

文章编号:1004-9045(2008)03-0264-04

SVM 方法在武汉市大雾预警预报中的应用

李才媛, 韦惠红, 邓 红

(武汉中心气象台, 武汉 430074)

摘要: 在简单介绍支持向量机(SVM)基本原理的基础上, 根据 2003—2006 年武汉市探空和地面气象要素资料, 使用支持向量机方法(SVM)和 CMSVM 应用软件平台, 通过对训练样本进行交叉验证和模型核参数的逐渐逼近, 建立了武汉市 0~24 h 大雾支持向量机预报模型。经过 2007 年试报, 该模型的 T_s 评分为 62.5%; 同时, 通过对 2007 年武汉市大雾特点和预警信号发布情况的分析发现, 该模型所选取的建模时段和设计的输出产品时间均比较合理, 能够较及时地为大雾预警预报提供参考依据。

关键词: 武汉市; 大雾; 预警预报; 支持向量机

中图分类号: P456 文献标识码: A

1 引言

支持向量机(Support Vector Machines, 简称 SVM)是近年来发展起来的一种机器学习方法。陈永义等^[1-2]根据 SVM 方法的原理, 建立了 CMSVM 应用平台, 并在气象预报领域进行了一些探讨性试验。SVM 使用了统计学习理论中的结构风险最小化原则, 因而具有良好的泛化能力, 在处理非线性特征的气象要素或天气现象的预报方面优势明显。目前, SVM 方法已经在降水、温度、冰雹、天空云量、气候趋势等方面得到初步成功的运用^[3-7]。

雾是影响大气水平能见度的一种常见天气现象, 其水平能见度在 1.0 km 以下的雾统称大雾。随着我国社会经济快速发展, 大雾对交通运输安全的影响越来越大, 造成的经济损失也越来越严重, 各级气象部门已将其作为一种主要的灾害性天气对外公开发布预警信号。因此, 提高大雾预报水平已成为天气预报业务的重点之一。由于雾是典型的小概率事件和离散事件, 其地理、季节和日变化特征较为明显, 因此雾的预报一直存在很大难度。雾的预报主要有天气学模型、经验指标、诊断分析、统计回归等方法^[8-10]。冯汉中等^[11]曾应用 SVM 方法对机场能见度预报进行过试验, 证明了该方法在小概率事件中具有特定优势。

武汉市大雾监测预报是城市防灾减灾应急气象服务的重要内容之一。为了提高武汉市大雾的预报水平, 本文利用 2003—2006 年该市探空和地面气象要素资料, 结合预报员经验, 尝试 SVM 方法, 建立了武汉市 0~24 h 大雾支持向量机预报模型(以下简称武汉市

大雾 SVM 预报模型); 经过 2007 年试报, 其 T_s 评分为 62.5%, 能够为大雾预警预报提供一些参考依据。

2 支持向量机(SVM)的基本原理

SVM 方法^[1]可概括地表述为: 给定训练样本 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$, 其中 $x_i \in R^N$, 为 N 维向量, $y_i \in \{-1, 1\}$ 或 $y_i \in \{1, 2, \dots, k\}$, 给出数据集: $x_{i+1}, x_{i+2}, \dots, x_m$ 通过训练学习建立分类模式 $M(x)$, 使其不但对训练样本能够正确分类, 而且具有较强的推广能力。即对于输入的数据 x_i , 通过模式运算得到正确的对应输出预报值 y_i 。

对于训练样本集的线性二类划分问题, 就是寻求函数

$$y=f(x)=\text{Sgn}[(w \cdot x)+b] \quad (1)$$

使对于 $i=1, 2, \dots, l$ 满足条件

$$y_i=f(x_i)=\text{Sgn}[(w \cdot x_i)+b] \quad (2)$$

其中 $w, x, x_i \in R^N, b \in R, w, b$ 为待确定的参数, Sgn 为符号函数。显然 $(w \cdot x)+b=0$ 为划分超平面, w 为其法方向向量。

对于线性可分离的问题, 满足条件如式(1)的线性决策函数不是唯一的。图 1 给出二维情况下满足条件的划分直线的分布区域图(图中 \bullet 分别代表不同的样本点)。落在虚线区域内的任一直线都可作为决策函数。哪一个是最优的决策函数, 就要对其进行判断。

