2012 IEEE 12th International Conference on Data Mining

Socialized Gaussian Process Model for Human Behavior Prediction in a Health Social Network

Yelong Shen
Department of Computer Science
Kent State University
yshen@cs.kent.edu

Ruoming Jin
Department of Computer Science
Kent State University
jin@cs.kent.edu

Dejing Dou, Nafisa Chowdhury
Computer and Information Science
University of Oregon
{dou,nafisa}@cs.uoregon.edu

Junfeng Sun
National Institutes of Health
junfeng.sun@nih.gov

Brigitta Piniewski
PeaceHealth Laboratories
BPiniewski@peacehealthlabs.org

David Kil
HealthMantic, Inc
david.kil@healthmantic.com

***Abstract*—Modeling and predicting human behaviors, such as the activity level and intensity, is the key to prevent the cascades of obesity, and help spread wellness and healthy behavior in a social network. In this work, we propose a Socialized Gaussian Process (SGP) for socialized human behavior modeling. In the proposed SGP model, we naturally incorporates human's personal behavior factor and social correlation factor into a unified model, where basic Gaussian Process model is leveraged to capture individual's personal behavior pattern. Furthermore, we extend the Gaussian Process Model to socialized Gaussian Process (SGP) which aims to capture social correlation phenomena in the social network. The detailed experimental evaluation has shown the SGP model achieves the best prediction accuracy compared with other baseline methods.**

I. Introduction

Obesity and overweight is becoming a major problem facing United States and across the world. Recent studies have shown obesity can spread over the social network [2], bearing similarity to the diffusion of innovation [3] and word of mouth effects in marketing [4]. Though the common belief is that the spreading comes from the social and cultural influence of poor health habits such as lack of exercise and fast food consumption, there has been lacking of scientific and quantitative study to elucidate how social relationship may contribute to macro-level human behaviors.

Recent advances in mobile technology and online social networks provide new opportunities to support healthy behaviors through lifestyle monitoring and online communities. Mobile devices can track and record the walking/jogging/running distance and intensity of an individual; social networks can help people to interact and participate various physical activities and exercise. Utilizing these technology, the recent YesIWell study conducted in 2010-2011 as collaboration between PeaceHealth Laboratories, SK Telecom Americas and University of Oregon to record the social network and daily physical activities for a group of 254 individuals for a period of ten months. The fundamental problems this study seeks to answer, which are also the key in understanding the determinants of human behavior, are as follows:

1. *Could social affiliations affect individual behaviors?*
2. *How can we leverage social network to help predict individual's behaviors?*

However it is nontrivial to determine that how much impact social influence could have on one's behavior. Human behaviors are constantly changing due to various reasons, including: *personal factor* (one person did not do any exercise in every Thursday this term because he took lots of courses on Thursday); *social correlation factor* (individuals tend to do sports together with their friends, i.e., playing basketball, rather than playing it alone), and *social influence factor* (one person does not like sports at beginning, but he will try to play basketball after most of his friends play everyday).

In the paper, we propose a novel Socialized Gaussian Process (SGP) for socialized human behavior modeling. The proposed Socialized Gaussian Process model naturally incorporates personal behavior factor and social correlation factor into a unified model, and social influence factor is implicitly modeled by updating the social correlation coefficient dynamically. we also devise an online prediction/inference scheme for the SGP model to predict individual's future behavior by learning from historical records and social network information.

