CORALS: Who are My Potential New Customers? Tapping into the Wisdom of Customers' Decisions

Ruirui Li
UCLA
rrli@cs.ucla.edu

Jyun-Yu Jiang
UCLA
jyunyu@cs.ucla.edu

Chelsea J.-T. Ju
UCLA
chelseaju@cs.ucla.edu

Wei Wang
UCLA
weiwang@cs.ucla.edu

ABSTRACT

Identifying and recommending potential new customers for local businesses are crucial to the survival and success of local businesses. A key component to identifying the right customers is to understand the decision-making process of choosing a business over the others. However, modeling this process is an extremely challenging task as a decision is influenced by multiple factors. These factors include but are not limited to an individual's taste or preference, the location accessibility of a business, and the reputation of a business from social media. Most of the recommender systems lack the power to integrate multiple factors together and are hardly extensible to accommodate new incoming factors. In this paper, we introduce a unified framework, CORALS, which considers the personal preferences of different customers, the geographical influence, and the reputation of local businesses in the customer recommendation task. To evaluate the proposed model, we conduct a series of experiments to extensively compare with 12 state-of-the-art methods using two real-world datasets. The results demonstrate that CORALS outperforms all these baselines by a significant margin in most scenarios. In addition to identifying potential new customers, we also break down the analysis for different types of businesses to evaluate the impact of various factors that may affect customers' decisions. This information, in return, provides a great resource for local businesses to adjust their advertising strategies and business services to attract more prospective customers.

CCS CONCEPTS

• **Information systems** **Location based services**; **Geographic information systems**.

KEYWORDS

Customer prediction; geographical preference; reputation reliance; pairwise ranking

ACM Reference Format:

Ruirui Li, Jyun-Yu Jiang, Chelsea J.-T. Ju, and Wei Wang. 2019. CORALS: Who are My Potential New Customers? Tapping into the Wisdom of Customers' Decisions. In *Proceedings of The Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3289600.3290995

WSDM '19, February 11–15, 2019, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-5940-5/19/02…$15.00
https://doi.org/10.1145/3289600.3290995

1 INTRODUCTION

Recommender system has attracted substantial attention from researchers since the last decade and has revolutionized the e-commerce industry. Various recommender systems have been developed to facilitate the matching between customers with appropriate products or services, such as movies on Netflix, music on Last.fm, and merchandises on Amazon. For customers, recommendations improve user experience by providing helpful suggestions to explore and discover relevant products or services. For providers, these recommendations increase the propensity of purchases from customers.

Over the past few years, the prevalence of GPS-enabled devices, such as smart phones, establish the prosperity of location-based social networks (LBSN), such as Foursquare, Yelp, and Facebook Local. LBSN attracts millions of users to share their social friendship and their locations via check-ins. For example, an average of 142 million users check in at local businesses via Yelp every month [42]. Foursquare has 55 million monthly active users and 8 million daily check-ins on the Swarm application [29]. Facebook Local, powered by 70 million businesses [6], facilitates the discovery of local events and places for over one billion active daily users [25]. The check-ins, which contain abundant hints of user preferences on locations, allow us to identify potential new customers for local businesses.

To identify potential new customers, the most crucial thing is to understand a customer's decision-making process. However, it is a complex process, and can be influenced by multiple factors. Most investigated factors are personal preferences and geographical convenience. Personal preferences are learned from customers' historical check-ins by applying collaborative filtering or matrix factorization techniques. The learned preferences, in return, help us find out new businesses which customers are interested in. In addition, check-in locations provide an ancillary resource to interpret customers' decisions from the perspective of geographical convenience. According to the Tobler's first law of geography and the law of demand, the propensity of a customer for a local business is inversely proportional to the distance between the customer and the business, which is similar to the probability of purchasing an item being inversely proportional to the cost.

There are also studies which show that customers prefer learning from local experts who know the neighborhood well and have firsthand experience [1, 36]. This is because that online reviews are becoming more and more influential in establishing and promoting the reputation of local businesses than ever before. The emergence of numerous review sites has created an unprecedented and ongoing online conversation about local businesses. Therefore, a business'

Table 1: Statistics and densities of six datasets.

Dataset	#(Users)	#(Items/businesses)	#(Ratings/check-ins)	Density
Netflix	48,189	17,770	100,480,507	1.17×10^{-1}
MovieLens 20M	138,000	27,000	20,000,000	5.37×10^{-3}
last.fm	1,892	17,632	92,834	2.78×10^{-3}
Yahoo! Music	1,000,990	624,961	262,810,175	4.20×10^{-4}
Yelp	1,029,432	144,072	5,099,750	6.94×10^{-6}
Foursquare	1,219,322	422,030	693,798	1.35×10^{-6}

reputation is more public and more accessible. Customers are able to see over the "wall" of corporate messaging at what lies behind. They can get a sense of a business' true essence through the shared experiences of other customers. These changes in marketing lead to a change in customers' habits. Customers are becoming more and more review-dependent. This is consistent with the study conducted by BrightLocal [3]. Compared with the trend in 2010, the number of people who search for local businesses before consumption is doubled in 2015 and 2016. Moreover, among all the participants in the study, 92% of the customers regularly or occasionally read online reviews, which help them judge whether a local business provides good services or not. Therefore, the impact of online reviews is non-negligible and growing.

Although identifying potential new customers is crucial for local businesses, it is still a very challenging task due to the following three reasons.

- **Data Sparsity**. To know and comment on a local business, a customer has to physically visit that business. Thus, the cost is higher than that of rating a movie or a song online. Even if a customer makes the effort to visit the business, he/she often does not check in due to privacy or safety concerns [38], let alone writing a review. Therefore, customers' check-in data is much sparser than other rating data for movies and music. Table 1 shows the statistics and the densities of four well-known movie and music rating datasets, together with two LBSN datasets, i.e., the Yelp challenge dataset and the Foursquare dataset. Here the density of a dataset is calculated by the number of ratings/check-ins divided by the product of the number of users and the number of items/businesses. The densities of Yelp and Foursquare datasets are much lower than the ones of Netflix, MovieLens, Last.fm, and Yahoo! music datasets. The extremely sparse check-in data makes it challenging for us to accurately model customers' preferences.
- **Geographical Influence**. The first law of geography states that everything is related to everything else, but near things are more related than distant things [30]. Many studies show that people tend to visit nearby local businesses or explore businesses near the ones that they have visited before [41]. Therefore, a challenge is how to estimate customers' activity trajectories or zones based on the sparse check-in data. Beyond this estimation, a more challenging fact is that the geographical influence is both customer-dependent and business-dependent. If a customer owns a car, he can visit a faraway business with less effort than ones who cannot drive and rely on public transits. On the other hand, the geographical factor has different impacts on different types of businesses. For example, customers tend to visit nearby fast-food businesses for convenience. However, they may be willing to travel farther to visit other types of businesses, such as museums, where they are more closely connected with cultures and get inspirations, or salons where they can have their hair cut and styled by professionals.
- **Reputation Influence**. Nowadays, more and more customers rely on online reviews to get a sense of the reputation of local businesses. These reviews implicitly influence customers' decisions towards visiting a business. The influence of reviews is also both customer-dependent and business-dependent. Different customers have different opinions on the same review. Moreover, similar reviews may have different impacts on different types of businesses. For example, a review such as "A bit of long wait" to a fast-food business may give other customers a very negative impression. However, the impact may be milder if the same comment is made on theme parks, such as Universal Studios.

