

## 基于用户需求的景点路线利益规划算法

王楠<sup>1,2,3</sup>, 周红磊<sup>1,3</sup>, 李金宝<sup>1,2,3</sup>, 黎玲利<sup>1,3</sup>

1. 黑龙江省数据库与并行计算重点实验室(黑龙江大学), 黑龙江 哈尔滨 150080;
2. 黑龙江大学电子工程学院, 黑龙江 哈尔滨 150080;
3. 黑龙江大学计算机科学技术学院, 黑龙江 哈尔滨 150080)

**摘 要:** 现有基于兴趣点(POI)路径规划的研究大部分只考虑 POI 的静态属性, 而热门景点拥堵以及用户产生的不满意情绪会造成旅游质量大大下降。为了提升用户旅游的满意度, 重点考虑了 POI 的动态属性, 提出基于用户需求的景点路线利益规划算法。首先, 设计了 GM(1,1)-马尔可夫景点人数预测算法, 通过引入预测残差以及概率转移矩阵, 使平均预测偏差比原 GM(1,1)算法降低 12.2%; 其次, 通过设计前向细化(FR)算法, 在满足用户解决需求的前提下减少用户不必要的访问地点和时间, 在相同的需求数下, 前向细化算法的平均解决需求时间比 TMT 算法降低 9.4%; 最后, 根据景点流行度、时间 KL 散度、地点访问次序以及路程时间等因素, 提出了景点路线利益规划算法, 在相同时间限制下景点路线利益算法平均拓展 Rank 1-5 的景点数量比 Time\_Based 算法提高 34.8%, 比 Rand\_GA 算法提高 47.3%。

**关键词:** 兴趣点; 景点利益; KL 散度; 路径规划

**中图分类号:** TP393

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-436x.2018089

## Algorithm for scenario benefit route planning based on user's requests

WANG Nan<sup>1,2,3</sup>, ZHOU Honglei<sup>1,3</sup>, LI Jinbao<sup>1,2,3</sup>, LI Lingli<sup>1,3</sup>

1. Key Laboratory of Database and Parallel Computing of Heilongjiang Province (Heilongjiang University), Harbin 150080, China
2. School of Electronic Engineering, Heilongjiang University, Harbin 150080, China
3. School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China

**Abstract:** Most of the existing research for point of interest route planning only consider the static properties of POI, however, the congestion of the hot spots and users' discontent may greatly reduce the travel quality. In order to increase the tourists' satisfaction, the dynamic attributes of POI was considered and a route planning algorithm based on user's requests was proposed. Firstly, Markov-GM(1,1) forecasting algorithm was designed to predict the number of people in each scenic spot. Markov-GM(1,1) could make the average predication error 12.2% lower than the GM(1,1) algorithm by introducing the predication residual. And then, the forward refinement (FR) algorithm was designed which could avoid visiting the unnecessary place and satisfy user's requests as well. The average solving time of forward refinement algorithm was 9.4% lower than TMT algorithm under the same amount of user's requests. Finally, based on the factors such as spot popularity, KL divergence of time, visiting order and distance et al, the scenic route profit planning algorithm which could make the number of Rank 1-5 spots 34.8% higher than Time\_Based algorithm and 47.3% higher than Rand\_GA algorithm.

**Key words:** point of interest, scenario benefit, KL divergence, route planning

收稿日期: 2017-09-13; 修回日期: 2018-03-29

通信作者: 李金宝, jbli@hlju.edu.cn

基金项目: 国家自然科学基金资助项目 (No. 61370222, No.61602159)

**Foundation Item:** The National Natural Science Foundation of China (No. 61370222, No.61602159)

## 1 引言

近年来,随着无线通信和传感器技术(如 GPS、智能手机以及平板)的发展,定位服务应用应运而生,其中,基于 GPS 移动设备的路径/路径规划(LBS)应用受到了人们的广泛关注,人们从各式的传感器以及社交网络中获取海量数据信息以辅助日常出行需求。

在信息技术蓬勃发展的大环境下,保障出行安全、提高游览效率、降低拥堵概率、减少等待时间成为目前首要解决的问题。用户在去往景点之前,会把该景点称为用户的兴趣点(POI, point of interest),用户怎样才能更快或更省距离地从现处位置到达 POI 成为首要关注的问题。一些文献研究了基于单目的地路径的规划算法,例如, Dijkstra 算法、Floyd-warshall 算法<sup>[1]</sup>等。但实际上单目的地路径规划已无法满足现有用户的出行需求,需涉及多地点路径规划。多地点路径规划可以归结到旅行商问题(TSP, travelling salesman problem)上, TSP 问题作为现有路径规划问题,已被研究者深入研究, TSP 的目的是发现一条路径可以到达用户想去的所有 POI,同时使路径距离最短。蚁群算法和遗传算法作为解决 TSP 问题的 2 种主要算法现已被广泛应用,但这 2 种算法主要用于解决静态图中的最短路径<sup>[2]</sup>。在现实中,路径的权值是经常变化的,如路程耗时权值会受到流量拥堵等因素的影响。所以现有的静态算法不能满足用户实时路径规划需求。

为了解决上述问题,本文提出了利益贪心算法。该算法在设计时考虑了如下因素。

### 1) 景点的流行度

景点的流行度为该景点对于用户的吸引程度,每个区域都有其景点流行度的排名。例如,在杭州旅游时,各大旅游网如携程旅行网,把西湖作为推荐度第一的景点代表,该景点更容易吸引用户前往。

### 2) 景点的合适访问时间

不同景点合适的访问时间也不同。例如,有些地方适合白天去,如灵隐寺。有些地方适合晚上去,如杭州西湖的音乐喷泉。

### 3) 地点的访问次序

由于每个地方的特有属性,地点的访问次序会存在一定顺序。例如,用户在餐馆吃完晚餐后回到酒店是一个较好的选择。

### 4) 路程时间(排队时间+景点间移动时间)

排队时间体现了用户在访问某个地点的拥挤

情况。因为用户的游玩地点满足随机性,即景点的排队人数满足正态分布  $N(\mu, \sigma^2)$ ,可以根据每个景点的人数到达速率  $\mu$  设置正态分布函数,同时减去服务速率来预测排队时间<sup>[3,4]</sup>。景点间的移动时间为用户在景点间移动所耗费的时间。

