

文章编号: 2095-2163(2019)01-0088-04

中图分类号: U491.111.2

文献标志码: A

基于信息熵动态拟合的路径通行时间预测方法

刘宇博¹, 李保珠²

(1 济南市历城第二中学, 济南 250104; 2 济南大学 信息科学与工程学院, 济南 250022)

摘要: 文章采用基于信息熵动态拟合方法实现通行时间预测。首先, 基于已有行程时间研究方法对各路段交通状态和通行时间计算, 随后采用最小二乘法获取待预测路径的历史典型相似路径样本, 进而获取待预测路径的各路段的通行参数并外推计算获取其通行时间, 最后, 基于信息熵的理念对通行路况进行动态拟合。实验表明该方法的可行性和准确性, 从而为交通出行路径的规划, 提供稳定可靠的指导建议。

关键词: 最小二乘法; 熵权赋值; 外推拟合数据; 路径通行时间预测

Travel time estimation and prediction for urban road based on dynamic fitting of information entropy

LIU Yubo¹, LI Baozhu²

(1 Jinan Licheng No.2 High School, Jinan 250104, China;

2 School of Information Science and Engineering, University of Jinan, Jinan 250022, China)

【Abstract】 The paper uses the method of dynamic fitting of information entropy to achieve prediction. Firstly, historical data collection is carried out for urban roads, and the traffic state and transit time of each road section are calculated based on the existing travel time research method. Subsequently, the least squares method is used to obtain the historical similar samples of the path to be predicted, and then the traffic parameters of each road segment to be predicted are obtained and extrapolated to obtain the transit time. Finally, based on the concept of information entropy, the traffic conditions are dynamically fitted. Experiments show the feasibility and accuracy of the method, and provide stable and reliable guidance for the planning of traffic routes.

【Key words】 least squares method; entropy weight assignment; extrapolation fitting data; path transit time prediction

0 引言

随着城市规模的不断扩大和城市人口持续增长, 市内交通出行压力越来越大^[1]。传统的路径通行时间预测, 往往基于当前各路段的采集信息, 通过平均通行时间进行总体行程时间预测。然而随着车辆的前进和时间的推移, 尤其下游路段往往路况多变^[2], 由此总体行程时间预测呈现出2个显著问题:

- (1) 车辆行程中途到达后期各路段的时间未知;
- (2) 未来一段时间内, 出行路径的路况信息未知^[3]。

针对上述问题, 本文的主要贡献如下:

从时间与空间相结合的角度, 基于历史典型相似路径样本数据。获取以历史时间为纵轴的各路段的通行时间。随后根据空间车辆的行进, 确定下一路段的到达时间, 以到达时间为横轴进行历史时间纵轴的数据截取, 根据截取数据采用信息熵来判断

路况的变化。

通过对不同通行路段具体通行时间的划分, 基于采集数据计算获取城市不同分类路段的通行时间, 并得到路段通行时间待定参数, 从而实现基于实时采集数据获取路段通行时间^[4]。随后基于信息熵实现路况变化状态的判别^[5], 若信息熵较小, 说明当前路况稳定, 可实时采集数据实现路径通行时间的预测, 若信息熵较大, 说明在车辆通行时间过程中路况状态变化, 此时实时采集数据失效。而历史典型相似路径数据更能体现其实际通行时间, 故通行时间更加依赖于历史典型相似路径数据, 从而实现基于路况变化的动态路径通行时间预测。

1 通行时间预测整体流程描述

传统的路径通行时间的预测, 往往根据当前的路况采集数据, 对某一路径的通行时间进行静态的预测。然而随着车辆的行进, 时间不断推移, 各路段的实时交通状态亦不断发生变化^[6], 由此基于静态

基金项目: 国家自然科学基金(61672262)。

作者简介: 刘宇博(2001-), 男, 济南市历城第二中学学生。主要研究方向: 智能计算、机器学习。

收稿日期: 2018-11-16

预测的路径通行时间预测与实际通行时间存在较大的差异。

本文采用基于信息熵动态拟合的路径通行时间预测整体流程如图 1 所示。一方面基于当前的采集数据进行稳定路况的通行时间预测^[7];另一方面基于历史数据对下游路段的未来通行时间进行预测,从而提高模型的整体预测精准度。

