Automatically Detecting Problematic Use of Smartphones

Choonsung Shin
Korea Electronics Technology Institute
1599 Sangam-dong, Mapo-gu, Seoul, South Korea
cshin@keti.re.kr

Anind K. Dey
Carnegie Mellon University, HCI Institute
5000 Forbes, Pittsburgh, USA, PA 15213
anind@cs.cmu.edu

ABSTRACT
Smartphone adoption has increased significantly and, with the increase in smartphone capabilities, this means that users can access the Internet, communicate, and entertain themselves anywhere and anytime. However, there is growing evidence of problematic use of smartphones that impacts both social and heath aspects of users' lives. Currently, assessment of overuse or problematic use depends on one-time, self-reported behavioral information about phone use. Due to the known issues with self-reports in such types of assessments, we explore an automated, objective and repeatable approach for assessing problematic usage. We collect a wide range of phone usage data from smartphones, identify a number of usage features that are relevant to this assessment, and build detection models based on Adaboost with machine learning algorithms automatically detecting problematic use. We found that the number of apps used per day, the ratio of SMSs to calls, the number of event-initiated sessions, the number of apps used per event initiated session, and the length of non-event-initiated sessions are useful for detecting problematic usage. With these, a detection model can identify users with problematic usage with 89.6% accuracy (F-score of .707).

Author Keywords
Health, detection, machine learning

ACM Classification Keywords
H.m. Information Systems: Miscellaneous.

INTRODUCTION
As smartphones are more widely used in daily life and their capabilities are constantly improving, the number of people who use smartphones and the extent of their usage continue to grow as well. According to a recent survey [5], 80% of the world's population has mobile phones and one-third of them own smartphones. In the United States, 43% of adults have smartphones and 62% of those aged 18-34 has smartphones [3]. This means a growing number of people are able to access the Internet and send emails and text messages anytime, and anywhere. In another trend, application usage has expanded due to simple, real-time access to the applications via *AppStores*. Americans, in particular, spend about 2.7 hours per day texting, browsing, and engaging in entertainment, *etc*. on their phones [4]. To put this in perspective, this is about one-third of the average person's period of sleep, distributed over the day. Smartphones provide their users with a wide range of opportunities and conveniences, and have become prevalent in our daily lives.

There is, however, growing concern that easy access to smartphones at any time and in any place may lead to overuse. While many users choose to make themselves unavailable on their smartphones at times [22], recently, a number of news articles have reported that people are increasingly "addicted" to their smartphones, tending to use their phones in inappropriate settings such as the bathroom (40% of smartphone users), bed (54%), during religious services (9%) and while driving (24%) [6,1,2,14]. Although using smartphones in such situations may seem necessary and helpful, these settings may not be appropriate for phone usage, as smartphones can cause people to neglect their surroundings and can have negative impacts on a user's daily life. For example, a simulation-based study showed that using smartphones while driving reduced performance due to a lack of driver attention [13]. Heavy phone users can also encounter social problems since they are frequently interrupted by their phones and unable to pay attention to others. 60% of users do not go one hour without checking their phone for notifications [1]. One study of college students in the UK found that 36% said they could not do without their phones, including 7% who said that their mobile phone had caused them to lose a relationship or job [24]. Therefore, the growing use, and in some cases overuse, of smartphones in our daily lives has the *potential to be quite problematic*, and this issue should be explored.

In spite of potential problems caused by smartphones, there is, as of yet, no agreed upon definition or sufficient study of the problematic use of smartphones. Instead, researchers interested in smartphone *addiction* have leveraged the characteristics of technological or Internet addiction. These have been classified by many as a behavioral addiction [11], as opposed to narcotics, which can cause physical addiction. Jang considered phone use that caused problems in one's physical life to be *mobile phone addiction* [16]. Bianchi and Phillips refer to this as *problematic use of smartphones* and developed and validated a problematic phone use scale based on perceived phone usage (*e.g.*, self-reports on the number of calls per day) [7]. Furthermore, Oulasvirta *et al.* found that mobile phones do seem to encourage "*habitual use*"—in particular, very short and frequent inspections of one's mobile phone throughout the day [20]. Subsequent studies based on perceived phone usage and the problematic phone use scale focused on identifying which behavioral factors contribute to problematic use [7,19]. Through these studies, it is commonly acknowledged that smartphone use can be problematic; thus we adopt the concept of problematic use and take a first step towards automatically detecting it with actual phone usage.

In our paper, we use descriptions from the psychology literature to *define problematic phone use* as including overuse, undesirable use that results in negative consequences in both personal and social aspects of one's life such as using the phone impulsively or without concerns for one's surroundings, and use that leads to feelings of withdrawal or psychological distress when unable to use a phone [7,21,27]. While problematic use is not necessarily considered to be addiction, habitual problematic use has the potential to lead to addiction [27], and deserves further study.

It is important to note that this characterization of problematic use is not meant to pathologize the user's behavior. The scientific community at large is still working to understand peoples' relationship to technology, a difficult task given the pace at which technology is progressing. Regardless, it is important to acknowledge that smartphone use can have negative consequences. This can be demonstrated explicitly when it comes to driving cars and operating heavy machinery, wherein one can harm themselves and others. It is also demonstrated when people use their devices in places where legislation has explicitly prohibited their use, such as planes or hospitals, due to safety concerns. With this acknowledgement, we hope this research works to provide insights into our ever-evolving relationship with our technology.

Previous research on problematic use has depended solely on self-reported phone usage. However, self-reports can be expensive to collect, are prone to inaccuracies and user bias, vary with personal circumstance [9], and provide information about a single moment in time. Furthermore, it is unclear how to use these self-reports and assessments repeatedly in order to understand how problematic use may be changing over time, as repeated exposure to the same assessments can compromise the quality of the responses. Therefore, it is difficult to gain a reliable understanding of how people may be using their phones, and how their usage might be problematic.

In our work, we examine the problematic use of smartphones in a more objective way, by using automatically collected data from the phone itself. Our focus is to build a system that can detect problematic use by gaining a more complete understanding of how people are using their smartphones. Rather than simply relying on perceived phone usage, we propose to collect contextual information from smartphones to gain an objective understanding of the characteristics of problematic usage. Thus, we expect that our system can be used to reliably and automatically detect problematic smartphone use, both continuously and repeatedly over time. This will support more responsive detection, allowing for more timely introductions of appropriate interventions.

