

基于支持向量机模式识别的大雾预报方法

贺 皓¹ 罗 慧²

(陕西省专业气象台, 陕西省气象局, 西安 710015)

摘要 选取 1971~2000 年 11~12 月大雾发生前近地面层的气象要素(气温、降水、能见度、风向风速、相对湿度、云量等 9 个预报因子),将支持向量机(SVM)方法应用于大雾预报。采用支持向量机方法,应用径向基函数,建立了陕西公路站点大雾 24 h 预报模型,并进行了大雾预报的模拟、训练,其寻优标准 TS 评分达到了理想的效果。

关键词 支持向量机 径向基函数 预测模型 分类 大雾预报方法

数学上有很多模式识别的方法,如概率统计中的判别分析、聚类分析、主成分分析等,随着计算机技术的发展和智能技术的进步,智能机器识别技能也得到了很好的发展。基于支持向量理论的支持向量机(SVM:Support Vector Machine)方法是一种新颖的小样本机器学习方法^[1,2],该方法建模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,它具有从海量的信息中,自动识别并提取关键信息的特点,适合处理本质上的非线性问题。文献[3]、[4]把 SVM 应用于冬季气候预测和山洪地质灾害预报中,对应的 SVM 分类模型和回归模型均具有良好的预报能力。文献[5]通过日最高气温、日最低气温和平均气温的预报试验表明,SVM 能够用于量化的气象要素客观预报。余健^[6]等把 SVM 方法用于高速公路交通安全气象影响评价体系,用径向基核函数建立的非线性降水分类模型优于用多项式核函数建立的线性降水分类预报,特别是资料减少时,非线性降水分类预报明显优于线性降水分类预报。基于此,把 SVM 方法用于大雾预报,是提高天气预报准确率的一个新途径。

1 SVM 大雾预测模型的设计

建模资料选取陕西省 1971~2000 年 11~12 月逐日地面观测资料,有气温、降水、能见度、风向风速、相对湿度、云量等。

1.1 预测对象的构造

公路交通气象服务是公共气象服务轨道的重要内容,陕西省公路布局可概括为“三纵四横五辐射”,外加 5 条省级高速公路联络线,路网总规模 5000 km,其中国家高速公路约 3500 km,省级高速公路约 1500 km。公路在陕西经济发展中发挥了重要作用,但困扰公路建设的气象灾害较多,特别是影响司机视程的大雾天气,是公路交通的主要灾害之一。本文选择横贯东西向的欧亚大陆桥连云港—霍尔果斯 G045 国道在陕西省经过的渭南、西安和宝鸡,西安—延安线的宜君和南北方向大动脉的安康为预测对象。这 5 个站点与陕西大雾地域特征相联系^[7~9],陕西大雾的气候概率是 7%,宜君是陕西的雾日之最,高达 17%,西安是关中盆地的中心,安康地处汉江中游,都是雾日最多的地方。

1.2 预测因子的构造

大雾的产生是在一定的环流背景下形成的,陕西省大雾主要是辐射雾和平流雾,近地面层的气象要素直接参与其形成。选取的 9 个因子分别是:

x_1 :08:00 气温。大雾形成时间一般在 08:00 前后,气温能直接使用。

x_2 :前一天日降水量,若某日有降水量,取值 1,反之取值 0。研究和观测事实表明,前期降水和次日的雾形成有很大关系。

x_3 :08:00 能见度。建立统计关系,低劣的能见度能否有滞后效应。

x_4 :08:00 风向。因风向是字符形式的资料,对风向做了如下处理:地面风观测取 16 个方位,从北风 N 以数字 1 代替开始,按顺时针方向,依次取 2, 3,⋯,直到北北西北 NNW 取 16 结束。静风 C 用 0 代替。这些数据最后按标准化处理,各个因子对预报对象的贡献是以概率的形式出现的,这样处理不影响模型效果。

x_5 :08:00 相对湿度。

x_6 :前一天的最低气温。

x_7 :08:00 风速。

x_8 :08:00 总云量。

x_9 :08:00 低云量。云和雾为同一概念,当低云接地时,为地面的雾。

上述 9 个预报因子,因量级各异,用下式进行标

准化处理:

$$x'_i = [x_i - \min(x_k)] / [\max(x_k) - \min(x_k)]$$

预报对象(大雾)按 1 和 -1 处理,1 为有大雾, -1 为无大雾。预报时效取 24 h。

1.3 预测模型的建立

将 5 个站 1971~2000 年 11~12 月共 30 年的资料分为 3 部分:①训练样本,也即建模样本 2000 个,约为 1971~1992 年 64 个月资料;②实验样本,用于测试用训练样本建立的 SVM 模型的预报能力,约为 6 年 20 个月 600 个样本;③检验样本,为了将最终确立的 SVM 预报模型对其做预报,以检验 SVM 模型的预报效果(推广能力)190 个,约为 2 年 6 个多月。表 1 给出向量机格式的部分数据。

表 1 向量机格式数据

预测对象	因子序号								
	1	2	3	4	5	6	7	8	9
-1	0.6991	1	0.3429	0	0.85	0.7033	0	1	0
-1	0.5929	0	0.3429	0	0.96	0.6053	0	0	0
1	0.5398	0	0.0143	0	0.98	0.5638	0	0	0
1	0.6018	1	0.1714	0	0.97	0.5875	0	0	0
-1	0.7227	1	0.4857	0	0.55	0.6914	0	0.8	0
-1	0.5811	1	0.5143	0.25	0.73	0.4985	0.4	0.3	0
-1	0.5221	1	0.3143	0	0.93	0.4985	0	1	0
-1	0.5133	0	0.1429	0.6875	0.95	0.5312	0.2	1	0

