



(12)发明专利申请

(10)申请公布号 CN 107220708 A

(43)申请公布日 2017.09.29

(21)申请号 201710484675.1

(22)申请日 2017.06.23

(71)申请人 哈尔滨工业大学

地址 150001 黑龙江省哈尔滨市南岗区西
大直街92号

(72)发明人 刘丹 王启松 刘志勇 刘昕
张岩 孙金玮

(74)专利代理机构 北京天奇智新知识产权代理
有限公司 11340

代理人 范光晔

(51)Int.Cl.

G06N 3/12(2006.01)

G06K 9/62(2006.01)

A61B 5/0476(2006.01)

A61B 5/00(2006.01)

权利要求书1页 说明书5页 附图3页

(54)发明名称

用于睡眠脑电分期特征选择的自适应模拟
退火遗传算法

(57)摘要

本发明公开了一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法,通过脑电信号进行睡眠分期,需要从脑电信号中提取大量特征参数,从中筛选出相对最优的特征参数组合用于睡眠脑电数学模型的建立。在已有的模拟退火遗传算法中,保留了遗传算法较强的总体搜索能力和模拟退火算法较强的局部搜索能力,以期提高产生优良个体的概率。但现有算法对迭代过程中的个体进行模拟退火操作时,在当前最优解的邻域内随机产生新解的机制存在致命的缺陷,本发明针对这一缺陷,不仅解决了传统模拟退火遗传算法邻域新解产生机制的迭代效率低、受邻域范围影响大等缺点,而且能够实现交叉概率和变异概率的自适应调整,同时采用加权方式进行适应度函数的设计。

1. 一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法,其特征在于,首先将自适应遗传算法交叉概率和变异概率的自适应调整机制加入到模拟退火遗传算法中;其次,采用基因优化算法代替模拟退火遗传算法中的邻域随机选择机制;最后,设计加权适应度函数,保证分类结果的正确率和相似性。

2. 根据权利要求1所述的用于睡眠脑电分期特征选择的自适应模拟退火遗传算法,其特征在于,具体步骤如下:

步骤一、特征参数编码:设初始特征向量为 $X = \{x_1, x_2, \dots, x_n\}$, n 为待选择特征参数个数;采用0/1对其进行编码,1代表选中,0代表未选中;例如, $S = \{1, 1, 0, 0, 1, 0, 0, \dots, 0\}$,表示选择了特征参数 x_1 、 x_2 和 x_5 三个特征参数;

步骤二、适应度函数设计:适应度函数具体描述如式(1)所示:

$$F(S_i) = \omega \times f(S_i) + (1 - \omega) \times c(S_i), i = 1, 2, \dots, n \quad (1)$$

其中, $f(S_i)$ 为第 i 个特征参数编码的分类精度, $c(S_i)$ 为第 i 个特征参数编码分类结果与目标结果的相似系数, ω 为调整系数,取0.6~0.8;

步骤三、自适应交叉操作:通过式(2)所示交叉概率 p_c ,调整机制进行自适应交叉操作,选择算术交叉Arithmetic Crossover算子;

$$p_c = \begin{cases} \frac{k_1(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_2, & f' < f_{\text{avg}} \end{cases} \quad (2)$$

其中, f_{\max} 为当前群体中适应度的最大值, f_{avg} 为当前群体的适应度均值, f' 为两个交叉个体中适应度值较大的, k_1 和 k_2 分别为(0, 1)区间的常数;

两个体 S_A^t 和 S_B^t 交叉运算后的新个体为:

$$\begin{cases} S_A^{t+1} = \alpha S_B^t + (1 - \alpha) S_A^t \\ S_B^{t+1} = \alpha S_A^t + (1 - \alpha) S_B^t \end{cases} \quad (3)$$

其中, α 为常数;

步骤四、自适应变异操作:通过式(4)所示变异概率 p_m ,调整机制进行自适应变异操作,选择高斯变异Gaussian Mutation操作算子;

$$p_m = \begin{cases} \frac{k_3(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_4, & f' < f_{\text{avg}} \end{cases} \quad (4)$$

