

基于遗传优化径向基概率神经网络的岩性识别应用

靳玉萍*, 李保霖

(西安科技大学 计算机科学与技术学院, 西安 710054)

(*通信作者电子邮箱 jyp9999@163.com)

摘要:岩性识别是测井数据解释中最关键的一环,但传统的岩性识别方法解释效率慢,精度低,受人为因素影响大。为此,提出一种遗传优化径向基概率神经网络(RBPNN)的岩性识别方法。该方法融合概率神经网络(PNN)和径向基函数神经网络(RBFNN)的优势来构造RBPNN,采用遗传算法搜索使得RBPNN训练法误差最小的最优隐中心矢量和相匹配的核函数控制参数,优化网络结构,提高收敛速度与精度,形成全结构遗传优化的RBPNN模型。实例应用表明,基于遗传优化RBPNN的岩性识别能够达到工程实际应用的规范标准,且是可行有效的,能够为油田地质勘探领域的岩性识别提供科学的理论支持与依靠。

关键词:岩性识别;径向基概率神经网络;遗传算法

中图分类号: TP183 **文献标志码:** A

Lithology identification based on genetic optimized radial basis probabilistic neural network

JIN Yuping*, LI Baolin

(College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an Shaanxi 710054, China)

Abstract: Lithology identification is the most critical procedure in the logging data interpretation field, while the traditional lithology identification methods have a lot of defects such as slow explain efficiency, low accuracy, and big influenced human factors. To resolve these problems, a new kind lithology identification method was put forward using genetic optimized Radial Basis Probability Neural Network (RBPNN). Probabilistic Neural Network (PNN) and the Radial Basis Function Neural Network (RBFNN) were combined to construct RBPNN. To optimize network structure, upgrade convergence speed and accuracy, Genetic Algorithm (GA) was used to search for the optimal hidden center vector and matching kernel function control parameters of the RBPNN structure which must satisfy minimum error of RBPNN training and form genetic optimized RBPNN network model. The case study shows that lithology identification based on genetic optimized RBPNN can achieve the actual application standards, and it is feasible and effective, it also can provide scientific theoretical supports and dependences for oil geological exploration field.

Key words: lithology identification; Radial Basis Probabilistic Neural Network (RBPNN); Genetic Algorithm (GA)

0 引言

地质条件、沉积环境的复杂性以及实际地层分布的非均质,使得采用交会图快速识别法、概率统计和模糊聚类等传统岩性识别方法^[1-2]不能满足实际生产需要。人工神经网络是由大量非线性神经元相互连接构成的模拟生物神经网络结构和功能的一种非线性人工智能动力学系统。目前,在岩性识别中得到广泛应用的误差反向传播神经网络(Back Propagation Neural Network, BPNN)虽然避免了传统岩性识别方法的许多弊端,但也存在着学习率与稳定性之间的矛盾,收敛速度慢,容易陷入局部最小点而无法得到最优解等问题^[3]。因此,本文构造径向基概率神经网络(Radial Basis Probabilistic Neural Network, RBPNN)模型。鉴于RBPNN隐层节点的不确定以及核控制参数选取困难问题,采用遗传算法优化RBPNN结构,提高了网络收敛速度和精度,增强了网络推广泛化能力。相对传统岩性识别方法,该方法在保证收敛速度和识别精度的同时,能够更加快速与准确地实现岩性信息识别,达到了工程实际应用的规范标准,对整个测井数据

的综合解释领域具有实用性。

1 全结构遗传优化RBPNN模型

1.1 RBPNN

RBPNN由概率神经网络(Probabilistic Neural Network, PNN)和径向基函数神经网络(Radical Basis Function Neural Network, RBFNN)融合而成,继承了概率神经网络收敛性能好、速度快的特点^[4],以及径向基函数神经网络结构简单、多类别模式间交错影响小、非线性映射能力强的优势。

