NAGAOKA UNIVERSITY OF TECHNOLOGY

DOCTORAL THESIS

Development of service industry evaluation indexes based on data / text mining

データ/テキストマイニングをベースとするサービス産業評価指標の開発

Author:
Elisa Claire ALEMÁN CARREÓN
Student No. 15340089

Supervisor:
Hirofumi NONAKA

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Knowledge Mining Laboratory
Department of Information Science and Control Engineering

February 2, 2022

Declaration of Authorship

I, Elisa Claire ALEMÁN CARREÓN, declare that this thesis titled, "Development of service industry evaluation indexes based on data / text mining" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"There are sadistic scientists who hurry to hunt down errors instead of establishing the truth "

Marie Curie

"The time you enjoy wasting is not wasted time. "

Bertrand Russell

"Study without desire spoils the memory, and it retains nothing that it takes in. "

Leonardo da Vinci

"Science isn't about why, it's about why not! "

Cave Johnson

NAGAOKA UNIVERSITY OF TECHNOLOGY

Abstract

Graduate School of Engineering
Department of Information Science and Control Engineering

Doctor of Philosophy

Development of service industry evaluation indexes based on data / text mining

by Elisa Claire ALEMÁN CARREÓN

Service industries make up for the majority of all businesses in the world. There is an ever growing need of analyzing consumer behavior and its relationship with the services businesses provide, at a larger and larger scale every day. While social studies in consumer psychology are usually performed in small scale using manual methods such as surveys or interviews, there are now tools for analyzing large quantities of data, either numerical or textual, if using natural language processing. However, studies using these tools in Japan are scarce, and as such I saw a need to perform studies to evaluate the relationship between services and their influence on customer behavior, namely purchase intention, actual purchase, and the satisfaction or dissatisfaction after having purchased a service or product. In this thesis I focus on two categories of services. First, I evaluate the effectivity of mere exposure to television adverts in leading consumers to change their purchase behavior. The results of this study suggest that television adverts have little to no effect on the actual purchase behavior, although there is the possibility for other psychological effects. Next, I analyze the effectivity of user-provided numerical data and contrast it to sentiment analysis to consider if customers are representing well their satisfaction through numbers. The results indicate that it is more effective to consider the text users provide to analyze thier satisfaction and sentiment. Last, I expand on this result and analyze the differences in customer satisfaction of Chinese and Western tourists in Japanese hotels using automated methodologies, natural language processing and machine learning. In this thesis, I succeeded in developing appropriate methodologies to evaluate the influence of services on customer behavior using large databases of customer interactions with the services provided, and expanded on the scarce literature analyzing business data from Japan.

Acknowledgements

This thesis could not have been possible without the guidance of my advisor Dr. Hirofumi Nonaka, who guided me and discussed my research with me since my time in Bachelors course up until this Ph.D. course, instilling a love for science that I wouldn't have found otherwise. I'd like to thank my labmates and coauthors who participated in efforts to publish these studies. I would like to show gratitude to Mr. Asahi Hentona and Mr. Hirochika Yamashiro who assisted with the analysis of the television advert data presented by the Nomura Research Institute, Ltd. I would like to show gratitude to Mr. Hugo Alberto Mendoza España, Prof. Toru Hiraoka, and Dr. Masaharu Hirota, who assisted me in my analyses of tourists in Japan. I would also like to show gratitude to Ms. Min Fan, Mr. Liangyuan Zhou, and Ms. Eerdengqiqige who gave us the irreplaceable commentary and discussion necessary to understand particular cultural aspects about the Chinese language and Chinese tourists that could influence the data's interpretation. Finally I'd like to thank Prof. Takashi Yukawa , Dr. Tessai Hayama, Dr. Yuta Nishiyama, and Dr. Tomohiro Shirakawa for their contributions towards the complete and final version of my thesis through their examination.

Contents

List of Figures

List of Tables

List of Abbreviations

API	**A**pplication **P**rogramming **I**nterface
GAN	**G**enerative **A**dversarial **N**etwork
MIC	**M**aximal **I**nformation **C**oefficient
NLP	**N**atural **L**anguage **P**rocessing
POS	**P**art **O**f **S**peech
SVM	**S**upport **V**ector **M**achine
SVC	**S**upport **V**ector **C**lassifier
TV	**T**ele**v**ision
U.K.	**U**nited **K**ingdom
URL	**U**niform **R**esource **L**ocator
U.S.A.	**U**nited **S**tates of **A**merica

Dedicated to my advisor and friends who supported me through all my struggles and mental health issues during this Ph.D. course, without whom I wouldn't be where I am now.

Chapter 1

Introduction

In today's post-industrial society, the service industry has become a majority of all industries in the world, with 70 - 80% of employment as part of the service industry.(United Nations Conference on Trade and Development, 2019) In Japan, specifically, 72.3% of the Gross Domestic Product (GDP) and 71% of employment were part of the service sector in 2015 (Japan Statistics Bureau and Communications, 2019). With a large industry comes a large customer base, with many differences in behavior that managers need to understand and analyze to better cater their services and make their businesses thrive. This has led to an increase in studies in the psychology, management and information retrieval fields to analyze customer behavior. However, there are not many studies using data from national and international customers using services in Japan. Moreover, many of the studies performed on Japanese data use traditional methodologies such as surveys and interviews for customers, which are time-consuming. Studies using the better alternative of information retrieval, big data analysis, machine learning, and natural language processing have also increased, albeit focusing on countries in the West. In fact, recent studies on social sciences 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 and Emler, 1978). Now, with the current wide usage of Web 2.0 and availability of big data analysis tools, and a better understanding of natural language processing, it has become possible to do these analyses for the Japanese service industry on a large scale. It is because of this that my thesis takes a focus on Japan when researching tools for evaluating the performance of services.

The service sector is defined as industries not directly concerned with the production of physical goods. This includes delivering physical goods to customers, such as in retail trade, promotional services such as adverts, lodging services, transportation services, among others. Contrasting with the production sectors, businesses in the service industry must observe the behavior of their consumers and attempt to positively influence it through improvements on service and promotion so that they will have engaged, interested and satisfied customers with the intention to either continue their purchases or recommend the services to others.

In social sciences and psychology, consumer behavior is studied in detail, dividing the different aspects and steps in which a person will interact with the service. For example, first, the customer must be aware of the service. Many consumer behavior analysts believe mere exposure (Zajonc, 1968) will affect consumer choice (Janiszewski, 1993). Others study brand recognition and perception fluency (Fang, Singh, and Ahluwalia, 2007), perception of the product or service (Gmuer, Siegrist, and Dohle, 2015), purchase intention (Armstrong, Morwitz, and Kumar, 2000; Morwitz, Steckel, and Gupta, 2007), the steps before the customer purchases a product or service. Then, many will also study the steps after consumption, namely consumer satisfaction (Hunt, 1975; Oliver, 1981; Oh and Parks, 1996), as well as how cultural aspects of the customer could affect these behaviors (Engel, Blackwell, and Miniard, 1990).

Understanding these has been commonly done with surveys and psychological studies of a small scale, with very localized and biased samples. My thesis focuses on these behaviors while using big sources of data, either collected by large research institutes like the Nomura Research Institute, Ltd., as well as the large fountain of data that is available online thanks to the use of Web 2.0. I used tools such as machine learning algorithms, text mining, natural language processing and statistical tests in order to handle large amounts of data without having to interact with any particular customer individually. This will help me develop tools for evaluating the performance of services in influencing customer behavior at a large and international scale.

With that purpose, my study follows the steps and behaviors in the consumption process mentioned above: the phases immediately before and after the experience of consuming the service: purchase intention and purchase decision, and satisfaction and dissatisfaction. My thesis focuses on two big sectors of the service industry in Japan for each of these stages. The structure of this focus during each phase of the consumption cycle is shown in Figure 1.1.

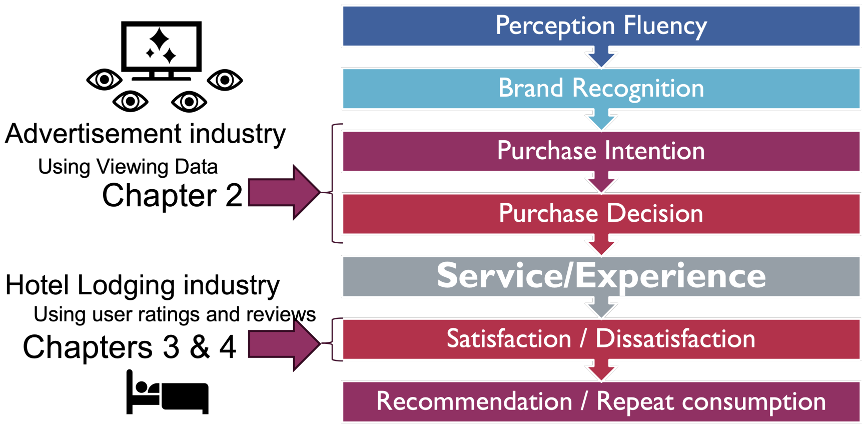


FIGURE 1.1: The consumption process and overview of the analysis chapter structure in this thesis.

First, since thousands of billions of Japanese yen are spent on TV advertisements each year [1], which are a service of promotion, I focus on customer behavior prior to experiencing the service and its relationship with adverts in Chapter 2. In Chapter 2 I focus on whether or not the introduction of mere exposure to adverts will influence the prediction accuracy of our models compared to cases where they were fed other data. Next, based on the fact that inbound international tourism has been increasingly affecting the Japanese economy (Jones et al., 2009), I focus on tourist behavior based on online reviews in Chapters 3 and 4. In these chapters, we explore the consumer behavior after the service experience, satisfaction and dissatisfaction, and which factors of the service contributed to that. In Chapters 3 and 4, my experiment is based on an entropy-based keyword extraction method for sentiment analysis and classification and natural language processing tools to examine the nature of the keywords and their relationship to the service provided.

[1]Dentsu, inc. 2017 Advertising Expenditures in Japan. Retrieved on May 2018 from http://www.dentsu.com/knowledgeanddata/ad_expenditures/pdf/expenditures_2017.pdf

Based on these studies, my research attempts to answer a question that's relevant to the current world of the large scale service industry:

Research Question 1.1: *How can we study consumer behavior at a large scale and evaluate the influence the services have on it?*

In my study, there's focus on the influence that services and their business strategies have on consumer behavior while understanding that the consumer behavior will, in turn, be a helpful tool to modify and adjust the services and business strategies to better fit those customers, completing a cycle of service performance. We will see this process examined and this question answered in different manners and methodologies across the following chapters. My studies will focus on the novelty of the exploration of these topics of consumer behavior as a chain of influences using large databases and contemporary methods of information science. This is accomplished both in each chapter separately and on this thesis as a whole. My study as a whole performs a novel exploration of the chain of consumer behavior before and after consumption and its relationship with the services and business strategies, using data from Japan that hasn't been analyzed in this way by other studies in the past. My thesis is novel in that it analyzes consumer behavior in Japan in different stages of consumption and makes a holistic analysis of the relationship between services and consumer behavior.

Chapter 2

Evaluating the performance of TV adverts' influence on purchase behavior

2.1 Introduction

It is generally thought that in order for companies to increase sales, they must somehow increase the purchase intention of their potential customers (Armstrong, Morwitz, and Kumar, 2000; Morwitz, Steckel, and Gupta, 2007). Historically this has been approached through many channels, but since the successful introduction of the television to the general public, it has been largely attempted via television commercial advertisements, and many companies invest heavily on these efforts. However, most studies to prove the effectiveness of these advertisements have been conducted on small sample groups, usually introducing a customer to a commercial advertisement and measuring their intentions to purchase a product before and after watching the advertisement with a survey (e.g. Khuong and Nguyen, 2015). Studies on the predictability of purchase behavior from purchase intention data have pointed out that many of these analyses have very different results (Morwitz, Steckel, and Gupta, 2007; Sun and Morwitz, 2010; Newberry, Klemz, and Boshoff, 2003), presumably because of small and non-representative samples, and controlled environments that do not reflect reality.

With the advent of Big Data and new methodologies in the field of information technology, there is a new and improved lens for advertisement research on the effects it can have on people outside controlled environments; however, its focus is mostly on similarly new advertisement online and in social media (Shareef et al., 2018; Gonzalez Camacho and Alves-Souza, 2018; Ramaboa and Fish, 2018; Wu et al., 2015), leaving behind the study of more traditional advertisement which has not declined in use since the increase of online advertisement. In response to this lack of current research in the field of television advertisement, I propose a machine learning approach to this problem, with a large database of the household television usage timelines of surveyed individuals and their answers regarding recent purchase intentions and purchase decision recalls at two points in time separated by 3 months, provided by the Nomura Research Institute, Ltd.

Now, observing the lack of large scale studies using actual populations that reflect reality following the traditional train of thought of the effects of mere exposure (Zajonc, 1968), the following question is posed:

Research Question 2.1: *Does mere exposure to advertising directly influence purchase intentions and purchase decisions in consumer behavior when observed in real populations?*

I propose collecting the accumulated number of seconds that a user has viewed a commercial advert related to a certain product and observe its effects on the users. With this data I propose training a number of models to predict the purchase intention and purchase recall of users based on the amount of accumulated seconds of being exposed to the advertisement for the related product in the survey, and then compare it to models that use demographic data of the users as a control. I propose to do this by unit of product, to observe the difference in marketing success from product to product, and by unit of user, to observe the rate of population that was potentially influenced by advertisements. This introduces both granularity, as I are using precise television viewing time and observing effects over time, and the potential of generalizing my prediction model to unknown new users or products after training.

2.1.1 Research objective

The objective of this experiment is to develop an updated methodology and a larger scale database to evaluate the performance of television adverts based on their mere exposure effect and perceptual fluency effect on purchasing behavior. For a long time, psychology based studies have been widely performed on small groups of people in very controlled environments that do not reflect customers in real life accurately, and they have been traditionally thought effective without criticism. I aim to measure the predictability in purchase behavior based on the time spent exposed to adverts of specific products in household televisions during the duration of 3 months, then compare it to the predictability in purchase behavior when using demographic data to provide a clearer answer to whether the heavy investment into TV advertising is actually having an effect on customers to purchase more. As control, I will also measure the predictability in purchase behavior based on demographic data and combining the two sources of data. In the case that the predictability is high enough compared to models that don't include exposure time, this methodology could be used as a measure for future sales. On the other hand, a low predictability in comparison to the control would create doubts on whether the mere exposure to advertisements on television is being effective.

2.2 Theoretical background

In previous research, there have been attempts to analyze the effects of adverts via mere exposure (Hekkert, Thurgood, and Whitfield, 2013), and many studies have replicated the original experiment by Zajonc unrelated to adverts in the field of psychology (Huang and Hsieh, 2013; Dechêne et al., 2009). Now, in addition to the focus on the mere exposure effect, there have been attempts to measure the effects of advertisements on brand recognition and perception fluency (Fang, Singh, and Ahluwalia, 2007), as well as its effects on the perception of the product (Gmuer, Siegrist, and Dohle, 2015). Fluency is defined as the level of ease or difficulty with which external information is processed (Schwarz, 2004). Previously it has been proven that it can produce bias, and it has been shown to affect the judgement of truth (Silva, Garcia-Marques, and Reber, 2017). For a long time, the perceptual fluency model has stated that repeated exposure leads to a more readily accessibility of the target brand in memory, which in turn must have an effect on the ability to recognize a brand in the future (e.g. Jacoby and Dallas, 1981). Most of the older research had arrived to a consensus that there is a positive influence (Reber, Winkielman, and

Schwarz, 1998; Seamon et al., 1995). More recent research, however, explores further whether these effects in memory are strictly related to positive emotional judgment on the brands or if they can also imply negative judgements based on the main objective of a product (Lee and Labroo, 2004).

Research of the direct effects of television advertisement has also been attempted. One study focuses on child obesity by using weight measurements (Boyland and Halford, 2013). An even more direct approach has been made in another study which has used brain imaging in order to explore the short-term and long-term memory effects of TV commercials (Rossiter et al., 2001). It should be noted that, as is to be expected in a brain imaging experiment, the participants observed the advert directly and more consciously than in mere exposure experiments.

Now, two of the main issues with these studies and others in television advertisement effects are that not only is the size of the samples in these experiments questionably small, but the environment is limited in that it becomes highly controlled. Control to this level does not reflect the reality of customers watching daily television in their homes and making purchases anymore, and the observations environment itself could affect the results. Controlled situations such as these are suitable to study the perception fluency and brand recognition stages of the consumption cycle explained in Chapter 1 since it is not necessary for a brand to be real and controlled designs can be studied without bias. However, these controlled environments are unfit to study purchase intention and purchase decision behaviors. These behaviors arise in real-world situations and environments from different background variables and factors that cannot be simulated in a laboratory environment.

In order to solve this limitation, my research is based on data science analysis methodology, such as machine learning algorithms trained from large samples of data from real-world environments, with data collected from users' actual television usage. Currently, other big data analysis on advertising is mostly focused on online advertisements (Wu et al., 2015; Stitelman et al., 2011), where, with the advance of current technology, a user is exposed to adverts placed near to the content they are currently consuming which are specifically targeting their interests (Perlich et al., 2013; Schwartz, Bradlow, and Fader, 2017), catching their attention (which is no longer mere exposure, but direct interaction), or a user is incentivized to watch an advertisement by blocking completely the content they were consuming until the advertisement is finished playing on screen. Most of the research in this area is focused on new ways to create online advertisements in social media (Shareef et al., 2017b) and suggestions or recommendations targeted to a user's interests (e.g. Jansen, Moore, and Carman, 2013; Zhang et al., 2014; Kannan, Ghinea, and Swaminathan, 2015; Choi et al., 2016) reducing the need of mere exposure advertisement while online. In addition to this, some studies have focused on testing the effects of online advertisement on customers (Alalwan, 2018; Lee and Hong, 2016; Shareef et al., 2017a).

While these new technologies made possible the analysis of online advertisement and social media, the focus has shifted and there is no research using these technologies to test the effectiveness of the mere exposure effect based advertisements which are still in use in other traditional means, such as billboards, or as I analyze in this study, television advertisement. This study is unique in that, using data from television advertisement in household environments and not online ads, I apply data science methodology to explore with a larger sample and a household environment, if there is an effect caused by mere exposure advertisement, and to what extent this effect happens. This study is also unique in that comparing the results of prediction models based on exposure time to those using demographic data used in previous literature, I can determine if there is an effect caused by exposure, or if purchase behavior is better decided by other external factors, such as income of each individual and their marital and parental statuses.

2.3 Methodology

As explained above, my approach is to train machine learning models based on the number of seconds of advertisement exposure and demographic data, to predict the effect on the customers purchase decisions measure their predictability. A high predictability based on exposure time would be useful for measuring and predicting sales in any industry. On the other hand, a low predictability in comparison to that of demographic data models would create doubts that the current advertisement based on mere exposure is effective.

My proposed method is explained in detail in the following sections.

2.3.1 Experiment design overview

First I will explain the general design of the experiments. Each experiment consists in creating a prediction model based on a dataset comprised of input features and previously known output labels. After the model is trained, it is able to make predictions of new output labels of unknown data if given new input values. In this study, I created many models by variating the training input features and output labels and compare their results. For the input data, I prepared datasets based on advertisement viewing time and demographic data. For the prediction targets, I prepared datasets for purchase intention and purchase decision behaviors. I also measured the predictability of each purchase behavior target either by unit of product, to observe the difference in marketing success from product to product, and by unit of user, to observe the rate of population that was potentially influenced by advertisements. Finally I utilized 3 different prediction models, Support Vector Machine (Cortes and Vapnik, 1995), XGBoost (Chen and Guestrin, 2016) and Logistic Regression (Walker and Duncan, 1967) in order to compare performance. These variations for each experiment are shown in Table 2.1 and each item will be explained in detail in the following sections. It is important to note that Purchase Intention is described to be used in the input vectors in my experiments in Table 2.1, but this was of course removed for the experiments in which it was the Prediction Target to avoid redundancies.

2.4 Data distribution

In this section I will describe the data I received from the Nomura Research Institute, Ltd., and the distribution and nature of products, adverts and prediction targets.

2.4.1 Surveyed products

The surveys of purchase behavior taken in January 2017 and March 2017 included 200 products, from which only 36 were matched to television adverts during the period between both surveys. Because most of the products are sold only in Japan, a general description of their nature and distribution is explained in Figure 2.1.

TABLE 2.1: Experiment variations.

Experiment Contents	**Variations**
Prediction Model Bases	• Product Based Models • User Based Models
Prediction Targets	• Purchase Intention • Purchase Decision
Input Data Variants	• Advert Viewing Time • Advert Viewing Time, Demographics, (and Purchase Intention) • Demographics (and Purchase Intention)
Prediction Models	• Support Vector Machine • XGBoost • Logistic Regression

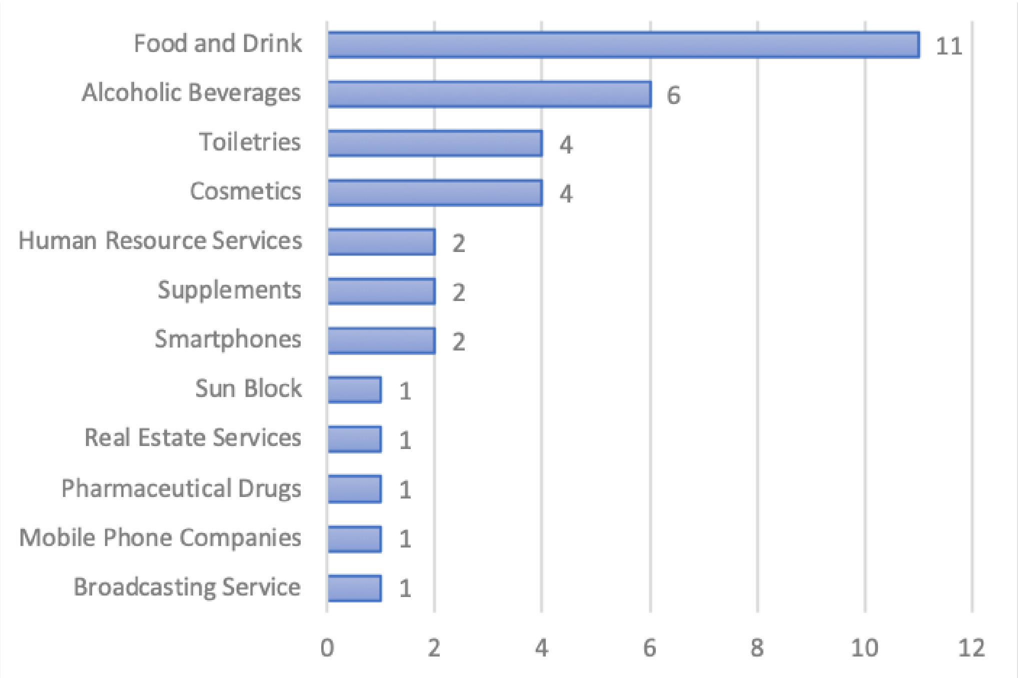


FIGURE 2.1: Products matched with advert viewing data.

2.4.2 Advert exposure and broadcasting data

The data I received from the Nomura Research Institute, Ltd. included the surveyees' household television viewing times and the program that was displayed when television was on. By matching this data with the adverts that were in between breaks from those programs for the products that were surveyed, I obtained the advert exposure time for each user for each product. In this study I explore the possibility of there being some difference in effect depending on the time slot, particularly the Primetime (19:00 to 23:00) time slot. In Figure 2.2 I show the broadcasting time distribution for the programs that displayed these 36 products during the period of time in between the survey in January 2017 and the survey in March 2017. In Figure 2.3 I show the total sum of exposure time in seconds across all users and products for the Primetime and otherwise time slots for each day of the week.

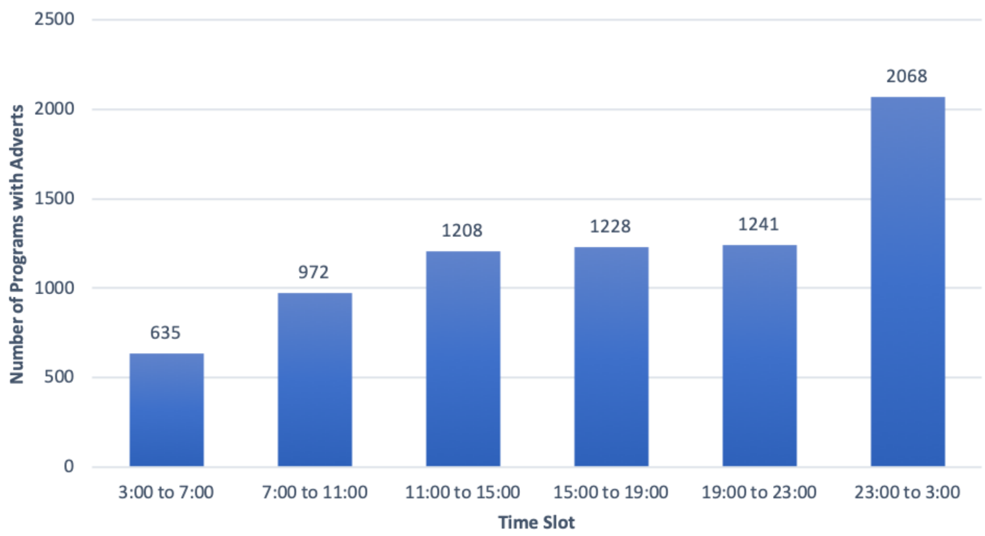


FIGURE 2.2: Programs including adverts broadcast time distribution

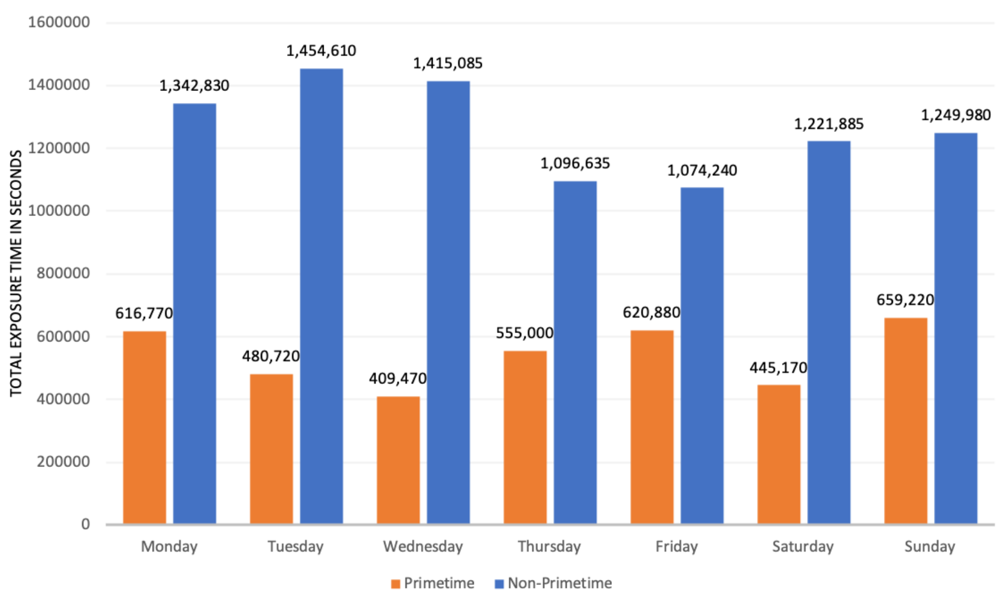


FIGURE 2.3: Advert Exposure Time for all users and products by Weekday and Time Slot

TABLE 2.2: Prediction model bases.

Prediction Model Base	**Description**
Product Based Prediction Models	For each product from 36 available in the survey, data from 3000 users was collected and paired with their labels.
User Based Prediction Models	For each user from 3000 available, data corresponding to the 36 products available in the survey was collected and paired with their labels.

2.4.3 Prediction model bases

In this study, I measured the predictability of each purchase behavior target either by unit of product, to observe the difference in marketing success from product to product, and by unit of user, to observe the rate of population that was potentially influenced by advertisements. After extracting the commercial advert viewing data using these parameters from the 3000 users that answered the survey, which includes purchase behavior questions from 200 products at two different points in time, only 36 products from those in the survey were linked to commercial adverts that were actually viewed by those same users. Thusly, I performed my experiments using the viewing data of 3000 users for these 36 products in the configurations explained before in Table 2.2.

2.4.4 Prediction targets

Purchase Intention and Purchase Decision

From the survey data provided by Nomura Research Institute Ltd., I can examine 3000 customer samples, of which I can extract the Purchase Intention and Purchase Decision answers at two points in time, one in January 2017, and another in March 2017, for 200 different products, 36 of which had advertisemnts in the same time period. Each time, the surveys inquire the customer if they have recently had an intention or desire to purchase a certain product (regardless of action on this desire), which corresponds to Purchase Intention; likewise, it inquires if they have recently had purchased a product, corresponding to the Purchase Decision element. I will inspect the effect of adverts on these two elements of a customer's purchase decisions and observe their change with time on the span of three months.

