

# 基于朴素贝叶斯分类的高速公路交通事件检测

张 轮, 杨文臣, 刘 拓, 施奕骋

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

**摘要:** 提出一种基于朴素贝叶斯分类的高速公路非重现交通事件检测算法. 将交通事件的检测看作是 0-1 分类问题, 采用交通波动理论建立交通事件的特征属性概念模型, 并利用分段离散化的方法将连续特征变量转换为离散特征变量, 设计基于朴素贝叶斯算法的交通事件分类器. 以典型高速公路的一条路段进行 VISSIM 仿真试验. 结果表明: 该算法的检测率高, 且在高强度状况下, 算法鲁棒性良好, 适用于高速公路交通事件检测系统.

**关键词:** 交通事件; 朴素贝叶斯; 特征离散; 模式识别; 高速公路运营

**中图分类号:** U492.8<sup>+</sup>5; TP391.9

**文献标志码:** A

## A Naive Bayesian Classifier-based Algorithm for Freeway Traffic Incident Detection

ZHANG Lun, YANG Wenchen, LIU Tuo, SHI Yicheng

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract:** This paper presents a naive Bayesian classifier-based algorithm for freeway non-recurrent traffic incident detection to enhance the accuracy and learning ability of intelligent traffic incident detection algorithm. The traffic wave theory is employed to establish a conceptual characteristic model of traffic incident, continuous characteristic variables are transferred into discrete characteristic variables via sub-discretization, and the naive bayesian-based traffic incident classifier is designed by regarding traffic incident detection as “0-1” classification problems. An experiment is carried on a section of a typical freeway, and the performance of the presented model and algorithm is validated via VISSIM simulation. Extensive simulation results show that the algorithm in freeway traffic incident detection system is of high accuracy and strong robustness even if the traffic volumes increase.

**Key words:** traffic incident; naive Bayesian; characteristic

discretization; pattern recognition; freeway operation

非重现高速公路交通事件是指高速公路上影响交通安全和畅通的偶发性事件, 如车辆交通事故、故障停车、货物散落等(不包含偶发性交通拥堵). 高速公路非重现交通事件的自动检测技术可使事件在最短时间内获得清除, 以防止事件后的灾难扩大, 是先进智能交通运输系统的关键技术.

高速公路交通流具有较强的不确定性和非线性, 难以提取标准的模式特征和建立精确的数学模型, 严重制约了基于模式识别及统计模型的传统检测算法的实施. 在与交通环境的交互过程中, 基于智能计算的交通事件检测算法挖掘交通事件知识, 无需精确数学模型且具有良好的非线性逼近能力, 是近年来研究的热点. 在众多智能交通事件检测算法中, 神经网络<sup>[1]</sup>、模糊聚类<sup>[2]</sup>、非参数回归<sup>[3]</sup>和支持向量机<sup>[4]</sup>占主导地位, 仿真结果表明交通事件智能检测方法能快速准确地判断交通事件. 但基于神经网络的检测缺乏有效样本集, 且权值及阈值参数的合理设置较难; 基于模糊聚类的检测采用专家经验知识确定模糊隶属参数, 且不具备学习功能; 依托大量的历史数据, 基于非参数回归的检测可滚动学习和更新事件知识, 但对搜索算法的效率提出挑战; 基于支持向量机的检测具有好的事件知识泛化能力, 结构直观, 但仍属于监督的学习算法.

