Research on Tourism Destination Marketing Strategy Optimization based on Big Data

Ce Zheng*, Mingdi Gao

College of Finance and Economics, Qinghai University, Xining, China

*3556867627@qq.com

Abstract. The advent of big data technology has introduced a duality of opportunities and challenges to the marketing of tourist destinations. The utilisation of big data enables tourist destinations to differentiate themselves from their competitors and rapidly attain a prominent position in the market, thereby conferring substantial economic and social advantages. This paper mainly uses big data technology to optimize the marketing strategy of tourist destinations. First, data mining technology is used to extract key information such as tourist behavior and preferences from multi-source data such as social media and travel platforms. Secondly, combined with the forecasting model, the dynamic changes of tourism demand are analyzed, the potential tourist groups are accurately predicted, and the factors affecting the attractiveness of the destination are identified. This study also applied machine learning algorithms to optimize precision marketing strategies to improve tourist engagement and promote the promotion effect of tourist destinations. In the experimental analysis part, the effectiveness of the optimization strategy is verified by comparing the actual cases and performance indicators.

Keywords: Personalized tutoring system, Neural network, Optimization method, Learning path.

1 Introduction

The rapid development of the Internet and the advent of the era of mass tourism have resulted in a notable shift in the way tourists obtain tourism information. In this new landscape, traditional media such as newspapers, magazines and radio have faced challenges in maintaining their competitive advantage, audience reach and scope of influence. The Internet marketing system, which is based on big data technology, is currently undergoing a process of maturation, resulting in the formation of a comprehensive marketing channel and system [1]. The dissemination of information is significantly faster and more extensive when using big data technology, which also enhances marketing effectiveness through precise user analysis and targeted personalised communication. Furthermore, the utilisation of big data-driven creativity, trending topic communication and visual experience has introduced a novel interactive experience for the audience, establishing itself as a highly valuable tool in the marketing of tourist destinations.

K. Zhang et al. (eds.), *Proceedings of the 5th International Conference on Economic Management and Big Data Application (ICEMBDA 2024)*, Advances in Economics, Business and Management Research 313,
https://doi.org/10.2991/978-94-6463-638-3_31

As the tourism industry enters a phase of accelerated growth, the marketing of tourist destinations has increasingly become a strategic priority. The extensive utilisation of big data technology has opened up new avenues for the marketing of tourist destinations. By employing data mining and user behaviour analysis, it is possible to reach potential tourists with greater precision and to enhance the brand awareness and market influence of tourist destinations [2]. Nevertheless, numerous challenges persist in the practical implementation of existing big data-driven marketing strategies. To illustrate, the integration and sharing of data resources is inadequate, the evaluation system for marketing performance is imperfect, there is a lack of comprehension and utilisation of the fundamental value of big data, and the innovation of marketing forms and content is an urgent necessity [3]. It is therefore evident that the optimisation of big data technology applications in the marketing of tourist destinations, particularly the enhancement of integrated marketing capabilities, the establishment of a rigorous performance evaluation system and the promotion of innovative marketing content and formats, will become pivotal strategies for tourist destinations to distinguish themselves in the context of intense competition.

The conventional marketing models are based on experience and a restricted data foundation, which makes it challenging to accurately comprehend the shifts in tourist demand. The integration and analysis of multi-channel data, including social media, online reviews, search engines and travel platform data, enables the application of big data technology to provide tourism destinations with deep market insights. This allows for the efficient allocation and accurate delivery of marketing resources [4]. Furthermore, big data can forecast the evolving patterns of travel demand, enabling destinations to enhance the provision of products and services and bolster their market competitiveness [5].

The objective of this study is to optimise the marketing strategy of tourism destinations through the utilisation of big data technology, with a view to enhancing the effectiveness of marketing activities through data-driven methods, thereby augmenting the brand value and market influence of the aforementioned destinations. The research will employ data mining, predictive analysis, sentiment analysis and other technologies, in conjunction with case analysis and experimental verification, to investigate the potential applications and optimisation pathways of big data technology in the field of tourism destination marketing.