其中, f_{\max} 为当前群体中适应度的最大值, f_{avg} 为当前群体的适应度均值, f' 为两个交叉个体中适应度值较大的, k_3 和 k_4 分别为(0, 1)区间的常数;

步骤五、个体模拟退火操作:对当前最优解 x_{best} 进行变异操作,轮流对每一位进行变异,可得到 n 个解空间 $\{S_1, S_2, \dots, S_n\}$;按照式(1)计算适应度值,并将 n 个 $F(S_i)$ 按照由大到小的顺序排列;采用基于排序的轮盘赌选择机制,根据幂律分布函数 $P_k = k^{-\tau}$ ($1 \leq k \leq n$)选择第 k 位的解为新解 x_{new} ;

步骤六、按照式(1)评价新解 x_{new} 的适应度,根据终止条件判断循环是否终止。

用于睡眠脑电分期特征选择的自适应模拟退火遗传算法

技术领域

[0001] 本发明涉及一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法,属于自适应模拟退火遗传算法技术领域。

背景技术

[0002] 特征选择是通过某种算法从全部特征参数集中筛选出少量的最适合于区分不同任务的特征参数。当特征参数数量较少时,可通过穷举法计算所有的特征参数组合来筛选最优特征参数。而当特征参数数量较多时,穷举法几乎不可能实现。随着模式识别算法的发展,特征参数选择已成为优化特征参数分类模型的重要手段。特征参数选择不仅可以简化所建立的特征分类模型,还可以最大程度地对输入变量进行优化选择,去除与输出变量无关的输入变量,进而提高分类模型的泛化特性和稳健性。

[0003] 目前,脑电信号特征参数选择算法应用较多的有聚类分析法和智能优化算法等。聚类分析法通过参数之间的可分离性、交互信息等参数选择。另一类重要的特征参数选择算法是智能优化算法,此类算法通过模拟生物进化论优胜劣汰等自然规律,以实际问题的最优解作为训练目标函数进行反复的迭代运算,筛选出最佳的特征参数组合。其中,遗传算法因具有全局最优、易于实现、鲁棒性较高等优势已经成功应用于许多领域。该算法通过编码方式将待选择变量模拟为遗传学中的染色体,通过遗传、交叉和变异的方式对染色体编码进行优化迭代,最终得到最优的结果。算法操作过程中以适应度函数作为目标函数,具有高效的并行计算特点,提高了选择效率。但是,该算法的不足之处在于:(1)需要根据实际优化问题自行设计适应度函数的类型和参数,没有通用的适应度函数标准;(2)易陷入局部最优;(3)收敛速度慢。近年来,人们从编码方式、控制参数的确定和交叉机理等方面对基本遗传算法进行了改进,改善了其易快速收敛到局部最优的缺点,提高了收敛速度。其中,最具代表性的是自适应遗传算法(Adaptive Genetic Algorithm,AGA)和模拟退火遗传算法(Simulated Annealing Genetic Algorithm,SAGA)。

[0004] 自适应遗传算法通过个体适应度值的当前趋势,对交叉概率和变异概率进行调整,克服了基本遗传算法过早收敛到局部最优的缺点。模拟退火遗传算法同时保留了模拟退火算法较强的局部搜索能力以及遗传算法较强的总体搜索能力,通过两种方法的优势互补,提高了寻优能力。但该算法在对迭代过程中的个体进行模拟退火操作时,在当前最优解的邻域内随机产生新解的机制存在致命缺陷,严重影响解的质量。

[0005] 综上所述,目前已有的模拟退火遗传算法特在特征选择过程中的邻域设置没有严格的定义,邻域区间的选择对于解得稳定性影响较大,严重的还会造成迭代结果不收敛。

发明内容

[0006] 本发明的目的是为了解决上述现有技术存在的问题,进而提供一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法。

[0007] 本发明的目的是通过以下技术方案实现的:

[0008] 一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法，

[0009] 首先将自适应遗传算法交叉概率和变异概率的自适应调整机制加入到模拟退火遗传算法中；其次，采用基因优化算法代替模拟退火遗传算法中的邻域随机选择机制；最后，设计加权适应度函数，保证分类结果的正确率和相似性。

