

基于蚁群优选的半监督主动协同睡眠分期方法研究

金 纯^{1,2}, 岳宗田¹, 邱 灿¹

(1. 重庆邮电大学 通信与信息工程学院 无线传输重点实验室, 重庆 400065; 2. 重庆金瓯科技发展有限公司, 重庆 400041)

摘 要: 基于脑电信号(EEG)的睡眠分期对于睡眠疾病的检测、预防和治疗有着重要的意义。传统的分期方法在使用 EEG 信号进行分期时存在冗余特征、标记样本需求量高这两方面的不足,影响了该方法的应用范围。为了提高该方法的适用范围,采用一种蚁群算法与半监督学习算法结合的睡眠分期算法(ACOTSS),根据对称 KL 距离筛选低置信度样本,利用主动学习策略和协同训练进行不同置信度样本的标注以提升低标注样本下的分类正确率。采用麻省理工的公开数据集验证算法的分期效果,结果表明 ACOTSS 算法在保证分期精度的同时,比 ALKLSS 和 LS-SVM 的分期性能分别提升了 16.83% 和 8.59%,证明该算法在低标注样本下具备可行性。

关键词: 脑电信号; 睡眠分期; 蚁群算法; 半监督学习

中图分类号: TP393.04

文献标志码: A

文章编号: 1001-3695(2018)05-1403-05

doi:10.3969/j.issn.1001-3695.2018.05.026

Research on semi-supervised active learning and co-training for sleep staging algorithm based on ACO

Jin Chun^{1,2}, Yue Zongtian¹, Qiu Can¹

(1. Wireless Transmission Key Laboratory, School of Communication & Information Engineering, Chongqing University of Posts & Telecommunications, Chongqing 400065, China; 2. Chongqing Jinou Science & Technology Development Co., Ltd., Chongqing 400041, China)

Abstract: The sleep staging based on EEG is important for detection, prevention and treatment of sleep disorder. The general algorithm of sleep staging has two disadvantages on redundant features, high demand of labeled samples, and it is difficult to use this algorithm in practical application. In order to improve the application range of the algorithm, this paper proposed the ACOTSS algorithm combined with ant colony optimization(ACO) and semi-supervised learning algorithm which used the symmetry KL distance to select low confidence samples, applied active learning strategy and co-training algorithm to improve classification accuracy for the low number of labeled samples with confidence samples of different degree. It used MIT data sets to validate the ACOTSS algorithm. The result demonstrates that it improves the accuracy by 16.83% and 8.59% than ALKLSS and LS-SVM. It demonstrates the feasibility of this algorithm for the low number labeled samples.

Key words: EEG; sleep staging; ACO; semi-supervised learning

0 引言

随着现代社会城市化进程的日益加快,激烈的职场压力与快速的生活节奏导致睡眠类疾病在都市人群中的发病率逐年攀升,且已成为困扰和威胁民众健康的一个突出问题,这引起了海内外医学界的普遍关注。因此,对睡眠状况进行分期研究从而发现病因也就成为了分析和治疗睡眠疾病的前提和重要环节。目前国际上普遍使用 R&K 睡眠分期规则,根据睡眠时脑电信号(EEG)的变化规律,将睡眠分为清醒期(wake)、非快速眼动睡眠期(non-rapid eye movement, NREM)和快速眼动睡眠期(rapid eye movement, REM),其中 NREM 又可分为 I、II、III、IV 期^[1]。睡眠分期的划分对于睡眠类疾病的研究具有重要意义,而由专家进行人工分期虽然精准,但在数据量过大的情况下易发生疲劳误判现象从而不具备现实操作性。因此,设计一个高效、可行的睡眠分期算法,在实现不同睡眠期准确划分的基础上,降低人工标注的工作量,这对于大数据时代下的

脑机接口研究来说尤为重要。

1 现有睡眠分期算法及其不足

现有的睡眠分期算法的步骤包括信号预处理、特征提取和分期分类三个环节,其主要思想是通过从预处理后的 EEG 信号中进行特征参数提取,并通过采用适当的分类器对该特征向量集合进行分类,从而实现睡眠分期的目的^[2]。

