Proceedings of the 55th Hawaii International Conference on System Sciences | 2022

A Sentiment Analysis of Star-rating: a Cross-Cultural Perspective

Yun Wan
University of Houston-Victoria
wany@uhv.edu

Makoto Nakayama
DePaul University
mnakayama@cdm.depaul.edu

Abstract

Consumer-generated reviews and ratings are critical for the tourism industry. The star rating distribution of services can significantly influence a consumer's decision-making and choice of services. We analyze the star rating distribution of restaurant reviews from three nations (Japan, China, and the U.S.) and find two distribution patterns: bimodal and unimodal. Then, we analyze the sentiment correlation with each star rating across the three cultures. We find the inconsistency of positive sentiment correlation with 5- and 4-star ratings generated by Japanese consumers. Possible contributing factors, including biases, national culture, and socioeconomic conditions, are discussed.

1. Introduction

The online economy is increasingly shaped by product and service ratings and reviews. When a customer cannot inspect the product or evaluate the service directly before purchase or consumption, they have to depend on others, especially so-called electronic word-of-mouth [1]. Online reviews, a major category of electronic word-of-mouth, are widely used by online merchants and consumers to facilitate online shopping. Empirical studies have found that online reviews significantly influence product or service sales [2], [3]. An online review is usually accompanied by a star rating. That is, a reviewer is expected to provide not only a text review of a product or service but also a 1–5 numeric star rating to indicate their overall evaluation of the product or service. A consumer provides a 5-star rating to indicate their best or most satisfactory experience and a 1-star rating to represent the worst. When the star ratings for a product or service are aggregated and averaged, the result is the average star rating. A study by Harvard Business School [4] showed that when an independent restaurant's average star rating increases by one star on Yelp, its revenue rises by 5–9%. Hence, star ratings have a substantial financial impact on product and service revenue.

Important as they are, online reviews are suspectable to noise such as bias (i.e., sequential bias and the Matthew effect [5]) and manipulation (i.e., fake reviews [6]). In addition, when consumers contribute reviews, the review itself is shaped and limited by the demographic background of the reviewer [7]. Meanwhile, the star rating distribution of a target product or service is also suspectable to self-selection bias and other factors [8].

Among research on this meta-topic, one stream focuses on the sentiment analysis of online reviews with the goal of mining desirable product/service features [9], evaluating product/service quality [10], and predicting review helpfulness [11] and star ratings [12]. However, relatively few studies explore this research topic from a cross-cultural perspective, such as the impact of national culture on consumer reliance on reviews [13] or the culture-of-origin effect [14]. Virtually no studies compare review sentiments and their correlation with star ratings from the cross-cultural perspective. For example, does a 5-star rating correlate with the same level of positive review sentiments across different cultures? Or does a 5- to 1-star rating schema mean decreasing positive sentiments or increasing negative sentiments in all cultures? One study found no link between rating and review sentiments using attribution theory [15].

This study bridges this research gap by examining the correlation between the star rating of reviews and their content sentiments as well as validating such correlations' consistency across three national cultures. Our research question is:

R.Q.: Do review sentiments correlate with star ratings across different cultures?

This research question has direct empirical implications for hospitability industries involving international customers. For example, when U.S. tourists search for a local restaurant in Tokyo, should they assume a Yelp Japan review of 5-star rating carries the same level of positive sentiment as a 5-star restaurant review in the United States? Should a 5-star-rated restaurant provide better overall service than a 4-star-rated restaurant?

To explore the above research question, we analyzed three datasets of online reviews and ratings on Japanese restaurants. The datasets were obtained from Yelp.com and Dianping.com. Yelp.com is a leading online review and rating portal in both North America,

URI: https://hdl.handle.net/10125/79906
978-0-9981331-5-7

Europe, and Japan for a wide range of service categories, including hospitability. Dianping.com is a similar leading Chinese online service review portal that covers 4.4 million merchants in all the major cities of China, with 310 million monthly visitors in 2018. Consumers from three countries mainly contributed reviews in these datasets, the U.S., Japan, and China, representing two major (Western and Eastern) cultural groups and covering consumers across all demographic groups in terms of age, gender, and income.

In the remainder of this paper, we first review the existing literature on review sentiments, star rating, and its distributions. Then, we explain our analysis of the Yelp and Dianping datasets and the findings.

2. Research background and literature

In the modern Internet-based economy, consumer-generated content such as online reviews and ratings is highly influenced by the service industry. Online reviews are frequently used by consumers to evaluate product and service quality. Star rating information is mainly used as ranking criteria by service providers or online information brokerage services to help consumers make informed decisions.