In addition to the three challenges above, given the heterogeneous information on check-in, location, online reviews, current works also lack an integrated analysis of personal preferences, geographical influence, and business reputation when modeling customers' decisions. To the best of our knowledge, this work is the first one considering all these factors under the scenario of recommending customers for local businesses. To be more specific, the main contributions of this work are as follows:

- We propose a customer recommendation model, CORALS, which, based on historical check-in information, integrates customers' personal preferences, geographical influence, and business reputation. In addition, the model is also capable of incorporating other factors such as expenses. Moreover, the model offers high interpretability by providing the quantitative importance of incorporated factors for different types of local businesses.
- We present a comprehensive empirical evaluation of our approach against 12 recommendation methods on two real-world datasets. The results show that our approach, CORALS, outperforms all baseline methods in suggesting potential new customers for local businesses in different cities.

2 METHODOLOGY

Table 2: List of symbols

Symbol	Description
$t_{b,i}$	personal preference of customer i on business b
$g_{b,i}$	geographical convenience of business b for customer i
$r_{b,i}$	reputation reliance of customer i on business b
w_b^g	the geographical influence weight on business b
w_b^r	the reputation influence weight on business b
$\mathbf{p}_b$	latent feature vector for business b
$\mathbf{q}_i$	personal preference feature vector for customer i
$\mathbf{u}_b$	business reputation vector for business b
$\mathbf{d}_i$	reputation reliance vector for customer i
η	learning rate
λ	regularization parameters
Θ	recommendation parameters
$\nabla\theta$	gradient of parameter θ
Φ	Gaussian mixture model parameters
$\mathbf{l}$	business location defined by latitude and longitude pair
M	number of Gaussian components

Table 2 lists the notations we use in this paper. We use bold letters for vectors and normal letters for scalars.

As we mentioned in the introduction, the key to addressing the recommendation problem is to accurately understand customers' decision-making processes. In this work, we decompose it into three main factors: a customer i's personal preference $t_{b,i}$ over a business b, the geographical convenience $g_{b,i}$ of business b for customer i, and customer i's reliance $r_{b,i}$ on business b's reputation. In addition, the contributions of $g_{b,i}$ and $r_{b,i}$ are given by w_b^g and w_b^r, respectively. Formally, the tendency of a customer i's visiting a business b is given by:

$$t_{b,i} + w_b^g g_{b,i} + w_b^r r_{b,i}. \tag{1}$$

Higher tendency indicates higher check-in likelihood.

Given an observed check-in from customer i on business b, denoted by (b, i), applying the idea of pair-wise comparisons, we sample another customer j who has not checked in at business b. Now, for a business b, we have an observed check-in (b, i) and a sampled unobserved check-in (b, j). It is logical to hypothesize that compared with customer j, customer i is more likely to visit business b. We construct the model by maximizing a posteriori over all observed and sampled check-ins:

$$C = \prod_{(b,i),j} p(i >_b j|\Theta)p(\Theta), \tag{2}$$

where Θ is a set of parameters, which define the model. $p(i >_b j|\Theta)$ gives the probability that a customer i prefers a business b more than another customer j does under the model. Formally,

$$p(i >_b j|\Theta) = \delta[(t_{b,i} - t_{b,j}) + w_b^g(g_{b,i} - g_{b,j}) + w_b^r(r_{b,i} - r_{b,j})], \tag{3}$$

where $\delta(x)$ is the sigmoid function:

$$\delta(x) = \frac{1}{1 + e^{-x}}. \tag{4}$$

In addition, the personal preference of customer i on business b is given by:

$$t_{b,i} = \mathbf{p}_b \cdot \mathbf{q}_i, \tag{5}$$

where $\mathbf{p}_b$ and $\mathbf{q}_i$ are business and customer vector representations in the preference hidden space, respectively. Similarly, the reputation reliance of a customer i on a business b is given by:

$$r_{b,i} = \mathbf{u}_b \cdot \mathbf{d}_i, \tag{6}$$

where $\mathbf{u}_b$ and $\mathbf{d}_i$ are business and customer vector representations in the reputation hidden space, respectively. Using Gaussian priors $\Theta \sim N(0, \lambda_\theta I)$ to model the parameters, we have

$$p(\Theta) = \frac{1}{\sqrt{2\pi}\sigma} e(-\frac{\|\Theta\|^2}{2\sigma^2}). \tag{7}$$

Substituting Equations 3, 4, 5, 6, 7 into the objective Equation 2, we can derive maximizing a posteriori as follows:

$$\begin{aligned} C &\propto \ln \prod_{(b,i),j} p(i >_b j|\Theta)p(\Theta) \\ &= \sum_{(b,i),j} \ln \delta[(t_{b,i} - t_{b,j}) + w_b^g(g_{b,i} - g_{b,j}) + w_b^r(r_{b,i} - r_{b,j})] + \ln p(\Theta) \\ &= \sum_{(b,i),j} \ln \delta[(t_{b,i} - t_{b,j}) + w_b^g(g_{b,i} - g_{b,j}) + w_b^r(r_{b,i} - r_{b,j})] - \lambda\|\Theta\|^2 \\ &= \sum_{(b,i),j} \{\ln \delta[(\mathbf{p}_b \cdot \mathbf{q}_i - \mathbf{p}_b \cdot \mathbf{q}_j) + w_b^g(g_{b,i} - g_{b,j}) + w_b^r(\mathbf{u}_b \cdot \mathbf{d}_i - \\ &\quad \mathbf{u}_b \cdot \mathbf{d}_j)] - \lambda\|\Theta\|^2\}, \end{aligned}$$

where λ is a set of regularization parameters for Θ. Here, the businesses' geographical convenience, $g_{b,i}$, and businesses' reputations, u_b, are inputs. These two factors will be discussed in Sections 2.1 and 2.2, respectively.