II. Problem Definition and Socialized Human Behavior Prediction

In this study, we consider two major data sources: 1) the social network $\mathcal{G} = (V, E)$, where $(i, j) \in E$ indicate users u_i and u_j are friends. Here, we consider the friend relationship is mutual and thus $\mathcal{G}$ is an undirected graph. 2) time series users' behaviors $\mathcal{X}$, where user u_i's sequential behavior is denoted as $\mathcal{X}^i = (x_1^i, x_2^i, ...x_T^i)$ with $x_t^i \in (-1, 0, 1)$. where $x_t^i = 1$ indicates user u_i does sports at day t, $x_t^i = -1$ indicates user u_i does not do sports at day t, while $x_t^i = 0$ indicates user u_i's record is missing at day t. Then, we define the *socialized human behavior prediction problem* as follows:

1550-4786 2012

DOI 10.1109/ICDM.2012.94

Given the social network $\mathcal{G}$ and individuals past behaviors until day t, $\mathcal{X}_{1..t} = (X^1_{1..t}, X^2_{1..t}, ..., X^N_{1..t})$, where $X^i_{1..t} = (x^i_1, x^i_2, ...x^i_t)$ and N is the number of users in the social network, socialized human behavior prediction problem is to predict the individuals behaviors at day $t+1$, i.e. $\mathcal{X}_{t+1}$.

Figure 1 illustrates the snapshots of user behaviors and social network at different points in time.

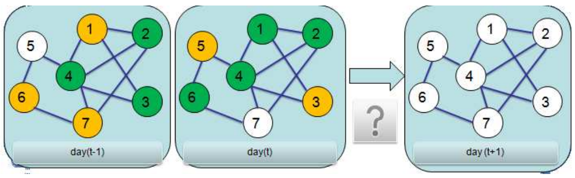

Figure 1: Snapshots of user behaviors and social network. The cycle with green color indicates the user does exercises at the time point, The cycle with yellow color indicates the user does not do exercises at the time point. White color indicates the user behavior unknown.

As illustrated in the figure, individual's future behaviors can be predicted from three aspects. First, **Personal behavior pattern:** personal behavior pattern is computed based on individuals' past behaviors, i.e., calculating the autocorrelation function. Because most of individuals have their own regular behavior cycle. Individual's historical behavior records can be leveraged for predicting his/her future behaviors. Second, **Social Correlation:** social correlation indicates that individuals tend to perform the same behaviors with their friends (homophily principle). Third, **Social Influence:** Social influence indicates that individual's behavior can also be influenced by his/her friends' past behaviors. In general, the socialized human behavior prediction problem could be formulated as:

$$(X^1_t, X^2_t, ..., X^N_t) = \mathcal{X}_t \sim p(\mathcal{X}_t|\mathcal{X}_{1:t-1}). \tag{1}$$

Considering there are three types of conditional independent factors: Personal factor, Social Correlation factor and Social Influence factor affect individual's future behavior, the formulation 1 can be rewritten as follows:

$$p(\mathcal{X}_t|\mathcal{X}_{1:t-1}) \propto \prod_{i=1..N} p(X^i_t|X^i_{1..t-1}; \theta_i) \prod_{i=1..N} p(X^i_t|X^i_{1..t-1}, X^{F(i)}_{1..t-1}; \phi_i) \prod_{(i,j)\in\mathcal{G}} p(X^i_t, X^j_t; \Omega) \tag{2}$$

where notation $F(i)$ represents the neighbors of user u_i in the social network. The posterior probability has three types of factor functions, corresponding to the intuitions. Specially personal factor $p(X^i_t|X^i_{1..t-1}; \theta_i)$ represents the posterior probability of u_i's behavior X^i_t at time t given the past behavior records $X^i_{1..t-1}$, θ_i is the parameter; social correlation factor $p(X^i_t, X^j_t; \Omega)$ reflects the correlation between pair of friends' behavior at time t, Ω is the social correlation parameter; social influence factor $p(X^i_t|X^i_{1..t-1}, X^{F(i)}_{1..t-1})$ indicates friends' influence on u_i's behavior at time t, where ϕ_i is the influence parameter;

Then we introduce a latent space $\mathcal{S}(S^i_t \in \mathcal{S})$, where S^i_t assumes to be the sufficient statistics for $X^i_{1..t}$. Thus, the Eqn. 2 can be rewritten,

$$p(S_t|S_{t-1}) = \frac{1}{Z} exp(\sum_{i=1..N} \theta_i f(S^i_t, S^i_{t-1}) + \sum_{i=1..N} \phi_i g(S^i_t, S^i_{t-1}, S^{F(i)}_{t-1}) + \sum_{(i,j)\in\mathcal{G}} \Omega_{ij} k(S^i_t = S^j_t)). \tag{3}$$

where Z is the partition function (normalization factor), the feature functions f, g, k correspond to personal factor, social influence factor and social correlation factor.

