

Locus recommendation using probabilistic matrix factorization techniques

Recomendación de locus para técnicas de factorización matricial probabilístico

Recomendação locus para técnicas de factorização de matrizes probabilísticas

Rachna Behl¹
Indu Kashyap²

Received: September 20th, 2020

Accepted: December 1st, 2020

Available: January 11th, 2021

How to cite this article: R. Behl, I. Kashyap. "Locus Recommendation Using Probabilistic Matrix Factorization Techniques," *Revista Ingeniería Solidaria*, vol. 17, no. 1, 2021, doi: <https://doi.org/10.16925/2357-6014.2021.01.10>

Research article. <https://doi.org/10.16925/2357-6014.2021.01.10>

1 Department of Computer Science and Engineering. Faculty of Engineering and Technology
email: rachnabel.fet@mriu.edu.in

Orcid: 0000-0002-8333-9178

2 Department of Computer Science and Engineering
FACULTY OF ENGINEERING AND TECHNOLOGY
email: indu.fet@mriu.edu.in

Orcid: 0000-0003-1884-0828



Abstract

Introduction: The present paper is the outcome of the research "Locus Recommendation using Probabilistic Matrix Factorization Techniques" carried out in Manav Rachna International Institute of Research and Studies, India in the year 2019-20.

Problem: Location Based recommendation systems (LBSNs) are mainstream these days due to the increasing pervasiveness of mobile devices. They provide recommendations to users based on the places which are frequently visited by them or by the people who are socially connected to them. These recommendation services use check-in information to mine users' patterns and provide interesting and attractive locations to different users. POI recommendation applications aim to provide personalized recommendations of places of interest to users to enrich their experiences.

Objective: The objective of the research is to recommend top- n locations by applying various probabilistic matrix factorization techniques.

Methodology: Matrix factorization is a model based collaborative technique for recommending new items to the users. These techniques are found to be effective and more accurate as they are based on discovering the latent features underlying the user-item rating matrix. In LBSN the rating matrix is generally sparse, thus the concept of matrix factorization is used to generate recommendations.

Results: Experimental results on two real-world LBSNs showed that PFM consistently outperforms PMF. This is because the technique is based on gamma distribution to model user and item matrix. Using gamma distribution is reasonable for check-in frequencies which are all positive in real datasets. However, PMF is based on Gaussian distribution that can allow negative frequency values as well.

Conclusion: The motive of the work is to identify the best technique for recommending locations with the highest accuracy and allow users to choose from a plethora of available locations; the best and interesting location based on the individual's profile.

Originality: A rigorous analysis of Probabilistic Matrix Factorization techniques has been performed on popular LBSNs and the best technique for location recommendation has been identified by comparing the accuracy viz RMSE, Precision@N, Recall@N, F1@N of different models.

Limitations: User's contextual information like demographics, social and geographical preferences have not been considered while evaluating the efficiency of probabilistic matrix factorization techniques for POI Recommendations.

Keywords: Collaborative filtering, Information Filtering, LBSN, Matrix Factorization, POI

Resumen

Introducción: El presente artículo es el resultado de la investigación "Recomendación de locus utilizando técnicas de factorización de matrices probabilísticas" llevada a cabo en el Instituto Internacional de Investigación y Estudios Manav Rachna, India, en el año 2019-20.

Problema: Los sistemas de recomendación basados en la ubicación (LBSN) son la corriente principal en estos días debido a la creciente omnipresencia de los dispositivos móviles. Ofrecen recomendaciones a los usuarios en función de los lugares que visitan con frecuencia o de las personas que están socialmente conectadas con ellos. Estos servicios de recomendación utilizan información de registro para analizar los patrones de los usuarios y proporcionan ubicaciones interesantes y atractivas para diferentes usuarios. Las aplicaciones

de recomendación de POI tienen como objetivo proporcionar recomendaciones personalizadas de lugares de interés a los usuarios para enriquecer sus experiencias.

Objetivo: El objetivo de la investigación es recomendar topn ubicaciones mediante la aplicación de diversas técnicas de factorización matricial probabilística.

Metodología: La factorización matricial es una técnica colaborativa basada en modelos para recomendar nuevos elementos a los usuarios. Se ha comprobado que estas técnicas son eficaces y más precisas, ya que se basan en el descubrimiento de las características latentes que subyacen a la matriz de calificación del usuario-elemento. En LBSN, la matriz de calificación es generalmente escasa, por lo que el concepto de factorización de la matriz se utiliza para generar recomendaciones.

Resultados: Los resultados experimentales en dos LBSN del mundo real mostraron que PFM supera consistentemente a PMF. Esto se debe a que la técnica se basa en la distribución gamma para modelar la matriz de usuario y artículo. El uso de la distribución gamma es razonable para las frecuencias de registro que son todas positivas en conjuntos de datos reales. Sin embargo, PMF se basa en una distribución gaussiana que también puede permitir valores de frecuencia negativos.

Conclusión: El motivo del trabajo es identificar la mejor técnica para recomendar ubicaciones con la mayor precisión y permitir a los usuarios elegir entre una gran cantidad de ubicaciones disponibles; la mejor e interesante ubicación según el perfil de la persona.

Originalidad: se ha realizado un análisis riguroso de las técnicas de factorización de matrices probabilísticas en LBSN populares y se ha identificado la mejor técnica para la recomendación de ubicación comparando la precisión, a saber, RMSE, Precision @ N, Recall @ N, F1 @ N de diferentes modelos.

Limitaciones: la información contextual del usuario, como las preferencias demográficas, sociales y geográficas, no se ha tenido en cuenta al evaluar la eficacia de las técnicas de factorización matricial probabilística para las recomendaciones de puntos de interés.

Palabras clave: filtrado colaborativo, filtrado de información, LBSN, factorización matricial, PDI

Resumo

Introdução: O presente artigo é o resultado da pesquisa "Recomendação de Locus usando Técnicas de Fatoração de Matrizes Probabilísticas" realizada no Instituto Internacional de Pesquisa e Estudos Manav Rachna, Índia no ano de 2019-20.

Problema: Os sistemas de recomendação com base em localização (LSBNs) são comuns hoje em dia devido à crescente difusão dos dispositivos móveis. Eles fornecem recomendações aos usuários com base nos locais que são frequentemente visitados por eles ou pelas pessoas que estão socialmente conectadas a eles. Esses serviços de recomendação usam informações de check-in para extrair os padrões dos usuários e fornecer locais interessantes e atraentes para diferentes usuários. Os aplicativos de recomendação de POI visam fornecer recomendações personalizadas de locais de interesse para os usuários para enriquecer suas experiências.

Objetivo: O objetivo da pesquisa é recomendar locais de topo aplicando várias técnicas de fatoração de matrizes probabilísticas.

