

# 城市环境下的用户移动行为建模概述

徐丰力, 李勇

(清华大学电子工程系, 北京 100084)

**摘要:** 针对城市环境成为移动通信、交通调度、疾病防控等领域的典型场景, 而用户的移动行为建模对这些关键领域有重要的应用与研究价值, 梳理、总结城市环境下移动行为建模的研究进展与现状, 为该领域的相关研究提供文献概述。首先讨论了城市环境移动行为建模问题面临的主要挑战及对应的核心科学问题, 即移动行为数据增强算法、城市结构感知的移动行为模式识别、多时空尺度的移动行为预测模型和移动数据隐私保护机制问题。进一步地, 围绕这些核心科学问题梳理总结了该领域近年来的发展脉络与最新研究成果, 为未来的研究工作奠定了基础。

**关键词:** 移动数据; 移动建模; 数据挖掘; 隐私保护机制

**中图分类号:** TP181

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-436x.2020147

## Survey on user's mobility behavior modelling in urban environment

XU Fengli, LI Yong

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

**Abstract:** Urban environment has become a typical scenario for areas of mobile communication, transportation scheduling, disease controlling and so on, and modelling user's mobility behavior plays an important role in these key applications. The research development in this area was combed and summarized, which provided a literature review for related works. Firstly, the main challenges in urban mobility modelling were discussed as well as the corresponding key scientific problems, which included mobility data augmentation, urban structure-aware mobility behavior discovering, multi-scale mobility behavior prediction and mobility data privacy protection. Furthermore, according to these key scientific problems, the recent developments and up-to-date scientific output in this area were summarized, which paved the way for future research.

**Key words:** mobility data, mobility modelling, data mining, privacy mechanism

### 1 引言

近年来, 随着移动网络的飞速发展和移动用户的迅猛增加<sup>[1]</sup>, 城市环境中的移动互联网逐步成为一个巨大的感知平台, 产生了刻画用户日常时空行为的海量“电子足迹数据”, 所形成的移动行为大数据记录了大规模用户在何时 (when)、何地 (where)、做什么事 (what) 等行为信息, 蕴含了丰富的社会、

经济和学术价值<sup>[2]</sup>。海量的智能终端和可穿戴设备连接到覆盖范围越来越小的移动网络基站或宽带接入点时, 运营商和服务提供商通过网络定位、GPS 等方法记录了移动用户在物理世界的细粒度移动行为, 这些数据得到了学术界和产业界越来越多的关注。例如, 在移动网络领域, 研究者提出了基于移动行为大数据的移动网络优化机制<sup>[3]</sup>; 在移动计算领域, 研究者正在建立移动大数据驱动的移动网络信

收稿日期: 2020-01-15; 修回日期: 2020-06-25

基金项目: 国家重点研发计划基金资助项目 (No.SQ2018YFB180012); 国家自然科学基金资助项目 (No.61971267, No.61972223, No.61861136003, No.61621091); 北京市自然科学基金资助项目 (No.L182038)

**Foundation Items:** The National Key Research and Development Program of China (No.SQ2018YFB180012), The National Natural Science Foundation of China (No.61971267, No.61972223, No.61861136003, No.61621091), Beijing Natural Science Foundation (No.L182038)

息平台<sup>[4]</sup>。然而，由于移动行为数据存在规模大、维度高、噪声强、隐私风险高等特点，对现有分析挖掘模型带来了很大的挑战，其内在价值仍远未被充分开发应用<sup>[5]</sup>。顺应生产力的发展趋势和新时代的要求，我国在近两年大力推动人工智能和大数据的实战应用<sup>[6]</sup>，明确要求产、学、研各界大力推动人工智能和大数据分析技术的快速发展，在总体技术和应用上达到世界先进水平，并指明人工智能和大数据分析技术将成为发展国民经济水平、改善民生的新途径。而在全世界范围内，城市环境成为诸多应用的典型场景。城市环境下的移动行为建模对移动网络优化、疾病防控、公共管理、城市规划等领域存在重要应用价值。基于以上背景，本文聚焦于城市环境下的用户移动行为建模问题，讨论了该问题面对的主要挑战和对应的研究框架，并梳理、总结了该领域的研究现状。

在世界范围内，城市环境移动行为数据的挖掘与应用受到了广泛的关注，前沿研究机构在这一领域开展了一系列的研究与探索。一方面，从群体角度看，移动大数据蕴含用户分布的时空规律和业务需求的时空特征，这些规律将有助于移动网络、交通系统、医疗系统等实现更加高效的资源适配和优化；另一方面，从个体角度看，在长时间的历史个体移动轨迹中，有可能找到移动行为的时空模式，实现对用户未来移动轨迹的预测，为下游应用提供有效支持。因此，海内外的研究者对个体、群体层面的移动模式识别和行为预测展开了大量研究。在关键移动行为模式识别方面，研究者提出在宏观层面用服从幂律分布的随机过程建模人群单次移动距离与移动时间间隔的统计特征<sup>[7]</sup>，在微观层面用频繁子序列挖掘、主成分分析、聚类分析、表征学习等成熟数据挖掘模型识别个体时空行为模式的解决策略<sup>[8]</sup>。在移动行为预测方法方面，个体移动轨迹预测与群体时空分布预测成为关注焦点，学术界提出了从时空行为数据中识别人群的宏观时空行为模式，用概率统计建模并对未来时空行为进行预测的解决思路<sup>[9]</sup>；此外，基于经典的马尔可夫链和隐藏马尔可夫链等序列挖掘模型也被广泛研究用于建模和预测个体的移动轨迹序列<sup>[10]</sup>。然而，相关研究工作针对的问题较分散，研究成果彼此间的关联、协同不足，未能形成统一的研究框架和知识体系，对下游应用的支撑能力存在不足。

## 2 城市环境移动行为建模问题与主要挑战

### 2.1 问题定义

城市环境下的用户移动行为建模问题处于多个前沿学科的交叉领域，其与信息技术、数据挖掘、城市科学、社会科学等经典研究领域存在深层次的耦合<sup>[11]</sup>。概括地说，该研究问题旨在利用城市环境下采集的海量用户移动行为数据，挖掘识别出个体、群体层面用户的移动行为模式，探索其与城市结构的相互作用机制，并构建能在下游领域直接应用的用户移动预测模型。这一研究问题在移动网络覆盖日益普及、大数据分析技术方兴未艾的今天具有重要意义，其研究成果对下游的无线网络优化、城市规划、疾病防控、公共安全等领域有重要应用价值，吸引了全球范围学术界与产业界的广泛关注<sup>[12]</sup>。然而，该领域的研究课题仍面临着诸多挑战，相关研究成果总结如下。

