**利用POI特有的地理影响力兴趣点推荐**

**摘要**---兴趣点(POI)推荐，即向用户推荐未访问的POI，是基于位置的社交网络的一个基本问题。POI推荐通过在POI中的地理影响与传统的项目推荐(如电影推荐)区别开来。现有的方法将两个POI之间的地理影响建模为两个POI在物理距离下被同一个用户共同访问的概率或倾向。这些方法假设POI之间的地理影响由其物理距离决定，未能捕捉到地理影响的不对称性以及跨POI的地理影响的高度变化。在本文中，我们利用POI特有的地理影响来改进POI推荐。我们利用三个因素对两个POI之间的地理影响进行建模：POI的地理影响，POI的地理敏感性及其物理距离。地理影响力可以捕捉POI对其他POI施加地理影响的能力，地理敏感性反映了POI在地理上受其他POI影响的倾向。在两个实际数据集上的实验结果表明，POI特有的地理影响显着地提高了POI推荐的性能。

**1 Introduction**

基于位置的社交网络(LBSNs)，如Foursquare和Gowalla，越来越受欢迎，弥补了现实世界和在线社交网络服务之间的差距[肖等人，2010年；Sun等人，2017年]。在LBSNs中，用户共享与位置信息相关的位置和内容，从而有助于理解用户的偏好和行为[BaO等人，2012年；Liu和熊，2013年；高ETal.，2015年；Wang等人，2015年a]。兴趣点（POI）推荐，即根据用户的登记记录向用户推荐未访问的POI（例如，餐馆，购物中心和剧院），在过去几年中获得了极大的研究兴趣[Li et al 。，2016; He et al。，2016; Zhang et al。，2016; Li et al。，2017]。

POI推荐最突出的特性之一是，POI的位置和目标用户是推荐的关键因素。例如，在Gowalla和Foursquare，90%的用户连续登录距离不到50公里(Liu等人，2017年)。因此，除了根据用户和POI之间的交互对用户的偏好进行建模之外，正如传统项目推荐中所做的那样，研究人员致力于利用POI之间的地理邻近度或地理影响来提高POI推荐的性能[Ye et al。2011; Lian et al。2014; 谢等人，2016]。

现有的利用地理影响进行POI推荐的方法大致可分为两种模式。假设相互靠近的POI具有相似的用户偏好，第一种方法利用地理邻近度来改善用户偏好的学习[Liu et al。2014; Li et al。2015; Xie et al。2016; Feng et al。2017]。在这些方法中，使用地理邻近度作为用户偏好的一种空间正则化。第二种方法明确地将POI之间的地理影响建模为两个POI在给定物理距离的情况下由同一用户共同访问的概率或倾向[Ye et al。2011; Cheng等，2012; 张和周，2013; Lian et al。2014; Saleem等2017]。各种形式的函数，如幂律函数和高斯分布，被用来捕捉共访问的概率分布与它们的物理距离有关。虽然上述方法在利用地理影响方面取得了一些成功，但它们无法捕捉到不同地理影响的高度差异。例如，如图1所示，Foursquare数据集中10个随机选择的POI表现出完全不同的地理影响，表明地理影响不能仅通过物理距离很好地捕获，因此地理影响应该是POI特定的。

在本文中，我们利用POI特定的地理影响来改进POI推荐。 我们使用三个因素模拟两个POI之间的POI特定地理影响：POI的地理影响，POI的地理敏感性和它们的物理距离。地理影响力捕捉了POI将访问者传播到其他POI的能力，地理敏感性反映了POI接收来自其他POI的访问者的倾向。例如，地铁站通常具有高度的地质影响，而餐馆通常具有较高的地质敏感性。这里，地理影响和地理敏感性是两个低维向量，两个POI之间的地理影响由一个POI的地理影响向量和另一个POI的地理敏感性向量的内积表示。

我们针对特定地理影响的POI模式有两个独特的好处：(1)POI之间的地理影响是不对称的，具有很高的灵活性来捕捉地理影响的高度变异性。（2）我们的模型不是使用POI交互矩阵直接建模POI特定的地理影响，而是通过两个低维向量对每个POI表示地理影响，显着减少了自由参数的数量[Wang et al。，2015b]。因此，我们的模型适用于严重数据稀疏问题的POI推荐。

