分层向量自回归的多通道脑电信号的特征提取研究

王金甲 [1] 陈 春 [1]

摘 要 有效的特征提取方法能提高脑机接口 (Brain-computer interface, BCI) 系统对脑电 (Electroencephalogram, EEG) 信号的识别率. 因脑电信号都是多通道的, 本文将分层向量自回归 (Hierarchical vector autoregression, HVAR) 模型用于脑电信号的特征提取, 并结合传统的线性支持向量机 (Support vector machine, SVM) 用于脑电信号识别. 该模型不仅克服了自回归 (Autoregression, AR) 模型只能用来提取单通道特征的局限性, 而且不再采用传统 VAR (Vector autoregression) 模型所有通道共用一个时滞的处理方法. 创新之处在于在传统的 VAR 模型基础上添加正则化思想, 有效地压缩参数空间, 实现合理的分层结构. 本文首次将 HVAR 模型用于由 Keirn 等采集并整理的脑电数据中. 实验结果证明 HVAR 模型在阶数较小的情况下 (2 阶) 与阶数较大 (6 阶) 的 AR 模型效果相当, 可见低阶的 HVAR 能很好地刻画脑电信号的时空关联关系, 这说明 HVAR 可能是刻画 EEG 信号的一种新颖的方法, 这对其他多通道时间序列分析都有借鉴意义.

关键词 脑机接口, 脑电信号, 分层向量自回归模型, 特征提取, 近邻梯度

引用格式 王金甲, 陈春. 分层向量自回归的多通道脑电信号的特征提取研究. 自动化学报, 2016, **42**(8): 1215−1226

DOI 10.16383/j.aas.2016.c150461

Multi-channel EEG Feature Extraction Using Hierarchical Vector Autoregression

WANG Jin-Jia[1] CHEN Chun[1]

Abstract Feature extraction and classification of electroencephalogram (EEG) signals is a core part of brain-computer interface (BCI). For multi-channel EEG signal and high dimension of feature vector of BCI system, a novel EEG signal recognition method called hierarchical vector autoregression (HVAR) is presented, which extracts EEG feature using regression coefficient of HVAR model and linear support vector machine (SVM). It overcomes the limitations of the autoregression (AR) model that can be used to extract the single channel EEG only, and effectively avoids the vector autoregression (VAR) model sharing a same delay for all channels. Our contribution is that regularization is added on the traditional VAR model and a reasonable hierarchical structure is adopted. It effectively compresses parameter space of VAR model. In this paper, HVAR model is used for EEG data classification for the first time. Experimental results show that the recognition accuracy of extracted feature of HVAR model using a 2 lag order multi-channel is higher than that of AR model of 6 lag order. So low-level HVAR model can describe the portrayed temporal relationship of EEG well. This shows HVAR may be a novel method to portray EEG signal, which has reference significance to other multi-channel time-series.

Key words Brain-computer interface (BCI), electroencephalogram (EEG), hierarchical vector autoregression (HVAR), feature extraction, proximal gradient method

Citation Wang Jin-Jia, Chen Chun. Multi-channel EEG feature extraction using hierarchical vector autoregression. *Acta Automatica Sinica*, 2016, **42**(8): 1215−1226

通过建立计算机与脑电波之间的联系从而实现大脑与外界信息的交换已经成为多年的一个研究热点[1−2]. 脑机接口 (Brain-computer interface, BCI) 是一种新的人机交互方式, 只需要通过大脑的活动信号将命令传给计算机或其他外部设备[3]. 这种不依赖外围神经肌肉的通信系统为完全瘫痪的人提供全新的生活方式.

为了能够正确有效地控制 BCI 系统, 必须通过一定的技术将不同的大脑活动转换为系统可以识别的模式从而发送不同的命令到外部设备中. BCI 系统的核心之一是脑电图 (Electroencephalogram, EEG) 的信号分类, 包括预处理、特征提取和选择、分类三个过程. 由此可见, 在 EEG 识别过程中, 特征提取起着非常重要的作用, 提取出来的特征会直

收稿日期 2015-07-20 录用日期 2016-02-18
Manuscript received July 20, 2015; accepted February 18, 2016
国家自然科学基金 (61473339), 中国博士后科学基金 (2014M561202), 河北省博士后专项基金 (B2014010005), 首批 "河北省青年拔尖人才" 项目 ([2013]17) 资助
Supported by National Natural Science Foundation of China (61473339), China Postdoctoral Science Foundation (2014M561202), Hebei Province Postdoctoral Special Foundation (B2014010005), Hebei Province Top Young Talents ([2013]17)
本文责任编委 田捷
Recommended by Associate Editor TIAN Jie
1. 燕山大学信息科学与工程学院 秦皇岛 066004
1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004

接影响系统的分类正确率. 目前, 脑信号的特征提取方法主要包括: 小波变换和小波包变换[4]、自回归 (Autoregression, AR) 模型参数估计[5]、时频分析 (Time-frequency analysis)[6−7]、共空间模式 (Common spatial patterns, CSP)[8] 等. 大致可以将上述方法分为参数法和非参数法两类. 非参数法需要大量的样本, 脑电信号潜在信息的泄露不可避免; 而参数法是从有限的记录数据中得到估计参数来描述随机信号, 基于该方法不需要对原始数据的产生做任何假设使得它在脑信号研究中得到广泛的应用[9]. 其中, AR 模型因其表达式简单、适用范围广等优点[10] 受到众多学者青睐. Huan 等使用 3 种不同的算法求解 AR 模型参数, 将得到的系数作为神经网络分类器的输入实现 EEG 信号分类[11]; Lawhern 等利用 AR 模型对连续的脑电信号进行特征提取, 与支持向量机 (Support vector machine, SVM) 结合进行分类, 在几个不同的人工制造的脑电信号上取得了接近 94 % 的正确率[12]; Chen 等将 AR 模型应用到提取 EEG 的相位和频率信息中[13]; Li 等在 AR 模型的基础上添加了 L_p 范数的约束条件来压缩噪声等外界干扰对脑电信号的分析[9]. 但是, AR 模型是针对单个通道进行分析的, 所以在对多通道脑电信号的处理上会忽略各个通道之间的相互作用, 不能有效地挖掘出通道之间的互信息. 相比之下, 针对多通道信号建立的多变量自回归模型 (Multivariate autoregressive, MVAR) 更适合用于脑信号的研究. Anderson 等对 4 个受试者 6 个通道的两类任务进行四种方法的特征提取, 结果表明使用 MVAR 系数表示不同的行为模式能取得更好的分类性能, 且结果更为稳定[14]; Pei 等在对左右手运动两类任务进行分类时首先用 MVAR 进行特征提取, 然后使用基于 Mahalanobis 距离的判别分类器进行分类, 实验结果证明, 在分类正确率、信噪比和互信息等方面, MVAR 模型都优于 AR 模型[15]. Hu 等将 AR、MVAR 模型运用于基于胳膊运动的分类任务上, 得到了同样的结论, 即 MVAR 模型因能够挖掘出肌电图通道之间的互信息可以提高分类正确率[16]; Wang 等提出了新的多元自适应自回归 (Multivariate adaptive autoregressive, MVAAR) 模型对左右手和脚三类任务进行特征提取, 结果表明 MVAAR 模型对于在线 BCI 系统是一种更为有效的特征提取方法[17]; Zhao 等结合 MVAR 模型和核主成分分析 (Kernel principal component analysis, KPCA) 降维方法用于 EEG 信号检测驾驶精神疲劳状态, 不仅提高了正确率, 而且加快了收敛速度[18]. Heger 等用 MVAR 模型提取特征, 分析脑信号信息流之间的联系[19]; Faes 等提出了扩充 MVAR (Extended MVAR, eMVAR) 框架, 并将其成功地应用在心血管变异时间序列和多通道的脑电图中[20]; Varotto 等运用 MVAR 模型分析使用频域线性索引连接的脑信号[21]; Panzica 等使用 MVAR 模型对皮质性肌阵挛患者的脑电图和肌电图进行分析[22]; Wang 等结合 MVAR 与 MPCA (Mutilinear principal component analysis) 降维方法对 UCI 数据库中的脑电数据进行分类, 分类器选用贝叶斯分类器, 取得了很好的分类效果[23]. 但 MVAR 模型系数是矩阵形式, 在进行分析时很容易出现高维小样本问题.

