**兴趣点推荐的地理配对排序矩阵分解模型**

**摘要----**兴趣点(POI)推荐是基于位置的社交网络(LBSNs)中的一个重要应用。与传统的推荐问题(如电影推荐)相比，地理影响是一个特殊的特征，在推荐POI中起着重要的作用。最近，POI建议中提出了各种将地理影响纳入协作过滤技术的方法。然而，以往的地理模型一直在努力解决地理上噪音大的POI问题，即遵循地理影响但不满足用户喜好的POI。我们观察到，同一地理区域的用户共享多个POI，因此我们提出了对地理噪声进行过滤的共同 地理影响。此外，我们提出了用于POI推荐的地理 - 成对排序矩阵因子分解（Geo-PRMF）模型，其将共同地理影响结合到个性化成对偏好排序矩阵分解模型中。我们对两个真实数据集，即Foursquare和Gowalla进行了实验，实验结果表明，该方法的性能优于现有的模型。

**1 Introduction**

兴趣点(POI)推荐一直受到基于位置的社交网络(LBSN)服务如Foursquare和Facebook Place的飞速发展的推动。一个典型的LBSN允许用户在他们的位置签入、交友和共享信息。LBSNs中的POI推荐旨在帮助用户通过LBSN服务在城市中探索新的和有趣的地方。例如，当你去购物时，你可以通过POI推荐很容易地找到市中心购物中心的详细信息和附近的食品店；这样做不仅可以改善用户的体验，同时也为商家提供了新的目标客户机会。

由于POI推荐的重要性，已经提出了解决这一任务的各种方法[1、8、11-13、18、20]。受传统推荐系统（例如，Netflix的电影推荐系统）的启发，构建用户POI矩阵，其将POI作为项目和用户的登记频率视为评级值。然后使用协同过滤技术来推荐POI。 此外，地理影响已被纳入提议的POI推荐系统以提高绩效的重要因素[1,2,11,16,19]。然而，以往旨在捕捉地理影响的模型一直在努力解决地理上嘈杂的聚居区问题。现有的地理影响模型受到地理上嘈杂的POI问题的困扰，因为他们建议新的POI接近用户已登记的POI，这完全取决于用户-POI的地理关系。这里，我们给出了一个地理上嘈杂的POI的例子。假设用户喜欢参观他家附近的商店和餐馆，因此在这些地方会产生许多入住。同时，酒店也位于用户的家附近。根据以前的地理影响模型，酒店应该被推荐，因为它靠近用户已经入住的POI。但是，人们住在自己的房子里，通常不想去附近的酒店。 因此，酒店被定义为地理上嘈杂的POI，其遵循地理影响但不满足用户的偏好。

在本文中，我们提出了共同地理的影响，以解决地理噪声的POI问题。我们观察到在相同区域中行动的用户共享许多POI。 例如，在同一所大学就读的两名学生可能彼此不认识，但可能会检查许多相同的兴趣点，例如大学周围的热门餐馆和夜总会。每个用户的登记行为都会增强每个商店的受欢迎程度，吸引更多人。受此观察的启发，我们提出了共同地理影响，假设用户在近距离区域遵循类似的访问模式。

在此基础上，提出了求解POI推荐问题的几何对排序矩阵分解(Geo-PRMF)模型.。受[19,20]的启发，我们对待用户“check-ins”作为隐式反馈，并通过个性化成对偏好排序学习系统。该首选项隐式嵌入成对(签入、未签入)，假定用户对签入POI的兴趣大于对未签入的POI的兴趣。我们利用共同地理的影响来细化偏好对集合，这降低了复杂性成本。具体地，我们的模型过滤地理上嘈杂的POI，这些POI在现有的地理影响模型[1，11，16]中没有解决。

本文的贡献总结如下。首先，我们提出了共同地理影响，以克服地理噪音的问题，阻碍以往的地理影响模型。此外，我们提出了Geo-PRMF模型，该模型将共同地理影响纳入个性化成对偏好排名模型，以学习用户偏好并且比最先进的模型表现更好。