目前有关 EEG 信号预处理相关的研究虽然很多,但绝大多数都将其视做单一且独立的方向进行研究,并不将其作为睡眠分期算法的一个环节来对其进行改进。而将信号预处理成为睡眠分期算法其中的一个环节时,大部分研究人员往往不对这一环节进行改进而采用目前已知且较为成熟的 EEG 预处理方法。例如在文献[3]中,作者采用了差异性系数法成功定位了眼电伪迹的产生范围,通过三层小波分解方法实现软阈值去噪。究其原因是因为对于 BCI 领域的研究人员来说,对原始信号进行预处理早已成为了目前信号去噪领域研究的一部分,而

收稿日期: 2016-12-26; 修回日期: 2017-02-23

作者简介: 金纯,男,重庆人,教授,博士,主要研究方向为蓝牙、物联网(zj@jinoux.com);岳宗田(1990-),男,硕士研究生,主要研究方向为物联网、信号处理;邱灿(1992-),男,硕士研究生,主要研究方向为物联网、室内定位。

不再作为睡眠分期方向的研究重点,睡眠分期所要关注的是如何从纯净信号中提取有效特征并将其进行有效的分期分类。因此本文不再将信号预处理作为研究重点,而对特征提取及分期分类环节进行改进。

在特征提取阶段,现有的研究方法主要分为单域特征提取和多域特征提取两种。对单域特征提取的研究较为常见,其根据某一特征域参数在不同睡眠阶段的变换规律来进行分期分类,该方法虽然容易操作,但由于 EEG 信号非稳定、非线性的特征,使得能够完美表征睡眠阶段变换规律的有效特征不应也不能通过单一特征域来获取^[4]。现有的多域特征提取方法虽然在某些方面弥补了前者的不足,但同时也可能因为特征参数过多,缺乏合理的筛选机制导致低敏感特征的混入而降低计算效率,致使“维数灾难”的发生。例如,文献[5,6]虽然是从多域特征中进行参数提取,但无一例外地忽略了特征参数自身的敏感性以及高维数特征集合所可能导致的“维数灾难”。此外,现有的研究往往缺少对特征集合进行分类效率分析、敏感特征选择这一环节,忽略了特征降维的实质,那就是凭借最少的特征参数来获取最高的分类精度。

睡眠分期中分类器的选择也是一个极其重要的环节。现有的睡眠分期研究几乎都专注于分类精度的提升,就算是采用特征降维策略来提高分类器的分类精度,往往也忽略了一个更重要的问题,那就是用于训练分类器的训练样本中标记样本的稀缺性。现有的睡眠分期研究中的分类方法往往采用监督学习算法,其好处是可以通过足够多的标记样本训练来获得分类精度较高的分类器,以保障分期分类的精度,例如文献[7~10]中,作者分别使用了最小二乘支持向量机(LS-SVM)、神经网络(NN)、K最近邻算法(KNN)以及朴素贝叶斯算法(NB)这四种监督学习算法。然而,由于睡眠 EEG 数据过于庞大使得人工标注大量样本的做法显得较为困难,标注量过低则无法获得性能较好的分类器;标注量大的话会导致人工标注成本上升,长时间作业的话也会致使标注错误的概率上升,从而降低分类精度。对此,本文从大数据研究的相关成果中获得启示,尝试将半监督学习算法引入到睡眠分期中,以实现降低标记样本需求量的目的。但是,要想将半监督学习算法应用于睡眠分期,需要解决的问题,那就是如何在缺少大量标记样本的情况下保障睡眠分期精度。

综上所述,针对目前多域特征睡眠分期算法中普遍存在的特征敏感性不一的问题,以及将半监督学习算法应用于睡眠分期所要面对的低标记样本下保障分类精度的问题,本文提出了一种基于蚁群优选的主动学习与协同训练相结合的半监督睡眠分期方法,即 ACOTSS 算法。其利用蚁群算法实现特征集合的优选子集提取,并通过协同训练和加权投票的方式实现了对半监督主动学习算法的改进,将其应用于优选子集的分类,以实现在保障睡眠分期分类精度的前提下,降低标记样本需求量。在仿真验证中,使用 ACOTSS 算法对 MIT-BIH 中所记录的真实睡眠 EEG 数据进行处理,通过与离线专家人工标注的分期结果进行对比来衡量该算法的具体性能。

2 半监督睡眠分期算法 ACOTSS

根据文献[11]的统计成果,能够表征睡眠信号阶段性变化的特征参数有数十种,分为时域、频域、时频域以及非线性特征四类。现有的多域特征提取方法过于直观,忽视了多域特征

