

引用格式:赵斌,韩晶晶,史覃覃,等.语义轨迹建模与挖掘研究进展[J].地球信息科学学报,2020,22(4):842-856. [Zhao B, Han J J, Shi T T, et al. Advancements in semantic trajectory modelling and mining[J]. Journal of Geo- information Science, 2020,22(4):842- 856.] DOI: 10.12082/dqxxkx.2020.190257

语义轨迹建模与挖掘研究进展

赵斌¹,韩晶晶¹,史覃覃¹,吉根林^{1*},刘信陶²,俞肇元³

1. 南京师范大学计算机科学与技术学院,南京 210023;2. 香港理工大学土地测量及地理资讯学系,香港 999077;
3. 南京师范大学地理科学学院,南京 210023

Advancements in Semantic Trajectory Modelling and Mining

ZHAO Bin¹, HAN Jingjing¹, SHI Tantan¹, JI Genlin^{1*}, LIU Xintao², YU Zhaoyuan³

1. Nanjing Normal University, School of Computer Science and Technology, Nanjing 210023, China; 2. The Hong Kong Polytechnic University, Department of Land Surveying and Geo-informatics, Hong Kong 999077, China; 3. Nanjing Normal University, School of Geography, Nanjing 210023, China

Abstract: A semantic trajectory is a combination of a spatiotemporal trajectory and semantic information. Besides spatiotemporal information, a semantic trajectory comprises movement states (e.g. speed, direction), external contextual information (e.g. temperature, spatial topological relationships), and social relationships (e.g. friend relationships, social activities) of moving objects. We can derive from semantic trajectories intentions, habits, emotions, and other high order semantic information, thus further discover the patterns, relationships, and rules of individual or collective mobility behaviors. Therefore, compared with spatiotemporal trajectories, semantic trajectories are more in line with the practical requirements of decision-making applications in terms of semantics, interpretation, feasibility, and so on. This paper reviews the key technologies of semantic trajectory mining. First, we introduce the concept of semantic trajectories, and summarize four classic types of semantic trajectory definitions according to semantic elements. Then, we introduce the main phases of semantic trajectory modeling, including preprocessing, trajectory segmentation, and semantic enrichment. Since semantic trajectories cannot be acquired from location-acquisition devices as spatiotemporal trajectories, semantic trajectories need to be obtained through modeling techniques. Thus, the basic idea is to combine spatiotemporal trajectories with semantic information to generate corresponding semantic trajectories. Next, we introduce the main tasks of semantic trajectory mining, including semantic trajectory pattern mining, semantic trajectory clustering, semantic trajectory classification, anomaly detection of semantic trajectories, and so on. For each mining task, this paper introduces the basic principles and related algorithms, and summarizes the main key technologies and challenges. Finally, this paper concludes with the existing challenges and promising research directions of semantic trajectory mining. Specifically, this paper discusses the important research issues of semantic trajectory modeling in aspects including modeling definition, semantic annotation technologies, and multi- source data

收稿日期:2019-05-27;修回日期:2019-10-31.

基金项目:国家重点研发计划项目(2017YFB0503500);国家自然科学基金项目(41971343、41571379)。[**Foundation items:** National Key Research and Development Program of China, No.2017YFB0503500; National Natural Science Foundation of China, No.41971343, 41571379.]

作者简介:赵斌(1978—),男,江苏南京人,博士,副教授,主要从事时空数据挖掘研究。E-mail: zhaobin@njnu.edu.cn

*通讯作者:吉根林(1964—),男,江苏海安人,博士,教授,主要从事数据挖掘研究。E-mail: glji@njnu.edu.cn

modeling. This paper also discusses the promising research issues of semantic trajectory mining such as semantic trajectory data management, classification and prediction, trajectory stream mining, privacy protection, multi-granularity mining, and evaluation methods.

Key words: semantic trajectory; semantic trajectory modeling; semantic trajectory mining; semantic enrichment; ontology knowledge base; semantic trajectory pattern mining; semantic trajectory clustering; semantic trajectory classification; semantic trajectory outlier detection

***Corresponding author:** Ji Genlin, E-mail: gjli@njnu.edu.cn

摘要: 语义轨迹是时空轨迹和语义信息融合的产物。除了含有时空信息以外,语义轨迹包括移动对象自身的运动状态(如速度、方向)、环境(如气温、空间拓扑关系)和社交关系(如好友关系、社交活动)等多方面信息。挖掘语义轨迹可以深入地发现个体或群体移动行为的意图、习惯、情感等高阶语义内容,从而深层次发现个体或群体移动行为的模式、关系和规律等。因而,相较于时空轨迹,语义轨迹在语义性、解释性、可行性等方面更符合决策分析应用的实践需求,具有更重要的研究意义和应用价值。本文对语义轨迹挖掘的关键技术进行了综述。首先,介绍语义轨迹的基本概念,并且根据语义元素类型的不同总结了4种常见的定义形式。其次,归纳了语义轨迹建模的基本阶段,包括预处理、轨迹分段和语义富化。由于语义轨迹无法像时空轨迹那样从位置感知设备中采集获得,因此语义轨迹是通过建模技术得到的,主要通过将语义信息和时空轨迹相融合生成相应的语义轨迹。然后,介绍语义轨迹挖掘的主要任务,包括语义轨迹模式挖掘、语义轨迹聚类、语义轨迹分类、语义轨迹异常检测等。针对每一项挖掘任务,介绍了有关的基本原理和相关算法,总结了主要的关键技术和挑战。最后,探讨了语义轨迹挖掘现存的研究难点和未来研究方向。从模型定义、语义标注技术、多源数据建模等方面,讨论了语义轨迹建模的重要研究问题;从语义轨迹数据管理、分类和预测、流式数据挖掘、隐私保护、多粒度挖掘、评价方法等方面,探讨了语义轨迹挖掘的未来研究问题。

关键词: 语义轨迹;语义轨迹建模;语义轨迹挖掘;语义富化;本体知识库;语义轨迹模式挖掘;语义轨迹聚类;语义轨迹分类;语义轨迹异常检测

1 引言

随着传感器技术、无线通信技术和全球定位技术的日益发展与成熟,采集海量的时空轨迹数据已经成为可能。时空轨迹是地理空间加上时间轴所形成的多维空间中的一条曲线,可以表示移动对象在一段较长时间范围内的位置变化^[1]。近年来,学术界针对时空轨迹挖掘开展了大量研究工作,包括运动模式挖掘^[2]、聚类分析^[3-4]、轨迹分类^[5]、异常检测^[6]等方面。但随着轨迹数据挖掘的深入研究,研究人员开始关注语义轨迹的挖掘问题。不同于时空轨迹,语义轨迹并非通过位置感知设备采集获得,而是通过语义轨迹标注技术将时空轨迹和语义信息结合产生的。因而,除了含有时空信息以外,语义轨迹包括移动对象自身的运动状态(如速度、方向)、环境(如气温、空间拓扑关系)和社交(如好友关系、社交活动)等多方面信息。基于语义轨迹的数据挖掘方法通过融合轨迹中的时空信息与语义信息,可以深入地发现移动行为中蕴含的意图、习惯、情感等高阶语义内容。不难发现,语义轨迹的挖掘结果相较于时空轨迹在语义性、解释性、可行性等方面更符合决策分析应用的实际需求,并且更

具有研究意义和应用价值。

语义轨迹数据挖掘在融合不同类型的应用场景信息和领域知识后可用于多种应用领域。例如,结合个人信息的语义轨迹可以用于个人位置推荐和预测。齐凌艳等^[7]基于语义轨迹为用户提供个性化的位置推荐。该方法将用户个性化特征、地标名称等语义信息融入用户移动轨迹,结合用户的兴趣、职业等个人信息为用户推荐附近的兴趣热点。Yao等^[8]提出一种基于语义轨迹的位置预测方法。采用结合位置、时间、用户偏好等多种信息的递归模型SERM,根据用户的历史轨迹发现用户移动的时空语义规律,从而进行位置预测。又如,结合环境信息的语义轨迹可用于环境监测。Zheng等^[9]在道路网络环境中将车辆轨迹与气象、POI(Point of Interest)等环境数据相结合,推断城市空气的污染程度,为城市空气质量检测与治理提供决策参考。再如,结合领域知识的语义轨迹可以用于事件检测与分析。Patroumpas等^[10]结合船舶航行的领域知识构建语义轨迹用于海事监控系统,检测船只的可疑状态或危险情况,以便为海事管理部门提供预警通知和决策支持。

本文对语义轨迹数据挖掘的研究情况进行了

总结。首先介绍常见的语义轨迹定义形式,它是语义轨迹查询、存储、检索和挖掘的数据组织形式;然后阐述语义轨迹的建模过程,主要包括轨迹分段和语义富化处理^[1];其次,分别介绍了语义轨迹的模式挖掘(包括频繁模式和周期模式)、聚类、分类和异常检测等研究情况;最后,探讨了语义轨迹挖掘现存的主要挑战和未来的研究方向。

