

基于多路可视图的健康与心梗患者 ECG 信号复杂网络识别*

马志怡¹⁾ 杨小冬^{1)†} 何爱军²⁾ 马璐¹⁾ 王俊³⁾

1) (中国矿业大学计算机科学与技术学院, 徐州 221116)

2) (南京大学电子科学与工程学院, 南京 210023)

3) (南京邮电大学地理与生物信息学院, 南京 210023)

摘要

可视图 (Visibility Graph, VG) 算法已被证明是将时间序列转换为复杂网络的简单且高效的方法, 其构成的复杂网络在拓扑结构中继承了原始时间序列的动力学特性。目前, 单维时间序列的可视图分析已趋于成熟, 但应用于复杂系统时, 单变量往往无法描述系统的全局特征。本文提出一种新的多元时间序列分析方法, 将心梗和健康人的 12 导联心电图 (Electrocardiograph, ECG) 信号转换为多路可视图, 以每个导联为一个节点, 两个导联构成可视图的层间互信息为连边权重, 将其映射到复杂网络。由于不同人群的全连通网络表现为完全相同的拓扑结构, 无法唯一表征不同个体的动力学特征, 根据层间互信息大小重构网络, 提取权重度和加权聚类系数, 实现对不同人群 12 导联 ECG 信号的识别。为判断序列长度对识别效果的影响, 引入多尺度权重度分布熵。由于健康受试者拥有更高的平均权重度和平均加权聚类系数, 其映射网络表现为更加规则的结构、更高的复杂性和连接性, 可以与心梗患者进行区分, 两个参数的识别准确率均达到 93.3%。

关键词: 心肌梗死, 多元时间序列, 多路可视图, 复杂网络

PACS: 05.10.-a, 87.80.-y

基金: 国家自然科学基金面上项目 (批准号: 61772532, 61876186 和 11774167) 资助。

† 通讯作者. E-mail: yxdnju@163.com

第一作者. E-mail: 1946714512@qq.com

1 引言

随着大数据的快速发展, 每时每刻有数以百万计的信息涌现, 多样和复杂的数据给各领域的研究者们带来巨大挑战。其中, 时间序列数据引起学者们的广泛关注。2006 年, Zhang 等^[1]将伪周期时间序列映射到复杂网络, 发现复杂网络作

为有效载体可以用于非线性时间序列的动力学分析。此后，复杂网络成为时间序列分析的有效工具，越来越受到学者们的关注^[2]。2008年，Lacasa等^[3]提出可视图算法，将时间序列的每一个数据点映射为网络中的节点，节点间是否连接取决于数据点之间的可视性。2009年，Luque等^[4]在其基础上提出结构更为简化的水平可视图算法，降低其时间复杂度。同年，Marwan等^[5-7]将相空间向量看作节点，通过向量之间的距离定义节点间的连接^[8]，提出递归网络。2012年，周婷婷等^[9]提出抗噪性能更好的有限穿越可视图。2013年，Sun等^[10]提出转换网络，主要思想是将时间序列转换为符号表示并将每个符号定义为一个节点，用不同符号之间的转移概率描述节点间的连接性^[11,12]。

可视图算法作为应用最为广泛的时间序列转换策略之一，被成功应用于地理学^[13,14]、医学^[15-20]、金融学^[21-23]等各个领域。然而，随着多传感技术的更新，单层网络对多通道时间序列的融合和复杂系统的描述存在一定的局限性，因此，多元时间序列分析逐渐受到关注^[24,25]。学者们对此进行大量研究，将注意力转向多层网络，提出从多元时间序列映射到多层复杂网络的诸多方法^[26-32]。2015年，Lacasa等^[33]提出多路可视图，开创了基于可视图的多元时间序列分析先河。在此基础上，Rafael等^[34]利用联合分布图分析环境变量间的相关性，证明层间互信息和平均边缘重叠可以描述大气污染物之间的相互作用。Gao等^[35]利用多路有限穿越可视图对基于SSMVEP的脑电信号进行分析，成功识别了个体的正常和疲劳行为。Cai等^[36]提出基于模体的多路可视图卷积神经网络，引入多模体熵作为区分不同睡眠阶段脑电图（Electroencephalogram, EEG）信号的定量指标。Samanta等^[37]提出基于不同电极间相互关联的聚类系数构建脑连接网络的新方法，采用基于自动编码器的深度特征提取技术从脑连接矩阵图像中提取有效特征，实现对不同运动脑电信号的分类。

心梗（Myocardial Infarction, MI）是全球范围内致死率最高的疾病之一^[38]。目前，心梗的临床诊断主要依赖医生对心电信号的经验分析，缺乏更加行之有效的手段。因此，学者们提出基于12导联分析的心梗自动检测算法。最早的方法是提取心电图中的时频域特征，揭示其演化与心肌梗死之间的联系。Dohare等^[39]提取12导联心电图的P波、QRS波和ST-T段的幅值、面积、均值、偏度和峰度等特征对心梗信号进行检测。Shanna等^[40]对心电信号进行6层连续小波分解并将每层小波系数的Renyi熵、分形维数和模糊熵作为特征进行心梗识别。Tripathy等^[41]提出基于傅里叶-贝塞尔级数展开的经验小波变换对多导联心电图进行多尺度分析。随着计算机技术的发展，神经网络成为提取心电图特征的有效工具。Acharya等^[42]提出基于11层CNN的模型，利用II导联的ECG信号准确检测心梗。熊等^[43]和Liu等^[44]将原始心电信号串联接入数据作为输入，利用神经网络方法进行特征提取从而实现了对心肌梗死的检测。然而，上述检测的方法或仅考虑单维时

间序列，忽略了 12 导联的多样性，或将 12 导联以串联方式送入网络，忽略了多导联之间的内在联系。针对以上不足，本文提出多路可视图方法，将 12 导联中的每一个导联作为多路可视图的一路，映射为一个网络节点，节点之间的层间互信息作为连边权重，并引入阈值构造无向加权网络，提取权重度和加权聚类系数两个参数，实现对心梗疾病的检测。该方法以受试者 12 导联作为研究对象，利用多路可视图对其进行联合分析，以研究心脏电系统的非线性动力学特征，对心梗患者的检测和诊断具有重要意义。

