**RELINE: Point-of-Interest Recommendations using Multiple Network Embeddings**

**摘要：**用户参与基于位置的社交网络（LBSN）的快速增长导致数据在全球范围内迅速增长。 访问和检索接近用户偏好的相关信息的需要是一个开放的问题，其不断为推荐系统带来新的挑战。由于稀疏性和冷启动问题，现有模型对兴趣点（POI）的推荐是不充分的。为了克服这些问题，文献中提出了许多模型，但大多数模型忽略了重要因素，例如：地理接近度，社会影响力，时间和偏好动态，这些因素在个性化推荐时解决了它们的准确性。在这项工作中，我们研究了这些问题，并提出了一个统一的模型，共同学习用户和POI动态。我们的提议被称为RELINE（对多个网络嵌入的推荐）。 更具体地说，RELINE通过将八个关系图嵌入到一个共享潜在空间中来捕获：i）社交，ii）地理，iii）时间影响，以及iv）用户的偏好动态。我们已经使用三个大的真实数据集在准确性方面评估了我们针对最先进方法的方法。 此外，我们已经研究了我们针对冷启动问题的方法的有效性。性能评估结果表明，与现有的最先进方法相比，实现了显着的性能提升。

1 INTRODUCTION

最近，在线社交网络（OSN）将地理信息纳入其内容，触发了新的功能，并引入了基于位置的社交网络（LBSN）的概念。在诸如Facebook Places，Foursquare，Yelp之类的网络中，用户可以与空间维度共享他们的兴趣，以基于他们最近的历史获得对可能有趣的地方的推荐。学习用户的历史是这些模型的关键任务，为兴趣点（POI）提供有意义的建议。不幸的是，稀疏性，异质性和多维度等因素构成了重大挑战，增加了问题的复杂性，对效率和效果产生了很大影响。

对该主题进行了广泛的研究，主要关注用户与用户位置二分网络上的地理信息的关系。 然而，这些方法未能消除稀疏性问题，因为它们错过了与用户交互相关的偏好动态和辅助信息。同样，其他作品通过社交网络关系丰富了用户位置信息，但仍然错过了上述所有因素。 此外，考虑到用户偏好的时间演变的研究仍然无法处理稀疏性，因为存在与用户相关的附加上下文信息，这些信息随着时间的推移而随着他们的偏好而变化。

在本文中，我们提出了一种技术，在一个统一的模型中，考虑社会，地理和时间影响，以及用户的偏好动态，将八个关系图嵌入到一个潜在的空间中。

**Spatial-based Behavior.** 最近的研究指出，在日常活动中用户的登记行为存在空间模式[2]。特别是，登记区域内的位置的用户很有可能参加邻近的位置。例如，位于工作地点或家附近的用户访问邻近位置的概率较高，而不是远距离的用户[6]。我们声称**用户倾向于在与任务相关的区域内的同一时间段内执行一系列活动**。例如，想要购物的用户通常在商场，超市，杂货店或者商店办理登记手续。面包店位于家附近，如图1（a）所示。此外，在周末，人们会在彼此靠近的俱乐部和餐馆中进行多次登记。因此，空间接近度应被视为重复的地理模式。而且，这些位置的关系可以被认为是与一个活动相关的位置的路线

**Temporal-based Behavior.** 通常，用户在其活动中维护固定的每日计划以及他们执行的签到。因此，在工作日，用户在接近工作6 - 5的位置进行登记，而从下午5点开始。 直到午夜，她在离家很近的地方办理登机手续，如图1（b）所示。这种模式在周末是相反的，因为用户往往会在办理入住手续的酒吧和餐馆。一些作品旨在通过关注时间转移来模拟这种行为，但他们**忽略了与用户活动相关的显性和隐性上下文信息**[12][17]。

**Preference Dynamics. 用户倾向于在时间内改变行为，这使得这个问题更加困难**。例如，用户可以在9月的每个周末参加俱乐部，在4月的周末参加电影院或餐馆，如图1（d）所示。在这两种情况下，同一用户交替进行他的登记行为，这应该被考虑在内。根据[6]，[21]的这种偏好演变可能是由于：1）新POI探索：与普通登记行为相反，用户倾向于访问新位置，2）用户体验：用户将根据他过去有过愉快经历的地方，3）受欢迎程度：某些地点往往在一段时间内更受欢迎而不是一年中其他时间，4）社会影响：朋友的意见偏倚用户的决定（如图1（c）所示）。 因此，用户倾向于在参加之前检查他们的朋友对位置的评估并跟随他们的领导。

**Motivation and Contribution.** 在本节中，受用户行为模式的驱动，我们总结了现有方法的局限性：

* 许多方法将POI视为常规节点，并且不捕获空间维度和它们之间的接近度
* 其他工作考虑地理影响，但忽略偏好动态解决所有准确性
* 捕捉动态时间的方法不同时处理空间维度
* 最后，同时考虑空间和时间行为的模型忽略了偏好的演化。