2.1. Reviews and their sentiments

The growth of user-generated content in the last 20 years has led to the accumulation of a large amount of opinion information, including blogs, tweets, wiki posts, and online reviews. An online review is opinion information on a target product or service shared by a consumer on an online platform [16]. Although opinion information is subjective, it is instrumental in helping individuals make decisions and has replaced the role of friends and family in the pre-Internet age.

In many cases, opinions are hidden in a large body of text information, making it difficult for humans to read, process, and summarize the opinions expressed in the text efficiently. Thus, automated opinion discovery and summarization systems have been created to serve the purpose, which has led to *sentiment analysis* or *opinion mining* [17].

Online reviews are one of the most important sources of opinions on products and services for businesses and consumers. The target entity of reviews is a product or service. The overall opinion of a review is provided by a star rating (i.e., 5- or 4-star for a positive review, 1- or 2-star for a negative review, and 3-star for neutral). The sentiment analysis of an online review is mainly a sentence-level sentiment classification. When both negative and positive sentiments appear in the same sentence, aspect-based sentiment analysis is used to identify the positive or negative opinions on target aspects such as product/service features and functions. Such analysis is useful for identifying customer preferences and product deficiencies [19].

2.2. Star ratings

The star rating is a 1–5 scale numeric score assigned by consumers to a target product or service to express their opinion when contributing a review. This practice is adopted by most online review platforms. In this study, we examine two online review platforms, Yelp.com and Dianping.com. Both started as restaurant review portals. They also adopt the same star rating system and use similar review preparation interfaces (Figure 1). Moreover, the rating scale (1 to 5 stars) and annotation are identical.

There are some minor interface differences (Figure 1). The Dianping platform explicitly reminds the reviewer to provide an overall rating (总体评价), while Yelp.com does not have such a reminder: it only asks reviewers to "select your rating." Dianping also does this using the prompt "please provide a star rating to the business (点击星星给商户打分)." In addition, the Dianping platform asks reviewers to provide a specific star rating on four service dimensions, right under the overall rating (Figure 1, bottom).

Dianping.com

Yelp.com

Additional service dimension rating on Dianping.com

Figure 1. Rating interface for Dianping (top) and Yelp (middle), and additional service dimension rating on Dianping (bottom)

Although the above differences are unlikely to cause any significant difference in overall restaurant rating by consumers, it does force Chinese reviewers to justify their overall rating in a more rational way.

2.3. Patterns of the star rating distribution

The common belief about the distribution pattern of a randomly polled service evaluation from customers is that it should follow a unimodal normal distribution. However, studies of the star rating distribution pattern of products and services from existing platforms such as Amazon, Airbnb, and Booking.com have found that most are bimodally distributed or follow an extreme distribution [18] [19].

Hu and colleagues [22] analyzed the distribution of product ratings for books, DVDs, and video categories on Amazon with data collected between February and July 2005. They found that 72–78% of the product ratings for books, DVDs, and videos are greater than or equal to 4 stars, showing an asymmetric bimodal (J-shaped) distribution and thus representing a poor proxy of product quality. By contrast, when they conducted a controlled experiment to ask randomly selected subjects to rate the same product, a unimodal rating distribution was observed.

The more comprehensive study of consumer reviews of 24 Amazon product categories by Schoenmüller et al. [23] found that the aggregated distributions of reviews in all categories exhibit a similar bimodal extreme distribution. They found that although 96% of the products across these 24 categories exhibit an extreme distribution, the percentage varies from 84% for Kindle products to 98% for products in the beauty, gourmet food, health, and pet categories.

In a study of book ratings and reviews on both Amazon.com and bn.com, Chevalier and Mayzlin [3] found that reviews are overwhelmingly positive on both websites. They also found that for most samples in the study, the impact of 1-star reviews is higher than the impact of 5-star reviews and that customers prefer to read review texts rather than just depending on summary statistics.

From an investigation of Amazon.com and 17 other platforms ranging from consumer goods to movies, music, and hospitality industries, it was found that 14 of these platforms have extremely distributed product or service ratings more than 50% of the time. Some platforms such as Amazon, Fandango, Airbnb, Yahoo! Songs, and online retailers have more than 90% product ratings extremely distributed [23].

Although the bimodal or extreme distribution is biased in presenting product or service quality, at least some consumers perceive them as more useful and enjoyable than moderate ratings [24]. For hedonic products, extreme rating distributions have been found to be able to reduce the perceived uncertainty of how accurately one can predict decision outcomes in terms of achieving the decision goal [25]. In other words, extreme star ratings may influence consumers' purchase decision-making and thus impact online sales of products and services.

