Analyzing Impact Factors for Smartphone Sharing Decisions Using Decision Tree

Tao Xu[1,2(✉)], Yun Zhou[3], Alexander Raake[4], and Xuyun Zhang[5]

[1] School of Software and Microelectronics, Northwestern Polytechnical University,
Xi'an 710072, Shaanxi, People's Republic of China
`xutao@nwpu.edu.cn`
[2] State Key Laboratory for Manufacturing Systems Engineering,
Xi'an Jiaotong University, Xi'an 710054, Shaanxi, People's Republic of China
[3] School of Education, Shaanxi Normal University,
Xi'an 710062, Shaanxi, People's Republic of China
`zhouyun@snnu.edu.cn`
[4] Audiovisual Technology Group, Institute for Media Technology,
University of Technology Ilmenau, 98693 Ilmenau, Germany
`alexander.raake@tu-ilmenau.de`
[5] Department of Electrical and Computer Engineering, University of Auckland,
Auckland 1010, New Zealand
`xuyun.zhang@auckland.ac.nz`

Abstract. The sharing activities of smart devices have been proved to occur commonly among users and discussed by many studies. In contrast to sharing behaviors of personal computer, smartphone sharing is impromptu. It occurs pervasively like on the street between strangers or at home between family members. The variety of sharing behaviors requires the investigation of the most impact factors for smartphone sharing, which underlies the design of usable privacy user interface. This work investigates and analyzes the influencing factors of smartphone users sharing decisions based on CART algorithm of Decision Tree. The data for analyzing is based on a survey involving 165 participants' responses. Results indicated the features belonging to the sharing attitudes and behaviors impacted users' sharing decisions mostly.

Keywords: User study · Sharing · Smartphone · Decision tree

1 Introduction

Smart devices sharing [4,10] has been proved to occur commonly among users and discussed by many studies, even though smartphones are personal objects. Compared with sharing behaviors of personal computer, smartphone sharing is impromptu. It occurs pervasively like on the street between strangers or at home between family members. The impromptu behaviors and various sharing decisions require the investigation of impact factors for smartphone sharing, which

underlies the design of usable privacy user interface. Due to the missing studies, this work investigates and analyzes the influencing factors of smartphone users sharing decisions using the CART algorithm of Decision Tree. First we collect data from a survey with 165 pieces of responses, obtaining users' attitudes towards tablet and smartphone sharing respectively. Next, we present the Decision Tree induction, CART algorithm and Gini index. The Gini index was used for analyzing and three users' sharing decisions were explored, including: the willingness to use multiple user accounts as the owner, the willingness to use multiple user accounts as the guest, and whether the individual taking actions to delete the trace history as the guest. Finally, results and a discussion were presented.

2   Related Work

2.1   Sharing Practices

Sharing practices analysis is the first step to investigate the sharing behaviors. In the work [4], Karlson et al. interviewed 12 smartphone users to explore the diversity of guest user categorizations and associated security constraints expressed by the participants. Hang et al. [1] conducted a focus group and a user study to analyze which data people are concerned of, which data people are willing to share and with whom people would share their device. The study [5] examined the privacy expectations of smartphone users by exploring participants' concerns with other people accessing the personal information and applications accessing this information via platform APIs.

Most of the time, it is daunting to identify the likely number of types and categorize behaviors into distinct types. Prior studies obtained sharing types by collecting data and analyzing distinct patterns from the observation. For example, Zhou et al. analyzed the relationship of sharer and sharee, sharing activities could be categorized as stranger, acquaintance, close people and kids [12]. Matthews et al. [6] organized sharing practices into a taxonomy of six sharing types, including borrowing, mutual use, setup, helping, broadcasting and accidental. This work involved a survey of 99 participants, a 21 days diary study with 25 participants and interviews with 24 participants. Behaviors are under the influence of multiple factors. In the work [11], the authors investigated the correlations between basic information, general privacy attitudes and sharing behaviors. However, with regard to sharing decisions, the analysis of the sharing types as well as the correlation and association is inadequate; it requires figuring out the influencing factors.

