**诚信声明**

本人声明：

本人所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名： 日期：

**毕业设计(论文)任务书**

设计(论文)题目：基于小波变换和支持向量机的癫痫脑电信号分类

学院：信息科学与技术学院 专业：自动化实验班 班级：自实1601

学生：陈帅华 指导教师： 宿翀 专业负责人： XXX

1. 设计（论文）的主要任务及目标

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX(小4，宋体)。

2、设计（论文）的主要内容

（1）XXXXXXXXXXXXXXX；

（2）XXXXXXXXXXXXXXXXXX；

（3）撰写毕业论文；

（4）翻译一篇5000汉字左右的外文文献。

3．设计（论文）的主要要求

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX(小4，宋体)。

4．主要参考文献

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX(小4，宋体)。

5．进度安排

|  |  |  |
| --- | --- | --- |
|  | 设计（论文）各阶段名称 | 起止日期 |
| 1 | 调研及查阅文献、翻译外文资料、撰写文献综述、开题报告 | 年月日--月日 |
| 2 | 对其中一种构型进行三维建模，初步进行流场模拟 |  |
| 3 | 流场模拟计算的完善及结果分析 |  |
| 4 | 撰写毕业论文、答辩准备 |  |

**基于小波变换和支持向量机的癫痫脑电信号分类**

**摘要：**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX（宋体，小四号）

**关键词：**EEG，小波变换，支持向量机，K最近邻算法，决策树，集成学习

注：

1、**摘要：（黑体，小四号，加粗）**

2、**关键词：（黑体，四号，加粗）**

3、1.25倍行距，段前0.5行

**Mixing Performance of Different Configurations of New Kneading Disc in a Co-rotating Twin Screw Extruder（Times New Roman，三号，加粗）**

**ABSTRACT：**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX（Times New Roman，小四号）

**KEY WORDS:** XXXXXXXX，XXXXXXX，XXXXXX，XXXXXX（Times New Roman，四号）

注：

1、**ABSTRACT：（Times New Roman，小四号，加粗）**

2、**KEY WORDS:（Times New Roman，四号，加粗）**

3、1.25倍行距，段前0.5行

4、标题实词首字母大写，关键词中实词首字母大写。

目 录

[前 言 1](#_Toc40221441)

[第1章 绪论 2](#_Toc40221442)

[第1.1节 课题研究的背景与意义 2](#_Toc40221443)

[第1.2节 国内外研究现状 3](#_Toc40221444)

[1.2.1、癫痫脑电信号特征提取研究现状 5](#_Toc40221445)

[1.2.2、癫痫脑电信号分类算法研究现状 5](#_Toc40221446)

[第1.3节 本文的创新点与研究难点 5](#_Toc40221447)

[1.3.1、 本文的创新点 5](#_Toc40221448)

[1.3.1、 本文的研究难点 6](#_Toc40221449)

[第1.4节 本文的主要内容与安排 6](#_Toc40221450)

[第2章 脑电相关知识与癫痫检测 8](#_Toc40221451)

[第2.1节 脑电相关知识与癫痫检测 8](#_Toc40221452)

[第2.2节 脑电信号的采集 8](#_Toc40221453)

[第2.3节 癫痫脑电信号的特征 8](#_Toc40221454)

[第3章 基于离散小波变换的脑电特征提取 8](#_Toc40221455)

[第3.1节 脑电信号数据来源及介绍 9](#_Toc40221456)

[3.1.1、波恩大学癫痫脑电数据集 9](#_Toc40221457)

[3.1.2、CHB-MIT头皮脑电图数据库脑电数据集 10](#_Toc40221458)

[第3.2节 (癫痫)脑电信号预处理 12](#_Toc40221459)

[3.2.1、波恩大学癫痫脑电数据集 12](#_Toc40221460)

[3.2.2、CHB-MIT头皮脑电图数据库脑电数据集 12](#_Toc40221461)

[第3.3节 小波变换理论介绍 13](#_Toc40221462)

[第3.4节 由离散小波变换进行脑电特征提取 15](#_Toc40221463)

[第3.5节 结果与分析 24](#_Toc40221464)

[第4章 基于多种机器学习算法的癫痫脑电信号分类 24](#_Toc40221465)

[第4.1节 支持向量机 24](#_Toc40221466)

[第4.2节 K最近邻算法 25](#_Toc40221467)

[第4.3节 决策树 25](#_Toc40221468)

[第5章 集成学习 26](#_Toc40221469)

[第5.1节 支持集成学习理论介绍 26](#_Toc40221470)

[第5.2节 集成学习1 27](#_Toc40221471)

[第5.3节 集成学习2 28](#_Toc40221472)

[结 论 28](#_Toc40221473)

[参考文献 29](#_Toc40221474)

[致 谢 31](#_Toc40221475)

**注：**

1、每章标题字体（黑体，四号，加粗）

2、每节标题及三级小标题字体（宋体，四号）

3、前言在第1章前。结论、参考文献、致谢不单独成章。

前 言

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX（宋体，小四，固定值22磅，首行缩进2字符）

第1章 绪论

第1.1节 课题研究的背景与意义

人类大脑中有上百亿个脑细胞，神经元与神经胶质细胞又共同组成了脑细胞，人类自身的一切活动过程正是靠大脑进行主导和调节的，以保证人类机体和周围环境的平衡，而这一主导和调节过程又正是依靠人类大脑内由大量神经元互相连接进而构成的极其庞大的神经网络实现的。这一种神经网络结构会在神经冲动出现的时候，使众多神经元被激活开始工作，进而传导电信号，故而大脑是靠通过发送传导电信号进行活动，控制神经的。

当人体处于正常的生理状况时，这一电信号的产生和传导过程是有迹可循的、有限的。但是当人体受到遗传因素的影响或者出现脑结构异常和脑功能的损伤时，可能会造成人体大脑的异常放电，这种放电形式比正常生理状况下的电信号传导距离要长，影响范围要广，同时也不受控制，可能还会造成短暂的大脑功能障碍。而这种异常的放电形式往往会诱发癫痫的产生。

癫痫也即民间常叫的“羊癫疯”，这种疾病发作的时候往往会对人体造成很大的伤害，尤其是对大脑造成的损伤，比如会造成人体大脑记忆力的下降和认知能力的下降，而如果该疾病频繁地发作，甚至会造成对大脑永久性的伤害。由世界卫生组织于2018年发布的报道可知，在全球范围内，罹患癫痫疾病的人数已经超过了5000万，而其中绝大一部分都出现在了中低收入国家，而在我们国家中，有900万以上的国人正深深地受到该疾病的困扰，此外癫痫病例增长的速度也十分的可怕，每年都会有60万例的增长[1]。而罹患癫痫之后，不仅患者的整个家庭要承受高昂的治疗费用，还可能会对患者的心理健康和未来的发展造成十分不利的影响，比如患者可能会遭受到社会隔离甚至是污名化，也可能会造成患者较低的教育成绩与糟糕的就业效果。故而针对癫痫的预防和鉴别诊断具有及其重要社会意义和现实意义。

但在实际的诊断过程中，诊断癫痫可能会存在很大的困难，出现这种情况的原因主要是由于有不少其他的疾病可能会展示出同癫痫疾病发作时十分类似的表现和病征，例如昏厥、换气过度综合征（Hyperventilation syndrome，HVS）、偏头疼和发作性睡病等等之类的疾病。目前癫痫的主要检查手段是脑电图（Electroencephalogram，EEG），该检查手段主要是在检查测试的时候，将精密的传感器通过黏黏的导电胶安装在人体的头皮之上，以获取人体大脑细胞互相发送信息的时候所产生的电信号。通过脑电图仪所记录的脑电信号，大脑的活动可以被清晰地呈现出来，若在脑电图监测期间出现癫痫发作的情况，那么脑电图上就会可以看到正常的大脑活动模式发生了改变，并可以看到癫痫发作过程中所独有的特征波。

但是通常由于脑电图本身是非常复杂的，人类的脑电波也是多种多样，而在临床上，主要靠医生在几分钟到几个小时之内测量脑电波，进而找到大脑之中异常脑电活动的依据，并且找出癫痫患者罹患的癫痫发作类型以及起源。但是这种方式十分依赖于医务人员自身的诊断经验和主观标准，进而可能会出现对患者的误诊断和漏诊断的情况，此外长时间采用这种方式会对医务人员的精力造成极大的消耗而且效率也会比较低，这对于越来越多的癫痫患者来说却是远远不够，因而寻找对癫痫脑电信号的识别、特征提取与分类方法对社会，对患者家庭和患者自身而言都有着极其重要的意义！

第1.2节 国内外研究现状

脑电信号具有不平稳和非线性的特性，是一种随着科学技术的进近年来，癫痫脑电信号检测与分类技术迅速发展，支持向量机、隐马尔科夫链、极限学习机、随机森林、深度学习和人工神经网络等方法也被广泛应用**错误!未找到引用源。错误!未找到引用源。错误!未找到引用源。错误!未找到引用源。**。

对于脑电信号特征的提取，针对癫痫脑电和正常脑电的不同特点，国内外研究人员对于脑电信号特征的提取也提出了很多方法。王海玉等人对其进行了总结，例如在时域上的方差分析、直方图分析等，在频域上的频谱分析、全息谱分析等方法**错误!未找到引用源。**，但无论是时频分析还是频域分析，都未充分体现EEG信号的不平稳和非线性的特性。李小兵等人指出时频分析方法充分考虑了信号的不平稳特性，在时间和频率上均有很好的分辨率**错误!未找到引用源。**。但由于人脑由大量神经元构成，组成了一个复杂的非线性系统，因此非线性分析方法也广泛用于脑电的分析**错误!未找到引用源。**，王海玉等人也指出Lyapunov指数、复杂度、样本熵、近似熵等都是目前研究脑电主要的非线性方法**错误!未找到引用源。**。

韩敏等人利用小波变换(WT)计算出每个频带小波系数的标准差、平均功率等指标，并将其作为脑电信号的特征，最后在对正常脑电信号和癫痫脑电信号的分类过程中取得了98.38%的分类精度**错误!未找到引用源。**。此论文最后得到了较高的分类精度，但是脑电信号特征维数较大，达到了22维；且分类过程中消耗的时间也相对更多。Yatindra Kumar等人将脑电信号进行离散小波分解得到各个频带的小波系数后求取了相对小波能量和小波熵作为脑电信号的特征，由于小波熵能衡量信号的有序度和无序度，反映非线性信号潜在的动力学特性，因此对于区分癫痫脑电信号有着很好的效果**错误!未找到引用源。**。Cheol Seung Yoo等人发现在癫痫发作过程中脑电信号的样本熵值下降，在癫痫发作终止时达到最小值，癫痫发作终止后，样本熵迅速恢复到癫痫发作前的数值**错误!未找到引用源。**。Julius Georgiou等人将具有低计算复杂度的排列熵作为自动癫痫检测特征进行了可行性研究，最终对正常脑电和癫痫脑电信号取得了93.55%的最佳平均辨别率**错误!未找到引用源。**。

