World Wide Web manuscript No.
(will be inserted by the editor)

Cultural Difference and Visual Information on Hotel Rating Prediction

Wei-Ta Chu and Wei-Han Huang

Received: date / Accepted: date

Abstract Due to the emergence of hotel social media platforms, how to discover interesting properties and utilize these discovered characteristics in hotel-related applications become important issues. In this work, we extend a large-scale hotel information collection to include heterogeneous hotel information, in order to facilitate multimodal and cross-culture analysis. With this rich dataset, we analyze various correlations between hotel properties and unveil interesting characteristics that would benefit hotel recommendation. We found that travelers from different cultural areas (countries) have different rating behaviors. In addition, beyond the scope of conventional text-based hotel analysis, we utilize visual analysis techniques to analyze hotel's cover photo, and investigate the relationship between rating behaviors and visual information. We adopt these correlations to predict hotel ratings, and verify that by considering visual information and cultural difference, prediction performance can be improved.

Keywords Hotel rating prediction · Cultural difference · Visual properties · Factorization machine

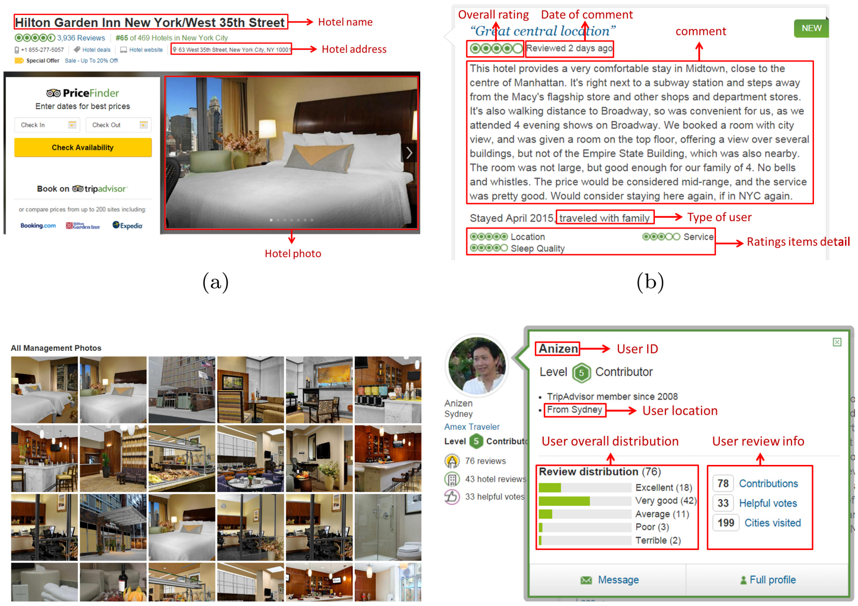
1 Introduction

With rapid development of network technology and wide applications of e-commerce, consumers are increasingly interested in booking hotels online. Web sites like TripAdvisor[1], agoda[2], Hotels.com[3], and Booking.com[4] attract millions of users. Tremendous amounts of travelers search and book hotels on these websites, and give ratings or comments after their stay. (1) The considerable number of users,

Wei-Ta Chu
National Chung Cheng University E-mail: wtchu@ccu.edu.tw

Wei-Han Huang
National Chung Cheng University E-mail: morewater0726@gmail.com

[1] http://www.tripavdisor.com
[2] http://www.agoda.com
[3] http://www.hotels.com
[4] http://www.booking.com



(a) (b) (c) (d)

Fig. 1: Examples of (a) hotel information, (b) rating information, (c) hotel photos, and (d) user information on TripAdvisor.

(2) various information about hotels, and (3) highly interactive hotel ratings make these hotel booking websites one of the richest platforms in the cyberspace.

Taking TripAdvisor as an example, we describe properties corresponding to the three aforementioned issues. First, TripAdvisor's traffic flow continues to show an upward tendency and it becomes one of the top 200 websites in the world, according to Alexa traffic statistics[5]. About the second and the third issues, Fig. 1 shows examples of various hotel information and user ratings. Fig. 1a shows hotel name and address. We can browse photos prepared by the hotel administration (Fig. 1c), and find ratings given by different types of travelers like family travelers or business travelers (Fig. 1b). Detailed user information is also available (Fig. 1d), like which country he/she comes from and how trustworthy he/she is.

Although there have been many works on hotel information analysis and hotel recommendation, many aspects of hotel-related information from rich and heterogeneous sources are still unexplored. We observe that hotels are one of the few places where different countries of people with different traveling purposes and with different cultural backgrounds would visit and generate various digital footprints. However, few works focus on how the cultural factor influences hotel ratings. We also envision the potential of visual information on hotel data analysis. Currently most hotel recommender systems work based on text-based information or metadata. It would be interesting to investigate how photos available on the web impacts hotel recommendation. Therefore, in this work we will collect various

[5] http://www.alexa.com/siteinfo/tripadvisor.com

types of data from TripAdvisor and especially investigate how hotel properties relate to user's ratings, how different users rate hotels, and how these discovered relationships benefit hotel rating prediction from the cultural and visual perspectives.

We will take hotel rating prediction as the main instance to verify effectiveness of the aforementioned factors. Hotel rating prediction could be directly related to personal hotel recommendation. For a user who wants to book a hotel with some constraints, like check-in date and room rate, the proposed rating prediction system predicts how this user would rate a specific hotel. In other word, we estimate how a user would rate a hotel before he/she really goes there. A hotel recommendation system, therefore, can be built by returning a list of hotels in descending order according to the predicted ratings. This scenario has actually been implemented by some hotel booking websites, e.g., the "Just for You" functionality at TripAdvisor, though we don't really know its technical details.

To enable the proposed studies, we extend an existing hotel dataset [1] by including travelers' personal information, especially the country property, and hotel's cover photos given by the hotel administration. This dataset provides the foundation for multimodal study that is beyond most current works. In addition to data collection, our primary contributions are twofold.

- Data analysis: We demonstrate several interesting characteristics based on statistics and multimodal analysis. More particularly, cultural difference is discovered through comparing rating trends of different countries' users. The relationship between visual concepts and hotel ratings is also investigated, as the first attempt among relevant hotel information studies.
- A novel hotel rating prediction system: In contrast to previous works mainly focusing on text-based metadata and prediction algorithms, we adopt a recently used recommendation algorithm (factorization machine [4]) with the consideration of cultural difference and visual information to achieve rating prediction from a different perspective.

The rest of this paper is organized as follows. We survey related works in Section 2. In Section 3, we introduce details of data collection and data analysis, where cultural difference and the relationship between visual concepts and hotel ratings will be discussed. In Section 4, we design a hotel rating prediction system considering visual information and cultural difference, and demonstrate its effectiveness. Finally, conclusion and future works are described in Section 5.

2 Related Works

2.1 Hotel Data Analysis

Many works have been proposed to study hotel information from various perspectives. Zhang et al. [5] showed that hotel room prices are considerably related to hotel's spatial location, and the traffic factor is more important than room prices in selecting tourism spots. Dai and Lin [6] analyzed mid-priced limited service hotels and found that travelers in such kind of hotels pay more attention on surrounding area, room service, and quality of facility. On the contrary, travelers in other types of hotels pay more attention on public area, restaurant, and meeting room facility. Hargreaves [30] studies several interesting rating and review characteristics from rating/review data for five Singapore hotels. She reported statistics of ratings for these five hotels, and grouped rating items based on factor analysis, showing rating variations for different hotels given by travelers with different travel intents. Later in [31], she further studied guest satisfaction ratings and conducted text analysis of customer reviews to better understand positive and negative sentiments on hotels. The most important attributes for guest satisfaction reported in [31] turned out to be rooms, value for money and location. Xiang et al. [43] demonstrated a number of findings obtained by large-scale statistical analysis on customer reviews for hotels in 100 largest U.S. cities. Interesting findings include, for example, the distribution of hotel numbers in different states, the distribution of ratings, primary words in hotel reviews, factor loadings of words, etc.

Recently, Banerjee and Chua [44] conducted an interesting study of two types of hotels, i.e., independent hotels and chain hotels. For each hotel type, hotels are categorized into four groups according to hotel locations, such as America and Europe, and travelers are categorized into five groups according to their travel types, such as business, couple, and family. Therefore, ratings of $5 \times 4 \times 2$ profiles parameterized by hotel type, traveler type, and hotel location were studied. This work is highly related to our work. However, we further consider the influence of visual information on hotel ratings. In contrast to investigate hotels in different continents, we investigate ratings given by users from different countries. It is thus possible to study cultural difference embedded in user ratings. Finally, we more finely study characteristics of different rating items, while their work mainly focused on the overall ratings.

Rahayu et al. [7] proposed a review and rating system for finding correlations between online reviews and ratings. For a hotel, some travelers may just give comments/ratings to facilities, and others may give comments/ratings to restaurants. However, the overall rating of a hotel is usually just average of these ratings, and thus ratings from different types of travelers are mixed. This system gives weights of different types of ratings, and provides more objective overall comments and ratings. Wei et al. [37] realized automatic hotel service quality assessment based on user comments by using fuzzy methods. Promising assessment performance can be obtained by especially importing trustworthy degree of comments.