中可能出现的冗余特征混入、敏感特征漏选的问题。因此,本文在现有多域特征提取的基础上,加入了使用蚁群算法来提前剔除冗余特征这一环节,实现了冗余特征的剔除以保留敏感特征,提高了特征子集在后续分类算法中的分类精度和计算效率。

当得到优选特征子集后,将其作为输入特征向量,通过本文所改进的半监督分类算法进行睡眠分期分类。要使得睡眠分期的半监督分类具备可行性,就需要保证其在低标记样本量下仍具备相当的分期精度,因此在半监督睡眠分期中仍不可避免地要用到人工辅助标注这一手段。文献[12]提出一种融合了主动学习策略的半监督分类算法(ALKLSS 算法),其利用对称 KL 距离筛选出低置信度样本并通过预先训练的分类器进行投票标注,在一定程度上提升了分类精度。但其也存在两个问题:a)该算法的投票分类器采用多个相同算法训练而来的分类器进行投票,这使得分类器之间缺乏差异性,造成多分类器对未标记样本的分类结果逐渐趋于一致,从而减弱了“投票”的作用;b)缺少专家辅助标注样本,使得分类精度不高。为此,本文在 ALKLSS 算法的基础上提出一种半监督主动协同分类算法,其融合了协同训练与主动学习策略,并根据 Tri-training 算法的研究成果,对其使用的多种分类器进行差异化选择,使其更加适应类似于睡眠分期样本这样的不平衡样本。

因此,本文所提 ACOTSS 算法的主要思想是:a)所选两个导联的 EEG 信号经过现有去噪算法的预处理后,通过蚁群算法测试从其中一导联数据中提取特征的分类敏感性并实现优选;b)利用优选结果保障分类精度的基础上,对另一导联数据中所提取的优选特征向量集合进行半监督分类,该过程融合了机器学习领域中的协同训练、半监督学习和主动学习策略,在保障分类精度的同时大大降低了标记样本的需求量。

综上所述,ACOTSS 算法的具体实现步骤如下:

a)将 A、B 两个导联 EEG 数据以时间 T 为长度划分为 N 个样本段,形成两个样本集合,即 $X_A = [x_1(t), x_2(t), \dots, x_N(t)]$ 和 $X_B = [x_1(t), x_2(t), \dots, x_N(t)]$,并根据预设特征提取 A 样本集合的多域特征值,得到特征值集合 $\text{set}_{L \times N}$,如式(1)所示。

$$\text{set}_{L \times N} = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1N} \\ F_{21} & F_{22} & \cdots & F_{2N} \\ \vdots & \vdots & & \vdots \\ F_{L1} & F_{L2} & \cdots & F_{LN} \end{bmatrix}_{L \times N} \quad (1)$$

其中:L 表示特征类型个数;N 表示样本段个数; F_{LN} 表示特征 L 中第 N 个样本段的值。

b)使用蚁群算法 ACO 对特征值集合 $\text{set}_{L \times N}$ 进行优选,其测试分类器采用 NB 算法,以式(2)所示的适应度值^[13]作为局部及全局解的量化函数,得到优选特征子集 set_{son} 。

$$f(k) = \frac{R(k)}{1 + \lambda \times n(k)} \quad (2)$$

其中:f 表示蚂蚁 k 生成解的适应度值;R(k) 表示该特征子集的分类正确率;n 表示此时子集的特征个数; λ 表示特征个数权重。

c)根据优选子集 set_{son} 所对应的特征,对 B 导联样本集合 X_B 进行特征提取。

d)根据式(3)和(4)所示的对称 KL 距离(SKL)方法依次对 X_B 中训练样本中的未标记样本集合 D_{UL} 进行计算,得到每一个样本的 SKL 并将该样本添加到自定的 Tlist 表中。

$$SKL(M \| N) = \frac{1}{2}KL(M \| N) + \frac{1}{2}KL(N \| M) \quad (3)$$

$$KL(M \| N) = \sum_{i=1}^s p(c_i | x) \times \ln \frac{p(c_i | x)}{p(c_j | x)} \quad (4)$$

