Guest editorial

1917

Guest editorial

Big data analytics and forecasting in hospitality and tourism

Introduction

With the rapid development of internet technology, the hospitality and tourism sector has joined the band of industries using big data analytics to enhance their regular operations. Modern information and communication technologies are generating a large amount of data about tourist and industry activities. Merchants and agencies have discovered methods for using these data resources to interact with potential consumers at every stage of their travels and use visitor statistics to make decisions and understand forecasts.

Over the last decade, interest has continued to grow in big data analytics and forecasting in hospitality and tourism. However, the challenge of extracting valuable knowledge from big data remains for practitioners and academics. This special issue offers the latest research from the field and suggestions for future research directions. It includes 17 studies around four themes: forecasting with big data, online review analytics, methodological innovation and application and using big data analytics to understand hotel and tourism phenomena and mechanisms.

Forecasting with big data

Hotel and tourism forecasting has caught the attention of industry practitioners because of industry dynamics, the need for data-driven decision-making and the development of big data (Li and Wu, 2019; Wu *et al.*, 2017). In this special issue, six studies focus on hotel and tourism forecasting based on big data. All of these studies use high-frequency data, but they emphasize different types of innovation that involve methodology, data and context. These studies shed new light on hotel and tourism forecasting with big data and propose possible directions for future research.

Liu *et al.* (2021) make the first attempt to introduce the LASSO-MIDAS model to tourism demand nowcasting. Nowcasting is an important topic in the field, focusing on predictions about the current and recent past. The release of statistical data is often delayed for a short period, such as a couple of weeks or months. With the development of high-frequency big data, tourism nowcasting has piqued the interest of industry and academic leaders. Mixed frequency models have been applied to tourism demand forecasting in previous studies, but Liu *et al.* (2021) are the first team to use the LASSO-MIDAS model for tourism nowcasting. They adopt visitor arrivals in Hong Kong as empirical cases. They also integrate daily Baidu Index data with monthly official statistical data into their proposed mixed frequency model is effective at producing accurate, reliable tourism nowcasts.

Tian *et al.* (2021) examine the forecasting ability of big data generated by search engines and social media. They used daily tourism demands for a visitor attraction in China as their case. In addition to the search engine data, they used information from three popular social media platforms in China (i.e. WeChat, Douyin and Weibo) and compared the number of

Welcome to the *International Journal of Contemporary Hospitality Management*'s special issue on Big Data Analytics and Forecasting in Hospitality and Tourism. I would like to sincerely thank our Guest Editors Professor Doris Chenguang Wu, Dr Ji Wu and Professor Haiyan Song for putting together this very strong and timely special issue. The articles included in this special issue should be well received by scholars, students and practicing managers in our field.



International Journal of Contemporary Hospitality Management Vol. 33 No. 6, 2021 pp. 1917-1921 © Emerald Publishing Limited 0959-6119 DOI 10.1108/IJCHM-06-2021-035 posts, followers, reads, comments, likes and forwards. The results show that search engine data are highly effective at decreasing forecasting error, whereas the forecasting ability of social media data is limited. The focus of this study is structured big data. In the future, researchers should consider exploring unstructured big data, such as texts or pictures, for hotel and tourism forecasting.

Ramos *et al.* (2021) adopt location data derived from technical communications between a public Wi-Fi network and mobile devices to conduct daily visitor forecasting in Palma, the capital of Mallorca. The big data from Wi-Fi networks provide insights into visitor movements from a geographic perspective and emphasize specific locations, such as a city, tourist attractions or beach areas. The results support the use of geographic location data from Wi-Fi networks to perform visitor forecasting. The authors also compare a support vector machine, ANN with multilayer perception, a recurrent neural network, state-space ARIMA and multiplicative state-space ARIMA with a long short-term memory model, which they find to be the most valuable in their case.

