



HOTFRED: A Flexible Hotel Fake Review Detection System

Michael Möhring¹(✉), Barbara Keller¹, Rainer Schmidt¹,
Matthias Gutmann¹, and Scott Dacko²

¹ School of Computer Science and Mathematics, Munich University
of Applied Sciences, 80335 Munich, Germany
{michael.moehring, barbara.keller,
rainer.schmidt}@hm.edu

² Warwick Business School, Warwick University, Coventry, UK
scott.dacko@wbs.ac.uk

Abstract. The importance to cope with online fake reviews in Tourism becomes more and more evident. In the hotel sector hoteliers as well as guests often struggle with the challenges to separate true and fake reviews from each other. Therefore, our research introduces HOTFRED - a flexible hotel fake review detection system - as part of an on-going research project. By combining different analytical approaches, the HOTFRED system indicates via an aggregated probability whether a review is true or fake. As the evaluation of the prototypical implementation showed, this approach can support to detect fake reviews. Many different stakeholders in the Tourism sector can profit from this automatic tool. Thus, hoteliers can take measures to safe their reputation, guests can benefit in their decision-making process and research might use the tool as an initial starting point for future research in the area of fake information.

Keywords: Online reviews · Fake reviews · Detection · Machine learning

1 Introduction

Online reviews are an important information source for decision making, for instance prior booking a hotel room [1]. However, not all of the information provided via online platforms is reliable. Online reviews at sites like Tripadvisor, Yelp or Google Places are not always written by real customers with a real experience of the hotel. Thus, some of the reviews are fictious and subsequently fake. Fake reviews with false information may lead to wrong decisions by tourists [1, 3]. In general, it is not easy for tourists to detect fake reviews [4]. They are not able the evaluate how trustworthy the provided information is. Additionally, both fake reviews with a negative as well as positive valence can be posted by a fake review writer [1].

In general, there are two ways tourists can detect hotel fake reviews before booking. They are typically either reviewing it manually based on different heuristics (e.g., [5]) or using software tools (e.g., [1, 6]). Due to sparse research on a combination and flexible implementation of different approaches for automatic fake review classification, the presented research aims to create a new system following the design science

research method [7]. Thus, this short paper is part of an ongoing research project addressing the research question “*How can a flexible hotel fake review detection system be designed?*”. In the following a short and comprehensive theoretical background section is given before the proposed solution approach is presented.

2 Background

Besides manual checks by tourists [5] different automatic approaches can be used to detect fake reviews in online environments. Past research showed different approaches on specific samples.

First, meta data of the behavior of the reviewer can be used to check the fake probability of reviews [2, 6, 8, 9]. For instance, if the reviewer writes a lot of hotel reviews in a short time period, this might help to identify fake reviews. Besides meta data, a fake detection can be carried out for example by investigating the writing style, grammar, spelling etc. [10]. Special data sources for training of fake classification models are provided by Yelp.de and through research using crowdsourcing approaches to write fake reviews [4, 10]. Furthermore, research focuses on the comparison of different textual styles between fake and non-fake classified texts [2, 4, 11]. The specific use of some phrases or special spelling issues can lead to the detection of fake reviews, too. Further, there are some online platforms (e.g., reviewmeta.com) providing assistance to identify fake reviews from Amazon. However, flexible hotel review-specific approaches applying different approaches with a scientific base are still missing. Thus, the research wants to address this gap by providing a flexible fake detection system based on different adjusted approaches.

