Nagaoka University of Technology

Graduate School of Engineering

Department of Information and Management Systems Engineering

Analysis of Emotional Judgement factors based on online reviews from tourists visiting Japan

(訪日観光客のオンラインレビューに基づく感情決定要因分析)

A thesis submitted in partial fulfilment of the requirement for the degree of Master of Engineering, January 2019.

Submitted by

Alemán Carreón Elisa Claire

Student No. 15340089

Supervised by

Professor Hirofumi Nonaka

Contents

Abstract

In recent years there has been a boom in international tourism in Japan, where recently there has been particularly an increasing number of Chinese tourists, among other international tourists. In response to this boom and its effect on the Japanese economy, there is a need for the development of market research tools that are cost and time effective and rely on the advantages of online reviews, which increasingly affect sales and the influx of potential customers alike. In contrast with previous research, there is a need to use these online reviews to determine emotional judgement factors, words that most closely reflect the demands of these customer bases and the emotional judgement attached to them. Using an entropy-based mathematical model and a machine learning algorithm, I determined a set of keywords representing these emotional judgement factors. With this method, I extracted the most influential keywords in emotional judgement from Chinese and English online reviews of Japanese hotels in the portal sites *Ctrip* and *TripAdvisor*. I identified that Chinese customers react positively to included breakfasts and big rooms, and that unsatisfied Chinese customers focus on the pricing of the hotel. Additionally, I found that cheaper hotels have a focus on transportation, compared to luxury hotels. I also found English speaking customers dislike pricey hotels and have a high dislike for dirty rooms and stains, and that they mainly focus positively on staff, cleanliness, and transportation availability. Because the process is based on probabilistic distributions and automatic algorithms, it has potential future applications in cross-language businesses, especially the hospitality industry where language barriers are a constant problem.

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 [1]. Particularly in Japan, however, international tourism has been increasingly exerting a stronger influence in the economy [2]. 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 [3]. 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 the customer response from several U.S. pizza selling companies [4]. Another study ranked products by using online reviews and sentiment analysis [5], while another used them to predict product sales [6]. Other uses for text mining techniques include patent document analysis, as was done in a study on patent scoring [7], and another on the extraction of technological terminology in patents [8]. It has also been used in the tourism and hospitality field to predict hotel demand by using web traffic data [9] or to make recommender systems [10]. 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 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 satisfac-

2

tion or dissatisfaction. Previous studies have found that there is a correlation between prominent reviews and sales figures [11, 12]. 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 [13]. 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 an availability of data from tourists writing in many languages. It is because of this important impact the hotel reviews have to consumers, and the highly polarized emotional content these reviews hold that I 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. 14, 15, 16, 17, 18, 19], while others focus on motivation, choice or behavior [e.g. 20, 21, 22]. 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 examining international tourism, they are limited by language barriers. In my study, I 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 tied to each emotional response.

In my study, I developed a text mining method to extract emotional judegment keywords and analyze the needs of Chinese customers of Japanese Hotels. I chose this combination of population and destination because I aim to develop a methodology that can be applied in any language and regardless of language barriers on a large scale. For comparison I chose to also study the reviews from English speaking people staying in Japan. The extracted emotional judegment keywords became the key to researching which topics and in

3

what order of priority the general Chinese tourist population 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, 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 emotional judegment in the satisfaction or dissatisfaction of Chinese customers of Japanese Hotels because of the language barrier that can be studied and the economic weight that these customers represent, as well as studying English speaking tourist populations for comparison. These emotional judegment 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.

I aim to create 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, I propose to use supervised machine learning to be trained on emotional judegment classification of 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, I also analyze the relevance of these keywords in different hotel price brackets to observe potential differences.

4

3. Related Work

In the hospitality field, satisfaction, motivation, and behavior have been studied for a long time using surveys [e.g. 16, 17, 18, 19, 20, 21, 22, 23]. More recently, studies have used big data in the hospitality industry [e.g. 9, 10], as well as data mining tools to analyze hotel reviews [e.g. 24, 25, 26]. However, most focus on either the sentiment classification without analyzing the words influencing that emotion, or they study the words without taking the emotions into account. Other studies focus on review helpfulness [27]; while others focus on summarization of reviews into smaller text [e.g. 28, 29]. There has not been yet a study that focuses on the conceptual words that lead directly to an emotional decision from the user. 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 [30]. 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 [31]. However, this study doesn't deterministically show which topics directly influence satisfaction and dissatisfaction or the order in which they do.