### 2.2 主要挑战

#### 2.2.1 数据质量差

移动行为数据往往从不同源头由不同设备采集汇聚而成。已有研究工作对移动数据质量带来的挑战已有较多讨论<sup>[13-14]</sup>，一方面，由于移动设备的数据采集能力较弱，覆盖个体规模巨大，行为维度多等原因，单一个体在单一维度上的移动数据记录十分稀疏<sup>[13]</sup>；另一方面，由于采集设备处理能力不均、移动网络数据丢失等原因，移动设备采集的数据普遍存在显著的数据噪声。并且，异源数据质量存在较大差异，难以直接统一建模<sup>[14]</sup>。比如，通过蜂窝基站定位和卫星信号采集的移动记录在空间精度、记录数目上存在较大差别。因此，如何在移动行为数据的挖掘中克服记录稀疏和噪声显著等数据质量问题是一个关键挑战。

#### 2.2.2 行为模式与城市环境存在复杂关联

以智能移动终端为媒介采集的移动行为数据具有跨物理域、网络域、社会域，覆盖多维度行为信息和非结构化的特点，且不同维度的行为之间存在复杂的耦合关系。在城市环境下，群体和个体移动行为与城市结构的关联关系是近年来研究的焦点之一<sup>[15]</sup>。随着城镇化的发展，城市空间逐渐被划分为功能各异的若干区域，从而形成了复杂的城市结构。而经典的移动行为模式挖掘算法的刻画能力较弱，只能捕捉较简单或预先设定的行为模式，难以自适应地、准确地识别出用户移动行为与城市结

构间的复杂关联<sup>[16]</sup>。

### 2.2.3 移动行为存在复杂的高阶时空关联

城市尺度下的真实移动行为数据中往往包含复杂的高阶时空相关性。举例而言,在时间维度上,个体的移动性受多种周期性行为的影响,当前所处位置常常与一天前同一时间或一周前同一时间所处位置相同;在空间维度上,彼此距离很远的地点也可能会因为地点性质相似而吸引相同的人群访问<sup>[17]</sup>。另一方面,这种高阶时空相关性是因人而异、因地而异、因时而异的,且往往与社交域、网络域的行为存在关联<sup>[18]</sup>。这种复杂的高阶时空关联对高质量的移动行为预测提出了巨大挑战,经典的马尔可夫链等简单序列模型无法有效建模高阶、多尺度的时空相关性。

### 2.2.4 数据隐私风险

在具有重要的学术与应用价值的同时,城市环境下的细粒度用户移动行为数据也存在严重的隐私风险。潜在攻击者通过简单观测个体移动数据即可对目标个体进行跟踪和监控,甚至进一步推测出其居住地址、工作单位、常去商铺等敏感信息。这些潜在的隐私风险不仅对个体信息安全构成了重要挑战,也妨碍了高质量移动数据在学术界、产业界的有效流通,从而无法充分发挥其内在价值。实现同时保障用户隐私、数据质量和应用价值的移动数据共享成为相关领域研究的重要挑战<sup>[19]</sup>。

## 2.3 核心科学问题

城市环境的移动行为建模是一个新兴的交叉领域。结合 2.2 节所诉的主要挑战,这一领域的核心科学问题可以总结为以下 4 个方面,其存在层层递进的关系,如图 1 所示。

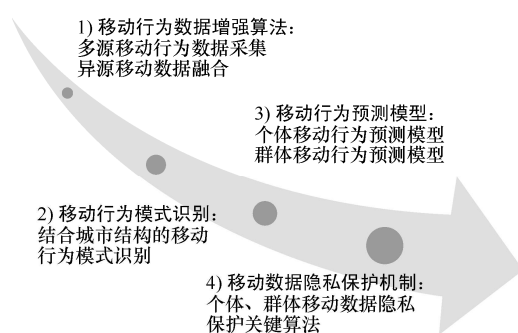


图 1 城市环境移动行为建模的 4 个核心科学问题

### 2.3.1 移动行为数据增强算法

随着移动通信技术的快速发展,城市环境下用户手持的智能终端和海量物联网设备成为采集移

动行为数据最有效的平台。然而,各渠道采集的移动行为数据均存在数据稀疏、数据准确度低等数据质量问题<sup>[14]</sup>。因此,有效融合异源采集的移动行为数据,从而实现移动数据增强是城市环境移动行为建模的一个关键课题。

用户移动行为模式的已有研究表明,个体移动行为存在两点基本共性:一方面,个体移动行为存在显著的周期性<sup>[20]</sup>;另一方面,不同个体间的移动则存在明显的差别<sup>[21]</sup>。因此,异源移动行为数据融合问题等价于基于移动行为模式判断不同渠道采集的移动数据是否属于同一用户的预测问题。具体而言,该科学问题涉及的主要研究内容如下。

1) 针对移动行为数据由于采集机制、感知设备、采集场景不同而导致的数据质量差异大、数据非结构化的特点,研究构建可以普遍适用于多种场景产生的移动行为数据的融合算法,实现基于互补的多源移动数据的融合增强<sup>[22]</sup>。

2) 在多维度移动行为数据(即除时空信息外,还包含与之关联的网络行为、社交行为等辅助维度)中,研究可以建模辅助维度信息的移动行为相似度测量机制,使其在不同来源的移动行为数据上具备足够的泛化能力和拓展能力,并充分利用其他辅助维度的信息<sup>[23]</sup>。

3) 针对移动行为数据由于终端数据处理能力较弱、记录丢失等原因不可避免地存在显著噪声的问题,探索识别并过滤数据噪声的方法,并研究基于移动行为规律补全缺失移动数据的算法,提出可靠方法来进一步提高移动数据在真实应用场景下的稳健性和应用价值<sup>[24]</sup>。

### 2.3.2 城市结构感知的移动行为模式识别

在城市环境下,识别出城市结构感知的移动行为模式是一个重点研究课题。随着城市功能分区的不断细化<sup>[25]</sup>,用户在城市空间中的移动往往有与城市结构相关的具体意图,如上班、购物等。因此,如何结合城市结构建模城市环境下的移动行为模式,并推断移动行为对应的意图成为一个重要的科学问题。其核心难点在于识别基于意图的移动行为模式,而非在物理空间中相似的移动轨迹。具体而言,这一问题对应的主要研究内容如下。

1) 通过对个体移动行为意图的理解和对城市结构信息的重新整合,研究可以有效表征个体移动意图的、不受物理空间远近约束的移动行为模式表示方法及与之匹配的模式识别算法<sup>[26]</sup>。