在对脑电信号进行特征提取后，由分类器根据特征向量进行分类。支持向量机作为一种监督学习算法，在分类和回归问题上被广泛地应用**错误!未找到引用源。错误!未找到引用源。错误!未找到引用源。错误!未找到引用源。**。肖文卿等人将支持向量机用于对正常小鼠和癫痫小鼠分类，并选取径向基核函数，在融合了多种特征向量后，得到了99.80%的分类准确率**错误!未找到引用源。**。吕健和吴艳在各自的研究中分别采用网格搜索等方法对支持向量机中的核函数参数以及惩罚因子进行了优化**错误!未找到引用源。错误!未找到引用源。**，最后的结果也证明了其较好的性能及泛化能力。

而Shahab Abdulla等人在他们的研究中除了使用支持向量机外，也采用了K最邻近算法(KNN)，该算法是机器学习算法中最为简单的一种，其关键在于对k的选取上，Shahab Abdulla等人将k定为1，在此研究中，支持向量机的性能优于KNN**错误!未找到引用源。**，这很可能是由于其k选取的过小，导致整体模型变得复杂，发生了过拟合。在Manjusha等人的研究中，为了解决KNN在一个大的训练集中花费较长时间寻找最近邻域的问题，他们采取了降维方法，但在最后对癫痫危险度分类中，使用KNN得到的准确度，灵敏度，特异性等指标较低，有待提高**错误!未找到引用源。**。

人工神经网络是基于人类大脑神经元结构和功能活动创立的数学模型，它以自身独有的非线性自适应信息处理能力被广泛应用，Yatindra Kumar等人在对癫痫脑电信号的分类中使用了BP神经网络，其结构为：输入层4个神经元，隐藏层6个神经元，输出层2个神经元，在对正常脑电和癫痫脑电的分类中取得了95%的分类准确率**错误!未找到引用源。**，但是该方法在处理大容量癫痫样本数据时，收敛速度会极为缓慢，对神经网络计算能力要求也十分苛刻，目前只能在最佳性能与计算复杂度间折衷考虑。

1.2.1、癫痫脑电信号特征提取研究现状

分为两部分，

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX [5]。（宋体，小四，固定值22磅，首行缩进2字符）

XXXXX[9-12]。

1.2.2、癫痫脑电信号分类算法研究现状

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX [12,16,18,20]。

第1.3节 本文的创新点与研究难点

* + 1. 本文的创新点

(1)采用两个数据集进行了研究验证

在癫痫脑电信号的特征提取与分类这一领域的过往研究之中，有很大一部分的研究都是针对波恩大学癫痫脑电数据集这一单一数据集进行的。但是考虑到如果仅仅针对该数据集进行信号的预处理、特征提取和信号的分类，那么最终得到的结论本身就不具备普遍性。本文考虑到这一点后，除去下载的波恩大学癫痫脑电数据集之外，又下载了波士顿儿童医院的CHB-MIT头皮脑电图数据库中的一部分数据集进行分析研究。本文通过对这两个数据集的研究进而使得最后得到的结论具备了一定的普遍性，对研究癫痫脑电信号的分类有一定的借鉴意义和实际应用意义。

(2)将集成学习方法应用于癫痫脑电信号分类，提高分类效果

本文将集成学习的方法应用在癫痫脑电信号分类这一领域之中，通过使用集成学习中的stacking方法，本文将若干个单一分类算法作为基学习器使用数据集进行训练，之后将通过基学习器得到的分类结果重新作为新的训练集来训练另一个新的分类算法，该分类算法得到的结果就作为最终的分类结果。

通过使用该算法，可以充分利用多个分类器进而得到更好的分类效果，同时提高了整体的泛化能力，很大程度上弥补了前人研究中使用单独使用某一模型时所带来的缺点。

* + 1. 本文的研究难点

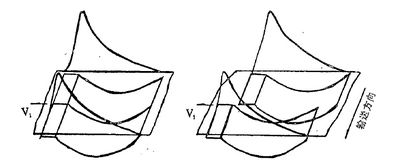
1. 对CHB-MIT头皮脑电图数据库数据的读取和预处理

由于该数据库中的数据均为edf格式，而该格式的数据文件不像txt或者csv文件那样可以直接进行读取，而是要在找到合适的api之后，然后调用再进行读取。针对该问题，经过一番调查、研究与学习之后，发现eeglab工具箱功能十分强大，它可以对数据进行读取、可视化、滤波、独立成分分析和消除伪影等等一系列操作，因此本文决定下载eeglab工具箱，在matlab环境之下，对edf格式的数据进行读取和预处理操作。

第1.4节 本文的主要内容与安排

论文中图片表示方式：

如图1·1所示为捏合盘元件的输送特性图。



(a) 单个捏合盘 (b) 一组捏合盘

图1·1 捏合盘元件输送特性（宋体，五号）

论文中公式表示方式：

 （1·1）

 （1·2）

（1·2）表示第1章中第2个公式

**论文中表格表示方式**：

表1·1 四种不同捏合块构型下流道的物理参数（宋体，五号，表格内文字皆用五号）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 捏合块  元件 | 导程/ | 螺杆外径/ | 机筒内径/ | 捏合盘厚度/ | 错列角/° | 中心距/ | 计算域长度/ |
| NKBLB |  |  |  |  |  |  |  |
| NKBRB |  |  |  |  |  |  |  |
| NKBLF |  |  |  |  |  |  |  |

续表1.1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NKBRF |  |  |  |  |  |  |  |
| NKBRF |  |  |  |  |  |  |  |

第2章 脑电相关知识与癫痫检测

第2.1节 脑电相关知识与癫痫检测

第2.2节 脑电信号的采集

第2.3节 癫痫脑电信号的特征

第3章 基于离散小波变换的脑电特征提取

第3.1节 脑电信号数据来源及介绍

本研究中所采用的EEG脑电数据集共有两个，分别是来源于波恩大学癫痫脑电数据库的数据集和来源于波士顿儿童医院的CHB-MIT头皮脑电图数据库，这两个数据集均是开源的，并且均可以直接在互联网上下载到，这两个数据集的下载地址分别为：<http://qcsdn.com/q/a/3118833.html>和<http://physionet.mit.edu/physiobank/database/chbmit/>。

3.1.1、波恩大学癫痫脑电数据集

对于来自于波恩大学癫痫脑电数据库的数据集，其一共由五组数据构成，分别用英文字母A～E表示。这五组数据均为从志愿者身上采取得到的，而采集方式则是使用上文中提及的国际标准的10-20系统来安置电极，对志愿者的脑电数据进行记录。这五组数据可以分为两部分：从五位健康志愿者身上采集到的两组数据(包括A、B两组)和从五位患有癫痫的志愿者身上采集到的三组数据(包括C、D和E三组)。而这两部分又分别是志愿者处于不同的状态下的时候，使用脑电图仪在志愿者大脑的不同部位记录采集到的，具体介绍详见下方表3·1所示。在A～E五组数据中，每一组数据中又分别均有100条单通道脑电记录，格式则为txt，其采样频率为173.61Hz，每一条脑电记录持续的时间则为23.6s，共有4097个数据点。值得注意的是，这些数据已经提前经过了带通滤波处理(其中滤波通带设置为0.53Hz~40Hz)，也已经经由专业人员进行人工伪迹的去除。

表3·1 波恩大学癫痫脑电数据集概述

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 组别 | A | B | C | D | E |
| 志愿者  身体状况 | 健康 | 健康 | 癫痫 | 癫痫 | 癫痫 |
| 志愿者状态 | 睁眼 | 闭眼 | 发作间期 | 发作间期 | 发作期 |
| 脑电信号  采集部位 | 大脑  表面 | 大脑  表面 | 与病灶相对脑半球的海马区 | 病灶区 | 病灶区 |

在matlab2018开发环境之下，针对A~E这五组数据中每一组的第一条脑电记录，进行脑电数据的读取与图像的绘制，读取之后得到的脑电信号时序图如图3·1所示。图3·1极其明确地展示了各组志愿者的脑电信号波形是怎么样的，通过观察这一张图可以发现A组和B组这两组的波形相似度较高，而C、D和E这三组则各不相同，和正常志愿者相比，可以发现C、D和E这三组的脑电信号时序图中存在着许多异常的波形，与正常人的脑电信号差异较大，这也为后续的脑电信号研究分类工作奠定了成功的基础。

通过研究图3·1之后，本研究将对癫痫脑电信号的分类工作共分为四种情况，其中第一种情况是针对A/E两组进行二分类，第二种情况是针对B/E两组进行二分类，第三种情况是针对C/E两组进行二分类，第四种情况则是对D/E两组进行二分类。

具体如表3·2所示。

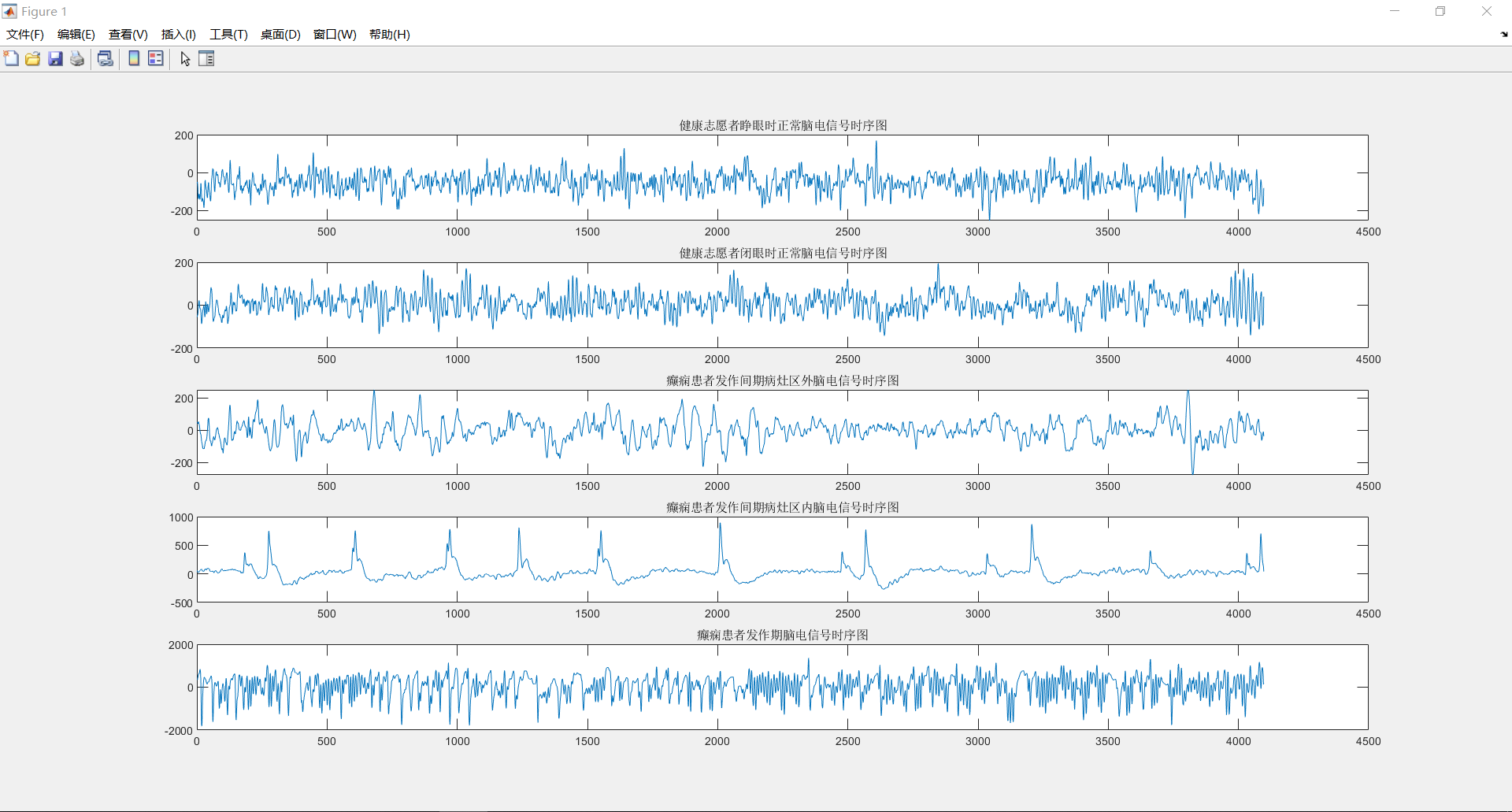


图3·1 A~E五组志愿者脑电信号时序图

表3·2 波恩大学癫痫脑电数据集四种分类任务

|  |  |  |
| --- | --- | --- |
| 情况 | 志愿者临床状态 | EEG子集 |
| 1 | 正常 vs 发作期 | A vs E |
| 2 | 正常 vs 发作期 | B vs E |
| 3 | 发作间期 vs 发作期 | C vs E |
| 4 | 发作间期 vs 发作期 | D vs E |

