

DOI:10.3969/j.issn.1005-202X.2023.04.008

医学信号处理与医学仪器

基于改进人工蜂群优化支持向量机的睡眠分期

熊馨¹, 吴迪¹, 张亚茹¹, 冯建楠¹, 易三莉¹, 王春武², 刘瑞湘³, 贺建峰¹

1. 昆明理工大学信息工程与自动化学院, 云南 昆明 650500; 2. 韩山师范学院物理与电子工程学院, 广东 潮州 521000; 3. 云南省第二人民医院临床心理科, 云南 昆明 650021

【摘要】本研究提出通过改进人工蜂群算法优化支持向量机(IMABC-SVM)进行睡眠分期。对提取的离散小波变换分解数据分量、时域特征、非线性特征、微状态特征,使用Relieff算法进行特征筛选,提取出最优特征矩阵,并由IMABC-SVM分类器对特征矩阵进行训练。为验证特征筛选与优化分类器效果,进行相关消融实验。结果表明IMABC-SVM方法精度可达89.97%。IMABC-SVM方法可为睡眠相关疾病的检测、预防和治疗提供有效的依据。

【关键词】睡眠分期;改进人工蜂群算法;支持向量机;Relieff;Lévy飞行

【中图分类号】R318;TN911.7

【文献标志码】A

【文章编号】1005-202X(2023)04-0440-08

Sleep staging using support vector machine optimized by improved artificial bee colony

XIONG Xin¹, WU Di¹, ZHANG Yaru¹, FENG Jiannan¹, YI Sanli¹, WANG Chunwu², LIU Ruixiang³, HE Jianfeng¹

1. College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. School of Physics and Electronic Engineering, Hanshan Normal University, Chaozhou 521000, China; 3. Department of Clinical Psychology, Yunnan Second People's Hospital, Kunming 650021, China

Abstract: A support vector machine (IMABC-SVM) optimized by an improved artificial bee colony algorithm is used to for sleep staging. The data components obtained by discrete wavelet transform, time-domain features, non-linear features and micro-state features are filtered using Relieff algorithm for obtaining the optimal feature matrix which is then trained by IMABC-SVM classifier. Some ablation experiments are conducted to verify the effectiveness of the feature selection and the optimized classifier. The experimental results show that the accuracy of IMABC-SVM reaches 89.97%. IMABC-SVM can provide a basis for the detection, prevention and treatment of sleep-related disorders.

Keywords: sleep staging; improved artificial bee colony algorithm; support vector machine; Relieff; Lévy flight

前言

睡眠是一个积极和受调节的过程,对身心健康具有重要的恢复功能。传统的睡眠分期通过多导睡眠图采集多种生理信号,由医学专家通过视觉进行手动划分,然而,这种方式不但成本高、效率低下,且不能大量普及^[1]。根据美国睡眠医学会于2007年修

订的R&K睡眠分期标准,睡眠状态分为5期:清醒期、浅眠1期、浅睡2期、深睡期、快速眼动期^[2]。睡眠分期为睡眠医学领域的研究重点,现今已有多位国内外学者对其进行研究,张泾周等^[3]从多种生物信号中提取17个特征,以BP神经网络作为分类器,准确率为79%。李谷等^[4]使用希尔伯特黄变换提取脑电(Electroencephalogram, EEG)信号多频段的能量作为特征,并使用最近邻分类器进行分类,准确率为81.7%。Güneş等^[5]基于K-means聚类的特征加权与C4.5决策树在睡眠阶段分类中的组合进行睡眠分期研究。Fraiwan等^[6]使用小波变换对EEG信号提取特征,由随机森林分类器分类,准确率为83%。Wang等^[7]利用决策树的支持向量机(Support Vector Machine, SVM)对心率数据进行睡眠分期。Long等^[8]通过心率数据的心率变异性判断睡眠情况。倪红波等^[9]根据脉率变异性特征进行睡眠分期,准确率

【收稿日期】2022-11-12

【基金项目】国家自然科学基金(82060329);云南省科技厅面上项目(202201AT070108)

【作者简介】熊馨,博士,讲师,研究方向:主动式脑机接口及模式识别、图像分析与智能辅助, E-mail: xiongxin840826@163.com

【通信作者】贺建峰,博士,教授,研究方向:医学成像仿真与图像处理分析、医疗信息融合与数据挖掘、医疗信息化管理、人工智能和计算机技术应用, E-mail: jfenghe@foxmail.com

为76%。李同庆等^[10]提取波形在每一帧数据的时长占比与平均幅值作为特征,并通过双向长短时记忆网络对特征进行训练,平均准确率为84.8%。陈玲玲等^[11]使用残差收缩网络设计高效的特征提取网络并在损失函数中基于重加权思想设计类别加权损失函数,将分期准确率提升到85.4%。Wang等^[12]采用转移学习和融合网络的深度学习算法对EEG信号进行自动睡眠分期。

脑功能信息十分复杂,单一特征无法诠释脑功能信息,不具可靠性,但复杂特征过于冗余,会导致分类效果低下^[13]。因此本研究利用多特征筛选算法筛选出最具代表性的特征集合进行睡眠分期任务,并通过改进人工蜂群优化算法来提升SVM分类器参数及性能,为探索大脑认知机制提供研究思路。

