

# 基于改进 Sage 滤波器的车辆行程时间预测模型

熊桂喜, 刘铭志

(北京航空航天大学, 北京 100083)

**摘要:**在现代 ITS 环境中, 公交车辆行程时间预测是实现公共交通智能化调度子系统、电子站牌显示子系统及公交信息服务子系统的必要条件。针对 Sage 滤波器自身的优缺点, 提出了一种基于车辆行程时间历史数据流信息的 Sage 滤波器, 并在此基础上建立了 BRT(Bus Rapid Transit)车辆行程时间预测模型。最后针对 2007 年 6 月 7 日北京市南中轴路大容量快速公交(BRT)线的实际数据进行了对比实验, 结果表明, 改进的 Sage 滤波器有效降低了原算法的误差。

**关键词:**车辆行程时间预测; Sage 滤波器; 流聚类

**中图分类号:** TP391.9

**文献标识码:** A

**文章编号:** 1673-629X(2008)09-0162-03

## BRT Vehicles Travel Time Prediction Model Based on Improved Sage Filter

XIONG Gui-xi, LIU Ming-zhi

(Beihang University, Beijing 100083, China)

**Abstract:** In modern ITS environment, vehicles travel time prediction is a necessary condition for the realization of the intelligent public transport scheduling subsystems, station electronic display subsystem and bus information service subsystem. First analyzed Sage filter's advantages and disadvantages, and presented an improved Sage filter based on the historical data samples of vehicle travel time. Then, on the basis of such algorithm, it built BRT vehicle travel time prediction model. Finally used actual data collected from BRT Transport of South Axis Street in Beijing on June 7, 2007 for experiment. The result shows that the improved Sage filter effectively reduce the error of the original algorithm.

**Key words:** vehicle travel time prediction; Sage filter; data stream clustering

## 0 引言

公交车辆行程时间预测在完善公交服务方面起着重要作用。研究公交车辆行程时间预测方法, 提高公交车辆行程时间预测精度, 对于提高公交服务质量, 进而提高公共交通吸引力、缓解交通拥挤具有重要意义。

目前, 各国的研究者们已经采用了很多方法用于行程时间预测, 例如时间序列法<sup>[1,2]</sup>、卡尔曼滤波模型<sup>[3]</sup>、马尔可夫链模型<sup>[4]</sup>、非参数回归模型<sup>[5,6]</sup>以及神经网络模型<sup>[7]</sup>等。这些方法和模型各有优缺点, 但当实际数据出现丢失、失真和异常时, 其预测结果往往会

出现不同程度的偏差。

针对这种情况, 提出了一种基于改进 Sage 滤波器的车辆行程时间预测模型。

## 1 改进的 Sage 滤波器

### 1.1 Kalman 滤波器

Kalman 滤波器模型属于线性动态模型, 认为车辆行程时间等于基本行程时间和随机误差的叠加, 基本行程时间由检测器收集的数据计算后得出, 随机误差由测量方程递推计算得出。Kalman 滤波器的系统模型如下<sup>[8]</sup>:

$$T_k = AT_{k-1} + \tilde{\omega}_{k-1} \quad (1)$$

$$Z_k = HT_k + v_k \quad (2)$$

其中: 式(1)为离散随机差分方程; 式(2)为量测方程。 $T_k$ : 状态变量, 即行程时间的预测值;  $A$ : 状态矩阵;  $Z_k$ : 观测变量, 即实际行程时间;  $H$ : 观测矩阵;  $\tilde{\omega}_k$ : 过程噪声或系统噪声, 为满足正态分布的白色噪声  $p(\tilde{\omega}) \sim N(0, Q)$ ;  $v_k$ : 观测噪声, 为满足正态分布的白色噪声  $p(v) \sim N(0, R)$  (这里假设过程噪声  $\tilde{\omega}_k$  和观测噪声  $v_k$  相互独立)。

### 1.2 Sage 滤波器

上述即为经典 Kalman 滤波器的模型, 该模型假设

收稿日期: 2007-12-30

基金项目: “十五”国家科技攻关计划(2005BA414B04)

