

基于特征加权的交通事件检测研究

王 晨^{1,2} 樊小红³

(1 长安大学 公路学院, 陕西 西安 710086; 2 陕西省高速公路建设集团公司, 陕西 西安 710086;

3 长安大学 电控学院, 陕西 西安 710064)

摘 要: 针对影响交通事件的特征参数较多, 参数之间有信息冗余, 影响检测效率的问题, 提出一种基于特征加权支持向量机的交通事件检测算法。影响交通事件的因素包括上下游交通流的交通密度、交通量、平均速度较多, 各个影响因素的影响大小是不同的, 根据可信间隔最大化确定影响权重。通过实测数据对算法的检测性能进行测试, 交通密度的权重最大, 说明事件发生时, 交通密度的变化影响最大, 与实际情况相符; 对于同质量的样本, 所提算法的检测率及误报率均优于标准支持向量机(SVM)算法。

关键词: 交通事件检测; 支持向量机; 特征加权; 分类间隔

中图分类号: TP31

文献标识码: A

文章编号: 1000-7180(2012)10-0121-03

Study on Feature Weighed Automatic Incident Detection

WANG Chen^{1,2}, FAN Xiao-hong³

(1 College of Highway, Chang'an University Xi'an 710086, China; 2 Shaanxi Expressway Construction Group Company Xi'an 710086, China; 3 College of Electronic and Control, Chang'an University, Xi'an 710064, China)

Abstract: A feature weighed support vector machine was proposed to solve for low detection of automatic incident detection caused by redundant feacture. Because the factors related to incident detection include occupy, traffic volume and average speed of the upstream and downstream, and the influence of each factor is different, the influence value was determined by the classe margin of the each feature. Using actual traffic data, the detection performance of the feature weighed SVM algorithm was tested. The results show that the weighed value of occupy is the biggest. It indicate that the influence of occupy is the biggest, which is consistent with the fact. For the same sample, the performance of the proposed algorithm is superior to the standard support vector machine (SVM).

Key words: automatic incident detection; support vector machine; feature weighed; classe margin

1 引言

交通事件检测一直是交通监控研究的热点之一。自从上世纪七十年代以来, 提出了各种交通事件检测算法。现有的交通事件检测算法大致可分为三类: 模式识别法、统计预测法、智能算法等。模式识别法、统计预测法都需要根据经验确定门限参数, 算法的移植性差。当前广泛使用的算法是智能算法, 主要有神经网络法、遗传算法、支持向量机、小波算法^[1-4]等。文献[2]中应用神经网络进行交通事件检测, 人

工神经网络具有良好的非线性逼近能力和自适应自学习的功能, 但是存在需要确定网络结构, 要求的训练样本大等问题。支持向量机(SVM)形式上类似于多层前向网络, 对小样本、非线性、高维问题具有良好的泛化能力, 而且, 可自动解决网络结构问题, 文献[3-4]中对支持向量机在事件检测中的应用进行了研究, 结果表明标准 SVM 能提高检测效率。

上述检测算法的输入变量的选取大都是依据经验, 选择事件发生时变化较明显的交通参数做为输入变量, 例如, 著名的 Califina 算法是以上、下游交

通的占有率为检测变量. 选择单一交通参数的变化, 信息量不够充分, 支持向量机则把所有可能相关的交通参数都做为输入变量, 通常取 $t, t+1, t+2$ 时刻上、下游交通的交通密度、平均速度、交通量共 18 个参数为输入变量, 输入参数较多, 对算法带来以下不利影响: (1) 算法要考虑所有因素的变化, 影响因素之间的信息冗余使得算法的识别精度受到影响. (2) 对算法的学习速度要求较高. (3) 算法平等对待各个参数的影响, 直接应用于参数影响大小有差别的问题中, 显然不合适. 基于以上原因, 文章提出采用基于特征加权^[5]的交通事件检测算法, 根据交通流三参数对事件检测的重要性不同, 确定各个属性特征的权值, 建立事件检测的非线性模型, 仿真实验证明了所提出算法的可行性及有效性.