The general social influence model is intractable due to the huge number of social influence parameters. We incorporate the social influence factor implicitly by modeling the dynamic of social correlation. In the paper, we model the dynamic of social correlation by simply updating the correlation coefficients along the time.

III. Socialized Gaussian Process Model

In this section, we propose a Socialized Gaussian Process (SGP) model which incorporates both personal factor and social correlation factor into a unified model. First, we give the formulation of Social Correlation based social influence (Implicit) model as follows:

$$p(X_t|X_{1:t-1}; \Theta, \Omega) \propto \int_{S_t, S_{t-1}} p(X_{1:t-1}|S_{t-1}) p(X_t|S_t) p(S_t|S_{t-1}; \Theta, \Omega) \tag{4}$$

$Where$

$$p(S_t|S_{t-1}; \Theta, \Omega) = \frac{1}{Z} exp(\sum_{i=1..N} \theta_i f(S^i_t, S^i_{t-1}) + \sum_{(i,j)\in\mathcal{G}} \Omega_{ij,t} k(S^i_t = S^j_t)). \tag{5}$$

$$\Omega_t \sim p(\Omega_t|\Omega_{t-1}, S_{t-1}). \tag{6}$$

In the general Social Correlation (Implicit) Social Influence model, we do not explicitly incorporate social influence factor in Eqn. 5. However the influence factor is implicitly captured by modeling the dynamic of social correlation in Eqn. 6. The parameters Ω_t for social correlation factor at time t can be viewed as a weighted graph, where $\Omega_{ij,t}$ is the weight of the edge (i, j) measures the correlation between u_i and u_j at time t.

For parameters Ω_t, we employ a temporally updating scheme to update Ω_t online. In the prediction phrase, Ω_t is estimated by previous social correlation $\Omega_t = \Omega_{t-1}$. In the updating phrase, Ω_t is updated according to the following formulation:

$$\Omega_{ij,t} = \frac{1}{W} exp(-\sum_{m=t-W..t} (X^i_m - X^j_m)^2); \tag{7}$$

A. Gaussian Process Model

Now we introduce how to estimate the emission probability $p(X_{1:t-1}|S_{t-1})$ and $p(X_t|S_t)$, and how to specify the dimension of latent space S ($S_t^i \in S$).

If latent space $\mathcal{S}$ is defined with K dimensional discrete states, S_t would be in K^N (N is the number of users in social network). The proposed model is still intractable, since there is no closed form to integral out S_t and S_{t-1} in the Eqn. 4.

Therefore, a nonparametric bayesian approach, gaussian process is employed to make the model tractable. Here latent space $\mathcal{S}$ is defined in a continuous real value space R. Then we define the emission probability $p(X_t|S_t)$ as follows:

$$p(X_t|S_t) = \prod_{i=1..N} \Phi(X_t^i * S_t^i) \tag{8}$$

where probit function Φ is the cumulate gaussian function, $\Phi(x) = \int_{-\infty}^{x} \mathcal{N}(\tau|0,1)d\tau$, where $\Phi(x) + \Phi(-x) = 1$.

