Hindawi
Security and Communication Networks
Volume 2018, Article ID 2790373, 13 pages
https://doi.org/10.1155/2018/2790373

*Research Article*

# A Heuristic Model for Supporting Users' Decision-Making in Privacy Disclosure for Recommendation

**Hongchen Wu**,[1] **Huaxiang Zhang**,[1] **Lizhen Cui**,[2,3] **and Xinjun Wang**[3]

[1]*School of Information Science and Engineering, Shandong Normal University, Jinan, China*
[2]*Shandong Provincial Key Laboratory of Software Engineering, Shandong University, Jinan, China*
[3]*School of Computer Science and Technology, Shandong University, Jinan, China*

Correspondence should be addressed to Hongchen Wu; wuhongchen@sdnu.edu.cn

Received 17 August 2017; Accepted 13 November 2017; Published 4 February 2018

Academic Editor: Xuyun Zhang



Privacy issues have become a major concern in the web of resource sharing, and users often have difficulty managing their information disclosure in the context of high-quality experiences from social media and Internet of Things. Recent studies have shown that users' disclosure decisions may be influenced by heuristics from the crowds, leading to inconsistency in the disclosure volumes and reduction of the prediction accuracy. Therefore, an analysis of why this influence occurs and how to optimize the user experience is highly important. We propose a novel heuristic model that defines the data structures of items and participants in social media, utilizes a modified decision-tree classifier that can predict participants' disclosures, and puts forward a correlation analysis for detecting disclosure inconsistences. The heuristic model is applied to real-time dataset to evaluate the behavioral effects. Decision-tree classifier and correlation analysis indeed prove that some participants' behaviors in information disclosures became decreasingly correlated during item requesting. Participants can be "persuaded" to change their disclosure behaviors, and the users' answers to the mildly sensitive items tend to be more variable and less predictable. Using this approach, recommender systems in social media can thus know the users better and provide service with higher prediction accuracy.

## 1. Introduction

In the era of Big Data, the proliferation of social media and Internet of Things (IoT) has foreseen the interconnection of countless number of items and smart devices, uniquely identifiable everyday with the storage and communication of information about the online users [1]. Recommender systems are important tools to help individuals find relevant items by providing suggestions for contents to be of use [2], but they collect user preferences in ways that are analogous to statistical database queries and could raise privacy problems [3–5]. For example, recommender systems can help people find a vacant taxi [6], offer a particular user a set of venues [7], and support real-time decision-making through the provision of travel tour recommendations [8]. However, the information collected for recommendation could be used in conjunction with other data sources to uncover identities and reveal personal details about a particular user [9]. Furthermore, to give accurate recommendations to support users, it is necessary to be as familiar with users as possible so that the recommender system can understand what type of item they want to buy [10, 11], what movie they want to watch [12], or to what music they want to listen [13]. In general, the more information that recommender systems have about individuals, such as personal feelings, idea, and comments, the better they will be able to evaluate them. This issue has raised user privacy concerns [14, 15], as people may not want their habits or views to be known to the recommender systems.

Interestingly, however, users' actions sometimes indicate otherwise, and in the web of resource sharing, there remains an interesting paradox in the privacy problem and recommender system literature. An important benefit offered by the recommender system is the opportunity to personalize offerings to users according to others' feelings [16]. Theories of self-disclosure suggest that users' willingness to disclose personal information is based on their assessments of the costs and benefits [17]. Thus, website managers that interact with users use several approaches to alter this cost-benefit tradeoff

and encourage self-disclosure. Some managers increase the subjective benefits of self-disclosure by offering rewards (e.g., coupons or gifts) in exchange for personal information and personal feelings. Other managers reduce the subjective costs of self-disclosure by posting extensive privacy policies that claim to protect user privacy [18]. However, users are hesitant to disclose personal information due to worries that their personal information could be inappropriately collected, maintained, accessed, or used by online merchants without their consent [19]. Xia et al. had utilized a scheme which supports content-based image retrieval before storing sensitive image on the cloud server [20]. Fu et al. proposed an efficient multikeyword fuzzy ranked search scheme over encrypted data with practically efficient and high accuracy [21] and built a user interest model with a searchable encryption scheme supporting personalized search and privacy preserve in cloud computing [22]. A secure multikeyword ranked search scheme was presented by Xia et al. with a "Greedy Depth-first Search" algorithm, which can be applied on encrypted data before outsourcing for privacy purpose [23]. In [24], a content-based scheme was designed to solve the problems of semantic search based on conceptual graphs over the encrypted data. Fu et al. also proposed an applicable and extensible search scheme, which could support multikeyword ranked search in parallel computing with privacy-preserving approach [25]. Liu et al. presented an adaptive method aiming at spatial-temporal efficiency in a heterogeneous cloud environment on sensitive data [26]. Several recent experiments have shown that users' privacy decision-making often cannot fully explained by logic or statistical models in terms of the perceived benefits and risks. Rather, their disclosure behaviors may be affected by various heuristics, such as the disclosure of information that other users have already disclosed, the withholding of information when items are requested in an unexpected order, or the default settings of the privacy policy applied in social media websites.

As a result, privacy has become a hot topic enablers of the social media and IoT vision.. Despite considerable contributions from previous scholarly works, little address possible impacts on consistency of users' disclosure. This paper demonstrates that users may express different disclosure behaviors from their previous disclosures due to our proposed heuristics approach, and we detect this inconsistency phenomenon with correlation analysis and decision-tree classifier; if the consecutive requests are less correlated, the prediction accuracy of whether they will disclose the subsequent requested information item is lower as well. This is an important phenomenon to study, because unexpected information requests can reduce user satisfaction and raise privacy concerns. Correlation analysis was used to detect the strength of dependence among users' disclosures, and the inconsistency was reported if the decision-tree classifier further showed that the prediction accuracy of users' next disclosures was very low when the model was informed from their previous actions. The obtained results are important for the current debate in the recommender system and privacy literature.

The paper is organized as follows. In Section 2, the related work is provided. Section 3 proposes the materials and methods, including several hypotheses and the decision-tree methods and correlation analysis in the heuristic approach. The experiment and its manipulation are described in Section 4. We present our conclusions in Section 5.

## 2. Related Works

This paper makes its scholarly contributions on the basis of previous literature on recommender systems and privacy, and it addresses a number of theoretical and practical gaps that have not been covered. We first introduce the related works on users' disclosure behaviors in recommendation environments. Furthermore, we present the heuristic model for detecting behavioral inconsistency by discussing how request types and orders may influence people's disclosure tendencies with correlation analysis and decision-tree classifier and whether users' features moderate these differential effects. Several possible explanations as to why people change their disclosure behaviors, as well as the basic estimation elements that we employ in this paper, are also presented in the next section. The findings of this paper could benefit researchers and recommender system developers who focus on user-centric aspects and, particularly, who aim to predict or prime users' disclosure behaviors in the system-preferred direction, for example, towards disclosing more personal information.

Users' privacy preferences are context-dependent and multidimensional [27]. Many users experience difficulties in managing their privacy settings, and some even avoid the hassle of changing their privacy settings altogether [28]. Acquisti and Grossklags showed that it is unrealistic to expect user rationality in privacy decisions, which may explain why users who may genuinely want to protect their privacy might fall prey to psychological distortions that are well documented in the behavioral literature [29]. Privacy has become a major issue among users who browse the Internet and receive benefits in exchange for their personal information. A recommender system can be applied for mutual benefit: users receive good suggestions on what to buy, which suit their needs and save time; managers can, in turn, apply user feedback, which can be used to place the most popular products in a more prominent place. In general, there is a tradeoff between the benefits and the potential privacy risks and trustworthiness threats in the crowdsourcing environment. For example, in recent years, several companies have been caught invading users' privacy [30], which increases user concerns regarding privacy protection. Unfortunately, users also have difficulties in deciding what information to disclose when making decisions [31]. Gross and Acquisti [32] and Baden et al. [33], who address privacy and rationality in individual decision-making, suggest that users should be able to manage their privacy. Empirical and theoretical research suggests that users often lack sufficient information to make privacy-sensitive decisions. Even with sufficient information, users are likely to trade long-term privacy for short-term benefits. Users have different individual privacy disclosure tendencies, and there is no single strategy that works for all of them; in other words, each user has his/her own privacy tendencies on specific items. Recent experiments and surveys have illustrated an obvious dichotomy between privacy