1 实验数据获取及处理

本研究数据来自科英布拉大学医院睡眠医学中心的ISRUC-Sleep公共数据集^[14],选取受试者20例,其中10例是健康人数据,另外10例是患有呼吸暂停疾病的患者睡眠数据,对每位受试者采集连续8h的睡眠数据。每个记录由来自19个通道的信号组成,本研究采用19个通道中的8个通道,包括EEG通道F3-A2、C3-A2、O1-A2、F4-A1、C4-A1、O2-A1,心电图(Electrocardiogram, ECG)通道LOC-A2、ROC-A1,采样频率均为200 Hz。由经验丰富的睡眠专家人工对每段30 s长的EEG数据进行标注。通过对数据进行每段30 s分段,总段数为18 177段,其中健康人10 737段,患者7 440段。EEG信号中的噪声是影响分期准确率的重要因素,因此采用FIR(Finite Impulse Response)数字滤波器消除50 Hz的工频干扰,过滤时使用0.3~35.0 Hz的高通滤波器^[15]。为保证原始信号相位不变,采用快速独立成分分析去除伪影。

本研究的算法流程如图1所示。首先,对原始的脑电和心电图数据进行数据切割,将每个数据段切割为30 s,并对每个数据段进行特征提取,这些特征包括离散小波变换(Discrete Wavelet Transform, DWT)、时域特征、非线性特征以及微状态特征。通过ReliefF算法对提取出的特征集合进行筛选,得到最优的特征集合。然后,通过改进的人工蜂群算法优化SVM分类器,对睡眠分期进行分类。

2 特征提取

特征提取的目的不仅是降低维数,而且能通过避免不必要或冗余的信息提取隐藏在信号中的更多有用主导信息。本研究对于数据分别进行DWT、时域特征、非线性特征以及微状态特征的提取。

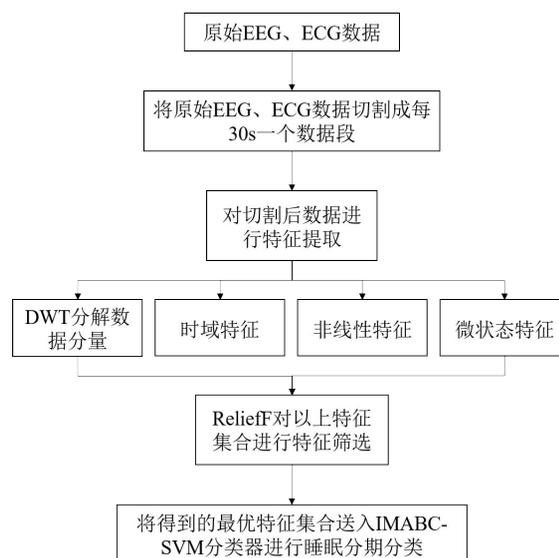


图1 算法流程图

Figure 1 Algorithm flowchart

2.1 DWT

DWT可识别多频率特性的非平稳信号,如EEG信号、ECG信号等^[16]。小波变换提供了时间域和尺度域的多分辨率分析,但连续小波变换系数冗余度较大,为在保留EEG及ECG信号信息的情况下尽量降低小波变换系数的冗余度,引入DWT。DWT具有无冗余分解以及准确重构的特点,可充分展现EEG及ECG信号的时频特征^[17]。

$$DWT(j, k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{+\infty} x(t) \phi\left(\frac{t - 2^j k}{2^j}\right) dt \quad (1)$$

其中, $x(t)$ 表示原信号; ϕ 表示小波基; j 和 k 分别表示尺度和平移参数。尺度参数 j 表示小波函数和尺度函数的缩放因子,平移参数 k 则表示小波函数和尺度函数在时间(或频率)上的平移量,可以理解为小波函数和尺度函数的相位。其中, $j = 0, 1, \dots, j - 1$; $k = 0, 1, \dots, 2^j - 1$ 。

利用DWT分解EEG以及ECG信号时,需选择合适的分解层次及小波函数。与10阶Symmlet小波(sym10)和4阶Coiflet小波(coif4)比较, Daubechies小波具有更高精度,且4阶Daubechies小波(db4)相比2阶Daubechies小波(db2)更好^[18]。因此,本研究采用db4小波对数据进行5层分解。

将EEG及ECG数据通过5层DWT分解获得的子带频率范围如下:A5分量包含 δ 频段(0~3 Hz)、D5分量包含 θ 频段(4~7 Hz)、D4分量包含 α 频段(8~15 Hz)、D3分量包含 β 频段(16~31 Hz)、D2分量包含 γ 频段(32~63 Hz)、D1分量包含高于64 Hz的频段^[19]。

2.2 时域特征

本研究选取的时域特征包括均值(Average

Value, AV)、方差值 (Variance, Var)、偏度值 (Skewness Value, Skew)、峰度值 (Kurtosis Value, Kurt)及Hjorth参数,用来衡量EEG及ECG信号的集中趋势、分散度、对称度以及峰值^[20-21]。均值、方差值、偏度值及峰度值定义如下:

$$AV = \frac{1}{N} \sum_{n=1}^N x_n \quad (2)$$

$$Var = \frac{1}{N} \sum_{n=1}^N (x_n - AM)^2 \quad (3)$$

$$Skew = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - AM)^3}{\sigma^3} \quad (4)$$

$$Kurt = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - AM)^4}{\sigma^4} \quad (5)$$