Wang et al. [1][8] proposed a multiple aspect ranking problem modeling the dependencies among aspects. The proposed system discovered latent aspects of user comments. The strength of latent aspects derived from a number of comments is then aggregated to form the overall ratings. Li et al. [35] modeled textual reviews as a generative process, and proposed the Aspect Identification and Rating model to identify latent aspects and predict ratings. They also considered that in short reviews, the aspects being mentioned may be imbalanced, and another model capturing mutual influence between aspect and rating was proposed. Chen et al. [39] proposed an interactive visualization for presenting summary of hotel reviews, which is discovered by latent Dirichlet allocation that enables topic clustering.

It is interesting to note that, given users' growing demands to use online reviews for travel planning, posting dishonest reviews to manipulate the online reputation of a hotel becomes a practical issue. In [32], they studied characteristics of manipulative and authentic negative reviews. From another perspective, Martin et al. [33] employed text mining to extract emotionality of hotel reviews, and found that influential writers are likely to use more affective words in terms of both emotion variety and intensity. Hwang et al. [41] proposed three types of features, i.e., content features, sentiment features, and writing quality features, associated with the latent Dirichlet allocation techniques to identify noteworthy hotel reviews. Minnich et al. [42] studied how ones can leverage information from multiple review hosting sites. They developed novel features to identify review discrepancies between three travel sites, and unveiled several interesting facts as well as found evidence of review manipulation.

2.2 Hotel Recommender Systems

As one of the most interesting targets in recommender systems, hotel recommendation has been studied for years. Most hotel recommender systems are based on user-item rating matrices, and many collaborative filtering techniques have been proposed [9]. The idea of collaborative filtering builds the foundation of many recommender systems. However, when the data are extremely sparse, recommendation performance may be seriously deteriorated. Saga et al. [13] proposed a system expressing user's behavior as a preference transition network. This network is represented as a directed graph, where each node is a hotel. Through the links between nodes, appropriate hotels are recommended to users. Guo [14] defined a new rating profile by incorporating the ratings of trustworthy neighbors and proposed new similarity measures to improve recommender systems. Xiong et al. [15] constructed a personalized recommendation system by analyzing customer's purchasing behaviors. In [16], a review-based hotel recommender system was developed by using the labeled Latent Dirichlet Allocation (LDA) [17] method to infer trip intents. Poriya et al. [18] built two recommender systems, a non-personalized recommender system and a collaborative recommender system, respectively. The former system is easier to implement, but all users are given the same recommendation without personalization. The latter is computationally effective, but much more user rating data are needed for training, and the cold-start problem may harm the system. Zhang et al. [10] combined collaborative filtering with a content-based method to solve the sparsity issue. They considered travel intents as additional information to overcome the cold-start problem. They utilized diverse techniques (i.e., SVD [11], PMF [12] and LFM [10]) to optimize the recommendation list.

Recently, Zhou et al. [34] proposed to use images as one of a user's profiles, and recommended hotels that are with images similar to the visual profile, as a new way to do hotel recommendation. Lin et al. [38] recorded the gesture information when users browse hotel reviews, and focused on the paragraphs of interest (identified based on gesture) to do text mining. Better recommendation performance was reported if the gesture-based user profiles are used. In [36], multiple preferences of customers are jointly considered to find a list of recommended hotels. This process was formulated as an optimization problem. The returned hotel list simultaneously minimizes user's search cost, maximizes the utility gained from the hotel, and recommends selected hotels in the top positions. Traub et al. [40] demonstrated a portal website that implements various state-of-the-art recommender algorithms. The developed system considers user interaction at large scale and provides hotel recommendation in real-time.

Although hotel data analysis and hotel recommendation have been studied for years, few works focus on how visual information and cultural factor impact ho-

Table 1: Statistics of the UIUC dataset [1] (top half) and our further crawled data (bottom half).

Property	Number
Number of users	1,320,773
Number of user comments	693,437
Number of user comment dates	693,437
Number of hotels	12,773
Number of rating items	9
Number of cover photos	12,773
Number of hotel address	12,773
Number of hotels with price information	11,941
Number of countries where users come from	213

tel information analysis. The work in [34] considered visual information as user's profile to facilitate hotel recommendation, but the proposed system was just a prototype, and the influence of visual factor on hotel recommendation is still unclear. In [5][6][30][31][43], interesting findings were discovered from various perspectives. However, the cultural factor, i.e., how users of different countries rate hotels, has not been clearly studied. We thus propose hotel information analysis from these two viewpoints, and take hotel rating prediction as the instance to verify effectiveness of these two factors.

3 Dataset Construction and Statistical Analysis

We will discuss data properties and the process of data construction in Section 3.1. In Section 3.2, we will show interesting statistics by various charts and tables. The relationships between visual information and hotel ratings are shown in Section 3.3.

3.1 Dataset Construction

We need a large-scale hotel collection associated with heterogeneous metadata to support the proposed studies. In this work we collect data from TripAdvisor, and focus on hotel properties in the following: (1) Hotel information including name, address, and other contact information (Fig. 1a); (2) hotel photos prepared by the hotel administration (Fig. 1c); (3) rating information like rating items, user comments and comment date (Fig. 1b); and (4) user information, especially his/her country (Fig. 1d).

To quickly build a convincing dataset, we crawl hotel-related information based on an existing large-scale hotel collection, i.e., UIUC dataset [1], which has been widely used in related studies [19][20][21]. Table 1 shows statistics of the UIUC dataset (the top half) and our further crawled data (the bottom half). There are totally over 12,000 hotels mostly from the USA. Over 1320K unique users give their comments or ratings, and there are over 693K user comments. There are nine different rating items to show how users rate hotels, including room, service, business service, cleanliness, check in/front desk, overall, value, location, and sleep quality. Table 2 shows detailed meanings of these nine rating items.

Table 2: Detailed meanings of rating items.

Rating item	Meaning
Rooms	Quality of bedding/mattresses
Service	Quality of facilities/amenities
Business service	Quality of internet access, work centers, printers and fax service
Cleanliness	Cleanliness of bathroom
Check in / front desk	Key pickup/access to the property
Overall	Overall condition of hotel
Value	C/P (Capability/Price) values
Location	Surrounding area of the hotel, traffic convenience
Sleep quality	Quality of sleep environment

From the viewpoint of multimedia research, the explosive number of user comments and hotel photos give rise to significant demands as well as research opportunities for data mining and image processing. Based on the hotel URLs available in the UIUC dataset, we develop a web crawler to collect cover photos and various metadata (the bottom half of Table 1). Fig. 2 shows the framework of our web crawler. Fig. 3 is the histogram showing numbers of hotels in different countries, in a descending order. Top sixteen countries containing more than 100 hotels in this database are shown. We clearly see that most hotels are located in the United States or Europe. Based on user IDs, we can get user location (in the representation of the city name). Given a city name, the corresponding country information was obtained by the Google Maps API[6]. Fig. 4 visualizes locations of users in this database by a heat map. Most users come from North America and Europe.

Note that the location information of user obtained from TripAdvisor is where a user resides. It may be possible that a Japanese residing in US, and our crawler automatically categories he/she as an American. Because there is no way for us to exactly know a user's nationality, we assume that in most cases resident region information coincides with nationality information. Based on a large-scale dataset, meaningful and believable trends can still be discovered even if there is sort of noise. In the following context, we equate *resident region*, *country*, and *nationality* for convenience, and thus can enable the proposed cultural factor study.

The advantages of using the UIUC dataset as the basis to collect data are worth mentioning. First, based on this dataset, we are able to quickly build a large-scale dataset with richer metadata. Second, comparing with other hotel booking websites, TripAdvisor has heterogeneous information, including both text-based metadata/comments and hotel cover photos. User's uploaded photos are also available, though we don't utilize them currently. We can thus construct a richer dataset, named CCU (National Chung Cheng University) hotel dataset, which would be useful in many hotel studies.

3.2 Data Analysis

We first show some statistics of the CCU hotel dataset, and investigate relationships between hotel properties. The relationships between ratings and user's cultural background as well as visual information are then discussed.

[6] https://developers.google.com/maps/

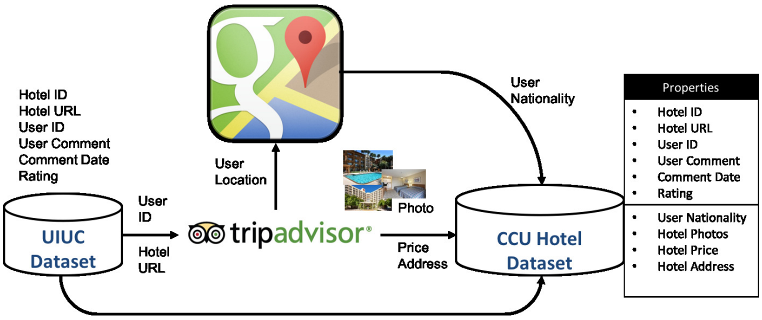


Fig. 2: The framework of our web crawler.

