

面向数据匮乏城市的下一个 POI 推荐方法^①

谭海宁^{②*} 姚迪^{***} 毕经平^{③*} 向徐^{***} 杨啸^{***}

(* 中国科学院大学 北京 100049)

(** 中国科学院计算技术研究所 北京 100190)

(*** 盲信号处理国家级重点实验室 成都 610041)

摘要 位置社交网络(LBSN)用户位置数据的分布不均衡,及某些用户出于对隐私安全的考量刻意隐藏自己部分位置信息等因素加剧了兴趣点(POI)推荐难度。就此本文提出了基于元学习的时空神经微分方程(ML-ODE)来进行有效的下一个 POI 推荐。该模型主要是将元学习的思想融入到 POI 推荐过程中,通过不同任务训练优化初始参数,将数据丰富城市中的泛化移动模式迁移到数据匮乏城市,达到优化 POI 预测任务的目的。该模型将神经微分方程用于 POI 推荐领域,定义连续的动态过程,可以接受任意时刻的输入数据,克服了大多数时序推荐模型静态离散化的时间间隔处理方式,更适用于 POI 序列推荐任务。在真实公开数据集 Foursquare 上的实验结果表明,ML-ODE 在 POI 推荐方面比当前主流的 POI 预测方法在 NDCG@N 指标上提升了超过 10%。

关键词 兴趣点(POI)推荐;位置社交网络(LBSN);元学习;神经微分方程;推荐系统

0 引言

随着各种具有定位服务的移动设备的普及,基于位置服务的移动社交网络(location-based social network, LBSN)发展迅速,具有位置信息的用户数据呈爆发式增长,如 Foursquare、Yelp、美团等电商服务平台^[1-2]。有效地利用具有位置信息的用户数据,探索用户兴趣点(point of interest, POI)已经成为移动社交网络数据挖掘领域的重要研究方向。其中,由于推荐系统是解决信息过滤和个性化服务问题的主要手段之一,其在用户位置挖掘服务中发挥着重要作用,面向用户兴趣点的推荐方法研究也成为移动社交网络挖掘领域的重中之重^[3-5]。

兴趣点表示用户感兴趣的已经访问或将会访问的地理位置^[1-2,6-7]。推荐的潜在兴趣点表示用户在将来某个时间点可能会访问的地理位置。作为面向

位置服务的移动社交网络数据挖掘领域的重要分支,POI 推荐可以预测用户行为和兴趣演化趋势,对城市交通的规划和商业模式的探索具有重要的现实意义。POI 推荐有着广泛的应用,例如大众点评网站的餐厅推荐^[4]、旅游路线规划^[5]、城市交通事件预警^[8-9]等。

当前大多数的 POI 推荐方法都假设 POI 数据是同分布的,将 POI 数据看成一个整体进行用户 POI 兴趣的挖掘工作,比如 STRNN^[1]、DeepMove^[10]、ST-LSTM^[11]等。这些方法将用户的签到 POI 数据看成一个整体,没有考虑轨迹数据的城市不均衡特性,缺乏对城市之间共性信息和城市本身特殊性信息进行区分的个性化建模,这样导致在数据匮乏的城市,由于 POI 历史记录少,对用户进行推荐难度大、效率低。本文分析了 Foursquare 收集的截止到 2014 年 1 月的用户签到 POI 数据的全球位置数据。从数据分

① 国家重点研发计划(2017YFC0820700)和国家自然科学基金(61472403)资助项目。

② 男,1991 年生,博士生;研究方向:时空数据挖掘,推荐系统,大数据;E-mail: tanhaining@ict.ac.cn。

③ 通信作者,E-mail: bjp@ict.ac.cn。

(收稿日期:2020-10-10)

布可以看出受到采集手段、隐私保护等因素的影响,用户 POI 签到数据在城市之间分布极其不均衡。北美、欧洲及各大洲沿海发达城市 POI 数据分布比较集中,其他城市 POI 数据分布非常稀疏。

因此,亟需探索如何在数据匮乏的城市进行 POI 推荐的有效方法^[12-14]。针对如何解决数据匮乏的问题,有少量的工作通过关联相关用户在目标城市外的历史信息进行推荐,比如 JIM^[15] 和 JFT^[16]。这些工作都是基于源城市和目标城市之间存在共同用户的前提下,将关联用户在源城市的行为信息进行迁移,完成关联用户在目标城市的 POI 兴趣发现和推荐任务。这类方法受限于事先设定的影响因素,无法对隐含因素进行建模,并且在城市之间的信息共享受共同用户这一假设制约,导致推荐方法应用范围受限,推荐效果不佳。

本文为了缓解数据分布不均衡等因素给 POI 推荐带来的影响,借助于元学习^[17]的思想,通过将数据丰富城市的泛化共性移动知识进行学习并迁移应用到数据匮乏城市的 POI 推荐任务中。与此同时,由于采集的 POI 数据的时间点并不是严格规整的,即并不是按照固定的时间间隔进行采集,这就造成 POI 数据之间时间间隔的不确定性。传统的序列化建模方法将时间间隔分割成不同的时间槽或者直接按照先后顺序进行建模,忽略了 POI 序列之间的连续时间信息,因此序列推荐效果受到严重制约。文献[18]在其工作 Neural-ODE 中指出了在序列化推荐任务中,连续时间建模对推荐效果的性能有明显提升,并提出了可以进行连续时间建模的新的深度网络,即神经微分方程。神经微分方程使用神经网络参数化隐藏状态的导数,而不是直接参数化隐藏状态。这里参数化隐藏状态的导数就类似构建了连续性的层级与参数,而不再是离散的层级。因此参数也是一个连续的空间,不需要再分层传播梯度与更新参数,这样可以更加高效地进行连续时间建模。本文受神经微分方程^[18]的启发,提出了基于元学习的时空神经微分方程(meta-learning based ordinary differential neural network, ML-ODE)进行有效的 POI 推荐。

本文的主要贡献如下。

(1) 借助于神经微分方程实现了对 POI 数据的连续时间序列建模,可以忽略传统方法对输入数据的时间间隔的约束。

(2) 在神经微分方程中融入时空注意力机制,借助注意力机制衡量不同时刻的签到行为与用户偏好的相关性,进而获得不同签到行为的权重系数。

(3) 提出了基于元学习机制的 POI 序列推荐框架,实现了将泛在共性移动知识从数据丰富城市迁移到数据匮乏城市。

(4) 在真实数据集上进行了大量实验,实验结果表明,本文提出的 POI 推荐算法可以有效地对数据匮乏城市的目标用户进行 POI 推荐。

本文结构如下:第 1 节介绍了面向位置社交网络 POI 推荐的相关工作,包括基于矩阵分解和基于深度学习的相关方法;第 2 节介绍了本文使用的基本符号和相关问题定义,并详细描述了本文提出的 POI 推荐框架 ML-ODE;第 3 节通过实验对本研究中提出的方法进行了有效验证,并报告相关的实验结果;第 4 节对本文的工作进行了总结并展望该技术的未来发展方向。

1 相关工作

本节详细介绍了目前主流的面向位置社交网络的 POI 推荐方法以及与神经微分方程相关的最新研究成果。目前,POI 推荐方法主要可以分为面向单城市的 POI 推荐方法和跨城市的 POI 推荐方法。

