Tourism Management Perspectives 24 (2017) 151–154

Contents lists available at ScienceDirect

Tourism Management Perspectives

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

Short communication

Are Yelp's tips helpful in building influential consumers?

João Guerreiro[a,b,*], Sérgio Moro[b,c]

[a] Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Lisboa, Portugal
[b] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisboa, Portugal
[c] ALGORITMI Research Centre, University of Minho, Guimarães, Portugal

ARTICLE INFO

Keywords:
eWOM
Online reviews
Fandom
Text mining
Support vector machine

ABSTRACT

In the cluttered environment of online reviews, consumers frequently have to choose the most trustworthy reviewers to help them in their purchasing decision. Such reviewers are influential in their community and co-create value among their peers.

The current research note studies the antecedents of fandom, particularly if contents of the message written by the reviewers predict the number of fans they might have in the future. 27,097 tips written by 16,334 users of Yelp are structured using text mining and a support vector machine algorithm is used to study the accuracy of such relation.

Results show that tips which may help consumers to avoid the service and tips that highlight the positive elements of the service are the most relevant in predicting the number of fans.

Findings may help managers to understand which type of messages may increase the reviewer's number of fans, thus increasing their influence in the network.

1. Introduction

One of the most challenging tasks of choosing a hotel or restaurant through online recommendations is whom and what to trust when it comes to decide. Today there is a plethora of online reviews about products and services that can be accessed through Booking, TripAdvisor and Yelp. However, in the cluttered environment of online reviews, consumers tend to heuristically choose some of them to reduce the consideration set of the possible alternatives (Chaiken & Ledgerwood, 2012). A recent study shows that the three most important factors that consumers evaluate when reading online reviews are the overall rating (66%), the ratio of positive and negative reviews (63%) and the amount of detail in the review (62%). Reviewer status (40%) is also one of the top factors considered when reading reviews (Statista, 2017).

Recommendation sites, as any other online social network, are built on value co-creation. Only some of the actors in the network create worthy value to others in the community, thus becoming more influential (Verhoef, Beckers, & van Doorn, 2013). Consumers read available information such as online reviews to form an overall opinion and reduce the risks of choosing one option over another (Bronner & de Hoog, 2011). However, in order to filter the most important reviews, credibility and trust in the reviewer are vital cues to access such probability of risk more accurately (Liu & Park, 2015). Trust is a two sided construct with both affective and cognitive mechanisms. Affective trust is generally established on the basis of how warm, open and friendly the reviewer is when it evaluates products or services (Johnson & Grayson, 2005; Johnson-George & Swap, 1982; Xu, 2014) while cognitive trust emanates from the expertise degree of the user (Cook & Wall, 1980; Moro, Rita, & Coelho, 2017; Xu, 2014). Perceived credibility is often influenced by the reviewers' reputation, which is often measured by the number of helpful votes, while trustworthiness may be measured in terms of the number of fans or followers a user has in the network (Xu, 2014). Trustworthy consumers with a huge number of fans are usually targeted by companies using seeded marketing campaigns (SMCs) to generate positive eWOM (Chae, Stephen, Bart, & Yao, 2016). Such consumers are generally influential in their community and, therefore, they are able to co-create value and boost loyalty and satisfaction among their peers (Sashi, 2012).

Although studies show that the number of fans is a proxy for trust in the information provider (Xu, 2014), there is still a need to understand the antecedents of fandom, particularly if the message written by the reviewers somehow predicts the number of fans they might have in the future. This research note has its main theoretical contributions focused on bridging such gap using a text-mining approach and a support vector machine (SVM) data mining algorithm to study the accuracy of such relation. Text mining has been successfully applied as a semi-automatic process to obtain a structured comprehension of the underlying terms

* Corresponding author at: Av. das Forças Armadas, 1649-026 Lisbon, Portugal.
E-mail address: joao.guerreiro@iscte-iul.pt (J. Guerreiro).

http://dx.doi.org/10.1016/j.tmp.2017.08.006
Received 29 July 2017; Accepted 9 August 2017

and topics in text (Miller, 2004). Its use extends from literature review analysis (Guerreiro, Rita, & Trigueiros, 2016) to online review analysis (Calheiros, Moro, & Rita, 2017). Likewise, SVMs have been used in previous studies to successfully predict positive and negative online reviews based on individual words (Dickinger, Lalicic, & Mazanec, 2017).

A specific kind of message was used to study how its content may predict the number of fans. Yelp currently has a kind of review posted by its users called *tips*, which are small reviews with a very specific kind of opinions. Unlike regular reviews, *tips* are often written using the smartphone application of Yelp (Yelp, 2017). Therefore, they include information about how the consumer felt immediately after using a service or buying a product. Such reviews may not be as rationalized as those written in the site, because consumers don't need to proactively login and write a full review. *Tips* are much more short and real-time driven than full reviews and they contain valuable information for consumers when they are about to choose a service provider. Therefore, the *tip* content should be more fit than full reviews to predict fans attraction.

