

程学伟,韩兆洲,2018. 市域脑卒中疾病与气象因素的关系及预测[J]. 气象,44(6):837-843.

## 市域脑卒中疾病与气象因素的关系及预测\*

程学伟 韩兆洲

暨南大学经济学院统计学系,广州 510632

**提 要:** 为了帮助医疗机构合理调配医务力量、床位和医疗药物,同时也帮助脑卒中高危人群及时采取干预措施,降低发病风险。本文对某市<sup>[1]</sup>四家医院 2013—2016 年脑卒中的就诊病例进行数据分析,将日就诊人数分为 6 个等级。然后,调取相应时段当地的逐日气象资料,采用支持向量机(SVM)和随机森林(RF)方法分别建立了日就诊人数预测模型和日就诊人数与气象因素的关系模型。研究表明:(1)脑卒中的日就诊人数为不平衡数据,这种数据特征将导致传统的预测模型正确率较低;(2)通过不断调整 SVM 预测模型的初始权重,经历了 4 次优化之后,使得日就诊人数的预测正确率从 52.46% 上升到 94.56%;(3)随机森林模型的结果显示,影响脑卒中发病率的三大气象因素分别是最高气温、最低气温和平均气温。基于机器学习模型的脑卒中疾病与气象因素的研究成果,提高了医疗气象统计模型的预报准确率,具有较高的应用和推广价值。

**关键词:** 脑卒中,气象因素,SVM,RF,非平衡数据

**中图分类号:** P49,TP181

**文献标志码:** A

**DOI:** 10.7519/j.issn.1000-0526.2018.06.011

## Relationship Between Stroke and Meteorological Factors and Its Prediction in Urban Areas

CHENG Xuewei HAN Zhaozhou

Department of Statistics, College of Economics, Jinan University, Guangzhou 510632

**Abstract:** The purpose of this paper is to help medical institutions in terms of a reasonable allocation of medical hands, beds and medical drugs, and also to help people in high stroke risk by mean of timely intervention, reducing the risk of stroke occurrence. The data of the cases of stroke in four hospitals from 2013 to 2016 were analyzed, and the number of daily patients in the hospital was divided into six grades. By taking the support vector machine (SVM) and the random forest (RF) methods, and using the corresponding local daily meteorological data collected, we established the prediction model for daily patients and the correlation model related to meteorological factors. The results showed that: (1) the number of daily patients with stroke was not balanced and the accuracy of traditional prediction model was relatively low. (2) By constantly adjusting the initial weight of the SVM prediction model, we found that, after 4 attempts of optimization, the prediction accuracy of the number of daily patients increased from 52.46% to 94.56%. (3) The results of the random forest model showed that the three meteorological factors affecting the incidence of stroke are the maximum temperature, the minimum temperature and the mean temperature. Thus, the research results of the correlation between stroke diseases and meteorological factors based on machine learning model could improve the prediction accuracy of medical meteorological statistical models, and have good application and promotion values.

**Key words:** stroke, meteorological factors, support vector mechine (SVM), random forest (RF), non-equilibrium data

\* 国家社会科学重点基金项目(15ATJ001)资助

2017 年 9 月 15 日收稿; 2017 年 12 月 11 日收修定稿

第一作者:程学伟,主要从事机器学习算法以及气象统计研究. Email:cxwpaper@163.com

[1]注:由于涉及脑卒中病例数据的隐私性,故本文隐去具体城市名

## 引言

脑卒中(俗称脑中风)是一种突然起病的脑血液循环障碍性疾病,也是脑血管疾病最严重的并发症。所以,脑卒中又被称为“人类沉默的杀手”。据统计,中国每年发生脑卒中的病人高达200万,每年脑卒中死亡120万。2012年,第三次全国死因调查显示(吕红霞和李志伟,2013),脑卒中致死率排名第一,高达22.45%。

