Quantitative ranking of satisfaction and dissatisfaction factors based on text mining of online reviews from Chinese and English speaking tourists visiting Japan

Elisa Claire Alemán Carreón[a,*], Hirofumi Nonaka[a], Toru Hiraoka[b]

[a]*Nagaoka University of Technology, Nagaoka, Japan*
[b]*University of Nagasaki, Nagasaki, Japan*

Abstract

Recently, there has been a boom in international tourism in Japan, where particularly Chinese tourists are increasing rapidly, among other international tourists. Because of its effect on the Japanese economy, there is a need for the use of newer research tools that are cost and time effective and use online reviews, which increasingly affect sales and the influx of customers. Using machine learning, we determined keywords representing satisfaction and dissatisfaction factors from Chinese and English online reviews of Japanese hotels in the portal sites *Ctrip* and *TripAdvisor*. We identified that Chinese customers react positively to included breakfasts and big rooms, and are unsatisfied with the pricing of the hotel. We also found English speaking customers dislike pricey hotels and have a high dislike for dirty rooms, stains, and cigarette smell, and that they mainly focus positively on staff, cleanliness, and transportation availability.

Highlights

- *Chinese customers tend to prefer included breakfasts, and rooms that are big and clean.*

- *Across all hotel prices, unsatisfied Chinese customers focus on the pricing of the hotel.*

- *English speaking customers are satisfied with the staff, cleanliness, and transportation availability.*

*Corresponding author
Email addresses:* s153400@stn.nagaokaut.ac.jp (Elisa Claire Alemán Carreón), nonaka@kjs.nagaokaut.ac.jp (Hirofumi Nonaka), hiraoka@sun.ac.jp (Toru Hiraoka)

- *English speaking customers dislike pricey hotels and have a high dislike for dirty rooms and cigarette smell.*

1. Introduction

The population of Chinese tourists has recently increased significantly worldwide over the years. This increase has led to a significant impact on several industries over the world, most directly the hotel industry, as their customer bases changed. In addition to this impact on business, there is a current increase in academic research across the world about Chinese tourist populations (Sun et al., 2017). Particularly in Japan, however, international tourism has been increasingly exerting a stronger influence in the economy (Jones et al., 2009). There were notable increases in the Chinese tourist population in Japan of 107.3% from 2014 to 2015, and a total number of tourists of 6,372,948 in 2016 (Japan National Tourism Organization, 2017). This increase in international tourism creates a need for cross-language, cost and time effective market research tools. In the hospitality industry, where language barriers are a constant problem, there are difficulties and costs when attempting to understand the needs of their customer base.

Until recent years, most of these market studies are performed based on the results of surveys or interviews. Currently, however, there are data mining and text mining techniques which potentially have the advantage over surveys and interviews. During recent years, several studies have successfully proved the validity and potential of using text and data mining analysis in the business management field. A study classified different kinds of Social Media posts and customer response from several U.S. pizza selling companies (He et al., 2013). Another study ranked products by using online reviews and sentiment analysis (Liu et al., 2017), while another used them to predict product sales (Fan et al., 2017). Other uses for text mining techniques include patent document analysis,

as was done in a study on patent scoring (Nonaka et al., 2014), and another on the extraction of technological terminology in patents (Nonaka et al., 2012). It has also been used in the tourism and hospitality field to predict hotel demand by using web traffic data (Yang et al., 2014) or to make recommender systems (Loh et al., 2003). Data mining and text mining techniques can not only cheaply and quickly gather tens of thousands if not more samples, but the source of this extensive amount of information also can be thought to be unaltered by human intervention in any part of the data extraction process.

An important benefit of using data and text mining in the hospitality field is that because of the current behavior of the customers in the information age, where the searching, choosing and booking process involves reading online reviews, and where customers write online reviews depending on their satisfaction or dissatisfaction. Previous studies have found that there is a correlation between prominent reviews and sales figures (Basuroy et al., 2003; Ye et al., 2009). There are also studies showing the direct impact of online hotel reviews on consumer consideration, that is, the decision process of many customers is affected either positively or negatively because of the reviews, with a stronger impact for lesser known hotels (Vermeulen & Seegers, 2009). Online reviews give researchers access to large databases of unfiltered customers opinions without the need to ask directly via surveys or interviews, and they also provide availability of data from tourists writing in many languages. It is because of the important impact the hotel reviews have on consumers, and because of the highly polarized emotional content these reviews hold that we have decided to use this data for analysis.

Many researchers have focused on studying the level of satisfaction of customers for predefined factors (e.g. Balbi et al., 2018; Kim et al., 2017; Truong & King, 2009; Wu & Liang, 2009; Shanka & Taylor, 2004; Kozak, 2002), while others focus on motivation, choice or behavior (e.g. Romão et al., 2014; Dongyang et al., 2015). However, most of the studies rely on making inferences about the larger population based on predefined factors, instead of extracting the factors directly from the data without human intervention. Not only that but when

3

examining international tourism, they are limited by language barriers. In our study, we aim to use the benefits of online reviews to mathematically extract the answer to these questions in a manner that doesn't rely on understanding the language to expand it further to a larger corpus. In addition, while most studies with text mining techniques and sentiment analysis techniques focus on either relevant topics or the sentiment value of reviews, this study merges both techniques to find the relevant topics and their ties to each emotional response.

In our study, we extracted satisfaction and dissatisfaction keywords and analyzed the needs of Chinese and English speaking customers of Japanese Hotels. We chose this combination of population and destination because we aim to apply machine learning techniques to overcome language barriers on a large scale. The extracted keywords are the key to researching which topics and in what order of priority the Chinese tourist and English speaking tourist populations consider when they write their reviews. This is, of course, valuable information for making business management decisions from the point of view of the hotel industry companies. Knowing which topics are not only largely discussed, but those that are most related to customer satisfaction and dissatisfaction, a company can improve customer service or facilities to increase profit. Using the methodology detailed in this study, such knowledge can become available to other researchers and companies by applying it in different environments.