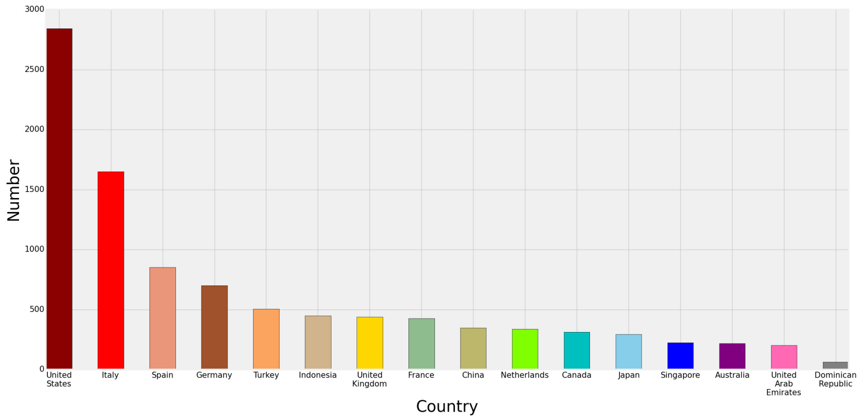


Fig. 3: Numbers of hotels in different countries. Only countries containing more than 100 hotels in the database are shown.

3.2.1 Hotel, User, and Rating Distributions

Do users have different preferences on rating items? Fig. 5 shows that more than 80% of users give ratings on overall, cleanliness, service, and value rating items. Less than 20% of users give ratings on business service, sleep quality, and check in/front desk rating items. The numbers of ratings in different items are imbalanced, which reflects rating behaviors in the real world. Most travelers are not afford or are not willing to spend time to carefully rate hotels in all nine aspects.

Fig. 6 shows the relationship between hotel prices and the overall ratings. Averages and standard deviations of hotel prices are illustrated in red, while original data points are shown in light gray in the background. We can see that averagely most hotel prices are below 200 US dollars. The average prices of score-5 hotels and score-4 hotels are about \$150 and \$120, respectively. Generally, standard deviations of prices on high-score hotels are larger than that of low-score hotels.

The box plot shown in Fig. 7 more clearly realizes the relationship between price and rating scores, by seeing the median prices and the first/third quartiles.

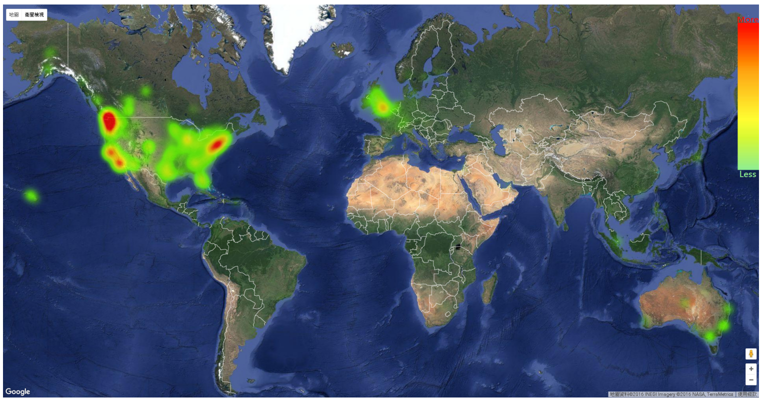


Fig. 4: The heat map showing where users come from.

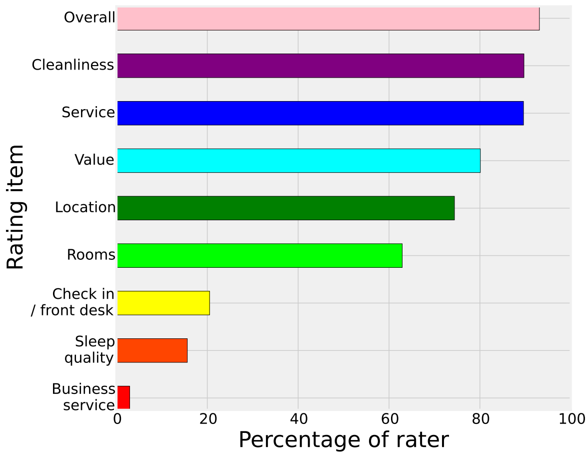


Fig. 5: Percentages of users who give ratings on different rating items.

Out of our expectation, the median prices of high-scored hotels (overall score larger than 4) are just slightly higher than lower-scored hotels. Interestingly, for high-scored hotels, the price differences between the first quartile and the third quartile are larger. This means that rating variations are larger for expensive hotels.

Fig. 8 shows the box plot of overall ratings for five big cities of the United States. We can see hotels in Denver and Houston have slightly lower median overall ratings than that in LA, Chicago and Boston, but generally the difference is not large. Considering the interval between the first quartile and the third quartile, we see that the intervals for Denver and Houston are smaller. Another interesting observation is that hotels in LA and Boston obtain at least 2 scores, showing that the least-quality hotels in these two cities are still given satisfactory overall ratings.

We would like to explore correlation between different rating items. To this end, we utilize GAP (generalized association plots) [22] to visualize magnitude of correlation and cluster similar rating items together. GAP is an exploratory

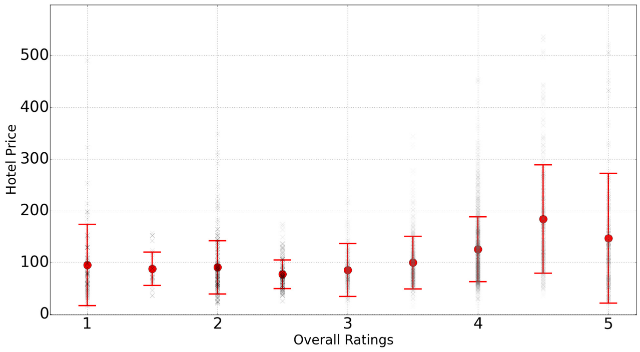


Fig. 6: The relationship between hotel prices and the overall ratings.

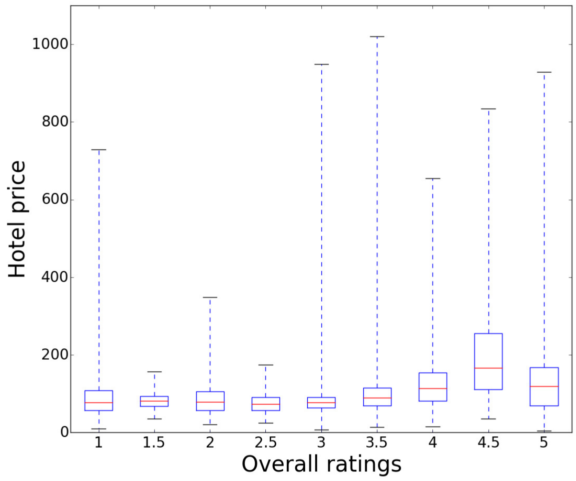


Fig. 7: The box plot showing hotel prices in different ranges of overall ratings.

data analysis tool for matrix visualization and clustering. It was designed to facilitate exploration of high-dimensional data with suitable color projection and clustering/seriation algorithms, without preset dimension reduction methods. Fig. 9 shows the Pearson correlation coefficients between nine rating items plus hotel price and comment dates. From this figure, we see that hotel price merely has correlations with other rating items. Most rating items except for sleep quality are highly positively correlated with each other. This indicates, for example, when a user gives a high score in business service, he/she would also give a high score in cleanliness. The sleep quality item is positively correlated with most other rating items, but is not correlated with business service and check in/front desk. This result is not beyond our expectation. User's comment dates are moderately negatively correlated with most rating items. Users usually first survey hotel ratings before booking, and tend to select hotels of high ratings. They may have high ex-

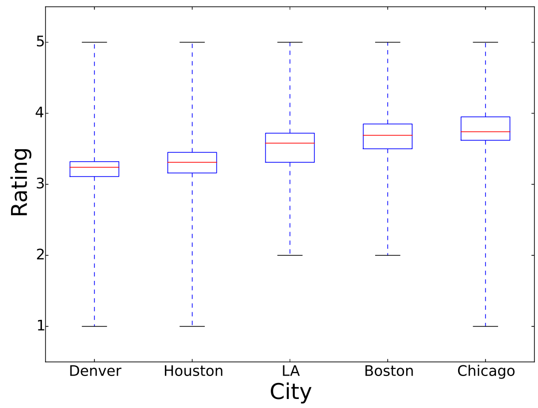


Fig. 8: The box plot showing ratings in five big cities of the United States.

