

毛宇清, 孙燕, 姜爱军, 等. 2011. 南京地区霾预报方法试验研究 [J]. 气候与环境研究, 16 (3): 273–279. Mao Yuqing, Sun Yan, Jiang Aijun, et al. 2011. An experimental study of haze prediction method in Nanjing [J]. Climatic and Environmental Research (in Chinese), 16 (3): 273–279.

南京地区霾预报方法试验研究

毛宇清¹ 孙燕² 姜爱军^{1, 3} 陈曲¹ 沈澄¹

1 南京市气象局, 南京 210009

2 江苏省气象台, 南京 210008

3 南京信息工程大学大气科学学院, 南京 210044

摘 要 支持向量机 (SVM) 方法是基于统计学理论的一种新的机器学习方法, 对解决小样本条件下的非线性问题非常有效。利用 2004~2007 年南京站的逐日常规观测资料以及同期南京市环境监测点的逐日污染物浓度资料, 使用 SVM 分类和回归方法分别建立了南京地区霾日分类预报模型和有霾日 14 时 (北京时间, 下同) 能见度预报模型。预报试验结果表明: 南京地区霾日的 SVM 分类预报结果, Ts (Threat scores) 评分均在 0.4 以上; 而有霾日 14 时能见度的 SVM 回归预报结果, 按能见度误差范围为 ± 3 km 算, 准确率均达到了 86% 以上; 加入当天 08 时新资料的订正预报模型, 其预报结果优于起始预报模型。二者的预报结果较为满意, 可以给实际业务预测提供参考。

关键词 霾 支持向量机 回归分析和预测

文章编号 1006-9585 (2011) 03-0273-07 **中图分类号** P456.4 **文献标识码** A

An Experimental Study of Haze Prediction Method in Nanjing

MAO Yuqing¹, SUN Yan², JIANG Aijun^{1,3}, CHEN Qu¹, and SHEN Cheng¹

1 *Nanjing Meteorological Bureau, Nanjing 210009*

2 *Jiangsu Meteorological Observatory, Nanjing 210008*

3 *College of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044*

Abstract The Support Vector Machine (SVM) is a new machine learning method based on the statistical learning theory and it is very useful to solve non-linear problems of short time series. Based on the daily observations obtained from the Nanjing meteorological station and the daily measured contamination data obtained from the Nanjing environmental quality monitoring station from 2004 to 2007, prediction models of haze-day classification and visibility at 1400 LST in haze days in Nanjing are built by the SVM method. The results show that the threat scores (Ts) of haze-day classification forecast are all over 0.4 and the precision of visibility at 1400 LST in haze days can reach 86% by considering 3 km as the error bounds. Otherwise, the forecast models which are revised by the new data at 0800 LST of that day are better than the beginning forecast models according to the results. Both SVM forecast models perform satisfactorily and can refer references to real business.

Key words haze, support vector machine, regression analysis and forecast

收稿日期 2009-12-20 收到, 2010-11-10 收到修定稿

资助项目 南京市科技计划项目 201001070, 南京市气象业务技术团队建设项目 NJ201003, 南京社会发展项目 200701100

作者简介 毛宇清, 女, 硕士, 工程师, 主要从事城市气象和短期天气预测研究。E-mail: maoyq1021@163.com

1 引言

影响能见度的天气现象主要有雾、降水、霾、烟、浮尘等。在中国气象局的《地面气象观测规范》中,霾天气被这样定义:“大量极细微的干尘粒等均匀地浮游在空中,使水平能见度小于 10 km 的空气普遍有混浊现象,使远处光亮物微带黄、红色,使黑暗物微带蓝色”。霾现象在我国日趋严重,已经成为一种新型灾害性天气,堪与北方沙尘暴相比(吴兑等,2006)。

