TAPER: A Contextual Tensor-Based Approach for Personalized Expert Recommendation

Hancheng Ge, James Caverlee, Haokai Lu
Department of Computer Science and Engineering
Texas A&M University
{hge,caverlee,hlu}@cse.tamu.com

ABSTRACT

We address the challenge of personalized recommendation of high-quality content producers in social media. While some candidates are easily identifiable (say, by being "favorited" many times), there is a long-tail of potential candidates for whom we have little evidence. Through careful modeling of contextual factors like the geo-spatial, topical, and social preferences of users, we propose a tensor-based personalized expert recommendation framework that integrates these factors for revealing latent connections between homogeneous entities (e.g., users and users) and between heterogeneous entities (e.g., users and experts). Through extensive experiments over geo-tagged Twitter data, we find that the proposed framework can improve the quality of recommendation by over 30% in both precision and recall compared to the state-of-the-art.

1. INTRODUCTION

Recommender systems are a cornerstone of how we engage online – by impacting the media we consume, the friends we connect with, and the products we purchase. A typical assumption in many recommender systems is to focus on *specific items* like movies, songs, or books as the basis of recommendation. In a separate direction, there are efforts to focus on *high-quality content producers* rather than specific items [21, 22]. These content producers – like creators of highly-rated Spotify playlists, Amazon's top reviewers, or media curators on platforms like Tumblr, Facebook, and Twitter – can potentially serve as conduits to high-quality curated items. Indeed, previous research has shown that in some cases item-based recommenders can be improved by biasing the underlying models toward the opinions of these "experts" [2].

While some high-quality content producers are easily identifiable (say, by being "favorited" or starred many times), there is a long-tail of potential candidates for whom we have little evidence. Hence, a natural question is whether we can identify these high-quality content producers – whom we shall refer to as *experts* in the rest of this paper – and recommend them to the right people. Such *personalized expert recommendation* faces a number of key challenges, though. First, many existing works have aimed at uncovering expert users in online systems – e.g., [4, 7, 9, 28, 32] – but typically without an emphasis on personalized recommendation. That is, these efforts have often attempted to explore general topic experts with broad appeal, e.g., the best doctors in Seattle or the top engineers in a certain field, rather than connecting users with personal experts. Second, personalized expert recommendation faces extreme sparsity since few users provide feedback on the quality of content producers. Third, there are typically complex relationships between users, candidate experts, and topics of interest.

RecSys '16, September 15-19, 2016, Boston , MA, USA

DOI: http://dx.doi.org/10.1145/2959100.2959151

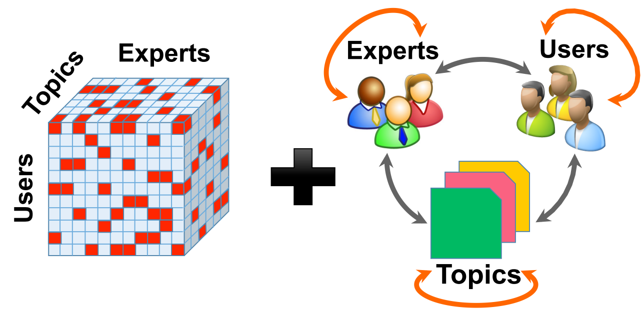


Figure 1: Personalized expert recommendation with contextual factors: augmenting the sparse tensor (left) with contextual factors (right) like the geo-spatial, topical, and social relationships between homogeneous entities (e.g., users and users) and between heterogeneous entities (e.g., users and experts).

Hence, we aim in this paper to tackle these challenges via a personalized expert recommendation framework called **TAPER** – for **T**ensor-based **A**pproach for **P**ersonalized **E**xpert **R**ecommendation. This proposed approach inherits the advantages of traditional recommender systems by making personalized recommendations based on the history of actions by similar users. In this way, specific personal experts can be recommended to individuals, rather than relying on globally-recognized (and less personalized) ones. While matrix factorization approaches have shown success in mitigating sparsity [15, 18], ultimately they are restricted to two-dimensional data (e.g., a user-expert matrix). In contrast, user preferences for experts may be impacted by many contextual factors including the topic of interest, the location of the user (and possibly of the expert), as well as social connections among users and experts, among many others. As illustrated in Figure 1, these user-expert-topic preferences naturally suggest a tensor-based approach where these multiple and varied relationships may augment the sparse tensor (on the left) by considering the relationships (on the right) between both homogeneous entities (e.g., users and users, experts and experts) and the relationships between heterogeneous entities (e.g., users and experts, topics and experts).

Through the TAPER framework, we explore questions like: What kinds of contextual factors impact preferences for personalized experts? How can these contextual factors be integrated into a tensor-

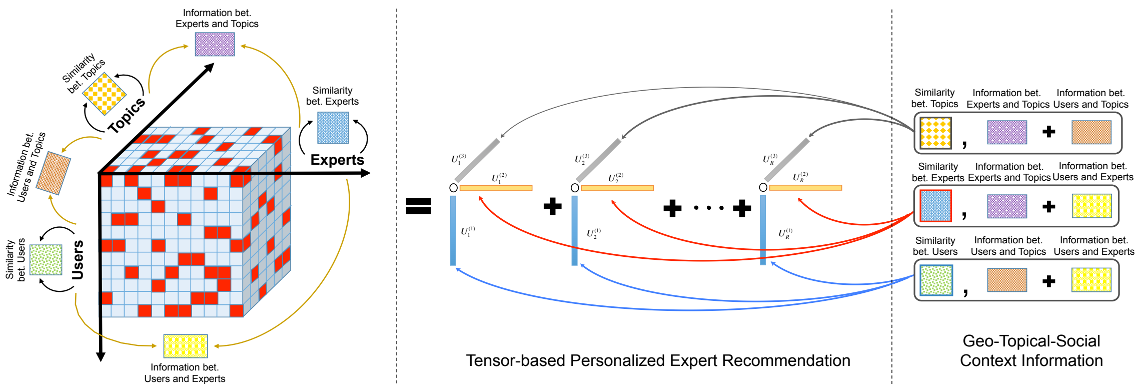


Figure 2: Overview of the proposed tensor-based personalized expert recommendation framework.

based personalized expert recommender? Do these factors result in higher quality recommendations than state-of-the-art methods? And are some contextual factors more important than others?

Related Work. Expert finding has been widely studied for many years. For instance, Weng et al. [28] proposed a PageRank-based approach to identify topic experts by applying both topical similarity between users and social link structure. Ghosh et al. [9] proposed the Cognos system to find topic experts by relying on Twitter lists. Zhang et al. [32] identified top experts in a Java forum by applying link analysis approaches such as PageRank and HITS. Hu et al. [12] proposed a more personalized recommendation by considering network semantic information, in addition to network topological measures for expert recommendation. Of course, there are many other efforts, including [10, 19, 25].