[0010] 具体步骤如下：

[0011] 步骤一、特征参数编码：设初始特征向量为 $X = \{x_1, x_2, \dots, x_n\}$ ， n 为待选择特征参数个数；采用0/1对其进行编码，1代表选中，0代表未选中；例如， $S = \{1, 1, 0, 0, 1, 0, 0, \dots, 0\}$ ，表示选择了特征参数 x_1 、 x_2 和 x_5 三个特征参数；

[0012] 步骤二、适应度函数设计：适应度函数具体描述如式(1)所示：

$$[0013] \quad F(S_i) = \omega \times f(S_i) + (1 - \omega) \times c(S_i), i = 1, 2, \dots, n \quad (1)$$

[0014] 其中， $f(S_i)$ 为第 i 个特征参数编码的分类精度， $c(S_i)$ 为第 i 个特征参数编码分类结果与目标结果的相似系数， ω 为调整系数，取0.6~0.8；

[0015] 步骤三、自适应交叉操作：通过式(2)所示交叉概率 p_c ，调整机制进行自适应交叉操作，选择算术交叉Arithmetic Crossover算子；

$$[0016] \quad p_c = \begin{cases} \frac{k_1(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_2, & f' < f_{\text{avg}} \end{cases} \quad (2)$$

[0017] 其中， f_{\max} 为当前群体中适应度的最大值， f_{avg} 为当前群体的适应度均值， f' 为两个交叉个体中适应度值较大的， k_1 和 k_2 分别为(0, 1)区间的常数；

[0018] 两个体 S_A^t 和 S_B^t 交叉运算后的新个体为：

$$[0019] \quad \begin{cases} S_A^{t+1} = \alpha S_B^t + (1 - \alpha) S_A^t \\ S_B^{t+1} = \alpha S_A^t + (1 - \alpha) S_B^t \end{cases} \quad (3)$$

[0020] 其中， α 为常数。

[0021] 步骤四、自适应变异操作：通过式(4)所示变异概率 p_m ，调整机制进行自适应变异操作，选择高斯变异Gaussian Mutation操作算子；

$$[0022] \quad p_m = \begin{cases} \frac{k_3(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_4, & f' < f_{\text{avg}} \end{cases} \quad (4)$$

[0023] 其中， f_{\max} 为当前群体中适应度的最大值， f_{avg} 为当前群体的适应度均值， f' 为两个交叉个体中适应度值较大的， k_3 和 k_4 分别为(0, 1)区间的常数；

[0024] 步骤五、个体模拟退火操作：对当前最优解 x_{best} 进行变异操作，轮流对每一位进行变异，可得到 n 个解空间 $\{S_1, S_2, \dots, S_n\}$ ；按照式(1)计算适应度值，并将 n 个 $F(S_i)$ 按照由大到小的顺序排列；采用基于排序的轮盘赌选择机制，根据幂律分布函数 $P_k = k^{-\tau}$ ($1 \leq k \leq n$)选择第 k 位的解为新解 x_{new} ；

[0025] 步骤六、按照式(1)评价新解 x_{new} 的适应度，根据终止条件判断循环是否终止。

[0026] 本发明通过脑电信号进行睡眠分期，需要从脑电信号中提取大量特征参数，从中筛选出相对最优的特征参数组合用于睡眠脑电数学模型的建立。在已有的模拟退火遗传算法中，保留了遗传算法较强的总体搜索能力和模拟退火算法较强的局部搜索能力，以期提

高产生优良个体的概率。本发明提出一种自适应模拟退火遗传算法,不仅解决了传统模拟退火遗传算法邻域新解产生机制的迭代效率低、受邻域范围影响大等缺点,而且能够实现交叉概率和变异概率的自适应调整,同时采用加权方式进行适应度函数的。相对于其它算法,本发明所提出的ASAGA算法的特征筛选效果最优,当特征参数维数为10时,分类正确率最高,达到了92%。