Prediction target data categorization

In order to explore the different effects commercial adverts may have on the purchase decisions of customers based on their answers from two different points in time, I have labeled each user in regard to each product with 6 categories (from 0 to 5), describing several patterns of behavior. For example, let's examine customers who answered they had purchased a product in January and then not in March, corresponding to category 0, in comparison to customers who purchased the product in March, corresponding to category 4: It is possible that, had category 0 customers were exposed to adverts in greater quantity than other users who still purchased the product and weren't exposed to as many adverts on the span of 3 months, this could mean that the advert was at least not effective, or in a worse scenario, off-putting. On the other hand, if the amount of advert exposure was minimal with category 0 customers and at the same time, customers in category 4 who actually recall having purchased the product in the March survey had been exposed to a large amount of adverts, it would prove to be an effective commercial advert campaign.

Although my approach for analysis is different, the above is a simple example of the importance of this distinction between behavior categories. The six categories for each element are explained in detail in Table 2.3 and Table 2.4.

TABLE 2.3: Category definition for Purchase Decision element.

Category	January Purchase Decision	March Purchase Decision
0	Yes	No
1	No	No
2	No	Yes
3	Yes	Yes
4	Yes/No	Yes
5	Yes/No	No

TABLE 2.4: Category definition for Purchase Intention element.

Category	January Purchase Intention	March Purchase Intention
0	Yes	No
1	No	No
2	No	Yes
3	Yes	Yes
4	Yes/No	Yes
5	Yes/No	No

The surveys included data of purchase intention and purchase decision at the times of January 2017 and March 2017 for 200 products, 36 of which were matched with television advert viewing data. I divided the data in 6 categories (0 to 5) in order to observe the changes in time for these purchase behaviors. The distributions of these categories are shown in Table 2.5. Note that categories 4 and 5, by their nature, are a sum of categories 2 and 3, and 0 and 1 respectively.

2.4.5 Input data

Advert viewing time

I extracted the viewing time for adverts of each product for each customer from the household television viewing data collected and provided by Nomura Research Institute Ltd. Now the data provided tells us if a user had their personal television turned on at the moment of a certain show. Using the information provided of which commercial advert was shown during which television show and how long they lasted, I extracted the number of accumulated seconds a user had the television on for the adverts of each product, and organized them into different weekdays. I called this the Weekday data configuration. For comparison, in a different model, I separated each weekday into two time slots. I did this to further analyze whether the time period regularly described as "Primetime" (19:00 to 23:00) had any different influence than other time slots. I called this the Weekday Time Slot data configuration. I show the detailed features in Table 2.6.

TABLE 2.5: Prediction target categories distribution.

	All products (200)		Advert matched products (36)	
Category	Actual Purchase	Purchase Intention	Actual Purchase	Purchase Intention
0	6%	8%	6%	8%
1	73%	59%	76%	58%
2	8%	9%	7%	8%
3	13%	24%	10%	26%
4	21%	33%	17%	35%
5	79%	67%	83%	65%

TABLE 2.6: Viewing time analysis elements.

Data Configuration	Advert Viewing Time in seconds Data Features
Weekdays	• Monday • Tuesday • Wednesday • Thursday • Friday • Saturday • Sunday
Weekday Time Slots	• Monday Primetime • Monday Non-Primetime • Tuesday Primetime • Tuesday Non-Primetime • Wednesday Primetime • Wednesday Non-Primetime • Thursday Primetime • Thursday Non-Primetime • Friday Primetime • Friday Non-Primetime • Saturday Primetime • Saturday Non-Primetime • Sunday Primetime • Sunday Non-Primetime

Demographic data

In order to perform control experiments, in which the prediction is either aided by, or designed only to be based on external factors from the advert exposure time, I performed experiments using the demographic information of each user collected at the time of the survey by Nomura Research Institute, Ltd. I used the age, sex, marital status, parental status and income bracket reported by each user. The answers and consequently the vector features are shown in detail in Table 2.7.

TABLE 2.7: Demographic data used in input vectors.

Survey Data	Possible Answers
Age	• 18 to 25 years old • 26 to 35 years old • 36 to 45 years old • 46 to 55 years old • 56 or older
Sex	• Male • Female
Marital Status	• Single • Married • Divorced or Widowed
Parental Status	• Parent • Not a Parent
Income Bracket	• Not disclosed • No Income • Under 1,000,000 yen • From 1,000,000 yen to 2,000,000 yen • From 2,000,000 yen to 3,000,000 yen • From 3,000,000 yen to 4,000,000 yen • From 4,000,000 yen to 5,000,000 yen • From 5,000,000 yen to 6,000,000 yen • From 6,000,000 yen to 7,000,000 yen • From 7,000,000 yen to 10,000,000 yen • From 10,000,000 yen to 15,000,000 yen • From 15,000,000 yen to 20,000,000 yen • Over 20,000,000 yen

2.4.6 Prediction models

In this study I chose 3 prediction models: Support Vector Machine, XGBoost and Logistic Regression. SVM and XGBoost are considered well performing supervised machine learning models in the machine learning field considering the size of the data available for this study. Logistic Regression is a statistical model commonly used for binary prediction that is also appropriate for the size of the data. I explain each of those models in more detail in the following sections.

Support Vector Machine

Support Vector Machines (later abbreviated SVM) are supervised machine learning models used in regression and classification problems (**svm**). Supervised learning

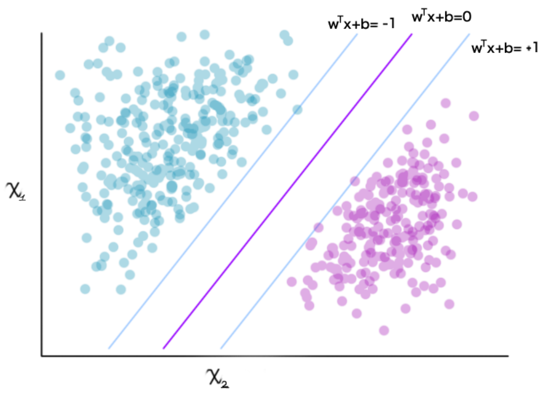


FIGURE 2.4: Two dimensional example of an SVM classification problem

meaning that the model trains on previously labeled data, and establishes a way to match the labels as accurately as possible for new unlabeled data to be analyzed. In a binary classification problem, also called a Support Vector Classifier (SVC), previously established binary labels are matched with a p-dimensional vector of input data. Each column or dimension in the vector expresses a feature in the input data, and each row of the vector is a different data point. After each data point is matched with a label, an SVM uses an algorithm to determine a (p-1)-dimensional hyperplane that separates the p-dimensional space in a way that minimizes error in classification, by maximizing the distance between the hyperplane and the nearest point in either classification. In this study I used the linear kernel for my SVC.

A two-dimensional example is shown in Figure 2.4.

In this study I used the linear kernel for the SVC, defined by the formula (2.1) below, where x is the input vector, w is the weight vector, and b is the bias vector. The dimensions of these vectors are such that $f(x)$ and b are the size of the sample size, and w is the size of the amount of features. The sign of the value of $f(x)$ determines which classification label y is applied, as shown in the formulas (2.2) and (2.3).

$$f(x) = w^\top x + b \tag{2.1}$$

$$f(x) \geq 0 \rightarrow y = +1 \tag{2.2}$$

$$f(x) \leq 0 \rightarrow y = -1 \tag{2.3}$$

The algorithm consists of, starting with a weight and bias vector comprised of zeroes, a randomly placed hyperplane is drawn. Each data point is tested for correct classification, and if the classification fails, the value of w is changed by a value of α as follows (2.4), finally achieving a value represented in the formula (2.5). Finally the distance to the nearest points, the support vectors, in either classification, called the margin, is calculated. During the SVC learning algorithm, each data point classified incorrectly alters the weight vector to correctly classify new data. These changes to the weight vector are greater for features close to the separating hyperplane. These features have stronger changes because they needed to be taken into account to classify with a minimal error. Sequentially, the weight vector can be interpreted as a numerical representation of each feature's effect on each class's classification process. Below I show the formula for the weight vector w (2.5), where x is the training data and each vectorized input x_i in the data is labeled y_i. Each cycle of the algorithm alters the value of w by α to reduce the number of wrong classifications. This equation shows the last value of α after the end of the cycle.

$$w \leftarrow w + \alpha sign(f(x_i))x_i \tag{2.4}$$

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{2.5}$$

This process is repeated so that the margin is maximized and the number of erroneous classifications are minimized.

XGBoost

Originally started as a research project by Tianqi Chen (Chen and Guestrin, 2016), XGBoost is an improved and optimized application of a Gradient Boosting Machine, or GBM, also called gradient tree boosting, or gradient boosted regression tree. A Gradient Boosted Regression Tree (GBRT) works by building an ensemble model from several weak learning machines which are just above random guessing in accuracy, in this case using Decision Trees. The misclassified results from these weak predictions are then weighted and added to a final strong learning machine. This process iteratively optimizes the misclassification cost in a functional gradient descent so that the final learning machine focuses on important factors from the training data for a stronger prediction model.

Logistic regression

The logistic model (Walker and Duncan, 1967) uses a logistic function to model a binary dependent variable. It is a form of regression in which the probability of the dependant variable being one of two possible values (0 or 1) is estimated from the independent variables.

The model in its most basic form is expressed by (2.6), where p is the probability of the binary label being 1, b is the base of the logarithm and exponent,β_0 is the y-inercept, and β_i are the coefficients for each independent variable x_i. However, the base of the logarithm and exponent b is usually e.

$$p = \frac{1}{1 + b^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i)}} \tag{2.6}$$

2.4.7 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 the training data in 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 2.8.

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

TABLE 2.8: Prediction outcomes.

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

I used the F_1 score, which measure is a harmonic mean between precision and recall. Precision, described in formula (2.7), lets us observe the rate of correct positive predictions from all the positive predictions, while Recall, detailed in formula (2.8), observes the rate of correct positive predictions from the total of actual positive data. The F_1 score in formula (2.9) 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.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{2.7}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{2.8}$$

$$F_1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{2.9}$$

2.5 Experiments

2.5.1 Model training

As explained in section 2.3.1, I designed the experiment by training variations of models depending on the input and output values. The combinations of configurations for the input data shown were explained in Table 2.9, and they give us a total of 15,180 possible inputs for experiment variations. The possible targets explained in Table 2.10 give us a total of 12 possible prediction targets. Together, I performed a total of 182,160 experiments per prediction model. Since I used 3 kinds of prediction models (SVM, XGBoost and Logistic Regression), I performed a total of 546,480 experiments in this study.

2.5.2 Experiment parameters

Each prediction model, SVM, XGBoost and the Logistic Regression function can have different parameters when fitting the data to the model. In this study the parameters were chosen broadly to make a general approach (not very specialized) to all the different configurations of the experiment that could take place. Because of the number of experiments explained in section 2.5.1, to choose parameters in a specific manner could unbalance one experiment in favor of the other. As such, I chose simple parameters that can apply to many cases.

The SVM experiments were performed with a linear kernel and a C value of 1. The C parameter allows for misclassification in exchange of a larger margin at small values, and it becomes stricter for larger values, perhaps causing overfitting if large

TABLE 2.9: Input Data configurations.

Prediction Model Base	**Prediction Target**	**Input Data**	**Units**
• Product Based Model • User Based Model	• Actual Purchase • Purchase Intention	• Advert Viewing Time Weekday Configuration • Advert Viewing Time Weekday Time Slot Configuration • Demographics (with Purchase Intention for Purchase Decision) • Advert Viewing Time Weekday Configuration and Demographics (with Purchase Intention for Purchase Decision) • Advert Viewing Time Weekday Time Slot Configuration and Demographics (with Purchase Intention for Purchase Decision)	• 36 products • 3000 users

TABLE 2.10: Possible Prediction Targets.

Prediction Targets	**Number of Categories**
Purchase Decision	6
Purchase Intention	6

enough. The XGBoost experiments were performed with a learning rate of 0.1, a maximum tree depth of 3, and 100 estimators. The Logistic Regression experiments were performed with unit weight per individual sample. A 5-Fold cross validation was performed for all of the models.

2.6 Results

Because of the large number of experiments performed in this study, I analyze the average performances for different variations of the model input and prediction output. In order to compare the performance across different variations, I performed *t*-tests and examined the p-values for statistical significance. The average performance results are detailed in section 2.6.1. The *t*-test comparisons are shown in section 2.6.2.

2.6.1 Prediction score averages

The F_1 scores for the SVM product based model for all 36 products were averaged for each variation of the experiment. The results are shown in Table 2.11. The average F_1 scores for the SVM user based model for all 3000 users are shown in Table 2.12.

Similarly, the XGBoost product based models average F_1 scores are shown in Table 2.13, and the user based models average F_1 scores are shown in Table 2.14.

Lastly, the Logistic Regression product and user based models average results are shown in Tables 2.15 and 2.16.

TABLE 2.11: SVM Product Based Models Average F_1 scores.

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	General Average	**0.293**	**0.292**	**0.489**	**0.495**	**0.497**	**0.413**
	0	0.000	0.000	0.211	0.225	0.218	0.131
	1	0.856	0.852	0.876	0.878	0.875	0.867
	2	0.000	0.000	0.327	0.343	0.343	0.204
	3	0.000	0.000	0.161	0.171	0.171	0.102
	4	0.000	0.000	0.446	0.441	0.441	0.267
	5	0.901	0.900	0.910	0.911	0.911	0.907
Purchase Intention	General Average	**0.252**	**0.248**	**0.273**	**0.275**	**0.276**	**0.265**
	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.570	0.558	0.590	0.590	0.596	0.581
	2	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.115	0.110	0.139	0.139	0.138	0.129
	4	0.166	0.156	0.226	0.226	0.233	0.202
	5	0.662	0.666	0.685	0.685	0.689	0.678
Both Targets	Total Average	**0.273**	**0.270**	**0.381**	**0.385**	**0.386**	**0.339**

2.6.2 Statistical analysis

In this study I performed a series of experiments where I trained different prediction models based on either advert viewing time or demographic data to predict purchase behaviors of purchase decision and purchase intention across 3000 users and 36 products. If I were able to predict purchase behaviors with models based on exposure time more reliably than with models based on demographic data, the obvious strategy for businesses would be to increase the number of adverts. On the other hand, if models based on exposure time had unreliable predictability in contrast to

TABLE 2.12: SVM User Based Models Average F_1 scores.

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.317**	**0.315**	**0.391**	**0.359**	**0.373**	**0.351**
	0	0.055	0.049	0.095	0.085	0.087	0.074
	1	0.745	0.755	0.854	0.789	0.812	0.791
	2	0.074	0.070	0.140	0.114	0.126	0.105
	3	0.076	0.071	0.108	0.107	0.121	0.096
	4	0.140	0.126	0.254	0.221	0.239	0.196
	5	0.812	0.822	0.893	0.840	0.855	0.844
Purchase Intention	**General Average**	**0.297**	**0.289**	**0.242**	**0.296**	**0.290**	**0.283**
	0	0.036	0.032	0.006	0.036	0.033	0.029
	1	0.553	0.553	0.534	0.548	0.553	0.548
	2	0.048	0.043	0.007	0.049	0.042	0.038
	3	0.217	0.195	0.093	0.217	0.198	0.184
	4	0.301	0.279	0.174	0.299	0.280	0.267
	5	0.629	0.634	0.639	0.626	0.632	0.632
Both Targets	**Total Average**	**0.307**	**0.302**	**0.316**	**0.327**	**0.331**	**0.317**

TABLE 2.13: XGBoost Product Based Models Average F_1 scores.

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.293**	**0.294**	**0.307**	**0.309**	**0.308**	**0.302**
	0	0.001	0.000	0.027	0.030	0.029	0.017
	1	0.847	0.849	0.756	0.755	0.752	0.792
	2	0.001	0.000	0.048	0.054	0.052	0.031
	3	0.003	0.003	0.043	0.044	0.045	0.028
	4	0.008	0.010	0.136	0.136	0.133	0.085
	5	0.898	0.900	0.835	0.835	0.833	0.860
Purchase Intention	**General Average**	**0.257**	**0.257**	**0.266**	**0.266**	**0.268**	**0.263**
	0	0.000	0.000	0.000	0.002	0.001	0.001
	1	0.574	0.571	0.573	0.565	0.572	0.571
	2	0.000	0.001	0.001	0.001	0.002	0.001
	3	0.121	0.122	0.136	0.136	0.142	0.131
	4	0.175	0.175	0.211	0.220	0.219	0.200
	5	0.670	0.674	0.673	0.669	0.674	0.672
Both Targets	**Total Average**	**0.275**	**0.275**	**0.287**	**0.287**	**0.288**	**0.282**

TABLE 2.14: XGBoost User Based Models Average F_1 scores.

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	General Average	**0.291**	**0.291**	**0.297**	**0.301**	**0.302**	**0.296**
	0	0.008	0.008	0.018	0.021	0.021	0.015
	1	0.771	0.769	0.752	0.758	0.760	0.762
	2	0.014	0.014	0.030	0.032	0.033	0.025
	3	0.027	0.029	0.042	0.048	0.049	0.039
	4	0.082	0.084	0.119	0.121	0.123	0.106
	5	0.841	0.842	0.825	0.827	0.825	0.832
Purchase Intention	General Average	**0.267**	**0.269**	**0.239**	**0.267**	**0.268**	**0.262**
	0	0.008	0.008	0.001	0.008	0.007	0.006
	1	0.538	0.543	0.533	0.542	0.542	0.540
	2	0.013	0.013	0.001	0.013	0.013	0.010
	3	0.154	0.161	0.085	0.155	0.159	0.143
	4	0.249	0.250	0.174	0.248	0.252	0.235
	5	0.638	0.637	0.643	0.636	0.636	0.638
Both Targets	Total Average	**0.279**	**0.280**	**0.268**	**0.284**	**0.285**	**0.279**

TABLE 2.15: Logistic Regression Product Based Models Average F_1 scores.

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	General Average	**0.293**	**0.293**	**0.500**	**0.513**	**0.508**	**0.421**
	0	0.000	0.000	0.226	0.234	0.237	0.139
	1	0.851	0.853	0.874	0.874	0.875	0.865
	2	0.000	0.000	0.343	0.368	0.355	0.213
	3	0.002	0.000	0.181	0.208	0.195	0.117
	4	0.006	0.004	0.462	0.480	0.475	0.286
	5	0.899	0.901	0.914	0.914	0.914	0.909
Purchase Intention	General Average	**0.259**	**0.256**	**0.289**	**0.294**	**0.292**	**0.278**
	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.575	0.572	0.608	0.611	0.611	0.595
	2	0.000	0.000	0.000	0.001	0.000	0.000
	3	0.125	0.118	0.159	0.170	0.163	0.147
	4	0.184	0.174	0.270	0.280	0.278	0.237
	5	0.671	0.668	0.696	0.703	0.701	0.688
Both Targets	Total Average	**0.276**	**0.274**	**0.394**	**0.404**	**0.400**	**0.350**

TABLE 2.16: Logistic Regression User Based Models Average F_1 scores.

Prediction Target	Category	Advert Viewing Weekday Time Slots	Advert Viewing Weekday Only	Demographics	Advert Viewing Weekday Time Slots and Demographics	Advert Viewing Weekday Only and Demographics	Total Average
Actual Purchase	**General Average**	**0.312**	**0.309**	**0.347**	**0.333**	**0.341**	**0.328**
	0	0.046	0.040	0.037	0.056	0.054	0.047
	1	0.740	0.750	0.846	0.771	0.794	0.780
	2	0.064	0.058	0.070	0.083	0.082	0.071
	3	0.074	0.066	0.064	0.089	0.091	0.077
	4	0.139	0.124	0.175	0.173	0.185	0.159
	5	0.806	0.815	0.888	0.827	0.844	0.836
Purchase Intention	**General Average**	**0.298**	**0.295**	**0.242**	**0.298**	**0.293**	**0.285**
	0	0.037	0.035	0.007	0.038	0.034	0.030
	1	0.555	0.557	0.535	0.554	0.556	0.551
	2	0.052	0.047	0.009	0.052	0.044	0.041
	3	0.220	0.204	0.093	0.220	0.207	0.189
	4	0.298	0.288	0.175	0.300	0.286	0.269
	5	0.628	0.635	0.636	0.628	0.634	0.632
Both Targets	**Total Average**	**0.305**	**0.302**	**0.294**	**0.316**	**0.317**	**0.307**

models based on demographic data, doubts would surface about the effectiveness of the hard investment in television advertising.

In order to analyze the change in predictability of purchase behavior I averaged the results of predictions across different variations of the experiments, detailed in the previous section, and then performed t-tests to observe the difference in performance between sets of results. I established 3 hypotheses to test for, explained below.

Hypothesis 2.1: *Advert viewing time based models perform differently from demographics based models.*

For this hypothesis, I performed a t-test using the results from models that include advert viewing time and the models that only include demographic data. More specifically, I tested the Weekday Time Slot model results against the Demographics models, and the Weekday Only models against the Demographics models. The p-values for each t-test are shown in Table 2.17 for Purchase Decision predictions and in Table 2.18 for Purchase Intention predictions. With these tests, I will examine the changes in predictability against demographic data, which I are using as the control data for my experiments. This will allow us to determine whether the advert viewing time based models are performing better or worse than the demographic models, and therefore conclude whether the advert viewing time is having an effect on customers purchase behavior or if it is decided by external factors.

Hypothesis 2.2: *Demographic and advert viewing based models perform differently from demographic based models.*

For this hypothesis, I performed a t-test between the results from models that include both advert viewing time and demographic data, and the models that only include demographic data. More specifically, I tested the Weekday Time Slot and Demographics model results against the Demographics models, and the Weekday and Demographics models against the Demographics models. The p-values for each t-test are shown in Table 2.19 for Purchase Decision predictions and in Table 2.20 for Purchase Intention predictions. With these tests, I will examine if adding the advert

TABLE 2.17: Hypothesis 1 t-test: p-values for Purchase Decision behavior.

Model	Base	Configuration	Purchase Decision Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	0.000	0.328	0.000	0.001	0.000	0.567
		Weekday Only	0.000	0.284	0.000	0.001	0.000	0.527
	user	Weekday Time Slot	0.000	0.000	0.000	0.000	0.000	0.000
		Weekday Only	0.000	0.000	0.000	0.000	0.000	0.000
XGBoost	product	Weekday Time Slot	0.000	0.005	0.000	0.015	0.000	0.013
		Weekday Only	0.000	0.004	0.000	0.015	0.000	0.011
	user	Weekday Time Slot	0.000	0.004	0.000	0.000	0.000	0.004
		Weekday Only	0.000	0.010	0.000	0.000	0.000	0.003
Logistic Regression	product	Weekday Time Slot	0.000	0.282	0.000	0.000	0.000	0.348
		Weekday Only	0.000	0.327	0.000	0.000	0.000	0.414
	user	Weekday Time Slot	0.012	0.000	0.236	0.032	0.000	0.000
		Weekday Only	0.374	0.000	0.007	0.757	0.000	0.000

TABLE 2.18: Hypothesis 1 t-test: p-values for Purchase Intention behavior.

Model	Base	Configuration	Purchase Intention Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	nan	0.803	nan	0.700	0.405	0.767
		Weekday Only	nan	0.694	nan	0.644	0.329	0.808
	user	Weekday Time Slot	0.000	0.026	0.000	0.000	0.000	0.234
		Weekday Only	0.000	0.028	0.000	0.000	0.000	0.534
XGBoost	product	Weekday Time Slot	0.324	0.983	0.041	0.804	0.598	0.969
		Weekday Only	0.324	0.982	0.203	0.811	0.606	0.992
	user	Weekday Time Slot	0.000	0.533	0.000	0.000	0.000	0.585
		Weekday Only	0.000	0.246	0.000	0.000	0.000	0.509
Logistic Regression	product	Weekday Time Slot	nan	0.673	0.803	0.582	0.222	0.724
		Weekday Only	nan	0.655	0.324	0.514	0.175	0.704
	user	Weekday Time Slot	0.000	0.017	0.000	0.000	0.000	0.364
		Weekday Only	0.000	0.010	0.000	0.000	0.000	0.947

viewing data to the demographic data causes any major changes, to determine if the predictions are being improved, worsened, or if they stay the same regardless of advert viewing.

Hypothesis 2.3: *Advert viewing time based models perform differently from demographic and advert viewing based models.*

For this hypothesis, I performed a t-test bewteen the results from models that include both advert viewing time and demographic data, and the models that only include advert viewing data. More specifically, I tested the Weekday Time Slot and Demographics model results against the Weekday Time Slot models, and the Weekday and Demographics models against the Weekday Only models. The p-values for each t-test are shown in Table 2.21 for Purchase Decision predictions and in Table 2.22 for Purchase Intention predictions. With these tests, I will examine if adding the demographic data to the advert viewing data causes any major changes, to determine if the predictions are being improved, worsened or if they stay the same regardless of demographic variances. By performing this last test, as well as the differences tested by Hypothesis 2.1 and Hypothesis 2.2, I can assume significant differences across all major 3 groups of data.

2.7 Discussion

2.7.1 Influence of TV adverts on Purchase Decision and Purchase Intention

Observing my results across models in the Tables of section 2.6.1, in general, I can observe that SVM models perform relatively better than XGBoost models and Logistic Models, and that the differences and directional change in averages between Advert Viewing Time based models, Demographics based models, and Advert Viewing Time and Demographics based models stay consistent across SVM, XGBoost and Logistic Regression models. That is to say, low predictability in Advert Viewing Time based models compared to Demographic data models stays constant regardless of the changes in performance across prediction techniques.

In Tables 2.11 and 2.12 I can observe this more closely. In general for Purchase Decision behavior, predictions using Advert Viewing Time only have a lower performance than the other models. Specially in categories 2 and 4 of the purchase behavior, I can see that the average predictability rises from 0 or close to 0, to a higher predictability in every case that demographic data is used for positive purchase behavior.

I can confirm this increase is statistically significant by observing the results of Hypothesis 2.1 in Table 2.17. For the most part, excluding negative purchase behavior, the data is significantly different at a 95% confidence level ($p < 0.05$) between models that use advert viewing time as a base for prediction and models that use demographic data as a base for prediction. Moreover, I can confirm that the changes in predictability between models that include both advert viewing time and demographic data, and models that only include demographic data are not statistically significant by observing the results of Hypothesis 2.2 in Table 2.19. In most cases, ($p > 0.05$), proving that the advert viewing time data did not influence the prediction scores significantly, and that whatever correct predictions were made were most likely based on the coefficients and weights of the demographic data. Finally, observing Hypothesis 2.3 in Table 2.21 also confirms the difference and increase of

TABLE 2.19: Hypothesis 2 t-test: p-values for Purchase Decision behavior.

Model	Base	Configuration	Purchase Decision Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	0.923	0.953	0.748	0.777	0.968	0.980
		Weekday Only	0.850	0.936	0.839	0.868	0.942	0.956
	user	Weekday Time Slot	0.056	0.000	0.000	0.837	0.000	0.000
		Weekday Only	0.139	0.000	0.033	0.038	0.078	0.000
XGBoost	product	Weekday Time Slot	0.758	0.973	0.581	0.936	0.991	0.993
		Weekday Only	0.811	0.929	0.703	0.897	0.924	0.956
	user	Weekday Time Slot	0.074	0.318	0.413	0.057	0.625	0.655
		Weekday Only	0.069	0.187	0.173	0.045	0.388	0.993
Logistic Regression	product	Weekday Time Slot	0.894	0.997	0.674	0.627	0.772	0.971
		Weekday Only	0.856	0.955	0.846	0.808	0.831	0.990
	user	Weekday Time Slot	0.000	0.000	0.007	0.000	0.823	0.000
		Weekday Only	0.000	0.000	0.017	0.000	0.187	0.000

TABLE 2.20: Hypothesis 2 t-test: p-values for Purchase Intention behavior.

Model	Base	Configuration	Purchase Intention Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	nan	0.942	nan	0.985	0.930	0.969
		Weekday Only	nan	0.986	nan	0.991	0.950	0.955
	user	Weekday Time Slot	0.000	0.099	0.000	0.000	0.000	0.118
		Weekday Only	0.000	0.029	0.000	0.000	0.000	0.402
XGBoost	product	Weekday Time Slot	0.002	0.924	0.850	0.996	0.900	0.960
		Weekday Only	0.115	0.990	0.364	0.916	0.916	0.986
	user	Weekday Time Slot	0.000	0.290	0.000	0.000	0.000	0.409
		Weekday Only	0.000	0.303	0.000	0.000	0.000	0.400
Logistic Regression	product	Weekday Time Slot	nan	0.957	0.331	0.856	0.884	0.916
		Weekday Only	nan	0.957	0.871	0.943	0.903	0.943
	user	Weekday Time Slot	0.000	0.028	0.000	0.000	0.000	0.313
		Weekday Only	0.000	0.016	0.000	0.000	0.000	0.869

TABLE 2.21: Hypothesis 3 t-test: p-values for Purchase Decision behavior.

