



(12)发明专利申请

(10)申请公布号 CN 108968915 A

(43)申请公布日 2018.12.11

(21)申请号 201810600753.4

(22)申请日 2018.06.12

(71)申请人 山东大学

地址 250061 山东省济南市历下区经十路
17923号

(72)发明人 魏守水 张志民 江兴娥 陈雷
刘振华

(74)专利代理机构 济南圣达知识产权代理有限
公司 37221

代理人 黄海丽

(51)Int.Cl.

A61B 5/00(2006.01)

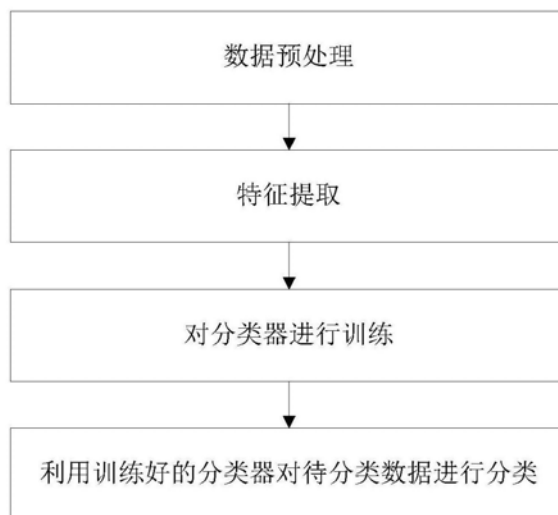
权利要求书4页 说明书15页 附图3页

(54)发明名称

基于熵特征及支持向量机的睡眠状态分类
方法及系统

(57)摘要

本发明公开了基于熵特征及支持向量机的睡眠状态分类方法及系统,包括:数据预处理:基于研究睡眠分期三个通道信号:脑电信号Fpz-Cz EEG,脑电信号Pz-Oz EEG和水平眼电图信号EOG;将每个通道的信号均按照设定时间间隔进行均匀分割,分割后得到若干个片段;特征提取:从每个片段中提取模糊熵、模糊测度熵和样本熵;特征融合;设计一对多支持向量机分类器,支持向量机分类器对待分类信号进行分类,可以实现2-6多分类。



1. 基于熵特征及支持向量机的睡眠状态分类方法,其特征是,包括:

步骤(1):数据预处理:

采集脑电信号Fpz-Cz EEG,脑电信号Pz-0z EEG和水平眼电图信号EOG;将采集的信号均按照设定时间间隔进行均匀分割,分割后得到若干个片段;

每个片段都有预先标记好的睡眠阶段,所述睡眠阶段包括:清醒阶段AWA、阶段一S1、阶段二S2、阶段三S3、阶段四S4和快速眼动期REM;

步骤(2):特征提取:从每个片段中提取三个熵特征:模糊熵、模糊测度熵和样本熵;

步骤(3):设计一对多支持向量机分类器,将提取的熵特征和预先标记好的睡眠阶段输入到一对多支持向量机分类器中,对一对多支持向量机分类器进行训练;

步骤(4):利用训练好的支持向量机分类器对待分类信号进行分类。

2. 如权利要求1所述的基于熵特征及支持向量机的睡眠状态分类方法,其特征是,所述模糊测度熵的计算过程如下:

(a) 假设 x 为脑电信号Fpz-Cz EEG、脑电信号Pz-0z EEG或者眼电信号EOG,对信号按照设定时间间隔进行分割,得到若干个时间序列,对于每一个时间序列 $\{x(i):1 \leq i \leq N\}$,通过重构序列生成 m 维局部序列段 X_i^m :

$$X_i^m = \{x(i), x(i+1), \dots, x(i+m-1)\} - x_0(i) \quad (1)$$

在(1)中, $i=1, \dots, N-m$, $\{x(i), x(i+1), \dots, x(i+m-1)\}$ 表示从第 i 个时间序列开始的连续的 m 个时间序列, $x_0(i)$ 表示 m 个时间序列的平均值;

计算对应信号的局部序列段 $XL_m(i) = X_i^m$;

通过去除整个信号 $\{x(i):1 \leq i \leq N\}$ 的平均值来定义全局序列段,

$$\text{即 } XF_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\} - \bar{x},$$

其中, \bar{x} 表示时间序列 $\{x(i):1 \leq i \leq N\}$ 的平均值, $x(i)$ 表示信号每间隔30秒分割之后得到的第 i 个信号片段, N 表示信号分割之后得到的总片段数;

(b) 计算信号的第 i 个局部序列段 $XL_m(i)$ 与第 j 个局部序列段 $XL_m(j)$ 的距离 $dL_m(i, j)$,如下:

$$dL_m(i, j) = d[XL_m(i), XL_m(j)] = d_{ij}^m \quad (2)$$

计算信号的第 i 个全局序列段 $XF_m(i)$ 和信号的第 j 个全局序列段 $XF_m(j)$ 之间的距离 $dF_m(i, j)$, $i, j=1 \sim N-m, j \neq i$,如下:

$$dF_m(i, j) = d[XF_m(i), XF_m(j)] = \max_{k \in (0, m-1)} \{|x(i+k) - \bar{x} - (x(j+k) - \bar{x})|\} \quad (3)$$

(c) 然后,计算信号的局部相似度 $DL_m(i, j)$ 和全局相似度 $DF_m(i, j)$ 如下:

$$DL_m(i, j) = \mu(dL_m(i, j), n, r) = \exp(-(dL_m(i, j)/r)^n) = D_{ij}^m \quad (4)$$

$$DF_m(i, j) = \mu(dF_m(i, j), n, r) = \exp(-(dF_m(i, j)/r)^n) \quad (5)$$

(d) 定义信号的局部相似度 $DL_m(i, j)$ 的平均值 $\phi L_m(n, r)$ 和全局相似度 $DF_m(i, j)$ 的平均值 $\phi F_m(n, r)$,如下:

$$\phi L_m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DL_m(i, j) \right) = \phi^m(n, r) \quad (6)$$

$$\phi F_m(n,r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DF_m(i,j) \right) \quad (7)$$

(e) 重复(a)到(d)的步骤,计算信号的局部相似度 $DL_{m+1}(i,j)$ 的平均值 $\phi L_{m+1}(n,r)$ 和全局相似度 $DF_{m+1}(i,j)$ 的平均值 $\phi F_{m+1}(n,r)$;

(f) 计算信号的局部模糊测度熵FuzzyLME $n(m,n,r,N)$ 和全局模糊测度熵FuzzyFME $n(m,n,r,N)$:

$$\text{FuzzyLME}n(m,n,r,N) = \ln \phi L_m(n,r) - \ln \phi L_{m+1}(n,r) \quad (8)$$

$$\text{FuzzyFME}n(m,n,r,N) = \ln \phi F_m(n,r) - \ln \phi F_{m+1}(n,r) \quad (9)$$

(g) 最后,信号的模糊测度熵FuzzyME $n(m,n,r,N)$ 计算如下:

$$\text{FuzzyME}n(m,n,r,N) = \text{FuzzyLME}n(m,n,r,N) + \text{FuzzyFME}n(m,n,r,N) \quad (10)。$$

3. 如权利要求1所述的基于熵特征及支持向量机的睡眠状态分类方法,其特征是,所述设计一对多支持向量机分类器的具体步骤为:

如果是完成睡眠阶段的二分类任务,则使用一个二分类支持向量机;

如果是完成睡眠阶段的三分类任务,则使用两个串联的二分类支持向量机;串联时,在后的二分类支持向量机的输入端与在前的二分类支持向量机的其中一个输出端连接;在前的二分类支持向量机的另外一个输出端和在后的二分类支持向量机的输出端均用于输出分类结果;

如果是完成睡眠阶段的四分类任务,则使用三个串联的二分类支持向量机,串联时,在后的二分类支持向量机的输入端均与在前的二分类支持向量机的其中一个输出端连接;三个串联的二分类支持向量机的其他输出端均用于输出分类结果;

如果是完成睡眠阶段的五分类任务,则使用四个串联的二分类支持向量机;串联时,在后的二分类支持向量机的输入端均与在前的二分类支持向量机的其中一个输出端连接;四个串联的二分类支持向量机的其他输出端均用于输出分类结果;

如果是完成睡眠阶段的六分类任务,则使用五个串联的二分类支持向量机;串联时,在后的二分类支持向量机的输入端均与在前的二分类支持向量机的其中一个输出端连接;五个串联的二分类支持向量机的其他输出端均用于输出分类结果。

4. 如权利要求1所述的基于熵特征及支持向量机的睡眠状态分类方法,其特征是,所述步骤(3)的具体步骤为:

步骤(301):统计训练睡眠标记数据中处于每个睡眠阶段的片段数;

步骤(302):按照片段数由多到少顺序对睡眠阶段进行排序;片段数排序第一的睡眠阶段设为A阶段;片段数排序第二的睡眠阶段为B阶段;片段数排序第三的睡眠阶段为C阶段;片段数排序第四的睡眠阶段为D阶段;片段数排序第五的为E阶段;

步骤(303):对分类任务进行分析,对一对多支持向量机进行选择:

如果分类任务是区分睡眠阶段和清醒阶段AWA,则采用一个二分类支持向量机;

如果分类任务是区分睡眠阶段、快速眼动期REM和非快速眼动期,则采用两个串联的二分类支持向量机;

如果分类任务是区分清醒阶段AWA、第一非快速眼动期、第二非快速眼动期和快速眼动期REM,则采用三个串联的二分类支持向量机;

如果分类任务是区分清醒阶段AWA、阶段一S1、阶段二S2、第三非快速眼动期和快速眼

动期REM,则采用四个串联的二分类支持向量机;