2) 基于大规模社交媒体中富含语义的个体签到数据（包括具体的签到时间、详细的地理位置和对应的兴趣点信息等）的支持，设计深度表征学习算法以实现无监督的移动语义信息挖掘，并提取同时保留物理空间移动行为模式和移动意图语义信息的低维向量表示<sup>[27]</sup>。

3) 研究意图感知的移动行为低维表征提取方式，通过聚类分析得到典型的移动行为模式，建立移动用户个体时空模型，并进一步分析不同的行为模式的特点及其与城市结构的关联<sup>[28]</sup>。

### 2.3.3 多时空尺度的移动行为预测模型

城市环境下的移动行为预测问题主要可以分为个体移动行为预测和群体移动行为预测两类。在个体移动行为预测方法方面，由于人类活动的规律性和复杂性，个体用户移动行为存在高阶时空相关性和显著的多尺度周期行为。同时，由于移动行为数据的产生和收集方式的限制，真实数据中往往存在大量噪声，而且通常具有稀疏性和复杂性的特点。因此，围绕这一科学问题的研究热点主要包含以下具体内容。

1) 基于递归神经网络、卷积神经网络等深度学习工具，建模个体移动行为在不同时空尺度上的高阶时空相关性，实现基于深度神经网络的移动行为预测模型<sup>[29]</sup>。

2) 通过引入注意力机制等，研究个体时空行为中的多尺度周期性（包括以日、周为周期的时间维度规律，甚至季节规律等），建立针对性的移动行为刻画模型以提高个体时空行为预测的准确性<sup>[30]</sup>。

3) 基于深度表征学习方法，研究综合利用个体移动行为数据中辅助维度信息（包括时间、空间信息和兴趣点以及文字评论等丰富的语义信息）的可行性，实现对于稀疏移动数据的合理使用，建立表现稳健、可以应用于复杂场景的一般模型。

在群体移动行为预测方面，由于高精度数据获取的困难性，当前研究的重心为基于粗粒度的群体时空分布数据预测动态的、细粒度的群体移动行为的映射问题。粗粒度的人口普查结果是当前最常用的群体时空分布的数据，但是人口普查数据存在空间解析度太低以及无法实时预测群体移动行为的问题导致其在现实应用中受到很大局限<sup>[1]</sup>。随着移动通信的普及，覆盖城市的大量移动蜂窝基站成为理想的探测人群分布传感器，即可以提供近乎实时的、精细到百米级别的群体移动行为数据。但是因

为电信数据的敏感性，这样理想的群体移动数据极难获得和使用，极大地阻碍了相关研究<sup>[31]</sup>。具体而言，围绕这一科学问题的主要研究内容如下。

1) 基于电信数据等高精度、大规模群体移动数据，分析城市范围内人群的时空分布特点和迁移规律，给出预测群体移动行为的关键因素和性能分析方法<sup>[31]</sup>。

2) 考虑群体移动行为的时空关联性，研究设计基于卷积网络、深度残差网络、生成对抗网络等深度学习模型的、从粗粒度人口普查分布向细粒度的人群分布的算法，并基于采集的真实时空行为大数据测试实际性能<sup>[32]</sup>。

3) 考虑群体移动行为的时间连续性，研究基于递归神经网络优化空间分布映射结果的有效性和实现动态映射实时估计的可行性。

### 2.3.4 移动数据隐私保护机制

在城市环境的移动行为建模中，隐私保护机制的主要目标是实现在保护用户隐私前提下的数据共享、发布与挖掘，为移动行为数据的有效流通提供基础。在这一目标下，群体和个体的移动行为数据中现行的隐私保护机制主要为匿名化处理和聚合化处理策略，即在共享移动行为数据前去除个体移动数据的用户标识或将群体移动数据聚合为集计形式。现有隐私保护机制的核心思路是阻止潜在攻击者获取目标用户的移动行为信息，从而消除数据集中个体用户的隐私风险。然而，近期研究工作表明用户移动行为存在强规律性、高唯一性等特点<sup>[21]</sup>。因此，攻击者有可能在匿名甚至聚合的移动行为数据中推断出个体用户的敏感信息。在此背景下，该科学问题衍生的主要研究内容为以下几点。

1) 分析移动用户行为模式中可能被攻击者利用的关键因素，建立针对匿名化处理和聚合化处理的去匿名攻击模型和轨迹恢复攻击模型，测量攻击模型在真实数据中的可行性。

2) 基于所提攻击模型分析真实移动网络大数据中的隐私风险，并探索影响隐私风险的关键因素，为实现可靠的数据隐私风险测量与保护算法奠定理论基础<sup>[33]</sup>。

3) 针对移动行为数据的典型应用场景，设计关键隐私保护算法及保障用户隐私的数据共享算法，从而为保护用户隐私的移动行为数据共享与流通奠定基础<sup>[34]</sup>。

### 3 现有研究与典型技术方案

依据 2.3 节总结的关键科学问题，本节梳理、概括了城市环境下移动行为建模的相关已有研究与典型技术方案，其中各部分研究成果的关联关系如图 2 所示。

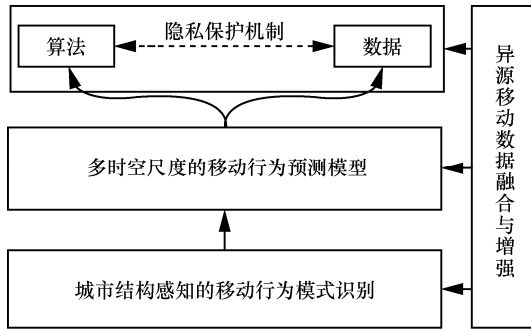


图 2 城市环境下移动行为建模主要研究内容间的关联关系

#### 3.1 异源移动数据融合与增强

伴随着移动网络应用的极大丰富，用户在网络空间中产生的行为记录往往关联到不同业务的用户标识。这些用户标识可能通过不同的移动设备进行登录，混杂在由不同数据源采集到的移动行为大数据集中。识别不同数据集中对应同一用户的不同身份标识，是实现异源移动数据融合、增强移动数据质量的关键问题。相关研究工作中，研究者提出了基于用户属性<sup>[35]</sup>、用户社交网络<sup>[36]</sup>和用户行为模式<sup>[37]</sup>的跨区域用户标识识别算法。由于移动数据是用用户行为的一个维度，本文重点讨论基于用户行为模

式的标识识别算法在异源移动数据融合中的应用。按采用的技术方案，这部分研究工作可以分为概率图模型<sup>[38]</sup>和深度学习模型<sup>[23]</sup>两大类。

