A Personalised Ranking Framework with Multiple Sampling Criteria for Venue Recommendation

Jarana Manotumruksa
University of Glasgow
Glasgow, Scotland, UK
j.manotumruksa.1@research.gla.ac.uk

Craig Macdonald, Iadh Ounis
University of Glasgow
Glasgow, Scotland, UK
first.lastname@glasgow.ac.uk

ABSTRACT

Recommending a ranked list of interesting venues to users based on their preferences has become a key functionality in Location-Based Social Networks (LBSNs) such as Yelp and Gowalla. Bayesian Personalised Ranking (BPR) is a popular pairwise recommendation technique that is used to generate the ranked list of venues of interest to a user, by leveraging the user's implicit feedback such as their check-ins as instances of positive feedback, while randomly sampling other venues as negative instances. To alleviate the sparsity that affects the usefulness of recommendations by BPR for users with few check-ins, various approaches have been proposed in the literature to incorporate additional sources of information such as the social links between users, the textual content of comments, as well as the geographical location of the venues. However, such approaches can only readily leverage one source of additional information for negative sampling. Instead, we propose a novel Personalised Ranking Framework with Multiple sampling Criteria (PRFMC) that leverages both geographical influence and social correlation to enhance the effectiveness of BPR. In particular, we apply a multi-centre Gaussian model and a power-law distribution method, to capture geographical influence and social correlation when sampling negative venues, respectively. Finally, we conduct comprehensive experiments using three large-scale datasets from the Yelp, Gowalla and Brightkite LBSNs. The experimental results demonstrate the effectiveness of fusing both geographical influence and social correlation in our proposed PRFMC framework and its superiority in comparison to BPR-based and other similar ranking approaches. Indeed, our PRFMC approach attains a 37% improvement in MRR over a recently proposed approach that identifies negative venues only from social links.

1 INTRODUCTION

With the emergence of Location-Based Social Networks (LBSNs) such as Foursquare and Yelp, users can search for interesting venues (e.g. restaurants and museums) to visit, share their location to their friends by making a check-in at the venue they have visited or leave a comment or rating to explicitly express their opinion about the venue. Such implicit and explicit sources of feedback provide rich information about both users and venues, and thus can be leveraged to study the user's movement in urban cities, as well as enhance the quality of personalised venue recommendations. Most existing venue recommendation systems (e.g. [3, 7, 16, 25, 28, 29]) apply Collaborative Filtering (CF) techniques to suggest relevant venues to users based on an assumption that similar users are likely to visit similar venues. Various venue recommendation approaches [3, 7, 15, 16] have been proposed that extend Matrix Factorisation (MF) [9], a popular CF-based technique that predicts a user's preference or rating on venues by exploiting explicit feedback (e.g. ratings and comments). Rankings of venue suggestions are then obtained based on the predicted user-venue rating generated by the MF-based model. However, in practice, users only focus on the top-K ranked list of venues, hence effective ranking-based models (e.g. learning-to-rank) that aim to generate accurate top-K venue suggestions are more useful than effective rating prediction-based models (i.e. regression models) [19]. From this point of view, MF-based approaches are not expected to perform as effectively as learning-to-rank models for the venue recommendation task [20]. In addition, explicit feedback is relatively sparse in LBSNs, which can degrade the effectiveness of the MF-based approaches [8].

To address the aforementioned challenges, various ranking-based approaches (e.g. [19]) have been proposed to leverage implicit feedback (e.g. check-ins), which is more abundant than explicit feedback [19], to generate accurate venue suggestions. Bayesian Personalised Ranking (BPR) [19] is a pairwise ranking-based model that is widely implemented and extended to leverage implicit feedback to generate the top-K venue recommendations (e.g. [14, 21, 25, 30]). The pairwise ranking criterion of the BPR model for venue recommendation is based on the assumption that a user prefers the visited venues observed from their historical check-ins over the non-visited ones. This idea results in a pairwise ranking loss function that tries to discriminate between a small set of visited venues and a very large set of all unvisited venues. Due to the imbalance between the user's visited venues and non-visited venues, the BPR model uniformly samples negative examples from the set of non-visited venues to reduce the training time.

As users have typically only visited a very small proportion of all venues in the LBSNs [21, 28], traditional BPR models typically suffer from the sparsity problem[1] that hinders the quality of the personalised venue suggestions. To mitigate the sparsity problem, various approaches have been previously proposed to leverage additional information such as social information [15, 21, 30], temporal influence [4], textual content of comments [16, 29] as well as geographical information [3, 11, 12, 24, 25, 28]. In particular, a common approach that enhances the performance of the BPR models under

[1] A common challenge in recommendation systems.

CIKM'17, November 6–10, 2017, Singapore.
 ISBN 978-1-4503-4918-5/17/11...$15.00
DOI: https://doi.org/10.1145/3132847.3132985

sparsity conditions is to extend the sampling criterion and pairwise ranking function of BPR to incorporate *additional* sources of information (e.g. social links [21, 30] and geographical information of venues [25]). However, a pairwise venue recommendation framework that seamlessly incorporates multiple types of additional information has not been previously proposed. Moreover, the various extended sampling criteria for BPR previously proposed in the literature [14, 21, 25, 30] are based on pre-defined assumptions and not on motivated by characteristics of users' movement and social interactions in LBSNs that have been observed in previous check-in studies [3, 28, 29] (This is further discussed in Section 3). Furthermore, such sampling criteria are not sufficiently flexible to incorporate additional sources of information. To address all of the aforementioned challenges, we propose a novel Personalised pairwise Ranking Framework with Multiple sampling Criteria (PRFMC) that incorporates multiple types of additional information to effectively sample negative examples and enhance the performance of the BPR model. In particular, our contributions are summarised below:

- We propose a novel Personalised pairwise Ranking Framework with Multiple sampling Criteria (PRFMC) for venue recommendation that exploits probabilistic models to effectively sample negative examples and generate personalised venues to users. In addition, PRFMC is sufficiently flexible to permit extension to incorporate multiple additional sources of information (This is further discussed in Section 4.1). To the best of our knowledge, our proposed framework (PRFMC) is the first study that extends BPR to incorporate multiple additional information.
- We propose a sampling criteria and pairwise ranking approach that applies the-state-of-the-art geographical and social probabilistic models: namely Multi-centre Gaussian and the power-law distribution models, to enhance the performance of the BPR model for venue recommendation. Our proposed approach differs from previous works [3, 28, 29] that exploit geographical influence and social correlation to directly enhance the user-venue rating prediction accuracy, whereas we leverage such influences to effectively sample negative examples as well as enhance the effectiveness of the BPR model.
- We conduct comprehensive experiments on three large-scale real-world datasets from Yelp, Brightkite and Gowalla to demonstrate the recommendation accuracy of PRFMC. The experimental results demonstrate that PRFMC consistently outperforms various state-of-the-art venue recommendation approaches in three datasets (Section 5).

The rest of this paper is organised as follows. We review related literature on Venue Recommendation in Section 2. Then, we provide the problem statement and describe the BPR model and extended BPR models for venue recommendation in Section 3, as well as some of their limitations. Our proposed PRFMC framework and its components are described in Section 4. The experimental setup for our experiments is detailed in Section 5, while comprehensive experimental results comparing the effectiveness of PRFMC with various state-of-the-art approaches are reported in Section 6 and concluding remarks follow in Section 7.

2 RELATED WORK

Conventional Recommendation systems. Matrix Factorisation (MF) is a collaborative filtering-based approach widely used to predict the ratings that users will give to items (e.g. movie and books), proposed in [9]. Traditional MF techniques aim to find *latent factors* of users and venues by leveraging the interactions between users and venues. Various existing MF-based approaches in the literature (e.g. [3, 12, 15, 16]) generate personalised venue recommendations by ranking the venues based on the predicted user-venue preference scores (e.g. rating). Such approaches can be identified as pointwise approaches [13]. Even though these approaches were designed for the venue prediction task of personalised ranking, none were directly optimised for ranking venues (i.e. focusing on getting the top-ranked suggestions that are relevant to users). Indeed, empirical studies have demonstrated that *pairwise* and *listwise* approaches are generally more effective than *pointwise* approaches for general information retrieval tasks such as web search [1, 2, 13].

