Discovering What You're Known For:
A Contextual Poisson Factorization Approach

Haokai Lu
Texas A&M University
hlu@cse.tamu.edu

James Caverlee
Texas A&M University
caverlee@cse.tamu.edu

Wei Niu
Texas A&M University
Wei@cse.tamu.edu

ABSTRACT

Discovering what people are known for is valuable to many important applications such as recommender systems. Unlike an individual's personal interests, what a user is known for is reflected by the views of others, and is often not easily discerned for a long-tail of the vast majority of users. In this paper, we tackle the problem of discovering what users are known for through a probabilistic model called Bayesian Contextual Poisson Factorization. Moving beyond just modeling user's content, it naturally models and integrates additional contextual factors, concretely, user's geo-spatial footprints and social influence, to overcome noisy online activities and social relations. Through GPS-tagged social media datasets, we find that the proposed method can improve known-for prediction performance by 17.5% in precision and 20.9% in recall on average, and that it can capture the implicit relationships between a user's known-for profile and her content, geo-spatial and social influence.

1. INTRODUCTION

Discovering what people are *known for* is valuable to many important applications, including recommender systems and question-answering sites. For example, item-based recommenders can be augmented to customize recommendations based on what knowledgeable users prefer, rather than relying on all users [2]. While an individual's *personal interests* are often reflected in the media she consumes and generates, what a user is *known for* is reflected by the views of others and is often not easily discerned. A few high-profile people are easily recognized – for example, a researcher may be interested in basketball, biking, and recommender systems, though mainly known for recommender systems. But there is a long-tail of the vast majority of users for whom we have only limited insight into what they are known for.

And yet accurate identification of a user's known-for profile is challenging. First, the content that a user chooses to post on social media is often noisy and ambiguous. What users are truly known for can be buried among posts about daily routines and personal interests. Second, many users post only infrequently, meaning extreme sparsity for the vast majority of all users. Third, while a natural step to overcoming this sparsity is the integration of additional contextual factors (e.g., social links between users), it is not clear how different contextual factors correlate with users' known-for profiles and how we can model these contextual influence and integrate them together.

Hence, in this paper, we tackle the problem of *discovering what users are known for* through a probabilistic factorization model called Bayesian Contextual Poisson Factorization (BCPF). Three of the key features of the proposed BCPF model are: (i) It is jointly learned on a small fraction of users whose known-for profiles are already known and the vast majority of users for whom we have little (or no) information. (ii) Moving beyond just modeling the content a user generates, it naturally models and integrates additional contextual factors that provide implicit linkages between users for improved known-for profile estimation. Concretely, we investigate the impact of geo-spatial footprints and social influence as additional contextual signals. (iii) It inherits the strengths of Bayesian Poisson factorization (BPF), a variant of probabilistic matrix factorization recently proposed in [9, 10], which demonstrates scalable inference for sparse data and outperforms traditional matrix factorization [9].

In summary, this paper makes the following contributions:
– First, we define the problem of discovering user's known-for profile in social media, and propose a probabilistic method called Bayesian Contextual Poisson Factorization. This model can capture the implicit relationships between user's known-for profile and her content, geo-spatial and social influence. We then develop an efficient approximate variational inference to learn the latent parameters of BCPF.
– Second, we evaluate the proposed method over two Twitter datasets and against several alternative baselines. Overall, we see a significant improvement of 17.5% in precision and 20.9% in recall on average over the next-best method.
– Finally, we study the inferred geo-spatial and social influence latent factors, and observe that the geo-spatial factors are able to capture the underlying distributions of user's known-for profile at different locations. We also show that friends have different social influence on users with respect to different topics, and that our model is able to learn this fine-grained social influence.

Problem Formulation. We assume there exist a set of users U in a social network, and a set of tags T used to describe what users are known for with respect to different topics or aspects. The **known-for profile** p_u of a user

$u \in U$ is defined to be a subset of tags from T which can best describe what this user is *known-for* by others, instead of what u is personally *interested-in*. To give an example: suppose Alice is known by others as a chef and frequently posts many food recipes, but she also posts news and personal related stories now and then as her personal interest. We associate Alice with tags "chef" and "recipe" instead of "news" to best describe her perceived image by others, and treat "chef" and "recipe" as the *known-for* profile for Alice.

With the above definition, we can define the task of discovering user's known-for profile in social media as follows: **User Known-For Profile Discovery.** Given the users U^{old} whose known-for profiles are already known (labeled users), identify the known-for profiles for the rest of the users U^{new} (unlabeled users) based on U^{old}. Equivalently, the task is the same as identifying the top-n tags from T for unlabeled users as their known-for profiles.

However, discovering user's known-for profile is a classic cold-start problem. Since unlabeled users do not have any identified tags in their profile, collaborative filtering techniques fail to infer the profiles of unlabeled users since the corresponding rows of the user-profile matrix are all zeros for unlabeled users. Thus, to overcome the cold-start situation, we propose to leverage a user's contextual information to extract implicit relationships between their profile and the context for the task as described in the following.

2. KNOWN-FOR PROFILE DISCOVERY

As a first pass, we can attack the problem of known-for profile discovery with Bayesian Poisson factorization (BPF), a variant of probabilistic matrix factorization recently proposed in [9, 10] to model implicit feedback for recommendation. Due to the assumption of Poisson distribution instead of traditional Gaussian distribution in modeling the discrete data, BPF can capture the long-tailed distribution of user behavior, enjoy scalable inference for sparse data and outperform traditional matrix factorization [9]. In our setting, since a user's known-for profile is inherently of discrete nature, i.e., either a tag exists in the profile or not, it is a natural choice to adopt a Poisson distribution instead of a Gaussian distribution to model user's contextual influence.