2. Research objective

The objective of this study is to determine the factors influencing the satisfaction or dissatisfaction of Chinese customers as well as studying English speaking customers of Japanese Hotels because of the language barrier that can be studied and the economic weight that these customers represent. These satisfaction and dissatisfaction factors can become the focal point for making improvements in tourism and service industries, increase the satisfaction of customers, and influence newer customers to write more satisfied online reviews that will in turn increase sales and attract new customers.

We propose to avoid language barrier limitations by using a process that is based on probability distributions and automatic algorithms, and that is not dependent on language-specific methodologies that limit its application in other environments. To do this, we propose to use a supervised machine learning methodology to be trained on the emotional classification of a sample online review texts and use it not only to classify other texts further but to analyze the weight towards classification that certain keywords hold in that trained machine. Furthermore, we also analyze the relevance of these keywords in different hotel price brackets to observe potential differences.

3. Related work

In the hospitality field, satisfaction, motivation, and behavior have been studied for a long time using surveys (e.g. Truong & King, 2009; Wu & Liang, 2009; Shanka & Taylor, 2004; Kozak, 2002; Romão et al., 2014; Dongyang et al., 2015; Chan et al., 2015). More recently, studies have used big data in the hospitality industry (e.g. Yang et al., 2014; Loh et al., 2003), as well as data mining tools to analyze hotel reviews (e.g. Al-Smadi et al., 2018; Browning et al., 2013; Xiang & Gretzel, 2010). However, most of these studies focus on either the sentiment classification without analyzing the relevant words tied to those sentiments, or they study the most relevant words without taking the emotions they largely represent into account. Other studies focus on review helpfulness (Ren & Hong, 2019); while others focus on summarization of reviews into smaller text (e.g. Hu et al., 2017; Amplayo & Song, 2017). There has not been yet a study that focuses on the conceptual words that are tied with the satisfaction or dissatisfaction of the userbase. For example, one study used a tool called "Word Cloud" to extract frequently used words inside the reviews which were considered important topics of concern to consumers (Hargreaves, 2015). However, the sentiment analysis is done at a hotel level, instead of a word level, so it is difficult to determine which topics are perceived positively or negatively. Another study used a heuristic frequency analysis to determine guest experience

topics (Xiang et al., 2015). However, this study doesn't deterministically show which topics are directly tied with satisfaction and dissatisfaction or the ranking order in which they do.

On the other hand, Zhang et al. (2011) performed a market analysis based on the sentiment analysis of product reviews using a Chinese sentiment dictionary called HowNet. The problem with using HowNet as a sentiment dictionary is that the words are classified based on ontology (Huang et al., 2008), where emotional words are usually those that describe emotions, such as "angry", "anger", "happiness", "content". While these words represent the emotions themselves, in an online review, it is unlikely that the general population writes their emotions explicitly; instead, it is more likely that words that do not describe the emotion but are intrinsically emotionally fueled are used. For example, one would not write "I am happy with the size of the room and the service"; instead what would be natural is "The room had lots of space for our luggage and the staff was friendly". HowNet is not able to determine that "space" is tied to positive emotions and that "staff" is being praised by the use of "friendly" in the sentence. These non-emotional words are directly representing the emotion of satisfaction or dissatisfaction in the online review. Now, in turn, this study methodology focuses on the influence of each word and their rank of importance in relation to the satisfaction or dissatisfaction expressed by the user, clearly defining not only their emotion but also the specific needs of Chinese customers in a concrete way.

There have been previous attempts at understanding the factors leading to satisfaction and dissatisfaction with text and data mining techniques as well. One study used a text-link analysis to understand which pairs of words were used more in positive or negative reviews (Berezina et al., 2016). However, the conclusions of this study are limited to their sample of manually classified 20 reviews, so it is difficult to extrapolate it to a larger population of reviews. Another study manually extracted factors that could lead to satisfaction or dissatisfaction from a sample and then continued to study the statistical differences between them in that sample (Zhou et al., 2014). However, this is limited

to their manually defined factors and that there are no further analyses to a larger number of online reviews because of language barriers. Another study used heuristic language-specific parsing algorithms to determine the frequency of words in a previously classified corpus (Xu & Li, 2016). The problem with this last method is that it is depending on the specific corpus of reviews classified as negative or positive from the moment they're input into the site into different input forms from *Booking.com* and that it is strictly only applicable to English texts. Our methodology not only applies the findings of the sampling to the full corpus, but it does in a way that overcomes the difficulties of language barriers and sentiment classification.

4. Methodology

We have extracted a large number of text reviews from a Chinese portal site *Ctrip*[1], as well as the travel site *TripAdvisor*[2] and determined the most commonly used words that would contribute the most to satisfaction and dissatisfaction opinions in a review using an entropy-based mathematical extraction method. These keywords extracted using entropy related to satisfaction or dissatisfaction not only allow us to perform a Support Vector Machine based emotional classification of the reviews but conceptual words in these lists bring insight into which concrete topics are the Chinese tourists concerned with. After classifying the sentences in the extracted reviews as signaling satisfaction and dissatisfaction with an optimized SVM, we have analyzed the weight value assigned to them by the SVM applying the methodology in previous research. We also observed the frequency of the terms in all of the reviews to extract the most utilized words in either kind of reviews. We show an overview of this methodology in Figure 1 (Alemán Carreón et al., 2018). Furthermore, we have also performed analyses for different price brackets to observe the change in demand for different kinds of hotels varying in luxury in the Chinese reviews.

---

[1]Ctrip: `www.ctrip.com/`

[2]TripAdvisor: `www.tripadvisor.com/`

4.1. Data collection

In this study we used the HTML parsing python library *BeautifulSoup*[3], and the local database management tool *SQLite*[4] using the python library *sqlite3* to automatize data managing processes.

