

汤驰¹ 冯文² 官满员¹ 施思¹ 吴文娟¹ 吴佳妮¹

SVM 方法在热带气旋风雨及温度预报中的应用

摘要

热带气旋的强弱和移动路径会直接影响到周围大气中气压、温度、露点等气象要素的变化。为更好地了解热带气旋对海口市的影响,通过收集影响海口市热带气旋关键因子,建立热带气旋风雨影响预报因子库,基于 SVM 方法对热带气旋在过程降水量、最大风速和平均温度进行趋势预报。结果表明,该方法对影响海口市热带气旋的过程降水量、最大风速和平均温度都有较好的预测效果,但对于超过 15 m/s 的最大风速和 200 mm 以上降水量级上存在一定的偏差,这可能与 SVM 模式中预报因子库中关键因子不全及模式的择中原理使结果趋于平均化相关。

关键词

支持向量机;热带气旋;预报

中图分类号 P49

文献标志码 A

收稿日期 2014-01-23

资助项目 国家自然科学基金(40821092)

作者简介

汤驰,男,硕士,助理工程师,主要从事大气环境与污染方向研究.tangchi1986@163.com

1 海口市气象局,海口,570100

2 海南省气象局,海口,570100

0 引言

海口地处热带滨海的低纬度热带北缘,东临南海,夏季多台风暴雨,据统计 1951—2011 年每年平均影响海口的台风为 4.6 个,主要集中在 6—10 月^[1]。狂风、暴雨是热带气旋的主要致灾因子,因此,有效地预报热带气旋的风雨影响,在海口市防灾减灾体系中有着十分重要的意义。

热带气旋的强弱和移动路径会直接影响到周围大气中气压、温度、露点等气象要素的变化,通过对相应历史气象要素资料进行分析,能有效地预测热带气旋的变化特征,反馈出未来热带气旋的发展动态及其影响。近年来发展起来一种新型的机器学习方法——支持向量机(Support Vector Machines, SVM)方法^[2],它是一种在统计学理论上发展起来的新的数据挖掘方法,被许多研究学者应用在气象预报预测的各个领域。文献[3-4]利用 SVM 方法建立四川盆地分区面雨量的分类推理模型,进行相应的预报试验,试验结果显示对应的 SVM 推理模型具有良好的预报能力;张礼平等^[5]利用自然正交函数(EOF)分解和 SVM 回归分析,对热带海表温度场预报降水场进行试验,证实了 SVM 对于短期气候预测的小样本、场预报具有一定的能力。本文利用 SVM 方法读取海口市气象观测地面自动站及周边各高空站各层气象资料,建立海口市热带气旋影响期间过程降水量及风力和温度的客观预报方法。

1 SVM 方法预报原理简介

20 世纪 90 年代中期,实现统计学习理论和原则的实用化算法——SVM 方法才被完整提出,并在模式识别等人工智能领域成功应用,受到广泛关注。

SVM 方法的基本思想是通过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最优(凸规划)问题,基于 Mercer 核展开定理,通过非线性映射 ψ ,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),使其可以在特征空间中应用线性学习机的方法,解决样本空间中的高维度非线性分类和回归等问题。简单地说就是实现升维和线性化。

一个向量维数无限增加时就表现为一个实函数,在实际问题中一般都是非线性问题。如果设样本空间为 X ,引入一个非线性映射 ψ ,

把样本空间通过映射到新的特征空间形成 $\psi(x)$, 将可积分函数 $\psi(x)$ 乘上一个新的二元函数 $K(x, y)$, 再作积分形成一个新的函数实现输出, 这个变换过程中的二元函数 $K(x, y)$ 称为积分变换的核函数. 对于一个对称正定的连续核称为 Mercer 核, Mercer 核函数 $K(x, y)$ 展开成一致收敛的函数项级数:

$$K(x, y) = \sum_i \lambda_i \varphi_i(x) \varphi_i(y),$$

其中 $\lambda_i, \varphi_i(x)$ 分别为核 $K(x, y)$ 的特征值和特征函数, 它们的个数可能有限或无穷. 通过构造特征空间的非线性升维映射, 最后求得最优超平面决策函数为

$$M(X) = \text{sgn} \left(\sum_{\text{支持向量}} a_i^* y_i (\psi(X) \cdot \psi(X_i) + b^*) \right),$$

其中 a_i^* 和 b^* 为确定最优划分超平面参数, 确定最优划分超平面的样本点称为支持向量. 从表达式可以看出只需要求出支持向量 X_i 及其支持的强度 a_i^* 和阈值 b^* , 通过核函数的计算, 即可得到原来样本空间的非线性划分输入值, 这样在实际计算中并不需要知道它们的显式表达式, 从而解决了因升维而产生的复杂计算问题.

