Ontology-based deep learning for human behavior prediction with explanations in health social networks

Nhathai Phan[a], Dejing Dou[a,*], Hao Wang[a], David Kil[b], Brigitte Piniewski[c]

[a] CIS Department, University of Oregon, Eugene, OR, USA
[b] HealthMantic, Inc., Los Altos, CA, USA
[c] PeaceHealth Laboratories, Vancouver, WA, USA

ABSTRACT

Human behavior modeling is a key component in application domains such as healthcare and social behavior research. In addition to accurate prediction, having the capacity to understand the roles of human behavior determinants and to provide explanations for the predicted behaviors is also important. Having this capacity increases trust in the systems and the likelihood that the systems actually will be adopted, thus driving engagement and loyalty. However, most prediction models do not provide explanations for the behaviors they predict.

In this paper, we study the research problem, *human behavior prediction with explanations*, for healthcare intervention systems in health social networks. We propose an ontology-based deep learning model (**ORBM**[+]) for human behavior prediction over undirected and nodes-attributed graphs. We first propose a bottom-up algorithm to learn the user representation from health ontologies. Then the user representation is utilized to incorporate *self-motivation*, *social influences*, and *environmental events* together in a human behavior prediction model, which extends a well-known deep learning method, the Restricted Boltzmann Machine. ORBM[+] not only predicts human behaviors accurately, but also, it generates explanations for each predicted behavior. Experiments conducted on both real and synthetic health social networks have shown the tremendous effectiveness of our approach compared with conventional methods.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Being overweight or obese is a major risk factor for a number of chronic diseases, including diabetes, cardiovascular diseases, and cancers. Once considered a problem only in high-income countries, overweight and obesity are now dramatically on the rise in low- and middle-income countries. Recent studies have shown that obesity can spread over the social network [5], bearing similarity to the diffusion of innovation [7] and word-of-mouth effects in marketing [12]. To reduce the risk of obesity-related diseases, regular exercise is strongly recommended, i.e., at least 30 min of moderate-intensity physical activity on 5 or more days a week [21]. However, there have been few scientific and quantitative studies to elucidate how social relationships and personal factors may contribute to macro-level human behaviors, such as physical exercise.

---

* Corresponding author.
   E-mail addresses: haiphan@cs.uoregon.edu (N. Phan), dou@cs.uoregon.edu (D. Dou), hao@cs.uoregon.edu (H. Wang), david.kil@healthmantic.com (D. Kil), BPiniewski@peacehealthlabs.org (B. Piniewski).

The Internet is identified as an important source of health information and may thus be an appropriate delivery vector for health behavior interventions [19]. In addition, mobile devices can track and record the distance and intensity of an individual's walking, jogging, and running. We utilized these technologies in our recent study, named YesiWell [24], conducted in 2010–2011 as a collaboration between PeaceHealth Laboratories, SK Telecom Americas, and the University of Oregon to record daily physical activities, social activities (i.e., text messages, social games, events, competitions, etc.), biomarkers, and biometric measures (i.e., cholesterol, triglycerides, BMI, etc.) for a group of 254 individuals. Physical activities were reported via a mobile device carried by each user. All users enrolled in an online social network, allowing them to befriend and to communicate with each other. Users' biomarkers and biometric measures were recorded via monthly medical tests performed at our laboratories. The fundamental problems this study seeks to answer, which are also the key in understanding the determinants of human behaviors, are as follows:

- How could social communities affect individual behaviors?
- Could we illuminate the roles of social communities and personal factors in shaping individual behaviors?
- How could we leverage personal factors and social communities to help predict an individual's behaviors?
- Could domain knowledge, e.g., ontologies, help us predict an individual's behaviors? If yes, then how?

It is nontrivial to determine how much impact social influences could have on someone's behavior. Our starting observation is that human behavior is the outcome of interacting determinants such as *self-motivation*, *social influences*, and *environmental events*. This observation is rooted in sociology and psychology, where it is referred to as *human agency in social cognitive theory* [1]. An individual's self-motivation can be captured by learning correlations between his or her historical and current characteristics. In addition, users' behaviors can be influenced by their friends on social networks through what are known as social influences. The effect of environmental events is composed of unobserved social relationships, unacquainted users, and the changing of social contexts [4].

Based on this observation we propose an ontology-based deep learning model, named **ORBM**$^+$, for human behavior prediction with explanations. Providing explanations for predicted human behaviors has the benefit to increase the trust in the intervention. It targets the intervention approaches to specific and truthful problems, to keep the users maintaining or improving their health status, and thus to increase the successful adaptation rate. Our model extends a well-used deep learning method, Restricted Boltzmann Machines (RBMs) [28], with domain ontologies [8]. The reason we utilize the ontologies is that they can help us generate better user representations, which is particularly important for human behavior prediction in health social networks. Another crucial reason is that common deep learning architectures, such as the RBMs [28], Convolutional Neural Networks (CNNs) [15], and Sum-Product Networks (SPNs) [25], take a flat representation of characteristics as an input, that might not reflect the domain knowledge of their differences. The characteristics are commonly in structural designs such as ontologies in the biomedical and health domain. Therefore, it would be better if a model could have the ability to learn the representations of individuals in health social networks from ontologies.

In ORBM$^+$ model, we first propose a bottom-up algorithm to learn the representation of users based on the ontologies of personal characteristics in the health domain. The key idea of our algorithm is that a representation of a concept will be learned by its own properties, the properties of its related concepts, and the representations of its sub-concepts. Our algorithm will learn a structure of representation that replicates the original structure of personal characteristics. This representation structure is further used to predict model human behaviors by modeling human behavior determinants in our health social network. Self-motivation can be captured by learning correlations between an individual's historical and current features. The effect of the implicit social influences on an individual is estimated by an aggregation function of the past of the social network. We further define a statistical and temporal smoothing function to capture social influences on individuals from their neighboring users. The environmental events such as competitions, are integrated into the model as observed variables that will directly affect the user behaviors. The effect of environmental events can be captured by learning the influences of unacquainted users and the evolving of the social network's parameters. To generate explanations for predicted behaviors, we adapt the Minimum Description Length (MDL) principle [26] to extract the key characteristics or components that cause users to do some specific behaviors. An explanation can be defined as a list of characteristics which maximize the likelihood of a behavior of a user.

Our main contributions are as follows:

- We study the research problem of *human behavior prediction with explanations* in health social networks, which is motivated by real-world healthcare intervention systems.
- We introduce **ORBM**$^+$, a novel ontology-based deep learning model, which can accurately predict and explain human behaviors.
- We propose a bottom-up algorithm to learn the individual representation, given structural designs of personal characteristics with ontologies. To our best knowledge, our algorithm is the first work to formally combine deep learning with ontologies in health informatics. It can be applied to other biomedical and health domains with ontologies available.
- An extensive experiment, conducted on real-world and synthetic health social networks, confirms the high prediction accuracy of our model and quality of explanations it generates.

In Section 2, we introduce the RBM and related works. We then introduce our YesiWell health social networks and the developed SMASH ontology in Section 3. We present our ontology-based deep learning algorithm for user representations
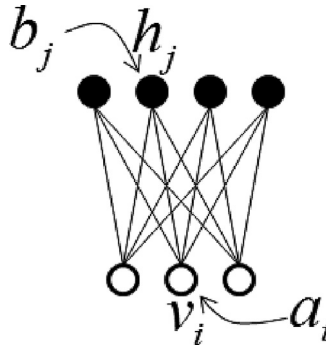
Fig. 1. The RBM.

in Section 4 and our human behavior prediction model in Section 5. The details of experimental evaluation are described in Section 6, and our work is concluded in Section 7.

2. The RBMs and related works

The Restricted Boltzmann Machine (RBM) [28] is a deep learning structure that has a layer of visible units fully connected to a layer of hidden units but no connections within a layer (Fig. 1). Typically, RBMs use stochastic binary units for both visible and hidden variables. The stochastic binary units of RBMs can be generalized to any distribution that falls in the exponential family [34]. To model real-valued data, a modified RBM with binary logistic hidden units and real-valued Gaussian visible units can be used. In Fig. 1, v_i and h_j are respectively used to denote the states of visible unit i and hidden unit j. a_i and b_j are used to distinguish biases on the visible and hidden units. The RBM assigns a probability to any joint setting of the visible units, **v** and hidden units, **h**:

$$p(\mathbf{v}, \mathbf{h}) = \frac{exp(-E(\mathbf{v}, \mathbf{h}))}{Z} \tag{1}$$

where $E(\mathbf{v}, \mathbf{h})$ is an energy function,

$$E(\mathbf{v}, \mathbf{h}) = \sum_i \frac{(v_i - a_i)^2}{2\xi_i^2} - \sum_j b_j h_j - \sum_{ij} \frac{v_i}{\xi_i} h_j W_{ij} \tag{2}$$

where ξ_i is the standard deviation of the Gaussian noise for visible unit i. In practice, fixing ξ_i at 1 makes the learning work well. Z is a partition function which is intractable as it involves a sum over the exponential number of possible joint configurations: $Z = \sum_{\mathbf{v}', \mathbf{h}'} E(\mathbf{v}', \mathbf{h}')$. The conditional distributions (with $\xi_i = 1$) are:

$$p(h_j = 1 | \mathbf{v}) = \sigma \left(b_j + \sum_i v_i W_{ij} \right) \tag{3}$$

$$p(v_i | \mathbf{h}) = \mathcal{N} \left(a_i + \sum_j h_j W_{ij}, 1 \right) \tag{4}$$

where $\sigma(.)$ is a logistic function, and $\mathcal{N}(\mu, V)$ is a Gaussian.