2.2   Solutions to Protect Privacy

Previous studies showed than users had concerns and negative emotions on the privacy leakage. Both their attitudes and behaviors implied that all-or-nothing mechanisms of privacy interfaces [2,7] on protecting privacy have not been useful

to assist individuals to control their personal information protection adequately. To bridge the gap between people's requirements and all-or-nothing mechanisms, new access control of exposing more options for users has been obtained the focus of researchers, the usability of which have been tested and investigated. The numbers of researches that revolve around exploring usability of new access control mechanisms and interfaces, as well as users' perception and feedbacks on the complex terminologies and interfaces, are increasing. For example, in the work [2], questions were designed and the usability of both all-or-nothing control and new access control mechanisms have been tested and the results showed that the new flexible access control gained more interests of participants.

Although the fine-grained access control is beneficial for users, there are still some problems requiring be considering and solving. Achieving new access control of exposing more options for users requires increased effort of users [9], and a heavy burden to understand terminologies and options would be on users when the number of apps increasing [10]. With regard to sharing activities, smartphone OSs have started to consider multiple user account or guest mode important since 2014, and integrated this feature into OSs. Android users can access multiple-user settings and employ a Guest/Profile/User account to share their smartphone in a safe way. Users of iOS, could define and block several interactive areas in Settings under the category of accessibility for the individual who is shared to use the phone. In this way, the owner of the phone could explicitly switch to this feature and disable user interface items that were not wanted to be used by a guest. Even though multiple user account or guest mode is beneficial for users, the burden of interactivity and understanding terminologies confuses the users. Therefore, to investigate the usability and finding out the issues existing in the usability has been considered essential. Among exploring the issues of usability, figuring out impacting factors of sharing decisions could provide the implications for the design of usable privacy interface, especially for informing the specific interface for sharing phone with the kid, the acquaintance, the close person and the stranger.

3   The Architecture of Analysis

Decision Tree (DT) induction [3] constructs a tree-like flowchart model where each nonleaf node denotes a test on a feature, each branch refers to an outcome of the test, and each leaf node corresponds to a class prediction. At each node, the algorithm selects the assumed "best" feature to partition the data into individual classes. Quinlan [8] developed a decision tree algorithm known as ID3 (Iterative Dichotomiser). Then Quinlan presented a successor of ID3 later, that is, C4.5. CART tree is also an algorithm used commonly. Decision Tree algorithms (e.g., ID3, C4.5, and CART) were originally used for classification. It is the learning of decision trees from class-labeled training tuples. The construction of decision tree classifiers does not require any prior knowledge or parameter setting. Therefore, it is appropriate for exploratory knowledge discovery. Besides, DT has the ability to handle multidimensional data and the acquired knowledge represented

in tree form is intuitive and easy to interpret. Usually, the training steps and classification speed are fast and the performance is good. DT induction has been widely used in many research and industry areas.

Commonly, the data sets used for analysis contain hundreds of features (also referred as attributes or dimensions). Many of these features may be redundant or irrelevant to the classification tasks. Keeping irrelevant or redundant features for analysis may be detrimental for the algorithm to analyze the data, as well as consume resources and slow down the processing speed. Therefore, rejecting these features is an essential step for further analysis. Feature subset selection (also known as attribute subset selection in data mining) could reduce the size of the dataset through removing irrelevant or redundant features. One of the feature subset selection methods is decision tree induction. When the decision tree induction method is employed for feature subset selection, a tree is constructed without the irrelevant features based on the given data. In this way, the set of features appearing in the tree form a reduced subset of features.

During tree construction, feature selection measures (or attribute selection measures, known as splitting rules) are used to select the feature that best partitions the tuples into distinct classes. A feature selection measure is a heuristic process which is for selection splitting criterion that "best" divides a given partition. Information gain, gain ration and Gini index are three most popular feature selection measures.
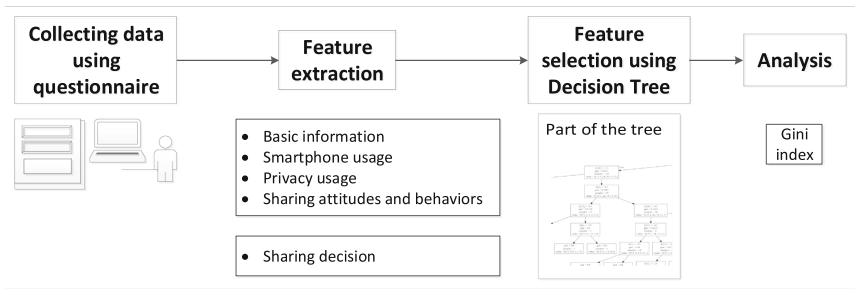