Metodologia: A fatoração de matrizes é uma técnica colaborativa baseada em modelo para recomendar novos itens aos usuários. Essas técnicas são consideradas eficazes e mais precisas, pois se baseiam na descoberta das características latentes subjacentes à matriz de classificação de itens do usuário. No LBSN, a matriz de classificação é geralmente esparsa, portanto, o conceito de fatoração da matriz é usado para gerar recomendações.

Resultados: os resultados experimentais em dois LBSNs do mundo real mostraram que o PFM supera consistentemente o PMF. Isso ocorre porque a técnica é baseada na distribuição gama para modelar a matriz do usuário e do item. Usar a distribuição gama é razoável para frequências de check-in que são todas positivas em conjuntos de dados reais. No entanto, o PMF é baseado na distribuição gaussiana que também pode permitir valores de frequência negativos.

Conclusão: O motivo do trabalho é identificar a melhor técnica para recomendar locais com a maior precisão e permitir que os usuários escolham entre uma infinidade de locais disponíveis; a melhor e interessante localização com base no perfil do indivíduo.

Originalidade: uma análise rigorosa das técnicas de Fatoração de Matriz Probabilística foi realizada em LBSNs populares e a melhor técnica para recomendação de localização foi identificada comparando a precisão viz RMSE, Precision @ N, Recall @ N, F1 @ N de modelos diferentes.

Limitações: as informações contextuais do usuário, como dados demográficos, preferências sociais e geográficas não foram consideradas ao avaliar a eficiência das técnicas de fatoração de matriz probabilística para recomendações de POI.

Palavras-chave: Filtragem colaborativa, Filtragem de informação, LBSN, Fatoração de matriz, POI

1. Introduction

With an increased development/advent in technology, location based social networks (LBSN), in particular Foursquare, Gowalla, Yelp, and Facebook Places are in much use these days. Through LBSN, users can publicize their geographical information and sentiments about various locations they have visited or in which they have an interest. The shared check-in information will help other users to identify popular locations and services such as restaurants and malls through POI recommendation; an important service of LBSN benefitting both LBSN users and POI owners. With mobile technology, users can identify their favorite POIs and can have an essence of improved user experience through accurate POI recommendations. POI owners can also be leveraged with POI recommendations to approach targeted customers.

An LBSN includes a geo-localized dimension to online social networks. The location dimension is now important and is the link between the real and the electronic world. Social and locative technology helps people to explore interesting activities around their arena by information drawn from their location and geocoded data [1]. The addition of location dimension generates a wide variety of relations between different users and venues: User-User connection, showing that two users are related to each other either because they are friends or the venues they have visited are similar; the connection between User-Place mentioning the spot that a particular user has visited; the Place-Place connection, which shows the relation between two places either due to their closeness or due to the same category they belong to. This is illustrated in Figure 1.

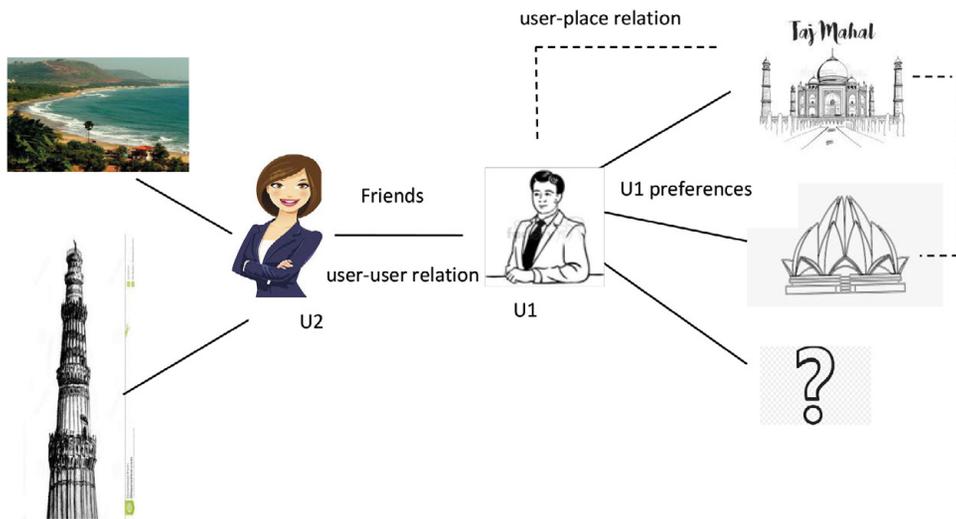


Figure 1. A typical LBSN
Source: own work

Traditional Recommender techniques have been endorsed by e-commerce markets, such as Amazon, Netflix, Flipkart, etc., however, the new user locale aspect in online social networks created tremendous gateways and challenges for existing recommendation systems. Due to this additional dimension, new approaches in recommender systems need to be developed to handle large heterogeneous data being continuously generated. The approaches may be based on different data sources and methodologies for enhancing recommendations. The following unique characteristics have to be considered when developing POI Recommender systems:

Geographical Influence: As mentioned by Tobler, the First Law of Geography defines that “Everything is related to everything else, but near things are more related than distant things” [2]. This law applies to LBSNs also as users tend to go to nearby locations more than distant ones. Moreover, geographically closer POIs might have similar features. Geographical Influence thus plays a crucial role in POI recommender systems.

Social Influence: Standard recommendation methods like collaborative filtering (CF), content-based recommendation (CB), and hybrid methods process information acquired from users' ratings and item characteristics and generate a list of recommendations based on these. However, in social networks, a users' behavior and their friends' check-ins might together decide the recommended list for a user.

Frequency Data and Sparsity: In traditional recommender systems, users provide explicit ratings to products (e.g., mobile phones, sports items, music, and so on), in the range [1-5]. Higher rating values for an item imply users liked the product more and also indicate user preferences. Unlike traditional recommender systems, in LBSN, a user's choices are measured by the check-in frequency for locations. A user-location rating matrix is then created (rating values here represent how many times a user has been to a particular location). The available frequency data is so rich and voluminous that choosing suitable ones can be overwhelming. Also, the check-in matrix is generally sparser than the user-item rating matrix and is a great barrier for POI recommendation. For example, the Netflix data set is around 99% sparse, whereas Gowalla is about 2.08×10^{-4} .

2. Literature review

In a recommender system, one of the significant elements that affect the quality of the recommendation is the algorithm applied for making recommendations. Popular recommendation algorithms are content-based (CB) methods [3], collaborative filtering [4,5,6,7] (CF) methods and Hybrid methods.

