

# Event Recommendation in Social Networks with Linked Data Enablement

Yinuo Zhang<sup>1</sup>, Hao Wu<sup>1</sup>, Vikram Sorathia<sup>2</sup> and Viktor K. Prasanna<sup>2</sup>

<sup>1</sup>*Department of Computer Science, University of Southern California, Los Angeles, CA, USA*

<sup>2</sup>*Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA*  
{yinuozha, hwu732, vsorathi, prasanna}@usc.edu

Keywords: Recommendation, Linked Data, Social Networks

Abstract: In recent years, social networking services have gained phenomenal popularity. They allow us to explore the world and share our findings in a convenient way. Event is a critical component in social networks. A user can create, share or join different events in their social circle. In this paper, we investigate the problem of event recommendation. We propose recommendation methods based on the similarity of an event's content and a user's interests in terms of topics. Specifically, we use Latent Dirichlet Allocation (LDA) to generate a topic distribution over each event and user. We also consider friend relationship and attendance history to increase recommendation accuracy. Moreover, we enable linked data as our data sources to collect contextual information related to events and users, and build an enhanced profile for them. As reliable resource, linked data is used to find structured knowledge and linkages among different knowledge. Finally, we conduct comprehensive experiments on various datasets in both academic community and popular social networking service.

## 1 Introduction

People live socially and keep connected in various ways. Social event is one of the essential components for networking. Celebrations, inaugurations, commencements, fund raising are all social events that serve for different purposes. People tend to refer to their friends or media for information of upcoming events. Nowadays, its main channel has shifted from newspaper, bulletin board and television to the internet, especially popular online social networks. For instance, users like to use the interactive interface of Facebook Events<sup>1</sup> to create events, invite friends and accept invitations. Significant portion of those events on Facebook are about parties and any other informal celebrations. Eventseer<sup>2</sup> represents another example that serve as a news forum for notifications of academic events, which are mainly conferences, workshops and seminars in various disciplines. Last.fm<sup>3</sup> is another example which contains various event information related to music, such as festivals, singer or band performance and fan party.

The offer of events is enormous online and there are usually many co-occurring activities even at the same location. Consequently, people find it difficult to keep track of the events that are of interest to them or worth spending time. To address this problem, recommendation models (Cornelis et al., 2005; Kayaalp et al., 2009; Klamma et al., 2009; Konstantas et al., 2009; Coppens et al., 2012; Minkov et al., 2010; Li et al., 2010; Daly and Geyer, 2011; De Pessemier et al., 2011) are designed to select relevant events that are most likely of interest to each individual user. A general approach of event recommendation is content based (Cornelis et al., 2005; De Pessemier et al., 2011), which aims to capture descriptive features of an event such as location, time and theme to match user interests. To characterize user interests, content-based approach leverage the past event attendance records of a user, as well as the user feedback such as the rating of events. The keywords that characterize user interests are then used as query to search on the future events for recommendation. One major problem of the keyword-based search is that it cannot fully capture the rich semantics of event content and how it matches user interests. Moreover, the success of

---

<sup>1</sup><http://www.facebook.com/events/>

<sup>2</sup><http://eventseer.net/>

<sup>3</sup><http://www.last.fm/events>

content-based approach largely depend on the user history records and user feedback. In this sense, the approach may suffer from data sparsity problem (Minkov et al., 2010) when dealing with new users who have inadequate history records and the user feedback of events is scarce.

In this paper, we adopt topic modeling method to bridge the semantic gap between events and user preferences. People tend to attend events with themes that match their personal interests. Therefore, the users' preferences of future events rely on underlying topics rather than word descriptions. In particular, Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is used to discover the underlying latent topics, in order to find events that best match user preferences in semantics. The influence of a user's connections is also considered. We look at the event attendance history of one's friends to find events that the user may be interested in. It is based on the intuition that friends with common interests are more likely to attend the same events. In an integrated manner, we learn a model to rank the future events by using the user attendance history for personalized recommendation. We also present a hybrid model combining the three topic modeling based approaches. We conduct comprehensive experiments on various datasets of academic community and popular social networks. The results show that our methods consistently outperform the baseline algorithm. The main contributions of our work can be summarized as:

- We present three event recommendation approaches based on topic modeling of event content and user profile.
- We propose a hybrid learning framework for recommendation which integrates topic similarity, user connection and attendance history.
- We enable linked data in constructing event and user profiles, as well as event history records.