贝叶斯分类算法具有智能统计和学习的能力, 20 世纪 80 年代, 贝叶斯方法首先被用于高速公路时间占有率等事件参数概率分布的预测<sup>[5]</sup>, 随后用于交通事件持续时间的预测<sup>[6]</sup>, 神经网络交通事件检测模型参数的学习更新<sup>[7]</sup>, 城市道路主干道交通事件检测<sup>[8]</sup>, 以及基于多源数据融合的城市道路交通状态估计<sup>[9]</sup>. 研究结果表明, 交通事件的贝叶斯检测

具有好的泛化学习能力,检测率很高,误报率低,但平均检测时间长.为解决贝叶斯算法实时响应特性不足,基于特征属性相互独立假设的朴素贝叶斯算法被提出,并已在数据挖掘、工业设计、生物医学等领域取得成熟的应用<sup>[10]</sup>,其对分类问题的有效性得到大量的实践检验.国内外已将朴素贝叶斯应用于交通出行活动识别<sup>[11]</sup>、交通状态数据挖掘<sup>[12]</sup>和交通事件持续时间预测<sup>[13]</sup>,但未见用于交通事件检测判别方面.

本文将交通事件的检测看作是一个 0-1 分类问题,根据交通波动理论合理选取交通事件检测的特征属性,采用区间离散化的方法将连续特征变量转换为离散特征变量.同时,引入朴素贝叶斯理论,结合交通事件检测的特征参数分布特性,设计基于朴素贝叶斯算法的交通事件分类器.对一条具有二级服务水平的高速公路路段进行试验,采用 Matlab 开发整套基于朴素贝叶斯的交通事件检测模型及算法,并用 VISSIM 仿真模拟不同交通状态下的事件场景,对本文模型及算法的效用进行评价.

## 1 朴素贝叶斯算法

朴素贝叶斯分类旨在使用概率表示各种事件的不确定性,是贝叶斯方法中实用性很高的一类智能学习算法.朴素贝叶斯分类的基本思想如下<sup>[14]</sup>:

(1) 每个数据样本用一个  $n$  维特征向量  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  表示,分别描述该样本  $n$  个特征属性  $A_1, A_2, \dots, A_n$  的度量.

(2) 假定有  $m$  个类  $C_1, C_2, \dots, C_m$ ,给定一个分类标号未知的数据样本  $\mathbf{X}$ ,分类器将预测  $\mathbf{X}$  属于具有最高后验概率的类(在条件  $\mathbf{X}$  下),即朴素贝叶斯分类将未知的样本  $\mathbf{X}$  分配给  $C_i (1 \leq i \leq m)$ ,当且仅当  $P(C_i|\mathbf{X}) > P(C_j|\mathbf{X})$ ,对任意的  $j=1, 2, \dots, m, j \neq i$ .其中,  $P(C_i|\mathbf{X})$  最大的类称为最大后验假定.根据贝叶斯定理,最大后验假定按下式计算:

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})} \quad (1)$$

(3) 对所有的分类,样本  $\mathbf{X}$  的概率  $P(\mathbf{X})$  可视为常数,因此,只需  $P(\mathbf{X}|C_i)P(C_i)$  最大,则  $P(C_i|\mathbf{X})$  最大.对  $\forall i$ ,如果  $C_i$  类的先验概率未知,则通常假定这些类是等概率的,即  $P(C_1) = P(C_2) = \dots = P(C_m)$ ,因此,最大后验假定的问题就转化为对  $P(\mathbf{X}|C_i)$  的最大化.  $P(\mathbf{X}|C_i)$  通常被称为给定  $C_i$  时样本  $\mathbf{X}$  的似然度,而使  $P(\mathbf{X}|C_i)$  最大的假设成为最大似然假设.否则,需要最大化  $P(\mathbf{X}|C_i)P(C_i)$ .此

时,类  $C_i$  的先验概率按下式计算:

$$P(C_i) = \frac{s_i}{s} \quad (2)$$

式中:  $s_i$  是属于类  $C_i$  的训练样本数,  $s$  是训练样本总数.

(4) 对给定的具有许多属性的数据集,计算  $P(\mathbf{X}|C_i)$  的时间开销会非常大.为降低计算  $P(\mathbf{X}|C_i)$  的时间复杂度,引入类条件独立的朴素假定,即对给定样本的类标号,假定属性值之间相互条件独立(在属性之间不存在依赖关系),则  $\mathbf{X}$  的似然度  $P(\mathbf{X}|C_i)$  按下式计算:

$$P(\mathbf{X}|C_i) = \prod_{k=1}^n P(x_k|C_i) \quad (3)$$

式中:概率  $P(x_1|C_i), P(x_2|C_i), \dots, P(x_n|C_i)$  可以由训练样本估计.

