

# User Preference Learning with Multiple Information Fusion for Restaurant Recommendation

Yanjie Fu <sup>\*</sup>      Bin Liu<sup>†</sup>      Yong Ge <sup>†</sup>      Zijun Yao<sup>◇</sup>      Hui Xiong<sup>†‡</sup>

## Abstract

If properly analyzed, the multi-aspect rating data could be a source of rich intelligence for providing personalized restaurant recommendations. Indeed, while recommender systems have been studied for various applications and many recommendation techniques have been developed for general or specific recommendation tasks, there are few studies for restaurant recommendation by addressing the unique challenges of the multi-aspect restaurant reviews. As we know, traditional collaborative filtering methods are typically developed for single aspect ratings. However, multi-aspect ratings are often collected from the restaurant customers. These ratings can reflect multiple aspects of the service quality of the restaurant. Also, geographic factors play an important role in restaurant recommendation. To this end, in this paper, we develop a generative probabilistic model to exploit the multi-aspect ratings of restaurants for restaurant recommendation. Also, the geographic proximity is integrated into the probabilistic model to capture the geographic influence. Moreover, the profile information, which contains customer/restaurant-independent features and the shared features, is also integrated into the model. Finally, we conduct a comprehensive experimental study on a real-world data set. The experimental results clearly demonstrate the benefit of exploiting multi-aspect ratings and the improvement of the developed generative probabilistic model.

**Keywords:** Restaurant Recommendation, Co-Matrix Factorization, Multi-information Fusion

## 1 Introduction

A large number of multi-aspect restaurant reviews are available on the Internet. These reviews have been frequently cited as the single most important factor that determines the customer’s choice of restaurants.

Thus, it is possible to exploit these multi-aspect reviews for personalized restaurant recommendations. However, it is a non-trivial task to develop a restaurant recommender system due to the unique characteristics of multi-aspect restaurant reviews.



Figure 1: The multi-response of restaurant rating.

Indeed, unlike traditional ratings of items, such as movies, the multi-aspect ratings of restaurants are often collected from the users at online review platforms. For example, the customers are allowed to give different ratings to a restaurant as shown in Figure 1, where the Shushi Ota restaurant received different ratings on Cuisine, Environment, and Overall from a user. The environment or cuisine rating represents the customer’s perception for a specific aspect of the service quality, while the overall rating indicates how much the customer is satisfied with the overall service provided by the restaurant. Thus, a restaurant may receive different ratings on various aspects from consumers. Different users may have different ratings on a specific aspect (e.g., overall, cuisine or environment) for a restaurant. As a result, personalized restaurant recommendations are expected to recommend restaurants to a customer which could reflect individual preferences in a collective way. In other words, the recommendation models should be able to learn the consumer’s preferences of all aspects by using multi-aspect ratings in order to produce effective personalized restaurant recommendations.

Also, different from movie or music recommendations, geographic factors also play an important role when people select a restaurant. Indeed, restaurants tend to co-locate at different regions in a city, rather than uniformly distributed. Thus, there is some intrinsic spatial property embedded in the restaurant ratings. In fact, the probability that a user chooses/rates a restaurant is inversely proportional to the geographic distance between the user and restaurant. As a result,

<sup>\*†◇+</sup>Management Science and Information Systems Department, Rutgers University, NJ, USA. Email: {yanjie.fu, binben.liu, zijun.yao, hxiong}@rutgers.edu.

<sup>†</sup>Computer Science Department, University of North Carolina at Charlotte, NC, USA. Email: yong.ge@uncc.edu.

<sup>‡</sup>Contact Author.

the restaurant recommendation should take these geographic factors into consideration.

There are some existing work on restaurant recommendations [4, 8, 21]. However, these methods are mostly based on user’s input and are only exploratory in nature. For instance, these methods have not considered the multi-aspect ratings and modeled them together. Also, geographic interrelationships between users and restaurants have not been effectively integrated into the recommendation method.

To this end, in this paper, we propose a systematic way to study restaurant recommendations by making a collective use of multi-aspect ratings, geographic information, and profile information of users and restaurants. Specifically, we model the multi-aspect ratings with a collaborative latent factor analysis framework. We represent user  $i$  with a vector of latent features  $u_i$ . Each item is represented with different vectors of latent features with respect to different aspects of rating. For instance, a restaurant is represented with a vector of latent features  $e_j$  with respect to environment ratings. Then, an environment rating  $g_{ij}$  is modeled as a function of  $u_i$  and  $e_j$ . Thus, the common user’s latent features connect the modeling of multi-aspect ratings together; that is, the user’s latent feature with respect to one aspect of rating is the same as that with respect to another aspect of rating. With this kind of shared user’s latent feature, the interrelationship among each aspect of rating is implicitly captured through the whole generative process. Furthermore, we consider the geographical proximity between a user and a restaurant, and integrate it into the model of overall ratings. In addition, due to such geographic factor, people in a spatial zone tend to have food in a group of restaurants, and barely visit and rate other restaurants. This phenomena creates the sparsity to the rating matrix. Thus, to overcome this sparsity challenge, we use the user/restaurant or rating-related profile information as shown in Table 2 and further tune the modeling of overall ratings. Finally, we conduct extensive experiments on a real-world data set. The experimental results show that our methods significantly outperform state-of-the-art collaborative filtering models in terms of various metrics, such as NDCG, MAE and RMSE.

## 2 Restaurant Recommendation

In this section, we introduce the collaborative latent factor analysis framework which incorporates multi-aspect ratings, profile information and geographic proximity together for restaurant recommendations.

### 2.1 Problem Definition

The problem of restaurant recommendations is to

Symbol	Size	Description
$\mathbf{Y}$	$I \times J$	overall ratings
$\mathbf{H}$	$I \times J$	cuisine ratings
$\mathbf{G}$	$I \times J$	environment ratings
$\mathbf{U}$	$I \times K$	user latent factor
$\mathbf{V}$	$J \times K$	restaurant latent factor
$\mathbf{E}$	$J \times K$	environment latent factor
$\mathbf{C}$	$J \times K$	cuisine latent factor
$\mathbf{x}$	$1 \times P$	user and restaurant shared features
$\mathbf{f}_i$	$1 \times M$	user-independent features
$\mathbf{f}_j$	$1 \times N$	restaurant-independent features
$R$	$1 \times \mathbb{R}$	region set, $r$ is a region in $R$
$\mu$	$\mathbb{R}^2$	mean of a region
$\Sigma$	$\mathbb{R}^{2 \times 2}$	covariance matrix of a region

