Ontology-based Deep Learning for Human Behavior Prediction in Health Social Networks

NhatHai Phan
University of Oregon, USA
haiphan@cs.uoregon.edu

Dejing Dou
University of Oregon, USA
dou@cs.uoregon.edu

Hao Wang
University of Oregon, USA
csehao@cs.uoregon.edu

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

Brigitte Piniewski
PeaceHealth Laboratories
BPiniewski@peacehealthlabs.org

ABSTRACT

Human behavior prediction is a key component to studying the spread of wellness and healthy behavior in a social network. In this paper, we introduce an *ontology-based Restricted Boltzmann Machine* (ORBM) model for human behavior prediction in health social networks. We first propose a bottom-up algorithm to learn the user representation from ontologies. Then the user representation is used to incorporate *self-motivation*, *social influences*, and *environmental events* together in a human behavior prediction model, which extends a well-known deep learning method, Restricted Boltzmann Machines (RBMs), so that the interactions among the behavior determinants are naturally simulated through parameters. To our best knowledge, the ORBM model is the first ontology-based deep learning approach in health informatics for human behavior prediction. Experiments conducted on both real and synthetic data from health social networks have shown the tremendous effectiveness of our approach compared with conventional methods.

Categories and Subject Descriptors

J.3 [**Computer Applications**]: LIFE AND MEDICAL SCIENCES—*Health*; H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Theory, Algorithms, Experimentation

Keywords

Ontology, deep learning, social network, health informatics

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 obesity can spread over the social network [6], bearing similarity to the diffusion of innovation [8] and word-of-mouth effects in marketing [13]. To reduce the risk of obesity-related diseases, regular exercise is strongly recommended (i.e., at least 30 minutes of moderate-intensity physical activity on 5 or more days a week [23]). 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.

The Internet is identified as an important source of health information and may thus be an appropriate delivery vector for health behavior interventions [21]. 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 [26], 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 friend and 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

BCB'15, September 09-12, 2015, Atlanta, GA, USA
 ISBN 978-1-4503-3853-0/15/09$15.00
DOI: http://dx.doi.org/10.1145/2808719.2808764.

in sociology and psychology, where it goes under the name *human agency in social cognitive theory* [4]. 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 [5].

Based on this observation we propose an ontology-based deep learning model for human behavior prediction. Our model extends a well-used deep learning method, Restricted Boltzmann Machines (RBMs) [32], with domain ontologies [9]. 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 [32], Convolutional Neural Networks (CNNs) [16], and Sum-Product Networks (SPNs) [27], take a flat representation of characteristics as an input. However, the characteristics are commonly in structural designs such as ontologies in the biomedical and health domain. Therefore, it would be better if a model can have the ability to learn the representations of individuals in health social networks from ontologies.

To address this issue, we 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 which replicates the original structure of personal characteristics. This representation structure is further used to model human behaviors in our health social network. Additionally, 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. Meanwhile, we define a statistical 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 which 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. The effectiveness of our model is verified by experiments on real and synthetic data from health social networks. Our main contributions of are as follows:

- We introduce an ontology-based deep learning model, ORBM, for human behavior prediction in health social networks.
- 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.
- Our experimental assessment on both real and synthetic data confirms that our model is very accurate in human behavior prediction.

In Section 2, we first introduce the related works and background in physical activity intervention, RBMs, and human behavior prediction. We then introduce the developed SMASH ontology in Section 3. We present our ontology-based deep learning algorithm for user representations in Section 4 and our human behavior prediction model in Section 5. The details of experimental evaluation are described in Section 6, and the work is concluded in Section 7.

2. RELATED WORKS AND BACKGROUND

2.1 Physical Activity Intervention

Regular physical activity decreases the risk of developing cardiovascular disease, diabetes, obesity, osteoporosis, some cancers, and other chronic conditions. Website-delivered physical activity interventions have the potential to overcome many of the barriers associated with traditional face-to-face exercise counseling or group-based physical activity programs. An Internet user can seek advice at any time, any place, and often at a lower cost compared with other delivery modalities [29]. In 2000, a set of articles that identified the potential of interactive health communications, including Internet and website-delivered interventions, for improving health behaviors were published [20, 24, 28]. Since then, over fifteen studies have been reported [36] that evaluate website-delivered interventions to improve physical activity. Better outcomes were identified when interventions had more than five contacts with participants and when the time to follow-up was short ($\leq$3 months; 60% positive outcomes), compared to medium-term (3-6 months, 50%) and long-term ($\geq$6 months, 40%) follow-up. A little over half of the controlled trials of website-delivered physical activity interventions have reported positive behavioral outcomes.

However, intervention effects were short-lived, and there was limited evidence of maintenance of physical activity changes. Although the website-delivered approaches reported positive results, research is needed to identify elements that can improve behavioral outcomes. Indeed, social networks have potential for being adopted, since they take the advantage of natural social relationships to deliver healthy behaviors. Furthermore, social networks can be a long-life environment, and thus the retention of participants could be improved. Human behavior prediction is a key component for further research since it offers us a powerful tool to understand the spread of physical activity in a health social network.

