

Liu, Yi; Liu, Huimin; Gao, Xuan; Li, Xiang; Han, Fangfei

Article

Categorizing review spam in tourism and hospitality

Annals of Tourism Research Empirical Insights

Provided in Cooperation with:

Elsevier

Suggested Citation: Liu, Yi; Liu, Huimin; Gao, Xuan; Li, Xiang; Han, Fangfei (2025) : Categorizing review spam in tourism and hospitality, Annals of Tourism Research Empirical Insights, ISSN 2666-9579, Elsevier, Amsterdam, Vol. 6, Iss. 2, pp. 1-4, <https://doi.org/10.1016/j.annale.2025.100186>

This Version is available at:

<https://hdl.handle.net/10419/337859>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

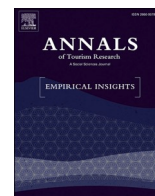
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Categorizing review spam in tourism and hospitality

Yi Liu^{a,b}, Huimin Liu^c, Xuan Gao^{a,*}, Xiang (Robert) Li^d, Fangfei Han^e

^a School of Tourism Management, Sun Yat-sen University, Guangzhou 510275, China

^b Key Laboratory of Sustainable Tourism Smart Assessment Technology, Ministry of Culture and Tourism of China, Zhuhai 519000, China

^c Lee Shau Kee School of Business and Administration, Hong Kong Metropolitan University, 30 Good Shepherd Street, Ho Man Tin, Kowloon, Hong Kong Special Administrative Region

^d School of Sport, Tourism and Hospitality Management, Fox School of Business, Temple University, 1810 N. 13th Street, Philadelphia, PA 19121, United States

^e Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China

ARTICLE INFO

Handling Editor: Ksenia Kirillova

Keywords:

Review spam
Online review
Fake review
Online platform
Typology

1. Introduction

Review spam is an inevitable concern when using tourists' reviews for analysis in tourism and hospitality. In 2022, TripAdvisor removed 1.3 million fake reviews (TripAdvisor, 2023). As technological advances in generative artificial intelligence (AI) now enable the creation of online reviews (Shukla & Goh, 2023), this concern becomes more serious (Li et al., 2023), because the prevalence of review spam can undermine consumer trust (Aye et al., 2013; Tuomi, 2021), harm business reputations (Wu et al., 2020), and reduce purchasing intent (Filiari, 2016).

Researchers currently use the terms “review spam” and “fake reviews” interchangeably, describing them as reviews deliberately created to promote a product's reputation (Shringi et al., 2022). However, we argue that not all review spam is deliberately faked but may exist in irrelevant or type error forms. To this point, Jindal and Liu (2007) identified three types of review spam: 1) false opinions, 2) brand-only reviews (brand-focused, not item-specific), and 3) non-reviews (ads or random text). However, the situation in tourism and hospitality remains unclear.

There are several ways to detect faked reviews, such as identifying consistency (Shan et al., 2021) and unusual posting frequency (Lim et al., 2010; Wang et al., 2012; Z. Wang et al., 2022). More recent studies

have also applied artificial intelligence (AI) and machine learning techniques for this task (Jabeur et al., 2023). But these studies hardly focus on the tourism and hospitality industries and also only focus on fake reviews, not the broader category of review spam. Recognizing this oversight as a research opportunity, this paper thus focuses on exploring the basic types of review spam in tourism and hospitality. This work seeks to provide a foundation for the future development of algorithmic detection tools.

2. Method

To accomplish this aim, we have conducted a pilot study (Liu et al., 2021) using reviews from destinations in China, with the conclusion that review spam is not a significant issue at present. Although the volume of review spam is small, there are more types than previously identified in earlier studies (e.g., Jindal & Liu, 2007). We first collected review data through web scraping from TripAdvisor and Booking.com for the top 10 attractions and top 3 hotels in six international destinations (New York, Los Angeles, Tokyo, Singapore, Hong Kong, Macau), with data spanning from January 2018 to April 2021. A total of 201,517 reviews were initially collected. After removing duplicate reviews (exactly same reviews) caused by crawling errors, we reduced the differences in reviews

* Corresponding author at: School of Tourism Management, Sun Yat-sen University, No. 135, Xingang Xi Road, Guangzhou, China.