Now S_{t-1}^i is a single real value in R, it could not be sufficient statistics for individual's past behaviors. We rewrite the Eqn. 4 and 5, by replacing S_{t-1}^i with $S_{1:t-1}^i$, i.e., $p(X_t|X_{1:t-1};\Theta,\Omega) =$

$$\int_{S_t,S_{1:t-1}} p(X_{1:t-1}|S_{1:t-1})p(X_t|S_t)p(S_t|S_{1:t-1};\Theta,\Omega) \tag{9}$$

$$p(S_t|S_{1:t-1};\Theta,\Omega) = \frac{1}{Z}exp(\sum_{i=1..N} f(S_t^i, S_{1:t-1}^i, \theta_i) + \sum_{(i,j)\in\mathcal{G}} -\Omega_{ij,t}(S_t^i - S_t^j)^2)). \tag{10}$$

where feature function $f(S_t^i, S_{1:t-1}^i, \theta_i)$ can be explicitly given by gaussian process model,

$$f(S_t^i, S_{1:t-1}^i, \theta_i) = -\begin{pmatrix} S_{1:t-1}^i \\ S_t^i \end{pmatrix}^T \begin{pmatrix} K_{1:t-1}^i & K_{(1:t-1),t}^i \\ {K_{(1:t-1),t}^i}^T & K_t^i \end{pmatrix} \begin{pmatrix} S_{1:t-1}^i \\ S_t^i \end{pmatrix} \tag{11}$$

where K^i is the covariance (kernel) matrix, with its element $K_{t,t'}^i$ defined by covariance function $K_{t,t'}^i = C(S_t^i, S_{t'}^i, \theta_i)$. Supposing $S_{1..t}^i$ is the stationary process, the covariance function is defined by the individual's behavior autocorrelation kernel,

$$C(S_t^i, S_{t'}^i, \theta_i) = \theta_1^i exp(mod((t-t'), cycle^i(t)) * \theta_3^i) + \theta_2^i exp((t-t')/cycle^i(t) * \theta_4^i) + \theta_5^i * I(t = t'). \tag{12}$$

where $cycle_t^i$ is computed by autocorrelation function(ACF),

$$cycle^i(t) = argmax_\tau \Psi(\tau) = \frac{1}{t-\tau} \sum_{n=1..t-\tau} X_n^i * X_{n+\tau}^i \tag{13}$$

Therefore, by replacing the term $f(S_t^i, S_{1:t-1}^i, \theta_i)$ in Eqn. 10 with the formulation 11, We can obtain the closed from of condition probability over S_t given $S_{1:t-1}$ as multivariate gaussian distribution. (Here we omit the mathematical details)

$$p(S_t|S_{1:t-1};\Theta,\Omega) = \mathcal{N}(S_t|\mu_t, \nu_t) \tag{14}$$

where the mean and covariance μ_t, ν_t are given as follows,

$$\begin{aligned} \mu_t = (\mathcal{V}_t + \mathcal{L}_t)^{-1}\mathcal{V}_t\mathcal{U}_t \qquad \nu_t &= (\mathcal{V}_t + \mathcal{L}_t)^{-1} \\ \mathcal{U}_t &= ({K_{(1:t-1),t}^i}^T {K_{t-1}^i}^{-1}(S_{1:t-1}^i))_{i=1..n}^T \\ \mathcal{V}_t = diag\{K_t^i - {K_{(1:t-1),t}^i}^T &{K_{t-1}^i}^{-1} K_{(1:t-1),t}^i\}_{i=1..n} \\ \mathcal{L}_{t,ii} = \sum_{j=1..n} \Omega_{ij,t} \qquad &\mathcal{L}_{t,ij} = -\Omega_{ij,t} \end{aligned} \tag{15}$$