文献[38]通过在物理空间中不同用户标识在同一设备或同一时空点上同时出现的频率，建立了用户标识之间关联强弱的无向有权图，其刻画了不同数据集的用户标识之间属于同一用户的概率。即在用户标识关联图上，连接越紧密的用户标识表示它们在物理空间中同时访问相同时空点的频率越高，因此有更大概率属于同一用户。基于这一思路，研究者构建了用户标识关系图上的贝叶斯推断模型，用于探测图上潜在的用户标识簇，从而在无监督信息的情况下识别对应同一用户的标识。

另一方面，文献[23]提出了一种基于孪生递归网络的用户标识关联方案。其核心思路是基于移动轨迹数据中个体移动的自相似性和不同个体间移动的差别性所提供的“弱标签”，将用户标识关联问题等效为异源移动轨迹相似度测量问题。这一等价转化使模型可以借助少量监督数据在轨迹层面推断用户标识间的关联，其模型结构如图 3 所示。核心组件包括共享多模态嵌入编码单元、孪生递归网络单元、交互式选择单元和全连接比较器，针对性地克服了基于移动数据的用户关联的关键挑战。①通过引入端到端学习的递归神经网络和孪生网络结构，实现移动轨迹特征的自动提取，取消了对人工特征工程的依赖，也使模型可以直接应用不同来源采集的移动数据；②为了充分利用移动网络中

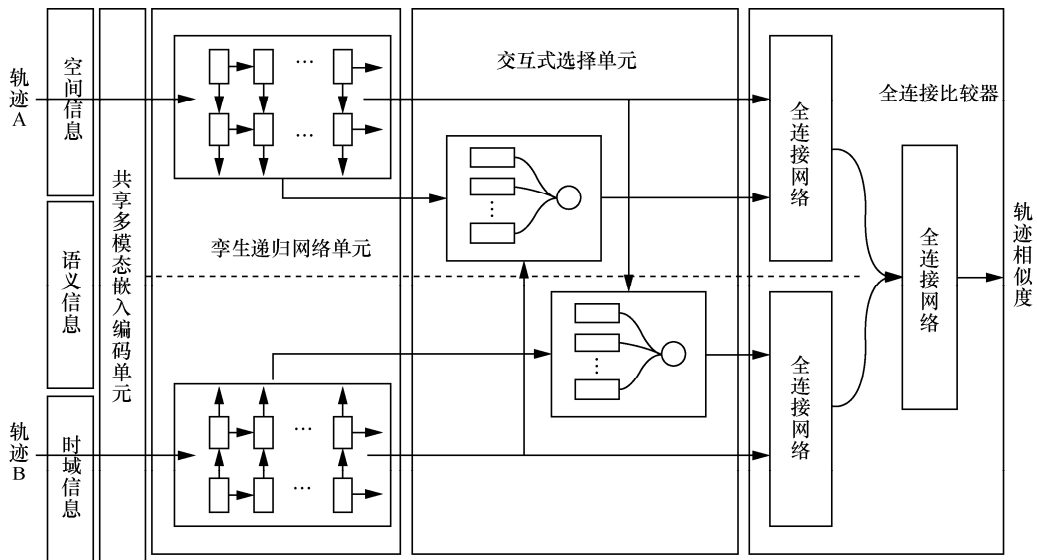


图 3 基于深度学习的轨迹相似度测量模型结构

的伴随移动行为数据的丰富辅助信息（如兴趣点和文字评论等多模态信息），该模型在递归神经网络之前设计了专门的共享多模编码模块，将移动行为数据中包括时间、空间、兴趣点、文本评论在内的多维信息进行综合表征和利用；③为了应对移动行为数据中的噪声影响，该模型引入了注意力机制，构建了交互式选择单元来对整个轨迹片段的特征进行进一步的有所侧重的选择。直观上，噪声移动数据的表征向量由于与整段移动轨迹的表征向量的相似度较低而被“过滤”，不能对移动行为相似度判断产生较大影响，同时真实的移动行为数据的表征向量被加以较高的注意力权重，从而进一步提高该模型提取相似移动轨迹特征的能力。

### 3.2 城市结构感知的移动行为模式识别

移动行为模式识别是一个由来已久的研究课题<sup>[16,39-40]</sup>。根据采用的技术路线不同，典型移动模式挖掘算法可以分为基于共有移动子序列挖掘和基于轨迹相似度测量两类。基于共有移动子序列挖掘的相关研究的主要目标是识别移动数据中多个个体共有的移动序列，其代表性工作有Giannott等<sup>[16]</sup>提出的T-pattern模型，用于识别移动数据集中频繁出现的移动子序列；Mamouli等<sup>[39]</sup>则研究了周期性频繁移动序列的识别问题；Lee等<sup>[8]</sup>提出分割-聚类法，首先识别轨迹中频繁出现的子序列，然后根据频繁移动子序列聚类移动行为模式。在移动轨迹的相似度测量的相关研究中，Zhang等<sup>[40]</sup>通过主成分分析法对移动行为数据进行隐特征提取，然后在隐空间中测量移动轨迹的相似性并进行聚类分析；Yao等<sup>[41]</sup>则通过循环神经网络为轨迹寻求表征向量，从而刻画轨迹在位置、速度、加速度上的相似程度；动态时间规整（DTW, dynamic time warping）<sup>[42]</sup>和最大共同子序列（LCSS, longest common sub-sequence）<sup>[43]</sup>模型则实现了对于轨迹序列时间差异和移动顺序的相似性建模。已有的研究工作利用物理空间远近或共同出现次数衡量移动行为差异，同一类模式局限于相近的物理空间，难以识别城市结构感知的移动行为模式。例如，具有相同目的但空间相隔较远的移动模式（比如从住宅区前往办公区上班的移动模式）无法被识别。

针对这一挑战，近期的2个工作研究了城市结构感知的移动模式识别问题<sup>[28,44]</sup>。文献[44]观察到城市结构感知的移动模式识别的难点在于移动意

图与物理空间的耦合，即现有移动轨迹相似度测量方法无法刻画用户移动行为在城市结构层面的相似性。基于这一观察，文献[44]进一步提出将用户的移动轨迹抽象为在不同城市结构上的时间划分方式，从而使移动行为与具体的物理位置解耦，在物理空间中相距很远的用户也可以提取相似的表征。在这种移动行为表征方式下，研究者设计了移动行为相似度测量方法与基于聚类分析的移动行为模式发现算法，并在真实数据中验证了算法的有效性。另一方面，文献[28]在其基础上设计了基于深度表征学习技术的保留城市结构语义信息的移动轨迹表征方法。该模型的核心思路是将移动行为数据中的时空信息与访问城市结构的行为信息分离，并将移动轨迹数据表示为用户在离散的时间片序列上在不同城市结构间的跳转方式。例如，用户可能上午8点处于住宅区，而11点处于办公区。然后，该模型采用深度表征学习技术，在抽象的移动行为序列上学习城市结构感知的移动行为模式的低维表征，并通过聚类分析的方式发现数据集中的主要移动行为模式，其流程框架如图4所示。这一方法采用新兴的深度表征学习技术，将移动行为模式投影到低维向量，实现了便捷、高效的移动行为模式相似度度量。同时其可以有效地保留移动行为模式在城市结构层面的语义，从而实现城市结构感知的移动行为模式识别。

