

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2019-03-29

Deposited version:

Post-print

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Guerreiro, J. & Moro, S. (2017). Are Yelp's tips helpful in building influential consumers?. *Tourism Management Perspectives*. 24, 151-154

Further information on publisher's website:

10.1016/j.tmp.2017.08.006

Publisher's copyright statement:

This is the peer reviewed version of the following article: Guerreiro, J. & Moro, S. (2017). Are Yelp's tips helpful in building influential consumers?. *Tourism Management Perspectives*. 24, 151-154, which has been published in final form at <https://dx.doi.org/10.1016/j.tmp.2017.08.006>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Are Yelp's tips helpful in building influential consumers?

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.

Keywords

eWOM; Online reviews; Fandom; Text mining; Support vector machine

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 assess 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 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.

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.

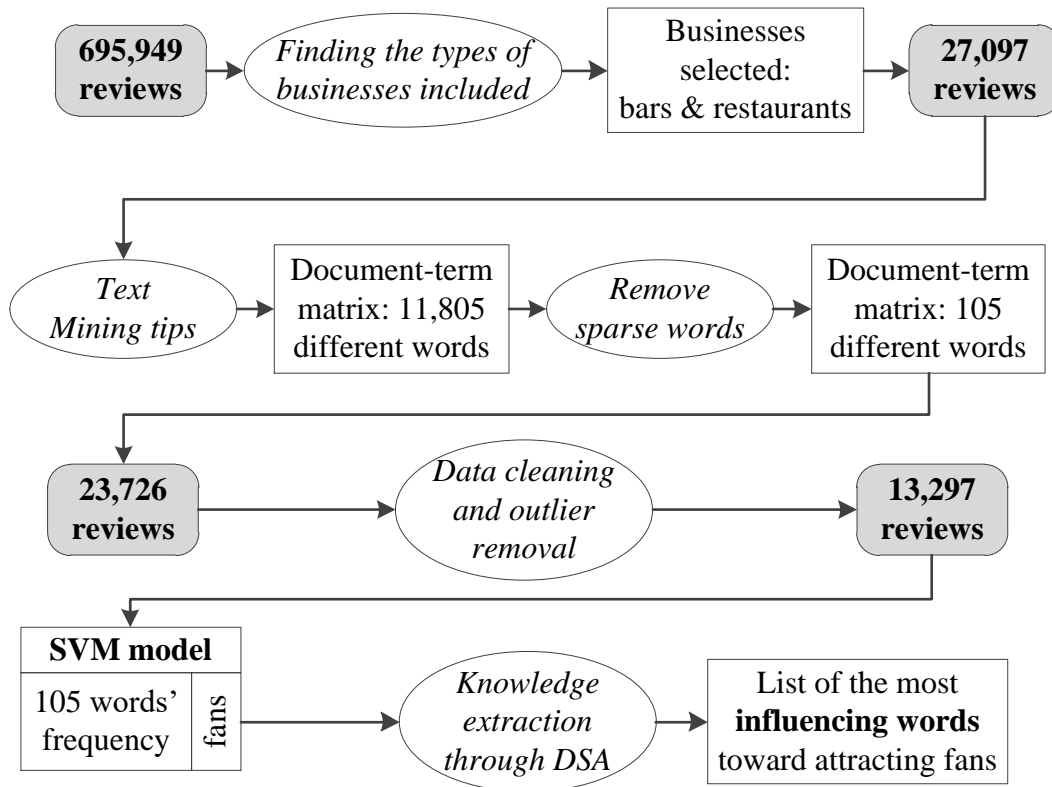


Figure 1 - Procedure undertaken.

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).

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 15% of relevance when attempting to explain the model of the number of fans.

Table 1 - Most relevant words for modeling number of fans.