2.3. Cross-cultural sentiment analysis

Consumer sentiment may be mediated by national culture, namely, the norms, behaviors, beliefs, customs, and values shared by the population of a sovereign nation. The correlation between national culture and customers' expectations and perceptions of services is well established in the literature (e.g., Crotts & Erdmann, 2000; Donthu & Yoo, 1998; Furrer et al., 2000; S. Huang & Crotts, 2019; Nakayama & Wan, 2018; Radojevic et al., 2019).

Hofstede and colleagues proposed six cultural dimensions in their series of national culture studies, namely, power distance, individualism vs. collectivism, masculinity vs. femininity, uncertainty avoidance, long- vs. short-term orientation, and indulgence vs. restraint. In the rich stream of cross-cultural research associated with national culture, Hofstede's cultural dimension framework [31], [32] is widely used to assess the impacts of national culture on consumer behaviors. The association between national culture and consumer expectation as well as service-related behavior such as satisfaction, service encounters, and service recovery have been consistently validated in the tourism research field, including many recent studies [33] [30], [34], [35].

Sentiment analysis could provide insights into star ratings and their different distribution patterns. For example, the strongest sentiment correlation is expected to be correlated with 5-star (positive sentiment) and 1-star (negative sentiment) ratings, respectively. Any inconsistency would mean that other factors such as bias and demographics play a mediating role to cause the deviation and the inconsistency details would reveal some clues. Hence, analyzing sentiment in a cross-cultural context could add culture-related insights when such inconsistency appears in certain cultural groups. This is our next step for the study.

3. Review dataset and summary statistics

In this study, we used Japanese restaurant reviews and ratings obtained from Yelp.com and Dianping.com (or three if we consider Yelp U.S. and Yelp Japan as separate websites) as the benchmark for the comparison. A Japanese restaurant is considered to be an upscale restaurant in both the U.S. and China, and many diners use online portals to check reviews before visiting. The datasets from Yelp and Dianping contain Japanese restaurant reviews, and each category has 50,000–100,000 reviews and ratings (Table 1). The number of restaurants in each dataset varies. However, we analyze individual reviews and ratings in our analysis rather than specific restaurants.

Table 1. Summary of Review Sources

Country	# Reviews	Language	City
U.S.	76,704	English	Phoenix, Las Vegas, others
Japan	55,159	Japanese	Tokyo, Osaka
China	88,880	Chinese	Shanghai

We used polarity and imbalance [23] as two scales to identify and measure the level of extreme distribution:

$$Polarity = \frac{Total\ number\ of\ 1\ \&\ 5\ star\ ratings}{Total\ number\ of\ ratings}$$

$$Imbalance = \frac{Total\ number\ of\ 4\ \&\ 5\ star\ ratings}{Total\ number\ of\ 1{,}2{,}4\ \&\ 5\ star\ ratings}$$

For a 5-point scale, a polarity measure above 40% implies an extreme distribution, whereas a polarity measure below 40% implies a non-extreme distribution (e.g., unimodal). An imbalance measure above 50% means more positive than negative reviews and below 50% indicates a majority of negative reviews. The frequencies of each star rating from each country are listed in Table 2.

Table 2. Star Rating Frequencies

Star Rating	Frequency		
	China	Japan	U.S.
5	47758	17160	31324
4	23753	24349	21898
3	10158	11748	10076
2	2881	1756	6494
1	4330	734	6562

The polarity and imbalance for ratings from the three countries are listed in Table 3.

Table 3. Polarity and Imbalance

Country	Polarity	Imbalance
U.S.	50%	80%
Japan	32%	94%
China	59%	91%

This result indicate that Japan Yelp restaurant review ratings demonstrate a non-extreme distribution (32% < 40%), while both U.S. and China Japanese restaurant review ratings have extreme distributions (50% and 59% > 40%). Of these, Chinese reviews have the most extreme distribution. Meanwhile, the ratings from all three countries show an imbalance (all toward positive) in distribution. Considering the extreme rating distribution of Yelp.com for restaurants is 72% [23], the non-extreme distribution of Japanese restaurants is an interesting exception.

4. Aspect-level sentiment classification

To further explore the cause and robustness of this difference, we conducted a sentiment analysis of reviews related to all rating levels. We expected 5-star and 1-star ratings to correspond to the highest positive and negative sentiment correlations on a 5-point scale, respectively.