面向单城市的 POI 推荐方法,主要有 FPMC^[19]、STRNN^[1]、DeepMove^[10]、AT-LSTM^[20]等。FPMC 通过引入基于马尔可夫链的个性化转移矩阵,进行时间信息和用户的长期兴趣偏好的捕捉,同时为了解决转移矩阵的稀疏局限性,引入了矩阵分解模型,可以减少参数并提高 POI 推荐效果。但是由于马尔可夫链的不同影响因素之间的独立性假设,并且受矩阵分解方法冷启动的影响,该方法具有一定的局限性。STRNN 通过扩展循环神经网络(recurrent neural network, RNN)^[21]提出了一种时空循环神经网络模型,该方法通过引入时间转移矩阵和空间距

离转移矩阵分别建模局部时间上下文和空间上下文,通过设置不同的窗口确定恰当的时空上下文邻居,分别计算时间距离 t_{dis} 和空间距离 l_{dis} 来作为转移矩阵的输入,规整后统一输入循环神经网络。DeepMove 提出了一种基于注意力机制的多模态循环神经网络,首先通过门控循环单元(gated recurrent unit, GRU)层建模用户即时偏好;其次通过历史注意力机制模块对用户的历史偏好进行建模;最后通过整合多层注意力机制提升用户移动模式预测效果。AT-LSTM 在长短期记忆网络(long short-term memory, LSTM)的基础上借助注意力机制进行时空信息长期偏好和短期偏好的捕捉,提升了推荐效果。

该类方法将用户的签到 POI 数据看成一个整体,没有考虑轨迹数据的城市不均衡特性,缺乏对城市之间共性和城市本身特殊性的区分性建模,这导致在数据匮乏的城市 POI 推荐效率低下。

面向跨城市 POI 推荐的方法主要有 JIM^[15]、ST-LDA^[22]、CTLM^[23] 等。文献[15]提出了一种可以用于跨城市的 POI 推荐的图模型 JIM,通过联合考虑 POI 内容、check-in 时间、POI 地理位置以及 POI 的流行度等因素,实现跨城市推荐。ST-LDA 是 JIM 工作的扩展,增加了对用户在不同城市的兴趣漂移因素的建模,通过叠加用户的自身兴趣偏好和用户在某城市的兴趣偏移实现对目标用户的个性化 POI 推荐。文献[23]提出了一种基于共同话题迁移学习模型的跨城市 POI 推荐模型,该方法将城市的 POI 按照各自属性信息分为城市间共有类(common topics)和城市内私有类(city-specific topics),将用户在别的城市与共有类的交互信息迁移到目标城市。

该类方法都是基于源城市和目标城市之间存在共同用户的前提下,将关联用户在源城市的行为信息进行迁移,导致推荐方法应用范围受限,推荐效果不佳。

神经常微分方程在模型训练的过程中,不使用其他 RNN 以及卷积神经网络(convolutional neural network, CNN)等直接参数化隐藏状态的处理方式,直接使用神经网络参数化隐藏状态的导数。因此,神经常微分方程的参数可以看成是一个连续的空间,不再需要分层进行梯度传播并更新参数。目前针对

神经常微分方程的工作还相对较少,主要有面向时序序列建模的 Neural-ODE^[18,24]、GRU-ODE^[25] 等。文献[18]提出神经常微分方程 Neural-ODE,这是一种新的深度神经网络,借助于常微分方程实现对序列数据的连续建模。GRU-ODE 结合 Neural-ODE 和 GRU 进行时间序列的连续性建模,根据输入离散的时间点观察值,通过 GRU-Bayes 模块更新 Neural-ODE 之间的隐状态向量。

该类方法目前还没有用来建模 POI 数据,并且不能直接迁移应用于跨城市 POI 推荐应用中。

针对现存 POI 推荐方法存在的问题,本文提出了基于神经常微分方程的跨城市 POI 推荐方法,将数据丰富城市的泛化共性行为知识迁移到数据匮乏的目标城市,不仅缓解了目标城市的数据匮乏问题,也实现了对序列数据的连续性建模。

2 时空神经常微分方程

本节将介绍所提出的面向数据匮乏环境的下一个 POI 推荐方法。

2.1 基本定义

对位置社交网络中的用户进行 POI 推荐时,通常用数值或者向量来表示网络中的实体。例如,在描述用户签到数据时,签到地点用经纬度数值二元组进行记录,时间用数值进行记录,用户的属性信息也会转换成属性向量进行表示。本文研究的问题是如何构建一个合适的用户 POI 兴趣学习模型,借助用户的历史签到记录,为其更好地进行 POI 推荐。在详细描述 POI 推荐算法之前,先给出一些基本的符号解释和定义。

定义 1 兴趣点(POI)^[22]。在位置社交网络中,一个兴趣点 $v_{i_k}^u$ 表示某一地理位置,通常由一对地理经纬度坐标以及相关的属性信息表示。

定义 2 签到行为(check-in activity)^[22]。在位置社交网络中,用户 u 的签到行为记录可以看成是一个三元组 $q_{i_k}^u$, 这里可以形式化为

$$q_{i_k}^u = (u, v_{i_k}^u, t_k)$$

其中, $q_{i_k}^u$ 表示用户在时刻 t_k 拜访了 POI $v_{i_k}^u$ 。

定义 3 签到序列(check-in sequence)。在位

置社交网络中,某用户 u 的签到序列由该用户一系列的历史签到行为构成,可以形式化为

$$Q_u = q_{t_1}^u, q_{t_2}^u, q_{t_3}^u, \dots, q_{t_k}^u$$

其中, k 表示用户 u 的历史签到行为总数。为了方便,所有用户的历史签到序列可以形式化为

$$Q = Q_{u_1}, Q_{u_2}, Q_{u_3}, \dots, Q_{u_{|U|}}$$

这里, U 表示用户集合。

定义 4 下一个 POI 推荐(next POI recommendation)。给定用户签到数据匮乏的目标城市集合 $D = D'_1, D'_2, D'_3, \dots, D'_M$, 用户签到数据丰富的源城市集合 $S = S'_1, S'_2, S'_3, \dots, S'_N$, 满足:

$$|Q^{U_{S'_n}}| \gg |Q^{U_{D'_m}}|, n \in 1, 2, \dots, N$$

$$m \in 1, 2, \dots, M$$

本文的目标是为目标城市 D'_m 中的用户 u 推荐其下一个可能感兴趣的 POI $v_{t_{k+1}}^u$, 可以将推荐过程形式化为

$$v_{t_{k+1}}^u = \operatorname{argmax}_{v_{t_{k+1}}^u} f(v_{t_{k+1}}^u | v_{t_1}^u, v_{t_2}^u, \dots, v_{t_k}^u, \theta)$$

