

DOI: 10.11992/tis.201804005

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180523.1350.002.html>

旅游路线规划研究综述

常亮, 孙文平, 张伟涛, 宾辰忠, 古天龙

(桂林电子科技大学 广西可信软件重点实验室, 广西 桂林 541004)

摘要: 旅游业的快速发展和用户分享内容的激增使得旅游领域的信息过载问题日益突出, 如何帮助游客在快速制定个性化游览路线的同时提升旅行体验, 成为当前旅游路线规划问题研究的关键。首先, 给出旅游路线规划问题的形式化定义; 然后, 将文献中的旅游路线规划求解方法分为基于精确数学建模的求解、基于用户生成内容的求解两大类, 对各类方法的关键技术和存在的主要问题进行了较为详细的考察; 最后, 给出一个旅游路线规划系统整体架构, 对其中存在的重点和难点问题进行了分析, 为旅游路线规划问题的研究提供理论支持的同时指明了下一步的研究方向。

关键词: 旅游路线规划; 位置服务; 轨迹挖掘; 个性化推荐; 上下文感知; 定向问题; 用户生成数据; 全域旅游
中图分类号: TP301 **文献标志码:** A **文章编号:** 1673-4785(2019)01-0082-11

中文引用格式: 常亮, 孙文平, 张伟涛, 等. 旅游路线规划研究综述[J]. 智能系统学报, 2019, 14(1): 82-92.

英文引用格式: CHANG Liang, SUN Wenping, ZHANG Weitao, et al. Review of tourism route planning[J]. CAAI transactions on intelligent systems, 2019, 14(1): 82-92.

Review of tourism route planning

CHANG Liang, SUN Wenping, ZHANG Weitao, BIN Chenzhong, GU Tianlong

(Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: With the rapid development of tourism and surge in user sharing, information overload in tourism has become an increasing problem. Currently, the key issue in tourism route planning is how to help tourists to develop personalized travel routes and enhance their travel experiences. In this paper, we first provide a formal definition of the tourism route planning problem. Then, we sort the methods used to solve this problem into two categories, i.e., those based on exact mathematical modeling and those based on user-generated content. We then present a detailed survey of the key technologies and difficulties of these methods. Lastly, we propose an overall framework for the tourism route planning system, analyze the key difficulties of the system, provide theoretical support for the study of tourism route planning, and suggest a future research direction.

Keywords: tourism route planning; location-based service; trajectory mining; personalized recommendation; context-aware; orientation problem; user-generated data; all-for-one tourism

在当前稳定的宏观经济和社会环境下, 国民旅游需求不断增加, 旅游活动持续升温, “全域旅游”的发展战略突破了传统景区与景点的资源观念, 延伸到农耕民俗、工业遗产等社会资源, 对旅游业的服务质量也提出了更高的要求。然而, 随

着可选地点的急剧增加, 如何根据用户需求帮助用户快速进行旅行路线规划, 成为全域旅游中亟待解决的难点问题, 使得相关旅游路线规划方法的研究成为当前旅游领域的研究前沿。

目前, 虽然旅行者在进行旅行规划时可以在互联网上方便地查看相关信息, 但仍然需要花费大量时间和精力^[1]。经常出现的现象是, 许多旅行者在事先花费了很多时间制定旅行路线, 但最终却又不得不选择跟团的形式进行旅行, 因此旅

收稿日期: 2018-04-04. 网络出版日期: 2018-05-24.

基金项目: 国家自然科学基金项目 (61572146, U1501252, U1711263); 广西创新驱动重大专项项目 (AA17202024); 广西自然科学基金项目 (2016GXNSFDA380006).

通信作者: 宾辰忠. E-mail: binchenzhong@guet.edu.cn.

行者对于旅行路线规划相关服务的需求日益迫切。

1 旅游路线规划理论

科学的旅游路线规划不仅有助于旅行者根据自己的时间和经费预算制定适合自己的游览路线,还能够提升旅行者的旅行体验,使得旅行者在旅行中有更多的时间和精力放在游览过程中。

在旅游路线规划问题的研究工作中,较早的工作大多集中在利用 OP 问题 (orienteeing problem) 作为基本问题,通过不同的变型对旅游路线规划问题进行建模求解。这类工作的重点是准确建模旅游路线规划问题中的多方面因素,比如用户约束、景点开放时间和出行交通方式等,最终能得到一个或多个满足用户约束的精确路线规划结果^[2]。但是,这类工作无法对现实生活中旅游路线规划问题的各种因素进行完全建模。一方面,由于旅游活动是一个动态的过程,在这个过程中有很多不确定的因素;另一方面,当兴趣点的地理范围较大时,不能再将兴趣点仅仅当作一个点进行建模,如一条观光河道,兴趣点的起点和终点可能相差很远。

随着互联网的飞速发展,与日常信息相关的各类用户生成内容迅速增多。在旅游领域中,形成了多种形式的旅游时空轨迹数据,例如:GPS 轨迹、北斗导航信息、签到记录等。这些数据与用户分享的大量旅游经历和旅行照片等数据,共同形成了旅游大数据。合理地利用这些数据进行旅游路线规划,是近期研究工作的一个热点。这类工作的优点是能够快速得到符合现实情况的可行解,帮助用户进行旅行规划,但难点在于合理利用多源数据准确地挖掘用户的历史行为轨迹^[3]。

2 旅游路线规划问题描述

游客到一个地方进行旅行时通常面临以下两个问题:首先是决定访问哪些景点,从而使自己的旅行变得更加有趣;其次是确定每个旅行日的路线,即确定对每个景点的访问顺序。这个过程需要考虑到多个参数和约束,如门票价格、天气条件等。

基于当前用户在进行旅行规划时所遇到的问题,旅游路线规划问题便应运而生。其实旅游规划问题并不是一个新的问题,最早可追溯到旅行商问题 (traveling salesman problem, TSP)。由于 TSP 属于 NP-Complete 问题,大量研究工作主要集中在如何进行启发式求解。