2 Related Work

In the context of marketing tourist destinations, the promotion of said destinations constitutes a pivotal element of the 4P theory. The 4P model was initially proposed by Jerry McCarthy [6] in the 1960s and comprises four fundamental elements: The four core elements of the 4P theory are product, price, place and promotion. Subsequently, the 4P theory has undergone further development and has become the fundamental framework of marketing mix theory. As a crucial component of the marketing strategy, the promotional aspect has a direct impact on the overall effectiveness of the marketing initiative.

In their analysis and comparison, McKay et al. have determined that the form of advertising is a significant determinant of advertising effectiveness.

Willians [7] posited that the initial phase of destination positioning entails the creation of an image of the destination that is not only based on its intrinsic characteristics but also takes into account the needs and preferences of the target market. In their analysis of the market positioning of different types of Korean golf tourists on tourist destinations, Kim et al. [8] considered the tourists' perception of the destinations in question. Their findings indicated a similarity in the perception image of these tourists to the destination. Mccleary et al. [9] identified two principal categories of factors influencing the image of tourist destinations external stimuli and the intrinsic characteristics.

3 Methodologies

3.1 Data Mining

Data mining is a foundational step of the entire model with the goal of extracting key information from multi-source data. Multi-source data includes social media, travel platforms, online reviews, and more. In order to effectively extract key information on tourist behavior and preferences, Term Frequency-Inverse Document Frequency is used (TF-IDF) analyzes visitor reviews and text data in social media to extract trending topics and keywords. The formula is represented as Equation 1.

$$TF - IDF(t,d) = TF(t,d) \times IDF(t), \tag{1}$$

where $TF(t,d)$ represents the frequency of the word t in document d, and $IDF(t)$ represents the inverse document frequency of the word t, which is used to measure the rarity of the word in all documents, and is calculated as Euation 2.

$$IDF(t) = log\frac{N}{|\{d \in D : t \in d\}|}. \tag{2}$$

The clustering algorithm was used to classify and group the behavior of tourists. K-Means is a commonly used clustering algorithm whose goal is to minimize the squared error within a cluster, denoted as Equation 3.

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 , \tag{3}$$

where C_i is the i-th cluster and μ_i is the centroid of the cluster.

3.2 Travel Demand Forecast

The time series prediction model is employed for the purpose of forecasting the evolving dynamics of tourism demand, as evidenced by an analysis of historical data. In this study, the autoregressive integrated moving average (ARIMA) model was employed for time series analysis with the objective of forecasting future trends in tourism demand. The ARIMA model is expressed as Equation 4.

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t, \tag{4}$$

where y_t is the time series value, ϕ and θ are the autoregressive coefficients and moving average coefficients, respectively, and ϵ_t are the error terms. HDFS (Hadoop Distributed File System) is used to store large-scale data, which is highly fault-tolerant and scalable, and is suitable for large-scale batch processing tasks. Use the MapReduce programming model to work with large data sets. Split complex tasks into multiple Map and Reduce steps and execute them in parallel to reduce the computing load on a node.

The application of tourism demand forecasting techniques enables the identification of potential groups of tourists. In order to achieve accurate population prediction, a logistic regression model is employed, which is primarily utilised for dichotomous problems, such as determining the level of interest among tourists in a specific destination. The formula used to calculate the predictive probability for the model is expressed as Equation 5.

$$P(y = 1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots \beta_n x_n)}}, \tag{5}$$

where β is the parameter of the model and x is the input feature vector.

Following the identification of potential groups of tourists, machine learning algorithms are employed to optimise precision marketing strategies. In this study, the eXtreme Gradient Boosting (XGBoost) algorithm was employed with the objective of enhancing the efficacy of marketing strategies. XGBoost is an ensemble learning method that optimises the objective function through the boosting of a tree model, which can be expressed as Equation 6.