南京是长江三角洲城市群三大中心城市之一。研究表明(孙燕等,2007):1970 年代中期至 1990 年代中期,南京地区工业大发展,空气质量下降,霾日急剧增加;1995~2000 年,霾出现的频率略有下降;2000~2005 年,霾日又明显增多。霾天气的出现,不仅使能见度变差,影响海、陆、空交通运输,增加交通事故频次,而且还造成空气质量下降,诱发鼻炎、支气管炎等多种呼吸道疾病,危害人体健康。因此,做好霾预报(包括是否出现霾及霾的强度预报),对建立霾的预警机制等有着重要意义。

目前,国内外学者对霾的判别、组成及气候特征等(Hachfeld and Jurgens, 2000; 段菁春等,2006; 童尧青等,2007; 李崇志等,2009)进行了诸多研究。国内关于城市霾的预报主要有经验预报法、统计预报法和数值预报法。经验预报法依靠掌握霾的变化特征及影响因素,总结出有利于霾天气出现的高空和地面形势,综合判断后进行预报(胡天玉等,2005)。数值预报法则是先通过空气质量数值预报模式计算污染物浓度分布和变化规律,再根据污染物浓度计算能见度(陈训来等,2008; 刘红年等,2009),该方法难度大、计算量大,并不适合单站预报使用。统计预报法是通过分析霾天气产生的气象条件,建立能见度与各种气象要素的统计关系,得到量化的预报方程,该方法简单易行,但由于导致霾天气出现的因素较多,影响因子复杂,目前的研究相对较少,尤其是霾日的分类预报(某天是否出现霾天气现象)尚未有较好的研究结果。

本文所采用的支持向量机(Support Vector Machine, SVM)方法是一种新的处理非线性分

类和回归的机器学习方法,它以统计学理论为基础,对小样本条件下的非线性映射具有优势(Vapnik, 1998; 陈永义等,2004),特别适合在信息量较大,而预报对象和预报因子的关系又不甚清楚的情况下使用(冯汉中等,2006)。本文将利用南京的常规气象观测资料和污染物浓度资料,尝试用 SVM 方法做南京地区霾日的分类预报和有霾日 14 时(北京时间,下同)的能见度预报,以期对实际霾天气的预测有一定的参考意义。

2 资料和方法

2.1 资料

气象资料来自 2004 年 1 月~2008 年 12 月南京国家基准气候站的逐日地面和高空观测资料;污染物浓度资料来自同期南京市 6 个环境质量监测点的 SO_2 、 NO_2 和 PM_{10} (空气动力学当量直径小于等于 $10\ \mu\text{m}$ 的颗粒物)逐日实测资料(单位: $\text{mg}\cdot\text{m}^{-3}$),计算出 6 个监测点的平均值代表南京地区的污染状况。

2.2 SVM 的基本思想

SVM 方法的基本思想是定义最优线性超平面,并把寻找最优线性超平面的算法归结为求解一个凸规划问题,进而基于 Mercer 核展开定理,通过非线性映射 φ ,把样本空间映射到一个高维乃至无穷维的特征空间(Hilbert 空间),使在特征空间中可以应用线性习机的方法解决样本空间中的高度非线性分类和回归等问题。图 1 给出了这一思路的大致框架(毛宇清等,2007)。

2.3 SVM 分类方法

分类问题用数学的形式可描述为:给定一组训练样本 $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)$, 其中 $\mathbf{x}_i \in \mathbf{R}^N$, 为 N 维向量, $y_i \in \{-1, 1\}$ 或 $y_i \in \{1, 2, \dots, k\}$; 另外给出待预报样本的预报因子数据集: $\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \dots, \mathbf{x}_m$, 寻求所对应的 y_i 输出值。当 $y_i \in \{-1, 1\}$ 时为最简单的二类划分,在本文中即为霾日和非霾日两类。

SVM 方法的核心概念是支持向量。所寻求的最优超平面是使两类达到无错分类,并且使参与分类的样本集到超平面的距离达到最大,这时样本点都在两条边界线之外,求最优划分超平面等价于求最大间隔。对于给定的训练样本集,根据

