

# 基于时序区分子图的阿尔茨海默症辅助诊断方法

信俊昌, 郭恩铭, 张嘉正

(东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

**摘 要:** 为解决现有区分子图方法在解决阿尔茨海默症辅助诊断上忽略脑网络动态连接变化的问题,提出一种基于时序区分子图的辅助诊断方法. 将功能磁共振成像经过处理后形成二值矩阵并使同一测试者的多张动态脑网络形成时序差异图,之后进行频繁差异子图挖掘、频繁差异序列挖掘,进而筛选出保留脑网络时序特性的生物标记物——时序区分子图. 获取 ADNI 公开数据集的一组数据进行实验,通过与现有的早期阿尔茨海默症辅助诊断方法进行大量的实验对比,证明本文方法的辅助诊断准确率在该数据集上提高了 12.7%,进而证明所提方法的有效性.

**关 键 词:** 阿尔茨海默症;时序区分子图;动态脑网络;功能磁共振成像;gSpan 算法

**中图分类号:** TP 391      **文献标志码:** A      **文章编号:** 1005-3026(2022)08-1089-08

## Aided Diagnosis Method of Alzheimer's Disease Based on Sequential Discriminative Subgraph

XIN Jun-chang, GUO En-ming, ZHANG Jia-zheng

(School of Computer Science & Engineering, Northeastern University, Shenyang 110169, China. Corresponding author: GUO En-ming, E-mail: 2671290806@qq.com)

**Abstract:** In order to solve the problem of ignoring the dynamic connection changes of the brain network in the aided diagnosis of Alzheimer's disease by the existing discriminative subgraph method, an aided diagnosis method based on the sequential discriminative subgraph was proposed. The functional magnetic resonance imaging (fMRI) is processed to form binary matrices, and multiple dynamic brain networks of the same subject form sequential difference graphs, and then frequent subgraph mining and frequent sequence mining are performed to screen out biomarkers (sequential discriminative subgraph) that retain the time sequence characteristics of the brain network. A set of data from the ADNI public data set was obtained for experiments, and a large number of experimental comparisons were conducted with the existing early Alzheimer's disease aided diagnosis methods, which proves that the aided diagnosis accuracy of this method is improved by 12.7% on this data set and the effectiveness of the proposed method.

**Key words:** Alzheimer's disease; sequential discriminative subgraph; dynamic brain network; functional magnetic resonance imaging (fMRI); gSpan algorithm

阿尔茨海默症 (Alzheimer's disease, AD) 是一种严重的神经退化疾病,也是一场困扰全球的记忆危机. 根据《中国阿尔茨海默症报告 2021》显示,60 岁及以上人群中 有 1 507 万例痴呆患者,其中 AD 患者高达 983 万例. 近年,功能磁共振成像 (functional magnetic resonance imaging, fMRI) 技术已经成为一种观察脑结构和功能的手段,并日

趋成熟与完善. 在 fMRI 图像中,人脑的不同脑区之间存在着连接与功能上的关联性,构建和分析脑网络可以更好地描述大脑的活动状态以及各个脑区之间的交互关系. 生物标记物可在分子水平上为疾病提供敏感而稳定的诊断手段<sup>[1]</sup>,通过寻找出两类具有特异性的标记物,用以阿尔茨海默症辅助诊断.

早期的阿尔茨海默症辅助诊断已被国内外广泛研究. 在基于机器学习方法进行阿尔茨海默症诊断中,Jo 等<sup>[2]</sup>采用深度学习方法构建卷积神经网络对从医学影像中提取的重要特征进行训练分类,从而进行诊断. Jie 等<sup>[3]</sup>提出一种图核方法计算脑网络相似性从而实现疾病诊断;之后大量的新型图核不断被构建来优化算法<sup>[4]</sup>. 近几年,越来越多的研究将脑网络的纵向数据融入研究,通过从脑网络中提取重要特征(如脑体积、相应皮质区域面积等),再通过不同样例不同时间点取样的方法获取到脑网络的动态特征信息,然后进行基于神经网络模型的 AD 诊断. Nguyen 等<sup>[5]</sup>提出使用 minimalRNN 模型进行时序信息填充,探究脑网络重要特征时序变化. Mansu 等<sup>[6]</sup>提出可解释 GNN 模型,充分利用动态脑网络的纵向数据判断脑网络重要特征对辅助诊断的影响.

图论领域知识被引入到对脑功能网络的研究中,使得复杂的脑以一种抽象化的模型被不断深入研究. Cao 等<sup>[7]</sup>就曾提出利用核磁共振脑网络的区分子图作为标记物诊断双相情感障碍问题. 区分子图不仅能够反映脑网络的拓扑结构,还能够更容易地捕捉到病变区域网络内连接模式并反映正常人与患者之间的差异,可以作为阿尔茨海默症辅助诊断的生物学标记物. 基于挖掘区分子图的算法也处于不断研发的进程中<sup>[8-9]</sup>. 尽管区分子图能够更好地反映脑网络的拓扑结构,还能够更容易地捕捉到病变区域网络内连接模式并反映正常人与患者之间的差异,可以作为神经精神疾病辅助诊断时的生物学标记物,但已有的研究成果主要基于提取静态脑网络的区分子图<sup>[10-11]</sup>. 静息态 fMRI 反映了被试者一段时间内的功能认知状态,构建静态脑网络往往会忽视时间维度上的信息,不能充分利用 fMRI 技术时间分辨率高的特点. 相反,使用不同时间段构建的动态脑网络,能够表现大脑的动态连接变化,进而反映每一时间段大脑功能活动信息. 为了更好地理解大脑功能活动需要充分利用时间信息,因此采用构建动态脑网络并且提出基于时序区分子图的方法进行阿尔茨海默症的辅助诊断.

