Context-aware Academic Collaborator Recommendation

Zheng Liu
HKUST
Hong Kong SAR.
zliual@cse.ust.hk

Xing Xie
Microsoft Research Asia
Beijing, China
xing.xie@microsoft.com

Lei Chen
HKUST
Hong Kong SAR.
leichen@cse.ust.hk

ABSTRACT

Collaborator Recommendation is a useful application in exploiting big academic data. However, existing works leave out the contextual restriction (i.e., research topics) of people's academic collaboration, thus cannot recommend suitable collaborators for the required research topics. In this work, we propose Context-aware Collaborator Recommendation (CACR), which aims to recommend high-potential new collaborators for people's context-restricted requests. To this end, we design a novel recommendation framework, which consists of two fundamental components: the Collaborative Entity Embedding network (CEE) and the Hierarchical Factorization Model (HFM). In particular, CEE jointly represents researchers and research topics as compact vectors based on their co-occurrence relationships, whereby capturing researchers' context-aware collaboration tendencies and topics' underlying semantics. Meanwhile, HFM extracts researchers' activenesses and conservativenesses, which reflect their intensities of making academic collaborations and tendencies of working with non-collaborated fellows. The extracted activenesses and conservativenesses work collaboratively with the context-aware collaboration tendencies, such that high-quality recommendation can be produced. Extensive experimental studies are conducted with large-scale academic data, whose results verify the effectiveness of our proposed approaches.

KEYWORDS

Collaborator Recommendation; Academic Data Mining; Context-aware Recommendation.

ACM Reference Format:
Zheng Liu, Xing Xie, and Lei Chen. 2018. Context-aware Academic Collaborator Recommendation. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3219819.3220050

1 INTRODUCTION

Nowadays, a number of academic search engines, e.g., Microsoft Academic Search, Google Scholar, ArnetMiner and CiteSeerX, have come into being, making it convenient to explore tremendous amount of digital academic materials, like scientific literatures and researchers' profiles. The rapid growth of data volume and variety requires more advanced tools to help with academic data's exploration, and huge effort has been dedicated to facilitate various types of meaningful applications. Representative works include: reference recommendation [7, 8, 20, 22, 32], scholars' profiling [28, 29] and academic impact estimation [5, 12].

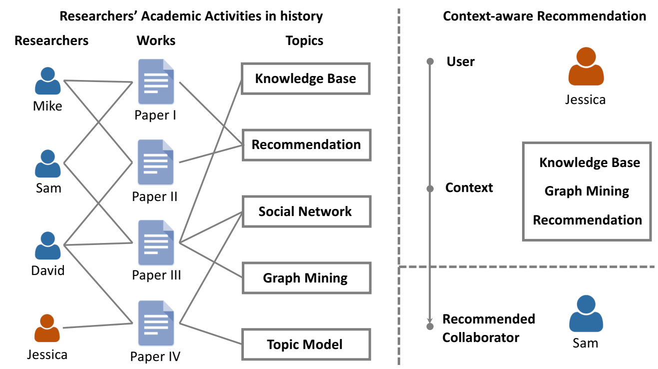


Figure 1: Toy Example for CACR.

In addition to the above applications, **academic collaborator recommendation** has long been regarded as an useful application in exploiting academic data, which aims to find the potential collaborators for a given researcher. In the past few years, several works have been proposed [3, 11, 23, 27] to solve such a problem. Despite the achieved progress, existing techniques are only able to recommend collaborators without contextual restriction, whereas incapable of making recommendations for the required topics. For example, current works cannot recommend appropriate candidates for a researcher to work with on research topics "Distributed System" and "Parallel Computing". Generally speaking, a researcher would determine her topics to work on, in advance of seeking her collaborators; that is to say, academic collaborations are "context-dependent" inherently. As a result, it is necessary to equip the academic collaborator recommendation with context-awareness.

In this paper, we focus on the **Context-aware Academic Collaborator Recommendation** (**CACR** for short) problem. Particularly, given a specific researcher r, together with her research topics T_0, (i.e., the contextual restriction), CACR finds the candidate fellows who are most likely to work with r on the required topics. Besides, to better meet the practical demands, CACR will recommend "new collaborators", i.e., those who have never collaborated with r in history. This is because researchers would usually expect a system to help with the expansion of their cooperation scopes, instead of simply retrieving the fellows whose collaboration relationships have been established. In this place, a toy example is presented as follows for the better illustration of CACR.

ACM ISBN 978-1-4503-5552-0/18/08...$15.00
https://doi.org/10.1145/3219819.3220050

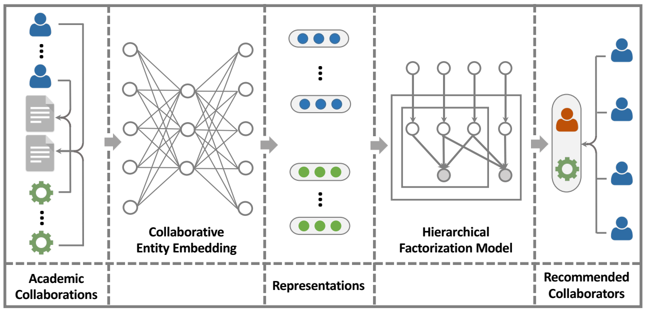


Figure 2: Infrastructure of CACR.

Example 1.1. *Suppose there are 4 researchers, Mike, Sam, David and Jessica, whose collaborative works and corresponding research topics are shown in Figure 1. Now, Jessica wishes to find someone whom she can work with on topics Knowledge Base, Graph Mining and Recommendation. According to the collaborations in history, Jessica has one unique collaborator, David; meanwhile, Mike and Sam have the same number of cooperations with David. As such, Mike and Sam's "social closenesses" to Jessica are roughly the same. However, it is also observed that Sam has conducted research on all the specified topics, which indicates that he is more likely to collaborate with Jessica for the given context. Jointly taking both aspects into consideration, CACR recommends Sam, instead of Mike, as Jessica's collaborator for the required topics.*

To solve the CACR problem, there are two challenges to overcome: 1) the extraction of researchers' context-aware collaboration tendencies; and 2) the identification of potential new collaborators for the context-restricted requests.

Firstly, as discussed in Example 1.1, CACR must rationally figure out researchers' collaboration tendencies under different contexts. Such a requirement can be further concretized into the following two consecutive tasks: **1)** to analyze researchers' relationships (or social-closeness), and the underlying semantics of research topics; **2)** to estimate researchers' collaboration tendencies on different topics. Treating the above tasks independently, many of the recent works can be applied. For example, graph-representation approaches in [6, 18, 26] can be adopted, which will map researchers and topics into compact vector spaces based on their co-occurrence relationships. The generated representations can be further used by supervised learning methods, like Logistic Regression or Factorization Machine [17, 21], for the inference of researchers' collaboration tendencies under different contexts. However, researchers and topics are mutually dependent in reality: a researcher's collaborators will be different given non-identical topics; similarly, a topic's coincident-topics will also be different given distinct researchers. For example, M. Jordan would work with I. Stoica on real-time system, but with D. Blei on LDA; meanwhile, the topic machine-learning tends to co-occur with LDA given D. Blei, whereas with real-time system given I. Stoica. However, the above simple workflow ignores such a mutual-dependency, thus cannot fully capture the collaboration tendencies and underlying semantics.

Secondly, it is necessary to develop an effective mechanism to identify potential new collaborators. In reality, people tend to repetitively collaborate with certain groups of fellow researchers; yet, new collaborations are much less frequent (to be demonstrated in Section 3). Because of such a property, non-collaborated fellows tend to be assigned with small and unreliable recommendation scores by the conventional approaches [30, 31, 33, 34], making it difficult to accurately identify potential candidates from the non-collaborated fellows.

