



面向车辆目的地推测的时空搜索优化

韩 磊¹, 於志勇^{1,2,3}, 朱伟平¹, 於志文⁴

(1. 福州大学 数学与计算机科学学院, 福州 350116; 2. 福建省网络计算与智能信息处理重点实验室, 福州 350116;
3. 空间数据挖掘与信息共享教育部重点实验室, 福州 350002; 4. 西北工业大学 计算机学院, 西安 710072)

摘 要: 在仅有车辆起始位置信息的情况下, 车辆目的地推测的准确率通常较低。针对该问题, 通过在城市道路摄像头的视频录像数据中进行时空搜索, 获取目标车辆更多的途经信息, 以更准确地推测出其目的地。为在相同的时空搜索次数下最大化目标车辆目的地推测的准确率, 设计基于概率的单一指标、基于概率和基尼指数的复合指标以及基于概率和信息增益的复合指标, 以评估不同时空搜索对于车辆目的地推测的效用, 并基于 3 种指标分别提出 CFMM-MidQuery、CFMM-UtilityQuery-Gini 和 CFMM-UtilityQuery-Info 算法。实验结果表明, 时空搜索有助于提高车辆目的地推测的准确率, 基于效益的复合指标较基于概率的单一指标评估效果更好, 在时空搜索次数相同的条件下, 两者目的地推测的准确率相差最高达 11.4%。

关键词: 目的地推测; 特征评估; 时空搜索; Markov 模型; 搜索优化

开放科学(资源服务)标志码(OSID):



中文引用格式: 韩磊, 於志勇, 朱伟平, 等. 面向车辆目的地推测的时空搜索优化[J]. 计算机工程, 2020, 46(2): 28-34.

英文引用格式: HAN Lei, YU Zhiyong, ZHU Weiping, et al. Spatiotemporal search optimization for vehicle destination inference[J]. Computer Engineering, 2020, 46(2): 28-34.

Spatiotemporal Search Optimization for Vehicle Destination Inference

HAN Lei¹, YU Zhiyong^{1,2,3}, ZHU Weiping¹, YU Zhiwen⁴

(1. College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China;
2. Fujian Provincial Key Laboratory of Networking Computing and Intelligent Information Processing, Fuzhou 350116, China;
3. Key Laboratory of Spatial Data Mining and Information Sharing, Ministry of Education, Fuzhou 350002, China;
4. School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China)

【Abstract】 The inference of vehicle destination can be inaccurate in the case that only the vehicle's starting position is available. To address this problem, this paper proposes an approach that spatiotemporally searches the video data of city road cameras to obtain more information about the route of the passing vehicle, so as to predict its destination more accurately. In order to maximize the accuracy of target vehicle destination inference under the same spatiotemporal search times, three types of indexes are designed: the probability-based single index, the probability and Gini index-based composite index, and the probability and information gain-based composite index, which are used to evaluate the utility of different spatiotemporal searches for vehicle destination. Further, the CFMM-MidQuery algorithm, the CFMM-UtilityQuery-Gini algorithm and the CFMM-UtilityQuery-Info algorithm are proposed based on the three indexes respectively. Experimental results show that spatiotemporal search can improve the accuracy of vehicle destination inference. The effect of the benefit-based composite indexes is better than that of the probability-based single index. The difference in inference accuracy is as high as 11.4% under the same spatiotemporal search times.

【Key words】 destination inference; feature evaluation; spatiotemporal search; Markov model; search optimization

DOI: 10.19678/j.issn.1000-3428.0053543

基金项目: 国家自然科学基金(61772136)。

作者简介: 韩磊(1992—), 男, 硕士研究生, 主研方向为群智感知; 於志勇(通信作者), 副教授、博士; 朱伟平, 博士研究生; 於志文, 教授、博士。

收稿日期: 2019-01-02 修回日期: 2019-03-13 E-mail: yuzhiyong@fzu.edu.cn

0 概述

随着我国城市化进程的不断加快, 城市建设取得了巨大成就, 但是, 城市人口密度大、人员流动性强的状况也引起了一些公共安全问题, 车辆目的地推测对解决城市公共安全问题具有重要意义。警察获知目标车辆某时刻从某地出发, 其需要尽可能准确地掌握该车辆的目的地以制定行动计划。然而, 在很多情况下, 由于数据丢失、隐私保护等原因, 使得准确推测该车辆的目的地十分困难。此时, 警察可以通过已知的目标车辆位置信息来推测其目的地, 但是由于已知车辆信息太少(在通常情况下, 警察只知道车辆的出发时刻和出发位置), 其目的地推测的准确率通常不高。

