

基于递归定量分析对 心率变异性进行睡眠分期的研究

王梦琪, 魏新江

(鲁东大学 数学与统计科学学院, 山东 烟台 264039)

摘要: 本文提出一种基于相关系数和互信息的特征选择算法, 利用递归定量分析方法对 16 名健康受试者的 RR 间期序列提取特征, 进而实现心率变异性的高精度睡眠分期。通过 Z 得分标准化减少不同受试者特征间的差异, 对每个特征建立相应的隐马尔可夫模型, 采用集体投票思想获得整体睡眠分期结果。实验表明, 递归定量分析方法所提取的信息与睡眠分期工作有关, 使用文中提出的特征选择方法可以有效地提高分类性能。

关键词: 递归定量分析; 心率变异性; 隐马尔可夫模型; 睡眠分期; 互信息

中图分类号: O213 **文献标志码:** A **文章编号:** 1673-8020(2021)02-0097-06

睡眠质量对人们的身心健康有很大的影响, 有相当多的病理与睡眠有关^[1], 睡眠阶段检测是解决睡眠质量评估问题的前期工作。起初, 睡眠分期主要依据脑电图 (Electroencephalogram, EEG) 信号进行研究^[2-5]。根据美国睡眠医学学会 (American Academy of Sleep Medicine, AASM) 的睡眠分期指南^[6], 多导睡眠图 (Polysomnography, PSG) 是检测睡眠阶段的高标准方式, 可以获得多种生理信号, 但是需要在睡眠中心的专用设备上进行监测, 其结果要求由专业和训练有素的人员诊断^[1]。目前, 许多学者已经使用其他方法获取高精度的生理信号进行睡眠阶段检测^[1, 7-8]。

心率变异性 (heart rate variability, HRV) 与自主神经系统活动密切相关, 该信号在不同睡眠阶段的变化趋势有显著差异^[9]。在日常生活中, 许多家用和便携式传感器易于获取心电信号, 如床垫传感器、T 恤、手部和胸部活动记录仪^[1, 7-8]。心率采集对受试者的睡眠影响较小, 所以从心率变化的角度获取睡眠信息有着十分重要的意义。近年来, HRV 信号对睡眠分期的研究主要从时域、频域和非线性域进行探索^[10-13]。文献 [10] 提取了不同频段的能量总和作为特征值, 以此验

证极低频段与高频段的功率谱能量之比能够更好地区分睡眠阶段的差异; 除采用时域、频域特征外, 文献 [11-12] 使用了去趋势波动分析、样本熵等非线性域特征探索不同睡眠阶段的差异性, 得出非线性特征识别性更高的结论。

睡眠是一个非常复杂的过程, 睡眠分期研究中关于非线性复杂动态系统特征的提取尚未得到充分探讨。递归定量分析 (recurrence quantification analysis, RQA) 是一种有前景的非线性动态系统分析方法。针对健康受试者和睡眠呼吸暂停综合症患者, 文献 [14] 表明 RQA 方法可以获取患者的心率动态。此外, 在脑电信号、心肺信号和呼吸信号的研究中, RQA 方法均被验证可以改进睡眠阶段识别系统^[15-17]。因此, RQA 方法用于 HRV 信号的睡眠分期有一定的研究价值。同时, 为消除高维特征的信息冗余, 特征选择是研究睡眠分期的一个必要环节。目前有三种特征选择方法, 分别是过滤、包装和嵌入方法, 它们主要考虑了特征之间的相关性和距离特性。事实上, 在选择特征时也要考虑特征与类的相关性, 这样所选取的特征能够更好地反映信号信息。

在现有研究基础上, 本文将 RQA 方法引入到 HRV 信号研究中, 提出一种考虑特征之间的相关

收稿日期: 2020-10-26; 修回日期: 2021-01-12

基金项目: 国家自然科学基金 (61973149)

第一作者简介: 王梦琪 (1997—), 女, 山东菏泽人, 硕士研究生, 研究方向为经济与社会统计。E-mail: wangmengqild@163.com

通信作者简介: 魏新江 (1977—), 男, 山东东营人, 教授, 硕士研究生导师, 博士, 研究方向为非线性系统控制、鲁棒控制等。E-mail: weixinjiang@163.com

