

# 基于灰色 Verhulst 和 HMM 的交通事故预测算法

郝 兵

(河套学院理学系, 内蒙古 巴彦淖尔 015000)

**摘要:** 道路交通事故预测是保证道路佳通安全的重要技术, 以往的道路交通事故预测方法往往具有预测精度不高和收敛速度慢的缺点, 为此, 设计了一种基于灰色 Verhulst 模型和隐形马尔科夫链的交通事故预测方法; 首先, 采用灰色 Verhulst 模型对观察事件在下一时刻的状态进行预测, 采用最小二乘估计法去估计模型中的参数, 将预测结果用于初始化 HMM 模型, 并采用前向算法和后向算法对 HMM 模型进行训练, 获得最终的初始分布矩阵、状态转移概率和观察概率分布矩阵, 然后采用最终的 HMM 模型进行交通事故预测; 仿真试验结果表明: 文中方法能有效地实现交通事故预测, 较其它方法相比, 具有预测精度高和收敛速度快的优点, 具有一定的优越性。

**关键词:** 交通事故预测; 隐形马尔科夫链; 灰色模型; 参数估计

## Traffic Accident Prediction Algorithm Based on Grey Verhulst and HMM

Hao Bing

(Science department, Hetao College, Bayannaoer 015000, China)

**Abstract:** Road traffic accident prediction is an important technology to guarantee the safety of road, the given road traffic accident prediction method has the low prediction accuracy and low convergence speed, therefore, a prediction method based on grey Verhulst model and HMM is proposed. Firstly, the grey Verhulst model is used to predict the state of the next time, the least square method is used estimate the parameters of the model, the predicting result is used to initialize the HMM model, and the forward and backward algorithm are designed to train the HMM model, and the final initial distribution matrix, state transferring probability and observing probability distribution matrix, and the final HMM model can be used to predict the traffic accident. The simulation result shows the method in this paper can effectively realize the traffic accident prediction, and compared with the other methods, it has the high predicting accuracy and quick convergence speed. Therefore, it has big priority.

**Keywords:** traffic accident prediction; hidden markov chain; grey model; parameter estimation

## 0 引言

随着经济生活水平的提高和汽车保有量的增加, 道路交通安全问题也日趋增加, 严重威胁了人民的生命和财产安全, 成为一个严重的社会问题<sup>[1]</sup>。

目前已有的交通事故预测方法主要包含: 基于灰色模型的方法<sup>[2]</sup>, 基于回归模型的方法<sup>[3]</sup>、基于 HMM 的方法<sup>[4]</sup>和基于灰色马尔科夫预测模型的方法<sup>[5-6]</sup>。

文献 [2] 设计了一个基于自回归综合移动模型 ARIMA 的交通事故预测模型, 通过差分等数据转换方法将非平稳序列转换为均值为 0 的平稳随机序列。文献 [3] 设计了一种基于残差灰色估计算法的交通事故预测方法。文献 [4] 设计了一组基于隐形马尔科夫链 HMM 的交通事故预测模型, 主要针对对交叉口的交通冲突进行分类实现事故预测。文献 [5] 提出了一种用于交通事故预测的灰色加权马尔科夫 SCGM (1, 1) 模型, 根据历史数据进行精确预测的优势进行交通事故预测。文献 [6] 在传统灰色模型和马尔科夫链的基础上, 建立了等维新信息无偏灰色马尔科夫预测模型。

上述工作均研究交通事故的预测, 但是仍存在一些问题, 如回归模型需要大量历史数据, 而基于马尔科夫链和灰色模型的方法需要考虑状态转移概率, 忽略了历史数据与当前时段数

据的关系强弱, 从而影响了预测的精度。为此, 文中提出了一种基于灰色 Verhulst 模型和 HMM 的组合预测模型。

## 1 交通车辆运行特征提取

由于在车辆运动的过程中, 具有运行冲突的车辆容易发生事故<sup>[7]</sup>, 因此, 将车辆运行过程中的冲突作为观察事件, 以交通道路和十字路口的摄像头拍摄的视频和图片作为特征提取的来源, 对其进行灰度处理, 以减弱车辆颜色对图像强度的影响, 具体的提取过程可以描述为:

- 1) 将具有冲突的两辆车的运动矢量相减, 得到运动矢量差;
- 2) 将运动矢量差定格到以地理坐标和 90° 角为划分界限的 4 个象限区域中;
- 3) 获得两车最近的快的距离来表示两车距离;
- 4) 两车的相距远近可以通过运动矢量与两车距离的比值来表示;
- 5) 结合车辆之间的位置关系和相距远近可以构成唯一的事故预测的观察特征向量。