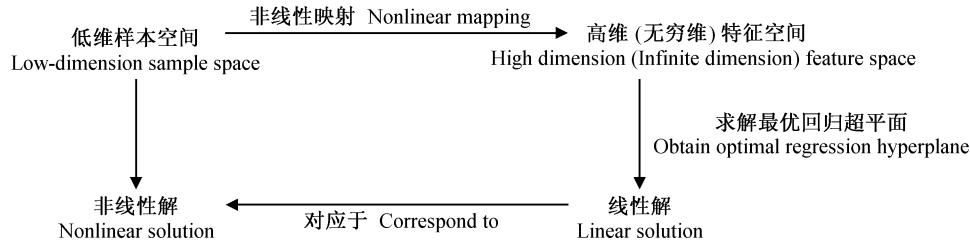


图1 SVM方法的主要思路

Fig. 1 Main idea of the SVM method

相关的理论和算法(陈永义等, 2004), 求得最优分类超平面线性决策函数为:

$$M(\mathbf{x}) = \text{Sgn} \left[\sum_{i=1}^L \alpha_i^* y_i (\mathbf{x} \cdot \mathbf{x}_i) + b^* \right], \quad (1)$$

其中, α_i^* 和 b 为确定最优划分超平面的参数, $(\mathbf{x} \cdot \mathbf{x}_i)$ 为两个向量的内积, L 为支持向量数。

应用核函数的展开定理, 利用非线性映射 φ , 把样本集映射到一个高维(以至于无穷维)的特征空间, 从而使得样本空间中的高度非线性问题在特征空间中应用 SVM 线性回归的方法得以实现。在整个求解过程中并不需要知道非线性映射的显示表达式, 只需将样本空间中的点 $\mathbf{x}$ 和 $\mathbf{x}_i$ 用映射的象 $\varphi(\mathbf{x})$ 和 $\varphi(\mathbf{x}_i)$ 代替, 这大大简化了计算。由此, (1) 式变为:

$$M(\mathbf{x}) = \text{Sgn} \left\{ \sum_{i=1}^L \alpha_i^* y_i [\varphi(\mathbf{x}) \cdot \varphi(\mathbf{x}_i)] + b^* \right\}. \quad (2)$$

根据 Mercer 定理, 可得 SVM 方法最终确定的非线性分类决策函数为:

$$M(\mathbf{x}) = \text{Sgn} \left[\sum_{i=1}^L \alpha_i^* y_i K(\mathbf{x}, \mathbf{x}_i) + b^* \right]. \quad (3)$$

2.4 SVM 回归方法

回归分析可以形式化为: 给定一组训练样本集 $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)$, 其中 $\mathbf{x}_i \in \mathbf{R}^N$, 为 N 维向量, 通常为预报因子值; $y_i \in \mathbf{R}$, 为预报对象值; 给出待预报样本的预报因子数据集: $\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \dots, \mathbf{x}_m$, 寻求与训练样本的输入输出拟合最优的函数关系 $y = f(\mathbf{x})$, 进而求出预报对象 y_i 的输出值。

与 SVM 分类方法类似, SVM 回归方法通过在特征空间中引入 ϵ ——不敏感误差函数, 定义最优线性回归超平面, 求解最优超平面同样等价于求最大间隔。对于给定的训练样本集, 最后求得最优回归超平面非线性决策函数为:

$$f(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b = \sum_{i=1}^L (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b. \quad (4)$$

2.5 核函数的选取

由于构造支持向量机的基础是 Mercer 定理, 作为建立支持向量机的核函数必须以满足 Mercer 定理的条件为前提, 因此本文我们选择下述径向基函数(满足 Mercer 定理条件)作为基函数建立 SVM 回归模型。径向基函数形为:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-r \|\mathbf{x} - \mathbf{x}_i\|^2), \quad (5)$$