最后，我们将POI特有的地理影响集成到一个捕捉用户偏好的标准模型中，形成了一种新的POI推荐方法。我们使用用户的登记记录训练我们的模型，并通过应用模型来“预测”他们可能在不久的将来访问的POI来验证推荐性能。我们对来自Foursquare和Gowalla的两个真实数据集进行了大量实验。Gowalla说明了我们模型的有效性。实验结果表明，POI特定的地理影响显着提高了POI推荐的表现，超越了最先进的POI推荐方法。

**2 Related Work**

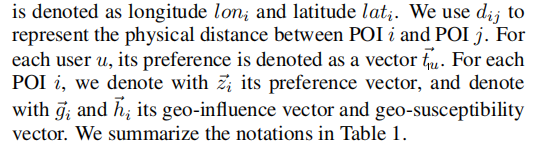
在本节中，我们将简要介绍POI推荐。POI推荐根据用户的登记记录向用户推荐未访问的POI。考虑到用户的签到是隐式反馈，现有方法通过拟合从签到计数[Lian et al。，2014]转换的分数或通过优化用户偏好对POI的成对排名来模拟签到[Li等，2015;2016年 赵等人，2017]。

由于用户签到的稀疏性，仅利用签到计数通常会导致性能不佳。可以结合辅助信息以缓解这种情况。例如，地理影响是最重要的因素之一，并且在在线推荐意义上不存在。现有的地理影响建模方法可分为两类，即全局方法[Ye et al。2011; Cheng等，2012; 张和周，2013; Lian et al。2014]和区域方法[Liu et al。2014; Li et al。2015; Xie et al。2016; Feng et al。2017]。

**全局方法**模拟了POI的共现与地理坐标之间的关系。Ye等人 [2011]和Lian等人 [2014]分别使用幂律分布和高斯分布来表征距离的地理影响。[Cheng et al。2012; Zhang and Chow，2013]通过固定分布捕获每个用户签到的散点图（例如，经度和纬度）。**区域方法**认为同一地理区域的POI对用户具有相似的吸引力。[Xie et al., 2016;Feng et al., 2017; Zhao et al., 2017]使用基于表示的学习方法并限制同一地区的POI共享相似的表示。[Liu et al., 2014; Li et al., 2015通过考虑其地理邻居的吸引力，直接计算目标POI的吸引力。然而，全局方法和区域方法提供了两种粗粒度的地理影响表示，忽略了POI特有的属性。我们在本文中讨论了这个问题。

此外，许多研究已经探索了其他信息以促进POI推荐绩效，例如社交关系[Tang et al，2013]，时间因素[Yuan et al，2013]和category [Zhang and Chow，2015]等。

**3 Preliminary**

我们分别用*U*和*I*表示用户集和POI集。 对于用户u和POI i，我们用cui表示用户访问POI i的次数，而wui是cui的缩放版本。用户访问过的所有POI形成他/她的登记历史记录，表示为Hu。对于每个POI i，其位置表示为经度loni和纬度lati。

**POI recommendation:** 给定一组具有登记历史H的用户U和具有位置信息（lon，lat）的一组POI，POI推荐为每个目标用户建议u∈U由POI组成的列表{i |i∈I}由以下组成：目标用户可能感兴趣但未访问该建议的POI。

**4 Model and Optimization**

在这一部分中，我们描述了POI推荐的模型。所提出的模型由两个部分组成，一个用于POI特定的地理影响，另一个用于用户/项目偏好的建模。该模型的主要新颖之处在于POI特定的地理影响。对于用户/项目偏好，我们将每个签到模型建模为从所有候选POI中选择一个目标POI的过程，避免由于将访问频率的数量直接建模为数字量而引起的偏差。接下来，我们描述POI特定的地理影响和用户偏好的建模。

**4.1 POI-Specific Geographical Influence**

对于目标POI j，我们考虑来自用户u的登记历史Hu中的每个POI i的地理影响。如图1所示，用户更喜欢访问相邻的POI，同时不同的POI具有其物理距离无法很好解释的特征。为了捕捉POI中地理影响的高度变化，我们将地理影响yij从POI i模拟为POI j