attitudes and actual behavior: users were willing to trade privacy for convenience and release personal information in exchange for relatively small rewards, but users were seldom willing to adopt privacy protective technologies [34]. Kobsa proposed a strategy of privacy protection that balances personalization and minimal use of users' personal data. Knijnenburg proposed a new approach by analyzing users' satisfaction, disclosure tendencies, and other characteristics and used justifications to help users make better disclosure decisions and increase users' willingness to disclose demographic and contextual information [35]. In [36], they adapted this justification-based approach to users' genders and disclosure tendencies. The researchers demonstrated there is no one-size-fits-all strategy in user disclosure personalization. Recently, scholars have proposed privacy personalization or adaptation as a highly dynamic, user-tailored, and context-aware approach to privacy decision support at an individual level [37]. A decentralized belief propagation-based method, PD-LBP, was proposed for multiagent task allocation in open and dynamic grid and cloud environments [38]. Chen et al. proposed a coverless information hiding method using Chinese character technology with high efficiency [39]. Reference [40] proposed a self-adaptive artificial bee colony algorithm based on the global best candidate for global optimization to a real privacy clustering problem based on the $K$-means technique. Yuan et al. proposed a coverless image steganography scheme based on scale invariant feature of secret communication [41]. In the broadest sense, a "privacy adaptation procedure" entails predicting users' privacy behaviors through decision-tree classifier [42], fuzzy clustering [43], correlation analysis [44, 45], or modified dimensionality reduction algorithm [46]. Our study is designed to give users privacy decision support and provide user-adaptive design, and we believe that good user privacy decision support can maintain high user satisfaction with the recommender system, gain trust regarding the items suggested, and lower the awareness of threats. Users in these conditions would probably disclose more personal information and thus help the recommender system get to know them better and create more accurate service. This study could be very useful for any online systems that apply user-adaptive interactions and user-centric-based strategies.

Many recent experiments have shown that users' privacy decision-making often cannot be well explained by assuming that a tradeoff between perceived benefits and risks is made in an entirely rational manner. Rather, their disclosure intentions and actual disclosures of personal information are impacted by various heuristics, which mainly include two aspects. First, research from a variety of literature has suggested that users' privacy decision-making can be affected by others' willingness, which is called the herding effect. In its most general form, herding can be defined as behavioral patterns that are correlated across users that share opinions towards the same items. In this circumstance, some users would probably release the item of information if their close friends do. In the crowdsourcing literature, personalization providers and website managers use personal information to tailor messages to individual users, but most users can elicit reactance when information techniques are deemed too intrusive, which counters the benefits while amplifying the cons of these tools. Users were more likely to disclose their information when they were warmed up with introductory questions, relative to other users that had answered the same questions without first answering the introductory questions. Applied to self-disclosure, research about herding suggests that if a large group of users are revealing several kinds of information, probably there is not great risk in doing so oneself. These findings imply that if users are surrounded by others who are revealing intimate details of their lives, they more easily conform to the prevailing norms of divulgence. Observing other users' willingness to disclose intrusive information, especially to admit to sensitive behaviors, may lead the users to be less concerned about request disapproval, which, in turn, increases disclosure. As a result, users' decision-making and judgment are inherently comparative in the herding effect, and sometimes they follow others' decisions without conscious awareness. Second, the order of sensitivity or type in which items are being asked influences users' privacy concern; for example, users from whom items of low sensitivity are requested first disclose more than users from whom highly sensitive items are requested first. On social media websites that are equipped with recommender systems, the so-called ordering effect could potentially be universal. It was found in the literature that based on the sequence of question intrusiveness, the disclosure is indeed affected by comparisons to previous disclosures. A user's likelihood of answering an intrusive question is affected by the contrast between the current and the previous questions in the information-requesting sequences [47]. The admission rates in a sensitivity-decreasing survey could be lower than those of a randomly ordered survey, although both are requesting same content from users [48]. From the perspective of personalization providers and website managers, a well-designed order may help them collect more useful knowledge of their users without raising too much privacy concern and thereby predict users' favorite products with higher accuracy. According to the above literature, we shall argue that users' disclosure behaviors are indeed affected by a variety of heuristics. It is very important to predict and make sense of users' reactions to website information-requesting strategies and determine how they respond to the continual stream of requests for personal information, which is an unavoidable feature of the personalization providers for the next-generation web.

However, each user has his/her favorable request-behavior information and withholds private information at an individual level, which indicates users may not obey the heuristics. Research on this is still rare and, currently, most websites still apply a one-size-fits-all information-requesting strategy for anyone who signs into the account. This, to some extent, may help the website owner more easily set the strategy of privacy collection, but some users may feel invaded by obligatorily signing the agreement without any alternative option. Our previous study had showed that users' behaviors towards personal information request could change due to unknown reason [49]. We believe that the level of users' knowledge to support their decision-making will determine how much they will be affected by the heuristics, which

further affect their resource-sharing willingness as well as the level of privacy concern and overall satisfaction. In the following, we propose the aspects that may impact users' disclosure decision-making in the form of several hypotheses and further set up our heuristic model.

## 3. Materials and Methods

*3.1. Hypothesis Development.* Based on current literature of privacy and heuristic shown above, we are forwarding several possible hypotheses and explanation under our heuristic model. As shown in Figure 1, the heuristic model introduces a workflow with several methods for detecting users' behavioral inconsistency in three aspects: request type and sensitivity, others' disclosure willingness, and crowdsourcing experience. Disclosure changes in users express changes in volume and variability, represented in hypotheses H1 and H2. The methods for detecting the inconsistencies are correlation analysis and decision-tree classifier, which are used to detect low correlation among users' answers and low prediction accuracy, demonstrated by hypotheses H3 to H6, respectively. Finally, the inconsistency and errors are visualized by determining the difference between younger users and older users, as well as between professional users and nonprofessional users.

The requested items vary in two categories: context and demographic, where the context category contains requests that are mainly related to a person's online experience, for example, "Would you mind letting us know how much time you spend using your phone browser?" and "Would you mind letting us know what is your homepage in your browser?", while the demographic category includes requests that are close to participants' daily lives, for example, "How many cars do you own?" and "Would you mind letting us know your kids' names?" Furthermore, some requests are sensitive (most participants would prefer not to disclose that information to us), whereas others are not sensitive (most participants would disclose that information to us more readily). If the participants agree to share an item, they must provide real answers for our confirmation. The requested items consist of 12 context items and 19 demographic items, which were used in a previous study [50]. We invited online users, including 50 students (27 males and 23 females, all below the age of 30) and 50 faculty members (24 males and 26 females, all over age 30) to volunteer for a pilot study in which all 31 items were classified into different degrees of sensitivity: high-sensitivity items (admission rate < 60%), low-sensitivity items (60% < admission rate < 80%), and nonsensitive items (admission rate > 80%). Specifically, the hypothesized effects on the type and sensitivity of requested items are listed as follows.

*(H1) Requesting Demographic Items First Will Increase the Users' Subsequent Disclosures versus Requesting Context Items First.* Users are more likely to share their demographic information than their online information because most of the users have difficulties managing their online profiles and can hardly evaluate the potential risks and benefits, but they may be more familiar with information about themselves. This hypothesis will be supported if the disclosure rates of subsequent items are higher when the demographic items are requested first than when the context items are requested first.

*(H2) Requesting Sensitive Items First Will Lower Users' Subsequent Disclosures.* Acquisti proposed that divulgence is affected by the order in which inquiries of varying intrusiveness are made and suggested that divulgence is anchored by the initial questions: users are particularly likely to divulge when questions are presented in decreasing order of intrusiveness and less likely when questions are presented in increasing order. We believe that requesting high-sensitivity items would raise users' privacy awareness and may even upset them, which would further lower their system satisfaction and reduce the amount of later-requested information that they are willing to disclose. We will reject this hypothesis if the results do not indicate that the disclosure rates of subsequent items are higher when the nonsensitive items are requested first than when the sensitive items are requested first.