目前, 在城市道路中分布着大量进行全天录像的摄像头, 可通过这些视频数据来获取更多关于目标车辆的途经信息, 进而提高车辆目的地推测的准确率。从视频中识别目标车辆的途径信息, 需要一定的人工和机器成本, 且摄像头数量规模庞大, 通过所有的视频数据进行全时全域搜索时难度巨大。因此, 通过尽可能少的时空搜索(通过搜索城市摄像头视频数据来查看某个时刻目标车辆是否途经某个位置)来准确推测出目标车辆的目的地, 引起了研究人员的广泛关注。

本文研究面向目的地推测的时空搜索优化问题, 通过对道路摄像头的所有视频录像进行时空搜索以获取车辆的途经信息, 从而提高车辆目的地推测的准确率。在此基础上, 使用基于概率的单一指标、基于概率和基尼指数的复合指标^[1]以及基于概率和信息增益的复合指标^[2]来评价时空搜索方法的性能。

1 相关工作

针对目标搜索问题, 文献[3]提出并建立了一种海上立体搜寻全局优化模型, 文献[4]就飞机发生事故失去动力后黑匣子落水点的位置问题进行了研究, 文献[5]以马航事件为背景, 基于贝叶斯方法提出一种针对特定区域的失踪目标搜索算法。

基于已知车辆位置信息判断车辆目的地是移动预测的目标, 针对该问题, 文献[6]提出了一种基于移动性近似的目的地预测算法, 该算法关注轨迹本身的运动行为以寻找潜在目的地, 然后利用移动梯度下降法来确定目标的目的地。文献[7]研究了个人社会信息和出行信息对出行目的地选择行为的影响。文献[8]通过调查GPS数据, 建立利用Markov模型进行目的地预测的概率模型。文献[9]根据人

类日常移动路线表现出的时间和空间的规律性, 使用拓展的连续路径模式挖掘算法从轨迹中提取运动模式, 并从该运动模式中构建模式树以进行目的地预测。此外, 在历史信息中(如道路状况、驾驶习惯^[10]和轨迹长度^[11-13]等)加入外部信息也可以提高预测的准确率。文献[14]提出了SMLP算法, 该算法在Markov模型的基础上利用网络内用户的相关性对位置进行预测。

主动感知指智能体采取一定行为以获取更多环境相关信息。在机器视觉^[15]和机器人学领域, 主动感知已经被广泛应用于定位和导航任务^[16]。此外, 主动感知还在同步定位和映射^[17]、越野驾驶^[18]、机器人探索^[19]、雷达目标追踪、声呐和光电探测器^[20]等方面得到应用。基于车辆起始位置, 通过时空搜索获取更多车辆途经信息以提高车辆目的地预测准确率的思想与主动感知^[21]相似。在一个主动感知过程中, 系统会根据环境状态主动执行感知行为^[22]以对环境进行认知。

本文针对城市中的目标车辆搜索问题, 提出一种面向目的地推测的时空搜索优化算法, 该算法主要包括移动预测和主动感知2个部分。传统目的地预测方法无车辆途经点搜索过程, 本文算法增加了这一过程, 以获取更多车辆途经信息。

2 问题定义

为了更好地定义面向目的地推测的时空搜索优化问题, 本文定义如下符号: 令摄像头视频集合 $D = \{d_1, d_2, \dots, d_{|D|}\}$, 位置集合 $L = \{l_1, l_2, \dots, l_{|L|}\}$, 行程集合 $N = \{n_1, n_2, \dots, n_{|N|}\}$, 行程时间序列 $T = \langle t_1, t_2, \dots, t_{|T|} \rangle$, 车辆集合 $C = \{c_1, c_2, \dots, c_{|C|}\}$, 车辆行程轨迹集合 $T_R = \{t_r(c_r, n_u)\}$, 其中, $t_r(c_r, n_u) = \langle (t_1, l_{t1}), (t_2, l_{t2}), \dots \rangle | c_r, n_u$ 。已知车辆 c_r 的出发时间 t_1 和出发位置 l_{t1} 后, 搜索时刻为 t_k , 搜索位置为 l_i , 时空搜索结果 $q(c_r, t_k, l_i)$, 当目标车辆 c_r 在 t_k 时刻出现在位置 l_i 时, $q(c_r, t_k, l_i) = 1$; 否则记为 $q(c_r, t_k, l_i) = 0$, 如式(1)所示。

$$q(c_r, t_k, l_i) = \begin{cases} 1, & (t_k, l_i) \in t_r(c_r, n_u) \\ 0, & \text{其他} \end{cases} \quad (1)$$

I_{Interval} 表示搜索时刻与车辆起始出发时刻的差值, 即 $I_{\text{Interval}} = t_k - t_1$ 。最后, 本文将 $q(c_r, t_k, l_i)$ 实例化后的 t_k, l_i 称为 time-location, 需要考虑 $(t_{|T|} - t_1) \times |L|$ 个候选 time-location。

