

基于支持向量机和 PCA 的脑电 α 波运动想象分类研究

蔡 靖¹, 刘光达¹, 王尧尧¹, 宫晓宇²

(1. 吉林大学 仪器科学与电气工程学院, 吉林 长春 130012; 2. 吉林大学 教育技术中心, 吉林 长春 130061)

摘要: 针对脑电信号(EEG)运动想象分类过程中弱相关特征量影响分类准确度的问题, 提出一种筛选方法, 该方法是基于 α 波和主成分分析(PCA)算法的。基于脑机接口(BCI)系统, 通过听觉诱发刺激产生向左和向右两种运动想象任务对应的脑电信号, 并对其做小波包分解处理, 然后进行脑电 α 频段信号的重构, 从而提取出 α 波形并对其进行统计特征提取。再结合 PCA 技术和支持向量机(SVM)方法, 实现弱相关特征的剔除和特征分类。根据筛选后的数据进行分类, 所得结果准确率更高, 信号分类的准确度由 90.1% 提高至 94.0%。

关键词: 小波包分解; 支持向量机; 运动想象; 主成分分析; 脑电信号

中图分类号: TN911.7; R318

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.211723

中文引用格式: 蔡靖, 刘光达, 王尧尧, 等. 基于支持向量机和 PCA 的脑电 α 波运动想象分类研究[J]. 电子技术应用, 2022, 48(6): 23-27.

英文引用格式: Cai Jing, Liu Guangda, Wang Yaoyao, et al. Classification of α wave motor imagery based on SVM and PCA[J]. Application of Electronic Technique, 2022, 48(6): 23-27.

Classification of α wave motor imagery based on SVM and PCA

Cai Jing¹, Liu Guangda¹, Wang Yaoyao¹, Gong Xiaoyu²

(1. College of Instrumentation & Electrical Engineering, Jilin University, Changchun 130012, China;

2. Educational Technology Center, Jilin University, Changchun 130061, China)

Abstract: A feature screening method based on alpha wave and principal component analysis was proposed to solve the problem that the weakly correlated feature quantity would affect the classification accuracy in EEG motor imagery classification. Based on brain computer interface system, the EEG signals corresponding to left and right motor imagination tasks were generated by auditory stimulation and processed by wavelet packet decomposition, and then the α band signals of the EEG were reconstructed, so as to extract the α waveforms and extract the statistical features. Combined with PCA technology and SVM method, the weak correlation features are eliminated and classified. According to the selected data, the accuracy of the results is higher, and the accuracy of signal classification is improved from 90.1% to 94.0%.

Key words: wavelet packet decomposition; SVM; motor imagery; PCA; EEG

0 引言

脑电信号 EEG 是大脑中神经元产生的生物电^[1], 不同的运动想象活动中, 大脑释放不同的脑电信号^[2]。脑电波按频率大小分为五大类: α 波(8~14 Hz)、 β 波(14~30 Hz)、 θ 波(4~8 Hz)、 δ 波(4 Hz 以下)和 γ 波(30 Hz 以上)^[3]。本文对脑电信号进行小波分解并提取 α 波^[4], 计算 α 波的多个信号特征, 利用 PCA 技术筛选出强相关特征量, 运用支持向量机进行运动想象分类^[5]。通过实验发现运用小波包变换和 PCA 技术后的分类准确率明显提高。

1 基本原理和算法

1.1 小波包分解

小波包分解以尺度函数 $\phi(t)$ 和小波函数 $\psi(t)$ 为基函数

数进行^[6], 两者的二尺度关系为:

$$\phi(t) = \sqrt{2} \sum_k h_{0k} \phi(2t-k) \quad (1)$$

$$\phi(t) = \sqrt{2} \sum_k h_{1k} \phi(2t-k) \quad (2)$$

式中, h_{0k} 、 h_{1k} 是滤波器系数。小波基为:

$$\varphi_{j,k}^{2i} = \varphi^{2i}((2k-t)/2^i)/\sqrt{2} = \sum_n h(n) \varphi_{j-1,2k-n}^i(t) \quad (3)$$

$$\varphi_{j,k}^{2i+1} = \varphi^{2i+1}((2k-t)/2^i)/\sqrt{2} = \sum_n g(n) \varphi_{j-1,2k-n}^i(t) \quad (4)$$