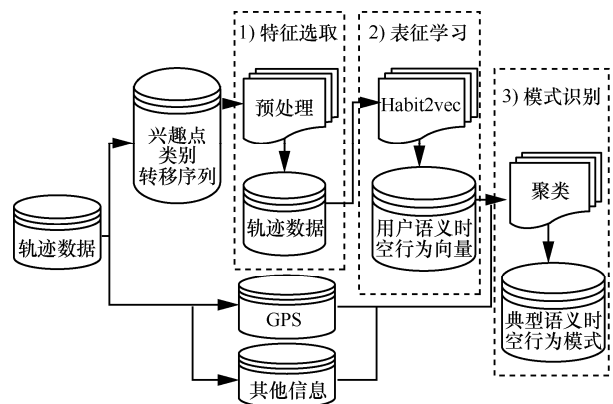


图4 城市结构感知的移动模式识别算法流程框架

### 3.3 多时空尺度的移动行为预测模型

移动行为预测的相关研究按照研究对象是个体移动用户还是群体移动用户可以分为两类。其中，传统的个体移动行为预测模型的准确度与实际应用要求存在差距。根据技术方案，已有研究工作又可以分为基于模式的方法和基于模型的方法<sup>[27,45]</sup>。

一方面, 基于模式的方法首先从移动数据发现其中蕴含的若干时空行为模式, 然后基于有限的时空行为模式进行移动行为预测, 此类方法局限于少数流行的移动模式而忽略了个体的偏好<sup>[27]</sup>。另一方面, 基于模型的方法大多采用经典的时间序列模型建模时空点之间的转移关系, 虽然可以根据个体数据学习不同模型, 但是受困于序列建模工具只能捕捉浅层低阶、时不变相关性的局限, 此类方法取得的预测性能依旧有限<sup>[45]</sup>。随着深度学习技术的日益成熟, 理论上可以建模时变、高阶、复杂转移特性的递归神经网络成为研究热点, 但是目前移动行为预测的大部分相关工作只是对递归神经网络的直接应用<sup>[46-47]</sup>, 在针对移动行为数据特性的研究和设计上十分不足。另一方面, 随着终端智能设备计算能力的快速提升, 城市环境下的移动行为数据往往包含了个体使用的移动应用、访问不同城市结构、社交互动等多维度的行为信息<sup>[48]</sup>。如何在基于深度学习的移动行为预测模型中结合这些复杂的行为信息来提升性能, 进一步预测个体移动对应的目的和行为类型, 是本文当前的重点研究方向。

文献[30]针对性地设计了基于深度神经网络的城市环境移动行为预测模型, 以实现高阶复杂时空相关性的准确建模。总体而言, 文献[30]所提方案首先设计了特征抽取模块来整合多维度信息输入, 然后利用递归神经网络建模高阶时变的状态转移, 并引入注意力机制来捕捉个体移动中的多尺度周期性, 最后是全连接分类映射网络, 具体的技术框架如图 5 所示。

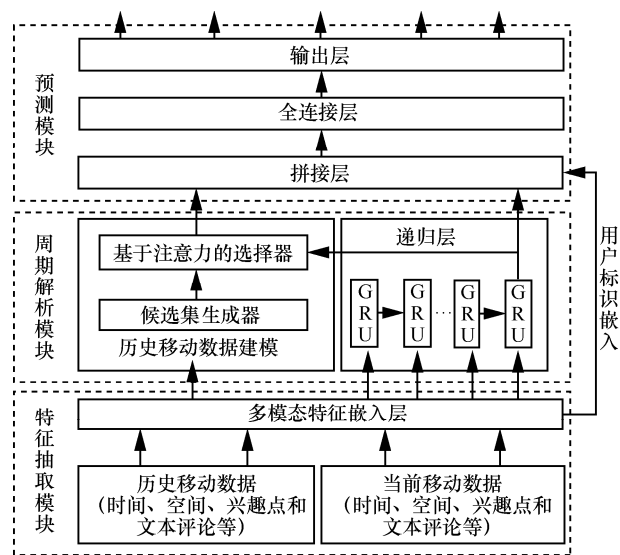


图 5 基于深度注意力机制的个体移动预测方法技术框架

文献[30]所提方案除预测模块外主要由 3 个部分构成。①为了综合利用轨迹数据中包括时间、空间、兴趣点和文本评论等在内的多维度信息, 该模型设计了包含 2 个组件的多模态编码模块: 第一个组件独立地对各个维度的特征进行编码, 得到各个维度特征的稠密表征; 第二个组件拼接和整合所有维度特征, 从而得到移动轨迹数据中基本元素的综合表示。基于多模态编码模块的表征能力, 模型将得到整合多维度信息的能力, 可以灵活地融合异构的可用信息。②为了建模个体移动过程中复杂的高阶状态转移关系, 该模型引入递归神经网络。由于原生的递归神经网络在训练过程中存在随着序列的延长出现梯度爆炸或者梯度消失而导致其训练困难、建模能力有限的问题, 该模型采用了多种专门的递归单元来克服这些问题, 其中长短时记忆 (LSTM, long short term memory) 递归神经网络和门控循环单元 (GRU, gated recurrent unit) 递归神经网络因其训练的高效性和表现的稳定性而被广泛使用于自然语言处理等复杂场景。③为了解决个体移动轨迹在较长时间尺度上表现出的较强的多尺度周期性, 该模型设计了基于注意力机制的周期提取单元。具体而言, 其首先将个体移动轨迹分为历史轨迹和当前轨迹两部分, 当前轨迹正常通过递归神经网络完成复杂转移关系的建模, 而历史轨迹部分则将通过专门的周期提取单元, 并将在当前轨迹的挑选下得到历史轨迹中的周期性部分来辅助轨迹预测。