4.2. Data processing

4.2.1. Text processing

Different from the English language, Chinese texts lack a separation between different words, and as such, when collecting these texts, they are all a single string of characters. To be able to perform a statistical analysis of each word, the Stanford Word Segmenter (Chang et al., 2008) program developed by the Stanford NLP Group[5] was implemented for this task using the python *nltk* library. During the segmentation of all the words in our corpus, irregularities occurred where the reviews were written in other languages or where unusual punctuation marks were used. We designated a list of characters that could be recognized as noise, such as punctuation marks, then cleared all the text in the corpus of these characters.
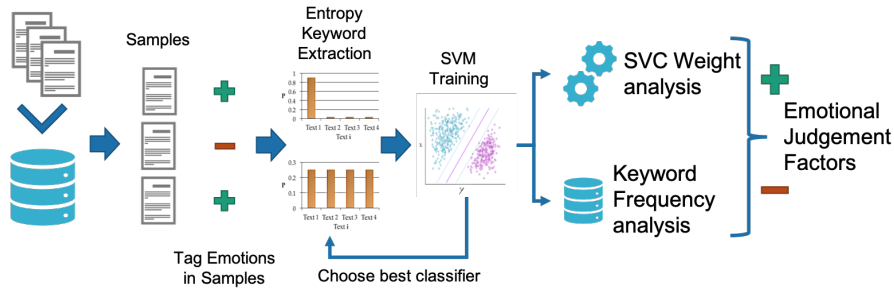

Figure 1: Overview of the methodology.

---

[3]Beautiful Soup: `https://www.crummy.com/software/BeautifulSoup/`

[4]SQLite: A local database management system with an SQL structure: `https://www.sqlite.org/`

[5]The Natural Language Processing Group at Stanford University

In the case of texts in English, however, only using spaces is not enough to correctly collect concepts. Because of variations and conjugations of words depending on the context and tense, a better segmentation is achieved by using lemmatization, which returns the dictionary form of each word. For this purpose, we used the *gensim* library with the English texts.

4.2.2. Sentiment analysis

4.2.2.1. Entropy based keyword extraction

In order to determine the words that clearly represent the user's satisfaction or dissatisfaction, we calculated the entropy value of each word in relation to each class. Shannon's Entropy, in the field of Information Theory, is defined to be the expected value of the information content in a signal (Shannon, 1948) or it can be thought of as the grade of impossibility of predicting an outcome. It is shown in formulas 1 and 2. Using this value we can observe the probability distribution of each word inside the corpus. A word that is included in many documents, it will have a high entropy value for that set of documents, since it becomes uncertain to predict in which document it will appear. Opposite to this, a word appearing in only one document will have an entropy value of zero, since it is completely predictable. We show this concept in Figure 2.

$$H = -\sum_{i=1}^{M}[P\log_2 P] \tag{1}$$

$$\lim_{P \to 0+} P\log_2 P = 0 \tag{2}$$

To apply this logic, we retrieved 50 reviews as a sample of our corpus and with the collaboration of a group of 5 Chinese students, tagged each sentence of a total of 159 as the classes satisfied or dissatisfied depending on the emotion that the text conveyed, then calculated the entropy values for each word in relation to the set of sentences from each class. In the case of English reviews, we sampled 665 reviews and with the collaboration of English speaking students manually tagged them by sentence, resulting in 2357 tagged sentences. Words

with higher entropy relating to the satisfaction set than to the dissatisfaction set by a factor of α were determined to be keywords tied with satisfaction in Chinese reviews of hotels. This is shown in formula 3. Likewise, words with higher entropy for the dissatisfaction set than the satisfaction set by a factor of α' were determined to be keywords tied to dissatisfaction in our texts. This is shown in formula 4.

$$H_P > \alpha H_N \tag{3}$$

$$H_N > \alpha' H_P \tag{4}$$

4.2.2.2. Support Vector Machine classification

Support Vector Machines are supervised machine learning models usually applied to classification or regression problems (Cortes & Vapnik, 1995). We use it to classify the rest of our corpus into emotional classes in this study. An SVM is trained to classify data based on previously labeled data, generalizing features of the data by defining a separating (p-1)-dimensional hyperplane in a p-dimensional space in which each dimension is a feature of the data. The separating hyperplane, along with the support vectors, divides the multi-dimensional space and minimize the error of classification. We show a two dimensional example in Figure 3.
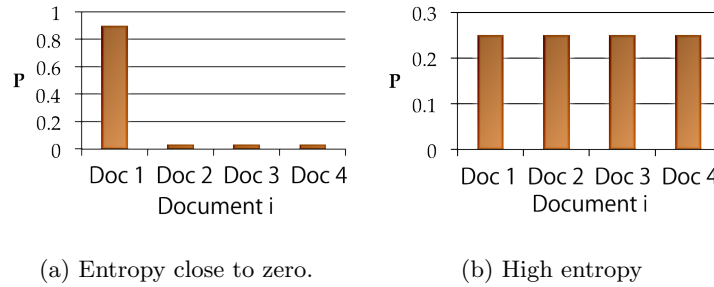

(a) Entropy close to zero.   (b) High entropy

Figure 2: Probabilities of a word j being contained in a document i.

In this study, we used the SVM classification process described in (Alemán Carreón et al., 2018) with a number of keyword lists, defined by our entropy calculations with different comparison coefficients, as the possible features; trained the SVMs and tested for each one using a K-fold Cross-Validation method (Kohavi, 1995). In each test we calculated the Precision, Recall and F_1-measure (Powers, 2011) for our predictions.

4.2.2.3. Model evaluation metrics

In order to measure the effectiveness of the training process and data, we performed what is called K-fold cross-validation. This means that after randomly shuffling and splitting the training data into k equal parts, k-1 of those parts are used for training, while the remaining one part is used in validation. Using the trained models, a prediction is made, and it is decided if such a prediction is correct or not, and counted and grouped as a True Positive, True Negative, False Positive or False Negative prediction. This is explained in Table 1.