Vapnik V N 提出一个间隔最大化原则^[1]。所谓间隔最大化原则是指寻求使间隔达到最大的划分为最优, 即对 w, b 寻优, 求得

$$\text{Max}[\text{Min}(\|x - x_i\| : x \in R^N, (w \cdot x)+b=0, i=1, \dots, l)]$$

对应最大间隔的划分超平面 (称为最优划分超平

收稿日期: 2008-06-11; 定稿日期: 2008-07-28

基金项目: 湖北省科技攻关项目(2006AA301B45)资助

作者简介: 李才媛, 女, 1956 年生, 正研级高级工程师, 主要从事业务天气预报工作。E-mail: lcy-whqxt@sina.com

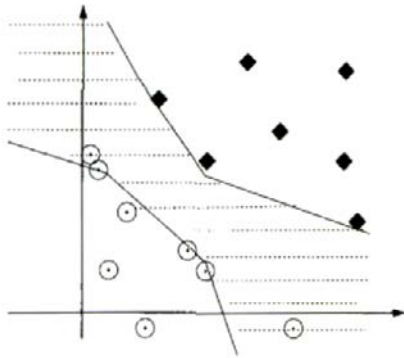


图 1 二维情况下满足条件的划分直线的分布区域图

面), 简称最优超平面, 如图 2 中的 L(图中两条平行虚线 l_1, l_2 为边界, 其距离的一半就是最大间隔)。

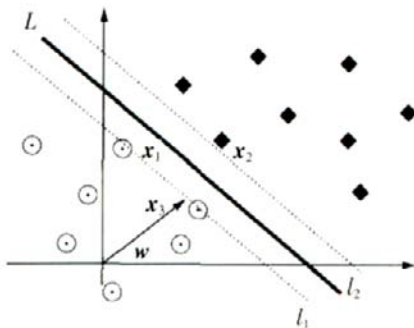


图 2 最优划分超平面示意图

最大间隔和最优超平面只由落在边界上的样本点完全确定, 而不依赖于所有点, 这样的样本点为支持向量, 如图 2 中的 x_1, x_2, x_3 样本点。

对于给定的训练样本集, 根据相关的理论和算法, 最终获得的线性支持向量机为

$$M(x) = \text{Sgn}[(w^* \cdot x) + b^*] = \text{Sgn}[\sum_{SV} \alpha_i y_i (x \cdot x_i) + b^*] \quad (3)$$

其中 w^*, b^* 为确定最优划分超平面的参数; $(x \cdot x_i)$ 为两个向量的点积, SV 为支持向量。

对于线性不可分的情况, 通过非线性映射把样本集映射入一个高维乃至无穷维的特征空间, 使在样本空间中高度非线性问题在高维空间中应用线性分类的方法得以实现。

由于在特征空间中采用的是线性分类方法, 所以在特征空间中的最优划分超平面分类函数的形式为

$$M(x) = \text{Sgn}[(w^* \cdot (x)) + b^*] \\ = \text{Sgn}[\sum_{i=1}^l \alpha_i y_i [(x) \cdot (x_i)] + b^*] \quad (4)$$

与(3)式相比, 这里只是用 $(x), (x_i)$ 代替了 x 和 x_i 。

根据 Mercer 定理, 式(4)最终转变为

$$M(x) = \text{Sgn}[(w^* \cdot \varphi(x)) + b^*] = \text{Sgn}[\sum_{SV} \alpha_i y_i K(x, x_i) + b^*] \quad (5)$$

式(5)就是 SVM 方法确定的最终非线性分类的决策函数。与(式 3)相比, 这里只是用 Mercer 核函数的计算代替了点积计算, 在整个求解过程中不需知道非线性映

射的显式表达式, 而是通过支持向量来表达预报因子与预报对象的关系。

3 武汉市大雾 SVM 预报模型的建立

3.1 武汉市大雾气候变化特征

周月华等^[12]通过分析 1951—2002 年历史气象资料, 揭示了武汉地区大雾的时空变化特征。李才媛等^[13]又分析了 1995—2006 年武汉市大雾的变化特征, 结果表明, 武汉市冬季大雾出现频率最高, 春、秋季次之, 夏季极少出现大雾; 大雾生成时间主要集中在 03—07 时, 其生成最大频率出现在 06 时左右。