The rest of the paper is organized as follows. We continue with discussion of related work in Section 2. We present our methodology in details In Section 3. We show a brief overview of linked data and the experiment setup in Section 4. Experimental results are presented in Section 5. Finally, we conclude our work and briefly discuss future work in Section 6.

## 2 Related Work

Event recommendation as a means of personalizing event information acquisition in social networks, has attracted increasing research attention in recent years. Most existing methods borrowed the ideas from information recommendation of other domains such as e-commerce (Linden et al., 2003), book (Guan et al., 2009), music (Chen and Chen, 2001) and photo sharing (Sigurbjörnsson and Van Zwol, 2008) websites. Content-based and collaborative filtering approaches which based on two mechanisms are employed for recommending future events. Daly and Geyer (Daly and Geyer, 2011) consider location and social information to filter events for recommendation. Cornelis et al. (Cornelis et al., 2005) propose an hybrid conceptual approach which leverages the merits of content-based and collaborative filtering. The approach recommends future events if they are similar to past ones that similar users have liked, which is an extension of Perny and Zucker's work (Perny and Zucker, 1999). However, the approach is only conceptual and the authors do not provide validation of the approach on experiments. Minkov et al. (Minkov et al., 2010) present a collaborative method called LowRank, which decomposes user parameters into shared and individual components for event recommendation in the setting of academic seminars. The method uses topic modeling to represent the past attendance activities of individual users and descriptions of the events as topic features. The experimental results demonstrate its superiority over basic content-based recommendation. However, the approach is only limited to one domain and it requests user feedbacks. In this paper, we aim to design a general and user feedback independent algorithms which can be applied to different settings of event recommendation.

Previous work also focuses on the aggregation, enrichment as well as personalized distribution of events from various web sources. Kayaalp et al. (Kayaalp et al., 2009) examine a social activity recommendation system for concert event. Concerts information is harvested from web sources using web services and scrapers. Recommendations are generated based on various features including user profiles, concert ratings, a social network structure, and activity properties. They build a complete event tracking system which is open to the integration of heterogeneous information resources. De Pessemerier et al. (De Pessemerier et al., 2011; Coppens et al., 2012) focus on representation of events as structured data. They build a highly-scalable event recommendation

platform for cultural events, which are collected and published as Linked Open Data with an RDF/OWL representation using the EventsML-G2 standard. This allows the incorporation of content-based filters for event distribution. However, those explorations do not capture the underlying topic of different events or activities quite well, and topics usually can provide a better description of an event in order to match user preferences. Also, their methods request user feedbacks which can affect flexibility. Moreover, the datasets they used are isolated which are in the form of XML. Sometimes the linkages among different datasets are quite critical for recommendation since one data source cannot contribute enough information. Different from their work, our methods explore both content and underlying topics to build event profiles. Also linked data is used as data source which is more flexible and inter-linkable. Furthermore, our methods only request the attendance history from users with any other feedbacks owing to the variety of linked data.

Social influence has great impact on a user’s decisions and actions. A user is likely to follow the actions of his/her friends with which they share common interest. This idea is embedded into various recommendation systems to improve the performance, including item recommendation in music sharing website (Konstas et al., 2009), product review rating prediction (Au Yeung and Iwata, 2011), interest targeting and friendship prediction (Yang et al., 2011). Based on the same idea, Klamma et al. (Klamma et al., 2009) explore academic events such as conferences and workshops and identify ones that might be of interest to individual researchers and can motivate cooperation between them. They propose a similarity measure based on the attendance records to generate recommendation. However, the content information such as conference name and description are not considered. In our work, we consider variety of datasets to evaluate our approaches. We also utilize the friendship which is the most influential relationship in social networks for event recommendation.