如果  $A_k$  是离散属性,则在给定类  $C_i$  下,属性取值  $x_k$  的概率按下式计算:

$$P(x_k|C_i) = \frac{s_{ik}}{s_i} \quad (4)$$

式中:  $s_{ik}$  是在属性  $A_k$  位上取值  $x_k$  且属于类  $C_i$  的训练样本数.

如果  $A_k$  是连续属性,则通常假定该属性服从高斯分布.在给定类  $C_i$  下,属性取值  $x_k$  的概率按下式计算:

$$P(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}^2}} \quad (5)$$

式中:  $g$  为正态分布概率密度函数,  $\mu$  为样本期望,  $\sigma$  为样本方差.

(5) 判定未知样本  $\mathbf{X}$  的分类.对每个类  $C_i$ ,分别计算  $P(\mathbf{X}|C_i)P(C_i)$ .样本  $\mathbf{X}$  被指派到类  $C_i$ ,当且仅当  $P(C_i|\mathbf{X}) > P(C_j|\mathbf{X})$ ,对任意的  $j=1, 2, \dots, m, j \neq i$ .

## 2 交通事件检测的朴素贝叶斯算法

### 2.1 算法构思

朴素贝叶斯算法有两个特征:一是可有效利用不完整的事件信息做出预测;二是可利用在不同时间段获得的交通事件信息滚动更新预测结果.这两个特征正好符合交通事件数据获取的内在特性.基于朴素贝叶斯分类的交通事件检测的理论依据是:由于在连续的交通流中,时间占有率等交通特征参数连续变化且具有特定的模式及相似性,可将非重现交通事件看作是连续交通流中随机扰动的“噪声”.道路中某个断面一旦发生交通事件,则受此交

通事件的影响,该断面及其邻近断面的交通参数将因此而发生改变,且该变化的模式将不同于平稳的交通流.因而,可将交通事件的检测看作“有或无”的二值分类问题,以邻近点交通参数的实时数据作为输入,采用朴素贝叶斯分类器识别其模式特征,进而判定交通事件的发生.

## 2.2 交通事件特征属性向量

表征交通流特性的状态参数主要有流量、密度和速度,其在道路交通信息采集系统中分别对应交通流量、时间占有率和行车速度三种数据. Kerner<sup>[15]</sup>的研究结果表明:以流量、密度和速度表征的高速公路交通状态分布在二维区域内,并不符合交通流中的基本关系曲线图. 为保证特征属性之间的相互独

立性,最多取三个状态参数中的两个参数. 由于依托单断面线圈难以采集速度信息,而且时间占有率数据能够最直接地反映道路交通拥堵程度,研究选取时间占有率( $O$ )和交通流量( $Q$ )作为交通事件检测的状态参数.

根据交通流的波动理论,当交通事件发生时,上游将会产生拥挤波,而下游将会产生疏散波,且波动将沿着空间和时间两个维度向外扩散,致使上、下游交通参数的模式特征发生改变. 如图 1 所示,假设在  $T$  时刻,  $D_g$  断面发生了交通事件,致使该断面的交通参数发生变化,则在时刻  $[T+\Delta T, T+2\Delta T, T+3\Delta T]$ ,此波动变化将扩散到断面  $[(D_{g-1}, D_{g+1}), (D_{g-2}, D_{g+2}), (D_{g-3}, D_{g+3})]$ .

