Predicting Smartphone Adoption in Social Networks

Le Wu[1], Yin Zhu[2], Nicholas Jing Yuan[3], Enhong Chen[1](✉), Xing Xie[3], and Yong Rui[3]

[1] University of Science and Technology of China, Hefei, China
cheneh@ustc.edu.cn
[2] Hong Kong University of Science and Technology, Hong Kong, China
[3] Microsoft Research, Beijing, China

Abstract. The recent advancements in online social networks and mobile devices have provided valuable data sources to track users' smartphone adoption, i.e., the usage of smartphones over time. An incisive understanding of users' smartphone adoption can benefit many useful applications, ranging from user behavior understanding to targeted marketing. This paper studies smartphone adoption prediction in social networks by leveraging the wisdom of an online world. A critical challenge along this line is to identify the key factors that underline people's adoption behaviors and distinguish the relative contribution of each factor. Specifically, we model the final smartphone status of each user as a result of three influencing factors: the social influence factor, the homophily factor, and the personal factor. We further develop a supervised model that takes all three factors for smartphone adoption and at the same time learns the relative contribution of each factor from the data. Experimental results on a large real world dataset demonstrate the effectiveness of our proposed model.

Keywords: Smartphone adoption · Social network · Social influence · Homophily

1 Introduction

Smartphones (e.g., iPhone and Android based mobile phones) are now ubiquitous in our daily lives. There were 1.82 billion smartphones being used worldwide at the end of 2013. Furthermore, according to a forecast by International Data Corporation, the smartphone market is expected to increase to 70.5% in 2017 in terms of all smart devices, including desktop PCs, portable PCs, tablets and smartphones [1].

With the expanding opportunities in the smartphone market, an incisive understanding of smartphone adoption among users has significant applications ranging from user behavior understanding in scientific disciplines [4,19] to targeted advertising for marketing strategies [7,10]. Thus, acceptance or adoption of smartphones has long been studied in the past from a variety of angles, such as cultural factors [23], technology needs [22] and perceived usefulness [18]. Nearly

T. Cao et al. (Eds.): PAKDD 2015, Part I, LNAI 9077, pp. 472–485, 2015.
DOI: 10.1007/978-3-319-18038-0_37

all these studies were based on traditional survey based approaches, e.g., by surveying hundreds of people. With both the time and money costs of collecting data, few researchers have attempted to investigate the smartphone adoption within a large-scale social network.

Luckily, with the recent advancements of online social networks and smartphones, an increasing number of people are sharing their daily lives with friends on these platforms through smartphones. Due to the mobile nature of the login devices, these *mobilized social networks* record the smartphone footprints of users. To illustrate this, we provide the following example. *Weibo* (weibo.com) is the leading microblog service in China. When a user posts a message in *Weibo*, the platform forwards an enriched message to all of the user's followers as shown in Fig. 1, which includes the post message, the timestamp and the sending device (iPhone). This device information creates valuable data sources to track smartphone adoption within a large-scale social network.

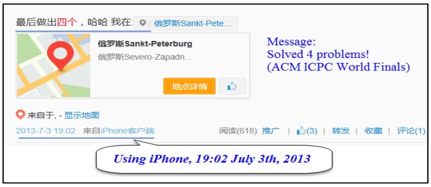


Fig. 1. A sample post from Weibo

As a matter of fact, even with the mobilized social network data, accurately understanding a user's smartphone adoption is still technically challenging from at least two aspects. On the one hand, there are various factors that underline person's decision-making process. How can we leverage them in a unified framework? Researchers have long identified three key factors for this process: the social influence factor that argues users are influenced by their social neighbors to make decisions [11,25]; the homophily factor refers to linked users performing similar decisions [15]; and the personal factor states users have their own personalized preferences [29,30]. On the other hand, though all these key factors help predict users' adoption behaviors, they lead to significantly different results [2,14]. Accurately understanding and distinguishing the relative contribution of each factor is critically important to guiding the firm's marketing strategy. E.g., if the social influence is responsible for users' decisions, then it is effective for the firm to incentivize several seed customers to trigger a cascade of information diffusion [11]. If the homophily factor dominates, then the firm can identify new potential customers based on each user's neighbors' decisions. If the final adoption behavior is driven by the personal factor, a better idea is to select the targeted customers based on their historical preferences for marketing. Nevertheless, few previous methods have incorporated all these principal factors together for product adoption prediction. Therefore, how to leverage all these key factors and distinguish them at the same time for smartphone adoption prediction remains pretty much open.

In order to solve both the data barrier and technical challenges mentioned above, in this paper, we propose a supervised machine learning model for smartphone adoption prediction. As a preliminary, we leverage a mobilized online social network to discover the smartphone usage patterns of a large group of networked users. Then, by borrowing the traditional user segmentation concept, we identify two groups of users based on their current smartphone status, i.e., potential first-time smartphone adopters and potential brand changers, respectively. After that, we develop a *S*upervised *H*omophily-*I*nfluence-*P*ersonality (SHIP) model for smartphone adoption prediction, in which the key factors that underline people's adoption are explicitly integrated. In fact, the proposed model can easily be extended to other product adoption tasks. Finally, the experimental results on 200K active mobile users show the effectiveness of our proposed model. To the best of our knowledge, this is the first comprehensive attempt to predict smartphone adoption from a social perspective with large-scale real world data.

