# Identifying emerging hotel preferences using Emerging Pattern Mining technique

Gang Li [a,1], Rob Law [b,2], Huy Quan Vu [a,3], Jia Rong [a,4], Xinyuan (Roy) Zhao [c,*]

[a] School of Information Technology, Deakin University, Vic 3125, Australia
[b] School of Hotel & Tourism Management, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[c] School of Business, Sun Yat-Sen University, Guangzhou, Guangdong, 510275, China

## HIGHLIGHTS

- A novel means for online review analysis identifies features of interest in hotel selection.
- Emerging Pattern Mining is utilized to identify those features.
- A dataset of 118,000 hotel reviews in Asia Pacific destinations was collected from TripAdvisor.

## ARTICLE INFO

## ABSTRACT

Hotel managers continue to find ways to understand traveler preferences, with the aim of improving their strategic planning, marketing, and product development. Traveler preference is unpredictable; for example, hotel guests used to prefer having a telephone in the room, but now favor fast Internet connection. Changes in preference influence the performance of hotel businesses, thus creating the need to identify and address the demands of their guests. Most existing studies focus on current demand attributes and not on emerging ones. Thus, hotel managers may find it difficult to make appropriate decisions in response to changes in travelers' concerns. To address these challenges, this paper adopts Emerging Pattern Mining technique to identify emergent hotel features of interest to international travelers. Data are derived from 118,000 records of online reviews. The methods and findings can help hotel managers gain insights into travelers' interests, enabling the former to gain a better understanding of the rapid changes in tourist preferences.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Most people prioritize accommodation when planning a trip, spending most of their planning time and effort on selecting the right option. Travelers have different expectations and/or preferences, depending on their destination, purpose and mode of travel, as well as previous accommodation experience (Liu, Law, Rong, Li, & Hall, 2013; Liu, Shi, & Hu, 2013). A comprehensive understanding of customer requirements can help hotel managers gain a lead in the market in terms of strategic planning, marketing, and product development (Wilkins, 2010). However, it is difficult to identify such crucial knowledge due to the complex decision-making process and the wide range of selection criteria (Li et al., 2013).

Of these criteria, the most important has to do with hotel features (i.e., attributes or factors) that most travelers seriously consider. The most valuable hotel features that significantly affect a traveler's selections include location, price, facilities, and cleanliness (Lockyer, 2005). Other features, such as the size and type of building, quality of service and a quiet environment, are important to some people (Albaladejo-Pina & Diaz-Delfa, 2009; Merlo & de Souza Joao, 2011). Merlo and de Souza Joao (2011) examine specific hotel features, such as air conditioning in bedrooms. Sohrabi, Vanami, Tahmasebipur, and Fazil (2012) present another list of important hotel features, including promenade, comfort, security, network, pleasure, news, recreational information, expenditure, room facilities, and car parking.

* Corresponding author. Tel.: +86 20 84112721; fax: +86 20 84036924.
E-mail addresses: gang.li@deakin.edu.au (G. Li), hmroblaw@polyu.edu.hk (R. Law), quan@tulip.org.au (H.Q. Vu), jiarong@tulip.org.au (J. Rong), zhaoxy22@mail.sysu.edu.cn, zhaoxinyuan@hotmail.com (X. (Roy) Zhao).
[1] Tel.: +61 3 9251 7434; fax: +61 3 9251 7604.
[2] Tel.: +852 3400 2181; fax: +852 2362 9362.
[3] Tel.: +61 432 411 359.
[4] Tel.: +61 3 925 17711.

Advances in Internet technology enable travelers to share their travel-related experiences, opinions, and concerns on many online platforms (Mack, Blose, & Pan, 2008). Thus, researchers are now shifting their attention to this data source as a way of mining traveler preference in a cheap, efficient, and nonintrusive manner. For instance, Stringam and Gerdes (2010) use a corpus-based approach to analyze guest comments on online hotel distribution sites as well as to identify frequently used words, patterns of word usage, and their relationship to hotel features rating. Furthermore, descriptive statistical data are used to assess the importance and effect on ratings of several features, including location, size of guest rooms, staff, facilities, and breakfast offerings (Stringam, Gerdes, & Vanleeuwen, 2010). Chaves, Gomes, and Pedron (2012) show that room, staff, location, cleanliness, friendliness, and helpfulness are the most frequently used words in online reviews of small and medium hotels in Portugal. Liu, Law, et al. (2013), Liu, Shi, et al. (2013) analyze comments collated from TripAdvisor.com and changes in hotel customers' expectations according to travel mode, using the association rule mining technique. Li et al. (2013) utilize the Choquet integral, a method of fuzzy decision support, to analyze the selection preferences of different groups of travelers in terms of several hotel features.

