

Internet Review Analysis of Foreign Visitors to Regional Cities in Japan

Yasushi Sugiyama ^{*}, Jianping Zheng [†], Tokuro Matsuo [†],
Hidekazu Iwamoto [‡], Teruhisa Hochin ^{*}

Abstract

In this paper, we analyzed and examined review data which is in various languages of foreign visitors to a local city where Hamamatsu city, Shizuoka prefecture, Japan is. Target data were in the categories of "hotel" and "tourist spots" in multilingual in Chinese and English. And the word appearance frequency and the result of emotion analysis were obtained from review data by using NLPiR which is a morphological analysis tool. By conducting knowledge discovery, the characteristics and differences of foreign travelers visiting local cities by nationality are clarified. By the method used in this study, local cities can increase the marketing ability to attract foreign tourists.

Keywords: Inbound marketing, Sightseeing resources, Local cities, Internet reviews, Text mining

1 Introduction

Toward the Tokyo Olympic Games 2020 in two years ahead, the number of foreign visitors to Japan continues to increase. The number of foreign visitors increased to 28,691 thousand (19.3% increase over the previous year) in 2017. In terms of nationality, the top three countries are China (7,355 thousand), Korea (7,140 thousand), and Taiwan (4,167 thousand). The number of foreign visitors to Japan from East Asia (including Hong Kong added to them) reached 21,291 thousand, closing the total number of 74.2%. From the perspective of economic ripple effects, the total consumption in 2017 was estimated to be 4,416.2 billion yen (an increase of 17.8% from the previous year). By nationality, China is ranked as the first in 1,694.7 billion yen (38.4% of the total) followed by Taiwan's 574.4 billion yen (13.0%), South Korea's 512.6 billion yen (11.6%) and Hong Kong's 34.16 billion yen (7.7%) From the viewpoint of consumption, the proportion of visitors from China and Taiwan is 51.4% of the total and the importance stands out [1]. On the

^{*} Kyoto Institute of Technology, Kyoto, Japan

[†] Advanced Institute of Industrial Technology, Tokyo, Japan

[‡] Josai International University, Chiba, Japan

other hand, the revitalization of local cities is one of the important issues in our country. It is known that the number of foreign tourists coming to local cities is very small compared to those visiting major cities such as Tokyo and Osaka. Governments and local governments have launched various policies such as the Japanese version of DMO to attract inbound to local cities, but it is still at the beginning stage.

In this research, in order to attract foreign tourists to local cities in Japan, we will consider marketing specialized in local cities. In recent years, the marketing method has been drawing attention by using reviews on the Internet review site [2] [3]. Travel related information is famous for travel information sites such as "Trip Advisor" and "Expedia" and bulletin boards such as "Yahoo!". Travelers will write impressions and experiences on various blogs, SNS, travel sites, and information sites in various countries. It is useful to understand the impression and evaluation of visitors to the country by collecting and analyzing such reviews [4]. As a typical research objective, the review is one piece of new information, its occurrence situation, and its impact. Studies and research on the relationship between consumers who exchange reviews and networks (connections) are also conducted.

In the rest of this paper, Section 2 mentioned the related work. Section 3 introduces the methodology of this survey. "Preliminary discussion," "Data collection," and "Data analysis." Section 4 mentioned the results of this analysis. Section 5 shows the discussion. Then, the conclusion of this paper is shown in Section 6.

2 Related Work

The sending and spreading of information via the Internet are effective for advertisements. It is widely known that it is used as marketing information by collecting and analyzing user reviews. This kind of study is conducted all over the world. Schuckert et al. [5] mentioned that online comments are essential to tourism. Travelers can directly comment on their travel experience through the Internet, so these comments on the Internet can be one of the important factors for future travelers. Xu et al. [6] explained that big data had been used for the analysis of travelers' comments. An analytical method, such as text mining, is useful for collecting data from the Internet. Hsiao et al. [7] identified important factors to develop guidelines for hotel services by using text mining. Text mining was applied to extract important words of service characteristics. They extracted travelers' online comments on twenty international hotels in Taipei and summarized ten service characteristics. They described that the facility is an important factor in developing hotel service.

This study also showed one of the examples that text mining is a useful tool to extract travelers' comments. Therefore, in this research, as effective knowledge for inbound marketing, we analyze multilingual reviews on the Internet and clarify features and differences by nationality.