Each user should have his/her own disclosure pattern, and this pattern will vary because of different individual intentions based on a risk-benefit analysis. Some users may disclose everything, while others may withhold everything; some users may always disclose information related to their work, but never share family information; some may like to connect their Facebook accounts with game apps to obtain a game bonus. Our previous study has indicated that users demonstrate fewer disclosures and less predictability in sharing behaviors when they lack knowledge for supporting the decisions. This paper extends this argument one step further by supposing that some users may change their privacy disclosure patterns in some heuristics, for example, after several personal information requests, and users may disclose more personal information.

The Spearman's rho test is a rank-based nonparametric statistical significance test that can be used to detect monotonic trends in a time series. It is a numerical measure of the strength of the connection between two random variables, and its value ranges from negative values (one value's increase or decrease is related to another value's decrease or increase) to positive values (one value's increase or decrease is related to another value's increase or decrease). A correlation coefficient is suggested as a technique for summarizing and objectively evaluating and assessing the goodness of fit of a hypothesized distribution and can be applied to assess the significance of trend in practice and calculate the strength of association between two random variances or the load for ranking data, such as analyzing the degree of dependence among items that were answered by participants in our experiment. Our data are binary ("0" represents "NO" and "1" represents "YES") and are very suitable for computing the Spearman's rho. For a sample of size $n$, the $n$ raw scores $X_i$ and $Y_i$ are converted to ranks $x_i$ and $y_i$, and the correlation of the two variances $\rho$ is computed from

$$\rho = 1 - \frac{6 \sum d_i^2}{n\left(n^2 - 1\right)}, \tag{1}$$

where $d_i = x_i - y_i$ is the difference between ranks. We hypothesize that age plays an important role in participants' decisions

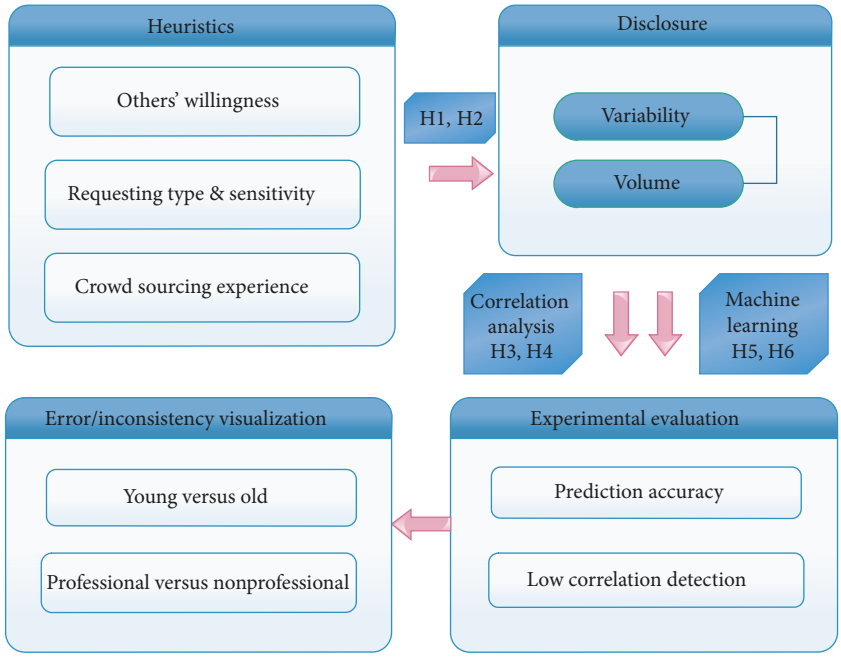


Figure 1: Hypothesized effects on inconsistency in users' behavioral disclosure by the heuristic model.

on disclosure behaviors, and we will mainly compare the results between participants who are below 30 and over 30. $x_i$ is the rank of requested item $i$ in ascending order of "No" answers from the younger participants, and $y_i$ is the rank of requested item $i$ in ascending order of "No" answers from the older participants. We can hypothesize the following.

*(H3) The Spearman's Rho among the Requested Items Answered by Experienced Participants Should Be Less Variable Than That of the Requested Items Answered by Less-Experienced Participants.* The older participants are expected to be more experienced people and are not expected to change their disclosure decisions because they know some disclosures would result in serious consequences; thus, their disclosures are more conservative. In contrast, the younger participants tend to disclose more information and change their minds on disclosures more easily because of their lack of social experiences. From the perspective of data quantity, younger participants have less valuable information compared with older participants. For one request such as "Could you tell us your house location?" a freshman would possibly disclose his apartment number at a university campus, while an older woman may not provide that information because her 5-year-old granddaughter may still live there. This hypothesis will be supported if we discover that the values of the Spearman's rho among older participants are more stable than those among younger participants.

*(H4) The Spearman's Rho Values between Two Closely Related Items and between Two Remote Items among Experienced Participants Should Be Less Variable Than among Less-Experienced Participants.* Although the 31 items are presented in random order, we still posit that the closer ranking of items was mostly presented to participants that are closer to each other versus other remote items according to our data. We hypothesize that older participants will rarely change their mind on information disclosure, so no matter the order in which the items are presented to older participants, the values of the Spearman's rho between items should not change enormously. For example, a mother who has given birth to four children would not disclose this information on her Facebook account, regardless of whether she had already been asked to disclose some nonsensitive information. However, a college freshman may connect his Facebook account with an online game account when he knows that a free bonus game would be provided. This hypothesis will be supported if we discover that the value of the Spearman's rho among older participants is less variable than among younger participants, even if the correlation was calculated between remote items or close items.

Decision tree has the powerful and effective ability to find a logical connection between the predicted item and the previous items by learning users' previous disclosures. In detail, each very last item was predicted by training a model on participants' answers to the previous items and then tested on their answers to the final item. The knowledge learned from this strategy can be represented as a tree model that contains the logical connections among their responses to requests. The precision is the rate of correctly classified samples, which is calculated as

$$\text{precision} = \frac{(\text{number of users who were accurately predicted})}{(\text{total number of users})}. \quad (2)$$

We run our decision-tree algorithm on the 31 disclosures of each individual item in the old dataset, which indicated

that the request order has an effect on the prediction accuracy. The request order also influences the variability of users' disclosure behaviors and further affects the prediction accuracy. We will also confirm participants' inconsistent disclosure behaviors when the correlation of answers to two requested items is low for the decision-tree analysis, and we hypothesize the following.

*(H5) The Request Type Will Influence the Prediction of Participants' Next Disclosures*. Users should be more likely to disclose demographic items than context items, so their disclosure behaviors should be less variable and thus more stable and predictable for the demographic items. This hypothesis will be supported if the prediction accuracy of the users' next disclosures is higher when the demographic items are requested first.

*(H6) Requesting Sensitive Items Will Affect the Prediction Accuracy of the Users' Next Disclosures*. Requesting high-sensitivity items will raise users' privacy concerns and thereby influence the knowledge learned from their previous behaviors, which may not be applicable to predicting their next disclosures. We will not reject this hypothesis if the prediction accuracy of the participants' next disclosure is higher when nonsensitive items are requested first.

*3.2. Heuristic Model.* As our heuristic model and its analysis rely on users answering a sequence of personal information requests, it is necessary to show the basic data structure used in the heuristic model.

*Requests(char RequestedItem1[Coin], char RequestedItem2[Coin], . . ., char RequestedItemN[Coin]). RequestedItemN* mainly consists of two categories: context and demographic, and there is an indicator that shows whether this request's sensitivity is high or low, which we call *Coin*. *Coin* = 1 indicates a highly sensitive context request or demographic request, which means most users would prefer not to disclose that information to us; otherwise, it is a low-sensitivity request, where most users would probably disclose that personal information to us more easily.

*Participant (bool Answer1, bool Answer2,. . ., bool AnswerN).* This data structure for participants only exists after they have finished answering the request sheet. The answer is marked "Yes" if this participant agrees to share the information; otherwise, it is marked "No." For easy recording, *Answerx* ($x = 1, 2, \ldots, n$) is denoted as "0" ("No, I would rather not disclose my information in response to this request") or "1" ("Yes, I will disclose my information in response to this request"). Additionally, the number of answers that each participant provided has to be the same number as the requests on the survey sheet that he/she had viewed; otherwise, we remove his/her answers due to incompleteness and disqualify his/her eligibility for payment. Furthermore, the total number of "0"s and "1"s from all participants in response to a request will determine its *Coin* value, which will be discussed at the end of Section 4.1.