Model	Base	Configuration	Purchase Decision Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	0.000	0.354	0.000	0.000	0.000	0.553
		Weekday Only	0.000	0.253	0.000	0.000	0.000	0.488
	user	Weekday Time Slot	0.000	0.000	0.000	0.000	0.000	0.000
		Weekday Only	0.000	0.000	0.000	0.000	0.000	0.000
XGBoost	product	Weekday Time Slot	0.000	0.005	0.000	0.009	0.000	0.013
		Weekday Only	0.000	0.003	0.000	0.013	0.000	0.009
	user	Weekday Time Slot	0.000	0.051	0.000	0.000	0.000	0.011
		Weekday Only	0.000	0.181	0.000	0.000	0.000	0.002
Logistic Regression	product	Weekday Time Slot	0.000	0.287	0.000	0.000	0.000	0.367
		Weekday Only	0.000	0.303	0.000	0.000	0.000	0.400
	user	Weekday Time Slot	0.005	0.000	0.000	0.001	0.000	0.000
		Weekday Only	0.000	0.000	0.000	0.000	0.000	0.000

performance between advert viewing time models and those that combine advert data with demographic data.

The exception to this rule is in category 1, where customers consistently answered "NO" in their purchase recall or purchase intention questions of the survey both in January 2017 and March 2017. Subsequently, this also influences category 5 results. Predictions seem to be high across all models when the customer has a negative purchase behavior. However, the *t*-test results for Hypotheses 2.1 and 2.3 show us that there is not a statistically significant difference between demographic data models and advert viewing time models. Because of these results, the factors that are influencing negative purchase decisions cannot be determined to be either advert viewing time or otherwise.

With these results in mind, it could be said that TV adverts are not a main factor in predicting whether a customer will change their purchasing behavior or not in a positive way, specially their purchase decision behavior. While the research based on the mere exposure effect would suggest otherwise, customers are observed to decide on their purchase without much predictability, except for their demographic data. It could be said that while there might be influence in the customer's knowledge of the brand, the data suggests that the amount of time exposed to TV adverts has no effect in the customers purchase decision behavior.

Other studies, using a controlled environment, have linked mere exposure with bias in consumer choice (Janiszewski, 1993). However, there is a possible explanation for these discrepancies in results. While controlled experiments show the TV adverts to their sample audience directly in most cases, in an uncontrolled environment of a customer's home, the customer is left free to ignore the advert and do something unrelated in the meanwhile (Abernethy, 1991). In the United Kingdom, there is a widely documented phenomenon involving TV advert timing and a surge in electricity caused by the use of electric kettles for preparing tea. This phenomenon is commonly called TV pickup, and has been documented for long (Bunn, 1982; Boait, Rylatt, and Wright, 2007). Similar to these cases, if the customers whose data were actively ignoring the adverts, the sample for training the prediction models would contain noise, altering the results. It stands to reason that without the influence of this active aversion would have on my learning model, it might correctly predict purchasing behavior as expected. However, this is more of a problem with the current TV advertisement model than with the methodology of this study. I will discuss this further in section 2.7.3 of this study.

2.7.2 Influence of TV adverts based on primetime

In my prediction model experiments, I used data from advertisement exposure during different time periods, days of the week and weekends. While I did this in order to observe differences in predictability for different time schedules available to different kinds of customers, especially during primetime television hours, I arrived to similar results for all time data configurations. I did not observe any difference in predictability based on Primetime television watching compared to other time periods. This could be interpreted as there being little influence in time periods and changes in purchasing behavior.

2.7.3 Implications for the TV advert industry

Based on the low results of predictability of purchase behavior by advert exposure, it can be observed that TV adverts have a low probability of achieving their main purpose: to increase sales. As was stated in section 2.7.1 of this study, there could be a large influence on this study's results from customers actively ignoring the adverts although they are being broadcast to their TVs. It is left to further discussion and research if adverts actually have the intended effect on customers when watched properly, or if this effect is not achieved anyway. In (Fang, Singh, and Ahluwalia, 2007) it is proposed that while the mere exposure of banner advertisement increases perceptual fluency, it doesn't have an effect on actual brand recognition compared to the control groups, for example. The existence or absence of influence by perceptual fluency on a customer's purchase decision hasn't been fully explored, but the consensus in the processing fluency model is that perceptual fluency influences brand judgement on some level, although it depends on the concept if the reception is positive or not (Lee and Labroo, 2004). The problem with these studies and the current consensus, as has been said previously in this study, is both that most experiments are done with relatively small sample sizes, and that there is a factor of uncertainty that comes with the physical avoidance of adverts in a customer's household.

With these things in mind, I consider both possibilities: either customers are attentive and the adverts have the expected influence in their short and long term memories in the case of repeated exposures (Rossiter et al., 2001); or the customers are inattentive of the advert and there might be some level of unconscious effect of mere exposure in their perception fluency (Fang, Singh, and Ahluwalia, 2007). I observed however in my results that there is no effect on purchase decision behavior. While it may be true and out of the reach of the data that the customers would have influence in their memory, there was no link observed between the time of advert exposure and the purchase decisions. This raises a concern for the TV advert industry. Regardless of the cause of my results, the main implication of this study is that currently, TV adverts are shown to have little to no effect on changes in purchase decision behavior. While thousands of billions of japanese yen are spent on TV advertisements each year [1], the effects observed in this study are negligible. Because of this, changes are necessary in the current TV advertisement model.

2.8 Limitations

In comparison with previous research regarding this topic, this study presents a much larger database, a sample of 3000 users for 36 different products and the previously unavailable household television viewing data increases the possibilities for studying the effects of advert exposure more realistically. In accordance with this size of data, I used SVM and XGBoost, which are considered well-performing machine learning algorithms in this level of magnitude. However, while I propose using machine learning algorithms as an effective method, I are still limited by the calculation times for each model. Top performing and state of the art models, such as Deep Neural Networks, with their variations and advancements, have been known to be used with similar magnitudes of data or to expand upon it by using GAN (Generative Adversarial Network) (Goodfellow et al., 2014), but their calculation time is far greater. Thus, Neural Networks are more appropriate for single models being

[1]Dentsu, inc. 2017 Advertising Expenditures in Japan. Retrieved on May 2018 from http://www.dentsu.com/knowledgeanddata/ad_expenditures/pdf/expenditures_2017.pdf

trained, instead of a performance comparison of a large array of models as I did in this study.

Another limitation of this study is the nature of the prediction targets collected by survey. While a person can be asked directly in a survey whether they would purchase an item (purchase intention) or if they had already in the recent past (Purchase Decision), research based on online shopping has access to the Purchase Decision data, and to the number of times a person looks at a products description page, or searches terms related to it. Television advertisement research, by its nature, is harder to connect to the actual behavior of the customers and can only be assumed to be equal to their reported behavior. There is also a limitation of the number of questions that a person might answer, and how honestly they might answer them with a survey of this magnitude.

In addition to this, because of the timing of the surveys being 3 months appart between January and March 2017, I can only examine the short-term effect of advertisements, and not the long-term effect across different years of constant advertisement exposure.

Furthermore, much of the data that could be used to inspect this matter further belongs to private institutions and in many cases, is treated as a company secret.

However, with the measurements of short-term effects of advertisement in a field where not much new research is done, my study can start to shed light on problems that could be having a large impact on the costs of many industries.

2.9 Conclusion and future work

In this study I analyzed the ability to predict purchasing behavior, namely Purchase Intention and Purchase Decision, based on the customers' time spent exposed to television adverts using machine learning algorithms, and compared it to the ability to predict the same behavior by using demographic data on its own and in combination with the exposure time data. Based on the low prediction results of Purchase Decision by exposure time models and the relatively high prediction results for demographic based models, as well as a non-significant difference between the demographic models and the combined models, I concluded that advertisement exposure has little to no effect in short-time purchase decision behavior.

I discussed possible influence by deliberate avoidance of advert cuts to prepare food or tea, and while some studies focus on the effect of attentive watching of adverts, other studies focus on the mere exposure effects, which would be achieved despite physical avoidance because of advert audio and simple proximity of the television. Both scenarios are in strong contrast with the results of my study, which shows little to no predictability in purchase behavior. Points left to research in future work are a deeper analysis of the predictable customers, looking for similarities or clusters within this class, as well as using newer and better performing deep learning algorithms when larger datasets are available.

TABLE 2.22: Hypothesis 3 t-test: p-values for Purchase Intention behavior.

Model	Base	Configuration	Purchase Intention Categories					
			0	**1**	**2**	**3**	**4**	**5**
SVM	product	Weekday Time Slot	nan	0.750	nan	0.713	0.359	0.738
		Weekday Only	nan	0.682	nan	0.633	0.300	0.766
	user	Weekday Time Slot	0.874	0.494	0.853	0.934	0.835	0.655
		Weekday Only	0.822	0.991	0.621	0.658	0.806	0.805
XGBoost	product	Weekday Time Slot	0.031	0.910	0.071	0.805	0.514	0.993
		Weekday Only	0.197	0.992	0.051	0.731	0.532	0.994
	user	Weekday Time Slot	0.946	0.634	0.803	0.838	0.885	0.757
		Weekday Only	0.586	0.886	0.941	0.764	0.880	0.841
Logistic Regression	product	Weekday Time Slot	nan	0.634	0.427	0.464	0.169	0.647
		Weekday Only	nan	0.618	0.324	0.468	0.139	0.652
	user	Weekday Time Slot	0.904	0.823	0.934	0.968	0.746	0.905
		Weekday Only	0.556	0.848	0.315	0.708	0.743	0.907

Chapter 3

Is it better to use scores or review texts for evaluating satisfaction?

3.1 Introduction

As the number of Chinese tourists visiting Japan increases, it is important for the hotel industry to conduct market research to analyze the needs of hotel guests. Under these circumstances, grasping needs through questionnaires and interviews has become the center of market research. However, surveys using such questionnaires and interviews have problems in terms of cost and real-time performance. On the other hand, with the spread of the Internet, there are many online reviews. Users will use the opinions of others as reference, and because of the large influence these have on purchase decisions (Vermeulen and Seegers, 2009; Sparks and Browning, 2011), these have also been actively used in the industry. Users' evaluations on many online review sites are numerous, and are divided into comment text as text information and a review score expressed numerically. The review scores are structured numerical information and are easy to use for analysis, so they are being used as evaluation indexes for users' products and services (Xie, Zhang, and Zhang, 2014; Bulchand-Gidumal, Melián-González, and González Lopez-Valcarcel, 2013; Zhou et al., 2014). On the other hand, text reviews are also being analyzed based on natural language processing. For example, research that analyzes sentiment of word-of-mouth using information theory (Amplayo and Song, 2017) as well as forecast of product sales (Fan, Che, and Chen, 2017) or the ranking of products based on sentiment analysisis (Liu, Bi, and Fan, 2017) is being conducted.

In customer trend analysis, what is used as an analysis index is extremely important. As mentioned above, in previous studies, customer behavior was analyzed using sentiment analysis only (Amplayo and Song, 2017; Liu, Bi, and Fan, 2017) and there are many studies that use only the review score points (Xie, Zhang, and Zhang, 2014; Bulchand-Gidumal, Melián-González, and González Lopez-Valcarcel, 2013; Zhou et al., 2014). Therefore, it is necessary to examine the relationship between these review scores and the sentiment analysis of comments. If the relationship between the review score and the sentiment analysis of the document is high, it would be easy and there would be a great merits to do analysis based only on the numerical information or using the numerical information as training data for future sentiment analysis. On the other hand, if the correlation is low, it is necessary to comprehensively evaluate the reviews by using the score and text information together. Alternatively, it is necessary to select a method that more reflects the user's emotions and opinions and use it as an evaluation index. In this way, investigating the relationship between the review score and the sentiment analysis of the comment text is extremely important in the analysis of the review.

Therefore, the next question is posed:

Research Question 3.1: *Is it better to use scores or review texts for evaluating satisfaction?*

In this study, I investigated the relationship between the review scores and the sentiment analysis of the comment text. In this study, I first collected a large number of review documents written about Japanese hotels and their review score from the online hotel review site *Ctrip* for Chinese tourists. Next, I trained an SVM using the entropy-based feature selection method developed by our researchers, and classified documents that express positive emotions and documents that express negative emotions. Since the feature vector characteristic of this emotion classification was constructed by keyword extraction based on entropy, the classification was performed using a language-independent method based on statistics. Furthermore, based on the emotion classification of each sentence in each review, the ratio of sentences expressing satisfaction and the ratio of dissatisfaction sentences to the full review were calculated in order to quantify emotions. Finally, the interrelationship between the review score and the emotional evaluation of the review was analyzed from Spearman's rank correlation coefficient and Kendall's rank correlation coefficient MIC. This will be described in detail below.

3.2 Literature Review

In previous research, targeting product reviews is the mainstream. For example, a study that applied Word Cloud to extract words that are often used by consumers (Hargreaves, 2015), and a market analysis using sentiment analysis of product reviews using an emotion dictionary called HowNet (Zhang et al., 2011). On the other hand, as for quantitative evaluation of the impact of hotel online reviews, there is a study that proves that hotel online reviews have a particular effect on customer motivation (Vermeulen and Seegers, 2009). There are other studies that have shown a relationship between sales and the textual information of reviews that are evaluated (Basuroy, Chatterjee, and Ravid, 2003). In the mentioned studies, the scores were not considered, and the relationship between the scores and the text information was not focused on.

3.3 Methodology

3.3.1 Preprocessing

At the crawling stage, I used the fact that the URL structure is determined by the hotel ID number, and so each hotel page can be automatically loaded. Next, scraping was performed, and the documents, IDs, review score, etc. of each review were acquired using the structure of the HTML code and saved in my database. Morphological analysis was also performed for the review. As a tool for morphological analysis of Chinese, Stanford Word Segmenter (Chang, Galley, and Manning, 2008) provided by The Stanford NLP Group of Stanford University was used.

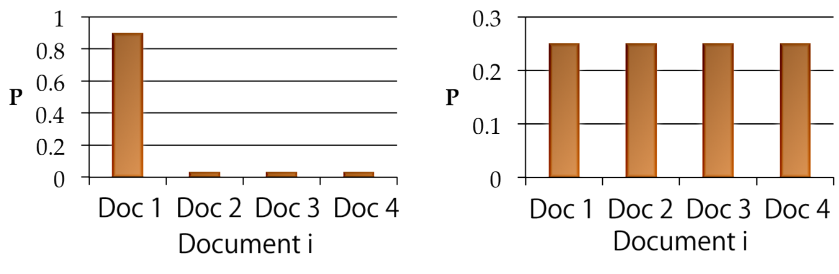
3.3.2 Sentiment Analysis

Samples were extracted from the data collected by online hotel reviews for sentiment analysis, and with the cooperation of three Chinese research students, each sentence in each review was tagged according to whether it expressed satisfaction or dissatisfaction. Manual classification of Positive and Negative was performed, and training data was created. Then, the words and emotion classification tags included in the sample review were trained by SVM, and the emotion classification of all the data in the population was performed. A feature vector characteristic of emotion classification by SVM was constructed by keyword extraction based on entropy, which will be described later. The details will be described below.

3.3.3 Entropy Based Keyword Extraction

In this study, I based the extraction of the keywords that are influenced by the users' emotional judgement on the calculation of an entropy value for each word. Speaking in Information Theory terms, Shannon's Entropy is the expected value of the information content in a signal (Shannon, 1948). Applying this knowledge to the study of words allows us to observe the probability distribution of any given word inside the corpus. For example, a word that keeps reappearing in many different documents will have a high entropy, given that predicting on which document it would appear becomes uncertain. On the contrary, a word that only was used in a single text and not in any other documents in the corpus will be perfectly predictable to only appear in that single document, bearing an entropy of zero. This concept is shown in Fig.3.1.

Based on the meaning of entropy explained above, keywords that will be considered positive will have a large entropy when they appear in many positive documents, and a smaller entropy in negative documents. The same will occur for negative keywords in the opposite documents. In this study I use the entropy values of keywords to perform a classification. First, I tagged a set of documents as positive or negative. Then, for each word j that appears in each document i, I counted the number of times a word appears in positive comments as N_{ijP}, and the number of times a word appears in negative comments as N_{ijN}. Then, as shown in the formulas below, I calculated the probability of each word appearing in each document shown below as P_{ijP} (3.1) and P_{ijN} (3.2).



(A) Entropy close to zero. (B) High entropy.

FIGURE 3.1: Probabilities of a word j being contained in a document i.

$$P_{ijP} = \frac{N_{ijP}}{\sum_{i=1}^{M} N_{ijP}} \tag{3.1}$$

$$P_{ijN} = \frac{N_{ijN}}{\sum_{i=1}^{M} N_{ijN}} \tag{3.2}$$

I then substitute these values in the next formula. I calculated the entropy for each word j in relation to positive documents as H_{Pj} (3.3), and the entropy for each word j in relation to negative texts as H_{Nj} (3.5). That is, as is shown in (3.4) and (3.6), all instances of the summation when the probabilities P_{ijP} or P_{ijN} are zero and the logarithm of these becomes undefined are substituted as zero into (3.3) and (3.5).

$$H_{Pj} = -\sum_{i=1}^{M}[P_{ijP}\log_2 P_{ijP}] \tag{3.3}$$

$$\lim_{P_{ijP}\to 0+} P_{ijP}\log_2 P_{ijP} = 0 \tag{3.4}$$

$$H_{Nj} = -\sum_{i=1}^{M}[P_{ijN}\log_2 P_{ijN}] \tag{3.5}$$

$$\lim_{P_{ijN}\to 0+} P_{ijN}\log_2 P_{ijN} = 0 \tag{3.6}$$

After calculating the entropies for each word, I adjusted for their α value by testing for the highest F-value. A positive keyword is determined when (3.7) is true, and likewise, a negative keyword is determined when (3.8) is true for the best performing α value, using these keywords as elements for training an SVM (Cortes and Vapnik, 1995). The performance was determined using a k-fold cross validation calculating the best F_1 value (Powers, 2011).

$$H_{Pj} > \alpha H_{Nj} \tag{3.7}$$

$$H_{Nj} > \alpha' H_{Pj} \tag{3.8}$$

3.3.4 Correlation Analysis

The following method was experimentally used to measure the correlation of the ratio of positive sentiment obtained by dividing the sentences judged as positive included in each review x to the review scores y.

Pearson correlation coefficient r

In order to measure the correlation of the sentiment ratio x, obtained by dividing the number of sentences judged as positive included in each review i_P by the total number of sentences i_T, to the score y, one of the methods used was Pearson's correlation coefficient r (3.10). The formula is shown below. The value of the formula (3.9) is substituted into the formula (3.10). When calculating the negative rate, substitute the number of sentences judged to be negative i_N into the formula (3.9).

$$x = \frac{i_P}{i_T} \tag{3.9}$$

$$r = \frac{\sum_{i=1}^{M}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{M}(x_i - \overline{x})^2(y_i - \overline{y})^2}} \tag{3.10}$$

Spearman's rank correlation coefficient ρ

In order to measure the correlation of the sentiment ratio x, obtained by dividing the number of sentences judged as positive included in each review i_P by the total number of sentences i_T, to the score y, since tI consider the score to be a ranked variable, I used Spearman's ranked correlation coefficient ρ (3.11) which is also based on Pearson's correlation coefficient. The formula is shown below. Substitute the value of the formula (3.9) into the formula (3.11).

$$r_s = \rho_{rg_X, rg_Y} = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}} \tag{3.11}$$

Kendal's rank correlation coefficient τ

Like Spearman's rank correlation coefficient, Kendal's rank correlation coefficient is used to investigate the relationship between the values that represent rank. The formula (3.12) is shown below. However, substitute the expression (3.13) and the expression (3.14) into the expression (3.12). Substitute the expression (3.9) for each.

$$\tau = (K - L) / \binom{n}{2} \tag{3.12}$$

$$L = \#\left\{ \{i, j\} \in \binom{[n]}{2} \mid \neg(x_i \lesseqgtr x_j, y_i \lesseqgtr y_j) \right\} \tag{3.13}$$

$$K = \#\left\{ \{i, j\} \in \binom{[n]}{2} \mid (x_i \lesseqgtr x_j, y_i \lesseqgtr y_j) \right\} \tag{3.14}$$

MIC

Pearson's correlation coefficient can only extract linear relationships. On the other hand, there is MIC (Maximal Information Coefficient) as a method to analyze the relationship between two variables including non-linearity (Reshef et al., 2011). It is an index for analyzing the relationship between variables including non-linearity based on the amount of mutual information, considering the two variables for which you want to analyze the relationship as random variables. Fig. 3.2 shows several examples of comparing the coefficients of MIC and Pearson. Pearson's coefficient is a value from 0 to 1 in a linear relationship, and it is determined whether it is a positive value or a negative value depending on the direction of inclination. On the other hand, in the case of MIC, even if it is a non-linear relationship, it can express the correlation using the value from 0 to 1 as long as there is a relationship. Examples of this are shown in Fig.3.2.

The procedure for calculating MIC will be described. First, for the variables X and Y to be analyzed, after plotting the two variables on the coordinate space, the space is divided by $a * b$ (Split the X direction into a parts, and the Y direction into b parts). Then, for each of the two variables, the cell existence probability can be calculated by dividing the number of sample points belonging to each cell by the total number of samples. That is, X and Y are regarded as random variables based on the existence probability in the cell. This makes it possible to calculate the Mutual Information for X and Y.

At this time, even if the original two variables have a non-linear relationship as well as a linear relationship, the dependency between the random variables becomes strong, so the mutual information takes a large value. Therefore, unlike Pearson's correlation coefficient, it is possible to extract a non-linear relationship. Now, with the MIC, the width and length of each cell are unequally spaced, so there are innumerable division methods (note that each cell has a maximum resolution). For the purpose of extracting non-linearity, it is necessary to find a dividing grid that maximizes the amount of mutual information as much as possible. In MIC, it is assumed that X is now divided into arbitrary a pieces and Y is divided into arbitrary b pieces. At this time, I find a division that maximizes the amount of mutual information in the division of $a * b$ by brute force. The formula (3.15) for mutual information is shown below.

$$I(X;Y) = \int_Y \int_X p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx\, dy \tag{3.15}$$

In this study, the MIC was calculated using the Python library minepy (Albanese et al., 2012). The Fig.3.2, which compares Pearson's correlation coefficient and MIC, was obtained from the minepy API site.

3.4 Results

The following describes what was specifically done in this study using the method described earlier.

3.4.1 Preprocessing

First, I collected 1,541,424 HTML files from May 2016 to September 2016 from *Ctrip*. Of the 1,541,424 HTML files, I collected data on 5,938 hotels in Japan. Among them, he scraped the comment text of 44,912 reviews. 286,109 sentences divided into sentence units are analyzed. In addition, the scores of the reviews were also collected.

3.4.2 Sentiment analysis performance

In order to create training data, a sample of reviews was randomly extracted from all the data, divided into sentences by expressing satisfaction or dissatisfaction, tagging work of 159 sentences was performed manually. The entropy was calculated from the sentences that become the training data.

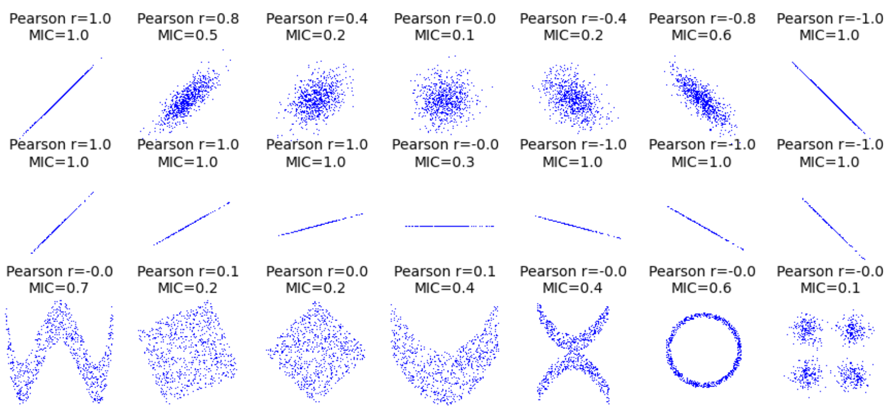


FIGURE 3.2: Comparison of minepy MIC and Pearson r in various cases

After calculating the entropy for the training data, the features with the maximum F_1 value were selected by evaluating α from 1.0 to 3.75 with a step size of 0.25 in order to obtain its optimum value. After training the SVM, the keyword list was selected based on the α which led to the maximum F_1 value as a result of 5-Fold Cross Validation ($k = 5$) for the evaluation data. Furthermore, both lists were combined to create a new list, and 5-Fold Cross Validation was performed in the same manner. The evaluation results are shown in Table 3.1. The feature that combines both has the highest F_1 value, and F_1 classifies all data with an SVM with a high accuracy of 0.95. Positive keywords are, for example, "热情" and "景色" (indicating "friendly" and "(good) scenery" respectively), and in the case of Negative keywords, there was an example of "价格" or "price" (indicating dissatisfaction because it is high).

TABLE 3.1: 5-fold Cross Validation performance results

Keyword list	C	$F_1\mu$	$F_1\sigma$
Positive keywords ($\alpha = 2.75$)	2.5	0.91	0.01
Negative keywords ($\alpha' = 3.75$)	0.5	0.67	0.11
Combination	**0.5**	**0.95**	**0.01**

TABLE 3.2: Correlation of sentiment analysis and score results

Sentence ratio	**Spearman's ρ**	**Kendall's τ**	**MIC**
Positive ratio	0.161	0.125	0.049
Negative ratio	-0.149	-0.122	0.0447

After learning with the optimal model, sentiment analysis was performed on unknown data, and the above-mentioned "Positive ratio" and "Negative ratio" were calculated. Since these values are the ratio of sentences expressing satisfaction and sentences expressing dissatisfaction, they were treated as numerical coefficients expressing emotions in each sentence, and correlation analysis was performed.

3.4.3 Correlation analysis

The positive and negative rates of each review in all the data and the review score of the reviews were analyzed using Spearman's rank correlation coefficient, Kendall's rank correlation coefficient, and MIC. The results are shown in Table 3.2.

3.5 Discussion

It was shown that both the Positive ratios and the Negative ratios were very low in relation to the review score in all the indicators.

Therefore, since the relationship between the result of sentiment analysis in the comment text and the review score is low, when analyzing the user's opinion from the review, considering both the content of the text and the numerical review score and their differences is very important.

In the past, although this relationship has not been shown, only review scores are often used as indicators of satisfaction and emotional evaluation. For example, the studies by Xie, Zhang, and Zhang (2014) and Bulchand-Gidumal, Melián-González, and González Lopez-Valcarcel (2013) used the scores for the hotel as a proxy for satisfaction, and Zhou et al. (2014) investigated the factors of satisfaction using a multivariate analysis to do this, but the dependent variable was the review score.

Based on the results of this study, it was suggested that it is not appropriate to use only the review score, but rather the sentiment analysis as an index to measure the satisfaction of tourists. This is why I will proceed to study satisfaction in Japan in Chapter 4.

3.6 Conclusion and future work

In this study, I investigated the relationship between the results of sentiment analysis in the text text of online hotel reviews and the evaluation points. I constructed a method with high classification performance ($F_1 = 0.95$) for sentiment analysis, and calculated the ratio of sentences classified as Positive and the ratio of sentences classified as Negative for each review. For the relationship, Spearman's rank correlation coefficient, Kendall's rank correlation coefficient, and MIC were used. As a result, all showed low values. Therefore, it was clarified that the relationship between the result of sentiment analysis in the comment text and the evaluation point, which is numerical information, is low. Therefore, it was considered that a more comprehensive evaluation was important. In the future, based on these results, I will investigate whether the result of sentiment analysis or the evaluation score expresses the user's opinion more after comparing with the analysis using multilingual information, and comprehensively reviewing. I will proceed with the development of analysis methods.

Chapter 4

Evaluating differences in Japanese hotel reviews by Chinese and Western tourists

4.1 Introduction

Inbound international tourism has been increasingly affecting Japanese economy (Jones et al., 2009). A year-on-year growth rate of 19.3% was observed in 2017, with 28,691,073 inbound tourists (Japan National Tourism Organization, 2019).

Japan's hospitality has been known historically to be of the highest quality. *Omotenashi*, which describes the spirit of Japanese hospitality, with roots in Japanese history and tea ceremony, is celebrated worldwide (Al-alsheikh and Sato, 2015). Consequently, it would stand to reason that tourists visiting Japan would have this hospitality as their first and foremost satisfaction factor. However, it is known that customers from different countries and cultures have different expectations (Engel, Blackwell, and Miniard, 1990). Thus, it could be theorized that their satisfaction factors should be different.

The Japanese tourist market is gradually becoming diverse because of multicultural tourist populations. This diversity means that the expectations when staying at a hotel will be varied. Cultural backgrounds have a decisive role in aspects of satisfaction and in the perceptions of quality (Mattila, 1999; Winsted, 1997), or behavioral intentions (Liu, Furrer, and Sudharshan, 2001), such as the difference in Westerners and Asians in their willingness to pay more (Levy, 2010). A difference in cultural background can also heavily influence customers' expectations, as well as their perceptions of quality, and the difference between these two is what expresses itself as satisfaction. This difference in expectations and perceptions of quality can be smaller or larger depending on the culture in reaction to the same service.

For a growing industry with increasing cultural diversity, it is essential to identify the cross-culture expectations of customers in order to provide the appropriate services, cater to these expectations to ensure and increase customer satisfaction, maintain a good reputation, and generate positive word-of-mouth.