2003—2006 年武汉市共出现大雾 33 次 (汉口测站资料), 其中冬季(12 月至次年 2 月)出现 17 次, 占全年大雾总次数的 51.5%; 春季(3—5 月)和秋季(9—11 月)共出现 14 次, 占 42.4%; 夏季(6—8 月)仅出现 2 次, 占 6.1%; 5 月、7—10 月无大雾出现记录。

3.2 挑选预报因子

由于 SVM 的最终决策函数只由支持向量所确定, 其复杂程度取决于支持向量的数目, 而不是样本空间的维数(因子数), 所以选取较多的因子可以更充分地表述大气的空间变化。针对大雾发生的基本天气气候条件^[13], 同时结合预报员的预报经验, 利用武汉市探空和地面气象要素资料, 构造了 100 多个预报因子, 希望能够较充分地描述因子群与预报对象(武汉市大雾)的关系。资料长度为 2003—2006 年, 具体资料包括: (1)武汉测站 08 时和 20 时的有关探空资料, 即 500、700、850、400、925 hPa 的高度、温度、露点、风向和风速; (2)武汉测站地面常规气象资料, 即 08、11、14、17、20 时的风速, 相对湿度, 温度, 24 小时变压, 24 小时变温和 24 小时变湿等。

3.3 确立预报对象和预报模型

预报对象为武汉市未来 0-24 h 有无大雾。若有大雾发生, 归为 +1; 无大雾发生归为 -1。由于 2003—2006 年武汉市大雾正样本少, 分季建模困难, 因此选择大雾出现相对比较集中的 11 月至次年 4 月(共有 31 个正样本)建模。采用中国气象局培训中心 SVM 应用研究小组开发的 CMSVM 应用软件, 建立 SVM 预报模型。建模时对所有样本进行归一化处理。由于构造支持向量机的基础是 Merer 定理, 作为建立支持向量机的核函数, 必须以满足 Merer 定理的条件为前提, 故本文以径向基函数作为基本函数建立 SVM 模型。径向基函数形为

$$K(x, x_i) = \exp(-r \|x - x_i\|^2) \quad (6)$$

在预报中, 基于高斯核通过训练学习后求得的决策函数为

$$M(x)=\text{Sgn}[\sum \alpha_i y_i K(x, x_i)+b]$$

$$=\text{Sgn}[\sum \alpha_i y_i \exp(-r\|x-x_i\|^2)+b] \quad (7)$$

其中 x_i 为支持向量的样本因子向量, x 为待预报因子向量, a, a^*, b 为建立 SVM 模型待确定的系数, r 为核参数, 求和运算只对支持向量进行。

3.4 选取预报模型参数

把处理好的样本数据集(资料时段为 2003—2006 年)分为训练样本、测试样本、检验样本三个部分。通过训练样本, 建立模型, 再利用测试样本不断优化模型, 建立预报模型后对检验样本进行试报。使用不同的训练样本、不同的核参数, 所建立的 SVM 预报模型的预报能力各不相同。在 CMSVM 应用平台, 抽取 30% 的训练样本, 选取常用的径向核函数, 通过调整核参数 g 和惩罚系数 c , 随机分别进行不同的 g, c 参数的逐渐逼近和测试; 再通过改变训练样本比例, 进行不同的交叉试验; 最终对武汉市大雾选择最优的 SVM 预报模型使其对于样本数据集符合率最大, 且具有较高的推广能力。

4 武汉市大雾 SVM 预报模型的效果检验

4.1 正样本 T_s 评分

武汉市大雾 SVM 预报模型于 2007 年在武汉中心气象台投入试运行。该模型在武汉市大雾可能出现的 1—4 月、11—12 月每天 21 时之后输出预报结果, 以便为预报员发布该市当晚到次日大雾预警预报提供参考依据。

为了检验武汉市大雾 SVM 预报模型的预报效果, 对大雾预报准确率采用正样本 T_s 评分。正样本 T_s 评分的计算式为

$$T_s = N_A / (N_A + N_B + N_C) \times 100\% \quad (8)$$

式(8)中, N_A 为预报正确次数, N_B 为空报次数, N_C 为漏报次数。若模型运行期间, 没有预报大雾, 也没有出现大雾, 不参加质量评定, T_s 表明模型对武汉市大雾预报的成绩。