1.1.1 RBPNN结构

本文构造的RBPNN为4层前馈结构,如图1所示。

输入层 输入层即为采集的样本数据,用 x_1, x_2, \dots, x_n 表示,其中 $x_i (i=1, 2, \dots, n)$ 为第 i 个样本节点。输入层节点的作用是传递信号到隐层。

模式层 主要由样本空间每个模式类别中心矢量组成(本文称隐中心矢量),结构上相当于RBFNN的隐层。节点的传输函数是既满足PNN又满足RBFNN的Parzen窗函

收稿日期:2012-07-27;修回日期:2012-09-03。 基金项目:国家自然科学基金资助项目(51134019)。

作者简介:靳玉萍(1975-),女,河南开封人,讲师,博士,主要研究方向:人工神经网络; 李保霖(1989-),女,陕西商洛人,硕士研究生,主要研究方向:人工神经网络。

数——径向基函数^[5],如式(1)所示:

$$f(\mathbf{x}) = \exp\left(-\left(\frac{\|\mathbf{x} - \mathbf{c}_t\|_2}{\sqrt{2}\sigma}\right)^2\right) \quad (1)$$

该函数的隐中心矢量 \mathbf{c}_t 表示某类别中第 t 个隐中心矢量,即代表了输入层到隐层的连接权重 \mathbf{w}_t ,也就是 $\mathbf{w}_t = \mathbf{c}_t$;式(1)中 σ 为核函数控制参数。

第一隐层的第 i 个节点的输出为:

$$f(\mathbf{x}_i) = \exp\left(-\left(\frac{\|\mathbf{x}_i - \mathbf{c}_t\|_2}{\sqrt{2}\sigma}\right)^2\right) \quad (2)$$

其中: \mathbf{x}_i 指第 i 个训练样本点节点, \mathbf{c}_t 为第 l 个类别的第 t 个隐中心矢量。

求和层 相当于 PNN 的隐层,其节点执行求和运算,即对第一隐层节点的输出,按隐中心矢量的类别不同进行有选择地加和。因此,第二隐层中每个节点与第一隐层的连接权重值根据隐中心矢量的类别不同,取值为 1 或 0。第二隐层的实际输出如式(3)所示:

$$h_k(\mathbf{x}_i) = \sum_{t=1}^{n_l} \left(\exp\left(-\left(\frac{\|\mathbf{x}_i - \mathbf{c}_t\|_2}{\sqrt{2}\sigma}\right)^2\right) \right) \quad (3)$$

其中: $h_k(\mathbf{x}_i)$ 为第 i 样本对应的第二隐层第 k 个节点的输出, n_l 为第 l 个类别的隐中心矢量数目。

输出层 RBPNN 的输出层相当于 RBFNN 的输出层,是对第二隐层的加权求和,如图 1 中的 $y_j (j = 1, 2, \dots, m)$ 。该层权值是网络训练的目标,若设 $h(\mathbf{x}_i) = (h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), \dots, h_m(\mathbf{x}_i))$, 则输出层的输出可表示为

$$\mathbf{y}_j = \mathbf{y}(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i) \mathbf{W} \quad (4)$$

其中 \mathbf{W} 为第二隐层到输出层的连接矩阵。

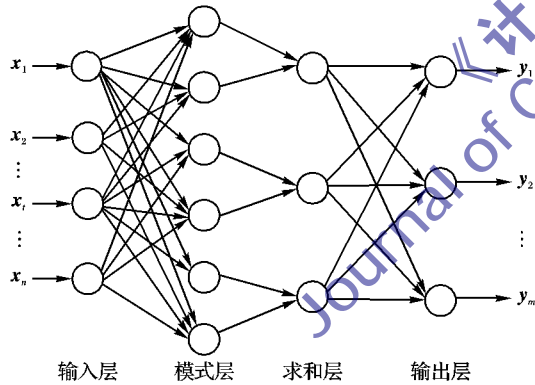


图1 RBPNN 结构

1.1.2 RBPNN 训练

首先,在不明确样本类别属性的情况下,采用无监督法对样本进行分类划分^[6];然后,优选隐中心矢量和相关控制参数,计算求和层 \mathbf{H} 输出;最后,通过递推正交最小二乘法学习训练网络,得到网络权值 \mathbf{W} ,确定网络模型。