*3.3. Prediction with Decision-Tree Analysis.* The experiment implements an orthogonal 2 × 3 × 2 × 2 between-subject design. Each participant is randomly assigned to one of the 24 conditions. Decision-tree analysis is used to predict users' next disclosure behaviors by learning their previous disclosures, due to its powerful and effective capability in finding logical connections between the predicted item and previous items. For example, suppose we predict request d530 based on previous requests d$xyx$ ($xyz$ range from 100 to 520 and d$xyx$ are all (0,1) values) by a decision-tree classifier, such as decision-tree algorithm, and we conduct a testing set confirmation with a supplied test set of the very last request with default parameters. The implementation of indicator prediction with decision-tree analysis is shown in Algorithm 1.

*3.4. Correlation Coefficient in Detecting Disclosure Changes.* The main study requires 12 items in each category *(Cs*, *Cn*, *Ds*, *Dn)*. These were selected from a set of 96 candidate items that were developed in a collaborative effort by the researchers and their colleagues. In an assay or instrument validation process, the reproducibility of the measurement from trial to trial is of interest. Unlike other validation processes, which are evaluated by using the Pearson correlation coefficient, the correlation coefficient of the paired $t$-test is a numerical measure of the strength of the connection between two random variables, and its value ranges from negative values (one value's increase or decrease instantly leads to another value's decrease or increase) to positive values (one value's increase or decrease instantly leads to another value's increase or decrease). The main drawback of the paired $t$-test is that it is more likely to reject a valuable conjuncture or a highly reproducible assay. The correlation coefficient is suggested as a technique for summarizing and objectively evaluating the information contained in probability plots and assessing the goodness of fit of a hypothesized distribution, and it will be very useful for ranking data, such as analyzing the degree of linear dependence among requests that were answered by participants in our experiment.

The dataset of participants' answer vectors is represented as $V$ (*answer*1, *answer*2, *answer*3,. . .,*answern*), where $n$ is the number of requests that we made to participants. For easier calculation, we use $ax$ to represent each *answerx* ($x = 1, 2, 3, \ldots, n$). Correlation coefficient matrix (CCM) is

$$\text{CCM} = \begin{bmatrix} C_{(1,2)} & C_{(1,3)} & C_{(1,4)} & & C_{(1,n-1)} & C_{(1,n)} \\ - & C_{(2,3)} & C_{(2,4)} & \cdots & C_{(2,n-1)} & C_{(2,n)} \\ - & - & C_{(3,4)} & & C_{(3,n-1)} & C_{(3,n)} \\ & \vdots & & \ddots & \vdots & \\ & - \quad - & & \cdots & C_{(n-2,n-1)} & C_{(n-2,n)} \\ & - \quad - & & & - & C_{(n-1,n)} \end{bmatrix}, \quad (3)$$

where $C(x_i, y_j)$ stands for the correlation coefficient between participants' answers to request the number $x_i$ and answers to the request number $y_j$. The implementation of correlation coefficient in our experiment is given in Algorithm 2.

**Input**: ParticipantSensitiveness data from *G*00, *G*01, *G*10, *G*11
**Output**: Build indicators' record $T$ with decision tree model
(1) **for** each group G*xy* **do** // classify each user group
(2) train dataset *Answers*[] by 10-folds-validation
(3) to predict indicators (*Ic*, *Id*) //train the data by turns
(4) Create a point Root //start to build the tree
(5) **if** all the participants in G*xy* belong to one class $C$
(6) **return** Root as a single leaf, label as $C$ //all users are same
(7) **Else** find best splits of subclasses $C\{c1, c2, \ldots, cn\}$
(8) with highest prediction accuracy $P$ //find the best splits
(9) **repeat** find further splits for each subclass $cx$
(10) **if** find higher prediction accuracy $P' > P$
(11) **return** to step (9) //for more specified splits
(12) **else** return the value record of (*Ic*, *Id*)
(13) $T = T + (Ic, Id)$ // add leaf to the tree model
(14) **if** *Answers*[] $\neq \emptyset$ //add all leaves to the tree
(15) *Answers*[] = *Answers*[] – *RequestedItemX*
(16) **where** *RequestedItemX* is the last column of *Answers*[]
(17) **return** to step (2) // the tree model is built
(18) **until** *Answers*[] = $\emptyset$

Algorithm 1: Indicator prediction with decision-tree analysis.

**Input**: Dataset of participants' answer vectors
**Output**: Correlation coefficient matrix (CCM)
(1) **for** all vectors $(a_{1i}, a_{2i}, a_{3i}, \ldots, a_{mi})$, where $i = 1, 2, \ldots, n$
(2) $m \leftarrow$ number of participants // store the population
(3) **if** some values of $a_i$ are missing // make sure no values miss
(4) $C_{ai} = N \times variance(a_i)$ // calculate by approximating
(5) $C_{ai} = \sum(a_i - \overline{a})^2 = \sum a^2 - 2n\overline{a}^2 + n\overline{a}^2 = \sum a^2 - n\overline{a}^2$
(6) $\{C\} \leftarrow$ all values of $C_{ai}$ and plot // ready for coefficient calculation
(7) **for** each two points $(i, C_{a,i})$ and $(j, C_{a,j})$:
(8) calculate the tilt angle $TA_{i,j} = |C_{aj} - C_{ai}|/|j - i|$
(9) Set$\{\widetilde{TA}\} = \emptyset$ // results for low correlation
(10) Set $\{\widehat{TA}\} = \emptyset$ // results for high correlation
(11) **for** all the values $TA_{i,j}$// all results should be calculated
(12) **if** $TA_{i-1,j} < TA_{i,j}$
(13) $\{\widetilde{TA}\} \leftarrow TA_{i,j}$ // adding low correlation results
(14) **else if** $TA_{i-1,j} > TA_{i,j}$
(15) $\{\widehat{TA}\} \leftarrow TA_{i,j}$ adding high correlation results

Algorithm 2: Calculation of correlation coefficient.

Dataset of participants' answer vectors = $V$ (*anwer*1, *answer*2, *answer*3,. . ., *answern*), where $n$ is the number of requests that we posted to participants. For easier calculation, we use $ax$ to represent each *answerx* ($x = 1, 2, 3, \ldots, n$). All values retained in $\{\widetilde{TA}\}$ are candidate changes that may indicate participants would like to disclose less information compared with the moment before seeing the request number $i$; in contrast, the values stored in $\{\widehat{TA}\}$ represent the candidate changes in which participants tend to disclose more information before finishing the request number $i$. Now, with the data structure and algorithm ready, we can perform our experiment.

## 4. Experiment and Discussion

*4.1. Data Preparation.* The experiment recruited 860 participants with unique IP addresses from a crowdsourcing Internet marketplace that enables workers and requestors to coordinate the use of human intelligence to perform tasks that computers are currently unable to perform. A total of 774 of the participants were qualified for further analysis, and the rest did not pass the cheating test. This dataset contains only the answers given by Chinese citizens from age 18 to 65, and the requested items were comprehensive, including 30 information disclosure decisions on 11 pieces of context data (e.g., homepage on the phone, online purchasing history,

etc.) and 19 pieces of demographic data (e.g., homing route, favorite music, etc.). In this dataset, the requested items are the most commonly requested items in daily life and can also be classified as sensitive, mild, and nonsensitive items, which are evenly distributed in descending order of sensitivity.

