Analyzing preferences differences between Chinese and English speakers in Japan hotel staying: A text mining method

Elisa Claire Alemán Carreón[a,*], Hirofumi Nonaka[a], Hugo Alberto Mendoza España[a], Toru Hiraoka[b]

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

Abstract

Abstract

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

Keywords: Sentiment Analysis, Hotels and Lodging, Machine Learning, Chinese, English, Preferences

1. Introduction

Recently, the Japanese economy has been more and more affected by an increase in inbound international tourism (Jones et al., 2009) with a Year-on-Year Growth Rate of 19.3% in 2017, with a total of 28,691,073 inbound tourists

*Corresponding author
Email addresses: s153400@stn.nagaokaut.ac.jp (Elisa Claire Alemán Carreón), nonaka@kjs.nagaokaut.ac.jp (Hirofumi Nonaka), s173330@stn.nagaokaut.ac.jp (Hugo Alberto Mendoza España), hiraoka@sun.ac.jp (Toru Hiraoka)

that year (Japan National Tourism Organization, 2019). From this total, the tourist population was mostly Asian (86.14%), with approximately a fourth of the total (25.63%) coming from China. Western countries, counting English speaking countries in addition to the whole of Europe make for 11.4% of the total, with a 7.23% of the total being countries where English is the official or the de facto national language. Specifically for Chinese tourists, the effect on international economies as well as the number of researchers interested in this phenomenon has been increasing as well (Sun et al., 2017).

However, up until recently studies on social sciences, and that includes tourist behavior, have been performed using surveys on populations that could be culturally biased for the western world (Nielsen et al., 2017; Jones, 2010; Gunaratne, 2009; Hogan & Emler, 1978). Those that do include Asian populations in their analysis, most commonly Chinese tourist behavior (e.g. Liu et al., 2019; Chang et al., 2010; Dongyang et al., 2015), and a few that compare Asian to western tourist behavior (e.g. Choi & Chu, 2000), are commonly survey or interview based studies with small samples, which while valid can have its limitations.

With the advent of Web 2.0 and customer review websites, researchers realized the benefits of online reviews for research, and their importance for sales (Ye et al., 2009; Basuroy et al., 2003), customer consideration (Vermeulen & Seegers, 2009) and perception of services and products (Browning et al., 2013), among other effects of online interactions between customers (e.g. Xiang & Gretzel, 2010; Ren & Hong, 2019). Consequentially, tourism research also began to use information collected online for data mining analysis, such as opinion mining (e.g. Hu et al., 2017), predicting hotel demand from online traffic (Yang et al., 2014), recommender systems (e.g. Loh et al., 2003), and more. Data mining and big data methodologies can increase the number of samples from the hundred or so samples manageable by researchers to the hundreds of thousands that are automatically analyzed by machines. This can not only help confirm existing theories, but also lead to finding new patterns and to knowledge discovery (Fayyad et al., 1996).

In this study we take advantage of the availability of enormous amounts of data online for Chinese and western English speaking tourist populations' online reviews for Japanese hotels to both confirm existing theories about their differences in behavior, as well as perform an exploration of the data to discover factors that could have been missed in the past. For this purpose, we use machine learning to automatically classify review sentences as positive or negative opinions of the hotel, and perform a statistical extraction of the topics that the customers of each population are most concerned about.

2. Research objective

The objective of this study is to determine the difference in preferences between Chinese speaking customers and English speaking customers of Japanese Hotels using text-mining techniques. In the past, survey based studies have provided a theoretical background for specific customer groups, and cultural and language differences often cannot be observed in a single study. We propose to use large scale data from online hotel reviews in Chinese and English to study these differences in a statistical manner.

These difference in preferences 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.

3. Theoretical background and hypothesis development

3.1. Customer satisfaction and dissatisfaction during hotel lodging

Customer satisfaction in tourism has been analyzed since decades past, Hunt (1975) having defining customer satisfaction as the realization or overcoming of expectations towards the service. Oliver (1981) defined it as an emotional response to the provided services in retail and other contexts, and Oh & Parks (1996) reviewed the psychological processes of customer satisfaction for the hospitality industry. It is generally agreed upon that satisfaction and dissatisfaction stem from the individual expectations of the customer, and as such, Engel et al. (1990) states that satisfaction and dissatisfaction are therefore influenced by each customer's background. This is why Western and Asian, specifically Chinese, customers can have very different factors of satisfaction and dissatisfaction since they have different backgrounds and cultures. These different backgrounds will lead to different expectations of the services that a hotel can provide for them, the experiences they should have while staying at a hotel, and the level of comfort that they will have, from the moment that they choose the hotel throughout their stay. These differing expectations in turn, will determine the differing factors of satisfaction and dissatisfaction for each kind of customer, as well as the order in which they prioritize them. Therefore we propose our first and second hypothesis:

Hypothesis 1: *Chinese and Western tourists have different satisfaction and dissatisfaction factors.*

Hypothesis 2: *Chinese and Western tourists have different priorities when it comes to similar satisfaction and dissatisfaction factors.*

Studies on customer satisfaction (e.g. Truong & King, 2009; Romão et al., 2014; Wu & Liang, 2009) commonly use the Likert scale (Likert, 1932) (e.g. 1 to 5 scale from strongly dissatisfied to strongly satisfied) to perform statistical analysis of which factors relate most to satisfaction on the same dimension as dissatisfaction (e.g. Chan et al., 2015; Choi & Chu, 2000). This leads to correlation analyses (either multivariate or single variable ones) where one factor can lead to satisfaction, while it is implied that the lack of it can lead to dissatisfaction. However, a binary distinction (satisfied or dissatisfied) could allow us to analyze the factors that solely correlate to satisfaction, as well as exploring factors which are solely linked to dissatisfaction. There are fewer examples of this approach, but studies have also done this in the past (e.g. Zhou et al., 2014). While it is true that this method can decrease the extent to which we can analyze degrees of satisfaction or dissatisfaction, it has the benefit that it can be applied to a large sample of text data via sentiment classification techniques.

Previous research has also focused on factors that are controllable by the hotel managers and staff, i.e. hotel services, staff behavior, or facilities (e.g. Shanka & Taylor, 2004; Choi & Chu, 2001), while the satisfaction might also be influenced by factors that are uncontrollable by the hotel staff, such as surroundings, location, language immersion of the country as a whole, or of touristic destinations, as well as integration of the hotel with tours available nearby, among other factors that can play a part in the customers choice behavior and satisfaction. This leads to our third hypothesis:

Hypothesis 3: *Satisfaction and dissatisfaction stems from both internal (managerial) and external (environmental) attributes of the hotel.*

3.2. Chinese and Western tourist behavior

In the past, tourist behavior analyzed from western samples and surveys was wrongly thought to be a representation of universal behavior across all cultures (Nielsen et al., 2017; Jones, 2010; Gunaratne, 2009; Hogan & Emler, 1978). Recently, however, with the rise of Chinese outbound tourism, both academic researchers and businesses have decided to study Chinese tourist behavior (Sun et al., 2017). This results in a number of studies focusing on only the behavior of this subset of tourists. To this day, cross-cultural studies and analyses for Asian and Western tourists have been scarce. A few examples are Choi & Chu (2000), where it was found that Western tourists visiting Hong Kong are satisfied more with room quality while Asians are satisfied with value for money; Bauer et al. (1993), where Westerners prefer the hotel health facilities while the Asian tourists were more inclined to peruse the Karaoke facilities of hotels, and both groups tend to have high expectations about facilities; Kim & Lee (2000), where American tourists were found to be individualistic and motivated by novelty, while Japanese tourists were found to be collectivist and motivated by prestige and family, with escape from routine and an increase in knowledge as a common motivator. One thing to note with the above cross-culture analyses is that they were performed before around the year 2000. The current Chinese economy boom should make an increase in influx, but the question that could be made is if that boom created a difference in the expectations of tourists and as such, of their satisfactions and dissatisfactions when traveling.

Other more recent studies, perhaps recognizing that samples being comprised of people from Western industrialized countries isn't representative, have gone further and studied people from many countries in their samples, and performing a more universal and holistic (not cross-culture) analysis. Choi & Chu (2001), for example, analyzed hotel guest satisfaction determinants in Hong Kong with surveys in English, Chinese and Japanese translations, with people from many countries in their sample. Choi & Chu (2001) found that staff service quality, room quality and value for money were the top satisfaction determinants. For another example, Uzama (2012) produced a typology for foreigners coming to Japan for tourism, without making distinctions for their culture, but their motivation in traveling in Japan. In another study, Zhou et al. (2014) analyzed hotel satisfaction using English and Mandarin online reviews from guests staying in Hangzhou, China coming from a number of different countries; although the factors were analyzed in general without much cross-cultural analysis, although their general satisfaction score was noticed to be different. Zhou et al. (2014) thus found that customers are universally satisfied by welcome extras, dining environments and special food services.

