

请参阅本出版物的讨论、统计和作者简介，网址为：  
<https://www.researchgate.net/publication/329001637>

基于 GPS 的自动体力活动和车内状态分类

和加速度计数据:使用机器的分级分类方法

学习技术

GIS 交易中的文章，2018 年 11 月

DOI : 10.1111 / tgis。 12485

引文

0

2 名作者:

李康宰

伊利诺伊大学厄巴纳-香槟分校

10 出版物 38 引文

见简介

该出版物的一些作者也在从事这些相关项目:

311 非紧急市政服务视图项目

使用加速度计和 GPS 数据视图项目的身体活动分类

读

22

关美宝

伊利诺伊大学厄巴纳-香槟分校

230 份出版物 8, 757 次引用

见简介

本页之后的所有内容都是由李康宰于 2018 年 11 月 17 日上传的。

用户请求增强下载的文件。

DOI : 10.1111 / tgis. 12485

研究文章

基于 GPS 和加速度计的自动身体活动和车内状态分类

数据:一种分层分类方法, 使用

机器学习技术

康在利·| mei po·关

伊利诺伊大学厄巴纳-香槟分校, 伊利诺伊州厄巴纳

一致

伊利诺大学厄巴纳香槟分校伊利诺信息研究所李康宰, 美国伊利诺伊州厄巴纳西克拉克 1205 号, 邮编: 61801。电子邮件: kasbiss@gmail.com

### Abstract

Due to the advancement of tracking technology, a large quantity of movement data has been collected and analyzed in various research domains. In human mobility and physical activity (PA) research, GPS trajectories and the capabilities of geographic information systems (GIS) facilitate a better understanding of the associations between PA and various environmental factors taking individuals' daily travels into account. PA research, however, needs to widen its focus from the intensity of PA to types of PA, which may provide useful clues for understanding specific health behaviors in particular geographic contexts. This study proposes and develops an algorithm to automatically classify PA types and in-vehicle status using GPS and accelerometer data. Walking, standing, jogging, biking and sedentary/in-vehicle statuses are identified through hierarchical classification processes based on machine learning and geospatial techniques. The proposed algorithm achieved high predictive accuracy on real-world GPS and accelerometer data. It can greatly reduce participants' and researchers' burdens by automatically identifying PA types and in-vehicle status for human mobility research, which is also known as travel mode imputation in transportation research.

由于跟踪技术的进步, 在各个研究领域收集和分析了大量运动数据( Eagle、Pentland 和 Lazer, 2009 年; 冈萨雷斯, 伊达尔戈和巴拉巴斯, 2008 年; 瑟瓦尔和艾萨克森, 2007 年)。作为跟踪设备之一, 全球定位系统( GPS )接收器已被广泛用于收集数据, 以增强我们对运动物体(如动物)时空动态的理解( Dodge 等人)。, 2013 年; Laube, Dennis, Forer 和 Walker, 2007 ), 人类( Kwan, 2004; 瑟瓦尔等人。 , 2011 年; 王、坤和柴, 2018 年), 或车辆(唐斯和霍恩, 2012 年; 费雷拉, Poco, Vo, Freire 和 Silva, 2013 年)。在体育活动( PA )研究中, GPS 轨迹和地理信息系统( GIS )的能力有助于更好地理解中等至剧烈体育活动( MVPA )与各种环境因素之间的联系, 如快步行走和跑步, 同时考虑到个人的日常旅行( Almanza、Jerrett、Dunton、Seto 和 Pentz, 2012 年; 内森·博鲁夫和奈因斯坦, 2012 年; 库珀等人。 , 2010 年; Jansen, Ettema, Pierik 和 Dijst, 2016 年; 拉乔维茨, 琼斯, 佩奇, 惠勒和库珀, 2012 年; 罗德里格斯等人。 , 2012 年; 特罗佩, 威尔逊, 马修斯, 克罗姆雷和梅利, 2010 年)。这些最近使用 GPS 轨迹的 PA 研究揭示了非居住区在人们健康行为和结果中的重要性, 以及居住区的重要性( Diez Roux & Mair, 2010 年; Perchoux, Chaix, Cummins 和 Kestens, 2013 年)。高分辨率 GPS 数据也可用于缓解不确定的地理环境问题, 因为它们有助于识别可能影响人们健康的各种非居住环境( Kwan, 2012a, b, 2013 年)。

PA 研究主要使用客观测量的加速度计数据来评估人们 PA 的强度, 如 MVPA, 因为客观 PA 测量比主观和自我报告的测量产生更重要的发现( Browning & Lee, 2017 年)。然而, PA 研究需要将其重点从 PA 的强度扩大到 PA 的类型, 这可能为理解特定地理环境下的特定健康行为提供有用的线索( Jankowska, Schipperijn 和 Kerr, 2015 年)。一些 PA 概念模型表明, 特定类型的 PA, 如步行和骑自行车, 与特定的环境或地理环境有关联, 如步道、安全人行道和辅助设施。然而, 需要更一致的经验证据来支持这种关联( Lokaitu - Sideris, 2006 年; Sallis, Bauman 和 Pratt, 1998 )。此外, 迄今为止, 很少有 PA 研究使用 GPS 和加速度计数据考虑了机动化交通模式, 如乘坐汽车(例如 Voss 等人)。 , 2016 年)。根据 Gordon - Larsen、Nelson 和 Beam ( 2005 年), 非机动车交通工具(如步行和骑自行车)在通勤中的高利用率主要表现在符合 PA 建议的年轻人群体中, 而不符合 PA 建议的年轻人比例较高与更高地利用机动交通工具(如公共交通工具和私家车)前往工作场所或学校有关。此外, 当人们呆在车内时, 各种环境因素对他们的 PA 的影响也可能比他们在外行走或跑步时小。因此, 为了更准确地估计人们暴露在各种环境中的情况, 识别和区分机动交通模式(基于车辆的运动)和 PA (非基于车辆的人类运动)非常重要。然而, 只有少数 PA 研究试图通过自动分类旅行模式和 PA 来识别人们是否在车辆上旅行(车内状态)( Ellis 等人)。 , 2014 年; 周, 2014 年)。

为了对该文献有所贡献, 本研究提出并开发了一种算法, 利用公众可获得的 GPS 和加速度计数据对 PA 类型和车内状态进行自动分类。基于机器学习技术的分层分类过程是本研究采用的创新方法。作为人工智能的一个分支, 机器学习通过基于使用大量输入数据的机器学习模型/算法的训练过程来改进对结果的预测。分层分类的引入是为了使用异构传感器数据, 如 GPS 和加速度计数据, 而不是仅使用其中一种数据, 共同识别更多的类别(用独特的标签来识别), 如本研究中的不同类型 PA。第 3 节详细讨论了分级分类过程。在本研究中, 三个组成部分构成了分级分类算法的框架:室内/室外分类、使用 GPS 数据的分类(室外)和使用加速度计数据的分类(室内)。机器学习技术使

基于从收集的 GPS 和加速度计数据中生成的数字或分类特征, 预测人们的 PA 类型和车内状态。作为可测量的特征, 特征是根据输入数据计算出的信息性可量化属性, 是机器学习中用于预测不同类别(例如, 本研究中的 PA 类型)的基本要素之一。关于预测的类别, 四种 PA 类型(骑自行车、跑步、步行、站立)、坐着(久坐状态)和在车辆中行驶(车内状态)通过开发的算法自动识别。因为跑步和骑自行车是美国年轻人最喜欢的两种户外 PA 类型(户外基金会, 2016 年), 提议的算法将试图识别这两种 PA 类型。

使用从高度和中度城市化地区的三个受试者收集的真实世界 GPS 数据, 对所提出的算法进行了定量和定性验证。这里的高度城市化地区是指建筑密度高、高楼多、交通繁忙的地区, 而中度城市化地区是指建筑密度适中、高楼少、交通适中的地区。本研究选择高度和中度城市化地区的理由是城市化程度(如高楼和交通拥堵)可能会影响 GPS 测量的准确性。例如, 在高度城市化地区, 由于高层建筑(用 GPS 术语来说称为城市峡谷)对 GPS 信号的阻碍, GPS 位置误差可能高于中度城市化地区。此外, 在高度城市化地区的某些高度拥挤路段, 行驶速度往往非常慢, 这可能会影响车内状态的分类精度, 因为当移动物体静止或移动缓慢时, 其 GPS 位置会变得不稳定。因此, 该算法的性能需要在这两类城市化水平不同的地区进行探索。

在这项研究中, 芝加哥被选为高度城市化地区的例子, 伊利诺伊州香槟被选为中度城市化地区的例子。芝加哥是美国第二大城市, 最大的摩天大楼高达 490 英尺(150 米)或更高, 而香槟没有摩天大楼(高层建筑和城市人居委员会, 2018 年)。芝加哥在 2014 年美国最大的城市地区中的总行程延误和总拥堵成本排名第三(Schrank、Eisele、Lomax 和 Bak, 2015 年)。关于总的旅行延误, 香槟在 2014 年有 1,966,000 小时的额外旅行时间, 远远少于芝加哥(302,609,000 小时)。

该算法通过自动分类 PA 类型和车内状态, 大大减轻了参与者记录旅行模式的负担, 因此对使用 GPS 的人员移动性研究很有用, 包括交通和公共健康研究。研究人员也可以从自动分类中受益, 因为不需要担心参与者丢失或不准确的旅行模式记录。自动行驶模式插补也将有助于识别与特定路线选择相关的行驶模式, 并有助于了解路线选择研究中的基本决策过程(Rab、Dill 和 Gliebe, 2012 年; 帕宾斯基, 斯科特和多尔蒂, 2009 年)。此外, 这项研究显示了公布的异构传感器数据(GPS 和加速度计数据)如何能够一起使用机器学习和地理空间技术提供准确和自动的 PA 分类。

本文的其余部分分为四个部分。第 2 节介绍了以前使用 GPS 和加速度计数据对 PA 类型和运输模式分类的研究。第 3 节描述了算法的组成部分, 重点是室内/室外分类和使用 GPS 数据的分类。第 4 节验证了所提出的算法, 第 5 节讨论了研究结果并给出了结论。

## 2 | PAST 对 PA 和运输方式分类的研究

已经进行了许多 PA 识别研究, 充分利用了智能感知能力的潜力

特别关注健康促进和管理的手机。特别是, 大多数智能手机中普遍配备的加速度计传感器使用机器学习技术, 以高精度为 PA 的自动识别带来了深刻的增强, 包括跑步、行走和坐着 (Anguita、Ghio、Oneto、Parra 和 Reyes - Ortiz, 2012 年; Arif, Bilal, Kattan 和 Ahamed, 2014 年; 夸皮斯, 维斯和摩尔, 2011 年; 韦斯等人