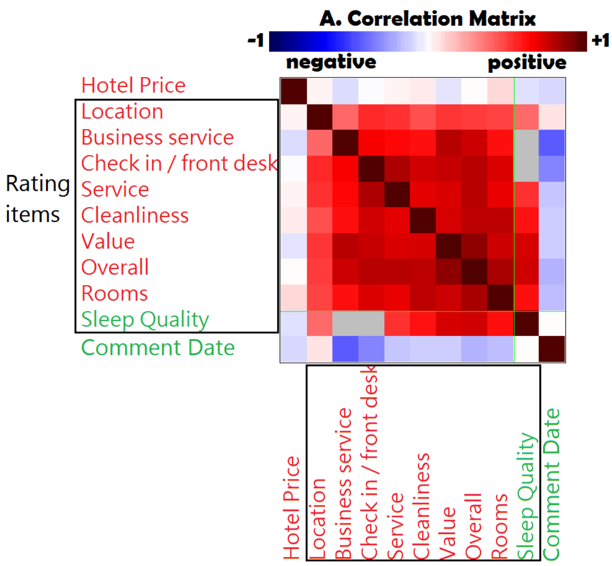
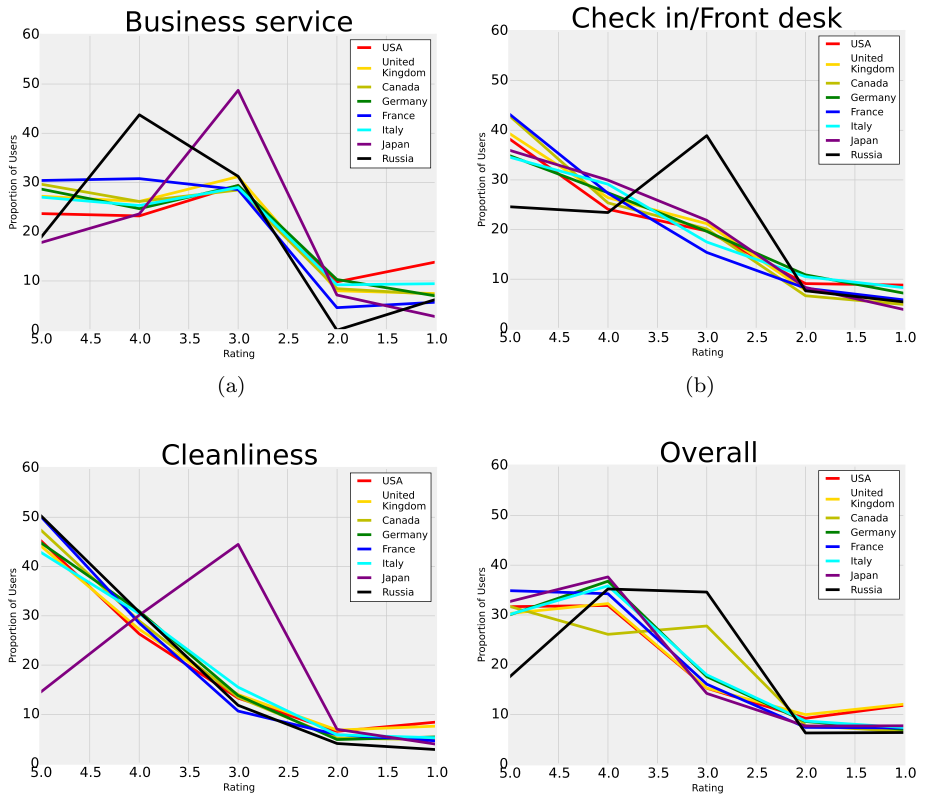


Fig. 9: Pearson correlation coefficients between nine rating items plus hotel price and user comment dates.

pectation that does not match with what these hotels have. Therefore, when they actually visit these hotels and experience the real situation, lower ratings may be given to reflect their disappointment.

3.2.2 User's Cultural Backgrounds and Hotel Ratings

Does travelers coming from different countries have different rating behaviors? To make our study focused, we especially consider travelers coming from countries in the G8 governmental political forum, i.e., Canada, France, Germany, Italy, Japan,



(a) (b) (c) (d)

Fig. 10: From left to right, top to down: the rating trends of different countries' travelers in terms of business service, check in/front desk, cleanliness, and overall.

Russia, UK, and US. Taking travelers from US as an example, for each hotel, we calculate the proportions of US travelers give 5 scores to 1 score on the business service rating item, respectively. The proportions of all hotels are then averaged to obtain the average trend of US travelers giving 5 scores to 1 score on the business service rating item. We call the average rating proportions as *the rating trend of US travelers in terms of business service*. The trends of different countries' travelers in terms of various rating items can be obtained similarly.

Fig. 10a shows the rating trends of different countries' travelers in terms of business service. We can see that the trends of Russian travelers (black) and Japanese travelers (purple) have peaks at 4 and 3 scores, respectively. On the other hand, roughly the same proportions of travelers from other countries give 5, 4, or 3 scores on the business service rating item. This means that Russian travelers and Japanese travelers relatively have higher expectations on business service and are carefully to give their highest scores.

Fig. 10b shows the rating trends in terms of check in/front desk. We can see again that Russian travelers (black) have behaviors distinct from other countries'

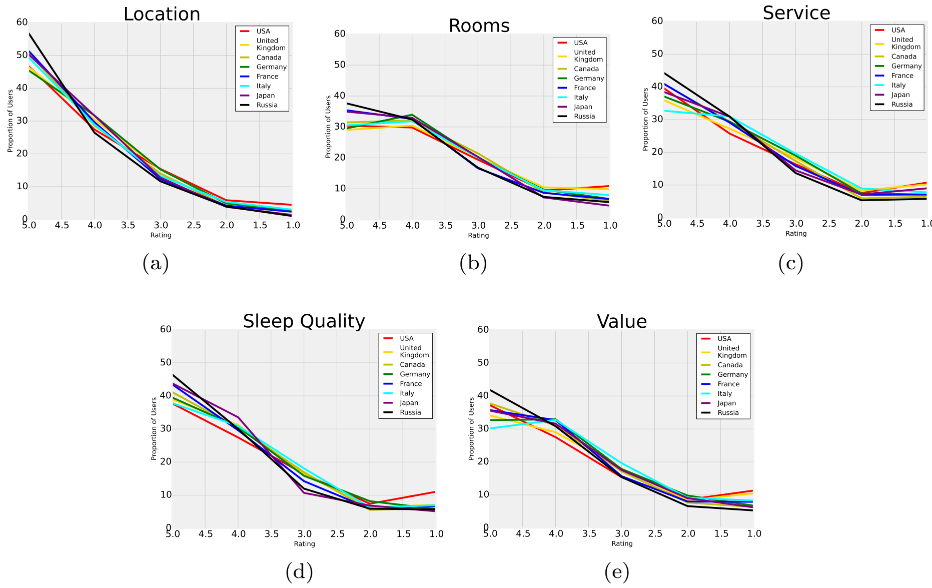


Fig. 11: The rating trends of different countries' travelers in terms of five other rating items.

travelers. They have relatively higher requirement on this item, and carefully give high evaluation (5 or 4) to hotels.

From Fig. 10c, we can see the rating trend of Japanese travelers in terms of cleanliness is quite different from other countries. They especially have high requirements in cleanliness.

Fig. 10d shows the rating trends in terms of the overall rating item. They can be categorized into four groups. The first group includes Japan, Germany, Italy and France. They have peaks at 4 scores. The second group includes US and UK, where peaks at 5 and 4 scores are almost the same. The third group includes Canada. Almost all Canadian travelers give scores higher than 3, and the proportion of 3 scores is much higher than other countries (except for Russia). The fourth group includes Russia. Russian travelers again merely give the highest score to the overall rating item, and most ratings concentrate on 4 and 3 scores.

Cultural difference does not exist in every rating item. Fig. 11 shows the rating trends in terms of the rest five rating items, where different countries' travelers have similar rating behaviors. The proportions basically decrease as the score decreases.

From Fig. 10 and Fig. 11, we see interesting and inconsistent rating trends given by travelers with different cultural backgrounds. Based on large-scale hotel rating data, this cultural difference in hotel ratings is explored for the first time in the literature and would give clues in developing an advanced hotel rating prediction system, which will be described in Section 4.

The aforementioned study shows how users with different cultural backgrounds rate worldwide hotels. In the following, we would like to investigate "how users with different cultural backgrounds rate hotels in different geographical areas." We focus on hotels in four areas with distinct cultures; meanwhile, the number of hotels in each area should be sufficient for study. Considering the distribution shown in Fig. 3, we mainly study on hotels in four areas: (1) North America (NA) including United States and Canada; (2) Europe (EU) including Italy, Spain, Germany, Turkey, United Kingdom, France, and Netherlands; (3) Far East (FE) including China and Japan; and (4) Middle East (ME) including United Arab Emirates. NA and EU have similar cultural backgrounds and can be viewed as the control pair; on the other hand, NA/EU, FE, and ME have distinct cultures (even FE and ME are both in Asia).

Fig. 12 shows rating trends in terms of business service, check in/front desk, cleanliness, and overall, for hotels located in North America and Europe. Rating trends in terms of other rating items are similar, and are omitted here due to space limitation. A few observations can be made. First, these rating trends may be slightly different from the trends shown in Fig. 10, which is not surprising. Rating trends for hotels in NA and EU are slightly different, too. For NA hotels, similar to Fig. 10, Russian give the most distinct rating behavior in check in/front desk, and Japanese give the most distinct rating behavior in cleanliness. However, Russian don't have distinct rating behavior on check in/front desk for EU hotels.

For FE and ME hotels, only rating trends in terms of business service, cleanliness, and overall, demonstrate cultural difference, and are shown in Fig. 13. Interestingly, the unique behavior of Japanese on cleanliness largely diminishes in these two areas. Jointly considering Fig. 10, Fig. 12, and Fig. 13, cultural difference exists in hotel rating, but the differences for hotels in different areas are varying. This makes rating prediction not only interesting but also challenging.