2 算法介绍

2.1 支持向量机

支持向量机是从线性可分情况下的最佳超平面发展而来的. 对于一组带有类别标记的训练样本集 $\{x_i\}$, $x \in R^l$, $y_i \in \{+1, -1\}$, $i=1, \dots, l$. 若超平面 $w \cdot x + b = 0$ 能将样本正确分为两类, 则最佳超平面应使两类样本到超平面最小距离之和最大. 最佳超平面通过求解下面的优化问题得到

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \epsilon_i \\ \text{s. t.} \quad & y_i[(w \cdot w_i) + b] \geq 1 \quad i = 1, 2, \dots, l \\ & \epsilon_i \geq 0 \quad i = 1, 2, \dots, l \end{aligned}$$

其中, ϵ_i 为误差项, C 为惩罚系数.

利用 Lagrange 乘子法可以把上述求解最佳超平面的问题转化为对偶问题:

$$\begin{aligned} \max \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j k(x_i \cdot x_j) \\ \text{s. t.} \quad & \sum_{i=1}^l \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \end{aligned}$$

相应的决策函数变为

$$f(x) = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i k(x_i, x) + b \right)$$

其中, $k(x_i, x) = \phi(x_i) \cdot \phi(x)$ 称为核函数, 核函数的选取应使其为特征空间的一个内积. 常用的核函数有:

线性核函数: $k(x_i, x) = x_i \cdot x$

高斯核函数: $k(x, x_i) = \exp(-\gamma_i \|x - x_i\|^2)$

Sigmoid 核函数: $k(x, x_i) = 1 / (1 + \exp[\beta(x \cdot$

$x_i) + b]$.

2.2 特征加权 SVM

标准的支持向量机是在假设样本向量的各维特征对分类的贡献是相同的, 标准的支持向量机的性能大大下降, 特征加权支持向量机是从特征重要性角度来考虑问题, 依据某种准则对数据集中的各个特征赋予一定的权重, 用特征权重拉伸标准的欧氏距离, 从而导致欧氏空间中点与点之间位置关系的改变, 以提高机器学习算法的学习能力. 特征权重的确定是该方法的关键. 文中采用分类间隔法^[6]确定特征权重. 步骤如下:

(1) for $i: 1$ to N (N 为特征向量维数).

(2) for $j: 1$ to n (n 为样本个数). 建立与样本 x_j 同类的样本集合 $C1$ 和异类的样本集合 $C2$, 计算 $C1, C2$ 内所有样本与 x_j 的可信间隔 $d_{ij}[C1], d_{ij}[C2]$. $d_{ij}^k = \exp(-|x_i^k - x_j^k|)$

(3) 用下式计算第 j 个样本第 i 个特征 w_i^j 权值:

$$w_i^j = |r_{ij}[C1] - r_{ij}[C2]|.$$

(4) 第 i 个特征权值 $w_i: w_i = \frac{1}{n} \sum_{j=1}^n w_i^j$

3 实验结果及分析

3.1 实验环境

交通事件检测的依据是事件发生时, 描述上下游交通状态的参数将发生急剧变化, 上游交通的交通密度变大, 平均速度变小, 而下游交通的交通密度变小, 平均速度变大. 所以选择描述交通流特征的参数为属性变量, 即选择 $t, t+1, t+2$ 时刻上下游交通的占有率、平均速度、交通量为属性变量. 实验数据使用的是 I-880 交通流数据, 每 30 s 记录一次交通流数据, 选取其中的 200 个样本, 事件样本数 13 个. 核函数选用 $K(x, x_i) = \exp(-\gamma_i \|x - x_i\|^2)$, 参数 $\gamma_i = 2$.

3.2 实验结果及分析

为了验证特征个数对事件检测算法检测精度的影响, 首先计算特征权重, 特征权重分布如图 1 所示.

横轴数字 1~18 所代表的变量分别如下:

1: t 时刻上游平均速度 $v_1(t)$

2: t 时刻上游占有率 $o_1(t)$

3: t 时刻上游交通量 $q_1(t)$

4: t 时刻下游平均速度 $v_2(t)$

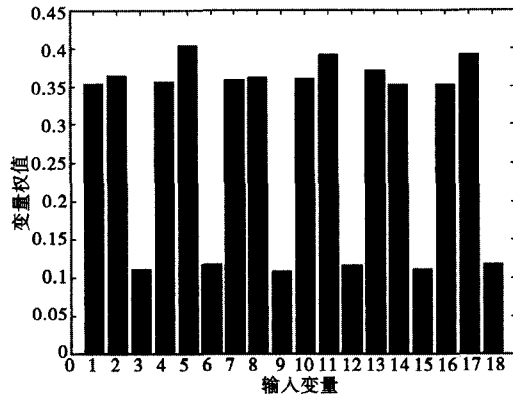


图1 特征权重分布图

- 5: t 时刻下游占有率 $o_2(t)$
- 6: t 时刻下游交通量 $q_2(t)$
- 7: $t+1$ 时刻上游平均速度 $v_1(t+1)$
- 8: $t+1$ 时刻上游占有率 $o_1(t+1)$
- 9: $t+1$ 时刻上游交通量 $q_1(t+1)$
- 10: $t+1$ 时刻下游平均速度 $v_2(t+1)$
- 11: $t+1$ 时刻下游占有率 $o_2(t+1)$
- 12: $t+1$ 时刻下游交通量 $q_2(t+1)$
- 13: $t+2$ 时刻上游平均速度 $v_1(t+2)$
- 14: $t+2$ 时刻上游占有率 $o_1(t+2)$
- 15: $t+2$ 时刻上游交通量 $q_1(t+2)$
- 16: $t+2$ 时刻下游平均速度 $v_2(t+2)$
- 17: $t+2$ 时刻下游占有率 $o_2(t+2)$
- 18: $t+2$ 时刻下游交通量 $q_2(t+2)$

其中 3、6、9、12、15、18 这 6 个变量权值相对较小,而且数值基本相等,其余 12 个变量权值相对较大,可见各时刻上、下游交通的交通量对交通事件检测的影响较小,所以在进行交通事件检测时,只需取各时刻上、下游交通的占有率和平均速度为输入变量即可。

根据权重大小不同进行分组仿真,检测结果如表 1 所示。可见,如果选择 t 、 $t+1$ 、 $t+2$ 时刻交通量、占有率、平均速度 18 个参数同时作为属性特征,检测率没有提高,误报率反而增大。而分别以各个时刻的交通量、占有率或平均速度单独作用作用时,信息量不充分,检测率较低,以 t 、 $t+1$ 、 $t+2$ 时刻的占有率、平均速度两参数为属性特征时,检测率最大,误报率相对较小。可见,根据特征权重大小选择事件检测参数是可取的。以 t 、 $t+1$ 、 $t+2$ 时刻的占有率、平均速度为属性特征,重新计算特征权重,应用特征加权法仿真,检测率为 0.909,误报率为 0.053 4。

表 1 特征个数对检测结果的影响

属性特征	检测率	误报率
各时刻交通量	0	0.123 0
各时刻平均速度	0.692 3	0.071 1
各时刻占有率	0.692 3	0.055 1
各时刻交通量、平均速度	0.846 1	0.124 6
各时刻交通量、占有率	0.769 2	0.129 5
各时刻占有率、平均速度	0.846 1	0.103 1
各时刻交通量、占有率、平均速度	0.846 1	0.135 9

4 结束语

本文将特征加权支持向量机引入到交通事件检测算法中,采用可信间隔确定特征权重,根据权重大小选择属性特征及加权系数。同标准的支持向量机检测算法相比,该方法能有效识别事件检测的有效特征,消除了属性特征间的信息冗余,提高了算法的检测效率。

参考文献:

- [1] Young-seon Jeong, Manoel Castro-Neto, Myong K Jeong. A wavelet-based freeway incident detection algorithm with adapting threshold Parameters [J]. Transportation Research Part C, 2011(19): 1-19.
- [2] 樊小红,荆便顺. 基于遗传算法的交通事件检测[J]. 长安大学学报:自然科学版, 2005, 25(4): 70-72.
- [3] Chen Shayan, Wang Wei, Henk Van Zuylen. Construct support vector machine ensemble to detect traffic incident [J]. Expert Systems with Applications, 2009 (36): 10976-10986.
- [4] 陈斌. 基于支持向量机的高速公路意外事件检测模型[J]. 中国公路学报, 2006, 19(6): 107-112.
- [5] 汪廷华,田盛丰,黄厚宽. 特征加权支持向量机[J]. 电子与信息学报, 2009, 31(3): 514-518.
- [6] 姜慧研,柴天佑. 基于可信间隔的特征选择方法研究[J]. 控制与决策, 2011, 26(8): 1229-1238.

作者简介:



王晨男, (1974—), 博士研究生。研究方向为交通规划、交通事件检测、人工智能。