Received January 1, 2017, accepted January 18, 2017, date of publication August 3, 2017, date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2735242

Design of A Universal User Model for Dynamic Crowd Preference Sensing and Decision-Making Behavior Analysis

MOLOUD ABDAR AND NEIL Y. YEN, (Member, IEEE)
School of Computer Science and Engineering, University of Aizu, Aizu 965-8580, Japan
Corresponding author: Neil Y. Yen (neil219@gmail.com)

ABSTRACT Sharing economy becomes an emerging issue in urban life. It is not a new phenomenon but an assembling of existing techniques to meet specific demands of users. It also points out a better way to implicitly collect users' contexts and to understand users than the conventional one that requires much user involvement (e.g., tedious inputs). A universal model, for this purpose, that supports dynamic analysis and mining of user-generated content (or contexts) is designed in this paper. Two major factors, sensing and analysis of crowd preference and their decision-making behavior, are especially targeted. This model formulates the given scenario that comprehensively illustrates the possible actors and correlated actions among them with a set of rules to enhance the machine learning results. This model outlines a detail process on pre-/post-process of the data, and indicates the core techniques for user modeling. The raw data collected from on-service website, i.e., Airbnb, are utilized for the preliminary examination of our proposal. We especially look at internal factors (e.g., nationality, gender, and age) and external factors (e.g., device, social media, and time) that reflect implicitly the difference on crowd's preference and behavior. Results after statistics-based machine learning reveal that the relation among users' internal and external factors share high similarity with their behavior patterns, and can be applied, considering particular features, for service provision to a specific type of crowds.

INDEX TERMS User understanding, crowd preference, human behavior, statistical analysis, human-centric computing, sharing economy.

I. INTRODUCTION

The 21st century stands for an era of innovation in most fields of study (e.g., medical science, computer science, economic, human science, and etc.). Significant progresses have been achieved and caused impacts to our living environment. One new phenomenon–Sharing Economy– that redefines the ownership among users receives wide discussions. Delivery service, for instance, such as package, mail, and food, may best describe the scenario. An important point under this scenario is decentralization. The service provider migrates from a firm to an individual, and a new item to an existing (seldom used) item. The above instance just indicates a simple case, and it can be widely applied to all the idling resources, no matter in the form of tangible or intangible, around our living environment with the support of urban technology.

Human behaviors can be observed via, and under, the sharing economy. Issues such as trust, preference, emotion and decision-making concerning human behavior have attracted significant attentions from researchers in recent years. Different types of datasets become available and specific behavior(s) may also be monitored according to data collected from sharing economy platforms (e.g., accommodation sharing, car sharing, food sharing [1], selfish sharing [2], etc.). Taking two studies above for instance. Researchers [3] have pointed out that the facial expressions may cause influence on the difference on decision on sharing economy platform that requires high interaction with users. Different expressions (e.g., joy, angry, sad, and neutral) on personal profile picture may lead to different feelings as well as purchase intention to potential users even other objective factors, such as star-rating and comments, are positive/negative. The other case [4] designed a framework for implementation of user-based relocation according to station-based one-way car sharing. Users, thus, have shown some implicit behaviors while using such sharing economy platforms, and these behavior data may be a new channel to achieve better understanding of end users.

Mining, and analysis as well, human behaviors is a way to achieve user understanding. The better we can understand the target users, the more we can serve them. Human behaviors can mainly be separated into external or internal contexts. A simple definition is that external contexts are factors from outside such as culture, device, social media, and time while internal contexts are those factors from inside such as nationality, gender, age, preferences or even prior experiences. All these contexts may lead to the changes on decision-making process. For instance, users may have additional concerns due to their age and gender, or users may change their willing to rent specific type of accommodation or car according to the situation. This phenomenon implicitly indicates that sometimes users may make decisions intentionally or unintentionally based on internal/external contexts, and thus mining such contexts to have better understanding may become an urgent issue especially in the scenario of sharing economy.

Users may possess similar behaviors if they share similar contexts, no matter from internal or external, in common. Access to these common behaviors are useful for researchers, as well as companies, to provide better services adaptive to specific users than without it. For instance, uses' behaviors may be different in accordance with their country, gender, age, and etc. Thus, in order to discover and identify those implicit human behaviors, machine learning methods, such as Decision Tree, SVM, Neural Network, Apriori and etc., were often applied to meet particular purposes. But however, most of the learning results are expected to obtain hidden patterns that are previously unaware. As a result, better services are prospective if human behaviors can be modeled.

Considering the growing popularity in sharing economy [5], [6], this study attempts to analyze decision-making process through a universal model that supports dynamic and crowd sensing and analysis of user-generated content as well as contexts effectively [7]. The model is named by DKDAR Model (abbreviation of Data, Knowledge, Decision, Algorithms and Rules). The behavior data extracted from the user-generated content and its related contexts is especially concentrated. This study particularly looks at the internal factors (e.g., nationality, gender, and age) and external factors (e.g., device, social media, and time) that reflect implicitly the difference on crowd's preference and behavior. This paper first outlines the process of modeling at abstract level, and then, comes up with a concrete example for verification the feasibility. Then a set of machine learning algorithms are implemented for retrieve the features on users' preference as well as behavior. The dataset, which is collected through crawler from Airbnb is taken as an instance to verify the proposed model.

Following the introduction, related studies are summarized in section II. Section III gives a definition and design on the universal model. Section IV discusses the findings from the data. Section V then concludes the work and points out potential issues and challenges that can be considered in the future.

II. LITERATURE REVIEW

Activities concerning sharing economy become popular, and meanwhile, indicates a new way for human understanding. It offers a peer-to-peer analysis scenario [5] compared to the conventional top-down one [8], and this change makes the analysis results become much closer to the real situation. This section summarizes the works concerning human behavior analysis via accommodation rental service, and comes up with the proposed work.

A. TYPES OF ACCOMMODATION RENTAL SERVICES ARE CHANGING

Types of accommodation rental are changing. The hotel, or hostel -a new form of hotel- is the main until the peer-to-peer accommodation rental services appear in recent. The traditional type of accommodation is in a top-down form. Users can book the rooms from the official website or other third-party sites such as TripAdvisor or Booking.com [9] directly, and the interaction is mainly between company and user. The new type of accommodation is in a decentralization form (peer-to-peer) that increases the interaction between users. The idea of sharing economy also prompts the development of peer-to-peer accommodation. There are many instances of peer-to-peer accommodation rental services such as Airbnb, Wimdu or 9Flats [10] that implement the idea of sharing economy. Airbnb has been recognized the most successful and the largest platform [11] that implements the idea of sharing economy.