性、特征与类之间的互信息的特征选择算法,以实现改善睡眠分期性能的目标。

1 方法和材料

1.1 数据来源和预处理

本研究使用 MIT-BIH 的多导睡眠数据库,其中 16 名男性受试者在波士顿贝斯以色列医院的睡眠实验室进行监测。该数据库包括 18 条多导睡眠记录,每条记录都包含不同导联监测得到的生理信号。本文仅对心电信号记录文件进行分析,获取 R 峰的位置,同时通过心跳标注文件获得不同睡眠阶段对应的睡眠标签。由于多导睡眠图数据以 30 s 为周期进行记录,所以对 R 峰序列执行 30 s 的循环操作来获取多个睡眠片段。将睡眠片段总数与睡眠标签总数进行比较,发现有 7 名受试者存在不一致现象,但是注释文件没有详细的描述,最终选择剩余的 9 名受试者作为实验对象。在睡眠片段内进行 R 峰值的差分处理,获取 RR(R wave-to-R wave) 间期序列,即 HRV 信号,并将 9 名受试者的睡眠阶段分为清醒、非快速眼动(non-rapid eye movement, NREM) 和快速眼动(rapid eye movement, REM),统计结果如表 1 所示。采用三次样条插值方法,对 RR 间期序列进行采样率为 5 Hz 的重采样,将原始的 RR 间期序列转换为等时间间隔序列。

表 1 9 名受试者的睡眠标签总数

Tab.1 Total number of sleep labels for 9 subjects

受试者	清醒/个	非快速眼动/个	快速眼动/个
slp01a	8	219	13
slp01b	180	155	25
slp02a	52	231	77
slp02b	108	133	29
slp03	135	495	74
slp04	162	535	23
slp32	394	246	0
slp41	229	461	90
slp48	214	515	31
slp60	286	386	31
slp67x	72	82	0

1.2 特征提取

1.2.1 文献特征

根据睡眠分期中关于特征提取的相关研究,

本文对 HRV 信号在时域、频域和非线性域上共提取了 17 个特征。采用原始的 RR 区间序列,提取的时域特征包括 RR 区间序列的均值和标准差、RR 区间差分序列的均方根和标准差、相邻的 RR 间期之差大于 50 ms 的个数与总个数的百分比。此外,基于经典的功率谱密度估计技术——快速傅里叶变换,计算的频域指标包括:RR 间期序列在 0~0.4 Hz 的功率值(TP)、0.15~0.4 Hz 的功率值(HF)、0.04~0.15 Hz 的功率值(LF)、0.003~0.04 Hz 的功率值(VLF)和 LF/HF;非线性域提取的四个识别睡眠阶段有良好性能的指标:样本熵、模糊熵、近似熵和去趋势波动分析特征。

1.2.2 递归定量分析

将信号表示为一个时间序列 $\{x(t), t = 1, 2, \dots, T\}$,其中 T 表示信号记录的总时间。如果观察其动态变化过程,有必要对一维时间序列进行重构,形成复杂的非线性动态系统。RQA 是研究动态系统复杂行为的有效方法,可以描述递归图的递归状态,以递归点和线段为计算单位,提取微观层面的信息。

递归图是原始信号重构后相空间的二维图形展示,能否充分提取原始信息取决于重构参数时间延迟 τ 和嵌入维数 m 。目前,互信息法和虚假最近邻点法分别是确定最佳时间延迟 τ 和嵌入维数 m 的常用方法^[18]。此外,阈值 ε 对递归矩阵的计算有较大的影响:如果 ε 选择太小,几乎没有递归点,递归结构不能有效反映系统性能;如果 ε 选择太大,可能会出现大量不存在的递归结构,同时可能会包含很多孤立点,导致提取的系统信息不符合实际情况。参考相关的递归定量分析文献发现,当阈值使递归率达到约 10% 时,能够确定递归定量分析指标的动态范围^[16]。

本文提取的 RQA 特征主要包括:递归率、确定率、平均对角线长度、最大对角线长度、递归熵、层状度、捕获时间、最大垂直线长度、第一和第二类型的递归时间、递归周期密度熵、聚类系数和传递性。

1.3 特征选择

通过不同的分析方法得到的特征具有很强的相关性,从而导致信息重复,为了更合理地去除特征间的冗余信息,选择合理的评估标准尤为重要。本文提出了相关性与互信息特征选择(correlation and mutual information feature selection, CMIFS)算