Content Based Recommendation: CB filtering recommends and suggests items using the items liked by users in past. Prediction is done by building users' and items' profile and finding items with similar content. [8] examined the combined impact of POI-related contents (e.g., category) and user opinions about a location (e.g., user comments) for POI recommendation. Their results demonstrated that content information significantly contributed to user behavior estimation. The SEAL (Sentiment-Enhanced Location search) system proposed by [9] is a fine-grained location search framework that is based on fine-grained user preferences. Their work emphasized collecting user feedback on venues and modeling user preferences by these fine-grained feedbacks. They used tensor factorization to study the positive and negative preferences of a user for personalized location ranking. Content-based recommendation systems, such as the one proposed by [10], match user preferences, discovered from users' profiles, with various location features, such as tags and categories to generate personalized recommendations displayed on a mini-map. Recent research explored the use of POI content information to reduce data sparsity. The Spatial topic model for POI recommendation, as suggested by [11], considered the spatial and textual features of user posts collected from Twitter. [12] focused on personal interests and local preferences associated with a POI's

content. [13] combined LDA and matrix factorization to study how POI associated tags affect venue recommendation. [14] proposed POI recommendation by using sentiment information and their technique demonstrated improved performance over the state-of-the-art approaches. [15] argued that content information, when combined with social correlations can generate effective recommendations and suggested a topic model based on this for offering better recommendations. A study by [16] considers location content information with its spatio-temporal patterns for recommending interesting locations. Content based technique is robust against the cold start issue for both new users and locations. These techniques work effectively so long as the newly added user or location has a description of the content. Most of the time the data utilized to generate recommendations is hard to capture and phony; thus, content-based systems have many limitations, such as the user might not be provided with novel recommendations as items are suggested based on the user profile. Moreover, due to limited user content, recommendations are not appropriate.

Collaborative Filtering Techniques: CF algorithms are the most accepted algorithms and are one of the influential algorithms in recommendation systems. Unlike CB methods, CF techniques do not need a description of items or users. Recommendations are generated by the opinions of a user's comparative taste and preferences to the active user. This section provides a literature review about significant POI recommender systems, mainly for collaborative filtering. Existing literature mentioned two classes of collaborative filtering strategies, memory-based and model-based [17].

Memory Based Collaborative Filtering: The memory-based CF computes similarities between the active user and other users (User-User Collaborative Filtering), or between items (item-item approach) to make recommendations. [18] studied the combined effect of social and geographical influences and incorporated both the features into the user-based CF framework for POI recommendation. Their experiment proved that user-oriented CF surpasses item-oriented CF for POI recommendation if geographical information is associated with a user-based CF model. However, adding social information brought little improvement in the model's performance. [19] extended upon the item-based CF method. However, calculating similarity is rather more difficult in memory-based techniques and the model's performance is confined by data sparsity. To deal with the aforementioned issues, model-based recommendation algorithms such as "latent semantic models" [20], "Bayesian models" [21], "clustering models" [22,23], and "matrix factorization models" [24,25,26] have been suggested and worked upon.

Model Based Techniques: Among the several CF technologies, **matrix factorization** is well known and great at managing high dimensional data and mitigating sparsity. According to [24], in this method, item and user vectors are defined in terms of vectors of factors inferred from item rating patterns. To reduce the data sparsity issue, auxiliary information such as time or social trust can be used with existing techniques. [27] proposed to combine the Matrix Factorization technique with the geographical and social influence for POI recommendation. Authors have modeled the geographical influence via the multicenter Gaussian model. The assumption is that a user's check-ins are generally located around multi centers. To find these multi centers, the authors have proposed a greedy clustering algorithm. Their model does not consider the extremely sparse frequency data. Further, there were problems in finding centers accurately and the outliers' effect was not dealt with. To overcome this problem, [28] presented a genetic algorithm-based Gaussian mixture model. Their model employed a genetic-based EM algorithm [29] to dissolve the effect of outliers in the mixture model. [30] utilized the mobility records of a user as implicit feedback and proposed a weighted matrix factorization for recommendation. They also added a spatial clustering phenomenon with matrix factorization. [31] analyzed the integrated effect of user preferences, geographical influences, and user mobility behavior. They predicted user preferences by combining probabilistic matrix factorization with a Poisson factor model. The STPMF (Spatial-Temporal Probabilistic Matrix Factorization) model developed by [43] captured users general preferences and multiple geographical features into PMF. [41] were the first ones to develop a model that learns the distribution of user preferences in a generative way. Their model considers two neural network components; one that recommends POI and another component judges if it is a true recommendation or not and helps in optimizing the proposed recommendation. The STACP (Spatio-Temporal Activity Center POI) novel recommendation model proposed by [42] considers the effect of spatial and temporal characteristics of a user jointly. This model trains the matrix factorization model in a static and temporal manner and forms spatio-temporal activity centers for users. [44] suggested a generic bayesian model that integrates the item content and social information using poisson matrix factorization to produce highly accurate recommendations.

3. Methodology

Collaborative Filtering experiences data sparsity and scalability issues as the number of users and items grow. Matrix factorization is a powerful solution to tackle and mitigate data sparsity and also reduce data dimensions, extract hidden features from data. Thus, matrix factorization is widely used in recommender systems owing to their characteristics. As a case study, two probabilistic matrix factorization techniques are trained on the Gowalla and Foursquare check-in data set with the aim of studying user mobility behaviour. User mobility patterns indicate the places a user is interested in visiting. This helps in identifying and recommending new locations to the user. The methodology to compute recommendations consisted of the following steps:

- Collect the Ratings of users on POIs.
- Find the relevant data.
- Compute the recommendations using the model-based techniques.
- Present the data.

3.1. Matrix Factorization Framework

Matrix factorization (MF) is a systematic Collaborative Filtering method. Given an $m \times n$ user-location rating matrix r to describe m users' frequency count on n POIs, the MF model learns an $m \times d$ user-latent feature matrix and a $d \times n$ item-latent feature matrix. Mathematically, it is represented as:

$$r_{m \times n} \approx u_{m \times d} I_{d \times n} \quad (1)$$

where,

- $r_{m \times n}$ is user-item rating matrix with m users and n locations,
- $u_{m \times d}$ is user-feature matrix with m users and d features, and
- $I_{d \times n}$ is location-feature matrix with n locations and d features

Factorization is performed in a manner that retains the properties and dependencies of the rating matrix. Matrix factorization is an optimization problem in which latent vectors are learned using the function mentioned in equation 2 on the set of available ratings:

$$\frac{1}{2} \sum_{i,j} (r_{ij} - \sum_{k=1}^d u_{ik} l_{kj}) + \frac{\lambda}{2} (\|u\|^2) + \frac{\lambda}{2} (\|l\|^2) \tag{2}$$

where,

$\lambda/2(\|u\|^2)$ and $\lambda/2(\|l\|^2)$ are L2 regularization terms that help to prevent over-fitting,
 r_{ij} is rating by user i to item j ,

u_{ik} and l_{kj} are instances of user-feature and location-feature matrix.

The MF framework is described in Figure 2.