2007 年, 武汉市共出现了 10 次大雾过程, 该模型均作出了正确预报, 未出现漏报, 空报 6 次, T_s 评分为 62.5%。该模型在出现大雾空报时, 武汉市多数情况下出现了轻雾, 如当年 12 月 22 日、25 日早晨, 武汉市及其周边县市出现了大范围轻雾。

4.2 模型的预警服务效果

2007 年武汉市的 10 次大雾过程, 均出现在雨后, 即大雾出现前 0~48 h 有降水, 且武汉周边地区上空大气稳定、水汽充沛、早晚温差大(辐射降温明显)。表 1 给出了 2007 年武汉市 10 次大雾的生成、消散时间及其它特征量。

表 1 2007 年武汉市全年 10 次大雾的生消时间及其它特征

序号	出现月份	生成时间/(日:时:分)	消散时间/(日:时:分)	持续时间/min	最小能见度/m
01	1	16:22:37	17:07:10	513	400
02	1	24:07:26	24:10:22	176	100
03	1	25:06:49	25:10:27	218	200
04	2	10:06:37	10:12:07	330	50
05	2	22:07:12	22:09:48	156	200
06	2	23:04:30	23:08:19	229	200
07	3	24:05:35	24:09:32	237	50
08	3	25:05:42	25:08:42	180	100
09	12	7:05:40	7:12:17	397	100
10	12	19:08:28	19:11:16	168	100

从表 1 中可看到, 2007 年武汉市的 10 次大雾, 除 1 月 16 日大雾生成在夜间 22 时之外, 其余均生成在早上, 且集中出现在 04—07 时(占 60%); 在 12 时之前消散的大雾次数占 80%, 在 08—11 时消散的占 60%; 大雾持续时间平均为 260.4 min, 1 月 16 日前半夜(20—24 时)生成的大雾持续时间最长, 达 513 min。以上分析结果表明, 2007 年武汉市大雾时空分布特征与近年来分析得到的有关武汉市大雾气候特征基本一致^[12-13]。

2007 年武汉中心气象台共发布气象灾害预警信号 79 次, 其中发布大雾预警信号 37 次, 占气象灾害预警信号发布总次数的 46.8%; 冬季(1、2、12 月)发布大雾预警信号 30 次, 占大雾预警信号发布总次数的 81.1%, 在 21 时以后发布大雾预警信号的次数占 90%。可见, 大雾是武汉市发生相对较为频繁的一种气象灾害, 尤其在冬季, 为大雾发生集中期, 其预警服务显得更为重要。通过对 2007 年武汉市大雾特点和预警信号发布情况的分析发现, 武汉市大雾 SVM 预报模型所选取的建模时段和设计的输出产品时间均比较合理, 能够较及时地为大雾预警预报提供参考依据。

5 结论与讨论

利用武汉市探空和地面气象要素资料, 根据支持向量机(SVM)的原理, 以武汉市未来 0~24 h 有无大雾作为预报对象, 建立了武汉市大雾 SVM 预报模型。该模型对 2007 年武汉市出现的 10 次大雾均做出了准确预报, 无漏报, 空报 6 次, T_s 评分为 62.5%; 经试用检验, 其预报服务效果较好。

由于 SVM 方法是通过支持向量来刻画预报因子与预报对象之间的关系, 是基于机器学习的“黑箱”输出最后预报结论, 且对预报因子的数量没有特别要求。因此, 对于大雾这种小概率、非线性、预报难度大的天气现象, SVM 方法有一定的优势, 在 2007 年武汉

市大雾预报应用试验中已得到证明。然而,大雾预报准确率的提高,仍需要在预报服务实践中不断总结经验,充实预报因子,优化预报模型,使 SVM 方法的优势得到进一步发挥。