其中, x_1, x_2, \dots, x_n 是 n 个样本点的取值, AM 为样本点的平均值, σ 是样本点的标准差。

Hjorth 参数包括平均功率、平均频率和频率带宽的变化,分别代表活动度、移动度及复杂度^[22]。其定义分别为:

$$Activity = \frac{1}{N} \sum_{n=1}^N x_i^2 \quad (6)$$

$$Mobility = \sqrt{\frac{1}{N} \sum_{n=2}^N (x_n - x_{n-1})^2} \quad (7)$$

$$Complexity = \sqrt{\frac{\frac{1}{N-2} \sum_{n=3}^N (x_n - 2x_{n-1} + x_{n-2})^2}{Mobility}} \quad (8)$$

其中, x_n, x_{n-1}, x_{n-2} 分别表示信号中第 $i, i-1, i-2$ 个采样点的值, N 表示信号的采样点数。

2.3 非线性特征

由于 EEG 数据为时间序列数据,具有非线性特征。因此对 EEG 数据提取近似熵 (ApEn)、样本熵 (SampEn)、模糊熵 (FuzzyEn)。

(1) 近似熵是度量当维数变化时,时间序列中形成新模式的概率大小,使用非负数来描述某时间序列的复杂度,近似熵与时间序列的复杂程度相关^[23]。

首先将时间序列 $\{x(t), t = 1, 2, \dots, N\}$ 转换成维数为 m 的向量 $\{x(t), t = 1, 2, \dots, N\}$ 。

然后计算 X_i 与 X_j 之间的距离:

$$d_{ij} = \max |x(t+k) - x(j+k)| \quad (9)$$

其中, $k = 0, 1, \dots, m-1$ 且 $j = 1, 2, \dots, N-m+1$ 。

接着计算每个向量的距离 $d_{ij} < R (R = r * SD)$ 的数目, r 为设定阈值, SD 为原一维时间序列的标准差,且该数目与总的向量个数 $N-m+1$ 的比值为 $C_i^m(r)$:

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (10)$$

最后令 $m = m+1$, 重复上述过程, 得到 $C_i^{m+1}(r)$ 和 $\phi^m(r)$, 计算近似熵:

$$ApEn = \phi^m(r) - \phi^{m+1}(r) \quad (11)$$

(2) 样本熵是条件概率的自然对数。样本熵数据长度相关性低, 且一致性较高^[24]。

$$\phi^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \ln C_i^m(r) \quad (12)$$

令 $m = m+1$, 重复上述过程。得到样本熵:

$$SampEn = \ln \phi^m(r) - \ln \phi^{m+1}(r) \quad (13)$$

(3) 模糊熵可以看作是样本熵与模糊隶属度函数的结合, 是一种用于衡量模糊集合复杂度的指标, 它可以描述模糊集合中不确定性的程度^[25]。

引入模糊隶属函数:

$$A(x) = \begin{cases} 1 \\ \exp \left[-\ln 2 \left(\frac{x}{r} \right)^2 \right] \end{cases} \quad (14)$$

当 $x = 0$ 时, $A(x)$ 为 1; 当 $x > 0$ 时, $A(x)$ 为 $\exp \left[-\ln 2 \left(\frac{x}{r} \right)^2 \right]$ 。

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (15)$$

令 $m = m+1$, 重复上述过程。得到模糊熵:

$$FuzzyEn = \ln \phi^m(r) - \ln \phi^{m+1}(r) \quad (16)$$

2.4 微状态特征

EEG 的最大优势为毫秒级的时间分辨率, 常规特征并不能利用 EEG 这一最大优势。在亚秒级分割 EEG, 以 80~120 ms 的间隔产生稳定且均匀的模式结果, 该模式结果为微状态, 微状态分析的目的是将记录的 EEG 时间样本分割成微状态类别, 同属一类的 EEG 数据具有十分接近的 EEG 地形图。不同微状态是由脑中不同神经元的活动导致的, 而不同的微状态反映了大脑不同的功能状态。依照 Michel 等^[26] 微状态分析步骤, 对数据采集的 EEG 数据进行微状态分析并提取微状态。

基于 EEG 数据的时间属性, 从 4 类微状态中导出 3 个参数, 即平均持续时间 (ms, 表示潜在神经元结构的稳定性)、发生频率 (/s, 表示潜在的神经源被激活的趋势)、时间覆盖率 (% , 表示每个微状态相对于其他微状态的覆盖时间)。由于微状态不同特征具有不同的单位, 因此需对其进行归一化处理, 数据被标准化为 $[0, 1]$ ^[27]。

2.5 ReliefF 算法特征筛选

特征参数对于分类器效果具有十分重要的作用, 若特征参数较少时, 易出现数据重叠等问题, 导致分类器效果较差; 若特征参数较多时, 特征冗余不

仅消耗大量时间和资源,而且会增加模型复杂度。因此,特征筛选剔除贡献度较低的冗余特征,保留贡献度高的特征参数,以提升模型运行效率与效果值。特征参数选择的核心在于复杂的组合优化问题。

