

黄翠银, 史旭明, 陆虹, 等. 三种智能计算低温冷害预报模型的对比分析研究[J]. 灾害学, 2020, 35(2): 48-54. [HUANG Cuiyin, SHI Xuming, LU Hong, et al. Comparative Analysis of Three Kinds of Intelligent Computational Prediction Models for Cold Damage Weather[J]. Journal of Catastrophology, 2020, 35(2): 48-54. doi: 10.3969/j.issn.1000-811X.2020.02.011.]

三种智能计算低温冷害预报模型的对比分析研究*

黄翠银¹, 史旭明², 陆虹¹, 周秀华¹, 金龙¹

(1. 广西壮族自治区气候中心, 广西南宁 530022; 2. 桂林航天工业学院理学院, 广西桂林 541004)

摘要: 为了探索不同的非线性智能计算预报建模方法在冬季低温冷害预报中的应用效果, 利用广西区域逐日平均气温和降水资料计算得出的1951-2018年冷湿指数作为预报量。通过计算该预报量与前期再分析资料中的各种物理量相关得出预报因子。首先利用逐步回归方法, 从初选的相关预报因子中, 客观选出9个预报因子, 再对未选入的预报因子作核主成分非线性降维计算, 选出方差贡献最大的核主成分因子, 作为各预报模型的输入因子。分别采用模糊神经网络方法(FNN)、遗传算法的神经网络集成(GA-ANN)预报方法以及粒子群算法的支持向量机(SVM-PSO)集成预报方法三种不同的智能计算预报建模方法建立预报模型, 对2012-2018年的20次低温冷湿天气进行预报对比试验。结果表明, 在同样的预报建模样本条件下, 模糊神经网络预报模型对20个独立样本预报平均绝对误差最小。进一步的计算还表明, 这三种智能计算预报模型均比同样预报建模样本和预报因子的线性回归预报平均绝对误差要小, 显示了非线性智能计算预报建模方法, 对于具有明显非线性变化特征的冬季低温冷害天气过程, 比线性预报方法有更好的预报能力。

关键词: 智能计算; 预报模型; 低温冷害; 广西

中国分类号: P642; X43; X915.5

文献标志码: A

文章编号: 1000-811X(2020)02-0048-07

doi: 10.3969/j.issn.1000-811X.2020.02.011

对于低温冷害天气, 学者们从多方面进行了研究, 取得了很多研究成果。丁一汇^[1]、李崇银^[2]、杨贵名^[3]、王东海等^[4]分析了2008年初冰冻雨雪灾害天气的特征及其在天气、动力、气候等方面的形成原因及其维持机制。李艳等^[5]研究发现2008年初环流系统具有明显的30~60 d低频振荡特征。李向红^[6]、史悦等^[7]指出必须有强冷空气、水汽输送、动力条件的相互配合才能造成强低温雨雪冰冻灾害。唐熠等^[8]给出广西重大低温雨雪冰冻天气概念模型有乌山阻高型、两槽一脊型和多波动型三种。布和朝鲁等^[9]揭示欧亚大陆大型斜脊斜槽系统是我国冬季大范围持续性极端低温事件形成和维持的主要关键环流系统。可见, 低温冷害天气是多种天气尺度系统相互作用影响造成的, 是非线性大气运动过程。

对低温冷害天气的预报研究也较多, 最常见的是模式降尺度法和逐步线性回归方法。针对我国冬季极端低温日数, 单机坤等^[10]采用逐步回归的统计降尺度方法, 建立了一个动力-统计相结合的降尺度预测方法, 对2011/2012年冬季极端低温日数的预测趋势基本正确。张伟等^[11]使用基于尺

度分离的多元线性回归方法建立预报模型, 对我国冬季持续性低温事件进行预报, 1979-2003年的预报检验TS评分为0.57, 具有一定的借鉴意义。陈官军等^[12]利用1983-2009年DERF2.0系统回算产品, 采用逐步回归方法建立预报模型对中国南方冬半年持续性低温指数作延伸期预报试验, 效果好于模式直接预报的2 m气温。神经网络等智能方法在降水、台风等预测有较多运用, 但对低温冷害天气的智能方法预测仍较少。2015年, 陆虹等^[13]使用粒子群-神经网络方法建立的低温雨雪过程预报模型, 预报效果明显比逐步回归方法好。鉴于此, 本文尝试利用模糊神经网络、遗传-神经网络集成及粒子群-支持向量机集成三种智能方法对低温冷害天气进行预测, 探索其预报效果并与传统逐步回归方法进行比较。