To address the above challenges, a two-stage framework is proposed in this work, which jointly represents researchers and topics based on their mutual-dependency, and extracts researchers' underlying characters for high-quality new collaborator recommendation. For the first stage, a novel representation network, namely **Collaborative Entity Embedding** (CEE), is proposed, which represents researchers and topics with a dual structure. Different from the existing methods which generate representations individually for the entities, CEE learns joint embeddings for researchers and topics with their mutual-dependency preserved, which in turn contributes to better estimation of the context-aware collaboration tendencies.

For the second stage, a probabilistic graphical model, **Hierarchical Factorization Model** (HFM), is devised, which finds new potential collaborators with the assistance of researchers' underlying characters: **activeness** (intensity of making collaborations) and **conservativeness** (concentration of collaborations). Intuitively, a more active but less conservative researcher exhibits greater tendency of establishing new collaborations, which can be clearly observed from data (demonstrated in Section 3). Both characters are truthfully estimated in HFM based on the probabilistic reasoning of collaborations' counts and occurrences, whose results work jointly with CEE's representations to generate high-quality new collaborator recommendation. To summarize, the following contributions are made in this paper.

- To the best of our knowledge, this is the first work that focuses on the CACR problem, which is important in practice.
- A novel representation network, CEE, is proposed, which jointly represents researchers and topics with their mutual dependency preserved.
- A probabilistic graphical model, HFM, is devised, which exploits researchers' activenesses and conservativenesses to make high-quality new collaborator recommendation.
- Comprehensive experiments are conducted with large-scale academic data, whose results verify the effectiveness of our proposed methods.

The rest of our paper is organized as follows. Definitions and infrastructure are presented in Section 2, and explorations for the academic data are demonstrated in Section 3. The collaborative entity embedding and hierarchical factorization model are introduced in Section 4 and 5, respectively; followed by related works reviewed in Section 7. Finally, conclusion is made in Section 8.

2 DEFINITIONS AND INFRASTRUCTURE

2.1 Problem Definition

Researcher. A researcher is associated with her published literatures, which reveal her research interests and academic collaborations with others. A pair of researchers are "New" to each other

$\mathbf{R}$	the whole researchers
$\mathbf{T}$	the whole topics
$\mathbf{A}$	the whole academic literatures
$\mathbf{U}^e$	CEE's encoding matrix for researchers
$\mathbf{U}^d$	CEE's decoding matrix for researchers
$\mathbf{V}^e$	CEE's encoding matrix for topics
$\mathbf{V}^d$	CEE's decoding matrix for topics
$\boldsymbol{\phi}_{ij}^{a}$	Context-aware collaboration tendency of r_i and r_j on a
$\boldsymbol{\sigma_i}$	Conservativeness of r_i
$\boldsymbol{\omega_i}$	Activeness of r_i

Table 1: Summary of Frequent-used Notations.

if they have never established any collaboration relationship in history.

Research Topic. A research topic is a keyword or phrase extracted from a specific literature (e.g., a word in the title or keyword list); meanwhile, a literature consists of one or more research topics , which jointly reflect its underlying category.

Context. A collaboration's context refers to the set of topics, which researchers jointly work on in their collaborated literature.

With the above preliminaries, the definition of Context-aware Academic Collaborator Recommendation is made as follows.

Definition 2.1. *(CACR) Given researcher r_0 and topics $\boldsymbol{T}_0$, CACR finds K (specified by requester) new collaborators from all candidates $\boldsymbol{R}$, who will work with r_0 on $\boldsymbol{T}_0$ with the highest probabilities.*

2.2 System Infrastructure

A two-stage framework is proposed for CACR, whose infrastructure is shown as Figure 2. Particularly, there are two major components: 1) the Collaborative Entity Embedding and 2) the Hierarchical Factorization Model, which are introduced as follows.

Collaborative Entity Embedding (CEE). CEE is a neural network which maps researchers and topics into their latent representations. CEE is trained with the academic collaboration records in history, where two types of properties are captured: 1) a pair of topics' correlation given specific researchers, and 2) a pair of researchers' co-occurrence given specific topics. The above properties encode topics' underlying semantics and researchers' context-aware collaboration tendencies, which are vital for CACR.

Hierarchical Factorization Model (HFM). HFM is a Bayesian network, which extracts researchers' **Activenesses** and **Conservativeness** for generating high-quality CACR. Particularly, the activeness indicates a researcher's intensity of academic collaboration, i.e., an active researcher is expected to make a large number of collaborations in the future; meanwhile, the conservativeness reflects the concentration of a researcher's collaborations: a conservative researcher confines her collaborations within a small group of fellows, whereas, a non-conservative one would frequently work with different fellows and continuously expand her collaboration scope. As such, a more active but less conservative researcher tends to make more new collaborations in the future. Given researchers' collaboration tendencies under specific contexts (estimated by CEE), HFM judiciously explains the following two observations: 1) the occurrences of researchers' historical collaborations, and 2) the counts of researchers' collaborations in history. With such a probabilistic formulation, the activeness and conservativeness can be truthfully extracted for each researcher, whereby helping to identify the potential new collaborators.

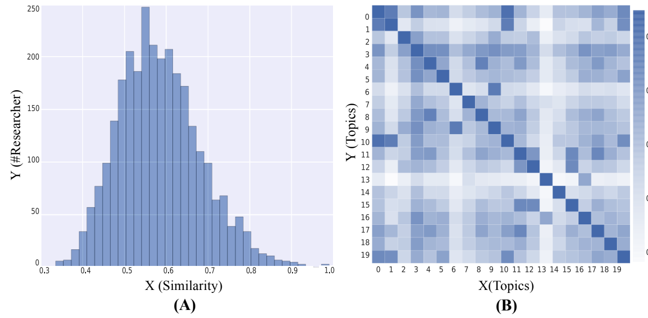


Figure 3: Topics' Similarity. (A): histogram of researchers' average topics' similarities; (B): heatmap of Philip Yu's topics' similarities.

Workflow. Our framework works with the following procedures (demonstrated as Figure 2). Firstly, researchers and topics are mapped into their latent representations with CEE, which encodes the context-aware collaboration tendencies and underlying semantics. Secondly, with the representations learned by CEE, HFM extracts the inherent activeness and conservativeness for each researcher. Finally, appropriate candidates for the CACR request are produced with rational consideration of researchers' context-aware collaboration tendencies, activenesses and conservativeness.

3 DATA EXPLORATION

In this section, comprehensive data exploration is performed on real-world academic data, which helps to better clarify our motivation and underlying mechanisms for the recommendation algorithms.

3.1 Exploration Settings

We adopt the DBIS dataset created in [24] for our data exploration, which includes a total of 72,902 literatures and 60,694 researchers from relevant communities of database and information-system. The research topics are extracted from titles of the literatures; with necessary cleaning and integration, e.g., stemming and stop-word removing, a total of 9,691 words are acquired, each of which is treated as a unique topic. For the adopted dataset, researchers' collaboration behaviors are analyzed for the following aspects.

Topics' Similarity. Researchers' collaboration tendencies under different contexts are analyzed with their topics' similarities. Given researcher r, similarity between two topics t_i and t_j is calculated as the cosine similarity of r's collaborators on both topics. For example, suppose that Jessica works with {Tom, Mike} on "recommendation-system" (RS) and {Adam, Mike} on "knowledge-graph" (KG), the similarity between RS and KG is calculated as: $(1 \times 1 + 1 \times 0 + 0 \times 1)/(\sqrt{2} \times \sqrt{2}) = 0.5$. Apparently, a large similarity indicates similar groups of collaborators under two topics.

