

基于 N 阶近邻路网的车辆行程时间估计模型

章登义, 欧阳黜霏, 吴文李

(武汉大学计算机学院, 湖北武汉 430072)

摘 要: 车联网的提出为智能交通的研究提供了新的交通信息收集技术. 针对短时交通中车辆的路网行程时间估计问题, 提出了基于 N 阶近邻的隐马尔科夫模型, 利用马尔科夫性质来解决道路行程时间的前后关联性问题, 同时考虑不同道路的异构性构建了 N 阶近邻路网模型来模拟路网间的交互影响. 针对短时交通中实时数据更新的问题, 提出基于道路关联性算法, 并结合车联网的采集技术给出了迭代更新模型的方法. 实验表明, 本文提出的方法在短时交通车辆行程时间预测中精度较高, 能够在车辆行进中做出实时预测.

关键词: 行程时间预测; 隐马尔科夫模型; 聚类

中图分类号: TP301 **文献标识码:** A **文章编号:** 0372-2112 (2015)12-2491-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2015.12.022

N -order Neighbor Road Network Based Vehicle Travel Time Estimation Model

ZHANG Deng-yi, OUYANG Chu-fei, WU Wen-li

(School of Computer, Wuhan University, Wuhan, Hubei 430072, China)

Abstract: The development of Internet of vehicles provides a new traffic information collection technique for the study of intelligent transportation. In this article, we propose an N -order hidden Markov model to approach the vehicle travel time prediction problem, utilizing the Markov nature to model the internship of road network. We also promote an N -order neighbor road network to address the heterogeneity of road. A non-trivial update algorithm is applied to handle the real time data approaching issue. We also prove the temporality of the N -order hidden Markov model in travel time prediction. Experimental results on authentic data indicate the effectiveness and accuracy of this approach.

Key words: travel time prediction; hidden Markov model; cluster

1 引言

近年来随着车联网技术的发展, 新的交通信息采集技术的应用使得交通数据的实时性、准确性日益提高, 这对如何充分利用交通数据解决车辆在路网中行程时间估计的问题提出了新的挑战^[1~4]. 以往的行程时间估计方法^[5~7]主要基于历史数据建立模型, 模型训练时间开销较大, 不能达到实时预测的需求, 因此本文主要研究如何基于局部路网的少量历史交通数据建模, 并且能够实时更新交通数据, 以便提供实时、准确的车辆路网行程时间预测.

已知的短时车辆行程时间预测方法分为以下几类: 一类是基于模拟仿真的方法, 包括 Dynamic MIT^[8]、Dynamic SMART^[9]. 基于仿真的车辆行程时间预测算法是建立在校准过后的仿真环境中的, 虽然可以到达较高的精确性和鲁棒性, 但是由于在真实的交通环境中, 如果需要进行动态的交通估计则整体计算性能将大幅下降,

因此不适合做实时预测的研究. 本文另外一类方法则是基于对历史和实时数据统计分析^[10], 如短时行程时间算法^[11], 线性模型^[12], 模式匹配算法^[13], 其中大部分算法相对简单具有良好的可控性, 然而由于交通网络复杂和随机的特性, 这一类算法不适合拥挤条件下对行程时间的预测. 还有一类是基于人工智能的算法^[14], 以人工神经网络算法为主包括前向网络 (feed-forward NNs)、反馈网络 (feed-back NNs) 以及自组织网络 (self-organization NNs) 等. 尽管人工神经网络方法在大规模路网的复杂交通环境下的行程时间预测方面具有相对较高的准确性, 但是它们通常需要应用特定的程序来进行大量运算^[15]. 特别是传统的人工神经网络在进行实时行程时间预测时是高度的非线性和非单调的. 因此, 基于实时信息采集的车辆行程时间预测模型会出现慢收敛和局部优化问题.

本文的主要贡献有: (1) 提出一种 (N -order hidden Markov model, N -HMM) 基于 N 阶近邻路网的隐马尔科

夫模型,通过建立 N 阶近邻的路网模型,建立状态依赖关系,得到模型所需的状态转移矩阵.(2)提出一种基于 Dirichlet Process Mixture Model (DPMM)的自适应聚类算法,通过对状态内分布进行似然估计的方法,计算出各 Dirichlet Process (DP)混合分布和最优混合系数.(3)提出一种利用实时数据动态预测车辆行程时间的方法,通过状态序列判断是否更新预测模型.