One promising approach for recommenders is to cast the problem as a matrix factorization problem, in which user preferences may be projected into a lower dimensional embedding space [15, 17, 29, 31]. In recent years, tensor factorization models – that are highly suitable for multiway data analysis as in our domain – have been applied in applications like tag recommendation [24], user group detection [16], link prediction [8], and anomaly detection [26]. There are two widely used low-rank decompositions – CANDECOMP/PARAFAC (CP) and the Tucker decomposition [14]. Karatzoglou et al. [13] studied multi-dimensional recommendations by leveraging contextual information to build a User-Item-Context tensor model. Hidasi and Tikk [11] developed an Alternating Least Squares (ALS) based tensor factorization approach for context-aware recommendations. Bhargava et al. [5] applied a tensor factorization-based approach to provide collaborative recommendations for points of interest (POI) involving multi-dimensions such as locations, activities and time. Similarly, Lu et al. [17] introduced a matrix-factorization approach for personalized expert recommendation, but restricted to two-dimensional data (e.g., a user-expert matrix) and only considering user's geo-spatial preferences on experts. In contrast, this work is the first to investigate tensor factorization for personalized expert recommendation by integrating user's preferences on experts from geo-topical-social contexts as well as valuable relationships between users, experts and topics, and to explore the impact of these contexts on such an approach.

2. PROBLEM STATEMENT

Let $\boldsymbol{L} = \{u_1, u_2, \ldots, u_N\}$ be a set of users where N is the total number of users, and $\boldsymbol{E} = \{e_1, e_2, \ldots, e_M\}$ be a set of experts where M is the total number of experts and $\boldsymbol{E} \in \boldsymbol{L}$. An expert e_i may have expertise in multiple topics expressed as $\boldsymbol{T} = \{t_1, t_2, \ldots, t_K\}$ where K is the total number of topics. A user may personally prefer some experts rather than others based upon the user's personal interest in their expertise. For example, Andy may prefer Bella in the topic of "Python programming", but have no opinion on Chris who may be a better Python developer. We denote the personal preferences of users towards experts in topics as a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N\times M\times K}$ where element $\boldsymbol{\mathcal{X}}(i,j,k)$ is binary, representing whether a user u_i prefers an expert e_j in a topic t_k.

We define the *personalized expert recommendation problem* as: Given a set of users $\boldsymbol{L}$ with partially observed preferences denoted as a tensor $\boldsymbol{\mathcal{T}}$ on experts $\boldsymbol{E}$ over topics $\boldsymbol{T}$, our goal is to recommend the top-k relevant experts to a user u_i.

Basic Recommendation by Tensor Factorization. As a natural first step, we can tackle this problem with a basic recommendation framework using tensor factorization. Let $U^{(1)} \in \mathbb{R}^{N\times R}$, $U^{(2)} \in \mathbb{R}^{M\times R}$ and $U^{(3)} \in \mathbb{R}^{K\times R}$ be latent factor matrices for users, experts, and topics, respectively, where $R \ll min(N, M, K)$ is the number of latent factors as the rank of a tensor. The basic tensor-based expert recommendation model can be defined as:

$$\begin{aligned} \underset{U^{(n)},\boldsymbol{\mathcal{X}}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{\mathcal{X}} - [\![U^{(1)}, U^{(2)}, U^{(3)}]\!]\|_F^2 + \frac{\lambda}{2}\sum_{n=1}^{3}\|U^{(n)}\|_F^2, \\ \text{subject to} \quad & \boldsymbol{\Psi} \circ \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{T}}, U^{(n)} \geq 0, n = 1, 2, 3 \end{aligned} \tag{1}$$

where $\boldsymbol{\mathcal{X}}$ denotes the complete preferences of users towards experts across topics, $\boldsymbol{\mathcal{T}}$ denotes the observed user's preferences on experts, $\|U^{(n)}\|_F^2$ is a Tikhonov regularization term used to avoid overfitting and provide a unique solution, $\circ$ is a Hadamard product operator and $\boldsymbol{\Psi}$ is a non-negative weight tensor with the same size as $\boldsymbol{\mathcal{X}}$ with $\boldsymbol{\Psi}(i,j,k) = 1$ indicating that we observe the selection of user u_i on expert e_j in a topic t_k, $\boldsymbol{\Psi}(i,j,k) = 0$ otherwise. This basic model estimates $\hat{\boldsymbol{\mathcal{X}}}$ that approximates the original (unknown) $\boldsymbol{\mathcal{X}}$ via learning optimal latent factor matrices $\{U^{(n)}, n = 1, 2, 3\}$. For each user and topic of interest, this model can recommend a ranked list of personalized experts.

Research Challenges. While this basic tensor factorization framework provides a first step toward personalized expert recommendation, it leaves open many critical questions:

- First, since the user-expert-topic tensor is necessarily sparse (meaning that even with factorization the underlying latent factors may be of poor quality), can we augment this basic approach with additional contextual preferences for experts by considering the relationships among both homogeneous entities (e.g., users and users, experts and experts) and heterogeneous entities (e.g., users and experts, topics and experts)?
- How can we integrate these contextual preferences into a tensor-based personalized expert recommendation framework?

Table 1: Dataset Summary

Data Type	Total Number of Records
Twitter Lists	11,322
Users (list creators)	10,559
Experts (list members)	8,417
List Relationships	117,187
Sparsity	0.13%

- How effective is the proposed tensor-based framework in comparison with other state-of-the-art baselines? And which contextual preferences have the most significant impact on the quality of personalized expert recommendations?

3. THE TAPER FRAMEWORK

We turn in this section toward constructing the contextual tensor factorization framework, as illustrated in Figure 2.

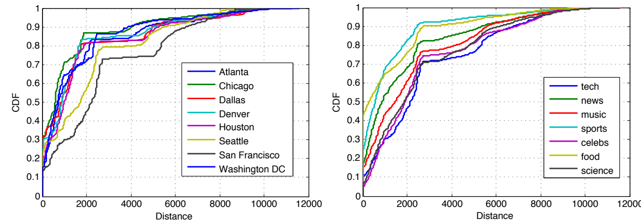
3.1 Evidence of Geo-Topical-Social Impact

We begin our investigation by examining the impact of three factors – geo-spatial, topical, and social context – on the observed preferences towards experts. Our goals in this section are to assess whether and to what degree these factors do affect how users select experts. Informed by these observations, we turn in the following to integrate them into the tensor-based factorization framework.