在一般情况下, 车辆到达目的地后就无法通过城市道路摄像头搜索到该车辆, 因此, 本文假设车辆当前行程结束后短时间内将无法通过摄像头视频数

据再次搜索到该车辆的踪迹。此外,查询视频录像的成本(包括人力和机器成本)远大于模型训练和目的地推测的计算成本,因此,完成车辆目的地推测任务的代价可通过时空搜索次数 N 来预估,使用 A_{Accuracy} 作为模型性能的评估度量, A_{Accuracy} 表示车辆 c_r 在行程 n_u 中目的地 l_x 的推测准确率,计算公式如下:

$$A_{\text{Accuracy}} = \frac{\#t_r(\cdot), l_x = l_{i|T_l}}{\#t_r(\cdot)} \quad (2)$$

其中, $\#t_r(\cdot)$ 表示轨迹条数。

问题(面向目的地推测的时空搜索优化)给出车辆的历史行程轨迹集合 T'_R 、车辆 c_r 行程 n_u 的起始时刻 t_1 、起始位置 l_{i1} ,以及当天可供时空搜索使用的道路摄像头视频录像数据 d_j 。优化目标是在 N 次时空搜索 $q(c_r, t_k, l_i)$ 的条件下最大化车辆 c_r 行程 n_u 的目的地 l_x 推测的准确率 A_{Accuracy} 。优化问题表达式如下:

$$\begin{aligned} & \text{give } c_r, n_u, (t_1, l_{i1}), T'_R, N, q(c_r, t_i, l_k) \text{ in } d_j \\ & \text{return } l_x, \text{ s. t. } \max A_{\text{Accuracy}} \end{aligned} \quad (3)$$

3 车辆目的地推测

假设已知车辆历史行程轨迹 T'_R ,在仅知道车辆某时刻从某地出发的情况下推测该车辆的目的地,方法有以下 2 种:根据车辆的起始位置信息推测出车辆目的地;通过时空搜索获取车辆途经信息,再结合车辆的起始位置信息推测车辆目的地。

对于第 1 种方法,本文基于简单一阶 Markov 模型进行实现,其推测车辆目的地的流程如图 1 所示。

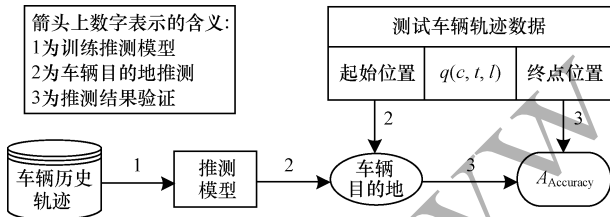


图 1 基于简单一阶 Markov 模型的车辆目的地推测
Fig.1 Vehicle destination inference based on simple first-order Markov model

设车辆位置转移概率(简称 M)表示车辆从一个位置移动到另一个位置的概率, M 计算公式如下:

$$M_{(t_1 \rightarrow t_{lT_l})}^{l_i \times l_x} = \{p(l_i, l_x) | (t_1 \rightarrow t_{lT_l})\} \quad (4)$$

$$p(l_i, l_x) | (t_1 \rightarrow t_{lT_l}) = \frac{\#t_r(\cdot), l_{i1} = l_i \& l_{i|T_l} = l_x}{\#t_r(\cdot), l_{i1} = l_i} \quad (5)$$

对于第 2 种方法,本文基于引入时空搜索的一阶 Markov 模型进行实现,其推测车辆目的地的流程如图 2 所示。

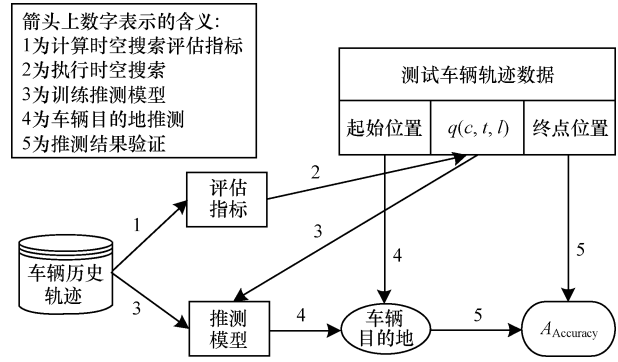


图 2 基于改进一阶 Markov 模型的车辆目的地推测
Fig.2 Vehicle destination inference based on improved first-order Markov model

此时车辆位置转移概率计算如下:

$$M_{(t_1 \rightarrow t_{lT_l}, q(\cdot))}^{l_i \times l_x} = \{p(l_i, l_x) | (t_1 \rightarrow t_{lT_l}, q(\cdot))\} \quad (6)$$

$$p(l_i, l_x) | (t_1 \rightarrow t_{lT_l}, q(\cdot)) = \frac{\#t_r(\cdot), l_{i1} = l_i \& q(\cdot) \& l_{i|T_l} = l_x}{\#t_r(\cdot), l_{i1} = l_i \& q(\cdot)} \quad (7)$$