Three counters, $\mathrm{Fa}_{\mathrm{SP}}^{X}$,$\mathrm{Fr}_{\mathrm{SP}}^{X}$, and $\mathrm{St}_{\mathrm{SP}}^{X}$, were initially set to 0. If one user agrees to share an item $x$ with his family members, $x$'s family sharing point counter $\mathrm{Fa}_{\mathrm{SP}}^{X}$ will be increased by 1. If one user agrees to share an item $x$ with his friends, $x$'s friends sharing point counter $\mathrm{Fr}_{\mathrm{SP}}^{X}$ will be increased by 1. If one user agrees to share an item $x$ with strangers, $x$'s stranger sharing point counter $\mathrm{St}_{\mathrm{SP}}^{X}$ will be increased by 1. The values of these three counters will determine the general value $\mathrm{Item}_{\mathrm{SP}}^{X}$ for item $x$. One of our previous studies confirmed that participants' willingness to share an item is highest when the recipient is a family member and decreases for friends and strangers, in this order. We invited an additional 300 online users to answer the 30 requested items, and the values of the three counters were $451:220:63 \approx 6:3:1$. We determine the sensitivity of an item by adding all of its sharing points for all item recipients:

$$\mathrm{Item}_{\mathrm{SP}}^{X} = \mathrm{Fa}_{\mathrm{SP}}^{X} \times 0.1 + \mathrm{Fr}_{\mathrm{SP}}^{X} \times 0.3 + \mathrm{St}_{\mathrm{SP}}^{X} \times 0.6. \quad (4)$$

Then, the items were ranked in ascending order; the top 10 items were regarded as nonsensitive items $\{N01, N02, \ldots, N10\}$; the bottom 10 items were regarded as sensitive items $\{S21, S22, \ldots, S30\}$; the remaining 10 items were regarded as mild items $\{M11, M12, \ldots, M20\}$. $\{M11, M12, \ldots, M17\}$, $\{S21, S22, \ldots, S26\}$, and $\{N01, N02, \ldots, N06\}$ are all demographic items, and the remaining items are all context items. When a participant is asked to disclose his information about the item, he can either deny the item request or share the information and reply with a *[REASON]*; we will check the authenticity later (the fake answers will be disqualified).

*4.2. Main Experiment.* In the main experimental section, we conduct an experiment to determine whether heuristics may lead to the inconsistency of participants' sharing behaviors. A total of 774 participants were randomly assigned to two situations (approximately half and half), where one is equipped with heuristics and the other is performed without the heuristics approach. In the heuristic-provided situation, each participant was given a chance to modify the sharing decision after either of the following heuristics was provided after the participant had denied an item request:

(1) If you disclose the information, your browsing experience and recommender quality could be optimized by [$XX$] percent.

(2) [$YY$] percent of the participants accept the item request, because . . .*[REASON]*.

Participants were randomly assigned to the following conditions, where the orders in terms of sensitivity and item type were different, and binary indicators $S = \{0, 1\}$, $T = \{0, 1\}$, and $H = \{0, 1\}$ were set to show which condition the participant had been assigned to. When $S = 0$ ($S = 1$), the participants were required to respond to the items in sensitivity-ascending (sensitivity-descending) order. When $T = 0$ ($T = 1$), the requested items were presented to the participants with the context (demographic) items requested first. When $H = 1$ ($H = 0$), the heuristics were (not) provided after a participant had denied a request. For example, the participants who had been assigned to the condition $\{0, 0, 1\}$ would be presented with the items in sensitivity-increasing and context-first order, for example, $\{N07{\sim}N10\}$, $\{N01{\sim}N06\}$, $\{M18{\sim}M20\}$, $\{M11{\sim}M17\}$, $\{S27{\sim}S30\}$, and $\{S21{\sim}S26\}$, with the support of the heuristics.

This experiment could help us find the most variable conditions that each participant may face and could be used to analyze their information disclosure and discover common knowledge. Items from the same category (such as $N08$ and $N10$; both are nonsensitive context items) were presented closer to each other than items from different categories (such as $N01$ and $N08$; context versus demographic). If we could use the Spearman's rho to determine that the strength of correlation among the items from the same category is higher than the correlation among the items from different categories, we could possibly say that participants' disclosure behaviors are less consistent with their previous disclosure behaviors. Furthermore, if the inconsistency only occurs in the heuristics-provided situation, we shall conclude that the heuristics have successfully "persuaded" the participants to disclose more personal information.

We separate the data of participants' answers according to $2 \times 2 \times 2 = 8$ different conditions with varying values of $\{S, T, H\}$ and collect the number of denials for the 30 requests as an 8-dimensional vector *ItemID* {|condition{0, 0, 0}|, |condition{0, 0, 1}|,. . ., |condition{1, 1, 1}|}, where |condition{$S$, $T$, $H$}| represents the number of denials this item has received from the participants in Condition{$S, T, H$}. The Spearman's rho was calculated among the items by loading the vectors, and we use it to detect any weakening inconsistencies between users' past behaviors and following behaviors.

The sign of the Spearman correlation coefficient indicates the direction of association between two measured items, $x$ (observed item) and $y$ (testing item); the Spearman correlation coefficient is positive if the number of denial answers for $x$ tends to increase when the number of "No" answers for $y$ also increases, and the coefficient is negative if the number of "No" answers for $x$ tends to increase as the number of "No" answers for $y$ decreases. A Spearman's correlation coefficient of zero indicates that there is no tendency. The closer $x$ and $y$ are to being monotone functions of each other, the closer the Spearman's correlation coefficient is 1 or −1. We categorize the strength of the correlation between those items according to the following standards.

*(1) $\rho > 0.6$ or $\rho < -0.6$.* In this situation, the strengths of the two items are very highly correlated, and the answers to the two items are highly consistent. The closer the absolute value of the Spearman's rho is to one, the more the number of "No" answers exceeds the number of "Yes" answers, or the more the number of "Yes" answers exceeds the number of "No" answers. In these circumstances, each participant was expected to strictly maintain his or her privacy calculus, such as never disclosing very sensitive items or disclosing everything. Thus, there will be more rejections if the testing item is more sensitive than the observed item. We call those

participants who keep only one privacy calculus or herding effect as consistent as possible.

*(2) $0.3 < \rho < 0.6$ or $-0.6 < \rho < -0.3$.* In this situation, the correlation of the answers to the two items is not as strong as in the previous situation. Although the absolute value of the Spearman's rho becomes lower, the number of participants who maintain their privacy calculus still exceeds the number of participants who change their disclosure pattern (e.g., someone who rejects the disclosure request from an item that has a denial rate of 45%, but agrees to share his information for an item that is rejected by 70% of the participants). These inconsistent phenomena probably indicate that several participants changed their privacy calculus.

*(3) $-0.3 < \rho < 0.3$.* In this situation, it is very difficult to determine participants' disclosure tendencies, and the strength of the correlation between the observed item and the testing item is considerably lower compared to the other situations. Thus, we cannot conclude that most of the participants hold only one privacy calculus, especially when the Spearman's rho is close to zero. If we look at the data, the number of inconsistent phenomena should exceed the number of consistent phenomena when the absolute value of the Spearman's correlation coefficient is close to 0.3 or less. We will determine these inconsistent phenomena and determine what may lead participants to disclose more information.

*4.3. Results and Discussion.* We first look at the denial rates of the 30 requests in all conditions. The denial rate is the percentage of participants who respond "No" to each request and is shown in Figure 2(a) (heuristic model applied) and Figure 2(b) (without heuristic model). It is interesting to note that some participants have indeed demonstrated inconsistent sharing behaviors.

Item 1~item 11 are all context items and are ordered according to increasing sensitivity, but the participants act similarly regardless of whether they answered the items with or without the support of heuristics. However, for demographic items (item 12~item 30, in sensitivity-ascending order), the participants tend to behave differently in the heuristic and nonheuristic environments; *their behavior is highly variable for the mild items. The heuristics successfully persuaded some participants to disclose more information*, and Figures 3(a) and 3(b) show that those *participants who lack related background knowledge for supporting their decision-making* (called amateur participants, APs) *demonstrate more varied behaviors than those participants with sufficient background knowledge* (called expert participants, EPs), who present consistent sharing behaviors, especially to the mild items 26~28.

We rank the variances and select the top 20%–40% of participants ($N = 152$) as APs and the bottom 60%–80% of participants ($N = 155$) as EPs. Generally, the context items (item 1~item 11) received considerably higher denial rates than the demographic items (item 12~item 30) in Figure 3. After checking the disclosed information and profiles, it is concluded that EPs are capable of utilizing more fruitful knowledge relating to the requested item. For example, an EP replied, "a senior website programmer" to the job-information request, while an AP responded, "a university freshman." Furthermore, EPs behave similarly on general disclosures, and most of the requested items have low denial-rate variances. Specifically, all the context items, regardless of whether they are requested before or after the demographic items, have denial-rate variances $< 0.0076$. Most of the demographic items have low variance (less than 0.0141), but there are three mild-demographic items that receive higher variances (higher than 2.1451). If we look at the APs' denial behaviors in Figure 3(a), these three items received considerably more information with the support of heuristics but were rejected in the circumstances without heuristics. *This interesting fact demonstrates that heuristics can help "persuade" participants who lack related knowledge to disclose more information only when the requests are not too sensitive*. However, in Figure 3(b), although EPs' sharing behaviors for the three mild-demographic items are more variable, their denials tend to be consistent.