本文将正则化思想添加在 MVAR 模型中, 并提出分层结构, 不仅能克服传统 MVAR 模型共用一个时滞的缺点, 而且由于正则化罚的使用有效地压缩了模型的参数空间, 将 MVAR 模型成功推广到高维小样本数据的应用上, 此外, 采用加速近邻梯度法求解模型加快了算法的收敛速度, 降低了时间复杂度.

1 LASSO-VAR 模型

1980 年, Sims 首先提出向量自回归 (Vector autoregression, VAR) 模型, 即上面提到的 MVAR 模型. 这种模型通过采用多方程联立的形式挖掘各个变量之间的线性关系. 定义 $\left\{\boldsymbol{y}_t \in \mathbf{R}^k\right\}_{t=1}^{T}$ 是长度为 T 的 k 维时间序列, p 阶的 VAR 模型具有如下形式:

$$\boldsymbol{y}_t = \boldsymbol{v} + B^{(1)}\boldsymbol{y}_{t-1} + \cdots + B^{(p)}\boldsymbol{y}_{t-p} + \boldsymbol{u}_t, \quad t = 1, \cdots, T \tag{1}$$

其中, p 是预先给定的最大时滞, $\boldsymbol{v} \in \mathbf{R}^k$ 是 k 维截距向量, $\left\{B^{(l)} \in \mathbf{R}^{k\times k}\right\}_{l=1}^{p}$ 是时滞为 l 的系数矩阵, B_{ij}^{l} 表示的是第 j 个时间序列前 l 个时刻的值对第 i 个时间序列当前时刻值的影响. $\left\{\boldsymbol{u}_t \in \mathbf{R}^k\right\}_{t=1}^{T}$ 定义的是均值为 0、同期协方差矩阵为 Σ_u 的白噪声, 即 $\boldsymbol{u}_t$ 满足以下两个条件: 1) $\mathrm{E}(\boldsymbol{u}_t) = 0$; 2) $\mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_\tau) = \begin{cases} 0, & t \neq \tau \\ \Sigma_u, & t = \tau \end{cases}$.

最小二乘估计 (Ordinary least square estimation, OLSE) 是求解线性回归模型参数估计最常用的方法. 使用最小二乘估计 VAR 模型的参数, 则参数估计可以描述为:

$$(\boldsymbol{v}, B) = \arg\min_{(\boldsymbol{v},B)} \frac{1}{2}\sum_{t=1}^{T}\left\|\boldsymbol{y}_t - \boldsymbol{v} - \sum_{l=1}^{p} B^{(l)}\boldsymbol{y}_{t-l}\right\|_2^2 \tag{2}$$

对于 $\mathrm{VAR}_k(p)$ 模型, 需要估计的参数个数为 k^2p, 也就是说, 参数空间随着时间序列的增加呈平方增长, 这对低频高维的时序数据很容易出现参数任意化问

题. 所以要想将 VAR 模型应用在高维数据上, 必须采取一定的正则化技术压缩参数空间.

Tibshirani 提出的 LASSO 模型[24] 是在最小二乘的基础上对系数的 1-范数施加约束, 不仅能够达到压缩系数的目的, 而且能够产生稀疏解, 从而降低参数空间.

为表示方便, 定义: $k \times T$ 维的响应矩阵 $Y = y_1, \cdots, y_T$, 表示所有时间序列不同时刻的值; $k \times kp$ 维的参数矩阵 $B = [B^{(1)} \cdots B^{(p)}]$, 表示不同序列间的相互影响; $kp \times 1$ 维的向量 $\boldsymbol{z}_t = \left[\boldsymbol{y}_{t-1}^{\mathrm{T}} \cdots \boldsymbol{y}_{t-p}^{\mathrm{T}}\right]^{\mathrm{T}}$, 表示所有时间序列滞后当前时刻 $1, 2, \cdots, p$ 个时刻的值; $kp \times T$ 维的协变量矩阵 $Z = [z_1 \cdots z_T]$, 此处, 我们需要用到 $\left\{\boldsymbol{y}_{-(p-1)}, \cdots, \boldsymbol{y}_0\right\}$ 来初始化 Z; $k \times T$ 维的扰动矩阵 $U = [\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_T]$ 以及 $T \times 1$ 维的全 1 向量 $\mathbf{1} = [1 \ \cdots \ 1]^{\mathrm{T}}$. 在最小二乘的基础上添加 L_1 罚, 也就是系数的绝对值之和. 这就是我们所说的 LASSO-VAR 模型:

$$\min \frac{1}{2} \left\| Y - \boldsymbol{v}\mathbf{1}^{\mathrm{T}} - BZ \right\|_2^2 + \lambda \|B\|_1 \tag{3}$$

其中, $\|A\|_2 = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} {a_{ij}}^2}$ 是矩阵 $A_{m \times n}$ 的 F-范数, $\|A\|_1 = \sum_i \sum_j |a_{ij}|$ 是矩阵 $A_{m \times n}$ 的 1-范数, λ 是正则化参数. LASSO-VAR 与 Bayesian 相比, 不仅可以使最小二乘解的某些项收缩至 0, 实现变量选择, 而且使得在回归参数个数 $k \times kp$ 接近甚至超过样本大小 $k \times T$ 的情况下也能得到可靠估计. 下一节我们将采用坐标下降法来求解该模型.

1.1　坐标下降法

记 $f(B, \boldsymbol{v}) = \frac{1}{2} \|Y - \boldsymbol{v}\mathbf{1}^{\mathrm{T}} - BZ\|_2^2 + \lambda \|B\|_1$, 将每一个时间序列的回归模型看作一个子问题, 那么可将目标函数分解为 k 个子问题. 第 j 个子问题的形式为:

$$f_j(B_{j\cdot}, v_j) = \frac{1}{2} \left\| Y_{jt} - v_j\mathbf{1}^{\mathrm{T}} - B_{j\cdot}Z \right\|_2^2 + \lambda \|B_{j\cdot}\|_1 \tag{4}$$

式中, v_j 和 $B_{j\cdot}$ 为未知变量. 首先通过求 $f_j(B_{j\cdot}, v_j)$ 关于 v_j 的偏导数并令其为 0 得到截距值 v_j. 即: $0 = \nabla_v f_j(B_{j\cdot}, v_j) = (Y_{jt} - v_j\mathbf{1}^{\mathrm{T}} - B_{j\cdot}Z)\mathbf{1}$, 从而求得: $v_j = \overline{Y}_j - \widehat{B}_{j\cdot}\overline{\boldsymbol{Z}}$. 所以截距向量 $\boldsymbol{v} = \overline{\boldsymbol{Y}} - \widehat{B}\overline{\boldsymbol{Z}}$. 其中, $\overline{\boldsymbol{Y}}$ 是 $k \times 1$ 维的响应矩阵的行均值向量, $\overline{\boldsymbol{Z}}$ 是 $kp \times 1$ 维的协变量矩阵的行均值向量. 接下来用坐标下降法求 $B_{j\cdot}$. 坐标下降法的基本思想是每次只优化一维变量, 其余变量看作常量, 这样优化问题可以很快完成, 且相关方程可以在变量循环中更新; 而通常情况下, 极小化的变量在众多参数中不随循环改变, 因此整个迭代过程将很快完成.

假设 Y 和 Z 都是中心化的, 则目标函数中不再包含截距向量. 使用坐标下降法逐个更新系数矩阵 B 的元素. 假设当前估计是 B 的 j 行 r 列的元素 B_{jr}, 那么将式 (4) 转换为单变量形式:

$$\begin{aligned} f_{jr}(B_{jr}) = &\frac{1}{2} \sum_t \left(Y_{jt} - \sum_{l \neq r} B_{jl} Z_{lt} - B_{jr} Z_{rt} \right)^2 + \\ &\lambda \sum_{l \neq r} |B_{jl}| + \lambda |B_{jr}| \end{aligned} \tag{5}$$

分 $B_{jr} > 0$, $B_{jr} < 0$, $B_{jr} = 0$ 三种情况对 B_{jr} 求导, 令 $R_t = Y_{jt} - \sum_{l \neq r} B_{jl} Z_{lt}$, 结果如下:

$$B_{jr} = \begin{cases} \dfrac{\sum\limits_t R_t Z_{rt} - \lambda}{\sum\limits_t {Z_{rt}}^2}, & \sum\limits_t R_t Z_{rt} > 0, \ \left| \sum\limits_t R_t Z_{rt} \right| > \lambda \\ \dfrac{\sum\limits_t R_t Z_{rt} + \lambda}{\sum\limits_t {Z_{rt}}^2}, & \sum\limits_t R_t Z_{rt} < 0, \ \left| \sum\limits_t R_t Z_{rt} \right| > \lambda \\ 0, & \left| \sum\limits_t R_t Z_{rt} \right| < \lambda \end{cases} \tag{6}$$

可见 B_{jr} 的坐标更新为软阈值形式: $B_{jr} = \frac{S(\sum_t R_t Z_{rt}, \lambda)}{\sum_t {Z_{rt}}^2}$. 其中, $S(x, \phi) = \mathrm{sgn}(x)(|x| - \phi)_+$, $(|x| - \phi)_+ = \max(|x| - \phi, 0)$.