Geo-Tagged Twitter Lists. We adopt geo-tagged Twitter lists as evidence of the revealed preferences of users for other users. A Twitter list allows a user u_i to label another user u_j with an annotation (e.g., news, food, technology). In isolation these lists support the curation of an individual user's information stream, but in the aggregate the list labels can encode what a target user is "known-for". Many efforts have demonstrated that these labels can provide a crowdsourced expertise profile of the target user [6, 7, 9, 23]. Concretely, we use a geo-tagged Twitter list dataset containing over 12 million crowd-generated lists and 14 million geo-tagged list relationships between list creators and members. We filter the lists to only keep US-based users in topics: news, music, technology, celebrities, sports, business, politics, food, fashion, art, science, education, marketing, movie, photography, and health. The dataset is summarized in Table 1. Both list creators and list members are associated with GPS coordinates.[1] We shall refer to list creators as *users* and members in the lists as *experts*.

Geo-Spatial Context. We begin by investigating the impact of distance on the experts selected by users. Figure 3(a) shows the cumulative distribution of the average distance between a user and the experts they have labeled, aggregated for eight different cities. In general, we see that users from different cities have different levels of locality. For example, users in San Francisco are more likely to select experts from a wider geographical scope than users based in Chicago. Specifically, almost 40% of users in Chicago have an average distance to their experts within 100 miles. However, only 14% of users in SF have an average distance within 100 miles. In a similar fashion, Figure 3(b) shows the cumulative distribution of the average distance between users and the experts they have labeled for seven different topics. For a fixed distance (e.g., 100 miles), the topic *food* has the largest probability. This implies that users interested in *food* are closer to their chosen experts while users interested in a broad topic like *celebrity* instead select experts with a wider geographical scope. Hence, we can conclude that *the*

[1]For those without GPS coordinates, their locations can be estimated with their tweets by an approach previously used for check-ins and geo-tagged images [20]. In order to simplify our study, we only focus on users with GPS coordinates.

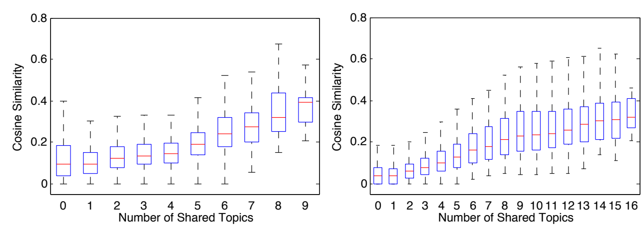


(a) Different Cities (b) Different Topics

Figure 3: The impact of distance on expert preferences by location (a) and by topic (b).

geo-spatial context of users, experts, and topics does indeed affect the preference for experts, and that these factors are impacted to varying degrees based on topic and on the particular locations of both users and experts.

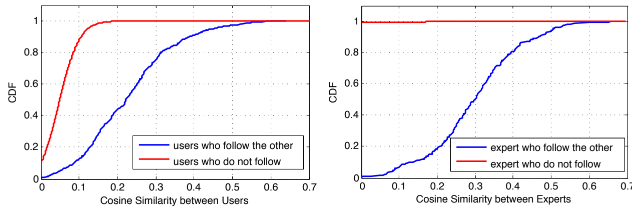
Topical Context. As we observed in Figure 3(b), different topics have different levels of locality. Here, we further investigate the impact of topical context on personalized preferences for experts. We begin by viewing each user as a vector of the experts they have selected, in which the element is 1 when the expert is listed by the user, otherwise 0. We then measure the impact of the number of shared topics between users (e.g., if two users have used three of the same topics in their lists, then we consider the number of shared topics between them to be three) on how similar are the users with respect to the experts they have selected. As we can see in Figure 4(a), users are more likely to select similar experts when they share more common topics. In other words, common topical interest impacts the choice of experts. In a similar fashion, we can view each expert as a vector of all users, in which the element is 1 when the expert is listed by the user, otherwise 0. We then measure the impact of the number of shared topics between experts (e.g., if two experts have been labeled by four of the same topics, then we consider the number of shared topics between them to be four) on how similar are the experts with respect to the users who have selected them. As we can see in Figure 4(b), experts who share more topics are more likely to be preferred by similar users. *We conclude that topical context has a strong impact on user preferences for experts.*



(a) Users-Users (b) Experts-Experts

Figure 4: The relationship between the number of shared topics and the similarity between users (a) and experts (b).

Social Context. Finally, the social connections among users and experts can be a strong indicator of shared topical interests, as well as an implicit signal that two users are more likely to be near each other. Researchers have found that social ties increase the likelihood of two users being near each other [3]. Figure 5 shows the cumulative distributions of the similarity of users and experts, respectively. Using the same approach introduced above, each user is represented as a vector of all experts; each expert is represented as a vector of all users. The cosine similarity is employed to calculate the similarity of users and experts. We can observe that users who follow the other generally have a larger similarity on selecting experts. For experts, we can see a similar pattern that those who follow the other are more likely to be selected by the same users. Through these observations, we can draw the conclusion that *social context strongly affects a user's preferences for experts.*



(a) Users-Users (b) Experts-Experts

Figure 5: The impact of social connections on selecting experts for users (a) and being selected for experts (b)

3.2 Integrating Contextual Preferences

Given this evidence of the significance of geo-topical-social context with respect to preferences on experts, a natural approach is to integrate them into the basic tensor-based expert recommendation framework as regularization terms. Intuitively, if two entities are similar, e.g., two users have similar preferences in recognizing experts across topics, the latent representations of these two entities should be similar. Hence, we can regulate latent representations of two similar entities to make them as close as possible. We denote S as a symmetric similarity matrix encoding contextual information between homogeneous entities (e.g., users and users, experts and experts), and then formally minimize the following loss function:

$$\begin{aligned}\Theta &= \frac{1}{2}\sum_{i,j} S(i,j)\|U_i^{(n)} - U_j^{(n)}\|^2 \\ &= \sum_i U_i^{(n)} D(i,i) U_i^{(n)T} - \sum_{i,j} U_i^{(n)} S(i,j) U_j^{(n)T} \\ &= \mathrm{tr}(U^{(n)T}(D-S)U^{(n)}) \\ &= \mathrm{tr}(U^{(n)T}\mathcal{L}U^{(n)})\end{aligned}$$

where $U_i^{(n)}$ is the ith row of the factor matrix $U^{(n)}$ for the nth-mode of a tensor $\boldsymbol{\mathcal{X}}, n \in \{1,2,3\}$, $tr(\cdot)$ is denoted as the matrix trace, D is a diagonal matrix with the element $D(i,i) = \sum_j S(i,j)$, and $\mathcal{L} = D - S$ is the graph Laplacian of the similarity matrix S. For the contextual information between heterogeneous entities (e.g., users and experts, topics and experts), we denote $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ as matrices encoding the contextual information between users and experts, users and topics and experts and topics, respectively. We then regulate them by directly adding regularization terms $\|\boldsymbol{A} - U^{(1)}U^{(2)^T}\|_F^2$, $\|\boldsymbol{B} - U^{(1)}U^{(3)^T}\|_F^2$, and $\|\boldsymbol{C} - U^{(2)}U^{(3)^T}\|_F^2$ into the basic framework. In order to simplify the parameter tuning, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are normalized to the same scale.