2.2 Human Behavior Prediction

Prediction of human social behavior has recently been studied, in such forms as analysis of user interactions on Facebook [37], activity recommendation [17], and user activity level prediction [41]. In [41], the authors focus on predicting users who have a tendency to reduce 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 and figure out the strategies to maintain such users. Social churn prediction has been studied in different applications, including online social games [14], Q & A forum [39], etc. The users in these applications usually have simple user behaviors. Meanwhile our models enrich the application area by incorporating various personal factors. The work that relates most closely to our study is the Socialized Gaussian Process Model (SGP) [31]. However, the SGP model is not designed to work with structural design of user characteristics.

2.3 Restricted Boltzmann Machines

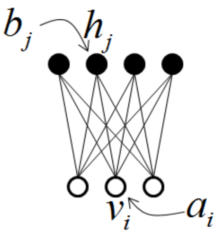


Figure 1: RBM

The Restricted Boltzmann Machines (RBMs) [32] 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 (Figure 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 [38]. To model real-valued data, a modified RBM with binary logistic hidden units and real-valued Gaussian visible units can be used. In Figure 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 units and hidden units. The RBM assigns a probability to any joint setting of the visible units, $\mathbf{v}$ and hidden units, $\mathbf{h}$:

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

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

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

where σ_i is the standard deviation of the Gaussian noise for visible unit i. In practice, fixing σ_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 (assuming $\sigma_i = 1$) are:

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

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

where $\sigma(.)$ is a logistic function, $\mathcal{N}(\mu, V)$ 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 [11]. The learning rule is:

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

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. The reconstructions are generated by starting a Markov chain at the data distribution. The hidden units can be updated by sampling Eq. 3, then updating the visible units by sampling Eq. 4.

To incorporate temporal dependencies into the RBM, the CRBM [34] adds autoregressive connections from the visible and hidden variables of an individual to his/her historical variables. The CRBM is effective at simulating the behavior of humans in the single agent scenario. However, it cannot

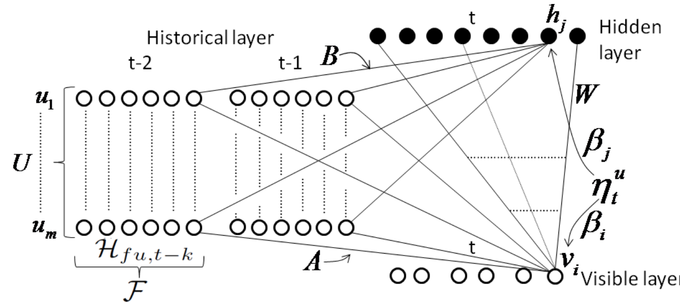


Figure 2: Social Restricted Boltzmann Machine [25].

capture the social influences on individual behaviors in the multiple agent scenario. Li et al. [18] 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, it is difficult to utilize personal characteristics in ctRBMs. Thus, the ctRBM cannot directly integrate personal characteristics with social influences to predict human behaviors.

2.4 Social Restricted Boltzmann Machine

In order to leverage the RBMs for human behavior prediction in social networks, we have proposed Social Restricted Boltzmann Machines (SRBM) [25]. Figure 2 illustrates the human behavior modeling in the SRBM model. The model includes three layers: the visible layer $\mathbf{v}$, hidden layer $\mathbf{h}$, and historical layer $\mathcal{H}$. Given a user, each visible variable v_i in the *visible layer* $\mathbf{v}$ corresponds to an individual feature f_i at time t. All the visible variables of all the users in the previous N time intervals $\{t-N,\ldots,t-1\}$ (i.e., $N < M$) are included in a *historical layer*, denoted by $\mathcal{H}_{t<}$. In addition, all the variables in the historical layer are called *historical variables*. Obviously, we will have $|\mathcal{F}| \times |U| \times N$ historical variables, where $\mathcal{F}$ is a set of individual features f. The hidden layer $\mathbf{h}$ consists of $|\mathbf{h}|$ hidden variables. In the SRBM model, visible layer, hidden layer, and historical layer are pair-wise fully bipartite-connected. The conditional distributions (assuming $\sigma_i = 1$) are:

$$p(h_{j,t} = 1|\mathbf{v}_t, \mathcal{H}_{t<}) = \sigma\big(\hat{b}_{j,t} + \sum_i v_{i,t} W_{ij}\big) \tag{6}$$

$$p(v_{i,t}|\mathbf{h}_t, \mathcal{H}_{t<}) = \mathcal{N}\big(\hat{a}_{i,t} + \sum_j h_{j,t} W_{ij}, 1\big) \tag{7}$$

where $\hat{b}_{j,t}$ and $\hat{a}_{i,t}$ are dynamic biases which are computed as:

$$\begin{aligned} \hat{b}_{j,t} &= b_j + \sum_{k\in\{1,\ldots,N\}} \sum_{f\in\mathcal{F}} \sum_{u\in U} B_{jfu,t-k} \mathcal{H}_{fu,t-k} + \beta_j \eta_t^u \\ \hat{a}_{i,t} &= a_i + \sum_{k\in\{1,\ldots,N\}} \sum_{f\in\mathcal{F}} \sum_{u\in U} A_{ifu,t-k} \mathcal{H}_{fu,t-k} + \beta_i \eta_t^u \end{aligned} \tag{8}$$

where β_i and β_j are parameters which present the ability to observe the social influences η_t^u from neighboring users of