研究表明,脑卒中中具有极高的致残率和较高的致死率,而这种疾病与诱发环境因素包括气温和湿度等之间存在密切的关系。国内外气象因素与心脑血管疾病的关系研究,我国最早可以追溯到20世纪80年代,王兴中和祝锦涛(1986)以合肥地区为例,研究了急性脑血管疾病的发病与异常气象的关系。许显福和裘建强(1990)、印佩芳等(1993)、杨贤为和邹旭恺(1998)分别探讨了气象因素和脑卒中发病之间的关系。初步证实了脑卒中发病与天气因素有关,但是具体什么天气因素与脑卒中发病相关,学者们仍然没有给出明确的结论。到了21世纪,梁丽英(2009)指出广州市脑卒中的日发病人数与当天的日平均气温、日最高气温、日最低气温、日平均水汽压、日平均露点温度呈显著负相关,与日平均气压呈显著正相关;凌佳(2013)利用半参数回归模型研究脑卒中的影响因素;方万里和马利娟(2014)利用2007—2010年宁波地区的脑卒中病例数据与气象资料,进行脑卒中的发病规律与气象因素的统计分析,发现脑卒中发病人数与前期的温差、高温和平均相对湿度情况呈负相关,与当期的气温和气压呈正相关。而在国外期刊上,Cowperthwaite and Burnett(2011)对2004—2008年155家美国医院的196439例脑卒中病人进行研究,并将脑卒中病人的发作类型分为缺血性、出血性、或短暂性脑缺血发作,研究结果表明天气和中风之间的季节性关联高度混淆。Wang et al(2017)对2004—2009年的中国济南地区1577例出血性脑卒中病人进行研究,发现出血性脑中风与环境温度近似呈线性关系,并给出了脑卒中的一个危险因素:低温。

当然,也有一些学者从医学机理的角度出发,分析气象因素对脑卒中发病的影响机制。张书余等(2013;2016)认为冷空气会影响血液的流动性,使得动脉粥样硬化以及血管收缩、心肌细胞受损。同时,

张书余等(2010)也指出高温、高湿天气将造成心脏排血量增加;强降温将导致肾上腺素水平上升,诱发冠状动脉痉挛甚至心肌梗死。阮焯等(2012;2013)、阮焯(2013)和赵娜娟等(2011)分别探讨了冷空气对人体血脂水平、血流变指标以及相关敏感指标的影响,证实了在冷空气活动影响心脑血管疾病的发病机制中,血管内皮素-1可能起到一定的作用;冷空气对HDL-C, ApoAL, TG及VLDL-C指标均有一定影响,这可能是造成心脑血管疾病的主要危险因素之一。

同时,对于脑卒中发病人数的预测以成为医学统计新的研究热点。董平和张敬华(2000)利用圆形统计探讨老年心血管病发病的季节规律,并说明发病的高峰时段大约在冬季12月至次年1月左右波动。随后,一些学者开始利用神经网络(谭英等,2013;闵晶晶等,2014)的方法来预测脑卒中的发病率,预测正确率分别为84.6%和60%。显然,传统的预测方法正确率还有待提升。而支持向量机具有一种自学习、自组织的非线性映射能力,高效地实现了从训练样本到预报样本的“转导推理”,大大简化了通常的分类和回归等问题。而且在医学数据中,存在大量的数据缺失、遗漏和误填,而支持向量机能够帮助我们抓住关键样本、“剔除”大量冗余样本,所以该算法具有较好的“鲁棒性”(即抗变换性,robustness)。因此,本文采用支持向量机来建立气象因素变化和脑卒中日均就诊人数的预测模型,为提高医疗气象统计模型的预报准确率提供参考依据。

## 1 资料与方法

### 1.1 资料来源

本文选用2013年1月至2016年12月期间,某市四家医院脑卒中就诊人数与相应的逐日气象资料作为研究对象。时间跨度为4年(共1461d),四家医院共接诊脑卒中(包括脑出血、脑梗塞)病人61877例,病例信息包括发病时间、诊断报告时间、性别、职业和年龄。XX气象站逐日气象资料包括8个基本因素:日平均气压(单位:hPa)、日最高气压(单位:hPa)、日最低气压(单位:hPa)、日平均气温(单位:℃)、日最高气温(单位:℃)、日最低气温(单位:℃)、日平均湿度(单位:%)和日最低湿度(单位:%)。