After integrating all regularization terms into Eq. (1), TAPER aims to solve the following optimization problem [2]:

$$\begin{aligned}\underset{U^{(n)},\mathcal{X}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{\mathcal{X}} - [\![U^{(1)}, U^{(2)}, U^{(3)}]\!]\|_F^2 + \frac{\lambda}{2}\sum_{n=1}^{3}\|U^{(n)}\|_F^2 \\ & + \frac{\gamma}{2}\sum_{n=1}^{3}\mathrm{tr}(Z^{(n)T}\mathcal{L}_n Z^{(n)}) + \frac{\beta}{2}(\|\boldsymbol{A} - U^{(1)}U^{(2)^T}\|_F^2 \\ & + \|\boldsymbol{B} - U^{(1)}U^{(3)^T}\|_F^2 + \|\boldsymbol{C} - U^{(2)}U^{(3)^T}\|_F^2), \\ \text{subject to} \quad & \boldsymbol{\Psi} \circ \boldsymbol{\mathcal{X}} = \mathcal{T}, U^{(n)} = Z^{(n)} \geq 0, n = 1,2,3\end{aligned} \quad (2)$$

where γ controls the weight of the contextual information between homogeneous entities (e.g., users and users, experts and experts), and β controls the weight of the contextual information between heterogeneous entities (e.g., users and experts, topics and experts). But how specifically should we model these contextual preferences for integration into the tensor-based framework? In the following, we consider geo-spatial, topical, and social preferences in turn.

3.2.1 Modeling Geo-Spatial Preferences

We first aim to model geo-spatial context for integration into the tensor-based factorization approach.

Relationships between Homogeneous Entities: Supported by the data-driven observations of the previous section and following Tobler's First Law of Geography [27] – which asserts that near things are more related than distant things – we propose to model user preferences as a function of distance. Concretely, we consider three graphs over homogeneous entities $\mathcal{G}_G^{uu}$, $\mathcal{G}_G^{ee}$ and $\mathcal{G}_G^{tt}$ for users, experts, and topics, respectively. For each graph, we treat "nearby" nodes as more alike if they are geographically closer. In this way, preferences on experts may be propagated to these nearby nodes in the tensor-based factorization approach.

To begin, consider the user-user graph $\mathcal{G}_G^{uu}$, where nodes are users and edges capture the affinity between users. We can define the adjacency matrix $\boldsymbol{H}_G \in \mathbb{R}^{N\times N}$ as:

$$\boldsymbol{H}_G(u_i, u_j) = exp(-\frac{Dist(u_i, u_j)^2}{2\alpha^2}),$$

where α is a decay constant (which we experimentally set to 20 miles) and $Dist$ is to measure the geographic distance between two users by using the Haversine formula. The affinity $\boldsymbol{H}_G(u_i, u_j)$ approaches one as two users are nearer each other.

Similarly, the expert-expert graph $\mathcal{G}_G^{ee}$ can be constructed, where nodes are experts and here edges represent the spatial similarity of pairs of experts. Rather than purely measuring the geographic nearness of two experts (which does not take into account the recognizability of experts in various locations), we consider how the users who have labeled those experts are distributed. The intuition is that experts who are preferred have higher popularity in the location of a user. Let $l(u_i) \in \mathbb{L}$ be a location of user u_i and $\phi_{e_i}^{l(u_i)}$ as the spatial popularity be the number of users selecting expert e_i in the location of user u_i across topics. The adjacency matrix $\boldsymbol{V}_G \in \mathbb{R}^{M\times M}$ can be defined as:

$$\boldsymbol{V}_G(e_i, e_j) = exp(-\frac{Dist(l(e_i), l(e_j))^2}{2\alpha^2} * \sum_{u_i\in U}(\phi_{e_i}^{l(u_i)} - \phi_{e_j}^{l(u_i)})^2),$$

where the first part in the exponential is to calculate the deviation of geo-spatial preferences between experts (so that experts who are nearby are considered "closer"), the second part is to consider the difference of their spatial popularities over all locations of users (to capture the nearness in terms of who is interested in them).[3]

The third graph – the topic-topic graph $\mathcal{G}_G^{tt}$ – has topics as nodes and edges that represent the geo-spatial correlation between topics. Intuitively, selecting experts could be impacted by the choice of topic. The spatial preference in the topic *food* is much more local than the topic *technology*. We aim to capture the similarity of spatial preference between topics. Let $d_{u_i}^{t_i}$ be the average distance between user u_i and a set of experts $E_{u_i}^{t_i}$ he/she recognizes in a topic t_i. The empirical distribution of spatial preference for a topic t_i can be obtained from calculations $\{d_{u_i}^{t_i}, u_i \in U\}$, which is denoted as Υ_{t_i}. We then apply the Kullback-Leibler divergence $D_{KL}(\cdot\|\cdot)$ to measure the closeness of empirical distributions between topics. Hence, we can define the adjacency matrix $\boldsymbol{W}_G \in \mathbb{R}^{K\times K}$ as:

$$\boldsymbol{W}_G(t_i, t_j) = 1 - D_{KL}(\Upsilon_{t_i}\|\Upsilon_{t_j}),$$

[2] An open question is how can we efficiently estimate the solution to this optimization problem since there is no closed form solution? In the supplementary material [1], we show how to solve this problem based on ADMM (Alternating Direction Method of Multipliers).

[3] In this case, we discretize the continental U.S. surface with a 1° by 1° geodesic grid, so we can map the location of a user (GPS coordinate) to a discrete region.

where $\boldsymbol{W}_G$ will approach to 1 if two topics have similar empirical distributions of spatial preferences.

Relationships between Heterogeneous Entities: In addition to the geo-spatial correlations between users, experts, and topics themselves, we can additionally consider the relationships across heterogeneous entities. Specifically, we consider the relationships between users and experts, as well as between users and topics.