Our approach is to use the validated questionnaires from previous studies on mobile phone use to establish whether or not recruited participants have problematic smartphone behaviors. We thus collected a wide variety of sensor and usage data from their smartphones for more than 3 weeks, on average. We analyzed this data to see if there were features that indicated whether a user was exhibiting problematic use. Finally, we interviewed a subset of our participants to better understand how they used their phones and whether they encountered problems with using their phones during their regular daily activities.

We found that the number of apps used and sessions per day positively correlated with problematic use. In addition, we also found that the combination of these features and other features—such as the number of event-initiated sessions, the ratio of SMSs to calls, the average length of sessions that were not initiated by incoming events and the number of apps used in sessions with events—were useful in detecting problematic usage. With a detection model using Adaboost with SVM classifiers, our detection model was 89.6% accurate in classifying problematic usage (F-measure of .707).

Next, we describe how previous studies have approached overuse or problematic use of smartphones. We then introduce our study design for collecting sensor and usage data from smartphones, and the use of subjective assessments for establishing problematic use. We describe which features of the automatically collected data are related to problematic use, and machine learning algorithms for detecting problematic use. Finally, we conclude with a discussion of future research on problematic use.

RELATED WORK

Here we describe past work in understanding people's smartphone usage behavior, and how that relates to addiction and problematic usage.

The psychology community has performed some significant research in an effort to understand mobile phone use. In 2005, even before smartphones became as prevalent as they are today, Bianchi and Phillips explored whether or not people's mobile phone use was related to gender and age, as well as a number of psychological factors [7]. In order to determine this, the authors developed the Mobile Phone Problem Usage Scale (MPPUS) as an indicator of problem use. This scale was formulated using a series of questions that ask people about their relationship with their mobile phone, incorporating current insights on behavioral and technological addiction. Whether or not excessive levels of technology use can be officially classified as "addiction" was outside the scope of their work and is a discussion still taking place within the psychological and medical communities. Regardless, it has been found that excessive and problematic use can have negative consequences, and this scale was developed and validated for gauging such mobile phone usage. Using this scale, Bianchi and Phillips found that using multiple linear regression, they could predict phone usage based on gender, age, self-esteem, extraversion, and neuroticism. It was also found that the MPUUS score was significantly correlated to the self-reported amount of mobile phone use time, number of calls made/received, and average monthly phone bill.

In 2007, Hooper and Zhou explored what they referred to as *addictive* mobile phone use by developing a survey to categorize the various ways people use their mobile phone [15]. They began by examining past research and identifying six categories of behavior: addictive, compulsive, habitual, mandatory, voluntary, and dependent. They then applied these categories to mobile phone usage and developed a questionnaire to capture how people's phone use relates to these categories. Administered to 184 participants, Hooper and Zhou found that while mobile phone behavior could fit in any one of the 6 categories, most frequently it was categorized as being dependent, voluntary, or mandatory. In other words, the participants felt that using smartphones was necessary to their lives, as found by others [1,14]. This categorization process is useful in understanding mobile phone use behavior and motivation. The authors note in their discussion that a factor analysis actually failed, and hypothesized that this was due to the different categorization types being so close to one another. Little exploration has been done to determine how well participants can self-identify their behavior, and if they can, what can be done with this information.

In 2009, Koo performed 2 studies on cell phone addiction with Korean adolescents. The first focused on developing a cell phone addiction scale [19], and the second on developing a prevention program for cell phone addiction [18]. In the first study, she prepared a 36-question assessment based on other studies, and then selected 20 among them, which were correlated but not redundant, for detecting cell phone addiction. She found that psychological measurements (self control and impulsiveness)—which describe how people control themselves—and perceived phone usage (*e.g.*, the self-reported number of calls and SMSs per day) are significantly correlated to her phone addiction scale. Koo introduced an informative lecture for students designed to reduce smartphone addiction, validating it with an experimental group and a control group [18]. After giving them the informative lecture, she found that the experimental group's self-esteem increased and smartphone use (*e.g.*, text messaging) significantly decreased.

Recently other studies have focused on analyzing how psychological factors affect smartphone usage. Ha *et al.* studied excessive cellular phone use among Korean adolescents [12]. They collected phone usage from self-reports and divided a group of cell phone users according to Bianchi and Phillips's MPPUS scale, into an upper group (top 30%) and a lower group (bottom 30%). They found that the upper group reported higher monthly smartphone bills and reported sending and/or receiving more SMSs than the lower group. In addition, they found that the upper group had more symptoms of depression than the lower group, particularly in terms of difficulty in expressing emotion, interpersonal anxiety, and lower self-esteem.

Takao *et al.* analyzed the relationship between addictive personality and problematic mobile phone use [26]. They found that high levels of self-monitoring, strong need for approval from others, and gender were related to problematic use of mobile phones through linear multiple regression. They found that although loneliness could not be used to predict problematic mobile phone use, it could predict overall use of smartphones, such as self-reported time spent on smartphones per week, number of people contacted regularly, and time spent using SMS per week. Shin *et al.* conducted an empirical study on mobile usage behavior focusing on smartphone addiction [25]. They adopted Hooper and Zhou's questionnaire [15], and examined factors such as self-monitoring, loneliness, self-efficacy, and two characteristics of smartphones (ability to change mood and social value). They found that self-monitoring, loneliness, and the smartphone characteristics were related to smartphone addiction, which also led to social withdrawal symptoms.

Oulasvirta *et al.* studied the habits of smartphone users through two studies [20]. In the first study, they analyzed application usage sessions from 136 smartphone users and 160 desktop computer users over a period of approximately 50 days. They found that smartphone users had more short sessions (less than 30s) than desktop users, and this usage was evenly spread over the day. In the second study, the researchers collected self-report data on smartphone usage from 12 students for 2 weeks, using the day reconstruction method [17]. They discovered several habitual patterns with various applications such as email, *Facebook*, update feeds, and headlines. Through these studies, the authors found that event or notification checking behaviors were repeated over time and that these and other similar habits contributed to increased use of smartphones.

