

# 基于模糊支持向量机的城市道路交通状态分类

李清泉<sup>1,2</sup>, 高德荃<sup>1,2</sup>, 杨必胜<sup>1,2</sup>

(1. 武汉大学 测绘遥感信息工程国家重点实验室, 武汉 430079; 2. 武汉大学交通研究中心, 武汉 430079)

**摘要:**运用模糊支持向量机方法进行城市道路交通状态分类,应用模糊隶属度处理了评价者的主观偏好表达和交通状态评价指标范围的不确定性问题;同时,通过集成支持向量机分类学习的能力修正了单纯模糊分类方法无法进行样本训练的局限,采用“一对一”方法进行交通状态多类分类。最后基于微观交通仿真数据,验证了模糊支持向量机方法在交通状态分类中能够减少样本数据噪声影响,提高分类训练学习的质量。

**关键词:**交通运输工程;交通状态;分类;模糊集;支持向量机

**中图分类号:**U491 **文献标识码:**A **文章编号:**1671-5497(2009)Sup. 2-0131-04

## Urban road traffic status classification based on fuzzy support vector machines

LI Qing-quan<sup>1,2</sup>, GAO De-quan<sup>1,2</sup>, YANG Bi-sheng<sup>1,2</sup>

(1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; 2. Transportation Research Center, Wuhan University, Wuhan 430079, China)

**Abstract:** This paper uses fuzzy support vector machines (FSVM) algorithm to classify urban road traffic status. Fuzzy membership handles linguistic expression bias of human perception and uncertainty in status parameter partition scale. Meanwhile, integrated SVM learning ability can solve the limitation that pure fuzzy classification method can't train sample data. The one against one approach in FSVM is applied for multi-class classification. Finally, a microcosmic simulation work is performed for an experiment data example, and the results indicates that FSVM method can reduce influence of noise sample data and provide better classification accuracy.

**Key words:** engineering of communication and transportation; traffic status; classification; fuzzy sets; fuzzy support vector machines

道路交通状态检测与分类一直是国内外交通工程研究的热点,交通状态的检测是根据实时的动态交通参数信息,准确、快速地判断出路网中交通流的运行状态,及时地发现路网中存在的交通拥挤。近几十年来不同领域的研究者从各自的角度对交通流状态的特性进行了分析,涌现出了大

量的研究成果。一些研究者将已有检测分类算法分为四类<sup>[1-6]</sup>:①直接比较算法。如早期的加利福尼亚系列算法和 McMaster 算法,指数平滑方法等;②时空预测算法。如基于状态空间转换模型的 cell transmission model,高斯混合隐马尔可夫模型 (Gaussian Mixture Hidden Markov

**收稿日期:**2009-03-13.

**基金项目:**“973”国家重点基础研究发展规划项目(2006CB705500);国家自然科学基金项目(60872132);高等学校博士学科点专项科研基金项目(20070486001).

**作者简介:**李清泉(1965-)男,教授,博士生导师.研究方向:交通地理信息系统. E-mail:qqli@whu.edu.cn

Models, GM-HMM); ③模式识别算法。如线性判别式分析, 扩展卡尔曼滤波和 Unscented 卡尔曼滤波, 贝叶斯交通状态估计模型, 模糊方法等; ④人工智能算法。如自组织神经网络, 构造性概率神经网络 (Constructive Probabilistic Neural Networks, CPNN) 等。传统的交通状态研究多是针对高速公路识别交通事件, 而较少整体考虑城市路网交通状态。本文综合运用模糊集和支持向量机方法, 提出应用模糊支持向量机方法对城市道路交通状态分类识别。

### 1 支持向量机

支持向量机<sup>[9]</sup>是建立在统计学习理论基础之上, 根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷, 以期获得最好的推广能力。近年来 SVM 在理论研究和算法实现方面都取得了突破性进展, 开始成为克服“非线性和维数灾难问题”、“过学习问题”和“局部极小点问题”等的有力手段。

SVM 方法是从线性可分情况下的最优分类面提出的, 对于线性不可分情况, 当用一个超平面不能把两类点完全分开时, 可以引入松弛变量  $\xi_i$  ( $\xi_i \geq 0, i = 1, 2, \dots, n$ ), 使超平面  $w^T x + b = 0$  满足:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad (1)$$