In 2017, Chinese tourists accounted for 25.63% of the tourist population. On the other hand, Western countries accounted for 11.4% of the total, and 7.23% were countries where English is the official or the de facto national language (Japan National Tourism Organization, 2019). The effect of Chinese tourists on international economies is increasing, along with the number of studies on this phenomenon, (Sun, Wei, and Zhang, 2017). Despite this, many tourist-behavior analyses have been performed only involving Western subjects. Yet, it is known that Western and Asian customers are heavily differentiated(Levy, 2010). As such, a knowledge gap existed until recent decades. Considering the numbers of inbound tourists in Japan and my team's language capabilities, my study focuses on Western and Chinese tourists.

In studies involving Asian populations in the analysis, Chinese-tourist behaviors have been evaluated most commonly (e.g. Liu et al., 2019; Chang, Kivela, and Mak, 2010; Dongyang, Mori, Hayashi, et al., 2015). The few studies reporting comparisons between Asian and Western tourists' behaviors (e.g. Choi and Chu, 2000) are typically survey- or interview-based, using small samples. These studies, although valid, can have limitations, namely, the scale and sampling. In the past, survey-based studies have provided a theoretical background for a few specific tourist populations of a single culture or traveling with a single purpose. These studies' limited scope often leads to difficulties in observing cultural and language differences in a single study. This creates a need for large-scale cross-cultural studies for the increasing Asian and Western tourist populations. It could be said that Westerners account for a smaller portion of the tourist population compared to Asians. However, according to Choi and Chu (2000), Westerners are known as "long-haul" customers, spending more than 45% of their budget on hotels. In comparison, their Asian counterparts only spend 25% of their budget on hotels. Therefore, it is essential to study Asian and Western tourist populations, their differences, and the contrast with the existing literature results.

However, with ever-increasing customer populations, this is hard to accomplish without extensive studies of the customer base. There is a need for an automated method for identifying these expectations at a large scale. My study intends to answer the need for such a methodology utilizing machine learning and natural language processing of large amounts of data. For this, I used a data-driven approach to my analysis, taking advantage of hotel review data. With this methodology, I explore the expectations and needs for the two most differing cultures currently interacting with the hospitality industry in Japan.

Owing to the advent of Web 2.0 and customer review websites, researchers realized the benefits of online reviews for research, sales (Ye, Law, and Gu, 2009; Basuroy, Chatterjee, and Ravid, 2003), customer consideration (Vermeulen and Seegers, 2009) and perception of services and products (Browning, So, and Sparks, 2013), among other effects of online interactions between customers (e.g. Xiang and Gretzel, 2010; Ren and Hong, 2019). Consequently, information collected online is being used in tourism research for data mining analysis, such as opinion mining (e.g. Hu, Chen, and Chou, 2017), predicting hotel demand from online traffic (Yang, Pan, and Song, 2014), recommender systems (e.g. Loh et al., 2003), and more. Data mining and machine learning technologies can increase the number of manageable samples in a study from hundreds to hundreds of thousands. These technologies can not only help confirm existing theories but also lead to finding new patterns and to knowledge discovery (Fayyad, Piatetsky-Shapiro, and Smyth, 1996).

In this study, I evaluate the satisfaction factors of two essential tourist populations that are culturally different from Japan: Chinese and Western tourists. I take advantage of the wide availability of online reviews of Japanese hotels by both Mainland Chinese tourists posting on *Ctrip* and Western, English-speaking tourists posting on *TripAdvisor*. Based on these data, I can confirm existing theories regarding the differences in tourists' behavior and discover factors that could have been overlooked in the past. I use machine learning to automatically classify sentences in the online reviews as positive or negative opinions on the hotel. I then perform a statistical extraction of the topics that most concern the customers of each population.

4.1.1 Research objective

With the knowledge that cultural background influences expectations in customers, which is the basis for satisfaction, it becomes important to know the difference in factors influencing satisfaction and dissatisfaction between the most differing and numerous tourist populations in a given area.

Research Question 4.1: *What are the differences in factors influencing satisfaction and dissatisfaction between Chinese and English-speaking tourists staying in Japanese hotels?*

This study aims to determine the difference in factors influencing satisfaction and dissatisfaction between Chinese and English-speaking tourists in the context of high-grade hospitality of Japanese hotels across several price ranges. I use machine learning to classify the sentiment in texts and natural language processing to study commonly used word pairings. More importantly, I also intend to measure how hard and soft attributes influence customer groups' satisfaction and dissatisfaction. I define hard attributes as attributes relating to physical and environmental aspects, such as the hotel's facilities, location, infrastructure, and surrounding real estate. In contrast, soft attributes are the hotel's non-physical attributes related to services, staff, or management.

4.2 Theoretical background and hypothesis development

4.2.1 Cultural influence in expectation and satisfaction

Customer satisfaction in tourism has been analyzed since decades past, Hunt (1975) having defined 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 and 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. As such, Engel, Blackwell, and Miniard (1990) states that each customer's background, therefore, influences satisfaction and dissatisfaction. It can also be said that satisfaction stems from the perceptions of quality in comparison to these expectations.

These differences in customers' backgrounds can be summed up in cultural differences as well. In the past, satisfaction and perceived service quality have been found to be influenced by cultural differences (e.g. Mattila, 1999; Winsted, 1997). Service quality perceptions have been studied via measurements such as SERVQUAL (e.g. Armstrong et al., 1997).

Previous studies on the dimensions of culture that influence differences in expectations have been performed in the past as well (e.g. Mattila, 2019; Levy, 2010; Donthu and Yoo, 1998), such as comparing individualism vs. collectivism, high context vs. low context, uncertainty avoidance, among other factors. While culture as a concept is difficult to quantify, some researchers have tried to use these and more dimensions to measure cultural differences, such as the six dimensions described by Hofstede (1984), or the nine dimensions of the GLOBE model (House et al., 1999).

These cultural dimensions are more differentiated in Western and Asian cultures (Levy, 2010). This study being located in Japan, it stands to reason that the differences in expectations between Western tourists and Asian tourists should be understood in order to provide a good service. However, even though geographically close, Japanese and Chinese cultures are both very different when it comes to customer service. This is why my study focuses on the difference between Chinese and Western customers in Japan. The contrasting cultural backgrounds between Chinese and Western customers will lead to varying expectations of the hotel services, the experiences they want to have while staying at a hotel, and the level of comfort that they will have. In turn, these different expectations will determine the distinct factors of satisfaction and dissatisfaction for each kind of customer and the order in which they prioritize them.

Because of their different origins, expectations, and cultures, it stands to reason Chinese and Western tourists could have completely different factors to one another. Therefore, it could be that some factors do not appear in the other reviews at all. For example, between different cultures, it can be that a single word can express some concept that would take more words in the other language. Therefore, I must measure their differences or similarities at their common ground as well.

4.2.2 Customer satisfaction and dissatisfaction towards individual factors during hotel stay

I reviewed the importance of expectations in the development of satisfaction and dissatisfaction and the influence that cultural backgrounds have in shaping these expectations. This is true for overall satisfaction for the service as a whole, as well as individual elements that contribute to satisfaction.

In this study, I study not overall customer satisfaction but the satisfaction and dissatisfaction that stem from individual-specific expectations, be they conscious or unconscious. For example, if a customer has a conscious expectation of a comfortable bed and a wide shower, and it is realized during their visit, they will be satisfied with this matter. However, suppose that same customer with a conscious expectation of a comfortable bed experienced loud noises at night. In that case, they can be dissatisfied with a different aspect, regardless of the satisfaction towards the bed. Then, the same customer might have packed their toiletries, thinking that the amenities might not include those. They can then be pleasantly surprised with good quality amenities and toiletries, satisfying an unconscious expectation. This definition of satisfaction does not allow us to examine overall customer satisfaction. However, it will allow us to examine the factors that a hotel can revise individually and how a population perceives them as a whole. In this study, I consider the definitions in Hunt (1975) that satisfaction is a realization of an expectation, and I posit that customers can have different expectations towards different service aspects. Therefore, in this study, I define satisfaction as the emotional response to the realization or overcoming of conscious or unconscious expectations towards an individual aspect or factor of a service. On the other hand, dissatisfaction is the emotional response to the lack of a realization or under-performance of these conscious or unconscious expectations towards specific service aspects.

Studies on customer satisfaction (e.g. Truong and King, 2009; Romão et al., 2014; Wu and 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, Hsu, and Baum, 2015; Choi and Chu, 2000). The Likert scale's use leads to correlation analyses where one factor can lead to satisfaction, implying that the lack of it can lead to dissatisfaction. However, a binary distinction (satisfied or dissatisfied) could allow us to analyze the factors that correlate to satisfaction and explore factors that are solely linked to dissatisfaction. There are fewer examples of this approach, but studies have done this in the past (e.g. Zhou et al., 2014). This method can indeed decrease the extent to which I can analyze degrees of satisfaction or dissatisfaction. However, it has the benefit that it can be applied to a large sample of text data via automatic sentiment detection techniques using artificial intelligence.

4.2.3 Japanese hospitality and service: *Omotenashi*

The spirit of Japanese hospitality, or *Omotenashi*, has roots in the country's history, and to this day, it is regarded as the highest standard (Ikeda, 2013; Al-alsheikh and Sato, 2015). There is a famous phrase in customer service in Japan: *okyaku-sama wa kami-sama desu*, meaning "The customer is god." Some scholars say that *omotenashi* originated from the old Japanese art of the tea ceremony in the 16th century, while others found that it originates in the form of formal banquets in the 7th-century (Aishima, Sato, et al., 2015). The practice of high standards in hospitality has survived throughout the years. Presently, it permeates all business practices in Japan, from the cheapest convenience stores to the most expensive ones. Manners, service, and respect towards the customer are taught to workers in their training. High standards are always followed to not fall behind in the competition. In Japanese businesses, including hotels, staff members are trained to speak in *sonkeigo*, or "respectful language," one of the most formal of the Japanese formality syntaxes. They are also trained to bow differently depending on the situation, where a light bow could be used to say "Please, allow me to guide you." Deep bows are used to apologize for any inconvenience the customer could have faced, followed by a very respectful apology. Although the word *omotenashi* can be translated directly as "hospitality," it includes both the concepts of hospitality and service (Kuboyama, 2020). This hospitality culture permeates every type of business with customer interaction in Japan. A simple convenience shop could express all of these hospitality and service standards, which are not exclusive to hotels.

It stands to reason that this cultural aspect of hospitality would positively influence customer satisfaction. However, in many cases, other factors such as proximity to a convenience store, transport availability, or room quality might be more critical to a customer. In this study, I cannot directly determine whether a hotel is practicing the cultural standards of *omotenashi*. Instead, I consider it as a cultural factor that influences all businesses in Japan. I then observe the customers' evaluations regarding service and hospitality factors and compare them to other places and business practices in the world. In summary, I consider the influence of the cultural aspect of *omotenashi* while analyzing the evaluations on service and hospitality factors that are universal to all hotels in any country.

Therefore, I pose the following research question:

Research Question 4.2a: *To what degree are Chinese and Western tourists satisfied with Japanese hospitality factors such as staff behavior or service?*

However, Japanese hospitality is based on Japanese culture. Different cultures interacting with it could provide a different evaluation of it. Some might be impressed by it, whereas some might consider other factors more important to their stay in a hotel. This point leads us to a derivative of the aforementioned research question:

Research Question 4.2b: *Do Western and Chinese tourists have a different evaluation of Japanese hospitality factors such as staff behavior or service?*

4.2.4 Customer expectations beyond service and hospitality

Staff behavior, hospitality and service, and therefore *Omotenashi*, are all soft attributes of a hotel. That is, they are non-physical attributes of the hotel, and as such, they are practical to change through changes in management. While it is important to know this, it is not known if the cultural differences between Chinese and Western tourists also influence other expectations and satisfaction factors, such as the hard factors of a hotel.

Hard factors are attributes uncontrollable by the hotel staff, which can play a part in the customers' choice behavior and satisfaction. Examples of these factors include the hotel's surroundings, location, language immersion of the country as a whole, or touristic destinations, and the hotel's integration with tours available nearby, among other factors.

Besides the facilities, many other aspects of the experience, expectation, and perception of the stay in a hotel can contribute to the overall satisfaction, as well as individual satisfactions and dissatisfactions. However, previous research focuses more on these soft attributes, with little focus on hard attributes, if only focusing on facilities (e.g. Shanka and Taylor, 2004; Choi and Chu, 2001). Because of this gap in knowledge, I decided to analyze the differences in cultures regarding both soft and hard attributes of a hotel.

This leads to two of my research questions:

Research Question 4.3a: *To what degree do satisfaction and dissatisfaction stem from hard and soft attributes of the hotel?*

Research Question 4.3b: *How differently do Chinese and Western customers perceive hard and soft attributes of the hotel?*

The resulting proportions of hard attributes to soft attributes for each population could measure how much the improvement of management in the hotel can increase future satisfaction in customers.

4.2.5 Chinese and Western tourist behavior

In the past, social science and tourism studies focused extensively on Western tourist behavior in other countries. Recently, however, with the rise of Chinese outbound tourism, both academic researchers and businesses have decided to study Chinese tourist behavior, with rapid growth in studies following the year 2007 (Sun, Wei, and Zhang, 2017). However, studies focusing on only the behavior of this subset of tourists are the majority. To this day, studies and analyses specifically comparing Asian and Western tourists are scarce, and even fewer are the number of studies explicitly comparing Chinese and Western tourists. One example is a study by Choi and Chu (2000), which found that Western tourists visiting Hong Kong are satisfied more with room quality, while Asians are satisfied with the value for money. Another study by Bauer, Jago, and Wise (1993) found that Westerners prefer hotel health facilities, while Asian tourists were more inclined to enjoy the Karaoke facilities of hotels. Both groups tend to have high expectations for the overall facilities. Another study done by Kim and Lee (2000) found American tourists to be individualistic and motivated by novelty, while Japanese tourists were collectivist and motivated by increasing knowledge and escaping routine.

One thing to note with the above Asian vs. Western analyses is that they were performed before 2000 and not Chinese-specific. Meanwhile, the current Chinese economic boom is increasing the influx of tourists of this nation. The resulting increase in marketing and the creation of guided tours for Chinese tourists could have created a difference in tourists' perceptions and expectations. In turn, if I follow the definition of satisfaction in Hunt (1975), the change in expectations could have influenced their satisfaction factors when traveling. Another note is that these studies were performed with questionnaires in places where it would be easy to locate tourists, i.e., airports. However, my study of online reviews takes the data that the hotel customers uploaded themselves. This data makes the analysis unique in exploring their behavior compared with Western tourists via factors that are not considered in most other studies. Furthermore, this study is unique in observing the customers in the specific environment of high-level hospitality in Japan.

More recent studies have surfaced as well. A cross-country study (Francesco and Roberta, 2019) using posts from U.S.A. citizens, Italians, and Chinese tourists, determined using a text link analysis that customers from different countries indeed have a different perception and emphasis of a few predefined hotel attributes. According to their results, U.S.A. customers perceive cleanliness and quietness most positively. In contrast, Chinese customers perceive budget and restaurant above other attributes. Another couple of studies (Jia, 2020; Huang, 2017) analyze differences between Chinese and U.S. tourists using text mining techniques and more massive datasets, although in a restaurant context.

These last three studies focus on the U.S.A. culture, whereas this study focuses on the Western culture. Another difference with this study is that of the context of the study. The first study (Francesco and Roberta, 2019) was done within the context of tourists from three countries staying in hotels across the world. The second study chose restaurant reviews from the U.S.A. and Chinese tourists eating in three countries in Europe. The third study analyzed restaurants in Beijing.

On the other hand, this study focuses on Western culture, instead of a single Western country, and Chinese culture clashing with the hospitality environment in Japan, specifically. Japan's importance in this analysis comes from the unique environment of high-grade hospitality that the country presents. In this environment, customers could either hold their satisfaction to this hospitality regardless of their culture or value other factors more depending on their cultural differences. My study measures this at a large scale across different hotels in Japan.

Other studies have gone further and studied people from many countries in their samples and performed a more universal and holistic (not cross-culture) analysis. Choi and Chu (2001) 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 and Chu (2001) found that staff service quality, room quality, and value for money were the top satisfaction determinants. As 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 many countries. The general satisfaction score was noticed to be different among those countries. However, a more in-depth cross-cultural analysis of the satisfaction factors was not performed. As a result of their research, 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 my 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 and Taylor (2004) found that English-speaking tourists in Perth, Australia were most satisfied with staff friendliness, the 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, Wei, and Zhang, 2017). Different researchers have found that Chinese tourist populations have several specific attributes. According to Ryan and Mo (2001) and their study of Chinese tourists in New Zealand, Chinese tourists prefer nature, cleanliness, and scenery in contrast to experiences and activities. Dongyang, Mori, Hayashi, 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 and 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 compared with other international tourists. Moreover, as stated by Gao, Zhang, and Huang (2017), who analyzed different generations of Chinese tourists and their connection to nature while traveling, Chinese tourists prefer nature overall. However, the younger generations seem to do so less than their older counterparts.

Although the studies focusing only on Chinese or Western tourists have a narrow view, their theoretical contributions are valuable. I can see that depending on the study and the design of questionnaires and the destinations; the results can vary greatly. Not only that, but while there seems to be some overlap in most studies, some factors are completely ignored in one study but not in the other. Since this study uses data mining, each factor's definition is left for hotel customers to decide en masse via their reviews. This means that the factors will be selected through statistical methods alone instead of being defined by the questionnaire. This method allows us to find factors that I would not have contemplated. It also avoids enforcing a factor on the mind of study subjects by presenting them with a question that they did not think of by themselves. This large variety of opinions in a well-sized sample, added to the automatic findings of statistical text analysis methods, gives my study an advantage compared to others with smaller samples. This study analyzes the satisfaction and dissatisfaction factors cross-culturally and compares them with the existing literature.

Undoubtedly previous literature has examples of other cross-culture studies of tourist behavior and may further highlight my study and its merits. A contrast is shown in Table 4.2. This table shows that older studies were conducted with surveys and had a different study topic. These are changes in demand (Bauer, Jago, and Wise, 1993), tourist motivation (Kim and Lee, 2000), and closer to my study, satisfaction levels (Choi and Chu, 2000). However, my study topic is not the levels of satisfaction but the factors that drive it and dissatisfaction, which is overlooked in most studies. Newer studies with larger samples and similar methodologies have emerged, although two of these study restaurants instead of hotels (Jia, 2020; Huang, 2017). One important difference is the geographical focus of their studies. While Francesco and Roberta (2019) , Jia (2020) and Huang (2017) have a multi-national focus, I instead focus on Japan. The focus on Japan is important because of its top rank in hospitality across all types of businesses. My study brings light to the changes, or lack thereof, in different touristic environments where an attribute can be considered excellent. The number of samples in other text-mining studies is also smaller than ours in comparison. Apart from that, every study has a different text mining method.

TABLE 4.1: Comparison between cross-culture or cross-country previous studies and my study.

	Bauer et.al (1993)	**Choi and Chu (2000)**	**Kim and Lee (2000)**	**Huang (2017)**	**Francesco and Roberta (2019)**	**Jia (2020)**	**Our study**
Comparison objects	Asians vs Westerns	Asians vs Westerners	Anglo-Americans vs Japanese	Chinese vs English-speakers	USA vs China vs Italy	Chinese vs US tourists	Chinese vs Westerners
Study topic	Changes in demand	Satisfaction Levels	Tourist Motivation	Dining experience of Roast Duck	Perception and Emphasis	Motivation and Satisfaction	**Satisfaction and Dissatisfaction**
Geographical focus	Asia Pacific region	Hong Kong	Global	Beijing	Multi-national	Multi-national	**Japan**
Industry	Hotels	Hotels	Tourism	Restaurant (Beijing Roast Duck)	Hotels	Restaurants	Hotels
Study subjects	Hotel managers	Hotel customers	Tourists arriving in airport	Diners online reviews	Hotel customers online reviews	Diners online reviews	Hotel customers online reviews
Sample method	surveys	surveys	survey	text mining	text mining	text mining	text mining
Number of samples	185 surveys	540 surveys	165 Anglo-American 209 Japanese	990 Chinese reviews 398 English reviews	9000 reviews (3000 per country)	2448 reviews (1360 Chinese) (1088 English)	**89,207 reviews** **(48,070 Chinese)** **(41,137 English)**
Study method	statistics	VARIMAX	MANOVA	Semantic Network Analysis	Text Link Analysis	Topic modeling (LDA)	**SVM, Dependency Parsing and POS tagging**
Subject nationality	Asians: China, Fiji, Hong Kong, Indonesia, Malaysia, Singapore, Taiwan, Guam, Tahiti, Thailand Westerners: Australia, New Zealand	Asians: China, Taiwan, Japan, South Korea, South-East Asia Westerners: North America, Europe, Australia, New Zealand	USA, Japan	English-speakers: U.K., U.S., Australia, New Zealand, Canada, Ireland Chinese-speakers: China	USA, China, Italy	USA, China	Chinese-speakers: China English-speakers: (U.K., U.S., Australia, New Zealand, Canada, Ireland)

4.2.6 Data mining, machine learning, knowledge discovery and sentiment analysis

In the past, social science and tourism studies focused extensively on Western tourist behavior in other countries. Recently, however, with the rise of Chinese outbound tourism, both academic researchers and businesses have decided to study Chinese tourist behavior, with rapid growth in studies following the year 2007 (Sun, Wei, and Zhang, 2017). However, studies focusing on only the behavior of this subset of tourists are the majority. To this day, studies and analyses specifically comparing Asian and Western tourists are scarce, and even fewer are the number of studies explicitly comparing Chinese and Western tourists. One example is a study by Choi and Chu (2000), which found that Western tourists visiting Hong Kong are satisfied more with room quality, while Asians are satisfied with the value for money. Another study by Bauer, Jago, and Wise (1993) found that Westerners prefer hotel health facilities, while Asian tourists were more inclined to enjoy the Karaoke facilities of hotels. Both groups tend to have high expectations for the overall facilities. Another study done by Kim and Lee (2000) found American tourists to be individualistic and motivated by novelty, while Japanese tourists were collectivist and motivated by increasing knowledge and escaping routine.

One thing to note with the above Asian vs. Western analyses is that they were performed before 2000 and not Chinese-specific. Meanwhile, the current Chinese economic boom is increasing the influx of tourists of this nation. The resulting increase in marketing and the creation of guided tours for Chinese tourists could have created a difference in tourists' perceptions and expectations. In turn, if I follow the definition of satisfaction in Hunt (1975), the change in expectations could have influenced their satisfaction factors when traveling. Another note is that these studies were performed with questionnaires in places where it would be easy to locate tourists, i.e., airports. However, my study of online reviews takes the data that the hotel customers uploaded themselves. This data makes the analysis unique in exploring their behavior compared with Western tourists via factors that are not considered in most other studies. Furthermore, my study is unique in observing the customers in the specific environment of high-level hospitality in Japan.

More recent studies have surfaced as well. A cross-country study (Francesco and Roberta, 2019) using posts from U.S.A. citizens, Italians, and Chinese tourists, determined using a text link analysis that customers from different countries indeed have a different perception and emphasis of a few predefined hotel attributes. According to their results, U.S.A. customers perceive cleanliness and quietness most positively. In contrast, Chinese customers perceive budget and restaurant above other attributes. Another couple of studies (Jia, 2020; Huang, 2017) analyze differences between Chinese and U.S. tourists using text mining techniques and more massive datasets, although in a restaurant context.

These last three studies focus on the U.S.A. culture, whereas my study focuses on the Western culture. Another difference with my study is that of the context of the study. The first study (Francesco and Roberta, 2019) was done within the context of tourists from three countries staying in hotels across the world. The second study chose restaurant reviews from the U.S.A. and Chinese tourists eating in three countries in Europe. The third study analyzed restaurants in Beijing.

On the other hand, my study focuses on Western culture, instead of a single Western country, and Chinese culture clashing with the hospitality environment in Japan, specifically. Japan's importance in this analysis comes from the unique environment of high-grade hospitality that the country presents. In this environment, customers could either hold their satisfaction to this hospitality regardless of their culture or value other factors more depending on their cultural differences. My study measures this at a large scale across different hotels in Japan.

Other studies have gone further and studied people from many countries in their samples and performed a more universal and holistic (not cross-culture) analysis. Choi and Chu (2001) 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 and Chu (2001) found that staff service quality, room quality, and value for money were the top satisfaction determinants. As 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 many countries. The general satisfaction score was noticed to be different among those countries. However, a more in-depth cross-cultural analysis of the satisfaction factors was not performed. As a result of their research, 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 my 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 and Taylor (2004) found that English-speaking tourists in Perth, Australia were most satisfied with staff friendliness, the 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, Wei, and Zhang, 2017). Different researchers have found that Chinese tourist populations have several specific attributes. According to Ryan and Mo (2001) and their study of Chinese tourists in New Zealand, Chinese tourists prefer nature, cleanliness, and scenery in contrast to experiences and activities. Dongyang, Mori, Hayashi, 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 and 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 compared with other international tourists. Moreover, as stated by Gao, Zhang, and Huang (2017), who analyzed different generations of Chinese tourists and their connection to nature while traveling, Chinese tourists prefer nature overall. However, the younger generations seem to do so less than their older counterparts.

Although the studies focusing only on Chinese or Western tourists have a narrow view, their theoretical contributions are valuable. I can see that depending on the study and the design of questionnaires and the destinations; the results can vary greatly. Not only that, but while there seems to be some overlap in most studies, some factors are completely ignored in one study but not in the other. Since my study uses data mining, each factor's definition is left for hotel customers to decide en masse via their reviews. This means that the factors will be selected through statistical methods alone instead of being defined by the questionnaire. My method allows us to find factors that I would not have contemplated. It also avoids enforcing a factor on the mind of study subjects by presenting them with a question that they did not think of by themselves. This large variety of opinions in a well-sized sample, added to the automatic findings of statistical text analysis methods, gives my study an advantage compared to others with smaller samples. This study analyzes the satisfaction and dissatisfaction factors cross-culturally and compares them with the existing literature.

Undoubtedly previous literature has examples of other cross-culture studies of tourist behavior and may further highlight my study and its merits. A contrast is shown in Table 4.2. This table shows that older studies were conducted with surveys and had a different study topic. These are changes in demand (Bauer, Jago, and Wise, 1993), tourist motivation (Kim and Lee, 2000), and closer to my study, satisfaction levels (Choi and Chu, 2000). However, my study topic is not the levels of satisfaction but the factors that drive it and dissatisfaction, which is overlooked in most studies. Newer studies with larger samples and similar methodologies have emerged, although two of these study restaurants instead of hotels (Jia, 2020; Huang, 2017). One important difference is the geographical focus of their studies. While Francesco and Roberta (2019) , Jia (2020) and Huang (2017) have a multi-national focus, I instead focus on Japan. The focus on Japan is important because of its top rank in hospitality across all types of businesses. My study brings light to the changes, or lack thereof, in different touristic environments where an attribute can be considered excellent. The number of samples in other text-mining studies is also smaller than ours in comparison. Apart from that, every study has a different text mining method.

TABLE 4.2: Comparison between cross-culture or cross-country previous studies and my study.

	Bauer et.al (1993)	**Choi and Chu (2000)**	**Kim and Lee (2000)**	**Huang (2017)**	**Francesco and Roberta (2019)**	**Jia (2020)**	**Our study**
Comparison objects	Asians vs Westerns	Asians vs Westerners	Anglo-Americans vs Japanese	Chinese vs English-speakers	USA vs China vs Italy	Chinese vs US tourists	Chinese vs Westerners
Study topic	Changes in demand	Satisfaction Levels	Tourist Motivation	Dining experience of Roast Duck	Perception and Emphasis	Motivation and Satisfaction	**Satisfaction and Dissatisfaction**
Geographical focus	Asia Pacific region	Hong Kong	Global	Beijing	Multi-national	Multi-national	**Japan**
Industry	Hotels	Hotels	Tourism	Restaurant (Beijing Roast Duck)	Hotels	Restaurants	Hotels
Study subjects	Hotel managers	Hotel customers	Tourists arriving in airport	Diners online reviews	Hotel customers online reviews	Diners online reviews	Hotel customers online reviews
Sample method	surveys	surveys	survey	text mining	text mining	text mining	text mining
Number of samples	185 surveys	540 surveys	165 Anglo-American 209 Japanese	990 Chinese reviews 398 English reviews	9000 reviews (3000 per country)	2448 reviews (1360 Chinese) (1088 English)	**89,207 reviews (48,070 Chinese) (41,137 English)**
Study method	statistics	VARIMAX	MANOVA	Semantic Network Analysis	Text Link Analysis	Topic modeling (LDA)	**SVM, Dependency Parsing and POS tagging**
Subject nationality	Asians: China, Fiji, Hong Kong, Indonesia, Malaysia, Singapore, Taiwan, Guam, Tahiti, Thailand Westerners: Australia, New Zealand	Asians: China, Taiwan, Japan, South Korea, South-East Asia Westerners: North America, Europe, Australia, New Zealand	USA, Japan	English-speakers: U.K., U.S., Australia, New Zealand, Canada, Ireland Chinese-speakers: China	USA, China, Italy	USA, China	Chinese-speakers: China English-speakers: (U.K., U.S., Australia, New Zealand, Canada, Ireland)

4.3 Methodology

I extracted a large number of text reviews from the site *Ctrip*, with mostly mainland Chinese users, and the travel site *TripAdvisor*. I then determined the most commonly used words that relate to positive and negative opinions in a review. I did this using Shannon's entropy to extract keywords from their vocabulary. These positive and negative keywords allow us to train an optimized Support Vector Classifier (SVC) to perform a binary emotional classification of the reviews in large quantities, saving time and resources for the researchers. I then applied a dependency parsing to the reviews and a Part of Speech tagging (POS tagging) to observe the relationship between adjective keywords and the nouns they refer to. I split the dataset into price ranges to observe the differences in keyword usage between lower-class and higher-class hotels. I observed the frequency of the terms in the dataset to extract the most utilized words in either review. I show an overview of this methodology in Figure 4.1, which is an updated version of the methodology used by Alemán Carreón et al. (2018). Finally, I also observed if the satisfaction factors were soft or hard attributes of the hotel.