Sharing economy platforms are new but has caused great impacts to the publics. Results in [12] indicated that sharing economy plays an effective role in solving the unemployment issue, and causes a great effects on tourism industries in many countries.

Interaction among users, especially the host and the guest in precise, becomes important while a decentralized accommodation rental is taken place. The contexts related to the users may be key factors to the success of such scenario. Following the instance above, study [13] discussed the importance of photos taken by the host and their impacts to the decisions made by guests on Airbnb. Furthermore, researchers [3] raised an issue of the seller's facial expressions and its impact on guests' selection and behavior on Airbnb. It indicates that different facial expression may cause difference on guests' feelings and their decision-making results.

B. CROWD SENSING IN SHARING ECONOMY (AIRBNB)

Sharing economy, as a socio-economic system and new phenomenon, is a coordinator in "the peer-to-peer-based activity". In this platform, users can share, gain, and/or give their tools, services and etc., by using "community-based online services" [5]. Crowd-sensing/-sourcing indicates a process on utilization, integration, and analysis of huge and

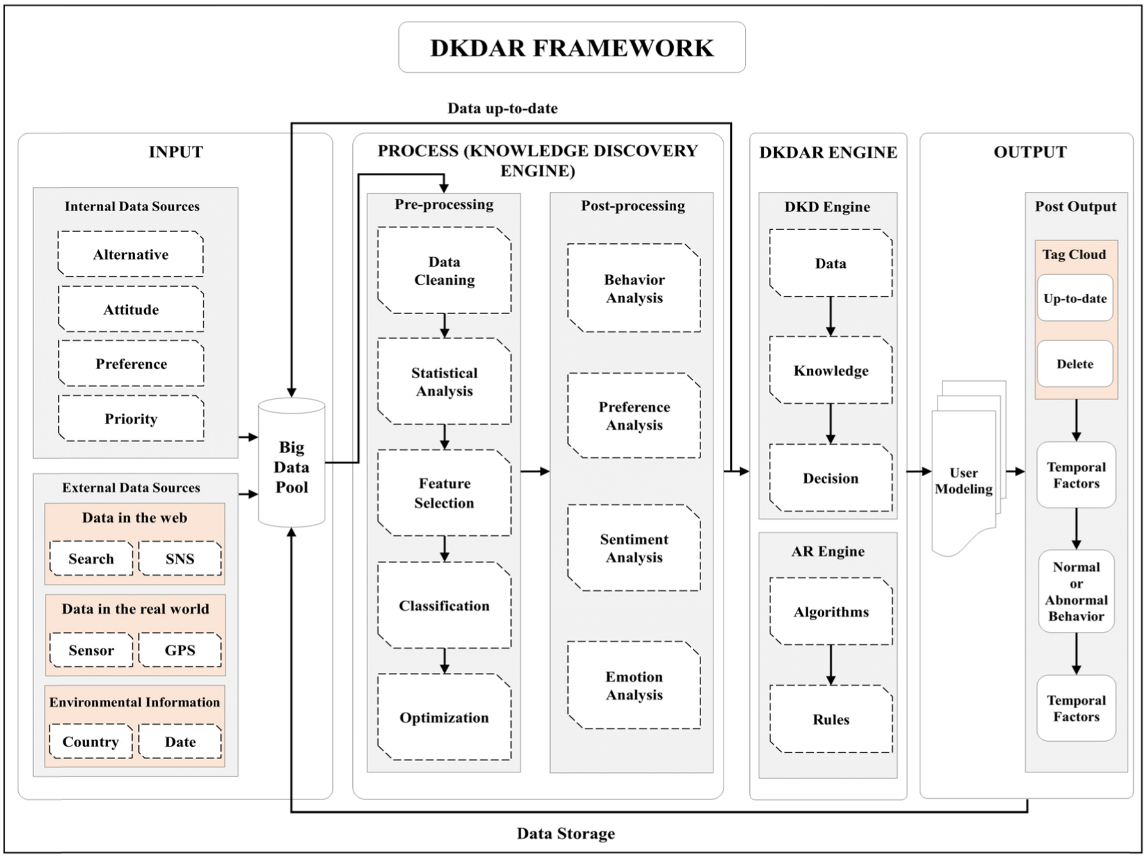

FIGURE 1. DKDAR framework for Human Behavior Analysis/Prediction.

heterogeneous data to explore, especially those hidden and implicit, information and experiences [14]. To better serve users who have particular needs, a comprehensive understanding to the target user, and user groups as well, is an urgent challenge [15].

Research [16] indicates that some municipalities are allowed to operate on Airbnb in their cities without imposing any regulation, whereas others are not allowed to operate without warranting. Research [17] then researched on user-generated brands (UGBs) to get a better understanding of this relatively new branding phenomenon. The discursive and visual results indicated that there is a significant relationship between members with the aim of sustainability discourse and inter-personal exchange in both brands. Finally, they argued that social media platforms have significant effect on member's trust with each other.

Research [18] points out that many Airbnb rentals are basically illegal that have expressed based on short-term rental regulations as study [16] mentioned that municipalities do not allow to such kind of activities. Other findings showed that some people prefer to use traditional accommodation because of security concerns and unpredictable experience According to obtained outcomes in [19], they reported that about 95 % of Airbnb properties had an average user-generated rating of either 4.5 or 5 stars (the maximum), and virtually none of them have less than a 3.5 star rating. The results revealed that the highest ratings (4.5 stars and above) on Airbnb are more than on TripAdvisor even though the average ratings on Airbnb and TripAdvisor were similar. Research [20] conducted an investigation on the effect of personal profiles and even to post pictures of individuals present in Airbnb. They expressed that non-black hosts charge about 12 % more than black hosts for the equivalent rental. Study [21] focused on various age groups in sharing economy and particularly in Airbnb. According to their outcomes, new generation are more interested to sharing economy compared to the other age groups. They also expressed that sharing economy as new phenomenon has strong impact on the human behavior.

III. DYNAMIC USER MODELING

Following the experiences and findings from sharing economy, this section goes further to discuss a universal model, namely (abbreviated Data, Knowledge, Decision, Algorithms, and Rules), that gives a clear scenario for user modeling. Two major parts, i.e., DKDAR framework (see Fig. 1) and an instance of data processing (see Fig. 2), are discussed.

