Your Neighbors Affect Your Ratings: On Geographical Neighborhood Influence to Rating Prediction

Longke Hu
lhu003@e.ntu.edu.sg

Aixin Sun
axsun@ntu.edu.sg

Yong Liu
liuy0054@e.ntu.edu.sg

School of Computer Engineering, Nanyang Technological University, Singapore

ABSTRACT

Rating prediction is to predict the preference rating of a user to an item that she has not rated before. Using the business review data from Yelp, in this paper, we study business rating prediction. A business here can be a restaurant, a shopping mall or other kind of businesses. Different from most other types of items that have been studied in various recommender systems (*e.g.,* movie, song, book), a business physically exists at a geographical location, and most businesses have geographical neighbors within walking distance. When a user visits a business, there is a good chance that she walks by its neighbors. Through data analysis, we observe that there exists *weak positive correlation* between a business's ratings and its neighbors' ratings, regardless of the categories of businesses. Based on this observation, we assume that a user's rating to a business is determined by both the intrinsic characteristics of the business and the extrinsic characteristics of its geographical neighbors. Using the widely adopted latent factor model for rating prediction, in our proposed solution, we use two kinds of latent factors to model a business: one for its intrinsic characteristics and the other for its extrinsic characteristics. The latter encodes the neighborhood influence of this business to its geographical neighbors. In our experiments, we show that by incorporating geographical neighborhood influences, much lower prediction error is achieved than the state-of-the-art models including Biased MF, SVD++, and Social MF. The prediction error is further reduced by incorporating influences from business category and review content.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval— *Information Filtering*

Keywords

Recommendation; Rating prediction; Matrix factorization; Yelp

1. INTRODUCTION

Recommender systems have attracted significant attention from both academia and industry since the last decade. Various recom-

mender systems have been developed to facilitate the matching between consumers (*i.e.,* users) with appropriate products or services (*i.e.,* items). Example items include songs, movies from content providers as well as books from E-commerce websites. Recently, recommender systems have also been applied to social network platforms (*e.g.,* Facebook, Twitter, Linkedin) for people recommendation, *e.g.,* recommending friends to a user or recommending who to follow. The prevalence of GPS-enabled devices (*e.g.,* smart phones) in the past few years further extends the landscape of recommender systems in location-based social networks (LBSN), exemplified by Foursquare and Gowalla.

Depending on the application, different recommendation problems have been defined and studied. The top-N item recommendation and rating prediction are two most widely studied categories of recommendation problems. On the one hand, top-N item recommendation tasks aim to recommend a user a list of items that she may be interested in. For example, in LBSN, *POI recommendation* aims to recommend unvisited POIs to users. Here POI stands for point-of-interest, referring to a focused geographic entity such as district and street, or a specific point location such as landmarks and restaurants [20]. Rating prediction, on the other hand, is to predict the preference rating of a user to a product or service that she has not rated before. The products or services with high predicted ratings are recommended to users. In this study, we are interested in the ***business rating prediction*** problem with business review data from Yelp.

Yelp is a business review site and has attracted 47 million reviews to local businesses since 2004. Example businesses in Yelp include restaurants, shopping malls, beauty & spas, etc.. The website reports that it had an average of approximately 117 million monthly unique visitors in the third quarter of 2013.[1] A Yelp user or Yelper can share her experience with a business by posting a review of the business and also a rating from 1 to 5 stars. Our task is therefore to *predict how many stars a user would give to a business that she has yet reviewed.*

At first glance, rating prediction of business is the same as prediction of user's rating to any other kind of items (*e.g.,* a song or a movie), with the only difference that the item here refers to a business. The key difference, between a business and other kind of items that have been studied in literature, is that a business physically exists at a specific geo-location with latitude/longitude coordinates. More importantly, most businesses (*e.g.,* restaurants and shops) are not geographically isolated from others. That is, when a user visits a business, there is a good chance that she walks by its neighbors if they are located within walking distance. The overall environment of that region might affect a user's view about a business and subsequently affect user's review and rating to the

---

[1]http://www.yelp.com/about

business. For example, the hygiene standard of a region might affect user's rating to many restaurants located in that region. On the other hand, a region distinguishes itself from other regions and becomes attractive to visitors often because there are a few good businesses co-located in the region. For business rating prediction, an interesting question here is: *Is there any correlation between a business's rating and the rating of its geographical neighbors*?

To answer the question above, we conducted data analysis on Yelp's business rating data. We observe that there does exist *weak positive correlation* between the rating of a business and the rating of its neighbors, regardless of the category of the business. Based on this observation, we incorporate geographical neighborhood influence into our business rating prediction model which is based on the widely adopted latent factor model realized by Matrix Factorization (MF). Together with influences from other factors including user reviews, business category, and business popularity, we show that the proposed model outperforms state-of-the-art baselines including Biased MF, SVD++ and Social MF [12, 17], measured by both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). To the best of our knowledge, this is the first study exploiting geographical neighborhood influence in business rating prediction. We note that geographical influence has been considered in POI recommendation and POI prediction. However, we argue that both POI recommendation and POI prediction are different research problems from business rating prediction (see Section 2 for more details). To summarize, the main contributions arise from this study are as follows:

- We conduct data analysis and observe that there exists weak positive correlation between a business's average rating and the average rating of its neighbors, regardless of the categories of businesses. This is an important observation that could be useful for not only business rating prediction, but also related studies, *e.g.,* POI recommendation and sentiment analysis of business reviews.

- We directly model the influence of geographical neighborhood into business rating prediction using matrix factorization. Specifically, for each business, we use two different latent factors to represent its intrinsic and extrinsic characteristics respectively. The influences from geographical neighbors of a business are then modeled as the linear combination of the latent factors for the extrinsic characteristics of its neighbors. In our recommender system, we have also considered other factors including business's review, category, and popularity. To the best of our knowledge, this is the first study that models geographical neighborhood influence into business rating prediction.

- We conducted extensive experiments to evaluate the effectiveness of incorporating influence of geographical neighborhood and other factors in business rating prediction and compared the prediction accuracy of the proposed model with an array of strong baseline models. We further show that the geographical distance between a user and a business adversely affects business rating prediction, although this is an important and effective factor in POI recommendation problems.

The rest of this paper is organized as follows. We review the related work on rating prediction, POI recommendation and POI prediction in Section 2. The data analysis is reported in Section 3. The details of the proposed business rating prediction model is presented in Section 4, followed by experimental evaluation in Section 5. Finally, Section 6 concludes this paper.

2. RELATED WORK

In this section, we briefly review the recent advances in recommender systems, mainly focusing on the applications of collaborative filtering on rating prediction. As one of the major contributions of this study is the influence of geographical neighborhood, we also review the related work employing geographical influence for POI recommendation and POI prediction. We then highlight the major differences between our research problem and these two problems.