本文所提算法能够有效减少用户的路程耗时,降低用户遇到长时间排队的概率,同时使用户尽可能多地游览流行度高的景点,使路线的利益达到最大化。

考虑用户在游览期间通常也会产生一些其他需求,例如,从景点 1 到景点 2,用户可能要解决的需求有取钱、加油和喝咖啡等,因此,需规划一条路线解决该需求集合,本文根据该问题提出了前向扩展细化算法从而使访问时间最短。

## 2 相关工作

现有的无线传感器技术给定位和路径规划服务提供了很大的便利,在物流、导航、工业等领域都有广泛的应用<sup>[5]</sup>。现有的导航方案和算法主要有 Dijkstra、Floyd、启发式搜索、蚁群以及遗传算法等<sup>[6]</sup>。蚁群算法为现有的主流算法,其原理是模拟自然界蚂蚁的觅食行为。蚂蚁觅食存在随机性,和景区中的游客游览极为相似。

Yu 等<sup>[7]</sup>通过在蚁群算法中引入特殊的遗传算子,并将两者结合,避免了蚁群算法本身所具有的局部搜索局限性,同时加快和加大了收敛的速度以及全局最优的能力。

旅游路径规划问题(TTDP)<sup>[8,9]</sup>在近几年受到相当大的关注,目前主要采用不同的启发式算法解决该问题。这些算法从不同的角度以及不同的变量和约束,产生了各种各样的模型。大多数 TTDP 模型中考虑的输入参数<sup>[10]</sup>包括:一组候选 POI,每个候选 POI 与一系列相关属性(如类型、位置、开放天数/小时、入场费等)。每个 POI 具有其“利润”,表示其相关重要性。每日还具有时间预算  $B$ ,代表游客希望在旅游景点上花费的时间(每日游览时间总和),即访问时间加上从 POI 转移到另一个 POI 的整体时间,应该在  $B$  之内。

TTDP 建模的目标是推导出一组近乎最佳的每日不相交的行程(有序访问 POI),每个行程包含可用(候选) POI 的子集,以便最大限度地提高旅游满意度(即整体收益)。推导出的旅游路线应该尊重用户约束/POI 属性,并满足观光所需的每日时间预算  $B$ 。因此,可以根据用户偏好来调整几个问

题参数，如 POI 利润可以通过客观和主观的加权函数进行计算，同样访问 POI 的时间通过其访问持续时间的平均值和用户对该特定 POI 的潜在兴趣得出<sup>[11]</sup>。

为了处理旅游规划的流程，现在出现了很多旅游路径规划软件 (trip planner)，例如，NileGuide、YourTour<sup>[12]</sup>等。国内的著名旅游网站有携程旅行网、阿里旅行、去哪儿网等，这些网站都能为游客提供丰富的旅游资源信息。这些软件首先把各景点的知名度进行排名，然后帮游客挑选有趣的景点，同时还能为用户规划好这些景点的访问顺序。此外，随着基于位置的社交网络 (LBSN)<sup>[13]</sup>的日益普及，人们在社交网络上记录各种足迹数据。其中，有关兴趣点和用户丰富有价值的信息，例如，POI 的物理坐标、类别、人气和签到偏好也被包含在这样的足迹数据中<sup>[14]</sup>，这些数据可以用于研究个性化的自助行程规划。现有可用的在线地图服务，例如，GoogleMaps、百度地图、商业 GPS 导航仪可以很方便地进行路径规划，但由于它们提供的只是静态距离最短或时间最短的行进路线，因此，其规划的旅行路线不能满足用户的个性化需求。现在以人为本的旅游模式越来越被人们所提倡，推荐排名靠前的旅游景点往往会出现人数爆棚的情况，同时不同景点的旅游顺序以及不同景点的访问时间都会影响游客旅游的质量，所以设计一个好的旅游路径规划系统具有重要意义。

本文算法基于 OP (orientation problem) 问题，通过情景利益设计动态规划算法及优化算法，可以使规划的路线更能够贴近用户实际需求，还可以提高用户的旅游质量。

### 3 系统框架和问题定义

#### 3.1 系统框架

本文利用智能手机、手环等设备上的三轴加速度计和三轴陀螺仪传感器搜集数据，将原始数据通过滤波降噪得到滤波数据，通过滤波数据得出步数，估计步长以及方向，再通过本文所提算法得到用户到达下个目的地的大致移动时间。然后将景点的历史客流量作为实验数据并采用灰色 GM(1,1)马尔可夫模型进行实时的客流量预测，系统框架如图 1 所示。景点人数预测模型涉及的数据包含了历史客流量数据以及实时的客流量数据，这些数据是通过微信公众号 (上海发布) 的接口获得，每隔 15 min 该公众号会更新一次景点人数数据。

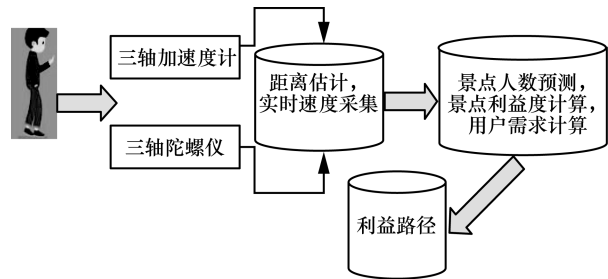


图 1 系统框架

数据以序列的方式存储，具有时间标识。具体的预测过程在 4.1 节详细说明。当预测完成后，根据利益函数对各景点利益进行计算。如果期间用户产生需求则进行需求的路径规划，最后得出一条完整的利益路径。

#### 3.2 问题定义

本文的研究问题可以分为 2 个部分：用户需求路径规划和基于利益的景点路径规划。

##### 3.2.1 用户需求路径规划

该问题是考虑解决用户在旅行过程中需求的路径规划。例如，用户要从图 2 所示的社交区域中的一个景点  $s$  出发，去往景点  $d$ ，但是在从  $s$  去往  $d$  的过程中有一些需求要解决，如需要取钱、买东西等。即给定一个 POI 网络  $M=(P,E)$ 、起点  $s$ 、终点  $d$  以及用户的需求集合  $R'$ ，多需求路径规划的目的是发现一条从  $s$  到  $d$  的总路程时间最短的路线，但是这需要在解决用户需求集合  $R'$  的前提下进行。表 1 为该问题涉及的一些符号定义。