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where $l(\cdot)$ is the loss function and $\Omega(\cdot)$ is the regular term, and the purpose is to control the complexity of the model to prevent overfitting. Set the producer and consumer of Kafka to write data from data sources to Kafka topics in real time. Configure a partition policy for Kafka to partition data based on data type and timestamp to improve data transmission efficiency. Use Kafka's replication and Log compaction features to ensure high availability and consistency of data transfer.

4 Experiments

4.1 Experimental Setup

In the experimental setup, the data is primarily sourced from the publicly available TripAdvisor Dataset, which provides a multi-source platform comprising guest ratings, reviews, and booking behaviour, thus illustrating how travellers experience and rate a destination. The multi-source data set encompasses a range of travellers' behaviours, preferences, reviews and points of interest, offering insights that can inform the optimisation of marketing strategies for tourist destinations.

4.2 Experimental Analysis

Log loss is an important metric to measure the uncertainty of classification models, especially for probabilistic output models. It evaluates the degree of deviation between the predicted probability of the model's output and the true classification labels. Figure 1 shows a comparison of three different methods, the 4P theory, the Destinations method, and the Log Loss of Ours, with the abscissa being the number of features. It can be seen that with the increase of the number of features, the log loss of each method gradually increases, which is consistent with the trend that the loss increases with the increase of the number of features. This graph shows the variation of the losses of the three methods with different numbers of features.

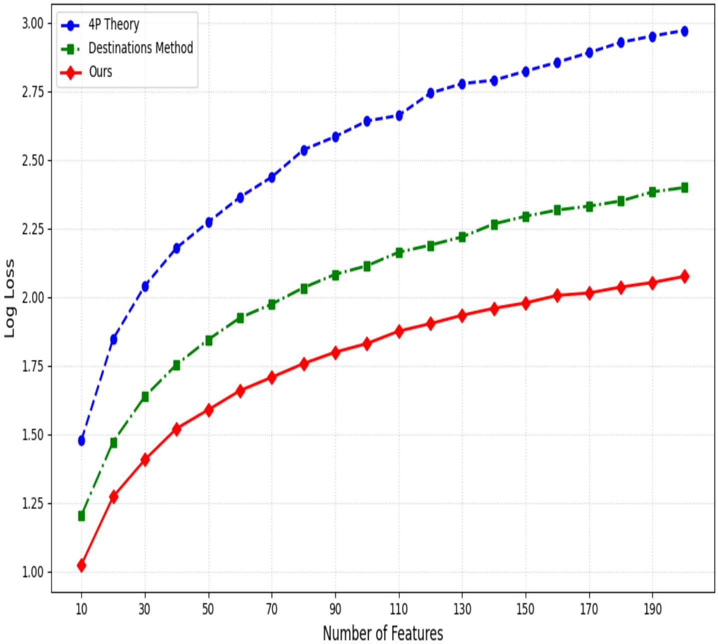

Fig. 1. Log Loss Comparison of Different Methods vs. Number of Features.

The Fowlkes-Malallows index is an evaluation index used to evaluate the effectiveness of clustering algorithms, especially for the task of clustering marketing customer groups. It measures the harmonized average of accuracy and recall between classification and real labeling.

Figure 2 shows a comparison of the Fowlkes-Mallows exponents for three different methods, with the abscissa being the regularization parameter (λ). With the change of regularization parameters, the Fowlkes-Mallows index of different methods performs differently in cluster analysis, and the effect of each method can be compared more intuitively through this graph.

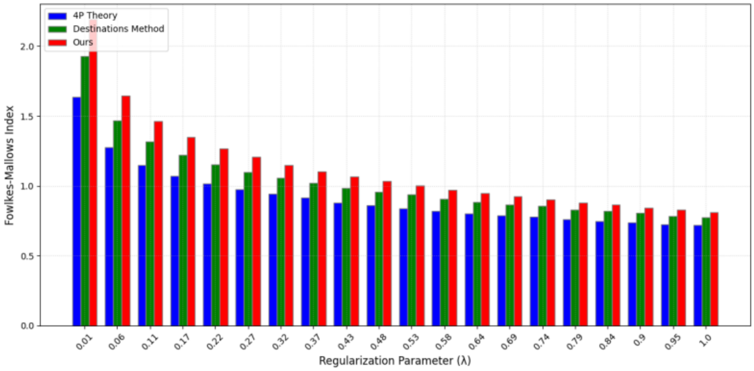

Fig. 2. Fowlkes-Mallows Index Comparison of Different Methods.