Ratio of New Collaborations. The ratio of new collaborations to the total is calculated to explore researchers' tendency of working with new collaborators. For example, researcher r_0 has 2 literatures a_1 and a_2, where collaborations are made with $\{r_1, r_2, r_3\}$ and $\{r_3, r_4\}$, respectively. Taking a_1 as history, r_0 makes one new collaboration in a_2 (with r_4), thus letting ratio equal to 1/2 (i.e., $|\{r_4\}|/|\{r_3, r_4\}|$). It's obvious that a larger ratio indicates a higher chance of making new collaborations in the future. In our exploration, 70% of literatures are partitioned into history, while new collaborations in the remaining 30% are selected for evaluation.

Ratio of New Collaborations w.r.t. Conservativeness & Activeness. The tendency of making new collaborations is further analyzed w.r.t. researchers' inherent characters: conservativenesses and activenesses. In our exploration, the conservativeness is measured with the entropy of historical collaborations (denoted by H); while the activeness is measured in terms of the collaborations' volume (denoted by V); e.g., given researcher r_0, who collaborates 2 times with r_1, and 1 time with r_2, it follows that H $= -(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}) = 0.637$, and V = 3. By definition, the entropy and conservativeness are negatively correlated: a larger H indicates a researcher's frequent collaborations with different fellows, thus reflecting lower conservativeness; meanwhile, a larger V means intense academic collaborations, thus showing greater activeness.

3.2 Exploration Analysis

Topics' Similarity. In Figure 3 (A), the average similarities of the top-20 frequently used topics are demonstrated for researchers with over 20 publications (a total of 3431). According to the presented result, the majority of similarities are below 0.8. Meanwhile, in Figure 3 (B), topics' similarities are presented as a heatmap for Philip Yu, who has the most publications (328) in the dataset. It is observed that many of the non-diagonal elements are small, reflecting that collaboration behaviors are significantly differed across research topics. (The diagonal elements represent topics' similarities w.r.t. themselves, thus equal to 1.) Both findings indicate that researchers tend to work with different groups of collaborators given distinct topics; therefore, it's necessary for collaborator recommendation to take contextual restriction into account.

Ratio of New Collaboration. Ratios of new collaboration (Ratios for short) are calculated for researchers with over 20 publications, whose histogram is shown as Figure 4 (A). It can be observed that ratios are relatively small for the majority of researchers, as the demonstrated result is densely distributed below 0.3. In other words, researchers' probabilities of working with new collaborators are significantly smaller than those with the existing ones.

Conservativeness & Activeness. Ratios' relationship with entropy is demonstrated in Figure 4 (B), where it is clearly observed that ratio and entropy are positively correlated. That is to say, researchers' of higher entropy, i.e., lower conservativeness, will be more likely to work with the new collaborators. Meanwhile, researchers' counts of collaborations are shown in Figure 4 (C), which follow the form of power-law distributions; i.e., the majority of researchers make limited amounts of collaborations, while some others generate exceedingly more. Besides, the counts and ratios are also positively correlated in Figure 4 (D), despite that the correlation is less significant than that of ratio and entropy. Therefore, it is expected that larger amounts of new collaborations can be produced by researchers with higher activenesses.

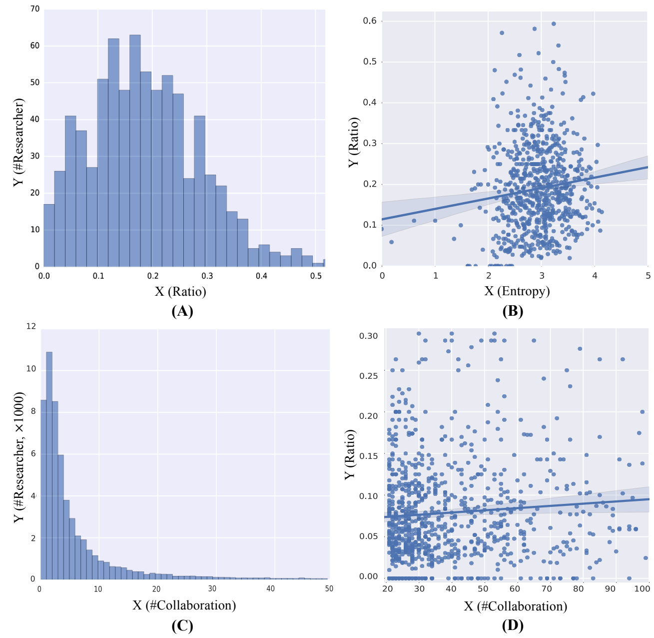


Figure 4: Ratio of New Collaboration. (A), (C): histograms of ratio and #collaboration; (B), (D): ratio's relationships with entropy and #collaboration.

Summary. The major findings of data exploration are summarized as follows. Firstly, researchers will work with different collaborators, given distinct research topics; therefore, collaborator recommendation must take context into consideration. Secondly, researchers tend to work with fixed groups of fellows, making new collaborations happen with small probabilities. Thirdly, researchers' low-conservativeness and high-activeness contribute to establishing new collaborations, thus necessary to leverage in the recommendation algorithm.

4 COLLABORATIVE ENTITY EMBEDDING

4.1 Framework of CEE

In this section, we devise Collaborative Entity Embedding (CEE) to represent topics and researchers in compact vector spaces. To fully capture researchers' context-aware collaboration tendencies and topics' underlying semantics, the extracted representations are desirable of maximizing two objectives, $\Omega_\mathbf{R}$ and $\Omega_\mathbf{T}$:

$$\Omega_\mathbf{R} = \sum_{a \in \mathbf{A}} \sum_{r \in \mathbf{R}_a} \sum_{r_c \in \mathbf{R}_a \setminus r} \log P(r_c | r, \mathbf{T}_a),$$
$$\Omega_\mathbf{T} = \sum_{a \in \mathbf{A}} \sum_{t \in \mathbf{T}_a} \sum_{t_c \in \mathbf{T}_a \setminus t} \log P(t_c | t, \mathbf{R}_a),$$

where $\mathbf{R}_a$, $\mathbf{T}_a$ are the researchers and topics in literature a. Apparently, $\Omega_\mathbf{R}$ and $\Omega_\mathbf{T}$ represent the generative log-likelihoods of researchers' and topics' co-occurrences w.r.t. the corresponding literatures' records (i.e., $\mathbf{R}_a$ and $\mathbf{T}_a$).

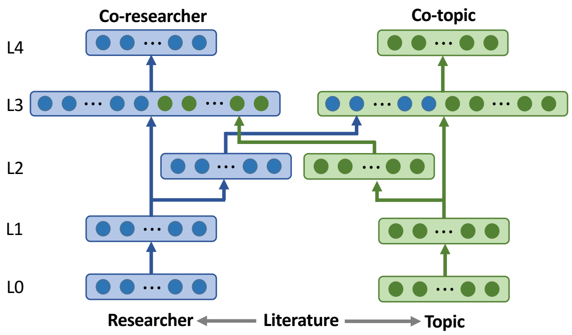


Figure 5: Architecture of Collaborative Entity Embedding.

To realize the above objectives, CEE adopts a dual-structure (shown as Figure 5), which naturally preserves the mutual-dependency while modeling entities' (i.e., researchers and topics) co-occurrence relationship. In particular, each side in CEE consists of five layers: one-hot layer L_0, encoding layer L_1, mean-pooling layer L_2, concatenating layer L_3, and decoding layer L_4. For each specific academic literature a, CEE works with the following steps.

• L_0 represents each entity (i.e., a researcher or topic in a) as a unique one-hot vector, whose non-zero entry identifies the corresponding entity.