如果分类任务是区分清醒阶段AWA、阶段一S1、阶段二S2、阶段三S3、阶段四S4和快速眼动期REM,则采用五个串联的二分类支持向量机;

步骤(304):利用选择好的一对多支持向量机分类器进行分类;

每个二分类支持向量机进行分类都是从睡眠片段数最多的睡眠阶段开始分类:将排序靠前的一类睡眠阶段归为一类,同时将其他剩余的睡眠阶段归为另一类;然后,针对其他剩余的睡眠阶段,再将排序靠前的一类睡眠归为一类,再将其他剩余的睡眠阶段归为另一类;以此类推,直至所有的睡眠阶段归类完毕,结束。

5.如权利要求4所述的基于熵特征及支持向量机的睡眠状态分类方法,其特征是,所述步骤(304)的具体步骤为:

如果是完成睡眠状态阶段的六分类任务,则使用五个串联的二分类支持向量机,所述五个串联的二分类支持向量机依次为:第一二分类支持向量机、第二二分类支持向量机、第三二分类支持向量机、第四二分类支持向量机和第五二分类支持向量机;

首先,将所有睡眠阶段的片段数均输入到第一二分类支持向量机中;输出值有两个分支,第一个分支的输出仅仅包括A阶段的片段数,第二个分支的输出包括B阶段、C阶段、D阶段和E阶段的片段数;

然后,将第一二分类支持向量机的第二个分支的输出值输入到第二二分类支持向量机中;第二二分类支持向量机的输出值有两个分支,第一个分支的输出仅仅包括B阶段的片段数,第二个分支的输出包括C阶段、D阶段和E阶段的片段数;

然后,将第二二分类支持向量机的第二个分支的输出值输入到第三二分类支持向量机中;第三二分类支持向量机的输出值有两个分支,第一个分支的输出仅仅包括C阶段的片段数;第二个分支的输出包括D阶段和E阶段的片段数;

然后,将第三二分类支持向量机的第二个分支的输出值输入到第四二分类支持向量机中;第四二分类支持向量机的输出值有两个分支,第一个分支的输出仅仅包括D阶段的片段数;第二个分支的输出包括E阶段的片段数;至此,一对多支持向量机训练结束。

6.如权利要求4所述的基于熵特征及支持向量机的睡眠状态分类方法,其特征是,所述步骤(304)的具体步骤为:

在进行A阶段、B阶段和C阶段三分类时,先将A阶段作为第一大类,将B阶段和C阶段作为第二大类,首先使用第一个二分类支持向量机进行分类分别输出A阶段和第二大类;然后使用第二个二分类支持向量机对第二大类进行分类,输出B阶段和C阶段;

同理,在进行A阶段、B阶段、C阶段和D阶段四分类任务时,将A阶段作为第一大类,将B阶段、C阶段和D阶段作为第二大类,使用第一个二分类支持向量机分类输出A阶段和第二大类;然后将B作为第三大类,将C和D作为第四大类,使用第二个二分类支持向量机对第二大类进行分类,输出B阶段和第四大类;最后使用第三个二分类支持向量机对第四大类进行分类,输出B阶段和C阶段;

以此类推,同样的原理完成五分类和六分类任务。

7.如权利要求1所述的基于熵特征及支持向量机的睡眠状态分类方法,其特征是,所述步骤(4)的具体步骤为:

待分类信号同样按照设定时间间隔进行均匀分割,分割后得到若干个片段,然后对每

个片段提取三个熵特征：模糊熵、模糊测度熵和样本熵；然后将三个熵特征均输入到训练好的支持向量机分类器中进行分类。

8. 如权利要求1所述的基于熵特征及支持向量机的睡眠状态分类方法，其特征是，所述步骤(2)之后，所述步骤(3)之前，还需要执行：

特征融合步骤：对模糊熵、模糊测度熵和样本熵特征分别进行归一化处理得到三个归一化的特征向量，将归一化后的特征向量分别赋予不同的权值，并串联成一个融合的高维特征向量，然后采用主成分分析PCA算法对高维特征向量进行降维处理，得到融合后的熵特征。

9. 基于熵特征及支持向量机的睡眠状态分类系统，其特征是，包括：存储器、处理器以及存储在存储器上并在处理器执行的计算机指令，所述计算机指令被处理器运行时，完成权利要求1-8任一项方法所述的步骤。

10. 一种计算机可读存储介质，其特征是，其上存储有计算机指令，所述计算机指令被处理器运行时，完成权利要求1-8任一项方法所述的步骤。

基于熵特征及支持向量机的睡眠状态分类方法及系统

技术领域

[0001] 本发明涉及睡眠分析研究,基于熵特征及支持向量机的睡眠状态分类方法及系统。

背景技术

[0002] 睡眠是一个复杂的生理过程,睡眠质量可以反映身体和精神的健康状况。随着现代生活压力的加大,睡眠呼吸暂停,失眠和嗜睡症已经成为普遍存在的问题并逐渐引起人们的广泛关注,有效地识别睡眠状态对分析和监测睡眠质量具有极大地帮助。多导睡眠图技术(PSG)通常被专家用于睡眠障碍的诊断,按照Rechtschaffen标准、Kales(R&K)标准或AASM(American academy of sleep medicine)标准,由经验丰富的睡眠专家通过经验及视觉判断来完成。在本发明中,我们采用R&K作为睡眠状态的分期标准,其总共包含6个睡眠状态,分别为清醒阶段AWA(Awake),阶段一S1,阶段二S2,阶段三S3,阶段四S4和快速眼动期REM(Rapid Eye Movement)。

[0003] 随着睡眠可穿戴设备的快速发展,对快速准确地睡眠状态分类提出了更高的要求,专家人工分期已不现实且容易出错,所以这种通过专家人工对睡眠阶段进行分期的方法逐渐被淘汰。随着近些年不断有人提出不同的睡眠阶段分期方法,如何实现睡眠阶段的快速准确自动分期是当今一个热门的话题。现阶段大多数睡眠阶段分期方法主要是采用脑电EEG、眼电EOG或肌电EMG信号的时域或频域特征。

[0004] 2010年,Chapotot F and Becq G发表了基于人工神经网络分类器设计了一种分类框架,对脑电EEG和肌电EMG信号的光谱和非线性候选特征进行提取,但是,其仅仅能够实现对睡眠或觉醒两个阶段的分类。

[0005] 2011年,KrakovskáA and MezeiováK采用脑电EEG、眼电EOG或肌电EMG信号的频率特征,其对睡眠的五个阶段分类准确度仅仅达到74.5%。

[0006] 现有技术需要解决的技术问题是:现有的睡眠状态分类方法的分类精确度偏低,无法用于临床。

发明内容

[0007] 为了解决现有技术的不足,本发明提供了基于熵特征及支持向量机的睡眠状态分类方法及系统;

[0008] 作为本发明的第一方面,提供了一种基于熵特征及支持向量机的睡眠状态分类方法;

[0009] 基于熵特征及支持向量机的睡眠状态分类方法,包括:

[0010] 步骤(1):数据预处理:

[0011] 采集脑电信号Fpz-Cz EEG,脑电信号Pz-Oz EEG和水平眼电图信号EOG;将采集的信号均按照设定时间间隔进行均匀分割,分割后得到若干个片段;

[0012] 每个片段都有预先标记好的睡眠阶段,所述睡眠阶段包括:清醒阶段AWA、阶段一

S1、阶段二S2、阶段三S3、阶段四S4和快速眼动期REM；

[0013] 步骤(2)：特征提取：从每个片段中提取三个熵特征：模糊熵、模糊测度熵(Fuzzy measure entropy)和样本熵；

[0014] 步骤(3)：设计一对多支持向量机分类器，将提取的熵特征和预先标记好的睡眠阶段输入到一对多支持向量机分类器中，对一对多支持向量机分类器进行训练；

[0015] 步骤(4)：利用训练好的支持向量机分类器对待分类信号进行分类。

[0016] 作为本发明的进一步改进，所述模糊测度熵的计算过程如下：

[0017] (a) 假设x为脑电信号Fpz-Cz EEG、脑电信号Pz-Oz EEG或者眼电信号EOG，对信号按照设定时间间隔进行分割，得到若干个时间序列，对于每一个时间序列 $\{x(i) : 1 \leq i \leq N\}$ ，通过重构序列生成m维局部序列段 X_i^m ：

$$[0018] \quad X_i^m = \{x(i), x(i+1), \dots, x(i+m-1)\} - x_0(i) \quad (1)$$

[0019] 在(1)中， $i=1, \dots, N-m$ ， $\{x(i), x(i+1), \dots, x(i+m-1)\}$ 表示从第i个时间序列开始的连续的m个时间序列， $x_0(i)$ 表示m个时间序列的平均值；

[0020] 计算对应信号的局部序列段 $XL_m(i) = X_i^m$ ；

[0021] 通过去除整个信号 $\{x(i) : 1 \leq i \leq N\}$ 的平均值来定义全局序列段，

[0022] 即 $XF_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\} - \bar{x}$ ，

[0023] 其中， \bar{x} 表示时间序列 $\{x(i) : 1 \leq i \leq N\}$ 的平均值， $x(i)$ 表示信号每间隔30秒分割之后得到的第i个信号片段，N表示信号分割之后得到的总片段数；

[0024] (b) 计算信号的第i个局部序列段 $XL_m(i)$ 与第j个局部序列段 $XL_m(j)$ 的距离 $dL_m(i, j)$ ，如下：

$$[0025] \quad dL_m(i, j) = d[XL_m(i), XL_m(j)] = d_{ij}^m \quad (2)$$

[0026] 计算信号的第i个全局序列段 $XF_m(i)$ 和信号的第j个全局序列段 $XF_m(j)$ 之间的距离 $dF_m(i, j)$ ， $i, j=1 \sim N-m$ ， $j \neq i$ ，如下：

$$[0027] \quad dF_m(i, j) = d[XF_m(i), XF_m(j)] = \max_{k \in (0, m-1)} \{|x(i+k) - \bar{x} - (x(j+k) - \bar{x})|\} \quad (3)$$