Regarding Western tourist behavior, a few examples can tell us what to expect when analyzing our data. Kozak (2002) found that British and German tourists' satisfaction determinants while visiting Spain and Turkey were hygiene and cleanliness, hospitality, the availability of facilities and activities, and accommodation services. Shanka & Taylor (2004) found that English speaking tourists in Perth, Australia were most satisfied with staff friendliness, efficiency of check-in and check-out, restaurant and bar facilities and lobby ambiance.

Regarding outbound Chinese tourists, academic studies about Chinese tourists have increased (Sun et al., 2017). In general Chinese tourist populations have been found to have specific attributes by different researchers. According to Ryan & Mo (2001) and their study of Chinese tourists in New Zealand, Chinese tourists prefer nature, cleanliness and scenery in contrast to experiences and activities. With some overlap, Dongyang et al. (2015) studied Chinese tourists in the Kansai region of Japan, and found that Chinese tourists are satisfied mostly with exploring the food culture of their destination, cleanliness and staff. Studying Chinese tourists in Vietnam, Truong & King (2009) found that Chinese tourists are highly concerned with value for money. According to Liu et al. (2019), Chinese tourists tend to have harsher criticism when compared with other international tourists. And according to Gao et al. (2017) who analyzed Chinese tourists and their connection to nature while traveling in different generations, Chinese tourists prefer nature overall, but the younger generations seems to do so less than their older counterparts.

Although the studies focusing on only Chinese tourists or only Western tourists have a narrow view, their theoretical contributions are valuable. We can see that depending on the study and the design of questionnaires, the as well as the destinations, the results can vary greatly. Not only that but while there seems to be both some overlap in most studies, some factors are completely ignored in one study while not in the other. However, with data mining, the definition of each subject or factor is left for the review writers to decide en masse, and be selected through statistical methods alone. This can open opportunities to find factors that the writers of this study would not have contemplated, or avoid enforcing a factor on the mind of reviewers by presenting them with a question that they didn't think of by themselves. In addition, this study could help us compare the satisfaction and dissatisfaction factors cross-culturally and compare them with the existing literature.

3.3. Data mining, knowledge discovery and sentiment analysis

In the current world, data is presented to us in larger and larger quantities. What was commonly only seen in very specialized large laboratories with super computers a couple of decades ago is now common data size for market and managerial studies, independent university students and any scientist that can connect to the Internet. Such quantities of data are available to study now more than ever, but it would be impossible for researchers to parse all of this data by themselves. As Fayyad et al. (1996) summarizes, data by itself is unusable until it goes through a process of selection, preprocessing, transformation, mining and evaluation, and only then it can be established as knowledge. With the tools available to us in the are of information science, algorithms can be used to detect patterns that would take researchers too long to recognize, which can be evaluated to generate knowledge. This process is called Knowledge Discovery in Databases.

Now, there are of course many sources of numerical data to be mined, but perhaps what is most available and interesting to managerial purposes is the resource of customers' opinions in text form. With the introduction of Web 2.0, a never before seen quantity of valuable information is being posted to the Internet at a staggering speed. Text mining then, has been proposed more than a decade ago to utilize this data (e.g. Rajman & Besançon, 1998; Nahm & Mooney, 2002), using what is called Natural Language Processing to parse language in a way that it can be analyzed by a computer, and improved on with the years. This has been used in the field of hospitality as well for many purposes, including satisfaction analysis from reviews (e.g Berezina et al., 2016; Xu & Li, 2016; Xiang et al., 2015; Hargreaves, 2015; Balbi et al., 2018), social media's influence on travelers (e.g. Xiang & Gretzel, 2010), review summarization (e.g. Hu et al., 2017), perceived value of reviews (e.g Fang et al., 2016), and even predicting hotel demand using web traffic data (e.g Yang et al., 2014).

More than only analyzing patterns within the text, researchers have found how to determine the sentiment behind a statement based on speech patterns, statistical patterns, and other methodologies. This is called sentiment analysis, or opinion mining, and a precursor of this method was attempted since decades before (Stone et al., 1966). With sentiment analysis, one could use patterns in the text to determine whether a sentence was being said with a positive opinion, a critical opinion, or even other ranges of emotions, depending on the thoroughness of the algorithm. Examples of sentiment analysis include ranking products through online reviews (e.g Liu et al., 2017; Zhang et al., 2011), predicting political poll results through opinions in Twitter (O'Connor et al., 2010), and so on. In the hospitality field, it has been used to classify reviewers' opinions of hotels in online reviews (Kim et al., 2017; Al-Smadi et al., 2018, e.g.).