其中: $p(c_i | x)$ 和 $p(c_j | x)$ 分别代表样本的类别后验概率, $i \neq j$, 且 $i, j \leq s, s$ 为类别数。

e) 对 Tlist 中的未标记样本按照其各自 SKL 距离进行升序排序, 选取前 ψ 个样本进行人工标注。

f) 将人工标注后的样本添加到标记样本集合 D_L 中, 同时在 D_{UL} 和 Tlist 中删去这些样本并将 Tlist 的剩余样本返回到 D_{UL} 中。

g) 完成 D_L 和 D_{UL} 的更新后, 对新的标记样本集合 D_L 进行三次 bootstrap 抽样, 产生三个子集 S_1 、 S_2 和 S_3 并将其作为训练集合, 利用不同的监督学习算法来训练三种初始分类器 C_1 、 C_2 、 C_3 。

h) 在第 t 次循环中, 对分类器 C_i ($i = 1, 2, 3$), 利用另外两个分类器 C_j 、 C_k ($j, k = 1, 2, 3$ 且 $j, k \neq i$) 对未标记样本 x 进行标记, 将满足 $\{x | x \in D_{UL} \text{ 且 } C_j(x) = C_k(x), j, k \neq i\}$ 的无标记样本加入到 S_i 中, 得到更新后的子集 $S_i = S_i + x$ ($i = 1, 2, 3$)。

i) 使用三个分类器分别对标记样本集合 D_L 进行分类, 得到各自的分类错误率 e_i ($i = 1, 2, 3$)。

j) 如果本轮错误率小于上一轮且满足式(5)所示的条件, 则利用 S_i 更新分类器 C_i ($i = 1, 2, 3$), 否则使用分层抽样法从 S_i 中抽取新的子集 S_{sub} 并用其代替 S_i 重新判断, 若满足则转至步骤 h), 否则终止迭代。

$$0 < \frac{e^t}{e^{t-1}} < \frac{\text{num}(S^{t-1})}{\text{num}(S^t)} < 1 \quad (5)$$

k) 对于测试样本集合 T_{UL} 使用如式(6)和(7)所示的加权投票法进行标注。

$$\text{label}_x = \text{label}(w_1 \times l_1 + w_2 \times l_2 + w_3 \times l_3) \quad (6)$$

$$w_i = \frac{e_i^{-1}}{e_1^{-1} + e_2^{-1} + e_3^{-1}} \quad i = 1, 2, 3 \quad (7)$$

其中: l_i 表示分类器 C_i 对该样本的标记结果; w_i 表示该分类器的分类权重; label 为标记函数, $i = 1, 2, 3$ 。

综上所述, 本文所提出的 ACOTSS 算法不但利用了蚁群算法实现了对多域特征集合的优化选取, 剔除了冗余特征、提高了计算效率, 还通过使用本文所改进的半监督学习算法, 在低标记样本情况下对大量未标记样本进行分类, 保证了其分类精度。接下来, 本文采用真实数据测试 ACOTSS 算法在不同标记样本比例下的分类精度变化情况, 并与现有的几种主流睡眠分期算法进行比较。

3 仿真实验分析

为了验证 ACOTSS 算法的有效性及其可行性, 本文使用 MIT-BIH 生理信息库中的公开测试数据 CAP sleep database 作为验证数据, 该数据统计了 108 位不同健康状况的测试者从前一天晚上 22:30 到第二天早上 7:13 将近九个小时的睡眠监测数据, 监测项目包含 10/20 系统法下的 10 导联 EEG 信号, 其采样频率为 512 Hz。此外, 该数据中还根据 R&K 标准请睡眠领域的相关专家对该数据不同睡眠阶段的变化情况进行了标注。在本实验中, 选取了一位编号为 n5 的健康测试者的 EEG 信号来作为测试数据, 通过滤波、归一化、小波去噪等预处理后得到如图 1 所示的纯净 EEG 信号, 并从中选取 2 导联

数据(C4 和 F4)分别作为特征优选验证及半监督分类数据。将此 2 导联数据分别按照每 $T = 3$ s 长度进行划分, 因此各得到了 15 520 条样本段。

