

# 基于可视图算法的数据序列网络特征分析

严 慧<sup>1</sup>

YAN Hui

## 摘 要

复杂网络原理为数据分析开辟了全新的研究途径。可视图算法因其简单、高效、易实现等特点广泛用于为不同数据构建复杂网络。文章以癫痫脑电信号以及学习投入数据为实验对象,采用五种可视图算法为这些数据构建复杂网络,提取网络的八种统计特征,分析不同可视图算法下正常脑电与癫痫脑电网络特征的分布差异以及两类学习投入数据网络特征的分布差异,最后挑选出具有高辨别力的网络特征用于癫痫脑电识别以及学习投入数据分类。研究表明,五种可视图算法下,共有13个网络特征识别癫痫脑电的准确率达到80%以上,最高可达88.0%;共有7个网络特征分类学习投入数据的准确率达到80%以上,最高可达88.1%。实验证明了基于可视图算法为不同数据构建复杂网络进而提取高辨别力网络特征的可行性,为基于网络特征进行信号识别与数据分类提供研究思路与实践参考。

## 关键词

复杂网络;可视图;网络特征;信号识别;数据分类

doi: 10.3969/j.issn.1672-9528.2026.04.023

## 0 引言

复杂网络是一种描述复杂系统中各个元素之间的相互关系的数学模型。将系统中的元素抽象为节点(如:人、

城市、通信设备),将元素之间的关系抽象为边(如:人际关系、交通线路、通信链路),构建出一个网络结构,用以研究整个系统的拓扑结构、动态行为以及元素之间的相互作用。

可视图算法是一种将数据序列转变为复杂网络的方法。这类方法在数据序列与复杂网络之间架起了一座桥梁,为数据分析开辟了新途径。Lacase等<sup>[1]</sup>在2008年首次提出可视图(visibility graph, VG),该算法可以将周期序列转变成规则网络,将随机序列转变成随机网络,将分形序列转变成

1. 金陵科技学院电子信息工程学院 江苏南京 211169

[基金项目] 2023年度高校哲学社会科学研究一般项目“基于复杂网络的应用型本科院校大学生学习投入度研究”(2023SJYB0584);2024年江苏省产学研合作项目“基于复杂网络的病理性脑电信号的特征提取与检测识别”(BY20240093)

[4] 曹来成,后杨宁,冯涛,等.面向动态博弈的k-匿名隐私保护数据共享方案[J].西安电子科技大学学报,2024,51(4):170-179.

[5] 杨邵成,蔡英,范艳芳.车载社交网中基于密文属性加密的车队车辆数据共享方案[J].汽车技术,2024(8):14-21.

[6] 葛胤池,张辉,孙浩航.基于隐空间扩散模型的差分隐私数据合成方法研究[J].计算机科学,2024,51(3):30-38.

[7] 高改梅,张瑾,刘春霞,等.基于区块链与CP-ABE策略隐藏的众包测试任务隐私保护方案[J].计算机应用,2024,44(3):811-818.

[8] 唐明坤,吴思竹,周佳茵,等.医学数据共享隐私保护中基于聚类的匿名化算法关键技术研究[J].医学信息学杂志,2023,44(6):65-71.

[9] 黄仁彦,杨立.数字化档案敏感信息混沌密钥加密算法仿真[J].计算机仿真,2025,42(1):249-252.

## 【作者简介】

史渊源(1982—),男,宁夏固原人,本科,高级工程师,研究方向:大数据分析应用, email:yuany5819@163.com.

万鹏(1988—),男,宁夏石嘴山人,本科,高级工程师,研究方向:数据管理与运营。

王亮(1981—),男,吉林松原人,本科,高级工程师,研究方向:大数据分析应用。

(收稿日期:2025-08-14 修回日期:2026-03-25)

无标度网络,从复杂网络角度描述序列特征。Luque 等<sup>[2]</sup>在 2009 年提出水平可视图 (horizontal visibility graph, HVG),发现水平可视图可用来判别时间序列随机性,因为任何一个随机序列都映射到一个具有  $P(k)=(1/3)(2/3)^{k-2}$  指数阶数分布的图上,而与随机序列的概率分布无关。周婷婷等<sup>[3]</sup>在可视图的基础上提出有限穿越可视图 (limited penetrable visibility graph, LPVG),并证明对于周期信号及混沌信号,有限穿越可视图比可视图具有更强的抗噪性能。Wang 等<sup>[4]</sup>证明了独立同分布随机数列映射为有限穿越水平可视图 (limited penetrable horizontal visibility graph, LPHVG) 时,该图具有指数级分布,与数列产生的概率分布无关。有限穿越水平可视图算法能有效区分混沌与非相关随机性,并能测量实时序列的全局演化特征。Wang 等<sup>[5]</sup>考虑到时间序列经常受到噪声干扰以及大多数节点与远处节点没有直接连接,提出滑动窗口有限穿越可视图 (sliding window limited penetrable visibility graph, SLPVG)。该算法为研究长时间序列的动态特性提供了一个简单而有效的工具。至今,可视图算法仍不断发展,并且应用不同的研究领域,如病理性脑电信号与心电信号检测<sup>[6-7]</sup>、基因探索<sup>[8]</sup>、经济预测<sup>[9]</sup>、气象研究<sup>[10]</sup>、电路检测<sup>[5]</sup>等。

本文以癫痫脑电信号以及学习投入数据为实验对象,分别采用五种可视图算法为数据构建复杂网络,提取网络的八种统计特征,通过 Kolmogorov-Smirnov 检验以及箱线图详细分析正常脑电与癫痫脑电网络特征的分布差异以及两类学习投入数据网络特征的分布差异,获取高辨别力的网络特征用于癫痫脑电识别以及学习投入数据分类。本文通过实验来证明基于复杂网络方法研究分析数据的可行性。

## 1 可视图算法

本节首先介绍最先提出的可视图算法,然后介绍可视图的后续改进算法,水平可视图、有限穿越可视图、有限穿越水平可视图,以及滑动有限穿越可视图。

### 1.1 可视图 (visibility graph, VG)

将一个数据序列置于直角坐标系中,若在一个序列点上能够直线观测到另一个序列点,不被中间序列点遮挡,则称这两个序列点可视,两者相互连接;反之,这两个序列点不可视,两者不相连。

两个序列点的可视准则 (连接准则) 表示为:

$$y_k < y_i + (y_j - y_i) \frac{t_k - t_i}{t_j - t_i}, \quad t_i < t_k < t_j \quad (1)$$

式中:  $(t_i, y_i)$  和  $(t_j, y_j)$  是序列中的任意两点;  $(t_k, y_k)$  是两点之间的任意一点。两个序列点  $(t_i, y_i)$  和  $(t_j, y_j)$  是否相连取决于式 (1) 是否成立。图 1 以 5 点序列为例,展示了序列点之间的连接以及对应的可视图。