### 1.2 资料处理

从前人的研究成果看,天气因素能够从某种程度上诱发脑卒中发病。为了从统计学上分析天气因素与就诊人数之间的相关关系,本文选取日就诊人数与 8 个天气因素进行复相关分析,复相关系数  $R = 0.107$ ,说明天气因素和日就诊人数之间存在弱相关关系。其  $P$  值为 0.033,在 0.05 的显著水平下,可以说天气因素和日就诊人数之间的相关关系是显著的。具体的复相关系数如图 1 所示。

从图 1 可以得出,虽然天气因素并不直接导致脑卒中的发病,但是之间的相关性还是存在的。为了进一步了解日就诊人数的分布特征,绘制频数分布折线如图 2 所示。

研究发现,2013—2016 年的 1461 天的脑卒中就诊人数不是简单的正态分布,而是呈现右偏分布,即日就诊人数处在  $[18, 60]$  占据了总天数的 84.74%。这种不平衡数据的存在,可能是导致传统的预测模型失效的因素之一。

### 1.3 支持向量机和随机森林方法

#### 1.3.1 支持向量机方法

支持向量机(SVM)是一种基于 VC 维理论和

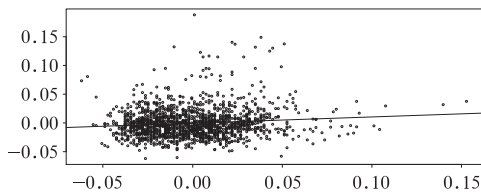


图 1 日就诊人数与天气因素的复相关系数图  
Fig. 1 Complex correlation coefficient between the number of daily patients and weather factors

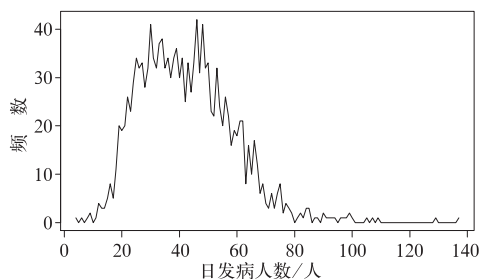


图 2 脑卒中日就诊人数频数分布折线图  
Fig. 2 Frequency distribution of the number of patients

结构风险最小原理的分类方法,根据有限样本在模型的复杂性和学习能力之间寻求最佳折中,以期获得最好的推广能力。而由 SVM 发展出来的回归方法称为支持向量回归(SVR)。

SVM 最基础的原理就是将空间中的两类点 ( $y = -1$  或  $y = 1$ ) 用超平面  $w^T x + b = 0$  分开(在严格线性可分的情况下,存在这样的超平面),而且希望这个超平面与两类的距离最大,也就是说,使得隔离带宽  $\rho = \frac{2}{\|w\|}$  最大。这等价于用 Lagrange 乘子法求下式的极小值。

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1] \tag{1}$$

根据得到的解  $w^*, b^*, \alpha^*$  得到最优分割超平面方程  $w^* T x + b^* = 0$ 。任意点  $(x)$  的函数值  $w^* T x + b^*$  的符号确定了该点的分类,或者说判别函数为  $\text{sgn}(w^* T x + b^*)$ 。

在实际的回归问题中,  $y$  的取值不仅仅是一和 1。令  $f(x) = w^T x + b$ , 希望  $y$  与  $f(x)$  的离差越小越好,SVR 问题还是归结于求使得  $\frac{\|w\|}{2} = \frac{w^T w}{2}$  最小的  $w$ ,但约束条件是  $|y_i - f(x_i)| \leq \epsilon$ , 这里  $\epsilon$  为某目标值。类似于 SVM,SVR 允许一些误差,这样就可以把上面的约束放宽为(对于大于 0 的  $\xi_i, \xi_i^*$ )  $y_i - f(x_i) \leq \epsilon + \xi_i$  及  $f(x_i) - y_i \leq \epsilon + \xi_i^*$ , 即得到 Lagrange 函数

$$L(w, b, \xi, \alpha, \eta) = \frac{1}{2} \|w\|^2 + C \sum_i (\xi_i + \xi_i^*) - \sum_i (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_i \alpha_i (\epsilon + \xi_i - y_i + w^T x_i + b) - \sum_i \alpha_i^* (\epsilon + \xi_i^* + y_i - w^T x_i - b) \tag{2}$$