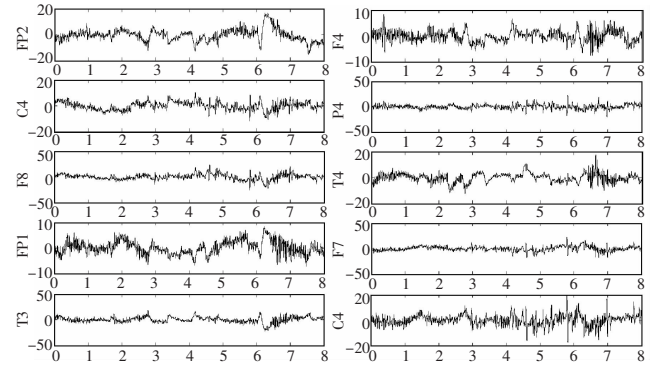


图1 预处理后的 10 导联 EEG 信号

本文使用 C4 导联 EEG 信号进行敏感特征提取, 以文献[14]的特征统计为基础, 将其中的小波时频参数换成更适于非线性信号的 S 变换系数。所以所提取的原始特征集合如下: 第 1~8 种是八个时域特征 ($X_1 \sim X_8$), 即均值、标准差、均方根、峰度、偏度、过零率、信息熵、Hjorthcomplexity 值; 第 9~14 种是六个频域特征 ($X_9 \sim X_{14}$), 对四个子节律波 (α 、 β 、 θ 、 δ 波段) 分别计算其总功率谱能量的规范化能量比, 即 $\alpha/(\delta + \theta)$ 、 $\delta/(\alpha + \theta)$ 、 $\theta/(\alpha + \delta)$ 、 $E(\alpha/\theta)$ 、 $E(\delta/\theta)$ 和 $\beta/(\delta + \theta + \alpha + \beta)$; 第 15~18 种是四个时频特征 ($X_{15} \sim X_{18}$), 即 S 变换系数的均值、标准差、方差及能量; 第 19~22 种是四个非线性特征 ($X_{19} \sim X_{22}$), 即 Sevcik 分形维数、样本熵、模糊熵及 Hurst 指数。

将此由 22 种特征参数所表示的特征向量集合通过蚁群算法来进行优化选取, 设置相关参数为: $\alpha = 0.8$, $\beta = 0$, $\tau_{ij}(0) = 1$, $\rho = 0.8$, $\varepsilon = 0.9$, $\lambda = 0.01$, $m = 15$ (参照“三步走”法^[14]中特征数量与蚂蚁数量之比为 1.5 设定), 采用原数据 (11 520 条) 的 75% 作为训练集合, 25% 作为测试集合进行测试。本实验共运行 10 次, 每次运行的迭代次数为 10, 因此得到了如表 1 所示的优选结果。

表1 ACO 特征优选结果

实验次数	分类正确率/%	适应度值	特征个数	运算时间/s
1	89.33	0.819 5	9	368.196
2	89.33	0.827 1	8	387.241
3	84.31	0.773 5	9	312.846
4	89.33	0.827 1	8	296.210
5	86.88	0.797 1	9	276.181
6	85.46	0.791 3	8	259.963
7	89.33	0.812 1	10	452.330
8	89.33	0.819 5	9	301.597
9	89.33	0.827 1	8	287.468
10	86.88	0.804 4	8	274.815
原始集合	79.25	0.649 6	22	986.996

由表 1 可知, 在本文的 10 次实验中, 有 5 次成功地将特征个数由 22 维降低到了 8 维, 压缩了 15 维; 而最差结果只有 1 次将特征维数压缩至 10 维, 与原始特征集合维数相比降低了 12 维。因此, 在这 5 次最高压缩结果中参考适应度值、运行时间等指标的情况下, 选择了第 9 次实验结果为优选特征子集 (setson = {1010000011100010010010}), 1 表示被选择, 0 表示未被选择, 即优选子集是 {均值, 均方根, S 变换系数均值, S 变换能量, $\alpha/(\delta + \theta)$, $\delta/(\alpha + \theta)$, $\theta/(\alpha + \delta)$, 模糊熵}。

在得到优选特征后, 本文使用 F4 导联信号的 11 520 条数据作为改进半监督算法的实验数据。在实验中, 选取 75% 的

数据作为训练数据,剩余 25% 的数据作为测试数据,而标记样本在训练数据中的比例分别为 15%、30%、45%、60% 及 75%。为了使算法中分类器之间的差异化更加明显,使用 NB 算法、分类决策树(C4.5)和径向基网络三个不同的分类算法,其中 Tlist 中的阈值 ψ 默认为标记样本数量的 5%。

随着训练集中标记样本比例的变化,算法所测得的分类正确率如表 2 所示,其中所测数据均为 10 次计算取平均值得出的结果。

表 2 不同标记样本比例下分类正确率

标记比例/%	wake	NREM				REM	Ac
		S_1	S_2	S_3	S_4		
15	77.86	75.68	71.98	66.31	69.33	68.56	71.62
30	80.03	80.12	75.64	88.24	72.96	76.11	78.85
45	94.63	88.41	82.37	84.15	68.32	89.54	84.57
60	94.89	90.26	84.10	84.96	74.47	90.90	86.43
75	95.55	94.73	84.19	89.75	81.77	93.17	89.86

