**An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks**

**ABSTRACT**

兴趣点（POI）推荐是基于位置的社交网络（LBSN）的重要服务，可以使用户和企业受益。近年来，已经提出了许多POI推荐系统，但是仍然缺乏系统的比较。在本文中，我们提供了对12种最先进的POI推荐模型的全面评估。从评估中，我们获得了几个重要的发现，在此基础上我们可以更好地理解和利用各种情景下的POI推荐模型。我们期待这项工作能够为读者提供有关POI推荐的前沿研究的整体情况。

1. **INTRODUCTION**

随着位置感知社交媒体的突出，人们可以轻松地共享与位置相关的内容。例如，Foursquare拥有超过5000万活跃用户，到2016年已经有超过80亿次登记到兴趣点(POI), Yelp拥有大约2100万用户和1.02亿有地理坐标的企业评论。

随着大量用户访问历史的可用性，POI推荐的问题已被广泛研究。已经发现，60％-80％的用户访问是在过去30天内没有访问过的POI [34]。POI建议可以极大地帮助用户找到他们感兴趣的新兴趣点，这对用户和企业都有益。但是，与其他推荐问题（例如产品，电影）相比，POI建议面临以下新挑战：

* **Rich contexts.** 首先，用户的移动性偏好受地理距离的影响：用户通常访问少数活动区域（例如，家附近或工作地点）内的POI。其次，用户可以每天访问相同的POI（例如，家庭，工作场所）。第三，用户的偏好取决于时间。 例如，用户很可能在清晨和深夜访问不同的地方。 第四，用户的访问偏好可能会受到社交关系的影响。 其他类型的上下文可能包括POI评论，POI社交帖子等。
* **Data scarcity problem.** 与其他推荐问题相比，POI建议遭遇更严重的数据稀缺问题。用户访问的POI数量通常只是所有POI的一小部分。例如，POI推荐的实验研究中使用的数据密度通常约为0.1％，而电影推荐的Netflix数据密度为1.2％[2]。

POI建议在过去五年中得到了广泛的研究关注，并提出了许多方法。这些研究在问题设置，推荐模型和评估数据方面有所不同。提出新提出的方法的论文经常报告实验研究，这些研究表明所提出的方法比某些数据集上的某些选定基线表现更好。然而，不清楚它们是否在不同类型的数据（例如，更稀疏）或不同类型的用户（例如，具有非常少的历史数据的用户）上表现更好。更糟糕的是，这些新提出的方法通常与使用类似框架的其他方法（例如，矩阵分解）进行比较，并且新方法在经验上不相互比较。

此外，这些提出的方法可以利用不同类型的上下文信息，并采用不同的框架来捕获用户偏好。对于利用相同类型的上下文信息或捕获用户对POI推荐的偏好的不同方法缺乏实证研究。这种状态使得难以确定哪种方法在特定环境中最合适。因此，显然需要一种基准，可以深入了解现有POI推荐方法的性能。

为了满足这一需求，我们设计了一个评估程序来评估12个代表性的POI推荐模型，包括最近提出的推荐模型，旨在从多个方面获得POI推荐模型的总体情况。具体来说，我们在不同来源和不同稀疏度的数据集以及具有不同大小的历史数据的用户上试验这些模型。该评估提供了关于这些POI推荐方法的相对价值以及这些模型的适用场景的新见解。我们还评估了POI推荐中用户偏好建模的不同推荐技术，例如矩阵分解，以及上下文信息的建模方法，例如地理上下文。此评估将提供有关哪种方法对每个组件表现更好的见解，以便在将来设计更准确的POI推荐方法。本文对12种代表性POI推荐模型进行了首次全方位评估。本文的其余部分安排如下：第2部分首先介绍了POI建议。随后，我们评估中的代表性模型被分类并在第3节中给出。第4节和第5节介绍了POI推荐模型的实验，从中发现了一些值得注意的发现。最后，我们将回顾第6节中的相关工作。

1. **POI RECOMMENDATION**

给定一组POI L，以及一组用户U，每个用户U与用户访问的一组POI Lu相关联，POI推荐的问题是为每个用户u∈U推荐新POI，即在集合L/ Lu中，很可能是用户u访问过的。POI建议受到诸如地理距离，社会关系和时间等丰富背景的显着影响。为了证明它们的影响，我们接下来展示了来自三个LBSN的数据集的一些统计分析结果，即Gowalla，Foursquare和Yelp。第4.1节介绍了这些数据的详细信息。

**Observation 1: spatial influence.** 我们将用户的连续签到视为POI之间的转换，并计算用户转换距离的分布。图1（a）显示了Gowalla，Foursquare和Yelp中过渡距离的累积分布函数（CDF）。我们可以看到，当距离很小时，所有三条曲线都会急剧上升。在Gowalla和Foursquare，90％的用户过渡距离不到50公里。这表明用户倾向于访问附近的POI。

**Observation 2: social influence.** 我们选择在德克萨斯州奥斯汀办理登机手续的Gowalla用户，并计算朋友之间以及随机抽样的非朋友之间的常见登记POI数量。对于朋友和非朋友，普通登记POI的平均数分别为5.69和0.91。特别地，图1（b）示出了共同登记POI的数量的CDF。我们可以看到超过60％的非朋友没有共同的登记POI，而朋友之间的数量只有16％。此外，大约85％的朋友拥有少于10个普通登记POI，超过80％的非朋友只有1个或没有登记POI。这些表明大多数朋友在他们的登记POI上有小的重叠，但重叠比非朋友大得多。

