

文章编号: 1001-0920(2009)08-1177-04

## 基于跳转模型的路网交通流预测

陈岳明, 萧德云

(清华大学 自动化系, 北京 100084)

**摘要:** 针对城市道路交通流复杂的非线性特征, 采用基于跳转的 ARIMA 模型研究交通流的变化规律, 以获得城市道路短期交通流的精确预测. 引入路口转弯比例矩阵来描述路网的交通流状态. 考虑在路网畅通和路网拥堵情况下路口转弯比例矩阵的不同特征, 对两种情况分别进行研究. 结合路段交通流预测模型, 对路网交通流进行预测. 仿真算例表明, 所提出的预测模型能较好地用于路网交通流预测.

**关键词:** 交通拥堵状态; 交通流预测; 跳转 ARIMA 模型; 转弯比例矩阵

**中图分类号:** U491

**文献标识码:** A

## Traffic network flow forecasting based on switching model

CHEN Yue-ming, XIAO De-yun

(Department of Automation, Tsinghua University, Beijing 100084, China. Correspondent: XIAO De-yun, E-mail: xiaody@mail. tsinghua. edu. cn)

**Abstract:** For the nonlinearity of complicated urban traffic flow, a switching autoregressive integrated moving average (ARIMA) model is proposed and employed to explore how traffic flow varies with time, so that accurate predictions can be obtained. The turning proportion matrix is introduced to describe the traffic network states. Considering the difference characters of the turning proportion matrix between congestion case and discongestion case, two formulas are proposed to forecast the network traffic flow. The simulation results show the effectiveness and feasibility of the proposed approach.

**Key words:** Congested traffic states; Traffic flow forecasting; Switching ARIMA model; Turning proportion matrix

### 1 引言

城市道路交通网络的实时交通流预测是实现交通诱导和控制的关键, 是智能交通系统的一个重要组成部分<sup>[1]</sup>. 由于交通流运行的复杂性和不确定性, 要获得准确的预测结果并不容易. 近年来, 学者们在该领域进行大量的研究, 取得了不少研究成果. 目前广泛应用的模型有历史均值法、回归预测法、时间序列法、神经网络预测法、卡尔曼滤波法、模糊控制法等<sup>[2-5]</sup>.

路段交通流序列在变化过程中可划分为上升态、下降态、高峰平台期和低谷平台期 4 个模态, 而对这些不同的模态用一个模型来描述往往是不够的. 跳转模型<sup>[6]</sup>引入一个隐含变量, 这个变量在不确定的时刻离散变化, 不同的值对应于不同时间序列的参数.

本文引入基于跳转的 ARIMA 模型来描述交通

流序列, 并进行路段交通流预测. 基于路口转弯比例的时间稳定性和数值稳定性的假设, 考虑路网畅通和路网拥堵对路口转弯比例矩阵的影响, 分别对这两种情况进行研究, 得出了畅通和拥堵情况下的路口转弯比例矩阵. 结合基于跳转 ARIMA 模型的路段交通流预测, 实现对路网交通流的预测.

### 2 基于跳转模型的路段交通流预测

设交通网络含有  $n$  个节点  $1, 2, \dots, n$  (即道路交叉口), 以  $l_{ij}$  表示连接节点  $i$  与  $j$  的路段, 以  $Y_i$  表示路段上的车辆数, 其中  $i$  为路段标号. 对于单独路段, 交通流量时间序列  $\{Y^k\}$  ( $k$  表示时刻) 常常表现为不规则的复杂行为, 注意到路段交通流序列在变化过程中可划分为上升态、下降态、高峰平台期和低谷平台期 4 个模态, 引入基于跳转的 ARIMA 模型<sup>[7]</sup>来描述交通流序列, 并进行路段交通流预测.

以  $S_k$  表示  $k$  时刻交通流序列所处的模态, 共有

收稿日期: 2008-07-29; 修回日期: 2008-10-14.