艾尔, 2016 年; 张, 麦库拉, 纽金特和郑, 2010 年)。例如, PA 分类被用于基于网络的应用程序中, 以监控儿童、肥胖人群或老年人的 PA, 并鼓励他们在日常生活中执行高效 Weiss 等人)。, 2016 年)。

由于基于纸或电话的旅行调查的低响应率和高不完整率,利用 GPS 轨迹自动识别不同的旅行模式引起了交通研究人员的关注。例如,郑、陈、李、谢和马(2010 年)提出了一种对四种旅行模式进行分类的方法——步行、驾车、乘公共汽车和骑自行车——方法是分割 GPS 轨迹,提取每个 GPS 段中的最大速度和加速度等特征,以了解人类的移动性,并将其显示在基于网络的地图应用上(表 1)。该研究提出了一种将 GPS 轨迹分为步行和非步行段的方法,强调步行模式作为向非步行模式过渡的重要性(例如,汽车、公共汽车或火车)。通过分割方法和决策树的使用,该研究达到了超过 75 % 的预测准确率。与 GPS 数据一起, GIS 数据被用来实现旅行模式分类的更高精度。在大多数情况下,已经使用了运输网络数据和相关基础设施(如车站、车站、入口)上的数据( Biljecki、Ledoux 和 Van Oosterom, 2013 年; 龚, 陈, 比亚洛斯托茨基和劳森, 2012 年; Witayangkurn, Horanont, Ono, Sekimoto 和 Shibasaki, 2013 年)。例如, Biljecki 等人。(2013 年)从 OpenStreetMap 中提取道路、铁路、公共汽车和电车网络、公共汽车站和火车站的位置,使用模糊逻辑方法检测 10 种运输模式,该算法的准确率达到 92 %。除了交通联系,莫西耶娃、杰苏伦和蒂莫曼斯(2010 年)提出了一个旅行模式推断系统,利用土地使用数据对七种交通模式进行分类。土地利用数据有助于提高某些类型土地利用(如铁路轨道)上可能出现的特定运输模式的预测精度。使用贝叶斯信念网络模型的传输模式分类达到了 95 % 的准确率。另一方面,一些研究没有使用 GIS 数据,因为这些数据可能无法用于许多研究领域。然而,一些研究能够达到中等至高的预测精度(肖宏亮、王福和吴晓燕, 2017 年; 朱先生等人。 , 2016 年)。

关于性能改进,如何提取特征会显著影响旅行模式分类的结果。在过去十年中,一些研究考虑了每次旅行的子片段的焦点特征,这些特征是通过在 GPS 轨迹上滑动的移动窗口获取的( Bolbol、Cheng、Tsapakis 和 Haworth, 2012 年; 道奇, 威贝尔和福罗奥坦, 2009 年; Van Dijk, 2018 年; 肖等人。 , 2017 年)。Bobol 等人。(2012 年)应用了一个固定大小的移动窗口,在多段 GPS 实例的速度和加速度值上滑动,使用支持向量机( SVM )对六种行驶模式进行分类的准确率达到 88 %。道奇等人。(2009 年)和肖等人。(2017 年)特别采用了 GPS 轨迹的焦点特性,通过计算落在滑动窗口内的 GPS 点的运动参数,分别达到了 82 % 和 91 % 的预测精度。van Dijk (2018)通过引入移动的空间和时间窗口,在旅行(移动)和活动(停留)分类中实现了超过 99 % 的预测准确性。

由于各种传感器的可用性不断提高,多传感器数据的组合使用最近被引入 PA 和旅行模式分类研究。帕特里克等人。(2008 年)开发了物理活动和位置测量系统( PALMS ),通过结合 GPS、加速度计和心率监测传感器,了解暴露生物学研究中与时间和空间相关的 PA 相关能量消耗。关于行驶模式识别,使用智能手机记录的加速度计、磁力计和陀螺仪数据提高了行驶模式检测的性能( Ellis 等人)。 , 2014 年; 冯·蒂莫曼斯, 2013 年; 方等人。 , 2016 年; Shafique & Hato, 2016 年; 周, 2014 年)。埃利斯等人。(2014 年)特别比较了几种机器学习模型对六种旅行模式分类的预测精度,包括公共汽车、汽车、坐着和步行,这些模式基于从两个训练有素的助手收集的不同环境下的 GPS 和加速度计数据。Evenson 和 Furberg (2017 年)为用户和研究人员开发了智能手机应用程序,利用 GPS 和加速度计数据自动预测 PA 类型。

在本研究中,利用其他研究人员收集的两种公开的公开传感器数据——GPS 和加速度计数据,开发了自动 PA 和车内状态分类算法(这些数据将在第 3 节中描述)。久坐是一种重要的固定行为,许多 PA 研究也是如此

**TABLE 1** Characteristic properties of past studies on PA types and transport modes

Author	Predictive accuracy	Classification algorithm or system	Sensor	Classified activities	Observation unit	GIS data use	Moving window
Zheng et al. (2010)	75.60%	Tree-based model	GPS	Walking, biking, car, bus	Segment	No	No
Biljecki et al. (2013)	91.60%	Fuzzy expert system	GPS	Walking, biking, car, bus, train, tram, underground, ferry, sailing boat, aircraft	Segment	Yes	No
Moiseeva et al. (2010)	95.40%	Bayesian belief network	GPS	Walking, running, biking, motorbike, car, bus, train	Segment	Yes	No
Xiao et al. (2017)	90.77%	XGBoost	GPS	Walking, biking, car, bus and taxi, subway, train	Segment	No	Yes
Zhu et al. (2016)	91.44%	Random forest	GPS	Walking, biking, car, bus	Segment	Yes	No
Bolbol et al. (2012)	88.00%	SVM	GPS	Walking, biking, car, bus, train, underground	Segment	No	Yes
Dodge et al. (2009)	82.00%	SVM	GPS	Pedestrian, biking, car, motorbike	Segment	No	Yes
van Dijk (2018)	99.40%	SVM and random forest	GPS	Move, stay	Point	Yes	Yes
Ellis et al. (2014)	91.90%	Random forest	GPS, accelerometer	Walking, biking, car, bus, sitting, standing	Point	No	Yes
Fang et al. (2016)	86.94%	SVM	Accelerometer, magnetometer, gyroscope	Walking, running, biking, vehicle, staying	Segment	No	Yes
Feng and Timmermans (2013)	85%	Bayesian belief network	GPS, accelerometer	Walking, running, biking, motorbike, car, bus, tram, subway	Point	No	Yes
Shafique and Hato (2016)	99.96%	Random forest	Accelerometer	Walking, biking, car, bus, train, subway	Segment	No	Yes
Zhou (2014)	Over 80%	Random forest	GPS, accelerometer	Walking, running, biking, in-vehicle, staying	Point	No	No



除了 MVPA 之外, 还进行了检查。这项研究的贡献在于如何利用两种不同的传感器数据来识别行驶模式。正如道奇等人在研究中采用的。(2009)和肖等人。(2017年), 本研究使用的焦点级别 GPS 轨迹操作器的优点在于, 它们可以帮助捕捉不同行驶模式下运动导数(例如速度、加速度)随时间的不同变化。特别是, 本研究在焦平面广泛使用 GPS 轨迹算子, 具有不同大小的移动空间和时间窗口, 如第 3.2 节所述。与 van Dijk (2018)使用空间和时间移动窗口的研究(如表 1 所示)相比, 该研究通过利用两种不同类型的开放式传感器数据, 提取更多的特征, 以最大限度地利用空间和时间移动窗口对更多旅行模式进行分类。

### 3 | a·弗瑞德利用公众可用的 GPS 和加速度计数据对 PA 和车内状态进行自动分类

本研究中的自动 PA 和车内状态分类是通过分层分类过程实现的, 以共同识别总共六个类别: 骑自行车、跑步、步行、站立、久坐和乘坐车辆(图 1 和 2)。分级分类已经被用来处理现实世界系统中关于从高层到低层分类的分级结构( Dumais & Chen, 2000; 麦克纳马拉, 罗斯利, 罗斯科, 艾伦和戴, 2015 年)。在本研究中, 分层分类方法预计将大大有助于在顶层对室内和室外活动进行准确分类, 并在较低层使用 GPS 和加速度计数据对基于车辆和非基于车辆的旅行模式进行准确分类(图 1)。这是因为 GPS 数据适用于识别基于车辆的运动与非基于车辆的人类运动,