2 多路可视图

2.1 构造方法

首先回顾一维可视图的构造^[3]。可视图算法的主要思想是将时间序列 $\{x_i\}_{i=1}^N$ 中的每一个数据点映射到复杂网络的一个节点，节点间有无连边取决于两节点是否满足可视性准则：对于任意两点 t_a 、 t_c ，若它们之间的连线没有被任何位于这两点之间的其他点所截断，则这两个节点“可视”，在复杂网络中可以连接成边，相邻两节点间一定“可视”。可视图算法的公式描述为

$$x_c < x_b + (x_a - x_b) \frac{t_b - t_c}{t_b - t_a}, a < b < c \quad (1)$$

其中， t_b 是位于 t_a 、 t_c 之间的任意节点。

多路可视图的构造建立在一维可视图基础之上，若将多元时间序列看作是多维的，则每个单变量时间序列为其中的一维。我们考虑一个 M 维实值序列 $\{x(t)\}_{t=1}^N$ ， $x(t) = (x^{[1]}(t), x^{[2]}(t), \dots, x^{[M]}(t))$ ，其中，上标对应序列所在维数， N 代表序列长度，每一维时间序列的长度相同。对每个单变量时间序列 $\{x^{[\alpha]}(t)\}_{t=1}^N$ ， $\alpha = 1, 2, \dots, M$ ，根据可视图算法构造可视图网络，将 M 维时间序列全部映射完成后构造的具有 M 层复杂网络的结构，即为多路可视图。图 1 以三维时间序列为例，描述了一个三路可视图的构造过程^[33]。多路可视图用其邻接矩阵向量 $A = \{A^{[1]}, \dots, A^{[\alpha]}, \dots, A^{[M]}\}$ 表示，其中，对于第 α 层的邻接矩阵 $A^{[\alpha]} = \{a_{ij}^{[\alpha]}\}$ ，当且仅当节点 i 和节点 j 在第 α 层连接时， $a_{ij}^{[\alpha]} = 1$ 。每一个节点 i 对应一个度向量 k_i ， $k_i = \{k_i^{[1]}, \dots, k_i^{[\alpha]}, \dots, k_i^{[M]}\}$ ， $k_i^{[\alpha]} = \sum_j a_{ij}^{[\alpha]}$ 。

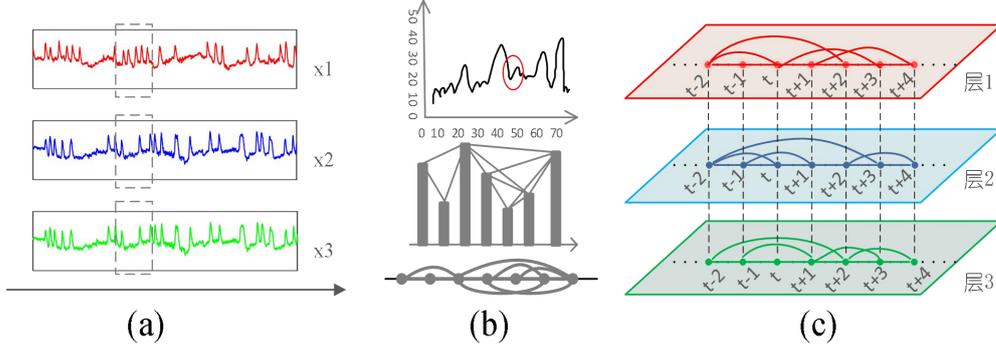


图 1 VG 算法构造多路可视图示意图 (a) 一个多维时间序列; (b) VG 算法示意图。从上至下依次为原始时间序列、红色圆圈包含片段的可视图算法过程及其构造后的网络; (c) 图(a)虚线框所含序列通过 VG 算法构造的多路可视图, 其中层 1、2、3 分别对应时间序列 x_1 、 x_2 、 x_3

Fig.1. Schematic diagram of constructing multiplex visibility graph using VG algorithm. (a) a multivariate time series; (b) schematic diagram of VG algorithm. From top to bottom are original time series, VG algorithm procedure for fragment contained in the red circle and the constructed network; (c) multiplex visibility graph of series in dotted box in figure (a) constructed by VG algorithm, where layers 1, 2 and 3 correspond to time series x_1 , x_2 and x_3 respectively.

2.2 多路可视图映射到复杂网络

2.2.1 层间互信息

我们用层间互信息 $I_{\alpha,\beta}$ ^[33] 量化多路可视图不同层之间度分布的相关性, I 表示为

$$I_{\alpha,\beta} = \sum_{k^{[\alpha]}} \sum_{k^{[\beta]}} P(k^{[\alpha]}, k^{[\beta]}) \log \frac{P(k^{[\alpha]}, k^{[\beta]})}{P(k^{[\alpha]})P(k^{[\beta]})} \quad (2)$$

其中, α , β 是多路可视图中的任意两层, 对应时间序列的度分布分别为 $P(k^{[\alpha]})$ 、 $P(k^{[\beta]})$, $P(k^{[\alpha]}, k^{[\beta]})$ 是在第 α 层寻找度为 $k^{[\alpha]}$ 的点和在第 β 层寻找度为 $k^{[\beta]}$ 的点的联合概率, 计算方法为

$$P(k^{[\alpha]}, k^{[\beta]}) = \frac{N_{k^{[\alpha]}, k^{[\beta]}}}{N} \quad (3)$$

其中, N 为每一层的序列长度, $N_{k^{[\alpha]}, k^{[\beta]}}$ 是在第 α 层和第 β 层上度值分别等于 $k^{[\alpha]}$ 和 $k^{[\beta]}$ 的节点数

$$\sum_{k^{[\alpha]}} \sum_{k^{[\beta]}} N_{k^{[\alpha]}, k^{[\beta]}} = N \quad (4)$$

层间互信息 $I_{\alpha,\beta}$ 量化了同一节点在两个不同层之间度的相关性, $I_{\alpha,\beta}$ 值越大, 两层之间的度分布越相关, 因此时间序列的结构也越相关, 表明两个时间序列可能具有更加相似的行为。

对于图 1 中所描述的三路可视图，将其每层视为网络中的一个节点，两层之间的层间互信息 I 作为节点间连边的权重，则该多路可视图被映射为具有三个节点的无向加权网络，且该网络是全连通的。映射后的网络如图 2 所示。

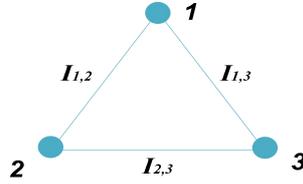


图 2 多路可视图映射后的网络。节点 1、2、3 对应多路可视图的第 1 层、第 2 层和第 3 层。 $I_{1,2}$ 、 $I_{1,3}$ 、 $I_{2,3}$ 分别为层 1 和层 2、层 1 和层 3、层 2 和层 3 之间的互信息，同时为对应两点连边的权重

Fig.2. Network mapped by multiplex visibility graph. Nodes 1,2 and 3 correspond to the 1th,2th and 3th layers of multiplex visibility graph. $I_{1,2}$, $I_{1,3}$, $I_{2,3}$ are the mutual information between layer 1 and layer 2, layer 1 and layer 3, layer 2 and layer 3 respectively, meanwhile, they are the weight of edges of corresponding two points.