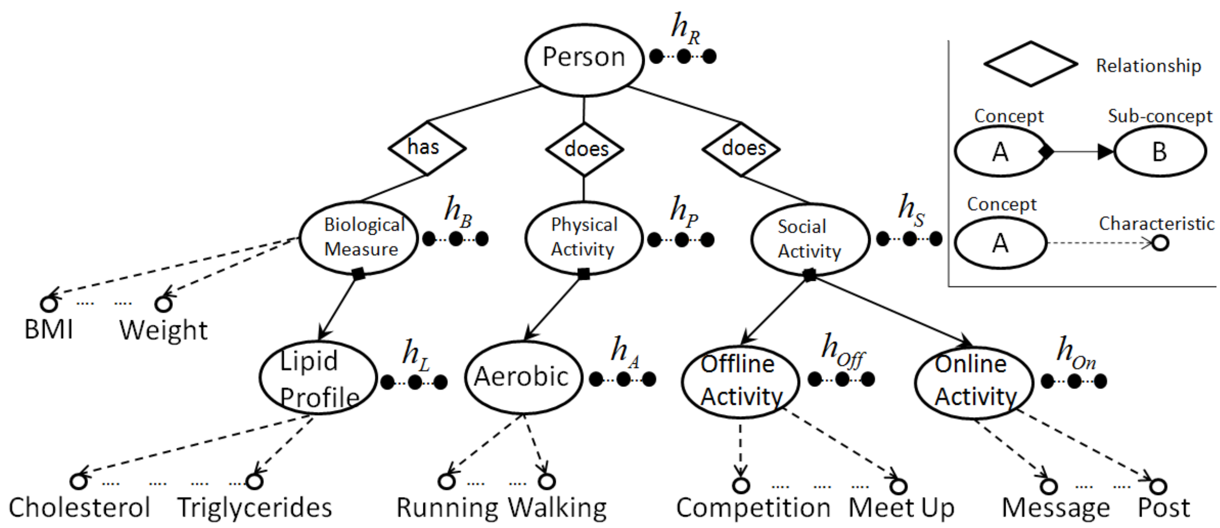


Figure 3: Partial View of the SMASH Ontology and its Hidden Variables.

user u given v_i and h_j. The energy function then becomes:

$$E(\mathbf{v}_t, \mathbf{h}_t|\mathcal{H}_{t<}, \theta) = \sum_{i\in\mathbf{v}} \frac{(v_{i,t} - \hat{a}_{i,t})^2}{2\sigma_i^2} - \sum_{j\in\mathbf{h}} \hat{b}_{j,t} h_{j,t} - \sum_{i\in\mathbf{v}, j\in\mathbf{h}} \frac{v_{i,t}}{\sigma_i} h_{j,t} W_{ij} \quad (9)$$

The SRBMs can accurately predict human behaviors since they capture the interactions among behavior determinants which are self-motivation, implicit and explicit social influences, and environmental events [4, 25]. In this paper, the SRBMs are adopted to predict physical activity behaviors in our YesiWell health social network.

3. THE SMASH ONTOLOGY

Ontology [10, 33] 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 [35]), Unified Medical Language System (UMLS [19]), and more than 300 ontologies in the National Center for Biomedical Ontology (NCBO [3]). 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 [7].

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 with 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, and support groups, etc. In the next step, these concepts and relationships are subsequently coded in the Web Ontology Language (OWL [1]) with *Protégé* [2]. 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, biomarker measures, physical activities, and social activities, in our SMASH ontology 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, that exchange thoughts and ideas, communicate information, and share emotions and experiences.
- Physical Activities: any bodily activity 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[1]. Figure 3 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 USER REPRESENTATION IN DEEP 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 Figure 3, the main concept is *Person*. With this concept, we have sub-concepts, relationships, and characteristics. Each person *has* a related concept *Biological Measure*

[1] `http://bioportal.bioontology.org/ontologies/SMASH`

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 $F_{\mathbb{S}}$, $\Theta_{\mathbb{S}} = \bigcup_{C \in C_{\mathbb{S}}} \mathbf{h}_C$ 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 $\Theta_{\mathbb{S}}$ in different ways by applying existing machine learning models such as linear or logistic regressions, and SVM. However, in this paper, we utilize the RBMs as a deep learning model since it very well fits to 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 (Figure 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:

$$p(h_j|V_{\mathbb{S}}, C_{\mathbb{S}}, F_{\mathbb{S}}) = \mathcal{N}\big(b_j + \sum_{v_i \in V_{\mathbb{S}} \cup \Psi_{\mathbb{S}} \cup \Theta_{\mathbb{S}}} v_i W_{ij}\big) \quad (10)$$

$$p(v_i|h_{\mathbb{S}}) = \mathcal{N}\big(a_i + \sum_{h_j \in h_{\mathbb{S}}} h_j W_{ij}\big) \quad (11)$$

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 [11], 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 [12] following the structure of the ontology H. For instance, in Figure 3, 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 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.