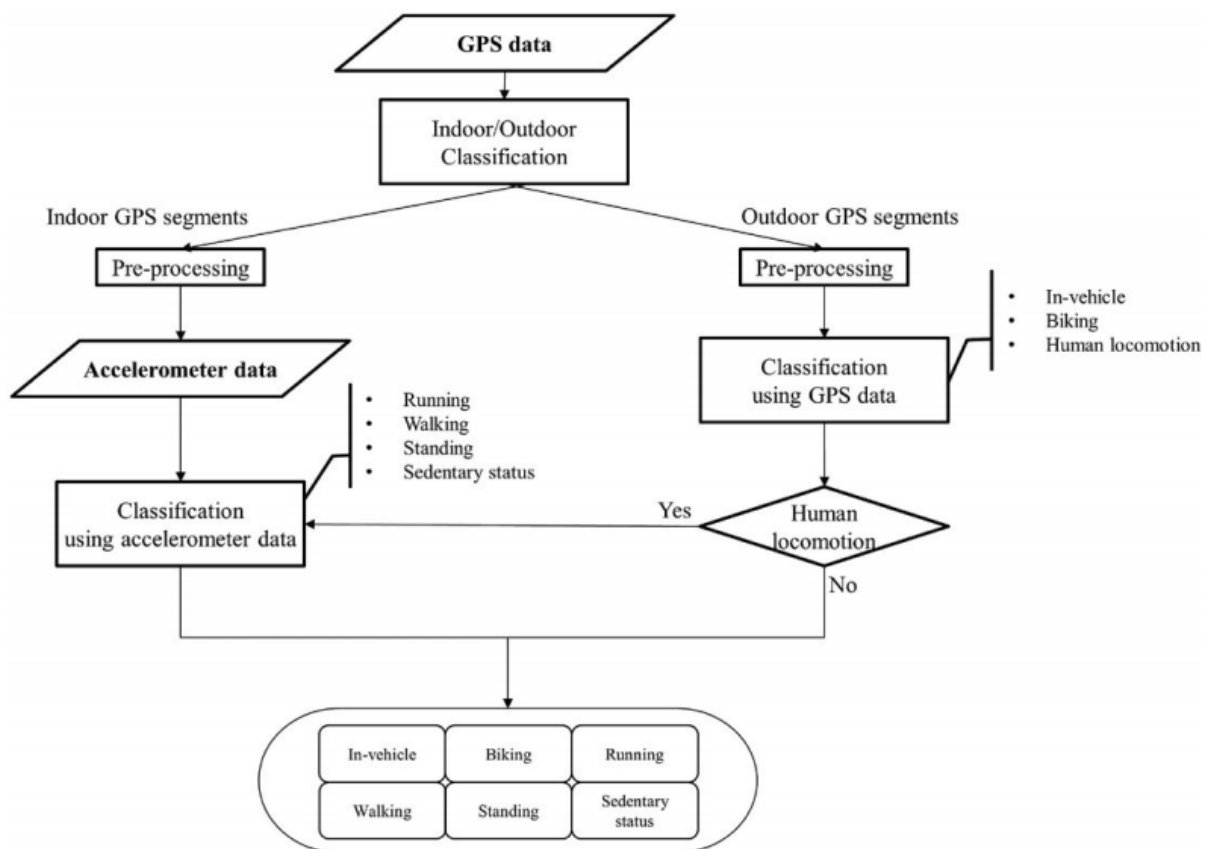


图 PA 和车内分类算法的 1 Flow 图

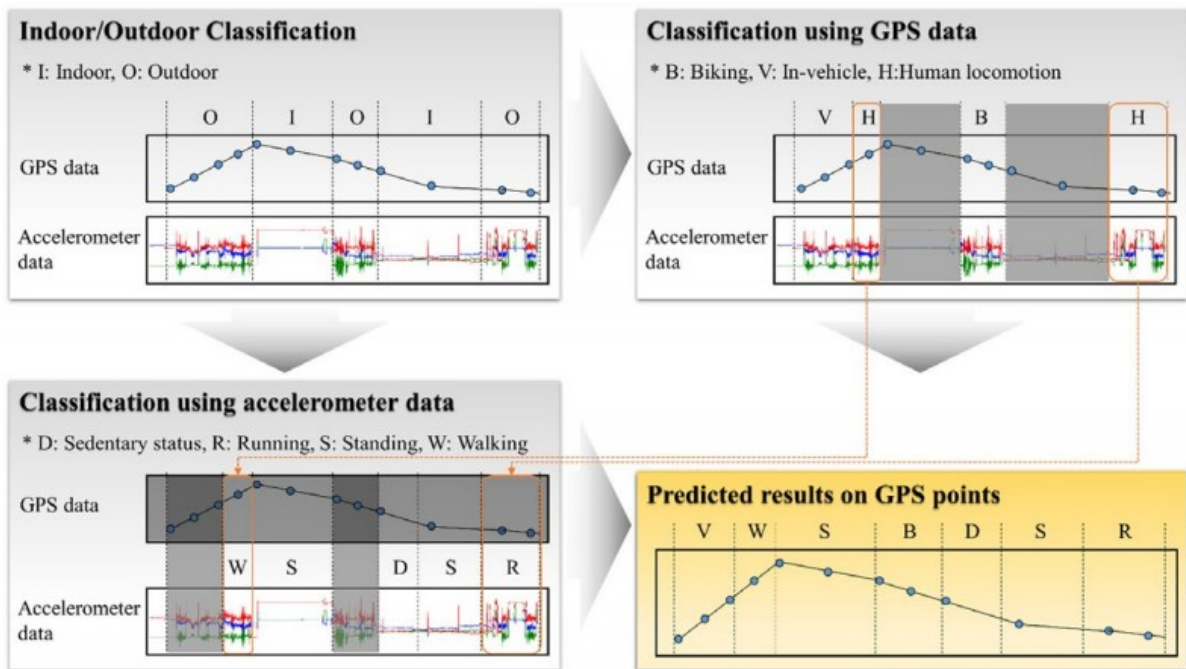


图 3 人类运动的三个分类组件和分级分类的 2 Details

而加速度计数据通常用于分类各种类型的非基于车辆的人体运动。特别地，一些 PA 类型(例如站立、坐着)不能使用 GPS 数据来表征，但是可以使用加速度计数据来识别。因此，分层分类方法允许使用加速度计数据在较低水平的非基于车辆的运动中识别这些 PA 类型(图 1)。除了这两个分类过程之外，GPS 点还在顶层被分类为室内或室外点，以便于在较低层进行准确的分类过程。例如，在室内环境中，像步行和站立这样的基本类型的 PA 可能会发生，而车内状态和骑自行车不太可能在室内发生。因此，室内 GPS 点应该经过一个分类过程，包括加速度计数据的分析。在本研究中，室内 GPS 点和加速度计数据通过“使用加速度计数据分类”过程在较低的级别进行处理，以预测特定的室内 PA 类型。为了对环境是室内还是室外进行分类，我们只使用了从 GPS 数据中获得的测量值，这些测量值不附带额外的传感器数据(例如，摄像机图像、光传感器级别)，以提高先前研究中实施的方法的性能(Lam 等人)，2013 年；坦登，萨伦斯，周，克尔和克里斯塔克斯，2013 年)。

在图 1 和图 2 所示的三级分类过程中，“室内/室外分类”过程首先将 GPS 轨迹划分为室内或室外段，因为室内 GPS 段由于室内环境中 GPS 信号的低精度和丢失，对于提取任何运动导数来说是不有效和不可靠的。然后，在此过程中识别的室外 GPS 段将经历“使用 GPS 数据分类”过程，该过程将 GPS 段分为基于车辆的运动(乘坐车辆、骑自行车)和非基于车辆的人类运动(人类运动)。使用 GPS 数据处理的室内/室外分类和分类应用了 GPS 轨迹算子的概念(Laube 等人)，2007 年)，如第 3.3 节所述。另一方面，室内 GPS 段不用于提取任何分类过程的特征，而是用于提供时间范围，在该时间范围内，通过“使用加速度计数据分类”预测特定的室内 PA 类型，如久坐状态和行走( Lee & Kwan, 2017 )。预处理之后，使用 GPS 或加速度计数据进行分类，以过滤出低质量的 GPS 点。在“使用 GPS 数据进行分类”过程中确定的三个类别中，PA 被进一步分类为特定 PA 类型，如运行

并逐步完成“使用加速度计数据进行分类”过程。通过这三级分类过程，识别出六个不同的 PA 类别:乘坐汽车旅行、骑自行车、跑步、步行、站立和



久坐。

### 3.1 | data 描述

来自四个不同来源的数据集用于三级分类过程(见表 2)。首先,由于没有公开可用的 GPS 数据可用于对室内和室外轨迹进行分类, GPS 数据是从居住在高度或中度城市化地区的三个人那里收集的,为期七天,带有水平精度稀释(HDOP)和“室内”或“室外”标签。精度稀释表示位置测量的精度,具体来说, HDOP 表示 GPS 数据水平位置的几何质量。高度城市化和中度城市化地区之间 GPS 精度的巨大差异可能会影响室内/室外识别。因此, GPS 和加速度计数据是从一个生活在高度城市化地区的受试者和两个生活在中度城市化地区的受试者收集的。这些数据用于第 4.1 节中报告的测试阶段。在第 3.3 节的室内/室外分类中,仅使用了来自三个主题中每一个主题的两天 GPS 轨迹,这些轨迹有许多室内和室外旅行的组合。

第二,地球生命项目 GPS 数据集(微软亚洲研究院)的 1.3 版用于使用 GPS 数据进行分类。GPS 数据集从 2007 年 4 月至 2012 年 8 月收集了 182 名受试者(郑、李、陈、谢、马, 2008 年;郑等人。 , 2010 年)。这些 GPS 轨迹大多以短时间间隔(例如 1 至 5s)被跟踪。来自 73 名参与者的 GPS 数据中,只有一部分有 11 种交通方式的标签。其中,研究中使用了标有七种交通方式的 GPS 轨迹——火车、公共汽车、汽车、出租车、自行车、步行和跑步。带有这七个运输模式标签的 GPS 点总数为 5, 063, 475 个,步行(35 %)占 GPS 点的很大比例,而跑步(0.03 %)占最小比例

部分。

第三,由于缺乏跑步模式的 GPS 数据,人们跑步时记录的开放式 GPS 数据是从 OpenStreetMap (OSM)和 TrackProfiler 获得的,并用于本研究。列车、出租车、公共汽车和汽车运输模式被合并到一个机动模式中,步行和跑步(来自 Geolife、OSM 和 TrackProfiler 的 GPS 数据)在研究中被归入一个人类运动类。

**TABLE 2** Description of GPS and accelerometer datasets

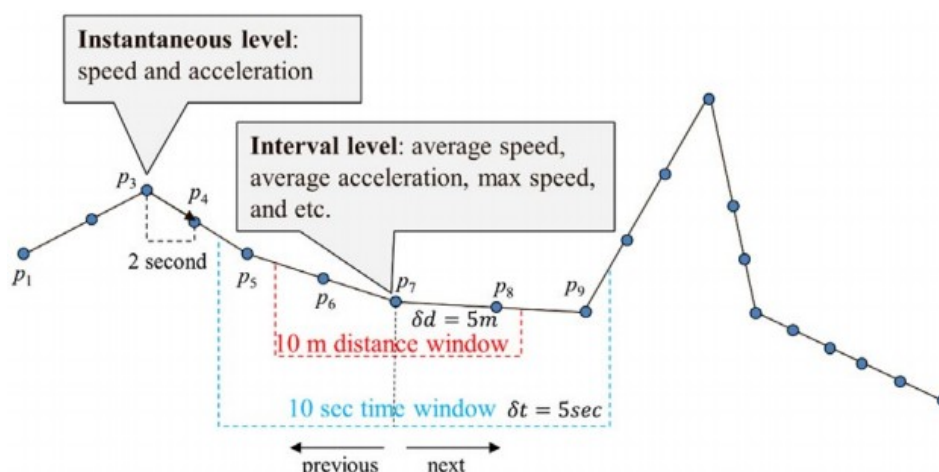
Dataset	No. of instances	Attributes	Mode	Sampling rate
GPS dataset collected from two subjects	58,172	Person ID, longitude, latitude, timestamp, HDOP, indoor/outdoor mode	Indoor (61%), outdoor (39%)	1 s
Geolife GPS dataset v1.3 (Microsoft Research Asia)	5,063,475	Person ID, longitude, latitude, timestamp, transport mode	Train (5%), bus (17%), car (8%), taxi (7%), bike (28%), walk (35%), run (0.03%)	Mostly 1–5 s
GPS dataset from OSM and TrackProfiler	10,435	Longitude, latitude, timestamp	Run (100%)	Variable, 1–17 s
Wireless sensor data mining (WISDM)'s accelerometer dataset v1.1	1,098,207	Person ID, physical activity mode, timestamp, acceleration (x, y, z)	Walking (39%), running (31%), upstairs (11%), downstairs (9%), sitting (6%), standing (4%)	20 Hz (20 samples/s)