Given a training set of state vectors, the weights and biases in an RBM can be learned following the gradient of *contrastive divergence*. The learning rules are:

$$\Delta W_{ij} = \langle v_i h_j \rangle_d - \langle v_i h_j \rangle_r$$
$$\Delta b_j = \langle h_j \rangle_d - \langle h_j \rangle_r \tag{5}$$

where the first expectation $\langle . \rangle_d$ is based on the data distribution and the second expectation $\langle . \rangle_r$ is based on the distribution of "reconstructed" data.

To incorporate temporal dependencies into the RBM, the CRBM [30] adds autoregressive connections from the visible and hidden variables of an individual to his/her historical variables. The CRBM simulates well human motion in the single agent scenario. However, it cannot capture the social influences on individual behaviors in the multiple agent scenario. Li et al. [17] proposed the *ctRBM* model for link prediction in dynamic networks. The ctRBM simulates the social influences by adding the prediction expectations of local neighbors on an individual into a dynamic bias. However, the visible layer of the ctRBMs does not take individual attributes as input. Thus, the ctRBM cannot directly predict human behaviors.

Meanwhile, social behavior has been studied recently, such as analysis of user interactions in Facebook [33], activity recommendation [16], and user activity level prediction [37]. In [37], the authors focus on predicting users who have a

Table 1
Personal attributes.

Behaviors	#joining competitions	#exercising days	#goals set
	#goals achieved	Σ(distances)	avg(speeds)
	#meet-up	#social games	
	Encouragement	Fitness	Followup
Social Communications (the number of inbox messages)	Competition	Games	Personal
	Study protocol	Progress report	Technique
	Social network	Meetups	Goal
	Wellness meter	Feedback	Heckling
	Explanation	Invitation	Notice
	Technical fitness	Physical	
Biomarkers	Wellness Score	BMI	BMI slope
	Wellness Score slope		

tendency to lower their activity levels. This problem is known as churn prediction. Churn prediction aims to find users who will leave a network or a service. By finding such users, service providers could analyze the reasons for the intended attrition and could determine strategies to maintain such users in different applications, including online social games [13], QA forums [35], etc. Our goal is not only to predict, but also to understand the roles of human behavior determinants, and to give explanations for predicted behaviors. In [2], the authors provide the WTFW model, which generates explanations for user-to-user links, but not for human behaviors.

This paper is an extension of our conference paper published in ACM BCB 2015 [23]. The major extensions we have engaged are: (1) We have improved our previous ORBM model, not only so that it more accurately predicts human behavior, but also, so that it can generate explanations for each predicted behavior. We introduce a new temporal-smoothing social influence function to better capture the evolving of social influence over time. We further incorporate physical activity-based social influence into our function. We also introduce a new algorithm to quantitatively estimate the effects and roles of human behavior determinants in predicted behaviors. (2) An extensive experiment has been conducted on both real-world and synthetic health social networks to validate the effectiveness of our model, the roles of human behavior determinants, and the quality of generated explanations.

3. YesiWell health social network and SMASH ontology

3.1. YesiWell health social network

Our health social network dataset was collected from Oct 2010 to Aug 2011, as a collaboration between PeaceHealth Laboratories, SK Telecom Americas, and University of Oregon, to record daily physical activities, social activities (i.e., text messages, competitions, etc.), biomarkers, and biometric measures (i.e., cholesterol, BMI, etc.) for a group of 254 individuals. Physical activities, including measurements of the number of walking and running steps, were reported every 15 min via a mobile device carried by each user. As mentioned in our Introduction, all users enrolled in an online social network, allowing them to befriend and communicate with each other. Users' biomarkers and biometric measures were recorded via daily/weekly/monthly medical tests performed at home (i.e., individually) or at our laboratories. In total, we have approximately over 7 million data points of physical exercise, over 21,205 biomarker and biometric measurements, 1371 friend connections, and 2766 inbox messages. Our longitudinal study was conducted for 10 months. Albeit that such might seem a short interval, when compared with public social networks, i.e., Twitter and Facebook, our health social network is a unique, solid, and comprehensive multi-dimensional social network. The YesiWell network contains rich information from social activities, physical activities, and biomarkers and biometric measures, availing us unique access to verify statements about physical activity with recorded physical activity, and to compare statements about health with clinical measures of health.

In this paper, 33 features are taken into account (Table 1). All the features are summarized daily and weekly. The features are designed to capture the self-motivation of each user. Some of the key measures are as follows:

- *Personal ability*: BMI, fitness, cholesterol, etc.
- *Attitudes*: the number of off-line events in which each user participates, individual sending and receiving messages, the number of goals set and achieved, Wellness-score [14], etc. Wellness-score is a measure to evaluate how well a user lives their life. Being active in social activities, setting and achieving more goals, and getting higher Wellness-score illustrate a healthier attitude of a user.
- *Intentions*: the number of competitions each user joins, the number of goals set, etc. We measure intent to exercise in terms of competitions joined and goals set.
- *Effort*: the number of exercise days, walking/running steps, the distances, and speed walked/run.
- *Withdrawal*: an increase of BMI slope and/or a decrease of Wellness-score [14] indicates a negative sign regarding the self-motivation. Users may give up.
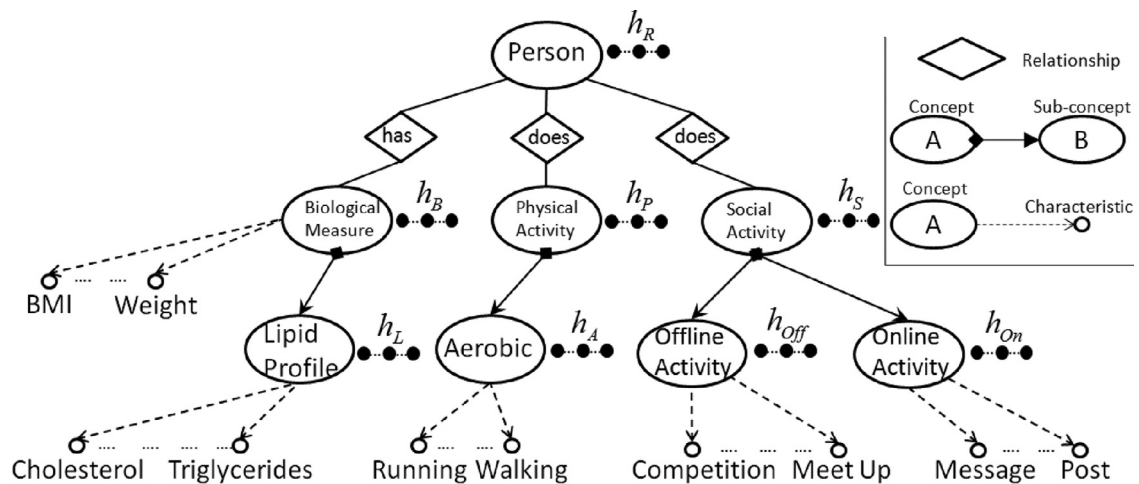
Fig. 2. Partial view of the SMASH ontology and its hidden variables.

3.2. SMASH ontology

Ontology [9,29] is the formal specification of concepts and relationships for a particular domain (e.g., genetics). Prominent examples of biomedical ontologies include the Gene Ontology (GO [31]), Unified Medical Language System (UMLS [18]), and more than 300 ontologies in the National Center for Biomedical Ontology (NCBO[1]). The encoded formal semantics in ontologies is primarily used for effective sharing and reusing of knowledge and data. They also can assist in the new research on systematic incorporation of domain knowledge in data mining, which is called semantic data mining [6].

We have developed an ontology for health social networks, in the SMASH (Semantic Mining of Activity, Social, and Health data) project, based on the YesiWell study. Our general workflow of ontology development can be described as a top-down (knowledge-driven), followed by a bottom-up (data-driven) validation and refinement approach. In the SMASH ontology, we have focused on defining concepts that are associated with sustained weight loss, especially the ones related to continued intervention with frequent social contacts. We first follow the traditional top-down design paradigm by identifying the core concepts of three modules in the SMASH system: social networks, physical activity, and health informatics. We specify the core concepts and relationships related to overweight and obesity in these modules, such as biomedical measures, trends, online and off-line events, competitions, social community, support groups, etc. In the next step, these concepts and relationships are subsequently coded in the Web Ontology Language (OWL[2]) with *Protégé*.[3] In the last step, we further validate and refine our ontology design through the data we collected from our distributed personnel devices and web-based social network platform in YesiWell.

The three modules in our SMASH ontology, biomarker measures, physical activities, and social activities, can be described as follows:

- Biomarkers: a collection of biomedical indicators that generally refer to biological states or conditions; in our case, specifically health conditions.
- Social Activities: a set of interactions between social entities, either persons or social communities, who exchange thoughts and ideas, communicate information, and share emotions and experiences.
- Physical Activities: any bodily activities involved in daily life. Some of the activities are conducted in order to enhance or maintain physical fitness and overall wellness/health.

The SMASH ontology has been submitted to the NCBO BioPortal.[4] Fig. 2 illustrates a partial view of the SMASH ontology and its hidden variables in the corresponding RBMs. (More details will be discussed in the next section.)