个性化旅游路线规划问题比 TSP 问题更加复杂。总体上是指游客在对多个兴趣点 (point of in-

terest, POI) 感兴趣的情况下,如何按照游客的相关约束以及对 POI 的兴趣度得到适合的旅行路线^[4]。尽管现阶段互联网中存在大量与旅游相关的信息,但对于一个访问陌生城市的游客来说,这仍然是一项具有挑战性的任务,尤其是其涉及的因素很多,例如每个景点所需的游览时间、景点的开放时间和景点之间的旅行距离等。旅游路线规划问题的关键是:在满足游客时间和花费的约束下选择更多匹配游客偏好的 POI 进行游览,最大化游客的满意度。在进行旅行规划时,要得到高质量的解决方案,除了需要考虑多方面的因素,还需要根据不同标准建立相应的评价模型^[5-6]。

本文将典型的旅游路线规划问题定义为 5 元组 $P = \langle \text{POIs}, \text{TrafficCost}, \text{Profit}, \text{TConstraint}, \text{FConstraint} \rangle$, 其中:

1) POIs 表示所有候选的 POI, 每个 POI 又具有多个属性,包括类型、位置、门票价格、开放时间等;

2) TrafficCost 表示在各个 POI 之间采用各种不同的交通方式所需要的旅行时间和费用,主要的交通方式包括公共交通、骑行、步行等;

3) Profit 表示游客游览每个 POI 所能获得的“收益”,通过对每个 POI 的客观打分以及游客的主观感受进行加权计算而得,其中游客的主观感受又主要取决于游客的个人偏好;

4) TConstraint 表示游客用于该次旅行的时间预算,包括游客该次旅行的总天数以及每天用于游览的时间数等;

5) FConstraint 表示游客用于该次旅行的费用预算。

对于天气状况这种影响旅行的因素,我们将其归类到 POI 的开放时间这个属性中。对于其他未考虑的因素,可以相应地对 5 元组进行扩展。

给定一个旅游路线规划问题之后,对该问题的求解是指找到关于各个 POI 访问日程和访问顺序的一套或多套方案,使得在满足游客的时间预算和费用预算等约束的前提下,游客所能获得的收益达到最大或者最佳。

3 旅游路线规划问题求解方法

目前,相关文献中存在许多对旅游路线规划问题进行求解的方法。本文将这些方法分为两大类:1) 对旅游路线规划问题进行精确的数学建模,通过规划求解得到较为精确的规划方案;2) 利用用户生成内容 (user generated content, UGC) 进行路线挖掘,并结合用户的喜好和相关约束得到一条或多条可行路线^[7]。

3.1 基于精确数学建模的求解

从建模角度进行求解时,关键是对旅游路线规划问题进行精准建模,可以通过对 TSP 模型加入不同参数和约束进行扩展得到不同的求解模型。这类工作又可以按照路线数量分为 2 类: 1) 求解出单条路线,找到能够满足用户旅行约束和用户对 POI 的偏好并且利润最大化的单程旅游线路; 2) 求解出多条路线。

3.1.1 单路线求解方法

旅游路线规划问题的单路线求解,可以利用单一目标旅行商问题增加收益目标进行变型建模,将节点之间的连接与收益和旅行成本相关联。其目标是:在所有节点的子集上找到一条回路,使得收益最大化,同时旅行成本最小化^[8]。

OP 问题是上述模型的一个变型,通常用于寻找在给定旅行花费的条件下使得总收益最大的路线。例如, Souffriau 等^[8]提出了 OP 问题在城市旅游中的应用,给出了一个综合人工智能和元启发式的方法。在已有的关于旅游路线规划的研究中绝大多数使用 OP 及其扩展建模不同变型。

具有时间窗口的 OP 变型是目前的一个研究热点。在该变型中考虑了对于图中的每个节点可能在不同的时间窗口内访问的情况,因此能够在建模时加入兴趣点的开放时间因素。例如, Gunawan 等^[9-10]提出了一种迭代本地搜索算法,使用贪心方法生成初始可行解,基于轮盘选择的方法插入非计划节点。在之后的工作, Gunawan 等^[11]进一步引入模拟退火算法,以较小的概率接受较差的解决方案,在一定程度上解决陷入局部最优的问题。时间窗对传统 OP 问题的性质及其解决算法有很大的影响。然而,因为不同景点可能在开放时间上有所不同,例如,大型灯光演出时间为夜晚,公园开放时间在白天,所以传统 OP 中通过重新对访问点排序来减少旅行时间的方法在这里并不适用。

具有时间依赖的 OP 问题在进行路线规划时将时间因素加入边的代价中,从而可以对旅行中在景点之间采用不同交通方式的情况进行建模。在此基础上, Verbeeck 等^[12]提出了一种基于蚁群算法的快速本地搜索元启发式方法,将蚁群算法的原理与包含时间依赖的本地搜索方法相结合,快速给出有效解决方案。通过实验表明,该算法能够在花费很少计算时间的情况下获得高质量的路线规划结果,保证在出现新的可用交通信息时快速更新路线,帮助游客快速到达目的地。

多目标 OP 问题是 OP 的多目标变型,每个节点(即 POI)可以被分配给不同的类别,例如文化、

历史、休闲、购物等,并且为每个类别提供不同的收益,在不违反最大旅行成本限制的情况下找到所有的高效解决方案。Schilde 等^[13]通过对多目标 OP 问题在城市旅游中的运用进行研究,开发和应用了 2 种用于多目标定向问题的启发式解决方法,这 2 种方法考虑到了每个旅游者在选择和访问兴趣点(例如博物馆、教堂)时对不同的类别可能有不同偏好的情况,使用帕累托蚁群优化算法将可变邻域搜索方法扩展到多目标情况。通过来自现实世界中几个城市的实例对 2 种算法进行了测试,结果表明,2 种方法对解决多目标定向问题都有很好的效果,能够根据不同游客对不同景点的偏好程度设计出使游客满意度最大的个性化旅游路线。

弧定向问题将传统 OP 中的收益不再放在节点中,而是放在边上,其中每条边只能访问一次,用边上的取值来表示景点的得分或者道路的状况。Lu 等^[14]将目标放在寻找风景最优美的路线上而不是距离最短的路线上,将道路网络视为空间网络,利用空间数据领域中的椭圆修剪和空间索引技术,提出了一系列元启发式算法来解决大规模道路网络中快速响应的问题;实验表明,该方法在推荐结果的准确性和效率上都有很大的提升。而在之后的工作中,该作者进一步提出了具有时间依赖的弧定向问题模型,在道路网络的边中同时融合旅行花费和吸引力值,在满足用户时间约束的前提下得到用户最喜欢的路线规划结果^[15]。