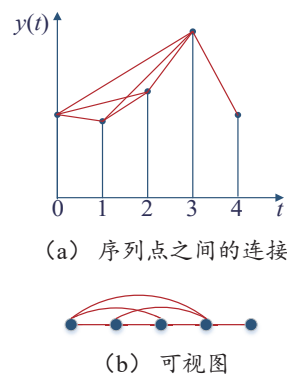


图 1 可视图 VG 示例

### 1.2 水平可视图 (horizontal visibility graph, HVG)

与 VG 算法不同, HVG 算法采用的是水平可视规则。若在一个序列点上能够水平直线观测到另一个序列点,不被中间序列点遮挡,则称这两个序列点水平可视,两者相互连接;反之,这两个序列点水平不可视,两者不相连。

两个序列点的水平可视准则 (连接准则) 表示为:

$$y_k < y_i, y_j, \quad t_i < t_k < t_j \quad (2)$$

式中:  $(t_i, y_i)$  和  $(t_j, y_j)$  是序列中的任意两点;  $(t_k, y_k)$  是这两点之间的任意一点。两个序列点  $(t_i, y_i)$  和  $(t_j, y_j)$  是否相连取决于式 (2) 是否成立。图 2 以 5 点序列为例,展示了序列点之间的连接以及对应的水平可视图。

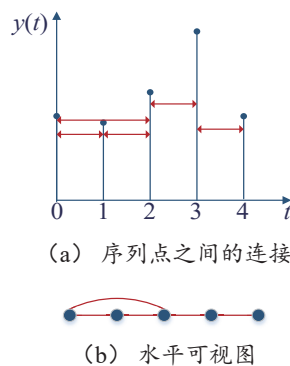


图 2 水平可视图 HVG 示例

### 1.3 有限穿越可视图 (limited penetrable visibility graph, LPVG)

LPVG 算法是在 VG 算法的基础上增设了有限穿越次数  $N$ 。相比于 VG 算法的可视准则, LPVG 算法的可视准则放宽要求,允许两个序列点之间的连线被中间序列点截断的次数不超过  $N$ 。

两个序列点的连接准则表示为:

$$\left\{ (t_k, y_k) \left| \frac{y_k - y_i}{t_k - t_i} - \frac{y_j - y_i}{t_j - t_i} \geq 0 \right. \right\} \leq N, \quad t_i < t_k < t_j \quad (3)$$

式中:  $(t_i, y_i)$  和  $(t_j, y_j)$  是序列中的任意两点;  $(t_k, y_k)$  是两点之间的任意一点;  $N$  是有限穿越次数。两个序列点  $(t_i, y_i)$  和  $(t_j, y_j)$  是否相连取决于式 (3) 是否成立。图 3 以 5 点序列

为例，展示了  $N=1$  时序点之间的连接以及对应的有限穿越可视图。

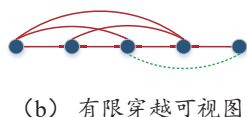
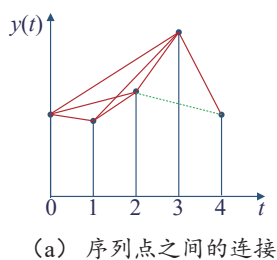


图 3 有限穿越可视图 LPVG 示例 ( $N=1$ )

#### 1.4 有限穿越水平可视图法 (limited penetrable horizontal visibility graph, LPHVG)

LPHVG 算法是在 HVG 算法的基础上增设了有限穿越次数  $N$ 。相比于 HVG 算法的可视准则，LPHVG 算法的可视准则放宽要求，允许两个序列点之间的水平连线被中间序列点截断的次数不超过  $N$ 。

两个序列点的连接准则表示为：

$$\left| \left\{ (t_k, y_k) \mid y_k \geq \min(y_i, y_j) \right\} \right| \leq N, \quad t_i < t_k < t_j \quad (4)$$

式中： $(t_i, y_i)$  和  $(t_j, y_j)$  是序列中的任意两点； $(t_k, y_k)$  是这两点之间的任意一点； $N$  是有限穿越次数。两个序列点  $(t_i, y_i)$  和  $(t_j, y_j)$  是否相连取决于式 (4) 是否成立。图 4 以 5 点序列为例，展示了  $N=1$  时序点之间的连接以及对应的有限穿越水平可视图。

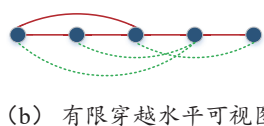
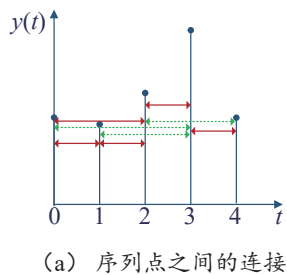


图 4 有限穿越水平可视图 LPHVG 示例 ( $N=1$ )

#### 1.5 滑动有限穿越可视图 (sliding limited penetrable visibility graph, SLPVG)

SLPVG 算法是在 LPVG 算法的基础上增加了一个滑动窗口，窗口的宽度为  $W$ ，仅需考虑窗口宽度内两序列点的连接关系。

两个序列点的连接准则表示为：

$$\left| \left\{ (t_k, y_k) \mid \left| \frac{y_k - y_i}{t_k - t_i} - \frac{y_j - y_i}{t_j - t_i} \right| \geq 0 \right\} \right| \leq N, \quad |t_i - t_j| < W, \quad t_i < t_k < t_j \quad (5)$$

式中： $(t_i, y_i)$  和  $(t_j, y_j)$  是序列中的任意两点； $(t_k, y_k)$  是这两点之间的任意一点； $N$  是有限穿越次数； $W$  是窗口宽度。两个序列点  $(t_i, y_i)$  和  $(t_j, y_j)$  是否相连取决于式 (5) 是否成立。图 5 以 5 点序列为例，展示了  $N=1, W=3$  时序点之间的连接以及对应的滑动窗口有限穿越可视图。

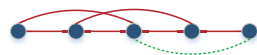
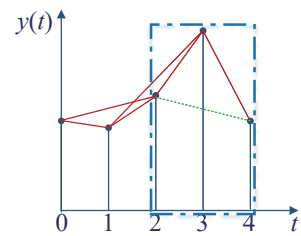
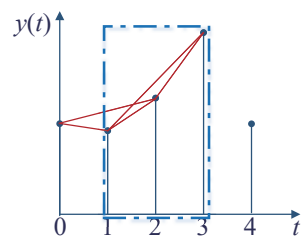
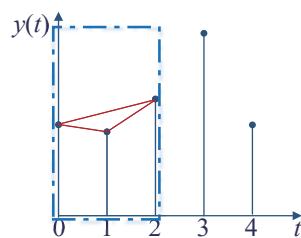


图 5 滑动有限穿越可视图 SLPVG 示例 ( $N=1, W=3$ )

