

SCATS 线圈数据短时多步双重预测方法

李琦¹, 姜桂艳^{1,2}

(1. 吉林大学 交通学院, 130022 长春; 2. 吉林大学 汽车仿真与控制国家重点实验室, 130022 长春)

摘要: 为了进一步改善悉尼自适应交通控制系统(Sydney coordinated adaptive traffic system, SCATS)线圈数据短时多步预测的效果,在对 SCATS 线圈数据进行预处理的基础上,将当前与之前若干时间间隔的交通数据及对应的时间点作为交通模式特征向量的构成要素,用欧式距离作为当前交通模式特征向量和历史交通模式特征向量相似性的测度指标,以多步预测结果的误差最小为目标选取近邻数,通过对交通模式之间距离的倒数正规化处理,确定了所选相似交通模式的未来交通参数的权重,设计了一种基于 k 近邻(k nearest neighbor, k -NN) 算法的短时多步双重预测方法,包括 SCATS 线圈数据的多步预测方法以及可预测步数在线估计方法,并采用某特大城市 SCATS 线圈实测数据进行了验证和对比分析. 结果表明,所提出的新方法能够进一步降低 SCATS 线圈数据短时多步预测的误差.

关键词: 交通运输工程; 悉尼自适应交通控制系统; 感应线圈; 短时交通预测; k 近邻算法

中图分类号: U491

文献标志码: A

文章编号: 0367-6234(2013)02-0123-06

Bi-level method of multi-step forecasting for short-term data of loop in SCATS

LI Qi¹, JIANG Guiyan^{1,2}

(1. College of Transportation, Jilin University, 130022 Changchun, China;

2. State Key Laboratory of Automotive Simulation and Control, Jilin University, 130022 Changchun, China)

Abstract: For the improvement of the effect of traffic multi-step forecasts using short-term data of loop in Sydney coordinated adaptive traffic system(SCATS), on the basis of data preprocessing, traffic data and time points at time t over a sampling period of n intervals were included in the traffic state feature vector, Euclidean distance was used to measure the closeness between current traffic state and historical traffic state, the number of nearest neighbors corresponding to the minimum error of travel multi-step forecasts was selected, and the weights of k -nearest neighbors were identified by normalizing the reciprocal of the distance between traffic states, a new bi-level method of multi-step forecasting using k Nearest Neighbor (k -NN) algorithm was designed, including a multi-step forecasting method and a predictable steps on-line estimation method. The validity of the proposed method was tested with data measured from a megacity. The results indicate that the proposed method can further improve the effect of short-term traffic multi-step forecasts.

Key words: traffic and transportation engineering; Sydney coordinated adaptive traffic system(SCATS); inductive loop; short-term traffic forecasts; k nearest neighbor (k -NN) algorithm

悉尼自适应交通控制系统(Sydney coordina-

ted adaptive traffic system, SCATS)在世界范围内得到了广泛的应用. 1986年上海在国内最早引入 SCATS 系统,随后天津、广州、沈阳、杭州、苏州、宜昌等城市也开始采用该系统控制城市交通^[1-2]. 目前,SCATS 线圈获取的动态交通数据(平均车头时距和交通流量)仅用于交通信号控制. 从数据共享的角度看,如果能够将其所获得的动态交通数据应用于该系统所覆盖道路的交通拥

收稿日期: 2012-06-04.

基金项目: 国家自然科学基金资助项目(51278257);

高等学校博士学科点专项科研基金资助项目(20110061110084);

浙江省自然科学基金资助项目(LY12F01013).

作者简介: 李琦(1985—),男,博士研究生;

姜桂艳(1964—),女,教授,博士生导师.

通信作者: 姜桂艳,jianggy@jlu.edu.cn.