Through these studies, it is clear that smartphone usage is growing and increasingly considered problematic. Previous studies have focused on overuse or problematic use and on finding behavioral factors that influence smartphone usage: self-esteem, impulsiveness, loneliness, and self-monitoring. However, these studies have primarily depended on the analysis of self-reported phone usage and were unable to automatically and objectively detect problem smartphone usage. As stated earlier, reliable data from self-assessments can be difficult to obtain, and repeated assessments can decrease quality even further. Furthermore, as a user's smartphone usage behavior can change over time and could, at times, be problematic or not, and the assessment mechanisms from past work cannot be applied continuously, past work will likely be unable to detect an individual's problematic use in a timely fashion. Lastly, as smartphones continue to acquire new and various functionalities and applications [8], self-report questionnaires and assessments will need to be constantly updated and re-validated to cover the increased and varied uses of smartphones.

In order to tackle these problems, we analyze smartphone usage by automatically collecting a wide range of contextual and usage information from smartphones, in addition to gathering psychological measurements. This approach provides us with objective data on how smartphones are being used and can be used to detect problematic behavior, repeatedly and in a timely manner.

RESEARCH QUESTIONS

We are interested in two questions about the problematic use of smartphones. First of all, we want to know which smartphone usage features are related to smartphone problematic use. As smartphones can be used in a variety of ways, usage patterns vary, and problematic behavior can be exhibited in multiple ways. Second, we are also interested in automatically identifying which smartphone users are exhibiting problematic usage. As usage problems may gradually change over time with repeated usage behaviors having negative consequences, it is important to identify problematic users as early as possible, so appropriate interventions can be introduced. Thus we have the following two questions:

- Can we identify phone context and usage features that correlate well with validated assessments of problematic use?
- If so, can we use those features to build a detection model to objectively and automatically identify which users are exhibiting problematic use?

STUDY DESIGN

To answer our questions, we designed a mixed-methods study in which we automatically collected data from participants' smartphones and measured their problematic usage. We now describe our recruitment process, our automated data collection process, and the smartphone problematic use assessment instruments we used.

Participants

We recruited 48 participants from our local university community (25), along with the Android marketplace (23) to have a more general participant population. Their ages ranged from 19 to 45 (mean of 26.7) with 34 males and 14 females. 25 are office workers, 20 are graduate students, and 3 are undergraduate students. We first asked them to fill out demographic information (*e.g.*, gender, age, occupation and education) and initial information about their phone usage. The phone usage information included how long they had used a smartphone and the reasons why they used their smartphones.

Data Collection and Phone Usage Extraction

We developed and deployed an Android phone app that unobtrusively collects data from participants' own smartphones, based on our data collection framework [10]. The deployment lasted for an average of 25.1 days (STD 6.6). The app collected a wide range of sensory data from users' smartphones including what apps were installed and in use, battery usage, events and notifications, and screen status data. See Table 1 for the full list. We validated that the app did not significantly impact the battery life of participant phones.

Table 1. Phone usage features and their sources.

Usage	Sources	Usage features extracted
General usage	Call log Screen status SMS log	-Number of calls and SMSs per day (*call, sms*) -Ratio of SMSs to calls (*smscall*) -Active hour per day (*activeHour*)
Battery usage	Battery info	-Battery charging frequency and battery level charged (*battChg* and *battLvl*)
Data usage	Network info	-Data received per day (*dataReceived*)
Push event usage	SMS log Call log Events	-Events received (*eventReceived*) -Response time to events (*respTimeToEvent*) -Probability responding to events (*respToEvent*)
Touch inputs	Touch events	-Number of clicks and scrolls occurred in apps used (*touchInput*)
Session usage	Screen status Event log App log	-Number of sessions per day (*session*) -Number of sessions with events, the number of sessions with/without apps used (*sessionWithEvent, sessionWithApp, sessionWithNoApp*) -Number of apps used with/without events per session (*sessionAppEvent, sessionAppNoEvent*) -Time length of sessions with events, or with/without apps used (*sessionTimeEvent, sessionTimeApp, and sessionTimeNoApp)*
App usage	App log	-Number of apps used per day (*appUsed*)

We extracted general phone usage features from the data collected. As collected via self-report in previous studies [7,19], we extracted general usage features, such as the average number of calls (*call*), SMSs (*sms*), and the total amount of time a phone's screen is on per day (*activeHour*).

In addition, as smartphones have a number of functionalities and apps, we extracted more diverse usage features of smartphones from the data collected.

- **Battery usage**: average level of battery when charged (*battLvl*), and the number of charging sessions per day (*battChg*), to indicate charging tendencies.
- **Network data usage**: amount of data received through mobile networks (*dataReceived*). Many smartphone apps receive network data from the Internet, and this usage indicates this type of app usage.
- **Session usage**: we define the interval between the screen turning on and the screen turning off as a *session*. A session indicates a unit of usage that involves app and event usage. During a session, a set of apps is executed, or no app is executed if users just check their events or the time. This usage data included the number of sessions

(*session*), the number of sessions without any events (*sessionWithApp*), the number of sessions with events (*sessionWithEvent*), and the number of sessions without apps used (*sessionWithNoApp*) per day. It also includes the session durations: *sessionTimeEvent*, *sessionTimeNoEvent*, and *sessionTimeNoApp*. We also use the number of apps used per session with or without events (*sessionAppEvent* and *sessionAppNoEvent*).

- **App usage**: number of apps used per day (*appUsed*), indicating how many apps users regularly used. Repeated use of an app in a session is aggregated into a single app use, while a single app used in different sessions is considered to be multiple apps, contributing to *appUsed.*
- **Touch inputs**: number of touch inputs (*touchInput*) shows how users click and scroll while interacting with apps, indicating the level of engagement with apps.
- **Push event usage**: we define the events sent from apps on a smartphone as *push events*. On smartphones, users frequently receive a number of push events (*e.g.*, new incoming SMS or email, upcoming events from calendar apps) from apps. This usage shows how many they receive per day (*eventReceived*). It also includes their response time to such events (*respTimeToEvent*) and how often they respond to the events (*respToEvent*).