2.1 Collaborative Filtering

Collaborative filtering (CF) is the most widely used technique in recommender systems [11, 12, 18, 19]. The underlying assumption of CF is that similar users would rate items similarly or similar items would receive similar ratings from users. Ratings here indicate users' preferences to items. There are mainly two kinds of approaches, namely, *memory-based* CF and *model-based* CF.

The basic idea of memory-based CF is to find similar users or items by using similarity measures, *i.e.,* user-based CF [1, 10] and item-based CF [8, 21]. The commonly used similarity measures are Pearson's correlation and cosine similarity, computed based on user-item ratings as well as other user-/item-specific features depending on the application. The similar users or items are also known as neighbors and therefore memory-based CF is also known as neighborhood-based CF. However "neighbor" in this context is defined by the chosen similarity measure. Different similarity measures lead to different sets of neighbors for a given user or item. In this paper, the neighborhood of a specific business refer to the businesses that are located physically close to it, *i.e.,* geographical neighborhood.

Model-based CF usually employs machine learning techniques to build models from the observed user-item ratings, and then to predict the unobserved ratings. Latent factor model is one of the most successful CF models, in which users and items are jointly mapped into a shared latent space of much lower dimensionality. As the most successful realization of latent factor model, matrix factorization (MF) [12, 14, 17] has been successfully applied to various recommender systems including music rating prediction for Yahoo! Music [11] and Last.fm [24], movie rating prediction for Netflix [12, 13] and Douban [17, 24], and personalized tweet recommendation [2].

Among various MF models proposed, SVD++ [12] is probably one of the most successful models. This model integrates the implicit feedback information from a user to items (*e.g.,* based on user's purchase history or browsing history). More specifically, the user vector of latent factors in this model is complemented by the latent factors of the items to which the user has provided implicit feedback. Recently, Ma [17] proposed a social regularization MF method, named Social MF, to employ the similar and dissimilar relationships between users and/or items to improve recommendation accuracy. The similarity between items is measured based on their ratings using Pearson's correlation or cosine similarity. In our experiments, the proposed models achieve better rating prediction accuracy than Social MF indicating that the influence from geographical neighborhood is more effective than the influence from neighbors chosen by rating-based similarity measures.

Based on MF, influence from other aspects of users or items besides the ratings can be flexibly and easily modeled. For example, Koenigstein *et al.* [11] incorporated rich item bias into MF model to capture the taxonomy information of music. Each music has multiple types of information such as track, album, artist and genre. In their proposed model, MF was extended by adding shared bias parameters for items linked by common taxonomy. Moreover, some other work has also shown that popularity is helpful in improving

the recommendation accuracy [7, 22]. To capture users' interests over tweet content, Chen *et al.* [2] proposed to use topic-level latent factors. Instead of directly asking whether a user is interested in a tweet, the model captures user's preference over the words in the tweet. In our problem setting, we consider the review from a user further elaborates about the rating. We model the words in review as latent factors and incorporate the review into rating prediction.

2.2 POI Recommendation and Prediction

POI recommendation has attracted significant attention recently with the popularity of LBSN [3, 6, 25, 26]. Because geographical neighborhood is the main focus in our study, there is a need to review recent advances in POI recommendation. The main research issue in POI recommendation is to accurately recommend unvisited POIs to users. Geographical influence and temporal influence are major considerations in POI recommendation [25, 26]. Geographical influence is based on the observation that users tend to visit nearby POIs of their home or office locations, as well as nearby locations of the POIs in their favor. Temporal influence is reflected by the observation that users check-in different types of POIs at different time slots of a day (*e.g.,* restaurant in lunch hour and bar in night). Moreover, POI recommendation has also considered social influence among friends [25].

POI prediction aims to predict which POI a user would visit next given her current location and time [4, 6]. Different from POI recommendation, the POI predicted for a user to be visited next may have been visited by the user before. Similarly, both geographical influence and temporal influence have been considered in the modeling. Liu *et al.* [16] proposed a category-aware POI prediction model by exploiting users' preference transition over location categories. In this model, MF is utilized to predict a user's preference transitions over categories and then the locations in the corresponding categories by considering geographical influences. The model was evaluated on Gowalla check-in data with three-level category hierarchy. In our model, we also consider category information of the business. However, we directly incorporate category latent factors into our factorization model to reflect the affinity of businesses linked by the categories they belong to. Note that, categories in the Yelp dataset are not organized in hierarchy.

In this paper, we study the impact of geographical neighborhood influence to business rating prediction. We argue that the geographical neighborhood influence in our research problem is essentially different from the geographical influence in POI recommendation and/or prediction. For POI recommendation/prediction, the geographical influence is more related to the cost of travel (*e.g.,* time cost or monetary cost) from a user's point of view. For rating prediction, on the other hand, we predict a user's degree of preference to a yet reviewed business. The cost of travel here is expected to be a less important factor. The contextual factors (*e.g.,* the current location, the time of visiting) which have demonstrated effective in POI recommendation/prediction are also less relevant to our research problem. In our research problem, geographical neighborhood influence is more related to the business environment created by the surrounding businesses to one business at a specific location. Next, we conduct data analysis of the Yelp data.
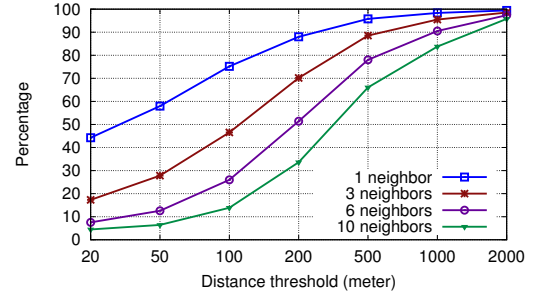
3. DATA ANALYSIS

3.1 Dataset

Our study is based on the recently released Yelp Dataset Challenge.[2] The data is sampled by Yelp from the greater Phoenix, AZ

Table 1: Number of reviews/ratings per user/business

Statistics	User	Business
Minimum number of reviews/ratings	1	3
Maximum number of reviews/ratings	473	540
Average number of reviews/ratings	5.00	19.93

Figure 1: Percentage of businesses having at least 1, 3, 6, 10 neighbors within a distance threshold from 20 to 2000 meters

metropolitan area from March 2005 to January 2013. It contains 11,537 businesses, 229,907 reviews by 43,873 users, and 8,282 check-in sets. This dataset was used as training data in ACM RecSys Challenge 2013.[3]

A *business* has a unique id, name, address, latitude longitude, its categories and some other attributes like city, state, and neighborhoods. However, we observe that the neighborhoods attribute is empty in the dataset. A *review* contains business id, user id, rating from 1 to 5 stars, date, review text, and voting. In this study, we do not use the voting feature for its less relevance to our research problem. Table 1 reports the minimum, maximum, and average number of ratings per user and per business respectively. A *check-in set* for a business contains the aggregated number of check-ins in every hour from Monday to Sunday. The largest number of check-ins to one business observed from the data is 22,977. Note that, the dataset does not provide detailed check-in records of which user checkin a business at what time, and there are only about 72% of businesses have check-in sets. The percentage becomes 85.13% among all businesses having at least 5 reviews. In other words, a business may have a number of reviews but no check-in records. The dataset does not provide much details of a *user*.

