Assessing language discrepancies between travelers and online travel recommendation systems: Application of the Jaccard distance score to web data mining

Sangwon Park, Dae-Young Kim *

University of Missouri, United States

ARTICLE INFO

ABSTRACT

By using a human-centric approach to online recommender systems, this research aims to estimate the language discrepancies of which travelers and destination marketers describe the travel experiences across 11 tourism destinations in USA. In order to address the research purpose, data has been collected from two different sources that reflect the views of travelers and service providers. Then, a set of text data mining methods (i.e., clustering analysis and Jaccard distance score) was applied to identify the language differences between travelers and CVB websites, according to the following categories: shopping, dining, nightlife/activities, and attractions. Some possible methodological extensions that can improve recommendation capabilities, and managerial implications of these findings are provided.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

The notion of smart tourism has recently gained attention from academics and practitioners. The concept aims to accelerate service innovation and improve tourism experience as well as enhance destination competitiveness by developing IT infrastructure and capabilities (Gretzel et al., 2015a). Particularly, the intelligent smart tourism system uses information aggregation and ubiquitous connectedness to facilitate travelers to obtain personalized information (Gretzel et al., 2015b). In this sense, the fundamental role of destination marketing organizations (DMOs) – that is, to understand travelers' expectations on visiting a destination and to offer tailored information and services – has become more crucial than ever before (Werthner and Klein, 1999).

Convention and visitors bureaus (CVBs) are important information brokers and disseminators in the local tourism industry and act as a layer of destination management in the U.S. With financial support from the local community, one of the critical goals of CVBs is to promote their destinations to both leisure and business travelers. As a result, providing useful/helpful information to travelers is an essential part in the CVBs' marketing activities and tasks (Kim et al., 2011; Stepchenkova et al., 2010). With the emergence of the Internet, many CVBs have adopted online applications that facilitate providing a substantial amount of information to travelers and, as a result, help plan their trips. Nonetheless, increased accessibility of destination-related information via the CVBs' websites may bring about "information overload",

which creates challenges for online travelers to find appropriate information and make choices (Choi et al., 2007; Kim, 2009). Besides, this information is often presented in a way that does not match how travelers search for information (Pan and Fesenmaier, 2006).

At present, the continuous evolution of information technology allows CVB websites to adopt recommender systems that can simplify the decision-making process for travelers (Fesenmaier et al., 2006). This system enables travelers to lessen search costs and cognitive efforts by identifying alternatives that meet the specific needs of online users and by offering information in a personalized way (Gretzel et al., 2012; Kabassi, 2010; Wind and Rangaswamy, 2001). Accordingly, the recommender systems should be human-centric in their design and functionality. This requires system-user interactions by understanding cognitive styles of online information seekers and adjusting the recommender system to address the needs/desires (Bauernfeind, 2003; Zins et al., 2003). Particularly, this research focuses on linguistic interactions between users and the system in the context of tourism (see Dann, 1996; Gretzel et al., 2012). Based upon the definition of linguistic interactions referring to the way in which language appears interactions in everyday to represent cognitive process (Couper-Kuhlen and Selting, 2001), providing destination information that online travelers actually required with proper language is a cornerstone of recommendation, which facilitates smart tourism.

Matching the language in tourism is important to fulfill the effective communication between travelers (or visitors) and destination marketers (hosts) (Xiang et al., 2008). Previous studies identified a number of cases of incongruent destination images projected by marketers and perceived by travelers (MacKay and Fesenmaier, 2000) as well as across different online travel resources (e.g., blogs, magazines, guides, and travel trade)

(Choi et al., 2007). The previous studies indicate the perceptions and/or images that destination travelers bring to mind do not seem to coincide with those highlighted by suppliers. Thus, an important research questions can be induced to test whether the experiences that travelers are looking for at the specific destinations ("perceived" or "expected") correspond to those promoted by marketers ("projected") as a foundation for implementing a human-centric approach to the recommender system.

Therefore, the main goal of this paper is to estimate the effectiveness of CVB websites by comparing the nature of the language travelers used to describe their expected trip experiences and the contents provided by destination marketers on CVB websites. This research used two types of text mining techniques to examine the language discrepancies: text clustering (Stepchenkova et al., 2009) and Jaccard distance score (Maedche et al., 2003). The findings of this study provide an extensive understanding of travel experiences expected by potential travelers and promoted by destination marketers across 11 tourism destinations, and assessment of language discrepancies focusing on multi-facets of travel products including "shopping", "dining", "night life and activity", and "attraction". Furthermore, this paper suggests applying the advanced data mining method to capture traveler preferences through analyzing textual data. As a result, the findings of this research suggests implications to develop more effective online recommender systems within the context of the tourism-related industry.