利用传统的统计分析方法得到的很多分类函数能够在样本集上轻易达到 100% 的正确率, 却在真实分类时表现欠佳 (即所谓的推广能力差, 或泛化能力差). 原因在于样本数相对于现实世界要分类的文本数来说简直九牛一毛, 经验风险最小化原则只在这占很小比例的样本上做到没有误差, 当然不能保证在更大比例的真实文本上也没有误差. SVM 方法的建模过程根本没有涉及概率测度, 而是依据间隔最大化原则进行统计建模, 它的最终决策函数只由少数支持向量确定, 计算的复杂性由影响该预报对象的最终支持向量数目和选用的核函数决定, 而非样本空间的维数, 这克服了在预测过程中因升维而产生的困扰, 也使 SVM 方法具有较好的鲁棒性.

2 资料处理及建模

2.1 资料的选取

SVM 回归方法是依据关键样本 (支持向量数) 来构造最终的决策函数, 这与传统的基于权重系数来建立各类因子与预报对象变化的统计方法 (如逐步回归、卡尔曼滤波、神经网络) 有显著区别. SVM 方法中单个因子和预报对象是否具有显著相关并不重要 (适合于解决本质上非线性的问题), 重要的是如何选择与预报对象有密切联系的因子集. 因此, 因子

预选库的广泛程度将直接影响到最终关键因子能否被完全选中.

为使所选资料更具有代表性和广泛性, 本文选取海口市 2006—2011 年东面 (站点编号 59758, 国家基准站) 和西面 (站点编号 59757, 国家一般站) 2 个地面站为所有地面要素作为预报因子, 同时收集海口市及其附近三亚市和三沙市 3 个高空站 5 个关键层所有气象要素为组合预报因子, 汇总建立预报因子预选库, 然后应用 SVM 回归方法建立预报模型, 寻找其与海口市热带气旋影响期间过程降水量、最大风力及温度之间的关系.

2.2 资料预处理

2.2.1 资料筛选

由于资料的不完整性, 为保证数据的连续和可对比性, 本文选取 2006 年 7 月—2011 年 5 月年间逐日连续性数据进行分析建模. 通过对比分析, 筛选出影响热带气旋最密切的 3 个高空站 (图 1) 5 个代表性层次 (200、500、700、850 和 950hPa) 的温度、露点、风向风速等各类基础气象要素统计纳入预报因子库进行预报分析; 地面站选用各类常规观测气象要素 (包括风向风速、温度、压强等) 及部分组合因子纳入预报因子库.

在选取的各类预报因子之间, 为保证有较高的收敛性, 首先对各类预报因子与预报对象之间进行相关性分析, 根据因子自由度的大小, 筛选出符合显著性水平最高 ($\alpha = 0.01$) 系数标准的因子为备选因子库成员, 然后在选出的因子之间再进行相关性检验, 有选择性地剔除掉相关性较大的因子, 最后将剩余因子作为最终预报因子库建立预报回归方程.

2.2.2 归一化处理

归一化是 SVM 方法中数据处理的重要环节, 不少学者在开始运用 SVM 方法时没有将数据进行归一化, 结果导致运行时间过长甚至出现死循环的情况, 因此, 为了使后面数据处理更加合理, 本文将样本数据进行归一化处理, 处理方法如下:

$$X_{ia} = \frac{X_{ia} - X_{\min}}{X_{\max} - X_{\min}},$$

其中, i 为样本序列号, a 为因子序列号, X_{\min} 为因子样本中的最小值, X_{\max} 为因子样本中的最大值.

通过归一化处理, 使整体样本数据统一在 0~1 之间, 为后期程序处理运行时收敛更加有效提供支持.