Specifically, given the binary occurrence r_{ut} of tag t in the known-for profile for user u, BPF assumes that r_{ut} is generated according to:

$$r_{ut} \sim Poisson(\theta_u^T \theta_t) \qquad (1)$$

where $\theta_u \in \mathbb{R}^K$ is the latent factor of u representing what u is known for in a latent space, and $\theta_t \in \mathbb{R}^K$ is the latent attribute of t. Each component of these latent parameters is drawn according to a Gamma distribution as follows:

$$\theta_{\cdot k} \sim Gamma(\lambda_a, \lambda_b) \qquad (2)$$

where λ_a and λ_b are the shape and rate parameter of the Gamma distribution, respectively. Given a binary matrix of known-for profiles, BPF can find each user's latent factors over the tag's latent attributes. These inferred latent factors can be further used to identify other related tags which may fit the user's known-for profile.

2.1 Bayesian Contextual Poisson Factorization

While BPF is able to reveal the latent factors for users whose known-for profiles are already known or partially known, it will fail to infer the known-for profiles of unlabeled users,

since these users are "zero rows" in the user-tag matrix. Hence, a natural approach to extend BPF to these missing users is through the integration of additional contextual information that may provide implicit linkages between users.

We refer to this extended framework as **Bayesian Contextual Poisson Factorization** since it inherits the strengths of BPF and is extended to integrate valuable contextual information. Intuitively, BCPF is designed to learn the influence of each contextual feature individually and combine them together for overall representation. Under the assumption of Poisson factorization, since the sum of Poisson random variables is still a Poisson distribution, we can linearly add different contextual influence together without changing the underlying probabilistic assumption.

In the rest of this section, we present how several important contextual factors can be modeled under BCPF – user content, geo-spatial impact, and social influence – and how these factors can be integrated into BCPF. Note that the proposed model is generalizable and other contextual factors could be incorporated as well. Concretely, we model users in a heterogeneous graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, consisting of four types of nodes $\mathbb{V} = (U, C, L, P)$. Each user u has a sample of the textual content that she posts online, denoted as $c_u \in C$, and a pair of geographical coordinates $(l_u^{lat}, l_u^{lng}) \in L$ indicating the approximate location of this user's posting activities. The known-for profile of a labeled user from U^{old} is denoted as $p_u \in P$. All of the above relations are directly relevant to users themselves, and can be regarded as user attributes. Furthermore, users may also have social relations with each other. Together, user attributes and social relations constitute all edges $\mathbb{V}$ between different entities.

Content factor. The first baseline factor is a user's content – that is, we assume that a user's known-for profile can be reflected by the usage of words in posts. Let $\mathbf{R^c}$ be a sparse matrix describing the usage of words from all users, where each element r_{uw} is the count of word w adopted by user u. We then choose to model every count r_{uw} using a Poisson distribution, with the intensity parameter factorizing over u's latent factor θ_u and w's latent factor θ_w as follows:

$$r_{uw} \sim Poisson(\theta_u^T \theta_w) \qquad (3)$$

where θ_u and θ_w are both K-dimensional non-negative vectors with Gamma priors specified in Equation 2.

Factorizing over the word matrix $\mathbf{R^c}$ gives us an unsupervised version of users' latent preference over words. However, for labeled users U^{old}, θ_u should also reflect their known-for profiles. For instance, if both users are labeled with "sports", it is likely that they may share some common words in their sports-related posts. Thus, to capture this intuition, θ_u should also be constrained by users' known-for profiles. Let $\mathbf{R^p}$ be a binary matrix describing known-for profiles for all users, where each element r_{ut} represents if user u is labeled with tag t. Similarly, we model each binary r_{ut} with a Poisson distribution, with its intensity parameter factorizing over θ_u and t's latent factor θ_t as follows:

$$r_{ut} \sim Poisson(\theta_u^T \theta_t) \qquad (4)$$

where θ_t is a K dimensional non-negative vector parametrized with Gamma priors. By sharing user's latent factor for both factorization 3 and 4, the model can not only reflect user's latent preference over common use of words, but also ensure that users with similar profiles have similar θ_u. Note however that a user's posts may be intertwined with other
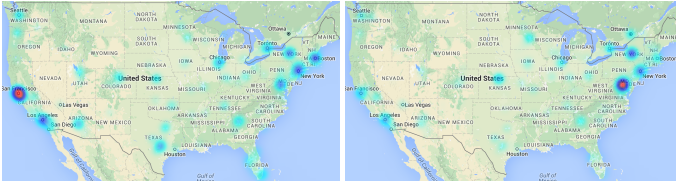
Figure 1: Geo-spatial distribution of users who are known for "entrepreneur" (left) and "politics" (right).

Figure 2: Proportion of followers who have the same tag in their known-for profile as the users. user1: red; user2: blue.

non-revealing texts, and thus can be noisy. To enhance the representation for unlabeled users, we also incorporate each user's geo-spatial footprints and social connections to further refine the model.

Geo-spatial factor. To demonstrate how geo-spatial location may reflect what a user is known for, we first show heat maps of users in US who are known for "entrepreneur" and "politics" in Figure 1 based on a sample of crowd-generated Twitter lists (described more fully in Section 3). As we can observe, for both topics, these well-known users are mostly distributed in a few areas, with a majority focused in San Francisco for "entrepreneur" and in Washington D.C. for "politics". This suggests that (i) for a specific tag, users do not distribute evenly across the entire geo-scope and may concentrate in a few areas; (ii) geo-spatial distributions of well-known users may vary by different topics.

Given the above observations, how can we use geo-spatial coordinates to enhance the representation of users? We notice from Figure 1 that users known for certain topical tags often appear in clusters, i.e., they concentrate in several discrete regions. In light of this, we introduce a region-dependent variable to represent a region's affinity over tags.