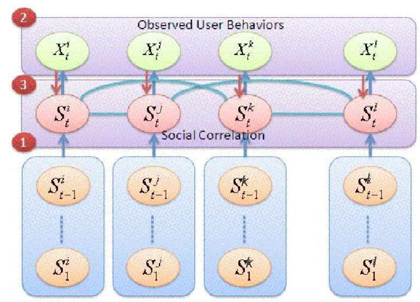

Figure 2: Social influence Gaussian Process for human behavior prediction

We also developed an online prediction/inference phases to verify the effective of the proposed socialized gaussian process model. In the online inference phase, the posterior probability of latent state S_t in the model is estimated given the observation of individuals' behavior X_t at time t. In the online prediction phase, we estimated the condition probability of X_t given the past individuals' latent state $S_{1..t-1}$. In the Figure 2, the graphical model for the socialized gaussian process is given, and the online prediction/inference phases are also illustrated. The first step is to calculate the condition probability of S_t given S_{t-1}. The second step is the online predicting phase. The third step is the online inference phase, which updates S_t by estimating the posterior probability of S_t given observation X_t.

Online Prediction: In the online prediction phase, socialized gaussian process model predicts the individuals' future behavior X_t at time t by estimating the condition probability of X_t given $S_{1..t-1}$, i.e., $p(X_t|S_{1:t-1};\Theta,\Omega,\alpha)$

$$\begin{aligned} &\propto \int_{S_t} p(S_t|S_{1:t-1}) \prod_{i=1..N} \Phi(X_t^i S_t^i) \\ &\propto \prod_{i=1..N} \Phi(\frac{X_t^i \mu_t^i}{\sqrt{\nu_t^i}}) \end{aligned} \tag{16}$$

where μ_t^i and ν_t^i are given in the Eqn. 15.

Online Inference: In the online inference phase, the posterior probability over S_t given the observation of individuals' behavior X_t is estimated,

$$p(S_t|X_t, S_{1:t-1}; \Theta, \Omega) \propto p(S_t|S_{1:t-1}) \prod_{i=1..N} \Phi(X_t^i S_t^i) \tag{17}$$

Unfortunately, the posterior is non-Gaussian. In practice, the first two moments of S_t are often used to construct a Gaussian approximation. Here, our approximation is based on a variational approach known as Adaptive Density Filter (ADF) [7]. Given a posterior distribution for S_t, ADF finds a Gaussian approximation that matches the first two moments of S_t. Specifically, let $\mathcal{N}(S_t; \mu_t^*, \nu_t^*)$ be the target gaussian, whose parameters μ_t^*, ν_t^* are chosen to minimize the Kullback Leibler divergence:

$$KL(\prod_{i=1..N} \Phi(X_t^i S_t^i)\mathcal{N}(S_t; \mu_t, \nu_t)||\mathcal{N}(S_t; \mu_t^*, \nu_t^*)) \tag{18}$$

The optimization problem can be solved by moment matching up to the second order, yielding:

$$\mu_t^{*i} = \mu_t^i + X_t^i \sqrt{\nu_t^i}\mathcal{V}(\frac{\mu_t^i X_t^i}{\sqrt{\nu_t^i}}). \tag{19}$$

$$\nu_t^{*i} = \nu_t^i[1 - \mathcal{W}(\frac{\mu_t^i X_t^i}{\sqrt{\nu_t^i}})]. \tag{20}$$

with $\mathcal{V}(x) = \frac{\mathcal{N}(x;0,1)}{\Phi(x)}, \mathcal{W}(x) = \mathcal{V}(x)(\mathcal{V}(x) + x)$.

Model Parameters: The model parameters $\Lambda = \{\Theta, \Omega\}$, Θ in the definition of covariance matrix in Eqn. 12 are hyper-parameters for personal factor modeling, which serve as hyper-parameters in gaussian process. We adopt a cross-validation method to determine Θ instead of performing learning strategy to infer hyper-parameters Θ. Since we found that setting the hyper-parameters manually could also achieve comparable performance.

IV. Experiments

In this section, we use the real world (YesIWell) user behavior data and the corresponding social network to empirically validate the effectiveness of the Socialized gaussian process model. We first elaborate on the experiment configurations on the data set, evaluation metrics and baselines. Then, we introduce the experimental results on individuals' behavior prediction, and how it could help improve prediction accuracy.