最后,无线传感器数据挖掘( WISDM )加速数据集的 1.1 版用于使用加速计数据处理进行分类 ( Kwapisz 等人。 , 2011 年)。在实验室条件下,从 36 名有时间戳的人身上收集三轴加速度,并以 20Hz 记录。每个实例(观察)都有一个 PA 标签(例如行走、跑步或站立)。由于在楼梯上的运动只占受试者每日 PA 的一小部分,因此在这项研究中,楼上和楼下的运动合并为步行。

### 3.2 | gps 轨迹算符

GPS 轨迹算符( Laube 等人。 , 2007 年)用于在室内/室外分类和使用 GPS 数据处理的分类中从 GPS 轨迹中提取特征。通过引入地图代数运算来捕捉空间和时间上的动态运动特征,提出了不同层次的 GPS 轨迹算子。操作员执行特定的数学和/或逻辑分析或计算。例如,在地图代数中,有四种类型的操作——局部的、焦点的、带状的和全局的——用于空间分析,以使用栅格数据生成最终地图( Tomlin, 1990 )。在四种不同级别的 GPS 操作符(瞬时、间隔、节间和全局)中,瞬时和间隔操作符用于获取动态运动的局部和焦点特征,并计算用于特征提取的运动导数(速度、加速度)。在机器学习中,特征提取是从数据中导出值(特征)以训练模型的过程。计算两个连续 GPS 点之间的瞬时水平上的运动导数,以捕获与这些 GPS 点中的每一个相关的局部运动特征。例如,假设给定的 GPS 轨迹以 2s 间隔记录(如图 3 所示),则基于  $p_3$  和  $p_4$  之间的距离和 2s 时间间隔(  $t$  ) 计算 GPS 点  $p_3$  的瞬时速度。

与瞬时运动导数相比,在考虑空间和时间维度的情况下,计算区间水平的运动导数。考虑到距离或时间窗范围内的所有 GPS 点,距离或时间窗沿着给定 GPS 轨迹移动,以计算区间移动导数,包括平均速度和最大加速度。考虑到交通状况,计算出的区间导数特别反映了不同交通模式之间运动特征的可区分的时间和空间变化。例如,假设图 3 中给定 GPS 轨迹的运输模式是平稳交通流中的“汽车”,基于  $p_6$ 、 $p_7$  和  $p_8$  的速度值计算 10m 距离窗口内 GPS 点  $p_7$  的平均速度,而基于  $p_5$ 、 $p_6$ 、 $p_7$ 、 $p_8$  和  $p_9$  的速度值计算 10s 时间窗口内相同 GPS 点  $p_7$  的平均速度。然而,如果  $p_7$  附近交通繁忙,10m 距离内考虑的 GPS 点的数量



图在瞬时和区间水平上计算的 3 Movement 导数

由于  $p_7$  附近有许多静止的 GPS 点,所以窗口将超过三个点,尽管窗口的大小仍然与平滑交通流中的窗口大小相同。另一方面,10 秒的时间窗口可能会缩小,因为在如此繁忙的交通中  $p_7$  附近有许多静止的

GPS 点, 这可能导致平均速度急剧下降。通过这种方式, 使用距离和时间窗的间隔移动导数可以以互补的方式捕获每个传输模式的不同焦点特性, 以提高性能。

### 3.3 | indoor/outdoor 分类法

由于室内 GPS 数据的用处有限, 需要将室外 GPS 点与室内 GPS 点分开。极端梯度提升(Chen & Guestrin, 2016 年), 遵循梯度提升的原则(Friedman, 2001 年), 用于根据从三个受试者收集的 GPS 数据中提取的三个选定特征对这两个标签(室内和室外)进行分类。Boosting 是一种前向阶段优化方法, 它使用每次迭代中学习到每个弱分类器的投票来生成强分类器。梯度增强使用树模型作为弱分类器, 并基于梯度的概念生成强分类器, 以使损失函数最小化。极端梯度增强尤其将正则化纳入交流计数, 以控制过度拟合, 从而提高性能。在特征提取中, 速度、加速度、HDOP 和缺失点的数量都被考虑在内, 因为 GPS 点可能有相当多的误差, 表现为位置和速度峰值, 并且可能由于信号丢失而没有被正确记录。考虑到室内/室外分类中 GPS 轨迹操作者的瞬时和间隔级别, 在所有 31 个特征中, 仅选择了三个重要特征 180 s 间隔级别的平均速度和 HDOP, 以及 120 s 间隔级别的缺失点数量(表 A, 附录)。

### 室内和室外 GPS 点的 3.4 | pre processing

GPS 数据的预处理包括确定有效的室内和室外 GPS 点。对于已识别的室内 GPS 点, 高 HDOP 的经度和纬度值(例如, 本研究中  $HDOP > 8$ )将被低 HDOP 的先前 GPS 点所取代, 因此许多任意分散的 GPS 尖峰可以在先前跟踪的位置上移动。由于信号丢失而丢失的室内 GPS 点将被先前带有室内标签的 GPS 点的位置所取代。关于识别出的室外 GPS 点, 高 HDOP 的 GPS 点被使用卡尔曼滤波器估计的那些点所取代, 以滤除噪声记录并产生精确的 GPS 点。卡尔曼滤波器的应用在第 4.1 节中描述。

### 3.5 | classification 使用全球定位系统数据

由于 Geolife GPS 数据针对每个人和旅行模式被划分成许多段, 因此需要将两次或两次以上时间间隔小于 1 分钟的连续旅行合并为一段, 作为特征提取的分析单元。合并过程考虑了两种不同模式之间的转换, 如步行和乘坐公共汽车, 以改进对交通模式的识别。例如, 2008 年 3 月 4 日和 2011 年 8 月 27 日, 人员 1 和人员 2 以不同的交通方式出行, 不同的出行被分成几个部分, 如表 3 所示。在这种情况下, 因为 Person 1 从第一行到第三行的前三段发生了

连续地, 它们被连接成一个片段。此外, 由于使用 GPS 数据进行分类的一个重要目标是准确地将车辆在车辆中行驶(车内模式)和骑自行车与 PA (类似于跑步)分开, 以便在下一阶段使用加速度计数据进行更好的分类, 因此步行/跑步和其他运输模式的连续段被合并到一个串联段中, 用于特征提取。

此外, 使用以下标准选择有效的 GPS 段: (1) 由平均记录时间间隔小于 10 秒的 GPS 点组成的段; 和 (2) 记录时间总和大于 3 分钟的片段。以较长时间间隔记录的 GPS 点无法确保特征的一致性, 因此

**TABLE 3** Examples of segments in Geolife GPS data

Row ID	Person ID	Start time	End time	Transport mode
1	1	04/03/2008 11:32:24	04/03/2008 11:46:14	walk
2	1	04/03/2008 11:47:14	04/03/2008 11:55:07	taxi
3	1	04/03/2008 11:55:24	04/03/2008 12:01:49	taxi
4	1	04/03/2008 16:00:00	04/04/2008 04:13:22	train

.....

2011 年 8 月 27 日 101 2 08 : 06 : 13 : 01 08 / 27 2011 年 8 : 01 : 37 步行

102 2 08 / 27 / 2011 15 : 01 : 59 08 / 27 / 2011 15 : 31 : 43 步行

注意。粗体字体表示连续的行程，并连接成一个段。

不包括具有 GPS 点的具有如此长时间间隔的段。此外，由于使用最大时间窗为 3 分钟的间隔级运算符，记录时间小于 3 分钟的片段被排除在外。

在瞬时和间隔水平上，每个 GPS 点总共生成 73 个特征，如表 4 所示。在瞬时水平上计算两个运动导数，速度和加速度，以及一组五个运动导数——平均速度、平均加速度、最大速度、最大加速度和速度变化率( Zheng 等人)。， 2010 年)——计算间隔级别的每个时间窗( 10 秒、20 秒等)和距离窗( 10m、20m 等)。除了瞬时导数和间隔导数之外，每个 GPS 点的记录时间也被提取出来并作为一个特征使用，因为它在分离时间受限的活动中发挥着重要作用。例如，人们很可能在白天骑自行车和跑步，而像公共汽车和汽车这样的机动交通工具通常在高峰时段使用。

随机森林( Breiman, 2001 年)用于基于从 Geolife GPS 数据集中提取的特征来识别基于车辆的运动(乘坐车辆和骑自行车)和非基于车辆的人类运动(人类运动)。衡量每个特征的平均基尼系数下降(表 B, 附录)，以检查重要特征。最重要的运动导数是区间水平上的平均速度和最大速度，而瞬时水平上没有运动导数显示出很高的重要性。

### 3.6 | classification 使用加速度计数据

基于 Lee 和 Kwan ( 2017 )开发的方法，使用加速度计数据在分类中识别跑步、行走、站立和久坐状态，当使用随机森林时，该方法具有 99.03 %的预测精度。室内 GPS 点的时间范围描绘了使用加速度计数据识别这种特定 PA 类型的边界。使用 GPS 进行分类时识别出的人类运动标签

数据也通过使用加速计数据的分类组件来生成特定的 PA 类型。

## 4 | results

### PA 的 4.1 | Performance 和基于真实数据的车内状态分类算法

PA 和车载分类算法是使用 R 统计计算软件实现的( R 统计计算基金会，2013 年)。R 支持名为 “Xgboost”( Chen & He, 2018 )和 “RandomForest”( Liaw & Wiener, 2002 )的机器学习模型包，分别用于室内/室外分类和使用 GPS 数据的分类。由于间隔级别的时间和距离窗口需要在使用 GPS 数据处理的室内/室外分类和分类中进行密集计算，因此使用了通过利用多核部署并行计算的 “doParallel”包

来提高计算性能( Calaway、Weston 和 Tenenbaum, 2014 年)。R 还用于从 GPS 和加速度计数据中提取所有特征。

TABLE 4 All 73 features for classification using GPS data (bold font: high importance)

GPS trajectory operator	Movement derivative
-------------------------	---------------------

瞬间的

速度

加速

记录时间(小时)

间隔时间窗

10 秒

平均速度、平均加速度、最大速度、最大加速度、速度变化率

20 秒平均速度、平均加速度、最大值

速度、最大加速度、速度变化率

30 秒平均速度，平均加速度，最大值

速度、最大加速度、速度变化率

60s 平均速度、平均加速度、最大值

速度、最大加速度、速度变化率

90s 平均速度、平均加速度、最大值

速度、最大加速度、速度变化率

120 秒平均速度，平均加速度，最大值

速度、最大加速度、速度变化率

180 s 平均速度、平均加速度、最大值

速度、最大加速度、速度变化率



距离窗

10m 平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

20m 平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

30m 平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

40m 平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

50m 平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

100 米平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

200 米平均速度, 平均加速度, 最大值

速度、最大加速度、速度变化率

使用 10 倍交叉验证评估了在室内/室外分类和使用 GPS 数据处理的分类中生成的两个学习模型的性能。对于使用 GPS 数据的分类, 考虑到每个标签的原始比例, 只有 218, 342 个实例被随机采样并用于交叉验证, 这是因为输入所有实例需要相当长的计算时间。使用极端梯度提升的室内/室外分类模型显示了总共 500 次迭代的 99.56 % 的预测精度, 室内和室外 GPS 点都被正确分类, 准确率超过 99 %, 如表 5 所示。