Algorithm 1: Parameter optimization with AdaGrad

Input: learning rate η, max iteration $iter_{max}$, regularization weights λ, max number of samples S_{max};
Output: Θ
1 **Initialization:** initialize Θ with Normal distribution $N(0,0.01)$, $iter = 0$, $\Theta_{opt} = \Theta$, $\mathrm{err}_{opt} = err_{vali}$;
2 **repeat**
3   **foreach** *observed check-in* (b, i) **do**
4   Counter $cnt = S_{max}$;
5   **while** $cnt > 0$ **do**
6   Randomly generate an unobserved customer j;
7   **if** $(t_{b,i} - t_{b,j}) + w_b^g(g_{b,i} - g_{b,j}) + w_b^r(r_{b,i} - r_{b,j}) > 0$ **then**
8   cnt- -;
9   **else**
10   **foreach** *involved* θ **do**
11           $\nabla\theta^{ts} = \frac{\partial J}{\partial \theta^{ts}}$;
12           $n_\theta^{ts+1} = n_\theta^{ts} + (\nabla\theta^{ts})^2$;
13           $\theta^{ts+1} = \theta^{ts} - \frac{\eta}{\sqrt{n_\theta^{ts} + \epsilon}} \nabla\theta^{ts}$;
14   break;
15   **if** $err_{vali} < err_{opt}$ **then**
16     $err_{opt} = err_{vali}$;
17     $\Theta_{opt} = \Theta$;
18   **else**
19     $\Theta = \Theta_{opt}$;
20   $iter$ + +;
21 **until** $iter > iter_{max}$;
22 **Return** Θ_{opt};

To optimize top ranked customers in the recommendation list, we apply the weighted approximate ranking strategy proposed in [33] to optimize precision@k. Algorithm 1 summarizes the optimization process. First, the parameters Θ are initialized using Normal distributions. The optimization process is iterative. In each iteration, it goes through each observed check-in in the training set. For each observed check-in (b, i), we sample a random customer j who has not visited business b. If the preference order between i and j on business b is correctly predicted using the current Θ, we randomly sample another customer to find a violation. This process repeats at most S_{max} times until we find such a violation. Once we find a violation, we update the corresponding parameter θ, $\theta \in \Theta$. After iterating through each check-in in the training set, we evaluate the performance using the validation set. If the performance increases, we accept the updates on Θ. Otherwise, we reject the updates. This step helps us avoid adopting over-fitting parameters on the training data. The optimization terminates when $iter$ reaches the maximum number of iterations.

As we mentioned in the introduction, since many businesses and customers have limited numbers of check-ins, the check-in data is extremely sparse. However, the parameters of these businesses and customers can be immensely useful and informative to the problem we want to optimize. To effectively leverage the sparse data, AdaGrad [8] is proposed to give a higher learning rate to the parameters that are more sparse in the data. We adopt this concept to adjust the learning rate adaptively for each individual parameter θ, which is shown in lines 10-13 of Algorithm 1. AdaGrad modifies the general learning rate η at each time step ts for every parameter θ based on the past gradients that have been computed for θ. n_θ^{ts+1} records the sum of the squares of the gradients with respect to θ up to the ts^{th} time step[1]. ϵ is a smoothing term that avoids division by zero. In this way, AdaGrad makes it such that parameters that are more sparse in the data have a higher learning rate which translates into a larger update for that parameter.

2.1 Geographical Convenience Inference

In this section, we discuss how to infer the geographical convenience of a business b for a user i, i.e., $g_{b,i}$, based on customer i's historical check-ins.

We apply a Gaussian mixture model (GMM) [27] to make the inference. A Gaussian mixture model is a weighted sum of M component Gaussian densities:

$$p(\mathbf{l}|\Phi) = \sum_{m=1}^{M} \alpha_m g(\mathbf{l}|\mu_m, \Sigma_m), \tag{8}$$

where $\mathbf{l}$ is a 2-dimensional location vector (i.e. latitude and longitude), α_m, $m = 1, ..., M$, are the mixture weights, and $g(\mathbf{l}|\mu_m, \Sigma_m)$ are the component Gaussian densities. Each component density is a 2-variate Gaussian function of the form,

$$g(\mathbf{l}|\mu_m, \Sigma_m) = \frac{1}{2\pi|\Sigma_m|^{1/2}} e^{-\frac{1}{2}(\mathbf{l}-\mu_m)'\Sigma_m^{-1}(\mathbf{l}-\mu_m)},$$

with mean location vector μ_m and covariance matrix Σ_m. The complete Gaussian mixture model is parameterized by the mean location vectors, covariance matrices and mixture weights from all component densities. These parameters are further collectively notated by Φ. For a particular customer, given a sequence of his N check-in locations, represented by N location vectors $L = \{\mathbf{l}_1, .., \mathbf{l}_N\}$, the GMM likelihood, assuming conditional independence between the location vectors, can be written as:

$$p(L|\Phi) = \prod_{n=1}^{N} p(\mathbf{l}_n|\Phi).$$

We use the Expectation-Maximization (EM) [7] algorithm to estimate the parameters. The EM algorithm begins with an initial model Φ, to estimate a new model $\bar{\Phi}$, such that $p(L|\bar{\Phi}) \geq p(L|\Phi)$. The new model then becomes the initial model for the next iteration and the process is repeated until convergence. In each EM iteration, re-estimation Equations 9, 10, and 11 are used to guarantee a monotonic increase in the model's likelihood value in the E-step.

$$\text{Mixture weights: } \bar{\alpha}_m = \frac{1}{N}\sum_{n=1}^{N} p(m|\mathbf{l}_n, \Phi), \tag{9}$$

$$\text{Location means: } \bar{\mu}_m = \frac{\sum_{n=1}^{N} p(m|\mathbf{l}_n, \Phi)\cdot \mathbf{l}_n}{\sum_{n=1}^{N} p(m|\mathbf{l}_n, \Phi)}, \tag{10}$$

$$\text{Variances: } \bar{\sigma}_m^2 = \frac{\sum_{n=1}^{N} p(m|\mathbf{l}_n, \Phi)\cdot \mathbf{l}_n^2}{\sum_{n=1}^{N} p(m|\mathbf{l}_n, \Phi)} - \bar{\mu}_m^2, \tag{11}$$

In the M-step, the posteriori probability for component m is given by

$$p(m|\mathbf{l}_n, \Phi) = \frac{\alpha_m g(\mathbf{l}_n|\mu_m, \Sigma_m)}{\sum_{m=1}^{M} \alpha_m g(\mathbf{l}_n|\mu_m, \Sigma_m)}.$$

[1]In one iteration, the same parameter may be optimized multiple times. Each optimization counts 1 time step.

To determine the number of Gaussian components M, we apply affinity propagation [9] to cluster each customer's check-ins. The number of clusters yields the number of Gaussian components.

After the GMM construction for a customer i, given the geographical location l_b of a business b, as shown in Equation 8, $p(\mathbf{l_b}|\Phi)$ gives the geographical convenience $g_{b,i}$ of the business b for each customer i.

2.2 Business Reputation Inference from Reviews

In this section, we discuss how to model the business reputation, $\mathbf{u}_b$, based on the reviews commented on the local businesses.

There are two main challenges. First, reviews differ in their lengths. Some reviews are informative and have more words while others are not. This challenge makes it difficult to model the business's reputation $\mathbf{u}_b$ in a fixed-length vector. Second, for a particular business, some reviews are older while others are more recent. They may have different influences on the reputation of the business.