[0028] (c) 然后，计算信号的局部相似度 $DL_m(i, j)$ 和全局相似度 $DF_m(i, j)$ 如下：

$$[0029] \quad DL_m(i, j) = \mu(dL_m(i, j), n, r) = \exp(-(dL_m(i, j)/r)^n) = D_{ij}^m \quad (4)$$

$$[0030] \quad DF_m(i, j) = \mu(dF_m(i, j), n, r) = \exp(-(dF_m(i, j)/r)^n) \quad (5)$$

[0031] (d) 定义信号的局部相似度 $DL_m(i, j)$ 的平均值 $\phi L_m(n, r)$ 和全局相似度 $DF_m(i, j)$ 的平均值 $\phi F_m(n, r)$ ，如下：

$$[0032] \quad \phi L_m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DL_m(i, j) \right) = \phi^m(n, r) \quad (6)$$

$$[0033] \quad \phi F_m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DF_m(i, j) \right) \quad (7)$$

[0034] (e) 重复(a)到(d)的步骤，计算信号的局部相似度 $DL_{m+1}(i, j)$ 的平均值 $\phi L_{m+1}(n, r)$ 和全局相似度 $DF_{m+1}(i, j)$ 的平均值 $\phi F_{m+1}(n, r)$ ；

[0035] (f) 计算信号的局部模糊测度熵 $FuzzyLME_n(m, n, r, N)$ 和全局模糊测度熵 $FuzzyFME_n(m, n, r, N)$ ：

[0036] $FuzzyLMEn(m, n, r, N) = \ln \phi L_m(n, r) - \ln \phi L_{m+1}(n, r)$ (8)

[0037] $FuzzyFMEEn(m, n, r, N) = \ln \phi F_m(n, r) - \ln \phi F_{m+1}(n, r)$ (9)

[0038] (g) 最后,信号的模糊测度熵FuzzyMEn(m,n,r,N)计算如下:

[0039] $FuzzyMEn(m, n, r, N) = FuzzyLMEn(m, n, r, N) + FuzzyFMEEn(m, n, r, N)$ (10)

[0040] 与模糊熵的计算步骤相比,模糊测度熵集成了时间序列的局部特性和全局特性,可以反映脑电或者眼电信号的整体复杂度。

[0041] 作为本发明的进一步改进,所述步骤(2)之后,所述步骤(3)之前,还需要执行:

[0042] 特征融合步骤:对模糊熵、模糊测度熵和样本熵特征分别进行归一化处理得到三个归一化的特征向量,将归一化后的特征向量分别赋予不同的权值,并串联成一个融合的高维特征向量,然后采用主成分分析PCA算法对高维特征向量进行降维处理,得到融合后的熵特征。

[0043] 作为本发明的进一步改进,所述设计一对多支持向量机分类器的具体步骤为:

[0044] 如果是完成睡眠阶段的二分类任务,则使用一个二分类支持向量机;

[0045] 如果是完成睡眠阶段的三分类任务,则使用两个串联的二分类支持向量机;串联时,在后的二分类支持向量机的输入端与在前的二分类支持向量机的其中一个输出端连接;在前的二分类支持向量机的另外一个输出端和在后的二分类支持向量机的输出端均用于输出分类结果;

[0046] 如果是完成睡眠阶段的四分类任务,则使用三个串联的二分类支持向量机,串联时,在后的二分类支持向量机的输入端均与在前的二分类支持向量机的其中一个输出端连接;三个串联的二分类支持向量机的其他输出端均用于输出分类结果;

[0047] 如果是完成睡眠阶段的五分类任务,则使用四个串联的二分类支持向量机;串联时,在后的二分类支持向量机的输入端均与在前的二分类支持向量机的其中一个输出端连接;四个串联的二分类支持向量机的其他输出端均用于输出分类结果;

[0048] 如果是完成睡眠阶段的六分类任务,则使用五个串联的二分类支持向量机;串联时,在后的二分类支持向量机的输入端均与在前的二分类支持向量机的其中一个输出端连接;五个串联的二分类支持向量机的其他输出端均用于输出分类结果。

[0049] 作为本发明的进一步改进,所述步骤(3)的具体步骤为:

[0050] 步骤(301):统计训练睡眠标记数据中处于每个睡眠阶段的片段数;

[0051] 步骤(302):按照片段数由多到少顺序对睡眠阶段进行排序;片段数排序第一的睡眠阶段设为A阶段;片段数排序第二的睡眠阶段为B阶段;片段数排序第三的睡眠阶段为C阶段;片段数排序第四的睡眠阶段为D阶段;片段数排序第五的为E阶段;

[0052] 步骤(303):对分类任务进行分析,对一对多支持向量机进行选择:

[0053] 如果分类任务是区分睡眠阶段和清醒阶段AWA,则采用一个二分类支持向量机;

[0054] 如果分类任务是区分睡眠阶段、快速眼动期REM和非快速眼动期(包括阶段一S1,阶段二S2,阶段三S3和阶段四S4),则采用两个串联的二分类支持向量机;

[0055] 如果分类任务是区分清醒阶段AWA、第一非快速眼动期(包括阶段一S1和阶段二S2)、第二非快速眼动期(包括阶段三S3和阶段四S4)和快速眼动期REM,则采用三个串联的二分类支持向量机;

[0056] 如果分类任务是区分清醒阶段AWA、阶段一S1、阶段二S2、第三非快速眼动期(包括

阶段三S3、阶段四S4)和快速眼动期REM,则采用四个串联的二分类支持向量机;

[0057] 如果分类任务是区分清醒阶段AWA、阶段一S1、阶段二S2、阶段三S3、阶段四S4和快速眼动期REM,则采用五个串联的二分类支持向量机;

[0058] 步骤(304):利用选择好的一对多支持向量机分类器进行分类;

[0059] 每个二分类支持向量机进行分类都是从睡眠片段数最多的睡眠阶段开始分类:将排序靠前的一类睡眠阶段归为一类,同时将其他剩余的睡眠阶段归为另一类;然后,针对其他剩余的睡眠阶段,再将排序靠前的一类睡眠归为一类,再将其他剩余的睡眠阶段归为另一类;以此类推,直至所有的睡眠阶段归类完毕,结束。

[0060] 作为本发明的进一步改进,所述步骤(304)的具体步骤为:

[0061] 如果是完成睡眠状态阶段的六分类任务,则使用五个串联的二分类支持向量机,所述五个串联的二分类支持向量机依次为:第一二分类支持向量机、第二二分类支持向量机、第三二分类支持向量机、第四二分类支持向量机和第五二分类支持向量机;

[0062] 首先,将所有睡眠阶段的片段数均输入到第一二分类支持向量机中;输出值有两个分支,第一个分支的输出仅仅包括A阶段的片段数,第二个分支的输出包括B阶段、C阶段、D阶段和E阶段的片段数;

[0063] 然后,将第一二分类支持向量机的第二个分支的输出值输入到第二二分类支持向量机中;第二二分类支持向量机的输出值有两个分支,第一个分支的输出仅仅包括B阶段的片段数,第二个分支的输出包括C阶段、D阶段和E阶段的片段数;

[0064] 然后,将第二二分类支持向量机的第二个分支的输出值输入到第三二分类支持向量机中;第三二分类支持向量机的输出值有两个分支,第一个分支的输出仅仅包括C阶段的片段数;第二个分支的输出包括D阶段和E阶段的片段数;

[0065] 然后,将第三二分类支持向量机的第二个分支的输出值输入到第四二分类支持向量机中;第四二分类支持向量机的输出值有两个分支,第一个分支的输出仅仅包括D阶段的片段数;第二个分支的输出包括E阶段的片段数;至此,一对多支持向量机训练结束。

[0066] 作为本发明的进一步改进,所述步骤(304)的具体步骤为:

[0067] 在进行三分类(如A阶段、B阶段和C阶段)时,先将A阶段作为第一大类,将B阶段和C阶段作为第二大类,首先使用第一个二分类支持向量机进行分类分别输出A阶段和第二大类;然后使用第二个二分类支持向量机对第二大类(也就是剩余的B阶段和C阶段)进行分类,输出B阶段和C阶段。

[0068] 同理,在进行四分类(如A阶段、B阶段、C阶段和D阶段)任务时,将A阶段作为第一大类,将B阶段、C阶段和D阶段作为第二大类,使用第一个二分类支持向量机分类输出A阶段和第二大类;然后将B作为第三大类,将C和D作为第四大类,使用第二个二分类支持向量机对第二大类进行分类,输出B阶段和第四大类;最后使用第三个二分类支持向量机对第四大类(也就是剩余的B阶段和C阶段)进行分类,输出B阶段和C阶段。

[0069] 以此类推,同样的原理完成五分类和六分类任务。

[0070] 上述技术每个二分类支持向量机进行分类都是将排序靠前的一类睡眠阶段归为一类,同时将其他剩余的睡眠阶段归为另一类,并从睡眠片段数多的睡眠阶段开始分类,有益效果是:节省了分类的时间,提高了分类的精度;节省分类时间的原理是:现有技术中支持向量机不是串联的,如果是六分类任务,则先利用第一个支持向量机完成A阶段与B阶段

的分类,然后再用第二个支持向量机完成A阶段与C阶段的分类,然后再用第三个支持向量机完成A阶段与D阶段的分类,然后,在利用第四个支持向量机完成A阶段与E阶段的分类。与本发明相比,现有技术的这种方式浪费时间,不如本申请分类速度快,且分类精度高。