However, unlike traditional information retrieval tasks, in the venue recommendation task, the recommender system needs to rank a large set of unvisited venues for each user, based on their historical feedback (i.e. check-ins or ratings on venues they previously visited) rather than small sets of candidate web documents. In addition, it is difficult to extract user-venue features as users typically visit a small set of venues in LBSNs. Hence, a listwise approach is less suited for venue recommendation. Instead, Rendle *et al.* [19] proposed a pairwise optimisation criteria, named Bayesian Personalised Ranking (BPR), which maximises a posterior estimation of *pairwise* ranking with Bayesian theory, in which an assumption is that for each user, each user's previously visited venues are preferred over their venues they have not visited.

Their empirical results demonstrated that BPR coupled with MF outperforms pointwise approaches [19]. Later, Rendle and Freudenthaler [18] proposed a non-uniform sampling approach extended from the BPR model to improve the convergence rate of the BPR learning algorithm for tag Recommendation systems. Our proposed PRFMC framework differ from Rendle and Freudenthaler's approach into two aspects: (1) PRFMC aims to enhance the effectiveness of the BPR model by incorporating multiple types of additional information using multiple sampling criteria, instead of the convergence rate and (2) we address venue recommendation rather than tag recommendation.

Venue recommendation with additional information. In contrast to non-spatial items such as movies, books and tags in conventional recommendation systems, the users of LBSNs must physically interact with the venues to consume their offered products or services (e.g. having lunch at restaurants). Various previously studies on check-in datasets on LBSNs have shown that user's movement on LBSNs can be captured by a power-law distribution [23, 26] or a multi-centre Gaussian distribution [3], while friends can influence users to visit novel venues and such behaviour can be captured by a power-law distribution model [28]. Previous literature has shown that the geographical information of venues (e.g. [3, 7, 12, 12, 23, 25, 27]) and social correlation [12, 15, 16, 21, 23, 28, 30] as well as textual content of comments (e.g. [17, 29]) are important factors to improve the effectiveness of venue recommendation systems. In particular, several approaches have been proposed to extend

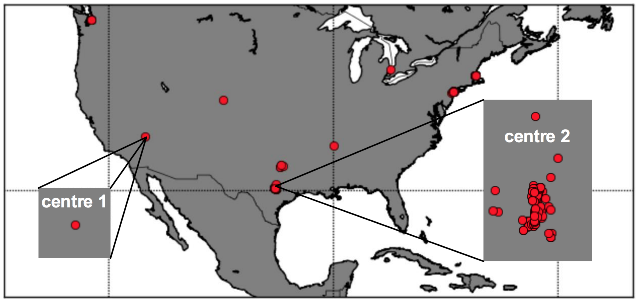

Figure 1: A typical user's multi-centres check-in behaviour sampled from the Brightkite dataset.

the BPR model to leverage additional information to enhance the effectiveness of the BPR model [14, 21, 25, 30]. However, these approaches can only incorporate one type of additional information and are not sufficiently flexible to incorporate other additional information. In addition, these approaches did not use either geographical influence or social correlation as explored in the previous studies mentioned above (Further discussed in Section 3.3-3.4). Recently, Loni *et al.* [14] proposed a pairwise ranking framework that extends the BPR model to leverage multiple types of implicit feedback (e.g. click and likes) for item recommendation. To the best of our knowledge, this is the state-of-the-art in BPR-based models that can incorporate multiple additional information. Later in Section 6, we use this model as a baseline to compare with our proposed PRFMC framework.

3 VENUE RANKING AND BPR

In this section, we first elicit the problem statement as well as the notations used in this paper (Section 3.1). Then, we briefly describe the Bayesian Personalised Ranking (BPR) model (Section 3.2) followed by the extended BPR models from the literature, which incorporate additional information and identify the limitations of these models (Section 3.3 and 3.4). Finally, Section 3.5 summarises the elicited limitations. Later, in Section 4, we describe our proposed framework that addresses these limitations.

3.1 Problem Statement

The task of venue recommendation is to generate a ranked list of relevant venues that a user might visit given his/her historical feedback (e.g. previously visited venues from their rating or check-in feedback). The historical feedback of users is represented as a matrix $R \in \mathbb{R}^{m \times n}$ where m and n are the number of users and venues, respectively. Let $r_{u,i} \in R$ denotes the rating or check-in frequency of user $u \in \mathcal{U}$ on venue $i \in \mathcal{V}$ where U and V are the set of users and venues in LBSN, respectively. Note that $r_{u,i} = 0$ means that user u has neither left a rating nor made a check-in at venue i. Social links are represented as a matrix $F \in \mathbb{R}^{m \times m}$ where F_u is the set of user u's friends.

In this paper, we define three different types of user's feedback: namely *observed*, *potential* and *unobserved* feedback. The *observed* feedback of user u is defined as the set of venues V_u^+ previously visited by user u, while the *unobserved feedback* of user u is defined as the complement $\mathcal{V}_u^- \in \mathcal{V} \backslash \mathcal{V}_u^+$. The *potential* feedback $\mathcal{V}_u^a$ of user u is defined as a type of additional feedback that can be obtained from an additional source of information a. For example, let $\mathcal{V}_{F_u}^s$ denote the social feedback that represents venues visited by the user u's friends but which user u has not visited before.

3.2 Bayesian Personalised Ranking

The Bayesian Personalised Ranking (BPR) model proposed by Rendle *et al.* [19] consists of a pairwise ranking function and a ranking-based optimisation criterion with a gradient-based learning algorithm for personalised venue recommendations. BPR creates user-venue tuples $D = \{(u,i,j) | i \in \mathcal{V}_u^+ \wedge j \in \mathcal{V}_u^-\}$ by uniformly sampling a user-venue pair (u,i) observed in R and a negative venue $j \in \mathcal{V}_u^-$ not observed in R. Indeed, BPR treats venue j sampled from the *unobserved feedback* as a negative example. However, we argue that this negative sampling criterion is not intuitive because venue j could be of interest to the user but he/she has not visited it yet (**Limitation 1**). Given a tuple $(u,i,j) \in D$, the BPR pairwise ranking function is defined as follows:

$$\hat{r}_{u,i,j}(\Theta) := \hat{y}_{u,i} > \hat{y}_{u,j}, i \in \mathcal{V}_u^+, j \in \mathcal{V}_u^- \quad (1)$$

where $\hat{r}_{u,i,j}(\Theta)$ is a pairwise ranking function that prefers venue i over venue j, Θ denotes a set of parameters and $\hat{y}_{u,i}$ is the predicted check-in frequency of user u in venue i, which can obtained from a matrix factorisation technique.

Given the tuples D, the BPR optimisation criterion is as follows:

$$BPR_{Opt}(D) = \underset{\Theta}{argmax} \sum_{(u,i,j) \in D} ln\ \sigma(\hat{r}_{u,i,j}(\Theta)) - \lambda \|\Theta\|^2 \quad (2)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is a logistic function and $\lambda \|\Theta_s\|^2$ is a regularisation term to prevent overfitting, where $\|.\|_F^2$ denotes the Frobenius norm. For each sampled tuple, the BPR algorithm updates parameters Θ with a Stochastic Gradient Descent approach based on the ranking criterion that venue i should be ranked higher than venue j (see [19] for further details).