Also in Japan, analysis and research on foreign traveler reviews and blog data by text mining and statistical methods are being conducted. These are useful for inbound business. Sugitani [2] pointed out the usefulness of reviews on the Internet "the effect of nonverbal clues in product evaluation." Okubo [8] showed a method for analyzing the language contents and extracting useful plan information for travel guide books and review sites. Tanabe [9] revealed that useful analysis could be obtained from the review analysis of the guests posted on the reservation site regarding strategy development of accommodation facilities.

3 Methodology

3.1 Preliminary Discussion

In this research, the purpose was to investigate what options are available for visitors. Then, we have selected "Hotels" and "Sightseeing spots," which seem to be very important items for visitors. And we also focused on whether there are any differences depending on nationality when foreign visitors to Japan visit local cities. Especially, Chinese and Taiwanese are important from the marketing point of view. As for the data collection aspect, we selected a regional city in Japan and investigated user reviews of Internet travel sites. The target city was Hamamatsu City, which is located between Tokyo and Osaka, and there is a Shinkansen station. The population is about 820 thousand. Manufacturing and agriculture are important industries. Local governments are actively implementing policies for inbound projects, and international conference halls are also substantial.

Table 1: Survey items of "Hotels"

Hotel	Facility			Restaurant	Conference room	Access to Station
	No. of Rooms	Parking lot	Hot spring			
A)	211	available	n/a	Breakfast only	n/a	3mins walking
B)	212	available	n/a	1 Restaurant (J)	2 rooms (10 seats x 2)	13mins walking
C)	224	available	n/a	4 Restaurants (J/Ch/Eur/Tea)	4 rooms (total 600 seats)	5mins walking
D)	322	available	n/a	4 Restaurants (J/Ch/Eur/Teppanyaki)	4 rooms (total 984 seats)	3mins walking
E)	138	n/a	n/a	1 Restaurant (J)	n/a	2mins walking
F)	71	available	n/a	n/a	n/a	0.5mins walking
G)	391	available	available	5 Restaurants (J/Ch/F/Buffer/Tea)	4 rooms (total 1070 seats)	30mins by taxi
H)	110	available	available	1 Restaurant	n/a	12mins by train
I)	192	available	n/a	6 restaurants (Jx2/Ch/Buffer/Bar/Lounge)	6 rooms (total 785 seats)	3mins walking

Table 2 : Survey items of "Sightseeing spots"

Tourism Resources	Overview
1) Lake Hamanako	10th biggest lake in Japan. Eel aquaculture is famous.
2) Hamamatsu Castle Park of Hamamatsu Castle.	Castle constructed in the 15th century.
3) Museums etc.	Hamamatsu is famous for automotive/instrument industry.
4) Historic buildings	Old station (Tenryu futamata Station) etc.
5) Temple, Shrine etc.	Ryutan temple, Houkou temple etc.
6) Nature, Parks etc	Hamanako nature park etc.

3.2 Data collection

A certain amount of data is necessary to perform data analysis. However, there are a limited number of review data by foreign visitors on Hamamatsu City, which can be collected on the Internet website. Therefore, as shown in Table 1, we collected data on nine hotels where foreigners stay relatively frequently. And review data on six sightseeing spots shown in Table 2, which are relatively frequent by foreigners, were investigated as well. As shown in Table 3, data collection of review data have been performed on several major travel sites, such as "Trip Advisor [10]", "Expedia [11]", "Agoda [12]", "Booking.com [13]" and "Ctrip.com [14]". As for the data collection method, we used "GooSeeker [15]" and "Octoparse [16]" which are Web API software tools. These tools easily realized the data scraping from the web site. The collected multilingual review data were classified in nationalities, such as China, Taiwan, Hong Kong and other nationalities (Korea, Singapore, Thailand, Europe, America, etc.). Samples of the collected data are shown in Table 4. These data include the country and region of visitors, evaluation of the hotel, impression of the hotel, visit time, purpose of stay and comments.