本研究采用 ReliefF 算法进行数据特征筛选。Kononenko^[28]于1994年提出 ReliefF 算法,该算法不仅能直接对多分类问题中的参数进行选择,搜索当前样本的各种近邻进行综合计算,并且算法运行效率高,运行结果令人满意。ReliefF 算法的核心是根据各特征和类型的相关性赋予特征不同的权重,其本质是一个矢量,而特征参数的权重是各特征的统计量指标之和,权重低于阈值的特征将被剔除。特征的权重与该特征的分类贡献度呈正相关^[29]。选取对分类贡献度高的特征组成特征参数子集即可优化选取特征^[30]。ReliefF 算法的主要过程如下:

①设特征集合设为 D , 抽样次数为 m , 特征统计量指标阈值为 δ , 最近邻样本个数 k ; 输出为各特征的统计量指标 T 。

②从 D 中随机抽取一个样本数据 E ; $H_j (j = 1, 2, \dots, k)$ 是从 E 的同类别样本集中找到的 k 个最近邻 $H_j (j = 1, 2, \dots, k)$, 再从每个类别不同的特征集中找出 k 个最近邻 $M_j(C) (j = 1, 2, \dots, k)$ 。

③计算每个特征的统计量指标,对于特征 A 的统计指标如下:

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, E, H_j)}{mk} + \sum_{C \notin \text{class}(E)} \frac{\frac{p(C)}{1 - p(\text{class}(E))} * \sum_{j=1}^k \text{diff}(A, E, H_j)}{mk} \quad (17)$$

其中, $p(C)$ 为该类别比例; $p(\text{class}(E))$ 为随机选择的某样本的类别比例。

$$\text{diff}(A, E_1, E_2) = \begin{cases} 0 & \\ \frac{|E_1[A] - E_2[A]|}{\max(A) - \min(A)} & \\ 1 & \end{cases} \quad (18)$$

若特征 A 是离散时,且 $E_1[A] = E_2[A]$ 时, $\text{diff}(A, E_1, E_2)$ 为 0; 若特征 A 是连续时, $\text{diff}(A, E_1, E_2)$ 为 $\frac{|E_1[A] - E_2[A]|}{\max(A) - \min(A)}$; 若特征 A 是离散时,并且 $E_1[A] \neq E_2[A]$ 时, $\text{diff}(A, E_1, E_2)$ 为 1。

将以上步骤进行 m 次,对各个特征的统计量指标进行计算统计,并按照从大到小的顺序对统计量进行排序,提取排序顺位较高的特征,并将提取到的特征集合转化为特征矩阵 W , 为后续分类任务做准备。

3 SVM 参数的优化

3.1 SVM

SVM 是一种经典的二分类有监督学习方法,其优点是克服过度学习和高维导致的计算复杂和局部极值,其基本的想法是将数据集不能线性分离的输入空间转换为数据集可能线性可分的高维特征空间^[31]。本研究使用带有径向基函数(RBF)的 SVM, RBF 核函数具有参数较少和非线性预测能力较好的特点,核函数见式(19):

$$K(x_i, x) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x\|^2\right) \quad (19)$$

其中, σ 是一个用户定义的参数,用于控制 RBF 核函数的宽度, $\|x_i - x\|$ 表示向量的欧几里得范数。

虽然传统的 SVM 分类器十分优秀,但对于 EEG、ECG 此类时间序列数据效果并不一定优秀,并且 SVM 的预测能力受到惩罚因子 C 和核函数 σ 的影响,惩罚系数 C 是对所有错误分类的数据惩罚程度的约束, C 越大,样本分离间隔越大,同时,也会增加泛化的风险;核函数 σ 反映了定义超平面的支持向量之间的关联,可能会导致分类器过拟合或欠拟合^[32]。因此,惩罚系数 C 和核函数 σ 的选取对 SVM 分类器至关重要^[33]。

3.2 人工蜂群算法优化 SVM (Support Vector Machine Optimized by Artificial Bee Colony, ABC-SVM)

人工蜂群算法是 Karaboga 等^[34]提出的一种群智能的全局优化算法,蜜蜂根据分工进行相应的活动,并将得到的信息送回蜂巢中进行信息比对,从众多信息中找到最优解决方式^[35]。人工蜂群算法通过模拟实际蜜蜂的蜂蜜收集机制将人工蜂群分为 3 个类别:引领蜂、跟随蜂和侦察蜂。首先,引领蜂利用先前的蜜源寻找新的蜜源并与跟随蜂分享蜜源信息;跟随蜂根据引领蜂共享的信息寻找新的蜜源;侦察蜂的使命就是寻找一个新优质蜜源,其在蜂房附近随机地寻找蜜源。算法目的为寻找蜜量最大的蜜源,其蜜源为 SVM 参数中惩罚系数 C 和核函数 σ 的最优值,该最优值可使 SVM 达到最佳效果^[36]。

该算法流程如下:首先对蜂群的种群数量、最大迭代次数、搜索范围即解的范围等进行初始化操作。设定最大迭代次数与最大搜索次数,并在搜索范围内随机产生 S_N 个蜜源,所产生的蜜源分别对应二维空间的一个二维向量,对应 SVM 的参数惩罚因子 C 和核函数 σ 。初始化后,总体样本在达到最大迭代次数或允许误差值之前,会重复对引领蜂、跟随蜂和侦察蜂的搜索过程。