其中, f 表示 POI 推荐模型, θ 表示模型相关参数, 模型整体框架图如图 1 所示。

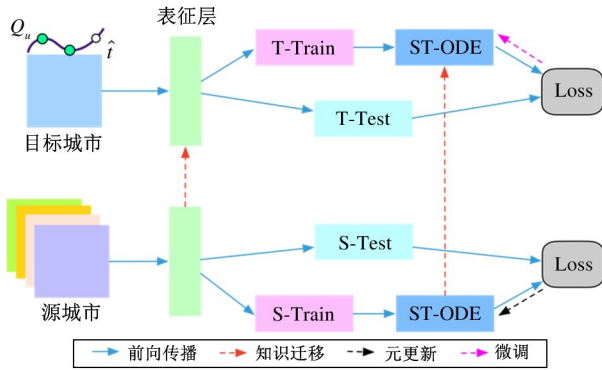


图 1 面向数据匮乏环境的下一个 POI 推荐框架

2.2 基于时空注意力机制的神经微分方程

本节将介绍提出的基于注意力机制的时空模型 ST-ODE。如图 2 所示,在给定用户 u 的一条轨迹 $Q_u = q_{t_1}^u, q_{t_2}^u, q_{t_3}^u, \dots, q_{t_k}^u$, ST-ODE 首先获取每一条签到记录的特征向量,然后通过时空神经微分方程建模轨迹序列的时空影响,最后通过时空注意力机制获取 u 在 t_{k+1} 时刻的状态向量。接下来,将分别介绍签到行为表征、时空神经微分方程以及时空注意力机制 3 个模块。

2.2.1 签到行为表征

在表征模块中,利用一个嵌入层获取签到行为序列中每一个签到行为的表征向量,如图 1 所示。在特征层面,仅考虑 POI 的类别、POI 位置以及时间戳 3 类特征,即 (c_k, l_k, t_k) 。其中 c_k 表示该签到行为所在 POI 的所属类别, l_k 表示该 POI 的所在位置,由经纬度坐标表示, t_k 表示发生该签到行为的时间点。

(1) 针对 POI 类别属性 c_k , 引入矩阵 $E_c \in \mathbb{R}^{N \times d_c}$, 其中, N 表示 POI 的类别数, d_c 表示 POI 类别表征向量的维度,这里矩阵 E_c 在多个城市之间进行共享。

(2) 针对 POI 地理位置属性 l_k , 引入矩阵 $E_l \in \mathbb{R}^{M \times d_l}$, 其中, M 表示 POI 的个数, d_l 表示 POI 位置属性表征向量的维度,通过经纬度坐标位置的 Geo-Hash 编码来初始化 POI 矩阵 E_l 。

(3) 针对时间戳 t_k , 引入矩阵 $E_t \in \mathbb{R}^{168 \times d_t}$, 其中 168 表示将一周分成 168 个时间槽 (7×24 h), 每一个时间槽的表征向量维度为 d_t 。

针对每一条签到记录,将 3 类特征的特征向量进行拼接,可以获得长度为 d 的表征向量 r_k , 其中, $d = d_c + d_l + d_t$ 作为后续时空神经微分方程的输入。

2.2.2 时空神经微分方程

在进行连续时间序列建模方面,本文借鉴文献[25]中的 GRU-ODE 方法,如图 2 所示。针对输入的签到记录表征向量 r_k , 通过基于 GRU 的 ODE 层进行签到行为之间的连续时间状态传播(传播层, Prop-ODE), 即将隐状态 h_{k-1} 从时刻 t_{k-1} 转移到时刻 t_k , 可以形式化为

$$h_{t_k}^- = \text{Prop-ODE}(h_{t_{k-1}}, t_{k-1}, t_k) \quad (1)$$

通过一层 GRU 层在观察点进行签到行为的更新(更新层, Update-GRU)。针对用户断断续续的签到行为,需要在观察到签到行为的时间点对隐状态进行更新。在本更新层中,针对给定的签到行为的表征向量 r_k , 在时间点 t_k , 可以通过式(2)进行更新。

$$h_{t_k}^+ = \text{Update-GRU}(h_{t_k}^-, r_k) \quad (2)$$

其中, $h_{t_k}^-$ 和 $h_{t_k}^+$ 表示在时刻 t_k 经历更新层前后的隐状态。在本文中, GRU 沿用文献[26]的工作, 具体

的公式推导过程在这里不再赘述。

由上述时空神经微分方程可以看出,构造目标函数时主要分为两部分。在 Prop-ODE 层,需要考虑经过传播后的隐状态与观察值之间的分布差异,这里损失函数可以表示为

$$p_{\text{Pre}} = \text{softmax}(f(\mathbf{h}_{t_k^-})) \quad (3)$$

$$L_{\text{Prop}} = - \sum_{i=1}^D \mathbf{v}_k[i] \log p_{\text{Pre}}[i] \quad (4)$$

其中, f 表示一个全连接层将隐状态向量 $\mathbf{h}_k$ 转换成一个 D 维向量, p_{pre} 表示根据隐状态向量推断的该时刻在所有 POI 上的概率分布, $\mathbf{v}_k \in \mathbb{R}^{D \times 1}$ 表示当前观察点的 POI $v_{i_k}^u$ 的 one-hot 表征向量。

在 Update-GRU 层,假设当前观测时刻的 POI v_k^u 的先验分布为 p_{obs} , 模拟贝叶斯更新过程, 获取在该观察点的用户隐状态 p_{Bayes} , 可以通过下式进行计算:

$$p_{\text{Bayes}} \propto p_{\text{Pre}} \cdot p_{\text{obs}} \quad (5)$$

用 p_{post} 表示在应用 Update-GRU 之后预测的概率分布, 则损失函数可以定义为

$$L_{\text{Update}} = D_{\text{kl}}(p_{\text{Baves}} \parallel p_{\text{post}}) \quad (6)$$

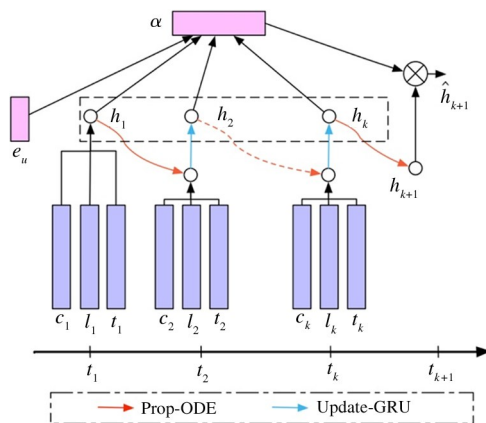


图2 基于时空注意力机制的神经常微分方程框架图

2.2.3 时空注意力机制

从直觉上来说,并非所有的历史签到都与用户的下一步行为密切相关,即需要更加注意用户信息偏好。但是,标准的神经常微分方程 GRU-ODE 网络无法检测其输入的哪一部分对下一个 POI 推荐至关重要。因此本文引入注意力机制,并提出了一种基于时空注意力的神经常微分网络来解决上述问题。时空注意力机制旨在捕获下一步用户行为与过

去签到行为之间的不同关联度。通过使用注意力机制,ST-ODE 还可以帮助选择用户喜欢的代表性签到行为,并为他们分配不同的权重。因此,可以更好地集成用户的历史签到行为表征,以有效地描述用户的兴趣偏好。

如图 2 所示,引入矩阵 $\mathbf{H} \in \mathbb{R}^{d \times k}$ 包含 ST-ODE 在所有评估点的隐状态向量 $\{\mathbf{h}_{t_1^+}, \mathbf{h}_{t_2^+}, \dots, \mathbf{h}_{t_k^+}\}$, 这里 d 表示隐状态向量的维度, k 表示用户签到到序列的长度。ST-ODE 借助于注意力权重向量 $\boldsymbol{\alpha}$ 来进行历史轨迹隐状态向量 $\{\mathbf{h}_{t^+}\}$ 的整合,可以形式化为

$$\hat{\mathbf{h}}_{k+1} = \sum_{i=1}^k \alpha_i \mathbf{h}_{t_i^+} \quad (7)$$

其中, α_i 表示第 i 条记录与下一步的签到行为的匹配程度, 可以通过下式进行权重矩阵的计算:

$$\alpha_i = \frac{\exp(g(\mathbf{h}_{t_i^+}, \mathbf{e}_u))}{\sum_{i=1}^k \exp(g(\mathbf{h}_{t_i^+}, \mathbf{e}_u))} \quad (8)$$

