

基于马尔科夫过程的行驶工况 构建中数据处理与分析

李友文¹, 石琴², 姜平²

(1. 合肥工业大学 机械与汽车工程学院, 安徽 合肥 230009; 2. 合肥工业大学 交通运输工程学院, 安徽 合肥 230009)

摘要:在利用马尔科夫过程构建行驶工况的方法中,将车辆的行驶状况看做一个随机过程,只需求得实验数据的状态转移矩阵,便可得到随机过程的统计性质,通过对实验数据进行状态划分、模型事件分类、模型事件集确定,最终可得到实验数据的转移矩阵。文章简述了马尔科夫构建行驶工况的基本原理及实验数据的处理方法和转移矩阵的求法,并利用 VBA 语言设计了一种数据分析处理程序。

关键词:马尔科夫过程; 模型事件; 行驶工况; VBA 语言

中图分类号:U121 **文献标识码:**A **文章编号:**1003-5060(2010)04-0491-04

Data disposal and analysis for driving cycle construction based on Markov process approach

LI You-wen¹, SHI Qin², JIANG Ping²

(1. School of Machinery and Automobile Engineering, Hefei University of Technology, Hefei 230009, China; 2. School of Transportation Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: In the process of driving cycle construction with Markov process approach, vehicular travel is treated as a stochastic process. The statistical character of the stochastic process is determined through the state transition matrix of experimental data. The final transition matrix of experimental data is obtained by analyzing state partition, model event partition, model event bin confirmation. The paper introduces basic principle of Markov process approach, the processing method of experimental data and the resolution method of transition matrix. Finally a data process program is designed with VBA language.

Key words: Markov process; model event; driving cycle; VBA language

马尔科夫过程是指某一随机过程在已知“现在”的条件下,具有“将来”与“过去”是独立的性质^[1]。这与城市车辆的行驶情况具有高度吻合性,因此用马尔科夫过程来开发行驶工况具有理论上的可行性。

同时在当今行驶工况研究领域,研究人员利用速度-时间数据构建行驶工况的时候都将其看作了确定的量,这与实际是不符的。因为通过研究数据和采集技术,会发现任意时刻的速度-时间

图都不具有确定性,而运用马尔科夫的性质可以很好地解决这一问题。

行驶工况的开发是一种综合运用多种试验方法和数学方法处理问题的过程。试验数据的处理工作是构建工况的基础,在工况的开发过程中占据重要的分量。

虽然不同工况构建方法具有不同的特点和表现内容,但它们的开发过程可以基本分为以下4个阶段,如图1所示。

收稿日期:2009-04-13;修改日期:2009-06-15

基金项目:国家自然科学基金资助项目(70771036);安徽省自然科学基金资助项目(070416244)

作者简介:李友文(1985-),男,安徽宿州人,合肥工业大学硕士生;

石琴(1963-),女,安徽蚌埠人,博士,合肥工业大学教授,博士生导师。



图 1 行驶工况开发的基本过程

数据处理主要是指对构成车速-时间曲线的实验原始数据的分析,从中提取出构建工况所需要的信息。不同的构建过程对应不同的数据处理方法,如利用主成分分析的方法构建行驶工况时,是利用运动学片断组合构建,需要划分每一个运动学片段,然后计算这些片段的特征值,形成具有特征值的片段数据库,并利用主成分分析法和聚类分析法构建最终的行驶工况^[2,3]。在利用马尔可夫方法构建行驶工况的过程中,需要对数据进行加速、减速、匀速和怠速片段的划分,每个被划分出的片段叫做模型事件,再将具有相似平均速度和平均加速度的片段收集在一起,称为模型事件集,全部模型事件集被定义为马尔科夫链状态空间中的“状态”。状态转移的条件概率被用来估计每个状态,转移概率用来形成转移矩阵^[4]。

1 构建行驶工况的基本原理

对实验数据进行分类,按照加速度的变化把速度分成 G 族,族号为 $g=1,2,\dots,G$,以及一系列未知参数 θ_k ,它们与观测值有关,可以被理解为总体的期望和方差。采用极大似然估计法(Maximum Likelihood Estimate,简称 MLE)作为划分的理论依据,并同时检验划分结果的合理性。该过程 MLE 函数如下:

$$L(\boldsymbol{\theta} | \bar{\mathbf{a}}) = \prod_{g=1}^G (\pi_g^{n_g} \sigma_g^{-n_g}) \exp \left\{ -\frac{1}{2} \sum_{g=1}^G \sum_{C_g} \frac{(a_i - \mu_g)^2}{\sigma_g^2} \right\} \quad (1)$$

其中, \mathbf{a} 表示加速度 a_i 的矢量; $\boldsymbol{\theta}$ 表示相应的参数的矢量; n_g 为 C_g 中观察点的数目; μ_g 为 g 族中观察点的期望值; σ_g 为 g 族中观察点的方差; C_g 为加速度属于 g 组的集合; π_g 表示加速度 a_i 在 g 族的概率。