[0071] 作为本发明的进一步改进,所述步骤(4)的具体步骤为:

[0072] 待分类信号同样按照设定时间间隔进行均匀分割,分割后得到若干个片段,然后对每个片段提取三个熵特征:模糊熵、模糊测度熵和样本熵;然后将三个熵特征均输入到训练好的支持向量机分类器中进行分类。

[0073] 作为本发明的第二方面,提供了基于熵特征及支持向量机的睡眠状态分类系统;

[0074] 基于熵特征及支持向量机的睡眠状态分类系统,包括:存储器、处理器以及存储在存储器上并在处理器执行的计算机指令,所述计算机指令被处理器运行时,完成上述任一方法所述的步骤。

[0075] 作为本发明的第三方面,提供了一种计算机可读存储介质;

[0076] 一种计算机可读存储介质,其上存储有计算机指令,所述计算机指令被处理器运行时,完成上述任一方法所述的步骤。

[0077] 与现有技术相比,本发明的有益效果是:

[0078] 本发明提出了一种新的睡眠阶段分类方法,通过提取脑电和眼电信号的模糊熵(FuzzyEn),模糊测度熵(FuzzyMEn)和样本熵(SampEn)及特征融合的理论,采用基于一对多的多分类支持向量机进行分类。统计配对样本T检验显示,应用的熵特征在每个睡眠阶段具有显著差异。本发明给出了2-6种状态的睡眠状态分类结果,并利用每个睡眠阶段的分类灵敏度和精度以及总体准确率和相应的Kappa系数来评估分类性能,结果表明该方法与现有的睡眠状态分类方法相比具有显著的优越性。

附图说明

[0079] 构成本申请的一部分的说明书附图用来提供对本申请的进一步理解,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。

[0080] 图1为本发明的流程图;

[0081] 图2(a)-图2(1)为样本sc4002e0的各个睡眠阶段的30s信号时长Fpz-Cz EEG和EOG信号示例;

[0082] 图3(a)-图3(c)为三通道信号不同睡眠状态下模糊熵、模糊测度熵和样本熵的箱形图;

[0083] 图4为受试者sc4002e0的睡眠状态分类,准确率为86.86%,Kappa系数为0.7424。从上到下分别是:专家注释,预测睡眠阶段,专家注释与相应预测标签之间不匹配的片段;

[0084] 图5为受试者st7121j0的睡眠阶段分类的准确率和Kappa系数分别为70.96%和0.5665。从上到下:专家注释,预测睡眠阶段,专家注释与相应预测标签之间不匹配的片段。

具体实施方式

[0085] 应该指出,以下详细说明都是例示性的,旨在对本申请提供进一步的说明。除非另有指明,本发明使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解的含义。

[0086] 需要注意的是,这里所使用的术语仅是为了描述具体实施方式,而非意图限制根据本申请的示例性实施方式。如在这里所使用的,除非上下文另外明确指出,否则单数形式也意图包括复数形式,此外,还应当理解的是,当在本说明书中使用术语“包含”和/或“包括”时,其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

[0087] 目标:睡眠质量可以反映一个人的身体及精神状态,有效的睡眠状态分类对于健康医疗具有重要的作用。本发明旨在提出一种基于熵特征以及支持向量机的简单有效的睡眠状态分类算法SCEF-SVM(Sleep stages Classification based on Entropy Features and Support Vector Machine classifier)。

[0088] 方法:本发明首次将模糊测度熵与模糊熵和样本熵结合应用于睡眠状态的分析 and 分类。我们选用Sleep-EDF数据库中的三路信号,包括Fpz-Cz EEG,Pz-Oz EEG和水平眼电图EOG,并将所有信号每30s分割共得到15166个片段,从中提取睡眠信号的熵特征。配对T检验结果显示,本研究所选用的三种熵值特征在6个睡眠阶段之间存在显著性差异,其中6个睡眠阶段包括清醒(AWA),阶段1(S1),阶段2(S2),阶段3(S3),阶段4(S4)和快速眼动期(REM)。我们设计了一种新型的多分类支持向量机(SVM),采用一对多的分类原则对提取的特征进行分类。本发明设计了三种不同的实验来测试SCEF-SVM的有效性和可靠性。

[0089] 主要结果:结果显示SCEF-SVM与当前最先进的分类方法相比具有显著的优越性,其2分类,3分类,4分类,5分类,6分类的分类准确率分别为98.19%,93.08%,91.07%,89.86%和88.49%,对应的Kappa系数同时保证了几乎完美的分类一致性。另外,我们还进行了5分类的10折交叉检验,得到的平均准确率和Kappa系数分别为90.18%和0.8485。

[0090] 意义:与最先进的分类方法相比,我们所提出的SCEF-SVM被证明是一种简单有效且可靠的睡眠分类方法。我们相信SCEF-SVM的提出将会进一步推动睡眠状态分类的发展。

[0091] 如图1所示,基于熵特征及支持向量机的睡眠状态分类方法,包括:

[0092] 步骤(1):数据预处理:

[0093] 在用于研究睡眠分期的国际标准数据库Sleep-EDF数据库的睡眠监测数据中选出八个样本,从每个样本中选出三个通道的信号:脑电信号Fpz-Cz EEG,脑电信号Pz-Oz EEG和水平眼电图信号EOG;将每个通道的信号均按照设定时间间隔进行均匀分割,分割后得到若干个片段;

[0094] 每个片段都设有预先标记好的睡眠阶段,所述睡眠阶段包括:清醒阶段AWA、阶段一S1、阶段二S2、阶段三S3、阶段四S4和快速眼动期REM;

[0095] 步骤(2):特征提取:从每个片段中提取三个熵特征:模糊熵、模糊测度熵和样本熵;

[0096] 步骤(3):搭建一对多支持向量机分类器,利用提取的熵特征和预先标记好的睡眠阶段,对支持向量机分类器进行训练;

[0097] 步骤(4):利用训练好的支持向量机分类器对待分类信号进行分类。

[0098] 所述睡眠监测数据包括:样本编号、样本被监测的开始时间、样本被监测的结束时间、样本被监测的通道和样本被监测通道的采样频率。

[0099] 样本被监测的通道,包括:脑电信号Fpz-Cz EEG,脑电信号Pz-Oz EEG、水平眼电图信号EOG、颞肌电信号包络、口鼻气流、直肠体温以及事件标记信号。

[0100] 所述脑电信号Fpz-Cz EEG,脑电信号Pz-Oz EEG和水平眼电图信号EOG的采样频率

均为100Hz。

[0101] 1、实验数据

[0102] 本发明中使用的实验数据来自Sleep-EDF数据库,该数据库包含8个健康男性和女性在没有服用任何药物情况下的睡眠监测数据。其中四个sc*记录(即sc4002e0,sc4012eo,sc4102e0,sc4112e0)中的每一个样本记录了七个通道的信号,包括Fpz-Cz EEG,Pz-Oz EEG和水平EOG,它们的采样频率均为100Hz,另外还包括四个通道的亚颞肌电信号包络,口鼻气流,直肠体温以及事件标记信号,它们的采样频率均为1Hz。除此之外,四个st*记录(即st7022j0,st7052j0,st7121j0and st7132j0)中的每一个样本记录了五个通道的信号,包括以100Hz采样的Fpz-Cz EEG,Pz-Oz EEG,水平EOG和颞肌电信号以及以1Hz采样的事件标记信号。所有记录都以EDF格式存储,每个PSG由训练有素的技术人员根据R&K手册进行睡眠状态标注。该数据库中将W,S1,S2,S3,S4,REM,MVT和'未评分'的睡眠阶段分别编码为0,1,2,3,4,5,6和9。另外需要注意的是,四个sc*序列记录了24小时的生理监测数据,而st*序列则只记录了夜间数据。

[0103] 为了便于发明的推广,在这项发明中,我们选择Fpz-Oz EEG,Pz-Oz EEG和水平EOG三路信号来分析和识别睡眠状态;这三个通道均以100Hz采样,为每个睡眠阶段提供相对详细的生理信息。图2给出了样本sc4002e0六个睡眠阶段30s的Fpz-Cz EEG和EOG信号的示例。另外需要说明的是,四个sc*样本的信号记录质量比较高,不存在缺失的信号,因此全部信号均被采集并应用于本项研究中。相比之下,四个st*序列中则存在少量无效信号,在进一步处理之前我们需要将这部分无效信号检测出来并去除掉。表1列出了8个记录的详细信息。

[0104] 表1:本研究中使用的数据库的详细信息

[0105]

序列	记录	开始时间	结束时间(第二天)	通道	采样频率
sc* series	sc4002e0	14:50:00	14:25:30	Fpz-Cz EEG	100 Hz
	sc4012eo	17:00:00	16:45:30	Pz-Oz EEG 水平 EOG	
	sc4102e0	15:56:00	15:45:30	EMG	1 Hz
	sc4112e0	15:20:00	14:30:30	体温 呼吸 事件标记	

[0106]

st* series	st7022j0	20:57:00	8:00:30	Fpz-Cz EEG	100 Hz
	st7052j0	21:45:00	8:30:30	Pz-Oz EEG	
	st7121j0	22:00:00	8:22:00	水平 EOG EMG	
	st7132j0	22:22:00	8:01:00	事件标记	1 Hz

[0107] 2、方法

[0108] 对于一个时间序列,产生新模式的概率越大意味着其复杂性越大,且熵被广泛用于评估时间序列的复杂度。换句话说,熵可以用来测量信号的复杂度或不规则性,且信号越复杂,其相应的熵值也就越大。例如,受试者在清醒状态下记录的脑电或眼电信号的熵明显大于受试者处于睡眠状态下记录的信号熵值,因为受试者在清醒状态下大脑和眼球的活跃程度明显高于睡眠状态,而当睡眠状态逐渐深入到非快速眼动期时,大脑的活跃性相应减弱,随着睡眠程度进一步加深,其对外部刺激的响应进一步减小,相应地,脑电和眼电信号的熵值也将进一步减小。因此,将熵作为特征进行分类的思想在理论上是可行的。

