



(12)发明专利

(10)授权公告号 CN 106371610 B

(45)授权公告日 2020.06.09

(21)申请号 201610846881.8

A61B 5/00(2006.01)

(22)申请日 2016.09.23

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 106371610 A

CN 105740887 A, 2016.07.06,

张毅等. 脑电信号的小波变换和样本熵特征提取方法.《智能系统学报》.2012,第7卷(第4期),第339-344页.

(43)申请公布日 2017.02.01

金纯等. 基于脑电的实时疲劳检测算法的研究.《科学技术与工程》.2015,第15卷(第34期),

(73)专利权人 重庆金瓯科技发展有限责任公司
地址 400041 重庆市九龙坡区二郎留学生
创业园D1栋8-2

第231-234页.

(72)发明人 金纯

审查员 李妮

(74)专利代理机构 重庆百润洪知识产权代理有限公司 50219

代理人 高姜

(51)Int.Cl.

G06F 3/01(2006.01)

G06K 9/62(2006.01)

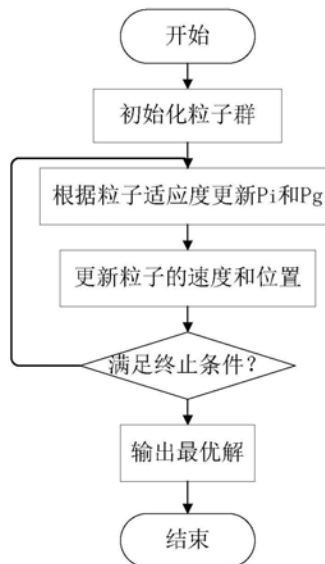
权利要求书6页 说明书16页 附图1页

(54)发明名称

一种基于脑电信号的驾驶疲劳的检测方法

(57)摘要

本发明涉及一种基于脑电信号的驾驶疲劳的检测方法,具体包括以下方面:(1)选取与疲劳驾驶状态相关的特征量;(2)疲劳驾驶状态的脑电信号功率谱分析;(3)疲劳驾驶状态的脑电信号样本熵分析;(4)疲劳驾驶状态的脑电信号Kc复杂度分析;(5)支持向量机(SVM)(6)最小二乘支持向量机(LS-SVM);(7)粒子群优化(PSO)算法设置LS-SVM的模型训练参数C和g。本发明通过对不同驾驶状态下提取到的脑电信号分别从功率谱的角度以及使用非线性动力学中的相关方法分别进行了研究,表明在准确性和可靠性上都取得较好的效果。



1. 一种基于脑电信号的驾驶疲劳的检测方法,其特征在于:具体包括以下方面:

(1) 选取与疲劳驾驶状态相关的特征量:是指通过对原始测量信号进行一系列相应处理后,得到一组由相关特征值组成的特征集,然后从该特征集中选取某些具有代表性、分类性能最佳的物理量作为区分不同行为的一组特征量,然后使用相关的分类识别方法,利用选取的特征量将不同的行为从特征空间中分离出来,因此对于学习分类而言都需要训练集样本:

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l \quad (1)$$

其中 $x_i \in X, X = R, y_i \in Y, Y = R, i = 1, \dots, l$;

(2) 疲劳驾驶状态的脑电信号功率谱分析:对处于不同状态下脑电信号的功率谱特性进行相关分析,将采集的脑电信号的长度选择为3秒,即对采集到的数据以每3秒截取一次并进行平均功率的计算,选取了从精力充沛状态过度到疲劳状态的脑电数据8分钟,将其截取为160段,并对截取到的这些时间段的脑电信号进行傅里叶变换:

$$G = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_T^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{2T} x_T^2(f) df \quad (2)$$

式中, f 为连续频率变量; $x_T(f)$ 为信号的频谱密度;

其中:

$$p_X(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} x_T^2(f) \quad (3)$$

在式(3)中显示了处于不同频段上的信号功率的分布情况,同时也可以表示单位时间内和单位频段上的信号的能量,因此该积分函数(3)又被称作为功率谱密度函数;经过傅里叶变换以后,得到 $x_n(e^{j\omega})$, 然后对频谱的模进行平方,最后得到功率谱密度的计算公式:

$$p_X(f) = \frac{1}{N} |X_N(f)|^2 \quad (4)$$

由上面可知信号在某个频段 h 的平均功率密度为:

$$G(h) = \frac{\int_{f_l}^{f_h} p_X(f) df}{f_h - f_l} \quad (5)$$

其中: f_h 为频段 h 的上限, f_l 为频段 h 的下限, $p_X(f)$ 为信号的功率谱密度;则不同频段的平均功率密度比值则定义为:

$$F = \frac{G(h_1)}{G(h_2)} \quad (6)$$

式中 h_1 和 h_2 为脑电信号的不同频段,将 $F(\alpha+\theta)/\beta, F\alpha/\beta$ 这两个相关节律波的平均功率密度的比值选作进行驾驶疲劳判别的两个特征指标;

(3) 疲劳驾驶状态的脑电信号样本熵分析:利用样本熵对脑电信号进行分析的具体步骤如下:计算选取长度为1000字符的数据作为一个数据窗口,以500字符的数据为移动步长,按以上步骤计算每次移动的数据的样本熵值,然后进行下1个数据窗口的样本熵值的计算,取2个数据窗口的样本熵的平均值作为时间序列的脑电数据样本熵值,通过对处于疲劳与清醒状态的脑电数据进行分析,得到不同状态下脑电信号的样本熵值;非疲劳状态时的

样本熵值大于疲劳状态下的样本熵值,表明样本熵可以用于区分不同的驾驶状态,进一步可以得出疲劳状态下脑电信号计算得到样本熵值主要集中在0.2~0.4,而清醒状态的样本熵值集中在0.5~0.8之间;由此可知随着驾驶疲劳程度的加深脑电信号的样本熵值逐渐降低的,因此可看出脑电信号的样本熵值是与人体精神状态之间存在相互对应关系,因此将其选作为用于驾驶疲劳识别的相关特征量;

(4) 疲劳驾驶状态的脑电信号Kc复杂度分析:首先采集得到的一段待分析的脑电时间序列,设为 $S(S_1, S_2, \dots, S_n)$,第一步是对该序列进行粗粒化处理,具体就是先求取该序列的平均值,然后对原序列的每个元素进行判断并重新赋值,如果序列元素 S_i 大于平均值时则赋值 $S_i' = 1$,如果序列元素 S_i 小于平均值时则赋值 $S_i' = 0$,然后得到重构序列 $S'(S'_1, S'_2, \dots, S'_n)$,利用 S_i' 重新对 S 中的元素进行赋值,则得到新的序列 $S_x(S_1, S_2, \dots, S_n)$,该新序列是一串由(0,1)字符组成的新序列;下一步则对重构序列进行子串划分与界定,将按照如下的规则进行:

在一个字符串 $S(S_1, S_2, \dots, S_n)$ 后再加一个字符串 $Q(q_1, q_2, \dots, q_n)$,这样就得到组合后新的字符串序列 SQ ,令 SQ 序列删去末尾的一个字符所得字符串命名为 SQ_v ,接下来需要对 Q 是否属于 SQ_v 的子串进行判断,如果属于,则把下一个字符 q_{n+1} 加到 Q 后面,增长一位再进行重复判断,直到增长后的 Q 不再属于 SQ_v 的一个子串,则用标记符号“.”进行序列的前后界定;下一步将“.”前的所有字符定义为新的 S ,重新以其后的字符开始构造 Q ,重复以上的判断和界定过程,直到整个字符串的序列判断结束;统计最终由“.”界定划分出的 S 的子串数目就是序列的Kc复杂度;

(5) 支持向量机:考虑样本 $\{(x_i, y_i)\}_{i=1}^N$,其中 x_i 代表第 i 个输入样本, y_i 则是对应输入的输出;首先假设由子集 $y_i = +1$ 代表的类和 $y_i = -1$ 代表的类是能够进行线性可分的,那么用于进行分离的超平面形式的决策曲面方程可以表示为如下:

$$w^T x + b = 0 \quad (7)$$

x ——输入变量;

w ——可调的权值向量;

b ——偏值;

那么式(7)则可以写成:

$$w^T x + b \geq 0, \text{ 当 } y_i = +1 \quad (8)$$

$$w^T x + b \leq 0, \text{ 当 } y_i = -1 \quad (9)$$

由式(8)和(9)则可合并为:

$$y_i (w^T x + b) \geq 0 \quad (10)$$

对于一个给定的权值向量 w 和偏值 b ,数据点到公式(1)定义的超平面的最近间隔称为分离边缘,用 ρ 表示,而支持向量机的目标则是寻找一个使得分离边缘 ρ 最大的超平面,

而这个决策曲面则被称为最优超平面;

假设 w_0 和 b_0 分别为得到的权值向量和偏值的最优值;那么此时,输入空间里表示的多维线性决策面的最优超平面则可以表示如下:

$$w_0^T x + b_0 = 0 \quad (11)$$

假如 x_k 对应于 y_k ,当满足 $y_k (w_0^T x_k + b_0) = 1$ 时,那么则称 (x_k, y_k) 是支持向量的,那么此时

支持向量到最优超平面的距离为：

$$r = \begin{cases} \frac{1}{\|w_0\|} & y_k = +1 \\ -\frac{1}{\|w_0\|} & y_k = -1 \end{cases} \quad (12)$$