3 Design of the System

A design science research approach was chosen [7, 12] to create a hotel fake review detection system with the aim to answer the proposed research question. Regarding the recommendations, the problem of fake review detection and need of a software solution were identified in Sect. 1 and 2 of the paper. Previous published research work (e.g., [2, 6, 8, 9, 11]) as well as setup several design workshops with several participants were reviewed to collect the relevant objectives of the solution. The primary objective of the system is to determine the probability of fake reviews for a given hotel using several analytical approaches. The system should gather data of the individual reviews collected from online review sites as well as information about the reviewers and the hotel itself. It should have the capability to integrate more analytical approaches stepwise over the time to improve accuracy and integrate current research results. Furthermore, components should be selectable and de-selectable case by case. Based on these objectives, a hotel fake review detection system based on different components has been created. At first, online hotel reviews and related meta data (such as hotel name, reviewers, etc. [2]) have to be collected through a web crawling tool [21] from online review sides like TripAdvisor. These data should be stored in a central database. Thus, central and fast accessible place for data access for the analytical components is shared.

In the following, the analytical components (*here*: (1) text mining-based classification and (2) spell checker) can fall back on the needed data to calculate the probability of fake reviews for a given hotel in a related time frame. The text mining-based classification (1) will use already classified hotel fake review data to calculate the probability of a fake by evaluating textual similarities. The spell checker (2) will calculate a probability based on the amount spelling and grammar issues. Furthermore, the reviewer behavior checker uses data (e.g. timings, hotels, etc. [2]) about the last written reviews of the reviewer to infer on fakes. The hotel environment checker uses data about the hotel to identify fake or incorrect information (e.g., location, stars, facilities). After all components (1 and 2) are analyzed, a scoring system [17] uses the individual probabilities to determine the final probability of fake reviews for a given hotel. The scoring system can run a weighted or unweighted average of the different probabilities. The weights can be adjusted based on trained models and validation after system use. The system architecture is summarized in the following Fig. 1 and allows analytical extensions in the future. Dotted components (reviewer behavior checker, hotel environment checker) are not implemented currently, but will follow up as a part of future research. For a first demonstration a prototype was implemented as explained in the following section.

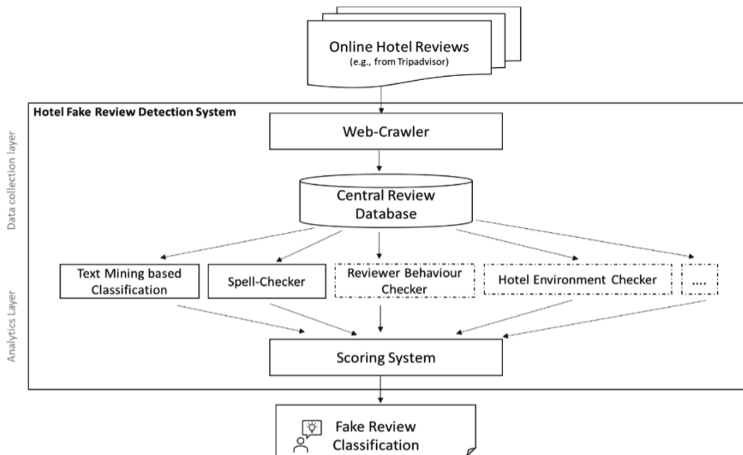


Fig. 1. Architecture of hotel fake detection system

4 First Prototype Development and First Evaluation

Prototypical Implementation: A flexible hotel fake detection system - called HOTFRED - was implemented within a first prototype according to general recommendations coming from previous research [13]. The prototype focuses on the main components to collect data on two major analytical components (1) text mining-based classification and (2) spell checker as well as the scoring system to provide the user (*here*: Tourist) an aggregated, comprehensive information. A web crawler tool [21] was

