TGTM: Temporal-Geographical Topic Model for Point-of-Interest Recommendation

Abstract: 基于位置的社交网络（LBSN）和微博-博客服务的广泛使用产生了大量用户的登记数据，其包括用户ID，文本内容，发布时间戳，地理信息等。兴趣点（POI）推荐是提供有趣地点的个性化推荐以增强LBSN中的用户体验的任务。在本文中，我们提出了2个新颖的时间位置内容感知POI推荐模型，它们共同整合辅助时间，文本和空间信息，以提高POI推荐的性能。具体地，我们利用时间信息将原始用户-POI登记频率矩阵划分为子矩阵，以便可以对类似时间场景中的行为进行分组。然后，我们利用（LDA）模型和空间坐标来推断POI。使用真实数据集进行的综合实验证明了我们的方法的优越性。

## 1 Introduction

鉴于用户群体的显着增长率和具有定位功能的大量移动设备，基于位置的社交网络服务已经变得非常受欢迎，包括Facebook Places，Google latitude，Twitter和Foursquare等在线网站。LBSN现在允许用户不仅以“签到”的形式共享他们的物理位置坐标和时间戳，而且还在他们访问过的POI上写下文本意见。使用位置日志数据挖掘和建模登记行为是 对用户和POI都很有价值，因为POI推荐可以帮助人们发现有吸引力的地方，并可能为POI的所有者培养更多潜在的业务。

为此，POI推荐已成为过去几年的热门研究课题。 POI建议中最关键的挑战之一是如何应对用户POI登记频率矩阵的极端稀疏性。此外，与仅考虑用户项目评级的传统建议不同，POI推荐系统要复杂得多，因为在这种情况下的移动行为可能是许多方面的混合。 特别是，当存在诸如评论或微博之类的富文本时，POI推荐应该是个性化的，位置内容感知和上下文依赖。有鉴于此，各种类型的信息构成了另一个重大挑战，即如何系统地以统一的方式将它们纳入其中。

直观地说，用户倾向于在工作日的中午附近的办公室附近的餐馆吃午餐。 但是在周末，用户可能会在晚上长时间访问酒吧以获得乐趣。 我们相信在类似的时间背景下可以更容易地找到移动模式。鉴于此，我们首先基于时间戳将原始用户POI信息划分为子组。然后，我们使用LDA算法[2]来推断用户的兴趣和POI的主题分布。在[4]中，作者发现用户的位移（2个连续登记位置之间的距离）趋势可以用幂律分布近似。考虑到文本和地理影响，在矩阵分解模型的基础上，我们可以推断出用户更喜欢POI的程度。

总之，本文的贡献有三个：

1. 我们从文本内容信息，地理坐标和时间影响的角度研究了用户隐性反馈登记行为与LBSN辅助信息之间的关系。
2. 我们提出了2种用于POI推荐的新概率矩阵分解模型，并且每种模型都包含上述三种类型的辅助信息。
3. 我们通过对从Twitter的API中提取的真实世界LBSN数据集进行综合实验来评估所提出的方法。 结果证明了我们的方法的有效性。

我们有通常的组织：调查，问题定义，提出的方法，实验和结论。

## 2 Related Work

### 2.1 Matrix Factorization

推荐系统中的协同过滤（CF）可以预测未消费项目的个性化偏好[1,19]。 协作过滤方法有两大类。基于邻域的解决方案通过寻找志同道合的用户[5,18]来预测用户的潜在兴趣，这些用户使用相似度函数（例如余弦距离或皮尔逊相关性）来计算相似的用户或项目。基于模型的[9,17]解决方案利用观察到的评级或标签来模拟用户-项目交互。基于模型的各种成功实现基于矩阵分解，其分解用户项目评级矩阵。在本文中，我们采用基于矩阵分解方法的方法，因此我们在此简要介绍。

基本上，矩阵分解（MF）方法将评级矩阵RR分解为一个用户特定矩阵UU和一个项目特定矩阵CC [10]。 然后，可以通过将两个分解矩阵相乘来近似原始评级矩阵，同时使用正则化项避免过拟合：



其中||⋅||F2表示Frobenius范数，Iij是指标函数，如果用户i对电影j进行评级则等于1，否则等于0。常数λ控制正则化的程度以减轻过拟合问题。概率矩阵分解（PMF）[16]给出了正则化的概率解释。

### 2.2 POI Recommendation

POI建议首先在GPS轨迹数据上进行了研究[23]。随着LBSN的日益普及，POI推荐引起了很多关注。有关利用地理影响的POI建议的大量文献。 现有研究发现，人们往往会去附近的地方进行相互观察。基于用户的CF和基于项目的CF在[11,20]中进行了研究。Ye等人[20]在基于用户的CF框架下共同探讨了社会和地理因素，以提出POI建议。最近有兴趣利用社会和地理属性来提高POI建议的有效性。在[3]中，他们提出了一种矩阵分解方案，通过高斯混合模型（GMM）结合地理和社会信息。在[6]中，作者发现，来自类似用户的偏好对于城镇内用户更为重要，而友谊对于城外用户更为重要。

刘等人[13]提出了一种基于图的方法，以综合方式利用时间和地理信息。 袁等人[21]以基于用户的CF方式结合时间影响，并且候选POI的最终偏好分数通过分别基于时间影响和地理影响计算的分数线性组合。

谈到文本内容信息，高等人[7]研究了LBSNs的内容信息。统一POI建议框架下的POI属性，用户兴趣和情绪指示。Liu和Xiong [12]研究了POI相关标签与聚合LDA模型的影响。Hu和Ester [8]根据主题建模方法调查了Twitter和Yelp的用户兴趣。 袁等[22]从空间，时间和内容方面共同模拟个人用户的移动行为。他们用概率生成模型捕获它们。