

刘海兵,卢秋芳.GMDH神经网络在雷电预警中的应用[J].陕西气象,2023(5):74-79.

文章编号:1006-4354(2023)05-0074-06

# GMDH神经网络在雷电预警中的应用

刘海兵,卢秋芳

(上饶市气象局,江西上饶 334000)

**摘要:**利用江西南昌地区南昌县、小蓝经开区、塔城乡和小莲村等站点多个时间序列的大气电场和雷电数据,选取南昌地区2018—2020年56个雷电活动过程中共计224个时间序列大气电场和经小波函数sym5和阈值分析法Rigorous Sure变换的大气电场信号进行训练和测试,建立自组织数据处理的群方法GMDH(group method of data handling)神经网络模型并应用于雷电活动的预警。结果表明:30~60 min GMDH神经网络模型预测值和实测值相关系数( $R$ )在0.7~0.85之间,经过小波变换处理后的大气电场数据预测准确率更高,预测值与实测值的相关系数接近于1,均方误差、均方根误差以及准则值( $P$ )均可控制在5%以内;借助提前获取的大气电场预测数据,通过大气电场强度剧烈的脉冲变化可以预示着雷电活动的发生,有效地提高雷电预警的准确率和时效性。

**关键词:**GMDH神经网络;雷电预警;小波函数;准则值

**中图分类号:**P49

**文献标识码:**A

随着神经网络(一种复杂非线性系统的启发式自组织建模算法)的发展,对非线性时间序列数据进行预测,可以用神经网络的非线性结构来建模和预测。从非线性本质上来发掘这类数据的特性,是一种很有发展前景的预测途径<sup>[1]</sup>。吴耿锋<sup>[2]</sup>利用GMDH(group method of data handling)神经网络结构,将降雨量的混沌参数提取后体现在GMDH结构中,并用GMDH模型对降雨量进行了预测。章劲秋等<sup>[3]</sup>针对电力企业电能供应量与用户需求量之间匹配度低的问题,按照GMDH算法原理设计了电力供应量智能分析系统。系统可以为电力企业节省大量的电能损耗,提高了电力供需的匹配度。朱新国等<sup>[4]</sup>为预测灌区地下水水位,提出基于混沌优化的GMDH神经网络预测方法建立预测模型,对地下水埋深进行预测。该方法预测精度较高,可用于对地下水水位的预测。措姆等<sup>[5]</sup>基于提高流域尺度的区域中

长期气象干旱预报的准确度,采用支持向量机和数据处理组合方法(GMDH)两种数据挖掘方法来识别时空气象相关性,并通过这两种方法对预测流域尺度的气象干旱预报性能进行对比,GMDH具有更好的预测稳定性。

雷电灾害是最为严重的自然灾害之一,高时效性的雷电预警方法可以减少雷电对人民生命和财产造成的损失。大气电场有着极强的非线性特征,大气电场的变化是多种相互影响的自然因素共同作用的结果,其中某些因素成因不明,变化复杂,使得大气电场数据表现出不定性和复杂性。对这样的数据进行预测,尤其是定量预测相当困难。大气电场数据采集是在一定地点按照一定的时间间隔进行记录,是一种与时间直接相关的数据,对大气电场这样的非线性时间序列数据进行预测,可以用神经网络的非线性结构来建模和预测。基于南昌地区(南昌县站、小蓝经开区、塔城

**收稿日期:**2022-11-05

**作者简介:**刘海兵(1980—),男,汉族,江苏泰兴人,硕士,高级工程师,从事雷电监测预警预报、天气短临预报研究。

**通信作者:**卢秋芳(1971—),男,汉族,江西宜春人,学士,高级工程师。研究方向为天气预报服务,短临预报研究。

**基金项目:**2019年江西省气象局科研项目

乡和小莲村)多大气电场仪站点多个时间序列联合监测的大气电场资料和闪电定位仪监测资料,从研究深度学习雷电临近预警模型出发,选取南昌地区 2018—2020 年 56 个雷电活动过程中 224 个时间序列大气电场和经小波函数  $\text{sym5}$  和阈值分析法 Rigorous SURE 变换的大气电场信号进行训练和测试,建立自组织数据处理的群方法 GMDH 神经网络模型并应用于雷电活动预警,借助提前获取的大气电场预测数据,通过大气电场强度剧烈的脉冲变化可以预示着雷电活动的发生,有效地解决雷电预警效果时效性欠佳的问题。

