An Analysis of Social Features Associated with Room Sales of Airbnb

Donghun Lee
Human Centered Computing Lab
Seoul National University
donghun@snu.ac.kr

Woochang Hyun
Human Centered Computing Lab
Seoul National University
woochang@snu.ac.kr

Jeongwoo Ryu
Human Centered Computing Lab
Seoul National University
jeongwoo@snu.ac.kr

Woo Jung Lee
Princeton University
wool@princeton.edu

Wonjong Rhee
Applied Data Science Lab
Seoul National University
wrhee@snu.ac.kr

Bongwon Suh
Human Centered Computing Lab
Seoul National University
bongwon@snu.ac.kr

CSCW '15 Companion, Mar 14-18 2015, Vancouver, BC, Canada
ACM 978-1-4503-2946-0/15/03.
http://dx.doi.org/10.1145/2685553.2699011

Abstract
Lately, collaborative consumption has emerged as an important socio-economic model because of its economic and environmental impacts. Airbnb, an online hospitality rental service provider, is a fast growing company that utilizes rich social communications. In this paper, we aim to quantitatively characterize collaborative consumption behaviors in Airbnb. We collected and analyzed a total of 4,178 room data, and investigated which features are more strongly associated with room sales. Besides the well-recognized room features like price, minimum stay, and amenities, our result shows that social features such as the responsiveness of host, the count of Wish List, the number of reviews, and the membership seniority are significantly associated with room sales. On the other hand, some of the conventional social features such as overall rating and number of references turned out to be not so critical for room sales.

Author Keywords
Collaborative consumption; Airbnb; data-driven research; social factor

ACM Classification Keywords
H.5.3 [Information Interfaces]: Group and Organization Interfaces - Collaborative computing, Web-based interaction, Computer-supported cooperative work

Category	Feature
Social Factor	#References #Reviews Total #Reviews Membership seniority Response time Response rate #Saved to Wish List User Scores (overall, value, accuracy, cleanliness, location, check-in, communication)
Room Factor	#Photos #Bedrooms #Bathrooms #Accommodation #Beds House type Room type Bed type Cancellation policy Minimum stay Price Deposit Cleaning fee Amenities (31 of them e.g. AC, Wi-Fi, etc.)

Table 1. Categorized features into two factors: social and room factors

Introduction

Collaborative consumption is a socio-economic model that enables individuals to collaboratively make use of goods and services via sharing, bartering, and gifting [2]. By the virtue of the proliferation of information technology that reduces transaction cost and makes connections between people, more and more people are participating in this system for the purpose of economic benefits and resource savings.

Previous studies on collaborative consumption suggest that reputation system for creating trust between buyers and sellers is crucial to peer-to-peer exchanges [1][3][6]. Ikkala and Lampinen [4], in this regard, presents that hosts divert their accumulated reputational capital into the rental price. While earlier works focused on the reputation system with qualitative approach, we aim to conduct a data-driven research to find which features are strongly associated with exchanges in collaborative consumption systems.

We focus on Airbnb, a peer-to-peer hospitality rental service for travelers, to investigate the effects of social features on users' willingness to purchase in collaborative consumption. We have chosen Airbnb because of the following two reasons. First, it is a good platform to conduct quantitative research because multiple features including rich social features. Second, it is one of the most cited examples of collaborative consumption services [7].

In this paper, we use a data-driven approach to explore the listings of Airbnb, and analyze which social features have actual influences on room sales.

Methodology and Datasets

We have collected Airbnb data of five cities in the United States. They are New York City, Chicago, Los Angeles, San Francisco, and Seattle. We have randomly selected 1,024 rooms per each city and we crawled the room data twice on August 1st and on October 1st.

We collected a total of 5,120 unique room data. Over the two months, 812 rooms were purged from the listing, thus removed from our analysis. 130 rooms were ignored since they didn't have complete data. As a result, a total of 4,178 room data were investigated. Because the price of rooms (Mean=$110, SD=83.7) and the number of saved to Wish List (Mean=229.6, SD=321.7) were skewed, those features are log normalized for the analysis.

Among collected features, we manually labeled 59 features into two categories, social factor and room factor. We were able to obtain 14 features in social factor as presented in Table 1. In this paper, we used these 59 features to analyze the significance of features on room sales.

