# Can You Form Healthy Habit? Predicting Habit Forming States through Mobile Phone

Yin Bai, Bin Xu, Shuyang Jiang, Haifeng Yang, Jian Cui

Department of Computer Science and Technology
Tsinghua University
Beijing, China
baiyin0429@gmail.com, xubin@tsinghua.edu.cn
695111463@qq.com, hellotony@live.cn, cuijian.hh@gmail.com

## ABSTRACT

Health-compromising behaviors are difficult to change since people do not behave in accordance with their intention. This paper aims at studying the extent to which a person's healthy habit forming process can be affected by mobile phone usage. We propose a novel healthy habit forming states predicting framework using mobile phone platform. First we present a definition for the healthy habit forming process consisting of several states. We define the social intervention types and user context data which are extracted from mobile phone sensor data. Then we make use of machine learning methods to study the correlation between these data and healthy habit forming states. Specifically, a predicting model called Habits Factor Graph(HaFG) is proposed to predict the habit forming states. To evaluate our work, an Android based prototype system is implemented. Experimental results show that the healthy habit forming states are predicted possibly from user context information with a fairly good accuracy (around 67%).

## Keywords

mobile phone sensor, mobile healthcare, mobile social network, factor graph, healthy habit

## 1. INTRODUCTION

Health intervention is one of classic methods in healthcare medical field to help people change habits. Interventions are offered by experts,doctors or friends and then the feedback informations are collected from subjects, for instance, by filling a questionnaire. The interventions efficiencies are examined and adjustments are taken by analyzing the feedbacks[7]. However, there exist several problems in this kind of method. First, the intervention types are limited and the feedbacks cannot be obtained immediately because of the communication obstacle. Second, the model is coarse-grained due to the lack of detail information. For example, the habit forming process not only is affected by health interventions but also depends on the user context data such as user activity and surrounding environment. Third, ex-

isting research works mainly focus on relationship between intervention activity and habit formation, but rarely try to predict the habit forming process using the context data. Based on these problems, some question are raised: can we predict the habit forming process using the data such as intervention activity and user context information? if the answer is yes, then a further question is, can we implement the method in a way of low cost and high facility? These are the main research goal in our work.

Fortunately, the fast development of smart phone and B-SN technologies provide new thoughts and methods to solve the problems. In this paper, we developed a mobile phone-based prototype system to collect context data such as interventions between users, living environment data and user physical activity data. We do data observations and find out the relationship between those information and user's healthy habit forming process, then try to predict healthy habit states by utilizing the relationship. The contributions of our work are as follows:

**1** We propose a quantitative description of healthy habit forming process.

**2** We propose HaFG predicting model to predict the healthy habit forming states using context data with a fairly good accuracy (about 67%).

**3** We implement Android-based prototype system including a mobile SNS application and mobile sensing module. The system significantly improves the intervention efficiency and reduces the cost compared to the previous habits intervention works.

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 introduces the definitions of problem and the data type. Section 4 talks about the data observation and proposes predicting model and learning process. Section 5 presents experimental results that validate the effectiveness of our method. Finally, Section 6 concludes this work and talks about future work.

## 2. RELATED WORKS

Research works in the field of psychology and medicine focus on the healthy habits changing process modeling and intervention influences examining. Health intervention methods are studied and the effectiveness are examined by dividing the people into different groups in [8] by S. Yajima. Schwarzer etc.[6, 7] proposed the Health Action Process Approach(HAPA) and evaluated the applicability of HAPA by replicating it across seven kinds of health behaviors of sam-

ples from different countries.

The BSN technologies are developed rapidly in recent years. Ma etc.[4] propose a method for assessing user's emotional states using mobile phone devices. A prototype system called SleepMiner is implemented to predict user's sleep quality using mobile phone data in [2]. Funf[1] is an Android platform developed by MIT media lab and the core concept of which is to provide an open source, reusable set of functionalities, enabling the collection, uploading, and configuration of a wide range of data types. Until recently, there are few works focus on healthy habits fostering using BSN technologies since that the process needs long time with different stages.