2.2.3 降水量等级化

根据预报和服务经验, 降水一般分为小雨、中

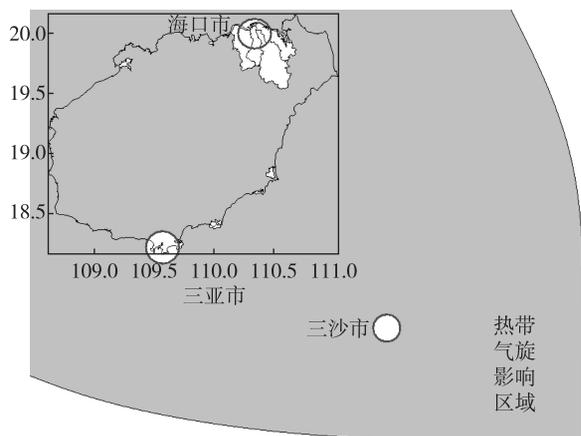


图1 三高空站点分布

Fig. 1 Distribution of three high-altitude stations

雨、大雨、暴雨等级别,因此,本文将降水量预报按照下面方式进行分级:[0~20) mm 为1级,[20~50) mm 为2级,[50~100) mm 为3级,[100~200] mm 为4级,大于200 mm 为5级.将降水量根据上述分级标准分类,基于SVM方法进行分级预报,验证降水预报的准确度.

2.3 基于SVM方法的模型建立

本文通过数据筛选后,得到过程降水量、最大风力和平均温度的最终预报因子库分别为24、25和26组进行建模分析.为保证样本数据最优化,随机选取样本集总量的70%、25%和5%分别作为训练、实验和检验数据,根据要求处理后输入SVM建模系统.

核函数的选取也十分重要,本文通过多组数据预测验证,3种气象要素的预测模式计算均选用径向基核函数,参数选取 $C=100$, $w=0.1$, $g=1$ 进行建模预测,然后根据需要选择合适的损失函数叠加上界和回归迭代最大次数建立预报方法.

3 SVM方法预报结果及检验

为检验SVM方法预报效果,本文首先通过在有限样本数据集中分段选取不同大小样本数进行比较分析,寻找最优样本数进行下一步预报建模;基于最优样本数,再次对历史热带气旋影响期间过程降水量、最大风速及平均温度做预报检验,评估该预报方法的推广能力.

3.1 不同样本数对预报结果的影响

许多学者对SVM方法研究表明,该向量机能对小样本数据有较好的预报效果,但并未有定量的结论.为了验证样本数对预报结果的定量影响,本文在

有限的样本条件下通过对样本数分段性递增,并检验其变化与预报相关系数分析.图2表明,在样本数为200个时,3类预报对象相关系数都有明显增加或减少的趋势,但随着样本数的继续增加,3类预报对象相关系数趋势都趋于一致性递增.其原因可能是由于样本数为200时,是SVM方法预报在对这3类要素预测时系统误差的最大点,因此,需尽量避开样本数在200左右时进行研究.本文将选取相关系数最好的样本数进行下面的预报分析.

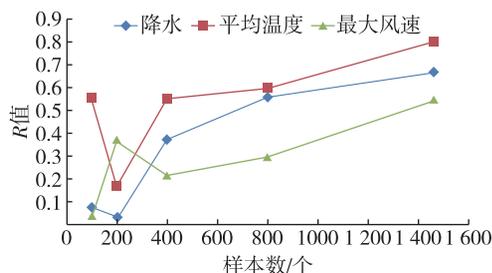


图2 3类预报对象样本数与预报相关系数对比

Fig. 2 Comparison of correlations between sample number and forecast accuracy for 3 factors

3.2 基于SVM方法预报模式对热带气旋影响要素的检验分析

为检验所建模型预报效果,实现预报业务化提供保障,本文选用近4年影响海口市的热带气旋资料与SVM方法预报结果比较分析,检验在热带气旋影响期间过程降水量、最大风速及平均温度的预报效果(图3).

从整体上看,最大风速与平均温度的预报趋势与实况均比较一致,能有效地反应出其未来的发展趋势,其中平均温度预报效果较好,最大风速超过15 m/s时,预报效果一般,但整体预报趋势比较一致,图3c中对于降水量超过200 mm及以上过程降水量的预报还存在一定的偏差,但200 mm以下预报效果和预报趋势都比较一致.对于较大降水量出现误差的原因与SVM模式预报因子库中关键因子不全有关(包括地形因子的考虑等),同时模式的择中原理使结果趋于平均化及降水量的不连续性(即前后数值跳动性大),是导致最终结果出现极值预报差异的原因.