A. Experiments Configuration

Human Physical Activities Dataset. The YesIWell study is conducted in 2010-2011 as collaboration among several health laboratories and universities to help people maintain active lifestyles and lose weight. The dataset is collected from 254 users, including personal information, social network, and their daily physical activities in ten months from October 2010 to August 2011.

The initial physical activities data, collected by a special electronic equipment for each user, is the number of steps and running distances. In our study, we transform the physical activities into the value $\{1, -1\}$, which indicates the user do exercises or not. In the YesiWell dataset, there are total 40 weeks' records for individual's daily activities. Since in the dataset, some users' daily records would be missing. Thus, we first give the basic analysis on the distribution of physical activities record number in figure 3. As shown in the figure 3, there are 14 users with their daily physical activities record number smaller than 10, and 8 users with their records number larger than 10 but smaller than 20. Therefore, to clean the data, we filtered the users whose daily physical activities record number is smaller than 80. We only use the rest 185 users for experiments. The general statistics of the users' daily physical activities data are shown in figure 4, which shows the distribution of sports ratio [1] and user number. i.e., Fifteen users do exercises in nearly 20 percent of their daily physical activities records.

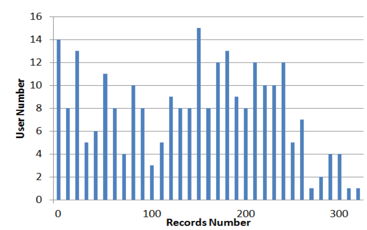

Figure 3: Distribution of the record number and user number

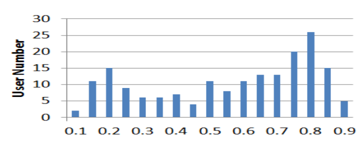

Figure 4: Distribution of the sports ratio and user number

The dataset contains 684 connections in the social network consisting of the 185 individuals, on average each individual has 4 friends in the social network. Figure 5 shows the distribution of friends number in the social network.

Evaluation metric. To verify the effectiveness of the proposed socialized gaussian process model, we conduct the experiments by predicting the future individual's physical activities according to their past behaviors and social network information. In the experiment, we select two weeks as the unit for prediction, i.e., leverage the previous 10 weeks daily records to predict the 11th and 12th weeks users' behaviors. We use the metric **accuracy** to measure the prediction quality between week t and $t+1$.

[1] sports ratio is defined as the percentage of days user exercised

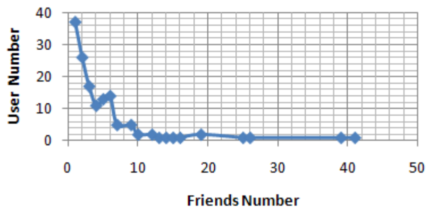

Figure 5: Distribution of friends number in the social network

$$accuracy = \frac{\sum_{i=1..N}\sum_{d\in\{t,t+1\}} I((X_d^i \neq 0) = \widetilde{X}_d^i)}{\sum_{i=1..N}\sum_{d\in\{t,t+1\}} I(X_d^i \neq 0)} \tag{21}$$

where X_d^i is the true user activity at day d for u_i, $\widetilde{X}_d^i$ denotes the predicted value. $(X_d^i \neq 0)$ indicates the physical activity record is not missing. I is the indication function, where $I(X) = 1$ when X is true, otherwise $I(X) = 0$. N is the number of users.

