# 特征选取

肖文卿等人在研究中，对小鼠的脑电信号进行了离散小波变换(DWT)，在得到小波系数后，分别计算出了正常小鼠和癫痫小鼠的原始脑电信号和小波系数的标准差，然后将两者进行了对比。结果表明，正常小鼠和癫痫小鼠的原始脑电信号和小波系数的标准差之间存在显著差异，可以作为两者分类的特征。本文参照该研究，将标准差选定作为人类癫痫脑电信号分类的特征之一。

Cheol Seung Yoo等人在研究中，根据脑电信号的样本熵在癫痫活动中降低，在癫痫发作终止时取得最小值，样本熵不仅反映了癫痫发作前后脑电信号的规律性和复杂性的不同状态，而且反映了癫痫发作前后脑电信号振幅的变化这一结论，基于样本熵对癫痫发作终止时间进行了自动检测。本文参照其关于癫痫发作时样本熵变化的结论，将样本熵选定作为癫痫脑电信号分类的另一特征。

# 方案确定

在本研究中，首先对脑电信号进行离散小波变换(DWT)得到各个频段上的小波系数，之后计算出原始脑电信号与小波系数的标准差和样本熵，对其进行对比之后，选择差异较为显著的作为之后分类器的输入。

对于分类，本研究计划将其分为单一属性的二分类和多属性的二分类。对于单一属性的二分类，即为每一个属性单独作为分类器的输入；对于多属性的二分类，即为融合多个单一属性组成特征向量矩阵作为分类器的输入。

对于分类算法，本研究采用机器学习中的三种方法，分别是支持向量机(SVM)、K最近邻算法(KNN)和决策树(DT),这三种方法均在matlab上进行设计实现。对于结果的评价指标，由于一项诊断试验需要具备能正确地鉴别患病和未患病的能力，以反映患病实际情况的准确程度，因此本研究选取准确度(accuracy)、灵敏度(sensitivity)和特异度(specificity)这三个指标来评价分类器的性能。

# 初步设计

本研究中使用的数据集来源于波恩大学电信中心数据库，该数据集中包含五组数据(用A-E表示)，其中只有E组对癫痫患者癫痫发作时的脑电信号进行了记录，而AB两组是健康志愿者不同状态下的正常脑电信号，CD两组分别是癫痫患者处于发作间期时对病灶区外内的脑电信号。因此在本研究中，将其分为四种情况的二分类，分别是A vs E、B vs E、C vs E和D vs E。

A－E这五组数据中，每一组中均包含100个持续23.6s的单通道脑电片段，每个片段包含4097个数据点,脑电信号采样频率为173.61Hz。为了增加样本的数量，首先对每个单通道脑电片段进行分组，具体为每1024个数据点为一组，因此A－E五组数据共可分为2000组脑电数据片段。

然后对这2000组脑电数据片段进行离散小波变换(DWT)，之后针对原始脑电数据片段与计算得到的小波系数，分别计算其标准差和样本熵。接着对以上提及的四种分类情况分别根据计算得到的标准差和样本熵来绘制出箱型图对比，选择出其中差异较大的特征作为后续分类器的输入。

分类算法均在matlab上设计实现，同时每次分类均进行若干次交叉验证。对于支持向量机(SVM)，对单一属性的二分类，选择线性核函数；对于多属性的二分类，选择RBF核函数，同时要对支持向量机的参数进行寻优。对于K最近邻算法(KNN)，其重点在于选择合适的K；KNN中对于距离的度量有以下常用的几种：欧几里得距离、余弦值（cos）, 相关度 （correlation）, 曼哈顿距离 （Manhattan distance）等，在本研究中，选取欧几里得距离。对于决策树(DT)，其重点在于生成决策树和对决策树进行剪枝以防止模型过拟合，在本研究中，通过matlab环境下的相关函数实现决策树的构建和剪枝进而实现分类。