## 2 灰色 Verhulst 模型交通事故初预测

### 2.1 灰色 Verhulst 模型

Verhulst 模型是由德国生物学家 Verhulst 在 1837 年提出的一种生物生长模型, 可以用于描述具有饱和状态的过程, 能克服灰色模型仅适合短时平稳预测的缺点, 目前已经在人口预测、繁殖预测、产品寿命预测等方面获得广泛的关注。

收稿日期: 2014-06-05; 修回日期: 2014-07-09;

作者简介: 郝 兵 (1976-), 男, 内蒙古巴彦淖尔人, 硕士, 讲师, 主要从事计算机软件方向的研究。

### 2.2 基于灰色 Verhulst 模型的交通事故预测

设在交通事故中采集的原始数据序列为：

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\} \quad (1)$$

其中,  $n$  为序列长度, 对其进行累减处理后, 得到的生成序列如下所示:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\} \quad (2)$$

其中, 任意元素  $X^{(1)}(k)$  可以通过下式获得:

$$x^{(1)}(k) = x^{(0)}(k) - x^{(0)}(k-1) \quad (3)$$

根据下式求取原始序列的近邻均值生成序列:

$$Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)\} \quad (4)$$

其中,  $z^{(1)}(k)$  可以通过下式获得:

$$z^{(1)}(k) = (z^{(0)}(k) + z^{(0)}(k-1))/2 \quad (5)$$

由此, 可以得到灰色 Verhulst 模型如下所示:

$$x^{(0)}(k) + az^{(1)}(k) = b(z^{(1)}(k))^2 \quad (6)$$

其中,  $a$  和  $b$  为参数, 则灰色 Verhulst 模型的白化方程可以表示为:

$$\frac{dx^{(0)}}{dt} + ax^{(0)} = b(x^{(0)})^2 \quad (7)$$

设参数向量为  $l = (a, b)^T$ , 并设定矩阵  $A$  和向量  $Y$  如下所示:

$$A = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & (z^{(1)}(n))^2 \end{bmatrix} \quad (8)$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix} \quad (9)$$

采用最小二乘法求解, 可以获得参数向量, 通过下式计算:

$$l = (A^T A)^{-1} A^T Y \quad (10)$$

则灰色 Verhulst 模型对应的白化方程解为:

$$x_i^{(1)} = \frac{ax^{(1)}(0)}{bx^{(1)}(0) + (a - bx^{(1)}(0))e^{at}} \quad (11)$$

其最终的对下一个时刻的交通事故预测值可以表示为:

$$x_{k+1}^{(1)} = \frac{ax^{(1)}(0)}{bx^{(1)}(0) + (a - bx^{(1)}(0))e^{at}} \quad (12)$$

由此, 通过上述过程可以判断出在某时刻的状态在下一状态的预测值, 从而可以判断出事故是否发生。

### 3 基于 HMM 的交通事故分类

在采用了灰色 Verhulst 模型获得了各观察事件的下一状态, 可以获得各状态的初始概率分布、转移分布, 采用其初始化 HMM 并对其进行训练, 从而获得最终的 HMM 模型实现交通事件的分类和预测。

#### 3.1 HMM 模型概述

隐型马尔科夫模型 (Hidden Markov model, HMM) 是一种能通过少量样本数据训练出可靠模型, 其主要包含马尔科夫链和一般随机过程两个随机过程, 其状态时不可观察, 只能通过观察事件和模式匹配原始, 将与观察信号最相似的状态作为匹配结果。

HMM 可以描述为五元组:  $\lambda = (N, M, \pi, A, B)$  :

1)  $\{\theta_1, \theta_2, \dots, \theta_N\}$  为马尔科夫链的状态序列,  $N$  为马尔科夫链中的状态总数;

2)  $\{v_1, v_2, \dots, v_M\}$  为观测序列集合,  $M$  为观测值数;

3) 初始概率矩阵为  $\pi$  ;

4) 状态转移概率  $A$ , 每个元素  $a_{ij} \in A$  表示状态  $\theta_i$  转移到  $\theta_j$  的概率;

5)  $B$  为观察值概率矩阵, 每个元素  $b_{ij} \in B$  表示  $t$  时刻时, 状态为  $\theta_i$  时观测值为  $v_k$  的概率。

#### 3.2 HMM 模型的训练和分类

初始化: 采用由灰色 Verhulst 模型得到的初始分布和转移概率初始化 HMM 模型  $\lambda_0$ , HMM 模型训练阈值  $th_0$ , 当前迭代次数为  $k = 1$  ;