从系数矩阵的求解过程中可以看出 LASSO-VAR 模型是将所有变量看作无关变量, 这也使得虽然通过添加 L_1 罚能得到稀疏解, 但是这种稀疏模式没有固定结构. 举例来说: 最大时滞为 4 的二元时间序列, 通过 LASSO-VAR$_2$(4) 得到的系数的可能稀疏模式如图 1 所示.

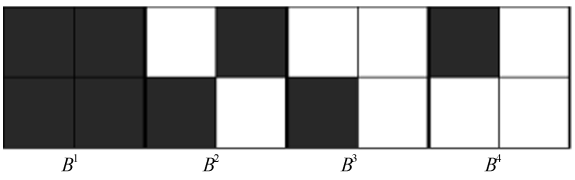

图 1　LASSO-VAR$_2$(4) 得到的稀疏模式示例图

Fig. 1　The sparse pattern for LASSO-VAR$_2$(4)

2　HVAR 模型

2.1　模型简介

HVAR (Hierarchical vector autoregression) 模型在 LASSO-VAR 的基础上, 通过对系数矩阵 B 的不同约束得到不同的分层结构, 可以生成一个具有

较高预测性能的可解释模型. 在讨论之前做以下定义:

$$B^{(l:p)} = [B^{(l)} \cdots B^{(p)}] \in \mathbf{R}^{k\times k(p-l+1)}$$
$$B_i{}^{(l:p)} = [B_i{}^{(l)} \cdots B_i{}^{(p)}] \in \mathbf{R}^{1\times k(p-l+1)}$$
$$B_{ij}{}^{(l:p)} = [B_{ij}{}^{(l)} \cdots B_{ij}{}^{(p)}] \in \mathbf{R}^{1\times (p-l+1)}$$

此外, 我们还定义了一个 $k\times k$ 维的最大时滞矩阵 L, L 中的每个元素定义如下: $L_{ij} = \begin{cases} \max\{l\}, & B_{ij}^l \neq 0 \\ 0, & B_{ij}^l = 0 \end{cases}$, $l = 1,\cdots,p$. 因此, L_{ij} 表示的是时间序列 j 对时间序列 i 的最长作用时间. 若 L 中的所有元素都相等即为普通的 $\mathrm{VAR}_k(p)$ 模型, 本文详细介绍三种分层结构.

1) 分量方式 HVARC: 将每一个时间序列 y_{jt} 看成一个边际方程, 每个方程有自己的最大时滞, 即: $L_{ij} = L_i$, $\forall j, i = 1,\cdots,k$. 若定义时滞矩阵: $L = \begin{bmatrix} 4\,4\,4 \\ 2\,2\,2 \\ 3\,3\,3 \end{bmatrix}$, 则该分层结构的时滞矩阵如图 2 所示. 图 2 表示所有时间序列对第一个时间序列的作用时间都为 4 个单位, 对第二个时间序列的作用时间为 2 个单位.

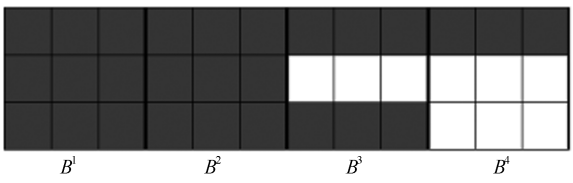

图 2 $\mathrm{HVARC}_3(4)$ 的分层时滞结构示例图

Fig. 2 A componentwise hierarchical lag structure: $\mathrm{HVARC}_3(4)$

2) 自身–其他分层结构 HVARO: 在 HVARC 的基础上增加假设条件: 自身作用时间不短于其他时间序列的作用时间, 且最多长于其他序列 1 个时间单位. 即: $L_{ij} = \begin{cases} L_i{}^{other}, & i \neq j \\ L_i{}^{other} \text{ 或者 } L_i{}^{other}+1, & i = j \end{cases}$, $i,j = 1,\cdots,k$. 表现在系数矩阵上是指: $B_{ij}^{\left([L_i{}^{other}+1]:p\right)} = 0, B_{ij}^{\left(L_i{}^{other}\right)} \neq 0$, $i \neq j$, $B_{ii}^{\left(L_i{}^{other}\right)} \neq 0$, 且对 $B_{ii}{}^{(other+1)}$ 没有约束, $B_{ii}{}^{([other+2]:p)} = 0$. 若定义时滞矩阵: $L = \begin{bmatrix} 4\,3\,3 \\ 2\,2\,2 \\ 2\,2\,3 \end{bmatrix}$, 则该分层结构的时滞矩阵如图 3 所示. 即: 第一个时间序列对其自身的作用时间为 4 个单位, 而其他时间序列对第一个时间序列的作用时间为 3 个单位.

3) 单个元素的分层结构 HVARE: 该分层结构比较灵活, 不同的时间序列彼此之间的作用时间不同. 即: $B_{ij}^{(L_{ij})} \neq 0$, $B_{ij}^{([L_{ij}+1]:p)} = 0$, $\forall i,j$. 若定义时滞矩阵: $L = \begin{bmatrix} 4\,1\,2 \\ 2\,0\,1 \\ 4\,4\,2 \end{bmatrix}$, 则时滞矩阵如图 4 所示. 彼此之间没有特定的约束关系.

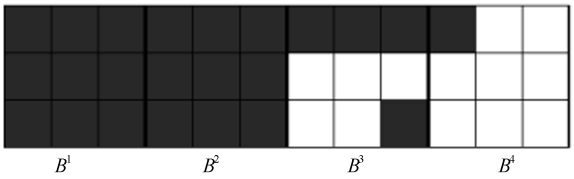

图 3 $\mathrm{HVARO}_3(4)$ 的分层时滞结构示例图

Fig. 3 An own-other hierarchical lag structure: $\mathrm{HVARO}_3(4)$

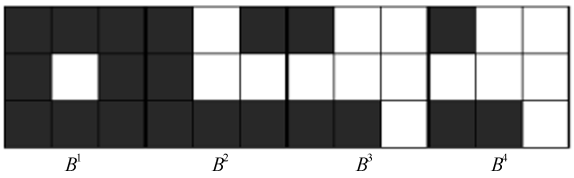

图 4 $\mathrm{HVARE}_3(4)$ 的分层时滞结构示例图

Fig. 4 An elementwise hierarchical lag structure: $\mathrm{HVARE}_3(4)$

为了实现上述三种分层结构, 我们采用分层组 LASSO 罚对系数加以约束. 分层组 LASSO 罚是指具有嵌套结构的组 LASSO 罚. 理论已经证明组 LASSO 罚[25] 将每一组的变量看成一块, 对块变量施加约束, 能够实现变量的组选择; 而嵌套组[26] 能够实现分层选择, 即某一组变量没有被选中, 则嵌套于该组的另一组变量也不被选中. 可以通过在 VAR 模型上添加不同的罚函数实现以上三种分层结构.

1) HVARC: 前面已经提到该分层结构是指每一个边际方程的阶数相同, 也就是要求满足这样的条件: 如果 $\widehat{B}_i^{(l)} = 0$, 那么 $\widehat{B}_i^{(l')} = 0, l' > l, i = 1,\cdots,k$. 即嵌套组 LASSO 罚可以实现这样的分层结构, 故只需在 VAR 模型的基础上添加嵌套组 LASSO 罚函数, 目标函数定义如下:

$$\min_{\boldsymbol{v},B}\left\{\frac{1}{2}\left\|Y - \boldsymbol{v}\mathbf{1}^{\mathrm{T}} - BZ\right\|_2^2 + \lambda\sum_{i=1}^{k}\sum_{l=1}^{p}\left\|B_i^{(l:p)}\right\|_2\right\} \tag{7}$$

随着 λ 的增加, i 越大, l 越小, 就会有 $\widehat{B}_i^{(l:p)} = 0$. 这样的罚更趋向于选择一个较小的最大时滞, 而不是简单地得到一个没有任何特殊结构的系数矩阵 B.