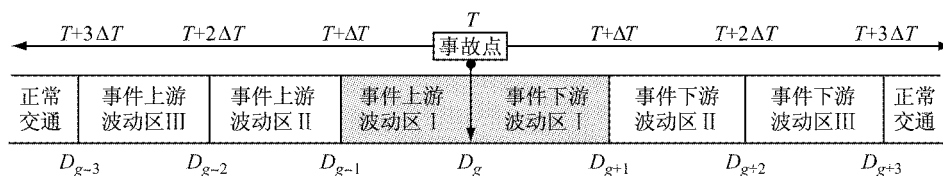


图 1 交通流的波动向上、下游的传递

Fig.1 Transmission of traffic wave to the upper and lower flow

为构建交通事件特征属性向量,在滞后时间段  $(\Delta T, \Delta T+\tau, \dots, \Delta T+b\tau)$  内,选取与事件断面相邻的上、下游检测点状态参数的检测数据作为特征属性. 其中,  $\tau$  是数据的采样时间间隔;  $b$  是滞后时间段带宽,取非负整数;  $\Delta T$  是交通事件波动从断面  $D_g$  传递到上、下游断面  $D_{g-1}$  或  $D_{g+1}$  的最大传递时间,按下式计算:

$$\Delta T = \frac{l}{V} \quad (6)$$

式中:  $l$  为事故断面到相邻检测断面间的距离;  $V$  为交通流的波动速度,若上、下游断面与事件发生断面的距离较近,事件检测的滞后时间段较长 ( $b > 1$ ) 且拥有一定的覆盖区间,则使用路段平均行驶速度  $v$  作为波动速度的估计值. 因此,交通事件的特征向量  $\mathbf{X} = (O_{\Delta T}^+, Q_{\Delta T}^+, O_{\Delta T+\tau}^+, Q_{\Delta T+\tau}^+, \dots, O_{\Delta T+b\tau}^+, Q_{\Delta T+b\tau}^+, O_{\Delta T}^-, Q_{\Delta T}^-, O_{\Delta T+\tau}^-, Q_{\Delta T+\tau}^-, \dots, O_{\Delta T+b\tau}^-, Q_{\Delta T+b\tau}^-)$  (7)

式中:  $O_{\Delta T}^+, Q_{\Delta T}^+$  分别表示在事件发生  $\Delta T$  后,上游检测断面的时间占有率和交通流量;  $O_{\Delta T}^-, Q_{\Delta T}^-$  则分别表示下游检测断面相应的时间占有率和交通流量.

## 2.3 交通事件特征属性的离散化

由于交通信息采集系统检测到的断面时间占有率和交通流量的交通数据是连续的数值,而且特征数据的单位和值域均不相同,不能将其直接应用于朴素贝叶斯算法. 同时,为降低交通事件检测算法的

计算复杂度,需对连续型的特征属性数据进行离散化预处理<sup>[16]</sup>,以将连续的特征数值转换为离散数值. 特征属性离散化的原理如下:假设交通参数的值域为  $[a_0, a_p]$ ,在  $a_0$  和  $a_p$  中插入  $p-1$  个数值将值域区间分为  $p$  段  $([a_0, a_1], [a_1, a_2], \dots, [a_{p-1}, a_p])$ ,则将落在第  $h$  个区间的连续数值  $x$  的离散取值标记为  $h$ ,即  $x=h$ ,若  $x \in [a_{h-1}, a_h]$ . 如此,可将值域为  $[a_0, a_p]$  的连续特征数据转换成值域为  $[1, p]$  的离散数据,如表 1 所示.

表 1 连续特征属性的离散取值

Tab.1 Discretization of continuous data of feature attributes

$x$ 的连续值区间	$x$ 的离散值
$[a_0, a_1]$	1
$[a_1, a_2]$	2
$\vdots$	$\vdots$
$[a_{p-1}, a_p]$	$p$

## 2.4 算法设计

用“0”和“1”分别表示交通事件  $Y$  的“不发生”和“发生”,即  $C_1=0, C_2=1$ . 记  $\mathbf{X}$  为交通事件的特征属性向量,  $C$  为其分类类别,训练数据集  $\{\mathbf{X}, C\}$  的样本规模为  $N$ ,采用朴素贝叶斯算法构造交通事件分类器的步聚如下:

**步骤 1** 学习训练数据的交通事件知识. 依托训练数据集,按式(4)计算每一个特征属性  $A_k$  取值  $h$

时属于给定类  $C_i$  的概率,即计算  $P(x_k=h|C=1)$  和  $P(x_k=h|C=0)$ ,并按式(2)计算交通事件的发生概率  $P(C=1)$ .具体计算过程如下:

定义指示函数

$$I(\phi) = \begin{cases} 1, & \phi \text{ 表达式为真} \\ 0, & \phi \text{ 表达式为假} \end{cases} \quad (8)$$

则根据式(4),交通事件发生或不发生的特征属性值的概率分别为

$$P(x_k = h | C = 1) = \frac{\sum_{i=1}^N I(x_k = h \cap C = 1)}{\sum_{i=1}^N I(C = 1)} \quad (9)$$

$$P(x_k = h | C = 0) = \frac{\sum_{i=1}^N I(x_k = h \cap C = 0)}{\sum_{i=1}^N I(C = 0)} \quad (10)$$

根据式(2),交通事件发生的概率

$$P(C = 1) = \frac{\sum_{i=1}^N I(C = 1)}{N} \quad (11)$$

$$P(C = 1 | \mathbf{X}) = \frac{\prod_{k=1}^n P(x_k | C = 1) P(C = 1)}{\prod_{k=1}^n P(x_k | C = 1) P(C = 1) + \prod_{k=1}^n P(x_k | C = 0) P(C = 0)} \quad (15)$$

**步骤3** 判断给定条件下交通事件是否发生.如果在交通特征属性  $\mathbf{X}$  下交通事件发生的后验概率大于某阈值  $\theta$ ,则判定此时有交通事件发生.交通事件判定的概念模型

$$Y = \begin{cases} 1, & P(C = 1 | \mathbf{X}) \geq \theta \\ 0, & P(C = 1 | \mathbf{X}) < \theta \end{cases} \quad (16)$$

**步骤4** 判断高速公路运营恶化等级  $L$ .交通事件引起的高速公路运行状况恶化可划分为两个等级,标记  $L_1$  为“行车中断”, $L_2$  为“行车缓慢”.在行车中断的状况下,事件邻近上游检测点的线圈长期处于被占用状态,其线圈时间占有率约为 100%.相反,下游检测点的时间占有率为零.同时,两个断面通过的车辆数近乎为零.为此,运营恶化等级判定的概念模型

$$L = \begin{cases} L_1, & O_{\Delta T+b\tau}^+ = 1 \text{ 和 } Q_{\Delta T+b\tau}^+ = 0 \\ L_2, & \text{其他} \end{cases} \quad (17)$$

由交通事件的朴素贝叶斯检测算法的推导过程,对含  $m$  个样本且每个样本含有  $n$  个特征属性的训练集,步骤1计算每一个特征属性值的条件概率

**步骤2** 计算给定条件下交通事件发生的后验概率.根据式(1),对新给定的测试数据  $\mathbf{X}=(x_1, x_2, \dots, x_n)$ ,在交通特征属性  $\mathbf{X}$  下交通事件发生的后验概率

$$P(C = 1 | \mathbf{X}) = \frac{P(\mathbf{X} | C = 1) P(C = 1)}{P(\mathbf{X})} \quad (12)$$

基于特征属性相互独立的假设,根据式(3),特征属性向量  $\mathbf{X}$  的概率

$$\begin{aligned} P(\mathbf{X}) &= \sum_{i=1}^m P(\mathbf{X} | C_i) P(C_i) = \\ &= \sum_{i=1}^m \prod_{k=1}^n P(x_k | C_i) P(C_i) = \\ &= \prod_{k=1}^n P(x_k | C = 1) P(C = 1) + \\ &= \prod_{k=1}^n P(x_k | C = 0) P(C = 0) \end{aligned} \quad (13)$$

同理,交通事件发生的似然度

$$P(\mathbf{X} | C = 1) = P(x_1, x_2, x_3, \dots, x_n | C = 1) = \prod_{k=1}^n P(x_k | C = 1) \quad (14)$$

将式(13)和(14)代入式(12),在交通特征属性  $\mathbf{X}$  下交通事件发生的后验概率

的时间复杂度为  $O(nm)$ ,而步骤2和步骤3的时间复杂度均为  $O(n)$ .因此,检测算法的时间复杂度为  $O(nm)$ .