### 3 Methodology

In this section, we go into full exploration of event recommendation modeling. We begin with setting up the problem. Suppose there are  $m$  users and  $n$  events in the network, the goal is to find relevant events that are of interest to a user  $u_i$ . We use similarity metrics to measure how likely a user  $u_i$  will attend an event  $e_j$ . The information we can use is the set of friends

Table 1: Notations used in this paper

Notation	Description
$u_i, i = 1 \dots m$	the $i$ th user in a set of $m$ users
$e_j, j = 1 \dots n$	the $j$ th event in a set of $n$ events
$F(u_i)$	the set of user $u_i$ ’s friends
$A(e_j)$	the set of users who have attended $e_j$

that  $u_i$  has, which is denoted by  $F(u_i)$ , and  $A(e_j)$ , the set of users who have attended  $e_j$ . We also summarize the notations used in this paper in Table 1.

Each event or user profile is usually represented as a text document with word descriptions. The ways of extracting word descriptions vary in different contexts. In general, meaningful properties such as title, description and location are extracted for events, while words from interests, description and activity history are used for users. To find events with coherent semantics that match user interests, we adopt topic modeling to uncover the underlying topics of each event, and user interests as well. We generate topic proportion of each document expressed by those words using Latent Dirichlet Allocation (LDA) (Blei et al., 2003). In LDA, a document is considered as a mixture of latent topics, and each word observation in the document is sampled from a multinomial distribution (the word mixture for a topic). Each topic is drawn from a multinomial distribution generated using the Dirichlet prior. LDA can capture the underlying structure of a document and reveal the latent topics.

We first present three event recommendation approaches based on semantic similarity, relationships between users, and attending history. In the first approach, we simply calculate the similarity between topic distributions over an event and a user profile, and the most similar events are recommended to corresponding users. In the second approach, friend relationships are considered for recommendation. The intuition is that users with same interests have large chance to attend same events. In the third approach, event attendance history is used to build a classifier for recommendation. Logistic regression is adopted in the classification phase. Finally, we present an hybrid approach that combine the above three approaches. The hybrid approach uses weighted sum for calculating the similarity between an event and a user. Next, we elaborate each method.

#### 3.1 Similarity Based Approach (SBA)

The basic idea for this approach is to capture the semantic similarity between a user and an event. Based

on the topic distribution similarity, those events with the highest similarity to a specific user are recommended.

Specifically, we generate topic distribution for each document using LDA. The topic distribution is in the form of a normalized vector denoted as  $\vec{\theta}$ . In order to find the events that are of interest to a specific user, the similarity between  $\vec{\theta}$  of an event and a user is calculated. To compute the similarity between two vectors, we adopt cosine similarity (Equation 1) in this paper for its simplicity, although various alternatives can be used.

$$S_1(u_i, e_j) = \cos(\vec{\theta}_{u_i}, \vec{\theta}_{e_j}) = \frac{\vec{\theta}_{u_i} \cdot \vec{\theta}_{e_j}}{\|\vec{\theta}_{u_i}\| \|\vec{\theta}_{e_j}\|}, \quad (1)$$

In Equation 1,  $S_1(u_i, e_j)$  is the recommendation score of event  $e_j$  for user  $u_i$ . Here we use cosine similarity between the topic distribution vectors of user  $u_i$  and event  $e_j$ . Once the score  $S_1$  is calculated, all the events w.r.t. a user are ranked in descending order according to the scores. In practice, only top- $k$  events are returned for recommendation, where  $k$  can be pre-defined as a query parameter (e.g., an option for user to choose  $k$  value).

### 3.2 Relationship Based Approach (RBA)

In this approach, we consider social influence that may have impact on users' attending an event. Users usually follow their friends to attend an event because of common interests or just for networking. We hence recommend a user with the events attended by his/her "friends". "Friendship" may refer to different relations in different contexts. For example, it is simply the friend relationship in general social networks such as Facebook. And it refers to co-authorship in academic social networks. In order to quantify the degree of sharing same interests, the similarity between the topic distributions of two users  $u_i$  and  $u_{i'}$  is calculated (Equation 2).

$$S_{uu}(u_i, u_{i'}) = \cos(\vec{\theta}_{u_i}, \vec{\theta}_{u_{i'}}) = \frac{\vec{\theta}_{u_i} \cdot \vec{\theta}_{u_{i'}}}{\|\vec{\theta}_{u_i}\| \|\vec{\theta}_{u_{i'}}\|} \quad (2)$$

Based on the similarity, the recommendation score of event  $e_j$  to user  $u_i$  is calculated as

$$S_2(u_i, e_j) = \frac{\sum_{u_k \in F(u_i) \cap A(e_j)} S_{uu}(u_i, u_k)}{|F(u_i) \cap A(e_j)|}, \quad (3)$$

where  $F(u_i) \cap A(e_j)$  represent user  $u_i$ 's friends who attend event  $e_j$ . Similar to the first approach, all events  $u_i$ 's friends attend are ranked in descending order of the score  $S_2$ , and only top- $k$  results are returned.