Ampountolas and Legg (2021) focus on hotel occupancy forecasting based on text analytics, using data from a hotel chain in the USA. They analyze a new data type, namely, text from social media (i.e. Twitter and SocialMention), to capture tourists' sentiment information. From a methodological perspective, the authors' proposed segmented machine learning gradient boosting approach improves forecasting accuracy, stability and robustness. The findings provide hotel practitioners valuable guidance on the formulation of an operational strategy for hotel demand forecasting revenue management.

Hu *et al.* (2021) contribute to the literature on tourism forecasting using search engine data. They examine and compare the forecasting ability of search queries via mobile devices and PCs. The empirical results of two tourist attractions in China show that PC search queries can improve tourism demand forecasting accuracy, whereas mobile device queries do not enhance performance. The authors offer evidence-based insights into tourism forecasting using search engine data.

Qiu *et al.* (2021) forecast hotel occupancy during periods of social unrest. In this study, crisis severity, which is captured and measured by social media data, is included in the modeling process to forecast hotel demand during a crisis. The social unrest in Hong Kong in 2019 is used as the case. The findings extend the application of big data in hospitality forecasting by showing that the proposed forecasting framework is effective for hotel forecasting during a crisis. Forecasting during unusual scenarios, such as a pandemic, is an important issue, and the authors encourage researchers to pay more attention to this topic in the future.

Online review analytics

User-generated content (UGC) is a primary category of tourism-related big data. Hospitality and tourism online reviews influence customer decision-making and are valuable information-based assets. Tourists examine online reviews in two steps. They browse through review ratings and process the information within the reviews. Then they decide whether to use the information based on the perceived helpfulness of the review. Focusing on the two steps of processing UGC, several studies have explored two fundamental research questions in the context of online reviews. First, which factors determine the rating of a review? Second, which factors determine the perceived helpfulness of a review?

To answer the first research question, Lai *et al.* (2021) examine the relationship between customer sentiment and online hotel ratings. They measure customer sentiment by gathering sentiment scores based on customer-generated reviews. Using threshold regressions, they establish that there is an asymmetric relationship between customer

1918

IICHM

33.6

sentiment and online hotel ratings. To address the second research question, Mariani and Borghi (2021) explore how the presence and depth of environmental disclosure in online reviews affect review helpfulness. By combining textual analysis and a negative binomial regression method, they find a significant positive relationship between environmental disclosure and review helpfulness. Lee *et al.* (2021) use several popular machine learning algorithms to predict review helpfulness and find XGBoost's performance to be best. They also reveal that reviewers' credibility is a fundamental factor in determining review helpfulness. In addition to customers' online reviews, employee reviews can provide innovative insights into tourism management. Shi and Chen (2021) investigate employee satisfaction in the hotel industry. By collecting employee reviews and using structural topic modeling and sentiment analysis, they find that workplace bullying, dirty work, organizational support and career advancement heavily influence employee satisfaction.

Methodological innovation and application

Current technologies for analyzing and converting big data into meaningful information for supporting decision-making are not widely available outside of large corporations. The massive volume and variety of big datasets extend beyond the scope of most common analytical tools. Most of the existing methods for analyzing big data are predefined in studies. They are not meant to develop a general understanding of tourists' movements, interests and experiences. Formal methods for analyzing big data for decisional support have yet to be developed, particularly in the tourism sector.

Four studies go beyond previous research by using design science methodology to develop and evaluate new analytical methods for turning unstructured big data into tourism-focused terms. Han *et al.* (2021) develop a textual analysis method called Tourism2vec to mine tourists' spatial-temporal behaviors by adapting a traditional word embedding technique. Their empirical analysis shows that Tourism2vec can help find tourist attractions. Li and Beldona (2021) propose a method for extracting tourists' revisit intentions from their online reviews automatically with a machine learning technique. Identified revisit intentions can serve as a novel measure for relationship marketing. Qin *et al.* (2021) explore and visualize the evolution of spatial-temporal patterns in China's hotel industry by using the space-time cube model. The exploratory spatial-temporal data analysis method may be a good tool for tourism and hotel management.

Finally, Lin *et al.* (2021) investigate how social media data collection methods can substitute or complement traditional survey methods. The authors compared the metrics and biases of social media analytics with that of a traditional visitor intercept survey about a tourist destination image. Their results show that social media data can serve as a complementary source of information to survey data.