• L_1 further maps each entity into the embedding space with an encoding matrix ($\mathbf{U}^e$ for researchers and $\mathbf{V}^e$ for topics); e.g., researcher r is embedded with the r-th row of $\mathbf{U}^e$, i.e., $\mathbf{U}^e_r$.

• L_2 aggregates each literature's information with mean-pooling operation. Specifically, $\mathbf{R}_a$ and $\mathbf{T}_a$ are represented as $\overline{\mathbf{U}}^e_{\mathbf{R}_a}$ and $\overline{\mathbf{V}}^e_{\mathbf{T}_a}$, which equal to $\sum_{r \in \mathbf{R}_a} \mathbf{U}^e_r / |\mathbf{R}_a|$ and $\sum_{t \in \mathbf{T}_a} \mathbf{V}^e_t / |\mathbf{T}_a|$, respectively.

• L_3 concatenates each entity's embedded vector with the literature's information extracted from the counter side. Such a vector produces a joint representation for the entity and its counterpart; e.g., $[\mathbf{U}^e_r, \overline{\mathbf{V}}^e_{\mathbf{T}_a}]$ and $[\mathbf{V}^e_t, \overline{\mathbf{U}}^e_{\mathbf{R}_a}]$ (denoted as $\mathbf{W}^r_{\mathbf{T}_a}$ and $\mathbf{W}^t_{\mathbf{R}_a}$, respectively), for researcher r and topic t from literature a.

• L_4 maps the joint representations from L_3 into decoded vectors (with matrix $\mathbf{U}^d$ for the researcher side and $\mathbf{V}^d$ for the topic side). For example, $\mathbf{W}^r_{\mathbf{T}_a}$ and $\mathbf{W}^t_{\mathbf{R}_a}$ are decoded into vectors $\theta^r_{\mathbf{T}_a}$ and $\theta^t_{\mathbf{R}_a}$, through operations $\mathbf{W}^r_{\mathbf{T}_a}\mathbf{U}^d$ and $\mathbf{W}^t_{\mathbf{R}_a}\mathbf{V}^d$, respectively. Entries of the decoded vector reflect the relative co-occurrence tendencies for the input entity. Specifically, r_j's co-occurrence tendency with r in a is calculated as $\exp(\theta^r_{\mathbf{T}_a}\{r_j\})$ (where $\theta^r_{\mathbf{T}_a}\{r_j\}$ is the r_j-th entry of $\theta^r_{\mathbf{T}_a}$), thus letting the co-occurrence probability equal to:

$$P(r_j|r, \mathbf{T}_a) = \frac{\exp(\theta^r_{\mathbf{T}_a}\{r_j\})}{\sum_{r' \in \mathbf{R}} \exp(\theta^r_{\mathbf{T}_a}\{r'\})}. \tag{1}$$

Similarly, t_j's co-occurrence probability with t in a equals to:

$$P(t_j|t, \mathbf{R}_a) = \frac{\exp(\theta^t_{\mathbf{R}_a}\{t_j\})}{\sum_{t' \in \mathbf{T}} \exp(\theta^t_{\mathbf{R}_a}\{t'\})}. \tag{2}$$

Remark 4.1. Entities' Augment. CEE is not limited to taking topics as context. In fact, extra type of contextual entities (e.g., affiliations) can be stacked in CEE with the same fashion as topics; i.e., the extra entities' embeddings are learned with their coincident researchers' and topics' average embeddings incorporated in the concatenation layer, and vice versa for researchers and topics.

Algorithm 1: Optimization of CEE

input : The whole academic literature $\mathbf{A}$, L2-reg λ, $\mathrm{Dim}_{\mathbf{R}}$ and $\mathrm{Dim}_{\mathbf{T}}$ for $\mathbf{U}^e$ and $\mathbf{V}^e$'s embedding dimensions.
output: Parameters of CEE: $\mathbf{U}^e$, $\mathbf{U}^d$, $\mathbf{V}^e$, $\mathbf{V}^d$.

begin
  $\mathbb{U}^e, \mathbb{V}^e \leftarrow$ Preliminary Training;
  Warm-start: $\mathbf{U}^e \leftarrow \mathbb{U}^e + \mathcal{U}^e$, $\mathbf{V}^e \leftarrow \mathbb{V}^e + \mathcal{V}^e$;
  repeat
    R-Step:
    $\mathbf{S}_{\mathbf{R}} \leftarrow$ mini-batch of training samples $(r_i, r_j, \mathbf{T}_a)$;
    AdamGD($\{\mathcal{U}^e, \mathcal{V}^e, \mathbf{U}^d\}$, $\mathbf{S}_{\mathbf{R}}$);
    T-Step:
    $\mathbf{S}_{\mathbf{T}} \leftarrow$ mini-batch of training samples $(t_i, t_j, \mathbf{R}_a)$;
    AdamGD($\{\mathcal{V}^e, \mathcal{U}^e, \mathbf{V}^d\}$, $\mathbf{S}_{\mathbf{T}}$);
  until Convergence;
return $\mathbf{U}^e$, $\mathbf{U}^d$, $\mathbf{V}^e$, $\mathbf{V}^d$;

4.2 Parameter Inference

In our work, CEE's parameters are inferred with warm-start, with embedding matrices $\{\mathbb{U}^e, \mathbb{U}^d\}$ and $\{\mathbb{V}^e, \mathbb{V}^d\}$ individually learned for researchers and topics in the very beginning. Specifically, the preliminary training is carried out to learn entities' 1st-order proximity-preserved representations (similar fashion as [18]), without considering their mutual-dependency; e.g., $\mathbb{U}^e_{r_i}\mathbb{V}^e_{r_j} > \mathbb{U}^e_{r_i}\mathbb{V}^e_{r_k}$, given that r_j, rather than r_k, has been r_i's collaborator. Based on the preliminary training result, $\mathbf{U}^e$ and $\mathbf{V}^e$ are set to be $\mathbb{U}^e + \mathcal{U}^e$ and $\mathbb{V}^e + \mathcal{V}^e$, respectively, with $\mathcal{U}^e$ and $\mathcal{V}^e$ randomly initialized.

Following the preliminary training, $\{\mathcal{U}^e, \mathbf{U}^d\}$ and $\{\mathcal{V}^e, \mathbf{V}^d\}$ are estimated to capture researchers' and topics' mutual-dependent co-occurrence relationship. Such a goal is formulated into the objective function Ω, which turns out to be the combination of $\Omega_{\mathbf{R}}$ and $\Omega_{\mathbf{T}}$:

$$\Omega = \Omega_{\mathbf{R}} + \Omega_{\mathbf{T}} + \lambda(\|\mathcal{U}^e\|_2^2 + \|\mathbf{U}^d\|_2^2 + \|\mathcal{V}^e\|_2^2 + \|\mathbf{V}^d\|_2^2), \tag{3}$$

where λ is the constant coefficient for L2-norm regularization.

Ω's optimization is carried out in an alternative manner. In particular, Ω is repetitively maximized w.r.t. $\{\mathcal{U}^e, \mathbf{U}^d\}$ and $\{\mathcal{V}^e, \mathbf{V}^d\}$ in two successive steps: R-Step and T-Step. In R-Step, mini-batches of training instances, $\mathbf{S}_{\mathbf{R}}$, are sampled from the whole academic literatures, based on which Ω is maximized w.r.t. $\{\mathcal{U}^e, \mathcal{V}^e, \mathbf{U}^d\}$ while $\{\mathbf{V}^d\}$ stay fixed. Similarly, in T-Step, based on the presented training samples, $\mathbf{S}_{\mathbf{T}}$, Ω is maximized w.r.t. $\{\mathcal{V}^e, \mathcal{U}^e, \mathbf{V}^d\}$ with fixed $\{\mathbf{U}^d\}$. Following the conventional treatment, optimizations in both steps are solved with the first-order gradient-based approach, where Adam [9] (denoted as AdamGD in Alg. 1) is adopted for the superior training effectiveness. The overall workflow of CEE's parameter inference is sketched with Alg. 1.