Because the methodology for finding patterns in the data is automatic and statistical in nature, it is both reliable, in that the algorithm will find a pattern by its nature, and unpredictable, in that because it has no intervention from the researchers in making questionnaires it can have different results from anything that the researchers could expect. This is why, much like actual mining, data mining is mostly exploratory in nature. One can never be sure that something can be found, but we can make predictions and estimates about where to find knowledge, and what kind of knowledge can be uncovered.

In this study, we can predict that several things might occur. Our data could show satisfaction and dissatisfaction factors that are universal and it could also find strictly cultural preferences, but we expect that both of these options will present themselves. We can also assert that, if the previous literature is correct in their findings, we could arrive to very similar results, albeit using a database several orders of magnitude larger. We can also expect that, because of the lack of questionnaire design and freedom of users to record their pleasures and grievances, it is possible that we will discover patterns that were previously unnoticed by researchers.

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 positive and negative opinions in a review using Shannon's entropy to extract keywords from their vocabulary. These positive and negative keywords allow us to perform a Support Vector Machine based emotional classification of the reviews in large quantities,

[1] Ctrip: `www.ctrip.com/`
[2] TripAdvisor: `www.tripadvisor.com/`

saving time and resources for the researchers. After classifying the sentences in the extracted reviews as emotionally positive or negative with an optimized SVM, we also observed their weight values in the machine, and 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).

4.1. Data collection

In the data collection stage for Chinese reviews in *Ctrip* a total of 5938 review pages of hotels in Japan were collected. From these pages, we extracted a total of 44,177 reviews, which were comprised of 572,218 separate sentences.

In the TripAdvisor data collection, we collected data from 21,154 different hotels. In total, we collected 295,931 reviews in English, which we then separated into 2,697,086 sentences using the *gensim* python library.

4.2. Text processing

Because Chinese text doesn't have spaces, in order to parse Chinese words on their own we used the Stanford Word Segmenter (Chang et al., 2008) program developed by the Stanford NLP Group[3]. In the case of texts in English, however, only using spaces is not enough to correctly collect concepts. Because

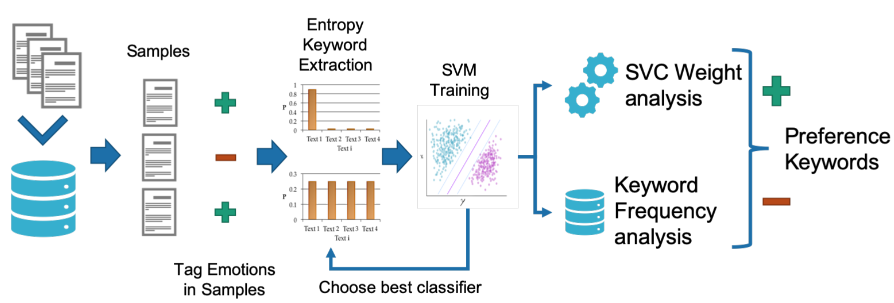


Figure 1: Overview of the methodology.

[3]The Natural Language Processing Group at Stanford University

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.

4.3. Sentiment analysis

The sentiment analysis was performed using the methodology described in Alemán Carreón et al. (2018). A group of keywords is determined by a calculation and comparison of Shannon's entropy (Shannon, 1948) between two classes, and then the keywords are used in a Support Vector Classifier (Cortes & Vapnik, 1995), optimizing the entropy comparison values to select the best performing classifier. The selected classifier's feature keywords would then clearly represent the user preferences leading to positive and negative emotions.

Shannon's entropy, in the field of Information Theory, is defined to be the expected value of the information content in a signal. 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 will have a high entropy value for that set of documents. Opposite to this, a word appearing in only one document will have an entropy value of zero. We show this concept in Figure 2.

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

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

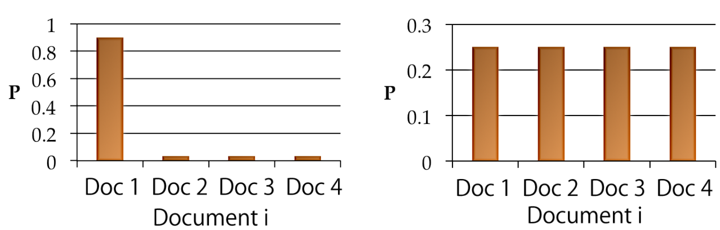
To apply this logic, we retrieved 50 reviews as a sample of our corpus and with the collaboration of a group of 5 Chinese students, which were split into 159 sentences. We then tagged each sentence as the classes positive or negative depending on the emotion that the text conveyed, then calculated the entropy values for each word in relation to the set of sentences from each class. In the case of English reviews, 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. Examples of positive and negative sentences in English and Chinese are shown in the Appendix, in Table A.6.