通过上述工作可以发现,在建模时考虑到多方面因素能够提高路线规划结果的准确性,如表 1 所示。此外,还有一些 OP 变型可用于建模旅游路线规划问题中更加具体的内容^[16-19],如可能需要多次访问或长时间访问一个 POI,这些变型对于提高具体问题推荐结果的准确性有很大帮助。

表 1 建模因素对比

Table 1 Modeling factors' comparison

文献	开放时间	交通时间	景点类型	道路状况
[10]	√			
[12]		√		
[13]			√	
[14]				√
[15]		√		√

3.1.2 多路线求解方法

用双目标 TSP 求解多路线的扩展变型被称为带利润的车辆路由问题。该问题中,不再是强制

性地访问整个节点集合,而是在访问节点时收集利润,且利润的收集分布在具有有限容量的几辆车上。团队定向问题是该问题的一个变型,多用于旅游路线规划问题的多路线求解,其目标是找到在最大长度限制条件下的 k 条路径(其中每个节点最多访问一次),并且具有最大的总收益。

带时间窗口的团队定向问题中加入了对于 POI 在特定的时间窗口进行访问的限制,从而可以适应更多场景。Lin 等^[20]提出了一个基于模拟退火算法的启发式算法,在每次迭代中,通过对当前解以相等的概率应用移动交换,插入或倒置其中的一个节点来获得相邻解,如果它比当前最佳找到的解决方案更具有收益,则新的解决方案被采用并成为当前的解决方案,这个概率会随着损失的增加而减少,应用上述方法进行一定数量的迭代之后,就会进一步优化用局部搜索方法找到目前最佳解。

带时间依赖和时间窗口的团队定向问题是指:给出一组节点和每对节点之间的旅行时间,其中每个节点与利润、访问时间和时间窗口相关联,目的是找到从起始节点到目的节点间的固定数量且不相交的路径,每条路径不超过给定的时间限制,在不违反其时间窗口的情况下通过访问所有路径中的节点来最大化收集总利润。Garcia 等^[21]提出了 2 种不同的方法来解决上述问题:1) 利用预先计算,得到所有 POI 对之间的平均旅行时间,约减掉时间依赖限制;2) 在旅行时间上加入时间依赖,但是该方法是基于周期性服务时间的简化假设,不符合现实中城市的交通网络。

此外还有一些用于模拟旅游路线规划问题的 TOP 变型,考虑到问题的更多属性或对不同属性的多个约束,Luo 等^[22]引入了一种用于 TOP 变型的启发式算法,该方法在旅行中插入节点时应用 2 种不同的优先级规则,算法在解决方案的质量和执行时间方面优于其他启发式算法,能够在较短时间内得到由精确算法求解实例中的最优解;Li 等^[23]制定了带容量约束和时间窗口的团队定向问题,增加了服务节点在有限时间可用性的约束,并使用整数线性规划求解方法获得了精确的解,然而这种方法不适合进行实时应用。

综上所述,利用 OP 的多种变型对旅游路线规划问题进行建模求解的方法,可以准确建模旅游路线规划问题中多方面的因素,如用户约束、景点开放时间和出行交通方式等,能得到一个或多个满足用户约束的精确路线规划结果,但是这种方法与现实生活中的旅游路线规划问题还有很

多不同。首先,由于旅游活动是一个动态的过程,有很多不确定的因素,无法进行准确建模,而恰是这些不确定因素可能对路线预测的准确性起着决定性的作用^[24]。此外,在对旅游路线规划问题的建模求解时,基于兴趣点的考虑,只是将兴趣点作为一个点,并没有考虑到兴趣点的实际大小,因此这种方法只适用于博物馆、公园、小广场等有固定出口且范围较小的景点,对于相对较大的景点,这种方法就会与实际情况有较大出入,如在游览一条观光河道时,兴趣点的起点和终点可能相差很远,这时再从起点进行路线规划就变得不切实际。

3.2 基于用户生成内容的求解

近些年,由于信息的传播和共享越来越便捷,互联网上积累了大量的集体智慧相关数据,影响着人类生活的许多领域,尤其是旅游业和旅游行为。研究表明,超过 87% 的客户依靠集体智慧为旅行做出决定,例如旅行者在预订住宿之前通常会查看相关的评分和评论^[25]。虽然许多旅游网站提供关于目的地和旅行路线的信息,但是整合和比较来自海量用户的不同类型信息需要大量时间和精力^[26]。

在旅游领域中,用户在进行一次旅行后,通常会分享自己的经验和评论,形成了包括用户评论、照片、签到数据、旅游游记和 GPS 轨迹等信息的大量用户生成内容,这些数据为便利行程计划提供了极大的机会^[27]。虽然一个单独的评论或者旅游游记中可能存在噪声或者偏见,但是将来自大量用户的贡献的内容作为一个整体可以有效地抓住一个景点的本质。因此,越来越多的研究利用空间分析和数据挖掘等技术对这些内容进行分析^[28],得到用户的相关偏好和历史轨迹信息,发现游客间的相似性,实现旅游路线的推荐^[29]。

3.2.1 利用用户 GPS 轨迹进行求解

随着配备 GPS 功能的设备数量不断增加,越来越多的轨迹被连续地产生和分享,也正在改变着人们与网络的交互方式。基于这些轨迹信息,一些应用问题变得可行,例如旅游路线规划问题, GPS 轨迹中包含丰富的信息,可以挖掘用户在一个位置花费的时间和对不同位置的访问顺序等,这些信息可以被用来挖掘指定区域中的热门景点和一般的旅行路线,从而进一步改进路线推荐。

Cui 等^[30]使用用户历史 GPS 信息挖掘用户的旅行习惯,提出了 2 种个性化的旅游路线推荐算法,能够在推荐时考虑用户的个人偏好,提高推荐结果的个性化程度。1) 首先使用协同过滤技术

估计用户的旅行行为频率,然后基于朴素贝叶斯模型生成一个符合用户旅行习惯的路线。2)在路线生成的过程中考虑了用户冷启动问题,利用所有用户的隐含因子向量的平均值作为未挖掘出用户旅行习惯的用户的隐含因子向量;此外,该方法还融合了用户旅行习惯中的最大可能出行距离,能够更好地限制生成路线的长度来满足用户的历史出行习惯。实验结果表明,这2种方法在召回率和准确率上均有更好的表现,但这两种方法在处理用户旅行习惯时,并没有考虑用户旅行习惯的序列性,也没有考虑用户旅行时的动态性,在生成路线结果的合理性上存在一定的偏差。