4.3.1 Data collection

In the *Ctrip* data collection, reviews from a total of 5774 hotels in Japan were collected. From these pages, I extracted a total of 245,919 reviews, from which 211,932 were detected to be standard Mandarin Chinese. Since a single review can have sentences with different sentiments, I separated sentences using punctuation marks. The Chinese reviews were comprised of 187,348 separate sentences.

In the *TripAdvisor* data collection, I collected data from 21,380 different hotels. In total, I collected 295,931 reviews, from which 295,503 were detected to be in English. Similarly to the Chinese data, I then separated these English reviews into 2,694,261 sentences using the *gensim* python library. For the language detection in both cases I used the *langdetect* python library.

However, to make the data comparisons fair, I filtered both databases only to contain reviews from hotels in both datasets, using their English names to do a search match. I also filtered them to be in the same date range. In addition, I selected only the hotels that had pricing information available. I extracted the lowest and highest price possible for one night as well. The difference in pricing can be from better room settings, such as double or twin rooms or suites, depending on the hotel. Regardless of the reason, I chose the highest-priced room since it can be an indirect indicator of the hotel's class. After filtering, the datasets contained 557 hotels in common. The overlapping date range for reviews was from July 2014 to July 2017. Within these hotels, from *Ctrip* there was 48,070 reviews comprised of 101,963 sentences, and from *TripAdvisor* there was 41,137 reviews comprised of 348,039 sentences.

The price for a night in these hotels ranges from cheap capsule hotels at 2000 yen per night to high-end hotels 188,000 yen a night at the far ends of the bell curve. Customers' expectations can vary greatly depending on the pricing of the hotel room they stay at. Therefore, I made observations on the distribution of pricing in my database's hotels and binned the data by price ranges, decided by consideration of the objective of stay. I show these distributions in Figure 4.2. The structure of the data after division by price is shown in Table 4.3. This table also includes the results of emotional classification after applying my SVC, as explained in 4.3.3. The first three price ranges (0 to 2500 yen, 2500 to 5000 yen, 5000 to 10,000 yen) would correspond to low-class hotels or even hostels on the lower end and cheap business hotels on the higher end. Further on, there are business hotels in the next range (10,000 to 15,000 yen). After that, the stays could be at Japanese style *ryokan* when traveling in groups, high-class business hotels, luxury love hotels, or higher class hotels (15,000 to 20,000 yen, 20,000 to 30,000 yen). Further than that is more likely to be *ryokan* or high class resorts or five-star hotels (30,000 to 50,000 yen, 50,000 to 100,000 yen, 100,000 to 200,000 yen). Note that because of choosing the highest price per one night in each hotel, the cheapest two price ranges (0 to 2500 yen, 2500 to 5000 yen) are empty, despite some rooms being priced at 2000 yen per night. Because of this, other tables will omit these two price ranges.

4.3.2 Text processing

I needed to analyze the grammatical relationship between words, be it English or Chinese, to understand the connections between adjectives and nouns. For all these processes, I used the Stanford CoreNLP pipeline developed by the Natural Language Processing Group at Stanford University (Manning et al., 2014). In order to separate Chinese words for analysis, I used the Stanford Word Segmenter (Chang, Galley, and Manning, 2008). In English texts, however, only using spaces is not enough to correctly collect concepts. The English language is full of variations and conjugations of words depending on the context and tense. Thus, a better segmentation is achieved by using lemmatization, which returns each word's dictionary form. For this purpose, I used the *gensim* library for the English texts.

A dependency parser analyzes the grammatical structure, detecting connections between words, and describing the action and direction of those connections. I show an example of these dependencies in Figure 4.3. This study uses the Stanford NLP Dependency Parser, as described by Chen and Manning (2014). A list of dependencies used by this parser is detailed by Marneffe and Manning (2008). In more recent versions, they use an updated dependency tag list from Universal Dependencies (Zeman et al., 2018). In my study, this step was necessary to extract adjective modifiers and their subject. I did that by parsing the database and extracting instances of a few determined dependency codes. One of these dependency codes is "amod", which stands for "adjectival modifier". This is used when an adjective modifies a noun directly (e.g., A big apple). The other dependency code I used was "nsubj", or nominal subject, the class's syntactic subject. I used this one for cases where the adjective is modifying the noun indirectly through other words (e.g., The apple is big). This dependency does not necessarily only include a combination of adjectives and nouns. However, it can also be connected with copular verbs, nouns, or other adjectives. I saw it necessary also to perform a Part of Speech (POS) tagging of these clauses.

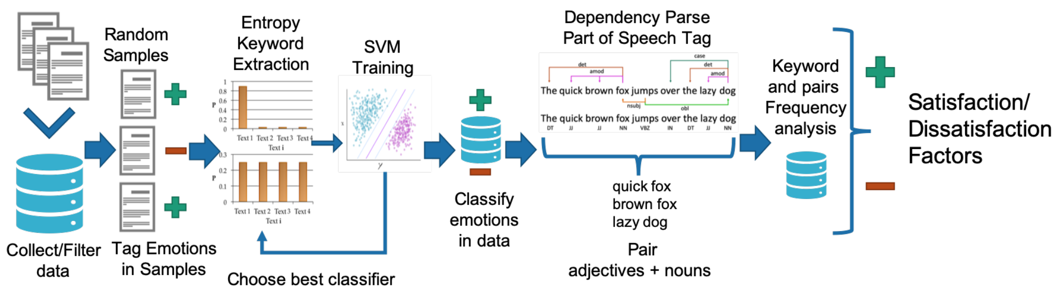
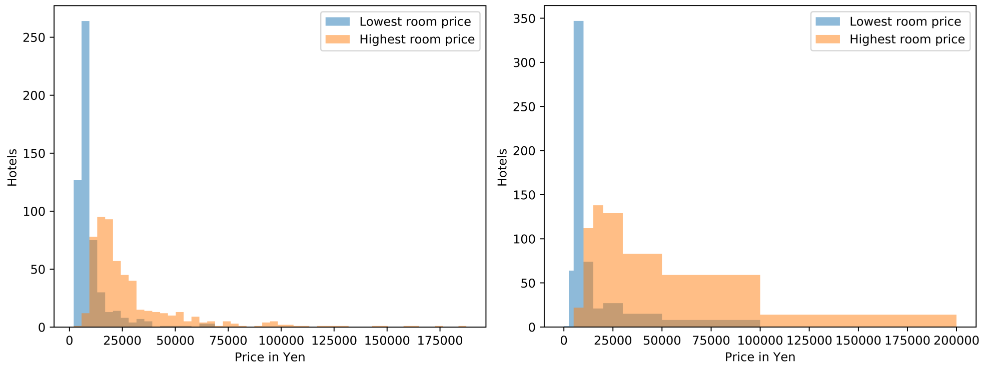


FIGURE 4.1: Overview of the methodology to quantitatively rank satisfaction factors.



(A) 50 equal lenght bins (B) manually set 9 price ranges

FIGURE 4.2: Price for one night distribution, blue: lowest price, orange: highest price.

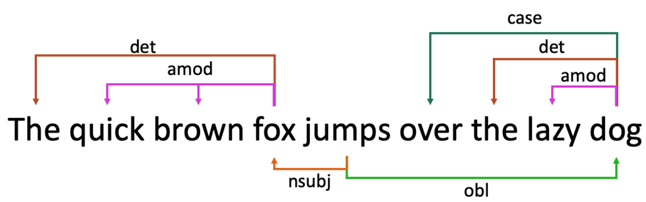


FIGURE 4.3: Example of dependency parsing.

A Part of Speech (POS) tagger is a program that assigns word tokens with tags identifying the part of speech. An example is shown in Figure 4.4. A Part of Speech is a category of lexical items that serve similar grammatical purposes, for example, nouns, adjectives, verbs, or conjunctions. In my study, I used the Stanford NLP POS tagger software, described by Toutanova and Manning (2000) and Toutanova et al. (2003), which uses the Penn Chinese Treebank tags (Xia, 2000).

FIGURE 4.4: Example of POS tagging with the Penn Treebank tags.

In this study, I were interested in identifying combinations of adjectives, some verbs, and nouns. I also needed to filter away bad combinations that were brought by the versatility of nominal subject dependencies. For this purpose, I identified the tags for nouns, verbs, and adjectives in Chinese and English, with the English tags being a bit more varied. What would be called adjectives in English corresponds more to stative verbs in Chinese, so I needed to extract those as well. I show a detailed description of the chosen tags in Table 4.4. I also show a detailed description of the tags I needed to filter. I selected these tags heuristically by observing commonly found undesired pairs in Table 4.5.

Once I had these adjective + noun or verb + noun pairs, I could determine what the customers referred to in their reviews. With what frequency they use those pairings positively or negatively.

4.3.3 Sentiment analysis using a Support Vector Classifier

The sentiment analysis was performed using the methodology described by Alemán Carreón et al. (2018). Keywords are determined by a comparison of Shannon's entropy (Shannon, 1948) between two classes by a factor of α for one class and α' for the other, and then they are used in an SVC (Cortes and Vapnik, 1995), optimizing keywords to select the best performing classifier using the F_1-measure (Powers, 2011). The selected SVC keywords would then clearly represent the user driving factors leading to positive and negative emotions. I also performed experiments to choose the best value of the parameter C used in the 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 can also increase false negatives. Inversely, high C values will likely result in minimal false negatives but a possibility of false positives. SVC performance results are displayed in Tables 4.6 and 4.7. Examples of tagged sentences are shown in Table 4.8.

TABLE 4.3: Collected data and structure after price range categorizing.

Price range	Data collected	Ctrip database	Tripadvisor database
0: All Prices	Hotels	557	557
	Reviews	48,070	41,137
	Sentences	101,963	348,039
	Positive sentences	88,543	165,308
	Negative sentences	13,420	182,731
1: 0 to 2500 yen	Hotels	0	0
	Reviews	0	0
2: 2500 to 5000 yen	Hotels	0	0
	Reviews	0	0
3: 5000 to 10,000 yen	Hotels	22	22
	Reviews	452	459
	Sentences	1,108	3,988
	Positive sentences	924	1,875
	Negative sentences	184	2,113
4: 10,000 to 15,000 yen	Hotels	112	112
	Reviews	2,176	2,865
	Sentences	4,240	24,107
	Positive sentences	3,566	11,619
	Negative sentences	674	12,488
5: 15,000 to 20,000 yen	Hotels	138	138
	Reviews	7,043	4,384
	Sentences	14,726	37,342
	Positive sentences	12,775	17,449
	Negative sentences	1,951	19,893
6: 20,000 to 30,000 yen	Hotels	129	129
	Reviews	11,845	13,772
	Sentences	24,413	115,830
	Positive sentences	21,068	55,381
	Negative sentences	3,345	60,449
7: 30,000 to 50,000 yen	Hotels	83	83
	Reviews	8,283	7,001
	Sentences	17,939	58,409
	Positive sentences	15,642	28,493
	Negative sentences	2,297	29,916
8: 50,000 to 100,000 yen	Hotels	59	59
	Reviews	16,670	9,646
	Sentences	36,255	81,940
	Positive sentences	31,638	38,217
	Negative sentences	4,617	43,723
9: 100,000 to 200,000 yen	Hotels	14	14
	Reviews	1,601	3,010
	Sentences	3,282	26,423
	Positive sentences	2,930	12,274
	Negative sentences	352	14,149

TABLE 4.4: Target Parts of Speech for extraction and pairing.

Language	POS Tag	Part of Speech	Examples
Chinese target tags	NN	Noun (general)	酒店 (hotel)
	VA	Predicative Adjective (verb)	干净 的 (clean)
	JJ	Noun modifier (adjectives)	干净 (clean)
	VV	Verb (general)	推荐 (recommend)
English target tags	NN	Noun (general)	room
	NNS	Noun (plural)	beds
	JJ	Adjective	big
	JJS	Adjective (superlative)	best
	JJR	Adjective (comparative)	larger
	VB	Verb (base form)	take
	VBP	Verb (single present)	take
	VBN	Verb (past participle)	taken
	VBG	Verb (gerund / present participle)	taking

TABLE 4.5: Filtered out Parts of Speech to aid pairing.

Language	POS Tag	Part of Speech	Examples
Commonly filtered tags	DT	Determiner	a, an
	PN	Pronoun	I, you, they
	CD	Cardinal Number	1, 2, 3, 4, 5
	PU	Punctuation	.!?
Chinese filtered tags	DEV	Particle	地 (Japan) (adverbial particle)
	NR	Noun (proper noun)	日本 (Japan)
	M	Measure word	个 (general classifier), 公里 (kilometer)
	SP	Sentence-final particle	他 (he), 好 (good)
	IJ	Interjection	啊 (ah)
English target tags	NNP	Noun (proper noun)	Japan
	PRP$	Possessive Pronoun	My, your, her, his
	WP	Wh-pronoun	What, who

TABLE 4.6: 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 ($\alpha = 2.75, \alpha' = 3.75$)	**Satisfaction**	**0.5**	**0.95**	**0.01**

TABLE 4.7: 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 ($\alpha = 1.5, \alpha' = 4.25$)	**Satisfaction**	**2**	**0.83**	**0.02**

TABLE 4.8: 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. My room was labeled 'non-smoking' but my duvet reeked of smoke.
		A bit pricey though

Shannon's entropy can be used to observe the probability distribution of each word inside the corpus. A word 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.

An SVC is trained to classify data based on previously labeled data, generalizing the data's features by defining a separating $(p-1)$-dimensional hyperplane in p-dimensional space. Each dimension is a feature of the data in this space. The separating hyperplane, along with the support vectors, divides the multi-dimensional space and minimizes classification error.

My study used a linear kernel for the SVC, as explained in Chapter 2 by the formula (2.1). Each training sentence is a data point, a row in the vector x. 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 hyperplane angle. The bias coefficient b determines its position.

During the SVC learning algorithm, each data point classified incorrectly alters the weight vector to correctly classify new data. These changes to the weight vector are greater for features close to the separating hyperplane. These features have stronger changes because they needed to be taken into account to classify with a minimal error. Sequentially, the weight vector shown in Chapter 2 in formula (2.5) can be interpreted as a numerical representation of each feature's effect on each class's classification process.

I tagged 159 Chinese sentences and 2357 English sentences as positive or negative for my training data. The entropy comparison factors α and α' were tested from 1.25 to 6 in intervals of 0.25. I applied this SVC to classify the rest of my data collection. Subsequently, the positive and negative sentence counts shown in Table 4.3 result from applying my SVC for classification.

4.4 Data Analysis

4.4.1 Frequent keywords in differently priced hotels

I observed the top 10 satisfaction and dissatisfaction keywords with the highest frequencies of emotionally positive and negative statements to study. The keywords are the quantitative rank of the needs of Chinese and English-speaking customers. I show the top 10 positive keywords for each price range comparing English and Chinese in Table 4.9. For the negative keywords, I show the results in Table 4.10.

I can observe that the most used keywords for most price ranges in the same language are similar, with a few changes in priority for the keywords involved. For example, in Chinese, I can see that the customers praise cleanliness first in cheaper hotels, whereas the size of the room or bed is praised more in hotels of higher class. Another example is that in negative English reviews, complaints about price appear only after 10,000 yen hotels. After this, it climbs in importance following the increase in the hotel's price.

4.4.2 Frequently used adjectives and their pairs

Some keywords in these lists are adjectives, such as the word "大 (big)" mentioned before. To understand those, I performed the dependency parsing and part of speech tagging explained in section 4.3.2. While many of these connections, I only considered the top 4 used keyword connections per adjective per price range. I show the most used Chinese adjectives in positive keywords in Table 4.11, and for negative Chinese adjective keywords in Table 4.12. Similarly, for English adjectives used in positive sentences I show the most common examples in Table 4.13, and for adjectives used in negative sentences in Table 4.14.

TABLE 4.9: English and Chinese comparison of the top 10 positive keywords.

Price range	Chinese keyword	Counts in Ctrip	English keyword	Counts in Tripadvisor
0: All Prices	不错 (not bad)	12892	good	19148
	大 (big)	9844	staff	16289
	干净 (clean)	6665	great	16127
	交通 (traffic)	6560	location	11838
	早餐 (breakfast)	5605	nice	11615
	近 (near)	5181	clean	9064
	地铁 (subway)	4321	helpful	5846
	购物 (shopping)	4101	excellent	5661
	推荐 (recommend)	3281	comfortable	5625
	环境 (environment)	3258	friendly	5606
3: 5000 to 10,000 yen	不错 (not bad)	139	good	206
	干净 (clean)	114	staff	181
	早餐 (breakfast)	112	clean	174
	大 (big)	76	nice	166
	交通 (traffic)	72	great	143
	地铁 (subway)	66	location	91
	近 (near)	55	comfortable	79
	地铁站 (subway station)	51	helpful	70
	远 (far)	41	friendly	64
	附近 (nearby)	34	recommend	59
4: 10,000 to 15,000 yen	不错 (not bad)	601	good	1399
	干净 (clean)	455	staff	1165
	大 (big)	348	great	961
	近 (near)	323	nice	808
	早餐 (breakfast)	270	location	800
	卫生 (health)	201	clean	656
	交通 (traffic)	196	excellent	412
	地铁 (subway)	164	friendly	400
	远 (far)	158	helpful	393
	附近 (nearby)	150	comfortable	391
5: 15,000 to 20,000 yen	不错 (not bad)	1925	good	2242
	干净 (clean)	1348	staff	1674
	大 (big)	1277	great	1414
	交通 (traffic)	1058	clean	1204
	近 (near)	1016	nice	1175
	地铁 (subway)	801	location	1109
	早餐 (breakfast)	777	comfortable	621
	地铁站 (subway station)	639	friendly	615
	附近 (nearby)	572	free	581
	购物 (shopping)	516	helpful	552
6: 20,000 to 30,000 yen	不错 (not bad)	3110	good	6550
	大 (big)	2245	staff	5348
	交通 (traffic)	1990	great	5074
	干净 (clean)	1940	location	4414
	近 (near)	1433	nice	3451
	地铁 (subway)	1073	clean	3364
	早餐 (breakfast)	1007	shopping	1992
	购物 (shopping)	979	helpful	1970
	周边 (surroundings)	837	comfortable	1941
	附近 (nearby)	825	friendly	1915
7: 30,000 to 50,000 yen	不错 (not bad)	2291	good	3407
	大 (big)	1913	staff	2867
	干净 (clean)	1159	great	2620
	交通 (traffic)	1105	location	2186
	近 (near)	935	nice	2160
	早餐 (breakfast)	846	clean	1750
	推荐 (recommend)	638	helpful	1147
	购物 (shopping)	636	train	1040
	周边 (surroundings)	552	subway	1034
	环境 (environment)	541	friendly	1001
8: 50,000 to 100,000 yen	不错 (not bad)	4451	great	4425
	大 (big)	3670	good	4350
	早餐 (breakfast)	2422	staff	3777
	交通 (traffic)	2012	nice	2991
	购物 (shopping)	1764	location	2439
	新 (new)	1634	clean	1655
	棒 (great)	1626	excellent	1555
	地铁 (subway)	1604	helpful	1313
	干净 (clean)	1577	comfortable	1246
	近 (near)	1354	friendly	1238
9: 100,000 to 200,000 yen	不错 (not bad)	375	great	1488
	大 (big)	315	staff	1277
	棒 (great)	189	good	994
	早餐 (breakfast)	171	nice	864
	环境 (environment)	157	location	799
	交通 (traffic)	127	excellent	631
	选择 (select)	112	beautiful	455
	推荐 (recommend)	109	large	404
	赞 (awesome)	101	helpful	401
	购物 (shopping)	98	wonderful	372

TABLE 4.10: English and Chinese comparison of the top 10 negative keywords.

Price range	Chinese keyword	Counts in Ctrip	English keyword	Counts in Tripadvisor
0: All Prices	价格 (price)	1838	pricey	462
	一般 (general)	1713	poor	460
	中文 (Chinese)	733	dated	431
	地理 (geography)	691	disappointing	376
	距离 (distance)	434	worst	327
	陈旧 (obsolete)	319	minor	258
	老 (old)	297	uncomfortable	253
	华人 (Chinese)	15	carpet	240
			annoying	220
			sense	220
3: 5000 to 10,000 yen	价格 (price)	31	worst	6
	一般 (general)	28	walkway	5
	距离 (distance)	11	unable	4
	地理 (geography)	10	worse	4
	中文 (Chinese)	9	annoying	3
	老 (old)	2	dirty	3
			funny smell	3
			poor	3
			renovation	3
			carpet	2
4: 10,000 to 15,000 yen	价格 (price)	98	dated	40
	一般 (general)	91	poor	29
	距离 (distance)	43	disappointing	26
	陈旧 (obsolete)	34	worst	24
	地理 (geography)	31	uncomfortable	23
	老 (old)	30	cigarette	22
	中文 (Chinese)	26	pricey	22
			minor	21
			paper	19
			unable	19
5: 15,000 to 20,000 yen	价格 (price)	296	poor	57
	一般 (general)	218	dated	41
	地理 (geography)	125	disappointing	38
	中文 (Chinese)	93	annoying	36
	距离 (distance)	84	worst	36
	陈旧 (obsolete)	43	cigarette	31
	老 (old)	26	rude	28
	华人 (Chinese)	3	uncomfortable	26
			paper	25
			pricey	24
6: 20,000 to 30,000 yen	一般 (general)	504	poor	136
	价格 (price)	472	dated	131
	地理 (geography)	164	pricey	120
	中文 (Chinese)	155	disappointing	112
	距离 (distance)	116	uncomfortable	103
	陈旧 (obsolete)	75	minor	93
	老 (old)	55	smallest	88
	华人 (Chinese)	2	worst	86
			cigarette	79
			annoying	70
7: 30,000 to 50,000 yen	价格 (price)	326	poor	92
	一般 (general)	311	pricey	92
	地理 (geography)	110	dated	65
	中文 (Chinese)	94	worst	64
	陈旧 (obsolete)	71	carpet	55
	距离 (distance)	68	uncomfortable	55
	老 (old)	45	dirty	51
	华人 (Chinese)	2	disappointing	50
			cigarette	46
			unable	43
8: 50,000 to 100,000 yen	价格 (price)	561	pricey	163
	一般 (general)	510	dated	150
	中文 (Chinese)	337	disappointing	129
	地理 (geography)	239	poor	124
	老 (old)	134	worst	98
	距离 (distance)	97	walkway	82
	陈旧 (obsolete)	90	carpet	71
	华人 (Chinese)	8	minor	63
			sense	63
			outdated	58
9: 100,000 to 200,000 yen	价格 (price)	54	pricey	40
	一般 (general)	51	sense	34
	中文 (Chinese)	19	minor	33
	距离 (distance)	15	lighting	20
	地理 (geography)	12	disappointing	19
	陈旧 (obsolete)	6	poor	19
	老 (old)	5	annoying	16
			mixed	15
			disappointment	14
			paper	14

TABLE 4.11: Top 4 words related to the mainly used adjectives in positive Chinese texts.

Price range	不错 (not bad)	大 (big)	干净 (clean)	近 (near)	新 (new)	棒 (great)
0: All Prices	不错 (not bad) : 12892 不错 酒店 (nice hotel) : 1462 不错 位置 (nice location) : 1426 不错 服务 (nice service) : 869 不错 环境 (nice environment) : 714	大 (big) : 9844 大 房间 (big room) : 3197 大 床 (big bed) : 772 大 酒店 (big hotel) : 379 大 超市 (big supermarket) : 232	干净 (clean) : 6665 干净 房间 (clean room) : 1224 干净 酒店 (clean hotel) : 737 干净 卫生 (clean and hygienic) : 464 干净 环境 (clean environment) : 61	近 (near) : 5181 近 酒店 (near hotel) : 453 近 桥 (near bridge) : 144 近 地铁站 (near subway station) : 122 近 站 (near station) : 108	新 (new) : 2775 新 设施 (new facility) : 363 新 酒店 (new hotel) : 246 新 装修 (new decoration) : 116 新 房间 (new room) : 53	棒 (great) : 3028 棒 酒店 (great hotel) : 463 棒 位置 (great position) : 218 棒 服务 (great service) : 168 棒 早餐 (great breakfast) : 164
3: 5000 to 10,000 yen	不错 (not bad) : 139 不错 酒店 (nice hotel) : 17 不错 位置 (nice location) : 16 不错 早餐 (nice breakfast) : 12 不错 服务 (nice service) : 8	大 (big) : 76 大 房间 (big room) : 11 大 床 (big bed) : 10 大 超市 (big supermarket) : 5 大 商场 (big market) : 3	干净 (clean) : 114 干净 房间 (clean room) : 21 干净 酒店 (clean hotel) : 10 干净 卫生 (clean and hygienic) : 6 干净 总体 (clean overall) : 4	近 (near) : 55 近 酒店 (near hotel) : 4 近 地铁 (near subway) : 2		棒 (great) : 11 棒 位置 (great position) : 2
4: 10,000 to 15,000 yen	不错 (not bad) : 601 不错 位置 (nice location) : 72 不错 酒店 (nice hotel) : 37 不错 服务 (nice service) : 34 不错 早餐 (nice breakfast) : 26	大 (big) : 348 大 房间 (big room) : 76 大 床 (big bed) : 30 大 社 (big club) : 26 大 空间 (big space) : 16	干净 (clean) : 455 干净 房间 (clean room) : 66 干净 卫生 (clean and hygienic) : 52 干净 酒店 (clean hotel) : 48 干净 打扫 (clean up) : 9	近 (near) : 323 近 酒店 (near hotel) : 27 近 站 (near station) : 14 近 地铁 (near subway) : 12 近 车站 (near the station) : 10	新 (new) : 37 新 设施 (new facility) : 9 新 装修 (new decoration) : 2 新 酒店 (new hotel) : 2	棒 (great) : 73 棒 位置 (great position) : 6 棒 房间 (great room) : 3 棒 水平 (great level) : 3 棒 温泉 (great hot spring) : 3
5: 15,000 to 20,000 yen	不错 (not bad) : 1925 不错 位置 (nice location) : 207 不错 酒店 (nice hotel) : 168 不错 服务 (nice service) : 131 不错 早餐 (nice breakfast) : 109	大 (big) : 1277 大 房间 (big room) : 316 大 床 (big bed) : 140 大 超市 (big supermarket) : 73 大 酒店 (big hotel) : 49	干净 (clean) : 1348 干净 房间 (clean room) : 234 干净 酒店 (clean hotel) : 161 干净 卫生 (clean and hygienic) : 92 干净 设施 (clean facilities) : 19	近 (near) : 1016 近 酒店 (near hotel) : 82 近 站 (near station) : 35 近 地铁站 (near subway station) : 34 近 桥 (near bridge) : 29	新 (new) : 234 新 设施 (new facility) : 47 新 酒店 (new hotel) : 25 新 装修 (new decoration) : 15 新 房间 (new room) : 10	棒 (great) : 241 棒 位置 (great position) : 33 棒 酒店 (great hotel) : 25 棒 服务 (great service) : 22 棒 早餐 (great breakfast) : 8
6: 20,000 to 30,000 yen	不错 (not bad) : 3110 不错 位置 (nice location) : 409 不错 酒店 (nice hotel) : 326 不错 服务 (nice service) : 206 不错 环境 (nice environment) : 183	大 (big) : 2245 大 房间 (big room) : 680 大 床 (big bed) : 198 大 酒店 (big hotel) : 102 大 空间 (big space) : 64	干净 (clean) : 1940 干净 房间 (clean room) : 360 干净 酒店 (clean hotel) : 203 干净 卫生 (clean and hygienic) : 137 干净 环境 (clean environment) : 21	近 (near) : 1433 近 酒店 (near hotel) : 164 近 地铁 (near subway) : 34 近 地铁站 (near subway station) : 31 近 车站 (near the station) : 27	新 (new) : 517 新 设施 (new facility) : 89 新 酒店 (new hotel) : 51 新 装修 (new decoration) : 24 新 房间 (new room) : 10	棒 (great) : 440 棒 酒店 (great hotel) : 51 棒 位置 (great position) : 45 棒 服务 (great service) : 23 棒 早餐 (great breakfast) : 20
7: 30,000 to 50,000 yen	不错 (not bad) : 2291 不错 位置 (nice location) : 277 不错 酒店 (nice hotel) : 274 不错 服务 (nice service) : 140 不错 环境 (nice environment) : 140	大 (big) : 1913 大 房间 (big room) : 643 大 床 (big bed) : 141 大 超市 (big supermarket) : 74 大 酒店 (big hotel) : 66	干净 (clean) : 1159 干净 房间 (clean room) : 224 干净 酒店 (clean hotel) : 146 干净 卫生 (clean and hygienic) : 71 干净 环境 (clean environment) : 16	近 (near) : 935 近 酒店 (near hotel) : 80 近 站 (near station) : 24 近 桥 (near bridge) : 20 近 山 (near mountain) : 12	新 (new) : 260 新 设施 (new facility) : 63 新 酒店 (new hotel) : 25 新 装修 (new decoration) : 15 新 房间 (new room) : 11	棒 (great) : 448 棒 酒店 (great hotel) : 68 棒 位置 (great position) : 34 棒 服务 (great service) : 24 棒 早餐 (great breakfast) : 14
8: 50,000 to 100,000 yen	不错 (not bad) : 4451 不错 酒店 (nice hotel) : 587 不错 位置 (nice location) : 415 不错 服务 (nice service) : 328 不错 早餐 (nice breakfast) : 251	大 (big) : 3670 大 房间 (big room) : 1340 大 床 (big bed) : 238 大 酒店 (big hotel) : 144 大 商场 (big market) : 88	干净 (clean) : 1577 干净 房间 (clean room) : 310 干净 酒店 (clean hotel) : 161 干净 卫生 (clean and hygienic) : 101 干净 服务 (clean service) : 13	近 (near) : 1354 近 酒店 (near hotel) : 88 近 桥 (near bridge) : 76 近 地铁站 (near subway station) : 35 近 铁 (Kintetsu) : 24	新 (new) : 1634 新 设施 (new facility) : 141 新 酒店 (new hotel) : 123 新 装修 (new decoration) : 57 新 斋 (new) : 22	棒 (great) : 1626 棒 酒店 (great hotel) : 281 棒 早餐 (great breakfast) : 112 棒 位置 (great position) : 96 棒 服务 (great service) : 86
9: 100,000 to 200,000 yen	不错 (not bad) : 375 不错 酒店 (nice hotel) : 53 不错 位置 (nice location) : 30 不错 环境 (nice environment) : 27 不错 服务 (nice service) : 22	大 (big) : 315 大 房间 (big room) : 131 大 面积 (large area) : 19 大 床 (big bed) : 15 大 卫生间 (big toilet) : 13	干净 (clean) : 72 干净 房间 (clean room) : 9 干净 酒店 (clean hotel) : 8 干净 卫生 (clean and hygienic) : 5	近 (near) : 65 近 酒店 (near hotel) : 8 近 地铁站 (near subway station) : 3 近 市场 (near market) : 3	新 (new) : 77 新 酒店 (new hotel) : 19 新 设施 (new facility) : 13 新 装修 (new decoration) : 3 新 位置 (new location) : 2	棒 (great) : 189 棒 酒店 (great hotel) : 36 棒 体验 (great experience) : 10 棒 服务 (great service) : 10 棒 早餐 (great breakfast) : 8

TABLE 4.12: Top 4 words related to the mainly used adjectives in negative texts.