Baselines. We compare the socialized gaussian process model with several ad-hoc methods for individual behavior prediction. Typically, the baseline methods are divided into two categories: personalized behavior prediction method and socialized behavior prediction method. Personalized method only leverages individual's personal past behavior records for future behavior prediction. Socialized method do not only use personal behavior record, but also incorporate friends' past behavior for prediction. Specifically, the following five prediction models are compared:

Behavior Pattern Search: Behavior Pattern Search(BPS) is an ad-hoc method for personalized behavior prediction. The main idea of BPS is searching the most similar behavior pattern from past behavior log for predicting future behavior. i.e.,

$$\begin{aligned} \widetilde{X}_d^i = X_{d-p}^i, where \\ p = argmax_{p=1..W1} \sum_{m=1..W2} X_{d-m}^i X_{d-m-p}^i \end{aligned} \tag{22}$$

where $W1$ is the size of the search window, and $W2$ is the size of the window pattern. In our experiments, $W1$ and $W2$ are set to 60 and 15 respectively could achieve best performance.

Logistical Autoregression: Logistical Autoregression (LAR) [1] utilizes logistical regression method to leverage historical activities for predicting future behaviors. i.e. $p(\widetilde{X}_d^i = 1) = logit(\alpha_0 + \sum_{m=1..W} \alpha_m X_{d-m}^i)$, where m is the lag length for autoregression (set to be 9 for all users), which is the determined by cross-validation. Results reported in all experimental results are based on the parameter configurations which produce the best results.

Socialized Logistical Autoregression: Socialized Autoregression (SLAR) method borrows the idea from [6], which combines the social influence and autoregression model together. SLAR models the social influence explicitly by incorporating friends' historical behaviors into the unified regression model, i.e.,

$$\begin{aligned} p(\widetilde{X}_d^i = 1) = logit(\alpha_0 + \sum_{m=1..W1} \alpha_m X_{d-m}^i + \\ \sum_{f\in F(i)} \pi_f^i \sum_{m=1..W2} \beta_m X_{d-m}^f) \end{aligned} \tag{23}$$

where parameters π_f^i indicates to what degree friend u_f's past behavior could effect u_i's behavior. $W1$ and $W2$ are the two lag length, which determined by cross-validation.

Personalized Gaussian Process: Personalized Gaussian Process (PGP) model does not incorporate the social network information, which purely utilized personal historical records for prediction. The observation of individual's behavior is discrete, especially binary. Thus, Binary Gaussian process used in our experiments [5]. In the setting of the personalized gaussian process, the covariance (kernel) matrix is defined as in Eqn. 17. The hyperparameters Θ in PGP is determined by cross-validation.

Socialized Gaussian Process: Socialized Gaussian Process (SGP) model incorporates both personalized historical behavior records and also dynamic social correlation information for future behavior prediction. In the SGP model, parameters α, in Eqn. 10 are estimated by solving the linear regression problem. The hyperparameters Θ is determined by validation like in PGP.

B. *Experiment Results*

In this subsection, we report the performance of different human behavior model for predicting individual's future behavior. The individuals' behavior records are divided according to time series. i.e., $T1 - T8$ indicates the records from the first week to the eighth week. Thus, we could evaluated the models in different time periods. As shown in the table I, We give the accuracy comparison with the five methods, Behavior Pattern Search (BPS), Logistical Autoregression (LAR), Socialized Logistical Autoregression (SLAR), Personalized Gaussian Process (PGP) and Socialized Gaussian Process (SGP). The columns in the table indicates different time period, i.e., $T17 - T18$ indicate individuals' behavior from $17th$ week to $18th$ week predicted.

As shown in Table I , Socialized Gaussian Process Model consistently outperforms the other four methods. The only exception is at week $T19 - T20$, PGP achieved a slightly better performance than SGP. The performance of Pattern Search method (BPS) generally is the consistently worst among all these method. One explanation is that it sensitive to the noisy data. It is interesting to notice that the performance of logistical Autoregression (LAR) is comparable with socialized logistical autoregression (SLAR) model which additionally incorporates friends' information into the model. Both LAR and SLAR method get worse performance than personalized Gaussian process model. There are two reasons to explain

Table I: Prediction Accuracy Comparison with different models(T9-T22)