**Observation 3: temporal influence.** 一方面，两个用户在时间上可能表现不同。 例如，一个人经常在午餐时间检查餐馆，而另一个人喜欢酒吧，并经常在午夜办理入住手续。另一方面，不同的POI具有不同的开放时间和高峰时间（例如，餐馆与酒吧），因此他们随时间的登记模式也不同。

1. **MODELS FOR EVALUATION（模型评估）**

在本节中，我们将介绍评估中包含的12个POI推荐模型。 它们代表了最先进的方法。它们包括（i）四种流行的推荐技术和（ii）五种类型的背景信息，如地理影响。 这些模型总结在表1中。接下来，我们根据它们的推荐技术对它们进行分组，并介绍它们是如何建模和合并上下文信息的。

* 1. **Matrix Factorization Models**

矩阵分解（MF）[17]将登记矩阵C∈RM×N分解为用户矩阵U∈RM×K和POI矩L∈RN×K，其中M，N和K分别是用户数，POI和潜在因素。每个用户i和POI j的潜在特征由ui和lj表示。因此，针对POI j的用户i的推荐得分被建模为内积，并且目标函数被表达为：



|| ·||F是矩阵的Frobenius范数，λ1和λ2是正则化参数。我们接下来介绍评估中考虑的基于MF的模型。

* + 1. **LRT（2013）**

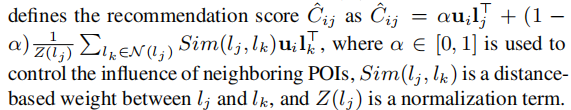
LRT [11]是一种时间增强的MF模型。 基于用户的登记行为随时间变化的观察，LRT针对不同的时隙通过不同的潜在向量对每个用户进行建模，并且从所有潜在向量计算最终推荐得分。

**Temporal influence.** 为了在不同时间对用户的偏好进行建模，LRT分别对每个时隙t的登记矩阵C(t)进行分解，其中t∈{0,1，...，23}是一天中的一小时。此外，受用户在近距离时间段的兴趣往往相似的直觉的启发，将正则化术语添加到MF的目标函数中，制定为

LRT将所有时段的推荐分数总结为最终得分，

* + 1. **IrenMF（2014）**

IRenMF [30]基于加权矩阵分解（WMF）[16,35]。IRenMF背后的直觉是(i)用户对相邻POI（位置级别影响）具有相似的偏好[15]，以及(ii)同一地理区域中的POI可以共享相似的用户偏好（区域级别影响）。

**Geographical influence.** 为了对位置级影响进行建模，用户i对POI j的推荐得分还包括相邻POI N(lj)的影响。在形式上，模型定义了推荐分数

为了对区域级影响进行建模，IRenMF首先根据其地理位置将所有POI聚类为G区域。通过假设来自相同区域的POI的潜在因子共享相同的稀疏模式，它们在目标函数中添加套索惩罚作为

* + 1. **GeoMF**

GeoMF [24]是一个地理WMF模型。 为了捕获空间聚类现象（即，同一用户访问的POI可能位于同一区域[46]），GeoMF通过对用户活动区域的建模和对地理空间的影响传播来整合地理影响。

**Geographical influence.** GeoMF将整个地理空间划分为R网格，每个网格代表一个地理区域。对于每个POI，其影响传播到周围区域，吸引附近的用户访问。特别地，引入了两个矩阵，即用户活动区域X∈RM×R和POI影响区域Y∈RN×R。条目表示POI j对区域r的影响，其中d(r，j)是POI j和区域r之间的距离，K(·)是标准正态分布，σ是标准偏差。Xir表示用户i出现在区域r中的可能性。用户i对POI j的地理偏好估计为

* + 1. **RankGeoFM**

RankGeoFM [22]是基于排名的MF模型，其(i)学习用户对POI的偏好排名，以及(ii)包括邻近POI的地理影响。

**Geographical influence.** 除了用户偏好矩阵U(1)之外，RankGeoFM使用另一个潜在矩阵U(2)来表示用户的地理偏好。因此，推荐分数被计算为第一项对用户偏好分数进行建模，而第二项对用户因为其邻居而喜欢POI的地理影响分数进行建模，其中N（lj）表示j的相邻POI，而wjk是基于距离的权重分配到POI k。