# 计算过程

特征提取流程如图 1所示，A－E五组数据共可分为2000组脑电数据片段，之后对这2000组脑电数据片段进行离散小波变换(DWT)，其中离散小波变换是以’db4’为小波基，进行五层小波分解，得到cD1、 cD2 、cD3、cD4、 cD5和 cA5六个频带上的小波系数，具体代码如图 2所示。

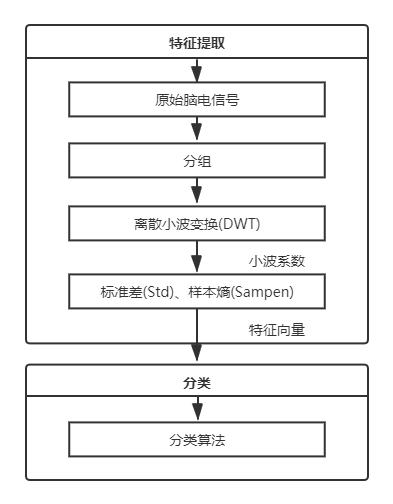


图 1　特征提取框图



图 2　 A组脑电信号小波系数计算

以下皆以AE两组二分类为例进行说明，在对AE两组的原始脑电信号与小波系数cD1~cD5、cA5读取，如图 2和图 3所示，之后选择出原始脑电信号与小波系数cD1~cD5，计算其标准差和样本熵，标准差计算如图 5所示，样本熵计算与其类似。