从表 2 可以看出,该算法所测的平均正确率随着标记样本比例的增大而呈现单调稳固的增长趋势,符合分类器的一般特性。但是,该数据也存在两个异常情况:a)随着标记样本所占比例的增大,wake 期分类正确率呈现出一种近似于阶跃性的增长态势,这是由于在本实验所用数据中,wake 数据一共只采集了大约 270 s 的样本量,这使得分类算法在少量样本量的情况下表现较为敏感,容易造成突发性的上升(下降)现象的发生;b)在 NREM 阶段的 S_3 和 S_4 期中,其分类正确率的增长趋势呈折线形上升态势,这是由于在 R&K 标准中 S_3 和 S_4 期差异性并非十分明显,如果从时域角度看也仅存在幅度方面的不同,这造成了算法对这两期的样本进行分类时容易出现较大概率的错分现象,所以在 AASM 标准中干脆把 S_3 和 S_4 进行合并,以抑制这类现象的发生。

表 3 对比了文献[12]所记录的几种不同的睡眠分期方法,包括 LS-SVM、SVM、NN 和回归树这四种方法,由于算法所使用的 EEG 信号导联、提取特征以及所采用的分类方法不同,所以最终的分期效果也不同。但从表中也可以看出,本文所提出的 ACOTSS 分期算法通过使用一导联数据验证优选特征,采用另一导联数据实现半监督分期的方法是一种可行性方法。其中,前四种算法所得出的分类正确率均是在占总数据量的 75% 的训练集合均为标记样本的情况下得出的,而类似 SVM 这样的监督学习算法虽然有着分类精度高的优点,但却需要大量的标记样本进行训练,这是非常耗时耗力的。本文所提算法对于标记样本的数量需求显然没有前四种算法那么大,表中 ACOTSS 算法所得出的结果是在标记样本量仅占训练集合 75% 的情况下得出的,即标记样本数量为总样本量的 56.25% (75% × 75%)。

表 3 多种睡眠分期方法对比

方法	特征数量	分类算法	平均正确率/%
Hsu 等人提出	6	NN + PNN + FNN	87.2
Pascualvaca 等人提出	56	SVM 多级分类	76.5
Fraiman 等人提出	20	随机森林分类器	74
高群霞等人提出	10	LS-SVM	92.93
本文方法	8	ACOTSS	89.86

最后,为了量化 ACOTSS 算法对于睡眠分期的性能提升量,本文使用相同标记样本比例下不同算法的精度差作为衡量算法优劣的指标,并得出算法的性能提升比率,如式(8)所示。因此,本文将 ALKLSS 算法以及上述四种算法中分类精度最高的 LS-SVM 算法与本文所提 ACOTSS 算法进行不同标记样本比例下的分类正确率及精度差比较,结果如图 2 和 3 所示。

$$\xi(a, b) = \frac{\sum_{i=\phi_0}^{\phi_n} (\varepsilon_a - \varepsilon_b)}{N_{\phi}} \quad (8)$$