3.1.2、CHB-MIT头皮脑电图数据库脑电数据集

对于来源于波士顿儿童医院的CHB-MIT头皮脑电图数据库的EEG脑电数据集，其则是由儿科患者的脑电图记录所构成，这些儿科患者均罹患有难治疗性癫痫。据了解，该数据库中共包含有22位受试者的23个病例记录(chb01~chb24，其中病例chb01在1年半之后重新进行了脑电图记录，并记为了chb21；而病例chb24则是在2010年12月份的时候添加到了该数据库当中，该病例的性别和年龄并未记录在其中，而是仅仅记录了该病例的癫痫发作情况)，该22位受试者则包括5例男性和17例女性，其年龄分别为3岁到22岁和1岁半到19岁。

该数据库中的数据和波恩大学癫痫脑电数据库的数据集一样，均根据国际标准的10-20系统来安置电极，对患者的脑电数据进行记录，这些所有采集到的数据均为以256Hz的采样率以16位分辨率进行采样。其中每一个案例均包含有若干个.edf数据文件(从最少的9个到最多的42个不等)，其中EDF(European Data Format)则为一种在生理信号(如脑电信号和心电信号等)领域中常用的一种标准数据格式，它可以对多通道数据进行存储，并且能够使得每一个通道的信号有不一样的采样频率。在大多数情况之下，每个.edf文件均正好包含有病患长达一个小时的数字化EEG信号，只有对少数的数据文件，其才会记录四个小时的EEG脑电信号，例如病例chb01有chb01\_01.edf~chb01\_46.edf共46个数据文件，每一个数据文件中均记录了病例chb01一个小时的EEG信号。这些数据文件中，大多数有23个脑电信号通道，只有少部分有24个或者26个通道。

本研究在matlab2018环境下使用eeglab工具箱对病例chb01下的数据文件chb01\_03进行了读取并对脑电图像进行了绘制,具体如图3·2所示。

在本研究中，采用该数据集的目的是为了对针对波恩大学癫痫脑电数据库的数据集提出的癫痫脑电信号分类方法进行验证以得到具有普遍性的结论，因此在本研究中仅选取部分病例的部分发作间期和发作期的脑电数据进行研究。经过研究，本研究选取病例chb01和chb03的FT9-FT10通道，并选取若干个记录有癫痫发作的数据文件中的发作间期和发作期各一部分时间(每次记录起始时间均为0s)来进行研究，详细情况如表3·3和表3·4所示(以下选取的数据文件，数据记录时间均为1个小时)。

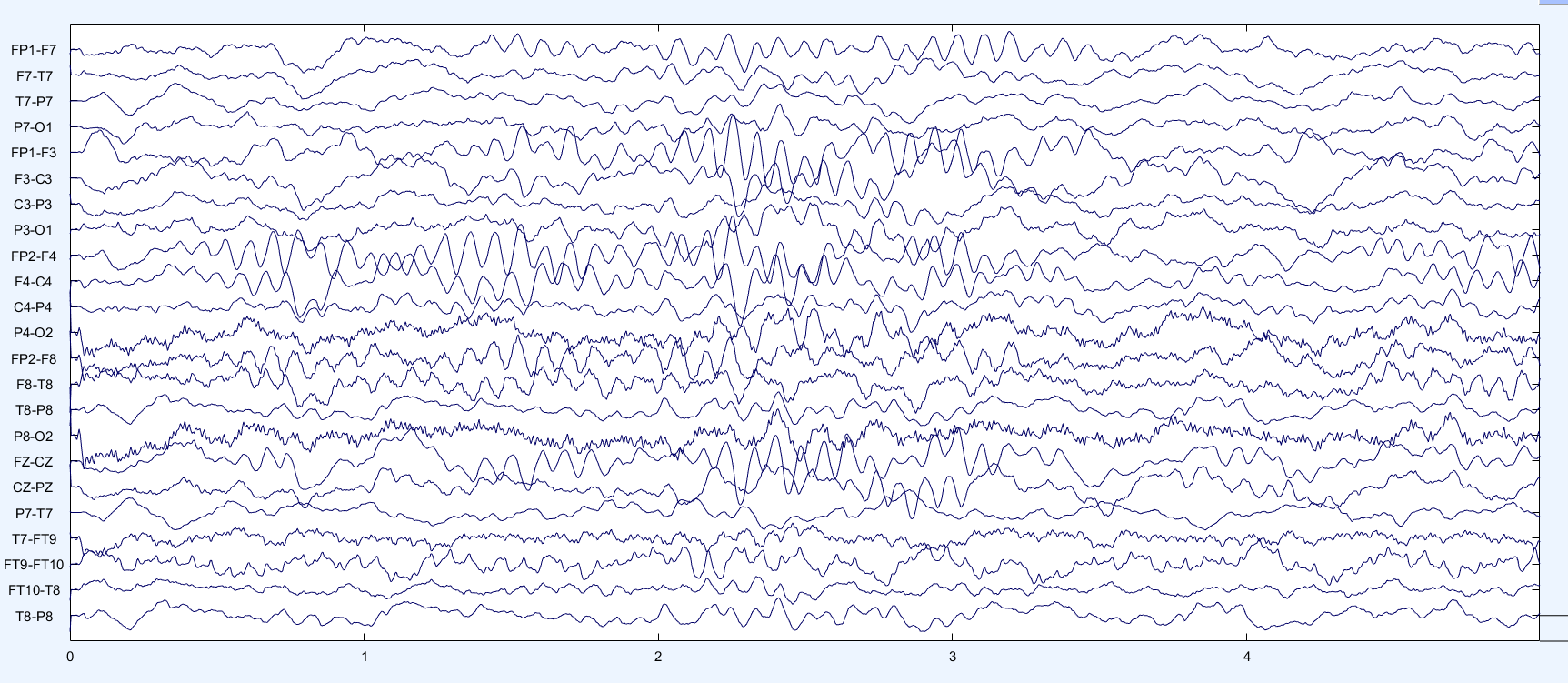


图3·2 chb01\_03脑电信号时序图

表3·3 病例chb01选取的数据文件及时间

|  |  |  |
| --- | --- | --- |
| 病例数据文件 | 选取发作间期时间(s) | 选择发作期时间(s) |
| chb01\_03 | 100~130 | 3001~3031 |
| chb01\_15 | 1000~1030 | 1737~1767 |
| chb01\_16 | 500~530 | 1025~1055 |
| chb01\_18 | 200~230 | 1750~1780 |
| chb01\_21 | 500~560 | 342~402 |
| chb01\_26 | 100~160 | 1882~1942 |
| 共计 | 240 | 240 |

表3·4 病例chb03选取的数据文件及时间

|  |  |  |
| --- | --- | --- |
| 病例数据文件 | 选取发作间期时间(s) | 选择发作期时间(s) |
| chb03\_01 | 2000~2040 | 368~408 |
| chb03\_02 | 150~180 | 748~778 |
| chb03\_03 | 2000~2030 | 452~482 |
| chb03\_04 | 300~340 | 2168~2208 |
| chb03\_34 | 520~560 | 1985~2025 |
| chb03\_35 | 1000~1030 | 2608~2638 |
| chb03\_36 | 100~130 | 1736~1766 |
| 共计 | 240 | 240 |

第3.2节 (癫痫)脑电信号预处理

3.2.1、波恩大学癫痫脑电数据集

对于该数据集，由于该数据集已经经过了通带为0.53Hz~40Hz的带通滤波和由专业人员进行的人工伪迹处理，因此本研究中不再进行额外的数据操作。

对于A~E五组数据中每一组中的每一条单通道脑电记录，本文按照时间顺序，以1024个数据点为一小组，那么一条单通道脑电记录就可以划分为四个小组，A~E这五组脑电数据则总共可以划分为2000个小组。在接下来的分析研究中，本文以一小组为单位进行分析操作。

3.2.2、CHB-MIT头皮脑电图数据库脑电数据集

对于该数据集，由于其还未经过任何处理，因此在通过使用eeglab工具箱将选中的病例chb01和chb03的数据文件读取之时，直接按照表3·3和表3·4中所选定的时间范围进行数据的读取。在进行了数据读取操作之后，本研究通过eeglab工具箱自带的数字滤波器对这些数据文件进行通带为0.1Hz~60Hz的带通滤波操作。通过这一操作，本研究达到了对所选数据文件滤除60Hz工频干扰的目的。

在下一步中，本研究直接将除了FT9-FT10这一通道外的其他通道清除，经过这两步的操作之后可以得到若干个期望的格式为edf的单通道数据文件，这些数据文件所包含的时间范围同样是表3·3和表3·4中需要的。

之后本研究以chb01\_03的发作间期和发作期的数据为例，通过使用EDFbrowser软件对这两部分的数据进行了读取，读取得到的脑电信号时序图分别如图3·3和图3·4所示。通过观察，显然可以发现这两张时序图的特征波和脑电信号幅值存在着较大的差异，这与波恩大学癫痫脑电数据库的数据集经由matlab读取后得到的D和E两组的癫痫脑电信号时序图有异曲同工之处。