综上所述,可以通过寻找一种可靠的生物标记物来进行早期的阿尔茨海默症的辅助诊断. 不同于机器学习领域的思想方法,区分子图的思想更能够保留脑网络的拓扑结构,有效地寻找出病变区域的差异性. 而本文提出的时序区分子图方法展现了脑网络的动态连接特性,能够有效提高阿尔茨海默症辅助诊断准确率.

# 1 时序区分子图在阿尔茨海默症辅助诊断中的应用

本文方法的整体技术路线一共分为五个步骤,包含动态脑网络构建、基于动态脑网络的频繁差异子图挖掘、基于频繁差异子图的频繁差异序列挖掘、频繁差异序列的筛选、基于时序区分子图的阿尔茨海默症辅助诊断.

## 1.1 动态脑网络的构建

首先进行数据预处理,从 ADNI 获得静息态 fMRI 数据之后,使用 Freesurfer 软件对数据进行预处理. 数据预处理流程如图 1 所示.

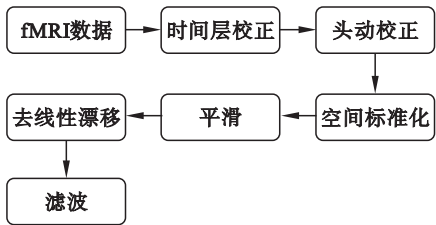


图 1 数据预处理流程图  
Fig. 1 Data processing flow chart

在完成数据预处理的工作后,进行动态脑网络的构建. 传统的脑功能网络多是基于标准模板而建立的脑区级功能网络. 以自动解剖图集(anatomical automatic labeling, AAL)模板<sup>[12-13]</sup>为例,首先将经过预处理的 fMRI 图像与 AAL 模板匹配,将匹配结果中与小脑相关的 26 个区域去除,对剩余的 90 个脑区的时间序列采用滑动时间窗处理<sup>[14]</sup>,将一个完整的时间序列划分为若干个窗口时间序列,采用皮尔逊偏相关系数度量<sup>[15]</sup>计算每个窗口中各个脑区之间的相关性,通过设定阈值得到多个时间段下的脑功能网络,由此构成了动态脑功能网络模型,如图 2 所示.

## 1.2 基于动态脑网络的频繁差异子图挖掘

通过 AAL 模板化之后,构成了  $90 \times 90$  的矩阵,用 0 表示 2 个脑区之间没有连接,用 1 表示 2 个脑区之间存在连接. 多张二值矩阵构成的连续图像可以代表动态脑网络.

首先需要构建差异图,将脑网络时序变化特征呈现出来,然后挖掘频繁出现的变化. 将相邻 2 个矩阵做差,保留时间在后的矩阵与前一时刻矩阵的差值矩阵. 1 表示在下一时刻 2 个脑区间出现增边现象, -1 表示出现减边现象, 0 表示在相邻的时间内边没有发生变化,经此处理能够使差异被放大,提高算法的效率.