Analysis and Results

Our goal in this paper is to figure out how significant effect each feature has on room sales. Our dataset contains a measure changing over two months, a quantity change in the number of reviews for each room (*review delta*). We used the *review delta* as the proxy of the minimum number of room reservations over time. This is because Airbnb allows its users to write a review only when he or she actually reserves a room. Since it was not possible to obtain actual room sales data, *review delta* could serve as a robust proxy of the minimum room sales.

	Coefficient
#Wish Lists	1.4257 ***
Essentials	1.0130 ***
Cleanliness score	0.9363 ***
Host's response time	0.8757 ***
Room type	0.6518 ***
Gym	0.6215 *
Shampoo	0.5736 ***
#Accommodation	0.1754 **
House type	0.0442 *
#Reviews	0.0240 ***
Host's response rate	0.0156 .
Total #review	0.0019 **
Membership seniority	-0.0022 ***
#Photos	-0.0118 .
Cleaning fee	-0.0145 ***
#References	-0.0648 .
Cancellation policy	-0.0766 .
#Bedrooms	-0.2421 .
Pets	-0.3093 .
Kitchen	-0.3656 *
Intercom	-0.4383 **
TV	-0.4697 ***
Minimum stay	-0.5578 ***
Communication score	-1.0831 ***
AC	-1.2525 ***
Adj. R-squared:	0.7976
p-value:	< 2.2e-16

Table 2. The result of multiple linear regression without intercept. Features in social factor are indicated in bold.
***: p<0.0001, **: p<0.001,
*: p<0.01, .: p<0.05

To quantitatively determine the significance of the features on predicting room sales, we examined the correlation between the room sales and the features by using multiple linear regression without intercept. Although the intercept of a linear model with intercept was statistically significant ($p = 0.000467$), it is not logical to predict the amount of room sales without any features of the rooms, thus the intercept is eliminated from the model. With the help of stepwise feature selection, our model was constructed with an adjusted R^2 of 0.7976, and the number of selected features is 25 with 9 out of 14 features in social factor. We found that features in social factor play a significant role in predicting the room sales. On the other hand, 16 out of 45 room features were selected. Popular room features such as Wi-Fi, dryer, and washer are not critical. The result shows that, to our surprise, conventional features have limited effect on the room sales in Airbnb. To better understand a social dynamics, we further analyzed the social features that outstand in our model.

Social Features That Matter

The Responsiveness of Hosts

The *response rate* and the *response time* represent how quickly and consistently a host reply to guests who want to stay in the listing. They can be generally interpreted as a quantitative measure of how active the host participates in social communication with their potential guests. As shown in the Table 2, hosts' responsiveness is a significant feature in our model to predict room sales. To understand how the responsiveness affects the dependent variable, *review delta*, we conducted a one-way ANOVA to test if there is any prominent difference between three response time groups (within an hour, within a few hours, within a day). As a result, the faster hosts respond to their customers, the more likely rooms were reserved in the same time period ($F(2, 4175) = 41.85$, $p < 2.0X10^{-16}$) as in Table 3. Each group shows statistically different means (all of the p-values in Tukey HSD test were below 0.05).

The Count of Wish List

'Save to Wish List' is a feature from Airbnb that allows users to bookmark their listings of interest. The number of saved to Wish List could be used as a quantitative measure of the public's interest on the room in the system. The change in the number of reviews (*review delta)* in the research period increases as the count of Wish List increases. Not only the Table 4 has shown these differences, but also the results of ANOVA and Tukey HSD indicated statistically different means ($F(3, 4174) = 93.78$, $p < 2.0X10^{-16}$). This implies that the count of Wish List, which is open to public, is significantly associated with the future room sales.

The Number of Reviews

Customer reviews are crucial component of reputation systems for online services and communities [5]. In Airbnb, guests who stayed in a room can qualitatively enrich the review by writing a free description of their experiences and scoring in seven categories. To our surprise, the quantity of reviews is more important than their ratings. Five out of the seven ratings (including overall rating) are excluded from our model, which implies they are not critical enough to predict room sales. On the other hand, a one-way ANOVA shows strong relationship between room sales and review numbers ($F(2, 4175) = 208.8$, $p< 2.0X10^{-16}$). Post-hoc test confirms that there are significant differences between each group as in Table 5 (all of the p-values in Tukey HSD test were below 0.01). This

Response time	N	Mean	SD
Within an hour	1971	8.19	5.10
Within a few hours	1804	7.18	4.74
Within a day	403	6.03	4.26

Table 3. Quick response facilitates room sales. Mean and SD are about review delta.

