

基于 AGA-DBSCAN 优化的 RBF 神经网络构造煤厚度预测方法

吴善杰 王新

中国矿业大学计算机科学与技术学院 江苏 徐州 221116

(1490354963@qq.com)



摘要 在构造煤厚度的预测中,经常出现因各种限制性因素而导致预测精度不高的问题,因此提出了利用自适应遗传算法优化密度聚类(DBSCAN)优化 RBF 神经网络参数的方法对构造煤厚度进行预测。首先,对采区三维地震属性数据进行预处理,采用主成分分析算法(PCA)对该数据降维并消除变量之间的线性相关性。然后,构建预测构造煤厚度的 RBF 神经网络模型,并利用 DBSCAN 获取最佳核心点数据,通过计算得到 k -means 聚类的初始聚类中心,以此优化 k -means 算法,进而得到 RBF 神经网络隐含层基函数最优的中心向量,提高该模型预测的精准性和鲁棒性。同时,针对遗传算法存在容易陷入局部最优的问题,通过随着进化次数的增多自适应地改变交叉率和变异率来改善遗传算法的全局和局部搜索能力,使之逃离局部最优,获得更优的进化结果。此外,为了增强模型的泛化能力,对模型权重参数加入了 L2 正则化项,有效避免了噪声对模型泛化能力的影响。最后,将该模型应用到芦岭煤矿 II 六采区 8# 煤层中,模型预测构造煤的厚度与实际地质资料具有较高的一致性。因此,所提构造煤厚度预测模型的实际预测精度较高,误差较小,可以推广到实际采区构造煤厚度的预测。

关键词: 主成分分析;遗传算法;密度聚类;中心向量;RBF 神经网络;构造煤;厚度预测

中图分类号 TP183

Prediction of Tectonic Coal Thickness Based on AGA-DBSCAN Optimized RBF Neural Networks

WU Shan-jie and WANG Xin

School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

Abstract In the prediction of tectonic coal thickness, the problem of low accuracy is often caused by various restrictive factors. Therefore, a method of optimizing the parameters of RBF neural networks by using adaptive genetic algorithm to optimize density clustering is used to predict the thickness of tectonic coal. Firstly, the 3D seismic attribute data of the mining area are preprocessed, and the PCA algorithm is used to reduce the dimension and eliminate the linear correlation between variables. Then a RBF neural network model for predicting the thickness of tectonic coal is constructed, the genetic algorithm is used to optimize the density clustering to obtain the best core point, and the initial clustering center of k -means clustering is further calculated to optimize the k -means algorithm, so that the RBF neural network implicit layer basis function is obtained. An excellent center vector increases the accuracy and robustness of the model prediction. At the same time, aiming at the problem that genetic algorithm is easy to fall into local optimal problem, the global and local search ability of the genetic algorithm is improved by adaptively changing the crossover rate and the mutation rate with the increase of the number of evolutions, so that it can escape the local best advantage and obtain better evolutionary results. The L2 regularization term is added to effectively avoid the influence of noisy data for generalization performance of the model. Finally, the proposed model is applied to the 8# coal seam of the No. 6 mining area of Luling Coal Mine. The predicted thickness of the model is highly consistent with the actual geological data. It is possible to promote the prediction of coal thickness in actual mining area.

Keywords Principal component analysis, Genetic algorithm, Density clustering, Center vector, Radial basis function neural network, Tectonic coal, Thickness prediction

到稿日期:2020-08-18 返修日期:2020-09-17 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(41704115,41774128);江苏省自然科学基金(BK20170273)

This work was supported by the National Natural Science Foundation of China(41704115,41774128) and Natural Science Foundation of Jiangsu Province(BK20170273).

通信作者:王新(wxgrin@cumt.edu.cn)

1 引言

煤与瓦斯突出是矿井三大主要动力灾害之一,主要表现为煤和瓦斯在井下瞬间涌出等现象,本质就是矿井采煤过程中所发生的一种复杂动力现象。构造煤(tectonic coal)是指煤层在构造应力作用下发生挤压、剪切、变形、破坏或强烈的韧塑性变形及流变迁移,是地球演化的产物,表示地应力作用的记录^[1-2]。许多研究表明煤矿瓦斯突出区域与构造煤存在着必然的联系^[3]。由于矿井开采强度的增大,矿井深度逐步增加,使得我国发生煤矿瓦斯突出灾害的频率和严重程度一直处于较高水平^[4]。因此,可以通过准确地预测构造煤厚度来标注瓦斯突出区域,进而减少事故的发生,这种措施在煤矿安全生产中具有非常重要的意义^[5]。传统的构造煤厚度预测方法有钻孔取芯方法,但是该方法具有取芯低等明显的缺点,无法保证煤矿的安全生产要求;还有通过对煤层建立楔形模型进行分析,使用瞬时频率的尖脉冲进行构造煤厚度预测,但是该方法只能预测薄层厚度,也满足不了实际生产的需求。