因此，需要在统一模型中考虑所有上述因素，其允许进一步理解用户的行为并使推荐个性化。我们的工作贡献如下：

* 为了克服稀疏性，我们在8个信息图上提出了一种概率加权策略。
* 我们提出了两种新的算法来提取过去历史记录中的路线和停留点
* 我们介绍了一种新的基于图形的方法，从这些加权图形中共同学习用户和POI嵌入到相同的潜在空间中，并提供个性化的POI推荐
* 我们的方法扩展了LINE模型[16]并学习了将大型单平面和二分图嵌入到低维空间中，
* 我们通过实验评估了我们的模型，该模型测量了对于i）所有用户，ii）用户冷启动，以及iii）位置冷启动 的POI建议的准确性

Extensions beyond the Conference Version. 这项工作是我们在第五届IEEE国际数据科学与高级分析会议（DSAA 2018）[3]上提出的工作的扩展版本。该期刊版本包含有关会议论文的若干增强功能。最重要的变化总结如下：

* 我们引入了两个新颖的网络：i）保持积分，代表用户保持最多的位置，以及ii）路线，访问POI时遵循的路径，
* 我们结合了两种新的算法来构建上述信息网络，并根据它们的重要性进行权衡。
* 我们共同捕获用户和POI顺序动态
* 通过研究与用户和位置的冷启动问题相关的领域中的两个重要主题，性能评估部分得到了显着扩展
* 我们在准确性方面比较了我们对其他最新技术方法的态度

**2 RELATED WORK**

在本节中，我们将讨论与POI推荐相关的研究。 特别是，我们分析了先前的工作如何利用n维网络来解决稀疏性和冷启动问题。这些网络包括社会关系、地理邻近性、时间距离和用户在过去历史中签入的偏好动态。

**POI Recommendations.** 轨迹点与用户偏好之间缺乏直接关系源于他们的登记记录，从而将研究工作推向了这个方向。近年来，随着LBSN的增加，用户可以登记位置，这导致他们的数据匿名访问以用于研究目的。在这种情况下，使用这些信息的推荐模型中引起了很多关注。大多数相关工作使用协同过滤（CF）[24]，基于内容的过滤（CB）[4]，[13]或混合[6]，[20]，[22]方法来了解用户对参加过的POI的偏好并对未访问的位置进行预测。前一种方法基于以下假设提供建议：访问相同POI的用户最有可能在将来访问相同位置。例如，袁等人通过登记历史利用用户之间的相似性，并使用协同过滤[24]。另一方面，基于内容的方法使用与用户或POI相关的附加信息来解决稀疏性问题。同样，高等人 [4]通过调查与POI属性相关的内容信息的类型以及登记记录来利用LBSN的内容信息。为了克服每个方法分别面临的问题，例如：i）将POI作为节点处理并忽略地理邻近度，以及ii）缺少诸如社会影响或时间演化之类的其他维度，引入了混合方法。 接下来我们详细讨论了使用其他信息网络的模型。

**Social Influence.** 由于登记记录不能总是克服稀疏性和冷启动问题，因此许多方法使用社交关系，假设用户倾向于跟随朋友的领导[6]。例如，Li等人[8]区分了三种类型的友谊，即i）链接，ii）共处，iii）邻近的朋友，并通过统一的框架利用他们的登记记录。首先，该模型学习目标用户和所有三种类型的朋友过去登记的常见POI。然后，它使用矩阵分解来最小化学习POI上的两个损失函数以个性化推荐。同样，Zhang等人[25]引入了另一个名为LORE（位置建议）的统一模型，它将序列模式与社会和地理影响相结合。该模型通过具有加性马尔可夫链（AMC）的动态图来利用POI对用户记录的顺序影响。最后，它将所有上述影响结合到一个模型中，产品融合规则方程。

**Geographical Influence.** 为了进一步增强用户知识并消除稀疏性，许多方法使用地理信息。不幸的是，其中一些将POI视为常规物品[11]，并且错过了地点之间的地理位置。另一方面，最近的研究[18]，[19]，[22]，[23]考虑了地理上的接近程度并将其视为“空间项目”。 特别是，Wang等人[19]关注用户购买的空间项目的顺序影响的重要性，并提出了将空间项目和目标用户的偏好的顺序影响融合到其中的顺序持久化项目REcommender系统（SPORE）。 通过概率主题区域统一模型的相同潜在空间。扩展了之前与空间项目推荐相关的工作，引入了另一个概率模型[18]，将地理影响，项目属性和用户评论联合起来，形成一个名为LSARS的统一框架。两个模型都支持用户愿意与之互动的主张。近似项目，因此他们将地理影响与其他因素结合使用。

尹等人[23]声称用户保持相同的偏好要么在他们的家乡或他们访问新的地区。作者使用POI的空间属性来缓解城外用户的稀疏性和冷启动问题。地理影响（例如区域属性）用作辅助信息以将用户与新位置自动关联。该模型名为SH-CDL，共同学习这些属性以及用户登记偏好的附加表示。通过这种方式，他们可以在其他地区搜索接近用户过去偏好的邻近项目。 在相同的内容中，Yang等[22]开发了一种***偏好和上下文嵌入****（PACE）*模型，该模型共同学习用户和位置的嵌入。该混合模型以统一的方式将半监督学习（SSL）与协同过滤（CF）联系起来。