Concretely, we assume that the geographical space is partitioned into $\mathcal{I}$ regions. Each user can then be assigned into the region where she belongs, which gives a matrix $\mathbf{R^{ul}}$. Each region corresponds to a K-dimensional location latent factor, denoted θ_l, indicating the region's affinity over tags. If we consider that each r_{ut} in the matrix R^p is only dependent on user's location, then r_{ut} can be generated through the following Poisson distribution:

$$r_{ut} \sim Poisson(\theta_{l_u}^T \theta_t) \qquad (5)$$

where l_u represents the region where user u belongs, and the intensity parameter is determined by the inner product of the corresponding location and tag latent factor. Similarly, each component of θ_l has a conjugate Gamma prior as specified in Equation 2. Thus, through the explicit handling of tag's dependence over locations, it is expected that θ_l could indicate which tags are mostly dominant for the corresponding region.

Social influence factor. Homophily in social networks [22] suggests that people tend to connect with others who are similar to themselves. Indeed, there already exists some works [19, 21] exploiting social relations for item recommendations. Here, we would like to explore how social relations may benefit user known-for profile discovery. An immediate problem is that user's social relations are often noisy, i.e., if a friend of a user is known for a tag, it does not necessarily mean that the user is also known for the tag. In Figure 2, we randomly select two Twitter users who are known for eight tags, and examine the proportion of followers who have the same tag in their known-for profiles. We can see that even
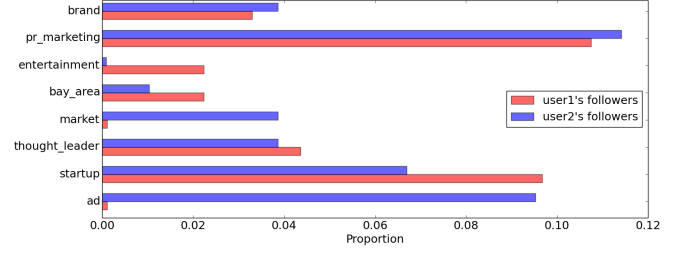
though user1 is known for "ad", very few followers of the user are also known for "ad". On the contrary, a significant portion of followers of user2 are known for "ad". This indicates that users may have different degree of social influence.

Thus, to capture this observation, we introduce a social influence parameter θ_f for each friend f of all users, indicating the degree of social influence of this friend on any user who follows her. Thus, if we have a total of M friends, the influence parameters constitute a vector of Θ^f with dimension M. Then, given user u's friends $\mathcal{N}_u$, if we consider that each r_{ut} in the matrix $\mathbf{R^p}$ is only dependent on user's friends' known-for profiles, then r_{ut} could be generated through the following Poisson distribution:

$$r_{ut} \sim Poisson(\sum_{f \in \mathcal{N}_u} \theta_f r_{ft}) \qquad (6)$$

where r_{ft} is the element in the matrix $\mathbf{R^{ft}}$ indicating if friend f has the tag t. The intensity parameter is obtained by aggregating all of the friends' latent influence for the user.

However, since each friend may be labeled with multiple tags, it is very likely that this friend has different levels of influence for different tags, as shown in Figure 2. In the figure, many people following user2 have the tag "pr marketing" and "ad". Thus, although user2 is also known for "entertainment", it is obvious that this user is more influential on "pr marketing" and "ad". As a result, if a new user starts to follow her, it is very likely that this user follows her because she is also known for "pr marketing" or "ad".

Hence, to capture this intuition, we assume that for each friend f of all users' friends, there is a latent factor θ_{ft} for each tag t. This parameter is used to capture the degree of influence that f has over t. Thus, if we have a total of N tags, the influence parameters constitute a M by N matrix Θ^{ft}. Thus, given user u's friends $\mathcal{N}_u$, similarly as above, r_{ut} can be generated through the following:

$$r_{ut} \sim Poisson(\sum_{f \in \mathcal{N}_u} \theta_{ft} r_{ft}) \qquad (7)$$

where the intensity parameter is obtained by aggregating all of the friends' latent influence for this user and the corresponding tag. Consequently, Equation 7 is able to model the above scenario where the new user is more likely to have the tag "pr marketing" or "ad" instead of "entertainment" if she follows user2. Similarly, each nonzero component of Θ^{ft} is put on Gamma priors for inference.

Integrating contextual factors. So far, we can generate independent known-for profiles through user's content, geographical coordinates and social relations. However, it is

Algorithm 1: Generative Process for BCPF
Input: hyper-parameter λ_a^w, λ_b^w, λ_a^u, λ_b^u, λ_a^l, λ_b^l, λ_a^t, λ_b^t, λ_a^{ft}, λ_b^{ft}

for each word w in the vocabulary, draw
   word topic factor $\theta_w \sim \text{Gamma}(\lambda_a^w, \lambda_b^w)$
for each user u, draw user content factor
   associated with tags, $\theta_u \sim \text{Gamma}(\lambda_a^u, \lambda_b^u)$
for each word w used by user u, draw
   the occurrence count $r_{uw} \sim \text{Poisson}(\theta_u^T \theta_w)$
for each discretized region l, draw region factor
   associated with tags, $\theta_l \sim \text{Gamma}(\lambda_a^l, \lambda_b^l)$
for each friend tag pair (f,t), draw
   friend social influence on t, $\theta_{ft} \sim \text{Gamma}(\lambda_a^{ft}, \lambda_b^{ft})$
for each tag t, draw tag factor associated with
   user content and region, $\theta_t \sim \text{Gamma}(\lambda_a^t, \lambda_b^t)$
for each tag t in user u's known-for profile, draw
   binary $r_{ut} \sim \text{Poisson}(\lambda_{ut})$, where λ_{ut} is determined
   by Equation 8.

natural to combine all types of contextual influence to form a better representation for users. Since the sum of Poisson distributions is still a Poisson distribution, we can combine the generative process specified by Equation 4, 5 and 7 together to obtain the overall intensity parameter of the sum Poisson distribution as:

$$\lambda_{ut} = \alpha\theta_u^T\theta_t + \beta\theta_{l_u}^T\theta_t + \gamma\sum_{f\in\mathcal{N}_u}\theta_{ft}r_{ft} \qquad (8)$$

where α, β and γ are the tradeoff weights for content, geospatial and social contributions, respectively. The entire generative process of BCPF is described in Algorithm 1.