2) HVARO: 该分层结构要求实现的是对角元素优于非对角元素, 所以只需在 HVARC 的基础上

对非对角元素进行再一次的惩罚, 故目标函数定义如下:

$$\min_{\boldsymbol{v},B}\left\{\frac{1}{2}\left\|Y-\boldsymbol{v}\mathbf{1}^{\mathrm{T}}-BZ\right\|_2^2+\lambda\sum_{i=1}^{k}\sum_{l=1}^{p}\left[\left\|B_i^{(l:p)}\right\|_2+\left\|B_{i,-i}^{(l)},B_i^{([l+1]:p)}\right\|_2\right]\right\} \tag{8}$$

其中, $B_{i,-i}^{(l)}=\left\{B_{ij}^{(l)}:j\neq i\right\}$, 并且约定 $B_i^{([p+1]:p)}=0$. 式 (8) 中罚函数与 HVARC 的罚函数的不同之处在于增加了第二项, 放宽了对对角元素 $B_{ii}^{(l)}$ 的约束, 这样就能保证在时滞为 l 时, 即使序列之间的交互影响已经为 0, 而自身影响可能并不为 0. 该模型能够确保自身的影响可能长于其他时间序列的影响 1 个时间单位.

3) HVARE: 该分层结构对时滞矩阵的结构没有硬性的约束, 任意时间序列对的时滞都可以不同, 所以只需对任意时间序列对的时滞矩阵进行约束, 故目标函数定义如下:

$$\min_{\boldsymbol{v},B}\left\{\frac{1}{2}\left\|Y-\boldsymbol{v}\mathbf{1}^{\mathrm{T}}-BZ\right\|_2^2+\lambda\times\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{l=1}^{p}\left[\left\|B_{ij}^{(l:p)}\right\|_2\right]\right\} \tag{9}$$

在这样的分层结构中, 因为每对时间序列都有自己的最大时滞, 因此 $\widehat{B}_{ij}^{(l:p)}=0$ 可能出现在时序对 (i,j) 的任意 l 值的情况下. 这样的分层结构是最灵活的.

前面已经讨论过截距向量与响应矩阵、协变量矩阵的关系. 此外仍假定响应矩阵与协变量矩阵都是中心化的, 所以在下面的具体求解过程中将不再讨论截距向量. 式 (7) ∼ (9) 都是基于分层组 LASSO 罚的, 在这儿引入函数 $\Omega_i(B_i^{(l:p)})$ 表示系数矩阵不同形式的范数, 即将式 (7) ∼ (9) 的罚函数统一用该函数表示, 那么可将三个凸优化问题统一为以下形式:

$$\min_{B}\left\{\frac{1}{2}\left\|Y-BZ\right\|_2^2+\lambda\sum_{i=1}^{k}\sum_{l=1}^{p}\Omega_i(B_i^{(l:p)})\right\} \tag{10}$$

因为每一个边际方程都是独立的, 所以可以将式 (10) 分解成 k 个如下的子问题求解:

$$\min_{B_i}\left\{\frac{1}{2}\left\|Y_i-B_iZ\right\|_2^2+\lambda\sum_{l=1}^{p}\Omega_i(B_i^{(l:p)})\right\} \tag{11}$$

2.2　加速近邻梯度法

下面详细讲述加速近邻梯度法求解式 (11). 加速近邻梯度法是由 Yurri Nesterov 于 1983 年首先提出的[27]. 该方法以负梯度的方向作为下降方向, 将上两代的迭代值做线性组合得到一个中间变量, 然后由该中间变量导出当代的最优解. 加速近邻梯度法通过对近邻梯度法的一个简单修正, 不仅保持了梯度法的简洁性, 更提高了算法的效率. 理论上已经证明了加速近邻梯度法的收敛率是 $1/k^2$. 本文选取的相邻两次迭代值线性组合的权重为 $\theta_m=2/(m+2)$, 步长 $t\leftarrow\sigma_1(Z)^{-2}$, 其中 $\sigma_1(Z)$ 是协变量矩阵 Z 的最大奇异值. 记 $L_i(B_i)=\frac{1}{2}\left\|Y_i-B_iZ\right\|_2^2$, 则有 $\nabla L_i(B_i)=(B_iZ-Y_i)\,Z^{\mathrm{T}}$, 式 (11) 的系数矩阵 B 的详细更新步骤见算法 1.

算法 1. 加速近邻梯度法求解式 (11)

输入. Y, Z, $\widehat{B}[0]$, λ, ε;

步骤 1. 初始化: $\widehat{B}[1]\leftarrow\widehat{B}[0]$; $\widehat{B}[2]\leftarrow\widehat{B}[0]$; $t\leftarrow\sigma_1(Z)^{-2}$, $i=1$;

步骤 2. 迭代直至收敛:

1) 更新中间变量 $\widehat{b}$, $\widehat{b}\leftarrow\widehat{B}_i[m-1]+\frac{m-2}{m+1}\left(\widehat{B}_i[m-1]-\widehat{B}_i[m-2]\right)$;

2) 更新当代的系数值 $\widehat{B}_i[m]$, $\widehat{B}_i[m]\leftarrow prox_{t\lambda\Omega_i}\left(\widehat{b}-t\left(\widehat{b}Z-Y_i\right)Z^{\mathrm{T}}\right)$;

3) 判断 $\left\|\widehat{b}-\widehat{B}_i[m]\right\|\leq\varepsilon$ 是否成立, 若成立, 内层循环结束, 转至步骤 3; 否则, $m\leftarrow m+1$, 转至 1);

步骤 3. 若 $i=k$, 算法结束; 否则 $i\leftarrow i+1$, 转至步骤 2;

输出. 系数矩阵 B.

三种不同分层结构的罚函数都是基于分层组 LASSO 罚, 由近邻算子的定义可知:

$$prox_{t\lambda\Omega_i}(u)=\arg\min\left\{\frac{1}{2}\left\|u-v\right\|_2^2+t\lambda\Omega_i(v)\right\} \tag{12}$$

将三种范数的近邻算子统一记为:

$$\widehat{x}=\arg\min\left\{\frac{1}{2}\left\|x-\widetilde{x}\right\|_2^2+\lambda\sum_{h=1}^{H}\omega_h\|x_{g_h}\|_2\right\} \tag{13}$$

其中, g_h 是第 h 组包含的变量的索引集合, 且 $g_1\subset\cdots\subset g_H$. 我们已经知道对于组 LASSO 罚, 组间是不重叠的, 它的近邻问题按组可分离, 近邻算子是广义软阈值算子. 而在分层结构中, 嵌套组的出现导致不能直接得到分层组 LASSO 罚的近邻算子. Jenatton 等已经证明对于分层组 LASSO 的近邻问题, 可以通过求其对偶范数的近邻问题得到[28]. 本文直接使用 Jenatton 等的结果. 见算法 2.

算法 2. 求解分层组 LASSO 罚的近邻算子

输入. $\widetilde{x}, \lambda, \omega_1, \cdots, \omega_H$;
步骤 1. 初始化: $r \leftarrow \widetilde{x}$;
步骤 2. 循环 $h = 1, \cdots, H$:
$r_{g_h} \leftarrow (1 - \lambda\omega_h / \|r_{g_h}\|)_+ r_{g_h}$;
输出. $\widehat{x} = r$.

3 实验结果与分析

3.1 实验数据说明

将 HVAR 模型应用到脑机接口的背景中. 实验采用的是由 Keirn 等采集的脑电数据, 详见网址: http://www.cs.colostate.edu/eeg/main/data/1989_Keirn_and_Aunon. 该数据集记录的是 7 位受试者在执行想象任务时的 EEG 数据. 每次实验要求受试者完成一种任务的想象, 时间持续 10 s, 采样频率是 250 Hz, 所以每一个样本的每个通道会产生 2 500 个采样点. 在一个周期中每个动作重复执行 5 次, 共 5 类任务. 采样期间受试者的 EEG 数据通过 7 个位于头部的电极记录, 采样电极的分布如图 5 所示. 每个受试者的样本数列于表 1.