[0109] 3、特征提取

[0110] (1) 模糊熵

[0111] 模糊熵表征的是一个模糊集的熵,表示一个序列不确定性信息。模糊熵的定义和计算过程如下:

[0112] (a) 对于一个时间序列 $\{x(i) : 1 \leq i \leq N\}$, 如公式 (1) 所示,通过重构序列生成 m 维局部序列段 X_i^m 。

[0113] (b) d_{ij}^m 定义为 X_i^m 和 X_j^m 之间的最大距离:

$$d_{ij}^m = d[X_i^m, X_j^m] = \max_{k \in (0, m-1)} \{|x(i+k) - x_0(i) - (x(j+k) - x_0(j))|\} \quad (11)$$

$(i, j = 1 \sim N - m, j \neq i)$

[0115] (c) D_{ij}^m 被定义为基于模糊函数 $\mu(d_{ij}^m, n, r)$ 的 X_i^m 和 X_j^m 的相似度:

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = \exp(-(d_{ij}^m)^n / r) \quad (12)$$

[0117] 其中, n 和 r 分别是指数函数边界的梯度和宽度。

[0118] (d) 定义 D_{ij}^m 的平均值为:

$$\phi^m(n, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \left(\frac{1}{N - m - 1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \right) \quad (13)$$

[0120] (e) 重复从 (a) 到 (d) 的过程,通过重构序列生成一个 $m+1$ 维矢量,并定义该函数:

$$\phi^{m+1}(n, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \left(\frac{1}{N - m - 1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^{m+1} \right) \quad (14)$$

[0122] (f) 那么模糊熵定义如下:

$$\text{FuzzyEn}(m, n, r) = \lim_{N \rightarrow \infty} [\ln \phi^m(n, r) - \ln \phi^{m+1}(n, r)] \quad (15)$$

[0124] 当 N 有限大时,模糊熵被定义为以下近似值:

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

[0126] (2) 样本熵

[0127] 样本熵是对近似熵的改进,用于评估生理时间序列信号的复杂度并诊断患病状态。它有两个优点:数据长度独立性和实现的可靠性。计算过程比较简单,如下所示。

[0128] 考虑相同的时间序列 $\{x(i) : 1 \leq i \leq N\}$, 按照 (1) 和 (11) 计算相应的 X_i^m 和 d_{ij}^m , 则样本熵表示为:

$$\text{SampEn} = -\log \frac{A}{B} \quad (17)$$

[0130] 在(17)中,A表示满足 $d_{ij}^{m+1} < r$ 的模板矢量对的数量,B表示满足 $d_{ij}^m < r$ 的模板矢量对的数量,其中 r 为 $0.15 \times \text{std}$, std 表示时间序列的标准差。所述时间序列是:脑电信号或眼电信号每隔30秒分割得到的一组信号片段。

[0131] (3) 模糊测度熵

[0132] 模糊测度熵是基于模糊集理论计算的,可以提高近似熵和样本熵的统计稳定性。模糊测度熵算法受Chen等人的研究启发,并在其中引入了模糊集来提高统计稳定性。

[0133] 本发明中,对应的脑电信号和眼电信号的模糊测度熵的计算过程如下:

[0134] (a) 假设 x 为脑电信号Fpz-Cz EEG或者脑电信号Pz-Oz EEG或者眼电信号EOG,计算对应脑电或者眼电信号的局部序列段 $XL_m(i) = X_i^m$ 。另外,通过去除整个 n 脑电或者眼电信号 $\{x(i) : 1 \leq i \leq N\}$ 的平均值来定义全局序列段,即 $XF_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\} - \bar{x}$,其中 \bar{x} 表示整个脑电或者眼电序列 $\{x(i) : 1 \leq i \leq N\}$ 的平均值。其中, $x(i)$ 表示脑电或者眼电信号中每30秒平均分割之后得到的第 i 个信号片段, N 表示信号分割之后得到的总片段数。

[0135] (b) 计算脑电或者眼电信号局部序列段 $XL_m(i)$ 和 $XL_m(j)$ 的距离,如下:

$$[0136] \quad dL_m(i, j) = d[XL_m(i), XL_m(j)] = d_{ij}^m \quad (2)$$

[0137] 另外,对应的脑电或者眼电全局序列段 $XF_m(i)$ 和 $XF_m(j)$ 之间的距离同时计算如下:

$$[0138] \quad dF_m(i, j) = d[XF_m(i), XF_m(j)] = \max_{k \in (0, m-1)} \{|x(i+k) - \bar{x} - (x(j+k) - \bar{x})|\} \\ (i, j = 1 \sim N-m, j \neq i) \quad (3)$$

[0139] (c) 然后,计算脑电或者眼电信号的局部相似度和全局相似度如下:

$$[0140] \quad DL_m(i, j) = \mu(dL_m(i, j), n, r) = \exp(-(dL_m(i, j)/r)^n) = D_{ij}^m \quad (4)$$

$$[0141] \quad DF_m(i, j) = \mu(dF_m(i, j), n, r) = \exp(-(dF_m(i, j)/r)^n) \quad (5)$$

[0142] (d) 定义上述 $DL_m(i, j)$ 和 $DF_m(i, j)$ 的平均值如下:

$$[0143] \quad \phi L_m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DL_m(i, j) \right) = \phi^m(n, r) \quad (6)$$

$$[0144] \quad \phi F_m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DF_m(i, j) \right) \quad (7)$$

[0145] (e) 重复(a)到(e)的步骤,计算 $\phi L_{m+1}(n, r)$ 和 $\phi F_{m+1}(n, r)$ 。

[0146] (f) 那么脑电或者眼电信号的局部模糊测度熵和全局模糊测度熵定义如下:

$$[0147] \quad \text{FuzzyLMEn}(m, n, r, N) = \ln \phi L_m(n, r) - \ln \phi L_{m+1}(n, r) \quad (8)$$

$$[0148] \quad \text{FuzzyFMEn}(m, n, r, N) = \ln \phi F_m(n, r) - \ln \phi F_{m+1}(n, r) \quad (9)$$

[0149] (g) 最后,整个脑电或者眼电信号的模糊度量熵计算如下:

$$[0150] \quad \text{FuzzyMEn}(m, n, r, N) = \text{FuzzyLMEn}(m, n, r, N) + \text{FuzzyFMEn}(m, n, r, N) \quad (10)$$

[0151] 与模糊熵的计算步骤相比,模糊测度熵集成了时间序列的局部特性和全局特性,可以反映脑电或者眼电信号的整体复杂度。

[0152] 所述步骤(2)之后,所述步骤(3)之前,还需要执行:

[0153] 特征融合步骤:对模糊熵、模糊测度熵和样本熵特征分别进行归一化处理得到三个归一化的特征向量,将归一化后的特征向量分别赋予不同的权值,并串联成一个融合的高维特征向量,然后采用主成分分析PCA算法对高维特征向量进行降维处理,得到融合后的熵特征。

[0154] 本发明选用一对多分类支持向量机作为分类器,本发明尝试了多种类型的支持向量机,比如LIBSVM,最后采用基于一对多的支持向量机分类器。对于多种不同类型的核函数,包括线性,多项式,径向基函数和sigmoid,本发明采用RBF作为核函数。根据我们已经进行的实验,提高总体准确性的一个技巧是每次从包含更多片段的睡眠阶段开始分类。

[0155] 5、实验和结果

[0156] 为了评估SCEF-SVM的分类性能,本发明使用MATLAB 2015a在Dell计算机上进行仿真实验,该计算机带有3.40GHz Intel Core i7-2600CPU和16.0GB RAM。实验包括三个主要部分:数据预处理,特征提取和SVM分类。为了测试SCEF-SVM的鲁棒,我们同时进行了5分类的10倍交叉检验。每个部分的实验设置和结果如下。

[0157] 5.1数据预处理

[0158] 本发明应用了三个通道数据,即Fpz-Cz EEG,Pz-Oz EEG和水平EOG。本发明主要考虑AWA,S1,S2,S3,S4和REM之间的分类,首先检测并去除四个st*序列中的无效信号并使用通带为0.5至30Hz的四阶巴特沃斯滤波器对原始信号进行滤波。由于专家每30秒对睡眠信号进行一次标注,所以数据每隔30秒分为一个片段,并顺序标号,每个片段包含3000个数据点。分类时,将标号为奇数的片段作为训练集,并将标号为偶数的片段作为测试集。表2中给出了每个睡眠阶段的片段数量。从表中可以看出,片段总数为15166个,训练集和测试集的片段数大致相同。此外,从表中可以看出AWA是整个记录的主要部分,占整个睡眠分期的一半以上,此外,S2则是人们处于睡眠状态中的主要阶段。在本研究中,按照睡眠片段数由多到少排列顺序为:AWA>S2>REM>S3>S4>S1,因此,在进行六分类的过程中,首先需要将AWA作为一类,同时将S2,REM,S3,S4和S1作为一类使用第一个二分类进行分类。然后将S2作为一类,将REM,S3,S4和S1作为另一类使用第二个串联的二分类支持向量机进行分类,以此类推,完成六分类的任务。

[0159] 表2:训练集和测试集中每个睡眠阶段的片段数目

睡眠分期	标签	训练集	测试集
AWA	0	4018	4015
S1	1	300	304
S2	2	1813	1808
S3	3	331	341
S4	4	315	312
REM	5	806	803
片段总数		7583	7583