2.2 Prediction

We have specified a Bayesian probabilistic model over the latent parameters Θ^u, Θ^w, Θ^l, Θ^{ft} and Θ^t, and the observed discrete data matrix $\mathbf{R^p}$, $\mathbf{R^c}$, $\mathbf{R^{ul}}$ and $\mathbf{R^{ft}}$. To predict the known-for profiles for new users, we need to estimate the posterior distributions of the latent parameters $p(\Theta^u, \Theta^w, \Theta^l, \Theta^{ft}, \Theta^t | \mathbf{R^p}, \mathbf{R^c}, \mathbf{R^{ul}}, \mathbf{R^{ft}})$ given the observed data. Once we have the posterior distributions of the latent factors, we can predict the known-for profiles for new users with the expectation of the sum of weighted Poisson distributions. This leads to computing the expectation of the intensity parameter λ_{ut}, which gives:

$$E(r_{ut}) = \alpha E(\theta_u)^T E(\theta_t) + \beta E(\theta_{l_u})^T E(\theta_t) + \gamma\sum_{f\in\mathcal{N}_u}E(\theta_{ft})r_{ft}$$

where all expectations are with respect to the posterior distributions. By making use of a new user's contextual information, we can predict the ranking of the tags for this user by their expected occurrence.

2.3 Learning with Variational Inference

Since it is not tractable to compute the exact posterior of the latent factors Θ, we propose to use variational methods [15] for approximate inference. Variational inference casts the approximation process as an optimization problem. By defining a freely parameterized family of distributions over latent variables, variational methods seek to fit its parameters so as to minimize the KL-divergence between the defined distribution and the posterior distribution. In our case, vari-

ational inference solves the following minimization problem:

$$q^*(\Theta) = argmin_q KL(q(\Theta)||p(\Theta|\mathbf{R^p}, \mathbf{R^c}, \mathbf{R^{ul}}, \mathbf{R^{ft}})) \quad (9)$$

where $q^*(\Theta)$ is the optimized variational distribution that is used as the proxy for the exact posterior. To facilitate the inference, we first introduce several auxiliary latent variables [9] for Equation 3 and 8. Specifically, let's assume $z_{uwk} \sim Poisson(\theta_{uk}\theta_{wk})$, where $\sum_k z_{uwk} = r_{uw}$, and $z_{utk}^c \sim Poisson(\alpha\theta_{uk}\theta_{wk})$, $z_{utk}^l \sim Poisson(\beta\theta_{l_uk}\theta_{tk})$, $z_{utf} \sim Poisson(\gamma\theta_{ft}r_{ft})$, where $\sum_k z_{utk}^c + \sum_k z_{utk}^l + \sum_{f\in\mathcal{N}_u} z_{utf} = r_{ut}$. Since the sum of a set of Poisson random variables is still a Poisson distribution, these auxiliary variables can still preserve the marginal distribution of r_{uw} and r_{ut} when marginalized out. Thus, our latent variables include latent parameters Θ and auxiliary latent variables Z_{uw}, Z_{ut}.

Before solving Equation 9, we need to derive the complete conditionals, i.e., the conditional distributions of a latent variable given all other variables, for each of Θ, Z_{uw} and Z_{ut}. Since each of Θ has conjugate Gamma priors, the complete conditionals for Θ are also Gamma distributions. For auxiliary latent variables, we follow the conclusion from [14] that given a vector Z_{ut} of Poisson distributed count, Z_{ut} is distributed as a multinomial conditioned upon the observed sum r_{ut}. Similarly, Z_{uw} is also a multinomial given r_{uw}. Thus, all complete conditionals can be derived.

Variational inference assumes $q(\Theta, Z_{uw}, Z_{ut})$ is in the same exponential family with the complete conditionals [10]. Thus, we employ the following mean field variational family, where each latent variable is independent with each other and governed by its own variational parameters:

$$q(\Theta^u, \Theta^w, \Theta^l, \Theta^{ft}, \Theta^t, Z_{uw}, Z_{ut}) = \prod_{u,k} q(\theta_{uk}|\eta_{uk}^a, \eta_{uk}^b)$$

$$\prod_{w,k} q(\theta_{wk}|\eta_{wk}^a, \eta_{wk}^b) \prod_{l,k} q(\theta_{lk}|\eta_{lk}^a, \eta_{lk}^b) \prod_{t,k} q(\theta_{tk}|\eta_{tk}^a, \eta_{tk}^b)$$

$$\prod_{f,t} q(\theta_{ft}|\eta_{ft}^a, \eta_{ft}^b) \prod_{u,w,k} q(z_{uwk}|\phi_{uwk}) \prod_{u,t,k} q(z_{utk}^c|\phi_{utk}^c)$$

$$\prod_{u,t,k} q(z_{utk}^l|\phi_{utk}^l) \prod_{u,t,f\in\mathcal{N}_u} q(z_{utk}|\phi_{utf})$$

where $\eta_.^a$ and $\eta_.^b$ are the shape and rate parameter of the variational Gamma distributions; and $\phi_{ut.}$ is the parameter of the variational Multinomial distributions. To obtain optimal values for $\eta_.^a$, $\eta_.^b$, $\phi_{uw.}$ and $\phi_{ut.}$, we need to solve the minimization in Equation 9. Since variational inference requires that the natural parameter of each $q(\cdot)$ is the expectation of the natural parameter of the corresponding complete conditional under $q(\cdot)$ [11], we just need to compute the expectation of the natural parameters of each complete conditional. To give an example, we show how to compute parameter η_{tk}^a and η_{tk}^b. The expectation of the natural parameters on the complete conditional of θ_{tk} gives:

$$\eta_{tk}^a = \lambda_t^a + \sum_u r_{ut}\phi_{utk}^c + \sum_u r_{ut}\phi_{utk}^l$$

$$\eta_{tk}^b = \lambda_t^b + \alpha\sum_u \frac{\eta_{uk}^a}{\eta_{uk}^b} + \beta\sum_u \frac{\eta_{l_uk}^a}{\eta_{l_uk}^b} \qquad (10)$$

where we have applied the fact that the expectation of a Gamma random variable is equal to the ratio of the shape parameter over the rate parameter. Similarly, we can derive the formula to compute other latent parameters. For auxiliary latent variable Z_{ut}, the expectation of its natural