2 Data Description and Problem Definition

Given a snapshot of a social network as a directed graph $G = < U, F, \mathbf{T} >$, where the node set $U = \{1, 2, ..., N\}$ is the users and F represents the relationships of users. $\mathbf{T} = [t_{ji}]_{N*N}$ is an edge strength matrix, where t_{ji} represents the tie strength from user j to user i. Specifically, if user i follows user j, then $(i, j) \in F$ and $t_{ji} > 0$, otherwise $t_{ji} = 0$. Since we mainly focus on the smartphone adoption of users, for ease of later explanation, a *mobile post* is defined as a post that is sent from a smartphone, rather than a PC client or a tablet. If a user sends more than τ mobile posts in a time period, we regard him/her as a *mobile user*.

Data Collection and Description. During the data crawling process, we collected the post streams of nearly 235 thousand users from January 2013 to July 2013 from *Weibo*. For data cleaning, we only selected mobile users (i.e., $\tau = 10$ empirically) and their associated relations. After pruning, we still had nearly 200 thousand users, 15 million edges, and 45 million post streams. Now we introduce how to infer each user's smartphone status from the continuous post streams. Similar to many smartphone marketing research [1], we treat each quarter as a time slice and further split each user's device streams into two time slices, i.e., the first quarter (2013Q1) and the second quarter (2013Q2) of 2013. Then we take the most popular device brand as the smartphone status of the user at that time. Note that a user may use several smartphones during a time slot, however, it is reasonable to discard the infrequent uses of other smartphone brands since users prefer those phones that they use the most frequently. Table 1 shows an example of the inferred smartphone status of two typical users.

Problem Definition. Generally, our goal is to predict the smartphone adoption status of users in time t based on the available data in the previous time $t-1$. In marketing research, a common practice is to first divide a broad target market

Table 1. Examples of two typical users' post device streams in Weibo

Alice				*Bob*			
Time slice	Timestamp	Sent device	Device status	Time slice	Timestamp	Sent device	Device status
2013Q1	20130211 20130212 ... 20130331	Web Weibo Web Weibo ... Web Weibo	$\Longrightarrow$ Desktop	2013Q1	20130211 20130220 ... 20130331	Nokia 5230 Nokia 5230 ... Nokia 5230	$\Longrightarrow$ Nokia
2013Q2	... 20130421 20130425 20130428 ...	... Web Weibo Samsung Galaxy S2 Samsung Galaxy S2 ...	$\Longrightarrow$ Samsung	2013Q2	... 20130419 20130422 20130423	... Nokia 5230 iPhone iPhone ...	$\Longrightarrow$ iPhone

into subsets of consumers and then design strategies to target each group of consumers. Following this approach, we divided users into two groups based on their smartphone status in $t-1$: potential first-time smartphone buyers and potential brand changers. The potential brand changers are those who have already used a smartphone in $t-1$ and their next action is deciding whether to change brands in t. E.g., as illustrated in Table 1, *Bob* is a potential brand changer as he had the *Nokia* smartphone in $2013Q1$ (i.e., time $t-1$). We regard those who do not use any smartphone in $t-1$ as potential first-time smartphone buyers. This assumption may not be accurate when applied to each person, but the overall trend is well supported by the high penetration of mobilized social networks in our everyday life. *Alice* is a potential first-time buyer as shown in Table 1. After segmenting users into these two groups, we set the target for each group as follows:

Task 1: First-time Buying Prediction. If a user is a potential first-time buyer in $t-1$, we predict whether she/he will buy a particular brand b in time t or not.

Task 2: Brand Change Prediction. If a user is a potential brand changer in time $t-1$, we predict whether this user will change to another brand in the next time period t.

We next assigned a label to each user based on the group information and the smartphone adoption status in t. E.g., *Alice* is a member of Task 1 and buys a Samsung in t, so she is a positively labeled user if we focus on predicting whether she will buy a *Samsung*. *Bob* is a positively labeled user in Task 2 as he changed from *Samsung* to *iPhone* in t. In summary, after user segmentation and label assignment for each task, these two tasks can be summarized in a unified prediction problem: Given a snapshot of a directed social network $G = < U, F, \mathbf{T} >$ with a positively labeled user set *UP* and a negatively labeled user set *UN* in time t, our goal is to predict the labels of all unknown users at time t as accurate as possible. In the next section, we focus on the model.