Table 1: Mathematical Notations

recommend restaurants to a user based on multi-aspect ratings and other side information (e.g., geographic and profile information) collected from a web site. Let  $U = \{u_1, u_2, \dots, u_i, \dots, u_I\}$  be a set of users, where each user has a location  $l_i$  and observable user-dependent features  $\mathbf{f}_i$  (e.g. religion, car ownership, etc.) which describe the user profile. Let  $V = \{v_1, v_2, \dots, v_j, \dots, v_J\}$  be a set of restaurants, where each restaurant has a location  $l_j = \{lon_j, lat_j\}$  in terms of longitude and latitude, and observable restaurant-dependent features  $\mathbf{f}_j$  (e.g. accessibility, franchise, etc.) which describe the restaurant characteristics. Different from traditional recommendation problems, we have multi-aspect ratings for each restaurant. For instance, there is a rating triplet:  $\langle \text{overall rating, cuisine rating, environment rating} \rangle$  given by user  $i$  for restaurant  $j$ . We denote the rating triplet as  $\langle y_{ij}, h_{ij}, g_{ij} \rangle$  for the pair of user  $i$  and restaurant  $j$ . Also, we refer  $i$  as user and  $j$  as restaurant in the following unless otherwise specified. In addition, there are shared features by users and restaurants for each rating. We denote such shared features of user  $i$  and restaurant  $j$  as  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . To summarize, we list these notations in Table 1. As can be seen, in addition to the user-restaurant ratings, we also have rich side information.

User-dependent Features: $\mathbf{f}_i$ (7 attributes) transportation, marital status, children, interest, personality, religion, activity
Restaurant-dependent Features: $\mathbf{f}_j$ (5 attributes) longitude, latitude, accessibility, franchise, other service
Share features: $\mathbf{x}$ (8 attributes) dress style, drink, price, ambience, smoke, parking, payment, cuisine

Table 2: Profiles of Restaurant Rating Events

### 2.2 Model Specification

Here, we first provide an overview of our proposed collaborative latent factor model for restaurant recommendations. Then we explain how to model the multi-aspect ratings with Bayesian Co-Nonnegative Matrix Factorization (BCoNMF), and how to compute the profile similarity and the geographic proximity. Finally, we integrate all them together into a uniform generative model via the proposed Geographic Bayesian Co-Nonnegative

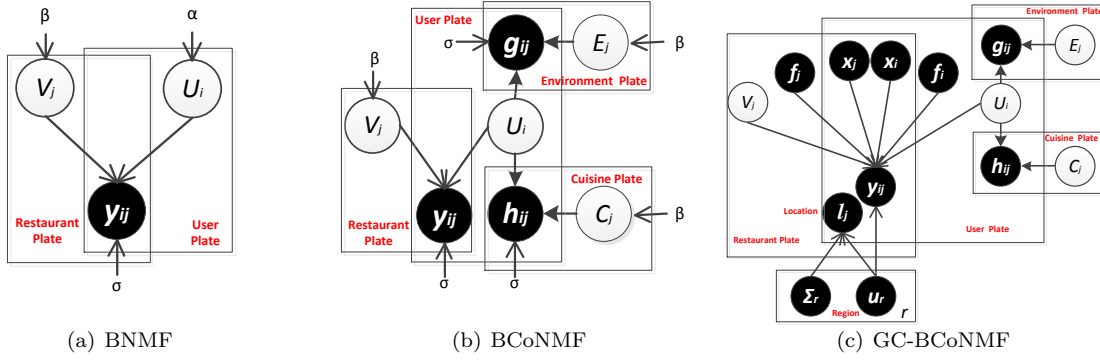


Figure 2: The graphical representation of the proposed GC-BCoNMF Model (c), BNMF model (a) and BCoNMF model. (the priors are not shown for the simplicity of description).

Matrix Factorization (GC-BCoNMF) model

### 2.2.1 Overview

By the utility theory [15], the rating of a restaurant can be regarded as the explicit response of the utility obtained from consumption in the restaurant, which is a measure of user satisfaction. A user’s decision process of choosing a restaurant can be viewed as a process that aims to maximize the utility of the choice. We argue that the utility a user  $i$  could gain from a restaurant  $j$  is heterogeneous and proportional to (1) the user’s personal preference hidden in the multi-aspect ratings,  $\gamma_{ij}$ ; (2) the profile similarity,  $\delta_{ij}$ ; (3) the geographic proximity,  $\rho_{ij}$  as

$$(2.1) \quad y_{ij} \propto \gamma_{ij} \cdot \delta_{ij} \cdot \rho_{ij}.$$

Figure 2(c) shows the decision process of user  $i$  to choose restaurant  $j$  in a generative way. First, we draw the user latent features  $\mathbf{u}_i$ . Then, the cuisine rating  $h_{ij}$  is generated by the interaction between user latent features  $\mathbf{u}_i$  and cuisine latent features  $\mathbf{c}_j$ , and the environment rating  $g_{ij}$  is generated by the interaction between user latent features  $\mathbf{u}_i$  and environment latent features  $\mathbf{e}_j$ . Furthermore, we model the user preference  $\gamma_{ij}$  as a linear combination of bias effect  $t$  and latent preference  $\mathbf{u}_i^\top \mathbf{v}_j$ . After incorporating the profile similarity and geographic proximity, the probabilistic generative process for user  $i$  to choose restaurant  $j$  can be expressed as Table 3.  $\mathbf{u}$ ,  $\mathbf{v}$ ,  $\mathbf{e}$  and  $\mathbf{c}$  can be drawn from some proper probabilistic distributions.

### 2.2.2 Modeling User Preference Via CoMF

Matrix Factorization based methods have been widely used to model the user-item relationship for recommendations. Bayesian Nonnegative Matrix Factorization (BNMF) is one of the matrix factorization based methods that can be applied for single aspect rating setting. Considering the simplest case: [user, restaurant, overall rating], where the overall rating of user  $i$  for restaurant  $j$  is  $y_{ij}$ , the overall rating  $y_{ij}$  can be factorized with  $U_i$  and