法,即对于提取的高维特征,计算特征与分类变量之间的互信息,以及特征之间的相关矩阵。当互信息值较大时,说明两个变量的相关性更强^[19];当两个变量相关系数大于 0.9 时,认为两者高度线性相关。依据上述两条准则,挑选出高度相关的配对变量,再根据两者的互信息值,保留互信息值较大的变量。

1.4 隐马尔可夫模型

研究表明,统计模型作为检测睡眠阶段的分类器有良好的分类性能,如判别分析、朴素贝叶斯和隐马尔可夫模型(hidden Markov model, HMM)^[1,12,19-22]。文献[21]验证了 HMM 作为睡眠分期分类器的可行性和有效性:不同睡眠状态的转变在信号上存在一定的相关关系,而 HMM 可以有效抓住这些时序上的特性,得到整夜睡眠状态的转移矩阵,使该模型有利于分析睡眠状况。因此,本文最终选择 HMM 作为睡眠阶段分类器。

隐马尔可夫模型可以在给定的观察序列下预测隐状态,是一种有效的识别工具。在本文中,隐状态分为清醒、NREM 和 REM 三个睡眠阶段,提取的特征序列为观察序列。针对单个特征序列,为了减小不同受试者间的差异,采用 Z 得分方法对特征进行标准化;同时,将特征的最小值与最大值的差值分成 10 等份,进行离散化处理,得到特征的符号序列^[20];然后,对特征符号序列建立隐马尔可夫模型,生成对应的分类结果;最后,利用集成投票的思想估计总体的隐状态序列,即睡眠分期结果。

2 结果讨论与分析

将 9 名受试者每个睡眠页面内的 RR 间期序列作为计算一组特征值的一维序列,计算结果按行排列生成特征矩阵。

时域特征是基于原始信号序列使用统计方法计算得到。关于 HRV 的频域测量过程,首先采用快速傅里叶变换,在整个时域内对能量有限的信号进行分解,并将其表示为不同频率的三角正弦波的线性叠加;然后,估计功率谱密度,累加不同频段对应的能量值,得到 5 个频域特征。如 1.2.1 节所述,非线性特征包括样本熵、近似熵、模糊熵和去趋势波动分析特征,其中计算去趋势分析特征时采用一阶多项式拟合去趋势阶段的趋势值。

此外,分别采用互信息法和虚假最近邻点法确定系统的最佳时间延迟 τ 和嵌入参数 m ,计算工具是 CRP(cross recurrence plot)工具箱中的函数 $mi(\cdot)$ 和 $fnn(\cdot)$ ^[23]。根据确定的参数值对 RR 间期序列进行处理,得到互信息图像(图 1),以及不同嵌入维数对应的邻接误差百分比(表 2)。

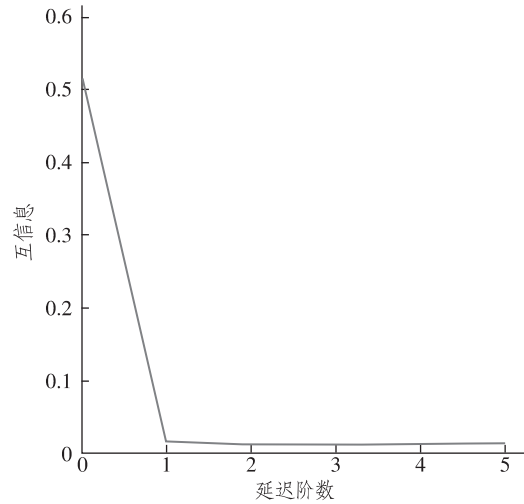


图 1 不同时间延迟下的互信息曲线
Fig.1 Mutual information curve of different time delays

表 2 不同嵌入维数的邻接误差百分比
Tab.2 The adjacency error percentage of different embedded dimension

嵌入维数	邻接误差百分比	嵌入维数	邻接误差百分比
1	1	9	0.055
2	0.96	10	0.025
3	0.77	11	0.015
4	0.355	12	0.015
5	0.18	13	0.01
6	0.08	14	0.02
7	0.07	15	0.01
8	0.03		

由互信息选取时间延迟的准则为:第一个极小值点对应的值为最佳结果^[18];当邻接误差百分比小于 5%(接近 0)或保持不变时,最小的嵌入维数为最佳值^[24]。由图 1 和表 2 中指标值的变化情况,确定最佳时间延迟 $\tau = 3$,最终嵌入维数 $m = 10$ 。分别计算阈值 ε 为 10%、20%、30% 对应的递归率。结果表明,当 $\varepsilon = 10\%$ 时,大部分的递归率约为 10%。因此,将 10% 作为阈值 ε 的最终结果,利用选择的最佳时延和嵌入维数值重构相空间,基于得到的递归图信息估算 RQA 指标。