Measures of accuracy are determined from these prediction outcomes. This process is then repeated k times and the measures taken are averaged. In this study, we used the F_1 score, which measure is a harmonic mean between precision and recall. Precision, described in formula (5), lets us observe the rate of correct positive predictions from all the positive predictions, while Recall, detailed in formula (6), observes the rate of correct positive predictions from the total of actual positive data. The F_1 score in formula (7) then can only be high when both of these measures are high simultaneously and will lower

Table 1: Prediction outcomes.

	Prediction is Correct	Prediction is Incorrect
Prediction is Positive	True Positive	False Positive
Prediction is Negative	True Negative	False Negative

11

substantially if they are not consistent. We use this score as it allows us to avoid overlooking data while maintaining accurate predictions. The overall accuracy (8), which doesn't account for overlooking of data, is also observed for reference but isn't used as the main evaluation of the model.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{5}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{6}$$

$$F_1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{7}$$

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Number\ Of\ Predictions} \tag{8}$$

4.2.3. Data analysis

4.2.3.1. SVM weight analysis

During the SVC learning algorithm, each point of data that is classified incorrectly causes a change in the weight vector to better classify new data correctly. These changes to the weight vector are strong for features that needed to be taken account of to classify with a minimal error, those contained in the support vectors, close to the separating hyperplane. Sequentially, the weight vector can be interpreted as a numerical representation of the effect each feature had for each class in the classification process. Because the weight vector gives value to the words that comprise the support vectors, the words with higher weights are thought to be closer to the dividing hyperplane, while still clearly and decisively belonging to one of the categories the hyperplane divides the high-dimensional space in. While other less ambiguously used words could be excluded from this weight assigning process, the words that are assigned weight values are thought to be words that while discussed in both emotional states, they are more strongly discussed in one of these emotional states. If anything, these are highly volatile words relating to the experience of the customers in the

hotel, which are important to consider as a hotel business to improve customer satisfaction. Below we show the formula for the weight vector (9).

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{9}$$

4.2.3.2. Keyword frequency analysis

Because the machine learning model used in this study is constructed from training data obtained by a relatively small sample of our data, after training the model and classifying the rest of the data, we studied the frequencies of each entropy based keyword in our complete data set. Words signifying a subject or topic that are relatively highly used will have more meaning when analyzing the emotional response of the user-base of Chinese tourists in Japan.

5. Results

5.1. Data set

In the data collection stage, a total of 1,541,424 HTML files were crawled, from which 5,938 were unique review pages of hotels in Japan. From these pages, we extracted a total of 44,912 reviews, which were comprised of 286,109 separate sentences. In our corpus, there were 23,443 different words used, from which 2802 were noise characters.

5.2. Sentiment analysis results

After having the training data tagged by a group of 5 Chinese student collaborators, we experimented with different comparison coefficients for the entropy values calculated from the satisfaction and dissatisfaction emotional classes. The mutually independent coefficients α and α' were tested from 1.25 to 6 in intervals of 0.25. The result was 40 lists, 20 for each emotional class. We repeated the process for the reviews in English that our colleagues tagged.

In the beginning, we experimented with different kernels for the SVM, as well as some Ensemble Learning methods, like the Boosting, Voting and Stacking.

We ultimately decided to use the linear kernel for the benefits of the weight vector obtainability. We also experimented with different parameters for the SVC, finding that the best performing value for C, a constant that affects the optimization process when minimizing the error of the separating hyperplane. Low values of C give some freedom of error, which minimizes false positives, but depending on the data it can increase false negatives. Inversely, high values of C will likely result in minimal false negatives, but a possibility of false positives. We found the ideal value for this parameter was C=0.5 in our final classifier for Chinese text, and 2 in the English text one.

We trained a different Support Vector Classifier with each of the lists, and we chose the best performing lists for each emotional class, resulting in a satisfaction emotion classifier (satisfied or not) and a dissatisfaction emotion classifier (dissatisfied or not), based on the results of a 5-fold cross-validation process in the Chinese reviews, and a 10-fold cross-validation process in the English reviews case, in which we calculated their accuracy and F-measure means and standard deviations. The number of k-folds was decided from the sample size. Since the sample for the Chinese texts was smaller, the number of validation data would be reduced if a higher value for k was used. After observing the misclassification behavior for the satisfaction emotion classifier, which mostly misclassified sentences with negative words, we decided to combine both keyword lists into a single large list to train the satisfaction emotion classifier. With this Combined list, the accuracy was of 0.92 ± 0.03 and $F_1 = 0.95 \pm 0.01$, both excellent results for classification. Table 2 shows the lists that had the best performance results in the case of Chinese text and Table 3 shows the best performance results in English texts.

Using both sets of keywords in the same SVC (named 'Combined') to classify the testing data, we obtained higher performance results overall. We then used these classifiers in the rest of the respective data to label positive emotion sentences. The sentences not belonging to the positive emotion class were considered to belong in the negative emotion class since the classification of the emotion of each sentence was binary in our training samples original classifica-
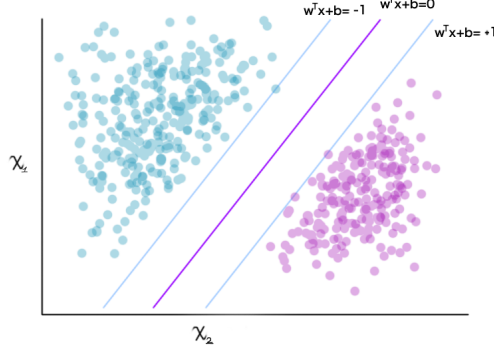
14

Figure 3: Two dimensional example of an SVM classification problem.

Table 2: Results of the best 5-fold cross-validation Chinese text classification performance tests.