则最终分类决策函数形为:

$$M(\mathbf{x}) = \text{Sgn} \left[\sum_{i=1}^L \alpha_i^* y_i \exp(-r \|\mathbf{x} - \mathbf{x}_i\|^2) + b^* \right], \quad (6)$$

最终回归决策函数形为:

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^L (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b \\ &= \sum_{i=1}^L (\alpha_i - \alpha_i^*) \exp(-r \|\mathbf{x} - \mathbf{x}_i\|^2) + b, \end{aligned} \quad (7)$$

其中, L 为支持向量数; $\mathbf{x}_i$ 为作为支持向量的样本因子向量; $\mathbf{x}$ 为待预报因子向量; α_i, α_i^*, b 为建立 SVM 模型待确定的系数, r 为核参数。可见, SVM 的最终决策函数只由少数的支持向量所确定, 落在两条边界线之外的所有样本点对最优超平面没有贡献。模型的复杂程度取决于支持向量的数目和核函数的计算, 而不是样本空间的维数, 从而在某种程度上避免了“维数灾”。

3 建立南京地区霾日 SVM 分类预报模型

霾天气的判定主要以能见度和相对湿度作为

标准,单一的能见度预报或相对湿度预报都不能用来作为预报霾天气是否会出现的依据,而分别对二者进行预报后再结合起来又会增加预报误差,故有必要选择一种直观简洁的方法。本文将首次利用 SVM 分类方法做南京地区霾日的分类预报试验。

3.1 预报对象

由于霾的复杂性和区域性,全国尚没有统一的定量判定指标,在严格执行《地面气象观测规范》的基础上,结合南京的实际情况,本文以地面观测记录为准,在天气现象一栏中有霾的就确定为一次霾记录,任意一天中只要有一次霾记录出现就记为一个霾日。以某天是否是霾日为预报对象,“是”记为“+1”,“否”记为“-1”,进行霾天气的分类预报研究。

3.2 预报因子的选择

SVM 预报模型的建立同因子与预报对象是否线性相关无明显的依赖关系,SVM 能通过高度非线性的特征映射来寻求支持向量,进而表征预报对象的演变趋势,因此因子的选择更注重因子与对象间的物理意义或天气学意义分析,及实际预报中积累的经验因子的引用(冯汉中和陈永义,2004)。霾天气产生的条件主要有两类,一是有利于霾天气形成的气象条件,二是作为霾天气形成的物质基础即大气污染物,因此霾预报需要综合考虑与这两方面有关的预报因子。

和雾形成的气象条件类似,有利于霾天气产生的地面形势是气压梯度小,水平风速弱;在垂直方向上,大气层结相对稳定,近地层存在逆温;混合层高度较低,不利于污染物的扩散。二者最大的区别在于霾天气的地面水平能见度比雾大,相对湿度比雾小。由于本文使用的是常规观测资料,针对上述特点,在地面资料中选取温度、气压、相对湿度、露点、风速作为预报因子,在探空资料中选取 1000、925、850、700、500 hPa 各层次的位势高度、温度、露点、风速作为预报因子。污染物预报因子采用的是 SO_2 、 NO_2 和 PM_{10} 的浓度。需要指出的是,由于 SVM 方法是通过支持向量构建预报模型,因此对预报因子的个数没有明显的限制(冯汉中和陈永义,2004)。

另外,考虑到在前一天 20 时至当天 08 时期间,受到天气系统的影响,夜间形成的接地逆温

未被破坏或者早晨可能会出现辐射雾,导致大气污染物浓度升高,对能见度等造成影响,在实际业务预测中,由于实测资料的更新,对同一个预报对象常要建立 2 个模型,分别做起始预报和订正预报。因此,本文利用上述要素构建了 2 组预报因子(均为 33 个),根据资料时间的不同反应出各要素的动态变化,最后用 SVM 分类方法分别建立 2 个预报模型。模型和因子的具体配置如表 1。