5 HIERARCHICAL FACTORIZATION MODEL

5.1 Formulation of HFM

Although researchers' collaboration tendencies on different contexts can be effectively captured with CEE, it's still challenging to predict potential new collaborators. This is because new collaborations usually take place with much smaller probabilities between the non-collaborated researchers, compared with the existing collaborators. In this section, the Hierarchical Factorization Model (HFM) is constructed on top of CEE, which takes full advantage of researchers' inherent characters, activeness and conservativeness, to generate accurate prediction for new collaborators. To estimate both characters, HFM probabilistically reasons the observed academic collaborations from two perspectives: Occurrence and Count.

Occurrence of Collaboration. HFM introduces the latent variable, conservativeness (σ), to capture the probability of each researcher's collaboration with different fellows. In particular, given r_i and r_j's context-aware collaboration tendency on $\mathbf{T}_a$, i.e., $P(r_j|r_i,\mathbf{T}_a)$ (denoted as ϕ^a_{ij}), the event of r_j's participation in r_i's work a (denoted as I^a_{ij}) is captured by the Bernoulli distribution $\mathrm{Ber}((\phi^a_{ij})^{\sigma_i\sigma_j})$, which follows:

$$P(\mathrm{I}^a_{ij}) = \begin{cases} (\phi^a_{ij})^{\sigma_i\sigma_j}, & \mathrm{I}^a_{ij}=1, \\ 1-(\phi^a_{ij})^{\sigma_i\sigma_j}, & \mathrm{I}^a_{ij}=0. \end{cases} \tag{4}$$

As $\phi^a_{ij}<1$ and $\sigma_i,\sigma_j>0$, smaller conservativenesses of both sides will increase the value of $P(\mathrm{I}^a_{ij})$, resulting in a larger chance of collaboration especially when ϕ^a_{ij} is small.

Count of Collaborations. HFM utilizes the latent variable of activeness, ω_i, to indicate her intensity of making academic collaborations. However, the activeness is not explicitly reflected in data, as the observed number of collaborations is jointly determined by researchers' activenesses and conservativeness; i.e., a higher-active but less-conservative researcher tends to make more collaborations. As a result, the factorized Poisson distribution, $\mathrm{Poi}(\omega_i\rho_i)$, is adopted for the reasoning of r_i's count of collaborations (denoted as N_i):

$$P(\mathrm{N}_i=k) = \frac{(\omega_i\rho_i)^k}{k!}e^{-\omega_i\rho_i}, \tag{5}$$

where ρ_i stands for the sum of r_i's tendencies on joining other researchers' works, i.e.,

$$\rho_i = \sum_{a\in\mathbf{A}}\sum_{r_j\in\mathbf{R}_a\setminus r_i}(\phi^a_{ij})^{\sigma_i\sigma_j}. \tag{6}$$

Notice that the calculation of Eq. 6 will be expensive due to the summation over all literatures in $\mathbf{A}$, sampling operation is adopted for its approximation:

$$\rho_i = \frac{\sum_{a\in\mathrm{sub}(\mathbf{A})}\sum_{r_j\in\mathrm{sub}(\mathbf{R}_a\setminus r_i)}(\phi^a_{ij})^{\sigma_i\sigma_j}}{\sum_{a\in\mathrm{sub}(\mathbf{A})}|\mathrm{sub}(\mathbf{R}_a\setminus r_i)|}|\sum\nolimits_{\mathbf{A}}\mathbf{R}_a\setminus r_i|, \tag{7}$$

where sub() stands for the sampled elements of relevant sets.

The graphical representation of HFM is illustrated with Figure 6, where the generative process is presented as follows.

- For each researcher $r_i\in\mathbf{R}$:
  draw conservativeness $\sigma_i\sim\mathrm{Gauss}(1,\tau^{-1})$;
  draw activeness $\omega_i\sim\mathrm{Gamma}(\alpha_i,\beta_i)$;
  draw count of collaborations $\mathrm{N}_i\sim\mathrm{Poi}(\omega_i\rho_i)$.
- For each researcher $r_j\in\mathbf{R}$:
  - for each literature $a\in\{a|r_j\in\mathbf{R}_a\}$:
    - for each fellow $r_i\in\mathbf{R}_a\setminus r_j$:
      draw collaboration occurrence $\mathrm{I}^a_{ji}\sim\mathrm{Ber}((\phi^a_{ji})^{\sigma_i\sigma_j})$;

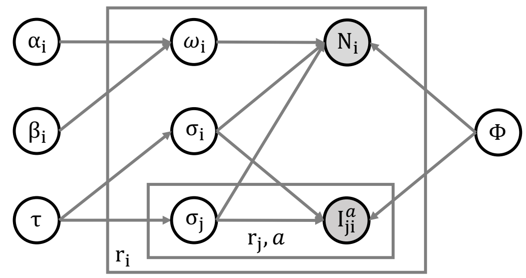


Figure 6: Graphical Model of HFM. Φ represents the context-aware collaboration tendencies acquired from CEE.

Notice that I^a_{ji} is only reasoned for the existed collaboration (i.e., $\mathrm{I}^a_{ji}=1$), as the causes of non-existed ones are indeterministic; in other words, the non-existed collaborations can be either resulted from researchers' conservativeness or their limited activenesses. Besides, it is set that $\alpha_i=\mathrm{N}_i$ and $\beta_i=|\mathbf{A}|/\sum_{\mathbf{A}}|\mathbf{R}_a\setminus r_i|$, which ensures the expectation of $\mathrm{Poi}(\omega_i\rho_i)$ equals to the observation N_i.

Given the above probabilistic formulation, the log-likelihood of HFM is derived as follows.

$$\mathcal{L} = \sum_{r_j}\sum_{a}\sum_{r_i}\sigma_i\sigma_j\log\phi^a_{ij} + \sum_{r_i}(\mathrm{N}_i\log\omega_i\rho_i - \omega_i\rho_i) + \sum_{r_i}-\tau\ \|\sigma_i-1\|_2^2+\varepsilon, \tag{8}$$

where the first three items are from the log-likelihoods of collaborations' occurrences, counts, and conservativeness's generation, while ε is the constant independent of latent variables. In this place, stochastic gradient descent is employed for the estimation of σ and ω, whose workflow is sketched as Alg. 2. Particularly, mini-batches of researchers, batch($\mathbf{R}$), are repetitively sampled from $\mathbf{R}$; and for each of the selected researchers, the relevant characters are updated w.r.t. their gradients, i.e., $\nabla_\omega\mathcal{L}$ and $\nabla_\sigma\mathcal{L}$. Such a process is continuously carried out until $\mathcal{L}$'s local optima is reached.

