

DOI:10.3969/j.issn.1005-202X.2021.11.013

医学信号处理与医学仪器

基于多重分形去趋势波动分析的脑电信号特征提取及分类方法

陈敬凯¹, 孟雪¹, 王常青¹, 钟亚鼎²

1. 安徽医科大学生物医学工程学院, 安徽 合肥 230032; 2. 安徽医科大学第一附属医院放射科, 安徽 合肥 230032

【摘要】目的:针对脑电信号普遍存在的数据维度高、难以预测的问题,提出一种多重分形去趋势波动分析特征提取方法与长短时记忆网络(LSTM)相结合的脑电信号分类方法。**方法:**首先对信号样本进行多重分形去趋势波动分析计算得到脑电信号样本的多重分形谱,计算广义Hurst指数 h_q 和广义维数 D_q 之间的函数关系;然后对多重分形谱进行分析,找出最具代表性的坐标值作为信号的特征向量;最后将其用于LSTM进行训练和分类测试。实验采用波恩大学采集的经过处理的癫痫脑电数据集。**结果:**当训练样本占总体样本比例超过10%之后,LSTM分类器的测试准确率均稳定在98%以上;当占比超过80%时LSTM分类器的测试准确率达到100%;即使训练样本较少时也有95%之上的准确率。**结论:**该算法有良好的准确率和稳定性。

【关键词】脑电信号;多重分形去趋势波动;长短时记忆网络;特征提取;信号分类

【中图分类号】R318;TP301.6

【文献标志码】A

【文章编号】1005-202X(2021)11-1387-05

Feature extraction and classification of electroencephalogram signal based on multifractal detrended fluctuation analysis

CHEN Jingkai¹, MENG Xue¹, WANG Changqing¹, ZHONG Yading²

1. School of Biomedical Engineering, Anhui Medical University, Hefei 230032, China; 2. Department of Radiology, the First Affiliated Hospital of Anhui Medical University, Hefei 230032, China

Abstract: Objective To propose a electroencephalogram (EEG) signal classification method based on the combination of feature extraction by multifractal detrended fluctuation analysis (MF-DFA) and long short-term memory network (LSTM) for solving the problems existing in EEG signal such as high data dimensionality and difficulty in prediction. **Methods** The multifractal spectrum of the EEG signal samples was firstly obtained by MF-DFA, and the functional relationship between the generalized Hurst exponent h_q and the generalized dimensionality D_q was calculated. Then the multifractal spectrum was analyzed to find the most representative coordinate value as the signal eigenvector. Finally, the obtained signal eigenvector was used for LSTM training and classification test. The experiment was carried out on a processed epileptic EEG data set collected by University of Bonn. **Results** When the training samples accounted for more than 10% of the total samples, the test accuracy of LSTM classifiers stabilized at 98% and above; and when the proportion was more than 80%, the test accuracy of LSTM classifier reached 100%. Even with a small number of training samples, the accuracy was higher than 95%. **Conclusion** The proposed algorithm has good accuracy and stability.

Keywords: electroencephalogram signal; multifractal detrended fluctuation; long short-term memory network; feature extraction; signal classification

前言

脑电活动与大脑的生长及发展状况有着密切的联系^[1-2]。对脑电信号的研究探索是认识脑活动机制、人的认知和学习机理机制、人的脑活动与机体活动的关系以及诊断脑部和精神疾病的重要手段,利用计算机技术对脑电信号进行处理和分析可以为医生提供快速有效的诊断依据^[3-5]。关于脑电信号的特征提取与分类方法已经有许多学者做出了很多成果。较常见的处理方法有时频域分析^[6]、小波变

【收稿日期】2021-05-15

【基金项目】国家自然科学基金(62001005);安徽省高校自然科学基金项目(KJ2017A209);安徽省自然科学基金(2008085QH425);安徽医科大学科研基金(XJ201811)

【作者简介】陈敬凯,硕士研究生,研究方向:医学信号处理,E-mail: 282078753@qq.com

【通信作者】孟雪,硕士,讲师,研究方向:生物医学信号及信息处理,E-mail: mengxue@ahmu.edu.cn

换^[7-8]、人工神经网络^[9-10]、支持向量机^[11-13]和非线性动力学分析^[14-15]等。但现有的脑电数据普遍存在维度高,数据难以预测解释等特点,如何提高分类准确率和算法的稳定性是脑电信号分类研究中需要继续思考的问题。