## 2 复杂网络的统计特征

复杂网络可以用邻接矩阵  $A=[a_{ij}]_{N \times N}$  表示。其中， $N$  表示网络节点数； $a_{ij}$  表示节点  $i$  与节点  $j$  之间的连接权重。使用可视图(VG)、水平可视图(HVG)、有限穿越可视图(LPVG)、有限穿越水平可视图(LPHVG)，以及滑动有限穿越可视图(SLPVG)算法所构建的复杂网络都是二进制无向网络。此类网络的特点是：邻接矩阵  $A$  是对称的，即  $A=A^T$  或  $a_{ij}=a_{ji}$ ；

$a_{ij}$  为 1 或 0, 分别代表节点  $i$  和节点  $j$  是否直接相连<sup>[11]</sup>。

为数据序列构建复杂网络之后, 可以提取网络的统计特征, 用于揭示网络的拓扑结构以及分析网络的连接类型。本节介绍复杂网络的八种统计特征, 度、集聚系数、传递性、最短路径长度、局部效率、全局效率、图能量、图拉普拉斯能量。

## 2.1 度

节点  $i$  的度是与该节点直接相连 (相邻) 的节点个数, 其公式为:

$$D_i = \sum_{j=1}^N a_{ij} \quad (6)$$

式中:  $N$  为网络中的节点数,  $a_{ij}$  取自邻接矩阵  $A=[a_{ij}]_{N \times N}$ 。一般来说, 某节点的度越大, 该节点就越重要。复杂网络的度均值越大, 节点之间的联系就越多。

## 2.2 集聚系数

集聚系数衡量的是网络的局部聚合度和信息传输的局部效率。节点  $i$  的集聚系数是与该节点相邻的节点相互连接的可能性, 是节点  $i$  的相邻节点之间实际连接的边数  $e_i$  与最大可能的连边数  $D_i(D_i-1)/2$  的比值, 也是包含节点  $i$  的三角形数量与以节点  $i$  为中心的连通三元组数量的比值, 其公式为:

$$C_i = \frac{2e_i}{D_i(D_i-1)} = \frac{\sum_{j \neq i, j=1}^N \sum_{k \neq i, j, k=1}^N a_{ij} a_{jk} a_{ki}}{\sum_{j \neq i, j=1}^N \sum_{k \neq i, j, k=1}^N a_{ij} a_{ki}} \quad (7)$$

## 2.3 传递性

传递性衡量了整个网络的聚合度和信息传输效率。传递性  $T_r$  是也称为复杂网络的全局集聚系数, 是网络中三角形数量的三倍与连通三元组数量的比值, 其公式为:

$$T_r = \frac{\sum_{i=1}^N \sum_{j \neq i, j=1}^N \sum_{k \neq i, j, k=1}^N a_{ij} a_{jk} a_{ki}}{\sum_{i=1}^N \sum_{j \neq i, j=1}^N \sum_{k \neq i, j, k=1}^N a_{ij} a_{ki}} \quad (8)$$

## 2.4 最短路径长度

一般情况下, 两节点之间具有多条路径, 最短路径是具有最少的连边, 代表信息传输的最优路径。在节点  $i$  和节点  $j$  的  $L$  条路径中, 最短路径长度定义为:

$$l_{ij} = \min \{l_{ij}^1, l_{ij}^2, \dots, l_{ij}^L\} \quad (9)$$

## 2.5 全局效率

如果网络不是完全连通状态, 就会出现断开的节点对, 这些节点之间的最短路径是无限长。因此提出效率的概念。全局效率  $E_G$  的定义公式为:

$$E_G = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i, j=1}^N \frac{1}{l_{ij}} \quad (10)$$

式中:  $1/l_{ij}$  是最短路径长度的倒数。全局效率越高, 网络中的信息传输效率就越高。

## 2.6 局部效率

由于集聚系数只是评估相邻节点之间的直接连接, 而不是多级连接, 因此提出局部效率的概念。节点  $i$  的局部效率定义公式为:

$$E_i = \frac{1}{D_i(D_i-1)} \sum_{j \in N_i} \sum_{k \neq j, k \in N_i} \frac{1}{l_{jk}} \quad (11)$$

式中:  $N_i$  是节点  $i$  的相邻节点形成的子图中的节点集;  $D_i$  是节点  $i$  的度;  $l_{jk}$  是节点  $k$  和  $j$  之间的最短路径长度。局部效率越高, 节点  $i$  与其相邻节点之间的信息传输效率就越高。

## 2.7 图能量

图能量定义为邻接矩阵  $A$  的特征值取模值后再做累加, 其公式为:

$$GE = \sum_{i=1}^N |\lambda_i| \quad (12)$$

式中:  $\lambda_1, \lambda_2, \dots, \lambda_N$  是邻接矩阵  $A=[a_{ij}]_{N \times N}$  的  $N$  个特征值。由于二进制无向网络的邻接矩阵是对称的, 所以特征值都为实数。图能量用来衡量网络的连接强度或拓扑复杂性。网络连接强度或拓扑复杂性越高, 图能量越高; 反之, 稀疏图的图能量较低, 信息传递效率较低。

## 2.8 图拉普拉斯能量

图拉普拉斯能量定义为图拉普拉斯矩阵的特征值偏离平均度取模后再做累加, 其公式为:

$$GLE = \sum_{i=1}^N |\gamma_i - \bar{D}| \quad (13)$$

式中:  $\gamma_1, \gamma_2, \dots, \gamma_N$  是拉普拉斯矩阵  $L=K-A$  的  $N$  个特征值;  $A$  是邻接矩阵;  $K$  是对角线度矩阵 (满足  $K_{ii}=D_i$ );  $\bar{D}$  是平均度。图拉普拉斯能量通过计算网络连通性偏离完美均匀性的程度来衡量结构的不规则性。通常, 图拉普拉斯能量越高, 网络的度分布越不均匀, 网络具有更短的路径长度, 更高的传输效率。

## 3 癫痫脑电信号网络特征分析

癫痫的主要病理特征是脑内神经元群的异常同步放电, 而脑电图作为直接捕捉这种电生理异常的标准技术, 能够实时记录神经元异常活动的时空特征。因此对癫痫脑电信号的研究成为神经科学、临床医学与生物工程等领域的重要课题。

### 3.1 癫痫脑电数据集

本实验数据选择德国波恩大学提供的波恩脑电数据集。该数据集包含 5 个子集。子集  $A$  和  $B$  分别采自健康人受试者睁眼和闭眼脑电; 子集  $C$  和  $D$  采自癫痫患者癫痫发作间期脑电; 子集  $E$  采自癫痫患者癫痫发作期脑电。每个子集包含 100 个数据片段, 每个片段的时间长度为 23.6 s, 数据点为 4 097 个, 采样频率为 173.61 Hz, 分辨率为 12 位。本实验将波恩脑电数据集分为两类, 一类为正常脑电 (子集  $A$  和  $B$ ),

