability, are among the most cited challenges. Political and social factors further influence travel decisions, with historical tensions and perceptions of nationalism sometimes deterring travel to Japan. For instance, negative media portrayals or strained political relations can suppress demand, whereas the allure of cultural affinity and perceived safety can mitigate these effects. Intrapersonal constraints, such as a lack of interest or perceived inconvenience, and interpersonal barriers, including disapproval from family or the absence of travel companions, also impact decision-making (Lin et al., 2017).

A previous study by Liu et al. (2023) developed an automated method to identify focus points, which represent the strongest tourism motivations for specific tourist spots, using reviews from inbound tourists. The study concentrated on Chinese tourists visiting Hokkaido, Japan, with the goal of uncovering the underlying motivations driving their interests in popular travel destinations. The researchers began by selecting ten well-known tourist spots in Hokkaido and gathering reviews from a travel-oriented website that catered to Chinese tourists. They then extracted keywords from the reviews specific to each spot and identified common n-gram patterns that encompassed these keywords.

To analyze these n-gram patterns, they applied a scoring system based on seven motivation factors that influence tourist behavior. These scores were then processed using Principal Component Analysis (PCA) to generate visual principal component charts that mapped the primary motivations associated with each tourist spot. Clustering techniques were subsequently applied to these n-gram patterns to categorize thematic content and clarify the interrelationships between different themes and tourist motivations.

Most of the previous studies rely on surveys or interviews, for example, those by Devesa et al. (2010), Park and Yoon (2009), Su et al. (2020), Chi and Phuong (2022), Annika Aebli and Taplin (2022), Hayashi and Fujihara (2008), M.Carvache-Franco et al. (2020), Valverde-Roda et al. (2022), Wen et al. (2019) and Jiang et al. (2020). These require significant time and labor.To address this, this study proposes using transformer models to automate the assessment of tourist motivations, reducing manual effort.

Focusing on Japanese destinations, this study adopts the seven motivational factors identified by Hayashi and Fujihara (2008). These factors facilitate the comparison and analysis of differences and preferences between Chinese and Japanese tourists. This framework was also utilized in previous research by Liu et al. (2023), further demonstrating its applicability for analyzing the motivations of Chinese tourists visiting Hokkaido.

2.2. Automatic Scoring of Tourism Reviews

Early attempts at automatic scoring of tourism reviews relied on keyword-based and rule-based approaches. For instance, keywords such as "beautiful," "peaceful," and "scenic" might be associated with positive sentiments about natural attractions, while words like "overcrowded" and "disappointing" could indicate negative experiences (Hu and Liu, 2004). Reviews were then scored based on the frequency of these keywords, with positive terms contributing to higher scores and negative terms leading to lower scores. However, these methods often struggled with nuances in language and context, leading to inaccuracies in scoring, especially in cases where reviews used sarcasm or complex language structures (Hu and Liu, 2004).

As the field advanced, statistical and machine learning methods became more prevalent in scoring tourism reviews, as shown in works by Ye et al. (2009) and (Alaei et al., 2019). These methods applied probabilistic models and classifiers to categorize text based on various features extracted from the data. For instance, in tourism reviews, features such as sentiment polarity, review length, and the presence of specific terms related to tourist attractions could be used to predict overall review scores (Ye et al., 2009). Support Vector Machines (SVM) and Naive Bayes classifiers were commonly used for this purpose, with SVMs demonstrating particularly high accuracy in sentiment classification tasks . These works emphasize the shortcomings of traditional survey-based methods, including biases like response bias and limited coverage, and highlight the advantages of NLP technologies. Unlike surveys, NLP techniques can process extensive, real-time datasets from sources such as social media and online reviews, enabling researchers to gain deeper, more authentic insights into tourist behavior and perceptions. Tourists often share their experiences and opinions freely in online platforms, which provides a more detailed understanding of their motivations and behaviors compared to the structured and potentially biased responses captured through surveys.

With the rise of deep learning, scoring tourism reviews saw significant improvements in accuracy and efficiency. Deep learning models, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), were applied to capture complex patterns and dependencies in text data. For instance, LSTM-based RNNs were used to analyze sequences of words in reviews, allowing the model to learn long-term dependencies and better understand the context of the text. CNNs, on the other hand, were effective in capturing local patterns and features within reviews, enabling accurate sentiment analysis and scoring (Phillips et al., 2015) (Zhang and Wallace, 2016) (Martín et al., 2018).

Additionally, deep learning techniques and sentiment have been utilized in tourism research by studies such as Mishra et al. (2021) and Gregoriades et al. (2023), offering new ways to analyze large-scale data from online sources like reviews and social media. These methods enable researchers to capture tourist behavior and preferences that might be missed in traditional survey approaches. By processing vast amounts of unstructured data, these techniques provide a more comprehensive understanding of tourist sentiments and experiences. However, despite their advantages, sentiment analysis methods have limitations. They primarily provide broad sentiment categorizations (e.g., positive, neutral, or negative) and lack the depth of insight that traditional surveys can deliver. Consequently, sentiment analysis serves best as a complementary approach rather than a full replacement for conventional surveys (Alaei et al., 2019). Furthermore, these techniques often require extensive annotated datasets for training, which can be both time-intensive and expensive to produce (Liu et al., 2023). This dependency on large, labeled datasets presents a scalability challenge for employing sentiment analysis on a broader scale without significant resource investment.

The introduction of transformer models, such as BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), marked a significant advancement in automatic scoring of tourism reviews. These models revolutionized natural language processing by capturing contextual information more effectively and efficiently than previous architectures. BERT, in particular, excelled in understanding the context of text data, allowing it to accurately score tourism reviews based on various factors such as sentiment, thematic relevance, and specific aspects of tourist experiences. Additionally, models like RoBERTa (Liu et al., 2019) further improved upon BERT's performance by optimizing the pretraining process and incorporating additional training data, resulting in even more accurate scoring of tourism reviews (Yanuar and Shiramatsu, 2020) (Chen et al., 2024).