3.3 Visual Information and Hotel Ratings

This section describes how we study the relationship between visual information and hotel ratings. First, we construct a classifier based on the state-of-the-art deep learning features to categorize hotel photos into indoor or outdoor. Second, we utilize visual concept detectors to detect semantic concepts embedded in photos, and describe each photo by a concept vector. Finally, we show the correlations between ratings and hotel photo properties.

3.3.1 Indoor-Outdoor Classification

We are wondering that, *when the hotel administrator selects a photo showing interior appearance of a hotel as the cover photo, he/she wants to show their high-quality rooms or business service, and travelers really give higher ratings on these rating items?* To verify this conjecture, we propose one of the first works analyzing visual information provided on hotel booking platforms.

We first construct a classifier to classify each hotel's cover photo into indoor or outdoor. To represent a photo, CNN (Convolutional Neural Network) features are extracted. The AlexNet framework [23] pre-trained by the ImageNet dataset is utilized to extract features. Unlike hand-crafted features, CNN features are automatically learnt from a large-scale image dataset, conveying responses of various filtering at multiple resolutions. We use the result of the first fully-connected layer (fc7) and obtain a 4096-dimensional CNN features to represent each photo. Based

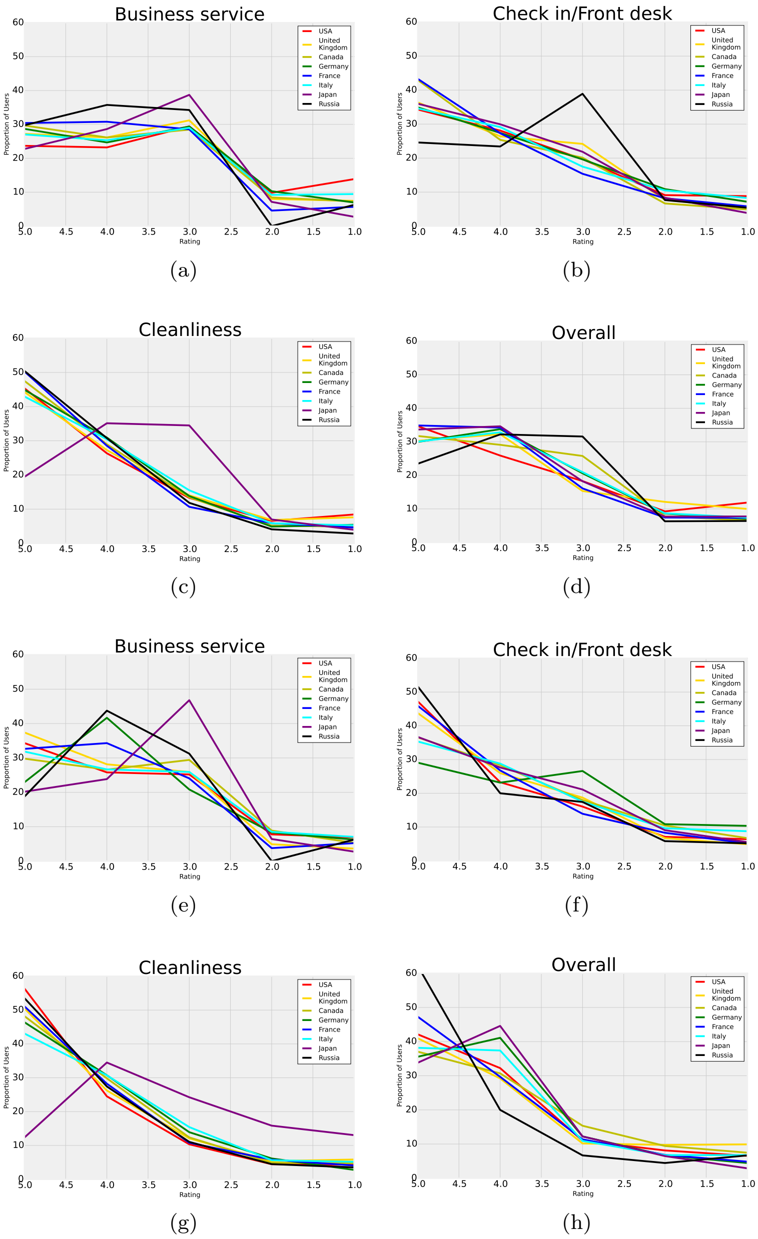


Fig. 12: The rating trends of different countries' travelers for hotels in North America (a) – (d) and Europe (e) – (h), respectively.

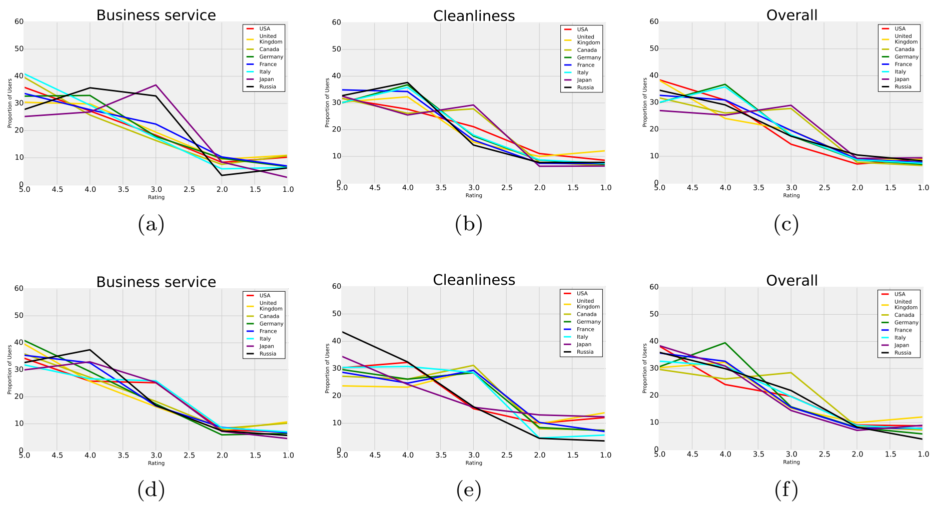


Fig. 13: The rating trends of different countries' travelers for hotels in Far East (a) – (c) and Middle East (d) – (f), respectively.

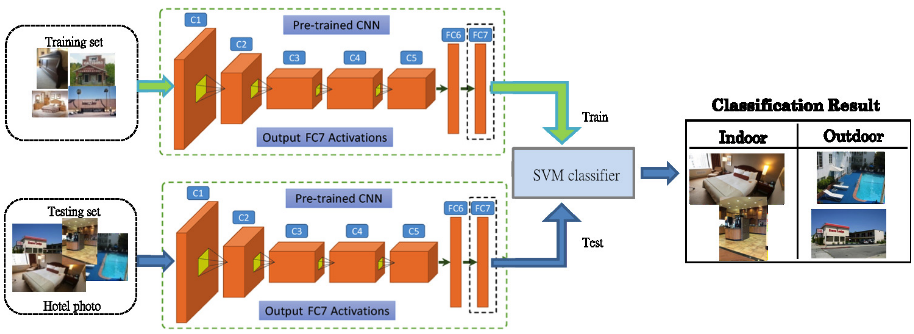


Fig. 14: The flowchart of indoor-outdoor image classification.

on CNN features extracted from the training data (1,124 outdoor photos and 6,000 indoor photos), we construct a classifier based on the support vector machine [24]. Fig. 14 shows the flowchart of indoor-outdoor image classification. With this classifier, we classify each hotel's cover photo into indoor or outdoor, and see whether the type of cover photo correlates with rating scores.

3.3.2 Semantic Concept Detectors

We are also curious about how visual semantic concepts correlate with rating behaviors. By evaluating the strength of visual concepts embedded in a photo, we can represent this photo as a concept score vector. In this work, we utilize the VIREO-374 concept detectors [3] to detect semantic concepts. The VIREO-374 concept detectors were built by support vector machines, based on bags of

Fig. 15: An example of semantic concept detection results. Top: a hotel's cover photo; bottom: the scores with respect to each visual concept.

local feature points as image representation. We extract scale-invariant feature transform (SIFT) descriptor from each photo, and a vocabulary of 500 visual words is used to quantize local feature points into visual words. With the visual vocabulary, each photo can be represented by a 500-dimensional vector, which is then input to the concept detectors to estimate how likely a visual concept is embedded in this photo.

It costs a lot of time to detect all 374 concepts. We, therefore, only detect 33 of 374 concepts that are related to hotels [25]. Fig. 15 shows an example of concept detection results of a hotel photo. Each bin within the histogram represents a concept score. We see that some concepts have high scores, like *outdoor*, *sky*, and *swimming pool*. Some results, however, are not so accurate, like *computer tv screen*. This reflects the limitation of visual concept detection, which is still an ongoing research direction pursued by many researchers.