Word	Percentage relevance
<i>bad</i>	1.83%
<i>staff</i>	1.72%
beer	1.48%
<i>closed</i>	1.45%
<i>buffet</i>	1.37%
sushi	1.37%
<i>don't</i>	1.36%
<i>perfect</i>	1.34%
<i>yum</i>	1.33%
<i>wait</i>	1.31%

Next, the eight words marked in italic from Table 1 are analyzed (“beer” and “sushi” were excluded as those are too narrow, and specifically related with types of food/drink). Fig. 2 shows the influence that each word has on the number of fans,

according to the model. The most relevant word for explaining tips is the word “bad”, which suggests that consumers trust those users who advise them on places to avoid. Generally, results show that a higher number of words imply a higher number of fans, until the number of fans reaches a certain stage, after which fans tend to decrease. Such behavior occurs for distinct words such as “buffet”, “wait”, “staff”, “don't”, and “bad”, although at different levels. Employing the word “bad” in the tip creates a small but steady effect on user's number of fans. Also the use of the “don't” word emphasizes such tendency for negative words, showing a more pronounced effect. However, that is not the case for “perfect”. In fact, it seems that using such word a few times tends to make users less popular, with the tendency reverting after three occurrences.

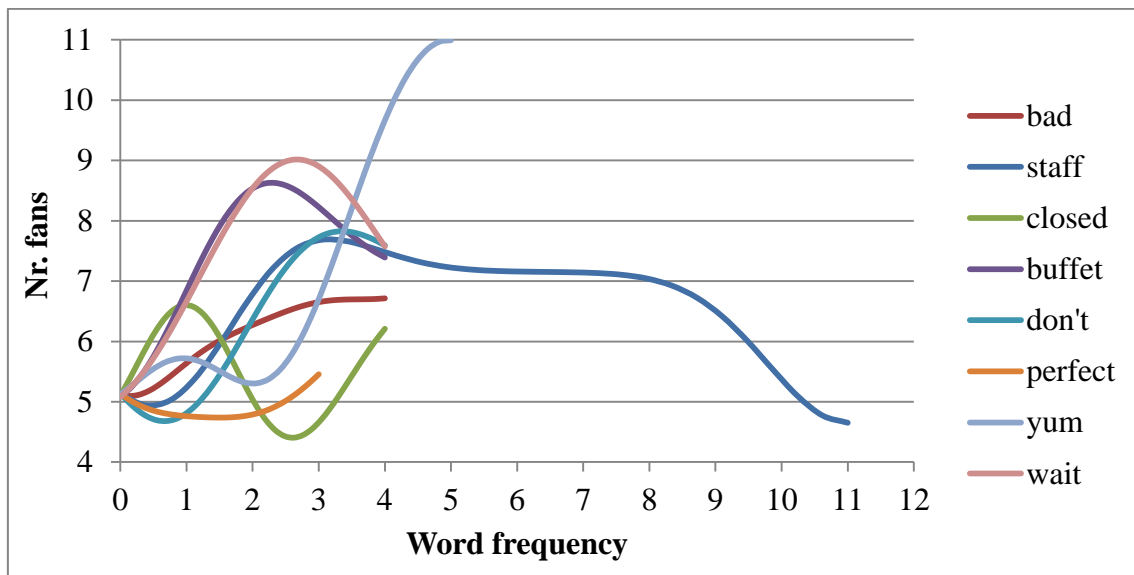


Figure 2 - Impact of the most relevant words.

The second most relevant word according to the model holds paramount implications to hospitality and tourism industry. “Staff” is the only word occurring more than five times which suggests that consumers trust users who talk about how they were treated and attended during the experience. Although the trend gradually fades, “staff” is still the word that maintains fans more engaged. Only the users that are continuously discussing “staff” in their tips (more than 9 occurrences) have a lower fan base. Also associated with customer service is the word “wait” which also reflects how consumers value tips that show them which restaurants or bars to avoid. Finally, the word “yum” has an interesting effect on fandom. “Yum” is a word that highlights how tasty the experience

has been for the user and holds the most pronounced amplification impact on the number of fans when compared to the remaining words.

Conclusions

The current research note used text mining to build a predictive model that is useful in explaining the number of fans of each reviewer through the most meaningful words from Yelp's tips. Text was structured into a document-term-matrix and frequent words were later used as independent variables. A support vector machine algorithm was used to predict the number of fans.