**2 Related Work**

在本节中，我们首先演示了POI推荐的最新进展。 然后，我们展示了先前的研究如何模拟地理影响。最后，我们探讨了我们提出的模型如何与先前的工作相关。

**POI Recommendation.** POI推荐最近引起了学术界的密切关注。大多数提出的方法使用协同过滤（CF）技术，包括基于内存和基于模型的方法，来推荐POI。[11,14,15]的研究人员使用基于用户的CF推荐POI，而其他研究利用基于模型的CF，包括矩阵分解（MF）技术[1,8,9,18]。具体而言，[8,9]的研究人员将签到模型建模为隐式反馈，并使用加权正则化MF进行POI推荐。与[8,9]中的研究人员不同，[3,7,19,20]中的研究人员通过成对排序方法模拟隐式反馈，该方法表现出更好的性能。

**Geographical Influence.** 地理影响在POI推荐中起着重要作用，因为用户在LBSN中的活动受到地理限制的限制。为了捕捉地理影响，研究人员假设POI的共现遵循特定的分布。 一方面，[1,4,16]的研究假设登记的POI遵循高斯分布并提出基于高斯分布的模型; [11,14]中的人使用幂律分布模型; [15]中的研究利用核密度估计模型来学习分布。另一方面，[8,9]的研究人员将地理影响纳入加权正则化MF模型。[19]中的工作将地理影响纳入排名模型，并提出POI推荐的分层地理成对排名。所提出的地理影响模型的核心思想基于以下直觉：用户更喜欢访问用户已登记的附近的新POI。

**Connection to Prior Work.** 之前的研究已经捕获了推荐新POI的地理影响，优先考虑与用户活动中心的接近程度或之前登记的POI。这产生了地理上嘈杂的POI的问题。我们提出了共同的地理影响来克服这个问题。 此外，由于成功使用成对偏好排名来模拟登记活动，我们提出了Geo-PRMF模型，该模型将共同地理影响纳入成对偏好排名模型以学习用户的POI偏好。

**3 Model**

在本节中，我们首先提出共同地理影响，以解决地理上嘈杂的POI问题。 然后，我们提出了Geo-PRMF模型，该模型将共同地理影响纳入成对偏好排序模型以推荐POI。

**3.1 Co-geographical Influence**

为了说明起见，我们定义了以下几个术语。

**Definition 1 (Geographical activity center)** 地理活动中心是基于地理影响而具有最高检入概率的POI。

**Definition 2 (Geographical neighbors)** 地理邻居是拥有地理活动中心的用户。

**Definition 3 (Geographically noisy POI)** 地理噪声POI是靠近用户地理活动中心的POI，但不是用户的首选。

图1演示了用户签入模式和地理干扰的POI问题。以前的研究[1，16]已经表明，大多数人在活动受限的地区生活和享受乐趣。根据这种地理特征，以往的工作构建了用户-POI地理关系：位于用户地理活动中心附近的POI是地理位置上的首选[1，16]。然而，如图1所示，这种假设很容易受到地理噪声的影响。一些POI在地理位置上靠近用户的地理活动中心，但它们与用户的登记模式不匹配，例如第1节中提到的酒店示例。

共同地理影响描述的是用户-用户地理关系，而不是用户-POI关系。我们观察到，地理上的邻国有许多共同的聚居区。具体来说，地理邻居之间的Jaccard相似度大约是随机用户之间的10倍。该模型不仅考虑了用户的地理特征，而且提取了两个用户之间的地理关系。我们发现用户的登记POI分布在一些活动中心[1,16]附近。因此，我们期望用户感兴趣的POI位于用户的地理邻居已签入的范围内。这有助于过滤掉地理上嘈杂的POI。结果，为用户设置的候选POI由POI组成，其中用户的地理邻居已经登记但他/她尚未登记。共同地理影响利用地理邻居之间的共同签入模式，过滤掉地理上有噪音的POI。