developed in Python to collect the review data from [tripadvisor.com](https://www.tripadvisor.com). The web crawler has to collect different data of the hotel (e.g., name, URL, class) and the review (e.g., date, review text, points) to run a proper fake review detection and related analysis. After data receiving via HTTPS, it is stored for further analysis within a MySQL database. As a first analytical component (1) a text mining-based fake detection approach was implemented according to the general text pre-processing recommendations [14]. Following, classified fake review data from Yelp was used as a data source for training the classification model [2]. This data set consists of pre-labeled examples regarding the filtered fake characters of hotel reviews written in English. Approximately, 14% of the data can be seen as filtered fake reviews. Existing research already used and validated this data source for e.g. validations [2]. After the evaluation of different classification algorithms (e.g., Support Vector Machines, Naïve Bayes Classifier, KNN), the Support Vector Machine has been chosen as a good fake review classifier based on the accuracy of the classification (e.g., combined metrics like precision, recall, F-score, etc.). For the second analytical component, (2) a spelling checker software tool was developed. This detection component of the system recognizes spelling mistakes based on the ideas of the Levenshtein Distance [15]. The software was programmed in Python. Therefore, the Python library `pyspellchecker` was used. The scoring system component can use the individual results of the finished analytical components to show a summarized view about the fake probabilities of the reviews for the given hotel.

Testing and Evaluation of the Prototype: For the demonstration of the detection system a touristic region in Italy was chosen. Thus, a full sample for all 3 stars and 4 stars hotels in Sorrento/Italy was selected. The sample contained $N = 35370$ reviews for 79 hotels from 3570 different users. For $N = 12$ hotels we found a high probability for having fake reviews within the given timeframe (examples are provided by the authors on request). This stands also in line with previous research, that states about 10%–20% of fake reviews on Yelp [16]. Furthermore, already recognized approaches were used and combined via a flexible scoring model. The accuracy of the trained models was ensured by quality metrics recommended in the literature. Furthermore, some workshops with researchers and potential users were performed based on the results to discuss the progress and application of the HOTFRED prototype of the proposed flexible system. For technical software testing and evaluations of system components (1) and (2), a Swagger interface through a FastAPI [18] deployment is currently under development. The system as well as its components (with deployment possibilities as microservices [20]) can be reachable via a user interface as well as a REST-API call. This function is also currently under development. Furthermore, HOTFRED is to be designed as a scalable system with a fast data processing.

5 Conclusion and Discussion

The detection of hotel fake reviews is an important topic for research and practice as well. On the one hand, tourists are afraid of taking unfavorable or wrong decisions based on fake reviews. On the other hand, hoteliers are afraid that fake reviews harm

their reputation. Therefore, the flexible HOTFRED fake detection system was implemented to cope with the challenges of fake reviews. This approach extends past research (e.g. [2, 4, 9]) in different ways. HOTFRED is designed as a flexible and open tool which enables review detection through different components and allows a case by case selection of these. Therefore, in practice different detection components can be used depending on a use-case specific evaluation. The components can be reached through a defined REST-API, which will be extended and in a currently on-going development project. At the moment, a combined detection approach using a new classified fake detection text model (1) as well as a spell checker (2) is used. In that components, in comparison to other approaches (e.g. [2]), we are using a spell checker focusing on grammar and a classified Yelp dataset not only for validation reasons but also to build a good textual classification model upon it. Additionally, further analytical components as depicted in Fig. 1 are under development. Research as well as practice can benefit from presented research.

Tourists can use the tool to evaluate easily and fast the probability of fake reviews for a given hotel. Business users (like hotel owners) can use HOTFRED to acquire fake review detection capabilities or to develop existing ones; this stands in-line with the current research discussion focusing on fake reviews (e.g., [19]).

Research can benefit from the new architecture which enables a fast as well as broad fake review detection system. At the moment, two interesting fake review detection components (textual and spell checker) are implemented some first preliminary evaluations for the prototype have been run. Additionally, considerations for further needed components have been done to enlarge the system in the future and enhance its predictive power. Research can build upon the results and use it for studies in different fields such as Tourism, information systems and machine learning.