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

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

The mutually independent coefficients α and α' were tested from 1.25 to 6 in intervals of 0.25. The result was 40 lists for each language, 20 for each emotional class. We trained a different Support Vector Classifier with each of the lists, and we chose the best performing lists for each emotional class and then combined the successful lists to train our best performing classifier. This resulted in a best performing positive emotion classifier (positive/non-positive) for each language.

For the performance tests, we used a 5-fold cross-validation (Kohavi, 1995) process in the Chinese reviews, and a 10-fold cross-validation process in the



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

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

English reviews case, in which we calculated their F-measure (Powers, 2011) means and standard deviations. The number of k-folds was decided from the sample size. Table B.7 shows the lists that had the best performance results in the case of Chinese text and Table B.8 shows the best performance results in English texts in the Appendix.

We then used these best performing classifiers in the rest of the respective data to label positive emotion sentences in a binary manner. The sentences not belonging to the positive emotion class were considered to belong in the negative emotion class. This resulted in 506,452 positive Chinese sentences, 65,766 negative Chinese sentences, 1,288,098 positive English sentences and 1,408,988 negative English sentences.

4.4. SVM weight analysis

During the SVM 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. Below we show the formula for the weight vector (5).

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{5}$$

4.5. Rank-biased Overlap

In order to compare the similarity between two ranked lists, most cases would call to action a statistical measure such as Kendall's τ. However, since the lists are not necessarily of the same length, or have necessarily the same contents, and are top-weighted (where the first rank is the most important to know, with less and less importance as the ranks continue), it is necessary to use another measure of similarity. Webber et al. (2010) proposed a Rank-biased Overlap measurement, which takes all of these factors into consideration. Because our Chinese keywords (their translations, at least) don't match our English keywords one to one, it is necessary to use this method. Webber's RBO produces 3 measurements: a minimum RBO, a residual RBO (from which one can know the maximum RBO), and an extrapolated RBO (where the list is assumed to continue in the same pattern of similarity towards infinity). The formula for the extrapolated RBO is shown in (6), Where S and T are listings, d is their depth, k is their evaluation depth, p is a parameter that controls the top-weightedness of the lists (or it could be thought as $1-p$ being the probability to stop looking at the next item in the list), and X being the overlap between the lists. The complete process is described at length by Webber et al. (2010).

$$RBO_{EXT}(S,T,p,k) = \frac{X_k}{k} \cdot p^k + \frac{1-p}{p} \sum_{d=1}^{k} \frac{X_d}{d} \cdot p^d \tag{6}$$

5. Data Analysis

5.1. Frequent keywords and their SVM weight values

In order to understand the preferences of Chinese speaking tourists and English speaking tourists when lodging in Japan, we study both the frequency of the words they use in relation to the number of total reviews, and their weight in the SVM classifiers. In order to know the relevance of a keyword as a preference for each group, we observed the frequencies of each entropy based keyword in our complete data set and their SVM weight value. The frequency of the keywords in the database shows the level of priority it has for customers, and the weight value allows us to observe the sentiment it relates to by its positive or negative sign. We ranked the keywords by frequency, and use their SVM weight for analyzing the related sentiment (positive weight means positive sentiment, and negative weight signifies negative sentiment).

We observed the top 10 words with the highest frequencies for keywords that were linked by entropy to satisfaction and dissatisfaction in emotionally positive and negative statements to study and quantitatively rank the needs of Chinese customers, as shown in the Tables 1 and 2 (however, the latter does not have more than 7 keywords available); and for English speaking customers, as shown in Tables 3 and 4.

There were many more keywords than shown in most cases, however, and some showed to be high in weight but low in frequency. This could mean they were useful for classification (the emotional reaction is more extreme) but aren't as important a preference for users (there aren't many cases in that the keyword was applicable). In the appendix, Table C.9 shows some keywords that have a relatively high weight value for both positive and negative extremes, and their translations in the relevant context. In Table C.10 we show keywords for the English classifier with high weight values as well.