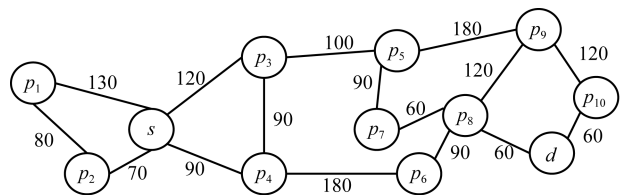


图 2 区域加权图的一个示例

表 1 符号定义

符号	描述
$r, R$	$r$ 代表某个需求, $R$ 代表需求集合
$R'$	$R'$ 代表用户的需求集合
$p, P$	$p$ 代表某个 POI, $P$ 代表 POI 集合
$s, d$	$s$ 代表起点, $d$ 代表终点
$\rho, \mu, \tau$	停留时间, 行走时间, 路程时间
$PSR(), RSP()$	某个需求的 POI 集合, 某个 POI 的需求
$M, E$	区域网络, 网络中的路径
$\varphi$	规划路径

**定义 1** Request (需求)。Request 代表用户所要做的一件事或一个服务,即需求。 $R=\{r_1,r_2,r_3,\dots,r_{|R|}\}$  表示需求集合。

**定义 2** POI (兴趣点)。POI 代表一个节点或一个地点。用  $p=\{p_1,p_2,p_3,\dots,p_p\}$  表示 POI 的集合。对于每个  $p$ , 函数  $c(p)$  代表其所属种类,如商场、饭店等。

**定义 3**  $PSR()$  和  $RSP()$  函数。对于给定的一个需求  $r \in R$  和一个  $p \in P$ ,  $PSR(r)$  代表能解决需求  $r$  的 POI 集合,  $PSR(r) \subseteq P$ 。  $RSP(p)$  代表兴趣点  $p$  所能解决的需求集合  $RSP(p) \subseteq R$ 。

**定义 4** POI 网络。 $M=(P,E)$ ,  $M$  代表区域社交网络,  $P$ 、 $E$  分别为其顶点和边集合。图 2 为区域加权图的一个示例。

**定义 5** 路径  $\varphi$ 。表示由一个或多个 POI 组成的路径,即对于  $\forall p' \in \varphi, p' \in P$ ,  $P$  为 POI 点集合。

**定义 6** 合法路径  $\varphi'$ 。给定用户的个性需求集合  $R'=\{r'_1,r'_2,\dots,r'_n\}$ , 设定  $R$  为总的需求集合, 所以  $R' \subseteq R$ 。当  $R' \subseteq PSR(\varphi')$ , 即路径  $\varphi'$  中包含所有用户的个性需求, 则称该路径为合法路径。

**定义 7** 路程时间  $\tau$ 。开始时给定起点  $s$  和终点  $d$ , 还有用户的个性需求集合  $R'=\{r'_1,r'_2,\dots,r'_n\}$ , 路程时间  $\tau(s,d,\varphi)$  为停留时间加上移动时间, 则有

$$\tau(s,d,\varphi) = \sum_{i=1}^m \rho(p_i) + \mu(s,p'_1) + \sum_{i=1}^{m-1} \mu(p'_i,p'_{i+1}) + \mu(p'_m,d) \quad (1)$$

### 3.2.2 基于利益的景点路径规划

考虑到现有的路径规划算法在动态环境下的结果准确性较差, 本文对此进行了深入的研究, 考虑了景点的 4 个属性, 各个属性定义如下。

**定义 8**  $pop(l_i)$ 。景点  $l_i$  的流行度, 代表该景点对于游客的吸引程度, 为

$$pop(l_i) = \frac{n(l_i)}{n_{total}} \quad (2)$$

则路线  $Route$  的流行利益度函数为

$$f_{pop}(Route) = \left( \prod_{i=1}^n pop(l_i) \right)^{\frac{1}{n}} \quad (3)$$

**定义 9**  $D_{KL}$ 。景点的访问时间散度如图 3 所示, 用散度表示用户访问景点的时间与景点合适访问时间分布的离散程度,  $D_{KL}$  越小代表访问时间越合适, 如式(4)所示。其中,  $f_i(t)$  为用户访问景点的时间分布, 如实线所示;  $G(t; \mu, \sigma^2)$  为景点的合适

访问时间分布, 如虚线所示。景点合适访问时间分布选取正态分布函数的原因是, 通过正态分布函数可直接显示出景点的合适访问时间分布与用户的访问景点的时间分布的差异性, 以此来计算散度。

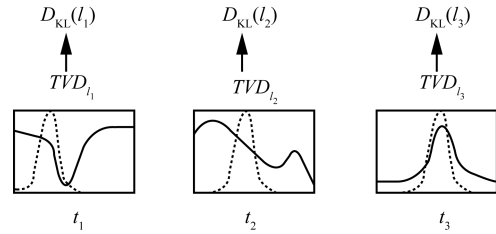


图 3 景点访问时间散度

$$\begin{aligned} D_{KL} &= (f_i(t) \parallel G(t; \mu, \sigma^2)) \\ &= \sum_x \left[ G(x; \mu, \sigma^2) \ln \frac{G(x; \mu, \sigma^2)}{f_i(x)} + \right. \\ &\quad \left. f_i(x) \ln \frac{f_i(x)}{G(x; \mu, \sigma^2)} \right] \quad (4) \end{aligned}$$

所以路线  $Route$  的合适访问时间的利益函数为

$$f_{visit}(Route) = \left( \prod_{i=1}^n D_{KL}(G(t; t_i, \sigma^2)) \frac{1}{pop(l_i)} \right)^{\frac{1}{n}} \quad (5)$$

**定义 10**  $visit\_order$ 。地点访问次序, 需要研究这种顺序并利用它们评估路线的质量。首先给定路线  $Route=\{(l_1,t_1),(l_2,t_2) \dots (l_n,t_n)\}$ , 然后计算 N-GRAM 的概率, 其中, N-GRAM 表示该序列的合理性, 概率越大代表次序越合理。

$$p_{bin}(s) = \left( \prod_{i=2}^n p(l_i | l_{i-1}) \right)^{\frac{1}{n}} \quad (6)$$

$$p_{tri}(s) = \left( p(l_1) p(l_2 | l_1) \prod_{i=3}^n p(l_i | l_{i-1} l_{i-2}) \right)^{\frac{1}{n}} \quad (7)$$