1. Draw user latent factor  $\mathbf{u}_i \sim P(\mathbf{u}_i; \Psi_{\mathbf{u}_i})$ .
2. Generate restaurant cuisine rating
  - a. Draw cuisine latent factor  $\mathbf{c}_j \sim P(\mathbf{c}_j; \Psi_{\mathbf{c}_j})$ .
  - b. Draw cuisine rating  $h_{ij} \sim P(\mathbf{u}_i^\top \mathbf{c}_j)$ .
3. Generate restaurant environment rating
  - a. Draw environment latent factors  $\mathbf{e}_j \sim P(\mathbf{e}_j; \Psi_{\mathbf{e}_j})$ .
  - b. Draw environment rating  $g_{ij} \sim P(\mathbf{u}_i^\top \mathbf{e}_j)$ .
4. Generate user preference
  - a. Generate bias effect  $t$
  - b. Draw restaurant latent factor  $\mathbf{v}_j \sim P(\mathbf{v}_j; \Psi_{\mathbf{v}_j})$ .
  - c. User-item preference  $\gamma_{ij} = t + \mathbf{u}_i^\top \mathbf{v}_j$ .
5. Compute profile similarity  $\delta_{ij}$ .
6. Compute geographic proximity  $\rho_{ij}$ .
7.  $y_{ij} \sim P(\mu_{ij})$  where  $\mu_{ij} = (t + \mathbf{u}_i^\top \mathbf{v}_j) \cdot \delta_{ij} \cdot \rho_{ij}$

Table 3: The generative process

$V_j$ , which represent the user latent features and item latent features respectively.  $U_i$  and  $V_j$  encode the affinity of user  $i$  and restaurant  $j$  in the latent space. Due to the non-negative rating, we further constrain user factors  $\mathbf{U}$  and restaurant factors  $\mathbf{V}$  to be non-negative. In BNMF, which is shown in Figure 2(a), we place exponential distribution on  $u_{ik}$  and  $v_{jk}$  as the empirical prior, and an inverse gamma distribution on  $\sigma^2$ . Then we assume rating  $y_{ij}$  is drawn from rectified normal distributions with mean  $\mathbf{u}_i^\top \mathbf{v}_j$  and variance  $\sigma^2$ . The generative process of  $y_{ij}$  follows: [1.] Draw  $u_{ik} \sim \alpha \cdot \exp(-\alpha \cdot u_{ik})$ ; [2.] Draw  $v_{jk} \sim \beta \cdot \exp(-\beta \cdot v_{jk})$  [3.] Generate  $y_{ij} \sim \mathcal{N}^R(\mathbf{u}_i^\top \mathbf{v}_j, \sigma^2)$ .

However, in our restaurant recommendation task, we may have multiple ratings given by individual user for each restaurant. Let us consider the rating triplet case: [user, restaurant, <overall rating, cuisine rating, environment rating>]. As the interaction between user plate and restaurant plate generates the overall rating  $y_{ij}$ , we assume that the interaction between user plate and cuisine plate generates the cuisine rating  $h_{ij}$  and the interaction between user plate and environment plate generates the environment rating  $g_{ij}$  as shown in Figure 2(b). Hence, we propose BCoNMF which has common user latent features as Table 4. In this way, BCoNMF co-factorizes the rating triplet via maximizing the likelihood of the observed multi-aspect ratings.

1. Generate latent factors
  - a. Draw  $u_{ik} \sim \alpha \cdot \exp(-\alpha \cdot u_{ik})$
  - b. Draw  $v_{jk} \sim \beta \cdot \exp(-\beta \cdot v_{jk})$
  - c. Draw  $c_{jk} \sim \beta \cdot \exp(-\beta \cdot c_{jk})$
  - d. Draw  $e_{jk} \sim \beta \cdot \exp(-\beta \cdot e_{jk})$
2. Generate rating triplet
  - a. Generate  $y_{ij} \sim \mathcal{N}(\mathbf{u}_i^\top \mathbf{v}_j, \sigma^2)$
  - b. Generate  $h_{ij} \sim \mathcal{N}(\mathbf{u}_i^\top \mathbf{c}_j, \sigma^2)$
  - c. Generate  $g_{ij} \sim \mathcal{N}(\mathbf{u}_i^\top \mathbf{e}_j, \sigma^2)$

Table 4: The generative process of BCoNMF

In addition, the user preference of a given restaurant is influenced by user personality and restaurant popularity. For example, a snack restaurant may have high popularity due to the local food culture in a specific region. Travellers may tend to give high ratings, even they actually do not favorite it, which is also called biased effect [5]. We add the bias effect  $t = \lambda \mathbf{f}$  to formulate user preference as  $\gamma_{ij} = \lambda \mathbf{f} + \mathbf{u}_i^\top \mathbf{v}_j$  where  $\mathbf{f} \in \mathbb{R}^{M+N}$  is a vectorized concatenation of user dependent features  $\mathbf{f}_i \in \mathbb{R}^M$  and restaurant dependent features  $\mathbf{f}_j \in \mathbb{R}^N$ ,  $t$  is regressed using both user dependent features and restaurant dependent features with  $\lambda$  as coefficient [3].

### 2.2.3 Modeling Profile Similarity

For the profiles of rating events in Table 2, shared features, which are actually the attributes attached to both user and restaurant when the rating event happens, can be used to understand user-restaurant similarity and improve the recommendation performances. Assuming that  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are user  $i$  and restaurant  $j$  observable shared features respectively, we model profile similarity  $\delta_{ij}$  as a matching score between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . We apply the symmetric Jensen-Shannon divergence to compute the context distance between user  $i$  and restaurant  $j$

$$(2.2) \quad D_{js}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2}D(\mathbf{x}_i||M) + \frac{1}{2}D(\mathbf{x}_j||M)$$

where  $M = \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j)$  and  $D(\cdot||\cdot)$  is the Kullback-Leibler distance. Then we define the matching score as  $\delta_{ij} = 1 - D_{js}(\mathbf{x}_i, \mathbf{x}_j)$ .

### 2.2.4 Modeling Geographic Proximity

We argue that the nearer the restaurant is from the user's location, the more likely the user will visit this restaurant [22, 7]. Due to dynamic mobility of human, a user may travel to different cities and rate restaurants there. In other words, the real-time location of user is unknown. But, unlike other POIs such as scenic spots, gas stations, restaurants have unique characteristics: (1) restaurants have strong cluster effect, i.e., they tend to locate together geographically. Since we assume that each restaurant is associated with a region  $r$ , the location of restaurant can be viewed as drawn from multivariate Gaussian distribution with region center as mean  $\mu_r$  and covariance  $\Sigma_r$ :  $l_j \sim \mathcal{N}(\mu_r, \Sigma_r)$ ; (2) restaurants interact with people through daily dining. Therefore, restaurants usually locate in a region with high human-