Table 1: Top 10 frequently used positive 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	8694	0.248
周边	periphery	8456	0.495
近	close	8372	0.028
交通	transportation	8264	0.586
附近	nearby	6619	0.495
地铁	subway	5386	0.180

5.2. Rank-biased Overlap

In order to calculate the similarity between the ranked lists shown in Tables 1 and 3, and the lists in Tables 2 and 4, we calculated the extrapolated Rank-biased Overlap between the English keyword lists and the English translation of the Chinese keyword lists at different cutoff points. In addition, we prepared and

Table 2: Frequently used negative 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
陈旧	old-fashioned	698	-0.000
距离	distance	349	0
老	old	311	0
华人	Chinese person	16	-0.238

Table 3: Top 10 positive English keywords in satisfaction sentences.

Word	Frequency	SVC Weight
staff	138677	0.537
clean	105971	1.886
location	103151	0.842
helpful	63558	1.999
comfortable	62793	1.724
friendly	57307	1.199
recommend	51433	1.158
train	45148	-0.000
free	42084	0.734
subway	38354	1.951

modified the lists so that words that are similar but would not overlap otherwise would overlap in this analysis. For example, in the English keyword lists, the word is 'pricey' while in the Chinese keywords it is translated as 'price', and as such it was changed to 'pricey' to match the English keywords. The results of this are shown in Table 5. In the case of Chinese dissatisfaction keywords, which is only comprised of 7 keywords, the list remains complete in all cutoff points bigger than 7, and then it is cut at the cutoff point of 5. Since the RBO measure can work regardless of the length of each list, this is not a problem in its calculation.

6. Results

6.1. Chinese tourists' satisfaction and dissatisfaction keywords

Analyzing satisfaction keywords of Chinese hotel reviews from Table 1, 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, the cleanliness of Japan's environment, streets without litter, potable water and their culture of respecting spaces. Our results also indicate that closeness of the hotel to scenic places is highly preferred by Chinese customers. Lower in the list of priorities, other positive factors that come into play when Chinese tourists choose a hotel to stay is the location in relation to public transport availability (such as the subway); and as mentioned before, environments, such as gardens or parks nearby; and services nearby, like places to go shopping.

One key component we found in Chinese customer preferences 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

Table 4: High negative weight English keywords in dissatisfaction sentences.

Word	**Frequency**	**SVC Weight**
pricey	3809	-1.614
carpet	3683	-0.507
slow	3177	-1.281
dirty	2943	-1.275
uncomfortable	2942	-2.423
stain	2787	-1.886
cigarette	2468	-0.435
curtain	2032	-0.224
paper	2029	-0.503
renovation	1898	-0.548

Table 5: RBO results between Chinese and English ranked keyword lists, $p = 0.9$.

List cutoff	**Emotion**	RBO_{EXT}
None	Satisfaction	0.222
	Dissatisfaction	0.299
Top 20	Satisfaction	0.221
	Dissatisfaction	0.293
Top 10	Satisfaction	0.209
	Dissatisfaction	0.285
Top 5	Satisfaction	0.221
	Dissatisfaction	0.321

inclusion in the hotel commodities, was very frequently used in positive texts compared to other food-related words.

Regarding dissatisfaction keywords in Table 2, we found that the most frequently criticized aspect of Japanese hotels was the price. 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. Another dissatisfaction factor is the word 'old', which can be referring to the age of the building, or the general design and look of the place being old-fashioned.

We can assert that Chinese customers value the room quality over transportation availability, that they are interested in included breakfast with the hotel stay, that they are concerned with value for money and the availability of Chinese language translations to have an easier time in the accommodation.

6.2. English speaking tourists' satisfaction and dissatisfaction keywords

The most important satisfaction keyword in English reviews (see Table 3) is 'staff', while lower in the list we can also observe 'helpful' and 'friendly', possibly referring to the staff as well, and since Japan is famous for it's customer service culture, this is not unexpected. Next on the list we can observe they value cleanliness, but a few more items in the list are regarding location of the hotel, and possibly the availability of nearby transportation, such as the subway or train. We can also observe that the word 'free' is present there, which after observing a few examples in the database, we concluded that it relates to the free amenities in a hotel room, such as cosmetics, soaps, coffee, tea, and so on.

The negative keywords of dissatisfied customer reviews (see Table 4) reveal an interesting picture. While the most used keyword is also related to the price of the hotels, most of the keywords relate to the hygiene of the hotel, namely dirty carpets, stains, cigarette smell in the room or curtains, and so on. We can also observe that the word 'renovation' is written lower in the list. Some cities in Japan (e.g. Osaka) are currently going under a large number of renovations, which are also extending to the hotel facilities. Customers staying in places in renovation or near a renovation construction can be expected to wake up to construction noises, have their view obstructed by metal bars, and other unpleasant experiences. Another keyword there is 'slow', which upon inspection of example reviews, we concluded that it reflects the speed of the Internet connection in the hotel rooms.