## 3 应用举例

### 3.1 案例描述

(1) 典型高速公路路段的交通事件 VISSIM 仿真模型

将 VISSIM 作为无偏的评价工具.采用 VISSIM 建立的具有二级服务水平的典型高速公路路段的 VISSIM 仿真模型如图2所示.高速公路路段的长度为 4 km.根据国内高速公路最高时速  $120 \text{ km} \cdot \text{h}^{-1}$  及最低时速为  $60 \text{ km} \cdot \text{h}^{-1}$  [17],路段上车流的平均运行速度为  $25 \text{ m} \cdot \text{s}^{-1}$ .在该路段上,每隔 1 km 布设一组线圈检测器,且根据采样定理,以 20 s 为数据采样间隔采集交通参数数据,即  $\tau=20$ .假设最低限速容许的检测误差为 20%,当事件发生断面的平均速度低于  $13 \text{ m} \cdot \text{s}^{-1}$  时,则该检测断面发生交通事件,如

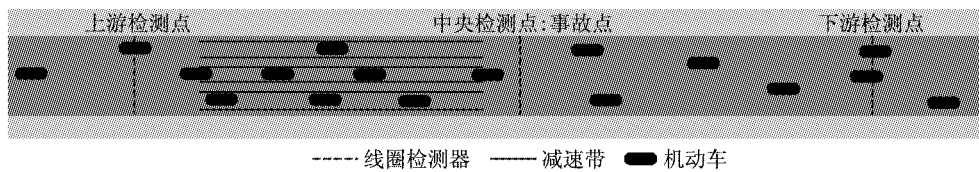


图 2 试验高速公路路段的交通事件 VISSIM 仿真模型

Fig.2 VISSIM-based traffic incident simulation model for the test freeway

图 2 所示,其在中央检测点(2 km)上游处,通过设置“减速带”的限制车速( $2\sim 13\text{ m}\cdot\text{s}^{-1}$ )模拟试验路段交通事件的随机发生.

(2) 仿真场景

为验证提出的交通事件检测算法在不同交通流量下及交通流波动情况下的检测效果,设计如下不同交通强度的试验场景:① 设置三种路段车辆发车率  $\lambda$  为 0.45,0.60 和 0.75,分别代表闲散、顺畅和繁忙三种交通状态,每种状态仿真时长设为 3 h;② 为模拟路段短时交通流到达的波动特性,在 VISSIM 的 Vehicle Inputs 的 Time intervals 文件中,设置 1 h 内 12 min 间隔车辆的发车百分比例依次为 20%—15%—25%—15%—25%;③ 根据二级服务水平高速公路直行车道的设计规范<sup>[17]</sup>,其通行能力为  $1\,600\text{ 辆}\cdot\text{h}^{-1}$ ,车头时距 MHT、驾驶员反应时间 MDT 的仿真参数校准为 1.8 s 和 1.5 s;④ 由于仿真效果受随机种子影响,在每个仿真小时内设置不同的随机种子.

3.2 试验设计

在每种交通状态下,采用 VISSIM 仿真各自生成 540 组原始数据,且为使训练数据保持均衡以学习事件知识,交通事件随机发生的概率为 0.25. 因此,通过 VISSIM 仿真试验共计生成 1 620 组原始的交通检测数据. 对每种交通状态,随机选取其中 70% 的数据作为训练数据集,余下的 30% 作为测试数据集.