**Temporal Influence.** 最近的最新研究表明，用户的行为也存在周期性，这也应该与所有上述因素一起遇到[6]，[10]，[26]。袁等人 [24]研究了不同时段的时空行为，声称用户倾向于在不同的时间段访问特定的位置。他们提出了一个统一的模型，结合CF和贝叶斯规则，提供考虑接近度和周期性的POI建议。同样，Kefalas等人[6]提出了另一种混合模型，它以统一的方式结合CF和CB，探索i）近似用户的偏好，ii）时间段内的文**本影响交替**，以及iii）**偏好动态演化**。结果表明用户登记行为的演变，因为他们受到所有因素的综合影响。结果表明，对于分别考虑每个因素的模型，精度的增量稳健性。在同一个方向，刘等人 [10]开发了另一种混合时空感知模型，该模型学习用户，时空模式和POI的联合表示。

另一方面，遵循像word2vec [14]，赵等人的嵌入模型 [26]提出了Geo-Temporal顺序嵌入秩（Geo-Teaser），它结合了i）时间POI嵌入，以及ii）空间分层成对排序。特别地，前半模型基于时间POI嵌入模型来学习POI表示，该模型使用用户的每日登记历史作为序列。后半模型基于分层成对偏好中的地理信息对POI进行排序。同样，Xie等人[20]提出了一个名为GE的统一的基于图的模型，它通过将顺序效应，空间影响，周期性和语义捕获到相同的潜在空间中来联合学习POI嵌入。

与现有工作相比，我们提出了一个模型，将所有上述因素都纳入一个统一模型。特别是，我们应用概率策略来权衡边缘对八个信息图的重要性。然后，我们引入一个基于图的方法，联合 从这些加权图表中学习用户和POI嵌入到相同的潜在空间中，并提供个性化的POI建议。此外，我们还研究了针对冷启动用户的方法的性能以及冷启动POI问题的准确性。与现有技术相比，所提出的技术显示出显着的性能提升。据我们所知，这是文学中第一次同时面对这两个问题的尝试。

**3 PRELIMINARIES AND PROBLEM DEFINITION**

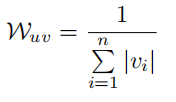
在本节中，我们将详细介绍问题定义，并讨论与我们的研究相关的一些基本概念。 图2描述了所有参与的网络，而表1列出了续集中最常用的符号。一些基本定义如下。

**Definition 1. (POI):** 是用户签入的唯一位置。POI表示为元组：<id、经度、纬度>。

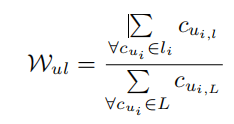
**Definition 2. (Check-in):** 是位于时间t的位置l的用户u的自报告定位，表示为元组：ci = <u，l，t>。一个签到只能由一个用户执行，但同一个用户可能有多个签到记录在配置文件中cui = {<li，ti>，...，<lj，tj>}。

**Definition 3. (Time Period):** 被定义为整个数据集的不同时间间隔，划分为相同大小的天，周或月。每个时段包含在该时间间隔ΔT期间执行的所有用户的签到。

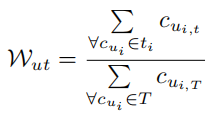
**Definition 4. (User-User graph):** 是一个描述用户社交网络的单独图。该图表示为Guu =（U∪V，Wuv），其中U和V是用户组，并且Wuv是其中的加权边缘的集合。该网络是用户之间友谊关系的无向图。因此，顶点之间的双向连接被描述为<u，v> = <v，u>，并且权重被定义为用户数量（v）的1的分数，他们是用户u的朋友：



**Definition 5. (User-POI graph):** 是一个有向的二分图，表示整个登记历史的用户和位置之间的关系。 特别是，它定义了目标用户的一个位置对所有其他位置的重要性。该图表示为Gul =（U∪L，Wul），其中U是用户集合，L是位置集合，Wul是其中加权边缘的集合。权重计算为分数用户访问所有签到位置的次数：



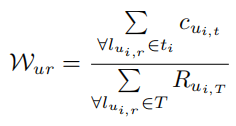
**Definition 6. (User-Time period graph):** 是一个有向二分图，表示用户在某个时间段的交互。 该图表示为Gut =（U∪T，Wut），其中U是用户集合，T是时间段集合，Wut是其中加权边缘的集合。权重计算为分数每个用户在一段时间内办理登机手续的次数，以防止他/她所做的所有登记：



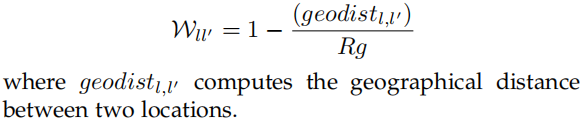
**Definition 7. (Route):** 是每个用户在时间段ΔT期间参加的POI的序列。例如，{t1：l1→l2→l3}表示用户u从位置l1移动到位置l2，然后在时间段t1期间在位置l3处结束他的路线。要提取每个用户的路线，我们首先根据时间缩短签到时间。然后，我们将数据集拆分为时间段，对于每个用户，我们在该时间间隔内构建路径序列。