由于车辆轨迹信息不足(只知车辆出发位置和出发时刻),基于简单一阶 Markov 模型推测车辆目的地的准确率不高,因此本文基于改进一阶 Markov 模型来推测车辆目的地。在时空搜索方式的选择上,本文提出基于概率的单一指标、基于概率和基尼指数的复合指标以及基于概率和信息增益的复合指标来评估不同时空搜索方式的性能,并基于 3 种指标分别实现算法 CFMM-MidQuery、CFMM-UtilityQuery-Gini 和 CFMM-UtilityQuery-Info。

3.1 CFMM-MidQuery 算法

设车辆 c_r 行程的起始时间为 t_1 ,起始位置为 l_i ,首先确定车辆 c_r 在 t_{mid} 时刻的 N 个可能性最大的位置,表示为 $l_q[1], l_q[2], \dots, l_q[j], \dots, l_q[N]$ 。然后执行时空搜索 $q(c_r, t_{\text{mid}}, l_q[j])$,得到 N 个返回结果,这些结果可能包含一个 1(即确定了目标车辆 c_r 在 t_{mid} 时刻的位置为 $l_q[j]$)或者所有结果都是 0(即 t_{mid} 时刻在 $l_q[1], l_q[2], \dots, l_q[j], \dots, l_q[N]$ 这些位置均未找到目标车辆 c_r),然后将 $q(c_r, t_{\text{mid}}, l_q[j])$ 的返回结果作为条件训练一阶 Markov 模型。该过程为 CFMM-MidQuery 算法过程,其伪代码如下。

算法 1 CFMM-MidQuery 算法

输入 T'_R (车辆历史行程轨迹), $\langle t_1, l_{i1} \rangle$ (c_r 的起始时刻和位置), t_{mid} (搜索时刻), N (时空搜索次数)

输出 车辆 c_r 的终点位置

1. 根据 T'_R 计算 t_1 时刻位置为 l_{i1} 的车辆在 t_{mid} 时刻的位置分布概率 LS_m
2. 遍历向量 LS_m , 返回 N 个概率最大的位置编号索引: $l_q[1], l_q[2], \dots, l_q[j], \dots, l_q[N]$
3. 执行时空搜索 $q(c_r, t_{\text{mid}}, l_q[j])$

4. 根据 $T'_R, <t_l, l_{li}>$ 和时空搜索结果计算终点位置分布概率 LS_c 。

5. 根据 LS_c 返回概率最大的位置

6. END

3.2 CFMM-UtilityQuery 算法

CFMM-MidQuery 算法需要先指定搜索时刻, 其不能说明哪个搜索时刻是最佳选择, 而在 CFMM-UtilityQuery 算法中, 按照效益值大小将不同的候选 time-location 进行排序, 选择效益大的 time-location 执行时空搜索, 然后将返回结果作为条件训练模型 (模型构建过程与 CFMM-MidQuery 相同)。在此基础上, 本文分别实现基于概率和基尼指数复合评估指标的 CFMM-UtilityQuery-Gini 算法与基于概率和信息增益复合评估指标的 CFMM-UtilityQuery-Info 算法。

3.2.1 CFMM-UtilityQuery-Gini 算法

假设目标车辆 c_r 的起始时间为 t_1 , 起始位置为 l_{li} 。首先, 基于概率和基尼指数复合评估指标计算 t_k 时刻搜索位置 l_s 对于推测车辆 c_r 目的地的效用, 计算公式如式(8)所示。

$$U_{Gini}^{utility}(t_k, l_s) = p(l_{li}, l_s) | (t_1 \rightarrow t_k) \times (1 - g_{gini}(t_k, l_s)) \quad (8)$$

式(8)由两部分组成: $p(l_{li}, l_s) | (t_1 \rightarrow t_k)$ 表示 $q(c_r, t_k, l_s) = 1$ 的概率, $(1 - g_{gini}(t_k, l_s))$ 用于评估以 $q(c_r, t_k, l_s) = 1$ 作为条件筛选训练数据后车辆 c_r 终点位置分布的混乱程度。若终点位置分布概率为 $P = \{p_1, p_2, \dots, p_k, \dots, p_M\}$, 则原始基尼指数计算公式如式(9)所示。

$$G_{Gini}(P) = \sum_{k=1}^M p_k(1 - p_k) = 1 - \sum_{k=1}^M p_k^2 \quad (9)$$

因此, 本文基尼指数计算公式如式(10)所示。

$$g_{gini}(t_k, l_s) = g_{gini} \{ p(l_{li}, l_s) | (t_1 \rightarrow t_{|T|}), q(c_r, t_k, l_s) = 1 \} \quad (10)$$

CFMM-UtilityQuery-Gini 算法伪代码如下:

算法2 CFMM-UtilityQuery-Gini 算法

输入 T'_R (车辆历史行程轨迹), $<t_l, l_{li}>$ (车辆 c_r 的起始时刻和位置), N (时空搜索次数)

输出 车辆 c_r 的终点位置

1. FOR $t_{mid} \leftarrow t_l$ TO $t_{|T|}$

2. 根据 T'_R 计算 t_l 时刻位置为 l_{li} 的车辆在 t_{mid} 时刻的位置分布概率 LS_m

3. FOR $l_{mid} \leftarrow l_l$ TO $l_{|L|}$

4. 根据 T'_R 计算 t_l 时刻位置为 l_{li} 、 t_{mid} 时刻位置为 l_{mid} 的车辆终点位置分布概率 LS_{me}

5. END FOR

6. 根据式(8)计算效益 $U_{Gini}^{utility}$

7. END FOR

8. 遍历矩阵 $(U_{Gini}^{utility})^{T \times L}$, 返回 N 个时刻和位置:

$<t_1, l_q[1]>, <t_2, l_q[2]>, \dots, <t_i, l_q[i]>, \dots, <t_N, l_q[N]>$

9. 执行时空搜索 $q(c_r, t_i, l_q[i])$ 并且返回搜索结果

10. 根据 $T'_R, <t_l, l_{li}>$ 和时空搜索结果计算终点位置分布概率 LS_c 。

11. 返回 LS_c 中概率最大的位置

12. END

3.2.2 CFMM-UtilityQuery-Info 算法

假设目标车辆 c_r 的起始时间为 t_1 , 起始位置为 l_{li} 。首先, 基于概率和信息增益复合评估指标计算 t_k 时刻搜索位置 l_s 对于推测车辆 c_r 目的地的效用, 其计算公式如式(11)所示。

$$U_{Info}^{utility}(t_k, l_s) = p(l_{li}, l_s) | (t_1 \rightarrow t_k) \times I_{info}(t_k, l_s) \quad (11)$$

其中, $I_{info}(t_k, l_s)$ 评估 $q(c_r, t_k, l_s) = 1$ 作为条件筛选训练数据后车辆终点位置分布的混乱程度。若终点位置分布概率为 $P = \{p_1, p_2, \dots, p_k, \dots, p_M\}$, 则熵计算公式如式(12)所示。

$$E(P) = - \sum_{k=1}^M p_k \lg(p_k) \quad (12)$$

因此, 本文信息增益计算公式如式(13)所示。

$$I_{info}(t_k, l_s) = E \{ p(l_{li}, l_{r|T_l}) | (t_1 \rightarrow t_{|T|}) \} - E \{ p(l_{li}, l_{r|T_l}) | (t_1 \rightarrow t_{|T|}), q(c_r, t_k, l_s) = 1 \} \quad (13)$$

CFMM-UtilityQuery-Info 算法伪代码如下:

算法3 CFMM-UtilityQuery-Info 算法

输入 T'_R (车辆历史行程轨迹), $<t_l, l_{li}>$ (车辆 c_r 的起始时刻和位置), N (时空搜索次数)

输出 车辆 c_r 的终点位置

1. 根据 T'_R 计算 t_l 时刻位置为 l_{li} 的车辆终点位置分布概率 LS_{sc}

2. 根据 LS_{sc} 和式(12)计算熵 E_l

3. FOR $t_{mid} \leftarrow t_l$ TO $t_{|T|}$

4. 根据 T'_R 计算 t_l 时刻位置为 l_{li} 的车辆在 t_{mid} 时刻的位置分布概率 LS_m

5. FOR $l_{mid} \leftarrow l_l$ TO $l_{|L|}$

6. 根据 T'_R 计算 t_l 时刻位置为 l_{li} 、 t_{mid} 时刻位置为 l_{mid} 的车辆终点位置分布概率 LS_{me}

7. 根据 LS_{me} 和式(12)计算熵 E_2

8. END FOR

9. 根据 E_l, E_2, LS_m 计算效益 $U_{Info}^{utility}$

10. END FOR

11. 遍历矩阵 $(U_{Info}^{utility})^{T \times L}$, 返回 N 个时刻和位置: $<t_1, l_q[1]>, <t_2, l_q[2]>, \dots, <t_i, l_q[i]>, \dots, <t_N, l_q[N]>$

12. 执行时空搜索 $q(c_r, t_i, l_q[i])$

13. 根据 $T'_R, <t_l, l_{li}>$ 和时空搜索结果计算终点位置分布概率 LS_c 。

14. 返回 LS_c 中概率最大的位置

15. END

4 实验结果与分析

4.1 原始数据描述

本文的实验数据来自滴滴出行平台 (<https://gaia.didichuxing.com>), 原始数据为成都市部分区域的滴滴订单数据, 采集区域的面积大小约为 $8.37 \text{ km} \times 8.35 \text{ km}$, 采集日期范围为 2016 年 11 月 1 日—2016 年 11 月 30 日, 原始数据信息如表 1 所示。