3.3 BPR with Geographical Influences

As mentioned in Section 2, the geographical information is an important factor that influences the users' decision on visiting novel venues, while the performance of BPR can be significantly decreased due to the sparsity problem. To alleviate this problem, Yuan *et al.* [25] extended the BPR model to incorporate geographical information (GBPR). They assumed that a user is likely to visit venue g if it is nearby to venues that the user has previously visited, $\mathcal{V}_u^+$. Given a user u and a venue they have visited $i \in \mathcal{V}_u^+$, GBPR samples venue g from the *potential feedback* $\mathcal{V}_{u,i}^g$, a set of geographical neighbours of venue i within a μ threshold distance, which the user u has not visited before, as a negative example to alleviate the sparsity problem. Then, they proposed a pair ranking function that prefers an unvisited neighbourhood venue $g \in \mathcal{V}_{u,i}^g$ over an unvisited venue $j \in \mathcal{V}_u^-$. The GBPR pairwise ranking function is defined as follows:

$$\hat{r}_{u,i,g,j}(\Theta) := \hat{y}_{u,i} > \hat{y}_{u,g} \wedge \hat{y}_{u,g} > \hat{y}_{u,j}, i \in \mathcal{V}_u^+, g \in \mathcal{V}_{u,i}^g, j \in \mathcal{V}_u^- \quad (3)$$

In previous studies examining users' movements on LBSNs [3, 28, 29], it has been shown that users are likely to visit venues nearby to a venue that they often visit (e.g. their office). However, GBPR [25] uniformly samples negative venues nearby to any previously visited venues, regardless of how few other venues they have visited in the same area.

To illustrate this, consider Figure 1, showing the character of a user in different cities (centres) of the USA. In centre 1, the user has only visited one venue, while he/she has visited various venues in centre 2. Hence, the user is more likely to visit venues nearby to venues in centre 2 rather than centre 1. However, we argue that GBPR's negative sampling approach uniformly samples negative venues nearby to previously visited venues regardless of the number of venues (visited) in the neighbourhood, which may lead to a non-optimal negative sampling approach (**Limitation 2**).

3.4 BPR with Social Correlations

Apart from the extended BPR model that incorporates geographical information as mentioned above, there are two recent works [21, 30] that have incorporated social information to sample negative examples based on different criteria. Zhao *et al.* [30] proposed a social BPR model (SBPR) that leveraged social links to sample negative examples. They assumed that users are likely to visit venues previously visited by their friends. The negative sampling criterion and ranking function of SBPR are similar to GBPR's (Equation (3)) but substitute $V^{g}_{u,i}$ with $V^{s}_{F_u}$ (i.e. a set of venues visited by the user u's friends but which user u has not visited before). Recently, Wang *et al.* [21] proposed a finer-grained social BPR that extended SBPR by considering a relationship between friends, in terms of Strong and Weak-ties (SWBPR): strong-ties are friends who share mutual friends while weak-ties are friends that do not share mutual friends. In doing so, their intuition is that venues previously visited by weak-tie friends might be of more interest to the user than venues previously visited by strong-tie friends because weak-tie friends are more likely to introduce novel venues. To illustrate their intuition, the authors assumed that strong-tie friends could be friends from the same high school so they share mutual friends and their preferences are likely to be similar. In contrast, weak-tie friends can introduce new venues that are more interesting. We summarise their proposed ranking criteria as follows:

$$\hat{r}_{u,i} > \hat{r}_{u,j}, \text{ if} \begin{cases} i \in \mathcal{V}^{+}_{u} \wedge j \in \mathcal{V}^{joint}_{u} & or \\ i \in \mathcal{V}^{joint}_{u} \wedge j \in \mathcal{V}^{weak}_{u} & or \\ i \in \mathcal{V}^{weak}_{u} \wedge j \in \mathcal{V}^{strong}_{u} & or \\ i \in \mathcal{V}^{strong}_{u} \wedge j \in \mathcal{V}^{none}_{u} & \end{cases} \tag{4}$$

where $\mathcal{V}^{joint}_{u}$ is a set of venues visited by at least one strong-tie and weak-tie friends of user u, $\mathcal{V}^{weak}_{u}$ and $\mathcal{V}^{strong}_{u}$ as the set of venues visited by at least one of weak-tie friends and strong-tie friends, respectively and $\mathcal{V}^{none}_{u}$ is a set of venues visited by neither user u nor his/her friends.

Similar to GBPR, the negative sampling criteria of SBPR and SWBPR do not rely on the social correlation explored in previous literature [28] (**Limitation 3**). As mentioned in Section 2, Zhang *et al.* [28] found that the social check-in frequency and similarity between friends greatly affects the user's behaviour to visit new venues. Moreover, we argue that GBPR, SBPR and SWBPR require a pre-defined sampling assumption to generate the *potential feedback* (e.g. $\mathcal{V}^{s}_{F_u}$ and $\mathcal{V}^{weak}_{u}$), which is not sufficiently flexible to permit extension to incorporate other types of additional information (**Limitation 4**). For instance, GBPR is not sufficiently flexible to permit extension to incorporate social information.

3.5 Summary of Limitations

To conclude, in the analysis of this section, we have identified four limitations of the negative sampling approaches used in BPR-based approaches in the literature:

Limitation 1: This limitation defines the inherent disadvantage of uniformly sampling negative examples from a set of unvisited venues (BPR's negative sampling criterion).
Limitation 2: Sampling approaches for which this limitation applies are based on pre-defined assumptions of how geographical patterns define appropriate negative venues to sample.
Limitation 3: Sampling approaches for which this limitation applies are based on pre-defined assumptions of how social interactions define appropriate negative venues to sample.
Limitation 4: Sampling approaches that are built upon pre-defined assumptions are not sufficiently flexible to incorporate different types of additional information.

4 VENUE RECOMMENDATION WITH SOCIAL AND GEOGRAPHICAL INFORMATION

In this section, we explain how we exploit geographical and social information to effectively sample negative feedback venues to enhance the effectiveness of BPR. In particular, in Section 4.1, we propose a novel Personalised Ranking Framework with Multiple Sampling Criteria (PRFMC) for venue recommendation systems. PRFMC aims to address **Limitations 1 & 4)** in Section 4.1. Sections 4.2 & 4.3 explain the components of PRFMC that address **Limitations 2 & 3**. Later, in Section 6, we demonstrate the effectiveness of PRFMC in comparison with various state-of-the-art venue recommendation systems.

4.1 Personalised Ranking Framework with Multiple Sampling Criteria

For a given user u and unvisited venue i, we calculate the user's preference score $s_{u,i}$ based on the product rule as follows:

$$s_{u,i} = \prod_{a \in A} P_a(i|u) \tag{5}$$

where $P_a(i|u)$ is the estimated probability that user u will visit venue i, which takes a source of additional information a into account. Note that the product rule has been widely used to fuse different probabilistic models for venues recommendations in previous works [3, 27–29] and has shown high robustness. Indeed, the higher the score, the more likely user u will visit venue i. Unlike those previous *pointwise* approaches (e.g. [28, 29]) that rank venues based on the score $s_{u,i}$ computed in Equation (5), we propose to leverage this score to effectively sample negative examples to enhance the effectiveness of BPR. Moreover, the user's preference score $s_{u,i}$ is sufficiently flexible to be extended to incorporate different types of additional information, such as textual comments, within A. The overall process of PRFMC is described in Algorithm 1. Later in Sections 4.2 & 4.3, we discuss the probabilistic models that can be combined into Equation (5).