注:第 1 列为预测对象,1 表示有雾,-1 表示无雾,第 2 列及其以后的数据表示标准化数据。

以径向基函数(满足 Mercer 定理条件,又称高斯核,简记为 RBF)作为核函数建立推理实验模型。径向基函数形式为:

$$K(x,x_i)=\exp(-r\parallel x-x_i\parallel^2)$$

在 SVM 模型识别的分类预报中,基于 RBF 核求得的最终决策函数为:

$$M(x)=\operatorname{sgn}\big[\sum_{\text{支持向量}}\alpha_iy_iK(x,x_i)+b\big]$$
$$=\operatorname{sgn}\big[\sum_{\text{支持向量}}\alpha_iy_i\exp(-r\parallel x-x_i\parallel^2)+b\big]$$

其中, x_i 作为支持向量的样本因子向量, x 为待预报因子向量, α_i 、 b 为建立 SVM 模型待定系数, r 为核参数,求和运算只对支持向量进行。

SVM 模型的参数寻优方案采用 TS 评分准则。

2 支持向量机模型试验效果分析

表 2 给出了各个公路站点大雾用 SVM 方法建

立推理模型的试验结果。

渭南在 2000 个建模训练样本中,获得支持向量 379 个,占到总样本的 19%,用最优模型对检验样本(190 个)进行分类的正确率 96.32%,预报成功率 96.55%,预报概括率 82.35%,TS 评分 80%。

西安在 2000 个建模训练样本中,获得支持向量 910 个,占到总样本的 45%,用最优模型对检验样本(190 个)进行分类的正确率 96.32%,预报成功率 94.44%,预报概括率 73.91%,TS 评分 70.83%。

宝鸡在 2000 个建模训练样本中,获得支持向量 259 个,占到总样本的 13%,用最优模型对检验样本(190 个)进行分类的正确率 97.37%,预报成功率 0%,预报概括率 0%,TS 评分 0%。只有 5 个有雾样本,没有一个被预报出来。

宜君在 2000 个建模训练样本中,获得支持向量 399 个,占到总样本的 20%,用最优模型对检验样本

(190 个)进行分类的正确率 90.53%,预报成功率 67.74%,预报概括率 72.41%,TS 评分 53.85%。

安康在 2000 个建模训练样本中,获得支持向量

257 个,占到总样本的 13%,用最优模型对检验样本 (190 个)进行分类的正确率 94.21%,预报成功率 86.54%,预报概括率 91.84%,TS 评分 80.36%。

表 2 SVM 模型对陕西省公路站点大雾预测试验结果

试验结果	渭南	西安	宝鸡	宜君	安康
最优模型中的参数(C)	100.03	100	100.067	100.02	100
最优模型中的参数(g)	0.091	3(u)	0.11	3.11(u)	0.141
建模样本数据中支持向量	379	910	259	399	257
检验样本	190	190	190	190	190
有雾样本	34	23	5	29	49
检验样本的气候概率/%	17.89	12.11	2.63	15.26	25.79
正确分类的样本	183	183	185	172	179
用最优模型分类的正确率/%	96.32	96.23	97.37	90.53	94.21
预报成功率/%	96.55	94.44	0	67.74	86.54
预报概括率/%	82.35	73.91	0	72.41	91.84
TS 评分/%	80	70.83	0	53.85	80.36

3 结语

SVM 方法是一种处理高度非线性问题的新的机器学习方法,具有更好的泛化能力和更高的计算效率。本文是 SVM 方法应用于大雾预报的初步尝试,除宝鸡因大雾的气候概率较低,用最优模型回代检验样本预报成功率较低外,其余站点预报成功率和 TS 评分都较高,SVM 预报模型可以在业务应用中进一步试用。

参考文献

[1] 陈永义,余小鼎,高学浩,等. 处理非线性分类和回归问题的一种新方法(I)——支持向量机方法简介[J]. 应用气象学报, 2004,15(3):345-353.

[2] 冯汉中,陈永义. 处理非线性分类和回归问题的一种新方法(II)——支持向量机方法在天气预报中的应用[J]. 应用气象学报,2004,15(3):355-364.

[3] 常军,李祯,朱业玉,等. 基于支持向量机(SVM)方法的冬季温度预测[J]. 气象科技,2005,33(增刊):100-104.

[4] 高留喜,杨晓霞,郇庆国,等. 支持向量机方法在山东山洪地质灾害预报中的应用试验[J]. 气象科技,2007,35(5):642-645.

[5] 陈永义,冯汉中. 支持向量机回归方法在实时业务预报中的应用[J]. 气象,2006,31(1):41-44.

[6] 贺皓,吕红,徐虹. 陕西省大雾的气候特征[J]. 高原气象,2004, 6(3):407-411.

[7] 贺皓,姜创业,徐旭然. 利用 MM5 模式输出产品制作雾的客观预报[J]. 气象,2002,27(9):41-43.

[8] 贺皓,刘子臣,徐虹,等. 陕西省高等级公路大雾的预报方法研究[J]. 陕西气象,2003,(1):7-10.

Fog Forecast Method Based on SVM and Model Identification

He Hao¹ Luo Hui²

(1 Shaanxi Specialized Meteorological Office, 2 Shaanxi Provincial Meteorological Bureau, Xi'an 710015)

Abstract: By means of Support Vector Machine (SVM) and nine meteorological element data near the surface layer before heavy fog occurring in December and November from 1971 to 2000 (air temperature, precipitation, visibility, wind speed, wind direction, relative humidity, and overall and low cloud cover), a 24-hour heavy fog forecast model for the highway in Shaanxi Province is developed. The model, based on the Gauss kernel function, has put to trial use and gets satisfactory TS scores through simulating and training in fog forecasting.

Key words: Support Vector Machine (SVM), kernel function, classification, fog forecast