每个引领蜂由式(20)产生一个新蜜

源 $v_i (i = 1, 2, \dots, S_N)$ 。

$$v_{ij} = x_{ij} + r(x_{ij} - x_{kj}) \quad (20)$$

其中, $i, k = 1, 2, \dots, S_N, j \in (1, 2)$, 且 $k \neq i$; r 为 $[-1, 1]$ 之间的随机数。

所有引领蜂进行搜寻过程后, 引领蜂会在蜂巢内招募跟随蜂, 跟随蜂根据轮盘赌法计算每个解的选择概率。

$$p_i = \frac{\text{fit}_i}{\sum_{k=1}^{S_N} \text{fit}_k} \quad (21)$$

其中, p_i 为第 i 个蜜源被选中的概率, fit_i 是第 i 个解的适应值。

当跟随蜂全部完成寻蜜过程后, 若某蜜源经过探索 limit 次后仍未被更新, 那就认为该蜜源陷入局部最优, 则蜜源将被剔除, 该蜜源对应的引领蜂变回侦查蜂。侦查蜂由式(21)产生新蜜源取代。

$$x_{ij} = x_j^{\min} + r_{ij}(x_j^{\max} - x_j^{\min}) \quad (22)$$

其中, $i, k = 1, 2, \dots, S_N, j \in (1, 2)$, x_{ij} 是第 i 个解的第 j 个维度, x_j^{\max} 和 x_j^{\min} 为第 j 个维度的上限和下限, r_{ij} 是一个 $[0, 1]$ 之间的随机数。然后返回开始步骤, 循环往复。

在食物源初始化或者每个食物源被分配给每个引领蜂后, 采用式(23)计算每个解的适应度, f_i 为第 i 个蜜源的目标函数值。

$$\text{fit}_i(t) = \begin{cases} \frac{1}{1 + f_i(t)}, & f_i(t) > 0 \\ 1 + |f_i(t)|, & f_i(t) < 0 \end{cases} \quad (23)$$

3.3 改进的人工蜂群算法优化 SVM (Support Vector Machine Optimized by Improved Artificial Bee Colony, IMABC-SVM)

3.3.1 Lévy 飞行策略 通常使用 Mantegna 方法计算 Lévy 分布, Lévy 飞行策略因其高效的全局搜索能力而被广泛采用^[37]。引入基于 Lévy 飞行的修正更新位置策略来弥补传统人工蜂群算法不能保持种群多样性, 易陷入局部最优的问题^[38]。

$$L(s) = s^{-\beta} \quad (24)$$

其中, s 为步长, β 为尺度参数, β 取值范围为 $1 \leq \beta \leq 3$ 。

在传统人工蜂群算法的邻域搜索增加 Lévy 飞行, 得到相应的 Lévy 解, 再由贪婪算法选择其中最好的解。

$$l_{ij} = v_{ij} + \alpha(v_{ij} - x_{kj})L(s) \quad (25)$$

其中, $\alpha = 0.001$, $L(s)$ 为 Lévy flight 所生成的值, β 取 1.5。

3.3.2 全局搜索因子 引领蜂是通过贪婪机制比较蜜

源的适应度值来确定之后蜜源搜索范围^[39]。根据标准的人工蜂群算法, 引领蜂搜索时只是将自身所探索的最优蜜源位置和当前蜜源位置进行比较, 此做法容易陷入局部最优的困境中。因此, 在传统人工蜂群算法的引领蜂搜索步骤中加入全局搜索因子能使当前最优蜜源位置信息加入到下一次位置更新中^[40]。

$$\text{new_}v_{ij} = x_{ij} + r(x_{mj} - x_{kj}) + \phi(x_{\text{best},j} - x_{ij}) \quad (26)$$

其中, $i, k, m = 1, 2, \dots, S_N, j \in (1, 2), k \neq m \neq i$; ϕ 为影响因子, 其范围为 $[0, 1]$; $x_{\text{best},j}$ 代表的是当前适应度值最高的蜜源。将全局搜索因子 $\phi(x_{\text{best},j} - x_{ij})$ 加入到算法中能使蜂群的搜索更加高效且迅速。

在标准的人工蜂群算法上加入 Lévy 飞行与全局搜索因子, 找到这两部分所对应的蜜源后进行比较, 保留较好的蜜源, 并继续进行标准的人工蜂群算法步骤, 以此迭代, 直至找到惩罚系数 C 和核函数 σ 对应的最优参数^[41]。之后, 将通过 ReliefF 特征筛选算法筛选得到的特征矩阵 W 送入 IMABC-SVM 分类器进行分类任务。

4 实验结果

为验证算法的高效性, 进行相应的消融实验。将实验数据划分为 4 个部分, 实验 A 进行的实验内容为仅采用 EEG 信号及其全特征矩阵; 实验 B 的实验内容为仅采用 EEG 信号并对其特征进行特征筛选; 实验 C 的实验内容为采用 EEG 与 ECG 双信号数据进行特征提取, 并利用全特征进行数据分类; 实验 D 的实验内容为采用 EEG 与 ECG 双信号数据进行特征提取, 并对提取到的特征集合进行特征筛选, 提取最优特征集合进行分类实验。算法初始各参数的设定如表 1 所示。

表 1 算法初始化参数

Table 1 Initialization parameters

参数名称	参数值
蜜源数	25
蜂群规模	25
最大限制次数	20
总迭代次数	200

为了凸显 IMABC-SVM 的算法优越性, 与最近邻分类器 (K-Nearest Neighbor, KNN)^[42]、随机森林分类器 (Random Forest, RF)^[43-44]、SVM 分类器、最小二乘 SVM 分类器 (LS-SVM)^[45]、ABC-SVM 分类器进行