在地震勘探中,地震属性(seismic attributes)是目前最主要的构造和岩性解释技术之一。当煤层中含有构造煤时,其弹性性质明显区别于正常煤层,含构造煤煤层中的地震属性也明显区别于不含构造煤煤层中的地震属性^[6-8]。因此,通过使用地震勘探技术获取构造煤相关的地震属性为预测矿区构造煤厚度提供了新的方法。已有研究表明,相关地震属性与构造煤的厚度之间具有非常复杂的非线性关系,而机器学习算法对于非线性特征关系的处理具有强大的可靠性能。径向基函数(Radial Basis Function, RBF)神经网络^[9]是一种具有单隐层的前馈型神经网络。该网络是将输入矢量扩展到高维空间的神经网络学习方法,具有准确的局部感知特性和强大的非线性逼近能力。RBF神经网络具有明显的优点,如更快的学习速率、不存在局部极小值以及较好的泛化性能等。但是传统的RBF神经网络也存在一些缺点,如隐含层节点数无法确定,中心向量参数不是最优从而导致预测效果不佳等问题。针对这些问题,本文通过引入混合聚类的思想来进行改进,进而构建优化的RBF神经网络预测模型。首先,利用遗传算法较好的寻优能力优化密度聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)关键参数得到最佳核心点数据。其次,通过层次聚类对最佳核心点数据构建凝聚层次树,以此获取最佳的 k -means聚类的初始聚类中心。然后,将 k -means聚类运算结果所获取的簇中心作为最优的RBF神经网络中心向量参数。在传统的遗传算法中,对变异和交叉概率通过使用平滑随进化次数的增加而改变的方法以获取更加健硕的进化结果。最后,将地震属性数据输入所构建的改进型RBF神经网络预测模型中,以提高采区构造煤厚度预测的精度和可靠性。

2 基本原理

2.1 遗传算法的基本原理

在DBSCAN中,参数eps和min_samples的确定经常出现不合理的情况,导致核心点对象大量增加或减少,进而使得聚类结果不稳定,因此需要对这两个参数进行优选。遗传算

法^[10](Genetic Algorithm, GA)是模拟生物学中自然遗传和达尔文进化理论而提出的并行随机优化算法,其特点是结合了适者生存和随机信息交换的思想,通过自然选择、交叉、变异等操作机制,实现种群的逐步进化。在寻优的过程中,遗传算法在解空间内随机产生多个种群基因并同时开始搜索,由适应度函数指导搜索方向,使其能够在复杂的搜索空间中快速寻求全局最优解。目前,遗传算法已经应用于很多领域,如参数的组合优化、图像的去噪处理、机器人路径优化以及数据的降维处理等。一般来说,遗传算法的执行过程可以视为一个循环,如图1所示。在所有的执行步骤中,编码设计、适应度函数设计和遗传操作最为关键。

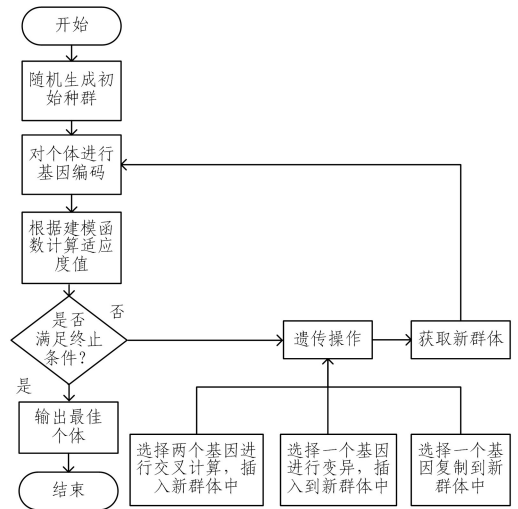


图1 遗传算法流程图

Fig. 1 Flowchart of GA

2.2 k -means 和 DBSCAN 聚类算法

k -means 聚类算法^[11]以 k 为参数,将数据集分为 k 个簇,并且使得每个簇内的对象具有较高的相似度,簇与簇之间距离最大化。相似度的计算则是根据一个簇中对象的平均值(被看作簇的中心)来进行。 k -means 聚类算法的核心思想为:首先从 m 个待聚类的样本中随机选取 k 个样本作为初始聚类中心,然后将其余的样本根据其与各中心的距离分配到最近的簇中,再进一步求出每个簇中样本的平均值作为新的中心。重复上述过程直到目标函数取得最优结果为止。

假设 x_j 表示簇号为 C_i 中的数据对象, c_i 表示簇号为 C_i 的中心(即簇中数据的平均值),则 k -means 聚类采用的目标函数即平方误差准则为:

$$E = \sum_{i=1}^k \sum_{x_j \in C_i} |x_j - c_i|^2 \quad (1)$$

其中 E 表示数据集中所有样本与其所在簇的中心的平方误差的综合。通过对该目标函数的优化,可以得到最终的聚类中心向量。

k -means 聚类算法的优点是聚类快速、容易实现,当数据集属于密集型且簇之间区别明显时,其效果较好,可以处理大型数据集;缺点是聚类个数 k 值通常需要预先给定,若 k 值给定不合理,则会导致聚类质量严重下降,并且 k -means 聚类算法对初始聚类中心点比较敏感,初始聚类中心选取的好坏会导致不同的聚类结果,好的聚类所选取的初始聚类中心应该是具有代表性的。传统的方法是随机选取初始聚类中心,虽

然方便快捷,但是不可避免会选到孤立点或样本边界点,进而容易导致聚类结果陷入局部最优,使得最终运算结果获取的簇中心不能成为最佳的 RBF 神经网络隐含层的中心向量参数。

DBSCAN 算法^[12-14]是从数据对象的分布密度出发,将密度邻近区域连接起来,只要邻近区域的密度(数据点的数目)超过某个阈值,就继续聚类,进而可以发现任意形状的聚类。有关概念介绍如下:

(1)数据密度。已知样本数据集 $D = \{x_1, x_2, \dots, x_n\}$, $D \in R^d$ 维的数据集,则将数据样本 x_i 的密度定义为 x_i 以 Eps 为邻域内所有样本的数量。数量越多说明该点的密度越大,反之密度越小。

(2) Eps 邻域。对于任意样本 $x_i \in D$, x_i 的 Eps 邻域 $N_{Eps}(x_i)$ 的定义为:以 x_i 为核心点,以 Eps 为半径所得到的 d 维超球体区域,可表示为:

$$N_{Eps}(x_i) = \{x_j \in D | 0 \leq \text{dist}(x_i, x_j) \leq Eps, j = 1, 2, \dots, n\} \quad (2)$$

其中 $\text{dist}(x_i, x_j)$ 表示数据集 D 中样本 x_i, x_j 之间的距离,用该距离来描述两个样本之间的相异度,表示如下:

$$\text{dist}(x_i, x_j) = \sqrt{(x_i^{(1)} - x_j^{(1)})^2 + (x_i^{(2)} - x_j^{(2)})^2 + \dots + (x_i^{(d)} - x_j^{(d)})^2} \quad (3)$$

其中 $x_i^{(k)}$ 表示样本 x_i 在第 k 维上的具体数据。

(3)核心点与边界点。当给定 $Eps, min_samples$ 参数后,如果样本 x_i 的 Eps 邻域内包含的样本数量 $|N_{Eps}(x_i)| \geq min_samples$,则称该样本 x_i 为核心点,即核心点邻域内的样本数量最少为 $min_samples$ 个,如果有的样本在某个核心点邻域内但不是核心点,那么该样本就叫做边界样本。

密度聚类具有发现任意簇、自动排除噪声点等优点,但是当数据密度差异较大时,聚类效果往往会差强人意,还可能出现聚类错误的情况。DBSCAN 聚类时还要人为设定 Eps 和 $min_samples$ 这两个重要参数,并且常常会因参数设置不合理而导致核心点数据骤减或剧增,进而导致聚类质量严重下滑。本文中引用改进的遗传算法对 Eps 和 $min_samples$ 两个参数进行寻优操作,已获取到真正的最佳核心点数据。

2.3 RBF 神经网络的基本原理

RBF 神经网络是 Moddy 等^[15]于 20 世纪 80 年代提出的一种具有单隐层的 3 层前馈型神经网络,分别为输入层、隐含层和输出层。其中输入层由信号源节点组成,隐含层的节点数主要视需要解决的实际问题而定,输出层即为该网络的输出预测结果。RBF 神经网络是一种将输入矢量扩展到高维空间的神经网络学习方法,具有准确的局部感知特性和强大的非线性逼近能力。RBF 神经网络拓扑图如图 2 所示。

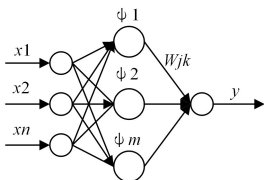


图 2 RBF 神经网络拓扑结构

Fig. 2 RBF neural network topology

RBF 神经网络的映射关系主要由两大部分组成(假设神经网络结构中输入层到输出层的节点数分别为 n, m, k):第一部分是从输入空间到隐含空间的非线性变换层,输出函数如式(4)所示;第二部分为从隐含空间到输出空间的线性合并层,输出函数如式(5)所示。

$$h_j(x) = \varphi_j(\|x - c_j\|) = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right), j = 1, 2, \dots, m \quad (4)$$

其中, $x = [x_1, x_2, \dots, x_n]$ 是 n 维输入矢量; $h_j(x)$ 是第 j 个隐含层节点的输出; $\varphi_j(*)$ 是隐含层神经元的变换函数,即径向基函数; c_j 是第 j 个隐含层节点的中心向量,与输入矢量 x 具有相同维数; σ_j 是第 j 个隐含层节点的宽度值,用来控制该基函数的径向作用范围,并决定了高斯函数的形状,可用来调节该网络的灵敏度。

$$y_k = \sum_{j=1}^m W_{jk} \cdot h_j(x), k = 1, 2, \dots, l \quad (5)$$

其中, y_k 是第 k 个输出单元的输出值(本文中 $k=1$); W_{jk} 是第 j 个隐含层节点到第 k 个输出层节点的连接权值。RBF 神经网络通过该连接权值有效地保证了其良好的预测及推广性能。

RBF 神经网络具有学习速度快、非线性映射能力强、逼近精度高、不存在局部极小值以及泛化能力强等优点,被广泛用于模式识别和时间序列分析等领域^[16]。因此,通过对煤层中相关地震属性数据的学习,RBF 神经网络可以找到地震属性与构造煤厚度之间复杂的非线性映射关系,进而可以对矿区中构造煤的厚度进行精准预测。

3 模型优化

3.1 遗传算法的改进

传统遗传算法对求解问题所有可能的解进行全局搜索,具有较强的全局搜索能力,但其局部搜索能力较弱,会导致种群在进化后期收敛速度慢且不能逼近全局最优解。因此,本文针对传统遗传算法的缺点进行相关改进。在进化初期,个体基因通常适应能力较差,需要选择较大的交叉概率来提高算法的全局搜索能力;在进化后期,随着个体基因适应能力的增强,需要较小的交叉概率来降低全局搜索能力。交叉概率变化的计算公式如式(7)所示:

$$E = \cos\left(\frac{t}{t_{\max}} \times \frac{\pi}{2}\right) \quad (6)$$

$$P_c = \begin{cases} \left(P_{c1} - \frac{(P_{c1} - P_{c2})(f' - f_{\text{avg}})}{f_{\max} - f_{\text{avg}}}\right) * E, & f' > f_{\text{avg}} \\ P_{c1} * E, & f' \leq f_{\text{avg}} \end{cases} \quad (7)$$

