Which phone will you get next: observing trends and predicting the choice

Yi Wang*, Hui Zang†, Pravallika Devineni‡, Michalis Faloutsos‡, Krishna Janakiraman† and Sara Motahari†
* Bourns College of Engineering, UC Riverside, USA  
Email: wangyi@cs.ucr.edu  
† Sprint, Burlingame, USA  
Email: huizang@gmail.com, {Janakiraman, Sara-Gatmir-motahari}@sprint.com  
‡Department of Computer Science, UNM, Albuquerque, USA  
Email: {pdevineni, michalis}@cs.unm.edu

Abstract—As the smartphone/cellphone market has exploded, the war on which smartphone platform will dominate has become fiercer than ever. In that vain, the goal of this paper is to answer two fundamental questions: What are the adoption trends for smartphones? And how can we estimate the demand for new smartphones? We answer these two questions by collecting a dataset of 3 million subscribers from a nationwide telecom operator. A key aspect of our work is that we have demographic information per user, such as income level, and age, which we correlate with phone usage patterns. Interestingly, we find that in all demographic groups, Android is leading platform top in all age groups and income levels. A key question is whether the “social influence” affects the choice of phone, which we find more pronounced in business plans. Finally, we develop a predictor to infer the phone a user will switch to considering: (a) the type of previous phone, (b) the social influence, and (c) the demographics of the user. Compared with the reference method, our predictor is effective in: (a) reducing the prediction error in number of phones by 1/3, and (b) in the case of minimizing phone costs, the monetary cost by half. Apart from its interest in observations, our work could help telecom operator forecast their inventory more accurately by pointing to the right properties to consider.

I. INTRODUCTION

What is the next phone a user will get? This is a critical question for the exploding smartphone market. In recent years, smartphone market has been driven by the mobile application market and mobile commerce. Estimates state that there are now more than 1.5 billion smartphones in the world and that 81% of U.S. mobile subscribers have 3G/4G subscriptions. Another study shows that the average life of a smartphone is 18 months. As people change phones frequently, it is critical for telecom operators to study the evolution and changing trends of smartphones and understand the factors that determine phone choice.

The problem we address here is how to predict the choice of next phone for users. Specifically, we focus on the following problem: given the properties of subscribers, such as demographics and previous phone choices, how can we predict which phone they will pick next? The input of the problem is a group of $k$ subscribers intending to switch phones and their properties, the output is different options (in terms of number of phones) available for these subscribers. In order to address this problem, we decompose the problem into two questions: 1) What are the trends in switching phones, and are they correlated to user properties, such previous phone or demographics? 2) Based on these trends, how can we predict the demand for new phones? Clearly, aggressive marketing campaigns, new phone releases and applications available on the phone can affect user preference, but this is not something that we incorporate in our problem formulation, as it is difficult to even assess the effectiveness of a marketing campaign for a new phone.

Despite its importance, predicting phone preference has not received a lot of attention in the research community. Previous efforts have studied the correlation between phone functionality and personal preference[1][2]. However, these early efforts have two limitations. First, some as are dated back to 2007, which is largely obsolete in the fast evolving world of smartphone technology. Second, some efforts focus on the transition of users towards smartphones. Finally, some studies rely on surveys that ask people about their preferences and not on actual phone adoption data. We discuss related work in more detail in section V.

Our work is partly motivated by the need for telecom operators to anticipate the demand for new phones. In fact, the work was proposed by the industry collaborators in our team as a problem of interest. Currently, it is difficult to predict what phone a subscriber may use next and to determine factors on which we can base that prediction. Inaccurately predicting demand for phones could lead to financial loss, especially when ordering expensive smartphones, which cost more than ordinary cellphones. An obvious baseline solution is to use historical trends, in other words, what customers wanted 6 months ago, and project that to the future. Finally, from a scientific point of view, it is interesting to see which factors affect phone preference.

The contribution of our work is two-fold. First, we study phone usage trends and identify user behaviors and properties that correlate with the choice of the next phone. Second, we develop a predictor, based on Bayesian Networks, to infer the phone a user will switch to considering: (a) type of previous phone, (b) social influence, and (c) demographics of the user. We conduct an extensive study on a nationwide cellular network dataset with 3 million subscribers. For each subscriber,
we know his/her age group, income level, previously used phones and the phone preference of his/her friends: (1) who are on the same bill plan, and (2) whom they frequently communicate with. We identify and study the actual phone switches that some of those subscribers do during the time of observation. Our contributions can be summarized as follows:

(1) The Android platform seems to be winning in every demographic group across all age groups and income levels.