Keyword List	Classifier emotion	C	Accuracy μ	Accuracy σ	F_1 μ	F_1 σ
Satisfaction keywords ($\alpha = 2.75$)	Satisfaction	2.5	0.87	0.02	0.91	0.01
Negative keywords ($\alpha' = 3.75$)	Dissatisfaction	0.5	0.85	0.05	0.67	0.11
Combined (α=2.75, α'=3.75)	**Satisfaction**	**0.5**	**0.92**	**0.03**	**0.95**	**0.01**

Table 3: Results of the best 10-fold cross-validation English text classification performance tests.

Keyword List	Classifier emotion	C	Accuracy μ	Accuracy σ	F_1 μ	F_1 σ
Satisfaction keywords ($\alpha = 1.5$)	Satisfaction	1.75	0.79	0.02	0.82	0.02
Dissatisfaction keywords ($\alpha' = 4.25$)	Dissatisfaction	3	0.70	0.04	0.80	0.03
Combined (α=1.5, α'=4.25)	**Satisfaction**	**2**	**0.81**	**0.02**	**0.83**	**0.02**

tion.

5.3. Keywords and their SVM weight values

In Table 4 we show some of our keywords that signify subjects or topics, and as such, help us understand the needs of the users. We show keywords that have a relatively high weight value for both positive and negative extremes, and their translations in the relevant context. In Table 5 we show keywords for the English classifier with high weight values.

5.4. Keyword frequencies

We observed the words with the highest frequencies for keywords in satisfied and dissatisfied statements to study and quantitatively rank the needs of Chinese customers, as shown in the Tables 6 and 7; and for English speaking customers, as shown in Tables 8 and 9.

In order to further analyze the data we obtained, aside from emotional classification, we also investigated the difference in emotional response from customers of hotels in different price brackets in the case of Chinese reviews. We separated hotel prices in 6 categories, and in Table 10 we show the words that have an interesting change in their ranking order when ordered by frequency in each price bracket. We include the price bracket definitions in Appendix A in Table A.11.

6. Discussion

6.1. Needs of the Chinese tourist market in Japan

Analyzing the results from the extracted keywords is the key to understanding Chinese and English speaking customers of Japanese hotels since they can be interpreted as rankings of their explicit needs and demands. Analyzing satisfaction keywords, we found that the most relevant subjects Chinese customers perceive positively are cleanliness and size, very possibly of the room they had stayed in. There is also the possibility that reviewers were praising, in general,

16

Table 4: High weight values for Chinese keywords signifying subjects or topics.

Word	Translation	Entropy List	SVC Weight
地方	region, local	Satisfaction	1.343
干净	clean	Satisfaction	0.638
大	big, wide	Satisfaction	0.624
交通	traffic, transportation	Satisfaction	0.586
热情	cordial, kindness	Satisfaction	0.495
周边	periphery	Satisfaction	0.495
景色	scenery	Satisfaction	0.495
推荐	recommendation	Satisfaction	0.495
日本	Japan	Satisfaction	0.495
早餐	breakfast	Satisfaction	0.495
附近	nearby	Satisfaction	0.495
中文	Chinese text	Dissatisfaction	-0.714
地理	geography	Dissatisfaction	-0.812
价格	price	Dissatisfaction	-1.505

Table 5: High weight values for English keywords signifying subjects or topics.

Word	Entropy List	SVC Weight
bathhouse	Satisfaction	2.000
museum	Satisfaction	2.000
meeting	Satisfaction	1.997
subway	Satisfaction	1.951
cozy	Satisfaction	2.000
convenience	Satisfaction	1.888
clean	Satisfaction	1.886
comfortable	Satisfaction	1.724
dirty	Dissatisfaction	-1.275
policy	Dissatisfaction	-1.463
prepay	Dissatisfaction	-1.517
pricey	Dissatisfaction	-1.614
sticky	Dissatisfaction	-2.000

Table 6: High weight Chinese keywords in satisfaction sentences.

Word	Translation	Frequency	SVC Weight
大	big	15470	0.624
干净	clean	12166	0.638
早餐	breakfast	10575	0.495
推荐	recommendation	8752	0.495
环境	environment, surroundings	8694	0.248

Table 7: High negative weight Chinese keywords in dissatisfaction sentences.

Word	Translation	Frequency	SVC Weight
价格	price	7636	-1.505
地理	geography	2238	-0.812
中文	Chinese language	1410	-0.714

Table 8: High weight English keywords in satisfaction sentences.

Word	Frequency	SVC Weight
staff	138677	0.537
clean	105971	1.886
location	103151	0.842
comfortable	62793	1.724
subway	38354	1.951
airport	36328	0.839
shopping	34355	1.308
onsen [*]	31084	0.094

[*] Onsen are Japanese hot springs

Table 9: High negative weight English keywords in dissatisfaction sentences.

Word	Frequency	SVC Weight
pricey	3809	-1.614
carpet	3683	-0.507
dirty	2943	-1.275
stain	2787	-1.886
cigarette	2468	-0.435

the cleanliness of Japan's environment, un-littered streets, potable water and their culture of respecting spaces. Our results are compatible with Ryan & Mo (2001) in their study of Chinese tourist behavior in New Zealand , where Chinese travelers were found to prefer traveling and exploring unfamiliar places that have a reputation for being clean and unpolluted environments, focusing on scenic beauty, history and culture. They also found complaints about the lack of Chinese language availability in the services they consumed. Liu et al. (2019) describe that the behavior studied in Australia reflected that Chinese tourists were more likely to appreciate touristic places for the scenery, architecture, landmarks, and popular spots, rather than appreciate cultural experiences or activities. Similar to the findings in Liu et al., our results indicate that closeness of the hotel to scenic places is highly tied to satisfaction. However, their research is on online reviews of touristic places, instead of hotels. With this difference, our results showed that breakfast inclusion and the size and cleanliness of rooms were also tied to satisfaction, aside from scenery. According to our data, other positive factors that come into play when Chinese tourists choose a hotel to stay is the location in relation to public transport availability; and as mentioned before, environments, such as gardens or parks nearby; and services nearby, like places to go shopping. Gao et al. (2017) argue that Chinese people are heavily influenced by religion and a culture that regards harmony and oneness with nature and humans, which in turn influences their tourism behavior. However, they also argue that there are generational differences and that the older generations that lived through the changes that changed the environment and caused pollution care more about the environment, and as such appreciate natural and scenic touristic spots more than younger generations.