## 1 资料来源

大气电场资料为江西省南昌市气象局提供的 2018—2020 年期间南昌地区地面电场仪(南昌县气象站、小蓝经开区、塔城乡以及小莲村)联合监测的大气电场时间序列数据,大气电场仪站点分布以及其主要参数指标见图 1 和表 1。

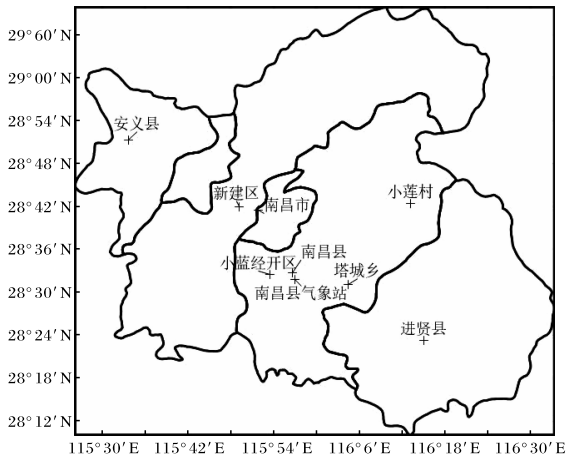


图 1 南昌地区大气电场仪站点分布图

(审图号为 GS(2019)3082 号)

表 1 南昌地区大气电场仪主要参数指标

性能指标	指标值
探测范围	0~±100 kV/m
探测距离( $R$ )	0~20 km
采样时间	1 s
测量误差	<5%

南昌地区闪电资料来源于 2018—2020 年江

西雷电监测网监测数据,包括闪电发生时间、地理位置(经、纬度)、闪电强度、闪电陡度、极性(正、负闪电)、定位站数和雷电流幅值等参数,用于筛选出南昌地区 2018—2020 年期间的 56 个雷电活动过程作为研究的样本数据。

## 2 基于 GMDH 神经网络的大气电场预测模型

### 2.1 GMDH 算法和分析步骤

GMDH 神经网络是一种复杂非线性系统的启发式自组织建模方法,是一种以多项式为基础通过不断筛选组合来辨识非线性系统的模型。主要是将网络的各输入端交叉组合产生一系列的活动神经元后<sup>[6]</sup>,再利用外准则从产生的一代神经元中选择若干与目标变量最接近的神经元,将被选出的神经元结合后产生新神经元,重复这个过程,直至新产生的一代神经元都比上一代神经元更加优秀<sup>[7-13]</sup>,则确认为最优模型被选出,迭代过程结束。主要步骤如下。

(1)确定样本数据,将输入输出样本数据划分为训练数据集和测试数据集。

(2)假定系统有  $m$  个输入变量  $x_i (i=1, 2, 3, \dots, m)$ ,一个输出变量  $y$ ,本文选取 K-G 多项式的离散形式作为参考函数(如式(1)),建立输入、输出变量关系, $a_0, a_i, a_{ij}, a_{ijk} \dots (i, j, k=1, 2, \dots, m)$ 为多项式的系数。

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (1)$$

(3)确定外准则:选取两个准则函数作为选取外准则,分别为以下两种。

#### ①预测误差平方和( $P$ )准则

$$P = \sum_{t=1}^m [y(t) - \hat{y}(t)]^2 \quad (2)$$

#### ②最小信息准则( $A$ )

$$A = m \ln S_k^2 + C + 2k \quad (3)$$

$$S_k^2 = \frac{1}{m} \sum_{t=1}^m [y(t) - \hat{y}(t)]^2 \quad (4)$$

式中  $\hat{y}(t)$  为中间模型在第  $t$  个样本数据的输出估计值, $y(t)$  为第  $t$  个样本数据的实际输出值, $m$  为样本的个数, $C$  为常量, $k$  为独立可调参数。