1 预报量和预报因子

1.1 预报量的计算

从文献可知^[13-14], 冷湿指数可表征华南地区

* 收稿日期: 2019-11-11 修回日期: 2020-01-10

基金项目: 国家自然科学基金项目(41565005); 国家重点研发计划重点专项项目(2018YFC1507700); 广西气象局延伸期气候预测创新团队项目

第一作者简介: 黄翠银(1977-), 女, 广西田东人, 高级工程师, 硕士, 从事气候分析与预测. E-mail: 358506774@qq.com

通讯作者: 史旭明(1981-), 男, 山西稷山人, 副教授, 硕士, 从事非线性统计. E-mail: 644721193@qq.com

的低温冷害天气。周秀华等^[14]对冷湿指数 PT 计算方法进行了修正, 使之更能体现我国华南区域冬季由低温和降水引起的低温冷害天气的程度。利用 1951-2018 年广西 92 站国家级地面气象站的冬季(12 月、1 月和 2 月)逐日平均气温和降水量数据, 按照周秀华等方法计算单日冷湿指数 PT 值。为了突出低温冷害天气的持续性, 再对逐日冷湿指数进行 3 天滑动平均。根据周秀华等定义的区域低温雨雪事件方法计算, 1951-2018 年冬季广西共有 319 例冷湿指数符合区域低温雨雪事件要求。因此, 本文以冷湿指数表征低温冷害天气, 以 1951-2018 年冬季的 319 例冷湿指数作为预报量。

1.2 预报因子的选取方法

1.2.1 预报因子资料及初选

冷湿指数表征某一区域低温雨雪影响轻重的程度, 主要由气温变化和降水量决定, 因此预报因子的选取也需要重点考虑选择和气温、降水密切相关的物理量。为此大气变量采用的是 NCEP/NCAR Reanalysis 逐日 $2.5^{\circ} \times 2.5^{\circ}$ 格点资料, 经度范围 $0^{\circ} \sim 360^{\circ}$, 纬度范围 $90^{\circ}S \sim 90^{\circ}N$, 物理量取气温场、高度场、海平面气压场、垂直速度场、风场、湿度场以及由风场计算得到的涡度场和散度场等。在实际预报中, 是由当天资料作第 2 d 的天气预报, 因而本文首先利用相关分析法计算预报量序列和前 1 d 各个物理量场的相关系数。从相关系数显著的区域选出相关场, 每个相关场至少含有 3 个相邻格点数据, 提取选定的相关场格点数据作平均即为一个因子。如果含有相邻相关场, 相关系数符号相同则再相加, 相反则相减, 由此产生新的组合相关因子, 以此提高因子的相关系数和稳定性。由以上方法初步选出 316 个相关因子作为预报因子。

1.2.2 预报因子的线性与非线性降维计算处理方法

因初选的预报因子较多, 故首先利用逐步回归方法对初选预报因子进行筛选, 同时考虑到被剔除的预报因子中可能包含与预报量相关的预报信息, 进一步利用核主成分分析方法^[15]对被逐步回归剔除的预报因子进行特征提取计算, 具体通过以下步骤对预报因子进行筛选:

(1) 分别筛查 316 个初选预报因子与预报量的相关系数, 以相关系数绝对值超过 0.3 并且达到或超过 0.01 显著性水平作为入选标准来选取, 最后共选出 41 个因子。

(2) 利用逐步回归方法 ($F = 3.0$) 在预报因子群中选 5 ~ 10 个预报因子, 最后筛选出 9 个预报因子。

(3) 因为利用逐步回归方法只有少数预报因子被选入预报模型, 仍有大量的较高相关预报因子

无法被选入, 进一步采用核主成分分析对被舍弃因子所包含的预报信息进行特征提取, 将提取的第一核主成分与逐步回归选入的 9 个因子共 10 个预报因子, 一起作为模型的输入矩阵。

通过以上步骤, 共选取了 500 hPa 温度场、地面温度场、散度场等共计 10 个预报因子。以 500 hPa 温度场为例说明预报因子的选取过程: 首先计算冷湿指数 PT 值序列与前一日 500 hPa 温度场的相关系数(图 1), 在我国及临近区域共计算出 3 个高相关区, 由相邻相关区同号相加、异号相减原则, 在 500 hPa 温度场上可以挑选出 5 个预报因子: 1 区、2 区、3 区、1 区和 2 区相减组合、1 区和 3 区相减组合, 它们的相关系数分别是 -0.21、0.28、0.26、-0.33、-0.34。通过 0.01 信度检验原则, -0.33 和 -0.34 的 1 区和 2 区相减组合、1 区和 3 区相减组合进入逐步回归方程中筛选, 最后相关系数为 -0.34 的 1 区和 3 区相减组合因子被选中。从图 1 可以看出, 1 区减 3 区表示这两个区域 500 hPa 温度差异的大小。1 区与 PT 值反相关, 说明 1 区温度越低 PT 值越高, 表示 1 区受冷空气影响气温低, 同理可推断 3 区受暖气团影响气温高, 两个区域之间的狭窄区域温度差大说明有锋区存在, 即在 $30^{\circ}N$ 附近存在东北-西南向的锋区。温度差异越大表示锋区越强烈, 次日锋线向南压冷空气南下引起广西降雨降温, 造成广西低温冷害天气, 该预报因子和广西低温冷害天气之间具有较明确的物理意义。