其中: i 表示节点号, j 表示分解级数; 设 $f(t)$ 表示时间信号, $p_j^i(t)$ 是小波包系数, 表示第 j 层上的第 i 个小波包。小波包分解的算法为^[7]:

$$\begin{cases} p_0^1(t) = f(t) \\ p_j^{2i-1} = \sum_k H(k-2t) p_{j-1}^i(t) \\ p_j^{2i} = \sum_k G(k-2t) p_{j-1}^i(t) \end{cases} \quad (5)$$

1.2 主成分分析(PCA)原理与算法

如果将 n 维空间中的 n 个样本 x_1, x_2, \dots, x_n 看成一个 n 维随机变量, 其矩阵的形式为 $X = [x_1; x_2; \dots; x_n]$, 对 X 的所有列取平均, 得:

$$\mathbf{m}_i = \frac{1}{n} \sum_{j=1}^n (x_i)_j \quad (6)$$

其中, n 表示样本的总数, \mathbf{m} 表示样本的均值^[8]。样本集 X 对应的协方差矩阵 S_i :

$$S_i = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (7)$$

对一个 n 维随机变量 x_i , 经过以下变换:

$$y_i = \mathbf{W}^T(\mathbf{x}_i - \mathbf{m}) \quad (8)$$

得到一个新的 n 维变量 y_i , 其矩阵形式为 $Y = [y_1; y_2; \dots; y_n]^T$, 其中 \mathbf{W} 为投影矩阵。将这一组变量带入下式便可以重构出原始数据:

$$\mathbf{x}_i = \mathbf{W}y_i + \mathbf{m} \quad (9)$$

其中, \mathbf{x}_i 表示重构出的原始数据。

1.3 支持向量机(SVM)原理与算法

若数据样本可以被一线性函数分开, 则其为线性可分的, 称该函数为超平面^[9]:

$$\gamma = \frac{\mathbf{x}_+ \mathbf{w}^T - \mathbf{x}_- \mathbf{w}^T}{\|\mathbf{W}\|} \quad (10)$$

其中, \mathbf{x}_+ 和 \mathbf{x}_- 分别表示两个正负支持向量。SVM 的思想是将间隔最大化, 即:

$$\begin{aligned} & \max_{w, b} \frac{2}{\|\mathbf{W}\|} \\ & \text{s.t. } \gamma_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i=1, 2, \dots, m \end{aligned} \quad (11)$$

最大化 $\frac{2}{\|\mathbf{W}\|}$ 相当于最小化 $\|\mathbf{W}\|$, 将式(11)转化成如下:

$$\begin{aligned} & \min_{w, b} \frac{1}{2} \|\mathbf{W}\|^2 \\ & \text{s.t. } \gamma_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i=1, 2, \dots, m \end{aligned} \quad (12)$$

此公式即为支持向量机。

2 脑电信号特征波提取与分类

2.1 小波包分解方法提取 α 波

由于 α 波处于 8~13 Hz 频段^[10], 频率分辨率很难满足提取的要求, 因此本文采用 db7 小波对脑电信号进行 7 层分解。

2.2 特征提取

2.2.1 均值

对 α 波数据求其均值, 计算均值如式(13)所示:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N x(i) \quad (13)$$