### 3.3 History Based Approach (HBA)

In this third approach, we consider recommendation as a classification problem based on event attendance history of each user. We train a logistic regression model for each user using the topic distributions of past attend events. As shown in Equation 4, the output of logistic function  $f_{u_i}$  on a future event is used as the recommendation score  $S_3$ .

$$S_3(u_i, e_j) = f_{u_i}(e_j) = \frac{1}{1 + e^{-z}} \quad (4)$$

where

$$z = \beta_0 + \beta_1 \theta_{e_j}^{(1)} + \dots + \beta_k \theta_{e_j}^{(k)} \quad (5)$$

In Equation 5,  $k$  is the number of topics,  $\theta_{e_j}$  represents the topic distribution vector for event  $e_j$ ,  $\theta_{e_j}^{(t)}$  is the value for topic  $t$  in the vector, and  $\vec{\beta} = [\beta_0, \beta_1, \dots, \beta_k]^T$  are the parameters for the logistic regression model of a specific user. Finally, top- $k$  results are recommended to user  $u_i$  based on the value of  $S_3(u_i, e_j)$  which is between 0 to 1.

### 3.4 A Hybrid Approach (SRH)

Each of the above three methods has its own pros and cons. When users in a social network are well connected and has strong ties between each other, RBA is favored in recommendation. When past event attendance history is adequate, HBA is better applicable. If neither conditions are true but the social network can provide rich user and event profiles, SBA may work best. In order to provide a satisfying recommendation in different social networks, we propose a hybrid approach which integrate all three methods with different weights. A hybrid score  $S(u_i, e_j)$  is generated as shown in Equation 6.

$$S(u_i, e_j) = \omega_1 S_1(u_i, e_j) + \omega_2 S_2(u_i, e_j) + \omega_3 S_3(u_i, e_j) \quad (6)$$

where  $\vec{\omega} = [\omega_1, \omega_2, \omega_3]^T$  are the weights for the three approaches proposed in previous sections. In order to set  $\vec{\omega}$  for different social networks, 10-fold cross-validation can be used to decide  $\vec{\omega}$ . Specifically, the dataset can be partitioned into 10 equal size subsets. One subset is used as the validation data and the other 9 subsets are used as training data for learning the weights. The process is repeated 10 times with each subset as the validation data. In the experiments, data from different social networks are used to illustrate how the weights are affected.

The base of all four approaches is topic modeling, and the results of topic modeling highly depend on the amount of document profiles and their keywords.

In next section, we will introduce linked data to discover new events and enrich the profile for users and events.

## 4 Linked Data Enablement

Using latent topics modeling techniques such as LDA, implicit semantics of documents (i.e., users and events) are extracted from the raw text in order to build a better recommender. Another way to improve the performance and accuracy for recommendation is to use semantic web techniques. Specifically, comprehensive ontologies and semantic queries are the common approaches. However, for domain like event recommendation, it is not possible to compute everything in a single ontology. It is also not efficient to store all related knowledge in one place. To address this issue, linked data provides a good alternative. Nowadays, the amount of information increases drastically on the web. Linked data is a type of structured data which are interlinked. It covers almost every subject(i.e., Geographic information, social networks, publications et al.) on the web. Linked data is also easy to be retrieved so that it can be used as a part of knowledge base.

For event recommendation domain, EventSeer2RDF<sup>4</sup> represents linked data version of the information on Eventseer.net. This enables easy access to the information related to variety of academic events including conferences and workshops using SPARQL queries. Most importantly, it not only stores past events records, but also provides future events information which can be recommended. Moreover, all academic profiles for users can also be retrieved. Besides eventseer, DBLP is also a good resource for academic information. Different from eventseer which provides event description in the form of “call for papers”, DBLP stores the information about past publications and co-authorship for papers. Fortunately, there is a linked data version of DBLP<sup>5</sup> on the semantic web. All publication records related to a author can be retrieved through its SPARQL endpoint.