需要在约束条件  $\alpha > 0$  且  $\eta > 0$  下,解  $\min_{w, b, \xi} \{ \max_{\alpha, \eta} L(w, b, \xi, \alpha, \eta) \}$  问题。对于非线性问题,可以基于对偶函数,猜测较为灵活的核函数  $K(x_i^T, x_j) = \Phi(x_i)^T \Phi(x_j)$ , 其中  $\Phi(x)$  为  $x$  的变换。

#### 1.3.2 SVM 针对不平衡数据的权重优化原理

非平衡数据是指数据集中某一类的样本数量明显少于其他类样本的数目。这类数据会带来一个问题,就是机器学习的泛化能力 (generalization ability) 的降低。针对不平衡数据的处理问题,2002 年 Chawla et al (2002) 提出了 SMOTE (synthetic

minority over-sampling technique)算法,对向上抽样策略进行改进。而本文在交叉验证环节,从分类预测的结果出发,找出机器学习的泛化误差来源,从而人为地优化各个初始类别的比重,使得机器学习更加具有目的性和针对性。这样优化的方法,可以减少泛化误差,提高 SVM 算法的分类精度。

### 1.3.3 随机森林方法

随机森林(RF)是一种基于统计学的机器学习模型,它利用 bootstrap 重抽样方法从原始样本中抽取多个样本,然后对每个 bootstrap 样本进行决策树建模,然后组合多棵决策树的预测,平权投票出最终的预测结果。

随机森林中的每一棵决策树,使用相应的袋外(out of bag, OOB)数据计算它的袋外数据误差,记为  $errOOB_1$ 。随机地对袋外数据所有样本的特征  $X$  加入噪声干扰(具体的操作方法就是随机地改变样本在特征  $X$  处的取值),再次计算它的袋外数据误差,记为  $errOOB_2$ 。假设随机森林中有  $N$  棵树,那么特征  $X$  的重要性为

$$Importance = \sum (errOOB_2 - errOOB_1) / N \quad (3)$$

用这个表达式来作为相应特征重要性的度量值是因为:若给某个特征随机加入噪声之后,袋外的准确率大幅度下降,则说明这个特征对于样本的分类结果影响很大,也就是说它的重要程度比较高。

## 2 优化建模与日就诊人数等级预测

经过数据统计发现,脑卒中日就诊人数最小值为 4,最大值为 134。基于此,本文将日就诊人数等分为六个等级(表 1),方便接下来构建 SVM 模型对日就诊人数进行预测和检验。

根据表 1 可知,脑卒中日就诊六类人数中,样本

表 1 脑卒中日就诊人数预报等级

Table 1 Grade prediction of the number of stroke patients

日就诊人数区间/人	天数/d	日就诊人数等级
[4,27)	248	A1
[27,50)	769	A2
[50,73)	387	A3
[73,96)	46	B1
[96,119)	9	B2
[119,142)	2	B3

数量呈现巨大差异,为典型的非平衡数据。这种在原始数据集分布就不均衡的数据,又称为本质非平衡数据。研究发现,传统的分类算法如决策树、Bagging、Boosting 对日就诊人数的预测正确率均为 50%以下,随机森林回归的预测正确率为 48%。在不平衡分类问题上,黄衍和查伟雄(2012)通过比较 20 个 UCI 数据集上随机森林和支持向量机的性能,证明了随机森林在不平衡数据上的表现显著逊色于支持向量机。因为 SVM 能够克服 RF 的等权投票机制,人为地优化各个类别的权重,从而达到更好的预测性能。

### 2.1 生成训练集和测试集

为了评价比较 SVM 算法,本文通过抽样将数据集分为训练集(train sample)和测试集(test sample),两者之间的比例为 3 : 1,即通过 3/4 的样本建立 SVM 模型,来预测另 1/4 样本的日就诊人数。为保持数据集分布,采用分层抽样,即在 A1、A2、A3、B1、B2、B3 各等级中分别抽取 1/4 作为测试集,剩下的 1095 个样本作为训练集。

### 2.2 建立 SVM 模型及预测判别

运用训练集建立 SVM 模型,所得结果如表 2 所示。

表 2 基于 SVM 的日就诊人数模型结果

Table 2 Results of daily number of patients based on SVM

参数			
SVM 类型	C 分类器	核函数	radial
成本	1	$\gamma$	0.125
支持向量数: 979 (7, 461, 186, 290, 34, 1)			