**Definition 8. (User-Route graph):** 是一个有向二分图，它强调路线对于每个用户的重要性。首先，我们使用算法1从给定的POI中提取路线(如定义1所述)。

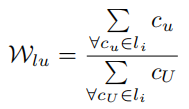
然后，我们将来自U的每个用户与来自集合R的一个路线相关联。该关系表示为Gur =（U∪R，Wur）。 两个节点之间边缘的权重计算为一个用户遵循一个特定路线到所有时间段内同一用户所做的总路径或路由的次数：



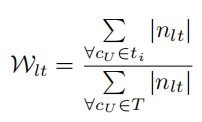
**Definition 9. (POI-POI graph):** 是指向两个位置之间空间接近度的双向二分图。 特别地，当且仅当一个用户在时间段t和距离范围Rg期间签入它们时，两个位置l和l’与边缘连接。根据该假设，我们构建表示为Gll0 =（L∪L，Wll’）的图，其中L是位置集，并且Wll’是考虑到它们的地理邻近度的两个位置节点之间的加权边的集合。此权重计算如下：



**Definition 10. (POI-User graph):** 也是一个有向二分图，它指出了位置和用户之间的关系。 该图的主要区别在于我们遵循的加权策略，因为位置对用户的影响与先前的方法不同。 特别地，该图表示为Glu =（U∪L，Wlu），其中L是位置集合，U是用户集合，并且Wlu是它们之间的加权边缘的集合。权重计算为用户访问位置l的次数与登记该位置的所有用户数的比例：

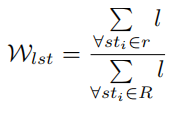


**Definition 11. (POI-Time Period graph):** 是一个有向二分图，表示一个时间段内位置的重要性。该图表示为Glt =（L∪T，Wlt），其中L是位置集合，T是时间段的集合，并且Wlt是它们之间的加权边缘的集合。权重被计算为在时间段ti期间由所有用户在位置l处执行的登记的次数与在相同位置的所有时间段期间的登记总次数的分数：



**Definition 12. (POI-Stay Points graph):** 是一个双向二分图，描述了用户某些位置的重要性。在每日计划中，用户遵循一些路线并在该路线序列的每个位置花费一些时间。每个位置的经过时间表示该位置对该特定用户的重要性。因此，在该位置上花费的时间越多，该位置的重要性越高。为了提取这些重要的位置，表示为“停留点”（sti），我们使用算法2。

然后我们构造Glst =（L∪ST，Wlst）图，其中L是位置集，ST是停留点集，Wlst是其中加权边的集合。权重计算为分数 一个位置被视为路线中的停留点的次数，与此位置在所有路线中被视为停留点的总次数相对应：



**Problem Definition:** “给定用户u，位置l和时间实例t（表示为Q（u，l，t））和登记历史记录，预测该目标用户的前N个未访问的邻近POI。”

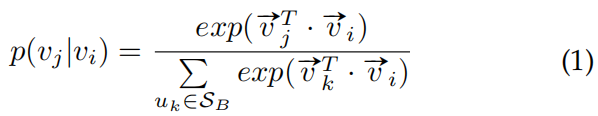
**4 PROPOSED APPROACH**

接下来，我们介绍RELINE，这是一个优化的解决方案，用于在相同的潜在空间中联合学习不同信息网络的图形嵌入，并且我们提出了POI推荐的统一框架。

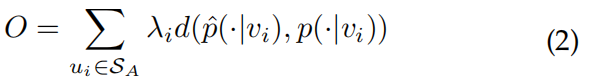
**4.1 Learning Embeddings in a Bipartite Graph**

直接连接到边ei,j和权重wi,j的节点由**一阶接近度**组成，即两个节点之间的局部成对接近度。另一方面，共享许多连接但是它们不与边缘直接相关的节点，它们属于相同的邻域，因此，最可能彼此相似。这些节点包括**二阶邻近度**，即两个未链接节点之间根据其网络结构的相似性。为了在单分图上提取这种接近度，LINE模型[16]学习将大图嵌入到低维空间中。通过这项工作，我们扩展了这个模型，以学习二分图节点上的嵌入。此外，我们的模型可以应用于所有类型的二分图，即有向/无向，加权/未加权或它们的组合。

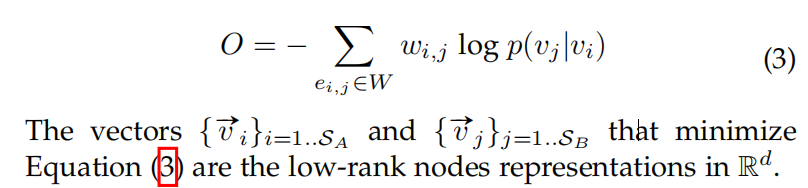
给定两个分离集G =（SA∪SB，W），SA中与SB共享许多连接但彼此不直接连接的节点最有可能具有相同的分布。一个节点vj∈SB的条件概率是通过节点∈SA生成的，例如：