另一类是癫痫脑电（子集  $C$ 、 $D$  和  $E$ ）。

3.2 脑电复杂网络构建与特征提取

由于每个脑电片段长度为 4 097 点，若直接构建复杂网络，网络将具有 4 097 个节点，过于复杂。所以首先将每个脑电片段分割为 8 段，每段 512 点，这样每个子集中含有 800 个脑电片段。最终获得 1 600 个正常脑电片段，2 400 个癫痫脑电片段，每个片段 512 点。使用 VG、HVG、LPVG、LPHVG、SLPVG 算法（LPVG、LPHVG、SLPVG 的有限穿越次数  $N=2$ ，SLPVG 的窗口宽度  $W=$  数据长度  $\times 60\%$ ）为每个脑电片段创建复杂网络，提取复杂网络的八种统计特征。表 1 中列出了八种网络特征的名称以及相应的缩写符号。

表 1 八种网络特征及其缩写

| 特征全称     | 特征缩写  |
|----------|-------|
| 度平均      | $D_M$ |
| 集聚系数均值   | $C_M$ |
| 最短路径长度均值 | $L_M$ |
| 局部效率均值   | $E_M$ |
| 传递性      | $T_r$ |
| 全局效率     | $E_G$ |
| 图能量      | GE    |
| 图拉普拉斯能量  | GLE   |

3.3 癫痫脑电网络特征分析

3.3.1 Kolmogorov-Smirnov 检验

Kolmogorov-Smirnov 检验（KS）是一种非参数统计检验方法，用于比较两类数据是否来自同一分布。一般情况下，若  $P$  值小于 0.05，说明两类数据分布存在差异，并且  $P$  值越小，说明两类数据分布差异越显著。表 2 展示在五种可视图算法下，正常脑电和癫痫脑电的八种网络特征经过 KS 检验后的  $P$  值。

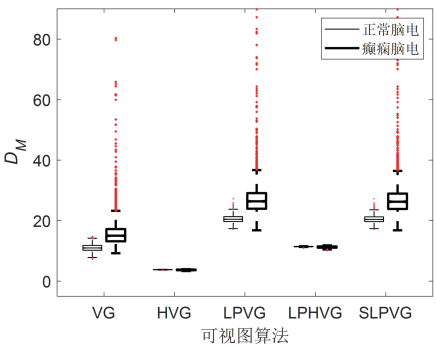
表 2 五种可视图类算法下正常脑电和癫痫脑电的八种网络特征经 KS 检验后的  $P$  值

| 特征    | VG        | HVG       | LPVG      | LPHVG     | SLPVG     |
|-------|-----------|-----------|-----------|-----------|-----------|
| $D_M$ | 0         | 5.89E-148 | 0         | 7.10E-125 | 0         |
| $C_M$ | 0         | 1.33E-144 | 3.91E-280 | 2.50E-209 | 6.08E-281 |
| $L_M$ | 1.51E-80  | 2.83E-246 | 1.38E-47  | 9.94E-222 | 5.24E-47  |
| $E_M$ | 8.47E-303 | 1.44E-140 | 4.07E-278 | 5.96E-204 | 1.61E-278 |
| $T_r$ | 4.18E-100 | 1.13E-132 | 1.69E-151 | 4.85E-127 | 6.06E-171 |
| $E_G$ | 8.51E-184 | 2.37E-258 | 2.77E-141 | 3.58E-248 | 6.44E-162 |
| GE    | 0         | 2.64E-161 | 0         | 1.07E-33  | 0         |
| GLE   | 0         | 4.11E-09  | 1.10E-218 | 4.22E-196 | 3.44E-221 |

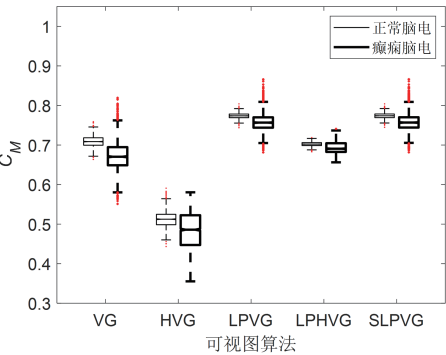
由表 2 可见，所有  $P$  值都远小于 0.05，有的特征  $P$  值甚至为 0（并非真正为 0，而是小于最小常量 2.225 1E-308），说明正常脑电和癫痫脑电的网络特征分布差异极为显著。表 2 中小于 1.0E-200 的  $P$  值已用粗体显示。

3.3.2 箱线图

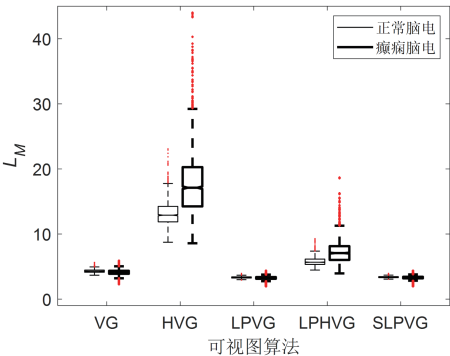
箱线图可以直观地展现两类数据的分布差异。图 6 的子图（a）~（h）分别展示了正常脑电与癫痫脑电的八种网络特征在五种可视图类算法下的箱线图。细箱线图代表正常脑电的网络特征，粗箱线图代表癫痫脑电的网络特征。矩形箱盒的中间横线代表数据的中位数，矩形箱盒的上横线代表数据的上四分位，矩形箱盒的下横线代表数据的下四分位，虚线的上端代表数据的上限，虚线的下端代表数据的下限，离散点代表离群值。可见，箱线图可以直观地展示正常脑电与癫痫脑电网络特征的分布差异。