分类效果对比实验。将总数据量70%数据用于训练,30%的数据用于验证模型训练效果。

根据表2的实验结果可以看出,在实验A和实验B的对照结果以及实验C和实验D的对照结果中,除了KNN分类器之外,其他分类器的运行时间均有所

缩短。这表明,特征筛选过程确实提高了数据分类效果。此外,从实验A和实验C以及实验B和实验D的结果来看,大多数基线分类器在同时使用EEG和ECG通道数据时都表现出较好的分类效果,优于仅使用EEG通道数据的分类效果。

表2 算法对比实验结果

Table 2 Experimental results of algorithm comparison

分类器	实验A		实验B		实验C		实验D	
	精度/%	运行时间/s	精度/%	运行时间/s	精度/%	运行时间/s	精度/%	运行时间/s
KNN	63.74	24.483	68.74	26.468	66.57	30.846	65.88	20.483
RF	67.39	25.157	73.66	22.210	68.48	24.811	73.30	18.157
SVM	67.12	27.106	73.27	21.684	73.43	25.119	72.34	18.106
LS-SVM	70.85	24.843	80.54	24.419	77.66	22.371	74.34	16.843
ABC-SVM	73.83	22.421	80.65	21.331	79.35	17.760	86.63	17.421
IMABC-SVM	76.92	22.773	85.35	20.202	81.95	18.368	89.97	11.773

5 讨论

人类睡眠分期的研究对于睡眠相关临床疾病的检测、预防和治疗都有着十分重要的意义。本研究采用ReliefF算法进行特征筛选,并通过改进人工蜂群优化算法来提升SVM分类器参数及性能。同时引入Lévy飞行与全局搜索因子能使标准人工蜂群算法摆脱局部最优的困境,且极大提高运行效率,通过算法改进后的分类器精度达到89.97%,相较于ABC-SVM等分类器都有所提升。IMABC-SVM分类器对于自动睡眠分期具有明显优势,与人工分期结果具有较高一致性。本研究所设计的针对SVM参数优化的多特征睡眠分期算法可以实现较为准确的睡眠分期,是一种有效的睡眠分期解决方案。

根据实验A和实验B以及实验C和实验D的对照结果,发现特征筛选过程确实提高了数据分类效果。此外,除KNN分类器外,其他分类器的运行时间均有所缩短,这表明特征筛选过程降低了数据的维度,减少了算法的运行压力,提高了运算效率。本研究还进行了两组分类实验:仅使用EEG通道数据和同时使用EEG、ECG通道数据。从实验A和实验C以及实验B和实验D的结果来看,大多数基线分类器对同时使用EEG、ECG通道数据具有较好的分类结果,这表明使用多类型数据可以帮助算法进行睡眠分期。此外,上述4组实验中,IMABC-SVM分类器的精度和运行时间在同一组别的实验结果中均为最优结果,这说明通过改进的人工蜂群算法优化的

SVM分类器(即IMABC-SVM分类器)具有卓越的性能。

人在睡眠过程中,还会产生肌电、眼电等生物电信号,这些生物电信号也可能对睡眠分期效果产生影响。因此,在未来的研究中可以结合其他生物电信号及其他生理数据来实现自动睡眠分期,以此得到更好的睡眠分期效果。另外,由于实验数据都是由专家人工标定,其本身就存在一定误差,而且睡眠的各阶段是连续的,这些因素均会对模型效果造成影响。因此,在未来的睡眠分期研究中,有必要获取更多的实验数据,并进行多次实验,以此保证本研究所提出的模型提升效果的稳定性和泛化能力。

本研究利用ReliefF算法对特征集合进行特征筛选,并通过改进人工蜂群优化算法来提升SVM分类器参数及性能,提高自动睡眠分期准确率,可为睡眠相关疾病的检测、预防和治疗提供有效的依据。

【参考文献】

- [1] 刘跃雷.一种新的人睡眠EEG自动分期方法研究[D].兰州:兰州大学,2010.
Liu YL. A new study on automatic human sleep EEG staging method [D]. Lanzhou: Lanzhou University, 2010.
- [2] 陈玲玲,毕晓君.多模态融合网络的睡眠分期研究[J].智能系统学报,2022,17(6):1194-1200.
Chen LL, Bi XJ. Sleep staging model based on multimodal fusion [J]. CAAI Transactions on Intelligent Systems, 2022, 17(6): 1194-1200.
- [3] 张泾周,周钊,滕炯华,等.基于神经网络的睡眠分期处理算法研究[J].计算机仿真,2010,27(8):141-144.
Zhang JZ, Zhou Z, Teng JH, et al. A sleep EEG segmentation process based on neural network [J]. Computer Simulation, 2010, 27(8): 141-144.