作者简介: 熊桂喜(1964-), 男, 湖北黄冈人, 副教授, 硕士, 研究方向为企业应用集成与智能交通系统。

系统噪声和观测噪声为零均值且统计特性已知的白色噪声,但是往往实际情况无法满足该假设,此时就会产生建模误差。

针对经典 Kalman 滤波器的缺陷, Sage 滤波器利用历史信息的平均来估计当前的系统噪声和观测噪声,再应用经典 Kalman 滤波器模型进行解算,获得系统状态的最优估值。在实际应用中,为了节省内存和减轻计算负荷,常常采用开窗法,即利用最新历元的新息向量或残差向量来估计当前的系统噪声和观测噪声<sup>[9]</sup>。公式如下:

$$Q_k = \frac{1}{m} \sum_{i=0}^{m-1} VT_{k-i} VT_{k-i}^T - AP_{k-1}A^T + P_k \quad (3)$$

$$R_k = \frac{1}{m} \sum_{i=0}^{m-1} VZ_{k-i} VZ_{k-i}^T - HP_k^T H^T \quad (4)$$

其中:  $VT_k = \hat{T}_k - \hat{T}_k^-$ ;  $VZ_k = H\hat{T}_k^- - Z_k$ ;  $m$ : 历元窗口宽度。

由式(3)可以看出,  $Q_k$  的求解式中含有  $P_k$ , 而求  $P_k$  又必须知道  $Q_k$ , 因而根本无法求解, 因此式(3)无解, 一般采用式(5)来近似代替式(3)<sup>[10]</sup>, 如下:

$$Q_k = \frac{1}{m} \sum_{i=0}^{m-1} VT_{k-i} VT_{k-i}^T \quad (5)$$

### 1.3 改进的 Sage 滤波器

从 Sage 滤波器的观测噪声估计式可以看出, 这种预测是对历史精度信息的平均, 其可靠性取决于当前历元观测精度与历史历元观测精度的一致性, 否则这种预测的代表性和可靠性就很难保证<sup>[11]</sup>。在观测噪声的估计中, 当观测值不出现异常数据时, 则预测值能在一定程度上反映当前历元观测信息的统计特性; 但当观测值中含有异常时, 估计的当前历元的观测噪声将会偏离观测信息的实际精度, 使滤波结果出现较大的偏差。因此, 需要对 Sage 滤波器进行改进。

考虑到公交运营具有重复性和规律性的特点, 即当前时刻观测值与历史上最近一段时期内对应时刻的观测值具有一定的相似性, 因而可以利用这种相似性来修正观测噪声。文中考虑利用经过聚类以后、对应时刻的车辆行程时间历史数据流信息对 Sage 滤波器的预测值和观测噪声估计式进行补充和改进(关于历史数据样本集聚类的相关参数说明将在下一节论述), 从而平滑异常观测值对预测结果的影响。修改后的结果如下:

$$\hat{T}_k^- = \alpha A \hat{T}_{k-1} + (1 - \alpha) CR_k \quad (6)$$

$$R_k = \frac{1}{m} \sum_{i=0}^{m-1} (\beta VZ_{k-i} VZ_{k-i}^T + (1 - \beta) VC_{k-i} VC_{k-i}^T) - HP_k^T H^T \quad (7)$$

其中:  $VC_k = CR_k - Z_k$ ;  $CR_k = \sum_{j=1}^N X_0(k)_j S(k)_j$ ;

$X_0(k)_j$  为对应时刻聚类样本集中的簇  $C_j$  的中心值;  $S(k)_j$  为对应时刻聚类样本集中的簇  $C_j$  的样本命中率;  $N$  为历史样本集窗口大小;  $\alpha$  和  $\beta$  为权值。

另外, Sage 滤波器需要利用上一状态的观测值对新的估计进行及时的纠正, 而实时数据会有丢失的情况出现, 这就会导致观测值的不连续性, 从而使得预测值得不到及时有效的修正; 此外, Sage 滤波器中预测初始状态通常也是未知的, 人为设置其大小往往会偏离实际。针对上两种情况, 这里考虑利用聚类以后、对应时刻的历史数据流的数值  $CR_k$  进行补充和替代, 从而提高预测的准确率。