其中, N 为样本点个数。

2.2.2 标准差

对 α 波信号求其方差, 如式(14)所示:

$$\sigma = \sqrt{\frac{1}{N} (x(i) - \bar{X})^2} \quad (14)$$

2.2.3 最值

对 α 波数据提取最大值和最小值作为两个特征进行提取, 计算公式如下:

$$\text{Max}(X_1, X_2, \dots, X_n) \quad (15)$$

$$\text{Min}(X_1, X_2, \dots, X_n) \quad (16)$$

2.2.4 中位数

对 α 波数据提取中位数作为特征进行提取, 计算公式如下:

$$\text{Median}(X_1, X_2, \dots, X_n) \quad (17)$$

2.2.5 功率谱

本文采用周期图法求整体功率谱密度, 其核心是傅里叶变换。由式(3)知, 要得到功率谱密度, 需要求得功率, 通过坐标变化可得功率谱密度。而通过对采样信号做傅里叶变换, 可以得到离散点的幅度, 再通过功率和幅度之间的关系, 便可得到功率:

$$\hat{S}_{\text{PRE}}(\omega) = \frac{1}{N} |U_N(\omega)|^2 \quad (18)$$

2.2.6 能量谱密度

通过傅里叶变换, 将能量信号分离为不同频域分量所对应的能量, 通过对 ω 积分得到信号的总能量, 由此, $|X(\omega)|^2$ 便可定义为能量谱密度, 也常称其为能量谱, 表示能量在某一频率上的分布集度:

$$E = \int_{-\infty}^{\infty} s^2(t) dt = \int_{-\infty}^{\infty} |s(f)|^2 dt \quad (19)$$

式中, E 为能量信号 $s(t)$ 的能量值, $|s(f)|^2$ 为能量谱密度。

2.2.7 自相关函数

随机信号 $x(t)$ 在不同的时刻具有一定相关性, 自相关函数描述随机信号 $x(t)$ 与 t 的取值间相关度。函数中具有周期性分量, 自相关函数能更好地反映这种周期性, 其定义为随机信号在两个不同时刻乘积的期望。

$$R_{xx}(\tau) = E\{x(t)x^*(t-\tau)\} \quad (20)$$

2.3 PCA 降维与 SVM 分类

本文共提取了 α 波数据的 8 个特征, 但并不是所有数据特征都对信号分类有着积极的作用, 因此本文采用 PCA 方法对特征数据进行降维, 将 8 维数据集降至 3 维, 这样不仅大大降低了分类的复杂度, 同时也提高了信号分类的可靠性。本文采用线性 SVM 对特征数据进行二分类, 利用降维后的 3 组特征分类得到左向和右向两种脑电信号。首先利用降维之前的 8 种特征进行分类, 再利用降维之后的 3 种特征进行分类, 对比分类效果。

3 实验过程与结果分析

3.1 数据采集及预处理

受试对象是一名 25 岁身体健康且无精神疾病的男性, 听觉正常, 测试环境温湿度正常, 安静无明显杂音。本次测试采用听觉诱发方式, 采集数据过程总时长 15 s。0~5 s 内, 由录音机播放预先录制好的音频指令, 指示受试者运动想象任务; 5~10 s 为放松准备时间; 10~13 s 进行脑电信号的采集, 计算机自动存储数据; 13 s 后采集的数据噪声较大, 不予采用。实验过程如图 1、图 2 所示。

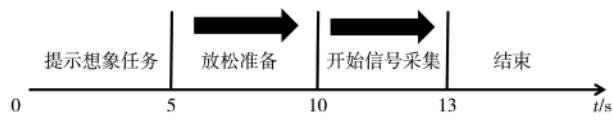


图 1 右向实验过程

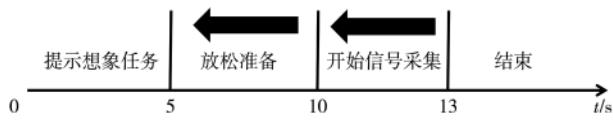


图 2 左向实验过程

采集到原始数据后, 由于脑电信号主要集中在 0~200 Hz 的低频段且存在 50 Hz 工频干扰, 因此需要对原始数据进行低通滤波和 50 Hz 陷波。本文设计了截止频率为 250 Hz 的低通滤波器和 50 Hz 的陷波器, 保留了富含脑电信号的主要频段并初步滤除了工频干扰, 以便后续进一步对脑电数据进行小波包分析。实验过程中采集了四导联脑电信号, 通过反复测量信号幅度并比对不同区域脑信号的特点及频率范围, 可以确定所测得的信号是真实的脑电信号, 如图 3 所示, 图中各点是脑电信号采集过程中常用的采集点, 其中 A_1 和 A_2 常用来做参考电极(耳夹电极), T_3 和 T_4 位于太阳穴位置, 信号强度较高, 也可用于脑电采集。考虑到大脑 C_3 区域的脑电信号对运动想象信号较为敏感, 故采用 C_3 区域采集的信号进行运动想象任务分析。