挤预测,对于改善交通控制、交通诱导和交通指挥的协调性具有重要的现实意义。

短时交通预测是指采样和分析时间间隔不大于 15 min 的交通预测,是 ITS 研究与应用的关键技术之一^[3]。然而,目前国内外对于短时交通预测的研究大部分局限于一步预测^[4-8],这种预测结果实际上只能修正信息采集与处理的时间延迟,是对当前时段交通状态的估计,无法支持交通管理者和交通出行者进行预见性决策。虽然有少数文献涉及了多步预测问题^[9-10],但文献[9]采用的是固定预测步数,不能针对某个时间段内交通数据序列的特性决定其可预测步数;文献[10]虽然可实现对交通数据序列的可预测程度的在线分析,但由于在 SCATS 中各个采样间隔内的交通数据不具有严格的时间可比性^[11],该方法不能直接用于 SCATS 线圈数据的多步预测,而且该方法在可预测步数的确定以及多步预测效果方面还存在一定的提升空间。

k 近邻(k nearest neighbor, k -NN) 算法是一种以最相样本和模式识别为基础的非参数回归方法,在分类和估计等领域已得到了广泛的应用^[12-14]。在特定区域内,交通数据序列是由区域内的社会经济、道路网络、车辆保有量、驾驶员特性、交通管制措施以及交通环境等因素共同作用的结果,同时具有随机性和相似性的特点,在一定程度上能够满足 k -NN 算法应用的需要,本文将以此为基础设计 SCATS 线圈数据的短时动态多步预测方法,在引用文献[11]提出的以虚拟时间序列构建方法对 SCATS 线圈数据进行预处理的基础上,以 k -NN 算法为基础,设计了 SCATS 线圈数据的多步预测方法以及可预测步数在线估计方法,并采用某特大城市 SCATS 线圈实测数据进行了验证和对比分析。

1 基于 k -NN 的 SCATS 线圈数据多步预测方法

1.1 多步预测基本方法设计

当前与之前若干时间间隔的交通数据构成的一段数据序列具有特定的数值大小、升降趋势等属性,可用于表征当前时间间隔的交通模式。此外,考虑到交通数据序列是随着时间的推移而不断演化的,可视作时间要素的函数,因此本文将交通模式的特征向量定义为

$$Y_t = [T_t, T_{t-1}, \dots, T_{t-d+1}, y_t, y_{t-1}, \dots, y_{t-d+1}] \quad (1)$$

其中: t 为当前时间间隔; Y_t 为第 t 个时间间隔的交通模式特征向量; d 为数据分析时间窗口宽度; T_t, T_{t-1}, T_{t-d+1} 分别为第 $t, (t-1), (t-d+1)$ 个时间间隔的时间点; y_t, y_{t-1}, y_{t-d+1} 分别为第 $t, (t-1), (t-d+1)$ 个时间间隔的交通数据。

k -NN 算法通常采用空间距离作为衡量不同模式匹配程度的手段。已有文献表明,不同的距离度量方法对目标估计结果的影响并不明显^[15]。因此,本文采用最为常用的欧式距离作为当前交通模式特征向量和历史交通模式特征向量相似性的测度指标。

近邻数 k 表示从历史数据库中选取的相似交通模式的个数,不同的 k 值会产生不同的估计结果。 k 值的选取在很大程度上与历史数据的具体情况相关,目前还没有统一的选取准则。已有的文献基本都是针对特定的数据环境,根据经验选择 k 值。本文将针对已有的实际数据集合,在其他条件不变的情况下,以多步预测结果的误差最小为目标选取 k 值。

设 H 为最大预测步数,以当前交通模式匹配出的 k 个相似模式为基础,利用相似模式所对应的后续 H 个时间间隔的交通数据,可实现对当前交通模式未来 H 步交通数据的直接估计,如图 1 所示。