Problematic Smartphone Use Assessment

We now introduce how we assessed problematic use of smartphones. Previous work has examined characteristics of smartphone addiction as an instance of behavioral addiction [11,16,19] and problematic use [7]. In these examinations, they have considered a range of criteria such as withdrawal symptoms, tolerance, conflict, compulsion, coping, and preoccupation due to the use of smartphones.

These criteria were determined by referring to the characteristics of behavioral and Internet addiction [7,19]. Although their definitions vary, the characteristics for identifying problematic use, or overuse, or compulsive use of smartphones, are similar to each other.

In order to measure problematic use of smartphones, we combined the previously validated assessments from Bianchi and Phillips [7] and Koo [19]. Bianchi's questionnaire contained 27 questions about social and emotional aspects of behavioral addiction, coping, and preoccupation, as well as dysfunction and tolerance [7]. Koo's phone addiction scale consisted of 20 questions about dysfunction, tolerance, and compulsion/persistence related to phone addiction behavior [19]. While these scales were used to measure problematic use, they were developed and applied independent of whether phone use was required for one's job. Rather they target one's emotional relationship to the phone and compulsion for using it. We created a comprehensive scale for the purposes of this study that includes all 27 questions from Bianchi and Phillips, as well as 7 questions from Koo that asked about compulsion and persistence. These 7 questions give insights into how often people keep their phone nearby and turned on, how often they check for messages and notifications, as well as how immediately they respond to messages, calls, and alerts. They were selected because they complemented Bianchi and Phillips' questions, to give a fuller picture of smartphone usage. Thus our final assessment contained 34 questions, each of which used a 5-point Likert scale. As with Bianchi and Phillips' MPPUS, the assessment score was the sum of the responses to the questions. Local participants filled out these questionnaires during our initial meeting, while marketplace participants filled them out upon downloading and launching the data collection app.

Individual Interview

At the conclusion of the study, we held short one-on-one interviews with each of the 25 local participants to gain a more qualitative understanding of how much they used and thought about their phones in their daily lives. Each participant was first asked to describe their regular weekly schedule and talk about how their phone plays a role in their various daily activities.

During the interview we surveyed the participants about their smartphone use using a 5-point Likert scale. This included questions about how often they used their smartphones in daily activities, with a range of answers between "never" to "always"; how much their phone use is related to their activities, with a range of "strongly unrelated" to "strongly related"; and whether they felt that they needed to reduce the use of their phone in a variety of different situations, with a range of "strongly disagree" to "strongly agree."

Lastly, during our exit interview, we asked the participants to state whether they experienced any personal or interpersonal problems as a result of using their smartphones, and to talk about whether or not they felt their level of usage was problematic.

RESULTS

Basic Statistics

We report the usage features from the 48 participants in Table 2. We found that, on average, participants spent 3.0 hours using their smartphones and executed apps 147.7 times per day. They made and received 34.2 calls and sent and received 54.1 SMSs per day. The average number of sessions per day across all participants was 89.9 (or about 1 every 16 minutes of a 24-hour day).

In addition, we also confirmed the reliability and internal validity/coherence of the questionnaires used in our study by measuring Chronbach's alpha coefficient. The value of 0.949 indicates the questions have high internal validity.

Influential Phone Usage to Problematic Usage

In order to build an automated detection system for problematic smartphone use, we analyzed the relationship between phone usage and problematic usage. We first identified features by performing a correlation and a linear regression analysis for all participants' data and assessments. We then divided the participants into two groups: those with high problematic use assessment scores

Table 2. Participant smartphone usage.

Usage features	*Mean*	*STD*
activeHour	3.0 hours	1.5 hours
appUsed	147.7	110.5
call	34.2	48.5
sms	54.1	61.6
smscall	57.1%	30.2%
battChg	3.1	2.3
battLvl	41.2%	15.9%
session	89.9	54.3
dataReceived	56.6 MB	63.6 MB
touchInput	535.0	903.0
eventReceived	177.6	220.1
respToEvent	86.3%	10.2%
respTimeToEvent	4.5s	3.5s
sessionWithEvent	43.3	32.5
sessionWithApp	27.4	28.6
sessionWithNoApp	19.5	24.5
sessionTimeApp	1.4 min	1.5 min
sessionTimeEvent	3.8 min	6.1 min
sessionTimeNoApp	0.9 min	1.4 min
sessionAppNoEvent	1.5	0.7
sessionAppEvent	2.4	2.1

and those with low scores, and compared the mean and information gain of features to identify additional distinguishing features for the high scoring group. We combined all of these features into a model for automatically detecting problematic use. We now discuss this process and our results in detail.

Analyzing Users for Problematic Usage

We first calculated the correlations between our phone usage features and the problematic usage scale. We found that *session* ($r = 0.310$, $p < 0.05$) and *appUsed* ($r = 0.286$, $p < 0.05$) were significantly correlated with smartphone problematic use: (see Fig. 1). Therefore, the users who were assessed as having greater problematic usage behaviors, used more apps, and had more interaction sessions.

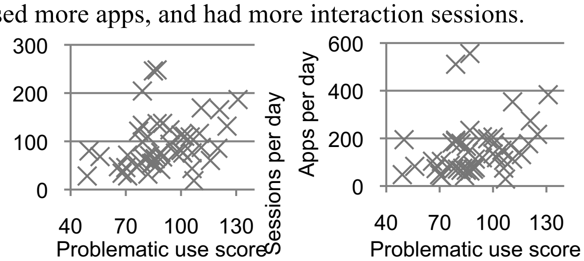

Fig. 1. Correlation between smartphone problematic use scores and phone usage features: sessions per day (left), and apps used per day (right).

Moreover, in order to understand how the *combination* of phone usage features are related to problematic smartphone use, we applied linear multiple regression to our data. For this purpose, we checked the feasibility of using each feature: normality and *multi co-linearity*. For normality, we used the *Shapiro –Wilk* test [23] and excluded *dataReceived* and *battChg* since they had non-normal distributions. We also removed *session, activeHour,* and *eventReceived* due to their *multi-colinearity* (*i.e.*, a high degree of correlation to other variables). We then calculated the linear relationship of the remaining 16 features between the assessment scores and the combinations of these features to identify the most significant features.