**3.2 Geo-Pairwise Ranking Matrix Factorization (Geo-PRMF) Model**

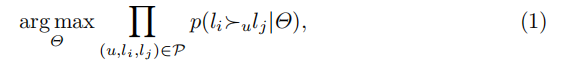
我们提出了Geo-PRMF模型，该模型将共同地理的影响纳入到一个成对排序模型中。由于在将登记活动建模为先前工作[7,18,20]中的隐式反馈的成对偏好排名成功，我们利用贝叶斯个性化排名标准[10]来学习用户对POI的偏好。此外，我们利用共同地理的影响将未评定的POI分类为可比较的POI和不相关的POI。我们假设用户的地理邻居签入的POI是可比较的，而其他的则是无关的。因此，我们只利用可比较的POI生成对的偏好集，抛弃无关的POI，从可比较的POI候选集中推荐POI。基于这一假设，我们提取了精化的配对偏好集和候选POI集，具体如下：

1. 我们为用户映射一个地理活动中心，并通过附近的中心识别顶部的k个地理邻居。
2. 我们认为只有用户的地理邻居签入的POI才是可比较的，任何其他的都是无关的。
3. 我们生成三元组(用户U，检入的POI li，可比的POI lj)作为细化的偏好集合P，并且将类似的POI作为候选集合。

然后，我们可以从细化的成对偏好集合中学习用户的偏好，并推荐来自候选POI集合的POI。

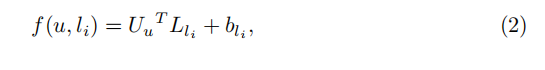
我们按如下方式制定POI推荐问题。设U为用户集合，L为POI的集合。用户u的成对偏好倾向于POI li而不是lj，定义为。Then, we define the pairwise preference set

, where L+u denotes the POIs where user u has checked-in, and Lcu denotes the POIs where geographical neighbors of user u have checked-in but u has not. 现在训练POI推荐系统是为了学习P中的成对偏好关系，

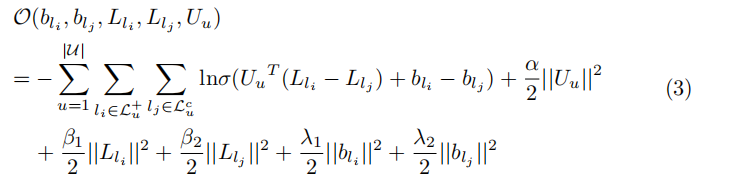


其中p()是用户偏好POI li而不是lj的概率，Θ表示模型的学习参数。

我们利用偏置MF对POI上的用户偏好进行建模。在此基础上，给出了用户u在POI li上的偏好评分函数，

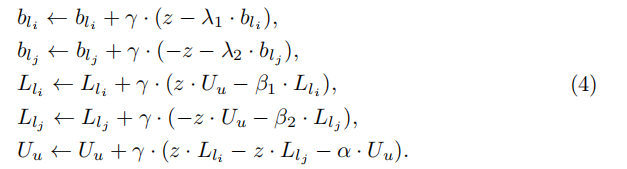


其中分别是用户u和POI li的潜在特征向量，而BLI是估计偏差。此外，我们通过sigmoid函数估计p()的概率函数，，其中σ是sigmoid函数。因此，通过最小化负对数似然，不难获得目标函数。



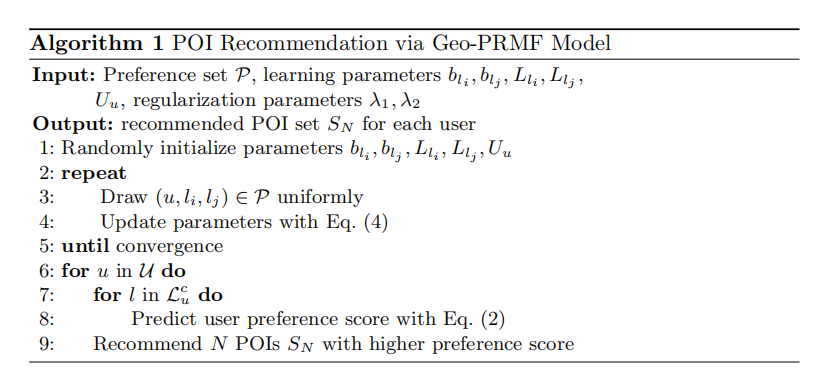
其中α，β1、β2、λ1和λ2是正则化参数。

我们采用随机梯度下降（SGD）方法来学习方程（3）中的参数。我们将公共表达式定义为。然后,参数被更新如下,



在学习参数之后，Geo-PRMF模型根据由等式（2）计算的分数来预测用户在给定POI处的登记偏好。我们首先根据登记偏好对候选集中的POI进行排名，然后为特定用户推荐前N个POI。 算法1演示了如何通过Geo-PRMF模型推荐POI。