The evolution of automatic scoring methods for tourism reviews has progressed from simple keyword-based approaches to sophisticated deep learning and transformer-based models. Each stage of development has brought improvements in accuracy, efficiency, and scalability, improving the analysis of tourist feedback. Transformer models like BERT and RoBERTa represent the current state-of-the-art, providing powerful tools for scoring tourism reviews accurately and effectively. These advancements have significantly enhanced the ability of stakeholders in the tourism industry to extract valuable insights from reviews, thereby improving customer experiences and guiding strategic decision-making processes. In this research, we plan to score the reviews by using seven different motivation factors by Hayashi and Fujihara (2008) instead of simply scoring the review by sentiment.

3. Materials and Methods

In this study, we propose a method for automatically scoring the tourism motivation factors in reviews of tourist sites. We used the same 500 Chinese tourist reviews extracted from Ctrip.com as Liu et al. (2023). These reviews were manually scored based on the seven types of tourism motivation factors by Hayashi and Fujihara (2008). Next, we fine-tuned pre-trained Transformer models (BERT, RoBERTa, and ELECTRA) using the scored data, employing five-fold cross-validation during this process. We compared the performance of these Transformer models. Finally, we used the best-performing Transformer model to predict the scores and compared these predictions with the human-assigned scores.

The research procedure is shown in Figure 1, and each step of the procedure is explained in Sections 3.1-3.3.

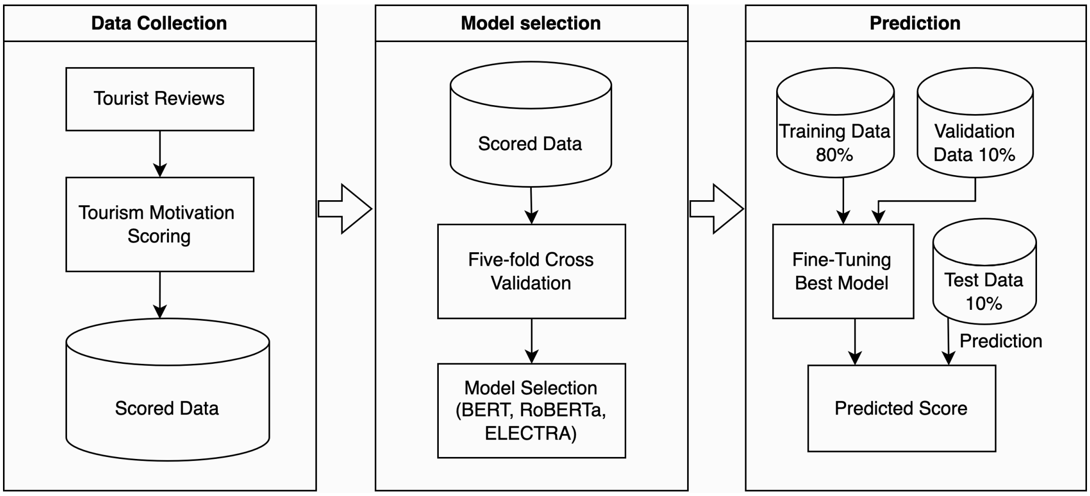

Figure 1. Flowchart of the Proposed Method

3.1. Dataset

The dataset by Liu et al. (2023) was collected by selecting popular Hokkaido spots from Ctrip.com [1] with lots of reviews. Ctrip.com is a well-known Chinese travel industry website. The site uses its own algorithm to provide a list of recommended tourist spots in Hokkaido based on spot ratings, number of reviews and other factors. From this list, we selected ten spots that ranked high and had a large number of reviews. By crawling the websites and extracting useful information such as content, ratings, and dates, they collected a total of 500 Chinese reviews from these spots. The dataset is summarized in Table 1.

Next, the impressions of the reviews were manually rated for each of the tourism motivation factors. The tourism motivation factors (stimulation, cultural observation, local communication, health recovery, nature experience, unexpectedness, and educating oneself) (Hayashi and Fujihara, 2008) proposed by Lin et al. were used.

Evaluators were responsible for rating the reviews on a five-point scale, where 5 indicates the strongest relationship between the review and the motivation factors and 1 indicates no relationship. Table 2 and Table 3 provide descriptions of the travel

[1]https://www.ctrip.com/

Table 1. Tourism Spots and the Numbers of Reviews

SpotName	Reviews
Asahiyama Zoo	50
Former Hokkaido Govt. Office	50
Hokkaido Shrine	50
Noboribetsu Hell Valley	50
Odori Park	50
Otaru Canal	50
Otaru Music Box Hall	50
Sapporo TV Tower	50
Shiroi Koibito Park	50
Tanukikoji Shopping Street	50
Total	500

motivation scales and scoring criteria, respectively. All reviews were scored by six Chinese-speaking evaluators, including both laypeople and tourism researchers. They had either visited Hokkaido or read introductions to its tourism spots and assessed the reviews based on tourism motivation factors. The average scores of these six evaluators were used in the experiment.

Table 2. Explanation of Tourist Motivation Scale, also used in the previous research by Liu et al. (2023)

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 while traveling
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 3. The Scoring Criteria, also used in the previous research by Liu et al. (2023)

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

3.2. Model Selection

The scored data was used for fine-tuning the transformer models, BERT, RoBERTa and ELECTRA.