2 智能交通中的信息采集方法

现有的 GPS/DR/MM 等的组合定位技术在定位精度、定位效率等方面还有所欠缺.在譬如隧道、室内停车场以及密集建筑物下等卫星覆盖不好的环境中,卫星定位系统效果不佳.而车联网中智能路侧技术可以用于精准车辆定位和车辆信息实时采集,是车联网中一项不可或缺的重要技术.因此,需要对车载定位技术进行改进以提高其运行的鲁棒性.在车联网环境下,基于车辆与基础设施之间的通信,车载单元的位置能够通过路侧单元发射和接受的无线射频信号的范围测量来预估.文献[16]中所提出的智能路侧采集方法可以提高交通数据采集的准确性和实时性,然而如何利用这种采集技术以提高行程时间预测的准确性,使得建立预测模型的开销可控,本文提出了基于该信息采集技术的车辆行程时间预测模型.本文提出的方法将综合解决道路关联性问题 and 实时数据动态更新的问题.

3 基于 N 阶近邻的隐马尔科夫模型

由于前后道路交通流具有关联性,即前一路段对后一路段的交通流具有影响,与马尔科夫性质类似,然而一阶马尔科夫模型仅仅是依赖于前一状态的,与实际道路交通受多条道路影响的情况不符.因此本文针对实际的路网,提出了可控的 N 阶近邻隐马尔科夫模型,研究首先给出了 N 阶路网模型的定义,对所有分段按时序性关联和相似性关联聚类,从而得到初始的状态矩阵;基于 Hierarchical Dirichlet Process-Hidden Markov Model (HDP-HMM)模型对隐状态内的线段分布进行推测,利用初始状态矩阵初始化超参;最后利用动态规划思想迭代产生优化模型.

3.1 N 阶近邻路网模型

交通路网模型是研究城市道路网络的基础^[17,18],针对不同的需求建立路网能够有效的简化数学模型.本节针对道路的关联性给出 N 阶近邻路网模型,并给出基于该路网模型的状态转移矩阵.图 1 中是英国伦敦某区域的路网流量图,对方框中的道路建立图 $G = (V, E, T)$,其中 V 为顶点集, $E \subseteq V \times V$, E 表示所有顶点间边的集合, T 表示该边在所有时刻的行程时间集.则一个顶点 $v_i \in V$ 表示一个十字路口或道路交汇点,边 e_{ij}

$= (v_i, v_j) \in E$ 表示图中一个从顶点 v_i 到顶点 v_j 的路段.定义边 e_{ij} 的一阶近邻为:

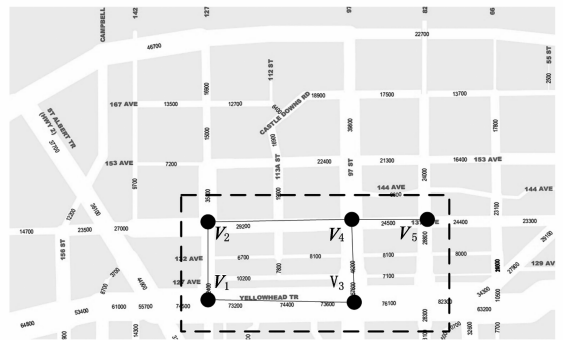


图1 路网图

$$E_i^{(1)} = \{e_{ij} | (j = i \vee i = j) \wedge \neg (j = i \wedge i = j)\} \quad (1)$$

其中 $E_i^{(1)}$ 表示与边 e_{ij} 共享顶点 v_i 或者 v_j 的所有同向边的集合,该集合不包含其反向边 e_{ji} .则图 1 中边 e_{34} 的一阶近邻为 $E_3^{(1)} = \{e_{13}, e_{45}, e_{42}\}$,其中边 e_{43} 为反向边被排除, e_{31} 和 e_{24} 不符合顶点共享规则.通过一阶近邻可以递归的定义 e_{ij} 的 N 阶近邻为:

$$E_i^{(n)} = \bigcup_{e \in E_i^{(n-1)}} E_j^{(1)} \quad (2)$$

依据式(2)中定义,则图 1 中边 e_{34} 的二阶近邻为:

$$\begin{aligned} E_3^{(2)} &= E_1^{(1)} \cup E_2^{(1)} \cup E_4^{(1)} = \{e_{21}, e_{34}\} \cup \{e_{54}, e_{21}\} \cup \{e_{24}\} \\ &= \{e_{21}, e_{24}, e_{34}, e_{54}\} \end{aligned}$$