其中, $g(\mathbf{h}_{i_t^+}, \mathbf{e}_u)$ 为注意力函数, 本文使用点积注意力来进行注意力的计算, 可以形式化为

$$g(\mathbf{h}_{i_i^+}, \mathbf{e}_u) = \frac{\mathbf{h}_{i_i^+}(\mathbf{e}_u)^T}{\sqrt{d}} \quad (9)$$

式中, $\mathbf{e}_u$ 表示用户 u 的上下文向量。本文根据用户访问过的 POI 的类别 (category) 信息进行 $\mathbf{e}_u$ 的初始化, $\mathbf{e}_u = \sum_{i=1}^k w_i \mathbf{e}_{i,c}$, w_i 表示类别权重, $\mathbf{e}_{i,c}$ 表示在第 i 条签到行为所在 POI 的类别表征向量。在训练过程中, 将其和参数进行同步更新。这里损失函数的构建与在 2.3 节中的在目标城市进行调优的损失函数一样, 可以形式化为

$$L_{\text{att}} = - \sum_{v^+} \sum_{v^-} \ln(\mathbf{p}_{v^-} - \mathbf{p}_{v^+}) \quad (10)$$

具体推导过程参照 2.3 节, 对一个含有 K 条记录的用户 u 来说, ST-ODE 的整体损失可以定义为

$$L_{\text{st}} = L_{\text{Prop}} + \gamma L_{\text{Update}} + \tau L_{\text{att}} \quad (11)$$

其中, γ 和 τ 表示加权目标函数的权重参数。

将 ST-ODE 模型整体算法总结如算法 1 所示。

算法 1 ST-ODE: 时空常微分方程

输入:

签到序列: $Q_u = q_{t_1}^u, q_{t_2}^u, q_{t_3}^u, \dots, q_{t_k}^u$

签到行为三元组: $q_{t_k}^u = (u, v_{t_k}^u, t_k)$

输出:

用户隐状态表征: $\mathbf{h}$

损失值: $loss$

INITIALIZE $time = 0, loss = 0$, final time T .

FOR $i = 1$ to k DO:

 #状态传播: 从 $time$ 到 t_i

$\mathbf{h}_{t_i^-} = \text{Prop-ODE}(\mathbf{h}_{i-1}, t_{i-1}, t_i)$

$time = t_i$

 #状态更新: 在发生签到行为的时刻

$\mathbf{h}_{t_i^+} = \text{Update-GRU}(\mathbf{h}_{t_i^-}, r_i)$

 根据式(8)进行历史隐状态向量整合

 根据式(11)进行 $loss$ 评估

END FOR

 #状态传播: 从 t_k 到 T

$\mathbf{h} = \text{Prop-ODE}(\mathbf{h}_k, t_k, T)$

 根据式(8)进行历史隐状态向量整合

RETURN $(\mathbf{h}, loss)$

2.3 移动行为知识迁移机制

本小节将介绍提出的基于元学习的移动知识迁移方法,借助于元学习将数据丰富城市的移动行为知识迁移到数据匮乏城市。

城市的移动行为模式随着时间的推移而发生变化,并且城市与城市之间也各不相同。但是,不同城市之间的用户也存在着诸多的共性,比如大家在上午都习惯于打卡咖啡店,在打卡完运动场之后可能更喜欢打卡餐厅。所以,可以将来自于多个城市的共性行为知识,比如空间邻近性和时间依赖性等,迁移到目标城市。

图3为基于元学习的时空常微分方程的参数更新过程。将 $\mathcal{O}$ 作为时空神经常微分方程的参数集合,因此 $\mathcal{O}$ 的相关参数中便隐含了时空移动知识。所以只需要传递更新参数 $\mathcal{O}$ 便可以实现共性行为知识的迁移。这里沿用元学习的思想来进行 ST-ODE 的参数学习,针对式(11),可以通过式(12)进行初始参数的更新。

$$\mathcal{O}_0 = \min_{\mathcal{O}_0} \sum_{S_i \in S} L_{st}(f_{\mathcal{O}_0 - \lambda \nabla_{\mathcal{O}} L_{st}^{S_i}(f_{\mathcal{O}_0})}) \quad (12)$$

其中, $\mathcal{O}_0$ 为 ST-ODE 的参数初始值, $L_{S_i}(f_{\mathcal{O}_0})$ 表示 ST-ODE 在源城市数据集 S_i 上的训练损失值。如在框架图中的每个源城市中,本文将会迭代更新参数。比如在数据集 S_i 中,按照式(13)进行参数更新。

$$\mathcal{O}_{S_i} = \mathcal{O}_0 - \lambda \nabla_{\mathcal{O}} L_{st}^{S_i}(f_{\mathcal{O}_0}) \quad (13)$$

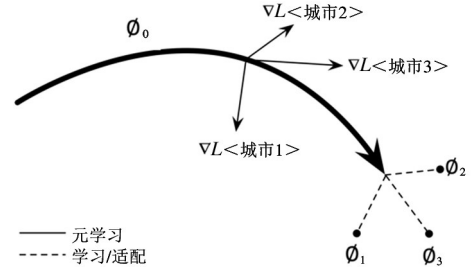


图3 基于元学习的参数优化框架图

由于模型的目标是为数据匮乏城市的用户进行 POI 推荐,所以需要针对目标城市进行参数优化。首先直接将在基于元学习的初始参数学习机制中学习到的初始参数 $\mathcal{O}_0$ 迁移到目标城市 D_i ; 其次,在参数优化过程中,根据 ST-ODE 的输出时刻 t_{k+1} 的状态向量 $\mathbf{h}_{k+1}$, 可以通过式(14)计算用户下一时刻 t_{k+1} 会访问某候选 POI v_i 的概率。

$$\mathbf{p}_{v_i} = (\mathbf{h}_{k+1})^T f(c_i, l_i, t_{k+1}) \quad (14)$$

其中, $f(\cdot)$ 表示表征函数,获取 POI 的相关特征的表征向量, $\mathbf{p}_{v_i}$ 表示用户在 POI v_i 上的概率值。

在推荐过程中,通过 $\arg\max_v(\mathbf{p})$ 为用户推荐最有可能访问的 POI v 。在训练过程中,参照 BPR-Loss 的个性化排名损失训练方式^[27]进行推荐任务损失函数的构建:

$$L_{rec} = - \sum_{v^+} \sum_{v^-} \ln \sigma(\mathbf{p}_{v^+} - \mathbf{p}_{v^-}) \quad (15)$$

其中, v^+ 表示用户访问过的 POI; v^- 表示到目前为止用户没有访问过的 POI; σ 表示非线性函数,即 $\sigma(x) = 1/(1 + e^{-x})$ 。因此在目标城市的参数训练过程中,损失函数可以重新定义为

$$L = L_{rec} + \beta L_{st} + \mu \|\mathcal{O}\| \quad (16)$$

同时,在目标城市中,参数 $\mathcal{O}_{D_i}$ 的更新过程可以改写为

$$\mathcal{O}_{D_i} = \mathcal{O}_0 - \lambda \nabla_{\mathcal{O}} L^{D_i}(f_{\mathcal{O}_0}) \quad (17)$$