与交叉算子相反的是,进化初期的变异概率应较小,以确保优秀的个体基因能够保存下来;进化后期则需要增大变异概率以便增强算法的局部搜索能力。变异概率变化的计算公式如式(8)所示:

$$P_m = \begin{cases} \left(P_{m1} - \frac{(P_{m1} - P_{m2})(f_{\max} - f)}{f_{\max} - f_{\text{avg}}}\right) * E, & f' > f_{\text{avg}} \\ P_{m1} * E, & f' \leq f_{\text{avg}} \end{cases} \quad (8)$$

其中, t 表示进化次数, t_{\max} 表示设定的最大遗传次数, f 是变

异个体具有的适应度值, f_{\max} 是种群中适应度值最大的个体, f_{avg} 代表平均适应度值, f' 是两个交叉个体中适应度值较大的个体。通过以上交叉与变异概率的平滑改变,使得遗传算法在保证优良个体被保存下来的前提下大大提高进化效率。

为了测试改进遗传算法的性能,将改进的遗传算法应用于 Rosenbrock 函数的寻优测试。同时为了比较改进遗传算法的性能,将其与未改进遗传算法和粒子群算法(Particle Swarm Optimization, PSO)进行了实验对比,其中 Rosenbrock 函数表达式如式(9)所示:

$$f(x) = \sum_{i=1}^n [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (9)$$

该测试中 Rosenbrock 函数的 n 取值为 5, $x \in [-100, 100]$ 。PSO 中参数 c_1 和 c_2 设置为 2, ω 为 1, GA, PSO, AGA 3 种优化算法分别运行 20 次所得结果如表 1 所列。

表 1 3 种算法对 Rosenbrock 函数寻优的结果

Table 1 Optimization results of Rosenbrock function by 3 algorithms

项目	GA	PSO	AGA
最大值	839.61	758.96	2.3654
最小值	9.1231	4.3823	0
平均值	171.44	138.98	0.2041
方差	193.61	212.13	0.4913

当 Rosenbrock 函数的 n 取值为 5 时,该函数有 5 个自变量,函数形式较为复杂。由表 1 所列的 3 种优化算法对 Rosenbrock 函数寻优,结果之间的差距较大。GA 与 PSO 算法在 20 次寻优结果中的方差分别为 193.61, 212.13, 方差数值都较大,表示传统的 GA 与 PSO 方法对函数寻优结果较差。文中改进的 AGA 算法具有良好的寻优性能,寻优结果的最大值为 2.3654,最小值为 0,均值为 0.2041,方差为 0.4913,寻优结果较为稳定,说明该算法避免了陷入局部最优的状况,且能够较好地找出全局最优解。

3.2 RBF 神经网络参数优化与算法描述

在 RBF 神经网络的训练学习中,主要是对基函数的数据中心向量 c_j , 方差 σ_j 和权值 W_{jk} 这 3 个参数进行确定,常见的学习算法有 4 种,分别为:随机选取中心法、自组织选取中心法、正交最小二乘法以及监督学习选取中心法,其中最常用的是自组织选取中心法。对于这些学习算法来说,关键的问题是基函数中心向量 c_j 的合理确定,因为该参数的取值是否合理直接影响到预测结果的好坏^[17-19]。

本文改进了 RBF 神经网络基函数中心向量的求解方法,即优化基函数中心向量的自组织选取中心法 K -means 聚类算法,以获取 RBF 神经网络的重要参数值。因为传统的 k -means 聚类在网络初始化阶段随机选取 k 个训练样本作为初始聚类中心,导致其聚类结果容易陷入局部最优并且极其不稳定,所以需要对其改进以便获取更加鲁棒的聚类中心向量。改进的 RBF 神经网络构造煤厚度预测模型算法描述如下。

(1)地震属性预处理。采用主成分分析(Principal Component Analysis, PCA)算法进行数据降维、消除相关性处理,保留结果集以备。

(2)充分利用全部输入输出样本的信息,建立样本之间的相似矩阵,采用模糊聚类最大矩阵元法^[20-21]圈定出 k 值以获取隐含层节点数。

(3)通过式(3)计算每个点与其他所有点之间的欧几里德距离。

(4)使用自适应遗传算法(Adaptive Genetic Algorithm, AGA)确定 DBSCAN 关键参数($Eps, min_samples$),使得 DBSCAN 算法获得更加准确有效的核心点数据,并且建立核心点与核心点距离小于半径 Eps 的点的映射。其中自适应遗传算法的适应度函数使用评价聚类效果的轮廓系数,该轮廓系数综合考虑了簇的密集性与分散性信息,能够很好地对聚类结果进行评判。

(5)通过高密度过滤得到所有最佳核心点数据,即计算每个核心点邻域内的核心点个数,如果个数大于 $min_samples$ 则为最佳核心点。

(6)通过层次聚类的构建凝聚树过程将最佳核心点数据凝聚成 k 个簇,最终以每个簇中所有数据的平均值(即质心)作为 k -means 聚类的初始聚类中心,通过 k -means 聚类结果得到的每个簇的质心可以作为 RBF 神经网络基函数的最优中心向量 c_j 。