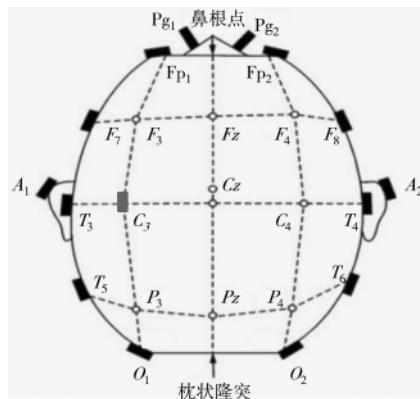


图 3 电极位置示意图

图 4 表示脑电 4 个通道提取出的电信号, 幅度采用归一化之后的幅度, 这样可以方便进行比较, 本文通过

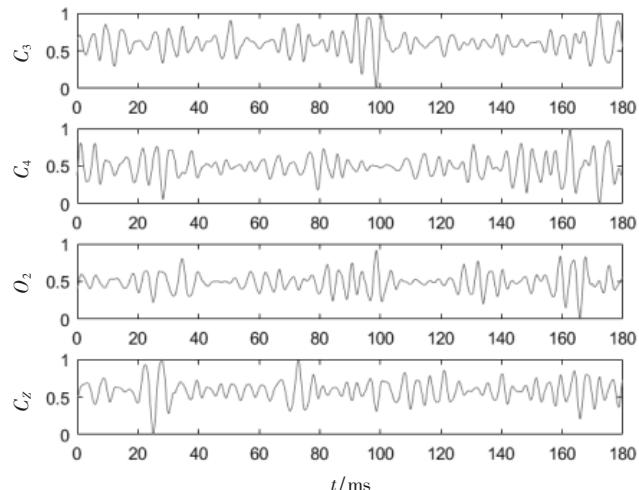


图 4 原始脑电波信号

对 C_3 通道的脑电信号进行数值分析, 来实现运动想象分类任务。

3.2 算法过程及数据分析

对 C_3 通道脑电进行 7 层小波包分解, α 波处于 8~13 Hz 频段^[11], 故采用节点 $[7, 5]$ 、 $[7, 6]$ 和 $[7, 7]$ 对信号进行重构。小波包分解树状图如图 5 所示, 为简明表示, 这里仅展示 3 层小波包分解的树状图, 重构得到的脑电 α 波左右手运动想象信号如图 6 所示。

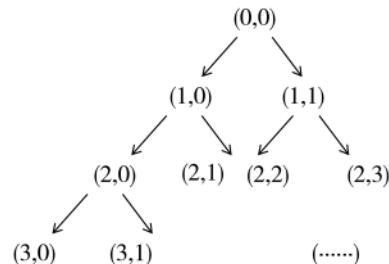
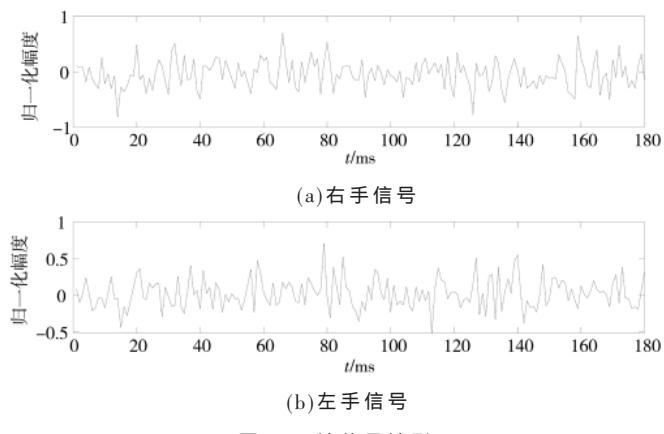
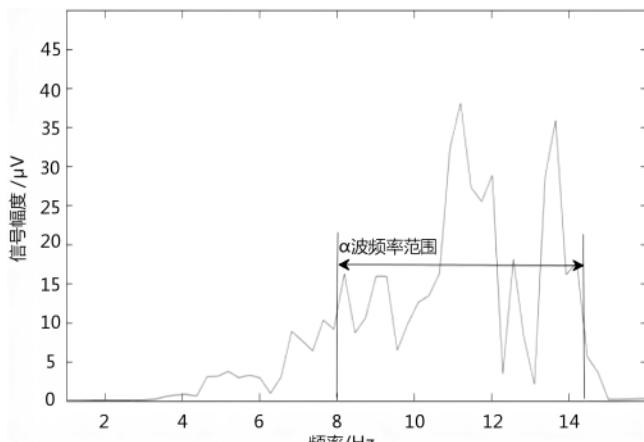