3.2.2 利用带地理标签的照片进行求解

目前,社交网站中存在大量带地理标签的照片数据,从这些照片数据中分析历史用户位置在地理空间中的分布特征和用户位置随时间的变化特征可以挖掘出用户的行进路线,这些路线可以加入到用于路线推荐的知识库中,通过这些挖掘出的路线帮助新用户进行旅游路线规划,可以提高路线规划的准确度和个性化程度。

利用带地理标签的用户照片进行旅游路线规划的工作是当前旅游路线规划问题研究领域中的一大热点。其中一类工作的重点是从地理标签照片中挖掘出隐含信息,进而得到用户的旅行习惯、移动模式或兴趣偏好等信息为用户进行路线规划。而另一类将重点放在从照片中挖掘序列特性,之后利用挖掘到的序列特性结合概率模型,得到从一个景点最有可能去往的下一个景点信息,最终生成路线规划结果,这类工作的推荐结果更倾向于路线的流行程度。

Sun等^[31]将重点放在挖掘旅游路线中的道路片段信息上,而不是挖掘用户相关信息,通过计算得到道路片段的流行度进行两个景点间的道路推荐。首先利用空间聚类对照片进行分类,而在噪音数据的处理上,提出一种熵过滤方法从照片数据中去除掉与旅行无关的照片,具体实现如式(1)、式(2):

$$E(u) = - \sum_i^{\text{Mon}(u)} P_i(u) \log P_i(u) \quad (1)$$

$$P_i(u) = \frac{D_i(u)}{\sum_i^{\text{Mon}(u)} D_i(u)} \quad (2)$$

式中: $D_i(u)$ 是在目标区域用户 u 在第 i 个月的拍照天数; $\text{Mon}(u)$ 是用户 u 在目标区域拍照的月数,作者使用一个阈值 $E(u)$,当超过这个值时,就

将该照片的拍摄者认定为本地居民,否则认为是游客拍摄的照片。最后利用DBSCAN算法从照片数据中识别出地标建筑,并按照流行度进行排序。实验数据表明,该方法能够给用户推荐一个包括合适景点且路线长度适中的旅游路线,但这种方法存在一定的缺点,主要体现在计算兴趣点流行度时并没有考虑游客的经验和知识。

Wei等^[32]提出了基于集体知识的路线推理框架。首先给定一个位置序列和时间跨度,通过以相互加强的方式聚合用户照片数据,得到流行的路线信息,之后路线算法根据用户指定的查询来构造top-k路线。算法可以在0.5s内找到前3个路线,与其相应的地理实况相比,距离误差小于300m。

Tai等^[33]使用关联规则挖掘和序列模式挖掘从用户带地理标签的照片中提取用户对热门景点的访问序列,从而进一步得到流行路线信息,之后基于用户的历史访问信息,从这些路线中挑选出最适合该用户的路线推荐给用户;Lu等^[34]使用从Panoramio中收集的大量带地理标签的照片,提出一个旅行路线生成算法,该方法考虑到在每个位置花费的时间、总旅行时间和用户偏好。

上述工作并没有将用户行为习惯、兴趣偏好和路线的流行度进行结合,路线规划结果的个性化程度不高。此外,在对用户进行路线规划时并没有过多地考虑上下文信息,如天气、访问时间、季节等,而这些因素往往影响了旅行者的访问习惯,进而使得推荐结果的准确性大大降低。文献^[35-36]基于以上不足出发,在利用用户照片数据的同时加入更多的上下文信息来提高路线推荐结果的准确性。例如,Arain等^[37]利用带地理标签的照片提取旅游景点语义信息的同时挖掘用户喜好信息,在进行推荐时考虑到用户的当前上下文信息,包括时空上下文信息、社交上下文信息和天气上下文信息。Huang^[38]基于游客对景点进行访问时在上下文存在相似性的原理出发,提出一种基于启发式的上下文相似度的计算方法,能够准确刻画用户间的上下文相似性,从而在对用户进行路线推荐的过程中加入上下文信息,给用户提供更加准确的推荐;这种方法不但能够用于照片数据,同样可以在用户GPS轨迹、签到信息等数据中使用,提供具有上下文感知的推荐结果。

3.2.3 利用用户签到数据进行求解

随着基于位置的服务的兴起,Facebook等社交网络提供了用户“签到”的服务,用户通过该服务可以将自己当前的访问地点与时间信息展现在自己的主页上。基于签到数据进行路线规划是近

年来基于位置服务中一个比较流行的研究领域。通过分析用户连续的签到数据,分析用户位置随时间的变化和签到位置在地理空间中的分布特征,挖掘签到位置在时间周期内的分布规律,从而可以得到用户的历史轨迹和访问时间等信息。此外,通过对某个特定地点的签到数据进行分析,还能够进一步挖掘出热门景点信息。

这类工作通常基于用户的出发点和目的地,并结合一些用户限制为用户进行推荐行程;在具体推荐算法上主要使用了基于人口统计学的推荐和基于模型的推荐。宋晓宇等^[39]基于用户签到数据提出了一种短时间体验式路线搜索算法,利用签到数据中连续两个景点签到的时间间隔来代表两个景点间的总时间代价,能够在短时间内让用户体验到多种类别特点的景点,得到一个满足用户要求且具有最大收益的路线作为推荐结果;然而,该方法会随着景点个数增加而增大空间的消耗程度,且没有考虑景点在不同时间段的流行程度问题。文献^[40]利用签到数据针对旅游中常见的结伴出行现象,提出一种群体旅游路线推荐问题,通过分析聚合用户偏好时通常采用的平均数策略与无痛苦策略在推荐结果方面存在的不足,针对搜索路线所具有的动态性特点,提出了一种动态聚合用户偏好的策略;基于该策略建立路线评价模型,对路线进行满意度评分,返回分值最高的路线,并验证了算法在不同参数设置下的有效性。