* + 1. **ASMF**

ASMF[20]是一个两步POI推荐框架，它(i)从用户朋友那里学习潜在的位置，和(ii)将潜在的位置结合到WMF中，以克服冷启动问题。

**Social influence.** 对于每个用户i，ASMF将三种类型的朋友（即社交朋友，位置朋友和邻近朋友）访问的位置视为他/她的潜在位置poti，并将小值α∈[0,1]分配给Cik，其中k∈poti。

**Categorical influence.** 在计算推荐分数时，ASMF使用基于类别的权重，即其中cj是POIj的类别，Qicj是用户i对cj的偏好，并且是调整参数。

**Geographical influence.** 基于距离的地理分数pGij与WMF的结果融合为总推荐分数，即，根据用户家庭与其签入POI之间的距离分布计算。

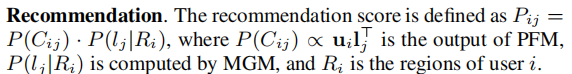
* 1. **Poisson Factor Models**

泊松因子模型（PFM）[31]是概率模型，其将用户-POI值机矩阵C分解为C~Poisson（ULT）。 我们在评估中包含两个基于PFM的模型。

* + 1. **MGMPFM**

MGMPFM [6]是一种融合PFM输出和地理建模方法的融合模型，即多中心高斯模型(MGM)。

**Geographical influence.** 基于观察用户的签到通常分布在几个中心（例如家庭和工作场所）的情况下，MGM使用多个高斯分布来学习每个用户的活动区域。



* + 1. **GeoPFM**

GeoPFM [26]的想法是用户的地理偏好和兴趣偏好相互影响，用户的偏好与他们两者相关。因此，GeoPFM共同学习用户的地理偏好和兴趣偏好。

**Geographical influence.** 潜在区域被整合到PFM中，由空间上的二维高斯分布表示，并且每个用户在区域上具有多项分布。

* 1. **Link-based Models**

LFBCA [41]是一个基于链接的模型，它构建一个图来模拟LBSN用户及其关系。在图中，用户偏好和社交影响都是由不同类型的边缘建模的。特别地，具有相似登记行为的用户被链接以对“相似性关系”进行建模，并且添加表示“友谊关系”的边缘以连接图中的朋友。基于构造的图形，为每个用户执行书签-着色算法算法（BCA）[3]以计算他/她与每个其他用户的相似性，然后执行基于用户的协同过滤(UCF) [1] 基于相似之处

* 1. **Hybrid Models(混合模型)**

混合模型组合了两种或更多种推荐方法的输出，并且每种方法模拟用户偏好或一种上下文信息。例如，MGMPFM（参见第3.2.1节）是PFM和MGM的混合模型。我们接下来介绍我们评估中包含的更多混合模型。请注意，对于推荐，USG线性地组合了上下文影响，而其他模型使用上下文影响的乘法作为最终推荐分数。

* + 1. **USG**

USG[46]同时为POI推荐模型提供用户偏好、社会影响和地理影响。

**User preference.** 该模型将UCF应用于模型用户偏好，其中两个用户之间的相似性是基于它们的公共检验来计算的。

**Social influence.** 为了利用社交影响力，USG建议基于朋友的协同过滤（FCF）根据类似的朋友提出POI推荐。朋友之间的相似性基于他们共同的登记POI和普通朋友。

**Geographical influence.** 给定POI j和用户i，地理影响被估计为基于用户的历史访问POI来访问j的概率，即，其中d(lj，lk)是lj和lk之间的距离。P r(·)估计用户行进距离d(lj，lk)的概率。

* + 1. **iGSLR**

iGSLR [54]利用地理偏好和社会影响力来推荐POI。

**Social influence.** 与USG类似，IGSLR还使用FCF来利用朋友的“签到”，其中朋友之间的相似性是基于他们的住宅的距离来计算的。在我们的数据集中，由于没有用户的居住位置，我们将用户最频繁的登录POI作为他们的住所。

**Geographical influence.** 对于每个用户，iGSLR使用核密度估计(KDE)从他/她的签入历史记录中学习距离分布。因此，用户i访问新POIj的概率是根据POIj与用户i访问的POI之间距离的KDE值估算的。

* + 1. **LORE**

不同于其他的模式，LORE[57]除了社会和地理上的影响外，还考虑了顺序的影响。

**Social influence**. 采用FCF来模拟社会影响，其中社会相似性被定义为iGSLR。

**Geographical influence.** 对于每个用户，LORE使用KDE在二维空间中建模登记概率分布。然后，基于其在登记概率分布上的位置来估计访问新POI的地理概率。

**Sequential influence.** LORE采用加性马尔可夫链（AMC）[40]来利用POI之间的连续影响。 用户访问POI的顺序概率基于所有用户访问的POI与目标POI之间的转移概率。

* + 1. **GeoSoCa**

Geosca[55]描述了三种类型的上下文信息，即地理、社会和分类相关。

**Geographical influence.** GeoSoCa还使用二维KDE进行地理建模。与LORE不同，其中σ由所有用户共享，GeoSoCa增加了本地（即用户相关的）带宽，使地理建模更加个性化。

**Social influence.** GeoSoCa估计用户社交登记频率的幂律分布（表示为fS (xij)）。 社交登记频率xij是指用户i的朋友在POI j上的登记频率。GeoSoCa使用fS (xij)的累积分布作为推荐中的社会影响。

**Categorical influence.** 与社会影响力建模类似，GeoSoCa估计用户的分类登记频率的幂律分布（表示为fC(yic)）。分类登记频率yic表示用户i在具有类别c的所有POI上的登记频率。fC(yic)的累积分布用作推荐中的分类影响。

**4. EVALUATION SETTING**(评价设定)

**4.1 Datasets**

我们的实验是在三个公共数据集中进行的。

**Gowalla dataset.** Gowalla登记数据于2009年2月至2010年10月在全球范围内生成。我们筛选出登记POI少于15的用户以及访客少于10人的POI。过滤的数据集包括18,737个用户，32,510个POI，1,278,274个签到。用户POI签到矩阵的稀疏度为99.865％。

**Foursquare dataset.** Foursquare数据[44]包括2012年4月至2013年9月的登记数据。我们使用美国境内（阿拉斯加和夏威夷除外）生成的记录，消除登记POI少于10的用户以及POI 访客少于10人。过滤的数据集包含24,941个用户，28,593个POI和1,196,248个签到。用户POI签到矩阵的稀疏度为99.900％。

**Yelp dataset.** Yelp数据包含大量地理标记业务（被视为POI）和几个城市内的评论。我们消除了少于10个登记POI的用户，以及访问者少于10个的POI。 这产生了一个包含30,887个用户，18,995个POI和860,888个评论的数据集。用户POI签到矩阵的稀疏度为99.860％。

我们将每个数据集分为训练集，调优集和测试集。 对于每个用户，我们使用最早的70％签到作为训练数据，最近20％签到作为测试数据，剩余10％作为调整数据。对于每个模型，我们根据调谐数据调整参数，以找到最大化Pre@10和Rec@10的最佳值（参见第4.2节），然后在测试数据中使用它们。参数设置在完整版[29]的附录B中提供。

