Future Generation Computer Systems 93 (2019) 952–961

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Multilevel learning based modeling for link prediction and users' consumption preference in Online Social Networks

Pradip Kumar Sharma, Shailendra Rathore, Jong Hyuk Park *

Department of Computer Science and Engineering, Seoul National University of Science and Technology, (SeoulTech), Seoul 01811, Republic of Korea

HIGHLIGHTS

- We explore users' behaviors in the OSN platform.
- Direct and latent models are proposed for link prediction and users' consumption preferences.
- To achieve high accuracy, a multilevel DBN learning based model is proposed.
- The proposed model offers significantly improved performance compared to other methods.

ARTICLE INFO

Article history:
Received 29 April 2017
Received in revised form 19 July 2017
Accepted 16 August 2017
Available online 25 August 2017

Keywords:
Online Social Networks
Link prediction
Deep learning

ABSTRACT

The problem with predicting links in Online Social Networks (OSNs) is having to estimate the value of a link that can represent the relationship between social media users. The evolution of the OSN is influenced by the structure of the social network and the interaction between the preferential behaviors of users that have long converged by sociologists. However, conventional methods treat these behaviors in isolation. Therefore, the roles of users' historical preferences and the dynamic structure of the social network are still not clear as to how these things affect the evolution of the OSN. Link prediction for new users who have not created a link or a small network is a fundamental problem in OSNs. To start creating social networks for such users, these behaviors can be used to recommend friends and user consumption preferences. In this paper, we propose novel direct and latent models to represent link prediction and a user's consumption preferences in an OSN platform. We also introduce a multilevel deep belief network learning-based model for link prediction and a user's consumption preferences to achieve high accuracy. To evaluate the performance of our model, we elaborated several performance measures and used datasets from Facebook, Amazon and Google+ to validate the accuracy. The result of our evaluation shows that our proposed model provides significantly improved performance for link prediction and user preferences over other methods.

1. Introduction

Nowadays, there is a large amount of information from social media due to the explosive growth of OSN websites. In OSNs, social connections such as follow-follower on Twitter, and Facebook' friends play a significant role in users' experiences and in the achievement of the OSN. To facilitate the development of social relationships between those who share similar activities, interests or real-life ties, OSNs offer an online platform. Users can use social media more often if the links of a user are well established. In this manner, individuals can stay connected with others and learn about the social preferences of their online friends and acquaintances. In this way, an excellent link prediction is required in order to permit OSNs to suggest valuable connections to users. In particular, for a new user who has not made a connection or has a small network, link prediction is considerably more critical because it can be used to accompany companions so that new users are able to create their social networks. For new users, a poor prediction may dispirit when using the platform. The connection is generally a mark with sign esteem that represents the type of opinion from one user to another, for example, the expression of support or contradict [1,2].

Linking in an OSN reflects the choice through individual preferences. As such, a link prediction method strategy can be produced to discover individual preferences according to the current network structure and available attributes of the individual. To reveal and represent individual preferences, utility analysis is usually

* Corresponding author.
E-mail addresses: pradip@seoultech.ac.kr (P.K. Sharma), rathoreshailendra@seoultech.ac.kr (S. Rathore), jhpark1@seoultech.ac.kr (J.H. Park).

http://dx.doi.org/10.1016/j.future.2017.08.031
0167-739X/

viewed as a standout amongst the most proficient tools. Regarding utility analysis, to connect with others, by maximizing their utility functions, individuals in a network are seen as intelligent agents who can make decisions. The benefits that can be derived from utility analysis were a key motivator in our developing a new link prediction method which is presented in this paper.

Apart from creating social links behavior of OSN users, their consumption preferences behavior that we must be taken into consideration. Users' consumption preferences behavior reflected in user–user interactions, such as buying some product, consuming of local food, etc. While individuals are faced with a dazzling array of potential consumer products, discovering user preferences is another task of behavioral prediction in an OSN platform. Individuals are likely to consume products that are locally well known among individuals with similar mentalities with similar consumption histories. However, sociologists have long recognized that the building of social bonds and the behaviors of users' consumer of products are not isolated [3]. Rather, their interactions lead to the evolution of OSNs, which leads to dynamic changes in user preferences and to the structure of the social network over time. The effect of social influence indicates that the future preferences of users are affected by the social network that surrounds them and people associate with individual who have similar preferences in the past.

According to the motivations outlined above, this research work aims to accomplish the following contributions:

- We explore users' behaviors in the OSN platform and several social features for link prediction and consumption preferences for new users who have not created a link or created a small network.
- We propose a novel direct model to represent link prediction and users' consumption preferences in the OSN platform.
- We also propose a novel latent model to represent link prediction and users' consumption preferences in OSN platform.
- To achieve high accuracy, we propose a multilevel deep belief network learning based model for link prediction and user's consumption preferences in the OSN platform.
- In addition, several performance matrices are elaborated to validate the accuracy of our proposed method.
- To evaluate our method using design performance matrices, we used Facebook, Amazon and Google+ datasets. The result of this evaluation demonstrates that our proposed model has provides significantly improved performance for link prediction and user preferences over other methods.

The rest of the paper is structured as follows: In Section 2, we discuss related work on link prediction, user consumption preference, and deep learning. In Section 3, we are present our new methodology to address link prediction and user consumption preferences in the OSN platform. In Section 4, we evaluate our proposed model based on different performance metrics. Finally, we present the conclusions of our research in Section 5.

2. Related works

2.1. Link prediction and user consumption preference

Link prediction In recent years, numerous strategies for link prediction have been proposed. These strategies can be categorized into different groups: methods based on probabilistic models, methods based on machine learning and methods based on similarity.

The method based on similarity is the most used method for link prediction. In this technique, each pair of nodes receives an index, which is defined as the comparability score between the two nodes. Based on the scores, all unobserved links are ranked, and unobserved links linking more similar nodes are assumed to have higher likelihoods of existence. Using fundamental topological features, we can define the similarity of the node. If the two nodes have correlated topological structures or more common characteristics, then they receive a higher similarity score [4,5]. Many reviews have discovered generous levels of topical closeness among people who are located near each other. For example, Luca Maria Aiello et al. [6] studied the prediction of friendship in social networks according to the presence of homology in three systems that combine the join labeling web-based social networking with OSNs. Structural similarity scores can be categorized into quasi-local, global and local indices. The local indices only use information about the neighbors of the nodes. Typical local indices include the resource allocation index, Adamic–Adar index, the index of Preferential addition, the Hub Promoted index, the depressed Hub index, the Leicht–Holme–Newman index, the Jaccard index, the Sorensen index, the Salton index, and common neighbors [7]. Quasi-local indices use more information than local indices. And do not need global topological information. This type of index includes the superimposed random walk, local random walk, and local path index [7]. Global indices such as the Matrix Forest Index, Leicht–Holme–Newman Index, and the Katz Index, need global topological information [7].