First, we propose to leverage the geo-spatial preferences between users and experts so that users are more likely to select experts who have a high popularity in the location of this user. The intuition is that the local popularity of an expert can be considered as a prior on what a user would prefer. For example, given a new user in Seattle with no expert preferences, we can default to a locally popular expert like Jeff Bezos. Concretely, we propose to improve the learning of latent matrices of users and experts as:

$$\mathcal{F}_{\boldsymbol{A}_G} = \|\boldsymbol{A}_G - U^{(1)}U^{(2)^T}\|_F^2,$$

where $\boldsymbol{A}_G$ is the adjacency matrix in which an element indicates the spatial popularity of an expert in the location of a user. By minimizing $\mathcal{F}_{\boldsymbol{A}_G}$, the recommender will prefer locally popular experts.

Similarly, the geo-spatial preferences between users and topics can also be leveraged so that a user is more likely to select experts on topics that have high popularity in the location of this user. Formally, we have the latent matrices of users and topics as:

$$\mathcal{F}_{\boldsymbol{B}_G} = \|\boldsymbol{B}_G - U^{(1)}U^{(3)^T}\|_F^2,$$

where $\boldsymbol{B}_G$ is the adjacency matrix in which an element indicates the spatial popularity of a topic in the location of a user. Again, by minimizing $\mathcal{F}_{\boldsymbol{B}_G}$, the recommender will prefer experts on topics that are more popular locally.

3.2.2 Modeling Topical Preferences

Beyond geo-spatial preferences, we next turn to models of topical preference. As we have observed in Section 3.1, the topical context can influence a user's preference for experts. Intuitively, users who have similar interests on topics tend to have similar preferences on experts.

Relationships between Homogeneous Entities: We begin by considering two graphs – one for the user-user graph $\mathcal{G}_T^{uu}$ and the other for the expert-expert graph $\mathcal{G}_T^{ee}$ in terms of topical preference.

In this case, the user-user graph $\mathcal{G}_T^{uu}$ has nodes that represent users but now edges represent the similarity between users in terms of topical preferences (rather than in terms of distance as in the geo-spatial case in the previous section). Inspired by the work [5], we define the adjacency matrix $\boldsymbol{H}_T$ as:

$$\boldsymbol{H}_T(u_i, u_j) = \frac{|T_{u_i} \bigcap T_{u_j}|}{|T_{u_i} \bigcup T_{u_j}|} * exp(\sum_{t \in T}(o_{u_i}^t - o_{u_j}^t)P_t),$$

where T_{u_i} is the set of topics a user u_i is interested in by labeling the lists with keywords related to these topics, $o_{u_i}^t$ denotes the number of experts a user u_i labeled in a topic t, and P_t denotes the probability of being interested in a topic t, which can be formally defined as $P_t = (\sum_{u \in U} o_u^t)/(\sum_{t \in T}\sum_{u \in U} o_u^t)$. The first part in $\boldsymbol{H}_T$ is used to measure how common two users share topics; the second part is applied to quantitively evaluate the similarity of their interests across all topics.

For the second graph – the expert-expert graph $\mathcal{G}_T^{ee}$ – we have nodes as experts and edges representing the similarity of experts towards their expertise. The adjacency matrix $\boldsymbol{V}_T$ is defined as:

$$\boldsymbol{V}_T(e_i, e_j) = \frac{|T_{e_i} \bigcap T_{e_j}|}{|T_{e_i} \bigcup T_{e_j}|} * exp(\sum_{t \in T}(\delta_{e_i}^t - \delta_{e_j}^t)P_t),$$

where T_{e_i} is the set of topics in which an expert e_i have expertise labeled by users, and $\delta_{e_i}^t$ is denoted as the number of times an expert e_i labeled by users in a topic t. The first part in $\boldsymbol{V}_T$ is used to measure how common two experts have the same expertise; the second part is applied to quantitively evaluate the similarity of their expertise across all topics.

Relationships between Heterogeneous Entities: In addition to the topical correlations between users and experts, themselves, we additionally consider the relationships across users and topics, as well as across experts and topics. Since each Twitter list is labeled with certain labels, the topic preferences of a user can be revealed by aggregating all of the labels he/she applies in the lists. As a result, we propose to leverage a user's topical preferences to improve the learning of latent matrices of users and topics as follows:

$$\mathcal{F}_{\boldsymbol{B}_T} = \|\boldsymbol{B}_T - U^{(1)}U^{(3)^T}\|_F^2,$$

where $\boldsymbol{B}_T$ is the affinity matrix where an element indicates whether a user is interested in certain topic by applying keywords related to this topic in her lists. Our goal is to minimize $\mathcal{F}_{\boldsymbol{B}_T}$ so that a user is more likely to select experts who have expertise on topics of interest to this user.

Meanwhile, an expert's expertise can be also found by aggregating all of the labels in all the lists this expert appears. Therefore, we leverage the expert's topic preferences to improve the learning of latent matrices of experts and topics as illustrated below:

$$\mathcal{F}_{\boldsymbol{C}_T} = \|\boldsymbol{C}_T - U^{(2)}U^{(3)^T}\|_F^2,$$

where $\boldsymbol{C}_T$ is the affinity matrix in which the element indicates the number of times that an expert has been recognized by users with respect to certain topics. Through minimizing $\mathcal{F}_{\boldsymbol{C}_T}$, an expert who has higher recognition in a topic is more likely to be selected by users who are interested in this topic.

3.2.3 Modeling Social Preferences

Finally, we consider how to model the social preferences of users with respect to personalized expert preferences. Concretely, we consider the social connections among users and experts:

Relationships between Homogeneous Entities: Similar to our previous efforts, we construct a graph $\mathcal{G}_S^{uu}$ in which nodes represent users and edges represent the pairwise similarity of users in terms of their social preferences, and a graph $\mathcal{G}_S^{ee}$ in which nodes represent experts and edges represent the pairwise similarity of experts with respect to their social connections. By applying the Jaccard coefficient, the adjacency matrix $\boldsymbol{H}_S$ for users is defined as:

$$\boldsymbol{H}_S(u_i, u_j) = \frac{|F_{u_i} \bigcap F_{u_j}|}{|F_{u_i} \bigcup F_{u_j}|},$$

where F_{u_i} represents the set of users u_i follows. As a similar fashion, we define the adjacency matrix $\boldsymbol{V}_S$ for experts as:

$$\boldsymbol{V}_S(e_i, e_j) = \frac{|F_{e_i} \bigcap F_{e_j}|}{|F_{e_i} \bigcup F_{e_j}|},$$

where F_{e_i} represents the set of users e_i follows. The intuition behind $\boldsymbol{H}_S$ and $\boldsymbol{V}_S$ is that users/experts who share more friends are more likely have similar behaviors on selecting experts/being selected by users.

