



HAL
open science

Comparing Marketing and Computer-Based Methods for Evaluating Online Reviews

Noudéhouéno Lionel Jaderne Houssou, Jeanne Lallement, Mickael Coustaty,
Luc Béal

► **To cite this version:**

Noudéhouéno Lionel Jaderne Houssou, Jeanne Lallement, Mickael Coustaty, Luc Béal. Comparing Marketing and Computer-Based Methods for Evaluating Online Reviews. Information and Communication Technologies in Tourism 2023, International Federation for IT and Travel & Tourism (IFITT), Jan 2023, Johannesburg, South Africa. pp.261-266, 10.1007/978-3-031-25752-0_28 . hal-03962270

HAL Id: hal-03962270

<https://hal-univ-rochelle.archives-ouvertes.fr/hal-03962270>





Submitted on 30 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Comparing Marketing and Computer-Based Methods for Evaluating Online Reviews

Noudéhouéno Lionel Jaderne Houssou^{1,4} , Jeanne Lallement² ,
Mickael Coustaty¹ , and Luc Béal³ 

¹ La Rochelle University, L3I Laboratory, 17042 La Rochelle, France
{noudéhouenou.houssou1,mickael.coustaty}@univ-lr.fr

² La Rochelle University, NUDD Laboratory, 17042 La Rochelle, France
jeanne.lallement@univ-lr.fr

³ Excelia Business School, CRIIM Laboratory, 17000 La Rochelle, France
beall@excelia-group.com

⁴ RMD Technologies, 17000 La Rochelle, France

Abstract. This short paper aims to compare humanities and computer-based online review analysis methods. In particular, we evaluate two classical methodologies coming from marketing and natural language processing fields. We assessed them through their ability to translate online reviews into synthetic evaluations reflecting consumers' overall feelings. Both methods were run in separate ways, then we confronted the results.

Keywords: Online reviews · Text mining · Content analysis

1 Introduction

The predominance of digital booking platforms in the tourism industry has made online reviews essential to get consumer insight and build e-reputation. Indeed, online reviews influence consumers who often give them greater credibility than expert reviews which they find more commercial and less persuasive. To extract value from them, online reviews are usually analyzed by domain experts and researchers from marketing or tourism management and computer science who process them using their specific methods. While marketing or tourism management researchers focus on online reviews' effects on market effectiveness and consumers' persuasion process, in computer science, attention is mainly put on text analysis. However, researchers from these different fields have to deal with the fact that online reviews are massive, not always consensual, and can be complex or ambiguous. The originality of this paper is to conduct an interdisciplinary work aiming at analyzing online reviews from humanities especially marketing and computer science perspectives, to understand how they can complement each other to manage the aforementioned limiting factors.

© The Author(s) 2023

B. Ferrer-Rosell et al. (Eds.): ENTER 2023, SPBE, pp. 261–266, 2023.

https://doi.org/10.1007/978-3-031-25752-0_28

2 Literature Review

There is a rich literature in humanities mainly in marketing and tourism management related to consumer reviews on Tripadvisor, Booking, Yelp, and other online booking platforms [1–3]. They all acknowledge the decisive role of reviews and comments on consumer choice. These online reviews are considered more trustworthy than commercial information [4] and therefore influence travelers' decision-making processes. In the particular case of a hotel room, identified by Karpik [5] as a singular good, the uncertainty is high. Therefore, consumers pay importance to judgment provided by others, non-experts, through online reviews. Two dimensions help the consumer in his choice: the arithmetic dimension, through the rating, and the expressive dimension, based on texts. Other research works [6–11] focused on perceived value and its dimensions (functional, price, emotional and social). A consensus emerges on the most persuasive attributes of online reviews, highlighting the importance of source credibility, volume, and valence (average rating) of reviews. While overall, the reviews left are quite positive, reflecting pleasant experiences, some research has focused on negative reviews which, compared to positive ones, are more important because they have a halo effect. Finally, research in services marketing emphasizes the importance of contact staff in the subjective evaluation of tourists.

Besides humanities, the abundance of data (Booking.com hosts more than 200 million authenticated reviews) also stimulates research in computer science concerning the automation of online reviews data collection, processing, and content analysis [4, 12, 13]. Research questions include, among others, how data is collected and annotated but also methods for analyzing reviews content or extracting sentiments.

To the best of our knowledge, there haven't been comparative studies between both approaches. This is why we aim to assess the relative accuracy of the computer-based method compared to a marketing expert's work in extracting sentiment from online travel reviews, and more precisely, translating a customer review into a synthetic evaluation (positive or negative).