附图说明

- [0027] 图1为相关系数算法特征参数排序图。
 [0028] 图2为遗传算法的特征参数排序图。
 [0029] 图3为自适应遗传算法的特征参数排序图。
 [0030] 图4为模拟退火遗传算法的特征参数排序图。
 [0031] 图5为自适应模拟退火遗传算法的特征参数排序图。
 [0032] 图6为特征参数选择算法分类对比图。

具体实施方式

[0033] 下面将结合附图对本发明做进一步的详细说明:本实施例在以本发明技术方案为前提下进行实施,给出了详细的实施方式,但本发明的保护范围不限于下述实施例。

[0034] 本实施例所涉及的一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法,首先将自适应遗传算法交叉概率和变异概率的自适应调整机制加入到模拟退火遗传算法中;其次,采用基因优化算法代替模拟退火遗传算法中的邻域随机选择机制;最后,设计加权适应度函数,保证分类结果的正确率和相似性。

[0035] 具体步骤如下:

[0036] 步骤一、特征参数编码:设初始特征向量为 $X = \{x_1, x_2, \dots, x_n\}$, n 为待选择特征参数个数。采用0/1对其进行编码,1代表选中,0代表未选中。例如, $S = \{1, 1, 0, 0, 1, 0, 0, \dots, 0\}$,表示选择了特征参数 x_1 、 x_2 和 x_5 三个特征参数。

[0037] 步骤二、适应度函数设计:适应度函数直接影响ASAGA算法的寻优结果,与实际问题的密切相关。为了提高睡眠分期正确率,本发明在适应度函数设计中同时考虑了睡眠分期结果的正确率,以及分类结果与目标结果的相似性。适应度函数具体可以描述为式(1)所示。

$$[0038] \quad F(S_i) = \omega \times f(S_i) + (1 - \omega) \times c(S_i), i = 1, 2, \dots, n \quad (1)$$

[0039] 其中, $f(S_i)$ 为第 i 个特征参数编码的分类精度, $c(S_i)$ 为第 i 个特征参数编码分类结果与目标结果的相似系数, ω 为调整系数,一般取0.6~0.8。

[0040] 步骤三、自适应交叉操作:通过式(2)所示交叉概率 p_c ,调整机制进行自适应交叉操作,选择算术交叉(Arithmetic Crossover)算子。

$$[0041] \quad p_c = \begin{cases} \frac{k_1(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_2, & f' < f_{\text{avg}} \end{cases} \quad (2)$$

[0042] 其中, f_{\max} 为当前群体中适应度的最大值, f_{avg} 为当前群体的适应度均值, f' 为两个交叉个体中适应度值较大的, k_1 和 k_2 分别为(0, 1)区间的常数。

[0043] 两个个体 S_A^t 和 S_B^t 交叉运算后的新个体为:

$$[0044] \begin{cases} S_A^{t+1} = \alpha S_B^t + (1-\alpha)S_A^t \\ S_B^{t+1} = \alpha S_A^t + (1-\alpha)S_B^t \end{cases} \quad (3)$$

[0045] 其中, α 为常数。

[0046] 步骤四、自适应变异操作:通过式(4)所示变异概率 p_m ,调整机制进行自适应变异操作,选择高斯变异(Gaussian Mutation)操作算子。与基本遗传算法编码的变异操作不同,高斯变异算子将一组服从正态分布的随机数代替原编码。由于变异后的解服从正态分布,因此该算法具有较强的局部搜索能力。

$$[0047] p_m = \begin{cases} \frac{k_3(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_4, & f' < f_{\text{avg}} \end{cases} \quad (4)$$

[0048] 其中, f_{\max} 为当前群体中适应度的最大值, f_{avg} 为当前群体的适应度均值, f' 为两个交叉个体中适应度值较大的, k_3 和 k_4 分别为(0,1)区间的常数。

[0049] 步骤五、个体模拟退火操作:对交叉、变异后的个体采用模拟退火算法进行局部寻优操作。与常规模拟退火算法不同的是,对当前最优解 x_{best} ,按照基因优化算法(GOA)产生新解 x_{new} 。具体方法是:对当前最优解 x_{best} 进行变异操作,轮流对每一位进行变异,可得到 n 个解空间 $\{S_1, S_2, \dots, S_n\}$ 。按照式(1)计算适应度值,并将 n 个 $F(S_i)$ 按照由大到小的顺序排列。采用基于排序的轮盘赌选择机制,根据幂律分布函数 $P_k = k^{-1}$ ($1 \leq k \leq n$)选择第 k 位的解为新解 x_{new} 。