表 1 预报因子和预报模型

Table 1 Forecast factors and prediction models

预报模型	预报因子	预报因子个数
起始预报(样本一/模型一)	前一天 14 时的地面资料、污染物浓度资料;前一天 20 时的地面资料、探空资料	共 33 个
订正预报(样本二/模型二)	前一天 20 时的地面资料;当天 08 时的地面资料、探空资料、污染物浓度资料	共 33 个

3.3 构建样本序列

用上述预报因子构建样本序列,剔除缺测和有问题的样本资料后,样本一的有效样本为 1762 个,样本二的有效样本为 1772 个。由于本文采用中国气象局培训中心开发的 CMSVM 软件平台建立模型,它对数据的文件格式有一定要求。首先要对各个因子进行归一化处理,这有利于避免各个因子间的量级差异,使每一个因子的数据落入区间 $[0, 1]$,归一化方法为: $(z - z_{\min}) / (z_{\max} - z_{\min})$,其中 $z_{\max}$ 、 $z_{\min}$ 分别为因子 z 的最大值和最小值。

用 2004~2007 年的样本资料训练最优分类预报模型,分别按照 75%、20%和 5%的比例构建训练集、实验集和检验集。建立检验集是为了将最终确立的 SVM 预报模型对其做预报,以检验模型的预报效果(推广能力),检验集的数据不参与训练学习及参数筛选等建模过程。用 2008 年的资料构建预报集,即用最优分类模型对 2008 年的霾日情况进行预报。

在样本一和样本二的训练集中,正样本即霾日样本的比例分别为 31.5%和 31.3%,而在预报集中,霾日样本的比例分别为 43.4%和 43.7%,可见 2008 年南京地区的霾日数较 2004~2007 年

有一定程度的增加。

3.4 SVM 分类预报模型及结果检验

用逐步逼近法最优化参数，得到南京地区霾日的 SVM 分类预报模型，最后通过 CMSVM 预报程序计算得出结果（如表 2），预报正确率是分类成功的样本数与所有样本数的比值，预报成功率是预报为正类且确为正类的样本数与预报为正类的所有样本数的比值，Ts（Threat scores）评分是预报为正类且实况为正类的样本数与正类样本数和负类中被预报为正类的样本数之和的比值。可以看出，由 SVM 方法建立的南京地区霾日分类预报模型，Ts 评分均在 0.4 以上，最大达到 0.58，预报结果较为满意。使用了当天 08 时新资料的订正预报模型（模型二），其预报正确率、成功率和 Ts 评分均高于起始预报模型（模型一）。从对 2008 年南京地区霾日（预报集）预报的结果来看，模型一漏报的次数相对较多，模型二空报的次数相对较多。另外，两组模型所用的支持向量分别占总样本数的 71.4% 和 65.2%，SVM 方法只有部分关键样本参与模型的构建，其他样本对模型的建立没有贡献。

4 建立南京地区有霾日 14 时能见度 SVM 回归预报模型

以上我们进行了南京地区霾日分类预报试验，但霾预报的最终目的还是为了要做霾的强度预报，大气的水平能见度是判断霾天气强度的主要依据，下面我们将用 SVM 回归方法做有霾日 14 时的能见度预报。

4.1 预报对象和预报因子的选择

孙燕等（2007）的研究表明，南京地区各级霾在一天当中以 11 时、14 时和 17 时出现最多，总计占霾出现总数的 77.5%，02 时、05 时和 08 时最少，仅占 5.4%，而中度霾在 14 时出现的最多。这主要是由于中午前后近地面的相对湿度条件比较有利。因此，本文选择霾日 14 时的能见度（单位： 10^{-1} km）作为 SVM 回归模型的预报对象。

由于能见度是霾天气判定的主要依据之一，因此有霾日的能见度预报因子也要综合考虑气象条件和大气污染状况，预报因子的选择同分类预报（见表 1），同样有两组样本，分别建立两个预报模型。