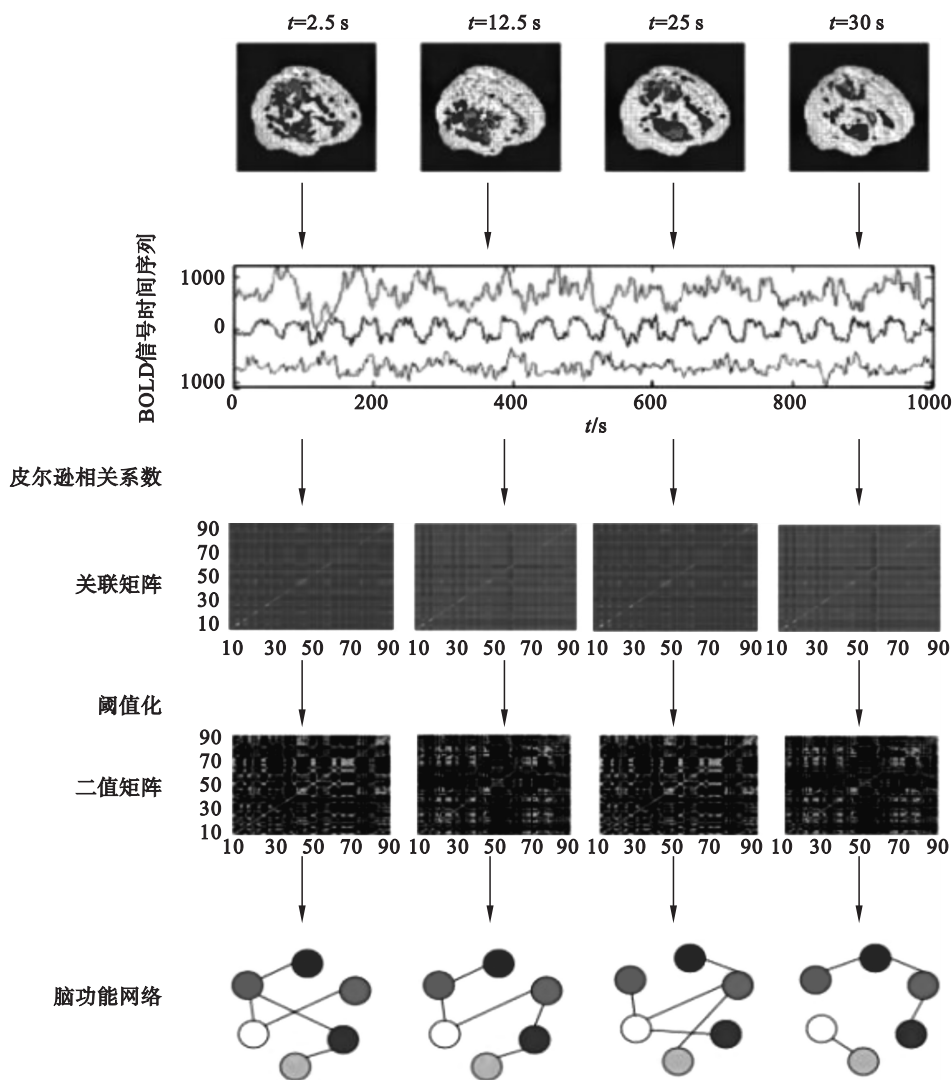


图 2 脑网络的构建  
Fig. 2 Construction of brain network

随后,进行频繁差异子图的挖掘.该算法的目标是找出所有满足支持度要求的差异子图.在频繁差异子图挖掘过程中,最重要的是子图同构的判定问题,这将直接影响算法的时间复杂度.因此通过最小 DFS (depth first search) 序的方法进行剪枝(即提早排除那些与已被搜索过的子图同构的子图),极大地提高了算法的执行效率.

频繁差异子图挖掘的大体步骤是由频繁边不断扩展,最终得到符合设定阈值的子图,以下详细介绍了本算法的不同环节的核心步骤.

1.2.1 频繁边挖掘

频繁边是指在图集中累计出现次数高于阈值要求的边的集合.通过筛选频繁边,对部分不满足阈值要求的子图进行剪枝,减少频繁差异子图挖掘过程中的枚举数量,有效控制搜索空间的大小.

假设  $G$  表示  $n$  张脑网络图的集合,  $g_1$  是寻找到的满足频繁差异子图阈值( $\eta$ )的差异子图,其在图集中共出现  $m$  次,那么由频繁差异子图定义

可知  $m \geq \eta \times n$ . 又因为差异子图  $g_1$  是由若干条边组成的,这些边在图集中出现的次数也应大于等于  $m$ ,由此可得所有频繁差异子图都是由频繁边组成的.同时考虑到在阈值相同的情况下频繁边数量远大于频繁差异子图的数量,为了控制算法整体的时间复杂度,需要将频繁边阈值( $\theta$ )设置为略大于  $\eta$  的合理值.所以,在执行频繁差异子图挖掘算法前需要对数据进行预处理,将那些不满足  $\theta$  要求的边删掉,提高挖掘算法的效率.具体挖掘算法如下所示:

算法 1 频繁边挖掘算法

Input: 图集  $G = \{g_1[N][N], g_n[N][N]\}$ , 最小支持度  $\theta$   
Output: 频繁边集合  $g[N][N]$   
1. for  $i = 1$  to  $N$  do  
2. for  $j = 1$  to  $N$  do  
3. cnt = 0; //计数变量  
4. for  $k = 1$  to  $n$  do



```
5.  if  $g_k[i][j] \neq 0$ 
6.  cnt = cnt + 1;
7.  if cnt >  $n \times \theta$ 
8.   $g[i][j] = 1$ ;
9.  else
10.  $g[i][j] = 0$ ;
11. Return g.
```

首先枚举图集中可能出现的所有边(第 1~2 行),然后分别对每条边统计其在图集中出现的次数(第 3~6 行),随后根据出现次数与最小支持度间的关系确定频繁边集合中是否加入该边(第 7~10 行),最后得到满足最小支持度要求的频繁边集合。

1.2.2 子图扩展算法

在进行频繁差异子图搜索过程中,需要构建子图搜索树来完整表示搜索空间,其中树上的每个节点存放 1 张全部由频繁边组成的联通子图,树的根节点表示空子图,即不含有任何一条边的子图,树中的孩子节点均是由其父亲节点扩展一条新边得到的.使用这种方法就可以不断对当前节点扩展新边来实现搜索树的生长,直至包含整个搜索空间。

对于搜索树中任意节点所包含着的联通子图  $g(V,E)$ ,将子图中所有节点  $V = \{v_1, v_2, \dots, v_n\}$  根据其发现顺序进行排序,最后发现的节点称为最右节点  $v_R^{NR}$ ,从最先发现的节点到最右节点的最短路径称为最右路径,记作  $R = (v_1, v_2, \dots, v_R)$ .对于加入子图中的新边  $e = (v_i, v_j)$ ,根据新边的 2 个顶点与当前联通子图的关系,可以将其分为前向边与后向边两类.其中,前向边是指起点  $v_i$  为最右顶点且终点  $v_j$  为最右路径上非最右顶点的任意节点的频繁边,而后向边则是指起点  $v_i$  为最右路径中的节点且终点  $v_j$  为不包含在当前子图中的新节点的频繁边.根据文献[8]可知,通过不断地对当前节点递归扩展其前向边与后向边,直至所有节点均被扩展,此时得到的搜索树完全包含搜索空间.因此,子图扩展算法仅需对当前节点的联通子图扩展前向边与后向边即可,具体算法如下所示:

算法 2 子图扩展算法

```
Input: 频繁边集合  $g[N][N]$ ,待扩展子图  $g_1(V_1,E_1)$ ;  $R$  为最后结点下标序号
Output: 扩展成功标志,扩展生成子图  $g_2(V_2,E_2)$ 
1.  $g_2 = g_1$ ;
2. for  $i = 1$  to  $(R - 1)$  do
3. if  $g[v_R][v_i] = 1$ 
```

```
4.  $g_2 \leftarrow E(v_R, v_i)$ ;
5. Return( True,  $g_2$  );
6. for  $i = 1$  to  $k$  do
7. for  $j = 1$  to  $N$  do
8. if  $v_j$  not in  $g_1$  and  $g[v_j][v_i] = 1$ 
9.  $g_2 \leftarrow E(v_j, v_i)$ ;
10. Return( True,  $g_2$  );
11. Return( False,  $g_2$  ).
```

首先将扩展生成子图赋值为待扩展子图(第 1 行);然后对于最右路径上的非最右节点,观察是否存在从最右节点连向该节点的频繁边,若存在则将该边加入生成子图并返回,否则继续执行算法(第 2~5 行);随后对于最右路径上的每个顶点,判断是否存在从当前节点出发的后向边,若存在则扩展并返回,否则继续执行算法(第 6~10 行);若此时算法仍未返回则认为搜索树中当前节点不再存在新的后继节点,返回失败标志(第 11 行).这样就按照先前向边再后向边的顺序完成子图扩展。

1.2.3 频繁差异子图挖掘

频繁差异子图是图集中出现次数大于阈值限制的差异子图的集合.这里所采用的频繁差异子图挖掘算法基于 gSpan 算法并结合实际进行多种剪枝操作,使得总体的时间复杂度进一步降低,具体算法如下。

算法 3 频繁差异子图挖掘算法

```
Input: 频繁边集合  $g[N][N]$ ,待扩展子图  $g_1(V_1,E_1)$ .
Output: 频繁差异子图搜索树
1. root = new node;
2. while extend_edge( $g_1$ )
3.  $g_2 = \text{extend\_edge}(g_1)$ ;
4. if check( $g_2$ ) = False
5. continue;
6. root ← create_tree( $g_2$ );
7. Return root.
```

该算法通过构建子图搜索树进行频繁差异子图的挖掘.其特征为:树中节点均代表 1 张满足频繁差异子图阈值要求的子图,树的根节点表示 1 张空子图,且对于树上任意 2 个相邻的节点,孩子节点都是父亲节点通过执行生长算法扩展一条边得到的.因此,频繁差异子图挖掘算法采用深度优先搜索的方法模拟搜索树的生长过程.首先,初始化根节点为空子图.然后,对于当前节点不断执行子图扩展算法.若扩展算法执行成功返回生长新边后得到的子图,则记录当前生长情况并将新得到的子

图作为当前节点递归执行挖掘算法.直至子图扩展算法无法加入新边时停止搜索并返回.这样就可以得到包含完整搜索空间的子图搜索树.

为了进一步提高算法执行效率,对于 gSpan 进行一系列改进,以优化频繁差异子图搜索过程,引进如下三种剪枝策略:①对于树中任意 2 个相邻的节点,孩子节点都是在父亲节点的基础上扩展得到的,假设父亲节点所代表的子图在图集中出现的次数为  $n_f$ ,孩子节点为  $n_s$ ,显然存在  $n_f \geq n_s$ .因此,可以在搜索树扩展过程中引入检查步骤,如果扩展得到了不满足约束的节点就执行剪枝操作,不再搜索它的孩子节点.②对于任意 1 张子图,它所对应的最小 DFS 序都是唯一确定的,根据子图扩展算法可以保证搜索树中存在一条从根节点出发按照最小 DFS 序进行扩展得到当前子图的路径.因此,对于从根节点到当前节点路径与其对应子图最小 DFS 序不一致的节点进行剪枝操作,不再进行扩展.③由于在构建子图搜索树时依次以每条频繁边作为根节点进行扩展,假设存在 2 条频繁边  $e_1$  和  $e_2$  ( $e_1 < e_2$ ),当以  $e_1$  作为根节点进行扩展时,已经将所有包含  $e_1$  的频繁子图全部搜索到,在后续以  $e_2$  作为根节点进行搜索时子图中不会出现  $e_1$  (可以证明,此时所有出现  $e_1$  的子图都不满足最小 DFS 序约束),因此在以某条频繁边作为根节点搜索结束后,应当将其从频繁边集中删除,从而达到剪枝的效果.