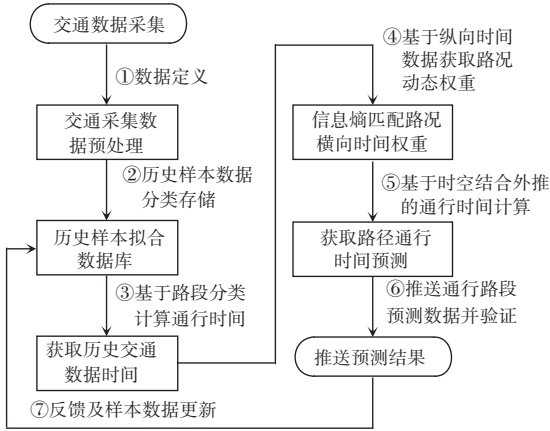


图 1 时间预测整体流程图

Fig. 1 Overall flow chart of time prediction

2 基于信息熵动态拟合的时间预测

针对路况随时间变化而导致实时采集数据失效的问题,本文采用信息熵对路况状态进行精细化拟合,从而提高通行时间预估的准确性。

2.1 获取历史典型相似路径

基于历史数据采集样本对当前路径的通行时间进行预测,从历史数据和当前采集数据 2 个维度进行。基于当前路况的当日已采集数据,采用非线性最小二乘优化算法对历史相似典型路径进行搜索,从而提高历史样本数据价值。

将当前路况采集特征数据与历史样本采集特征数据,即计算每个历史采集样本 X' 与当前采集数据 X 的距离,采用非线性最小二乘优化算法:

$$f(X, X') = \min \left\{ \sum_{j=1}^{\lfloor t/\tau \rfloor} \sum_{i=1}^u \|x_i - x'_i\|^2 \right\} \quad (1)$$

其中, u 代表采集样本特征个数; x_i 为当前采集样本 X 的属性值; x'_i 为历史采集数据 X' 的属性值,进入路段的时间点 t 在采集数据时间间隔为 τ 时取整。

2.2 基于信息熵的路况动态拟合

随时间的推移,路况信息变化较大,故本文借鉴信息熵的概念,对路况的通行时间进行客观权重衡量。具体表现为,当路况一直处于顺畅状态或拥堵

状态时,路径的通行时间均相对稳定,而当路况发生变化时导致后期各路段的通行时间变化较大,此时同类历史交通数据的路径通行时间的参考价值更大^[8]。通过求路径通行时间的信息熵来确定实时采集数据的路段通行数据与历史典型相似路径通行时间的权重。

设某样本 X 有 m 个同类日采集样本数据, n 个数据采集时间点,用 $x_{(i,j)}$ 表示第 i 个样本的第 j 个时刻点的车辆通过该路段的行驶时间,则评价指标构成的矩阵 R 为:

$$R = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & \cdots & x_{(1,n)} \\ x_{(2,1)} & x_{(2,2)} & \cdots & x_{(2,n)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{(m,1)} & x_{(m,2)} & \cdots & x_{(m,n)} \end{bmatrix} \quad (2)$$

数据归一化预处理得:

$$y_{(i,j)} = \frac{[x_{(i,j)} - \min x_{(i,j)}]}{[\max x_{(i,j)} - \min x_{(i,j)}]} \quad (3)$$

此时, $y_{(i,j)} \in [0, 1]$, 则第 j 个时刻点的历史车辆通过时间的熵值为:

$$H_j = - \frac{1}{\ln \cdot m} \sum_{i=1}^m \frac{y_{(i,j)}}{\sum_{i=1}^m y_{(i,j)}} \ln \frac{y_{(i,j)}}{\sum_{i=1}^m y_{(i,j)}} \quad (4)$$

此时,熵权的大小代表着第 j 个时刻点,历史车辆通过该路段时间的变化量。若 H_j 较小,则说明该路段在前 j 个时刻点的通行时间变化较大,此刻路况变化较大。反之,若 H_j 较大,则说明该路段在前 j 个时刻点的通行时间变化较小,此刻在路况平稳期。

因此,基于时间数据推进,进入路段的时间点 t 在采集数据时间间隔为 τ 时取整,其前后通行时间变化的权重具体计算公式为:

$$T_t = T_a \cdot w_j + T_b(1 - w_j) \quad (5)$$

其中, w_j 表示的车辆在 j 时刻及其之前该路段的通行变化, T_a 为历史典型相似路径的该路径均值通过时间,并以 $(1 - w_j)$ 的概率维持基于短期相似路段数据拟合的通行时间 T_b ,从而获得该路段的预估通行时间 T_t 。

3 实验及分析

实验内容主要通过 2 方面进行,一方面,基于实时采集数据和历史交通数据,以信息熵动态拟合路径通行时间,计算其拟合数据;另一方面,基于预测结果分析,说明本文方法的优越性。