Tourism phenomena and mechanisms from the big data perspective

As multidimensional mass data acquisition and storage technology have matured, structured and unstructured user behavior data have increased. Tourist behavior theory and mechanism research have become more important. Big data analytics can reveal the rules behind user behaviors and help researchers examine the relationships in hotel and tourism phenomena. Three studies on big data analysis focus on theories and phenomena in the fields of hospitality and tourism. Yi *et al.* (2021) consider the relationships between hotels and Airbnb by investigating the multidimensional effects of neighborhood hotels on Airbnb listings' popularity. They find both cooperative and competitive relationships, which vary across geographical locations. Lee and Kim (2021) use the decision tree method to analyze which determinants influence hotel bookings and consider the geographical distance

Guest editorial

1919

IJCHM 33,6 between tourists' origins and destinations. Their results show that the choice of packaged products is the foremost determinant for longer-stay hotel consumers. Finally, Azevedo (2021) uses the Flickr application to identify photos of a specific destination and treats the photos as a proxy for estimating the zonal visitation rate for the destination.

1920 Conclusion

The 17 studies in this special issue contribute to different aspects of hospitality and tourism. The authors discuss forecasting with big data, online review analytics, methodological innovation and application and using big data analytics to understand hotel and tourism phenomena and mechanisms. The availability of big data has grown rapidly over the last decade and has become significant to the supply and demand sides of hospitality and tourism. Industry leaders and scholars are increasingly interested in how to extract valuable information from big data. Thus, continuous research is encouraged. First, in addition to structural data types, more unstructured data should be adopted for hotel and tourism forecasting. They can come from multiple sources, such as textual and picture mining, from the supply side (i.e. industry and government) and the demand side (i.e. social media). Second, novel modeling techniques, such as deep learning and ensemble models, should be combined with big data for probabilistic and scenario forecasting to enhance accuracy forecasts, especially during periods of significant uncertainty. Third, big data analytics can go beyond forecasting and business intelligence; the methods offer more than improved predictions and visualization. This type of analytics goes beneath the surface of the data to link a set of explanatory variables to a tourism phenomenon or outcome. Big data analytics can help managers and researchers better understand phenomena and theories in the tourism industry. Finally, methodologically speaking, big data enable us to leverage predictions/forecasting and causal analysis (Agarwal and Dhar, 2014). State-of-the-art big data research can draw on a variety of techniques, such as machine learning, classical statistics and econometrics, to design methods and experiments for testing hypotheses, developing new theories and creating large-scale value for tourism research.

Doris Chenguang Wu and Ji Wu

School of Business, Sun Yat-Sen University, Guangzhou, China, and Haiyan Song School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong, China

References

- Agarwal, R. and Dhar, V. (2014), "Big data, data science, and analytics: the opportunity and challenge for is research", *Information Systems Research*, Vol. 25 No. 3, pp. 443-448.
- Li, G. and Wu, D.C. (2019), "Introduction to the special focus: tourism forecasting new trends and issue", *Tourism Economics*, Vol. 25 No. 3, pp. 305-308.
- Wu, D.C., Song, H. and Shen, S. (2017), "New developments in tourism and hotel demand modeling and forecasting", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 1, pp. 507-529.

About the authors Doris Chenguang Wu, PhD, is a Professor of Tourism Forecasting in the School of Guest editorial Business at the Sun Yat-Sen University, China. Her research interests include tourism demand forecasting and tourism impact analysis.



Ji Wu, PhD, is an Associate Professor of Management Science in the School of Business at the Sun Yat-Sen University, China. Her research interests include big data analytics and consumer behavior management. Ji Wu is the corresponding author and can be contacted at: wuji3@mail.sysu.edu.cn



Haiyan Song, PhD, is Chan Chak Fu Professor in International Tourism in the School of Hotel and Tourism Management at The Hong Kong Polytechnic University. His research interests are in tourism demand modeling and forecasting, tourism supply chain management and wine economics.

1921