- [4] 李谷, 范影乐, 李轶, 等. 基于脑电信号 Hilbert-Huang 变换的睡眠分期研究[J]. 航天医学与医学工程, 2007, 20(6): 458-463.
Li G, Fan YL, Li Y, et al. Automatic sleep stage classification based on Hilbert-Huang transform method of EEG[J]. Space Medicine & Medical Engineering, 2007, 20(6): 458-463.
- [5] Güneş S, Polat K, Yosunkaya S, et al. A novel data pre-processing method on automatic determining of sleep stages: K-means clustering based feature weighting[J]. Complex Syst Appl ICCSA, 2009: 112-117.
- [6] Fraiwan L, Lweesy K, Khasawneh N, et al. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier[J]. Comput Meth Prog Bio, 2012, 108(1): 10-19.
- [7] Wang JS, Shih GR, Chiang WC. Sleep stage classification of sleep apnea patients using decision-tree-based support vector machines based on ECG parameters[C]//Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics. Hong Kong: IEEE, 2012: 285-288.
- [8] Long X, Fonseca P, Haakma R, et al. Automatic detection of overnight deep sleep based on heart rate variability: a preliminary study[C]//2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Chicago: IEEE, 2014: 50-53.
- [9] 倪红波, 邓军权, 施尚南, 等. 脉率变异性睡眠分期方法[J]. 浙江大学学报(工学版), 2017, 51(3): 572-576.
Ni HB, Deng JQ, Shi XN, et al. Sleep staging methods using pulse frequency variability[J]. Journal of Zhejiang University (Engineering Science), 2017, 51(3): 572-576.
- [10] 李同庆, 邹俊忠, 张见, 等. 基于周期分割的睡眠自动分期研究[J]. 计算机工程与应用, 2019, 55(9): 94-99.
Li TQ, Zou JZ, Zhang J, et al. Research of automatic staging of sleep based on period segmentation[J]. Computer Engineering and Applications, 2019, 55(9): 94-99.
- [11] 陈玲玲, 毕晓君. 基于残差收缩网络的睡眠脑电分期[J]. 仪器仪表学报, 2022, 43(2): 148-155.
Chen LL, Bi XJ. Sleep EEG staging based on the residual shrinkage network[J]. Chinese Journal of Scientific Instrument, 2022, 43(2): 148-155.
- [12] Wang H, Guo H, Zhang K, et al. Automatic sleep staging method of EEG signal based on transfer learning and fusion network[J]. Neurocomputing, 2022, 488: 183-193.
- [13] 梁明晶, 王璐, 温昕, 等. 多特征融合的脑电情绪分类[J]. 计算机工程与应用, 2023, 59(5): 155-159.
Liang MJ, Wang L, Wen X, et al. EEG emotion classification based on multi-feature fusion[J]. Computer Engineering and Applications, 2023, 59(5): 155-159.
- [14] Khalighi S, Sousa T, Santos JM, et al. ISRUC-Sleep: a comprehensive public dataset for sleep researchers[J]. Comput Meth Prog Bio, 2016, 124: 180-192.
- [15] Ahmadi N, Pei Y, Carrette E, et al. EEG-based classification of epilepsy and PNES: EEG microstate and functional brain network features[J]. Brain Inform, 2020, 7(1): 1-22.
- [16] 彭成, 王松松, 贺婧, 等. 基于离散小波变换和随机森林的轴承故障诊断研究[J]. 计算机应用研究, 2021, 38(1): 101-105.
Peng C, Wang SS, He J, et al. Research on bearing fault diagnosis based on discrete wavelet transform and random forest[J]. Application Research of Computers, 2021, 38(1): 101-105.
- [17] Subasi A. Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction[J]. Comput Biol Med, 2007, 37(2): 227-244.
- [18] 陈仁祥, 黄鑫, 杨黎霞, 等. 基于卷积神经网络和离散小波变换的滚动轴承故障诊断[J]. 振动工程学报, 2018, 31(5): 883-891.
Chen RX, Huang X, Yang LX, et al. Rolling bearing fault identification based on convolution neural network and discrete wavelet transform[J]. Journal of Vibration Engineering, 2018, 31(5): 883-891.
- [19] 袁思念, 叶继伦, 张旭, 等. 基于小波变换与人工神经网络的麻醉深度计算方法研究[J]. 生物医学工程学杂志, 2021, 38(5): 838-847.
Yuan SN, Ye JL, Zhang X, et al. An anesthesia depth computing method study based on wavelet transform and artificial neural network[J]. Journal of Biomedical Engineering, 2021, 38(5): 838-847.
- [20] Diykh M, Li Y, Wen P. EEG sleep stages classification based on time domain features and structural graph similarity[J]. IEEE Trans Neural Syst Rehabil Eng, 2016, 24(11): 1159-1168.
- [21] Şen B, Peker M, Çavuşoğlu A, et al. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms[J]. J Med Syst, 2014, 38(3): 1-21.
- [22] 刘朝宁, 黄琪, 高剑波, 等. 癫痫共病焦虑抑郁的相关因素与脑电背景活动非线性特征[J]. 中国神经精神疾病杂志, 2020, 46(11): 676-682.
Liu ZN, Huang Q, Gao JB, et al. Analysis of related factors of epilepsy comorbidity anxiety and depression and research on the nonlinear characteristics of EEG background activities[J]. Chinese Journal of Nervous and Mental Diseases, 2020, 46(11): 676-682.
- [23] Pincus SM. Approximate entropy as a measure of system complexity[J]. Proc Natl Acad Sci, 1991, 88(6): 2297-2301.
- [24] Richman JS, Lake DE, Moorman JR. Sample entropy[M]//Methods in Enzymology. United States: Academic Press, 2004, 384: 172-184.
- [25] 易芳吉, 钟丽莎, 李章勇. 基于SVM分类器的癫痫脑电时空特征提取方法的研究[J]. 重庆邮电大学学报(自然科学版), 2022, 34(3): 444-450.
Yi FJ, Zhong LS, Li ZY. Research on the spatio-temporal feature extraction of epileptic EEG based on SVM classifier[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2022, 34(3): 444-450.
- [26] Michel CM, Koenig T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review[J]. Neuroimage, 2018, 180: 577-593.
- [27] Lehmann D, Ozaki H, Pál I. EEG alpha map series: brain micro-states by space-oriented adaptive segmentation[J]. Electroencephalogr Clin Neurophysiol, 1987, 67(3): 271-288.
- [28] Kononenko I. Estimating attributes: analysis and extensions of RELIEF[C]//ECML, 1994, 94: 171-182.
- [29] 黄辉. 基于Relieff的多标签特征选择算法研究[D]. 广州: 广东工业大学, 2018.
Huang H. Research of multi-label feature selection algorithm based on Relieff[D]. Guangzhou: Guangdong University of Technology, 2018.
- [30] 张小清, 王晨曦, 吕彦, 等. 基于Relieff的层次分类在线流特征选择算法[J]. 计算机应用, 2022, 42(3): 688-694.
Zhang XQ, Wang CX, Lü Y, et al. Hierarchical classification online streaming feature selection algorithm based on Relieff algorithm[J]. Journal of Computer Applications, 2022, 42(3): 688-694.
- [31] 周鹏, 李向新, 张翼, 等. 基于主成分分析和支持向量机的睡眠分期研究[J]. 生物医学工程学杂志, 2013, 3(6): 1176-1179.
Zhou P, Li XX, Zhang Y, et al. Research on individual sleep staging based on principal component analysis and support vector machine[J]. Journal of Biomedical Engineering, 2013, 3(6): 1176-1179.
- [32] 刘靖洁, 陈桂明, 刘小方, 等. 基于遗传算法的SVM参数组合优化[J]. 计算机应用与软件, 2012, 29(4): 94-96.
Liu QJ, Chen GM, Liu XF, et al. Genetic algorithm based SVM parameter composition optimization[J]. Computer Applications and Software, 2012, 29(4): 94-96.
- [33] Aboalayon KA, Faezipour M. Multi-class SVM based on sleep stage identification using EEG signal[C]. 2014 IEEE Healthcare Innovation Conference (HIC). Seattle: IEEE, 2014: 181-184.
- [34] Karaboga D, Akay B. A comparative study of artificial bee colony algorithm[J]. Appl Math Comput, 2009, 214(1): 108-132.
- [35] Karaboga D, Basturk B. On the performance of artificial bee colony (ABC) algorithm[J]. Appl Soft Comput, 2008, 8(1): 687-697.
- [36] 朱志洁, 张宏伟, 王春明. 基于人工蜂群算法优化支持向量机的采场底板破坏深度预测[J]. 重庆大学学报, 2015, 38(6): 37-43.
Zhu ZJ, Zhang HW, Wang CM. Prediction of floor damaged depth in working area based on support vector machine and artificial bee colony algorithm[J]. Journal of Chongqing University, 2015, 38(6): 37-43.
- [37] 刘紫阳, 庞志华, 陶佩, 等. 记忆增强的莱维飞行引力搜索算法[J]. 计算机仿真, 2022, 39(1): 312-317.
Liu ZY, Pang ZH, Tao P, et al. Memory enhancement levy flight