Table 3: Data collection in detail

Language	English/Chinese/Taiwanese
Number of corrected data	about 1500
Data correction website	TripAdvisor, https://www.tripadvisor.com Expedia, https://www.expedia.com/ Agoda, https://www.agoda.com/ Booking.com, https://www.booking.com/ Ctrip, https://www.ctrip.com/

Table 4: Sample of collected data (a part) for Hotel D)

Nationality	Score	Impression	Time	Purpose	Comment
unanswered	4	Beautiful Hotel	May,2017	Couple Trip	We stayed here for one night
Suvanjoya, Malaysia	4	5-star std	May,2017	Trip with Family	big hotel; my room at 32th
Kuala Lumpur, Malaysia	4	Hotel beside station	Jun,2017	Travel with friends	It is the tallest hotel
Normal, IL	5	Wonderful hotel	May,2017	Business Trip	Very convenient to the station
Hong Kong, China	5	Beautiful view	Mar,2017	Trip with family	The hotel is the highest in
Newberry, UK	5	My favorite hotel	Mar,2017	Business Trip	I have stayed here for 3 nights

3.3 Data analysis

For data analysis, freeware of text mining using elemental technology of natural language processing is used, NLPIR [17] (Chinese, Taiwan compatible). This tool can be used for the analysis of nouns, verbs, adjective appearance frequency, keyword extraction, emotion analysis and so on. NLPIR has high operability as a morphological analyzer in Chinese and has many clients including universities, laboratories, banks, and ICT companies. We analyzed

the collected data with NLPIR and created an analysis report of "hotels," "sightseeing spots," and "foods and drinks." The analysis was also made by nationality. In addition, morphological analysis was performed on the entire collected data. The whole sentence, the keyword, the noun, the verb, the appearance frequency of the adjective, and the outline of the emotion analysis result were obtained. In this research, we focused on nouns, verbs, and adjective appearance frequency and emotion analysis results. Then, it is clarified the results of the appearance frequency of nouns, verbs, and adjectives.

As for the analysis of Hotel D) by the 45 Chinese visitors review data, the numbers of nouns extracted by analysis were Hotel (55), Hamamatsu (18), Breakfast (11), Room (10), Location (9), Japan (8), Equipment (8), Geography (8), Sensation (8), and Restaurants (7).

Then, in words extracted, we omit words not used for analysis. Words that can be omitted are "Hotel," "Hamamatsu," "Japan," and "Sense." Furthermore, different words representing the same meaning are treated as the same word. For example, even if expressions such as "Location" and "Geography" are different, these words are treated as "location."

After the treatment mentioned above, the number of the words "location" is 17, as an example. We performed the same treatment to the other words. As a result, in the noun, location (17), breakfast (11), room (10), facility (8), and restaurant (5) were extracted. For the verb, there were service (9), recommend (6), walk (5), and value (3). And for the adjective, there were good (17), advantage (11), convenient (10), and clean (4). It is necessary to consider that there was a large difference in appearance frequency of words due to difference in collected data amount on each survey target. In case of Hotel D), there were 45 reviews and including extracted words of "Hotel" (17) and "Breakfast" (11). In the case of Hotel F), there were 11 reviews, including extracted words of "hotel" (7) and "breakfast" (4). Therefore, it is useful to perform data normalization based on the population of reviews. That is, when data X is given, it is normalized by $Y = (X - \min) / (\max - \min)$. As a result, the maximum value is 1, while the minimum value is 0. The value is rounded to two decimal places.

These analysis results are shown in Table 5.

Table 5: Appearance Frequency of words for Hotel D)

Nouns	No. of word appearances	After normalization
location	17	1.00
breakfast	11	0.50
room	10	0.42
facility	8	0.25
restaurant	5	0.00
Verbs	No. of word appearances	After normalization
service	9	1.00
recommend	6	0.60
walk	5	0.40
value	3	0.00
Adjectives	No. of word appearances	After normalization
good	17	1
advantage	11	0.54
convenient	10	0.46
clean	4	0

Concerning emotion analysis, the proportion of positive emotion and negative one could be evaluated by sentences. In Table 6, the emotion analysis result for Hotel D) is shown. The positive is 83.97%, which corresponds to 373 words evaluated as “positive” and to 71 words evaluated as “negative” in the whole text of reviews written by 45 Chinese visitors.