本研究利用多重分形去趋势波动分析算法(Multifractal Detrended Fluctuation Analysis, MF-DFA)来提取多尺度特征,并将其与长短时记忆网络(Long Short-Term Memory Network, LSTM)结合起来对脑电信号进行分类,最后通过相关的实验来验证该方法的有效性和可行性。实验流程如图1所示。

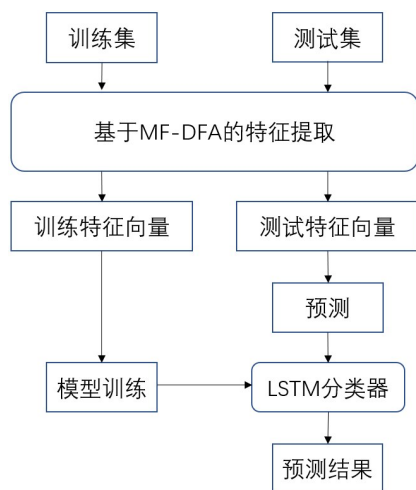


图1 分类实验流程图

Fig.1 Flowchart of classification experiment

1 MF-DFA 分析

MF-DFA 是由 Kantelhardt 等^[16]提出的一种可以用于分析脑电信号多尺度特征的分析方法。通过该方法可以得到信号样本的多重分形谱,即广义 Hurst 指数 hq 与广义维数 D_q 之间的函数关系,然后从中找出信号样本类间差异较大的标量作为特征向量用于分类器分类^[17-19]。

对于长度为 N 的时间序列 $\{x_k\}$, $k = 1, 2, \dots, N$, MF-DFA 计算步骤如下:

第一步,计算序列样本 $\{x_k\}$ 的平均值:

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N x_k \quad (1)$$

第二步,确定信号样本的累计离差:

$$Y(i) = \sum_{k=1}^i [x_k - \bar{x}] \quad (2)$$

其中, $i = 1, 2, \dots, N$ 。

第三步,将第二步所得的累计离差序列 $Y(i)$ 划分成 N_s 个小区间。其中 $N_s = \text{int}\left(\frac{N}{s}\right)$ 。如果 N 不能整除 s ,

$Y(i)$ 将会有一段数据没有被使用。为了能够充分利用数据样本而不造成数据丢失,需要对序列的剩余部分重复这一划分过程。最终得到 $2N_s$ 个等长小区间,将数据样本的所有数据都充分利用起来以达到最佳效果。

第四步,将第三步中划分所得的每个等长小区间内的 s 个点进行最小二乘法的 k 阶多项式拟合:

$$y_v(i) = a_1 i^k + a_2 i^{k-1} + \dots + a_k i + a_{k+1} \quad (3)$$

其中 $i = 1, 2, \dots, s; k = 1, 2, \dots$ 。

第五步,计算均方误差。设区间为 $v = 1, 2, \dots, 2N_s$, 计算均方误差 $F^2(s, v)$:

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s \{Y[(v-1)s+i] - y_v(i)\}^2 \quad (4)$$

当 $v = N_s + 1, N_s + 2, \dots, 2N_s$, 计算均方误差 $F^2(s, v)$:

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s \{Y[N - (v - N_s)s + i] - y_v(i)\}^2 \quad (5)$$

第六步,对去趋势后的 $F^2(s, v)$ 取平均值,则可得 q 波动函数 $F_q(s)$:

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(s, v)]^{\frac{q}{2}} \right\}^{\frac{1}{q}} \quad (6)$$

其中, q 为任意不为零的实数, $F_q(s)$ 随着 s 的增大,以幂律关系递增,即 $F_q(s) \propto s^{h(q)}$ 。则对应每一个 s ,都有一个对应的函数值 $F_q(s)$,对于 $\ln[F_q(s)] - \ln s$ 函数关系图中的斜率即为广义 Hurst 指数 hq 。不同阶波动函数下 F_q 和 s 间的关系如图2所示。 q 阶广义 Hurst 指数 hq 与阶数 q 的关系如图3所示。

