International Journal of Hospitality Management 47 (2015) 14–24

Contents lists available at ScienceDirect

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhosman

Hotel location evaluation: A combination of machine learning tools and web GIS

Yang Yang[a,1], Jingyin Tang[b,2], Hao Luo[c,*], Rob Law[d,3]

[a] School of Tourism and Hospitality Management, Temple University, Philadelphia, PA 19122, United States
[b] Department of Geography, University of Florida, Gainesville, FL 32611, United States
[c] Sun Yat-sen Business School, Sun Yat-sen University, Guangzhou, Guangdong 510275, China
[d] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

ARTICLE INFO

Keywords:
Web GIS
Hotel location
Spatial decision making
Machine learning

ABSTRACT

The need for a reliable, unbiased, and objective assessment of hotel location has always been important. This study presents a new approach to evaluate potential sites for proposed hotel properties by designing an automated web GIS application: Hotel Location Selection and Analyzing Toolset (HoLSAT). The application uses a set of machine learning algorithms to predict various business success indicators associated with location sites. Using an example of hotel location assessment in Beijing, HoLSAT calculates and visualizes various desirable sites contingent on the specified characteristics of the proposed hotel. The approach shows considerable potential usefulness in the field of hotel location evaluation.

1. Introduction

Evaluating and assessing a location site is an important aspect when establishing a new hotel to secure long-term business prosperity. Once located, it is nearly impossible for the hotel to relocate; owing a considerable sunk cost for hotel establishment and operation. A large body of literature has been devoted to deciphering hotel location patterns and mechanisms through different perspectives; and several models have been advocated including mono-centric models (Egan and Nield, 2000; Yang et al., 2012), agglomeration models (Canina et al., 2005; Kalnins and Chung, 2004), and multi-dimension models (Baum and Haveman, 1997; Urtasun and Gutiérrez, 2006). Various empirical efforts have also been conducted to recognize those factors as influencing decisions on where to locate new hotels such as star rating, years after opening, service diversification, ownership, agglomeration effects, public service infrastructure and transport accessibility (Yang et al., 2014).

In conventional hotel location prediction models, linear regression has been dominantly used to recognize the preferred site with considerable potential (Yang et al., 2014). However, the limitations associated with simple linear regression result in several common drawbacks such as poor prediction accuracy because of the over-fitting problem, inability to adequately incorporate interactions and nonlinearity among variables, and failure to consider spatial dependency and spatial heterogeneity. Furthermore, although numerous theoretical and empirical models have been constructed by scholars, the majority are fairly complicated and difficult for practitioners to understand because of the heavy use of sophisticated mathematical or statistical knowledge. Therefore, the practical value and applicability of these models become relatively limited, and this phenomenon impedes the wide application of these "scholarly" models to solve real hotel location problems (Yang et al., 2014). To better facilitate the decision making of hotel location selection, a huge demand has arisen for the ability to transfer knowledge from these "scholarly" location models to knowledge with great practical values.

Over the past decade, numerous applications in information technology (IT) have been introduced in the hospitality industry to handle routine operational problems (O'Connor and Murphy, 2004; Rob and Giri, 2005). These applications have been used to improve employee productivity and increase revenue (Siguaw et al., 2000), facilitate information exchange between hotels and customers (Chung and Law, 2003), and propose appropriate room pricing across different distribution channels (O'Connor, 2003). Despite the prevalence of IT applications, few have incorporated geographical information system (GIS) technology in hospitality management. Considering that locational information is critical in

* Corresponding author. Tel.: +86 20 8411 2561; fax: +86 20 8411 3687.
E-mail addresses: yangy@temple.edu (Y. Yang), sugar1987cn@gmail.com (J. Tang), luohao6@mail.sysu.edu.cn (H. Luo), rob.law@polyu.edu.hk (R. Law).
[1] Tel.: +01 215 204 8701; fax: 01 215 204 8705.
[2] Tel.: +01 352 294 7521; fax: 01 352 392 8855.
[3] Tel.: +852 3400 2181; fax: +852 2362 9362.

http://dx.doi.org/10.1016/j.ijhm.2015.02.008
0278-4319/

determining performance and operation of hotels and in targeting specific market segments (Egan and Nield, 2000), GIS applications from a spatial perspective impart valuable insights on the competitive advantages/disadvantages of hospitality properties (Kisilevich et al., 2013).

To echo Yang et al.'s (2014) plea for more research attention to the operational hotel location models and address the research gap in hotel location analysis and IT applications, we design and present a web-based GIS application for hotel location analysis. By utilizing several machine learning models such as projection pursuit regression, artificial neural network, support vector regression, and boosted regression, we are able to overcome various data-related problems in the simple linear regression models, such as the non-linearity of relationships, the presence of noise, and the absence of necessary information on function form. As a web GIS application, HoLSAT enables the visualization of location prediction based on the calibrated models, and allows scenario analysis contingent on different characteristics of proposed hotels. Based on the characteristics specified by the users, the program can automatically visualize the predicted indicator of business success for each location following the calibrated models. Therefore, as an IT application coupled with machine learning tools, HoLSAT greatly improves the decision-making capability of hotel investors in assessing the location site of the proposed hotels.