Price range	一般 **(general)**	陈旧 **(obsolete)**	老 **(old)**
0: All Prices	一般 (general) : 1713 一般 设施 (general facilities) : 137 一般 服务 (general service) : 115 一般 酒店 (average hotel) : 106 一般 早餐 (average breakfast) : 97	陈旧 (obsolete) : 319 陈旧 设施 (obsolete facilities) : 184 陈旧 设备 (obsolete equipment) : 18 陈旧 房间 (outdated room) : 10 陈旧 酒店 (outdated hotel) : 10	老 (old) : 297 老 酒店 (old hotel) : 74 老 设施 (old facility) : 58 老 店 (old shop) : 15 老 装修 (old decoration) : 11
3: 5000 to 10,000 yen	一般 (general) : 28 一般 设施 (general facilities) : 5 一般 早餐 (average breakfast) : 3 一般 味道 (general taste) : 2 一般 效果 (general effect) : 2		老 (old) : 2
4: 10,000 to 15,000 yen	一般 (general) : 91 一般 设施 (general facilities) : 10 一般 位置 (general location) : 8 一般 酒店 (average hotel) : 6 一般 早餐 (average breakfast) : 5	陈旧 (obsolete) : 34 陈旧 设施 (obsolete facilities) : 17 陈旧 家具 (obsolete furniture) : 2 陈旧 设备 (obsolete equipment) : 2	老 (old) : 30 老 酒店 (old hotel) : 8 老 设施 (old facility) : 7 老 建筑 (old building) : 3
5: 15,000 to 20,000 yen	一般 (general) : 218 一般 设施 (general facilities) : 23 一般 酒店 (average hotel) : 21 一般 早餐 (average breakfast) : 14 一般 卫生 (general hygiene) : 8	陈旧 (obsolete) : 43 陈旧 设施 (obsolete facilities) : 25 陈旧 设备 (obsolete equipment) : 3 陈旧 酒店 (outdated hotel) : 2	老 (old) : 26 老 酒店 (old hotel) : 11 老 设施 (old facility) : 7 老 外观 (old appearance) : 2
6: 20,000 to 30,000 yen	一般 (general) : 504 一般 设施 (general facilities) : 42 一般 酒店 (average hotel) : 37 一般 服务 (general service) : 34 一般 早餐 (average breakfast) : 21	陈旧 (obsolete) : 75 陈旧 设施 (obsolete facilities) : 42 陈旧 设备 (obsolete equipment) : 7 陈旧 装修 (old decoration) : 3 陈旧 酒店 (outdated hotel) : 2	老 (old) : 55 老 酒店 (old hotel) : 9 老 设施 (old facility) : 8 老 店 (old shop) : 3 老 房间 (old room) : 3
7: 30,000 to 50,000 yen	一般 (general) : 311 一般 设施 (general facilities) : 23 一般 服务 (general service) : 22 一般 早餐 (average breakfast) : 19 一般 酒店 (average hotel) : 15	陈旧 (obsolete) : 71 陈旧 设施 (obsolete facilities) : 43 陈旧 设备 (obsolete equipment) : 5 陈旧 房间 (outdated room) : 3	老 (old) : 45 老 酒店 (old hotel) : 11 老 设施 (old facility) : 7 老 店 (old shop) : 3 老 房间 (old room) : 2
8: 50,000 to 100,000	一般 (general) : 510 一般 服务 (general service) : 39 一般 设施 (general facilities) : 32 一般 早餐 (average breakfast) : 30 一般 酒店 (average hotel) : 25	陈旧 (obsolete) : 90 陈旧 设施 (obsolete facilities) : 53 陈旧 房间 (outdated room) : 5 陈旧 感觉 (Stale feeling) : 2	老 (old) : 134 老 酒店 (old hotel) : 34 老 设施 (old facility) : 26 老 装修 (old decoration) : 9 老 店 (old shop) : 7
9: 100,000 to 200,000	一般 (general) : 51 一般 服务 (general service) : 7 一般 早餐 (average breakfast) : 5 一般 位置 (general location) : 2 一般 房间 (average room) : 2	陈旧 (obsolete) : 6 陈旧 设施 (obsolete facilities) : 4	老 (old) : 5 老 设施 (old facility) : 2

TABLE 4.13: Top 4 words related to the mainly used adjectives in positive English texts.

Price range	**good**	**clean**	**comfortable**	**helpful**	**free**	**large**	**firendly**	**great**
0: All Prices	good : 19148 good location : 1985 good service : 1042 good breakfast : 942 good hotel : 874	clean : 9064 clean room : 3596 clean hotel : 969 clean bathroom : 282 clean everything : 200	comfortable : 5625 comfortable bed : 1919 comfortable room : 1098 comfortable stay : 272 comfortable hotel : 238	helpful : 5846 helpful staff : 2927 helpful concierge : 304 helpful desk : 110 helpful service : 74	free : 4318 free wifi : 773 free shuttle : 286 free drink : 234 free bus : 225	large : 4104 large room : 1256 large hotel : 268 large bathroom : 202 larger room : 192	friendly : 5606 friendly staff : 3819 friendly service : 169 friendly hotel : 73 friendly person : 63	great : 16127 great location : 2313 great view : 1099 great service : 841 great hotel : 802
3: 5000 to 10,000 yen	good : 206 good location : 30 good value : 19 good english : 10 good place : 7	clean : 174 clean room : 55 clean bathroom : 14 clean place : 12 clean hotel : 6	comfortable : 79 comfortable bed : 21 comfortable room : 9 comfortable futon : 8 comfortable stay : 3	helpful : 70 helpful staff : 36	free : 35 free wifi : 10 free tea : 4 free raman : 2 free toothbrush : 2	large : 31 large room : 7 large area : 2 large size : 2	friendly : 64 friendly staff : 53 friendly everyone : 2 friendly service : 2	great : 143 great location : 21 great view : 14 great place : 13 great experience : 5
4: 10,000 to 15,000 yen	good : 1399 good location : 159 good breakfast : 87 good hotel : 71 good service : 67	clean : 656 clean room : 247 clean hotel : 74 clean bathroom : 20 clean everything : 14	comfortable : 391 comfortable bed : 123 comfortable room : 90 comfortable hotel : 26 comfortable stay : 20	helpful : 393 helpful staff : 206 helpful concierge : 20 helpful desk : 10 helpful service : 4	free : 271 free wifi : 53 free breakfast : 15 free service : 12 free drink : 11	large : 250 large room : 84 large bathroom : 20 larger room : 12 large hotel : 10	friendly : 400 friendly staff : 292 friendly service : 15 friendly hotel : 7 friendly person : 6	great : 961 great location : 158 great service : 51 great hotel : 43 great place : 35
5: 15,000 to 20,000 yen	good : 2242 good location : 242 good hotel : 116 good breakfast : 113 good service : 108	clean : 1204 clean room : 440 clean hotel : 133 clean bathroom : 38 clean everything : 26	comfortable : 621 comfortable bed : 219 comfortable room : 99 comfortable stay : 30 comfortable hotel : 20	helpful : 552 helpful staff : 301 helpful desk : 11 helpful concierge : 9 helpful reception : 5	free : 581 free wifi : 109 free shuttle : 35 free bus : 30 free breakfast : 27	large : 349 large room : 85 large suitcase : 18 larger room : 18 large hotel : 17	friendly : 615 friendly staff : 444 friendly hotel : 12 friendly service : 8 friendly most : 7	great : 1414 great location : 199 great view : 81 great hotel : 68 great place : 61
6: 20,000 to 30,000 yen	good : 6550 good location : 703 good service : 331 good english : 304 good breakfast : 303	clean : 3364 clean room : 1379 clean hotel : 379 clean bathroom : 95 clean everything : 77	comfortable : 1941 comfortable bed : 658 comfortable room : 359 comfortable stay : 100 comfortable hotel : 82	helpful : 1970 helpful staff : 1019 helpful concierge : 79 helpful desk : 42 helpful receptionist : 17	free : 1186 free wifi : 269 free breakfast : 68 free coffee : 57 free drink : 38	large : 1257 large room : 329 large hotel : 87 larger room : 81 large bed : 43	friendly : 1915 friendly staff : 1311 friendly service : 51 friendly person : 21 friendly hotel : 19	great : 5074 great location : 881 great service : 249 great hotel : 232 great view : 220
7: 30,000 to 50,000 yen	good : 3407 good location : 380 good breakfast : 191 good service : 182 good english : 155	clean : 1750 clean room : 725 clean hotel : 197 clean bathroom : 61 clean everything : 36	comfortable : 1000 comfortable bed : 345 comfortable room : 193 comfortable hotel : 49 comfortable stay : 47	helpful : 1147 helpful staff : 607 helpful concierge : 53 helpful service : 20 helpful desk : 17	free : 933 free drink : 145 free wifi : 129 free coffee : 45 free bus : 38	large : 580 large room : 174 larger room : 32 large hotel : 30 large bed : 28	friendly : 1001 friendly staff : 715 friendly service : 24 friendly hotel : 13 friendly person : 13	great : 2620 great location : 393 great view : 162 great hotel : 134 great service : 114
8: 50,000 to 100,000 yen	good : 4350 good location : 406 good service : 296 good hotel : 196 good breakfast : 191	clean : 1655 clean room : 648 clean hotel : 156 clean bathroom : 48 cleanliness : 40	comfortable : 1246 comfortable bed : 425 comfortable room : 266 comfortable stay : 56 comfortable hotel : 51	helpful : 1313 helpful staff : 589 helpful concierge : 108 helpful service : 28 helpful desk : 26	free : 1072 free shuttle : 181 free wifi : 172 free bus : 127 free service : 65	large : 1233 large room : 442 large hotel : 109 large bathroom : 58 larger room : 38	friendly : 1238 friendly staff : 810 friendly service : 51 friendly hotel : 20 friendly person : 12	great : 4425 great location : 506 great view : 436 great service : 267 great hotel : 241
9: 100,000 to 200,000 yen	good : 994 good location : 65 good service : 56 good breakfast : 53 good hotel : 40	clean : 261 clean room : 102 clean hotel : 24 cleanliness : 8 clean place : 7	comfortable : 347 comfortable bed : 128 comfortable room : 82 comfortable stay : 16 comfortable hotel : 10	helpful : 401 helpful staff : 169 helpful concierge : 35 helpful everyone : 7 helpful team : 5	free : 240 free wifi : 31 free breakfast : 19 free drink : 16 free bus : 14	large : 404 large room : 135 large bathroom : 38 large hotel : 15 large bed : 12	friendly : 370 friendly staff : 194 friendly service : 18 friendly everyone : 7 friendly person : 4	great : 1488 great location : 155 great view : 155 great service : 101 great hotel : 80

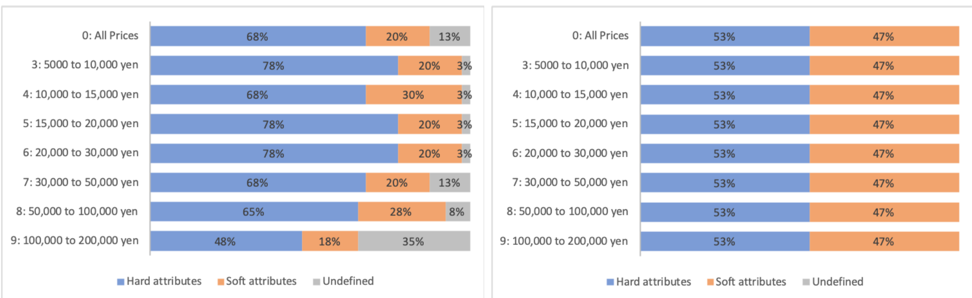
TABLE 4.14: Top 4 words related to the mainly used adjectives in negative English texts.

Price range	poor	dated	worst	dirty	uncomfortable
0: All Prices	poor : 460 poor service : 55 poor breakfast : 41 poor quality : 27 poor english : 24	dated : 431 outdated : 128 outdated room : 20 outdated hotel : 10 outdated bathroom : 7	worst : 327 worst hotel : 43 worst experience : 18 worst part : 15 worst service : 10	dirty : 188 dirty carpet : 34 dirty room : 23 not dirty : 7 dirty bathroom : 6	uncomfortable : 253 uncomfortable bed : 63 uncomfortable pillow : 20 uncomfortable mattress : 8 uncomfortable night : 8
3: 5000 to 10,000 yen	poor : 3		worst : 6 worst room : 2	dirty : 3	uncomfortable : 2
4: 10,000 to 15,000 yen	poor : 29 poor breakfast : 3 poor service : 3 poor conditioning : 2 poor view : 2	dated : 40 outdated : 11 outdated decor : 2 outdated room : 2	worst : 24 worst hotel : 4 worst experience : 2	dirty : 11 dirty floor : 2	uncomfortable : 23 uncomfortable bed : 4 not uncomfortable : 2 uncomfortable night : 2 uncomfortable pillow : 2
5: 15,000 to 20,000 yen	poor : 57 poor service : 10 poor breakfast : 6 poor hotel : 5 poor experience : 3	dated : 41 outdated : 8	worst : 36 worst hotel : 8 worst experience : 3 worst part : 2 worst service : 2	dirty : 14 dirty room : 2	uncomfortable : 26 uncomfortable bed : 7 uncomfortable pillow : 2
6: 20,000 to 30,000 yen	poor : 136 poor breakfast : 15 poor service : 14 poor english : 9 poor quality : 9	dated : 131 outdated : 31 outdated room : 6 outdated hotel : 2	worst : 86 worst hotel : 11 worst part : 7 worst breakfast : 5 worst experience : 5	dirty : 67 dirty room : 10 dirty carpet : 8 dirty bathroom : 3 dirty chair : 2	uncomfortable : 103 uncomfortable bed : 24 uncomfortable pillow : 11 uncomfortable night : 4 uncomfortable experience : 3
7: 30,000 to 50,000 yen	poor : 92 poor service : 8 poor breakfast : 7 poor english : 7 poor connection : 5	dated : 65 outdated : 17 outdated hotel : 4 outdated bathroom : 2 outdated decor : 2	worst : 64 worst hotel : 10 worst room : 3 worst service : 3 worst part : 2	dirty : 51 dirty carpet : 11 dirty room : 7 dirty clothe : 2 dirty luggage : 2	uncomfortable : 55 uncomfortable bed : 20 uncomfortable mattress : 6 uncomfortable pillow : 5 uncomfortable room : 5
8: 50,000 to 100,000 yen	poor : 124 poor service : 16 poor breakfast : 9 poor quality : 9 poor english : 6	dated : 150 outdated : 58 outdated room : 9 outdated furniture : 6 outdated hotel : 4	worst : 98 worst hotel : 9 worst experience : 5 worst part : 3	dirty : 36 dirty carpet : 12 dirty room : 3 dirty cup : 2 dirty rug : 2	uncomfortable : 33 uncomfortable bed : 7
9: 100,000 to 200,000 yen	poor : 19 poor service : 4 poor choice : 2 poor experience : 2	dated : 3 outdated : 2	worst : 12 worst experience : 2	dirty : 6	uncomfortable : 8 little uncomfortable : 2

4.4.3 Determining hard and soft attribute usage

To further understand the differences in satisfaction and dissatisfaction in Chinese and Western customers of Japanese hotels, I classified these factors as either hard or soft attributes of a hotel. I define hard attributes as matters regarding the hotel's physical or environmental aspects, such as facilities, location, or infrastructure. Some of these aspects would be impractical for the hotel to change, such as its surroundings and location. Others can be expensive to change, such as matters requiring construction costs, which are possible but would require significant infrastructure investment. On the other hand, soft attributes are the non-physical attributes of the hotel service and staff behavior that are practical to change through management. For example, the hotel's services or the cleanliness of the rooms are soft attributes. For my purposes, amenities, clean or good quality bed sheets or curtains, and other physical attributes that are part of the service and not the hotel's physical structure are considered soft attributes. Thus, I can observe the top 10 satisfaction and dissatisfaction keywords and determine whether they are soft or hard attributes.

I manually labeled each language's top keywords into either hard or soft by considering how the word would be used when writing a review. If the word described unchangeable physical factors by the staff or management, I consider them hard. If the word implied an issue that could be solved or managed by the hotel staff or management, I consider it soft. For adjectives, I looked at the top four adjective and noun pairings used in the entire dataset and counted the usage percentage in each context. If it was not clear from the word or the pairing alone, I declared it undefined. Then, I added the counts of these words in each category. A single word with no pairing is always deemed 100% in the category it corresponds to. I add the partial percentages for each category when an adjective includes various contexts. The interpretation of these keywords is shown in the Tables 4.15 and 4.16. I can see the summarized results for the hard and soft percentages of positive and negative Chinese keywords in Figure 4.5. For the English keywords, see Figure 4.6.



(A) Positive keywords (B) Negative keywords

FIGURE 4.5: Hard and soft attributes from the top Chinese keywords for all price ranges

4.5 Results

4.5.1 Experimental results and answers to research questions

My research questions were related to two issues. Based on research questions 4.2a and 4.2b, the objective of this study was to determine the differences in how Chinese and Western tourists perceive Japanese hotels, whose hospitality and service are influenced by the *omotenashi* culture.

TABLE 4.15: Determination of hard and soft attributes for Chinese keywords.

Keyword Emotion	Keyword	Attribute Category
Positive Keywords	不错	50% hard, 25% soft, 25% undefined
	大	100% hard
	干净	25% hard, 75% soft
	早餐	100% soft
	交通	100% hard
	棒	25% hard, 50% soft, 25% undefined
	近	100% hard
	购物	100% hard
	环境	100% hard
	地铁	100% hard
	卫生	100% soft
	新	50% hard, 25% soft, 25% undefined
	推荐	100% undefined
	选择	100% undefined
	地铁站	100% hard
	远	100% hard
	附近	100% hard
	周边	100% hard
	赞	100% undefined
Negative Keywords	价格	100% soft
	一般	50% hard, 50% soft
	中文	100% soft
	距离	100% hard
	地理	100% hard
	陈旧	100% hard
	老	75% hard, 25% soft
	华人	100% soft

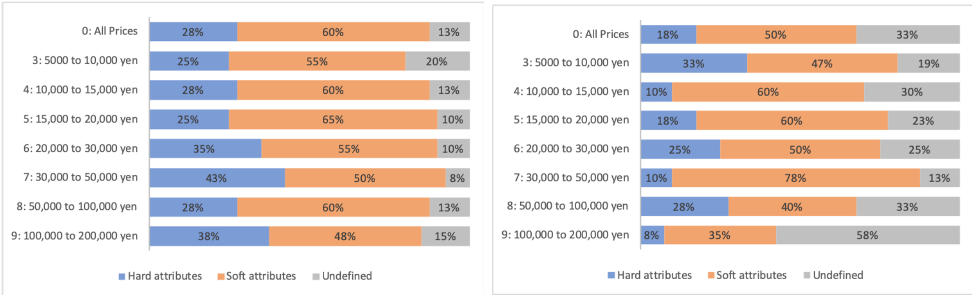
Observing the top-ranking positive keywords in Chinese reviews, as shown in Tables 4.9 and Table 4.11, it was revealed that, while service, cleanliness, and breakfast were praised in most hotels, the location was more important when observing the pairings. Hard attributes were abundant lower on the lists. The negative keywords in Table 4.10 indicate that a lack of a Chinese-friendly environment was perceived, although there were more complaints about hard attributes such as the building's age and the distance from other convenient spots. However, most complaints were about the hotel's price, which included all of the price ranges; therefore, the price was the primary concern for Chinese customers with different travel purposes.

On the other hand, the word "staff" is the second or third in the lists of satisfaction factors in English-written reviews in all the price ranges. This word is followed by a few other keywords lower in the top 10 list, such as "helpful" or "friendly". When I look at the pairings of the top-ranked keyword "good" in Table 4.13, I find that customers mostly praise the location, service, breakfast, or English availability. When I look at the negative keyword "poor" and its pairings in Table 4.14, I see that it is also service-related concepts that the Western tourists are disappointed with.

I can also observe some keywords that are not considered by their counterparts. For example, English-speaking customers mentioned tobacco smell in many reviews. However, it was not statistically identified as a problem for their Chinese counterparts. On the other hand, although they appear in both English and Chinese lists, references to "购物 (shopping)" are more common in the Chinese lists across hotels of 15,000 yen to 200,000 yen per night. Meanwhile, the term "shopping" appeared solely in the top 10 positive keywords list for English speakers who stayed in rooms

TABLE 4.16: Determination of hard and soft attributes for English keywords.

Keyword Emotion	Keyword	Attribute Category
Positive Keywords	good	25% hard, 50% soft, 25% undefined
	great	50% hard, 25% soft, 25% undefined
	staff	100% soft
	clean	100% soft
	location	100% hard
	nice	50% hard, 25% soft, 25% undefined
	excellent	25% hard, 50% soft, 25% undefined
	helpful	100% soft
	comfortable	25% hard, 50% soft, 25% undefined
	shopping	100% hard
	beautiful	25% hard, 75% soft
	friendly	100% soft
	train	100% hard
	large	100% hard
	free	100% soft
	subway	100% hard
	recommend	100% undefined
	wonderful	50% soft, 50% undefined
Negative Keywords	pricey	100% soft
	worst	25% hard, 50% soft, 25% undefined
	dated	75% hard, 25% undefined
	poor	100% soft
	walkway	100% hard
	sense	100% undefined
	unable	100% soft
	disappointing	50% soft, 50% undefined
	minor	100% undefined
	worse	100% undefined
	annoying	75% hard, 25% undefined
	lighting	100% soft
	uncomfortable	100% soft
	carpet	100% soft
	dirty	75% soft, 25% undefined
	cigarette	100% soft
	funny smell	100% soft
	rude	100% soft
	smallest	75% hard, 25% undefined
	mixed	100% undefined
	renovation	100% hard
	paper	100% undefined
	disappointment	100% undefined
	outdated	75% hard, 25% undefined



(A) Positive keywords

(B) Negative keywords

FIGURE 4.6: Hard and soft attributes from the top English keywords for all price ranges

priced 20,000–30,000 yen per night.

With these results, I can observe that both Chinese and English-speaking tourists in Japan have different priorities. However, both populations consider the hotel's location and transport availability (subways and trains) nearby as secondary but still essential points in their satisfaction with a hotel. The Chinese customers are primarily satisfied with the room quality in spaciousness and cleanliness and the service of breakfast.

For research questions 4.3a and 4.3b, I considered how customers of both cultural backgrounds evaluated the hard and soft attributes of hotels. My study discovered that Chinese tourists mostly positively react to the hotel's hard attributes, albeit the negative evaluations are more uniform than the positive evaluations, with a tendency of 53 % towards hard attributes. On the other hand, English-speaking tourists were more responsive to soft attributes, either positively or negatively. In the case of negative keywords, they were more concerned about the hotel's soft attributes.

One factor that both populations had in common is that, when perceiving the hotel negatively, the "老 (old)," "dated," "outdated," or "陈旧 (obsolete)" aspects of the room or the hotel were surprisingly criticized across most price ranges. However, this is a hard attribute and is unlikely to change for most hotels.

4.5.2 Chinese tourists: A big and clean space

I found that mainland Chinese tourists were mainly satisfied by big and clean spaces in Japanese hotels. The adjectival pairings extracted with dependency parsing and POS tagging (Table 4.11) imply big and clean rooms. Other mentions included big markets nearby or a big bed. Across different price ranges, the usage of the word "大 (big)" increased with the increasing price of the hotel. When inspecting closer by taking random samples of the pairs of "大 空间 (big space)" or "大 面积 (large area)," I notice that there were also many references to the public bathing facilities in the hotel. Such references were also implied by a word pairing "棒 温泉 (great hot spring)."

In Japan, there are the so-called "銭湯 (*sentō*)," which are artificially constructed public bathing facilities, including saunas and baths with unique qualities. On the other hand, there are natural hot springs, called "温泉 (*onsen*)." However, they are interchangeable if natural hot spring water is used in artificially made tiled bath facilities. It is a Japanese custom that all customers first clean themselves in a shower and afterward use the baths nude. It could be a cultural shock for many tourists but a fundamental attraction for many others.

Chinese customers are satisfied with the size of the room or bed; however, it is not trivial to change this. In contrast, cleanliness is mostly related to soft attributes when I observe its adjectival pairings. I can observe pairs such as “干净 房间 (clean room)” at the top rank of all price ranges and thereupon “干净 酒店 (clean hotel),” “干净 总体 (clean overall),” “干净 环境 (clean environment),” and “干净 设施 (clean facilities),” among other examples. In negative reviews, there was a mention of criticizing the “一般 卫生 (general hygiene)” of the hotel, although it was an uncommon pair. Therefore, I can assert that cleanliness was an important soft attribute for Chinese customers, and they were mostly pleased when their expectations were fulfilled.

A key soft satisfaction factor was the inclusion of breakfast within the hotel. While other food-related words were extracted, most of them were general, such as “food” or “eating,” and were lower-ranking. In contrast, the word “早餐 (breakfast),” referring to the hotel commodities, was frequently used in positive texts compared to other food-related words across all price ranges, albeit at different priorities in each of them. For this reason, I regard it as an important factor. From the word pairs of the positive Chinese keywords in Table 4.11, I can also note that “不错 (not bad)” is paired with “不错 早餐 (nice breakfast)” in four of the seven price ranges with reviews available as part of the top four pairings. It is only slightly lower in other categories, although it is not depicted in the table. Thus, I consider that a recommended strategy for hotel management is to invest in the inclusion or improvement of hotel breakfast to increase good reviews.

4.5.3 Western tourists: A friendly face and absolutely clean

From the satisfaction factors of English-speaking tourists, I observed at least three words were related to staff friendliness and services in the general database: “staff,” “helpful,” and “friendliness.” The word “staff” is the highest-ranked of these three, ranking second for satisfied customers across most price ranges and only third in one of them. The word “good” mainly refers to the location, service, breakfast, or English availability in Table 4.13. Similar to Chinese customers, Western customers also seemed to enjoy the included breakfasts regarding their satisfaction keyword pairings. However, the relevant word does not appear in the top 10 list directly, in contrast to their Chinese counterparts. The words “helpful” and “friendly” are mostly paired with “staff,” “concierge,” “desk,” and “service.” By considering the negative keyword “poor” and its pairings in Table 4.14, I realized once again that Western tourists were disappointed with service-related concepts and reacted negatively.