- gravitational search algorithm[J]. Computer Simulation, 2022, 39(1): 312-317.
- [38] 马卫, 朱嫻. 基于莱维飞行扰动策略的麻雀搜索算法[J]. 应用科学学报, 2022, 40(1): 116-130.
- Ma W, Zhu X. Sparrow search algorithm based on levy flight disturbance strategy[J]. Journal of Applied Sciences, 2022, 40(1): 116-130.
- [39] 黄英双, 曹辉. 改进人工蜂群算法优化支持向量机及应用[J]. 计算机应用与软件, 2021, 38(2): 258-263.
- Huang YS, Cao H. Support vector machine optimized by improved artificial bee colony algorithm and its application [J]. Computer Applications and Software, 2021, 38(2): 258-263.
- [40] 蒋伟, 陈照光. 基于改进的人工蜂群算法的微电网储能系统容量优化配置[J]. 上海电力大学学报, 2021, 37(5): 415-421.
- Jiang W, Chen ZG. Optimization of microgrid energy storage system capacity based on improved artificial bee colony algorithm[J]. Journal of Shanghai University of Electric Power, 2021, 37(5): 415-421.
- [41] Abdullah NA, Ibrahim MA, Haider AS. Genetic algorithm as a key parameter of SVM parameter optimization and feature selection for acute Leukemia diagnosis[J]. Univ Aden J Nat Appl Sci, 2020, 24(2): 385-393.
- [42] Cover T, Hart P. Nearest neighbor pattern classification[J]. IEEE Trans Inf Theory, 1967, 13(1): 21-27.
- [43] Phan TN, Kuch V, Lehnert LW. Land cover classification using Google Earth Engine and random forest classifier-the role of image composition[J]. Remote Sens, 2020, 12(15): 2411.
- [44] Jackins V, Vimal S, Kaliappan M, et al. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes[J]. J Supercomput, 2021, 77: 5198-5219.
- [45] Suykens JA, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Process Lett, 1999, 9(3): 293-300.

(编辑:谭斯允)