当  $\xi_i < 1$  时样本点  $x_i$  仍旧被正确分类, 而当  $\xi_i \geq 1$  时样本点  $x_i$  被错分。为此, 引入以下目标函数:

$$\phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \quad (2)$$

式中:  $C$  是一个正常数, 称为惩罚因子。

此时 SVM 可以通过二次规划(对偶规划)来实现:

$$\left\{ \begin{array}{l} \max \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i^T x_j) \\ \text{s. t} \quad a_i \geq 0, i = 1, \dots, n \\ \sum_{i=1}^n a_i y_i = 0 \end{array} \right. \quad (3)$$

用核函数  $K(x_i, x_j)$  代替最优分类平面中的点积  $x_i^T x_j$ , 相当于把原特征空间变换到了某一新的特征空间。最终得到相应的分类判别函数式为:

$$f(x) = \text{sgn} \left( \sum_{i=1}^n a_i y_i K(x_i, x) + b \right) \quad (4)$$

支持向量机中常用的核函数主要有四类, 分别为线性、多项式、径向基和 Sigmoid。

### 2 基于模糊支持向量机的交通状态分类

#### 2.1 模糊支持向量机模型

支持向量机分类方法不必知道因变量和自变量之间的关系, 通过对样本的学习即可获得因变量和自变量之间非常复杂的映射关系。模糊集合是一种处理不精确、不完全信息的方法, 能有效地以类似人的概念处理模式来表达类别问题。与 SVM 建模方法相比, 模糊建模的优势在于可以充分反映经验数据, 它所建立的模型简单。在标准 SVM 的分类算法中, 对所有训练样本的重要程度是同等对待的, 但实际上每个样本的作用是不一样的。用模糊支持向量机方法来对交通状态的分类<sup>[10]</sup>, 其主要目标和思想为: 针对支持向量机对训练样本内的噪音和孤立点的敏感性, 在支持向量机中引入模糊参数, 使得在构造目标函数时, 不同的样本有不同的贡献, 从而减弱了噪音和孤立点对分类的影响, 使学习算法在代价敏感数据或含噪声数据的情况下更加具有稳健性。

基于模糊支持向量机, 交通状态分类输入的模糊样本为:  $(x_1, y_1, \mu(x_1)), (x_2, y_2, \mu(x_2)), \dots, (x_n, y_n, \mu(x_n)), i = 1, 2, \dots, n$ 。

式中:  $\mu(x_i)$  为隶属度,  $0 < \mu(x_i) \leq 1$ 。 $\mu(x_i)$  表示了交通状态特征参数模糊化后的模糊预选规则, 度量了该样本隶属某类别的可靠程度; 同时, 在支持向量机的训练过程中, 说明每个训练数据对支持向量机学习所起的权重作用是不同的。通过模糊隶属度, 模糊支持向量机的最优分类面为下面目标函数的最优解

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \mu_i \xi_i \quad (5)$$

约束条件为

$$\left\{ \begin{array}{l} y_i(w \cdot x_i + b) + \xi_i \geq 1 \\ \xi_i \geq 0, i = 1, \dots, n \end{array} \right. \quad (6)$$

式中:  $C$  为惩罚函数因子;  $\xi_i$  是松弛变量, 为目标函数分类的误差度量。因而,  $\mu_i \xi_i$  表示重要性不同的样本错分程度的度量。

在支持向量机方法中, 参数  $C$  用来控制样本偏差与机器推广能力之间的平衡,  $C$  越大, 惩罚

就越大,对错分样本的约束程度就越大,得到的分类间隔越小。在模糊支持向量机中,设置  $C$  为一个较大的值,如果取所有的隶属度  $\mu(x_i)$  为 1,则与标准支持向量机的方法一样。因此在模糊支持向量机方法中,根据不同输入数据对分类结果影响的不同,对不同的样本赋予不同的隶属度  $\mu(x_i)$ ,达到对不同样本采用不同程度的惩罚作用。