4. Ontology-based representation learning

In this section, we present our algorithm to learn user representations based on concepts and characteristics (properties) in ontologies. The representational primitives of ontologies are typically concepts, characteristics (datatype properties), and relationships (object properties). For instance, in Fig. 2, the main concept is *Person*. With this concept, we have sub-concepts,

---

[1] http://www.bioontology.org/.
[2] http://www.w3.org/TR/owl-ref/.
[3] http://protege.stanford.edu/.
[4] http://bioportal.bioontology.org/ontologies/SMASH.

relationships, and characteristics. Each person *has* a related concept *Biological Measure* which contains a set of characteristics such as BMI, weight, slope, wellness, etc. They are related to another concept, *Social Activity*, which contains sub-concepts, such as *Offline Activity* and *Online Activity*.

Given a health social network ontology H, the first step to utilize the formal semantics in deep learning is to learn the representation of all concepts, sub-concepts, and relationships given the characteristics. Our key hypothesis is that a concept or a sub-concept $\mathbb{S} \in H$ can be represented by its own characteristics, its sub-concepts, and its related concepts. In essence, $\mathbb{S} \in H$ is represented by a set of learnable hidden variables $\mathbf{h}_{\mathbb{S}}$. The learning process of the $\mathbf{h}_{\mathbb{S}}$ is as follows. $\mathbb{S}$ is composed of a set of characteristics $V_{\mathbb{S}}$, a set of sub-concepts $C_{\mathbb{S}}$, and a set of relationships $F_{\mathbb{S}}$. Let us denote $\Psi_{\mathbb{S}} = \bigcup_{F \in F_{\mathbb{S}}} V_F$ as the union of all characteristics from its relationships $F_{\mathbb{S}}$, and $\Theta_{\mathbb{S}} = \bigcup_{C \in C_{\mathbb{S}}} \mathbf{h}_C$ as the union of all the hidden variables from its sub-concepts $C_{\mathbb{S}}$. The hidden variables $\mathbf{h}_{\mathbb{S}}$ can be learned from $V_{\mathbb{S}}$, $\Psi_{\mathbb{S}}$, and $\Theta_{\mathbb{S}}$ by utilizing the RBMs as a deep learning model, since it very well fits our goals, which aim to generate a deeper analysis for human behavior prediction. By using the RBMs, the $\mathbf{h}_{\mathbb{S}}$ is considered as a hidden layer, and all the variables $v_i \in V_{\mathbb{S}} \cup \Psi_{\mathbb{S}} \cup \Theta_{\mathbb{S}}$ are considered as a visible layer in a RBM (Fig. 1). The conditional probabilities of an $h_j \in \mathbf{h}_{\mathbb{S}}$ and $v_i \in V_{\mathbb{S}} \cup \Psi_{\mathbb{S}} \cup \Theta_{\mathbb{S}}$ are given by:

$$p(h_j | V_{\mathbb{S}}, C_{\mathbb{S}}, F_{\mathbb{S}}) = \mathcal{N}\left(b_j + \sum_{v_i \in V_{\mathbb{S}} \cup \Psi_{\mathbb{S}} \cup \Theta_{\mathbb{S}}} v_i W_{ij} \right) \tag{6}$$

$$p(v_i | h_{\mathbb{S}}) = \mathcal{N}\left(a_i + \sum_{h_j \in h_{\mathbb{S}}} h_j W_{ij} \right) \tag{7}$$

where a_i and b_j are static biases, and W_{ij} is a parameter associated with h_j and v_i. By denoting $\mathbf{v}_{\mathbb{S}} = V_{\mathbb{S}} \cup \Psi_{\mathbb{S}} \cup \Theta_{\mathbb{S}}$, the energy function of the RBM for $\mathbb{S}$ is:

$$E(\mathbf{v}_{\mathbb{S}}, \mathbf{h}_{\mathbb{S}} | \theta) = \sum_{v_i}^{\mathbf{v}_{\mathbb{S}}} \frac{(v_i - a_i)^2}{2\sigma_i^2} + \sum_{h_j}^{\mathbf{h}_{\mathbb{S}}} \frac{(h_j - b_j)^2}{2\sigma_j^2} - \sum_{v_i, h_j}^{\mathbf{v}_{\mathbb{S}}, \mathbf{h}_{\mathbb{S}}} \frac{v_i}{\sigma_i} \frac{h_j}{\sigma_j} W_{ij}$$

By using contrastive divergence [10], we can train this RBM and learn all the parameters which are used to estimate the hidden variables $\mathbf{h}_{\mathbb{S}}$. In fact, $\mathbf{h}_{\mathbb{S}}$ can be considered as the representation of $\mathbb{S}$. Note that we use normal distributions for the hidden variables in $\mathbf{h}_{\mathbb{S}}$, because they will be used to learn the representation of the parent concepts of $\mathbb{S}$, denoted $P_{\mathbb{S}}$. $P_{\mathbb{S}}$ may contain real-valued characteristics (datatype properties). So the consistency in the learning process is guaranteed. The representations of all the concepts and sub-concepts can be learned by applying the bottom-up greedy layer-wise algorithm [11] following the structure of the ontology H. For instance, in Fig. 2, we can learn all the representations in the following order: $\mathbf{h}_L$, $\mathbf{h}_A$, $\mathbf{h}_{Off}$, $\mathbf{h}_{On}$, first; then $\mathbf{h}_B$, $\mathbf{h}_P$, $\mathbf{h}_S$, and $\mathbf{h}_R$, finally.

Let us denote the root concept (e.g., *Person*) and its representation $\mathbf{h}_R$, which also is an individual representation. Different applications may have different settings. The challenge becomes how we organize the training data so that individual representation can be learned. In fact, the data of each user u will be collected in a set of time intervals T, denoted by $D_u = \{K_1^u, \ldots, K_T^u\}$ where K is the set of all personal characteristics at all the concepts and sub-concepts. D_u will be used to train the model. After training the model, for every $t \in T$ we can navigate the K_t^u following the ontology structure to estimate the representation of root concept $\mathbf{h}_R$ which is also the representation of user u at time t. We train the model for each user independently. Each user will have different representations at different time intervals. In the next section, we show how to use those ontology-based user representations (i.e., RBMs) for human behavior prediction.

Our algorithm can learn the representations of individuals given different ontologies in different domain applications. In fact, our Ontology-based Representation Learning algorithm can learn the representation of any concept in a given ontology. Therefore, as long as we have an ontology, which represents or includes a "Person" concept, we can learn the representations of individuals. By offering this advanced ability, our model also can represent the hierarchical concept structures among biomedical and health domains.

5. ORBM$^+$: human behavior prediction with explanations

In this section, we present how to conduct human behavior prediction with explanations based on user representations, which have been learned from the SMASH ontology. Given an *online social network* $G = \{U, E\}$ where U is a set of all users and E is a set of edges. Every edge $e_{u,m}$ exists in E if u and m friend each other in G; otherwise $e_{u,m}$ does not exist. Each user has a set of individual representation features $\mathcal{F} = \{f_1, \ldots, f_n\}$. In essence, $\mathcal{F}$ is the $\mathbf{h}_R$ which has been learned for each user in the previous Section. The social network G grows from scratch over a set of time points $T = \{t_1, \ldots, t_m\}$. To illustrate this, we use $E_T = \{E_{t_1}, \ldots, E_{t_m}\}$ to denote the topology of the network G over time, where E_t is a set of edges which have been made *until* time t in the network, and $\forall t \in T : E_t \subseteq E_{t+1}$. For each user, the values of individual features in $\mathcal{F}$ also change over time. We denote the values of individual features of a user u at time t as $\mathcal{F}_u^t$. At each time point t, each user u is associated with a binomial behavior $y_u^t \in \{0, 1\}$. y_u^t could be "decrease/increase exercise," or "inactive/active in exercise." y_u^t will be clearly described in our experimental results section.
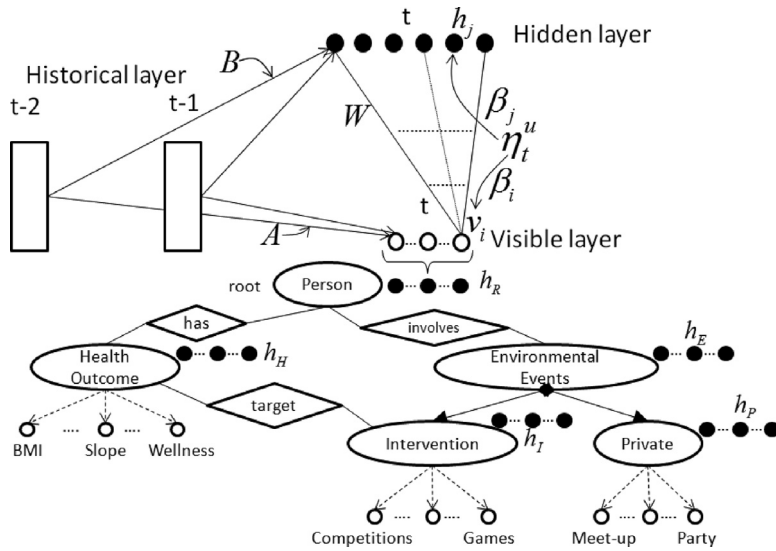
Fig. 3. The ORBM$^+$ model.

Problem formulation: Given the health social network in M timestamps $T_{data} = \{t - M + 1, \ldots, t\}$, we would like to predict the behavior of all the users in the next timestamp $t + 1$. More formally, given $\{\mathcal{F}_u^t, y_u^t, E_t | t \in T_{data}, u \in U\}$ we aim at predicting $\{y_u^{t+1} | u \in U\}$.