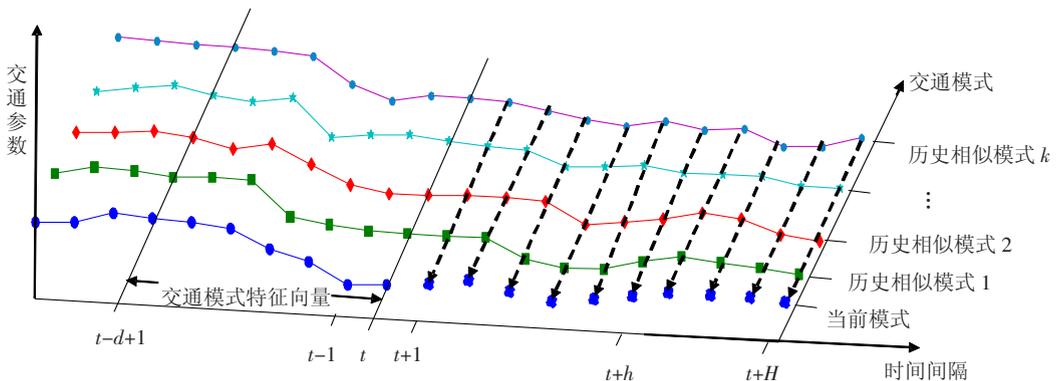


图 1 基于相似模式的未来 H 步交通数据估计示意图

从理论上讲,所筛选出的 k 个相似模式与当前交通模式的相似程度越高,其后续 H 个时间

间隔的交通数据对当前交通模式未来 H 步交通数据估计结果的影响越大。因此,可通过对交通模式间

距离的倒数进行正规化处理,确定 k 个相似交通模式后续 H 个时间间隔交通数据的加权值,计算公式为

$$\hat{y}_{t+h} = \frac{1}{k} \sum_{q=1}^k w_q \hat{B}_q, 1 \leq h \leq H; \quad (2)$$

$$w_q = D_q^{-1} / \sum_{i=1}^k D_q^{-1}. \quad (3)$$

其中: k 为搜索出的相似交通模式个数; \hat{y}_{t+h} 为当前交通模式所对应的未来第 h 步交通数据的估计值; \hat{B}_q 为第 q 个相似交通模式所对应的后续第 h 个时间间隔的交通数据; w_q 为第 q 个相似交通模式所对应后续 H 个时间间隔交通数据的权重; D_q 为第 q 个相似交通模式与当前交通模式之间的欧式距离.

1.2 多步预测改进方法设计

基于 k -NN 的 SCATS 线圈数据多步预测基本方法需要标定的参数包括数据分析时间窗口宽度 d 和近邻数 k ,不同的 (d, k) 会产生不同的预测结果.因此,当前模式未来第 h 步交通数据的预测误差 δ_h 可视为 (d, k) 的函数,即

$$\delta_h = f_h(d, k). \quad (4)$$

H 步交通数据预测的误差之和 δ 可表示为

$$\delta = \sum_{h=1}^H \delta_h = \sum_{h=1}^H f_h(d, k). \quad (5)$$

(d, k) 标定的目标是使得 δ 达到最小,即

$$\delta_{\min} = \min \sum_{h=1}^H \delta_h = \min \sum_{h=1}^H f_h(d, k). \quad (6)$$

如果每一步预测误差 δ_h 都为最小值,则误差之和 δ 必定为最小值.但在基于 k -NN 的 SCATS 线圈数据多步预测基本方法中, H 步交通数据预测采用相同的 (d, k) ,不能确保每一步交通数据的预测误差 δ_h 都能达到最小值.因此, H 步交通数据预测的误差之和 δ 还有进一步改进的空间.为了能够实现误差之和 δ 最小的目标,本文对未来每一步预测分别进行标定 $(d, k)_h$,并使得 δ_h 达到最小,即

$$\min \delta_h = f_h(d, k)_h. \quad (7)$$

在进行第 h 步预测时, $(d, k)_h$ 的具体标定的方法为:1) 初始设置 $d = 1$;2) 近邻数 k 从 1 开始逐渐增加,在确定的 d 下获取误差最小的 k 值;3) $d = d + 1$,重复步骤 2,直至获取对应最小误差的 (d, k) .

选择应用最广的平均相对误差^[16] 作为第 h 步交通数据预测结果的评价指标,其值越小,说明方法的性能越优.考虑到相同日期(星期一,星期二, ..., 星期日)的交通模式具有较强的相似

性^[11],对于每一检测器,在建立多步预测方法时,以相同日期的交通模式为基础.