其中的加号表示 x_k 在最优超平面的正面，负号则表示最优超平面的负面，式(12)可以得到分类间隔 r ：

$$\rho = 2r = \frac{2}{\|w_0\|} \quad (13)$$

由上式得到的最优解则对应于分类间隔最大的情况，其等价于凸二次最优化问题：

$$\begin{cases} \min \frac{1}{2} w^T \bullet w \\ \text{s. t. : } y_i(w^T x_i + b) \geq 1, i=1, \dots, m \end{cases} \quad (14)$$

由上式得到的最优解则是由拉格朗日(Lagrange)函数的鞍点决定：

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m a_i [y_i(w^T x_i + b) - 1] \quad (15)$$

a_i, a_j ——拉格朗日乘子；

由最优解KKT条件可得：

$$\left. \begin{aligned} \frac{\partial L(w, b, a)}{\partial b} = 0 &\Rightarrow \sum_{i=1}^m a_i y_i = 0 \\ \frac{\partial L(w, b, a)}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^m a_i x_i y_i \end{aligned} \right\} \quad (16)$$

根据KKT(Karush-Kuhn-Tucker)条件，相应的拉格朗日乘子 a_i 应满足如下条件：

$$a_i [y_i(w^T x_i + b) - 1] = 0 \quad (17)$$

从上式可以看出只有支持向量的 a_i 不为零；

根据核函数变换，则式(15)和(17)可转化为对偶问题求解，即：

$$\left\{ \begin{aligned} W(a) &= \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j x_i^T x_j \\ \text{s.t. } \sum_{i=1}^N y_i a_i &= 0 \quad a_i \geq 0, i=1, \dots, N \end{aligned} \right. \quad (18)$$

然后由公式(16)可以求出最优权值向量 w_0 ，以及利用式子(17)可以得到 b_0 ，于是关于待测样本 x ，通过只算 $\text{sgn}[(w^T x) + b]$ 就可以判断其所属的类别；

在分类超平面的定义中引入松弛变量 $\xi_i \geq 0$ ，即允许少量误差的存在，那么公式(10)的约束条件变为如下形式：

$$y_i (w^T x + b) \geq 1 - \xi_i, i = 1, 2, \dots, N \quad (19)$$

此时的目标函数则变为:

$$\begin{cases} \min \frac{1}{2} w^T \bullet w + C \sum_{i=1}^N \xi_i \\ y_i (w^T x + b) \geq 1 - \xi_i, i = 1, 2, \dots, N \end{cases} \quad (20)$$

其中C为惩罚因子且 $C \geq 0$,当选取的参数C比较大时,表示支持向量机模型对训练样本的质量具有较高的信心,相反,当选取的参数C比较小时,则认为训练样本中存在噪声;由公式(20)得到相应的式子如下所示:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N a_i [y_i (w^T x_i + b) + \xi_i - 1] - \sum_{i=1}^N \beta_i \xi_i \quad (21)$$

那么根据KKT条件转化为其对偶问题如下:

$$W(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j x_i^T x_j \quad (22)$$

$$s.t. \sum_{i=1}^m y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, m \quad (23)$$

将输入空间 R^n 中的点利用非线性映射 $\Phi: R^n \rightarrow H$ 映射到高维的Hilbert空间当中去,然后再使用线性可分的方法进行最优分类面的求解,而在这个求解的过程中,仅仅涉及到内积的运算,并且计算量也只依赖于训练样本数的个数,因此这个过程就是如何找到一个核函数 $K(x_i, x_j)$, $x_i, x_j \in R^n$,使得该核函数满足 $K(x_i, x_j) = \Phi^T(x_i) \Phi(x_j)$,这样就可以保证即使在不知道变换函数 Φ 形式的情况下也能够求得最终的结果,同时能够很好地避免维数灾难的发生,因此在最大间隔情况下,非线性支持向量机的目标函数为:

$$W(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(x_i, x_j) \quad (24)$$

同样可得知,非线性支持向量机的最优化问题如下所示:

$$\begin{cases} \min \frac{1}{2} w^T \bullet w + C \sum_{i=1}^N \xi_i \\ s.t.: \sum_{i=1}^m y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, m \end{cases} \quad (25)$$

它的对偶问题则为:

$$\begin{cases} \max W(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(x_i, x_j) \\ s.t.: \sum_{i=1}^N y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, N \end{cases} \quad (26)$$

相应得到的分类函数为:

$$f(x) = \text{sgn}[(w^T \Phi(x)) + b] = \text{sgn}[\sum_{i=1}^N y_i a_i K(x_i, x_j) + b] \quad (27)$$

(6) 最小二乘支持向量机 (LS-SVM): 最小二乘支持向量机算法的描述如下: 首先选择一个非线性的变换 $\varphi(\bullet)$, 把 n 维输入和 1 维输出的样本向量 $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, $x_i \in \mathbb{R}$, $y_i \in \mathbb{R}, i=1, 2, \dots, N$, 从原空间映射到高维特征空间, 则分类 LS-SVM 的优化问题则变为:

$$\begin{cases} \min \frac{1}{2} w^T \bullet w + \frac{C}{2} \sum_{i=1}^N \xi_i^2 \\ y_i [w^T \bullet \varphi(x_i) + b] = 1 - \xi_i, i=1, 2, \dots, N \end{cases} \quad (28)$$

即将支持向量机的不等式约束改为等式约束, 构造拉格朗日函数进行求解:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N a_i [y_i (w^T x_i + b) - 1 + \xi_i] \quad (29)$$

其中 a_i —— 拉格朗日乘子;

最优解满足 KKT 条件则: 进而可以得到下面的方程组:

$$\begin{cases} w = \sum_{i=1}^N a_i y_i \varphi(x_i) \\ \sum_{i=1}^N a_i y_i = 0 \\ a_i = C \xi_i \\ y_i [w^T \varphi(x_i) + b] - 1 + \xi_i = 0 \end{cases} \quad (30)$$

对于 $i=1, 2, \dots, N$, 通过式子 (30) 消去 w 和 ξ 得到如下线性系统:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

式中:
$$\begin{cases} Z = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, \dots, \varphi(x_N)^T y_N] \\ Y = [y_1, y_2, \dots, y_N]^T \\ 1 = [1, 2, \dots, N]^T \\ a = [a_1, a_2, \dots, a_N] \end{cases} \quad (31)$$

令 $\Omega = ZZ^T + C^{-1}I$, 求解可得到:

$$\begin{cases} a = \Omega^{-1} (Y - b1) \\ b = \frac{1^T \Omega^{-1} Y}{1^T \Omega^{-1} 1} \end{cases} \quad (32)$$

最终得到的决策函数如下:

$$f(x) = \sum_{i=1}^N a_i K(x, x_i) + b; \quad (33)$$

(7) 粒子群优化PSO算法设置LS-SVM的模型训练参数C和g:采用最小二乘法支持向量机进行模型建立的过程来确定两个参数,即惩罚参数C和核函数参数g,粒子群算法具体如下:首先在可解空间内初始化一群粒子,而每一个粒子都同时拥有位置、速度和适应度值三个特征,而其中的适应度值的大小则决定该粒子的优劣,同时粒子的移动主要是通过跟踪个体极值Pbest、群体极值Gbest以及过去速度的惯性来更新个体位置,其中个体极值Pbest是指单个粒子所经历的适应度值最优的位置,群体极值Gbest是指种群中所有粒子所经历适应度值最优的位置,粒子每迭代一次就移动一次位置,其适应度值就相应的需要重新计算一次,然后通过对新粒子个体极值的适应度值、群体极值的适应度值与迭代之前的极值比较来确定是否更新粒子的个体极值和群体极值的位置;

假设在一个D维的空间中,其中 $X = (X_1, X_2, \dots, X_n)$ 为由n个粒子组成的种群,其中向量 $X_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ 表示第i个粒子,对于第i个粒子,它的适应度值一般就是根据粒子 X_i 在目标函数中的值,速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{id})^T$,其个体极值为 $P_i = (p_{i1}, p_{i2}, \dots, p_{id})^T$,种群的全局极值为 $P_g = (p_{g1}, p_{g2}, \dots, p_{gd})^T$

在每一次的迭代过程中,每个粒子都是通过个体极值、全局极值以及惯性速度来更新它的速度和位置信息,其更新公式(34)和(35)如下所示:

$$v_{id}^{k+1} = wv_{id}^k + c_1r_1(p_{id}^k - x_{id}^k) + c_2r_2(p_{gd}^k - x_{id}^k) \quad (34)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (35)$$

式(34)中,w为惯性权重,当w较大时有利于群体在较大的范围内进行搜索,从而避免陷入寻找局部最优,而较小的w能够保证算法能够快速收敛到最优值; $d=1, 2, \dots, D, i=1, 2, \dots, n, k$ 为当前迭代次数; v_{id} 为粒子的速度; c_1 和 c_2 为非负常数,称为加速度因子,是每一个粒子在 P_i 和 P_g 位置的随机加速权重, r_1 和 r_2 是 $[0, 1]$ 之间的随机数;为了防止粒子群的盲目搜索,一般把粒子群的位置限定在 $[-X_{max}, X_{max}]$ 和速度限定在 $[-V_{max}, V_{max}]$ 的范围内;

根据上面的分析,使用粒子群算法优化参数C和g,具体步骤如下:

- (a) 初始化粒子群并随机产生所有粒子的位置和速度并确定粒子的 P_i 和 P_g ;
- (b) 将每个粒子当前最优位置 P_i 及种群中所有粒子所经历的最好位置 P_g 进行比较,如果这个粒子的位置最优,则将其设置为当前的最好位置 P_g ,否则, P_g 保持不变;
- (c) 更新粒子的速度、位置;
- (d) 若没有达到结束条件,条件为:预设的运算精度、迭代次数,便返回步骤(b);
- (e) 开始下一次的迭代计算,否则取当前的 P_g 为最优解。

一种基于脑电信号的驾驶疲劳的检测方法

技术领域

[0001] 本发明属于交通驾驶技术领域,具体涉及一种基于脑电信号的驾驶疲劳的检测方法

背景技术

[0002] 如何从与疲劳状态紧密相关的众多特征中挑选出合适的特征量,对于进行驾驶疲劳的准确识别具有非常重要的作用,同时为了实现对驾驶疲劳进行客观而准确的识别,往往在对提取出来的脑电信号进行相关处理后,得到选取的相关的特征向量后,再使用某些分类器根据这些特征向量进行判别。支持向量机是常用的方法之一,它在实际应用中虽然能够取得不错的效果,但是它需要借助二次规划的方法进行求解,在样本数量比较大的情况下需要耗费大量的内存以及运算时间。本发明通过对提取到的不同驾驶状态的脑电信号分别从功率谱的角度以及采用非线性动力学的相关方法进行了研究,探讨了不同的驾驶状态与这些特征值之间的存在的对应关系,进行相关分析比较后选取了脑电信号的Kc复杂度、样本熵以及相关节律波的平均功率的比值等多个特征值构成了特征向量空间,作为构成驾驶疲劳识别的主要特征空间,然后采用最小二乘支持向量机对提取的样本信号的特征向量进行驾驶疲劳识别模型的训练,从而避免了解二次规划的问题,提高了模型训练的速度,同时针对LS-SVM参数通常依靠经验进行选取的问题,使用了粒子群算法(PSO)进行了相应的优化处理,最后通过实验进行了分析,表明使用PSO进行参数优化的LS-SVM进行驾驶疲劳的识别,在准确性和可靠性上都取得较好的效果,并且在实际应用中具有较好的效果。

[0003] 支持向量机(Support Vector Machine, SVM)理论作为一种新型的机器学习方法,由Vapnik等人依据统计学理论中的VC维理论以及结构风险最小化原理提出来的,它是一种监督式的学习方法,被广泛用于对统计分类以及回归分析问题中,它在解决样本数较小、非线性以及高维模式的识别问题中有着特有的优势。它的主要思想如下:

[0004] (1) 它专门是针对于样本有限的情况下使用的机器学习方法,在实现结构风险最小化的原则下,即对于给定的数据通过在逼近的精度以及逼近函数的复杂性之间寻找一个比较好的折衷,以期望获得最好的推广能力。

[0005] (2) 它最终是通过将要解决的问题转化为一个凸二次规划的问题进行求解,最后得到它的全局最优解。

[0006] (3) 它将实际问题通过非线性变换转换到高维特征空间,通过在高维空间中构造线性决策函数以实现原空间中的非线性决策函数,这能够巧妙的解决维数问题,从而能够保证具有较好的推广能力。

发明内容

[0007] 本发明提供一种基于脑电信号的驾驶疲劳检测的方法,通过对不同状态下的脑电信号进行相应分析后,选取了Kc复杂度、样本熵以及相关节律波的平均功率的比值等多个特征值构成了相应的特征向量空间,同时针对支持向量机需要利用二次规划进行计算的问

题,提出了使用LS-SVM的方法进行模型的训练,提高了模型训练的速度同时针对LS-SVM参数通常依靠经验进行选取的问题,使用了粒子群算法(PSO)进行了相应的优化处理,最后通过实验进行了分析,表明使用PSO进行参数优化的LS-SVM进行驾驶疲劳的识别,在准确性和可靠性上都取得较好的效果。

[0008] 为了实现上述目的,本发明技术方案如下。

[0009] 一种基于脑电信号的驾驶疲劳的检测方法,具体包括以下方面:

[0010] (1) 选取与疲劳驾驶状态相关的特征量:特征选择是指通过对原始测量信号进行一系列相应处理后,得到一个由一组相关特征量组成的特征集,然后从该特征集中选取某些具有代表性、分类性能最佳的物理量作为区分不同行为的一组特征量,然后使用相关的分类识别方法,利用选取的特征量将不同的行为从特征空间中分离出来,因此对于学习分类而言都需要训练集样本:

$$[0011] \quad T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l \quad (1)$$

[0012] 其中 $x_i \in X = R^n, y_i \in Y = R, i = 1, \dots, l$;

[0013] 对于符合要求的训练集就是如何选择满足要求的特征量,具体是:

[0014] (a) 需要选取适当数量的能够描述不同行为模式的样本作为训练集,但是又不能选取过多的训练集,因为采用过多的训练集会导致训练时间的增长;

[0015] (b) 在输入样本中,代表每一样本的行为特征量需要尽量地能描述出该行为,因此选取的特征量不能太少;

[0016] (c) 为了避免复杂运算,用来描述输入样本的特征量不能什么都包括,以致于训练后的模型过于敏感;

[0017] (2) 疲劳驾驶状态的脑电信号功率谱分析:对处于不同状态下脑电信号的功率谱特性进行相关分析,将采集的脑电信号的长度选择为3秒,即对采集到的数据以每3秒截取一次并进行平均功率的计算,选取了从精力充沛状态过度到疲劳状态的脑电数据8分钟,将其截取为160段,并对截取到的这些时间段的脑电信号进行傅里叶变换:

$$[0018] \quad G = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_T^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{2T} x_T^2(f) df \quad (2)$$

[0019] 式中, f 为连续频率变量; $x_T(f)$ 为信号的频谱密度;

[0020] 其中:

$$[0021] \quad p_X(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} x_T^2(f) \quad (3)$$

[0022] 在式(3)中显示了处于不同频段上的信号功率的分布情况,同时也可以表示单位时间内和单位频段上的信号的能量,因此该积分函数又被称作为功率谱密度函数;经过傅里叶变换以后,得到 $x_n(e^{jf})$, 然后对频谱的模进行平方,最后得到功率谱密度的计算公式:

$$[0023] \quad p_X(f) = \frac{1}{N} |X_N(f)|^2 \quad (4)$$

[0024] 由上面可知信号在某个频段 h 的平均功率密度为:

$$[0025] \quad G(h) = \frac{\int_{f_l}^{f_h} p_X(f) df}{f_h - f_l} \quad (5)$$

[0026] 其中： f_h 为频段h的上限， f_l 为频段h的下限， $p_x(f)$ 为信号的功率谱密度；

[0027] 则不同频段的平均功率比值则定义为：

$$[0028] \quad F = \frac{G(h_1)}{G(h_2)} \quad (6)$$

[0029] 式中 h_1 和 h_2 为脑电信号的不同频段，将 $F_{(\alpha+\theta)/\beta}$ ， $F_{\alpha/\beta}$ 这两个相关节律波的平均功率的比值选作进行驾驶疲劳判别的两个特征指标；

[0030] (3) 疲劳驾驶状态的脑电信号样本熵分析：利用样本熵对脑电信号进行分析的具体步骤如下：计算选取长度为1000的数据作为一个数据窗口，以500个数据为移动步长，按以上步骤计算每次移动的数据的样本熵值，然后进行下1个数据窗口的样本熵值的计算，取2个数据窗口的样本熵的平均值作为时间序列的脑电数据样本熵值，通过对处于疲劳与清醒状态的脑电数据进行分析，得到不同状态下脑电信号的样本熵值；非疲劳状态时的样本熵值大于疲劳状态下的样本熵值，表明样本熵可以用于区分不同的驾驶状态，进一步可以得出疲劳状态下脑电信号计算得到样本熵值主要集中在0.2~0.4，而清醒状态的样本熵值集中在0.5~0.8之间；由此可知随着驾驶疲劳程度的加深脑电信号的样本熵值逐渐降低的，因此可看出脑电信号的样本熵值与人体精神状态之间存在相互对应关系，因此本发明将其选作为用于驾驶疲劳识别的相关特征量；

[0031] (4) 疲劳驾驶状态的脑电信号Kc复杂度分析：首先采集得到的一段待分析的脑电时间序列，设为 $S(S_1, S_2, \dots, S_n)$ ，第一步是对该序列进行粗粒化处理，具体就是先求取该序列的平均值，然后对原序列的每个元素进行判断并重新赋值，如果序列元素 S_i 大于平均值时则赋值 $S_i' = 1$ ，如果序列元素 S_i 小于平均值时则赋值 $S_i' = 0$ ，然后得到重构序列 S' (S_1' , S_2' , ..., S_n')，重新利用 S_i' 重新对S中的元素进行赋值，则得到新的序列 $S(S_1, S_2, \dots, S_n)$ ，该新序列是一串由(0,1)字符组成的新序列；下一步则对重构序列进行子串划分与界定，将按照如下的规则进行：