## 2 车辆行程时间历史数据流聚类和管理

改进的 Sage 滤波器模型需要车辆行程时间历史数据流信息的支持, 然而数据流具有按时间顺序、快速变化、海量和潜在无限等特点, 因此不可能存储整个数据流或者对其扫描多次。为有效处理和利用流数据, 文中采用类似流聚类<sup>[12]</sup>的方法对每天内的车辆行程时间历史数据流进行清理和规约, 以获得一段时期内的历史样本集; 然后借鉴流数据分类的分类器系综方法<sup>[12]</sup>的思想, 对不断增加的历史样本集进行管理。

### 2.1 车辆行程时间历史数据流聚类

文中以天为单位, 对车辆行程时间历史数据流进行分段聚类, 从而获得每天的样本集, 进而得到一段时期内的历史样本集。

对于给定某天内的历史数据流  $X = \{x_i \mid x_i \in R\}$ , 首先按照如下方法定义其簇  $C_i$  的相关参数计算公式: 簇和簇  $C_j$  之间距离公式

$$d(C_i, C_j) = \sum_{x \in C_i} \sum_{x' \in C_j} |x - x'| \quad (8)$$

簇  $C_i$  半径公式(其中  $x_0$  为簇的中心)

$$R(C_i) = \frac{\sum_{x \in C_i} |x - x_0|}{n} \quad (9)$$

基于以上公式, 车辆行程时间历史数据流聚类算法流程如下:

- (1) 计算最大阈值  $M$ , 其值等于数据流内所有数据的方差;
- (2) 对于簇  $C_i$ , 分别计算与其直接邻近的簇  $C_{i-1}$  和  $C_{i+1}$  的距离, 找到两者中与簇  $C_i$  距离最小的簇, 记为  $C'_i$ ; 如果把簇  $C_i$  和簇  $C'_i$  进行合并后, 其结果中的簇半径不大于最大阈值  $M$ , 则合并簇  $C_i$  和簇  $C'_i$ , 否则不进行合并;
- (3) 循环, 直至所有的簇都不能再合并为止;
- (4) 遍历所有的簇, 判断并删除离群点;
- (5) 如果步骤(3)中还能发现离群点, 则重新计算

最大阈值  $M$ , 其值等于各个簇半径的均值, 然后转步骤(2); 否则结束整个聚类过程。

对于簇  $C_i$ , 只要不满足如下条件之一, 即可判定为离群点<sup>[13]</sup>:

① 簇  $C_i$  中数据对象数目少于最小阈值;

②  $d(C_i, C_{i-1}) > R(C_i) + R(C_{i-1})$  且  $d(C_i, C_{i+1}) > R(C_i) + R(C_{i+1})$ 。

## 2.2 历史样本集管理

随着时间的推移, 以天为单位聚类以后所形成的历史样本集会不断聚集和增长, 因此需要对其进行管理, 从而去除冗余的, 或者不再能反应当前历史趋势的样本集数据。文中以每个样本集为基本单位, 以样本集里面的簇为最小单位, 对整个历史样本集进行管理。

对于新增加的样本集, 按照如下过程对整个历史样本集进行添加、更新和删除:

(1) 基于当前整个的历史样本集数据, 用改进的 Sage 滤波器模型对当前车辆行程时间历史数据流进行预测, 并计算相对误差  $\delta$ 。

(2) 如果误差  $\delta$  大于最大允许值, 则分两种情况分别处理: 如果当前历史样本集的数目还未达到最大历史样本集窗口大小  $N$ , 则直接把新增的样本集添加到历史样本集中; 否则, 分别计算历史样本集里面已有的样本集的命中率(其大小等于该样本内所有簇平均命中率的均值), 剔除命中率最小的样本集后, 再把新样本集加入到历史样本集中。

(3) 否则, 如果误差  $\delta$  小于等于最大允许值, 则遍历当前整个历史样本集里面的所有样本集, 分别与当前车辆行程时间数据流进行比照, 更新其相应簇  $C_i$  的命中率  $S_i$ 。

## 3 预测及结果分析

文中针对 2007 年 6 月 7 日全天北京市南中轴路大容量快速公交(BRT)线(共计 17 个站台, 全程 31.7 公里, 90 辆运营车辆)、各个站台感应线圈采集到的数据, 分别运用 Kalman 滤波器、Sage 滤波器以及改进的 Sage 滤波器对全天车辆的全程行程时间进行预测, 其对比结果如表 1 所示。