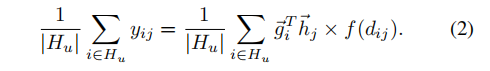
在这里，矢量~gi捕捉到了POI的地理影响，即POI将其访问者扩展到其他POI的能力；矢量Hj反映了POI j的地理敏感性，即POI接收其他POI访客的倾向；dij是POI i与POI j之间的物理距离。

方程式(1)背后的理由如下：

* 首先，f(dij)反映了同一个用户访问两个POI的概率，因为他们的物理距离是dij。在本文中，我们考虑四种类型的函数，如表2所示。通常，f(dij)随着dij的增加而减小，捕获用户更喜欢访问地理上相邻的POI的现象。此外，f(dij)的参数可以在训练POI推荐模型之前进行预训练，也可以与POI推荐模型的训练一起进行。
* 我们模拟了访问的POI与目标POIj之间的相互作用，即。通过这种方式，POI之间的地理影响是不对称的，提供了捕捉POI之间地理影响的高度变化的灵活性。 此外，我们的模型不是使用POI交互矩阵直接建模POI特定的地理影响，而是通过每个POI的两个低维向量表示地理影响，显着减少了自由参数的数量。 因此，我们的模型适用于遭受严重数据稀疏性问题的POI推荐。
* POI特定的地理影响yij基本上捕获了两个POI物理距离和内在特征的联合效应。对于目标POI j，地理上相邻且影响POI将导致高y，而远程但影响POI（或相邻但影响较小的POI）将导致相对较小的y。影响分数y也取决于目标POI的内在特征~h。在给定相同的访问POI i的情况下，这使得不同的目标POI的y不同。换句话说，我们模型中的影响力得分就所涉及的POI而言是POI特定的。

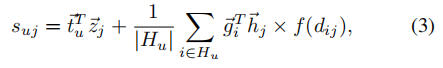
给定用户u和Eq（1）的访问POI的集合hu，我们考虑所有这些访问POI的影响并模

拟Hu对目标POI j的整体地理影响。



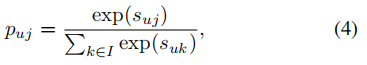
**4.2 Preference Modeling and Recommendation**

为了推断用户对目标POI的偏好，我们考虑了用户偏好和地理影响的影响。具体来说，我们将用户u对POI j的偏好表示为suj，其给出为



采用矩阵分解的方法，对用户u的偏好与POI j偏好之间的交互作用进行了建模。

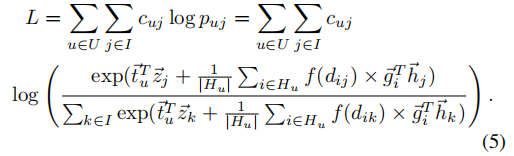
请注意，用户的签到记录了他们在POI的访问频率，这是一种隐含的用户偏好。 因此，与直接适合cuj的传统推荐不同，我们将每个签到模型建模为从所有候选POI中选择一个目标POI的过程。通过这种方式，我们的模型避免了将访问频率的数量直接建模为数字量所引起的偏差。具体地，用户u优选POI j的概率puj被建模为



其中，我是POI的集合，分母是给定用户u对所有POI的规范化。

从等式（4）观察到，对于给定用户u，{puj}用作多项分布的一组参数。因此，用户u访问POI j的行为被建模为决策过程的结果，其中用户从所有候选中挑选一个POI j。然后可以方便地将签到解释为从用户的偏好分布{puj}中抽取的样本。

我们最大限度地提高了观察用户签入的日志可能性：

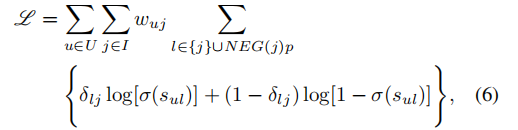


由于登记计数cuj经常显示非常偏斜的分布，因此与少数流行POI的交互可能主导对数可能性。因此，我们用缩放版本wuj替换cuj以缓解问题。具体来说，我们采用对数形式函数[Lian et al。2014]，由给出，其中e是缩放参数。 对于未访问的POI，我们只需设置wuj = 0。