Figure 1 shows the partial graph structure of an example “event”(i.e. event\_16594) in eventseer linked data. Information in linked data is represented as rdf triples  $\langle \text{subject}, \text{predicate}, \text{object} \rangle$ . In this example, four triples exist in the partial rdf graph. As can be seen, predicates “based\_near” and “dtstart” show location and starting time for that

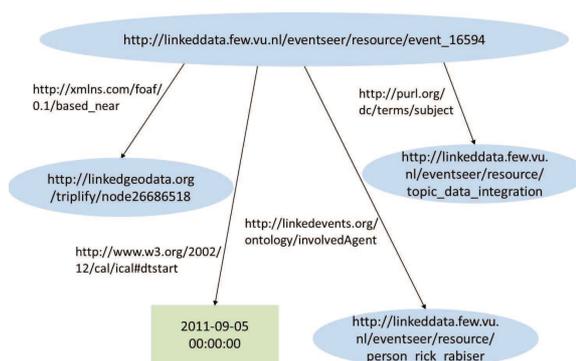


Figure 1: Event Representation in Linked Eventseer Data

event. “Persons” related to that event can be extracted from “involvedAgent” predicate. Moreover, predicate “subject” provides the related “topics” of that event. In this example, person Rick Rabiser is involved in event 16594 while the event is related to topic data integration.

Similar to “event”, “person” also has several properties. As shown in Figure 2, “Topic\_interest” provides the topics related to that person, while “knows” lists all persons who share the same interests with that person. In our context, “knows” is a way to identify friendship relations among academic persons.

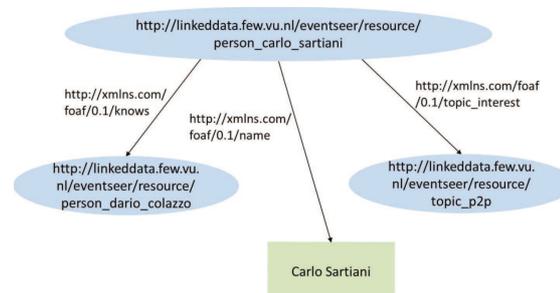


Figure 2: Person Representation in Linked Eventseer Data

Another example is enrich user profile using DBLP linked data. The detailed publications for a specific person can be found through “creator” property. All meaningful keywords in his/her publication titles then contribute to his/her profile. In next section, we will evaluate our methods on academic event recommendation with linked data enablement.

## 5 Experiments and Results

In this section, we conduct comprehensive experiments to compare the four methods proposed in Section 3. Specifically, two sets of data are used. The first dataset is for academic event recommendation.

<sup>4</sup><http://linkeddata.few.vu.nl/eventseer>

<sup>5</sup><http://www4.wiwiw.fu-berlin.de/dblp>

Open linked version of Eventseer and DBLP data are the sources for academic events and users. In total, profiles consisting of 10020 events and 26508 user are generated. To our best knowledge, we are the first to use open linked data for dynamic event recommendation. The second dataset comes from the most popular social network - Facebook. The reason we use Facebook is that it has explicit representation of social events which does not appear in other social networks such as Twitter or LinkedIn. In total, 1088 users and 4040 events are crawled through Facebook Graph API Explorer<sup>6</sup> to recommend social network events for Facebook users. In Section 5.1 and 5.2, the results on two datasets are presented sequentially.

## 5.1 Academic Event Recommendation

Our proposed methods are firstly evaluated on academic event dataset. Each event to recommend can be a future conference or a future workshop. Eventseer is such a resource that provides detailed information of future events and prospective attendees. In order to generate latent topic models on events and users, each of them is viewed as a document for LDA process. Topic distribution over document is considered as the feature space. In our experiments, linked eventseer data is used to extract keywords for each event and user. It is also the resource to find friendship relationship and participation information. Moreover, linked DBLP data provides more keywords for users since it has all publication records. Mean Average Precision(MAP) (Manning et al., 2008) is the metrics for evaluating four proposed methods. Average Precision for each user is defined as

$$AP = \frac{1}{n} \sum_{i=1}^n prec(k_i) \quad (7)$$

where  $prec(k_i)$  is the precision at rank  $k$ . It is defined as the number of correct records up to rank  $k$ , then divided by  $k$ . And  $n$  is the number of correct answers, while  $k_i$  represents the rank of each correct answer. For example, given a ranked list in which 1,3,4 are correct answers while 2,5 are not. The average precision for this list is  $(1 + \frac{2}{3} + \frac{3}{4})$ .