On the other hand, Zhang et al. performed a market analysis based on the sentiment analysis of product reviews using a Chinese sentiment dictionary called HowNet [32]. The problem with using HowNet as a sentiment dictionary is that the words are classified based on ontology [33], 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 positive and that

5

"staff" is being praised by the use of "friendly" in the sentence. These non-emotional words are directly influencing the emotional judegment 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 emotional judegment taken by the user, clearly defining not only the emotions 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 [34]. 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 [35]. 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 [36]. 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. My 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.

Previous research in the field of linguistics and psychology has been made in an effort to detect and classify emotion in Chinese texts, using emotional verbs and other linguistic features in texts [e.g. 37, 38]. Many studies are mainly based on the Sinica Treebank [39], Sinica Corpus [40], Penn Chinese Treebank [41], and Chinese Gigaword Corpus [42], which contain texts on subjects like science, philosophy, literature, art, financial newsletters, articles, among other formal texts. While using these can be effective in classifying emotional states

in similarly formal or literary texts, there are no appropriate training corpora for this current endeavor, which includes informal internet slang and a focus on service or product satisfaction. This is why I have decided to train my model using only the collected data. That is to say, the segmentation of Chinese words using these corpora has been done successfully in other hotel review studies [e.g. 43], and so I have decided to employ them as well.

4. Methodology

I 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 emotional judegment behind positive and negative opinions in a review using an entropy-based mathematical extraction method. These keywords extracted using entropy related to emotional judegment 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 positive and negative with an optimized SVM, I have analyzed the weight value assigned to them by the SVM. For words that are not part of the support vector this is equal to 0; however, the support vector lets us observe the words that, while potentially close to the border between positive and negative sentiments, provide a strong and clear distinction in the emotional classification. This will be explained further in section 4.2.3.1. These words, especially those signifying a subject or topic, allow us to analyze the writing tendencies of Chinese customers in either positive or negative reviews. I also observed the frequency of the terms in all of the reviews to extract the most utilized words in either kind of reviews. I show an overview of this methodology in Figure 1. I repeated this experiment in a corpus of reviws of Japanese hotels written in English for comparison. Further-

---

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

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

7

more, I 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.

4.1. Data Collection

In this study I 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 [20] 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 my corpus, irregularities occurred where the reviews were written in other languages or where unusual punctuation marks
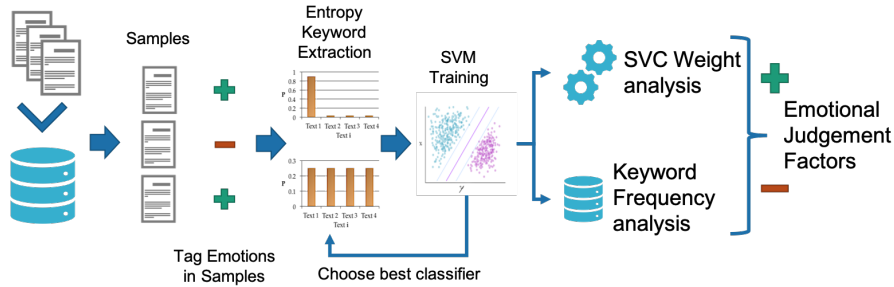

Figure 1: Overview of the Methodology

8

were used. I 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.

In the case of English text, 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 I 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 impact the user's emotional judegment, I 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 [44] or it can be thought of as the grade of impossibility of predicting an outcome. Using this value I 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. I show this concept in Figure 2.
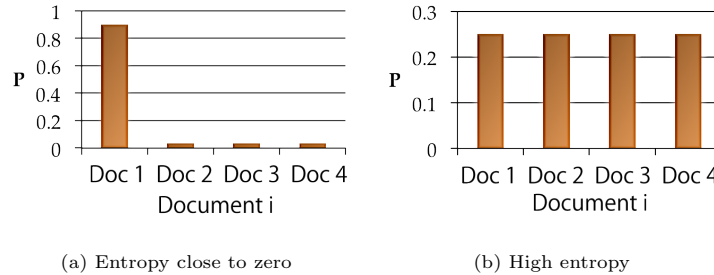

(a) Entropy close to zero   (b) High entropy

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

9

To apply this logic, I retrieved 50 reviews as a sample of my corpus and with the collaboration of a group of 5 Chinese students, tagged each sentence of a total of 159 as the classes positive or negative 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, I 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 positive set than to the negative set by a factor of α were determined to be keywords influencing positive emotional judegment in Chinese reviews of hotels. Likewise, words with higher entropy for the negative set than the positive set by a factor of α' were determined to be keywords related to the negative emotional judegment in my texts.