在睡眠阶段识别过程中,文献[11-12]已充分证明了时域、频域、熵和去趋势波动分析特征的

可行性,本文仅讨论 RQA 特征在不同睡眠阶段是否存在显著差异。首先对提取的 RQA 特征进行正态性检验,结果中 P 值均小于显著性水平 0.05,可推断 RQA 特征的分布不满足正态性。因此,选择非参数 $K-W$ 检验来判断清醒阶段、NREM 阶段和 REM 阶段的 RQA 特征差异,检验结果表明:RQA 特征能够有效区分不同睡眠阶段 ($P < 0.05$)。13 个 RQA 特征的均值 \pm 标准差结果如表 3 所示,其中平均对角线长度 L 、最大对角线长度 L_{\max} 、捕获时间 TT 、最大垂直线长度 V_{\max} 和第二类型的递归时间 $T(2)$ 的标准差较大,即特征取值波动范围较大,表明不同的受试者间存在明显的个体差异。

表 3 RQA 特征的统计分析

Tab.3 The statistic analysis of RQA features

RQA 特征	清醒期	NREM	REM
<i>REC</i>	0.301 \pm 0.225	0.185 \pm 0.189	0.229 \pm 0.216
<i>DET</i>	0.989 \pm 0.008	0.987 \pm 0.026	0.985 \pm 0.052
<i>L</i>	28.164 \pm 10.664	24.514 \pm 9.39	25.248 \pm 10.456
L_{\max}	117.37 \pm 11.616	108.6 \pm 25.93	107.89 \pm 27.159
<i>ENT</i>	3.29 \pm 0.558	2.89 \pm 0.688	3.039 \pm 0.752
<i>LAM</i>	0.972 \pm 0.12	0.881 \pm 0.272	0.889 \pm 0.274
<i>TT</i>	13.823 \pm 10.975	8.262 \pm 7.172	10.282 \pm 10.32
V_{\max}	42.13 \pm 30.352	26.92 \pm 25.871	31.87 \pm 28.746
$T(1)$	1.957 \pm 1.066	2.259 \pm 1.928	1.97 \pm 1.1
$T(2)$	22.976 \pm 10.633	16.02 \pm 9.552	17.491 \pm 11.262
<i>Trec</i>	0.628 \pm 0.207	0.526 \pm 0.263	0.542 \pm 0.254
<i>Clust</i>	0.725 \pm 0.125	0.615 \pm 0.175	0.647 \pm 0.187
<i>Trans</i>	0.731 \pm 0.122	0.652 \pm 0.164	0.684 \pm 0.171

进一步,将提取的 30 个特征进行标准化和离散化处理,以减弱不同受试者的睡眠差异,并将处理后的特征应用到隐马尔可夫模型的后续分析中。将所有特征构造为四个子集:特征集 A 只包含时域和频域特征;第二个子集包含时域、频域和非线性特征,定义为集合 B ;第三个特征子集包含所有特征,为特征集 C ;第四个子集只包含 RQA 特征,命名为特征集 D 。为了与文献 [25] 的研究结果对比,本文使用相同的特征子集 A (剔除 RR 区间差分序列的标准差) 建立分类模型,计算性能评估指标 Kappa 系数,特征选择前后的评估指标对比结果如图 2 所示。此外,利用 CMIFS 算法进行特征选择,基于选择的特征集拟合隐马尔可夫模型,所得到的性能指标如表 4 所示。