3.3.3 Correlation between Visual Information and Hotel Ratings

Based on the aforementioned visual information, we discuss the relationship between them and rating items. Table 3 shows average scores and score differences in terms of different rating items when hotel's cover photos are indoor or outdoor, respectively. The score difference is calculated by subtracting average score of hotels

Table 3: Average scores of different rating items when hotel's cover photos are indoor or outdoor, respectively. The p-values show how significantly the scores differ in hotels with indoor cover photos and with outdoor cover photos.

	Rooms	Service	Business service	Cleanliness	Check in/front desk
Indoor	3.62	3.76	3.30	3.95	3.73
Outdoor	3.53	3.68	3.26	3.83	3.65
Difference	+0.09	+0.08	+0.04	+0.12	+0.08
p-value	**0.0070**	**0.0026**	0.4947	**0.0000**	0.0200
	Overall	Value	Location	Sleep Quality	
Indoor	3.62	3.72	3.98	3.75	
Outdoor	3.52	3.65	4.05	3.66	
Difference	+0.10	+0.07	-0.07	+0.09	
p-value	**0.0001**	**0.0031**	**0.0012**	**0.0000**	

with outdoor cover photos from average score of hotels with indoor cover photos. The statistical significance of such score difference is shown by p-values, where we especially emphasize the p-values smaller than 0.01.

An interesting observation from Table 3 is that, for the hotels with outdoor cover photos, the average score of the location rating item is consistently higher than hotels with indoor cover photos (p-value much less than 0.01). On the contrary, hotels with indoor cover photos generally have higher average scores in room, service, cleanliness, overall, value, and sleep quality rating items. This shows hotel's cover photos selected by the hotel administration somehow correlate with the promotion purposes or user's feelings. On the other hand, scores in the business service item seem irrelevant to the type of cover photo.

To show the correlations between visual concepts and rating items, conceptually we can construct a linear regressor to predict rating scores based on visual concept scores. For example, given a photo, we detect 33 visual concept scores and use them to predict the cleanliness score. The correlation between predicted values and actual values is measured by the squared multiple correlation coefficient R^2:

$$R^2 = SS_{Reg}/SS_{Total} \tag{1}$$

where SS_{Reg} is the sum of squared difference between regression results and the mean, and SS_{Total} is measure of total variation.

$$SS_{Reg} = \sum_i (\hat{y}_i - \bar{y})^2 \tag{2}$$

$$SS_{Total} = \sum_i (y_i - \bar{y})^2 \tag{3}$$

where y_i is the actual value of a rating item, $\hat{y}_i$ is the value estimated by regression, and $\bar{y}$ is the average of all y_i's. Here R^2 is a measure of the strength of the linear relationship between rating score y and 33 concepts $x_1, x_2, ..., x_{33}$ mentioned in Section 3.3.2. The value R^2 is nonnegative. When the value of the multiple correlation R^2 is close to 0, the regression equation barely predicts y better than chance. A value of R^2 close to 1 indicates a very good fit [26].

Table 4 shows multiple correlations [27] between each rating item and all concepts. We can see that the overall rating item has the highest R^2 value (> 0.6) and

Table 4: Multiple correlations between each rating item and all visual concepts.

Rating item	R^2 value	p-value
Overall	0.651	0.001
Rooms	0.505	0.003
Cleanliness	0.479	0.002
Value	0.451	0.002
Service	0.407	0.001
Location	0.212	0.003
Sleep Quality	0.089	0.069
Check in/front desk	0.072	0.030
Business Service	0.024	0.032

the smallest p-value (< 0.01). Statistically, the overall rating item has strong correlation with the detected visual concepts. Rooms, service, cleanliness, and value rating items have moderate correlation with these visual concepts, and the rest of rating items are weakly correlated with these visual concepts.

Both this result and the cultural difference mentioned previously are interesting findings and are first unveiled in the literature. In the following, we will take hotel rating prediction as an example to show how these findings can be used to provide better hotel recommendation.

4 Hotel Rating Prediction

In this section, we especially verify the effectiveness of cultural difference and visual information on hotel rating prediction. In Section 4.1, we discuss the basis of recommender system development and factorization machine, which is the state-of-the-art method for building a recommender system. We then report the influences of visual information and cultural difference on the recommender system in Section 4.2.

4.1 Recommender System

Recommender system is an application of information filtering [28][29], with the main function of predicting user's preference or suggesting candidate items to users. However, the data sparsity problem often impedes recommender systems. This problem refers to the difficulty in finding sufficient reliable similar users since most active users only rated a small portion of the items. In order to address this problem, we build a recommender system based on factorization machine [4], which was proposed to largely keep the sparsity problem off.

Factorization machine is a predictor like SVM, but it has better ability to address the sparsity problem because it models all nested variable interactions. Given various types of features, factorization machine captures weights of all single and pairwise interactions between variables. Assume that the data matrix is $\boldsymbol{X} \in \mathbb{R}^{N \times D}$, where the i-th row $\boldsymbol{x}^{(i)} = (x_1^{(i)}, ..., x_D^{(i)}) \in \mathbb{R}^D$ describes one record (including a user rating a hotel, plus the hotel information and user behaviors, in our case) with D real-valued variables. Usually, a two-way factorization machine model can be defined as:

$$\hat{y}(\boldsymbol{x}) = w_0 + \sum_{i=1}^{D} w_i x_i + \sum_{i=1}^{D} \sum_{j=i+1}^{D} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j \tag{4}$$

where the model parameter $w_0 \in \mathbb{R}$ is the global bias, w_i models the strength of the i-th variable, and $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$ denotes the dot product of $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ of size k, and models the interaction between the i-th variable and the j-th variable, i.e.,

$$\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = \sum_{f=1}^{k} v_{i,f} v_{j,f} \tag{5}$$

where $k \in \mathbb{N}_0^+$ is a hyperparameter that defines the dimensionality of the factorization.

Given a query vector $\boldsymbol{x}$, the value $\hat{y}(\boldsymbol{x})$ is the prediction result obtained by considering weights of different variables as well as the interactions between variables. A factorization machine has a closed model equation and can be computed in linear time. The model parameters $\theta = (w_0, \boldsymbol{w}, V)$, where $\boldsymbol{w} = (w_1, ..., w_D)$, and $V = \{\boldsymbol{v}_i\} \in \mathbb{R}^{D \times k}$, can be learned efficiently by the gradient descent method [4]:

$$\frac{\partial}{\partial \theta} \hat{y}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \theta \text{ is } w_0 \\ x_i & \text{if } \theta \text{ is } w_i \\ x_i \sum_{j=1}^{n} v_{j,f} x_j - v_{i,f} x_i^2 & \text{if } \theta \text{ is } v_{i,f} \end{cases} \tag{6}$$

Through reformulating the third term of eqn. 4, the third case of eqn. 6 can be derived. The value $\sum_{j=1}^{n} v_{j,f} x_j$ is independent of i and thus can be precomputed (when computing $\hat{y}(\boldsymbol{x})$).

We reduce the hotel recommendation task into predicting rating score, especially on the overall rating item, and formulate it as a regression problem. More particularly, we would like to build a function $T : \mathbb{R}^D \rightarrow \mathbb{R}$ that outputs a real-valued number $\hat{y}$ (predicted score) by giving a D-dimensional vector $\boldsymbol{x}$. We adopt the libFM library [4] to build this function, based on the training data $\{\boldsymbol{x}, y\}$ randomly selected from the CCU hotel dataset. Taking the overall rating as an instance, the rating score would be one to five, i.e., $y = \{1, 2, 3, 4, 5\}$. The data vector consists of the following subvectors:

- User vector $\boldsymbol{x}_U$: A binary vector indicating which user is rating or giving comments to a hotel. For example, the $\boldsymbol{x}_U$ subvector of the first record $\boldsymbol{x}^{(1)}$ in Fig. 16 is equal to $(1, 0, 0, \dots)$, which indicates that the first record is given by the first user in the database.
- Hotel vector $\boldsymbol{x}_H$: A binary vector indicating which hotel is rated. For example, the $\boldsymbol{x}_H$ subvector of the fourth record $\boldsymbol{x}^{(4)}$ in Fig. 16 is equal to $(0, 0, 1, 0, \dots)$, which indicates the third hotel in the database is rated.
- Date vector $\boldsymbol{x}_D$: A real number indicating the date this record was given. It is represented as the number of months since January, 2001. For example, the $\boldsymbol{x}_D$ subvector of the fourth record $\boldsymbol{x}^{(4)}$ in Fig. 16 is equal to 6, which indicates this record was given in June, 2001.
- Price vector $\boldsymbol{x}_P$: A real number indicating price of the room a user had stayed. For example, from $\boldsymbol{x}^{(1)}$, we can see that the first user paid 243 US dollars for the room of the first hotel.

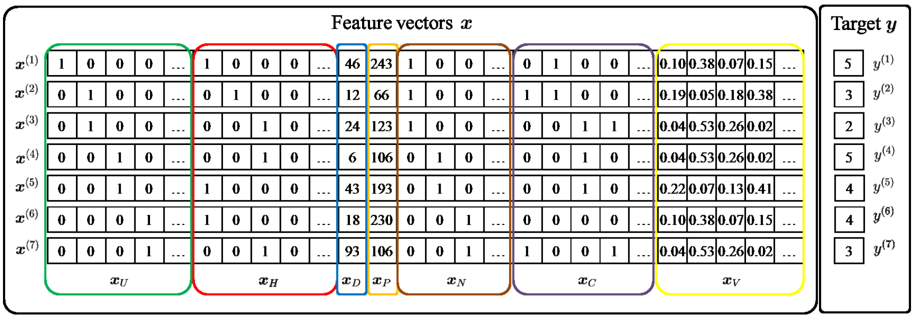


Fig. 16: An example of sparse real-valued feature vectors $\boldsymbol{x}$ created from user rating records. Each row represents a data record $\boldsymbol{x}^{(n)}$ with its corresponding target $y^{(n)}$.