表 1 原始数据信息

Table 1 Raw data information

字段	示例	备注
司机编号	glox.jrrltBMvCh8nxqktdr2dtopmlH	—
订单 ID	jkkt8kxniolvIFuns9qrrlvst@iqnpkwz	—
时间戳/s	1 501 584 540	unix 时间戳
经度/(°)	104.112 25	GCJ-02 坐标系
纬度/(°)	30.686 30	GCJ-02 坐标系

4.2 实验数据预处理

在数据集中, 不同车辆的 GPS 点是无序的, 需要对数据集中车辆订单 ID 和时间使用二级排序算法生成出租车的 GPS 点序列。由于原始数据时间精度为 3 s, 如果同一行程轨迹相邻两 GPS 点的时间相差超过 3 s, 则认为该轨迹数据部分缺失, 按照 3 s 的时间精度补全缺失数据。如果缺失时长不超过 60 s, 则在该段缺失时间内车辆被认为是匀速直线行驶; 若缺失时间超过 60 s, 则认为该轨迹缺失严重, 将其舍弃。

在排好序的出租车 GPS 点序列中, 每辆车每分钟包括若干个 GPS 坐标点, 且不同分钟内 GPS 点的个数不同, 因此, 本文以分钟为单位将时间离散化, 每分钟仅取时间最小的一条位置数据表示车辆在该时间的位置。最后根据数据范围将该区域划分成大小约为 $0.93 \text{ km} \times 0.93 \text{ km}$ 的方格^[23] (共计 81 个网格), 将车辆的轨迹数据投影到网格中, 一个网格就代表一个位置区域。

4.3 结果分析

本文基于相同的实验数据集, 测试不同阶 Markov 模型推测车辆目的地的准确率, 结果如图 3 所示。从图 3 可以看出, 当 Markov 模型阶数达到 3 时, 车辆目的地推测的准确率达到最高 (2 阶 Markov 模型和 3 阶 Markov 模型在推测车辆目的地时的准确率基本一致), 然后随着模型阶数的增加其推测准确率逐渐降低, 因此, 当有多个时空搜索返回结果为 1 时, 本文仅选择一个 time-location 参与模型的训练与预测。

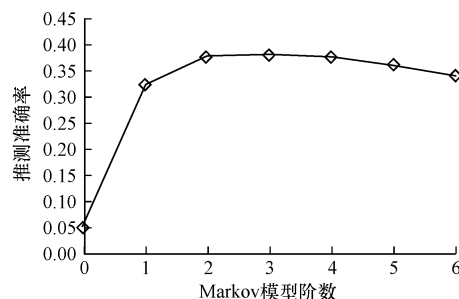


图 3 不同阶数 Markov 模型的推测性能对比

Fig. 3 Comparison of inference performance of Markov models with different orders

为研究相同数据集下车辆目的地推测准确率与 I_{Interval} 的关系, 本文基于二阶 Markov 模型改变 I_{Interval} (固定起始时间, 改变中间搜索时刻), 测试模型推测准确率, 结果如图 4 所示。从图 4 可以看出, 随着 I_{Interval} 的增大, 车辆目的地推测的准确率也逐渐提高。因此, 当有多个时空搜索返回结果为 1 时, 本文仅选择 I_{Interval} 最大的 time-location 参与模型的训练与预测。

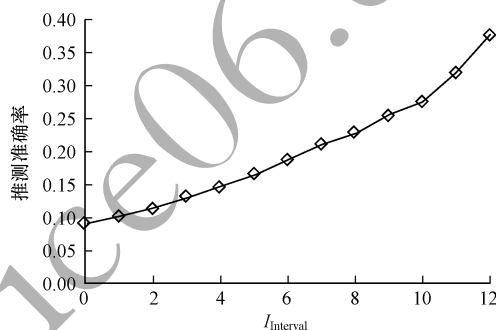


图 4 不同 I_{Interval} 时的二阶 Markov 模型推测准确率

Fig. 4 Inference accuracy of second-order Markov model with different I_{Interval}

图 5 所示为不同行程时长下的轨迹统计情况, 由图 5 可以看出, 随着时长的增加, 大于该时长的轨迹比例逐渐减小, 即在搜索时刻逐渐增大时, 通过摄像头视频录像找到车辆的可能性将降低。