2.2.2 权重度分布熵

权重度分布熵^[45]表征复杂网络的拓扑结构信息，反应网络的连接复杂度。当网络的连边较多、结构更为规则时，其权值分布较均匀，表现为权重度分布熵较大；反之，当连边较少、结构不规则时，复杂网络的平均度较小，某些权值高频出现，权重度分布熵较小。节点 i 的权重度分布熵表示为

$$E(i) = -\sum_{j=1}^n a(i, j) \log(a(i, j)) \neq 0 \quad (5)$$

式中， $a(i, j)$ 表示节点 i 和节点 j 连边权重的概率分布, 公式为

$$a(i, j) = \frac{I_{ij}}{\sum_{m=1}^n I_{im}} \quad (6)$$

其中， I_{ij} 为多路可视图的第 i 层和第 j 层之间的层间互信息，根据 2.2 中的映射准则，在将多路可视图映射为复杂网络时，其表示节点 i 和节点 j 之间的权重。为便于理解、阅读方便，下文在提到节点 i 和节点 j 之间的权重时均用 w_{ij} 表示。 n 代表与节点 i 相连的节点数。

具有 N 个节点的加权网络的平均权重度分布熵可表示为

$$\bar{E} = \frac{1}{N} \sum_{i=1}^N E(i) \quad (7)$$

3 十二导联 ECG 信号的多路可视图

3.1 数据描述

我们将本文多路可视图运用到十二导联 ECG 信号的分析中。数据来自 PhysioNet 网站 (<https://www.physionet.org>) 的 PTB 心电数据库。PTB 数据库是由德国国家计量学会提供的 ECG 数据库^[46]，该数据库包含 52 名健康人及 238 名不同疾病患者（其中包括 148 名心梗患者、18 名心力衰竭患者、15 名束支性传导阻滞患者、14 名心律失常患者以及少数其他疾病的患者），每个人对应 1-5 条记录，每条记录包含十二导联 ECG 信号（3 个肢体标准导联 I、II、III，3 个加压肢体导联 avr、avf、avl 和 6 个胸导联 v1-v6），表 1 描述了该十二导联在检测心脏疾病时控制的具体位置。PTB 心电数据库使用的采集设备由 14 个用于 ECG、1 个用于呼吸以及 1 个用于线路电压的通道组成，输入电压为 $\pm 16\text{mV}$ ，补偿偏置电压为 $\pm 300\text{mV}$ ，分辨率为 16 位，每个采样值为 2 个字节，采样频率为 1kHz。

表 1 十二导联名称及其对应的心脏位置

Table 1. Name of the 12-lead and their corresponding cardiac location

导联名称	心脏部位
I、avl	心脏左室前侧壁
II、III、avf	心脏后壁、下面
avr	心室内腔
v1、v2	右室
v3、v4	室间隔
v5、v6	左室

我们选取 PTB 数据库中健康人和心梗患者的十二导联 ECG 信号进行分析，所有的心梗患者数据中，除去无年龄信息的受试者外，有 26 名 70 岁以上的心梗患者没有与之年龄匹配的健康人群，有 13 名 30 岁以下的健康人群没有与之匹配的心梗患者。将剩余的所有受试者根据年龄分为 4 组，4 组受试者的年龄范围分别为 60-69 岁（包含 7 名健康受试者和 47 名心梗患者）、50-59 岁（包含 11 名健康受试者和 30 名心梗患者）、40-49 岁（包含 6 名健康受试者和 23 名心梗患者）和 30-39 岁（包含 10 名健康受试者和 4 名心梗患者）。PTB 数据集已对心电信号进行预处理，所以在本研究中，无需对数据进行再次处理。将每个受试者的十二导联数据片段各选择其 5000 个顺序数据点进行实验。

3.2 多尺度分析

Costa^[47]等人提出时间序列多尺度粗粒化的方法。对于一个长度为 N 的时间序列 $\{x(i), i=1, 2, \dots, N\}$ ，根据粗粒化过程定义时间尺度，得到粗粒化时间序列 $\{y^s(j), j=1, 2, \dots, N/s\}$ ，即

$$y^s(j) = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x(i), 1 \leq j \leq \frac{N}{s} \quad (8)$$

其中, s 代表尺度因子, N 为原始时间序列长度。粗粒化的时间序列长度等于原始序列长度 N 除以尺度因子 s 。当 $s=1$ 时, 序列 $y^{(1)}$ 与原始时间序列等值。

我们首先对两类人群原始十二导联 ECG 信号进行粗粒化, 初始序列长度为 $N=5000$, 尺度因子 s 的取值范围为 1 至 50。经实验, 当尺度因子选择过大即序列长度很短时, 所采集的信号不能准确反映受试者的心脏状态。在不同尺度因子 s 下, 分别将获得的新序列构成多路可视图, 并映射到复杂网络, 提取平均权重分布熵。