Will presenting the requests to the participants in different orders of type affect disclosure? The results of four comparisons in Figure 2 of the two lines for condition $\{|S|, 0, |H|\}$ versus condition $\{|S|, 1, |H|\}$ have suggested that *participants act similarly regardless of whether the context items are requested before or after the demographic items*. As a result, we conclude that hypothesis 1 is not supported. However, comparing condition $\{0, |T|, |H|\}$ and condition $\{1, |T|, |H|\}$, participants will disclose more information for mild items when the nonsensitive items are requested beforehand (only 60 denials) than when the sensitive items are requested beforehand (as many as 109 denials). Therefore, we conclude that *presenting the requests to the participants in sensitivity-ascending order is a better plan*, and hypothesis 2 is definitely supported. The request order in terms of sensitivity does have an effect on the numbers of denials of subsequent nonsensitive items. Very highly sensitive items increased participants' privacy awareness and, thus, the participants were probably more at ease when nonsensitive items were requested.

Examining only denials of the items is not sufficient because we want to determine whether the correlations among participants' tendencies towards items change due to the provided heuristics. As mentioned before, the 30 requests are presented to the participants in eight conditions with different values of $S$, $T$, and $H$. All 30 items were presented only once to the 774 participants, and we recorded the denials of the items for each condition in vectors of length $2 \times 2 \times 2 = 8$. The correlation between each pair of items was calculated with their vectors accordingly.

Generally, the absolute value of Spearman's correlation coefficient ranges from $[0.3, 0.6]$ (397 samples for APs, M = 62.07%, SD = 7.25%; 310 samples for EPs, M = 53.33%, SD = 4.76%) to $[0.6, 1]$ (192 samples for APs, M = 77.13%, SD = 3.05%; 352 samples for EPs, M = 59.02%, SD = 1.05%). In these situations, participants all behave consistently; that is, their previous and subsequent disclosures were consistent even with the support of the heuristics. Specifically, for EPs, we observed that most of the correlations fall into the range of $[0.6, 1]$ (46.10% of the samples when the observed item and the testing item are in the same category, 49.27% of the

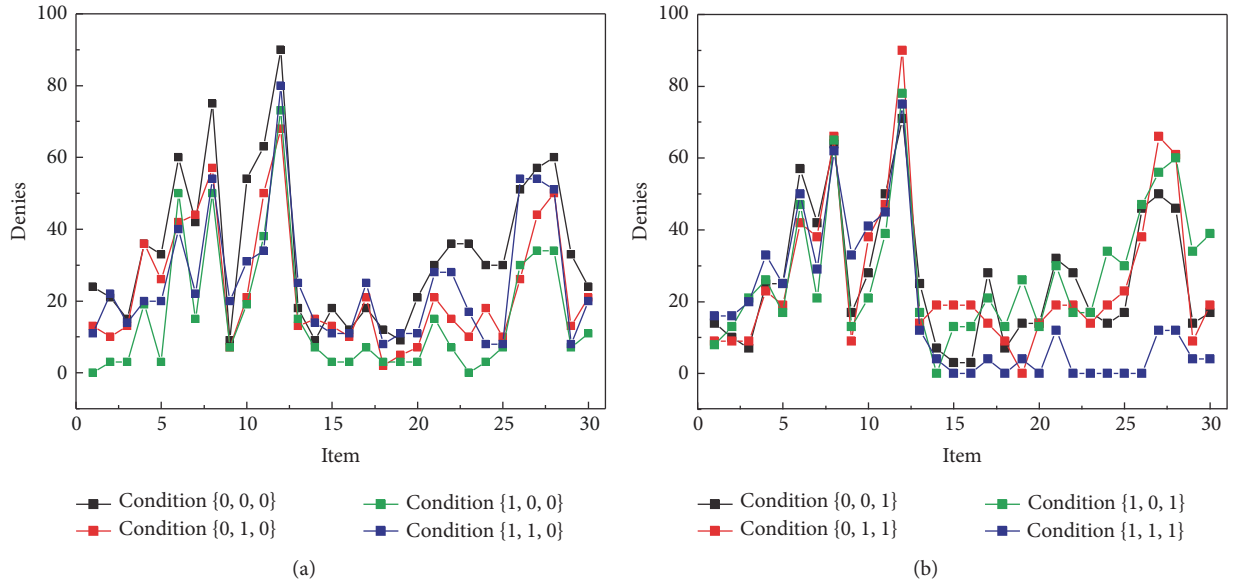


Figure 2: The numbers of denials from the participants for each requested item across eight different conditions.

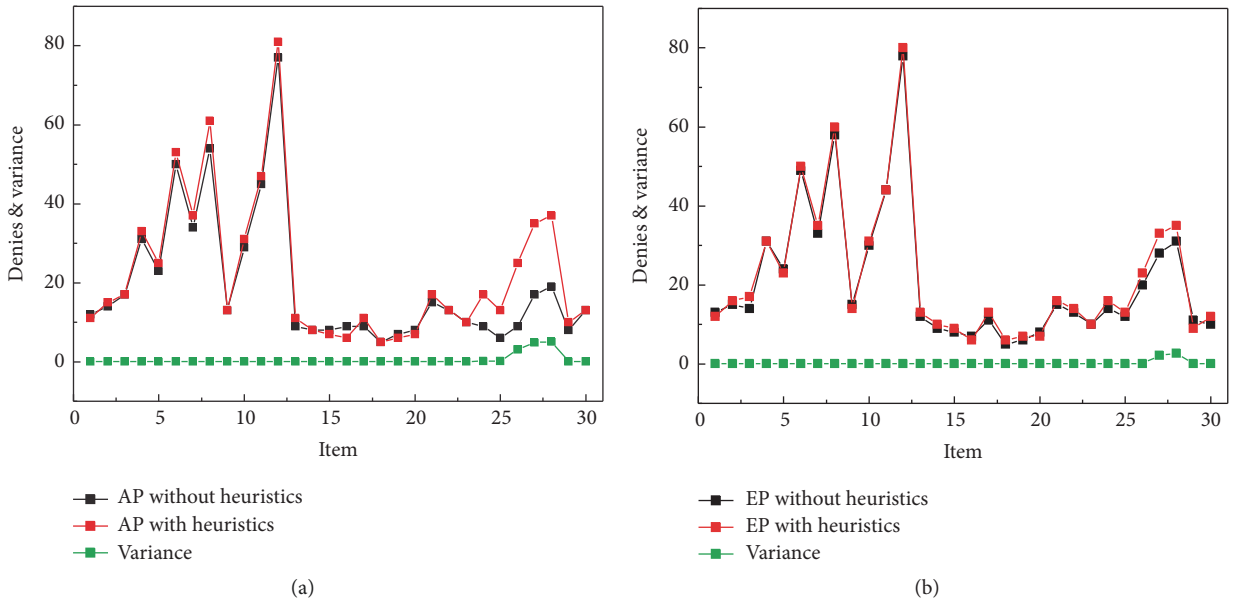


Figure 3: Numbers of denials of each requested item in all conditions for amateur APs (a) and EPs (b). Red (black) squares and lines show that this condition is (not) applied with heuristics model, and green squares and lines represent the variance of disclosure.

samples when the two items are in different categories), and only less than 5% of the samples belong to $[0, 0.3]$. For APs, 37.26% of the correlations fall in the range from 0.6 to 1 when the observed item and testing item belong to the same category, and 42.51% of the correlations fall in the range from 0.3 to 0.6 when the observed item and testing item belong to different categories. As a result, we conclude that *most of EPs maintain their information-sharing tendencies more seriously than the APs do*, and hypothesis 3 that the correlations among the requested items answered by experienced participants are less variable than those among the requested items answered by less-experienced participants is supported. *The possible reason is that EPs would like to make their own conservative decisions on what to disclose and are hardly influenced by the heuristics. However, APs were more easily influenced by the heuristics, and they may have felt delighted that some suggestions of what to disclose were provided.*