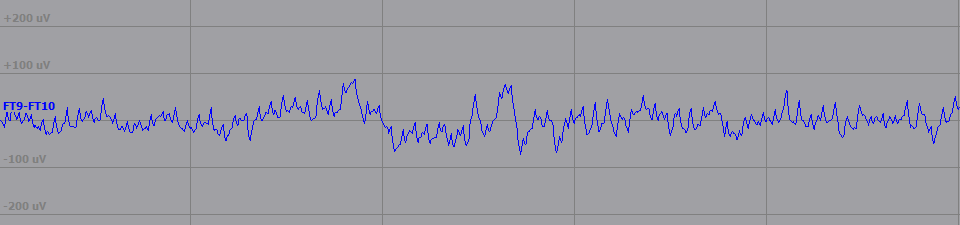


图3·3 chb01\_03发作间期脑电信号时序图

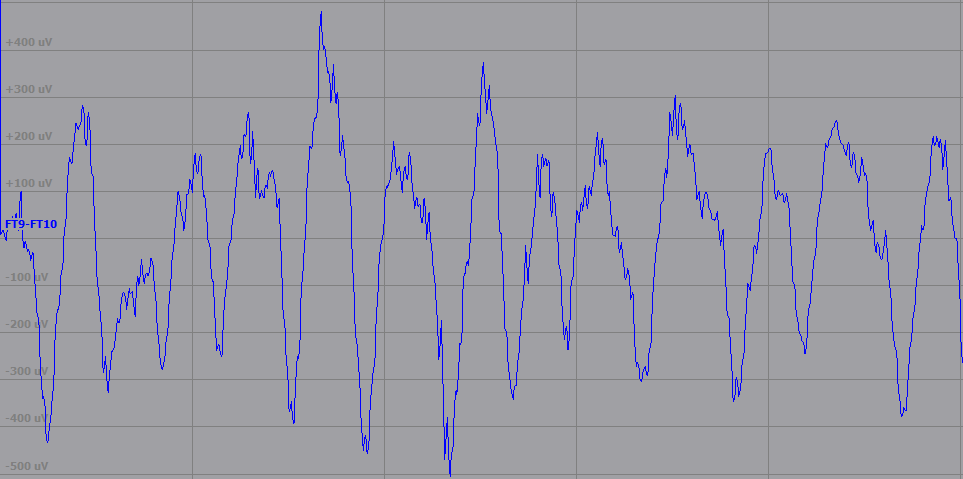


图3·4 chb01\_\_03发作期脑电信号时序图

由于matlab或者python中处理数据时通常用到的数据格式是txt格式或者是csv格式而并非本文上文中提及的edf格式，因此还需要对数据文件的数据格式进行转换。本研究中通过使用EDFbrowser软件，实现了将数据由edf格式到txt格式的转换，这样就可以直接和波恩大学癫痫脑电数据库的数据集进行相同的操作，十分方便。

在将这些数据文件的格式由edf转换为txt之后，对于病例chb01(病例chb03同理，也进行相应的操作)，本研究在其发作间期和发作期均各随机挑选了240s的时间，共计480s，不同于将波恩大学癫痫脑电数据集以1024个数据点分为一个小组，此处以256个数据点即1s的时间为单位进行分组，因此病例chb01的脑电数据共可以分为480个小组，在接下来的研究分析中，均以小组为单位进行操作。

第3.3节 小波变换理论介绍

小波变换与傅里叶变换的关系

小波变换(wavelet transform)是由傅里叶变换(Fourier Transform)发展改良而来的。根据以往学习过的知识， 我们可以知道傅里叶变换是以一个无限长的三角函数作为它的基函数，进而把一个在时域上不属于周期信号的一个连续信号转换为一个在频域上非周期的一个连续的信号，其中傅里叶变换的公式如所示。但是对于像生物医学信号(例如本文中主要讨论的脑电信号)这种自然界大量存在着的有着非平稳、非线性特性的信号，傅里叶变换处理的时候就会出现无法忽视的缺陷，例如对于一段未知的信号，傅里叶变换只可以获取到该信号大体上是由哪一些频率成分构成的，然而傅里叶变换对于这一段未知信号中的各个组成成分出现的时间并不能确定，故而可能会出现这样一个现象，两个信号的频谱图可能十分相似，但是这两个信号在时域上却可能相差甚远。



因此之后就出现了对傅里叶变换进行改进的方法-短时傅里叶变换(Short-time Fourier Transform，STFT)。短时傅里叶变换对于傅里叶变换改进之处在于“加窗”，对于一个时域信号，短时傅里叶变换将其分解为大量的小过程，而这些小过程同时又是等长度的，由于对该时域信号进行了大量的分解，因此可以近似认为每一个小过程都是平稳的，接下来再将该时域信号通过加的窗分为一段一段的子部分，针对这些分好段的时域信号再进行傅里叶变换，通过这种方法就可以明确频率成分随时间变换的情况，其中短时傅里叶变换的公式如所示，其中即为窗口函数。但是对于窗口函数而言，一旦其选定之后，窗口的形状便被固定，切断了窗口框的宽度与频率之间的联系；此外施加的小窗在太宽或者是太窄的情况下都会出现不同的问题，例如当所加窗的框太窄时，自然而然便会导致信号过少，致使得到的频域信号频率分辨率差；而当所加窗的框太宽的时候便会导致时域信号的时间分辨率很低。



而小波变换则是克服了傅里叶变换和短时傅里叶变换的缺点，将傅里叶变换中的基函数(一个无限长的三角函数)替换为了一个有限长度会衰减的小波基函数，该小波基函数可以对窗口进行调变，例如当频率增高的时候，时间窗口的宽度就会相应地变窄，进而达到提高频率分辨率的目的；而当频率降低的时候，时间窗口的宽度就会相应地便宽，已达到提高时间分辨率的目的。因此可以看出小波变换对于非稳定信号的时频分析处理有着极大的优势。

小波变换包括连续小波变换(Continuous Wavelet Transform，CWT)和离散小波变换(Discrete Wavelet Transform，DWT)

CWT公式，其中称之为小波母函数



DWT

框图 离散小波的多层分解过程



第3.4节 由离散小波变换进行脑电特征提取

小波基的选择

上文中已经对离散小波变换进行了详细的介绍，小波变换就是将一个信号分解成为一系列的小波，而其中非常重要的一个步骤就是对小波基的选取。

但是小波基的种类十分繁多，比如常见的Symlet(symN)小波、Daubechies(dbN)小波和Coiflet(coifN)小波等，本研究对这些小波中若干个典型的小波基进行了选取，其图像分别如图3·5、图3·6和图3·7所示。当对小波基进行不同的选择之后，在对一段相同的信号进行分析处理的时候，最终得到的处理效果也可能迥然不同。