(1) 交通事件检测的朴素贝叶斯算法的参数确定

根据式(6)计算,可得事件的最大传递时间  $\Delta T$  为 40 s,即事件断面引起的交通波动将在 40 s 后传递到上、下游检测点. 滞后时间段带宽可采用经典的专家法或试错法确定,但这类方法没有普适性. 研究采用相关系数法确定滞后带宽<sup>[18]</sup>,亦即通过计算历史数据中各交通参数与事件断面速度之间的相关系数,来确定其是否为输入层,相关系数  $r$  按下式计算:

$$r = \frac{\sum_{i=1}^M (z_i - \bar{z})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^M (z_i - \bar{z})^2 \sum_{i=1}^M (v_i - \bar{v})^2}} \quad (18)$$

式中: $z_i$  是事件发生后的上、下游的交通参数(时间占有率或交通流量), $\bar{z}$  是相应交通状态参数的均值; $v_i$  是事件断面的速度, $\bar{v}$  是事件断面速度的均值; $M$  是事件样本的数量.

在事件发生后,上、下游的交通流量和时间占有率与事件断面速度相关系数的计算结果如表 2 所示. 可以看出:自事件发生后的第三阶段,各交通参数与当前事件断面速度的相关系数已趋于稳定,这是因为在事件波动效应持续一段时间后,断面交通流自组织形成了较稳定的形态特征. 因而,选取滞后带宽  $b$  为 3.

表 2 交通状态参数与事件断面速度的相关系数(20 s 采样间隔)

Tab.2 Correlation coefficients of traffic status parameters and the speed at the traffic incident spot(20 s sampling interval)

$z$	$r$					
	$b=6$	$b=5$	$b=4$	$b=3$	$b=2$	$b=1$
交通流量	-0.538 5	-0.549 6	-0.548 7	-0.551 5	-0.579 1	-0.593 2
时间占有率	-0.565 8	-0.570 4	-0.577 1	-0.575 8	-0.615 8	-0.635 3

根据式(7),研究选取交通事件断面的上、下游检测点在滞后时间段(40 s,60 s,80 s)内的断面时间占有率数据和交通流量数据,共计 12 个特征值组成交通事件特征属性向量. 设定交通事件后验概率的阈值  $\theta$  为 0.60,即事件的后验概率大于 0.60 时判定此时发生交通事件.

(2) 试验方案

计算机数值的检测算法采用 Matlab 2008Ra 编程. 交通事件检测存在四种可能结果:无事件正确检测、有事件正确检测、无事件错误报警和有事件漏检. 为验证基于朴素贝叶斯的交通事件检测算法(NBTIA)的效用,选取检测率(detection rate, DR)、

误报率(false alarm rate, FAR)、漏检率(miss detection rate, MDR)和平均检测时间(average detection time, ADT)为性能指标,并与基于BP神经网络的交通事件检测算法(BPTIA)<sup>[19]</sup>对比,对试验案例进行检测。

### 3.3 试验结果

三种状态下的交通事件检测结果如表3所示。与BPTIA算法相比,文中提出的NBTIA算法检测

率高、实时性更好,且随着交通状态升高,误报率、漏检率和平均检测时间的性能稳定,算法鲁棒性良好。数据分析结果表明:事件条件下的交通参数特征与路段处于较高饱和状态下的交通参数特征之间的差异不显著,BPTIA检测算法的性能受到有效样本的质量影响;而NBTIA算法不仅适用于特征属性信息不完整的交通事件,所构建的事件贝叶斯分类器更具有滚动学习的能力。

表3 三种状态下NBTIA和BPTIA交通事件检测算法的性能比较

Tab.3 The performance index comparison NBTIA and BPTIA for three traffic status

交通状态	检测算法	$R(W)$	$E(U)$	DR/%	FAR/%	MDR/%	ADT/s
闲散( $\lambda=0.45$ )	BPTIA	36(38)	4(124)	94.74	3.22	5.26	0.92
	NBTIA	38(38)	2(124)	100.0	1.61	0.00	0.76
顺畅( $\lambda=0.60$ )	BPTIA	43(46)	6(116)	93.48	5.17	6.52	0.92
	NBTIA	46(46)	3(116)	100.0	2.59	0.00	0.77
繁忙( $\lambda=0.75$ )	BPTIA	39(42)	7(120)	92.86	5.83	7.14	0.94
	NBTIA	41(42)	4(120)	97.62	3.30	2.38	0.77