Table 6: Emotion analysis results for Hotel D)

Positive	%	Negative	%
Easy	9.97%	Angry	0.10%
Good	74.00%	Sad	2.52%
		Fear	0.75%
		evil	11.00%
		shock	1.26%
Total	83.97%	Total	16.03%

The results of emotion analysis for the reviews by Chinese visitors in nine hotels are shown in Figure 1. The most positive value is Hotel B) of 91.36%, and the lowest value is Hotel H) of 68.82%.

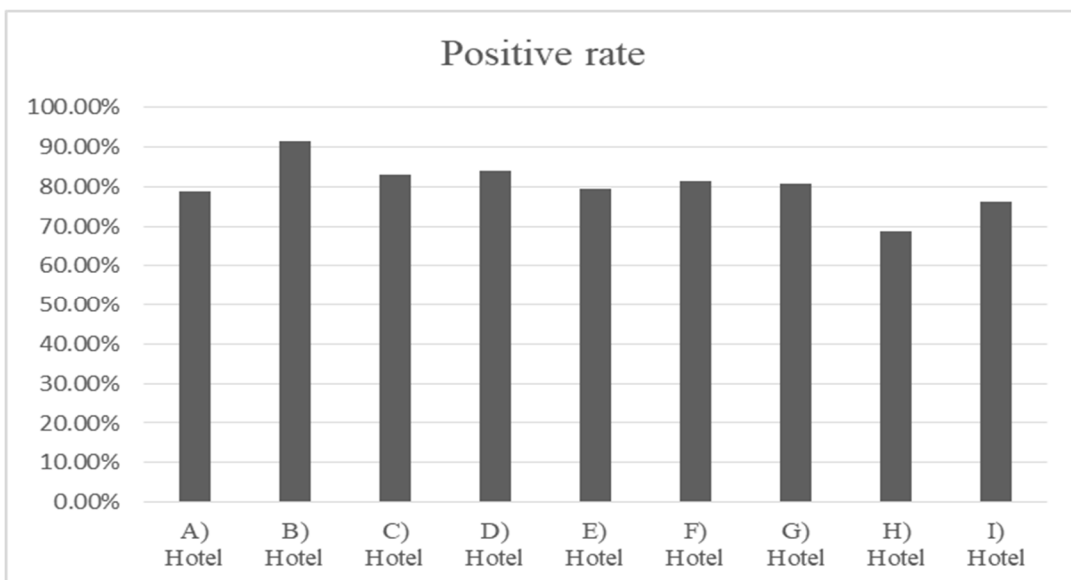


Figure 1 : Emotion Analysis Results

4 Results

4.1 Features of Chinese visitor's interest in a hotel

The result of the appearance frequency after normalization for the extracted words is shown in Table 7. The number "0" in the table means the word was not extracted. We found several features of Chinese visitors to Hamamatsu city in this analysis. Chinese visitor's concerns about the hotel were "room (0.68)," "location (0.34)," and "breakfast (0.3)." It is clarified that the room is the most expected factor. The next point is for the visitors who stayed at Hotel G) and H). It is important for them to experience "Hot spring." "Hot spring" attracts certain foreign visitors. For Hotel B), the emotion analysis gave the highest score (91.36%) in nine hotels, the "employee" could be an important factor.

Table 7: Appearance frequency after normalization

Hotel	Breakfast	Room	Price	Station	Location	Facility	Em- ployee	Hot Spring
A)	0.43	1.00	0.14	1.00	0.43	0	0	0
B)	0.27	0.91	0.36	0	0	0	1.00	0
C)	0.10	0.50	0	0	0	1.00	0	0
D)	0.50	0.42	0	0	1.00	0.25	0	0
E)	0	0.29	0	1.00	0.29	0	0	0
F)	0.40	1.00	0	0.40	0.40	0	0	0
G)	0	0.60	0	0	0	0.38	0	1.00
H)	0	0.67	0	0	0	0	0	1.00
I)	1.00	0.70	0	0.10	0.90	0	0	0
Ave	0.30	0.68	0.06	0.28	0.34	0.18	0.11	0.22

4.2 Comparison of interests between Chinese and Taiwanese for hotel

The analysis results of the Chinese and Taiwanese about Hotel A) are shown in Table 8. The first point is emotional analysis of positive rate and the evaluation score. It seems that Taiwanese slightly strictly evaluated the hotel comparing with Chinese. The Positive ratio of Taiwanese is 2.31% less than that of Chinese. The average evaluation score of Taiwanese is 0.2 point less than that of Chinese, too.