We found that the combination of these phone usage features was significantly related to problematic smartphone use. This regression result was quite significant ($r^2 = 0.490$, $F(11, 36) = 3.747$, $p < 0.005$) with the most influential features being *call*, *appUsed, sesssionWithEvent, sessionAppEvent* and *sessionTimeNoApp*, as shown in Table 4, which shows the name, regression coefficient β, t distribution value, and p value of each usage feature.

Table 3. Most significant results from the multiple regression.

Explanatory variables	β	t	p
call	**-.529**	**-3.188**	**.003**
appUsed	**1.642**	**2.119**	**.041**
sms	.239	1.120	.270
smscall	-.329	-1.418	.165
respToEvent	.163	1.056	.298
sessionWithEvent	**1.366**	**2.875**	**.007**
sessionWithApp	.693	1.610	.116
sessionAppNoEvent	-.152	-.953	.347
sessionAppEvent	**.505**	**2.170**	**.037**
sessionTimeNoEvent	-.195	-1.348	.186
sessionTimeNoApp	**-.285**	**-2.072**	**.046**

One feature (*appUsed*) found from our earlier correlation was identified in this regression, yet other features, such as the number of calls per day (*call*), the number of sessions triggered by a push event (*sessionWithEvent*), the number of apps used per event session (*sessionAppEvent*), and the length of sessions in which no apps were used (*sessionTimeNoApp*) also contributed to describing problematic smartphone use of the participants. Thus participants with higher problematic use scores tend to have more sessions that are related to checking events. They also use more apps when there is an event to be checked, and have shorter sessions in which they do not use any apps.

Analyzing User Groups with Problematic Usage

In order to know how users assessed as having problematic use are different from the other users, we divided our participants into 2 groups, those with high and low problematic use assessment scores. In past work, this partitioning was performed by considering either the top 8% [19] or top 30% [12] of high scoring users, Although there is no accepted guideline about how to perform this partitioning, one study found that the appropriate percentage was 16% [24]. We use a more robust method for obtaining the ground truth of unlabeled samples: using unsupervised learning to cluster samples that minimize the differences within a cluster, and maximize the differences across clusters.

As shown in Fig. 2, we selected 8 participants (top 17%) as having high problematic use. Two of them were graduate students and the remaining 6 were office workers. The average assessment scores of the high and low problematic group were 118.1 (STD 7.5) and 84.2 (STD 14.4) respectively. The scores of the 2 groups were significantly different in a two sample t-test ($T(46) = 5.999$, p = 000).

As can be seen in Fig. 3, among phone usage features, the number of SMSs, apps used per day, and event sessions are significantly different between the two groups. The results of a two sample T-test were: *sms*, $T(46) = 2.736$, $p < 0.009$; *appUsed*, $T(46) = 2.255$, $p < 0.029$; and *sessionWithEvent*, $T(46) = 2.945$, $p < 0.005$. In other words, the high problematic use group made/received more SMSs, used a higher number of apps per day, and had more sessions triggered by events each day.

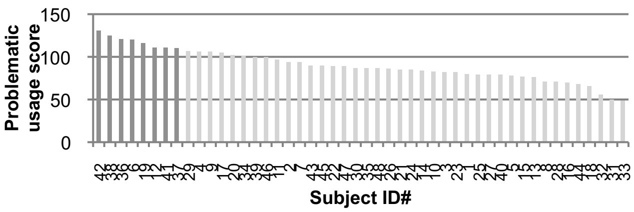

Fig. 2. Assessment of problematic smartphone use.

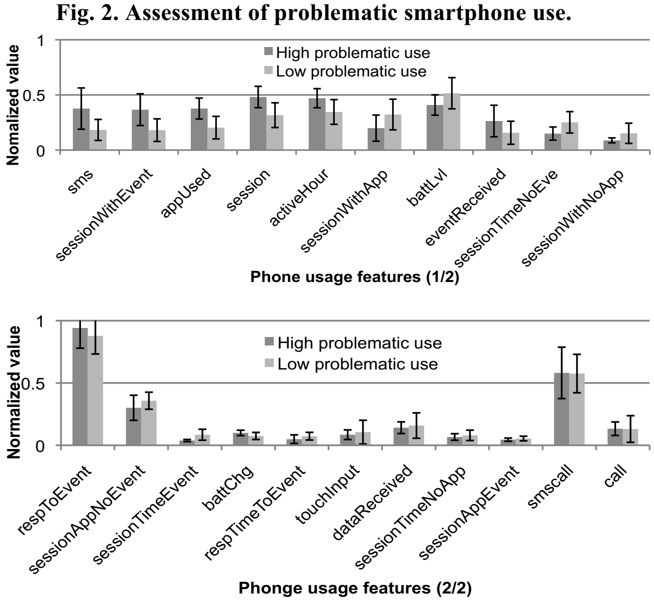

Fig. 3. Comparisons of the normalized phone usage features between the high and low problematic use groups.

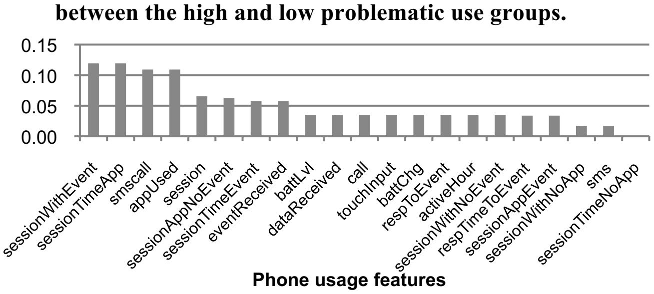

Fig. 4 Information gain of normalized phone usage features

In addition, we measured the information gain of each feature to know how valuable individual features are for differentiating between the two groups. As all features are continuous values, we discretized them with 3 bins with equal frequency. We used the InforGainAttributeEvaluation and Ranker search method in Weka [28]. As shown in Fig. 4, the features most influential to this partitioning are *sessionWithEvent, sessionTimeApp, smscall, appUsed, session, and sessionAppNoEvent*.

Using the 4 analyses (correlation, multiple linear regression, mean and information gain), we identified a set of promising features for objectively assessing problematic use: *sms*, *call*, *smscall*, *session*, *appUsed*, *sessionAppEvent*, *sessionAppNoEvent*, *sessionTimeApp*, and *sessionTimeNoApp,* and *sessionWithEvent*.