One key component we found in Chinese customer satisfaction is the inclusion of breakfast within the hotel. This can be inferred from the high frequency with which this keyword was included in the sentences emotionally classified as positive. While other food-related words were extracted, most of them were general in nature, like "food" or "eating", and in a lower ranking. In contrast, the word "breakfast", which is referring to a specific time and very possibly its

inclusion in the hotel commodities, was very frequently used in positive texts compared to other food-related words. This specific keyword is previously unmentioned in studies regarding the eating behavior of Chinese tourists. While these studies on eating behavior are very scarce, both a study on the food preferences of Chinese tourists in Australia (Chang et al., 2010) and an article on Chinese food culture (Ma, 2015) state that even in foreign countries, Chinese people have a habit to eat Chinese food, probably because of a need for familiarity. However, the analysis of our data points to the fact that more than Chinese food as a specific concept, hotel reviews are written with general words that represent food. We would like to further study specific Chinese tourists food preferences, such as the contents of said breakfast and other meals in future work to observe the level of necessity of these particular needs.

Now, when analyzing dissatisfied responses, the most frequently criticized aspect of Japanese hotels was the price. This was true across all of the price brackets, from the cheapest hotels to the most luxurious ones. In their study of Chinese tourists in Vietnam, Truong & King (2009) also found both the tendency to be concerned with value for money. It is possible that concerns over value for money are a general cultural trait in the Chinese population. Applying this previous finding to our data, negativity about the pricing of hotels across all price brackets can mean they believe the price is too high for the service that they receive, regardless of their satisfaction or dissatisfaction with the service. Our results suggest that Chinese customers are expecting lower prices for whatever services they receive.

Another important negative factor can be the availability or lack thereof of Chinese translations. Chinese customers can feel lost when they don't understand directions or instructions, either written or spoken; however, according to our data, most customers can be thought to have complained about written translations.

Additionally, as shown in Table 10, there is a positive reaction to "Japanese-style", which we infer in this context can mean rooms or possibly ceremonies, in luxury rate hotels. Another interesting point is that the expectation for hotel

staff increases as the price of the hotel rises. Furthermore, while "recommendation" of a hotel can be seen across all price brackets, customers searching for low priced hotels are relatively more inclined to recommend that hotel to others, based on a comparison of frequency ranking of each word for each price bracket.

6.2. Comparison with English speaking tourists in Japan

The results with the English speaking population are similar to the results with the Chinese speaking population, but there are some noteworthy differences that can be explored. For example, while Chinese customers seem to focus positively on the cleanliness of the room first and foremost, English speaking customers seem to consider the behavior of the staff before it. However, when we consider the negative associations, the price is the first concern for both kinds of customers, but there's a high concern for cleanliness in a negative connotation, such as complaints about stains, dirty rooms or carpets. A thing to note is that there is a high number of complaints about cigarette smell, in comparison to the Chinese customers in which it isn't a keyword within our extraction methods. Previous research states that 49 - 60% of Chinese men (and 2.0 - 2.8% of women) currently smoke or have smoked before, taken from a sample of 170,000 Chinese adults in 2013-2014, which is high compared to many English speaking countries (Zhang et al., 2019; World Health Organization, 2015).

In previous research, it has been found that Chinese tourists are harsher in their negative touristic attraction reviews than other nationalities (Liu et al., 2019). Taking this into account, and the large difference in frequency between complaints about price and other elements, Chinese customers can be said to be dissatisfied with price in a disproportionate way. English speaking tourists' dissatisfaction factors, on the other hand, have less of a gap in frequency, and as such can be said to be more concerned about cleanliness related factors than price.

On the positive reviews side, there is a high number of mentions for nearby transportation methods, such as the subway, or access from the airport. This is also different when compared to Chinese customers, who don't mention trans-

portation services as frequently for viewing a hotel in a positive way. Another difference that is noteworthy is that Chinese customers appreciate included breakfast in hotels, but for the English speaking customers it seems to not be a priority.

7. Conclusions and future work

In this study, with the purpose to understand emotional responses of Chinese customers of Japanese hotels and understand their needs, we extracted keywords from their reviews from the Chinese portal site *Ctrip* using entropy calculations from a manually classified sample of our data; then we used these keywords in machine learning experiments. Using the keywords to train a linear kernel Support Vector Classifier, we obtained the highest performance ($F = 0.95 \pm 0.01$ and $Accuracy = 0.92 \pm 0.03$) using both our positive and negative keywords for a binary emotional classifier.

Using the weight vectors of our classifiers, as well as frequency of the words in our dataset, we found that Chinese customers have a preference for big and clean rooms, big thermal baths or bathhouses, expect good value for money regardless of price, that there is a lack of Chinese text translation and that they prefer hotels where breakfast is included. We also found that the friendliness of the staff increases in priority as the price and luxury of the hotel increases and that transportation is more important for customers looking for cheaper options compared to the more luxurious ones.

For comparison, we also repeated the experiments on reviews written by English speaker tourists staying at Japanese hotels from *TripAdvisor*. Observing the results we identified that English speaking customers hold a bigger contempt and negative reactions for dirty or stained rooms, but that the main concern is also the pricing of the hotels. We also found that they prefer friendly staff and transportation available from the airport or subway.

Our results show the needs of a large and increasing segment of the customers. This would have otherwise presented difficulties with traditional meth-

ods because of language and culture barriers. The hotel industry can make use of these results to better the response from Chinese customers by focusing on improving topics they find important, such as cleanliness, the inclusion of breakfast, availability of Chinese translations in text format and so on. By doing this, satisfied users will write new reviews and thus influence future customers positively.