In the formulas below I show my process for the calculation the the entropy of each word i in each document j in the positive set of documents H_{Pj} (1) and the entropy for the negative set H_{Nj} (3) below, using the probabilities of a word being in a positive document P_{ijP} (5) and likewise for negative documents P_{ijN} (6). N_{ijP} is the number of times a word i is included in a document j from the positive set and likewise for N_{ijN} with regard to the negative set. Lastly I show the keyword determining comparison formula using the mutually independent coefficients α and α' in the formulas (7) and (8) respectively.

$$H_{Pj} = -\sum_{i=1}^{M}[P_{ijP} \log_2 P_{ijP}] \tag{1}$$

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

$$H_{Nj} = -\sum_{i=1}^{M}[P_{ijN} \log_2 P_{ijN}] \tag{3}$$

$$\lim_{P_{ijN} \to 0+} P_{ijN} \log_2 P_{ijN} = 0 \tag{4}$$

10

$$P_{ijP} = \frac{N_{ijP}}{\sum_{i=1}^{M} N_{ijP}} \tag{5}$$

$$P_{ijN} = \frac{N_{ijN}}{\sum_{i=1}^{M} N_{ijN}} \tag{6}$$

$$H_{Pj} > \alpha H_{Nj} \tag{7}$$

$$H_{Nj} > \alpha' H_{Pj} \tag{8}$$

4.2.2.2. Support Vector Machine Classification

Support Vector Machines are supervised machine learning models usually applied to classification or regression problems [45]. I use it to classify the rest of my corpus into emotional classes in my 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. I show a two dimensional example in Figure 3.

In my study, I used the linear kernel of the SVM classification process, defined by the formula (9) below. Each training sentence is a point of data, a row in the vector x, where each column represents a feature, in my case the quantities of each of the keywords in that particular sentence. The labels of previously known classifications (1 for positive, 0 for negative) for each sentence comprise the $f(x)$ vector. The Weight Vector w is comprised of the influences each point has had in the training process to define the angle of the hyperplane and the bias coefficient b determines its position, as explained in section 4.2.3.1. After the hyperplane is drawn, the classification of the new data is done as follows in formula (10) and (11).

$$f(x) = w^{\top} x + b \tag{9}$$

11

$$f(x_{new}) \geq 0 \rightarrow y_{new} = +1 \tag{10}$$

$$f(x_{new}) \leq 0 \rightarrow y_{new} = -1 \tag{11}$$

At the beginning of the process, the initial weights for each word are set to zero. Then, as a random separating vector is drawn, each data point x_i is tested and if the classification for it fails, the value for w is changed by a value of α as follows (12). This α is not to be confused with the entropy comparison α and α' in section 4.2.2.1.

$$w \leftarrow w + \alpha sign(f(x_i))x_i \tag{12}$$

The process is then repeated until the classification error is minimized as much as possible. The resulting function for classification can be expressed as follows (13). In the end, although the weight vector is the same size as the number of feature words, weight is only assigned to words that are explicitly inside the support vectors that are close to and guide the separating hyperplane.

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i (x_i^\top \cdot x) + b \tag{13}$$

In the field of corpus study and Natural Language Processing, each of the features of a data point is the number of times that a word is included in a document. In my study I used a number of keyword lists, defined by my entropy calculations with different comparison coefficients, as the possible features; trained the SVMs implementing the Support Vector Classifier included in the python library *scikit-learn*[6]; managed the vector mathematics with the mathematical python library *numpy*[7]; and tested for each one using the K-fold Cross-Validation method, which has been proven to provide good results in small

---

[6]Scikit-Learn `http://scikit-learn.org/`

[7]Numpy. `http://numpy.org/`

samples [46]. In each test I calculated the Precision, Recall and F_1-measure [47]
for my predictions.