Researchers have yet to meet the increasing demand of hotel managers for more accurate knowledge on the hotel preferences of travelers. Several limitations that prevent researchers from identifying such knowledge are listed below.

Identifying Emerging Features

Traveler preference is unpredictable and dynamic. For example, travelers once preferred having a telephone in their room. During that time, charging for telephone usage used to be a significant source of revenue, but usage has declined to a point wherein investing in this facility resulted in losses for many hotels that offer this facility (Huettel, 2010). Today, hotels gain significant customer satisfaction by offering free wireless Internet (Bulchand-Gidumal, Melian-Gonzalez, & Lopez-Valcarcel, 2011). These changes in travelers' concerns can affect the performance of hotel businesses. As such, managers must effectively identify features that are becoming important to travelers. However, efforts to address this issue have been limited.

User Identification for Feature Improvement

Different types of travelers have different expectations of hotel features (Liu, Law, et al., 2013; Liu, Shi, et al., 2013). Some aspects may be important to all travelers, whereas others may be significant only to a specific subgroup. A clear picture of such differences could benefit hotel managers. For instance, if travelers from Western countries prefer clubbing facilities, managers can design appropriate business solutions to improve those features and meet the specific expectations of this group. However, this aspect has received little research attention.

The identification of emerging features is different from traditional approaches to hotel feature analysis, because analysts have no prior knowledge on what features should be included in the study. Large data samples are also required to identify emerging changes in customer preference patterns. Traditional research methods, such as surveys, opinion polls or focus groups, are inadequate. Therefore, resorting to available online data, such as online reviews generally expressed as textual comments, is necessary. These reviews contain abundant information on user opinions, experiences, or concerns, and are considered potential goldmines from which tourism researchers can gain insights into the behavior of travelers (Pan, MacLaurin, & Crotts, 2007).

The analysis of hotel features treats each feature as an item, and a set of hotel features associated with a traveler is an item set. Identifying emerging changes in traveler response to such features is typically formulated as a problem of Emerging Pattern Mining (EPM). Originally proposed by Dong and Li (1999), EPM can capture emerging trends in time-stamped databases or sharp contrasts between data sets or groups. This technique is mainly applied in bioinformatics (Li, Liu, Downing, Yeoh, & Wong, 2003; Li & Wong, 2002) and remains an active topic in computer science (Li & Yang, 2007; Yu, Chen, & Tseng, 2011). By using EPM, researchers can identify emerging hotel features.

The current study aims to fill the current research gap by introducing the EPM technique to establish emerging hotel features. The term "hotel features" includes any entity or concept that concerns travelers when reviewing a hotel. In our case study, we first construct a comprehensive list of candidate hotel features from a large collection of text-based online reviews ($N \approx 118,000$). We use the EPM technique to identify emerging features that currently receive more attention from international travelers. We also construct a set of user profiles to assist hotel managers in improving the features available in their properties. The method and the findings of this study are potentially valuable to hotel managers who want to gain insights into travelers' concerns and find ways to adapt to rapid changes in the tourism market.

The rest of the paper is organized into sections. Section 2 summarizes the methods used and attempts made to analyze hotel features. Section 3 presents the review framework used for creating a hotel features list from textual comments, and a detailed description of EPM concepts used to identify emerging features. Section 4 demonstrates the effectiveness of the proposed method in a case study. Finally, Section 5 concludes our study and offers suggestions for future research directions.

## 2. Related work

This section reviews existing studies that utilize hotel features to explore traveler preferences. We also present a critical analysis of the limitations of these studies and our research objectives.

### 2.1. Hotel features analysis

Several studies have analyzed hotel features to acquire knowledge on traveler preference. A popular method for data collection is using survey questionnaires, wherein hotel features are represented by short-answer questions or a set of keywords. Table 1 summarizes the hotel features included in traditional studies.

Due to the increasing interaction among travelers, studies are increasingly utilizing observation data collected from online resources (e.g., blogs, travel websites, and social media) through online reviews. Traditional, statistics-based data analysis models are ineffective in extracting information from text-based reviews and comments. Thus, new techniques and approaches have been proposed. For instance, manual content analysis is used to study traveler characteristics and communications about Australia as a tourism destination (Carson, 2008; Wenger, 2008). Another study employs the narrative structure analysis to identify key marketing elements, including characterization, space categorization, and evaluation of the product experience (Tussyadiah & Fesenmaier, 2008). Manual methods are time consuming and incapable of obtaining the overall differences in traveler preferences. Although automated approaches, such as corpus-based semantic analysis (Rayson & Garside, 2000) and stance-shift analysis (Davidson & Skinner, 2010), are also employed, such methods require users to have a background in linguistics and access to expensive software (Capriello, Mason, Davis, & Crotts, 2013). Thus, tourism researchers