No research is without limitations. First, not all possible detection algorithms could have been implemented and evaluated so far. This aspect is going-to be addressed in future research. Furthermore, it is technically not feasible to have a system with 100% correctness. Aiming to reduce the failure rate, the flexible fake detection system uses different analytical components at the same time. The system is implemented by a prototype. In the future it has to be tested more and confronted with new and original classified fake data to ensure a good evaluation and accuracy. In general, it is hard to get actual fake data to evaluate several systems. Due to the on-going research project, it is planned to further evaluate and extend our detection system by the following steps. At first, more analytical components will be integrated (e.g., hotel environment checker, reviewer behavior checker as recommend also in the literature (e.g. [2])) as well as several software tests will be run. Also, more complex models and combinations of it (e.g. through neural networks) are well-suited opportunities for future improvements and adaptations. Additionally, a qualitative evaluation of the system with experts from the hospitality sector as well as different tourist groups are intended. Furthermore, it is considered to extend the tool to collect review data from different review platforms like Yelp and Google Places and show a combined analysis and results view. After extensions, it is planned to make the tool public for research and

practice as well and collect further feedback to expand promising future research directions.

References

1. Casalo LV, Flavian C, Guinaliu M, Ekinci Y (2015) Do online hotel rating schemes influence booking behaviors? *Int J Hosp Manag* 49:28–36
2. Rayana S, Akoglu L (2015) Collective opinion spam detection: bridging review networks and metadata. In: *Proceedings of the 21th ACM sigKDD*
3. Choi S, Mattila AS, Van Hoof HB, Quadri-Felitti D (2017) The role of power and incentives in inducing fake reviews in the tourism industry. *J Travel Res* 56(8):975–987
4. Yoo KH, Gretzel U (2009) Comparison of deceptive and truthful travel reviews. In: *Proceedings of ENTER*
5. Möhring M, Keller B, Dacko S, et al (2019) Reducing value co-destruction in tourism: an exploration of consumer strategies to detect fake. In: *Proceedings of the Naples forum on service science*
6. Hooi B (2015) BIRDNEST: Bayesian inference for ratings-fraud detection. In: *Proceedings of the international conference on data mining*
7. Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. *MIS Q* 28(1):75–105
8. Ye J, Kumar S, Akoglu L (2016) Temporal opinion spam detection by multivariate indicative signals. In: *Proceedings of the tenth international AAAI conference on web and social media*
9. Lee K, Ham J, Yang SB, Koo C (2018) Can you identify fake or authentic reviews? An fsQCA approach. In: *Proceedings of ENTER*
10. Barbado R, Araque O, Iglesias CA (2019) A framework for fake review detection in online consumer electronics retailers. *Inf Process Manag* 4(56):1234–1244
11. Mukherjee A, Venkataraman V, Liu B, Glance NS (2013) What yelp fake review filter might be doing? In: *Proceedings of ICWSM*, pp 409–418
12. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee SA (2007) Design science research methodology for information systems research. *J Manag Inf Syst* 24(3):45–77
13. Naumann JD, Jenkins AM (1982) Prototyping: the new paradigm for systems development. *MIS Q* 6(3):29–44
14. Tan AH (1999) Text mining: the state of the art and the challenges. In: *Proceedings of the PAKDD workshop*
15. Miller FP, Vandome AF, McBrewhster J (2009) *Levenshtein distance*. Alpha Press, Duesseldorf
16. Luca M, Zervas G (2016) Fake it till you make it: reputation, competition, and yelp review fraud. *Manag Sci* 62(12):3412–3427
17. Rudin C, Ustun B (2018) Optimized scoring systems: toward trust in machine learning for healthcare and criminal justice. *Interfaces* 48(5):449–466
18. FASTAPI (2020) FastAPI framework, high performance, easy to learn, fast to code, ready for production. <https://fastapi.tiangolo.com/>

19. Dacko S, Schmidt R, Möhring M, Keller B (2020) Dealing with fake online reviews in retailing. In: Retail futures. Emerald Publishing Limited
20. Thönes J (2015) Microservices. IEEE Softw 32(1):116–116
21. Mitchell R (2018) Web scraping: collecting more data from the modern web. O'Reilly Media, Sebastopol

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.