2 语义轨迹定义

语义轨迹的定义形式源自于时空轨迹,但信息更丰富,类型更多样,结构更复杂。时空轨迹是轨迹数据中最基本的形式。按照轨迹序列中元素类型的不同,原始轨迹通常采用2种基本定义形式:轨迹点序列和轨迹边序列^[12]。虽然语义轨迹的定义也采用序列形式,但为了对移动对象的移动行为进行语义层次的抽象表达,需要在定义中加入语义信息。然而,因为语义信息数据来源广泛、表现形式多样、概念层次不一,所以导致了语义轨迹的定义比时空轨迹更复杂。

根据语义轨迹中信息单元语义层次的不同,语义轨迹定义分为4类:基于停留(Stop)和移动(Move)的语义轨迹定义、基于地理空间对象的语义轨迹定义、基于轨迹片段(Episode)^[13]的语义轨迹定义和基于事件的语义轨迹定义。

2.1 基于停留和移动的语义轨迹定义

移动对象在运动过程中具有2种常见的基本状态,即停留和移动^[14]。在时空场景下,移动对象的这两种状态可以采用子轨迹来表示,分别是“停留”子轨迹和“移动”子轨迹。于是,移动对象的语义轨迹可以定义为由“停留”和“移动”子轨迹构成的序列,形如 $\{begin, move_1, stop_1, \dots, move_n, stop_n, end\}$ 。其中,“begin”和“end”分别是移动对象的开始和结束状态,“move”和“stop”分别为移动和停留子轨迹。典型的研究工作文献[15]–[20]。

2.2 基于地理空间对象的语义轨迹定义

虽然基于停留和移动的语义轨迹定义相较于时空轨迹已经表现出了移动对象运动行为方面的语义性,但停留和移动状态仍属于移动行为的低阶语义信息。在此基础上,研究人员提出了基于地理空间对象的轨迹定义。该方法根据时空轨迹与地理空间对象的空间拓扑关系对移动行为进行

建模,将语义轨迹表示成由地理空间对象组成的序列,形如 $\{go_1, \dots, go_n\}$ 。go为地理空间对象,以二元组 $\langle region, interval \rangle$, $region = \{(x_1, y_1), \dots, (x_n, y_n)\}$, $interval = \{t_{begin}, t_{end}\}$ 形式表示。常见的地理空间对象为兴趣区域(Region of Interest, ROI)^[21–22]。基于此想法,Manaa等^[23]按照城市基础设施的不同功能定义了5种常见的基础设施对象(Infrastructure Objects),然后基于这些对象定义了语义轨迹。

2.3 基于轨迹分段的语义轨迹定义

2013年Fileto等^[24]提出了基于“轨迹片段”概念的语义轨迹定义方法^[25]。轨迹片段是指移动对象的某种特征持续保持稳定状态的运动片段。它既可以描述移动行为的运动特征(如速度、加速度、方向等),也可以描述环境特征(如温度、具有空间拓扑关系的ROI、附近的好友等)。由于轨迹片段在语义信息表达的广度和深度方面都优于前2种语义轨迹的定义,因此可以采用轨迹片段序列的形式定义语义轨迹^[24–28],形如 $\{episode_1, episode_2, \dots, episode_n\}$ 。图1展示了基于轨迹片段的语义轨迹的定义形式。

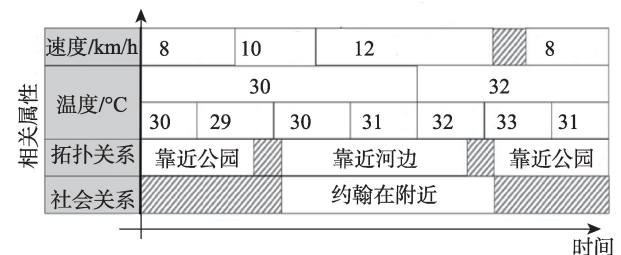


图1 基于轨迹片段的多层语义轨迹定义方法示例^[28]

Fig. 1 Example of multi-layered representation of a trajectory and context based on episodes

2.4 基于事件的语义轨迹定义

上述3种语义轨迹定义主要是从空间语义、时间语义和属性语义的角度对移动对象运动行为进行建模。然而,移动对象运动行为中还包含另一种重要的高阶语义信息——事件。事件包含了移动对象状态的改变、空间位置的变化、对象属性的演变等多方面内容。围绕事件组织的空间、时间和属性等信息更能准确、有效地反映事件对象的移动行为。

现有的研究工作主要以个体或者群体的“活动(Activity)”作为事件进行研究。常见的活动有上班、娱乐、购物、就餐等^[22–23,26,29]。基于这些活动信息可以将语义轨迹表示为活动事件的序列,形如 $\{(st_1, activity_1), (st_2, activity_2), \dots, (st_n, activity_n)\}$ 。其

中, st 为活动事件的时空属性, 定义为四元组 (x, y, t, POI) , 分别是经度、纬度、时间和兴趣位置点; 而 $activity$ 代表活动事件的语义标签。

3 语义轨迹建模方法

语义轨迹建模是语义轨迹分析挖掘的前提与基础。由于语义轨迹无法像时空轨迹一样从位置感知设备中采集获得, 因此语义轨迹建模的主要任务是完成时空轨迹的语义化处理, 即通过融合语义信息为时空轨迹生成相应的语义轨迹。其中, 语义信息主要来源于移动对象的个体属性、地理环境和领域知识3个方面。

语义轨迹建模主要包括数据预处理、轨迹分段和语义富化3个阶段(图2): ① 数据预处理用于解决轨迹数据可能存在的数据质量问题, 包括定位数据不准确、数据不完整、数据缺失、数据不一致等; ② 轨迹分段按照移动对象运动状态的不同对原始轨迹进行划分和识别, 发现轨迹中的“停留”和“移动”片段; ③ 基于轨迹片段识别的结果, 通过融合外在的语义信息完成语义富化的处理任务。

在上述3个阶段中, 除了数据预处理之外, 其余阶段都是针对语义轨迹必要的建模处理过程。

3.1 轨迹分段

无论是时空轨迹还是语义轨迹, 在轨迹预处理中都包含轨迹分段处理。但2种轨迹进行分段处理的目的和方法并不相同。时空轨迹的分段处理主

要依据时间间隔、轨迹形状和语义信息进行轨迹分段, 常用于轨迹压缩^[30-31]等预处理阶段。而语义轨迹的分段主要依据移动行为的属性特征进行轨迹分段, 为轨迹语义富化做准备, 且有助于后续语义信息的组织和语义轨迹的表示。被识别出的每个轨迹分段将被赋予一个语义标签, 如运动状态、常见的交通模式^[32]等。

目前, 最常见的轨迹分段方式是按照移动对象的运动状态将轨迹划分成不同片段, 如停留片段(Stop Episode)和移动片段(Move Episode)。这样语义轨迹可以定义成停留片段和移动片段交替出现的片段序列。例如, Wang等^[33]采用栅格的方法检测重要的停留点, 将GPS轨迹分割成停留/移动交替出现的序列结构。由此可见, 轨迹分段问题可以转变为“停留片段”和“移动片段”的识别问题。

3.1.1 停留片段识别

识别移动轨迹中停留片段需要在结合环境信息和应用领域的前提下通过综合分析运动状态而实现。需要指出的是, “停留”并非是指移动对象完全静止, 而是在邻域范围内的“逗留”。与移动片段相比, 主要区别在于是否属于跨区域的移动行为。所以, 仅从运动状态的变动性方面是难以判断的, 如何结合环境和应用领域才是准确识别停留片段的关键。

在早期研究中, 轨迹分段研究主要关注轨迹中停留片段的识别问题。研究人员从移动行为的固有属性(如速度、方向等)入手, 提出了多种识别停留片段的方法。Ashbrook等^[34]利用定位信号的状