其中: ε 表示分类正确率(精度); ξ 表示算法性能的提升比率; N_{ϕ} 表示抽取精度的个数。

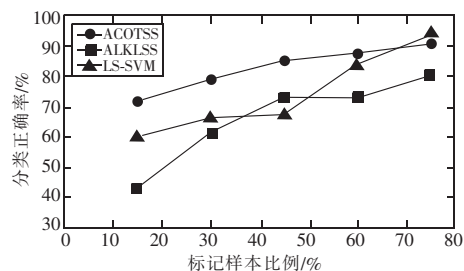


图 2 不同算法分类精度对比

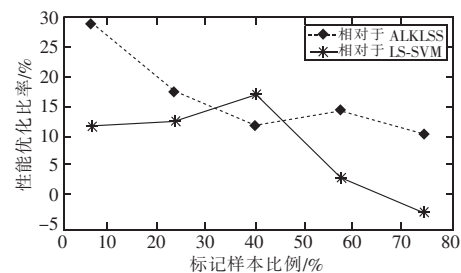


图 3 ACOTSS 算法性能优化比率

从图 2 可以看出,随着标记样本比例的增大,三种算法的分类正确率均呈现上升趋势,当比例达到 75% 时,LS-SVM 算法的分类正确率最高,ACOTSS 算法其次,ALKLSS 算法最差。但是 LS-SVM 算法的训练过程对于标记样本量的要求很高,即当标记量足够高时可以获得较高的分类正确率,但随着标记样本比例的降低,其分类正确率呈现较大斜率的下降趋势,表明其在中低标记样本量下分类性能较差。而当标记样本比例低于 65% 时,ACOTSS 算法的分类正确率在另外两者之上,且其整体正确率曲线下降较为平缓,表明其在中低标记样本比例下仍具备较高的可靠性。而从图 3 可以看出,虽然 ACOTSS 算法在不同比例下的精度提升比呈现负增长趋势,但仍全程优于 ALKLSS 算法,且只有在比例高于 65% 之后负于 LS-SVM 算法。由此可以看出,ACOTSS 算法在睡眠分期领域内可以实现保障分类精度的要求,且由式(8)计算可以得出其在同等标记样本比例下分类正确率相较 ALKLSS、LS-SVM 而言,分别平均提高了 16.83% 和 8.59%。

图 4 为三种不同算法的运行时间比较结果。

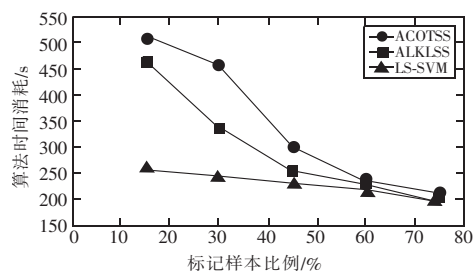


图 4 不同算法的运行时间对比

从图 4 可以看出,由于增加了分期算法的运算步骤,使得 ACOTSS 算法的执行时间相比其他两者而言相对增加,其在中低标记样本比例下的时间消耗均高于 ALKLSS 和 LS-SVM,在 15% ~ 60% 标记样本比例下差别尤其明显,且三种算法的运行时间均呈现不断递减趋势。究其原因在于 ACOTSS 算法通过增加特征优选及样本筛选、训练机制来保障其在低标记样本量

下的分期精度,通过牺牲部分时间资源来换取不同置信度样本的优化识别及标记,而这样的代价对于睡眠样本这种人工分期成本高的数据来说是值得的。在综合考虑了分期正确率、标记样本需求量及运行时间这三方面的因素来看,ACOTSS 算法比起其他两种方法适用范围更广且更加具备可行性。

4 结束语

在睡眠分期领域中,现有的分期算法均未从降低训练集中标记样本需求这一角度来进行研究,使得标注样本所消耗的资源以及在标注过程中因疲劳可能造成的误判现象被忽视。因此,本文在保障睡眠分期正确率的前提下,提出一种蚁群优选与半监督学习相结合的睡眠分期算法,即 ACOTSS 算法。该算法适用于 R&K 分期标准下多导联 EEG 信号的睡眠分期,依靠蚁群算法实现特征优选,再采用协同训练与主动学习相结合的半监督分类算法,有效地在保障分类正确率的前提下降低了标记样本需求量。利用 MIT-BIH 数据库中的数据进行仿真测试,从分类正确率、性能优化比率和运行时间这三个方面来衡量该算法的有效性、可行性,并与其他算法进行比较。实验结果表明,相比 LS-SVM、ALKLSS 算法而言,ACOTSS 算法大大降低了标记样本需求量,且在中低标记样本量下分类性能较好,适用于像睡眠 EEG 信号这样的非线性信号,这为未来的半监督睡眠分期研究提供了参考。

参考文献:

- [1] Ebrahimi F, Setarehdan S K, Nazeran H. Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs [J]. *Biomedical Signal Processing & Control*, 2015, 18(4): 69-79.
- [2] Hassan A R, Bhuiyan M I H. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features [J]. *Journal of Neuroscience Methods*, 2016, 271(9): 107-118.
- [3] Khatun S, Mahajan R, Morshed B I. Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel EEG data [J]. *IEEE Journal of Translational Engineering in Health & Medicine*, 2016, 4(3): 1-5.
- [4] Zhu Guohun, Li Yan, Wen P P. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal [J]. *IEEE Journal of Biomedical & Health Informatics*, 2014, 18(6): 1813-1821.
- [5] 周鹏,李向新,张翼,等. 基于 PCA 和 SVM 的个性化睡眠分期研究 [J]. 中国科技论文在线, [2012-01-20]. <http://www.paper.edu.cn/releasepaper/content/201201-785>.
- [6] 李玉平,夏斌. 样本不平衡的睡眠数据分期研究 [J]. *微型机与应用*, 2016, 35(18): 55-57.
- [7] Djordjevic V R, Gerla V, Lhotska L, et al. EEG based sleep staging in term newborns [C]//Proc of World Congress on Medical Physics and Biomedical Engineering. 2012: 348-351.
- [8] Sen B, Peker M, Cavusoglu A, et al. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms [J]. *Journal of Medical Systems*, 2014, 38(3): 1-21.
- [9] Radha M, Garcia-Molina G, Poel M, et al. Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal [C]//Proc of the 36th International Conference on Engineering in Medicine and Biology Society. 2014: 1876-1880.
- [10] Hassan A R, Bhuiyan M I H. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating [J]. *Biomedical Signal Processing & Control*, 2015, 24(5): 1-10.
- [11] Mora A M, Fernandes C M, Herrera L J. Sleeping with ants, SVMs, multilayer perceptrons and SOMs [C]//Proc of International Conference on Intelligent Systems Design and Applications. 2010: 126-131.
- [12] Aboalayon K A I, Faezipour M, Almuhammadi W S, et al. Sleep stage classification using eeg signal analysis: a comprehensive survey and new investigation [J]. *Entropy*, 2016, 18(9): 272.
- [13] 刘建峰,吕佳. 融合主动学习的改进贝叶斯半监督分类算法研究 [J]. *计算机测量与控制*, 2014, 22(6): 432-434.
- [14] Pan Li, Zheng Hong, Zhang Zuxun, et al. Genetic Feature Selection for Texture Classification [J]. *地球空间信息科学学报: 英文版*, 2004, 7(3): 162-166, 173.
- [15] 李玉平,夏斌. 样本不平衡的睡眠数据分期研究 [J]. *微型机与应用*, 2016, 35(18): 55-57, 61.
- [16] Krishnanand K R, Panigrahi B K, Rout P K, et al. Application of multi-objective teaching-learning-based algorithm to an economic load dispatch problem with incommensurable objectives [C]//Proc of the 2nd International Conference on Swarm, Evolutionary and Memetic Computing. Berlin: Springer-Verlag, 2011: 697-705.
- [17] Rao R V, Patel V. An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems [J]. *International Journal of Industrial Engineering Computations*, 2012, 3(4): 535-560.
- [18] Rajasekhar A, Rani R, Ramya K, et al. Elitist teaching-learning opposition based algorithm for global optimization [C]//Proc of IEEE International Conference on Systems, Man, and Cybernetics. 2012: 1124-1129.
- [19] Rao R V, Patel V. An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems [J]. *Scientia Iranica*, 2013, 20(3): 710-720.
- [20] 于坤杰,王昕,王振雷. 基于反馈的精英教学优化算法 [J]. *自动化学报*, 2014, 40(9): 1976-1983.
- [21] 拓守恒,雍龙泉,邓方安. “教与学”优化算法研究综述 [J]. *计算机应用研究*, 2013, 30(7): 1933-1938.

(上接第 1389 页)

- [2] Rao R V, Savsani V, Vakharia D P. Teaching-learning-based optimization: a novel optimization method for continuous non-linear large scale problems [J]. *Information Sciences*, 2012, 183(1): 1-15.
- [3] Rao R V, Savsani V. Mechanical design optimization using advanced optimization techniques [M]. London: Springer-Verlag, 2012.
- [4] Rao R V, Patel V. Multi-objective optimization of heat exchangers using a modified teaching-learning-based optimization algorithm [J]. *Applied Mathematical Modelling*, 2013, 37(3): 1147-1162.
- [5] Rao R V, Patel V. Multi-objective optimization of two stage thermoelectric cooler using a modified teaching-learning-based optimization algorithm [J]. *Engineering Applications of Artificial Intelligence*, 2013, 26(1): 430-445.
- [6] Toğan V. Design of planar steel frames using teaching-learning based optimization [J]. *Engineering Structures*, 2012, 34(1): 225-232.
- [7] Satapathy S C, Naik A. Data clustering based on teaching-learning-based optimization [C]//Proc of the 2nd International Conference on Swarm, Evolutionary and Memetic Computing. Berlin: Springer-Verlag, 2011: 148-156.