则路线  $Route$  的访问次序的利益函数为

$$f_{order}(Route) = \frac{P_{bin}(Route) + P_{tri}(Route)}{2} \quad (8)$$

**定义 11**  $Route_i(Move_i+Queue_i)$ 。移动时间和排队时间分别为

$$Queue_i = \frac{length_i}{v_{server}} \quad (9)$$

$$Move_i = \frac{route_i}{v_{move}} \quad (10)$$

$$Route(time) = Queue_i + Move_i \quad (11)$$

则路线 *Route* 的路程时间利益函数为

$$f_{\text{time}}(\text{Route}) = \alpha(\text{Queue}_{l_i} + \text{Move}_{l_i}) \quad (12)$$

根据景点的上述 4 个属性可以得出路径的利益函数 *Fitness(Route)*, 其中,  $\alpha$  为时间利益调整参数。

$$\begin{aligned} \text{Fitness}(\text{Route}) = & \text{pop}(\text{Route}) - \alpha \text{Route}(\text{time}) + \\ & \text{Fitness}_{\text{visit}}(\text{Route}) + \text{Fitness}_{\text{order}}(\text{Route}) \end{aligned} \quad (13)$$

### 4 处理算法

本文涉及 3 个主要算法：基于灰色马尔可夫的人数预测算法、前向细化算法以及基于利益最大化的路径规划算法。

#### 4.1 基于灰色马尔可夫的人数预测算法

景点人数预测为本文景点利益路径规划算法的前提, 式(2)中的  $n(l_i)$  为景点  $l_i$  的预测人数。灰色模型的优点是<sup>[15]</sup>: 1)不需要大量的样本; 2)样本不需要有规律性分布; 3)计算工作量小; 4)定量分析结果和定性分析结果一致; 5)可进行短期和中长期预测; 6)该模型预测准确率高。在实际应用中, 系统可以收集景区部门的专业实时统计数据, 这样的数据具有更高的可靠性和精确度。

本文在灰色模型的基础上加入了残差作为修

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & \dots & -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) \\ 1 & \dots & 1 \end{bmatrix}$$

其中,  $Y$  和  $B$  中的  $n$  为数据序列长度, 序列长度  $n$  将会影响  $\hat{\alpha}$ , 从而影响预测值的精度。把残差值作为序列项进行累加, 重建偏差值的 GM(1,1)模型, 式(18)为 GM(1,1)模型。由于残差具有正负, 如表 2 所示, 所以建立马尔可夫转移矩阵。

表 2 残差状态

状态	状态值	残差值
1	1	$\delta(t) > 0$
2	0	$\delta(t) = 0$
3	-1	$\delta(t) < 0$

$$P_{ij} = P\{X(t+s) = j | X(s) = i\} \quad (20)$$

根据  $P_{ij}$  和表 3 可以得出  $3 \times 3$  的概率转移矩阵为

$$\begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \quad (21)$$

正项, 残差定义为预测值与真实值之间的差值。

首先设置原始数据序列为以下形式的序列:  $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ , 为了预测其未来的数据, 需要对数据进行累加, 得到累加序列  $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$ , 其中,  $n$  为序列长度。再根据  $X^{(1)}(t)$  建立灰色模型所对应的微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (14)$$

当  $t = t_0$  时,  $x^{(1)} = x^{(1)}(t_0)$ , 求解该微分方程, 可以得出该方程的解为

$$\hat{x}^{(1)}(t) = \left[ x^{(0)}(t_0) - \frac{u}{a} \right] e^{-a(t-t_0)} + \frac{u}{a}, t = 1, 2, \dots, n \quad (15)$$

然后根据最小二乘法进行求解

$$\hat{\alpha} = \begin{pmatrix} a \\ u \end{pmatrix} = (B^T B)^{-1} B^T Y_n \quad (16)$$

其中,  $Y$  和  $B$  分别为

$$Y = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T \quad (17)$$

$$\hat{x}^{(0)}(t) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \quad (18)$$

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & \dots & -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) \\ 1 & \dots & 1 \end{bmatrix}^T \quad (19)$$

表 3 残差状态转移

状态转移	$S_1 \rightarrow S_1$	$S_1 \rightarrow S_2$	$S_1 \rightarrow S_3$
转移概率	$p(s_1   s_1)$	$p(s_2   s_1)$	$p(s_3   s_1)$
符号	$p_{11}$	$p_{12}$	$p_{13}$

把残差值和马尔可夫状态转移矩阵作为修正和转移参数加入灰色模型, 建立灰色马尔可夫模型为

$$\begin{aligned} \hat{x}^{(0)}(t+1) = & [x^{(0)}(1) - \frac{u}{a}] e^{-at} (1 - e^{at}) + \delta(t+1) \cdot \\ & [(y^{(0)}(1) - \frac{u'}{a'}) e^{-a't} (1 - e^{a't})] \end{aligned} \quad (22)$$

#### 4.2 前向细化算法

前向细化算法 (forward refinement) 在  $k$ -near 算法和 TMT (transferred moving time) 算法的基础上进行了扩展。路径演示如图 4 所示, 可能会出现如下情况: 用户从  $s$  出发去往  $d$ , 用户的需求为喝咖啡、加油以及取钱。根据上述 2 个算法得出的 *Route* 结果是 {coffee, 加油站, 商场}, 但是发现商

场中包含取钱和喝咖啡这 2 个需求，所以可以把 coffee 这个地点去掉，直接从  $s$  出发到达加油站加油，然后到商场喝咖啡和取钱。这样可以减少 POI 的访问次数，减少停留时间。前向细化算法如算法 1 所示。