RBPNN 的训练过程如下:

对于 L 个训练样本,可以得到计算网络实际输出向量矩阵 \mathbf{Y} ,如式(5)所示:

$$\mathbf{Y} = \mathbf{H} \mathbf{W} \quad (5)$$

其中: \mathbf{H} 为第二隐层的输出矩阵, \mathbf{W} 为第二隐层与输出层的连接权重矩阵。

\mathbf{W} 可根据式(6)进行计算:

$$\mathbf{W} = \mathbf{R}^{-1} \bar{\mathbf{Y}} \quad (6)$$

其中 \mathbf{R} 、 $\bar{\mathbf{Y}}$ 可根据式(7)~(8)进行计算。

$$\mathbf{H} = \mathbf{Q} \begin{bmatrix} \mathbf{R} \\ 0 \end{bmatrix} \quad (7)$$

其中: \mathbf{H} 根据正交分解定理^[7-8] 分解得到正交矩阵 \mathbf{Q} 和上三角阵 \mathbf{R} 。

$$\mathbf{Q}^T \mathbf{Y} = \begin{bmatrix} \bar{\mathbf{Y}} \\ \hat{\mathbf{Y}} \end{bmatrix} \quad (8)$$

其中: \mathbf{Y} 为期望信号矩阵, $\bar{\mathbf{Y}}$ 与 $\hat{\mathbf{Y}}$ 分别为正交化后的期望信号矩阵和误差驻留矩阵。

1.2 遗传优化 RBPNN 结构

隐中心矢量的优化选择和核函数控制参数的设计是 RBPNN 的关键。若将所有样本都作为隐中心矢量,可以得到满意的网络收敛精度,但会使网络结构过于庞大,训练时间和测试时间延长。核函数控制参数的作用机理与人类视觉神经系统相似,宽度选择过大,只能看到目标的概貌;反之,得到的只能是细微特征。核函数控制参数和隐中心矢量在结构上是相互依赖的^[9-10],遗传优化 RBPNN 的实质是利用遗传算法的全局优化性,搜索 RBPNN 结构空间中满足最低训练误差所形成的结构曲面上的最优隐中心矢量和相匹配的核函数控制参数的过程^[11]。

1.2.1 遗传算法的构造

1) 编码方式的确定。由隐中心矢量与相应控制参数之间的依赖关系,采用二进制与实数混合编码方式构造染色体。其中,前 $L-1$ 位表示初始隐中心矢量,1 表示该样本被选择,0 表示不被选择,最后一位实数为核函数控制参数。

2) 适应度函数的构造。根据染色体的编码方式,适应度函数定义如下:

$$f(e, \mathbf{x}_i) = \frac{C(s+1)}{2^s N_c} \quad (9)$$

其中: $s = \text{sign}(\varepsilon - e)$; \mathbf{x}_i 为当前个体; C 为常数; ε 是给定的误差标准; N_c 表示当前染色体中隐中心矢量数目; l 是染色体的长度; e 为个体对应的实际输出误差,可以通过递推正交最小二乘法得到,如式(10)。

$$e = \|\mathbf{Y}_{\text{exp}} - \mathbf{H} \mathbf{W}\|_2 \quad (10)$$

其中 \mathbf{Y}_{exp} 为期望输出信号。

3) 选择(繁殖)策略。采用赌轮选择及最佳个体保持的选择策略进行繁殖^[12]。

由 $P_{si} = f_i \cdot \left(\sum f_i \right)^{-1}$ 计算当前种群中每个个体的选择概率;产生 $(0, 1)$ 内的一个随机数 r , 若 $P_{s1} + P_{s2} + \dots + P_{s(i-1)} \leq r < P_{s1} + P_{s2} + \dots + P_{si}$, 则个体 X_i 被选择到配对库;否则,放弃。当前种群中的最优个体直接作为下一代种群成员。