5.2 New Collaborator Recommendation

With activeness and conservativeness estimated for all researchers, the probability for r_j's collaboration with r_i on context $\mathbf{T}_a$ (denoted as Ψ^a_{ij}) can be directly analytically derived from Eq. 4 and 5 (similar deduction with Theorem2 in [15]):

$$\Psi^a_{ij} = 1-\exp\left(\frac{-\omega_j(\phi^a_{ij})^{\sigma_i\sigma_j}}{\sum_{a\in A}|\mathbf{R}_a|}\right) \simeq \frac{\omega_j(\phi^a_{ij})^{\sigma_i\sigma_j}}{\sum_{a\in A}|\mathbf{R}_a|} \propto \omega_j(\phi^a_{ij})^{\sigma_i\sigma_j}. \tag{9}$$

According to Eq. 9, a higher $\omega_j(\phi^a_{ij})^{\sigma_j}$ would mean r_j is more likely to join in r_i's research on $\mathbf{T}_a$, which makes it the indicator of recommendation's suitableness. As a result, the recommendation for a given request is made as:

$$\mathrm{Rec}(r_i,\mathbf{T}_a) = \text{Top-K}\{\omega_j(\phi^a_{ij})^{\sigma_j}|\mathbf{R}\setminus r_i\}. \tag{10}$$

Remark 5.1. Relationship with CEE. The collaboration tendency generated by CEE, i.e., ϕ^a_{ij}, will be equivalent to the result

Algorithm 2: Inference of HFM

input : The whole academic literatures **A**, the results of CEE: Φ, the hyper parameter τ.

output : The estimation of ω and σ.

begin
- Initialization of $\{\omega_i, \sigma_i|\mathbf{R}\}$;
- **repeat**
  - sample batch($\mathbf{R}$) from $\mathbf{R}$;
  - **for** $r_i \in$ batch($\mathbf{R}$) **do**
    - $\sigma_i \leftarrow \sigma_i + \eta \nabla_{\sigma_i} \mathcal{L}$;
    - $\omega_i \leftarrow \omega_i + \eta \nabla_{\omega_i} \mathcal{L}$;
- **until** Convergence;

return $\{\omega_i, \sigma_i|\mathbf{R}\}$;

of HFM, if $\sigma = 1$ (as initialized) and ω is identical for all the researchers. As a result, HFM can be regarded as a generalization of CEE, with researchers' activenesses and conservativeness taken into account.

Remark 5.2. Relationship between $\phi^{\mathrm{a}}_{\mathrm{ij}}$, $\mathrm{P}(\mathrm{I}^{\mathrm{a}}_{\mathrm{ij}})$ and $\Psi^{\mathrm{a}}_{\mathrm{ij}}$. All three items are semantically correlated as they illustrate r_i and r_j's collaboration from different perspectives. In particular, ϕ^a_{ij} indicates r_i's relative tendency of getting r_j involved in her work a, while her conservativeness and both characters of r_j are not considered. $P(\mathrm{I}^a_{ij})$ explains the absolute chance of collaboration, given that r_j is not subject to her constraint of activeness. Finally, Ψ^a_{ij} reasons the collaboration's tendency considering that r_j will only make a limited amount of collaborations due to her constraint of activeness. It is clear that Ψ^a_{ij} comprehensively characterizes the collaboration behavior, thus providing a more meaningful recommendation basis.

6 EXPERIMENTAL STUDY

6.1 Experiment Settings

6.1.1 Data and Implementation. In our experiments, a larger-scale dataset, AMiner [29], is adopted, which provides adequate training samples for the evaluation. Specifically, there are a total of 3,194,405 academic literatures; with necessary cleaning and integration (e.g., stemming and elimination of stopping words), 61,801 keywords are selected from the titles, each of which is treated as a unique topic. Meanwhile, 313,928 researchers, who have participated in no less than 5 literatures, are filtered for the experiments. The dataset is partitioned such that 80% of the academic literatures are used for the training phase, with 20% left for evaluation. For each of the literature to be evaluated, the collaboration between two researchers, r_i and r_j, is classified as "Exist" or "New", based on whether they have ever collaborated in training set (despite that the collaboration contexts are different). Algorithms are implemented in Python and Theano, and tested over the platform with one Xeon(R) E5-2680 v4 CPU and Nvidia Tesla M60 GPU cards.

6.1.2 Evaluation Objectives. Experiments are carried out for the following two primary objectives: 1) the **Context-restricted Collaborator Prediction**,which evaluates CEE's effectiveness on identifying researchers for the given context; 2) the **New Collaborator Prediction**, which tests CEE+HFM's performance on predicting the new collaborators. For both objectives, **Recall@K** is employed as the metric of effectiveness , whose definition is made as:

$$\text{Recall@K} = \frac{\text{\# of true collaborators in the top-K list}}{\text{the total \# of the true collaborators}}. \tag{11}$$

Such a metric is widely adopted for recommendation system's evaluation, such as [4, 30, 31, 33].

6.1.3 Evaluation Methods. To evaluate CEE's performance on identifying context-restricted collaborators, the following representative baseline methods are employed for the comparison.

• Logistic Regression (**LR**). LR is one of the most popular supervised learning approach, which is also widely adopted in recommendation system. In our experiments, LR's input features are formed as the concatenation of researcher's embedding and the averages of topics' embeddings, which are imported from CEE's warm-start matrices, $\mathbb{U}^e$ and $\mathbb{V}^e$.

• Structural Network Embedding (**SNE**). SNE stands for the methods designed for encoding the structural information of entities and their relations [2, 13]. When applied to the CACR problem, researchers are regarded as entities, and collaborations on specific contexts are treated as relations. Notice that there are literally infinite number of relations in CACR due to topics' combination, conventional methods, like [2, 13], cannot be directly adopted. Instead, we use a similar network architecture as Doc2Vec [10], which jointly learns embeddings $\mathcal{S}$ for the representation of researchers and topics, and the mapping matrix $\mathcal{M}$ for the prediction of collaborators; e.g., $r_j = \text{argmax}\{\mathcal{M}(\mathcal{S}_{r_i} + \mathcal{S}_{\mathbf{T}_a})\}$, given that r_j worked with r_i on $\mathbf{T}_a$.

• Task Specific Embedding (**TSE**). TSE is proposed in [4] for author identification problem. Briefly, TSE consists of three levels: the 1st level represents each source of contextual information (e.g. keywords, venues) with embedding learning approaches, e.g., [16, 18], to preserve the corresponding proximity (which does not consider the mutual-dependency between entities); and in the experiments, these embeddings are learned independently for researchers and topics. In the 2nd level, the extracted embeddings of all sources are integrated with different weights; finally, the 3rd level learns model parameters for specific classification tasks given the integrated embedding. Similar hierarchical structure is also adopted by [19] while learning a joint representation for networks with different views.

In addition, the context-independent (**CI**) collaborators are acquired with CEE's warm-start matrices $\{\mathbb{U}^e, \mathbb{U}^d\}$, which provides candidates without considering the contextual restrictions.

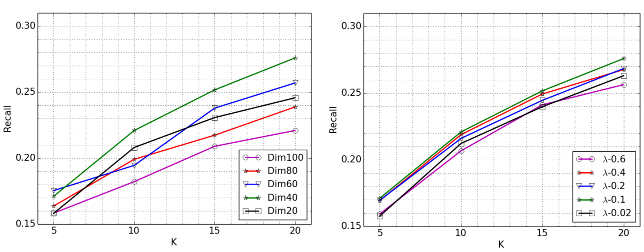
The following hyper-parameters are shared across all the comparison methods: 1) the embedding dimensions, Dim; and 2) the L2-norm regularization coefficient λ. In our experiments, Dim is ranged from {10, 20, 30, **40**, 50, 60}[1]; and λ is varied within {0.02, **0.1**, 0.2, 0.4, 0.6}. The bold numbers indicate the default values for effectiveness comparison, and others are used for the evaluation of parameter sensitivity. In the experiments, we use vanilla Softmax rather than negative sampling for all the comparison methods.

Meanwhile, HFM's performance in predicting new collaborators is compared with the one which directly uses CEE's result. What's

[1] In our experiments, researchers and topics are represented with the same dimension.