计算每组数据中健康人和心梗患者的平均权重分布熵, 结果如图 3 所示。

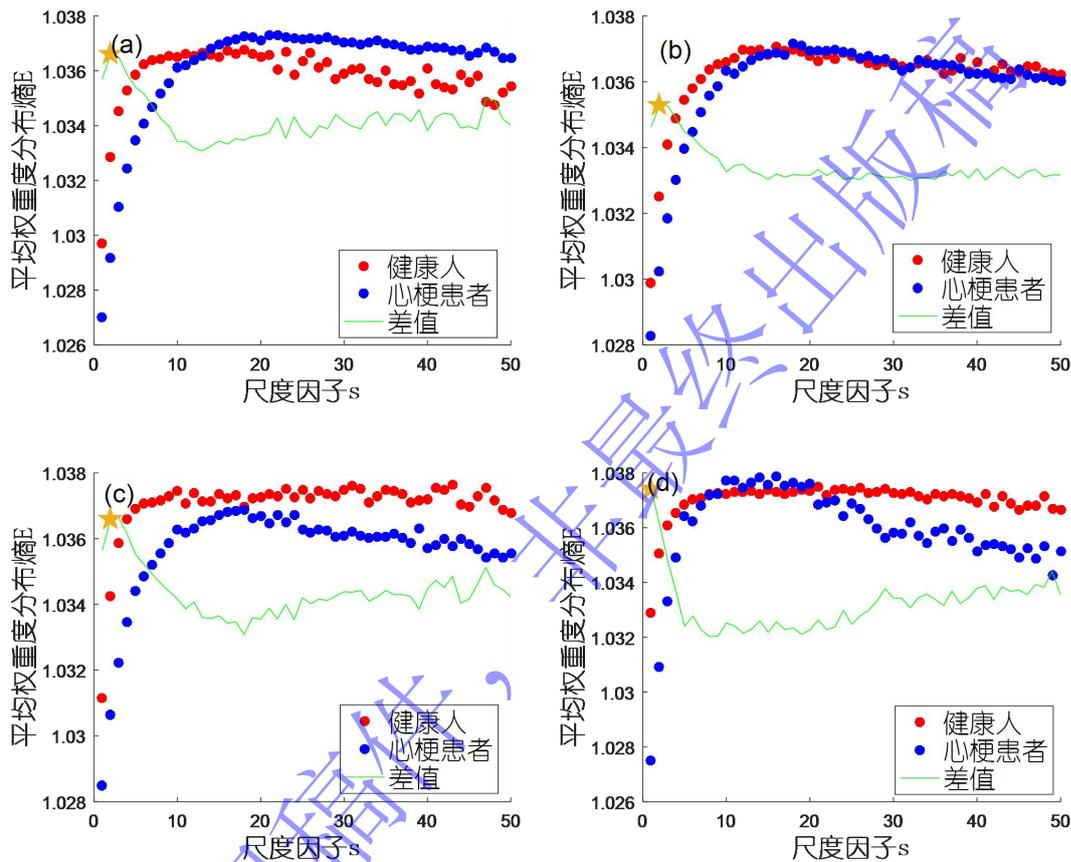


图 3 健康人和心梗患者在不同尺度因子 s 下平均权重分布熵。四组受试者对应的年龄阶段分别为 (a) 60-69 岁; (b) 50-59 岁; (c) 40-49 岁; (d) 30-39 岁。绿色实线为两类人群平均权重分布熵按比例缩放后的差值, 星号表示差值最大处

Fig.3. Average weighted distribution entropy of healthy people and patients with MI under different scale factor s . Age of four groups are (a) 60-69 years old; (b) 50-59 years old; (c) 40-49 years old; (d) 30-39 years old respectively. The solid green line is the difference of average weighted distribution entropy of two groups after scaling, and asterisk represents the maximum difference.

图 3 描述了四组不同年龄阶段的健康人与心梗患者在尺度因子增大时平均权重分布熵的变化情况。由于短序列更容易产生相似的度分布, 两层之间拥有更大的互信息值, 全连通网络的权重分布更加均匀, 因此, 随着尺度因子 s 在 1-10

范围内的增加，两类人群的平均权重度分布熵均明显增大，但当序列粗粒化到一定程度时，该变化消失，平均权重度分布熵值保持稳定。

观察图 3 中的差值曲线，对于不同年龄阶段的两类人群，当 s 取值约为 1-10 时，随着尺度因子的增加，平均权重度分布熵的差值曲线呈下降趋势，而此曲线在 s 大于 10 时出现波动。因此，健康人和心梗患者的十二导联 ECG 信号在短序列下的识别较为困难。此外，我们发现，当 s 为 1-10 时，不同年龄阶段健康受试者的平均权重度分布熵总是大于心梗患者，表明健康受试者对应复杂网络的权值分布更为平均，拥有更加规则的结构。

4 重构网络

全连通网络具有完全相同的网络拓扑结构，无法对两类人群 12 导联 ECG 信号的潜在动力学结构进行直观判断。因此，我们引入阈值，重构映射网络。由于重构网络的结构对于每一个个体都是唯一的，并且在不同人群之间存在较大差异，可以提供一种更加简单和直观的方法来描绘复杂系统中不同的动力学特性。

4.1 构造方法

因为连边权重由多路可视图的层间互信息决定，我们定义阈值为最大层间互信息值的常数倍。对于初始全连通网络，当网络中两点的连边权重 I 小于给定阈值时，删除该权重对应的连边。否则，保留该连边。如果常数 c 取值过小，阈值也相应的取较小值时，两类人群的重构网络中小于给定阈值的权重寥寥无几，删去有限连边后，两类人群的重构网络仍然是近似全连通的，网络参数差异较小。因此，我们令 c 的取值范围在 0.4 至 0.9 之间。为了获得更加精确的阈值区间，我们将 c 的步长设置为 0.01。

基于以上考虑，定义阈值计算公式为

$$\omega = c * \max(\{I_{\alpha,\beta}\}), c \in [0.4, 0.9] \quad (9)$$

其中， $\{I_{\alpha,\beta}\}$ 是多路可视图中所有层间互信息值的集合， c 为常数，表示倍数。通过调节 c 的大小，可以获得最优的阈值区间。

常数 c 的取值决定了重构网络中连边的数量，十二导联代表的节点之间度分布的相关性取决于每两个节点之间的层间互信息。通过重构网络，部分连边被删除，这一相关性表现在代表十二个导联的节点属性上。因此，我们提取描述节点属性的平均权重度 $\langle d \rangle$ 。平均权重度越大，表明该导联与其余导联之间的度分布相关性越高。复杂网络平均权重度 $\langle d \rangle$ 的计算公式为

$$\langle d \rangle = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N w_{ij} \quad (10)$$

其中， w_{ij} 为节点 i 和节点 j 之间的权重， N 代表节点总数。

在不同阈值下，分别对健康受试者和心梗患者的全连通网络进行重构，并计算重构网络的平均权重度，结果如图 4 所示。寻找使两类人群平均权重度差值最大的阈值区间，在此区间内，两类人群重构网络的潜在动力学特征差异最为明显，识别效果最佳，因此可以进一步精确阈值范围。