第七步,计算质量指数:

$$\tau(q) = qh(q) - 1 \quad (7)$$

第八步,计算广义维数:

$$D_q = \frac{\tau(q)}{q - 1} \quad (8)$$

通过上述计算过程,可得出在波动函数不同的矩阶数 q 下信号样本的多重分形谱,如图4所示。

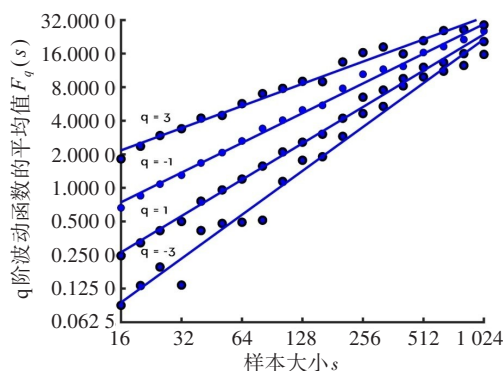


图2 不同阶波动函数下 F_q 和 s 间的关系

Fig.2 Relationship between F_q and s under different order wave functions

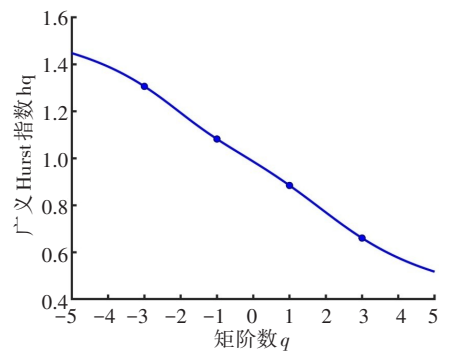


图 3 广义 Hurst 指数 h_q
Fig.3 Generalized Hurst exponent h_q

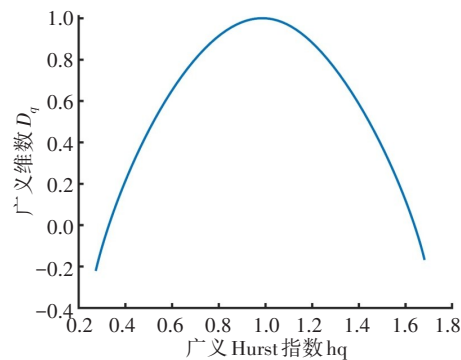


图 4 信号样本的多重分形谱
Fig.4 Multifractal spectrum of signal samples

2 LSTM分类训练与测试

LSTM 是对循环神经网络 (Recurrent Neural Network, RNN)进行改进之后的结果^[20-21],因此,它的网络结构和模型参数都与 RNN 很相像。LSTM 和 RNN 在深度学习中常用于处理时序信息。

本研究采用 LSTM 对 MF-DFA 所提取的多重分形特征向量进行分类训练和测试。使用数学工具 Matlab 中的深度学习神经网络设计工具箱来设计 LSTM 分类模型,结构如图 5 所示。

LSTM 的核心第一层的特征序列输入层和第二层的 LSTM 处理层和最后一层的分类输出层,第一层输入层将脑电信号的特征向量导入神经网络。第二层 LSTM 层对特征向量进行分析,找出其中的相关性用于数据分类,最后一层的分类输出层会输出分类结果。

LSTM 的运作从序列输入层输入特征值序列开始,然后是 LSTM 层对输入的特征序列进行分析。LSTM 的用途是对信号样本进行分类,所以该网络的末尾是一个分类输出层 classoutput。中间加了一个 dropout 层是为了防止过拟合的情况出现,此处 dropout 的可能性参数为 0.5。

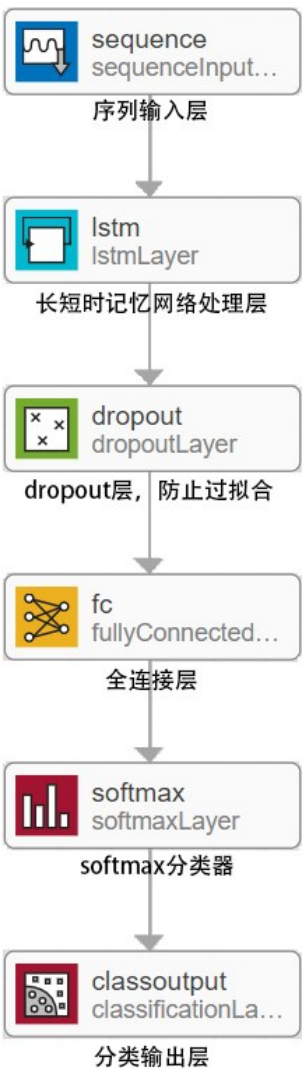


图 5 LSTM 结构示意图
Fig.5 Schematic diagram of long short-term memory network (LSTM) structure

3 实验结果与分析

3.1 数据描述

实验数据为波恩大学医院临床采集的癫痫脑电数据集。该数据集由 5 个子集组成。每个子集由 100 个数据样本构成。每个子集对应一种类别的脑电信号,分别为 Z 类、O 类、N 类、F 类、S 类。