5. HUMAN BEHAVIOR PREDICTION WITH ONTOLOGY-BASED RBMS

In this section, we present how to conduct human behavior prediction with our Ontology-based RBM (ORBM) model. 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 result section.

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 $\{y_u^{t+1} | u \in U\}$.

Self-motivation and Environmental Events. Self-motivation is composed of many dimensions including attitudes, intentions, effort, and withdrawal which can all affect the motivation that an individual experiences [30]. In our YesiWell study, individual features are specially 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/receiving messages, the number of goals set and achieved, Wellness-score [15], etc. Wellness-score is a measure to evaluate how well a user lives his or her life. In general, being active in social activities, setting and achieving more goals, and obtaining higher Wellness-score results in better attitudes among users.
- *Intentions*: the number of competitions each user joins, the number of goals set, etc. Users are intent on exercising, and some join competitions and set additional goals.
- *Effort*: the number of days for exercise, walking/running steps, the distances, and speed.
- *Withdrawal*: BMI slope, Wellness-score slope [15] , etc. The increase of BMI slope or decrease of Wellness-score indicates negative signs in self-motivation. The users may temporarily give up their progress.

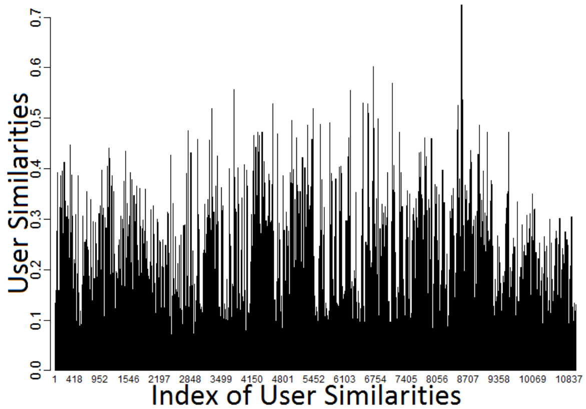


Figure 4: A sample of user similarity distributions.

Additionally, the effect of environmental events is composed of unobserved social relationships, unacquainted users, and the changing of social context [5, 25]. In other words, a user can be influenced by any users via any features in health social networks. It is hard to exactly define the influences of environmental events. Fortunately, the dynamic of the SRBM model [25] offers us a good solution to capture the flexibility of implicit social influences, as well as self-motivation. In fact, the individual features of a user, denoted as $\mathcal{F}_u$, can be considered as the visible variables in the SRBMs (Figure 2). 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 Eq. 8 as well.

The quantitative environmental events, such as the number of competitions and meet-up events, are included in individual characteristics. Therefore, the effect of environmental events is better embedded into the model. Next, we will incorporate the social influence into the SRBM model [25].

Social Influences. It is well-known that individuals tend to be friends with people who perform similar behaviors as them (*homophily principle* [22]). Extended from our work in [25], we define user similarity given two neighboring users u and m as a *cosine function* of their individual features (i.e., $\mathbf{v}^u$ and $\mathbf{v}^m$) and their hidden features (i.e., $\mathbf{h}^u$ and $\mathbf{h}^m$). The user similarity between u and m at time t, denoted $s_t(u, m)$, is defined as:

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

where $cos_t(\cdot)$ is the cosine similarity function, $cos_t(u, m|\mathbf{v})$ is the cosine similarity of $p(\mathbf{v}_t^u|\mathbf{h}_t^u, \mathcal{H}_{t<}^u)$ and $p(\mathbf{v}_t^m|\mathbf{h}_t^m, \mathcal{H}_{t<}^m)$, and $cos_t(u, m|\mathbf{h})$ is the cosine similarity of $p(\mathbf{h}_t^u|\mathbf{v}_t^u, \mathcal{H}_{t<}^u)$ and $p(\mathbf{h}_t^m|\mathbf{v}_t^m, \mathcal{H}_{t<}^m)$.

Figure 4 illustrates a sample of user similarity spectrum (i.e., $s_t(\cdot, \cdot)$) 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). This quality of similarities demonstrates the social influences of local neighbors on individuals.

Definition 1. *The explicit social influence η_t^u of a user u at time t is defined as an exponential similarity average of the cumulative distribution function (CDF) of the instant similarity over the user similarity spectrum, i.e.,*

$$\eta_t^u = \frac{1}{|\mathbb{Z}_t^u|} \sum_{m=1}^{|U|} l_t(u, m) \times p\big(s_t \leq s_t(u, m)\big) \tag{12}$$

where $\mathbb{Z}_t^u = \sum_{m=1}^{|U|} l_t(u, m)$, and the indicator function l_t is 1 if user u is connected to user m until time t (i.e., $e_{u,m} \in E_t$), and 0 otherwise. 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 similarity is less than or equal to the instant similarity $s_t(u, m)$.

The effect of explicit social influences, η_t^u, is integrated to the dynamic biases of visible and hidden variables (Eq. 8) in the SRBMs.

Inference and Learning. Inference in the ORBM model is no more difficult than in the SRBMs. The states of the hidden variables are determined by both the input they receive from the visible variables and the input they receive from the historical variables. The conditional probability of hidden and visible variables at time interval t can be computed as in Equations 6, 7, and 8. The energy function is similar to the SRBMs (i.e., Eq. 9).