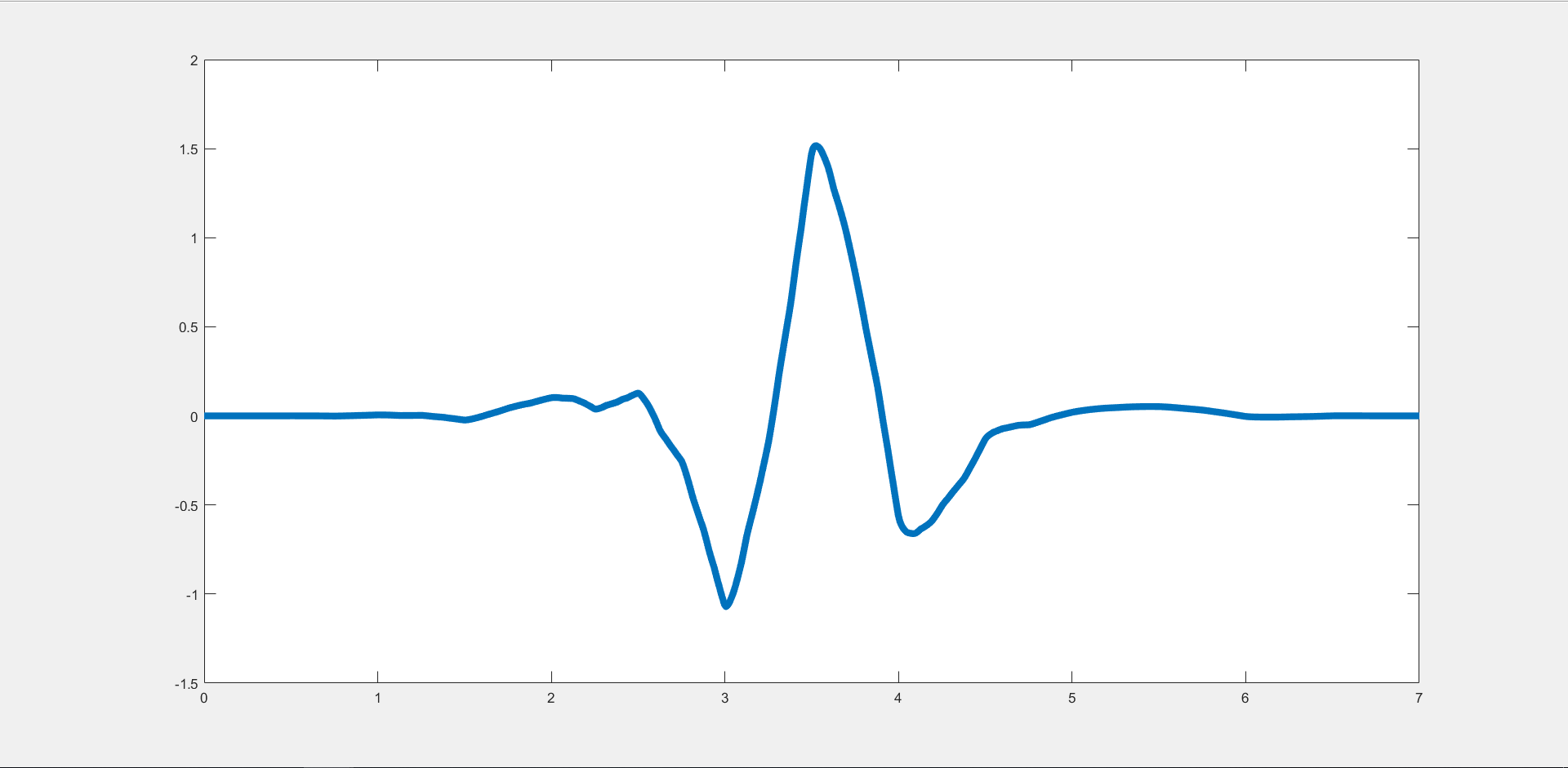


图3·5 Sym4小波基函数图像

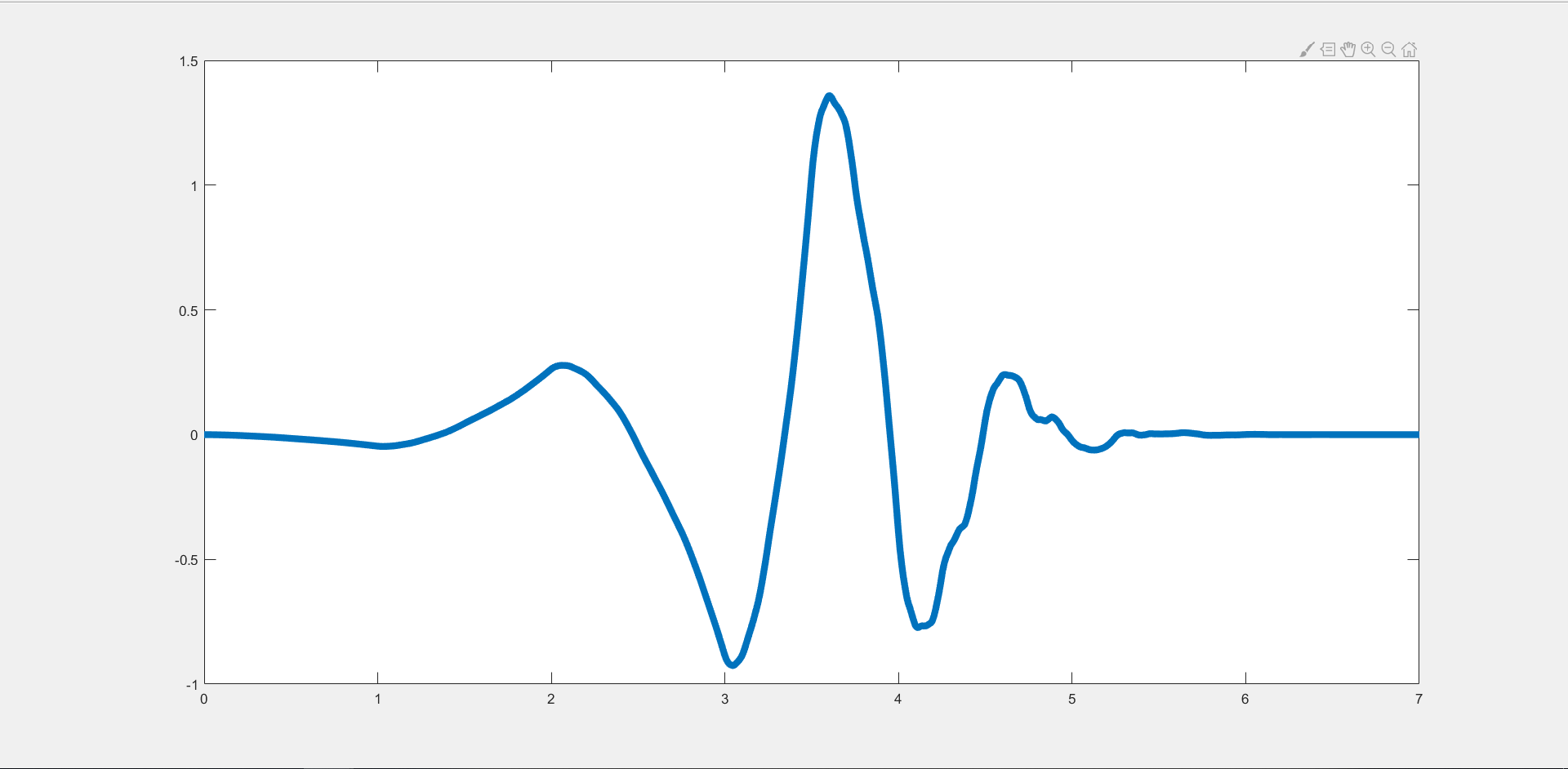


图3·6 Db4小波基函数图像

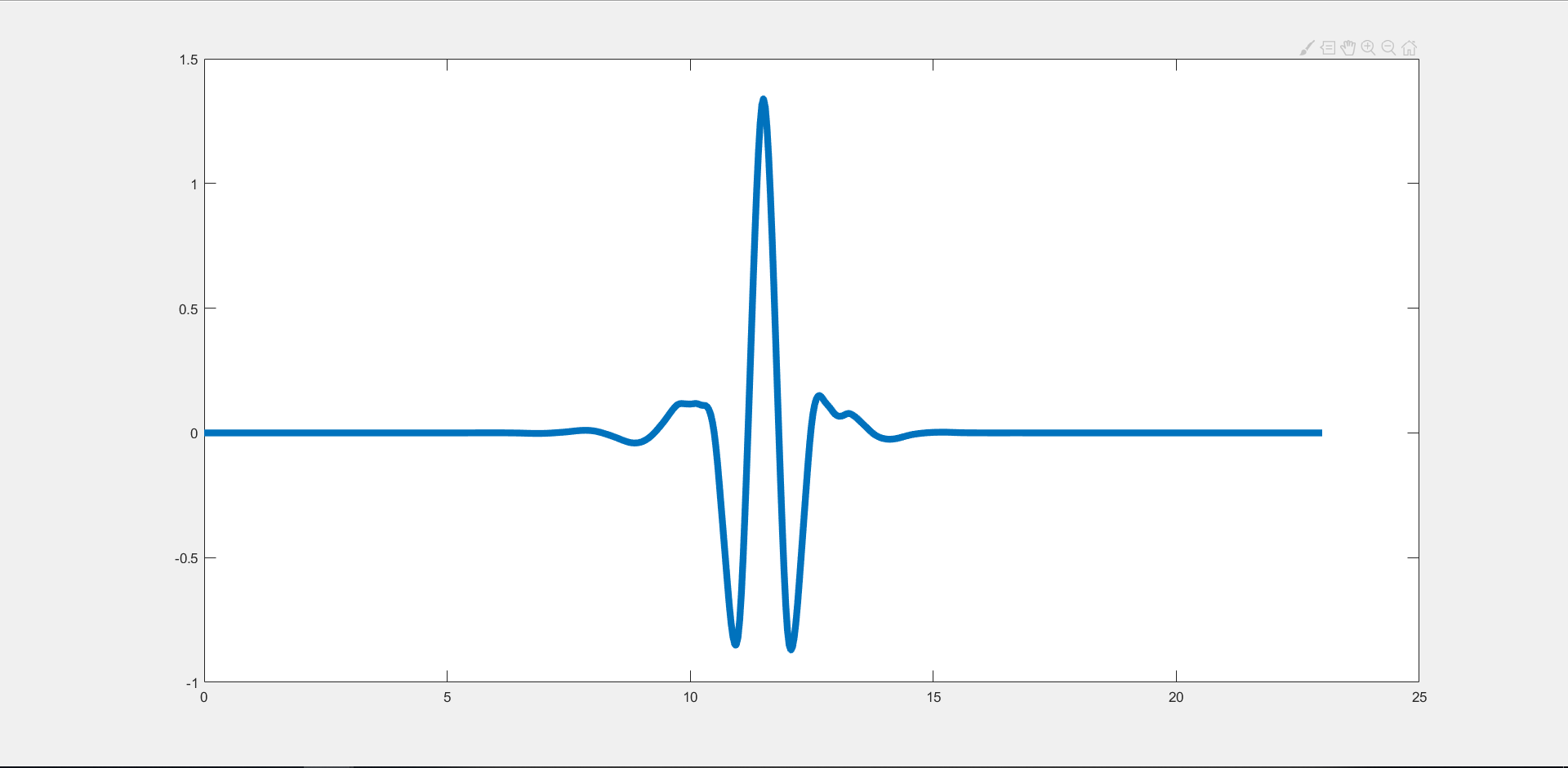


图3·7 Coif4小波基函数图像

小波基的选取要从正交性、支撑长度、对称性和正则性等几个方面的因素进行考虑，同时还要考虑所要处理的信号和小波基函数波形的相似性，如果两者的相似性越高，那么经过离散小波变换之后得到的小波系数就会越大，而如果相似性越低，那么小波系数就会越小。本研究应该尽量选取与癫痫脑电信号最为相似的小波基，综合以上所有的因素考虑之后，在图3·5、图3·6和图3·7这三张图像之中选择了’db4’小波基函数进行后续的离散小波变换。

癫痫脑电信号的特征提取

本研究已经确定要使用’db4’为小波基来进行5层的离散小波变换(DWT)，进而对脑电信号进行处理提取特征值，经过离散小波变换我们便可以得到待处理信号在不同频率范围上的近似系数和细节系数。

根据对波恩大学癫痫脑电数据库的数据集描述可知，采集该数据集时的采样频率为173.61Hz，由香农采样定理(Shannon Sampling Theorem)的相关描述可知，波恩大学癫痫脑电数据库的数据集的最大有用频率应该是采样频率的一半。那么可进一步得出由离散小波变换进行分解的频率带应该是0~86.81Hz。具体的分解过程如图3·8所示，对图3·8中提及的符号的详细描述如表3·5所示，cD1和cA1等与表3·5中的cA5和cD5类似，均为相应层数分解后得到的小波系数。

离散小波变换分解的过程如下：首先，在第一层分解中，输入的整个一维EEG脑电信号同时经过一个低通滤波器和一个高通滤波器，这两个滤波器将输入的EEG信号的频率分为了相等的两部分得到了详细小波系数cD1(频带为0~43.4Hz)和cA1(43.4~86.8Hz)，经过滤波之后，信号的时间分辨率减半，而信号的频率分辨率变为原来的两倍；然后cA1继续进行第二层的分解，同样也是经过一个低通滤波器和一个高通滤波器，进而得到cD2(频带为0~21.7Hz)和cA2(频带为21.7~43.4Hz)。该过程重复进行五次之后，可以得到cA5、cD1、cD2、 cD3 、cD4和 cD5六个小波系数，而这六个小波系数覆盖了0~86.8Hz的整个频带。按照这一流程，本研究将上文中经过分组的2000小组脑电信号记录均分别进行了五层离散小波变换，这样便可得到各为2000组的cA5、cD1、cD2、 cD3 、cD4和 cD5小波系数。

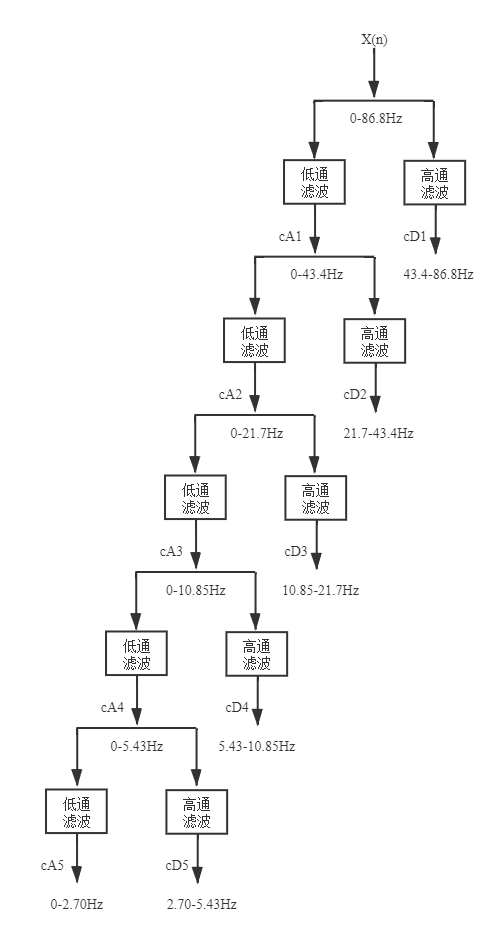


图3·8 基于离散小波变换的EEG信号分解过程

表3·5 离散小波变换分解过程符号解释

|  |  |
| --- | --- |
| 名称 | 描述 |
| X(n) | 一维离散输入信号 |
| cA5 | 第五层分解的近似小波系数 |
| cD5 | 第五层分解的近似小波系数 |

本研究以波恩大学癫痫脑电数据集中A组脑电数据中的第一条脑电记录的第一个小组(上文提及的依时间顺序以1024个数据点为一组)为例，将其脑电时序图与经过离散小波变换之后得到的六个小波系数进行了绘制，如图3·9中所示。