3 Methodology

3.1 Data Analysis in Marketing

Content analysis is widely used in social science for interpreting texts. This task relies on the use of a strict methodology, which cuts and classifies the text into semantic units in order to gain a better understanding of the object of study. In this case, we chose the method from Spiggle [14], where the author proposes a content analysis approach divided into 7 steps: categorization, abstraction, comparison, dimensionalization, integration, iteration, and refutation. In categorization, the researcher cuts out and organizes the text by a code system. In our case, the codes are derived from the classic criteria used for hotel classification which include but are not limited to the description of the room, the services offered to the customers, or the role of the contact staff. The first code, made a

priori using the literature on the field, was completed a posteriori by the analysis of the reviews. This results in the definition of more than 40 different categories that characterize the content of the reviews. These initial categories were then grouped during the abstraction stage, according to the frequency of occurrence, into 17 higher-level categories that we detail in Table 1. In the third step, we compare these high-level categories to each other. This helps us to highlight the features that make the overall experience positive or negative. Finally, in the integration step, the researcher attempts to identify the grounded theory in the results, before the iteration step.

Table 1. Higher-level categories according to the frequency of occurrence

| Equipment's criteria | Relational and external criteria |
|--|--|
| Quality of bedding, quality of food, nature and quantity of food, size of room, presence of parking, additional equipment (spa, swimming pool, gym, digital equipment), nature and presence of welcome products (bathrobe, soap), characterization of reception, appreciation of the decoration (recentness, modernity, aesthetic taste) | Interaction with contact staff, accumulation of experience, presence of surrounding nature, sensory appreciation (smell, touch, temperature, security, cleanliness), surrounding noise, location close to the expected points (public transport or place of visit), appreciation of value for money, situational situation (covid, date) |

3.2 Data Analysis in Computer Science

The computer-based analysis relies on a syntactic rule-based approach combined with a pre-trained language model, similar to [15], to extract aspects (hotel services and features) and opinions from customer reviews. The main idea is to take advantage of the grammatical structure of the sentences in reviews. Indeed, a brief analysis of the reviews shows that aspects are mainly nouns or noun groups and opinions correspond to adjectives which are sometimes associated with a modifier like an adverb or a negation word. Moreover, aspects are generally either directly preceded by opinions or part of a verbal phrase where the opinion is the complement and the aspect is the subject. These observations led to the definition of the following four rules:

1. *Subject + Verb + complement* \Rightarrow *aspect + opinion*
2. *Adjective + noungroup* \Rightarrow *aspect + opinion*
3. *Adverb + adjective* \Rightarrow *modifier + opinion*
4. *Negation word + adjective* \Rightarrow *modifier + opinion*

Before applying these different rules to reviews, we start with a classical pre-processing step (splitting reviews into sentences and tokens and then removing unnecessary items like numbers or URLs). The remaining words are processed through a grammatical parser that is able to detect the grammatical function of

each word and the connections between different words. More precisely, it labels words as subjects, verbs, complements, adverbs, and so on. Then, it creates a graph where nodes correspond to words and edges indicate the grammatical links (the subject of a verb for example). Figure 1 illustrate a graph obtained after parsing a sentence.

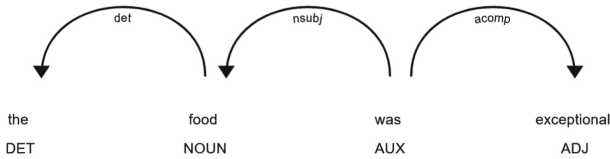


Fig. 1. Graph generated by the parser.

Once the graph is generated, we look for patterns corresponding to the predefined rules inside it. When a pattern is matched, we extract the aspect and the opinion with its negation or adverb separately before regrouping them. Then the general sentiment is obtained by aggregating the polarities corresponding to the different aspects and opinions extracted. The parser, the polarity generator, and the pre-trained language model are provided by Spacy¹. We would like to mention that this approach is not designed to handle complex, ambiguous, or implicit sentiments.

3.3 Data Collection

The reviews exploited are in french and have been scrapped from Booking.com². The collected data is related to the Seine-Maritime department (in the northwest of France). We randomly selected 32 hotels (17% of hotels in Seine-Maritime): 3 were located in the countryside, 9 in small touristic towns, and 21 in the largest cities. The average hotel size in the sample is 45 rooms, 20 hotels are of upper middle-scale class (3 stars), 4 upscale (4 stars) and 8 are mid-scale or economy class (2 and 1 star). Ten reviews, dated between 2011 and 2022, were randomly selected for each hotel in the sample for a total of 320 reviews analyzed.