(4)将所选择参考函数中的每一项作为算法初始输入变量,以2018年6月30日大气电场样本数据为例, $x_1, x_2$ 为输入变量, $f(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2$ 作为K-G多项式,输入变量为: $u_1 = a_1x_1, u_2 = a_2x_2, u_3 = a_3x_1^2, u_4 = a_4x_2^2, u_5 = a_5x_1x_2, u_6 = a_0$ 。

(5)自组织自适应建立 $y_k = f_k(u_i, u_j)$ 第一层中间模型,再根据内准则运用最小二乘估计在训练数据集上算出 $y_k$ 的参数<sup>[14]</sup>。

(6)选择最优复杂度模型:在测试数据集时,利用步骤(3)所确定的外准则对上层中间模型进行筛选,选择外准则值小的模型作为下一层的输入变量,不断重复进行。

(7)将选择外准则值最低或不再减小的模型作为最优模型<sup>[15-16]</sup>。

## 2.2 大气电场资料质量控制

受大气电场仪自身灵敏度和安装环境等因素的影响,实测的大气电场信号中常夹杂着一些其他非雷暴电场信号,也就是噪声信号,会对大气电场资料的分析及应用造成很大的困难。在具体应用大气电场资料时,需要对原始信号进行小波变换处理,保留雷暴天气电场的变化趋势和雷电引起的电场变化。利用小波分析方法对大气电场信号进行去噪,经过多种小波函数和阈值分析方法的对比后,利用小波函数sym5和阈值分析法Rigorous Sure对原始时间序列的电场信号作小波变换<sup>[17-19]</sup>。选取2018年6月30日00—24时南昌县气象站大气电场数据进行个例分析,该时段雷电活动过程较为频繁,且原始大气电场信号上叠加有一定的噪声,采用上述小波分析方法处理后,得到小波变换后的信号波形(图2)。由图2可见,大气电场信号在进行小波变换处理后,波形重叠部分明显减少,波形更为平滑且突出显示了雷电活动过程中大气电场的主要变化趋势。表明采用该方法可以较为显著地降低信号中的噪声,为数据的后期分析和处理起到质量控制的作用。

## 2.3 GMDH神经网络模型仿真试验

GMDH神经网络是将观测样本分为训练数据和测试数据,并进行数据分组,在训练样本上建

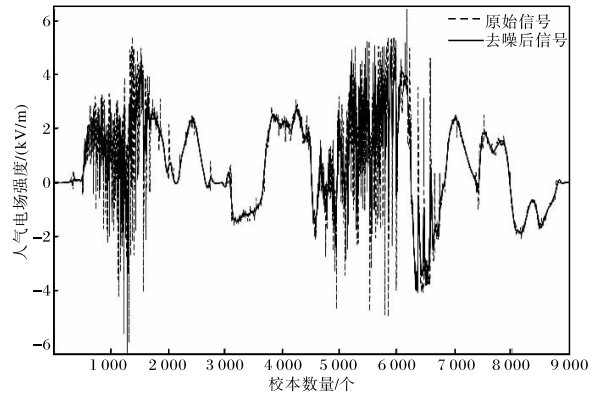


图2 2018-06-30T00—24南昌县气象站大气电场原始信号和小波变换去噪后信号叠加图

立模型,并利用测试数据通过外准则来进行模型筛选。基于上述理论研究,将南昌地区2018—2020年发生的56次强对流天气过程共计224个时间序列的大气电场数据和经小波函数变换后的电场数据作为研究样本,通过充分的样本数据量,建立地面大气电场GMDH神经网络预测模型。分别将原始信号和小波变换信号的50%样本数据集作为训练数据,50%的数据作为测试数据<sup>[20]</sup>。训练数据用于建模,测试数据仅用于选择最优复杂度模型<sup>[21]</sup>。经检验,所有站的大气电场预测值和实测值相关系数 $R$ 在0.8~1.0之间,均方误差(MSE,用 $E$ 表示)、均方根误差(RMSE)以及准则值( $P$ )均可控制在5%以内,是可接受水平。下面基于GMDH神经网络模型,以2018年6月29—30日雷暴天气过程中南昌县站的大气电场数据作为样本数据集进行个例分析。