5.1. Self-motivation and environmental events

In order to model self-motivation of a user u, we first fully bipartite connect the hidden and visible layers via a *weight matrix W* (Fig. 3). Then each visible variable v_i and hidden variable h_j will be connected to all the historical variables of u, denoted by $\mathcal{H}_{fu,t-k}$ where $f \in \mathcal{F}$ and $k \in \{1, \ldots, N\}$. These connections are presented by the two weight matrices A and B (Fig. 3). Each historical variable $\mathcal{H}_{fu,t-k}$ is the state of feature f of the user u at time point $t - k$. Note that all the historical variables are treated as additional *observed* inputs. The hidden layer can learn the correlations among the features and the effect of historical variables to capture self-motivation. This effect can be integrated into a dynamic bias:

$$\hat{b}_{j,t} = b_j + \sum_{k \in \{1, \ldots, N\}} \sum_{f \in \mathcal{F}} B_{jfu,t-k} \mathcal{H}_{fu,t-k} \tag{8}$$

which includes a static bias b_j, and the contribution from the past of the social network. B_j is a $|\mathbf{h}| \times |U| \times N$ weight matrix which summarizes the autoregressive parameters to the hidden variable h_j. This modifies the factorial distribution over hidden variables: b_j in Eq. 3 is replaced with $\hat{b}_{j,t}$ to obtain

$$p(h_{j,t} = 1 | \mathbf{v}_t, \mathcal{H}_{t_<}) = \sigma \left(\hat{b}_{j,t} + \sum_i v_{i,t} W_{ij} \right) \tag{9}$$

where $h_{j,t}$ is the state of hidden variable j at time t, the weight W_{ij} connects v_i and h_j. The self-motivation has a similar effect on the visible units. The reconstruction distribution in Eq. 4 becomes

$$p(v_{i,t} | \mathbf{h}_t, \mathcal{H}_{t_<}) = \mathcal{N} \left(\hat{a}_{i,t} + \sum_j h_{j,t} W_{ij}, 1 \right) \tag{10}$$

where $\hat{a}_{i,t}$ is also a dynamic bias:

$$\hat{a}_{i,t} = a_i + \sum_{k \in \{1, \ldots, N\}} \sum_{f \in \mathcal{F}} A_{ifu,t-k} \mathcal{H}_{fu,t-k} \tag{11}$$

The effect of environmental events is composed of unobserved social relationships, unacquainted users, and the changing of social context [4,22]. In other words, a user can be influenced by any users via any features in health social networks. It is hard to precisely define the influences of environmental events. It is difficult to define straightforward statistical distribution functions, e.g., exponential function, Gaussian function, etc., to precisely present the influences of meet-up events, social games, competitions, etc., due to unobserved/hidden factors, such as undeclared friend connections and self-motivation. In fact, of all the actual friend relationships among participants, just a portion is observed in the YesiWell dataset, since users
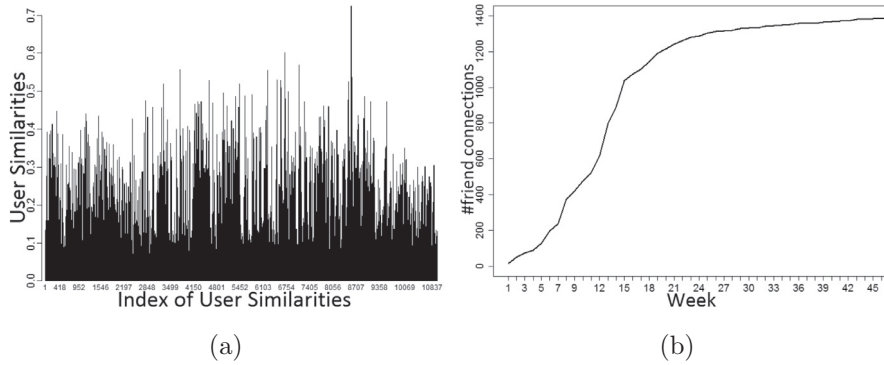
Fig. 4. A sample of cosine similarities (a) and cumulative number of friend connections (b) in our dataset.

participated in many off-line events. Inspired by the SRBM model [22], we model the flexibility of implicit social influences, as well as self-motivation, as follows. The individual features of a user, denoted as $\mathcal{F}_u$, can be considered as the visible variables in the ORBM$^+$ model (Fig. 3). Given a user u, each visible variable v_i and hidden variable h_j are connected to all historical variables of all other users. It is similar to the self-motivation modeling: The influence effects of each user, and the social context on the user u, are captured via the weight matrices A and B. These effects can be integrated into the dynamic biases $\hat{a}_{i,t}$ and $\hat{b}_{j,t}$ in Eqs. 11 and 8, as well. The dynamic biases in Eqs. 8 and 11 become:

$$\hat{b}_{j,t} = b_j + \sum_{k \in \{1,...,N\}} \sum_{f \in \mathcal{F}} \sum_{u \in U} B_{jfu,t-k} \mathcal{H}_{fu,t-k} \qquad (12)$$

$$\hat{a}_{i,t} = a_i + \sum_{k \in \{1,...,N\}} \sum_{f \in \mathcal{F}} \sum_{u \in U} A_{ifu,t-k} \mathcal{H}_{fu,t-k} \qquad (13)$$

The quantitative environmental events, such as the number of competitions and meet-up events, are included as individual characteristics. Therefore, the effect of environmental events is better embedded into the model. Next, we will incorporate the social influence into our ORBM$^+$ model.

5.2. Explicit social influences

It is well-known that individuals tend to be friends with people who perform behaviors similar to theirs (*homophily principle*). In addition, as shown in [24], users differentially experience and absorb physical exercise-based influences from their friends. Therefore, the explicit social influences in health social networks can be defined as a function of the homophily effect and physical exercise-based social influences. Let us first define user similarity as follows.

Given two neighboring users u and m, a simple way to quantify their similarity is to applying a *cosine function* of their individual representations (i.e., $\mathbf{v}^u$ and $\mathbf{v}^m$) and hidden features (i.e., $\mathbf{h}^u$ and $\mathbf{h}^v$). The user similarity between u and m at time t, denoted $s_t(u, m)$, is defined as:

$$s_t(u, m) = cos_t(u, m | \mathbf{v}) \times cos_t(u, m | \mathbf{h}) \qquad (14)$$

where $cos_t(\cdot)$ is a cosine similarity function, i.e.,

$$cos_t(u, m | \mathbf{v}) = \frac{p(\mathbf{v}_t^u | \mathbf{h}_t^u, \mathcal{H}_{t_<}^u) \cdot p(\mathbf{v}_t^m | \mathbf{h}_t^m, \mathcal{H}_{t_<}^m)}{\| p(\mathbf{v}_t^u | \mathbf{h}_t^u, \mathcal{H}_{t_<}^u) \| \, \| p(\mathbf{v}_t^m | \mathbf{h}_t^m, \mathcal{H}_{t_<}^m) \|}$$

$$cos_t(u, m | \mathbf{h}) = \frac{p(\mathbf{h}_t^u | \mathbf{v}_t^u, \mathcal{H}_{t_<}^u) \cdot p(\mathbf{h}_t^m | \mathbf{v}_t^m, \mathcal{H}_{t_<}^m)}{\| p(\mathbf{h}_t^u | \mathbf{v}_t^u, \mathcal{H}_{t_<}^u) \| \, \| p(\mathbf{h}_t^m | \mathbf{v}_t^m, \mathcal{H}_{t_<}^m) \|}$$

Fig. 4 a illustrates a sample of a user similarity spectrum of all the edges in our social network over time. We randomly select 35 similarities of neighboring users for each day in ten months. Apparently, the distributions are not uniform, and different time intervals present various distributions. To well qualify the similarity between individuals and their friends, it potentially requires a *cumulative distribution function* (CDF). In addition, our health social network is developed from scratch. As time goes by, each participant will have more connections to other users (Fig. 4b). Thus a temporal smoothing is needed to better capture the explicit social influences. Eventually, we propose a *statistical explicit social influence*, denoted η_t^u, of a user u at time t as follows:

$$\eta_t^u = \alpha \eta_{t-\tau}^u + (1 - \alpha) \frac{1}{|\mathbb{F}_t^u|} \sum_{m \in \mathbb{F}_t^u} \psi_t(m, u) \times p\big(s_t \le s_t(u, m)\big)$$

Table 2
Learning rules.