Problematic Use Detection

To *automatically* identify which users have problematic usage, we used these phone usage features in a machine-learning algorithm. As we had 48 users each with data from an average of 25.1 days of deployment, we used 3 different models based on Adaboost and the selected features, using for classifiers: naïve Bayes (NB), Support Vector Machines (SVM), logistic regression. A fourth model used Adaboost and NB, but with general usage features that indicate how heavily users used their smartphone such as active hour, the number of sessions, the number of SMSs, and the number of apps used per day. The Adaboost approach iteratively produces improved models that more heavily weight the impact of classification errors made by the previous iteration. It thus only requires a small number of training samples to estimate its parameters and supports a stronger classification for diverse samples.

We first preprocessed the data and measured the detection accuracy for 48 users. We discretized the phone usage features with 3 bins with equal frequency as before. We then excluded 4 features (*call*, *sms*, *session*, and *sessionTimeApp*), which were highly correlated to other features, identified using the cfsSubSetEval feature selector in Weka [28]. The selected features from the classifier were: the number of apps used per day (*appUsed*), the ratio of SMS to calls (*smscall*), the number of event sessions (*sessionWithEvent*), the number of apps used without events per session (*sessionAppNoEvent*), the number of apps used with events per session (*sessionAppEvent*), and the length of sessions without apps used (*sessionTimeNoApp*). Interestingly, the ratio of SMSs over calls (*smscall*) was not significantly higher for the high problematic use group, as seen in Fig. 3. However, this feature was important for detection, when the ratio was either very high or very low. It can be seen that problematic use is not the same as heavy use of a smartphone such as increased SMSs, calls and sessions, but was more complex, involving events, apps, sessions and session durations.

We evaluated the classification models with cross validation per user (48-fold cross validation), due to the small number of samples. We compared their performance to the Most Frequently Used (MFU) strategy, which selects the most frequently occurring class (*i.e.*, all participants are classified as not being problematic users).

MFU and Adaboost with NB and general usage features performed poorly: accuracies of 83.3% and 72.9%, respectively. The models with the selected features outperformed these: with SVM (89.6%) and logistic regression and NB (both 87.5%. We also calculated the F-measure, a more meaningful measure of accuracy that takes into account false negatives and false positives. MFU and Adaboost with NB and general usage features had F-measures of 0 (precision of 0, recall of 0) and .235 (precision of .222, recall of .25), respectively. The F-measures for Adaboost with the selected features were: .667 for logistic regression (precision of .6, recall of .75); .700 for NB (precision of .583, recall of .875); .707 for SVM (precision of .667, recall of .75) While the numeric gains over MFU may seem small, it should be pointed out that MFU detects *none* of the problematic users, and, therefore, offers no classification value. The best performing models, Adaboost with NB and SVM with chosen features, correctly classified 7 and 6 out of 8 problematic users, and 35 and 37 out of 40 non-problematic users, respectively. They significantly outperformed MFU and Adaboost with NB with general features (two sample t-Tests $T(94)=7.10$, $p = 0.000$, and $t(94)=9.47$, $p=0.000$, respectively).

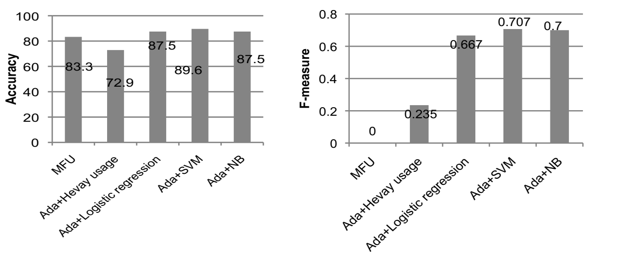

Fig. 5. Results of the detection models. Left: accuracy, right: F-measure.

S42, the only problematic user misclassified by both models, is an office worker. He used apps (*appUsed)* and had many event sessions (*sessionWithEvent*) but his ratio of SMSs to calls (*smscall*) was neither extremely high nor low. Four of the five users misclassified by the NB model as problematic users were students while the other was an office worker. Three of the students (S20, S3, and S15) frequently used SMSs resulting in very high ratios of SMSs to calls (*smscall*), while the remaining student, S18, did not. Instead S18 had shorter sessions in which apps were used (*sessionTimeApp*). The misclassified office worker (S30) used a number of apps per day (*appUsed*) and made a lot of daily calls, for a very low ratio of SMSs to calls (*smscall*).

Individual Interviews

We were only able to conduct interviews with our 25 local (and mostly student) participants and not the anonymous 23 participants recruited through the Android Marketplace. We used the interviews to verify that participants' use of their phones did not change as a result of being in our study. In the final interviews, our local participants stated that they used their phones in a variety of daily activities to communicate with their friends, retrieve information from the Internet, schedule and remember events/tasks, and enjoy entertainment, *etc*. Fig. 6 reports on 5-point Likert scale responses to how often they used their phone in various situations, whether they felt that usage is related to the situations, and whether they felt they needed to reduce smartphone use in these situations. They frequently used their phones waiting for a bus or car, when they wake up, and when they watch TV/DVD. They sometimes used their phones while driving or when in the bathroom, both instances where usage is likely not appropriate. While they agreed that they sometimes used their phones even when their smartphones were not necessary, overall they *did not* feel they needed to reduce their smartphone use, except while studying and driving.

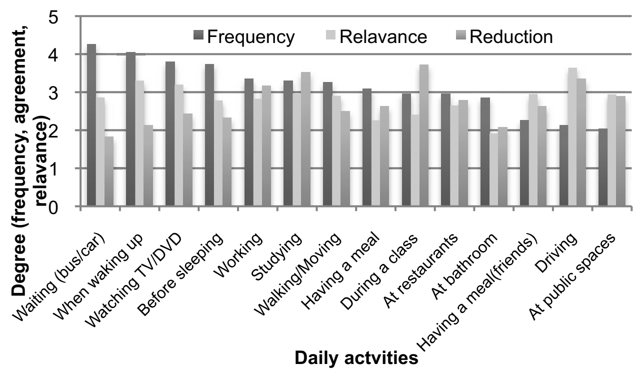

Fig. 6. Smartphone usage in daily activities.