Another soft attribute that is high on the list for most of the price ranges is the word “clean”, so I examined its word pairings. Customers largely praised “clean rooms” and “clean bathrooms” and also referred to the hotel in general. When observing the negative keyword frequencies for English speakers, I can find words such as “dirty” and “carpet” as well as word pairings such as “dirty carpet,” “dirty room,” and dirty bathroom.” Along with complaints about off-putting smells, I could conclude that Western tourists had high expectations about cleanliness when traveling in Japan.

An interesting detail of the keyword ranking is that the word “comfortable” was high on the satisfaction factors, and “uncomfortable” was high on the dissatisfaction factors. The words were paired with nouns such as “bed,” “room,” “pillow,” and “mattress,” when they generally referred to their sleep conditions in the hotel. It seems that Western tourists were particularly sensitive about the hotels’ comfort levels and whether they reached their expectations. The ranking for the negative keyword "uncomfortable" is similar across most price ranges except the two most expensive ones, where this keyword disappears from the top 10 list.

Albeit lower in priority, the price range of 15,000 to 20,000 yen hotels also includes "free" as one of the top 10 positive keywords, mainly paired with "Wi-Fi." This price range corresponds to business hotels, where users would expect this feature the most.

4.5.4 Tobacco, an unpleasant smell in the room

A concern for Western tourists was uncleanliness and the smell of cigarettes in their room, which can be regarded as soft attributes. Cigarette smell was an issue even in the middle- and high-class hotels, of which the rooms were priced at more than 30,000 yen per night. For hotels with rooms priced above 50,000 yen per night, however, this problem seemed to disappear from the list of top 10 concerns. Tobacco was referenced singularly as "cigarette", but also in word pairs in Table 4.14 as "funny smell." By manually inspecting a sample of reviews with this keyword, I noticed that the room was often advertised as non-smoking; however, the smell permeated the room and curtains. Another common complaint was that there were no nonsmoking facilities available. The smell of smoke can completely ruin some customers' stay, leading to bad reviews, thereby lowering the number of future customers.

In contrast, Chinese customers seemed not to be bothered by this. Previous research has stated that 49–60 % of Chinese men (and 2.0–2.8 % of women) currently smoke or smoked in the past. This was derived from a sample of 170,000 Chinese adults in 2013–2014, which is high compared to many English-speaking countries (Zhang et al., 2019; World Health Organization, 2015).

Japan has a polarized view on the topic of smoking. Although it has one of the world's largest tobacco markets, tobacco use has decreased in recent years. Smoking in public spaces is prohibited in some wards of Tokyo (namely Chiyoda, Shinjuku, and Shibuya). However, it is generally only suggested and not mandatory to lift smoking restrictions in restaurants, bars, hotels, and public areas. Many places have designated smoking rooms to keep the smoke in an enclosed area and avoid bothering others.

Nevertheless, businesses, especially those who cater to certain customers, are generally discouraged by smoking restrictions if they want to maintain their clientele. To cater to all kinds of customers, including Western and Asian, Japanese hotels must provide spaces without tobacco smell. Even if the smoke does not bother a few customers, the lack of such a smell will make it an appropriate space for all customers.

4.5.5 Location, location, location

The hotel's location, closeness to the subway and public transportation, and availability of nearby shops proved to be of importance to both Chinese and English-speaking tourists. In positive word pairings in Tables 4.11 and 4.13, I can find pairs such as "不错 位置 (nice location)," "近 地铁站 (near subway station)," "近 地铁 (near subway)" in Chinese texts and "good location," "great location," and "great view" as well as single keywords "location" and "shopping" for English speakers, and "交通 (traffic)," "购物 (shopping)," "地铁 (subway)," and "环境 (environment or surroundings)" for Chinese speakers. All of these keywords and their location in each population's priorities across the price ranges signify that the hotel's location was a secondary but still important point for their satisfaction. However, since this is a hard attribute, it is not often considered in the literature. By examining examples from the data, I recognized that most customers were satisfied if the hotel was near at least two of the following facilities: subway, train, and convenience stores.

Japan is a country with a peculiar public transportation system. During rush hour, the subway is crowded with commuters, and trains and subway stations create a confusing public transportation map for a visitor in Tokyo. Buses are also available, albeit less used than rail systems in metropolitan cities. These three means of transportation are usually affordable in price. There are more expensive means, such as the bullet train *shinkansen* for traveling across the country and taxis. The latter is a luxury in Japan compared to other countries. In Japan, taxis provide a high-quality experience with a matching price. Therefore, for people under a budget, subway availability and maps or GPS applications, as well as a plan to travel the city, are of utmost necessity for tourists, using taxis only as a last resort.

Japanese convenience stores are also famous worldwide because they offer a wide range of services and products, from drinks and snacks to full meals, copy and scanning machines, alcohol, cleaning supplies, personal hygiene items, underwear, towels, and international ATMs. If some trouble occurs, or a traveler forgot to pack a particular item, it is most certain that they can find it.

Therefore, considering that both transportation systems and nearby shops are points of interest for Chinese and Western tourists, and perhaps offering guide maps and information about these as an appeal point could result in greater satisfaction.

4.6 Discussion

4.6.1 Western and Chinese tourists in the Japanese hospitality environment

To date, scholars have been correcting my historical bias towards the West. Studies have determined that different cultural backgrounds lead to different expectations, which influence tourists' satisfaction. In other words, tourists of a particular culture have different leading satisfaction factors across different destinations. However, Japan presents a particular environment; the spirit of hospitality and service, *omotenashi*, which is considered to be of the highest standard across the world. My study explores whether such an environment can affect different cultures equally or whether it is attractive only to certain cultures.

My results indicate that Western tourists are more satisfied with soft attributes than Chinese tourists. As explained earlier in this study, Japan is well known for its customer service. Respectful language and bowing are not exclusive to high-priced hotels or businesses; these are met in convenience stores as well. Even in the cheapest convenience store, the level of hospitality is starkly different from Western culture and perhaps unexpected. In higher-priced hotels, the adjectives used to praise the service ranged from normal descriptors like "good" to higher levels of praise like "wonderful staff," "wonderful experience," "excellent service," and "excellent staff." Furthermore, Kozak (2002) and Shanka and Taylor (2004) have also proven that hospitality and staff friendliness are two determinants of Western tourists' satisfaction.

However, the negative English keywords indicate that a large part of the dissatisfaction with Japanese hotels stemmed from a lack of hygiene and room cleanliness. Although Chinese customers had solely positive keywords about cleanliness, English-speaking customers deemed many places unacceptable to their standards, particularly hotels with rooms priced below 50,000 yen per night. The most common complaint regarding cleanliness was about the carpet, followed by complaints about cigarette smell and lack of general hygiene. Kozak (2002) also proved that hygiene and cleanliness were essential satisfaction determinants for Western tourists. However, in the previous literature, this was linked merely to satisfaction. In contrast, my research revealed that words related to cleanliness were mostly linked to dissatisfaction. I could assert that Westerners had a high standard of room cleanliness compared to their Chinese counterparts.

According to previous research, Western tourists are already inclined to appreciate hospitality for their satisfaction. When presented with Japanese hospitality, this expectation is met and overcome. In contrast, according to my results, Chinese tourists were more concerned about room quality rather than hospitality, staff, or service. However, when analyzing the word pairs for "不错 (not bad)" and "棒 (great)," I can see that they praise staff, service, and breakfast. By observing the percentage of hard to soft attributes in Figure 4.5, however, I discover that Chinese customers were more satisfied with hard attributes compared to Western tourists, who seemed to be meeting more than their expectations.

It could be considered that Chinese culture does not expect high-level service initially. When an expectation that is not held is met, the satisfaction derived is less than that if it was expected. In contrast, some tourists report a "nice surprise": when an unknown need is unexpectedly met, there is more satisfaction. It is necessary to note the difference between these two reactions. The "nice surprise" reaction fulfills a need unexpectedly. Perhaps the hospitality grade in Japan does not fulfill a need high enough for the Chinese population, thereby resulting in less satisfaction. For greater satisfaction, a need must be met. However, the word "not bad" is at the top of the list in most price ranges, and one of the uses is related to service. Thus, I cannot conclude that they were not satisfied with the service. Instead, they held other factors at a higher priority; thus, the keyword frequency was higher for other pairings.

Another possibility occurs when I observe the Chinese tourists' dissatisfaction factors. Chinese tourists may have expectations about their treatment that are not being met, even in this high-standard hospitality environment. This could be because Japan is monolingual and has a relatively large language barrier to tourists (Heinrich, 2012; Coulmas and Watanabe, 2002). While the Japanese effort to accommodate English speakers is slowly developing, efforts for Chinese accommodations can be lagging. Chinese language pamphlets and Chinese texts on instructions for the hotel room and its appliances and features (e.g., T.V. channels, Wi-Fi setup, etc.), or the treatment towards Chinese people, could be examples of these accommodations. Ryan and Mo (2001) also found that communication difficulty was one of the main reasons Chinese customers would state for not visiting again. However, this issue is not exclusive to Japan.

My initial question was whether the environment of high-grade hospitality would affect both cultures equally. This study attempted to determine the answer. It is possible that Chinese customers had high-grade hospitality and were equally satisfied with Westerners. In that case, it appears that the difference in perception stems from a psychological source; expectation leads to satisfaction and a lack of expectation results in lesser satisfaction. There is also a possibility that Chinese customers are not receiving the highest grade of hospitality because of cultural friction between Japan and China.

It is unclear which of these two is most likely from my results. However, competing in hospitality and service includes language services, especially in the international tourism industry. Better multilingual support can only improve the hospitality standard in Japan. Considering that most of the tourists in Japan come from other countries in Asia, multilingual support is beneficial. Proposals for this endeavor include hiring Chinese-speaking staff, preparing pamphlets in Chinese, or having a translator application readily available with staff trained in interacting through an electronic translator.

4.6.2 Hard vs. soft satisfaction factors

As stated in section 4.2.2, previous research has mostly focused on the hotel's soft attributes and their influence on customer satisfaction (e.g., Shanka and Taylor, 2004; Choi and Chu, 2001). Examples of soft attributes include staff behavior, commodities, amenities, and appliances that can be improved within the hotel. However, hard attributes are not usually analyzed in satisfaction studies. It is important to consider both kinds of attributes. If the satisfaction was based on soft attributes, a hotel can improve its services to attract more customers in the future. Otherwise, if the satisfaction was related more to hard attributes overall, hotels should be built considering the location while minimizing other costs. Because the satisfaction factors were decided statistically in my study via customers' online reviews, I can see the importance of the hard or soft attributes in their priorities.

Figure 4.5 shows that, in regards to Chinese customer satisfaction, in general, 68 % of the top 10 keywords are hard factors; in contrast, only 20 % are soft factors. The rates are similar for most price ranges except the highest-priced hotels. However, two of these soft attributes are all concentrated at the top of the list ("不错 (not bad)," "干净 (clean)"), and the adjective pairs related to soft attributes of "不错 (not bad)" are also at the top in most price ranges. Chinese tourists may expect spaciousness and cleanliness when coming to Japan. The expectation may be due to reputation, previous experiences, or cultural backgrounds. I can compare these results with previous literature, where traveling Chinese tourists choose their destination based on several factors, including cleanliness, nature, architecture, and scenery (Ryan and Mo, 2001). These factors found in previous literature could be linked to the keyword "环境 (environment or surroundings)" as well. This keyword was found for hotels priced at more than 20,000 yen per night.

In contrast, English speakers are mostly satisfied with the hotels' soft attributes. Figure 4.6 shows that soft attributes are above 48 % in all price ranges, the highest being 65 % in the price range of 15,000 to 20,000 yen per night, which corresponds to, for example, affordable business hotels. The exception to this is the hard attribute that is the hotels' location, which is consistently around the middle of the top 10 lists for all price ranges.

For both customer groups, the main reason for dissatisfaction was pricing, which can be interpreted as a concern about value for money. However, English-speaking customers complained less about the price in lower-priced hotels. In contrast, Chinese customers consistently had "价格 (price)" as the first or second-most concern across all price ranges. A study on Chinese tourists found that they had this concern (Truong and King, 2009). However, my results indicate that this has more to do with the pricing of hotels in Japan than with Chinese culture. In general, Japan is an expensive place to visit, thereby impacting this placement in the ranking. Space is scarce in Japan, and capsule hotels with cramped spaces of 2 x 1 meters cost around 3000 to 6000 yen per night. Bigger business hotel rooms are relatively expensive, ranging from 5000 to 12,000 yen per night. For comparison, hotels in the USA with a similar quality can charge half the price.

Around half of the dissatisfaction factors for both Chinese and Western customers are caused by issues that could be improved; this is true for all price ranges. The improvements could be staff training (perhaps in language), hiring professional cleaning services for rooms with cigarette smoke smells, or improving the bedding; however, these considerations can be costly. However, once the hotel's location and construction are set, only a few changes can be made to satisfy Chinese customers further. As mentioned previously, Chinese language availability is a soft attribute that can be improved with staff and training investment.

Western tourists are mainly dissatisfied with soft attributes. This is revealed by a low satisfaction level of 35 % in the highest price range where undefined factors are the majority and a maximum of 78 % in the price range of 30,000 to 50,000 yen per night in a hotel. Improvement scope for Western tourists is more extensive than that for their Chinese counterparts. As such, it presents a larger investment opportunity.

4.6.3 Satisfaction across different price ranges

In previous sections of this study, I mentioned the differences reflected in hotel price ranges. The most visible change across differently priced hotels is the change in voice when describing satisfaction. I noticed this by observing the adjective-noun pairs and finding pairs with different adjectives for the same nouns. For example, in English, words describing nouns such as "location" or "hotel" are "good" or "nice" in lower-priced hotels. In contrast, the adjectives that pair with the same nouns for higher-priced hotels are "wonderful" and "excellent." In Chinese, the change ranges from "不错 (not bad)" to "棒 (great)" or "赞 (awesome)." I can infer that the level of satisfaction is high and influences how customers write their reviews. Regarding the negative keywords, however, the change ranges from "annoying" or "disappointing" to "worst."

In this study, I follow the definition of satisfaction by Hunt (1975), where meeting or exceeding expectations produces satisfaction. Conversely, the failure of meeting expectations causes dissatisfaction. I can assume that a customer that pays more for a higher-class experience has higher expectations. For example, in a highly-priced hotel, any lack of cleanliness can lead to disappointment. In the case of English-speaking customers in the 30,000-–50,000 yen per night price range, cigarette smell is particularly disappointing. However, I consistently see customers with high expectations for high-class hotels reacting even more positively when satisfied. In the positive case, expectations appear to be exceeded in most cases, judging from their reactions.

I argue that these are two different kinds of expectations: logical and emotional. In the first case, customers are determined that the service must not fall below a specific standard; for example, they can be disappointed with unhygienic rooms or cigarette smell. In contrast, in the second case, customers have a vague idea of having a positive experience but do not measure it against any standard. For example, they expect a pleasant customer service experience or a hospitable treatment by the staff at a high-class hotel. Regardless of their knowledge in advance, positive emotions offer them a perception of exceeded expectations and high satisfaction. Thus, hospitality and service enhance the experience of the customers.

There are interesting differences between Chinese and English-speaking tourists in their satisfaction to differently priced hotels. For example, Chinese tourists have

"购物 (shopping)" as a top keyword in all the price ranges. In contrast, English-speaking tourists mention it only as a top keyword in the 20,000—30,000 yen price range. It is widely known in Japan that many Chinese tourists visit Japan for shopping. Tsujimoto (2017) analyzed the souvenir purchasing behavior of Chinese tourists in Japan and showed that common products besides food and drink are: electronics, cameras, cosmetics, and medicine, among *souvenir* items representative of the culture or places that they visit Japan Tourism Agency (2014). Furthermore, Chinese tourists' choice to shop in Japan is more related to the quality of the items rather than their relation to the tourist attractions. My results suggested that Western tourists were engaging more in tourist attractions rather than shopping activities compared to Chinese tourists.

Another interesting difference is that English-speaking tourists start using negative keywords about the hotel's price only if it concerns hotels of 15,000 yen or more; thereafter, the more expensive the hotel, the higher the ranking. In contrast, for Chinese customers, this keyword is the top keyword across all price ranges. Previous research suggests that value for money is a key concern for Chinese and Asian tourists (Choi and Chu, 2000; Choi and Chu, 2001; Truong and King, 2009), whereas Western customers are more concerned about hospitality (Kozak, 2002).

While some attributes' value changes depending on the hotel's price range, some other attributes remain constant for each culture's customers. For example, appreciation for staff from English-speaking tourists is ranked close to the top satisfaction factor in all the price ranges. Satisfaction for cleanliness by both cultures constantly remains part of the top 10 keywords, except for the most expensive one, where other keywords replace keywords related to satisfaction or cleanliness in the ranking; however, they remain still high on the list. Chinese tourists have a high ranking for the word "早餐 (breakfast)" across all price ranges as well. As discussed in section 4.5.5, transportation and location are also important for hotels of all classes and prices. While the ranking of attributes might differ between price ranges, hard and soft attribute proportions also appear to be constant within a 13 % margin of error per attribute. This suggests that, from a cultural aspect, the customers have a particular bias to consider some attributes more than others.

4.6.4 Cross-culture analysis of expectations and satisfaction

The basic premise of this study is that different cultures lead to different expectations and satisfaction factors. This premise also plays a role in the differentiation between the preferences of hard or soft attributes.

In Donthu and Yoo (1998), subjects from 10 different countries were compared with respect to their expectations of service quality and analyzed based on Hofstede's typology of culture (Hofstede, 1984). The previous study states that, although culture has no specific index, five dimensions of culture can be used to analyze or categorize a country in comparison to others. These are *power distance*, *uncertainty avoidance*, *individualism—collectivism*, *masculinity–femininity*, and *long-term–short-term orientation*. In each of these dimensions, at least one element of service expectations was found to be significantly different for countries grouped under contrasting attributes (e.g., individualistic countries vs. collectivist countries, high uncertainty avoidance countries vs. low uncertainty avoidance countries).

However, Hofstede's typology has received criticism from academics, particularly for the fifth dimension that Hofstede proposed, which was later added with the alternative name *Confucian dynamic*. Academics with a Chinese background criticized Hofstede for being misinformed on the philosophical aspects of Confucianism as well as considering a difficult dimension to measure (Fang, 2003). Other models, such as the GLOBE model, also consider some of Hofstede's dimensions and replace them with others, making a total of nine dimensions (House et al., 1999). The *masculinity—femininity* dimension, for example, is proposed to be instead of two dimensions: *gender egalitarianism* and *assertiveness*. This addition of dimensions avoids assuming that assertiveness is either masculine or feminine, which stems from outdated gender stereotypes. Such gender stereotypes have also been the subject of critique on Hofstede's model(Jeknić, 2014). I agree with these critiques and thus avoid considering such stereotypes in my discussion.

For my purposes of contrasting Western vs. Chinese satisfaction stemming from expectations, these dimensions could explain why Chinese customers are generally satisfied more often with hard factors while Westerners are satisfied or dissatisfied with soft factors.

The backgrounds of collectivism in China and individualism in Western countries have been studied previously (Gao, Zhang, and Huang, 2017; Kim and Lee, 2000). These backgrounds as well as the differences in these cultural dimensions could be the underlying cause for differences in expectations. Regardless of the cause, however, measures in the past have proven that such differences exist (Armstrong et al., 1997).

The cultural background of Chinese tourists emphasizes their surroundings and their place in nature and the environment. Chinese historical backgrounds of Confucianism, Taoism, and Buddhism permeate the thought processes of Chinese populations. However, scholars argue that the changes in generations and their economic and recent history attaches less importance to these concepts in their lives (Gao, Zhang, and Huang, 2017). Nevertheless, one could argue a Chinese cultural attribute emphasizes that the environment and the location affect satisfaction rather than the treatment they receive.

A more anthropocentric and individualistic Western culture could correlate more of their expectations and priorities to the treatment in social circumstances rather than the environment. According to Donthu and Yoo (1998), highly individualistic customers, in contrast to collectivist customers, have a higher expectation of empathy and assurance from the provider, which are aspects of service, a soft attribute of a hotel.

Among other dimensions in both models, I can consider uncertainty avoidance. Customers of high uncertainty avoidance carefully plan their travel and thus have higher expectations towards service. In contrast, customers of lower uncertainty avoidance do not take risks in their decisions and thus face less disappointment with different expectations. However, according to Xiumei and Jinying (2011), the difference between China and the USA in uncertainty avoidance is not clear when measuring with the Hofstede typology and the GLOBE typology. While the USA is not representative of Western society, uncertainty avoidance may not cause the difference in hard-soft attribute satisfaction between Chinese and Western cultures. Differences in another factor, power distance, were also noted when using Hofstede's method compared to the GLOBAL method; therefore, power distance was not considered for comparison.

4.6.5 Implications for hotel managers

My study reached two important conclusions: one about hospitality and cultural differences and another about managerial decisions towards two different populations. Overall, Chinese tourists did not attach much importance to hospitality and service factors. Instead, they focused on the hard attributes of a hotel. In particular, they were not satisfied with hospitality as much as Western tourists were; otherwise, they felt that basic language and communication needs were not met; thereby, they were not much satisfied. Western tourists were highly satisfied with Japanese hospitality and preferred soft attributes to hard ones.

The other conclusion is that managerial decisions could mostly benefit Western tourists, except for language improvements and breakfast inclusion could satisfy both groups. As mentioned earlier in this study, Westerners are "long-haul" customers, spending more of their budget on lodging than Asian tourists (Choi and Chu, 2000). With bigger returns on managerial improvements, I recommend investing in improving attributes that dissatisfy Western customers, such as cleanliness and removing tobacco smell. In addition, breaking the language barrier is one of the few strategies to satisfy both groups. Recently, Japan has been facing an increase in Chinese students as well as students of Western universities. Hiring students as part-time workers could increase the language services of a hotel.

To satisfy both customer types, hotel managers need to invest in cleanliness, deodorizing, and making hotel rooms tobacco-free. It could also be recommended to invest in breakfast inclusion and multilingual services and staff preparedness to deal with Chinese and English speakers. Western tourists were also observed to have high comfort standards, which could be managerially improved for better reviews. Perhaps it could be suggested to perform surveys of the bedding that is most comfortable for Western tourists. However, not all hotels can invest in all of these factors simultaneously. My results suggest that satisfying cleanliness needs could satisfy both customer types. I suggest investing in making the facilities tobacco-free. My results are also divided by price ranges; thereby, a hotel manager could consider which analysis suits their hotel the most. Hard attributes are difficult to change; however, improvements in service can be made to accompany these attributes. For example, transportation guides for foreigners that might not know the area could increase satisfaction.

The managers must consider their business model for implementing the next strategy. One option could be attracting more Chinese customers with their observed low budgeting. Another could be attracting more big-budget Western customers. For example, investing more in cleanliness could improve Western customers looking for high-quality lodging satisfaction, even for an increased price per night. On the other hand, hotels might be deemed costly by Chinese customers wherever such an investment is made.

4.7 Limitations

In this study, I analyzed keywords based on whether they appeared on satisfied reviews or dissatisfied ones. Following that, I attempted to understand these words' context by using a dependency parser and observing the related nouns. However, a limitation is that it analyzed solely the words directly related to each keyword and did not search for further connections. This means that if the words were used in combination with other keywords, I did not trace the effects of multiple contradicting statements. For example, in the sentence "The room is good, but the food is lacking," I extracted "good room" and "lacking food" but did not consider the fact that both occurred in the same sentence.

This study analyzed the differences in customers' expectations at different levels of hospitality and service factors by dividing my data into price ranges. However, in the same price range, for example, the highest one, I can find both a Western-style five-star resort and a high-end Japanese style *ryokan*. Services offered in these hotels are of high quality, albeit very different. Nevertheless, most of my database was focused on the middle range priced hotels, the services of which are comparably less varied.

An essential aspect of this study is that I focused on the satisfaction and dissatisfaction towards the expectations of individual aspects of the hotels. This gave us insight into the factors that hotel managers can consider. However, each customer's overall satisfaction was not measured since it would require methods that are out of the scope of this study. Another limitation is that further typology analysis could not be made because of the nature of the data collected (for example, Chinese men and women of different ages or their Westerner counterparts).

4.8 Conclusion and future work

In this study, I analyzed the differences in satisfaction and dissatisfaction between Chinese and English-speaking customers of Japanese hotels, particularly in the context of Japanese hospitality, *omotenashi*. I extracted keywords from their online reviews on *Ctrip* and *TripAdvisor* using Shannon's entropy calculations. I used these keywords for sentiment classification via an SVC. I then used dependency parsing and part of speech tagging to extract common pairs of adjectives and nouns as well as single words. I divided these data by sentiment and hotel price range (most expensive room/night).

I found that Western tourists were most satisfied with staff behavior, cleanliness, and other soft attributes. However, Chinese customers had other concerns for their satisfaction; they were more inclined to praise the room, location, and hotel's convenience. I found that the two cultures had different reactions to the hospitality environment and the prices. Thus, I discussed two possible theories on why Chinese tourists responded differently from Westerners in the environment of *omotenashi*. One theory is that, although they were treated well, their experience was deteriorated by language or culture barriers. The second possible theory is that they reacted to hospitality differently since they did not have the same expectations. I theorized that a lack of expectations could result in lessened satisfaction than that to the same service if expected. On the other hand, even when they held high expectations in a high-priced hotel, Japanese hospitality exceeded Western tourists' expectations, judging by their vocabulary for expressing their satisfaction. I considered that Western tourists were more reactive to hospitality and service factors Chinese tourists.

Lastly, I measured the satisfaction and dissatisfaction factors, that is, a hotel's hard and soft attributes. Hard attributes are physical and environmental elements, and as such, are impractical elements to change. In contrast, soft attributes can be changed via management and staff by an improvement in services or amenities. I found that, for satisfaction, Western tourists favored soft attributes in contrast to Chinese tourists, who were more interested in the hard attributes of hotels across all the price ranges consistently. For dissatisfaction, Western tourists were also highly inclined to criticize soft attributes, such as cleanliness or cigarette smell in rooms. In contrast, Chinese tourists' dissatisfaction derived from both hard and soft attributes evenly.

One approach for hotel managers is to work to satisfy Chinese tourists more, who dedicate a lower percentage of their budget to hotels but are more numerous. They are less satisfied with soft attributes but have an identifiable method for improving satisfaction by lessening language barriers and providing a satisfactory breakfast. Another approach was focused on the cleanliness, comfort, and tobacco-free space expected by Western tourists. “Long-haul” Western tourists, who spend almost half of their budget on hotels with this strategy, were favored. Although Westerners are less in number than Chinese tourists, it could be proven that they have more substantial returns. This is because Chinese customers also favor cleanliness as a satisfaction factor, and both populations could be pleased.

In future work, I plan to investigate these topics further. I plan to extend my data to research different trends and regions of Japan, different kinds of hotels, and customers traveling alone or in groups, whether for fun or for work. Another point of interest in this study’s future work is to use word clusters with similar meanings instead of single words.

Chapter 5

General Discussion

5.1 Conclusion

At the beginning of this thesis, I stated the objective of exploring the influences of services and business decisions on consumer behavior across the phases immediately prior and posterior to the experience of a service. To accomplish this, I developed appropriate methodologies and evaluated the influence of services on customer behavior using large databases of customer interactions in the service industry of Japan either provided by a research institute or text-mined using the availability of user-provided data in Web 2.0. This effectively provides numerical evaluation indexes of the performance of these services. In Chapter 2 I analyzed the effectivity, or rather lack thereof, of television adverts in influencing purchase intention and behavior in Japanese television. Then in Chapter 3, I considered whether numerical indexes already available as review scores for Japanese hotels online were appropriate for analyzing customer satisfaction, and the results showed that it is necessary to use the review text in order to fully understand the satisfaction of a customer leaving a review. In Chapter 4 I followed through that necessity to analyze the review texts and evaluated the satisfaction factors of Chinese and Western tourists in Japan and their difference, discussing possible cultural influences on that difference. In this thesis, I have succeeded in presenting a case for evaluating the performance of services and their interaction with consumer behavior to discover possible strategies to improve sales or performance of businesses in Japan.

5.2 Limitations

The studies in this thesis were limited each and detailed within each chapter, but to summarize, there were technical limitations that prevented from performing a deeper study correlating all the stages of consumption under one specific service. There were also physical, social, and time limitations, for example, calculation times for more complex machine learning algorithms, language barriers during the international analyses, as well as time limitations within the publishing of my work.

5.3 Future Work

This thesis focused on the steps in the consumption process immediately prior and posterior to the consumption and experience of a service, using different data sources for each step because of the limitations in data retrieval of this magnitude. However, for future work, it would be appropriate to develop a methodology that allows for the exploration of the whole consumption process for a single service or product, starting from the perception of a service, brand recognition, purchase intention, purchase decision, consumption, satisfaction and dissatisfaction, and recommendation or repetition.

Aside from this, each chapter provides details on future work that could be done to delve deeper into the topics seen in my studies. However, an aspect that has always escaped my research is the differences in consumer behavior in different regions of Japan. In the future, it would be ideal to use methodologies such as spatial autocorrelation to perform geographic-aware analyses of these large sets of data and develop an automatic methodology for understanding more complex relations between customers and companies and their services.