[0050] 步骤六、按照式(1)评价新解 x_{new} 的适应度,根据终止条件判断循环是否终止。

[0051] 实验验证:

[0052] 为了验证自适应模拟退火遗传算法的有效性,以MIT-BIH多导睡眠数据库的样本为分析对象,分别提取25个特征参数。将25个特征参数依次记为 $\{x_i, i=1, 2, \dots, 25\}$ 。将自适应模拟退火遗传算法(ASAGA)与现有的相关系数法(CCA)、普通遗传算法(GA)、自适应遗传算法(AGA)以及模拟退火遗传算法(SAGA)的特征选择效果进行对比验证。将数据库中所有样本分为5份,分别依次选择其中4份为校正样本,剩余1份为验证样本。轮流执行5次,以5次分类结果正确率的均值作为评价特征参数选择结果的标准。

[0053] (1) 基于相关系数的特征选择

[0054] 相关系数的特征选择算法(Correlation Coefficient Algorithm, CCA)是将各特征参数按其于睡眠状态相关系数绝对值由大到小的顺序进行排列,依次选择前1~25个参数进行特征分类。其中,排名前10的特征参数如图1所示。

[0055] (2) 基于智能优化算法的特征选择

[0056] 随机产生一个初始特征向量 $X = \{x_1, x_2, \dots, x_n\}$,分别运行GA, AGA, SAGA, ASAGA等算法各100次,统计每个特征参数被选中的频率,按照从大到小的顺序排列,依次选择前1~25个参数进行特征分类。各自排名前10的特征参数如图2~图5所示。

[0057] 5种特征参数选择算法分别依次选择1~25个参数的分类结果对比如图6所示。其中,当参数个数为5、10、15、20和25时,各特征参数选择算法的分类正确率如表1所示。由图6和表1可知,采用相关系数法进行特征参数筛选时,分类结果的正确率随着特征参数数量的

增加呈上升趋势。采用GA、AGA、SAGA以及ASAGA等优化算法的结果总体呈现先上升后下降的趋势,造成这种现象的原因可以归结为:当参与建模的参数过多时,出现了过拟合现象,使得模型的分类误差增加,分类正确率降低。另一方面也说明优化算法可以从所有特征参数中筛选出少量的最适应建模的特征参数组合。由图6可知,GA、AGA以及SAGA在特征参数选择的效果上差别较小,筛选出的最佳特征组合的特征维数与分类正确率也较为接近。相对于其它四种算法,本发明所提出的ASAGA算法的特征筛选效果最优,当特征参数维数为10时,分类正确率最高,达到92%。

[0058] 表1各特特征参数选择算法在不同特征维数下的分类正确率对比

[0059]

特征选择 方法	不同特征维数下的分类正确率(%)				
	5	10	15	20	25
CCA	77.83	81.17	83.83	86.17	86.83
GA	80.17	85.67	87.67	87.17	87.00
AGA	84.83	89.00	87.83	87.33	87.17
SAGA	83.67	87.83	88.83	87.17	87.83
ASAGA	88.17	92.00	89.83	87.83	87.17

[0060] 以上所述,仅为本发明较佳的具体实施方式,这些具体实施方式都是基于本发明整体构思下的不同实现方式,而且本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求书的保护范围为准。