2 基于 k -NN 的 SCATS 线圈数据可预测步数在线估计方法

2.1 可预测步数在线估计的基本原理

在应用上一节提出的方法时,只要程序不停止运行,就可以对 SCATS 线圈数据进行无限多步的预测.但实证分析表明,在特定时点上进行多步预测时,每一步的预测误差与预测步数之间存在正相关关系,满足期望误差的预测步数不可能无限多.为此文献[10]最先提出了交通数据序列的动态可预测性的概念,并针对城市快速路线圈的 5 min 交通数据,设计了一种可预测步数在线估计方法.其基本方法是:首先,分析实际交通数据序列的动态特性,建立关联数据特征指标向量,并获得每个时点处的关联数据特征指标向量值;其次,在指定期望预测误差的情况下,运用一定的预测方法获得每个时点处可满足预测误差要求的预测步数;再次,运用 BP 神经网络建立关联数据特征指标向量与可预测步数之间的映射关系;最后,在实际运用过程中,根据所获得的动态交通数据,在线计算当前时点的关联数据特征指标向量值,并据此估计此时的可预测步数.本文以上述文献为基础,充分利用交通参数时间序列中所包含的可预测性信息,直接用交通参数的时间数据序列本身(包括时间点序列和交通参数数据序列)构成交通模式的特征向量,并基于 k -NN 算法建立新的估计方法.

2.2 基于 k -NN 的 SCATS 线圈数据可预测步数在线估计方法设计

基于 k -NN 的 SCATS 线圈数据可预测步数在线估计同样包括交通模式特征向量的定义、交通模式相似性测度方法的选择、近邻数 k 的确定以及基于相似模式的预测步数估计等 4 个步骤,具体方法与 1.1 节基本一致,不同之处在于 1.1 节中交通模式对应的目标值为后续 H 个时间间隔的交通数据,而此处为可预测步数.由于可预测步数是离散变量,因此选择平均绝对误差^[16] 作为可预测步数估计结果的评价指标,该值越小说明方法性能越优.

3 基于 k -NN 的 SCATS 线圈数据短时多步双重预测方法工作流程

由于交通参数的多步预测结果是可预测步数估计方法标定的数据基础,因此本文所设计方法

的标定流程是基于 k -NN 的 SCATS 线圈数据多步预测方法在前, 基于 k -NN 的 SCATS 线圈数据可预测步数在线估计方法在后. 在实际应用过程中, 则应先估计目标交通参数在当前时间间隔的可预测步数, 再对可预测步数内的交通参数进行估计. 因此, 本文所设计的基于 k -NN 的 SCATS 线圈数据短时多步双重预测方法的在线应用流程如图 2 所示.

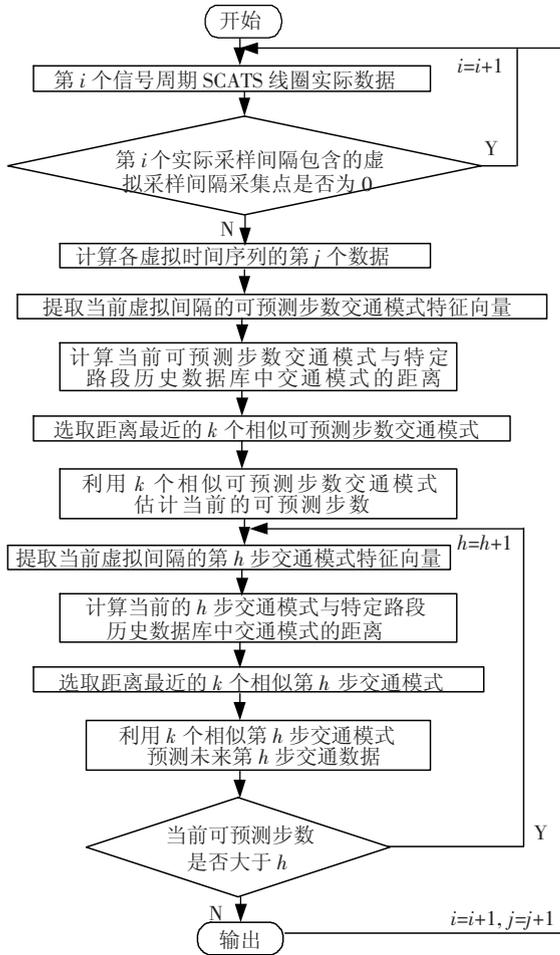


图 2 多步双重预测方法的在线应用流程

4 实证分析

实证数据来源于国内某大城市主干道的 SCATS 线圈, 2009 年 5 月至 7 月每天采样 24 h. 实验区域为该主干道 A—G 连续 7 个交叉口, 双向共 12 条路段, 具体路线如图 3 所示.