4.2.2.3. Model Evaluation Metrics

In order to measure the effectiveness of the training process and data, I performed what is called a K-fold cross-validation. This means that after randomly shuffling and splitting my 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, I used the F_1 score, which measure is a harmonic mean between precision and recall. Precision, described in formula (14), lets us observe the rate of correct positive predictions from all the positive predictions, while Recall, detailed in formula (15), observes the rate of correct positive predictions from the total of actual positive data. The F_1 score in formula (16) then can only be high when both of these measures are high simultaneously and will lower substantially if they are not consistent. I use this score as it allows us to avoid overlooking data while maintaining accurate predictions. The overall accuracy (17), which doesn't account for overlooking of data, is also observed for reference but isn't used as the main evaluation of the model.

Table 1: Prediction outcomes

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

13

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{14}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{15}$$

$$F_1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{16}$$

$$Accuracy = \frac{TruePositives + TrueNegatives}{NumberOfPredictions} \tag{17}$$

4.2.3. Data Analysis

4.2.3.1. SVM Weight Analysis

As stated earlier in section 4.2.2.2, each point of data that is classified incorrectly causes a change in the weight vector to better locate the separating hyperplane and 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, or in my case, each of these normally ambiguous words, had for the classification process and the class it has a decisive influence in. 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 I show the formula for the weight vector (18).

14

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{18}$$

4.2.3.2. Keyword Frequency Analysis

Because the machine learning model used in my study is constructed from training data obtained by a relatively small sample of my data, after training the model and classifying the rest of my data, I studied the frequencies of each entropy based keyword in my 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. Keywords with lower frequencies can be interpreted to be emotional keywords to a small sector of the user-base, but this information is important nevertheless from a mathematical and a marketing standpoint.

5. Results

5.1. Data set

In my 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, I extracted a total of 44,912 reviews, which were comprised of 286,109 separate sentences. In my corpus, there were 23,443 different words used, from which 2,802 were noise characters.

5.2. Sentiment Analysis Results

After having my training data tagged by a group of 5 Chinese student collaborators, I experimented with different comparison coefficients for the entropy values calculated from the 'positive' and 'negative' 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. I repeated the process for the reviews in English that my colleagues tagged.

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

ultimately decided to use the linear kernel for the benefits of the weight vector obtainability. I 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. I found the ideal value for this parameter was C=0.5 in my final classifier for Chinese text, and 2 in the English text one.

I trained a different Support Vector Classifier with each of the lists, and I chose the best performing lists for each emotional class, resulting in a positive emotion classifier (positive or not positive) and a negative emotion classifier (negative or not negative), 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 I 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 positive emotion classifier, which mostly misclassified sentences with negative words, I decided to combine both keyword lists into a single large list to train the positive 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 my testing data, I obtained higher performance results overall. I 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 my training samples original classification.
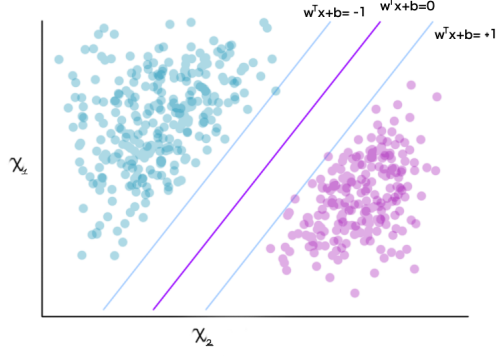
16

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
Positive keywords ($\alpha = 2.75$)	Positive emotion	2.5	0.87	0.02	0.91	0.01
Negative keywords ($\alpha' = 3.75$)	Negative emotion	0.5	0.85	0.05	0.67	0.11
Combined (α=2.75, α'=3.75)	**Positive emotion**	**0.5**	**0.92**	**0.03**	**0.95**	**0.01**

5.3. Keywords and their SVM Weight Values

In Table 4 I show some of my keywords that signify subjects or topics, and as such, help us understand the needs of the users. I 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 I show keywords for the English classifier with high weight values.

5.4. Keyword Frequencies

I observed the words with the highest frequencies for keywords in positive and negative statements to study the needs of Chinese customers, as shown in the Tables 6 and 7. The highest frequencies for positive and negative keywords are shown in Tables 8 and 9.