<b>Overall Rating</b>	$y_{ij} \sim \mathcal{N}(y_{ij} \mu_{ij}^y, \sigma^2)$
<b>Environment Rating</b>	$g_{ij} \sim \mathcal{N}(g_{ij} \mu_{ij}^g, \sigma^2)$
<b>Cuisine Rating</b>	$h_{ij} \sim \mathcal{N}(h_{ij} \mu_{ij}^h, \sigma^2)$
<b>Overall Utility</b>	$\mu_{ij}^y = (t + \mathbf{u}_i \cdot \mathbf{v}_i) \cdot \delta_{ij} \cdot \rho_{ij}$
<b>Environment Utility</b>	$\mu_{ij}^g = \mathbf{u}_i^\top \mathbf{e}_j$
<b>Cuisine Utility</b>	$\mu_{ij}^h = \mathbf{u}_i^\top \mathbf{c}_j$
<b>Profile similarity</b>	$\delta_{ij} = 1 - D_{js}(\mathbf{x}_i, \mathbf{x}_j)$
<b>Geographic proximity</b>	$\rho_{ij} = (\frac{d_0}{d_0 + d(i,j)})^e$
<b>User latent factors</b>	$u_{ik} \sim \text{Exp}(\alpha_k)$
<b>Restaurant latent factors</b>	$v_{jk} \sim \text{Exp}(\beta_k)$
<b>Environment latent factors</b>	$e_{jk} \sim \text{Exp}(\beta_k)$
<b>Cuisine latent factors</b>	$c_{jk} \sim \text{Exp}(\beta_k)$
<b>Bias effect</b>	$t = \lambda \mathbf{f}$
<b>Variance</b>	$\sigma^2 \sim \text{Inv} - \text{Gamma}(\sigma^2 a, b)$

Table 5: GC-BCoNMF: a hierarchical view

mobility density so that they can be easily accessed by people. Then, the distance between a user and restaurant  $j$  can be estimated as the distance between the center of region and  $l_j$ . To this end, we first cluster restaurants into  $\mathbb{R}$  regions. Then we select the region  $r$  where the restaurant  $j$  is associated, use  $r$  to represent the user activity area, and model the user-restaurant distance as  $d(i, j) = \|\mu_r - l_j\|_2$ . Finally, the geographic proximity influence can be evaluated by using a power-law like parametric term  $\rho_{ij} = (\frac{d_0}{d_0 + d(i,j)})^e$ , where  $d_0$  as a parameter. Basically, this means that the probability user  $i$  visits restaurant  $j$  decays as the power-law of the distance between them [7].

### 2.2.5 GC-BCoNMF

By combining user preference, profile and geographical proximity, we design the GC-BCoNMF model as a uniform four-layer hierarchical framework shown in Table 5. From top to the bottom, the first layer specifies how to generate the rating triplet based on overall, environment and cuisine utilities that user  $i$  obtains from restaurant  $j$ . The second layer describes how to extract the overall, environment and cuisine utilities. The third layer details the calculation of profile similarity and geographic proximity. Finally, the fourth layer presents all the latent factors, bias effect and variance.

### 2.3 Model Estimation

Let us denote all parameters by  $\Psi = \{\mathbf{U}, \mathbf{V}, \mathbf{E}, \mathbf{C}, \sigma^2, \lambda\}$ , and the hyperparameters by  $\Omega = \{\alpha, \beta, a, b\}$ , and the observed data collection by  $\mathcal{D} = \{y_{ij}, g_{ij}, h_{ij}, l_j\}^{I_{ij}}$  where  $l_j$  is the location of restaurant  $j$  and  $I_{ij}$  is the indicator function which equals to 1 if user  $i$  rates restaurant  $j$ . Given the  $\Omega$  and  $\mathcal{D}$ , our goal is to find the best parameters  $\Psi$  that maximize  $Pr[\mathcal{D}|\Psi, \Omega]$ . According to Table 5, the posterior probability is

$$(2.3) \quad P(\Psi|\mathcal{D}, \Omega) \propto P(Y|UV) \cdot P(G|UE) \cdot P(H|UC) \cdot P(U|\alpha) \cdot P(V|\beta) \cdot P(E|\beta)P(C|\beta)P(\sigma^2)$$

The region over each restaurant is observable.

$P(l_j|\mu_r, \Sigma_r)$  is a constant and can be ignored in posterior. In Equation 2.3,  $Y, G, H$  are generated by

$$(2.4) \quad \begin{aligned} P(Y|UV) &= \prod_{i=1}^I \prod_{j=1}^J \frac{1}{\sigma} \exp \left( -\frac{(y_{ij} - [(\lambda f + UV_{ij}) \cdot \delta_{ij} \cdot \rho_{ij}])^2}{2\sigma^2} \right) \\ P(G|UE) &= \prod_{i=1}^I \prod_{j=1}^J \frac{1}{\sigma} \exp \left( -\frac{(g_{ij} - UE_{ij})^2}{2\sigma^2} \right) \\ P(H|UC) &= \prod_{i=1}^I \prod_{j=1}^J \frac{1}{\sigma} \exp \left( -\frac{(h_{ij} - UC_{ij})^2}{2\sigma^2} \right) \end{aligned}$$

Moreover, since the elements in  $U, V, G, H$  also follow an exponential distribution, we define

$$(2.5) \quad \begin{aligned} P(U|\alpha) &= \prod_{i=1}^I \prod_{k=1}^K \alpha \cdot \exp(-\alpha \cdot U_{ik}) \\ P(V|\beta) &= \prod_{j=1}^J \prod_{k=1}^K \beta \cdot \exp(-\beta \cdot V_{kj}) \\ P(E|\beta) &= \prod_{j=1}^J \prod_{k=1}^K \beta \cdot \exp(-\beta \cdot E_{kj}) \\ P(C|\beta) &= \prod_{j=1}^J \prod_{k=1}^K \beta \cdot \exp(-\beta \cdot C_{kj}) \end{aligned}$$

Hence, by inferring Equation 2.3, we have the log of the posterior for the GC-BCoNMF model:

$$(2.6) \quad \begin{aligned} \mathcal{L}(U, V, E, C, \sigma^2, \lambda|Y, G, H, \alpha, \beta) &= \sum_{i=1}^I \sum_{j=1}^J \left[ -\frac{1}{2} \ln \sigma^2 - \frac{[y_{ij} - (\lambda f + UV_{ij}) \cdot \delta_{ij} \cdot \rho_{ij}]^2}{2\sigma^2} \right] \\ &+ \sum_{i=1}^I \sum_{j=1}^J \left[ -\frac{1}{2} \ln \sigma^2 - \frac{[g_{ij} - UE_{ij}]^2}{2\sigma^2} \right] \\ &+ \sum_{i=1}^I \sum_{j=1}^J \left[ -\frac{1}{2} \ln \sigma^2 - \frac{[h_{ij} - UC_{ij}]^2}{2\sigma^2} \right] \\ &- \sum_{i=1}^I \sum_{k=1}^K \alpha \cdot U_{ik} - \sum_{j=1}^J \sum_{k=1}^K \beta \cdot V_{kj} - \sum_{j=1}^J \sum_{k=1}^K \beta \cdot E_{kj} - \sum_{j=1}^J \sum_{k=1}^K \beta \cdot C_{kj} \\ &- ((a+1) \ln \sigma^2 + \frac{b}{\sigma^2}) \end{aligned}$$

For simplicity, let  $\Delta_1 = \lambda f$ ,  $\Delta_2 = \delta_{ij} \cdot \rho_{ij}$ ,  $\Delta_3 = \sum_{k' \neq k}^K U_{ik'} V_{k'j}$ ,  $\Delta_4 = \sum_{k' \neq k}^K U_{ik'} E_{k'j}$ ,  $\Delta_5 = \sum_{k' \neq k}^K U_{ik'} C_{k'j}$ . We apply iterated conditional modes (ICM) [19] method and obtain the updating rules for the parameters set  $\Psi$ :

$$(2.7) \quad \begin{aligned} U_{ik} &= \frac{\sum_{j=1}^J \left[ \frac{V_{kj} \Delta_2 (y_{ij} - \Delta_1 \Delta_2 - \Delta_2 \Delta_3)}{+ E_{kj} (g_{ij} - \Delta_4) - C_{kj} (h_{ij} + \Delta_5)} \right] - \alpha \sigma^2}{\sum_{j=1}^J [\Delta_2^2 V_{kj}^2 + E_{kj}^2 + C_{kj}^2]} \\ V_{kj} &= \frac{\sum_{i=1}^I (y_{ij} - \Delta_1 \Delta_2 - \Delta_2 \Delta_3) \Delta_2 U_{ik} - \beta \sigma^2}{\sum_{i=1}^I \Delta_2^2 U_{ik}^2} \\ E_{kj} &= \frac{\sum_{i=1}^I U_{ik} (g_{ij} - \Delta_4) - \beta \sigma^2}{\sum_{i=1}^I U_{ik}^2} \\ C_{kj} &= \frac{\sum_{i=1}^I U_{ik} (h_{ij} - \Delta_5) - \beta \sigma^2}{\sum_{i=1}^I U_{ik}^2} \\ \sigma^2 &= \frac{[y_{ij} - (\Delta_1 + UV_{ij}) \Delta_2]^2 + [g_{ij} - UE_{ij}]^2 + [h_{ij} - UC_{ij}]^2 + b}{\frac{3IJ}{2} + a - 1} \end{aligned}$$

In addition, a gradient descend method is applied

to update  $\lambda$ :  $\lambda^{t+1} = \lambda^t - \eta \frac{\partial(-\mathcal{L})}{\partial \lambda}$  to minimize  $-\mathcal{L}$ .

## 2.4 Recommendation

After the parameters are learned, GC-BCoNMF predicts the overall rating of restaurant  $j$  given by user  $i$  as  $\mathbb{E}(y_{ij}|\Psi) = (\lambda f + \mathbf{u}_i^\top \mathbf{v}_j) \cdot \delta_{ij} \cdot \rho_{ij}$ , which is used for restaurant recommendations. The larger the predicted value, the more likely the user will choose the restaurant.

## 3 Experimental Results

This section presents an empirical evaluation of the proposed GC-BCoNMF model on a real-world dataset.

### 3.1 Data Description

We use a restaurant dataset which covers seven months period from July, 2010 to February, 2011 from three main areas in Mexico: Ciudad Victoria, San Luis Potosi, Cuernavaca. In addition to the user-restaurant rating records, there are many rich associated attributes for each rating event as described in Table 2. The dataset comprises 1,165 ratings from 138 users for about 130 restaurants; and the rating sparsity is 93.5%. The rating values range from 1, 2, to 3, where 1 indicates that the user does not like the restaurant, and 3 denotes a high preference. Due to the privacy issue, it is extremely difficult to get a larger restaurant data which contains multi-aspect ratings, the detailed user profile and restaurant profile information. Even the data we use in this paper has small size, it possesses characteristics of commonly-used recommendation data. For instance, the average number of ratings is about 8.4 per user and 8.9 per restaurant; half of the ratings are on the 38 best rated restaurants. A small number of users, restaurants or cities dominate the ratings whereas many users, restaurants or cities obtain only a few. Those observations present the power law distribution, which is usually found on recommendation data.

### 3.2 Evaluation Metrics

We will evaluate the model with both prediction accuracy and top-N recommendation performance.

**Prediction Accuracy:** We adopt two metrics to measure the prediction accuracy: (1) the Mean Absolute Error,  $MAE = \sum_{i,j} |y_{ij} - \hat{y}_{ij}|/N$ ; (2) the Root Mean

Square Error,  $RMSE = \sqrt{\sum_{i,j} (y_{ij} - \hat{y}_{ij})^2/N}$ , where  $y_{ij}$  and  $\hat{y}_{ij}$  denote the observed overall rating and the predicted overall rating, and  $N$  denotes the total number of the tested data. The smaller the value of MAE or RMSE, the more precise a recommendation.

**Top-N Recommendation:** We also evaluate the algorithms in terms of ranking. We present each user with  $N$  restaurants sorted by their predicted overall rating and evaluate based on Normalized Discounted Cumulative Gain (NDCG). The NDCG@N metric assumes that highly relevant restaurant should appear earlier in

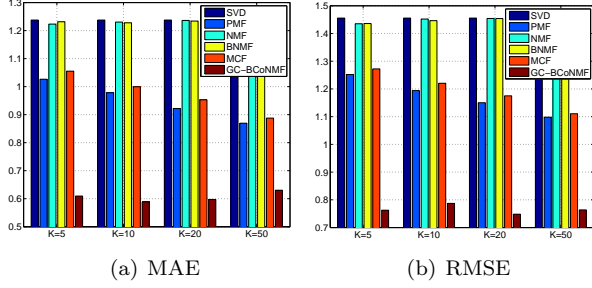


Figure 3: The average MAE and RMSE values at different latent dimensions  $K$ .

the recommendation list. Discounted Cumulative Gain (DCG) at  $N$  is defined by

$$DCG[n] = \begin{cases} rel_1 & \text{if } n = 1 \\ DCG[n-1] + \frac{rel_n}{\log_2 n}, & \text{if } n \geq 2 \end{cases}$$

Given the ideal DCG  $IDCG$ , NDCG@ $N$  can be computed as  $NDCG[n] = \frac{DCG[n]}{IDCG[n]}$ .

### 3.3 Evaluation Methods

To show the effectiveness of GC-BCoNMF, we compare the recommendation performances of our model against following baseline algorithms: (1) **Singular Value Decomposition (SVD)**[12]. (2) **Probabilistic Matrix Factorization (PMF)**[18]. (3) **Nonnegative Matrix Factorization (NMF)**[13]. (4) **Bayesian Nonnegative Matrix Factorization (BNMF)** [19]. (5) **Multi-criteria Collaborative Filtering (MCF)** [1], which is delicately designed for multi-criteria (also called multi-rating or multi-aspect) recommendation.

To understand the contribution of different factors, we test the following methods: (1) **Bias Effect (BE-BCoNMF)**: only considering the bias effect in BCoNMF (Section 2.2.2). (2) **Profile Similarity (PS-BCoNMF)**: only incorporating profile similarity (Section 2.2.3) into BCoNMF. (3) **Geographic Proximity (GP-BCoNMF)**: only fusing geographic proximity (Section 2.2.4) to BCoNMF. (4) **Combination (GC-BCoNMF)**: the proposed method (Section 2.2.5).

### 3.4 Implementation Details

We set  $\lambda_U = 0.01$  and  $\lambda_V = 0.01$  for PMF. For GC-BCoNMF, we set  $\alpha = 1.3$ ,  $\beta = 1.1$ ,  $a = 1$ ,  $b = 1$  for hyperparameters and  $\lambda = 0.1$ . Also, we set the number of regions  $\mathbb{R} = 3$  (the number of cities in the data). We pre-cluster the restaurants into three regions and compute  $\mu_r$  and  $\Sigma_r$ . We use 5-fold cross validation.

### 3.5 Overall Performances

Here, we present the performance comparison of overall rating on accuracy and top- $N$  recommendation.

**Prediction Accuracy.** We compare MAE and RMSE of our method against baseline methods with four latent dimensions  $K=5, 10, 20, 50$ . The MAE comparisons are shown in Figure 3(a). Overall, GC-BCoNMF

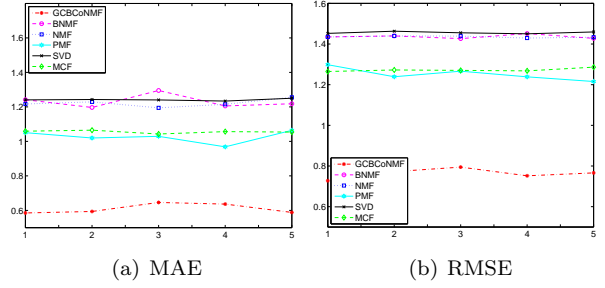


Figure 4: MAE and RMSE values in each fold ( $K=5$ ). outperforms baselines with significant margins, because the MAE values of GC-BCoNMF (average in 0.6) are much lower than the MAE values of SVD, PMF, NMF, BNMF or MCF (average in 1.1). In addition, in all the baselines, MCF and PMF works slightly better than SVD, NMF and BNMF in all  $K$  value settings. Figure 3(b) shows the RMSE comparisons: GC-BCoNMF outperforms all the competing models. For example, the average RMSE of the five baselines is about 1.3 whereas the RMSE of GC-BCoNMF is only 0.7. Figure 4(a) and 4(b) further show the MAP and RMSE of each approaches in each test fold during cross validation with latent dimension  $K=5$ . In cross validation, GC-BCoNMF is stable and consistently outperforms other comparable baselines. Specifically, we conduct a series of paired T-tests with 0.95 confidence at different  $K$ . The results show that the improvement of GC-BCoNMF on both MAE and RMSE are statistically significant.

**Top- $N$  Recommendation.** Figure 5 shows the top- $n$  recommendation performances of GC-BCoNMF and five baselines in terms of NDCG@ $N$  ( $N=1, 3, 5$  respectively). In the experiment, we choose those who rate more than  $N$  restaurants. Then, we apply all the six methods to recommend a  $N$ -length restaurant list to those users and calculate the NDCG@ $N$  of the Top- $N$  recommendation against its ground truth. As can be seen in Figure 5(a), 5(b) and 5(c), GC-BCoNMF still achieves the best results with list length  $N=1, 3, 5$  and latent dimension  $K=5, 10, 20$ . Similarly, Figure 5(d) demonstrates the NDCG@5s of GC-BCoNMF in each test round are consistently higher than other comparable methods when  $K=5$ . The paired T-tests with 0.95 confidence of different NDCG@ $N$  also proves the statistical significance of GC-BCoNMF’s improvements.

The above analysis shows the strength of exploiting rating triplet as well as combining the profile similarity and geographic influence.

### 3.6 The Study on Average Rating

We also study the performance of predicting user average rating and restaurant average rating under GC-BCoNMF and other baselines. We set the latent dimension  $K=5$  and calculate the average rating



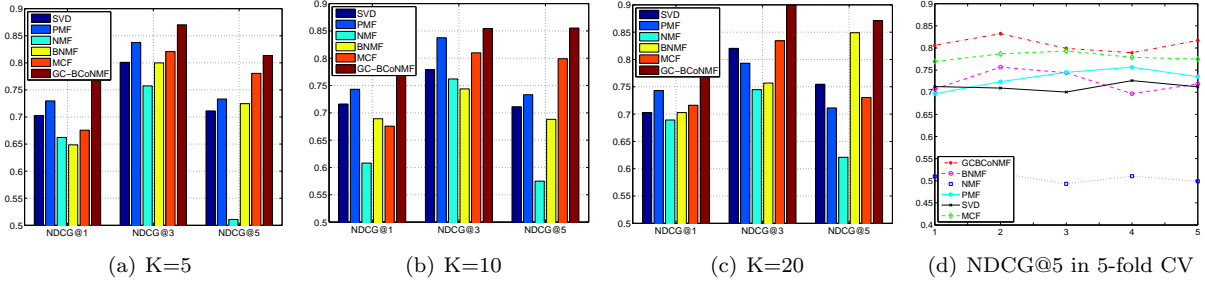


Figure 5: Average NDCG @N at different latent dimensions  $K$  and NDCG@5 during cross validation ( $K=5$ ).

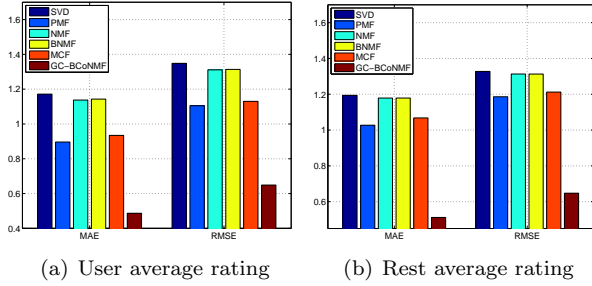


Figure 6: Performance comparison on user average rating and restaurant average rating ( $K=5$ ).

per user and per restaurant. Figure 6 shows the performance comparison: GC-BCoNMF achieves the lowest RMSE (0.648) and MAE (0.487) on user average rating shown in Figure 6(a), and also the lowest RMSE (0.647) and MAE (0.512) on restaurant average rating comparing with baselines shown in Figure 6(b).