在实际的更新参数过程中根据情况进行多次迭代,最终获得具有普适性的初始参数。知识迁移部分算法如算法2所示。

算法2 ML-ODE: 时空行为知识迁移

输入:

源城市集合: $S_1, S_2, \dots, S_N$

目标城市集合: $D_1, D_2, \dots, D_N$

输出:

适配于相关目标城市的元参数: $\mathcal{O}_0$

在目标域完成 POI 推荐任务

INITIALIZE $\mathcal{O}_0$

WHILE model is not convergent DO:

 FOR city S_i in $\{S_1, S_2, \dots, S_N\}$ DO:

 根据式(13)更新参数 $\mathcal{O}_{S_i}$

 根据式(11)评估 $L_{st}^{S_i}(f_{\mathcal{O}_{S_i}})$

 END FOR

 根据式(11)更新参数 $\mathcal{O}_0$

END WHILE

FOR city D_i in $\{D_1, D_2, \dots, D_N\}$ DO:

 根据式(17)更新参数 $\mathcal{O}_{D_i}$

 根据训练的模型完成推荐任务

END FOR

2.4 复杂度分析

基于时空神经微分方程的推荐方法是否能够进行实际应用的关键在于算法的时间复杂度,在这里对 ML-ODE 的时间复杂度进行详细分析。在对多个时间序列进行批处理过程中,对所有时间序列的观测时间进行排序,对于每个唯一时间点 t_k , 创建一个具有观测值的时间序列的列表用于后续模块。ML-ODE 的时空开销主要存在于表征向量获取以及推荐排名两部分。在表征向量获取方面,ML-ODE 的主循环在时序列表唯一的时间点上迭代,在 Prop-ODE 步骤中,针对单个列表,传播所有隐藏状态,时间复杂度为 $O(|\mathcal{O}_{\text{Prop-ODE}}|)$, 这里 $|\mathcal{O}_{\text{Prop-ODE}}|$ 表示 Prop-ODE 模块的参数数量,Update-GRU 更新和损失计算也是仅对在特定具有观测值的时间序列执行,时间复杂度为 $O(|\mathcal{O}_{\text{Update-GRU}}|)$, $|\mathcal{O}_{\text{Update-GRU}}|$ 表示 Update-GRU 模块的参数数量;在推荐排名部分,时间复杂度为 $O(d^2)$, 这里 d 表示表征向量的维度为常量,因此 ML-ODE 的整体时间复杂度为 $O(|\mathcal{O}|)$, $|\mathcal{O}|$ 表示状态向量获取模块的整体参数数量,可以应用于较大规模 POI 推荐任务中。

3 实验与结果分析

本节将详细介绍实验的数据集及相关实验过程,对比不同推荐模型和方法在公开数据集上的

POI 推荐效果,并详细分析各类模型的性能表现。

3.1 数据来源

本文使用的数据集来自于 Foursquare^[28]。Foursquare 是在位置社交网络领域的一个大规模的社交网站,允许用户在不同的地理位置进行签到行为(check-ins),本文使用 2012 年 9 月到 2013 年 9 月之间 Foursquare 公开的去隐私化的用户全球规模签到数据集。据统计,这个数据集包含 33 278 683 条用户签到数据,涵盖用户量为 266 909,POI 数量为 3 680 126(涉及全球 77 个国家中的 415 个城市)。在这 415 个城市中,每个城市至少包含 10 000 以上的签到数据。本文列举了几个在实验过程中整理的城市签到数据样例,如表 1 所示。图 4 和图 5 展示了数据集统计表中东京和旧金山(SF)的城市 POI 分布热力图,可以看出城市间之间 POI 分布极不平衡。

在实验评估过程中,按照用户的平均签到条目数来区分数据丰富城市和数据匮乏城市,在这里分别指定洛杉矶(LA)和旧金山为数据匮乏城市,其余城市划分为数据丰富城市。为了验证本文方法在数

表 1 城市签到数据统计

城市	用户	POI	签到行为	均值
纽约	17 385	66 660	581 544	33.45
洛杉矶	8969	21 556	162 658	18.14
旧金山	5455	11 076	87 906	16.11
芝加哥	6870	21 880	184 453	26.87
华盛顿	6420	13 828	131 444	20.47
东京	12 613	88 546	1 100 216	87.23

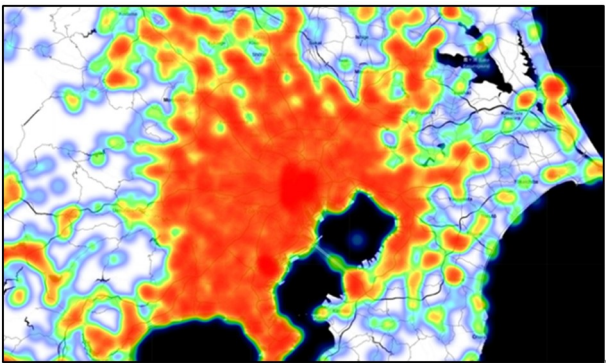


图 4 东京 POI 分布热力图

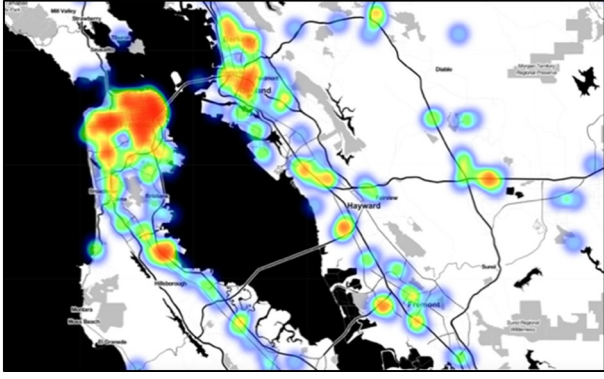


图5 旧金山 POI 分布热力图

据匮乏区域的效果,在数据匮乏城市只使用 1 个月的 POI 数据进行实验。除此之外,在实际评估过程中,按照 7 : 1 : 2 的比例划分训练集、验证集和测试集。

3.2 对比方法

本文实验中选取了目前比较流行的效果较好的 POI 推荐模型,主要通过命中率(HR)和归一化折损累计收益(NDCG)等指标来进行各类算法的实验对比,下面概述在本文中对比的主流算法。

MF-BPR^[27]: 该算法提出了一种基于贝叶斯的改进矩阵分解模型,主要是通过贝叶斯的个性化排序来优化 MF^[29],通过建模用户访问过的 POI 数据来建模用户与 POI 的隐式关系。

CML^[30]: 该算法将度量学习(metric learning)与协同过滤(collaborative filtering)进行结合来解决用户物品推荐的问题,提出了协同度量学习模型,学习用户和物品之间潜在关系的同时评估用户和用户、物品和物品之间的相似性,借助用户隐式评价进行 Top-K 推荐。