则对应的最优分类面的决策函数为

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right\} \quad (7)$$

式中:  $0 \leq \alpha_i \leq \mu(x_i)C$ ;  $K(x_i, x)$  为核函数。

按照上述方法求得的模糊最优分类函数与传统 SVM 的最优分类函数几乎完全相同,原因在于模糊最优分类面仍采用使分类间隔最大化、使分类错误最小化的思想。所不同的是,在 FSVM 中将惩罚参数模糊化,对不同的样本使用不同的惩罚参数。选择不同的核函数,可以实现输入空间中不同类型的线性或非线性决策面的学习。

## 2.2 交通状态多类分类

道路交通状态识别是一个多类分类问题,由于传统的支持向量机方法在分类问题上只考虑了二值分类的问题,所以需要扩展 SVM 建立多个支持向量机分类器。目前,构造 SVM 多类分类器的方法主要有两类:一类是直接法,直接在目标函数上进行修改,将多个分类面的参数求解合并到一个最优化问题中,通过求解该最优化问题“一次性”实现多类分类。这种方法计算复杂度比较高,实现起来比较困难,只适合于小型问题中;另一类是间接法,主要是通过组合多个二分类器来实现多分类器的构造,常见的方法有一对多法(one-against-all)、一对一法(one-against-one)等<sup>[11]</sup>。一对多法训练时依次把某个类别的样本归为一类,其他剩余的样本归为另一类。一对一法是在  $k$  个样本中,分别选取 2 个不同类别构成一个 SVM 子分类器,相当于将多类问题转化为多个两类问题来求解。

## 2.3 交通状态模糊隶属度

综合国内外城市道路交通状态划分情形,本文将城市道路交通状态模式分为 5 个等级水平:畅通、基本畅通、拥挤、拥堵、堵塞。道路交通状态指标应该以各种交通参数为基础,因此交通状态模式的语言描述需要进一步数量映射。本文以道路行程速度为评价参数指标,在不同道路等级下

对交通状态模式进行按速度域范围的划分,实现交通状态模式从定性的语言描述过渡到指标数量化约定表达,如表 1 所示。

表 1 交通状态等级

Table 1 Level of road traffic status

分类等级	A	B	C	D	E
模糊描述	畅通	基本畅通	拥挤	拥堵	堵塞
平均行	快速路 >60	[40,60]	[30,40]	[20,30]	<20
程速度	主干路 >45	[35,45]	[25,35]	[15,25]	<15
(km/h)	次干路 >35	[25,35]	[18,25]	[12,18]	<12
	支路 >25	[18,25]	[12,18]	[8,12]	<8

对于大众而言,交通状态是一个感觉量,在确定交通状态时存在模糊性。这种模糊性既受道路环境的影响,也受人们心理认知的影响,具体表现首先是交通区域的差异性。在不同道路上,人们对同一交通量下的实际交通状况会有不同的感受;其次是使用者的差异性。不同的道路使用者对交通状态的认同存在差异,即使是同一个使用者在同一条道路上在不同的时间、不同的心态下对交通状态的理解也会发生变化。因此,针对交通状态表达的模糊特性,运用模糊数学的方法进行交通特征表达,计算其交通状态的隶属度,以改善支持向量机分类过程中的训练学习性能。

## 3 实验分析

实验样本数据通过 VISSIM4.2 环境仿真获得,一共有 300 组交通速度参数数据。交通速度包括路段平均速度和路段行程速度,路段行程速度作为交通状态分类判别参考依据。300 组速度交通数据分为训练样本集和测试样本集,每组 150 个样本。支持向量机实验环境选用 LIBSVM<sup>[12]</sup>,其多类分类方法运用 One-against-One 策略。首先对表征交通状态的速度进行模糊化处理,接着使用四种核函数分别进行实验。分类精度分别是:线性为 90.7%,多项式为 92.0%,径向基为 90.7%,Sigmoid 为 92.7%。从分类结果看,四种核函数的支持向量机的训练精度相差无几,而在分类结果精度中,采用 Sigmoid 核函数的训练和分类效果都是最好,采用多项式的次之,采用线性和径向基核函数两者分类精度相同。但是四种核函数最终分类精度的差异并不很大说明了支持向量机分类方法具有比较好的稳定性。