Algorithm	Learning rules
Contrastive divergence	$\Delta W_{ij} = \sum_t \left(\langle v_{i,t} h_{j,t} \rangle_d - \langle v_{i,t} h_{j,t} \rangle_r \right)$
	$\Delta a_i = \sum_t \left(\langle v_{i,t} \rangle_d - \langle v_{i,t} \rangle_r \right); \ \Delta b_j = \sum_t \left(\langle h_{j,t} \rangle_d - \langle h_{j,t} \rangle_r \right)$
	$\Delta A_{ifu,t-k} = \sum_t \left(\langle v_{i,t} \mathcal{H}_{fu,t-k} \rangle_d - \langle v_{i,t} \mathcal{H}_{fu,t-k} \rangle_r \right)$
	$\Delta B_{jfu,t-k} = \sum_t \left(\langle h_{j,t} \mathcal{H}_{fu,t-k} \rangle_d - \langle h_{f,t} \mathcal{H}_{fu,t-k} \rangle_r \right)$
	$\Delta \beta_i = \sum_t \left(\langle v_{i,t} \rangle_d - \langle v_{i,t} \rangle_r \right) \eta_t^u; \ \Delta \beta_j = \sum_t \left(\langle h_{j,t} \rangle_d - \langle h_{j,t} \rangle_r \right) \eta_t^u$
	$\frac{\partial C(\theta)}{\partial s_j} = -\sum_t (y_t - \hat{y}_t) h_j; \ \frac{\partial C(\theta)}{\partial c} = -\sum_t (y_t - \hat{y}_t)$
Back-propagation	$\frac{\partial C(\theta)}{\partial W_{ij}} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j) v_i$
	$\frac{\partial C(\theta)}{\partial a_i} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j) W_{ij}$
	$\frac{\partial C(\theta)}{\partial b_j} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j)$
	$\frac{\partial C(\theta)}{\partial A_{ifu,t-k}} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j) W_{ij} \mathcal{H}_{fu,t-k}$
	$\frac{\partial C(\theta)}{\partial B_{jfu,t-k}} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j) \mathcal{H}_{fu,t-k}$
	$\frac{\partial C(\theta)}{\partial \beta_i} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j) W_{ij} \eta_t^u$
	$\frac{\partial C(\theta)}{\partial b_j} = -\sum_t (y_t - \hat{y}_t) s_j h_j (1 - h_j) \eta_t^u$

where $\mathbb{F}_t^u$ is a set of friends of user u, from the beginning *until* time t. $\psi_t(m, u)$ is the physical exercise-based social influence of m on u at time t, which is derived by using the CPP model [24]. s_t is the similarity between two arbitrary neighboring users in the social network at time t. $p(s_t \leq s_t(u, m))$ represents the probability that the similarity is less than or equal to the instant similarity $s_t(u, m)$. α and τ are two parameters to control the dynamics of η.

5.3. Inference, learning, and prediction

Inference in the ORBM$^+$ is no more difficult than in the RBM. The states of the hidden variables are determined both by the inputs they receive from the visible variables and from the historical variables. The conditional probability of hidden variables at time interval t can be computed, as in Eqs. 8 and 9. The combination of the implicit and explicit social influences can be viewed as a linear adaptive bias: $\hat{a}_{i,t}$ in Eq. 11 becomes

$$\hat{a}_{i,t} = a_i + \sum_{k \in \{1,...,N\}} \sum_{f \in \mathcal{F}} \sum_{u \in U} A_{ifu,t-k} \mathcal{H}_{fu,t-k} + \beta_i \eta_t^u$$

$$\hat{b}_{j,t} = b_j + \sum_{k \in \{1,...,N\}} \sum_{f \in \mathcal{F}} \sum_{u \in U} B_{jfu,t-k} \mathcal{H}_{fu,t-k} + \beta_j \eta_t^u$$

where β_i is a parameter which presents the ability to observe the explicit social influences η_t^u of user u given v_i.

The energy function becomes:

$$E(\mathbf{v}_t, \mathbf{h}_t | \mathcal{H}_{t_<}, \theta) = \sum_{i \in \mathbf{v}} \frac{(v_{i,t} - \hat{a}_{i,t})^2}{2\xi_i^2} - \sum_{j \in \mathbf{h}} \hat{b}_{j,t} h_{j,t} - \sum_{i \in \mathbf{v}, j \in \mathbf{h}} \frac{v_{i,t}}{\xi_i} h_{j,t} W_{ij} + \lambda |\theta = \{A, B, W, \beta\}|_1 \tag{15}$$

Contrastive divergence is used to train the ORBM$^+$ model. The updates for the symmetric weights, W, the static biases, a and b, the directed weights, A and B, are based on simple pairwise products. The gradients are summed over all the training time intervals $t \in T_{train} = T_{data} \setminus \{t - M + 1, \ldots, t - M + N\}$. The learning rules are summarized in Table 2.

On top of our model, we put an output layer for the user behavior prediction task. Our goal is to predict whether a user *will increases* or *will decreases* physical exercise levels. Thus the softmax layer contains a single output variable $\hat{y}$ and binary target values: 1 for increases, and 0 for decreases. The output variable $\hat{y}$ is fully linked to the hidden variables by weighted connections S, which includes $|\mathbf{h}|$ parameters s_j. The logistic function is used as an activation function of $\hat{y}$ as follows:

$$\hat{y} = \sigma \left(c + \sum_{j \in \mathbf{h}} h_j s_j \right)$$

where c is a static bias. Given a user $u \in U$, a set of training vectors $X = \{\mathcal{F}_u^t, E_t | t \in T_{train}\}$, and an output vector $Y = \{y_t | t \in T_{train}\}$, the probability of a binary output $y_t \in \{0, 1\}$ given the input x_t is as follows:

$$P(Y|X, \theta) = \prod_{t \in T_{train}} \hat{y}_t^{y_t} (1 - \hat{y}_t)^{1-y_t} \tag{16}$$

where $\hat{y}_t = P(y_t = 1|x_t, \theta)$.

A loss function to appropriately deal with the binomial problem is *cross-entropy error*. It is given by

$$C(\theta) = - \sum_{t \in T_{train}} \left(y_t \log \hat{y}_t + (1 - y_t) \log(1 - \hat{y}_t) \right) \tag{17}$$

In the final stage of training, *Back-propagation* is used to fine-tune all the parameters together. The derivatives of the objective $C(\theta)$ with respect to all the parameters over all the training cases $t \in T_{train}$ are summarized in Table 2. In the prediction task, we need to predict the y_{u+1}^u without observing the $\mathcal{F}_u^{t+1}$. In other words, the visible and hidden variables are not observed at the future time point $t + 1$. Thus we need a causal generation step to initiate these variables. Causal generation from a learned ORBM$^+$ model can be done just like the learning procedure. In fact, we always keep the historical variables fixed and perform alternating Gibbs sampling to obtain a joint sample of the visible and hidden variables from the ORBM$^+$ model. To start alternating Gibbs sampling, a good choice is to set $\mathbf{v}_t = \mathbf{v}_{t-1}$, (i.e., $\mathbf{v}_{t-1}$ is a strong prior of $\mathbf{v}_t$). This picks new hidden and visible variables that are compatible with each other and with the recent historical variables. Afterward, we aggregate the hidden variables to evaluate the output $\hat{y}$.

5.4. Explanation generation

The success of human behavior intervention does not only depend on its accuracy in inferring and exploring users' behaviors, but it also relies on how the deployed interventions are perceived by the users. Explanations increase the transparency of the intervention process and contribute to users' satisfaction, and are facilitative in engaging users to the program.

The essence of explanations is to know the key characteristics or components that cause users to do some specific behaviors. In other words, an explanation can be defined as a list of characteristics which maximize the likelihood of a behavior being engaged by a user or a set of users. Recall that the log likelihood $\log P(Y|X, \theta)$ that the users will do Y is maximized given all the characteristics X. Thus, X should be the best explanation for Y. However, X usually contains many characteristics in high dimensional data, that significantly reduces the interpret-ability of X. We need a solution to address the trade-off between the likelihood maximization and the complexity of X. Given a selection model f, the problem can be written as:

$$X^* = \arg \min_{X' \subseteq X} f\left(\log P(Y|X', \theta), X' \right) \tag{18}$$

To address this issue, we utilize the *Minimum Description Length* (MDL) [26]. In the two-parts of MDL, we first use the explanation X' to encode the observed data Y, and we then encode the explanation itself. We denote the encoding length of Y given X' by $L(Y, X')$, and the encoding length of X' by $L(X')$. As in [26], $L(Y, X')$ is the negative log-likelihood $\log P(Y|X', \theta)$ and $L(X')$ is given by $|X'|\log(|X|)$. MDL minimizes the combined encoding:

$$MDL = L(Y, X') + L(X') = - \log P(Y|X', \theta) + |X'| \log(|X|) \tag{19}$$

The complexity of explanation generating is NP-Hard. It is given by the following lemma.

Lemma 1. *Finding the optimal explanation problem and the minimal value of the selection function is NP-Hard.*

Proof. Given a set of elements $\{1, 2, \ldots, m\}$ (called the universe) and n sets whose union comprises the universe, the set cover problem [20] is to identify the smallest number of sets whose union still contains all elements in the universe.

We can easily reduce the finding optimal explanation problem to the set cover problem by setting that $x \in X$ is an item and X is the universe. x and X will be mapped into a likelihood space Y as: $x :\rightarrow Y_x = \left(P(y_1|x, \theta), \ldots, P(y_k|x, \theta) \right)$, and $X :\rightarrow Y_X = \left(P(y_1|X, \theta), \ldots, P(y_k|X, \theta) \right)$. Y_X can be considered as a new universe, and Y_x is an item in that universe. The problem now becomes finding the smallest number of items Y_x whose union still covers the universe Y_X. That means we must have $\forall i : P(y_i|X', \theta) = P(y_i|X, \theta)$ where $Y_{X'}$ is the smallest union of items. In doing so, finding the optimal explanation is as hard as solving the corresponding instance of the set cover problem. This proves the NP-Hardness property of the given problem. □

Since the finding optimal explanation problem is NP-Hard, we apply a heuristic greedy algorithm, which will add a new characteristic into the existing explanation so that the selection model MDL is minimized. Afterward, we are able to select the optimal explanation with a minimal MDL score.

6. Experimental results

We have carried out a series of experiments in a real health social network to validate our proposed ORBM$^+$ model (source codes and data[5]). We first elaborate about the experiment configurations, evaluation metrics, and baseline approaches. Then, we introduce the experimental results.