We can use *contrastive divergence* [11] for training the ORBM. The updates for the symmetric weights, W, as well as the static biases, a and b, have the same form as Eq. 5. However, they have a different effect because the states of the hidden and visible variables are now influenced by the implicit and explicit social influences. The updates for the directed weights, A and B, are also based on simple pairwise products. The gradients are summed over all the training time intervals $t \in T_{train} = T_{data} \backslash \{t-M+1, \ldots, t-M+N\}$. We train the ORBM for each user independently. At any time we update the parameters, we will update the explicit social influences for all the users.

Human Behavior Prediction. On top of the ORBM model, we put a softmax layer for the user behavior prediction task. Our goal is to predict whether a user is *active* or *inactive* in physical exercises. Thus the softmax layer contains a single output variable $\hat{y}$ and binary target values: 1 for active, and 0 for inactive. The output variable $\hat{y}$ is fully linked to the hidden variables by weighted connections S which includes $|\mathbf{h}|$ parameters s_j. We use the logistic function as an activation function to saturate the two target values, i.e.,

$$\hat{y} = \sigma(c + \sum_{j \in \mathbf{h}} h_j s_j) \tag{13}$$

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:

$$P(Y|X, \theta) = \prod_{t \in T_{train}} \hat{y}_t^{y_t} (1 - \hat{y}_t)^{1 - y_t} \tag{14}$$

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

$$C(\theta) = -\sum_{t \in T_{train}} \Big(y_t \log \hat{y}_t + (1 - y_t) \log(1 - \hat{y}_t)\Big) \tag{15}$$

Table 1: Personal Characteristics.

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

As this last step of the training, *Back-propagation* is used to fine-tune all the parameters together including the parameters in user representation learning based on ontologies (i.e., Section 4). The derivatives of the objective $C(\theta)$ with respect to all the parameters over all the training cases $t \in T_{train} = T_{data} \setminus \{t-M+1, \dots, t-M+N\}$.

Causal Generation. In the prediction task, we need to predict the y^u_{t+1} without observing $\mathcal{F}^{t+1}_u$. 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 trained ORBM can be done just like the learning procedure. 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. To start alternating Gibbs sampling, a good choice is to set $\mathbf{v}_t = \mathbf{v}_{t-1}$, since $\mathbf{v}_{t-1}$ can be considered as 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 variable $\hat{y}$.

6. EXPERIMENTS

We have carried out a series of experiments using datasets from both real-world and synthetic health social networks to validate our proposed ORBM model (source codes and data[2]). We first elaborate the experiment configurations on the data sets, evaluation metrics, and baseline approaches. Then, we introduce the experimental results.

The YesiWell data and Experiment Configurations. The YesiWell social network data were collected from Oct 2010 to Aug 2011 as a collaboration between PeaceHealth Laboratories, SK Telecoms 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 information of the number of walking and running steps, were reported via a mobile device carried by each user. All users enrolled in an online social network allowing them to friend and communicate with each other. Users' biomarkers and biometric measures are recorded via daily/weekly/monthly medical tests performed at home (i.e., individually) or at our laboratories.

[2]https://www.dropbox.com/s/ciajf63c89besbk/ORBM.rar?dl=0

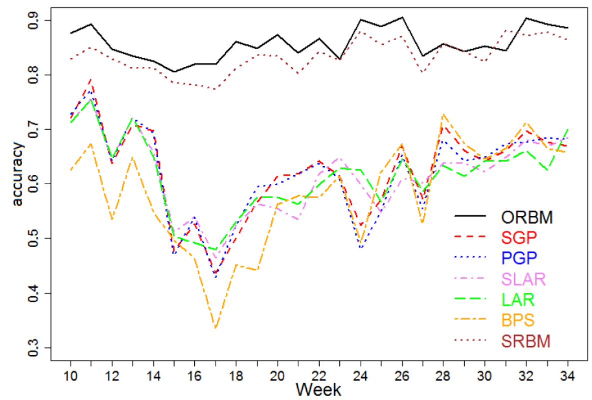


Figure 6: ORBM vs baselines in terms of accuracy.

Table 2: ORBM vs state-of-the-art models in terms of accuracy in the whole data set.