PRME^[15]: 该算法使用针对推荐的排序度量嵌入方法对用户的个性化签到序列进行建模,此模型整合了签到的序列信息、个人偏好和地理影响,以提高推荐性能。

STRNN^[11]: 该算法在循环神经网络的基础上增加了时间敏感性,捕捉在签到行为之间的时间和空间位置信息变化。

DeepMove^[10]: 该算法在循环神经网络的基础上增加了用户兴趣注意力机制,考虑用户签到序列时序关系的基础上还考虑了用户的长期兴趣。

ST-LSTM^[11]: 该算法是 LSTM 的变式,在 LSTM 中增加了时间和距离两种门限机制来捕捉用户签到行为序列之间的时空关系进行 POI 推荐。

GRU^[26]: 该算法也是循环神经网络的一种,可以看成是 LSTM 变体,结构更加简单,主要解决了 RNN 网络中的长依赖问题。

ST-ODE: 该算法是 ML-ODE 的弱化版本,在不使用元学习机制下,应用时空神经网络常微分方程进行推荐,旨在比较分析元学习机制在该任务中是否有效。

Single-FT: 该算法是 ML-ODE 的变式,与 ML-ODE 参数优化机制不同的是,Single-FT 利用单个源城市对目标城市时空模型的初始化参数进行更新。

Multi-FT: 该算法也是 ML-ODE 的变式,与 ML-ODE 参数优化机制不同的是,Multi-FT 将多个源城市的数据进行混合之后无差别地对目标城市的时空模型进行初始化参数更新。

No-ATT: 该算法也是 ML-ODE 的变式,与 ML-ODE 不同的是,No-ATT 不使用时空注意力机制进行隐状态的聚合,直接使用 ST-ODE 的最后输出隐状态作为推荐模块的输入。

ML-ODE: 该算法是本文提出的基于元学习的神经网络常微分方程,将神经网络常微分方程用于时空关系挖掘。除此之外,借助元学习进行初始参数的优化和学习,实现在数据匮乏城市的 POI 推荐任务。

可以看出,在比较的方法中,MF-BPR 和 CML 是推荐系统领域比较通用(general)的模型,应用比较广泛。而 STRNN、DeepMove 和 ST-LSTM,都是要根据签到序列的上下文信息(context-aware)进行时空关系挖掘的推荐模型。在后期的实验评估过程中将会详细地分析各类模型在 POI 推荐任务上的表现。

3.3 评价指标

在性能对比实验中,采用两种推荐指标来评估推荐模型的性能:命中率(HR@N)和归一化累计增益(NDCG@N)。

$$HR@N = \frac{NumberOfHits@N}{GT} \quad (18)$$

其中,GT 表示所有的测试集合,NumberOfHits@N

表示推荐列表的 Top-N 子集中属于测试集合个数的总和。

$$NDCG@N = \frac{DCG@N}{IDCG} \quad (19)$$

这里 DCG 表示折损累计增益,是在累计增益 CG 的基础上引入了位置影响因素,计算公式如下:

$$DCG@N = \sum_{i=1}^N \frac{2^{rel_i} - 1}{\log_2(i + 1)} \quad (20)$$

可以看出,推荐的相关性越多,DCG 值越大,相关性强的排在推荐列表前面的话,推荐效果越好,

DCG 值越大。在式(19)中, IDCG 表示推荐系统针对某一用户返回的最好推荐列表,即假设返回结果按照相关性排序,最相关的结果排在最前面的情况下的 DCG 值。由定义可以看出 NDCG 的值为(0, 1]。

3.4 推荐任务

表 2 是各种对比方法在 HR@5、HR@10、NDCG@5 以及 NDCG@10 指标上的实验结果。可以看出本文方法与其他对比方法相比在推荐任务方面表现更佳,下面对实验结果进行详细分析。

表 2 各类方法在 POI 推荐任务上的结果统计

类型	方法	城市	HR@5	NDCG@5	HR@10	NDCG@10
一般方法	MF-BPR	SF	0.03456	0.02304	0.06283	0.04409
		LA	0.03312	0.02801	0.06103	0.04334
	CML	SF	0.05872	0.04912	0.14871	0.10594
		LA	0.05541	0.04647	0.14276	0.10012
上下文感知	PRME	SF	0.04076	0.03247	0.10189	0.06719
		LA	0.03972	0.03032	0.09535	0.06197
	ST-RNN	SF	0.05006	0.03903	0.13649	0.08359
		LA	0.04718	0.03827	0.12902	0.08143
	DeepMove	SF	0.06012	0.04997	0.15096	0.10889
		LA	0.05982	0.04921	0.14872	0.10727
	ST-LSTM	SF	0.07043	0.05819	0.16762	0.12440
		LA	0.06801	0.05654	0.16426	0.12026
	GRU	SF	0.05118	0.04092	0.13765	0.08503
		LA	0.04831	0.03972	0.13033	0.08209
方法变式	ST-ODE	SF	0.07601	0.06827	0.18102	0.14821
		LA	0.07413	0.06423	0.17704	0.14454
	Single-FT	SF	0.08122	0.07031	0.20872	0.16252
		LA	0.08086	0.06769	0.20224	0.16157
	Multi-FT	SF	0.09092	0.07368	0.23125	0.17098
		LA	0.08923	0.07121	0.22767	0.17003
	No-ATT	SF	0.09722	0.08301	0.25125	0.19799
		LA	0.09586	0.07969	0.24872	0.19503
ML-ODE	SF	0.09722(+14.74%)	0.08301(+22.37%)	0.25125(+23.36%)	0.19799(+29.43%)	
	LA	0.09586(+10.98%)	0.07969(+17.28%)	0.24872(+15.13%)	0.19503(+25.96%)	

(1) 上下文信息的重要性。由实验结果可以看出,利用时空上下文信息的方法(如 PRMF、STRNN、DeepMove、ST-LSTM、ST-ODE 等)在推荐任务上要优于没有利用上下文信息的方法(MF-BPR 和 CML)。可以看出上下文信息在推荐任务中起着至关重要的作用,有助于模型捕捉用户签到行为时空上的序列关系,提升推荐效果。文献[31]也通过大量的实验

论证了上下文信息对 POI 推荐的重要性。

(2) 神经网络的优势。由实验结果可以看出,基于矩阵分解的方法(如 MF-BPR 和 PRME)在推荐效果上欠佳,而基于神经网络的方法(如 STRNN、GRU、DeepMove、ST-LSTM、ST-ODE 等)在推荐任务上要更好。可以看出单纯的矩阵分解方法在捕捉用户的高阶关系上效果较差,而神经网络由于可以建

模用户更高阶的潜在关系,推荐效果要优于传统的矩阵分解方法。

(3)连续时间建模的优势。由实验结果可以看出,将签到行为序列看成等间隔进行建模的方法(如 STRNN 和 DeepMove)在推荐效果上要略逊于时间敏感的推荐方法(如 ST-LSTM 和 ST-ODE),签到行为的时间依赖也是用户兴趣点推荐不可忽视的因素。并且本文提出的基于 Neural ODEs 的方法可以建模连续时间序列,从实验结果可以看出,ST-ODE 进一步提升了时间敏感类方法在 POI 推荐上的性能。

(4)时空注意力机制的优势。由实验结果可以看出,借助于时空注意力机制的 ML-ODE 在推荐效果上要优于不使用注意力机制的 No-ATT 方法,主要原因在于时空注意力机制对用户偏好的权重进行了更好地建模,并克服了在长时间序列传播过程中的知识损失,提升了 POI 推荐效果。