在群体移动行为预测方面, 由于数据的缺乏和现有模型的局限性, 如何准确地预测人类群体的时空分布一直是学术界和公共管理部门关心的难点问题<sup>[31,49-50]</sup>。已有研究工作根据主要使用的数据来源可以分为两部分: 第一部分主要基于遥感数据并综合使用土地资源等数据, 使用统计方法完成人口的空间分布估计<sup>[51]</sup>, 此类方法因为数据本身的限制只能完成简单的夜间人群空间分布估计, 不可能实现实时的群体移动行为预测; 另一方面, 随着移动设备的普及, 大规模的电信信令数据为研究者提供了实时的群体分布采样数据, 极大地推动了群体移动行为的研究<sup>[52]</sup>。基于电信系统记录的用户人数和实际群体分布间存在的幂律关系, 相关研究综合考虑土地功能等辅助数据用统计模型的方式建立了电信数据到群体分

布的映射关系。然而，由于统计模型刻画复杂关联关系的能力十分有限，此类研究难以在预测准确性上取得进一步发展<sup>[53]</sup>。随着深度学习在计算机视觉方向的成功应用，部分学者考虑将真实的地理空间转换为二维平面空间，从而可以引入计算机领域视觉成熟的工具和方法<sup>[53-54]</sup>。

受计算机视觉领域基于神经网络的超分辨率技术的启发，文献[32]将群体移动行为预测问题转化为基于粗粒度人群分布数据的超分辨率问题。如图 6 所示，借助城市区域的空间网格化，其把城市区域的群体空间分布视为一张特殊的图片，给定该区域的粗粒度人口普查数据相当于给定一张模糊的图片，该区域的细粒度人群移动行为则相当于一张高清的图片。在计算机视觉领域中，图像超分辨率问题即是通过图片的低分辨率结果来生成该图片的高分辨率结果，因此存在重要的借鉴价值。随着深度学习的发展，图像超分辨率模型在近期取得了较大进展。不同于经典的立方插值等简单的公式插值法，基于深度卷积网络和对抗生成网络的模型在图片质量上都取得了巨大的进步。

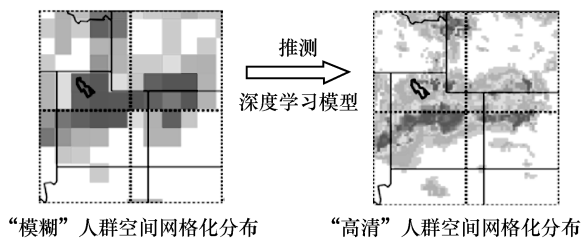


图 6 群体移动行为预测问题示意

文献[32]所提模型采用了深度卷积网络和对抗生成网络来建模人群分布的空间映射关系，基本的研究思路如图 7 所示，可分为 3 个部分。①数据融合：考虑到城市功能分布、交通系统等城市空间结构对群体移动行为（粗粒度人口分布）的影响，其同时融合兴趣点分布图、城市交通分布图等多种信息源作为辅助数据输入。②静态空间分布估计：其首先基于深度卷积网络刻画人群分布的复杂空间映射关系，同时辅助使用对抗生成网络来捕捉空间分布的高频细节，避免空间分布估计结果过于平滑而失去部分细节。③动态分布估计：不同于普通的图片超分辨率问题本质上是静态的一对一映射学习，因为该研究希望从静态的粗粒度人群分布得到动态的群体移动行为预

测，其本质上是动态的一对多映射。为了解决时间因素带来的动态变化难题，其引入了递归神经网络来对人群分布的时间特性进行建模。

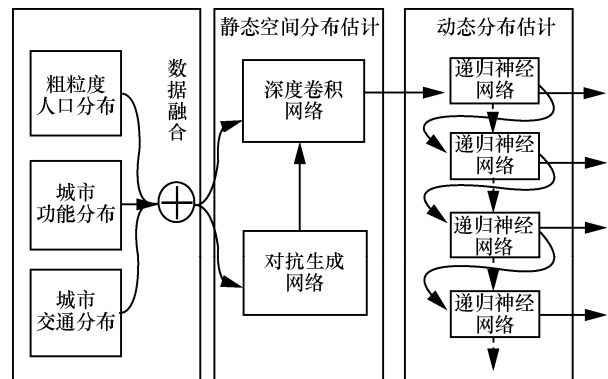


图 7 基于时空关联建模的群体移动行为预测框架

### 3.4 移动数据隐私保护机制

在面向数据共享的隐私保护方案中， $k$  匿名<sup>[55]</sup>和差分隐私<sup>[56]</sup>是当前最广泛采用的数据隐私保护算法。 $k$  匿名算法的目标是使攻击者无法基于任意背景知识，实现对匿名用户的唯一识别。因此，其一般采用数据泛化的方法使任一用户的数据与至少  $k-1$  个其他用户的数据无法区分，从而保证每个用户的匿名性。在此基础上， $l$ -多样<sup>[57]</sup>和  $t$ -相似<sup>[58]</sup>框架被陆续提出以提供更强的隐私保护。另一方面，差分隐私策略一般采取添加噪声的方式，使数据库在包含和不包含目标用户时不会表现出显著差异<sup>[56,59]</sup>。

在移动数据隐私保护方面，已有研究显示移动行为的高唯一性和强规律性使匿名个体移动数据中可能存在显著隐私风险<sup>[20-21]</sup>。具体地，攻击者可能基于少量的额外移动数据（如对目标用户住宅区和办公区的观测），准确匹配出目标用户在匿名数据集集中的移动轨迹<sup>[21]</sup>。其可以细分为去匿名攻击和概率推断攻击 2 种<sup>[60]</sup>，如图 8 所示。面对这一隐私风险，近期的研究工作<sup>[60]</sup>设计了基于移动轨迹合并的隐私保护算法。其核心思想是通过时空泛化的方法实现不同用户的移动轨迹间的合并，从而保证匿名个体移动数据满足  $k$  匿名要求，而攻击者无法基于任意外部信息实现对目标用户的攻击。这一方法的重点问题在于如何将用户群体分为若干不少于  $k$  个用户的匿名组，从而使时空泛化处理后的数据可用性损失最小。文献[61]提出了一种启发式的组合优化方法，实现了在数据可用性和用户隐私间的高效取舍。