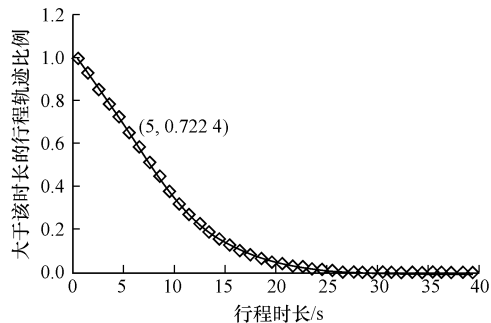


图 5 不同行程时长下的轨迹统计情况

Fig. 5 Track statistics under different travel durations

由图4可以看出,随着 I_{Interval} 的增大,车辆目的地推测的准确率逐渐提高;由图5可以看出,随着搜索时刻的增大,行程时长大于该搜索时刻的轨迹比例逐渐减小。因此,CFMM-MidQuery算法很难选择合适的中间搜索时刻。本文基于二阶Markov模型,测试相同搜索次数下不同中间搜索时刻时车辆目的地推测的准确率,结果如图6所示,从图6可以看出,在相同搜索次数下, I_{Interval} 为5 min时,CFMM-MidQuery算法推测车辆目的地的准确率最高。

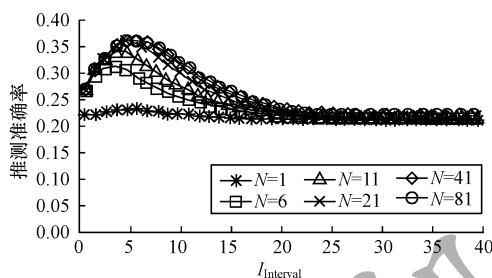


图6 不同搜索时刻下CFMM-MidQuery算法推测准确率
Fig.6 Inference accuracy of CFMM-MidQuery algorithm at different search times

由图6可以看出,当搜索时刻为5 min时,CFMM-MidQuery算法的性能较好,因此,将CFMM-MidQuery算法的搜索时间定为5 min,本文称之为CFMM-MidQuery-5算法。比较CFMM-MidQuery-5、CFMM-UtilityQuery-Info、CFMM-UtilityQuery-Gini 3种算法,结果如图7所示。从图7可以看出,当 N 较小时,CFMM-UtilityQuery-Gini和CFMM-MidQuery-5算法推测车辆目的地的准确率没有太大区别,且2种算法略优于CFMM-UtilityQuery-Info算法;当 N 进一步增大时,CFMM-UtilityQuery算法推测车辆目的地的准确率明显高于CFMM-MidQuery-5算法,而CFMM-UtilityQuery-Info、CFMM-UtilityQuery-Gini 2种算法在推测车辆目的地的准确率方面无明显区别。基于概率单一指标的方法只考虑当前时空搜索找到目标车辆的概率,不考虑当前时空搜索对车辆目的地推测的效用,并且该指标需要指定搜索时刻,限制了时空搜索范围,导致时空搜索次数较高时无法执行更合适的时空搜索;信息增益和基尼指数都是评价车辆位置分布混乱程度的度量,都能在客观上反映车辆位置的概率分布情况,因此,这2种复合指标在推测车辆目的地的准确率方面无明显区别。

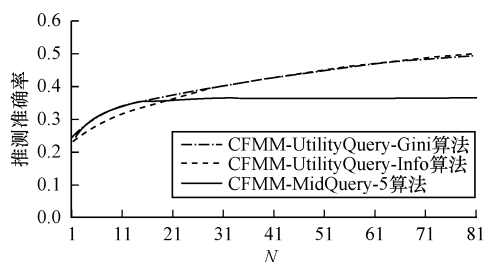


图7 不同搜索次数下3种算法车辆目的地推测准确率对比
Fig.7 Comparison of inference accuracy of vehicle destination using 3 algorithms under different search times

5 结束语

获取目标车辆更多的行程途经信息可以提高车辆目的地推测的准确率。本文提出基于概率和基尼指数的复合指标与基于概率和信息增益的复合指标,不仅考虑当前时空搜索下找到目标车辆的可能性,还兼顾当前时空搜索对目的地推测的长期效用。复合指标无需指定搜索时间,可以将不同时间下的空间搜索进行比较,解决了概率单一指标需要指定搜索时间的问题。实验结果表明,在高搜索次数条件下,复合指标的评估效果明显优于概率单一指标。

本文假设车辆行程结束后短时间内无法通过视频数据再次搜索到该车辆的踪迹,该假设过于理想,下一步将基于车辆一天中的完整轨迹在给定时刻进行车辆位置搜索,以解决上述假设的局限性问题。

参考文献

- [1] KALMIJN W. Gini coefficient [EB/OL]. [2018-12-20]. https://link.springer.com/referenceworkentry/10.1007/978-94-007-0753-5_1168.
- [2] BERRAR D, DUBITZKY W. Information gain [M]. Berlin, Germany: Springer, 2013.
- [3] XIN Shengwei. Research on global optimization model and simulation of joint aeronautical and maritime search [D]. Dalian: Dalian Maritime University, 2012. (in Chinese)
邢胜伟. 海上立体搜寻全局优化模型及仿真研究 [D]. 大连: 大连海事大学, 2012.
- [4] JING Liangping, MU Shengdong, JIANG Lixin, et al. Research on the problem of black box landing point based on dynamics [J]. Science and Technology Information, 2015, 13(1): 205-206. (in Chinese)
晋良平, 穆盛东, 蒋栋鑫, 等. 基于动力学探究黑匣子降落点问题 [J]. 科技资讯, 2015, 13(1): 205-206.
- [5] YU Mei, XU Zijian. Optimal search algorithm for missing target based on Bayesian approach [J]. Computer and Modernization, 2016(10): 21-24. (in Chinese)