其中顶点和的嵌入向量分别表示为，。 因此，对于每个节点∈SA，等式（1）将条件分布p(·| )定义为集合SB中的所有对应节点。然后，对于每个边缘，存在一个重量，这意味着该平局的强度。 为了保持SA中未链接节点的接近度，我们让所描述的条件分布近似于经验分布，具有以下目标函数：

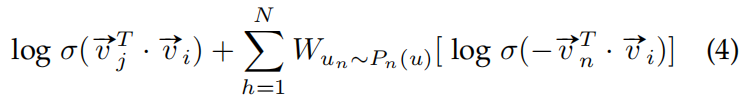


其中d（·|·）表示条件和经验分布的Kullback-Leibler散度，λi是调整节点vi的重要性的正则化参数。为简单起见，我们将此参数设置为等于每个节点的out-degree。因此，等式（2）对应于以下目标函数的最小化：

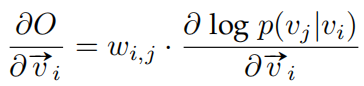


**4.2 Model Optimization trough Negative Sampling**

为了避免计算需要等式（3）中整个节点集合的条件概率p（·| vi），我们在每个边缘上应用负抽样（NEG）[14]。特别地，我们使用每个边缘的噪声分布来对N个负边缘进行采样，如以下目标函数中所述：

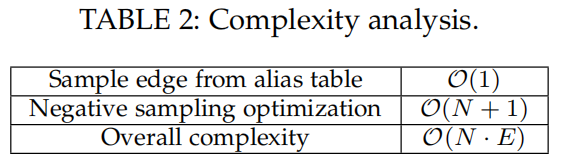


其中σ(x)是具有输出值[0-1]的S形函数，而是噪声分布，其中负样本用[14]中的经验调整的一元分布选择，例如 那集合中的每个节点出现都与所有其他节点出现无关。因此，选择节点作为负样本与该功率的出度有关。为了进一步改进方程（4）的解，我们应用基于异步随机梯度下降（ASGD）的'Hogwild'算法[15]。特别是，每次对边进行采样时，我们计算节点相对于相应嵌入向量的梯度，如下所示：



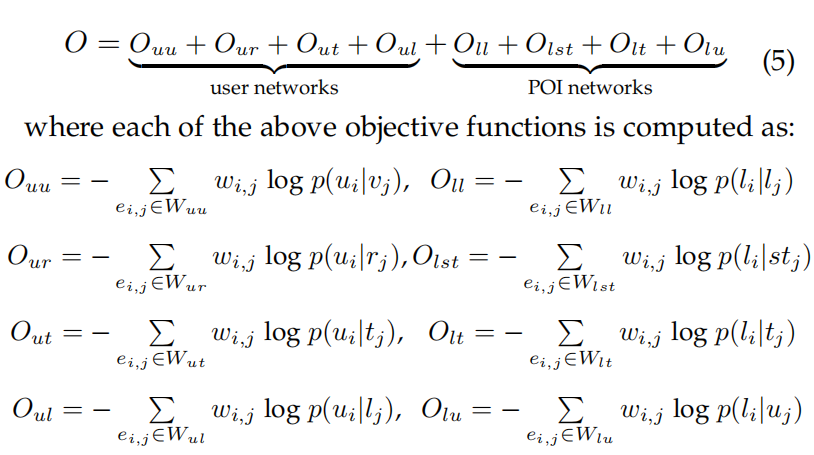
请注意，节点vi的梯度乘以与该边相关的权重。因此，调整模型的学习速率可能由于权重的有效性而导致问题。一方面，如果根据边缘乘以小权重值选择较大的学习速率，则可能会对具有较大权重的梯度发生“过度拟合”。另一方面，如果权重较小的梯度可能会出现“欠拟合”，如果选择小学习率用于边缘乘以大权重值。

为了平衡学习率或我们的模型，我们采用了[16]中提出的抽样方法。特别地，我们采样随机边，其中表示特定网络中所有权重的总和，然后我们检查特定采样边缘落入的间隔，即。最后，我们根据[7]使用别名表绘制采样边缘，最终将复杂度降低到O(1)。 表2列出了边缘采样优化的全部复杂性。



**4.3 Joint Learning of Graph Dynamics**

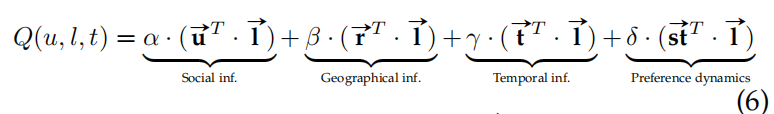
给定输入的二分图，下一步是将它们集成到我们的模型中。 与用户关系对应的图形是：用户-用户，用户-兴趣点，用户-路径和用户-时间段。另一方面，对应于与其他网络的位置关联的图表是：POI-POI，POI-用户，POI-重要位置和POI-时间段。我们通过最小化目标函数，共同整合这些图的嵌入对应于用户和POI的关系：