According to the definition of average precision, the value is highly related to the query context. For example, the worst case and random result for recommending  $N$  records with only one correct are  $\frac{1}{N}$  and the inverse of harmonic mean of  $N$ . In order to be consistent, a fixed number of records (i.e., 20) are returned for different sets of experiments. Recall is not used in the experimental evaluation because the

dataset does not contain all the attended events for each user. As a result, the absolute recall value cannot be calculated.

### 5.1.1 SBA

Each event and user is represented as a document. Keywords for events are extracted from EventSeer2RDF, a linked data repository for eventseer. Keywords for users are extracted from both EventSeer2RDF and D2R DBLP Bibliography Database. A topic distribution for each user and event is calculated. Based on the distribution, cosine similarity between each event and user pair is calculated. In this experiment, two sets of data are used. The first set consists of 20 events and 140 users, while the second set consists of 5000 events and 26508 persons. LDA is processed on both datasets separately. For the first set, all events are returned in the recommendation results. For the second set, the similarities between the topic distributions over the same set of events and users as in the first set are computed. All 20 events are ranked for each user in descending order of the similarity. Figure 3 shows the performance of SBA on these two datasets. Each point is the averaged MAP over 140 users. We also vary the parameters for topic modeling. In detail, the number of latent topics is set as 6 values (25, 50, 75, 100, 125, 150) and the number of Gibbs sampling iterations is empirically set as 500.

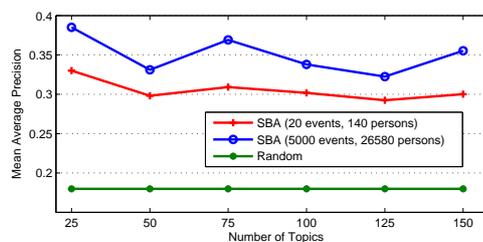


Figure 3: MAP of SBA

As observed from Figure 3, SBA outperforms random method on both datasets. Also SBA using large dataset has higher precision than that with small one. It is because large dataset has more “documents” and keywords for LDA to process. As a result, LDA has sufficient training set to provide a good topic model on each user and event. Another observation is that MAP does not differ much as the number of topics for LDA varies. The performance of topic modeling is relatively stable as the number of topics increase from 25 to 150.

<sup>6</sup><http://developers.facebook.com/tools/explorer>

### 5.1.2 RBA

Friendship also plays an important role for recommendation. Now we investigate how it affects the precision of recommending academic events. The second data set generated in SBA is used for topic modeling. Top 20 events are returned for 140 users based on their RBA scores.

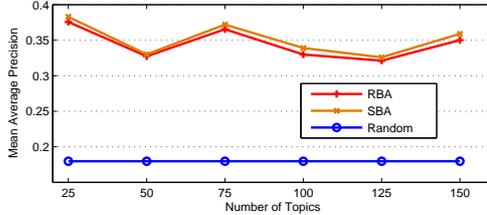


Figure 4: MAP of RBA

Figure 4 shows that RBA always has higher precision than the random method regardless of the number of topics for LDA process. However, SBA performs slightly better than RBA under the same setting. This can be explained by the property of the academic dataset. The friendship relations in eventseer are not as dense as that in the tradition social networks. As a result, friendship is not a good indicator for recommending events under this context.

### 5.1.3 HBA

Logistic regression is adopted in the third method. Specifically, recommendation is considered as a classification problem. The feature space is the topic distribution over each event. The whole data set (10020 events and 26508 users) with 25 topics is used in LDA process. Figure 5 shows the performance of HBA compared with random method. The number of training events varies from 5000 to 10000. The test set is a fixed set of 20 events. They are ranked in the descending order of HBA scores. MAP is also used to measure the precision of both approaches.