(5)知识迁移的必要性。根据表 2 针对知识迁移的必要性进行实验论证,通过 ML-ODE 的几个进阶版本进行对比实验。ST-ODE 仅利用目标城市的数据进行训练,即没有基于元学习机制进行参数的初始化。Single-FT 是利用单个源城市对目标城市的参数进行更新。Multi-FT 是将多个源城市的数据进行混合之后无差别地对目标城市的时空模型进行初始化更新。由实验结果可以看出,将多个城市视为不同的推荐任务来借助元学习进行初始参数学习的 ML-ODE 方法具有更加优异的表现,NDCG@10 在两个城市的性能提升均超过了 25%。Multi-FT 的方法较 Single-FT 表现更好,可能是由于多个城市数据混合之后在一定程度上缓解了由于数据分布差异性对学习移动知识的通用性的影响。除此之外,ML-ODE 比 Multi-FT 表现更佳,说明了元学习框架在知识迁移方面的优势,通过多任务的分解可以更好地学习模型的初始化参数。

3.5 模型参数设置

本小节将详细描述 ML-ODE 各种参数的设置情况并分析其对模型性能的影响。在 ML-ODE 对比实验中,对所有方法都是采用 5 折交叉验证,并且中间状态传播过程中的隐状态维度均设置为 50。

在参数训练过程中,dropout 的取值范围为 $[0, 0.1, 0.2, 0.3]$,weight decay 的取值范围为 $[0.1, 0.03, 0.01, 0.003, 0.001, 0.0001, 0]$,learning rate 的取值范围为 $[0.001, 0.003, 0.0001]$ 。通过 5 折交叉验证,每次留 20% 的验证数据集对上述参数训练的 ML-ODE 参数进行评估,模型表现最优时的 dropout 值为 0.2,weight decay 的值为 0.001,learning rate 的值为 0.003。

除此之外,在进行 Top-N 推荐任务的过程中,通过将参数 N 在取值范围 $[1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50]$ 内进行变化,观察不同模型的表现情况,实验结果如图 6 ~ 图 9 所示。可以看出模型在 N 的取值范围内表现均优于其他对比实验方法。同时,相较于 $HR@N$ 指标,ML-ODE 在 $NDCG@N$ 上的提升效果更加明显。将图 8 和图 9 进行对比发现,在旧金山数据集上,ML-ODE 的 $NDCG@N$ 的提

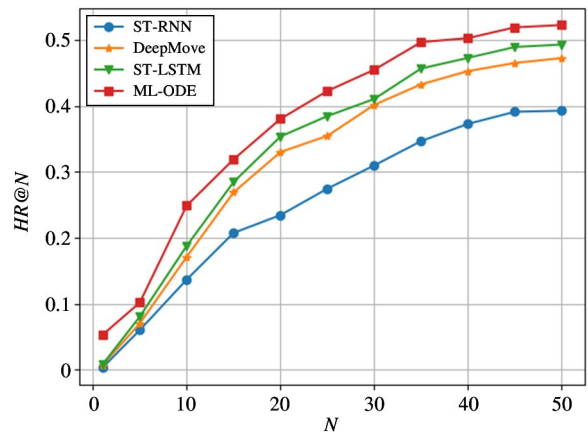


图 6 在旧金山数据集上的 $HR@N$ 结果

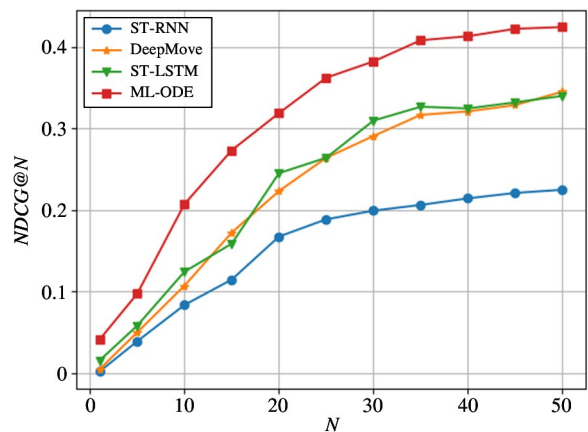


图 7 在旧金山数据集上的 $NDCG@N$ 结果

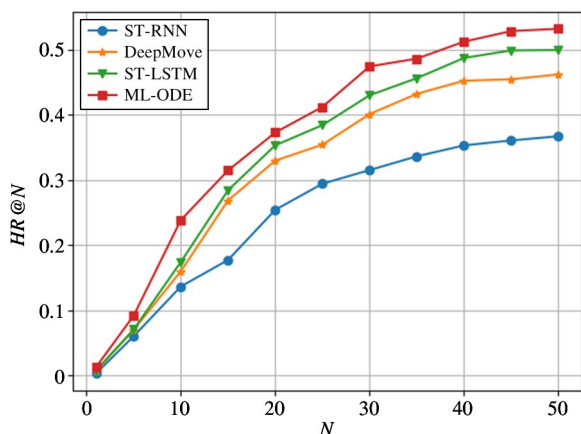


图 8 在洛杉矶数据集上的 HR@N 结果

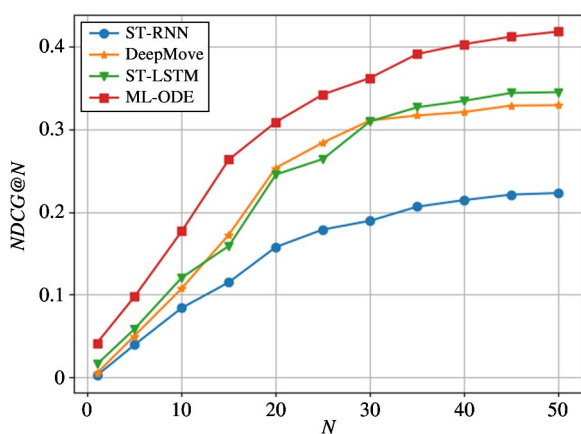


图 9 在洛杉矶数据集上的 NDCG@N 结果

升效果更加明显。在洛杉矶数据集上也发现了类似现象,说明在排序结果中,高关联度的结果出现在更靠前的位置,具有更高的累计增益值。

4 结 论

针对兴趣点推荐,本文提出了一种基于元学习的时空 Neural ODEs 模型(ML-ODE)。首先,利用基于时空神经微分方程的时空关系学习模块进行用户的时空兴趣建模。其次,在模型中借助基于贝叶斯的更新模块根据用户的签到记录对时空神经微分方程预测的用户隐状态进行更新,更好地学习用户签到行为的时空序列依赖。除此之外,借助于元学习机制进行时空神经微分方程的参数初始化学学习,在进行参数初始化的过程中,有效地将数据丰富城市的移动行为知识进行迁移,提升模型在数据匮乏

城市的 POI 推荐效果。最后,在真实的公开 LB-SNs 数据集上进行实验,实验结果有效地验证该方法的推荐效果,并证明了 ML-ODE 在 HR@N 和 NDCG@N 指标上相比当前先进的兴趣点推荐技术有了明显提高。