**4.2 Evaluation Metrics**

为了评估模型，我们使用4种广泛使用的度量，即精度（Pre@K），召回（Rec@K），归一化折扣累积增益（nDCG@K）和平均精度（MAP@K）。据我们所知，以前的工作都没有使用所有4个实验指标。指标的正式定义包含在完整版的附录A中[29]。

**4.3 Performance Evaluation Procedure**

为了系统地评价所有的模型，我们设计了一个全面的评估过程，包括以下四个部分。

***4.3.1 Evaluation on Different Types of Data***

为了评估不同数据特性对精度的影响，我们设计了以下两个实验。

**Different datasets.** 我们评价了Gowalla、Foursquare和Yelp数据上的所有模型。我们从5、10、20到50变化k。

**Data density（数据密度）.** 为了研究训练数据密度的影响，对于Foursquare和Gowalla数据集中的每一个，我们生成具有不同密度水平的训练集，即0.0010,0.0008,0.0006,0.0004和0.0002，通过随机消除登记矩阵的非零项 （使其更稀疏）或随机将数据从调整集移动到训练集（使其更密集）。 请注意，Foursquare数据集的密度为0.0010，因此我们只能生成密度小于0.0010的训练数据。

***4.3.2 Evaluation for Different Types of Users***

我们设计了三个实验来研究不同类型的用户属性对推荐模型的影响。

**Number of check-in POIs of users.** 预计检入POI的数量将影响推荐模型的准确性。我们根据训练数据中检入POI的数量将用户分成几组。特别是，我们将Gowalla用户分为五组：“<15”，“15-30”，“30-50”，“50-100”和“> 100”，其中分别包含6164,7201,2979,1672和721个用户。我们还将Foursquare用户分为五组：“<10”，“10-20”，“20-30”，“30-50”和“> 50”，其中分别包含6045,9689,4882,3341和984个用户。我们使用所有用户训练模型，并分别在不同的用户组上进行评估。

**Activity range of users.** 由于地理因素在POI建议中很重要，我们设计此实验来研究用户活动范围对推荐模型的影响。特别是，我们将Gowalla和Foursquare数据的用户分为五组：“<10”，“10-50”，“50-200”，“200-800”和“> 800”，基于他们之间的平均距离登记POI（以公里为单位），反映活动范围。在Gowalla，五组中分别有3664,3263,3536,4213和4061名用户。 在Foursquare中，五个组中分别有6814,3189,3419,5939和5580个用户。 请注意，这些组中的登记POI数量相似，因此不会影响结果。

***4.3.3 Evaluation for Different Modeling Methods***

如第3.4节所述，这些混合POI推荐模型在如何模拟用户偏好和每种类型的上下文信息（尤其是地理和社会影响）方面存在差异。如果我们可以评估这些方法的各个组成部分来回答诸如POI建议的用户偏好建模组件有多好等问题，那将非常有用。我们设计了以下实验来比较模型中使用的用户偏好、地理和社会组件。这些单独建模组件的名称如表1所示。

**Comparing geographical modeling methods.** 我们评估了这些模型中使用的6种地理建模方法。此外，我们还评估具有不同登记POI数量的用户的准确性。请注意，很难将地理和社会部分与基于MF的模型和GeoPFM隔离开来，因此我们不在此处包含它们。

**Comparing social modeling methods.** 我们评价了在这些模型中使用的5种类型的社会建模方法。我们还评估了不同数量的朋友的用户的准确性。

**Comparing user preference modeling methods.** 我们评估了7种用户偏好建模方法，没有包含任何上下文信息(例如地理信息和社会信息)。

***4.3.4 Scalability Evaluation***

可伸缩性也是推荐系统实际利益的重要维度。因此，我们评估模型的训练和查询（即推荐）可伸缩性。据我们所知，这是第一项评估POI推荐模型的培训和查询可扩展性的工作。

**Time complexity analysis.** 我们分析了不同模型的训练和查询时间复杂性，并在全文附录F[29]中总结了结果。

**Training scalability.** 为了探索训练可扩展性，我们使用20％，40％，60％和80％的Gowalla数据集作为训练集来测试7个模型的训练时间（即MGMPFM，LRT，LFBCA，IRenMF，GeoMF，RankGeoFM和GeoPFM）。

**Querying scalability.** 为了探索查询可扩展性，我们将Gowalla用户分为五组：“<15”，“15-30”，“30-50”，“50-100”和“> 100”，基于用户检查的数量-在POI中，测试每个组的模型的平均查询时间。

1. **EVALUATION RESULTS**

在本节中，我们将展示实验结果。请注意，对于某些实验，nDCG和MAP上的结果与精度和召回相似，我们不会显示它们以节省空间。

**5.1 Performance on Different Types of Data**

**5.1.1 Performance on Different Datasets**

图2,3和4分别描述了12个模型与Gowalla，Foursquare和Yelp的top-K建议的总体比较。请注意，Foursquare数据没有社交信息，因此我们只报告这些方法的结果而不使用社交信息。此外，GeoSoCa和ASMF仅在Yelp上进行评估，因为Gowalla和Foursquare没有所需的分类信息。

对于每个模型，Gowalla和Foursquare数据的准确度相似。 但是，所有模型在Yelp上表现更差。 例如，Gowalla和Foursquare上RankGeoFM的Pre @ 5分别为0.069和0.063，而Yelp上的值仅为0.032。这可能是因为Yelp用户的活动范围大于Gowalla和Foursquare用户（见图1（a）），这使得很难对用户的地理偏好进行建模（在第5.2.2节中讨论）。考虑到相对性能，大多数模型在三个数据集中表现一致。 我们讨论他们的相对表现如下。