5.4 In closing

The research within this thesis answers the research question presented in Chapter 1 (Research Question 1.1) by providing several methodologies and collecting literature on consumer behavior that can be used to have a deeper understanding of how to connect the information engineering field and behavioral studies.

Bibliography

Abernethy, A. (1991). "Television Exposure: Programs vs. Advertising". In: *Current Issues And Research In Advertising* 13.1–2, pp. 61–77. DOI: 10.1080/01633392.1991.10504959.

Aishima, Toshimi, Yoshinobu Sato, et al. (2015). "The origin of Japanese omotenashi in Man-yo-shu". In: *Business & accounting review* 16, pp. 103–122.

Al-alsheikh, Abdulelah and Yoshinobu Sato (2015). "Characteristics of the Hospitality, Omotenashi in the Traditional Japanese Inn: A Case Study of Kagaya". In: *Business & accounting review* 16, pp. 123–142.

Alalwan, A. (2018). "Investigating the impact of social media advertising features on customer purchase intention". In: *International Journal of Information Management* 42, pp. 65–77. DOI: 10.1016/j.ijinfomgt.2018.06.001.

Albanese, Davide et al. (Dec. 2012). "minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers". In: *Bioinformatics* 29.3, pp. 407–408. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bts707. eprint: https://academic.oup.com/bioinformatics/article-pdf/29/3/407/17105876/bts707.pdf. URL: https://doi.org/10.1093/bioinformatics/bts707.

Alemán Carreón, Elisa Claire et al. (Feb. 4, 2018). "Emotional Contribution Analysis of Online Reviews". In: *Proceedings of the 2018 International Conference on Artificial Life and Robotics (ICAROB2018)*. Vol. 23. Beppu, Japan, pp. 359–362. DOI: 10.5954/ICAROB.2018.OS5-3.

Amplayo, Reinald Kim and Min Song (2017). "An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews". In: *Data & Knowledge Engineering* 110, pp. 54–67. ISSN: 0169-023X. DOI: 10.1016/j.datak.2017.03.009. URL: https://www.sciencedirect.com/science/article/pii/S0169023X16301525.

Armstrong, J., V. Morwitz, and V. Kumar (2000). "Sales forecasts for existing consumer products and services: Do purchase intentions contribute to accuracy?" In: *International Journal of Forecasting* 16.3, pp. 383–397. DOI: 10.1016/s0169-2070(00)00058-3.

Armstrong, Robert W et al. (1997). "The importance of cross-cultural expectations in the measurement of service quality perceptions in the hotel industry". In: *International Journal of Hospitality Management* 16.2, pp. 181–190.

Basuroy, S., S. Chatterjee, and S. Ravid (2003). "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets". In: *Journal Of Marketing* 67.4, pp. 103–117. DOI: 10.1509/jmkg.67.4.103.18692.

Bauer, Thomas, Leo Jago, and Brian Wise (1993). "The changing demand for hotel facilities in the Asia Pacific region". In: *International Journal of Hospitality Management* 12.4, pp. 313–322. DOI: 10.1016/0278-4319(93)90048-E.

Boait, P., R. Rylatt, and A. Wright (2007). "Exergy-based control of electricity demand and microgeneration". In: *Applied Energy* 84.3, pp. 239–253. DOI: 10.1016/j.apenergy.2006.09.001.

Boyland, E. and J. Halford (2013). “Television advertising and branding: Effects on eating behaviour and food preferences in children”. In: *Appetite* 62, pp. 236–241. DOI: 10.1016/j.appet.2012.01.032.

Browning, Victoria, Kevin Kam Fung So, and Beverley Sparks (2013). “The influence of online reviews on consumers’ attributions of service quality and control for service standards in hotels”. In: *Journal of Travel & Tourism Marketing* 30.1-2, pp. 23–40. DOI: 10.1080/10548408.2013.750971.

Bulchand-Gidumal, Jacques, Santiago Melián-González, and Beatriz González Lopez-Valcarcel (2013). “A social media analysis of the contribution of destinations to client satisfaction with hotels”. In: *International Journal of Hospitality Management* 35, pp. 44–47. ISSN: 0278-4319. DOI: 10.1016/j.ijhm.2013.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0278431913000728.

Bunn, D. (1982). “Audience Presence During Breaks in Television Programs”. In: *Journal of Advertising Research* 22.5, pp. 35–39.

Chan, Andrew, Cathy H.C. Hsu, and Tom Baum (2015). “The Impact of Tour Service Performance on Tourist Satisfaction and Behavioral Intentions: A Study of Chinese Tourists in Hong Kong”. In: *Journal of Travel & Tourism Marketing* 32.1-2, pp. 18–33. DOI: 10.1080/10548408.2014.986010.

Chang, P., M. Galley, and C. Manning (2008). “Optimizing Chinese Word Segmentation for Machine Translation Performance”. In: *Proceedings of the Third Workshop On Statistical Machine (Statmt ’08)*. Columbus,Ohio,USA, pp. 224–232. URL: http://nlp.stanford.edu/pubs/acl-wmt08-cws.pdf.

Chang, R., J. Kivela, and A. Mak (2010). “Food preferences of Chinese tourists”. In: *Annals Of Tourism Research* 37.4, pp. 989–1011. DOI: 10.1016/j.annals.2010.03.007.

Chen, Danqi and Christopher D. Manning (2014). “A fast and accurate dependency parser using neural networks”. In: *Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 740–750.

Chen, T. and C. Guestrin (2016). “XGBoost: A scalable tree boosting system”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’16)*. San Francisco, California, USA, pp. 785–794. DOI: 10.1145/2939672.2939785.

Choi, I. et al. (2016). “Collaborative filtering with facial expressions for online video recommendation”. In: *International Journal of Information Management* 36.3, pp. 397–402. DOI: 10.1016/j.ijinfomgt.2016.01.005.

Choi, Tat Y. and Raymond Chu (2000). “Levels of satisfaction among Asian and Western travellers”. In: *International Journal of Quality & Reliability Management* 17.2, pp. 116–132. DOI: 10.1108/02656710010304537.

— (2001). “Determinants of hotel guests’ satisfaction and repeat patronage in the Hong Kong hotel industry”. In: *International Journal of Hospitality Management* 20.3, pp. 277–297. ISSN: 0278-4319. DOI: 10.1016/S0278-4319(01)00006-8. URL: http://www.sciencedirect.com/science/article/pii/S0278431901000068.

Cortes, C. and V. Vapnik (1995). “Support-vector networks”. In: *Machine Learning* 20.3, pp. 273–297. DOI: 10.1007/bf00994018.

Coulmas, Florian and Makoto Watanabe (2002). “Japan’s nascent multilingualism”. In: *Contributions To The Sociology Of Language* 87, pp. 249–274.

Dechêne, A. et al. (2009). “Mix me a list: Context moderates the truth effect and the mere-exposure effect”. In: *Journal of Experimental Social Psychology* 45.5, pp. 1117–1122. DOI: 10.1016/j.jesp.2009.06.019.

Dongyang, Zhang, Takeshi Mori, Kenta Hayashi, et al. (2015). "A Study on Preferences and Behavioral Patterns of Chinese Tourists in Kansai Region, Japan". In: *Konan Economic Papers* 55.1-2, pp. 31–46. DOI: 10.14990/00001507.

Donthu, Naveen and Boonghee Yoo (1998). "Cultural influences on service quality expectations". In: *Journal of service research* 1.2, pp. 178–186.

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

Fan, Zhi-Ping, Yu-Jie Che, and Zhen-Yu Chen (2017). "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis". In: *Journal of Business Research* 74, pp. 90–100. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2017.01.010. URL: https://www.sciencedirect.com/science/article/pii/S0148296317300231.

Fang, Tony (2003). "A critique of Hofstede's fifth national culture dimension". In: *International journal of cross cultural management* 3.3, pp. 347–368.

Fang, X., S. Singh, and R. Ahluwalia (2007). "An Examination of Different Explanations for the Mere Exposure Effect". In: *Journal of Consumer Research* 34.1, pp. 97–103. DOI: 10.1086/513050.

Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). "From data mining to knowledge discovery in databases". In: *AI magazine* 17.3, pp. 37–37. DOI: 10.1609/aimag.v17i3.1230. URL: https://www.aaai.org/ojs/index.php/aimagazine/article/download/1230/1131.

Francesco, Galati and Galati Roberta (2019). "Cross-country analysis of perception and emphasis of hotel attributes". In: *Tourism Management* 74, pp. 24 –42. ISSN: 0261-5177. DOI: 10.1016/j.tourman.2019.02.011. URL: http://www.sciencedirect.com/science/article/pii/S0261517719300408.

Gao, Jun, Chaozhi Zhang, and Zhuowei (Joy) Huang (2017). "Chinese tourists' views of nature and natural landscape interpretation: a generational perspective". In: *Journal of Sustainable Tourism* 26.4, pp. 668–684. DOI: 10.1080/09669582.2017.1377722.

Gmuer, A., M. Siegrist, and S. Dohle (2015). "Does wine label processing fluency influence wine hedonics?" In: *Food Quality and Preference* 44, pp. 12–16. DOI: 10.1016/j.foodqual.2015.03.007.

Gonzalez Camacho, L. and S. Alves-Souza (2018). "Social network data to alleviate cold-start in recommender system: A systematic review". In: *Information Processing & Management* 54.4, pp. 529–544. DOI: 10.1016/j.ipm.2018.03.004.

Goodfellow, Ian et al. (2014). "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Gunaratne, Shelton A. (Mar. 2009). "Globalization: A Non-Western Perspective: The Bias of Social Science/Communication Oligopoly". In: *Communication, Culture and Critique* 2.1, pp. 60–82. ISSN: 1753-9129. DOI: 10.1111/j.1753-9137.2008.01029.x. eprint: 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". In: *American Journal Of Marketing Research* 1.4, pp. 208–214.

Heinrich, Patrick (2012). *The making of monolingual Japan: Language ideology and Japanese modernity*. Vol. 146. Multilingual matters.

Hekkert, P., C. Thurgood, and T. Whitfield (2013). "The mere exposure effect for consumer products as a consequence of existing familiarity and controlled exposure". In: *Acta Psychologica* 144.2, pp. 411–417. DOI: 10.1016/j.actpsy.2013.07.015.

Hofstede, Geert (1984). *Culture's consequences: International differences in work-related values*. Vol. 5. London: Sage.

Hogan, Robert T and Nicholas P Emler (1978). "The biases in contemporary social psychology". In: *Social Research*, pp. 478–534.

House, Robert J et al. (1999). "Cultural influences on leadership and organizations: Project GLOBE". In: *Advances in global leadership* 1.2, pp. 171–233.

Hu, Ya-Han, Yen-Liang Chen, and Hui-Ling Chou (2017). "Opinion mining from online hotel reviews – A text summarization approach". In: *Information Processing & Management* 53.2, pp. 436–449. DOI: 10.1016/j.ipm.2016.12.002.

Huang, Jue (2017). "The dining experience of Beijing Roast Duck: A comparative study of the Chinese and English online consumer reviews". In: *International Journal of Hospitality Management* 66, pp. 117 –129. ISSN: 0278-4319. DOI: 10.1016/j.ijhm.2017.07.003. URL: http://www.sciencedirect.com/science/article/pii/S0278431917301251.

Huang, Y. and P. Hsieh (2013). "The mere exposure effect is modulated by selective attention but not visual awareness". In: *Vision Research* 91, pp. 56–61. DOI: 10.1016/j.visres.2013.07.017.

Hunt, John D (1975). "Image as a factor in tourism development". In: *Journal of travel research* 13.3, pp. 1–7.

Ikeda, Nobuhiro (2013). "Omotenashi: Japanese Hospitality as the Global Standard". In: *Management Of Service Businesses In Japan*. World Scientific. Chap. 9, pp. 145–154.

Jacoby, L. and M. Dallas (1981). "On the Relationship Between Autobiographical Memory and Perceptual Learning". In: *Journal of Experimental Psychology: General* 110, pp. 306–304. DOI: 10.1037/0096-3445.110.3.306.

Janiszewski, C. (1993). "Preattentive Mere Exposure Effects". In: *Journal of Consumer Research* 20.3, pp. 376–392. URL: http://www.jstor.org/stable/2489354.

Jansen, B., K. Moore, and S. Carman (2013). "Evaluating the performance of demographic targeting using gender in sponsored search". In: *Information Processing & Management* 49.1, pp. 286–302. DOI: 10.1016/j.ipm.2012.06.001.

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

Japan Statistics Bureau, Ministry of Affairs and Communications (2019). *STATISTICAL HANDBOOK OF JAPAN*. Table 3.2. URL: https://www.stat.go.jp/english/data/handbook/pdf/2019all.pdf.

Japan Tourism Agency (2014). *Consumption trend survey for foreigners visiting Japan*.

Jeknić, Ranka (2014). "Gender equality, young women and culture in the context of the "masculinity/femininity" as a dimension in Geert Hofstede's model of "national culture"". In: *Zbornik radova Pravnog fakulteta u Splitu* 51.3, pp. 681–696.

Jia, Susan (Sixue) (2020). "Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews". In: *Tourism Management* 78, p. 104071. ISSN: 0261-5177. DOI: 10.1016/j.tourman.2019.104071. URL: http://www.sciencedirect.com/science/article/pii/S0261517719302687.

Jones, Dan (2010). "A WEIRD View of Human Nature Skews Psychologists' Studies". In: *Science* 328.5986, pp. 1627–1627. ISSN: 0036-8075. DOI: 10.1126/science.

328.5986.1627. eprint: https://science.sciencemag.org/content/328/5986/1627.full.pdf. URL: https://science.sciencemag.org/content/328/5986/1627.

Jones, Thomas et al. (2009). "Prefectural branding in Japan – Tourism, National Parks and the Shinshu brand". In: *Place Branding and Public Diplomacy* 5.3, pp. 192–201. DOI: 10.1057/pb.2009.13.

Kannan, R., G. Ghinea, and S. Swaminathan (2015). "What do you wish to see? A summarization system for movies based on user preferences". In: *Information Processing & Management* 51.3, pp. 286–305. DOI: 10.1016/j.ipm.2014.12.001.

Khuong, M. and T. Nguyen (2015). "The Effects of Television Commercials on Customers Purchase Intention – A Study of Milk Industry in Ho Chi Minh City, Vietnam". In: *Journal of Economics, Business and Management* 3.9, pp. 851–857.

Kim, Chulwon and Seokho Lee (2000). "Understanding the Cultural Differences in Tourist Motivation Between Anglo-American and Japanese Tourists". In: *Journal of Travel & Tourism Marketing* 9.1-2, pp. 153–170. DOI: 10.1300/J073v09n01_09.

Kozak, Metin (2002). "Measuring Tourist Satisfaction With Multiple Destination Attributes". In: *Tourism Analysis* 7.3-4, 229–240(12). DOI: 10.3727/108354203108750076.

Kuboyama, Tetsuo (2020). ""Omotenashi" Must Comprise Hospitality and Service". In: *Serviceology for Services*. Springer Singapore, pp. 34–53. DOI: 10.1007/978-981-15-3118-7_3. URL: https://doi.org/10.1007/978-981-15-3118-7_3.

Lee, A. and A. Labroo (2004). "The Effect of Conceptual and Perceptual Fluency on Brand Evaluation". In: *Journal of Marketing Research* 41.2, pp. 151–165. DOI: 10.1509/jmkr.41.2.151.28665.

Lee, J. and I. Hong (2016). "Predicting positive user responses to social media advertising: The roles of emotional appeal, informativeness, and creativity". In: *International Journal of Information Management* 36.3, pp. 360–373. DOI: 10.1016/j.ijinfomgt.2016.01.001.

Levy, Stuart E. (2010). "The hospitality of the host: A cross-cultural examination of managerially facilitated consumer-to-consumer interactions". In: *International Journal of Hospitality Management* 29.2. Special issue on Event Studies, pp. 319–327. ISSN: 0278-4319. DOI: 10.1016/j.ijhm.2009.10.003. URL: https://www.sciencedirect.com/science/article/pii/S027843190900111X.

Likert, Rensis (1932). "A technique for the measurement of attitudes". In: *Archives of psychology* 22, p. 55.

Liu, Ben Shaw-Ching, Olivier Furrer, and Devanathan Sudharshan (2001). "The relationships between culture and behavioral intentions toward services". In: *Journal of service research* 4.2, pp. 118–129.

Liu, Yang, Jian-Wu Bi, and Zhi-Ping Fan (2017). "Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory". In: *Information Fusion* 36, pp. 149–161. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2016.11.012. URL: https://www.sciencedirect.com/science/article/pii/S1566253516301580.

Liu, Yi et al. (2019). "Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations". In: *Tourism Management* 71, pp. 337–347. ISSN: 0261-5177. DOI: 10.1016/j.tourman.2018.10.004. URL: http://www.sciencedirect.com/science/article/pii/S0261517718302395.

Loh, Stanley et al. (2003). "A Tourism Recommender System Based On Collaboration And Text Analysis". In: *Information Technology & Tourism* 6.3, pp. 157–165. DOI: 10.3727/1098305031436980.

Manning, Christopher D. et al. (2014). "The Stanford CoreNLP Natural Language Processing Toolkit". In: *Proceedings of the Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60. URL: http://www.aclweb.org/anthology/P/P14/P14-5010.

Marneffe, Marie-Catherine de and Christopher D. Manning (2008). *Stanford typed dependencies manual*. Tech. rep. Stanford University. URL: https://nlp.stanford.edu/software/dependencies_manual.pdf.

Mattila, Anna S (1999). "The role of culture and purchase motivation in service encounter evaluations". In: *Journal of Services Marketing* 13, pp. 376–389.

Mattila, Anna S. (2019). "A commentary on cross-cultural research in hospitality & tourism inquiry (invited paper for 'luminaries' special issue of International Journal of Hospitality Management)". In: *International Journal of Hospitality Management* 76. HOSPITALITY LUMINARIES ISSUE, pp. 10–12. ISSN: 0278-4319. DOI: 10.1016/j.ijhm.2018.06.007. URL: https://www.sciencedirect.com/science/article/pii/S027843191830505X.

Morwitz, V., J. Steckel, and A. Gupta (2007). "When do purchase intentions predict sales?" In: *International Journal of Forecasting* 23.3, pp. 347–364. DOI: 10.1016/j.ijforecast.2007.05.015.

Newberry, C., B. Klemz, and C. Boshoff (2003). "Managerial implications of predicting purchase behavior from purchase intentions: a retail patronage case study". In: *Journal of Services Marketing* 17.6, pp. 609–620. DOI: 10.1108/08876040310495636.

Nielsen, Mark et al. (2017). "The persistent sampling bias in developmental psychology: A call to action". In: *Journal of Experimental Child Psychology* 162, pp. 31 – 38. ISSN: 0022-0965. DOI: 10.1016/j.jecp.2017.04.017. URL: http://www.sciencedirect.com/science/article/pii/S0022096517300346.

Oh, Haemoon and Sara C Parks (1996). "Customer satisfaction and service quality: a critical review of the literature and research implications for the hospitality industry". In: *Hospitality Research Journal* 20.3, pp. 35–64.

Oliver, Richard L (1981). "Measurement and evaluation of satisfaction processes in retail settings". In: *Journal of retailing* 57.3, pp. 25–48.

Perlich, C. et al. (2013). "Machine learning for targeted display advertising: transfer learning in action". In: *Machine Learning* 95.1, pp. 103–127. DOI: 10.1007/s10994-013-5375-2.

Powers, D. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". In: *Journal Of Machine Learning Technologies* 2.1, pp. 37–63. URL: http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf.

Ramaboa, K. and P. Fish (2018). "Keyword length and matching options as indicators of search intent in sponsored search". In: *Information Processing & Management* 54.2, pp. 175–183. DOI: 10.1016/j.ipm.2017.11.003.

Reber, R., P. Winkielman, and N. Schwarz (1998). "Effects of Perceptual Fluency on Affective Judgments". In: *Psychological Science* 9.1, pp. 45–48. DOI: 10.1111/1467-9280.00008.

Ren, Gang and Taeho Hong (2019). "Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews". In: *Information Processing & Management* 56.4, pp. 1425–1438. DOI: 10.1016/j.ipm.2018.04.003.

Reshef, David N. et al. (2011). "Detecting Novel Associations in Large Data Sets". In: *Science* 334.6062, pp. 1518–1524. DOI: 10.1126/science.1205438. eprint: https:

//www.science.org/doi/pdf/10.1126/science.1205438. URL: https://www.science.org/doi/abs/10.1126/science.1205438.

Romão, João et al. (2014). "Determinants of trip choice, satisfaction and loyalty in an eco-tourism destination: a modelling study on the Shiretoko Peninsula, Japan". In: *Ecological Economics* 107, pp. 195–205. DOI: 10.1016/j.ecolecon.2014.07.019.

Rossiter, J. et al. (2001). "Brain-Imaging Detection of Visual Scene Encoding in Long-term Memory for TV Commercials". In: *Journal of Advertising Research* 41.2, pp. 13–21. DOI: 10.2501/jar-41-2-13-21.

Ryan, C. and X. Mo (2001). "Chinese visitors to New Zealand - Demographics and perceptions". In: *Journal Of Vacation Marketing* 8.1, pp. 13–27. DOI: 10.1177/135676670200800103.

Schwartz, E., E. Bradlow, and P. Fader (2017). "Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments". In: *Marketing Science* 36.4, pp. 500–522. DOI: 10.1287/mksc.2016.1023.

Schwarz, N. (2004). "Metacognitive Experiences in Consumer Judgment and Decision Making". In: *Journal of Consumer Psychology* 14.4, pp. 332–348. DOI: 10.1207/s15327663jcp1404_2.

Seamon, J. et al. (1995). "The Mere Exposure Effect Is Based On Implicit Memory: Effects of Stimulus Type, Encoding Conditions, and Number of Exposures on Recognition and Affect Judgements". In: *Journal of Experimental Psychology: Learning, Memory And Cognition* 21.3, pp. 711–721.

Shanka, Tekle and Ruth Taylor (2004). "An investigation into the perceived importance of service and facility attributes to hotel satisfaction". In: *Journal of Quality Assurance in Hospitality & Tourism* 4.3-4, pp. 119–134. DOI: 10.1300/J162v04n03_08.

Shannon, C. (1948). "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27.3, pp. 279–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shareef, M. et al. (2017a). "Content design of advertisement for consumer exposure: Mobile marketing through short messaging service". In: *International Journal of Information Management* 37.4, pp. 257–268. DOI: 10.1016/j.ijinfomgt.2017.02.003.

Shareef, M. et al. (2017b). "Social media marketing: Comparative effect of advertisement sources". In: *Journal of Retailing And Consumer Services*. DOI: 10.1016/j.jretconser.2017.11.001.

Shareef, M. et al. (2018). "Advertisements on Facebook: Identifying the persuasive elements in the development of positive attitudes in consumers". In: *Journal of Retailing And Consumer Services* 43, pp. 258–268. DOI: 10.1016/j.jretconser.2018.04.006.

Silva, R., T. Garcia-Marques, and R. Reber (2017). "The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth". In: *Consciousness and Cognition* 51, pp. 53–67. DOI: 10.1016/j.concog.2017.02.016.

Sparks, Beverley A. and Victoria Browning (2011). "The impact of online reviews on hotel booking intentions and perception of trust". In: *Tourism Management* 32.6, pp. 1310–1323. ISSN: 0261-5177. DOI: 10.1016/j.tourman.2010.12.011. URL: https://www.sciencedirect.com/science/article/pii/S0261517711000033.

Stitelman, O. et al. (2011). "Estimating the effect of online display advertising on browser conversion". In: *Proceedings of the Fifth International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD 2011)*. San Diego, California, USA, pp. 8–16.

Sun, B. and V. Morwitz (2010). “Stated intentions and purchase behavior: A unified model”. In: *International Journal of Research In Marketing* 27.4, pp. 356–366. DOI: 10.1016/j.ijresmar.2010.06.001.

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

Toutanova, Kristina and Christopher Manning (2000). “Enriching the knowledge sources used in a maximum entropy part-of-speech tagger”. In: *Proceedings of the 2000 Joint SIGDAT Conference EMNLP/VLC, 63-71, 2000*.

Toutanova, Kristina et al. (2003). “Feature-rich part-of-speech tagging with a cyclic dependency network”. In: *Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology-volume 1*. Association for Computational Linguistics, pp. 173–180.

Truong, T. and B. King (2009). “An evaluation of satisfaction levels among Chinese tourists in Vietnam”. In: *International Journal Of Tourism Research* 11.6, pp. 521–535. DOI: 10.1002/jtr.726. URL: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jtr.726.

Tsujimoto, Noriko (2017). “The purchasing behavior of Chinese tourists at popular visiting areas in Japan”. In: *Journal of Global Tourism Research* 2.2, pp. 99–104.

United Nations Conference on Trade and Development (2019). *Gross domestic product: GDP by type of expenditure, VA by kind of economic activity, total and shares, annual*. URL: http://unctadstat.unctad.org/wds/TableViewer/tableView.aspx?ReportId=95.

Uzama, Austin (2012). “Yokoso! Japan: Classifying Foreign Tourists to Japan for Market Segmentation”. In: *Journal of Hospitality Marketing & Management* 21.2, pp. 132–154. DOI: 10.1080/19368623.2011.615016.

Vermeulen, Ivar E. and Daphne Seegers (2009). “Tried and tested: The impact of online hotel reviews on consumer consideration”. In: *Tourism Management* 30.1, pp. 123–127. ISSN: 0261-5177. DOI: 10.1016/j.tourman.2008.04.008. URL: https://www.sciencedirect.com/science/article/pii/S0261517708000824.

Walker, S. and D. Duncan (1967). “Estimation of the Probability of an Event as a Function of Several Independent Variables”. In: *Biometrika* 54.1–2, pp. 167–179. DOI: 10.2307/2333860.

Winsted, Kathryn Frazer (1997). “The service experience in two cultures: A behavioral perspective”. In: *Journal of Retailing* 73.3, pp. 337–360.

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

Wu, C. et al. (2015). “Location-aware service applied to mobile short message advertising: Design, development, and evaluation”. In: *Information Processing & Management* 51.5, pp. 625–642. DOI: 10.1016/j.ipm.2015.06.001.

Wu, Cedric Hsi-Jui and Rong-Da Liang (2009). “Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants”. In: *International Journal of Hospitality Management* 28.4, pp. 586–593. DOI: 10.1016/j.ijhm.2009.03.008.

Xia, Fei (2000). *The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0)*. Tech. rep. IRCS Technical Reports Series. 38. Institute for Research in Cognitive Science, University of Pennsylvania. URL: http://repository.upenn.edu/ircs_reports/38.

Xiang, Zheng and Ulrike Gretzel (2010). "Role of social media in online travel information search". In: *Tourism Management* 31.2, pp. 179–188. DOI: 10.1016/j.tourman.2009.02.016.

Xie, Karen L., Zili Zhang, and Ziqiong Zhang (2014). "The business value of online consumer reviews and management response to hotel performance". In: *International Journal of Hospitality Management* 43, pp. 1–12. ISSN: 0278-4319. DOI: 10.1016/j.ijhm.2014.07.007. URL: https://www.sciencedirect.com/science/article/pii/S027843191400125X.

Xiumei, SHI and WANG Jinying (2011). "Cultural distance between China and US across GLOBE model and Hofstede model". In: *International Business and Management* 2.1, pp. 11–17.

Yang, Yang, Bing Pan, and Haiyan Song (2014). "Predicting hotel demand using destination marketing organization's web traffic data". In: *Journal of Travel Research* 53.4, pp. 433–447. DOI: 10.1177/0047287513500391.

Ye, Qiang, Rob Law, and Bin Gu (2009). "The impact of online user reviews on hotel room sales". In: *International Journal of Hospitality Management* 28.1, pp. 180–182. DOI: 10.1016/j.ijhm.2008.06.011.

Zajonc, R. (1968). "Attitudinal effects of mere exposure". In: *Journal of Personality And Social Psychology* 9.2 Pt.2, pp. 1–27. DOI: 10.1037/h0025848.

Zeman, Daniel et al. (2018). "CoNLL 2018 shared task: Multilingual parsing from raw text to universal dependencies". In: *Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to universal dependencies*, pp. 1–21.

Zhang, Haiping et al. (2011). "Feature-level sentiment analysis for Chinese product reviews". In: *2011 3rd International Conference on Computer Research and Development*. Vol. 2. IEEE, pp. 135–140. DOI: 10.1109/ICCRD.2011.5764099.

Zhang, M et al. (2019). "Prevalence of smoking and knowledge about the smoking hazards among 170,000 Chinese adults: a nationally representative survey in 2013-2014." In: *Nicotine & tobacco research: official journal of the Society for Research on Nicotine and Tobacco* ntz020. DOI: 10.1093/ntr/ntz020.

Zhang, Y. et al. (2014). "Bid keyword suggestion in sponsored search based on competitiveness and relevance". In: *Information Processing & Management* 50.4, pp. 508–523. DOI: 10.1016/j.ipm.2014.02.004.

Zhou, Lingqiang et al. (2014). "Refreshing hotel satisfaction studies by reconfiguring customer review data". In: *International Journal of Hospitality Management* 38, pp. 1–10. ISSN: 0278-4319. DOI: 10.1016/j.ijhm.2013.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0278431913001801.