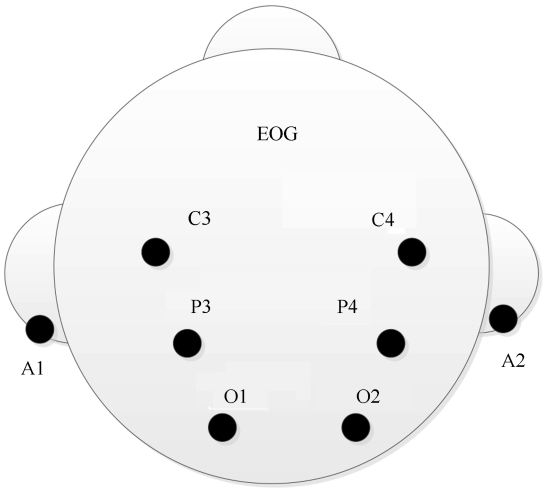

图 5 采样电极放置图
Fig. 5 Electrode placement

表 1 受试者的样本数
Table 1 Trails of subjects

受试者	样本数
1	10
2	5
3	10
4	10
5	15
6	10
7	5

图 5 中, 电极 C3, C4, P3, P4, O1, O2, EOG 记录 7 个通道的 EEG 信号, 电极 A1, A2 之间通过一个放大器 (Grass7P511) 连接, 该放大器的模拟滤波器的带宽为 0.1 Hz ~ 100 Hz, 这两个电极的记录是电乳突的参考电位.

从表 1 可知, 受试者 1, 3, 4, 6 分别执行了两个周期的任务, 受试者 2, 7 是执行了一个周期的任务, 受试者 5 则是执行了 3 个周期的任务. 前面已提到每个受试者在每个周期中共执行 5 类任务, 详细介绍如下:

1) 基准任务: 在该任务下, 要求受试者尽可能地放松, 尽量不要去想任何事情. 所以可将该任务看作脑电图的基准.

2) 乘法计算任务: 要求受试者计算一个复杂的乘法问题, 在计算过程中, 受试者不能发声或有明显的动作, 每次的乘法问题都是不重复的, 因此受试者不可能直接给出答案, 而且也已经证实受试者在每次的执行过程中都没有在 10 s 之内给出答案, 所以在采样期间受试者一直处于计算状态.

3) 字母组合任务: 要求受试者不发声地给朋友或亲人写一封信. 因为任务重复 5 次, 所以在后边几次的任务中要求受试者回想在先前的信中是否有遗漏的字母.

4) 几何图旋转任务: 在该任务下, 首先给受试者 30 s 的时间学习画一幅复杂的 3 维图, 30 s 之后, 画板撤掉, 然后要求受试者想象该物体沿着某个轴旋转.

5) 视觉计算任务: 要求受试者想象将数字按顺序写在一块黑板上.

3.2 实验过程

电极 EOG 是用来检测受试者是否有眨眼行为, 已经有相关实验证明该通道对任务分类正确率没有影响[29], 因此在实验中我们只用了 6 个通道的值. 图 6 是受试者 1 的 5 类任务的 6 个通道的时序图.

图 6 的每一幅子图都是长度为 2 500 的 6 元时间序列, 由图可见, 这 6 元时间序列基本上都是平稳的. 实验中用到的是各个任务的 10 个样本, 从表 1 可知, 受试者 2 和 7 每一类任务只有 5 个样本, 而受试者 4 的数据中包含了未知的无效值, 故可用的是受试者 1, 3, 5, 6 的数据. 对于受试者 5, 我们用的是他的前 10 个样本, 重新对 1, 3, 5, 6 编号为 1, 2, 3, 4. 此外, 我们用到了 0.5 s 的时间窗, 所以将原始长度为 2 500 的时间序列分割成 20 段, 每个时间段中包含了 125 个采样点. 考虑到每个人的思维方式不同, 所以对每个受试者单独进行实验.

首先对这 20 段分别用 VAR、LASSO-VAR、HVARC、HVARO、HVARE 5 种方法进行特征提取, 那么每类任务的每一个样本就会得到 20 种特征模式, 因此每类想象任务共得到 200 种特征

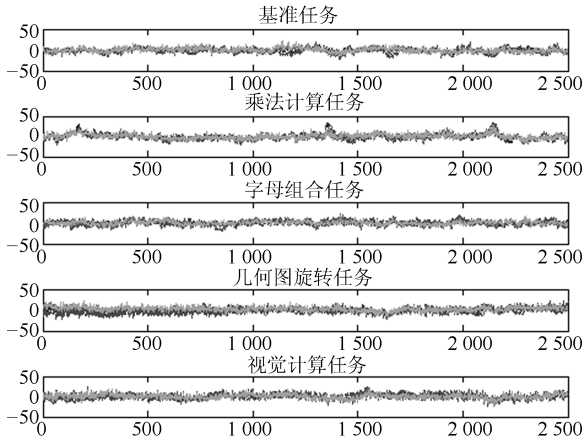

图 6　受试者 1 的分类任务 6 个通道的多元时序图

Fig. 6　The sequence diagram of each channel for Subject 1

模式, 也就是说每类任务都会有 200 个样本. 然后将不同的任务进行两两组合, 不同的任务组合方式见表 2; 接着将组合后的 400 个样本随机分成大小相等的训练集和测试集, 共进行 10 次试验; 最后将 10 个测试集上得到的正确率求平均. 分类器我们选择线性支持向量机. 实验中每种方法的阶数都选择 2 ∼ 7.

3.3　实验结果与讨论

实验中使用网格法选择正则化参数 λ. 网格的深度与长度分别设置为 30 和 10. 首先我们详细分析受试者 1 的实验结果. 受试者 1 的不同任务组合方式, 在 5 种不同特征提取方法的不同时滞下得到的 10 次试验的平均分类正确率如表 3 ∼ 7 所示.

从表 3 ∼ 7 可见, 不同的任务组合方式所对应的最大时滞是不同的, 但总的看来, 当阶数等于 2 时, 不同的特征提取方法至少在 5 种任务组合方式下能获得最高正确率. 由此我们可认为受试者 1 的最大时滞是 2. 为了对 10 次实验的分类结果有更直观的认识, 本文以第 5 种和第 3 种任务组合方式为例进行详细分析. 图 7 和图 8 表示在阶数为 2 时, 五种不同方法分别在这两种情况下得到的分类结果的箱线图.

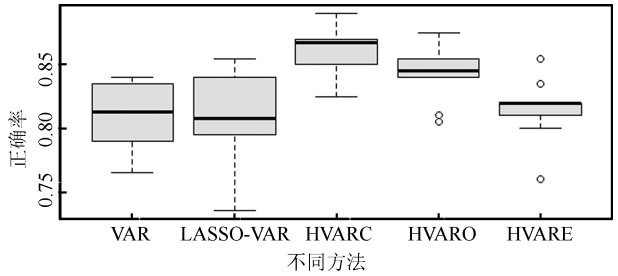

图 7　第 5 种任务组合方式对应的不同方法的正确率箱线图

Fig. 7　The boxplot of the fifth task combination for different methods

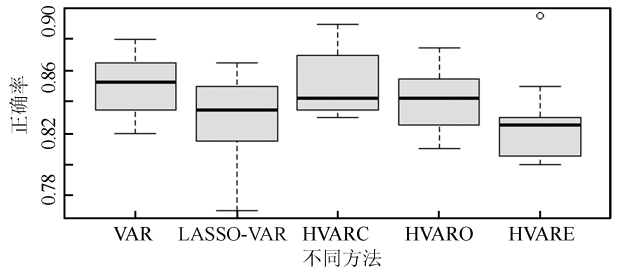

图 8　第 3 种任务组合方式对应的不同方法的正确率箱线图

Fig. 8　The boxplot of the third task combination for different methods

表 2　任务组合方式

Table 2　Patterns of task combinations

任务编号	组合方式	任务编号	组合方式
1	基准、乘法计算	6	乘法计算、几何图旋转
2	基准、字母组合	7	乘法计算、视觉计算
3	基准、几何图旋转	8	字母组合、几何图旋转
4	基准、视觉计算	9	字母组合、视觉计算
5	乘法计算、字母组合	10	几何图旋转、视觉计算