表 1 三种滤波器误差对比

	Kalman 滤波器	Sage 滤波器	改进的 Sage 滤波器
平均误差率	10.06	9.89	9.06
最大误差率	65.28	65.25	32.71

通过对比可以看出, 经典的 Kalman 滤波不能抵制异常观测值的影响, 当观测值中含有异常扰动时, 经典 Kalman 滤波的效果比较差, 甚至可能发散。Sage 滤波器利用历史信息的平均来估计当前的系统噪声和观测

噪声, 能在一定程度上近似当前历元的噪声统计特性, 从而达到改善滤波的效果, 但当存在观测异常和动态扰动异常时, 会使得估计偏离实际的噪声统计特性, 使滤波结果出现偏差, 甚至发散。改进的 Sage 滤波器通过聚类以后、对应时刻的车辆行程时间历史数据流信息对 Sage 滤波器的观测噪声估计式进行修正, 使之自适应于观测信息, 取得了很好的滤波结果。

## 4 结束语

针对 Sage 滤波器自身的优缺点, 提出一种改进的 Sage 滤波器, 其主要思想是利用经过聚类处理以后、对应时刻的车辆行程时间历史数据流信息, 对 Sage 滤波器的预测值以及观测噪声估计式进行补充和改进, 从而平滑异常观测数据对预测结果的影响; 同时, 通过利用该历史数据对实际采集过程中丢失的数据进行补充, 并及时修正预测结果。之后, 提出了对车辆行程时间历史数据流信息进行聚类和管理的方法。最后, 在实际数据基础上, 通过与经典 Kalman 滤波器以及 Sage 滤波器的预测结果进行对比, 可以看出改进的 Sage 滤波器避免了观测异常和动态扰动异常的影响, 从而提高了车辆行程时间预测的准确性。

## 参考文献:

- [1] Ishak S, Ai - Deek H. Performance evaluation of short - term time - series traffic prediction model[J]. Journal of Transportation Engineering, 2002, 128(6): 490 - 498.
- [2] D' Angelo M P, Al - Deek H M, WANG M C. Travel - time prediction for freeway corridors[R]. Transportation Research Record 1676. Washington, DC: [s. n.], 1999: 184 - 191.
- [3] Kuchipudi, Mouly C, Chien, et al. Development of a hybrid model for dynamic travel time prediction[C]//Transportation Research Board, Annual Meeting. Washington DC: [s. n.], 2002.
- [4] Lin, Weihua, Amit, et al. Arterial travel time estimation for advanced traveler information systems[C]//Proceedings of the 82th Annual Meeting of the Transportation Research Board. Washington D. C., USA: National Academies Press, 2003.
- [5] Robinson, Steve J P. Modeling urban link travel - time using data from inductive loop detectors[C]//World Conference on Transport Research. Istanbul, Turkey: [s. n.], 2004.
- [6] Bajwa S U I, Chung E, Kuwahara M. A travel time prediction method based on pattern matching technique[C]//21st ARRB and 11th REAAA Conference. Cairns, Australia: [s. n.], 2003.
- [7] Van Lint W C. Reliable Travel Time Prediction for Freeways

(下转第 169 页)

```

.....
SavePath= SaveUpFilesPath '存放上传文件的目录
for each formName in upload.file '列出所有上传了的文件
set ofile= upload.file(formName) '生成一个文件对象
if ofile.filesize>(MaxFileSize*1024) then msg= "文件大小超过
了限制,最大只能上传"& CStr(MaxFileSize) & "K的文件!"
Fmime= obj.filetype("file") '得到文件的类型 SaveData= obj.
Save
Fileto db("file")
- 保存文件 -
ofile.SaveToFile Server.mappath(fileName)'保存文件
msg= "上传文件成功!"
%>