3.1 数据拟合实现过程

为更好地体现本文提出方法的预测性能,针对

济南市经十路的不同路段进行数据采集。总体路径分布为济南东高速口附近到泉城公园附近,路径信息如图2所示。



图2 济南东高速口到泉城公园路径示意图

Fig. 2 Road map of Jinan East Express way to Quancheng Park

采集历史数据为2018年3月5日到2018年9月2日26周的交通采集数据,实验数据为2018年9月3号(周一)分别在6:30~9:30、13:30~16:30和18:00~21:00,采用线圈检测器采集数据,以5 min为一个信号检测周期。

以2018年9月3号(周一),7:00从济南东高速口到泉城公园路段的通行时间为例,此时基于当前各路段的检测数据,计算获取基于实时采集数据的路径通行时间预测,见表1。

表1 历史短期相似路段样本拟合部分数据

Tab. 1 Historical short-term similar road segment sample fitting part data

	潘庄路段 367 m	高新区路段 558 m	燕山立交桥路段 452 m	泉城公园路段 579 m
T/s	16	68	88	116

随后,以基于熵权法的路况动态拟合未来路况变化。首先,根据历史交通数据对不同路段进行纵向的路况动态权值计算,同样以7:00出发的济南东高速口到泉城公园路段的部分典型路段为例,见表2。

表3为7:00~8:10济南东高速口到泉城公园路段的部分典型路段的纵向历史通行时间,由于出行时间为7点钟,此时各路段均为畅通状态。通过时空外推拟合获取到达各路段的时间点及对应通行时间,表3的阴影车辆行驶时间段的横向时间截断获取历史数据的纵向各通行时间,采用熵权法对计算实时采集数据与历史典型相似路径的权重比例,由于到达潘庄路段基于历史数据的通行时间基本无变化,故此时信息熵为0,故仅基于实时采集数据给出

通行时间预测,而到达高新区路段基于历史数据的通行时间变化较大,此时信息熵为0.6647,到达燕山立交桥路段的信息熵为0.8942,到达泉城公园路段的信息熵为0.9371,故随着时间的推移,较远的路段通行时间变动较大,此时采用基于信息熵比例的动态数据拟合法,即高新区路段、燕山立交桥路段和泉城公园路段的通行时间分别为:

$$T_8^s = 0.6647 \times 75 + (1 - 0.6647) \times 68 = 72.6529 \text{ s}$$

$$T_{14}^s = 0.8942 \times 97 + (1 - 0.8942) \times 88 = 96.0478 \text{ s}$$

$$T_{22}^s = 0.9371 \times 126 + (1 - 0.9371) \times 116 = 125.3710 \text{ s}$$

表2 同类历史交通数据的部分典型路段通行时间

Tab. 2 Partial road sections of similar historical traffic data s

	潘庄 路段 367 m	高新区 路段 558 m	燕山立交桥 路段 452 m	泉城公园 路段 579 m
7:00	18	34	41	52
7:05	18	38	52	63
7:10	19	41	55	65
7:15	26	66	68	82
7:20	28	75	82	92
7:25	32	89	88	108
7:30	33	93	91	115
7:35	34	96	94	123
7:40	36	96	97	122
7:45	43	95	102	120
7:50	48	99	101	126
7:55	51	108	110	125
8:00	48	101	109	129
8:05	44	105	112	130
8:10	45	102	111	126

本文以信息熵度量变化路况信息,实现实时采集数据与历史数据的综合预测。

3.2 预测结果与数据分析

为验证本文所述方法的优越性,根据实验采集数据分别设计以下实验路径:

行程1:9月3日7:00出发,从济南东高速口到泉城公园路段;

行程2:9月3日14:00出发,从济南东高速口到泉城公园路段;

行程3:9月3日19:00出发,从济南东高速口到泉城公园路段。

对上述行程分别采用4种方法进行预测数据对比。

(1)传统方法预测(某导航APP实现);

- (2) 基于实时采集数据的短期相似路段拟合预测;
- (3) 基于历史典型相似路径的拟合预测;
- (4) 本文预测,预测时间与预测误差见表 3。

表 3 实验数据预测值及预测误差

Tab. 3 Experimental data prediction value and prediction error

行程	行程 1		行程 2		行程 3	
	T/s	$e/\%$	T/s	$e/\%$	T/s	$e/\%$
方法 1	3 356	21	3 356	7.5	5 283	17.5
方法 2	3 777	11	3 011	3.6	4 879	8.5
方法 3	4 473	5	3 272	4.8	4 308	4.2
方法 4	4 387	2.9%	3 168	1.5	4 373	2.8
实际数据	4 264	-	3 122	-	4 497	-