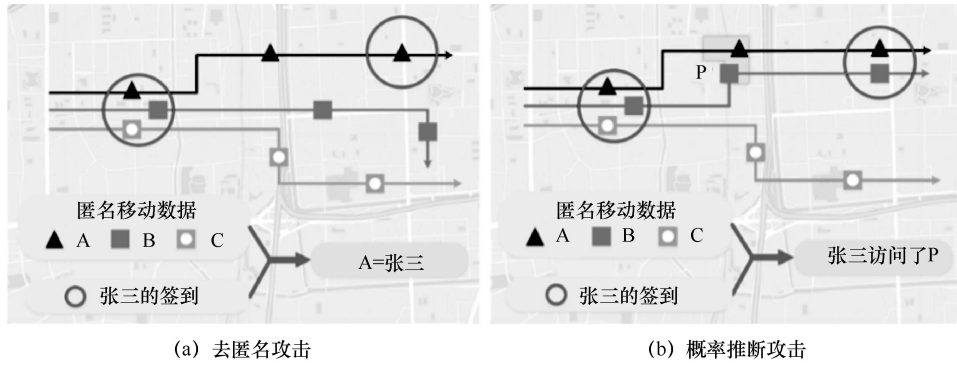


图 8 匿名个体移动数据隐私攻击模型示意

## 4 结束语

在全球城市化不断推进、大数据时代已经到来的背景下，城市环境下的移动行为建模对于诸多关键应用有重要价值。本文讨论了这一问题的重点挑战，构建了针对性的研究框架。此外，本文进一步梳理总结了移动行为数据增强算法、城市结构感知的移动行为模式识别、多时空尺度的移动行为预测模型和移动数据隐私保护机制 4 个关键研究问题的发展脉络与研究现状，为该领域的未来研究打下了基础。

### 参考文献:

[1] DEVILLE P, LINARD C, MARTIN S, et al. Dynamic population mapping using mobile phone data[J]. Proceedings of the National Academy of Sciences, 2014, 111(45): 15888-15893.

[2] 单留举, 王晓东, 马英运. 基于大数据的用户学习偏好建模及应用[J]. 计算机应用与软件, 2016, 33(1): 77-80.

SHAN L J, WANG X D, MA Y Y. User learning preference and application based on big data[J]. Computer Applications and Software, 2016, 33(1): 77-80.

[3] ZHENG K, YANG Z, ZHANG K, et al. Big data-driven optimization for mobile networks toward 5G[J]. IEEE Network, 2016, 30(1): 44-51.

[4] 尹浩, 乔波. 大数据驱动的网络信息平面[J]. 计算机学报, 2016, 39(1): 126-139.

YIN H, QIAO B. Big data-driven network information plane[J]. Chinese Journal of Computers, 2016, 39(1): 126-139.

[5] NAIMI A I, WESTREICH D J. Big data: a revolution that will transform how we live, work, and think[J]. Mathematics & Computer Education, 2013, 47(17): 181-183.

[6] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.

LI G J, CHEN X Q. Big data research: a major strategic field of future technology and economic and social development-research status and

scientific thinking of big data[J]. Bulletin of Chinese Academy of Sciences, 2012, 27(6): 647-657.

[7] BROCKMANN D, HUFNAGEL L, GEISEL T. The scaling laws of human travel[J]. Nature, 2006, 439(7075): 462.

[8] LEE J G, HAN J, WHANG K Y. Trajectory clustering: a partition-and-group framework[C]//ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2007: 593-604.

[9] JIANG S, YANG Y, GUPTA S, et al. The TimeGeo modeling framework for urban motility without travel surveys[J]. Proceedings of the National Academy of Sciences, 2016, 113(37): e5370.

[10] QIAO S, SHEN D, WANG X, et al. A self-adaptive parameter selection trajectory prediction approach via hidden Markov models[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(1): 284-296.

[11] CAMAGNI R, GIBELLI M C, RIGAMONTI P. Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion[J]. Ecological Economics, 2002, 40(2): 199-216.

[12] SCHRANK D, EISELE B, LOMAX T. TTI's 2012 urban mobility report[J]. The Texas A&M University System, 2012, 4(1): 1-64.

[13] GOODSPEED R, YAN X, HARDY J, et al. Comparing the data quality of global positioning system devices and mobile phones for assessing relationships between place, mobility, and health: field study[J]. JMIR mHealth and uHealth, 2018, 6(8): e168.

[14] LOPEZ A J, SEMANJSKI I, GAUTAMA S, et al. Assessment of smartphone positioning data quality in the scope of citizen science contributions[J]. Mobile Information Systems, 2017, 2017(1): 1-11.

[15] BASSOLAS A, BARBOSA-FILHO H, DICKINSON B, et al. Hierarchical organization of urban mobility and its connection with city livability[J]. Nature Communications, 2019, 10(1): 1-10.

[16] GIANNOTTI F, NANNI M, PINELLI F, et al. Trajectory pattern mining[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 330-339.

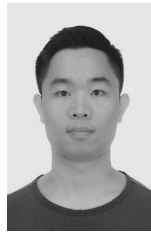
[17] HASAN S, SCHNEIDER C M, UKKUSURI S V, et al. Spatiotemporal patterns of urban human mobility[J]. Journal of Statistical Physics, 2013, 151(1-2): 304-318.

[18] FAN C, LIU Y, HUANG J, et al. Correlation between social proximity and mobility similarity[J]. Scientific Reports, 2017, 7(1): 1-8.