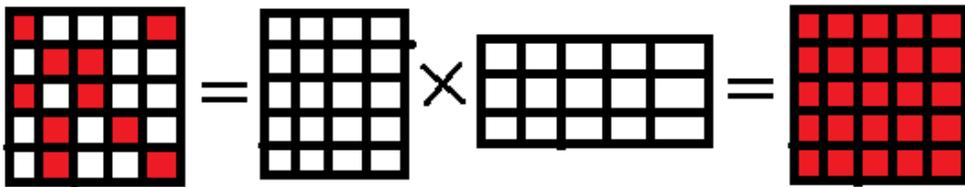


Figure 2. Matrix Factorization Framework
 Source: own work

Some of the Matrix Factorization models as depicted in literature are Singular Value Decomposition [32, 33], Probabilistic Matrix Factorization (PMF) [25, 34], and Non-Negative Matrix Factorization [35].

In this section we will discuss two latent factor models:

3.1.1. Probabilistic Matrix Factorization (PMF)

PMF is a robust statistical framework with a Bayesian perspective on the model matrix R . In this model, entries of R are assumed to be normally distributed around the inner product $\langle u_i, l_j \rangle$, through a common variance. Letting I_{ij} be 1 if the entry was observed and 0 if the data value was not present, we can write the likelihood of the entries of R as follows.

$$p(R|U, L, \sigma^2) = \prod_{i=1}^m \prod_{j=1}^n [N(R_{ij}|u_i^T l_j, \sigma^2)]^{I_{ij}} \tag{3}$$

where,

$N(x|\mu, \sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2 , and

I_{ij} is the indicator function that is equal to 1 if user i rated item j and equal to 0 otherwise.

The primary assumptions of this likelihood are as follows:

the entries of R are independent

each entry is normally distributed, and

the entries share a common variance σ^2 . These assumptions may or may not be appropriate for certain applications and will need to be considered more closely in practice.

To explain the full Bayesian model, prior distributions in the matrices U and L have been used with the following form.

$$p(U|\sigma_u^2) = \prod_{i=1}^m N(u_i | 0, \sigma_u^2 I) \quad p(L|\sigma_l^2) = \prod_{j=1}^n N(l_j | 0, \sigma_l^2 I) \quad (4)$$

In these priors, the following are the assumptions:

- a) The rows of U and L are uncorrelated,
- b) are normally distributed, and
- c) have common variance.

With additional prior information, more informative prior distributions can be constructed to have an in-depth correlation between observations or features in the design matrix R .

3.1.2. Probabilistic/Poisson Factor Model (PFM)

Poisson factorization is one of the probabilistic models of users and items. Each observation of the user-POI rating matrix is assumed to be taken from a Poisson distribution; "an exponential family of distribution over positive integers whose parameter is a linear combination of the corresponding user preferences and item attributes". In [36, 37, 38], in contrast to classical matrix factorization in which Gaussian Distribution is used to generate both positive and negative examples, the user and item matrices in PFM are generated from the gamma distribution as it is not feasible to have negative frequency visit counts to a place in the real world. The Poisson factor model segments the user-POI check-in count matrix Y as $Y \sim \text{Poisson}(U \cdot L)$. That is, for each user-POI response y_{ij} , Poisson distribution is assumed over the mean f_{ij} : $y_{ij} \sim \text{Poisson}(f_{ij})$. The mean matrix F is divided into two matrices U of $M \times K$ and L of $N \times K$. Each element u_{ik} of $U_{m \times k}$ indicates the preference of user i for "feature" k ,

and each element l_{ik} of L reflects the affinity of item j to feature k . In addition, empirical Gamma Distribution priors are placed with u_{ik} and l_{jk} priors. The steps are defined as:

1. Generate user latent factor u_{ik} from Gamma Distribution.
2. Generate item latent factor l_{jk} from Gamma Distribution
3. Generate y_{ij} from Poisson distribution.

4. Results

The performance of the two approaches has been experimentally verified. All the approaches have been implemented in Python. The techniques have been applied to the real world Gowalla Dataset and Foursquare Dataset [39]. The Gowalla check-ins were captured for a period from February 2009 to October 2010. The records of users with less than 15 check-ins and POIs with less than 10 visitors are filtered. The new processed data has 18,737 users, 32,510 POIs and 1,278,274 check-ins, resulting in a matrix that is 99.86% sparse.

The Foursquare data contains user check-ins for a period of 17 months from April 2012 to September 2013. These check-ins are generated from within the United States (except Alaska and Hawaii). For this dataset also, records of users with less than 10 check-in POIs, as well as those POIs with less than 10 visitors are eliminated. The processed and cleaned dataset has 24,941 users, 28,593 POIs and 1,196,248 check-ins. The sparsity of user-POI check-in matrix is 99.900%.

A snippet of Gowalla and Foursquare datasets is displayed in Figure 3.

0	0	1287503727-0
0	1	1287411463-0
0	2	1287330123-0
0	3	1287314765-0
0	4	1287226242-0
0	5	1286899083-0
0	6	1286892131-0
0	1	1286883880-0
0	7	1286870240-0

Figure 3. user check-ins

Source: own work

Each Dataset has three columns showing user-id, POI-id, timestamp of check-in made. For this study, we used 70% data as training set and 30% as test set.

4.1. Evaluation of Probabilistic Techniques

The comparison of two techniques: PMF and PFM in terms of evaluation metrics is performed in this section. It compares two traditional approaches on two real LBSNs Gowalla and Foursquare. A statistical analysis of the performance evaluation metrics such as Precision@N, Recall@N, and RMSE and F1@N is performed.

RMSE

“Root mean square error computes the mean value of all the differences squared between the true and the predicted ratings and then proceeds to calculate the square root out of the result” [40]. It is calculated as:

$$\sqrt{\frac{\sum_{i=1}^n (r-r')^2}{N}} \quad (5)$$

where, r is observed rating, and r' is the actual rating.

The RMSE has the same unit of the variable r .

The RMSE graph of PMF for Gowalla and Foursquare Dataset is shown in Figure 4 and Figure 5 respectively.