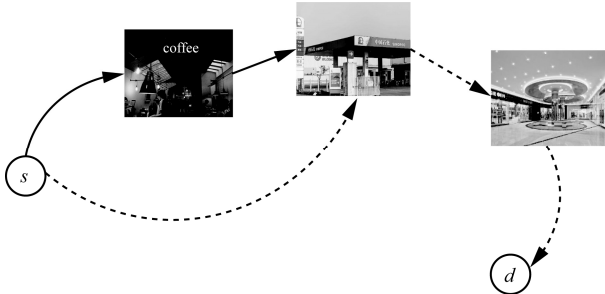


图 4 路径演示

**算法 1** 前向细化算法

输入  $s$ 、 $d$ 、 $M$ 、 $R'$ 、 $R$ 、 $k$

输出  $\min\_Route(s,d)/s$  到  $d$  的最短路径

procedure  $FR(s,d)$

- 1) 初始化  $M, P_{result}, Route = \{s\}$
- 2) while  $R \neq \{\emptyset\}$  do
- 3) for each  $p_i \in |M|$
- 4) SORT( $|PSR(p_i) \cap R'|$ )
- 5)  $A[k] = |PSR(p_i) \cap R'|$
- 6) 把  $A[k]$  插入  $A[0 \dots j]$  已排序数组中
- 7) for each  $p_j \in |M|$
- 8) if  $|PSR(p_j) \cap R'| = \max(A)$  do:
- 9) 比较  $TMT(p_i, s)$  和  $TMT(p_j, s)$
- 10) if  $TMT(p_j, s) > TMT(p_i, s)$  do:
- 11) 把  $p_j$  加入集合  $Route$  并令  $s = p_j$
- 12) 移除  $R'$  中的元素  $PSR(p_j) \cap R'$
- 13) else
- 14) 把  $p_i$  加入集合  $Route$  并令  $s = p_i$
- 15) 移除  $R'$  中的元素  $PSR(p_i) \cap R'$
- 16) end if
- 17) end for
- 18) end for
- 19) end for
- 20) end while
- 21) return  $Route$
- 22) end procedure

以图 2 为例，算法 1 首先初始化  $Route = \{s\}$ ，对于区域网络  $M$  中的所有 POI 进行  $|PSR(p_i) \cap R'|$  的排序，然后选择最大的 POI，可以发现其为  $p_6$ 、 $p_9$ ，

因为  $|PSR(p_6) \cap R'| = |PSR(p_9) \cap R'| = 2$ ，计算  $d$  到  $p_6$ 、 $p_9$  的距离，选择最短的距离点为  $p_9$ ，启动细化机制，然后把  $p_9$  加入  $Route$  集合。移除  $R'$  中的  $r_7$ 、 $r_8$  需求。启动细化机制把  $s$  移向  $p_9$ ，继续执行步骤 2)，对剩下的 POI 进行排序，可以得出剩下所有的 POI 都只能提供一个  $R'$  的需求。在这些 POI 中选择距离  $p_9$  最近的点，为  $p_{10}$ 。然后启动细化机制判断  $p_{10}$  与  $p_9$  距离  $s$  的距离， $TMT(p_9, s) > TMT(p_{10}, s)$  则  $p_9$  不需要移到  $p_{10}$ ，因为  $p_9$  距离  $s$  比  $p_{10}$  距离  $s$  远，该机制是前向扩展所以不需要来回移动。把  $R'$  中的  $r_3$  移除，然后继续扩展  $p_9$ ，发现  $TMT(p_5, p_9) > TMT(p_6, p_9)$ ，选择  $p_5$  作为候选点，启动机制  $\omega$ ，然后把  $p_5$  加入  $Route$ 。把  $R'$  中的  $r_6$  移除，然后继续扩展  $p_5$ ，选择  $p_3$ ，然后就直接到起点  $s$ 。根据算法 1 可以得出  $Route = \{s, p_3, p_5, p_9, p_{10}, d\}$ 。路程时间  $\tau$  为 820，比算法 TMT 提高了 270 个时间单位。

**4.3 基于利益最大化的路径规划算法**

本算法根据利益函数得出基于利益最大化的贪心算法，给定用户签到数据、用户的初始地点以及时间  $(l_q, t_q)$ 、用户需游览的景点总数  $\omega$ ，该算法可利用上述信息建立一条具有  $\omega$  个点的路线使  $Fitness(Route)$  利益最大化。算法 2 为基于利益最大化的路径规划算法。

**算法 2** 基于利益最大化的路径规划算法

输入  $DB(Route)$ 、 $\omega$ 、start\_location =  $(l_q, t_q), t$

输出 profit\_Route

procedure  $Fitness(DB(Route))$

- 1) 初始化  $Route = (l_1 = l_q, t_1 = t_q)$
- 2) for  $i$  to  $\omega$  do:
- 3)  $c_i = \{l_c | l_{i-1} \rightarrow l_c \text{ in } Route\ DB\}$
- 4)  $f_{max} = 0$
- 5) for each  $l_c \in c_i$
- 6)  $Route_{tmp} = Route + \langle (l_c, t_c) \rangle$
- 7) 计算  $Fitness f(Route_{tmp})$
- 8) if  $f(Route_{tmp}) > f_{max}$  do:
- 9)  $Route = Route_{tmp}$
- 10)  $f_{max} = f(Route_{tmp})$
- 11) end if
- 12) end for
- 13) end for
- 14) return  $Route$
- 15) end procedure

基于利益最大化的路径规划算法为贪心算法，

需要证明其满足最优子结构和贪心选择性。

1) 最优子结构证明

使用反证法，假设不存在贪心策略的最优子结构，则存在全局最优解利益值  $F_p$ ，且  $F_p$  中用户经过的某几个节点的路径并不是利益最大的路径。假设它们的利益值和为  $F_1$ ，剩下部分节点的利益值和为  $F_t$ ，即  $F_p = F_1 + F_t$ 。因为经过这些节点的路径不是最短路径，所以总能找到某条路径使经过这些点的利益值更大，即  $F'_1 < F_1$ 。则存在  $F'_1 + F_t < F_1 + F_t$ ，即存在一个新的最优解比原解更小，与假设矛盾，则该算法最优子结构得证。

2) 贪心选择性证明

设  $M = \{l_1, l_2, \dots, l_n\}$  是具有利益权值的序列，且权值的大小按从大到小排列，每次选择都是从  $M$  中选择利益最大的地点。

采用归纳法证明。

当  $N=1$  时，假设  $\max(M) = l_i$ ，则有  $Fitness(l_i) > \forall_{p \in N \& p \neq i} Fitness(l_p)$ 。