In future work we plan to investigate further into this topic, extending our data set and researching for different trends for different regions of Japan and in different kinds of hotels, as well as between customers traveling alone or in groups, for fun or for work. Another point for the future of this study is to use word clusters with similar meanings instead of single words. As mentioned before, we would also study the food preferences of Chinese tourists, as well as the impact of the recommendation in different price brackets. Additionally, it would be interesting to study further into more specific emotions than the satisfaction and dissatisfaction classifications we performed in our study.

Table 10: Words that change in importance between price brackets in positive sentences.

Word	Translation	Observation
交通	traffic, transportation	Higher ranking for hotels with lower prices, while not as high in luxury hotels.
环境	environment, surroundings	Higher ranking for luxury hotels and in the lowest priced hotels, but not as high in the middle price brackets.
日式	japanese-style	Only high ranking in luxury hotels.
人员	staff, personnel	Rises in ranking as the hotel price rises.
推荐	recommend	Included in all hotels, but higher ranking in lower priced hotels.

References

Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., & Qawasmeh, O. (2018). Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Information Processing & Management*, *56*. doi:`10.1016/j.ipm.2018.01.006`.

Alemán Carreón, E. C., Nonaka, H., Hiraoka, T., Kumano, M., Ito, T., & Hirota, M. (2018). Emotional contribution analysis of online reviews. In *Proceedings of The 2018 International Conference on Artificial Life and Robotics (ICAROB2018)* (pp. 359–362). Beppu, Japan volume 23. doi:`10.5954/ICAROB.2018.OS5-3`.

Amplayo, R. K., & Song, M. (2017). An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. *Data & Knowledge Engineering*, *110*, 54–67. doi:`10.1016/j.datak.2017.03.009`.

Balbi, S., Misuraca, M., & Scepi, G. (2018). Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management*, *54*, 674–685. doi:`10.1016/j.ipm.2018.04.009`.

Basuroy, S., Chatterjee, S., & Ravid, S. (2003). How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal Of Marketing*, *67*, 103–117. doi:`10.1509/jmkg.67.4.103.18692`.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, *25*, 1–24. doi:`10.1080/19368623.2015.983631`.

Browning, V., So, K. K. F., & Sparks, B. (2013). The influence of online reviews on consumers' attributions of service quality and control for service standards in hotels. *Journal of Travel & Tourism Marketing*, *30*, 23–40. doi:`10.1080/10548408.2013.750971`.

Chan, A., Hsu, C. H., & Baum, T. (2015). The impact of tour service performance on tourist satisfaction and behavioral intentions: A study of Chinese tourists in Hong Kong. *Journal of Travel & Tourism Marketing*, *32*, 18–33. doi:`10.1080/10548408.2014.986010`.

Chang, P., Galley, M., & Manning, C. (2008). Optimizing Chinese word segmentation for machine translation performance. In *Statmt '08 Proceedings Of The Third Workshop On Statistical Machine* (pp. 224–232). Columbus,Ohio,USA. URL: `http://nlp.stanford.edu/pubs/acl-wmt08-cws.pdf`.

Chang, R., Kivela, J., & Mak, A. (2010). Food preferences of Chinese tourists. *Annals Of Tourism Research*, *37*, 989–1011. doi:`10.1016/j.annals.2010.03.007`.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. doi:`10.1007/bf00994018`.

Dongyang, Z., Mori, T., Hayashi, K. et al. (2015). A study on preferences and behavioral patterns of Chinese tourists in Kansai region, Japan. *Konan Economic Papers*, *55*, 31–46. doi:`10.14990/00001507`.

Fan, Z.-P., Che, Y.-J., & Chen, Z.-Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*, *74*, 90–100. doi:`10.1016/j.jbusres.2017.01.010`.

Gao, J., Zhang, C., & Huang, Z. (2017). Chinese tourists' views of nature and natural landscape interpretation: a generational perspective. *Journal of Sustainable Tourism*, *26*, 668–684. doi:`10.1080/09669582.2017.1377722`.

Hargreaves, C. (2015). Analysis of hotel guest satisfaction ratings and reviews: An application in Singapore. *American Journal Of Marketing Research*, *1*, 208–214.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal Of Information Management*, *22*, 464–472. doi:`10.1016/j.ijinfomgt.2013.01.001`.

Hu, Y.-H., Chen, Y.-L., & Chou, H.-L. (2017). Opinion mining from online hotel reviews – a text summarization approach. *Information Processing & Management*, *53*, 436–449. doi:`10.1016/j.ipm.2016.12.002`.

Huang, S., Shih, Y., & Chen, K. (2008). *E-HowNet: the Expansion of HowNet*. Beijing, China: In The First National HowNet Workshop.

Japan National Tourism Organization (2017). *April 2017 Foreigner Visits Statistical Data [Data file in Excel format divided by nationality, month and year]*. Technical Report Japan National Tourism Organization. URL: `http://www.jnto.go.jp/jpn/statistics/visitor_trends/` (in Japanese).

Jones, T., Nagata, S., Nakajima, M., & Masuyama, K. (2009). Prefectural branding in Japan – tourism, national parks and the Shinshu brand. *Place Branding and Public Diplomacy*, *5*, 192–201. doi:`10.1057/pb.2009.13`.

Kim, K., joung Park, O., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*, *123*, 362–369. doi:`10.1016/j.techfore.2017.01.001`.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference On Artificial Intelligence*, *14*, 2.

Kozak, M. (2002). Measuring tourist satisfaction with multiple destination attributes. *Tourism Analysis*, *7*, 229–240(12). doi:`10.3727/108354203108750076`.

Liu, Y., Bi, J.-W., & Fan, Z.-P. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic

fuzzy set theory. *Information Fusion*, *36*, 149–161. doi:`10.1016/j.inffus.2016.11.012`.