对于使用 GPS 数据的分类, 随机森林以 500 棵树达到了 94.47 % 的预测精度。表 6 中的混淆矩阵表明, 人类运动和车内类别的识别准确率高达 90 % 以上。骑自行车的分类准确率最低( 77.90 % ), 尤其是 1, 833 和 1, 061

带有自行车标签的实例分别被错误地预测为人类运动和车内运动。在这三个类别中, 人类运动显示出最高的预测精度( 96.57 % ), 这反过来又有望提供更准确的分类人类运动实例, 在使用加速度计数据的分类中被进一步识别为四种 PA 类型之一。

为了在从三个地方收集的真实数据、GPS 和加速度计数据上验证所提出的算法

测试了自由生活条件下的受试者。这三个科目都是男性、年轻人以及没有健康问题的本科生或研究生。大学生是合适的科目，他们可以利用城市地区的各种机会。特别是，这三名受试者在日常生活中经常通过在校园里散步、锻炼或者在他们的居住区跑/骑自行车来参与 MVPA。他们被给予智能手机，或者使用他们自己的智能手机在 7 天、8 天或 28 天内记录 GPS 轨迹和加速度计数据，如表 7 所示。受试者 1 特别参与了为期近 1 个月的数据收集，通过可视化探索更多具有预测结果的活动。这三名受试者都被要求收集 GPS 和加速度计数据，将手机放在裤子口袋里。然而，受试者 3 错误地将所提供的智能手机放在了他的一个夹克口袋里，这导致了使用加速度计数据进行分类时的预测准确性差，因此从受试者 3 收集的加速度计数据没有用于本研究。此外，来自受试者 2 的加速度计数据最初用廉价手机( LG 领域)收集了一周，都是错误的。所有三个坐标轴的值都有变化，预测结果大多是坐着的

**TABLE 5** Confusion matrix of indoor/outdoor classification using extreme gradient boosting

		<i>Actual class</i>	
		Indoor	Outdoor
Predicted class	Indoor	35,140	106
	Outdoor	147	22,740
Accuracy (%)		99.58	99.54

**TABLE 6** Confusion matrix of classification component using GPS data

		<i>Actual class</i>		
		Human locomotion	In-vehicle	Biking
Predicted class	Human locomotion	89,593	5,755	1,833
	In-vehicle	2,949	106,466	1,061
	Biking	235	250	10,200
Accuracy (%)		96.57	94.66	77.90

自由生活条件下从三名受试者收集的 GPS 和加速度计数据的表 7 Description

主题 ID 设备数据记录时间

1 LG G3、三星 Galaxy Alpha GPS 和加速度计 28 天

2 三星 Galaxy Alpha GPS 和加速度计 8 天

3 LG 领域全球定位系统 7 天

即使受试者走路时的状态。因此，受试者 2 被要求使用三星 Galaxy Alpha 再次收集数据，这导致了更好的分类结果。利用 HDOP 以 1 秒的间隔记录 GPS 轨迹。此外，还收集了详细记录受试者交通方式和 PA 类型的活动日志。由于所开发的算法需要基于受试者的活动、交通方式和 PA 类型的非常详细的记录进行

评估, 因此要求每个受试者通过检查可视化 GPS 轨迹来详细说明他每秒的活动日志。详细的 3 天活动日志、GPS 轨迹和加速度计数据随后被用作验证算法的测试数据集。

由于一些 GPS 点由于其低精度和固有的位置误差而被排除在外, 因此卡尔曼滤波器被用来精确地估计无效 GPS 点的纬度和经度值。卡尔曼滤波器是应用最广泛、性能最好的平滑方法之一, 用于减少影响 GPS 点衍生测量精度的 GPS 随机误差( Jun 等人)。, 2006 年; 格雷瓦尔等人。 , 2011 年)。卡尔曼滤波 R 中的函数 “DLMs - mooth” 在本研究中用于估计缺失或排除的 GPS 点, 并平滑 GPS 轨迹 ( Petris, 2009 年)。关于卡尔曼滤波器, Petris ( 2009 ) 的简单动态线性模型所建议的参数值被应用于默认[表 C(c) ], 附录中观测噪声和系统噪声的方差。[表 C(b) 、 ( d ) , 附录]也测试了较高的系统噪声方差, 以观察弱(系统噪声方差= 1.0 )或强(系统噪声方差= 0.01 )滤波器的影响, 这使得测试 GPS 轨迹或多或少偏离原始记录(图 4 )。系统噪声方差为 0.1 的卡尔曼滤波具有最佳预测精度, 因此被应用于测试 GPS 轨迹, 以进一步评估 PA 和车载状态分类算法。

该算法在现实世界 GPS 轨迹上的性能在识别六种 PA 类型时达到了 96.20 % 的精度(表 8 )。预测准确率最低的 PA 类型正在运行( 69.98 % ); 在使用 GPS 数据的分类中, 189 个 GPS 点被错误地归类为自行车。步行和站立是结果中最准确的两种类型(分别为 98.25 % 和 97.83 % )。还计算了该算法的灵敏度、特异性、阳性预测值和阴性预测值(见表 9 )。

#### 4.2 | predicted pa 类型和车载状态及其与 GPS 轨迹的可视化

为了定量验证 PA 和车内状态分类算法, 基于加速度计数据和 GPS 轨迹中的时间戳, 将开发的算法预测的 PA 类型和车内状态与收集的 GPS 点相结合(见图 5 )。使用 ArcGIS 10.4 在地图上可视化 GPS 轨迹和预测 PA 类型, 以评估连续出行的预测结果及其关系

具有周围地理环境的船只。

汽车、公共汽车和自行车大多显示出正确的分类。尽管在道路交叉口和公共汽车站附近有速度较慢的等待点, 但是对汽车和公共汽车行驶模式的预测显示出了有希望的结果。然而, 与中等城市化地区的公共汽车相比, 城市地区的一些公共汽车出行轨迹被错误地归类为自行车。自行车还显示了一些错误的车内分类结果, 这可以通过第 4.1 节中的适度预测精度来解释。步行、久坐状态(坐着)和站立也通过使用加速度计数据的分类, 被随机森林准确地分类。从一名受试者收集短期(例如 10 分钟)跑步数据, 跑步过程中的一些 GPS 点为

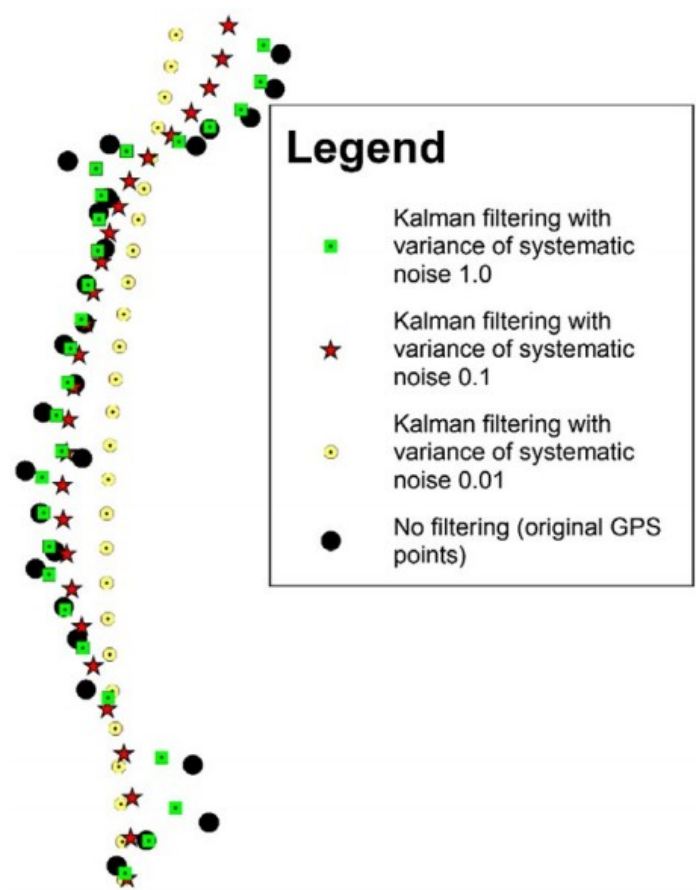


图 GPS 轨迹上的 4 Testing 卡尔曼滤波，系统噪声有三个不同的方差值

**TABLE 8** Confusion matrix of PA and in-vehicle status classification on free-living condition GPS and accelerometer dataset

		Actual class					
		Running	Walking	Sitting	Standing	In-vehicle	Biking
Predicted class	Running	464	19	0	0	0	0
	Walking	3	6,725	1,363	8	27	0
	Sitting	0	80	65,889	0	101	2
	Standing	0	0	1,019	2,889	17	0

车内

骑自行车

7

189

19

2

47

0

56

0

1, 662

0

175

1, 779

准确度( %) 69.98 98.25 96.44 97.83 91.98 90.95

错误地归类为骑自行车。一次地面地铁出行被检测为车内出行，该算法没有考虑车内类别。还检查了一系列日常活动，如图 5 右下角所示。一名受试者开车去了一个娱乐场所，对这个人来说，步行、坐着和站立都是室内和室外娱乐场所的代表。

TABLE 9 Sensitivity, specificity, positive predictive value, and negative predictive value for PA and in-vehicle status classification algorithm

Measures	Running	Walking	Sitting	Standing	In-vehicle	Biking
Sensitivity (%)	69.98	98.25	96.44	97.83	91.98	90.95
Specificity (%)	99.98	98.15	98.71	98.70	99.62	99.76
Positive predictive value (%)	96.07	82.76	99.72	73.61	84.54	90.31