(7)根据(6)中 k -means 聚类结果获得的质心之间的距离确定 RBF 神经网络隐含层节点的扩展常数 σ_j 。扩展常数设定为 $\sigma_j = \mu d_j$, μ 为重叠系数(本文取值为 1)。

$$d_j = \min_j \|c_j - c_i(k)\| \quad (10)$$

其中, d_j 表示第 j 个数据中心与其他最近的数据中心之间的距离。

(8)隐含层与输出层之间的连接权值 W_{jk} 通过引用自适应动量估计优化算法进行监督训练获取。

通过使用 AGA-DBSCAN 优化 k -means 算法获得最优的 RBF 神经网络隐含层参数,进而获得高效且鲁棒的 RBF 神经网络模型,理论上它可以作为预测构造煤厚度的预测模型。改进型 RBF 神经网络的构造煤厚度预测算法整体流程如图 3 所示。

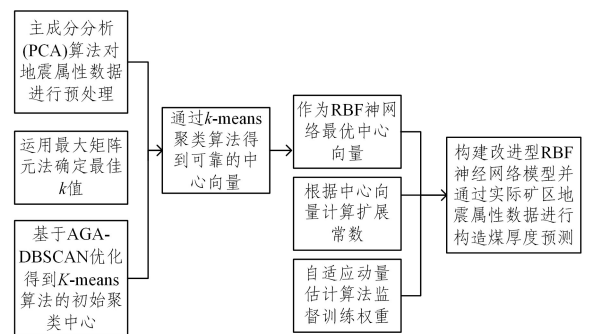


图 3 基于 RBF 的构造煤厚度预测算法流程

Fig. 3 Flowchart of tectonic coal thickness prediction algorithm based on RBF

3.3 预测模型建立

3.3.1 测试数据

为了建立适合构造煤厚度预测的改进型 RBF 神经网络模型,确定实验所需要的数据尤为重要。由于构造煤的分布会影响煤层的地震属性,使其明显区别于正常煤层,因此一些学者开始运用地震属性预测构造煤的厚度与构造煤的分布,并且取得了不错的效果。因此,本文同样将地震属性进行处

理作为实验数据集。首先建立构造煤的模型结构图,如图4所示,煤层的顶底板参数分别为:最薄处为0m,最厚处为10m,原生煤厚度也是如此;直接顶或底是泥岩,厚度为2m,老顶或底是砂岩。根据褶积模型,我们利用50Hz的Ricker子波获得正演地震剖面,如图5所示。利用构造煤模型正演剖面提取出相关的地震属性。

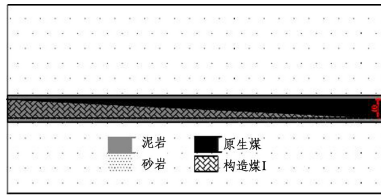


图4 构造煤模型

Fig. 4 Model of tectonic coal

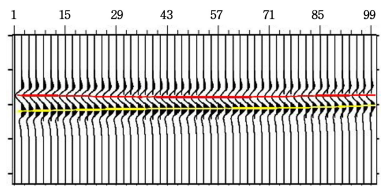


图5 正演地震剖面

Fig. 5 Forward seismic profile

通过上述所建立的煤层楔形模型以及获取的正演地震剖面,能够进一步提取出地震属性数据来进行实验测试。由于地震属性数据具有维度高、噪声多等特点,因此实验过程中使用PCA来进行降维去噪处理,PCA算法中取累积贡献率高于95%的主成分属性作为预测模型的输入属性,为后续实验奠定了基础。

在8#煤层中,我们收集的地震属性为:PoststackAmplitude, ApparentPolarity, Bandwidth, CosineofInstPhase, Dip, Dominantfrequency, InstAcceleration, InstAmp, InstFrequency, InstPhase, InstQ, LambertianReflectance, MostNegativeCurvature, RelativeAI, ResponseFrequency, ResponsePhase, Strike, Sweetness, ThinBedIndicator, WeightedFrequency, 共计20维。该高维地震属性数据经过PCA算法处理后,将原始的20维数据降为现在的10维数据。

3.3.2 参数设置

在改进的遗传算法中,基因编码采用二进制进行编码,共21位,前17位是邻域 Eps ,范围为(0.001,15),后4位是 $min_samples$,范围为(1,31),初始种群数 $S=30$,初始交叉概率 $P_{c1}=0.9$,初始变异概率 $P_{m1}=0.02$,自适应交叉概率 $P_{c2}=0.3$,自适应变异概率 $P_{m2}=0.05$,最大迭代次数 $t_{max}=100$ 。使用模糊聚类中最大矩阵元法获取RBF神经网络隐含层神经元个数 $k=10$,该值也作为 k -means聚类算法的聚类个数。此外,为了探究改进型RBF神经网络模型的预测性能,将改进后的和未改进的RBF神经网络模型与传统的BP神经网络模型以及用 K -折交叉验证方法优化的支持向量机(Support Vector Machines, SVM)模型分别进行预测实验结果的对比。其中,BP神经网络的结构设置为:输入层神经元为10,隐藏层神经元为10,输出层神经元为1,传递函数选用tansing与

logsig函数。学习率设为0.01,最大训练次数设为600,性能度量函数采用均方差函数(Mean Squared Error, MSE)与判定系数(Coefficient of Correlation, R^2)。使用sklearn库中SVM的回归预测模型(Support Vector Regression, SVR),并采用10-折交叉验证进行优化。