通过 N 阶近邻路网模型的定义,我们可以对任意目的地路段计算其 N 阶近邻.基于该路网模型我们可以定义每条道路的交通状态集合为 $Z, Z = \{z_i^n\}$ 用于表示每条边的交通状态(即道路拥堵情况), n 表示状态的个数.通过模型可以进一步得到状态转移过程以及与状态对应的时间序列 O 即每条边的实际路网行程时间,用 $O = \{o_i^1, o_i^2, \dots, o_i^n\}$ 表示,状态之间的转移是由状态转移概率模型来表示.

3.2 N 阶状态依赖模型

隐马尔科夫模型可用 $HMM = \{S, A, B, \pi\}$ 表示. S 表示状态序列集, $S = \{s_i^1, s_i^2, \dots, s_i^n\}$ 用于表示每条边在不同时刻 t 的交通状态序列, $s_i^t = \{z_i^n\}$; A 表示状态转移概率矩阵, $A = \{a^{xy}\}, 1 \leq x, y \leq n, a^{xy} = P(s_i^t = z_i^y | s_i^{t-1} = z_i^x)$ 表示从状态 z_i^x 转移到状态 z_i^y 的概率; B 表示输出概率矩阵; $B = \{b^x(o)\}, 1 \leq x \leq n, b^x(o) = P(o \in O_i^1 | s_i^1 = z_i^x)$ 表示在状态 z_i^x 下输出该观测值 o 的概率; π 表示初始概率矩阵; $\pi = \{\pi^x\}, 1 \leq x \leq n, \pi^x = P(s_i^1 = z_i^x)$ 表示时序的初始状态为 z_i^x 的概率.

图 1 中边 e_{34} 的一阶近邻为 $E_3^{(1)} = \{e_{13}, e_{45}, e_{42}\}$ 因此说明 t 时刻中边 e_{34} 的道路状态 S_4 不仅受前时刻 S_4^{t-1} 的影响,同时也受到 e_{34} 的一阶近邻的所有边的影响,依据此情况对图 1 中各边建立状态依赖关系可得到如图 2

所示的一阶状态转移模型. t 时刻中边 e_{34} 的道路状态 S_4^t 依赖于前一时刻的状态 $\{S_1^{t-1}, S_2^{t-1}, S_3^{t-1}, S_4^{t-1}\}$. 因此可以得到基于 N 阶近邻的隐马尔科夫模型 $n\text{HMM} = \{S, C, D, \pi\}$. S 表示状态序列集, $S = \{s_i^1, s_i^2, \dots, s_i^t\}$ 用于表示每条边在不同时刻 t 的交通状态序列, $s_i^t = \{z_i^t\}$; C 表示状态转移概率矩阵, $C = \{c_{i,j}^{xy}\}, 1 \leq x, y \leq n, 1 \leq i \leq |E|, j \in E_i^{(n)}, c_{i,j}^{xy} = P(s_i^t = z_i^t | s_i^{t-1} = z_i^{t-1}, \dots, s_j^{t-1} = z_j^{t-1})$ 表示从状态 z_i^{t-1} 转移到状态 z_j^t 的概率; D 表示输出概率矩阵, $D = \{d_i^x(o)\}, 1 \leq i \leq |E|, 1 \leq x \leq n, d_i^x(o) = P(o \in O_i^t | s_i^t = z_i^t)$ 表示在状态 z_i^t 下输出该观测值 o 的概率; π 表示初始概率矩阵, $\pi = \{\pi_i^x\}, 1 \leq i \leq |E|, 1 \leq x \leq n, \pi_i^x = P(s_i^1 = z_i^x)$ 表示时序的初始状态为 z_i^x 的概率.