We can assert that English speaking tourists value staff friendliness and location convenience in relation to transport, but are concerned about any kind of decline in room quality both visually and regarding the smell and air quality, being quick to judge any sort of remains of cigarette smell.

6.3. Comparison of Chinese and English speaking tourists' preferences

The extrapolated Rank-biased Overlap (shown in Table 5) ranged from 0.20 to 0.22 in satisfaction lists at different cutoff points, and from 0.28 to 0.32 in the dissatisfaction lists. This means that, while there is some similarity, the preferences are fundamentally different if we consider them as top-weighted, that is, the first elements are the most important in the lists, and therefore in their similarity measurement as well.

From observation however, we can assert that both Chinese and English speaking tourists in Japan have different priorities, but consider the location of the hotel and the availability of transport nearby, such as subway or trains, as a secondary but still important point in their satisfaction of a hotel. The Chinese customers are primarily satisfied with the room quality in spaciousness and cleanliness, while the English speaking customers are easily upset by any lack of cleanliness and smoke smell from cigarettes. English speaking tourists on the other hand value staff friendliness over room quality when considering their satisfaction.

We also can observe some keywords that aren't considered by their counterparts. For example, Chinese tourists are very satisfied with breakfast inclusion, while English speaking customers are satisfied with free amenities. On the other hand, English speaking customers mentioned tobacco smell in many reviews, while it wasn't statistically identified as a problem at all for their Chinese counterparts.

7. Discussion

7.1.

Acknowledgements

During our research, apart from the group of Chinese students that collaborated with the training data and emotion tagging process, we received the commentary and discussion necessary to understand certain cultural aspects that could influence the interpretation of the data, from empirical and anecdotal Chinese sources, as well as collaboration in understanding the meaning behind the words we extracted by our dear colleagues, whom we would like to show gratitude to, Mr. Zhou Liangyuan, Ms. Fan Min and Ms. Eerdengqiqige.

Funding: This work was supported by the Japan Construction Information Center Foundation (JACIC).

Declarations of interest: none

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.

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.

Bauer, T., Jago, L., & Wise, B. (1993). The changing demand for hotel facilities in the asia pacific region. *International Journal of Hospitality Management*, *12*, 313–322. doi:10.1016/0278-4319(93)90048-E.

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.

Choi, T. Y., & Chu, R. (2000). Levels of satisfaction among asian and western travellers. *International Journal of Quality & Reliability Management*, *17*, 116–132. doi:10.1108/02656710010304537.

Choi, T. Y., & Chu, R. (2001). Determinants of hotel guests' satisfaction and repeat patronage in the hong kong hotel industry. *International Journal of Hospitality Management*, *20*, 277–297. URL: http://www.sciencedirect.com/science/article/pii/S0278431901000068. doi:10.1016/S0278-4319(01)00006-8.

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.

Engel, J., Blackwell, R., & Miniard, P. (1990). *Consumer Behavior (6th edition)*. Hinsdale, Illinois, USA: Dryden Press.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498–506. doi:https://doi.org/10.1016/j.tourman.2015.07.018.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*, 37–37. URL: https://wvvw.aaai.org/ojs/index.php/aimagazine/article/download/1230/1131. doi:10.1609/aimag.v17i3.1230.

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.

Gunaratne, S. A. (2009). Globalization: A Non-Western Perspective: The Bias of Social Science/Communication Oligopoly. *Communication, Culture and Critique*, *2*, 60–82. URL: https://doi.org/10.1111/j.1753-9137.2008.01029.x. doi:10.1111/j.1753-9137.2008.01029.x. arXiv:http://oup.prod.sis.lan/ccc/article-pdf/2/1/60/21366916/jcccrit0060.pdf.

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

Hogan, R. T., & Emler, N. P. (1978). The biases in contemporary social psychology. *Social Research*, (pp. 478–534).

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.

Hunt, J. D. (1975). Image as a factor in tourism development. *Journal of travel research*, *13*, 1–7.

Japan National Tourism Organization (2019). *Nationality / Monthly Foreign visitors to Japan (2003-2019)*. Technical Report Japan National Tourism Organization. URL: https://www.jnto.go.jp/jpn/statistics/since2003_visitor_arrivals.pdf (in Japanese).