3.3.3 测试结果

在实验过程中通过选取90个降维后的地震属性数据进行模型的训练与测试,其中70%作为训练集,对模型参数进行训练;剩余30%作为测试集,测试改进的RBF神经网络构造煤厚度预测模型。遗传算法进化结果如图6所示;训练集损失值随迭代次数变化的曲线如图7所示;测试集经模型预测的拟合曲线如图8所示。

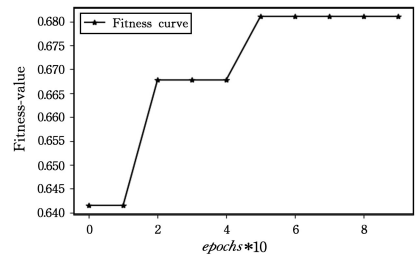


图6 适应度值变化曲线

Fig. 6 Fitness value curve

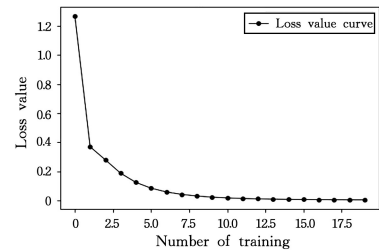


图7 训练集损失值变化曲线

Fig. 7 Loss value curve of training set

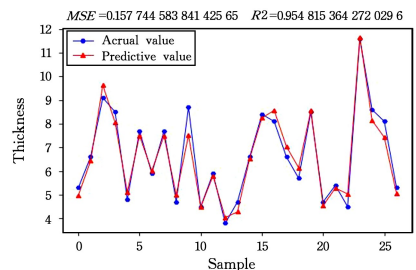


图8 构造煤厚度预测拟合曲线

Fig. 8 Fitting curve of tectonic coal thickness prediction

由图6可知,遗传算法在进化到50次时,适应度值达到最大,为0.68,此时获得的参数值分别为: $Eps=0.42$, $min_samples=20$ 。由图7可知,训练在100次左右开始收敛,最终损失值为0.0003,训练精度较高。由图8可知,预测结果有很好的拟合效果,拟合精度达0.954,均方误差为0.15。首先将改进的RBF神经网络构造煤厚度预测模型与未改进的RBF神经网络模型进行对比验证。取90个数据样本进行实验,构建模型,每个模型分别运行10次,并且每次均随机抽取70%作为训练集,30%作为测试集,预测结果如图9所示。

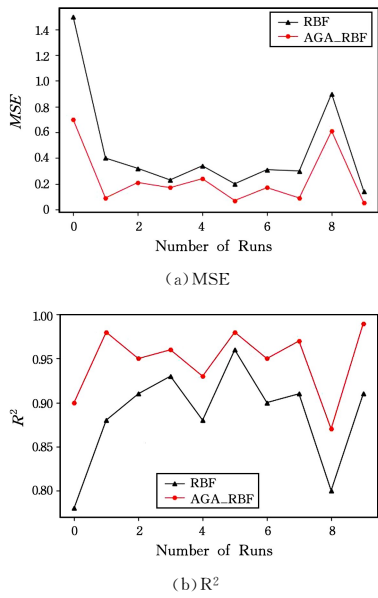


图9 改进算法与原算法预测结果比较

Fig. 9 Predictive results comparison between improved algorithm and original algorithm

由图9可知,改进的RBF神经网络构造煤厚度预测模型在准确率方面要优于未改进的RBF神经网络模型,且在10次实验中的预测结果都比原RBF神经网络模型更好。因此,可以得出以下结论:本文提出的改进方案在原理和效果上具有明显的提升,可以为进一步的验证分析做出指导。

为了进一步验证改进模型所具有的优秀预测性能,将改进的RBF神经网络构造煤厚度预测模型与经过K-折交叉验证优化的SVM模型以及传统的BP神经网络模型进行实验,用预测结果MSE与 R^2 两个指标进行比较与分析。地震属性数据的选取及分配方式同前面保持一致。其中SVM采用10-折交叉验证方法进行参数优选,预测结果如图10所示。

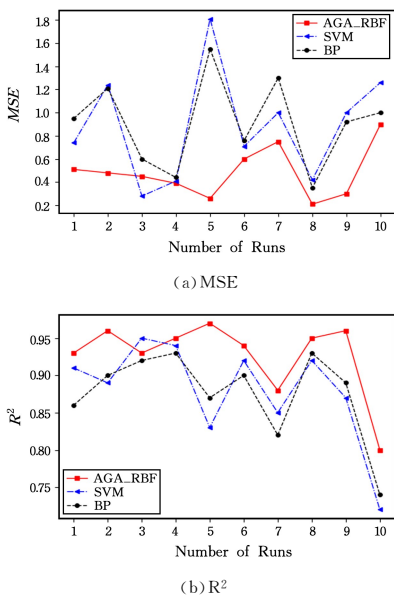


图10 改进算法与其他算法预测结果比较

Fig. 10 Predictive results comparison between improved algorithm and other algorithms

由图10可知,采用AGA-DBSCAN优化的RBF神经网络构造煤厚度预测模型的预测准确率要高于使用K-折交叉验证优化的SVM模型与传统的BP神经网络模型的预测准确率,且在10次实验中,改进的AGA-RBF模型的平均预测准确率达到92%以上。因此,本文提出的基于AGA-DBSCAN优化的RBF神经网络构造煤厚度预测模型在定量预测构造煤的厚度上是可行的,并且预测效果较好。