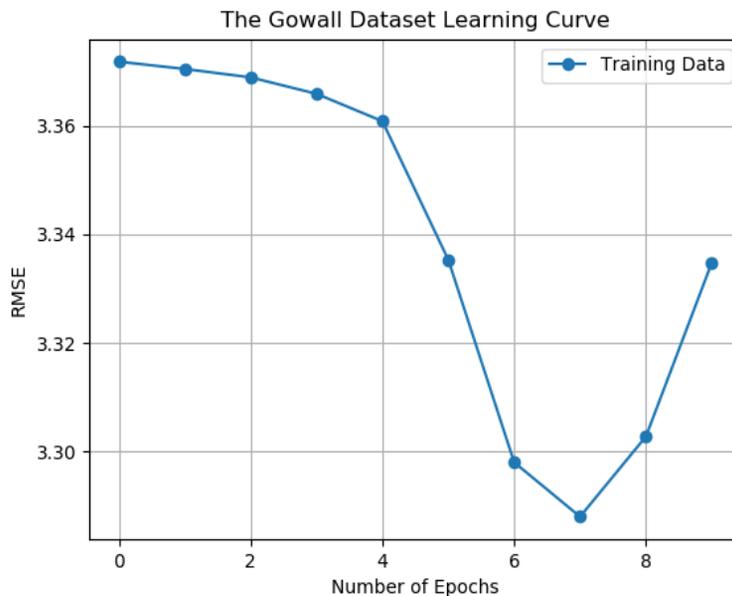


Figure 4. RMSE Graph-PMF(Gowalla)
Source: own work

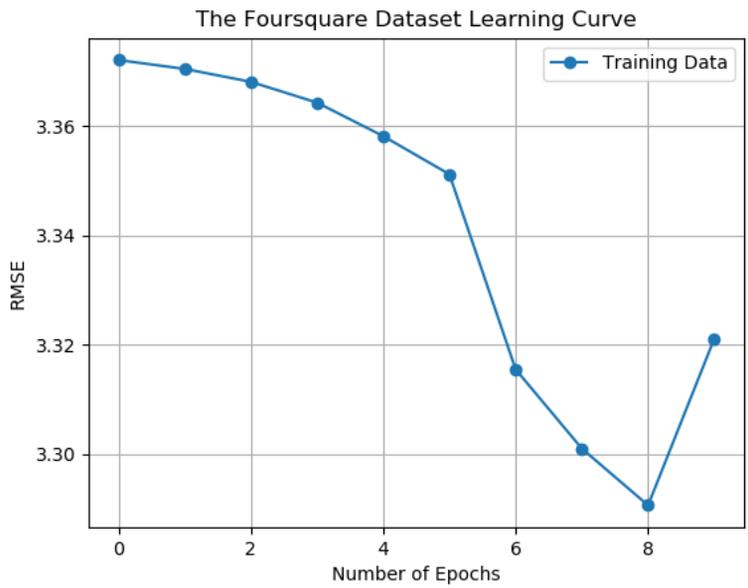


Figure 5. RMSE Graph-PMF(Foursquare)
Source: own work

Figure 6 and Figure 7 display the behaviour of RMSE for Gowalla and Foursquare using PFM.

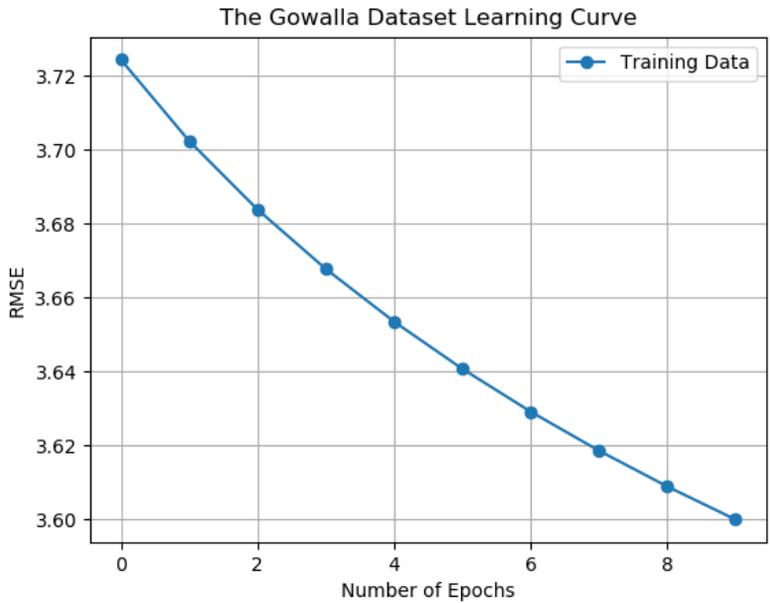


Figure 6. RMSE Graph-PFM(Gowalla)
Source: own work

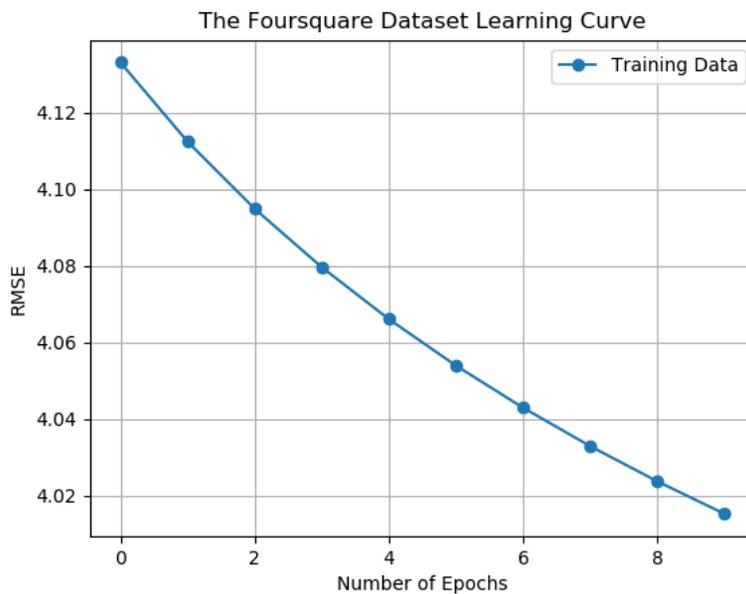


Figure 7. RMSE Graph-PFM(Foursquare)
Source: own work

Values of RMSE for PMF over Gowalla Dataset shows a declining trend with values ranging from 3.38 to 3.29 as the number of epochs increases from 0 to 7. At the 7th iteration, RMSE increases from 3.29 to around 3.34. Similar trends are exhibited in Foursquare Dataset with values decreasing from 3.38 to 3.32 and increasing at 8th epoch from 3.27 to 3.32. However, for PFM RMSE is always decreasing with the number of epochs and RMSE value is in the range from 3.72-3.60 for Gowalla and between 4.13-4.01 for Foursquare dataset.

Precision@N

Precision and recall are traditional evaluation metrics in machine learning algorithms and for document retrieval tasks. In the recommender system's context, we most likely wanted to recommend top-N POIs to the user. It makes more sense to compute precision and recall metrics in the first N POIs instead of all the POIs. Thus, the concept of *Precision@N* and *Recall@N* is used where N is a user-defined integer, set by the user to satisfy the top-N recommendations objective. It is calculated as:

$$\text{Precision@N} = (\# \text{ of recommended POIs @N that are relevant}) / (\# \text{ of recommended POIs @N})$$

Table 1 and Table 2 shows the results obtained for Precision@N for the two techniques when N is varied between 5, 10, 15 and 20, and their comparison is made using graphs shown in Figure 8 and Figure 9 respectively.