We did not expect a substantial difference in the variance of the correlations between the EPs and the APs. Thus, we conclude that hypothesis 4 is not supported. According to the results on the denials, participants are only primed by those items that they are more familiar with, such as low-sensitivity demographic items, and they may change their disclosure tendency slightly according to the heuristics model

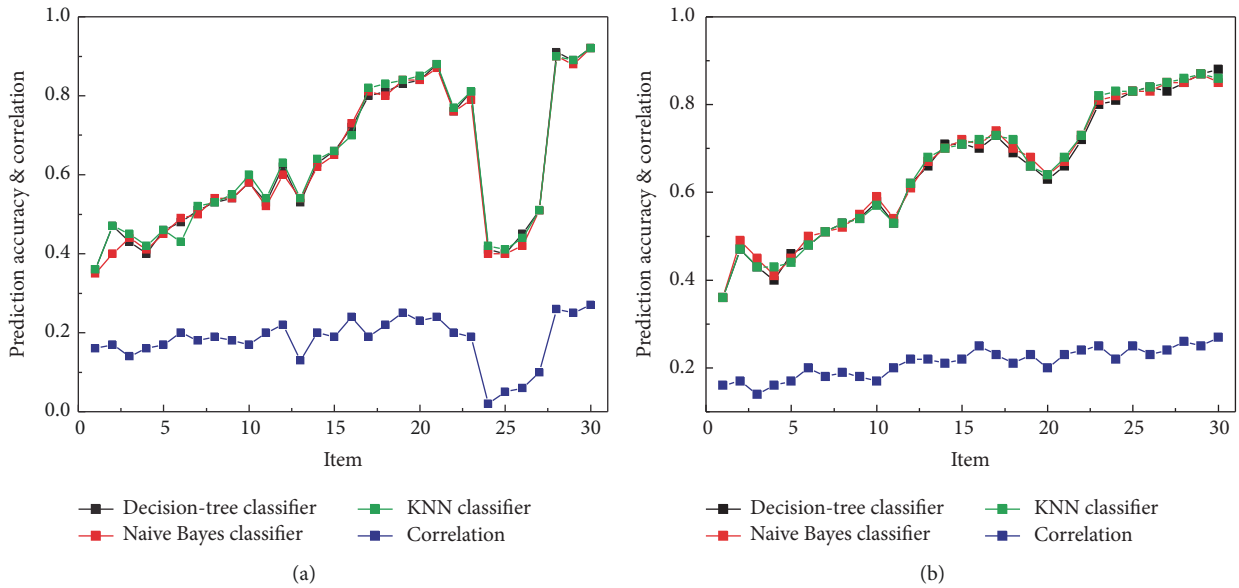


FIGURE 4: In AP users' dataset the prediction accuracy decreases when the correlation is very low, while in EP users' dataset the prediction accuracy increases when more items were loaded and no decreasing fact was found.

for high-sensitivity context items. H5 and H6 are partially supported. One possible reason behind this phenomenon is that *APs lack social experiences and are unaware of the risks and benefits of information disclosure*. They probably lack sufficient knowledge to consider the benefits and risks at the beginning of the request process but are then primed with a new disclosure tendency with the support of the heuristics model and may answer subsequent requests more carefully.

If the APs were primed with a new disclosure tendency after the heuristics, the accuracy of predictions for the next disclosures for APs should be lower than the accuracy of predictions for EPs, especially for those items with low correlations. We predict the participants' subsequent disclosures based on their previous disclosures and record the errors of the predicted disclosures with respect to the real disclosures. The accuracy of prediction for each item is the number of errors from the predictions for this item multiplied by the total number of predictions for this item. Three machine learning methods, namely, decision-tree classifier, naïve Bayes classifier, and KNN classifier, were applied to the dataset, where the value of correlation is within the range of $[0, 0.3]$. The prediction accuracy was shown in Figure 4.

*A general fact is that more loaded items in the training set will guarantee higher prediction accuracy of the machine learning methods*. Figure 4(a) has shown the correlation and the prediction accuracy in the AP users' dataset. As found before, AP users behaved less consistently towards the mild-demographic items so that the correlation becomes very low, and the prediction accuracy decreases along with the lowered correlation. However, EP users disclose their information with stronger correlation and the prediction accuracy was increasing as always. The strength of correlation between the observed item and testing item is indeed monotonous with the accuracy of predictions for the participants' next disclosures on the testing items learned by the observed item. *Specifically, the heuristic model leads AP users to behave less consistently in their disclosures, which further reduces the prediction accuracy to the mild-demographic items, while EP users change their disclosure tendency much less and maintain the prediction accuracy stable*.

## 5. Conclusions

This paper demonstrates that some users can be persuaded to alter their disclosure tendencies in a system-preferred direction, for exmple, to disclose more information, so that the recommender system can know the users better and provide service with higher prediction accuracy. In detail, users with sufficient background knowledge will maintain a stable decision-making calculus in the information disclosures, while others who lack knowledge to support their decision-making could be persuaded by the heuristic model to disclose more information. In this paper, we conclude that users disclose more information for the mild items when they lack background knowledge. This is an interesting discovery for researchers or web owners who want to obtain more information from their participants or customers. Without the knowledge that supports their decision-making, users may disclose information with less consideration of the potential risks and benefits. However, if the order in which the information is requested is not well designed, the user relies on a privacy calculus and divulges less personal information. Generally, users with knowledge enough to support their decisions are considered separately from experienced users. The behaviors of experienced users were found to be less varied according to the correlation analysis and machine learning results. Lower correlation could indicate low prediction accuracy of users' next disclosure behaviors. We believe some participants are more cognitive in nature, and knowledge obtained from their previous disclosures

could not be applied to predict their subsequent disclosures. Privacy awareness can be increased or inhibited using subtle interactive primers such as a low-quality request order; this increased awareness may make participants more likely to rely on a privacy calculus (i.e., reasoned decision behavior). Once these amateur users are primed with this privacy calculus, which is inconsistent with their previous disclosures, they are more likely to rely on it in when faced with future requests, making it difficult to predict the responses to the next requests.

The heuristic model could be applied to dataset which possibly involving time-related inconsistency, such as disclosure, emotion, and willingness. For example, if the model trained by the mentioned decision-tree classifier is no longer applicable to the overtime dataset, such as lower prediction accuracy, there should be an inconsistency occurred. That could be indicating users would like to disclose less amount of privacy, would no longer viewing their favorite movies, or would like to try a new music they had not yet listen to. With the success of detecting user changes in information disclosure decision-making by the heuristic model, we are convinced that user disclosure behaviors can be steered in a desired and preferred direction, thereby causing the recommender system to get to know its users better and faster; thus, in a social media environment, the cold-start period could be shortened, and more accurate and personalized suggestions could be given. In our future work, we will further categorize the requests into temporary requests, for example, those items for which the participants may have more varied answers, and permanent requests, for example, those items for which participants make strict decisions on disclosing the information and rarely give ambiguous answers. Hopefully, more interesting phenomena will be found.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61702312, no. 61572295); the Natural Science Foundation of Shandong Province of China (no. ZR2017BF019, no. ZR2017ZB0420); the National Key R&D Program (no. 2016YFB1000602); the project of Shandong Province Higher Educational Science and Technology Program (no. J17KB178).

## References

[1] J. H. Ziegeldorf, O. G. Morchon, and K. Wehrle, "Privacy in the internet of things: threats and challenges," *Security and Communication Networks*, vol. 7, no. 12, pp. 2728–2742, 2014.

[2] J. Zhou, M. Tang, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan, and S. Lee, "Social network and tag sources based augmenting collaborative recommender system," *IEICE transactions on Information and Systems*, vol. 98, no. 4, pp. 902–910, 2015.

[3] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, no. 99, 2016.

[4] K. Xu, Y. Guo, L. Guo, Y. Fang, and X. Li, "My privacy my decision: Control of photo sharing on online social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 2, pp. 199–210, 2017.

[5] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015.

[6] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: a recommender system for finding passengers and vacant taxis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2390–2403, 2013.