在未来的工作中,将进一步考虑高效的面向稀疏 POI 签到数据的 POI 推荐方法,并将从以下几个方面进行尝试:(1)结合分布式机器学习技术,将 ML-ODE 进行分布式算法扩展,结合 GPU、TPU 等高效计算方式,让分析超大规模 POI 数据成为可能;(2)解决 ML-ODE 动态更新问题,在实际应用场景中,用户的签到行为是动态变化的,而目前大多数的推荐方法都是建立在离线静态 POI 签到数据上,所以对网络中动态 POI 签到行为进行合理化建模,实现合理的在线参数更新机制,值得更多的关注;(3)计划引入用户的长期兴趣偏好,或者结合主题模型等用户长期兴趣建模方法,进一步提升兴趣点推荐系统的性能。

参考文献

- [1] Liu Q, Wu S, Wang L, et al. Predicting the next location: a recurrent model with spatial and temporal contexts [C]//International Conference on Artificial Intelligence, Phoenix, USA, 2016: 194-200
- [2] 任星怡,宋美娜,宋俊德. 基于用户签到行为的兴趣点推荐[J]. 计算机学报, 2017, 40(1): 28-50
- [3] Zhang J D, Chow C Y. iGSLR: personalized geo-social location recommendation: a kernel density estimation approach [C]//Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, USA, 2013: 334-343
- [4] 刘树栋, 孟祥武. 基于位置的社会化网络推荐系统 [J]. 计算机学报, 2015, 38(2): 322-336
- [5] Endo Y, Nishida K, Toda H, et al. Predicting destinations from partial trajectories using recurrent neural network [C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining, Jeju, Korea, 2017: 160-172
- [6] Du N, Wang Y, He N, et al. Time-sensitive recommendation from recurrent user activities [C]//Advances in Neural Information Processing Systems, Cambridge, USA, 2015: 3492-3500

- [7] 廖国琼, 姜珊, 周志恒, 等. 基于位置社会网络的双重细粒度兴趣点推荐[J]. 计算机研究与发展, 2017, 54(11): 2600-2610
- [8] Yao H X, Liu Y D, Wei Y, et al. Learning from multiple cities: a meta-learning approach for spatial-temporal prediction[C] // The World Wide Web Conference, San Francisco, USA, 2019: 2181-2191
- [9] Wang L Y, Geng X, Ma X J, et al. Crowd flow prediction by deep spatio-temporal transfer learning[EB/OL]. <http://arXiv.org/abs/1802.00386>, (2018-06-09), [2020-10-08]
- [10] Zhou Y, Huang Y. DeepMove: learning place representations through large scale movement data[C] // International Conference on Big Data, Seattle, USA, 2018: 2403-2412
- [11] Zhou P P, Zhao H F, Liu Y C, et al. Where to go next: a spatio-temporal gated network for next POI recommendation[C] // Proceedings of the International Conference on Artificial Intelligence, Honolulu, USA, 2019: 5877-5884
- [12] Wei Y, Zheng Y, Yang Q. Transfer knowledge between cities[C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016: 1905-1914
- [13] Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7167-7176
- [14] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 22(10): 1345-1359
- [15] Yin H Z, Zhou X F, Shao Y X, et al. Joint modeling of user check-in behaviors for point-of-interest recommendation[C] // Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 2015: 1631-1640
- [16] Lin X, Zhang M, Zhang Y F. Joint factorizational topic models for cross-city recommendation[C] // Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data, Beijing, China, 2017: 591-609
- [17] Chelsea F, Pieter A, Sergey L. Model-agnostic meta-learning for fast adaptation of deep networks[C] // Proceedings of the 34th International Conference on Machine Learning-Volume, Sydney, Australia, 2017: 1126-1135
- [18] Chen R T Q, Rubanova Y, Bettencourt J, et al. Neural ordinary differential equations[C] // Advances in Neural Information Processing Systems, Montreal, Canada, 2018: 6571-6583
- [19] Feng S S, Li X T, Zeng Y F, et al. Personalized ranking metric embedding for next new POI recommendation[C] // Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015: 2069-2075
- [20] Huang L W, Ma Y T, Wang S B, et al. An attention-based spatiotemporal LSTM network for next POI recommendation[J]. *IEEE Transactions on Services Computing*, 2019, 14(6): 1585-1597
- [21] Zhang Y, Dai H, Xu C, et al. Sequential click prediction for sponsored search with recurrent neural networks [EB/OL]. <http://arXiv.org/abs/1404.5772>, (2014-04-23), [2020-10-08]
- [22] Yin H, Zhou X, Cui B, et al. Adapting to user interest drift for POI recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(10): 2566-2581
- [23] Li D C, Gong Z G, Zhang D F. A common topic transfer learning model for crossing city POI recommendations [J]. *IEEE Transactions on Cybernetics*, 2018, 49(12): 4282-4295
- [24] Rubanova Y, Chen R T Q, Duvenaud D K. Latent ordinary differential equations for irregularly-sampled time series[C] // Advances in Neural Information Processing Systems, Vancouver, Canada, 2019: 5320-5330
- [25] De Brouwer E, Simm J, Arany A, et al. GRU-ODE-Bayes: continuous modeling of sporadically-observed time series[C] // Advances in Neural Information Processing Systems, Vancouver, Canada, 2019: 7379-7390
- [26] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. <http://arXiv.org/abs/1406.1078>, (2014-06-03), [2020-10-08]
- [27] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[C] // Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, Montreal, Canada, 2009: 452-461

- [28] Yang D Q, Qu B Q, Yang J, et al. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach [C] // The World Wide Web Conference, San Francisco, USA, 2019: 2147-2157
- [29] Mnih, Salakhutdinov R. Probabilistic matrix factorization [C] // Advances in Neural Information Processing Systems, Vancouver, Canada, 2008: 1257-1264
- [30] Hsieh C K, Yang L Q, Cui Y, et al. Collaborative metric learning [C] // International World Wide Web Conference, Perth, Australia, 2017: 193-201
- [31] 彭宏伟, 靳远远, 吕晓强, 等. 一种基于矩阵分解的上下文感知 POI 推荐算法 [J]. 计算机学报, 2019 (8):1797-1811

A next POI recommendation method for data-poor cities

Tan Haining^{***}, Yao Di^{**}, Bi Jingping^{**}, Xiang Xu^{***}, Yang Xiao^{***}

(^{*} University of Chinese Academy of Sciences, Beijing 100049)

(^{**} Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

(^{***} National Key Laboratory of Science and Technology on Blind Signal Processing, Chengdu 610041)

Abstract

In location-based social networks (LBSNs), the uneven distribution of users' check-in data and the fact that some users deliberately hide some of their location information for privacy and security concerns aggravate the difficulty of point of interest (POI) recommendation. Aiming at this issue, a meta-learning based ordinary differential neural network (ML-ODE) is proposed to carry out effective POI recommendation tasks. ML-ODE leverages the meta-learning mechanism to optimize the parameters of recommendation model. ML-ODE utilizes different tasks to initial parameters, during which the generalized mobility knowledge in the data-rich area is transferred to the data-poor area to achieve the purpose of optimizing the POI recommendation. The network defines a continuous dynamic process that can accept input data that is sporadically-observed. It overcomes the static discretization constraints of most time series prediction models and is more suitable for POI recommendation. The experimental results on the Foursquare real public dataset show that ML-ODE has better performance in POI recommendation than the current state-of-the-art POI recommendation method on the index of NDCG@N which has been increased by more than 10%.

Key words: point of interest (POI) recommendation, location-based social network (LBSN), meta-learning, neural ordinary differential equation, recommender system