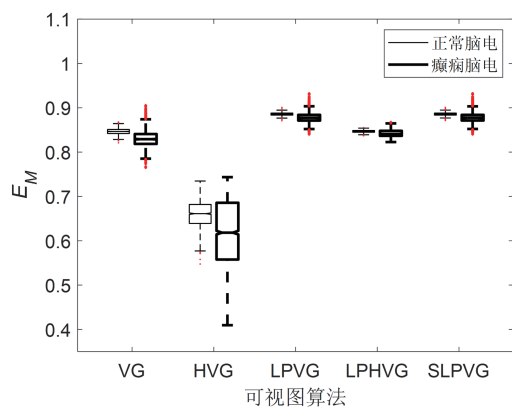
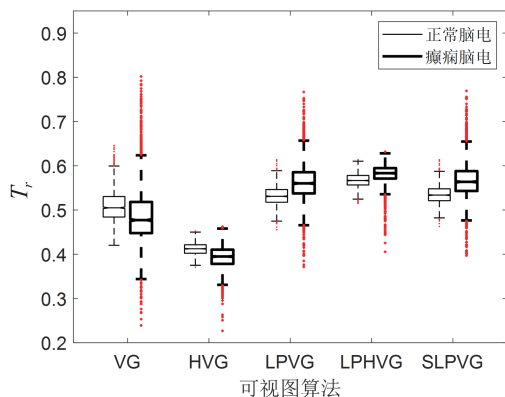
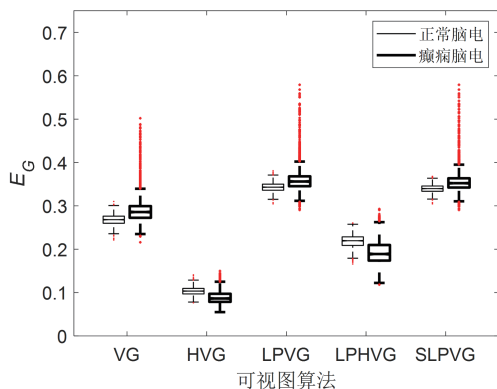
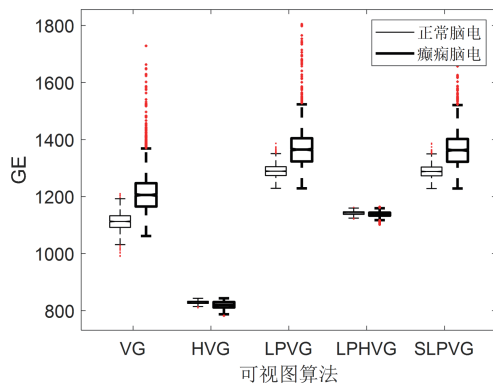
（a）度均值（ $D_M$ ）的箱线图



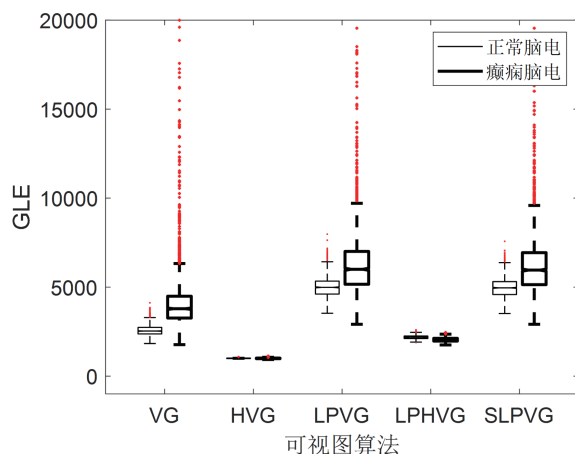
（b）集聚系数均值（ $C_M$ ）的箱线图



（c）最短路径长度均值（ $L_M$ ）的箱线图

(d) 局部效率均值 ( $E_M$ ) 的箱线图(e) 传递性 ( $T_r$ ) 的箱线图(f) 全局效率 ( $E_G$ ) 的箱线图

(g) 图能量 (GE) 的箱线图



(h) 图拉普拉斯能量 (GLE) 的箱线图

图6 五种可视图类算法下正常脑电与癫痫脑电网络特征的箱线图

### 3.3.3 网络特征分析

图6的每个子图展现的是每个特征的箱线图,分别对应表2中的每一行。通常情况下,表2中的 $P$ 值越小,正常脑电与癫痫脑电网络特征的箱线图就分离得越远,网络特征分布差异越显著。下文将结合表2中的 $P$ 值以及图6中的箱线图来分析正常脑电与癫痫脑电网络特征的分布差异。

对于 $D_M$ 特征, VG、LPVG、SLPVG算法下的 $P$ 值为0。从图6(a)可见, VG、LPVG、SLPVG算法下,粗箱盒的下四分位远高于细箱盒的上四分位。所以这三种算法下,正常脑电与癫痫脑电的 $D_M$ 特征的分布差异更显著。

对于 $C_M$ 特征, VG、LPVG、SLPVG算法下的 $P$ 值都小于 $4.0E-280$ 。从图6(b)可见, VG算法下,细箱盒的下四分位高于粗箱盒的上四分位; LPVG、SLPVG算法下,细箱盒的下四分位与粗箱盒的上四分位基本平齐。所以这三种算法下,正常脑电与癫痫脑电的 $C_M$ 特征的分布差异更显著。

对于 $L_M$ 特征, HVG、LPHVG算法下的 $P$ 值都小于 $1.0E-220$ 。从图6(c)可见, HVG、LPHVG算法下,粗箱盒的下四分位与细箱盒的上四分位基本平齐。所以这两种算法下,正常脑电与癫痫脑电的 $L_M$ 特征的分布差异更显著。

对于 $E_M$ 特征, VG、LPVG、SLPVG算法下的 $P$ 值都小于 $1.0E-275$ 。从图6(d)可见, VG、LPVG、SLPVG算法下,细箱盒的下四分位与粗箱盒的上四分位基本平齐。所以这三种算法下,正常脑电与癫痫脑电的 $E_M$ 特征的分布差异更显著。

对于 $T_r$ 特征, SLPVG算法下的 $P$ 值为 $6.06E-171$ ,小于其他四种算法下的 $P$ 值。从图6(e)可见, SLPVG算法

下，粗箱盒的下四分位更接近于细箱盒的上四分位。所以SLPVG 算法下，正常脑电与癫痫脑电的  $T_r$  特征的分布差异更显著。

对于  $E_G$  特征，HVG、LPHVG 算法下的  $P$  值都小于  $1.0E-245$ 。从图 6（f）可见，HVG、LPHVG 算法下，细箱盒的下四分位与粗箱盒的上四分位基本平齐，而其他三种算法下，粗箱盒的下四分位低于细箱盒的上四分位。所以 HVG、LPHVG 算法下，正常脑电与癫痫脑电的  $E_G$  特征的分布差异更显著。

对于 GE 特征，VG、LPVG、SLPVG 算法下的  $P$  值为 0。从图 6（g）可见，VG、LPVG、SLPVG 算法下，粗箱盒的下四分位远高于细箱盒的上四分位。所以这三种算法下，正常脑电与癫痫脑电的 GE 特征的分布差异更显著。

对于 GLE 特征，VG、LPVG、SLPVG 算法下的  $P$  值都小于  $1.0E-215$ 。从图 6（h）可见，VG 算法下，粗箱盒的下四分位远高于细箱盒的上四分位；LPVG、SLPVG 算法下，粗箱盒的下四分位高于细箱盒的中位线。所以这三种算法下，正常脑电与癫痫脑电的 GLE 特征的分布差异更显著。

表 2 中的  $P$  值越小，箱线图中的细箱盒与粗箱盒分离得越远，相应特征分类正常脑电与癫痫脑电的能力越强。

表 3 展现的是选用  $P$  值小于  $1E-200$  的网络特征分类正常脑电与癫痫脑电的准确率，使用了四种机器学习模型，决策树（decision tree,DT），K 近邻（K-nearest neighbor,KNN），支持向量机（support vector machine,SVM），以及贝叶斯分类器（Naïve Bayes, NB），并采用十折交叉验证获得最终的准确率。由表 3 可见，准确率能达到 80% 以上的特征有：VG 算法下的  $D_M$ 、 $C_M$ 、 $E_M$ 、GE、GLE；LPVG 算法下的  $D_M$ 、 $C_M$ 、 $E_M$ 、GE；SLPVG 算法下的  $D_M$ 、 $C_M$ 、 $E_M$ 、GE。