We used IBM Watson Explorer Content Analytics 11.0 (WCA) to measure the review sentiments in all three languages. WCA uses TAKMI (Text Analysis and Knowledge Mining), which contains an accurate sentiment detector combined with a document processor based on the Unstructured Information Management Architecture standard. It can analyze content in different languages, including Chinese, English, and Japanese [36]. This provides a consistent measurement standard to compare the sentiments in different languages.

IBM WCA provides a direct measurement of the correlations between star rating and positive or negative sentiments. Correlation is “a measure of how strongly a facet value is related to the current query or selection criteria” [36]. It is the ratio of (a) the review proportion containing a sentiment expression given a query criterion (i.e., a star rating) over (b) the review proportion containing the query criterion given all the reviews [37]. Tables 4 to 6 list the sum of aspect sentiment correlations. We highlight the highest sentiment correlation in both the positive and the negative columns in each table.

Table 4. Sentiment for Chinese Reviews

Star Rating	Sentiment		
	Positive	Negative	Overall
5	**320.7**	75.4	396.2
4	207.5	191.2	398.7
3	110.7	332.9	443.6
2	43.9	302.4	346.3
1	63.2	**522.0**	585.2

Table 5. Sentiment for Japanese Reviews

Star Rating	Sentiment		
	Positive	Negative	Overall
5	177.6	58.1	235.7
4	**232.8**	98.8	331.6
3	128.1	123.6	251.7
2	26.4	144.9	171.3
1	6.2	**149.1**	155.2

Table 6. Sentiment for the U.S. Reviews

Star Rating	Sentiment		
	Positive	Negative	Overall
5	**293.2**	69.0	362.1
4	287.7	128.0	415.7
3	179.0	231.0	410.0
2	104.6	371.5	476.0
1	68.6	**462.7**	531.3

It was evident from the results that the 5-star and 1-star ratings all corresponded to the highest positive and negative sentiments correlations, except that the Japanese positive sentiment has the highest correlation with the 4-star rating. This is unusual because it indicated that when Japanese customers have the strongest positive feeling toward the service of a restaurant, they are most likely to provide a 4-star instead of a 5-star rating.

To compare the difference in sentiment between star ratings across three nations/languages, we normalized each sentiment correlation $C_{i,j}$ in Tables 4 to 6:

$$Normalized\ C_{i,j} = \frac{C_{i,j} - C_{\min j}}{C_{\max j} - C_{\min j}}$$

Here, subscript i indicates the star rating (1 to 5) and j indicates the three sentiment categories (positive, negative, and overall). The normalized results for each sentiment categories are illustrated in Figures 2 to 4.

We compared the normalized distribution of the positive sentiment correlation among the 5- to 1-star ratings between China, Japan, and the U.S. (Figure 2). We confirmed that Chinese review sentiments followed a J-curved distribution [22]. By contrast, Japanese rating distributions followed a 180-degree mirror-reversed J-curved distribution, with 4-star ratings correlated with the strongest positive sentiment and 1-star ratings correlated with the lowest positive sentiment. The distribution of U.S. ratings seemed to be the most rational because it followed the strict decreasing positive sentiment correlation from 5- to 1-star ratings.

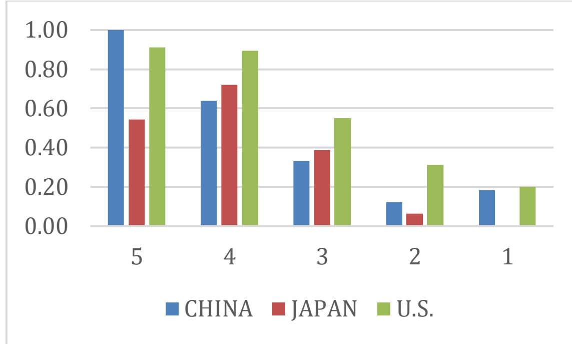

Figure 2. Normalized positive sentiment correlation with the star ratings

The comparison of the normalized negative sentiment correlation with the star rating demonstrated some interesting patterns, too (Figure 3). Both U.S. and Japanese reviews showed a consistently increasing negative sentiment correlation with decreasing star ratings, which is rational consumer behavior. Chinese reviews showed slight irregularity, with a higher negative sentiment for a 3-star rating than a 2-star rating. This means Chinese reviewers may demonstrate more negative sentiment but provide better ratings when they experience mediocre services. This behavior is like Japanese reviewers assigning a 4-star rating to demonstrate more positive sentiments than a 5-star rating.