实验中采用了不同的方法和测试环境进行分类。分类算法主要为:模糊集、支持向量机和模糊支持向量机。后面两种选用 Sigmoid 核函数。多

类方法对比为 one-against - all(OAA)和 one-against - one(OAO)。分类实验结果如图 1 所示。

图 1 中, SVM(OAO)和 SVM(OAA)为 Canu 等人<sup>[13]</sup>提供的支持向量机多类分类方法。在五种不同分类方法对比中,应用模糊分类精度最差,FSVM 分类精度最好。从直接单纯通过支持向量机对交通速度数据进行分类效果看,使用支持向量机分类精度比模糊分类有较大的提升。SVM(OAO)和 SVM(OAA)的多类分类结果比 FSVM 要差,是因为在 LIBSVM 中对支持向量机的分类参数先进行了交叉寻优的过程。在实验中,SVM(OAO)和 SVM(OAA)采用的分类参数一致,分类精度差别很小。

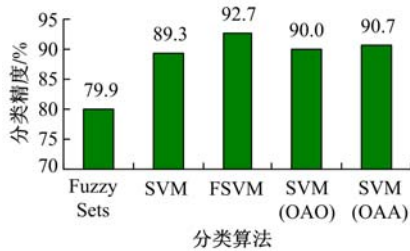


图 1 不同算法分类误差

Fig.1 Classification error with different algorithm

通过模糊方法、支持向量机、FSVM 等分类方法的横向比较,集成支持向量机的分类准确率比模糊分类有了较为明显的提高,也优于单一支持向量机。通过实验研究,说明了模糊支持向量机方法通过模糊隶属度考虑了样本和分类标签之间的约束,提高了分类的准确度,验证了基于模糊支持向量机进行道路交通状态多类分类是可行的。

#### 4 结束语

城市道路交通状态类别的语言定性表达具有模糊性,而在使用微观交通特征参数进行定量度量时存在参数区间划分的不确定性。本文以模糊支持向量机分类集成算法进行道路交通状态识别研究,集成算法用隶属度对交通特征参数表达进行模糊化处理,并采用“一对一”法识别交通状态的多类类属。最后以仿真数据为例进行实验分析。通过结果比较表明,采用模糊支持向量机集成分类的方法在分类精度上比单纯模糊分类有明显的改善,提高了分类学习的能力。

#### 参考文献:

[1] 白竹. 城市主干路交通异常状态自动判别方法研究[D]. 长春: 吉林大学, 2006.

[2] Sun X T, Mufioz L, Horowitz R. Highway traffic state estimation using improved mixture kalman filters for effective ramp metering control[C]// Proceedings of the 42nd IEEE Conference on Decision and Control, Hawaii, USA, 2003: 3413-3417.

[3] Porikli F, Li XK. Traffic congestion estimation using HMM models without vehicle tracking[C]// IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004: 188-193.

[4] 江龙辉. 城市道路交通状态判别及拥挤扩散范围估计方法研究[D]. 长春: 吉林大学, 2007.

[5] Pesti G, Abbas M M, Chaudhary N. Traffic state classification in condition-responsive traffic control[C]// Proceedings of 11th International Conference on Intelligent Engineering Systems, 2007: 33-37.

[6] Hegyi A, Girimonte D, Babuska R, et al. A comparison of filter configurations for freeway traffic state estimation[C]// Proceedings of the IEEE Intelligent Transportation Systems Conference, Toronto, Canada, 2006: 1029-1034.

[7] Yang H, Qiao F X. Neural network approach to classification of traffic flow states[J]. Journal of Transportation Engineering, 1998, 124 (6): 521-525.

[8] Jin X, Srinivasan D, Cheu R L. Classification of freeway traffic patterns for incident detection using constructive probabilistic neural networks[J]. IEEE Transactions on Neural Networks, 2001, 12 (5): 1173-1187.

[9] Vapnik V. An overview of statistical learning theory [J]. IEEE Transaction on Neural Networks, 1999, 10 (5): 988-999.

[10] Lin C F, Wan S D. Fuzzy support vector machines[J]. IEEE Trans on Neural Networks, 2002, 13(2): 464-471.

[11] Hsu C W, Lin C J. A comparison of methods for multi-class support vector machines[J]. IEEE Transaction on Neural Network, 2002, 13(2): 415-425.

[12] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J/OL]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.

[13] Canu S, Grandvalet Y, Guigue V, et al. SVM-KM Toolbox[J/OL]. Available: <http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.htm>, 2005.