E-mail addresses: liuyi89@mail.sysu.edu.cn (Y. Liu), huliu@hkmu.edu.hk (H. Liu), gaox225@mail2.sysu.edu.cn (X. Gao), robertli@temple.edu (X. (R. Li)), hanff5@mail2.sysu.edu.cn (F. Han).

<https://doi.org/10.1016/j.annale.2025.100186>

Received 27 March 2025; Received in revised form 7 July 2025; Accepted 7 July 2025

Available online 12 July 2025

2666-9579/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 2
Comparison of review spam across time periods and platforms.

Profiles of review spam	Total number	Booking.com hotels	TripAdvisor hotels	TripAdvisor tourist attractions	TripAdvisor (in total)
Collected reviews	22,000	6300	3253	12,447	15,700
Detected review spam	546 (2.48 %)	61 (1.0 %)	104 (3.2 %)	381 (3.1 %)	485 (3.1 %)
Numbers and percentages for each type of review spam					
A1 – Meaningless title	166 (30.4 %)	0	40 (38.5 %)	126 (33.1 %)	166 (34.2 %)
A2 – Irrelevant title	84 (15.4 %)	0	37 (35.6 %)	47 (12.3 %)	84 (17.3 %)
B1 – Meaningless content	16 (2.9 %)	5 (8.2 %)	0	11 (2.9 %)	11 (2.3 %)
B2 – Irrelevant topic	63 (11.5 %)	7 (11.5 %)	3 (2.9 %)	53 (13.9 %)	56 (11.5 %)
B3 – Irrelevant object	29 (5.3 %)	0	3 (2.9 %)	26 (6.8 %)	29 (6.0 %)
B4 – Mismatched sentiment	134 (24.5 %)	48 (78.7 %)	12 (11.5 %)	74 (19.4 %)	86 (17.7 %)
B5 – False review	2 (0.4 %)	0	0	2 (0.5 %)	2 (0.4 %)
B6 – Duplicate review	2 (0.4 %)	0	0	2 (0.5 %)	2 (0.4 %)
B7 – Grammatical error	11 (2.0 %)	0	1 (1.0 %)	10 (2.6 %)	11 (2.3 %)
B8 – Non-tourist review	27 (4.9 %)	0	2 (1.9 %)	25 (6.6 %)	27 (5.6 %)
B9 – Inconsistent with title	1 (0.2 %)	1 (0.0 %)	0	0	0
C1 – Duplicate review with the same user ID	4 (0.7 %)	0	4 (3.8 %)	0	4 (0.8 %)
C2 – Pre-trip review	7 (1.3 %)	0	2 (1.9 %)	5 (1.3 %)	7 (1.4 %)

Note: Percentages in the first two rows represent the proportions of review spam in each online review dataset. In subsequent rows, percentages reflect the distribution of each type of review spam within each dataset.

Table 3
Four-category classification of review spam in tourism and hospitality contexts.

Four-category classification	Review spam types	Percentage	
False reviews	B5 – False review	0.4 %	2.8 %
	B6 – Duplicate review	0.4 %	
	C1 – Duplicate review with the same user ID	0.7 %	
Irrelevant reviews	C2 – Pre-trip review	1.3 %	32.2 %
	A2 – Irrelevant title	15.4 %	
	B2 – Irrelevant topic	11.5 %	
Non-reviews	B3 – Irrelevant object	5.3 %	40.2 %
	A1 – Meaningless title	30.4 %	
	B1 – Meaningless content	2.9 %	
	B7 – Grammatical error	2.0 %	
Contradictory reviews	B8 – Non-tourist review	4.9 %	24.7 %
	B4 – Mismatched sentiment	24.5 %	
	B9 – Inconsistent with title	0.2 %	

This typology differs from that in non-tourism contexts [Jindal and Liu \(2007\)](#) and highlights that review spam is broader than fake reviews ([Choi et al., 2017](#)). The patterns are similar across tourism and hospitality. Our study provides a baseline for developing detection tools, recommend type-specific algorithmic approaches. Platforms can improve spam detection algorithms based on the classification in this study, particularly by optimizing the detection of specific spam types, and leverage these classifications to enhance their review screening mechanisms, improving content quality and user trust. Businesses can use the classification to more accurately identify review spam and improve customer satisfaction and brand image by addressing or eliminating such reviews. For users, this study offers an effective tool to help identify and avoid review spam, especially when choosing tourism products or services. By understanding the types of review spam, users can assess review content, avoid being misled, and make more informed purchasing decisions.