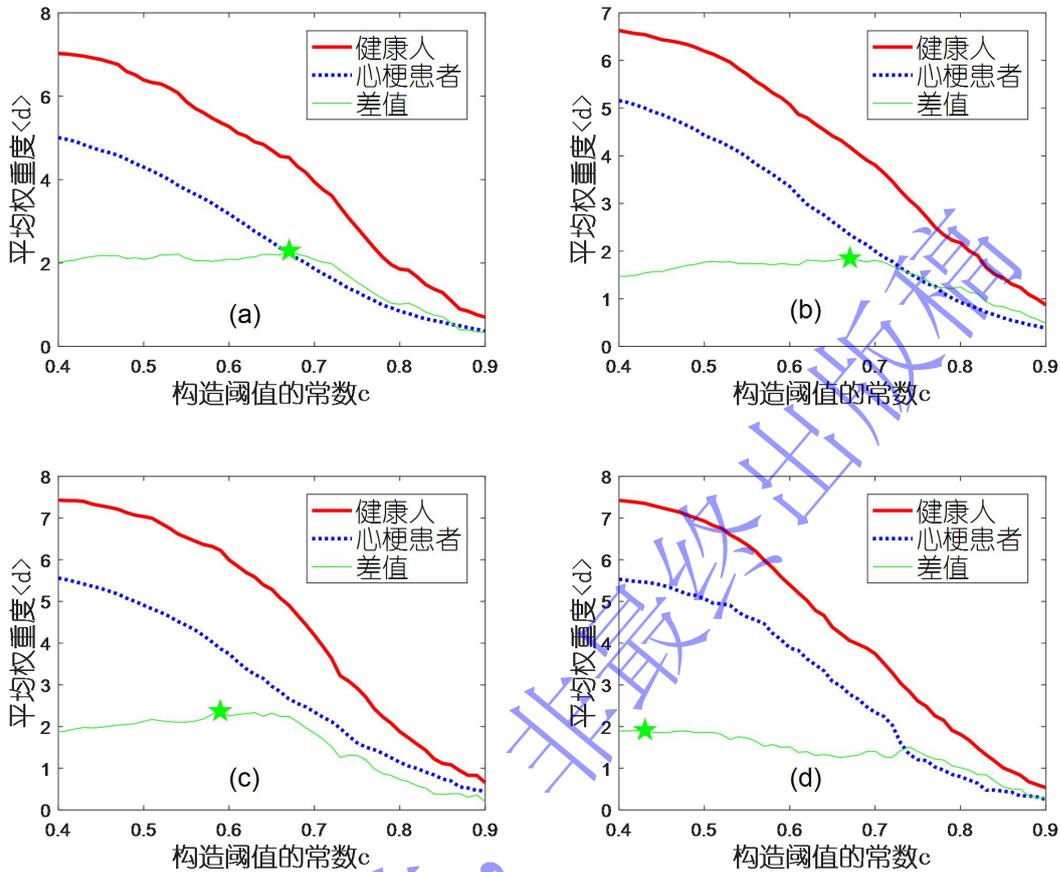


图 4 健康受试者与心梗患者在常数 c 不同取值时的平均权重度。四组受试者对应的年龄阶段分别为 (a) 60-69 岁；(b) 50-59 岁；(c) 40-49 岁；(d) 30-39 岁。绿色实线为两类人群平均权重度的差值，星号表示差值最大处，其值见表 2

Fig.4. Average weighted degree of healthy subjects and myocardial infarction patients at different values of constant c . Age of four groups are (a) 60-69 years old; (b) 50-59 years old; (c) 40-49 years old; (d) 30-39 years old. The solid green line is the difference of the average weighted degree of two groups, and the asterisk represents the maximum difference, see table 2 for specific values.

根据图 4 可以发现，随着常数 c 的增大，心梗患者和健康人群平均权重度均逐渐减小并接近 0。这是由于当倍数 c 取值较大时，阈值对应为较大值，导致多数连边被删除，网络拓扑结构中易产生孤立点。此外，我们发现，不同年龄阶段健康人群的平均权重度总是大于心梗患者，表明健康人的重构网络具有更加规则的结构，与图 3 所得结论相呼应，证明健康人对应的多路可视图网络具有更高的复杂性和连接性。

为了精确阈值区间，我们在图 4 中作出两类人群平均权重度的差值曲线。可以看到，当常数 c 在 0.67 范围以内时，心梗患者和健康人群平均权重度 $\langle d \rangle$ 的差值曲线基本保持稳定。 c 较大时，两类人群平均权重度的差距逐渐减小，即网络拓扑结构越来越相似。表 2 列出了 2.1 中 4 组健康人与心梗患者在平均权重度差值最大时的常数 c 值。根据表 2 中的数据并结合图 4，我们得出，当倍数 c 取值在 0.43-0.67 范围时，健康人和心梗患者对应重构网络的平均权重度差异明显，网络拓扑结构上的差异也最为显著。

表 2 四组健康人-心梗患者对应重构网络的平均权重度差值最大时的常数 c

Table 2. Constant c of four groups when difference of average weighted degree of reconstructed network is the maximum

	60-69 岁	50-59 岁	40-49 岁	30-39 岁
倍数 c	0.67	0.67	0.59	0.43
差值	2.2895	1.8389	2.3592	1.9015

4.2 拓扑结构

为探究十二导联 ECG 信号的潜在动力学特征，作出重构网络拓扑图。由 4.1，当 c 取值在 0.43-0.67 范围内时，两类人群的网络参数存在显著差异。由于 $c=0.43$ 时可删除最少的连边重构网络，其与原始网络的结构最相似。因此，我们取 0.43 作为 c 值，计算阈值并重构网络，健康人和心梗患者的网络拓扑图如图 5 所示。通过观察，两类人群的网络拓扑结构存在显著差异。健康受试者的网络拓扑图拥有更加复杂的结构，节点连接更为紧密，表明健康人多路可视图的层间互信息较大，十二导联 ECG 信号之间具有更加相似的结构；而心梗患者对应的网络拓扑图节点之间的连接相对稀疏，甚至出现对应导联 i 和导联 avr 的两个孤立节点，这两个导联与其他导联的度分布相关性较低。

此外，网络拓扑图中节点的颜色和大小与该节点权重度成正比，计算公式为

$$d_i = \sum_{j=1}^N w_{ij} \quad (11)$$

其中， w_{ij} 代表节点 i 和节点 j 之间的权重， N 代表节点数目。

由图 5 可以看出，无论健康人还是心梗患者，胸导联的权重度均大于肢体导联，在进行 ECG 信号识别时起到关键作用。胸导联由于距离人体心脏更近，对心脏电活动的反应更为强烈。因此，在对心梗进行检测时，这六个导联是需要重点关注的对象。对比图 5(a) 和图 5(b)，健康受试者和心梗患者的 v_5 、 v_6 导联对应节点的大小和颜色差异最大，我们或可以据此联合表 1 判断，心梗对该患者的左心室影响较为严重。因此，网络拓扑图或为心梗定位提供新的思路，然而，其实际效果及判断准确率等仍需进一步的论证。