[0032] 在一个字符串 $S(S_1, S_2, \dots, S_n)$ 后再加一个字符串 $Q(q_1, q_2, \dots, q_n)$ ，这样就得到组合后新的字符串序列SQ，令SQ序列删去末尾的一个字符所得字符串命名为 SQ_v ，接下来需要对Q是否属于 SQ_v 的子串进行判断，如果属于，则把下一个字符加到Q后面，增长一位再进行重复判断，直到增长后的Q不再属于 SQ_v 的一个子串，则用标记符号“.”进行序列的前后界定；下一步将“.”前的所有字符定义为新的S，重新以其后的字符开始构造Q，重复以上的判断和界定过程，直到整个字符串的序列判断结束；统计最终由“.”界定划分出的S的子串数目就是序列的Kc复杂度；

[0033] 由于Kc复杂度的计算对数据的长度没有特定的要求，但是数据长度过长将影响计算的效率，而数据长度过短将不具有代表性，通过多次实验，最后本发明选择数据长度为1000，得到的不同状态下脑电信号的Kc复杂度值结果；处于不同驾驶状态的Kc复杂度值存在着一定的差异性，其中非疲劳驾驶状态的Kc复杂度值主要集中在0.15~0.23，而处于疲劳驾驶状态得到的Kc复杂度值主要集中在0.08~0.14左右，说明受试者在非疲劳状态下进行驾驶操作的大脑复杂性高于疲劳状态下进行驾驶操作的大脑复杂性，因此表明Kc复杂度可以作为驾驶疲劳检测的另一个指标。

[0034] (5) 支持向量机：考虑样本 $\{(x_i, y_i)\}_{i=1}^N$ ，其中 x_i 代表第i个输入样本， y_i 则是对应输入的输出；首先假设由子集 $y_i = +1$ 代表的类和 $y_i = -1$ 代表的类是能够进行线性可分的，那么用

于进行分离的超平面形式的决策曲面方程可以表示为如下：

[0035] $w^T x + b = 0$ (7)

[0036] x ——输入变量；

[0037] w ——可调的权值向量；

[0038] b ——偏值；

[0039] 那么式(1)则可以写成：

[0040] $w^T x + b \geq 0$, 当 $y_i = +1$ (8)

[0041] $w^T x + b \leq 0$, 当 $y_i = -1$ (9)

[0042] 由式(8)和(9)则可合并为：

[0043] $y_i (w^T x + b) \geq 0$ (10)

[0044] 对于一个给定的权值向量 w 和偏值 b , 数据点到(1)定义的超平面的最近间隔称为分离边缘, 用 ρ 表示, 而支持向量机的目标则是寻找一个使得分离边缘 ρ 最大的超平面, 而这个决策曲面则被称为最优超平面；

[0045] 假设 w_0 和 b_0 分别为得到的权值向量和偏值的最优值；那么此时, 输入空间里表示的多维线性决策面的最优超平面则可以表示如下：

[0046] $w_0^T x + b_0 = 0$ (11)

[0047] 假如 x_k 对应于 y_k , 当满足 $y_k (w_0^T x_k + b_0) = 1$ 时, 那么则称 (x_k, y_k) 是支持向量的, 那么此时支持向量到最优超平面的距离为：

[0048]
$$r = \begin{cases} \frac{1}{\|w_0\|} & y_k = +1 \\ -\frac{1}{\|w_0\|} & y_k = -1 \end{cases} \quad (12)$$

[0049] 其中的加号表示 x_k 在最优超平面的正面, 负号则表示最优超平面的负面, 式(11)可以得到分类间隔 r ：

[0050] $\rho = 2r = \frac{2}{\|w_0\|}$ (13)

[0051] 由上式得到的最优解则对应于分类间隔最大的情况, 其等价于凸二次最优化问题：

[0052]
$$\begin{cases} \min \frac{1}{2} w^T \bullet w \\ \text{s. t. : } y_i (w^T x_i + b) \geq 1, i = 1, \dots, m \end{cases} \quad (14)$$

[0053] 由上式得到的最优解则是由拉格朗日 (Lagrange) 函数的鞍点决定：

[0054] $L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N a_i [y_i (w^T x_i + b) - 1]$ (15)

[0055] 式中,

[0056] a_i, a_j ——拉格朗日乘子；

[0057] 由最优解条件 (KKT 条件) 可得：

$$[0058] \left. \begin{aligned} \frac{\partial L(w, b, a)}{\partial b} = 0 &\Rightarrow \sum_{i=1}^m a_i y_i = 0 \\ \frac{\partial L(w, b, a)}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^m a_i x_i y_i \end{aligned} \right\} \quad (16)$$

[0059] 根据KKT (Karush-Kuhn-Tucker) 条件, 相应的拉格朗日乘子 a_i 应满足如下条件:

$$[0060] a_i [y_i (w^T x_i + b) - 1] = 0 \quad (17)$$

[0061] 从上式可以看出只有支持向量的 a_i 不为零;

[0062] 根据核函数变换, 则式 (14) 和 (16) 可转化为对偶问题求解, 即:

$$[0063] \left\{ \begin{aligned} W(a) &= \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j x_i^T x_j \\ s.t. \sum_{i=1}^N y_i a_i &= 0 \quad a_i \geq 0, i=1, \dots, N \end{aligned} \right. \quad (18)$$

[0064] 然后由 (16) 可以求出最优权值向量 w_0 , 以及利用式子 (11) 可以得到 b_0 , 于是关于待测样本 x , 通过只算 $\text{sgn}[(w^T x) + b]$ 就可以判断其所属的类别;

[0065] 但是实际应用当中的大部分问题都是属于线性不可分的情况, 而造成这样的原因也是各种各样的, 可能在样本的产生时造成一定误差, 也可能在进行统计记录, 标记的错误导致样本数据的误差等, 这样也可能使得样本出现线性不可分的情况; 为了处理当存在少量样本无法用线性方法分开的情况, 因此在分类超平面的定义中引入松弛变量 $\xi_i \geq 0$, 即允许少量误差的存在, 那么 (14) 的约束条件变为如下形式:

$$[0066] y_i (w^T x + b) \geq 1 - \xi_i, i=1, 2, \dots, N \quad (19)$$

[0067] 此时的目标函数则变为:

$$[0068] \left\{ \begin{aligned} \min \frac{1}{2} w^T \bullet w + C \sum_{i=1}^N \xi_i \\ y_i (w^T x + b) \geq 1 - \xi_i, i=1, 2, \dots, N \end{aligned} \right. \quad (20)$$

[0069] 其中 C 为惩罚因子且 $C \geq 0$, 当选取的参数 C 比较大时, 表示支持向量机模型对训练样本的质量具有较高的信心, 相反, 当选取的参数 C 比较小时, 则认为训练样本中存在噪声; 由 (20) 得到相应的式子如下所示:

$$[0070] L(w, b, a) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N a_i [y_i (w^T x_i + b) + \xi_i - 1] - \sum_{i=1}^N \beta_i \xi_i \quad (21)$$

[0071] 那么根据KKT条件转化为其对偶问题如下:

$$[0072] W(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j x_i^T x_j \quad (22)$$

$$[0073] s.t. \sum_{i=1}^m y_i a_i = 0, 0 \leq a_i \leq C, i=1, \dots, m \quad (23)$$

[0074] 将输入空间 R^n 中的点利用非线性映射 $\Phi: R^n \rightarrow H$ 映射到高维的Hilbert空间当中去, 然后再使用线性可分的方法进行最优分类面的求解, 而在这个求解的过程中, 仅仅涉及到内积的运算, 并且计算量也只依赖于训练样本数的个数, 因此这个过程就是如何找到一个核函数 $K(x_i, x_j)$, $x_i, x_j \in R^n$, 使得该核函数满足 $K(x_i, x_j) = \Phi^T(x_i) \Phi(x_j)$, 这样就可以保证

即使在不知道变换函数 Φ 形式的情况下也能够求得最终的结果,同时能够很好地避免维数灾难的发生,因此在最大间隔情况下,非线性支持向量机的目标函数为:

$$[0075] \quad W(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(x_i, x_j) \quad (24)$$

[0076] 同样可得知,非线性支持向量机的最优化问题如下所示:

$$[0077] \quad \begin{cases} \min \frac{1}{2} w^T \bullet w + C \sum_{i=1}^N \xi_i \\ \text{s.t.}: \sum_{i=1}^m y_i a_i = 0, \quad 0 \leq a_i \leq C, \quad i=1, \dots, m \end{cases} \quad (25)$$

[0078] 它的对偶问题则为:

$$[0079] \quad \begin{cases} \max W(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(x_i, x_j) \\ \text{s.t.}: \sum_{i=1}^N y_i a_i = 0, \quad 0 \leq a_i \leq C, \quad i=1, \dots, N \end{cases} \quad (26)$$

[0080] 相应得到的分类函数为:

$$[0081] \quad f(x) = \text{sgn}[(w^T \Phi(x)) + b] = \text{sgn}[\sum_{i=1}^N y_i a_i K(x_i, x_j) + b] \quad (27)$$

[0082] (6) 最小二乘支持向量机:最小二乘支持向量机算法的描述如下:首先选择一个非线性的变换 $\varphi(\bullet)$,把n维输入和1维输出的样本向量 $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, $x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i=1, 2, \dots, N$,从原空间映射到高维特征空间,则分类LSSVM的优化问题则变为:

$$[0083] \quad \begin{cases} \min \frac{1}{2} w^T \bullet w + \frac{C}{2} \sum_{i=1}^N \xi_i^2 \\ y_i [w^T \bullet \varphi(x_i) + b] = 1 - \xi_i, \quad i=1, 2, \dots, N \end{cases} \quad (28)$$