(2) Demographics and information about the previous phone is strongly correlated with the phone one uses. For example, 90% of subscribers who are both young and have high income prefer to use smartphones, while only 65% of subscribers who are both old and have low income have smartphones.

(3) Social influence affects phone choice for subscribers but mostly on business plans (i.e. business phones provided by companies to their employees). We find that 75% of business plan subscribers use the same phone as their peers.

(4) We develop a predictor using Bayesian Networks to infer which phone users will prefer when they switch phones.

We formulate two problems: (a) minimizing the number of wrong guesses in terms of the absolute number of phones, and (b) minimizing the cost of ordering phones erroneously. Compared with a baseline method, which relies on historical trends, our approach reduces the prediction errors by 1/3 in terms of phone numbers and can reduce the monetary cost by half.

The scope of this work is mostly to shed light on the possibilities for improving prediction accuracy. First, we want to show that several user properties are correlated with phone preference. Second, our work suggests that if we consider these properties carefully, we can predict more accurately the next phone that users will want.

II. Dataset and Features

We give an overview of our dataset, define features and discuss its possible limitations to the result of this paper.

A. Dataset introduction

**Phone-switching history.** Our dataset consists of over 3 million subscribers who switched phones from January to June, 2012. We know the demographic properties of subscribers from the third party, including age group, income level, their previous and current phones and the lifetime of their previous phones. Of those subscribers, 90% use only one previous phone while 10% have more than one phone. Table I shows a concrete example where the subscriber 49 is young, his income level is Medium and he used Samsung Galaxy III from 07/01/2011 to 01/31/2012 and switched to iPhone 4s on 02/01/2012.

**Defining phone types.** There are more than 200 types of phones in the dataset. We cluster these phones based on their operating system (OS) into six groups: (a) iOS, (b) Android, (c) Blackberry, (d) Palm, (e) Windows and (f) feature phones. The feature phone, which we sometimes refer to as “cellphone” group whose operating systems is of none the previous five, and include mostly non-smartphones. Note that in the rest of the paper, when we mention the term “phone”, we do not mean a particular phone but the name of the group the phone belongs to. In Table I, Phone 1 is an Android phone because Samsung Galaxy III is installed with Android OS and Phone 2 is an iOS phone.

**Call Data Records and bill plan.** First, we collect Call Data Records (CDRs) generated by the 3 million subscribers in March 2012 represented as CDR-Mar. This dataset contains all their communication records and in each record, there are a caller, a callee, the direction of calling, etc. Using these records, we construct a call graph where nodes are subscribers and edges represent two subscribers calling each other at least once.

We use the term a **call friend** to refer to a 1-degree neighbor on the call graph. We also have the information regarding bill plans, that is phones that are on the same plan (e.g. family or a business plan) for the 3 million subscribers in June 2012 represented as BILL-Jun. We use the term a **plan friend** to refer to subscribers who are on a cellular plan with the same bill.

B. Definition of features

The goal of this paper is to infer the phone choice of subscribers. Here, we consider the following features that we will evaluate to see if they are useful for the inference.

**Demographic properties.** For a subscriber, we have two types of demographic properties: age group and income level. We have four age groups: young(18-30), middle(31-45), middle old(46-65) and old(>66) which make up 17.8%, 43.8%, 29.5% and 8.6% of the whole population. We have three income levels: low, medium and high which constitute 33.1%, 33.3% and 33.6%, respectively.

**Previous phone preference.** We consider two features regarding the phone that was used before a switch: (a) the type of phone used before the change, which we denote by $PH_{prev}$ and (b) the lifetime of the phone, which we denote by $Len$. In our dataset of people that switch phones, the distribution among the six types of $PH_{prev}$ is: Android: 42.5%, Cell: 39.7%, Blackberry: 14.8%, iOS: 2.1%, Palm: 0.4% and Windows: 0.4%. The length, $Len$, is the duration from the start day to the end day pertaining to the use of $PH_{prev}$ in terms of months. Its Probability Density Function (PDF) distribution is shown in Figure 2. In the Table I, $PH_{prev}$
is “Android” and its \textit{Len} is 7 months. We evaluate if these features can help us predict the choice of the new phone.