Cho等^[41]基于用户签到行为的时间特征,利用人类运动的周期性原理,设计出了一个能够基于当前时间进行位置推荐的路线推荐系统,认为短期出行在空间和时间上的周期性不受社会网络结构的影响,而长途旅行则更多地受到来自社会网络关系的影响。鉴于此,笔者借助时间概率分布函数和社会关系提出了一个人类流动模型,相比现有的人类流动模型具有更好的预测性能,从而可以预测用户未来运动的位置和相关动态,提高了对用户进行路线规划结果的准确性。但该模型并没有考虑突发状况对于出行活动的影响,在路线预测的可靠性上也还有待提高。

Rahimi等^[42]通过研究用户签到数据中空间和时间的周期性进一步扩展了现有工作,提出了2种新的推荐算法,即基于用户空间旅行行为的概率分类推荐方法和基于用户历史行为的概率分类推荐方法,然后使用时间概率分布对用户推荐感兴趣类别的位置。实验结果表明,该方法在召回率和准确率上可以实现超过15%的提升。

3.2.4 利用多类型用户生成内容进行求解

利用单一类型的用户生成内容进行旅游路线

规划时,往往存在很大的不确定性,需要加入多方面的限制进行预测和判断,很难保证所挖掘到的用户轨迹的准确性。因此,综合利用多种类型的用户生成内容,更加准确地挖掘用户历史轨迹对用户进行旅行路线推荐的方法成为当下研究的热点。

Guo等^[43]利用一种多源社交媒体融合的方法从多方面整合零碎的旅游信息对用户进行路线推荐。利用信息熵的方法计算一个词在一条评论中所占的比重,从而进一步得到一条评论在所有评论所占的比重,去掉景点的无效评论信息,将游记中景点出现的顺序作为用户对于景点的访问序列,之后利用序列模式挖掘算法从用户游记中挖掘流行路线,最后基于用户评论和景点图片之间的相似性以及从游记中挖掘的流行路线,得到多源信息间的相关性,实现对用户路线推荐。但是这种方法在进行路线推荐时,并没有考虑到一条路线的时间和花费约束,也没有考虑到用户的个性化偏好和需求,因此推荐结果的准确性和个性化程度不高。

Chen等^[44]利用基于地理位置的社交网络中的地理标签图像和用户签到信息,提出了一个名为Scenic Planner的新型框架用于旅行路线推荐。该框架包括风景路网建模和景区路线规划:首先,通过从地理标签图像和签到数据中提取相关信息,丰富道路网络,为每条路段分配适当的景观评分,为景区道路网建模;之后,应用启发式算法在满足用户指定约束(包括起点、目的地和总新行驶距离)的情况下迭代地添加路段使得总景点评分最大化;最后,通过现实世界中的3个数据集验证了框架的效率和效能。虽然该框架在一定程度上加速了道路网络中节点距离的计算,但并未考虑时间因素,如在一天中的不同时间或者在不同季节的建模问题,不能完全整合用户的旅游偏好。

Chen等^[45]抽取多源信息的特征后利用机器学习的方法进行训练,从而获得用户历史行为习惯,对用户进行路线推荐。首先抽取出景点的分类、流行度、平均访问时间和总访问时间作为景点特征,使用K-means方法对景点进行聚类,通过景点信息获得兴趣点的排名信息,来解决路线规划中的起点和终点问题;在旅游游记中利用马尔科夫链模型挖掘兴趣点之间的转移特性,最终结合景点排名和景点间的转移特性推荐出一条路线。该方法与其他路线推荐方法相比在时间和准确性上有很大提高,但是这种方法并没有考虑到用户的个性化需求和旅行路线规划的动态性。例

如：一个用户在一个景点花费的时间过长时，去往下一个景点的概率可能会发生相应的改变。

现阶段越来越多的工作基于这些用户生成内容出发，从中挖掘用户的历史行为习惯和出行轨迹、旅行路线等信息，其中一些具有代表性的工作如表 2 所示，这些工作的优点是可以避免传统基于建模方法的复杂度问题，在路线规划结果上

更贴近于现实生活，符合用户行为习惯。然而，在处理用户数据上也会花费一定的时间，并且当前存在的主要问题是，推荐的结果主要停留在热门的景点和路线上，在推荐时并没有过多地考虑到用户的偏好和用户所处地点和时间的上下文信息(如天气、访问季节等)，推荐结果的个性化程度和准确性方面有很大的提升空间。

表 2 基于 UGC 进行旅游路线规划工作对比

Table 2 Comparison of UGC based tourism route-planning

应用主要数据类型	挖掘的知识内容	应用方法	代表文献
用户照片、签到信息	签到位置分布特征、景点信息	启发式算法	[44]
用户评论、旅游游记	热门景点、景点属性、用户位置轨迹	自然语言处理、序列模式挖掘算法	[43]
景点信息、旅游游记	用户偏好特征、用户位置轨迹	监督学习、分类算法、马尔可夫模型	[45]
用户照片	景点信息、位置分布、社会关联性	关联规则挖掘、马尔可夫链模型	[35-36]
用户签到信息	签到频率、社会关系、位置移动模式	相似性计算、协同过滤	[46]
用户 GPS 轨迹	位置轨迹关联性、位置地理分布	序列挖掘、基于模型推荐	[47]

3.3 旅游路线规划问题的重点和难点

在进行旅游路线规划问题求解时，除了要对用户建立详细的用户画像以及挖掘用户的行为习惯和相关偏好之外，还需要考虑用户的上下文信息和相关旅游信息，以便进一步提高推荐结果的准确性。基于这种认识，在对 3.1 节、3.2 节中相关工作进行考察的基础上，设计了图 1 所示的旅游路线规划系统整体架构。

有的多数工作只考虑到以固定的起点和终点进行规划，用户被限制在一组预定义的地点中，这与现实的场景相差较大，显然使用固定集合中预定义的位置(例如 POI 和酒店)计算的旅行成本不能支持动态的使用场景，因此应该对研究的问题进一步泛化，其中旅程的开始、结束位置可以是目的地城市中的任意位置，在运行时才被确定。此外，在建模的过程中加入更多的用户约束和用户上下文信息，如天气、位置和季节信息等，能够在贴近用户需求的同时极大地提高路线推荐的准确性；在进行精准建模时还可以参照一些其他领域的研究工作，利用其中的一些优化算法，捕获旅游路线规划问题中密切相关的约束，如建模参数等。