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)——支持向量机简介[J].应用气象学报,2004,15(2):345-354.
- [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(2):355-365.
- [3] 王建生,熊秋芬.支持向量机方法在单站降水预报中的应用探讨[J].暴雨灾害,2007,26(2):159-162.
- [4] 陈晓艳,赵玉金,孙文英,等.支持向量机方法作温度预报试验[J].贵州气象,2006,30(1):31-33.
- [5] 吴爱敏,郭江勇,张洪芬,等.支持向量机在冰雹预报中的应用[J].干旱气象,2005,23(4):41-45.
- [6] 熊秋芬,胡江林,陈永义.天空云量预报及支持向量机和神经网络方法比较研究[J].热带气象学报,2007,23(3):255-260.
- [7] 张礼平,陈永义,周筱兰.支持向量机(SVM)及其在预测中的应用[J].热带气象学报,2006,22(3):278-282.
- [8] 吴洪,柳崇键,邵洁,等.北京地区大雾形成的分析和预报[J].应用气象学报,2000,11(1):124-127.
- [9] 张燕光.华南沿海春季大雾的分析和预报[J].气象,1999,25(2):53-55.
- [10] 陈晓红,方冲.安徽省省级大雾预报业务系统[J].气象,2005,31(2):61-64.
- [11] 冯汉中,陈永义,成永勤,等.双流机场低能见度天气预报方法研究[J].应用气象学报,2006,17(1):94-99.
- [12] 周月华,王海军,吴义城.增暖背景下武汉地区雾的变化特征分析[J].气象科技,2005,33(6):509-511.
- [13] 李才媛,韦惠红,王东阡.近 10 年武汉市大雾变化特征及 2006 年一次大雾个例分析[J].暴雨灾害,2007,26(3):241-245.

Application of SVM Method in Warning and Forecasting of heavy fog in Wuhan City

LI Cai-yuan, WEI Hui-hong, Deng Hong

(Wuhan Central Meteorological Observatory, Wuhan 430074)

Abstract: Based on the fundamental rules of the Support Vector Machine (SVM) and the conventional observation data of Wuhan from 2003 to 2006, the cross-checking of train samples and the gradual approaching of core parameters of model are made to establish the 0-24 hour's forecasting model of the heavy fog of Wuhan city with SVM method and the application software platform of Chinese Meteorological Support Vector Machine (CMSVM). The T_s score of this model is 62.5% during its initial forecasting in 2007. The features and warning information issuing of the heavy fog in Wuhan in 2007 are analyzed to show that both the model-building date design and the time of output are proved to be reliable to provide the heavy fog warning and forecasting in good time.

Key words: Wuhan city; Heavy fog; Warning and forecasting; Support Vector Machine

今年 7 月湖北省襄樊市出现近 50 年来特大暴雨

2008 年 7 月 22 日 02—20 时,地处鄂西北的襄樊市出现了一次特大暴雨过程,其日降水量突破了自 1959 年该市有气象观测记录以来的最大值。此次暴雨过程的特点是降水强度大、时段集中、持续时间长,给当地农业生产、道路交通等造成严重灾害。

强降水始于 22 日 02 时 50 分,当日该市普降暴雨或大暴雨,其中襄樊市区、襄阳城关、局部乡镇降了特大暴雨。全市人工观测站资料显示,当天 02—20 时 18 小时降水量,襄樊市所辖各县(市)均在 90 mm 以上,其中,襄樊市区降水量达 293.9 mm,襄阳城关超过 300 mm。

另据加密自动气象站雨量资料,上述 18 小时雨量,襄樊市 98 个自动站,有 93 站超过 50 mm,50 站超过 100 mm,6 站超过 200 mm,3 站超过 250 mm。此次过程最大雨量(301.4 mm)出现在襄阳城关;最大 6 h 雨强为 222.0 mm,06-12 时出现在襄阳城关;最大 3 h 雨强为 136.2 mm,09-12 时出现在襄阳城关;最大 1 h 雨强为 61.6mm,09-10 时出现在襄阳城关,强降水中心位于襄阳城关及襄樊市区。

此次特大暴雨过程造成直接经济损失达 19.1 亿元,其中农业直接经济损失 10.9 亿元。全市 102 个乡镇 208.14 万人受灾,转移安置灾民 126 335 人,因灾死亡 2 人,伤病 553 人;农作物受灾面积达 17 万公顷,绝收 5.62 万公顷;43 662 间房屋受到不同程度损坏,其中倒塌 7 093 间;该市 6 条省级公路、5 条县乡级道路被迫中断,水毁路基或坍塌 80 处 40 481m³。此外,这次特大暴雨过程还造成城区严重内涝、交通受阻、房屋进水等,严重影响当地人民群众的正常生活。

(襄樊市气象台 王 芹)