表 2 中 SVM 类型项目说明样本模型的类别为 C 分类器模型;SVM 核函数项目说明本模型所使用的核函数为高斯内积函数且参数  $\gamma$  的取值为 0.125;成本项目说明本模型确定的约束违反成本为 1。模型找到了 979 个支持向量,六个等级各自对应的支持向量个数为 7,461,186,290,34 和 1。根据以上建立的模型,对测试集进行预测,其预测的混淆矩阵如表 3 所示。

从混淆矩阵可以发现,如果不对分类的权重进行优化的话(默认比重为 1 : 1 : 1 : 1 : 1 : 1),SVM 针对不平衡数据的预测结果也比较差,预测正确率仅为 52.46%,而且几乎把日就诊人数都预测为 A2 等级。

表 3 日就诊人数的预测结果的混淆矩阵  
Table 3 Confusion matrix of predictions about the number of patients

	A1	A2	A3	B1	B2	B3
A1	0	62	0	0	0	0
A2	0	190	2	0	0	0
A3	0	95	2	0	0	0
B1	0	12	0	0	0	0
B2	0	2	0	0	0	0
B3	0	1	0	0	0	0

### 2.3 优化 SVM 模型及预测判别

从混淆矩阵和数据分析特征可知,每组的样本量不一样,如果采用等权赋值的话,预测结果就较差。所以下面将对六个类别进行不等权赋值优化,即提高前三类的权重,降低后三类的权重。具体优化的混淆矩阵如表 4 所示,预测正确率即对应的初始分类权重如表 5 所示。

表 4 4 次优化对应的混淆矩阵

Table 4 Confusion matrix corresponding to four optimizations

①	A1	A2	A3	B1	B2	B3	②	A1	A2	A3	B1	B2	B3
A1	7	53	2	0	0	0	A1	30	30	2	0	0	0
A2	0	191	1	0	0	0	A2	5	184	3	0	0	0
A3	0	77	20	0	0	0	A3	0	62	35	0	0	0
B1	0	11	1	0	0	0	B1	1	9	1	1	0	0
B2	0	2	0	0	0	0	B2	0	2	0	0	0	0
B3	0	1	0	0	0	0	B3	0	1	0	0	0	0
③	A1	A2	A3	B1	B2	B3	④	A1	A2	A3	B1	B2	B3
A1	54	8	0	0	0	0	A1	61	1	0	0	0	0
A2	4	184	4	0	0	0	A2	0	183	9	0	0	0
A3	0	19	78	0	0	0	A3	0	8	89	0	0	0
B1	0	6	0	6	0	0	B1	0	1	0	11	0	0
B2	1	1	0	0	0	0	B2	0	0	0	0	2	0
B3	0	1	0	0	0	0	B3	0	1	0	0	0	0

表 5 优化的分类权重与预测正确率  
Table 5 Optimized classification weight and predictive accuracy

分类权重	预测错误数	正确率
100 : 100 : 100 : 10 : 2 : 1	148	59.56%
1000 : 1000 : 1000 : 100 : 2 : 1	116	68.31%
100000 : 100000 : 100000 : 1000 : 2 : 1	44	87.98%
1000000 : 2000000 : 2000000 : 10000 : 2000 : 1	20	94.56%

结合表 4 和表 5,可以得到:

(1) 由表 3 的混淆矩阵可知,前三类的特征变量不明显。需要调高模型前三类的权重,于是将分类权重从 1 : 1 : 1 : 1 : 1 : 1 优化为 100 : 100 : 100 : 10 : 2 : 1。正确率从 52.46% 上升到了 59.56%,且由表 4 混淆矩阵①可得,日就诊人数等级为 A1 的,有 7 个预测正确,终于实现了 A1 与 A2 的特征区分。A3 被预测正确的个数已由原来的 2 个上升为 20 个;

(2) 由混淆矩阵①可知,适当升高前三类的比重是有效果的,但是 A1 与 A3 中的错误数依旧远高

于正确数。于是将比重从 100 : 100 : 100 : 10 : 2 : 1 优化为 1000 : 1000 : 1000 : 100 : 2 : 1,这时的错误个数减少了 32,正确率上升了近 10%。但是 A1 与 A3 的错误数还是很高,证明 SVM 模型还是不能够很好地区分前三类的具体特征;