[0084] 即将支持向量机的不等式约束改为等式约束,构造拉格朗日函数进行求解:

$$[0085] \quad L(w, b, a) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N a_i [y_i (w^T x_i + b) - 1 + \xi_i] \quad (29)$$

[0086] 其中 a_i ——拉格朗日乘子;

[0087] 最优解满足KKT条件则:进而可以得到下面的方程组:

$$[0088] \quad \begin{cases} w = \sum_{i=1}^N a_i y_i \varphi(x_i) \\ \sum_{i=1}^N a_i y_i = 0 \\ a_i = C \xi_i \\ y_i [w^T \varphi(x_i) + b] - 1 + \xi_i = 0 \end{cases} \quad (30)$$

[0089] 对于 $i=1, 2, \dots, N$,通过式子(30)消去 w 和 ξ 得到如下线性系统:

$$[0090] \quad \begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (31)$$

$$[0091] \quad \text{式中:} \begin{cases} Z = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, \dots, \varphi(x_N)^T y_N] \\ Y = [y_1, y_2, \dots, y_N]^T \\ 1 = [1, 2, \dots, N]^T \\ a = [a_1, a_2, \dots, a_N] \end{cases} \quad (32)$$

[0092] 令 $\Omega = ZZ^T + C^{-1}1$, 求解5-13可得到:

$$[0093] \quad \begin{cases} a = \Omega^{-1}(Y - b1) \\ b = \frac{1^T \Omega^{-1} Y}{1^T \Omega^{-1} 1} \end{cases} \quad (33)$$

[0094] 最终得到的决策函数如下:

$$[0095] \quad f(x) = \sum_{i=1}^N a_i K(x, x_i) + b; \quad (34)$$

[0096] (7) 粒子群优化 (PSO) 算法设置LS-SVM的模型训练参数C和g:采用最小二乘法支持向量机进行模型建立的过程来确定两个参数,即惩罚参数C和核函数参数g,不同参数对模型性能以及泛化能力有不同的影响;因此,在建模的时候,选取适当的参数能够极大地改善识别的准确度,接下来具体研究LS-SVM的参数影响:惩罚参数C可以理解为对超出样本数据的一个管制制度,它直接影响模型的复杂程度和稳定性,惩罚参数C能够使模型在复杂度和训练误差之间做出一个平衡,使模型的推广能力得到保证,若惩罚参数C过小,则对超出的样本数据惩罚就小,训练的误差就会变大,模型的推广能力就会变差,当使用模型对新的数据进行学习时,错分率就会很高,而当惩罚参数C过大时,训练学习精度相应提高,但会对训练样本产生过度的依赖,对新数据的认识度不够,导致错分率增加,也会影响模型的推广能力,此外,惩罚参数C值也会对算法处理样本中的离群点(数据集合中一些元素和其他元素很不相似或不一致)产生影响,选取合适的值就能在一定程度上抗干扰,使模型稳定;

[0097] 核函数参数g对支持向量的性能有很大的影响,对于高斯核中的核宽度 δ 而言,它反映了支持向量之间的相关程度, δ 过小,支持向量之间的联系比较松弛,机器学习相对复杂,推广能力得不到保证, δ 过大,支持向量之间的影响过强,预测模型很难达到足够的精度,所以在选取参数时,应综合考虑他们之间的相互影响;

[0098] 因此,只有恰当的选择LS-SVM的参数才能得到最小的经验误差,同时提高模型的训练速度以及模型的泛化能力,使得预测结果更为准确;其中传统的模型参数选取主要通过交叉验证法以及经验选择方法等,但是使用经验选择法通常需要获知该重逢的经验知识,在通过判断进行参数的设置,要求比较高,因此一般不采用,而交叉验证法较为普遍,是大多数所采用的方法,便于实现,步骤简单,但是通过这种方法获得的参数精确度一般,并且需要花费较长的时间,本发明采用智能优化算法中的粒子群算法进行对模型参数的寻优,使其获得较好的性能,粒子群算法具体如下:首先在可解空间内初始化一群粒子,而每一个粒子都同时拥有位置、速度和适应度值三个特征,而其中的适应度值的大小则决定该粒子的优劣,同时粒子的移动主要是通过跟踪个体极值Pbest、群体极值Gbest以及过去速度的惯性来更新个体位置,其中个体极值Pbest是指单个粒子所经历的适应度值最优的位置,群体极值Gbest是指种群中所有粒子所经历适应度值最优的位置,粒子每迭代一次就移动一次位置,其适应度值就相应的需要重新计算一次,然后通过对新粒子个体极值的适应

度值、群体极值的适应度值与迭代之前的极值比较来确定是否更新粒子的个体极值和群体极值的位置；

[0099] 假设在一个D维的空间中,其中 $X = (X_1, X_2, \dots, X_n)$ 为由n个粒子组成的种群,其中向量 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$ 表示第i个粒子,对于第i个粒子,它的适应度值一般就是根据粒子 X_i 在目标函数中的值,速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$,其个体极值为 $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})^T$,种群的全局极值为 $P_g = (p_{g1}, p_{g2}, \dots, p_{gD})^T$

[0100] 在每一次的迭代过程中,每个粒子都是通过个体极值、全局极值以及惯性速度来更新它的速度和位置信息,其更新公式(35)和(36)如下所示:

$$[0101] \quad v_{id}^{k+1} = wv_{id}^k + c_1r_1(p_{id}^k - x_{id}^k) + c_2r_2(p_{gd}^k - x_{id}^k) \quad (35)$$

$$[0102] \quad x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (36)$$

[0103] 式(35)中,w为惯性权重,当w较大时有利于群体在较大的范围内进行搜索,从而避免陷入寻找局部最优,而较小的w能够保证算法能够快速收敛到最优值; $d=1,2,\dots,D$, $i=1,2,\dots,n$,k为当前迭代次数; v_{id} 为粒子的速度; c_1 和 c_2 为非负常数,称为加速度因子,是每一个粒子在 P_i 和 P_g 位置的随机加速权重, r_1 和 r_2 是 $[0,1]$ 之间的随机数;为了防止粒子群的盲目搜索,一般把粒子群的位置限定在 $[-X_{max}, X_{max}]$ 和速度限定在 $[-V_{max}, V_{max}]$ 的范围内;

[0104] 根据上面的分析,本发明使用粒子群算法优化参数C和g,具体步骤如下:

[0105] (a) 初始化粒子群并随机产生所有粒子的位置和速度并确定粒子的 P_i 和 P_g ;

[0106] (b) 将每个粒子当前最优位置 P_i 及种群中所有粒子所经历的最好位置 P_g 进行比较,如果这个粒子的位置最优,则将其设置为当前的最好位置 P_g ,否则, P_g 保持不变;

[0107] (c) 更新粒子的速度、位置;

[0108] (d) 若没有达到结束条件,为:预设的运算精度、迭代次数,便返回步骤(b);

[0109] (e) 开始下一轮的迭代计算,否则取当前的 P_g 为最优解。

[0110] 该发明的有益效果在于:本发明通过对脑电信号从功率谱角度,非线性动力学的相关方法等多个角度进行了其与疲劳状态的相关性,最后从中选取了多个特征值构成了采用LS-SVM进行驾驶疲劳识别的向量空间,解决了仅仅从单一角度进行分析带来的识别准确度的不高的问题,同时使用粒子群算法对最小二乘向量机参数进行了优化,解决了依靠经验选取得到参数导致的准确度精度偏低的情况,最后通过实验进行了分析,表明使用PSO进行参数优化的LS-SVM进行驾驶疲劳的识别,在准确性和可靠性上都取得较好的效果。

附图说明

[0111] 图1为本发明实施例中所采用的PSO优化算法流程图。

具体实施例

[0112] 上面结合附图和具体实施例进一步阐述本发明。

[0113] (1) 选取与疲劳驾驶状态相关的特征量:特征选择是指通过对原始测量信号进行一系列相应处理后,得到一个由一组相关特征量组成的特征集,然后从该特征集中选取某些具有代表性、分类性能最佳的物理量作为区分不同行为的一组特征量,然后使用相关的分类识别方法,利用选取的特征量将不同的行为从特征空间中分离出来,因此对于学习分

类而言都需要训练集样本：

$$[0114] \quad T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l \quad (1)$$

[0115] 其中 $x_i \in X = \mathbb{R}^n, y_i \in Y = \mathbb{R}, i = 1, \dots, l$;

[0116] 对于符合要求的训练集就是如何选择满足要求的特征量，具体是：

[0117] (a) 需要选取适当数量的能够描述不同行为模式的样本作为训练集，但是又不能选取过多的训练集，因为采用过多的训练集会导致训练时间的增长；

[0118] (b) 在输入样本中，代表每一样本的行为特征量需要尽量地能描述出该行为，因此选取的特征量不能太少；

[0119] (c) 为了避免复杂运算，用来描述输入样本的特征量不能什么都包括，以致于训练后的模型过于敏感；

[0120] (2) 疲劳驾驶状态的脑电信号功率谱分析：对处于不同状态下脑电信号的功率谱特性进行相关分析，将采集的脑电信号的长度选择为3秒，即对采集到的数据以每3秒截取一次并进行平均功率的计算，选取了从精力充沛状态过度到疲劳状态的脑电数据8分钟，将其截取为160段，并对截取到的这些时间段的脑电信号进行傅里叶变换：

$$[0121] \quad G = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_T^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{2T} x_T^2(f) df \quad (2)$$