基金项目: 国家“十五”科技攻关项目(2005BA414B09); 国家“十一五”科技支撑项目(2007BAK12B15).

作者简介: 陈岳明(1981—), 男, 湖南岳阳人, 博士生, 从事异常交通分析、交通控制的研究; 萧德云(1945—), 男, 福建仙游人, 教授, 博士生导师, 从事故障诊断、智能交通等研究.

$M$  个可能的状态,表示为  $1, 2, \dots, M$ . 这里  $M = 4$ , 分别对应于波谷态、上升态、波峰态和下降态. 注意到这些状态不能任意跳转,只能从波谷态到上升态,从上升态到波峰态,从波峰态到下降态,从下降态到波谷态. 以符号  $l_{S_k}$  表示状态  $S_k$  的持续时间. 假设状态转移方程具有如下形式:

$$p(S_{k+1} = j | S_k = i \text{ at } l_i) = b_{ij}(l_i). \quad (1)$$

其中:  $b_{ij}(l_i)$  是矩阵  $\mathbf{B}(l)$  的第  $i$  行第  $j$  列元素,

$$\mathbf{B}(l) = \begin{bmatrix} 1 - g_1(l_1) & g_1(l_1) & 0 & 0 \\ 0 & 1 - g_2(l_2) & g_2(l_2) & 0 \\ 0 & 0 & 1 - g_3(l_3) & g_3(l_3) \\ g_4(l_4) & 0 & 0 & 1 - g_4(l_4) \end{bmatrix}, \quad (2)$$

$$g_i(l_i) = \text{Sigm}\left(\frac{l_i - \mu_i}{\sigma_i}\right), \quad i = 1, 2, 3, 4. \quad (3)$$

式中:  $\text{Sigm}(x)$  是 Sigmoid 函数<sup>[8]</sup>, 参数  $\mu_i$  和  $\sigma_i$  为控制状态  $S_k = i$  持续时间的均值和方差.

对于一般的 ARIMA 模型, 若已知时间序列  $\{Y^k\}$  当前及以前时刻的值, 则对  $Y^{k+1}$  的最佳估计为

$$\hat{Y}^{k+1} = \sum_{n=1}^p \Phi_n Y^{k+1-n} + \sum_{m=1}^q \Theta_m e^{k+1-m}. \quad (4)$$

其中:  $Y^k, \dots, Y^{k+1-p}$  表示  $Y^{k+1}$  的前  $p$  个值, 组成了模型的自动回归部分;  $\Phi_1, \dots, \Phi_p$  表示自动回归因子的系数;  $e^k, \dots, e^{k+1-q}$  是零均值白噪声过程, 组成了模型的滑动平均部分;  $\Theta_1, \dots, \Theta_q$  表示滑动平均因子的系数, 这里差分因子系数取为 0.

对于跳转 ARIMA 模型, 若能预先知道  $k+1$  时刻的状态  $S_{k+1}$ , 则交通流预测算法可简单写成

$$\hat{Y}^{k+1|S_{k+1}} = \sum_{n=1}^p \Phi_n(S_{k+1}) Y^{k+1-n} + \sum_{m=1}^q \Theta_m(S_{k+1}) e^{k+1-m}. \quad (5)$$

然而, 隐含状态  $S_{k+1}$  往往是事先未知的, 但给定当前和以前时刻的时间序列值, 隐含状态  $S_{k+1}$  的概率分布是可计算的. 因此基于跳转 ARIMA 模型的预测算法为

$$\hat{Y}^{k+1} = \sum_{S_{k+1}=1}^M \hat{Y}^{k+1|S_{k+1}} p(S_{k+1} | Y^{k:0}, e^{k:0}). \quad (6)$$