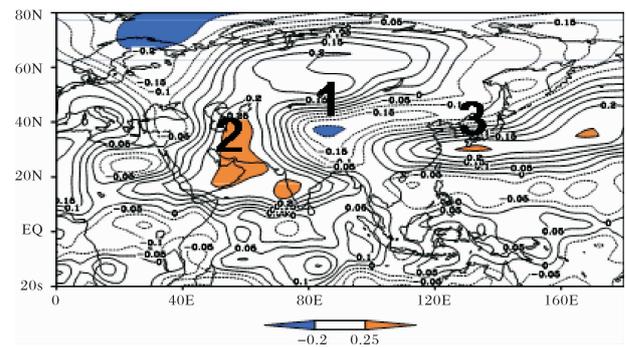


图 1 500 hPa 温度场和冷湿指数的相关分布

2 三种智能计算预报建模方法

在冷湿指数预报量序列(样本长度为 319)中, 本文将 1951-2011 年的冷湿指数样本作为预报建模样本, 2012-2018 年的冷湿指数样本作为独立预报样本进行预报检验, 即建模样本 299 个, 预报试验独立样本 20 个。利用同样的 299 个预报建模样本和 10 个预报因子, 分别使用模糊神经网络方法(FNN)、遗传算法的神经网络集成(GA-ANN)方法以及粒子群算法的支持向量机(SVM-PSO)集成预

报方法,共三种不同的智能计算预报建模方法建立冷湿指数预报模型,对20个独立样本进行预报试验。在每一种智能方法的逐次预报中,所有参数保持不变,使独立样本的预报与实际预报一致。

2.1 模糊神经网络冷湿指数预报

2.1.1 模糊神经网络预报建模

本文中的模糊神经网络模型采用4层前馈网络结构^[16],其第一层为模型的输入层,第二层为隶属度生成层,第三层为模糊推理层,第四层为反模糊化输出层。

在输入层中,以10个预报因子作为输入,节点数为10。

在隶属度生成层中,采用高斯函数作为隶属度函数,其表达式如下:

$$\mu_{ij} = \exp\left(\frac{-(x_i - a_{ij})^2}{\sigma_{ij}^2}, i=1, \dots, n; j=1, \dots, m\right) \quad (1)$$

式中: m 是模糊分割数, n 为输入节点数; a_{ij} 和 σ_{ij}^2 分别为隶属度函数的中心和宽度。

在模糊推理层中,采用相乘方式来进行模糊逻辑“and”运算:

$$\pi_j = \mu_{1j} \times \mu_{2j} \times \dots \times \mu_{nj} = \prod_{i=1}^n \mu_{ij}, j=1, \dots, m \quad (2)$$

在输出层中,采用中心平均反模糊方法进行运算:

$$y = \omega_1 \pi_1 \times \omega_2 \pi_2 \times \dots \times \omega_m \pi_m = \sum_{j=1}^m \omega_j \pi_j \quad (3)$$

其中 $\omega_j (j=1, \dots, m)$ 为输出层连接权值。这里利用BP算法对连接权值 ω_j 、隶属度函数的中心 a_{ij} 和宽度 σ_{ij}^2 进行学习训练。

首先定义模型的输出误差(即目标函数)为:

$$E_p = \frac{1}{2}(y - Y)^2 \quad (4)$$

式中: E_p 为平均误差函数, y 为模型的实际输出值, Y 为期望输出值,学习目标使 E_p 达到最小。根据梯度下降法推出3个参数的学习公式为:

$$a_{ij}(t+1) = a_{ij}(t) - 2\alpha(y - Y)\omega_j \times \prod_{i=1}^n \mu_{ij} \times (x_i - a_{ij})/\sigma_{ij}^2; \quad (5)$$

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) - 2\alpha(y - Y)\omega_j \times \prod_{i=1}^n \mu_{ij} \times (x_i - a_{ij})/\sigma_{ij}^3; \quad (6)$$

$$\omega_j(t+1) = \omega_j(t) - \alpha(y - Y)\pi_j \quad (7)$$

在模型学习过程中,首先利用样本训练集对模型进行学习训练,其主要计算步骤如下:

(1) 设学习率 α 为0.9,模型训练误差阈值为 e 。

(2) 训练初始时刻,利用随机数对连接权值 ω_j 、隶属度函数的中心 a_{ij} 和宽度 σ_{ij}^2 进行初始化。

(3) 利用式(5)~式(7)对3个参数进行学习训练。

(4) 计算模型的实际输出 y 与期望输出 Y 的误

差 E_p 。

当误差 $E_p > e$ 时,返回到(3);否则模型学习结束,利用最后确定的参数和连接权建立模糊神经网络预报模型。

2.1.2 模糊神经网络预报结果

采用以上模糊神经网络预报建模方法对20个独立样本进行预报试验,得到表1的预报结果。20个独立样本的冷湿指数实况平均值为51.84,预报平均值46.48,比实况值低5.36,平均绝对误差为14.76,平均相对误差30.63%。其中,PT预报值偏高7例,偏低13例。把相对误差分成0%~30%、30%~50%、50%~100%、>100%4个范围,分别有12、5、3、0例,集中于误差较小的范围。仅3例相对误差超过50%,最大94.85%,说明预报准确率高,预报性能稳定。从以上预报结果的误差分析来看,模糊神经网络预报建模方法对冷湿指数的预报误差较小,能作为广西区域低温冷害天气预报的借鉴方法。