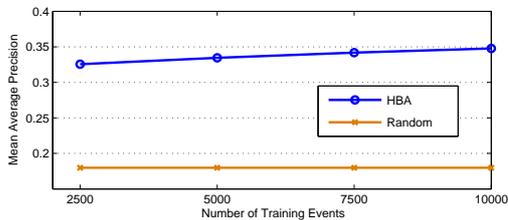


Figure 5: MAP of HBA

As can be seen in Figure 5, HBA always outperforms the random method under different numbers of training event sets. In detail, HBA is twice more precise than the random method as the number of training

Table 2: MAP of all methods

Method	Random	SBA	RBA	HBA	SRH
MAP	0.1799	0.3308	0.3274	0.3416	0.3664

events reaches 10000. Another observation is that the result of HBA becomes more precise as the number of training events increases. However, the improvement on the precision is only 0.03 when the size increases from 2500 to 10000. This is also caused by the property of dataset itself. Specifically, each user only participates a few events. As a result, the ground truth matrix for training is very sparse. In most case, training negatives are received.

### 5.1.4 SRH

Finally, the hybrid approach SRH is evaluated on the academic dataset. In this experiment, the second data set in SBA experiment is used to generate latent topic models with the number of topics as 50. Same query semantics is adopted as top 20 events are returned in the answer set. Three combinations of weight for  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  (0.2, 0.3, 0.5; 0.5, 0.2, 0.3; 0.3, 0.5, 0.2) are selected in the experiments. The MAP value for SRH in Table 2 uses the setting of 0.2, 0.3, 0.5 which has the highest precision among the three sets. As can be seen from the table, the hybrid method SRH with a proper set of weights outperforms all other three methods as well as the random method. The reason is that with tuning the weights, the hybrid method can best fit its sub-method to the properties of the dataset in order to provide a better recommendation than any of them.

## 5.2 Facebook

Apart from academic events, we also applied our methods to event recommendation in one general and popular social network - Facebook. Facebook provides a very useful tool - Graph API. It is the core of Facebook Platform, enabling developers to read from and write data into Facebook. It presents a simple, consistent view of the Facebook social graph, uniformly representing objects in the graph (e.g., people, photos, events, and pages) and the connections between them (e.g., friend relationships, shared content, and photo tags).

Specifically, an access token is generated on Graph API Explorer. One token corresponds to one user. Given this token, the information of all friends of a specific user as well as all events which his/her friends attend are crawled. The keywords for topic

modeling are extracted from such information to build the profiles for both users and events. Two ways are exploited to generate event sets. Starting from a user, the events list of her/him is retrieved using Graph API. The words contributing the event profile can be extracted based on the event id. Second method is based on the search function provided by Graph API. For example, all events related to keyword “USC” can be retrieved through the link

<https://graph.facebook.com/search?q=USC&type=event>

The second method significantly enlarges the dataset since it does not depend on specific users.

For the experimental setting, 1088 users and 4040 events are crawled from Facebook Graph API. In total, 16499 unique keywords are extracted from those users and events raw texts. Same as academic event recommendation, LDA is used to generate topics and distributions over each user and event. Specifically, 300 iterations are adopted. For the parameters,  $\beta = 0.01$  and  $\alpha = 1$ . Figure 6 shows MAPs of SBA, RBA, HBA and SRH ( $\omega_1 = 0.3, \omega_2 = 0.5, \omega_3 = 0.2$ ). Each point is generated using the average MAP over 100 users. The recommendation result includes 20 events related to those 100 users. For HBA, the number of training events are 4000.

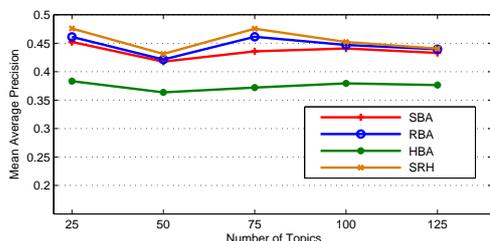


Figure 6: MAP on Facebook dataset

As can be seen from Figure 6, SRH with  $\vec{\omega} = [0.3, 0.5, 0.2]^T$  outperforms all three other methods using Facebook data. Among the three methods, RBA has the highest precision. This is because Facebook has a more well-developed friendship relation network which can be utilized to find potential common interests among different users. HBA has lowest precision because the attendance history matrix is still quite sparse. Most friends of the user used in the experiments are not so active in terms of attending events. However, the precision is still over twice higher than the random method (with precision as 0.1799). Another observation is MAP for all methods are not so sensitive to the number of topics. As a result, 25 topics are enough for recommendation in order to reduce computational cost.