## 3. PROCESS AND PROBLEM DEFINITION

### 3.1 Healthy Habit Forming Process

We first give the definition of the habit forming process shown as Figure 1. For quantitatively analyzing the healthy habit forming process, we define seven definite states which includes "Intention", "Planning", "Maintenance", "Recovery", "Initiative" from HAPA theory and two additional states as "Success" and "Quit". "Success" means that the user has already developed a healthy habit successfully, and the "Quit" denotes that the user quit from the healthy habit fostering process with failure. We define each state explicitly and quantitatively for each habit so that we can examine user's state in definitive way.
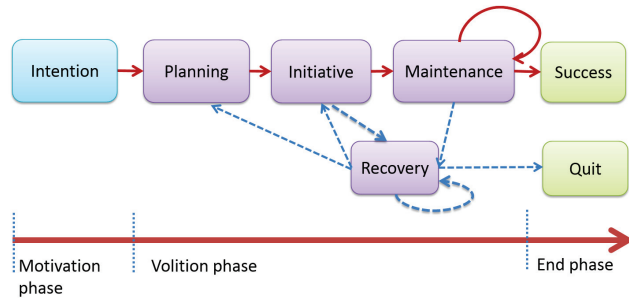


**Figure 1: Habit forming process**

The process is defined to fit any kind of healthy habit. Without loss of generality, we choose "Early-Sleep-and-Early-Wake(ESEW)" healthy habit to examine our work since it is one of common and important healthy habit. We give the definition of each state for ESEW as follows.

Users are initially in "Intention" state and are required to fill questionnaires designed by us. In our experiments, we only invite the people who have problems of sleep habits and have a willing to make a change. When subjects join the experiments, the real health forming process starts and the subjects' states are transformed into "Planning". The completion extent of the everyday plan are defined by three types, as "accomplished", "failed" and "partially accomplished" respectively. We use rules defined by us to confirm the exact state for each user every day. States are transformed from one to another according to the completion extent and the previous day's states.

### 3.2 Problem Description

In this subsection we formally define the problem. First we describe a mobile social network including the subjects and activities between them. A mobile social network can be defined as $G = (V, E)$, where $V$ is the set of $|V|$ subjects and $E \subset V \times V$ is the set of mobile social activities between those subjects. Specifically, $E_{ij}^t$ represents social intervention features between subject $i$ and $j$. $e_k$ is the value of feature $j$ of two subjects represented by edge $i$.

A set of intervention features are designed and extracted from mobile phone platform, which are shown as Figure 2. Interventions mainly refer to the online social interactions using our client application. First we define three kinds of "Entity" including "Status", "News" and "Article" respectively. "Status" means the one's history state information; "News" refers to text information send by users talking about their state; "Articles" are some healthy habit related articles recommended to user from server. Interventions are accomplished by using these entities as bridge. Specifically, we define four types of interventions as follows:

a) Comment: comment means any messages written by a user to reply to "News", "Article" or "Status" of other users. We count on the message number each user sent in each day and the number a user received each day as a feature.

b) Evaluate: there are two kinds of evaluations in our model, as "praise" and "criticize" respectively. User can choose one of them as evaluation to others' state. We also count the number of each kind of evaluation as a feature.

c) Alarm: alarm means the time reminder which users can send to each other. We use the alarm number one received in each day as a feature.

d) Share: users can share health related articles to each other. System will count the sharing number of each user in each day as feature.

Those intervention activities collectively referred to as "Intervention feature".