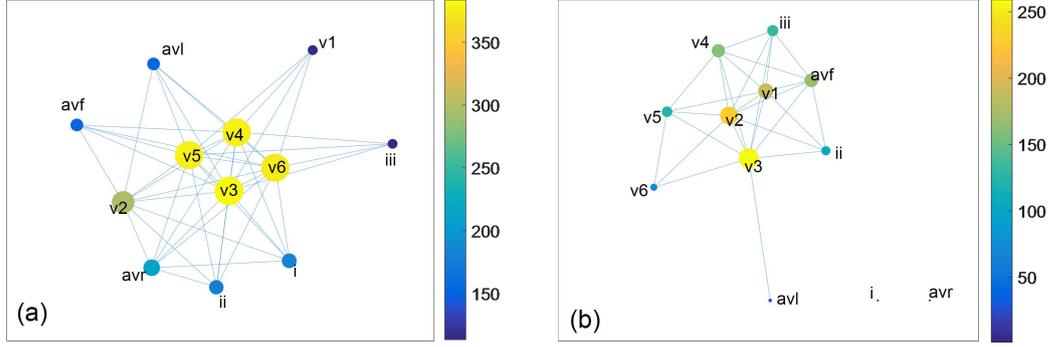


图 5 常数 $c=0.43$ 时重构网络的拓扑图: (a) 健康受试者; (b) 心梗患者。各节点对应的导联已在图中标明。节点大小和颜色对应应该节点的权重度。图右侧颜色柱表示节点权重度

Fig.5. Topology of the reconstructed network of (a) healthy subjects and (b) patients with MI at $c=0.43$. The corresponding leads of each node have been indicated in the figure. Size and color are relative to the weighted degree of the node. The color bar on the right of the figure indicates the value of weighted degree.

4.3 参数分析

为进一步研究两类人群重构网络的拓扑结构,我们计算重构网络的平均权重度 $\langle d \rangle$ (公式 10) 和平均加权聚类系数 C_w , 其中 C_w 量化了网络的连接性, 反应网络微观结构的规则性, 计算公式为

$$C_w = \frac{\sum_{i \in N} C_w^i}{N} \quad (12)$$

其中, N 代表网络中的节点数, C_w^i 是每个节点 i 的加权聚类系数

$$C_w^i = \frac{\sum_{j,l \neq i \in N} (w_{ij} w_{jl} w_{il})^{\frac{1}{3}}}{k_i(k_i - 1)} \quad (13)$$

其中, k_i 为节点 i 的度, w_{ij} 、 w_{jl} 、 w_{il} 分别代表节点 i 、 j 、节点 l 、 j 以及节点 i 、 l 之间的权重值。为避免实验偶然性,我们对 4 组健康人与心梗患者数据进行实验并作出箱图 (图 6), 对结果进行 t 检验, 发现 P 值远小于 0.05, 证明健康受试者和心梗患者的平均权重度和平均加权聚类系数之间存在显著差异。

根据图 6 我们可以得出, 相较心梗患者, 健康受试者拥有更高的平均权重度和平均加权聚类系数。从混沌理论的角度来看, 心脏作为自然界中最为复杂的非线性系统之一, 人体各系统的耦合作用使得心电信号呈现连续的上下波动, 形成复杂的时间序列。心梗患者由于心脏部分衰竭导致功能丢失, 其复杂性与正常人相比降低, 该结论与本文一致。此外, 由于心梗患者心脏功能减弱, 在不同心肌位置测得的心电信号表现为形态各异的结构, 构成多路可视图后的层间互信息较小, 在重构网络时, 有更多的连边被删除。因此, 心梗患者对应复杂网络的连接

性低，且各节点拥有较小的权重度。

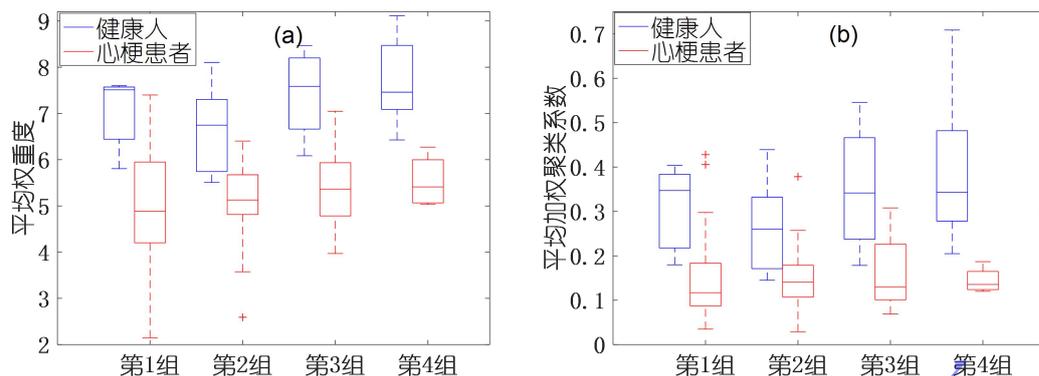


图6 健康人和心梗患者在最佳阈值下的(a)平均权重度和(b)平均加权聚类系数。第1组-第4组受试者的年龄范围分别为60-69岁、50-59岁、40-49岁以及30-39岁

Fig.6. (a) Average weighted degree and (b) average weighted clustering coefficients of healthy and myocardial infarction patients under optimal thresholds. Subjects in groups 1 to 4 correspond to ages of 60-69 years old, 50-59 years old, 40-49 years old and 30 to 39 years old respectively.