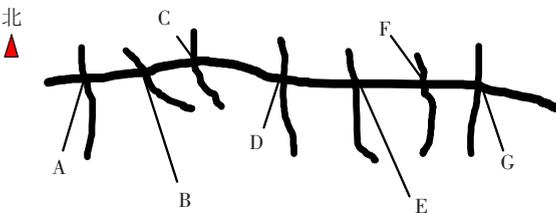


图 3 实验区域道路示意图

利用文献[11]提出的虚拟时间序列构建方法, 确定 SCATS 线圈数据虚拟采样间隔时长为 150 s. 因此, 后续分析以 150 s 时间尺度为例加以说明. 选取 5、6 月份所有星期一的数据作为标定集合, 7 月份所有星期一的数据作为测试集合. 短时多步预测结果的可接受误差阈值 z 应依据用户需求进行设定, 本文参照文献[10]的方法定为 20%. 将文献[9]与文献[10]的相应方法作为对比方法, 分别对固定步数多步预测结果、可预测步数估计结果以及可变量步数多步预测结果进行对比分析, 以评价本文设计方法的有效性.

以交叉口 B 西进口线圈的星期一数据集合为例, 以前述约定的标定数据集合为基础, 经过实验计算, 最终确定的平均车头时距与交通流量第 h 步预测所对应的参数及平均相对误差如表 1 所示.

由表 1 可看出, 随着预测步数的增加, 平均车头时距和交通流量最终确定的数据分析时间窗口宽度均整体上呈递增的趋势, 而近邻数没有明显的趋势性. 将本文所设计的基于 k -NN 的 SCATS 线圈数据多步预测基本方法简称为本文基本方法, 将所设计的基于 k -NN 的 SCATS 线圈数据多步预测改进方法简称为本文改进方法. 以前述约定的测试数据集合为基础, 文献[9]方法、文献[10]方法、本文基本方法以及本文改进方法的平均车头时距和交通流量 1~30 步预测结果的平均相对误差分别如图 4 和图 5 所示.

从图 4 和图 5 可以看出: 1) 平均车头时距和交通流量 1~30 步预测结果的平均相对误差总体上呈递增的趋势, 说明预测误差与预测步数之间存在一定的正相关关系. 2) 在对平均车头时距和交通流量进行 1~30 步预测时, 参与比较的各种方法在预测误差方面存在明显且稳定的差异, 本文基本方法和本文改进方法的平均相对误差均小于文献[9]方法和文献[10]方法, 说明本文所设计的两种多步预测方法具有相对优势. 3) 本文改进方法的平均相对误差小于本文基本方法, 说明本文所设计的改进方法能够进一步改善多步预测的效果.

以前述约定的标定数据集合为基础, 经过实验计算, 最终确定平均车头时距和交通流量可预测步数在线估计所对应的 d 分别为 6 和 12. 以欧式距离为基础, 在选取不同近邻数 k 时, 平均车头时距与交通流量可预测步数估计的平均绝对误差如图 6 所示.

表 1 平均车头时距与交通流量第 h 步预测所对应的参数及平均相对误差

预测步数	平均车头时距			交通流量			预测步数	平均车头时距			交通流量		
	时间窗口	近邻个数	平均相对误差/%	时间窗口	近邻个数	平均相对误差/%		时间窗口	近邻个数	平均相对误差/%	时间窗口	近邻个数	平均相对误差/%
1	7	12	13.01	12	13	14.04	16	14	11	15.90	16	15	18.72
2	7	17	13.12	12	20	14.38	17	11	17	16.24	14	17	19.17
3	8	11	13.15	13	18	15.16	18	11	15	16.61	16	19	19.57
4	9	9	13.18	13	18	15.29	19	14	13	16.92	18	12	20.17
5	8	11	13.27	14	17	15.47	20	11	12	17.08	17	19	20.66
6	9	14	13.34	14	12	15.63	21	12	11	17.21	17	13	20.91
7	7	15	13.59	14	15	15.76	22	13	16	17.67	18	15	21.24
8	9	11	14.01	15	14	15.78	23	13	19	17.98	18	11	21.67
9	10	16	14.31	14	11	15.82	24	12	12	18.16	18	9	22.07
10	10	12	14.53	15	17	16.11	25	13	16	18.24	18	14	22.35
11	10	14	14.79	15	19	16.75	26	14	14	18.36	19	18	22.60
12	11	15	15.18	15	15	17.22	27	14	12	18.41	19	11	22.89
13	12	19	15.45	17	16	17.72	28	14	15	18.69	19	17	23.31
14	11	17	15.47	15	14	18.07	29	13	13	18.75	20	15	23.78
15	13	16	15.55	16	10	18.26	30	15	12	18.92	20	11	24.21