We also specified the context of subjects as "Attribute feature" using mobile phone data. Features of subject $i$ in day $t$ is denoted as $X_i^t$. In which, $x_{ij}^t$ is the value of feature $j$ of subject $i$ in day $t$. In this study user attribute mainly includes user activity data and living environment information. User activity data refers to the user physical activity such as "Running", "Walking", "Sitting" and "Standing". We extract these information from mobile accelerometers raw
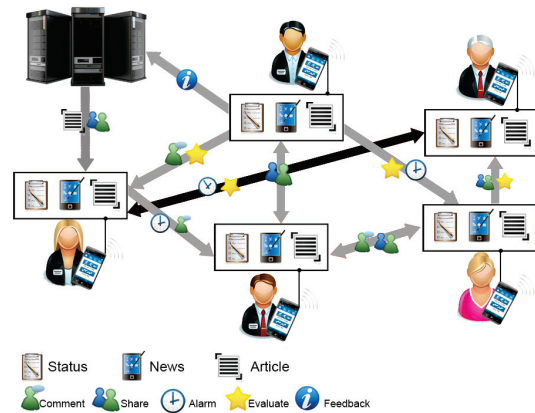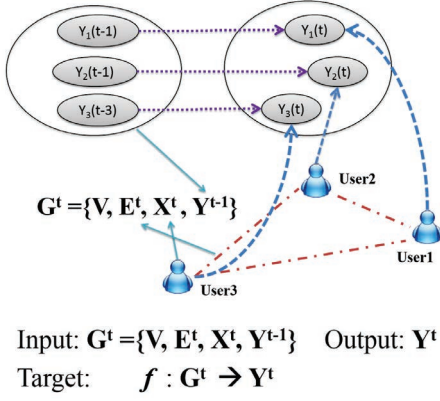


**Figure 2: Health interventions on mobile devices**

**Figure 3: Habit forming states predicting problem**



**Figure 4: Intervention correlation distribution**

data; Living environment information includes the average sound/light intensities of user's surroundings, which are collected from phone microphone and light sensor periodically.

Besides the two kinds of features talked above, we also use user's previous day's healthy habits state as one feature. We denote this with $Y_i^{(t-1)}$, which represents the state of the user $i$ at the day of $t-1$.

Based on the above discussion, states of a person can be modeled as a directed graph $G^t = \{V, E^t, X^t, Y^{(t-1)}\}$, where $V = \{v_1, v_2, ..., v_n\}$ is the set of subjects. $E \subseteq V \times V$ is the set of links between subjects. Each directed link $e_{ij} = (v_i, v_j) \in E$ indicates a social activity vector between subject $v_i$ and subject $v_j$ in day $t$. $x_{ij}^t$ is the value of attribute feature $j$ of subject $i$ in day $t$.

Finally, the predicting problem can be defined as:

**Healthy Habit Forming States Predicting**

Let $<1,...,t>$ be a sequence of time stamps with a particular time granularity (one day in our work). Given a feature set $X^t$, the previous day's healthy habit fostering states $Y^{(t-1)}$, and the social activity $E^t$. The task is to find a predictive function:

$$f : G^t \rightarrow Y^t \qquad (1)$$

such that we can forecast the healthy habit forming states at time $t$; $Y^t$ is the state of user in day $t$. The problem definition is shown as Figure 3.

## 4. DATA OBSERVATIONS AND PREDICTING MODEL DESCRIPTION

In this section we will present some significant observations over a real-world dataset captured by our prototype system. First, we compare the healthy habits developing speed between active and inactive people and find significant differences between subjects. As shown in Figure 4, the active users spend shorter time to reach "succuss" state than inactive group, which proves that the intervention activities benefits to habits change.