Table 1. Precision Comparison for Gowalla

	Precision@5	Precision@10	Precision@15	Precision@20
PMF	9.60e-4	1.45e-3	1.47e-3	1.65e-3
PFM	2.20e-2	2.31e-2	1.86e-2	1.61e-2

Source: own work

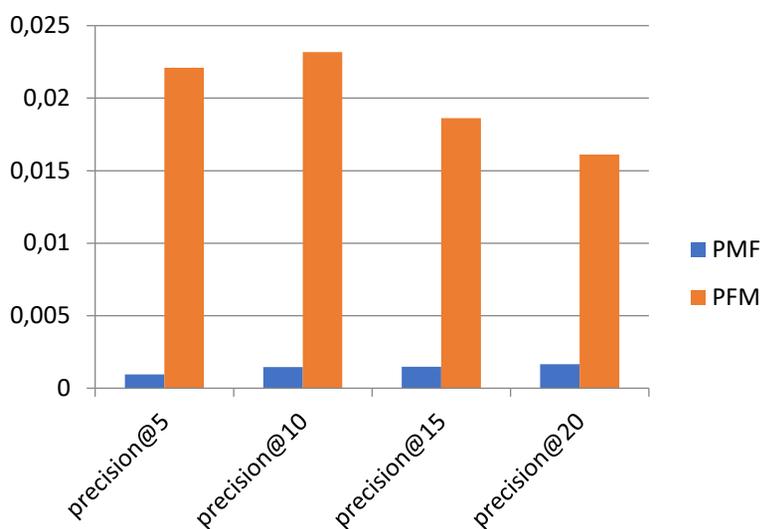


Figure 8. Precision@N Comparison Graph(Gowalla)

Source: own work

Table 2. Precision Comparison for Foursquare

	Precision@5	Precision@10	Precision@15	Precision@20
PMF	9.70e-4	1.33e-3	1.35e-3	1.62e-3
PFM	2.52e-2	2.10e-2	1.76e-2	1.55e-2

Source: own work

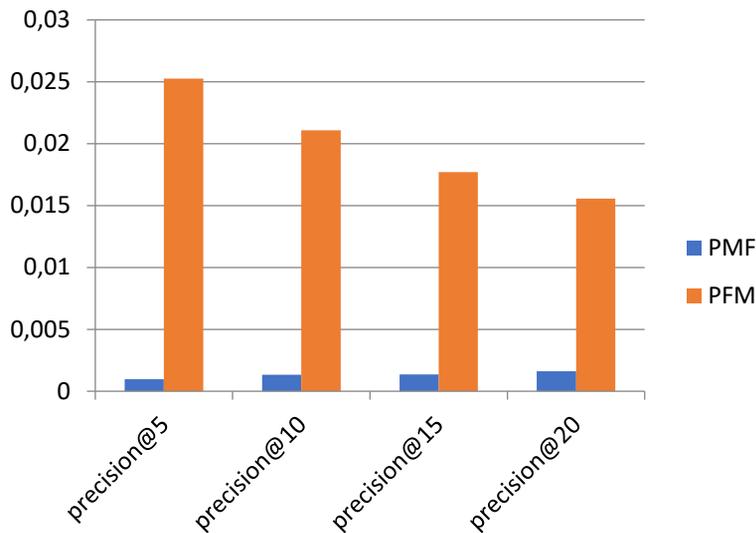


Figure 9. Precision@N Comparison Graph (Foursquare)
Source: own work

Precision@N drops gently with increasing N in both the datasets. PMF is reported to have lower Precision@N values. We have taken N values as 5, 10, 15 and 20 and Precision@N for Gowalla data set for different values of N is reported as $9.60e-4$, $1.45e-3$, $1.47e-3$, $1.65e-3$ respectively. Whereas, for Foursquare data set these values are $9.70e-4$, $1.33e-3$, $1.35e-3$, $1.62e-3$. However, for PFM Precision@N is reported as $2.20e-2$, $2.31e-2$, $1.86e-2$, $1.61e-2$ for different N respectively for Gowalla and a similar trend in values is observed for Foursquare dataset also.

Recall@N

Recall measure relevancy of POIs retrieved from all the predictions.

$$\text{Recall@N} = (\# \text{ of recommended POIs @N that are relevant}) / (\text{total } \# \text{ of relevant POIs})$$

Table 3 and Table 4 show the recall values of both the techniques for Gowalla and Foursquare and related comparison graphs are shown in Figure 10 and Figure 11.

Table 3. Recall Comparison for Gowalla

	Recall@5	Recall@10	Recall@15	Recall@20
PMF	$2.93e-4$	$9.91e-4$	$1.51e-3$	$2.43e-3$
PFM	$9.40e-3$	$2.04e-2$	$2.48e-2$	$2.86e-2$

Source: own work

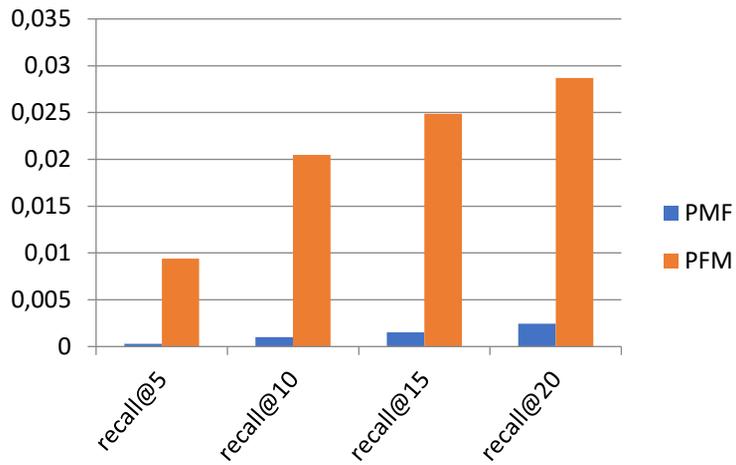


Figure 10. Recall@N Comparison Graph (Gowalla)
Source: own work

Table 4. Recall Comparison for Foursquare

	Recall@5	Recall@10	Recall@15	Recall@20
PMF	2.79e-4	9.71e-4	1.40e-3	2.31e-3
PFM	9.39e-3	1.91e-2	2.52e-2	2.89e-2