3 The Proposed SHIP Model

Researchers have long converged on the idea that there are three principal factors that drive people's adoption decisions: the social influence factor and the homophily factor that lead to correlated user behaviors among linked users and the personal factor that states users' unique preferences [15,25,29]. Obviously, each of these three factors can exploit a specific part of users' decisions. In addition, as illustrated before, different factors result in significantly different marketing strategies [2,14]. Thus, simply aggregating all these factors for prediction will not be the best choice. A better idea is to distinguish the relative effect of each factor in the decision making process. In the following, we propose a Supervised Homophily-Influence-Personlity (SHIP) model that can automatically learn the contribution of each factor for users' smartphone adoption. Next we describe how to construct the SHIP model step by step.

Overview of Smartphone Adoption Function. For each user i, we explicitly model the smartphone adoption status p_i as a combination of the three key factors:

$$p_i = (1-\alpha)\sum_{j\in F_i} t_{ji}[(1-\beta)p_j + \beta u_{ji}] + \alpha b_i, \tag{1}$$

where p_i is the predicted smartphone adoption probability that ranges from 0 to 1. F_i are the users that i follows in this network. t_{ji} represents the strength between i and j. If i follows j, then $t_{ji} = \frac{1}{|F_i|}$, otherwise $t_{ji} = 0$. We have two parts in this equation, the first part captures the social network effect (including social influence and homophily) and the second part (b_i) mimics the personal bias. Specifically, for user i and any user j that i links ($j \in F_i$), i's adoption probability p_i is balanced by the influence of j's adoption status p_j (social influence) and the homophily effect u_{ji}, where β controls the relative contribution of these two factors in social networks. α ($0 \leq \alpha \leq 1$) is a parameter that controls the relative effect of the social network and personal bias. The larger the α, the more personal preference plays a role in the task.

Since for each pair of linked users, we have a vector $\boldsymbol{e_{ji}}$ that captures the various features between them, we model the homophily, i.e., the similarity between each pair of linked users as:

$$u_{ji} = s(\boldsymbol{w}\cdot(\boldsymbol{e_{ji}})) = s(\sum_k w_k \times e_{jik}), \tag{2}$$

where e_{jik} is the k-th element of $\boldsymbol{e_{ji}}$. Similarly, for each user i, we have $\boldsymbol{x_i}$, which captures her various characteristics. Then the personal bias can be defined as:

$$b_i = s(\boldsymbol{v}\cdot x_i) = s(\sum_k v_k \times x_{ik}). \tag{3}$$

In the above two equations, $s(l)$ can be set as any monotonically increasing function. As $\forall i \in U, 0 \leq p_i \leq 1$, for fair comparison of the different effects, these values are better ranges in $[0,1]$. Thus, a natural idea is to set $s(l)$ as a logistic function $s(x) = \frac{1}{1+e(-x)}$.

Note that the proposed smartphone adoption probability function (Eq.(1)) has close relationship with the recent progress in supervised random walk based models. These models incorporated the node and edge features to supervise the random walk process for node classification tasks [3,28]. E.g., the works of

[3] utilized the social influence factor and [28] further extended this work by incorporating the personal bias. Nevertheless, the homophily factor is neglected by all these previous works in the modeling process, while we explicitly depict the homophily factor between each pair of linked users. In other words, the previous works for node classification can be seen as special cases of our models, e.g., our work is reduced to the models proposed by Zeng et al. when excluding the homophily factor [28].

Optimization Function Construction. Based on Eq. (1) , in order to get the final label preference p_i for each user i, we have to learn four parameters $\theta = [\alpha, \beta, \boldsymbol{w}, \boldsymbol{v}]$ in the training process. As we have a set of users' labels in the training data at time t, an intuitive idea is to train a supervised model that automatically learns the parameters θ such that all labeled positive users in the training data have larger probabilities than the labeled negative ones. Next, we model the objective learning function as:

$$\min_{\boldsymbol{\theta}} L = \sum_{i \in UP} \sum_{j \in UN} h(p_j - p_i) + \lambda[\boldsymbol{w}'\boldsymbol{w} + \boldsymbol{v}'\boldsymbol{v}], \tag{4}$$

where the first term models the goodness for fitting the data and the second term controls model complexity. Since j is a negatively labeled user and i a positively one, the larger the p_i the better and the smaller the p_j the better. Based on the above, we empirically set $h(x)$ as:

$$h(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{1+e^{-cx}} & \text{if } x \geq 0. \end{cases} \tag{5}$$

Thus if $p_j - p_i > 0$, the loss value is about 1. Otherwise, it approximates to zero.