Fig. 1. The architecture of analysis

In this exploratory and preliminary work, we use the CART algorithm to test and find the most relevant features, which impact users' sharing decisions. Therefore, the Gini index will be discussed in the section of Results and Discussion. The architecture of analysis is as shown in Fig. 1, containing four steps: collecting data using questionnaire, extracting features, selecting features as the impacting factors using decision tree, and analyzing the impacting factors. First, we propose a questionnaire to obtain users' attitudes towards phone sharing and their sharing decisions. Then, the features are extracted and the values of features are collected based on participants' responses. Next, we employ the data to construct DT induction and measure Gini index. Finally, we analyze the impacting factors through the Gini index of each category of the feature.

4   Survey Participants and Data Collecting

In our previous work [12], we conducted a fine-grained survey with 165 participants, including German and Chinese to study users' attitudes towards smartphone sharing as the owner and the guest, and their behaviors on such control mechanism. To go a step further, to know at which level the factors influencing on sharing activities, in this latest work, we use the decision tree algorithm to process the data.

4.1   Participants

165 participants' responses were collected and made into a dataset. In this dataset, we had 69.1% males and 30.9% females. Their ages were distributed and covered almost all age groups. However, 66.7% were in the range of 18 to 24. There was a bias towards higher education levels and 94.0% of participants were under a higher education.

4.2   Data Collecting

The dataset were divided into five following parts:

(a) Basic information including (1) age, (2) gender, (3) nationality, (4) educational level
(b) Smartphone usage including (1) budget for smartphone, (2) price of owned phone, (3) numbers of years' experience using smartphone, (4) smartphone OS, (5) numbers of apps installed, and (6) the time spent on smartphone daily
(c) Privacy usage including (1) importance and (2) sensitivity of personal information stored on smartphone
(d) Sharing attitudes and behaviors, including (1) frequency of sharing behaviors, (2) concern levels when lending the phone, (3) the time let others use the phone, and (4) necessity of using multiple user accounts as the owner
(e) Sharing Decision, including (1) the willingness to use multiple user accounts as the owner, (2) the willingness to use multiple user accounts as the guest, and (3) whether taking actions to delete the trace history as the guest.

The features (or could be also called as variables in this paper) and the scales of measurement that are used for analyzing are listed as below:

(a) age, ordinal
(b) gender, categorical
(c) nationality, categorical
(d) education level, ordinal
(e) budget for smartphone, ordinal
(f) price of owned phone, ordinal
(g) numbers of years' experience using smartphone, ordinal
(h) smartphone OS (Operating System), categorical

  (i) numbers of apps installed, ordinal
  (j) the time spent on smartphone daily, ordinal
 (k) the importance of personal information stored on smartphone, interval
  (l) the sensitivity of personal information, interval
(m) frequency of sharing behaviors, interval
 (n) concern levels when lending the phone, interval
 (o) the time let others use the phone, ordinal
 (p) necessity of using multiple user accounts as the owner, interval.

We listed the following the classes and the scales of measurement that are used for analyzing:

(a) the willingness to use multiple user accounts as the owner, interval
(b) the willingness to use multiple user accounts as the guest, interval
(c) whether taking actions to delete the trace history as the guest, categorical.

The former sixteen items are selected as features, that is, the impacting factors. The last three items are treated as the classes.

5   Results and Discussion

In this work, we construct the decision trees that are built using basic information, smartphone usage, privacy usage and sharing attitudes as features and three sharing decisions as three classes. The decision tree has been built using basic information, smartphone usage, privacy usage and sharing attitudes as features and the first sharing decision as the class, which is a multi-class decision tree. As shown in Fig. 2, with regard to feature ranking, we plot a tree-like structure based on the first four layers of this decision tree and present the Gini index. The X[number] refers to a feature, which has been listed as shown on the right. The features that were selected and ranked by the tree induction had a higher weight, which means these features determined the users' sharing decisions. Figure 2 indicated the features as shown on the right impacted the willingness to use multiple user accounts as the owner, the answer of which were organized with the Likert items based on seven-point scale with anchor points of "1-weak/7-strong".

Figure 3 indicated the features as shown on the right impacted the willingness to use multiple user accounts as the guest, the answer of which were organized with the Likert items based on seven-point scale with anchor points of "1-weak/7-strong".