SGP	PGP	SLAR	LAR	BPS	SRBM	ORBM
0.677	0.684	0.663	0.662	0.675	0.83	**0.859**

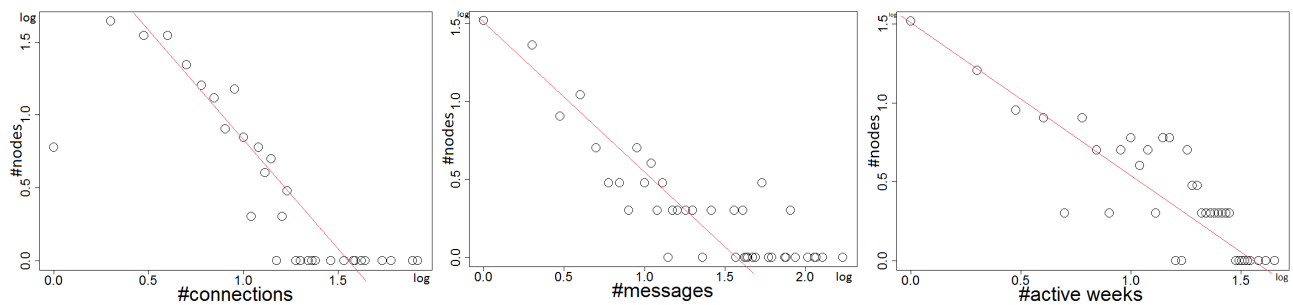
In total, we have 30 features taken into account (Table 1). All the features are weekly summarized. Figure 5 illustrates the distributions of friend connections, and the number of received messages in the health social network. They clearly follow the Power law distribution. The number of hidden units and the number of previous time intervals N respectively are set to 200 and 3. In the user representation learning, the number of hidden units at all the concepts and sub-concepts in the SMASH 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} [11] is used for maximum likelihood learning. We train the model for each user independently.

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

$$accuracy = \frac{\sum_{i=1..|U|} I(y_i = \hat{y}_i)}{|U|} \qquad (16)$$

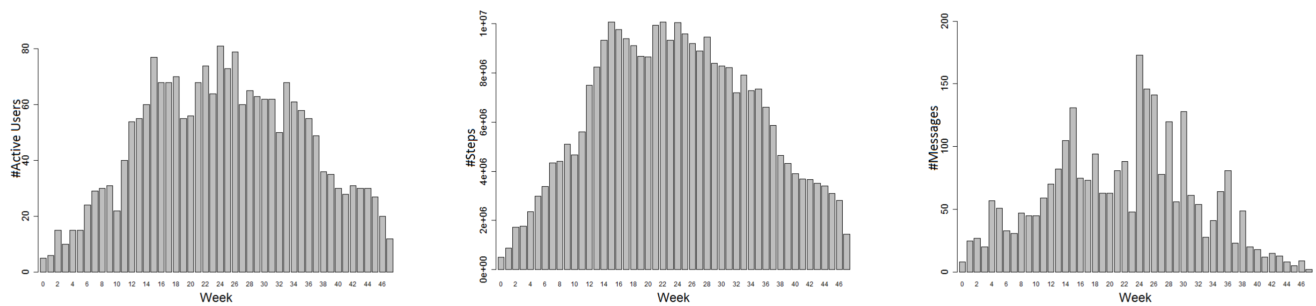
where y_i is the true user activity of the user u_i, and $\hat{y}_i$ denotes the predicted value, I is the indication function.

Competitive Prediction Models. We compare the ORBM model with the conventional methods reported in [31]. 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 individuals' past behavior records and his or her friends' past behaviors for prediction. Specifically, five models reported in [31] are Socialized Gaussian Process (**SGP**) model, Socialized Logistical Autoregression (**SLAR**) model, Personalized Gaussian Process (**PGP**) model, Logistical Autoregression (**LAR**) model, and Behavior Pattern Search (**BPS**) model. In addition, we consider a **SRBM** model [25] which is similar to ORBMs in the learning and inference process but does



(a) (b) (c)

Figure 5: The distributions of friend connections (a), inbox messages (b), and active users (c) in YesiWell study.



(a) The number of active users (b) The number of walking and running steps (c) The number of received messages

Figure 7: Users' activities of our health social network over time.

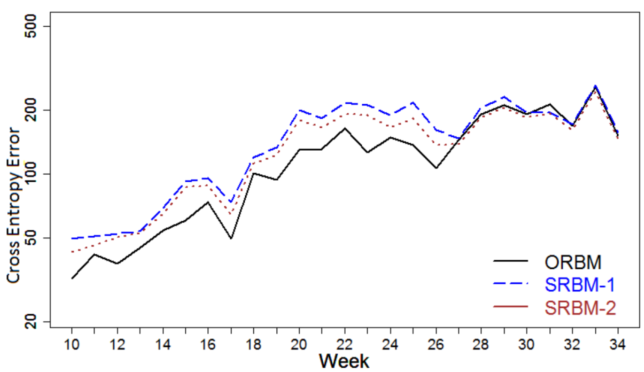


Figure 8: ORBM vs SRBM in terms of cross-entropy errors (i.e., Equation 15). Lower is better.

not use ontologies to generate user representations. The SRBM model treats all personal features flatly as visible units in the model because it does not consider semantics (e.g., concept hierarchy) in the ontology.

6.1 ORBM vs State-of-The-Art Models

Figure 6 and Table 2 show the accuracy comparison over 25 weeks (i.e., weeks 10^{th}-34^{th}) with the YesiWell dataset. It is clear that the ORBM outperforms the other models. One of the most interesting points is the accuracy curves of the models. The accuracies tend to drop in the middle period of the study. To explain this, we illustrate the summarized activities of all the users in the whole dataset in Figure 7. In Figure 7a, we can see that the number of active users is significantly higher in the middle weeks than the beginning and ending weeks. Thus many users will become active even though they have never been active before. If we only used the exercise status of the users in the beginning weeks, we would not have enough information to predict the users' behavior in the middle weeks. The existing models do not capture the influences of environmental events. As a result, they give rise to a lot of noises in the prediction results. Consequently, they have low and unstable performances at the middle weeks.