注: $R$ 为正确识别个数, $W$ 为交通事件总数, $E$ 为错误识别个数, $U$ 为非交通事件总数。

NBTIA算法的部分数据测试结果如表4所示。在整个交通事件检测过程中,仅漏判了一个事件数据。在该事件发生时,事件发生断面的车流速度为 $12.27 \text{ m} \cdot \text{s}^{-1}$ ,计算的事件后验概率值为0.5581,这可能是因为研究假定的交通事件发生的速度阈值为 $13 \text{ m} \cdot \text{s}^{-1}$ ,且计算的事件的后验概率值与0.60的判定阈值接近所致。

表4 NBTIA算法部分测试数据的检测结果

Tab.4 The detection results by part of test data

$v/(\text{m} \cdot \text{s}^{-1})$	$P(C=1 X)$	$v/(\text{m} \cdot \text{s}^{-1})$	$P(C=1 X)$
34.32	0.0043	23.04	0.0737
25.06	0.0104	12.27	0.5581
7.68	0.9998	10.09	0.9842
27.41	0.1241	28.28	0.0445
2.55	1.0000	20.69	0.1080

## 4 结语

将朴素贝叶斯分类算法应用于高速公路非重现交通事件的检测,提出了一种基于朴素贝叶斯分类的高速公路交通事件检测算法。本文引入了“0-1”二值分类思想,提出了交通事件的朴素贝叶斯检测算法的理论依据,并遵循交通流的波动理论,以事件检测断面在滞后时间段内交通参数的检测数据为基础,构建了交通事件特征属性向量的概念模型。采用区间分段离散的方法将连续型的特征属性离散化,结合朴素贝叶斯算法和交通事件检测的特征,构造了交通事件的朴素贝叶斯分类器,提出事件检测流程。对一条典型高速公路的路段进行试验,开发了

Matlab交通事件检测程序,在不同强度的交通场景下,采用VISSIM仿真提供的交通事件基础数据,分析本文算法的检测性能。仿真结果表明基于朴素贝叶斯的交通事件检测算法具有良好的事件检测能力,平均检测时间短,且随着路况交通强度升高,算法性能稳定。在未来研究中,将采用实际高速公路的交通事件管理数据对本文算法的性能进行验证。

## 参考文献:

- [1] Srinivasan D, Jin X, Cheu R L. Adaptive neural network models for automatic incident detection on freeways [J]. Neurocomputing, 2005, 64: 473.
- [2] Hawas Y E. A fuzzy-based system for incident detection in urban street networks [J]. Transportation Research Part C: Emerging Technologies, 2007, 15(2): 69.
- [3] Tang S, Gao H. Traffic-incident detection-algorithm based on nonparametric regression [J]. IEEE Transactions on Intelligent Transportation Systems, 2005, 6(1): 38.
- [4] Cheu R L, Srinivasan D, Teh E T. Support vector machine models for freeway incident detection [C]// The IEEE Conference on Intelligent Transportation Systems. Madrid: IEEE, 2003: 238-243.
- [5] Levin M, Krause G M. Incident detection: a Bayesian approach [J]. Transportation Research Record, 1978, 682: 52.
- [6] Ozbay K, Noyan N. Estimation of incident clearance times using Bayesian networks approach [J]. Accident Analysis & Prevention, 2006, 38(3): 542.
- [7] Abdulhai B, Ritchie S G. Enhancing the universality and transferability of freeway incident detection using a Bayesian-based neural network [J]. Transportation Research Part C: Emerging Technologies, 1999, 7(5): 261.

(下转第595页)