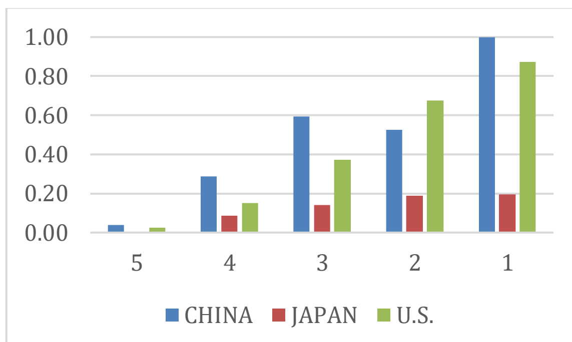

Figure 3. Normalized negative sentiment correlation with the star ratings

The comparison of the overall sentiment correlation with the star ratings revealed interesting new patterns (Figure 4). The first pattern is consumers from China and the U.S. showed more overall sentiments (positive and negative sentiments combined) when they provided lower star ratings than higher star ratings. By contrast, Japanese customers were the opposite. This indicates that Chinese and U.S. consumers tended to be more passionate when justifying an extremely low rating, while Japanese consumers tended to be more emotional when justifying a better experience.

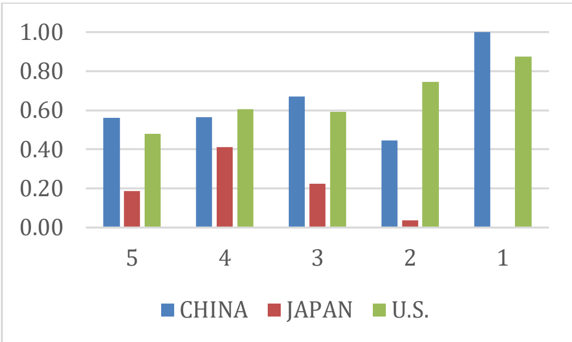

Figure 4. Normalized overall sentiment correlation with the star ratings

The second pattern is Chinese and U.S. consumers were generally more emotional or passionate in their reviews than Japanese customers. This is dramatic in the 1-star rating, of which Chinese and U.S. reviewers both demonstrated the highest sentiment correlations, while Japanese customers showed almost none. The ratios of sentiment correlations among Chinese, Japanese, and U.S. reviewers are listed in Table 7.

Table 7. Ratio Comparison

Country	Sentiment Ratio		
	Positive	Negative	Overall
China	1.31	2.48	1.89
Japan	1.00	1.00	1.00
U.S.	1.63	2.20	1.92

5. Discussion

5.1. Bias, cognitive dissonance, and adjustment

Most studies thus far have attributed the star rating distribution to self-selection bias, which arises from biased sampling. That is, the characteristics of the individuals cause them to select themselves in a specific group, creating abnormal or undesirable conditions in the group [22]. For example, suppose only those consumers motivated by extreme emotion (love or hate) bother to contribute their ratings and reviews about a product or service. In that case, we have self-selection bias because the supposedly unbiased sampling of collective ratings comes from extremely motivated individuals instead of the general population.

Hu et al. [22] argued that the J-shaped extreme distribution could be caused by a combination of two self-selection biases: *underreporting bias* (people with moderate views are less passionate about spending the time and effort to report their ratings) and *purchase bias* (people who purchase a product are more likely to write positive reviews). They also concluded that because of self-selection bias, average ratings on Amazon.com are not representative of true product quality.

Our sentiment correlation analysis supported self-selection bias in 1-star ratings by both Chinese and U.S. consumers because they received the highest overall sentiment in both cases.

Related to purchase bias and as an alternative explanation, cognitive dissonance, which refers to having inconsistent thoughts, beliefs, or attitudes, mainly as relating to behavioral decisions and attitude change [38], was found to contribute to the positive extreme. One type of dissonance called post-decision dissonance could explain purchase bias that leads to the extreme distribution of online review ratings. Because cognitive dissonance is perceived as unpleasant, the individual wants to reduce it after the decision-making process, such as the purchase of a product or consumption of a service. As a result, they tend to share their opinions, such as posting positive reviews for the product or service, to justify their decision and reduce dissonance, as evidenced by recent findings [39].

Our sentiment correlation analysis provides evidence of cognitive dissonance in 5-star ratings for all culture groups because the overall sentiments of 5-star ratings in all of them are less than those for 4- or 3-star ratings. This means that consumers may only use the 5-star rating to justify their decision and reduce dissonance, hence lacking strong sentiment.