Source: own work

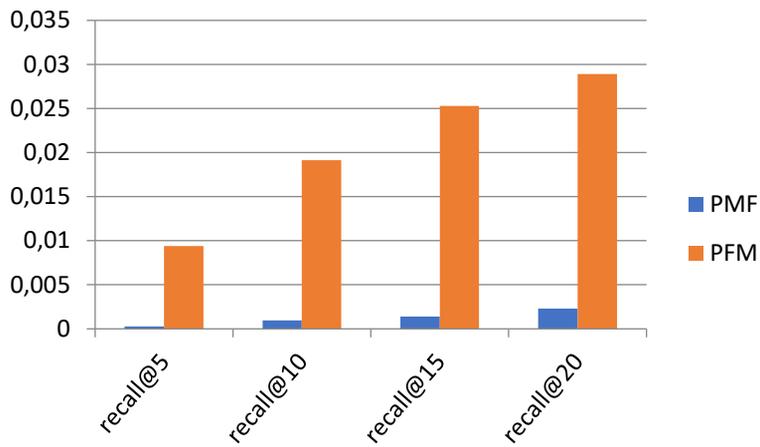


Figure 11. Recall@N Comparison Graph (Foursquare)
Source: own work

Recall@N increases gradually as N increases from 5 to 20, on the other hand, values for PMF are small when compared with PFM for two datasets. Recall values for two techniques for different N values show an increasing trend. Compared to PFM, the PMF technique has lower Recall@N. As N varies Recall@N for PMF is 2.93e-4, 9.91e-4, 1.51e-3, 2.43e-3 for Gowalla and 2.79e-4, 9.71e-4, 1.40e-3 and 2.31e-3 for Foursquare respectively. For Gowalla dataset Recall@N using PFM is 9.40e-3, 2.04e-2, 2.48e-2, 2.86e-2 respectively.

F1@N

It is computed as the harmonic mean of precision and recall.

$$F1@N = 2 / \left(\frac{1}{Precision@N} + \frac{1}{Recall@N} \right) \quad (5)$$

Table 5, Table 6 and Table 7 show the precision, recall and F1 values of PMF and PFM for Gowalla and Foursquare Dataset.

Table 5: Comparative Evaluation of Precision@N for Gowalla and Foursquare DataSet

Dataset	Method	Precision@5	Precision@10	Precision@15	Precision@20
Gowalla	PMF	9.60e-4	1.45e-3	1.47e-3	1.65e-3
	PFM	2.20e-2	2.31e-2	1.86e-2	1.61e-2
Foursquare	PMF	9.70e-4	1.33e-3	1.35e-3	1.62e-3
	PFM	2.52e-2	2.10e-2	1.76e-2	1.55e-2

Table 6. Comparative Evaluation of Rrecision@N for Gowalla and Foursquare DataSet

Dataset	Method	Recall@5	Recall@10	Recall@15	Recall@20
Gowalla	PMF	2.93e-4	9.91e-4	1.51e-3	2.43e-3
	PFM	9.40e-3	2.04e-2	2.48e-2	2.86e-2
Foursquare	PMF	2.79e-4	9.71e-4	1.40e-3	2.31e-3
	PFM	9.39e-3	1.91e-2	2.52e-2	2.89e-2

Table 7. Comparative Evaluation of F1@N for Gowalla and Foursquare DataSet

Dataset	Method	F1@5	F1@10	F1@15	F1@20
Gowalla	PMF	4.49e-4	1.17e-3	1.49e-3	1.97e-3
	PFM	1.31e-2	2.17e-2	2.12e-2	2.06e-2
Foursquare	PMF	4.33e-4	1.12e-3	1.37e-3	1.90e-3
	PFM	1.36e-2	2.00e-2	2.08e-2	2.02e-2

In summary, from the experimental investigation over Gowalla and Foursquare Data set, it can be concluded that the PFM exhibits unvarying performance and is more effective over real-world datasets. Moreover, PFM is much better, more effective, and provides more accurate top-N recommendations in contrast to PMF.

5. Discussion and conclusion

As location-based social networks are gaining prevalence these days, personalized POI recommendation services are boosting up and procuring the attention of industries and scholars. In addition, to help users explore new places, these services are a boon to LBSN providers and helping them to increase their profits. In this paper, we illustrate the task of POI recommendation in LBSNs using matrix factorization. We first describe the unique characteristics of POI recommender systems that distinguish them from traditional recommender systems. We have investigated in detail two traditional probabilistic techniques for the POI recommendation problem. Accuracy comparisons are presented and experimental results on Gowalla and Foursquare datasets indicate that PFM outperforms PMF in accuracy. From the results it can be inferred that PFM is best for the check-in datasets as in real scenarios, check-in frequencies cannot be negative.

Our current and future research plans are inclined to improve the performance of the POI recommendation by exploring more influential factors, and its applicability under realistic scenarios. Specifically, in future work, we plan to extend this work by considering social factors and also by adding time factors that can reflect different contexts about user interests.

6. References

1. R. Wilken, "Places nearby: Facebook as a location-based social media platform," *New Media & Society*, vol.16, no. 7, 1087-1103, 2014.
2. W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic geography*, vol. 46, sup.1, 1970, 234-240.
3. M. J. Pazzani, D. Billsus, "Content-Based Recommendation Systems," in Brusilovsky P., Kobsa A., Nejdl W. (eds) *The Adaptive Web*. Lecture Notes in Computer Science, vol. 4321, 2007, Springer, Berlin, Heidelberg, doi: https://doi.org/10.1007/978-3-540-72079-9_10.
4. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, June 2005, doi: 10.1109/TKDE.2005.99.
5. Z. Huang, D. Zeng and H. Chen, "A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce," *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 68-78, Sept.-Oct. 2007, doi: [10.1109/MIS.2007.4338497](https://doi.org/10.1109/MIS.2007.4338497).
6. X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.* 2009, Article 4, January 2009, doi: <https://doi.org/10.1155/2009/421425>.
7. S. Vucetic and Z. Obradovic, "Collaborative filtering using a regression-based approach," *Knowledge and Information Systems*, vol. 7, no. 1, 1-22, 2005.
8. H. Gao, J. Tang, X. Hu, and H. Liu, "Content-aware point of interest recommendation on location-based social networks," *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pp. 1721-1727, Austin, TX, USA, January 2015.
9. D. Yang, D. Zhang, Z. Yu, and Z. Yu, "Fine-grained preference-aware location search leveraging crowdsourced digital footprints from LBSNs," *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*, Association for Computing Machinery, New York, NY, USA, 479-488, 2013, doi: <https://doi.org/10.1145/2493432.2493464>.
10. M. H. Park, J. H. Hong, S.B. Cho, "Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices," in Indulska J., Ma J., Yang L.T., Ungerer T., Cao J. (eds) *Ubiquitous Intelligence and Computing. UIC 2007*. Lecture Notes in Computer Science, 2007, vol 4611. Springer, Berlin, Heidelberg, doi: https://doi.org/10.1007/978-3-540-73549-6_110.