```

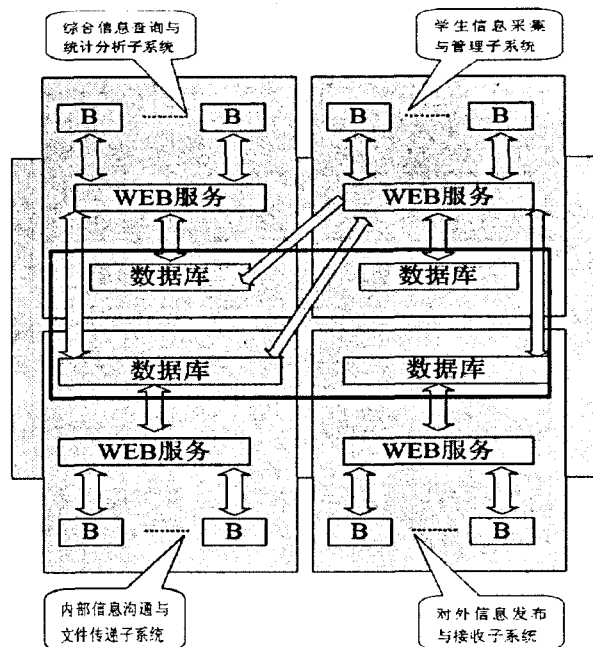


图4 学生工作管理系统结构示意图

说明:B表示(浏览器)用户;四个子系统既可以独立运行,又相互联系,彼此共享相应数据。

(2)系统同时具有远程备份数据库的功能,下面是实现该功能页面 Admin\_Database.asp 中的关键部分:

```

<% .....
input type= text size= 20 name= bkfolder value= Databackup '备
份相对路径目录,如果目录不存在,将自动创建

```

```

.....
<td height="40" colspan="3"><input name="submit" type=
submit value="开始备份"><% If ObjInstalled= false Then re-
sponse.Write "disabled"%>>
.....
response.write "<center>备份数据库成功,备份的数据库为 "&
bkfolder & "\" & bkdbname & ".mdb</center>"返回备份
成功信息
%>

```

## 5 结束语

学生信息管理系统中,涵盖了学生信息管理工作的绝大部分功能,满足了用户对数据查询、添加、修改和删除的基本需要,提出了基于数据库和自建用户权限体系的系统安全策略,基本满足用户对数据安全的要求,并且尽量做到程序设计人性化,方便用户进行操作。

学生信息管理系统是当今高校管理信息化和办公自动化的一个重要组成部分,该项目作为江苏省教育科学“十一五”规划课题中的重点课题(批准号 B-a/2006/01/022)的子项目和徐州建筑职业技术学院重点科研课题(院字[2007]14-Yjk06-8)开发成功,正式投入使用以来,运行状况良好,使学校的管理和教学更加趋于科学和高效,极大地提高了工作效率,具有较高的推广价值,这对于建设高校现代化的管理和决策无疑发挥了积极的作用。

## 参考文献:

- [1] 刘芳,龚明.基于ASP技术的动态站点的研究与开发[J].电脑与信息技术,1999(1):14-16.
- [2] 李华斌.Active Server Pages 程序设计与开发[M].北京:中国水利水电出版社,2000.
- [3] 王玉莲.基于ASP技术的学生信息管理系统的研究与开发[J].计算机工程及应用,2004(11):218-220.
- [4] Dwight J,Erwin M,Niles R.CGI 开发使用手册[M].北京:机械工业出版社,1998.
- [5] Schach S.Software Engineering With Java[M].袁兆山译.北京:机械工业出版社,2003.

(上接第164页)

- [D]. Delft, Netherlands: Delft University Press, 2004.
- [8] 杭明升,杨晓光,彭国雄.基于卡尔曼滤波的高速道路行程时间的动态预测[D].同济大学学报:自然科学版,2002,30(9):1069-1072.
- [9] 崔先强.噪声协方差矩阵加权估计的 Sage 自适应滤波器[J].测绘科学,2002,27(2):26-30.
- [10] 曾翠娟,王忠,兰竹,等. GPS 动态定位的自适应卡尔曼滤波算法研究[J].导航,2006,42(1):39-48.
- [11] 杨元喜,何海喜,徐天河.论动态自适应滤波[J].测绘学报,2001(4):294-298.
- [12] Han Jiawei.数据挖掘概念与技术[M].范明,孟小峰,译.北京:机械工业出版社,2007:251-345.
- [13] 倪巍伟,陆介平,陈耿,等.基于k均值分区的数据流离群点检测算法[J].计算机研究与发展,2006,43(9):1639-1643.