在输入数据样本中取出50%的数据作为建模样本,从样本中的 $k$ 个数据中任取两个 $x_i, x_j$ , $y$ 为输出,用最小二乘法建立如2.1节中的公式(1)的处理单元,可以提取出 $k(k-1)/2$ 个基本单元,将当前样本所有基本单元输出方差的均值设为阈值 $E_g$ ,这样在第一步的处理中就可以将误差较大的单元剔除,将余下的样本中相应变量分别带入上述处理单元,算出处理单元的输出与实际输出的方差 $E$ ,将各单元的方差 $E$ 与 $E_g$ 比较,保留方差低于阈值的单元并记录这些单元中的最小方差,这样就得到了第一层的单元<sup>[22]</sup>。将全部数据中的相应变量代入第一层单元进行计算,得到

第一层的输出  $Y$ ; 将  $Y$  作为第二层的输入, 重复上述两步, 可以得到第二层的  $k$  个处理单元, 第二层的输出  $Y$  以及第二层的最小方差  $E_m$ , 若第二层的最小方差小于第一层, 则第二层建立成功, 继续进行下一层。假设进行到第  $L+1$  层, 发现该层的最小方差相对于上一层更小, 且外准则值最低或不再减小时, 则终止建模, 并将第  $L$  层中方差最小的那个单元作为输出单元。最后, 将与输出单元相关的上层单元逐层连接, 其余无关单元被排除在网络结构外, 至此, 得到了以样本为基础的 GMDH 神经网络<sup>[23]</sup>。GMDH 网络的结构是在训练中确定的, 通过训练可以舍弃那些性能不好的神经元<sup>[24]</sup>。

将 2018 年 6 月 29 日 23:00—30 日 02:30 雷暴天气过程中南昌县气象站的大气电场数据作为样本数据集输入网络模型后, 训练数据集为 6 月 29 日 23:00—30 日 01:30 的大气电场数据, 6 月 30 日 01:30—02:30 的数据为测试数据集, 每次顺次输入 2 个数据, 并将预测的第 3 个数据与实际值比较, 预测和误差值仿真的变化曲线见图 3。实验结果表明, GMDH 神经网络模型的大气电场信号预测与实测值基本相符(图 4)。从图 4 可以看出, 后半个小时相对前半个小时预测结果值的误差值增大, 但是大气电场的变化趋势与实测值基本一致。从大气电场信号预测与实测值的相关性检验结果(图 5)可以看出, GMDH 神经网络模型在 01:30 之后的大气电场预测准确率较高, 相关系数接近 0.9。

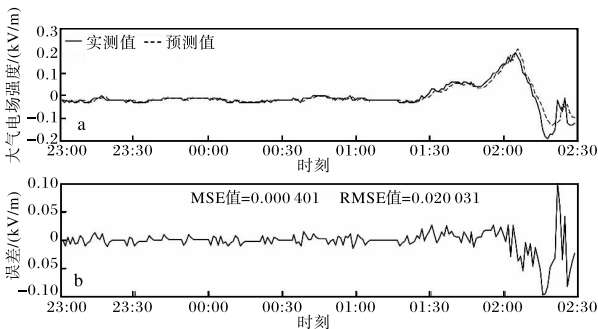


图 3 2018-06-29T23:00—2018-06-30T02:30 南昌县气象站 GMDH 网络模型训练和测试阶段预测和实测大气电场强度(a)及误差(b)仿真曲线