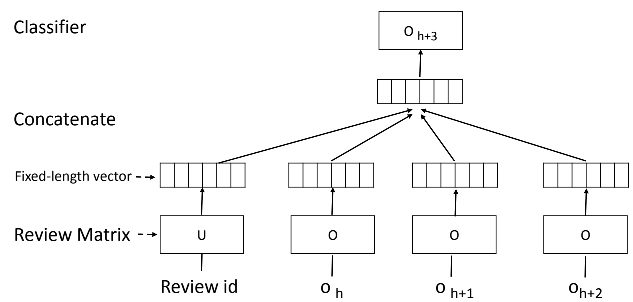


Figure 1: The framework for learning review vector

To solve the first challenge, we apply a distributed memory model proposed in [16]. Figure 1 shows the framework for the vector learning task, which is to predict a word given other words in a context. Formally, given a sequence of training words o_1, o_2, o_3, ..., o_H, the objective of the model is to maximize the average log probability

$$\frac{1}{H}\sum_{h=k}^{H-k} logp(o_h|o_{h-k}, ..., o_{h+k}).$$

The prediction task is performed via a multiclass classifier, i.e., softmax. Then, we have:

$$p(o_h|o_{h-k}, ..., o_{h+k}) = \frac{e^{y_{o_h}}}{\sum_o e^{y_o}}.$$

Each y_{o_h} is the un-normalized log probability for each output word o_h, calculated as:

$$y_{o_h} = V_0 + Vz(o_{h-k}, ..., o_{h+k}, U),$$

where V_0 and V are the softmax parameters. z is constructed by a concatenation of a review vector and word vectors from O. Both review and word vectors are trained using stochastic gradient descent (SGD) and the gradient is obtained via back propagation. At each step of SGD, we sample a fixed-length context from a random review, compute the error gradient and update the parameters in the model. Once the parameters get converged, we obtain the dense representation of each review. In order to address the impact of the chronological order of the reviews, we use the vector of the most recent review as the reputation vector of the business, $\mathbf{u}_b$.

3 EXPERIMENTS

In this section, we conduct extensive experiments on two real-world datasets to evaluate the performance of CORALS.

3.1 Datasets and Experimental Settings

The experiments are conducted on two datasets. One is the most recently released dataset from the Yelp challenge. The other is the Foursquare dataset. The Yelp dataset contains interactions between customers and businesses, with 4.1M reviews and 947K tips by 1M users for 144K businesses. We investigate the recommendation tasks in 7 large cities. The Foursquare dataset contains interactions between customers and businesses in Los Angeles and New York. Table 3 shows the statistics for the 9 cities in the two datasets.

Unfortunately, some businesses do not accumulate an adequate amount of check-ins. Moreover, some customers lack sufficient check-ins to infer their preferences. We follow the data cleaning strategy in [13] and filter out businesses and customers whose check-ins are less than 20 for the Yelp dataset. For the Foursquare dataset, we follow the cleaning steps in [2] and remove business and customers that have less than 8 check-ins. For each business, its check-ins are sorted in chronological order based on the timestamps. The first 50% of the check-ins are used as the training data. The following 20% are used for validation and the remaining 30% are used as the test data for evaluation. Table 4 shows the parameter settings of CORALS in the experiments. These parameters are tuned by grid search.

3.2 Baselines

To compare our approach with others, the following 12 methods are adopted as baselines.

- **Weighted Regularized MF (WRMF)**. WRMF [15] minimizes the square error loss by assigning both observed and unobserved check-ins with different weights based on matrix factorization.
- **Maximum Margin MF (MMMF)**. MMMF [32] minimizes the hinge loss based on matrix factorization.
- **Bayesian Personalized Ranking MF (BPRMF)**. BPRMF [26] optimizes Area Under the Curve (AUC) based on pairs of observed check-ins and sampled unobserved check-ins.
- **CofiRank**. CofiRank [31] optimizes the estimation of a ranking loss based on Normalized Discounted Cumulative Gain (NDCG).
- **CLiMF**. CLiMF [28] optimizes a different ranking-oriented loss, i.e., Mean Reciprocal Rank (MRR) loss.
- **WARP**. In [33], Weighted Approximate-Rank Pairwise loss is proposed to optimize precision@k. WARP loss differs from AUC loss in updating parameters. WARP keeps drawing negative samples until getting a disordered prediction or reaching a cutoff value.
- **kOS**. $k-$Order Statistic loss is proposed in [34] and provides a variant that optimizes precision@k.
- **USG**. USG [41] is a collaborative filtering method. It utilizes social and geographical information to improve recommendations.
- **GeoMF**. GeoMF [20] is a geographically weighted matrix factorization model.
- **Rank-GeoFM**. Rank-GeoFM [18] incorporates geographical and temporal information to provide recommendations.
- **ASMF**. ASMF [17] utilizes geographical information, social information, and attributes of businesses to enhance the accuracy of recommendations.
- **ARMF**. ARMF [17] extends ASMF by applying ranking losses.

Among these 12 baseline methods, WRMF is a point-wise matrix factorization method while MMMF and BPRMF are pair-wise based. CofiRank, CLiMF, WARP, kOS focus on optimizing top ranked positions. USG, GeoMF, Rank-GeoFM, ASMF, and ARMF utilize additional information, such as check-in locations, social relationship, businesses' attributes, and temporal information to improve the accuracy of recommendations. All parameters in baselines are tuned based on their guidelines.

In addition to the above baselines, we also implement CORALS [2] with two other gradient-based parameter optimization strategies, i.e. SGD and RMSprop [12].

- **CORALS-SGD**. CORALS-SGD applies SGD to conduct optimizations. All parameters share the same learning rate.
- **CORALS-RMSprop**. RMSprop [12] is applied to optimize learning rates adaptively. It addresses the issue of radically diminishing learning rates in AdaGrad.

3.3 Recommendation Performance

In this section, we evaluate the performances of CORALS and its variants against the 12 baseline methods. Mean Average Precision (*MAP*) is adopted as the evaluation metric. Given a ranked list rl of potential new customers, the average precision for a business b is:

$$ap_b = \frac{1}{\omega} \sum_{pos=1}^{|rl|} precision(pos) * rel(pos) \tag{12}$$

where ω is the number of new customers who visit a business b in the test set, pos denotes the position in the ranked list rl and $|rl|$ gives the total number of potential new customers in rl. Customers are ranked decreasingly based on how likely they will come in rl. $precision(pos)$ is the precision of a cut-off rank list from 1 to pos, and $rel(pos)$ is an indicator function that equals to 1 if the customer visits b in the test set, 0 otherwise. For example, three new customers visit a business b (i.e., $\omega = 3$) in the test set and they are ranked at position 2, 4, and 7 in rl, respectively. Therefore, $ap_b = \frac{1}{3}(\frac{1}{2} + \frac{2}{4} + \frac{3}{7})$. The mean average precision is the average of the average precision of all businesses.

$$MAP = \sum_{b=1}^{|B|} ap_b / |B| \tag{13}$$

MAP ranges from 0 to 1, and a higher value indicates a better performance in recommendation.

Table 5 shows the recommendation performances of different methods on the nine cities from the two datasets. The top seven rows show the performances based on the cities in the Yelp dataset, while the bottom two rows show the performances based on the cities in the Foursquare dataset. In addition, we further show the average recommendation performances for the top (10%) and tail (10%) businesses[3] in each city to demonstrate how each method performs when there is a relatively rich or poor amount of check-ins, respectively. For example, WRMF achieves 0.026 on average for all

[2] To distinguish the parameter learning algorithms used in CORALS and its variants, we also call CORALS CORALS-AdaGrad.