Algorithm 1 Learning Algorithm for PRFMC

```
1: Input: users 𝒰, venues 𝒱, visited venues 𝒱_u^+ and social links F_u for each u ∈ 𝒰
2: Output: Θ = {P ∈ ℛ^{m×d}, Q ∈ ℛ^{n×d}, b ∈ ℛ^n}
3: P ∼ U(0, 1), Q ∼ U(0, 1)
4: 𝒯 ← 0 // iteration number
5: repeat
6:     for 𝒯 ← 1 to |𝒰| do
7:         u ← draw a random user from 𝒰
8:         i ← draw a random visited venue from 𝒱_u^+
9:         j, k ← draw random unvisited venues from 𝒱_u^-
10:        if s_{u,k} > s_{u,j} then
11:            swap j and k
12:        end if
13:        Compute gradients of P_u, Q_i, Q_j, Q_k, b_i, b_j, b_k
14:        // Equation (11 - 15)
15:        Updated the above parameters
16:        // Equation (10)
17:    end for
18: until convergence
```

To tackle **Limitations 1 & 4**, we uniformly sample two unvisited venues $j, k \in \mathcal{V}_u^-$ and then calculate the user's preference score $s_{u,j}$ and $s_{u,k}$ (see Algorithm 1 Lines: 9-11). Then, the PRFMC pairwise ranking function is defined as follows:

$$\hat{r}_{u,i,j,k}(\Theta) := \begin{cases} \hat{y}_{u,i} \succ \hat{y}_{u,j} \wedge \hat{y}_{u,j} \succ \hat{y}_{u,k}, & \text{if } s_{u,j} > s_{u,k} \\ \hat{y}_{u,i} \succ \hat{y}_{u,k} \wedge \hat{y}_{u,k} \succ \hat{y}_{u,j}, & \text{otherwise} \end{cases} \tag{6}$$

As mentioned above in Section 3.4, previous sampling approaches [21, 25, 30] that generate *potential feedback* (e.g. $\mathcal{V}_{u,i}^g$ or $\mathcal{V}_u^s$) based on a particular pre-defined sampling criterion are not sufficiently flexible to incorporate different types of additional information (**Limitation 4**). In contrast, our proposed PRFMC is more flexible, since to incorporate additional sources of information we can simply define a new probability component $P_a(i|u)$, where a is the additional source of information, within Equation (5). Indeed, to extend the sampling approaches proposed by [21, 25, 30] to incorporate additional information we need to 1) adjust the sampling criterion, then 2) adjust the pairwise ranking function and re-calculate Equations (7)-(8) and (11)-(15). However, with PRFMC, we need only to extend the preference score function $s_{u,i}$ in Equation (5) to incorporate additional probabilistic models.

Based on our proposed pairwise ranking function, the objective of PRFMC can be optimised by maximising the value of the Area Under the ROC curve (AUC), which is a technique widely used to optimise pairwise ranking approaches in the literature [19, 21, 24, 25, 30]. In particular, a large AUC value indicates that the venues previously visited by a user V_u^+ are likely to be ranked higher than venues the user has not visited before V_u^-, and non-visited venues with higher preference score $s_{u,i}$ are more likely to be ranked higher than the non-visited ones with a lower score. Let Θ denote the set of all parameters to be optimised, which consists of the latent factors of users $P \in \mathcal{R}^{m \times d}$ and venues $Q \in \mathcal{R}^{n \times d}$ where d is the number of latent dimensions, and $b \in \mathcal{R}^n$ is the venues' check-in frequency bias parameter.

For each user $u \in \mathcal{U}$, the likelihood function of PRFMC can be expressed as follows:

$$\mathcal{L}(\Theta) = \prod_{u \in \mathcal{U}} \Bigg(\prod_{i \in \mathcal{V}_u^+} \prod_{j \in \mathcal{V}_u^-} P(\hat{r}_{u,i} \succ \hat{r}_{u,j} \mid \Theta) \prod_{j \in \mathcal{V}_u^-} \prod_{k \in \mathcal{V}_u^-} P(\hat{r}_{u,j} \succ \hat{r}_{u,k} \mid \Theta) \Bigg) \tag{7}$$

The likelihood function in Equation (7) aims to optimise the value of Area Under the ROC Curve (AUC) (i.e. maximising the probability that venue $i \in \mathcal{V}_u^+$ is ranked higher than venue $j \in \mathcal{V}_u^-$ and that venue j is ranked higher than venue $k \in \mathcal{V}_u^-$). To optimise the AUC likelihood function, we approximate the probability function P using the sigmoid function $\sigma(x)$, so that the likelihood function is differentiable. Then, following common practice [19], our proposed likelihood function of PRFMC can be formulated as follows:

$$\mathcal{J}(\Theta) = \underset{\Theta}{argmax} \sum_{u \in \mathcal{U}} \Bigg[\sum_{i \in \mathcal{V}_u^+} \sum_{j \in \mathcal{V}_u^-} ln(\sigma(\hat{r}_{u,i} - \hat{r}_{u,j})) + \sum_{j \in \mathcal{V}_u^-} \sum_{k \in \mathcal{V}_u^-} ln(\sigma(\hat{r}_{u,j} - \hat{r}_{u,k})) \Bigg] - \lambda_p \sum_{u \in \mathcal{U}} \|P_u\|_F^2 - \lambda_q \sum_{i \in \mathcal{V}} \|Q - i\|_F^2 - \lambda_b \sum_{i \in \mathcal{V}} b_i^2 \tag{8}$$

In Equation (8), regularisation terms are added to avoid overfitting where $\lambda_p, \lambda_q, \lambda_n$ are regularisation parameters and $\|.\|_F^2$ denotes the Frobenius norm. We use matrix factorisation to predict $\hat{r}_{u,i}$, the check-in frequency of user u on venue i based on their historical check-ins, obtained by calculating the dot product of the latent factors of the user P_u and the venue Q_i, as follows:

$$\hat{r}_{u,i} = P_u^T Q_i + b_i = \sum_{f=1}^{d} p_{u,f} \times q_{i,f} + b_i \tag{9}$$

Recall that d is the number of latent factors and b_i is the check-in frequency model parameter for venue i.

Note that our proposed framework *PRFMC* allows flexibility in using more-sophisticated MF-based check-in prediction models or other predictive models for calculating $\hat{r}_{u,i}$ in Equation (9) (e.g. Tensor Factorisation model [22]). Finally, we use Stochastic Gradient Descent (SGD) to find a local maximum of the objective function (Equation (8)). In particular, for each iteration (Algorithm 1 Lines: 13-15), given a random feedback tuple of user u who has visited venue i, but not visited venue j and k, $(u, i, j, k) \in D = \{(u, i, j, k) | i \in \mathcal{V}_u^+ \wedge j, k \in \mathcal{V}_u^-\}$, we update the model parameter $\theta \in \Theta$ based on the gradient of its corresponding parameter $\frac{\partial \mathcal{J}}{\partial x}$ while fixing the others, until convergence, as follows:

$$\theta^{(\mathcal{T}+1)} = \theta^{(\mathcal{T})} + \eta^{(\mathcal{T})} \cdot \frac{\partial \mathcal{J}}{\partial \theta}(\theta^{(\mathcal{T})}) \tag{10}$$

The gradients of latent factor matrices P_u, Q_i, Q_j, Q_k and venue bias b_i, b_j, b_k are calculated as follows:

$$\frac{\partial \mathcal{J}}{\partial P_u} = \delta(\hat{r}_{u,j} - \hat{r}_{u,i})(Q_i - Q_j) + \delta(\hat{r}_{u,k} - \hat{r}_{u,j})(Q_j - Q_k) - \lambda_p P_u \tag{11}$$

$$\frac{\partial \mathcal{J}}{\partial Q_i} = \delta(\hat{r}_{u,j} - \hat{r}_{u,i}) P_u - \lambda_q Q_i \qquad \frac{\partial \mathcal{J}}{\partial b_i} = \delta(\hat{r}_{u,j} - \hat{r}_{u,i}) - \lambda_b b_i \tag{12}$$

$$\frac{\partial \mathcal{J}}{\partial Q_j} = (\delta(\hat{r}_{u,k} - \hat{r}_{u,j}) - \delta(\hat{r}_{u,j} - \hat{r}_{u,i}))P_u - \lambda_q Q_j \quad (13)$$

$$\frac{\partial \mathcal{J}}{\partial b_j} = (\delta(\hat{r}_{u,k} - \hat{r}_{u,j}) - \delta(\hat{r}_{u,j} - \hat{r}_{u,i})) - \lambda_b b_j \quad (14)$$

$$\frac{\partial \mathcal{J}}{\partial Q_k} = -\delta(\hat{r}_{u,k} - \hat{r}_{u,j})P_u - \lambda_q Q_k \qquad \frac{\partial \mathcal{J}}{\partial b_k} = \delta(\hat{r}_{u,k} - \hat{r}_{u,j}) - \lambda_b b_k \quad (15)$$