假设当  $N = \omega$  时， $Fitness(Route)$  也成立，其中， $Fitness(Route) = Profit(Route_p)$  且  $Route = \{l_1, l_2, \dots, l_p\}$ 。当  $N = p+1$  时， $Fitness(Route) = Profit(Route_p) + Fitness(l_{p+1}) > Profit(Route_p) + Fitness(\forall_{p \in N \& p \neq i} l_i)$ ，所以  $Fitness(Route)$  成立，即该算法贪心选择性成立。

5 实验与分析

本文实验环境为上海地区景点的模拟环境，实验参数如表 4 所示。其中，每个地点都具有其所特有的服务项，同时为了方便计算，统一把每个地点的服务需求时间设置为 3 min。网络区域点的个数和用户需求种类的个数以及总服务种类的个数都按照实际的情况进行设置，同时这些属性设置对于本文的算法结果不会产生实质性的影响，即不会改变算法相较于其他算法的效率。

表 4 实验参数

实验数据	默认值
需求服务时间/min	3
网络区域中点的个数	100
用户的需求种类个数	10
总服务种类个数	25
$\alpha$	0.02
$k$	2

对于景点人数预测，本文搜集了从 2017 年 1 月

12 日—4 月 18 日的上海方塔园游客数据，分别对于其数据进行模拟和预测，得出平均每个时刻的估计值与真实值的偏差。

将残差值进行 GM(1,1) 建模，得到残差值的 GM(1,1) 模型。然后将未加入残差参数项的预测值和加入偏差参数项的预测值进行比较，结果如图 5 所示。由图 5 可以看出，本文加入残差项的 GM(1,1) 模型比原来的模型偏差更小且稳定。

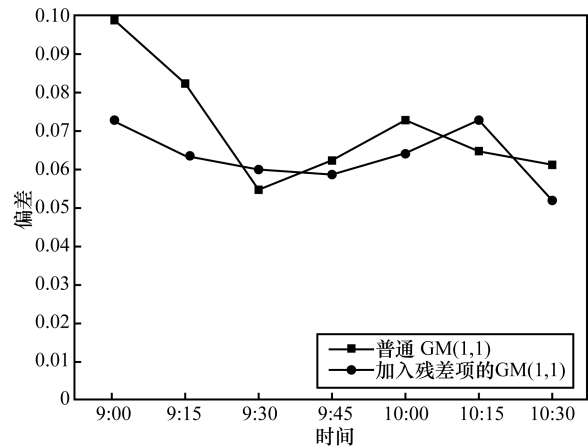


图 5 加入残差项的 GM(1,1) 与普通 GM(1,1) 比较

本文对不同的数据序列长度  $n$  值进行实验。由于公众号接口的数据从 7:00 开始统计，为了保证实验数据的预测准确性，选取序列长度 6~10 作为实验参数，图 6 为  $n$  值分别为 6、8、10 时的实验结果。由图 6 可以看出，当  $n=8$  时，数据预测性能较好。原因是  $n=6$  时数据序列长度较短，对于数据的预测效果较差；但是  $n=10$  时，即在 9:30 时会产生景区人数激增现象，造成整体的数据序列波动较大，这对于预测效果造成很大影响。

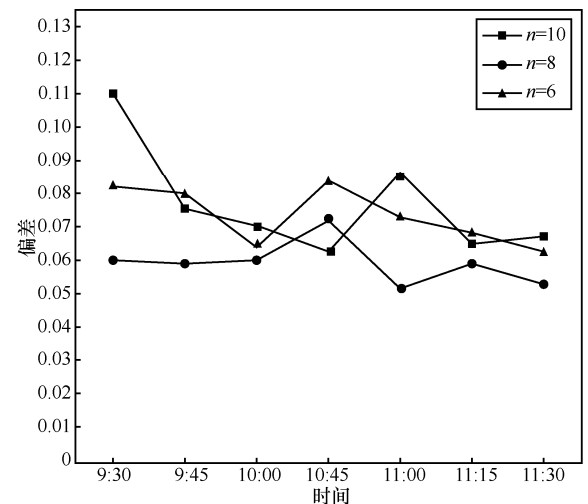


图 6 不同  $n$  值的模拟值偏差

将本文的前向细化算法与文献[2]的  $k$ -near 和 TMT 算法进行比较, 实验评价的主要标准为路线移动时间和地点停留时间之和。分别根据  $k$  值以及用户需求个数来得出实验结果。把  $k$  的默认值设为 2, 该值一般都默认作为实验中设定的值,  $\alpha$  为路程时间利益的调整参数, 可以根据其他属性对  $\alpha$  进行调整, 所以本文把  $\alpha$  设置为 0.02, 其目的是为了均衡路程时间相较于其他的利益因素。本文对景点的利益参数进行设置, 把景点的流行度、时间 KL 散度、访问次序都分为 3 个区间, 每个区间对应不同的利益值。同时为了便于计算, 统一把每个景点的排队服务时间都设置为 5 人/分钟, 景点的平均单位时间到达人数为 2~8 人/分钟。景点预测人数越多代表平均单位时间到达人数越多, 其中, Rank 1-2 为 6~8 人/分钟的随机, Rank 3-5 为 4~6 人/分钟的随机, Rank 6-9 为 2~3 人/分钟的随机。对于流行度、时间 KL 散度、访问次序分别设置成 (5,4,3) 这 3 个利益级别, 是为了说明 3 个属性具有相同的重要性和参考度, 这些属性的取值对实验结果以及和其他算法比较性能不会产生显著影响。后一项路程时间的利益值为  $time \times \alpha$ , 利益参数如表 5 所示。

表 5 利益参数

利益项	取值
景区流行度	Rank 1-2(5)、Rank 3-5(4)、Rank 6-9(3)
时间 KL 散度	[0,3](5)、(3,6)(4)、(6,10)(3)
访问次序	[0.7,1](5)、[0.4,0.6](4)、[0-0.3](3)
时间限制/min	120、240、360
排队服务时间 (QS_Time) (人·分钟 <sup>-1</sup> )	5
景点人数到达速度 $\mu$ (人·分钟 <sup>-1</sup> )	2~8
路程时间	$time \times \alpha$

图 7 和图 8 为  $k$  在 2~7 条件下的各实验比较结果。为了让实验结果更具有代表性, 本文在图上随机设定一个起点、一个终点。3 种算法分别从路程时间和程序运行时间的结果进行比较, 可以得出本文所提算法较  $k$ -near 和 TMT 算法效果更好。