3.3.2 数据预处理

在利用用户生成内容进行旅游路线规划时，数据的来源主要集中在相关网站和用户的个人分享，得到的初始数据存在很多噪音信息或者数据缺失的现象，而数据的数量和准确性直接关系到所生成推荐路线的质量，因此如何对获取到的数据进行精确的预处理是此类工作的一个研究重点。噪音数据处理的结果会直接影响最终的推荐结果，例如，对常住居民拍摄的照片和游客拍摄的照片进行区分，能够更加准确地挖掘出游客历史旅行习惯^[33]；在使用用户历史 GPS 轨迹挖掘轨迹信息进行路线规划时，由于受 GPS 精度和信号干扰等因素影响，导致原始 GPS 数据中可能存在一些异常点，而这些异常点属于噪音数据，会影响后续用户轨迹挖掘时的精度和准确性，可以基

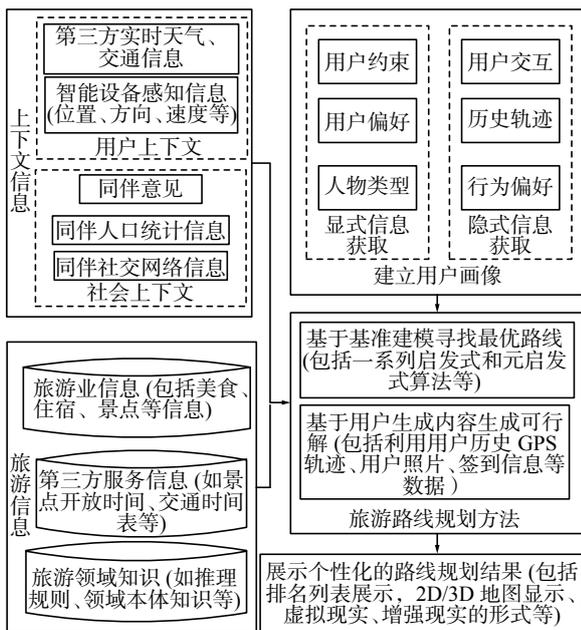


图 1 旅游路线规划系统整体架构

Fig. 1 Tourism route-planning system framework

3.3.1 贴近现实进行精准建模

在对旅游路线规划问题进行建模求解时，现

于序列中相邻点的移动距离小于最大行进距离的原理;利用轨迹点之间的欧式距离和游客最大移动速度剔除掉序列中的异常点。

3.3.3 位置轨迹挖掘与用户偏好特征提取

在利用用户生成内容对用户进行旅游路线规划时,用户的位置移动轨迹能够在一定程度上反应出用户的旅行习惯和偏好行为,因此如何从这些数据中准确地获取到用户的相关信息(历史轨迹、偏好信息等)是此类工作关键问题之一,无论在应用开发还是学术研究中,准确挖掘用户信息都起到至关重要的作用,在未来的研究工作中,应该是重点关注的内容之一。目前利用用户生成内容提取的用户旅行活动特征主要包括用户旅行活动的地理空间分布特性、用户轨迹中访问地点的挖掘和序列特征、用户照片中地标建筑和访问次序的挖掘。在推荐时考虑到不同用户可能有不同需求,为了给用户推荐更高质量旅游路线,挖掘到用户的位置轨迹和偏好特征信息后结合更多的上下文信息,如用户的当前位置、景区实时流量、突发事件报告和用户行为习惯等,能够进一步提高预测的准确性。此外,来源于社交网络中的旅游数据在文本描述方面通常较为简短,所以语义稀疏性较高,因此如何有效解决旅游数据的语义稀疏问题以准确获取游客偏好也是此类工作的一个重点和难点问题,如近期 Kou^[48]、Cheng^[49]等利用短文本中的空间和时间等特征提出了一些具有代表性的主题模型抽取方法。

3.3.4 路线快速生成与实时更新

在当前大数据的时代背景和全域旅游的发展战略下,用户生成数据激增,可选择的旅行地点和内容也急剧增多,这些都为快速实现路线规划带来很大的困难。在旅游路线规划的算法设计中最重要的目标之一便是对用户查询的实时响应,目前解决这一问题的有效途径之一是通过并行计算技术,例如在启发式和元启发式算法中对好的相邻解的局部搜索进行并行计算或者在空间中划分多个子空间,并行地在每一个子空间中运行启发算法,因此并行计算技术是未来快速旅游行程推荐的重要研究方向之一。此外,现有解决方案中没有考虑到用户偏离原始计划路线情形,尽管这种偏离极有可能发生,例如用户自身状况改变、突发社会事件、景区流量控制等,因此需要加入动态重调度功能实时检测当前路线是否偏离,若偏离则呈现新的路线调度。

3.3.5 融合多源信息实现精准推荐

利用多源的用户生成内容进行路线规划,可

以更加准确地挖掘用户历史轨迹和偏好等信息,能够有效提高对用户进行路线推荐时的时效率和精确性,例如:通过用户的历史GPS信息和用户签到信息挖掘用户的历史轨迹要比从用户照片中挖掘相关信息具有更高的可靠性,而通过用户照片挖掘用户历史访问景点信息要比前两者具有更高的可靠性。除了利用照片数据挖掘用户历史轨迹,还可以利用景点的文本描述信息,从中提取出景点的分类信息、流行程度,平均访问时间和总访问时间,将其作为景点特征信息进行景点聚类来提高推荐结果的准确性。

3.3.6 用户隐私保护

伴随信息时代的快速发展,用户的隐私问题得到了越来越多的关注,在基于位置的服务中,用户的隐私保护问题一直是领域内的热点问题,更是关系到用户是否使用该服务的决定性因素。在对用户进行路线推荐时,需要用户主动共享位置信息来获取用户的移动特征或行为偏好,其中涉及用户的隐私问题主要体现在2个方面:这些信息可能被非法使用,有些隐私信息是用户不希望被获取到的。此外,在位置信息的传播过程中也可能会导致用户隐私的泄露,这主要是因为网络中位置信息的传播是以明文的形式,容易被非法的第三方机构获取。因此,如何权衡用户隐私的保护与利用,为用户提供良好的路线规划服务的同时保护好用户的隐私是当前基于用户历史轨迹进行路线规划的一个重点问题,在未来用户在位置共享的选择上,可能会更加趋于自主化,而位置信息的传播也会进行相应的加密。