Related to self-selection bias, minor contributing factors such as reviewer adjustment and the adjustment effect could enhance the impact of the extreme distribution, although not decisively [23]. For example, review fraud could contribute to both positive and negative extreme ratings because fake reviews tend to either boost a product or criticize its competitors' products.

The lack of strong overall sentiments in a 5-star rating could indicate the existence of an adjustment effect, but the strong sentiment correlation with a 1-star rating indicates otherwise. Hence, there is no consistent evidence in this study to support this explanation.

5.2. Cultural influences

According to Hofstede's cultural dimension theory, Japanese culture has a far higher *uncertainty avoidance* index (92) than China (30), the U.S. (46), and other Western cultures. Consumers from a national culture with high uncertainty avoidance have a distinctive preference in many aspects, such as travel attributes [40]. An important feature of a high uncertainty avoidance culture is avoiding confrontation, which could even lead to no action on unsatisfactory services for Japanese customers [41]. This is consistent with our findings that only 1.3% of all the ratings provided by Japanese reviewers were 1-star ratings (Table 1).

Most importantly, high uncertainty avoidance correlates with an essential Japanese cultural feature called Kaizen (改善, かいぜん). In Japanese culture, Kaizen means seeking continual improvement, eliminating defects, and trying to find better ways in everything we do. In other words, the pursuit of perfection. Kaizen led to Total Quality Management and other revolutionary management practices in Japan [42].

From a Kaizen perspective, no restaurant service is perfect, so it is likely Japanese reviewers are reluctant to provide a 5-star rating, which means perfect, to a restaurant, even though they like the service very much. In other words, a Japanese customer may provide a restaurant either a 4- or a 5-star rating as the highest rating, and both correlate with the highest level of positive sentiments. However, most choose to provide a 4-star rating due to influence of Kaizen. This explains the unusual pattern in Table 5.

Additional evidence from other online review platforms would provide more information to allow us to offer a better interpretation, especially for Japan. One such platform could be the Japanese-only restaurant review site tabelog.com: most restaurant ratings on the platform are around 3.5 out of 5, another unimodal case.

5.3. Socioeconomic impact

The overall stronger sentiment manifested in Chinese reviews indicated the possible impact of the confluence of national culture and changing socioeconomic status, such as rising consumerism, on consumer review and rating characteristics. Recent studies have found that customer rage incidents have risen in Eastern collectivistic cultures such as China with emerging consumerism and rising wealth [43]. The stronger Chinese consumer sentiments in reviews and the distinction from Japanese consumer sentiments, even though they belong to the same Eastern culture group, support such observations.

6. Conclusion

This study demonstrates that online star ratings may be inconsistent with their corresponding sentiment levels. Self-selection bias motivates consumers to post reviews with polarized ratings, while varying national culture characteristics and changing socioeconomic status may nudge the rating in the reverse direction.

Our findings indicate that consumers must use the rating distribution and average rating information in an international destination with a grain of salt, especially as the rating data are contributed mostly by local customers. They also suggest that hospitality businesses should incorporate the possible cultural influence into their interpretation of review and rating information on a service in different cultural and socioeconomic contexts.

Our research is limited to only three culture groups and restaurant categories. The external validity of our results thus needs further confirmation using different cultural groups and service categories. The challenge remains for data scientists to infer accurate quality evaluation from reviews and ratings. However, more factors may be considered in the future, such as platform adjustment and rating spam [44]. Consider this fact: the acclaimed Ono's sushi restaurant in Tokyo, which was featured in the documentary *Jiro Dreams of Sushi* [45] and is considered to be the best sushi restaurant in the U.S., only receives a 4-star rating on Yelp Japan. Many foreign tourists could thus miss it if they use Yelp Japan to search for a nearby restaurant.

7. References

[1] Y. Wan, M. Nakayama, and N. Sutcliffe, "The impact of age and shopping experiences on the classification of search, experience, and credence goods in online shopping," *Inf. Syst. E-bus. Manag.*, vol. 10, no. 1, pp. 135–148, 2012.

[2] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," *J. Mark.*, vol. 74, no. 2, pp. 133–148, 2010, doi: 10.1509/jmkg.74.2.133.

[3] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *J. Mark. Res.*, vol. 43, no. 3, pp. 345–354, 2006.

[4] M. Luca, "Reviews, reputation, and revenue: The case of Yelp. com," No. 12-016, 2016. [Online]. Available: https://ssrn.com/abstract=1928601.

[5] Y. Wan, "The Matthew Effect in social commerce: The case of online review helpfulness," *Electron. Mark.*, vol. 25, no. 4, pp. 313–324, 2015, doi: 10.1007/s12525-015-0186-x.