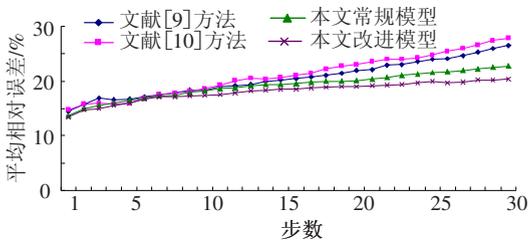


图 4 平均车头时距 1 ~ 30 步预测结果的平均相对误差

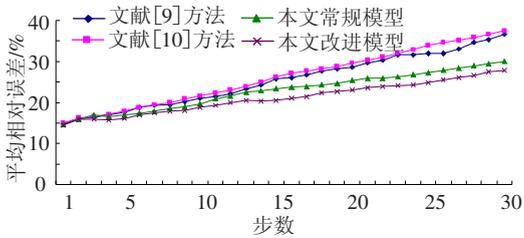


图 5 交通流量 1 ~ 30 步预测结果的平均相对误差

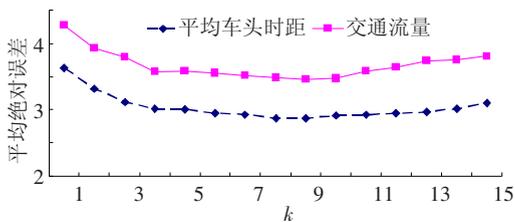


图 6 平均车头时距与交通流量可预测步数估计的平均绝对误差

从图 6 中可以看出,当 k 分别取 8 和 9 时,平均车头时距和交通流量可预测步数估计的平均绝对误差最小. 因此,在本文中,将平均车头时距和流量可预测步数估计的近邻数 k 分别选定为 8 和 9. 以前述约定的测试数据集为基础,采用本文方

法和文献[10]方法,对实验区域 SCATS 线圈数据可预测步数进行估计,平均车头时距可预测步数的平均绝对误差分别为 2.83、3.61,交通流量可预测步数的平均绝对误差分别为 3.21、4.77,本文方法明显优于文献[10]方法. 为了对可变步数多步预测的效果进行评价,以测试数据集为基础,计算全部时间间隔对应的可预测步数内各步预测结果的总体平均相对误差. 由于文献[9]没有设计可预测步数的估计方法,在进行评价时,分别引入文献[10]和本文设计的可预测步数估计方法,并分别简称为文献[9]与文献[10]组合方法以及文献[9]与本文组合方法. 相应方法的固定 30 步预测效果与可变步数多步预测效果如表 2 所示.

由表 2 可看出:1) 平均车头时距和交通流量可变步数多步预测的总体平均相对误差都优于对应方法的固定步数 30 步预测方法,说明进行可预测步数估计是非常必要的. 2) 本文所设计的基本方法和改进方法的平均车头时距和交通流量多步预测的总体平均相对误差均优于文献[9]与本文的组合方法,进一步表明了本文所设计方法的有效性. 3) 采用文献[9]与本文的组合方法,平均车头时距和交通流量的总体平均相对误差均优于文献[9]与文献[10]的组合方法,进一步说明了本文所设计的可预测步数估计方法的有效性. 4) 在 5 种可变步数多步预测的方法中,本文改进方法

的预测效果优于其他 4 种方法,表明本文所设计的改进方法能够进一步降低 SCATS 线圈数据短时多步预测的误差.