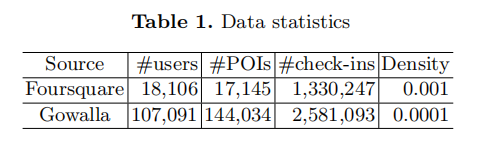
复杂性分析。推荐POI有两个步骤：模型训练和项目推荐。训练Geo-PRMF模型的复杂性是O（d·| S |），与BPR-MF模型[10]的顺序相同，其中d表示潜在因子向量维数和| S | 表示样本数。Geo-PRMF在项目推荐步骤上优于其他模型。对于一般的基于MF的推荐模型，项目推荐步骤的时间复杂度为O（| U |·| L |·d）.Geo-PRMF模型的项目推荐时间复杂度为O（| U |·| Lc |·d）与| Lc | 表示用户的候选POI的平均数量。作为| Lc | 远小于| L |，Geo-PRMF在项目推荐步骤中比其他模型消耗更少的计算。



**4 Experiment**

**4.1 Data Description and Experimental Setting**

实验中使用了两个真实世界的数据集：[5]中的Foursquare数据和[4]中的Gowalla数据。我们从2010年3月到10月从两个数据集中提取数据，将登记的POI过滤少于5个用户，然后选择已签入10次以上作为样本的用户。表1显示了数据统计信息。 我们随机选择每个用户签到的80％作为训练数据，并将剩余的20％用于测试数据。 在[8,20]之后，我们使用精度和召回来测量模型性能。



**4.2 Baseline Methods**

鉴于该方法的目的是为POI推荐建立一个有效的基于MF的模型，我们选择Biased-MF[6]和BPR-MF[10]作为基本的可比模型。此外，为了说明我们提出的模型在捕获地理影响方面的优势，我们将它与融合MF与多中心高斯模型（MGMMF）[1]和具有地理影响的联合模型和MF（GeoMF）[8]进行比较， 捕获地理影响的最先进的POI推荐方法。

**4.3 Experimental Results**

在下文中，我们演示了基线模型和我们提出的Geo-PRMF模型之间的精度@N和召回@N的性能比较。 我们将潜在因子向量维度设置为20，用于所有比较模型。我们针对前5个和前10个POI推荐任务评估两个数据集的不同模型。 图2显示了获得的结果，我们从中得出以下观察结果。1)提出的Geo-PRMF模型在捕获地理影响方面优于MGMMF模型和GeoMF模型，通过滤除地理噪声的POI，取得了最佳的性能。与最佳基准竞争对手相比，Geo-PRMF模型在精度@ 5和召回@ 5时至少提高了5％，并且在精度为10时至少提高了7％，并且在两个数据集中召回了@ 10。2)Geo-PRMF、MGMMF和GeoMF均优于BiasedMF和BPR-MF，证明了捕获地理影响的有效性。

**5 Conclusion and Future Work**

本文针对POI推荐问题，提出了Geo-PRMF模型。我们首先提出了共同地理的影响，这减少了地理噪音的POI，并显著缩小了特定用户的候选集。此外，我们还提出了Geo-PRMF模型，该模型将协同地理的影响纳入了一个成对排序模型中。最后，我们对两个实时LBSN数据集进行了详细的实验，以验证我们提出的模型。实验结果表明，我们提出的Geo-PRMF模型优于最先进的模型。

今后，我们将从以下几个方面对Geo-PRMF模型进行改进.我们可以设计一种自适应的方法来选择活动中心的数量来提高性能。此外，我们还可以考虑用户的评论或位置类别特性，以进一步提高总体推荐性能。此外，最近在LBSNs[17]中出现了一个新的应用程序，它使用签入数据来挖掘商业机会。我们将考虑利用签入特性来增强业务挖掘应用。