[3] Businesses are sorted based on their check-in numbers. Top businesses are the ones that have more check-ins, while tail businesses are the ones that have fewer check-ins.

Table 3: Business and customer statistics

Dataset	Yelp							Foursquare	
City	Charlotte	Cleveland	Las Vegas	Madison	Phoenix	Pittsburgh	Toronto	Los Angeles	New York
# of Customers	69,005	5,578	432,399	26,083	314,610	51,422	58,377	501,940	717,382
# of Businesses	10,652	9,960	282,204	3,895	43,482	8,037	20,849	215,614	206,416

Table 4: Parameter settings

Para.	η	ϵ	$iter_{max}$	S_{max}	λ_p	λ_q	λ_d	λ_{w^g}	λ_{w^r}	$\|\mathbf{p}\|$	$\|\mathbf{q}\|$	$\|\mathbf{d}\|$
Value	10^{-3}	10^{-6}	5×10^2	5×10^2	1	1	1	10^{-1}	10^{-1}	10^2	10^2	2×10

businesses in Charlotte. It achieves 0.058 and 0.014 on average for the top and tail businesses in Charlotte, respectively. We observe that the more check-ins we have for businesses, the more accurate recommendations we can achieve. This observation applies to businesses in almost all nine cities under the 15 methods. This is because the more check-ins we have for businesses, the more accurately we can infer the style, the geographical influence, and the reputation of the businesses.

MMMF, BPRMF, CofiRank, CLiMF, WARP, and kOS achieve better recommendation performances than WRMF in general. This verifies that methods achieving low prediction errors do not necessarily have high recommendation accuracies. In other words, directly optimizing the predicted check-ins may not always provide the best recommendation lists to businesses. CofiRank, CLiMF, WARP, and kOS further outperform MMMF and BPRMF due to their optimizing strategies. They optimize NDCG, MRR, precision@k, and precision@k, respectively, which all focus on better optimizing the top-ranked customers on the list. BPRMF, which optimizes AUC, focuses on optimizing the entire list of customers. CofiRank, CLiMF, WARP, and kOS outperform USG, which shows the advantage of the learning-to-rank recommendation methods. Even without utilizing location and social information, they can accurately infer customer preferences and achieve good recommendation performances. In general, WARP achieves the best recommendation performance among the 7 methods in the upper table, where only check-in information is utilized to infer customer preference.

GeoMF, Rank-GeoFM, ASMF, and ARMF outperform WRMF, MMF, BPRMF, CofiRank, CLiMF, and kOS in general. It shows that incorporating ancillary information compensate for the sparsity issue in location-based recommendation tasks. The performance of Rank-GeoFM is not as good as the one of GeoMF. This is because Rank-GeoFM, which incorporates temporal information, intends to predict the next point of interest (POI) to visit, while the task in this work is to predict new customers for POIs. GeoMF achieves better *MAP* than ASMF and ARMF. This might be because ASMF and ARMF focus on utilizing social information, while learning geographical influence might be a better way to improve recommendation performances in location-based tasks.

CORALS-AdaGrad or its variants outperform all 12 baseline methods with few exceptions, which demonstrates the effectiveness of CORALS-AdaGrad. In particular, CORALS-AdaGrad increases the mean *MAP* by 51% and 33% against WARP and GeoMF, respectively. Bold numbers in Table 5 indicate the winners for the same city and the same group of the business. In summary, CORALS-AdaGrad or its variants win in all scenarios except in Las Vegas where WARP and ASMF score slightly better.

3.4 Geographical Preference Inference

In this section, we will use examples to verify that the geographical influence is both business-dependent and customer-dependent.

We first use three case examples to show the geographical influence on different types of local businesses. We select three local businesses in Phoenix, i.e. the Phoenix Art Museum, a branch of McDonald's, and Alo Cafe. Figures 2a, 2b, 2c show the locations of the three businesses, represented by a blue mark each, together with the heat maps of their visitors. The location of a visitor is estimated by the average of all locations he/she has visited. There are two interesting observations. First, Phoenix Art Museum has more check-ins than McDonald's and Alo Cafe do. Second, the majority of the check-ins of McDonald's and Alo Cafe come from their nearby regions while the visitors of Phoenix Art Museum are scattered in the entire Phoenix. In addition, the number of museums in Phoenix is much fewer than the numbers of fast-food businesses and cafes. The rationale behind the observations is that people tend to get services from nearby businesses if the services are available since it takes less effort. However, for some business that is only available in a remote location, the customers may be more tolerant of traveling a long distance. Therefore, businesses such as fast-food and cafes get influenced more by the geographical convenience than businesses like museums. In CORALS, parameter w_b^g is used to model the geographical influence on a business b. Higher values of w_b^g indicate greater influences on the geographical convenience. In Section 3.6, we show a detailed analysis of w_b^g on various types of businesses.

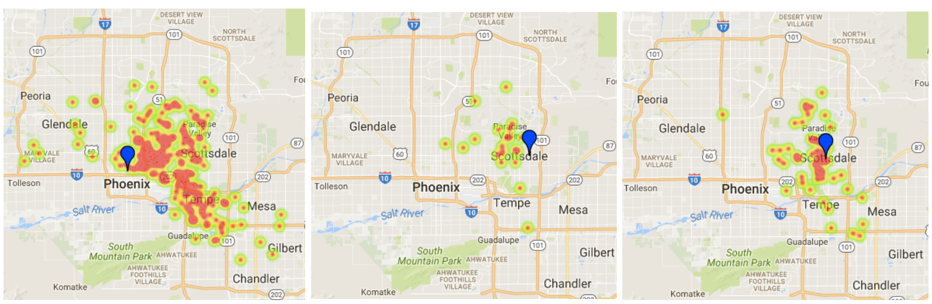
Then, we study the geographical influence on individual customers. We randomly sample two customers from Las Vegas and plot their check-ins in Figures 4a and 4b, respectively. We observe that the two customers have their own exploration preferences. User 1 tends to explore the main street in Las Vegas, while user 2 not only explores the main street but also checks in at the northwestern region of Las Vegas. Given a local business b, represented by the black marker, GMM tells $g_{b,u_1} < g_{b,u_2}$, which indicates that business b is more geographically convenient for user 2. The geographical convenience information, embedded in the GMM, helps CORALS better understand customers' decision-making processes from the perspective of the convenience of the local businesses.

Note that for each customer, we group his/her check-ins by affinity propagation to derive the number of components in the GMM. Figure 3 shows the customer percentage distributions over the number of exploration centers in different cities. We observe that most customers have only one or two exploration centers. The rationale behind it is that most customers explore around their workplaces or/and residences, which is consistent with the findings in the previous study [5].

Table 5: Recommendation performance (MAP). The upper table shows the performances of methods using only check-in information, and the lower table demonstrates the performances of methods using both check-in and heterogeneous information. `Mean` represents the average performance on all businesses in a city. `Top` represents the average performance on the top 10% businesses that have more check-ins, and `Tail` represents the average performance on the tail 10% businesses with fewer check-ins.