为精确检验降水预报效果,本文对降水量进行等级化划分,分级后SVM方法对热带气旋的预报效果如表1所示.由表1可以看出,整体预报效果都在60%以上,其中2级的预报效果最好,达72%,这一

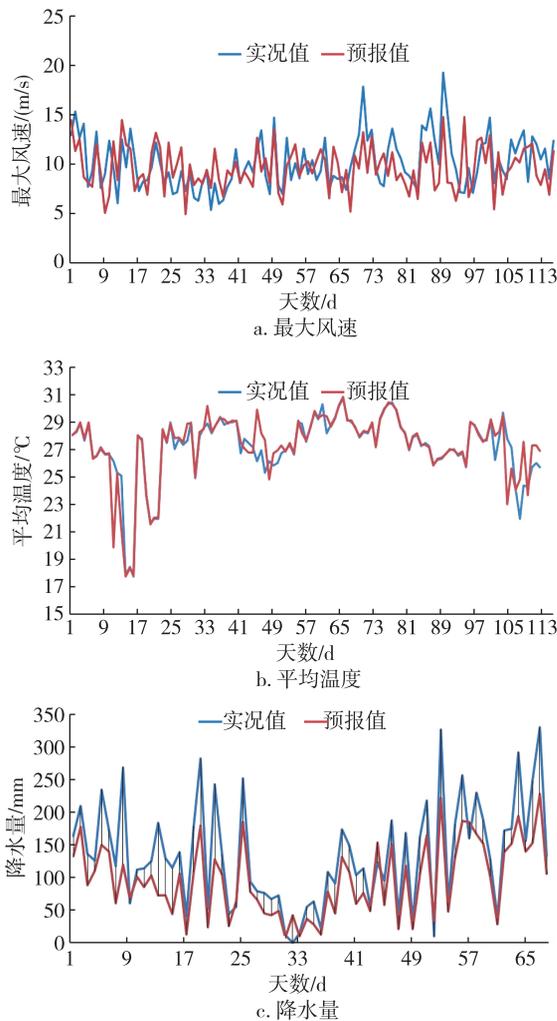


图3 实况值与SVM预报值比较

Fig. 3 Comparison between measured value and SVM forecasting value for maximum wind speed, average temperature, and precipitation

预报效果能为热带气旋影响海口市期间降水预报提供有效的预报基础。

表1 SVM方法分级预报降水准确率

Table 1 Forecasting accuracy of SVM method for different precipitation grades

级别	1级	2级	3级	4级	5级
准确率	0.68	0.72	0.66	0.63	0.65

从热带气旋对于降水产生的误差分析看,出现偏差主要存在以下2种情况:

1) 热带气旋路径比较复杂.从预报结论和实况数据比较分析来看,在对于热带气旋路径比较复杂情况下,预报出现误差较大.如1986年18号台风

wayne、1996年18号台风willie及2001年14号台风fitow,这些台风路径都相对比较复杂,3次预报效果均偏差较大,因此,对于较复杂路径的热带气旋,预测需要考虑的预报因子更加精细和完整,应加大所要考虑的因子范围和精细程度,才能使预测效果有所改善.

2) 登陆的热带气旋.另一个预报结果误差较大的是登陆海南岛的热带气旋.如1980年在文昌清澜一带登陆的14号台风ruth、1984年在文昌龙楼镇登陆的1号台风ike以及2008年在文昌龙楼镇登陆的17号台风higos,这些都是直接登陆海南岛的台风.由于登陆台风必须要充分考虑到地形条件和建筑条件的复杂程度等,它们都会一定程度上影响到预测结果,但在预报因子库中都没有得到充分的诠释,因此,这些条件对本研究的最终结果有一定程度的影响.

4 结论及展望

1) 基于SVM方法的热带气旋风雨影响及温度预报在不同样本数时效果不同,在总样本数为200时不确定性最大,在条件允许范围内应尽量选用更多样本数从而达到更高的准确度.

2) SVM方法对热带气旋在过程降水量、最大风速和平均温度的趋势预报方面都有很好的效果,能有效地反映出其未来的发展趋势.其中平均温度预报效果较好,但最大风速超过15 m/s时,预报效果一般,对于降水量超过200 mm及以上过程降水量的预报还存在一定的偏差,这与SVM模式预报因子库中关键因子不全有关(包括地形因子的考虑等),而模式的择中原理使结果趋于平均化及降水量的不连续性(即前后数值跳动性大),是导致最终结果出现极值预报差异的原因.

总体而言,基于SVM方法的热带气旋风雨预测及温度预报对评估热带气旋的影响有一定的指导意义,只要不断完善预报因子库,找到影响的关键支持向量,该方法的预报能力将会有更好的体现.