表1 模糊神经网络预报模型20个独立样本预报结果

| 序号 | 实况值 | 预测值 | 误差 | 相对误差/% |
|----|-------|-------|--------|--------|
| 1 | 37.53 | 35.43 | -2.1 | 5.6 |
| 2 | 31.15 | 40.74 | 9.59 | 30.8 |
| 3 | 25.77 | 23.24 | -2.54 | 9.85 |
| 4 | 24.03 | 32.0 | 7.97 | 33.17 |
| 5 | 23.07 | 40.65 | 17.57 | 76.16 |
| 6 | 86.91 | 30.39 | -56.52 | 65.03 |
| 7 | 94.07 | 56.41 | -37.67 | 40.04 |
| 8 | 53.06 | 66.83 | 13.77 | 25.96 |
| 9 | 21.19 | 41.29 | 20.1 | 94.85 |
| 10 | 32.57 | 38.01 | 5.44 | 16.69 |
| 11 | 52.44 | 34.81 | -17.63 | 33.62 |
| 12 | 51.56 | 39.13 | -12.43 | 24.1 |
| 13 | 40.27 | 59.84 | 19.58 | 48.61 |
| 14 | 59.62 | 42.42 | -17.2 | 28.85 |
| 15 | 57.99 | 53.28 | -4.71 | 8.12 |
| 16 | 42.79 | 40.8 | -1.99 | 4.66 |
| 17 | 73.4 | 53.96 | -19.44 | 26.49 |
| 18 | 87.52 | 76.02 | -11.5 | 13.14 |
| 19 | 78.87 | 75.93 | -2.94 | 3.73 |
| 20 | 62.97 | 48.47 | -14.51 | 23.03 |
| 平均 | 51.84 | 46.48 | 14.76 | 30.63 |

2.2 遗传-神经网络冷湿指数集成预报

2.2.1 遗传-神经网络预报建模

本文设计采用遗传算法优化三层BP神经网络^[17]模型,首先将三层BP网络模型的各层节点、连接权、阈值按顺序组成一个遗传个体;进一步利用实数与二进制相混合编码方法对种群个体进行优化计算,构建了遗传-神经网络集合预报模型。其中关键细节具体如下:

(1) 适应度函数计算

首先对每个遗传个体进行解码, 再将其与输入训练样本一起进行隐层激活函数输出计算:

$$b_i = f_i(\sum_{h=1}^m a_h v_{hi} + \theta_i) \quad (8)$$

和网络输出计算:

$$\hat{y}_j = f(\sum_{i=1}^p w_{ij} b_i + \gamma_j) \quad (9)$$

式中: v_{hi} 为输入层至隐层连接权系数, 而 w_{ij} 为隐层至输出层的连接权系数, θ_i 和 γ_j 分别为输入层至隐含层和隐含层至输出层的阈值, m 为输入节点个数, p 为隐节点个数。则适应度函数定义为模型总体误差 E 的倒数:

$$F(x) = \frac{1}{E} = \frac{1}{\sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}} \quad (10)$$

(2) 进化操作计算

选择算子是采用轮盘赌选择计算, 其计算公式为:

$$P_k = F_k(x) / \sum_{k=1}^K F_k(x) \quad (11)$$

式中: P_k 为每个遗传个体的概率值, $F_k(x)$ 为每个遗传个体的适应度, 种群数为 K 。通过比较, 可以在概率条件下保证每一个适应度大的个体以较大可能遗传到下一代。交叉算子是以概率 P_c 进行多点交叉; 变异算子则是以概率 P_m 对遗传个体的基因与其它遗传个体作等位的基因替换。

采用上述三个遗传算子进行反复的进化操作进算, 直到进化到事先设定的 N 代后结束。最后对进化到第 N 代的遗传种群的每个个体进行解码, 获得全部神经网络个体。对每个个体赋予相同的权重, 得到遗传 - 神经网络集合预报模型的预测值。

2.2.2 遗传 - 神经网络预报结果

利用遗传 - 神经网络集合预报模型对 20 个独立样本进行预测, 预测结果见表 2。20 个独立样本预报的 PT 平均值是 60.22, 比实况值高 8.38, 其中 16 例预报值比实况值高, 4 例比实况值低。平均绝对误差 17.77, 平均相对误差 44.6%。相对误差在 0% ~ 30%、30% ~ 50%、50% ~ 100%、> 100% 的范围分别有 10、5、1、4 例, 低误差个例较多, 但也存在少量高误差个例。从误差分析可见, 该方法的预测值存在一定的规律性, 总体趋势偏高, 利用这一特征加以订正, 则可以提高其在实际业务中的参考性。

2.3 粒子群 - 支持向量机冷湿指数集成预报

2.3.1 支持向量回归方法

在处理实际天气预报等一些非线性问题时, 可在支持向量回归^[18]算法 (Support Vector Regression, SVR) 中利用核函数将非线性问题转化为线性问题进行回归估计。