其中, T 为行程预测时间, e 为预测误差, 由表 3 可以看出, 在行程 2 时的通行路径均处于平稳期, 即路况分别处于顺畅状态, 传统预测方法在时间上与实际通行时间相差不大, 预测误差维持在 7% 左右。然而当预测时长面临如行程 1 的路况进入早高峰状态, 或者行程 3 路况开始脱离晚高峰的状态时, 传统预测方法基于当前的采集数据难以根据时间的推移对路径通行时间进行有效预测, 而本文方法结合了实时采集数据和历史数据, 分别结合了实时采集数据时效性和历史数据稳定性的优点, 即使路况状态明显变化时, 预测误差仍然在 2% 左右, 可有效实现对未来短期内的城市道路通行时间的预测。综上所述, 本方法具有显著的实用性和优越性。

4 结束语

本文通过对通行路径的定点检测数据获取不同路段的历史交通数据, 并基于分类分段的城市道路计算获取各路段不同状态下的待定参数和历史通行时间^[9]。基于最小二乘法获取实验数据的历史典型相似路径, 随着车辆通行时间的外推, 实时路况不

断发生变化, 此时依赖于实时采集数据难免导致预测的片面性, 故需结合同类样本的历史典型相似路径, 以信息熵度量变化路况信息, 从而实现实时采集数据与历史数据的综合预测, 提高路径通行时间预测的准确性^[10]。实验结果表明, 本文提出的基于信息熵动态拟合的时间预测方法可很好地拟合短期内的城市交通路况, 对路径的通行时间预测相对准确, 从而为出行路径选择提供有效的参考与行程规划。

参考文献

- [1] 《中国公路学报》编辑部. 中国交通工程学术研究综述·2016 [J]. 中国公路学报, 2016, 29(6): 1-161.
- [2] 刘燕. 基于车流量检测的交通灯控制系统设计[J]. 智能计算机与应用, 2018, 8(1): 148-150.
- [3] HARMS I M, BROOKHUIS K A. Dynamic traffic management on a familiar road; Failing to detect changes in variable speed limits [J]. Transportation Research Part F: Psychology and Behaviour. 2016, 38: 37-46
- [4] LADINO A, KIBANGOU A Y, DE WIT C C, et al. A real time forecasting tool for dynamic travel time from clustered time series [J]. Transportation Research Part C: Emerging Technologies. 2017 (80): 216-238.
- [5] CHEN Meng, YU Xiaohui, LIU Yang. Mining moving patterns for predicting next location [J]. Information Systems. 2015, 54(C): 156-168.
- [6] 晏雨婵, 白璘, 武奇生, 等. 基于多指标模糊综合评价的交通拥堵预测与评估[J/OL]. 计算机应用研究: 1-7 [2018-10-31]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20181011.1259.032.html>.
- [7] 戴洪波, 曾献辉. 基于 BP 神经网络的高速公路短时交通流预测 [J]. 智能计算机与应用, 2015, 5(4): 36-38, 41.
- [8] HAO Peng, BAN Xuegang (Jeff), YU J W. Kinematic equation-based vehicle queue location estimation method for signalized intersections using mobile sensor data [J]. Journal of Intelligent Transportation Systems. 2015(3): 256-272
- [9] 孙兆祖. 路径行程时间及其可靠性研究[D]. 成都: 西南交通大学, 2017.
- [10] 王雪菲, 丁维龙. 一种面向高速公路大数据的短时流量预测[J/OL]. 计算机应用: 1-7 [2018-10-31]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20180920.1020.034.html>.

(上接第 87 页)

- [5] 郑振, 甘新年, 王丽媛. 基于自适应波束形成的鱼雷对潜目标识别技术[J]. 智能计算机与应用, 2016, 6(2): 31-33, 37.
- [6] 韩慧鹏, 梁红, 胡旭娟. 自适应 IIR 陷波器在信号检测中的应用 [J]. 弹箭与制导学报, 2008, 28(2): 315-317.
- [7] 刘家亮, 王海燕, 姜喆, 等. 垂直线列阵结构对 PTRM 阵处理空间增益的影响 [J]. 鱼雷技术, 2010, 18(4): 263-267.

- [8] 李春龙, 刘莹. 一种高斯色噪声混响背景的宽带信号检测算法 [J]. 科学技术与工程, 2011, 11(3): 480-483.
- [9] 盛会平, 黄定东, 黄勇. 一种改进的分数阶 Fourier 变换目标检测算法 [J]. 舰船电子工程, 2012, 32(4): 45-47, 67.
- [10] 柳超, 李秀友, 黄勇. 优化的多模型粒子滤波机动微弱目标检测跟踪方法 [J]. 信号处理 2015, 31(9): 1131-1137.