Jones, D. (2010). A weird view of human nature skews psychologists' studies. *Science*, *328*, 1627–1627. URL: https://science.sciencemag.

org/content/328/5986/1627. doi:10.1126/science.328.5986.1627. arXiv:https://science.sciencemag.org/content/328/5986/1627.full.pdf.

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, C., & Lee, S. (2000). Understanding the cultural differences in tourist motivation between anglo-american and japanese tourists. *Journal of Travel & Tourism Marketing*, *9*, 153–170. doi:10.1300/J073v09n01_09.

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.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, .

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.

Nahm, U. Y., & Mooney, R. J. (2002). Text mining with information extraction. In *Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases* (pp. 60–67). Stanford CA.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31 – 38. URL: http://www.sciencedirect.com/science/article/pii/S0022096517300346. doi:10.1016/j.jecp.2017.04.017.

O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings Of The Fourth International AAAI Conference On Weblogs And Social Media* (p. 122–129). 11(1-2).

Oh, H., & Parks, S. C. (1996). Customer satisfaction and service quality: a critical review of the literature and research implications for the hospitality industry. *Hospitality Research Journal*, *20*, 35–64.

Oliver, R. L. (1981). Measurement and evaluation of satisfaction processes in retail settings. *Journal of retailing*, .

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.

Rajman, M., & Besançon, R. (1998). Text mining-knowledge extraction from unstructured textual data. In *Advances in data science and classification* (pp. 473–480). Springer.

Ren, G., & Hong, T. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information 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. doi:10.1016/j.ecolecon.2014.07.019.

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.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis.*. MIT press.

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. URL: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jtr.726. doi:10.1002/jtr.726.

Uzama, A. (2012). Yokoso! japan: Classifying foreign tourists to japan for market segmentation. *Journal of Hospitality Marketing & Management*, *21*, 132–154. doi:10.1080/19368623.2011.615016.

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.

Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)*, *28*, 20. URL: http://codalism.com/research/papers/wmz10_tois.pdf. doi:10.1145/1852102.1852106.

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.

Zhou, L., Ye, S., Pearce, P. L., & Wu, M.-Y. (2014). Refreshing hotel satisfaction 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. Sentiment analysis training data examples

Table A.6: Examples of positive and negative sentences used for training SVM

Language	Emotion	Sentences
Chinese	Positive	酒店 的 服务 很 好 和 我 住 过 的 所有 日本 酒店 一样 各 种 隐形 服务 非常 厉害 (translated as: ”The service of the hotel is very good. All the services of the Japanese hotels I have stayed in are extremely good.”)
		有 一 个 后门 到 地铁站 非常 近 周边 也 算 方便 酒店 服务 和 卫生 都 很 好 (translated as: ”There is a back door to the subway station very close to it. The surrounding area is also convenient hotel service and health are very good”)
	Negative	酒店 旁边 很 荒凉 连个 便利 店 都 要 走 很远 (translated as: ”The hotel is very bleak, and you have to go very far to go to the nearest convenience store.”)
		唯一 不 足 是 价格 太高 (translated as: ”The only negative is that the price is too high.”)
English	Positive	It was extremely clean, peaceful and the hotel Hosts made us feel super welcome
		Location is very good, close to a main road with a subway station, a bakery, a 7 eleven and a nice restaurant that is not too expensive but serves good food
	Negative	The only downside: our room was labeled ’non-smoking’ but our duvet reeked of smoke.
		A bit pricey though

Appendix B. Entropy keyword extraction experiment results

As explained on section 4.3, we performed experiments with different entropy values to extract keywords from the vocabulary. Then we chose the best performing classification machine based on those keywords as shown in the Tables B.7 and B.8. We also performed experiments to choose the best value of the parameter C used in SVC. C is 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.

Table B.7: Best performing SVC 5-fold cross-validation Chinese text classifiers.

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

Table B.8: Best performing SVC 10-fold cross-validation English text classifiers.

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

Appendix C. Keywords with high SVM weights regardless of frequency

There were many more keywords than shown in Tables 3 and 4. Some showed to be high in weight but low in frequency. This could mean they were useful for classification but aren't as important a preference for users. Table C.9 shows some keywords that have a relatively high weight value for both positive and negative extremes, and their translations in the relevant context. In Table C.10 we show keywords for the English classifier with high weight values as well.

Table C.9: Chinese keywords with high SVM weight values regardless of frequency.

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 C.10: English keywords with high SVM weight values regardless of frequency.

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