In addition, most of the participants stated that they were frequently proactive in checking their phones to see whether there were any events from apps. They received a number of events that required them to take action: most often, SMSs, but also email, *Facebook* and *Twitter*.

The 4 problematic usage participants that we interviewed also tended to use their phones heavily. S6 said, "I used my phone too much while in class and bed." He said that he used his phone when his classes were boring. He said that he checked Facebook and Twitter for 1-2 hours before sleeping. He also said, "My girlfriend complained to me that I used SMSs too much, and didn't pay attention to her when we meet." S12 preferred to use SMSs over making calls to communicate with her friends. She stated, "I use SMSs too much" and said, "I'm pretty fast in texting." S19 stated, "I'm using SMSs and the calendar app too much." She estimated that she checks her phone every hour to see whether there are any events from apps.

Many of the participants in the non-problematic use group *also* reported that they used their phones too much during some daily activities and were sometimes concerned about their use of their phones. Five participants (S11, S17, S20, S21, and S24) thought that they were frequently checking events and three participants (S7, S17, and S25) said they used their phone to pass time. Two participants (S7, and S24) said they used SMSs too much and four participants (S16, S20, S23, and S24) said they used game apps too much. Furthermore five participants (S2, S4, S14, S17, and S21) worried about low battery on their phones whenever they missed an opportunity to charge their phone, since that would restrict their use of SMSs/calls. Four participants (S10, S16, S21, and S17) used their phone while driving, yet they only sent/received SMSs when their cars were not moving. Three participants (S2, S3, and S21) frequently used streaming music services, which uses a lot of network data. Finally, three participants (S3, S17, and S20) had also been told by friends or parents that they were using their phones too much, while texting or playing games. Among our 25 interviewees, 13 had some indication that they may be problematic users, while only 4 were actually problematic according to the questionnaire assessment. The two best classifiers detected all but one of the problematic users from this group and classified the remaining non-problematic participants correctly.

In addition, our interviews elicited additional features that could contribute to improving the detection of problematic users. One feature is the specific apps that were being used frequently; several users pointed out that they were using specific apps too much such as games, and SMS. A second feature is changes to settings; users who did not change their phone alert settings (*e.g.*, vibrate, lowered volume) said they were more responsive to phone calls and events.

DISCUSSION

We are very interested in understanding how specific phone usage features are related to problematic use of smartphones, and how they may be leveraged in order to detect problematic use. As seen in the previous section, we can successfully build a system that detects whether an individual would be assessed as having high problematic use. Here we discuss the relevance of particular usage features, how to address the issues of outliers, the need for repeated assessment, and how to apply a detection system.

We found that problematic usage patterns not only rely on heavy use of smartphones but also involve other usage features. Initially, we found that higher values for individual features were related to problematic use. In the comparison of the two groups, the problematic users tended to use more SMSs and apps, and had more sessions per day than the participants who were assessed as being non-problematic users. However, among them only the number of apps used per day was selected in our classification model, along with a set of more complex features. The participants assessed as having problematic usage also tended to dominantly use either SMSs or calls and had more sessions triggered by events than other sessions. It is interesting to note that Koo's study found that the number of self-reported calls and SMSs a subject made or received was related to problematic phone use [19], but we found that the number of apps, the ratio of SMSs to calls, and event sessions were more indicative of problematic use. Similarly, we found that the self-reports of our interviewees were not completely reflective of their assessment. On one hand, 2 of 4 problematic smartphone users thought they had problematic use, while the other 2 were told by friends or family members that their use was problematic. On the other hand, 13 of the non-problematic users self-reported problematic behavior or were told by others that they needed to reduce their smartphone usage. Rather than rely on these less accurate self-reports, we take a more objective, automated approach in detecting problematic use.

We found that people in the problematic use group had shorter non-event-initiated interaction sessions than those in the non-problematic use group. However, there was no distinction in the number of apps used in such sessions. This means that people in the high problematic use group spent less time per app than those in the low problematic use group. Further, they tended to use more apps per event-initiated session compared to the low problematic use group. In future work, it will be useful to explore the details of what users do during event-initiated and non-event-initiated interaction sessions. This may help us understand whether users in the higher group are simply more efficient in looking at apps and/or whether they feel a need to look at certain apps when a notification prompts them to use their phone, even in the absence of a reason to look at those apps.

While the best detection system was quite accurate at 89.6%, it needs to be adapted to address the misclassifications. One user in the problematic group used apps and had many sessions, but was not correctly classified. The five users in the non-problematic group misclassified by the NB approach had similar usage patterns as the high problematic users, as described earlier. In order to detect these users, more elaborate models dedicated for those that were misclassified need to be developed. We expect that repeating this analysis with a larger sample size will both improve the quality of our model and help to identify a more diverse set of users. However, despite these misclassifications, the best classifiers significantly outperformed MFU (detects no problematic usage) and the high usage (detected only 2 of the 8 problematic users) approaches.

Our two best performing approaches had similar results, with the SVM model slightly outperforming the NB model. There is a clear tradeoff between these. The NB model was better at classifying problematic users (missing 1 *vs*. 2 for the SVM), while the SVM model was better at classifying non-problematic users (missing 3 *vs.* 5 for the NB). Which model to use, depends on tolerances for misclassifications for the problematic and non-problematic users.

Past work has used a single subjective assessment to determine whether a subject is exhibiting problematic phone use. However, to determine whether some form of intervention is needed, the detection of problematic usage would likely need to be repeated over time. First, repeated assessment would support the identification of a usage trend towards more or less problematic usage. Second, combining the results of multiple detections would minimize the impact of noise in the detection process and noise in subject behavior. Fortunately, as our detection approach is objective and automated, it can be repeated as frequently as desired. It remains as future work to determine how to identify concerning trends, to determine a re-assessment frequency and how to combine multiple assessments into a single *actionable* determination of problematic usage.