首先创建当前子树根节点 root 用于继续向下搜索和作为返回值(第 1 行);然后不断调用子图扩展算法 extend\_edge 对当前节点所含子图进行扩展并将结果保存(第 2~3 行);随后判断扩展得到的子图是否满足最小 DFS 序和频繁差异子图阈值  $\eta$  的约束,若不满足则不再对其进行扩展(第 4~5 行);接下来对满足约束的子图递归调用频繁差异子图挖掘算法 create\_tree 并作为当前节点的孩子节点(第 6 行).这样就实现了频繁差异子图挖掘的过程(见图 3).

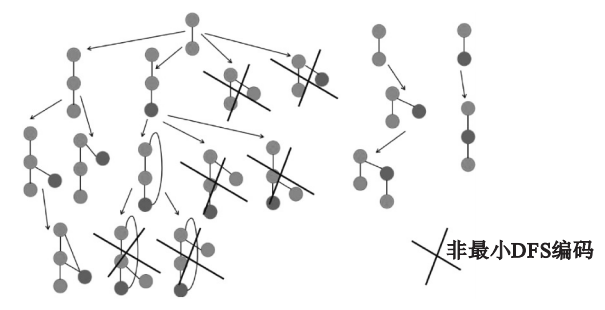


图 3 频繁差异子图挖掘过程  
Fig. 3 Mining process of frequent difference subgraph

### 1.3 基于频繁差异子图的频繁差异序列挖掘

频繁差异序列保留了动态脑网络的特征,利用差异子图序列反映一段时间内脑网络连接模式的变化.之前的工作得到了所有满足支持度要求的频繁差异子图,为了进一步研究频繁差异序列,需要进行回溯标记.

在回溯标记的过程中,构建了 match 数组,其中  $match[i][j]$  表示的是第  $i$  张差异图中是否出现了第  $j$  张差异频繁子图,若出现则赋值为 True,反之是 False.

算法 4 为频繁差异序列挖掘算法.首先创建深度优先搜索树的根节点(第 1 行);然后针对每一个样本全部脑网络差异图搜寻所存在的频率差异子图,若当前脑网络差异图中包含该频繁差异子图,则以其作为搜索起点调用 markCount 算法递归扩展其孩子节点(第 2~5 行);待所有测试数据完成匹配后返回搜索树根节点(第 6 行).这样就实现了频繁差异序列的挖掘,后期可以使用深度优先搜索算法根据实际情况筛选出搜索树中满足频繁差异序列阈值( $\zeta$ )要求的节点,具体操作为枚举所有从根节点出发到树上其他节点的路径,若当前节点计数值  $cnt \geq \zeta \times n$  ( $n$  为图集中脑网络数量),则将该路径加入到频繁差异序列挖掘结果中.

算法 4 频繁差异序列挖掘算法  
Input: 图集与频繁差异子图匹配数组  $match[m][n]$   
Output: 频繁差异子图搜索树

1. root = new node;
2. for  $i = 1$  to  $m$  do
3. for  $j = 1$  to  $n$  do
4. if  $match[i][j] = \text{True}$
5. root ← markCount( $i + 1, 1$ );
6. Return root.

算法 4 中采用的递归匹配算法具体过程见算法 5.首先判断若已将当前样本所有时序脑网络图匹配完成则结束递归并返回(第 1 行);然后若当前节点不存在则创建新节点,否则更新当前节点计数值,表示满足当前序列的样本数量加 1(第 2~3 行);随后遍历当前差异图中所有频繁差异子图,分别将其作为序列的新增元素并递归判断该样本下一张差异图(第 4~6 行);最后返回根节点信息(第 7 行).这样就完成了所有可能的频繁差异子图序列挖掘,按照阈值进行筛选即可得到所需频繁差异序列.

算法 5 递归匹配算法  
Input: 图集与频繁差异子图匹配数组  $match[m]$

[ $n$ ],图集中起始位置  $i$ ,递归层数  $c$

Output:频繁差异子图搜索树

1. if check( $i$ ) = False Return;
2. if root = null root = new node;
3. else if root! = null root←root + 1;
4. for  $k = 1$  to  $n$  do
5. if match[ $i$ ][ $k$ ] = True
6. root←markCount( $i + 1, c + 1$ );
7. Return root.

1.4 频繁差异序列的筛选

为了进行阿尔茨海默症的辅助诊断,需要找到患者或正常人所特有的频繁差异序列作为辅助诊断的标记物,这就要求将双方共有的频繁差异序列剔除.将频繁差异序列定义为一个  $n$  元组 [ $x_1, \cdots, x_n$ ],其中  $x_i (i \in n)$  代表一个频繁差异子图的编号,  $x_i, x_j (i \in n \cap j \in n \cap i < j)$  之间存在着  $x_i$  所代表的频繁差异子图出现的时刻早于  $x_j$  所代表的频繁差异子图的关系.

通过筛选出不具有区分性的频繁差异序列后,能够得到来自于 AD 患者和正常对照组 (normal control, NC) 的频繁差异序列集合  $S$  (AD 序列集) 和  $Q$  (NC 序列集),其中  $a (a = [x_1, \cdots, x_n]) \in S$ , 代表任一 AD 患者序列集中序列元素,  $b (b = [y_1, \cdots, y_m]) \in Q$ , 代表任一 NC 序列集中序列元素.

如算法 6 所示,频繁差异序列的筛选即需要排除掉出现在  $S, Q$  两个集合中的序列,还要设定区分阈值  $h$ ,来满足具有较明显区分性的要求.