4.4 与其他方法对比结果

为了评价所提出算法的性能，将心梗信号作为正样本，健康信号作为负样本，并使用平均权重度和平均加权聚类系数两个指标进行分析，计算敏感性(sensitivity, SE)、特异性(specificity, SP)和准确率(accuracy, AC)，有

$$SE = \frac{TP}{TP + FN} \quad (14)$$

$$SP = \frac{TN}{TN + FP} \quad (15)$$

$$AC = \frac{TP + TN}{TP + FP + TN + FN} \quad (16)$$

式中，TP、FN、TN、FP表示分类正确的正样本个数、被错误分类为负样本的正样本个数、分类正确的负样本个数以及被错误分类为正样本的负样本个数。分类准确率代表了该算法对两类样本的总体分类正确率。

本文提出的基于多路可视图的心梗检测方法与现有的心梗检测方法结果进行了对比，基于PTB数据集的对比结果如表3所示。由表中数据可知，本文方法提取的参数权重度和加权聚类系数分别取得了94.4%、95.83%的敏感性和88.24%、82.35%的特异性，相比文献^[48-50]的检测方法，本文仅在特异性参数上低于Sun等^[49]提出的使用五阶多项式拟合，确定74维特征空间，然后应用粒子群优化器将权重建模为高斯分布后进行分类的方法。相比文献^[48]将多导联数据分别使用HMM提取特征后堆叠特征，本文考虑了多导联间的关联性，利用关联指数层间互信息将其构造为网络后进行特征提取，仅使用了一个网络特征即完成了心梗的检测，在使用特征个数上优于已有方法，尤其较文献^[50]中进行10个特征融合有了较大的提升。综上，本文方法相较已有方法存在优势，且分类准确率更高，达到93.3%，

能够实现心梗的有效检测。

表 3 本方法与其他现有方法的比较

Table 3. The comparison between proposed method and other existing methods

分类方法	特征个数	结 果		
		敏 感 性	特 异 性	准 确 率
HMM+GMM ^[48]	4	85.71%	79.82%	82.50%
KNN+集成学习 ^[49]	2	92.3%	88.1%	NA
随机森林 ^[50]	10	NA	NA	87%
AdaBoostM1 ^[50]	10	NA	NA	90%
平均权重度（本文算法）	1	94.4%	88.24%	93.3%
平均加权聚类系数（本文算法）	1	95.83%	82.35%	93.3%

5 结论

本文将十二导联 ECG 信号构造为多路可视图，通过层间互信息将其映射到复杂网络，为分析十二导联信号的多尺度特征，我们计算网络的多尺度权重度分布熵，发现其可以用作检测心梗患者的有效工具。然而，由于全连通网络具有完全相同的网络拓扑结构，无法唯一表征不同个体的动力学特征，因此，我们对映射网络进行重构。计算重构网络的平均权重度和平均加权聚类系数的结果表明，与健康人群相比，心梗患者的重构网络具有更低的复杂性和连接性，此方法在 PTB 数据库上的分类准确率达到 93.3%，实现了对心梗患者和健康人群的有效分类。此外，我们发现，随着患者年龄的增加，其重构网络平均权重度的均值减小，表明老年患者的重构网络拥有更低的边权重，十二导联之间相似的动力学行为减少，患病更加严重。因此，老年心梗患者的诊断和治疗亟需我们的关注。

多路可视图为多元时间序列分析提供了一种新颖、准确且易于扩展的方法。通过计算一些基础的多路参数，如层间互信息和平均边缘重叠，可以对多元时间序列进行联合分析，获取不同状态序列之间的内在联系，以实现非线性动力学系统的非线性动力学特性进行完整而全面的分析。此外，在使用多路可视图方法进行多元时间序列分析时，可以任意选择变量个数，这使得多通道联合分析变得更加灵活和高效。然而，本文在识别单维时间序列的内在特征时存在局限，仅关注了不同序列之间的整体相似性，因此，我们期待满足反应复杂系统时间演化和表征时间序列局部波动的新的映射方法的提出。

参考文献

[1] Zhang J, Small M 2006 *Phys. Rev. Lett.* **96** 238701

[2] Zou Y, Donner R V, Marwan N, Donges J F, Kurths J 2019 *Phys. Rep.* **787** 1

- [3] Lacasa L, Luque B, Ballesteros F, Luque J, Nuno J C 2008 *Proc. Natl. Acad. Sci.* **105** 4972
- [4] Luque B, Lacasa L, Ballesteros F, Luque J 2009 *Phys. Rev. E* **80** 046103
- [5] Donner R V, Zou Y, Donges J F, Marwan N, Kurths J 2010 *New J. Phys.* **12** 033025
- [6] Marwan N, Donges J F, Zou Y, Donner R V, Kurths J 2009 *Phys. Lett. A* **373** 4246
- [7] Xu X, Zhang J, Small M 2008 *Proc. Natl. Acad. Sci.* **105** 19601
- [8] Donges J F, Donner R V, Trauth M H, Marwan N, Schellnhuber H J 2011 *Proc. Natl. Acad. Sci.* **108** 20422
- [9] Zhou T T, Jin N D, Gao Z K, Luo Y B 2012 *Acta Phys. Sin.* **61** 030506 (in Chinese) [周婷婷, 金宁德, 高忠科, 罗跃斌 2012 物理学报 **61** 030506]
- [10] Sun X R, Small M, Zhao Y, Xue X P 2013 *Physica A* **24** 1054
- [11] Zhang J Y, Zhou J, Tang M, Guo H, Small M, Zou Y 2017 *Sci. Rep.* **7** 7795
- [12] Kulp C W, Chobot J M, Freitas H R, Sprechini G D 2016 *Chaos* **26** 073114
- [13] Zou Y, Small M, Liu Z, Kurths J 2014 *New J. Phys.* **16** 013051
- [14] Donner R V, Donges J F 2012 *Acta Geophys* **60** 589
- [15] Bajestani G S, Behrooz M, Khani A G, Nouri-Baygi M, Mollaei A 2019 *Comput. Methods Programs Biomed.* **177** 277
- [16] Tang X Y, Xia L, Liao Y Z, Liu W F, Peng Y H, Gao T X, Zeng Y J 2013 *Clin. EEG Neurosci.* **44** 150
- [17] Supriya S, Siuly S, Wang H, Cao J L, Zhang Y C 2016 *IEEE Access* **4** 6554
- [18] Bhaduri S, Ghosh D 2015 *Clin. EEG Neurosci.* **46** 218
- [19] Wang L, Long X, Arends J B, Aarts R M 2017 *J. Neurosci. Method* **290** 85
- [20] Kong T J, Shao J, Hu J Y, Yang X, Yang S Y, Malekian R 2021 *Sensors* **21** 1870
- [21] Yi N 2014 *M.S. Thesis* (Guangzhou: Jinan University) (in Chinese) [易娜 2014 硕士学位论文 (广州: 暨南大学)]
- [22] Shi W B, Shang P J, Wang J, Lin A J 2014 *Physica A* **403** 35
- [23] Jiang W, Wei B Y, Zhan J, Xie C H, Zhou D Y 2016 *Comput. Ind. Eng.* **101** 260
- [24] Azami H, Escudero J 2017 *Physica A* **465** 261
- [25] Han Y F, Jin N D, Zhai L S, Ren Y Y, He Y S 2019 *Physica A* **518** 131
- [26] Gao Z K, Liu M X, Dang W D, Ma C, Hou L H, Hong X L 2021 *chem. Eng. J.*