**Hybrid models.** 在混合模型中，USG表现出比iGSLR，LORE和GeoSoCa更好的推荐质量。以Yelp的结果为例，就Pre@5而言，USG的表现优于其他三个，分别为105.93％，6.71％和43.29％。一个可能的原因是尽管iGSLR，LORE和GeoSoCa利用地理，社交，顺序和分类信息来间接表征用户偏好，但他们仍然错过了USG所做的用户偏好的直接建模（即，使用UCF）。因此，对于混合模型，最好在用户偏好建模之上利用其他信息来获得POI建议。

**IRenMF, GeoMF and RankGeoFM.** 他们是前3名最佳模型，而RankGeoFM通常表现最佳。 与USG相比，RankGeoFM和IRenMF在所有4个指标和不同的K值方面分别在Gowalla上实现了大约14％和10％的改善。 Gowalla和Yelp的GeoMF也分别比USG高出5％-10％和15％-20％。所有这三个模型都是针对隐式反馈数据而设计的。这表明将用户签到建模作为隐式反馈更适合POI建议。

**Comparing IRenMF and GeoMF.** 在Gowalla和Foursquare数据上，IRenMF优于GeoMF，但在Yelp数据上更差。这可能是因为IRenMF假设用户倾向于访问其访问位置附近的POI，而GeoMF将这些POI视为负样本。图1（a）显示Gowalla和Foursquare中用户登记POI之间的距离小于Yelp，这意味着用户更有可能访问Gowalla和Foursqaure附近的POI。因此，与Yelp相比，IRenMF的假设更可能保持在Gowalla和Foursquare上，这导致它们在这些数据集上的不同表现。

**ASMF.** ASMF的表现与USG相似（差异<5％）。 但是，它不如其他地理增强型MF模型好。例如，在所有评估指标和K值上，GeoMF比ASMF好8％-23％。 这可能是因为ASMF专注于利用社交信息，而学习地理影响可能是改善MF模型的更好方法。

**LRT.** LRT在这些型号中表现最差。 它只考虑时间信息，而不对地理和社会影响进行建模。此外，基于时隙划分登记矩阵使得数据更稀疏，对学习用户偏好造成负面影响。

**LFBCA.** Ye等[46]表明基于链路的方法的性能不如USG。 然而，在我们的评估中，LFBCA（也是一种基于链接的方法）在Gowalla上与USG达到了相似的准确度，在Yelp上比USG差5％-10％。这是因为LFBCA不仅利用了朋友的信息， 但也考虑其他空间相似用户的影响。 这告诉我们，对于基于链接的模型，用户关于其空间行为的关系也应被视为“链接”。

**Poisson Factor Models.** 在实验中，GeoPFM始终击败MGMPFM。它们之间的区别在于GeoPFM共同学习地理影响和用户偏好，而MGMPFM简单地融合了两个组件的输出。 这意味着联合学习是比单独建模更好的利用上下文信息的方法。

***5.1.2 Performance on Different Density***

图5显示了在不同数据密度下Gowalla和Foursquare的准确度。LRT和iGSLR不包括在这里，因为它们的性能在之前的实验中并不好。Yelp的结果在质量上相似并省略。

**GeoPFM.** 通常，除了GeoPFM之外，每个模型都会受到较低数据密度的严重影响。 例如，当数据密度从0.0008降至0.0002时，这些模型在Pre @ 10方面的性能在Gowalla上降低了至少35％。尽管当数据密集时，GeoPFM不如RankGeoFM，IRenMF和GeoMF好，但它 对数据密度低是非常强大的。就Pre@10而言，当密度从0.0008降至0.0002时，GeoPFM仅损失13.66％和12.25％。一种可能的解释是GeoPFM使用分层方式来描述用户偏好，即，每个用户对潜在区域具有偏好并且对每个区域内的POI具有偏好。这可能有助于克服数据稀缺问题。

**RankGeoFM.** 当密度大于0.0008时，RankGeoFM报告最佳性能，将第二好的（即IRenMF）超出5％-10％。然而，当密度下降到0.0002时，其精度急剧下降，这比GeoMF和GeoPFM差。这可能是因为RankGeoFM学会将正例排名高于负例，但可用于学习稀疏数据排名的正例较少。结果不太可靠。

**GeoMF.** GeoMF的性能略差于IRenMF和RankGeoFM在0.0008和0.0010，但在密度为0.0002时变为最有效的模型，在两个数据集上均优于第二最佳的5％-10％。因此，对于稀疏数据，GeoMF优于其他模型。

**Comparing USG and LFBCA.** 结果表明，USG在稀疏数据（0.0002-0.0006）上的表现优于LFBCA至少7％，而在密集数据（0.0010）上的表现略差2％-3％。这可能是因为用户偏好建模，LFBCA 还考虑间接类似用户，例如，类似用户的类似用户。间接用户可能在数据密集时提供更多信息，而在数据稀疏时导致偏差。