4 学习投入数据网络特征分析

大学生的学习投入是用来衡量学生为达到学习目标在学习过程中持续努力的程度。研究表明，增加学习投入不仅有助于提高学业成绩，更能培养良好的思维习惯和受益终身的学习能力。

4.1 获取数据

本实验使用 Schaufeli 等 7 开发的学习投入量表。该量表

表 3  $P$  值小于  $1E-200$  的网络特征分类正常脑电与癫痫脑电的准确率

| 算法特征  |       | DT          | KNN         | SVM         | NB          |
|-------|-------|-------------|-------------|-------------|-------------|
| VG    | $D_M$ | 0.861±0.017 | 0.855±0.016 | 0.857±0.021 | 0.857±0.016 |
|       | $C_M$ | 0.807±0.025 | 0.813±0.016 | 0.812±0.015 | 0.813±0.016 |
|       | $E_M$ | 0.793±0.018 | 0.800±0.017 | 0.808±0.016 | 0.804±0.015 |
|       | GE    | 0.858±0.015 | 0.854±0.020 | 0.858±0.022 | 0.858±0.020 |
|       | GLE   | 0.873±0.015 | 0.872±0.021 | 0.873±0.010 | 0.875±0.011 |
| HVG   | $L_M$ | 0.756±0.016 | 0.754±0.020 | 0.752±0.023 | 0.758±0.019 |
|       | $E_G$ | 0.751±0.018 | 0.754±0.019 | 0.754±0.023 | 0.756±0.029 |
| LPVG  | $D_M$ | 0.878±0.014 | 0.878±0.012 | 0.873±0.014 | 0.877±0.015 |
|       | $C_M$ | 0.811±0.010 | 0.814±0.021 | 0.817±0.019 | 0.816±0.014 |
|       | $E_M$ | 0.811±0.020 | 0.815±0.017 | 0.815±0.014 | 0.816±0.018 |
|       | GE    | 0.815±0.026 | 0.815±0.021 | 0.817±0.013 | 0.816±0.020 |
|       | GLE   | 0.749±0.021 | 0.748±0.038 | 0.738±0.025 | 0.752±0.020 |
| LPHVG | $C_M$ | 0.787±0.021 | 0.791±0.015 | 0.786±0.017 | 0.790±0.019 |
|       | $L_M$ | 0.731±0.013 | 0.731±0.016 | 0.735±0.018 | 0.741±0.019 |
|       | $E_M$ | 0.789±0.028 | 0.787±0.017 | 0.787±0.021 | 0.792±0.023 |
|       | $E_G$ | 0.757±0.024 | 0.757±0.011 | 0.759±0.023 | 0.759±0.021 |
| SLPVG | $D_M$ | 0.879±0.016 | 0.879±0.017 | 0.876±0.020 | 0.880±0.018 |
|       | $C_M$ | 0.808±0.015 | 0.812±0.018 | 0.817±0.021 | 0.817±0.021 |
|       | $E_M$ | 0.806±0.020 | 0.816±0.026 | 0.816±0.012 | 0.817±0.019 |
|       | GE    | 0.816±0.015 | 0.816±0.021 | 0.816±0.012 | 0.817±0.019 |
|       | GLE   | 0.750±0.021 | 0.750±0.016 | 0.739±0.026 | 0.753±0.027 |

包含动机、精力、专注三个维度,设有17个问题,采用7分制,7代表总是/每天,1代表从不。面向金陵科技学院电子信息工程学院的学生,使用该量表收集学习投入数据,最终获得696条有效数据。这些数据根据学生一学期的综合学业成绩分为两类,一类学业成绩大于等于80分,样本数量为332个,另一类学业成绩小于80分,样本数据为364个。

#### 4.2 学习投入数据复杂网络创建与特征提取

首先对原始数据进行预处理。每条样本包含17个数据点,数值位于1~7之间。计算每条样本在七个数值下的数量,再与样本的最大值、最小值、中位数、众数、均值构成12个数据点。为突出数据变化趋势,重复12个数据点构成24个数据点。使用VG、HVG、LPVG、LPHVG、SLPVG算法(LPVG、LPHVG、SLPVG的有限穿越次数 $N=1$ ,SLPVG的窗口宽度 $W=$ 数据长度 $\times 60\%$ )为每条数据创建复杂网络,提取复杂网络的八种统计特征,如表1所示。

#### 4.3 学习投入数据网络特征分析

##### 4.3.1 Kolmogorov-Smirnov 检验

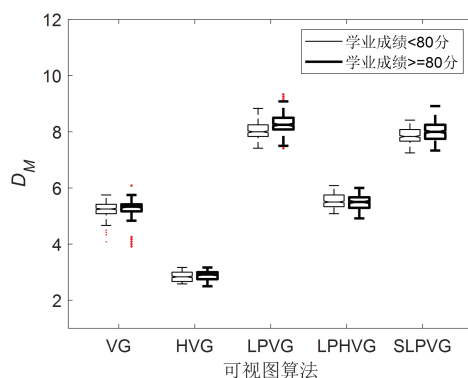
表4展示的是在五种可视图算法下,学业成绩大于等于80分与小于80分的两类学习投入数据的八种网络特征经过KS检验后的 $P$ 值。可见,所有 $P$ 值都小于0.05,有的特征 $P$ 值甚至小于 $1E-40$ ,说明两类学习投入数据网络特征的分

表4 五种可视图类算法下两类学习投入数据的八种网络特征经KS检验后的 $P$ 值

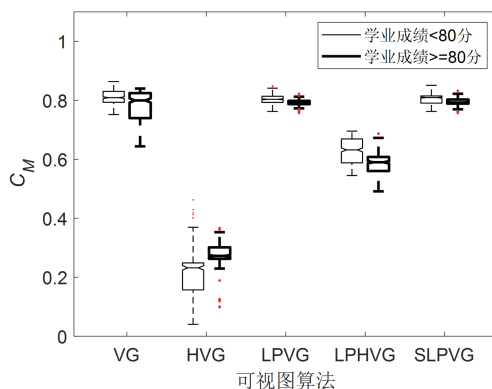
| 特征    | VG       | HVG             | LPVG     | LPHVG           | SLPVG           |
|-------|----------|-----------------|----------|-----------------|-----------------|
| $D_M$ | 7.63E-08 | 2.31E-04        | 4.50E-18 | 1.97E-07        | 9.35E-10        |
| $C_M$ | 2.87E-20 | <b>4.43E-48</b> | 8.33E-38 | <b>4.55E-48</b> | 6.14E-32        |
| $L_M$ | 6.06E-09 | 3.41E-19        | 9.90E-15 | 4.80E-13        | 2.41E-04        |
| $E_M$ | 5.75E-22 | <b>1.45E-47</b> | 1.25E-35 | <b>2.10E-48</b> | 3.98E-30        |
| $T_r$ | 3.40E-06 | <b>1.97E-48</b> | 6.62E-12 | 1.28E-34        | 1.88E-13        |
| $E_G$ | 1.79E-09 | 2.75E-18        | 5.76E-20 | 6.58E-15        | 4.49E-08        |
| GE    | 2.54E-32 | <b>1.43E-66</b> | 0.028    | 1.21E-13        | 1.12E-06        |
| GLE   | 5.02E-32 | 1.33E-22        | 1.28E-24 | 2.84E-16        | <b>2.19E-57</b> |