A. DKDAR FRAMEWOR

(1) INPUT includes three components, namely External Data Sources (EDS), Internal Data Sources (IDS) and Big Data

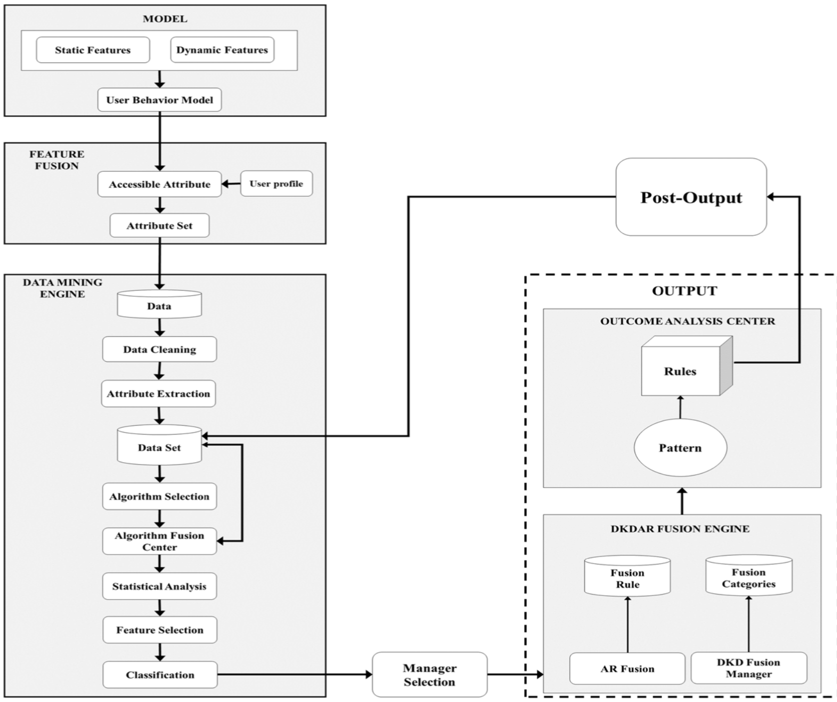

FIGURE 2. An instance of data processing under DKDAR.

Pool (BDP), for the universal data collection. IDS is simple. It refers to those data from inside user. Instances such as characteristics, attitude, preference, and etc., are included. EDS refers to those data from outside such as the data that user gives or receives on the Web, the data sensed through user's carryon devices, and surrounded environmental information. BDP then refers to a data storage pool that include a set of rules to categorize retrieved data from IDS and EDS.

(2) Process (Knowledge Discovery Engine) includes two stages, pre-processing and post-processing. Pre-processing stage consists of five steps to ensure the quality of the data to be further analysis. Noises, e.g., empty columns, incomplete records, etc., need to be removed as many collected data, especially raw data, is useless and may lead to extra costs while being mined. After removal of the noises from raw data, we go further to conduct the statistical analysis and feature selection processes. Association rules are revised to meet the needs of classification in our scenario. In addition to the results obtained from the data, a set of new rules to be applied onto the data are created. After above four steps, the optimization is then conducted. In our case, we especially focus on how the incomplete the data to be excluded at cleaning step, how the teacher nodes should be determined at feature selection step, how the rules to be set and refined at classification step, and how the performance of above-mentioned steps to be improved. Post-processing stage includes four main steps with different aspects and purposes on the collected data. The analysis on sentiment, emotion, behavior and preference are particularly concentrated. In the beginning, the interaction data, such as conversation, reservation details, user profiles, on Airbnb is retrieved. Although such data includes the text and multimedia contents, the text is considered the entry while DKDAR model is designed.

(3) Dkdar Engine is implemented by DKD engine and AR engine. The main purpose of DKD engine is to properly transfer data to decision. This engine attempts to support end users, especially users who are highly involved in the sharing economy scenarios, to take proper action(s) while a choice needs to be made. The main purpose of AR engine is to create rules to support the process of DKD engine. A bulk of rules may be applied through those obtained internal/external data to be integrated with the existing learning rules. Through the support of these two sub-engines under DKDAR engine, the sketch of user model can be drafted.

(4) Output is the last stage of the whole model. Its main purpose is to implement the user model by integrating additional factors. Our model generates tag cloud for each user. It is also a learning process as tags that present the characteristics of users may change over time. And as mentioned, temporal factor is essential while conducting the normalization to all the tags. One dynamic temporal model [14] is applied. Then, abnormal behaviors are checked, which is also a crucial part in this stage. Behavior patterns are categorized and stored in different groups, and irrelevant data, or noise, is also removed if necessary. This step may help identify the mostly-applied behaviors, common behaviors, and special behaviors while the sharing economy scenario is performed. With such results, the index to each category can be made

TABLE 1. Rule extraction through Association rule.

Rule ID	Condition	Prediction
1.	19/04/2013 < data first booking < 24/05/2014	affiliate_provider = google
2.	data first booking >24/05/2014 and country_destination = US	date_account_created > 06/08/2013
3.	date_first_booking > 24/05/2014 and affiliate_provider = direct and first_affiliate_tracked = untracked	date_account_created > 06/08/2013
4.	data first booking > 24/05/2014 and age <= 28	date_account_created > 06/08/2013
5.	19/04/2013 date_first_booking < 24/05/2014 and signup_flow 5.000 and affiliate_channel = sem-brand and signup_app = Web	affiliate_provider = google

TABLE 2. Rule extraction through Apriori algorithm.

Rule ID	Consequent	Antecedent	Support%	Confidence%
1.	Affiliate_channel = Direct	Affiliate_provider = Direct and First_browser = Chrome and First_ affiliate_tracked = Untracked and Country_destination = US	12.766	100
2.	Affiliate_provider = Direct	First_browser = Chrome and Affiliate_channel = Direct and Country_destination = US	18.807	99.807
3.	Signup_app = Web	Signup_method = Facebook and First_ affiliate_tracked = Untracked and Country_destination = US and Language = en	17.083	81.896
4.	Country_destination = US	First_device_type = Windows Desktop and Affiliate_provider = Direct and Signup_app = Web and Language = en	19.889	80.369
5.	First_device_type = Mac Desktop	First_browser = Safari and Affiliate_provider = Direct and Affiliate_channel = Direct and Signup_app = Web and Language = en	14.527	99.694

TABLE 3. Rule extraction through C5.0 algorithm.

Rule ID	Ruleset
1.	If age>36 and Signup_method = Facebook and Affiliate_provider = Direct and First_ affiliate_tracked = Untracked and First_browser = Safari and Country_destination = DE THEN 'Male'
2.	If age>28 and First_ affiliate_tracked = Untracked and Signup_app = Moweb and First_browser = Safari and Country_destination = FR THEN 'Male'
3.	If First_ affiliate_tracked = Linked and First_device_type = iPhone THEN 'Female'
4.	If age<= 30 and Signup_method = basic and Language = en and Signup_app = Web and First_device_type = Mac Desktop and First_browser = Safari THEN 'Female'
5.	If 25< age<= 56 and Signup_method = Facebook and Language = en and Affiliate_channel = Sem-brand and First_ affiliate_tracked = omg and First_device_type = Windows Desktop First_browser = Chrome THEN 'Male'

and open to support development of external services and/or platforms.