[0160] 5.2特征提取

[0162] 如5.1节所示,数据每30秒分为一个片段,所有8个样本共有15166个片段。在数据预处理之后,计算三个通道每个片段的模糊熵,模糊测度熵和样本熵。对于模糊熵,嵌入维数 $m=2$,阈值 $r=0.15*\text{std}$ (std是信号的标准偏差),模糊因子Fuzzy Power $n=2$,时间滞后tau设置为1。对于模糊测度熵,相应的参数 $m=2$,r也为 $0.15*\text{std}$,对于样本熵也有相同的参数设置。

[0163] 图3(a)-图3(c)给出了三个熵的箱形图,从中可以清楚地看出熵值随着睡眠时间的增长而变化,清醒阶段的熵一般高于所有其他睡眠阶段。两个脑电信号的三个不同熵值具有相似的变化趋势,即随着睡眠从S1加深到S4,熵值逐渐减小,直到REM状态呈上升趋势。此外,模糊测度熵和样本熵在这些不同睡眠阶段之间表现出比模糊熵更大的差异,表明这两个特征更有利于实现睡眠状态分类。相反,眼电信号的熵则表现出不同的趋势,即S2阶段的熵高于其他睡眠阶段,而REM阶段的熵最低。

[0164] 表3给出了熵特征配对样本T检验的结果,其中给出了所有可能配对之间的相关性和差异性检验。结果显示模糊测度熵和样本熵对所有可能的配对都具有显著性差异,而模糊熵的配对样本T检验结果表明REM-S2,S1-S2,S2-S3,S2-S4和S3-S4之间没有显著性差异,因此模糊测度熵和样本熵的分类性能更好。同时,表3中也给出了所有配对之间的相关性,并且三个熵特征都显示出S1和S2之间以及S2和S3之间有较少的相关性。

[0165] 表3:熵值特征配对样本T检验

配对	模糊熵		模糊测度熵		样本熵	
	Cor.sig	T.sig	Cor.sig	T.sig	Cor.sig	T.sig
AWA-S1	.000	.000	.627 [#]	.000	.416 [#]	.000
AWA-S2	.000	.000	.000	.000	.000	.000
AWA-S3	.153 [#]	.000	.001	.000	.017	.000
AWA-S4	.846 [#]	.000	.000	.000	.001	.000
AWA-REM	.000	.000	.000	.000	.054	.000
REM-S1	.451 [#]	.000	.003	.000	.000	.000
REM-S2	.004	.140 [*]	.000	.000	.000	.000
REM-S3	.088	.000	.010	.000	.000	.000
REM-S4	.000	.000	.046	.000	.000	.000
S1-S2	.456 [#]	.109 [*]	.823 [#]	.000	.860 [#]	.000
S1-S3	.000	.000	.121 [#]	.000	.000	.000
S1-S4	.000	.001	.000	.000	.000	.000
S2-S3	.249 [#]	.404 [*]	.840 [#]	.000	.122 [#]	.000
S2-S4	.000	.185 [*]	.951 [#]	.000	.622 [#]	.000
S3-S4	.000	.181 [*]	.000	.000	.000	.000

Cor.sig< 0.05 表示具有显著相关性; T.sig< 0.05 表示具有显著性差异
#: 没有显著相关性; *:没有显著性差异

[0168] 5.3 SVM分类

[0169] 本发明采用支持向量机进行分类,将分类准确率,灵敏度,精度和Kappa系数 κ 作为分类性能指标,用于评价SCEF-SVM的分类性能。其中, κ 值分别在0.00-0.20,0.21-0.40,0.41-0.60,0.61-0.80和0.81-1.00之间对应于轻微的,直接的,中等的,实质性的和几乎完美的一致性。为了呈现所提出的方法的详细分类性能,我们进行了C=2-6种睡眠阶段分类。

[0170] a、2分类

[0171] 对于2分类,AWA(清醒)标记为0而其他睡眠阶段(睡眠)标记为1。在2分类的结果中,4015个睡眠片段中有3928个被正确分类为清醒状态(标签0),同时3568个睡眠片段中有3518个被正确地分类为睡眠状态(标签1)。此外,87个片段被错误地分类为标签1,而50个片

段被错误分类为标签0。分类结果显示清醒状态(标签0)的分类精度为98.74%，睡眠状态(标签1)的分类精度为97.59%。同时，清醒状态(标签0)的灵敏度为97.83%，而睡眠状态(标签1)的灵敏度为98.60%。因此，两分类的总体分类准确率为98.19%，相应的Kappa系数为0.9638，对于区分睡眠和清醒两种状态SCEF-SVM具有几乎完美的准确度和一致性。

[0172] b、3分类

[0173] AWA被标记为0，非快速眼动期NREM(S1,S2,S3,S4)被标记为1，快速眼动期REM被标记为2。结果显示4015个睡眠片段中有3928个被正确分类为标签0，2765个睡眠片段中有2401个被正确地分类为标签1，并且803个睡眠片段中的729个被正确地分类为标签2。此外，70个AWA睡眠片段被错误分类为NREM并有17个被错误分类为REM，同时，49个NREM睡眠片段被错误地归类为AWA，另外315个被错误地归类为REM，而当涉及到REM睡眠时期时，1个被错误归类为AWA，另外73个被错误归类为NREM。分类结果显示AWA,NREM(S1,S2,S3,S4)和REM的分类精度分别为98.74%，94.38%和68.71%，相应的灵敏度分别为97.83%，86.84%和90.78%。总体分类准确率为93.08%，Kappa系数 κ 为0.8817，详见表5。

[0174] c、4分类

[0175] 对于4分类，AWA被标记为0，同时S1和S2被标记为1，而S3和S4被标记为2，NREM被标记为3。为了简单起见，这里没有给出每个睡眠阶段的正确分类以及错误分类数，但每个睡眠阶段相应的精确度，灵敏度以及总体准确度(91.07%)和Kappa系数(0.8578)可以在表5中看到。

[0176] d、5分类

[0177] 对于5分类，我们给出了详细的分类结果，表4给出了分类结果的混淆矩阵，其中对角线上的单元格对应正确分类的片段，其他单元是错误分类的片段数。表4还列出了每个睡眠阶段的相应精确度和灵敏度，以及总体准确度(89.86%)和Kappa系数(0.8425)。值得注意的是，S1阶段的精确度(50.85%)和灵敏度(39.14%)随着睡眠种类的逐渐增多而逐渐变差，这意味着S1阶段鉴别是一项具有挑战性的工作。

[0178] 表4:5分类混淆矩阵

[0179]

标注 预测	AWA	S1	S2	S3S4	REM	Sen(%) / Pre(%)
AWA	3928	30	10	9	1	97.83/98.74
S1	53	119	43	7	12	39.14/50.85
S2	18	54	1546	99	100	85.51/85.09
S3S4	3	5	101	532	1	81.47/82.87
REM	13	96	108	6	689	85.80/75.55

[0180] e、6分类

[0181] 本研究还进行了6种睡眠状态的分类，表5给出了6分类的性能指标，结果表明总体灵敏度，精确度，准确性(88.49%)和Kappa系数(0.8222)都逐渐变差但仍具有竞争力。

[0182] 表5:C=2-6类分类的每个阶段分类的灵敏度和精确度

[0183]

	C = 2	C = 3	C = 4	C = 5	C = 6	
	Sen(%) / Pre(%)	Sen(%) / Pre(%)	Sen(%) / Pre(%)	Sen(%) / Pre(%)	Sen(%) / Pre(%)	
AWA	97.83/98.74	97.83/98.74	97.83/98.74	97.83/98.74	97.83/98.74	
S1	98.60/97.59	86.84/94.38	82.62/86.26	39.14/50.85	38.16/52.97	
S2				85.51/85.09	85.51/85.09	
S3			84.84/78.92	81.47/82.87	57.48/52.41	
S4					75.32/83.04	
REM			90.78/68.71	84.56/77.16	85.80/75.55	85.80/75.55
准确率			98.19	93.08	91.07	89.86
κ	0.9638	0.8817	0.8578	0.8425	0.8222	

[0184] 注意: Sen: 灵敏度, Pre: 精度.

[0185] 6.1 分类10折交叉检验

[0186] 为了进一步测试SCEF-SVM的有效性,我们进行了10折交叉检验以评估5分类的平均准确度。在这项研究中我们总共进行了10次实验,每次实验中随机选择15166个片段中的90%作为训练数据,其余10%片段作为测试数据。结果如表6所示,10次实验的平均准确率为90.18%,Kappa系数为0.8485。我们应该注意到10折交叉检验的平均准确率与4.3d中得到的结果相比略有改善,这表明所提出的方法SCEF-SVM对于睡眠分类是高效的。

[0187] 表6: 5分类10折交叉检验

[0188]

	1	2	3	4	5	6	7	8	9	10	Average
Acc (%)	90.44	90.51	89.72	89.45	90.57	90.71	90.38	90.84	89.78	89.45	90.18
κ	0.8526	0.8529	0.8416	0.8377	0.8549	0.8566	0.8515	0.8581	0.8420	0.8367	0.8485