**Social features.** We are interested in understanding how social influence affects phone choices. Recall that we defined earlier two kinds of social circles: call friends and plan friends. Let us consider the call friends first. Among that group, there is a dominant phone type, which is used by the majority of the call friends, and which we denote as call-neighborhood majority phone $PH_{call}$. We refer to the people that use the call-neighborhood majority phone as call influence neighborhood. For example, if 6 out of 10 call-neighborhood friends have a Blackberry, then the size of the call influence neighborhood is 6 and the call-neighborhood majority phone $PH_{call}$ is a ‘Blackberry’.

For plan friends, we can have a similar definition. We use the term plan phone $PH_{plan}$ to refer to the phone that is used by the majority of the plan friends. We define the plan influence neighborhood to be the plan friends with the majority phone. Let us consider a real example from our data. Subscriber 49 switches phone on 02/01/2012 and his $PH_{plan}$ is the most popular phone used by his plan friend on 01/31/2012. We divide plans into two categories based on number of subscribers. Plans with more than eight subscribers are business plans and its subscribers are labeled as business subscribers. Plans with less than or equal to eight subscribers are non-business plans, whose subscribers are labeled as non-business subscribers. As we see in Section III, the business subscribers seem to be more influenced by the majority phone in the plan, compared to non-business subscribers.

**Features of communication.** We also consider communication patterns which include total call frequency of the subscriber, average call duration, total number of short messages sent and total bytes of data consumed. We study the usefulness of features extracted from CDR-Mar in predicting the next phone. However, we found that they were not statistically significant in how they correlated with phone choices, so we do not discuss them further in the paper. This study is omitted due to space limitations.

**Limitation.** Since the dataset provider introduced iPhones only from October 2011, the data collected might not be able to correctly predict iPhone switching patterns. 2.1% of subscribers have an iPhone as their previous phone while for most iPhone users, we cannot know their next phone choices. This relatively small sample could introduce artifacts in the results.

### III. Observations and Trends

We now present a study on: (a) basic trends and statistics of phone usage, and (b) how features and properties correlate with the choice of a new phone.

**A. Basic trends and statistics**

We study basic trends in phone usage such as the distribution of different types of phones, and the frequency with which people change phones.

#### Can your demographics tell your phone preference?

We explore the correlation between the demographics of a subscriber and his/her phone preference. We calculate the conditional probability: $\text{Prob}(PH_{curr}|Age)$, which represents if a subscriber is using the phone $PH_{curr}$ given the age group. We perform similar calculation with income level and show results in Figure 1.

1) Android is the top choice for all age groups and income levels. No matter which group a subscriber belongs to, the probability of using Android phone is always above 40%.

2) For the young group, the probability of using iPhones is twice that of using cellphones. These numbers are about 30% and 15%. For the high income group, the probability of using iPhone is three times that of using cellphones and the two numbers change to 35% and 12%. The high income group is more likely to use iPhones than the young group.

3) The probability of using smartphones for subscribers who are both young and have high income is 90%, while the probability for those who are old and have low income is only 65%.

#### The low income group switches phones more frequently than the high income group.

We plot the probability density distribution of the lifetime of the used phone, denoted by $\text{Len}$, over all subscribers in Figure 2 (left). As we see, there are two peaks in the plot: one located at 12 months and the other at 22 months. The second one can be explained by the likeliness of the subscriber to switch to another phone towards the end of the a two year contract. We are intrigued by the first peak.

We break down the distribution by plotting the versions over various age groups and income levels of subscribers, as we plot
in Figure 2 (right). We find that the trends for various income levels provide useful insight. For low income subscribers, the first peak is more pronounced than the second one. For the high income subscribers, the second peak is as pronounced as the first one. We find that among low income subscribers that keep a phone less than 6 months have about 70% probability that the phone is a feature phone. Going on a limb, we conjecture a possible explanation for this phenomenon. It could be that the low income subscribers tend to use short-term prepaid or pay-as-you-go phones, which may be feature phones, which are often provided for such contracts at really low prices. By contrast, high income subscribers could be less likely to get pay-as-you-go phones, and thus are likely use the same phone longer as per their contract.

B. Subscriber phone switching patterns

True or False: “The longer you use a phone, the more you stick to it?” We explore the correlation between the lifetime of the previous phone, Len and the probability that one uses the same phone by calculating the probability Prob(PH_{curr} = PH_{prev}|Len = l) in Figure 3.