针对 Z 类和 S 类两种类别的脑电信号进行分类实验,其中 Z 类数据为正常人的脑电波,而 S 类数据为癫痫患者发病时的脑电波。两类样本示例如图 6 所示。

3.2 脑电信号的特征提取

此次实验是对特征提取方法和分类器的有效性和可行性进行探索,因此首先利用 MF-DFA 对脑电信号进行特征提取,然后进行脑电信号分类。从 Z 类和 S 类中各取两个样本进行 MF-DFA 计算所得到的多重分形谱如图 7 所示。

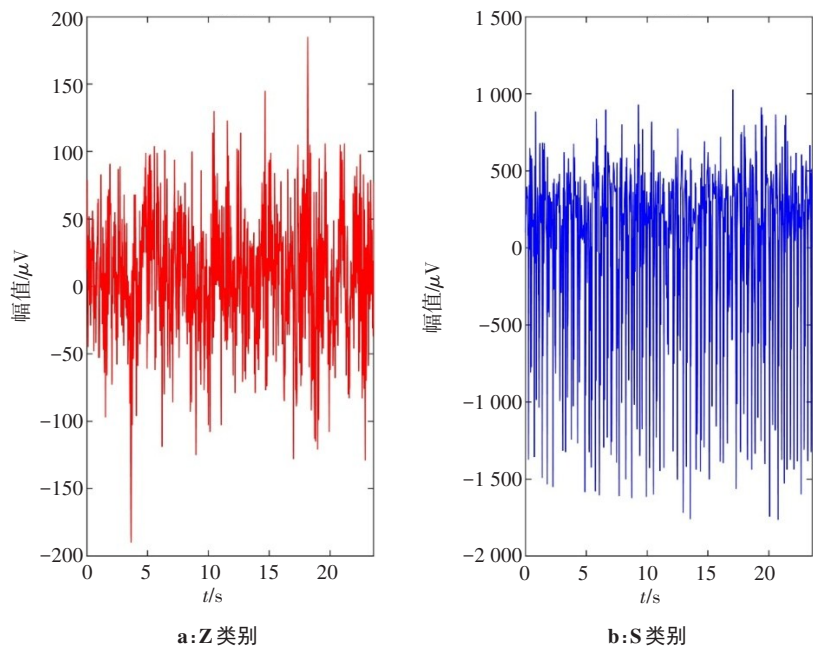


图6 Z与S两类脑电信号示例

Fig.6 Examples of two types of EEG signals, namely Z-type and S-type

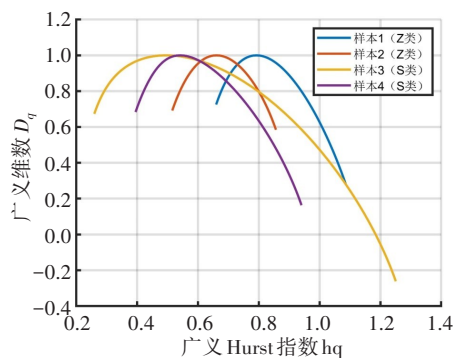


图7 Z类和S类中各取两个样本所得的多重分形谱

Fig.7 Multifractal spectrum obtained by taking two samples in each of Z-type and S-type

4个样本的多重分形谱中,根据 D_q 和 h_q 两者的变化,可以看到两个类别的样本有显著的不同。Z类样本中 h_q 的最大值和最小值之差要小于S类样本,同时 h_q 最大值和最小值对应的 D_q 值之差也要小于S类样本。每个样本数据的多重分形谱中可找到 h_q 最大值对应的点、 h_q 最小值对应的点以及 D_q 最大值对应的点,这3个点基本可以反映出该样本数据的多重分形特征。从物理意义上讲, h_q 最大值对应的点及最小值对应的点分别对应多重分形中的最大和最小奇异指数,分别展现了脑电信号在低概率测度子集和高概率测度子集的奇异程度。 D_q 最大值对应的点对应的 h_q 值,则为奇异性的值。值越小说明信号越平滑,反之,说明信号细节越复杂。因此,取这3个点的坐标作为该样本的特征,如样本1的特征值为(1.086 5,0.272 1)、(0.790 1,1.000 0)、(0.659 9,0.725 3)。

3.3 分类实验结果

将100个Z类和100个S类样本数据按照不同比例划分训练集和测试集进行分类。LSTM分类结果如表1所示。

表1 LSTM分类结果

Tab.1 LSTM classification results

训练数量(Z:S)	测试数量(Z:S)	测试集上的准确率/%
5:5	95:95	95.79
10:10	90:90	98.33
50:50	50:50	98.00
80:80	20:20	100.00
100:100	20:20	100.00