- [27] Tanizawa T, Nakamura T, Taya F, Small M 2018 *Physica A* **512** 437
- [28] Gao Z K, Fang P C, Ding M S, Jin N D 2015 *Exp. Therm Fluid Sci.* **60** 157
- [29] Bianconi G 2013 *Phys. Rev. E* **87** 062806
- [30] Nicosia V, Bianconi G, Latora V, Barthelemy M 2013 *Phys. Rev. Lett.* **111** 058701
- [31] Battiston F, Nicosia V, Latora V 2014 *Phys. Rev. E* **89** 032804
- [32] Quesada D, Valverde G, Larrañaga P, Bielza C 2021 *Eng. Appl. Artif. Intell.* **103** 104301
- [33] Lacasa L, Nicosia V, Latora V 2015 *Sci. Rep.* **5** 15508
- [34] Carmona-Cabezas R, Gómez-Gómez J, Ariza-Villaverde A B, de Ravé E G, Jiménez-Hornero F J 2020 *Atmos. Pollut. Res.* **11** 205
- [35] Gao Z K, Guo W, Cai Q, Ma C, Zhang Y B, Kurths J 2019 *Chaos* **29** 073119
- [36] Cai Q, An J, Gao Z 2021 *Brain Sci. Adv.* **6** 355
- [37] Samanta K, Chatterjee S, Bose R 2020 *IEEE Sensors Lett.* **4** 1
- [38] Thygesen K, Alpert J S, White H D 2007 *J. Am. Coll. Cardiol.* **50** 2173
- [39] Dohare A K, Kumar V, Kumar R 2018 *Appl. Soft Comput.* **64** 138
- [40] Sharma M, San Tan R, Acharya U R 2018 *Comput. Biol. Med.* **102** 341
- [41] Tripathy R K, Bhattacharyya A, Pachori R B 2019 *IEEE Sensors J.* **19** 11437
- [42] Acharya U R, Fujita H, Oh S L, Hagiwara Y, Tan J H, Adam M 2017 *Inf. Sci.* **415** 190
- [43] Xiong P, Xue Y P, Liu M, Du H M, Wang H R, Liu X L 2020 *J. Biomed. Eng.* **37** 142 (in Chinese) [熊鹏, 薛彦平, 刘明, 杜海曼, 王洪瑞, 刘秀玲 2020 生物医学工程学报 **37** 142]
- [44] Liu W H, Huang Q J, Chang S, Wang H, He J 2018 *Biomed. Signal Process. Control* **45** 22
- [45] Zhang H Y, Meng Q F, Du L, Liu M M 2019 *Chin. J. Biomed. Eng.* **38** 273 (in Chinese) [张汉勇, 孟庆芳, 杜蕾, 刘明敏 2019 中国生物医学工程学报 **38** 273]
- [46] Goldberger A L, Amaral L, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K, Stanley H E 2000 *Circulation* **101** e215
- [47] Costa M, Goldberger A L, Peng C K 2002 *Phys. Rev. Lett.* **89** 068102

- [48] Chang P C, Lin J J, Hsieh J C, Weng J 2012 *Appl. Soft Comput.* **12** 3165
- [49] Sun L, Lu Y P, Yang K T, Li S Z 2012 *IEEE Trans. Biomed. Eng.* **9** 3348
- [50] Wang D L 2020 *M.S. Thesis* (Hangzhou: Hangzhou Danzi University) (in Chinese)
- [王丹俐 2020 硕士学位论文 (杭州: 杭州电子科技大学)]

ECG signals' complex network recognition for health and Myocardial Infarction based on multiplex visibility graph*

Zhiyi Ma¹⁾ Xiaodong Yang^{1)†} Aijun He²⁾ Lu Ma¹⁾ & Jun Wang³⁾

1) (School of Computer Science and Technology, China University of Mining and Technology, Xuzhou
221116, China)

2) (School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China)

3) (School of Geographic and Biologic Information, Nanjing University of Posts and
Telecommunications, Nanjing 210023, China)

Abstract

The Visibility Graph algorithm has been proved to be a simple and efficient method to transform time series into complex network and has been widely used in time series analysis because it can inherit the dynamic characteristics of original time series in topological structure. Now, visibility graph analysis of univariate time series has driven to mature, however, most complex systems in real world are multi-dimensional, it is difficult for univariate analysis to describe the global characteristics when applied to multi-dimensional series. In this paper, a novel method for multivariate time series analysis is proposed. For patients with myocardial infarction and healthy subjects, the 12-lead electrocardiogram signals of each individual are considered as a multivariate time series, which is transformed into a multiplex visibility graph through visibility graph algorithm and then mapped to fully connected complex network. Each node of the network corresponds to a lead and the inter-layer mutual information between visibility graph of two leads represents the weight of edges. Due to the fully connected network of different groups show identical topological structure, the dynamic characteristics of different individuals cannot be uniquely represented. Therefore, we

reconstruct the fully connected network according to inter-layer mutual information, when the value of inter-layer mutual information is less than the threshold we set, edge corresponding to the inter-layer mutual information is deleted. We extract average weighted degree and average weighted clustering coefficient of reconstructed networks so as to recognize the 12-lead ECG signals of healthy subjects and myocardial infarction patients. Moreover, multiscale weighted distribution entropy is also introduced to analyze relationship between the length of original time series and final recognition results. Due to higher average weighted degree and average weighted clustering coefficient of healthy subjects, their reconstructed networks show a more regular structure, higher complexity and connectivity, and could be distinguished from patients with myocardial infarction, whose reconstructed networks are sparser. Experiment results show that the identification accuracy of both parameters, average weighted degree and average weighted clustering coefficient, reach 93.3%, which can distinguish the 12-lead ECG signals of healthy people and patients with myocardial infarction, and realize the automatic detection of myocardial infarction.

Keywords: Myocardial Infarction, multivariate time series, multiplex visibility graph, complex network

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61772532 , 61876186, 11774167).