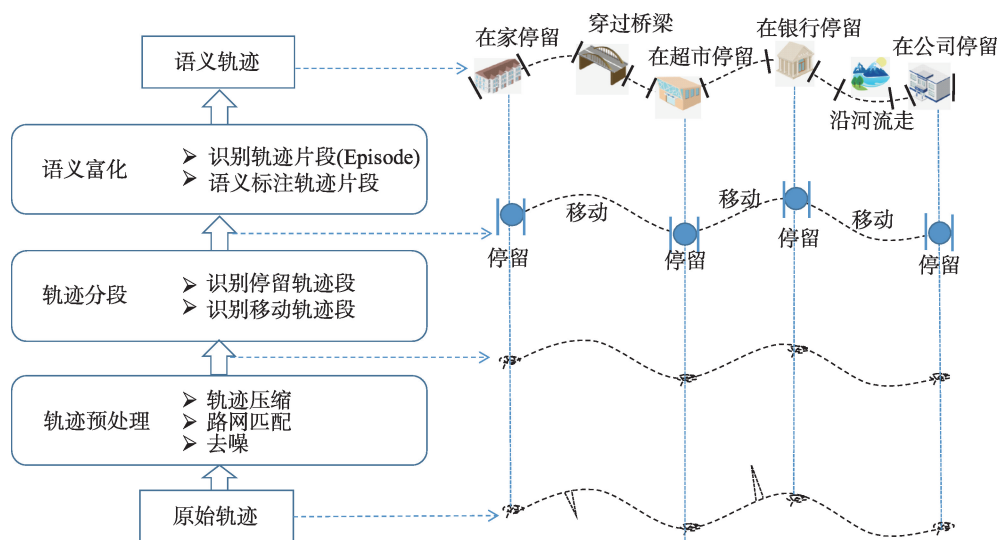


图2 语义轨迹建模过程

Fig. 2 Semantic trajectory modeling processes

态变化(GPS是否关闭或者车辆是否进入车库)和移动速度(速度是否为零)识别停留片段。但是,由于定位信号误差和测速效果不佳等原因,而无法识别实际的停留片段。Krumm等^[35]调整了识别停留片段的规则,如移动物体在5 min内无位移或者保持至少5 min的低速运动,则判断此段轨迹为停留片段。Andrienko等^[36]采用更长的时间阈值(如2 h)识别移动对象的访问位置。Zheng等^[37]提出了更严苛的要求,即停留片段的空间距离和持续时间同时满足相应的阈值要求。除了采用设定时空阈值的方法识别停留片段以外,研究人员还设计了利用轨迹中速度和方向的变化识别停留片段的方法。Palma等^[38]将停留片段设定为轨迹中移动速度低于平均速度的片段。Rocha等^[39]利用渔船捕捞停船前船向突变的特性识别渔船作业时的停留片段。此外,结合移动场景中的地理环境信息和应用领域信息也可以识别原始轨迹中的停留片段。Alvares等^[40]利用交通领域信息识别停留片段,例如利用路口交通灯识别车辆的停留区域。

3.1.2 移动片段识别

不同于停留片段的识别问题,移动片段的识别主要考虑移动对象交通状态的变化与交通方式的改变。然而,个体移动行为的多样性和复杂性为移动片段识别问题带来了挑战。个体在城市范围内可以选择的交通方式多种多样,有时无法通过单一运动特征区分不同的交通方式,如在城市环境中公交车和私家车在移动速度方面的差别并不明显。即使使用单一的交通方式,也可能出现运动状态的改变,而造成错误的识别。例如,车辆行驶至路口处等待交通灯时的短暂停留,这可能使得完整的移动片段被切割成多个片段。所以,这种个体移动行为中的多样性和复杂性是移动片段识别面临的主要难点。目前已有一些相关的研究工作。Liao等^[41]利用高斯混合模型识别出轨迹中行走、高速和低速3种不同速度的移动片段。Zheng等^[42]除了考虑移动速度之外,还综合考虑了加速度和速度变化率等因素发现轨迹中不同交通方式(包括自行车、公交车和自驾等)的移动片段。此外,利用道路信息也可以识别轨迹中的移动片段。Yan等^[43]利用交通网络和道路类别等语义信息识别出移动片段。

3.2 语义富化处理

语义富化处理是指时空轨迹语义标注的过程,即根据行为属性、环境信息和领域知识为识别出的

轨迹片段标注合适的语义标签。由于语义信息来源广泛、表现形式多样、概念层次不一,因而要为轨迹片段标注全面、完整、丰富的语义信息是语义富化处理面临的主要挑战。其中,有2个关键问题需要解决:①如何根据信息的类型和特征采用合适的计算方法建立语义信息和轨迹的匹配关系。在此方面,研究人员针对具有几何特征的地理对象和具有时空属性的社交网络分别提出了针对性的语义标注方法;②如何组织复杂、多样、丰富的语义信息。目前,主要采用知识库的技术将语义信息进行领域化和模块化组织,用以支撑语义富化处理。后续将从语义标注的信息来源和信息组织方式2个方面总结现有的研究工作。

3.2.1 语义标注的信息来源

语义标注的主要信息来源是地理空间中的地理对象。按照其几何信息的不同,可以分为兴趣区域、兴趣线段和兴趣点^[44]。其中,兴趣区域和兴趣点主要用于标注停留片段,而兴趣线段用于标注移动片段。常见兴趣区域标注方法主要采用空间连接(Spatial Join)算法发现与原始轨迹具有拓扑关系的兴趣区域。然而兴趣点的标注方法完全不同,在兴趣点的密集区域中,停留片段可能与多个兴趣点相关,再加上GPS采样率较低使得识别有效的POI更加困难。Yan等^[45]在轨迹数据质量有限的情况下采用隐式马尔可夫模型为停留片段标注POI类别。另一方面,兴趣线段标注方法常采用全局地图匹配算法确定移动片段的所在路段,并且推断交通方式(如步行、骑自行车、乘地铁等)。Yan等^[44]提出了“点一路段距离(Point-segmentdistance)”的度量方法替代全局地图匹配算法中原有的距离函数,更适用于密集的道路网络、平行道路、交叉路口等地理场景中的兴趣线段标注问题。

近年来,研究人员也尝试采用非地理几何信息开展移动行为的标注研究,尤其是来自于社交网络(如推特、微博等)的信息。2015年,Wu等^[46]结合静态位置信息(如POI)与动态事件信息(如微博或者推特的消息)标注移动行为的主题、目的等。He等^[47]开发了基于位置的社交网络语义标注技术。此方法不仅可以为兴趣点标注类别标签,还可以标注情感标签。Soares等^[48]针对轨迹语义标注任务开发了一个交互式平台VISTA。该系统通过可视化交互手段帮助领域专家处理轨迹数据、结合语义信息与轨迹特征进行轨迹的语义标注工作。

3.2.2 语义信息的组织方式

虽然轨迹语义标注所需的语义信息来源广泛、形式多样。但语义标注方法仍然存在不足,即语义信息缺乏系统性的组织、明确的语义定义和准确的关系描述。语义信息主要来源于移动行为的单一属性(如运动特征、兴趣区域、交通方式等),缺少面向领域的信息来源,以及系统的、完整的组织方式。因而导致标注的信息不全面,尤其缺乏属性间的关系定义与描述,不便于语义轨迹的推断、分析和挖掘。而基于本体的标注方法恰好可以弥补上述研究方法的不足。目前,按照本体类型的不同,基于本体的语义标注方法可以分为2种:基于领域本体(Domain Ontology)的标注方法和基于模块化本体(Modular Ontology)的标注方法。

(1) 基于领域本体的标注方法

基于领域本体的建模主要依靠领域本体对于领域概念及其关系的描述能力,为原始轨迹增强特定领域的语义信息。Baglioni等^[14]提出了一种语义轨迹建模的本体论方法,以“停留”和“移动”概念为基础采用本体的形式定义语义轨迹,并且利用本体的推理能力进一步增强轨迹的语义信息。Renso等^[17]扩展了Baglioni的本体定义,提出了核心本体(Core Ontology)的概念,形式化描述人类移动行为中停留、移动、时间、地点、模式等概念,进一步丰富了语义轨迹的定义。Hu等^[49]采用地理本体将语义轨迹表示为位置序列和线段的组合。Fileto等^[23]提出了基于轨迹片段的语义轨迹建模的Baquara本体模型。与以往方法不同之处在于,该模型利用链接数据(Linked data)补全轨迹的语义信息。该类型数据在本质上是属于语义网的,主要数据来源于DBpedia、LinkedGeoData和GeoCodes等。Vandecasteele^[19]在语义轨迹建模中引入了事件的语义信息。2015年,Fileto等^[26]对Baquara进行了扩展,提出了Baquara本体框架。该方法将层次结构引入到轨迹数据模型中,这其中包括移动分段层次、移动模式层次和包含层次等。

(2) 基于模块化本体的标注方法

虽然基于领域本体的建模方法可以为原始轨迹提供丰富的领域语义信息,但是移动行为的语义信息往往比较多源,并非是单一领域本体可以完全支持的。所以,近年来学术界开展了模块化本体建模方法的研究工作。模块化本体由3类子本体构成,分别是轨迹本体、地理本体和应用本体。轨迹本体描述轨迹在空间、时间和轨迹自身方面的概念