4 实例分析

4.1 研究区介绍

本文选取淮北矿业集团芦岭煤矿II六采区8#煤层作为研究区域,通过结合实际的地震属性,对构造煤厚度进行预测。该矿区面积达 24km^2 ,煤层厚度较大,构造煤较发育,瓦斯含量较高,极易发生煤与瓦斯突出。因此,淮北矿业集团的芦岭煤矿成为研究煤与瓦斯突出课题的理想区域。芦岭煤矿中,8#煤层埋深为 $448\sim 930.73\text{m}$,平均埋深 642.13m ,煤层厚度为 $3.8\sim 16.2\text{m}$,平均厚度为 10.24m 。构造煤厚度为 $3.8\sim 11.6\text{m}$,平均厚度为 6.7m ,钻孔共有20口,详细情况如表2所列。

表2 8#煤层钻孔揭露煤层信息

Table 2 Drills in 8# coal seam exposing coal information (单位:m)

钻孔编号	构造煤厚度	底板标高	煤层厚度
L44	8.1	926.8	10.6
L50	8.4	812.99	11.8
91-5	6.6	671.15	13
2002-4	11.6	723.85	14.1
2002-5	9.1	794.95	12.7
2010-11	8.2	877.25	11.1
2012-1	5.3	643.51	7.4
2014-5	8.6	930.71	11
L43	8.7	807.92	10.7
06-4	4.3	457.81	7.4
91-2	8.5	565.95	11.5
92-8	4.5	592.75	6.6
94-2	3.8	600.35	3.8
91-1	5.9	458.57	8.4
92-2	4.8	589.6	7.6
94-5	5.4	582.85	8.7
2002-3	5.7	760.35	11.5
94-1	4.7	505.6	9.8
94-3	7.7	448.58	16.2
99-1	4.4	673.61	11

4.2 构造煤厚度预测

利用上述预测模型,对芦岭煤矿II六采区8#煤层的构造煤厚度进行预测。由于已知的20口钻孔数据较少,难以保证预测模型的可靠性,因此将钻孔附近范围为 $15\text{m}\times 15\text{m}$ 的地震属性数据提取出来作为训练集(共有2142个数据),以此增加训练集数量。8#煤层所收集的地震属性维数为20,首先使用主成分分析算法对该数据进行降维,处理后数据维数为10,然后采用 2142×10 的数据训练构造煤厚度预测模型。将SVM模型、BP神经网络模型以及改进的AGA-RBF神经网络模型分别用于整个芦岭8#煤层的构造煤厚度预测中,并且将这3种方法预测出的20口钻孔的构造煤厚度提取出来与实际真实数据进行对比。20口钻孔构造煤的预测值与

预测误差如表 3 所列。

表 3 3 种方法钻孔处预测值对比

Table 3 Predicted value comparison of three methods at drills

(单位: %)

钻孔名	构造煤厚度	SVM		BP		AGA_RBF	
		预测值	绝对误差	预测值	绝对误差	预测值	绝对误差
L44	8.1	8.70	0.60	8.40	0.30	8.32	0.22
L50	8.4	7.90	0.50	8.11	0.29	8.53	0.13
91-5	6.6	6.45	0.15	6.50	0.10	6.50	0.10
2002-4	11.6	12.6	1.00	11.5	0.10	11.0	0.60
2002-5	9.1	9.40	0.30	8.88	0.22	9.00	0.10
2010-11	8.2	8.31	0.11	8.10	0.10	7.90	0.30
2012-1	5.3	5.50	0.20	5.10	0.20	5.23	0.07
2014-5	8.6	8.51	0.09	8.90	0.30	8.60	0.00
L43	8.7	8.91	0.21	8.20	0.50	8.69	0.01
06-4	4.3	4.20	0.10	4.60	0.30	4.70	0.40
91-2	8.5	8.60	0.10	8.80	0.30	8.50	0.00
92-8	4.5	4.80	0.30	4.40	0.10	4.63	0.13
94-2	3.8	3.77	0.03	4.00	0.20	3.72	0.08
91-1	5.9	7.50	1.60	5.60	0.30	5.70	0.20
92-2	4.8	4.20	0.60	4.70	0.10	4.80	0.00
94-5	5.4	5.55	0.15	5.70	0.30	5.20	0.20
2002-3	5.7	5.60	0.10	5.58	0.12	5.90	0.20
94-1	4.7	4.66	0.04	4.90	0.20	4.85	0.15
94-3	7.7	7.77	0.07	7.55	0.15	7.60	0.10
99-1	4.4	4.80	0.40	4.65	0.25	4.30	0.10
平均值	6.71	6.88	0.33	6.70	0.22	6.68	0.15

由表 3 可以看出,使用改进的 AGA_RBF 神经网络模型对 20 口钻孔的构造煤厚度预测的绝对误差较小,平均误差为 0.15,其中最大误差为 0.6,最小误差为 0,预测结果比较稳定;使用 10-折交叉验证优化的 SVM 模型预测精度略差一些,平均误差为 0.33,最大误差为 1.6,最小误差为 0.03;而传统 BP 神经网络在对构造煤厚度的预测结果中平均误差为 0.22,最大误差为 0.5,最小误差为 0.10,虽然其预测效果要好于 SVM 模型,但差于改进的 AGA_RBF 神经网络模型。因此,使用本文提出的改进型 RBF 神经网络模型对构造煤厚度的预测效果更好,具有明显优势。