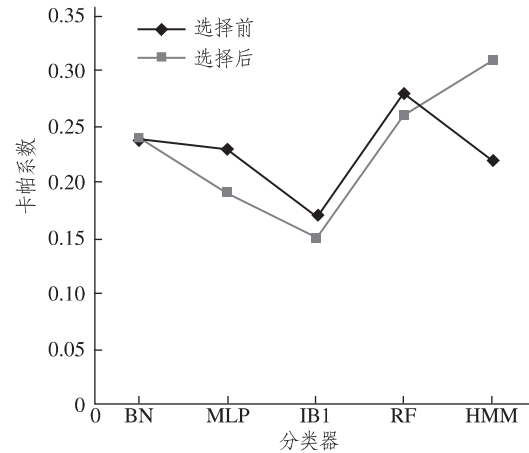


图 2 卡帕系数的比较

Fig.2 Comparison of Kappa coefficient

表 4 选择的特征子集的评估指标

Tab.4 The evaluation indicators of the selected feature subset

特征集	准确率/%	Kappa 系数
A	59.44	0.31
B	67.88	0.37
C	68.69	0.38
D	63.06	0.29

从图 2 可以看出:如果只考虑时域和频域特征,HMM 分类器得到的结果并不是最优的,准确率仅为 58.71%;如果使用本文提出的 CMIFS 特征选择算法建立 HMM,特征集 A 的一致性显著提高。文献 [25] 利用基于相关性的特征子集选择 (correlation-based feature subset selection, CFS) 算法,但四个分类器的 Kappa 系数没有得到更好的结果。由表 4 可以看出,根据本文的睡眠分期系统得到的一致性指标结果均优于文献 [25] 中机器学习分类器的结果。此外,如果仅使用 RQA 特征来进行睡眠分期,评估指标并不理想;而采用所有特征时,其性能是最佳的。

同样地,采用隐马尔可夫模型分类器评估 HRV 的睡眠宏观结构,文献 [20] 得到的分期性能优于本文的结果,总体 Kappa 系数为 0.42 \pm 0.1493,总体准确率为 79.43% \pm 8.83%。特别指出,文献 [20] 仅采用了时变自回归模型提取时频域特征。因此,优化特征提取器提取时频域特征将是本文睡眠分期系统的改进方向。

3 结语

心率变异性分析在睡眠分期研究中具有重要的应用价值。到目前为止,HRV 信号在时域、频

域和非线性域得到了广泛的研究,但是并没有从动态系统相空间的角度来描述其变化。本文加入递归定量分析特征构建睡眠分期系统,从HRV序列中共获取30个特征,使用改进的CMIFS特征选择算法确定了22个特征用于建立HMM模型。通过对每个受试者的特征进行归一化和离散化处理,HMM分类器识别清醒期、NREM和REM阶段的科恩Kappa系数为0.38,总体准确率为68.69%。未来的工作中,在本文的研究基础上使用时频分析技术将有助于获得更好的睡眠分期性能。

参考文献:

- [1] MIGLIORINI M, BIANCHI A M, NISTICÒ D, et al. Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors [C] // 32nd Annual International Conference of the IEEE EMBS Buenos Aires, 2010: 3273–3276.
- [2] LARSEN L E, WALTER D O. On automatic methods of sleep staging by EEG spectra [J]. *Electroencephalography and Clinical Neurophysiology*, 1970, 28(5): 459–467.
- [3] SMITH J R, KARACAN I. EEG sleep stage scoring by an automatic hybrid system [J]. *Electroencephalography and Clinical Neurophysiology*, 1971, 31(3): 231–237.
- [4] LARSEN L E, RUSPINI E H, MCNEW J J. A test of sleep staging systems in the unrestrained chimpanzee [J]. *Brain Research*, 1972, 40(2): 319–343.
- [5] GATH I, SCHWARTZ L. Syntactic pattern recognition applied to sleep EEG staging [J]. *Pattern Recognition Letters*, 1989, 10(4): 265–272.
- [6] IBER C, ANCOLI-ISRAEL S, CHESSON A L, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, American Academy of Sleep Medicine [EB/OL]. [2020-10-05]. <https://rtsleepworld.com/wp-content/uploads/2018/03/9.-Morphing-Manual-Adult-Rules.pdf>.
- [7] BIANCHI A M, MENDEZ M O. Automatic detection of sleep macrostructure based on a sensorized T-shirt [C] // 32nd Annual International Conference of the IEEE EMBS Buenos Aires, 2010: 3606–3609.
- [8] AKTARUZZAMAN M, RIVOLTA M W, KAEMACHARYA R, et al. Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification [J]. *Computers in Biology and Medicine*, 2017, 89: 212–221.
- [9] TRINDER J, KLEIMAN J, CARRINGTON M, et al. Autonomic activity during human sleep as a function of time and sleep stage [J]. *Journal of Sleep Research*, 2001, 10(4): 253–264.
- [10] HOSSEN A, ÖZER H, HEUTE U. Identification of sleep stages from heart rate variability using a soft-decision wavelet-based technique [J]. *Digital Signal Processing*, 2013, 23(1): 218–229.
- [11] VIGO D E, DOMINGUEZ J, GUINJOAN S M, et al. Nonlinear analysis of heart rate variability within independent frequency components during the sleep-wake cycle [J]. *Autonomic Neuroscience*, 2010, 154(1/2): 84–88.
- [12] EBRAHIMI F, SETAREHDAN S K, AYALA-MOYEDA J, et al. Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals [J]. *Computer Methods and Programs in Biomedicine*, 2013, 112(1): 47–57.
- [13] GENG D Y, ZHAO J, WANG C X, et al. A decision support system for automatic sleep staging from HRV using wavelet packet decomposition and energy features [J]. *Biomedical Signal Processing and Control*, 2020, 56: 1–9.
- [14] SMIETANOWSKI M, SZELENBERGER W, TRZEBSKI A. Nonlinear dynamics of the cardiovascular parameters in sleep and sleep apnea [J]. *Journal of Physiology and Pharmacology*, 2006, 57(11): 55–68.
- [15] SONG I H, LEE D S, KIM S I. Recurrence quantification analysis of sleep electroencephalogram in sleep apnea syndrome in humans [J]. *Neuroscience Letters*, 2004, 366(2): 148–153.
- [16] ROLINK J, KUTZ M, FONSECA P, et al. Recurrence quantification analysis across sleep stages [J]. *Biomedical Signal Processing and Control*, 2015, 20: 107–116.
- [17] TERRILL P I, WILSON S J, SURESH S, et al. Application of recurrence quantification analysis to automatically estimate infant sleep states using a single channel of respiratory data [J]. *Medical and Biological Engineering and Computing*, 2012, 50(8): 851–865.
- [18] TENG G Y, ZHOU X J, YANG C L, et al. Ultrasonic detection method of micro defects in thick-section CFRP [J]. *Optics and Precision Engineering*, 2018, 26(12): 3108–3117.
- [19] PALACIOS M, VALLVERDU M, FRIEDRICH H, et al. Hidden Markov models and mutual information analysis to characterize nonlinear dynamics in heart rate variability [C] // *IEEE Engineering in Medicine and Biology*