11. B. Hu, and M. Ester, "Spatial topic modeling in online social media for location recommendation," in *RecSys '13: Proceedings of the 7th ACM conference on Recommender systems*, pp. 25-32.
12. H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen, "LCARS: a location-content-aware recommender system," *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13)*, Association for Computing Machinery, New York, NY, USA, 221–229, 2013, doi: <https://doi.org/10.1145/2487575.2487608>.
13. B. Liu, and H. Xiong, "Point-of-interest recommendation in location based social networks with topic and location awareness," *Proceedings of the 2013 SIAM international conference on data mining*, pp. 396-404, May 2013, Society for Industrial and Applied Mathematics.
14. D. Yang, D. Zhang, Z. Yu and Z. Wang, "A sentiment-enhanced personalized location recommendation system," *Proceedings of the 24th ACM conference on hypertext and social media*, pp. 119-128, 2013.
15. B. Hu and M. Ester, "Social Topic Modeling for Point-of-Interest Recommendation in Location-Based Social Networks," *2014 IEEE International Conference on Data Mining*, Shenzhen, 2014, pp. 845-850, doi: 10.1109/ICDM.2014.124.
16. Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat- Thalmann, "Time-aware point-of-interest recommendation" *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '13)*, Association for Computing Machinery, New York, NY, USA, 363–372, 2013, doi: <https://doi.org/10.1145/2484028.2484030>.
17. J. S. Breese, D. Heckerman, C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Microsoft Research Microsoft Corporation One Microsoft Way Redmond, WA, 98052*, 1998.
18. M. Ye, P. Yin, W. C. Lee and D. L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (SIGIR '11)*, Association for Computing Machinery, New York, NY, USA, 325–334, 2011, doi: <https://doi.org/10.1145/2009916.2009962>.
19. J. J. Levandoski, M. Sarwat, A. Eldawy and M. F. Mokbel, "LARS: A Location-Aware Recommender System," *2012 IEEE 28th International Conference on Data Engineering*, Washington, DC, 2012, pp. 450-461, doi: [10.1109/ICDE.2012.54](https://doi.org/10.1109/ICDE.2012.54).
20. T. Hofmann, and J. Puzicha, "Latent class models for collaborative filtering," *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI*, vol. 2, pp. 688-693.

21. K. Miyahara and M. J. Pazzani, "Collaborative filtering with the simple bayesian classifier", in *Pacific Rim International Conference on Artificial Intelligence*, pp. 679-689, 2000, Springer, Berlin, Heidelberg.
22. Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting Geographical Neighborhood Characteristics for Location Recommendation," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14)*. Association for Computing Machinery, New York, NY, USA, 739–748, doi: <https://doi.org/10.1145/2661829.2662002>.
23. J. Zhu, C. Wang, X. Guo, Q. Ming, J. Li, and Y. Liu, "Friend and POI recommendation based on social trust cluster in location-based social networks", *EURASIP Journal on Wireless Communications and Networking*, 2019, no. 1, 89.
24. Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," in *Computer*, vol. 42, no. 8, pp. 30-37, Aug. 2009, doi: [10.1109/MC.2009.263](https://doi.org/10.1109/MC.2009.263).
25. R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization," *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07)*, Curran Associates Inc., Red Hook, NY, USA, 1257–1264, 2007.
26. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13*, pp. 556-562, 2001, Denver, CO, United States.
27. C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," *Proceedings of Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI)*, pp. 17–23, Toronto, ON, Canada, July 2012.
28. S. Zhao, I. King and M. R. Lyu, "Capturing geographical influence in POI recommendations," in *International Conference on Neural Information Processing*, pp. 530-537, 2013, Springer, Berlin, Heidelberg.
29. N. D. Thang, C. Lihui and C. C. Keong, "An outlier-aware data clustering algorithm in mixture models," *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)*, Macau, 2009, pp. 1-5, doi: [10.1109/ICICS.2009.5397571](https://doi.org/10.1109/ICICS.2009.5397571).
30. D. Lian., C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation," *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 831-840, August, 2014.

31. B. Liu, H. Xiong, S. Papadimitriou, Y. Fu and Z. Yao, "A General Geographical Probabilistic Factor Model for Point of Interest Recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 5, pp. 1167-1179, 1 May 2015, doi: 10.1109/TKDE.2014.2362525.
32. B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Application of dimensionality reduction in recommender system-a case study", WebKDD-2000, Minnesota Univ Minneapolis Dept of Computer Science.
33. H. Polat and W. Du, "SVD-based collaborative filtering with privacy," *Proceedings of the 2005 ACM symposium on Applied computing*, pp. 791-795.
34. H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: social recommendation using probabilistic matrix factorization", *Proceedings of the 17th ACM conference on Information and knowledge management (CIKM '08)*, Association for Computing Machinery, New York, NY, USA, 931-940, 2008, doi: <https://doi.org/10.1145/1458082.1458205>.
35. D. Cai, X. He, J. Han and T. S. Huang, "Graph Regularized Nonnegative Matrix Factorization for Data Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548-1560, Aug. 2011, doi: 10.1109/TPAMI.2010.231.
36. P. Gopalan, L. Charlin, D.M. Blei, "Content-based recommendations with poisson factorization," *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems*, December 2014, vol. 2, pp. 3176-3184, Montreal, Quebec, Canada.
37. P. Gopalan, J.M. Hofman, and D.M. Blei, "Scalable recommendation with poisson factorization," *arXiv preprint arXiv:1311.1704*, 2013.
38. H. Ma, C. Liu, I. King, and M. R. Lyu, "Probabilistic factor models for web site recommendation," *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (SIGIR '11)*, Association for Computing Machinery, New York, NY, USA, 265-274, July, 2011, doi: <https://doi.org/10.1145/2009916.2009955>.
39. Y. Liu, T. N. Pham, G. Cong, and Q. Yuan, "An experimental evaluation of point-of-interest recommendation in location-based social networks," *Proc. VLDB Endow.*, vol. 10, no. 10, pp. 1010-1021, June, 2017, doi: <https://doi.org/10.14778/3115404.3115407>.
40. T. Chai, and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?- Arguments against avoiding RMSE in the literature", *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247-1250, 2014, doi:10.5194/gmd-7-1247-2014.

41. Zhou, Fan, Ruiyang Yin, Kunpeng Zhang, Goce Trajcevski, Ting Zhong, and Jin Wu, "Adversarial point-of-interest recommendation", In *The World Wide Web Conference*, pp. 3462-34618, 2019.
42. H.A. Rahmani, M. Aliannejadi, M. Baratchi, F. Crestani, Joint Geographical and Temporal Modeling Based on Matrix Factorization for Point-of-Interest Recommendation, In: Jose J. et al. (eds) *Advances in Information Retrieval. ECIR 2020. Lecture Notes in Computer Science*, vol 12035. Springer, Cham. https://doi.org/10.1007/978-3-030-45439-5_14, 2020.
43. Li, Huayu, Richang Hong, Zhiang Wu, and Yong Ge. "A spatial-temporal probabilistic matrix factorization model for point-of-interest recommendation." *Proceedings of the 2016 SIAM international conference on data mining*, pp. 117-125. Society for Industrial and Applied Mathematics, 2016.
44. Eliezer de Souza, da Silva, H. Langseth, & H Ramampiaro, "Content-based social recommendation with poisson matrix factorization.", *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 530-546. Springer, Cham, 2017.