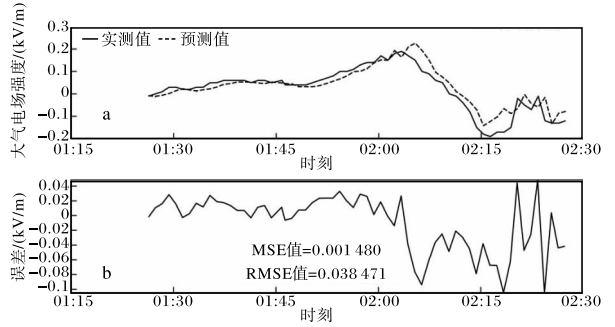


图 4 2018-06-30T01:30—02:30 南昌县气象站 GMDH 网络模型预测阶段预测和实测大气电场强度(a)及误差(b)仿真曲线

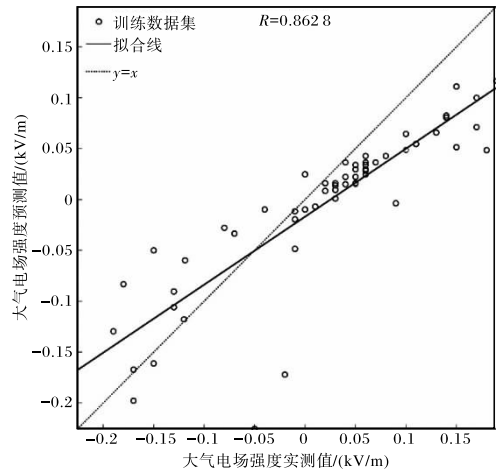


图 5 南昌县气象站 GMDH 网络模型大气电场预测和实测值相关性检验

利用小波函数 sym5 和 Rigorous SURE 阈值分析法, 对南昌县气象站 2018 年 6 月 29 日 23:00—30 日 02:30 的原始大气电场信号作小波变换, 再通过构造完成的 GMDH 神经网络模型进行训练和测试后, 其训练和测试阶段的预测和误差值仿真曲线见图 6。从 GMDH 神经网络模型 01:30 之后的大气电场信号预测及误差值仿真曲线(图 7)可看出, 相对于未经过小波变换的大气电场信号, 预测与实测值更加相符, MSE 和 RMSE 均可控制在 5% 以内, 在可接受水平之内。从经过小波变换后的大气电场信号预测与实测值的相关性检验结果(图 8)可以发现, 预测准确率相对小波变换前高,  $R$  已接近于 1。

将 6 月 30 日 01:30—02:30 南昌县气象站大气电场仪 15 km 范围内发生的雷电活动与大气电场预测值相结合分析发现, 01:30—01:50 时间

段内大气电场变化较为平稳,电场仪 15 km 范围内未出现雷电活动,01:50—02:30 间段内大气电场强度出现大幅波动,说明附近条件成熟时雷云中电荷量瞬时发生突变,引起地面大气电场形成脉冲变化,在 02:25—02:30 发生首次雷电。

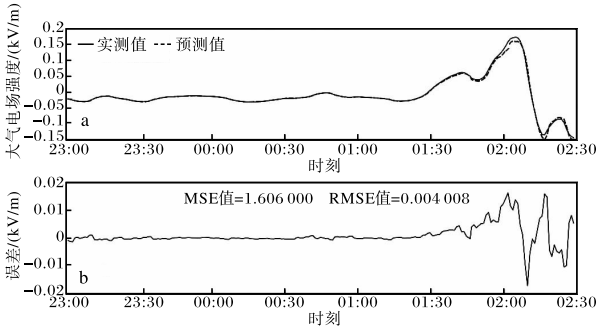


图 6 2018-06-29T23:00—2018-06-30T02:30 南昌县气象站 GMDH 网络模型训练和测试阶段预测和实测大气电场强度(a)及误差(b)仿真曲线(小波变换)