其中

$$\begin{aligned} p(S_{k+1} | Y^{k:0}, e^{k:0}) &= \\ & \sum_{S_k=1}^M \sum_{l_{S_k}=1}^L p(S_{k+1} | S_k, l_{S_k}) p(l_{S_k} | Y^{k:0}, \\ & e^{k:0}) p(S_k | Y^{k:0}, e^{k:0}), \quad (7) \\ p(S_k | Y^{k:0}, e^{k:0}) &= \end{aligned}$$

$$\frac{p(Y^k | Y^{k-1:0}, e^{k-1:0}, S_k) p(S_k | Y^{k-1:0}, e^{k-1:0})}{\sum_{S_k=1}^M p(Y^k | Y^{k-1:0}, e^{k-1:0}, S_k) p(S_k | Y^{k-1:0}, e^{k-1:0})}, \quad (8)$$

$$\begin{aligned} p(l_{S_k} = d | Y^{k:0}, e^{k:0}) &= \\ & \begin{cases} p(l_{S_{k-1}} = d - 1 | Y^{k-1:0}, e^{k-1:0}), & d \geq 2; \\ \sum_{S_{k-1}=1}^M \sum_{l_{S_{k-1}}=1}^L p(S_k | S_{k-1}, l_{S_{k-1}}) p(l_{S_{k-1}} | Y^{k-1:0}, \\ e^{k-1:0}) p(S_{k-1} | Y^{k-1:0}, e^{k-1:0}), & d = 1. \end{cases} \quad (9) \end{aligned}$$

式(7)和(9)中的  $L$  是状态持续时间的最大值.

当有新的观测值到达时, 预测误差序列可由下式计算:

$$e^{k+1} = Y^{k+1} - \hat{Y}^{k+1}. \quad (10)$$

### 3 路网畅通情形下的交通流预测

路网中的车行转弯比例用转移矩阵  $\mathbf{A}_k$  描述为

$$\mathbf{A}_k = (a_{ij}^k)_{n \times n}, \quad (11)$$

其中  $a_{ij}^k$  表示在  $k$  时刻从路段  $i$  转移到路段  $j$  的车辆比例.

车行转弯比例矩阵  $\mathbf{A}_k$  一般可通过实测数据直接获得, 也可用交叉口进出的流量来反算出交叉口转弯比例<sup>[9]</sup>. 利用上海市交叉口流量采集系统, 对 2005 年 3 月 200 个交叉口各车道的流量, 进行左、直、右 3 个方向的转弯比例分析, 发现交叉口进口各方向转弯比例呈现明显的时间稳定性和数值稳定性<sup>[10]</sup>. 假定路口转弯比例的时间稳定性和数值稳定性也适用于一般通畅情况下的交通网络.

路网道路交通状况用向量  $\mathbf{Y}_k$  描述为

$$\mathbf{Y}_k = (Y_1^k, Y_2^k, \dots, Y_n^k)^T, \quad (12)$$

其中  $Y_i^k$  表示  $k$  时刻第  $i$  条路段上的车辆总数. 基于稳定转弯比例, 假设  $a_{ij}^k = a_{ij}^0$  是与时间点  $k$  无关的常数, 可得

$$Y_j^{k+1} = \sum_{i=1}^n a_{ij}^k Y_i^k, \quad (13)$$

$$\begin{aligned} \mathbf{Y}_{k+1} &= (Y_1^{k+1}, Y_2^{k+1}, \dots, Y_n^{k+1})^T = \\ & \left( \sum_{i=1}^n a_{i1}^k Y_i^k, \sum_{i=1}^n a_{i2}^k Y_i^k, \dots, \sum_{i=1}^n a_{in}^k Y_i^k \right)^T = \\ & \mathbf{A}_k^T (Y_1^k, Y_2^k, \dots, Y_n^k)^T = \mathbf{A}_k^T \mathbf{Y}_k. \quad (14) \end{aligned}$$

运用基于跳转 ARIMA 模型的预测算法对路段交通流量  $Y_i^k$  进行预测, 基于路口转弯比例的稳定性假设, 由式(14)便可估计未来路网中的交通状况. 然而, 路口转弯比例的稳定性假设只对一般畅通情况下的路网适用, 因为当下游路段发生拥挤时, 路口转弯比例矩阵随着拥堵程度会发生变化.