Experiment configurations. In our study, we take into account 33 personal attributes (Table 1). The personal social communications include 2766 inbox messages, which are categorized into 20 different types. Fig. 5 illustrates the distributions
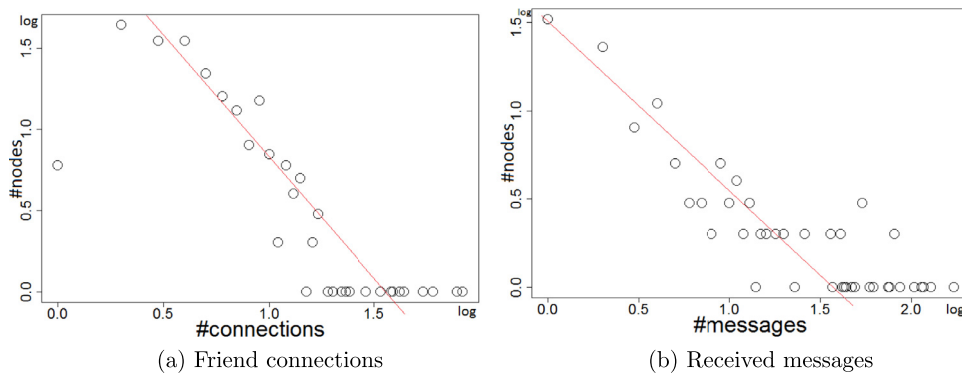

(a) Friend connections   (b) Received messages

Fig. 5. Some distributions in our dataset.

of friend connections, and #received messages in our data. They clearly follow the Power Law distribution. Note that, given a week, if a user exercises more than she did in the previous week, she is considered to be increasing exercise; otherwise, she will be considered to be decreasing exercise. The number of hidden units, and the number of previous time intervals N, respectively are set to 200 and 3. In the individual representation learning, the number of hidden units at all the concepts and sub-concepts in the ontology will double the number of visible units. The weights are randomly initialized from a zero-mean Gaussian with a standard deviation of 0.01. All the learning rates are set to 10^{-3}. A contrastive divergence CD_{20} is used to maximize likelihood learning. We train the model for each user independently.

Evaluation metrics. In the experiment, we leverage the previous 10 weeks' records to predict the behaviors of all the users (i.e., increase or decrease exercises) in the next week. The prediction quality metric, i.e., *accuracy*, is as follows:

$$accuracy = \frac{\sum_{i=1..|U|} I(y_i = \hat{y}_i)}{|U|}$$

where y_i is the true user activity of the user u_i, and $\hat{y}_i$ denotes the predicted value, and I is the indication function.

Competitive prediction models. We compare the ORBM$^+$ model with the conventional methods reported in [27]. The competitive methods are divided into two categories: personalized behavior prediction methods and socialized behavior prediction methods. Personalized methods only leverage individuals' past behavior records for future behavior predictions. Socialized methods use both one person's past behavior records and his or her friends' past behaviors for predictions. Specifically, the five models reported in [27] are the Socialized Gaussian Process (**SGP**) model, the Socialized Logistical Autoregression (**SLAR**) model, the Personalized Gaussian Process (**PGP**) model, the Logistical Autoregression (**LAR**) model, and the Behavior Pattern Search (**BPS**) model.

We also consider the RBM related extensions, i.e., the **CRBM** [30] and **ctRBM** [17], as competitive models. The CRBM can be directly applied to our problem by ignoring the implicit and explicit influences in our SRBM$^+$ model. Since the ctRBM cannot directly incorporate individual attributes with social influences to model human behaviors, we only can apply its social influence function into our model. In fact, we replace our statistical explicit social influence function with the ctRBM's social influence function. We call this version of ctRBM a *Socialized ctRBM* (**SctRBM**). We also compare the ORBM$^+$ model with our previous works, SRBM in [22] and ORBM in [23].

6.1. Validation of the ORBM$^+$ model for prediction

Our task of validation focuses on two key issues: (1) Which configurations of the parameters α and τ produce the best-fit social influence distribution? and (2) Is the ORBM$^+$ model better than the competitive models, in terms of prediction accuracy? We carry out the validation through two approaches. One is to conduct human behavior prediction with various settings of α and τ. By this we look for an optimal configuration for the statistical explicit social influence function. The second validation is to compare our ORBM$^+$ model with the competitive models in terms of prediction accuracy.

Fig. 6a illustrates the surface of the prediction accuracy of the ORBM$^+$ model, with variations of the two parameters α and τ on our health social network. We observed that the smaller values of τ tend to have higher prediction accuracies. This is quite reasonable, since the more recent behaviors have stronger influences. The temporal smoothing parameter τ has similar effects to a time decay function [37]. Meanwhile, the middle range values of α offer better prediction accuracies. Clearly, the optimal setting values of α and τ are 0.5 and 1 respectively.

To examine the efficiency, we compare the proposed ORBM$^+$ model with the competitive models, in terms of human behavior prediction. Fig. 6c shows the accuracy comparison over 37 weeks in our health social network. It is clear that the ORBM$^+$ outperforms the other models. The accuracies of the competitive models tend to drop in the middle period of the study. All the behavior determinants and their interactions potentially become stronger, since all the users improve their activities, such as walking and running, participating in more competitions, etc. (Fig. 7) in the middle weeks. Absent or
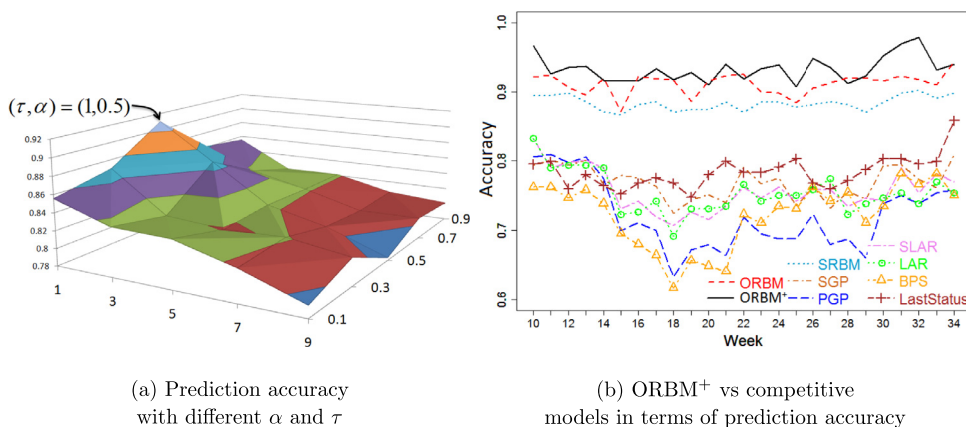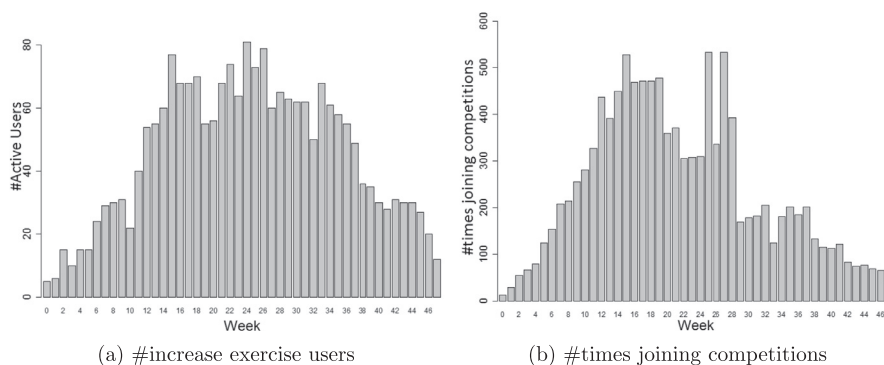
(a) Prediction accuracy
with different α and τ

(b) ORBM$^+$ vs competitive
models in terms of prediction accuracy

Fig. 6. Validation of the ORBM$^+$ model.

(a) #increase exercise users

(b) #times joining competitions

Fig. 7. The distributions of users' activities.

insufficient modeling of one of the determinants, or of one of their interactions, results in a low and unstable prediction performance. Therefore, competitive models do not well capture the social influences and environmental events. Meanwhile, the ORBM$^+$ model comprehensively models all the determinants. So, the correlation between the personal attributes and the implicit social influences can be detected by the hidden variables. Thus, much information has been leveraged to predict individual behaviors. In addition, our prediction accuracy stably increases over time. That means our model well captures the growing of our health social network (Fig. 4b). Consequently our model achieves *higher prediction accuracy* and a more *stable* performance. The ORBM$^+$ model outperforms our previous ORBM model because of the new temporal-smoothing social influence function, which better captures the explicit social influences in our health social network. Overall, the ORBM$^+$ model achieves the best prediction accuracy in average as **0.9332**.

6.2. The effectiveness of user representation learning

In Fig. 6b, the ORBM$^+$ outperforms the SRBM, which does not learn user representation from the SMASH ontology, by 4.94% (0.8941 vs. 0.9435). This is because ontology-based user representation provides better features, which encode semantics of the data structure from domain knowledge to the ORBM$^+$ model. Even when we stack another hidden layer on the SRBM, the accuracy and the cross-entropy error are not (or are not much) improved, compared with the effect of ontology-based user representation. Consequently, we can conclude that: (1) The SMASH ontology helps us to organize data features in a suitable way; (2) Our algorithm can learn meaningful user representations from ontologies; and (3) Meaningful user representations can further improve accuracies of deep learning approaches for human behavior prediction.