B. INSTANCE OF DATA PROCESS

Following the descriptions of proposed DKDAR model, this sub-section provides a concrete instance that demonstrates the process. Fig. 2 is a quick summary of scenario while a set of data is inputted to our model.

Table 1, 2, and 3 demonstrate three separate sets of rules to be applied based on Association rule, Apriori, and C5.0 methods. It is obvious that each of them comes up with rules different to each other though same input is given. For instance, it is observed that the factor 'gender' may be connected more to other factors as well as user's previous behavior patterns in C5.0 algorithm than in other two algorithms. In addition, the retrieved features are much clear, and meets the expectation from our study.

Comparing the performance with other existing works is not the main purpose in this study. Instead, this study targets to generate rules for human behavior mining and analysis on sharing economy platform. It is more like an attempt to verify whether behaviors on urban platforms can also be modeled through existing techniques, and to make precise prediction to support its users. This approach (simple rules) helps us to understand the existing relationship/relationships among features much more easily than using all the complex rules. These simple rules are being recorded in the proposed AR engine under DKDAR model, and may be adjusted based on follow-up given data from time to time.

IV. CROWD PREFERENCE SENSING AND DECISION-MAKING BEHAVIOR ANALYSIS

This section provides the analysis results based on the DKDAR model. Three subsections are included: first, source

TABLE 4. Airbnb new user booking dataset.

Feature name	Description
ID	User id
Date-account-created	The date of account creation
Timestamp-first-active	Timestamp of the first activity, note that it can be earlier than date-account-created or date-first-booking because a user can search before signing up.
Date-first-booking	Date of first booking
gender	Male, Female, Unknown and Other
age	All of ages
Signup-method	-
Signup-flow	The page a user came to signup up from
language	International language preference
Affiliate-channel	What kind of paid marketing?
Affiliate-provider	Where the marketing is e.g. google, craigslist, other?
First-affiliate-tracked	What's the first marketing the user interacted with before the signing up?
Signup-app	-
First-device-type	-
First-browser	-
Country-destination	-

of the dataset is discussed; second, obtained results on crowd preference sensing and decision-making behavior analysis are detailed; and third, a conclusion is discussed.

A. THE DATASET

Collecting correct and useful data for analysis is always the challenge for researchers. There are two ways to retrieve the data in general. One is to request the dataset from the open repository such as UCI Learning repository, and other similar websites. The other is to collect the data by researchers themselves from different sources. In our case, a crawler was designed to retrieve open data from Airbnb, one of the representative platform that implements the sharing economy scenario. The contexts related to users, such as conversation, transaction, profile, etc., were collected from October 2016. To avoid any illegal issue while conducting this research, we declare that our crawler did not collect any privacy data from the source as well as distribute users' personal information on any public channels. We collected a bulk of data that includes 213,451 records with 16 features as shown in Table 4, and these are the target to be analyzed and discussed.

B. RESULTS

This study targets the exploration of dynamic crowd preference and decision-making behavior analysis based on users' previous behavior/behaviors on Airbnb.

Users may have background experiences, such as attitude or preliminary information, about the target before getting into the decision-making process. The preliminary information may vary because it depends tightly on the service provider. Users can only try to achieve better understanding. But however, the user's attitude may be different. The attitude is formed mainly based on instinct, i.e., personality, and external contexts, i.e., interactions with others. Both of them may be performed in conscious or unconscious ways, and may implicitly cause influence to the attitude. Thus, the statistics discussed in this section are then summarized to reach a better understanding.

TABLE 5. The number of gender grouped by country-destination.

Country (Abbrev)	Male	Female
AU	188	**207**
CA	**477**	455
DE	**416**	358
ES	677	**853**
FR	1335	**1962**
GB	682	**881**
IT	699	**1091**
NDF	26719	**31048**
NL	**278**	254
PT	69	**78**
US	19457	**22694**
Other	**3443**	3160

One significance is that our dataset includes both internal factors (e.g., nationality, gender, and age) and external factors (e.g., device, social media, and time) retrieved from Airbnb. These factors are applicable to achieve understanding of crowd preference, and find out why specific decisions were, or are going to be, made. For example, we can categorize our users into different age groups (e.g., Group A, Group B, and etc.). After that, by using internal factors and external factors we can find out crowd preference towards specific country. Accordingly, we can understand that which country can be suggested for which age group. Thus, we have discussions based on the correlation among retrieved factors.

1) GENDER AND COUNTRY-DESTINATION

Finding the correlation among gender and country-destination is to obtain relative attractiveness between them. With the removal of the noise, 124,543 users with "NDF" tag and 10,094 users with "other" tag, number of users who made the reservation via Airbnb (for both male users and female users) and probably travel to the destination is illustrated in Table 5.

Table 5 reveals that female users are more than male users in 8 cases, while 4 cases (CA, DE, NL and Other) on the opposite. And among all these 8 countries in which female users are greater than male users, our records show that these female users are mainly from US, FR and the IT respectively.

2) GENDER AND SIGNUP-METHOD

The Airbnb currently allows users to sign up simply through SNS (i.e., Facebook, WeChat, Google, and Twitter) and Email. Sign-up method may reflect the internal personality of a specific user. The statistics can be found in Fig. 3.

The statistics have indicated that the basic sign-up method is still the most commonly-applied method than SNS-based one while accessing the Airbnb from both female and male users. Although it is found that the 'unknown' gender still possesses high number of users, we can count them directly since the gender does not affect the sign-up method. It should be noted that number of female users and male users are very similar in basic method and Facebook but 'unknown' users

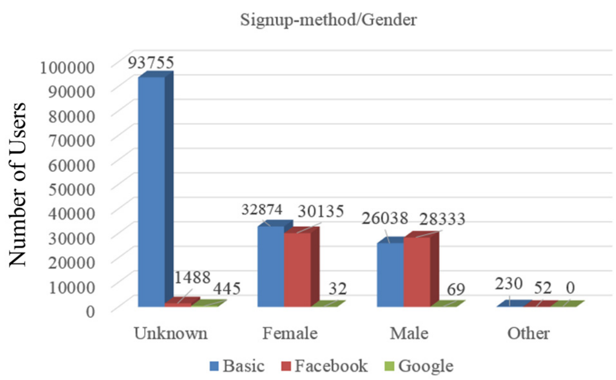

FIGURE 3. Distribution of signup-method grouped by gender.

TABLE 6. The number of users grouped by signup-app for each gender.