[0189] 6.2 本发明SCEF-SVM与当前最新的睡眠分类方法的准确性比较

[0190] 为了验证本发明SCEF-SVM的分类性能,表7给出了该方法与现有方法在进行2-6分类的准确性比较。在表7中提到了四种方法,包括模糊逻辑迭代系统,基于ANN算法的PSD (Power Spectral Density) 方法,DVG (Difference Visibility Graphs) 方法以及基于熵和频谱边缘频率特征的分类方法,均使用与本发明中相同的数据库。表7列出了该五种方法在使用Sleep-EDF数据库进行睡眠状态分类的准确性结果,可以发现其余四种方法的准确率均低于本发明中提出的SCEF-SVM得到的分类结果。其中,Berthomier等人使用自动睡眠分期系统在Sleep-EDF数据库中完成2-5分类的任务,结果显示其分类准确率均比本发明中的SCEF-SVM分类结果要低。另外,Ronzhina等人使用Welch方法来估计EEG信号的功率谱密度(PSD),并且将ANN架构应用于睡眠状态分类,由表7可以发现其2-6分类结果均比该发明中的结果要差。同时,Zhu等人将DVG (Difference Visibility Graphs) 应用到脑电信号的研究中,并从每个HVG (Horizontal Visibility Graphs) 和DVG中获得MD (Mean degree),并且提取了与睡眠脑电信号相关联的七个最优DD (Degree distribution) 来进行2-6分类。同样的,他们应用了与本发明中使用的相同的数据库以及相同的训练集和测试集构造方法,考虑到在他们的研究中仅使用单个脑电通道,因此其结果十分具有竞争力,但分类准确率仍比本发明中的方法要差。此外,Nakamura等人通过提取脑电信号的多尺度熵和谱边缘频

率特征,并采用一对一支持向量机来进行2-6分类,其分类结果同样显示在表7中。由表7可知,本发明中提出的SCEF-SVM方法在进行自动睡眠状态分类中与目前提出的方法相比具有十分明显的优越性。

[0191] 表7:准确率比较

[0192]

分类	模糊逻辑迭代系统	PSD 特征	DVG	熵值及边缘频率特征	SCEF-SVM
$C=2$	95.4%	96.90%	97.9%	97.40%	98.19%
$C=3$	88.3%	90.31%	92.6%	94.50%	95.12%
$C=4$	74.5%	84.05%	89.3%	91.00%	91.07%
$C=5$	71.2%	81.55%	88.9%	88.60%	89.86%
$C=6$	-	76.70%	87.5%	86.60%	88.49%

[0193] 6.3进一步核实

[0194] 为了进一步验证SCEF-SVM的有效性,我们针对5分类设计了不同的训练和测试实验。本发明进行了8次实验,每次实验选择8名受试者中的其中一名作为测试对象,其余7名构成训练集,目的是将训练数据和测试数据完全分开。5类分类的结果如表8所示,它表明:如果将训练数据和测试数据完全独立,那么分类的平均准确率(75.34%)和 κ 系数(0.6228)都会变差。

[0195] 表8:独立训练集和测试集

样本	片段	准确率	κ
[0196] sc4002e0	2828	86.86%	0.7424
sc4012eo	2847	82.65%	0.6805
sc4102e0	2856	73.74%	0.4858
sc4112e0	2779	91.80%	0.7982
st7022j0	944	64.85%	0.5924
[0197] st7052j0	1034	68.38%	0.5817
st7121j0	1026	70.96%	0.5665
st7132j0	852	63.50%	0.5352
平均	1895	75.34%	0.6228

[0198] 预测的睡眠阶段与专家注释相比较的例子在图4和图5中给出,其中最上面的对应于专家注释,中间的对应于预测标签,而底部的代表错误分类的片段。

[0199] 本发明提出了一种基于熵值特征以及一对多支持向量机的新型睡眠状态分类算法SCEF-SVM。在本发明中,应用一种全新的熵值特征,也就是模糊测度熵,来研究脑电信号以及眼电信号。通过使用模糊测度熵、模糊熵以及样本熵,本发明发现熵值特征在不同的睡眠阶段具有显著性差异,特别是在清醒状态时的熵值特征明显高于睡眠阶段,而且在睡眠

过程中不同阶段同样具有显著性差异。因此,本发明选择从两路脑电以及一路眼电中提取信号的熵值特征并进行2-6种状态的睡眠阶段分类。本发明选用Sleep-EDF数据库中的8例样本并将每路信号每30秒分段,共得到15166个信号片段。对于分类器的选择,本发明设计了一个新型的一对多支持向量机分类器,该分类方法由多个二分类支持向量机完成多分类的目标,并且在每个二分类中遵循一对多的分类原则。最终结果显示本发明提出的分类算法SCEF-SVM对于5分类能够得到89.86%的分类精度以及0.8425的Kappa系数。

[0200] 实验证明本发明所设计出来的SVM分类器相对于传统的多分类器来说,如LIBSVM,具有显著地优越性。而且,实验结果表明SCEF-SVM的分类精度是目前在使用国际标准数据库Sleep-EDF并且实验数据超过14000个片段的所有睡眠状态分类算法中最好的。除此之外,我们还进行了10折交叉检验,结果证明了该算法的有效性以及鲁棒性。因此,本发明提出的SCEF-SVM是一种简便并且有效的睡眠状态分类算法,本发明为相关产业的发展提供了新的方向和理论依据。

[0201] 以上所述仅为本申请的优选实施例而已,并不用于限制本申请,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

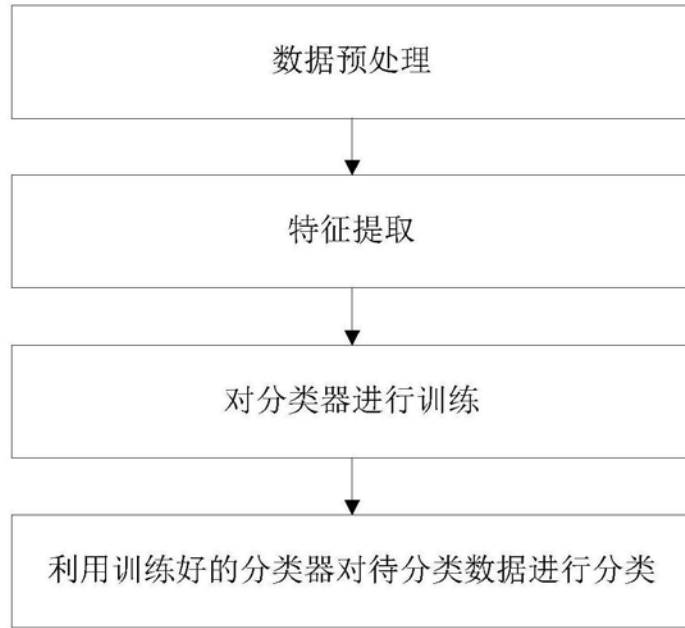
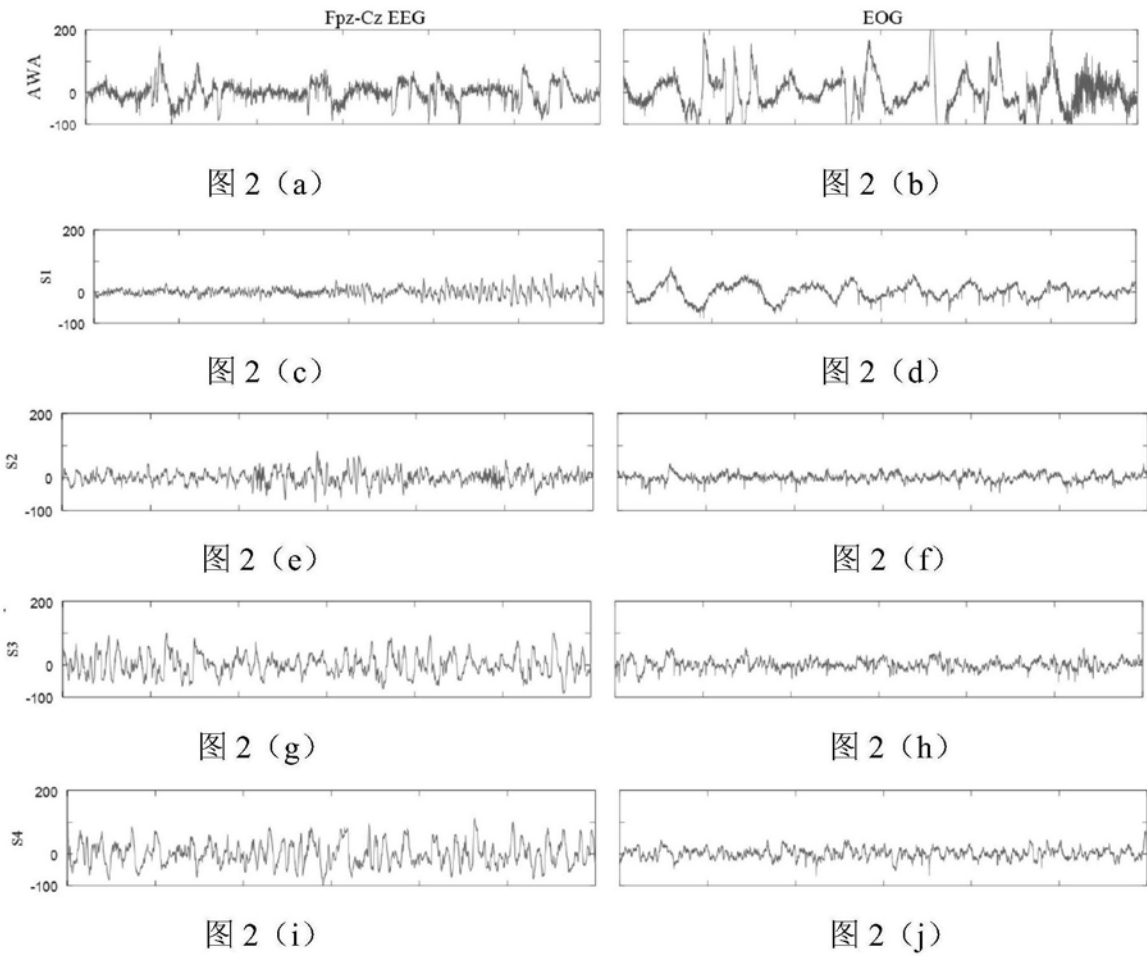