6.3. Synthetic health social network

To illustrate that the ORBM$^+$ model can be generally applied on different datasets, we perform further experiments on a synthetic health social network. To generate the synthetic data, we use the software Pajek[6] to generate graphs under the
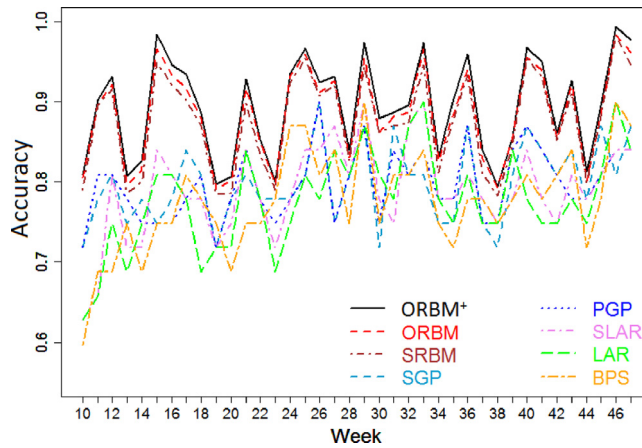
---

[6] http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Fig. 8. Accuracies on the synthetic data.

Table 3
Dynamic biases of human behavior determinants.

Determinants (output variable)	Dynamic biases
Self-motivation ($\hat{y}_{t+1,self}$)	$\hat{b}_{j,t} = b_j + \sum_{k=1}^{N} \sum_{f \in \mathcal{F}} B_{jfu,t-k} \mathcal{H}_{fu,t-k}$ $\hat{a}_{i,t} = a_i + \sum_{k=1}^{N} \sum_{f \in \mathcal{F}} A_{ifu,t-k} \mathcal{H}_{fu,t-k}$
Implicit social influence ($\hat{y}_{t+1,im}$)	$\hat{b}_{j,t} = b_j + \sum_{k=1}^{N} \sum_{f \in \mathcal{F}} \sum_{\substack{m \in U \\ m \neq u}} B_{jfm,t-k} \mathcal{H}_{fm,t-k}$ $\hat{a}_{i,t} = a_i + \sum_{k=1}^{N} \sum_{f \in \mathcal{F}} \sum_{\substack{m \in U \\ m \neq u}} A_{ifm,t-k} \mathcal{H}_{fm,t-k}$
Explicit social influence ($\hat{y}_{t+1,ex}$)	$\hat{b}_{j,t} = b_j + \beta_j \eta_t^u$ $\hat{a}_{i,t} = a_i + \beta_i \eta_t^u$
Environmental events ($\hat{y}_{t+1,env}$)	only use #competitions,#meet-up events, and #social games in the ORBM$^+$ model

Scale-Free/Power Law Model, which is a network model whose node degrees follow the Power Law distribution, or at least do asymptotically. However the vertices in the current synthetic graph do not have individual features similar to the real-word YesiWell data. An appropriate solution to this problem is to apply a graph-matching algorithm to map pairwise vertices between the synthetic and real social networks. In order to do so, we first generate a graph with 254 nodes, in which the average node degree is 5.4 (i.e., similar to the real YesiWell data). Then, we apply PATH [36], which is a very well-known and efficient graph-matching algorithm, to find a correspondence between vertices of the synthetic network and vertices of the YesiWell network. The source code of the PATH algorithm is available in the graph-matching package, GraphM.[7] Then, we can assign all the individual features and behaviors of real users to corresponding vertices in the synthetic network. Consequently, we have a synthetic health social network that imitates our real-world dataset. Fig. 8 shows the accuracies of the conventional models, SRBM model, the ORBM model, and the ORBM$^+$ model on the synthetic data. We can see that the ORBM$^+$ model still outperforms the other models in terms of the prediction accuracy.

6.4. *Validation of behavior determinants and explanations*

6.4.1. *Reliability of behavior determinants*

One of our main goals is to validate the reliability of the human behavior determinants that are learned in our model. We illustrate this in a comparative experiment, which is to say: each determinant is independently used in the ORBM$^+$ model to predict behaviors of the users. The ORBM$^+$ model provides a natural way to address this. The effects of human behavior determinants are reflected via the ways we formulate the dynamic biases. To evaluate the effect of self-motivation on predicted behaviors, we compute the output variable $\hat{y}_{t+1}$ by using the dynamic biases $\hat{b}_{j,t}$ and $\hat{a}_{i,t}$, which are in the forms of Eqs. (8) and (11). This means we only use the self-motivation effect to predict the behaviors of the users. Similarly, we can evaluate the effects of *implicit social influences*, *explicit social influences*, and *environmental events*. The corresponding dynamic biases for each human behavior determinant are summarized in Table 3. We use $\hat{y}_{t+1,self}$, $\hat{y}_{t+1,im}$, $\hat{y}_{t+1,ex}$, and $\hat{y}_{t+1,env}$ to denote the output variable $\hat{y}_t$ given corresponding determinants. With regard to the effect of *environmental events*, we use the number of joined competitions, meet-up events, and social games to evaluate the output variable $\hat{y}_{t+1,env}$.
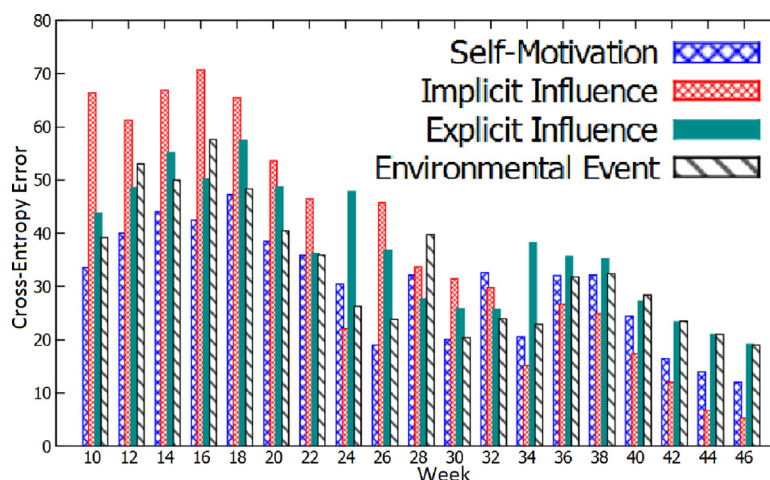
Fig. 9. The cross-entropy errors of the human behavior determinants.

Fig. 9 illustrates the cross-entropy errors in Eq. 16 of four determinants over 37 weeks of our data. We observed that self-motivation is more reliable than other determinants in terms of motivating users to increase their exercises in the first 10 weeks. This is because it achieves the best cross-entropy errors. Meanwhile, the implicit social influence is not effective in this period, since the users have not developed enough implicit relationships (i.e., highest cross-entropy errors). However, from the 24th week, the implicit social influence becomes one of the most reliable determinants, since it achieves the lowest cross-entropy errors. This phenomenon shows the important role of the implicit social influences in health social networks. In the meantime, the explicit social influences and environmental events behave as connecting factors, which not only influence the behaviors of users, but also associate the self-motivation and the implicit social influences together. In addition, the evolution of the determinants suggests a strong interaction among them, since there are no either absolutely reliable nor absolutely unreliable determinants.

In obvious, we discover three meaningful observations: (1) The ORBM$^+$ model enables the modeling of expressive correlations between determinants; (2) The self-motivation is especially important at the beginning; and (3) The implicit social influence will become one of the most reliable determinants, if the users have enough time to develop their relationships.

6.4.2. Human behavior explanations

We have shown that the ORBM$^+$ model not only achieves a significantly higher prediction accuracy, compared with the conventional models, but it also offers a deep understanding of the human behavior determinants. That is a breakthrough in human behavior modeling in health social networks. In this section, we focus on giving explanations for human behaviors. The concerned questions are: "which are the key characteristics leading to the increasing or decreasing exercises of the users in the next week?" and "do these key characteristics offer a higher prediction accuracy with more interpretable results compared with the original personal characteristics?"

For instance, given the 9th week, the users will be classified into two groups which are predicted as increase exercise users and predicted as decrease exercise users in the next week. We then extract the key characteristics for each group by applying our algorithm based on MDL (Section 5.4). Fig. 10a illustrates the scores of MDL the users who are predicted as increase exercises in the next week. We achieve the minimal value of the MDL given the list of characteristics[8] {1, 0, 4, 11, 19} which is equivalent to {#exercise days, #competitions, Wellness-score, #competition messages, #progress reports}. Meanwhile, for the users for whom the model predicted a decrease in exercise, we obtain the minimal value of the MDL given the list of characteristics {4, 3, 22, 20, 16} which is equivalent to {Wellness-score, BMI slope, #invitations, #heckling messages, #goal messages} (Fig. 10b).