4 结束语

旅游业的快速发展和日益严重的“信息过载”问题,使得旅游路线规划问题得到了广泛关注和应用。虽然在已有的路线规划问题研究中存在很多求解方法,但传统的旅游路线规划在推荐结果的质量和速度上都存在很多不足,而全域旅游和智慧旅游等战略的提出以及用户分享内容的激增,给旅游路线规划问题带来更多机遇的同时也带来了巨大挑战,基于用户生成内容进行旅行路线规划的方法成为当前研究的热点,但仍有一些问题有待解决。本文从旅游路线规划问题建模出发,分析了当前研究工作中对问题进行建模求解的现状和不足。在此基础上,引出了基于用户生成内容进行旅游路线规划的研究,从求解方法的不同角度详细综述了目前旅游路线规划问题的研究进展。在深入、细致地进行分类总结的基础

上,指出其中的问题或不足。围绕旅游路线规划系统整体架构,对面临的重点和难点问题进行了分析和讨论,为下一步旅游路线规划问题的研究提供理论支持,同时指出了未来该领域研究的重点方向。

参考文献:

- [1] 常亮,曹玉婷,孙文平,等.旅游推荐系统研究综述[J].*计算机科学*,2017,44(10):1-6.
CHANG Liang, CAO Yuting, SUN Wenping, et al. Review on tourism recommendation system[J]. *Computer science*, 2017, 44(10): 1-6.
- [2] GAVALAS D, KONSTANTOPOULOS C, MASTAKAS K, et al. A survey on algorithmic approaches for solving tourist trip design problems[J]. *Journal of heuristics*, 2014, 20(3): 291-328.
- [3] ZHOU Xiaolu, WANG Mingshu, LI Dongying. From stay to play-a travel planning tool based on crowdsourcing user-generated contents[J]. *Applied geography*, 2017, 78: 1-11.
- [4] ABBASPOUR R A, SAMADZADEGAN F. Time-dependent personal tour planning and scheduling in metropolises[J]. *Expert systems with applications*, 2011, 38(10): 12439-12452.
- [5] GAVALAS D, KASAPAKIS V, KONSTANTOPOULOS C, et al. A personalized multimodal tourist tour planner[C]//Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia. Melbourne, Australia, 2014: 73-80.
- [6] BAO Jie, ZHENG Yu, WILKIE D, et al. Recommendations in location-based social networks: a survey[J]. *GeoInformatica*, 2015, 19(3): 525-565.
- [7] ZHENG Bolong, SU Han, ZHENG Kai, et al. Landmark-based route recommendation with crowd intelligence[J]. *Data science and engineering*, 2016, 1(2): 86-100.
- [8] SOUFFRIAUX W, VANSTEENWEGEN P, VERTOMMEN J, et al. A personalized tourist trip design algorithm for mobile tourist guides[J]. *Applied artificial intelligence*, 2008, 22(10): 964-985.
- [9] GUNAWAN A, LAU H C, VANSTEENWEGEN P, et al. Well-tuned algorithms for the team orienteering problem with time windows[J]. *Journal of the operational research society*, 2017, 68(8): 861-876.
- [10] GUNAWAN A, LAU H C, LU Kun. An iterated local search algorithm for solving the orienteering problem with time windows[M]//OCHOA G, CHICANO F. Evolutionary Computation in Combinatorial Optimization. Cham: Springer, 2015: 61-73.
- [11] GUNAWAN A, LAU H C, LU Kun. SAILS: hybrid algorithm for the team orienteering problem with time windows[C]//Proceedings of the 7th Multidisciplinary International Scheduling Conference. Prague, Czech Republic, 2015: 276-295.
- [12] VERBEECK C, SØRENSEN K, AGHEZZAF E H, et al. A fast solution method for the time-dependent orienteering problem[J]. *European journal of operational research*, 2014, 236(2): 419-432.
- [13] SCHILDE M, DOERNER K F, HARTL R F, et al. Meta-heuristics for the bi-objective orienteering problem[J]. *Swarm intelligence*, 2009, 3(3): 179-201.
- [14] LU Ying, SHAHABI C. An arc orienteering algorithm to find the most scenic path on a large-scale road network[C]//Proceedings of the 23th SIGSPATIAL International Conference on Advances in Geographic Information Systems. Seattle, Washington, 2015: 51-62.
- [15] LU Ying, JOSSE G, EMRICH T, et al. Scenic routes now: efficiently solving the time-dependent arc orienteering problem[C]//Proceedings of 2017 ACM Conference on Information and Knowledge Management. Singapore, 2017: 487-496.
- [16] LU Yongliang, BENLIC U, WU Qinghua. A memetic algorithm for the orienteering problem with mandatory visits and exclusionary constraints[J]. *European journal of operational research*, 2018, 268(1): 54-69.
- [17] GAO Huiji, TANG Jiliang, HU Xia, et al. Exploring temporal effects for location recommendation on location-based social networks[C]//Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong, China, 2013: 93-100.
- [18] XU Ying, HU Tao, LI Ying. A travel route recommendation algorithm with personal preference[C]//Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. Changsha, China, 2016: 390-396.
- [19] 柯良军,章鹤,尚可,等.一类求解带时间窗的团队定向问题的改进蚁群算法[J].*计算机科学*,2012,39(4):214-216.
KE Liangjun, ZHANG He, SHANG Ke, et al. Improved ant colony optimization approach for the team orienteering problem with time windows[J]. *Computer science*, 2012, 39(4): 214-216.
- [20] LIN S W, YU V F. A simulated annealing heuristic for the team orienteering problem with time windows[J]. *European journal of operational research*, 2012, 217(1): 94-107.
- [21] GARCIA A, VANSTEENWEGEN P, ARBELAITZ O, et al. Integrating public transportation in personalised electronic tourist guides[J]. *Computers and operations research*, 2013, 40(3): 758-774.