Method	WRMF			MMMF			BPRMF			CofiRank			CLiMF			WARP			*k*OS		
City	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail
Charlotte	0.026	0.058	0.014	0.028	0.049	0.028	0.029	0.049	0.029	0.031	0.052	0.024	0.034	0.058	0.024	0.044	0.060	0.036	0.038	0.057	0.024
Cleveland	0.041	0.086	0.029	0.039	0.072	0.025	0.040	0.073	0.030	0.043	0.078	0.042	0.050	0.081	0.043	0.055	0.077	0.046	0.053	0.085	0.041
Las Vegas	0.004	0.012	0.001	0.009	0.016	0.004	0.009	0.016	0.005	0.013	0.024	0.009	0.013	0.022	0.008	0.017	0.025	**0.015**	0.014	0.023	0.010
Madison	0.067	0.136	0.043	0.066	0.134	0.042	0.058	0.122	0.034	0.063	0.129	0.037	0.054	0.107	0.031	0.061	0.115	0.039	0.058	0.115	0.038
Phoenix	0.004	0.010	0.002	0.008	0.015	0.006	0.008	0.015	0.006	0.011	0.020	0.008	0.011	0.020	0.009	0.020	0.026	0.015	0.015	0.022	0.013
Pittsburgh	0.028	0.067	0.015	0.027	0.053	0.014	0.027	0.054	0.013	0.031	0.065	0.017	0.037	0.073	0.020	0.044	0.071	0.033	0.040	0.069	0.027
Toronto	0.009	0.019	0.005	0.011	0.018	0.009	0.011	0.017	0.009	0.014	0.025	0.011	0.014	0.022	0.011	0.021	0.031	0.019	0.019	0.029	0.015
Los Angeles	0.005	0.007	0.003	0.005	0.006	0.004	0.009	0.009	0.011	0.010	0.008	0.006	0.008	0.013	0.004	0.011	0.019	0.006	0.009	0.012	0.004
New York	0.002	0.003	0.001	0.003	0.004	0.001	0.005	0.005	0.007	0.004	0.004	0.006	0.005	0.005	0.006	0.005	0.007	0.005	0.004	0.005	0.003

Method	USG			GeoMF			Rank-GeoFM			ASMF			ARMF			CORALS-AdaGrad			CORALS-RMSprop			CORALS-SGD		
City	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail	Mean	Top	Tail
Charlotte	0.029	0.039	0.021	0.035	0.058	0.024	0.035	0.041	0.027	0.027	0.049	0.021	0.037	0.084	0.021	**0.056**	**0.087**	**0.050**	0.055	**0.087**	0.046	**0.056**	**0.087**	0.048
Cleveland	0.048	0.075	0.031	0.044	0.089	0.029	0.043	0.068	0.038	0.047	0.081	0.034	0.056	0.110	0.051	**0.091**	0.169	**0.059**	0.090	**0.171**	0.053	0.085	0.164	0.044
Las Vegas	0.008	0.019	0.004	0.017	0.024	0.012	0.011	0.018	0.010	**0.018**	**0.029**	0.011	0.010	0.016	0.005	0.014	0.026	0.010	0.014	0.026	0.010	0.014	0.026	0.010
Madison	0.063	0.104	0.047	0.077	0.148	0.038	0.063	0.112	0.044	0.072	0.151	0.048	0.089	0.184	0.043	0.116	0.192	0.091	**0.121**	0.210	0.095	0.118	**0.212**	**0.105**
Phoenix	0.010	0.017	0.006	0.020	0.023	0.017	0.014	0.016	0.011	0.017	0.023	0.012	0.016	0.019	0.011	**0.021**	0.029	**0.018**	0.020	**0.030**	0.016	0.020	0.029	**0.018**
Pittsburgh	0.030	0.055	0.023	0.038	0.069	0.032	0.042	0.047	0.030	0.047	0.071	0.030	0.041	0.090	0.033	**0.057**	0.115	**0.035**	**0.057**	**0.116**	**0.034**	0.055	0.115	0.033
Toronto	0.014	0.026	0.010	0.022	0.030	0.021	0.016	0.020	0.013	0.018	0.025	0.014	0.012	0.037	0.004	**0.027**	0.038	**0.025**	0.026	**0.040**	0.022	0.026	0.038	0.024
Los Angeles	0.020	0.025	0.017	0.021	0.021	0.017	0.008	0.011	0.006	0.009	0.010	0.007	0.008	0.011	0.004	0.021	**0.028**	**0.023**	**0.022**	0.025	**0.023**	0.019	0.024	0.019
New York	0.005	0.003	0.006	0.010	0.008	0.008	0.003	0.004	0.002	0.006	**0.009**	0.005	0.003	0.003	0.003	**0.012**	0.008	**0.012**	0.011	0.008	0.010	**0.012**	**0.009**	0.011



(a) Phoenix Art Museum (b) One branch of McDonald's (c) Alo Cafe

Figure 2: Customer heat maps for three local businesses in Phoenix

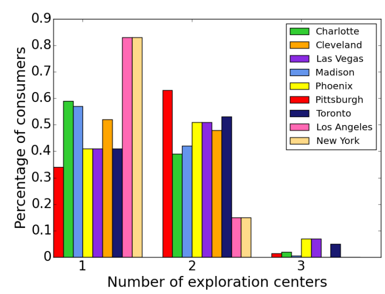
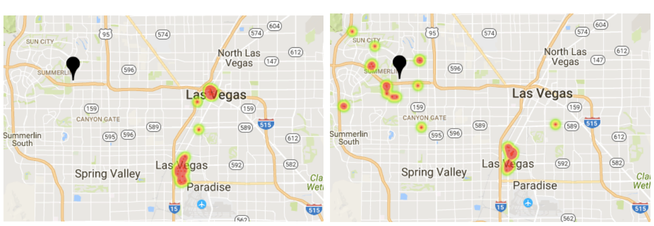


Figure 3: Exploration Center Distribution



(a) User 1's explorations (b) User 2's explorations

Figure 4: Explorations of two customers in Las Vegas

3.5 Reputation Influence Analysis

In this section, we investigate how the *MAP* performance of CORALS changes with the number of reviews considered when constructing businesses' reputation vectors. First, we do not incorporate any reviews, notated as 0 reviews. Then, we use 1, 3, 5, 7, 9, and 11 most recent reviews to construct the reputation vectors of businesses, respectively. Figure 5 shows the performance of CORALS (measured by *MAP*) on the nine cities. In particular, the performance on the Yelp dataset is plotted in solid lines, while the performance on the Foursquare dataset is plotted in dashed lines. When we ignore review information in the model, the performance is relatively poor. As long as we incorporate the information of the most recent review, the performance improves. For example, the performance increases from 0.081 to 0.097 for Madison. However, when we incorporate more reviews to construct reputation vectors, the performance gain is marginal. This is mainly due to the fact that customers only read a few latest reviews to perceive the reputation of the local business.

3.6 Analysis on Contributions of Geographical Convenience and Reputation Reliance

In this section, we analyze to what extent the geographical convenience and online reviews affect customers' decisions in visiting various types of local businesses.