**LORE**. LORE并不直接模拟用户偏好，但它仍然受益于更高的密度。这表明社会影响等背景信息对数据密度也很敏感。

***5.1.3 Performance on Other Datasets***

此外，我们还对其他一些数据集进行了大量实验。 这些数据集的结果在性质上相似，并在完整版的附录C和D中提供[29]。

**5.2 Performance on Different Users**

**5.2.1 Users with Different Numbers of Check-in POIs**

图6显示了具有不同登记POI数量的不同用户组的结果。 考虑社交信息的模型仅在Gowalla上进行评估，Yelp上的结果相似且省略。 我们做出以下观察：（1）USG，LFBCA，IRenMF，GeoMF和RankGeoFM的表现总体上优于其他模型，其中IRenMF，GeoMF和RankGeoFM在两个数据集中通常优于其他两个模型5％-15％。 对于登记POI较少的用户群，改进甚至更大。这一观察结果证实了5种模型对主动和冷启动用户的优越性。（2）LORE经历了最大的增长，因为登记POI的数量增加，如图6（a）所示 特别是，对于用户组“<15”，LORE的Pre @ 10仅为USG的一半，而对于用户组“100 <”，它与USG的Pre @ 10相当。 值得注意的是，使用地理，社会和时间信息的其他模型的增加并不像LORE那样重要。 因此，LORE对活跃用户的改进很大程度上归功于其顺序影响建模。

**5.2.2 Users with Different Activity Ranges**

图7显示了Gowalla和Foursquare上具有不同活动范围的用户组的结果。Yelp的结果在定性上是相似的并且被省略。我们进行了两次观察：（1）RankGeoFM，GeoMF，IRenMF，USG和LFBCA在所有基础上均大幅超越其他（> 30％），RankGeoFM通常表现最佳。（2）所有具有地理建模的模型在Rec@10的两个数据集中经历从第一组（<10）到最后一组（800 <）的至少10％的损失。这意味着更大范围的用户活动导致侧面 对地理增强的POI建议的影响。对于具有广泛活动范围的用户，地理模型可能无法正常工作。

**5.2.3 Tourist Users**

为了评估特殊类型用户（即游客）的模型，我们将模型与另外两个数据集上的简单基线进行比较。 我们发现推荐最受欢迎的POI的基线比我们评估的大多数模型表现更好，并且在为游客推荐时与RankGeoFM相当。详情可以在完整版的附录E中找到[29]。

5.3 Different Modeling Methods

5.3.1 Geographical Modeling

图8和图9显示了6种地理建模方法对Gowalla数据的评价结果。类似的结果可以在Foursquare和Yelp上找到，因此被省略了。

**PD & PD+.** PD是USG的地理组成部分（参见第3.4.1节），PD +是PD的改进版本[50]。 如图8和9所示，相对于不同的K值和用户组，PD +优于所有其他模型至少25％，并且PD在大多数情况下是第二好的方法。

**MGM.** MGM是MGMPFM的地理组成部分（见第3.2.1节）。图8显示，MGM的整体性能不令人满意，而图9显示，它对超过100个签入POI的用户表现良好，在pre@10时优于pd 10%，在rec@10时仅差1%。这表明当用户具有许多登记POI时，他/她的POI的分布更可能与MGM的假设一致，即，遵循多中心高斯分布。

**1dKDE.** 1 dKDE是iGSLR的地理组成部分(见3.4.2节)。它在我们的实验中表现最差。

**2dKDE & AKDE.** 2dKDE和AKDE分别是LORE和GeoSoCa的地理组成部分（见3.4.3节和3.4.4节）。AKDE始终优于2dKDE 15％-40％。这是因为AKDE为每个用户使用内核函数的全局和个性化带宽，因此可以更好地执行。与PD相比，前四个用户组（<100）的AKDE较差8％-17％，而最后一个用户组（> 100）变得相似（差异<2％）。 这意味着，与MGM类似，AKDE对于活跃用户也更好。

**5.3.2 Social Modeling**

图10和11显示了5种社交建模方法的结果：FBCA。FBCA是LFBCA的社会组成部分(见第3.3节)。一般来说，FBCA是最好的社交建模方法，在Gowalla的召回率方面表现优于第二好的CI-FCF，但其精确度值相似。这表明基于链接的方法可能是建模社会影响力的良好选择。具体来说，图11显示FBCA对于冷启动用户更强大，但对活跃用户无效。这是因为对于朋友很少的用户，FBCA可以利用他/她朋友的朋友等，这有助于为冷启动用户聚合更多信息，而其他模型只考虑用户的直接朋友。但是，对于活跃用户，由于包含太多间接朋友的影响，FBCA可能不准确。

**CI-FCF & CN-FCF.** CI-FCF和CN-FCF是USG的两个社会组成部分（参见第3.4.1节）。 从图10中我们可以看出，就Pre@5而言，CI-FCF比CN-FCF更有效22.77％。这意味着基于登记POI和频率的空间相似性优于FCF方法的在线相似性，例如共同朋友的数量 这一结果与之前的工作一致[46]。

**D-FCF.** D-FCF是iGSLR和Lore的社会组成部分(见第3.4.2和3.4.3节)。图10显示了在所有方法中它是最差的。

**SC.** SC是GeoSoCa的社会组成部分（参见第3.4.4节）。就Pre @ 10而言，SC的表现优于CN-FCF和D-FCF 15.23％和20.87％，而CI-FCF的表现优于CI-FCF，CI-FCF也为社会登记提供建议。这意味着CI-FCF可能是社交登记聚合的更好选择。

值得注意的是，在图11中，所有模型的召回值在开始时增加，但之后略有下降。 这是因为拥有更多朋友的用户也倾向于在更多POI处办理登机手续。 具体而言，第一组（<2）中的用户平均有21.57个登记POI，而最后一组（20 <）的这个值是38.80。