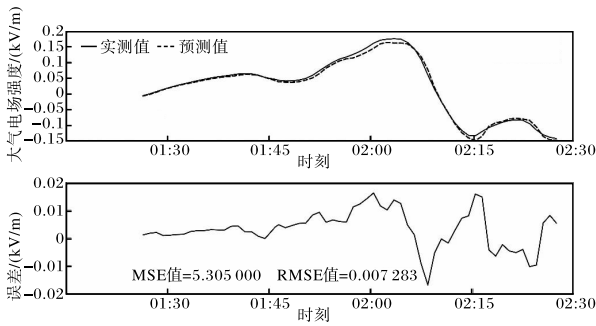


图 7 2018-06-30T01:30—02:30 南昌县气象站 GMDH 网络模型预测阶段预测和实测大气电场强度(a)及误差(b)仿真曲线(小波变换)

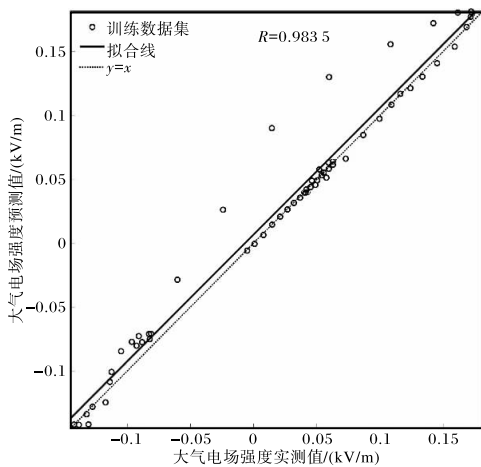


图 8 南昌县气象站 GMDH 网络模型预测和实测值相关性检验(小波变换)

图 9 为南昌县气象站 GMDH 网络准则值曲线。从图 9 可以看出,网络模型训练层数为 3 层,其外准则值收敛到第 2 层时已经达到最小值,找寻全局神经元的输出误差均方差最小值是这一学习算法的最佳停止准则,因此从第二层逆向递推得到最优数学模型。

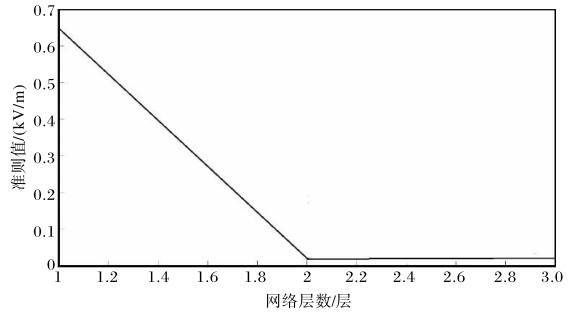


图 9 南昌县气象站 GMDH 网络准则值曲线

### 3 结论

本文基于江西南昌地区多站点多时间序列的大气电场,以研究深度学习雷电临近预警模型为目的,选取南昌地区 2018—2020 年 56 个雷电活动过程中共计 224 个时间序列的原始大气电场信号和经小波函数变换的大气电场信号进行训练和测试,建立 GMDH 神经网络模型并应用于雷电活动的预警,在神经网络模型对大气电场信号的训练、测试、预测和相关性检验过程中,总结得出以下结论。

(1)经过多种小波函数和阈值分析方法的对比后,发现采用小波函数 sym5 和 Rigorous SURE 阈值分析法对时间序列的原始电场信号进行小波变换,为大气电场数据的后期分析和处理起到质量控制作用。

(2)在已建立的 GMDH 神经网络模型的训练和预测过程中,均方误差(MSE)、均方根误差(RMSE)、相关系数(R)和准则值(P)是检验预测效果的有效参数。经过小波变换后的大气电场信号,GMDH 网络模型预测和实测值相关性 R 更接近 1,预测效果更好;MSE、RMSE 以及 P 均可控制在 5%以内,是可接受的水平。