为了最小化等式(5)中呈现的目标函数，首先我们将所有单面和二部图的边合并在一起，然后，在每个步骤中，我们通过对新边进行采样来更新模型。采样边缘的概率对应于与该边缘相关的权重。这样，我们的模型就图表的内部和外部顶点以及权重影响遍历异质二分图。 如算法3所示，模型的训练是联合和动态完成的。

**4.4 Unified Model for POI Recommendations**

当前一部分中呈现的所有嵌入已经被学习并且给出关于时间戳t处的位置l中的用户u的预测请求Q（u，l，t）时，我们将这些值投影到相应的时间。周期t，路线r和停留点st，地理距离小于10公里。我们声称一个用户愿意去附近的地点。因此，给定超出该范围的推荐，参与的概率非常小。然后，我们对该距离中该用户的Top@n未访问候选POI进行排名。 每个未访问位置的预测分数计算如下：



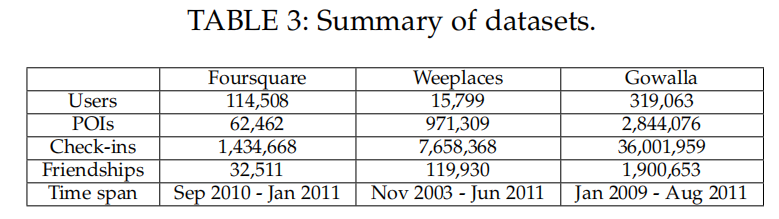
其中是用户u的嵌入，是位置l的嵌入，是此登记所属的路径r的嵌入,是时间段t的嵌入 特定的登记入住，是住宿点的嵌入。 此外，RELINE在同一潜在空间中共同学习来自不同信息网络的嵌入。因此，学习的POI嵌入捕获前面部分中呈现的所有参与网络的信息，例如uu,ul,ll等。这样，我们的目标是通过使用附加信息网络来消除稀疏性。特别是，我们共同学习社会影响的动态，地理影响，时间影响，以及用户的偏好动态。最后，α，β，γ和δ是正则化参数，它们将每个信息网络的重要性分别定义到我们的模型中。

**5 PERFORMANCE EVALUATION**

在本节中，我们将重点关注所提方法的性能。 特别是，我们将我们的技术与以前的技术进行比较 我们的方法的源代码可以在https://github.com/thedx4/RELINE上找到。

**5.1 Datasets and Techniques**

我们使用了三个真实世界的数据集：i）Foursquare [20]，ii）Weeplaces [1]，和iii）Gowalla [1]。它们的主要特征如表3所示。所有数据集都包含用户的登记历史记录，包括时间戳和地理信息。此外，它们还包含有关用户友谊的社交关系的信息。数据集分别跨越5,91和31个月的时间段。



在图3中，我们呈现了所有三个数据集的分布。如图所示，数据集遵循用户签到次数和特定位置访问次数的幂律分布。根据幂律分布，有少量用户有很多签到(短头)和许多用户有一个小的登记记录 (长尾)。类似地，同样的原则适用于一方面，有大量访问的热门位置，另一方面访问次数较少的位置。此外，所有数据集都是冷启动问题的一个很好的例子[5]，即向具有小历史记录的用户推荐新位置。

此外，我们提出了图4中所有数据集的时间分布。特别是，图4（a）显示了每周签到的分布情况。值得注意的是，用户在周四到周日比一周的休息时间更活跃，由于用户在业余时间办理登机手续的事实，这个数字反映了他们在一周结束时用来执行更多登记的趋势。如图4（b）所示，相同的原则代表白天签到的分配。我们再次观察到用户在下午13:00更活跃。到凌晨2点，高峰时间是晚上18点。 而且他们倾向于在夜间比在早上更多地点办理登机手续。

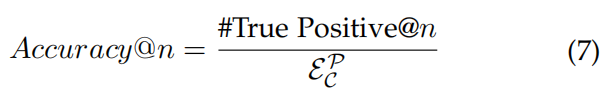
在实验中，我们评估了以下技术的性能：

* **Rank-GeoFM [9]:** 是一种MF统一模型，它可以学习用户的偏好并同时结合邻近POI的空间影响。模型的第二项使用与目标PO I相关的基于距离的加权进行正则化。
* **ASMF [8]:** 是另一个MF模型，由两步程序组成。首先，该模型学习已被朋友访问的候选POI（社交影响），然后考虑地理影响来应用基于分类的权重。
* **GE [20]:** 是图形因子分解（GF）方法，根据该方法计算每对节点的两个联合概率分布。第一个与邻接矩阵有关，第二个与嵌入有关。该方法将四个二分信息网络嵌入到相同的潜在空间中，以通过动态更新用户的信息来预测下一个未访问的POI。
* **PACE [22]:** 是一种基于用户偏好和上下文嵌入的半监督学习框架，它共同学习用户的嵌入和位置嵌入。
* **Versions of RELINE:** 为了评估每个网络的影响，我们使用了三个版本的模型。特别是，我们从一个简单化的版本开始，然后进行更多增强的替代方案。
* **RELINEV1:** 是reline的简化版本，它只包含社会影响信息，即
* **RELINEV2:** 是前面的模型丰富了地理影响。
* **RELINEV3：**是最后一个版本，它丰富了前一个模型的时间影响。

