



Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip



Sérgio Moro^{a,b,*}, Paulo Rita^{c,d}, Joana Coelho^e

^a Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisboa, Portugal

^b ALGORITMI Research Centre, University of Minho, Portugal

^c Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Lisboa, Portugal

^d NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

^e Instituto Universitário de Lisboa (ISCTE-IUL), ISCTE Business School, Lisboa, Portugal

ARTICLE INFO

Article history:

Received 9 November 2016

Received in revised form 16 April 2017

Accepted 22 April 2017

Available online xxxx

Keywords:

Customer feedback

Customer reviews

Online reviews

Knowledge extraction

Data mining

Modeling

Sensitivity analysis

Las Vegas

ABSTRACT

This study presents a data mining approach for modeling TripAdvisor score using 504 reviews published in 2015 for the 21 hotels located in the Strip, Las Vegas. Nineteen quantitative features characterizing the reviews, hotels and the users were prepared and used for feeding a support vector machine for modeling the score. The results achieved reveal the model demonstrated adequate predictive performance. Therefore, a sensitivity analysis was applied over the model for extracting useful knowledge translated into features' relevance for the score. The findings unveiled user features related to TripAdvisor membership experience play a key role in influencing the scores granted, clearly surpassing hotel features. Also, both seasonality and the day of the week were found to influence scores. Such knowledge may be helpful in directing efforts to answer online reviews in alignment with hotel strategies, by profiling the reviews according to the member and review date.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The Online Travel Agencies (OTA) are now the most used tool of travel booking, both for the means of transport and accommodation (Mauri & Minazzi, 2013) and, consequently, online reviews have been exponentially increasing its use and impact in the hospitality industry over the last years, due to the social media and technological evolution. In fact, nowadays potential hotel customers search for online feedback before travelling and base their purchase decisions on online reviews (Mauri & Minazzi, 2013). Therefore, electronic word-of-mouth (eWOM), which according to Henning-Thurau et al. (2004, pp. 39) is defined as “any positive or negative statement made by potential, actual or former customers about a product or company, which is made available to a multitude of people and institutions via the internet”, has become a huge aspect when travelling, since currently every consumer has access to the internet and can easily express either positive or negative feedback. Most importantly, it is an online tool to be used when others seek for advice as part of the decision-making process, such as where to stay, especially in hospitality industry, as consumers are purchasing

an experience and cannot predict its evaluation (Sparks & Browning, 2011). Moreover, holidays can be considered as a high risk and involvement purchase, due to its usual personal importance and also high value of money (Papathanassis & Knolle, 2011). Service quality is a determinant of the customer's perceptions and their feedback. The ideal would be that the target's expectations meet the perceptions, which will directly influence a positive word-of-mouth, contributing for a development of reputation and trust (Corbitt, Thanasankit, & Yi, 2003). Hence, research contributions that unveil and provide in-depth understanding on the features that have the most impact on customer feedback are valuable for sustainable decision making.

Previous studies have been conducted by various researchers in order to understand and explain the influence and impact of online reviews in the hospitality industry. One of the most common methods used include the analysis of variance (ANOVA) technique, which is offered in many data analysis' solutions such as the IBM SPSS software. For example, Vermeulen and Seegers (2009) adopted the ANOVA for testing whether or not the user-generated online reviews influence the consumer choice. In a parallel line of research, Jeong and Jeon (2008) also used the ANOVA for analyzing the impact of five relevant features (hotel ownership, stars, number of rooms, room rates, and popularity index) in scoring New York hotels on TripAdvisor's nine rating items (e.g., location; cleanliness). Their results show that both the

* Corresponding author at: ISCTE-IUL, Av. das Forças Armadas, 1649-026 Lisboa, Portugal.

E-mail address: sergio.moro@iscte.pt (S. Moro).

number of stars and room rates influence the rating items from TripAdvisor. A similar study focused on analyzing the relationship between the hotel specific rating items used by Expedia (service, condition, cleanliness, and comfort) in the hundred largest US cities. Again, statistical tools and methods were adopted, including the ANOVA (Stringam, Gerdes, & Vanleeuwen, 2010). Additionally, Sparks and Browning (2011) went further on their research and studied the fact that a consumer generated quantitative rating could be associated together with the actual written review. In a more recent data-driven study, it has been shown through regression models that the financial benefits of an online review from TripAdvisor conceal intrinsic value to the hospitality industry (Neirotti, Raguseo, & Paolucci, 2016). Nevertheless, the majority of previous recent studies are focused on the impact of the text review itself, applying text mining techniques, which aim to extract meaningful knowledge from a variety of textual data and find relationships and patterns within such unstructured information (Calheiros, Moro, & Rita, 2017).