- BERT (Devlin et al., 2019), the Bidirectional Encoder Representations from Transformers, has a multilingual variant. Based on the Transformer architecture (Vaswani et al., 2017), BERT processes text bidirectionally, capturing contextual relationships more effectively than traditional left-to-right or right-to-left models. It is pre-trained on two tasks: Masked Language Modeling (predicting masked words using context) and Next Sentence Prediction (determining logical connections between sentence pairs).
- RoBERTa (Liu et al., 2019), or Robustly Optimized BERT Approach, enhances BERT by optimizing the pre-training process. It extends training on a larger dataset, removes Next Sentence Prediction, and employs dynamic masking during Masked Language Modeling, improving its robustness and language understanding.
- ELECTRA (Clark et al., 2020) introduces a novel pre-training approach called replaced token detection. Instead of masking words, a generator replaces some tokens with incorrect ones, and a discriminator learns to identify them. This method is more sample-efficient than Masked Language Modeling, allowing ELECTRA to match or surpass BERT and RoBERTa's performance with the same data.

To ensure the robustness and reliability of the model's performance, five-fold cross-validation is used in this experiment. Five-fold cross-validation involves dividing the dataset into five subsets, where the model is trained on four subsets and validated on the remaining one. This process is repeated five times, each time with a different subset used as the validation set.

Then, we evaluated the performance using metrics such as R2, MAE (Mean Absolute Error), MSE (Mean Squared Error), and accuracy to compare the Transformer models. The highest performing model was chosen for the next step of our analysis.

3.3. Prediction

We fine-tuned the best-performing model using 80% of the scored data for training, 10% for validation, and the remaining 10% for testing. Allocating 80% of the data to training is a widely adopted practice in machine learning research, as it provides sufficient data for the model to learn meaningful patterns while reserving portions for validation and evaluation Goodfellow et al. (2016). While it is possible to allocate a larger portion for testing, the relatively small size of our dataset necessitated maximizing the data used for training. This decision helps mitigate the risk of overfitting, which can occur when the training set is too small Hawkins (2004), Ying (2019).

We used Intraclass correlation coefficients to calculate the similarity between the predicted scores and the scores assigned by evaluators. The Intraclass Correlation Coefficient (ICC) measures the reliability or consistency of measurements made by multiple observers or instruments on the same subject. It quantifies how much of the total variability in the data is due to differences between groups (or classes) versus differences within groups. This is done to ensure the consistency and dependability of measurements across different observers or instruments. For example, in the case of our research the ICC can determine whether the scores are consistent between human evaluators and the model's predictions. A high ICC value, close to 1, indicates high reliability, meaning that measurements within the same group are very similar, while a low ICC value, close to 0, suggests low reliability, with significant variability within groups. Negative ICC values indicate that within-group variability exceeds between-group variability, signifying poor reliability.

4. Experiments

4.1. Model Selection

We used the scored reviews to fine-tune the transformer models, and the performance results are shown in Tables 4, 5 and 6.

The RoBERTa model demonstrated the best performance in predicting tourism motivations, with R2 scores ranging from 0.4701 to 0.6196 for the motivation factors "Culture" and "Nature" as shown in Tables 4, 5, and 6.

For behavioral sciences, Cohen (2013) proposed 'small', 'medium' and 'large' magnitudes for R2. These values are 0.02, 0.13, and 0.26, respectively. However, he states that these are general definitions and the interpretation of what constitutes a "large" predictive relationship depends heavily on the context, subject matter, and specific research objectives.

Henseler et al. (2009) provide the following rules for interpreting R2 for hospitality and tourism research: 0.75 (substantial), 0.50 (moderate) and 0.25 (weak). This means that the highest scoring motivation factors (culture, nature) would show approximately moderate predictive relationship.

Conversely, the R2 scores for "Stimulation" (0.2496), "Health" (0.2906), and "Self-expansion" (0.2231) are lower, indicating a weak predictive relationship. These results suggest that the model faces greater difficulty in accurately predicting these more abstract and subjective motivation factors. The lower R2 scores for these factors can be attributed to the fact that they often involve context-dependent experiences that are harder to capture in written reviews, as noted in previous NLP studies on sentiment analysis and abstract concepts Pang et al. (2008), Cambria and White (2014).

These findings align with the broader understanding in NLP and machine learning that abstract and subjective concepts—such as excitement (Stimulation), health recovery (Health), and personal growth (Self-expansion)—are more difficult to predict accurately due to their complex, context-sensitive nature and the absence of explicit linguistic markers Cambria and White (2014).

4.2. Prediction

We chose the best-performing model, RoBERTa, to predict the test data, which comprised 10% of the scored data. To evaluate the degree of consistency between human-assigned scores and predicted scores, we used intraclass correlation coefficients (ICCs). The ICC scores are shown in Table 7. The average ICC score was 0.5476 and according to Koo and Li (2016), ICC values between 0.50 and 0.75 represent moderate reliability, values between 0.75 and 0.90 indicate good reliability, and values above 0.90 are considered excellent. This shows that the results indicate a moderate reliability.

Motivation factors such as "Cultural Observation" (ICC = 0.6361), "Local Communication" (ICC = 0.6870), and "Experiencing Nature" (ICC = 0.6893) demonstrated higher agreement between predicted and human-assigned scores. These results sug-

Table 4. The Performance of BERT Model

Tourism Motivation Factor	R2	MAE	MSE	Accuracy
Stimulation	0.1715	0.3197	0.1894	0.7860
Cultural observation	0.5042	0.3435	0.2051	0.7360
Local communication	0.2391	0.3376	0.1803	0.7560
Health recovery	0.2084	0.4224	0.2600	0.6460
Experiencing nature	0.5336	0.4381	0.3180	0.6460
Unexpectedness	0.2449	0.3775	0.2246	0.7000
Educating oneself	0.1530	0.3771	0.2051	0.7060
Average	0.2935	0.3737	0.2261	0.7109

Table 5. The Performance of RoBERTa Model

Tourism Motivation Factor	R2	MAE	MSE	Accuracy
Stimulation	0.2496	0.2983	0.1733	0.7920
Cultural observation	0.5889	0.3271	0.1696	0.7820
Local communication	0.4021	0.2885	0.1407	0.8260
Health recovery	0.2906	0.3850	0.2332	0.7240
Experiencing nature	0.6196	0.3975	0.2608	0.6900
Unexpectedness	0.3631	0.3334	0.1894	0.7600
Educating oneself	0.2231	0.3645	0.1875	0.7360
Average	0.3910	0.3421	0.1935	0.7586