4) 遗传算子的设计。交叉算子采用实数基因位作为固定交叉点的多点交叉方式。变异算子对于二进制编码和实数编码分别采用基于变异概率的每个基因位逻辑取反和高斯变异方式来确定。

1.2.2 全结构遗传优化 RBPNN 的实现

根据 1.2.1 节关于遗传算法进行 RBPNN 结构优化的介绍,本文遗传算法优化的 RBPNN 的流程如图 2 所示。

2 工程实例

本文选用长庆油田某区域样本作为实验数据,样本大小为 42。其中,用 30 个数据来训练网络,其余的 12 个用来测试网络通用性。输入层由 4 个神经元组成,分别选择自然伽玛(natural Gamma Ray, GR)、自然电位(Spontaneous Potential, SP)、声波(acoustic wave)和密度测井(density logging)四种测井数据。将测井识别值作为研究对象,分别为网络的 3 个输

出层节点,即灰白色粗砂岩、深灰色泥岩和灰色泥质粉砂岩。

对网络期望输出编码^[13]设计如下:

灰白色粗砂岩 (0.99,0.01,0.01);

深灰色泥岩 (0.01,0.99,0.01);

灰色泥质粉砂岩 (0.01,0.01,0.99)。

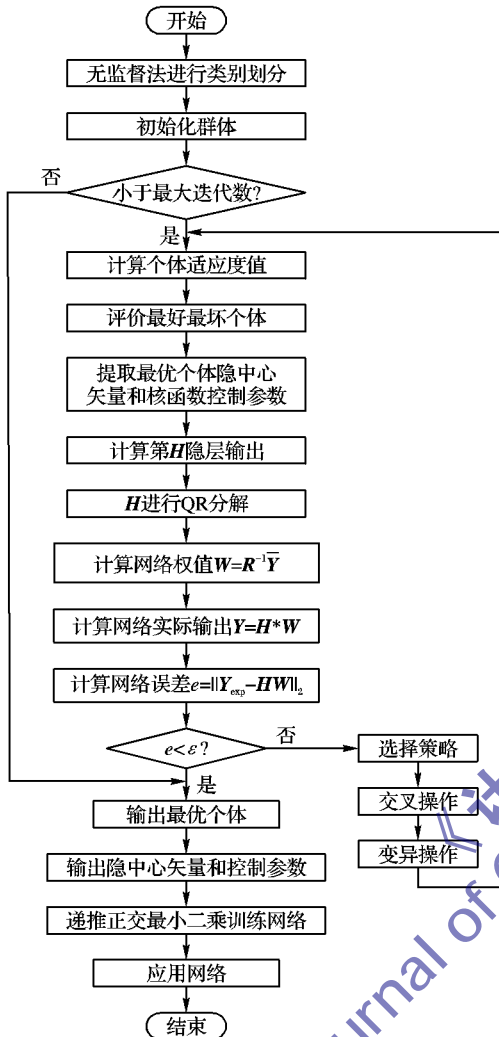


图2 全结构遗传优化RBPNN流程

2.1 RBPNN模型的构造

1) 控制参数设置。首先,对于无类别样本,无监督学习算法确定各样本的类别属性;其次,确定有关控制参数。在此,某些参数是在实验过程中不断寻优取得,如种群规模 $popsize = 30$, $chromlength = 31$,交叉概率 $pc = 0.7$,变异概率 $pm = 0.005$,算法执行的最大迭代数 $maxgeneration = 50$,个体适应度的收敛误差标准 $\varepsilon = 0.85$ 。

2) 遗传算法全结构优化RBPNN,得到最后的优生种群及相对应的个体适应度值如图3所示。

3) 优选隐中心矢量和核函数控制参数。对最后优生种群的分析:隐中心矢量和核函数控制参数都已经趋于稳定数值,其最优解如下:

0.7392 0.0057 0.0906 0.1364

0.9247 0.9267 0.8882 0.5848

0.8610 0.9606 0.8941 0.5734

0.0000 0.6845 0.1648 0.6258

0.7360 0.6636 0.4854 0.3722

相匹配的核函数控制参数 $\sigma = 0.3975$ 。

4) 应用遗传算法优选的隐中心矢量和核函数控制参数,

采用递推正交最小二乘算法训练网络,得到网络权值如下:

0.5139 0.0174 -0.0551

0.0103 0.3800 -0.0913

0.0975 -0.2649 1.0135

第49代优生种群及个体适应度:

0011001000010100010100110001100000	0.3975	适应度值:	0.0123
0011000100001100000010100101111000	0.3975	适应度值:	0.0156
0011000100001000001100101001000000	0.3975	适应度值:	0.0156
0011000100001010001100100000101000	0.3975	适应度值:	0.0156
0011001000001000001100101001000000	0.3975	适应度值:	0.0100
0000101000001000001010010111100000	0.3975	适应度值:	0.0156
0011000100001000001010010111100000	0.3975	适应度值:	0.0278
0011000100001000001100100000101000	0.3975	适应度值:	0.0204
0011000100001000001100101001000000	0.3975	适应度值:	0.0204
0011000100001010001100100000101000	0.3975	适应度值:	0.0100
0011000100001010001100100000101000	0.3975	适应度值:	0.0278
0001001000001000001100100000100000	0.3975	适应度值:	0.0156
0011000100001001101010010011100000	0.3975	适应度值:	0.0204
0011000100001010001010010000100000	0.3975	适应度值:	0.0156
0011000100001010001100101001000000	0.3975	适应度值:	0.0204
0011000100001010001100101001000000	0.3975	适应度值:	0.0156
0011000100001010001100101001000000	0.3975	适应度值:	0.0100
0011000100001010001100101001000000	0.3975	适应度值:	0.0278
0011000100001000001011001010010000	0.3975	适应度值:	0.0204
0011000100001000001011001010010000	0.3975	适应度值:	0.0400
0011000100001010001100100000101000	0.3975	适应度值:	0.0400
0011000100001010001101001011010000	0.3975	适应度值:	0.0123
0011000100001000001001010010000000	0.3975	适应度值:	0.0156
0011000100001000001110000101100000	0.3975	适应度值:	0.0204
0011000100001000001010010111000000	0.3975	适应度值:	0.0156
0011000100001101000101001000001000	0.3975	适应度值:	0.0204
0011000100001000001011001010010000	0.3975	适应度值:	0.0204
0011000100001000001010010100101000	0.3975	适应度值:	0.0278
0011000100001010001100101001000000	0.3975	适应度值:	0.0204
0010001000001010001100101001000000	0.3975	适应度值:	0.0278

图3 最优种群与个体

2.2 数据校验

针对预留的12组测试数据,采用训练好的RBPNN模型识别岩性,并与期望输出进行了对比,如表1所示。

表1 测试样本的实际输出与测试数据期望输出对比

实际输出	测试样本期望		
	灰白色粗砂岩	深灰色泥岩	灰色泥质粉砂岩
0.9909	0.0039	0.0195	0.99
0.9882	0.0007	0.0331	0.99
0.9775	-0.0019	0.0439	0.99
0.9852	0	0.0359	0.99
0.0024	0.9458	0.0857	0.01
0.0072	0.9574	0.0167	0.01
0.0083	0.9630	-0.0036	0.01
0.0077	0.9768	-0.0109	0.01
0.0029	0.0145	0.9879	0.01
0.0317	0.1332	0.9773	0.01
0.0392	0.1193	0.9485	0.01
0.0181	0.1074	0.9803	0.01