- [22] LUO Zhixing, CHEANG B, LIM A, et al. An adaptive ejection pool with toggle-rule diversification approach for the capacitated team orienteering problem[J]. *European journal of operational research*, 2013, 229(3): 673–682.
- [23] LI Zhenping, HU Xianman. The team orienteering problem with capacity constraint and time window[C]//Proceedings of the 10th International Symposium on Operations Research and Its Applications. Dunhuang, China, 2011: 157–163.
- [24] HJALAGER A M. 100 Innovations that transformed tourism[J]. *Journal of travel research*, 2015, 54(1): 3–21.
- [25] MURPHY H C, CHEN Mengmei, COSSUTTA M. An investigation of multiple devices and information sources used in the hotel booking process[J]. *Tourism management*, 2016, 52: 44–51.
- [26] BANERJEE S, CHUA A Y K. In search of patterns among travellers' hotel ratings in TripAdvisor[J]. *Tourism management*, 2016, 53: 125–131.
- [27] ZHOU Xiaohu, XU Chen, KIMMONS B. Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform[J]. *Computers, environment and urban systems*, 2015, 54: 144–153.
- [28] 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 Limin, GAO Xu, WU Bin, et al. . Discovering common behavior using staying duration on semantic trajectory[J]. *Journal of computer research and development*, 2017, 54(1): 111–122.
- [29] 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 Qiang, ZHANG Fengli, WANG Ruijin, et al. Trajectory big data: a review of key technologies in data processing[J]. *Journal of software*, 2017, 28(4): 959–992.
- [30] CUI Ge, LUO Jun, WANG Xin. Personalized travel route recommendation using collaborative filtering based on GPS trajectories[J]. *International journal of digital earth*, 2018, 11(3): 284–307.
- [31] SUN Yeran, FAN Hongchao, BAKILLAH M, et al. Road-based travel recommendation using geo-tagged images[J]. *Computers, environment and urban systems*, 2015, 53: 110–122.
- [32] WEI Lingyin, ZHENG Yu, PENG W C. Constructing popular routes from uncertain trajectories[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China, 2012: 195–203.
- [33] TAI C H, YANG Denian, LIN L T, et al. Recommending personalized scenic itinerary with geo-tagged photos[C]//Proceedings of 2008 IEEE International Conference on Multimedia and Expo. Hannover, Germany, 2008: 1209–1212.
- [34] LU X, WANG C, YANG J M, et al. Photo2Trip: generating travel routes from geo-tagged photos for trip planning[C]//Proceedings of the 18th International Conference on Multimedia. Florence, Italy, 2010: 143–152.
- [35] KURASHIMA T, IWATA T, IRIE G, et al. Travel route recommendation using geotagged photos[J]. *Knowledge and information systems*, 2013, 37(1): 37–60.
- [36] KURASHIMA T, IWATA T, IRIE G, et al. Travel route recommendation using geotags in photo sharing sites[C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Toronto, Canada, 2010: 579–588.
- [37] ARAIN Q A, MEMON H, MEMON I, et al. Intelligent travel information platform based on location base services to predict user travel behavior from user-generated GPS traces[J]. *International journal of computers and applications*, 2017, 39(3): 155–168.
- [38] HUANG Haosheng. Context-aware location recommendation using geotagged photos in social media[J]. *International journal of geo-information*, 2016, 5(11): 195.
- [39] 宋晓宇, 许鸿斐, 孙焕良, 等. 基于签到数据的短时间体验式路线搜索[J]. *计算机学报*, 2013, 36(8): 1693–1703.
SONG Xiaoyu, XU Hongfei, SUN Huanliang, et al. Short-term experience route search based on check-in data[J]. *Chinese journal of computers*, 2013, 36(8): 1693–1703.
- [40] 宋晓宇, 闫玉奇, 孙焕良, 等. 基于签到数据的群体旅游路线推荐[J]. *计算机科学与探索*, 2015, 9(1): 51–62.
SONG Xiaoyu, YAN Yuqi, SUN Huanliang, et al. Group trip recommendation based on check-in Data[J]. *Journal of frontiers of computer science and technology*, 2015, 9(1): 51–62.
- [41] CHO E, MYERS S A, LESKOVEC J. Friendship and mobility: user movement in location-based social networks[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, CA, USA, 2011: 1082–1090.
- [42] RAHIMI S M, WANG Xin. Location Recommendation based on periodicity of human activities and location categories[M]//PEI Jian, TSENG V S, CAO Longbing, et al. *Advances in Knowledge Discovery and Data Mining*. Berlin, Heidelberg: Springer, 2013: 377–389.
- [43] GUO Tong, GUO Bin, OUYANG YI, et al. CrowdTravel: scenic spot profiling by using heterogeneous crowd-sourced data[J]. *Journal of ambient intelligence and humanized computing*, 2017(5): 1–10.
- [44] CHEN Chao, CHEN Xia, WANG Zhu, et al. ScenicPlan-

- ner: planning scenic travel routes leveraging heterogeneous user-generated digital footprints[J]. *Frontiers of computer science*, 2017, 11(1): 61–74.
- [45] CHEN Dawei, ONG C S, XIE Lexing. Learning points and routes to recommend trajectories[C]//Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. Indianapolis, USA, 2016: 2227–2232.
- [46] LU E H C, CHEN C Y, TSENG V S. Personalized trip recommendation with multiple constraints by mining user check-in behaviors[C]//Proceedings of 20th International Conference on Advances in Geographic Information Systems. Redondo Beach, CA, USA, 2012: 209–218.
- [47] ZHENG Yu, ZHANG Lizhu, XIE Xing, et al. Mining interesting locations and travel sequences from GPS trajectories[C]//Proceedings of the 18th International Conference on World Wide Web. Madrid, Spain, 2009: 791–800.
- [48] KOU Feifei, DU Junping, LIN Zijian, et al. A semantic modeling method for social network short text based on spatial and temporal characteristics[J]. *Journal of computational science*, 2017.
- [49] CHENG Xueqi, YAN Xiaohui, LAN Yanyan, et al. BTM: topic modeling over short texts[J]. *IEEE transactions on*

knowledge and data engineering, 2014, 26(12): 2928–2941.

作者简介:



常亮,男,1980年生,教授,博士,中国计算机学会高级会员,主要研究方向为数据与知识工程、形式化方法、智能系统。主持国家自然科学基金项目 1 项、广西自然科学基金项目 1 项。发表学术论文 70 余篇,被 SCI、EI 收录 60 余篇。



孙文平,男,1990年生,硕士研究生,主要研究方向为机器学习、智能规划。



张伟涛,男,1993年生,硕士研究生,主要研究方向为机器学习、推荐系统。