表 3　VAR 模型不同时滞的平均正确率

Table 3　Average accuracy rate using VAR model with different order

不同时滞 \ 任务	1	2	3	4	5	6	7	8	9	10
2	**0.83**	0.73	**0.85**	**0.85**	**0.81**	**0.91**	**0.84**	**0.77**	**0.74**	**0.78**
3	0.79	0.74	0.81	0.82	0.80	0.88	0.80	0.73	0.73	0.75
4	0.75	**0.75**	0.80	0.82	0.80	0.86	0.80	0.71	0.71	0.70
5	0.73	0.71	0.78	0.78	0.81	0.86	0.77	0.70	0.65	0.70
6	0.73	0.72	0.75	0.76	0.78	0.84	0.76	0.68	0.68	0.69
7	0.71	0.67	0.71	0.73	0.77	0.81	0.72	0.65	0.65	0.65

表 4 LASSO-VAR 模型不同时滞的平均正确率

Table 4 Average accuracy rate using LASSO-VAR model with different order

不同时滞 \ 任务	1	2	3	4	5	6	7	8	9	10
2	0.79	0.70	**0.83**	0.79	**0.81**	**0.88**	**0.83**	**0.70**	**0.72**	**0.67**
3	**0.80**	**0.73**	0.79	**0.80**	0.80	0.86	0.78	0.65	0.65	0.65
4	0.79	**0.73**	0.82	0.79	0.78	0.85	0.76	0.65	0.63	0.64
5	0.77	0.69	0.79	0.76	0.76	0.83	0.71	0.67	0.61	0.64
6	0.76	0.71	0.78	0.73	0.75	0.84	0.75	0.66	0.62	0.63
7	0.76	0.66	0.76	0.73	0.73	0.84	0.75	0.63	0.58	0.61

表 5 HVARC 模型不同时滞的平均正确率

Table 5 Average accuracy rate using HVARC model with different order

不同时滞 \ 任务	1	2	3	4	5	6	7	8	9	10
2	**0.82**	0.73	**0.85**	0.80	**0.86**	**0.90**	**0.84**	**0.73**	**0.70**	**0.70**
3	0.81	0.75	0.82	**0.84**	0.84	**0.90**	**0.84**	0.70	0.68	0.69
4	0.80	**0.76**	0.84	0.82	0.83	0.89	0.82	0.69	0.66	0.68
5	0.78	0.73	0.81	0.79	0.85	0.87	0.78	0.69	0.64	0.64
6	0.80	0.73	0.81	0.79	0.82	0.88	0.81	0.70	0.65	0.66
7	0.78	0.73	0.81	0.79	0.80	0.87	0.81	0.67	0.63	0.66

表 6 HVARO 模型不同时滞的平均正确率

Table 6 Average accuracy rate using HVARO model with different order

不同时滞 \ 任务	1	2	3	4	5	6	7	8	9	10
2	0.81	0.72	**0.84**	0.81	0.84	0.90	**0.86**	**0.71**	**0.72**	**0.72**
3	**0.82**	**0.76**	0.83	**0.82**	**0.85**	**0.91**	0.83	0.69	0.69	0.69
4	**0.82**	0.74	**0.84**	0.80	0.83	0.88	0.82	0.67	0.68	0.67
5	0.81	0.71	0.82	0.81	0.83	0.87	0.82	0.67	0.64	0.65
6	0.81	0.72	0.81	0.79	0.81	0.87	0.82	0.67	0.67	0.66
7	0.79	0.71	0.81	0.80	0.81	0.87	0.82	0.65	0.62	0.64

表 7 HVARE 模型不同时滞的平均正确率

Table 7 Average accuracy rate using HVARE model with different order

不同时滞 \ 任务	1	2	3	4	5	6	7	8	9	10
2	**0.79**	0.70	**0.83**	0.80	**0.82**	**0.87**	**0.83**	**0.69**	**0.71**	0.68
3	0.77	0.72	0.80	**0.82**	0.81	**0.87**	0.79	0.64	0.68	**0.69**
4	0.78	**0.73**	0.80	0.80	0.79	0.86	0.77	0.65	0.65	0.64
5	0.78	0.71	0.78	0.79	0.80	0.85	0.75	0.66	0.64	0.64
6	0.78	0.71	0.76	0.79	0.80	0.85	0.77	0.68	0.64	0.66
7	0.78	0.68	0.76	0.80	0.76	0.86	0.77	0.64	0.64	0.65

从上边两个箱线图来看, 不同任务组合对应的最佳特征提取方法是不一样的. 对于第 5 种任务组合, 较好的特征提取方法是 HVAR 方法, 以 HVARC 为最佳, 在 10 次实验中, 无论是最小值、最大值、上下四分位点还是中位数都远远高于其他四种方法, 但是 HVARC 方法得到的正确率的分布不如 VAR 均匀. 而对于第 3 种任务组合, 则是 VAR 的效果更好些, HVARC 次之. 两种方法 50 % 的值的分布范围大致相同, 但总的来看仍属 VAR 的分类正确率较高. 在两种情况下, 没有固定稀疏结构 LASSO-VAR 的分类性能都不好, 而 HVARO 和 HVARE 都在不同程度上出现了离散点.

对其他受试者进行同样分析, 得到受试者 2、3、4 的最大时滞也都为 2, 将各个受试者在不同方法下得到的分类结果绘于图 9, 每个受试者都选用各自对应的阶数. 图中横坐标代表 10 种不同任务组合方式, 纵坐标代表不同方法 10 次实验结果均值.

从图 9 我们可得到以下结论: 1) 因每个人的思维方式不同, 所以不同的受试者对应的最佳的特征提取方法不同; 2) 对于不同的受试者, 最优的任务组合方式不同; 3) 在传统的 VAR 模型的基础上添加合适的分层结构在某些识别任务中能提高分类性能, 像受试者 1 的第 5 种任务组合方式以及受试者 3 的第 4 种任务组合方式. 结合 4 位受试者来看, 在阶数为 2 时, 具有分层结构的 HVAR 模型并没有表现出明显的优势, 甚至在绝大多数情况下得到的分类结果都不如普通的 VAR 模型高. 此外, 将 EEG 信号的不同通道不同时刻的值看作单一变量的 LASSO-VAR 方法, 虽然能有效地压缩参数空间, 但是作为 EEG 信号的特征提取方法不能得到较为满意的识别效果.

图 9 不能很好地看出各种方法的差异性, 我们进一步采用重复测量方差分析的方法来探究不同特征提取方法的显著性, 并重点分析受试者 1 的实验结果. 首先对不同任务组合方式的分类结果进行方差齐次性检验, 得到的 p 值分别为: 0.1696、0.002571、0.08715、0.3105、0.2482、0.5554、0.6208、0.2263、0.5239、0.913. 由此可知第 2 种任务组合方式的结果不满足方差齐性的要求, 所以不再对其进行重复性方差分析, 对其余 9 种任务组合方式进行分析得到的 p 值为: 1.5E−08、6.59E−08、8.87E−05、2.18E−09、1.52E−05、0.00013、0.0307、0.000102、0.00258. 可见差