Signup-app	Unknown	Female	Male	Other
Android	**3425**	819	1208	2
Moweb	**3253**	1806	1199	3
Web	**78964**	56284	47208	261
iOS	**10046**	4132	4825	16

are mostly in basic method. Then, female users used SNS (i.e., Facebook) more than male users, and both female and male prefer to use Facebook than Google while accessing the Airbnb. We may come to a conclusion that the usability and/or connectivity of Facebook might be better than Google if the integrity to other external services is mentioned.

3) GENDER AND SIGNUP-APP

Following the results above, we go further to investigate the correlation between gender and signup-app. The signup-app indicates the approach while signing in/up. In general, Airbnb allows users to sign in/up via different mobile device or PC. We will then find out which of them is the most acceptable, or comfortable/suitable, for this purpose. The number of users grouped by signup-app is illustrated in Table 6.

Table 6 demonstrated that the Web (i.e., browser) is the most common-used channel while doing the sing in/up. The reason might be because only full functions are provided by Airbnb on the Web, and the mobile Apps on both iOS and Android only offer partial functions for the users.

4) GENDER AND FIRST-DEVICE-TYPE

First-device-type is another important factor. It may reveal the implicit behavior patterns as well as the decision-making process. Fig. 4 illustrates the statistics from our records. It is worth mentioning that 4 types, i.e., Mac Desktop, Windows Desktop, iPad and iPhone, are major while accessing the Airbnb online service. And the Mac-based system/device is considered much better than Windows-based one. In addition, female users are greater than male users in all 4 of these systems/devices. Distribution of first-device-type grouped by gender is presented in Fig. 4:

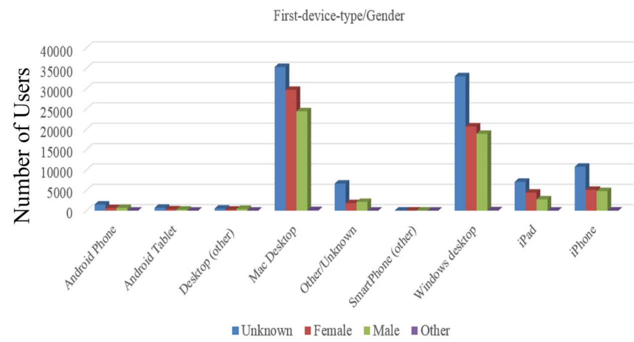

FIGURE 4. Distribution of first-device-type grouped by gender.

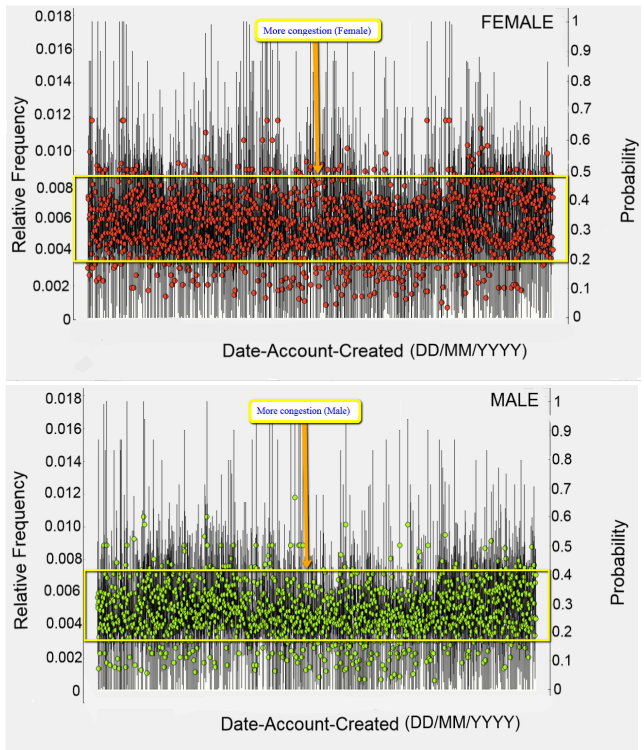

FIGURE 5. Distribution of 'date-account-created' grouped by 'gender' (relative frequencies) for both gender.

5) GENDER, DATE-ACCOUNT-CREATED AND DATE-FIRST-BOOKING

Another important factor is the time for decision-making process. We assume that most users on Airbnb created their accounts are for making the reservation. According to this assumption, we can go further to investigate the time difference between the account is created (date-account-created) and the reservation is made (date-first-booking) and its relation to users' genders. The correlations among gender, 'date-account-created' and 'date-first-booking' are illustrated in Fig. 5 and Fig. 6. The vertical axis has two sides that the left one is for 'Relative Frequency', the right one is for 'Probability' and the horizontal axis is for 'date-account-created' and 'date-first-booking'.

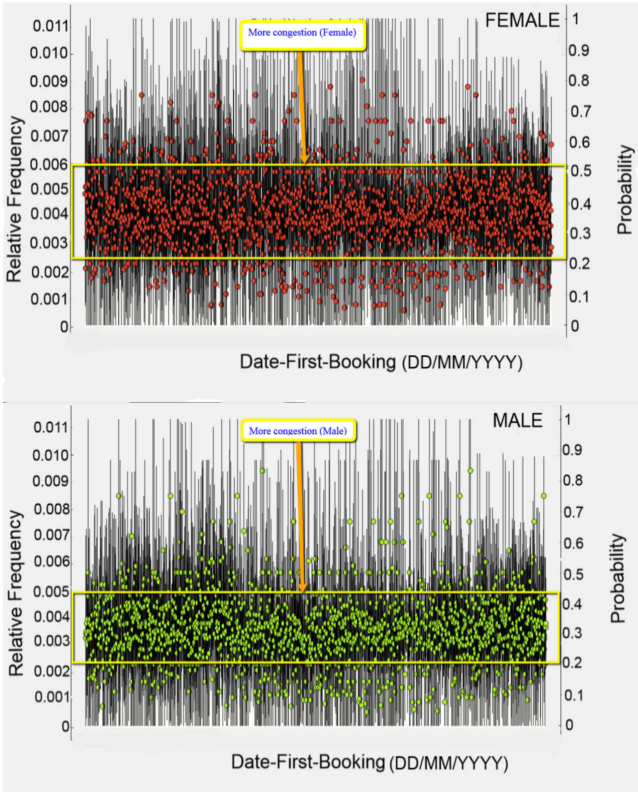

FIGURE 6. Distribution of 'date-first-booking' grouped by 'gender' (relative frequencies) for both gender.

According to the results, female users tend to ask more questions (about the properties and facilities) than male users (about the direction and location of the property) before making the reservations. Female users also tend to share their opinions via social media before making the decisions, and this is also reflected that most female users prefer to use Facebook and other social network services as their sign in/up methods. The main reason may because that it is easier to share with their connections. However, male users are totally on the opposite side.