We look into five types of local businesses, i.e. fast-food, bar, cafe, salon, and museum in the two largest cities, i.e., Phoenix and Las

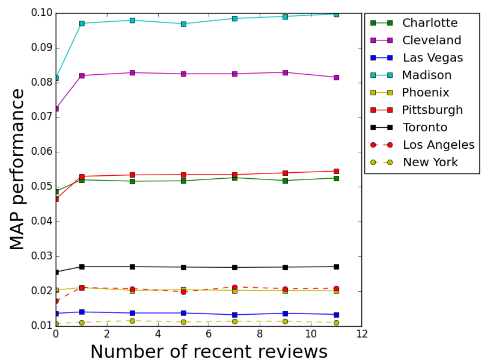


Figure 5: MAP performance over number of recent reviews

Table 6: Influential factors study for businesses in Phoenix

Phoenix	Fast-food	Bar	Cafe	Salon	Museum
Number of businesses	144	302	394	28	8
Geographical influence	0.316	0.313	0.321	0.306	0.248
Review influence	0.101	0.127	0.133	0.241	0.234

Table 7: Influential factors study for businesses in Las Vegas

Las Vegas	Fast-food	Bar	Cafe	Salon	Museum
Number of businesses	112	280	350	27	9
Geographical influence	0.302	0.29	0.298	0.297	0.232
Review influence	0.070	0.073	0.088	0.203	0.149

Vegas, in terms of the number of customers and businesses. The type of the business is inferred from the name of the business. For each type of businesses, we look into their geographical influence weights w^g and reputation influence weights w^r, and calculate the type-wise median of the influence weights. Table 6 shows the analysis based on the businesses in Phoenix. There are 144 fast-food restaurants, 302 bars, 394 cafes, 144 salons, and 8 museums. The geographical influences of fast-food restaurants, bars, and cafes are all around 0.316. For salons, the low geographical influence weight, 0.306, indicates that customers are willing to travel a little bit farther for better haircare services. For museums, which are fewer in quantity, customers have to travel farther compared with other types of businesses. The geographical influence decreases to 0.248. Moreover, the reputation also has distinct influences on different types of businesses. The reviews on fast-food restaurants, bars, and cafes have a relatively small influence on customers' decisions since customers care more about the convenience of these types of local businesses. For museums and salons, where customers care more about the experiences, reviews have a stronger influence. Table 7 shows the same analysis based on the businesses in Las Vegas, which is consistent with most discoveries in Phoenix. There is one interesting discovery about the geographical influence on bars in Las Vegas, which indicates that customers in Las Vegas are willing to take more effort in visiting faraway bars compared with fast-food restaurants, cafes, and even salons. The rationale behind it is that there are many attractive shows and events in Las Vegas bars.

4 RELATED WORK

Many recent studies [4, 20, 35, 37, 41, 43, 47] show that there is a strong correlation between customers' check-in activities and geographical distances. Thus leveraging geographical influences to improve the recommendation accuracy has been noticed by most of the current location-based recommendation work. For example, Cheng et al. [4] first detect multiple centers for each customer based on their check-in histories. Then it recommends a business with the probability that is inversely proportional to the distance between the location of the business and the nearest customer center. In [41], geographical influence is modeled by a power-law distribution between the check-in probability and the pair-wise distance of two check-ins. [21, 45] utilize the kernel density estimation to study customers' check-ins and avoid employing a specific distribution. [24] exploits geographical neighborhood information by assuming that customers have similar preferences on neighboring POIs and POIs in the same region may share similar user preferences.

Collaborative filtering algorithm is also used to fuse the check-in and geographical information, such as POI popularity, social influence, temporal influence, and content information [10, 11, 14, 17, 19, 22, 41, 43, 44, 46]. [11] conducts sentiment analysis on customers' comments to infer how good a POI is. [14, 43] apply topic models to incorporate content information. [17] uses friendship to identify potential check-ins and optimizes POI recommendations based on observed, potential, and unobserved check-ins. [40] investigates the temporal matching between POI popularity and customer regularities to recommend POIs. To model the dynamic and sequential preferences of customers, Xie et al. [38] developed a graph-based embedding model to learn the representations of POIs and recommend POIs. [23] developed a bi-weighted low-rank graph model to learn customer interests and their sequential preferences in a coherent way. PACE [39] explores the use of deep neural networks for learning user preferences over POIs.

The proposed method, CORALS, not only provides an integrated analysis of the joint effect of multiple factors, i.e., personal preference, geographical influence, and business reputation, applying the state-of-the-art learning-to-rank strategy, but also aims at proposing an explainable and flexible framework to look into the importance of integrated factors.

5 CONCLUSION AND FUTURE WORK

In this work, we study the problem of recommending new customers to local businesses in LBSN. We look into the customers' decision-making processes and propose a model, CORALS, which integrates customers' personal preferences, geographical influence, and businesses' reputation. We conduct extensive experiments to demonstrate the effectiveness of CORALS comparing to 12 different baseline methods on two real-world datasets. CORALS is flexible to incorporate new features, such as the average expense on the businesses and the customers' tolerances of the expenses. The social network information can also be easily integrated with weighted negative sampling. Moreover, CORALS can quantify the importance of incorporated factors for different types of local businesses.

ACKNOWLEDGMENTS

This project was supported by NIH U01HG008488 and Yelp Inc.. We would like to thank Shuo Song, Bin Bi, and Zijun Xue for their insightful comments and discussions. We also thank the reviewers for their constructive feedback.

REFERENCES

[1] Judd Antin, Marco de Sá, and Elizabeth F. Churchill. 2012. Local experts and online review sites. In *CSCW '12, Seattle, WA, USA, February 11-15, 2012 - Companion Volume.*

[2] Jie Bao, Yu Zheng, and Mohamed F. Mokbel. 2012. Location-based and preference-aware recommendation using sparse geo-social networking data. In *SIGSPATIAL '12, Redondo Beach, CA, USA, November 7-9, 2012.*

[3] BrightLocal. 2016. Local Consumer Review Survey. https://www.brightlocal.com/learn/local-consumer-review-survey/.

[4] Chen Cheng, Haiqin Yang, Irwin King, and Michael R. Lyu. 2012. Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks. In *AAAI '12, July 22-26, 2012, Toronto, Ontario, Canada.*

[5] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *SIGKDD '11, San Diego, CA, USA, August 21-24, 2011.*

[6] Josh Constine. 2017. Facebook relaunches Events app as Facebook Local, adds bars and food. https://techcrunch.com/2017/11/10/facebook-local/.

[7] Arthur P Dempster, Nan M Laird, and Donald B Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)* (1977).

[8] John C. Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research* 12 (2011), 2121–2159. http://dl.acm.org/citation.cfm?id=2021068

[9] Brendan J Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *science* 315, 5814 (2007).

[10] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2013. Exploring temporal effects for location recommendation on location-based social networks. In *RecSys '13, Hong Kong, China, October 12-16, 2013.*

[11] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2015. Content-Aware Point of Interest Recommendation on Location-Based Social Networks. In *AAAI '15, January 25-30, 2015, Austin, Texas, USA.* 1721–1727.