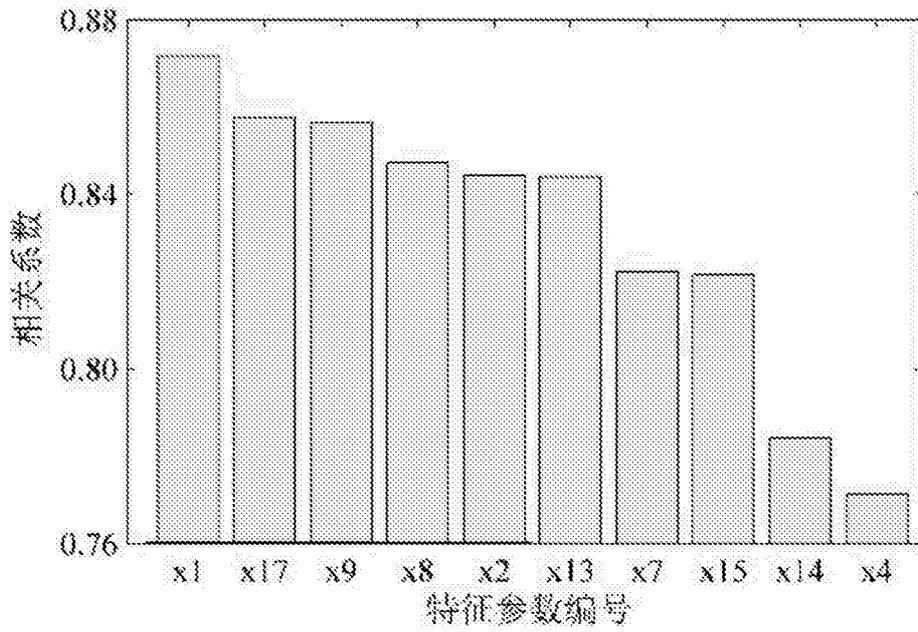


图1

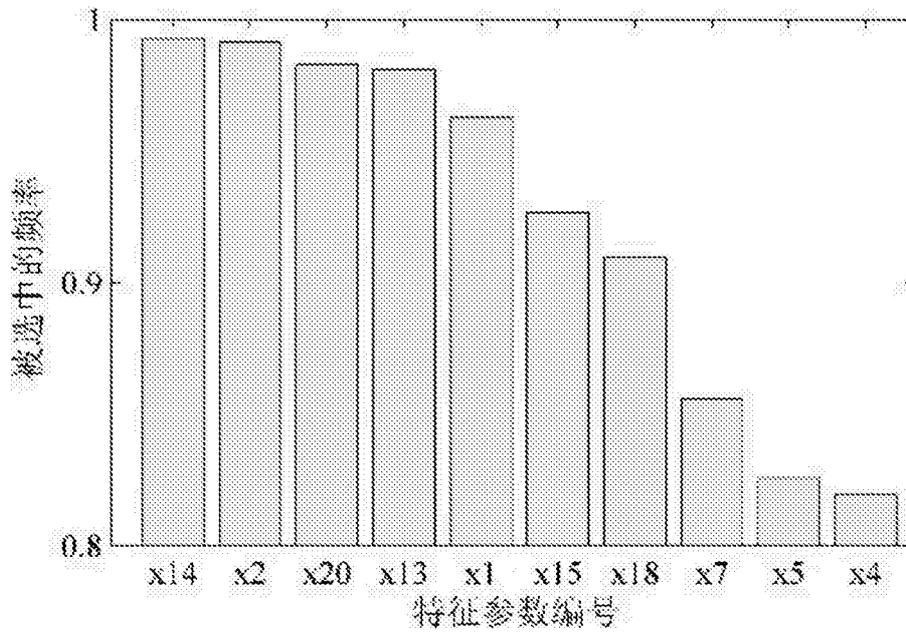


图2

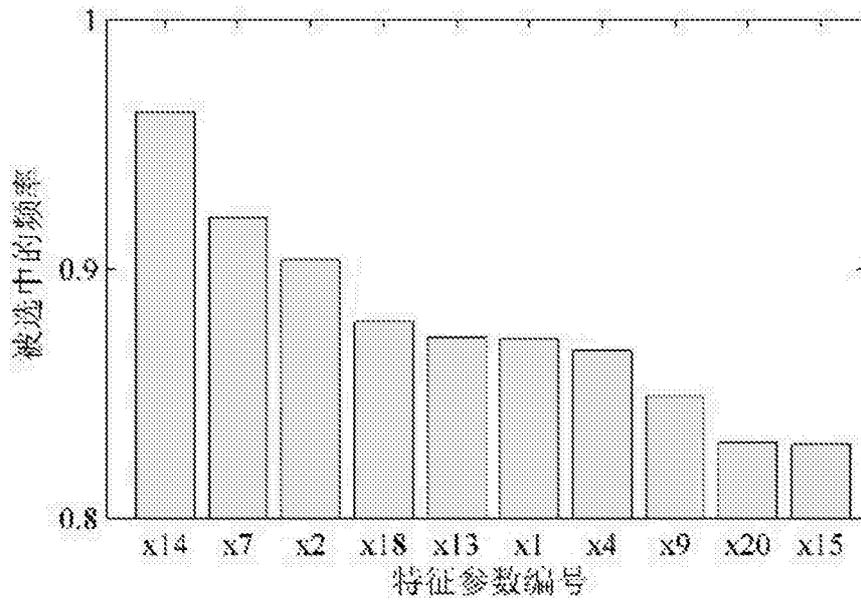


图3

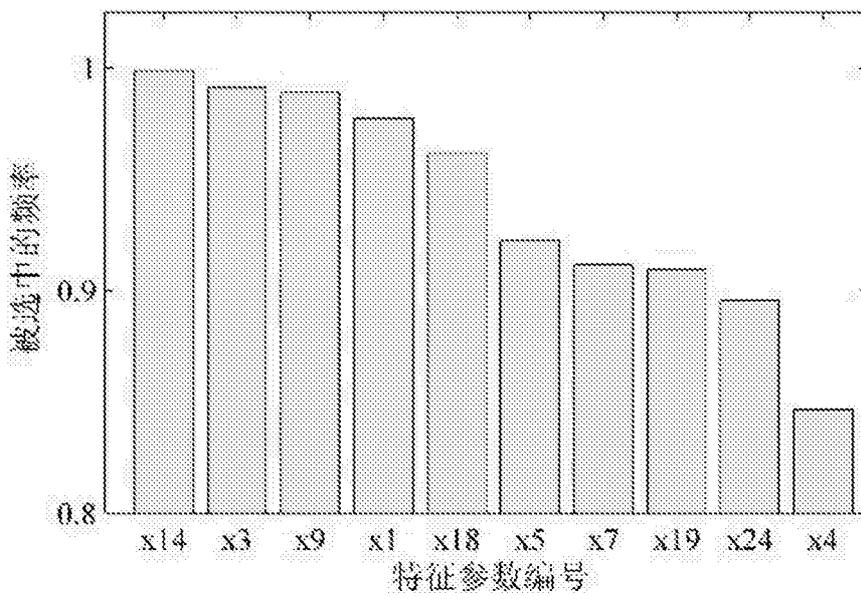


图4

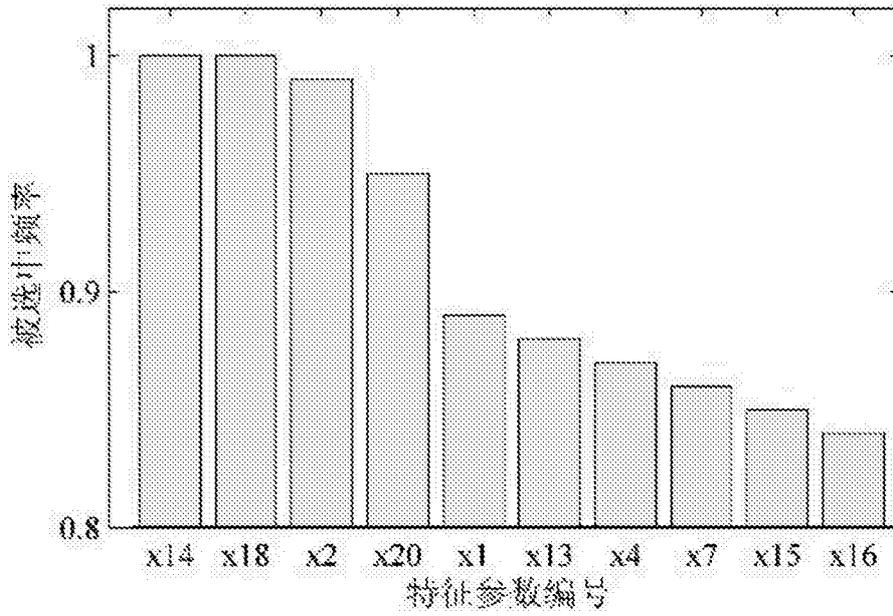


图5

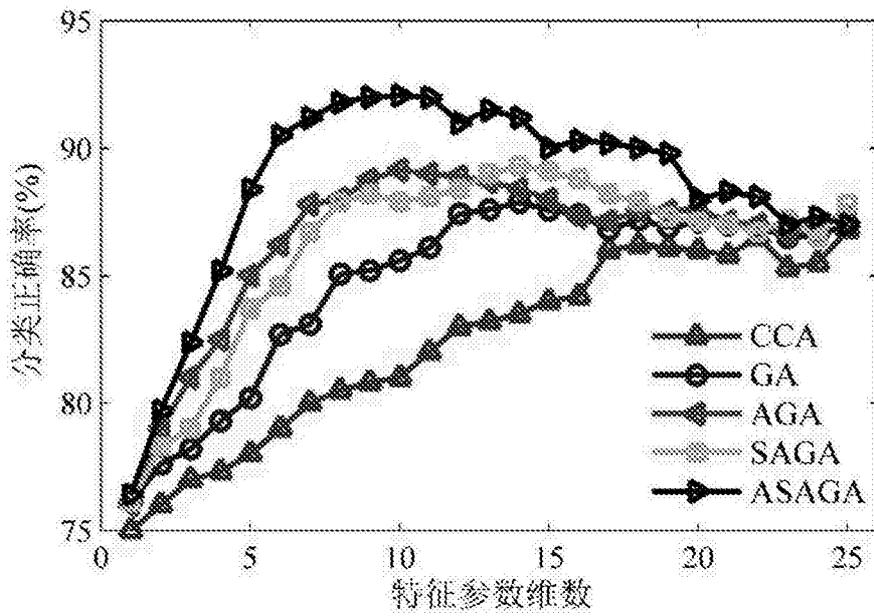


图6

专利名称(译)	用于睡眠脑电分期特征选择的自适应模拟退火遗传算法		
公开(公告)号	CN107220708A	公开(公告)日	2017-09-29