Model Learning. We apply the power iteration method to solve the optimization problem in Eq.(4) [17]. Specifically, we write the derivatives of each parameter of θ as:

$$\begin{aligned} \frac{\partial L}{\partial \alpha} &= \sum_{i,j} \frac{\partial h(\delta_{ij})}{\partial \delta_{ij}} \left(\frac{\partial p_j}{\partial \alpha} - \frac{\partial p_i}{\partial \alpha}\right), & \frac{\partial L}{\partial \boldsymbol{w}} &= \sum_{i,j} \frac{\partial h(\delta_{ij})}{\partial \delta_{ij}} \left(\frac{\partial p_j}{\partial \boldsymbol{w}} - \frac{\partial p_i}{\partial \boldsymbol{w}}\right) + 2\lambda \boldsymbol{w}, \\ \frac{\partial L}{\partial \beta} &= \sum_{i,j} \frac{\partial h(\delta_{ij})}{\partial \delta_{ij}} \left(\frac{\partial p_j}{\partial \beta} - \frac{\partial p_i}{\partial \beta}\right), & \frac{\partial L}{\partial \boldsymbol{v}} &= \sum_{i,j} \frac{\partial h(\delta_{ij})}{\partial \delta_{ij}} \left(\frac{\partial p_j}{\partial \boldsymbol{v}} - \frac{\partial p_i}{\partial \boldsymbol{v}}\right) + 2\lambda \boldsymbol{v}. \end{aligned} \tag{6}$$

According to Eq. (1) of the predicted adoption rate, we have:

$$\begin{aligned} \frac{\partial p_i}{\partial \alpha} &= -\sum_{j \in F_i} t_{ji}[(1-\beta)p_j + \beta u_{ji}] + (1-\alpha) \sum_{j \in F_i} t_{ji}(1-\beta)\frac{\partial p_j}{\partial \alpha} + b_i, \\ \frac{\partial p_i}{\partial \beta} &= (1-\alpha)[\sum_{j \in F_i} t_{ji}[-p_j + (1-\beta)\frac{\partial p_j}{\partial \beta} + u_{ji}]], \qquad \frac{\partial p_i}{\partial \boldsymbol{w}} = (1-\alpha)[\sum_{j \in F_i} t_{ji}[(1-\beta)\frac{\partial p_j}{\partial \boldsymbol{w}} + \beta \frac{\partial u_{ji}}{\partial \boldsymbol{w}}]], \\ \frac{\partial p_i}{\partial \boldsymbol{v}} &= (1-\alpha) \sum_{j \in F_i} t_{ji}[(1-\beta)\frac{\partial p_j}{\partial \boldsymbol{v}}] + \alpha \frac{\partial u_i}{\partial \boldsymbol{v}}. \end{aligned} \tag{7}$$

Now it is easy to determine the remaining derivative of $\frac{\partial u_{ji}}{\partial \boldsymbol{w}}$ and $\frac{\partial b_i}{\partial \boldsymbol{v}}$:

Algorithm 1. Parameter Learning Process for the Proposed SHIP Model

Initialize $\alpha^0 = \beta^0 = 0.5$, $\boldsymbol{w}$, $\boldsymbol{v}$ and $\mathbf{p^0}$ with small positive values;
for $k = 1; k \leq K; k++$ **do**
  //Part 1:Given $\theta^{(k-1)} = [\alpha^{(k-1)}, \beta^{(k-1)}, \boldsymbol{w}^{(k-1)}, \boldsymbol{v}^{(k-1)}]$, calculate $\mathbf{p}^k$;
  while *not converged for* $\mathbf{p}^k$ **do**
    for *each user* $i \in U$ **do**
      Calculate p_i^k based on Eq.(1);
  //Part 2: Given $\mathbf{p}^k$, calculate the following equations;
  while *not converged* **do**
    for *each user* $i \in U$ **do**
      compute the equations in (7);
  //Part 3: Given $\mathbf{p}^k$, calculate θ^k based on Eq.(6) ;
  $\alpha^k = \alpha^{(k-1)} - step\_size * \frac{\partial^k L}{\partial \alpha}$, $\boldsymbol{w}^{\boldsymbol{k}} = \boldsymbol{w}^{(\boldsymbol{k-1})} - step\_size * \frac{\partial^k L}{\partial \boldsymbol{w}}$;
  $\beta^k = \beta^{(k-1)} - step\_size * \frac{\partial^k L}{\partial \beta}$, $\boldsymbol{v}^{\boldsymbol{k}} = \boldsymbol{v}^{(\boldsymbol{k-1})} - step\_size * \frac{\partial^k L}{\partial \boldsymbol{v}}$;
Return $\mathbf{p}^K, \theta^K = [\alpha^K, \beta^K, \boldsymbol{w}^K, \boldsymbol{v}^K]$;

$$\frac{\partial u_{ji}}{\partial \boldsymbol{w}} = \frac{\partial s(\boldsymbol{w} \cdot \boldsymbol{e_{ji}})}{\partial \boldsymbol{e_{ji}}} \boldsymbol{e_{ji}}, \qquad \frac{\partial b_i}{\partial \boldsymbol{v}} = \frac{\partial s(\boldsymbol{w} \cdot \mathbf{x_i})}{\partial (\boldsymbol{w} \cdot \mathbf{x_i})} \mathbf{x_i} \tag{8}$$

Convergence Analysis. Algorithm 1 shows the entire optimization process of our proposed model. There are two power iterations as shown in Part 1 (Eq. (1)) and Part 2 (Eq. (7)) of the algorithm. For all of these equations, they could be rewritten as a unified form as $\mathbf{z_i} = (1-d)\sum_{k \in F_i} t_{ki}\mathbf{z_k} + d\mathbf{y}$. This unified representation defines a linear system problem and its closed form is $\mathbf{Z} = \alpha(\mathbf{I} - (1-\alpha)\mathbf{T})^{-1}\mathbf{Y}$. This closed form satisfies the convergence condition of Gauss-Seidel iterative method [16]. In conclusion, all of the iterations can be solved in linear time with a convergence guarantee.