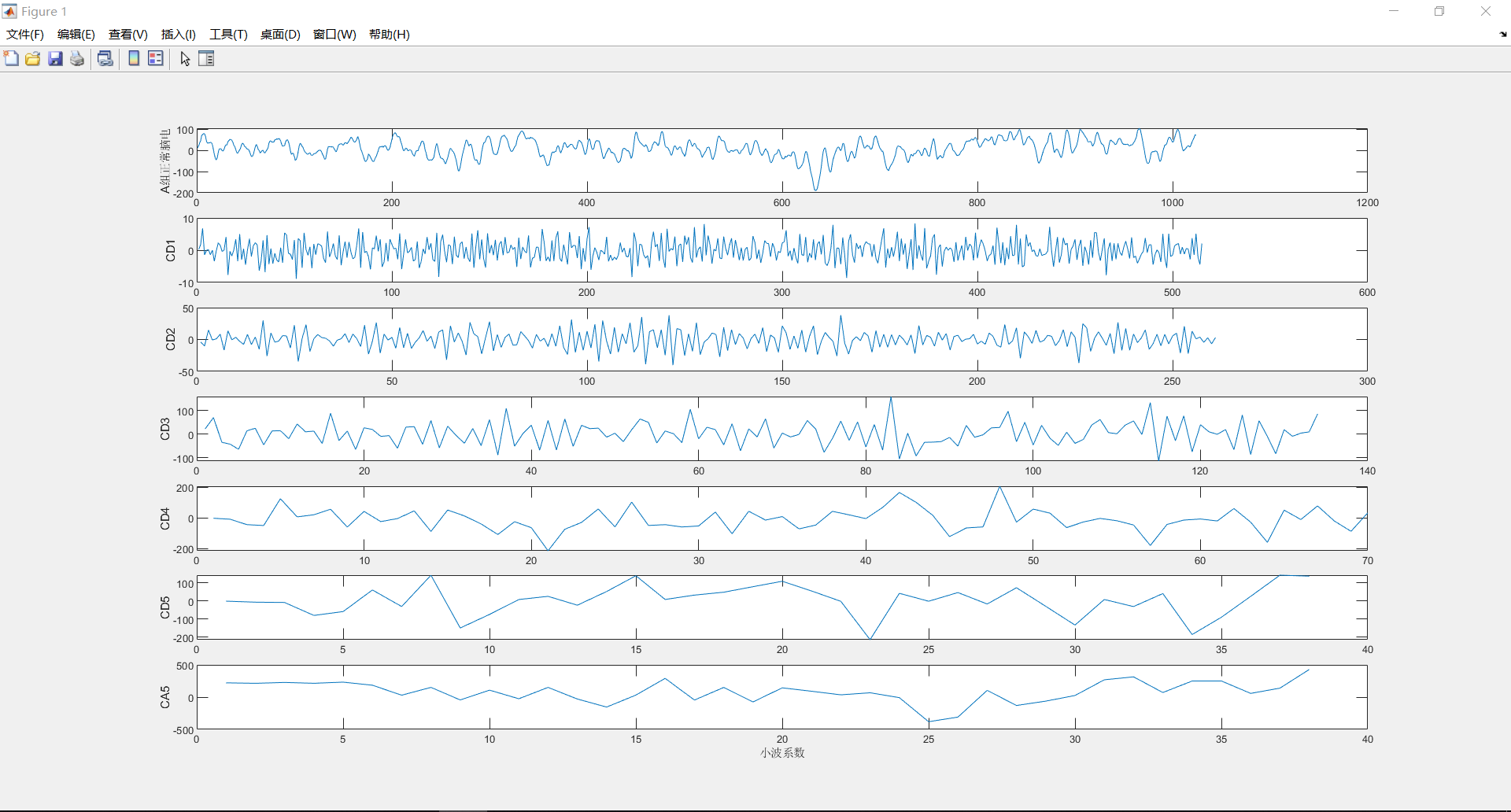


图3·9 A组脑电信号小波系数图像

特征选择与提取

在肖文卿等人的研究中，（改为引用文献，文献18）由于当小老鼠在癫痫发作时脑电信号的标准差是会低于小老鼠在正常生理情况下的脑电信号标准差，因此他们对小老鼠在两种不同状态之下的脑电信号自身的标准差进行了提取，并将其作为分类器的输入来对正常小鼠和癫痫小鼠进行了分类。

本研究参考了这一思想，对波恩大学癫痫脑电数据集中经过分组得到的2000组原始脑电信号记录和各为2000组的cD1、cD2、 cD3 、cD4和 cD5小波系数计算其标准差，这样便可得到共计6组(原始脑电信号、cD1、cD2、 cD3 、cD4和 cD5)标准差，每一组中则有2000个标准差。

此处以表3·2中的第一种情况即A vs E两组进行二分类为例，根据计算得到的原始脑电信号、cD1、cD2、 cD3 、cD4和 cD5的标准差绘制了图3·10所示的箱型图。经过对比可以发现，正常志愿者和正处于癫痫发作期的志愿者的自身脑电信号标准差值和cD1~cD5这五组小波系数的标准差值均有着十分明显的差别，而且处于癫痫发作期志愿者自身的EEG与cD1~cD5这五组小波系数的标准差值均大于正常志愿者的标准差值。所以根据图3·10，本研究选取志愿者自身的脑电信号标准差值与cD1~cD5这五组小波系数的标准差值作为接下来研究中对癫痫脑电信号进行分类的特征。

对于表3·2中的其他三种情况，同样进行上述类似的操作，并且分别经过与上述类似的六种方差的对比，也进行了相同的选择，并将其分别作为各自进行分类时的特征。

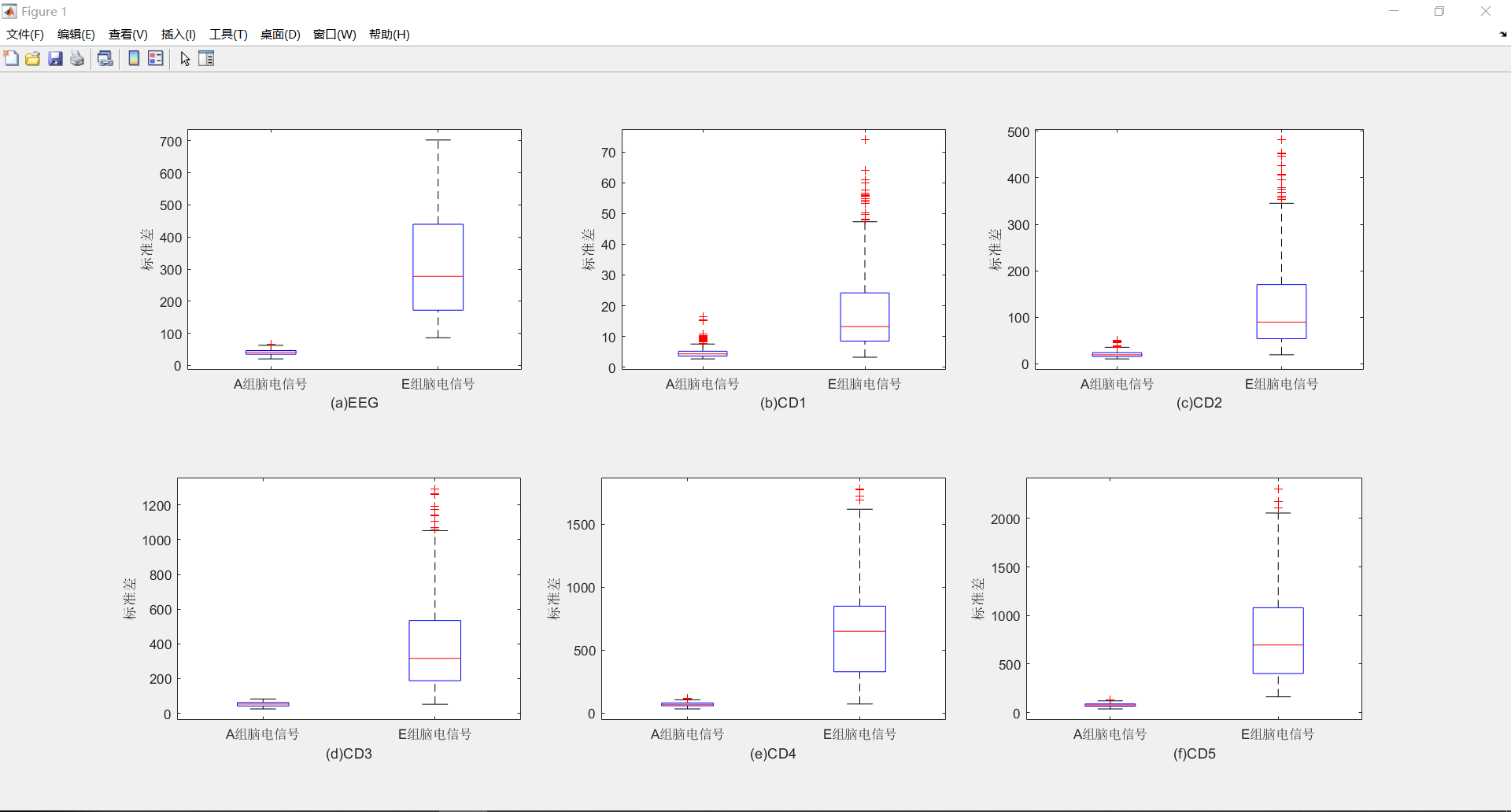


图3·10 AE两组标准差箱型图对比

对于经过预处理和时间选择的波士顿儿童医院的CHB-MIT头皮脑电图数据库的EEG脑电数据集，与上述对波恩大学癫痫脑电数据集进行的操作一样，也分别对病例chb01和chb03的各480组的脑电数据中的每一组也别进行了五层的离散小波变换，并且相应得到了cA5、cD1、cD2、 cD3 、cD4和 cD5这六组小波系数，也计算得到了原始脑电信号与这六组小波系数的标准差。在此处，本研究以病例chb01为例，同上述内容进行相同的操作，绘制了病例chb01在发作间期和发作期时的原始脑电、cD1、cD2、 cD3 、cD4和 cD5这五组小波系数的标准差相对比的箱型图，具体如图3·11所示。

观察图3·11可以发现，病例chb01当分别处于发作间期和发作期时，这两个阶段的自身脑电信号标准差值和cD1~cD5这五组小波系数的标准差值均有着十分明显的差别，而且其处于发作期的自身的EEG与cD1~cD5这五组小波系数的标准差值均大于其处于发作间期时的标准差值，而这一结论与对图3·10的分析一致。因此本研究选取病例chb01自身的脑电信号标准差值与cD1~cD5这五组小波系数的标准差值作为接下来研究中对癫痫脑电信号进行分类的特征。对于病例chb03的脑电信号数据也进行与病例chb01类似的操作，此处不再赘述。