Algorithm 2: Variational inference for BCPF

Input: hyper-parameter λ_a^w, λ_b^w, λ_a^u, λ_b^u, λ_a^l, λ_b^l, λ_a^t, λ_b^t, λ_a^{ft}, λ_b^{ft}, data matrix $\mathbf{R^p}$, $\mathbf{R^c}$, $\mathbf{R^{ul}}$, $\mathbf{R^{ft}}$

repeat

  Initialize all variational parameters $\eta_{\cdot}^a$, $\eta_{\cdot}^b$, $\phi_{ut\cdot}$, $\phi_{uw\cdot}$

  for each r_{uw}

    Update ϕ_{uwk} for each k with

      $\phi_{uwk} \sim \exp(\psi(\eta_{uk}^a) - \ln(\eta_{uk}^b) + \psi(\eta_{wk}^a) - \ln(\eta_{wk}^b))$

    Normalize $\phi_{uw\cdot}$

  for each r_{ut}

    Update ϕ_{utk}^c for each k, ϕ_{utk}^l for each k, and

      ϕ_{utf} for each $f \in \mathcal{N}_u$ with Equation 11

    Normalize $\phi_{ut\cdot}$

  for each w, update η_{wk}^a and η_{wk}^b for each k with

    $\eta_{wk}^a = \lambda_w^a + \sum_u r_{uw}\phi_{uwk}$, $\eta_{wk}^b = \lambda_w^b + \sum_u \frac{\eta_{uk}^a}{\eta_{uk}^b}$

  for each u, update η_{uk}^a and η_{uk}^b for each k with

    $\eta_{uk}^a = \lambda_u^a + \sum_w r_{uw}\phi_{uwk} + \sum_t r_{ut}\phi_{utk}^c$

    $\eta_{uk}^b = \lambda_u^b + \sum_w \frac{\eta_{wk}^a}{\eta_{wk}^b} + \sum_t \frac{\eta_{tk}^a}{\eta_{tk}^b}$

  for each l, update η_{lk}^a and η_{lk}^b for each k with

    $\eta_{lk}^a = \lambda_l^a + \sum_t \sum_u I(l_u = l)r_{ut}\phi_{utk}^l$

    $\eta_{lk}^b = \lambda_l^b + \sum_t \sum_u I(l_u = l)\frac{\eta_{tk}^a}{\eta_{tk}^b}$

  for each t, update η_{tk}^a and η_{tk}^b for each k with Equation 10

  for each f and each t, where $r_{ft} > 0$

    Update η_{ft}^a and η_{ft}^b with

      $\eta_{ft}^a = \lambda_{ft}^a + \sum_u I(f \in \mathcal{N}_u)r_{ut}\phi_{utf}$

      $\eta_{ft}^b = \lambda_{ft}^b + \sum_u I(f \in \mathcal{N}_u)r_{ft}$

until Converge

Table 1: Geo-tagged Twitter data.

Loc	# of users	# of tweets	# of following links
US	10,552	317,436	24,676
World	19,776	594,929	30,853

Loc	# of tags	# of records	sparsity
US	1,011	85,994	0.81%
World	1,456	136,625	0.47%

datasets (see Table 1), one in the US and one across the world. We also crawled about 30 recent tweets for each user and sampled her social relations.

Experimental Setup. For evaluation, we randomly partition all users into 60% for training, 30% for testing, and 10% for cross-validation. For the dimension K of the latent factors, we empirically select 100 for all methods. For the tradeoff weights α, β and γ in Equation 8 for BCPF, we set them to 1, 1.2 and 0.2 via cross-validation. To initialize the hyper-parameters of the Gamma priors in BCPF, we follow [10] and set them to 0.3 plus a small random variations on geo-spatial and social latent factors for sparse solutions; for user content and word variational parameters, we adopt LDA [5] and use document topic distribution and topic word distribution to initialize η_{uk} and η_{wk}, respectively. In the modeling of geo-spatial influence, it is assumed that the geographical space is partitioned into discrete regions. Here, we adopt k-means clustering instead of a simple gridding by clustering users' geo-coordinates. We choose a clustering-based approach because the geo-spatial distribution of users exhibits a clustering effect, as shown in Figure 1. For evaluation metrics, we adopt Precision@k (Prec@k) and Recall@k (Rec@k). Prec@k measures the percentage of the correctly identified tags over the top k predicted tags. Rec@k represents what percentage of true tags can emerge in top k predicted tags. Both measures are averaged over all testing users. In our experiments, we select k to be 5 and 10.