[0122] 式中， f 为连续频率变量； $x_T(f)$ 为信号的频谱密度；

[0123] 其中：

$$[0124] \quad p_X(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} x_T^2(f) \quad (3)$$

[0125] 在式(3)中显示了处于不同频段上的信号功率的分布情况，同时也可以表示单位时间内和单位频段上的信号的能量，因此该积分函数又被称作为功率谱密度函数；经过傅里叶变换以后，得到 $x_n(e^{jf})$ ，然后对频谱的模进行平方，最后得到功率谱密度的计算公式：

$$[0126] \quad p_X(f) = \frac{1}{N} |X_N(f)|^2 \quad (4)$$

[0127] 由上面可知信号在某个频段 h 的平均功率密度为：

$$[0128] \quad G(h) = \frac{\int_{f_l}^{f_h} p_X(f) df}{f_h - f_l} \quad (5)$$

[0129] 其中： f_h 为频段 h 的上限， f_l 为频段 h 的下限， $p_X(f)$ 为信号的功率谱密度；

[0130] 则不同频段的平均功率比值则定义为：

$$[0131] \quad F = \frac{G(h_1)}{G(h_2)} \quad (6)$$

[0132] 式中 h_1 和 h_2 为脑电信号的不同频带，将 $F_{(\alpha+\theta)/\beta}$ ， $F_{\alpha/\beta}$ 这两个相关节律波的平均功率的比值选作进行驾驶疲劳判别的两个特征指标；

[0133] (3) 疲劳驾驶状态的脑电信号样本熵分析：利用样本熵对脑电信号进行分析的具体步骤如下：计算选取长度为1000的数据作为一个数据窗口，以500个数据为移动步长，按以上步骤计算每次移动的数据的样本熵值，然后进行下1个数据窗口的样本熵值的计算，取2个数据窗口的样本熵的平均值作为时间序列的脑电数据样本熵值，通过对处于疲劳与清

醒状态的脑电数据进行分析,得到不同状态下脑电信号的样本熵值;非疲劳状态时的样本熵值大于疲劳状态下的样本熵值,表明样本熵可以用于区分不同的驾驶状态,进一步可以得出疲劳状态下脑电信号计算得到样本熵值主要集中在0.2~0.4,而清醒状态的样本熵值集中在0.5~0.8之间;由此可知随着驾驶疲劳程度的加深脑电信号的样本熵值逐渐降低的,因此可看出脑电信号的样本熵值与人体精神状态之间存在相互对应关系,因此本发明将其选作为用于驾驶疲劳识别的相关特征量;

[0134] (4) 疲劳驾驶状态的脑电信号Kc复杂度分析:首先采集得到的一段待分析的脑电时间序列,设为 $S(S_1, S_2, \dots, S_n)$,第一步是对该序列进行粗粒化处理,具体就是先求取该序列的平均值,然后对原序列的每个元素进行判断并重新赋值,如果序列元素 S_i 大于平均值时则赋值 $S'_i = 1$,如果序列元素 S_i 小于平均值时则赋值 $S'_i = 0$,然后得到重构序列 S' (S'_1, S'_2, \dots, S'_n),重新利用 S'_i 重新对 S 中的元素进行赋值,则得到新的序列 $S(S_1, S_2, \dots, S_n)$,该新序列是一串由(0,1)字符组成的新序列;下一步则对重构序列进行子串划分与界定,将按照如下的规则进行:

[0135] 在一个字符串 $S(S_1, S_2, \dots, S_n)$ 后再加一个字符串 $Q(q_1, q_2, \dots, q_n)$,这样就得到组合后新的字符串序列SQ,令SQ序列删去末尾的一个字符所得字符串命名为 SQ_v ,接下来需要对 Q 是否属于 SQ_v 的子串进行判断,如果属于,则把下一个字符加到 Q 后面,增长一位再进行重复判断,直到增长后的 Q 不再属于 SQ_v 的一个子串,则用标记符号“.”进行序列的前后界定;下一步将“.”前的所有字符定义为新的 S ,重新以其后的字符开始构造 Q ,重复以上的判断和界定过程,直到整个字符串的序列判断结束;统计最终由“.”界定划分出的 S 的子串数目就是序列的Kc复杂度。

[0136] 为了更清晰说明计算过程,举例进行说明,比如100101,则其复杂度可由下列步骤计算:

[0137] 第一个字符→1.

[0138] 2、 $S=1, Q=0, SQ=10, SQ_v=1, Q$ 不属于 $SQ_v \rightarrow 1.0.$

[0139] 3、 $S=10, Q=0, SQ=100, SQ_v=10, Q$ 属于 $SQ_v \rightarrow 1.0.0$

[0140] 4 • $S=10, Q=01, SQ=1001, SQ_v=100, Q$ 不属于 $SQ_v \rightarrow 1.0.01.$

[0141] 5. $S=1001, Q=0, SQ=10010, SQ_v=1001, Q$ 属于 $SQ_v \rightarrow 1.0.01.0$

[0142] 6. $S=1001, Q=01, SQ=100101, SQ_v=10010, Q$ 属于 $SQ_v \rightarrow 1.0.01.01$

[0143] 在第六步完成了整个字符串判断,结果中“.”界定划分出 S 的子串数目是4个,所以这个序列的复杂度 $c(n) = 4$,即 $c(100101) = 4$ 。

[0144] 由于结果并无明显的上界,不易进行比较和分析,所以Lempel和Ziv又对 S_i 的(0,1)字符串序列进行了进一步的研究,发现几乎所有的序列 $c(n)$ 都趋向于一定值:

$$[0145] \quad \lim_{n \rightarrow \infty} c(n) = b(n) = n / \log n \quad (7)$$

[0146] 所以 $c(n)$ 体现的是随机序列的渐近行为,为了能够更加直观得到利用复杂度来对序列的结果进行分析,则利用趋向的定值 $b(n)$ 对 $c(n)$ 进行归一化处理,得到新的复杂度结果即如下所示:

$$[0147] \quad C(n) = c(n) / b(n) \quad (8)$$

[0148] 由于Kc复杂度的计算对数据的长度没有特定的要求,但是数据长度过长将影响计

算的效率,而数据长度过短将不具有代表性,通过多次实验,最后本发明选择数据长度为1000,得到的不同状态下脑电信号的Kc复杂度值结果;处于不同驾驶状态的Kc复杂度值存在着一定的差异性,其中非疲劳驾驶状态的Kc复杂度值主要集中在0.15~0.23,而处于疲劳驾驶状态得到的Kc复杂度值主要集中在0.08~0.14左右,说明受试者在非疲劳状态下进行驾驶操作的大脑复杂性高于疲劳状态下进行驾驶操作的大脑复杂性,因此表明Kc复杂度可以作为驾驶疲劳检测的另一个指标。

[0149] (5) 支持向量机:考虑样本 $\{(x_i, y_i)\}_{i=1}^N$,其中 x_i 代表第 i 个输入样本, y_i 则是对应输出的输出;首先假设由子集 $y_i=+1$ 代表的类和 $y_i=-1$ 代表的类是能够进行线性可分的,那么用于进行分离的超平面形式的决策曲面方程可以表示为如下:

$$[0150] \quad w^T x + b = 0 \quad (9)$$

[0151] x ——输入变量;

[0152] w ——可调的权值向量;

[0153] b ——偏值;

[0154] 那么式(1)则可以写成:

$$[0155] \quad w^T x + b \geq 0, \text{ 当 } y_i = +1 \quad (10)$$

$$[0156] \quad w^T x + b \leq 0, \text{ 当 } y_i = -1 \quad (11)$$

[0157] 由式(10)和(11)则可合并为:

$$[0158] \quad y_i (w^T x + b) \geq 0 \quad (12)$$

[0159] 对于一个给定的权值向量 w 和偏值 b ,数据点到(1)定义的超平面的最近间隔称为分离边缘,用 ρ 表示,而支持向量机的目标则是寻找一个使得分离边缘 ρ 最大的超平面,而这个决策曲面则被称为最优超平面;

[0160] 假设 w_0 和 b_0 分别为得到的权值向量和偏值的最优值;那么此时,输入空间里表示的多维线性决策面的最优超平面则可以表示如下:

$$[0161] \quad w_0^T x + b_0 = 0 \quad (13)$$

[0162] 假如 x_k 对应于 y_k ,当满足 $y_k (w_0^T x_k + b_0) = 1$ 时,那么则称 (x_k, y_k) 是支持向量的,那么此时支持向量到最优超平面的距离为:

$$[0163] \quad r = \begin{cases} \frac{1}{\|w_0\|} & y_k = +1 \\ -\frac{1}{\|w_0\|} & y_k = -1 \end{cases} \quad (14)$$

[0164] 其中的加号表示 x_k 在最优超平面的正面,负号则表示最优超平面的负面,式(13)可以得到分类间隔 r :

$$[0165] \quad \rho = 2r = \frac{2}{\|w_0\|} \quad (15)$$

[0166] 由上式得到的最优解则对应于分类间隔最大的情况,其等价于凸二次最优化问题:

$$[0167] \quad \begin{cases} \min \frac{1}{2} w^T \bullet w \\ \text{s. t. : } y_i(w^T x_i + b) \geq 1, i=1, \dots, m \end{cases} \quad (16)$$