**5.2 Evaluation Methodology**

我们考虑将每个目标用户的签到划分为三组：（i）训练集，即总签入的80％并被视为已知信息，（ii）探测集，即 10％用于测试我们的模型，以及（iii）验证集是调整超参数的其余10％。它认为，和。因此，对于每个目标用户，我们仅基于中的POI生成推荐。

对于评估，我们测量[19]中提出的Accuracy @ n。特别地，对于作为查询Q(u，l，t)给出的每个，我们利用等式（6）计算该l的预测分数以及目标用户的所有未访问的邻近POI。我们将预测得分排列到一个列表中，然后我们选择top@n POI。如果地面真值出现在top@n，那么我们已经正确地预测了该位置（即真正的正），否则我们的预测是错误的。为了计算top @ n的总体准确度，我们将所有预测测试用例平均为：



**5.3 Impact of Information Networks**

接下来，我们检查每个信息网络对整体预测准确性的影响。特别是，我们探讨了浓缩对前n个预测的有益程度。本节中考察的参与者网络是：i）社会影响，ii）地理影响，iii）时间影响，以及iv）用户的偏好动态。 因此，我们将RELINE与RELINEV1，RELINEV2和RELINEV3三种模型进行比较，这些模型在5.1节中描述。值得注意的是，只要我们嵌入更多信息网络，准确度也会提高，如表4（a）-4（c）的每一栏所示。因此，我们的模型逐渐缓解了稀疏性问题，因为它探索了更多关于用户或POI的信息。此外，每个模型的准确性随着n增加，这意味着模型很好地适合用户的行为。

**5.4 Comparison with Other Techniques**

在本节中，我们将模型与其他最先进的方法在前n个预测[n = 1,5,10,15,20]与3个大数据集的准确性方面进行比较。特别是，我们检查所有模型的性能，为i）所有用户，ii）冷启动用户和iii）冷启动POI提供POI建议。

**5.4.1 Accuracy over all users**

首先，我们检查模型在向所有用户提供建议的同时如何执行，而不考虑历史大小或数据稀疏性。据观察，RELINE明显优于所有方法，如图5所示。

与a）学习用户偏好然后结合Rank-GeoFM等地理影响的方法相比，或b）学习ASMF等社交网络签到并忽略i）POI的顺序，ii）时间影响，以及iii 由于我们的方法在相同的潜在空间中探索更丰富的信息，因此偏好动态，性能要高得多。另一方面，通过多个信息网络（如GE和PACE）学习用户和POI嵌入到同一潜在空间的方法错过了其他重要因素，例如i）周期性，ii）偏好演变，以及iii）顺序重要性。与我们的方法相比，签到并获得更低的准确性。我们强调所有模型都显示出更高的准确性，而POI的数量很少，登记活动更密集，如图5（a）所示。此外，当数据集稀疏且要访问多个可用POI时，例如图5（b）-5（c）中呈现的其他两个数据集，准确度较低。这一发现支持了这样的主张，即学习尽可能多的信息网络的用户和POI嵌入，增加了模型将用户与POI相关联的能力，并最终提高了准确性。

**5.4.2 Accuracy over the cold-start Users**

接下来，我们将检查模型在冷启动用户方面的有效性，并与其他方法进行比较。这个概念最初是在[5]中引入的，指的是历史较短的用户。很明显，由于缺乏足够的信息，为这些用户提供支持是一项艰巨的任务。在这种情况下，我们进行了实验，仅向冷启动用户提供建议，我们在准确性方面比较了模型的性能，如图6所示。由于只有ASMF和GE支持冷启动建议，我们将模型与这两种方法。此外，我们观察到与前一部分的实验相比的总体准确性降低了，这是正常的，因为我们没有太多信息。即使ASMF学习目标用户的社交网络访问的POI，然后使用分类加权策略w.r.t优化结果。地理影响，与我们的方法相比，其准确性显着降低。类似地，GE共同捕获空间影响，顺序效应，周期性和语义到同一潜在空间的性能较低，因为它没有考虑：1）每个序列中停留点的重要性，2） 偏好动态，3）时间效应，4）社会影响。相反，我们使用来自八个加权信息网络的与用户和POI相关的辅助信息，这对有效性具有显着影响。

**5.4.3 Accuracy over cold-start POIs.**

接下来，我们将研究一个名为冷启动POI的类似问题。在这种情况下，目标是向在少于15次签到的POI中至少办理一次登记的用户推荐未访问的POI。因此，我们不仅检查模型在新用户上的行为方式，还检查新位置进入系统时的行为方式。简单来说，我们想要检查新位置是否属于前n个推荐。