及其关系;地理本体描述与位置相关的地理概念及其关系;应用本体描述应用领域中的概念及其关系。

Wannous等^[50]为了使原始轨迹拥有丰富的语义信息,将不同来源的领域本体和空间本体进行整合得到新的本体,然后利用该本体中的规则对语义轨迹进行建模。Nogueira等^[18,51]提出了QualiTraj本体用于刻画原始轨迹的各种运动特征,尤其是衍生性的特征,如速度、加速度、方向等。Nogueira等^[24]在QualiTraj本体基础上提出了信息描述能力更强的新本体,即语义轨迹片段(Semantic Trajectory Episodes, STEP)本体。它不仅可表达基本的运动特征,还可在更高的语义层次上表达移动轨迹的环境特征。Manaa等^[22]采用本体方法对原始轨迹进行层次式概念建模。按照逻辑层次的不同,轨迹本体分为:几何本体、地理本体和应用领域本体。相应地,基于这些本体设计出了不同语义层次的轨迹数据模型。Nogueira等^[2]提出了基于STEP本体的语义轨迹标注框架FrameSTEP。该方法可以计算每个轨迹片段的各种物理运动和空间几何特征,并利用外部可靠资源(如OpenStreetMap和LinkedGeoData地理知识库)标注轨迹的环境特征。

基于本体知识库的语义信息组织方式的优点十分明显,可以利用本体强大的知识描述和推理能力以及知识库强大的资源优势,为轨迹提供丰富、多样、准确的语义信息,为后续的轨迹数据挖掘提供强有力的支持。但是,此类方法在语义轨迹建模中也存在不足。首先,本体知识库的构建需要领域专家的参与,并且需要被学术界和工业界广泛认可,这在实践应用中通常难以达成;其次,在缺乏权威的专业领域本体知识库的情况下,无法为原始轨迹提供全面、高质量的语义信息;最后,基于本体的语义轨迹与领域事件之间的关系描述不够充分,轨迹与领域事件的结合程度欠佳。

4 语义轨迹挖掘

语义轨迹挖掘是利用数据挖掘方法(关联、聚类、分类、异常检测等)从海量语义轨迹中挖掘有价值的知识和潜在模式,揭示时空场景下个体/群体移动对象的活动行为的规律。语义轨迹挖掘不仅可以直观地了解人类移动行为,还可以洞察移动的意图、习惯、情感等,为位置预测、活动推荐、通勤调查等实践应用提供决策支持。

与时空轨迹挖掘不同的是,语义轨迹挖掘方法

不仅需要考虑时间和空间属性,还需要考虑轨迹中的语义信息。按照挖掘任务的不同,语义轨迹挖掘的主要任务包括:语义轨迹模式挖掘、语义轨迹聚类、语义轨迹分类、语义轨迹异常检测等,如图3所示。

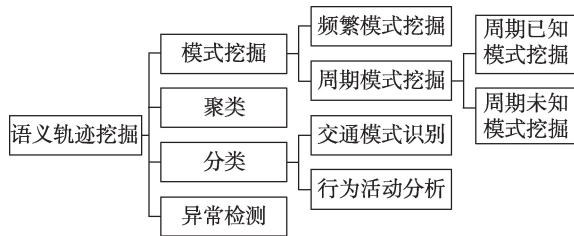


图3 语义轨迹挖掘任务

Fig.3 Tasks of semantic trajectory mining

4.1 语义轨迹模式挖掘算法

现有的语义轨迹主要采用序列结构进行轨迹定义,基于此结构的模式挖掘很自然地受到传统序列模式挖掘研究^[52]的影响。不同之处在于,语义轨迹模式挖掘需要在时空背景下挖掘轨迹中语义元素的序列规律。常见的语义轨迹模式挖掘包括频繁模式挖掘和周期模式挖掘。语义轨迹频繁模式挖掘是发现语义轨迹数据集中频繁出现的时空序列模式。例如,Zhang等^[53]通过挖掘发现在伦敦不少人下班后去酒吧的习惯,而语义轨迹的周期模式挖掘是挖掘个体或者群体移动行为具有明显时间周期性规律的时空序列模式,如每天上下班的通勤行为、野生动物的季节性迁徙等。

语义轨迹模式挖掘包含2个关键问题:模式定义和挖掘算法设计。模式定义主要基于移动对象个体/群体移动行为的关键性特征,通常采用序列结构进行表示;而挖掘算法设计的主要任务是如何降低挖掘算法的计算复杂度,提升挖掘性能。

4.1.1 语义轨迹频繁模式挖掘算法

轨迹频繁模式的研究开始于时空轨迹的相关研究工作。Monreale等^[54]在国际顶级会议“知识发现与数据挖掘”(Knowledge Discovery & Data Mining, KDD)上提出了基于时空轨迹挖掘地理区域关联性的方法,并用于地理位置预测。此后,时空轨迹频繁模式挖掘问题被广泛研究。与时空轨迹的频繁模式研究不同,由于语义轨迹的频繁模式中融合了地理空间、时间和语义等信息,因而语义轨迹的模式定义和挖掘算法设计变得更复杂。语义轨迹频繁模式研究需要解决的核心问题是,如何表示模式定义中的语义信息,以及如何与时空信息相结合的

问题。在此方面,研究人员开展了一系列工作。Ying等^[55]挖掘语义轨迹中“地理-语义”的频繁模式进行位置预测。Ying等^[56]基于先前的研究工作在频繁模式挖掘中引入时间信息,挖掘语义轨迹中“地理-时间-语义”的频繁模式,预测用户的移动行为,显著提高了预测的准确性。Zhang等^[53]提出算法Splitter用于挖掘语义轨迹中的细粒度序列模式,该算法将位置的语义维连同空间和时间维度一起加入到细粒度序列模式挖掘中。细粒度序列模式是指满足空间紧凑性、语义一致性、时间连续性的模式。该序列模式挖掘分为2步:①检索出一组轨迹片段支持的空间粗模式;②自顶向下细分每个粗模式,从密集紧凑的轨迹片段集中得到细粒度序列模式。Choi等^[57]为了发现局部区域中人们的移动行为规律挖掘语义轨迹中区域序列模式。该模式是指在局部空间中频繁但全局空间内不一定频繁的语义轨迹模式。Wu等^[58]发现传统的语义轨迹频繁模式研究中很少涉及时间约束,或者时间约束的要求又过于严苛。这造成挖掘得到的频繁模式缺少灵活性,适用性不高。针对此问题提出了近似到达时间约束下的语义轨迹频繁模式挖掘问题。该研究工作满足了用户对频繁模式在到达时间方面的要求。Wan等^[59]提出SMOPAT方法用于挖掘私家车轨迹数据中的频繁模式,用于研究和分析城市个体的出行目的。

4.1.2 语义轨迹周期模式挖掘算法

轨迹周期模式的研究工作最早来自于序列周期模式的研究。自从1998年Coward等^[60]研究生物遗传基因序列中的周期模式开始,学术界针对序列的周期模式开展了广泛的研究工作,提出了完全周期模式、部分周期模式、完美周期模式、不完美周期模式、同步周期模式和异步周期模式等各种周期模式^[61]。此后,学术界开始关注时空环境下的移动周期模式。由于移动对象的运动环境复杂,轨迹形态多样,并且空间跨度较大,无法借鉴序列的周期模式定义和方法用于轨迹周期模式挖掘中。

周期模式挖掘的主要挑战在于周期行为的不确定性。无论是个体还是群体的周期行为都不可能严格按照准确时刻进行对齐,普遍存在偏移、失真、缺失等情况。此外,时间周期具有多粒度特性(如,日、周、月、季、年),针对不同的挖掘任务如何选择合适的周期粒度也是周期模式挖掘研究中的难点。

目前,周期模式挖掘可以分为周期已知和周期

未知两种研究情况。周期已知模式的研究首先设定移动行为的时间周期,然后依据此时间约束将时空轨迹的空间位置转化成地理空间区域的序列,从中挖掘周期性模式。Mamoulis等^[62]定义了基于区域的周期性模式挖掘问题,用于分析具有时间规律的移动行为。在设定时间周期之后,将每个周期的轨迹转换成区域序列,从区域序列集中挖掘周期模式。Cao等^[63]在Mamoulis研究的基础上利用区域序列挖掘带有移位或失真的周期性模式,利用自下而上和自上而下2种策略提取轨迹数据中的长周期模式。Yuan等^[64]提出MPAD方法挖掘移动对象的多粒度周期性活动。其中,移动对象的活动是指由移动对象、具有语义含义的热点区域、起止时间构成的序列。研究人员引入多粒度模型描述移动对象的时空信息,利用空间优先和时间优先的算法将对象活动的时空信息转化成不同粒度的数据,由此从原始轨迹数据中挖掘移动对象不同粒度的周期模式,发现移动用户的行为习惯。