- 于美,徐子健.基于贝叶斯方法的失踪目标优化搜索算法[J].计算机与现代化,2016(10):21-24.
- [6] SIMMONS R, BROWNING B, ZHANG Y L, et al. Learning to predict driver route and destination Intent [C]//Proceedings of IEEE Intelligent Transportation Systems Conference. Washington D. C., USA: IEEE Press, 2006: 127-132.
- [7] GAO Jingxin, JUAN Zhicai, NI Anning. Modeling and applications of traveler destination choice behavior based on Bayesian network [J]. Journal of Systems and Management, 2015, 24(1): 32-37. (in Chinese)
高晶鑫, 隗志才, 倪安宁. 基于贝叶斯网络的出行者目的地选择行为建模与应用[J]. 系统管理学报, 2015, 24(1): 32-37.
- [8] GAO Jingxin, JUAN Zhicai, DONG Yingrong. The model and simulation process design for residents travel destination choice[EB/OL]. [2018-12-20]. <https://www.ixueshu.com/document/dcbe02a0d190bb3fe6c6ecc99b596103318947a18e7f9386.html>. (in Chinese)
高晶鑫, 隗志才, 董英荣. 居民出行目的地选择模型与仿真流程设计[EB/OL]. [2018-12-20]. <https://www.ixueshu.com/document/dcbe02a0d190bb3fe6c6ecc99b596103318947a18e7f9386.html>.
- [9] CHEN Ling, LÜ Mingqi, CHEN Gencai. A system for destination and future route prediction based on trajectory mining[J]. Pervasive and Mobile Computing, 2010, 6(6): 657-676.
- [10] ZIEBART B D, MAAS A L, DEY A K, et al. Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior [C]//Proceedings of UBIComp '08. New York, USA: ACM Press, 2008: 322-331.
- [11] HORVITZ E, KRUMM J. Some help on the way: opportunistic routing under uncertainty [C]//Proceedings of ACM Conference on Ubiquitous Computing. New York, USA: ACM Press, 2012: 371-380.
- [12] KRUMM J, HORVITZ E. Predestination: inferring destinations from partial trajectories [C]//Proceedings of International Conference on Ubiquitous Computing. Berlin, Germany: Springer, 2006: 243-260.
- [13] KRUMM J, HORVITZ E. Predestination: where do you want to go today? [J]. IEEE Computer, 2007, 40(4): 105-107.
- [14] YU Ruiyun, XIA Xingyou, LI Jie, et al. Position prediction algorithm for mobile users based on social relations in participatory perception systems[J]. Chinese Journal of Computers, 2015, 38(2): 374-385. (in Chinese)
于瑞云, 夏兴有, 李婕, 等. 参与式感知系统中基于社会关系的移动用户位置预测算法[J]. 计算机学报, 2015, 38(2): 374-385.
- [15] UPTON K. A modern approach[J]. Manufacturing Engineer, 1995, 74(3): 111-113.
- [16] ZHOU H J, SAKANE S. Mobile robot localization using active sensing based on Bayesian network inference[J]. Robotics and Autonomous Systems, 2007, 55(4): 292-305.
- [17] DAVISON A J, MURRAY D W. Simultaneous localization and map-building using active vision [M]. Washington D. C., USA: IEEE Computer Society, 2002.
- [18] PATEL K, MACKLEM W, THRUN S, et al. Active sensing for high-speed offroad driving [C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2005: 262-280.
- [19] PEDERSENT L, WAGNER M D, APOSTOLOPOULOS D, et al. Autonomous robotic meteorite identification in Antarctica [C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2001: 23-52.
- [20] HERO A O, KASTELLA K, CASTANON D, et al. Foundations and applications of sensor management [M]. Berlin, Germany: Springer, 2008.
- [21] BAJCSY R. Active perception [J]. Proceedings of the IEEE, 1988, 76(8): 966-1005.
- [22] GOSANGI R, GUTIERREZOSUNA R. Active temperature modulation of metal-oxide sensors for quantitative analysis of gas mixtures [J]. Sensors and Actuators B: Chemical, 2013, 185: 201-210.
- [23] XUE A Y, ZHANG R, ZHENG Y, et al. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction [C]//Proceedings of IEEE International Conference on Data Engineering. Washington D. C., USA: IEEE Press, 2013: 254-265.

编辑 吴云芳