Results show that tips with negative words which may help consumers to avoid the service (“bad”, “don't” and “closed”), and tips which include words alerting consumers for customer service quality (“staff” and “wait”) are among the most relevant to explain the number of fans. Positive words are also present as predictors of fandom. If users point to a place with a “yummy” food, they have generally more fans, probably waiting for good advices on where to go on their next visit.

According to our findings, the most trusted consumers in recommendation platforms use mainly avoidance and positive advice words in their messages based on their past experiences. Previous research shows that satisfaction with previous experiences and service provider performance are antecedents of cognitive trust (Johnson & Grayson, 2005). Our findings suggest that consumers may trust users who write tips that work as a proxy to evaluate such satisfaction and performance indexes.

The study may help managers to monitor more closely the users who frequently report avoidance and positive advices in their recommendation messages as these may predict the users' number of fans grabbed through eWOM, and thus their influence in the network.

References

- Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. N. Srivastava, & M. Sahami (Eds.), *Text mining: Classification, clustering, and applications* (pp. 71–94). Boca Raton: CRC Press.
- Broner, F., & de Hoog, R. (2011). Vacationers and eWOM: Who posts, and why, where, and what? *Journal of Travel Research*, 50(1), 15–26.
- Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*. <http://dx.doi.org/10.1080/19368623.2017.1310075>.
- Chae, I., Stephen, A. T., Bart, Y., & Yao, D. (2016). Spillover effects in seeded word-of-mouth marketing campaigns. *Marketing Science*. <http://dx.doi.org/10.1287/mksc.2016.1001> (October, mksc.2016.1001).
- Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 246e266–). London, UK: Sage Publications Ltd.
- Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfilment. *Journal of Occupational Psychology*, 53(1), 39–52. <http://dx.doi.org/10.1111/j.2044-8325.1980.tb00005.x>.
- Cortez, P. (2014). *Modern optimization with R*. Springer.
- Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225, 1–17.
- Dickinger, A., Lalicic, L., & Mazanec, J. (2017). Exploring the generalizability of discriminant word items and latent topics in online tourist reviews. *International Journal of Contemporary Hospitality Management*, 29(2), 803–816. <http://dx.doi.org/10.1108/IJCHM-10-2015-0597>.
- Guerreiro, J., Rita, P., & Trigueiros, D. (2016). A text mining-based review of cause-related marketing literature. *Journal of Business Ethics*, 139(1), 111–128.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, 58(4), 500–507. [http://dx.doi.org/10.1016/S0148-2963\(03\)00140-1](http://dx.doi.org/10.1016/S0148-2963(03)00140-1).

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, 43(6), 1306–1317. <http://dx.doi.org/10.1037/0022-3514.43.6.1306>.

Lee, C. H., & Cranage, D.a. (2014). Toward understanding consumer processing of negative online word-of-mouth communication: The roles of opinion consensus and organizational response strategies. *Journal of Hospitality and Tourism Research*, 38. <http://dx.doi.org/10.1177/1096348012451455>.

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151. <http://dx.doi.org/10.1016/j.tourman.2014.09.020>.

Miller, T. W. (2004). *Data and text mining: A business application approach*. New Jersey, NJ: Prentice-Hall, Inc.

Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip. *Tourism Management Perspectives*, 23, 41-52. <http://dx.doi.org/10.1016/j.tmp.2017.04.003>.

Sashi, C. M. (2012). Customer engagement, buyer-seller relationships, and social media. *Management Decision*, 50(2), 253–272. <http://dx.doi.org/10.1108/00251741211203551>.

Statista (2017). Most important aspects and features of online reviews in the U.S. 2017.

Vapnik, V., Guyon, I., & Hastie, T. (1995). Support vector machines. *Machine Learning*, 20, 273–297.

Verhoef, P. C., Beckers, S. F. M., & van Doorn, J. (2013). Understand the perils of cocreation. *Harvard Business Review*, 91(9).

Xu, Q. (2014). Should i trust him? The effects of reviewer profile characteristics on eWOM credibility. *Computers in Human Behavior*, 33, 136–144. <http://dx.doi.org/10.1016/j.chb.2014.01.027>.

Yelp (2017). Yelp Challenge Dataset 2017. Retrieved from https://www.yelp.com/dataset_challenge.