### 3.7 The Study at Different Influential Factors

Here, we evaluate the influence of different factors: bias effect (BE-BCoNMF), profile similarity (PS-BCoNMF), geographic proximity (GP-BCoNMF) and combination (GC-BCoNMF). As shown in Figure 7: (1) combination of the three factors achieve the lowest MAE and RMSE comparing with other individual factors; (2) in all the individual factors, profile similarity perform best and achieve around 0.63 in MAE and 0.80 in RMSE; (3) bias effect contribute less than profile similarity but much larger than geographic proximity; (4) geographic proximity contribute least with MAE at 0.77 and RMSE at 1 but still lower than other five baselines. This experiment comes to a conclusion: combination > profile similarity > bias effect > geographic proximity.

### 3.8 The Study on Cuisine and Environment

Beside overall ratings, we evaluate the prediction accuracy of cuisine ratings and environment ratings and check if GC-BCoNMF can correctly capture the relations among rating triplet. Figure 8(a) and Figure 8(b) report the MAEs and RMSEs of cuisine rating prediction comparing with other five baselines. With average in 0.75 on RMSE and 0.6 on MAE, GC-BCoNMF out-

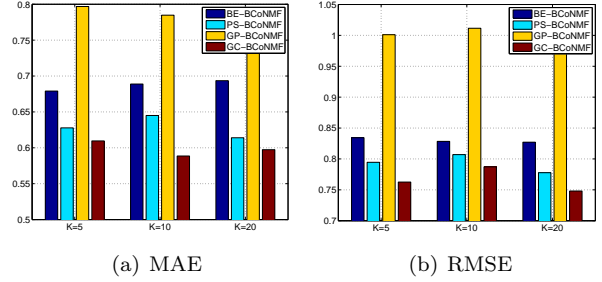


Figure 7: MAE and RMSE (influential factors).

Predictor	Coefficient	Std. Error	t value	Pr(>  t )
cuisine	0.48936	0.03330	14.70	<2e-16 ***
environment	0.51071	0.03502	14.59	<2e-16 ***

Residual standard error: 0.5104 on 498 degrees of freedom Multiple R-squared: 0.9524, Adjusted R-squared: 0.9523 F-statistic: 4987 on 2 and 498 DF, p-value: < 2.2e - 16

Table 6: The coefficient estimation regressed from overall, cuisine, environment ratings

performs the five baselines, whose RMSEs are larger than 1.08 and MAEs are larger than 0.85. Similarly, in the MAEs and RMSEs of environment ratings, GC-BCoNMF is better than SVD, PMF, NMF, BNMF and MCF with significant margins as shown in Figure 8(c) and Figure 8(d). The consistent high accuracy in both cuisine and environment ratings proves the correctness of jointly factorizing the restaurant rating triplet. We also proposed a linear regression method with a hypothesis: for each restaurant with overall rating to be predicted, we assume the cuisine and environment ratings of this restaurant for every user are available. Then, by assuming  $overall = a_1 * cuisine + a_2 * environment$ , we are able to predict overall ratings shown in Table 6. This method can achieve 0.5416553 in RMSE. The performance is comparable to GC-BCoNMF, which assumes that overall, cuisine and environment ratings are partially known in reality.

### 3.9 The Study on Different Cities

Here, we study the performance of the proposed model at different cities. For citywide restaurant recommendation, the geographic isolation between cities and the lack of ratings in small cities lead to the sparseness and im-

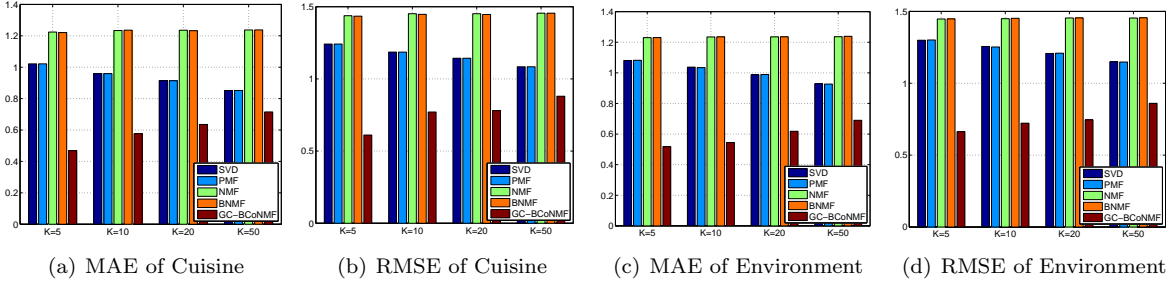


Figure 8: MAE and RMSE of cuisine rating and environment rating comparing with other methods.

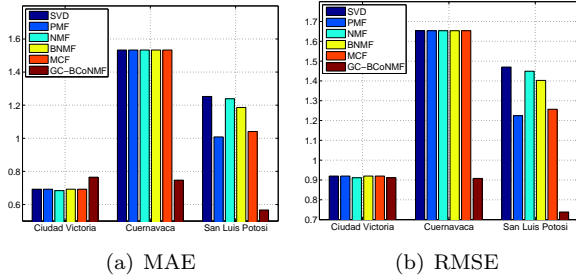


Figure 9: MAE and RMSE on different cities (K=5).

balance challenge. A good recommender system should not only works in big cities with lots of rating data, but also in small cities. Thus, we first report the MAE and RMSE of the six methods on three cities. As shown in Figure 9, in general, our proposed model achieves the lowest MAE and RMSE in most of the three cities. Whereas, the five baselines only perform relatively good in Ciudad Victoria, but achieve much higher MAE and RMSE in Cuernavaca and San Luis Potosi than that of GC-BCoNMF. The above shows the robust performance of GC-BCoNMF in multiple cities.

### 3.10 Summary

Traditional matrix factorization methods assume the overall ratings are generated by user and restaurant latent factors. However, in restaurant recommendation, this assumption fails to comprehensively describe the latent generative structure of restaurant review data. We consider the intercorrelation among overall, cuisine and environment ratings contains users' in-depth interest for restaurant. GC-BCoNMF takes into account both rating triplet and geospatial influence and establishes a latent generative structure for restaurant review data. The incorporation of profile similarity as side information overcomes the data sparseness challenge led by citywide geo-isolation. Hence, we can observe significant improvements against all the baseline algorithms.