我们再次仅使用支持冷启动POI问题的模型来评估我们的方法。在图7中，我们展示了针对所有三个数据集的所有模型的结果。显然，我们的模型优于比较方法，因为在所有上述因素中，它将POI作为一系列路径进行探索。特别是，当我们提取用户的路线时，所有冷启动POI都与其他邻近的POI相关联。此外，根据用户在该位置上花费的时间，其中一些被视为“停留点”。然后，我们权衡用户和冷启动POI之间的边缘，作为一个重要的边缘，如果该用户在那里花费了很多时间。这样，我们通过在学习阶段使用它们与图上其他节点的关系影响来处理冷启动POI。 值得注意的是，如图7（a）-7（b）所示，当POI的数量较小时，两种比较模型都获得较低的前1预测准确度，因为它们探索未访问过的用户的朋友签到 冷启动POI，或者他们基于的探索POI序列用户检查冷启动POI的频率。在GE使用的POI-POI图中，几乎没有签到的POI没有连接到其他POI。 最后，两个比较模型似乎在系统上的许多POI上获得更高的准确度，例如图7（c）中指出的Gowalla数据集。

**5.5 Parameter Tuning**

在本节中，我们将研究参数调整的重要性。特别是，我们研究了以下因素的影响：1）向模型添加信息网络，2）样本数量S，3）嵌入维数d，以及4）时间段大小T，以我们的模型的性能表示 准确性。

**5.5.1 Impact of Samples and Dimensions number**

在这里，我们展示了为实现样本数量和维度选择最佳候选参数而进行的实验。每个数据集的结果显示在表5中。我们对top@ 10的发现表明我们的模型不受维数d的极大影响。对于Foursquare和Weeplaces，精度随着速率和维度的增加而增加，直到d = 100，对于Gowalla，d = 140。此后，准确度不会发生显着变化。与维度相反，我们的模型对样本数量敏感（S）。在达到收敛点之前，我们的模型随着S的大小不断提高其准确性，然后改善很差。此外，网络边缘之间的连接数越高，样本的数量越多。为了获得更高的精度，我们将Foursquare的样本数量设置为100，Weeplaces为200，Gowalla为300，以及前面讨论的维数。

**5.5.2 Impact of Time Period**

接下来，我们检查时间间隔鈭员对顶部@n预测的变化值的总体准确性的影响。ΔT对我们的模型至关重要，因为它用于构建多个图形，如用户路线，用户时间段，POI-时间段等。当提取用户路线边缘时，如果在小的时间间隔期间没有足够的登记数据，则难以实现该用户与其他候选POI的相关性。为了克服这个问题，我们使用不同的尺寸来检测ΔT以获得更高的精度。表6分别显示了每个数据集的结果。值得注意的是，每个子表中存在精度达到其最大值然后逐渐减小的点。 原因是，当ΔT太小时，数据很少而且精度很低。另一方面，当ΔT较大时，存在与目标用户相关的太多节点，这导致过度拟合。 因此，我们将Foursquare，Weeplaces和Gowalla的大小分别设置为20,40和15。

**5.5.3 Tuning parameters α, β, γ, and δ**

在图8中，我们给出了根据所有数据集的准确度调整参数α，β，γ和δ的结果。如等式（6）所示，每个参数分别对应于社会，地理和时间影响以及偏好动态。为简单起见，我们将每个参数p设置为[0-1]和所有其他参数之间的值，等于（1-p）/ 3。观察到在每个图中存在精度满足的交叉点。 此外，还有一个峰值点，我们的模型获得更高的精度，用于训练我们的模型。 此外，每个有影响力的网络的重要性是多样化的，试图使模型适应用户的行为。

如图8所示，有影响的参数可能不同。例如，关于Foursquare，用户受其社交网络的影响很大，如图8（a）所示。因此，他们倾向于访问他们的朋友访问的位置。另一方面，Weeplaces用户倾向于每天遵循特定路线，因此他们到新位置的移动受地理因素的影响很大，如图8（b）所示。最后，在具有许多签到的数据集中，如Gowalla，很明显，在预测下一个位置时捕获偏好动态影响更为重要，如图8（c）所示。

**6 CONCLUSIONS**

用户在OSN中的互动和他们提供的大量信息的快速增长，使研究人员找到了能够检索接近他们过去偏好的个性化信息的模型。不幸的是，这些模型是由稀疏性和冷启动问题解决的。与这些网络中发布的内容相关的地理信息触发了新的功能和研究方向，以解决这两个问题。有许多模型使用社交网络影响或用户当前位置的地理邻近度来提供POI建议。然而，这些模型错过了时间维度或用户偏好动态的影响，这对于个性化检索的信息是至关重要的。

在本文中，我们提出了一个新的模型，该模型在提供POI建议的同时考虑了所有上述因素。特别地，我们的模型使用概率加权策略来对应于用户和POI关系的8个信息图。 然后，它使用基于图形的方法，将这些加权图形中的用户和POI嵌入联合学习到相同的潜在空间，并提供个性化的POI建议。我们从准确性的角度考察社会，地理，时间和偏好动态的影响。 我们将我们的方法与四种最先进的模型进行比较，这些模型测量i）所有用户，ii）冷启动用户和iii）冷启动位置的建议的准确性。我们的方法明显优于最先进的方法。