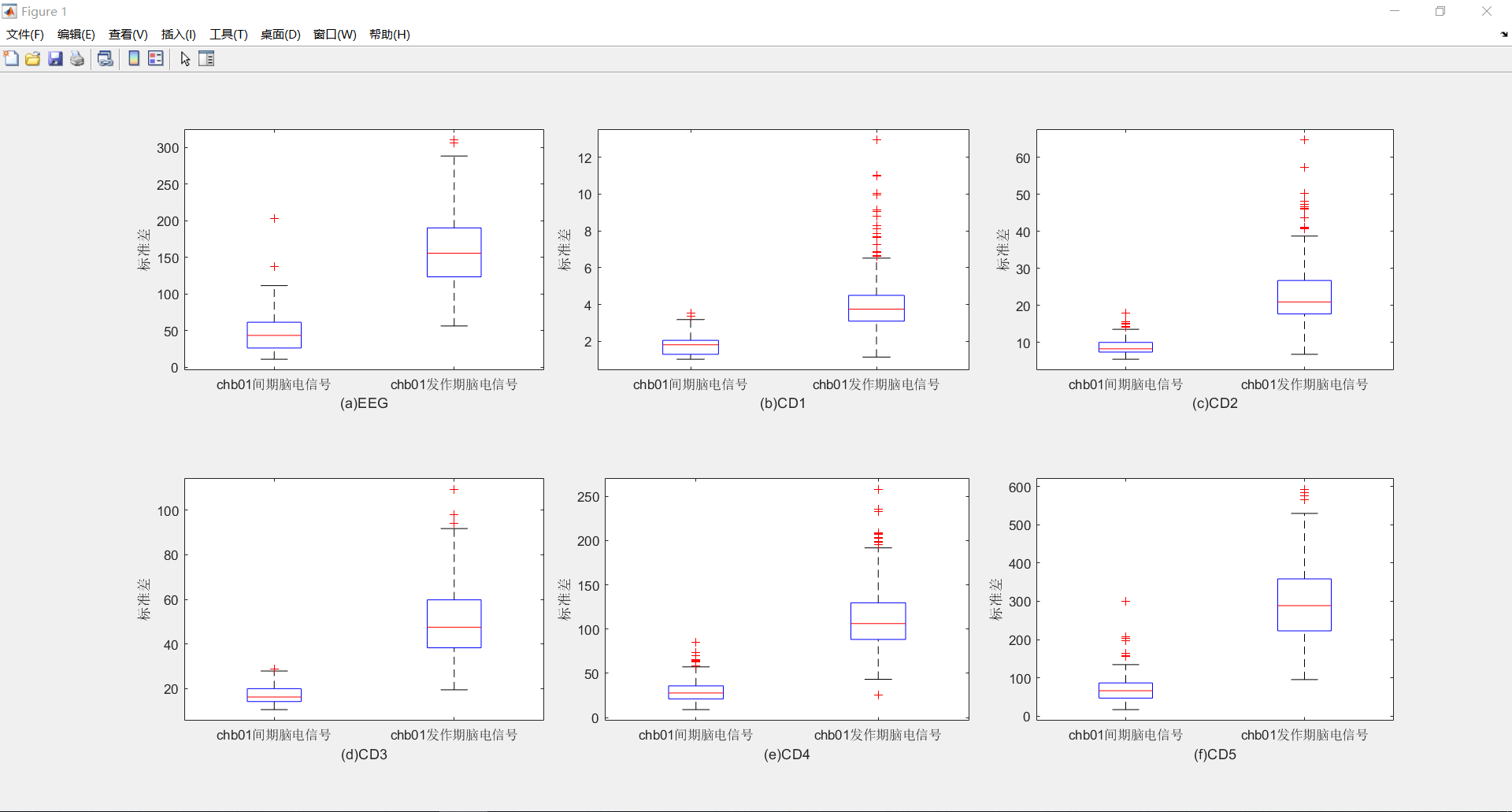


图3·11 病例chb01发作间期和发作期脑电信号方差箱型图对比

样本熵(Sample Entropy)最早由Richman等人提出，它从数据长度独立性和相对无故障的实现两个方面对近似熵(Approximate Entropy)进行了改进，它经常用来评估生理时间序列信号的复杂性，诊断疾病状态。根据Cheol Seung Yoo等人的研究，他们某某某发现在人体在癫痫发作过程中，脑电信号的样本熵值会下降，在癫痫发作终止时达到最小值，癫痫发作终止后，样本熵迅速恢复到癫痫发作前的数值。根据这一结论，本研究将脑电信号的样本熵这一非线性特征选择作为除去标准差之外的对脑电信号分类的另一个特征。

假定一个时间序列为，其时间间隔是一个固定的值，其中N为该序列中包含数据的个数。定义一组向量序列，该序列的维数是m，而，其中。定义距离函数为切比雪夫距离(Chebyshev distance)，即



定义样本熵为：



其中为阈值，，而则是对每一个值，均对的数目进行了计算与统计，并将该数目记作，对于，有：



而和也可以经过和上述类似的操作得到，只是在构成向量序列的时候与有所不同，在此处则是将原来的m维增加到了m+1维，即与，而其他的步骤与上面的步骤一样。

从中对样本熵的定义可以看出，的值始终小于或者等于的值，因此将始终为零或者正值。在研究中通常采用的参数值，其中代表信号的标准偏差。在本研究中，采用的参数。

同样与上文中对波恩大学癫痫脑电数据集中经过分组得到的2000组原始脑电信号记录和五组各为2000组的小波系数计算标准差类似，此处可计算得到共计6组(原始脑电信号、cD1、cD2、 cD3 、cD4和 cD5)样本熵，每一组中则有2000个样本熵。此处同样以表3·2中的第一种情况即A vs E两组进行二分类为例，根据计算得到的原始脑电信号、cD1、cD2、 cD3 、cD4和 cD5的样本熵绘制了图3·10所示的箱型图。

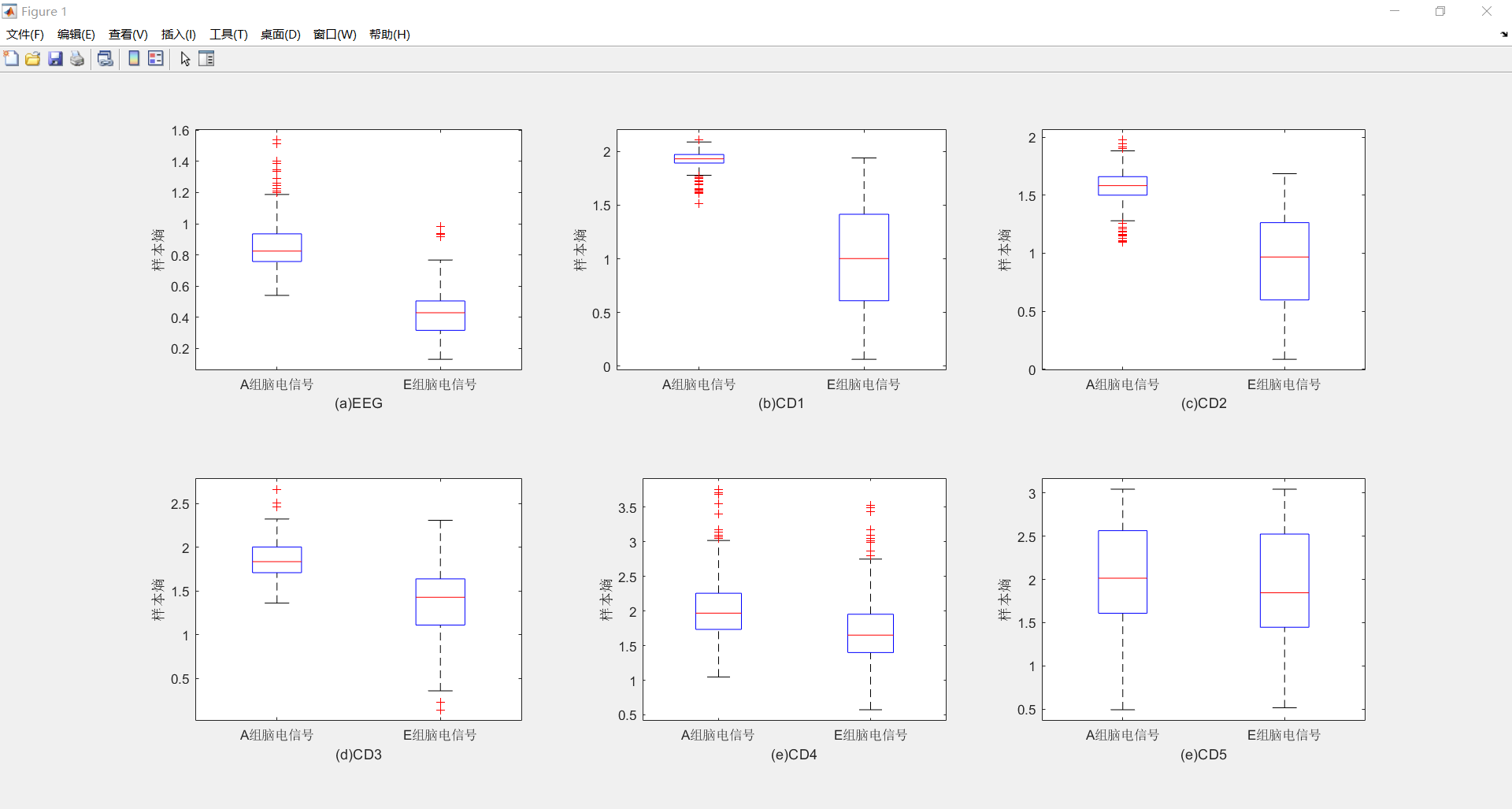


图3·12 AE两组样本熵箱型图对比

观察过图3·12可发现，与上文中对标准差作为特征时的选择不同，此处正常志愿者和正处于癫痫发作期的志愿者只有自身脑电信号样本熵值和cD1~cD3这四组小波系数的样本熵值均有着十分明显的差别，而且处于癫痫发作期志愿者自身的EEG与cD1~cD3这五组小波系数的样本熵值均小于正常志愿者的样本熵值，这与Cheol Seung Yoo等人的研究结论一致。所以本研究选取志愿者自身的脑电信号与cD1~cD3这四组数据的样本熵值作为接下来研究中对癫痫脑电信号进行分类的特征，其余三种情况的二分类也进行了相似的选择。

同样，经过预处理和时间选择的波士顿儿童医院的CHB-MIT头皮脑电图数据库的EEG脑电数据集计算样本熵值时与其计算标准差值时的方法类似，也可分别计算处病例chb01和chb03发作间期和发作期时自身的EEG与cD1~cD5这五组小波系数的样本熵值，并可分别绘制出如图3·13和图3·14所示的箱型图。

通过图3·13和图3·14，对于病例chb01和chb03分别选择小波系数cD3~cD5三组小波系数的样本熵值和自身脑电信号与小波系数cD3的样本熵值作为其分类的特征。