[0168] 由上式得到的最优解则是由拉格朗日 (Lagrange) 函数的鞍点决定:

$$[0169] \quad L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m a_i [y_i(w^T x_i + b) - 1] \quad (17)$$

[0170] 式中,

[0171] a_i, a_j ——拉格朗日乘子;

[0172] 由最优解条件 (KKT条件) 可得:

$$[0173] \quad \left. \begin{aligned} \frac{\partial L(w, b, a)}{\partial b} = 0 &\Rightarrow \sum_{i=1}^m a_i y_i = 0 \\ \frac{\partial L(w, b, a)}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^m a_i x_i y_i \end{aligned} \right\} \quad (18)$$

[0174] 根据KKT (Karush-Kuhn-Tucker) 条件, 相应的拉格朗日乘子 a_i 应满足如下条件:

$$[0175] \quad a_i [y_i (w^T x_i + b) - 1] = 0 \quad (19)$$

[0176] 从上式可以看出只有支持向量的 a_i 不为零;

[0177] 根据核函数变换, 则式 (16) 和 (18) 可转化为对偶问题求解, 即:

$$[0178] \quad \begin{cases} W(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j x_i^T x_j \\ \text{s.t. } \sum_{i=1}^m y_i a_i = 0 \quad a_i \geq 0, i=1, \dots, N \end{cases} \quad (20)$$

[0179] 然后由 (18) 可以求出最优权值向量 w_0 , 以及利用式子 (13) 可以得到 b_0 , 于是关于待测样本 x , 通过只算 $\text{sgn}[(w^T x) + b]$ 就可以判断其所属的类别;

[0180] 但是实际应用当中的大部分问题都是属于线性不可分的情况, 而造成这样的原因也是各种各样的, 可能在样本的产生时造成一定误差, 也可能在进行统计记录, 标记的错误导致样本数据的误差等, 这样也可能使得样本出现线性不可分的情况; 为了处理当存在少量样本无法用线性方法分开的情况, 因此在分类超平面的定义中引入松弛变量 $\xi_i \geq 0$, 即允许少量误差的存在, 那么 (16) 的约束条件变为如下形式:

$$[0181] \quad y_i (w^T x + b) \geq 1 - \xi_i, i=1, 2, \dots, N \quad (21)$$

[0182] 此时的目标函数则变为:

$$[0183] \quad \begin{cases} \min \frac{1}{2} w^T \bullet w + C \sum_{i=1}^N \xi_i \\ y_i (w^T x + b) \geq 1 - \xi_i, i=1, 2, \dots, N \end{cases} \quad (22)$$

[0184] 其中 C 为惩罚因子且 $C \geq 0$, 当选取的参数 C 比较大时, 表示支持向量机模型对训练样本的质量具有较高的信心, 相反, 当选取的参数 C 比较小时, 则认为训练样本中存在噪声; 由 (22) 得到相应的式子如下所示:

$$[0185] \quad L(w, b, a) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N a_i [y_i (w^T x_i + b) + \xi_i - 1] - \sum_{i=1}^N \beta_i \xi_i \quad (23)$$

[0186] 那么根据KKT条件转化为其对偶问题如下：

$$[0187] \quad W(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j x_i^T x_j \quad (24)$$

$$[0188] \quad s.t. \sum_{i=1}^m y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, m \quad (25)$$

[0189] 将输入空间 R^n 中的点利用非线性映射 $\Phi: R^n \rightarrow H$ 映射到高维的Hilbert空间当中去，然后再使用线性可分的方法进行最优分类面的求解，而在这个求解的过程中，仅仅涉及到内积的运算，并且计算量也只依赖于训练样本数的个数，因此这个过程就是如何找到一个核函数 $K(x_i, x_j)$ ， $x_i, x_j \in R^n$ ，使得该核函数满足 $K(x_i, x_j) = \Phi^T(x_i) \Phi(x_j)$ ，这样就可以保证即使在不知道变换函数 Φ 形式的情况下也能够求得最终的结果，同时能够很好地避免维数灾难的发生，因此在最大间隔情况下，非线性支持向量机的目标函数为：

$$[0190] \quad W(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(x_i, x_j) \quad (26)$$

[0191] 同样可得知，非线性支持向量机的最优化问题如下所示：

$$[0192] \quad \begin{cases} \min \frac{1}{2} w^T \bullet w + C \sum_{i=1}^N \xi_i \\ s.t.: \sum_{i=1}^m y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, m \end{cases} \quad (27)$$

[0193] 它的对偶问题则为：

$$[0194] \quad \begin{cases} \max W(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j K(x_i, x_j) \\ s.t.: \sum_{i=1}^N y_i a_i = 0, 0 \leq a_i \leq C, i = 1, \dots, N \end{cases} \quad (28)$$

[0195] 相应得到的分类函数为：

$$[0196] \quad f(x) = \text{sgn}[(w^T \Phi(x)) + b] = \text{sgn}[\sum_{i=1}^N y_i a_i K(x_i, x_j) + b] \quad (29)$$

[0197] 虽然支持向量机的方法在很多应用中都能取得不错的效果，但是在SVM的求解过程中实际上是通过使用二次规划的方法来进行求解的，而在二次规划的求解过程中，当样本数目很大时，计算将耗费大量的机器内存以及运算时间。因此，在SVM的研究的基础上，提出了最小二乘支持向量机，它采用训练误差的平方代替松弛变量，同时将不等式约束变成等式约束，从而问题的求解则变成了对线性方程组的求解，从而避免了解二次规划的问题，并且能够提高求解的速度

[0198] (6) 最小二乘支持向量机：最小二乘支持向量机算法的描述如下：首先选择一个非线性的变换 $\phi(\bullet)$ ，把 n 维输入和1维输出的样本向量 $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ ， $x_i \in R^n, y_i \in R, i = 1, 2, \dots, N$ ，从原空间映射到高维特征空间，则分类LSSVM的优化问题则变为：

$$[0199] \quad \begin{cases} \min \frac{1}{2} w^T \bullet w + \frac{C}{2} \sum_{i=1}^N \xi_i^2 \\ y_i [w^T \bullet \varphi(x_i) + b] = 1 - \xi_i, \quad i=1, 2, \dots, N \end{cases} \quad (30)$$

[0200] 即将支持向量机的不等式约束改为等式约束,构造拉格朗日函数进行求解:

$$[0201] \quad L(w, b, a) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N a_i [y_i (w^T x_i + b) - 1 + \xi_i] \quad (31)$$

[0202] 其中 a_i ——拉格朗日乘子;

[0203] 最优解满足KTT条件则:进而可以得到下面的方程组:

$$[0204] \quad \begin{cases} w = \sum_{i=1}^N a_i y_i \varphi(x_i) \\ \sum_{i=1}^N a_i y_i = 0 \\ a_i = C \xi_i \\ y_i [w^T \varphi(x_i) + b] - 1 + \xi_i = 0 \end{cases} \quad (32)$$

[0205] 对于 $i=1, 2, \dots, N$,通过式子(32)消去 w 和 ξ 得到如下线性系统:

$$[0206] \quad \begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (33)$$

$$[0207] \quad \text{式中:} \begin{cases} Z = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, \dots, \varphi(x_N)^T y_N] \\ Y = [y_1, y_2, \dots, y_N]^T \\ 1 = [1, 2, \dots, N]^T \\ a = [a_1, a_2, \dots, a_N] \end{cases} \quad (34)$$

[0208] 令 $\Omega = ZZ^T + C^{-1}I$,求解可得到:

$$[0209] \quad \begin{cases} a = \Omega^{-1}(Y - b1) \\ b = \frac{1^T \Omega^{-1} Y}{1^T \Omega^{-1} 1} \end{cases} \quad (35)$$

[0210] 最终得到的决策函数如下:

$$[0211] \quad f(x) = \sum_{i=1}^N a_i K(x, x_i) + b; \quad (36)$$

[0212] 由上面可以知道,最小二乘支持向量机是支持向量机的一种改进,其中最小二乘支持向量机把复杂的非线性二次规划问题化解成线性问题来进行求解,是支持向量机在二次损失函数下的一种形式,它既发扬了支持向量机本身的优点,也发挥出自己一些独特的优点,例如计算相对简单,计算速度提高,减少计算机内存等优点,虽然LSSVM算法相比其他方式有着更强的泛化能力,但它对模型本身的内部参数有着很强的依赖性,其中参数值的选取很大程度的关系着模型的性能以及它的识别精度。

[0213] (7) 粒子群优化(PSO)算法设置LS-SVM的模型训练参数 C 和 g :采用最小二乘法支持向量机进行模型建立的过程来确定两个参数,即惩罚参数 C 和核函数参数 g ,不同参数对模型性能以及泛化能力有不同的影响;因此,在建模的时候,选取适当的参数能够极大地改善

识别的准确度,接下来具体研究LS-SVM的参数影响:惩罚参数C可以理解为对超出样本数据的一个管制制度,它直接影响模型的复杂程度和稳定性,惩罚参数C能够使模型在复杂度和训练误差之间做出一个平衡,使模型的推广能力得到保证,若惩罚参数C过小,则对超出的样本数据惩罚就小,训练的误差就会变大,模型的推广能力就会变差,当使用模型对新的数据进行学习时,错分率就会很高,而当惩罚参数C过大时,训练学习精度相应提高,但会对训练样本产生过度的依赖,对新数据的认识度不够,导致错分率增加,也会影响模型的推广能力,此外,惩罚参数C值也会对算法处理样本中的离群点(数据集合中一些元素和其他元素很不相似或不一致)产生影响,选取合适的值就能在一定程度上抗干扰,使模型稳定;