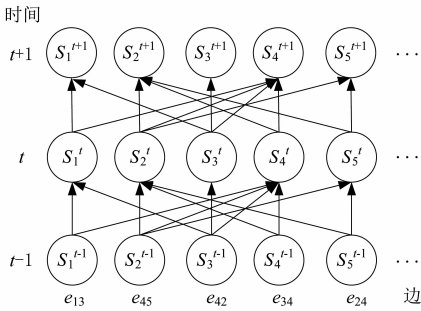


图2 一阶状态依赖模型

车辆行程时间的估计依赖于隐马尔科夫模型中产生时间序列 $O = \{o_i^1, o_i^2, \dots, o_i^t\}$ 的概率为最大, 在已知 N 阶近邻隐马尔科夫模型 $h = \{S, C, D, \pi\}$, 时间序列 O 和状态集 S 情况下其概率公式如下:

$$P(O, S | h) = \pi_i^x d_i^x(o_i^1) \prod_{t=2}^T c_{i,j}^{xy} d_i^x(o_i^t) \quad (3)$$

针对本文中的情况, 如果要通过基于 N 阶近邻的隐马尔科夫模型实时预测车辆行程时间, 则必须解决以下两个问题:

- (1) 已知时间序列 $O = \{o_i^1, o_i^2, \dots, o_i^t\}$ 如何得到状态序列 $S = \{s_i^1, s_i^2, \dots, s_i^t\}$ 及其状态集 $Z = \{z_i^n\}$.
- (2) 如何利用实时监测数据更新已有的隐马尔科夫模型.

3.3 基于 DPMM 的聚类

城市道路的车辆的行程时间是一个具有波动性的分布, 以往的方法仅仅用拥堵和非拥堵状态区分道路交通状态, 同时已有的隐马尔科夫模型的状态是通过采样得到的固定值, 即 $Z = \{z_i^n\}$ 中 n 为人为设定值, 该方法不适用于城市道路交通复杂多变的情况, 本文利用聚类方法自适应的从时间序列中得到状态个数 n , 并给出了寻找最佳状态序列 S 的算法.

3.3.1 时序数据处理

由于每条边 e_{ij} 上采集到的行程时间序列数据 $O = \{o_i^1, o_i^2, \dots, o_i^t\}$ 并不是在时间 t 上的均匀分布, 即某些

时段的数据分布稀疏性较大甚至有数据缺失的现象发生, 本文对采集到的时间序列数据进行了压缩和归一化的处理. 假设采样时间为 $M = 30$ 天, 时间间隔为 $\beta = 5$ 分钟, 则采样点个数 $t = \left\lceil \frac{M \cdot 24 \cdot 60}{\beta} \right\rceil$ 将包含 8640 个点. 压缩公式如式(4):

$$\bar{O} = \{\bar{o}_i^1, \bar{o}_i^2, \dots, \bar{o}_i^f\} \quad (4)$$

则压缩后的时间序列 \bar{O} 在一个时间周期 $P = 24$ 小时内其包含的采样点数为 $f = \left\lceil \frac{P}{\beta} \right\rceil = 288$ 个.

3.3.2 DPMM 自适应聚类

高斯混合模型(GMM)^[19]常常用来处理复杂的概率密度分布, 对于时间序列 $\bar{O} = \{\bar{o}_i^1, \bar{o}_i^2, \dots, \bar{o}_i^f\}$ 的概率密度分布函数, 假设其服从高斯混合分布, 且包含 n_i 个聚类每个聚类都是一个单独的高斯分布, 则其定义如下:

$$\text{GMM}_i(o) = \sum_{k=1}^{n_i} m_{i,k} \cdot N(o | \mu_{i,k}, \delta_{i,k}^2) \quad (5)$$

其中 $m_{i,k}$ 表示第 k 个高斯分布的混合系数, 且满足 $\sum_{k=1}^{n_i} m_{i,k} = 1$, 第 k 个高斯分布 $N(o | \mu_{i,k}, \delta_{i,k}^2)$ 其均值为 $\mu_{i,k}$ 方差为 $\delta_{i,k}^2$.

然而由于聚类个数 n_i 为事先指定, 如果 n_i 过小则会造成观测序列中的特征被覆盖, 如果 n_i 过大则会导致过拟合. 因此需要自适应的寻找最优聚类个数. 同时对于本文中观测序列实时变化的情况, $Z = \{z_i^n\}$ 并不是一个有穷状态而是一个潜在状态无穷的集合. 本文引入无限状态分布狄利克雷混合分布(DPMM)^[20]来表示观测序列的概率密度分布. 则其混合分布函数如下:

$$m_{i,k} | \alpha \sim \text{GEM}(\alpha), z_i^n | m_{i,k} \sim m_{i,k} \\ s_i^t | s_i^{t-1} \sim \text{Multinomial}(m_{i,k}), \bar{o}_i^f | s_i^t \sim \text{DP}(\bar{o}_i^f) \quad (6)$$