6. Limitations

This study has two limitations. First, the manual analysis is prone to bias and human error, and the data were collected from only two platforms, which limits the generalizability of the findings. Second, platform-specific biases exist, such as the rule setting of TripAdvisor led to a large volume of the Meaningless review and Pre-trip review, which may impact the identification of review spam. Future research should explore alternative methods, more platforms, and the relationship

between review spam types and outcomes.

CRedit authorship contribution statement

Yi Liu: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Huimin Liu:** Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Xuan Gao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis, Data curation. **Xiang (Robert) Li:** Writing – review & editing, Conceptualization. **Fangfei Han:** Visualization, Formal analysis, Data curation.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Acknowledgement

This work has been fully supported by National Natural Science Foundation of China (42271182).

References

Ayeh, J. K., Au, N., & Law, R. (2013). “Do we believe in TripAdvisor?” examining credibility perceptions and online travelers’ attitude toward using user-generated content. *Journal of Travel Research*, 52(4), 437–452.

Choi, S., Mattila, A. S., Van Hoof, H. B., & Quadri-Felitti, D. (2017). The role of power and incentives in inducing fake reviews in the tourism industry. *Journal of Travel Research*, 56(8), 975–987.

Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research*, 58, 46–64.

Jabeur, S. B., Ballouk, H., Arfi, W. B., & Sahut, J.-M. (2023). Artificial intelligence applications in fake review detection: Bibliometric analysis and future avenues for research. *Journal of Business Research*, 158, Article 113631.

Jindal, N., & Liu, B. (2007). Analyzing and detecting review spam. In *IEEE international conference on data mining [Icdm 2007: Proceedings of the seventh IEEE international conference on data mining]*. 7th IEEE international conference on data mining, Omaha, NE.

Li, H., Gao, H., & Song, H. (2023). Tourism forecasting with granular sentiment analysis. *Annals of Tourism Research*, 103, Article 103667.

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., & Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on information and knowledge management*.

Liu, Y., Meng, L., Li, X., & Guo, J. (2021). On the Characteristics of Tourism Review Spam: A Comparison of Tourist Scenic Spot and Hotel Reviews. *Tourism Forum*, 14 (03), 31–43. <https://doi.org/10.15962/j.cnki.tourismforum.202103022>

Shan, G., Zhou, L., & Zhang, D. (2021). From conflicts and confusion to doubts: Examining review inconsistency for fake review detection. *Decision Support Systems*, 144, Article 113513.

- Shringi, S., Sharma, H., & Suthar, D. (2022). Fitness-Based Grey Wolf Optimizer Clustering Method for Spam Review Detection. *Mathematical Problems in Engineering*, 2022.
- Shukla, A. D., & Goh, J. M. (2023). *Fighting fake reviews: Authenticated anonymous reviews using identity verification*. Business Horizons.
- Tripadvisor. (2023). **2023 Tripadvisor review transparency report**. <https://www.tripadvisor.com/TransparencyReport2023#group-section-Fake-Reviews-HZjJZOxSZ4>.
- Tuomi, A. (2021). Deepfake consumer reviews in tourism: Preliminary findings. *Annals of Tourism Research Empirical Insights*, 2(2), Article 100027.
- Wang, G., Xie, S., Liu, B., & Yu, P. S. (2012). Identify online store review spammers via social review graph. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4), 1–21.
- Wang, Z., Wei, W., Mao, X.-L., Guo, G., Zhou, P., & Jiang, S. (2022). User-based network embedding for opinion spammer detection. *Pattern Recognition*, 125, Article 108512.
- Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 132, Article 113280.