Table 6. The Performance of ELECTRA Model

Tourism Motivation Factor	R2	MAE	MSE	Accuracy
Stimulation	0.1007	0.3239	0.2077	0.7560
Cultural observation	0.4701	0.3588	0.2175	0.7340
Local communication	0.2799	0.3124	0.1704	0.7880
Health recovery	0.1664	0.4274	0.2741	0.6380
Experiencing nature	0.4621	0.4716	0.3708	0.6200
Unexpectedness	0.2301	0.3813	0.2295	0.7020
Educating oneself	0.1389	0.3834	0.2080	0.6780
Average	0.2640	0.3798	0.2397	0.7023

gest that the model effectively captures explicit and well-defined patterns in reviews. On the other hand, factors like “Stimulation” (ICC = 0.3398) and “Self-expansion” (ICC = 0.4561) showed lower agreement, reflecting the difficulty of predicting abstract motivations.

The moderate average ICC score highlights the model’s capacity to replicate human evaluations for certain motivation factors while indicating areas where further refinement is necessary. These findings align with the broader understanding that subjective and abstract concepts are more challenging for machine learning models to interpret accurately Pang et al. (2008) Cambria and White (2014).

Table 7. Intraclass Correlation Coefficients Score

Tourism Motivation Factor	Score
Stimulation	0.3398
Cultural observation	0.6361
Local communication	0.6870
Health recovery	0.4911
Experiencing nature	0.6893
Unexpectedness	0.5341
Educating oneself	0.4561
Average	0.5476

5. Results and Discussion

5.1. Results and Discussion

In this study, we fine-tuned Transformer models BERT, RoBERTa and ELECTRA to predict tourism motivations from Chinese reviews. The results indicate that RoBERTa is the best-suited model for our case.

The R2 score measures the proportion of variance in the target variable explained by the model. It ranges from 0 to 1, with higher values indicating better performance. From the R2 scores of the RoBERTa model, Table 5, the motivation factors ”Nature” and ”Culture” exhibited higher score, while the factors ”Stimulation”, ”Health” and ”Self-exp.” showed lower score. This difference in prediction performance can be attributed to several factors:

- Feature Patterns and Keywords:

  Reviews related to ”Nature” and ”Culture” often contain distinct and rich feature patterns or keywords. These reviews may describe specific attributes such as scenic beauty, historical sites, or cultural events, which provide clear signals that the model can easily learn from and associate with the respective motivation factors.

  In contrast, ”Stimulation”, ”Health” and ”Self-exp.” may involve more abstract and subjective experiences that are less explicitly mentioned in reviews. These factors can encompass a wide range of activities and feelings, making it harder for the model to identify consistent patterns and accurately predict them.

- Review Content and Context:
  The content of reviews for "Nature" and "Culture" is often more focused and concrete, describing tangible experiences and observations. This specificity helps the model to better understand and predict the underlying motivation factors.
  Reviews for "Stimulation", "Health" and "Self-exp." might be more diverse and context-dependent, covering a broad spectrum of personal experiences and emotional responses. The variability and subjectivity in these reviews can lead to less accurate predictions.
- Data Distribution and Representation:
  There might be a more balanced and abundant representation of "Nature" and "Culture" related reviews in the dataset, providing the model with sufficient examples to learn effectively.
  Conversely, "Stimulation", "Health" and "Self-exp." related reviews might be underrepresented or unevenly distributed, leading to less robust model training and lower prediction accuracy.
- Language and Expression:
  The language used in reviews for "Nature" and "Culture" is likely to be more descriptive and consistent, using common phrases and terms that are easier for the model to recognize and learn.
  Reviews addressing "Stimulation", "Health" and "Self-exp." may involve more varied and finely detailed language, reflecting personal perceptions and individual interpretations, which poses a challenge for the model to accurately capture and predict.

Examples of reviews and prediction scores are shown in Table 8. Reviews related to the nature motivation factor often include clear descriptions of natural elements, such as "beautiful garden," "ducks," "lush trees," and "wooden stairs." These reviews contain numerous keywords directly related to nature, making it easier for the Transformer model to recognize and predict this motivation factor accurately. Reviews describing cultural experiences, such as the history and exhibits of the Otaru Music Box Hall, frequently mention terms associated with cultural knowledge and experiences. This explicit connection to cultural content allows the Transformer model to accurately identify and predict the culture motivation factor. Reviews related to the stimulation motivation factor often lack clear keywords or phrases indicating excitement or thrilling experiences. Instead, they may describe experiences that some individuals find exciting based on personal feelings, making it difficult for the model to predict this factor accurately. This subjectivity and lack of consistent terminology pose challenges for the model. For health recovery and self-exp. motivation factors, reviews typically do not contain explicit keywords related to health recovery or self-growth. For example, while some people might feel relaxed and stress-free when seeing cute animals, these reviews may not explicitly state this connection. Similarly, experiences like encountering a hot spring with a strong sulfur odor, which is uncommon in China, may surprise Chinese tourists and contribute to their sense of self-expansion. However, such nuanced experiences are challenging for the model to recognize without clear linguistic indicators.

In summary, the varying performance across different motivation factors is largely due to the presence or absence of explicit keywords and the subjective nature of certain experiences. The Transformer models perform better with factors that have clear, consistent linguistic patterns, while more abstract or subjective factors pose greater challenges.