Meanwhile, the ORBM and SRBM, two deep learning models, well capture the self-motivation, influences of environmental events, and social influences which become stronger at the middle weeks (i.e., all the activities such as social communications and physical activities are improved). In addition, the correlation between the personal features and the hidden social influences can be adequately detected by the hidden variables. Thus, much information has been leveraged to predict individual behaviors. Our models not only achieve higher prediction accuracy but also the performance is stable over time. Overall, the ORBM model achieves **0.859** accuracy. Meanwhile, the SRBM model achieves **0.83** (Table 2).

6.2 The Effectiveness of User Representation Learning using the SMASH Ontology

In Figure 6, the ORBM and SRBM models both achieve higher prediction accuracy than state-of-art prediction models because of deep learning. To show the advantage of using the SMASH ontology in the ORBM model, we further compare ORBM with SRBM in more detail. The ORBM outperforms the SRBM-1 and the SRBM-2 ('1' and '2' indicate the number of hidden layers) by 3% (0.859 vs 0.83). We do not distinguish SRBM-1 and SRBM-2 in Figure 6 since they achieves the same results. From another perspective, the ORBM model has better cross-entropy errors (i.e., Equation 15) compared with the two SRBM models. The detailed comparisons are shown in Figure 8. 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-1 to be the SRBM-2, 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 software Pajek[3] to generate graphs under the 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 and average node degree is 5.4 (i.e., similar to the real YesiWell data). Then, we apply PATH [40] 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[4]. 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 which imitates our real-world dataset. Figure 9 shows the accuracies of the conventional models, SRBM model, and the ORBM model on the synthetic data. We can see that the ORBM model still outperforms the conventional models in terms of the prediction accuracy. In Figure 9, the SRBM is used to indicate SRBM-1 and SRBM-2 which achieve the same accuracy in terms of prediction.

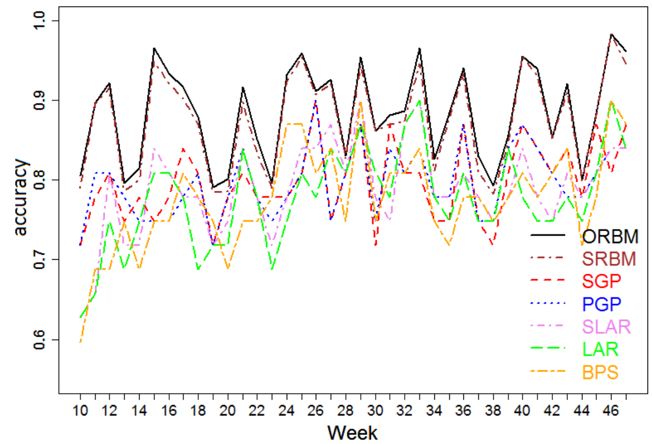


Figure 9: Accuracies on the synthetic data.

7. CONCLUSIONS AND FUTURE WORKS

This paper introduces ORBM, a novel ontology-based deep learning model for human behavior prediction in health social networks. We contribute several novel techniques to deal with health social network ontologies, self-motivation, social influence, and environmental event modeling. We first propose a bottom-up algorithm to learn user representations given the ontologies. Then, we build up a deep learning model to incorporate human behavior determinants which are self-motivation, social influences, and environmental events from user representations. Our empirical analysis over real and synthetic health social networks illustrates that our ORBM model predicts the future activity levels of users more accurately and stably than conventional methods. More importantly, the experiment also emphasizes the three meaningful observations: 1) the SMASH ontology helps us to organize data features in an suitable way; 2) our algorithm can learn meaningful user representations from ontologies; and 3) meaningful user representations could further improve accuracies of deep learning approaches.

Our work can be extended in several directions. First, we can leverage the ontologies to generate descriptive explanations for predicted behaviors. Second, the approach explored in this paper is rooted on the RBM [32]. However, other alternatives are possible, which can be based on CNNs [16] or Sum-Product Networks [27]. We plan to explore and compare these different strategies in future work.

[3] `http://vlado.fmf.uni-lj.si/pub/networks/pajek/`
[4] `http://cbio.ensmp.fr/graphm/`

Acknowledgment

This work is supported by the NIH grant R01GM103309 to the SMASH project. We thank Xiao Xiao, Rebeca Sacks, and Ellen Klowden for their contributions.

8. REFERENCES

[1] OWL Web Ontology Language. `http://www.w3.org/TR/owl-ref/`.

[2] The Protégé Ontology Editor and Knowledge Acquisition System. `http://protege.stanford.edu/`.

[3] The National Center for Biomedical Ontology. `http://www.bioontology.org/`.

[4] A. Bandura. Human agency in social cognitive theory. *The American Psychologist*, 44(9):1175–84, 1989.

[5] N. Christakis. The hidden influence of social networks. In *TED2010*, 2010.

[6] N. Christakis and J. Fowler. The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*, 357:370–9, 2007.

[7] D. Dou, H. Wang, and H. Liu. Semantic Data Mining: A Survey of Ontology-based Approaches. In *ICSC'15*, pages 244–251, 2015.