In order to further analyze the data I obtained, aside from emotional classification, I also investigated the difference in emotional response from customers of hotels in different price brackets in the case of Chinese reviews. I separated hotel prices in 6 categories, and in Table 10 I show the words that have an interesting change in their ranking order when ordered by frequency in each price bracket. I include the price bracket definitions in A in Table 12.

Now, some words, despite not being highly frequent in the complete dataset, were extracted in my sample as keywords when evaluating their emotional entropy. I show some interesting examples of these in Table 11, as they can be keywords for a niche demand.

6. Discussion

6.1. Needs of the Chinese tourist market in Japan

Analyzing the extracted keywords is the key to performing a market study of Chinese customers of Japanese hotels since they can be interpreted as explicit needs. Analyzing keywords of positive emotional judegment, I 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

18

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
Positive keywords ($\alpha = 1.5$)	Positive emotion	1.75	0.79	0.02	0.82	0.02
Negative keywords ($\alpha' = 4.25$)	Negative emotion	3	0.70	0.04	0.80	0.03
Combined (α=1.5, α'=4.25)	**Positive emotion**	**2**	**0.81**	**0.02**	**0.83**	**0.02**

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

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

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

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

Table 6: High weight Chinese keywords in positive 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 negative 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 positive 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 negative 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

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

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

Table 11: Interesting keywords that could represent niche demand

Word	Translation	Observation
烤肉	Yakiniku (Japanese grilled meat)	While other keywords describe food in very general terms, this keyword is a specific dish that is classified as positive by the SVC.
花园, 庭院, 园林	Three different words for garden parks, flower gardens, courtyards	There is a very specific search for either gardens inside the hotel or nearby the hotel, across all price brackets, but specially in higher priced ones.

that reviewers were praising, in general, the cleanliness of Japan's environment, un-littered streets, potable water and their culture of respecting spaces. my results are compatible with Ryan & Mo's study of Chinese tourist behavior in New Zealand [48], 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.

Another key component I 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 my sentences emotionally classified as positive. While other food-related words were extracted, most of them were general in nature, like "food" or "eating". In contrast, the word "breakfast", which is referring to a specific time and possibly a specific kind of food, 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 [49] and an article on Chinese food culture [50] 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 my data points to the fact that more than Chinese food as a specific concept, hotel reviews are written with general words that represent food. I 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.

According to my 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; environment, such as gardens or parks nearby; and services nearby, like places to go shopping.

Now, when analyzing negative responses, the most frequently criticized aspect of Japanese hotels was the price. This was true across all of the price

23

brackets, from the cheapest hotels to the most luxurious ones. In their study of Chinese tourists in Vietnam, Truong & King [16] 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 my 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. my 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 my data, most customers can be thought to have complained about written translations.

Additionally, as shown in Tables 10 and 11, there is a positive reaction to "Japanese-style", which I infer in this context can mean rooms or possibly ceremonies, in luxury rate hotels, and there is a possibility that there are words that can be extracted to research niche needs for tourists that travel for food purposes. 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 cusomers 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 if we consider frequency, although the order is reversed if we consider the SVC weight assigned to

24

the word. However, when we consider the negative assosiations, price is the first concern for both kinds of customers, but there's a high concern for cleanliness, such as complaints about stains, dirty rooms or carpets. A thing to note is that there is a high number of complaints for cigarette smell, in comparison to the Chinese customers in which it isn't a keyword. 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. 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.

6.3. Implications for the Japanese Hotel Industry

These results can be utilized directly as a market study, which shows the needs of this large and increasing segment of the customers in a direct manner. This would have otherwise presented difficulties with traditional methods 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.

Not only that but the tool I have developed, with some adjustments, might be utilizable in other contexts for analyzing Chinese tourism in Japan in general, and my methodology can be used as well with other populations and languages. That is, I have developed a potential tool for automatic market research, which is able to reduce costs and difficulties that are presented when using surveys and interviews, which are the most common tool in the marketing research ambit, both in academia and in the industry, as well as translations and interpreters, which are not only costly, but there is also an overwhelming demand for skilled translators when there are not as many of them available.

25

7. Conclusions and Future Work

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

Using the weight vectors of my classifiers, as well as frequency of the words in my dataset, I 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. I 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 I also repeated the experiments on reviews written by English speaker tourists staying at Japanese hotels from *TripAdvisor*. Observing the results I 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. I also found that they prefer friendly staff and transportation available from the airport or subway.