Liu, Y., Huang, K., Bao, J., & Chen, K. (2019). Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations. *Tourism Management*, *71*, 337 – 347. URL: `http://www.sciencedirect.com/science/article/pii/S0261517718302395`. doi:`https://doi.org/10.1016/j.tourman.2018.10.004`.

Loh, S., Lorenzi, F., Saldaña, R., & Licthnow, D. (2003). A tourism recommender system based on collaboration and text analysis. *Information Technology & Tourism*, *6*, 157–165. doi:`10.3727/1098305031436980`.

Ma, G. (2015). Food, eating behavior, and culture in Chinese society. *Journal Of Ethnic Foods*, *2*, 195–199. doi:`10.1016/j.jef.2015.11.004`.

Nonaka, H., Kobayahi, A., Sakaji, H., Suzuki, Y., Sakai, H., & Masuyama, S. (2012). Extraction of the effect and the technology terms from a patent document. *Journal of Japan Industrial Management Association*, *63*, 105–111.

Nonaka, H., Kubo, D., Kimura, M., Hiraoka, T., Ota, T., & Masuyama, S. (2014). Correlation analysis between financial data and patent score based on hits algorithm. In *Proceedings of the IEEE Technology Management Conference (ITMC)* (pp. 1–4).

Powers, D. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal Of Machine Learning Technologies*, *2*, 37–63. URL: `http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf`.

Ren, G., & Hong, T. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Informa-*

tion Processing & Management, *56*, 1425–1438. doi:`10.1016/j.ipm.2018.04.003`.

Romão, J., Neuts, B., Nijkamp, P., & Shikida, A. (2014). Determinants of trip choice, satisfaction and loyalty in an eco-tourism destination: a modelling study on the Shiretoko Peninsula, Japan. *Ecological Economics*, *107*, 195–205.

Ryan, C., & Mo, X. (2001). Chinese visitors to New Zealand - demographics and perceptions. *Journal Of Vacation Marketing*, *8*, 13–27. doi:`10.1177/135676670200800103`.

Shanka, T., & Taylor, R. (2004). An investigation into the perceived importance of service and facility attributes to hotel satisfaction. *Journal of Quality Assurance in Hospitality & Tourism*, *4*, 119–134. doi:`10.1300/J162v04n03_08`.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 279–423. doi:`10.1002/j.1538-7305.1948.tb01338.x`.

Sun, Y., Wei, Y., & Zhang, L. (2017). International academic impact of Chinese tourism research: a review based on the analysis of SSCI tourism articles from 2001 to 2012. *Tourism Management*, *58*, 245–252. doi:`10.1016/j.tourman.2016.03.008`.

Truong, T., & King, B. (2009). An evaluation of satisfaction levels among Chinese tourists in Vietnam. *International Journal Of Tourism Research*, *11*, 521–535. doi:`10.1002/jtr.726`.

Vermeulen, I., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*, 123–127. doi:`10.1016/j.tourman.2008.04.008`.

World Health Organization (2015). Who global report on trends in prevalence of tobacco smoking 2015. URL: `https://apps.who.int/iris/bitstream/handle/10665/156262/9789241564922_eng.pdf` ISBN: 978-9241564922.

Wu, C. H.-J., & Liang, R.-D. (2009). Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants. *International Journal of Hospitality Management*, *28*, 586–593. doi:`10.1016/j.ijhm.2009.03.008`.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*, 179–188. doi:`10.1016/j.tourman.2009.02.016`.

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120–130. doi:`10.1016/j.ijhm.2014.10.013`.

Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International journal of hospitality management*, *55*, 57–69. doi:`10.1016/j.ijhm.2016.03.003`.

Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, *53*, 433–447. doi:`10.1177/0047287513500391`.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*, 180–182. doi:`10.1016/j.ijhm.2008.06.011`.

Zhang, H., Yu, Z., Xu, M., & Shi, Y. (2011). Feature-level sentiment analysis for Chinese product reviews. *In Proceedings of the 2011 3Rd International Conference On Computer Research And Development*, *2*, 135–140. doi:`10.1109/iccrd.2011.5764099`.

Zhang, M., Liu, S., Yang, L., Jiang, Y., Huang, Z., Zhao, Z., Deng, Q., Li,
Y., Zhou, M., Wang, L. et al. (2019). Prevalence of smoking and knowl-
edge about the smoking hazards among 170,000 Chinese adults: a nation-
ally representative survey in 2013-2014. *Nicotine & tobacco research: offi-
cial journal of the Society for Research on Nicotine and Tobacco*, *ntz020*.
doi:`10.1093/ntr/ntz020`.

Zhou, L., Ye, S., Pearce, P. L., & Wu, M.-Y. (2014). Refreshing hotel satisfac-
tion studies by reconfiguring customer review data. *International Journal of
Hospitality Management*, *38*, 1–10. doi:`10.1016/j.ijhm.2013.12.004`.

Appendices

Appendix A. Price bracket analysis details

In the Keyword Frequency Analysis in section 5.4 and Table 10, we examined the frequency of certain keywords across different price brackets. We show these ranges in Table A.11. These price brackets were selected by observing different kinds of hotels and their level of luxury, along with their price. Hostels and capsule hotels depending on their location are often in the first two price brackets. Business hotels can range from the third to fourth, while more luxurious hotels are included in the fifth and sixth ranges. The price data was mined as the average price per night in a hotel using the data available in *Ctrip*, and not for each review as it might vary by the number of nights and the kind of room chosen.

Table A.11: Price per night ranges used for analysis.

Price bracket Number	Price bracket in japanese yen
Price Bracket 1	0 yen to 5000 yen
Price Bracket 2	5000 yen to 10,000 yen
Price Bracket 3	10,000 yen to 15,000 yen
Price Bracket 4	15,000 yen to 20,000 yen
Price Bracket 5	20,000 yen to 100,000 yen
Price Bracket 6	100,000 yen to 500,000 yen