(3) 明确了提高前三类的比重对正确率有正向影响,于是将比重从 1000 : 1000 : 1000 : 100 : 2 : 1 优化为 100000 : 100000 : 100000 : 1000 : 2 : 1,这次权重优化的幅度更大一点,前三类都提高了两个数量级。错误数减少了 72,比上次优化的效果还好一点,正确率从 68.31% 上升为 87.98%。从混淆矩阵③可知,此时的 A1, A2, A3 基本能够正确区分,并且 B1 的预测正确率也已经达到了 50%;

(4) 从混淆矩阵④可知,虽然前三类的正确率不错,但是 B2 和 B3 的错误率 100%,而且错的比较离谱。于是在第四次优化的过程中,增大前三类权重的同时,适当增大第四类、第五类的比重,将优化权重调为 1000000 : 2000000 : 2000000 : 10000 : 2000 : 1,错误个数为 20 个,说明一年 366 d 中,仅有 20 d 预测错误,正确率达到了 94.56%。并且除

了 B3 这一组的预测差距比较大之外,其他组的错误基本保持在上下等级浮动。这一预测结果对于医疗机构合理调配医务力量、床位和医疗药物,具有现实性的指导意义。

### 3 探索影响日就诊人数的天气因素—建立随机森林模型

虽然前一节给出的日就诊人数的预测,解决了医疗机构面临的问题。但是对于具体的病人来说,除了能够得到合理的救治之外,并没有给出具体的防范措施。于是,本节将建立随机森林模型,探索气象因素中的重要因素,帮助脑卒中高危人群及时采取干预措施。利用 2013—2016 年的脑卒中就诊人数与逐日气象数据,建立 70 棵树的随机森林模型,给出每个变量的重要性(表 6)。

表 6 影响日就诊人数的气象因素的重要性  
Table 6 Importance of meteorological factors affecting the daily number of patients

	IM	INP	I
$P_{ave}$	8.072040	40829.58	5
$P_{max}$	8.538240	40424.15	6
$P_{min}$	8.959362	44126.04	4
$T_{ave}$	7.406853	45547.79	3
$T_{max}$	9.829383	48362.74	1
$T_{min}$	7.403011	47071.93	2
$R_{ave}$	4.920590	37541.62	8
$R_{min}$	4.807281	39507.11	7

注: $P_{ave}$ ,  $P_{max}$ ,  $P_{min}$  分别表示日平均、最高和最低气压; $T_{ave}$ ,  $T_{max}$ ,  $T_{min}$  分别表示日平均、最高和最低气温; $R_{ave}$ ,  $R_{min}$  分别表示日平均、最低湿度。

Note:  $P_{ave}$ ,  $P_{max}$ ,  $P_{min}$  are the pressures of daily average, maximum and minimum;  $T_{ave}$ ,  $T_{max}$ ,  $T_{min}$  are the temperatures of daily average, maximum and minimum;  $R_{ave}$  and  $R_{min}$  are the humidities of daily average, minimum.

表 6 中的 IM 和 INP 指标都是重要性值的度量方式,数值越大,代表重要性越高。IM(increase in MSE)表示指标表示精度平均减少值,即表示误差的增加,准确性的降低;INP(increase in node purity)表示节点不纯度平均减少值,其实就是 RSS(回归平方和)的减少;I 是 Importance,为相应特征重要性的度量值,对应变量的重要值越大,则说明变量对于模型进行分类越重要。本文的重要性排序主要以第二个指标为准,发现排名前三的分别是日最高气温、日最低气温和日平均气温。鉴于此,我们发现温度是诱发脑卒中就诊的最重要因素,气压次之,温

度相关性较差。因此,脑卒中高危人群需要在温度变化异常的天气采取相应的预防措施,降低发病的风险。

## 4 结 论

本文从机器学习算法的角度,对脑卒中的日就诊人数进行预测,并通过随机森林的方法找出影响脑卒中发病的重要气象因素。主要结论如下:

(1) 初期的数据挖掘发现,日就诊人数与 8 个气象因素呈现弱相关。但脑卒中的日就诊人数为不平衡数据,从折线图上反映出其为典型的右偏数据类型,这种数据特征是导致传统的预测模型预测脑卒中日就诊人数等级准确率较低的原因。

(2) 通过不断优化的 SVM 模型,使得日就诊人数等级的预测正确率达到 94.56%,一年之中仅有 20 d 的日就诊人数等级预测错误。可以说,这种预测结果基本解决了医疗机构所面临的医务力量、床位和医疗药物的调配,具有一定的应用和推广价值。

(3) 最后,建立了日就诊人数关于气象变量的随机森林模型,探索出影响脑卒中发病率的三大重要气象因素:日最高气温、日最低气温、日平均气温。这将提醒脑卒中高危人群在温度变化异常的天气时,及时采取干预措施,降低发病风险。

## 参考文献

- 董平,张敬华,2000.用圆形统计探讨老年心血管病发病的季节规律[J].数理统计与管理,19(5):31-33.
- 方万里,马利娟,2014.宁波地区脑卒中发病规律与气象诱因统计分析[J].中国卫生统计,31(1):137-138.
- 黄衍,查伟雄,2012.随机森林与支持向量机分类性能比较[J].软件,33(6):107-110.
- 梁丽英,2009.广州市气象因素和空气污染与脑卒中发病的相关性研究[D].广州:广州医学院.
- 凌佳,2013.基于半参数回归模型的脑卒中发病率影响因素分析[J].数理医药学杂志,26(1):17-20.
- 吕红霞,李志伟,2013.睡眠障碍与脑卒中的相关性研究进展[J].现代医药卫生,29(17):2620-2622.
- 闵晶晶,丁德平,李津,等,2014.北京急性脑血管疾病与气象要素的关系及预测[J].气象,40(1):108-113.
- 阮焯,2013.冷空气活动对心脑血管疾病相关指标影响的初步研究[D].兰州:兰州大学.
- 阮焯,牛静萍,张莉,等,2012.冷空气活动对心脑血管疾病发生相关敏感指标影响的研究[J].环境与健康杂志,29(6):496-498.
- 阮焯,牛静萍,张莉,等,2013.冷空气对心脑血管疾病患者血流变指标的影响[J].环境与健康杂志,30(9):773-776.

- 谭英,耿德勤,黄水平,2013.用人工神经网络建立缺血性脑卒中复发的预测模型[J].中国卫生统计,30(5):687-689.
- 王兴中,祝锦涛,1986.急性脑血管疾病的发病与季节和异常气象的关系(合肥地区2142例分析)[J].安徽医学,7(5):1-3.
- 许显福,裘建强,1990.急性脑卒中与气象要素的探讨[J].浙江气象科技,11(1):33-35.
- 杨贤为,邹旭恺,1998.北京地区脑卒中发病率的气象条件研究[J].气象,24(9):51-54.
- 印佩芳,马辛宇,袁军,等,1993.脑卒中与天气过程的关系[J].气象,19(12):44-47,53.
- 张书余,牛静萍,罗斌,等,2013.冷空气对心脑血管疾病的影响及其机制研究[M].北京:气象出版社.
- 张书余,王宝鉴,谢静芳,等,2010.吉林省心脑血管疾病与气象条件关系分析和预报研究[J].气象,36(9):106-110.
- 张书余,张夏琨,崔世杰,等,2016.冷空气对人心血管系统及相关影响因素的自然实验研究[J].气象,42(10):1256-1262.
- 赵娜娟,牛静萍,阮焯,等,2011.冷空气活动对人群血脂水平的影响[J].环境与健康杂志,28(11):953-955.
- Chawla N V, Bowyer K W, Hall L O, et al, 2002. SMOTE: synthetic minority over-sampling technique[J]. J Artif Intell Res, 16(1): 321-357.
- Cowperthwaite M C, Burnett M G, 2011. An analysis of admissions from 155 United States hospitals to determine the influence of weather on stroke incidence[J]. J Clin Neurosci, 18(5): 618-623.
- Wang Qinzhou, Gao Cuilian, Liu Hongen, et al, 2017. Hypertension modifies the short-term effects of temperature on morbidity of hemorrhagic stroke[J]. Sci Total Environ, 598:198-203.