There are also some other observations from data context, such like people are more interested in "Evaluate" intervention than other types; "Share" intervention shows more effects on early states like "Maintenance", etc. Here we do not present these results due to the page limit. To sum up, we find out that the interventions and contexts correlates with
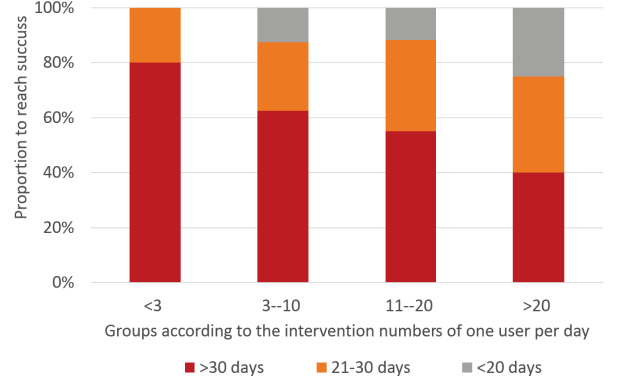
healthy habits changing to some extent. That also gives us the motivation to predict the healthy habits changing based on mobile data.

To predict the healthy habits, we need to incorporate the different factors into a unified model. We describe our solution in next subsection.

### 4.1 Predicting Model

Based on the above observation, we present Habits Factor Graph(HaFG) model to deal with the state prediction problem. HaFG is proposed based on factor graph [3] from machine learning theory, and is mostly used in prediction problem. In HaFG, we transfer prediction problem, a global function, into product of factors which weigh the effect of different features to the prediction result. The relationships between local function factors consist of a graph structure. We first define the following three factor functions.

(1) Attribute factor: $f(y_i^t, \mathbf{x}_i^t)$ represents the posterior probability of the healthy habit forming state $y_i^t$ given the the attribute vector $\mathbf{x}_i^t$. Specifically, $y_i^t$ represents the state of subject $v_i$ at time $t$, and $\mathbf{x}_i^t$ refers to the attributes of subject $v_i$ at time $t$. In particular, we define attribute factor as follows:

$$f(y_i^t, \mathbf{x}_i^t) = \frac{1}{Z_\alpha} exp\{\alpha^T \mathbf{f}(y_i^t, \mathbf{x}_i^t)\} \qquad (2)$$

where $\alpha$ is a weighting vector, $Z_\alpha$ is a normalization factor and $\mathbf{f}$ is a vector of feature functions.

(2) Intervention Factor: g(y,G) denotes the correlation between intervention activities, where G(y) is the set of correlated relationships to y. That is, g(y,G) denotes the correlation between nodes via the edge on the graph model, which represents the healthy habits intervention between two subjects in our model. In particular, the intervention factor is defined as follows.

$$g(y_i^t, G(y_i^t)) = \frac{1}{Z_\beta} exp\{\sum_{y_j^t \in G(y_i)} \beta^T \mathbf{g}(y_i^t, y_j^t)\} \qquad (3)$$

where $\beta$ is a weighting vector, $Z_\beta$ is a normalization factor and $\mathbf{g}(y_i^t, y_j^t)$ can be defined as a vector of indicator functions.

(3) State Correlation Factor:

$$h(y_i^t, H(y_i^{t-1})) = \frac{1}{Z_\gamma} exp\{ \sum_{y_i^{t-1} \in H(y_i^t)} \gamma^T \mathbf{h}(y_i^t, y_i^{t-1}) \} \quad (4)$$

where $\gamma$ is a weighting vector, $Z_\gamma$ is a normalization factor and $H(y_i^{t-1})$ is the set of correlated relationships to $y_i^{t-1}$.

Based on Equations (1)-(4), the joint distribution $p(Y|G)$ is

$$p(Y|G) = \prod_i f(y_i^t, \mathbf{x}_i^t), g(y_i^t, G(y_i^t)), h(y_i^t, H(y_i^{t-1})) \quad (5)$$

Learning the HaFG model is to estimate a optimum parameter configuration $\theta = (\alpha, \beta, \gamma)$ to maximize the log-likelihood function of $p(Y)$. Since the maximization of log-likelihood cannot be directly calculated, we employ the Loopy Belief Propagation(LBP) to approximate marginal probabilities $p(y_i|\theta)$ and $p(y_i, y_j|\theta)$ according to [5].