Baselines. We compare BCPF with the following baselines:
– *k-Nearest Neighbors.* In this baseline, we extract content features from user's sampled tweets, and apply kNN to find n most similar users to the testing user u. We then compute tag's score according to $s_{ut} = \sum_i^n w_{ui}r_{it}$, where w_{ui} is the similarity between user u and i, and select top k tags for prediction. We use two approaches to compute textual similarity, one with the bag-of-words model (kNN-BoW), another with the topic model (kNN-LDA) [5].
– *One-Vs-Rest multi-label ranking* [4]. We train a One-Vs-Rest multi-label classifier with bag-of-words on user's tweets. Logistic regression is used as the classification method to have a probabilistic output for ranking tags.
– *Wsabie* [29]. Wsabie is an embedding-based model which learns a mapping from a feature space to the joint space with the tags. Here, we train the model with WARP loss on bags-of-words of user's tweets, and use the learned user and tag embeddings for ranking tags.
– *Content-based Poisson Factorization (C-PF).* This is the method where we only keep the content factor in BCPF. It can be considered a form of CTPF proposed in [10] where we treat user's tweets as documents, and user's known-for profiles as binary ratings. We also ignore the topic offsets since the testing users are in complete cold-start situation.
– *Geo-spatial CPF (GC-PF).* This is the method in which

parameter of the complete conditional leads to

$$\phi_{utk}^c \sim \alpha \exp(\psi(\eta_{uk}^a) - \ln(\eta_{uk}^b) + \psi(\eta_{tk}^a) - \ln(\eta_{tk}^b))$$

$$\phi_{utk}^l \sim \beta \exp(\psi(\eta_{l_uk}^a) - \ln(\eta_{l_uk}^b) + \psi(\eta_{tk}^a) - \ln(\eta_{tk}^b)) \quad (11)$$

$$\phi_{utf} \sim \gamma r_{ft} \exp(\psi(\eta_{ft}^a) - \ln(\eta_{ft}^b))$$

where $\psi(\cdot)$ is the digamma function; and we have applied the fact that $E(\ln(X)) = \psi(\eta^a) - \ln(\eta^b)$ when $X \sim Gamma(\eta^a, \eta^b)$. Similarly, we could obtain the formula for Z_{uw}. We then use coordinate ascent to update each variational parameter by turns to obtain the locally optimal values. The overall update algorithm is shown in Algorithm 2.

The complexity of Algorithm 2 is largely determined by the product of K and the maximum number of non-zeros of the matrix $\mathbf{R^c}$, $\mathbf{R^p}$ and $\mathbf{R^{ft}}$. Thus, learning is quite efficient for a sparse data matrix. Here, we omit detailed complexity analysis due to the space limit.

3. EXPERIMENTAL EVALUATION

In this section, we conduct several experiments to evaluate the proposed BCPF for user known-for profile discovery.

Data. Our data is based on a sample of about 12 million Twitter lists collected from 2013 to 2014. Twitter lists [3, 8] are crowd-generated, for which a labeler can put a user on a tagged list, if the labeler thinks the user is known for the topic indicated by the tags. Thus, if a user is labeled by different labelers with certain tags, for example, "chef" and "recipe", then we consider this user is known for these topics. In our experiments, we use the threshold of three labelers to determine the existence of a tag in user's known-for profile. In addition, we also filter out infrequent tags with fewer than 20 users to focus on quality tags. We randomly sample two

we keep both content and geo-spatial factor in BCPF.

– *Social influenced CPF (SC-PF).* This method keeps both content and social influence factor in BCPF.

3.1 Effectiveness of BCPF

How well does the proposed BCPF perform compared to alternative baselines? In Table 2, we report Prec@k and Rec@k for all methods. Overall, BCPF gives the best results for all metrics and both datasets. Specifically, it gives an average improvement of 17.5% in precision and 20.9% in recall over the best content based method C-PF. This indicates the superiority of BCPF by exploiting geo-spatial and social influence factors other than user's content.

In content-based methods, neighborhood-based methods generally provide the worst performance of all since only local information can be used in prediction. Here, LDA based kNN performs better than BoW based kNN, since LDA can take advantage of the global topic information in users' posts. Supervised methods, however, give relatively better performance than neighborhood methods, as shown by One-Vs-Rest multi-label ranking and Wsabie. Both methods are trained with multiple labeled tags for each user on the BoW features, where Wsabie obtains embeddings for each user and tag. However, One-Vs-Rest clearly outperforms Wsabie, although Wsabie is more efficient and requires less computation time. C-PF, however, gives the best performance of all content-based methods, indicating the superiority of joint modeling by learning labeled users' profiles on content features and unlabeled users' texts.

Furthermore, we can see from the table that both GC-PF and SC-PF provides better performance against only content-based C-PF, respectively. Specifically, GC-PF gives an average improvement of 8.07% in precision and 11.1% in recall, respectively. SC-PF gives an average improvement of 10.9% in precision and 11.2% in recall, respectively. This indicates that geo-spatial features and parameterized social influence can both improve the identification of user's known-for profile. Given the overall improvement of the combination of these features, we can also conclude that these factors are able to complement each other.

3.2 Geo-spatial Factor Analysis

In the modeling of geo-spatial influence on user's known-for profile, we discretize the US/world area into geographical regions, and treat user's location as the region where she belongs. An important question here is how to select the number of regions $\mathcal{I}$. As we can imagine, if $\mathcal{I}$ is too small, the model may not be able to reflect the geo-spatial distribution of user's known-for profile; if it is too large, regions may be too small, thus leads to sparse observations. To that end, we select $\mathcal{I}$ for k-means clustering from 10 to 200, and run GC-PF with each value for ten random initializations of parameters. In Figure 3, we show how precision and recall changes with respect to the number of regions for both datasets. We can see that as $\mathcal{I}$ goes up, precision generally also increases. However, the performance plateaus when it is large. We attribute this to: first, when the number of regions increases, we obtain finer-grained geo-spatial characterization of tags; second, when it is large enough, a finer-grained discretization does not help distinguish tag's local distributions. It may even result in sparse counts in regions and also increases model complexity, thus degrading the final performance. Note that it requires less number of
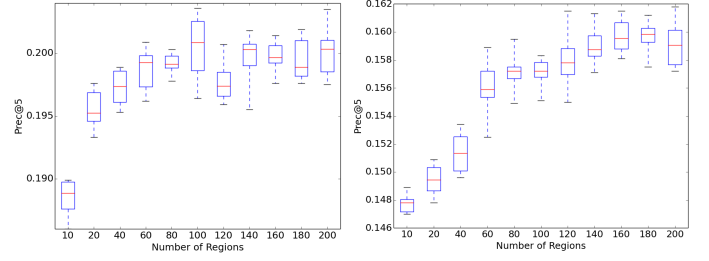

Figure 3: Box plot for precision@5 with respect to the number of regions. Left: US; Right: world.