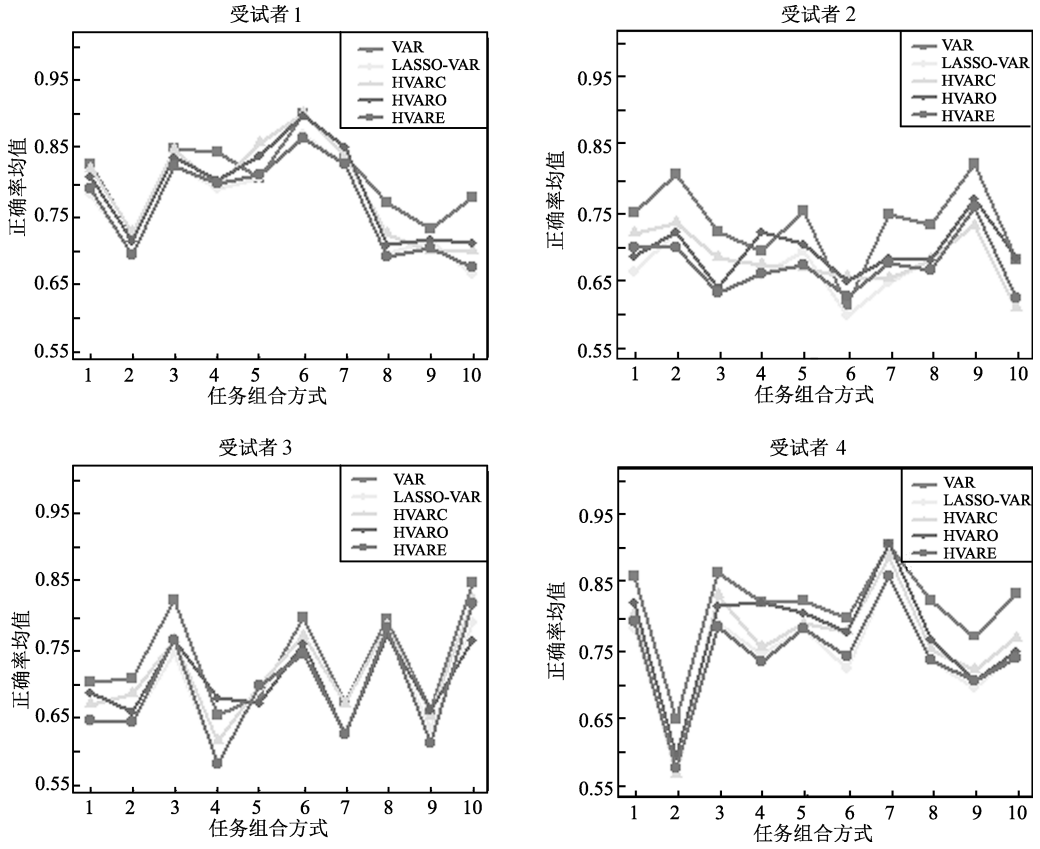

图 9　不同受试者的分类结果

Fig. 9　The results for all subjects

异都较为显著.

表 8 是本文用到的五种特征提取方法与文献 [11] 中方法的分类正确率的比较. 文献 [11] 的分类器选择神经网络. 表中 AR-BG 表示用 Burg 算法求解 AR 系数.

从表 8 可知, 采用多通道的 VAR 相关方法在多数情况下的分类性能都优于单通道的 AR 模型. 此外, 文献 [11] 中 AR 相关方法使用的阶数是 6, 而本文中所有方法选择的阶数都是 2, 由此说明, 多通道 VAR 的相关方法因其能挖掘出通道之间的互信息, 可以在较短的时间范围内就得到最优的分类结果. 对于受试者 1、2、4, 本文的方法得到的最佳任务组合方式与 AR 相同.

在阶数为 2 的时候, VAR 模型的参数空间还不太大, 在这种情况下, 我们并没有看出 HVAR 相关方法的优势. 下面我们分析阶数为 6 时各种方法的实验结果. 图 10 是 4 个受试者在阶数为 6 时五种方法对不同任务的分类性能的箱线图.

从图 10, 我们可以发现: 当参数空间增大时, 对于受试者 1、3、4, 在传统的 VAR 模型的基础上添加合适的分层结构有助于分类性能的提高, 但是所有特征提取方法在受试者 1 的实验中都出现了正偏的情况, 也就是说对于不同的识别任务各种方法都没有太好的稳定性; 从受试者 3 的实验结果来看, HVAR 相关方法与 VAR 的性能基本持平, 就最大值看 HVAR 的三种方法都要远高于 VAR; 对于受试者 4, HVAR 的性能远优于 VAR, 尤其以 HVARC 效果最好, 值分布比较均匀, 也就是说该方法对于不同的任务组合都有较好的识别正确率, 稳定性很好. 总的看来, HVARO 和 HVARC 两种方法的适用范围较广, 在所有受试者上都能取得较好的实验结果, 这说明当参数空间增大的时候, 具有分层结构的 VAR 模型更有优势, 这也反映出在多通道的信号处理中, 不同通道的信号间的最长相互作用时间是不一样的, 且自影响大于交互影响. 但正如前面提到, 每个人的思维方式不同, 对于受试者 2, HVAR 模型的分类正确率反而不如 VAR 高. 此外, 将 EEG 信号的不同通道不同时刻的值看作单一变量的 LASSO-VAR 方法, 虽然能有效地压缩参数空间, 但是在一定程度上会降低分类性能.

图 10 表明在参数空间大的时候, 在传统 VAR 模型上添加罚函数的 HVAR 模型与 VAR 模型的性能相当, 这是因为在阶数大的情况下, 两种模型的参数求解误差都比较大, 造成分类效果相当. 但是, 我们知道 VAR 模型只有在 $T > kp$ 的情况下才能求解参数, 故当时间序列之间的相互作用时间比较长的时候必然会造成 VAR 模型不再适用, 而 HVAR 模型并不受该条件的约束, 因此我们猜想虽然高阶 HVAR 模型对本文的脑电数据并没有实用性, 但是基于其不受 $T > kp$ 条件的约束, 该方法对处理多通道时间序列的分析会有一定的借鉴意义.

4 结论

在 EEG 识别过程中, 特征提取起着非常重要的作用, 提取出来的特征会直接影响系统的分类正确率. 本文对多通道的 EEG 信号建立 HVAR 模型, 将不同的分层结构模型得到的系数作为特征, 并用 Keirn 等采集并整理的脑电数据进行检验, 结果表明, 在阶数远低于文献 [11] 给出的单通道 AR 的最

表 8 不同特征提取方法的结果总结

Table 8 Summary of classification results for all subjects

		AR-BG	VAR	LASSO-VAR	HVARC	HVARO	HVARE
受试者 1	平均值	0.78	0.81	0.77	0.79	0.79	0.77
	最大值	0.83	0.91	0.88	0.91	0.90	0.87
	最佳任务组合方式	乘法计算、几何图旋转	乘法计算、几何图旋转	乘法计算、几何图旋转	乘法计算、几何图旋转	乘法计算、几何图旋转	乘法计算、几何图旋转
受试者 2	平均值	0.74	0.73	0.67	0.68	0.69	0.67
	最大值	0.89	0.82	0.76	0.74	0.77	0.76
	最佳任务组合方式	字母组合、视觉计算	字母组合、视觉计算	字母组合、视觉计算	字母组合、视觉计算	字母组合、视觉计算	字母组合、视觉计算
受试者 3	平均值	0.67	0.73	0.68	0.71	0.70	0.68
	最大值	0.77	0.84	0.78	0.82	0.77	0.81
	最佳任务组合方式	字母组合、几何图旋转	几何图旋转、视觉计算	几何图旋转、视觉计算	几何图旋转、视觉计算	字母组合、几何图旋转	几何图旋转、视觉计算
受试者 4	平均值	0.77	0.82	0.75	0.77	0.78	0.75
	最大值	0.93	0.91	0.86	0.89	0.91	0.86
	最佳任务组合方式	乘法计算、视觉计算	乘法计算、视觉计算	乘法计算、视觉计算	乘法计算、视觉计算	乘法计算、视觉计算	乘法计算、视觉计算

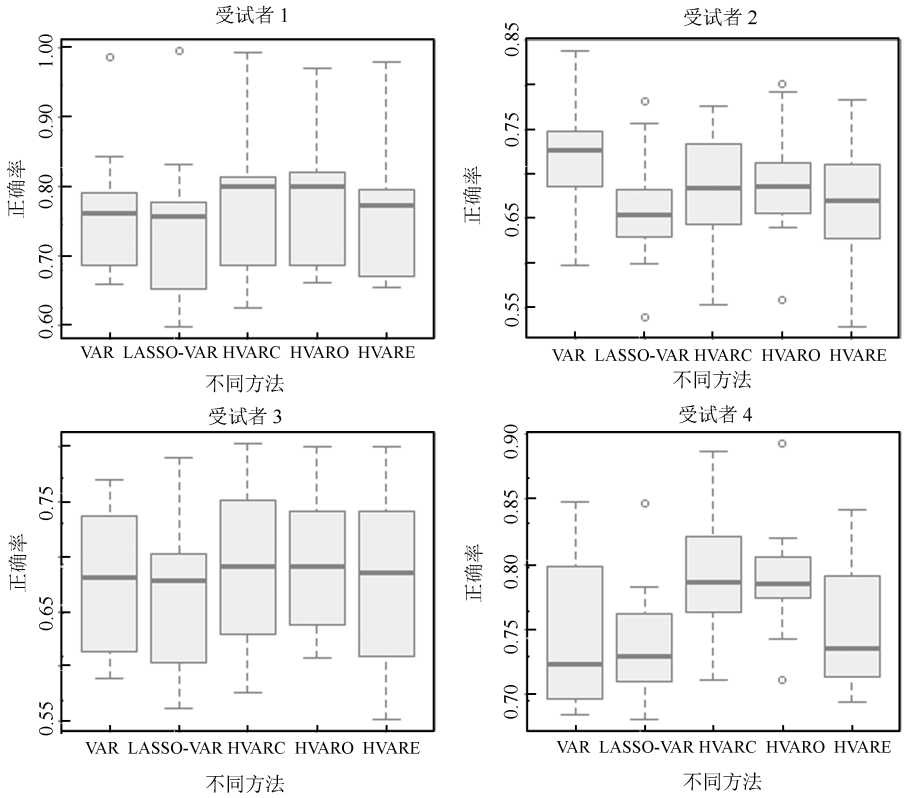

图 10　阶数为 6 时不同方法的分类结果箱线图

Fig. 10　The boxplots of all methods using six order for all subjects

优阶数时, 多通道 VAR 相关方法就可以得到与 AR 模型相近的分类结果. 在高阶的时候, 参数求解误差较大, 分类效果一般, 但 HVAR 模型与 VAR 模型相比优势在于其不受 $T > kp$ 条件的约束, 由于正则化思想的加入, 该模型能够实现变量选择, 剔除无关变量, 压缩参数空间, 在一定程度上降低参数求解误差, 故对于相互作用时间较长的多通道时间序列, HVAR 模型不失为一种新颖的方法.