4 Experiments

4.1 Experimental Settings

We conduct experiments on the collected Weibo data as described in Section 2. We focus on predicting smartphone adoption in 2013Q2 (time t) based on the smartphone status in 2013Q1 (time $t-1$). Given a snapshot of the social network, for each task, we randomly split users into five equal parts and each time we select 80% of the users as labeled users for training and the remaining 20% users are used for prediction. We conduct five-fold cross validation and report the average results. In fact, we only choose the leading four brands (i.e., iPhone, Samsung, Nokia and Xiaomi) in Task 1 for prediction as the remaining brands take less than 1% market share. The detailed data statistics can be found in Table 2. As shown in this table, the data is very unbalanced, for most tasks, the number of negative records is much larger than that of the positive records.

Table 2. Dataset statistics of the two tasks. #P: the number of labeled positive users, #N: the number of labeled negative users (#P+#N=#users). $P\_ratio = \frac{\#P}{\#P+\#N}$.

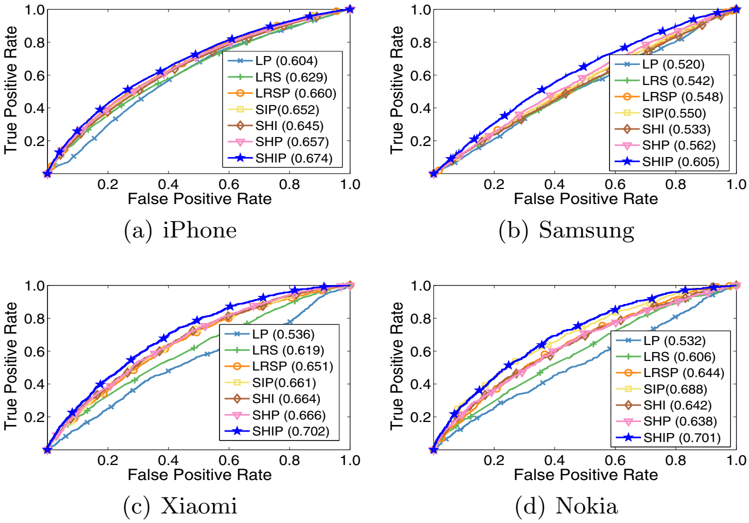
Task	Users	Edges	Brand	#P	#N	P_ratio
Task 1	12,306	125,876	iPhone	5,152	7,154	41.9%
			Samsung	854	11,452	6.94%
			Xiaomi	458	11,848	3.72%
			Nokia	313	11,993	2.54%
Task 2	144,567	15,829,075	/	22,192	122,374	15.35%

For the evaluation, we first use the AUC (i.e., Area Under the ROC Curve) measure, which is especially useful for evaluating the performance of an unbalanced dataset [27]. A random guess would result in an AUC value around 0.5 and the larger the value the better the performance. In addition, as we focus on the most likely positive users of the test data, which can be used for marketing. We measure the relative gain of the precision as $Rel\_gain@N = \frac{Pre@N}{P\_ratio} - 1 = \frac{\#hits}{N*P\_ratio} - 1$. This measure evaluates how the proposed models improve the precision compared to random guess. A random guess will lead to a Rel_gain@N result of 0.0 and the larger the value the better the performance.

Table 3. Summarization of different kinds of features

Type	Feature Description	Type	Feature Description
Social	# of followers that have positive labels in $t-1$	Profile	gender, location
			is this user a verified account
	# of followers that have negative labels in $t-1$		#followers, #followees, #friends
	# of friends that have positive labels in $t-1$		#posts that the user sent in $t-1$
	# of friends that have negative labels in $t-1$		#posts that the user sent in $t-1$
Edge	whether they are friends	Brand	
	# of co-followers that have positive labels in $t-1$		the brand the user used in
	# of co-friends that have positive labels in $t-1$		$t-1$ (only available in Task
	# of co-followers that have negative labels in $t-1$		2)
	# of co-friends that have negative labels in $t-1$		