最后，给定用户u，我们根据用户访问POI j的概率puj推荐未访问的POI。 每个用户的推荐列表由未访问的POI中具有最高概率puj的前n个POI组成

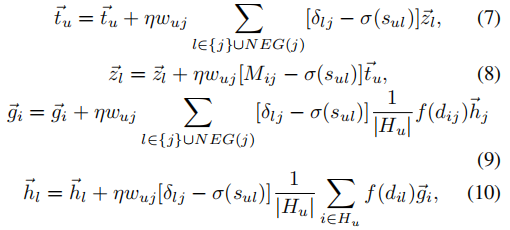
**4.3 Optimization**

我们现在介绍四类潜在因素的优化，包括和地理函数中的参数。

我们采用[Mikolov等，2013]中提出的负抽样方法来最大化对数似然L.对于每个访问的POI，我们根据一些噪声分布对K负检查进行抽样。 具体来说，我们使用以下目标函数来代替L。

其中NEG（j）表示相对于POI j的负POI的集合。 δlj是指示符，如果l = j则等于1，否则为0，并且σ（·）是S形函数。

采用随机梯度上升(SGA)算法对新的目标函数进行优化。在每一次迭代中，我们随机抽样一个由比率ζ设置的小批对来进行优化。采样概率与缩放的登记计数成比例，即wuj。 如果对（u，j）对进行采样，则潜在向量将更新如下：



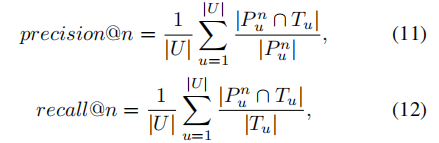
其中i∈Hu，l∈j∪NEG（j），η是学习率。

**5 Experiment**

**5.1 Datasets**

我们使用Foursquare [Cho et al，2011]和Gowalla [Yuan et al。，2013]的两个真实数据集进行评估。我们通过删除访问次数少于10个POI的用户以及访问过的POI来预先处理Foursquare中的签到少于10个用户。在Gowalla中，消除阈值设置为40.在预处理之后，由Foursquare数据集中的88,193个POI的6,118个用户生成172,961个签到，以及由Gowalla数据集中的3,585个POI的1,624个用户生成115,890个签入。两个数据集中的每个POI都与其经度和纬度相关联。此外，在Foursquare数据集中，每个POI由8个类别和240个子类别标记。 对于每个用户，我们按时间顺序对他/她的签到进行排序，并将其早期70％的签到作为培训数据，接下来的15％作为验证数据，最后15％作为测试数据。

**5.2 Evaluation Metrics**

我们采用了两个广泛使用的评估指标[Lian等人，2014年；Liu等人，2017年]，即精确@n和Recall@n，其中n是推荐列表中的POI数量。

其中Pnu是用户u推荐列表中前n个POI的集合，而Tu是用户u的POI基础真值集。| X | 表示集合x的基数。 对于每个度量，我们在实验中考虑n的7个值（1,2,3,5,10,15,20）。

**5.3 Methods in Comparison**

为了方便起见，我们使用GeoIE作为建议的POI推荐方法的名称。我们通过将GeoIE与以下具有代表性的POI推荐方法进行比较，来评估GeoIE的性能：

* UCF + G [Ye et al,2011]：它使用幂律函数来捕获距离的登记概率，然后结合基于用户的协同过滤方法。
* MGM PFM[Cheng等人,2012]：它使用多中心高斯模型捕捉用户的签入分布，然后结合概率因子模型。
* GeoMF[Lian等人,2014年]：它通过增加原始用户和POI的潜在因素和用户的活动区域以及POI的影响区域来扩展MF。
* RankGeoFM [Li et al,2015]：这是一个基于排名的MF模型，它通过考虑邻近POI的吸引力来包括地理影响。根据POI推荐的最新评论[Liu et al，2017]，RankGeoFM是表现最佳的方法之一，并且是最先进的POI推荐方法之一。
* geo-teaser[赵等人,2017]：它结合了一个时态POI嵌入模型和一个地理层次的两两排序方法。

**5.4 Experimental Setting**

我们将缩放参数e设置为10.我们在执行优化时为每个潜在向量放置一个L2正则化项，并将正则化系数设置为0.02。 潜在向量的维数为32.在每次迭代中，负样本的数量K为10，采样率ζ为0.2，学习率为0.001。