The answer is no, as the probability does not display strictly positive relationships as the Len increases. In Figure 3, the plot has two parts divided by a turning point at 2 years. Before the two year point, the plot keeps flat, showing that there is no obvious relationship between the two variables. After the point, the plot rises as the duration increases. We find that 80% of subscribers using a phone more than 2 years are business subscribers. One possible reason could be that they require some specific functionality of the phone or need to follow the company policy, so they continue using that phone for a long time.

Is the previous phone correlated to the choice of the next phone? In order to answer this question, we calculate the probability that a subscriber uses a specific phone given the type of his/her latest used phone, which is denoted as Prob(PH_{curr}|PH_{prev}). The result is shown in Table III.

Android phones dominate as the probability, Prob(PH_{curr} = Android|PH_{prev}) is about 45% for both Android and any other type of previous phone. Naturally, iOS phones are the second most dominant phones. In particular, subscribers previously using the Palm and Blackberry phones are equally likely to choose iOS and

Android as their next phone. Also, 90% of smartphone users continue to prefer smartphones. In fact, the interesting group here is the 10% of smartphone users that opt to go for feature phones, possibly to reduce associated costs with expensive devices and data plans. At the same time, 64% of cellphone users switch to smartphones.

C. The effect of Social Influence

How does social influence affect phone switching? We are interested in studying the affect of social influence on a subscriber’s phone preference. In other words, we want to see whether one adopts their friend’s choice when choosing their next phone. Here we answer this question by checking the influence of plan and call friends on the subscriber as defined in the Section III. To determine this, we calculate the probability of PH_{plan}=PH_{curr} and PH_{call}=PH_{curr}.

We plot the average probability Prob(PH_{plan}=PH_{curr}) over subscribers with influence neighborhood sizes ranging from 2 to 20 in Figure 4. When the influence neighborhood size is more than 8, we observe a strong positive correlation between the probability and size of the neighborhood. This shows that it is highly probable for subscribers on a business plan to choose the same phone as their friends. Subscribers whose plan influence neighborhood size is about 15 have greater than 65% chance that they have the same phone as their plan friends. Of all the business subscribers, 75% use the same phone as their plan phones. In contrast, subscribers on a non-business plan will not be affected by the influence of their plan friends. When the neighborhood size is less than 8, the plot fluctuates and does not display a rising trend. Only 50% of non-business subscribers use the same phone as PH_{plan}.
For problem 1, we select phone \( j \) with the maximum \( p_{ij} \) as the output. For problem 2, given the probability of phone choices of \( k \) subscribers, the sum of probabilities \( p_{ij} \) over \( k \) subscribers is the predicted number for a specific phone \( j \), denoted as \( NP_j \):

\[
NP_j = \sum_{i=1}^{k} p_{ij}
\]

where \( j \in \{\text{iOS}, \text{Android}, \text{Plam}, \text{Blackberry}, \text{Windows}, \text{cell-phone}\} \).

We calculate \( NP_j \) over six types of phones as the output.

**Predicting individual choice is hard.** Our initial results suggest that it is hard to achieve high accuracy irrespective of the algorithm employed. The prediction accuracy is roughly around 60%. Intrigued, we investigated the source of the errors by analyzing the confusion matrix. We find that major portion in the misclassification is contributed by iOS to Android and vice versa.

Given that telecommunications operators order in bulk for their subscribers and the inaccurate prediction of iOS to Android and vice versa cancels each other out. In the rest of this section, we focus only on Problem 2.

**Formulations:** We identify two formulations for the second problem: minimizing the error in the predicted number and minimizing the cost in of the mis-predicted phones.

**A. Absolute Difference:** For a phone \( j \), it is the difference between the predicted number \( NP_j \) and the actual number \( N_j \).

\[
AD_j = |N_j - NP_j|, \quad AD = \sum_j AD_j
\]

Note that AD without \( j \) represents the summation of ADs over six types of phones.

**B. Absolute Cost Difference:** The cost of inaccurately predicting an iPhone is much higher than that of a cellphone. Such costs are not considered by AD, so we propose another metric called Absolute Cost Difference. ACD takes the difference among phone costs into account and uses a weight based on \( COST \) to calculate the difference.

\[
ACD_j = |N_j - NP_j| \times COST_j, \quad ACD = \sum_j ACD_j
\]

where \( COST_j \) is over-estimated cost, \( O - COST \) when \( N_j > NP_j \). Otherwise, \( COST_j \) is under-estimated cost, \( U - COST \).