## 4 Related Work

Related work can be grouped into two categories. The first one includes the work on restaurant recommendations. In the second category, we present the related

collaborating filtering methods.

Some previous studies on restaurant recommendation relied on knowledge based approach [21], which used knowledge about users and restaurants to generate recommendations. For example, work [8, 21] requested favorite restaurants, demographic information or filtering rules and reasoned out recommended restaurants by applying similarity based retrieval. Knowledge-based approaches utilize implicit assumptions on important features by knowledge engineering and are independent of user ratings. However, since the knowledge is unchanged and reused, the recommendations remain static. By considering rich rating event profiles including user and restaurant description, work [4] recommended restaurants using a rule-based strategy. In contrast, restaurant recommender system can also be developed in term of users' past preferences via matrix factorization (MF), such as SVD [11], PMF [18], Bayesian probabilistic matrix factorization (BPMF) [17], NMF [13], BMNF [19]. One drawback of MF is the so-called ramp-up problem, in which MF performs poorly when the recommender system is not initialized with a sufficient number of ratings. Therefore, more studies such as [6, 3, 5, 23, 9] incorporated side and prior information into latent factor models by exploring user and item observable features, or even textual data.

More recent studies explored geographic patterns to better understand the geographic influence on the decision process and provide more accurate location based recommendation. For example, previous works [24, 22] applied the collaborative filtering method to recommend locations based on trajectory data or check-in data. More recently, the work [7, 16] considered the multi-center of user check-in patterns, user mobility, and incorporated the inverse effect of geographic proximity as well as social influence for POI recommendation.

In addition, our work also has a connection with multi-criterion recommendation systems. For example, the work [14] reduced the problem to a multiple criteria ranking task and apply an adapted skyline query algorithm. Paper [1, 2] extended the single criteria CF to include multi-criteria ratings in the calculation of the



similarity between users or items by applying average similarity or worst-case similarity schema. Besides, [1] also proposed to use standard CF to predict single criteria rating, then a weighted linear regression function was applied to aggregate multiple criteria ratings to a global rating. More recent works [10, 20] proposed to view rating matrix as a pairwise relationship between two participants and simultaneously factorized multiple rating matrices into latent space. Although [20] generalize the problem of collective factorization, they only provide the model inference for a  $U \times I \rightarrow R1 \times R2$  case, which only suit for two rating matrices. Besides, [10, 20] lack of an integrated analysis of the interinfluence of multi-rating, geography and side information. In our paper, we adopted and adapted the idea of CoMF and explored multi-aspect ratings.

In summary, while there are some work on restaurant recommendation, we provide a systematic way for restaurant recommendation by collectively exploiting multi-aspect ratings, geographic information, and profile information of users and restaurants.

## 5 Conclusion

In this paper, we provided an integrated analysis of multiple factors which collectively influence the restaurant selection process of customers. Specifically, we proposed a hierarchical framework to learn the inner preference with multiple information fusion for restaurant recommendation. The proposed collaborative probabilistic factor analysis framework strategically takes rating triplet, profile similarity and geographic influence into the consideration. There are several advantages of the proposed recommendation method. First, the model could exploit the multi-aspect ratings of restaurants to uncover the user preference more precisely. Second, the method incorporates profile similarity, bias effect, and geographic proximity influence as well to overcome the geographical isolation challenge. It can balance the recommendation accuracy in different cities when it comes to a city-wide case. Finally, experimental results on real-world restaurant rating data validated the performances of the proposed method.

**Acknowledgement.** This research was partially supported by National Science Foundation (NSF) via grant numbers CCF-1018151 and IIS-1256016. Also, it was supported in part by Natural Science Foundation of China (71329201).

## References

- [1] G. Adomavicius and Y. Kwon. New recommendation techniques for multicriteria rating systems. *Intelligent Systems, IEEE*, 22(3):48–55, 2007.
- [2] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Trans. Inf. Syst.*, 23:103–145, 2005.
- [3] D. Agarwal and B.-C. Chen. Regression-based latent factor models. In *KDD '09*, 2009.
- [4] V. Blanca, G. Gabriel, and P. Rafael. Effects of relevant contextual features in the performance of a restaurant recommender system. In *3rd Workshop on Context-Aware Recsys*, 2011.
- [5] B.-C. Chen, J. Guo, B. Tseng, and J. Yang. User reputation in a comment rating environment. In *KDD '11*, 2011.
- [6] T. Chen, Z. Zheng, Q. Lu, W. Zhang, and Y. Yu. Feature-based matrix factorization. *arXiv preprint arXiv:1109.2271*, 2011.
- [7] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *AAAI*, 2012.
- [8] P. Dwivedi and N. Chheda. A hybrid restaurant recommender. *International Journal of Computer Applications*, 55(16):20–25, 2012.
- [9] Y. Ge, Q. Liu, H. Xiong, A. Tuzhilin, and J. Chen. Cost-aware travel tour recommendation. In *KDD*, pages 983–991. ACM, 2011.
- [10] L. Hong, A. S. Doumith, and B. D. Davison. Co-factorization machines: modeling user interests and predicting individual decisions in twitter. In *WSDM*, pages 557–566. ACM, 2013.
- [11] Y. Koren. Factorization meets the neighborhood: a multi-faceted collaborative filtering model. In *KDD '08*, 2008.
- [12] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42:30–37, 2009.
- [13] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS '00*, 2000.
- [14] H.-H. Lee and W.-G. Teng. Incorporating multi-criteria ratings in recommendation systems. In *IEEE IRI 2007*, pages 273–278, 2007.
- [15] B. Li, A. Ghose, and P. G. Ipeirotis. Towards a theory model for product search. In *WWW '11*, 2011.
- [16] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. *KDD '13*, pages 1043–1051, 2013.
- [17] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *ICML '08*, 2008.
- [18] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2008.
- [19] M. N. Schmidt, O. Winther, and L. K. Hansen. Bayesian non-negative matrix factorization. In *ICA '09*, 2009.
- [20] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. In *KDD*, pages 650–658. ACM, 2008.
- [21] S. Trewin. Knowledge-based recommender systems. *Encyclopedia of library and information science*, 69(Supplement 32):69, 2000.
- [22] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR '11*, 2011.
- [23] L. Zhang, D. Agarwal, and B.-C. Chen. Generalizing matrix factorization through flexible regression priors. In *RecSys '11*, 2011.
- [24] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *WWW*, pages 1029–1038, 2010.