Table 8. Review and Prediction Score

Motivation Factor	Review	Review (English)	Human-assigned Score	Prediction Score
Nature	花园很美，池塘里有鸳鸯，自由自在的游着，树木葱郁，树下让人觉得很凉爽，楼里是木质楼梯，很怀旧，室内展览了北海道的历史，有志愿者在面，给人亲切的感觉，让人很舒服自在。	The garden is beautiful, with mandarin ducks swimming freely in the pond. The lush trees provide a cool, refreshing shade. Inside the building, the wooden stairs give a nostalgic feel. There are exhibits on the history of Hokkaido, and the presence of volunteers creates a warm and welcoming atmosphere, making visitors feel comfortable and at ease.	3.00	3.13
Culture	这家小樽音乐盒堂位于日本北海道小樽北一硝子商业街尽头的童话十字路口街角，是一栋具有一百多年历史的二层砖木结构的老建筑，门口的一座蒸汽大座钟是其显着的标志。这家音乐盒堂主要展示、销售音乐盒，这些音乐盒形态各异、色彩丰富、大小不等，令人大开眼界。在二楼的廊道里，还有一间小型的留声机展室，里面摆放了各式各样的留声机和其他一些机械乐器。	The Otaru Music Box Hall is located at the end of Kitaichi Glass Street in Otaru, Hokkaido, Japan, at the Fairy Tale Crossroads. It is a two-story brick and wood building with over a hundred years of history. The prominent landmark is a large steam clock at the entrance. This music box hall mainly displays and sells music boxes of various shapes, colors, and sizes, which are truly eye-opening. In the hallway on the second floor, there is a small phonograph exhibition room displaying various phonographs and other mechanical musical instruments.	3.00	3.32
Stimulation	地铁出口到神宫门口还有一段步行距离。这段杳无人烟的路走得真不容易，昏暗的路灯映衬着两旁高耸的大树，透露着丝丝神秘与诡异，总觉得黑暗的丛林里有一双眼睛在凝视。	There is a walking distance from the subway exit to the entrance of the shrine. The path is desolate and walking it is not easy. Dim streetlights cast shadows on the tall trees on both sides, exuding a sense of mystery and eeriness. It always feels like a pair of eyes are staring from the dark forest.	3.00	1.75
Health	小动物们太可爱了，超萌，有机会还来	The little animals are so cute and adorable. I will definitely come back if I have the chance.	2.67	1.48
Self-exp.	地狱谷就在温泉酒店那，走近地狱谷就能闻到浓浓的硫磺味道;因为下雪的关系，走道有点滑，最好扶着扶手走，免得摔倒。足汤那上边有一片空地，适合一家大小玩耍	Hell Valley is right next to the hot spring hotel. As you approach Hell Valley, you can smell the strong sulfur odor. Due to the snow, the paths are a bit slippery, so it's best to hold onto the handrails to avoid falling. There is an open space above the foot bath area, suitable for family fun.	2.00	0.95

The ICC score in Table 7 also indicates that the motivation factors "Cultural observation," "Local communication," and "Experiencing nature" exhibit higher agreement, with ICC scores of 0.6361, 0.6870, and 0.6893, respectively. These high ICC scores reflect a strong correlation between human and model predictions for these categories, underscoring the model's ability to capture the context associated with these motivation factors. In contrast, "Stimulation" demonstrates the lowest agreement, with an ICC score of 0.3398, suggesting that this factor is more complex and harder for the model to predict accurately. Other factors such as "Health recovery," "Unexpectedness," and "Educating oneself" show moderate agreement, with ICC scores of 0.4911, 0.5341, and 0.4561, respectively. This pattern in the ICC scores mirrors the results observed in the R2 scores.

The variation in agreement across different factors highlights the importance of further refining the model for certain categories and suggests potential areas for future improvement. Overall, the RoBERTa model shows promise in automating the scoring process, with particular strengths in specific motivation factors. These results can provide a foundation for enhancing the model's accuracy and reliability, ultimately contributing to more efficient and effective analysis of tourist reviews.

Furthermore, we can use the predicted tourism motivation scores to analyze the focus points of tourists at each spot. This way, we can confirm that the preferences of tourists obtained by automatic scoring match those by human evaluators. We performed Principal Component Analysis (PCA), a statistical method that captures the most important features of the data by identifying principal components that maximize variance, effectively summarizing essential patterns. This analysis was applied to both the predicted and human-assigned scores for three tourist spots in Hokkaido to identify the attention-grabbing aspects of each spot. The results are shown in Figure 2. The horizontal axis represents the seven tourism motivation factors, while the vertical axis represents the values of the first principal component (Comp.1) weights. The one with the highest weights indicating the most important motivations for that spot.

The automatically scored results are similar to the human-assigned scores. The highest tourism motivation score for Asahiyama Zoo is related to nature, suggesting that Chinese visitors experience a sense of nature when surrounded by animals and natural scenery at the zoo. For Tanukikoji Shopping Street, the highest motivation is culture, indicating that Chinese tourists might perceive Japanese culture through shopping stores and restaurants. At Hokkaido Shrine, Chinese tourists feel the nature the most, which may be related to the beautiful natural environment of the shrine.

However, some differences exist between RoBERTa and human-assigned scores. The motivation factors of Stimulation, Local, and Health predicted by RoBERTa are lower than the human-assigned scores. This discrepancy may be due to the more diverse context and human feelings associated with these factors, making it more challenging for RoBERTa to accurately capture and predict the scores.

These findings are consistent with recent studies leveraging data analysis and deep learning methods in tourism research. Firstly, Mishra et al. (2021) and Gregoriades et al. (2023) demonstrated the value of deep learning and natural language processing techniques such as sentiment analysis and topic modeling to predict tourist revisit intentions, emphasizing the importance of combining advanced computational methods with traditional data sources to understand tourist preferences. Moreover, the results align with the research by Liu et al. (2023), which emphasized the utility of natural language processing tools in analyzing tourist behavior. The findings validate the proposed automated approach's ability to replicate human assessments of tourist motivations, particularly in scenarios involving large datasets. Moreover, the high performance in "Culture" and "Nature" factors reflects the explicit linguistic markers present in tourist reviews, supporting frameworks like Hayashi and Fujihara (2008). The ability of RoBERTa to effectively predict these motivation factors aligns with studies such as Yanuar and Shiramatsu (2020), where transformer-based models were applied to analyze linguistic patterns in customer feedback. These models demonstrated robustness in capturing explicit content, reinforcing the role of more advanced computational methods in supplementing conventional survey approaches.