5.3.3 User Preference Modeling

我们比较了POI推荐模型中使用的7种用户偏好建模方法，包括基于用户的协同过滤(UCF)，基于位置的书签着色算法（LBCA）矩阵分解（MF），泊松因子模型（PFM），加权矩阵分解(WMF)，非负矩阵分解（NMF）[25]和贝叶斯个性化排序（BPRMF）[36]。仅关注用户偏好建模，所有这些方法都基于用户POI签到矩阵，而不使用任何上下文信息。图12和13描绘了这些方法的准确性。

**LBCA.** LBCA是3.3节中基于链接的模型，在用户 - 用户图中没有社交链接。 LBCA是用户偏好建模的最佳方法。例如，就Pre@5而言，它分别比UCF，WMF和BPRMF分别高出11％，7％和13％。一个可能的原因是它利用直接和间接类似用户（即，类似用户的类似用户），因此可以聚合更多信息用于推荐。

**MF.** 矩阵因子分解在传统推荐问题中表现良好，因POI建议而失败。 这是因为POI建议中的数据比传统的推荐问题更稀疏，并且MF不适合签入，这是隐式反馈数据。 该发现与先前工作中报道的结果一致[22]。

**PFM & NMF.** 它们在POI建议方面表现不佳。这也解释了为什么基于PFM的MGMPFM和GeoPFM在我们以前的实验中是不好的。

**UCF, WMF & BPRMF.** 它们表现相似（差异<3％），是用户偏好建模的最佳方法之一。例如，它们在Pre @ 5上的表现优于NMF和PFM超过140％。 此外，它们对于冷启动用户和活跃用户都更好。这也解释了为什么基于这些方法的模型，例如USG，IRenMF，GeoMF和RankGeoFM，表现良好。

**5.4 Scalability**

**5.4.1 Training Scalability**

表2显示了POI推荐模型在不同大小的训练集上的训练时间。请注意，我们不评估USG，iGSLR，LORE和GeoSoCa的培训可扩展性，因为它们不需要培训。对于ASMF，Gowalla数据的训练时间（没有类别信息）与训练WMF相同，因此省略。对于LFBCA，我们报告书签着色算法（BCA）的训练时间。 为了比较具有不同训练数据大小的每个模型的可扩展性，我们在图14（a）中显示相对训练时间，其被定义为与使用20％训练数据的运行时间相比的比率。

我们观察到（1）GeoMF几乎不受训练集大小的影响，变化小于2％。这是因为GeoMF的训练时间基于|| C-XYT|| 0而不是NC（表7，附录F [29]）。|| C-XYT|| 0的值不受训练数据大小的影响。因此，GeoMF比其他模型更具可扩展性。（2）MGMPFM，LRT，IRenMF，RankGeoFM和GeoPFM与训练数据的大小成线性比例。 在这些模型中，LRT比其他模型更具可扩展性。 当训练数据的大小增加300％（从20％到80％）时，LRT的训练时间仅增加50％，而IRenMF和RankGeoFM分别增加170％和120％。 MGMPFM和GeoPFM的表现与LRT，IRenMF和RankGeoMF相似且更差。(3)LFBCA的训练时间随训练数据的大小呈超线性增长。因此，LFBCA并不适合于非常大的数据集。

5.4.2 Querying Scalability

表3显示了不同用户组的所有模型的每个用户的平均查询时间，每个用户组具有不同数量的登记POI。在实验中，查询时间表示在所有POI上计算用户u的推荐分数的平均时间。查询时间中不包括POI排名。对于GeoSoCa，由于Gowalla数据中没有分类信息，我们测试其地理和社会组件（即GeoSo）的查询时间。为了将不同方法的可扩展性与登记POI的数量进行比较，我们在图14（b）中显示了相对查询时间，其定义为与第一个用户组的时间成本相比的比率（即<15））。

**iGSLR.** iGSLR具有最差的推荐可扩展性。 特别是，POI少于15的用户仅需1.958s，而POI超过100的用户仅需1292.566。这是因为iGSLR的查询时间随着用户登记POI的数量而立方增加（表7，附录F [29]）。因此，iGSLR不适合活跃用户。

**Hybrid models vs. Joint-learning models.** 一方面，混合模型的查询时间通常随着用户的登记POI数量而增加（MGMPFM除外）。例如，GeoSo的查询时间增加了10倍，从4.977秒（<15）到42.828秒（100 <）。原因是混合模型基于用户访问的POI分别对地理信息进行建模。在推荐时，地理组件需要迭代所有用户访问的POI，因此查询时间通常是访问POI的数量的多项式。另一方面，所有联合学习模型（例如，IRenMF和GeoMF）的查询时间对于不同的用户组是恒定的。这是因为这些模型共同将用户的偏好和上下文信息嵌入到固定大小的矩阵中，因此查询时间是相同的。因此，联合学习模型的查询可扩展性通常优于混合模型。

**MGMPFM.** 对于所有5个用户组，其查询时间在3.4s到6.5s之间变化，这比其他混合模型更具可扩展性。这是因为MGMPFM（即MGM）的地理建模方法基于用户的高斯中心Gu 而不是访问过的POI。即使对于活跃用户，他们的登记POI也可能位于少数几个中心附近。因此，MGMPFM可以更好地扩展到活跃用户。

**5.5 Summary of New Insights**

从评估中，我们观察到许多有趣的发现，这些发现在任何现有工作中从未报道过。这些发现对于理解POI推荐模型很重要，这有助于我们为特定场景（例如，稀疏数据）选择和设计合适的模型。我们总结了下面的主要发现。