[8] R. Fichman and C. Kemerer. The illusory diffusion of innovation: An examination of assimilation gaps. *Information Systems Research*, 10(3):255–275, 1999.

[9] T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[10] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-computer Studies*, 43(5):907–928, 1995.

[11] G. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[12] G. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.

[13] J. Huang, X.-Q. Cheng, H.-W. Shen, T. Zhou, and X. Jin. Exploring social influence via posterior effect of word-of-mouth recommendations. In *WSDM'12*, pages 573–582, 2012.

[14] J. Kawale, A. Pal, and J. Srivastava. Churn prediction in mmorpgs: A social influence based approach. In *CSE'09*, pages 423–428, 2009.

[15] D. Kil, F. Shin, B. Piniewski, J. Hahn, and K. Chan. Impacts of social health data on predicting weight loss and engagement. In *O'Reilly StrataRx Conference, San Francisco, CA*, October 2012.

[16] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[17] K. Lerman, S. Intagorn, J.-K. Kang, and R. Ghosh. Using proximity to predict activity in social networks. In *WWW'12 Companion*, pages 555–556, 2012.

[18] X. Li, N. Du, H. Li, K. Li, J. Gao, and A. Zhang. A deep learning approach to link prediction in dynamic networks. In *SDM'14*, pages 289–297, 2014.

[19] D. Lindberg, B. Humphries, and A. McCray. The Unified Medical Language System. *Methods of Information in Medicine*, 32(4):281–291, 1993.

[20] B. Marcus, C. Nigg, D. Riebe, and L. Forsyth. Interactive communication strategies: implications for population-based physical activity promotion. *American Journal of Preventive Medicine*, 19(2):121–6, 2000.

[21] A. Marshall, E. Eakin, E. Leslie, and N. Owen. Exploring the feasibility and acceptability of using internet technology to promote physical activity within a defined community. *Health Promotion Journal of Australia*, 2005(16):82–4, 2005.

[22] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[23] R. Pate, M. Pratt, S. Blair, and et al. Physical activity and public health. a recommendation from the centers for disease control and prevention and the american college of sports medicine. *The Journal of the American Medical Association*, 273(5):402–7, 1995.

[24] K. Patrick. Information technology and the future of preventive medicine: potential, pitfalls, and policy. *American Journal of Preventive Medicine*, 19(2):132–5, 2000.

[25] N. Phan, D. Dou, B. Piniewski, and D. Kil. Social restricted boltzmann machine: Human behavior prediction in health social networks. In *ASONAM'15*, 2015 (To appear).

[26] N. Phan, D. Dou, X. Xiao, B. Piniewski, and D. Kil. Analysis of physical activity propagation in a health social network. In *CIKM'14*, pages 1329–1338, 2014.

[27] H. Poon and P. Domingos. Sum-product networks: A new deep architecture. In *IEEE ICCV Workshops*, pages 689–690, 2011.

[28] J. Prochaska, M. Zabinski, K. Calfas, J. Sallis, and K. Patrick. Pace+: interactive communication technology for behavior change in clinical settings. *American Journal of Preventive Medicine*, 19(2):127–31, 2000.

[29] L. Ritterband, L.A.G.-Frederick, D. Cox, A. Clifton, R. West, and S. Borowitz. Internet interventions: in review, in use, and into the future. *Professional Psychology: Research and Practice*, 34:527–34, 2003.

[30] R. M. Ryan and E. L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54 – 67, 2000.

[31] Y. Shen, R. Jin, D. Dou, N. Chowdhury, J. Sun, B. Piniewski, and D. Kil. Socialized gaussian process model for human behavior prediction in a health social network. In *ICDM'12*, pages 1110–1115, 2012.

[32] P. Smolensky. Information processing in dynamical systems: Foundations of harmony theory. chapter Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, pages 194–281. 1986.

[33] R. Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1):161–197, 1998.

[34] G. Taylor, G. Hinton, and S. Roweis. Modeling human motion using binary latent variables. In *NIPS'06*, pages 1345–1352.

[35] The_gene_ontology_consortium. Creating the gene ontology resource: design and implementation. *Genome Research*, 11(8):1425–1433, August 2001.

[36] C. Vandelanotte, K. Spathonis, E. Eakin, and N. Owen. Website-delivered physical activity interventions: A review of the literature. *American Journal of Preventive Medicine*, 33(1):54–64, 2007.

[37] B. Viswanath, A. Mislove, M. Cha, and K. Gummadi. On the evolution of user interaction in facebook. In *WOSN'09*, pages 37–42, 2009.

[38] M. Welling, M. Rosen-Zvi, and G. Hinton. Exponential family harmoniums with an application to information retrieval. In *NIPS'04*, pages 1481–1488, 2004.

[39] J. Yang, X. Wei, M. Ackerman, and L. Adamic. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. In *ICWSM'10*, 2010.

[40] M. Zaslavskiy, F. Bach, and J.-P. Vert. A path following algorithm for graph matching. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 5099, pages 329–337. 2008.

[41] Y. Zhu, E. Zhong, S. Pan, X. Wang, M. Zhou, and Q. Yang. Predicting user activity level in social networks. In *CIKM'13*, pages 159–168, 2013.