In particular, the alignment with previous research underscores the potential of transformer models to reduce manual effort and provide scalable solutions in tourism analytics, which has been an issue found in many previous studies (Devesa et al., 2010), (Park and Yoon, 2009), (Su et al., 2020), (Chi and Phuong, 2022), (Annika Aebli and Taplin, 2022), (Hayashi and Fujihara, 2008), (M.Carvache-Franco et al., 2020), (Valverde-Roda et al., 2022), (Wen et al., 2019), (Jiang et al., 2020) and (Liu et al., 2023). This scalability is crucial for handling the increasing volume of unstructured data, such as online reviews, in tourism research. Furthermore, the ability to replicate human assessments suggests that these models can complement traditional methods, bridging gaps in time and resource constraints.

However, challenges in predicting abstract motivations highlight limitations similar to those identified by Alaei et al. (2019), where subjective sentiments were harder to quantify using automated approaches alone. The variability in how tourists express such motivations suggests that future studies should explore integrating additional data sources (e.g., images, videos) or advanced architectures like GPT-based models to enhance predictive accuracy. This limitation is consistent with findings in related NLP research, where abstract and subjective concepts often pose challenges due to their reliance on context and implicit expressions.

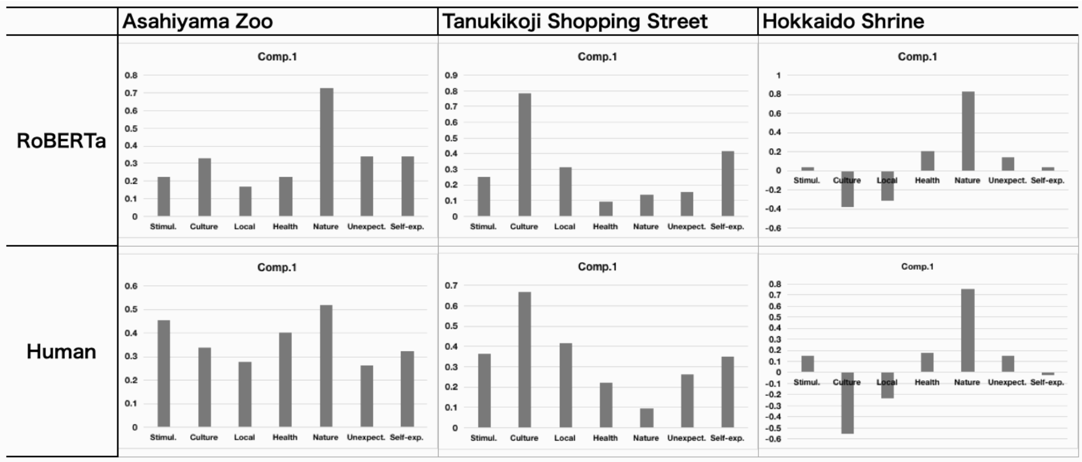

Figure 2. The PCA Results of Tourist Spots

5.2. Implications

The findings from this study have several significant implications for both the field of tourism research and the practical management of tourist destinations. Firstly, the use of advanced Transformer models such as BERT, RoBERTa, and ELECTRA for predicting tourism motivation factors from online reviews demonstrates the potential of natural language processing (NLP) techniques in tourism analytics.

Unlike traditional survey-based approaches, these models can process large-scale unstructured textual data, offering new opportunities to explore tourism dynamics. The findings reinforce the relevance of established frameworks, such as Hayashi and Fujihara (2008)'s categorization of motivations, while also identifying areas where automated methods can augment or refine these models. For example, the high predictive accuracy for motivations like "Culture" and "Nature" supports past research emphasizing the linguistic explicitness of these factors, as discussed by Pearce (2012) and Hu and Liu (2004).

By accurately identifying motivation factors from tourist reviews, stakeholders can gain deeper insights into what attracts tourists to specific spots, enabling more targeted and effective marketing strategies. For destination managers, the ability to automatically score and analyze tourist reviews provides a scalable and efficient tool for monitoring visitor satisfaction and preferences. This can lead to better-informed decisions regarding resource allocation, the development of new attractions, and improvements to existing facilities. The ability of models like RoBERTa to analyze motivations efficiently allows for real-time feedback, enabling more targeted marketing campaigns and personalized tourist experiences. For instance, insights into motivations such as "Culture" and "Nature" can guide promotional strategies for regions rich in cultural heritage or natural attractions, such as Hokkaido.

Moreover, understanding the specific aspects that appeal to tourists, such as cultural and natural attractions, allows for the creation of more tailored and engaging experiences, ultimately enhancing tourist satisfaction and promoting positive word-of-mouth. The high accuracy in predicting certain motivation factors, such as "Culture" and "Nature" suggests that these aspects are particularly significant for tourists. This insight can guide policymakers in prioritizing the conservation and promotion of cultural and natural resources, which are critical to maintaining a competitive edge in the tourism market.

However, the findings also reveal areas where automated methods lag behind traditional approaches. For instance, manual evaluations still outperform models in identifying highly subjective motivations, such as "Stimulation" and "Self-Expansion." This aligns with the broader understanding in NLP research that abstract concepts are harder to model accurately due to their contextual and personal nature Cambria and White (2014).