References

1 Wang Xing-Yu, Jin Jing, Zhang Yu, Wang Bei. Brain control: human-computer integration control based on brain-computer interface. *Acta Automatica Sinica*, 2013, **39**(3): 208−221
(王行愚, 金晶, 张宇, 王蓓. 脑控: 基于脑−机接口的人机融合控制. 自动化学报, 2013, **39**(3): 208−221)

2 Fu Yun-Fa, Wang Yue-Chao, Li Hong-Yi, Xu Bao-Lei, Li Yong-Cheng. Direct brain-controlled robot interface technology. *Acta Automatica Sinica*, 2012, **38**(8): 1229−1246
(伏云发, 王越超, 李洪谊, 徐保磊, 李永程. 直接脑控机器人接口技术. 自动化学报, 2012, **38**(8): 1229−1246)

3 McFarland D J, Wolpaw J R. Brain-computer interfaces for communication and control. *Communications of the ACM*, 2011, **54**(5): 60−66

4 Yang B H, Yan G Z, Yan R G, Wu T. Adaptive subject-based feature extraction in brain-computer interfaces using wavelet packet best basis decomposition. *Medical Engineering & Physics*, 2007, **29**(1): 48−53

5 Shannon M, Zen H, Byrne W. Autoregressive models for statistical parametric speech synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, **21**(3): 587−597

6 Sun Hui-Wen, Fu Yun-Fa, Xiong Xin, Yang Jun, Liu Chuan-Wei, Yu Zheng-Tao. Identification of EEG induced by motor imagery based on Hilbert-Huang transform. *Acta Automatica Sinica*, 2015, **41**(9): 1686−1692
(孙会文, 伏云发, 熊馨, 杨俊, 刘传伟, 余正涛. 基于 HHT 运动想象脑电模式识别研究. 自动化学报, 2015, **41**(9): 1686−1692)

7 Fu Yun-Fa, Xu Bao-Lei, Li Yong-Cheng, Li Hong-Yi, Wang Yue-Chao, Yu Zheng-Tao. Recognition of actual grip force movement modes based on movement-related cortical potentials. *Acta Automatica Sinica*, 2014, **40**(6): 1045−1057
(伏云发, 徐保磊, 李永程, 李洪谊, 王越超, 余正涛. 基于运动相关皮层电位握力运动模式识别研究. 自动化学报, 2014, **40**(6): 1045−1057)

8 Mousavi E A, Maller J J, Fitzferald P B, Lithgow B J. Wavelet common spatial pattern in asynchronous offline brain computer interfaces. *Biomedical Signal Processing and Control*, 2011, **6**(2): 121−128

9 Li P Y, Wang X R, Li F L, Zhang R, Ma T, Peng Y H, Lei X, Tian Y, Guo D Q, Liu T J, Yao D Z, Xu P. Autoregressive model in the *Lp* norm space for EEG analysis. *Journal of Neuroscience Methods*, 2015, **240**: 170−174

10 Jain N, Dandapat S. Constrained autoregressive (CAR) model. In: Proceedings of 2005 Annual IEEE India International Conference. Chennai, India: IEEE, 2005. 255−257

11 Huan N J, Palaniappan R. Neural network classification of autoregressive features from electroencephalogram signals

for brain-computer interface design. *Journal of Neural Engineering*, 2004, **1**(3): 142−150

12 Lawhern V, Hairston W D, McDowell K, Westerfield M, Robbins K. Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. *Journal of Neuroscience Methods*, 2012, **208**(2): 181−189

13 Chen L L, Madhavan R, Rapoport B I, Anderson W S. Real-time brain oscillation detection and phase-locked stimulation using autoregressive spectral estimation and time-series forward prediction. *IEEE Transactions on Biomedical Engineering*, 2013, **60**(3): 753−762

14 Anderson C W, Stolz E A, Shamsunder S. Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Transactions on Biomedical Engineering*, 1998, **45**(3): 277−286

15 Pei X M, Zheng C X. Feature extraction and classification of brain motor imagery task based on MVAR model. In: Proceedings of 2004 International Conference on Machine Learning and Cybernetics. Shanghai, China: IEEE, 2004. 3726−3730

16 Hu X, Nenov V. Multivariate AR modeling of electromyography for the classification of upper arm movements. *Clinical Neurophysiology*, 2004, **115**(6): 1267−1287

17 Wang J, Xu G Z, Wang L, Zhang H Y. Feature extraction of brain-computer interface based on improved multivariate adaptive autoregressive models. In: Proceedings of the 3rd International Conference on Biomedical Engineering and Informatics. Yantai, China: IEEE, 2010. 895−898

18 Zhao C L, Zheng C X, Zhao M, Tu Y L, Liu J P. Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Systems with Applications*, 2011, **38**(3): 1859−1865

19 Heger D, Terziyska T, Schultz T. Connectivity based feature-level filtering for single-trial EEG BCIS. In: Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing. Florence, Italy: IEEE, 2014. 2064−2068

20 Faes L, Erla S, Porta A, Nollo G. A framework for assessing frequency domain causality in physiological time series with instantaneous effects. *Philosophical Transactions of the Royal Society A-Mathematical, Physical and Engineering Sciences*, 2013, **371**(1997): 20110618

21 Varotto G, Fazio P, Rossi Sebastiano D, Duran D, D′Incerti L, Parati E, Sattin D, Leonardi M, Franceschetti S, Panzica F. Altered resting state effective connectivity in long-standing vegetative state patients: an EEG study. *Clinical Neurophysiology*, 2014, **152**(1): 63−68

22 Panzica F, Camafoglia L, Framceschetti S. EEG-EMG information flow in movement-activated myoclonus in patients with Unverricht-Lundborg disease. *Clinical Neurophysiology*, 2014, **125**(9): 1803−1808

23 Wang J J, Zhang Y N. A novel method of multi-channel feature extraction combining multivariate autoregression and multiple-linear principal component analysis. *Journal of Biomedical Engineering*, 2015, **32**(1): 19−24

24 Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 1996, **58**(1): 267−288

25 Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 2006, **68**(1): 49−67

26 Zhao P, Rocha G, Yu B. The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics*, 2009, **37**(6A): 3468−3497

27 Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Image Sciences*, 2009, **2**(1): 183 − 202

28 Jenatton R, Mairal J, Obozinski G, Obozinski G, Bach F. Proximal methods for hierarchical sparse coding. *The Journal of Machine Learning Research*, 2011, **12**: 2297−2334

29 Keirn Z A, Aunon J I. A new mode of communication between man and his surroundings. *IEEE Transactions on Biomedical Engineering*, 1990, **37**(12): 1209−1214

王金甲 燕山大学信息科学与工程学院教授. 主要研究方向为信号处理和模式识别. 本文通信作者.
E-mail: wjj@ysu.edu.cn
(**WANG Jin-Jia** Professor at the School of Information Science and Engineering, Yanshan University. His research interest covers signal processing and pattern recognition. Corresponding author of this paper.)

陈 春 燕山大学信息科学与工程学院硕士研究生. 主要研究方向为信号处理.
E-mail: xkz1124357698@sina.cn
(**CHEN Chun** Master student at the School of Information Science and Engineering, Yanshan University. Her main research interest is signal processing.)