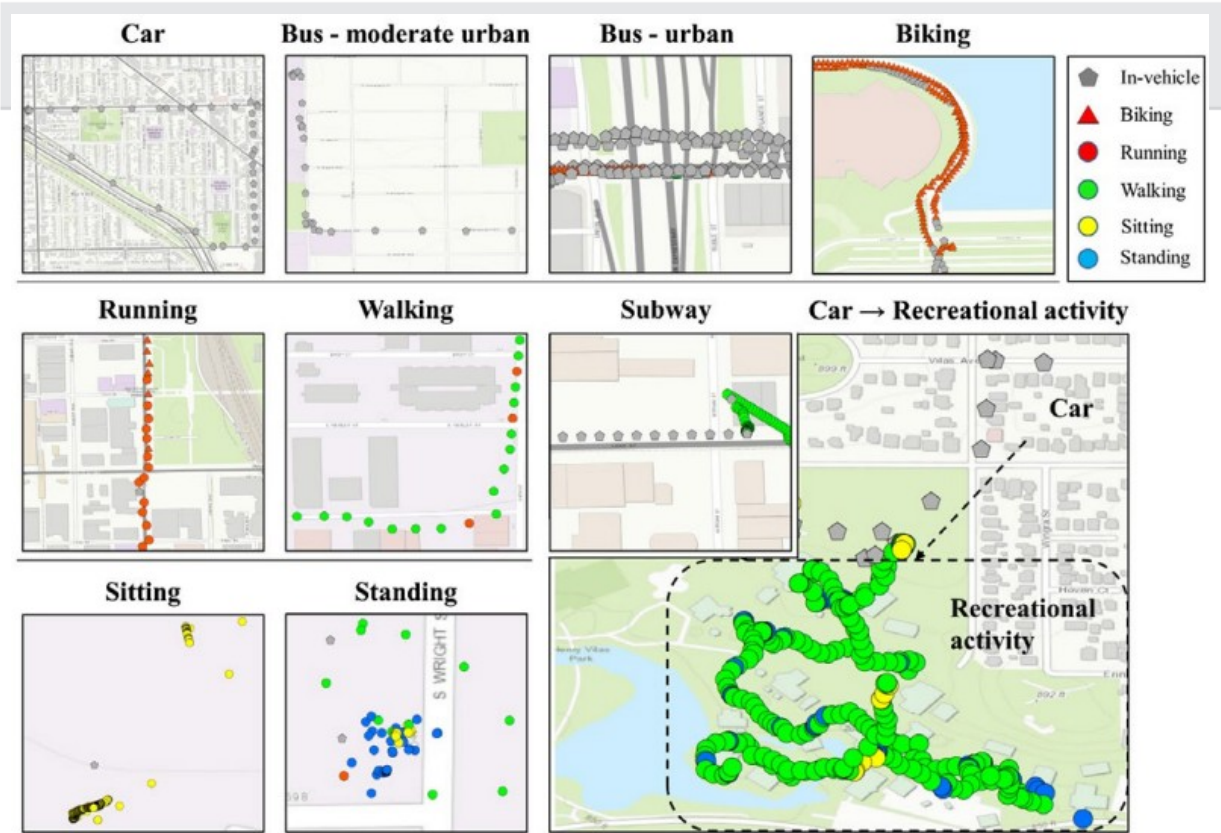




图5 Visualization 预测了自由生活条件下的六种分类 PA 类型和车内状态

## 5 | discussion 和结论

在本研究中,利用公开的 GPS 和加速度计数据开发的分级分类算法显示出优异的性能,能够根据真实世界的 GPS 和加速度计数据准确预测车内状态和 PA 类型。公开提供的 GPS 和加速度计数据显示有潜力自动准确地对人们的车内状态和 PA 类型进行分类,以便使用客观识别的 PA 进行研究。结果还表明,在没有任何分割方法的情况下,车载状态、自行车和 PA 可以用 GPS 轨迹算子成功地分类。根据我们的简单探索,还发现表 6 中自行车分类的相对较低精度是由于一些 GPS 点偏离了它们

最初的轨迹,可能是由于位置错误和实际骑车模式造成的,在商业区,甚至可能是目的地附近,尽管我们在这项研究中没有具体分析这些轨迹。

在使用 GPS 数据进行分类时,瞬时水平特征都不重要。这表明,基于不同大小的时间和距离窗口计算的间隔级运动描述符

优于从两个连续 GPS 点计算出的瞬时水平运动描述符,这在许多交通模式分类研究中被广泛使用。此外,表 4 中的 20 个重要特征表明,提取特征不仅要考虑时间维度,还要考虑空间维度。通过时间窗和距离窗计算出的平均速度和最大速度在更好地预测车内状态、骑自行车和人类运动方面发挥了重要作用。

使用异构传感器数据集对车内状态和 PA 进行分级处理是一种新颖的方法,在广泛的领域有着巨大的应用潜力。例如,Prelicean、Gidyfalvi 和 Susilo (2017 年)强调,关于旅行模式检测的跨学科解决方案不应局限于一个研究领域。他们还强调,当前验证新算法和使用无法广泛共享的数据集的研究趋势阻碍了跨学科研究。在这方面,这项研究为交通和 PA 研究中的研究人员提供了指导,通过使用公开的异构传感器数据集的训练模型,以可扩展的方式设计旅行模式分类算法。此外,我们将通过网页向感兴趣的研究人员或从业人员提供该算法。同样重要的是,预测的 PA 类型和带有 GPS 轨迹的车内状态可能有助于预测其他活动相关或上下文信息,包括活动本身和活动目的,这对解决 UGCoP 非常有用。例如,公交车站周围受试者的 GPS 点,加上预测的“站立”或“坐着”,可以被解释为等待公交车的时间。如果耦合的 GPS 轨道沿着特定的环境,具有特定的目的,例如娱乐设施,那么预测的“步行”标签也可以通过系统假设和逻辑推理,按目的进一步分离。因此,从预测结果中推断出的信息有助于推进 PA 和运输研究。此外,本研究还深入探讨了 PA 和运输模式分类如何应用于涉及人员流动分析的各种研究领域(例如,空气质量;Tanio 等人.,2016 年),特别是关于算法如何在自由生活条件下收集的传感器数据集上工作,以及预测结果如何表示。

在本研究中,使用基于分层分类过程的算法对从三个受试者收集的 GPS 和加速度计数据进行分类。分类结果被可视化以增强受试者对日常活动和旅行的了解。该算法以一系列可信的车内状态和 PA 类型高精度地对所有活动进行分类:在中度城市化地区乘坐汽车或公共汽车,骑自行车和坐在建筑物中,以及站在公共汽车站周围。特别是,在使用 GPS 数据的分类中,GPS 轨迹操作者的区间水平有助于捕捉不同时间和距离窗口内汽车和公共汽车行驶的特征。这有助于正确分类速度较慢的等待点

道路交叉口和公共汽车站。此外,娱乐活动中不受控制的运动被合理地解释为不同姿势的组合,如行走、站立和坐着,这些都是娱乐活动的特征。更重要的是,该算法能够以高预测精度检测室内 PA。

然而,只有当收集的传感器数据质量可靠时,所开发的算法才能实现高性能。内置加速度计传感器质量低,可能会记录不稳定的加速度值,这可能不利于计算精确 PA 分类的特征( Lee & Kwan, 2017 )。智能手机的放置也会极大地影响使用加速度计数据进行分类的结果。因此,智能手机筛选程序和智能手机放置指导对于确保算法的性能是必要的。此外,用于训练模型的可公开访问的数据都不包括任何个人信息,如人口信息或社会经济状况,这表明算法最适合的特定人群子群。个人层面的特征不仅可以解释年龄、性别和种族方面的个体内部差异,还可以解释日常 PA 模式、健康状况和交通方式方面的个体内部差异,这反过来又共同构成了

训练数据。然而,由于地球生命全球定位系统数据集、OSM 全球定位系统数据集和 WISDM 加速度计数据不提供任何个体或群体特征,因此该算法能否对其他群体实现相同的性能还不确定。

本研究中开发的分类算法有一些局限性,需要在未来的研究中解决。首先,应该进行额外的处理,以尽量减少不正确分类的 PA 类型。举例来说,城市地区的公共汽车出行分类结果不一致,而中等城市化地区的公共汽车出行则被完全分类。不正确的分类在很大程度上是由于 GPS 点的低定位精度造成的。此外,一些错误预测的 PA 类型也在旅行中间歇出现。在这种情况下,需要进行后处理,用正确的 PA 标签替换错误的 PA 标签。

此外,分类算法的设计需要改进,以获得更高的预测精度。分级分类将错误向下传播到其他分类过程中。例如,在使用加速度计数据的 PA 预测中,在使用 GPS 数据的分类中,一些跑步点先前被误分类为自行车;因此,使用加速度计数据的 PA 分类不能干预预测,也不能将自行车标签更改为跑步标签,即使使用加速度计数据的 PA 分类正确地对其进行了分类。因此,未来的研究应该解决如何有效利用分级分类过程的预测结果。

最后,机动化运输模式应该进一步分类,以考虑不同的机动化模式。这项研究将火车、公共汽车、汽车和出租车合并成单一的机动运输模式,因为机动运输的细分在 PA 的强度和类型方面没有任何意义。然而,对于交通或社会科学等其他研究领域来说,电动交通方式的区分,无论是公共的还是私人的,都是很重要的( Biljecki 等人)。, 2013 年; 冯·蒂莫曼斯, 2013 年; Shafique & Hato, 2016 年)。因此,本研究中的单一电动运输模式需要在未来的这类研究中进行细分和指定。

## ORCID

康在利·[http://orcid.org/0000\\_0002\\_2857\\_6496](http://orcid.org/0000_0002_2857_6496)·关美波·[http://orcid.org/0000\\_0001\\_8602\\_9258](http://orcid.org/0000_0001_8602_9258)

## 参考

马萨诸塞州阿拉木图市、马萨诸塞州杰勒特市、新泽西州敦顿市、新泽西州濠户市和马萨诸塞州彭茨市 ( 2012 年)。使用卫星、GPS 和加速度计数据对社区设计、绿色和儿童体育活动进行研究。健康与地点, 18, 46 - 54。安吉塔, 吉奥, 阿, 奥内托, 帕拉, X, & 雷耶斯-奥尔蒂斯, J. L. ( 2012 )。智能手机上的人类活动识别

多类硬件友好支持向量机。《第四次环境辅助生活国际研讨会记录》(第 216 - 223 页)。西班牙维多利亚-加斯泰兹。

麻省阿里夫, 比尔, 麻省, 卡特坦, 麻省, 和阿哈迈德, 麻省理工学院( 2014 )。使用智能手机加速度传感器进行更好的身体活动分类。医疗系统杂志, 8, 95。

李杜, H. & 范奥斯特罗姆, F. Biljecki ( 2013 )。基于运输模式的移动轨迹分割和分类。国际地理信息科学杂志, 27, 385 - 407。博尔特, A, 程, T, Tsapakis, I, & Haworth, J. ( 2012 )。使用移动窗口 SVM 分类从稀疏 GPS 数据推断混合运输模式。计算机、环境和城市系统, 36, 526 - 537。博鲁夫, 新泽西州, 内森, 新泽西州和新泽西州( 2012 )。使用 GPS 技术(重新)检查基于地方的健康研究中“邻近”的操作定义。国际健康地理杂志, 11, 22。布里曼( 2001 )。随机森林。机器学习, 45, 5 - 32。