根据表1测试数据将识别结果应用到岩性识别软件系统中,得到三维可视化结果如图4所示。

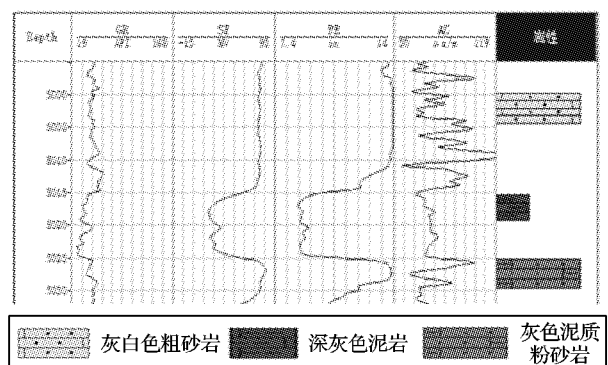


图4 测试数据岩性识别图示

为进一步验证该方法的泛化能力,挑选长庆油田安塞区域19口井中1000个样本作为测试数据。其中,棕红色泥岩549个,棕红色细砂岩360个,杂色砾岩91个。对这组数据采用BPNN和RBPNN进行岩性识别应用对比,结果如表2所示。

从表 2 可以看出,本文的全结构遗传优化 RBPNN 较传统 BPNN 在识别正确率和识别时间上都取得了一定幅度的提

高。将该方法应用到岩性识别中是切实可行且有效的,测试结果能够达到工程实际的规范要求。

表 2 岩性识别应用对比

模型	棕红色泥岩		棕红色细砂岩		杂色砾岩		总体正确率/%	时间/min
	识别数	正确率/%	识别数	正确率/%	识别数	正确率/%		
BPNN	447	81.00	312	86.7	71	78.1	83.0	3.47
RBPNN	513	93.44	316	87.7	80	87.9	90.9	1.54

3 结语

1) 本文的遗传优化 RBPNN 是结合 PNN 和 RBFNN 的优点构造的 4 层前馈 RBPNN, 引入遗传算法在整个种群范围内, 搜索使得递推正交最小二乘学习方法误差最小的最优隐中心矢量和相匹配的核函数控制参数, 增强了网络结构的稳定性, 提高了网络收敛的速度和精度。

2) 由测试数据应用结果对比可知, 基于遗传优化 RBPNN 的岩性识别, 相对于其他识别方法有着显著的速度快、准确率高、结构稳定、推广泛化能力好等优点。经多次实验论证, 该方法能够达到实际工程应用的标准要求, 为油田地质勘探和测井数据解释领域提供了科学的方法支持和理论依靠。

参考文献:

- [1] 张涛, 莫修文. 基于交会图与模糊聚类算法的复杂岩性识别[J]. 吉林大学学报, 2007, 37(1): 109-113.
- [2] 刘秀娟, 陈超, 曾冲, 等. 利用测井数据进行岩性识别的多元统计方法[J]. 地质科技情报, 2007, 26(3): 109-112.
- [3] 潘伟, 陈玉梁. 两种神经网络方法在岩性识别方面的应用[J]. 国外测井技术, 2008, 23(4): 47-50.

- [4] 蔡曲林. 基于概率神经网络的模式识别[D]. 长沙: 国防科学技术大学, 2005: 9-11.
- [5] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2002: 135-160.
- [6] 漆随平, 于慧彬, 刘涛, 等. 基于径向基概率神经网络的气象参数状态识别[J]. 自动化仪表, 2008, 29(8): 5-7.
- [7] 程去鹏, 张凯院, 徐仲. 矩阵论[M]. 西安: 西北工业大学出版社, 2001: 196-215.
- [8] 赵文波, 黄得双. 基于无监督学习算法训练径向基概率神经网络[J]. 模式识别与人工智能, 2003, 16(4): 442-447.
- [9] 赵温波. 径向基概率神经网络研究[D]. 合肥: 中国科学技术大学, 2003: 1-121.
- [10] 李伦波, 马广富. 基于 RBPNN 的退化交通标志图像的识别算法[J]. 吉林大学学报: 工学版, 2008, 38(6): 1429-1433.
- [11] 赵温波, 黄得双, 郭璘. 径向基概率神经网络的遗传优化[J]. 中国科学技术大学学报, 2003, 33(6): 733-741.
- [12] 潘勇, 郭晓东. 一种基于遗传算法改进的粒子群优化算法[J]. 计算机应用与软件, 2011, 28(9): 222-224.
- [13] 曾喆昭, 王耀南. 基于正交神经网络算法的多传感器数据融合方法[J]. 传感技术学报, 2007, 20(6): 1368-1370.