Lastly, automating the analysis of tourist feedback reduces reliance on time-intensive manual methods, like those used by Devesa et al. (2010), Park and Yoon (2009), Su et al. (2020), Chi and Phuong (2022), Annika Aebli and Taplin (2022), Hayashi and Fujihara (2008), M.Carvache-Franco et al. (2020), Valverde-Roda et al. (2022), Wen et al. (2019), Jiang et al. (2020) and Liu et al. (2023), freeing up resources for strategic decision-making. Destination managers can use these insights to design tailored itineraries, enhance visitor engagement, and address unmet tourist expectations. Utilizing advanced NLP techniques enables stakeholders in the tourism industry to gain meaningful insights into visitor preferences, facilitating more informed decisions and more effective marketing strategies. Further exploration is recommended to build on these findings and overcome the noted limitations, contributing to the development of tourism analytics and management.

6. Conclusions

In this study, we proposed and evaluated a method for automatically scoring tourism motivation factors from Chinese tourist reviews using Transformer models—BERT, RoBERTa, and ELECTRA. Our results indicate that RoBERTa performs best, particularly for the "Culture" and "Nature" categories, highlighting the potential of NLP techniques in tourism research and management. The high accuracy and strong R^2 scores for "Culture" and "Nature" suggest that these motivations contain clearer patterns, making them easier for models to learn. In contrast, the lower accuracy for more abstract factors like "Stimulation," "Health," and "Self-exp." underscores the challenges in predicting subjective motivations, emphasizing the need for further methodological improvements.

This research has important implications for both academia and industry. For researchers, it advances the application of NLP in tourism studies. For practitioners, automated analysis of tourist reviews provides valuable insights for destination management, marketing, and enhancing visitor satisfaction. However, limitations remain, including the small dataset, subjectivity in manual scoring, and difficulties in predicting abstract motivations. Addressing these challenges in future research will be key to refining the method and improving its generalizability.

In conclusion, this study demonstrates the feasibility of using Transformer models to analyze tourism motivations, reducing human labor while allowing for more scalable and objective analyses. By leveraging advanced NLP techniques, tourism stakeholders can better understand visitor preferences, leading to more data-driven decision-making and targeted marketing. Further research should expand on these findings to enhance tourism analytics and management.

7. Limitations and Future Directions

7.1. Limitations

Despite its promising results, this study has several limitations. First, it focused solely on tourist spots in Hokkaido, Japan, which may limit the method's applicability to other regions. Future research will expand to additional locations across Japan and internationally, incorporating reviews from a more diverse range of tourists to assess the model's generalizability. Second, the dataset used for training and evaluation consisted of only 500 reviews from Chinese tourists. This relatively small and specific sample may not fully capture broader tourist populations. Future studies should integrate larger, more diverse datasets to validate model performance across different cultural and linguistic contexts. Additionally, the manual scoring of reviews based on seven tourism motivation factors introduces a degree of subjectivity. Although averaging scores from six evaluators helps mitigate bias, inconsistencies may still arise. Establishing more standardized evaluation criteria could improve the reliability of training data and enhance model accuracy.

Moreover, certain motivation factors, such as "Stimulation," "Health," and "Self-exp.," are inherently abstract and subjective, making them harder for the model to predict. Tourists may express these motivations in highly variable ways, posing challenges for NLP-based analysis. Addressing this issue may require more advanced language models or additional contextual information to refine prediction accuracy. Finally, this study primarily relied on textual data from reviews, without incorporating other potential sources such as images, videos, or social media posts. A multimodal approach could offer deeper insights into tourist motivations and improve the robustness of the model's predictions. While this study demonstrates the potential of Transformer models in analyzing tourism motivations, further research is needed to overcome these limitations, refine methodologies, and enhance applicability across diverse contexts.

7.2. Future Directions

Building on this study's findings, several avenues for future research can enhance the automatic scoring of tourism motivation factors from tourist reviews.

- **Expanding the Dataset:** Future research should incorporate a larger and more diverse dataset, including reviews from various nationalities and regions. This would improve model generalizability and validate its performance across different contexts.
- **Exploring Other Transformer Models:** Further exploration of Transformer architectures, fine-tuning hyperparameters, and experimenting with training techniques could enhance performance. Addressing challenges in predicting abstract motivations like "Stimulation" and "Health" may require integrating multimodal data (e.g., images, videos) or leveraging advanced models like GPT-based architectures.
- **Longitudinal Studies:** Analyzing motivation shifts over time, especially in response to events like COVID-19, can provide valuable insights for destination managers adapting to changing tourist preferences.
- **Real-Time Analysis and Applications:** Developing real-time tools for automatic review analysis can provide immediate feedback, helping refine marketing strategies and improve tourist experiences.
- **Cross-Language and Cross-Cultural Studies:** Expanding research to multiple languages and cultural backgrounds would help identify universal and region-specific motivation factors, offering a more comprehensive understanding of global tourist behavior.
- **Other Tourism Motivation Frameworks:** While this study utilizes Hayashi et al.'s framework, integrating alternative models—such as the push-pull theory (Dann, 1981), hierarchy of needs, or socio-psychological motivations (Crompton, 1979)—could provide deeper insights into tourist decision-making.

While this study demonstrates the potential of Transformer models in predicting tourism motivations, refining automated approaches and expanding datasets will be crucial for developing robust, globally adaptable frameworks.

References

Alaei, A. R., Becken, S., and Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of travel research*, 58(2):175–191.

Annika Aebli, M. V. and Taplin, R. (2022). A two-dimensional approach to travel motivation in the context of the covid-19 pandemic. *Current Issues in Tourism*, 25(1):60–75.

Cambria, E. and White, B. (2014). Jumping nlp curves: A review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, 9(2):48–57.

Chen, Q., Liu, R., Jiang, Q., and Xu, S. (2024). Exploring cross-cultural disparities in tourists' perceived images: a text mining and sentiment analysis study using lda and bert-bilstm models. *Data Technologies and Applications*.

Chi, N. T. K. and Phuong, V. H. (2022). Studying tourist intention on city tourism: the role of travel motivation. *International Journal of Tourism Cities*, 8(2):497–512.

Clark, K., Luong, M.-T., Le, Q. V., and Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators.