参考文献

References

[1] 施思,吴文娟.近61年影响海口的热带气旋气候特征分析[J].海南气象,2013,5(1):22-24
SHI Si, WU Wenjuan. Climate characteristics of tropical cyclones affecting Haikou in past 61 years [J]. Hainan Meteorology, 2013, 5(1): 22-24

[2] 吕庆平,罗坚,朱坤,等.基于SVM的气候持续法在热

- 带气旋路径预报中的应用试验[J].海洋预报,2009,26(1):76-83
- LÜ Qingping, LUO Jian, ZHU Kun, et al. Experiments on predicting tracks of tropical cyclones with climatic persistence method based on support vector machine [J]. Marine Forecast, 2009, 26(1): 76-83
- [3] 陈永义, 俞小鼎, 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): 346-353
- CHEN Yongyi, YU Xiaoding, GAO Xuehao, et al. A new approach for non-linear classify and non-linear regression. I: Introduction to support vector machine [J]. Journal of Applied Meteorological Science, 2004, 15(3): 346-353
- [4] 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): 355-365
- FENG Hanzhong, CHEN Yongyi. A new approach for non-linear classify and non-linear regression. II: Application of support vector machine to weather forecast [J]. Journal of Applied Meteorology, 2004, 15(3): 355-365
- [5] 张礼平, 陈永义, 周筱兰. 支持向量机(SVM)及其在天气预报中的应用[J]. 热带气象学报, 2006, 22(3): 278-282
- ZHANG Liping, CHEN Yongyi, ZHOU Xiaolan. Applications of support vector machines in the fields prediction [J]. Journal of Applied Meteorology, 2006, 22(3): 278-282
- [6] 方辉. 支持向量机的研究与发展[J]. 大庆师范学院学报, 2007, 27(5): 21-23
- FANG Hui. Research and development of support vector machine [J]. Journal of Daqing Normal University, 2007, 27(5): 21-23
- [7] 冯汉中, 陈永义. 支持向量机回归方法在实时业务预报中的应用[J]. 气象, 2005, 31(1): 41-44
- FENG Hanzhong, CHEN Yongyi. Application of support vector machine regression method forecasting in weather forecast [J]. Meteorological Monthly, 2005, 31(1): 41-44
- [8] 王宝书. 支持向量机方法在预报白山流域夏季定量降水中的应用[C]//第26届中国气象学会年会, 2009
- WANG Baoshu. Application of support vector machine method in quantitative precipitation forecast for the summer in White Mountain Basin [C] // The 26th Chinese Meteorological Society Annual Meeting, 2009
- [9] 樊高峰, 张勇, 柳苗, 等. 基于支持向量机的干旱预测研究[J]. 中国农业气象, 2011, 32(3): 475-478
- FAN Gaofeng, ZHANG Yong, LIU Miao, et al. Study of drought prediction based on support vector machine [J]. Chinese Journal of Agrometeorology, 2011, 32(3): 475-478
- [10] 王在文. 基于非线性支持向量机方法的奥运场馆气象预报[C]//第27届中国气象学会年会, 2010
- WANG Zaiwen. The weather forecasting of Olympic venues based on the method of nonlinear SVM [C] // The 27th Annual Meeting of Chinese Meteorological Society, 2010
- [11] 滕卫平, 胡波, 滕舟, 等. SVM回归法在西太平洋热带气旋路径预报中的应用研究[J]. 科技通报, 2012, 28(11): 49-53
- TENG Weiping, HU Bo, TENG Zhou, et al. The SVM application on the path forecast of tropic cyclones of the West Pacific [J]. Bulletin of Science and Technology, 2012, 28(11): 49-53

Application of SVM method in weather forecasting during tropical cyclone process

TANG Chi¹ FENG Wen² GUAN Manyuan¹ SHI Si¹ WU Wenjuan¹ WU Jiani¹

¹ Haikou Meteorological Bureau of Hainan Province, Haikou 570100

² Hainan Meteorological Bureau, Haikou 570100

Abstract The strength and movement path of tropical cyclones will directly affect the changes of the ambient atmosphere in aspects of air pressure, temperature, dew point and other weather elements. To better understand the impact of tropical cyclones on climate of Haikou city, this paper collects the key factors of tropical cyclones affecting Haikou city, thus establishes the impact predictor pool, and forecasts the trends in precipitation, maximum wind speed and the average temperature based on the Support Vector Machines (SVM) method. The results show that this method generally works well in trend prediction of the three weather elements during tropical cyclone process affecting the Haikou city. But the method show some deviation in forecast accuracy for wind speed higher than 15m/s and rainfall greater than 200 mm, which may be due to the lack of some key factors in predictor pool and the averaging process in principle of SVM model.

Key words support vector machines (SVM); tropical cyclone; forecast