In future work I plan to investigate further into this topic, extending my data set and researching for different trends across time 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, I would also like to study the food preferences of Chinese tourists, as well as the impact of the recommendation in different price brackets. Additionally, it would be

26

interesting to study further into more specific emotions than the satisfaction and dissatisfaction classifications I performed in my study.

Acknowledgements

27

References

References

565

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

[2] T. Jones, S. Nagata, M. Nakajima, K. Masuyama, Prefectural branding in Japan – tourism, national parks and the Shinshu brand, Place Branding and Public Diplomacy 5 (3) (2009) 192–201. `doi:10.1057/pb.2009.13`.

570

[3] Japan National Tourism Organization, April 2017 foreigner visits statistical data [data file in excel format divided by nationality, month and year], Tech. rep., Japan National Tourism Organization, (in Japanese) (2017).
URL `http://www.jnto.go.jp/jpn/statistics/visitor_trends/`

[4] W. He, S. Zha, L. Li, Social media competitive analysis and text mining: A case study in the pizza industry, International Journal Of Information Management 22 (3) (2013) 464–472. `doi:10.1016/j.ijinfomgt.2013.01.001`.

575

[5] Y. Liu, J.-W. Bi, Z.-P. Fan, Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory, Information Fusion 36 (2017) 149–161. `doi:10.1016/j.inffus.2016.11.012`.

580

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

585

[7] H. Nonaka, D. Kubo, M. Kimura, T. Hiraoka, T. Ota, S. Masuyama, Correlation analysis between financial data and patent score based on hits

algorithm., in: Proceedings of the IEEE Technology Management Conference (ITMC), 2014, pp. 1–4.

[8] H. Nonaka, A. Kobayahi, H. Sakaji, Y. Suzuki, H. Sakai, S. Masuyama, Extraction of the effect and the technology terms from a patent document, Journal of Japan Industrial Management Association 63 (2E) (2012) 105–111.

[9] Y. Yang, B. Pan, H. Song, Predicting hotel demand using destination marketing organization's web traffic data, Journal of Travel Research 53 (4) (2014) 433–447. doi:10.1177/0047287513500391.

[10] S. Loh, F. Lorenzi, R. Saldaña, D. Licthnow, A tourism recommender system based on collaboration and text analysis, Information Technology & Tourism 6 (3) (2003) 157–165. doi:10.3727/1098305031436980.

[11] S. Basuroy, S. Chatterjee, S. Ravid, How critical are critical reviews? the box office effects of film critics, star power, and budgets, Journal Of Marketing 67 (4) (2003) 103–117. doi:10.1509/jmkg.67.4.103.18692.

[12] Q. Ye, R. Law, B. Gu, The impact of online user reviews on hotel room sales, International Journal of Hospitality Management 28 (1) (2009) 180–182. doi:10.1016/j.ijhm.2008.06.011.

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

[14] S. Balbi, M. Misuraca, G. Scepi, Combining different evaluation systems on social media for measuring user satisfaction, Information Processing & Management 54 (4) (2018) 674–685. doi:10.1016/j.ipm.2018.04.009.

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

29

[16] T. Truong, B. King, An evaluation of satisfaction levels among Chinese tourists in Vietnam, International Journal Of Tourism Research 11 (6) (2009) 521–535. `doi:10.1002/jtr.726`.

[17] C. H.-J. Wu, R.-D. Liang, Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants, International Journal of Hospitality Management 28 (4) (2009) 586–593. `doi:10.1016/j.ijhm.2009.03.008`.

[18] T. Shanka, R. Taylor, An investigation into the perceived importance of service and facility attributes to hotel satisfaction, Journal of Quality Assurance in Hospitality & Tourism 4 (3-4) (2004) 119–134. `doi:10.1300/J162v04n03_08`.

[19] M. Kozak, Measuring tourist satisfaction with multiple destination attributes, Tourism Analysis 7 (3-4) (2002) 229–240(12). `doi:10.3727/108354203108750076`.

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

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

[22] Z. Dongyang, T. Mori, K. Hayashi, et al., A study on preferences and behavioral patterns of Chinese tourists in Kansai region, Japan, Konan Economic Papers 55 (1-2) (2015) 31–46. `doi:10.14990/00001507`.