Table 3: Top ranking tags for different areas obtained from latent location factors and tag factors.

Los Angeles	D.C.	SF	Chicago
la	dc	sf	chicago
celebrity	politics	geek	chi
star	progressive	bay area	illinois
tv	baltimore	dev	pr
artist	us	technology	advertising
actor	conservative	mobile	social media
famous people	government	startup	marketing
entertainment	politico	software	marketer

regions, i.e., larger partition of areas, for the US dataset than world to get to the best average performance. The reason is that the world data may be more diverse in terms of the geo-spatial distribution of user known-for profiles, thus requiring finer-grained geo-segments to reach the best performance.

To further examine the inferred region factors, we show in Table 3 the top ranking tags associated with selected areas. In particular, we compute the affinity score between a location and a tag by taking the expected inner product of the latent factor of the location and the tag, denoted as $E(\theta_l^T \theta_t)$. We can see from the table that, generally, the top ranking tags can reflect the common knowledge of the characteristics of those selected areas. For example, it is very likely that a user from the Washington D.C. area is well known for her political activities/comments. Thus, these inferred region factors are capable of nudging user's known-for profile toward her region characteristics.

3.3 Social Influence Factor Analysis

We have seen from the previous experiments that social influence plays an importance role in representing user's known-for profile by looking at her connections. In order to get a further understanding of how social influence varies for different friends and tags, we consider the following methods using social influence: (i) naive SC-PF (nSC-PF), where we replace θ_f with a single parameter θ_{naive} in Equation 6 for all friends and tags; (ii) user weighted SC-PF (uSC-PF), which models social influence by a vector of dimension M, with each element indicating the level of a friend's social influence for all tags (see Equation 6); and (iii) SC-PF, which models each friend's social influence for every tag of hers by matrix Θ_{ft}. We compare these methods with C-PF to see which method models social influence the best. Note that we also include the content factor as a basic representation for users since not all users have social connections.

258

Table 2: Comparison of performance with alternative methods. BCPF generally gives the best performance by integrating the contextual influence of textual, geo-spatial and social factors, and these factors are able to complement each other.

| Method | US | | | | World | | | |
| | Precision | | Recall | | Precision | | Recall | |
	@5	@10	@5	@10	@5	@10	@5	@10
kNN-BoW	0.114	0.098	0.120	0.181	0.083	0.072	0.099	0.151
kNN-LDA	0.150	0.127	0.155	0.226	0.120	0.100	0.137	0.200
One-Vs-Rest	0.166	0.140	0.176	0.250	0.136	0.112	0.161	0.227
Wsabie	0.159	0.137	0.169	0.244	0.116	0.101	0.134	0.201
C-PF	0.188	0.160	0.194	0.279	0.147	0.126	0.168	0.246
GC-PF	0.199	0.174	0.212	0.312	0.159	0.138	0.186	0.277
SC-PF	0.209	0.177	0.217	0.309	0.165	0.138	0.189	0.270
BCPF	**0.222**	**0.188**	**0.240**	**0.336**	**0.173**	**0.147**	**0.201**	**0.295**
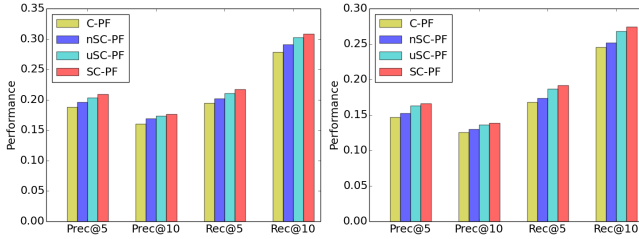
Figure 4: Comparative performance of the methods modeling social influence. SC-PF performs best by making it dependent on both friends and tags. Left: US; Right: world.

Table 4: User's inferred social influence θ_f by uSC-PF and social influence θ_{ft} for top ranking tags by SC-PF.

User	θ_f	top-1/θ_{ft}	top-2/θ_{ft}
@UnderTheBar	0.00290	entrepreneur 0.0230	health 0.0226
@skwigg	0.00425	fit 0.0217	nutrition 0.0190
@mollyblock	0.0143	marketer 0.104	pr 0.0226
@bikehugger	0.210	bicycle 0.435	cycling 0.303

As we can see from Figure 4, overall, social connection is able to improve the performance for predicting user's known-for profile. Specifically, we can make the following two conclusions: (i) it is generally better to differentiate social influence for different friends, rather than treating all friends with the same influence; (ii) it is generally better to differentiate social influence for different tags of each friend, rather than treating all tags of each friend with the same influence.

Table 4 shows the inferred social influence for different users and their top-ranking tags, which further exemplifies the observation that users may have different social influence on their followers with respect to different tags. In particular, user @UnderTheBar shows little influence on his followers' known-for profiles even if he may have tens of thousands of followers. This indicates that users who follow @UnderTheBar are not necessarily also known for "entrepreneur" or "health". On the contrary, @mollyblock has much larger social influence on her followers for her tag "marketer", in-

dicating that users following @mollyblock are more likely to also be known for "marketer", even if @mollyblock has fewer followers than @UnderTheBar. One explanation for this observation is that people connect with others for different reasons on social media, and that popularity itself is not necessarily an indicator of one's social influence.