将 Request 个数从 6 到 10 进行实验, 结果如图 9 所示。本文所提算法的效果更好是因为引入了细化机制概念, 从而把不必要的环节删除。

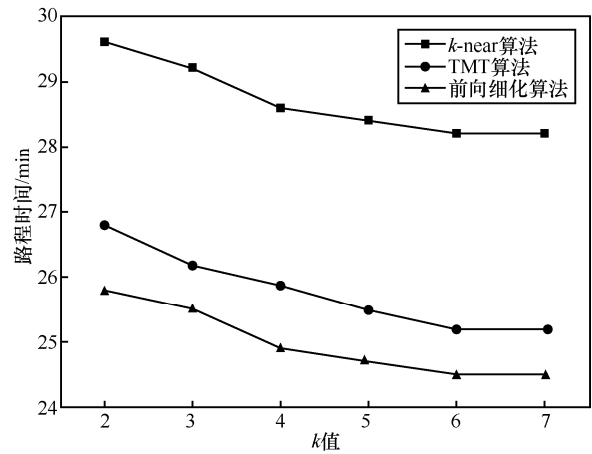


图 7 不同 k 值下的路程时间结果比较

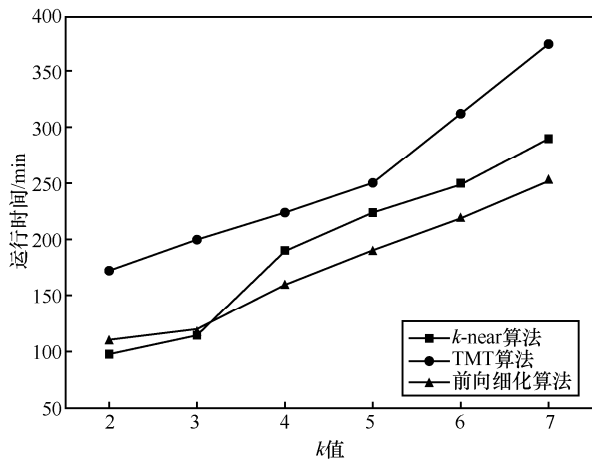


图 8 不同 k 值下的程序运行时间结果比较

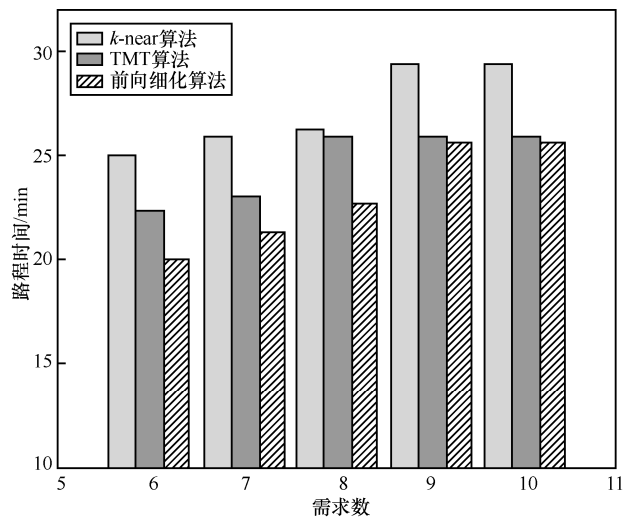


图 9 不同 Request 值下的路程时间结果比较

本文基于利益最大化算法 (Profit 算法) 在需游览完随机  $|Route|$  个景点的前提下, 在不同时间点 (8:00、12:00 以及 18:00) 与文献[7]的 Rand\_GA 算

法和文献[16]的 Time\_Based 算法进行比较。这样比较的意义在于在不同时间点，景点具有不同的流行度（因为景点流行度具有时效性，有些景点在早晨吸引人，有些景点在下午或晚上吸引人），造成景点的人数产生变化，即造成路程时间产生变化，结果如图 10 所示。

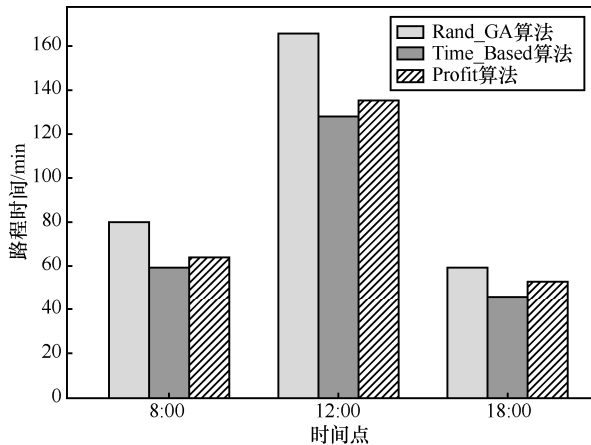


图 10 不同时间点下的路程时间

由图 10 可以看出，Time\_Based 算法的时间比本文所提算法时间略优，主要原因是 Time\_Based 算法只考虑时间的规划而不考虑用户的体验度。Rand\_GA 算法的路程时间高的原因是它没有考虑动态的流量这个因素。

图 11 为 3 种算法在不同时间点下的路径利益。由图 11 可以看出，本文所提算法所产生路径的利益值（为式(13)的  $Fitness(Route)$  值）比其他 2 种算法高，这是因为本文考虑了多重属性的利益，而 Time\_Based 只考虑了路程时间，Rand\_GA 只考虑了路径长度。同时在不同时间限制内，本文所提算法能更多地拓展排名靠前的景点，因为本文考虑了景点流行度等因素，更符合出行需求，结果如图 12 所示。

## 6 结束语

本文研究了基于用户需求的利益路径规划算法，提出了基于灰色马尔可夫的人数预测算法、前向细化算法和基于利益的路径规划算法。其中，景点游客量预测算法采用了灰色模型和隐马尔可夫转移矩阵的结合来预测景区的人数，把预测残差项作为修正项加入原有预测模型中，可以克服灰色模型的本身缺陷（对于波动大的随机序列预测效果差），预测结果可以达到预期效果。基于用户多需求路径规划，在 TMT 算法的基础上提出了细