另外,周期未知模式的研究在不预设时间周期的前提下检测移动行为的周期性。Li等^[65]研究混合周期模式的挖掘问题,即在相同地点存在多个周期模式并存的情况。研究人员提出了一种基于概率模型的周期检测方法MPDA。该方法采用核密度方法从移动对象的轨迹中发现高密度区域,然后生成访问区域的时间序列,最后采用“过滤-精炼”(filter-refine)框架检测混合周期性。Yuan等^[66]为了研究社交媒体用户的周期性移动行为,提出了一种无参数的贝叶斯模型方法PRED对不同区域间的周期性行为进行建模。该方法无需先验知识(区域数量、规模和周期时长等信息),并且对噪声数据低敏感。Yi等^[67]研究发现,基于位置检测的周期模式挖掘方法容易受到周期行为“波动”的影响导致检测效果不佳。于是,研究人员提出基于“行为意图”的周期模式挖掘模型MIRE,通过构建的行为意图序列发现周期性模式。

上述研究普遍具有一个共同特点,无论是周期已知还是周期未知的模式挖掘大多面向单个移动对象进行研究,而面向群体移动行为的周期模式挖掘研究相对较少。群体移动行为的多样性以及无法长期跟踪的特点为挖掘群体周期性模式带来了挑战。

4.2 语义轨迹聚类算法

轨迹的聚类研究开展得比较早,主要应用于群体移动模式的挖掘研究。由Laube等^[2]提出了伴随

模式flock。随后学术界在此方向上开展了大量研究工作,相继提出了convoy、swarm、traveling、companions。此外,还提出了聚集模式,包括gathering、snowball、evolving group。目前,基于时空轨迹的群体移动模式的研究工作已相当成熟。研究人员针对群体运动形态的特点设计了各种群体移动模式,并且提出了相应的挖掘算法。但是,此类研究有一个共同的缺点,挖掘得到的轨迹模式主要是由时空轨迹构成的簇,缺乏语义信息,解释性差。而语义轨迹的聚类包含了必要的语义信息,可以较好地弥补此方面的不足。

与传统的轨迹聚类研究相同,语义轨迹度量是此类研究的核心问题。由于时间、空间、语义信息的量纲不同,因而采用简单距离函数定义的度量方法无法实现令人满意的聚类效果。所以,根据挖掘任务的应用需求,从时间、空间、语义信息逻辑关系层面出发设计语义轨迹的聚类算法,成为了主要的研究思路。Furtado等^[68]提出了语义轨迹的多维相似性度量方法MSM。此方法以语义轨迹的时间、空间、语义属性的相似性为基础,计算语义轨迹的全局相似性。MSM方法需要度量多维轨迹序列的相似性,同时计算2条轨迹中所有元素的最佳匹配得分,在综合考虑所有元素的匹配情况之后计算语义轨迹的整体相似性。郭黎敏等^[69]提出了基于语义轨迹层次聚类的行为模式挖掘方法。首先挖掘每个移动对象的频繁语义行为模式,然后定义语义行为模式之间的相似性度量方法,最后采用层次聚类的方法对移动对象进行聚类,找出具有相似行为模式的移动对象群体。Choi等^[57]为了发现局部区域中群体的频繁移动行为,提出了基于语义轨迹的DBSCAN聚类算法,该方法的聚类对象不是传统的数据点,而是语义轨迹在地理空间中折线形状的线段序列。Cai等^[70]利用带位置属性的照片数据研究群体的共同移动行为,采用语义标注技术结合城市、时间、天气等语义信息将原始轨迹转换为语义轨迹,提出多维语义轨迹的相似性度量方法,并且采用基于OPTICS的聚类算法对相似语义轨迹进行分组。

4.3 语义轨迹分类方法

语义轨迹分类问题主要是,基于移动对象的特征构建分类模型,预测空间位置或者识别对象类别等。目前,大多数研究都是从原始轨迹中提取移动行为的速度、加速度、方向、路网环境等语义信息作为分类特征,然后构建分类模型用于交通模式识

别、行为活动分析等应用。

时空轨迹分类研究主要面临2个挑战:①缺少带标签的轨迹训练数据。识别轨迹中的运动状态往往需要人为参与,这造成标签标识过程效率低下,无法获得大量足够的训练数据。所以如何建立合理的轨迹分类模型以应对轨迹标签缺少的情况是此分类研究的挑战之一;②如何综合移动行为中的不同特征提高分类性能是此类研究的另一个挑战。由于运动行为复杂性,往往单一特征无法准确识别,如速度、路线等。因而综合多方面、多类型、多模态的特征是语义轨迹分类器设计的关键。

Patterson等^[71]较早地开展了基于移动数据识别交通模式的研究工作,构建结合地图、交通常识等语义信息的粒子过滤器分类模型,识别步行、驾车和乘坐公共汽车3种交通模式。此后,Stenneth等^[72]充分利用移动设备产生的GPS数据和交通网络的环境信息(包括实时公交车的位置、轨道、车站等信息)构造分类特征,识别6种不同的交通模式,具体包括汽车、公交车、轻轨、步行、自行车和静止。Biljecki等^[73]针对一条轨迹中可能混合多种交通模式的情况,采用“分段—识别”的计算框架解决此问题。该方法通过检测轨迹中交通模式可能转换的拐点分割时空轨迹,然后逐一识别每段子轨迹的交通模式,轨迹段如果相邻并且交通模式相同可以进行合并。采用此方法可以实现十种不同交通模式的识别,并且还解决了轨迹中部分片段数据缺失的交通模式识别问题。

除了交通模式识别之外,研究人员还利用语义轨迹分类方法开展活动行为分析方面的研究工作。Yin等^[74]为了发现用户行为目的提出了基于动态贝叶斯网络的推断模型。该模型分为3层:底层为原始无线局域网信号数据,可转换得到原始轨迹;中间层为原始轨迹对应的地理位置列表;顶层对应于用户的活动。该方法采用动态贝叶斯网络从活动序列中推断行为意图。Boukhechba等^[75]研究实时应用场景的活动识别问题,采用增量式的分析方法从原始GPS轨迹中提取带语义信息的空间位置,然后推断用户的活动类型。

4.4 语义轨迹异常检测算法

与其他轨迹挖掘研究一样,最早的研究工作是关于时空轨迹的异常检测问题,即基于轨迹的空间属性、时间属性、几何形状等进行距离度量,发现异常的轨迹^[76]或者轨迹的异常部分^[77]。然而,语义轨

迹的异常检测研究更关注移动对象异常行为的检测问题。目前,主要研究思路是从原始轨迹中提取速度、加速度、方向等运动属性,结合领域规则和应用知识判断异常行为。因而,如何针对特定应用领域定义异常是研究该问题的关键。在此方面,研究人员结合实际应用场景开展了相关研究工作。Lei^[78]研究船舶的异常行为检测问题。他提出了用于海事轨迹建模和异常检测的计算框架MT-MAD,该方法充分考虑到异常行为常出现在非正常海域,或者轨迹序列中的非正常环节,因此通过识别空间、序列、行为3种异常特征构建异常行为检测模型,利用海事自动检测系统(Automatic Identification System, AIS)生成的真实数据验证了系统的有效性和可行性。

5 语义轨迹分析挖掘系统

语义轨迹分析挖掘应用大多集中在车辆、船舶等交通工具的管理领域。例如,Pu等^[79]提出一种交互式轨迹分析系统T-Watcher,用于监测与分析城市的交通状况。该系统从3个层次视图(区域、道路、车辆)交互式地检测并分析车辆轨迹数据,挖掘车辆的移动模式,从而为交通管理者提供可靠的决策支持。Wang等^[80]也提出一种基于轨迹的交互式交通评估系统。该系统基于路网的动态轨迹查询模型实现轨迹的拓扑查询,结合交通领域知识分析交通状况。Patroumpas等^[10,81]设计并开发了一个海事监管系统,实时获取船只的轨迹数据,采用语义轨迹处理、流数据分析、复杂事件处理等技术监控并分析船只运行状况,检测船只异常行为,及时向海事管理部门提供全面、准确的决策支持服务。

此外,也有研究人员开发了通用的语义轨迹挖掘平台。例如,Bogorny等^[82]提出用于语义轨迹挖掘的软件体系结构Weka-STPM,这是第一个涵盖完整轨迹挖掘过程的语义轨迹分析工具。该系统支持数据预处理、语义富化处理、轨迹挖掘和可视化等功能,并且可以帮助用户自动或半自动地执行完整的数据挖掘过程。