表 2 遗传 - 神经网络集成预报模型 20 个独立样本预报结果

| 序号 | 实况值 | 预测值 | 误差 | 相对误差/% |
|----|-------|-------|--------|--------|
| 1 | 37.53 | 70.72 | 33.19 | 88.45 |
| 2 | 31.15 | 36.89 | 5.74 | 18.43 |
| 3 | 25.77 | 30.95 | 5.18 | 20.1 |
| 4 | 24.03 | 33.87 | 9.84 | 40.95 |
| 5 | 23.07 | 62.29 | 39.22 | 169.96 |
| 6 | 86.91 | 49.19 | -37.72 | 43.4 |
| 7 | 94.07 | 52.42 | -41.65 | 44.28 |
| 8 | 53.06 | 39.07 | -13.99 | 26.37 |
| 9 | 21.19 | 44.98 | 23.79 | 112.25 |
| 10 | 32.57 | 69.32 | 36.75 | 112.85 |
| 11 | 52.44 | 73.9 | 21.46 | 40.93 |
| 12 | 51.56 | 52.06 | 0.5 | 0.97 |
| 13 | 40.27 | 81.86 | 41.59 | 103.29 |
| 14 | 59.62 | 65.61 | 5.98 | 10.03 |
| 15 | 57.99 | 76.54 | 18.55 | 31.99 |
| 16 | 42.79 | 45.01 | 2.22 | 5.19 |
| 17 | 73.4 | 74.68 | 1.28 | 1.75 |
| 18 | 87.52 | 90.72 | 3.2 | 3.65 |
| 19 | 78.87 | 91.82 | 12.95 | 16.42 |
| 20 | 62.97 | 62.47 | -0.5 | 0.79 |
| 平均 | 51.84 | 60.22 | 8.38 | 44.6 |

设训练样本集 $T = \{x_i, y_i, i = 1, 2, \dots, n\}$, 则非线性估计函数为

$$f(x) = \omega \varphi(x) + b \quad (12)$$

其中 $\omega \in R^d$ 为权值向量, $b \in R$ 为阈值。引入 ε - 不敏感损失函数:

$$L_\varepsilon(f(x), y) = \max\{0, |y - f(x)| - \varepsilon\} \quad (13)$$

通过拉格朗日变换, 可求得最优 lagrange 乘子 a_i 和 a_i^* 。设支持向量为 $(a_i - a_i^*) \neq 0$ 的训练样本, 支持向量个数为 S , 最后回归函数转化为:

$$f(x) = \sum_{i=1}^S (a_i - a_i^*) K(x_i, x_j) + b \quad (14)$$

2.3.2 粒子群 - 支持向量回归模型

由于在 SVR 模型中的惩罚参数 C 、核函数参数 σ 和不敏感损失函数 ε 的选择缺少很好的理论指导计算方法, 往往靠人工经验或测试的经验方法来原因。为此本文尝试利用粒子群算法对 SVR 模型中待定的 3 个参数进行优化选择^[19]。在进化计算开始时, 利用均匀分布对种群进行初始化, 第 i 个粒子的位置向量和飞行速度分别为 $X_i = (X_{i1}, X_{i2}, \dots, X_{iD})$ 和 $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ 。每个粒子通过跟踪个体最优解 $P_i = (P_{i1}, P_{i2}, \dots, P_{iD})$ 和全局最优解 $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})$, 并根据以下公式进行反复的更新计算。

$$v_{ij}^{(t+1)} = \omega^{(t)} v_{ij}^{(t)} + u_1 r_1^{(t)} (p_{ij}^{(t)} - X_{ij}^{(t)}) + u_2 r_2^{(t)} (p_{gj}^{(t)} - X_{ij}^{(t)}); \quad (15)$$

$$\omega^{(t)} = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \cdot t; \quad (16)$$

$$X_{ij}^{(t+1)} = v_{ij}^{(t+1)} + X_{ij}^{(t)} \quad (17)$$

式中: $i = 1, 2, \dots, m$, m 为种群个数; 其中惯性

权重的最大值和最小值分别为 ω_{\max} 与 ω_{\min} ；当前迭代次数和最大迭代次数分别为 t 与 t_{\max} ；通常加速因子 $u_1, u_2 \in [0, 2]$ ； r_1 和 r_2 是 $0 \sim 1$ 之间均匀分布的随机数。

反复进化计算到第 T_{\max} 代(事先设定)后, 对最后一代粒子种群个体进行解码得到 m 个支持向量回归个体, 并对每个支持向量机个体给予同样的权系数, 从而得到等权的粒子群-支持向量回归集合预报模型。

2.3.3 粒子群-支持向量回归集合预报结果

利用上述粒子群-支持向量回归集合预报模型对 20 例独立样本进行了预报试验, 预报结果见表 3。PT 指数预报平均值 56.3, 比实况值高 4.46, 其中 10 例样本预报偏高, 10 例样本预报偏低; 平均绝对误差 17.65, 平均相对误差 43.63%。相对误差在 $0\% \sim 30\%$ 、 $30\% \sim 50\%$ 、 $50\% \sim 100\%$ 、 $>100\%$ 的范围分别为 11、4、2、3 例, 低误差个例占绝大多数。从平均误差和逐个误差分布可见, 粒子群-支持向量回归集合预报模型可作为广西低温冷害天气预报的参考方法。