### 4 拥堵条件下的路网交通流预测

对于道路拥堵的研究,可分为城市出行易达性和道路服务水平两个方面.本文采用道路饱和度作为定义道路是否拥堵的指标,即

$$\sigma_i = Y_i / C_i. \tag{15}$$

其中: $Y_i$  为第  $i$  条道路上的车辆数, $i$  表示道路的编号; $C_i$  为道路  $i$  的拥堵容限(常数),表示在第  $i$  条道路上车辆拥堵与畅通的分界线.当  $\sigma_i \geq 1$  时,则判定第  $i$  条道路拥堵.

需要说明的是,交叉口仍有可能向拥堵路段转移车辆.设  $C_i^*$  为最大容限,即第  $i$  条道路能容下的最大车辆数.此时路段已经堵死,外部车辆无法转入该条道路.

当某条道路上的车辆数逐渐增大,以至达到拥堵容限时,道路开始呈现拥堵态,车辆之间的距离变小,导致相互干扰加剧.在交叉口处,此时转移比例系数不再是常数,而是与下游车辆有关的函数,即

$$a_{ji} = h(\sigma_i). \tag{16}$$

当路段畅通时,转移比例稳定,有

$$h(\sigma_i) = a_{ji}^0, \sigma_i \in [0, 1]. \tag{17}$$

当路段车辆达到路段最大容限,即  $\sigma_i = C_i^* / C_i$  时,转移比例为 0,有

$$h(C_i^* / C_i) = 0. \tag{18}$$

根据式(17)和(18),用线性函数去近似逼近  $\sigma_i \in [1, C_i^* / C_i]$  的曲线,可得

$$h(\sigma_i) = -\frac{a_{ji}^0}{C_i^* - C_i} (C_i \sigma_i - C_i^*), \tag{19}$$

$$\sigma_i \in [1, C_i^* / C_i].$$

式(19)描述了当拥堵路段没有受到其他拥堵路段的影响时,路段转移比例与其道路饱和度之间的关系.

设上游路段  $j$  邻接多个下游拥堵路段  $i_1, i_2, \dots, i_s$  和多个畅通路段  $k_1, k_2, \dots, k_t$ , 设  $i_1, i_2, \dots, i_s$  道路的饱和度分别为  $\sigma_1, \sigma_2, \dots, \sigma_s$ , 且都大于等于 1. 根据式(19)分别求出单下游路口饱和和情况下路段  $i_1, i_2, \dots, i_s$  的期望转移比例系数,记为  $a_{j i_1}, a_{j i_2}, \dots, a_{j i_s}$ , 则有

$$a_{j i_r} = -\frac{a_{j i_r}^0}{C_{i_r}^* - C_{i_r}} (C_{i_r} \sigma_{i_r} - C_{i_r}^*), \tag{20}$$

$$r = 1, 2, \dots, s.$$

在多下游路段拥堵的情形下,期望转移比例系数表示由本道路饱和度确定出的理想转移比例,反映了交通网络对不同道路的需求分配,但由于不同拥堵路段之间会有相互干扰,这样的交通需求不能完全得到满足.因此需要考虑各拥堵路段之间的相互作用.

设  $i_r$  和  $i_m (1 \leq r, m \leq s)$  是交叉口下游任意两个拥堵路段,从  $i_r$  路段被退回来并重新分配到  $i_m$  的期望车辆比例为

$$a_1 = (a_{j i_r}^0 - a_{j i_r}) \frac{a_{j i_m}^0}{1 - a_{j i_r}^0}. \tag{21}$$