其中 α 为用户定义的超参数, $\sum_{k=1}^{n_i} m_{i,k} = 1$, 观测数据服从分布 $\text{DP}(\bar{o}_i^f)$. 为了找到最佳的聚类个数 n_i , 我们给出自适应的聚类算法 1:

算法 1 自适应聚类算法

输入: 时间序列线段 $\bar{O} = \{\bar{o}_i^1, \bar{o}_i^2, \dots, \bar{o}_i^f\}$

输出: DP 混合分布 $\text{DP}(\bar{o}_i^f)$

步骤:

DPMM preDP = null, newDP = null;

preLH = $-\infty$, newLH = $-\infty$;

int $n = 0$;

$Q = \bigcup_{x=1}^f \bar{o}_i^x$

将 Q 等分为 α 个子集 $Q[1], \dots, Q[\alpha]$

while(newLH \leq preLH)

```

preDP = newDP;
preLH = newLH; newLH = 0;
n = n + 1;
for x = 1...α
train = Q/Q[1]; test = Q[1];
newDP = EstimateDP(train, n);
newLH = newLH + EvalLH(test, newDP);
end
end
DP( $\bar{o}_i^t$ ) = preDP;
return DP( $\bar{o}_i^t$ );

```

算法首先将原压缩序列等分为 α 份, 然后计算新的 DP 分布 newDP 和它的似然估计 newLH. 在 10 ~ 14 行中将 $\alpha - 1$ 份序列作为训练集剩下一份为测试集, 通过 EstimateDP(train, n) 的 EM 算法计算得到 newDP, 同时通过 newDP 和测试集的似然估计得到新的 newLH. 算法迭代直到找到极大似然估计得到最终分布也就同时得到了最佳聚类个数 n_i , 算法的时间复杂度为 $O(\alpha n)$.

3.4 模型更新算法

通过上述方法我们建立了一个基于 N 阶路网模型的隐马尔科夫模型, 并得到了其状态状态转移矩阵. 然而, 该模型是基于历史数据训练的, 在 OD 路段行驶过程中对于实时采集到的数据没有处理能力, 因此不能随着车辆行驶而动态的预测车辆与目的地之间的行程时间, 为解决这一问题本文提出模型更新算法.

步骤 1 给定一个预测目标 OD 路段, 对其建立 N 阶路网模型 $G = (V, E, T)$, 其中 $E = (e_1, \dots, e_n)$ 通过式 2 得到其 n 阶近邻, 进而得到图 2 中状态依赖过程 $s_i^t = z_i^t | s_1^{t-1} = z_1^t, \dots, s_j^{t-1} = z_j^t$, 利用算法 1 可以得到状态集 $Z = \{z_i^n\}$.

步骤 2 当车辆行驶在边 e_1 上时利用模型预测得出各边在对应时刻的状态 s_i^t 下的行程时间序列 $Y = (Y_{e_2}, \dots, Y_{e_n})$, 通过实时采集技术得到各段实时行程时间序列 $T = (T_{e_2}, \dots, T_{e_n})$.

步骤 3 将各边的时间序列 $T = (T_{e_2}, \dots, T_{e_n})$ 加入式 6 计算聚类, 若有新分布 newDP 且产生新分布的边为下一路段 e_2 , 则该段预测时间以 $Y_{e_2} = T_{e_2}$ 表示, 否则若没有新分布产生, 更新分布 newDP 段重新估计该边行程时间 Y_{e_n} .

步骤 4 当车辆驶入下一路段后, 则更新 $E = (e_2, \dots, e_n)$ 通过 2 式重新估计路网模型 $G = (V, E, T)$, 更新状态 $S = \{s_1^1, s_1^2, \dots, s_i^t\}$, 重复步骤 2、3.

当迭代更新完 $T = (T_{e_2}, \dots, T_{e_n})$ 后 OD 路段间的车辆行程时间 = $\sum_{k=1}^n Y_{e_k}$. 在步骤 3 中是否更新模型是依

据实时数据加入后是否能产生新分布决定, 而当产生新的分布时多数情况下表明路段上的行程时间耗时与平时有异, 也就是发生拥堵情况, 因此如果产生新分布的 T_{e_n} 来自下一路段, 可直接用来更新预测时间 Y_{e_n} .