表 3 粒子群-支持向量机集成预报模型
20 个独立样本预报结果

| 序号 | 实况值 | 预测值 | 误差 | 相对误差/% |
|----|-------|--------|--------|--------|
| 1 | 37.53 | 115.86 | 78.33 | 208.72 |
| 2 | 31.15 | 33.13 | 1.99 | 6.38 |
| 3 | 25.77 | 25.94 | 0.16 | 0.63 |
| 4 | 24.03 | 23.95 | -0.08 | 0.32 |
| 5 | 23.07 | 58.45 | 35.38 | 153.31 |
| 6 | 86.91 | 48.74 | -38.17 | 43.92 |
| 7 | 94.07 | 57.88 | -36.19 | 38.47 |
| 8 | 53.06 | 28.84 | -24.22 | 45.65 |
| 9 | 21.19 | 48.16 | 26.97 | 127.26 |
| 10 | 32.57 | 58.51 | 25.94 | 79.64 |
| 11 | 52.44 | 50.43 | -2.0 | 3.82 |
| 12 | 51.56 | 47.06 | -4.5 | 8.72 |
| 13 | 40.27 | 65.29 | 25.03 | 62.15 |
| 14 | 59.62 | 43.61 | -16.02 | 26.86 |
| 15 | 57.99 | 76.59 | 18.6 | 32.07 |
| 16 | 42.79 | 50.31 | 7.52 | 17.56 |
| 17 | 73.4 | 69.54 | -3.86 | 5.25 |
| 18 | 87.52 | 88.7 | 1.18 | 1.34 |
| 19 | 78.87 | 77.3 | -1.57 | 1.99 |
| 20 | 62.97 | 57.62 | -5.35 | 8.49 |
| 平均 | 51.84 | 56.3 | 4.46 | 43.63 |

3 不同预报建模的对比分析

3.1 三种智能预报建模效果对比

从平均误差来看, 模糊神经网络预报建模明

显比另两种智能方法效果要好, 粒子群-支持向量机集成预报方法次之, 遗传-神经网络集成预报方法在三种智能方法中平均误差最大。FNN 预测结果的平均绝对误差和平均相对误差分别比 SVM-PSO 集成预报方法低 2.89 和 13%, 比 GA-ANN 预报方法低 3.01 和 13.97%。从误差的变化幅度来看, 三种智能方法绝大多数独立样本相对误差小于 50%, FNN 预报建模震荡最小, 相对误差介于 $3.73\% \sim 94.85\%$ 之间, GA-ANN 预报建模次之, 为 $0.79\% \sim 169.96\%$, SVM-PSO 集成预报建模震荡最大, 从 0.32% 跨到 208.72% , 所以稳定性仍然是模糊神经网络预报建模最好。

3.2 多元线性回归方法对比分析

以上三种方法是非线性智能方法对低温冷害天气的探索预报, 为分析其与传统线性预测方法的优劣, 进一步采用多元线性回归方法对冷湿指数进行预报。建立多元回归方程时, 与三种智能方法一样, 利用同样的 10 个预报因子、299 个建模样本和 20 个独立预报样本。取 $F=3$, 复相关系数为 0.684, 剩余标准差为 25.439, 回归方程为:

$$Y = -5.856 + 23.128X_1 + 0.272X_2 - 1.813X_3 - 648359.431X_4 - 1976133.790X_5 - 760971.949X_6 - 1193229.425X_7 - 3324878.825X_8 - 620305.585X_9 - 230218.981X_{10} \quad (18)$$

表 4 多元线性回归方程的冷湿指数独立样本预报结果

| 序号 | 实况值 | 预测值 | 误差 | 相对误差/% |
|----|-------|-------|--------|--------|
| 1 | 37.53 | 72.31 | 34.79 | 92.7 |
| 2 | 31.15 | 43.36 | 12.21 | 39.21 |
| 3 | 25.77 | 25.56 | -0.22 | 0.84 |
| 4 | 24.03 | 38.69 | 14.66 | 61.01 |
| 5 | 23.07 | 65.7 | 42.63 | 184.76 |
| 6 | 86.91 | 55.2 | -31.72 | 36.49 |
| 7 | 94.07 | 60.1 | -33.98 | 36.12 |
| 8 | 53.06 | 47.25 | -5.81 | 10.95 |
| 9 | 21.19 | 50.05 | 28.85 | 136.15 |
| 10 | 32.57 | 76.06 | 43.49 | 133.53 |
| 11 | 52.44 | 75.48 | 23.04 | 43.94 |
| 12 | 51.56 | 58.86 | 7.31 | 14.16 |
| 13 | 40.27 | 82.66 | 42.39 | 105.28 |
| 14 | 59.62 | 66.43 | 6.81 | 11.42 |
| 15 | 57.99 | 74.99 | 17.0 | 29.32 |
| 16 | 42.79 | 46.31 | 3.52 | 8.23 |
| 17 | 73.4 | 71.81 | -1.59 | 2.16 |
| 18 | 87.52 | 85.46 | -2.06 | 2.36 |
| 19 | 78.87 | 88.76 | 9.89 | 12.54 |
| 20 | 62.97 | 63.61 | 0.64 | 1.02 |
| 平均 | 51.84 | 62.43 | 10.59 | 48.11 |

表 4 是多元线性回归预测方法的预测结果, 20 例冷湿指数预报平均值是 62.43, 比实况值高 10.59, 是 4 种方法中与实况值偏差最大的。平均绝对误差 18.13, 平均相对误差 48.11%, 也是 4 种方法中平均误差最大的, 分别比以上三种智能