The rest of this paper is organized as follows: Section 2 reviews the relevant literature on GIS applications in hospitality management, especially those that are used in location analysis; and Section 3 introduces the technical design and programing framework of the web GIS application. Section 4 elaborates different algorithms embedded in the system to predict the future performance of each location for the proposed hotel; Section 5 presents the estimation and cross-validation results of various algorithms based on a sample of star-rated hotels in Beijing, as well as a detailed demonstration of how this system works; conclusions are drawn in Section 6.

2. Literature review

2.1. Hotel location models

In a systematic literature review, Yang et al. (2014) discussed various theoretical, empirical, and operational models used in hotel location studies. Based on their definition, the theoretical model provides theoretical insight to elucidate the hotel location process under certain conditions. Ashworth and Tunbridge (1990) proposed a tourist–historic city (THC) model and described a comprehensive typology of hotel locations within a historical city, including traditional city gates, railway station/approach roads, main access roads, "nice" locations, transition zones and urban periphery on motorway, and airport transport interchanges. Using the THC model, Bégin (2000) studied the hotel location pattern in Xiamen, China. The results highlighted a clustering of cheap/downscale hotels in the historical center as well as new hotel constructions in the transition zone between the old downtown and the emerging CBD. Egan and Nield (2000) proposed a mono-centric model to explain the spatial hierarchy of hotels in terms of the distance to the city center: luxury hotels choose central locations, whereas budget hotels locate at the edge of the city. They ascribed this result to the revenue associated with locations and affordability of hotels. Shoval (2006) discovered the hotel location hierarchy shaped by two geographies of lodging demand in Jerusalem, including the hotel area for individual tourism and that for organized tourism. Yang et al. (2012) empirically test the mono-centric hotel location model with historical hotel entrance data in Beijing. They suggested several factors explaining the mono-centric pattern, such service diversification, ownership, agglomeration effect, public service infrastructure, and road accessibility. Moreover, to understand the co-location of heterogeneous/homogeneous hotels, a set of hotel location models focus on the agglomeration process of new hotel entrants (Baum and Haveman, 1997; Ingram and Inman, 1996; Kalnins and Chung, 2004). Gains and losses stemming from location clustering depend on the relative strength and nature of agglomeration and/or competition effects.

To define the rule and criteria used in evaluating the potential site for new hotel entrants, different methods have been applied. First, the decision rule can be extracted by observing and estimating the location choice of existing hotels (Canina et al., 2005; Yang et al., 2012). In this choice model, the probability of choosing a specific site can be defined as a function of hotel characteristics and location attributes associated with each alternative site. Second, superior locations can be underscored by studying the historical data of existing hotels' performance. An ideal location is regarded to be associated with high RevPAR (Tsang and Yip, 2009), high productivity/efficiency (Assaf and Agbola, 2011), high room rate (Enz et al., 2008), high occupancy (Jeffrey et al., 2002), and low failure rate (Ingram and Baum, 1997). Yang et al. (2014) conducted a comprehensive review on past literature of hotel location analysis, and several types of location associated with better business performances were highlighted (p. 216–218). The results from these empirical models enable hoteliers to apply pre-existing hotel location rules to make operational hotel location decisions by statistical prediction (Yang et al., 2014). To reduce the estimation bias from the conventional prediction methods, Biemer and Kimes (1991) proposed a refined hotel location prediction model using a three-step bootstrap procedure. Moreover, for prediction purposes, a cross-or external validation is necessary to test the performance of out-of-sample prediction and avoid the over-fitting problem of estimates (Biemer and Kimes, 1991; Kimes and Fitzsimmons, 1990). Lastly, the validity of statistical prediction is rooted in the assumption of stability, which states that market and economic conditions are stationary over a reasonably long period. Major structural changes in the market and in economy tend to render substantial bias in this statistical prediction.

Another way to obtain the hotel location rule is from individual evaluation (Yang et al., 2014). Unlike statistical prediction from existing hotels' historical data, the individual evaluation relies on the knowledge and insights from surveyed hoteliers/investors (from the supply side) and potential guests (from the demand side). Adam and Amuquandoh (2013, 2014) conducted a survey on hotel owners in Ghana, and the results disclosed several location selection factors, such as laws and regulatory frameworks, neighborhood characteristics, physical site characteristics, and socio-cultural factors. In Newell and Seabrook (2006)'s survey of hotel investors, several location factors were highly ranked when making hotel investment decision. These factors include site attributes, hotel supply, demand volatility, and number of domestic visitors. Juan and Lin (2011) had a panel survey and highlighted six hotel location selection dimensions: factor endowments, demand conditions, firm strategy, related and supporting industries, government and chance.

2.2. GIS applications in hospitality and tourism management

GIS is defined as a computerized system used for the storage, retrieval, mapping, and analysis of geographic data. Considering that tourism is an essentially spatial phenomenon by the movement of tourists, Farsari and Prastacos (2004) reviewed GIS applications in tourism studies and indicated the great potential of GIS usage. Based on their review, the major uses of GIS include the