Taiwanese and Chinese have many common characteristics. They are interested in the same factors, such as "room," "station," "location" etc. However, although Chinese were interested in "price," it is not much important concern by Taiwanese. They are interested in "restaurants," "parking lots" etc. It is estimated this tendency that since Taiwanese visitors would come by private tours, they have to make an arrangement for their parking lot and restaurant by themselves. As Chinese visitors would come by group tours, all necessary accommodations are arranged by tour organizers. It is an interesting finding to understand the difference between Chinese and Taiwanese.

In the analysis of "Adjectives," the common words extracted were "convenient" and "clean." Furthermore, the Chinese had a nondescript impression of "good," while Taiwanese evaluated more specifically "comfortable" and "kind" for services and hospitality.

Table 8: Comparison Chinese and Taiwanese for Hotel A)

Nationality	Chinese	Taiwanese
Emotional analysis "Positive"	78.92%	76.61%
Evaluation score average	4.73	4.53
Score distribution	No. of data:35 5point:22 4-5point:10 3-4point:3	No. of data:18 5point:5 4-5point:11 3-4point:2
Appearance frequency		

Nouns	room:(1.00) station:(1.00) location:(0.43) price:(0.14) breakfast:(0.43)	room:(1.00) station:(0.57) location:(0.57) restaurant:(0.43) breakfast:(0.14) parking lots:(0.14)
Verbs	service:(1.00) recommend:(1.00)	service:(1.00) offer:(0.50) shopping:(0) examine:(0)
Adjective	good:(0.71) convenient:(1.00) clean:(0.29) optimal:(0)	comfortable:(1.00) kind:(0.60) clean:(0.60) convenient:(1.00)

4.3 Comparison between Chinese/Taiwanese/Hong Kong and Other nationalities for sightseeing spots

Table 9 shows the emotion analysis results for sightseeing spots by nationality. First, it is clear that East Asian (including Chinese, Taiwanese, and Hong Kong) scores were always higher than other nationalities. As a further feature, there are no major differences between East Asian people and the other nationality for the three parks including “Hamamastu Castle,” while about 10% difference has been observed for museums, beaches, and temples. Knowing the difference in feelings to sightseeing spots due to the difference in nationality is an indication of the design of appropriate tourist route planning.

Table 9: Emotion analysis results for sightseeing spots by nationality

Place	Chinese, Taiwanese, Hong Kong	Other nationalities (Australia, India, Singapore, Canada, etc.)	Δ
Hamamatsu Air Park	76.19%	75.23%	0.96%
Hamamatsu Castle	77.52%	75.15%	2.37%
Hamamatsu Flower Park	85.90%	82.22%	3.68%
Hamamatsu Museum of Musical Instruments	76.14%	66.10%	10.04%
Nakatajima Beach	80.65%	71.26%	9.39%
Ryotan-ji Temple	88.81%	78.24%	10.57%

5 Discussion

5.1 Insight of result of the analysis

This study revealed that the interest and expectation for hotels and sightseeing spots are not always the same by nationality. As shown in Table 7, "room," "location," and "breakfast" are important for Chinese tourists. Therefore, in order to increase the satisfaction level of Chinese visitors, it is considered firstly the easiest way to raise the quality of "breakfast."

Also, it may be difficult to change the "location" of the hotel, but it is possible to renovate the "room." Furthermore, there is a possibility that improvement of transportation to the "station" and improvement of the service of "employee" could increase the evaluation from Chinese customers. According to the comparison between Chinese and Taiwanese tourists, in Table 8, the Taiwanese evaluation score was much severe than the Chinese. It seems that quality of "Room" and "Services" for both Chinese and Taiwanese are the most important things. They are expecting good quality levels of them. In addition, it has to be noted that Chinese tourists are "price" conscious, and Taiwanese tourists may want to be supported for their private trips to Japan. The different characteristics of Chinese tourists and other nationality's ones as mentioned above are the analysis results from the review for Hamamatsu City of the Internet Web site. Therefore, the same result is not always obtained in other cities in Japan. However, by collecting and analyzing such reviews on the Internet, it is possible to understand the awareness of foreign tourists in local cities as well and to clarify appropriate marketing approaches for the targeted foreign tourists.