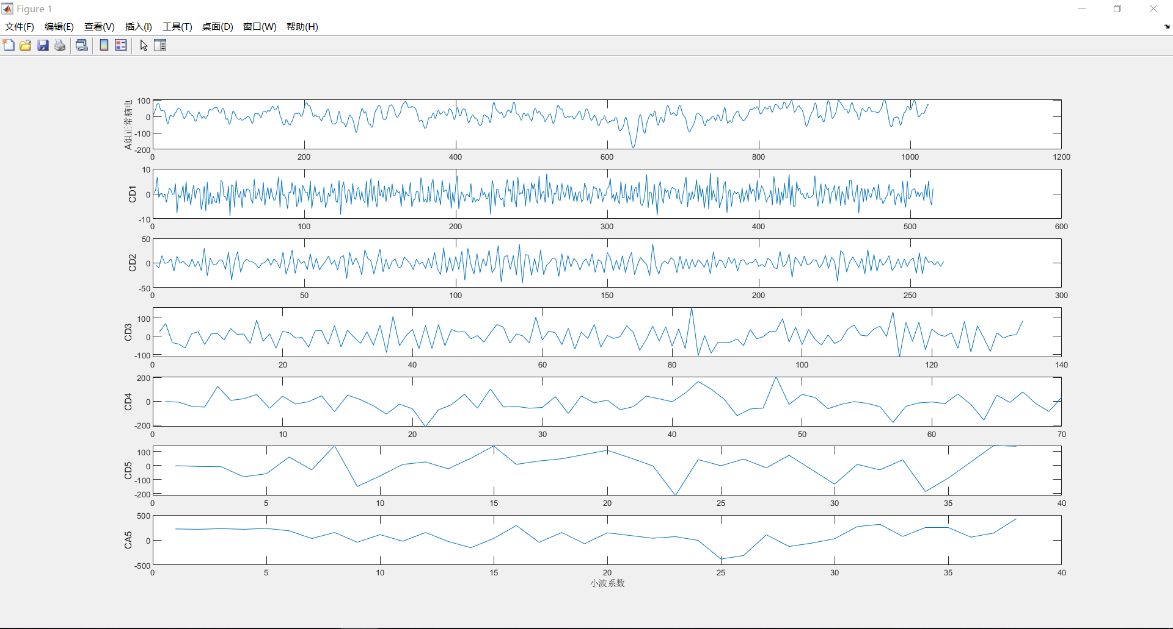


图 3　 A组脑电信号小波系数

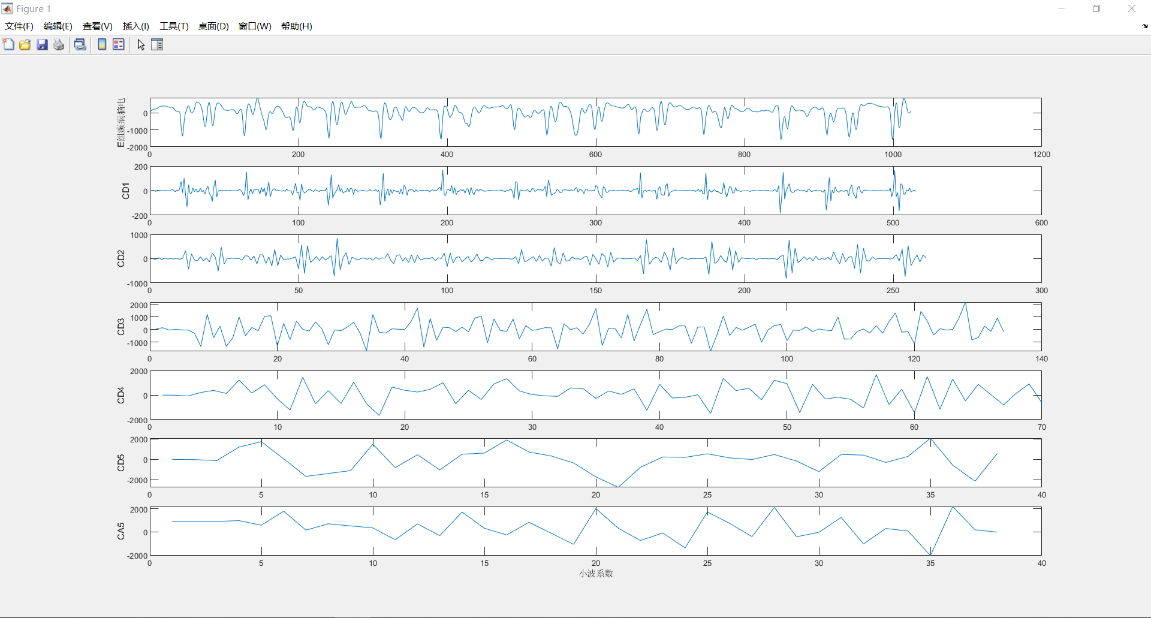


图 4　 **E**组脑电信号小波系数

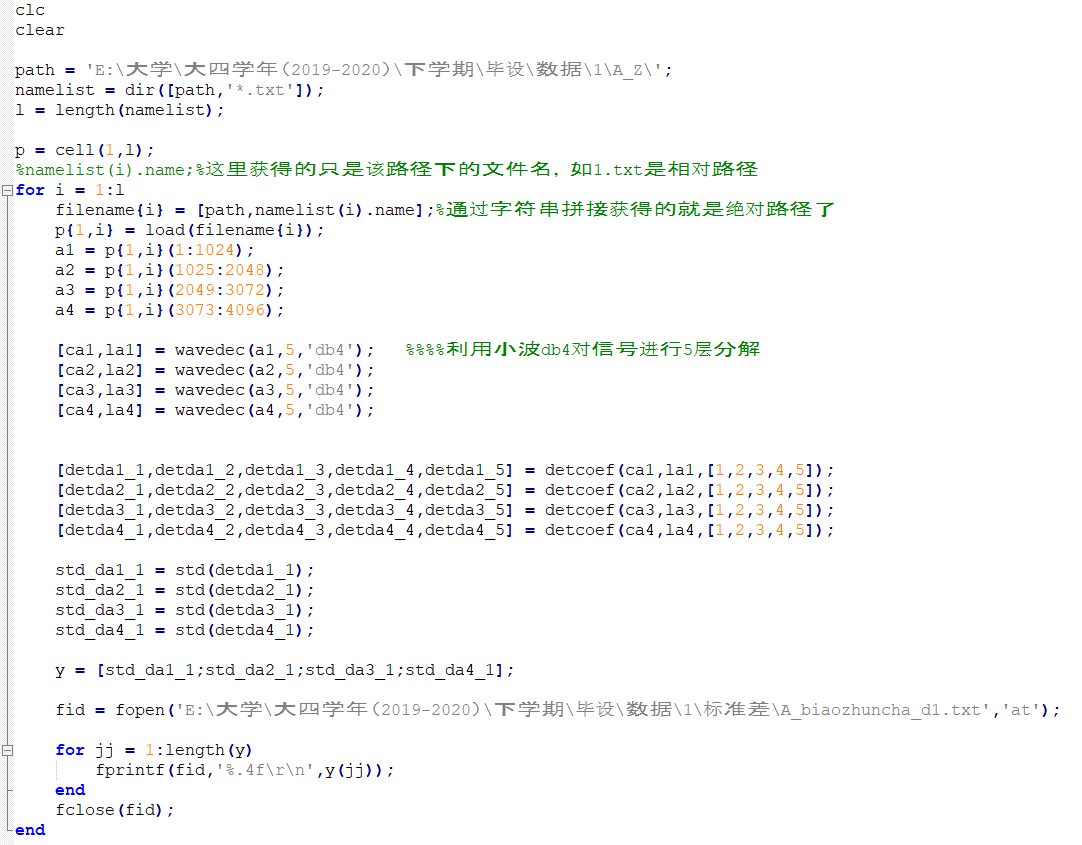


图 5　A组小波系数cD1标准差计算