[12] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. 2012. Overview of mini-batch gradient descent. http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

[13] Bo Hu and Martin Ester. 2013. Spatial topic modeling in online social media for location recommendation. In *RecSys '13, Hong Kong, China, October 12-16, 2013.*

[14] Bo Hu and Martin Ester. 2014. Social Topic Modeling for Point-of-Interest Recommendation in Location-Based Social Networks. In *ICDM '14, Shenzhen, China, December 14-17, 2014.* 845–850.

[15] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *(ICDM '08), December 15-19, 2008, Pisa, Italy.*

[16] Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *ICML '14, Beijing, China, 21-26 June 2014.*

[17] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. 2016. Point-of-Interest Recommendations: Learning Potential Check-ins from Friends. In *SIGKDD '16, San Francisco, CA, USA, August 13-17, 2016.*

[18] Xutao Li, Gao Cong, Xiaoli Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. 2015. Rank-GeoFM: A Ranking based Geographical Factorization Method for Point of Interest Recommendation. In *SIGIR '15, Santiago, Chile, August 9-13, 2015.* 433–442.

[19] Defu Lian, Yong Ge, Fuzheng Zhang, Nicholas Jing Yuan, Xing Xie, Tao Zhou, and Yong Rui. 2015. Content-Aware Collaborative Filtering for Location Recommendation Based on Human Mobility Data. In *ICDM '15, Atlantic City, NJ, USA, November 14-17, 2015.* 261–270.

[20] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *KDD '14, New York, USA - August 24 - 27, 2014.*

[21] Moshe Lichman and Padhraic Smyth. 2014. Modeling human location data with mixtures of kernel densities. In *KDD '14, New York, USA - August 24 - 27, 2014.*

[22] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. 2013. Learning geographical preferences for point-of-interest recommendation. In *KDD '13, Chicago, IL, USA, August 11-14, 2013.*

[23] Yanchi Liu, Chuanren Liu, Bin Liu, Meng Qu, and Hui Xiong. 2016. Unified Point-of-Interest Recommendation with Temporal Interval Assessment. In *SIGKDD '16, San Francisco, CA, USA, August 13-17, 2016.*

[24] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting Geographical Neighborhood Characteristics for Location Recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014.* 739–748. https://doi.org/10.1145/2661829.2662002

[25] Maddy Osman. 2018. 28 Powerful Facebook Stats Your Brand Can't Ignore in 2018. https://sproutsocial.com/insights/facebook-stats-for-marketers/.

[26] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI '09, Montreal, QC, Canada, June 18-21, 2009.*

[27] Douglas A. Reynolds. 2009. Gaussian Mixture Models. In *Encyclopedia of Biometrics.*

[28] Yue Shi, Alexandros Karatzoglou, Linas Baltrunas, Martha Larson, Nuria Oliver, and Alan Hanjalic. 2012. CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering. In *RecSys '12, Dublin, Ireland, September 9-13, 2012.*

[29] Craig Smith. 2016. By the numbers: 20 important Foursquare Stats. http://expandedramblings.com/index.php/by-the-numbers-interesting-foursquare-user-stats/.

[30] Waldo R Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46, sup1 (1970).

[31] Markus Weimer, Alexandros Karatzoglou, Quoc V. Le, and Alexander J. Smola. 2007. COFI RANK - Maximum Margin Matrix Factorization for Collaborative Ranking. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007.*

[32] Markus Weimer, Alexandros Karatzoglou, and Alexander J. Smola. 2015. Improving maximum margin matrix factorization. *Machine Learning* 72, 3 (2015).

[33] Jason Weston, Samy Bengio, and Nicolas Usunier. 2011. WSABIE: Scaling Up to Large Vocabulary Image Annotation. In *IJCAI '11, Barcelona, Catalonia, Spain, July 16-22, 2011.*

[34] Jason Weston, Hector Yee, and Ron J. Weiss. 2013. Learning to rank recommendations with the k-order statistic loss. In *RecSys '13, Hong Kong, China, October 12-16, 2013.*

[35] Xian Wu, Yuxiao Dong, Baoxu Shi, Ananthram Swami, and Nitesh V. Chawla. 2018. Who will Attend This Event Together? Event Attendance Prediction via Deep LSTM Networks. In *Proceedings of the 2018 SIAM International Conference on Data Mining, SDM 2018, May 3-5, 2018, San Diego, CA, USA.* 180–188.

[36] Xian Wu, Yuxiao Dong, Jun Tao, Chao Huang, and Nitesh V. Chawla. 2017. Reliable fake review detection via modeling temporal and behavioral patterns. In *2017 IEEE International Conference on Big Data, BigData 2017, Boston, MA, USA, December 11-14, 2017.* 494–499.

[37] Xian Wu, Baoxu Shi, Yuxiao Dong, Chao Huang, Louis Faust, and Nitesh V. Chawla. 2018. RESTFul: Resolution-Aware Forecasting of Behavioral Time Series Data. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018.* 1073–1082.

[38] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning Graph-based POI Embedding for Location-based Recommendation. In *CIKM '16, Indianapolis, IN, USA, October 24-28, 2016.*

[39] Carl Yang, Lanxiao Bai, Chao Zhang, Quan Yuan, and Jiawei Han. 2017. Bridging Collaborative Filtering and Semi-Supervised Learning: A Neural Approach for POI Recommendation. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017.* 1245–1254. https://doi.org/10.1145/3097983.3098094

[40] Zijun Yao, Yanjie Fu, Bin Liu, Yanchi Liu, and Hui Xiong. 2016. POI Recommendation: A Temporal Matching between POI Popularity and User Regularity. In *ICDM '16, December 12-15, 2016, Barcelona, Spain.*

[41] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik Lun Lee. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR '11, Beijing, China, July 25-29, 2011.*

[42] Yelp. 2017. Yelp for Business Owners. https://biz.yelp.com/support/what_is_yelp.

[43] Hongzhi Yin, Bin Cui, Yizhou Sun, Zhiting Hu, and Ling Chen. 2014. LCARS: A Spatial Item Recommender System. *ACM Trans. Inf. Syst.* 32, 3 (2014).

[44] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat-Thalmann. 2013. Time-aware point-of-interest recommendation. In *SIGIR '13, Dublin, Ireland - July 28 - August 01, 2013.*

[45] Jia-Dong Zhang and Chi-Yin Chow. 2013. iGSLR: personalized geo-social location recommendation: a kernel density estimation approach. In *SIGSPATIAL '13, Orlando, FL, USA, November 5-8, 2013.*

[46] Jia-Dong Zhang and Chi-Yin Chow. 2015. GeoSoCa: Exploiting Geographical, Social and Categorical Correlations for Point-of-Interest Recommendations. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015.* 443–452. https://doi.org/10.1145/2766462.2767711

[47] Xuchao Zhang, Liang Zhao, Arnold P Boedihardjo, Chang-Tien Lu, and Naren Ramakrishnan. 2017. Spatiotemporal Event Forecasting from Incomplete Hyper-local Price Data. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* ACM, 507–516.