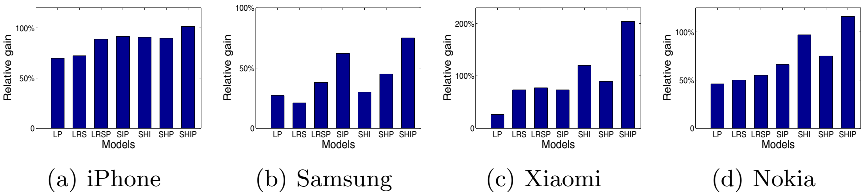
Baselines. To the best of our knowledge, few researchers have tried to explore the smartphone adoption problem with real world collected data. However, we can borrow several classic models that are widely used for the binary class prediction task in a social network: the first category builds classifiers using the extracted graph information as features, and the second category directly propagates the existing labels via random walks in this graph [5]. In the first kind, we choose the logistic regression (LR) model. Specifically, we implemented the LRS baseline that purely relies on Social network features and the LRSP baseline which uses both Social network features and the user Profiles. For the second kind, we choose label propagation (LP), which performs node class prediction based on partially labeled data in a graph [31]. Specifically, LP can be seen as an unsupervised version of our model that only utilizes the graph structure information, i.e., the social influence factor in our model.



(a) iPhone (b) Samsung

(c) Xiaomi (d) Nokia

Fig. 2. Comparison of the AUC results of different models for Task 1



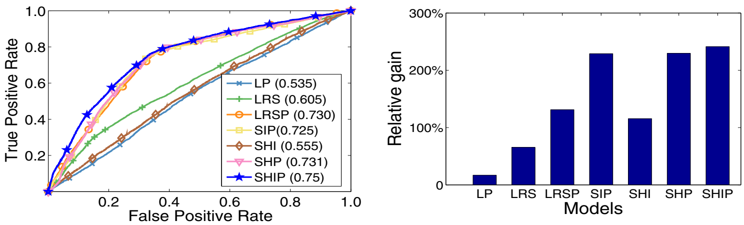
(a) iPhone (b) Samsung (c) Xiaomi (d) Nokia

Fig. 3. Comparison of the Rel_gain@100 results of different models for Task 1

Also, to demonstrate the fitness of the three proposed factors in our model, we compare the proposed SHIP with three related models: SHI (Supervised Homophily-Influence), SHP (Supervised Homophily-Personlity) and SIP (Supervised Influence-Personality). Please note that the simplified SIP model can be seen as a superior version of the work proposed in [28], which can automatically learn the relative importance parameter α between influence and personality. In both LR and our proposed models, we have the regularization parameter λ. As the dataset is very large, choosing λ in a reasonable range (e.g., [0.01, 100]) has little impact on the final prediction results. For the remaining experiments, we empirically set $\lambda = 10$. We summarize the profile, edge, and social features we used in this paper in Table 3.

4.2 Experimental Results

Overall Performance. Task 1 focuses on predicting whether a user will buy a particular brand as first-time buying behavior. Fig. 2 reports the AUC results of different models, where each brand's prediction result is shown in the sub figure and the detailed AUC value is followed by each method in the legend. First, we observe that all models have better performance than a random guess.(i.e., an AUC value of 0.5) Among them, our proposed SHIP model is better than all baselines with regard to all brands, followed by the three related models (i.e., SIP, SHI and SHP), indicating the superiority of our proposed model and the importance of combining the three key factors together for predicting smartphone adoption. Although the overall trend is the same, the detailed AUC results vary. Among all brands, "whether to choose a Samsung for a first-time buying" is the hardest to predict and the best AUC result is only 0.605. One possible reason is that Samsung has too many device types, ranging from high-end smartphones that compete with the iPhone to entry-level smartphones. The reasons why people buy Samsung smartphones vary and are harder to predict. For the other brands, the AUC reaches about 0.7 for SHIP. The average improvement is 3% to 10% over LRSP and 15% to 30% for the remaining baselines. Similar trend can be found for the Rel_gain@100 comparison as shown in Fig. 3. Based on the above analysis, we conclude that the proposed SHIP can help better capture the decision process for first-time buying behavior, thus generating better results than other baselines and related models for Task 1.



(a) AUC (b) Rel_gain@100

Fig. 4. Overall comparison of Task 2: brand change prediction

In Task 2, we predict whether users will change brand in the next time period. Fig. 4 reports both the AUC and the Rel_gain@100 for different models in Task 2. The overall trends are the same as Task 1. For both metrics, SHIP performs the best, followed by our two related models (i.e., SIP and SHP) and LRSP baseline. However, the related SHI model and the LRS baseline, which do not consider the user preference factor, perform badly. In other words, after adding the user preference factor (i.e., the user profile features as shown in Table 3 and the brand feature), the performance improvement is very significant. E.g., the improvement of LRSP over LRS is 20.66% for AUC and 100% for Rel_gain@100, the improvement of SHIP over SHI is 35% for AUC and more than 100% for Rel_gain@100. Why is the improvement so significant after adding the user preference factor? We leave the explanations for the next section.