The computational complexity of our proposed PRFMC framework consists of the calculation of MF, our proposed pairwise learning algorithm as well as the preference score function (Equation (5)). In particular, the training time of MF scales linearly with the number of check-ins in R [9]. Regarding the complexity of our proposed pairwise learning algorithm, the computation of each gradient is $O(d)$ (Equations (11)-(15)), where d is the number of latent factors. Since the probabilistic models in Sections 4.2 & 4.3 can be pre-computed the complexity of the scoring function is $O(1)$. The total complexity of PRFMC is $O(\mathcal{T} \cdot |\mathcal{U}| \cdot d)$, where $\mathcal{T}$ is the number of iterations and $|\mathcal{U}|$ is the number of users. In this respect, the computational complexity of PRFMC is equivalent to BPR, GBPR, SBPR and SWBPR and our proposed framework PRFMC is similar efficient and scalable to large datasets. In the next section, we describe how to integrate state-of-the-art probabilistic models into PRFMC.

4.2 A Negative Sampling Criterion with Geographical Influence

As discussed in Section 3.3, Yuan *et al.* [25] enhanced the effectiveness of BPR by sampling negative examples from unvisited venues nearby a previously visited venue i, $\mathcal{V}^g_{u,i}$. We argued in Section 3.3 that their proposed sampling criterion ignores the users' geographical movement, which has been widely explored in previous literature [3, 23, 26, 27, 29], and can lead to a non-optimal sampling approach. To address **Limitation 2**, we propose a novel sampling criterion that takes the users' geographical movement into account, which is captured by leveraging the probabilistic model (Multi-centre Gaussian [3] model). In particular, we use this model to estimate the preference score $s_{u,i}$ in Equation (5).

A previous study [3] on users' behaviour in LBSNs using check-in datasets have found that users typically visit venues located around several centres (e.g. home, office and travel places), and hence the probability of a user visiting a venue is inversely proportional to the distance from its nearest centre. To capture these users' movements, we apply the Multi-centre Gaussian model (MGM) proposed by Cheng *et al.* [3] to calculate the probability of a user u, visiting venue i, given a multi-centre of the user C_u as follows:

$$P_m(i|C_u) = \sum_{c_u \in C_u} P(i \in c_u) \frac{f^{\alpha}_{c_u}}{\sum_{j \in C_u} f^{\alpha}_j} \frac{\mathcal{N}(i|\mu_{c_u}, \sigma_{c_u})}{\sum_{j \in C_u} \mathcal{N}(i|\mu_j, \sigma_j)} \quad (16)$$

Equation (16) consists of a marginalisation of the product of three terms, namely:

- $P(i \in c_u)$, $\propto 1/dist(i, c_u)$, is inversely proportional to the distance between venue i and the centre c_u.
- $\frac{f^{\alpha}_{c_u}}{\sum_{j \in C_u} f^{\alpha}_j}$ denotes the normalised effect of check-in frequency r_{u,c_u} on the centre c_u, where $\alpha \in (0, 1]$ controls the check-in frequency property (i.e. the smaller α is the less significant effect on the check-in frequency).
- $\frac{\mathcal{N}(i|\mu_{c_u}, \sigma_{c_u})}{\sum_{j \in C_u} \mathcal{N}(i|\mu_j, \sigma_j)}$ denotes the probability of a venue belonging to the centre c_u, where $\mathcal{N}(i|\mu_{c_u}, \sigma_{c_u})$ is the probability density function of a Gaussian distribution, while μ_{c_u} and σ_{c_u} correspond to the mean and covariance distances of venues located around the centre c_u.

Next, we use a greedy clustering algorithm, proposed by Cheng *et al.* [3], to find the multi-centres of a user C_u. For each user u, we start from the most visited venue of the user in $\mathcal{V}^+_u$, and combine all other visited venues from $\mathcal{V}^+_u$ whose distance is less than κ kilometres from the selected venue, into a given region. If the ratio of the total check-in number of venues in this region to the user's total check-in number is greater than a threshold ϕ, we set these check-in venues as a region and determine the most visited check-in venue as the centre of the region. Algorithm 2 shows the procedure for discovering the multiple centres of all users.

Algorithm 2 Multi-centre Discovering Algorithm [3]

for $u \in \mathcal{U}$ **do**
  Sort all venues in $\mathcal{V}^+_u$ according to visiting frequency
  $\forall i \in \mathcal{V}^+_u,\ v_i.centre = -1$
  $centre\_list = \emptyset, centre\_no = 0$
  for $i = 1 \rightarrow |\mathcal{V}^+_u|$ **do**
    $centre\_no++,\ centre = \emptyset$
    $centre.total\_freq = 0$
    $centre.add(v_i), centre.total\_freq += v_i.freq$
    for $j = i + 1 \rightarrow |V^+_u|$ **do**
      if $v_i.centre == -1 \wedge dist(v_i, v_j) \leq \kappa$ **then**
        $v_j.centre = centre\_no, centre.add(v_j)$
        $centre.total\_freq += v_j.freq$
      end if
    end for
    if $centre.total\_freq \geq u.total\_freq \times \phi$ **then**
      $centre\_list.add(centre)$
    end if
  end for
  return centre_list for u
end for

4.3 A Negative Sampling Criterion with Social Correlation

Previous works [3, 21, 28–30] have shown that friends can influence each other (i.e. they are likely to visit similar venues). As argued in Section 3.4, previous works by Zhao *et al.* [30] and Wang *et al.* [21] sampled the negative venues from venues previously visited by the user's friends (e.g. $V^s_{F_u}$ and $\mathcal{V}^{weak}_u$) based on their proposed pre-defined sampling assumptions (**Limitation 3**), which did not take social interactions previously observed in other works [28] into account. Indeed, Zhang *et al.* [28, 29] found that users are more likely to visit venues that their friends often visited and similarly friends are also likely to visit similar venues and such social interactions follow the power-law distribution. To address **Limitation 3**, we propose to apply the social relevance model based on the power-law distribution proposed by Zhang *et al.* [28] to effectively sample negative examples to enhance the effectiveness of BPR. Note that our contribution in this section differs from that of Zhang *et al.* [28], since we apply the social relevance model to effectively sample negative examples, while Zhang *et al.* used this model to predict a user's rating on unvisited venues. Later in Section 6, we demonstrate that our proposed sampling approach significantly outperforms several social-based BPR approaches. The social relevance model consists of three steps: social aggregation, distribution estimation of social check-in frequency and social relevance score computation.

Step 1: Social aggregation. Given a user u and an unvisited venue i, we aggregate the check-in frequency of user u's friends on venue i, as follows:

$$x_{u,i} = \sum_{f \in F_u} r_{f,i} \tag{17}$$

Then we transform the social check-in frequency into normalised relevance based on the social check-in frequency distribution, which is learned from the historical check-in of all users.

Step 2: Distribution estimation of social frequency. In real-world datasets, the social check-in frequency random variable x follows a power-law distribution [28], the probability density function of which is defined by:

$$f_{So}(x) = (\beta - 1)(1 + x)^{-\beta},\ x \geq 0,\ \beta > 1 \tag{18}$$

where β is estimated by the check-in matrix R and the social links matrix F, as follows:

$$\beta = 1 + |\mathcal{U}||\mathcal{V}| \left[\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{V}} ln(1 + \sum_{f \in F_u} r_{f,i}) \right]^{-1} \tag{19}$$

Step 3: Social relevance score computation. The estimated probability density function f_{So} in Equation (18) is monotonically decreasing with respect to the social check-in frequency x, but the social relevance score should be monotonically increasing with regard to the social check-in frequency, because users who have friends with whom they have common visited venues should have high social relevance scores. Thus, we define the social relevance score of $x_{u,l}$ in Equation (17) based on the cumulative distribution function of f_{So}, given by:

$$P_s(i|u) = \int_0^{x_{u,i}} f_{So}(z)dz = 1 - (1 + x_{u,i})^{1-\beta} \tag{20}$$

such that $P(i|u)$ is monotonically increasing with respect to the social check-in frequency $x_{u,i}$. Moreover, based on the cumulative distribution probability $P(i|u)$ in Equation (20), the social check-in frequency $x_{u,i}$ is transformed into a social relevance score that reflects the relative position of $x_{u,i}$ in all the social check-in frequencies of users on venues.