算法 6 频繁差异序列筛选算法

Input: AD 患者和 NC 组的频繁序列集合  $S, Q, t, s$  分别代表  $S, Q$  集中序列个数,  $p, q$  分别代表序列  $i$  在 AD 和 NC 中出现的次数,  $\lambda_i, \mu_i$  分别是赋予序列  $i$  的两个权重.

Output: AD 和 NC 的频繁差异序列集合  $S, Q$

1. sort( $S$ ); sort( $Q$ );
2. point1 = 1; point2 = 1;
3. While point1 ≤  $t$  and point2 ≤  $s$
4. if  $S[\text{point1}] = Q[\text{point2}]$
5.  $S = \text{del}(S[\text{point1}])$ ;
6.  $Q = \text{del}(Q[\text{point2}])$ ;
7. else if  $S[\text{point1}] < Q[\text{point2}]$
8. point1←point1 + 1;
9. else
10. point2←point2 + 1;
11.  $S = S + Q$ ;
12. for  $i$  in  $S$

13. [ $p, q$ ] = get\_frequency( $i$ );
14. for  $i$  in  $S$
15. if  $|p_i \times \lambda_i - q_i \times \mu_i| < h$
16.  $S = \text{del}(i)$ ;
17. [ $S, Q$ ] = split( $S$ );
18. Return( $S, Q$ ).

将得到的频繁差异序列先进行排序(第 1 行),再进行去除出现在 AD 序列集和 NC 序列集中的公共部分(第 2 ~ 10 行).之后进行赋权优化区分序列的操作,使得到的频繁差异序列拥有更好的区分性(第 11 ~ 16 行).然后将总集中 AD 的区分序列和正常组的区分序列分离,得到具有区分性的检测标记物(第 17 ~ 18 行).

1.5 基于时序区分子图的阿尔茨海默症诊断

在完成频繁差异序列的筛选之后,就能得到时序区分子图,也就是阿尔茨海默症辅助诊断的生物学术标记物.由于样本分为 AD 和 NC,因此设置  $S = \{S_1, S_2, \cdots, S_n\}, Q = \{Q_1, Q_2, \cdots, Q_n\}$  分别为 AD 和 NC 的区分序列.在辅助检验之前需要为所有区分序列设置权值,其中  $W_{S_i} = \frac{\max(p_{S_i}, q_{S_i})}{p_{S_i} + q_{S_i}}$  ( $p_{S_i}$  表示  $S_i$  在 AD 中出现的次数,  $q_{S_i}$  表示  $S_i$  在 NC 中出现的次数) 表示赋给每一个  $S$  中序列的权重,  $W_{Q_i} = \frac{\max(p_{Q_i}, q_{Q_i})}{p_{Q_i} + q_{Q_i}}$  ( $p_{Q_i}$  表示  $Q_i$  在 AD 中出现的次数,  $q_{Q_i}$  表示  $Q_i$  在 NC 中出现的次数) 表示赋给每一个  $Q$  中序列的权重.

现有区分序列集合  $X = \{X_1, X_2, \cdots, X_m\}$ , 对于任一测试样例,需要进行与  $X$  的比对,为其 AD 检测赋权重  $W_{AD} = \{ \sum W_{s_j} | X_i \in S, X_i = S_j \}$ , 为其 NC 检测赋权重  $W_{NC} = \{ \sum W_{q_j} | X_i \in Q, X_i = Q_j \}$ , 最后判断,若  $W_{AD} > W_{NC}$  则预测为 AD 患者;若  $W_{AD} < W_{NC}$  则预测为正常.

2 实验结果与讨论

使用 MATLAB R2021a 进行实验中部分代码编写,其中包括实验分析图的绘制,功能性磁共振图像的预处理、AAL 模板匹配.其余部分使用 Visual Studio 2019 编译器、C++ 语言完成,其中包括频繁差异子图的挖掘、频繁差异序列的挖掘、频繁差异序列的筛选、阿尔茨海默症辅助诊断.

2.1 实验数据集

实验数据来自 ADNI 公开数据集 (<http://adni.loni.usc.edu>) 的功能磁共振数据. ADNI 数据库是由研究员 Weiner 博士领导的一个非营利



性组织,其主要目标是测试 MRI、PET、其他生物标记物以及临床检验和神经心理学评估是否可以结合起来进行 MCI 和早期 AD 进展的测量. 本实验获取了 178 例静息态功能性磁共振数据,其中包括 89 例患有阿尔茨海默症的病人和 89 例正常

对照组,由于考虑了脑网络动态特征,选取了同样样本的 5 个 TR 时间内的扫描数据,以构成时序脑网络. 对每个数据样本的一个 TR 时间内截取 140 层,且全部样本采用 Philips3.0T 扫描仪获取,具体信息如表 1 所示.