## 5. EXPERIMENTS

To evaluate our work, we implement a C/S structured prototype system with an Android application as client. The raw sensor data are collected and all the intervention functionalities defined in section 2 are implemented in client application. Subjects are required to run the application on their phone in a whole day and fill out the questionnaire every morning about yesterdays' sleep and wake up time.Data analysis works are completed in the server module.

Our experiments are carried out with 30 subjects for 30 successive days, sampling phone-based context data covering social activities, user sleep-wake time records, physiological activities, surroundings, and user feedback records. During the daytime, all subjects are required to switch on the client application installed on their phone. Every morning, they also need to fill the questionnaire of sleep and wake up time. In order to ensure that subjects carefully finish the questionnaire, we validate each questionnaire submitted every day. Validation results show that along the experiments, more than 95% of the questionnaires are filled out carefully by subjects. The experimental results are shown in Table 1.

**Table 1: Performance of HaFG prediction.(%)**

| Factors used | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| State | 50.42 | 66.23 | 72.31 | 69.13 |
| + Intervention | 63.22 | 81.23 | 87.23 | 84.13 |
| + Attributes | 57.26 | 62.64 | 66.62 | 64.56 |
| All | 67.26 | 85.86 | 84.65 | 85.25 |

### 5.1 Discussion

We evaluate the contribution of different factors defined in our model as follows. We first remove all the other factors and keep state correlation factor only, and then add each of the other factors into the model. Then we evaluate the performance improvement one by one. We find that each factor is useful but the contribution is different. Specifically, only using previous state feature, the accuracy achieves about 50%. After we add intervention feature, the accuracy increases 13%, that reveals the fact that state factor have a relationship with habit forming process and help improving the prediction accuracy. The attributes factor can also help

improve the prediction performance but not strong as interventions factor. By adding all the features into our assessing model together, it achieves a fairly good performance(67%). Note that in fact, there also some problems exist in our experiments which will bring uncertainties. For example, the phone sensor system is limited to provide a sensor data in very high accuracy, which may affect the performance of attribute feature. Additionally, there are some other factors possibly affecting the habit forming process which are not obtained by our system.

## 6. CONCLUSIONS

In this paper, we propose a novel approach to predict healthy habit forming states using mobile phone. We first define the habit forming process based on psychology theory. We extract several features from mobile social network and mobile phone sensor data, and then propose a HaFG model for forecasting habit forming states using these features. We train our model using these context features, the experimental result shows that the proposed HaFG model can significantly improve the performance for prediction. For different people, the effects of intervention might be different. So in future we will focus on some special groups of people such as the young people to help them developing healthy habits using mobile phone.

## 7. ACKNOWLEDGMENTS

## 8. REFERENCES

[1] N. Aharony. *Social fMRI: Measuring, Understanding, and Designing Social Mechanisms in the Real World.* PhD thesis, Massachusetts Institute of Technology, 2011.

[2] Y. Bai, B. Xu, Y. Ma, G. Sun, and Y. Zhao. Will you have a good sleep tonight?: sleep quality prediction with mobile phone. In *BodyNets 2012*, pages 124–130, 2012.

[3] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transaction on Information Theory*, 47:498–519, Feb. 2001.

[4] Y. Ma, B. Xu, Y. Bai, G. Sun, and R. Zhu. Daily mood assessment based on mobile phone sensing. In *BSN 2012*, pages 142–147, 2012.

[5] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *UAI*, pages 467–475, 1999.

[6] R. Schwarzer. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, 57(1):1–29, 2008.

[7] R. Schwarzer and A. Luszczynska. How to overcome health-compromising behaviors. *European Psychologist*, 13(2):141–151, 2008.

[8] S. Yajima, T. Takano, K. Nakamura, and M. Watanabe. Effectiveness of a community leaders' programme to promote healthy lifestyles in tokyo, japan. *Health Promotion International*, 16(3):235–243, 2001.