Different studies are aligned through similar conclusions regarding the fact that text mining applications to social media data (i.e. any online platform where customers can exchange information) can provide significant insights on the human behavior and interaction (e.g., He, Zha, & Li, 2013). However, while several studies are known using data mining for sentiment classification and opinion mining (e.g., Schuckert, Liu, & Law, 2015), none was found up to the present adopting a quantitative approach on modeling tourists' reviews through advanced data mining techniques for extracting the influence of hotels' and users' features on the score provided by users. Nevertheless, the quantitative score is the first relevant information users see when they search for feedback information on their next stay (O'Connor, 2010). Understanding which profiles of users are most likely to result in poorer scores may help to shape strategies for choosing the users to whom to answer in TripAdvisor, as answering all users is time-consuming and requires significant human effort (Nguyen & Coudounaris, 2015). Thus, such directed effort can lead to an improvement in positive eWOM, as the responses may be framed for specific users. Additionally, identifying the features influencing scores granted may help to profile users, helping to identify outlier behaviors and possible reputation attacks (Buccafurri, Lax, Nicolazzo, & Nocera, 2014). Since users are influenced by hotels (Casalo, Flavian, Guinaliu, & Ekinci, 2015), including hotel features in a unique model allows to obtain explanatory knowledge intersecting both dimensions. Hence, the present study aims at filling such research gap by focusing on online reviews' quantitative features such as number of stars of the hotel and number of helpful votes the user has received in order to build a predictive model of the tourists' score on the hotels. The knowledge built upon such model may help to shed some light on what drives the rating of a hotel, potentiating meaningful information to support managerial decisions.

The proposed data mining approach is an attempt to answer the following research questions: Can the score of an online hospitality review be predicted using as input only quantitative data? What are the features that influence most the review scores in hospitality? How does each of those features affect the score and can this knowledge be useful for hotel managers?

Concluding, the main goals and contributions of this study are as follows:

- Creating a model that predicts the review score based on quantitative features of the user/reviewer and the hotel, as well as the period of time of the specific stay;
- Contributing to research on customers' feedback and online reviews by providing a novel approach on the used data, the quantitative features, as opposed to the most common analyses of the reviews' text itself;

- Understanding how users are inherently influenced by hotels' features when submitting numerical scores besides text comments on online platforms, such as TripAdvisor.

The next section describes the background concepts, such as the history and evolution of online reviews, as well as the methods for knowledge extraction from data, its dimensions and its use in the industry. Section 3 discusses the materials (e.g. input dataset) and procedures that were applied in the experiment. Then, the results are shown and a critical discussion takes place on the findings section. Finally, the main conclusions of this research are drawn.

2. Theory

2.1. Online reviews

In 2004, Tim O'Reilly coined the term Web 2.0 as the network connecting all devices to which individual users contribute largely by sharing their experiences in numerous ways, therefore becoming one of the most relevant sources of the internet through the so called user-generated contents (O'Reilly & Battelle, 2009). Such internet evolution effectively became a global revolution, including the tourism and hospitality industry by adding new online sources of information to the existing hotel and tourism companies' websites, implying users are becoming key-players in influencing others through their online reviews (Law, Buhalis, & Cobanoglu, 2014).

Traditional websites have therefore evolved by increasing interactivity level to keep pace with Web 2.0 new demands. However, in this new information-driven era, specialized user-content sites and applications such as wikis, forums, blogs, social networks and especially online reviews' sites for the case of tourism and hospitality have underpinned a new paradigm in which the user is at the center of the network, leading to a mutual exchange and sharing of values (Liburd, 2012). As Zeng and Gerritsen (2014, pp. 27) pointed out, "leveraging off social media to market tourism products has proven to be an excellent strategy".

Several studies are found based on online reviews for tourism and hospitality, especially to analyze how exchanges of information influence directly the consumer choices regarding a certain hotel (e.g., Park & Nicolau, 2015), with most of them concluding that an exposure to an online hotel positive review will increase the average probability of that consumer to book a room in the same hotel. Features such as the number of stars have shown to positively influence the score granted by users on online reviews (Hu & Chen, 2016). In fact, users expect higher rated hotels (i.e., with a higher number of stars) to have more positive reviews, according to Phillips, Zigan, Silva, and Schegg (2015). The latter study goes further on the analysis by revealing that larger hotel units with higher number of rooms do not directly translate into high revenue. By building an artificial neural network model, Phillips et al. (2015), managed to obtain a unique and valuable model explaining the intersection of a few hotel and regional characteristics, with the number of reviews. However, the same study did not include in its model the features of each individual user, as it was aimed for a granularity at the hotel level. Fang, Ye, Kucukusta, and Law (2016) confirmed through an econometric model that user/reviewer characteristics affect the perceived value of the reviews made, proving that user features should also be accountable when modeling online reviews' scores.

The recent study by Kim, Kim, Park, and Park (2017), comparing both TripAdvisor scores and traditional customer satisfaction through travel intermediaries, found out that online reviews play a more significant role in explaining hotel performance metrics than traditional feedback. Such finding can be linked to users' perceptions, as a vast majority of them believe in online reviews published on platforms such as TripAdvisor, being directly influenced by scores granted by other users, even though reputation attacks seem to occur often in the

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات