Received December 28, 2018, accepted January 17, 2019, date of publication January 29, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895824

Semantic-Enhanced and Context-Aware Hybrid Collaborative Filtering for Event Recommendation in Event-Based Social Networks

MINGHUA XU[1] AND SHENGHAO LIU[2]

[1]School of Journalism and Information Communication, Huazhong University of Science and Technology, Wuhan 430074, China

[2]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Shenghao Liu (shenghao_liu@hust.edu.cn)

This work was supported in part by the Huazhong University of Science and Technology Social Science Major Project, under Grant 2019WKZDJC019.

ABSTRACT The fast development of *event-based social networks* (EBSN) provides a convenient platform for recruiting offline participants via online event announcements. Given its ever-increasing new events, how to accurately recommend users their most preferred ones is a key to the success of an EBSN. In this paper, we propose a *semantic-enhanced and context-aware hybrid collaborative filtering* for event recommendation, which combines semantic content analysis and contextual event influence for user neighborhood selection. In particular, we first exploit the latent topic model for analyzing event description text and establish each user a long-term interest model and short-term interest model from her event registration history. We next establish each event an influence weight to jointly represent its social impact among users and its semantic uniqueness among events. For one user, we select her neighbors according to their long-term interest similarities weighted by events' influences. For new event recommendation, we construct a user-event rating matrix based on users' short-term interest models and for each user, we compute event rating predictions from her neighbors' ratings. The experiments based on the real-world dataset demonstrate the superiority of our algorithm over the peer schemes.

INDEX TERMS Hybrid collaborative filtering, event semantic analysis, event influence weight, event recommendation, event-based social networks.

I. INTRODUCTION

Recently, many *event-based social networks* (EBSNs), such as Meetup and Douban Event,[1] have emerged for helping people to announce upcoming events online and recruit offline participants [1]–[4]. Given the ever-increasing of new events in EBSNs, it becomes harder for users to quickly find out their mostly preferred events [5]–[7]. For example, Meetup currently has 16 million users with more than 300,000 events announced per month [1]. Although search engines can help users to find their interested events with keywords, it is still a challenging task for users to match their preferences with appropriate events. Even worse, some users are unable to clearly express their interests. In response to the pressing demands, a good event recommendation system is much required for EBSNs.

Event recommendation in EBSNs, though shares some similarities with recommending general items like book, movie, music and etc. [8]–[11], presents some special characteristics like the *cold start* problem and *social relation* issue [12]–[18]. The cold start problem refers to that an event cannot be actually *'consumed'* and evaluated before its commencement time. In practice, an EBSN often provides a kind of *registration* mechanism for users to register a new event. For example, in Douban Event, a user can register to **'join'** in one event or register to indicate that she is **'interested'** in the event. However, the two actions are exclusive to each other: A user cannot register as both to one event at the same time.

The social relations are particular important in EBSNs [16], [19]. Users who have attended a same event

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

[1]Meetup: www.meetup.com; Douban Event: www.douban.com

might establish some *explicit* social relations and even become friends; While some users would like to attend an event simply because that her friends decide to do so. On the other hand, users registering a same event may also indicate a kind of *implicit* social relations by their similar preferences to the event. The classical *collaborative filtering* (CF) recommendation can help to catch such social relations by *user neighborhood construction*: that is, users with the close social relations to a user are selected as her neighbors [20]. While their neighbors' registration information of an event are used for her *prospective rating prediction*.

However, the CF based merely on event co-registration might not be enough to profile users' preferences. Fortunately, most EBSNs also provide each event a piece of text description, which not only states the event expense, time and location, but also provides an in-depth introduction about the event type and its selling points [2]. With the recent advances of *natural language processing* (NLP) [21], many tools have been developed for analyzing text to obtain a kind of latent topic distribution as an event feature vector [22], which can help profile a user's interest model from her historical registrations and can also be used for *content-based* recommendation [23].

In this paper, we propose a *semantic-enhanced and context-aware hybrid collaborative filtering* (scHCF) for event recommendation, which consists of two parts: user neighborhood construction and event rating predication. We first extract each event semantic feature vector from its text description. Based on the historical registration, we establish each user a *long-term interest model* (LTIM) as a weighted sum of her registered events' features. Furthermore, we propose a *context-aware event influence* for each event to jointly represent its social impact among users and its semantic uniqueness among events. For one user, we select her neighbors according to their long-term interest similarities weighted by events' influences. For event rating prediction, we construct a user-event rating matrix based on users' *short-term interest models* (STIM) and for each user we compute event rating predictions from her neighbors' ratings. We compare the proposed scHCF scheme with peer schemes via experiments on two real datesets crawled from Douban Event for two cites, Beijing and Shanghai, in China. The experiment results validates its superiority over the peer schemes.

Our contributions are summarized as follows:

- Propose to establish a time attenuated LTIM from event semantic analysis;
- Propose a new context-aware event influence weight for user neighborhood construction;
- Propose to apply a STIM for prospective event rating prediction;
- Experiment on the real EBSN datasets for performance comparisons.

The rest of the paper is structured as follows: Section II provides a brief review on the related work. The proposed scHCF scheme is presented in Section III and experimented in Section IV. The paper is concluded in Section V.

II. RELATED WORK

The CB recommendation algorithms have been extensively researched in the last decades [24]–[31]. The basic idea is to establish a user's interest or preference model through her historical actions, like having purchased some books. As some items are with text descriptions, the *Latent Dirichlet Allocation* (LDA) and the *Term Frequency–Inverse Document Frequency* (TF–IDF) method have been employed to map the item descriptions and the user interests into a same vector space, such that their vector distance can be computed and used for rating items [22], [28], [32]–[35]. For example, Narducci *et al.* [28] propose a content-based recommender system which is able to generate cross-lingual recommendations for books and movies by exploiting the TF-IDF technique to map different language documents into the same space of Wikipedia concepts. Macedo *et al.* [22] propose a context-aware approach for event recommendation by leveraging several contextual signals as features to rank events, in which the TF-IDF technique is applied to analysis events' description. Krestel *et al.* [33] introduce an approach for recommending tags of resources in order to improve search performance, which overcomes the cold start problem for tagging new resources by leveraging the LDA technique for eliciting latent topics from resources and recommending tags to the new resources based on the topics' similarities. Wang *et al.* [35] present a novel sematic-based friend recommendation system, which models users' daily lives as life documents and applies LDA algorithm to extract users' life styles. Although these CB recommendation algorithms exploit the LDA or TF-IDF for establishing users' interest models, they have not take into consideration users' social relations that may be embedded in the co-registration of same events.

The CF recommendation algorithms is another classic recommendation algorithm which pays attention to various relations in between users and items. In CF algorithm, the core step is the neighborhood user construction, which helps a recommendation system to match users with the items that their neighborhood users would also like to choose [12]–[14], [36]–[41]. For example, Sun and Chen [12] propose a social event recommendation method by exploiting users' social and collaborative friendships which are aggregated to identify the acquaintances of a user and events relevant to the preferences of the acquaintances. Gao *et al.* [13] introduce a novel Bayesian latent factor model that not only considers individual preferences, but also takes social group influence into consideration for event recommendation. Gu *et al.* [14] propose a context aware matrix factorization model to tackle with the cold-start problem, which captures users' preferences by explicit contextual features and combines them with the matrix factorization model.

Recently, some hybrid recommendation algorithms combining both the CB and CF technique have been proposed [23], [42]–[50]. For example, Khrouf and Troncy [42] propose a weighted hybridization using a linear combination of recommendation scores calculated through the content similarity and CF technique, respectively. Hsieh *et al.* [23] propose a user-centric recommendation model that exploits users' diverse personal digital traces for extracting users' interests via the LDA technique and then makes recommendation which can be fine-tuned to meet a user's interests on the target platform. Wei *et al.* [50] design a deep neural network to extract the items' features and make use of the CF model to take the content features into rating prediction for cold start items.

III. SEMANTIC-ENHANCED AND CONTEXT-AWARE HYBRID COLLABORATIVE FILTERING

A. OVERVIEW

The proposed scHCF predicts unregistered new events' ratings for each individual user upon her online access to the EBSN and recommends the top-N unregistered new events with the N-highest predicted ratings. Upon the recommendation time, we divide the timeline into two consecutive parts: each containing the so-called *new events* and *history events*. The new events are those events that have not yet started. Notice that new events might be announced online long time ago and some users might have already registered. History events are those that have already completed. Furthermore, among history events, we also identify a subset of most recently finished events, called *recent events*.

The proposed scHCF contains two parts: user neighborhood construction and event rating prediction. In the former phase, we first establish users' long-term interest models and compute events' influence factors from history events, which are then exploited to construct the user neighborhood for each individual user. In the latter phase, we establish users' short-term interest models from recent events and construct a user-event rating matrix for new events already registered by some users; While the rating prediction for a user's unregistered new event is computed from her neighbors' predicted ratings. The framework of scHCF algorithm is illustrated in Fig. 1.

B. USER NEIGHBORHOOD CONSTRUCTION

For each user, the user neighbor construction is to select the her mostly similar users. To do so, we resort to event semantics to first establish user interest models. In Douban Event, each event announcement also includes a text description to describe what the event is about, along with the event location and time. We apply the *Latent Dirichlet Allocation* (LDA) technique [51] to analyze each event text description and obtain an *event feature vector*. For an event, let $\boldsymbol{e}$ and t_e denote its feature vector and its commencement time.

For a user u, let $\mathcal{L}_{jo}$ and $\mathcal{L}_{in}$ denote the list of the history events that the user u had registered as 'join' and 'interested', respectively. An element $(\boldsymbol{e}, t_e) \in \mathcal{L}_{jo}$ indicates that the event $\boldsymbol{e}$ with commencement time t_e was registered as 'join' by the user. For each user, we establish a *long-term interest model* (LTIM) that combines her history event features in a time-attenuated way, which is computed by

$$\boldsymbol{u} = \alpha_{jo} \sum_{\boldsymbol{e} \in \mathcal{L}_{jo}} \omega_e \boldsymbol{e} + \alpha_{in} \sum_{\boldsymbol{e} \in \mathcal{L}_{in}} \omega_e \boldsymbol{e}, \tag{1}$$

where α_{jo} and α_{io} are the system weights for discriminating the two online registration types. Usually, we argue that *'attending to an event'* would be more important to *'interesting to an event'*. We use the training data to decide the appropriate values of α_{jo} and α in our experiments. The event weight ω_e is to measure how a past event would influence the user interest in the current recommendation. As a user's interest might be changing with time, we argue that the older the past event, the less influence to the current recommendation. So we apply an exponential time decay model to compute ω_e as follows:

$$\omega_e = \exp(-\lambda(\tau - t_e)), \tag{2}$$

where τ is the current recommendation time, and λ is the decay coefficient, deciding the speed of attenuation. In our experiments, we notice that most of users normally do not attend more than three events in one month. So the time unit is set as one month, i.e., τ - t_e refers to the months in between two events' commencement time, and we use training data to decide an appropriate value of λ in our experiments.

Many offline events are kind of social activities, which might involve lots of people interactions. The long-term interest model $\boldsymbol{u}$ in Eq. (1) is merely based on the event content semantics, which might not be enough to describe people social relations. We argue that including social relations for event recommendation would augment its performance. To do so, we resort to exploit users' event participation information as well as their ratings to events. For M users and N_{hst} history events, we construct a *user-event historical rating matrix* $\boldsymbol{A} \in \mathbb{R}^{M \times N_{hst}}$ to represent the similarities between users' interests and events' features. An element $a_{ij} \in \boldsymbol{A}$ is computed by:

$$a_{ij} = \begin{cases} \dfrac{\boldsymbol{u}_i \cdot \boldsymbol{e}_j}{\|\boldsymbol{u}_i\| \|\boldsymbol{e}_j\|}, & \text{if } u_i \text{ registered } e_j, \\ 0, & \text{Otherwise.} \end{cases} \tag{3}$$

Notice that we use the commonly used cosine distance between a user LTIM and an event feature as the user's rating to the event. This is because the original dataset from Douban Event does not provide the direct rating mechanism. Yet only using the registration type might be too coarse for describing a user's preference to an event, as we argue that in most cases it is the event content that mostly attracts a registration.

Each row of $\boldsymbol{A}$ represents a user's ratings to her registered events, which can indicate her preferences in some particular events to some extent. Two users who had registered a same event with similar ratings are normally considered as with similar preferences and close relations. On the other hand, some kind of popular events with exceptional large amount

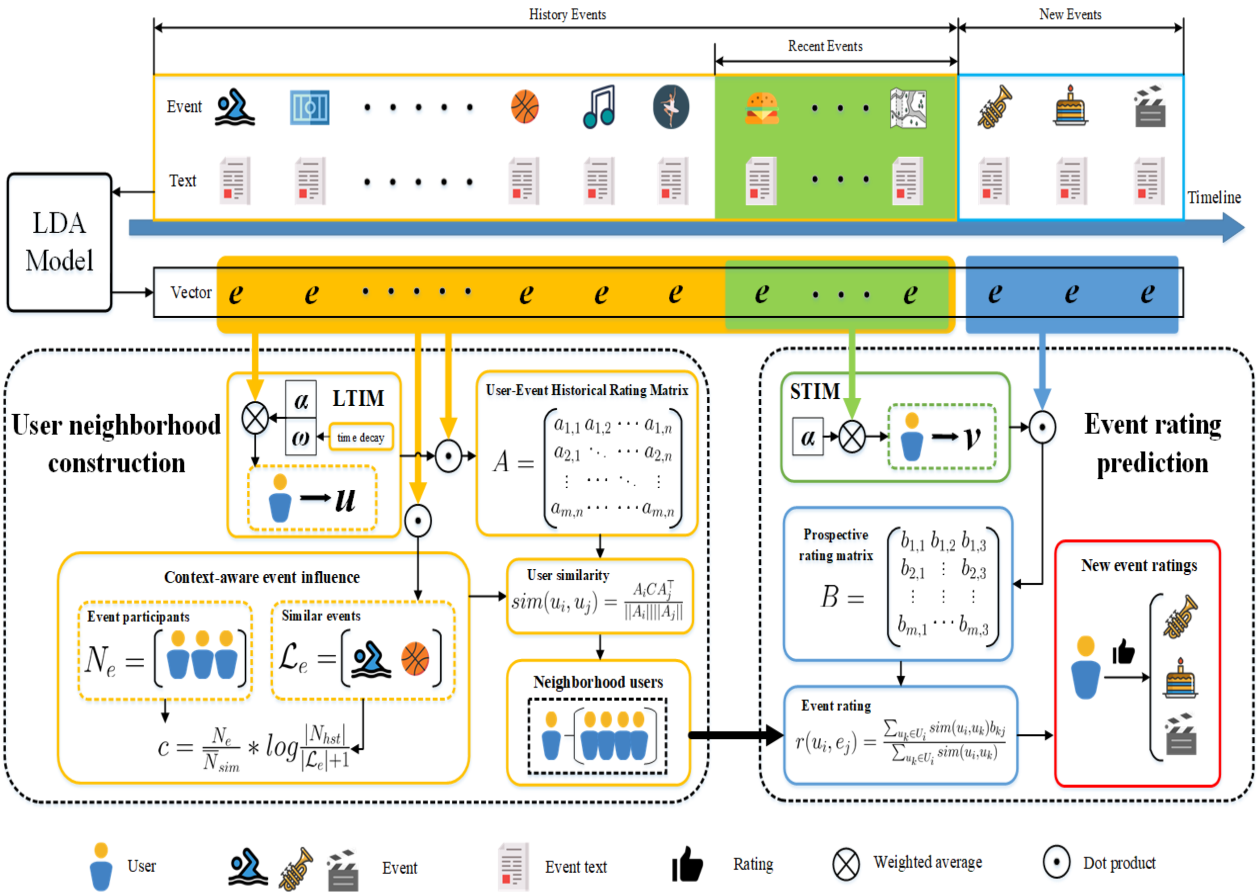


FIGURE 1. **Illustration of the framework of the proposed scHCF scheme, which consists of two parts: user neighborhood construction and event rating prediction. After applying the *Latent Dirichlet Allocation* (LDA) technique, each event obtains an even feature vector from its text description. (i) In the user neighborhood construction, a *long term interest model* (LTIM) is first established for each user based on her registered history events, and a user-event historical rating matrix is constructed according to the similarities in between the users' LTIMs and events' features. A context-aware event influence is then computed for each event based on the event's participant information, and the most similar users are selected for each user as her neighborhood from the historical rating matrix yet weighted by the influence weights. (ii) In the event rating prediction, a *short-term interest model* (STIM) is first established for each user from her recent events' features, and a user-event prospective rating matrix is constructed according to the similarities in between the users' STIM and new events' features. The user' prospective ratings to her unregistered new events are then computed from her neighbors' prospective ratings and the top-*N* new events with the highest ratings are recommended to the user.**

of participants, like concerts and shows, might not be representative enough to describe the participants' social relations; While some events with fewer participants may better represent the participants' interest focuses and describe their social relations. To this end, we define each history event a context-aware *influence weight* to represent its implicit impact on users' social relations. For a past event e, let N_e denote the number of its participants. Furthermore, we compute the semantic similarity between two events e_i and e_j by:

$$sim(\boldsymbol{e}_i, \boldsymbol{e}_j) = \frac{\boldsymbol{e}_i \cdot \boldsymbol{e}_j}{\|\boldsymbol{e}_i\| \|\boldsymbol{e}_j\|}. \tag{4}$$

We say that the two events are semantic similar, if $sim(\boldsymbol{e}_i, \boldsymbol{e}_j) \geq \sigma$, where σ is a predefined threshold. For an event e, let $\mathcal{L}_e$ denote the list of similar events.

For each history event e, we define its context-aware influence weight c by

$$c = \frac{N_e}{\overline{N}_{sim}} \times \log \frac{N_{hst}}{|\mathcal{L}_e| + 1}, \tag{5}$$

where $|\mathcal{L}_e|$ is the number of similar events and $\overline{N}_{sim}$ is the average number of participants of the similar events:

$$\overline{N}_{sim} = \frac{1}{|\mathcal{L}_e| + 1} \sum_{e' \in \mathcal{L}_e} N_{e'} \tag{6}$$

According to Eq. (5), the influence of an event is not only related to its number of participants, but also related to the average participants of its similar events as well as the portion of such similar events in all the history events. An event is considered more influential if its participants are much more than its similar events and these events are not too many. For all N history events, we construct a diagonal weight matrix $\boldsymbol{C} \in \mathbb{R}^{N \times N}$, where the diagonal elements are the corresponding event influence weights and all non-diagonal elements equal zero.

We then compute an event influence weighted user similarity between two users for neighborhood construction as follows:

$$sim(u_i, u_j) = \frac{\boldsymbol{A}_i \boldsymbol{C} \boldsymbol{A}_j^\top}{\|\boldsymbol{A}_i\| \|\boldsymbol{A}_j\|}, \tag{7}$$

where $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$ are the ith and jth row vector of the rating matrix $\boldsymbol{A}$, respectively. For each user u_i, we sort its similar users according to their similarity values in a decreasing order and choose the top K similar users to construct her neighborhood, denoted by $\mathcal{U}_i$.

C. EVENT RATING PREDICTION

For one user, event recommendation can be made upon each of her accesses to the system. We notice that users' accesses are in an asynchronous way: some accessing the system in the daytime, yet some in the nighttime. Furthermore, as new events are normally announced long before their actual commencement time, users' asynchronous accesses make it possible that some users have already registered new events, while some not yet. As such, for one user we can exploit her neighborhood users' registration information for her event recommendation. On the other hand, as new events have not yet been actually 'consumed', we again resort to the similarity between event semantic feature and user interest model to obtain a kind of prospective ratings to new events.

For the prospective ratings, we propose to use the *short term interest models* (STIMs) which are established with the similar procedure to the LTIM, yet only based on the *recent events* that have passed not more than T time unit ago. For a user u, let $\mathcal{S}_{jo}$ and $\mathcal{S}_{io}$ denote the list of recent events that the user u had registered as 'join' and 'interested', respectively. Her short term interest model $\boldsymbol{v}$ is computed by:

$$\boldsymbol{v} = \alpha_{jo} \sum_{e \in \mathcal{S}_{jo}} \boldsymbol{e} + \alpha_{in} \sum_{e \in \mathcal{S}_{in}} \boldsymbol{e}, \tag{8}$$

where the system parameters α_{jo} and α_{in}, respectively, are set the same as those in the LTIM. Notice that in the STIM, we do not include time decayed weights, as recent events are close to the recommendation time.

We next build the *user-event prospective rating matrix* $\boldsymbol{B} \in \mathbb{R}^{M \times N_{new}}$ for M users and N_{new} events as follows: An element b_{ij} is the prospective rating given by user u_i to a new event e_j, which is computed by

$$b_{ij} = \begin{cases} \dfrac{\boldsymbol{v}_i \cdot \boldsymbol{e}_j}{\|\boldsymbol{v}_i\| \|\boldsymbol{e}_j\|}, & \text{if } u_i \text{ registered } e_j, \\ 0, & \text{Otherwise.} \end{cases} \tag{9}$$

For a user u_i, we compute her prospective rating $r(u_i, e_j)$ to her unregistered new event e_j by

$$r(u_i, e_j) = \frac{\sum_{u_k \in \mathcal{U}_i} sim(u_i, u_k) b_{kj}}{\sum_{u_k \in \mathcal{U}_i} sim(u_i, u_k)}, \tag{10}$$

where u_k is a neighborhood user of u_i, $sim(u_i, u_k)$ the similarity between the two users computed by Eq. (7), and b_{kj} the prospective rating of user u_k given to event e_j. We then sort the unregistered new events for user u_i in a decreasing order of prospective rating and select the first N new events to compose her recommendation list.

IV. EXPERIMENT

A. EXPERIMENT DATASETS

We have crawled our experiment datasets from Douban Event for two main cities, Beijing and Shanghai, in China, ranging from May 1st, 2016 to May 1st, 2017. The Beijing dataset includes 16411 events, 193536 users, 260206 'join' user-event (U-E) pairs and 347705 'interested' U-E pairs. The Shanghai dataset includes 14928 events, 158236 users, 204736 'join' U-E pairs and 312323 'interested' U-E pairs. To respect the actual event timeline and ensure enough testing samples, we divide the two datasets into two parts: The training datasets are from May 1st, 2016 to March 1st, 2017, and the testing datasets are from March 1st, 2017 to May 1st, 2017. In both testing datasets, we select the users who have *joined* more than three events as the testing users and their joined events as the testing events. There are 3302 testing events and 3680 testing users in Beijing, and 3264 testing events and 3324 testing users in Shanghai. Table 1 summarizes the statistics of the two datasets.

TABLE 1. Statistics of dataset.

Beijing	User	Event	'Join' U-E Pair
	193536	16411	260206
	'Interested' U-E Pair	Testing user	Testing event
	347705	3680	3302
Shanghai	User	Event	'Join' U-E Pair
	158236	14928	204736
	'Interested' U-E Pair	Testing user	Testing event
	312323	3324	3264

B. EVALUATION METRICS

We adopt five commonly used performance metrics for recommendation evaluation: P@n (Precision at position n), MAP (Mean Average Precision), Recall, F1 and Coverage. For a user u_i ($i = 1, \ldots, M$) in the testing set, let L_i denote her recommendation list and N be the list length. Let $\mathcal{H}_i$ denote the set of events that the user u_i has registered as 'join', which are called her positive events.

Both P@n and MAP are used to measure the hit rate with taking top n position of positive events into consideration. P@n is defined as follows:

$$P@n = \frac{\sum_{i=1}^{M} \sum_{j=1}^{n} \mathbb{I}(L_i^{(j)} \in \mathcal{H}_i)}{M \times n}, \tag{11}$$

where $\mathbb{I}(\cdot)$ is an indicator function and $L_i^{(j)}$ is the jth event in the user u_i's recommendation list.

MAP is the mean of the *average precision* (AP) scores over all testing users, where AP is calculated by:

$$AP_i = \frac{\sum_{n=1}^{N} P@n \cdot \mathbb{I}(L_i^{(n)} \in \mathcal{H}_i)}{|\mathcal{H}_i|}, \tag{12}$$

where $L_i^{(n)}$ denotes the nth recommended event in the list L_i. $|\mathcal{H}_i|$ represents the number of events that had been actually attended by the u_i in the testing set. Thus, MAP is defined by

$$MAP = \frac{\sum_{u_i \in \mathcal{U}_{tst}} AP_i}{|\mathcal{U}_{tst}|}, \tag{13}$$

where $\mathcal{U}_{tst}$ denotes the set of testing users.

Recall reflects the proportion of events that users have actually attended in the top-n place. Take user u_i for example, her recall $R_i(L)$ is defined by

$$R_i(L) = \frac{d_i(L)}{|\mathcal{H}_i|} \tag{14}$$

where $d_i(L)$ indicates the number of u_i's attended events in the top-n places of the recommendation list L_i, and $|\mathcal{H}_i|$ the total number of u_i's attended events. The mean recall is obtained by averaging the individual recall over all users with at least one relevant event.

The F1 metric is used to evaluate the joint effectiveness of the Recall and Precision:

$$F1 = \frac{2PR}{P+R}, \tag{15}$$

where P and R are the Precision and Recall, respectively.

The Coverage evaluates the ability of a recommendation system to explore testing events and it is defined by:

$$Coverage = \frac{|L_1 \bigcap L_2 \bigcap \ldots \bigcap L_N|}{|\mathcal{E}_{tst}|}, \tag{16}$$

where $\mathcal{E}_{tst}$ is the set of testing events.

C. PARAMETER SETTING

In the proposed scheme, the parameters α_{jo} and α_{in} weight the users' choices to an event, and the parameter λ decides the speed of attenuation of past events to the current interest model. We experiment on the training datasets to choose their appropriate values. For α_{jo} and α_{in}, we use a grid search approach by repeating the experiments for different value steps: We first keep $\alpha_{jo} = 0$ and change α_{in} from 0.1 to 1. Fig.2 plots the experiment results against the values of α_{in}. We observe that both P@1 and P@3 perform better for small values of α_{in}, which is in accordance to our intuition that 'join' registration could be more important than 'interested' registration. Also the results show that $\alpha_{in} = 0.3$ is a suitable value. We next redo these experiments with $\alpha_{in} = 0.3$ and change the value of α_{jo} from 0.5 to 1. The results are drawn in Fig.3, where P@1 and P@3 achieve better performance for $\alpha_{jo} \geq 0.8$, yet other metrics have slight changes with different α_{jo} values. As such, we choose $\alpha_{jo} = 0.9$ and $\alpha_{in} = 0.3$ and normalize them to $\alpha_{jo} = 0.75$ and $\alpha_{in} = 0.25$ in our experiments. Fig.4 plots the experiment results where we change λ from 0.3 to 1.5 at a step of 0.1. We observe that if $0.5 \leq \lambda \leq 1$, P@1 gets the maximum value, that is, the attenuation speed is about half a month to one month. In addition, although P@3, recall and F1 get their maximum values when $\lambda = 1.3$, the performance only decreases slightly. As one month is also

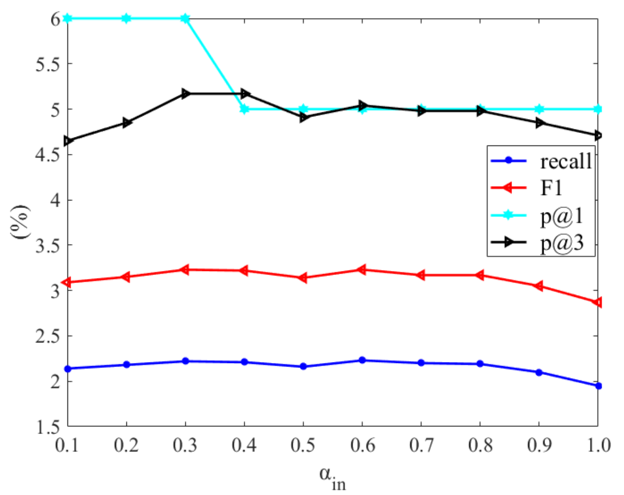


FIGURE 2. The recommendation performance with $\alpha_{jo} = 1$ and α_{in} changing from 0.1 to 1.

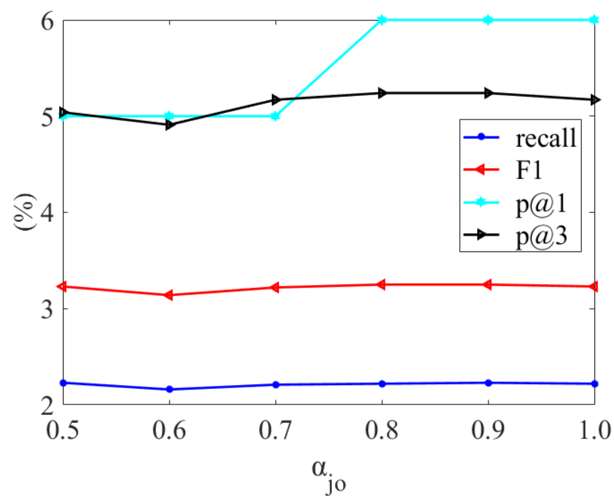


FIGURE 3. The recommendation performance with $\alpha_{in} = 0.3$ and α_{jo} changing from 0.5 to 1.

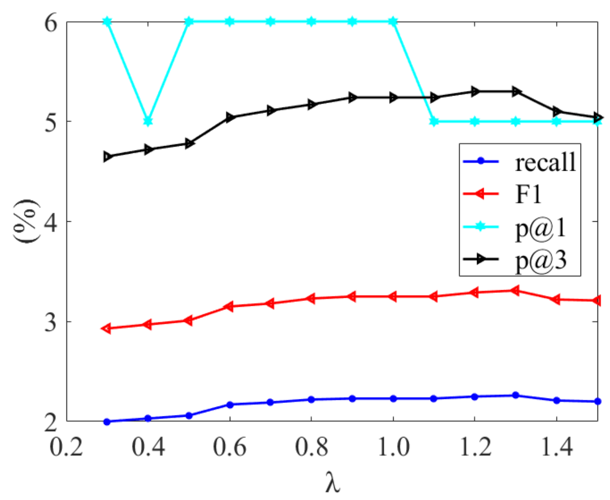


FIGURE 4. The recommendation performance with λ changing from 0.3 to 1.5.

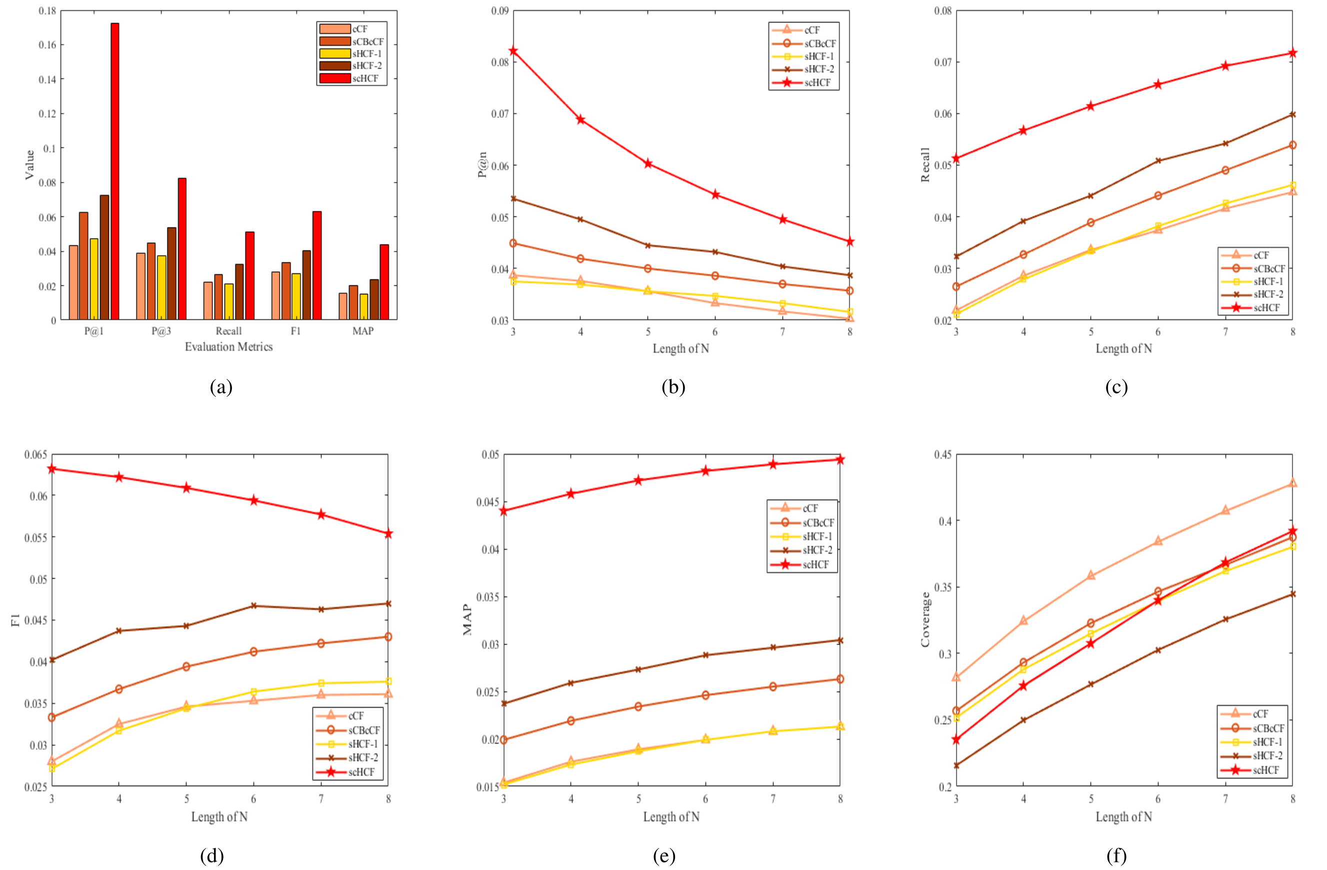


FIGURE 5. Beijing: Experiment results of P@n, Recall, F1, MAP and Coverage with different lengths of recommendation lists. (a) Top-3 result of Beijing. (b) *P@n* of Beijing. (c) *Recall* of Beijing. (d) *F1* of Beijing. (e) *MAP* of Beijing. (f) *Coverage* of Beijing.

with an easily understandable semantics, we choose $\lambda = 1$, i.e., one month, in our experiments.

D. COMPARISON SCHEMES

We compare the proposed scHCF scheme with the following representative peer schemes:

cCF: This a context-aware CF scheme. We first construct a zero-one matrix $\boldsymbol{R}$ according to users' registration to events and compute the context-aware similarity between two users u_i and u_j by

$$sim_{ctx}(u_i, u_j) = \frac{\sum_{e \in \mathcal{E}_i \cap \mathcal{E}_j} \frac{1}{\log(1+|\mathcal{U}_e|)}}{\|\boldsymbol{R}_i\| \|\boldsymbol{R}_j\|}, \quad (17)$$

where $\boldsymbol{R}_i$ is the ith row of $\boldsymbol{R}$, $\mathcal{E}_i$ the registered events by user u_i and $\mathcal{U}_e$ the number of registered users to event e. We argue that in general the more the participants in one event, the loose the social relation in between two participants in this event. So instead of using the simple dot product in the cosine distance, we include a participant-related penalty factor in the numerator. After selecting the similar users, the rest for recommending event for one user is with the same procedure as that in the classic CF scheme.

sCBcCF: This is a hybrid recommendation scheme consisting of semantic-enhanced CB and context-aware CF. Besides using Eq. (17), it also establishes the LTIM $\boldsymbol{u}$ for each user according to Eq. (1) and computes the semantic similarity $sim_{smt}(u_i, u_j)$ between two users according to

$$sim_{smt}(u_i, u_j) = \frac{\boldsymbol{u}_i \cdot \boldsymbol{u}_j}{\|\boldsymbol{u}_i\| \|\boldsymbol{u}_j\|}. \quad (18)$$

Note that Eq. (18) does not include context-aware event influence as that in Eq. (7). The user similarity is then obtained as the average of sim_{ctx} and sim_{smt} for neighborhood construction, and the rest is the same as that in cCF.

sHCF-1: This is a variant of the proposed scHCF. It also establishes a LTIM and STIM for each user based on the event semantic analysis. Yet its LTIM does not include a time attenuation factor to weight a past event, that is,

$$\boldsymbol{u}' = \alpha_{jo} \sum_{\boldsymbol{e} \in \mathcal{L}_{jo}} \boldsymbol{e} + \alpha_{in} \sum_{\boldsymbol{e} \in \mathcal{L}_{in}} \boldsymbol{e}. \quad (19)$$

Based on the LTIM $\boldsymbol{u}'$, it also constructs a user-event historical rating matrix $\boldsymbol{A}' \in \mathbb{R}^{M \times N_{hst}}$. For neighborhood construction, it does not include the event influence weight in user

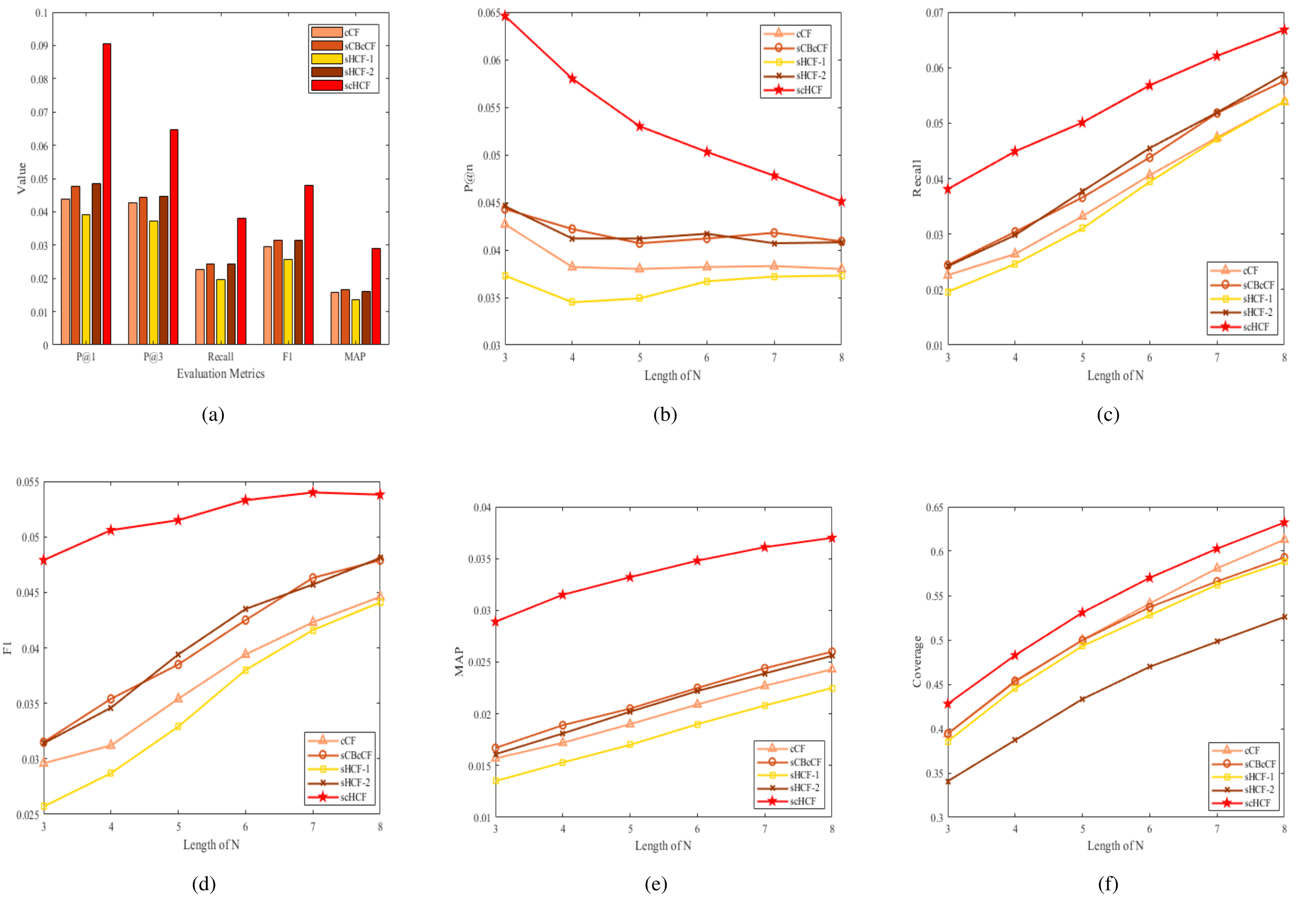


FIGURE 6. Shanghai: Experiment results of P@n, Recall, F1, MAP and Coverage with different lengths of recommendation lists. (a) Top-3 result of Shanghai. (b) *P@n* of Shanghai. (c) *Recall* of Shanghai. (d) *F1* of Shanghai. (e) *MAP* of Shanghai. (f) *Coverage* of Shanghai.

similarity computation, that is,

$$sim'(u_i, u_j) = \frac{A_i' A_j'^{\mathsf{T}}}{\|A_i'\| \|A_j'\|}. \tag{20}$$

The rest of sHCF-1 is the same as that in scHCF.

sHCF-2: This is also a variant of the proposed scHCF. Compared with scHCF, the only difference is that it does not include the event influence weight in the user similarity computation, that is, without including $\boldsymbol{C}$ in Eq. (7).

E. EXPERIMENT RESULTS

Figs. 5(a) and 6(a) present the experiment results of P@1, P@3, Recall, F1 and MAP for Beijing and Shanghai datasets, respectively, where we set the length of recommendation list to three, i.e., $N = 4$. It can be observed that the proposed scHCF outperforms all the other schemes in all evaluation metrics in both cities. The sHCF-1 scheme performs worse than the other schemes in almost all evaluation metrics except P@1 in Beijing, and it also has the worst performance in all evaluation metrics in Shanghai. Notice that except the cCF scheme, all the others exploit event semantics for establishing users' interest models. However, the sHCF-1 scheme does not consider the time attenuation factor in the interest model construction. This indicates that the users' interests normally change over time and the older historical event registration may not be well enough to represent users' up-to-date interests in the recommendation time.

Figs. 5(b)-5(e) and Figs. 6(b)-6(e) compare the schemes' performance for Beijing and Shanghai datasets, respectively, when setting the length of the recommendation list from three to eight. We notice that both cCF and sHCF-1 are the two second worst schemes in all performance metrics, except the coverage. This is not unexpected as that the cCF only exploits the social relations among users who had registered in the same events, yet without considering the event semantics for better describing users' interests. When we compare the sCBcCF scheme with the cCF and sHCF-1 scheme, we can find that it performs better than them. This is because that the sCBcCF scheme takes both of the event description semantics and user social relations into consideration. Furthermore, the sHCF-2 scheme achieves the second best performance among the comparison schemes, which also takes the event description semantics into consideration for establishing time attenuated interest model. These observations indicate that applying event description semantics to establish users' interest models can help to improve recommendation performance, and a time-attenuated interest model could be even better.

In Fig.5, we also observe that the scHCF scheme performs the best among all comparing schemes in terms of all performance metrics except the Coverage and in all recommendation list length. The improvement is due to that the scHCF scheme establishes two user interest models, namely, LTIM and STIM, for user neighborhood construction and event rating prediction, respectively. Furthermore, it applies a novel context-aware event influence weight for user neighborhood construction, which can help to better describe the social relations in between users registering a same event by including not only event participant numbers but also event description semantics. However, the scHCF scheme does not perform the best in terms of Coverage, which might be due to the diversity of Beijing events. When recommending events to users, only a part of events that have similar features would be recommended and some dissimilar events are ignored, leading to a low Coverage. In addition, the cCF scheme does not consider the user interest model from event semantics at all; its focus on history registration may help to recommend some dissimilar events, even with a small portion, to users. We note that although the Coverage of cCF is the best from event viewpoint, its recommendation precision is much low from individual user viewpoint.

In Fig.6, besides some similar observation as the performances in Fig.5, we observe that the proposed scheme scHCF outperforms the other peer schemes, even in the Coverage evaluation metric. In a short summary of our experiments, the performance of the proposed scHCF scheme outperforms the peer schemes in terms of almost all evaluation metrics. This first suggests that the event semantics are useful for establishing users' interest models, yet better in a time attenuated way. Furthermore, the context-aware event influence also plays an important part in recommendation.

V. CONCLUSION

In this paper, we have proposed the scHCF scheme which combines the social relationship with content information. Besides these two kinds of factors, we construct different user interest model for different tasks and take the temporal impact and context aware event influence into consideration. Experiments from a real EBSN, Douban Event, have validated its superiority over the peer schemes in terms of better recommendation results.

However, there are still some challenges for our research. In our experiments, we have not trained the parameters automatically, which costs much more time to find appropriate parameter values. Furthermore, we notice that when selecting neighborhood users for target user, the system need to traverse all users, which will cost a lot of time, if there are too many users. Therefore, in our future work, we shall investigate some approaches for parameter training and Matric Factorization(MF) to simplify the recommendation system and improve the recommendation performance.

REFERENCES

[1] G. Liao, Y. Zhao, S. Xie, and P. S. Yu, "An effective latent networks fusion based model for event recommendation in offline ephemeral social networks," in *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manage. (CIKM)*, 2013, pp. 1655–1660.

[2] H. Ding, C. Yu, G. Li, and Y. Liu, "Event participation recommendation in event-based social networks," in *Social Informatics—SocInfo* (Lecture Notes in Computer Science), vol 10046, E. Spiro and Y. Y. Ahn, Eds. Cham, Switzerland: Springer, 2016.

[3] X. Liu, Q. He, Y. Tian, W.-C. Lee, J. McPherson, and J. Han, "Event-based social networks: Linking the online and offline social worlds," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 1032–1040.

[4] F. Hao, S. Li, G. Min, H. C. Kim, S. S. Yau, and L. T. Yang, "An efficient approach to generating location-sensitive recommendations in ad-hoc social network environments," *IEEE Trans. Services Comput.*, vol. 8, no. 3, pp. 520–533, May 2015.

[5] W. Zhang, J. Wang, and W. Feng, "Combining latent factor model with location features for event-based group recommendation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 910–918.

[6] T. J. Ogundele, C.-Y. Chow, and J.-D. Zhang, "Eventrec: Personalized event recommendations for smart event-based social networks," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, May 2017, pp. 1–8.

[7] S. Liu, B. Wang, and M. Xu, "Serge: Successive event recommendation based on graph entropy for event-based social networks," *IEEE Access*, vol. 6, pp. 3020–3030, 2018.

[8] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Phys. Rep.*, vol. 519, no. 1, pp. 1–49, 2012.

[9] F. Zhang, "A personalized time-sequence-based book recommendation algorithm for digital libraries," *IEEE Access*, vol. 4, pp. 2714–2720, 2016.

[10] F. Xia, Z. Chen, W. Wang, J. Li, and L. T. Yang, "MVCWalker: Random walk-based most valuable collaborators recommendation exploiting academic factors," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 364–375, Sep. 2014.

[11] L. Hu, Y. Wang, Z. Xie, and F. Wang, "Semantic preference-based personalized recommendation on heterogeneous information network," *IEEE Access*, vol. 5, pp. 19773–19781, 2017.

[12] Y.-C. Sun and C. C. Chen, "A novel social event recommendation method based on social and collaborative friendships," in *Proc. Int. Conf. Social Inform.* Springer, 2013, pp. 109–118.

[13] L. Gao, J. Wu, Z. Qiao, C. Zhou, H. Yang, and Y. Hu, "Collaborative social group influence for event recommendation," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, 2016, pp. 1941–1944.

[14] Y. Gu, J. Song, W. Liu, L. Zou, and Y. Yao, "Context aware matrix factorization for event recommendation in event-based social networks," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Oct. 2016, pp. 248–255.

[15] T. J. Ogundele, C.-Y. Chow, and J.-D. Zhang, "SoCaST: Personalized event recommendations for event-based social networks: A multi-criteria decision making approach," *IEEE Access*, vol. 6, pp. 27579–27592, 2018.

[16] Z. Wang, Y. Zhang, Y. Li, Q. Wang, and F. Xia, "Exploiting social influence for context-aware event recommendation in event-based social networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[17] Z. Wang, Y. Zhang, H. Chen, Z. Li, and F. Xia, "Deep user modeling for content-based event recommendation in event-based social networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 1304–1312.

[18] L. Tang and E. Y. Liu, "Joint user-entity representation learning for event recommendation in social network," in *Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE)*, Apr. 2017, pp. 271–280.

[19] Z. Qiao, P. Zhang, Y. Cao, C. Zhou, L. Guo, and B. Fang, "Combining heterogenous social and geographical information for event recommendation," in *Proc. AAAI*, vol. 14, 2014, pp. 145–151.

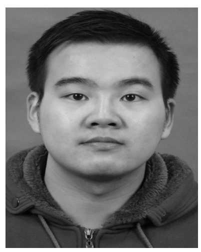
[20] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web* (Lecture Notes in Computer Science), vol 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer, 2007.

[21] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48–57, May 2014.

[22] A. Q. Macedo, L. B. Marinho, and R. L. T. Santos, "Context-aware event recommendation in event-based social networks," in *Proc. 9th ACM Conf. Recommender Syst. (RecSys)*, 2015, pp. 123–130.

[23] C.-K. Hsieh, L. Yang, H. Wei, M. Naaman, and D. Estrin, ''Immersive recommendation: News and event recommendations using personal digital traces,'' in *Proc. 25th ACM Int. Conf. World Wide Web*. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016, pp. 51–62
[24] M. J. Pazzani and D. Billsus, ''Content-based recommendation systems,'' in *The Adaptive Web* (Lecture Notes in Computer Science), vol 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer, 2007.
[25] I. Cantador, A. Bellogín, and D. Vallet, ''Content-based recommendation in social tagging systems,'' in *Proc. 4th ACM Conf. Recommender Syst.*, 2010, pp. 237–240.
[26] S. Puglisi, J. Parra-Arnau, J. Fornè, and D. Rebollo-Monedero, ''On content-based recommendation and user privacy in social-tagging systems,'' *Comput. Standards Interfaces*, vol. 41, pp. 17–27, Sep. 2015.
[27] T. Rutkowski, J. Romanowski, P. Woldan, P. Staszewski, R. Nielek, and L. Rutkowski, ''A content-based recommendation system using neuro-fuzzy approach,'' in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2018, pp. 1–8.
[28] F. Narducci *et al.*, ''Concept-based item representations for a cross-lingual content-based recommendation process,'' *Inf. Sci.*, vol. 374, pp. 15–31, Dec. 2016.
[29] S. Debnath, N. Ganguly, and P. Mitra, ''Feature weighting in content based recommendation system using social network analysis,'' in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 1041–1042
[30] M. Soleymani, A. Aljanaki, F. Wiering, and R. C. Veltkamp, ''Content-based music recommendation using underlying music preference structure,'' in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun./Jul. 2015, pp. 1–6.
[31] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, P. Piazzolla, and M. Quadrana, ''Content-based video recommendation system based on stylistic visual features,'' *J. Data Semantics*, vol. 5, no. 2, pp. 99–113, 2016.
[32] F. Zhao, Y. Zhu, H. Jin, and L. T. Yang, ''A personalized hashtag recommendation approach using LDA-based topic model in microblog environment,'' *Future Gener. Comput. Syst.*, vol. 65, pp. 196–206, Dec. 2016.
[33] R. Krestel, P. Fankhauser, and W. Nejdl, ''Latent dirichlet allocation for tag recommendation,'' in *Proc. 3rd ACM Conf. Recommender Syst.*, 2009, pp. 61–68.
[34] N. Lee, E. Kim, and O. Kwon, ''Combining TF-IDF and LDA to generate flexible communication for recommendation services by a humanoid robot,'' *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 5043–5058, 2018.
[35] Z. Wang, J. Liao, Q. Cao, H. Qi, and Z. Wang, ''Friendbook: A semantic-based friend recommendation system for social networks,'' *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 538–551, Mar. 2015.
[36] E. Minkov, B. Charrow, J. Ledlie, S. Teller, and T. Jaakkola, ''Collaborative future event recommendation,'' in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2010, pp. 819–828.
[37] C.-L. Liao and S.-J. Lee, ''A clustering based approach to improving the efficiency of collaborative filtering recommendation,'' *Electron. Commerce Res. Appl.*, vol. 18, pp. 1–9, Jul./Aug. 2016.
[38] J. Zhang, Q. Peng, S. Sun, and C. Liu, ''Collaborative filtering recommendation algorithm based on user preference derived from item domain features,'' *Phys. A, Stat. Mech. Appl.*, vol. 396, pp. 66–76, Feb. 2014.
[39] C. Luo, W. Pang, Z. Wang, and C. Lin, ''Hete-CF: Social-based collaborative filtering recommendation using heterogeneous relations,'' in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2014, pp. 917–922.
[40] X. Yang, Y. Guo, Y. Liu, and H. Steck, ''A survey of collaborative filtering based social recommender systems,'' *Comput. Commun.*, vol. 41, pp. 1–10, Mar. 2014.
[41] J. Liu, M. Tang, Z. Zheng, X. Liu, and S. Lyu, ''Location-aware and personalized collaborative filtering for Web service recommendation,'' *IEEE Trans. Services Comput.*, vol. 9, no. 5, pp. 686–699, Sep./Oct. 2016.
[42] H. Khrouf and R. Troncy, ''Hybrid event recommendation using linked data and user diversity,'' in *Proc. 7th ACM Conf. Recommender Syst.*, 2013, pp. 185–192.
[43] L. Hu, G. Song, Z. Xie, and K. Zhao, ''Personalized recommendation algorithm based on preference features,'' *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 293–299, Jun. 2014.
[44] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, ''Context-based collaborative filtering for citation recommendation,'' *IEEE Access*, vol. 3, no. 1, 2015.
[45] O. Barkan and N. Koenigstein, ''ITEM2VEC: Neural item embedding for collaborative filtering,'' in *Proc. IEEE 26th Int. Mach. Learn. Signal Process. (MLSP)*, Sep. 2016, pp. 1–6.
[46] M. Al-Hassan, H. Lu, and J. Lu, ''A semantic enhanced hybrid recommendation approach: A case study of e-government tourism service recommendation system,'' *Decis. Support Syst.*, vol. 72, pp. 97–109, Apr. 2015.
[47] O. Kaššák, M. Kompan, and M. Bieliková, ''Personalized hybrid recommendation for group of users: Top-N multimedia recommender,'' *Inf. Process. Manage.*, vol. 52, no. 3, pp. 459–477, 2016.
[48] W. Chen, Z. Niu, X. Zhao, and Y. Li, ''A hybrid recommendation algorithm adapted in e-learning environments,'' *World Wide Web*, vol. 17, no. 2, pp. 271–284, 2014.
[49] M. Nilashi, O. B. Ibrahim, and N. Ithnin, ''Hybrid recommendation approaches for multi-criteria collaborative filtering,'' *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3879–3900, 2014.
[50] W. Jian, J. He, K. Chen, Y. Zhou, and Z. Tang, ''Collaborative filtering and deep learning based recommendation system for cold start items,'' *Expert Syst. Appl.*, vol. 69, pp. 29–39, Mar. 2017,
[51] D. M. Blei, A. Y. Ng, and M. I. Jordan, ''Latent Dirichlet allocation,'' *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

MINGHUA XU received the B.A. degree from the School of Journalism and Information Communication, Huazhong University of Science and Technology (HUST), in 2002, and the M.A and Ph.D. degrees from Social Science Department, National University of Singapore, in 2005 and 2010, respectively. She is currently an Associate Professor with the School of Journalism and Information Communication, HUST. She has published over 30 peer-reviewed conference and journal papers. Her research interests include the theory and practice of information diffusion in social networks, recommendation systems and applications, and information communication theories.



SHENGHAO LIU received the B.S. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2016, where he is currently pursuing the Ph.D. degree. His research interests include recommendation systems and applications.