- 27th Annual Conference, 2005: 1-4.
- [20] MENDEZ M O, MATTEUCCI M, CERUTTI S, et al. Sleep staging classification based on HRV: time-variant analysis [C] // 31st Annual International Conference of the IEEE EMBS Minneapolis, 2009: 9-12.
- [21] JIANG C H, LI J W, FENG H Q, et al. Application of hidden Markov model (HMM) in sleep staging [J]. Shandong Journal of Biomedical Engineering, 2003, 22 (2): 4-7.
- [22] GIRI E P, ARYMURTHY A M, FANANY M I, et al. Sleep stages classification using shallow classifiers [C] // International Conference on Advanced Computer Science and Information Systems, 2015: 297-301.
- [23] MARWAN N. Cross recurrence plot toolbox for MATLAB [EB/OL]. [2020-08-17]. http://tocsy.pik-potsdam.de/CRPtoolbox/crp_man.pdf.
- [24] YANG C L, FENG W, BIAN C L, et al. Micro defects detection in metallic materials based on recurrence analysis of ultrasonic backscattering signal [J]. Optics and Precision Engineering, 2019, 27(4): 932-944.
- [25] NOVIYANTO A, ISA S, WASITO I. Selecting features of single lead ECG signal for automatic sleep stages classification using correlation-based feature subset selection [J]. International Journal of Computer Science Issues, 2011, 8(5): 139-148.

Sleep Staging for Heart Rate Variability Based on Recurrence Quantification Analysis

WANG Mengqi, WEI Xinjiang

(School of Mathematics and Statistics Science, Ludong University, Yantai 264039, China)

Abstract: To realize high-precision sleep staging for heart rate variability, a feature selection algorithm based on correlation coefficient and mutual information was proposed in this paper, and recurrence quantification analysis method was used to extract features from the RR interval sequences of 16 healthy subjects. By utilizing Z-score standardization, the differences among features of different subjects were reduced, and a hidden Markov model was established for each feature. Then, the integrated voting idea was executed to obtain the overall sleep staging results. The experiment result indicates that the extracted information of recurrence quantification analysis is relevant to sleep staging work, which effectively improves the classification performance by using the proposed feature selection method.

Keywords: recurrence quantification analysis; heart rate variability; hidden Markov model; sleep staging; mutual information

(责任编辑 顾建忠)