[0214] 核函数参数g对支持向量的性能有很大的影响,对于高斯核中的核宽度 δ 而言,它反映了支持向量之间的相关程度, δ 过小,支持向量之间的联系比较松弛,机器学习相对复杂,推广能力得不到保证, δ 过大,支持向量之间的影响过强,预测模型很难达到足够的精度,所以在选取参数时,应综合考虑他们之间的相互影响;

[0215] 因此,只有恰当的选择LS-SVM的参数才能得到最小的经验误差,同时提高模型的训练速度以及模型的泛化能力,使得预测结果更为准确;其中传统的模型参数选取主要通过交叉验证法以及经验选择方法等,但是使用经验选择法通常需要获知该重逢的经验知识,在通过判断进行参数的设置,要求比较高,因此一般不采用,而交叉验证法较为普遍,是大多数所采用的方法,便于实现,步骤简单,但是通过这种方法获得的参数精确度一般,并且需要花费较长的时间,本发明采用智能优化算法中的粒子群算法进行对模型参数的寻优,使其获得较好的性能,粒子群算法具体如下:首先在可解空间内初始化一群粒子,而每一个粒子都同时拥有位置、速度和适应度值三个特征,而其中的适应度值的大小则决定该粒子的优劣,同时粒子的移动主要是通过跟踪个体极值Pbest、群体极值Gbest以及过去速度的惯性来更新个体位置,其中个体极值Pbest是指单个粒子所经历的适应度值最优的位置,群体极值Gbest是指种群中所有粒子所经历适应度值最优的位置,粒子每迭代一次就移动一次位置,其适应度值就相应的需要重新计算一次,然后通过对新粒子个体极值的适应度值、群体极值的适应度值与迭代之前的极值比较来确定是否更新粒子的个体极值和群体极值的位置;

[0216] 假设在一个D维的空间中,其中 $X = (X_1, X_2, \dots, X_n)$ 为由n个粒子组成的种群,其中向量 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$ 表示第i个粒子,对于第i个粒子,它的适应度值一般就是根据粒子 X_i 在目标函数中的值,速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$,其个体极值为 $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})^T$,种群的全局极值为 $P_g = (p_{g1}, p_{g2}, \dots, p_{gD})^T$

[0217] 在每一次的迭代过程中,每个粒子都是通过个体极值、全局极值以及惯性速度来更新它的速度和位置信息,其更新公式(37)和(38)如下所示:

$$[0218] \quad v_{id}^{k+1} = wv_{id}^k + c_1r_1(p_{id}^k - x_{id}^k) + c_2r_2(p_{gd}^k - x_{id}^k) \quad (37)$$

$$[0219] \quad x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (38)$$

[0220] 式(37)中,w为惯性权重,当w较大时有利于群体在较大的范围内进行搜索,从而避免陷入寻找局部最优,而较小的w能够保证算法能够快速收敛到最优值; $d=1, 2, \dots, D, i=1, 2, \dots, n, k$ 为当前迭代次数; v_{id} 为粒子的速度; c_1 和 c_2 为非负常数,称为加速度因子,是每一个粒子在 P_i 和 P_g 位置的随机加速权重, r_1 和 r_2 是 $[0, 1]$ 之间的随机数;为了防止粒子群的盲目搜索,一般把粒子群的位置限定在 $[-X_{max}, X_{max}]$ 和速度限定在 $[-V_{max}, V_{max}]$ 的范围内;

[0221] 根据上面的分析,本发明使用粒子群算法优化参数C和g,具体步骤如下:

[0222] (a) 初始化粒子群并随机产生所有粒子的位置和速度并确定粒子的 P_i 和 P_g ;

[0223] (b) 将每个粒子当前最优位置 P_i 及种群中所有粒子所经历的最好位置 P_g 进行比较,如果这个粒子的位置最优,则将其设置为当前的最好位置 P_g ,否则, P_g 保持不变;

[0224] (c) 更新粒子的速度、位置;

[0225] (d) 若没有达到结束条件,如:预设的运算精度、迭代次数,便返回步骤(b);

[0226] (e) 开始下一次的迭代计算,否则取当前的 P_g 为最优解。

[0227] 上面提出的最小二乘支持向量的模型训练的过程中,惩罚系数C和核函数参数g都需要由实验者依据其经验进行事先确定,然而这两个参数选择的合适与否与建模后的识别准确度息息相关。本发明分别采用交叉验证法以及采用粒子群优化的方法分别进行了相关实验,寻找出相应的参数(C,g),其中在使用PSO-LSSVM的算法中参数设置,初始化种群规模为30,最大迭代次数为200, $c_1=c_2=2$,惯性权重 $w_0.9$ 递减至0.4,初始化惩罚因子C的取值范围是[0.001,500],核函数参数g的取值范围是[0.1,1000],最大速度限制为10,由此得到的最优参数对比表如下所示:

| | 最优C | 最优g |
|------------|-------|------|
| [0228] SVM | 23.12 | 7.83 |
| PSO-LSSVM | 52.73 | 1.26 |

[0229] 选取由5名体型、特征各不相同,身体健康的志愿者进行模拟驾驶状态,分别采集处于不同状态下的脑电数据,并对采集到的脑电数据进行相应的后,采用第四章提供的特征提取方法提取不同状态下特征值,由此组成相关的特征向量 $X_i = (F_{\alpha/\beta}, F_{(\alpha+\theta)/\beta}, \text{SampEn}, Kc \text{复杂度})$,然后将实验中所取得的总体样本数据的75%作为训练样本,并对其进行分类,然后将剩余的25%作为测试样本。分别使用SVM、LSSVM以及PSO-LSSVM算法进行实验,得到对比分析结果。

[0230] 在选取相同测试集下,使用SVM进行分类的识别准确率达到89.34%,而使用LSSVM进行分类的准确率为85.47%,使用P粒子优化算法对惩罚系数C和核函数参数g进行优化后,使用PSO-LSSVM的对驾驶疲劳进行识别的准确率有了进一步的提高达到92.67%,表明,使用粒子算法对参数进行优化能够提高识别的准确率。

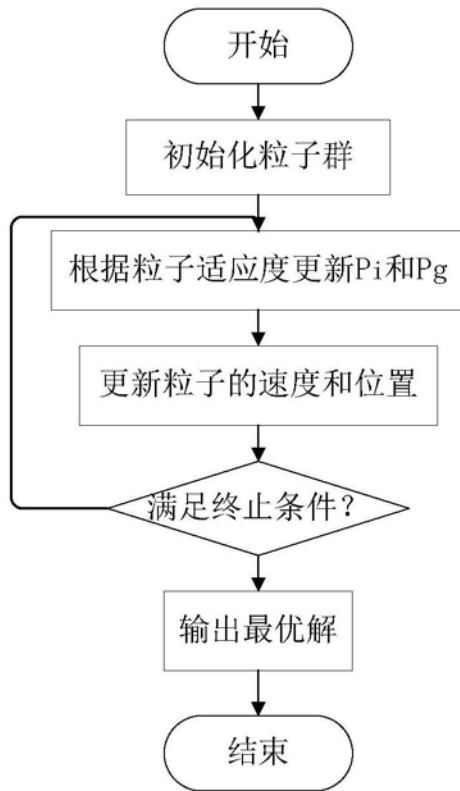


图1

| | | | |
|----------------|--|---------|------------|
| 专利名称(译) | 一种基于脑电信号的驾驶疲劳的检测方法 | | |
| 公开(公告)号 | CN106371610B | 公开(公告)日 | 2020-06-09 |
| 申请号 | CN201610846881.8 | 申请日 | 2016-09-23 |
| [标]申请(专利权)人(译) | 重庆金瓯科技发展有限公司 | | |
| 申请(专利权)人(译) | 重庆金瓯科技发展有限公司 | | |
| 当前申请(专利权)人(译) | 重庆金瓯科技发展有限公司 | | |
| [标]发明人 | 金纯 | | |
| 发明人 | 金纯 | | |
| IPC分类号 | G06F3/01 G06K9/62 A61B5/00 | | |
| CPC分类号 | A61B5/72 A61B2503/22 G06F3/015 G06F2203/011 G06K9/6269 | | |
| 审查员(译) | 李妮 | | |
| 其他公开文献 | CN106371610A | | |
| 外部链接 | Espacenet SIPO | | |

摘要(译)

本发明涉及一种基于脑电信号的驾驶疲劳的检测方法，具体包括以下方面：(1)选取与疲劳驾驶状态相关的特征量；(2)疲劳驾驶状态的脑电信号功率谱分析；(3)疲劳驾驶状态的脑电信号样本熵分析；(4)疲劳驾驶状态的脑电信号Kc复杂度分析；(5)支持向量机(SVM)(6)最小二乘支持向量机(LS-SVM)；(7)粒子群优化(PSO)算法设置LS-SVM的模型训练参数C和g。本发明通过对不同驾驶状态下提取到的脑电信号分别从功率谱的角度以及使用非线性动力学中的相关方法分别进行了研究，表明在准确性和可靠性上都取得较好的效果。

