

基于模糊支持向量机的高速公路交通事件的自动检测

焦军彩

(南京工业大学 数学系, 江苏 南京 210009)

摘要:利用支持向量机的全局优化、适应性强、泛化性能好等优点,针对实时交通流数据的随机性、高维、非线性和时变等特性,将模糊支持向量机应用于高速公路交通事件检测问题中。在识别阶段利用60组实测数据训练模糊支持向量机,利用60组实测数据进行测试,测试结果表明,利用FSVM进行交通事件检测,识别率达到96.7%。

关键词:模糊支持向量机;交通事件自动检测;隶属函数

中图分类号:U495 **文献标识码:**A **文章编号:**1671-5322(2009)04-0070-03

由于交通事件引起的拥挤是突发的、事先无法预测的,所以极易引起交通瓶颈和二次交通事故,因此及时检测出交通事件并发出警告是高速公路交通管理系统中的重点之一。目前已经有的事件检测算法,大体可以分为模式识别方法和统计预测方法。典型的模式识别算法主要有:加利福尼亚算法和McMaster算法。代表性的统计预测算法有:标准正常偏差法、贝叶斯算法等。近年来,又出现了一些新的算法,其中有突变论法、神经网络法、图像处理法、支持向量机法(Support Vector Machines—SVM)等。利用人工神经网络训练数据进行事件检测有着一定的优势,然而,神经网络需要大量数据进行再训练才能使其具有好的移植性。视频图像处理应用于交通事件自动检测是近年来的新技术,然而,相机探测范围和数据处理时间的限制依然是其运用于实时交通事件检测的主要问题。支持向量机基于结构风险最小化原理,同时还采用了二次规划、拉格朗日算法,其解是全局最优解,具有很高的泛化能力,因此,利用支持向量机进行事件的检测可有效的提高检测率和降低误检率。然而,SVM在找最优分类面时会因为一些奇异离群点而与真正的最优平面有较大的差别,从而降低SVM的泛化能力^[1-2]。

高速公路交通流,特别是在发生事件的情况下,是高度非线性、时变、不能精确建模或实时定量表示的,许多交通概念具有重要但不精确的含

义。本文利用支持向量机的全局优化、适应性强、泛化性能好等优点,针对实时交通流数据的随机性、高维、非线性和时变等特性,将模糊支持向量机应用于高速公路交通事件检测问题中。

1 算法研究

SVM是一种基于统计学习理论的机器学习算法,SVM方法是从线性可分的情况下的最优分类面提出的,在讨论这种条件下分类方法的基础上,逐渐将问题推广到非线性可分的情况。

SVM在找最优分类面时会因为一些奇异离群点而与真正的最优平面有较大的差别,从而降低了SVM的泛化能力。这是因为SVM在分类过程中样本是以同等地位输入的,但实际样本对分类的作用大小不一。我们知道权向量解中只有支持向量,非支持向量对权向量解并没有影响。如果奇异点成为支持向量则严重影响SVM的泛化能力。因此,SVM如果能够正确估计样本对分类作用的大小,就会降低奇异点在分类中的贡献。为此,我们将模糊数学引入其中,提出了FSVM。根据样本对分类贡献的不同赋予相应的隶属度,从而减小噪声的影响,提高SVM的分类能力。

对于最简单的两分类问题,首先对数据进行预处理:事先选择一个适当的隶属函数,对所有样本进行模糊化,得到每一个样本 x_i 的隶属度 μ_i 。于是训练集便成为模糊训练集:

收稿日期:2009-06-02

作者简介:焦军彩(1974-),女,河北藁城人,讲师,硕士,主要研究方向为模糊数学。

$$(x_1, y_1, \mu_1), (x_2, y_2, \mu_2), \dots, (x_n, y_n, \mu_n)$$

$$x_i \in R^m, \quad y_i \in \{-1, 1\},$$

$$0 \leq \mu_i \leq 1 \quad i = 1, 2, \dots, N$$

对于上述新的训练集,为了得到最优分类超平面,需要解决以下优化问题:

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2}(w^T w) + c \sum_{i=1}^N \mu_i \xi_i \\ \text{s.t.} \quad & y_i [w \cdot \Phi(x_i) + b] \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad i = 1, 2, \dots, N \end{aligned}$$

其中 C 为惩罚参数, ξ_i 是我们引入的松弛变量,它度量一个数据点对模式可分的理想条件的偏离程度。 Φ 将 x_i 从 R^m 映射到高维特征空间。

上面优化问题的解释也由以下拉格朗日函数的鞍点给出:

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2}(w^T w) + c \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \alpha_i [y_i (w \cdot \Phi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^N \beta_i \xi_i \quad (1)$$

其中 α_i, β_i 为拉格朗日乘子。 w, b, ξ 须满足下列条件:

$$\frac{\partial L(w, b, \xi; \alpha, \beta)}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i \Phi(x_i) = 0$$

$$\frac{\partial L(w, b, \xi; \alpha, \beta)}{\partial b} = - \sum_{i=1}^N \alpha_i y_i = 0$$

$$\frac{\partial L(w, b, \xi; \alpha, \beta)}{\partial \xi} = \mu_i c - \alpha_i - \beta_i = 0$$

应用以上三式到(1)式,可将(1)转化为:

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

而(2)式中对应的最优分类函数变为:

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i^0 y_i K(x_i, x) + b_0 \right)$$

其中, $b_0 = y_k - \sum_{x_i \in S_{\text{SV}}} \alpha_i^0 y_i K(x_i, x_k) \quad k \in \{i | \alpha_i > 0\}$

通过 FSVM 最后得到的分类函数对重要数据的分类精度显著提高,并且有较强的抗噪音能力。

2 数据分析和结果分析

当高速公路某一路段发生事件时,在出事地点上下游的一定范围内将出现交通流反常:上游车辆因交通受阻而减速,占有率增大;下游车辆因稀少而加速,占有率减小等等。因此,本文以上下游两个固定检测器采集到的路段的数据信息为输入变量,包括:检测时间段内的平均流量(流率)、

平均事件占有率、平均速度。数据来源于参考文献[3],共120组数据,前60组为训练样本,后60组为测试样本。训练样本 $x_i = (x_{i1}, x_{i2}, \dots, x_{i6})$, $i = 1, 2, \dots, 60$ 是一六维向量,其中 x_{ik} ($k = 1, 2, \dots, 6$) 分别表示:上游的流量、上游的占有率、上游的速度、下游的流量、下游的占有率、下游的速度。将这60组训练样本数据代入模糊支持向量机模型,得到如下规划问题:

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

在 FSVM 中样本点的隶属度的确定很重要,也是应用 FSVM 方法首先要解决的问题。我们知道如果样本点距离模式类中心越远则它属于这类的可能性越小,属于奇异点的可能性就越大,FSVM 应该赋予它较低的隶属度。所以我们用下列方法确定隶属度:

$$\mu_i = e^{-\frac{(x_i - m_c)^2}{\sigma^2}} \quad i \in \{i | y_i = 1\}$$

其中: m_c 为模式类中心, σ 为宽度参数。类似可以定义另一模式类。

核函数采用高斯径向基函数(RBF): $K(x_i, x_j) = \exp(-\gamma(x_i - x_j)^2)$, γ 为参数。

通过 Matlab 编程,求解得到 α_i 的最优解 α_i^0 。通过对支持向量机的分析,我们知道,只有非零的 α_i 对应的向量为支持向量,因此我们得到8个支持向量 x_i ($i = 1, 2, \dots, 8$):

$$\begin{aligned} & (0.90909, 1.89E-10, 0.92286, 0.6875, 1, 0) \\ & (0.454545, 0.163574, 0.55768, 0.4375, \\ & 0.052626, 0.8148469) \\ & (0.090909, 0.293655, 0.191725, 0.4375, \\ & 0, 0.8511408) \\ & (0.454545, 0.278252, 0.278302, 0.5625, \\ & 0.1129074, 0.8910722) \\ & (0.545455, 0.308825, 0.327454, 0.75, \\ & 0.289618, 0.8419527) \\ & (0.363636, 0.269805, 0.219912, 0.6875, \\ & 0.0894078, 0.8576034) \\ & (0.727273, 0.217217, 0.397173, 0.875, \\ & 0.1797774, 0.8408655) \\ & (0.363636, 0.162834, 0.345698, 0.625, \\ & 0.1415109, 0.7676298) \end{aligned}$$