步骤 1: 根据灰色 Verhulst 模型得到的  $P(\lambda_k | o_t)$  来初始化  $P(\text{HMM}_k | o_t)$ , 根据贝叶斯定理, 根据下式计算观察值转移概率矩阵  $B = [b_{ij}]_{N \times M}$  :

$$P(o_j = v_j | s_j = \theta_j) = \frac{P(s_j = \theta_j | o_j = v_j)P(o_j = v_j)}{\pi_j} \quad (13)$$

其中,  $P(s_j = \theta_j | o_j = v_j)$  的值为  $P(\text{HMM}_k | o_t)$ ,  $P(o_j = v_j)$  为观察周期的倒数。

步骤 2: 采用前向评估算法计算在时刻  $t$  时, 在以往的观察序列为  $o_1, o_2, \dots, o_t$  时, 状态值为  $\theta_j$  的概率, 可以通过下列公式获得:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, s_t = \theta_i | \lambda_t) \quad (14)$$

其可以通过下列过程估计:

$$\alpha_1(i) = P(o_1, s_1 = \theta_i | \lambda) \quad (15)$$

$$\alpha_{t+1}(i) = \left( \sum_{j=1}^N \alpha_t(j) a_{ij} \right) b_j(o_{t+1}) \quad (16)$$

则根据前向观点, 预测值可以估计为:

$$P(Y | \lambda) = \sum_{i=1}^N \alpha_T(i) = \sum_{j=1}^N \left\{ \left[ \sum_{i=1}^N \alpha_{T-1}(i) a_{ij} \right] b_j(o_T) \right\} \quad (17)$$

步骤 3: 设  $\beta_t(i)$  表示模型  $\lambda$  在  $t$  时刻, 当观察序列为  $o_{t+1}, o_{t+2}, \dots, o_T$  时, 处于状态为  $\theta_i$  的概率, 则结合前向评估算法和后向评估算法可以得到:

$$P(O | \lambda) = \sum_{i=1}^N \alpha_t(i) \beta_t(i) \quad (18)$$

步骤 4: 以最大化后验概率为目标, 目标函数可以表示为:

$$J = -\log P(\lambda_1 | O, \lambda) = -\log \frac{P(\lambda_1, O | \lambda)}{P(O | \lambda)} \quad (19)$$

将式 (19) 分别对  $\pi_i$  和  $a_{ij}$  求导, 可以得到新的模型为:

$$\begin{cases} \pi'_i = r_1(i) = \sum_{j=1}^N s(i, j) \\ a'_{ij} = \sum_{i=1}^{T-1} s_i(i, j) / \sum_{i=1}^{T-1} r_i(i) \\ b'_k = \sum_{i=1}^{T-1} r_i(j) d_i(j, k) / \sum_{i=1}^{T-1} r_i(i) \\ \lambda' = (\pi', a'_{ij}, b'_k) \end{cases} \quad (20)$$

其中,  $s(i, j)$  表示模型在时刻  $t$  位于状态  $i$ , 而在时刻  $t+1$  位于状态  $j$  的概率, 可以通过下式求取:

$$\begin{aligned}
 s(i, j) &= P(s_t = \theta_i, s_{t+1} = \theta_j | O, \lambda) \\
 &= \alpha_i(i) a_{ij} b_j(o_{t+1}) \beta_j(i) / P(O | \lambda) \\
 &= \frac{\alpha_i(i) a_{ij} b_j(o_{t+1}) \beta_{j+1}(i)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(i) a_{ij} b_j(o_{t+1}) \beta_{j+1}(i)} \quad (21)
 \end{aligned}$$

$r_1(i)$  表示状态  $i$  转移到  $j$  的期望次数,  $r_t(j) d_t(i, j)$  表示从状态输出为  $j$  的期望次数。

步骤 5: 判断模型是否已经收敛:

$$| \lg(P(O | \lambda')) - \lg(P(O | \lambda)) | \leq th \quad (22)$$

如果式 (22) 满足, 则模型已经收敛, 可以将新的观察数据直接输入监测模型, 得到分类结果作为最终的预测结果;

否则  $\lambda = \lambda'$ , 并输入下一个观察样本数据并返回步骤 2 继续进行训练。

### 4 实验分析

为了验证文中方法的优越性, 以 1985-2012 年某地级市交通事故统计数据作为实验数据, 将 1985-2012 年的事故数据划分为训练样本进行分析, 如表 1 所示。