4.2 构建样本序列

剔除掉无霾日、缺测和有问题的样本后，两组资料分别获得有效样本 397 个和 396 个。用 2004~2007 年的样本资料训练最优回归模型，由于资料长度有限，未另安排检验集，用实验集代替。训练集和实验集（检验集）的比例分别约为 85% 和 15%。仍用 2008 年的资料构建预报集，即用最优回归模型对 2008 年有霾日的 14 时能见度大小进行预测。

4.3 SVM 回归预报模型及结果检验

表 3 为 SVM 最优回归模型对南京地区有霾日的 14 时能见度预报结果。从对检验集和预报集的预报可以看出，按能见度误差范围为 ± 3 km 算，准确率均达到了 86% 以上，若按误差范围为 ± 2 km 算，准确率也均达到了 68% 以上。和分类预报相同，使用了当天 08 时新资料所建立的模型二，其预报结果普遍好于模型一。

表 2 SVM 对南京地区霾日的分类预报
Table 2 Predictions for classification of haze days in Nanjing by the SVM

	最优分类模型参数		漏报次数		空报次数		预报正确率		预报成功率		Ts 评分	
	<i>c</i>	<i>g</i>	检验集	预报集	检验集	预报集	检验集	预报集	检验集	预报集	检验集	预报集
模型一	300	0.05	10	69	14	54	64.29%	66.21%	61.54%	62.24%	0.48	0.42
模型二	140	0.3	6	48	13	60	71.83%	70.33%	66.67%	64.91%	0.58	0.51

表 3 SVM 对南京地区有霾日的 14 时能见度预报
Table 3 Predictions for visibility at 1400 LST in haze days in Nanjing by the SVM

	最优回归模型参数			绝对差/ 10^{-1} km		均方差/ 10^{-1} km		准确率（按 3 km 误差算）		准确率（按 2 km 误差算）	
	<i>c</i>	<i>g</i>	<i>w</i>	检验集	预报集	检验集	预报集	检验集	预报集	检验集	预报集
模型一	3	2.2	0.1	15.89	15.32	18.65	18.20	86.67%	91.89%	68.44%	73.27%
模型二	200	0.01	1	15.12	14.57	18.74	17.38	88.63%	91.18%	70.45%	77.45%

虽然考虑了一定的人工观测误差后, SVM 方法对南京地区有霾日 14 时的能见度预报准确率较为满意, 但观察发现, 预报值主要集中在 40~80 (10^{-1} km) 之间, 这说明最优回归模型对极值尤其是对重度霾和严重霾天气的预报能力较弱。分析误差产生的原因, 一方面在于训练样本长度较短, 不能使预报模型包含比较完备的信息, 另外, 各季节的样本数在训练集、检验集和预报集中的比例分配不均也有可能是原因之一, 如在预报集中 (2008 年) 夏季样本的比例约占 31%, 而在训练集和检验集 (2004~2007 年) 中仅为约 17%。因此, 在实际的业务预测中需要不断增加新的样本参与训练, 才能不断改善预报模型。

4.4 与逐步回归方法的结果对比

利用相同的资料, 用逐步回归方法建立南京地区有霾日 14 时的能见度预报模型如下, 模型一:

$$y = 73.4864 + 2.01x_5 - 0.20719x_8 - 38.8786x_{13} + 0.1674x_{20}, \quad (8)$$

其中, x_5 表示前一天 14 时的地面风速, x_8 表示前一天 20 时地面相对湿度, x_{13} 表示前一天 14 时的 PM10 浓度, x_{20} 表示前一天 20 时的 925 hPa 露点; 模型二:

$$y = 61.8591 - 1.6537x_{10} - 85.6671x_{13} + 1.6007x_{15} - 1.0281x_{16} - 0.26054x_{28}, \quad (9)$$