然而  $i_m$  路段也是拥堵路段,故有一部分车辆将被退返到  $i_r$  路段,退返率为

$$D_{\text{退}}^{i_m} = 1 - a_{j i_m} / a_{j i_m}^0. \tag{22}$$

从  $i_r$  路段重新分配到  $i_m$  的车辆,将有  $a_2$  的比例被退返回  $i_r$  路段,则

$$a_2 = (a_{j i_r}^0 - a_{j i_r}) \frac{a_{j i_m}^0}{1 - a_{j i_r}^0} \left(1 - \frac{a_{j i_m}}{a_{j i_m}^0}\right) = (a_{j i_r}^0 - a_{j i_r}) \frac{a_{j i_m}^0 - a_{j i_m}}{1 - a_{j i_r}^0}. \tag{23}$$

对下标  $m$  求和,因为拥堵被重新分配到其他路段的车辆将有  $a_2^*$  的比例又被退回  $i_r$  路段,则

$$a_2^* = \frac{a_{j i_r}^0 - a_{j i_r}}{1 - a_{j i_r}^0} \sum_{m=1, m \neq r}^s (a_{j i_m}^0 - a_{j i_m}). \tag{24}$$

另一方面,考虑从其他拥堵路段  $i_m$  重新分配到  $i_r$  路段的比例.从  $i_m$  路段被退回并重新分配到  $i_r$  的期望车辆比例为

$$a_3 = (a_{j i_m}^0 - a_{j i_m}) \frac{a_{j i_r}^0}{1 - a_{j i_m}^0}. \tag{25}$$

$i_r$  路段上的车辆保留率为

$$D_{\text{留}}^{i_r} = a_{j i_r} / a_{j i_r}^0. \tag{26}$$

对下标  $m$  求和,从其他拥堵路段真正分配到  $i_r$  路段的车辆比例为

$$a_3^* = \sum_{m=1, m \neq r}^s (a_{j i_m}^0 - a_{j i_m}) \frac{a_{j i_r}^0}{1 - a_{j i_m}^0} \frac{a_{j i_r}}{a_{j i_r}^0} = a_{j i_r} \sum_{m=1, m \neq r}^s \frac{a_{j i_m}^0 - a_{j i_m}}{1 - a_{j i_m}^0}. \tag{27}$$

综上所述,便得到多下游路段拥堵时,从  $j$  向拥堵路段  $i_r$  的转移比例系数  $a_{j i_r}^*$ , 它是期望转移比例系数、被退回  $i_r$  的再分配车辆比例和其他拥堵路段再分配给  $i_r$  的车辆比例三部分之和,即

$$a_{j i_r}^* = a_{j i_r} + a_2^* + a_3^*, r = 1, 2, \dots, s. \tag{28}$$

非拥堵下游路段和上游路段的转移比例,在车辆重新分配过程中不会涉及到退回的情形.从上游路口  $j$  到下游路口  $k_u$  的转移比例系数可表示为

$$a_{j k_u}^* = a_{j k_u}^0 + \sum_{r=1}^s (a_{j i_r}^0 - a_{j i_r}) \frac{a_{j k_u}^0}{1 - a_{j i_r}^0}, \tag{29}$$

$$u = 1, 2, \dots, t.$$

滞留在上游路口  $j$  的最终转移比例系数为

$$a_{j j}^* = a_{j j}^0 + \sum_{r=1}^s (a_{j i_r}^0 - a_{j i_r}) \frac{a_{j j}^0}{1 - a_{j i_r}^0}. \tag{30}$$

至此,上下游各路口的最终转弯比例均得到确定.

### 5 算例分析

构造交通网络如图 1 所示. 在预测之前, 需要确定模型实验时各参数值. 跳转 ARIMA 模型的参数分为两部分: 跳转模型参数和 ARIMA 模型参数. 对于跳转模型参数, 需要确定  $\mu_i$  和  $\sigma_i$ , 根据对交通流序列的观察, 设定的参数值如表 1 所示.