每个族包括具有加速、减速、匀速和怠速 4 种片段的模型事件。再次利用极大似然估计法,将具有相似速度的模型事件集中在一起,称为模型事件集,将每个模型事件集定义为马尔科夫链的一个状态 P_i ,所有的模型事件集组成马尔科夫链状态空间。可以得到状态转移矩阵:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix}_{K \times K} \quad (2)$$

其中, p_{rs} 为状态 S_r 到状态 S_s 的条件概率,称为转移概率。状态转移矩阵具有如下的性质:

- (1) 所有的元素有非负性。
- (2) 每行元素之和为 1,即:

$$\sum_{s=1}^K p_{rs}(\tau) = 1, \quad r = 1, \dots, K \quad (3)$$

第 1 个性质是很显然的,第 2 个性质说明各个状态之间相互独立,且状态有限可列。

2 数据的划分及转移矩阵的求法

实验中利用非接触式智能测速仪采集到的速度信号,通过 AM-2600S 汽车性能实验处理系统的数据库,可以形成连续的速度-时间关系图,并根据人为需要提供不同的距离步距和时间步距的文本文件。无论是瞬态工况还是模态工况,行驶工况的最终表达形式都为速度-时间曲线,且时间步长通常为 1 s,因此选用时间步距为 1 s 的文本文件,对其进行数据处理。

将实验数据导入 Excel 中,设实验数据总体为 n ,相邻时间的速度之差(单位时间速度差)可看作车辆的单位加速度 a_{t-1} :

$$a_{t-1} = \frac{dv}{dt} = \frac{\Delta v}{\Delta t} = \frac{v_t - v_{t-1}}{3.6}, \quad 2 \leq t \leq n \quad (4)$$

利用加速度-时间实验数据划分为代表加速、减速、匀速、怠速的很多模型事件。划分过程中注意以下事项:

(1) 对于 $0 < |a_{t-1}| \leq 0.1 \text{ m/s}^2$ ^[5] 的情况,即 $0 < |v_t - v_{t-1}| \leq 0.36 \text{ km/h}$,将该处加速度近似看作 $a_{t-1} = 0 \text{ m/s}^2$ 。

(2) 在出现(1)中情况时,若对应速度同时为 0 km/h,该处应归为怠速情况,称为怠速优先性。

(3) 每个数据不可以重复使用。

计算每个模型事件的平均速度,将平均速度处于相同速度区间的模型事件归纳在同一个模型事件集中。模型事件集的确定,标志着马尔科夫过程的状态空间的确定。对于整个数据而言,既有本状态内部的转移,又包含各个状态之间的转移。为了实现编程的可行化,利用颜色的变化来实现,即将处于同一个事件集的数据用同一种颜色来标示,这样状态之间的转化就被认为是颜色之间的转化。记录各种转化的次数,用来计算转移概率。在马尔可夫过程中,状态的转移概率为条件概率,如状态 r 到状态 s 的转移概率为 $P\{X_r = s | X_{r-1} = r\}$,利用极大似然函数(MLE),得到状态转移概率方程为:

$$P_{rs} = N_{rs} / \sum_s N_{rs} \quad (5)$$

其中, N_{rs} 表示从状态 r 转到状态 s 的频率; $\sum N_{rs}$ 表示从状态 r 转到所有状态的频率。

3 行驶状态特征参数的选择

计算相应的行驶状态特征参数,是用来对整个实验过程中车辆行驶的基本情况进行定性的说明,同时可以对构建行驶工况的合理性与正确性进行评估和选择。本方法采用多组特征参数,包括每个模型事件的平均速度、平均加速度,每个模型事件集的平均速度、平均加速度、最大速度、最小速度、最大加速度和最小加速度等,见表 1 所列,其中 i 为模型事件个数, j 为模型事件集的个数。

表 1 行驶状态特征参数分类

特征参数	物理意义
avg v_i	各模型事件平均速度
avg a_i	各模型事件平均加速度
avg v_j	各模型事件集平均速度
avg a_j	各模型事件集平均加速度
max v_i	各模型事件集最大速度
min v_i	各模型事件集最小速度
max a_j	各模型事件集最大加速度
min a_j	各模型事件集最小加速度
⋮	⋮

由于模型事件和模型事件集的相关处理在第 2 部分已经完成,只要通过 VBA 编写处理程序,利用冒泡法筛选出各种模型事件及模型事件集中速度与加速度的最大值和最小值,同时也可以计算出它们的平均值^[6,7]。

4 程序系统设计及应用

遵循数据划分原则,程序的设计流程如图 2 所示。在程序系统的设计过程中,需要说明的是:

(1) 确定数据总数是为了避免进入死循环,各模型事件起始点和结束点的正确划分是程序编

写的关键。

(2) 将速度与加速度对应起来, v_{t-1} 对应 a_{t-1} 表明了 $t-1$ 时刻到 t 时刻的速度 v 的变化情况。

(3) 在确定模型事件时,由于怠速情形具有优先性,应先划分出怠速片段,然后对剩下的数据进行其它速度片段的划分。

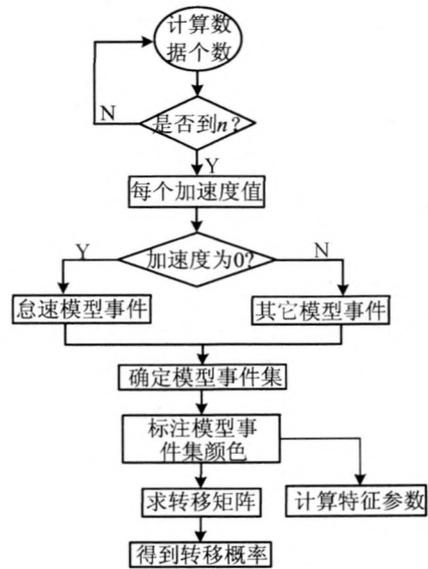


图 2 程序流程框图

应用本系统对所采集的合肥市典型道路的车辆行驶数据进行了处理和分析。实验选取了合肥市胜利路、屯溪路和明光路等 5 条典型代表性道路进行速度-时间数据的采集。按照行驶工况的构建过程,把实验数据按照加、减速度分成若干独立的片段,把具有相似的速度片段(加速度暂未考虑,将在后期修正时加入)组合在一起,划分为 6 个状态,分别是(按平均速度):

- 0~10 km/h, 10~20 km/h,
- 20~30 km/h, 30~40 km/h,
- 40~50 km/h, 大于 50 km/h。

计算出每个状态的特征参数和状态转移概率,程序运行结果见表 2 和表 3 所列。

表 2 实验数据状态特征参数值

状态	max v	min v	avg v	max a	min a	avg a
1	21.59	0	2.086 01	2.711 111	-2.202 78	0.037 24
2	42.60	0.46	15.384 93	2.2	-2.202 78	-0.045 93
3	55.24	0.46	24.982 08	2.202 771	-2.202 78	-0.014 34
4	62.82	2.30	34.682 06	1.752 778	-2.191 11	-0.015 95
5	67.75	18.03	45.106 90	1.244 444	-2.266 67	0.016 01
6	75.33	33.76	57.732 30	1.594 444	-1.913 89	-0.001 36

表 3 状态转移概率

状态	1	2	3	4	5	6	%
1	82.575 02	15.585 67	1.839 30	0	0	0	
2	16.800 92	59.263 52	22.439 59	1.495 97	0	0	
3	3.240 22	16.759 78	58.324 02	19.664 80	1.899 44	0.111 73	
4	0.676 59	5.142 08	16.508 80	60.487 14	16.373 48	0.811 91	
5	0	0.516 79	3.746 77	11.369 51	71.059 43	13.307 49	
6	0	0	0.934 58	1.869 16	10.046 73	87.149 53	

表 2 中字母的含义见表 1 中各种特征参数的定义,速度单位为 km/h,加速度单位为 m/s^2 。表 3 中数字 1~6 分别代表状态 1~状态 6。

通过总数相加、数据查询等必要验证,数据分析结果准确、可信,真实反应了实验数据的内容。

5 结束语

运行结果证明,该程序可以很好地满足对实验数据处理的要求,为工况的最终构建提供了参照和依据,应成为最终形成的工况开发软件的重要组成部分。

由于 VBA 具有和 VB 相似的语言结构和几乎相同的集成开发环境(IDE),开发相应的 VB 与 VBA 接口程序就可以利用 VB 调用 VBA 程序处理 Excel 中的数据,同时利用 VB 和 MATLAB 的混合编程直接输出最终的行驶工况图。

正是由于车辆行驶的不确定性,因此可将行驶工况看作一个随机过程。对实验数据划分结果,利用对数似然比 χ^2 来检验行驶工况各状态间的非独立性,根据数理统计的 χ^2 表,同时利用 SPSS14.0 统计学软件,计算出的 χ^2 值远远大于 χ^2 表中的值^[8],所以汽车行驶工况各状态之间相

互独立的假设不成立,因此将马尔科夫过程论运用在汽车行驶工况研究中是完全可行的。

[参 考 文 献]

- [1] 杜雪樵,惠 军. 随机过程 [M]. 合肥:合肥工业大学出版社,2006:15-45.
- [2] 朱西产,李孟良,马志雄,等. 车辆行驶工况开发方法 [J]. 江苏大学学报(自然科学版),2005,26(2):110-112.
- [3] 李孟良,朱西产,张建伟,等. 典型城市车辆行驶工况构成的研究 [J]. 汽车工程,2005,27(5):558-560.
- [4] A T 巴鲁查-赖特. 马尔科夫过程论初步及其应用 [M]. 杨纪珂,吴立德,译. 上海:上海科学技术出版社,1979:132-166.
- [5] Lin Jie. A Markov progress approach to driving cycle development[D]. University of California, Davis,2002.
- [6] 蔡士源. Excel VBA 语法字典 [M]. 北京:中国青年出版社,2005:106-112.
- [7] 谭浩强,袁 玫,薛淑斌. Visual Basic 程序设计 [M]. 北京:清华大学出版社,2005:8-100.
- [8] Lin Jie, Niemeier D A. Estimating regional air quality vehicle emission inventories; constructing robust driving cycles [J]. Transportation Science, 2003,37(3):330-346.

(责任编辑 吕 杰)