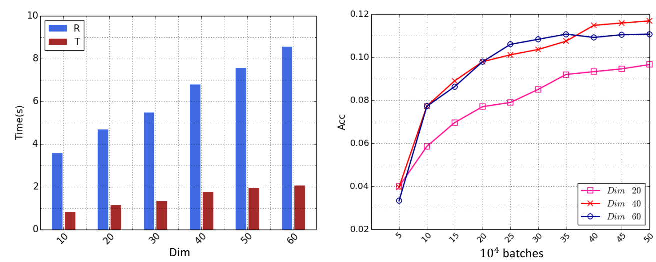
	Dim-20				Dim-40				Dim-60			
	Recall@K				Recall@K				Recall@K			
	5	10	15	20	5	10	15	20	5	10	15	20
CI	0.0707	0.0824	0.0887	0.0945	0.0865	0.0961	0.1044	0.1092	0.0883	0.0952	0.1092	0.1150
LR	0.1161	0.1336	0.1436	0.1520	0.1122	0.1282	0.1591	0.1732	0.1057	0.1367	0.1495	0.1646
SNE	0.1380	0.1753	0.2012	0.2146	0.1409	0.1828	0.2067	0.2227	0.1577	0.1879	0.2044	0.2206
TSE	0.1443	0.1870	0.2120	0.2287	0.1535	0.2018	0.2290	0.2486	0.1607	0.1922	0.2153	0.2339
CEE	**0.1582**	**0.2081**	**0.2307**	**0.2458**	**0.1711**	**0.2209**	**0.2517**	**0.2759**	**0.1726**	**0.2158**	**0.2379**	**0.2539**

Table 2: Performances of Predicting Context-restricted Collaborators. Dim-20, Dim-40, and Dim-60 demonstrate the performances where embedding dimensions are set to 20, 40 and 60, respectively.



(A) **(B)**

Figure 7: Impact of Hyper Parameters. (A): impact of embedding dimensions; (B): impact of L2-norm regularization.



(A) **(B)**

Figure 8: Training Scalability. (A): time costs of R-Step and T-Step for every 100 batches, batch-size=100; (B): testing accuracy's growth.

more, to evaluate the individual impact of activeness and conservativeness, performances are also tested with only one character (either activeness or conservativeness) taken into consideration.

6.2 Context-restricted Collaborator Prediction

6.2.1 Effectiveness. The performances of all the comparison methods are summarized as Table 2, where embedding dimensions are set to be 20, 40, and 60; while, the size of recommendation, K, is set to be 5, 10, 15 and 20. According to the presented results, CEE demonstrates its superiority in predicting context-restricted collaborators, as the generated recalls are clearly higher than those of others. At the same time, all context-aware approaches show better performances than CI. Such a result is consistent with our finding in data exploration, that researchers tend to collaborate with different fellows given distinct topics, making it necessary to equip the recommendation algorithm with context-awareness. With further analysis of the experimental results, the following three insights can be drawn.

Firstly, both CEE and TSE generate better performances, as compared with SNE. The most notable distinction among these methods is that both CEE and TSE extract topics' semantics while learning the recommendation model, which in turn provide more meaningful representations for collaborations' contexts; whereas, such an operation is not carried out in SNE. Therefore, it is desirable of incorporating topics' semantic extraction into the learning of recommendation model.

Secondly, it is observed that CEE produces even better performances than TSE. Different from TSE, where researchers' and topics' representations are learned independently, representation learning are conducted with mutual adaption in CEE: while learning topics' representations, the coincident researchers are taken into account, and vice-versa for the learning of researchers' representations. In this way, mutual-dependency between researchers' and topics' co-occurrence relationships are naturally preserved, which in turn contributes to a more accurate recommendation.

Thirdly, LR generates the lowest performances among all the comparison methods. Despite that topics' semantics are utilized by LR, its loosely-coupled structure (i.e., the representation and recommendation learning are conducted in two separate steps) prevents the extracted semantics from being adapted by researchers' collaborations, which largely impairs its performances. Similar observations were also reported in previous works, e.g., [30, 31].

6.2.2 Impact of Hyper Parameters. The impacts of hyper parameters, Dim and λ, are demonstrated in Figure 7 (A) and (B), respectively. According to the presented results, superior performance is achieved where Dim = 40 and $\lambda = 0.1$. In addition, it is obvious that Dim exerts greater impacts on CEE's performance, as the generated results are greatly diversified for different number of dimensions. A larger Dim will introduce stronger discriminative power, but it might also lead to over-fitting; as such, a proper value needs to be carefully selected through grid search.

6.2.3 Scalability. The scalability of training CEE is demonstrated in Figure 8, where time costs of R and T steps (in Alg. 1) are tested with growing numbers of dimension. The time costs of processing 100 training batches are around a few seconds, and grows

	Recall@K			
	5	10	15	20
New	0.0228	0.0331	0.0388	0.0433
Exist	0.3156	0.3695	0.3968	0.4136
N2E Ratio	**7.22%**	**8.96%**	**9.78%**	**10.47%**

Table 3: CEE's Performances w.r.t. New and Exist. New: new collaborators; Exist: the collaborators whom have worked with before.

linearly with Dim. Besides, the testing accuracy of ϕ_{ij}^a converges with roughly $5{\times}10^5$ batches processed.

6.3 New Collaborator Prediction

6.3.1 New vs. Exist. We first evaluate CEE's performance w.r.t. different types of collaborators, whose results are shown in Table 3. Compared with Exist (the ones whose collaborations on different contexts are included in the training set), CEE's performance is significantly lower when the collaborator to be predicted is a new one. Such a result is consistent with our analysis in Section 1 and 3, which further justifies the necessary of HFM.

6.3.2 Effectiveness and Effect of Characters. HFM's performance in predicting new collaborators are demonstrated in Figure 9 (A), whose generated recommendation is significantly more accurate than that of CEE. To further study the effect of each individual character, two comparison methods, conservativeness-adapted CEE and activeness-adapted CEE, are introduced, which adapts CEE's recommendation based on conservativeness and activeness, respectively. Particularly, recommendations of the above methods are made w.r.t. $(\phi_{ij}^a)^{\sigma_j}$ and $\omega_j \phi_{ij}^a$, instead of $\omega_j (\phi_{ij}^a)^{\sigma_j}$ in HFM where both characters are taken into account. Compared with the raw output of CEE (Raw), the recommendation performances are clearly improved with the adoption of both characters. Such an observation is consistent with our finding in data exploration (Section 3). Moreover, it can be observed that activeness plays a more important role in predicting new collaborators, as its performance is relatively higher than that of the conservativeness-adapted one. Despite that activeness is not as closely correlated with the new collaboration ratio as conservativeness (demonstrated in Section 3), a higher activeness suggests more intense academic collaborations of a researcher, thus increasing the overall chances of generating new collaborators.

6.3.3 Impact of Hyper Parameter. The impact of hyper parameter τ is evaluated in Figure 9 (B), whose value is tested for 0.01, 0.2, 0.5, and 1.0. According to the presented result, HFM achieves the best performance when τ is set to 0.2. While increasing the value of τ, the conservativeness σ will be further confined around 1, forcing the resulted model purely rely on activeness, and inevitably impairing its performance. Meanwhile, a too small τ is also inadvisable, as the generated performance is even worse. A possible explanation about the degraded performance is that a too small τ will make the model over-emphasized on conservativeness, which in turn leads to less accurate estimation of σ and ω.