表 1 ADNI 数据集详情  
Table 1 Details of ADNI data set

数据集	样本数(男/女)	年龄/岁	MMSE	切片厚度/mm	TE/ms	TR/s
AD	48/41	72.5 ± 15	21.24 ± 3.37	3.31	30	3.0
NC	46/43	75.1 ± 13	28.95 ± 1.56	3.31	30	3.0

2.2 参数影响分析

在提出的基于时序区分子图的诊断方法中,存在 4 个可变参数:频繁边阈值  $\theta$ 、频繁差异子图阈值  $\eta$ 、频繁差异序列阈值  $\zeta$ 、区分阈值  $h$ . 其中频繁边阈值  $\theta$  和频繁差异子图阈值  $\eta$  分别表示在进行频繁差异子图挖掘过程中可进行差异子图扩展的边满足的最小支持度,以及得到的频繁差异子图满足的最小支持度. 频繁差异序列阈值  $\zeta$  表示在进行频繁差异序列挖掘的过程中满足的最小支持度. 区分阈值  $h$  表示挖掘出来的频繁差异序列经过赋权成为区分序列所需要满足的最小支持度. 为了评价这些参数在 AD 的辅助诊断过程中

对分类性能的影响,分别依次选用不同的参数值来计算每个参数对分类精度的影响.

测试使用频繁边阈值  $\theta$  ( $\theta = 0.150, 0.152, 0.154, 0.156, 0.158, 0.160$ ) 和频繁差异序列阈值  $\zeta$  ( $\zeta = 3, 4, 5, 6, 7, 8, 9, 10$ ) 取不同值时频繁差异子图阈值  $\eta$  ( $\eta = 0.07, 0.08, 0.09, 0.10$ ) 对 AD 和 NC 分类结果的影响,在实验过程中分别变化 3 个参数  $\theta, \zeta$  及  $\eta$ , 在每一张图像中,保持  $\theta$  不变,观察  $\zeta$  和  $\eta$  对准确率的影响. 判断两组数据集中参数  $\eta$  对 AD 诊断的分类精度影响如图 4 所示. 由于区分阈值可通过权重的更改保持定值,这里  $h = 1$ .

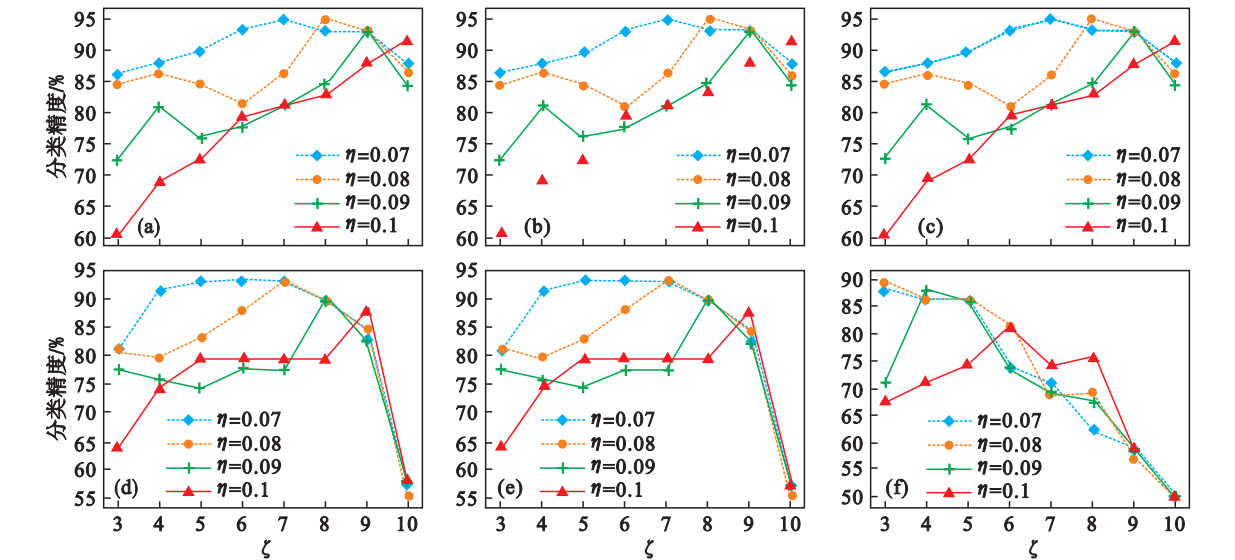


图 4 参数对分类精度的影响

Fig. 4 Influence of parameters on classification accuracy

(a)— $h=1, \theta=0.150$ ; (b)— $h=1, \theta=0.152$ ; (c)— $h=1, \theta=0.154$ ; (d)— $h=1, \theta=0.156$ ;  
(e)— $h=1, \theta=0.158$ ; (f)— $h=1, \theta=0.160$ .

由于频繁边阈值  $\theta$ 、频繁差异子图阈值  $\eta$ 、频繁差异序列阈值  $\zeta$  控制着最后得到的区分序列的数量,如果将阈值设定的相对小,会导致结果中存在区分性过小的区分序列,这反而会干扰高区分性的区分序列对预测结果的影响程度;如果阈值设定过大,会导致找出的区分序列过少,同样影响

分类精度. 由图 4 可见,当  $\zeta, h$  保持不变时,随着  $\theta$  的变化分类精度变化不明显,当  $\theta = 0.160$  时会出现在  $\zeta$  取值大于 7 的范围内区分性降低的问题. 当保持  $\theta, \zeta$  不变时,可以发现  $\eta = 0.07$  时分类精度效果最优,且在  $\zeta = 6$  或  $\zeta = 7$  时出现接近 95% 的分类精度,这说明在物理条件允许的前提

下,频繁差异子图越多,组合而成的频繁差异序列也就越多,便于后续依据频繁序列的差异性进行筛选,从而提高预测准确性. 由于受到数据测试集规模影响,相关参数的变化鲁棒性较低,在参数小浮动变化范围内,分类精度变化较为明显. 该问题可通过大数据集解决.