6 结论与展望

语义轨迹是通过融合时空轨迹与语义信息得到的。对于语义轨迹的挖掘可以更深入地发现个体或者群体移动行为的意图、习惯、情感等高阶语

义内容,从而深层次发现个体或群体的移动行为模式、关系和规律,使人们可以深度认知移动行为、社交行为,甚至周遭地理环境的变动和演化。

语义轨迹研究经过近10年的发展取得了很大成果。在建模方面,形成了完整的基于本体知识库的语义轨迹建模方法,深度结合领域知识和实际应用,为移动对象在行为和活动方面的分析研究提供了有力的支持。在挖掘方面,语义轨迹在模式挖掘、聚类、分类和异常检测等经典的数据挖掘研究维度上都提出了相应的算法,并且在交通、海事等实际应用中,语义轨迹挖掘都提供了有力的方法指导和技术保证。尽管如此,在语义轨迹挖掘中挑战与机遇并存,本节将从研究挑战和研究趋势两方面对语义轨迹挖掘研究进一步阐述。

6.1 面临的挑战

目前语义轨迹挖掘研究存在的挑战如下。

(1)缺少统一的语义轨迹模型。语义轨迹中的语义信息来源丰富多样,包括移动对象的属性、应用场景、环境、领域知识等。但这些语义信息形式多样、尺度不一、关系复杂,无法采用统一的建模方法进行处理和标注。即使可以采用本体技术改善这一过程,但由于构建本体的代价和复杂性过高,因而导致多源语义信息的处理、标注、融合的困难很大。不利于后续语义轨迹的建模和挖掘工作。

(2)缺乏标准、公认、高质量的适合语义轨迹的语义信息。随着互联网的发展和成熟,产生了众多基于Web的开源知识库,如维基百科、百度百科、Wikidata等。但是,面向地理空间并且可以提供地理服务的开源知识库较少,这极大影响了语义轨迹的建模工作,这也是目前语义轨迹研究遇到了重要挑战。

(3)语义轨迹丰富性的需求和信息融合的实际困难之间存在矛盾。语义轨迹中的时间属性、空间属性和语义属性在量纲、形式、质量等方面均不相同,无法直接地进行融合处理,因而为后续的分析挖掘工作带来了挑战。例如,在语义轨迹的聚类研究中,面对不同类型的属性,如何定义一个兼顾语义属性和非语义属性要求的相似性度量方法是此类型研究中的重要研究问题。

(3)缺乏成熟的语义轨迹挖掘理论框架。与传统的时空轨迹数据挖掘相比,语义轨迹挖掘研究还远未成熟。现有的语义轨迹挖掘方法主要沿袭了传统时空轨迹挖掘的思路和方法。因而还缺乏一个令人满意的语义轨迹挖掘的理论框架,需要解决

一系列问题,具体包括:待挖掘的语义轨迹模式的种类、模式分类方法以及与之适合的挖掘算法等。未来构建语义轨迹挖掘理论框架可以更好地理解语义轨迹模式所具备的表达能力,同时有助于提出更有效的实践应用技术。

6.2 未来的研究趋势

在语义轨迹建模方面,未来的研究工作将侧重以下4点:

(1)新的语义轨迹定义方式。在现有的语义轨迹定义中,基于轨迹片段的多层语义轨迹定义方法涵盖的信息最丰富、最完整、最系统。但是,语义信息在多粒度表示方面明显不足,无法支持高阶、复杂的多层次语义信息表达。未来研究中,应该引入层次结构,丰富现有的语义信息组织方式。这样才能有效支撑未来多粒度的语义轨迹挖掘研究。

(2)面向新型数据的标注方法。为语义轨迹标注方法扩展新的数据来源。近年来,智能视频监控技术快速发展,视频覆盖的人群和范围日益扩大,使得监控视频成为了新的移动数据来源。公共场所的监控视频中不仅包含移动个体/群体的轨迹数据,还包含各种行为数据、空间拓扑关系、活动信息等。这对传统的语义轨迹标注技术提出了挑战,也带来新的机遇。

(3)基于知识图谱的语义标注方法。现有的语义轨迹标注方法主要使用本体知识库获得完整、系统、准确、丰富的语义信息。但是,此类方法最主要的问题是知识库的构建门槛较高,需要领域专家的参与,并且需要被学术界和工业界广泛认可。近年来,知识图谱的技术已经日趋成熟。它可以建立具有语义处理能力与开放互联能力的知识库。但与传统知识库不同的是,知识图谱可以利用互联网上的开放数据、公开信息以及链接关系聚集成为知识,使得信息资源易于计算、理解和评价。与语义轨迹建模方法结合,可以极大丰富语义轨迹的信息来源,增强移动行为的语义化的解释与表达,并且语义信息可持续更新。

(4)基于多源数据的建模方法。由于基于单一移动数据的轨迹挖掘研究中常存在数据代表性问题,因而学术界越来越重视多源移动数据的挖掘研究。但是,不同的移动对象所处的应用场景和地理环境差异性较大,采用现有的建模方法生成的语义轨迹在数据模型、粒度和表达形式方面都可能存在差异。因而不便于对多源移动数据进行融合处理

和分析挖掘。未来可以结合本体对齐、实体链接、冲突解决、关系推演等数据融合技术提出适合多源移动数据的语义轨迹建模方法。

在语义轨迹挖掘技术方面,未来研究工作将侧重以下7点。

(1)语义轨迹的管理问题。由于语义信息的加入,语义轨迹的数据规模、复杂程度、管理难度远远大于传统的时空轨迹。这对基于语义轨迹的数据挖掘应用实践带来了巨大挑战。为了应对实际的数据管理需求,未来研究可以从模型定义、预处理、存储与索引和查询分析处理等方面对语义轨迹的管理问题开展研究。其中,尤其包括语义轨迹的分布式存储、相似性度量、近似查询和多维索引等热点技术问题。

(2)语义轨迹的分类和预测研究。现有的语义轨迹预测仍然主要聚焦于时空层面的研究问题,如位置预测、轨迹预测、路线预测、目的地预测等。在分类和预测方法上,现有的语义轨迹主要以时空特征为主而以语义特征为辅构建分类模型来预测群体/个体的移动行为。未来应该着重针对移动行为特征在表达、描述、提取等问题上开展研究,聚焦移动行为层面的研究问题,如移动行为的意图、习惯、情感等。

(3)面向语义轨迹的流式挖掘处理。研究面向实时应用场景的语义轨迹挖掘方法。语义轨迹挖掘研究可以应用于人类移动性、城市动态性等研究工作中。但随着城市中大量且频繁的聚集活动对城市管理水平和综合治理能力提出的严峻考验,依靠传统的离线挖掘方法已经无法应对实时应用场景的需要。因而,未来研究工作应该更加关注于实时应用中的语义轨迹生成、处理、挖掘的方法,满足实时挖掘算法在时间和空间复杂度方面的更高要求。

(4)隐私保护

研究隐私保护的语义轨迹数据挖掘方法。语义轨迹中的语义信息比时空属性更容易暴露移动对象的隐私。因而,隐私保护问题是语义轨迹数据建模与挖掘过程中的一个不可回避的问题,现有的挖掘方法中较少研究,应该成为语义轨迹未来重要的研究方向之一。

(5)多粒度数据挖掘

语义轨迹中包含的时间粒度、空间粒度和语义

粒度对挖掘结果有直接影响。多粒度的语义轨迹挖掘方法在不同粒度水平上可以发现不同的感兴趣模式,并且自适应地支持多层次、多尺度、多维度分析。例如,在城市间、城市和亚城市3个不同尺度上,人群移动模式各有不同,因而反映出在人地关系方面的移动规律也不一样。但是,当前多粒度的语义轨迹挖掘研究开展得较少,这应该是未来一个重要的研究方向。

(6)评测指标与数据集

评测方法是现有数据挖掘领域中的重要研究内容。它是验证挖掘算法性能的主要手段。现有的语义轨迹挖掘中,还未出现符合语义轨迹特点的标准评测指标及其数据集。目前,大部分方法仍然借用传统数据挖掘中的评测指标。例如,部分语义轨迹聚类研究沿用传统聚类算法的评测指标。但由于引入了语义信息,挖掘结果具有很强的模糊性和不确定性,因而传统的确定性的评价指标无法有效评估语义挖掘算法的性能。在未来研究中,应该考虑符合语义轨迹数据特点和应用需求的评测指标及其标准数据集。

(7)与应用领域结合

广泛结合实际应用和领域。语义轨迹挖掘研究不同于传统的时空轨迹挖掘研究之处在于语义性方面。而语义性并非单一的,而是多样性的。语义轨迹挖掘应该结合实际应用和领域知识,这样才能充分体现语义的优势,解决传统轨迹挖掘的不足。

参考文献(References):