3.2 Observations

Tobler's First Law of Geography states "Everything is related to everything else, but near things are more related than distant things" [23]. Next, we present 3 observations made from the data with respect to geographical neighborhood.

OBSERVATION 1. *Most businesses have **neighbors within a short geographical distance** from their locations. More than 44% of businesses have one neighbor next to it within 20 meters and 95% of businesses have one neighbor within 500 meters.*

Based on the business latitudes and longitudes, we calculate the geographical distance between two businesses using Haversine formula.[4] Figure 1 plots the percentage of businesses having at least 1, 3, 6, and 10 neighbors respectively, within a geographical distance threshold ranging from 20 to 2000 meters. It shows that most businesses are not isolated geographically from others. Specifically,
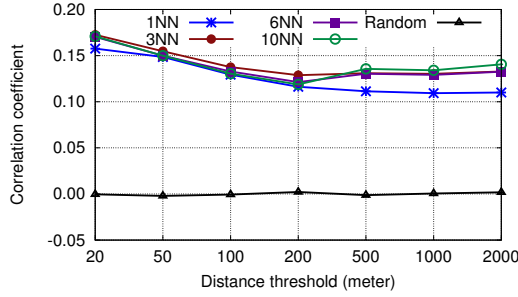
Figure 2: Pearson's correlation between a business's rating and the average rating of its {1, 3, 6, 10} nearest neighbors (NN), within a distance threshold.

44.3% of businesses have one neighbor next to it within 20 meters. The percentage rises to 95.6% if the distance threshold is set to 500 meters, a distance for a 6-minute walk.[5] Within this walking distance, about 80% of businesses have at least 6 neighbors, and 66% have at least 10 neighbors.
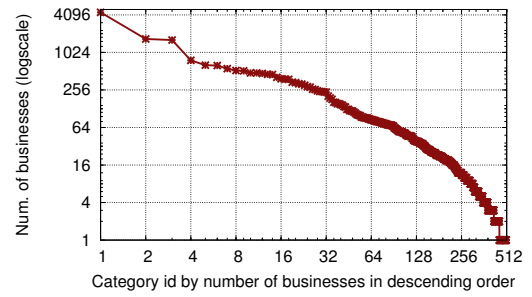
OBSERVATION 2. *The average rating of a business is **weakly positively correlated** with the average rating of its neighbors.*

Before exploring a new place, we often receive advice on which region is famous for good food (*e.g.,* a few famous restaurants collocating in that region), or advice on which region to avoid for shopping because the shops there are infamous for cheating. This phenomenon of "things of one kind come together" is also reflected from the positive correlation between the ratings of businesses and their geographical neighbors.

We calculate the average rating of a business from all its user reviews. Hereafter, for easy presentation, we refer this average rating simply as *the rating of a business* or *a business's rating* when the context is clear. Figure 2 plots the Pearson's correlation coefficient between a business's rating and the average rating of its 1, 3, 6, and 10 nearest neighbors, at different distance thresholds from 20 to 2000 meters. For a given distance threshold, if a business has no neighbors within the distance, then this business is excluded from the computation. If a business has fewer than the number of nearest neighbors specified (*e.g.,* 10NN) within the distance threshold, then the ratings of all its nearest neighbors within the distance threshold are averaged. For reference, for each business participated in the computation within a distance threshold, we also randomly sample a business from the dataset (regardless of the distance between the two) and then compute the Pearson's correlation between the two sets of ratings, labeled "Random" in the plot.

From Figure 2, we observe that the rating of a business is weakly positively correlated with the average rating of its nearest neighbor(s). Pearson's correlation coefficient is in the range of 0.109 to 0.173. The correlation is relatively stronger within a smaller distance (*e.g.,* 20 or 50 meters) and becomes stable for distance threshold of 500 meters or larger. In comparison, correlation coefficient between the ratings of the same set of businesses and their randomly selected counterparts is in the range of -0.002 to 0.002, *i.e.,* no correlation, as expected.

We have two interpretations of the weak positive correlation between ratings of businesses and their neighbors. First, the rating of a business should *mainly* depend on the intrinsic characteristics

---

[5]Human preferred walking speed is about 1.4 m/s or 5km/h. http://en.wikipedia.org/wiki/Preferred_walking_speed

Figure 3: Business category distribution

of the business (*e.g.,* quality of products/services), not its neighbors. Second, a business is not geographically independent from its neighbors. When a user visits one business, she has at least glanced over its neighbors if not visited them. These neighbors give her the sense of the surrounding environment of the business (*e.g.,* hygiene standard). Such extrinsic characteristics may affect her view of the business, leading to the weak positive correlation. Based on the weak positive correlation, we expect more accurate business rating prediction can be achieved by considering the influence of the geographical neighbors of a business.

OBSERVATION 3. *The weak positive correlation in ratings is **independent of the categories** of the businesses and/or their neighbors.*
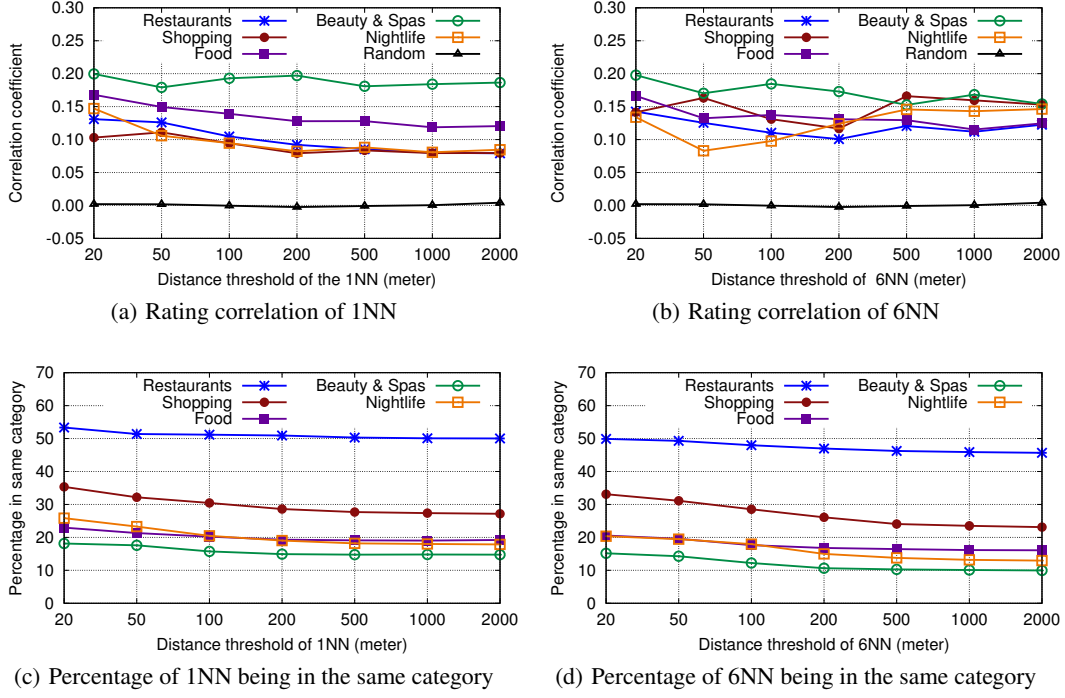
A business in Yelp is assigned one or more category labels (*e.g.,* restaurant, shopping). Plotted in Figure 3, the number of businesses in categories demonstrate a power-law like distribution, with a few categories each containing a large number of businesses and many small categories each has only one or two businesses. Restaurants, the largest category, covers nearly 40% of the 11,537 businesses in the dataset. Next, we study the 5 largest categories as representative examples. They are: Restaurants (4,503), Shopping (1,681), Food (1,616), Beauty & Spas (764), and Nightlife (640), where the numbers in parentheses indicate the number of businesses in each category.

For each of the 5 categories, we calculate the rating correlation between the businesses in the category and their neighbors, regardless of the categories of their neighbors because the neighbors here are defined geographically. The category-wise rating correlations with 1NN and 6NN are plotted in Figures 4(a) and 4(b) respectively. The lines labeled "random", plotted for reference purpose, report the correlation between the ratings of businesses in any of the 5 categories and randomly selected businesses from the dataset (regardless of their categories and geographical distance). Shown in Figure 4(a), businesses in Beauty & Spas demonstrate slightly stronger rating correlation with their 1NN compared with businesses in Nightlife and Shopping. When the number of nearest neighbors is enlarged to 6, this pattern becomes less apparent, particularly when the distance threshold reaches 500 meters, plotted in Figure 4(b). In both Figures 4(a) and 4(b), weak positive correlation is observed for ratings of businesses in all 5 categories across all distance thresholds.[6]

Figures 4(c) and 4(d) show the percentage of the nearest neighbors that are also in the same category. The two figures show very similar patterns: (i) Restaurants are more likely collocated with restaurants with about 50% chance that the neighbors of a restaurant are also restaurants; (ii) Beauty & Spas on the other

---

[6]Similar observation also holds for 3NN and 10NN. The figures are not plotted due to page limit.

(a) Rating correlation of 1NN

(b) Rating correlation of 6NN

(c) Percentage of 1NN being in the same category

(d) Percentage of 6NN being in the same category

Figure 4: Category-wise rating correlation with nearest neighbors (1NN, 6NN), and the percentage of neighbors in the same category

hand is much more distinctive from their neighbors. Below 20% of their neighbors are also in the same category. With respect to Figures 4(a) and 4(b), we argue that weak positive rating correlation between businesses and their neighbors is independent of whether the neighbors are in the same category.

In summary, the three observations made from the data suggest that geographical neighborhood has influences on the rating of a business. Next, we incorporate geographical neighborhood influence into business rating prediction.

4. BUSINESS RATING PREDICTION

The problem of rating prediction has been well formulated in literature. We use r_{ui} to denote the review rating that user u gives to item i (*i.e.*, a business). r_{ui} is in the range of 1 to 5 stars with more stars indicating higher preference. Given the existing ratings made by m users to n items, the task is to predict the unknown rating $\hat{r}_{ui}$, if user u has not rated item i before. In the following, we first briefly introduce matrix factorization and then present our proposed model by incorporating various influences into the prediction. Table 2 lists the notations used in this paper.

4.1 Matrix Factorization

Our proposed method is based on the latent factor model realized by matrix factorization. Through matrix factorization, each user and each item is associated with a f-dimensional vector, where $f \ll \min(m, n)$. The inner product of a user vector $\mathbf{p}_u \in \mathbb{R}^{f \times 1}$ and an item vector $\mathbf{q}_i \in \mathbb{R}^{f \times 1}$ is used to approximate the user's preference to the item (see [13] for a detailed introduction of matrix factorization). Accordingly, the predicted rating of user u to item i is computed using

$$\hat{r}_{ui} = \mathbf{p}_u^\top \mathbf{q}_i,$$

where $\mathbf{p}_u$ and $\mathbf{q}_i$ can be learned from the user-item rating matrix with known ratings. However, users may have certain degree of bi-

Table 2: Notations and semantics

N_i	Set of geographical neighbors of item i
C_i	Set of categories item i belongs to
R_i	Set of words in item i's review
$\mathcal{K}$	Set of (u, i) pairs with known r_{ui} ratings
$r_{ui}, \hat{r}_{ui}$	Observed and predicted ratings of user u to item i
μ	Mean of all known r_{ui} ratings
b_u, b_i	Bias parameters for user u and item i, respectively
$\mathbf{p}_u$	Latent factors of user u
$\mathbf{q}_i$	Latent factors of item i for its intrinsic characteristics
$\mathbf{v}_i$	Latent factors of item i for its extrinsic characteristics
$\mathbf{d}_c$	Latent factors of category c
$\mathbf{q}_w$	Latent factors of review word w
ρ_i	Normalized popularity of item i
β_i	Popularity weighting parameter for item i
$\tau_{u,i}$	Normalized geo-distance between user u and item i
β_u	Geo-distance weighting parameter for user u

ases: some users are more lenient and some are very strict about ratings. Similarly, items may also have some degree of biases because of location or branding for example. To achieve more accurate rating prediction, Biased MF extends the basic matrix factorization by considering the biases,

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \mathbf{q}_i, \qquad (1)$$

where μ is the average rating of all known ratings; b_u and b_i are the user bias and item bias, respectively. Learning the unknown parameters $\mathbf{p}_u$, $\mathbf{q}_i$, b_u, and b_i is an optimization problem to minimize the regularized squared error on the set of known ratings $\mathcal{K}$.

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 \left(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2 \right) + \lambda_2 \left(b_u^2 + b_i^2 \right)$$

In this equation, λ_1 and λ_2 are regularization parameters used to avoid overfitting. Both stochastic gradient descent (SGD) and alternating least squares (ALS) algorithms can be used to solve the optimization function and learn the parameters [12, 13]. In this paper, we adopt SGD to learn the parameters following the algorithm presented in [12].

4.2 Incorporating Neighborhood Influence

Based on our observations in Section 3.2, most businesses have neighbors within a short geographical distance, and more importantly, the rating of a business is weakly positively correlated with the rating of its neighbors. These observations suggest that considering the geographical neighborhood influence may improve the accuracy of business rating prediction.

In this paper, to model users' rating behavior on businesses, we first assume that a user's rating to a given business is determined by its *intrinsic characteristics* and the *extrinsic characteristics* of its geographical neighbors. For a business i, we use $\mathbf{q}_i$ and $\mathbf{v}_i$ to model its intrinsic and extrinsic characteristics, respectively. More specifically, $\mathbf{q}_i$ models the intrinsic characteristics of a business (*e.g.,* taste of food and quality of service) observable by users who have interacted with the business. $\mathbf{v}_i$ models the extrinsic characteristics of a business (*e.g.,* hygiene standard) in influencing its geographical neighbors observable by the "pass-by" visitors.

Let N_i be the set of geographical neighbors of a business i, satisfying certain selection criteria (*e.g.,* the top-6 nearest neighbors). Let $n \in N_i$ be a neighbor of business i. Incorporated with influence from geographical neighbors, the predicted rating is now computed with both $\mathbf{q}_i$ and $\mathbf{v}_n$'s, shown in Equation 2.

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \left(\mathbf{q}_i + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n \right) \qquad (2)$$

In above equation, parameter $\alpha_1 \in [0, 1]$ controls the importance of geographical neighborhood influence in business rating prediction, $| \bullet |$ denotes the cardinality of a set. Accordingly, the objective function is updated with regularization components for $\mathbf{v}_n$, shown in Equation 6 in Table 3, where λ_3 is the newly introduced regularization parameter, similar to λ_1 and λ_2.

Note that, by considering $\mathbf{v}_n$'s in the prediction, Equation 2 partially addresses the cold-start problem for newly established businesses that do not appear in training data. Although $\mathbf{q}_i$ of a new business is empty, the $\mathbf{v}_n$'s of its geographical neighbors are not empty if they appear in training data. Therefore Equation 2 can be applied to make rating prediction of an existing user to a newly established business by using $\mathbf{p}_u$, b_u and $\mathbf{v}_n$'s.

4.3 Incorporating Category Influence

Analyzed in Section 3.2, a business in Yelp may belong to one or more categories. The category of a business usually reflects the characteristics of a business, *e.g.,* product/service offered by the business or the way the business is conducted. Intuitively, users may use different criteria to evaluate businesses in different categories. For example, the criteria commonly used for reviewing restaurant (*e.g.,* taste of food) cannot be used to review businesses in beauty & spas category. Moreover, a recent study has also shown that POI recommendation achieves better accuracy by considering the categories of the POIs [16].

In our model, we introduce category latent factors to exploit business categories for more accurate business rating prediction. For a business category c, it is associated with a latent vector $\mathbf{d}_c \in \mathbb{R}^{f \times 1}$. Let C_i be the set of categories a business i belongs to. By incorporating the category influence, the predicted rating $\hat{r}_{ui}$ is now de-

fined in Equation 3, where $\alpha_2 \in [0, 1]$ is a parameter that controls the importance of category influence in rating prediction. The objective function is updated accordingly, see Equation 7 in Table 3.

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \left(\mathbf{q}_i + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) \quad (3)$$

4.4 Incorporating Review Content

In Yelp, when giving a rating to a business from 1 to 5 stars, the user usually writes a review. Typically, the review elaborates the reason behind the rating and partially reflects the characteristics of the business. Collectively, words in all reviews to a business provide a much better description about the business than the learned vector of "latent" factors. However, in order to make use of the review words in the prediction model, the review words have to be mapped to the same f-dimensional vector space. Similar to the topic level decomposition of tweets for tweet recommendation in [2], we decompose the latent factors of a business $\mathbf{q}_i$ into a combination of latent factors of review words. Let R_i be the set of words that appear in business i's review, the decomposition is shown as follows, where $\mathbf{q}_w$ denotes the latent factors of word w.

$$\mathbf{q}_i \Rightarrow \frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w \qquad (4)$$

With the decomposition, the predicted rating is now shown in Equation 5. The objective function is shown in Equation 8 in Table 3.

$$\begin{aligned} \hat{r}_{ui} = {} & \mu + b_u + b_i \\ & + \mathbf{p}_u^\top \left(\frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) \end{aligned} \quad (5)$$

4.5 Popularity and Geo-distance Influences

Next, we discuss two features that have been studied in POI recommendation, namely *popularity* and *geographical distance*. In Section 1, we argue that the key difference between a business with other kind of items that have been studied in rating prediction is that a business physically exists at a geographical location. Although we argue that business rating prediction is different from POI recommendation, it remains interesting to investigate whether the features explored in POI recommendation are useful here.

Regional popularity is mentioned as one of the unique characteristics in LBSNs which distinguish POI recommendation from other recommendation tasks [15]. For example, businesses located in downtown area are likely to receive more visits than those in suburban. In the Yelp dataset, a business has check-in sets and reviews, both are indicators of business popularity. Note that, a user may write one review to a business but may check-in at the business multiple times. However, the Yelp dataset only provides the aggregated check-in numbers on hourly basis from Monday to Sunday but not user-specific check-ins. Some of the businesses in the dataset do not have check-in sets. We therefore simply sum up the number of reviews and the number of check-ins of a business to be its popularity.

In POI recommendation, geographical distance is a major consideration because users tend to visit nearby POIs [25, 26]. With users' review data, we estimate a user's 'home location' by using the recursive grid search algorithm proposed in [5]. Then the geographical distance between a user and an unrated business can be easily calculated based on the estimated user location and the location of the business.

Table 3: Objective functions for incorporating neighborhood influence, category influence, review words and other factors

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 \left(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2 \right) + \lambda_2 \left(b_u^2 + b_i^2 \right) + \lambda_3 \sum_{n \in N_i} \|\mathbf{v}_n\|^2 \tag{6}$$

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*, \mathbf{d}_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 \left(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2 \right) + \lambda_2 \left(b_u^2 + b_i^2 \right) + \lambda_3 \left(\sum_{n \in N_i} \|\mathbf{v}_n\|^2 + \sum_{c \in C_i} \|\mathbf{d}_c\|^2 \right) \tag{7}$$

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*, \mathbf{d}_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 \left(\|\mathbf{p}_u\|^2 + \sum_{w \in R_i} \|\mathbf{q}_w\|^2 \right) + \lambda_2 \left(b_u^2 + b_i^2 \right) + \lambda_3 \left(\sum_{n \in N_i} \|\mathbf{v}_n\|^2 + \sum_{c \in C_i} \|\mathbf{d}_c\|^2 \right) \tag{8}$$

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*, \mathbf{d}_*, \beta_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 \left(\|\mathbf{p}_u\|^2 + \sum_{w \in R_i} \|\mathbf{q}_w\|^2 \right) + \lambda_2 \left(b_u^2 + b_i^2 + \beta_i^2 + \beta_u^2 \right) + \lambda_3 \left(\sum_{n \in N_i} \|\mathbf{v}_n\|^2 + \sum_{c \in C_i} \|\mathbf{d}_c\|^2 \right) \tag{9}$$

For simplicity, we model both popularity and geographical distance as a rating bias z.

$$z = \beta_i \rho_i + \beta_u \tau_{u,i} \tag{10}$$

In above equation, ρ_i is the normalized popularity of item i by taking the common logarithm of its raw popularity value; $\tau_{u,i}$ is the normalized distance between user u and business i by taking the common logarithm of the geographical distance. The two parameters β_i and β_u are the weighting parameters for business i and user u respectively; both are learned from the training data. With rating bias z, the predicted rating is shown in Equation 11.

$$\hat{r}_{ui} = \mu + b_u + b_i + z$$
$$+ \mathbf{p}_u^\top \left(\frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) \tag{11}$$

Note that, we do not replace the user bias b_u and item bias b_i by rating bias z. The reason is that the rating bias $\beta_i \rho_i + \beta_u \tau_{u,i}$ captures biases specific to popularity and geographical distance respectively, while b_u and b_i are used to capture bias from all unknown factors. The objective function considering both popularity and geographical distance is shown in Equation 9 in Table 3.

5. EXPERIMENTS

We now conduct experiments on the Yelp dataset to evaluate the proposed models and compare the proposed models with state-of-the-art baselines.

5.1 Experimental Setting

Dataset. We use the Yelp dataset that has been studied in Section 3.1 in our experiments. The preprocessing of the dataset in our experiments includes removal of businesses and users having fewer than 10 reviews, stop words removal and stemming in reviews. After the preprocessing, we have $113,514$ ratings by $3,965$ users to $3,760$ businesses. For each user, we sort her ratings in chronological order. The first 70% of ratings are used for training, and the remaining 30% for testing. As the result, we have $79,309$ ratings to build the matrix factorization model for the prediction of the remaining $34,205$ ratings. The data sparsity is 99.47%.

Evaluation Metric. We adopt two popular evaluation metrics, namely, *Mean Absolute Error* (MAE) and *Root Mean Square Error* (RMSE). The smaller MAE or RMSE value means better rating prediction accuracy. In the following equations, $\mathcal{T}$ is the set of user-item rating pairs (u,i) used in testing.

$$MAE = \frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} |r_{ui} - \hat{r}_{ui}|$$

$$RMSE = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} (r_{ui} - \hat{r}_{ui})^2}$$

Baseline Methods. We compare the proposed models with the following 8 baseline methods. All these baseline methods are implemented by using the MyMediaLite library [9].

1. *Global Mean*: this method predicts an unknown rating to be the average of all known ratings, *i.e.*, $\hat{r}_{ui} = \mu$.

2. *User Mean*: this method utilizes the mean rating of each user to predict the missing values for the corresponding user.

3. *Item Mean*: this method uses the mean rating of each item to predict the missing values for the corresponding item.

4. *User KNN*: this method is the user-based collaborative filtering. The unknown ratings are predicted by considering the ratings given by similar users. User similarity is computed by cosine similarity of ratings and $k = 150$ in our experiments. We have also evaluated Pearson's correlation as user similarity which gives poorer results compared with cosine similarity.

5. *Item KNN*: this method is the item-based collaborative filtering. The unknown ratings are predicted by considering the ratings received by similar items. Item similarity is computed by cosine similarity of ratings and $k = 150$.

6. *Biased MF*: this is the MF model with user and item biases briefly described in Section 4.1. Biased MF is widely used as a baseline in recommender systems.

7. *SVD++*: this model considers implicit feedback from users for rating prediction [12]. SVD++ usually offers a superior accuracy and is considered as a state-of-the-art matrix factorization algorithm for recommendation.

8. *Social MF*: this model considers implicit social information between items and/or users. The implicit social information can be derived from most similar and dissimilar users/items using Pearson's correlations or cosine similarity of ratings. As our model considers influences from nearest geographical neighbors, for a fair comparison, we include the implicit

social information from most similar items in *Social MF*.[7] Following [17], the most similar items are selected by using Pearson's correlation of ratings greater than 0.75.

Proposed Methods. We extended Biased MF to incorporate influences from multiple factors: geographical neighborhood (*N*), business category (*C*), review content (*R*), business popularity (*P*), and geographical distance (*D*). The proposed methods are denoted using the letters in parentheses to indicate the influences considered in each method.

9. *N-MF*: this method incorporates geographical neighborhood influence (Section 4.2, Equation 2).

10. *NC-MF*: this method incorporates both geographical neighborhood and category influences (Section 4.3, Equation 3).

11. *NCR-MF*: this model incorporates neighborhood, category and review content influences (Section 4.4, Equation 5).

12. *NCRP-MF*: this model incorporates neighborhood, category, review content and popularity influences, by setting $z = \beta_i \rho_i$ in Equation 10 (Section 4.5).

13. *NCRPD-MF*: this model incorporates all factors: neighborhood, category, review content, popularity and geographical distance, $z = \beta_i \rho_i + \beta_u \tau_{u,i}$ (Section 4.5).

We also evaluate another two methods: *CRP-MF* and *CRPD-MF*. These two methods do not incorporate geographical neighborhood influence but incorporate influences from other factors (*i.e.*, *C*, *R*, *P*, and *D*) indicated by the method names.

Parameter Setting. We performed 5-fold cross-validation on the training set to empirically set the hyperparameters. The number of latent factors $f = 20$; the relative importance of neighborhood and category influences are set to $\alpha_1 = 0.8$, $\alpha_2 = 0.6$; The regularization parameters: $\lambda_1 = 0.8$, $\lambda_2 = 0.4$, $\lambda_3 = 0.6$. The latent factors are learned by SGD with initial learning rate $\gamma = 0.008$, which decreases by a factor of 0.9 after each iteration (see Appendix A). The same parameters are used in all methods for fair comparison for all our proposed methods and the baseline methods whenever applicable. For example, the number of latent factors is also set to 20 in baseline methods *Biased MF*, *SVD++* and *Social MF*. For geographical neighborhood influence, by default, the proposed methods use the 6 nearest neighbors for each business. For all the methods based on matrix factorization, the reported results are averaged over 5 runs to avoid the impact of initialization in parameter learning.

5.2 Experimental Results

We first compare the proposed methods with baseline methods and then evaluate the two schemes for defining the set of geographical neighbors for a business. Lastly, we evaluate the proposed methods with cold-start setting.

5.2.1 Method Comparison

The prediction errors measured by MAE and RMSE of all methods are reported in Table 4 with best results highlighted in boldface. We make four observations from the results.

First, incorporating geographical neighborhood influence into business rating prediction greatly reduces prediction errors measured

[7] Note that, our model can also be extended to include the other kinds of implicit social information as in *Social MF*. However, the main purpose of our experimental evaluation is to evaluate the influence from geographical neighbors in comparison to the influence from similar items measured by rating-based similarity.

Table 4: MAE and RMSE of all methods, the lower the better.