图 5 树状分解

图 6 α 波信号波形

对 α 波信号进行频谱分析, 发现其频率范围主要分布在 8~13 Hz, 如图 7 所示。这证明了通过小波包方法提取的 α 波是可靠的。

图 7 α 波幅频特性

提取多个脑电 α 波的数据特征^[12]，均值、标准差、最大值、最小值、中位数、功率谱、能量谱、自相关函数共计 8 个数据特征。为使各特征数据离相关，使用 PCA 方法将 8 维数据降维至 3 维，如图 8、图 9 所示，降维后的数据彼此相互正交，更具代表性。

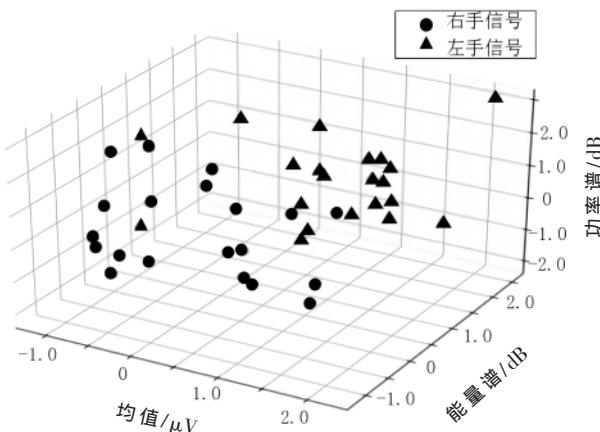


图 8 PCA 三特征运动想象分类

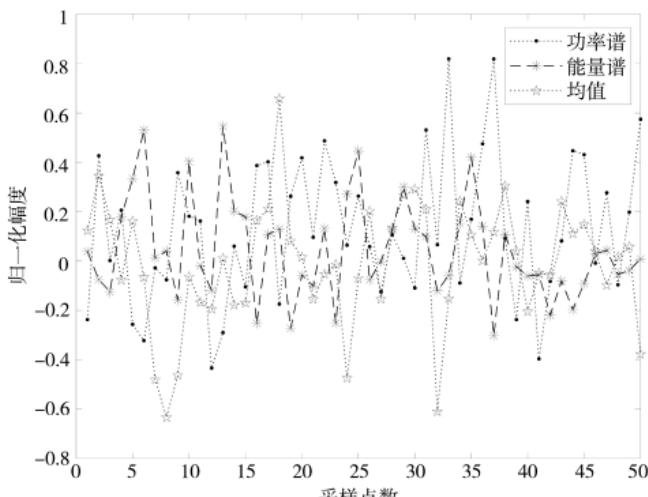


图 9 相关特征波形

3.3 分类效果比较

如图 10、图 11 所示不同灰度分别表示左右手信号脑电 α 波的特征点数据(上方数据点表示右手信号，下方则表示同一次实验中的左手信号)。对比可以发现降维之后的分类准确率明显上升。首先用降维之前的数据特征进行 SVM 分类，然后利用 SVM 方法对剩余的 3 个特征进行线性分类^[13]，分类结果投影到第一和第二个特征所在的平面，信号分类的准确度由 90.1% 提高至 94.0%，如表 1 所示。