其中, x_{10} 表示当天 08 时的地面风速, x_{13} 表示当天 08 时的 PM10 浓度, x_{15} 表示当天 08 时的 1000 hPa 温度, x_{16} 表示当天 08 时的 1000 hPa 露点, x_{28} 表示当天 08 时 700 hPa 露点。

预报方程均通过了显著性检验。可见, 对霾天气时能见度有显著贡献的气象因子主要是地面风速、地面相对湿度以及对流层中低层的温度和露点。而在污染物因子中, PM10 的浓度对能见度的影响最大, 这也证实了细粒子污染是影响大气能见度的主要因素之一, 特别是 PM2.5 和 PM10 对霾的形成影响更大, 但由于我们目前缺乏这方面的资料, 在预报方程中无法体现。

将 2 个回归模型分别对预报集进行预报, 按能见度误差范围为 ± 3 km 算, 准确率模型一为 92.14% 和 91.56%, 与 SVM 方法相当, 但若按能见度误差范围为 ± 2 km 算, 准确率均在 70% 以下。可见, 用 SVM 回归方法建立的南京地区有霾

日 14 时的能见度预报模型, 其预报结果优于逐步回归方法, 对实际业务预测有一定的参考价值。

5 小结

SVM 方法通过核函数实现从样本空间到高维特征空间的非线性映射, 通过支持向量来刻画预报因子与预报对象之间的非线性关系, 从而解决本质上的非线性问题。模型的建立对预报因子的个数没有明显的限制, 计算的复杂程度只取决于支持向量的个数而不是样本空间的维数, 只有部分关键样本才对最优模型有贡献。

试验结果表明, 南京地区霾日的 SVM 分类预报模型, Ts 评分均在 0.4 以上, 最大达到 0.58; 而 SVM 回归模型对有霾日的 14 时能见度预报结果显示, 按能见度误差范围为 ± 3 km 算, 准确率均达到了 86% 以上, 若按误差范围为 ± 2 km 算, 准确率也均达到了 68% 以上, 优于逐步回归方法; 使用了当天 08 时新资料的订正预报模型 (模型二), 其预报结果好于起始预报模型 (模型一)。这些结果均表明, 利用支持向量机方法所建立的南京地区霾预报模型, 预报结果较为满意, 可以给实际业务提供参考。

然而, 本文的工作也存在一些不足之处。如在本文中预报因子的选择仅注重了天气学意义分析, 而物理意义不够明晰, 同时, 由于形成霾天气的原因尚不是十分清晰, 这也给预报因子的选择带来一定的难度。另外, 虽然 SVM 方法对预报因子的个数没有明显的限制, 但引入分辨率不强的因子有增加噪声的可能, 对模型的预报效果造成影响。在运用 SVM 方法进行预报时可以通过引入数值预报产品中对预报对象有明确物理意义的因子, 剔除一些意义不大的因子, 从而改善预报效果。

参考文献 (References)

- 陈永义, 俞小鼎, 高学浩, 等. 2004. 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, 15 (3): 345 - 354. Chen Yongyi, Yu Xiaoding, Gao Xuehao, et al. 2004. A new method for non-linear classify and non-linear regression I: Introduction to support vector machine [J]. Journal of Applied Meteorological Science (in Chinese),