5 EXPERIMENTAL SETUP

In the remainder of the paper, we evaluate the effectiveness of our proposed PRFMC framework by comparing with state-of-the-art venue recommendation approaches. In particular, we aim to address two research questions, which we now elicit. Firstly, as argued by **Limitations 2 & 3**, previous works sample negative venues based on pre-defined assumptions with respect to the type of additional information that they use (e.g. users like venues previously visited by friends). However, as argued in Section 3, an effective negative sampling approach should build upon known results for identifying the user's movement and social interactions. Hence, our first research question is:

RQ1 *Can we effectively sample negative venues by leveraging the geographical influence and social correlation?*

Furthermore, as discussed in Section 2, no previous attempt has combined the negative sampling approaches to enhance the performance of venue ranking approaches. Hence, our second research question is the following:

RQ2 *Is a negative sampling approach based on multiple criteria more effective than a sampling approach with a single criterion in improving the quality of venue suggestions?*

Note that **Limitations 1 & 4** have been addressed in the PRFMC framework discussed in Section 4.1 and do not require experimental verification. In the remainder of this section, we describe the experimental setup in terms of datasets (Section 5.1), baselines (Section 5.2) and algorithm parameters (Section 5.3). The experimental results and analysis follow in Section 6.

Table 1: Statistics of three datasets

	Yelp	Brightkite	Gowalla
Number of users	40,228	25,063	72,953
Number of venues	34,932	48,177	131,328
Number of ratings or check-ins	987,050	3,309,555	3,487,258
Number of social links	1,598,096	33,290	330,762
% density of User-Venue matrix	0.0702	0.2740	0.0363

5.1 Datasets & Measures

All our experiments are conducted using publicly available large-scale LBSN datasets. In particular, to show the generalisation of our proposed framework across multiple LBSN platforms and sources of feedback evidence, we use two check-ins datasets from Gowalla and Brightkite[2], and a rating dataset from Yelp[3]. For each dataset, we conduct experiments using a 5-fold cross-validation, where each fold has 60% training, 20% validation and 20% test instances (check-ins/ratings). Due to the high sparsity of the datasets, we follow the common practice from previous works [6, 10, 19, 25, 29] to filter out users/venues with less than 10 interactions. Table 1 summarises the statistics of the filtered datasets.

For each dataset, we measure the quality of the ranked venue recommendations in terms of Mean Average Precision (MAP), Normalised Discounted Cumulative Gain (NDCG) and Mean Reciprocal Rank (MRR), which are widely used in the recent recommendation literature [14, 21, 25, 30]. In particular, MAP and MRR consider the ranking nature of the task, by taking into account the rank(s) of the venues that each user has previously visited/rated in the produced ranking, while NDCG goes further by considering the check-in frequency/rating value of the user as the graded relevance label. Lastly, significance tests are conducted using a paired t-test with $p < 0.01$.

5.2 Baselines

In this paper, we propose a novel Personalised Ranking Framework with Multiple sampling Criteria (PRFMC_{MS}) that consists of

[2] https://snap.stanford.edu/data/
[3] https://www.yelp.com/dataset_challenge

two components: namely Multi-centre Gaussian and Social power-Law distribution models. We compare the effectiveness of each component (i.e. PRFMC_M incorporates geographical information and PRFMC_S incorporates social links) with state-of-the-art venue recommendation approaches that incorporate similar additional sources of information. In particular, we compare PRFMC_{MS}, with a number of baselines, which can be grouped into categories, namely: traditional BPR, geo-based approaches, social-based approaches and hybrid approaches combining social- and geo-based BPR. In the following we explain our implementation of each baseline in details. All baselines and our proposed PRFMC framework are implemented using LibRec [5], a Java library for recommendation systems.

5.2.1 Traditional BPR.

BPR. This is the classical pairwise ranking approach, coupled with matrix factorisation for user-venue rating/check-in frequency prediction proposed by Rendle *et al.* [19].

5.2.2 Geo-based approaches.

GMG. This is a Multi-center Gaussian Model that incorporates geograpical influence proposed by Cheng *et al.* [3]. Recommendations are generated by ranking all venues according to the score computed by Equation (16) (see Section 4.2 for further details).

GBPR. This is a state-of-the-art BPR model that incorporates the geographical influence model proposed by Yuan *et al.* [25]. Their model assumes that neighbourhood venues of venues previously visited by users should be ranked higher than the distant ones (see Section 3.3 for further details).

5.2.3 Social-based approaches.

SPLD. This is a Social Power-Law Distribution model that incorporates social influences proposed by Zhang *et al.* [28]. In particular, venue recommendations are generated by ranking all venues according to the score computed by Equation (20).

SBPR. This is a Social BPR model that leverages social information proposed by Zhao *et al.* [30]. Their model's ranking criterion assumes that venues previously visited by the user's friends should be ranked higher than venues neither the user nor his/her friends visited (see Section 3.4 for further details).

SWBPR. A state-of-the-art BPR model that is extended from SBPR proposed by Wang *et al.* [21]. This model considers *Strong* and *Weak* Social ties of the user's friends. Their ranking criterion assumes that venues visited by weak tie friends should be ranked higher than venues visited by strong tie friends, because weak tie friends are likely to introduce novel and diverse venues (again, Section 3.4 provides further details).

5.2.4 Hybrid (social & geo)-based approaches.

GeoSo. A state-of-the-art probabilistic model that incorporates both geographical and social influences proposed by Zhang *et al.* [28]. To permit a fair evaluation, we have re-implemented their GeoSoCa approach to consider only geographical and social information, and ignore the categorical properties of venues, in common with our proposed approach that also does not consider categories.

GSBPR. This model combines **GBPR** and **SBPR** together by assuming that the neighbourhood venues visited by the user's friends should be ranked higher than the distant ones. The optimisation criterion of this model is $BPR_{Opt}(D_{gs})$, where:

$$D_{gs} = \left\{(u,i,k,j) \mid i \in V_u^+ \wedge k \in V_{u,i}^g \cap V_{F_u}^s \wedge j \in V_u^- \right\}.$$

Indeed, D_{gs} contains tuples (u,i,k,j) where user u has visited venue i, k is neighbouring venue of venue i that the user has not visited but his/her friends have visited, and j is a venue never visited by neither user u nor by his/her friends.

BPRMC. This is a state-of-the-art BPR model that can simultaneously incorporate multiple sampling approaches (i.e. GBPR and SBPR) based on a pre-defined weight of each sampling approach proposed by Loni *et al.* [14]. This approach is a suitable baseline, as it permits a fair comparison of our proposed PRFMC framework with another that considers multiple sampling approaches.

5.3 Recommendation Parameter Setup

To permit a fair comparison, our proposed PRFMC framework and all of the BPR-based baselines deploy Matrix Factorisation (MF) as the prediction function. Following common practice [15, 16, 21, 25], the MF's parameters are set as follows: the dimension of the latent factors $d = 10$, and the regularisation parameters $\lambda_p, \lambda_q, \lambda_b = 0.001$. To the fullest extent possible, we apply the parameters used by the baselines and probabilistic models (**GMG**, **SPLD** and **GeoSo**) when these were applicable, i.e. when the values reported in the corresponding papers were recommended for the datasets we use in this paper. For instance, following [3], we set MGM's parameters as follows: $\phi = 0.02$, the distance threshold $\kappa = 15$ and the frequency control parameter $\alpha = 0.2$. SWBPR's parameters are determined using the validation set for each fold. Similarly, for other approaches not previously reported on these datasets, we determine the values for their parameters using the validation set for each fold.