4 Results

We started with a comparison of the qualitative classification of the reviews. This consists in highlighting the similarities and differences between the criteria found in the two methods (classification step of the marketing approach vs aspects obtained in the rule-based approach). We observe that the elements of categorization highlighted by the manual classification are similar to the computer-based method. However, the rule-based method appears to be more effective in

¹ <https://spacy.io/models/fr>.

² <https://www.booking.com/>.

extracting hotels' aspects related to tangible amenities (bed, shower, elevator, carpark...) or contact personnel (reception, housekeeping, waiter) but its effectiveness decrease in interpreting aspects when the reviews punctuation, syntax and spelling are poor. Nevertheless, the rule-based approach remains relevant and able to manage millions of reviews, which can not be handled by the marketing method.

In the second step, we propose a quantitative analysis to reflect the customer's feelings ("positive or negative") regarding his stay in the hotel. We compared the sentiment analysis output provided by the marketing and the computer-based approaches. Results are presented in Table 2.

In a global overview, the two compared methods concur 93% of the time. This means that the expert and the automatic system globally agree on the decision about reviews sentiment analysis. The difference can be explained by the fact that the computer-based method is less effective when extracting opinions if instead of adjectives, idioms, verbs, or hyperboles are used in the review to express a point of view.

Finally, these results show the necessity to borrow technics and knowledge from the humanities in order to improve the computer-based processing of reviews. Particularly to deal with complex ambiguous or implicit sentiments.

Table 2. Comparison results between marketing and computer-based approaches output in terms of number and proportion of agreement/disagreement

| | Number | Percentage |
|-------------------------------------|--------|------------|
| Agreement in positive evaluation | 284 | 94% |
| Disagreement in positive evaluation | 17 | 6% |
| Agreement in negative evaluation | 15 | 79% |
| Disagreement in negative evaluation | 4 | 21% |

References

1. Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., Weaven, S.: Market segmentation and travel choice prediction in spa hotels through TripAdvisor's online reviews. *Int. J. Hosp. Manag.* **80**, 52–77 (2019). <https://doi.org/10.1016/j.ijhm.2019.01.003>
2. Chang, Y.C., Ku, C.H., Chen, C.H.: Social media analytics: extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *Int. J. Inf. Manage.* **48**, 263–279 (2019)
3. Banerjee, S., Chua, A.Y.: In search of patterns among travellers' hotel ratings in TripAdvisor. *Tour. Manage.* **53**, 125–131 (2016)
4. Marine-Roig, E.: Content analysis of online travel reviews. In: Xiang, Z., et al. (eds.) *Handbook of e-Tourism*, pp. 1–26. Springer, Berlin (2022). https://doi.org/10.1007/978-3-030-05324-6_31-1.pdf
5. Karpik, L.: *L'économie des singularités*, vol. 2. Gallimard Paris (2007)

6. Eid, R.: Integrating Muslim customer perceived value, satisfaction, loyalty and retention in the tourism industry: an empirical study. *Int. J. Tour. Res.* **17**(3), 249–260 (2015)
7. Gallarza, M.G., Saura, I.G.: Value dimensions, perceived value, satisfaction and loyalty: an investigation of university students' travel behaviour. *Tour. Manage.* **27**(3), 437–452 (2006)
8. Gallarza, M.G., Maubisson, L., Rivière, A.: Replicating consumer value scales: a comparative study of EVS and PERVAL at a cultural heritage site. *J. Bus. Res.* **126**, 614–623 (2021)
9. Petrick, J.F.: Development of a multi-dimensional scale for measuring the perceived value of a service. *J. Leis. Res.* **34**(2), 119–134 (2002)
10. Prebensen, N.K., Xie, J.: Efficacy of co-creation and mastering on perceived value and satisfaction in tourists' consumption. *Tour. Manage.* **60**, 166–176 (2017)
11. Wu, C.H.J., Liang, R.D.: Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants. *Int. J. Hosp. Manag.* **28**(4), 586–593 (2009)
12. Piris, Y., Gay, A.C.: Customer satisfaction and natural language processing. *J. Bus. Res.* **124**, 264–271 (2021)
13. Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W.: A survey on aspect-based sentiment analysis: tasks, methods, and challenges. arXiv preprint [arXiv:2203.01054](https://arxiv.org/abs/2203.01054) (2022)
14. Spiggle, S.: Analysis and interpretation of qualitative data in consumer research. *J. Consum. Res.* **21**(3), 491–503 (1994)
15. Rana, T.A., Cheah, Y.N.: A two-fold rule-based model for aspect extraction. *Expert Syst. Appl.* **89**, 273–285 (2017)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