[6] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and Yelp review fraud," *Manage. Sci.*, vol. 62, no. 12, pp. 3412–3427, 2016.

[7] S. Bakhshi, P. Kanuparthy, and E. Gilbert, "Demographics, weather and online reviews: A Study of Restaurant Recommendations," *Proc. 23rd Int. Conf. World wide web - WWW '14*, pp. 443–454, 2014, doi: 10.1145/2566486.2568021.

[8] N. Hu, J. Zhang, and P. A. Pavlou, "Overcoming the J-shaped distribution of product reviews," *Commun. ACM*, vol. 52, no. 10, pp. 144–147, 2009.

[9] W. Ariyasriwatana and L. M. Quiroga, "A thousand ways to say 'Delicious!'—Categorizing expressions

of deliciousness from restaurant reviews on the social network site Yelp," *Appetite*, vol. 104, pp. 18–32, 2014, doi: 10.1016/j.appet.2016.01.002.

[10] W. Duan, Q. Cao, Y. Yu, and S. Levy, "Mining online user-generated content: Using sentiment analysis technique to study hotel service quality," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2013, pp. 3119–3128, doi: 10.1109/HICSS.2013.400.

[11] T. L. Ngo-Ye, A. P. Sinha, and A. Sen, "Predicting the helpfulness of online reviews using a scripts-enriched text regression model," *Expert Syst. Appl.*, vol. 71, pp. 98–110, 2017, doi: 10.1016/j.eswa.2016.11.029.

[12] J. Qiu, C. Liu, Y. Li, and Z. Lin, "Leveraging sentiment analysis at the aspects level to predict ratings of reviews," *Inf. Sci. (Ny).*, vol. 451, pp. 295–309, 2018.

[13] R. Y. Kim, "Does national culture explain consumers' reliance on online reviews? Cross-cultural variations in the effect of online review ratings on consumer choice," *Electron. Commer. Res. Appl.*, vol. 37, 2019, doi: 10.1016/j.elerap.2019.100878.

[14] M. Nakayama and Y. Wan, "Is culture of origin associated with more expressions? An analysis of Yelp reviews on Japanese restaurants," *Tour. Manag.*, vol. 66, pp. 329–338, 2018, doi: 10.1016/j.tourman.2017.10.019.

[15] J. Jiang, U. Gretzel, and R. Law, "Do Negative Experiences Always Lead to Dissatisfaction? — Testing Attribution Theory in the Context of Online Travel Reviews," in *Information and Communication Technologies in Tourism 2010*, Springer, Vienna, 2010, pp. 297–308.

[16] S. M. Mudambi and D. Schuff, "Research note: What makes a helpful online review? A study of customer reviews on Amazon. com," *MIS Q.*, vol. 34, no. 1, pp. 185–200, 2010.

[17] B. Liu, "Sentiment analysis and subjectivity.," *Handb. Nat. Lang. Process.*, vol. 2, no. 2010, pp. 627–666, 2010.

[18] V. S. Jagtap and K. Pawar, "Analysis of different approaches to sentence-level sentiment classification," *Int. J. Sci. Eng. Technol.*, vol. 2, no. 3, pp. 164–170, 2013.

[19] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, 2015.

[20] N. Hu, P. A. Pavlou, and J. J. Zhang, "On Self-Selection Biases in Online Product Reviews.," *MIS Q.*, vol. 41, no. 2, pp. 449–471, 2017.

[21] W. W. Moe and D. A. Schweidel, "Online product opinions: Incidence, evaluation, and evolution," *Mark. Sci.*, vol. 31, no. 3, pp. 372–386, 2012.

[22] J. J. Heckman, "Sample selection bias as a specification error," *Econometrica*, vol. 47, no. 1, pp. 153–161, 1979.

[23] V. Schoenmüller, O. Netzer, and F. Stahl, "The Extreme Distribution of Online Reviews: Prevalence, Drivers, and Implications," No. 18-10, 2018. Accessed: May 09, 2020. [Online]. Available: https://ssrn.com/abstract=3100217at:https://ssrn.com/abstract=3100217.

[24] S. Park and J. L. Nicolau, "Asymmetric effects of online consumer reviews," *Ann. Tour. Res.*, vol. 50, pp. 67–83, 2015.

[25] W. Chu, M. Roh, and K. Park, "The effect of the dispersion of review ratings on evaluations of hedonic versus utilitarian products," *Int. J. Electron. Commer.*, vol. 19, no. 2, pp. 95–125, 2015.