拉削, J, 迪尔, J, & 格利贝, J. ( 2012 )。骑自行车的人在哪? 利用显示的偏好 GPS 数据开发路线选择模型。交通研究 A 部分:政策与实践, 46, 1730 - 1740。

布朗宁( 2017 )。“绿色”最能预测身体健康的距离是多少? 系统综述

有 GIS 缓冲区分析的文章。国际环境研究与公共卫生杂志, 14, 675。

南卡罗来纳州韦斯顿市加来威市&特纳鲍姆市( 2014 年)。并行操作:为并行包创建每个并行适配器。从 <https://rdrr.io/cran/doParallel/>取回

陈, T, & 盖斯林, C. ( 2016 )。Xgboost :一个可扩展的树提升系统。《第 22 届 ACM SIGKDD 知识发现和数据挖掘国际会议记录》(第 785 - 794 页)。加利福尼亚州旧金山:加勒比海协会。陈东升和何东升( 2018 )。极端梯度提升。从 [https://cran.r-project.org/web/packages/](https://cran.r-project.org/web/packages/xgboost/) xgboost /渐晕/xgboost.pdf 中检索

纽约州库珀市, 纽约州佩奇市, 华盛顿州惠勒市, 纽约州格里夫市, 纽约州戴维斯市, 马萨诸塞州希尔斯登市和纽约州贾戈市( 2010 年)。绘制步行去学校的地图

使用加速度计结合全球定位系统。美国预防医学杂志, 38, 178 - 183。高层建筑和城市生境理事会。( 2018 年)。摩天大楼中心。从 [ter.com/city/chicago](http://ter.com/city/chicago) 摩天大楼检索

戴斯·鲁( 2010 年)。邻里和健康。纽约科学院年鉴, 1186, 125 - 145。道奇, S, Bohrer, G, Weinzierl, R, 戴维森, 南卡罗来纳州, Kays, Douglas, D, ...Wikelski, M. ( 2013 )。环境数据

自动轨迹注释( EnvDATA )系统:将动物轨迹与环境数据联系起来。运动生态学, 1, 3。

魏贝尔·道奇( 2009 )。揭示运动的物理:比较不同类型运动物体运动特征的相似性。计算机、环境和城市系统, 33, 419 - 434。唐斯, 法学硕士和霍纳, 麻省理工学院( 2012 )。用于可视化和分析车辆跟踪数据的概率潜在路径树。运输地理学杂志, 23, 72 - 80。

杜美思, & 陈, H. ( 2000 )。网页内容的分级分类。《第 23 届国际 ACM SIGIR 信息检索研究与发展年会记录》(第 256 - 263 页)。希腊雅典: ACM。美国南卡罗来纳州彭特兰郡鹰镇& Lazer, D. ( 2009 )。利用手机数据推断友谊网络结构。美国国家科学院学报, 106, 15274 - 15278。

埃利斯, K, 戈伯乐, 马歇尔, 兰克立特, 斯陶丹迈尔, J, & Kerr, J. (2014)。使用 GPS、加速度计和机器学习算法识别挑战性环境中的主动旅行行为。公共卫生, 2, 39 - 46。埃文森, 英国, 弗伯格, 研发(2017)。移动应用:跟踪身体活动和位置的数字日记。英国运动医学杂志, 51, 1169 - 1170。

方, S H、廖、H H、范、Y X、陈、K H、黄、J W、陆、Y D、& Tsao, Y. (2016)。使用智能手机上的传感器对运输模式进行分类。传感器, 16, 1324。

冯涛, &蒂姆曼斯, 新泽西州(2013)。使用 GPS 和加速度计数据进行运输模式识别。交通研究 C 部分: 新兴技术, 37, 118 - 130。

纽约州费雷拉市、新泽西州 Poco 市、纽约州 Vo 市、新泽西州 Freire 市和康涅狄格州 Silva 市(2013 年)。大时空城市数据的视觉探索:纽约出租车出行研究。IEEE 可视化和计算机图形学交易, 19, 2149 - 2158。弗里德曼(2001)。贪婪函数逼近:一种梯度增强机器。统计年鉴, 29, 1189 - 1232。龚, H, 陈, 比亚洛斯托茨基, 东, 和劳森, 北卡罗来纳州(2012)。一种用于纽约市旅行模式检测的 GPS/GIS 方法。

计算机、环境和城市系统, 36, 131 - 139。

冈萨雷斯, 麻省, 伊达尔戈, 加州, & Barabasi, 洛杉矶(2008)。了解个人的移动模式。自然, 453, 779 - 782。

戈登-拉森, 纳尔逊, 麻省, 和梁, K. (2005)。年轻人的积极交通、体育活动和体重状况之间的联系。肥胖研究, 13, 868 - 875。

格雷瓦尔, 麻省(2011)。卡尔曼滤波。在洛夫里奇先生(埃德。), 国际统计科学百科全书(第 705 - 708 页)。德国柏林:斯普林格。

J. Schipperijn, Jankowska, M.和 Kerr. J. (2015)。在体育活动和久坐行为研究中使用 GPS 数据的框架。《运动与运动科学评论》, 43、48 页。

扬森, 麻省, 艾特玛, Pierik, F, & Dijst, 麻省(2016)。体育设施、购物中心或家庭:在什么位置

对成年人的身体活动很重要吗? 横断面研究。国际环境研究与公共卫生杂志, 13, 287。

江俊杰, 关斯勒, 林瑞杰和奥格莱(2006)。平滑方法, 以最小化全球定位系统随机误差对行驶距离、速度和加速度分布估计的影响。运输研究记录, 1972 年, 141 - 150。关文平(2004)。时间地理研究中的 GIS 方法:人类活动模式的地理计算和地理可视化。Geografiska Annaler B, 86, 267 - 280。

关文平(2012a)。GIS 如何帮助解决社会科学研究中不确定的地理环境问题。地理信息系统年鉴, 18, 245 - 255。

关文平(2012 年)。不确定的地理环境问题。美国地理学家协会年鉴, 102, 958 - 968。

关敏(2013 年)。超越空间(正如我们所知):迈向隔离、健康和可及性的时间整合地理。美国地理学家协会年鉴, 103, 1078 - 1086。

郭培思, 张志平, 张志平, 张志平( 2011 )。使用手机加速计的活动识别。ACM SigKDD 探索通讯, 12, 74 - 82。

纽约拉乔维茨, 琼斯, 美联社, 页, 南太平洋, 惠勒, 北太平洋, &库珀, 南太平洋( 2012 )。全球定位系统能告诉我什么

我们了解不同类型的城市绿地对儿童体育活动的贡献? 健康与地方, 18, 586 - 594。

林美思、葛伯乐、陈美思、奥利弗、孟达兰、马歇尔、林美思、…克尔( 2013 )。测量花在户外的时间

使用可佩戴的照相机和全球定位系统。《第四届国际传感和普及成像会议记录》(第 1 - 7 页)。加州圣地亚哥。

劳贝, 丹尼斯, T, 福尔, P, &沃克, M. ( 2007 )。超越地理空间生命线的快照动态分析。计算机、环境和城市系统, 31, 481 - 501。

李克勤和关文平( 2017 )。使用智能手机加速度计数据对自由生活条件下的身体活动进行分类, 并探索预测结果。计算机、环境和城市系统, 67, 124 - 131。文学硕士和文学硕士( 2002 )。随机森林分类和回归。R 新闻, 2, 18 - 22。娄开头-菱铁矿, A. ( 2006 )。走路安全吗? 邻里安全和安保考虑因素及其对走路。CPL 书目, 20, 219 - 232。

麦克纳马拉、克罗斯利、罗斯科、艾伦、洛克、戴杰( 2015 )。作文评分的分级分类方法。评估写作, 23, 35 - 59。

新泽西州耶苏伦市莫西瓦市和纽约州蒂姆曼斯市( 2010 年)。活动旅行日记的半自动归集:全球使用定位系统跟踪、提示回忆和上下文敏感的学习算法。运输研究记录, 2183, 60 - 68。

户外基金会。( 2016 年)。2016 年户外娱乐参与在线报告。检索自 <https://www.室外-foundation.org/pdf/ResearchParticipation2016Topline.pdf>

马萨诸塞州帕宾斯基市, 马萨诸塞州斯科特市和新泽西州多尔蒂市( 2009 年)。探索路线选择决策过程:比较

使用基于人的 GPS 获得的计划和观察路线。交通研究 F 部分:交通心理学和行为, 12, 347 - 358。

帕特里克, K, 克尔, J, 诺曼, G, 瑞安, 萨利, J, 克鲁格, I, …安斯沃思, B. ( 2008 )。地理空间测量和物理活动分析:物理活动位置测量系统( PALMS )。流行病学, 19, S63。康明斯、康明斯、康明斯、康明斯、康明斯、康明斯、康明斯( 2013 年)。流行病学中环境博览会的概念化和测量:考虑与日常活动相关的活动空间。健康与地点, 21, 86 - 93。彼得里斯( 2009 )。动态线性模型贝叶斯分析的 R 包。从 <ftp://ftp.math.ethz.ch/SFS/pub/软件/RCRAN/web/软件包/dlm/插画/dlm.pdf> 中检索

美国加州大学预科, Gidyfalvi, G. & Susilo, YO. ( 2017 )。运输模式检测:对适用性和可靠性的深入审查。运输评论, 37, 442 - 464。

统计计算基金会。( 2013 年)。R :统计计算的语言和环境。奥地利维也纳: R 统计计算基金会。



罗德里格兹, 乔, G- H, 埃文森, 康威, T. L .科恩, Ghosh - Dastidar, B ., Lytle, L. A. ( 2012 ). 出去

关于:建筑环境与青春期女性身体活动行为的关联。健康与地点, 18, 55 - 62。

《美国马萨诸塞州鲍曼市萨利斯市和普拉特市》( 1998 年)。促进体育活动的环境和政策干预。美国预防医学杂志, 15, 379 - 397。

罗马克斯·史兰克, 艾泽勒, B, T, & Bak, J. ( 2015 )。2015 年城市交通记分卡。从 <https://static.tti.tamu.edu/tti.tamu.edu/文档/移动性记分卡-2015.pdf> 中检索

夏菲克, 麻省, 和海涛, 东( 2016 )。不同智能手机数据采集频率的旅行模式检测。传感器, 16, 716。

苏福尔, 纽约&艾萨克森( 2007 )。跟踪数字时代的游客。旅游研究年鉴, 34, 141 - 159。苏福尔, N, 瓦尔, H....., G, 艾萨克森, M, 奥斯瓦尔德, F, Edry, T, ...Heink, J. ( 2011 )。全球定位的使用