Table 4. The learned relative weight of each factor

Factor	Weight Representation	Task 1				Task 2
		iPhone	Samsung	Xiaomi	Nokia	
Influence	$(1-\alpha)\times(1-\beta)$	0.822	0.644	0.540	0.692	0.157
Homphily	$(1-\alpha)\times\beta$	0.081	0.138	0.181	0.219	0.42
Personality	α	0.097	0.218	0.278	0.089	0.423

Impacts of the Parameters. As shown in Eq. (1), α and β are two important parameters that control the relative effects of the three key factors for decision-making. We summarize the learned relative weight of each factor of the two tasks in Table 4. As shown in this table, in Task 1, the personality effect (α) and the homophily effect (i.e., $(1-\alpha)*\beta$) for all brands are very small while the relative contribution of social influence for all brands is larger than 50%. That is to say, users are easily influenced by social neighbors for first-time buying behavior. In contrast to this, the social influence effect is very small in Task 2 (i.e., 15.7%) while there is a high impact of the personality factor and the homophily effect for brand change behavior. In other words, users are not easily influenced by social neighbors for changing brand. Their brand change behavior is more liked caused by the homophily effect and their own preferences.

Table 5. Part of $\boldsymbol{v}$ in Task 2: the weight of the feature "the brand the user used in $t-1$"

Brand	iPhone	Samsung	Xiaomi	Nokia
Weight	-0.555*	-0.269*	-0.222*	0.266*

*Pass the T test at the confidence level of 0.005.

For brand manufactures, they would like to explore the inherent reasons that may prevent customer loss, i.e., the brand change behavior in Task 2. As explained before, after adding the user preference factor, the performance improvement is prominent in Task 2. Also, the user personality effect contributes more than 40% to brand change behavior. So we will focus on the user personality effect of Task 2 in this section. Specifically, each dimension of parameter $\boldsymbol{v}$ controls the importance of the corresponding user related feature for smartphone adoption (Eq.(3)). The larger the absolute value of this dimension, the greater the corresponding feature weights for smartphone adoption. In Task 2, we have two kinds of user personality features: user profile features and the brand feature, which describes the brand a user used in $t-1$. To our surprise, all profile features' weights are around 0 and the weight of the brand feature dominates. Next, we try to use this brand feature only in the logistic regression model and the AUC reaches 0.7229, while LRSP's result is 0.7300. The improvement is less than 1% when adding so many user profiles and social features, which also indicates social neighbors' smartphone adoption status does not have a large impact on users' choices of changing smartphones. Thus, we argue, the most prominent factor that determines whether a user will change brands later is the current brand she/he uses. A user's decision on whether to change brands follows the overall brand loyalty. If most people that uses a particular brand in the current time period are likely to change brand in the next time, then this user is also likely to change without a discussion. Table 5 shows the learned weights of the brand feature in Task 2. Among all the listed brands, iPhone users are most loyal. They do not like to change to another brand in the next time period, followed by Samsung and Xiaomi.

5 Related Work

Smartphone usage mining has attracted considerable attentions due to the rapid growth of the smartphone market in recent years. Some researchers revealed the correlation between mobile phone usage and user profiles [12,20,21]. Others attempted to consider the factors that affect people's choices when adopting a mobile device from various perspectives, such as culture [23], technology needs [22] and perceived usefulness [18]. Among them, Harsha et al. found compelling evidence of social influence in the purchase of mobile phones by sample surveys from Asian countries [8]. However, nearly all these works relied on small-scale questionnaires without considering the smartphone adoptions in a large-scale social network.

Our work is closely related to the problem of production adoption prediction in social networks. Generally, some models purely utilized user's profiles in social networks for product adoption prediction [29,30]. Others further incorporated the aggregated features extracted from social networks to boost product adoption performance [6,9]. However, the global product diffusion process among linked users is rarely analyzed, not to mention distinguishing the relative contributions of each factor. While the importance of distinguishing various factors underlining people's correlated decisions in social networks has been well recognized, the related work mainly focused on the homophily factor and the social influence factor that lead to correlated user behaviors [2,14]. The proposed solutions either estimated the upper bound of each factor or needed additional group information of users. On the contrary, our proposed model explicitly balances the correlated user behaviors and each user's own preference. Also, the relative performance of each factor can be learned automatically in the training process.

Our proposed model is also related to the node classification task, i.e., predict the classes of unlabeled nodes with partially labeled nodes in this graph [5,13,26]. A basic assumption of these models is the label correlations in the network, thus we can propagate the labels with respect to the intrinsic graph structure [24,31]. To leverage both the social network structure and the edge features, in recent years, [3] first proposed a supervised random walk algorithm that guides label propagation, where the social influence factor is explicitly modeled. Zeng et al. [28] further extended the supervised random walk model for user affiliation prediction by incorporating both the social influence and the user bias factors. Nevertheless, the homophily effect between linked users was neglected by all these works. Thus our model can be seen as generalizing the recent advances of these related methods in node classification tasks. Moreover, in contrast to these previous approaches, our proposed model can automatically learn the relative effect of each factor while others needed to tune the parameters manually.