## 6 Conclusions

In this paper, we investigate the problem of event recommendation. We propose four methods involving two machine learning techniques (i.e., LDA and logistic regression) which can extract implicit semantics from the raw data of events and users. We also retrieved the explicit semantics by enabling open linked data (e.g., linked eventseer and linked DBLP) in the recommendation process. Finally, we conduct comprehensive experiments both academic events (i.e., conference and workshops) and social networking events (i.e., social activities on Facebook). The results show that the hybrid approach SRH outperforms all other three methods with a proper selection of weights. Moreover, all four methods have higher recommendation precisions than the random method on both datasets.

One future direction is to automate the process of choosing weights for SRH. Some machine learning techniques such as  $n$ -fold cross-validation can be adopted. Another direction is to focus on the computational aspect of the recommendation algorithms. The reason is that dynamism exists everywhere in the social networks and the recommended events should be updated with the times. How to provide not only accurate but also prompt recommendations is a challenging problem to investigate.

## ACKNOWLEDGEMENTS

This work is supported by Chevron Corp. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

## REFERENCES

- Au Yeung, C. and Iwata, T. (2011). Strength of social influence in trust networks in product review sites. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 495–504. ACM.
- Blei, D., Ng, A., and Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022.
- Chen, H. and Chen, A. (2001). A music recommendation system based on music data grouping and user interests. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 231–238. ACM.
- Coppens, S., Mannens, E., De Pessemier, T., Geebelen, K., Dacquin, H., Van Deursen, D., and Van de Walle, R.

- (2012). Unifying and targeting cultural activities via events modelling and profiling. *Multimedia Tools and Applications*, pages 1–38.
- Cornelis, C., Guo, X., Lu, J., and Zhang, G. (2005). A fuzzy relational approach to event recommendation. In *Proceedings of the Indian International Conference on Artificial Intelligence*.
- Daly, E. M. and Geyer, W. (2011). Effective event discovery: using location and social information for scoping event recommendations. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 277–280, New York, NY, USA. ACM.
- De Pessemier, T., Coppens, S., Geebelen, K., Vleugels, C., Bannier, S., Mannens, E., Vanhecke, K., and Martens, L. (2011). Collaborative recommendations with content-based filters for cultural activities via a scalable event distribution platform. *Multimedia Tools and Applications*, pages 1–47.
- Guan, Z., Bu, J., Mei, Q., Chen, C., and Wang, C. (2009). Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 540–547. ACM.
- Kayaalp, M., Özyer, T., and Özyer, S. T. (2009). A collaborative and content based event recommendation system integrated with data collection scrapers and services at a social networking site. In *ASONAM*, pages 113–118.
- Klamma, R., Pham, M. C., and Cao, Y. (2009). You never walk alone: Recommending academic events based on social network analysis. In *Complex (1)*, pages 657–670.
- Konstas, I., Stathopoulos, V., and Jose, J. M. (2009). On social networks and collaborative recommendation. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 195–202, New York, NY, USA. ACM.
- Li, H., Tian, Y., Lee, W.-C., Giles, C. L., and Chen, M.-C. (2010). Personalized feed recommendation service for social networks. In *SocialCom/PASSAT*, pages 96–103.
- Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80.
- Manning, C. D., Raghavan, P., and Schtze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- Minkov, E., Charrow, B., Ledlie, J., Teller, S. J., and Jaakkola, T. (2010). Collaborative future event recommendation. In *CIKM*, pages 819–828.
- Perny, P. and Zucker, J. (1999). Collaborative filtering methods based on fuzzy preference relations. *Proceedings of EUROFUSE-SIC*, 99:279–285.
- Sigurbjörnsson, B. and Van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on World Wide Web*, pages 327–336. ACM.
- Yang, S., Long, B., Smola, A., Sadagopan, N., Zheng, Z., and Zha, H. (2011). Like like alike: joint friendship and interest propagation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 537–546. ACM.