(3)通过南昌地区 2018—2020 年发生的 56 次强对流天气过程共计 224 个时间序列的大气电场样本数据建立的神经网络模型,预测准确率在

0.7~0.85之间,预测时间在0.5~1h之间,预测的时间越短,预测准确率越高,经过小波变换处理后的数据GMDH神经网络模型预测准确率更高。

(4)本文是基于南昌地区大气电场数据作为建立神经网络的样本数据,此神经网络预测模型在其他地区的预测效果有待于进一步开展研究。

#### 参考文献:

- [1] 罗威,罗焯泓,王威. 基于机器学习的韶关地区短期日平均气温研究[J]. 陕西气象,2022(6):21-25.
- [2] 吴耿锋,彭虎,储阅春,等. 具有混沌特征的GMDH神经网络在降雨量预测中的应用[J]. 小型微型计算机系统,2000(2):24-26.
- [3] 章劲秋,俞阳,朱君,等. 基于GMDH算法的电力供需量智能分析系统设计[J]. 电子设计工程,2019,27(15):113-117.
- [4] 朱新国,张展羽,刘莉. 基于混沌优化GMDH网络的灌区地下水水位预测[J]. 河海大学学报(自然科学版),2010,38(3):317-321.
- [5] 措姆,加勇次成,红梅. 利用数据挖掘方法探索流域尺度气象干旱预报的研究[J]. 四川环境,2018,37(4):65-70.
- [6] 崔霞,高建华. 基于GMDH神经网络的组合预测模型[J]. 小型微型计算机系统,2016,37(6):1164-1167.
- [7] 陈建平,杨宜民,张会章,等. 一种基于GMDH模型的神经网络学习算法[J]. 云南大学学报(自然科学版),2008(6):569-574.
- [8] 张瑛. 基于自组织数据挖掘的应用技术成果转化系统研究[D]. 南京:南京航空航天大学,2012:16.
- [9] 陈小强,许仙珍,蔡璐璐,等. 基于GMDH神经网络的超临界机组过热蒸汽温度预测模型及仿真研究[J]. 热力发电,2014,43(6):102-107.
- [10] 肖进,贺昌政. 基于动态分类器集成的客户流失预测模型研究[J]. 预测,2010,29(5):57-62.
- [11] 潘国荣,谷川. GMDH神经网络算法在变形预测中的应用[J]. 大地测量与地球动力学,2008(3):54-58.
- [12] 白丽华. 自组织数据挖掘在能源消耗预测中应用[J]. 商场现代化,2006(17):13-14.
- [13] 周晓华,黄玲,刘胜永. GMDH方法在中长期电力负荷预测中的应用[J]. 电气应用,2008(6):67-69.
- [14] 唐志恩. 数据挖掘在经济预测中的应用研究[D]. 成都:电子科技大学,2008:128.
- [15] 韩愈. 基于改进GMDH的应用技术成果转化系统模型构建[D]. 南京:南京航空航天大学,2014.
- [16] 周莉. GMDH结构突变搜索模型及实证研究[D]. 成都:电子科技大学,2006:35.
- [17] 余海,张廷龙,刘文杰,等. 基于小波分析对地面电场信号的去噪及闪电信息提取[J]. 海南大学学报(自然科学版),2020,38(1):67-74.
- [18] 李鹏,郑毅,张义军. 闪电瞬态电场信号波形去噪方法[J]. 强激光与粒子束,2007(12):2055-2059.
- [19] 余蓉,杜牧云,杜九三. 大气电场仪资料的小波去噪研究[J]. 成都信息工程大学学报,2016,31(2):175-179.
- [20] 杨良怀,柳乔凡,范玉雷. 运动腕表心率序列相似性检测[J]. 浙江工业大学学报,2019,47(5):581-590.
- [21] 李小飞. 基于自组织数据挖掘的GDP预测研究[D]. 杭州:浙江工商大学,2012:41.
- [22] 王佳兴. 基于GMDH方法的设备剩余寿命预测[D]. 武汉:武汉科技大学,2010:22.
- [23] 张慧,刘湘南,黄刚. 基于模拟退火遗传算法的GMDH网络模型[J]. 华中师范大学学报(自然科学版),2013,47(2):162-166.
- [24] 陈洪,陈森发. 基于模糊GMDH网络的交通流量预测模型[J]. 南京理工大学学报(自然科学版),2010,34(1):46-50.