Timely detection of problematic use is important, as it could allow for some type of early intervention to help users be aware of their behavior and perhaps avoid this problematic use in the future. As our approach can be reapplied frequently and at low inconvenience to the user, it increases the likelihood that it could detect problematic use quickly after such behavior is exhibited. In future work, we will work with experts in technological addiction and phone use to identify appropriate interventions that can be applied, depending on the severity of the problematic behavior. Koo studied the use of a lecture-based intervention with students who were assessed as being addicted to their phones [18]. The intervention was effective, with students reporting increased self-esteem and reduced phone usage.

Finally we highlight a few limitations of our work. First, the number of participants in our study was small (48), and with almost half being students at the same campus. While we attempted to address this issue by also recruiting half of our subjects from the Android marketplace, we understand this is a limitation of our work. By increasing the size and diversity of our study population, we expect to see greater variations in smartphone usage patterns and we will be able to better validate our detection approach. Despite this, our approach is quite promising for the detection of problematic use. Secondly, we observed participants for a relatively short period of time (approximately 3.5 weeks, on average). While participants reported that their phone usage during the study period was indicative of their normal routine, looking at longer-term data could yield some interesting insights. Particularly as users' behaviors change according to their schedules or changes in activities, friends *etc*., we would expect to see changes in their phone usage.

CONCLUSION

In this paper, we designed a mixed-methods study to understand and detect smartphone problematic use by collecting a wide range of context and usage data from users' smartphones, and assessing them with a validated problematic usage instrument. We identified features that were relevant to detecting problematic usage: the number of apps used per day, the number of event-initiated sessions, the ratio of SMSs to calls, the length of interaction sessions without an incoming event, and the number of apps used in sessions with incoming events. We trained a number of detection models using these features and found that the best-performing model could distinguish between higher and lower problematic use subjects with an accuracy of 89.6%. This work is a promising step towards being able to objectively and automatically detect problem behavior with technology. Important future work includes improving our detection accuracy, working with a larger population, and incorporating our detection method into clinical workflow, to identify when to offer an intervention and to determine the nature of that intervention.

ACKNOWLEDGEMENTS

This research was supported by a grant from the R&D Program (Implementing the Next Generation Smart Learning Center) funded by the Ministry of Trade, Industry & Energy (MOTIE), Republic of Korea and was also supported by NSF grant IIS-1017429.

REFERENCES

1. http://annapetlina.wordpress.com/2011/11/08/smartphone- addiction/
2. http://articles.cnn.com/2012-03-06/
3. http://blog.nielsen.com/nielsenwire/?p=29786/
4. Microsoft Tag, The Growth of Mobile Marketing and Tagging, March 21, 2011. http://tag.microsoft.com/community/blog/t/the_growth_of_mobile_marketing_and_tagging.aspx
5. http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats
6. http://www.cnn.com/2010/TECH/mobile/10/18/
7. Bianchi, A., & Phillips. J.G. Psychological predictors of problem mobile phone use, CyberPsychology & Behavior 8, 1(2005), 39-51.
8. Bødker, M., Hedmanm, J., & Gimpel, G. The User Experience of Smart Phones: A Consumption Values Approach, 8th global mobility roundtable conference, 2009.
9. Del Boca, F. K., & Noll, J. A. Truth or consequences: the validity of self-report data in health services research on addictions, Addiction, 95, 11s3(2000), S347-S360.
10. Ferreira, D., Dey, A.K & Kostakos, V. Understanding human-smartphone concerns: a study of battery life, In Proc of Pervasive 2011, 19-33.
11. Griffiths, M. Internet addiction: Does it really exist? In J. Gackenbach (Ed.), Psychology and the internet: Intrapersonal, Interpersonal, and Transpersonal Implications, Academic Press Inc. (1998), 61-75.
12. Ha, J.H., Chin, B., Park, D.-H., Ryu, S.-H., & Yu, J. Characteristics of excessive cellular phone use in Korean adolescents, CyberPsychology & Behavior 11, 6(2008), 783-4.
13. Haigneya, D., and Westermanb, S. J. Driving Mobile (cellular) phone use and driving: a critical review of research methodology, Ergonomics 44, 2(2001).
14. Harris Interactive Poll. Lookout Mobile Mindset Study, May 2012.
15. Hooper, V. & Zhou, Y. Addictive, Dependent, Compulsive? A Study of Mobile Phone Usage, BLED 2007, Paper 38.
16. Jang, H.J. The psychological characteristic of adolescents addictive using cellular phone, unpublished master's thesis, Sungshin Women's University, Seoul, 2002.
17. Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. A survey method for characterizing daily life experience: The Day Reconstruction Method (DRM), Science 306, 5072(2004), 1776-1780.
18. Koo, H. Development and Effects of a Prevention Program for Cell Phone Addiction in Middle School Students, Journal Korean Academy of Child Health Nursing 17, 2(2011), 91-99.
19. Koo, H. Development of a cell phone addiction scale for Korean adolescents, Journal of Korean Nursing 39, 6(2009), 818-828.
20. Oulasvirta, A., Ratternbury, T., Ma, L. & Raita, E. Habits make smartphone use more pervasive, PUC, 16, 1 (2012), 105-114.
21. Rush, S. (2011). Problematic use of smartphones in the workplace : an introductory study. Thesis, Central Queensland
22. Salovaara, A., Lindqvist, A., Hasu, T., & Hakkila, J. The Phone Rings but the User Doesn't Answer: Unavailability in Mobile Communication. MobileHCI 2011, 503-512.
23. Shapiro, S. S., & Wilk, M. B. An analysis of variance test for normality. Biometrika, 52, 3/4(1965), 591-611.
24. Sheffield, D. Mobile phone use and perceptions of stress, Proceedings of Division of Health Psychology Annual Conference; September 13-15. University of Essex, UK, 2006.
25. Shin, H.-K., Lee, M. S., & Kim, H. G. An Empirical Study on Mobile usage Behavior, Informatization Policy 18, 3(2011), 50-68.
26. Takao, M., Takahashi, S., & Kitamura, M. CyberPsychology & Behavior 12, 5(2009), 501-507.
27. Walsh, S.P., White, K.M. & Young, R.M. Needing to connect: The effect of self and others on young people's involvement with their mobile phones. Australian Journal of Psychology, 62, 4(2010). 194-203.
28. Weka, http://www.cs.waikato.ac.nz/ml/weka/