表1第一行表示从两个类别中各取5个数据作为训练样本,另外的190个样本作为测试集合,准确率为95.79%。以此类推,最后一行表示200个数据全部用来训练,再从中挑选20个数据进行测试,最后所得的准确率为100%。

从表1中可以看出,当训练样本达到20个,占样本总数的10%,算法的准确率就可高达98%。当训练样本超过160个,占样本的总数的80%,准确率达到100%。可以看到分类器在不同比例的训练集和测试集中都有良好的表现并且具有较高的分类准确率和稳定性,该方法的有效性和可行性得到了验证。

4 结 语

对脑电信号进行特性分析和分类研究对人类认知大脑运行机制和处理相关疾病具有重大而深远的意义。实验使用波恩数据集中的癫痫患者和健康者的脑电信号样本进行二分类,对算法的有效性和可行性进行了验证。通过对每个脑电数据样本进行特征提取来实现通过少量的数据反映样本的特性,大大降低分类器学习过程中的计算量,也可以防止无意义数据对分类器的影响。MF-DFA 提取出的特征向量能够很好地反映信号样本的差异。并且,在特征提取的时间成本上,MF-DFA 分析算法要远小于其多种特征分析算法组合。因此,该方法还存在计算量少、特征数据少、耗时少,且物理意义清晰的优势。LSTM 一般直接用来对时间信号样本进行分类训练和测试,不过这样会因为时间信号纬度高体量大特点,使得网络的训练成本非常高,需要耗费大量的时间进行迭代的同时,成功率也只有 90%^[22]。本研究将 MF-DFA 和 LSTM 结合起来用于脑电信号的分类实验,将分类器的训练成本极大降低,成功率也大幅提升。该结果为癫痫疾病的精确诊断提供了有效的辅助信息,也为脑电信号的异常检测提供参考依据。