4 实验

仿真实验基于城市公共数据集, 包含了从 2007 年 4 月至 2008 年 5 月 uptown 的全区域交通 GPS 数据. 本文将选取图 1 中部分街道作为训练模型的历史数据, 另外交通实时采集数据将通过一个综合交通仿真平台 Vis-sim5.2 结合真实环境数据模拟产生, 如图 3 所示.

首先利用平台进行交通仿真, 通过现场数据对模型校准直至仿真时间与真实时间误差小于 1%. 其次利用 Network Simulator 网络仿真软件进行通信仿真, 对道路中每个车辆入路网时分配一个 ID. 通过这个唯一的 ID 来识别该车辆, 以及其在路网中的运动轨迹. 利用路侧设备统计每个路段的交通流并统计实时行程时间.

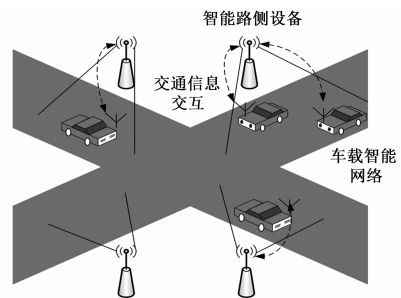


图3 路侧仿真模型

选定车辆 A , 路段 OD 总路段数为 m , 当 A 进入第一路段即 OD1 时, 统计其它 $m - 1$ 个路段的实时行程时间, 并将其加入聚类算法计算, 复杂度为 $O(\alpha n)$, 如果有新聚类产生则以实时时间替代预测时间, 否则实时数据加入历史数据中计算, 并更新预测时间, 其时间复杂度也为 $O(\alpha n)$. 因此, 车辆 A 在路段上的总时间复杂度为 $O(\alpha nm)$.

在此基础上我们进行了六组实验, 首先分析了训练数据与参数的选择对模型的影响, 然后测试了在不同阶数下 N -HMM 模型预测的准确性和性能.

系统的硬件平台采用惠普服务器, 具体配置为 8 个 3.2GHz Intel Core i7-2600 CPU, 16GB 内存, 2TB SCSI 磁盘组包含 4 个 7200rpm 磁盘. 平台内部编程使用 VB 语言完成.

4.1 参数分析

训练数据集和模型中相关参数的选取会对模型的预测产生一定的影响, 因此本节设置两组实验分别进行考察. 首先定义准确性的衡量标准平均平方差总和 (ASSL), 对于边 e_n 有:

$$ASLL(e_i) = \frac{1}{m_i} \cdot \sum_{j=1}^{m_i} (Y_j - T_j)^2 \frac{ASLL_{N-HMM}}{ASLL_{N-HMM}}$$

其中 m_i 表示测试数据集中边的总数, Y_j 表示预测时间, T_j 表示实际行程时间. 由此可以定义总的 ASLL:

$$ASLL = \frac{1}{|E|} \cdot \sum_{e_i \in E} ASLL(e_i)$$

模式隐马尔科夫预测方法在时间预测上具有较高的精确性, 本文以模式隐马尔科夫预测方法为基准方

法定义 ASLL Loss Ratio 为 $\frac{ASLL_{N-HMM}}{ASLL_{基准}}$. 在图 4(a) 中我们采用了三种不同的数据集作为训练数据对 9 月期间的行程时间进行预测, data1 是全年的数据集, data2 为全年前两个月数据集, data3 为八、九月份的数据集. 测试结果表明选取离预测时间点较近的少量数据集预测效果更好. 在图 4(b) 中我们选取不同的时间间隔 $\beta = 5$ 分钟、 $\beta = 15$ 分钟和 $\beta = 30$ 分钟与基准方法的预测效率进行比较, 实验证明时间间隔越小预测效果越好, N-HMM 方法的 ASLL 仅为基准方法的 45%.