Figure 4 indicated the features as shown on the right impacted whether users were taking actions to delete the trace history as the guest, the answer of which were organized with four options of "Yes, I tried to go something, like logging out, deleting the number I dialed", "No, I did nothing, because I thought I don't need to take actions", "No, I did nothing, but I thought I need to take actions", and "No, I did nothing because I didn't become aware of privacy issues".

Figure 5 indicated the features as shown on the right impacted the three classes as stated above. Among these features, the following three features were mostly impacting users' sharing decisions:
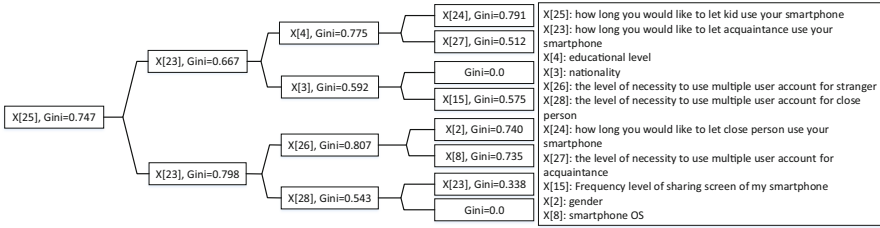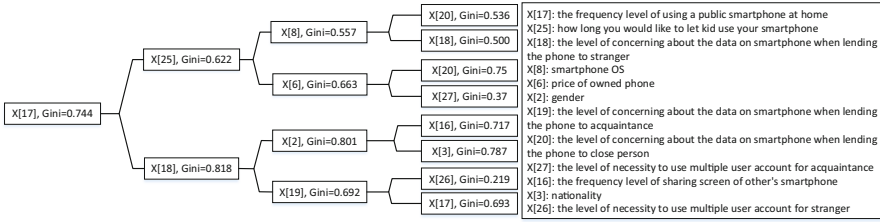
Fig. 2. Part of the tree for the class 1.
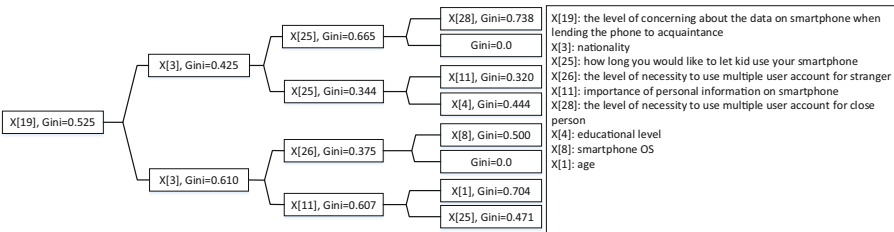

Fig. 3. Part of the tree for the class 2.
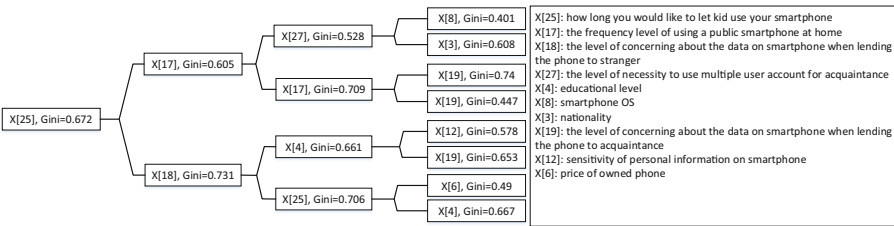

Fig. 4. Part of the tree for the class 3.

Fig. 5. Part of the tree for all classes.

(a) how long the individual would like to let kid use the smartphone
(b) the frequency level of using a public smartphone at home
(c) the level of concerning about the data on smartphone when lending the phone to stranger.

6   Conclusion

This work investigates and analyzes the influencing factors of smartphone users sharing decisions using the CART algorithm of Decision Tree. The data for analyzing is based on a survey involving 165 participants' responses. The Gini index was used for analyzing and three users' sharing decisions were explored, including: the willingness to use multiple user accounts as the owner, the willingness to use multiple user accounts as the guest, and whether the individual taking actions to delete the trace history as the guest. Results sharing decisions indicated the following features impacted these sharing decisions mostly: "how long the individual would like to let kid use the smartphone", "the frequency level of using a public smartphone at home" and "the level of concerning about the data on smartphone when lending the phone to stranger".