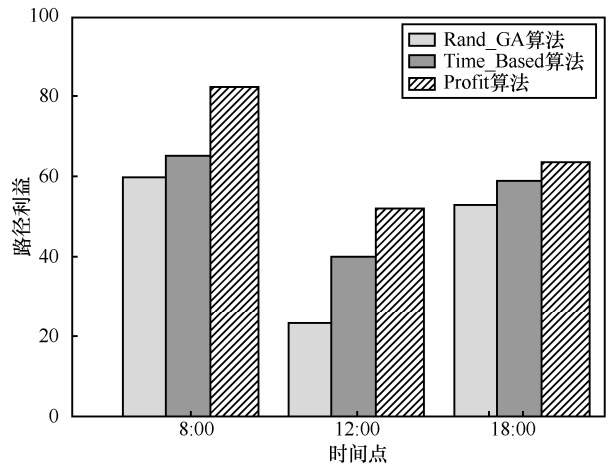


图 11 不同时间点下的路径利益

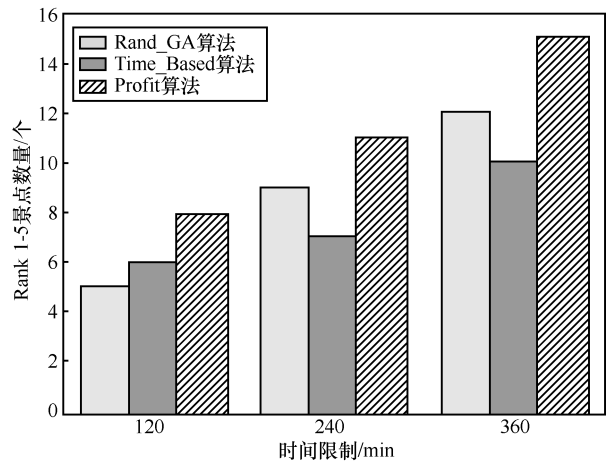


图 12 不同时间限制下的拓展景点数量

化机制，通过从后往前扩展，挑选需求最大的地点作为候选点，实验效果有了很大提高。本文基于多目的地路径规划算法考虑了地点的多个属性：景点流行度、时间 KL 散度、地点访问次序以及路程时间，每个属性代表利益，证明了该算法满足贪心选择策略并根据贪心策略选择利益最大的路径实现局部最优解，最终得到近似优化解，取得了较好的效果。

## 参考文献：

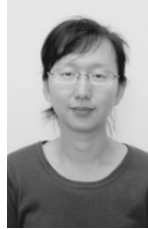
- [1] GOERIGK M, SCHMIDT M. Line planning with user-optimal route choice[J]. European Journal of Operational Research, 2016, 259(2): 1-23.
- [2] LU E H C, CHEN H S, TSENG V S. Efficient approaches for multi-requests route planning in urban areas[C]//IEEE 14th International Conference on Mobile Data Management (MDM). 2013: 36-45.
- [3] LI H, YANG T. Queues with a variable number of servers[J]. European Journal of Operational Research, 2000, 124(3): 615-628.
- [4] BRAJDIC A, HARLE R. Walk detection and step counting on uncon-

- strained smartphones[C]//The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2013: 225-234.
- [5] CHEN C, CHEN X, WANG Z, et al. Scenic planner: planning scenic travel routes leveraging heterogeneous user-generated digital footprints[J]. *Frontiers of Computer Science*, 2017, 11(1): 1-14.
- [6] VANAJAKSHI L, SUBRAMANIAN S C, SIVANANDAN R. Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses[J]. *Intelligent Transport Systems Let*, 2009, 3(1): 1-9.
- [7] YU Y, SZETO K Y. Minimize the average mean first passage time of random walk in complex networks by genetic algorithm[C]// *Evolutionary Computation*. 2016: 2352-2359.
- [8] CARRABS F, CERRONE C, CERULLI R, et al. A novel discretization scheme for the close enough traveling salesman problem[J]. *Computers & Operations Research*, 2017, 78(2): 163-171.
- [9] GAVALAS D, KASAPAKIS V, KONSTANTOPOULOS C, et al. Scenic route planning for tourists[J]. *Personal & Ubiquitous Computing*, 2017, 21(1): 137-155.
- [10] DE PAOLA A, FERRARO P, GAGLIO S, et al. Context-awareness for multi-sensor data fusion in smart environments[M]// *IA 2016 Advances in Artificial Intelligence*. Springer International Publishing. 2016: 377-391.
- [11] HSIEH H P, LI C T, LIN S D. Exploiting large-scale check-in data to recommend time-sensitive routes[C]//The ACM SIGKDD International Workshop on Urban Computing. 2015: 55-62.
- [12] VANSTEENWEGEN P, SOUFFRIAU W, BERGHE G V, et al. The city trip planner: an expert system for tourists[J]. *Expert Systems with Applications*, 2016, 38(6): 6540-6546.
- [13] QUERCIA D, SCHIFANELLA R, AIELLO L M. The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city[C]//ACM Conference on Hypertext and Social Media. 2014: 116-125.
- [14] 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.
- [15] 陈淑燕, 陈家胜. 一种改进的灰色模型在交通量预测中的应用[J]. *公路交通科技*, 2014, 21(2): 81-83.  
CHEN S Y, CHEN J S. Application of an improved grey model in traffic prediction[J]. *Highway Traffic Technology*, 2014, 21(2): 81-83.
- [16] 许博闻. 环境感知的智能场景分析技术的研究[D]. 哈尔滨: 黑龙

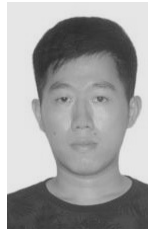
江大学, 2015.

XU B W. Research on intelligent scene analysis technology of environmental perception [D]. Harbin: Heilongjiang University, 2015.

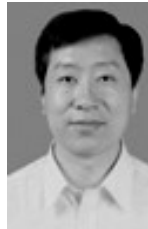
#### [作者简介]



王楠 (1980-), 女, 黑龙江哈尔滨人, 黑龙江大学博士生, 主要研究方向为数据挖掘、无线传感器网。



周红磊 (1992-), 男, 浙江金华人, 黑龙江大学硕士生, 主要研究方向为物联网。



李金宝 (1969-), 男, 黑龙江庆安人, 博士, 黑龙江大学教授、博士生导师, 主要研究方向为无线传感器网络、社交网络、移动计算。



黎玲利 (1986-), 女, 四川广元人, 博士, 黑龙江大学副教授, 主要研究方向为数据质量、大数据管理。