In link prediction strategies, machine learning strategies are also exploited. For link prediction in complex networks, a supervised classification aggregation method has been proposed by Manisha Pujari et al. [8]. A continuous-time regression model was proposed Duy Q. Vu et al. [9]. The model can integrate time-varying regression coefficients and time-dependent network statistics. In the link prediction task, in order to use potential information in a large number of unbound node pairs in networks, Zhengzhong Zeng et al. [10] introduced a technique that consolidates semi-supervised learning. Yu-lin He et al. [11] proposed a link-based prediction algorithm based on the set on the basis of a weighted average operator. In this method, to obtain final prediction scores for information-based link prediction algorithms, the method assigns weights to nine local algorithms and aggregates their results. Based on optimization of ant colonies, an unsupervised structural link prediction algorithm proposed by Ehsan Sherkat et al. [12]. Based on multi-resolution community sharing, Jingyi Ding et al. [13] have proposed a technique for predicting potential links. Bolun Chen et al. [14] proposed an approximate algorithm for predicting links related to nodes of interest.

In recent years, some other methods that are based on multi-relational networks and probabilistic models have been proposed. Yang Yang et al. [15] proposed a probabilistic technique in heterogeneous multi-relational networks for link prediction. They also presented an unsupervised learning technique for connection expectation in heterogeneous systems. Victor Stroele et al. [16] presented a technique that uses clustering approaches in multi-relational social scientific networks for proving maximum flow measurements. However, regarding link prediction, none of these techniques in multi-relational networks take the influence between the different relationships. Whereas, for probabilistic models, a stochastic subject model for predicting the link on the graphs assigned and assigned to the nodes proposed by Nicola Barbieri et al. [17]. Their model predicts links and produces explanations for each predicted link. Using a constraint-based genetic algorithm, to detect human movement and advanced a method of labeling human movement in a social network is presented by Fuyuan Hu et al. [18]. However, in real world applications, it is difficult to know the distribution of the appearance of the links in advance which is necessary in the probabilistic model.

User consumption preference: The behavior of the social link and users' consumption behaviors are not isolated, rather they have

a mutually reinforcing relationship. More specifically, the social influence hypothesis shows that individuals tend to associate and link with other users who have comparative inclinations and properties, and the impact recommends that the connections between users would assist influence users to carry on correspondingly with their companions. Hence, the researchers contended that we can use one sort of information data source for the rest of the prediction assignment. Among them, the social recommendation system uses information about the structure of the social network to mitigate the issue of saving data in conventional referral systems [19,20]. Integrating users' historical consumer preferences is another way to suggest potential links between users. For more accurate prediction of link, a technique to exploit users' historical preferences is proposed by Tang et al. [21]. They contended that users with comparable historical interests will probably construct social connections later on. Gong et al. proposed to argue the social network in a network of social attributes, so that the attribution of the node and the information of the social network could be informed [22]. Nevertheless, all of these methods were based on a static assumption about the OSN platform and could not be used to model the evolution of OSNs.

2.2. Deep learning

With several levels of abstraction to learn data representations, deep learning is an approach that permits computational models with numerous of layers of processing defined by Yann LeCun et al. [23]. Deep learning is a model with more complex training schemes and architectures compared to artificial neural networks. Deep learning could several layers hidden with each layer can have distinctive functions; whereas, a neural network includes of an input, a hidden and an output layers. As compare to traditional neural networks, deep learning training scheme is different. To guarantee the strength of deep learning models, new model parameters such as dropping out which specifies the extent of neurons to be arbitrarily disregarded amid model preparing are introduced.

For prediction and feature extraction, deep leaning can be used in supervised and unsupervised ways to build a deep neural network and an auto-encoder model. It takes a significant amount of effort to transform the raw data into meaningful functionality to guarantee the achievement of traditional machine learning procedures. To extract features in other domains, unsupervised depth learning has been widely used in an effective manner [24].

3. Proposed methodology

The purpose of this research is to consider a newly signed users in OSN who has created no social link or user who has created very few social links. In general, users perform two types of behavior on most OSN platforms in which they create social links and consume products. Given a non-directed social network graph, we have attempted to determine the users on the given graph to which new users will connect and will prefer the consumption of products. In this section, we formulate the problem of link prediction and user preferences and describe the solution.

3.1. Problem statement

In an OSN platform, there are usually two sets of entities: a set of products $P = (|P| = N)$ and a set of users $U = (|U| = M)$. In general, the consumption of a product refers to the interaction between a user and a product. Different OSN platforms have different sets of products. In an OSN platform, most of the users engage in both types of behavior, which leads to dynamic changes in social network structures and users' preferences over time.

Let $G = (U, E)$ be a given non-directed OSN graph, where U denotes a collection of existing users in the OSN graph, and E defines the collection of edge $e\langle u, w\rangle$ that describes non-directed links between users u and w, where $u, w \in U$. Generally, OSN users provides other attributes related to their personal information, such as place of employment, organization, college, city, and age. Therefore, we created a vector of users' personal attributes that is represented as $Q = \langle q_1(u), q_2(u), \ldots, q_m(u)\rangle$. Apart from attributes, all of the links of user u in a set of his or her friends are gathered. This is represented as $F(u) = \{f | f \in U \cap e\langle u, f\rangle \in E\}$. By using the attributes vector Q and all links $F(u)$, we can represent an existing user u as a tuple $u : \langle q_1(u), q_2(u), \ldots, q_m(u), F(u)\rangle$, where some element of the tuple may have a value of null value. This is because there is the possibility that user u does not have all attributes or reveal set of his or her friends. There is no pre-exist links are available for new users, usually, they are asked by OSNs for providing personal information at the time of sign up. Therefore, a new user s where no existing links (i.e., no friend sets) and only attribute vector is available, can be denoted as tuple $s : \langle q_1(s), q_2(s), \ldots, q_m(s)\rangle$. For selecting desirable users $u \in U$ from the set of existing users to construct a link by s, a candidate set of existing users (i.e., J) is classified into two sets of de-linked users (i.e., K) and linked users (i.e., I); that is $J = I \cup K$, where $J \in U$. It is assumed that the users in the set I have a great probability for being linked by s as compared with users in the set K. In the paragraph below, we describe the problem of predicting links with new users.

For a given OSN graph $G = (U, E)$, where for every user, a friend set and an attribute vector is defined as $u : \langle q_1(u), q_2(u), \ldots, q_m(u), F(u)\rangle$, and $J \subseteq U$ denotes a set of existing user candidates, $s : \langle q_1(s), q_2(s), \ldots, q_m(s)\rangle$ represents an attribute vector of given new user s, then prediction of users in J with them, a link may not be created by s is labeled as K (de-linked-users) and a link may be created is labeled as I (linked-users).

More precisely, we represented the two types of user behavior each time t as two the matrices of a social link matrix $S^t \in R^{MxM}$ and a consumption matrix $P^t \in R^{MxN}$. If user u consumes item e at t, P^t_{ue} represents the rating preference score; otherwise the value is 0, which represents no preference over the time t. In the same way, the value of S^t_{uw} is 1 if there is a connection between user u and user w at time t; otherwise the value is 0.