- 15 (3): 345-354.
- 陈训来, 冯业荣, 范绍佳, 等. 2008. 离岸型背景风和海陆风对珠江三角洲地区灰霾天气的影响 [J]. 大气科学, 32 (3): 530-542. Chen Xunlai, Feng Yerong, Fan Shaojia, et al. 2008. Effect of the off-shore background flow and sea-land breezes on haze weather over the Pearl River Delta region [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 32 (3): 530-542.
- 段菁春, 毕新惠, 谭吉华, 等. 2006. 广州灰霾期大气颗粒物中多环芳烃粒径的分布 [J]. 中国环境科学, 26 (1): 6-10. Duan Jingchun, Bi Xinhui, Tan Jihua, et al. 2006. The particle diameter distribution of polycyclic aromatic hydrocarbons (PAHs) in atmospheric particle during haze period in Guangzhou [J]. China Environmental Science (in Chinese), 26 (1): 6-10.
- 冯汉中, 陈永义. 2004. 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, 15 (3): 355-365. Feng Hanzhong, Chen Yongyi. 2004. A new method for non-linear classify and non-linear regression II: Application of support vector machine to weather forecast [J]. Journal of Applied Meteorological Science (in Chinese), 15 (3): 355-365.
- 冯汉中, 陈永义, 成永勤, 等. 2006. 双流机场低能见度天气预报方法研究 [J]. 应用气象学报, 17 (1): 94-99. Feng Hanzhong, Chen Yongyi, Cheng Yongqin, et al. 2006. A study on the forecast method of the low visibility weather of Shuangliu Airport [J]. Journal of Applied Meteorological Science (in Chinese), 17 (1): 94-99.
- Hachfeld B, Jurgens N. 2000. Climate patterns and their impact on the vegetation in a fog driven desert: The Central Namib Desert in Namibia [J]. Phytocoenologia, 30 (3-4): 567-589.
- 胡天玉, 杨显宇, 谢巨伦. 2005. 雷州半岛霾的初步分析与预报 [J]. 广东气象, 12 (1): 21-22. Hu Tianyu, Yang Xianyu, Xie Julun. 2005. Preliminary analysis and prediction of haze on Leizhou Peninsula [J]. Guangdong Meteorology (in Chinese), 12 (1): 21-22.
- 李崇志, 于清平, 陈彦. 2009. 霾的判别方法探讨 [J]. 南京气象学院学报, 32 (2): 327-332. Li Chongzhi, Yu Qingping, Chen Yan. 2009. Discussion of haze distinction method [J]. Journal of Nanjing Institute of Meteorology (in Chinese), 32 (2): 327-332.
- 刘红年, 胡荣章, 张美根. 2009. 城市灰霾数值预报模式的建立与运用 [J]. 环境科学研究, 22 (6): 631-636. Liu Hongnian, Hu Rongzhang, Zhang Meigen. 2009. Development and application of urban haze numerical forecast model [J]. Research of Environmental Sciences (in Chinese), 22 (6): 631-636.
- 毛宇清, 王咏青, 王革丽. 2007. 支持向量机方法应用于理想时间序列的预测研究 [J]. 气候与环境研究, 12 (5): 676-682. Mao Yuqing, Wang Yongqing, Wang Geli. 2007. An application on prediction and analysis for ideal time series based on the SVM method [J]. Climatic and Environmental Research (in Chinese), 12 (5): 676-682.
- 孙燕, 魏建苏, 俞剑蔚. 2007. 南京灰霾天气的气候特征分析 [J]. 中国人口资源与环境, 17 (5): 179-182. Sun Yan, Wei Jiansu, Yu Jianwei. 2007. Analysis of climate characters of haze weather in Nanjing [J]. China Population, Resources and Environment (in Chinese), 17 (5): 179-182.
- 童尧青, 银燕, 钱凌, 等. 2007. 南京地区霾天气特征分析 [J]. 中国环境科学, 27 (5): 584-588. Tong Yaoqing, Yin Yan, Qian Ling, et al. 2007. Analysis of the characteristics of hazy phenomena in Nanjing area [J]. China Environmental Science (in Chinese), 27 (5): 584-588.
- Vapnik V N. 1998. Statistical Learning Theory [M]. New York: John Wiley & Sons, Inc., 733pp.
- 吴兑, 毕雪岩, 邓雪娇, 等. 2006. 珠江三角洲气溶胶云造成的严重灰霾天气 [J]. 自然灾害学报, 15 (6): 77-83. Wu Dui, Bi Xueyan, Deng Xuejiao, et al. 2006. Heavy haze weather resulting from aerosol cloud over Pearl River Delta [J]. Journal of Nature Disasters (in Chinese), 15 (6): 77-83.