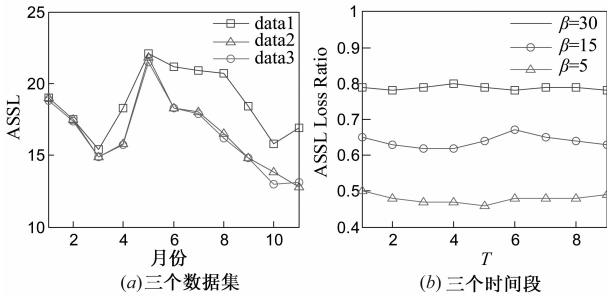


图4 参数分析

4.2 时间预测

选定了数据集和时间间隔后, 本节对 N-HMM 模型的阶数选择进行了实验, 图 5(a) 中本文选取了不同阶的模型计算其近邻边的基数, 当阶数高于 6 之后, 假设模型中单个边的状态数为 4, 则模型可能的邻近边将达到 $4 \cdot 4^{12}$, 计算量较大, 因此尽量选择小于 5 的 N-HMM 进行预测. 在图 5(b) 中定义 couple ratio 为 $\frac{ASLL_{N-HMM}}{ASLL_{N-HMM}}$, $ASLL_{N-HMM}$ 表示不使用 N 阶近邻的原始隐马尔科夫模型, 实验证明 N 阶近邻模型对预测效率的提升比较明显, 2 阶近邻的 ASLL 仅为原始模型的 72%.

本文对动态更新算法的收敛性也进行了实验, 设定一段总数为 $x = 20$ 的 OD 路段进行预测, 选取 $\alpha = 10$, $\beta = 5$ 分别对 2 阶和 3 阶近邻下的动态更新模型按路段迭代计算其 ASLL 值, 图 6 实验表明在不同阶数下动态更新算法仍然能够保持收敛.

本文对模型更新情况下的预测也进行了实验, 在图 7(a) 中对模型更新时间和模型训练时间进行了比较, 在阶数小于 5 的情况下模型更新时间和训练时间都

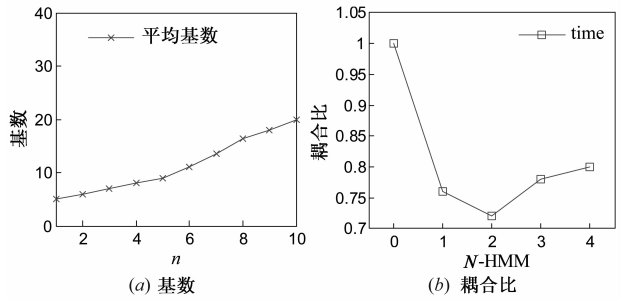


图5 阶数选择

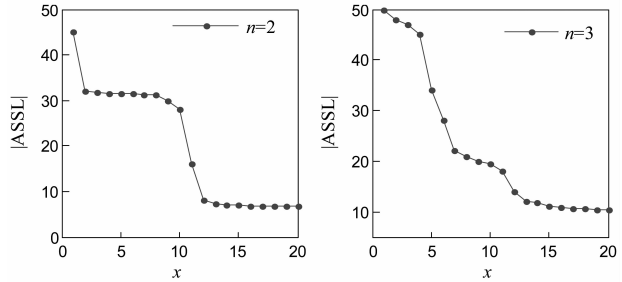


图6 收敛性分析

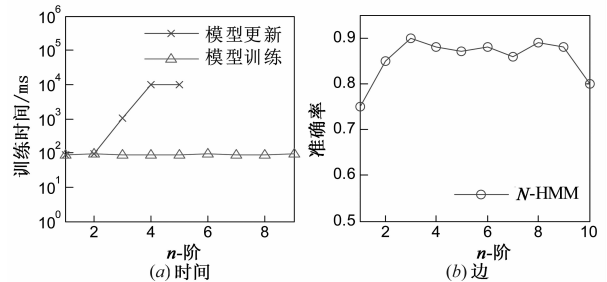


图7 模型更新

在可接受范围内, 能够在车辆驶入下一路段前完成模型更新和训练. 图 7(b) 中选定了一段总数为 20 的 OD 路段进行预测, 通过比较实际预测时间 T_j 和模型预测时间 Y_j 计算平均的准确率, x 轴表示经过的路段数据递增的情况下, 平均预测效果. 模型预测效率随着路段数量的增加基本趋于稳定, 因此模型更新后依然有较高的预测能力. 但是由于迭代算法的局部最优解问题, 在路段数量较大后, 模型的预测性能略有影响, 因此选定模型的参数对能否取得全局最优解比较重要.

5 结论

N 阶近邻的隐马尔科夫模型能够对车辆行程时间做出有效预测, 当车辆在行进过程中实时采集数据传送过来后, 模型能够自适应的计算路网模型中状态转移过程, 通过模型更新算法使得模型及时的给出路段行程时间. 实验证明模型具有较高的时效性和精确性.