Model	T9-T10	T11-T12	T13-T14	T15-T16	T17-T18	T19-T20	T21-T22
BPS	0.6435	0.6347	0.6965	0.6993	0.6982	0.7058	0.6808
LAR	0.6946	0.6583	0.7435	0.7367	0.7388	0.7421	0.7096
SLAR	0.6837	0.6646	0.7421	0.7343	0.7371	0.7432	0.7084
PGP	0.6707	0.7080	0.7717	0.7428	0.7509	**0.7640**	0.7402
SGP	**0.6983**	**0.7376**	**0.7902**	**0.7524**	**0.7583**	0.7628	**0.7431**

Table II: Paired t-Test(2-tail) results

	PGP	SLAR	LAR	BPS
SGP	$1.91e-3$	$2.86e-4$	$4.23e-4$	$8.75e-8$

why SLAR method is unsuccessful in our experiments. First, incorporating additional parameters would increase the model complexity, and also increase the risk of being overfitting. Second, for human's daily behavior, friends would be willing to do together at the same time, rather than following friends' previous behaviors.

In the experiments, BPS, LAR and PGP are all personalized behavior prediction models. However, PGP significantly outperforms the other two methods by exploiting individual's personal behavior records information. Socialized Gaussian Process (SGP) achieves further improvement based on PGP by incorporating the dynamic social correlation information. In terms of accuracy, SGP method improves the performance averagely as high as 1.79%, 4.24%, 4.45% and 10.04% in contrast to PGP, SLAR, LAR and BPS, respectively.

Finally, to validate the statistical significance of our experiments, we perform the paired t-test (2-tail) over the accuracy of the experiential result. As shown in Table II, all the t-test results are less than 0.01, which means the improvements of SGP over other methods are statistically significant.

V. Conclusion and Future Work

In this paper, we have developed a new social influence model, referred as Socialized Gaussian Process (SGP) for socialized human behavior modeling in a health social network. The detailed experimental evaluation has demonstrated the effectiveness and accuracy of the SGP model. In the future work, we plan to expand this model to study how the biometrics and biomarkers, such as BMI, HDL, LDL, triglycerides, etc., can be predicted through the physical activities. We also plan to study whether this SGP can be used (adopted) to model the general time series and network structure data, such as biological and/or economical system.

Acknowledgment

The authors appreciate the anonymous reviewers for their extensive and informative comments to help improve the paper. Yelong Shen and Ruoming Jin's work in the paper is partially supported by National Science Foundation under CAREER Award IIS-0953950. Dejing Dou and Nafisa Chowdhury's research is partially supported by the NSF grant IIS-1118050 and the NIH grant R01EB007684. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NSF or NIH, or the U.S. Government.

References

[1] Barry C. Arnod and C. A. Robertson. Autoregressive logistical regression. *Journal of Applied Probabilty*, 26:524–531, 1989.

[2] Nicholas A. Christakis and James H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.

[3] Robert G. Fichman and Chris F. Kemerer. The illusory diffusion of innovation: An examination of assimilation gaps. *Information Systems Research*, 10:255–275, 1999.

[4] Junming Huang, Xue-Qi Cheng, Hua-Wei Shen, Tao Zhou, and Xiaolong Jin. Exploring social influence via posterior effect of word-of-mouth recommendations. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 573–582, New York, NY, USA, 2012. ACM.

[5] Malte Kuss and Carl Edward Rasmussen. Assessing approximate inference for binary gaussian process classification. *J. Mach. Learn. Res.*, 6:1679–1704, December 2005.

[6] Wei Pan, Manuel Cebrian, Wen Dong, Taemie Kim, James Fowler, and Alex Pentland. Modeling dynamical influence in human interaction patterns. *Work*, abs/1009.0:19, 2010.

[7] David H. Stern, Ralf Herbrich, and Thore Graepel. Matchbox: large scale online bayesian recommendations. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 111–120, New York, NY, USA, 2009. ACM.