从而得到最优分类函数:

$$f(x) = \text{sign}(\sum_{i=1}^8 \alpha_i^0 y_i K(x_i, x) + b_0)$$

将后 60 组数据代入最优分类函数,见表 1。

表 1 分类结果

Table 1 Results of classification

所属 类	判别 结果	所属 类	判别 结果	所属 类	判别 结果	所属 类	判别 结果
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	0	1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	0	1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	0	-1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	0	-1
1	1	1	1	1	1	0	-1
1	1	1	1	1	1	0	-1

传统高速公路交通事件自动检测方法的评价指标是在线情况下在一段时间内测试得到的结果,鉴于本文是离线测试,因此,提出一种新的评价指标——识别准确率:

$$\text{识别准确率} = \frac{\text{准确识别出的测试样本数}}{\text{总的测试样本数}}$$

从上表的识别结果可以看到,在 60 组测试样本中,有两个有事件的样本被错误地识别为没有事件发生,识别准确率为 96.7%。以上的实验结

参考文献:

[1] 刘伟铭. 高速公路系统控制方法[M]. 北京:人民交通出版社,1998.
[2] 杨纶标,高英仪. 模糊数学原理及应用[M]. 广州:华南理工大学出版社,2002.
[3] 王琪. 基于神经网络和支持向量机的高速公路交通事件检测[D]. 成都:西南交通大学,2006.

Approach to Automatic of Highway Detection Incident Based on Fuzzy support vector machine

JIAO Jun-cai

(Department of Mathematics Nanjing University of Technology, Jiangsu Nanjing 210009, China)

Abstract:Support Vector Machine(SVM) has the advantages of global solutions, good adaptability, high generalization in theory. Due to the randomness, high dimension, nonlinear, time - variant of the traffic flow data, Fuzzy support vector machine(FS-VM) applied to high way incident detection In the stage of traffic incident recognition, sixty train data was used to train the FS-VM, sixty test data was used to testify the effect of the recognition model. The results show that the recognition rate is approximate to 96.7%, which also proved the feasibility of the approach proposed in this paper.

Keywords:fuzzy support vector machine;traffic incident automatic detection;Subjecton function

(责任编辑:沈建新;校对:张英健)

果表明,相比支持向量机,利用模糊支持向量机得到的分类函数对数据的分类精度有显著提高,并且有较强的抗噪声能力,解决了传统支持向量机抗干扰能力较差的特点。

为了检验模糊支持向量机的效果,笔者采用相同的训练数据和测试数据,利用支持向量机模型进行检测。利用 Matlab 对支持向量机和模糊支持向量机两种方法的检测结果进行比较(表 2)。

表 2 SVM 与 FSVM 检测结果比较

Table 1 Test result comparison of SVM and FSVM

检测方法	识别准确率/%	误检率/%
SVM	95.8	4.2
FSVM	96.7	3.3

3 总结

本文将模糊理论和支持向量机理论相结合,构建了一种基于模糊支持向量机的预测模型,以上下游的平均流量(流率)、平均占有率、平均速度组成的 6 维数据信息作为输入变量,对模糊支持向量机进行训练,根据样本对分类贡献的不同赋予相应的隶属度,求出支持向量,从而构造判别函数,对高速公路交通事件进行自动检测。通过对该算法进行离线的性能评价,测试结果表明,利用模糊支持向量机进行交通事件自动检测是可行的,且优于一般的检测方法,从而为事件自动检测提供了一种更科学有效的方法。