参考文献

- [1] Yunfei Fang. A new model for lane reservation problem with time - dependent travel times [A]. 10th IEEE International Conference on IEEE[C]. Washington; IEEE, 2013. 367 - 372.
- [2] Tongyu Zhu. Travel time prediction for float car system based on time series [A]. International Conference on Advanced Communication Technology[C]. Washington; IEEE, 2010. 1503 - 1508.
- [3] Mengdan Gao. Analysis of travel time patterns in urban using taxi gps data [A]. IEEE International Conference on Green Computing Communications[C]. Washington; IEEE, 2013. 512 - 517.
- [4] Till Wohlfarth. A data-mining approach to travel price forecasting [A]. International Conference on Machine Learning and Applications Workshops[C]. Washington; IEEE, 2011. 84 - 89.
- [5] Jiankai Wang. Vehicle reidentification with self-adaptive time windows for real-time travel time estimation [J]. IEEE Trans on ITS, 2013, 15(2): 540 - 562.
- [6] Ding H, Trajcevski G, Scheuermann P, et al. Querying and mining of time series data: experimental comparison of representations and distance measures [A]. International Conference on Very Large Data Bases[C]. San Fransisco: Morgan Kaufmann, 2008. 1542 - 1552.
- [7] Shangyao Yan. A car pooling model and solution method with stochastic vehicle travel times [J]. IEEE Trans on ITS, 2013, 15(1): 47 - 56.
- [8] 曹天扬, 申莉. 一种应用路阻函数的路段行驶时间预测方法 [A]. 第八届中国智能交通年会论文 [C]. 北京: 电子工业出版社, 2014. 533 - 539.
- [9] Jungme Park, Yi Lu. Intelligent trip modeling for the prediction of an origin-destination traveling speed profile [J]. IEEE Trans on ITS, 2014, 15(3): 1039 - 1046.
- [10] J Park, D Li, YMurphey, J Kristinsson, et al. Real time vehicle speed prediction using a neural network traffic model [A]. IEEE International Joint Conference on Neural Network[C]. Washington; IEEE, 2011. 2991 - 2996.
- [11] Teh Y, Jordan M. Hierarchical Bayesian Nonparametric Models with Applications. Bayesian Nonparametrics Principles and Practice [M]. London: Cambridge University Press, 2009. 1 - 47.
- [12] Wang P, Wang H, et al. Finding semantics in time series [A]. International Conference on Special Interest Group on Management of Data[C]. New York; ACM, 2011. 385 - 396.
- [13] Won P, Melek M, Golnaraghi F. A kalman particle filter based position and orientation estimation method using a position sensor inertial measurement unit hybrid system industrial electronics [J]. IEEE Trans on Industrial Electronics, 2010, 57(5): 1787 - 1798.
- [14] Chen Y, Hu B, Keogh E. DTW-D: time series semi-supervised learning from a single example [A]. SIGKDD[C]. New York; ACM, 2013. 383 - 391.
- [15] Paisley J, Wang C, et al. Nested hierarchical dirichlet processes [J]. ArXiv Preprint ArXiv, 2012, 12(10): 136 - 158.
- [16] Xu Z, MacEachern S, Xu X. Modeling non-gaussian time series with nonparametric bayesian models [J]. IEEE Trans on PAMI, 2015, 37(2): 372 - 382.
- [17] Van Lint. Reliable real-time framework for short-term freeway travel time prediction [J]. Journal of transportation engineering, 2006, 132(12): 921 - 932.
- [18] Rice J, Van Zwet. A simple and effective method for predicting travel times on freeways [J]. IEEE Trans on ITS, 2004, 5(3): 200 - 207.
- [19] Zhang X, Rice J A. Short-term travel time prediction [J]. Transportation Research Part C: Emerging Technologies, 2003, 11(3): 187 - 210.
- [20] Fei X, Eisenman S, Mahmassani H, et al. Application of dynamart-x to the maryland chart network for real-time traffic management center decision support [A]. 12th World Congress Intelligent Transportation System [C]. Washington; IEEE, 2005. 341 - 347.

作者简介



章登义 男, 1965 年 5 月出生于湖北省荆州市. 现为武汉大学计算机学院教授、博士生导师.



欧阳黼霏 (通讯作者) 男, 1988 年出生于湖北荆州. 现为武汉大学博士研究生, 主要研究数据库、数据挖掘.

E-mail: whuxiaouou@whu.edu.cn