6) GENDER AND AGE

The last, but not least, important factor is the gender in the dataset. But the statistics showed that users without registering their gender reach 40% up (around 95,668 users out of 21,453 users) than those who registered. This indeed cause difficulty in analysis but we attempted to focus on only the rest users with gender information. Statistics of gender on collected dataset is presented in Fig. 7:

According to Fig. 3 and Fig. 7, it is observed that female users may have higher density than male users. When we checked the dataset, we found that most of records do not have any value and some of them have illogical values. For instance, difference between the youngest user (1 year-old) and the eldest user (2014 years-old) reaches a huge gap that cannot be existed. In order to solve this issue, records without value or with incorrect values were removed from the dataset. We found that the density related to male user is higher than

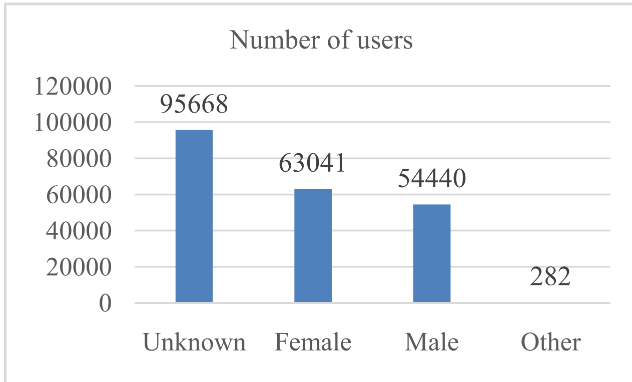

FIGURE 7. Statistics of gender on collected dataset (original dataset).

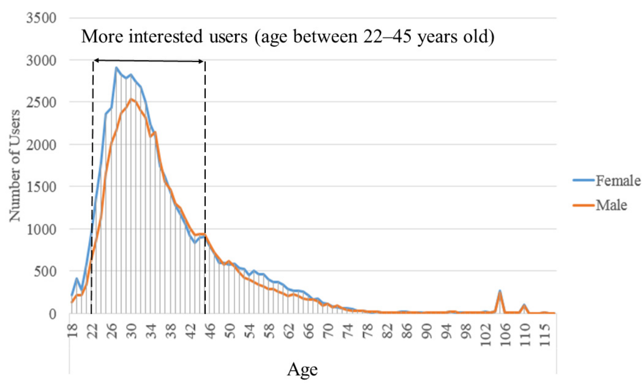

FIGURE 8. Relationship between gender and age for those active users.

female user although the number of female users is greater than male users.

To make precise analysis, users whose age under 18 and with unknown or other genders were removed from the dataset. Purpose of this step is to find out the activeness of different age group in sharing economy based on their gender. Also, we consider that users more than 18 years-old are more applicable to participate in sharing economy scenarios than those users under this age line (see Fig.8). The revised statistics show that users with age between 22-45 are more interested in sharing economy platform (at least Airbnb in this study).

C. DISCUSSION

This sub-section focuses on the analysis of gender factor and its relationship with country-destination, signup-method, signup-app, and first-device-type in Airbnb. According to the results presented in Fig. 8 above, we observed that even number of female users (57164 users) is greater than number of male users (50374 users), in some age groups male users are more active than female users as is illustrated in Fig. 9 below.

According to Fig. 9, we observed that male users with ages between 39-49 years old are more interested than female users in these ages. Another finding indicated that when users are older than 70 years old, both female and male users have almost similar behavior in terms of their activities in

TABLE 7. **Distribution of country-destination grouped by gender (relative frequencies) and probabilities.**

Country (Abbrev)	Normalized (relative) frequencies				Probabilities			
	Unknown	Female	Male	Other	Unknown	Female	Male	Other
AU	**0.004**	0.003	0.003	**0.004**	0.147±0.033	**0.451±0.047**	0.400±0.046	0.002±0.005
CA	0.012	0.007	0.009	**0.022**	0.182±0.023	0.395±0.029	**0.418±0.030**	0.005±0.004
DE	0.007	0.006	0.008	**0.013**	0.145±0.024	0.388±0.033	**0.464±0.034**	0.004±0.004
ES	0.016	0.014	0.012	**0.018**	0.158±0.017	**0.742±0.024**	0.367±0.023	0.002±0.002
FR	0.037	0.032	0.025	**0.058**	0.170±0.012	**0.488±0.016**	0.339±0.015	0.004±0.002
GB	**0.017**	0.014	0.013	0.013	0. 166±0.017	**0.465±0.023**	0.368±0.023	0.002±0.002
IT	0.021	0.018	0.013	**0.022**	0.178±0.017	**0.500±0.022**	0.320±0.020	0.002±0.002
NDF	0.284	0.479	**0.480**	0.290	0.084±0.002	**0.485±0.004**	0.429±0.004	0.001±0.000
NL	0.006	0.004	0.005	**0.013**	0.168±0.030	0.385±0.039	**0.442±0.040**	0.006±0.006
PT	0.001	0.001	0.001	**0.004**	0.146±0.055	**0.446±0.078**	0.401±0.077	0.006±0.012
US	**0.514**	0.371	0.366	0.455	0.178±0.003	**0.438±0.004**	0.381±0.004	0.002±0.000
Other	0.080	0.051	0.064	**0.085**	0.178±0.009	0.388±0.011	**0.431±0.011**	0.003±0.001

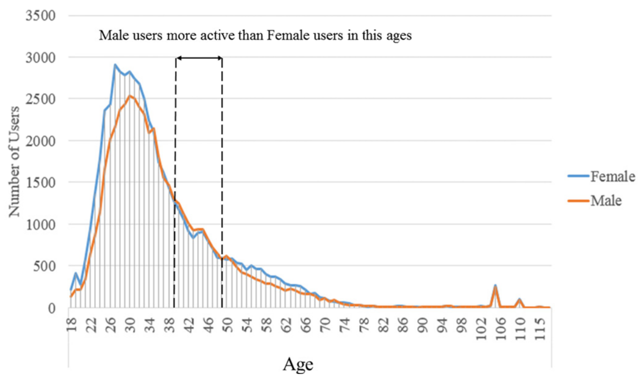

FIGURE 9. **The male users more active than female users.**

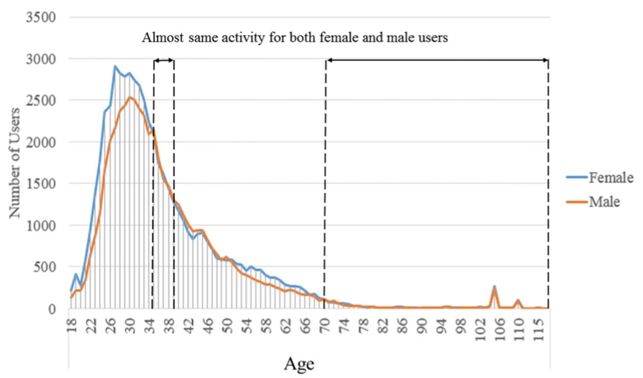

FIGURE 10. **Similar behavior in different age groups.**

Airbnb. But we also found that female and male ages between 35-39 years old had a similar behavior as is presented in Fig. 10.