图1



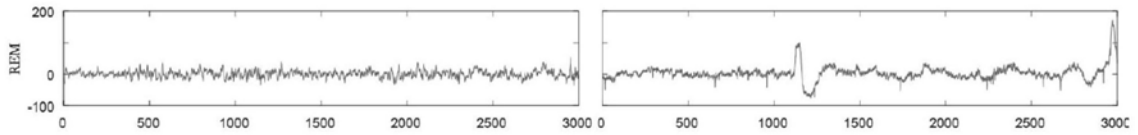


图 2 (k)

图 2 (l)

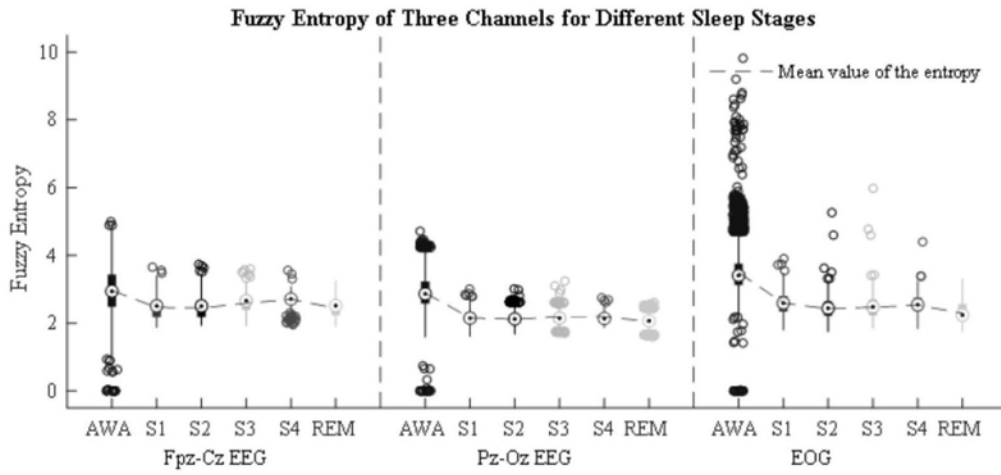


图3 (a)

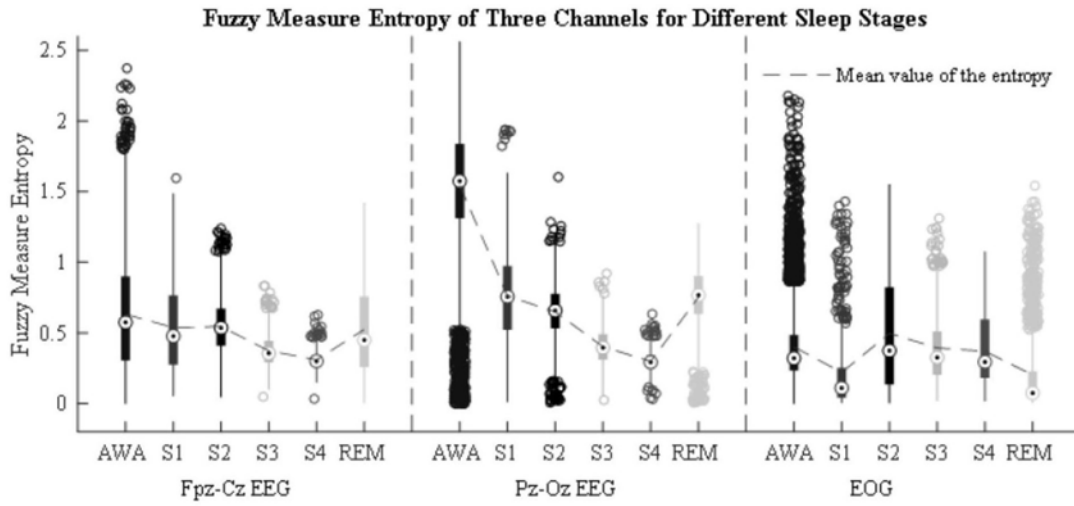


图3 (b)

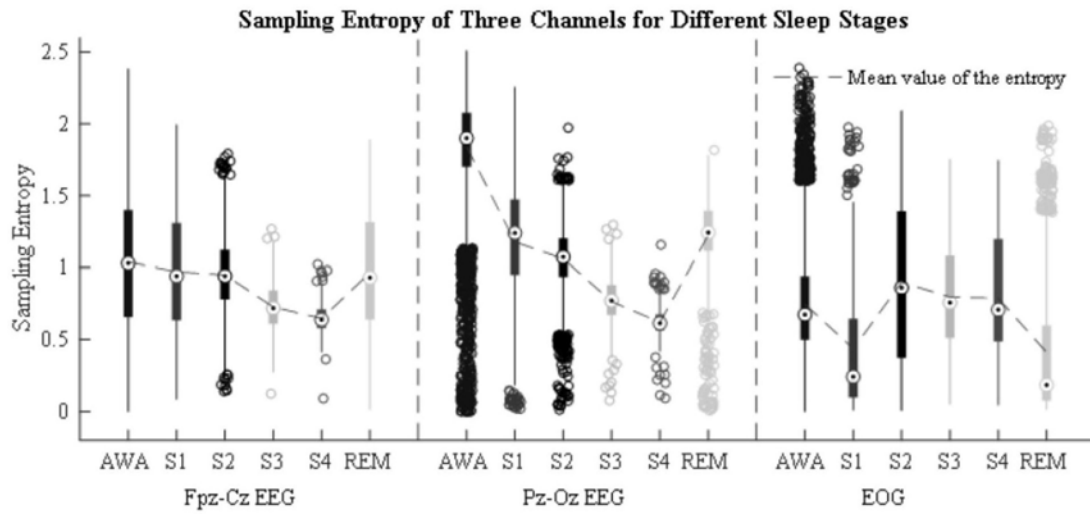


图3(c)

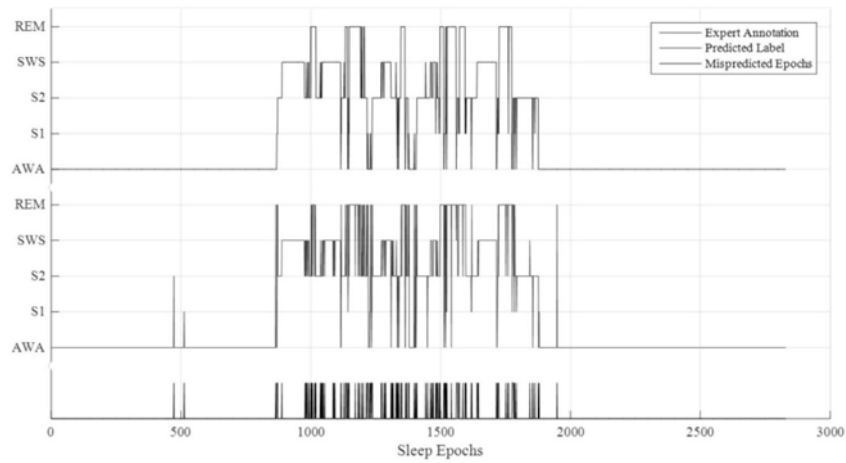


图4

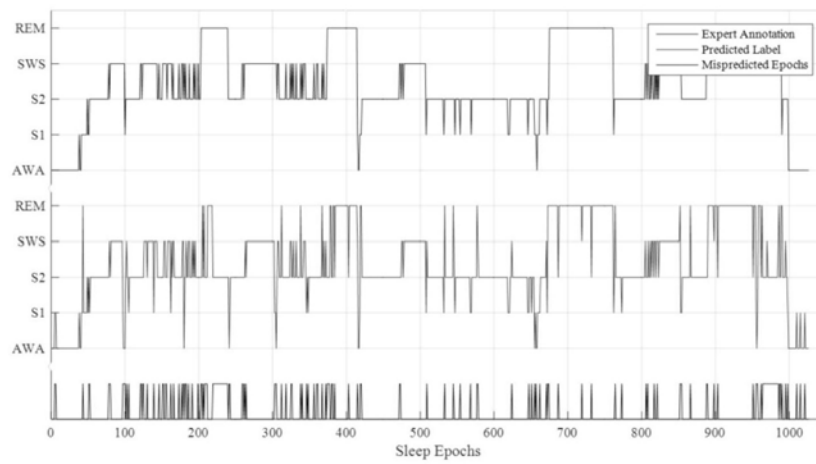


图5

专利名称(译)	基于熵特征及支持向量机的睡眠状态分类方法及系统		
公开(公告)号	CN108968915A	公开(公告)日	2018-12-11
申请号	CN201810600753.4	申请日	2018-06-12
[标]申请(专利权)人(译)	山东大学		
申请(专利权)人(译)	山东大学		
当前申请(专利权)人(译)	山东大学		
[标]发明人	魏守水 张志民 江兴娥 陈雷 刘振华		
发明人	魏守水 张志民 江兴娥 陈雷 刘振华		
IPC分类号	A61B5/00		
CPC分类号	A61B5/4812 A61B5/4818		
代理人(译)	黄海丽		
外部链接	Espacenet SIPO		

摘要(译)

本发明公开了基于熵特征及支持向量机的睡眠状态分类方法及系统，包括：数据预处理：基于研究睡眠分期三个通道信号：脑电信号Fpz-Cz EEG，脑电信号Pz-Oz EEG和水平眼电图信号EOG；将每个通道的信号均按照设定时间间隔进行均匀分割，分割后得到若干个片段；特征提取：从每个片段中提取模糊熵、模糊测度熵和样本熵；特征融合；设计一对多支持向量机分类器，支持向量机分类器对待分类信号进行分类，可以实现2-6多分类。