【参考文献】

- [1] ZHANG X, WU D R, DING L Y, et al. Tiny noise, big mistakes: adversarial perturbations induce errors in brain-computer interface spellers[J]. Natl Sci Rev, 2021, 8(4): 78-90.
- [2] ZHAO X, WANG Y B, ZHANG Y, et al. Propofol-induced anesthesia alters corticocortical functional connectivity in the human brain: an EEG source space analysis[J]. Neurosci Bull, 2021, 37(4): 563-568.
- [3] 巫嘉陵, 高忠科. 脑机接口技术及其在神经科学中的应用[J]. 中国现代神经疾病杂志, 2021, 21(1): 3-8.
WU J L, GAO Z K. Brain-computer interface technology and its application in neuroscience[J]. Chinese Journal of Contemporary Neurology and Neurosurgery, 2021, 21(1): 3-8.
- [4] 葛松, 徐晶晶, 赖舜男, 等. 脑机接口: 现状、问题与展望[J]. 生物化学与生物物理进展, 2020, 47(12): 1227-1249.
GE S, XU J J, LAI S N, et al. Brain-computer interface: current status, problems and prospects[J]. Progress in Biochemistry and Biophysics, 2020, 47(12): 1227-1249.
- [5] 伏云发, 丁鹏, 苏磊, 等. 国内外脑电分析处理软件现状分析及发展趋势[J]. 昆明理工大学学报(自然科学版), 2021, 46(1): 54-67.
FU Y F, DING P, SU L, et al. Analysis and development trend of EEG analysis and processing software at home and abroad[J]. Journal of Kunming University of Science and Technology (Natural Science Edition), 2021, 46(1): 54-67.
- [6] 武侠, 钟楚鹏, 丁玉琬, 等. 利用时频分析研究非相位锁定脑电活动[J]. 心理科学进展, 2018, 26(8): 1349-1364.
WU X, ZHONG C P, DING Y L, et al. Application of time-frequency analysis in investigating non-phase locked components of EEG[J]. Advances in Psychological Science, 2018, 26(8): 1349-1364.
- [7] DASH D P, KOLEKAR M. Hidden Markov model based epileptic seizure detection using tunable Q wavelet transform[J]. J Biomed Res, 2020, 34(3): 70-179.
- [8] SLIMEN I B, BOUBCHIR L, MBARKI Z, et al. EEG epileptic seizure detection and classification based on dual-tree complex wavelet transform and machine learning algorithms[J]. J Biomed Res, 2020, 34(3): 151-161.
- [9] AFRAKHTEH S, MOSAVI M R, KHISHE M, et al. Accurate classification of EEG signals using neural networks trained by hybrid population-physic-based algorithm[J]. Int J Automat Comput, 2020, 17(1): 108-122.
- [10] CUI S, DUAN L J, GONG B, et al. EEG source localization using spatio-temporal neural network[J]. China Commun, 2019, 16(7): 131-143.
- [11] 张宪法, 郝矿荣, 陈磊. 免疫多域特征融合的多核学习 SVM 运动想象脑电信号分类[J]. 自动化学报, 2020, 46(11): 2417-2426.
ZHANG X F, HAO K R, CHEN L. Multi-core learning SVM motor imagery EEG signal classification based on immune multi-domain feature fusion[J]. Acta Automatica Sinica, 2020, 46(11): 2417-2426.
- [12] 刘宝, 蔡梦迪, 薄迎春, 等. 一种基于 PSO-CSP-SVM 的运动想象脑电信号特征提取及分类算法[J]. 中南大学学报(自然科学版), 2020, 51(10): 2855-2866.
LIU B, CAI M D, BO Y C, et al. A feature extraction and classification algorithm based on PSO-CSP-SVM for motor imagery EEG signals[J]. Journal of Central South University (Science and Technology), 2020, 51(10): 2855-2866.
- [13] 王玉满, 姜威, 刘治, 等. 基于共空间模式算法和支持向量机二重分类的癫痫发病预测[J]. 生物医学工程学杂志, 2021, 38(1): 39-46.
WANG Y X, JIANG W, LIU Z, et al. Prediction of epilepsy based on common spatial model algorithm and support vector machine double classification[J]. Journal of Biomedical Engineering, 2021, 38(1): 39-46.
- [14] 雷灵, 杨勇, 侯娜, 等. 基于非线性时空动力学的意识障碍患者脑电图研究[J]. 中国生物医学工程学报, 2021, 40(1): 60-70.
LEI L, YANG Y, HOU N, et al. EEG study of patients with disorder of consciousness based on nonlinear spatiotemporal dynamics[J]. Chinese Journal of Biomedical Engineering, 2021, 40(1): 60-70.
- [15] 翟晓雪, 张皓. 非线性动力学分析方法在神经康复领域中的应用进展[J]. 中国康复医学杂志, 2019, 34(4): 483-486.
ZHAI X X, ZHANG H. The application progress of nonlinear dynamic analysis methods in the field of neurorehabilitation[J]. Chinese Journal of Rehabilitation Medicine, 2019, 34(4): 483-486.
- [16] KANTELHARDT J W, ZSCHIEGNER S A, KOSCIELNY-BUNDE E, et al. Multifractal detrended fluctuation analysis of nonstationary time series[J]. Phys A, 2002, 316(1): 87-114.
- [17] 张石, 余黎煌, 王雅凡, 等. 基于多重分形去趋势波动分析的视网膜图像分割[J]. 东北大学学报(自然科学版), 2019, 40(2): 158-163.
ZHANG S, SHE L H, WANG Y F, et al. Retinal image segmentation based on multifractal detrending fluctuation analysis[J]. Journal of Northeastern University (Natural Science Edition), 2019, 40(2): 158-163.
- [18] WANG J, SHAO W, JUNSEOK K. Combining MF-DFA and LSSVM for retina images classification[J]. Biomed Signal Proces, 2020, 60: 1-8.
- [19] YUAN Z, PENG T T, AN D, et al. Rolling bearing fault diagnosis based on adaptive smooth ITD and MF-DFA method[J]. J Low Freq Noise V A, 2020, 39(4): 1-19.
- [20] GAO S, HUANG J, LI Y R, et al. A forecasting model for wave heights based on a long short-term memory neural network[J]. Acta Oceanol Sin, 2021, 40(1): 62-69.
- [21] LIU P X, ZHU Z S, YE X F, et al. Conditional random field tracking model based on a visual long short term memory network[J]. Journal of Electronic Science and Technology, 2020, 18(4): 308-319.
- [22] 鲁杰, 杨晓栋, 彭靖宇, 等. 基于 LSTM 的运动想象脑电信号分类方法[J]. 电子设计工程, 2021, 29(4): 88-92.
LU J, YANG X D, PENG J Y, et al. Mental imagery EEG signal classification method based on LSTM[J]. Electronic Design Engineering, 2021, 29(4): 88-92.

(编辑: 薛泽玲)