Furthermore, the country-destination is taken as a target. Note that "normalized (relative) frequencies" calculates the scale of given data, for each factor, by percentage of the dataset. "Probability estimators" computes probabilities of values of class variable through (1) p(C=c) for unconditional probabilities, and (2) p(C=c|V=v) for conditional probabilities where notation c is a class and v is a feature value.

Distribution of country-destination grouped by gender (relative frequencies) and probabilities is presented in Table 7:

Table 7 indicates that the normalized (relative) frequencies related to male users in three countries including CA, DE and NL are more than the normalized (relative) frequencies related to female users in these countries. According to the related probabilities presented in Table 7, we can see that number of female users is greater than male users except CA, DE, and NL. Normalized (relative) frequencies indicate that "Other" has larger number than other factors. Distribution of signup-method grouped by gender (relative frequencies) and probabilities are indicated in Table 8.

Female users have higher interests (see related probabilities presented in Table 8) in using Facebook and Basic while male users have more in using Google on information collection. We believe that the trust level between genders should be different. According to this evidence, we would argue that female users were more sensitive due to the different type of their activities as well as the level of their trust in the decision-making process. Female users prefer to share their question(s) with their friends but male users may prefer search in some search engine such as Google. In order to get more information about accommodation, female users would check further details before decision-making processes. For this reason, they like use some social media like Facebook as we discussed earlier. In the next step, distribution of signup-app grouped by gender (relative frequencies) and probabilities is illustrated in Table 9.

According to the results presented in Tables 6 and 9, Android and iOS used by male users more than female users whereas Moweb and Web used female users (see both normalized (relative) frequencies as well as probabilities presented in Table 9). Furthermore, we concentrated on first-device-type and its relation/relations with gender. According to Table 10 and Normalized (relative) frequencies, we can observe that first-device-type as Android Phone, Desktop (Other), other/Unknown, Windows Desktop and iPhone is applied more by male users than female users, while number

TABLE 8. Distribution of signup-method grouped by gender (relative frequencies) and probabilities.

Sign-method	Normalized (relative) frequencies				Probabilities			
	Unknown	Female	Male	Other	Unknown	Female	Male	Other
Basic	**0.941**	0.490	0.452	0.768	0.236±0.003	**0.420±0.004**	0.341±0.004	0.003±0.000
Facebook	0.056	0.509	**0.547**	0.232	0.016±0.001	**0.505±0.004**	0.478±0.004	0.001±0.000
Google	**0.003**	0.001	0.001	0.000	0.333±0.078	0.206±0.067	**0.461±0.082**	-

TABLE 9. Distribution of signup-app grouped by gender (relative frequencies) and probabilities.

Sign-app	Normalized (relative) frequencies				Probabilities			
	Unknown	Female	Male	Other	Unknown	Female	Male	Other
Android	0.020	0.014	**0.023**	0.009	0.147±0.014	0.343±0.019	**0.509±0.020**	0.001±0.001
Moweb	**0.025**	0.021	0.015	0.013	0.180±0.015	**0.505±0.020**	0.314±0.019	0.001±0.001
Web	0.886	0.895	0.868	**0.906**	0.135±0.002	**0.465±0.003**	0.398±0.003	0.002±0.000
iOS	0.069	0.070	**0.094**	0.071	0.117±0.006	0.404±0.010	**0.477±0.010**	0.002±0.001

TABLE 10. Distribution of first-device-type grouped by gender (relative frequencies) and probabilities.

First-device-type	Normalized (relative) frequencies				Probabilities			
	Unknown	Female	Male	Other	Unknown	Female	Male	Other
Android Phone	0.011	0.009	0.011	**0.013**	0.152±0.020	0.403±0.027	**0.442±0.027**	0.002±0.003
Android Tablet	**0.008**	0.005	0.004	0.000	0.195±0.030	**0.458±0.038**	0.346±0.037	-
Desktop (Other)	0.007	0.004	0.008	**0.009**	0.146±0.025	0.317±0.033	**0.534±0.036**	0.003±0.004
Mac Desktop	0.456	0.477	0.455	**0.536**	0.132±0.003	**0.470±0.004**	0.396±0.004	0.002±0.000
Other/Unknown	0.028	0.028	**0.038**	0.013	0.117±0.010	0.399±0.015	**0.483±0.016**	0.001±0.001
Smart-Phone (Other)	**0.000**	**0.000**	**0.000**	**0.000**	0.077±0.084	**0.615±0.153**	0.308±0.145	**-**
Windows Desktop	0.342	0.327	**0.343**	0.308	0.137±0.003	**0.447±0.005**	0.414±0.005	0.002±0.000
iPad	0.063	**0.070**	0.051	0.036	0.138±0.008	**0.523±0.011**	0.338±0.011	0.001±0.001
iPhone	0.085	0.081	**0.089**	0.085	0.135±0.007	**0.438±0.009**	0.425±0.009	0.002±0.001

of female users are more than male users in the rest of cases. More information about distribution of first-device-type grouped by gender is given in Table 10:

Related probabilities presented in Table 10 indicated that male users have greater probabilities for Android Phone, Desktop (Other), and other/Unknown than female users. Whereas female users have higher probabilities than male users in rest of cases. A significant point about iPhone in Table 10 is that even though male users have higher normalized (relative) frequency than female users, but we observed that related probability for female users is greater than related probability for male users.

V. CONCLUSION

Success of sharing economy creates many opportunities for the researchers in all fields. From the viewpoint of computer science, the platforms, i.e., Airbnb, that implement the idea of sharing economy generated a huge number of data, especially user-generated data, that can be used to achieve better understanding of human beings. With this concern, this paper designs a universal user model that supports the dynamic crowd sensing on user preference, and analysis of correlated decision-making behavior. This study especially concentrates on the investigation of relationships between internal factors (e.g., nationality, gender, and age) and external factors (e.g., device, social media, and time). Results indicated that female users are more sensitive than male users on decision-making process according to the statistics of external features such as time (e.g., date-account-created and date-first-booking), social media (e.g., Facebook) and device (e.g., first-device-type, signup-method and signup-app). In addition, female users are more active than male users in sharing economy platform, at least on Airbnb, and willing to interact with other users.