Managerial contributions of the current research may help managers to carefully plan how to answer *tips* that may increase the number of fans. *Tips* that include negative opinions and that are expressed by a user with a huge influence in the network, may negatively influence the company reputation if not properly addressed (Lee & Cranage, 2014). Those *tips* will eventually become more influential and reach a wider target than *tips* that don't have such word markers.

2. Materials and methods

The current study used a dataset available from Yelp (2017) from which a sample was extracted that contained only *tips* from bars and restaurants. The sample extracted 27,097 *tips* from 1,536 bars and restaurants rated by 16,334 users.

The procedure for preparing and analyzing the data gathered is detailed in Fig. 1. The ellipses show the steps toward building a data mining model that could be useful in explaining the number of fans of each reviewer through the most meaningful words from Yelp's *tips*. The white squares display outcomes obtained during the procedure, while the grayed rounded squares show the dataset through the data preparation stage, a stepping stone toward modeling (Moro et al., 2017). After *tips* have been extracted to a *json* data file, unstructured text was transformed into a document-term-matrix (DTM) that was later reduced in its sparsity to decrease the influence of outlier frequencies (Blei & Lafferty, 2009). After such transformation, the final dataset kept 13,297 *tips* with frequent occurring words.

The support vector machine (SVM) was chosen for modeling, as it offers a non-linear machine learning algorithm to distinguish data by defining separating hyperplanes (Vapnik, Guyon, & Hastie, 1995). SVM was fed with the words gathered by text mining techniques, which enable to extract patterns of knowledge from unstructured data, such as the textual contents of *tips* (Calheiros et al., 2017). Finally, the data-based sensitivity analysis (DSA) was adopted for extracting the relative relevance of each feature in terms of its contribution to the model (Cortez & Embrechts, 2013). All experiments were conducted using the R statistical tool (https://cran.r-project.org/), as it offers packages specifically implemented for the tasks undertaken, including text mining ("*tm*") and data mining ("*rminer*") (Cortez, 2014).

3. Results and discussion

The accuracy of the final SVM model obtained with the 105 most frequent words (i.e., the ones that were not considered sparse, with a sparsity level below 99%) was measured through the mean absolute error (MAE) and the normalized MAE (NMAE) (Hyndman & Koehler, 2006). The former represents the average deviation of the predicted number of fans from the real value, while the latter is the normalized deviation considering the amplitude of the outcome variable, the number of fans. The results presented a MAE of 7.67 and a NMAE of 12.79%, with the latter showing a relatively low error, thus validating the model for subsequent knowledge extraction.

The DSA enabled to offer an overview on the level of contribution of each word contained in *tips* to the number of fans. Table 1 shows only the ten most relevant words. Two direct observations can be made: first, all words are concise, which highlights the synthesis factor associated with the *tips*, when compared to reviews; second, there is a high degree of dispersion in terms of the relevance of each individual word, as even the most relevant word only contributes to less than 2% of the model. Nevertheless, the ten words represented in Table 1 encompass almost

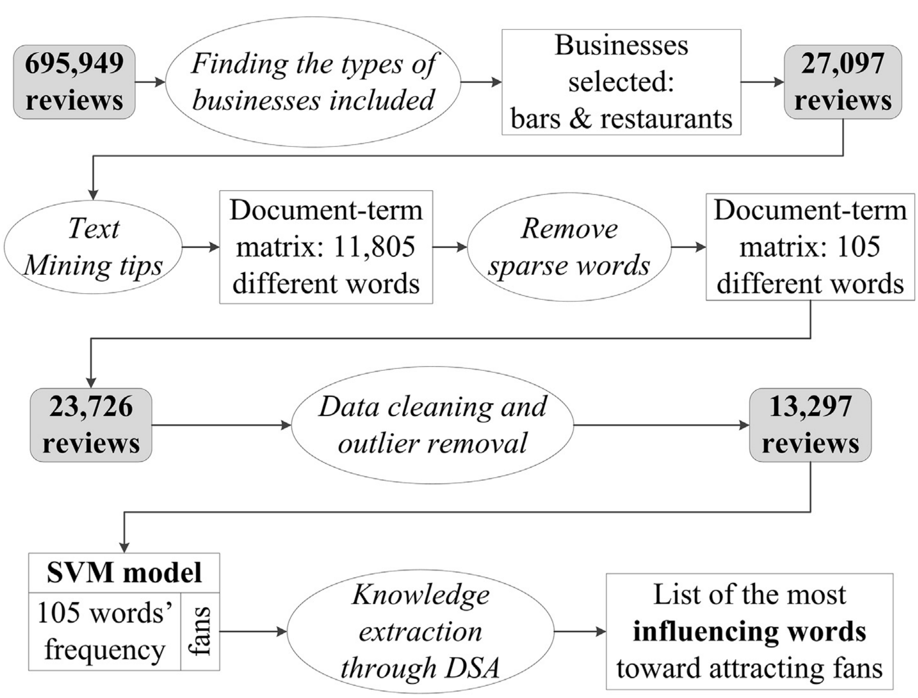

Fig. 1. Procedure undertaken.