Relationships between Heterogeneous Entities: In order to take advantage of the social connections between different entities (e.g., user-expert) into the expert recommendation, we leverage the following social relationships from a user to an expert to improve the learning of latent matrices of users and experts, as shown below:

$$\mathcal{F}_{\boldsymbol{A}_S} = \|\boldsymbol{A}_S - U^{(1)}U^{(2)^T}\|_F^2, \tag{3}$$

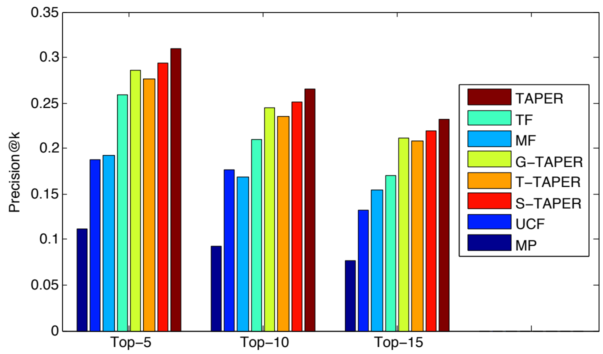


Figure 6: Precision@k: Comparing TAPER versus Alternative Methods.

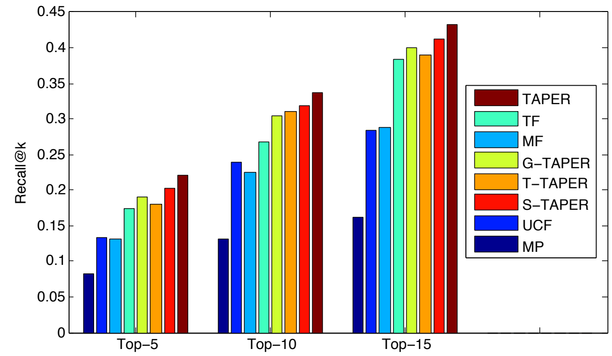


Figure 7: Recall@k: Comparing TAPER versus Alternative Methods.

where $\boldsymbol{A}_S$ is a matrix where the element indicates if a user follows an expert. In this way, the recommender can prefer experts who are followed by this user through minimizing $\mathcal{F}_{\boldsymbol{A}_S}$. Note that this is a very strong signal, and unlikely to be present for most users.

4. EXPERIMENTS

In this section, we investigate (i) the effectiveness of TAPER versus alternatives; (ii) the impact of specific contextual factors (e.g., geo-spatial, topical, and social); (iii) how incorporating evidence of experts for whom a user is *not interested* affects recommendation quality; and (iv) the impact of the amount of training data.

4.1 Experimental Setup

To evaluate the performance of the proposed framework, we randomly split experts for a user into 50% for training and 50% for testing. For latent factor dimension, we empirically choose 20 for all methods after testing various settings {5, 10, 20, 30, 40, 50, 100} for a tradeoff between accuracy and the computational cost. For the number of experts that a user does not pick up in the Twitter lists, we empirically select 350 through all experiments we conducted. The effects of this number will be further discussed in Section 4.3. Three positive parameters are involved in the experiments: λ, γ and β in Eq. (2). λ is the regularization parameter used to avoid overfitting. γ is to control the contribution of contextual information between homogeneous entities. β is to control the contribution of contextual information between heterogeneous entities. As a common way, we employ the cross-validation to tune these parameters.[4] Concretely, we empirically set $\lambda = 0.1$, $\gamma = 0.1$ and $\beta = 0.01$ for general experiments, respectively. Their effects on the performance of the proposed framework is evaluated in [1].

We adopt Precision@k and Recall@k as our evaluation metrics. Precision@k represents the percentage of correctly recommended experts out of the top-k recommendations; Recall@k represents the percentage of experts emerging in the top-k recommendations. Both of them have been widely used to evaluate the quality of recommendation. In our experiments, we test for k at 5, 10, and 15.

4.2 Baselines

We consider seven baselines in addition to the proposed TAPER approach. The first three baseline are classical recommender system approaches:

- *Most Popular (**MP**)*: This baseline recommends the most listed experts in a topic to all users.
- *User-based Collaborative Filtering (**UCF**)*: We adopt a user-based recommendation framework [33] to recommend personal experts, which discovers user's implicit preferences towards experts by aggregating similar users.
- *Matrix Factorization (**MF**)*: This baseline computes a user's preferences on experts for each topic by a pair-wise latent matrix factorization model trained by stochastic gradient descent [17].

The next four baselines are simplified variants of the proposed TAPER approach, all building on tensor factorization:

- *Tensor Factorization (**TF**)*: In analogy to matrix factorization, tensor factorization computes each user's preferences for experts by considering users, experts, and topics simultaneously. This basic tensor factorization model corresponds to Eq. (1).
- *Geo-based TAPER (**G-TAPER**)*: This model is a variant of the basic tensor factorization model, but only integrates geo-spatial preferences: that is, $\boldsymbol{H}_G$, $\boldsymbol{V}_G$, $\boldsymbol{W}_G$, $\boldsymbol{A}_G$ and $\boldsymbol{B}_G$ in Eq. (2).
- *Topical-based TAPER (**T-TAPER**)*: This model only integrates topical preferences: that is, $\boldsymbol{H}_T$, $\boldsymbol{V}_T$, $\boldsymbol{B}_T$ and $\boldsymbol{C}_T$ in Eq. (2).
- *Social-based TAPER (**S-TAPER**)*: This variant only considers social preferences: that is, $\boldsymbol{H}_S$, $\boldsymbol{V}_S$ and $\boldsymbol{A}_S$ in Eq. (2).

Finally, we consider the proposed TAPER approach:

- *Contextual Personalized Expert Recommendation (**TAPER**)*: This is the proposed framework, incorporating all three types of contextual information among users, experts, and topics in Eq. (2). Specifically, we let $\boldsymbol{H} = \boldsymbol{H}_G * \boldsymbol{H}_T * \boldsymbol{H}_S$, $\boldsymbol{V} = \boldsymbol{V}_G * \boldsymbol{V}_T * \boldsymbol{V}_S$, $\boldsymbol{W} = \boldsymbol{W}_G$, $\boldsymbol{A} = \boldsymbol{A}_G + \boldsymbol{A}_S$, $\boldsymbol{B} = \boldsymbol{B}_G + \boldsymbol{B}_T$ and $\boldsymbol{C} = \boldsymbol{C}_T$.