### 2.3 分类性能比较

为了测试基于时序区分子图方法在诊断阿尔茨海默症上的分类性能,本文选取基于静态区分子图<sup>[9]</sup>和动态脑网络分类器<sup>[15]</sup>方法进行对比. 基于同一数据集,所有方法全都进行了参数调优,最

终分类结果见表 2.

由表 2 可知,时序区分子图方法在区分 AD 患者和正常人的性能上远远优于其他几种对比方法. 基于时序区分子图的诊断方法最高的分类准确率为 94.8%,而与之对比的最好的方法为动态脑网络分类器方法,最高分类准确率为 82.1%. 实验结果表明所提出的基于时序区分子图的方法可以在提取脑网络生物学标志物上得到更好的标志物,同时时序脑网络信息也能更好地反映 AD 患者和正常人的差异,从而更好地进行 AD 患者的辅助诊断.

表 2 不同方法的分类性能比较  
Table 2 Comparison of classification performance of different methods

方法	准确度/%	敏感度/%	特异度/%
静态区分子图	77.8	79.25	71.6
动态脑网络分类器	82.1	79.9	70.8
时序区分子图	94.8	89.65	93.5

## 3 结 语

针对已有基于区分子图的阿尔茨海默症诊断方法多是建立在静态区分子图上的,忽略了脑网络的动态连接结构变化,提出了基于时序区分子图的阿尔茨海默症辅助诊断,它既能够保留脑网络的时序信息以及拓扑结构信息,还能够通过彰显差异性的方式,降低图挖掘的复杂度,进而充分利用数据信息建立模型. 通过大量实验证明,提出的时序区分子图方法能够更全面地保留大脑活动的信息,提高了对阿尔茨海默症患者辅助诊断的准确性.

### 参考文献:

[1] 李楠,许均瑜. 蛋白质组学对于阿尔茨海默症诊断标记物筛选的研究进展[J]. 基因组学与应用生物学,2016,35(5):1108-1114.  
(Li Nan, Xu Jun-yu. Proteomic screening for biomarkers of Alzheimer's disease[J]. *Genomics and Applied Biology*, 2016,35(5):1108-1114.)

[2] Jo T, Nho K, Saykin A J. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data [J]. *Frontiers in Aging Neuroscience*, 2019, 11: 1-14.

[3] Jie B, Zhang D, Wee C Y, et al. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification[J]. *Human Brain Mapping*, 2014, 35(7): 2876-2897.

[4] 汪新蕾,王之琼,王中阳,等. 面向阿尔茨海默病的脑网络多频融合图核[J]. 计算机学报,2020,43(1):64-77.  
(Wang Xin-lei, Wang Zhi-qiong, Wang Zhong-yang, et al. Multi-frequency fused graph kernel of brain network for Alzheimer's disease[J]. *Chinese Journal of Computers*,

2020,43(1):64-77.)

[5] Nguyen M, He T, An L J, et al. Predicting Alzheimer's disease progression using deep recurrent neural networks[J]. *NeuroImage*, 2020, 220(10): 117203.

[6] Mansu K, Jaesik K, Jeffrey Q, et al. Interpretable temporal graph neural network for prognostic prediction of Alzheimer's disease using longitudinal neuroimaging data [C]//BIBM 2021. Houston, 2021: 1381-1384.

[7] Cao B K, Zhan L, Kong X N, et al. Identification of discriminative subgraph patterns in fMRI brain networks in bipolar affective disorder [C]//International Conference on Brain Informatics and Health. London, 2015: 105-114.

[8] Kong X, Yu P S, Wang X, et al. Discriminative feature selection for uncertain graph classification [C]// SIAM International Conference on Data Mining. Austin, 2013: 82-93.

[9] Yan X, Han J. gSpan: graph-based substructure pattern mining [C]// 2002 IEEE International Conference on Data Mining. Maebashi: IEEE, 2002: 721-724.

[10] Mrzic A, Meysman P, Bittremieux W, et al. Grasping frequent subgraph mining for bioinformatics applications[J]. *BioData Mining*, 2018, 20(11): 181-189.

[11] Suresh A T, Jereesh A S. Associated subgraph mining in biological network [C]// 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA). Coimbatore, 2017: 80-85.

[12] Fischl B, Satal D H, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain [J]. *Neuron*, 2002, 33(3): 341-355.

[13] Tzouriomazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain [J]. *Neuroimage*, 2002, 15(1): 273-289.

[14] 卢思成. 基于动态网络的阿尔茨海默病计算机辅助诊断 [D]. 沈阳: 东北大学, 2019.  
(Lu Si-cheng. Computer-aided diagnosis of Alzheimer's disease based on dynamic network [D]. Shenyang: Northeastern University, 2019.)

[15] Benesty J, Chen J, Huang Y, et al. Pearson correlation coefficient [C]//Noise Reduction in Speech Processing. Berlin, Heidelberg: Springer, 2009: 1-4.