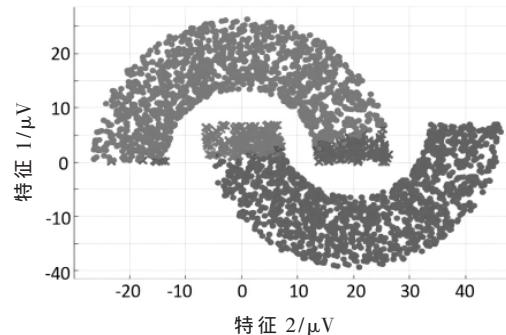


图 10 8 特征 SVM 分类示意图

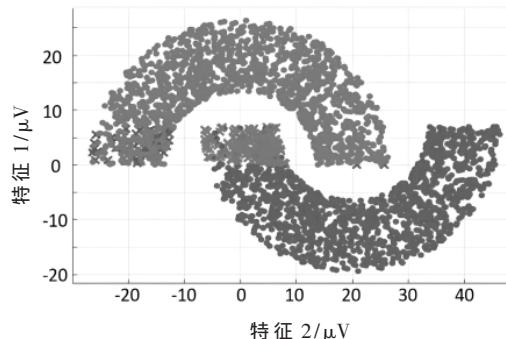


图 11 3 特征 SVM 分类示意图

表 1 PCA 前后分类准确率对比

| 特征数 | 是否经过 PCA 降维 | 分类准确率/% |
|-----|-------------|---------|
| 8 | 否 | 90.1 |
| 3 | 是 | 94.0 |

4 结论

为了提高基于脑电信号的运动想象分类准确率，本文将主成分分析方法用于脑电特征降维，并辅助以小波分析方法实现了运动想象脑电分类，实验结果证明本文提出的方法能明显提高分类准确率，有较高的可行性。

此外，本文采用支持向量机方法进行脑电运动想象数据二分类任务，而没有采用通常使用的 BP 神经网络^[14]，因为支持向量机方法更适用于二分类任务^[15]，此方法在简单的脑电分类任务中也有着较强的参考价值。

参考文献

[1] 林伟成. 基于 EEG 信号的认知任务模式分类研究[D]. 杭州：杭州电子科技大学，2013：12–13.

[2] ZHAI W W, YANG Y N, LU SH Y, et al. Brain computer interface system of the upper limb rehabilitation training robot[J]. *Biomedical Engineering Research*, 2019, 38(3): 269–274.

[3] 姚健康, 熊根良. 小波包分析和 FastICA 相结合对单通道脑电信号的去噪研究[J]. *现代电子技术*, 2021, 44(7): 60–65.

[4] 李昭阳, 伏云发. 基于 EEG 微状态方法的视觉想象识别研究[J]. *计算机工程与科学*, 2021, 43(3): 465–472.

[5] 白雪冬. 小儿病毒性脑炎中采用 BEAM、EEG 诊断的临床价值研究[J]. *世界最新医学信息文摘*, 2019, 19(51): 189–190.

[6] BOZHOKIN S V, SUSLOVA I B. Wavelet-based analysis of spectral rearrangements of EEG patterns and of non-stationary correlations[J]. *Physica A: Statistical Mechanics and its Applications*, 2015, 421(1): 151–160.

[7] LEE W L, TAN T, LEUNG Y H. An improved P300 extraction using ICA-R for P300-BCI speller[C]//International Conference of the IEEE Engineering in Medicine & Biology Society, 2013: 7064–7067.

[8] 姚红革, 杜亚勤, 刘洋. 基于小波分析和 BP 神经网络的图像特征提取[J]. *西安工业大学学报*, 2008, 28(6): 568–572.

[9] 白宇辰, 尹华超, 刘三军. 基于小波分析的改进 KDJ 指标[J]. *电子科技*, 2013, 26(8): 13–15.