6 EXPERIMENTAL RESULTS

Table 2 reports the effectiveness of various approaches in term of the MAP, NDCG and MRR measures on the three different datasets. The grouped columns of the table correspond to the grouping of baseline approaches based upon the sources of additional information, as discussed in Section 5.2, along with the corresponding implementation of PRFMC.

Firstly, on inspection of Table 2, we note that the relative venue recommendation quality of the baselines on the three datasets in terms of two measures are consistent with the results reported for the various baselines in the corresponding literature [14, 21, 25, 30]. For instance, GBPR outperforms BPR by 3-9% across three datasets [25] and SWBPR outperforms SBPR by 0.22-25% across three datasets. Note that previous works [14, 21, 30] used different datasets, while our reimplementations of their proposed approaches obtain relatively similar improvements. We now analyse in turn each group of approaches based upon the source of additional information employed.

Models with Geographical Influence. Within the Geo-based group of columns in Table 2, we compare PRFMC_M with MGM and GBPR, which are the probabilistic model and extended BPR model that

Table 2: Performance in terms of MAP, NDCG and MRR of various approaches. For each type of additional information and evaluation measure, the best performing result is highlighted in bold and $*$ indicates significant differences in terms of paired t-test with $p < 0.01$, comparing to the best performing result. Percentage differences compared to BPR are denoted by Δ.

Dataset	Measure		Geo-based			Social-based				Hybrid geo- & social-based			
		BPR	MGM	GBPR	PRFMC$_M$	SPLD	SBPR	SWBPR	PRFMC$_S$	GeoSo	GSBPR	BPRMC$_{GS}$	PRFMC$_{MS}$
Yelp	MAP	0.1974	0.0080*	0.2037*	**0.2071**	0.0011*	0.2014*	0.2051*	**0.2101**	0.0062*	0.2016*	0.2042*	**0.2109**
	Δ		-95.94%	3.18%	4.89%	-99.44%	2.03%	3.89%	6.45%	-96.86%	2.12%	3.44%	6.81%
	NDCG	0.3253	0.1575*	0.3467*	**0.3587**	0.1246*	0.3451*	0.3521*	**0.3625**	0.1581*	0.3432*	0.3519*	**0.3690**
	Δ		-51.58%	6.58%	10.25%	-61.70%	6.07%	8.21%	11.43%	-51.41%	5.48%	8.15%	13.41%
	MRR	0.2186	0.0197*	0.2343*	**0.2402**	0.0031*	0.2275*	0.2384*	**0.2506**	0.0137*	0.2295*	0.2361*	**0.2492**
	Δ		-90.96%	7.19%	9.91%	-98.59%	4.11%	9.07%	14.68%	-93.73%	5.01%	8.04%	14.01%
Gowalla	MAP	0.0703	0.0511*	0.0724*	**0.0826**	0.0003*	0.0722*	0.0758*	**0.0843**	0.0503*	0.0724*	0.0732*	**0.0933**
	Δ		-27.35%	3.01%	17.50%	-99.59%	2.79%	7.91%	19.99%	-28.41%	3.07%	4.11%	32.76%
	NDCG	0.2485	0.2307*	0.2578*	**0.2929**	0.1054*	0.2669*	0.2678*	**0.2894**	0.2259*	0.2592*	0.2717*	**0.3174**
	Δ		-7.17%	3.71%	17.83%	-57.60%	7.40%	7.76%	16.43%	-9.11%	4.28%	9.32%	27.69%
	MRR	0.0881	0.1142*	0.0951*	**0.1259**	0.0010*	0.0877*	0.1098*	**0.1364**	0.0779*	0.0958*	0.0906*	**0.1510**
	Δ		29.61%	7.93%	43.00%	-98.85%	-0.42%	24.63%	54.82%	-11.53%	8.78%	2.87%	71.47%
Brightkite	MAP	0.1561	0.0459*	0.1607*	**0.1854***	0.0749*	0.1518*	0.1528*	**0.1649**	0.1132*	0.1470*	0.1525*	**0.1857**
	Δ		-70.62%	2.94%	18.76%	-52.05%	-2.77%	-2.11%	5.62%	-27.46%	-5.81%	-2.34%	18.95%
	NDCG	0.3026	0.1997*	0.3109*	**0.3618**	0.2124*	0.3042*	0.3007*	**0.3165**	0.2816*	0.2939*	0.3055*	**0.3647**
	Δ		-34.00%	2.73%	19.56%	-29.81%	0.53%	-0.62%	4.58%	-6.95%	-2.88%	0.96%	20.50%
	MRR	0.1738	0.0786*	0.1841*	**0.2170**	0.1202*	0.1688*	0.1692*	**0.1916**	0.1577*	0.1614*	0.1696*	**0.2193**
	Δ		-54.77%	5.88%	24.86%	-30.84%	-2.89%	-2.67%	10.24%	-9.25%	-7.14%	-2.41%	26.16%

incorporate geographical influence, respectively. We observe that PRFMC$_M$ consistently and significantly outperforms MGM and GBPR for MAP, NDCG and MRR across all datasets. This implies that our proposed negative sampling approach that considers the user's movement captured by the Multi-centre Gaussian model (MGM) is more effective than the GBPR approach [25], which itself relies on a pre-defined assumption on the likely relevance of neighbouring venues, as summarised by **Limitation 2**. In particular, PRFMC$_M$ can enhance the effectiveness of the BPR model by approximately 4-43% for three metrics across the three datasets.

Models with Social Correlation. Next, we consider the social-based column group, to compare the effectiveness of PRFMC$_S$ with SPLD, SBPR and SWBPR. Trends that are similar in nature to those observed for the geo-based approach group are observed, in that the PRFMC$_S$ significantly outperforms the probabilistic model (SPLD) and extended BPR models that incorporate social information (i.e. SBPR and SWBPR) based on the pre-defined sampling assumptions that venues previously visited by friends are likely to be visited **Limitation 3**. Interestingly, the relatively low results for SBPR and SWBPR across MAP, NDCG and MRR on the Brightkite dataset are likely due to the sparsity of the social links between the users in Brightkite LBSN (see Table 1). In contrast, PRFMC$_S$ can improve the effectiveness of BPR, whereas SBPR and SWBPR both do not. Indeed, we find that sampling negative venues using the power-law distribution model is more effective than the pre-defined sampling criteria proposed by [21, 30]. Moreover, exploiting the power-law distribution model to sample negative venues is more useful to enhance the quality of venue recommendation than simply ranking venues according to the score computed by SPLD model. In particular, PRFMC$_S$ can enhance the effectiveness of the BPR model by 5-54% for three metrics across the three datasets.

Together, the analyses conducted individually for the geo- and social-based models allow us to conclude that for research question RQ1, leveraging the geographical influence and social correlation through PRFMC increases the various ranking metrics by approximately 4-43% and 4-54%, for the geo- and social-based negative sampling approaches, respectively, and thereby overall significantly outperforms the MGM, GBPR, SPLD, SBPR and SWBPR approaches.

Hybrid geo- and social-based models. Next, we consider the deployment of hybrid models that combine both geo- and social-based additional sources of information within the negative sampling. In doing so, we compare our proposed framework PRFMC with BPRMC and GSBPR. In particular, we compare our proposed framework that is comprised of geographical and social components PRFMC$_{MS}$, with the state-of-the-art BPR models that can incorporate multiple sampling criteria BPRMC$_{GS}$.