[23] A. Chan, C. H. Hsu, T. Baum, 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 (1-2) (2015) 18–33. `doi:10.1080/10548408.2014.986010`.

[24] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, O. Qawasmeh, Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features, Information Processing & Managementdoi:10.1016/j.ipm.2018.01.006.

[25] V. Browning, K. K. F. So, B. Sparks, The influence of online reviews on consumers' attributions of service quality and control for service standards in hotels, Journal of Travel & Tourism Marketing 30 (1-2) (2013) 23–40. doi:10.1080/10548408.2013.750971.

[26] Z. Xiang, U. Gretzel, Role of social media in online travel information search, Tourism Management 31 (2) (2010) 179–188. doi:10.1016/j.tourman.2009.02.016.

[27] G. Ren, T. Hong, Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews, Information Processing & Managementdoi:10.1016/j.ipm.2018.04.003.

[28] Y.-H. Hu, Y.-L. Chen, H.-L. Chou, Opinion mining from online hotel reviews – a text summarization approach, Information Processing & Management 53 (2) (2017) 436–449. doi:10.1016/j.ipm.2016.12.002.

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

[30] C. Hargreaves, Analysis of hotel guest satisfaction ratings and reviews: An application in Singapore, American Journal Of Marketing Research 1 (4) (2015) 208–214.

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

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

[33] S. Huang, Y. Shih, K. Chen, E-HowNet: the Expansion of HowNet, In The First National HowNet Workshop, Beijing, China, 2008.

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

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

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

[37] S. Y. M. Lee, Y. Chen, C.-R. Huang, A text-driven rule-based system for emotion cause detection, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Association for Computational Linguistics, 2010, pp. 45–53.

[38] Y. Chen, S. Y. M. Lee, S. Li, C.-R. Huang, Emotion cause detection with linguistic constructions, in: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, 2010, pp. 179–187.

[39] F. Chen, P. Tsai, K. Chen, C. Huang, The construction of Sinica Treebank,

32

Computational Linguistics And Chinese Language Processing 4 (2) (1999) 87–104.

[40] C. Huang, K. Chen, A Chinese corpus for linguistics research, in: Proceedings of the 1992 International Conference on Computational Linguistics (COLING-92), Nantes, France, 1992, pp. 1214–1217.

[41] N. Xue, F. Xia, F. Chiou, M. Palmer, The Penn Chinese TreeBank: Phrase structure annotation of a large corpus, Natural Language Engineering 11 (2) (2005) 207–238. `doi:10.1017/s135132490400364x`.

[42] C. Huang, A. Kilgarriff, Y. Wu, C. Chiu, S. Smith, P. Rychlý, M. Bai, K. Chen, Chinese sketch engine and the extraction of collocations, in: Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing, Jeju, Korea, 2005, pp. 48–55.

[43] C.-H. Chen, W.-P. Lee, J.-Y. Huang, Tracking and recognizing emotions in short text messages from online chatting services, Information Processing & Management 54 (6) (2018) 1325–1344. `doi:10.1016/j.ipm.2018.05.008`.

[44] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (3) (1948) 279–423. `doi:10.1002/j.1538-7305.1948.tb01338.x`.

[45] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. `doi:10.1007/bf00994018`.

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

[47] D. Powers, Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation, Journal Of Machine Learning Technologies 2 (1) (2011) 37–63.
URL `http://www.flinders.edu.au/science_engineering/fms/`

725   `School-CSEM/publications/tech_reps-research_artfcts/TRRA_`
`2007.pdf`

[48] C. Ryan, X. Mo, Chinese visitors to New Zealand - demographics and perceptions, Journal Of Vacation Marketing 8 (1) (2001) 13–27. `doi:10.1177/135676670200800103`.

730 [49] R. Chang, J. Kivela, A. Mak, Food preferences of Chinese tourists, Annals Of Tourism Research 37 (4) (2010) 989–1011. `doi:10.1016/j.annals.2010.03.007`.

[50] G. Ma, Food, eating behavior, and culture in Chinese society, Journal Of Ethnic Foods 2 (4) (2015) 195–199. `doi:10.1016/j.jef.2015.11.004`.

Appendices

Appendix A   Price bracket Analysis Details

In my Keyword Frequency Analysis in section 5.4 and Table 10, I examined the frequency of certain keywords across different price brackets. I show these ranges in Table 12. 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 12: 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