- [1] 许佳捷,郑凯,池明旻,等.轨迹大数据:数据、应用与技术现状[J].通信学报,2015,36(12):97-105. [Xu J J, Zheng K, Chi M M, et al. Trajectory big data: Data, applications and techniques[J]. Journal on Communications, 2016,36(12):97-105.]
- [2] 吉根林,孙鸿艳,赵斌.时空轨迹群体运动模式挖掘研究进展[J].南京航空航天大学学报,2016,48(5):615-624.[Ji G L, Sun H Y, Zhao B. Research progress in group moving patterns mining of spatio-temporal trajectories[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2016,48(5):615-624.]
- [3] 龚玺,裴韬,孙嘉,等.时空轨迹聚类方法研究进展[J].地理科学进展,2011,30(5):522-534. [Gong X, Pei T, Sun J, et al. Review of the research progresses in trajectory clustering methods[J]. Progress in Geography, 2011,30(5):522-534.]
- [4] 牟乃夏,徐玉静,张恒才,等.移动轨迹聚类方法研究综述

- [J].测绘通报,2018,1(1):1-7. [Mou N X, Xu Y J, Zhang H C, et al. A review of the mobile trajectory clustering methods[J]. Bulletin of Surveying and Mapping, 2018,1(1):1-7.]
- [5] 赵竹珺,吉根林.时空轨迹分类研究进展[J].地球信息科学学报,2017,19(3):289-297. [Zhao Z J, Jin G L. Research progress of spatial-temporal trajectory classification[J]. Journal of Geo-information Science, 2017,19(3): 289-297.]
- [6] 毛嘉莉,金澈清,章志刚,等.轨迹大数据异常检测:研究进展及系统框架[J].软件学报,2017,28(1):17-34. [Mao J L, Jin C Q, Zhang Z G, et al. Anomaly detection for trajectory big data: advancements and framework[J]. Journal of Software, 2017,28(1):17-34.]
- [7] 齐凌艳,陈荣国,温馨.基于语义轨迹停留点的位置服务匹配与应用研究[J].地球信息科学学报,2014,16(5):720-726. [Qin L Y, Chen R G, Wen X. Research on the LBS matching based on stay point of the semantic trajectory [J]. Journal of Geo-information Science, 2014,16(5):720-726.]
- [8] Yao D, Zhang C, Huang J H, et al. SERM: A recurrent model for next location prediction in semantic trajectories [C]. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore: ACM, 2017:2411-2414.
- [9] Zheng Y, Liu F, Hsieh H P. U-air: When urban air quality inference meets big data[C]. The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago: ACM, 2013:1436-1444.
- [10] Patroumpas K, Artikis A, Katzouris N, et al. Event recognition for maritime surveillance [C]. Proceedings of the 18th International Conference on Extending Database Technology. Brussels: OpenProceedings.org, 2015:629-640.
- [11] Yan Z X. Towards semantic trajectory data analysis: A conceptual and computational approach[C]. Proceedings of the VLDB 2009 PhD Workshop. Lyon: VLDB Endowment, 2009.
- [12] 吉根林,赵斌.时空轨迹大数据模式挖掘研究进展[J].数据采集与处理,2015,30(1):47-58. [Jin G L, Zhao B. Research progress in pattern mining for big spatio-temporal trajectories[J]. Journal of Data Acquisition and Processing, 2015,30(1):47-58.]
- [13] Mountain D, Raper J. Modelling human spatio-temporal behavior: a challenge for location-based services[C]. Proceedings of the 6th International Conference on GeoComputation. Australia: Springer, 2001.
- [14] Spaccapietra S, Parent C, Damiani M L, et al. A conceptual view on trajectories[J]. Data & Knowledge Engineering, 2008,65(1):126-146.
- [15] Baglioni M, Macêdo J, Renso C, et al. An ontology-based approach for the semantic modelling and reasoning on trajectories[C]. Proceeding in Advances in Conceptual Modelling-Challenges and Opportunities, ER Workshops. Barcelona Spain: Springer, 2008:344-353.
- [16] Yan Z X, Jose M, Christine P, et al. Trajectory ontologies and queries[J]. Transactions in GIS, 2010,12(s1):75-91.
- [17] Baglioni M, Macêdo J, Renso C, et al. Towards semantic interpretation of movement behavior[C]. Proceedings of the 12th AGILE Conference Advances in GIScience. Hannover: Springer, 2009:271-288.
- [18] Renso C, Baglioni M, Macêdo J, et al. How you move reveals who you are: Understanding human behavior by analyzing trajectory data[J]. Knowledge and Information Systems, 2013,37(2):331-362.
- [19] Nogueira P T, Braga R B, Martin H. An ontology-based approach to represent trajectory characteristics[C]. The 5th International Conference on Computing for Geospatial Research and Application. USA: IEEE, 2014:102-107.
- [20] Vandecasteele A, Devillers R, Napoli A. From movement data to objects behavior using semantic trajectory and semantic events[J]. Marine Geodesy, 2014,37(2):126-144.
- [21] Xu J Q, Ralf H G. A generic data model for moving objects[J]. GeoInformatica, 2013,17(1):125-172.
- [22] Boulmakoul A, Karim L, Bouziri A E, et al. A system architecture for heterogeneous moving-object trajectory metamodel using generic sensors: Tracking airport security case study[J]. IEEE Systems Journal, 2015,9(1):283-291.
- [23] Manaa M, Akaichi J. Ontology-based modeling and querying of trajectory data[J]. Data & Knowledge Engineering, 2017,111:58-72.
- [24] Fileto R, Krüger M, Pelekis N, et al. Baquara: A holistic ontological framework for movement analysis using linked data[C]. The 32th International Conference on Conceptual Modeling, ER 2013. Hong-Kong: Springer, 2013: 342-355.
- [25] Nogueira T P, Martin H. Querying semantic trajectory episodes[C]. Proceedings of the 4th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems. Bellevue: ACM, 2015:23-30.
- [26] Guc B, May M, Saygin Y, et al. Semantic annotation of GPS trajectories[C]. The 11th AGILE International Conference on Geographic Information Science. Spain: Springer, 2008:1-9.
- [27] Fileto R, May C, Renso C, et al. The baquara2 knowledge-based framework for semantic enrichment and analy-

- sis of movement data[J]. *Data & Knowledge Engineering*, 2015,98:104-122.
- [28] Nogueira T P, Braga R B, Oliveira C T, et al. FrameSTEP: A framework for annotating semantic trajectories based on episodes[J]. *Expert Systems with Applications*, 2018,92:533-545.
- [29] Hage W R, Malaisé V, Vries G, et al. Abstracting and reasoning over ship trajectories and web data with the simple event model (SEM) [J]. *Multimedia Tools and Applications*, 2012,57(1):175-197.
- [30] Douglas D H, Peucker T. Algorithms for the reduction of the number of points required to represent a line or its caricature[J]. *The Canadian Cartographer*, 1973,10(2):112-122.
- [31] Hershberger J, Snoeyink J. Speeding up the Douglas-Peucker Line-Simplification algorithm[R]. *Proceedings of International Symposium on Spatial Data Handling*, 2000: 134-143.
- [32] Zheng Y, Li Q N, Chen Y K, et al. Understanding mobility based on GPS data[C]. *The 10th International Conference on Ubiquitous Computing*. Seoul: ACM, 2008:312-321.
- [33] Wang Y, McArthur D. Enhancing data privacy with semantic trajectories: a raster-based framework for GPS stop/move management[J]. *Transactions in GIS*, 2018,22(4):975-990.
- [34] Ashbrook D, Starner T. Using GPS to learn significant locations and predict movement across multiple users[J]. *Personal and Ubiquitous Computing*, 2003,7(5):275-286.
- [35] Krumm J, Horvitz E. Predestination: Inferring destinations from partial trajectories[C]. *The 8th International Conference on Ubiquitous Computing*. Orange County: ACM, 2006:243-260.
- [36] Andrienko G L, Andrienko N V, Wrobel S. Visual analytics tools for analysis of movement data[J]. *SIGKDD Explorations*, 2007,9(2):38-46.
- [37] Zheng Y, Zhang L Z, Ma Z X, et al. Recommending friends and locations based on individual location history [J]. *ACM Transaction on the Web*, 2011,5(1):1-44.
- [38] Palma A T, Bogorny V, Kuijpers B, et al. A clustering-based approach for discovering interesting places in trajectories[C]. *Proceedings of the 2008 ACM Symposium on Applied Computing*. Fortaleza: ACM, 2008:863-868.
- [39] Rocha J, Times V C, Oliveira G, et al. Db-SMoT: A direction-based spatio-temporal clustering method[C]. *The 5th IEEE International Conference on Intelligent Systems*. London: IEEE, 2010:114-119.
- [40] Alvares L O, Bogorny V, Kuijpers B, et al. A model for enriching trajectories with semantic geographical information [C]. *The 15th ACM International Symposium on Geographic Information Systems*. Washington: ACM, 2007:22.
- [41] Liao L, Fox D, Kautz H A. Location-based activity recognition using relational Markov networks[C]. *Proceedings of the 9th International Joint Conference on Artificial Intelligence*. Edinburgh: Professional Book Center, 2005: 773-778.
- [42] Zheng Y, Chen Y K, Li Q N, et al. Understanding transportation modes based on GPS data for web applications [J]. *ACM Transactions on the Web*, 2010,4(1):1-36.
- [43] Yan Z X, Chakraborty D, Parent C, et al. SeMiTri: A framework for semantic annotation of heterogeneous trajectories[C]. *The 14th International Conference on Extending Database Technology*. Sweden: ACM, 2011:259-270.
- [44] Yan Z X, Chakraborty D, Parent C, et al. Semantic trajectories: Mobility data computation and annotation[J]. *ACM Transaction on Intelligent Systems and Technology*, 2013,4(3):49(1-38).
- [45] Parent C, Spaccapietra S, Renso C, et al. Semantic trajectories modeling and analysis[J]. *ACM Computing Surveys*, 2013,45(4):42(1-32).
- [46] Wu F, Li Z H, Lee W C, et al. Semantic annotation of mobility data using social media[C]. *Proceedings of the 24th International Conference on World Wide Web*. Switzerland: ACM, 2015:1253-1263.
- [47] He T, Yin H, Chen Z Y, et al. A spatial-temporal topic model for the semantic annotation of POIs in LBSNs [J]. *ACM Transaction on Intelligent Systems and Technology*, 2016,8(1):12(1-24).
- [48] Soares A, Rose J, Etemad M, et al. VISTA: A visual analytics platform for semantic annotation of trajectories[C]. *The 22th International Conference on Extending Database Technology*. Lisbon: OpenProceedings.org, 2019:570-573.
- [49] Hu Y J, Janowicz K, Carral D, et al. A geo-ontology design pattern for semantic trajectories[C]. *The 11th International Conference on Spatial Information Theory*. Scarborough: Springer, 2013:438-456.
- [50] Wannous R, Malki J, Bouju A, et al. Modelling mobile object activities based on trajectory ontology rules considering spatial relationship rules[C]. *Modeling Approaches and Algorithms for Advanced Computer Applications*, 2013:249-258.
- [51] Nogueira T P, Martin H. Qualitative representation of dynamic attributes of trajectories[C]. *Agile Conference on Geographic Information Science*, 2014.
- [52] Pei J, Han J W, Mortazavi-Asl B, et al. PrefixSpan: Mining sequential patterns by prefix-projected growth[C]. *Proceedings of the 17th International Conference on Data Engineering*. Heidelberg: IEEE Computer Society, 2001:

- 251-224.
- [53] Zhang C, Han J W, Shou L D, et al. Splitter: Mining fine-grained sequential patterns in semantic trajectories[J]. *PVLDB*, 2014,7(9):769-780.
- [54] Monreale A, Pinelli F, Trasarti R, et al. Where next: A location predictor on trajectory pattern mining[C]. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris: ACM, 2009:637-646.
- [55] Ying J J C, Lee W C, Weng T C, et al. Semantic trajectory mining for location prediction[C]. *The 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*. Chicago: ACM, 2011:34-43.
- [56] Ying J J C, Lee W C, Tseng V S. Mining geographic-temporal-semantic patterns in trajectories for location prediction[J]. *ACM TIST*, 2013,5(1):2(1-33).
- [57] Choi D W, Pei J, Heinis T. Efficient mining of regional movement patterns in semantic trajectories[J]. *PVLDB*, 2017,10(13):2073-2084.
- [58] 吴瑕,唐祖锴,祝园园. 近似到达时间约束下的语义轨迹频繁模式挖掘[J]. *软件学报*, 2018,29(10):3184-3204. [Wu X, Tang Z K, Zhu Y Y. Frequent pattern mining with approximate arrival-time in semantic trajectories[J]. *Journal of Software*, 2018,29(10):3184-3204.]
- [59] Wan C C, Zhu Y M, Yu J D, et al. SMOPAT: mining semantic mobility patterns from trajectories of private vehicles[J]. *Information Sciences*, 2018,429:12-25.
- [60] Coward E, Drablos F. Detecting periodic patterns in biological sequences[J]. *Bioinformatics*, 1998,14(6):498-507.
- [61] 高强,张凤荔,王瑞锦,等. 轨迹大数据: 数据处理关键技术研究综述[J]. *软件学报*, 2017,28(4):959-992. [Gao Q, Zhang F L, Wang R J, et al. Trajectory big data: A review of key technologies in data processing[J]. *Journal of Software*, 2017,28(4):959-992.]
- [62] Mamoulis N, Cao H P, Kollios G, et al. Mining, indexing, and querying historical spatiotemporal data[C]. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington: ACM, 2004:236-245.
- [63] Cao H P, Mamoulis N, Cheung D W. Discovery of periodic patterns in spatiotemporal sequences[J]. *IEEE Transactions on Knowledge Data Engineering*, 2007,19(4):453-467.
- [64] Yuan G, Zhao J, Xia S X, et al. Multi-granularity periodic activity discovery for moving objects[J]. *International Journal of Geographical Information Science*, 2017,31(3):435-462.
- [65] Li J, Wang J J, Zhang J F, et al. A probabilistic approach to detect mixed periodic patterns from moving object data [J]. *GeoInformatica*, 2016,20(4):715-739.
- [66] Yuan Q, Zhang W, Zhang C, et al. PRED: Periodic region detection for mobility modeling of social media users [C]. *Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. Cambridge: ACM, 2017:263-272.
- [67] Yi F, Yin L B, Wen H, et al. Mining human periodic behaviors using mobility intention and relative entropy[C]. *The 22nd Pacific-Asia Conference Advances in Knowledge Discovery and Data Mining*. Melbourne: Springer, 2018:488-499.
- [68] Furtado A S, Kopanaki D, Alvares L O, et al. Multidimensional similarity measuring for semantic trajectories[J]. *Transactions in GIS*, 2016,20(2):280-298.
- [69] 郭黎敏,高需,武斌,等. 基于停留时间的语义行为模式挖掘[J]. *计算机研究与发展*, 2017,54(1):111-122. [Guo L M, Gao X, Wu B, et al. Discovering common behavior using staying duration on semantic trajectory [J]. *Journal of Computer Research and Development*, 2017,54(1):111-122.]
- [70] Cai G C, Lee K, Lee I. Mining mobility patterns from geotagged photos through semantic trajectory clustering [J]. *Cybernetics and Systems*, 2018,49(4):234-256.
- [71] Patterson D J, Liao L, Fox D, et al. Inferring high-level behavior from low-level sensors. *The 5th International Conference on Ubiquitous Computing*. Seattle: Springer, 2003:73-89.
- [72] Stenneth L, Wolfson O, Yu P S, et al. Transportation mode detection using mobile phones and GIS information [C]. *The 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*. Chicago: ACM, 2011:54-63.
- [73] Biljecki F, Ledoux H, Oosterom P V. Transportation mode-based segmentation and classification of movement trajectories [J]. *International Journal of Geographical Information Science*, 2013,27(2):385-407.
- [74] Yin J, Chai X Y, Yang Q. High-level goal recognition in a wireless LAN[C]. *Proceedings of the 19th National Conference on Artificial Intelligence, 6th Conference on Innovative Applications of Artificial Intelligence*. California: AAAI Press, 2004:578-584.
- [75] Boukhechba M, Bouzouane A, Bouchard B, et al. Online recognition of people's activities from raw GPS data: Semantic trajectory data analysis[C]. *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*. Greece: ACM, 2015:40(1-8).
- [76] Knorr E M, Ng R T, Tucakov V. Distance-based outliers: algorithms and applications[J]. *VLDB Journal*, 2000,8(3-

- 4):237-253.
- [77] Lee j g, Han J W, Li X L. Trajectory outlier detection: A partition-anddetect framework[C]. Proceedings of the 24th International Conference on Data Engineering. Mexico: IEEE Computer Society, 2008:140-149.
- [78] Lei P R. A framework for anomaly detection in maritime trajectory behavior[J]. Knowledge Information Systems, 2016,47(1):189-214.
- [79] Pu J S, Liu S Y, Ding Y, et al. T-watcher: A newvisual analytic system for effective traffic surveillance[C]. The 14th International Conference on Mobile Data Management. Milan: IEEE Computer Society, 2013:127-136.
- [80] Wang F, Chen W, Wu F R, et al. A visual reasoning approach for data-driven transport assessment urban roads [C]. The 2014 IEEE Conference on Visual Analytics Science and Technology. Paris: IEEE Computer Society, 2014:103-112.
- [81] Patroumpas K, Alevizos E, Artikis A, et al. Online event recognition from moving vessel trajectories[J]. GeoInformatica, 2017,21(2):389-427.
- [82] Bogorny V, Avancini H, Paula B C, et al. Weka-STPM: A software architecture and prototype for semantictrajectory data mining and visualization[J]. Transaction in GIS, 2011,15(2):227-248.