表 2 相应方法的固定 30 步预测效果与可变步数多步预测效果

方法	总体平均相对误差/%	
	平均车头时距	交通流量
文献[9]固定步数 30 步预测效果	19.4	25.3
文献[10]固定步数 30 步预测效果	21.5	27.1
本文基本方法固定步数 30 步预测效果	17.5	20.2
本文改进方法固定步数 30 步预测效果	16.9	19.3
文献[9]与文献[10]组合方法的可变步数多步预测效果	15.4	17.1
文献[9]与本文组合方法的可变步数多步预测效果	14.6	16.1
文献[10]方法的可变步数多步预测效果	16.1	17.6
本文基本方法的可变步数多步预测效果	13.9	14.8
本文改进方法的可变步数多步预测效果	13.5	14.6

5 结 论

1) 在基于 k -NN 的 SCATS 线圈数据多步预测基本方法中,各步交通数据预测采用相同的参数,不能确保每一步交通数据的预测误差都能达到最小,从而提出了 SCATS 线圈数据短时多步双重预测方法.

2) 交通参数可变步数多步预测的总体平均相对误差优于固定步数预测方法,说明进行可预测步数估计是非常必要的.

3) 短时交通参数一步预测难以支持交通管理者和交通出行者制定具有预见性的对策与措施,多步预测是该领域的发展趋势.

4) 采用我国某特大城市的实测数据进行验证和对比分析,结果表明所设计的新方法能够进一步降低 SCATS 线圈数据短时多步预测的误差.

参考文献

- [1] 全永燊. 城市交通控制[M]. 北京:人民交通出版社, 1989.
- [2] 彭信林,王宁鸣,周剑峰. SCATS 数据采集系统的设计与实现[J]. 计算机工程,2008,34(24):256-260.
- [3] 姜桂艳. 道路交通状态判别技术与应用[M]. 北京:人民交通出版社,2004.
- [4] 陆海亭,张宁,黄卫. 短时交通流预测方法研究进展[J]. 交通运输工程与信息学报,2009,7(4):84-91.
- [5] SMITH B L, WILLIAMS B M, OSWALD R K. Comparison of parametric and nonparametric models for traffic flow forecasting [J]. Transportation Research: Part C Emerging Technologies, 2002, 10(4):257-321.
- [6] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results [J]. Journal of Transportation Engineering, 2003, 129(6):664-672.
- [7] 姚智胜,邵春福. 道路交通状态预测研究[J]. 哈尔滨工业大学学报,2009,41(4):247-250.
- [8] HOSSEINI S H, MOSHIRI B, RAHIMI-KIAN A, et al. Short-term traffic flow forecasting by mutual information and artificial neural networks [C]//2012 IEEE International Conference on Industrial Technology. Athens: IEEE, 2012: 1136-1141.
- [9] DOUGHERTY M S, COBBET M R. Short-term inter-urban traffic forecasts using neural networks [J]. International Journal of Forecasting, 1997, 13:21-31.
- [10] 姜桂艳,常安德,牛世峰,等. 基于 BP 神经网络的交通数据序列动态可预测性分析方法[J]. 北京工业大学学报,2011,37(7):1019-1026.
- [11] 姜桂艳,李琦,常安德,等. 基于 SCATS 线圈数据的交通状态在线定量评价与预测方法:中国,200910217820.5 [P]. 2010-06-16.
- [12] MACK Y P. Local properties of k -NN regression estimates [J]. SIAM, Journal Algebraic and Discrete Methods, 1981, 2(3):311-323.
- [13] DENOEU X T. A k -nearest neighbor classification rule based on Dempster-Shafer theory [J]. IEEE Transactions on Systems Man and Cybernetics, 1995, 5:804-813.
- [14] TRIGUERO I, DERRAC J, GARCÍA S, et al. A taxonomy and experimental study on prototype generation for nearest neighbor classification [J]. IEEE Transactions on Systems, Man, and Cybernetics: Part C Applications and Reviews, 2012, 42(1):86-100.
- [15] ZHANG S, JANK W, SHMUELI G. Real-time forecasting of online auctions via functional k -nearest neighbors International [J]. Journal of Forecasting, 2010, 26:666-683.
- [16] 薛薇. SPSS 统计分析方法及应用[M]. 北京:电子工业出版社,2004. (编辑 魏希柱)