ID	Method	MAE	RMSE
1	Global Mean (μ)	0.8854	1.0962
2	Item Mean	0.8369	1.0939
3	User Mean	0.8599	1.0838
4	Item KNN	0.8208	1.0574
5	User KNN	0.8110	1.0429
6	Biased MF	0.8237	1.0483
7	SVD++	0.8120	1.0352
8	Social MF	0.8123	1.0303
9	N-MF	0.7952	1.0110
10	NC-MF	0.7929	1.0096
11	NCR-MF	0.7923	1.0078
12	NCRP-MF	**0.7920**	**1.0072**
13	NCRPD-MF	0.7958	1.0132
14	CRP-MF	0.7956	1.0138
15	CRPD-MF	0.8062	1.0191

by both MAE and RMSE. All the proposed methods with geographical neighborhood influence (*i.e.,* methods 9 - 13) outperform all baseline methods (methods 1 - 8). The best prediction accuracy is achieved by *NCRP-MF* which considers geographical neighborhood (N), business category (C), review content (R), and business popularity (P). With geographical neighborhood influence alone, *N-MF* outperforms all baselines including state-of-the-art methods *SVD++* and *Social MF*. This result suggests that geographical neighbors are more effective than the "neighbors" derived from rating-based similarity as in *Social MF*.

Second, compared with geographical neighborhood, further considering factors like business category (*C*), review content (*R*), and business popularity (*P*) leads to relatively small additional reduction in prediction errors. The geographical distance factor (*D*) adversely affects business rating prediction accuracy. Method *NCRPD-MF* performs the worst among all methods with geographical neighborhood influence measured by both MAE and RMSE. This result suggests that the geographical distance factor introduces noise in the prediction model. Such result supports our earlier discussion that geographical distance between users and items is less relevant to the problem of business rating prediction.

Third, without incorporating geographical influence, *CRP-MF* performs poorer than most methods with geographical neighborhood influence including *N-MF*, *NC-MF*, *NCR-MF*, and *NCRP-MF*. The poorer performance of *CRP-MF* against *N-MF* suggests that the geographical neighborhood influence is more effective than the combination of the three factors (*C*, *R*, and *P*) in business rating prediction. The much poorer performance of *CRPD-MF* in comparison with *CRP-MF* again suggests the adverse effect of geographical distance factor (*D*). On the other hand, the effectiveness of geographical neighborhood influence is also reflected from the better performance of *NCRPD-MF* compared with *CRPD-MF*.
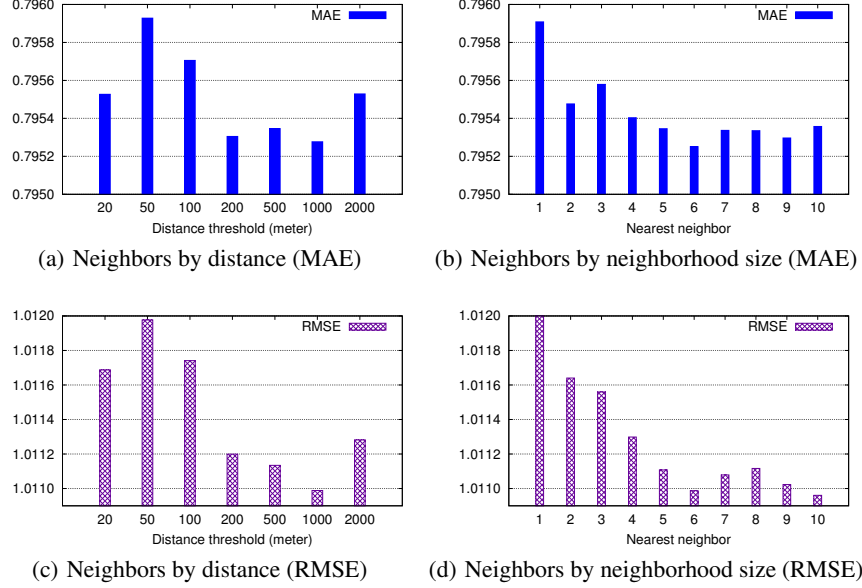
Last, among the 8 baseline methods, *SVD++* and *Social MF*, the two state-of-the-art methods, perform the best evaluated by RMSE. By MAE measure, *User KNN* gives surprisingly low error rate and the two methods *SVD++* and *Social MF* remain among the three best methods.

5.2.2 Impact of Neighborhood Size

Given a business, its geographical neighbors can be defined based on either a geographical distance threshold or the number of nearest neighbors. In this set of experiments, we evaluate the two schemes on their impact to the prediction error of the *N-MF* method. The

(a) Neighbors by distance (MAE)

(b) Neighbors by neighborhood size (MAE)

(c) Neighbors by distance (RMSE)

(d) Neighbors by neighborhood size (RMSE)

Figure 5: Prediction error of *N-MF* with two schemes for determining geographical neighbors of a business

N-MF method is selected as the method for evaluation because it only considers the geographical neighborhood factor.

Figures 5(a) and 5(c) respectively plot MAE and RMSE of the *N-MF* method by defining geographical neighbors with a distance threshold ranging from 20 to 2000 meters. Figures 5(b) and 5(d) respectively plot the MAE and RMSE of the method by taking the number of nearest neighbors ranging from 1 to 10. The *y*-axes of the two sets of figures are plotted in the same scale for easy comparison. From the two sets of figures, either the distance threshold of 1000 meters or the top-6 nearest neighbors give the best prediction accuracy by considering both MAE and RMSE. Plotted in Figure 1, with a distance threshold of 1000 meters, more than 90% of businesses have at least 6 neighbors and about 84% of the businesses have at least 10 neighbors. Because some businesses may have a large number of neighbors within 1000 meters, we choose to use 6 nearest neighbors mainly for minimizing computational cost. We have also evaluated the method by using at most 6 nearest neighbors within a 1000-meter distance threshold. However, poorer prediction accuracy was obtained compared with simply using 6 nearest neighbors without a distance threshold.

5.2.3 Cold-Start Business Rating Prediction

Cold-start is a challenging issue in any recommender systems when there is no enough information about either users or items. We now evaluate the performance of the proposed methods in predicting ratings by *existing users* to *newly established* businesses. Recall that in our data preprocessing (see Section 5.1), businesses and users with fewer than 10 reviews were removed. In this set of experiments, we try to predict the ratings by existing users in the training data to the businesses that were removed in data preprocessing. That is, from existing training data, we have the user vector of latent factors $\mathbf{p}_u$, but not the item vector $\mathbf{q}_i$ because the "newly established" businesses were not the training data. In total, there are $20,395$ ratings made by $3,319$ existing users to $6,939$ new businesses. This collection of ratings will be used as test set.

Among all baseline methods, *Global Mean* and *User Mean* can be easily applied. *Item Mean*, *User KNN* and *Item KNN* cannot be applied as we assume that the newly established businesses have no

Table 5: MAE and RMSE in cold-start setting

Method	MAE	RMSE
Global Mean	1.0319	1.2749
User Mean	0.9963	1.2566
Biased MF	1.0020	1.2539
N-MF	0.9956	1.2538
NC-MF	**0.9936**	**1.2535**

reviews yet. Both *Biased MF* and *SVD++* are reduced to $\mu + b_u$ because b_i and $\mathbf{q}_i$ are both unknown (see Equation 1). *Social MF* is not applicable here because it relies on similar items by some similarity functions. Among the proposed methods, both the latent factors for the extrinsic characteristics of geographical neighbors (*i.e.,* $\mathbf{v}_n$'s of existing businesses) and business category (*i.e.,* $\mathbf{d}_c$) can be utilized in prediction. Here, we assume the category of the newly established business is known (*e.g.,* restaurant or bookstore). The prediction is made by Equation 3 after removing b_i and $\mathbf{q}_i$ from the equation.