The Shapley value evaluates the contribution of each input feature to the model's prediction outcome. In big data marketing analytics, Shapley values can be used to assess the importance of each feature in precision marketing. Figure 3 shows a comparison of Shapley values for three different methods. Boxplots provide the distribution of each method on Shapley values, including medians, interquartile ranges, and possible outliers, allowing for a visual comparison of the performance of each method in terms of feature contribution.

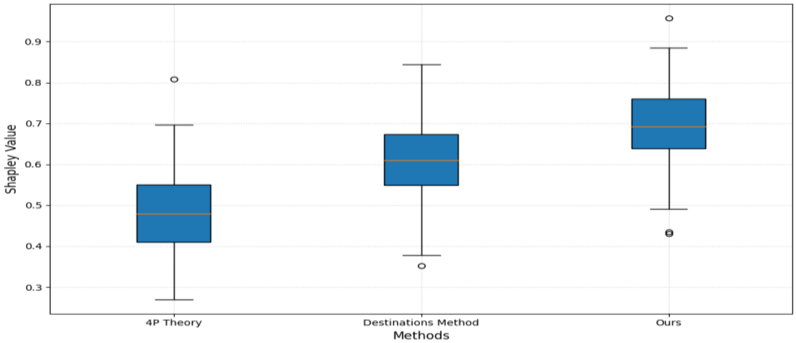

Fig. 3. Shapley Value Comparison of Different Methods.

5 Conclusion

In conclusion, our method has outperformed the traditional 4P theory and the Destinations method on multiple evaluation indicators. In the comparison of Log Loss and Fowlkes-Malallows, our method showed better accuracy and clustering performance, while in Shapley value analysis, our method showed a higher feature contribution, proving its effectiveness in optimizing precision strategies. Future work could further explore more sophisticated machine learning models, as well as incorporate a wider variety of data sources including real-time social media stream data, to improve dynamic responsiveness and personalized service levels of destination marketing strategies.

References

1. Arbolino, Roberta, et al. "Multi-objective optimization technique: A novel approach in tourism sustainability planning." Journal of Environmental Management 285 (2021): 112016.
2. Kumar, Prem. "Digital marketing in hospitality and tourism." University of South Florida (USF) M3 Publishing 17.9781732127593 (2021): 3.
3. Xie, Gang, Yatong Qian, and Shouyang Wang. "Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach." Tourism Management 82 (2021): 104208.
4. Ivars-Baidal, Josep A., et al. "Measuring the progress of smart destinations: The use of indicators as a management tool." Journal of Destination Marketing & Management 19 (2021): 100531.
5. Gao, Ying, et al. "Service design for the destination tourism service ecosystem: a review and extension." Asia Pacific Journal of Tourism Research 27.3 (2022): 225-245.
6. MacKay, Kelly J., and Malcolm C. Smith. "Destination advertising: Age and format effects on memory." Annals of Tourism Research 33.1 (2006): 7-24.
7. Uysal, Muzaffer, Joseph S. Chen, and Daniel R. Williams. "Increasing state market share through a regional positioning." Tourism Management 21.1 (2000): 89-96.
8. Kim, Samuel Seongseop, Hejin Chun, and James F. Petrick. "Positioning analysis of overseas golf tour destinations by Korean golf tourists." Tourism Management 26.6 (2005): 905-917.
9. Baloglu, Seyhmus, and Ken W. McCleary. "A model of destination image formation." Annals of tourism research 26.4 (1999): 868-897.