结束语 本文利用实际矿区煤层中的地震属性数据来进行构造煤厚度的定量预测。首先对所获取的高维地震属性数据进行降维、去噪处理,然后针对传统方法的不足构建了改进的 RBF 神经网络构造煤厚度预测模型,使模型结构与参数达到最优,并进行了淮北矿业集团的芦岭煤矿 II 六采区 8# 煤层构造煤厚度预测实验。实验结果表明,其预测精度和性能明显优于未改进的 RBF 神经网络模型和传统的 BP 神经网络模型以及运用 K-折交叉验证优化的 SVM 模型,且预测误差较小,结果合理可靠,因此可以用于实际矿区构造煤厚度的预测,从而为解决煤矿瓦斯突出问题提供了一种新方法。此外,模型的预测结果还受到地震属性信噪比的影响。如果噪声数据在训练样本中含量较多,将会导致预测效果严重不佳,因此还需要在降低地震属性数据信噪比的方法上进行更进一步的科学研究。

参考文献

[1] SHU L Y, WANG K, QI Q X, et al. Key structural body theory

of coal and gas outburst[J]. Chinese Journal of Rock Mechanics and Engineering, 2017, 36(2): 347-356.

- [2] YU S, BO J, HEWEN L, et al. Variations in stress-sensitive minerals and elements in the tectonic-deformation Early to Middle Permian coals from the Zhuxianzhuang mine, Anhui Province [J]. Journal of Geochemical Exploration, 2018, 188: 11-23.
- [3] LI X J, SHEN Z H, LIU Y. Experimental study on the influence of pore structure of structural coal and primary structure coal on adsorption characteristics in northwestern Guizhou [J]. Journal of Mining and Safety Engineering, 2017, 34(1): 170-176.
- [4] ZHAO Y P, SHI L Q, GAO W F, et al. Statistical analysis of coal mine accidents in China's coal mine transformation and development period [J]. Coal Technology, 2016(9): 321-324.
- [5] CHEN T J, WANG X, GUAN Y W. Quantitative prediction of structural coal thickness based on SVR and seismic attributes [J]. Chinese Journal of Coal, 2015, 40(5): 1103-1108.
- [6] TANG Z. The regional tectonic evolution and its control mechanism for structural development in Yangquan coalfield [D]. Xuzhou: China University of Mining and Technology, 2018: 46-79.
- [7] YANG Y L, CHEN Y Q, GAO J X, et al. Experimental study on the effect of temperature on the gas adsorption capacity of media [J]. Mining Safety and Environmental Protection, 2016, 43(4): 6-9.
- [8] ZHANG H J, ZHANG L, WANG D, et al. Gas emission characteristics of tectonic coal and microscopic explanation of pore structure [J]. Journal of China Coal Society, 2018, 43(12): 3404-3410.
- [9] ZHANG Y G, WANG Y, YANG R J. Weather forecasting model based on radial basis neural network [J]. Guiyang: Journal of Guizhou University (Natural Science Edition), 2018, 35(1): 69-72.
- [10] ZHANG P F, XU X, GE X Q. A hardware resource allocation method for multi-antenna ground station based on improved genetic algorithm [J]. Computer Engineering & Science, 2017, 39(6): 1155-1163.
- [11] ZHANG X L, CUI N N, YANG T, et al. A layered adaptive fast K-means algorithm [J]. Application Research of Computers, 2016, 33(2): 421-443, 427.
- [12] SITTEL F, STOCK G. Robust density-based clustering to identify metastable conformational states of proteins [J]. Journal of Chemical Theory & Computation, 2016, 12(5): 1-4.
- [13] SONI N, GANATRA A. AGED (automatic generation of Eps for DBSCAN) [J]. International Journal of Computer Science and Information Security, 2016, 14(5): 536-559.
- [14] WANG X L, HUANG L, YANG P, et al. Dynamic RBF neural networks for model mismatch problem and its application in flotation process [J]. CIESC Journal, 2016, 3(67): 897-902.
- [15] JGOUTA M, NSIRI B. GNSS Positioning performance analysis using PSO-RBF estimation model [J]. Transport and Telecommunication Journal, 2017, 18(2): 146-154.
- [16] LIAN J M, LEE Y G, SCOTT D, et al. Self-organizing radial basis function network for real-time approximation of continuous-

time dynamical system[J]. IEEE Transactions on Neural Networks, 2008, 19(3):460-474.

- [17] SHI K, QIN H, SIMA C, et al. Dynamic barycenter averaging kernel in RBF networks for time series classification[J]. IEEE Access, 2019, 7:47564-47576.
- [18] WEI W, WANG G B. Detection and recognition of air targets by unmanned aerial vehicle based on RBF neural network[J]. Ship Electronic Engineering, 2018, 38(10):37-40.
- [19] CHEN D. Research on traffic flow prediction in the big data environment based on the improved RBF neural network[J]. IEEE Transactions on Industrial Informatics, 2017, 13(4):2000-2008.
- [20] WU C M, FAN J L. Maximal matrix element method for determining the number of hidden nodes of RBF neural networks[J]. Computer Engineering and Applications, 2004, 20:77-79.
- [21] CHEN T, WANG X. Thickness prediction of tectonically deformed coal using calibrated seismic attributes: A case study

[C]//ASEG Extended Abstracts 2016; 25th International Geophysical Conference and Exhibition. Australian Society of Exploration Geophysicists, 2016:817-821.



WU Shan-jie, born in 1993, postgraduate. His main research interests include intelligent information processing and machine learning.



WANG Xin, born in 1978, Ph.D, associate professor. Her research interests include intelligent information processing and machine learning.