测量具有不同认知功能的老年人外出活动的系统。老龄化与社会, 31, 849 - 869。

纽约州纳泽尔市泰尼奥市、Gütschi 市、南卡罗来纳州卡尔梅尔市、罗哈斯-鲁达市、新文惠森市、伍德科克市( 2016 年)。空气污染会否定骑自行车和散步对健康的好处吗? 预防医学, 87, 233 - 236。丹登, 彭丽萍, 萨伦斯, 周正康, 克尔, J, &克里斯塔克斯, 美国民主党( 2013 )。幼儿园儿童的室内和室外时间。美国预防医学杂志, 44, 85 - 88。

汤林特区( 1990 )。地理信息系统和制图建模。新泽西州恩格伍德悬崖:普伦蒂斯大厅。罗培德, 彭丽萍, 威尔逊, 马修斯, 克罗米莱, 康乃尔, 梅莱, 史瑞杰( 2010 )。建筑环境和基于位置

身体活动。美国预防医学杂志, 38, 429 - 438。

范迪杰( 2018 )。利用多个移动窗口从 GPS 数据中识别活动旅行点。计算机, 环境和城市系统, 70, 84 - 101。

马萨诸塞州沃斯市、新泽西州西姆斯-古尔德市、马萨诸塞州阿什市、马萨诸塞州麦凯市、加利福尼亚州普格市和马萨诸塞州温特斯市( 2016 年)。公共交通使用和体育活动

居住在社区的老年人:结合 GPS 和加速度计来评估交通相关的身体活动。运输与健康杂志, 3, 191 - 199。

王建军, 关颖珊, 麻省理工, & Chai, Y. ( 2018 )。一种创新的基于环境的晶体生长活动空间方法

精神暴露评估:一项利用芝加哥收集的 GIS 和 GPS 轨迹数据的研究。国际环境研究与公共卫生杂志, 15, 703。

威斯, 通用, 洛克哈特, 普利克卡尔, 麦休, 麻省理工, 罗南, 麻省理工, 和蒂姆科, 新泽西州 (2016)。行动跟踪器:智能手机

基于活动识别系统来改善健康和福祉。《2016 年 IEEE 数据科学与高级分析国际会议记录》(第 682 - 688 页)。加拿大蒙特利尔: IEEE。威特扬库尔, A, 霍兰诺, T, 小野, N, Sekimoto, Y, & Shibasaki, R. (2013)。旅行重建和运输

对来自移动电话的低数据速率 GPS 数据的模式提取。第十三届国际城市规划和城市管理计算机会议记录。荷兰乌得勒支。

肖日志, 王永义, 傅克英, 吴福福(2017)。使用基于树的集成分类器从轨迹数据中识别不同的运输模式。ISPRS 国际地理信息杂志, 6, 57。

张 s, 麦库拉, P, 纽金特, C, &郑, H. (2010)。使用智能手机的加速度计监控活动

分级分类。《2010 年第六届智能环境国际会议记录》(第 158 - 163 页)。马来西亚吉隆坡。

郑、Y、陈、Y、李、Q、谢、X、& Ma、W Y (2010)。了解基于 GPS 数据的网络应用的交通模式。ACM 网上交易, 4, 1。

郑, Y, 李, Q, 陈, Y, 谢, X, & Ma, W Y (2008)。基于 GPS 数据了解移动性。《第十届泛在计算国际会议记录》(第 312 - 321 页)。韩国首尔: ACM。周, 十(2014)。调查建筑环境与使用场所的年轻人主动旅行之间的关系

基于技术(未发表的博士论文)。伊利诺伊大学厄巴纳-香槟分校, 伊利诺伊州厄巴纳-香槟分校。

朱、Q、朱、M、李、M、傅、M、黄、Z、甘、Q、&周、Z. (2016)。从原始全球定位系统中识别运输模式

数据。《第二届国际青年计算机科学家、工程师和教育工作者会议记录》(第 395 - 409 页)。中国哈尔滨。

如何引用这篇文章: 李 K, 关 M - P。基于 GPS 和加速度计数据的自动身体活动和车内状态分类:一种使用机器学习技术的分级分类方法。地理信息系统中的交易。2018 年; 00 : 1 - 28。https://doi.org/10.1111/tgis.12485

## APPENDIX

**TABLE A** All 31 GPS features and relative importance for indoor/outdoor classification (bold font: three selected important features)

GPS trajectory operator	Movement derivative	Relative importance	Importance order
Instantaneous	Velocity	0.001	25
	Acceleration	0.000	31
Interval	HDOP	0.001	24
	10 s	0.003	14
	average velocity	0.000	30
	average acceleration	0.003	16
	number of missing GPS points	0.010	5
	HDOP	0.002	20
	20 s	0.000	29
	average velocity	0.001	28
	average acceleration	0.004	13
	number of missing GPS points	0.002	17
	HDOP	0.002	21
	30 s	0.009	6
	average velocity	0.002	18
	average acceleration	0.001	23
	number of missing GPS points	0.001	27
	HDOP	0.004	10
	60 s	0.003	15
	average velocity	0.004	11
	average acceleration	0.002	19
	number of missing GPS points	0.001	26
	HDOP	0.008	8
	90 s	0.020	4
	average velocity	0.002	22
	average acceleration	0.050	3
	number of missing GPS points	0.006	9
	HDOP	0.488	1
	120 s	0.004	12
	average velocity	0.009	7
	average acceleration	0.356	2
	number of missing GPS points		
	HDOP		
	180 s		
	average velocity		
	average acceleration		
	number of missing GPS points		
	HDOP		

GPS 轨迹算子移动导数相对重要性顺序瞬时速度加速度表 B 所有 73 个 GPS 特征和使用 GPS 数据分类的相对重要性(粗体字体:前 20 个重要特征)

Interval	Recorded time (h)	825	35
Time window			
10 s	average velocity	776	37
	average acceleration	343	67
	max velocity	754	38
	max acceleration	409	60
	change rate of velocity	207	73
20 s	average velocity	1,671	15
	average acceleration	415	59
	max velocity	946	31
	max acceleration	456	55
	change rate of velocity	297	71
30 s	average velocity	1,612	16
	average acceleration	440	57
	max velocity	1,118	23
	max acceleration	533	49
	change rate of velocity	485	52

(Continued)





TABLE B (Continued)

GPS trajectory operator	Movement derivative	Relative importance	Importance order
60 s	average velocity	4,524	8
	average acceleration	524	50
	max velocity	2,975	10
	max acceleration	791	36
	change rate of velocity	595	44
90 s	average velocity	4,641	7
	average acceleration	588	46
	max velocity	7,693	3
	max acceleration	1,005	29
	change rate of velocity	1,015	28
120 s	average velocity	9,007	2
	average acceleration	598	43
	max velocity	5,543	4
	max acceleration	1,089	25
	change rate of velocity	1,239	19
180 s	average velocity	9,088	1
	average acceleration	737	40
	max velocity	5,429	5
	max acceleration	1,247	18
	change rate of velocity	2,562	11

(Continued)



TABLE B (Continued)

GPS trajectory operator	Movement derivative	Relative importance	Importance order
<i>Distance window</i>			
10 m	average velocity	1,215	21
	average acceleration	311	69
	<b>max velocity</b>	<b>1,229</b>	<b>20</b>
	max acceleration	389	62
	change rate of velocity	232	72
20 m	average velocity	1,103	24
	average acceleration	346	66
	max velocity	472	53
	max acceleration	421	58
	change rate of velocity	326	68
30 m	average velocity	890	32
	average acceleration	358	65
	max velocity	662	42
	max acceleration	460	54
	change rate of velocity	384	63
40 m	<b>average velocity</b>	<b>1,346</b>	<b>17</b>
	average acceleration	370	64
	max velocity	841	34
	max acceleration	499	51
	change rate of velocity	448	56

(Continued)





TABLE B (Continued)

GPS trajectory operator	Movement derivative	Relative importance	Importance order
50 m	average velocity	1,209	22
	average acceleration	392	61
	max velocity	2,260	12
	max acceleration	567	47
	change rate of velocity	595	44
100 m	average velocity	1,767	14
	average acceleration	556	48
	max velocity	2,208	13
	max acceleration	707	41
	change rate of velocity	740	39
200 m	average velocity	4,840	6
	average acceleration	845	33
	max velocity	3,989	9
	max acceleration	1,040	27
	change rate of velocity	966	30

**TABLE C** Test results of the application of Kalman filtering to improve performance of physical activity and in-vehicle status classification

(a) No filtering methods (predictive accuracy: 95.26%)							
Predicted class	Actual class						
	Running	Walking	Sitting	Standing	In-vehicle	Biking	
Running	414	21	0	0	0	0	0
Walking	3	6,716	1,363	8	33	0	0
Sitting	0	80	65,855	0	86	4	4
Standing	0	0	1,019	2,789	4	0	0
In-vehicle	15	28	81	156	1,684	774	774
Biking	231	0	0	0	0	1,178	1,178
Accuracy (%)	62.14	98.12	96.39	94.41	93.19	60.23	60.23
(b) Kalman filtering with variance of observation noise 0.8 and variance of systematic noise 0.01 (predictive accuracy: 96.06)							
Predicted class	Actual class						
	Running	Walking	Sitting	Standing	In-vehicle	Biking	
Running	625	21	0	0	0	0	0
Walking	25	6,708	1,363	8	41	0	0
Sitting	7	80	65,924	0	390	4	4
Standing	0	0	1,019	2,945	17	0	0
In-vehicle	6	36	4	0	1,179	43	43
Biking	0	0	8	0	180	1,909	1,909
Accuracy (%)	94.27	98.00	96.50	99.73	65.25	97.60	97.60

(Continued)



**TABLE C** (Continued)

(c) Kalman filtering with variance of observation noise 0.8 and variance of systematic noise 0.1 (predictive accuracy: 96.20%)							
Predicted class	Actual class						
	Running	Walking	Sitting	Standing	In-vehicle	Biking	
Running	464	19	0	0	0	0	0
Walking	3	6,725	1,363	8	27	0	0
Sitting	0	80	65,889	0	101	2	2
Standing	0	0	1,019	2,889	17	0	0
In-vehicle	7	19	47	56	1,662	175	175
Biking	189	2	0	0	0	1,779	1,779
Accuracy (%)	69.98	98.25	96.44	97.83	91.98	90.95	90.95
(d) Kalman filtering with variance of observation noise 0.8 and variance of systematic noise 1.0 (predictive accuracy: 95.61%)							
Predicted class	Actual class						
	Running	Walking	Sitting	Standing	In-vehicle	Biking	
Running	331	8	0	0	0	0	0
Walking	1	6,702	1,363	8	34	0	0
Sitting	0	80	65,879	0	101	2	2
Standing	0	0	1,019	2,790	5	0	0
In-vehicle	5	42	57	155	1,667	401	401
Biking	326	13	0	0	0	1,553	1,553
Accuracy (%)	49.92	97.91	96.43	94.48	92.25	79.40	79.40

[查看出版物统计信息](#)