**TABLE IV**

| UNDER- AND OVER-ESTIMATED COSTS OF 6 TYPES OF PHONES |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                 | iOS            | Android        | Blackberry     | Windows        | Palm           | Cell           |
| \( U - COST \)  | 10             | 10             | 10             | 10             | 10             | 10             |
| \( O - COST \)  | 200            | 150            | 100            | 150            | 100            | 50             |

Table IV lists the over-estimated and under-estimated costs of six kinds of phones. The numbers in Table IV are inputs from domain experts. \( O - COST \) is set according to the price of a phone. If a phone ordered by the operator is not consumed by the subscriber, the operator cannot return it back to the manufacturer. Thus, the cost is roughly proportional to the price of a phone. For the under-estimated cost, it is a constant. If the operator finds no phone matching subscriber’s demands in store, it can be ordered later from the manufacturer and the cost is proportional to the delivering cost, which is set to a
fixed value in this paper.

**Baseline Algorithm.** We employ an algorithm based on Historical Trend Analysis, denoted by Histro, as our baseline. If x% of subscribers in the training dataset, TRAIN have $PH_{\text{current}} = \text{Android}$, Histro reflects the same percentage over s subscribers in testing dataset, TEST by multiplying s with x%. The product is the predicted number of Android users in TEST. We have been told that this is not very far from common business practices in predicting inventory. To the best of our knowledge, no one did the similar work before us, so we decide to use Histro as the baseline algorithm

**Results:**

The Bayesian Network approach reduces Absolute Difference (AD) by 30% compared to Histro. The absolute difference for Bayesian Network is around 120K while it is 180K in Histro. From Figure 5, it is clear that Bayesian Network performs better than Histro in terms of aggregated result. In Table V, the AD in Bayesian Network for Android and Cellphone is about 2 and 1.5 times lower than in Histro.

**Adjustment with a cost matrix** The second problem formulation tries to minimize the Absolute Cost Difference, instead of focusing on reducing the Absolute Difference. In order to meet this requirement, we construct a cost matrix, which is defined as follows:

$$c_{ij} = (O - COST_j)/(O - COST_i)$$

where $c_{ij}$ is the cost of misprediction of phone i with another phone j.

If the $O - COST_j > O - COST_i$, the penalty cost $c_{ij}$ is higher because ordering a more expensive phone that is not consumed by a subscriber means higher financial loss. With this matrix, we anticipate that the inaccuracy in predicting the number of smartphones, especially iPhones, will be reduced.

The Bayesian Network approach with a cost matrix (Baye+Mat) reduces the Absolute Cost Difference by 50% compared to Histro. The results of cost matrix applied on the top of Bayesian Network are shown in Table V. Comparing with the output of Bayesian Network without the cost matrix, the inaccuracy in iOS and Android phones reduces but increases in feature phones. However, the cost of inaccurately predicting a feature phone is much lower than that of a smartphone and the total Absolute Cost Difference is reduced by 1/2 (See Figure 5).

**Summary.** Our work leads to the following conclusions.

First, the performance of Histro is typically worst among all the prediction methods regarding the absolute difference and the absolute cost difference. It fails to predict the iOS, Android, Blackberry and non-smart cellphones, but Histro is good at predicting the number of Palm and Windows phones. Second, our Bayesian-Network-based algorithms show better performance than Histro for all the phones except Palm and Windows phones. We believe that this could be a consequence of having too few Palm and Windows phones in our training dataset. Third, our Logistic Regression-based method performs comparably to the Bayesian Networks approach for the absolute difference. As a side note, the time of building its model is much longer than that of training a Bayesian Network. With a sample of 20,000 records for training, the training time of Bayesian Network implemented in the WEKA toolkit was 2 seconds, while that of Logistic Regression was 51 seconds.

Given that the time is in the order of seconds, this may not be an important factor for this size of problems. Finally, we see that one has to formulate the problem of interest, since reducing Absolute Cost Difference is significantly different from Absolute Difference. Clearly, using the right cost matrix is important, but this is an straightforward calculation for a mobile provider, as they know exactly the relative costs of pre-ordering different types of phones.

**V. RELATED WORK**

Predicting the purchase intent of a phone has attracted much attention in the fields of Sociology, Management and Computer Science. Functionality and mobile applications are important features influencing the purchase of a cellphone [1]. Mokhlis et al.[2] found that innovative features such as built-in camera significantly affect personal choice in the smartphone. Park et al. [6] showed the strong correlation between phone adoption and mobile applications in Korea. Haverila [7] shows that business functionality has a significant correlation with repurchase intent.