申请号	CN2017110484675.1	申请日	2017-06-23
[标]申请(专利权)人(译)	哈尔滨工业大学		
申请(专利权)人(译)	哈尔滨工业大学		
当前申请(专利权)人(译)	哈尔滨工业大学		
[标]发明人	刘丹 王启松 刘志勇 刘昕 张岩 孙金玮		
发明人	刘丹 王启松 刘志勇 刘昕 张岩 孙金玮		
IPC分类号	G06N3/12 G06K9/62 A61B5/0476 A61B5/00		
CPC分类号	A61B5/0476 A61B5/4812 A61B5/7264 G06K9/6229 G06N3/126		
外部链接	Espacenet SIPO		

摘要(译)

本发明公开了一种用于睡眠脑电分期特征选择的自适应模拟退火遗传算法，通过脑电信号进行睡眠分期，需要从脑电信号中提取大量特征参数，从中筛选出相对最优的特征参数组合用于睡眠脑电数学模型的建立。在已有的模拟退火遗传算法中，保留了遗传算法较强的总体搜索能力和模拟退火算法较强的局部搜索能力，以期提高产生优良个体的概率。但现有算法对迭代过程中的个体进行模拟退火操作时，在当前最优解的邻域内随机产生新解的机制存在致命的缺陷，本发明针对这一缺陷，不仅解决了传统模拟退火遗传算法邻域新解产生机制的迭代效率低、受邻域范围影响大等缺点，而且能够实现交叉概率和变异概率的自适应调整，同时采用加权方式进行适应度函数的设计。

$$p_c = \begin{cases} \frac{k_1(f_{\max} - f')}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ k_2, & f' < f_{\text{avg}} \end{cases} \quad (2)$$