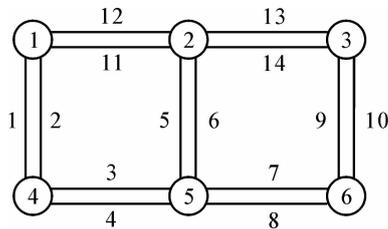


图 1 实验路网结构

表 1 跳转模型的参数

	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$\mu_i$	8	6	34	34
$\sigma_i$	4	4	4	4

实验中选取 ARIMA(4,0,0), 即 ARIMA 模型中不包括差分操作和滑动平均项. 对于这种 ARIMA 模型参数的学习, 可直接利用线性回归法求得参数. 路段拥堵容限向量取  $C = 4000 * \text{ones}(1,14)$ , 最大容限向量取  $C^* = 6000 * \text{ones}(1,14)$ , 单位均为辆/h. 利用模型进行预测, 选取路段 1 进行预测结果观察.

路段 1 交通流采集数据的间隔为 15 min, 因此一天有 96 个数据. 实验中选取两天 192 个数据进行预测. 路段 1 上的交通流数据取自北京市交通管理 UTC/SCOOT 系统, 选取前三门大街作为实验点.

根据上面设定的参数, 按基于跳转 ARIMA 模型对测试数据进行预测, 预测结果如图 2 所示.

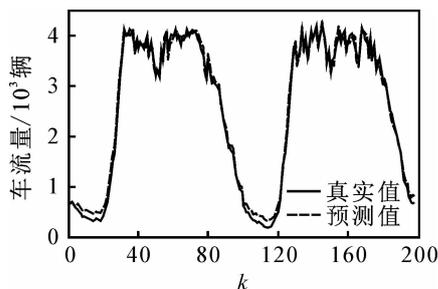


图 2 交通流的真实值和预测值

模型预测精确水平的衡量标准通常采用一些预测精度指标. 常用的指标有预测误差绝对值均值 (MAE), 预测误差均方根 (RMSE) 和预测误差百分比绝对值均值 (MAPE). 各项指标分别为

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t| \quad (31)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \quad (32)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right| \times 100\% \quad (33)$$

其中:  $\hat{y}_t$  和  $y_t$  是  $t$  时刻的预测交通流和真实交通流,  $T$  是预测周期. 实验中采用 RMSE 指标来衡量预测精度.

为评价基于跳转 ARIMA 模型的预测精度, 实验给出了另外 4 种模型的预测结果并作比较. 这 4 种模型分别是: 随机游走、历史平均、带信息的历史平均和  $K$  近邻. 比较结果如表 2 所示.

表 2 跳转 ARIMA 模型与 4 种模型预测结果比较

模 型	RMSE	模 型	RMSE
随机游走	123.46	$K$ 邻近	119.65
历史平均	152.55	跳转 ARIMA 模型	104.57
带信息的历史平均	110.95		

### 6 结 论

考虑到交通流序列在变化过程中可划分为上升态、下降态、高峰平台期和低谷平台期 4 个模态, 本文引入基于跳转的 ARIMA 模型来实现路段交通流预测. 基于路口转弯比例的时间稳定性和数值稳定性的假设, 考虑路网畅通和路网拥堵对路口转弯比例矩阵的影响, 分别对这两种情况进行研究, 得出了畅通和拥堵情况下的路口转弯比例矩阵, 并结合基于跳转 ARIMA 模型的路段交通流预测, 实现对路网交通流的预测.

这种预测方法的优点在于可实现对路网交通流的预测, 能方便地得到城市路网的交通流状态. 由于建模分析是从宏观角度考虑, 在处理一些宏观的交通问题时有其优势. 预测短时交通流是借助于路口车辆转弯比例矩阵的时间稳定性和数值稳定性, 由此得到的结果形式比较简单, 且容易操作.

#### 参考文献 (References)

[1] 陆化普, 李瑞敏, 朱茵. 智能交通系统概论[M]. 北京: 中国铁道出版社, 2004.  
(Lu H P, Li R M, Zhu Y. Conspectus of intelligent transportation system [M]. Beijing: China Railway Press, 2004.)

[2] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: A theoretical basis and empirical results [J]. J of Transportation Engineering, 2003, 129(6): 664-672.