6 Conclusion

In this paper, we have proposed a SHIP model for predicting smartphone adoption in a social network. Our model identified the three key factors in the decision-making process and can automatically distinguish the relative contributions of each factor. Experimental results on a large-scale dataset showed the strong prediction power of our model. An incisive conclusion is that the potential first-time smartphone buyers are largely influenced by social neighbors' choices while a user's decision on whether to change to another brand follows overall brand loyalty. In fact, the proposed model is also generally applicable to other node classification tasks. In the future, we plan to apply our model to other smartphone markets, and we will study the adoption in a finer granularity of time periods.

Acknowledgments. This research was partially supported by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010) and the National High Technology Research and Development Program of China (Grant No. 2014AA015203).

References

1. Idc's forcast research. http://www.idc.com/getdoc.jsp?containerId=prUS24314413
2. Aral, S., Muchnik, L., Sundararajan, A.: Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. PNAS **106**(51), 21544–21549 (2009)
3. Backstrom, L., Leskovec, J.: Supervised random walks: predicting and recommending links in social networks. In: WSDM, pp. 635–644. ACM (2011)
4. Baum, W.M.: Understanding behaviorism: Science, behavior, and culture. HarperCollins College Publishers (1994)
5. Bhagat, S., Cormode, G., Muthukrishnan, S.: Node classification in social networks. In: Social Network Data Analytics, pp. 115–148. Springer (2011)
6. Bhatt, R., Chaoji, V., Parekh, R.: Predicting product adoption in large-scale social networks. In: CIKM, pp. 1039–1048. ACM (2010)
7. Boe, B.J., Hamrick, J.M., Aarant, M.L.: System and method for profiling customers for targeted marketing (2001). US Patent 6,236,975
8. De Silva, H., Ratnadiwakara, D., Zainudeen, A.: Social influence in mobile phone adoption: Evidence from the bottom of the pyramid in emerging asia. Information Technologies & International Development **7**(3) (2011)
9. Guo, S., Wang, M., Leskovec, J.: The role of social networks in online shopping: information passing, price of trust, and consumer choice. In: EC, pp. 157–166. ACM (2011)

10. Iyer, G., Soberman, D., Villas-Boas, J.M.: The targeting of advertising. Marketing Science **24**(3), 461–476 (2005)
11. Kleinberg, J.: Cascading behavior in networks: Algorithmic and economic issues. Algorithmic Game Theory **24**, 613–632 (2007)
12. Lins, L., Klosowski, J.T., Scheidegger, C.: Nanocubes for real-time exploration of spatiotemporal datasets. IEEE Trans. VCG **19**(12), 2456–2465 (2013)
13. London, B., Getoor, L.: Collective classification of network data. Data Classification: Algorithms and Applications, 399 (2014)
14. Ma, L., Krishnan, R., Montgomery, A.: Homophily or influence? an empirical analysis of purchase within a social network (2010)
15. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology, 415–444 (2001)
16. Meijerink, J.A., van der Vorst, H.A.: An iterative solution method for linear systems of which the coefficient matrix is a symmetric m-matrix. Mathematics of Computation **31**(137), 148–162 (1977)
17. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web (1999)
18. Park, Y., Chen, J.V.: Acceptance and adoption of the innovative use of smartphone. Industrial Management & Data Systems **107**(9), 1349–1365 (2007)
19. Skinner, B.F.: Science and human behavior. Simon and Schuster (1953)
20. Smith, A.: 46% of american adults are smartphone owners. Pew Internet & American Life Project (2012)
21. Smith, A.: Smartphone ownership-2013 update. Pew Internet & American Life Project (2013)
22. Van Biljon, J., Kotzé, P.: Modelling the factors that influence mobile phone adoption. In: SAICSIT, pp. 152–161. ACM (2007)
23. Van Biljon, J., Kotzé, P.: Cultural factors in a mobile phone adoption and usage model. Journal of Universal Computer Science **14**(16), 2650–2679 (2008)
24. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. IEEE Trans. KDE **20**(1), 55–67 (2008)
25. Xiang, B., Liu, Q., Chen, E., Xiong, H., Zheng, Y., Yang, Y.: Pagerank with priors: an influence propagation perspective. In: IJCAI, pp. 2740–2746. AAAI Press (2013)
26. Xu, H., Yang, Y., Wang, L., Liu, W.: Node classification in social network via a factor graph model. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds.) PAKDD 2013, Part I. LNCS, vol. 7818, pp. 213–224. Springer, Heidelberg (2013)
27. Yan, L., Rober, D., Mozer, M.C., Wolniewicz, R.: Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic. In: ICML, pp. 848–855 (2003)
28. Zeng, G., Luo, P., Chen, E., Wang, M.: From social user activities to people affiliation. In: ICDM, pp. 1277–1282. IEEE (2013)
29. Zhang, Y., Pennacchiotti, M.: Predicting purchase behaviors from social media. In: WWW, pp. 1521–1532 (2013)
30. Zhang, Y., Pennacchiotti, M.: Recommending branded products from social media. In: Recsys, pp. 77–84. ACM (2013)
31. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. NIPS **16**, 321–328 (2003)