##### 4.3.2 箱线图

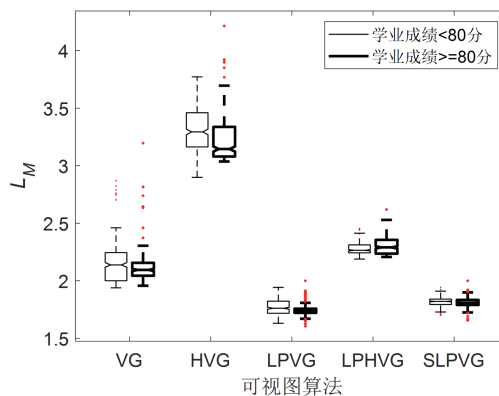
图7(a)~(g)分别展示了两类学习投入数据的八种网络特征在五种可视图类算法下的箱线图。细箱线图代表学业成绩小于80分的学习投入数据网络特征,粗箱线图代表学业成绩大于等于80分的学习投入数据网络特征。



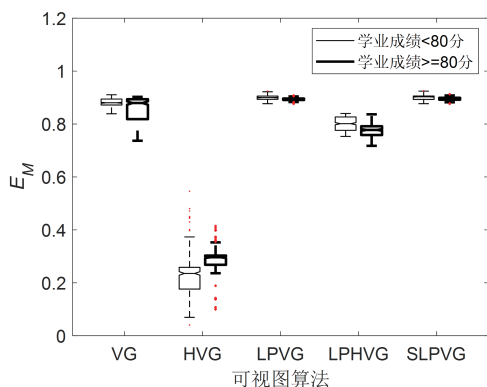
(a) 度均值 ( $D_M$ ) 的箱线图



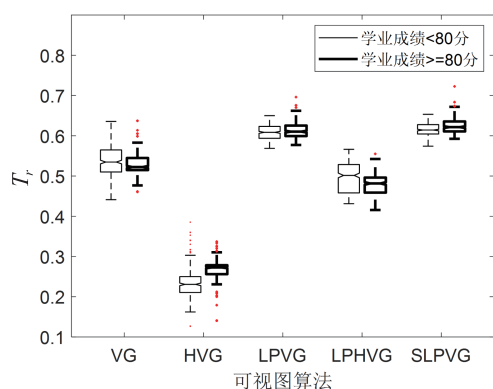
(b) 集聚系数均值 ( $C_M$ ) 的箱线图



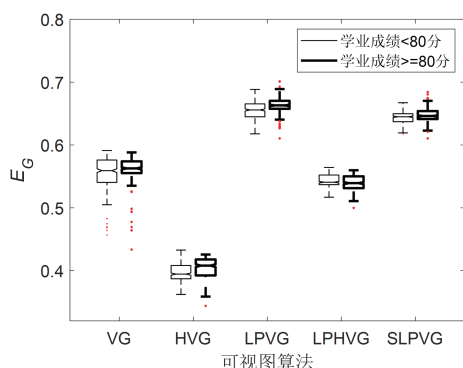
(c) 最短路径长度均值 ( $L_M$ ) 的箱线图



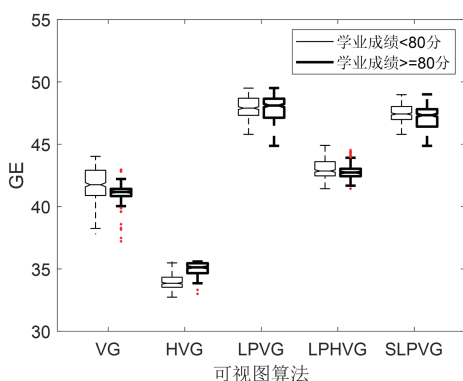
(d) 局部效率均值 ( $E_M$ ) 的箱线图



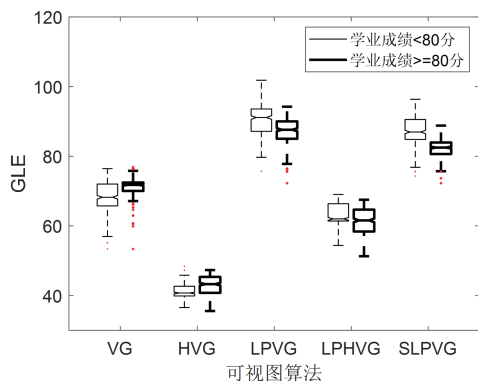
(e) 传递性 ( $T_r$ ) 的箱线图



(f) 全局效率 ( $E_G$ ) 的箱线图



(g) 图能量 (GE) 的箱线图



(h) 图拉普拉斯能量 (GLE) 的箱线图

图7 五种可视图类算法下两类学习投入数据网络特征的箱线图

### 4.3.3 网络特征分析

通常情况下, 表4中的 $P$ 值越小, 两类学习投入数据网络特征的箱线图分离得越远, 特征分布差异越显著。下文将结合表4中的 $P$ 值以及图7中的箱线图来分析两类学习投入数据网络特征分布差异。

对于 $D_M$ 特征, LPVG算法下的 $P$ 值最小。从图7(a)可见, LPVG算法下, 粗箱盒的下四分位高于细箱盒的中位线。所以LPVG算法下, 两类学习投入数据的 $D_M$ 特征的分布差异更显著。

对于 $C_M$ 特征, HVG、LPHVG算法下的 $P$ 值小于 $4.0E-45$ 。从图7(b)可见, HVG算法下, 粗箱盒的下四分位高于细箱盒的上四分位; LPHVG算法下, 粗箱盒的上四分位低于细箱盒的中位线。所以HVG和LPHVG算法下, 两类学习投入数据的 $C_M$ 特征的分布差异更显著。

对于 $L_M$ 特征, HVG算法下的 $P$ 值最小。从图7(c)可见, HVG算法下, 细箱盒的下四分位高于粗箱盒的中位线。所以HVG算法下, 两类学习投入数据的 $L_M$ 特征的分布差异更显著。

对于 $E_M$ 特征, HVG、LPHVG算法下的 $P$ 值小于 $4.0E-45$ 。从图7(d)可见, HVG算法下, 粗箱盒的下四分位高于细箱盒的上四分位; LPHVG算法下, 粗箱盒的上四分位低于细箱盒的中位线。所以HVG和LPHVG算法下, 两类学习投入数据的 $E_M$ 特征的分布差异更显著。