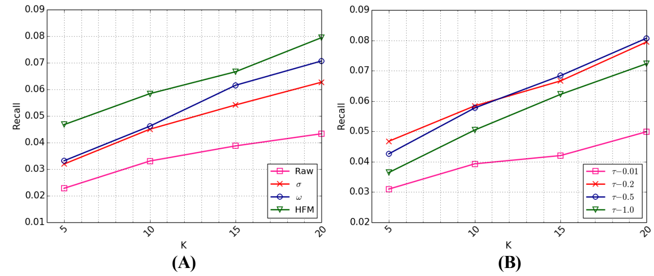


Figure 9: Prediction of New Collaborators. (A): performances of Raw CEE (Raw), Conservativeness-adapted CEE (σ), Activeness-adapted CEE (ω), and HFM. (B): impact of τ.

6.4 Summary

Major points of the experimental study are summarized as follows.

- The exploitation of topics' underlying semantics, the incorporation of researchers' and topics' mutual-dependency, and the tight-coupled structure for representation and recommendation learning are crucial for capturing researchers' context-aware collaboration tendencies. All these factors jointly lead to CEE's superior performance in the experiments.
- HFM significantly improves raw CEE's performance on predicting new collaborators; besides, both characters, activeness and conservativeness, contribute to better predictions of new collaborators, thus necessary to be incorporated in HFM.

7 RELATED WORK

In this section, related works are reviewed from two perspectives: academic data exploitation and context-aware recommendation.

Academic Data Exploitation. With the rapid growth of academic data, a large variety of applications have been developed to facilitate its exploitation. Representative works include (but not limited to) collaborator recommendation [3, 11, 23, 27], reference recommendation [7, 8, 20, 22, 32], scholars' profiling [28, 29] and academic impact estimation [5, 12]. However, existing works on collaborator recommendation would simply provide context-independent results, whereas unable to find collaborators of specific research topics. Although some other representation learning [4] or general-purpose recommendation [30, 31] works can be adapted for the task of CACR, their performances of recommending potential new collaborators are limited because of two defects: 1) the mutual-dependency between researchers and topics is ignored, and 2) researchers' inherent characters, activenesses and conservativeness, are not taken into consideration.

Context-aware Recommendation. Context-aware recommendation system [1, 25] has been intensively studied in recent years. Particularly, the context-awareness means that recommendations are generated w.r.t. specific contextual information, e.g., time and spatial location. Such a property is highly emphasized for mobile applications [14, 25, 35], as users' preferences are easily influenced by their environments. Meanwhile, context-awareness is also desirable for academic collaborator recommendation, as researchers usually look for collaborators with specific topics, and researchers' collaboration tendencies are greatly diversified given different research topics. Notice that other contextual information, like affiliation or age, may also plays a role in making even better recommendation. The additional contextual information can be processed in a similar way as topics (as discussed in Remark 4.1), whose impact will be further studied in our future work.

8 CONCLUSION

In this work, we propose Context-aware Academic Collaborator Recommendation (CACR), which produces high-potential new collaborators for the required research topics. In CACR, we design the Collaborative Entity Embedding network, which jointly represents researchers and research topics as compact vectors based on their co-occurrence relationship, whereby capturing researchers' context-aware collaboration tendencies and topics' underlying semantics. Furthermore, we develop the Hierarchical Factorization Model, where researchers' activenesses and conservativenesses are effectively exploited to generate high-quality recommendation result. Experiments on large-scale academic data verify the effectiveness of our proposed approaches.

Acknowledgment. The work is partially supported by the Hong Kong RGC GRF Project 16207617, G-HKUST603/14 , Hong Kong ITC ITS/212/16FP grant,National Grand Fundamental Research 973 Program of China under Grant 2014CB340303, the National Science Foundation of China (NSFC) under Grant No. 61729201, Science and Technology Planning Project of Guangdong Province, China, No. 2015B010110006, Webank Collaboration Research Project, and Microsoft Research Asia Collaborative Research Grant

REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *RecSys 2015*.
[2] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS 2013*.
[3] H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles. Collabseer: a search engine for collaboration discovery. In *JCDL 2011*.
[4] T. Chen and Y. Sun. Task-guided and path-augmented heterogeneous network embedding for author identification. In *WSDM 2017*.
[5] Y. Dong, R. A. Johnson, and N. V. Chawla. Will this paper increase your h-index?: Scientific impact prediction. In *WSDM 2015*.
[6] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *KDD 2016*.
[7] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles. Context-aware citation recommendation. In *WWW 2010*.
[8] W. Huang, S. Kataria, C. Caragea, P. Mitra, C. L. Giles, and L. Rokach. Recommending citations: translating papers into references. In *CIKM 2012*.
[9] D. Kinga and J. B. Adam. Adam: A method for stochastic optimization. In *ICLR 2015*.
[10] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *ICML 2014*.
[11] J. Li, F. Xia, W. Wang, Z. Chen, N. Y. Asabere, and H. Jiang. Acrec: a co-authorship based random walk model for academic collaboration recommendation. In *WWW 2014*.
[12] L. Li and H. Tong. The child is father of the man: Foresee the success at the early stage. In *KDD 2015*.
[13] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu. Learning entity and relation embeddings for knowledge graph completion. In *AAAI 2015*.
[14] Q. Liu, H. Ma, E. Chen, and H. Xiong. A survey of context-aware mobile recommendations. *International Journal of Information Technology & Decision Making*, 12(01):139–172, 2013.
[15] Z. Liu and L. Chen. Worker recommendation for crowdsourced q&a services: A triple-factor aware approach. *VLDB 2017*.
[16] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS 2013*.
[17] N. M. Nasrabadi. Pattern recognition and machine learning. *Journal of electronic imaging*, 16(4):049901, 2007.
[18] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *KDD 2014*.
[19] M. Qu, J. Tang, J. Shang, X. Ren, M. Zhang, and J. Han. An attention-based collaboration framework for multi-view network representation learning. In *CIKM 2017*.
[20] X. Ren, J. Liu, X. Yu, U. Khandelwal, Q. Gu, L. Wang, and J. Han. Cluscite: Effective citation recommendation by information network-based clustering. In *KDD 2014*.
[21] S. Rendle. Factorization machines. In *ICDM 2010*.
[22] T. Strohman, W. B. Croft, and D. Jensen. Recommending citations for academic papers. In *SIGIR 2007*.
[23] Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han. Co-author relationship prediction in heterogeneous bibliographic networks. In *ASONAM 2011*.
[24] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB 2011*.
[25] Y. Sun, N. J. Yuan, Y. Wang, X. Xie, K. McDonald, and R. Zhang. Contextual intent tracking for personal assistants. In *KDD 2016*.
[26] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *WWW 2015*.
[27] J. Tang, S. Wu, J. Sun, and H. Su. Cross-domain collaboration recommendation. In *KDD 2012*.
[28] J. Tang, L. Yao, D. Zhang, and J. Zhang. A combination approach to web user profiling. *TKDD*, 5(1):2, 2010.
[29] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: extraction and mining of academic social networks. In *KDD 2008*.
[30] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *KDD 2011*.
[31] H. Wang, N. Wang, and D.-Y. Yeung. Collaborative deep learning for recommender systems. In *KDD 2015*.
[32] X. Yu, Q. Gu, M. Zhou, and J. Han. Citation prediction in heterogeneous bibliographic networks. In *SDM 2012*.
[33] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma. Collaborative knowledge base embedding for recommender systems. In *KDD 2016*.
[34] H. Zhao, Q. Yao, J. Li, Y. Song, and D. L. Lee. Meta-graph based recommendation fusion over heterogeneous information networks. In *KDD 2017*.
[35] H. Zhu, E. Chen, H. Xiong, K. Yu, H. Cao, and J. Tian. Mining mobile user preferences for personalized context-aware recommendation. *TIST*, 5(4):58, 2015.