Cohen, E. (1972). Toward a sociology of international tourism. *Social Research*, 39(1):164–182.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Crompton, J. L. (1979). Motivations for pleasure vacation. *Annals of Tourism Research*, 6(4):408–424.

Dann, G. M. (1981). Tourist motivation an appraisal. *Annals of Tourism Research*, 8(2):187–219.

Devesa, M., Laguna, M., and Palacios, A. (2010). The role of motivation in visitor satisfaction: Empirical evidence in rural tourism. *Tourism Management*, 31(4):547–552.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dunn Ross, E. L. and Iso-Ahola, S. E. (1991). Sightseeing tourists' motivation and satisfaction. *Annals of Tourism Research*, 18(2):226–237.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Gregoriades, A., Pampaka, M., Herodotou, H., and Christodoulou, E. (2023). Explaining tourist revisit intention using natural language processing and classification techniques. *Journal of Big Data*, 10(1):60.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1):1–12. PMID: 14741005.

Hayashi, Y. and Fujihara, T. (2008). 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.

Henseler, J., Ringle, C. M., and Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing*, pages 277–319. Emerald Group Publishing Limited.

Hsu, C. and Huang, S. (2007). Travel motivation: A critical review of the concept's development. *Tourism management: Analysis, behaviour and strategy*, pages 14–27.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Huang, S. and Hsu, C. H. (2009). Travel motivation: linking theory to practice. *International journal of culture, tourism and hospitality research*, 3(4):287–295.

Jiang, S., Scott, N., Tao, L., and Ding, P. (2020). Chinese tourists' motivation and their relationship to cultural values. In *Culture and Cultures in Tourism*, pages 202–214. Routledge.

JNTO (2024). 2024 visitor arrivals and japanese overseas travelers. `chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.jnto.go.jp/statistics/data/_files/20240821_1530-1.pdf`.

Koo, T. K. and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163.

Lin, P. M., Qiu Zhang, H., Gu, Q., and Peng, K.-L. (2017). To go or not to go: Travel constraints and attractiveness of travel affecting outbound chinese tourists to japan. *Journal of Travel & Tourism Marketing*, 34(9):1184–1197.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach.

Liu, Z., Masui, F., Eronen, J., Terashita, S., and Ptaszynski, M. (2023). A new approach to extracting tourism focus points from chinese inbound tourist reviews after covid-19. *Sustainability*, 15(11).

Martín, C. A., Torres, J. M., Aguilar, R. M., and Diaz, S. (2018). Using deep learning to predict sentiments: case study in tourism. *Complexity*, 2018(1):7408431.

M.Carvache-Franco, W.Carvache-Franco, O.Carvache-Franco, Hernández-Lara, A. B., and Buele, C. V. (2020). Segmentation, motivation, and sociodemographic aspects of tourist demand in a coastal marine destination: a case study in manta (ecuador). *Current Issues in Tourism*, 23(10):1234–1247.

Mishra, R. K., Urolagin, S., Jothi, J. A. A., Neogi, A. S., and Nawaz, N. (2021). Deep learning-based sentiment analysis and topic modeling on tourism during covid-19 pandemic. *Frontiers in Computer Science*, 3.

Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.

Park, D.-B. and Yoon, Y.-S. (2009). Segmentation by motivation in rural tourism: A korean case study. *Tourism management*, 30(1):99–108.

Pearce, P. (2012). *The Ulysses factor: Evaluating visitors in tourist settings*. Springer Science & Business Media.

Phillips, P., Zigan, K., Santos Silva, M. M., and Schegg, R. (2015). The interactive effects of online reviews on the determinants of swiss hotel performance: A neural network analysis. *Tourism Management*, 50:130–141.

Salim, E. and Ravanel, L. (2023). Last chance to see the ice: visitor motivation at montenvers-mer-de-glace, french alps. *Tourism Geographies*, 25(1):72–94.

Su, D. N., Nguyen, N. A. N., Nguyen, Q. N. T., and Tran, T. P. (2020). The link between travel motivation and satisfaction towards a heritage destination: The role of visitor engagement, visitor experience and heritage destination image. *Tourism Management Perspectives*, 34:100634.

The Asahi Shimbun (2024). Inbound tourists hit record high for first six months of year. `https://www.asahi.com/ajw/articles/15353063`.

The Japan Times (2024). Inbound tourism numbers hit record high, with japan set to achieve 2025 goal. `https://www.japantimes.co.jp/news/2024/04/17/japan/society/record-high-inbound-travelers/`.

UN Tourism (2023). International tourism and covid-19. `https://www.unwto.org/tourism-data/global-and-regional-tourism-performance`. Accessed on: May 1, 2023.

Valverde-Roda, J., Viruel, M., Castaño Prieto, L., and Sánchez, M. (2022). Interests, motivations and gastronomic experiences in the world heritage site destination of granada (spain): satisfaction analysis. *British Food Journal*, 125.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wen, J., Huang, S. S., and Ying, T. (2019). Relationships between chinese cultural values and tourist motivations: A study of chinese tourists visiting israel. *Journal of Destination Marketing & Management*, 14:100367.

Xiang, Z., Schwartz, Z., Gerdes, J. H., and Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44:120–130.

Yanuar, M. R. and Shiramatsu, S. (2020). Aspect extraction for tourist spot review in indonesian language using bert. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 298–302. IEEE.

Ye, Q., Zhang, Z., and Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications*, 36(3):6527–6535.

Ying, X. (2019). An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing.

Zeng, B. (2021). Pattern of chinese tourist flows in japan: a social network analysis perspective. In *Tourism Spaces*, pages 42–64. Routledge.

Zeng, B. and He, Y. (2019). Factors influencing chinese tourist flow in japan–a grounded theory approach. *Asia Pacific Journal of Tourism Research*, 24(1):56–69.

Zhang, Y. and Wallace, B. (2016). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.