We first discuss the effectiveness of GSBPR, BPRMC$_{GS}$ & PRFMC$_{MS}$ in comparison with each of their constituent geo-based and social-based component baselines. In particular, from Table 2 we observe that the results of GSBPR are generally not higher than both of its constituents that each consider only one sampling criterion (i.e. GBPR and SBPR). This implies that simply combining the sampling criteria (as done by GSBPR) is not a suitable approach. In contrast, BPRMC$_{GS}$ is more effective than *GSBPR* at combining multiple sampling criteria. Moreover, by comparing BPRMC$_{GS}$ with GBPR and SBPR, we find that, for three metrics in the Yelp dataset, BPRMC$_{GS}$ outperforms the extended BPR models that consider only a single sampling criterion (i.e. GBPR and SBPR). However, for the Gowalla and Brightkite datasets, the effectiveness of BPRMC$_{GS}$ greatly decreases when one of the constituent sampling criterion is not effective. For instance, regarding the results of GBPR, SBPR and BPRMC$_{GS}$ in terms of MAP and MRR in the Brightkite dataset, we observe that when the performance of SBPR decreases, the effectiveness of BPRMC$_{GS}$ also decreases. A similar observation is found for BPRMC$_{GS}$ in terms of MRR in the Gowalla dataset. These results imply that BPRMC_{GS} cannot distinguish the effectiveness of its combined sampling criteria.

Next, we compare the effectiveness of PRFMC that considers different sampling criterion (i.e. PRFMC_M, PRFMC_S and PRFMC_{MS}). The results show that our proposed framework PRFMC_{MS} which samples negative examples based on both geographical influence and social correlation – captured by the Multi-centre Gaussian model and the power-law distribution, respectively – outperforms both PRFMC_S and PRFMC_M, across all three metrics on all three datasets, with a single exception, namely: AUC for the Brightkite dataset, where PRFMC_M slightly outperforms PRFMC_{MS}. This single exception is likely explained by the comparative underperformance of PRFMC_S for that metric and dataset. Overall, the strong results for PRFMC_{MS} demonstrate the effectiveness of PRFMC in combining different types of sampling criteria. In addition, unlike BPRMC_{GS}, the effectiveness of PRFMC_{MS} does not decrease if one of the fused sampling criteria is not effective.

Hence, in response to research question RQ2, we find that our PRFMC framework provides a significant benefit across various datasets and measures, compared to various existing state-of-the-art single criterion negative sampling approaches as well as probabilistic models (i.e. MGM, SPLD and GeoSo). Indeed, among the results reported in Table 2, all of the highest improvements over the classical BPR baseline, for all three measures on all three datasets, are by the PRFMC_{MS} hybrid negative sampling approach. Indeed, for the Gowalla dataset, PRFMC_{MS} attains a 71% improvement over the MRR of BPR, as well as 37% and 59% improvements in MRR over the recently proposed SWBPR [21] and GBPR [25] approaches (Table 2: 0.1098→0.1510; 0.0951→0.1510), respectively.

7 CONCLUSIONS

In this paper, we explored various techniques to effectively sample negative examples to improve the effectiveness of the BPR model for venue recommendation on LBSNs. In particular, we proposed a novel Personalised Ranking Framework with Multiple sampling Criteria (PRFMC) to incorporate different sources of additional information. In addition, we proposed negative sampling approaches that exploit existing probabilistic models (i.e. Multi-centre Gaussian and the power-law distribution models) in a new manner, namely to consider previously observed users' movement and social interactions, when sampling negative training instances.

Our comprehensive experiments on three large-scale datasets on Yelp, Gowalla and Brightkite LBSNs demonstrate the effectiveness of our proposed framework (PRFMC) as well as the sampling approaches for venue recommendation, which are superior to various state-of-the-art venue recommendation approaches. For instance, on the Gowalla dataset, PRFMC_{MS} attains a 37% improvement in MRR over the recently proposed SWBPR approach [21]. Moreover, these improvements are attained without increased computational complexity compared to the baseline approaches. For future work, we plan to apply more sophisticated probabilistic models to capture the semantic influence of textual contents of comments left by the user's friends to further improve the effectiveness of our proposed sampling approach.

REFERENCES

[1] Christopher J. Burges, Robert Ragno, and Quoc V. Le. 2007. Learning to Rank with Nonsmooth Cost Functions. In *Proc. of NIPS.* 193–200.

[2] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *Proc. of ICML.* 129–136.

[3] 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 *Proc. of AAAI.* 17–23.

[4] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2013. Exploring temporal effects for location recommendation on location-based social networks. In *Proc. of RecSys.* 93–100.

[5] Guibing Guo, Jie Zhang, Zhu Sun, and Neil Yorke-Smith. 2015. LibRec: A Java library for recommender systems. In *Proc. of UMAP.*

[6] Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *Proc. of SIGIR.*

[7] Longke Hu, Aixin Sun, and Yong Liu. 2014. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *Proc. of SIGIR.*

[8] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Proc. of ICDM.* 263–272.

[9] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.

[10] Huayu Li, Yong Ge, and Hengshu Zhu. 2016. Point-of-Interest Recommendations: Learning Potential Check-ins from Friends. In *Proc. of KDD.*

[11] Xutao Li, Gao Cong, Xiao-Li Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. 2015. Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation. In *Proc. of SIGIR.* 433–442.

[12] 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 *Proc. of SIGKDD.* 831–840.

[13] Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331.

[14] Babak Loni, Roberto Pagano, Martha Larson, and Alan Hanjalic. 2016. Bayesian Personalized Ranking with Multi-Channel User Feedback. In *Proc. of RecSys.* 361–364.

[15] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *Proc. of WSDM.* 287–296.

[16] Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis. 2016. Regularising factorised models for venue recommendation using friends and their comments. In *Proc. of CIKM.* 1981–1984.

[17] Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis. 2017. Matrix Factorisation with Word Embeddings for Rating Prediction on Location-Based Social Networks. In *Proc. of ECIR.* Springer, 647–654.

[18] Steffen Rendle and Christoph Freudenthaler. 2014. Improving pairwise learning for item recommendation from implicit feedback. In *Proc. of WSDM.* 273–282.

[19] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proc. of UAI.* 452–461.

[20] 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 *Proc. of RecSys.* 139–146.

[21] Xin Wang, Wei Lu, Martin Ester, Can Wang, and Chun Chen. 2016. Social Recommendation with Strong and Weak Ties. In *Proc. of CIKM.* 5–14.

[22] Lina Yao, Quan Z Sheng, Yongrui Qin, Xianzhi Wang, Ali Shemshadi, and Qi He. 2015. Context-aware Point-of-Interest Recommendation Using Tensor Factorization with Social Regularization. In *Proc. of SIGIR.*

[23] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proc. of SIGIR.* 325–334.

[24] Haochao Ying, Liang Chen, Yuwen Xiong, and Jian Wu. 2016. PGRank: Personalized Geographical Ranking for Point-of-Interest Recommendation. In *Proc. of WWW.* 137–138.

[25] Fajie Yuan, Guibing Guo, Joemon Jose, Long Chen, and Haitao Yu. 2016. Joint Geo-Spatial Preference and Pairwise Ranking for Point-of-Interest Recommendation. In *Proc. of ICTAI.* 46–53.

[26] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. 2013. Time-aware point-of-interest recommendation. In *Proc. of SIGIR.*

[27] Jia-Dong Zhang and Chi-Yin Chow. 2013. iGSLR: personalized geo-social location recommendation: a kernel density estimation approach. In *Proc. of SIGSPATIAL.* 334–343.

[28] Jia-Dong Zhang and Chi-Yin Chow. 2015. GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proc. of SIGIR.* 443–452.

[29] Jia-Dong Zhang, Chi-Yin Chow, and Yu Zheng. 2015. ORec: An opinion-based point-of-interest recommendation framework. In *Proc. of CIKM.* 1641–1650.

[30] Tong Zhao, Julian McAuley, and Irwin King. 2014. Leveraging social connections to improve personalized ranking for collaborative filtering. In *Proc. of CIKM.*