(上接第 332 页)

到了基本 HSA 中, 有效利用了 DEA 的收敛速度快、方向性强的优点, 解决了基本 HSA 寻优速度慢、后期停滞的问题。实验表明, DEHSA 是一种比基本 HSA、ES 和 NDCWPSO 算法求解性能更优的算法。

参考文献:

- [1] 王凌. 车间调度及其遗传算法[M]. 北京: 清华大学出版社, 2003.
- [2] BIEGEL J, DAVERN J. Genetic algorithms and job shop scheduling[J]. Computers and Industrial Engineering, 1990, 19(1/2/3/4): 81-91.
- [3] 陈琦, 马向阳. 求解有限产能作业车间调度问题的改进蚁群算法[J]. 计算机工程与应用, 2012, 48(5): 232-235, 243.
- [4] 宋晓宇, 朱云龙, 尹朝万, 等. 求解模糊 Job Shop 调度问题的改进禁忌搜索算法[J]. 沈阳建筑大学学报: 自然科学版, 2006, 22(5): 841-845.
- [5] 朱霞. 一种求解作业车间调度的文化粒子群算法[J]. 计算机应用研究, 2012, 29(4): 1234-1236.
- [6] SHA D Y, HSU C-Y. A hybrid particle swarm optimization for job-shop scheduling problem[J]. Computers and Industrial Engineering, 2006, 51(4): 791-808.
- [7] 贺仁杰, 陈宇宁, 姚锋, 等. 求解柔性作业车间调度问题的知识型协同演化方法[J]. 计算机集成制造系统, 2011, 17(2): 310-315.
- [8] 李莉. 基于自适应参数与混沌搜索的 PSO 算法求解柔性作业车间调度问题[J]. 计算机应用, 2012, 32(7): 1932-1934, 1950.
- [9] 王伟玲, 李俊芳, 王晶. 求解多目标作业车间调度问题的双种群

遗传算法[J]. 计算机集成制造系统, 2011, 17(4): 808-815.

- [10] 刘爱军, 杨育, 朱明华, 等. 基于人机协同优化配置的多目标动态车间调度[J]. 系统工程, 2010, 28(3): 46-52.
- [11] GEEM Z W, KIM J H, LOGANATHAN G V. A new heuristic optimization algorithm: harmony search[J]. Simulation, 2001, 76(2): 60-68.
- [12] GEEM Z, KIM J, LOGANATHAN G. Harmony search optimization: application to pipe network design[J]. International Journal of Model Simulation, 2002, 22(2): 125-133.
- [13] 李亮, 迟世春. 新型和声搜索算法在土坡稳定分析中的应用[J]. 水利与建筑工程学报, 2007, 5(3): 1-6, 24.
- [14] GEEM Z W. Optimal cost design of water distribution networks using harmony search[J]. Engineering Optimization, 2006, 38(3): 259-277.
- [15] 武磊, 潘全科, 桑红燕, 等. 求解零空闲流水线调度问题的和声搜索算法[J]. 计算机集成制造系统, 2009, 15(10): 1960-1967.
- [16] 李宁, 刘建芹, 贺毅朝. 基于和声搜索算法求解组合优化问题[J]. 计算机应用, 2012, 32(4): 1041-1044.
- [17] STORN R, RICE K. Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces[J]. Journal of Global Optimization, 1997, 11(4): 341-359.
- [18] 刘志雄, 杨光祥. 基于演化策略算法的作业车间调度优化[J]. 计算机工程, 2010, 36(19): 8-10, 19.
- [19] 乔佩利, 马丽丽, 郑林. 基于改进粒子群算法的车间作业调度问题研究[J]. 哈尔滨理工大学学报, 2011, 16(2): 35-39.