To validate the interpret-ability of key characteristics extracted by our algorithms and models, we compare the decision trees extracted by interpretable classifiers from original personal characteristics and human behavior determinants in terms of classification and prediction accuracies. The interpretable classification models we used are: (1) *Decision trees* (DT) [3], which is a classical interpretable classifier; and (2) *Ism_td* [32], which is the state-of-the-art tree-based interpretable classification model. Table 4 illustrates the accuracies of the classification on predicted behaviors and the prediction, given actual behaviors of the users. The tree's depth is the maximal number of layers from the root of decision trees, which are generated by the DT and Ism_td algorithms. We can consider that, given the same prediction accuracy, the smaller the depth of the tree, the more interpretable the tree is. It is also true that, given the same depth, more accurate the tree in terms of
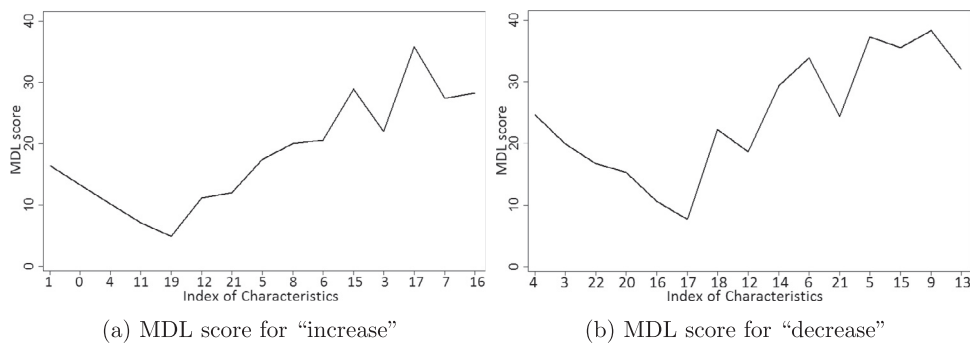
---

[7] http://cbio.ensmp.fr/graphm/.

[8] https://www.dropbox.com/s/nodra74cku06oeo/IndexCharac.txt?dl=0.

(a) MDL score for "increase"

(b) MDL score for "decrease"

Fig. 10. The scores of the MDL for predicted behaviors of the users at the 9th week.

Table 4
Classification and prediction accuracies of interpretable classifiers in the whole dataset.

Tree's depth		5		10		15		20	
Classifier		DT	Ism_td	DT	Ism_td	DT	Ism_td	DT	Ism_td
Personal	Classification	0.554	0.607	0.604	0.657	0.617	0.685	0.625	0.707
Characteristics	Prediction	0.494	0.558	0.569	0.57	0.587	0.605	0.593	0.631
Key	Classification	**0.684**	**0.881**	**0.756**	**0.907**	**0.777**	**0.925**	**0.778**	**0.936**
Characteristics	Prediction	**0.635**	**0.801**	**0.728**	**0.838**	**0.748**	**0.886**	**0.751**	**0.860**

Bold numbers indicate better accuracies.

classification and prediction, the more interpretable the tree is. In Table 4, we can see that by applying the DT and Ism_td algorithms on key characteristics, they achieve significantly better classification and prediction accuracies, given different values of tree's depth. In this point of view, the key characteristics extracted by our algorithm can be used to generate better interpretable decision-trees compared with the original personal characteristics. An example of generated decision trees from the Ism_td model is made available here.[9]

7. Conclusions

This paper introduces ORBM$^+$, a novel ontology-based deep learning model for human behavior prediction and explanation in health social networks. By incorporating all human behavior determinants - self-motivation, implicit and explicit social influences, and environmental events - our model predicts the future activity levels of users more accurately and stably than conventional methods. We contribute novel techniques to deal with structural domain knowledge (i.e., ontologies) and human behavior modeling. Our experiments in real-world and synthetic health social networks discover several meaningful insights: (1) User representations based on ontologies can further improve accuracies of deep learning approaches for human behavior prediction; (2) The ORBM$^+$ model expressively represents all the determinants and their correlations; and (3) Personal key characteristics which are extracted by our model are reliable, and they can be used to generate better explanations compared with original personal characteristics for human behaviors.

Acknowledgments

References

[1] A. Bandura, Human agency in social cognitive theory, Am. Psychologist 44 (9) (1989) 1175–1184.
[2] N. Barbieri, F. Bonchi, F. Manco, Who to follow and why: link prediction with explanations, in: KDD '14, 2014, pp. 1266–1275.
[3] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth International Group, 1984.
[4] N. Christakis, The hidden influence of social networks, TED2010, 2010. URL http://www.ted.com/talks/nicholas_christakis_the_hidden_influence_of_social_networks.
[5] N. Christakis, J. Fowler, The spread of obesity in a large social network over 32 years., N. Engl. J. Med. 357 (2007) 370–379.
[6] D. Dou, H. Wang, H. Liu, Semantic data mining: asurvey of ontology-based approaches, in: ICSC'15, 2015, pp. 244–251.
[7] R. Fichman, C. Kemerer, The illusory diffusion of innovation: an examination of assimilation gaps., Inf. Syst. Res. 10 (3) (1999) 255–275.

---

[9] https://www.dropbox.com/s/69sx8ijbwydmpt3/IsmtdTree.pdf?dl=0.

[8] T. Gruber, A translation approach to portable ontology specifications, Knowl. Acquisition 5 (2) (1993) 199–220.
[9] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing? Int. J. Hum.-Comput. Stud. 43 (5) (1995) 907–928.
[10] G. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800.
[11] G. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[12] J. Huang, X.-Q. Cheng, H.-W. Shen, T. Zhou, X. Jin, Exploring social influence via posterior effect of word-of-mouth recommendations, in: WSDM'12, 2012, pp. 573–582.
[13] J. Kawale, A. Pal, J. Srivastava, Churn prediction in mmorpgs: a social influence based approach, in: CSE'09, 2009, pp. 423–428.
[14] D. Kil, F. Shin, B. Piniewski, J. Hahn, K. Chan, Impacts of social health data on predicting weight loss and engagement, in: O'Reilly StrataRx Conference, San Francisco, CA, 2012.
[15] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[16] K. Lerman, S. Intagorn, J.-K. Kang, R. Ghosh, Using proximity to predict activity in social networks, in: WWW'12 Companion, 2012, pp. 555–556.
[17] X. Li, N. Du, H. Li, K. Li, J. Gao, A. Zhang, A deep learning approach to link prediction in dynamic networks, in: SDM'14, 2014, pp. 289–297.
[18] D. Lindberg, B. Humphries, A. McCray, The unified medical language system, Methods Inf. Med. 32 (4) (1993) 281–291.
[19] A. Marshall, E. Eakin, E. Leslie, N. Owen, Exploring the feasibility and acceptability of using internet technology to promote physical activity within a defined community., Health Promotion J. Aust. 2005 (16) (2005) 82–84.
[20] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, New York, USA, 1995.
[21] R. Pate, M. Pratt, S. Blair, Physical activity and public health. A recommendation from the centers for disease control and prevention and the american college of sports medicine., J. Am. Med. Assoc. 273 (5) (1995) 402–407.
[22] N. Phan, D. Dou, B. Piniewski, D. Kil, Social restricted Boltzmann machine: human behavior prediction in health social networks, in: ASONAM'15, 2015, pp. 424–431.
[23] N. Phan, D. Dou, H. Wang, D. Kil, B. Piniewski, Ontology-based deep learning for human behavior prediction in health social networks, in: ACM BCB'15, 2015, pp. 433–442.
[24] N. Phan, D. Dou, X. Xiao, B. Piniewski, D. Kil, Analysis of physical activity propagation in a health social network, in: CIKM'14, 2014, pp. 1329–1338.
[25] H. Poon, P. Domingos, Sum-product networks: a new deep architecture, in: IEEE ICCV Workshops, 2011, pp. 689–690.
[26] J. Rissanen, A universal prior for integers and estimation by minimum description length., Ann. Stat. 14 (5) (1983) 416–431.
[27] Y. Shen, R. Jin, D. Dou, N. Chowdhury, J. Sun, B. Piniewski, D. Kil, Socialized Gaussian process model for human behavior prediction in a health social network, in: ICDM'12, 2012, pp. 1110–1115.
[28] P. Smolensky, Information processing in dynamical systems: foundations of harmony theory, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press, 1986, pp. 194–281.
[29] R. Studer, V.R. Benjamins, D. Fensel, Knowledge engineering: principles and methods, Data knowl. Eng. 25 (1) (1998) 161–197.
[30] G. Taylor, G. Hinton, S. Roweis, Modeling human motion using binary latent variables, in: NIPS'06, 2006, pp. 1345–1352.
[31] The gene ontology consortium, Creating the gene ontology resource: design and implementation, Genome Res. 11 (8) (2001) 1425–1433, doi:10.1101/gr.180801.
[32] A. Van Assche, H. Blockeel, Seeing the forest through the trees: Learning a comprehensible model from an ensemble, in: ECML'07, vol. 4701, 2007, pp. 418–429.
[33] B. Viswanath, A. Mislove, M. Cha, K. Gummadi, On the evolution of user interaction in facebook, in: WOSN'09, 2009, pp. 37–42.
[34] M. Welling, M. Rosen-Zvi, G. Hinton, Exponential family harmoniums with an application to information retrieval, in: NIPS'04, 2004, pp. 1481–1488.
[35] J. Yang, X. Wei, M. Ackerman, L. Adamic, Activity lifespan: an analysis of user survival patterns in online knowledge sharing communities, ICWSM'10, 2010.
[36] M. Zaslavskiy, F. Bach, J.-P. Vert, A path following algorithm for graph matching, in: IEEE Trans. Pattern Anal. Mach. Intell., 5099, 2008, pp. 329–337.
[37] Y. Zhu, E. Zhong, S. Pan, X. Wang, M. Zhou, Q. Yang, Predicting user activity level in social networks, in: CIKM'13, 2013, pp. 159–168.