4.3 Results

We begin by investigating the quality of TAPER versus each of the baselines. We adopt 10-fold cross validation and report the average precision and recall over 10 test runs in Figure 6 and Figure 7. Overall, the proposed personal expert recommendation framework TAPER performs the best among all alternative baseline methods in both precision and recall. From Figures 6 and 7, we can observe that TAPER consistently outperforms the baseline methods MP, UCF, MF, and TF with an average improvement of 42.2% over the best of these four methods in precision and 33.5% in recall. Concretely, TAPER performs better than TF with an average improvement of 14.1% in precision and 17.8% in recall, indicating the superiority of a tensor factorization model integrating rich contextual preferences. Moreover, TAPER has a better performance than MF with an average improvement of 26.6% in precision and 25.8% in recall, which is significantly higher than a related matrix factorization approach in [17].

We observe that MP and UCF perform the worst among all approaches. We attribute the poor performance of the UCF approach to sparseness – the low density of data can lead to poor recommendations, whereas both MF and TF can leverage the low-rank approximation of user preferences towards experts. Overall, we

[4] We find that parameter settings of $\gamma < 1$ and $\beta < 0.1$ lead to fairly stable precision and recall, indicating the stability of TAPER to these regularization parameters.

Table 2: What Impact Does Contextual Preference Have on Each Approach? Here we compare contextual preferences of heterogeneous entities versus homogeneous entities.

Metric	Scenario	G-TAPER-Het	G-TAPER-Hom	T-TAPER-Het	T-TAPER-Hom	S-TAPER-Het	S-TAPER-Hom	TAPER-Het	TAPER-Hom
Precision	Top-5	0.271	0.282	0.262	0.274	0.289	0.278	0.296	0.306
	Top-10	0.234	0.242	0.227	0.232	0.248	0.241	0.254	0.260
	Top-15	0.204	0.208	0.192	0.199	0.214	0.201	0.219	0.229
Recall	Top-5	0.183	0.182	0.177	0.178	0.187	0.180	0.199	0.201
	Top-10	0.286	0.297	0.289	0.304	0.310	0.304	0.313	0.329
	Top-15	0.391	0.388	0.386	0.381	0.402	0.398	0.416	0.411

observe that the best method achieves a precision of around 0.3 and a recall of around 0.4 As Ye et al. [30] have observed, the effectiveness of recommenders with sparse datasets is usually low.

We also observe that adding additional contextual factors improves the basic tensor factorization (TF) approach. Concretely, G-TAPER gives an average improvement of 8.1% in precision and 7.3% in recall over TF. This indicates that the geo-spatial preferences among users, experts, and topics can help identify similar users and distinguish popular experts according to their spatial popularity. T-TAPER performs slightly better than TF with an average improvement of 5.2% in precision and 4.9% in recall, implying that the topical preferences among users and experts can help improve the performance of personalized expert recommendation via tensor factorization. Furthermore, S-TAPER gives an average improvement of 13.2% in precision and 10.6% in recall. This indicates that the social ties of users and experts can help find users with similar behaviors on selecting experts, which provide more significant contributions to the personal expert recommendation than the geo-spatial and topical. Recall that social ties implicitly capture latent geo-spatial and topical preferences.

The Impact of Contextual Preferences. We have seen that geo-spatial, topical, and social preferences can be integrated into tensor factorization for improved personalized expert recommendation. In this section, we aim to dig deeper into the impact of geo-spatial, topical, and social signals on the quality of personalized recommendation. What impact do heterogeneous and homogeneous contextual preferences have? Are these impacts equal across approaches? For this experiment, we add the suffixes *Het* and *Hom* to indicate which variant of the proposed framework is at study. For instance, G-TAPER-Hom represents the model G-TAPER by only leveraging the geo-spatial preferences between homogeneous entities including $\boldsymbol{H}_G$, $\boldsymbol{V}_G$, and $\boldsymbol{W}_G$. Similarly, we consider variations of T-TAPER, S-TAPER, as well as the full TAPER.

As it can be seen in Table 2, TAPER-Hom outperforms TAPER-Het with an average improvement of 3.4% in precision and 1.6% in recall, indicating that in general, the contextual preferences between homogeneous entities plays a more important role than between heterogeneous entities since such information can help identify similar users and further improve the quality of the recommendation. G-TAPER-Hom has a better performance than G-TAPER-Het with an average improvement of 3.0% in precision and 1.6% in recall. T-TAPER-Hom gives an average improvement of 3.5% in precision and 1.9% in recall. These results indicate that the geo-spatial and topical preferences between homogeneous entities contribute more to the proposed framework. However, it is surprising that S-TAPER-Het performs better than S-TAPER-Hom with an average improvement of 4.3% over S-TAPER-Hom in precision and 2.4% in recall. This implies that the following relationships between users and experts are a strong signal, and confirms that if a user is already following this expert, it is very likely that this user will include this expert on the list [17]. In addition, we also observe that the social preferences are more significant in contributing to the proposed framework than other factors.

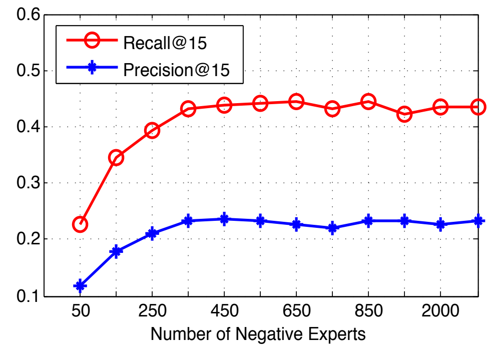
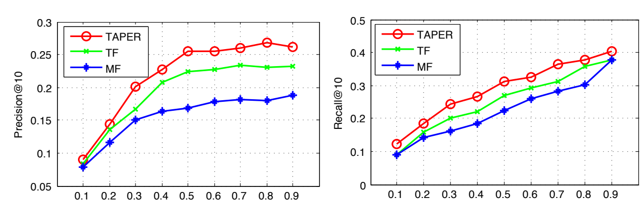


Figure 8: Effect of Number of Negative Experts.