Although the results have shown that our newly-proposed model is applicable, further actions to make improvement are still required. We would continuously work on the direction of human behavior analysis for the development of well-being in society, and increase the supports to other existing sharing economy services based on the proposed model in the future.

REFERENCES

[1] P. M. Falcone and E. Imbert, "Bringing a sharing economy approach into the food sector: The potential of food sharing for reducing food waste," in *Food Waste Reduction and Valorisation*. Cham, Switzerland: Springer, 2017, pp. 197–214.

[2] R. Belk, "Selfish sharing," in *NA—Advances in Consumer Research*, vol. 44, P. Moreau and S. Puntoni, Eds. Duluth, MN, USA: Association Consumer Research, 2016, pp. 37–41. [Online]. Available: http://acrwebsite.org/volumes/1021384/volumes/v44/NA-44

[3] A. Fagerstrøm, S. Pawar, V. Sigurdsson, G. R. Foxall, and M. Yani-de-Soriano, "That personal profile image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behavior on AirBnb," *Comput. Human Behavior*, vol. 72, pp. 123–131, Jul. 2017.

[4] A. B. Brendel, B. Brauer, and B. Hildebrandt, "Toward user-based relocation information systems in station-based one-way car sharing," in *Proc. 22nd Amer. Conf. Inf. Syst.*, San Diego, CA, USA, 2016, pp. 1–10.

[5] J. Hamari, M. Sjöklint, and A. Ukkonen, "The sharing economy: Why people participate in collaborative consumption," *J. Assoc. Inf. Sci. Technol.*, vol. 61, no. 9, pp. 2047–2059, 2016.

[6] C. Lutz, C. P. Hoffmann, E. Bucher, and C. Fieseler, "The role of privacy concerns in the sharing economy," *Inf., Commun. Soc.*, vol. 16, pp. 1–21, Jun. 2017.

[7] Y. Zheng, F. Liu, and H. Hsieh, "U-Air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2013, pp. 1436–1444.

[8] S. Dumais, R. Jeffries, D. M. Russell, D. Tang, and J. Teevan, "Understanding user behavior through log data and analysis," in *Ways of Knowing in HCI*. New York, NY, USA: Springer, 2014, pp. 349–372.

[9] J. Pesonen and K. Pasanen, "A closer look at tourist information search behaviour when travelling abroad: What is the role of online marketing in choice of destination?" in *Information and Communication Technologies in Tourism*. Cham, Switzerland: Springer, 2017, pp. 431–443.

[10] T. A. Weber, "Intermediation in a sharing economy: Insurance, moral hazard, and rent extraction," *J. Manag. Inf. Syst.*, vol. 31, no. 3, pp. 35–71, Jul. 2014.

[11] M. Möhlmann, "Collaborative consumption: Determinants of satisfaction and the likelihood of using a sharing economy option again," *J. Consum. Behaviour*, vol. 14, no. 3, pp. 193–207, May 2015.

[12] B. Fang, Q. Ye, and R. Law, "Effect of sharing economy on tourism industry employment," *Ann. Tourism Res.*. vol. 57, pp. 264–267, Mar. 2016.

[13] E. Ert, A. Fleischer, and N. Magen, "Trust and reputation in the sharing economy: The role of personal photos in AirBnb," *Tourism Manag.* vol. 55, pp. 62–73, Aug. 2016.

[14] Z. Xu *et al.*, "Building knowledge base of urban emergency events based on crowdsourcing of social media," *Concurrency Comput., Practice Exp.*, vol. 28, no. 15, p. 4038- 4052, 2016.

[15] J. Imgraben, A. Engelbrecht, and K.-K. R. Choo, "Always connected, but are smart mobile users getting more security savvy? A survey of smart mobile device users," *Behaviour Inf. Technol.*, vol. 33, no. 12, pp. 1347–1360, 2014.

[16] G. Quattrone, D. Proserpio, D. Quercia, L. Capra, and M. Musolesi, "Who benefits from the sharing economy of AirBnb?" in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 1385–1394.

[17] N. Yannopoulou, M. Moufahim, and X. Bian, "User-generated brands and social media: Couchsurfing and AirBnb," *Contemp. Manage. Res.*, vol. 9, no. 1, p. 85, 2013.

[18] D. Guttentag, "AirBnb: Disruptive innovation and the rise of an informal tourism accommodation sector," *Current Issues Tourism*, vol. 18, no. 12, pp. 1192–1217, 2015.

[19] G. Zervas, D. Proserpio, and J. Byers, "A first look at online reputation on AirBnb, where every stay is above average," Boston Univ., Boston, MA, USA, Working Paper, 2015, vol. 28. [Online]. Available: http://ssrn.com/abstract=2554500(11/12/2016)

[20] B. G. Edelman and M. Luca, "Digital discrimination: The case of AirBnb.com," Harvard Business School NOM Working Paper 14-054, 2014, pp. 14–54. [Online]. Available: http://www.hbs.edu/faculty/Publication%20Files/Airbnb_92dd6086-6e46-4eaf-9cea-60fe5ba3c596.pdf(10/12/2016)

[21] M. Abdar and N. Y. Yen, "Sharing economy and its effect on human behaviour changes in accommodation: A survey on AirBnb," *Int. J. Soc. Humanistic Comput.*, vol. 2, nos. 3–4, pp. 203–218, 2017.

MOLOUD ABDAR received the bachelor's degree in computer engineering from Damghan University, Iran, in 2015. He is currently pursuing the master's degree with the University of Aizu, Aizu, Japan. He has several papers about data mining, machine learning, and user modeling in some refereed international journals and conferences. His research interests include data mining, artificial intelligence, machine learning, social media, big data, and human-centered computing. He is a member of International Association of Engineers. He is also very active in three international conferences and some referred international journals, including the *Future Generation Computer Systems*, the IEEE Access, the *Neurocomputing*, and the *Journal of Internet Technology*, as a Reviewer.

NEIL Y. YEN received the Ph.D. degree in human sciences from Waseda University, Japan, and the degree in engineering from Tamkang University, New Taipei, Taiwan, in 2012. His Doctorate degree at Waseda University was funded by the Japan Society for the Promotion of Science under RONPAKU Program. He has been with the University of Aizu, Aizu, Japan, as an Associate Professor since 2012. He has been involved extensively in an interdisciplinary field of research, where the themes are in the scope of big data science, computational intelligence, and human-centered computing. He is a member of the IEEE Computer Society, the IEEE System, Man, and Cybernetics Society, and the Technical Committee of Awareness Computing. He has been actively involved in the research community by serving as a Guest Editor, an Associate Editor, and a Reviewer for international referred journals and as the Organizer/Chair of the ACM/IEEE-sponsored conferences, workshops, and special sessions.