Reported in Table 5, the proposed method *NC-MF* achieves the best prediction accuracy by considering both geographical neighborhood and business category. Utilizing geographical influence alone, *N-MF* is the second best performing methods by both MAE and RMSE. Both *User Mean* and *Biased MF* are better than *Global Mean* with the former achieves better MAE and the latter achieves better RMSE. In summary, the proposed methods predict more accurate ratings for existing users to newly established businesses in cold-start setting.

6. CONCLUSION

To the best of our knowledge, this is the first study on geographical neighborhood influence to users' business rating behavior. We believe the observation that a business's rating is weakly positively correlated with its geographical neighbors' rating is an important one. Based on this observation, we model a business with two vectors of latent factors one for its intrinsic characteristics and the other for its extrinsic characteristics (or its influence to its geographical neighbors). In our experiments, we show that by incorporating ge-

graphical neighborhood influence, the proposed methods outperform state-of-the-art methods. Other factors like business category, popularity, and review content can further improve the rating prediction accuracy. Nevertheless, the geographical distance between a user and a business adversely affects the prediction accuracy, although it is an important factor in POI recommendation and POI prediction. The incorporation of geographical neighborhood influence also partially enables our methods to better handle cold-start situation, for rating prediction of newly established businesses based on both their geographical neighbors and business categories. As a part of our future work, we are interested in investigating the influence of geographical neighborhood in POI recommendation and sentiment analysis of business reviews.

7. REFERENCES

[1] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *UAI*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.

[2] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu. Collaborative personalized tweet recommendation. In *SIGIR*, pages 661–670. ACM, 2012.

[3] 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.

[4] C. Cheng, H. Yang, M. R. Lyu, and I. King. Where you like to go next: Successive point-of-interest recommendation. In *IJCAI*, pages 2605–2611. AAAI Press, 2013.

[5] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring millions of footprints in location sharing services. In *ICWSM*, pages 81–88. The AAAI Press, 2011.

[6] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *KDD*, pages 1082–1090. ACM, 2011.

[7] P. Cremonesi, Y. Koren, and R. Turrin. Performance of recommender algorithms on top-n recommendation tasks. In *RecSys*, pages 39–46. ACM, 2010.

[8] M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.*, 22(1):143–177, 2004.

[9] Z. Gantner, S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. MyMediaLite: A free recommender system library. In *RecSys*, pages 305–308. ACM, 2011.

[10] R. Jin, J. Y. Chai, and L. Si. An automatic weighting scheme for collaborative filtering. In *SIGIR*, pages 337–344, 2004.

[11] N. Koenigstein, G. Dror, and Y. Koren. Yahoo! music recommendations: Modeling music ratings with temporal dynamics and item taxonomy. In *RecSys*, pages 165–172. ACM, 2011.

[12] Y. Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *KDD*, pages 426–434. ACM, 2008.

[13] Y. Koren. Collaborative filtering with temporal dynamics. In *KDD*, pages 447–456. ACM, 2009.

[14] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[15] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *KDD*, pages 1043–1051. ACM, 2013.

[16] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *CIKM*, pages 733–738. ACM, 2013.

[17] H. Ma. An experimental study on implicit social recommendation. In *SIGIR*, pages 73–82. ACM, 2013.

[18] S. Moghaddam, M. Jamali, and M. Ester. ETF: Extended tensor factorization model for personalizing prediction of review helpfulness. In *WSDM*, pages 163–172. ACM, 2012.

[19] N. Natarajan, D. Shin, and I. S. Dhillon. Which app will you use next?: Collaborative filtering with interactional context. In *RecSys*, pages 201–208. ACM, 2013.

[20] A. Rae, V. Murdock, A. Popescu, and H. Bouchard. Mining the web for points of interest. In *SIGIR*, pages 711–720, 2012.

[21] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, pages 285–295. ACM, 2001.

[22] H. Steck. Item popularity and recommendation accuracy. In *RecSys*, pages 125–132. ACM, 2011.

[23] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46:234–240, 1970.

[24] D. Yang, T. Chen, W. Zhang, Q. Lu, and Y. Yu. Local implicit feedback mining for music recommendation. In *RecSys*, pages 91–98. ACM, 2012.

[25] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, pages 325–334. ACM, 2011.

[26] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *SIGIR*, pages 363–372. ACM, 2013.

APPENDIX

A. PARAMETER ESTIMATION

All the objective functions (*e.g.,* Equations 6, 7, 8, and 9) in the proposed models share the same form. Next, we detail the parameter estimation for Equation 9 (where $z = \beta_i \rho_i + \beta_u \tau_{u,i}$) as an example using Stochastic Gradient Descent (SGD) algorithm [12]. Let e_{ui} be the error associated with the prediction $e_{ui} = r_{ui} - \hat{r}_{ui}$. The parameters are learned by moving in the opposite direction of the gradient with a learning rate γ in an iterative manner. In our experiments, learning rate γ is set to 0.008 in the first iteration and is decreased by a factor of 0.9 after each subsequent iteration.

$$b_u \leftarrow b_u + \gamma \cdot (e_{ui} - \lambda_2 \cdot b_u)$$

$$b_i \leftarrow b_i + \gamma \cdot (e_{ui} - \lambda_2 \cdot b_i)$$

$$\beta_i \leftarrow \beta_i + \gamma \cdot (e_{ui} \cdot \rho_i - \lambda_2 \cdot \beta_i)$$

$$\beta_u \leftarrow \beta_u + \gamma \cdot (e_{ui} \cdot \tau_{u,i} - \lambda_2 \cdot \beta_u)$$

$$\mathbf{p}_u \leftarrow \mathbf{p}_u + \gamma \cdot \left(e_{ui} \cdot \left(\frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n \right. \right.$$
$$\left. \left. + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) - \lambda_1 \cdot \mathbf{p}_u \right)$$

$$\forall w \in R_i : \quad \mathbf{q}_w \leftarrow \mathbf{q}_w + \gamma \cdot \left(e_{ui} \cdot \frac{1}{|R_i|} \cdot \mathbf{p}_u - \lambda_1 \cdot \mathbf{q}_w \right)$$

$$\forall n \in N_i : \quad \mathbf{v}_n \leftarrow \mathbf{v}_n + \gamma \cdot \left(e_{ui} \cdot \frac{\alpha_1}{|N_i|} \cdot \mathbf{p}_u - \lambda_3 \cdot \mathbf{v}_n \right)$$

$$\forall c \in C_i : \quad \mathbf{d}_c \leftarrow \mathbf{d}_c + \gamma \cdot \left(e_{ui} \cdot \frac{\alpha_2}{|C_i|} \cdot \mathbf{p}_u - \lambda_3 \cdot \mathbf{d}_c \right)$$