4. RELATED WORK

User profiling is an important task, with many efforts building user's topic interests for personalized search [7, 24], targeted advertising [1] and social media systems [6, 27, 34]. For example, Dou et al. [7] built user interest profile by pre-defined topic categories for personalized webpage re-ranking. Ahmed et al. [1] proposed a statistical framework to extract user's dynamic interest profile for behavioral targeting. Chen et al. [6] built topic profiles of users with bag-of-words model to recommend conversations in Twitter. Zhao et al. [34] improved user's topic interest profile by joint factorization of user's topic matrix and behaviors. User profiles are also used to provide recommendations for online activities such as commenting on news stories [25] and mentions in micro-blogging systems [27]. Besides user's topic interest profile, other works [16, 17, 23] focus on inferring user's attribute and demographics profile such as gender and education. For example, Mislove et al. [23] proposed community detection based approaches to infer user's college and major. Li et al. [16] presented a weakly-supervised approach to extract user's job and education. Our work is different from the above works in that we tackle the problem of discovering what users are known for instead of user's topic interests.

Contextual information in social media has been widely used in previous works [10, 13, 19, 21, 26, 28, 31, 33] to learn better user profiles and improve task performance. A seminal work by Singh and Gordon [26] proposed collective matrix factorization to simultaneously factor several matrices encoding contextual information for better prediction of user-movie ratings. Wang and Blei [28] proposed collaborative topic regression to learn latent user preference by modeling both ratings and content. Similarly, Gopalan et al. [10] also modeled both user's ratings and content but with Poisson factorization. Jamal et al. [13] proposed a context dependent factor model to learn general latent factors of entities in social networks for better recommendation. Temporal information has also been used in [31] to learn both user-oriented topics and time-oriented topics. Other contextual information used to learn user preference includes domain-specific communities [33] and social relations [19,

21] which are used to regularize latent factors between socially connected users. In our work, instead of using social regularization for latent factors, we learn both user and tag dependent social influence from the data with Poisson distribution for fine-grained modeling of social relations.

Geographical footprints have also been widely explored in many location-based applications [12, 18, 20, 30, 32, 35]. One of the most popular applications is POI recommendation, where geographical influence is combined with user preference for better performance [18, 30]. Other works have used geographical influence for rating prediction in Yelp [12], activity recommendation with GPS history [35], expert recommendation [20] and event-based group recommendation [32]. In contrast, we are focused on geo-spatial influence for user known-for profile discovery in social media.

5. CONCLUSION

In this paper, we tackled the problem of discovering what users are known for in social media. By integrating user's textual, geo-spatial and social influence, we proposed Bayesian Contextual Poisson Factorization to overcome the noisiness of user's posting activities and social relations. Experimental results showed that BCPF can improve known-for prediction by 17.5% in precision and 20.9% in recall on average. We also showed that user's connections have varying social influence for different topics, confirming our fine-grained modeling of social influence by making it dependent on both users and topics. In our future work, we are interested in exploring the impact of additional contextual signals beyond the textual, geo-spatial and social influence studied here.

6. REFERENCES

[1] A. Ahmed, Y. Low, M. Aly, V. Josifovski, and A. J. Smola. Scalable distributed inference of dynamic user interests for behavioral targeting. In *SIGKDD*, 2011.

[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] 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.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., 2006.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.

[6] J. Chen, R. Nairn, and E. Chi. Speak little and well: recommending conversations in online social streams. In *SIGCHI*, 2011.

[7] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *WWW*, 2007.

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

[9] P. Gopalan, J. M. Hofman, and D. M. Blei. Scalable recommendation with poisson factorization. *arXiv preprint arXiv:1311.1704*, 2013.

[10] P. K. Gopalan, L. Charlin, and D. Blei. Content-based recommendations with poisson factorization. In *NIPS*, 2014.

[11] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *JMLR*, 2013.

[12] L. Hu, A. Sun, and Y. Liu. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *SIGIR*, 2014.

[13] M. Jamali and L. Lakshmanan. Heteromf: recommendation in heterogeneous information networks using context dependent factor models. In *WWW*, 2013.

[14] N. Johnson, A. Kemp, and S. Kotz. *Univariate Discrete Distributions*. John Wiley and Sons, 2005.

[15] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine learning*, 1999.

[16] J. Li, A. Ritter, and E. H. Hovy. Weakly supervised user profile extraction from twitter. In *ACL*, 2014.

[17] R. Li, C. Wang, and K. C.-C. Chang. User profiling in an ego network: co-profiling attributes and relationships. In *WWW*, 2014.

[18] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *SIGKDD*, 2013.

[19] X. Liu and K. Aberer. Soco: a social network aided context-aware recommender system. In *WWW*, 2013.

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

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

[22] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[23] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: inferring user profiles in online social networks. In *WSDM*, 2010.

[24] F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In *WWW*, 2006.

[25] E. Shmueli, A. Kagian, Y. Koren, and R. Lempel. Care to comment?: recommendations for commenting on news stories. In *WWW*, 2012.

[26] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. In *SIGKDD*, 2008.

[27] B. Wang, C. Wang, J. Bu, C. Chen, W. V. Zhang, D. Cai, and X. He. Whom to mention: expand the diffusion of tweets by@ recommendation on micro-blogging systems. In *WWW*, 2013.

[28] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *SIGKDD*, 2011.

[29] J. Weston, S. Bengio, and N. Usunier. Wsabie: Scaling up to large vocabulary image annotation. In *IJCAI*, 2011.

[30] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, 2011.

[31] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang. A temporal context-aware model for user behavior modeling in social media systems. In *SIGMOD*, 2014.

[32] W. Zhang, J. Wang, and W. Feng. Combining latent factor model with location features for event-based group recommendation. In *SIGKDD*, 2013.

[33] X. Zhang, J. Cheng, T. Yuan, B. Niu, and H. Lu. Toprec: domain-specific recommendation through community topic mining in social network. In *WWW*, 2013.

[34] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi. Improving user topic interest profiles by behavior factorization. In *WWW*, 2015.

[35] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *WWW*, 2010.