3.2. Direct modeling for link prediction and user preference

Direct link prediction modeling: Given a set J of existing users that present friends and attributes, and a new user s who shares his or her certain attributes of his or her profile, the basic idea exploits all available information to define existing users from J as the linked-users (I) who shows the similarity with $s$because many existing works have verified that the probability of connecting people with one another on OSN is depends on the similarity between them. As such, we modeled the probability of the friend, that provides a probability measurement of the creation of a link by s with $u \in J$, by calculating their similarity according to their information such as u's friends' attribute vector, u's friend set, and s and u's attribute vector. A link prediction model that combines multiple social features ω_{us} to define the probability of friends is trained by generating a set of training data. To generate a training dataset, the information is gathered from several user pairs. Here, multiple social features y_i and label l_i are corresponded to each pair of users. By default l_i is assigned the value of 0 only when l_i is assigned the value 1 if a friendship exists between two users. To ensure that the parameterized combination of social functionality describes the connectivity model between users, we aimed to form a set of x parameters using the dataset. By taking the social characteristics parameterized by the trained x, e can calculate the probability of friendship between u and s. Finally, using this probability measure,

it is determined for the new user s whether u belongs to the set K or I. Therefore, the optimization problem can be solved by constructing the link prediction model as follows:

$$MinF\,(x) = \frac{1}{2}\,\|x\|^2 + \partial \sum_{j=1}^{o} \alpha_j,\ \begin{cases} \alpha_j \geq 0 \\ z_i \left\langle x, y_j \right\rangle \geq 1 - \alpha_j. \end{cases} \tag{1}$$

Where j denotes the jth pair, o represents the total number of user pairs, ∂ is a constant and α_j $(j = 1, \ldots, o)$ are loose variables for optimization.

Direct user preference modeling: The main motives for creating a new consumption record are as follows: In traditional collaborative filtering [1] models assume that a user likes to consume products that are locally popular among users who have a similar historical consumer preference. Furthermore, the theory of social influence suggests that users are likely to be influenced by the preferences of neighbors of the social network to make decisions in preference to consumption. We formulated user consumption preference as:

$$P_{ue}^{t} = (1 - \varphi_u)\, g\,(u, e, t) + \varphi_u h\,(u, e, t)\,, 0 \leq \varphi_u \leq 1 \tag{2}$$

where P_{ue}^{t} signifies the expected consumption of user u to item e. For each user, using non-negative parameter φ_u, and balanced the functions h and g, which capture social influence and collaborative filtering for the predicting consumption. The balance parameters were customized and vary from user to user because users can have their own decisions by balancing these two aspects. For example, some users prefer to receive suggestions about their friends' consumption while others like to follow their own historical preferences to make future consumption. Particularly, the function g captures the predicted consumption score using user usage history so that all collaborative filtering models like latent factor models or object-based collaborative filtering, can be applied [25–27].

Specifically, $h\,(u, e, t)$ is a function that models the effect of social influence on the consumer's preference for the product e at time t. The effect of social influence indicates the dissemination of information on the social networks that lead people to consume local popular objects among their friends. This effect has been considered to be a foundation for many important social applications, such as maximizing social influence for viral marketing. With this effect of direct social influence, the future preference of the user on the product e is directly influenced by the historical consumer accounts of the decisions of his social neighbors of the same product:

$$h\,(u, e, t) = \frac{\sum_{t'=1}^{t-1} \sum_{d \in V_u^{t'}} Z_{de}^{t'} S_{ud}^{t} P^{t'}_{ue}}{\sum_{t'}^{t-1} \sum_{d \in V_u^{t'}} Z_{de}^{t'} S_{ud}^{t}}. \tag{3}$$

Where Z_{de}^{t} is by default, equal to 0, and it will equal to 1 if d consumes e at time t. Till t, u connects the set of users V_u^t. S_{ud}^{t} signifies the social influence of a pair of users depending on the structure of the social network. We assumed that the strength of social influence differs because the network structure dynamically changes. Here we referred to the widely used Adamic/Adar metric [28] to measure the proximity which varies with time as follows:

$$S_{ud}^{t} = \frac{1}{\sum_{b \in V_u^{t-1} \cap V_d^{t-1}} log\left(\left|V_b^{t-1}\right|\right)}. \tag{4}$$

3.3. Latent modeling for link prediction and user preference

Latent link prediction modeling: For the latent modeling of a new user, we consider the information is available u's friend set. The link $e\,\langle u, s\rangle$ will possibly be generated if the friends of u and s are similar. We call this relationship a latent relationship which is created through s and the friends of u. If u's friends have attributes that are similar to s then one latent attribute link is formed.

Let us consider that existing user u has three friends e_1, e_2, e_3 and e_1 shares their college and company attributes with s. This leads to two latent links among u and s. s also links to other friends of u, (e.g., e_2 and e_3) through various attributes. We also assumed that there may be many disconnections between the friends of u and s over the attributes. In order to model the latent relation betweens and u, we needed to quantify these latent links and the disconnections between the friends of u and s. Usually, if there are fewer disconnections and more latent links then there will be a higher likelihood of being friends between s and u. Consequently, we penalize the latent relationship if there is a disconnection and incentive the latent relationship if there is latent link between s and u. Based on this scheme, we computed the latent relationship between s and u by $b - \partial c$, where ∂ is a parameter for penalizing value, c and b are the number of disconnections and latent links . As a result, we calculated the latent relationship score as follows:

$$ls = \frac{1}{1 + e^{-\varphi(b - \partial c)}}. \tag{5}$$

Where φ is an exponential parameter. For determining whether $e\,\langle u, s\rangle$ will exist or not, the latent relationship score is important when most of the features of s and u cannot be obtained and s reveals only a few attributes.

Latent user preference modeling: For each item e and each user u, the consumption preference of users' behaviors at time t, latent representation among them can be expressed as follows:

$$cp\,(P|X, Y) = \prod_{t=1}^{T} \prod_{u=1}^{M} \prod_{e=1}^{N} \gamma\left\{P_{ue}^{t} | \left\langle X_u^t, Y_e \right\rangle, \varepsilon_P^2\right\}^{Z_{ue}^{t}}. \tag{6}$$

Where, variance ϵ^2 and mean ϑ, $\gamma\left(\vartheta, \epsilon^2\right)$ signify a normal distribution. By default, Z_{ue}^{t} is zero, which refers to an indicator variable. At time t, if user u consumes an item e, then Z_{ue}^{t} is equal to 1. With the latent matrix X^t at time t, $X_u^t \in R^{Qx1}$ refers to the latent consumption vector of u. In an item latent matrix $Y \in R^{NxQ}$, $Y_e \in R^{NxQ}$ represents the item latent vector. In the equation above, to model changes in user preference, we assumed that the latent matrix of users varies with time. Consequently, we can summarize the latent preferences of the users in a matrix sequence of latent consumption preference $X = \left\{X^1, \ldots, X^t, \ldots, X^T\right\}$. Prior to the latent variables, a typical approach is added to the limited observed preference data to avoid overfitting. On the item latent matrix, we added the zero-mean Gaussian as follows:

$$cp\left(Y|\epsilon_Y^2\right) = \prod_{e=1}^{N} \gamma\left(Y_e|0, \epsilon_Y^2 E\right). \tag{7}$$

With limited consumption data, our objective was to model the evolution of the matrix X of the latent consumption matrix of users. For each user, social influence and previous latent consumption preference influence u user's future latent interest, which is illustrated below. As we used the latent representation of consumer consumption interests, at time window $t = 2, 3, \ldots, T$, we modeled the two latent interest effects of each user as:

$$cp\left(X_u^t\right) = \gamma\left(X_u^t | \overline{X_u^t}, \epsilon_Y^2 E\right) \tag{8}$$

$$\overline{X_u^t} = (1 - \vartheta_u)\, X_u^{(t-1)} + \vartheta_u \sum_{\omega \in \gamma_u^{(t-1)}} \frac{S_{u\omega}^{t-1}}{N_u^{t-1}} X_\omega^{(t-1)},$$

$$\forall u \in X, 0 \leq \vartheta_u \leq 1. \tag{9}$$

Where, at time $t-1$, $S_{u\omega}^{t-1}$ indicates the influence of user ω to u, and $N_u^{t-1} = \sum_{\omega \in \gamma_u^{(t-1)}} S_{u\omega}^{t-1}$ represents a normalization constant on all the friends of u which guarantees that $\sum_{\omega \in \gamma_u^{(t-1)}} \frac{S_{u\omega}^{t-1}}{N_u^{t-1}} = 1$.

In a social space, the social influence score $S_{u\omega}^{t-1}$ represents the similarity of these two users. For each user, we describe a latent structure factor F_u^{t-1}, as it is likely to impose on users who have a similar latent structure factor with a higher influence resistance score. Thus, we defined the resistance influence score as follows:

$$S_{u\omega}^{t-1} = j\left(\left\langle F_u^{t-1}, F_\omega^{t-1}\right\rangle\right) = \frac{1}{1 + exp\left(-\left\langle F_u^{t-1}, F_\omega^{t-1}\right\rangle\right)}. \tag{10}$$

Where, to limit the influence score between 0 and 1, we used the logistic function $j(a) = \frac{1}{1+ex(-a)}$. Initially, the social network has not yet been set up at $t = 1$. As such, the latent consumption preferences of each user are determined only by his or her own consumption preferences without any social influence. Therefore, we considered a zero-mean Gaussian distribution of the latent vectors of the users. Thus, the latent consumption matrix is as follows:

$$cp\left(X|\epsilon_X^2, \epsilon_{X1}^2\right) = \prod_{u=1}^{X} \gamma\left(X_u^1|0, \epsilon_{X1}^2 E\right) \prod_{t=2}^{T} \gamma\left(X_u^t|\overline{X_u^t}, \epsilon_Y^2 E\right). \tag{11}$$

3.4. Multilevel deep belief network model

To address the link prediction and user consumption preference prediction for the new user in an OSN platform, we used the unsupervised learned Deep Belief Network (DBN). At the top of the DBN link prediction and user consumption preference, we added a linear output unit layer for class labels. A class label is represented by each output unit and the output unit with the highest value should be the sample's label. At the bottom of the DBN, when we input a sample vector ω, the output layer functions as a linear classifier and the value is as follows:

$$DL_i(\omega) = \sum_{\mathfrak{L}_j \in \mathbb{N}_{top}} \mathbb{Z}_{ji}\mathfrak{L}_j. \tag{12}$$

Where $\mathbb{N}_{top}$ is the activation of the hidden layers of the upper restricted Boltzmann machine in the DBN and $\mathbb{Z}$ represents the weights between the output layer and $\mathbb{N}_{top}$. We computed the $\mathbb{N}_{top}$ as follows:

$$p(\mathfrak{L}_i = 1|g) = \delta\left(\rho_i + \sum_j g_j \mathbb{Z}_{ji}\right) \tag{13}$$

Where, ρ_i is the bias of hidden unit i, $g\left(g_1, \ldots, g_j, \ldots\right)$ is the input to the visible layer, and $\delta(y) = \frac{1}{(1+e^{-y})}$. When we input a vector $\mathfrak{L}(\mathfrak{L}_1, \ldots, \mathfrak{L}_i, \ldots)$ in the hidden layer, the binary state g_j of each visible unit was set to 1 with probability by:

$$p\left(g_j = 1|\mathfrak{L}\right) = \delta\left(\mathfrak{B}_j + \sum_i \mathfrak{L}_i \mathbb{Z}_{ji}\right). \tag{14}$$

Where $\mathfrak{B}_j$ is the bias of visible unit j.

Then, at the input vector ω, the class label possibility is

$$p\left(label = DL_j|\omega\right) = \frac{e^{DL_j}}{\sum_i e^{DL_j}}. \tag{15}$$

By minimizing the entropy loss error as $-\sum_j \mathcal{F}_j log \mathfrak{l}_j$, this classifier is learned; where, $\mathcal{F}_j$ and $\mathfrak{l}_j$ are the class label and the value of predicted link. When we updated the weights of the output layer, the unsupervised DBN was set to achieve better classification performance.

4. Performance evaluation

In this section, we present the details of the data preparation, assessment metrics and evaluation of our methodology. We considered three different methods that we compared with our method based on different performance metrics.

4.1. Data representation and experimental setup

We selected two different real-world networks datasets, Facebook [29] and Amazon [30], to evaluate the performance of our method. Facebook's dataset consists of Facebook's "friends lists". Using Facebook app, Facebook to collect this data. The dataset was comprised of ego networks, profiles, and circle features. The data set included eight ego networks, and each set of ego network data consisted of friendships (non-directed links); IDs; and attributes, such as occupation, current city, education, gender, etc. On the other hand, the Amazon dataset consists of user comments on Amazon products. We used 6 months of data to perform our experiments. The dataset consisted of attributes like information about users, products, user preferences, and ratings.

We use Google+ dataset [31] to identify the relationship between each user profile of Facebook and Amazon. Using the 'share circle' feature, the data are collected from users who have shared their circles. The dataset consists of ego networks, circles and profile features. To extract the common Facebook–Amazon users from Google+ and other two datasets, we examine Google+ profiles to find out matched Facebook–Amazon users.

We learned our dataset and features using Naïve Bayes, C4.5, and the Support Vector-based Classifier (SVC) classifiers [32,33] to compare our method with other state-of-art methods. We extracted the features using the Python programming language and learned in Matlab. We carried out our experimental analysis on an Intel i7 3.40 Ghz system with 16 GB of RAM.

4.2. Evaluation metrics

To evaluate the performance of our proposed method, we used the AUC score, precision, recall and F-measure performance metrics.

AUC score: To assess the quality of the results of user consumption preferences and link prediction, we usually used the Area under Curve (AUC) score. The AUC score is used for quantifying the accuracy of a prediction method. We defined the AUC score for the link prediction and the preference of user consumption as:

$$AUC = \frac{(p + 0.5q)}{r}. \tag{16}$$

Where p is the number of times the existing links/consumption preferences have a higher score, q is the number of times they score the same, and r denotes the number of independent comparisons.

Precision: If n user pairs represent the existing edges in the T_H set of user pairs that have top-H maximum scores, then the precision of the result is defined as

$$p\text{-}measure = \frac{n}{H}. \tag{17}$$

Recall: In the network, if n existing edges is predicted by the method from $\mathbb{N}$ number of existing edges, then the recall of the result is defined as

$$r\text{-}measure = \frac{n}{\mathbb{N}}. \tag{18}$$

F-measure: The F-measurement can be defined based on the *p-measure* and *r-measure* of a link prediction/ consumption preference result as

$$F\text{-}measure = \frac{2x\,(p\text{-}measure)\,x\,(r\text{-}measure)}{(p\text{-}measure) + (r\text{-}measure)} \tag{19}$$

4.3. Link prediction performance

In this section, we provide the results of our two proposed models for the prediction of the link. In particular, we call the direct

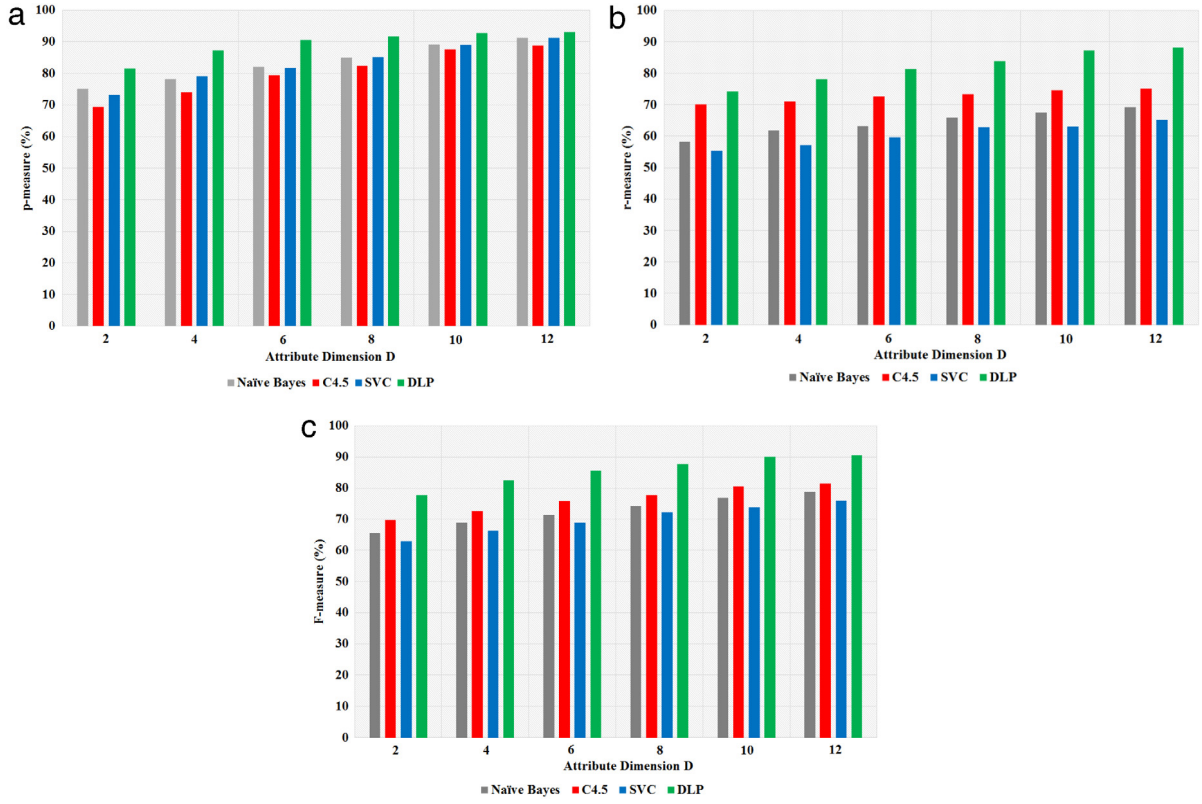


Fig. 1. Comparison results of direct link prediction model.

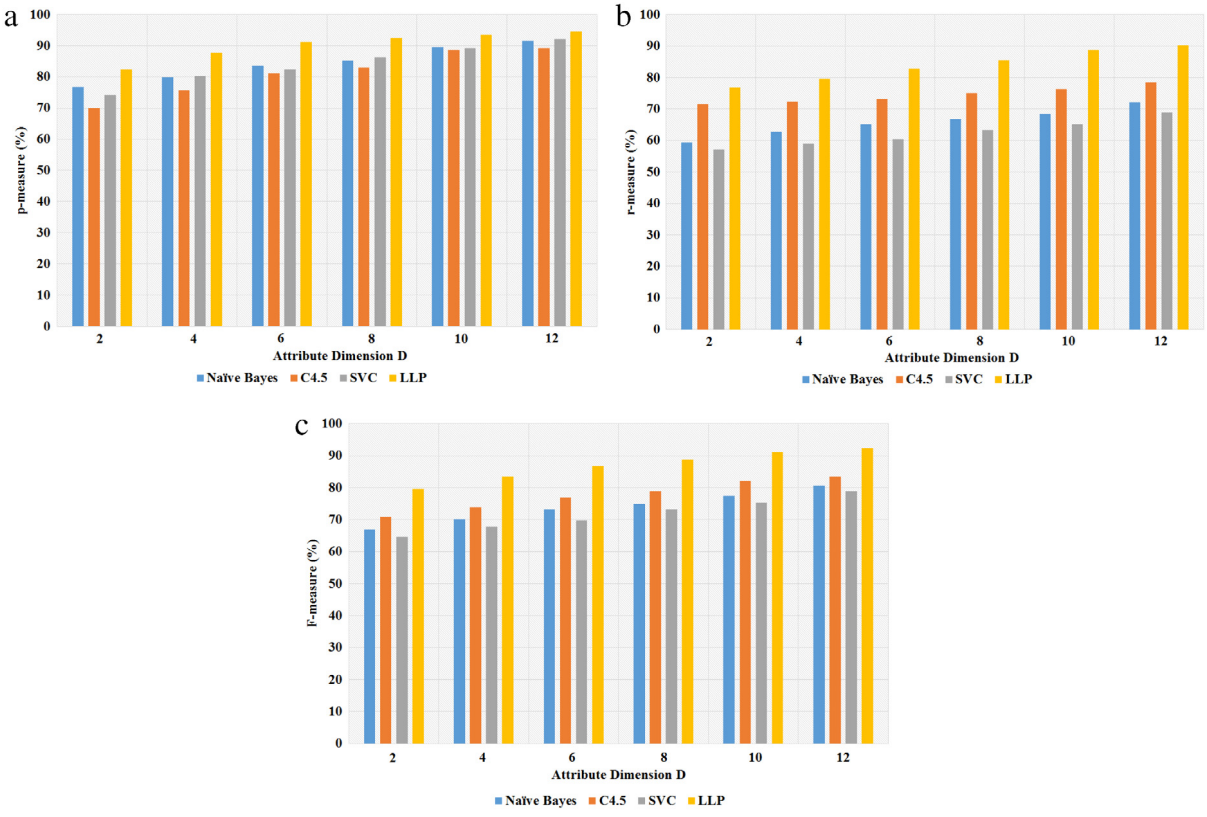


Fig. 2. Comparison results of latent link prediction model.

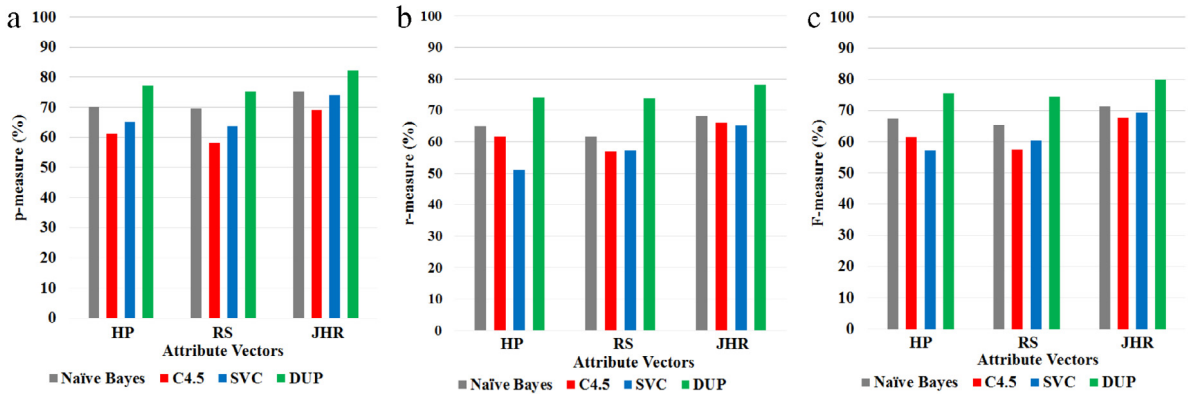


Fig. 3. Comparison results of direct user preference prediction model.

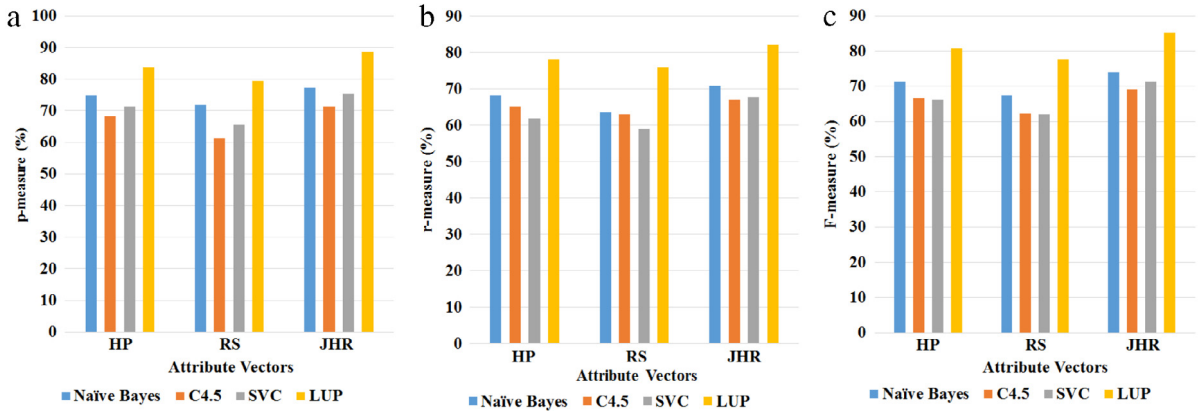


Fig. 4. Comparison results of latent user preference prediction model.

and latent prediction of the link introduced in Section 3 the Direct Link Prediction (DLP) and Latent Link Prediction (LLP) methods.

In the task of link prediction, our goal is to classify potential linked users. Due to the limitation of high performance resources, it is very difficult to consider all the users as friend candidates because the size of users is huge. So, we considered 200 negatively linked users who are not connected to it until the window of test time randomly. To select the high potential users linked to each test user, we combined users that linked positively with negatively coupled ones. We obtained the average results of all the performance metrics by repeating each process 20 times. Figs. 1 and 2 show the comparison metrics of these link prediction models on our datasets. The attribute dimensions were set to D = (2, 4, 6, 8, 10, 12). Based on the results obtained from the datasets, DLP and LLP behaved better than other methods. Their performance improvements were significant on all other models. On average, the improvement of the F1-measure of DLP and LLP on the best baseline was about 10%–15%. The results empirically validated that it is reasonable and efficient from an evolutionary point of view.

4.4. User preference prediction performance

To analyze the performance of our two proposed models for the prediction of user preferences, we report the experimental results. We named the direct user preference and latent user preference models as the DUP and LUP models.

Figs. 3 and 4 show the experimental results of different models on the datasets with the Historical Preferences (HP), Reception Suggestions (RS), and joint vector JHR (by considering both HP and RS) attribute vectors. We made several observations, which are as follows: First, DUP and LUP work better than other methods, which indicates the effectiveness of incorporating time and social networking information for predicting user preferences. However, DUP and LUP work well on this task on the HP attribute vector in comparison with the RS attribute vector. The results show that the user gives more preference to his or her own historical data compared to suggestions from friends on a product. Second in JHR, the proposed models offer superior performance over all others. This shows that our proposed model is effective in predicting consumer preference from an evolutionary point of view. On average, the improvement of the F1-measure of DUP and LUP on the best baseline was about 9–16%.

4.5. Joint prediction model performance

So far the evaluation results show that the level of accuracy of the direct and latent prediction methods for link prediction and consumer preference using multilevel DBN is superior to other methods. To increase the accuracy of prediction by using our proposed model, we joined both the direct and latent models of 'CLP' and 'CUP' for link prediction and user consumption preference, and verified the accuracy of the prediction. Figs. 5 and 6 show the results of joint model as well as other ones for link prediction and user consumption preferences. In link prediction, CLP achieved 94% accuracy for the F1-measure at the attribute dimension of 12 and an AUC score of 0.945. On the other hand, in predicting user consumption preference, CUP achieved 89% accuracy for the F1-measure at an attribute dimension of 12 and a 0.901 AUC score.

Given the results of predicting social ties and consumer consumption behaviors, we concluded predicting social network links

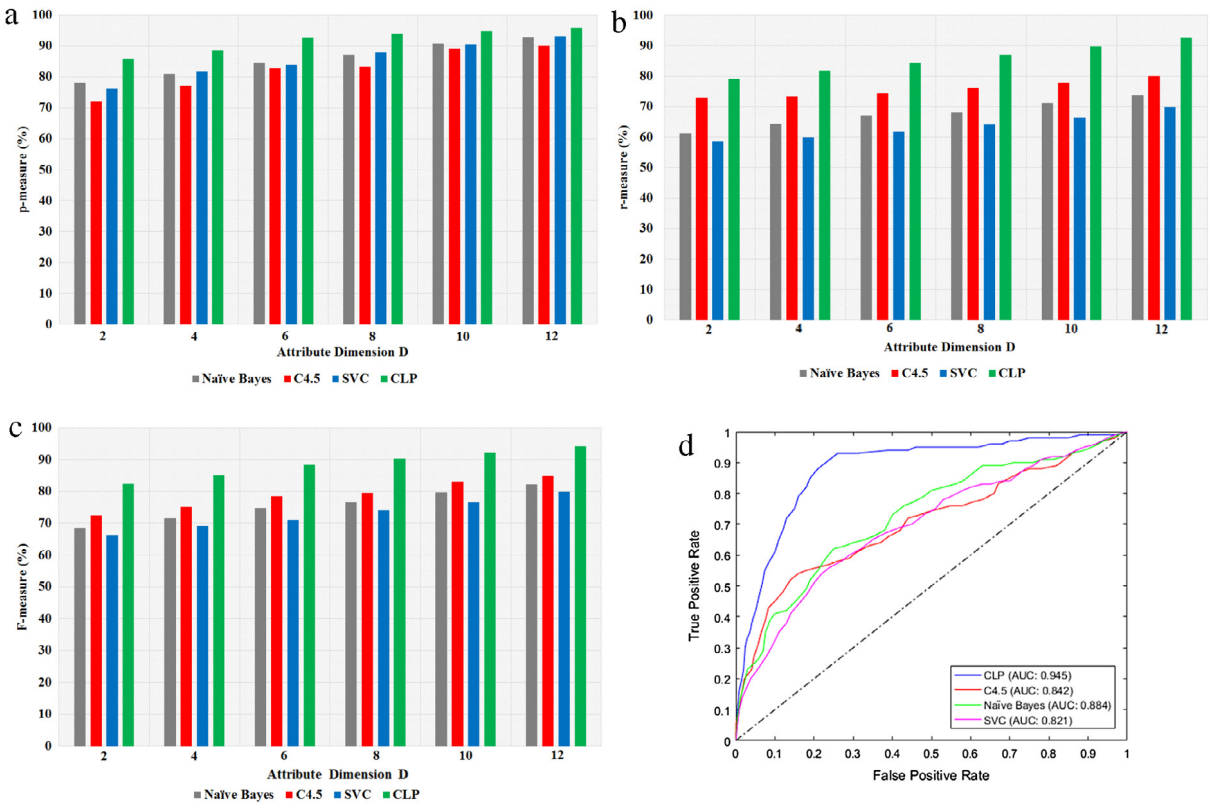


Fig. 5. Comparison results of joint link prediction model.

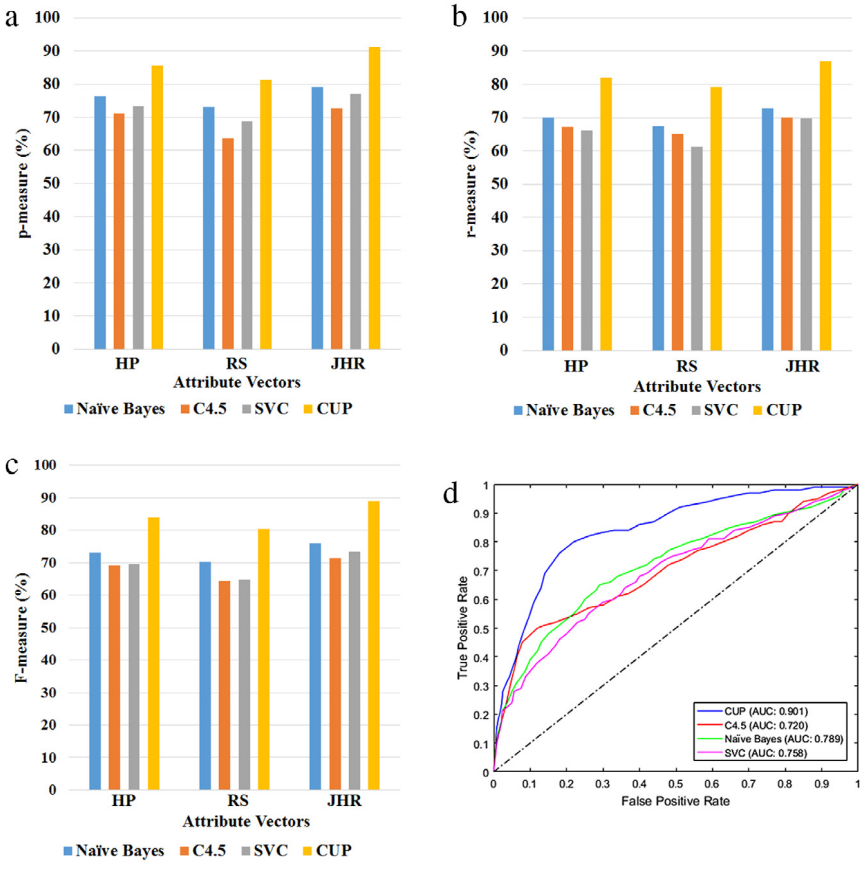


Fig. 6. Comparison results of joint user preference prediction model.

and consumer consumption behaviors are mutually beneficial. From an evolving perspective, jointly modeling them would benefit both tasks. From among our proposed models, CLP and CUP always provided the best performance. Compared with other methods, we concluded that our proposed model is more effective and accurate for modeling the evolution of users' consumption preferences and link prediction behaviors. The final results of our evaluation empirically demonstrate that our method has better predictive power.

5. Conclusions

In this paper, we have proposed novel models for the prediction of links and users' consumption preferences for new users in OSNs. We proposed direct and latent representations that directly presume the users' behaviors, which are represented by their historical behaviors and latent observable behaviors, in order to illustrate the evolution of user behaviors in OSN platforms. To achieve high accuracy, we also proposed a multilevel deep belief network learning approach. We evaluated and compared our proposed model with Naïve Bayes, C4.5, and SVC methods using real world datasets. For link prediction, CLP (i.e., a joint direct and latent link prediction model) achieved 94% accuracy for the F1-measure at the attribute dimension of 12 and an AUC score of 0.945. CUP (i.e., joint direct and latent user consumption preference model) achieved 89% accuracy for the F1-measure at an attribute dimension of 12 and a 0.901 AUC score for predicting user consumption preferences. Our proposed model achieved about 10–15% greater improvements over other methods.

For future work, we intend to try some other approaches for multilevel deep learning and to try and find some other feature that implies the values of user behaviors.

Acknowledgements

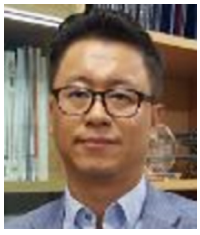
This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2013-0-00684) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

[1] D. Lee, Personalizing information using users' online social networks: A case study of citeulike, J. Inf. Process. Syst. 11 (1) (2015) 1–21.
[2] A. Corbellini, D. Godoy, C. Mateos, S. Schiaffino, A. Zunino, DPM: A novel distributed large-scale social graph processing framework for link prediction algorithms, Future Gener. Comput. Syst. (2017). http://dx.doi.org/10.1016/j.future.2017.02.025.
[3] S. Aral, L. Muchnik, A. Sundararajan, Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks, Proc. Natl. Acad. Sci. 106 (51) (2009) 21544–21549.
[4] M.W. Ahn, W.S. Jung, Accuracy test for link prediction in terms of similarity index: The case of WS and BA models, Physica A 429 (2015) 177–183.
[5] M. Hoffman, D. Steinley, M.J. Brusco, A note on using the adjusted rand index for link prediction in networks, Social Networks 42 (2015) 72–79.
[6] L.M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, F. Menczer, Friendship prediction and homophily in social media, ACM Trans. Web (TWEB) 6 (2) (2012) 1–33.
[7] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (6) (2011) 1150–1170.
[8] M. Pujari, R. Kanawati, Supervised rank aggregation approach for link prediction in complex networks, in: Proceedings of the 21st International Conference on World Wide Web, ACM, 2012, pp. 1189–1196.
[9] D.Q. Vu, D. Hunter, P. Smyth, A.U. Asuncion, Continuous-time regression models for longitudinal networks, in: Advances in Neural Information Processing Systems, 2011, pp. 2492–2500.
[10] Z. Zeng, K.J. Chen, S. Zhang, H. Zhang, A link prediction approach using semi-supervised learning in dynamic networks, in: Advanced Computational Intelligence, ICACI, 2013 Sixth International Conference on, IEEE, 2013, pp. 276–280.
[11] Y.L. He, J.N. Liu, Y.X. Hu, X.Z. Wang, OWA operator based link prediction ensemble for social network, Expert Syst. Appl. 42 (1) (2015) 21–50.
[12] E. Sherkat, M. Rahgozar, M. Asadpour, Structural link prediction based on ant colony approach in social networks, Physica A 419 (1) (2015) 80–94.
[13] J.Y. Ding, L.C. Jiao, J.S. Wu, Y.T. Hou, Y.T. Qi, Prediction of missing links based on multi-resolution community division, Physica A 417 (1) (2015) 76–85.
[14] B.L. Chen, L. Chen, B. Li, A fast algorithm for predicting links to nodes of interest, Inform. Sci. 329 (2016) 552–567.
[15] Y. Yang, N. Chawla, Y. Sun, J. Hani, Predicting links in multi-relational and heterogeneous networks, in: Data Mining (ICDM), 2012 IEEE 12th International Conference on, IEEE, 2012, pp. 755–764.
[16] V. Stroele, G. Zimbrao, J.M. Souza, Group and link analysis of multi-relational scientific social networks, J. Syst. Softw. 86 (2013) 1819–1830.
[17] N. Barbieri, F. Bonchi, G. Manco, Who to follow and why: link prediction with explanations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014, pp. 1266–1275.
[18] F.Y. Hu, H.S. Wong, Labeling of human motion based on CBGA and probabilistic model, Int. J. Smart Sens. Intell. Syst. 6 (2) (2013) 583–609.
[19] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, S. Yang, Social contextual recommendation, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM, 2012, pp. 45–54.
[20] M. Jiang, P. Cui, F. Wang, W. Zhu, S. Yang, Scalable recommendation with social contextual information, IEEE Trans. Knowl. Data Eng. 26 (11) (2014) 2789–2802.
[21] J. Tang, H. Gao, X. Hu, H. Liu, Exploiting homophily effect for trust prediction, in: Proceedings of the sixth ACM International Conference on Web Search and Data Mining, ACM, 2013, pp. 53–62.
[22] N.Z. Gong, A. Talwalkar, L. Mackey, L. Huang, E.C.R. Shin, E. Stefanov, et al., Joint link prediction and attribute inference using a social-attribute network, ACM Trans. Intell. Syst. Technol. 5 (2) (2014) 1–20.
[23] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[24] M. Längkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recognit. Lett. 42 (2014) 11–24.
[25] R. Salakhutdinov, A. Mnih, Probabilistic matrix factorization, Nips 1 (1) (2007) 1–8.
[26] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: Proceedings of the 10th International Conference on World Wide Web, ACM, 2001, pp. 285–295.
[27] W. Li, X. Li, M. Yao, J. Jiang, Q. Jin, Personalized fitting recommendation based on support vector regression, Human-Centric Comput. Inf. Sci. 5 (1) (2015) 1–11.
[28] D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks, J. Assoc. Inf. Sci. Technol. 58 (7) (2007) 1019–1031.
[29] J. Leskovec, Social circles: Facebook. https://snap.stanford.edu/data/egonets-Facebook.html. (Accessed 24 April 2017).
[30] J. Leskovec, Web data: Amazon reviews. https://snap.stanford.edu/data/web-Amazon.html. (Accessed 24 April 2017).
[31] J. Leskovec, Social circles: Google+. https://snap.stanford.edu/data/egonets-Gplus.html. (Accessed 24 April 2017).
[32] P. Wang, B. Xu, Y. Wu, X. Zhou, Link prediction in social networks: the state-of-the-art, Sci. China Inf. Sci. 58 (1) (2015) 1–38.
[33] A. Malhotra, L. Totti, W. Meira Jr., P. Kumaraguru, V. Almeida, Studying user footprints in different online social networks, in: Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on, IEEE, 2012, pp. 1065–1070.

Mr. Pradip Kumar Sharma He is a Ph.D. scholar at the Seoul National University of Science and Technology. He works in the Ubiquitous Computing & Security Research Group under the supervision of Prof. Jong Hyuk Park. Prior to beginning the Ph.D. program, he worked as a software engineer at MAQ Software, India. He worked on a variety of projects, proficient in building large-scale complex data warehouses, OLAP models and reporting solutions that meet business objectives and align IT with business. He received his dual Masters degree in Computer Science from the Thapar University, in 2014 and the Tezpur University, in 2012, India. His current research interests are focused on the areas of ubiquitous computing and security, cloud computing, SDN, SNS, and IoT. He is also reviewer of Journal of Supercomputing (JoS).

Mr. Shailendra Rathore He is a Ph.D. student in the Department of Computer Science at Seoul National University of Science and Technology (SeoulTech.), Seoul, South Korea. Currently, he is working in Ubiquitous Computing Security (UCS) Lab under the supervision of Prof. Jong Hyuk Park. His broadly research interest includes Information and Cyber Security, SNS, Digital Forensic, IoT. Previous to joining Ph.D. at SeoulTech, he has worked as an Executive-Technology at Crompton Greaves Global R & D, Mumbai, India from June, 2013 to July, 2014. He received his M.E. in Information Security from Thapar University, Patiala, India and B.Tech. in Computer Engineering from Rajasthan Technical University, Kota, Rajasthan, India.



Dr. James J. (Jong Hyuk) Park received Ph.D. degrees in Graduate School of Information Security from Korea University, Korea and Graduate School of Human Sciences from Waseda University, Japan. From December, 2002 to July, 2007, Dr. Park had been a research scientist of R&D Institute, Hanwha S&C Co., Ltd., Korea. From September, 2007 to August, 2009, He had been a professor at the Department of Computer Science and Engineering, Kyungnam University, Korea. He is now a professor at the Department of Computer Science and Engineering and Department of Interdisciplinary Bio IT Materials, Seoul National University of Science and Technology (SeoulTech), Korea. Dr. Park has published about 200 research papers in international journals and conferences. He has been serving as chairs, program committee, or organizing committee chair for many international conferences and workshops. He is a founding steering chair of some international conferences–MUE, FutureTech, CSA, UCAWSN, etc. He is editor-in-chief of Human-centric Computing and Information Sciences (HCIS) by Springer, The Journal of Information Processing Systems (JIPS) by KIPS, and Journal of Convergence (JoC) by KIPS CSWRG. He is Associate Editor / Editor of 14 international journals including 8 journals indexed by SCI(E). In addition, he has been serving as a Guest Editor for international journals by some publishers: Springer, Elsevier, John Wiley, Oxford Univ. press, Hindawi, Emerald, Inderscience. His research interests include security and digital forensics, Human-centric ubiquitous computing, context awareness, multimedia services, etc. He got the best paper awards from ISA-08 and ITCS-11 conferences and the outstanding leadership awards from IEEE HPCC-09, ICA3PP-10, IEE ISPA-11, and PDCAT-11. Furthermore, he got the outstanding research awards from the SeoulTech, 2014. Dr. Park' s research interests include Human-centric Ubiquitous Computing, Vehicular Cloud Computing, Information Security, Digital Forensics, Secure Communications, Multimedia Computing, etc. He is a member of the IEEE, IEEE Computer Society, KIPS, and KMMS.