IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 10, 2019, accepted June 23, 2019, date of publication July 5, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927026

Search Personalization in Folksonomy by Exploiting Multiple and Temporal Aspects of User Profiles

KEEJUN HAN[1], MUN Y. YI[1], AND JUNGEUN KIM[2]

[1]Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea
[2]Intelligence Information Research Division, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

Corresponding author: Jungeun Kim (je_kim@etri.re.kr)

This work was supported by the Korea Government Institute for Information Communications Technology Promotion (IITP) through the development of an advanced open data distribution platform based on international standards under Grant MSIT 2017-0-00253.

ABSTRACT Social tagging data, also known as folksonomy, are a valuable indication for the user's understanding of a resource. The nature of folksonomy data in which a user annotates a resource with their opinions provides immense potential to contribute to search personalization. The challenge lies in extracting interests from the folksonomy data and building accurate user profiles while maintaining their characteristics. Furthermore, the current state-of-the-art technologies that utilize folksonomy for search personalization have not fully exploited both multiple and temporal aspects in user profiles. In this paper, we propose a search personalization framework that constructs a user profile network with identification of the *multiple topics* of the user and the *temporal values* of tags. Then, the user profile network is further explored through a *link analysis technique for the network* to score the tags by their importance. The performance of the proposed framework is evaluated against various state-of-the-art folksonomy-based personalization models and it consistently outperforms all of the compared models under the conditions of the best combination of ranking functions and link analysis techniques.

INDEX TERMS Search personalization, folksonomy, user profile, multiple topics, temporality, link analysis.

I. INTRODUCTION

Nowadays, tags are commonly used by users in various services such as Instagram [1] and Facebook [2]. Compared to the early stage of social tagging services, users have recently become more aware of tags and voluntarily annotate their contents with the tags in order to represent their topics (*i.e.*, interests) in summarized forms. Therefore, the rise of social tagging data, also known as folksonomy, has been noted again as an effective means of acquiring users' topics and improving results of search personalization [3]–[7]. The improved performances result from the characteristics of folksonomy data that not only provide a direct connection between resources and users but also reveal topics of the users in more explicit ways [8].

Prior works that sought to build user profiles using folksonomy data mainly focused on suggesting methods to adapt existing IR models such as the Vector Space Model (VSM) [9] and the unigram language model [10]. Despite potential advantages of applying the classical bag-of-words models in IR to folksonomy for personalization, there is room for further improvement by incorporating important features of user data: users possess more than one topic, and the topics are drifting, not fixed [11].

In fact, the features are not mutually exclusive but often occur simultaneously. Let us suppose that there is a user who has been a fan of action movies for many years. Lately, his environment caused him to also be interested in family movies as well as action movies. This example implies that even though the number of recent tags about family movies is relatively small, they should be more weighted than the old ones so that the user profile can stay relevant to his current topics. At the same time, the profile is still required to store the old topic, action movies, as it is still one of his main interests. This also addresses another aspect of a user profile: we need to keep multiple topics of users to reflect various information needs. Prior works partially have considered multiple [12], [13] and temporal

The associate editor coordinating the review of this manuscript and approving it for publication was Yin Zhang.

aspects [11], [14], [15] of user topics. However, to the best of our knowledge, no researches have attempted to fully exploit both aspects at once, nor did they extend this concept for search personalization in folksonomy.

To attain our goal, instead of applying traditional ways in folksonomy of building the user profile, we represent the user profile where vertices correspond to tags that the user has annotated and edges are defined as co-occurrences of tags. That is, if a user writes a tag in any resource, that tag becomes a vertex in the network. Multiple tags (i.e., vertices) written on the same resource will be connected via edges in the network. We call this a user profile network throughout the rest of the paper. This user profile representation allows us to solve the aforementioned issues, *i.e.*, multiplicity and temporality features of user topics. Specifically, it enables us to (1) discover multiple topics intuitively through a community detection method [16]–[18], (2) reflect the temporal aspects of user topics in the process of building the user profile network, and finally (3) learn the importance of tags for users from the network by various link analysis techniques [19]–[21]. Furthermore, it can be used to solve the ambiguity problem in folksonomy since it enables us to figure out which context the tag was used by analyzing the tag connectivity. Lastly, based on the importance scores of the tags for the users from the network analysis, we conduct a personalized search by calculating the relevance between the user and resources for given queries. We describe the process of the search personalization in Section IV. Note that the proposed framework can be used for any folksonomy datasets with tuples <user, tag, time>.

To verify the effectiveness of our framework, we proceed thorough evaluations on three actual folksonomy datasets: CiteULike, MovieLens, and Delicious. For the best experimental setting, our results surpass all of the state-of-the-art folksonomy-based search personalization models and a variant of the proposed model in terms of MRR, Personalization Gain (P-Gain), and Precisions at top k documents (Pr@K).

In summary, our contributions in this paper can be summarized as follows:

- We build a user profile that is aware of multiple topics and their drifting by constructing and analyzing folksonomy-based user profile networks through community detection and edge weighting techniques.
- We further investigate several link analysis techniques for the user profile network to discover which of the techniques perform best in our search personalization to measure the importance of tags for the user.
- Based on the proposed user profile, we present a new framework for search personalization that achieves higher effectiveness compared to state-of-the-art folksonomy-based search personalization models.

The remainder of this paper is structured as follows: Section II reviews related studies and Section III describes the procedure of building user profiles for calculating the importance scores of tags for each user. Section IV shows a search personalization framework based upon the proposed user profiles. Section V demonstrates the effectiveness of the proposed framework through experiments and presents the results. Lastly, Section VI concludes the study and suggests future research directions.

II. RELATED WORKS

We review a number of relevant studies on building temporal and multiple topic-aware user profiles, and search personalization in folksonomy. In Sub-section II-B, we also introduce state-of-the-art folksonomy-based search personalization models that are used for comparison in our experiment.

A. TEMPORAL AND MULTIPLE TOPICS-AWARE USER PROFILES

Regarding discovering multiple topics of users, one of the most common approaches is to use a human-generated ontology such as the Open Directory Project [11]. However, this approach may have poor coverage, and a low level of granularity, and may lack novel terms. Moreover, manually-created ontologies demand time-consuming effort [22]. As an alternative, recent approaches [15], [22] have researched learning multiple topics by means of topic modeling, such as LDA [23], [24]. Despite the promising performance of LDA in personalization, an exact inference process of topic models is intractable. User profile can be constructed by exploiting user's own activities on their social networks as well as the folksonomy activities [25]. Meanwhile, in our framework, the process of determining communities to which a certain tag belongs is easily tractable by visualizing user profile networks. Furthermore, LDA was designed to perform well on a large text corpus; however, folksonomy only holds a limited number of the words to be analyzed.

Regarding the temporal aspect of user topics, there was an attempt to model on short and long term topics of users by analyzing query log data [11]. An adaptation of LDA has been also devised to represent short- and long-term user topics [15]. In [26], temporal aspect is captured by applying a variant of forgetting function, and the aspect is aggregated with other aspects through matrix factorization. It is worth noting that one of the most similar work to ours for the temporal feature is a technique proposed in [14] since it also controls edge weights after they build a folksonomy-based network. In their work, however, the consideration for multiple topics of users is overlooked, and calculating the initial weights is rather a simple summation of co-occurrence, which is too naive to reflect the topic drifting of user behaviors up to a certain accuracy. Also, the work only focuses on a user study of building user profiles, and it was not extended for search personalization.

In this paper, in contrast to [11], [14], we exploit a community detection technique to detect multiple topics of a user profile by representing the folksonomy as a set of vertices and edges rather than a large text corpus. Furthermore, unlike [15], [22], as discovering temporal topics of users by adapting LDA, we propose an edge weighting technique to differentiate the importance for each vertex while conducting link analysis for the network.

B. SEARCH PERSONALIZATION IN FOLKSONOMIES

In general, the procedure of search personalization is to (1) construct a user profile that consists of a pair of terms and its values to represent the topics of the user, (2) build a resource profile with a pair of terms and its values to represent the topics of the resource, and (3) re-rank the user relevance resources to be highly located in the final ranking by computing the similarity between the resources and the user profile given the initial ranking by a user query. To achieve improved personalization results, it is important to construct precise user profiles that can represent user topics and to design more sophisticated ranking functions that measure relevance between users and resources.

The available user profiles in folksonomy are generally based on models such as the Vector Space Model (VSM) or BM25 model. In [9], they considered the tag frequency (TF), and inversed user frequency (IUF) which are an adaptation of the TF-IDF model. Similarly, another model, proposed in [10], is a hybrid of the TF-IDF and BM25 models. The work proposed in [27] is a simple yet effective personalization algorithm that exploits tags for resources based on the term frequency. However, they only utilize the values of the user tag frequency and set all tag frequencies from resources as 1 to empower the user profile when measuring similarity.

Recent studies [9], [28], [29] propose a Normalized Term Frequency (NTF) model that both emphasizes the importance of tag frequency and minimizes biases on active users by adopting proportions of user tags. The philosophy for this model is that if a user writes a certain tag more frequently, it denotes that the user is more interested in that tag.

There have been other approaches that aggregate ratings and tags to further improve the performance of search personalization [30], [31]. In [30], rating information including not only 'like', but also 'dislike' is used to weight the importance of tags. Similarly, in [31], the weight of extreme-rating such as 1 or 5 (out of 5) shared by similar users is strengthened during the ranking algorithm to highlight the user's explicit characteristics for personalization. Despite its potential to improve personalization, the rating information should be given to implement the approaches.

In this paper, instead of conventional approaches, we adopt diverse analysis techniques to consider the network topology in folksonomy and utilize the latent semantics among the tags in the network. We also evaluate the proposed framework against other state-of-the-art models presented in the aforementioned studies.

III. BUILDING USER PROFILES

User profile network of user u is defined as an undirected weighted network $G_u = (V, E)$, where V is a set of tags that the user has annotated on resources and $E \in V \times V$ is a set of co-occurrence relationships between the tags. Let $w_{i,j}$, be a weight of edge between v_i and v_j ($v_i \neq v_j$) where the weight is obtained by counting the number of co-occurrences between tags. Multiple topics of the user are discovered through community detection. Then, multiple topics and temporal information for tags are exploited for updating a user profile network, G_u. Let s_i be a set of importance scores obtained from link analysis associated with the tags t_i in the network G_u. Final user profile of the user u is denoted by P_u, which is a vector of tag t_i: updated importance score s_i' pairs. In the subsequent sub-sections, we describe each step of building user profiles proposed in this work.

A. DISCOVERING MULTIPLE TOPICS

In this step, we build a primitive user profile network where the weights of edges are yet to apply neither temporality nor multiplicity aspects of topics in user profile. We explain how to: (1) build the initial folksonomy-based user profile network and (2) detect multiple topics from the network by exploiting a community detection technique.

1) CONSTRUCTING PRIMITIVE USER PROFILE NETWORK

The user's history of tag annotation for resources is stored in folksonomy data and it is used to build a primitive tag-based network. Note that the network is built with tags given from a user u, rather than global tag network from all users. This is due to the fact that building separated sub networks for every user is a better estimation for user preference [22]. In our initial user profile network, vertices and edges denote tags and co-occurrence relations between two tags, respectively. In other words, two vertices are connected if user used two tags to annotate one resource. This co-occurrence definition is commonly used in prior works [12]–[14], [32], [33] to build networks because of its simplicity and computability.

2) DETECTING MULTIPLE TOPICS

To identify multiple topics of a user, we conduct community detection on the primitive user profile network. In our framework, any community detection techniques can be applied as long as they can support an undirected and a weighted network format. We adopt a community detection technique called modularity optimization which enables to find high modularity partitions of a network [18]. It is also known as a fast and robust method in detecting communities. This method calculates the gain in modularity ΔQ by putting an isolated vertex v_i into a community C as follows:

$$\Delta Q = \left[\frac{\sum_{in} w_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} w_{incident} + k_i}{2m}\right)^2\right] - \left[\frac{\sum_{in} w_{in}}{2m} - \left(\frac{\sum_{tot} w_{incident}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right] \quad (1)$$

where $\sum_{in} w_{in}$ is the sum of the weights of the links inside C, $\sum_{tot} w_{incident}$ is the sum of the weights of the links incident to vertices in C, k_i is the sum of the weights of the links incident to vertex v_i, $k_{i,in}$ is the sum of the weights of the links from v_i to vertices in C and m is the sum of the weights of all the links in the network.

After the community detection, all vertices in the network belong to specific communities and a set of vertices within the same community represents a certain topic [12]. A topic that a vertex v_i belongs to is labeled by c_i. Let c_i be an integer value which indicates the topic label where the vertex v_i belongs. The topics labels, c_i, are exploited for edge weighting process described below.

B. EDGE WEIGHTING

In the proposed user profile network, vertices indicate tags that user u chose to annotate resources. Regarding to edges, temporal information is applied to weight edges in a network. This is executed by implementing a decay function: Each time a new resource is inserted, the weight of each edge in the network decreases by a small percentage of its current value unless the edge exists in the new resource. Specifically, our earlier work [33] observed a noteworthy characteristic of the co-occurrence relationships: When two vertices are used together frequently, the relevance between two vertices significantly increases, compared to used together few times. Based on this observation, we made the following assumption:

1) ASSUMPTION 1

For two tags t_i and t_j, if the two tags are used together frequently, increment of relationship of the tags t_i and t_j is greater than linear growth.

Furthermore, the following assumption reflects multiplicity information that obtained during community detection into user profiles:

2) ASSUMPTION 2

For two tags t_i and t_j, if their topic labels, c_i and c_j, are identical, the relationship of the tags t_i and t_j becomes strengthened.

The sub-steps of edge weighting process with temporality and multiplicity of topics are given as follows:

- Based on annotated time information of each tag, tags are listed in chronological order. A list of the resources where user attached a tag is also sorted in the same temporal order as they were added. We denote this list of resources as L.
- Starting with the first resource in the list L, the m number of tags: $t_1, t_2, \ldots, t_m$ that used to annotate the resource is added to a network. For every existing tag t_i, create a corresponding vertex v_i, if v_i dose not exist.
- If an edge between v_i and v_j does not yet exist, we add an edge with a co-occurrence weight of 1.
- If an edge between v_i and v_j does already exist, we update the edge weight w_{ij} between the two vertices as follows:

$$w_{ij} = \begin{cases} (1+\alpha)\log(exp(w_{ij})) & \text{if } c_i = c_j \\ \log(exp(w_{ij})) & \text{otherwise;} \end{cases} \tag{2}$$

  where c_i and c_j ($i \neq j$) indicate the topic labels of v_i and v_j, respectively, and $exp(w_{ij})$ reflects our Assumption 1. α is a control parameter to reflect the Assumption 2. Note that α is $a \in [0, 1]$, and is fixed at 0 if c_i and c_j are not identical. In Sub-section V-E, $\alpha = 0.8$ is proven to be the best choice, and thus, this is used as a default value for our experiments.
- Before moving to the next resource in the list L, the edges in the network reduce as follows:

$$w_{ij} = w_{ij} - \beta \cdot w_{ij} \tag{3}$$

  where $\beta \in [0, 1]$ is a constant decay factor to weight the impact of temporal aspect. As β becomes smaller, weights of old tags diminish. In Sub-section V-E, $\beta = 0.4$ proved to be the best performance, and thus this is used as a default value for our experiments.
- Lastly, once all of the resources annotated by the user are visited, edge weights in the user profile network are normalized as follows:

$$\widehat{w}_{ij} = \frac{w_{ij}}{max(w_{ij})} \tag{4}$$

  where $max(w_{ij})$ indicates maximum w_{ij} in the network.

C. CALCULATING IMPORTANCE SCORES OF TAGS

In this step, we describe calculating importance scores for tags in a user profile network by exploiting five link analysis techniques: PageRank, HITS, Closeness, Eccentricity, and Betweenness.

It is important to note that edges in the proposed user profile network is undirected and weighted. Thus, we adapted PageRank to exploit the weight of the edge, $\widehat{w}_{ij}$, between two vertices v_i and v_j, on undirected network as follows:

$$PR(v_i) = (1-d) + d \cdot \sum_{v_j \in B_i} \frac{\widehat{w}_{ij}}{\sum_{v_k \in L_j} \widehat{w}_{ij}} PR(v_j) \tag{5}$$

where B_i is the set of vertices that point to v_i and L_j is the set of vertices that vertex v_j points to. d is a damping factor to implement the random suffer model. In this paper, d is set to 0.85 as suggested from [34].

Another iterative algorithm, called HITS, measures importance for vertices in the network by computing two separate values for each vertex: Authority and Hub. Authority update rule is defined as:

$$auth(v_i) = \sum_{j=1}^{n} hub(v_j) \tag{6}$$

where n is the number of vertices linked to v_i and v_j is a vertex linked to v_i.

Hub update rule is defined as:

$$hub(v_i) = \sum_{j=1}^{n} auth(v_j) \tag{7}$$

where n is the number of vertices v_i linked to and v_j is a vertex which v_i is linked to.

For centrality-based link analysis techniques, we first define the length of the shortest distance (i.e., $d_G(v_i, v_j)$)

between v_i and v_j which incorporates edge weights in the network, as follows:

$$d_G(v_i, v_j) = \min\left(\frac{1}{(\widehat{w}_{ih})^{\lambda}} + \cdots + \frac{1}{(\widehat{w}_{hj})^{\lambda}}\right) \tag{8}$$

where λ is set to be 1 herein to have the same effect as Dijkestra's algorithm [35]. Based on the definition, the closeness centrality represents the average distance from a given vertex to all other vertices in the network and is defined as:

$$C_C(v_i) = \frac{1}{\sum_{j\in V} d_G(v_i, v_j)} \tag{9}$$

The eccentricity centrality represents the distance from a given vertex to the farthest vertex in the network and is defined as:

$$C_E(v_i) = \frac{1}{max_{j\in V} d_G(v_i, v_j)} \tag{10}$$

The betweenness centrality measures how often a vertex appears on the shortest paths between vertices in the network and is defined as:

$$C_B(v_i) = \sum_{v_i\neq v_j\neq v_k\in V} \frac{\sigma_{v_j,v_k}(v_i)}{\sigma_{v_j,v_k}} \tag{11}$$

where σ_{v_j,v_k} is the total number of shortest distances from vertice v_j to vertice v_k and $\sigma_{v_j,v_k}(v_i)$ is the number of those distances that pass through v_i.

In order to further differentiate scores of vertices in user profile network, the scores s_i obtained from the five link analysis techniques are finally updated to s_i' by multiplying normalized tag frequency as follows:

$$s_i' = s_i \cdot \frac{N_i}{N} \tag{12}$$

where N is the total number of resources tagged by a target user, N_i is the number of times a user u uses a tag t_i to annotate a resource.

Finally, a user profile of the target user u is created based on the definition as follows:

Definition 1: A user profile of user u, represented by $\boldsymbol{P_u}$, is a vector in the form of tag:value pairs as follows:

$$\boldsymbol{P_u} = (t_1 : s_1', t_2 : s_2', \ldots, t_n : s_n') \tag{13}$$

where t_i ($i = 1, \ldots, n$) is a tag that exists in user profile network G_u, and n is the total number of tags in G_u.

IV. SEARCH PERSONALIZATION

In order to rank the resources for search personalization, the user profile, P_u, is used to measure the relevance score to resource profile, P_r, given a query q. Similar with the user profiles, a resource profile P_r of resource r is created based on the definition as follows:

Definition 2: A resource profile for the resource r, represented by $\boldsymbol{P_r}$, is a vector in the form of tag:value pairs as follows:

$$\boldsymbol{P_r} = (t_1 : p_1, t_2 : p_2, \ldots, t_n : p_n) \tag{14}$$

where t_i ($i = 1, \ldots, n$) is a tag being used to describe resource r, n is the total number of tags for r, and p_i indicates frequency in which resource r possesses the tag t_i. For p_i, we use the normalized tag frequency to measure the relevance degree of t_i to resource r, implying that the higher the value of p_i, the more salient tag t_i for the resource r.

$$p_i = \frac{M_i}{M} \tag{15}$$

where M is the total number of users who annotate the resource r and M_i is the number of users using the tag t_i to annotate the resource r.

The query q is also represented as a vector based on the following definition:

Definition 3: A query profile for query q, denoted by $\boldsymbol{P_q}$, is a vector of terms as follows:

$$\boldsymbol{P_q} = (t_1^q, t_2^q, \ldots, t_m^q) \tag{16}$$

where t_i^q ($i = 1, \ldots, m$) is a query term and m is the total number of terms in the query.

In recent studies on folksonomy based personalization [9], [10], [28], several ranking functions were proposed to measure relevance. In this paper, we chose three state-of-the-art ranking functions to estimate how robust our proposed framework is compared to the existing ranking functions. The ranking functions are: cosine function [9], scalar ranking function [10], and fuzzy ranking function [28].

Based on the VSM model, the cosine similarity measure was modified to compute the ranking score as follows:

$$score(r, q, u) = sim(\boldsymbol{P_r}, \boldsymbol{P_u}) \cdot sim(\boldsymbol{P_r}, \boldsymbol{P_q}) \tag{17}$$

where *sim* function calculates cosine similarity between two vectors.

On the other hand, scalar function eliminates the length normalization factor in the cosine function and is denoted as follows:

$$score(r, q, u) = (\boldsymbol{P_r} \cdot \boldsymbol{P_u}) \cdot (\boldsymbol{P_r} \cdot \boldsymbol{P_q}) \tag{18}$$

A different approach, called fuzzy ranking function, is introduced in [28] which computes the query relevance between a query and resources, and the user relevance at first, and then combines the two relevance scores to achieve the final ranking score as follows:

$$score(r, q, u) = \frac{\frac{s}{m} \cdot \frac{\sum p_i}{m} + \frac{\sum l_x \cdot s_i'}{m}}{2} \tag{19}$$

where s is the number of terms matching the query, m is the number of query terms, and

$$l_x = \begin{cases} p_i + (1 - s_i')(1 - p_i), & p_i \in (0, 1),\ s_i' > 0 \\ 1, & p_i = 1,\ s_i' > 0 \\ 0, & p_i = 0,\ s_i' > 0 \end{cases} \tag{20}$$

where p_i is the value for tag t_i from resource P_r and s_i' is the value for tag t_i from user profile P_u. The ranking in the final list is based on the relevance score (i.e., $score(r, q, u)$), implying that the greater value of the score, the higher the position of the resource in the list.

V. EXPERIMENTS

A. EXPERIMENTAL SETUP

We first describe how we made use of the folksonomy datasets for evaluation. We then proceed to present the details of our experiment setups.

1) DATASETS

To evaluate our framework, CiteULike[1] MovieLens[2], and Delicious[3] datasets are used.

The first dataset is CiteULike, a publicly-available dataset that contains social bookmarking features to researchers. We only used the annotations of users who have more than 100 tags with more than a month of use, in order to incorporate the temporal characteristics of the user topics drifting as time passes. After tag trimming, the test bed contained 124,236 documents and 63,559 individual tags. The average distinct number of tags per user was 64.82, while each resource had an average of 2.87 tags as its annotation.

To further evaluate our framework, we used another larger dataset, MovieLens, which includes 95,580 tuples of users, movies, and tags; this contains a breakdown of 71,567 users and 10,681 movies. While the user had an average of 10.62 tags, each resource had an average of 9.09 tags.

Lastly, we also used Delicious dataset, which is a social bookmarking web service for web-pages. It consists of 1,867 users with 105,000 bookmarks. For all of the datasets, we used five-fold cross validation to split training and testing datasets. We used the training data to build user and resource profiles. We then used the test data as input queries to evaluate the performance of search personalization based on the profiles. Those three datasets are often used for various folksonomy-based retrieval tasks, ensuring the reproducibility of the proposed approach.

In the experiment on the CiteULike dataset, we used a commercial search engine to detect the position of the target article in the search list returned by query tags. We limited the search engine to search only the resources within the CiteULike website. We downloaded the top 100 documents in the result list returned by the search engine, and discarded cases where the target resource was not found in the result list. By doing so, we can filter less relevant resources to the queries before we begin our experiment, and this process allows us to emulate a user using a real Web search system [10]. After the download, the average position of the target resource on the result list was 12.43.

For the MovieLens and Delicious dataset, ranking functions were used to calculate the relevance between query and resources instead of using a ranking retrieved by a search engine, as the targetting resources are either outdated or unavailable for automatic access with crawler. It is worth mentioning that our evaluation framework is similar to those in [9], [10], [28], implying that experiment results remain stable in a conventional evaluation framework for personalization in folksonomy.

[1]http://www.citeulike.org
[2]http://grouplens.org
[3]https://grouplens.org/datasets/hetrec-2011/

2) EVALUATION METRICS

To evaluate the proposed framework, we used five metrics in total: Mean Reciprocal rank (MRR), Personalization Gain (P-Gain) [22], and Precision at top 10, 20, 30 documents [31].

First, MRR measures how well a personalization approach rank resources. It is denoted as follows:

$$MRR = \frac{1}{n} \times \sum_{i=1}^{n} \frac{1}{POS_i} \tag{21}$$

where n is the number of queries and POS_i is the position of the target resource in the result list. The larger MRR is, the faster the user can access the resources in the list.

Second, P-Gain compares the number of times the personalization approach improves the ranking with the number of times it worsens compared to the baseline. This can be denoted as follows:

$$P - Gain = \frac{\#better - \#worse}{\#better + \#worse} \tag{22}$$

based on this metric, a value of 0 indicates no change in the rankings through personalization, a positive value indicates an improvement and a negative value indicates a degradation in performance. We selected NTF to be the baseline of P-Gain as it was reported to be the best-performing model for folksonomy-based search personalization in [28].

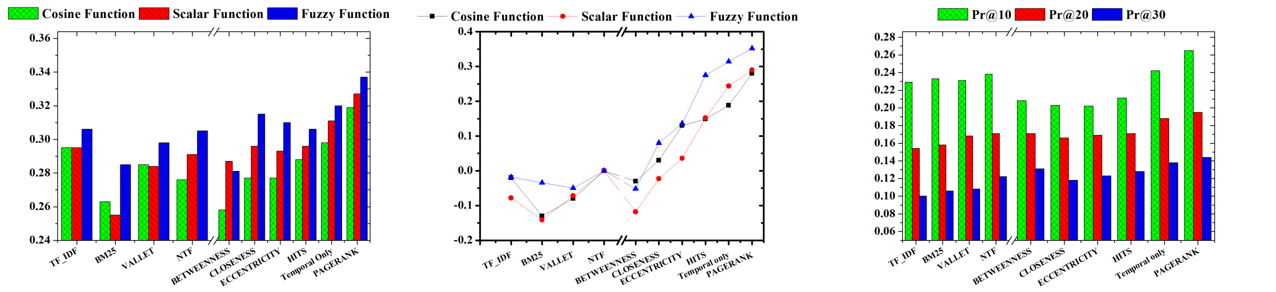
Lastly, Precision at top 10, 20, 30 documents, denoted as Pr@10, Pr@20, and Pr@30, are used to measure the accuracy of the proposed model in terms of precision. Pr@k is the proportion of top-k ranked documents that are relevant.

3) BASELINES

We used TF-IDF [9], BM25 [9], Vallet [10], and NTF [28] as baselines to construct user and resource profiles. These models are the current mainstream approaches in search personalization in folksonomy. In addition, we also investigated Betweenness, Closeness, Eccentricity, HITS, and PageRank to calculate the tag importance score, which considers temporal and multiple aspects. We then compared their performances with respect to the above three metrics.

We also compared two state-of-the-art approaches, Distinctness [31] and MPU (multi-level user profiling) [30], with the proposed model for MovieLens dataset. Those approaches exploited addition user information, that is user rating data. In both approaches, when building the profiles for personalization, rating information were adopted for weighting important edges. As the other two datasets do not contain user rating data, therefore, those approaches were only built for the performance evaluation in MovieLens.

Lastly, we built a variant of the proposed model, called temporal only, that excludes multiplicity aspect to observe the performance gap as each feature is added to the proposed approach.



(a) MRR Result. (b) P-Gain Result. (c) Pr@10,20,30 Result on Fuzzy Function.

FIGURE 1. Performance evaluation of CiteULike.

Moreover, the performances on the five evaluation metrics above were evaluated by using all ranking functions introduced in Section IV. This in-depth investigation was designed to show that our framework is superior if achieving a consistently improved performance compared to any of the ranking functions.

B. RESULTS ON CITEULIKE DATASET

Fig. 1-(a) shows the MRR scores of the personalization models on CiteULike. In general, we can observe that fuzz ranking function promises more reliable performance for search personalization than cosine and scalar functions. This is due to the fact that this function is more fine-tuned for folksonomy-based personalization. On the fuzz ranking function, the proposed framework, which adopts Closeness, Eccentricity, HITS, and PageRank, outperforms all of the baselines by 6.06%, 4.48%, 6.38%, and 6.70% on average, respectively. Similarly, those link analysis techniques also outperform the baselines on scalar function. However, Betweenness does not provide a significant improvement compared to the baselines on none of the ranking functions. Nonetheless, the improvements obtained from the other link analysis techniques are a strong indication that reflecting the temporality and multiplicity of topics of a user profile can yield a positive effect on search personalization in folksonomy. Specifically, among the link analysis techniques, PageRank shows a consistent improvement over the three ranking functions followed by HITS. The MRR results show that, on average, regardless of the ranking functions, the personalization based on PageRank and HITS outperforms the baselines by 7.25% and 4.22%, respectively. Notably, temporal only shows comparable performance against all baselines while the proposed model considering both multiplicity and temporality always shows the best performance. These results demonstrate that each aspect contains a piece of meaningful information and properly unifying these two aspects is crucial for the personalized search. Note that all of the improvements were found to be significant through a Wilcoxon test with a confidence level of 0.01.

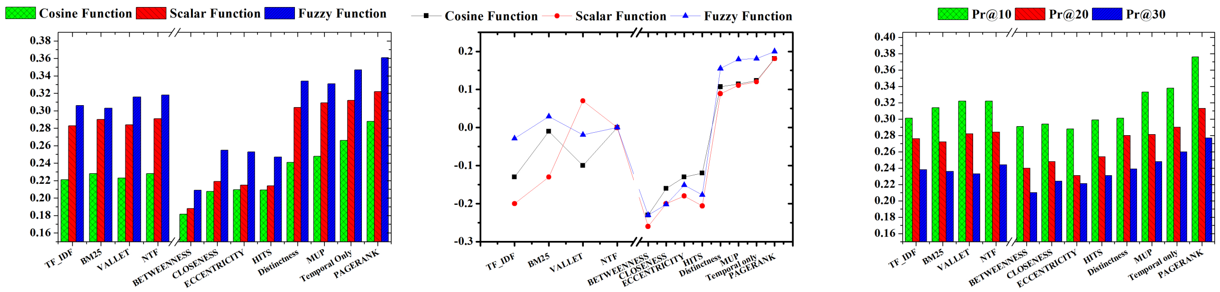
By looking at the P-Gain scores in Fig. 1-(b), we can be more confident that the proposed framework fully benefits from exploiting temporal and multiple aspects. In particular, Eccentricity, HITS, and PageRank significantly outperform the baseline, which is set to NTF in 10.06%, 19.2%, and 30.73%, while others often exhibit inferior performance to NTF. Indeed, in related studies [36], [37], they also report that PageRank is more effective than the centrality methods for certain tasks, especially, ranking the vertices in the network.

As fuzzy ranking function shows the best performance in two metrics, we additionally show precision scores on this function at top 10, 20, and 30 resources in the ranked list in Fig. 1-(c). We observe that PageRank outperforms all the other methods for all k. Based on the above observations, we can see that PageRank is the most stable link analysis technique to calculate importance scores for vertices, regardless of ranking functions. Due to the page limit, we do not include the precision results of the other two functions; however, we observed that similar tendencies were shown in both functions.

C. RESULTS ON MOVIELENS DATASET

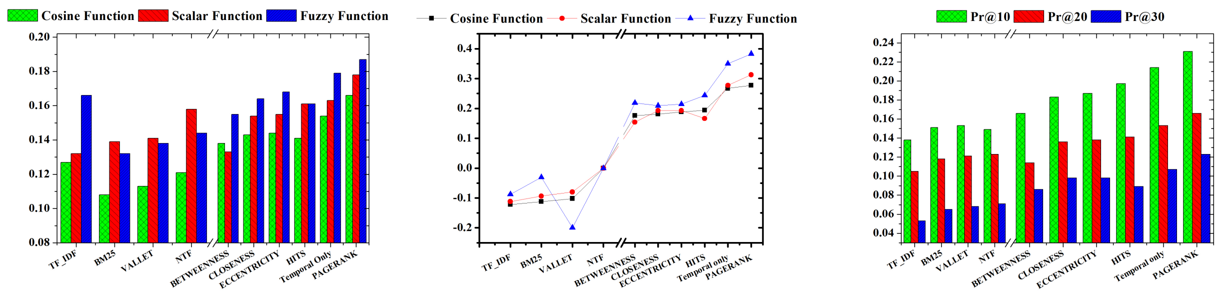
Fig. 2-(a) shows the MRR scores of the personalization models on MovieLens. Interestingly, we observed a marked drop in the HITS model in terms of MRR although it performs as effectively as PageRank on CiteULike. Furthermore, all of the link analysis techniques except PageRank show lower performances than the baselines. Fig. 2-(b) also presents a similar tendency in terms of P-Gain. Except for the PageRank model, all of the link analysis techniques show lower performance than the baseline, which is the NTF model. Fig. 2-(c) illustrates that precision at 10, 20, and 30 on fuzzy ranking function show the same tendency as the performance based on link analysis techniques, except for PageRank, is lower than the baselines. Among baselines, MUP shows the best performance while Distinctness shows the second best performance in general. However, in all settings, the proposed algorithm always surpasses these two state-of-the-art algorithms, demonstrating the superiority of the proposed one.

To explain this tendency, we observed that the average network density is 0.02 and 0.199 for MovieLens and CiteULike, respectively. Furthermore, the average number of connected components is 80.8 and 14.2 for MovieLens and CiteULike, respectively. Those values indicate that the user profile networks in MovieLens are mostly sparser and less connected



(a) MRR Result. (b) P-Gain Result. (c) Pr@10,20,30 Result on Fuzzy Function.

FIGURE 2. **Performance evaluation of MovieLens.**



(a) MRR Result. (b) P-Gain Result. (c) Pr@10,20,30 Result on Fuzzy Function.

FIGURE 3. **Performance evaluation of delicious.**

compared to in CiteULike. In this case, values related to measuring centrality become 0 for most of the tags in the network. Consequently, the tags are overlooked in calculating the relevance between the user and resource profiles, causing underperformance of personalization based on Betweenness, Closeness, and Eccentricity models. In addition, for the HITS model, since vertices are less connected, some of the tags in the network are unreachable, causing a value of 0 for those tags. However, PageRank has a random walk feature. This allows the model to allocate a certain value to a disconnected tag by jumping to the tag from the distant one. Due to this feature, the PageRank model still performs well when the network is sparse. In terms of MRR, on average, the personalization based upon PageRank model outperforms the baselines and other ranking analysis models by 10.55% and 49.51%. In terms of P-Gain, it outperforms the baselines and other ranking analysis models by 18.79% and 43.28% on average, respectively. The improvements are significant according to a Wilcoxon test with a confidence level of 0.01. Based on this observation, we can conclude that our personalization framework has the potential to improve search personalization in folksonomy; however, careful consideration is also required for adopting a ranking function and link analysis technique to ensure reliable performance of the chosen folksonomy data.

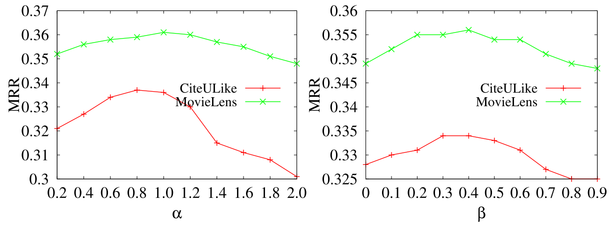
D. RESULTS ON DELICIOUS DATASET

Fig. 3-(a) shows the MRR scores of the personalization models on Delicios. Similar to the MovieLens datasets, we observed a marked drop for all of the link analysis techniques except PageRank show lower performances than the baselines. Fig. 3-(b) also presents a similar tendency in terms of P-Gain. Except for the PageRank model, all of the link analysis techniques show lower performance than the baseline, which is the NTF model. Fig. 3-(c) illustrates that precision at 10, 20, and 30 scores on fuzzy ranking function show the same tendency as the performance based on link analysis techniques, except for PageRank, is lower than the baselines.

By observing results based on all of the datasets, we have confirmed that PageRank performs robustly due to its feature of random walk and is the best option to be adapted, in order to outperform existing state-of-the-art approaches. In fact, it is worth noting that this observation is specifically meaningful in folksonomy network as most of the user-resource network based on folkosonomy is sparse. In practice, few users annotate most resources on folksonomy and users in long-tail either annotate few resources or leave nothing.

E. THE IMPACT OF PARAMETERS

In the previous chapter, we introduced parameters α and β while building user profiles. The value of α controls the impact of the multiplicity aspect into building user profiles. The value of β controls the impact of the temporal aspect of topics in the decay function. If $\beta = 0$, the decay function is deactivated. To evaluate the impact of the parameters, we vary one parameter while fixing the other at random value. After we find the best value, we then vary the remaining parameter while the other is fixed at the best value. Fig. 4-(a) illustrates



(a) The Impact of α. (b) The Impact of β.

FIGURE 4. The effect of varying parameters.

the effect of varying parameter α in terms of MRR[4]. When comparing the plots on two datasets, we observe that the performance peaks are approximately between 0.8 and 1.0, slowly decreasing afterwards.

Similarly, Fig. 4-(b) presents the effect of varying parameter β in terms of MRR. When comparing two plots on two datasets, we observe the performance peaks approximately between 0.3 and 0.4 and slowly decreases afterwards. We set the parameters to its best performing value in our experiment. In addition, the performance of the proposed framework without the decay function, when $\beta = 0$, is lower than the average of the performances when β is set to be the best performing value. This implies that incorporating the information of temporal and multiple properties of user topics is beneficial to improve the performance of personalization in folksonomy.

VI. FUTURE WORKS AND CONCLUDING REMARKS

In this paper, we presented a novel search personalization framework that exploits folksonomy and its characteristics by representing a user profile as a user profile network. The network consists of a set of tags annotated by the user and edges between the tags. In particular, in the proposed framework, the temporal and multiple properties of user topics are considered by strengthening the edges of the user profile network that own the properties. Then, link analysis techniques are applied to rank the tags in the network by their importance for search personalization.

We conclude that our search personalization framework has proven to be the most effective model compared to the other folksonomy-based personalization models in our experiments. We also explored several link analysis techniques for ranking the tags in the network. In our experiment, the performance for each link analysis technique varies, but PageRank was shown to be constantly effective due to its random walk feature. The performance of our framework also depends on the unique characteristics of chosen folksonomy, implying that ranking function and link analysis techniques for achieving the best performance can be different for each folksonomy. Nonetheless, the proposed framework shows promising results for tapping into the potential of essential aspects of folksonomy for the enhancement of search personalization.

[4]We conducted the tests by using other evaluation metrics as well, and the results looked almost the same.

In our future study, we aim to further improve our framework by (1) circumventing the problem of disconnectedness, which exhibits in the centrality-based link analysis techniques, (2) learning to rank the importance of user tags based on various topological features of user profile networks, and (3) efficiently updating the user profile network. One might concern the privacy issue. However, coping with the privacy issue is one large research area itself [38], [39]. Furthermore, those studies are applicable to our research since they are orthogonal to our framework. Since this problem is beyond the scope of this paper, we leave it as future work as well.

REFERENCES

[1] A. K. Singh, N. K. Nagwani, and S. Pandey, "TAGme: A topical folksonomy based collaborative filtering for tag recommendation in community sites," in *Proc. 4th Multidisciplinary Int. Social Netw. Conf.*, 2017, Art. no. 27.

[2] A. Oeldorf-Hirsch and S. S. Sundar, "Posting, commenting, and tagging: Effects of sharing news stories on facebook," *Comput. Hum. Behav.*, vol. 44, pp. 240–249, Mar. 2015.

[3] O. Tsur and A. Rappoport, "What's in a hashtag?: Content based prediction of the spread of ideas in microblogging communities," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, 2012, pp. 643–652.

[4] D. Zhou, X. Wu, W. Zhao, S. Lawless, and J. Liu, "Query expansion with enriched user profiles for personalized search utilizing folksonomy data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1536–1548, Jul. 2017.

[5] H. Yu, B. Zhou, M. Deng, and F. Hu, "Tag recommendation method in folksonomy based on user tagging status," *J. Intell. Inf. Syst.*, vol. 50, no. 3, pp. 479–500, 2017.

[6] Z. Xu, O. Tifrea-Marciuska, T. Lukasiewicz, M. V. Martinez, G. I. Simari, and C. Chen, "Lightweight tag-aware personalized recommendation on the social Web using ontological similarity," *IEEE Access*, vol. 6, pp. 35590–35610, 2018.

[7] M. M. Rahman and N. A. Abdullah, "A personalized group-based recommendation approach for Web search in E-learning," *IEEE Access*, vol. 6, pp. 34166–34178, 2018.

[8] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *J. Inf. Sci.*, vol. 32, no. 2, pp. 198–208, 2006.

[9] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu, "Exploring folksonomy for personalized search," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2008, pp. 155–162.

[10] D. Vallet, I. Cantador, and J. M. Jose, "Personalizing Web search with folksonomy-based user and document profiles," in *Proc. Eur. Conf. Inf. Retr.*, 2010, pp. 420–431.

[11] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui, "Modeling the impact of short- and long-term behavior on search personalization," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2012, pp. 185–194.

[12] S. Lee, M. Ko, K. Han, and J.-G. Lee, "On finding fine-granularity user communities by profile decomposition," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2012, pp. 631–639.

[13] A. Yeung, C. Man, N. Gibbins, and N. Shadbolt, "A study of user profile generation from folksonomies," in *Proc. World Wide Web Workshop Social Web Knowl. Manage.*, 2008, pp. 1–8.

[14] E. Michlmayr and S. Cayzer, "Learning user profiles from tagging data and leveraging them for personal (ized) information access," in *Proc. 16th Int. World Wide Web Conf. Workshop Tagging Metadata Social Inf. Org.*, 2007, pp. 1–7.

[15] T. Vu, A. Willis, S. N. Tran, and D. Song, "Temporal latent topic user profiles for search personalisation," in *Proc. Eur. Conf. Inf. Retr.*, 2015, pp. 605–616.

[16] J. Kim, S. Lim, J.-G. Lee, and B. S. Lee, "LinkBlackHole: Robust overlapping community detection using link embedding," *IEEE Trans. Knowl. Data Eng.*, to be published.

[17] J. Kim, J.-G. Lee, and S. Lim, "Differential flattening: A novel framework for community detection in multi-layer graphs," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, 2017, Art. no. 27.

[18] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, 2008, Art. no. P10008.

[19] U. Brandes, "A faster algorithm for betweenness centrality," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001.

[20] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM* , vol. 46, no. 5, pp. 604–632, 1999.

[21] J. Kim, M. Choy, D. Kim, and U. Kang, "Link prediction based on generalized cluster information," in *Proc. 23rd Int Conf. World Wide Web*, 2014, pp. 317–318.

[22] M. Harvey, F. Crestani, and M. J. Carman, "Building user profiles from topic models for personalised search," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage.*, 2013, pp. 2309–2314.

[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[24] T. Vu, D. Q. Nguyen, M. Johnson, D. Song, and A. Willis, "Search personalization with embeddings," in *Proc. Eur. Conf. Inf. Retr.*, 2017, pp. 598–604.

[25] S. Goel and R. Kumar, "Folksonomy-based user profile enrichment using clustering and community recommended tags in multiple levels," *Neurocomputing*, vol. 315, pp. 425–438, Nov. 2018.

[26] L. Luo, H. Xie, Y. Rao, and F. L. Wang, "Personalized recommendation by matrix co-factorization with tags and time information," *Expert Syst. Appl.*, vol. 119, pp. 311–321, Apr. 2019.

[27] M. G. Noll and C. Meinel, "Web search personalization via social bookmarking and tagging," in *The Semantic Web*. Berlin, Germany: Springer, 2007, pp. 367–380.

[28] Y. Cai and Q. Li, "Personalized search by tag-based user profile and resource profile in collaborative tagging systems," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 969–978.

[29] Y. Cai, Q. Li, H. Xie, and H. Min, "Exploring personalized searches using tag-based user profiles and resource profiles in folksonomy," *Neural Netw.*, vol. 58, pp. 98–110, Oct. 2014.

[30] Q. Du, H. Xie, Y. Cai, H.-F. Leung, Q. Li, H. Min, and F. L. Wang, "Folksonomy-based personalized search by hybrid user profiles in multiple levels," *Neurocomputing*, vol. 204, pp. 142–152, Sep. 2016.

[31] G. E. Lee, K. Han, and Y. Y. Mun, "Incorporating distinct opinions in content recommender system," in *Proc. AIRS*, 2015, pp. 109–120.

[32] A. Byde, H. Wan, and S. Cayzer, "Personalized tag recommendations via tagging and content-based similarity metrics," in *Proc. 13th Int. AAAI Conf. Web Social Media*, 2007, pp. 1–3.

[33] S. Kim, K. Han, M. Y. Yi, S. Cho, and S. Kim, "Exploiting knowledge structure for proximity-aware movie retrieval model," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage.*, 2014, pp. 1847–1850.

[34] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," Stanford InfoLab, Tech. Rep. 1999-66, 1999.

[35] T. Opsahl, F. Agneessens, and J. Skvoretzc, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Netw.*, vol. 32, no. 3, pp. 245–251, 2010.

[36] Y. Ding, E. Yan, A. Frazho, and J. Caverlee, "PageRank for ranking authors in co-citation networks," *J. Assoc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2229–2243, 2009.

[37] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel, "Co-authorship networks in the digital library research community," *Inf. Process. Manage.*, vol. 41, no. 6, pp. 1462–1480, 2005.

[38] G. Guo, T. Yang, and Y. Liu, "Search engine based proper privacy protection scheme," *IEEE Access*, vol. 6, pp. 78551–78558, 2018.

[39] Y. S. Yoon, H. Zo, M. Choi, D. Lee, and H.-W. Lee, "Exploring the dynamic knowledge structure of studies on the Internet of Things: Keyword analysis," *ETRI J.*, vol. 40, pp. 745–758, Dec. 2018.

KEEJUN HAN received the B.S. degree in information technology from Australian National University, in 2010, and the M.S. degree in knowledge service engineering from Korea Advanced Institute of Science and Technology, in 2013, where he is currently pursuing the Ph.D. degree with the Graduate School of Knowledge Service Engineering. His research interests include recommender systems and bioinformation retrieval.

MUN Y. YI received the Ph.D. degree in information systems from the University of Maryland at College Park. He is currently a tenured Professor with the Korea Advanced Institute of Science and Technology. His current research focuses on knowledge engineering, biomedical information processing, and human–computer interaction. He currently serves as an Associate Editor for the *International Journal of Human-Computer Studies* and a Senior Editor of *AIS Transactions on Human–Computer Interaction*.

JUNGEUN KIM received the B.S. degree in information and communication engineering from Inha University, in 2011, and the M.S. and Ph.D. degrees in knowledge service engineering from KAIST, in 2013 and 2017, respectively. He is currently a Senior Researcher with the Intelligence Information Research Division, Electronics and Telecommunications Research Institute. His research interests include graph big data mining, big data analysis with the distributed processing platforms, and recommender systems.