* RankGeoFM，IRenMF和GeoMF在不同的数据集和用户类型上优于其他模型，RankGeoFM通常表现最佳。这些模型（1）基于隐式反馈模型，例如WMF和基于排名的MF，以及（2）考虑地理信息。相比之下，不具备这两种特性的LRT表现更差。此外，没有基于用户-POI登记矩阵（例如，iGSLR）直接建模用户偏好的模型不具吸引力（图2,3,4,6和7）。当登记数据密集时（即稀疏度大于0.0004），RankGeoFM是最佳模型，其后是IRenMF和GeoMF。 但是，对于稀疏数据，GeoMF优于所有其他方法（即，稀疏度低至0.0002）（图5）。
* GeoPFM是稀疏度变化最不敏感的模型，而当数据变得稀疏时，所有其他模型的准确度会急剧下降（图5）。
* 当用户拥有更多的POI时，LORE经历了最大的增长，这主要归功于它的顺序建模(图6)。
* 由于用户活动范围更大，所有地理模型的准确性都会下降(图7)。
* PD +是所有用户组的最佳地理建模方法，PD通常表现第二好。 这表明利用幂律分布是地理建模的有效解决方案（图8和9）。
* 模拟个性化二维登记分布（即AKDE和MGM）的方法优于PD，仅适用于具有超过100个登记POI的用户。因为个性化的二维分布需要更多数据来精确捕获用户的行为。因此，它更适合活跃用户（图9）。
* FBCA和CI-FCF是社交建模的两个最佳模型。FBCA为有1位朋友的用户提供最佳服务，而CI-FCF对于有超过5位朋友的用户更有效。基于共同朋友（即CN-FCF）和地理距离（即DFCF）的FCF方法没有显示出有希望的性能（图10和11）。
* LBCA在用户偏好建模方面表现最佳（超过7％），其次是UCF，WMF和BPRMF，其中差异在6％以内。这4个模型比其他方法更好。 这也解释了为什么基于这些偏好建模方法的模型（例如，RankGeoFM和GeoMF）优于其他模型。因此，为了在未来开发新的POI推荐模型，将LBCA，UCF，WMF或BPRMF扩展为推荐更为有希望（图12和13）。
* 对于培训可扩展性，GeoMF是最具可扩展性的模型，其训练时间与训练数据的大小不变; LFBCA超级线性地扩展到训练数据的大小; 其他模型与训练数据的大小成线性关系（图14(a)）。
* 为了查询可扩展性，关于用户的登记POI数量，联合学习模型（即LRT，IRenMF，GeoMF，RankGeoMF和GeoPFM）通常比混合模型（例如，iGSLR和GeoSoCa）更具可扩展性（ 图14（b））。

1. RELATED WORK

在本节中，我们将回顾现有的POI推荐研究。我们首先根据他们使用的方法对它们进行分类，然后介绍具有不同设置的其他POI推荐问题。有关其他POI推荐问题的相关工作的详细信息包含在完整版的附录G中[29]。表4首次从推荐问题，推荐模型和使用的背景信息中全面概述了这些研究。

**6.1 POI Recommendation**

**Collaborative Filtering.** 在关于POI建议的早期工作之一[46]，Ye等人应用基于用户的CF（UCF）和基于Friend的CF（FCF）来模拟他们模型中的用户偏好（包括在我们的评估中）。随后，UCF和FCF已被用于许多POI推荐模型[45,46,54,57,58]。此外，基于项目的CF(ICF)也用于POI建议[19,37]，并且ICF表现得比UCF差[46]。

**Link-based Methods.** 在POI建议中，Ying等[49]提出了一个基于HITS的模型，该模型考虑了用户和POI之间的联系，以及Noulas等人。[34]包括用户-用户友谊和用户POI链接。王等人 [41]进一步包括空间相似用户之间的关系作为用户-用户图中的边缘，并将它们用于推荐。

**Factorization Models.** 不同的MF方法已经被用于POI推荐，例如MF [11]，WMF [30,24,20]，BNMF [25]和基于排名的MF [22]。此外，Bhargava等[4]也应用张量分解来模拟用户偏好。

**Probabilistic Models.** 概率模型通常用图形表示来描述变量之间的相互作用，例如POI建议中地理影响和用户兴趣之间的相互影响。一方面，泊松因子模型（PFM）首先被用于POI推荐[6]然后 扩展到包括上下文信息[26]；另一方面，空间主题模型用于模拟用户的潜在兴趣并整合用户的上下文偏好[18,14,25,59]。

在我们的评估中，每种方法中最具代表性和最先进的模型都会被考虑.

**6.2 Other POI Recommendation Problems**

接下来我们将介绍POI推荐问题的三个变体，它们将附加信息作为输入的一部分。 它们旨在满足用户的更多特定需求。更多细节可以在完整版[29]的附录G中找到。

**Next POI Recommendation.** 给定用户和他/她的当前位置，下一个POI推荐旨在推荐用户可能在下一个时间间隔（例如，在接下来的6个小时内）访问的新POI [7,18,9,60， 13,28]。 大多数下一个POI推荐模型使用连续签入之间的顺序信息来推荐。另外，为位置预测开发的模型也可以应用于下一个POI推荐[32,38,51,53,33,43]。

**Time-aware POI Recommendation.** 考虑到用户偏好随时间变化，给定用户一个时间（例如，下午5点），时间感知POI推荐返回当时最可能被访问的新POI给用户[50,52,22,8]。

**In-town/Out-of-town POI Recommendation.** 给定用户的家乡和当前位置或城市，城镇/城外POI推荐向用户返回新的POI，当用户在城外或城外时使用不同的推荐策略[10,47， 48,42]。 除了最早的工作[10]，城镇/城外POI推荐的所有其他模型都依赖于内容信息（例如，标签和类别）。

还有其他类型的POI推荐，例如类别感知[27]和需求感知[51,53] POI推荐。