The Impact of Negative Experts. We next turn to the impact of "negative experts", that is, to incorporate evidence of experts for whom a user is *not interested*. We seek to understand if these negative experts can be used as evidence in addition to the positive relationships investigated so far (e.g., by exploiting the geo-spatial preferences of users for experts). For example, knowing that a user is interested in a California Politics expert, but not interested in an expert on US National Politics may convey strong information about the preferences for that user on topics of regional (but not necessarily national) issues. To test the impact of these negative experts, we run the following experiment: first, we take the task of recommending the top-15 experts to users, and then we vary the number of negative experts. We vary the number of negative experts from 50 to 3,000 for each user by randomly selecting experts whom this user do not put in the Twitter list. Figure 8 demonstrates the impact of an increasing number of negative experts on the precision and recall of personalized expert recommendation. First, we observe that both precision and recall increase as the number of negative experts increases. This indicates that this signal of not being interested can provide some additional information beyond the positive relationships exploited so far. Second, we observe that the precision and recall curves flatten once the number of negative experts is larger than 350. Since there are nearly 9,000 experts in the dataset, the probability of false negative samples is small when only selecting a tiny part of them. However, this probability will increase as the number of negative experts grows, further affecting the quality of recommendations.



(a) Precison@10 (b) Recall@10

Figure 9: Varying the Amount of Training Data

Varying the Amount of Training Data. For the experiments so far, we have relied on cross-validation over a random split of experts into a training half and a testing half. Here, we explore the impact of varying the amount of training data on the task of personalized expert recommendation. Does the proposed approach still perform better than alternatives even with low amounts of training data? Do precision and recall plateau at some point? We vary the fraction of training data from 10% to 90% and evaluate TAPER versus a baseline tensor factorization method (TF) and a non-negative matrix factorization (MF) method on the task of personalized recommendation of the top-10 experts to each user. As we can see in Figure 9, the proposed framework TAPER consistently outperforms both MF and TF in precision and recall, across all fractions of training data. We also observe that the precision curves for all methods plateaus around 30%, indicating that good results may be achieved with even less training data. Naturally, the recall of all methods consistently increases as the training data increases, since recall is more sensitive to the number of testing samples.

5. CONCLUSION

We have studied the problem of personalized expert recommendation through a tensor-based exploration of geo-spatial, topical, and social context across users, experts and topics. Through a Twitter dataset, we have seen that the proposed framework can improve the quality of the recommendation by over 30% in both precision and recall compared to state-of-the-art baselines. In our continuing work we are interested to integrate additional contextual signals (e.g., are temporal factors important?) and to explore alternative settings (e.g., LinkedIn and Amazon). One promising direction is to evaluate the quality of downstream applications that can be built over these personalized expert models; for example, what impact does integrating these expertise signals into item-based recommenders and under what scenarios do they work well?

6. ACKNOWLEDGEMENTS

This work was supported in part by NSF grant IIS-1149383 and a Google Research Award. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

7. REFERENCES

[1] Supplementary material. http://students.cse.tamu.edu/hge/papers/recsys16_expert_supp.pdf.

[2] X. Amatriain, N. Lathia, J. M. Pujol, H. Kwak, and N. Oliver. The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. In *SIGIR*, 2009.

[3] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *WWW*, 2010.

[4] K. Balog, L. Azzopardi, and M. De Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR*, 2006.

[5] P. Bhargava, T. Phan, J. Zhou, and J. Lee. Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data. In *WWW*, 2015.

[6] P. Bhattacharya, S. Ghosh, J. Kulshrestha, M. Mondal, M. B. Zafar, N. Ganguly, and K. P. Gummadi. Deep twitter diving: Exploring topical groups in microblogs at scale. In *CSCW*, 2014.

[7] Z. Cheng, J. Caverlee, H. Barthwal, and V. Bachani. Who is the barbecue king of texas?: a geo-spatial approach to finding local experts on twitter. In *SIGIR*, 2014.

[8] D. M. Dunlavy, T. G. Kolda, and E. Acar. Temporal link prediction using matrix and tensor factorizations. *TKDD*, 2011.

[9] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: crowdsourcing search for topic experts in microblogs. In *SIGIR*, 2012.

[10] I. Guy, U. Avraham, D. Carmel, S. Ur, M. Jacovi, and I. Ronen. Mining expertise and interests from social media. In *WWW*, 2013.

[11] B. Hidasi and D. Tikk. Fast als-based tensor factorization for context-aware recommendation from implicit feedback. In *MLKDD*. 2012.

[12] D. Hu and J. L. Zhao. Expert recommendation via semantic social networks. *ICIS*, 2008.

[13] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *RecSys*, 2010.

[14] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM*, 2009.

[15] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 2009.

[16] Y. Liu, F. Shang, L. Jiao, J. Cheng, and H. Cheng. Trace norm regularized candecomp/parafac decomposition with missing data. *Cybernetics*, 2014.

[17] H. Lu and J. Caverlee. Exploiting geo-spatial preference for personalized expert recommendation. In *RecSys*, 2015.

[18] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM*, 2008.

[19] D. W. McDonald and M. S. Ackerman. Expertise recommender: a flexible recommendation system and architecture. In *CSCW*, 2000.

[20] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *ICDL*, 2004.

[21] A. Pal, A. Herdagdelen, S. Chatterji, S. Taank, and D. Chakrabarti. Discovery of topical authorities in instagram. In *WWW*, 2016.

[22] S. A. Paul, L. Hong, and E. H. Chi. Who is authoritative? understanding reputation mechanisms in quora. *arXiv preprint arXiv:1204.3724*, 2012.

[23] V. Rakesh, D. Singh, B. Vinzamuri, and C. K. Reddy. Personalized recommendation of twitter lists using content and network information. In *ICWSM*, 2014.

[24] S. Rendle and L. Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *WSDM*, 2010.

[25] P. Serdyukov, H. Rode, and D. Hiemstra. Modeling multi-step relevance propagation for expert finding. In *CIKM*, 2008.

[26] J. Sun, D. Tao, and C. Faloutsos. Beyond streams and graphs: dynamic tensor analysis. In *SIGKDD*, 2006.

[27] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, pages 234–240, 1970.

[28] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM*, 2010.

[29] B. Yang and S. Manandhar. Tag-based expert recommendation in community question answering. In *ASONAM*, 2014.

[30] M. Ye, X. Liu, and W. Lee. Exploring social influence for recommendation: a generative model approach. In *SIGIR*, 2012.

[31] H.-F. Yu, C.-J. Hsieh, I. Dhillon, et al. Scalable coordinate descent approaches to parallel matrix factorization for recommender systems. In *ICDM*, 2012.

[32] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *WWW*, 2007.

[33] D. Zhou, B. Wang, S. M. Rahimi, and X. Wang. A study of recommending locations on location-based social network by collaborative filtering. In *Adv. in Artificial Intelligence*. 2012.