在将我们的方法与基线进行比较之前，我们首先评估具有四个不同地理函数的方法的性能，如表2所述。相应的GeoIE命名为GeoIE-PL-PRE，GeoIE-PL，GeoIE-EXP和GeoIE-HB。图2和图3分别显示了四种变体GeoIE在Foursquare数据集和Gowalla数据集上的性能。可以在两个图中观察到，指数函数实现是描述本文中使用的两个数据集上的地理影响和距离之间的关系的最佳选择。本文所使用的两个数据集对两个数据集的影响和距离。指数函数的优越性在于指数函数具有更多的参数，因而可以灵活地捕捉地理影响的高度变化。通过比较GeoIE-PL-PRE和GeoIE-PL，我们发现它们实现了类似的性能。然而，需要指出的是，在我们的模型中用其他潜在因素优化地理函数的参数将导致更快的收敛速度。

**5.5 Effectiveness of GeoIE**

我们选择GeoIE-exp作为我们的方法的代表，并将其与最先进的方法进行比较。对Foursquare数据集和Gowalla数据集的性能比较分别在图4和图5。据观察，GeoIE-EXP始终优于竞争基线方法。

图4显示MGM + PFM在Foursquare上的表现优于UCF + G. 然而，在Gowalla上，UCF + G优于MGM + PFM，除非k = 1，如图5所示。它们承担不同的地理功能，并分别在两个数据集上取得了良好的性能。与通过线性插值融合两个独立模型的UCF + G和MGM + PFM不同，GeoMF将用户偏好和地理影响集成到一个统一模型中。潜在因素在其参数学习过程中可以相互影响，从而产生更好的性能。然而，GeoMF分解了用户的零签入，除了非零签入。这可能会在一定程度上削弱其绩效。

RankGeoMF考虑了邻近POI的地理影响，并利用一种基于排序的方法来挖掘负样本.在这两个数据集上，它的性能都优于GeoMF。然而，邻近POI的吸引力不能直接作为目标POI吸引用户的能力。 Geo-Teaser通过线性插值集成了基于表示学习的方法和地理上分层的成对排序方法。 我们可以观察到它略好于RankGeoMF。

总之，实验结果清楚地表明POI特定的地理影响改善了POI推荐。 POI特定地理影响优于其他利用地理方法的方法，这为我们提供了两个关键含义：(1)POI推荐的地理影响太大，难以用幂律函数或指数函数等简单函数进行量化；（2）尽管POI建议存在数据稀疏性问题，但引入两个额外的向量来表征POI特定的地理影响可以获得显着的改善。

**5.6 POI’s Geo-Influence and Geo-Susceptibility**

为了捕获POI特定的地理影响，我们引入了两个潜在的向量，即每个POI的地理影响和地理敏感性。接下来，我们研究不同类别POI的属性强度。由于空间有限，我们只需从每个类别中选择一个子类别进行案例研究。具体来说，我们计算每个子类别的两种类型潜在向量的平均范数，并将这些范数除以最大值，将它们在0到1之间进行缩放。我们在图6中给出了这些范数。

可以观察到，公共汽车站具有最大的地理影响，这表明它有强大的能力将用户传播到其他POI。然而，这个有影响力的子类别的地理敏感性并不是最大的。同时，酒吧和美术馆的两个属性都处于低位，这表明它们不是用户旅行选择中的重要位置。这些观察结果表明每个POI都具有特定的地理影响和地理敏感性，这也是不对称的。这也验证了我们的假设，即必须对POI特定的地理影响进行建模。

**6 Conclusions**

在本文中，我们利用POI特定的地理影响来改进POI推荐。我们使用三个因素来模拟两个POI之间的地理影响：POI的地理影响，POI的地理敏感性以及它们的物理距离。地理影响力可以捕捉POI对其他POI施加地理影响的能力，地理敏感性反映了POI在地理上受其他POI影响的倾向。我们的模型自然地捕捉了POI之间的不对称地理影响，提供了高度灵活性，以捕捉POI之间的地理影响的高度变化。作为未来的工作，有希望通过POI之间的相互作用直接学习地理影响的多样性。