Appendix

Basic information

1. Which of the following group best describes your age?
2. What is your gender?
3. Which of the following best describes the highest level of your education?

Smartphone usage

1. Which of the following best describes your budget to buy a smartphone?
2. Which of the following best describes the price of your smartphone?
3. How long have you been using a smartphone generally?
4. Which smart device operating system (OS) are you using?
5. How many apps do you install by yourself on your phone?
6. How much time on average do you spend on using apps (free or paid) and internet on your smart devices per week?

Privacy usage

1. Please rate the importance of personal information on your smartphone.
2. Please rate the sensitivity of personal information on your smartphone.

Sharing attitudes and behaviors

1. Please select which best describes the frequency of your sharing behaviors (lending my smartphone, borrowing other's smartphone, sharing screen of my smartphone, sharing screen of other's smartphone, using a public smartphone at home).

2. Please describe how much you are concerned about your data on your smart phone when you lending your phone to stranger, someone you know but not close (neighbor, coworker, etc.), closed people (like good friend, brother, wife, etc.), kid (kid of yours, friend's, relative's, etc.).
3. Please indicate how long you would like to let the people use your smart phone (stranger, someone you know but not close, closed people, kid).
4. How necessary do you think it's to use multiple user account/guest mode for each person (stranger, someone you know but not close, closed people, kid).

Sharing Decision

1. Please indicate how much for each device you would like to use the one-button function to clear the guests' trace histories (to protect guest's privacy), instead that the guests clear and delete their traces manually.
2. Please indicate how much for each device you would like to use the one-button function to clear your trace histories as the guest to protect your privacy, in case you may forget to log out or delete browser histories.
3. Did you take actions to delete/erase your trace history before you return smartphone?

References

1. Hang, A., Von Zezschwitz, E., De Luca, A., Hussmann, H.: Too much information! User attitudes towards smartphone sharing. In: Nordic Conference on Human-Computer Interaction: Making Sense Through Design, pp. 284–287 (2012)
2. Hayashi, E., Riva, O., Strauss, K., Bernheim Brush, A.J., Schechter, S.: Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In: Eighth Symposium on Usable Privacy and Security, p. 2 (2012)
3. Karimi, K., Hamilton, H.J.: Generation and interpretation of temporal decision rules. Int. J. Comput. Inf. Syst. Ind. Manag. Appl. **3**(1), 314–323 (2010)
4. Karlson, A.K., Bernheim Brush, A.J., Schechter, S.: Can i borrow your phone? Understanding concerns when sharing mobile phones. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1647–1650 (2009)
5. King, J.: How come i'm allowing strangers to go through my phone? Smartphones and privacy expectations. SSRN Electron. J. (2014)
6. Matthews, T., Liao, K., Turner, A., Berkovich, M., Reeder, R., Consolvo, S.: "She'll just grab any device that's closer": a study of everyday device & account sharing in households. In: CHI Conference on Human Factors in Computing Systems, pp. 5921–5932 (2016)
7. Mazurek, M.L., Arsenault, J.P., Bresee, J., Gupta, N., Ion, I., Johns, C., Lee, D., Liang, Y., Olsen, J., Salmon, B.: Access control for home data sharing: attitudes, needs and practices. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 645–654 (2010)
8. Quinlan, J.R.: Simplifying Decision Trees. Academic Press Ltd., London (1987)
9. Smetters, D.K., Good, N.: How users use access control. In: Symposium on Usable Privacy and Security, SOUPS 2009, Mountain View, California, USA, July 2009, pp. 1–12 (2009)

10. Zhou, Y., Piekarska, M., Raake, A., Tao, X., Xiaojun, W., Dong, B.: Control yourself: on user control of privacy settings using personalization and privacy panel on smartphones. Procedia Comput. Sci. **109**, 100–107 (2017)
11. Zhou, Y., Raake, A., Xu, T., Zhang, X.: Users' perceived control, trust and expectation on privacy settings of smartphone. In: Wen, S., Wu, W., Castiglione, A. (eds.) CSS 2017. LNCS, vol. 10581, pp. 427–441. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69471-9_31
12. Zhou, Y., Xu, T., Raake, A., Cai, Y.: Access control is not enough: how owner and guest set limits to protect privacy when sharing smartphone. In: Stephanidis, C. (ed.) HCI 2016. CCIS, vol. 617, pp. 494–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40548-3_82