#Wish Lists	N	Mean	SD
Btw 0 ~ 10	82	4.60	4.10
Btw 11 ~ 100	1680	6.25	4.59
Btw 101 ~ 1000	2290	8.42	4.86
More than 1000	126	10.48	5.47

Table 4. More number of Wish Lists facilitates room sales. Mean and SD are about review delta.

#Reviews	N	Mean	SD
Btw 0 ~ 10	1171	5.62	4.29
Btw 11 ~ 100	2815	8.03	4.81
More than 100	192	12.2	5.23

Table 5. More number of room reviews facilitates room sales. Mean and SD are about review delta.

Membership Seniority (days)	N	Mean	SD
Btw 0 ~ 100	246	7.67	5.18
Btw 101 ~ 1000	2851	7.76	4.93
More than 1000	1081	6.95	4.76

Table 6. It takes longer time for old users to sell their room. Mean and SD are about review delta.

indicates rooms with more reviews are more likely to be booked overtime. One explanation could be that most reviews in Airbnb are highly positive, and thus the scores have less distinguishing power.

The Membership Seniority

The membership seniority (Mean=699 days, SD=439) was negatively correlated as in Table 2, suggesting that the host who registered lately would have higher room sales. Hosts are categorized into three groups as shown in Table 6. A one-way ANOVA shows significant difference among groups on predicting room sales ($F(2, 4175) = 10.7$, $p < 2.32X10^{-5}$). Post-hoc test confirms that there is significant difference between groups of users who registered more/less than 1000 days. Further study is needed to investigate what makes these differences between groups and the relationship between the seniority and other features such as response time, the number of Wish Lists and reviews.

Conclusions and Discussion

In this paper, the effect of social and room features of Airbnb service to the room sales was analyzed. Our result shows that social features such as the responsiveness of hosts, the number or reviews, the membership seniority, and the count of Wish List are significantly associated with room sales as well as the room features like price, minimum stay, and amenities. Surprisingly, conventional social features such as overall rating and number of references are not as critical to the room sales. Room features involving Wi-Fi, dryer, and washer are not significant, either. The result points out that the important features of modern collaborative consumption systems, and provides useful insights on future collaborative consumption system design. Some limitations apply to this study. The dataset was partial and the variation of review number cannot indicate the precise number of sales but only the comparative tendency. Some qualitative features (e.g. impressions from photos or locational benefits) were ignored, too.

Acknowledgement

This work was supported by the Energy Efficiency & Resources Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the MOTIE, Republic of Korea. (*No. 20132010101800*)

References

[1] Belk, R., You Are What You Can Access: Sharing and Collaborative Consumption Online. *Journal of Business Research*, *67*(8) (2014) 1595-1600.

[2] Botsman, R., Rogers, R. What's Mine Is Yours: The Rise of Collaborative Consumption. Harper Collins, 2010.

[3] Gheitasy, A., et al. Designing for Online Collaborative Consumption: A Study of Sociotechnical Gaps and Social Capital. In *Human-Computer Interaction. Applications and Services* (2014) 683-692.

[4] Ikkala, T. and Lampinen, A. Defining the Price of Hospitality: Networked Hospitality Exchange via Airbnb. *CSCW'14 Companion*, ACM (2014) 173-176.

[5] Jøsang, A., et al. A Survey of Trust and Reputation Systems for Online Service Provision. *Decision support systems*, *43*(2) (2007) 618-644.

[6] Leonard, L. N., and Jones, K. Consumer-to-Consumer e-Commerce Research in Information Systems Journals. *Journal of Internet Commerce 9*(3-4) (2010) 186-207.

[7] Teubner, T. Thoughts on the Sharing Economy. *Proc. of the International Conference on e-Commerce.* Vol. 11 (2014) 322–326.