[10] 张德丰. *MATLAB 小波分析*[M]. 北京: 机械工业出版社,

2009.

[11] 吕赵明, 张颖江. 改进 GOA-sVM 算法的异常流量识别[J]. *湖南科技大学学报(自然科学版)*, 2019, 34(4): 90–96.

[12] 贺王鹏, 杨琳, 王芳, 等. 基于 TQWT 的癫痫脑电信号的识别[J]. *生物医学工程研究*, 2017, 36(4): 346–350.

[13] GHADERYAN P, ABBA A, SEDAAGHI M H. An emcient seizure prediction method using KNN-based undersampling and linear frequency the measures[J]. *Journal of Neum-science Methods*, 2014, 232: 134–142.

[14] FISHER R S, BOAS W, BLUME W, et al. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy(ILAE) and the intemational bureau for the epilepsy(IBE)[J]. *Epilepsia*, 2005, 46(4): 470–472.

[15] Wu Wei, Gao Xiaorong, Gao Shangkai. One-versus-therest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case[C]//2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. Shanghai, China: IEEE, 2005: 2387–2390.

(收稿日期: 2021-05-07)

作者简介:

蔡靖(1980-), 男, 硕士, 主要研究方向: 精密仪器与机械。

刘光达(1965-), 男, 博士, 教授, 主要研究方向: 医疗仪器。

王尧尧(1999-), 男, 硕士研究生, 主要研究方向: 生物信号检测。



扫码下载电子文档

(上接第 22 页)

过程追溯管理平台、农产品身份证信息管理平台、农业投入品管理系统等, 推行农产品质量安全监管联动模式, 推进农产品质量安全移动监管系统与农用地图对接融合, 实现农产品质量安全从传统监管向信息化监管转变。二是农业金融风险管控, 以数字技术提升农业贷款、保险的风险管控能力, 为农村金融供给双方提供专业信息服务。通过农村信用体系建设, 打通信用与金融的对接通道, 涉农金融机构开展网上信用建档、评定、授信等工作, 结合当地产业特征和风险防范需要, 创新创优信贷产品, 构建农业保险线上化、智能化、电子化、数据化服务模式, 有效激活农村金融市场活力。

参考文献

[1] 中国共产党第十九届中央委员会第五次全体会议公报[Z]. 2020.

[2] 农业农村信息化专家咨询委员会. 中国数字乡村发展报告(2019)[R]. 2019.

[3] 农业农村信息化专家咨询委员会. 中国数字乡村发展报告(2020 年)[R]. 2020.

[4] 农业农村部市场与信息化司, 农业农村部信息中心. 2020 全国县域数字农业农村发展水平评价报告[R]. 2020.

[5] 前瞻产业研究院. 2020 年中国粮食供需安全研究报告[R]. 2020.

[6] 农业农村部网站. 发展智慧农业 建设数字乡村[DB/OL]. (2020-04-xx)[2022-02-22]. http://www.jhs.moa.gov.cn/zlyj/202004/t20200430_6342836.htm.

[7] 国务院网站. 全国 98% 以上的农业经营主体仍是小农户[DB/OL]. (2019-03-xx)[2022-02-22]. http://www.gov.cn/xinwen/2019-03/01/content_5369755.htm.

[8] 商务部. 主要消费品供需状况统计调查分析报告[R]. 2019.

[9] 光明网. 我国农产品加工产业科技贡献率达到 63% [DB/OL]. (2021-03-xx)[2022-02-22]. <https://m.gmw.cn/bajia/202103/25/34714169.html>.

(收稿日期: 2022-02-22)

作者简介:

宋姝媛(1992-), 通信作者, 女, 硕士, 工程师, 主要研究方向: 数字城市、数字经济, E-mail: songshy120@163.com。

范国浩(1991-), 男, 硕士, 工程师, 主要研究方向: 数字城市、数字经济、数据要素化。

王岩(1989-), 男, 博士, 工程师, 主要研究方向: 数据要素、数字城市、数字经济、工业互联网。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所