5.2 Profitable Strategy of Local Cities

If a Chinese tourist would be a target customer, it is most effective to provide a clean room and good breakfast at reasonable prices. It is also important for the hotel to provide Taiwanese tourists with comfortable service and space for rooms and restaurants. In addition, various information related to travel may be pleased. In terms of sightseeing spots, tourists from China and Taiwan are interested in visiting temples and shrines. Also, parks are expected sightseeing spots for guests of all nationalities. Proposals for services that take into account the characteristics according to nationality may produce high profits.

Continuity is the most important in business, and it could build a stable business base by increasing customer loyalty and repeat rate. For customers, it is important to enhance the quality of service and provide it at reasonable prices.

6 Conclusion

In this paper, we collected and analyzed the review data of foreign travelers visiting Japan as a basic knowledge to attract foreign visitors to local cities and showed the results of the traveler's emotional analysis. Reviews on Hamamatsu City in Shizuoka prefecture, a regional city in Japan, from travel sites on the Internet by using Web API tools. Target data were in the categories of "hotel" and "tourist spots" in multilingual in Chinese and English. And the word appearance frequency and the result of emotion analysis were obtained from review data by using NLPiR which is a morphological analysis tool. Then, knowledge discovery was conducted from the analysis result. As a result, the characteristics and differences of foreign travelers visiting local cities by nationality are clarified. By the method used in this study, local cities can increase the marketing ability to attract foreign tourists. As a result, it may lead to the revitalization of the local areas.

In this research, since it is limited to the data of Hamamatsu City in Shizuoka prefecture, we would continue to collect and accumulate data continuously and increase the accuracy of the analysis.

References

- [1] Ministry of Land, Infrastructure Transport and Tourism, “White paper of Tourism in 2017”, 2018
- [2] Y.Sugitani, “Effectiveness of Review on the Internet: Effects of Nonverbal Clues in Product Evaluation,” *Sophia University Economic Theory*, pp. 47-58, 2009.
- [3] T. Morita, A. Izawa, A. Nagateyama, “Image analysis of cities by text mining using free description data,” *Proceedings of JSCE, D3, Vol.68, No. 5*, pp. 315-323, 2012.
- [4] A. Takasaka, “The way of inbound strategy making use of regional characteristics,” *JRI I Journal Reviews*, Vol. 7, No. 37, 2016J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, pp. 68–73,1892.
- [5] M. Schuckert, X. Liu and R. Law, “Hospitality and Tourism Online Reviews: Recent Trends and Future Direction,” *Journal of Travel & Tourism Marketing*, Vol. 32, No.5, pp. 608-621, 2015.
- [6] X. Xu and X. Li, “The Antecedents of Customer Satisfaction and Dissatisfaction toward Various Types of Hotels: A Text Mining Approach,” *International Journal of Hospitality Management*, Vol. 55, pp. 57-69, 2016.
- [7] YH. Hsiao, MC. Chen, and MK. Lin, “Kansei Engineering with Online Review Mining for Hotel Service Development, “*Proceedings of 6th IIAI International Congress on Advanced Applied Informatics*,” pp. 29-34, 2017.
- [8] R. Ohkubo, “A Study of Destination Images of Foreign Tourists to Japan by Analyzing Travel Guidebook and Review Site,” *Journal of the City Planning Institute of Japan*, Vol. 49, No. 3, pp. 573-578, 2014.
- [9] W. Tanabe, “A Study on User Review Analysis to Support Planning Strategies of Accommodation,” *Tokyo City University, Journal of the center for information studies*, Vol. 9, pp. 91-101, 2008.
- [10] TripAdvisor, <https://www.tripadvisor.com>, Sept. 2017.
- [11] Expedia, <https://www.expedia.com/>, Sept. 2017.
- [12] Agoda, <https://www.agoda.com/>, Sept. 2017.
- [13] Booking.com, <https://www.booking.com/>, Sept. 2017.
- [14] Ctrip, <https://www.ctrip.com/>, Sept. 2017.
- [15] GooSeeker, <http://www.gooseeker.com>, Sept. 2017.
- [16] Octoparse, <http://www.bazhuayu.com/>, Sept. 2017.
- [17] NLPIR, <http://ictclas.nlpir.org/>, Sept. 2017.