对于 $T_r$ 特征, HVG算法下的 $P$ 值最小, 为 $1.97E-48$ 。从图7(e)可见, HVG算法下, 粗箱盒的下四分位高于细箱盒的上四分位。所以HVG算法下, 两类学习投入数据的 $T_r$ 特征的分布差异更显著。

对于 $E_G$ 特征, LPVG算法下的 $P$ 值最小。从图7(f)可见, LPVG算法下, 粗箱盒的下四分位高于细箱盒的中位线。所以LPVG算法下, 两类学习投入数据的 $E_G$ 特征的分布差异更显著。

对于GE特征, HVG算法下的 $P$ 值最小, 为 $1.43E-66$ 。从图7(g)可见, HVG算法下, 粗箱盒的下四分位高于细箱盒的上四分位。所以HVG算法下, 两类学习投入数据的GE特征的分布差异更显著。

对于GLE特征, SLPVG算法下的 $P$ 值最小, 为 $2.19E-57$ 。从图7(h)可见, SLPVG算法下, 细箱盒的下四分位高于粗箱盒的上四分位。所以SLPVG算法下, 两类学习投入数据的GLE特征的分布差异更显著。

表4中的 $P$ 值越小, 箱线图粗箱盒与细箱盒分离得越远, 相应特征分类两类学习投入数据能力越强。表5展现的是选用 $P$ 值小于 $1E-40$ 的网络特征分类两类学习投入数据, 同样使用了四种机器学习模型, 决策树(DT), K近邻(KNN), 支持向量机(SVM), 以及贝叶斯分类器(NB), 并采用十折交叉验证获得最终的准确率。由表5

表 5  $P$  值小于  $1E-40$  的网络特征分类两类学习投入数据的准确率

| 算法特征  |       | DT          | KNN         | SVM         | NB          |
|-------|-------|-------------|-------------|-------------|-------------|
| HVG   | $C_M$ | 0.875±0.037 | 0.838±0.047 | 0.858±0.036 | 0.852±0.027 |
|       | $E_M$ | 0.881±0.035 | 0.840±0.049 | 0.866±0.030 | 0.849±0.036 |
|       | $T_r$ | 0.849±0.047 | 0.822±0.038 | 0.830±0.027 | 0.859±0.029 |
|       | GE    | 0.845±0.016 | 0.831±0.047 | 0.816±0.036 | 0.819±0.025 |
| LPHVG | $C_M$ | 0.846±0.048 | 0.862±0.045 | 0.793±0.044 | 0.815±0.065 |
|       | $E_M$ | 0.828±0.046 | 0.849±0.039 | 0.751±0.052 | 0.759±0.052 |
| SLPVG | GLE   | 0.831±0.046 | 0.800±0.042 | 0.796±0.057 | 0.799±0.040 |

可见, 准确率能达到 80% 以上的特征有: HVG 算法下的  $C_M$ 、 $E_M$ 、 $T_r$ 、GE; LPHVG 算法下的  $C_M$ 、 $E_M$ ; SLPVG 算法下的 GLE。

5 结论

本文以癫痫脑电以及学习投入数据为实验对象, 分别使用五种可视图算法为这些数据构建复杂网络, 提取网络的八种统计特征, 通过 KS 检验以及箱线图详细分析了不同可视图算法下正常脑电与癫痫脑电网络特征的分布差异以及两类学习投入数据网络特征的分布差异, 并挑选分布差异显著的网络特征完成癫痫脑电识别以及学习投入数据分类。实验结果表明, 在癫痫脑电识别中, 准确率达到 80% 以上的特征有: VG 算法下的  $D_M$ 、 $C_M$ 、 $E_M$ 、GE、GLE 特征, LPVG 算法下的  $D_M$ 、 $C_M$ 、 $E_M$ 、GE 特征, 以及 SLPVG 算法下的  $D_M$ 、 $C_M$ 、 $E_M$ 、GE 特征。在学习投入数据分类中, 准确率达到 80% 以上的特征有: HVG 算法下的  $C_M$ 、 $E_M$ 、 $T_r$ 、GE 特征, LPHVG 算法下的  $C_M$ 、 $E_M$  特征, 以及 SLPVG 算法下的 GLE 特征。本文通过实验证明了基于可视图算法为数据构建复杂网络进而提取高辨别力网络特征的可行性, 后续研究可以拓展至有向可视图、加权可视图, 以及其他典型的网络特征。

参考文献:

[1] LACASA L, LUQUE B, BALLESTEROS F, et al. From time series to complex networks: the visibility graph[J]. Proceedings of the national academy of sciences, 2008, 105(13): 4972-4975.

[2] LUQUE B, LACASA L, BALLESTEROS F, et al. Horizontal visibility graphs: exact results for random time series[J]. Physical review E, 2009, 80(4): 046103.

[3] 周婷婷, 金宁德, 高忠科, 等. 基于有限穿越可视图的时间序列网络模型 [J]. 物理学报, 2012, 61 (3): 86-96.

[4] WANG M G,VILELA A L M, DU R J,et al. Exact results of the limited penetrable horizontal visibility graph associated to random time series and its application[J]. Scientific reports,

2018[2026-03-19].<https://www.nature.com/articles/s41598-018-23388-1>.

[5] WANG S L, LI P, CHEN G W,et al. Sliding limited penetrable visibility graph for establishing complex network from time series[J]. Chaos: an interdisciplinary journal of nonlinear science, 2024, 34: 043145.

[6] KUTLUANA G, TURKER I. Classification of cardiac disorders using weighted visibility graph features from ECG signals[J]. Biomedical signal processing and control, 2024, 87: 105420.

[7] ZHU G H, LI Y, WEN P. Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm[J]. Computer methods and programs in biomedicine, 2014, 115(2): 64-75.

[8] KOSMIDIS K, HUETT M T. DNA visibility graphs[J]. Physica a-statistical mechanics and its applications, 2023, 626: 129043.

[9] HU Y T, XIAO F Y. A novel method for forecasting time series based on directed visibility graph and improved random walk[J]. Physica a-statistical mechanics and its applications, 2022, 594: 127029.

[10] LIU J L, YU Z G, ZHOU Y. A cross horizontal visibility graph algorithm to explore associations between two time series[J]. Chaos solitons & fractals, 2024, 181: 114674.

[11] YAN H, CHEN Z Y, ZHU F, et al. A study of complex network features for electrocardiograms and its applications in atrial fibrillation recognition[J]. Biomedical signal processing and control, 2025, 102: 107295.

【作者简介】

严慧 (1982—), 女, 江苏南京人, 硕士研究生, 讲师、教师, 研究方向: 信号处理以及复杂网络。

(收稿日期: 2025-09-07 修回日期: 2026-04-03)