2. Literature review

2.1. Online recommendation system

Ricci (2002) defined recommendation systems as applications associated with online platforms to suggest products/services and offer travelers personalized information to help with their decision-making process. Considering the complex nature of travel planning that involves numerous decision tasks – not only a destination but also, e.g., accommodations, activities, restaurants – travelers are subject to experience an excess of information over their capability in the process to make diverse decisions. In this context, online recommendation systems have great potential for the usability in not only reducing the search costs but also improving decision qualities (e.g., Häubl and Dellaert, 2004; Häubl and Trifts, 2000). The online system can assist this task by matching the consumers' needs and preferences through providing tailored services and available options.

Studies on the development of online recommender systems can be categorized into two classes of research focuses: (1) process of the systems by which the recommendation systems operate with certain algorithms according to different contexts and (2) outcome of the systems based upon different individual and situational features (Fesenmaier et al., 2006). The recommendation systems vary in sophistication, ranging from simple retrieval or filtering approaches to comprehensive computing systems (Spiekermann and Paraschiv, 2002). There are basically two classifications: content-based and collaborative filtering systems (Yeh and Cheng, 2015). The assumption of content-based filtering is that characteristics of an item determine the user's preferences of the item (Ricci, 2002). Specifically, the content-based filtering approach provides a user with suggested products/services that are similar to those s/he has purchased or searched in the past. The systems attempt to match the attributes of the products/services with the characteristics of the users stored in the data base. To the contrary, collaborative filtering (or social filtering) systems infer the behaviour of users toward products/services from other users who show similar interests or preferences and mimic social processes (Breese et al., 1998). This application assumes that the evaluation or opinions of others are an important information source that travelers use in their decision-making process (Gavalas et al., 2014).

The later aspect of the recommendation research investigating outcome of the systems is directly related to understandings of information processing and evaluations as well as decision making behaviors (Kabassi, 2010). For example, the consumer styles inventory has been

applied to comprehend travel decision making styles in a way to envisage different information sources and contents travelers searched as well as attributes of the destinations they preferred (Zins et al., 2003). Gretzel et al. (2012) proposed a theoretical framework of destination recommender systems, suggesting the design components should be responsive to travelers' needs in terms of personal characteristics of the travelers (e.g., demographics and personality), situational needs and constraints (e.g., travel party and lengths of stay) and aspects of the decision-making process (e.g., the specificity of the choice task and decision frames). The focus on the traveler as the user of the system is highlighted by anticipating user needs and offering recommended alternatives according to specific consumption contexts (Buhalis and Amaranggana, 2015).

In brief, the constant findings of those previous studies argue that a human-centric approach is extremely important to make the recommender systems helpful and successful as decision-making support tools (Chung et al., 2015; Gretzel et al., 2006). Design and functionality of human-centric computing require an intensive understanding of the individual behaviors so that the system enhances the ability to fulfill the interactions between the recommender systems and users. Zins et al. (2003) stated that the adjustment of the recommender system interface to fit a user's cognitive style is vital for enhancing the quality of the interaction. With the development of information technology, travelers are able to assert their needs for information, which are formed within their individual contexts. DMOs that manage contents/design of the destination websites become a primary agent that establishes a basic lens to represent a destination and experiential aspects as well as a process by which travelers gain information (Pan and Fesenmaier, 2006). Thus, accomplishing the axiom "speaking the right language" is an important aspect in the online recommender system, which addresses the slogan of user-centered design: "Recommender systems are about people, not machines" (Ricci, 2002; pp. 57). The statement emphasizes issue of the product description language. That is, even if the recommendation system is well-developed in the engineering aspect, users will have challenges when information presented (or destination descriptions) is too terse or does not fit their needs. In this vein, to address the research question of this study, this research estimates language discrepancies between expressed by travelers and marketers. Due to the textual format of data analyzed in this research, a set of text mining approaches has been used to examine the differences. The following sections discuss the methods of text mining in general and Jaccard distance score approach in particular.

2.2. Text mining techniques

Text mining, also known as text data mining, is the process of deriving useful information from a text dataset (Feldman and Sanger, 2007). Machine learning, data mining, and information retrieval techniques have enabled the text mining field to advance dramatically during the past decade. The development of the data mining field brought about a diverse set of text mining techniques, such as text categorization, text clustering, concept extraction, sentiment analysis, and entity relation modeling (Feldman and Sanger, 2007; Ikonomakis et al., 2005). These text mining techniques have been largely applied to various fields. For example, in marketing, text mining was used in the context of customer relationship management in order to develop prediction models for customer attrition (Coussement and Van Poel, 2008). The fuzzy cluster technique was used to classify customers based on their historical loyalty analysis (Simha and Iyengar, 2006). In customer preference prediction analysis, Bayesian-based cluster models were adopted to predict the active user's preference (Adomavicius and Tuzhilin, 2005; Breese et al., 1998). It was also used in gaining tourism knowledge as to how people search information online by clustering similar phases according to their meanings (Xiang et al., 2007). Xiang et al. (2015) tried to comprehend the associations of hotel guest experience with satisfaction by analyzing traveler reviews with text mining analytics.