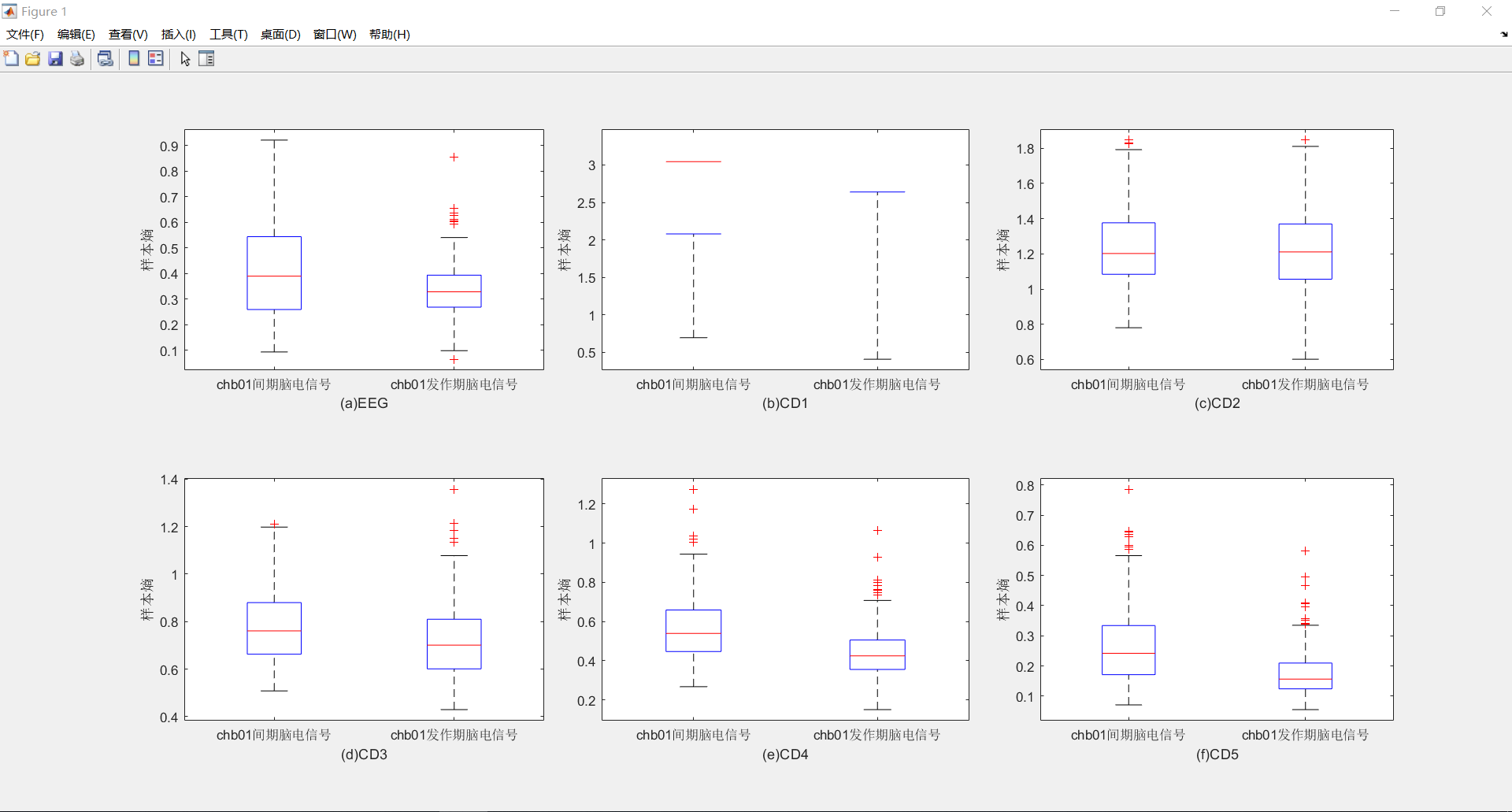


图3·13 病例chb01发作间期和发作期脑电信号方差箱型图对比

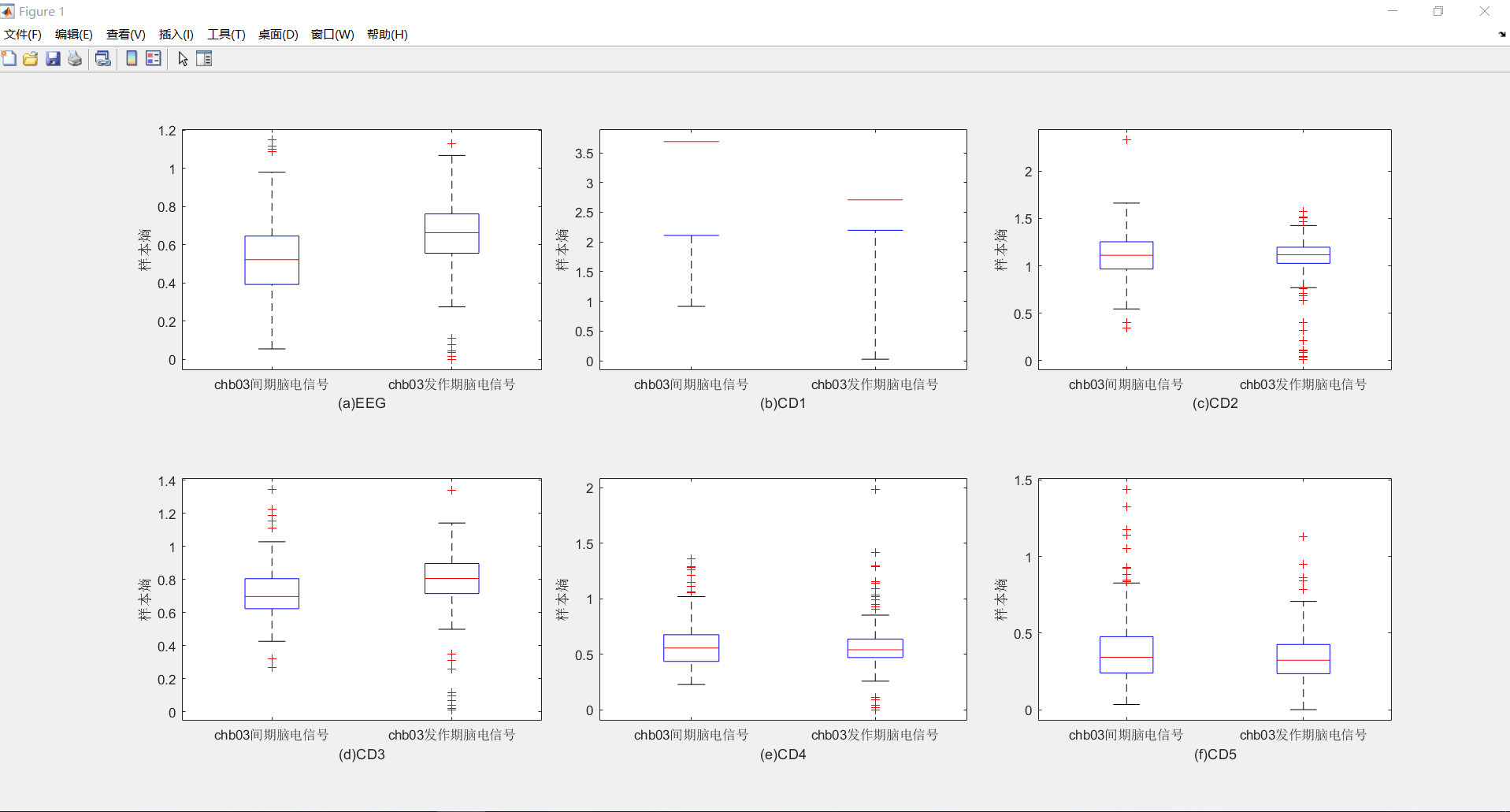


图3·14 病例chb03发作间期和发作期脑电信号方差箱型图对比

第3.5节 结果与分析

本章的主要内容和目的是基于离散小波变换(DWT)对脑电数据集进行特征的提取。首先本文对要用到的两个数据集进行了详细的介绍并进行了相应的预处理操作，得到了可用于分析研究的数据。然后对傅里叶变换、短时傅里叶变换和小波变换均进行了介绍，并对连续小波变换(CWT)和离散小波变换(DWT)的原理进行了阐述

第4章 基于多种机器学习算法的癫痫脑电信号分类

第4.1节 支持向量机

第4.2节 K最近邻算法

第4.3节 决策树

第5章 集成学习

第5.1节 支持集成学习理论介绍

第5.2节 集成学习1

第5.3节 集成学习2

结 论

本课题采用XXXXXXXXXXXXXXXXXXXXXXXXXX得到以下结论：

1、XXXXXXXXXXXXXXXXXXXXXX

2、XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX（宋体，小四，固定值22磅，首行缩进2字符）

参考文献

1. 唐颖莹,陆璐,周东.中国癫痫诊断治疗现状[J].癫痫杂志,2019,5(03):161-164.

[2] 刘光知．螺杆元件排列组合的混炼效果研究[J]．中国塑料，1993，7(3)：58．

[3] 耿孝正，张沛．塑料混合及设备[M]．北京：中国轻工业出版，1992：1．

[4] 李鹏，耿孝正，马秀清．啮合同向双螺杆挤出过程组合流道（捏合块+螺纹元件）三维流场分析[J]．中国塑料，2001，15(7)：1．

[5] Li Shaofan，Liu WK．Mesh free and particle methods and their applications[J]．ApplMech Rev，2002，55(3)：1-34．

[6] 张平亮．双螺杆挤出机的进展及其应用[J]．工程塑料应用，2005，33(5)：2-3．

[7] 李鹏，耿孝正．同向啮合双螺杆挤出机捏合块流道三维流场分析[J]．中国塑料，2000，14(3)：75．

[8] 王文飞．啮合同向双螺杆挤出机捏合块组合的研究[D]．北京：北京化工大学，2011．

[9] 邱国庆．同向旋转双螺杆挤出机螺旋状捏合盘元件混炼机理研究[D]．北京：北京化工大学，2010．

[10] Yang．H．H，Manas-Zloczower．I．Flow field analysis of the kneading disc region in a co-rotating twin screw extruder[J]．Polymer Engineering & Science，1992，32(19)：1411-1417．

[11] Shaffiq A．Jaffer，Victor L．Bravo，Philip E．Wood，Andrew N．Hfwmak．Experimental validation of numerical simulations of the kneading disc section in a twin screw extruder[J]．Polymer Engineering & Science，2000，40 (4)：892-901．

[12] White．J．L，Chen Ziyun．Simulation of non-isothermal flow in modular co-rotating twin screw extruder[J]．Polymer Engineering & Science，1994，34(3)：229-237．

[13] D．J．Van Der Wal，D．Goffart，E．M．Klomp，H．W．Hoogstraten, L．P．B．M．Janssen．Three-dimensional flow modeling of a self-wiping co-rotating twin screw extruder．Part II： the kneading section[J]．Polymer Engineering & Science，1996，36(7)：912-924．

[14] Yasuya Nakayama，EijiTakeda，TakashiShigeishi，HidekiTomiyama，Toshihisa Kajiwara．Melt-mixing by novel pitched-tip kneading disks in a co-rotating twin-screw extruder[J]．Chemical Engineering Science，2011，66 (1)：103-110．

[15] Takeshi Ishikawa，Shin-Ichi Kihara，Kazumori Funatsu．3-D numerical simulations of nonisothermal flow in co-rotating twin screw extruders[J]．Polymer Engineering＆Science．2000，40(2)：357-364．

[16] V．L．Bravo，A．N．Hrymak．Numerical simulation of pressure and velocity profiles in kneading elements of a co-rotating twin screw extruder[J]．Polymer Engineering & Science，2000．40(2)：539．

[17] Yoshinaga，Makoto，Katsuki，Shingo，Miyazaki，Masam．Mixing mechanism of three-tip kneading block in twin screw extruders[J]．Polymer Engineering & Science，2000，40(1)：168-178．

[18] 李鹏，耿孝正．啮合同向双螺杆挤出机捏合块流道流场分析[D]．北京：北京化工大学，2000．

[19] 刘青峰，马秀清．同向双螺杆挤出过程不同螺杆构型的混合性能分析[J]．橡胶技术与装备，2006，32(9)：1-6．

[20] Xian-Ming Zhang，Lian-Fang Feng，Wen-Xing Chen，Guo-Hua Hu．Numerical simulation and experimental validation of mixing performance of kneading discs in a twin screw extruder[J]．Polymer Engineering & Science，2009，49(9)：1772-1783．

[21] 尹燕玲，耿孝正，马秀清．啮合同向双螺杆挤出过程新型混合元件——六棱柱元件实验研究[J]．中国塑料，2002，16(6)：76-80．

（宋体，小四，固定值22磅，首行缩进2字符，至少15篇中文文献，5篇外文文献）

致 谢

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXxXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

（宋体，小四，固定值22磅，首行缩进2字符）