Most literature so far considers usage functions and functional features of phones such as input style, touch-screen and keyboard. By contrast, our work considers the social demographics of a subscriber, such as age group and income level. Also, we use features like phone-switching history and neighborhood based on call history of subscribers to study the correlation between choice of previous and current phones.

Feature selection plays a key role in the accuracy of prediction. Previous efforts listed below inspire us to use demographics and social network-related features. Okazaki et al. [8] revealed the effects of demographic characteristics, including age, gender, marital status and occupation, on the adoption of mobile contents. Lu et al. [9] found that the social influences are potential but not direct determinants on the adoption of wireless Internet service.

Predicting phone selection-based on previous usage patterns and currently available options is still in its initial stages. Strauts [5] identified factors to predict cell phone and landline usage through a social survey. Stanford [4] mined purchase
TABLE V
RESULTS OF PREDICTIONS BY HISTRO, LOGISTIC REGRESSION (LOGI) AND BAYESIAN NETWORK (BAYE) WITH/WITHOUT A COST MATRIX (MAT). ↓ AND ↑ MEAN UNDER- AND OVER-ESTIMATING.

<table>
<thead>
<tr>
<th></th>
<th>iOS</th>
<th>Android</th>
<th>Blackberry</th>
<th>Windows</th>
<th>Palm</th>
<th>Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>419,297</td>
<td>909,296</td>
<td>93,368</td>
<td>6,454</td>
<td>6,940</td>
<td>384,942</td>
</tr>
<tr>
<td>Histro NP</td>
<td>459,712</td>
<td>959,943</td>
<td>84,973</td>
<td>6,554</td>
<td>4,732</td>
<td>304,853</td>
</tr>
<tr>
<td>Histro AD</td>
<td>40,415</td>
<td>50,647</td>
<td>8,395</td>
<td>100</td>
<td>2,208</td>
<td>80,559</td>
</tr>
<tr>
<td>Logi NP</td>
<td>478,500</td>
<td>895,812</td>
<td>87,800</td>
<td>2,644</td>
<td>3,312</td>
<td>351,229</td>
</tr>
<tr>
<td>Logi AD</td>
<td>50,203</td>
<td>13,484</td>
<td>5,568</td>
<td>3,810</td>
<td>3,628</td>
<td>33,713</td>
</tr>
<tr>
<td>Baye NP</td>
<td>479,900</td>
<td>895,937</td>
<td>87,583</td>
<td>2,747</td>
<td>3,268</td>
<td>350,862</td>
</tr>
<tr>
<td>Baye AD</td>
<td>60,603</td>
<td>13,359</td>
<td>5,785</td>
<td>4,193</td>
<td>3,186</td>
<td>34,080</td>
</tr>
<tr>
<td>Baye+Mat NP</td>
<td>370,975</td>
<td>860,894</td>
<td>121,859</td>
<td>4,777</td>
<td>2,934</td>
<td>458,858</td>
</tr>
<tr>
<td>Baye+Mat AD</td>
<td>48,322</td>
<td>48,402</td>
<td>28,491</td>
<td>2,163</td>
<td>3,520</td>
<td>73,916</td>
</tr>
</tbody>
</table>

history to predict adoption of mobile computing, which is similar to our work in terms of dataset. Both of these works used Logistic Regression for their prediction approaches. Nedevschi [3] proposed use of Bayesian Network for prediction. Here, we consider both Bayesian Networks and Logistic Regression and we also expand the problem to address the absolute cost difference formulation.

VI. CONCLUSION

In this paper, we present a case study about phone preference of subscribers on a dataset provided by the mobile carriers. We conduct an extensive study real data from a nationwide cellular network dataset with 3 million subscribers. First, we study phone usage trends and identify user behaviors and properties that correlate with the choice of the next phone. Second, we develop a predictor to infer the type of the next phone a user will choose. Our predictor considers a several types of information that are largely available to mobile carriers: (a) type of previous phone, (b) social influence, and (c) demographics of the user, phone choices.

From a scientific point of view, our study attempts to explore how predictable is the phone-selection of users. Our study provides a first step in this direction. We identify user properties that are correlated with phone preference, and we provide some algorithmic solutions that seem to be able to predict phone-selection with reasonable success.

REFERENCES