非线性方法高 3.37、0.36、0.48 和 17.48%、3.51%、4.48%。相对误差在 0% ~ 30%、30% ~ 50%、50% ~ 100%、> 100% 的范围分别有 10、4、2、4 例, 大于 50% 的高误差个例是四种方法中最多的。相对误差最小为 0.84%、最大 184.76%, 变化幅度比粒子群 - 支持向量机集成预报方法稍小, 但远大于模糊神经网络方法, 也比遗传-神经网络方法大。

为了进一步比较三种智能非线性预报方法和多元线性回归方法对低温冷害天气预报的准确率和稳定度, 把 4 种预报方法的绝对误差分为 < 10、10 ~ 20、20 ~ 30、30 ~ 40、> 40 等 5 个等次, 如图 2 所示。在最小的绝对误差 < 10 的范围内, 4 种预报方法预报样本次数为 8 ~ 10 次, 其中粒子群-支持向量机集成方法次数最多为 10 次; 在次小误差 10 ~ 20 范围内, 模糊神经网络出现次数远比另三种方法多, 共有 9 次, 其余三种方法为 2 ~ 3 次。在最小的两个误差等级 0 ~ 20 中, 模糊神经网络方法有 17 例, 占预报总样本的 85%, 其余三种方法均是 12 例, 占总样本 60%。可见, 预报精度上, 模糊神经网络预报建模方法远好于其它三种方法。较大的绝对误差 20 ~ 30 和 30 ~ 40 范围内, 模糊神经网络方法各有 1 次, 另三种方法为 5 ~ 7 次。在最大的误差 > 40 等级中, 模糊神经网络方法和粒子群 - 支持向量机方法各有 1 次, 遗传-神经网络方法 2 次, 线性回归方法最多为 3 次, 分别占总预报样本数的 5%、10% 和 15%, 即在预报精度较低的高误差中线性回归方法占总预报样本数的比重最大。

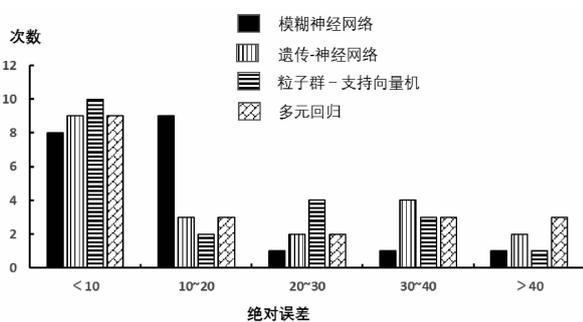


图 2 4 种预报方法 20 个独立样本预报结果绝对误差比较

综上所述, 在反映总体预报水平的平均误差中, 三种非线性智能计算预测方法均好于传统的多元线性回归预报方法; 而在表征预报准确度较高的低误差范围内, 模糊神经网络方法出现次数最多, 在预报准确率较低的高误差中, 线性回归方法次数最多。因而对于非线性的低温冷害天气的预报, 多元线性回归方法不能很好地体现大气运动的特点, 存在一定的局限性。而三种智能预报方法因为运用非线性智能计算方法, 能够在一定

程度上体现低温冷害天气是由多尺度天气系统相互作用共同影响而产生的特点, 拥有比较大的优势, 从而做出更准确的预报。

4 小结

(1) 为了探索不同的非线性智能计算预报建模方法在广西冬季低温冷害天气预报中的应用效果, 把 1951-2011 年的 299 例冷湿指数作为预报建模样本, 2012-2018 年的 20 例冷湿指数作为独立样本进行预报试验。挑选与冬季气温和降水关系密切的气温场、高度场、湿度场等物理量场中的高相关因子, 经过逐步回归和核主成分分析方法筛选出 10 个因子作为预报因子。

(2) 分别使用模糊神经网络方法 (FNN)、遗传算法的神经网络集成 (GA-ANN) 方法以及粒子群算法的支持向量机 (SVM-PSO) 集成预报方法三种非线性智能计算预报建模方法建立冷湿指数预报模型。试验结果表明, 三种方法的平均绝对误差分别为 14.76、17.77 和 17.65, FNN 误差远比另两种方法小, 而 SVM-PSO 集成预报方法误差比 GA-ANN 方法稍小, 即 FNN 效果最好, SVM-PSO 次之, GA-ANN 在三种方法中稍差。逐个误差分析显示, 预报精度和稳定度仍然是 FNN 最好, SVM-PSO 和 GA-ANN 相差不多。

(3) 利用多元线性回归方法对同样的建模样本和预报因子进行预报, 20 例独立样本预报的平均绝对误差为 18.13, 比三种智能预报建模方法高, 在高误差范围内出现的预报个例也最多, 因而预报的总体效果、预报精度和稳定度比三种非线性智能方法要差。

低温冷害天气是多尺度天气系统相互作用引起的极端灾害天气, 是大气的非线性运动过程, 相比于传统的多元线性回归方法, 非线性智能方法预报建模能更好地体现天气系统非线性运动的特点, 从而作出更准确的预报。今后将尝试运用这三种智能计算建模方法, 结合数值预报产品, 在实际业务中进行预报, 并加以跟踪检验订正, 以提高其在广西低温冷害天气预报中的参考性。

参考文献

- [1] DING Yihui, WANG Zunya, SONG Yafang, et al. Causes of unprecedented freezing disaster in January 2008 and its possible association with global warming [J]. Acta Meteor Sinica, 2008, 22 (4): 538-557.
- [2] 李崇银, 杨辉, 顾薇. 中国南方雨雪冰冻异常天气原因的分析 [J]. 气候与环境研究, 2008, 13(2): 113-122.
- [3] YANG Guiming, KONG Qi, MAO Dongyan, et al. Analysis on the Long-Lasting freezing rain and snowstorm event at the beginning of 2008 [J]. Acta Meteor Sinica, 2010, 24(3): 380-396.