- [19] LI H, MA D, MEDJAHED B, et al. Data privacy in the emerging connected mobility services: architecture, use cases, privacy risks, and countermeasures[J]. *SAE International Journal of Transportation Cybersecurity and Privacy*, 2019, 2(1): 49-61.
- [20] SONG C, QU Z, BLUMM N, et al. Limits of predictability in human mobility[J]. *Science*, 2010, 327(5968): 1018-1021.
- [21] DE MONTJOYE Y A, RADAELLI L, SINGH V K. Unique in the shopping mall: on the reidentifiability of credit card metadata[J]. *Science*, 2015, 347(6221): 536-539.
- [22] YOU L, ZHAO F, CHEAH L, et al. A generic future mobility sensing system for travel data collection, management, fusion, and visualization[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 1(1): 1-12.
- [23] FENG J, ZHANG M, WANG H, et al. DPLink: user identity linkage via deep neural network from heterogeneous mobility data[C]//*The World Wide Web Conference*. New York: ACM Press, 2019: 459-469.
- [24] HOTEIT S, CHEN G, VIANA A, et al. Filling the gaps: on the completion of sparse call detail records for mobility analysis[C]//*Proceedings of the Eleventh ACM Workshop on Challenged Networks*. New York: ACM Press, 2016: 45-50.
- [25] HOYT H. The structure and growth of residential neighborhoods in American cities[M]. Washington: US Government Printing Office, 1939.
- [26] WU F. Mining heterogeneous data for semantic understanding of mobility data[D]. Pennsylvania: Pennsylvania State University, 2018, 1(1): 1-120.
- [27] CHEN M, LIU Y, YU X. NLPMM: a next location predictor with Markov modeling[C]//*Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Berlin: Springer, 2014: 186-197.
- [28] CAO H, XU F, SANKARANARAYANAN J, et al. Habit2vec: trajectory semantic embedding for living pattern recognition in population[J]. *IEEE Transactions on Mobile Computing*, 2019, 19(5): 1096-1108.
- [29] LIU W, SHOJI Y. DeepVM: RNN-based vehicle mobility prediction to support intelligent vehicle applications[J]. *IEEE Transactions on Industrial Informatics*, 2019, 16(6): 3997-4006.
- [30] FENG J, LI Y, ZHANG C, et al. Deepmove: predicting human mobility with attentional recurrent networks[C]//*Proceedings of the 2018 World Wide Web Conference*. New York: ACM Press, 2018: 1459-1468.
- [31] XU F, ZHANG P, LI Y. Context-aware real-time population estimation for metropolis[C]//*Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York: ACM Press, 2016: 1064-1075.
- [32] ZONG Z, FENG J, LIU K, et al. DeepDPM: dynamic population mapping via deep neural network[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2019: 1294-1301.
- [33] GAO J, SUN L, CAI M. Quantifying privacy vulnerability of individual mobility traces: a case study of license plate recognition data[J]. *Transportation Research Part C: Emerging Technologies*, 2019, 104: 78-94.
- [34] YOUNG M, RODRIGUEZ L, KELLER E, et al. Beyond open vs. closed: balancing individual privacy and public accountability in data sharing[C]//*Proceedings of the Conference on Fairness, Accountability, and Transparency*. [S.l.:s.n.], 2019: 191-200.
- [35] GOGA O, LOISEAU P, SOMMER R, et al. On the reliability of profile matching across large online social networks[C]//*Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2015: 1799-1808.
- [36] NARAYANAN A, SHMATIKOV V. De-anonymizing social networks[C]//*2009 30th IEEE Symposium on Security and Privacy*. Piscataway: IEEE Press, 2009: 1-19.
- [37] DE MONTJOYE Y A, HIDALGO C A, VERLEYSEN M, et al. Unique in the crowd: the privacy bounds of human mobility[J]. *Scientific Reports*, 2013, 3: 1376.
- [38] WANG H, LI Y, WANG G, et al. You are how you move: linking multiple user identities from massive mobility traces[C]//*Proceedings of the 2018 SIAM International Conference on Data Mining*. [S.l.:s.n.], 2018: 189-197.
- [39] MAMOULIS N, CAO H, KOLLIOS G, et al. Mining, indexing, and querying historical spatiotemporal data[C]//*ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2004: 236-245.
- [40] ZHANG K, LIN Y R, PELECHRINIS K. Eigen Transitions with hypothesis testing: the anatomy of urban mobility[C]//*International AAAI Conference on Web and Social Media*. Palo Alto: AAAI Press, 2016: 486-495.
- [41] YAO D, ZHANG C, ZHU Z, et al. Trajectory clustering via deep representation learning[C]//*International Joint Conference on Neural Networks*. Piscataway: IEEE Press, 2017: 3880-3887.
- [42] VLACHOS M, HADJIELEFTHERIOU M, GUNOPULOS D, et al. Indexing multi-dimensional time-series with support for multiple distance measures[C]//*ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2003: 216-225.
- [43] GAO Q, ZHOU F, ZHANG K, et al. Identifying human mobility via trajectory embeddings[C]//*International Joint Conference on Artificial Intelligence*. [S.l.:s.n.], 2017:1689-1695.
- [44] XU F, XIA T, CAO H, et al. Detecting popular temporal modes in population-scale unlabelled trajectory data[J]. *Proceedings of the ACM on IMWUT*, 2018, 2(1): 46.
- [45] CHENG C, YANG H. Where you like to go next: successive point-of-interest recommendation[C]//*International Joint Conference on Artificial Intelligence*. [S.l.:s.n.], 2013: 2605-2611.
- [46] MATHEW W, RAPOSO R, MARTINS B. Predicting future locations with hidden Markov models[C]//*ACM Conference on Ubiquitous Computing*. New York: ACM Press, 2012: 911-918.

- [47] DU N, DAI H, TRIVEDI R, et al. Recurrent marked temporal point processes: embedding event history to vector[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 1555-1564.
- [48] ZHANG C, ZHANG K, YUAN Q, et al. GMove: group-level mobility modeling using geo-tagged social media[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 1305-1314.
- [49] DU H, YU Z, YI F, et al. Recognition of group mobility level and group structure with mobile devices[J]. IEEE Transactions on Mobile Computing, 2018, 17(4): 884-897.
- [50] YU Z, YI F, LV Q, et al. Identifying on-site users for social events: mobility, content, and social relationship[J]. IEEE Transactions on Mobile Computing, 2018, 17(9): 2055-2068.
- [51] DONG C, CHEN C L, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.
- [52] VHADURI S, POELLABAUER C. Hierarchical cooperative discovery of personal places from location traces[J]. IEEE Transactions on Mobile Computing, 2018, 17(8): 1865-1878.
- [53] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]// Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 1655-1661.
- [54] VANDAL T, KODRA E, GANGULY S, et al. DeepSD: generating high resolution climate change projections through single image super-resolution[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2017: 1663-1672.
- [55] SWEENEY L. k-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.
- [56] DWORK C. Differential privacy: a survey of results[C]//International Conference on Theory and Applications of Models of Computation. Berlin: Springer, 2008: 1-19.
- [57] MACHANAVAJHALA A, GEHRKE J, KIFER D, et al. l-diversity: privacy beyond k-anonymity[C]//22nd International Conference on Data Engineering. Piscataway: IEEE Press, 2006: 24.
- [58] LI N, LI T, VENKATASUBRAMANIAN S. t-closeness: privacy beyond k-anonymity and l-diversity[C]//2007 IEEE 23rd International Conference on Data Engineering. Piscataway: IEEE Press, 2007: 106-115.
- [59] RASTOGI V, NATH S. Differentially private aggregation of distributed time-series with transformation and encryption[C]//Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2010: 735-746.
- [60] GRAMAGLIA M, FIORE M. Hiding mobile traffic fingerprints with glove[C]//Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies. New York: ACM Press, 2015: 26.
- [61] XU F, TU Z, HUANG H, et al. No more than what I post: preventing linkage attacks on check-in services[C]//The World Wide Web Conference. New York: ACM Press, 2019: 3405-3412.

## [作者简介]



徐丰力 (1993- )，男，广东深圳人，清华大学博士生，主要研究方向为移动数据挖掘、城市计算、数据隐私。



李勇 (1985- )，男，湖南长沙人，博士，清华大学副教授、博士生导师，主要研究方向为数据挖掘、城市计算、移动计算。