- Nationality vector $\boldsymbol{x}_N$: A binary vector indicating which country the user comes from. Users in the database come from totally 213 countries, and thus $\boldsymbol{x}_N$ is 213-dimensional. The $(0, 0, 1, \dots)$ subvector $\boldsymbol{x}_N$ of $\boldsymbol{x}^{(6)}$ indicates that this user is from the third country in our database, i.e., France.
- Comment vector $\boldsymbol{x}_C$: A binary vector indicating which important words were used in a record. From the CCU hotel dataset, we calculate each word's TF-IDF value (term frequency/inverse document frequency) and determine fifty most "important" words by selecting ones with the highest TF-IDF values. The subvector $\boldsymbol{x}_C$ is thus a 50-dimensional binary vector. For example, in $\boldsymbol{x}^{(7)}$, the user gave a comment where the first and the fourth important words appear in the comment.
- Visual concept vector $\boldsymbol{x}_V$: A real-valued vector indicating the distribution of scores of 33 visual concepts mentioned previously. For example, the subvector $\boldsymbol{x}_V$ of the first record $\boldsymbol{x}^{(1)}$ shows that strength of the second visual concept is relatively higher than other concepts (0.38 vs. other values smaller than 0.20).

The aforementioned subvectors are concatenated as a long data vector $\boldsymbol{x}^{(i)}$. The five-fold cross-validation scheme is adopted to evaluate the rating prediction performance, i.e., at each run 80% of the data records are randomly selected as the training data, and the remaining 20% are for testing, and the average performance of five different runs is finally reported. Note that we can eliminate some subvectors on the purpose of evaluating the influence of different factors. In the experiments, we construct the factorization machine based on concatenation of all subvectors. At testing, we can evaluate if only basic information is available, i.e., testing vectors are formed by concatenating $\boldsymbol{x}_U$, $\boldsymbol{x}_H$, $\boldsymbol{x}_D$, and $\boldsymbol{x}_P$. To show the influence of visual information, for example, we can evaluate test vectors formed by $\boldsymbol{x}_U$, $\boldsymbol{x}_H$, $\boldsymbol{x}_D$, $\boldsymbol{x}_P$, and $\boldsymbol{x}_V$. Detailed experimental results are given in the following.

4.2 Influences of Visual Information and Cultural Difference

We employ three metrics to evaluate rating prediction performance: mean absolute deviation (MAD), mean square error (MSE), and Pearson correlation. Comparing the predicted overall ratings with the ground truths, we calculate MAD and MSE by:

$$MAD = \frac{1}{M}\sum_{i=1}^{M} |\hat{y}^{(i)} - y^{(i)}|, \tag{7}$$

$$MSE = \frac{1}{M}\sum_{i=1}^{M} (\hat{y}^{(i)} - y^{(i)})^2, \tag{8}$$

where $\hat{y}^{(i)}$ is the predicted value, $y^{(i)}$ is the ground truth, M is number of tests, and $|\cdot|$ is the absolute value of a number.

The definition of Pearson correlation is:

$$\rho_{\boldsymbol{y},\hat{\boldsymbol{y}}} = \frac{cov(\boldsymbol{y},\hat{\boldsymbol{y}})}{\sigma_{\boldsymbol{y}}\sigma_{\hat{\boldsymbol{y}}}} \tag{9}$$

where $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ are vectors of predicted values and ground truths, respectively. The value $cov(\cdot)$ is the covariance, i.e., $cov(\boldsymbol{y},\hat{\boldsymbol{y}}) = E[(\boldsymbol{y} - \mu_{\boldsymbol{y}})(\hat{\boldsymbol{y}} - \mu_{\hat{\boldsymbol{y}}})]$, and $\sigma_{\boldsymbol{y}}$ and $\sigma_{\hat{\boldsymbol{y}}}$ are the standard deviations of $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$, respectively.

By varying the information embedded in the test vectors, we evaluate the influences of different factors on rating performance. Table 5 shows prediction performance of the overall rating item with varying information. The first row shows prediction performance when we predict scores given by a user ($\boldsymbol{x}_U$) to a hotel ($\boldsymbol{x}_H$), assuming that we have acknowledged the price information of this hotel ($\boldsymbol{x}_P$) and known when the user wants to stay ($\boldsymbol{x}_D$). Note that, when training, $\boldsymbol{x}_P$ indicates price of the room a user stays, and $\boldsymbol{x}_D$ indicates when the user gave ratings. When testing, we view $\boldsymbol{x}_P$ as the average room price of the targeted hotel $\boldsymbol{x}_H$, and view $\boldsymbol{x}_D$ as the date when the user wants to stay in this hotel. The first row of Table 5 shows that if only such basic information is available, the prediction performance is not satisfactory (MAD=2.867 is quite large when rating scores range from 1 to 5).

If we further consider the hotel's cover photo or nationality, the prediction performance can be largely improved (MAD decreases from 2.867 to 1.132, and from 2.867 to 1.230, respectively), as shown in the second and the third rows of Table 5. These verify the effectiveness of considering visual information and nationality in hotel rating prediction. Moreover, if both clues are jointly considered, more performance improvement can be gained (MAD decreases to 1.083). If all vectors mentioned in Section 4.1 are jointly considered, the best performance can be obtained (the last row of Table 5). For clarity, we show performance gains of different settings in terms of MSE in the last column. By further considering visual information, for example, the improved MSE is calculated as $\frac{9.793-2.828}{9.793} \times 100\% = 71.12\%$.

In addition to the overall rating item, from Section 3.2.2 we see that cultural difference can also be seen from the business service, check in/front desk, and cleanliness rating items. We especially show the influences of different factors on cleanliness prediction in Table 6. Generally better performance can be obtained when more factors are jointly considered.

Previous works on hotel recommendation focused on prediction from text-based review comments. Wang et al. [1] proposed latent aspect rating analysis (LARA) to investigate both opinion ratings on topical aspects (e.g., location, service of a hotel) and the relative weights reviewers have placed on each aspect based on textual

Table 5: Overall rating predication performance.

	MAD	MSE	Pearson corr.	Imp. MSE
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P$	2.867	9.793	0.024	NA
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_V$	1.132	2.828	0.068	71.12%
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_N$	1.230	2.325	0.166	76.26%
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_V + \boldsymbol{x}_N$	1.083	1.611	0.304	83.55%
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_V + \boldsymbol{x}_N + \boldsymbol{x}_C$	0.821	1.132	0.373	88.44%

Table 6: Cleanliness rating predication performance.

	MAD	MSE	Pearson corr.	Imp. MSE
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P$	1.228	2.636	-0.010	NA
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_V$	0.967	1.783	0.325	32.36%
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_N$	1.000	1.666	0.287	36.80%
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_V + \boldsymbol{x}_N$	1.020	1.504	0.356	42.94%
$\boldsymbol{x}_U + \boldsymbol{x}_H + \boldsymbol{x}_D + \boldsymbol{x}_P + \boldsymbol{x}_V + \boldsymbol{x}_N + \boldsymbol{x}_C$	0.917	1.164	0.384	65.83%

Table 7: Rating prediction performance comparison between the proposed method and LARAM.

	MSE	Pearson corr.
LARAM [8]	1.234	0.228
Our method (overall rating)	1.132	0.373

review content. Later, the same research group proposed a unified generative model to improve LARA, and applied LARA to a wide range of applications, including rating recommendation [8]. The hotel collection same as our CCU hotel dataset was used in [8], and thus we can fairly compare our work with the results shown in [8]. Table 7 shows the performance comparison between our method and the LARA model (LARAM) [8], in terms of MSE and Pearson correlation. As can be seen, our proposed method is competitive or even better. This again shows the effectiveness of considering visual information and country information.

5 Conclusion

In this work, we present one of the first attempts to study how cultural factors and visual information influence hotel ratings. These novel perspectives enable new ways to develop better hotel recommendation systems, and may inspire behavioral studies on cultural difference in some ways. To facilitate the proposed studies, a web crawler is developed to extend an existing hotel collection to construct our CCU hotel dataset. Based on this large-scale dataset, several interesting characteristics are unveiled, e.g., hotel price is not tightly related to user's ratings; difference of rating behavior in five big cities of the United States. More importantly, we found that there is clear cultural difference in rating behaviors, like the uniqueness of Japanese to judge cleanliness, and the uniqueness of Russian to judge check in/front desk. In addition to text information, we also show that visual information derived from the hotel's cover photo can benefit rating prediction, which is a novel perspective in the field of hotel recommendation. By jointly considering the cultural factor and the visual factor, a hotel rating prediction system is built, and the influences of visual information and cultural difference are extensively verified.