- [4] WANG Donghai, LIU Chongjian, LIU Ying, et al. A preliminary analysis of features and causes of the snow storm event over the southern areas of china in January 2008[J]. *Acta Meteor Sinica*, 2009, 23(3): 374-386.
- [5] 李艳, 张金玉, 李旭, 等. 两次典型极端低温过程低频特征分析[J]. *高原气象*, 2018, 37(5): 1341-1352.
- [6] 李向红, 陆虹, 伍经, 等. 影响广西两次低温灾害及物理机制对比分析[J]. *灾害学*, 2018, 33(4): 60-64, 77.
- [7] 史悦, 郑建盟, 张万诚, 等. 2016年1月云南低温雨雪灾害的原因分析[J]. *灾害学*, 2017, 32(4): 208-213.
- [8] 唐熠, 韦健, 周文志, 等. 广西重大低温雨雪冰冻天气过程概念模型分析[J]. *灾害学*, 2013, 28(2): 25-30.
- [9] 布和朝鲁, 彭京备, 谢作威, 等. 冬季大范围持续性极端低温事件与欧亚大陆大型斜脊斜槽系统研究进展[J]. *大气科学*, 2018, 42(3): 656-676.
- [10] 单机坤, 梁潇云, 孙林海, 等. 冬季极端低温日数预测方法研究[J]. *高原气象*, 2016, 35(6): 1609-1614.
- [11] 张伟, 江静. 我国冬季持续性低温事件预报模型的建立[J]. *气象科学*, 2016, 36(4): 517-523.
- [12] 陈官军, 魏凤英, 姚文清, 等. 基于低频振荡信号的中国南方冬半年持续性低温指数延伸期预报试验[J]. *气象学报*, 2017, 75(3): 400-414.
- [13] 陆虹, 翟盘茂, 覃卫坚, 等. 低温雨雪过程的粒子群-神经网络预报模型[J]. *应用气象学报*, 2015, 26(5): 513-524.
- [14] 周秀华, 陆虹, 金龙, 等. 适用于我国华南广西区域冬季的冷湿指数计算方法研究[J]. *灾害学*, 2019, 34(2): 31-36.
- [15] 陈刚, 丁慧玲. 基于主成分分析的模糊时间序列模型的平稳化算法[J]. *控制与决策*, 2017, 32(9): 1643-1648.
- [16] V Vapnik, O Chapelle. Bounds on error expectation for support vector machines[J]. *Neural Computation*, 2000, 12(9): 2013-2036.
- [17] 金龙. 神经网络气象预报建模理论与应用[M]. 北京: 气象出版社, 2004.
- [18] 黄小燕, 史旭明, 刘苏东, 等. 模糊神经网络方法在热带气旋强度预报中的应用研究[J]. *高原气象*, 2009, 28(6): 1408-1413.
- [19] 董良才, 徐子奇, 宓为建. 基于遗传算子粒子群算法的拖轮动态调度[J]. *数学的实践与认识*, 2012, 42(6): 122-133.

Comparative Analysis of Three Kinds of Intelligent Computational Prediction Models for Cold Damage Weather

HUANG Cuiyin¹, SHI Xuming², LU Hong¹, ZHOU Xiuhua¹ and JIN Long¹

(1. *Guangxi Climate Center, Nanning 530022, China;*

2. *Faculty of Science, Guilin University of Aerospace Technology, Guilin 541004, China)*

Abstract: In order to explore the application effects of different nonlinear intelligent computational forecast models in the prediction of cold damage weather in winter, the daily average temperature and precipitation data of Guangxi region to calculate the cold and wet index of 1951-2018 as the forecast amounts are used. The forecast factors are obtained by calculating the forecast amounts and the physical quantities in the previous reanalysis data. First, nine forecast factors are selected from the relevant primary predictors through the stepwise regression method. The unselected forecast factors are subjected to calculate of nonlinear dimensionality reduction by the kernel principal component, and the nuclear master with the largest variance contribution is selected. And then, ten forecast factors serve as input factors for prediction models. Three prediction models are established to predict 20 cold damage weather in 2012-2018 respectively, those are fuzzy neural network method (FNN), genetics algorithmic neural network integration (GA-ANN) forecast method and particle swarm optimization support vector machine (SVM-PSO) integrated forecast method. The forecast results show that FNN's average absolute error is the smallest for 20 independent samples under the same forecast modeling conditions. Further calculations also show that the average absolute errors of the three intelligent computational prediction models are smaller than the linear regression forecasting's with the same samples and factors. It indicates that the nonlinear intelligent forecasting modeling methods have better forecast abilities than the linear forecast method for the cold damage weather in winter, which has obvious nonlinear variation characteristics.

Key words: intelligent computing; forecast model; cold damage weather; Guangxi