[7] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen, "Lcars: a location-content-aware recommender system," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (SIGKDD '13)*, pp. 221–229, Chicago, Illinois, USA, August 2013.

[8] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "Mobile recommender systems in tourism," *Journal of Network and Computer Applications*, vol. 39, no. 1, pp. 319–333, 2014.

[9] G. Han, L. Shu, S. Chan, and J. Hu, "Security and privacy in Internet of things: methods, architectures, and solutions," *Security and Communication Networks*, vol. 9, no. 15, pp. 2641-2642, 2016.

[10] B. Zhang, Z. Huang, J. Yu, and Y. Xiang, "Trust computation for multiple routes recommendation in social network sites," *Security and Communication Networks*, vol. 7, no. 12, pp. 2258–2276, 2014.

[11] S. S. Li and E. Karahanna, "Online recommendation systems in a B2C E-commerce context: A review and future directions," *Journal of the Association for Information Systems*, vol. 16, no. 2, pp. 72–107, 2015.

[12] A. Azaria, A. Hassidim, S. Kraus, A. Eshkol, O. Weintraub, and I. Netanely, "Movie recommender system for profit maximization," in *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13)*, pp. 121–128, October 2013.

[13] K. Lee and K. Lee, "Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4851–4858, 2015.

[14] L. Chen, R. Lu, K. Alharbi, X. Lin, and Z. Cao, "ReDD: Recommendation-based data dissemination in privacy-preserving mobile social networks," *Security and Communication Networks*, vol. 8, no. 7, pp. 1291–1305, 2015.

[15] A. Kobsa, "Privacy-enhanced web personalization," *Lecture Notes in Computer Science*, vol. 4321, pp. 628–670, 2007.

[16] M. D. Ekstrand, D. Kluver, F. M. Harper, and J. A. Konstan, "Letting users choose recommender algorithms: An experimental study," in *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15)*, pp. 11–18, September 2015.

[17] I. Kaur, T. K. Bhatia, and A. Verma, "Agile Model Based Sentiment Analysis from Social Media," *International Journal of Computer Science and Information Security*, vol. 15, p. 433, 2017.

[18] G. Di Crescenzo, D. L. Cook, A. McIntosh, and E. Panagos, "Practical and privacy-preserving information retrieval from a database table," *Journal of Computer Security*, vol. 24, no. 4, pp. 479–506, 2016.

[19] K. A. Stewart and A. H. Segars, "An empirical examination of the concern for information privacy instrument," *Information Systems Research*, vol. 13, no. 1, pp. 36–49, 2002.

[20] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren, "A privacy-preserving and copy-deterrence content-based image

retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2016.

[21] Z. Fu, X. Wu, C. Guan, X. Sun, and K. Ren, "Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2706–2716, 2016.

[22] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 9, pp. 2546–2559, 2016.

[23] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 340–352, 2016.

[24] Z. Fu, F. Huang, X. Sun, A. V. Vasilakos, and C. Yang, "Enabling semantic search based on conceptual graphs over encrypted outsourced data," *IEEE Transactions on Services Computing*, vol. PP, no. 99, 2016.

[25] Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, "Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing," *IEICE Transactions on Communications*, vol. E98B, no. 1, pp. 190–200, 2015.

[26] Q. Liu, W. Cai, J. Shen, Z. Fu, X. Liu, and N. Linge, "A speculative approach to spatial-temporal efficiency with multi-objective optimization in a heterogeneous cloud environment," *Security and Communication Networks*, vol. 9, no. 17, pp. 4002–4012, 2016.

[27] P. G. Kelley, L. F. Cranor, and N. Sadeh, "Privacy as part of the app decision-making process," in *Proceedings of the 31st Annual CHI Conference on Human Factors in Computing Systems: Changing Perspectives (CHI '13)*, pp. 3393–3402, May 2013.

[28] A. Acquisti, "Privacy in electronic commerce and the economics of immediate gratification," in *Proceedings of the 5th ACM Conference on Electronic Commerce (EC '04)*, pp. 21–29, May 2004.

[29] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *IEEE Security & Privacy*, vol. 3, no. 1, pp. 24–30, 2005.

[30] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation E-commerce: Privacy preferences versus actual behavior," in *Proceedings of the 3rd ACM conference on Electronic Commerce*, pp. 38–47, October 2001.

[31] M. J. Culnan and R. J. Bies, "Consumer privacy: Balancing economic and justice considerations," *Journal of Social Issues*, vol. 59, no. 2, pp. 323–342, 2003.

[32] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM workshop on Privacy*, pp. 71–80, November 2005.

[33] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin, "Persona: An online social network with user-defined privacy," in *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication (SIGCOMM '09)*, pp. 135–146, August 2009.

[34] B. P. Knijnenburg and A. Kobsa, "Helping users with information disclosure decisions: Potential for adaptation," in *Proceedings of the ACM International conference on Intelligent user*, pp. 407–416, March 2013.

[35] B. P. Knijnenburg and A. Kobsa, "Making decisions about privacy: Information disclosure in context-aware recommender systems," *ACM Transactions on Interactive Intelligent Systems (TiiS '13)*, vol. 3, no. 3, article no. A20, 2013.

[36] B. P. Knijnenburg, "Simplifying privacy decisions: towards interactive and adaptive solutions," in *Proceedings of the Recsys 2013 Workshop on Human Decision Making in Recommender Systems*, pp. 40-41, 2013.

[37] M. Benisch, P. G. Kelley, N. Sadeh, and L. F. Cranor, "Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs," *Personal and Ubiquitous Computing*, vol. 15, no. 7, pp. 679–694, 2011.

[38] Y. Kong, M. Zhang, and D. Ye, "A belief propagation-based method for task allocation in open and dynamic cloud environments," *Knowledge-Based Systems*, vol. 115, pp. 123–132, 2016.

[39] X. Chen, S. Chen, and Y. Wu, "Coverless information hiding method based on the Chinese character encoding," *Journal of Internet Technology*, vol. 18, no. 2, pp. 91–98, 2017.

[40] Y. Xue, J. Jiang, B. Zhao, and T. Ma, "A self-adaptive artificial bee colony algorithm based on global best for global optimization," *Soft Computing*, pp. 1–18, 2017.

[41] C. Yuan, Z. Xia, and X. Sun, "Coverless image steganography based on SIFT and BOF," *Journal of Internet Technology*, vol. 18, no. 2, pp. 209–216, 2017.

[42] B. Liu, J. Lin, and N. Sadeh, "Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help?" in *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*, pp. 201–211, April 2014.

[43] H. Zhang and J. Lu, "Creating ensembles of classifiers via fuzzy clustering and deflection," *Fuzzy Sets and Systems*, vol. 161, no. 13, pp. 1790–1802, 2010.

[44] G. Pallapa, S. K. Das, M. Di Francesco, and T. Aura, "Adaptive and context-aware privacy preservation exploiting user interactions in smart environments," *Pervasive and Mobile Computing*, vol. 12, pp. 232–243, 2014.

[45] H. Zhang, L. Cao, and S. Gao, "A locality correlation preserving support vector machine," *Pattern Recognition*, vol. 47, no. 9, pp. 3168–3178, 2014.

[46] Y. Wang, H. Zhang, and F. Yang, "A Weighted Sparse Neighbourhood-Preserving Projections for Face Recognition," *IETE Journal of Research*, vol. 63, no. 3, pp. 358–367, 2017.

[47] S. T. Margulis, "Privacy as a social issue and behavioral concept," *Journal of Social Issues*, vol. 59, no. 2, pp. 243–261, 2003.

[48] A. N. Joinson, A. Woodley, and U.-D. Reips, "Personalization, authentication and self-disclosure in self-administered Internet surveys," *Computers in Human Behavior*, vol. 23, no. 1, pp. 275–285, 2007.

[49] H. Wu, X. Wang, Z. Peng, and Q. Li, "Div-clustering: Exploring active users for social collaborative recommendation," *Journal of Network and Computer Applications*, vol. 36, no. 6, pp. 1642–1650, 2013.

[50] H. Wu, B. P. Knijnenburg, and A. Kobsa, "Improving the prediction of users disclosure behavior by making them disclose more predictably?" in *Proceedings of the Symposium on Usable Privacy and Security (SOUPS '14)*, 2014.