In the future, we will collect more hotel properties like travel intent, or more user information, from TripAdvisor or other hotel ratings website, to more deeply investigate rating behaviors. Other interesting correlations between hotel properties and more visual information, like the ones extracted from user's uploaded photos, are also interesting.

References

1. Wang, H.-N., Lu Y., and Zhai, C.-X.: Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach, In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 783–792 (2010)
2. Szummer, M., and Picard R.W.: Indoor-Outdoor Image Classification. In Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Database, pages 42–51 (1998)
3. Jiang, Y.-G., Yang, J., Ngo, C.-W., and Hauptmann, A.-G.: Representations of Keypoint-Based Semantic Concept Detection: A Comprehensive Study. IEEE Transactions on Multimedia, vol. 12, no. 1, pages 42–53 (2010)
4. Rendle, S.: Factorization Machines. In Proceedings of IEEE International Conference on Data Mining, pages 995–1000 (2010)
5. Zhang, H.-L., Zhang, J., Lu. S.-J., and Liu, Z.-H.: Analysis of Spatial Autocorrelation and Price Determinants in Hotel Room Prices. In Proceedings of International Conference on Geoinformatics, pages 1–5 (2010)
6. Dai, W., and Lin, Q.: Service Design by Utilization of Content Analysis: A Case of Mid-priced Limited Service Hotels in China. In Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing, pages 13–18 (2014)
7. Rahayu, D.A., Krishnaswamy, S., Alahakoon, O., and Labbe, C.: RnR: Extracting Rationale from Online Reviews and Ratings. In Proceedings of IEEE International Conference on Data Mining Workshops, pages 358–368 (2010)
8. Wang, H.-N., Lu, Y., and Zhai, C.-X.: Latent Aspect Rating Analysis without Aspect Keyword Supervision. In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 618–626 (2011)
9. Schafer, J.B., Konstan, J., and Riedl, J.: Recommender Systems in E-Commerce. In Proceedings of ACM International Conference on Electronic Commerce, pages 158–166 (1999)
10. Zhang, K., Wang, K.Q., Wang, X.L., Jin, C.Q., and Zhou, A.Y.: Hotel Recommendation based on User Preference Analysis. In Proceedings of IEEE International Conference on Data Engineering Workshops, pages 134–138 (2015)
11. Paterek, A.: Improving Regularized Singular Value Decomposition for Collaborative Filtering. In Proceedings of ACM International Conference on Knowledge Discovery and Data Mining, pages 39–42 (2007)
12. Salakhutdinov, R., and Mnih, A.: Probabilistic Matrix Factorization. In Proceedings of Annual Conference on Neural Information Processing Systems, pages 1257–1264 (2007)
13. Saga, R., Hayashi, Y., and Tsuji, H.: Hotel Recommender System Based on User's Preference Transition. In Proceedings of IEEE International Conference on Systems, Man and Cybernetics, pages 2437–2442 (2008)
14. Guo, G.: Improving the Performance of Recommender Systems by Alleviating the Data Sparsity and Cold Start Problems. In Proceedings of International Joint Conference on Artificial Intelligence, pages 495–496 (2013)
15. Xiong, Y.-N., and Geng, L.-X.: Personalized Intelligent Hotel Recommendation System For Online Reservation. In Proceedings of IEEE International Conference on Management and Service Science, pages 1–5 (2010)
16. Hariri, N., Zheng, Y., Mobasher, B., and Burke, R.: Context-Aware Recommendation Based On Review Mining. In Proceedings of International Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, (2011)
17. Blei, D.M., Ng. A.Y., and Jordan, M.I.: Latent Dirichlet Allocation. Journal of Machine Learning Research, vol. 3, pages 993–1022 (2003)

18. Poriya, A., Patel, N., Bhagat, T., and Sharma, R.: Non-Personalized Recommender Systems and User-based Collaborative Recommender Systems. International Journal of Applied Information Systems, vol. 6, no. 9, pages 22–27 (2014)
19. Evangelopoulos, G., Zlatintsi, A., Potamianos, A., Maragos, P., Rapantzikos, K., Skoumas, G., and Avrithis, Y.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2012)
20. Liu, B., and Zhang, L.: A Survey of Opinion Mining and Sentiment Analysis. Mining Text Data, Springer, pages 415–463 (2012)
21. Lu, Y., Castellanos, M., Dayal, U., and Zha, C.-X.: Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach. In Proceedings of International Conference on World Wide Web, pages 347–356 (2011)
22. Chen, C.-H.: Generalized Association Plots: Information Visualization via Iteratively Generated Correlation Matrices. Statistica Sinica, vol. 12, pages 7–29 (2002)
23. Krizhevsky, A., Sutskever, I., and Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. In Proceeding of Annual Conference on Neural Information Processing Systems, pages 1106–1114 (2012)
24. Chang, C.-C., and Lin, C.-J.: LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, Article No. 27 (2011)
25. ESL vocabulary, http://www.manythings.org/vocabulary/lists/a/words.php
26. Multiple Regression Model, http://www.unesco.org/webworld/idams/advguide/Chapt5.htm
27. Abdi, A.: Multiple Correlation Coefficient, Encyclopedia of Measurement and Statistics, pages 648–651 (2007)
28. Pazzani, M.J.: A Framework for Collaborative, Content-based and Demographic Filtering. Artificial Intelligence Review, vol. 13, no. 5-6, pages 393–408 (1999)
29. Resnick, P., and Varian, H.R.: Recommender Systems, Communications of the ACM, vol. 40, no. 3, pages 56–58 (1997)
30. Hargreaves, C.A.: A Comparative Analysis of Hotel Ratings and Reviews: An Application in Singapore. American Journal of Marketing Research, vol. 1, no. 3, pages 118–129 (2015)
31. Hargreaves, C.A.: Analysis of Hotel Guest Satisfaction Ratings and Reviews: An Application in Singapore. American Journal of Marketing Research, vol. 1, no. 4, pages 208–214 (2015)
32. Banerjee, S., and Chua, A.Y.K.: A Study of Manipulative and Authentic Negative Reviews. Proceedings of International Conference on Ubiquitous Information Management and Communication, Article No. 76 (2014)
33. Martin, L., Sintsova, V., and Pu, P.: Are Influential Writers More Objective? An Analysis of Emotionality in Review Comments. Proceedings of International Conference on World Wide Web, pages 799–804 (2014)
34. Zhou, J., Albatal, R., and Gurrin, C.: Applying Visual User Interest Profiles for Recommendation and Personalisation. Proceedings of International Conference on Multimedia Modeling, pages 361–366 (2016)
35. Li, H., Lin, R., Hong, R., and Ge, Y.: Generative Models for Mining Latent Aspects and Their Ratings from Short Reviews. Proceedings of International Conference on Data Mining, pages 241–250 (2015)
36. Rianthong, N., Dumrongsiri, A., and Kohda, Y.: Improving the Multidimensional Sequencing of Hotel Rooms on an Online Travel Agency Web Site. Electronic Commerce Research and Applications, vol. 17, pages 74–86 (2016)
37. Wei, X., Luo, X., Li, Q., Zhang, J., and Xu, Z.: Online Comment-Based Hotel Quality Automatic Assessment Using Improved Fuzzy Comprehensive Evaluation and Fuzzy Cognitive Map. IEEE Transactions on Fuzzy Systems, vol. 23, no. 1, pages 72–84 (2015)
38. Lin, K.-P., Lai, C.-Y., Chen, P.-C., and Hwang, S.-Y.: Personalized Hotel Recommendation Using Text Mining and Mobile Browsing Tracking. Proceedings of IEEE International Conference on Systems, Man, Cybernetics, pages 191–196 (2015)
39. Chen, Y.-S., Chen, L.-H., and Takama, Y.: Proposal of LDA-based Sentiment Visualization of Hotel Reviews. Proceedings of International Conference on Data Mining Workshops, pages 687–693 (2015)
40. Traub, M., Kowald, D., Lacic, E., Schoen, P., Supp, G., and Lex, E.: Smart Booking without Looking: Providing Hotel Recommendations in the TripRebel Portal. Proceedings of International Conference on Knowledge Technologies and Data-driven Business, Article No. 50 (2015)
41. Hwang, S.-Y., Lai, C.-Y., Chang, S., Jiang, J.-J.: The Identification of Noteworthy Hotel Reviews for Hotel Management. Pacific Asia Journal of the Association for Information Systems, vol. 6, no. 4, pages 1–17 (2014)

42. Minnich, A.J., Chavoshi, N., Mueen, A., Luan, S., and Faloutsos, M.: TrueView: Harnessing the Power of Multiple Review Sites. Proceedings of International World Wide Web, pages 787–797 (2015)
43. Xiang, Z., Schwartz, Z., Gerdes Jr., J.H., and Uysal, M.: What can Big Data and Text Analytics Tell Us about Hotel Guest Experience and Satisfaction? International Journal of Hospitality Management, vol. 44, pages 120–130 (2015)
44. Banerjee, S., and Chua, A.Y.K.: In Search of Patterns among Travellers' Hotel Ratings in TripAdvisor. Tourism Management, vol. 53, pages 125–131 (2016)