表 1 历史交通事故数据年份事故

年份	事故次数	年份	事故次数	年份	事故次数
1985	320	1994	1409	2004	4533
1986	412	1995	1534	2005	5002
1987	496	1996	1658	2006	5323
1987	585	1997	1789	2007	5832
1988	634	1998	1923	2008	6333
1989	742	1999	2045	2009	6742
1990	879	2000	2214	2010	7142
1991	935	2001	3009		
1992	1134	2002	3592		
1993	1352	2003	4034		

为了对文中方法的优越性进行验证, 将文中方法用于交通事故预测, 并与文献 [7] 和文献 [8] 进行对比, 文中参数设置为: HMM 模型训练阈值  $th = 0.1$ , 迭代次数最大值  $T = 100$ , 3 种方法得到的事故预测次数对比如图 1 所示。

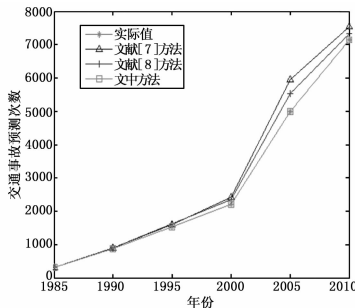


图 1 预测值与实际对比

从图 1 中可以看出, 文中方法与实际值的预测曲线基本完

全拟合, 预测的事故次数较为精确, 而文献 [7] 方法的预测平均误差为 9.57%, 而文献 [8] 方法的预测平均误差为 5.73%。

3 种方法得到的预测误差收敛曲线如图 2 所示。

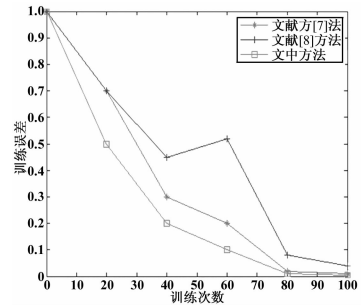


图 2 训练误差对比

从图 2 所示的训练误差对比图中可以看出, 文献 [7] 和文献 [8] 方法的训练误差在迭代到 100 次时分别为 0.1 和 0.04, 但是始终未能收敛, 文中方法得到的训练误差最终为 0.05, 且在迭代到 80 次时就趋于收敛, 显然, 文中方法具有预测精度高的优点。

### 5 结论

道路交通事故预测是交通系统安全管理和防范的重要手段, 为此, 文中设计了一种基于灰色 Verhulst 模型和隐形马尔科夫链 HMM 的道路交通事故预测模型, 首先采用灰色 Verhulst 模型对任意状态的下一时刻状态进行预测, 采用最小二乘法进行系数估计, 然后将灰色 Verhulst 模型的预测结果用于初始化 HMM 的模型参数, 采用前向和后向的算法进行评估, 获得最终的 HMM 模型, 最后, 采用仿真试验对预测次数和训练误差进行分析, 实验证明了文中方法的有效性。

#### 参考文献:

- [1] 薛大维. 道路交通事故勘察处理 [M]. 北京: 机械工业出版社, 2011.
- [2] Yang Z Q. Highway traffic accident predication based on SVR trained by genetic algorithm [J]. Advanced materials research, 2012: 433 (440): 5886-5889.
- [3] 钟连德, 孙小端. 高速公路事故预测模型 [J]. 北京工业大学学报, 2009 (7): 55-59.
- [4] 薛大维, 宋成举. 道路交通事故次数组合预测模型 [J]. 交通科技与经济, 2013, 1 (15): 19-21.
- [5] 张 杰, 刘小明, 贺玉龙, 陈永胜. ARIMA 模型在交通事故预测中的应用 [J]. 北京工业大学学报, 2007, 2 (33): 1295-1299.
- [6] 宁 彬, 谷 琼. 基于残差灰色估计交通事故数量挖掘算法仿真 [J]. 计算机仿真, 2013, 6 (30): 178-181.
- [7] 周 君, 程 琳, 周凌云, 等. 基于 HMM 的交叉口交通事件预测研究 [J]. 交通运输系统工程与信息, 2003, 6 (13): 52-29.
- [8] 赵 玲, 许宏科. 基于灰色加权马尔可夫 SCGM (1, 1) 的交通事故预测 [J]. 计算机工程与应用, 2012, 48 (31): 11-14.
- [9] 赵 玲, 许宏科. 基于新维无偏灰色马尔可夫的交通预测 [J]. 计算机工程与应用, 2013 (7): 35-38.
- [10] 李宏魁, 谢正光, 王 轲, 王天庆. 一种新的视频停车事件检测算法 [J]. 计算机测量与控制, 2013, 21 (11): 3040-3042.