[26] N. Donthu and B. Yoo, "Cultural influences on service quality expectations," *J. Serv. Res.*, vol. 1, no. 2, pp. 178–186, 1998, doi: 10.1177/109467059800100207.

[27] J. C. Crotts and R. Erdmann, "Does national culture influence consumers' evaluation of travel services? A test of Hofstede's model of cross-cultural differences," *Manag. Serv. Qual. An Int. J.*, vol. 10, no. 6, pp. 410–419, Dec. 2000, doi: 10.1108/09604520010351167.

[28] O. Furrer, B. S. C. Liu, and D. Sudharshan, "The Relationships between Culture and Service Quality Perceptions: Basis for Cross-Cultural Market Segmentation and Resource Allocation," *J. Serv. Res.*, vol. 2, no. 4, pp. 355–371, 2000, doi: 10.1177/109467050024004.

[29] T. Radojevic, N. Stanisic, and N. Stanic, "The culture of hospitality: From anecdote to evidence," *Ann. Tour. Res.*, vol. 79, Nov. 2019, doi: 10.1016/j.annals.2019.102789.

[30] S. Huang and J. Crotts, "Relationships between Hofstede's cultural dimensions and tourist satisfaction: A cross-country cross-sample examination," *Tour. Manag.*, vol. 72, no. December 2018, pp. 232–241, 2019, doi: 10.1016/j.tourman.2018.12.001.

[31] G. H. Hofstede and G. H. Hofstede, *Culture's Consequences: International Differences in Work-related Values*. SAGE, 1984.

[32] G. Hofstede, *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE Publications, 2001.

[33] M. Schuckert, X. Liu, and R. Law, "A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently," *Int. J. Hosp. Manag.*, vol. 48, pp. 143–149, 2015, doi: 10.1016/j.ijhm.2014.12.007.

[34] S. (Sixue) Jia, "Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews," *Tour. Manag.*, vol. 78, 2020, doi: 10.1016/j.tourman.2019.104071.

[35] P. Stamolampros, D. Dousios, N. Korfiatis, and E. Symitsi, "The joint effect of consumer and service providers' culture on online service evaluations: A response surface analysis," *Tour. Manag.*, vol. 78, 2020, doi: 10.1016/j.tourman.2019.104057.

[36] W.-D. J. Zhu *et al.*, *IBM Watson Content Analytics: Discovering Actionable Insight from Your Content*. IBM Redbooks, 2014.

[37] M. Nakayama and Y. Wan, "Cross-Cultural

Examination on Content Bias and Helpfulness of Online Reviews : Sentiment Balance at the Aspect Level for a Subjective Good," in *HICSS 52*, 2019, vol. 6, pp. 1154–1163.
[38] L. Festinger, *A theory of cognitive dissonance*, vol. 2. Stanford university press, 1957.
[39] J. Berger, "Word of mouth and interpersonal communication: A review and directions for future research," *J. Consum. Psychol.*, vol. 24, no. 4, pp. 586–607, 2014.
[40] R. B. Money and J. C. Crotts, "The effect of uncertainty avoidance on information search, planning, and purchases of international travel vacations," *Tour. Manag.*, vol. 24, no. 2, pp. 191–202, 2003, doi: 10.1016/S0261-5177(02)00057-2.
[41] J.-H. Huang, C.-T. Huang, and S. Wu, "National character and response to unsatisfactory hotel service," *Int. J. Hosp. Manag.*, vol. 15, no. 3, pp. 229–243, 1996.
[42] R. Recht and C. Wilderom, "Kaizen and culture: on the transferability of Japanese suggestion systems," *Int. Bus. Rev.*, vol. 7, no. 1, pp. 7–22, 1998.
[43] P. G. Patterson, M. K. Brady, and J. R. McColl-Kennedy, "Geysers or Bubbling Hot Springs? A Cross-Cultural Examination of Customer Rage From Eastern and Western Perspectives," *J. Serv. Res.*, vol. 19, no. 3, pp. 243–259, Aug. 2016, doi: 10.1177/1094670516649057.
[44] M. Schuckert, X. Liu, and R. Law, "Insights into Suspicious Online Ratings: Direct Evidence from TripAdvisor," *Asia Pacific J. Tour. Res.*, vol. 21, no. 3, pp. 259–272, Mar. 2016, doi: 10.1080/10941665.2015.1029954.
[45] D. (Director) Gelb, K. (Producer) Iwashina, and T. (Producer) Pellegrini, *Jiro Dreams of Sushi [Documentary]*. Magnolia Pictures, 2012.