计算出结果之后，绘制出箱型图进行对比，之后可以选择出10维特征向量，每维800个数据，每一维单独作为分类器的输入进行上述四种情况的分类；每个分类均进行5折交叉验证，每个分类进行3次，最后的结果取平均值。图 6中展示了使用支持向量机对AE两组根据小波系数**cD1**的标准差进行分类的代码，其他情况也是类似，此处不在展示。



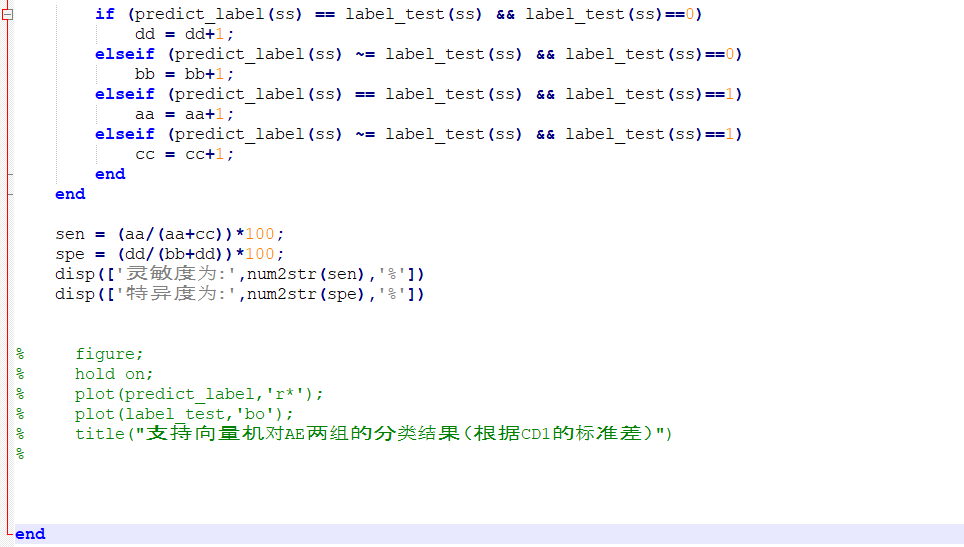


图 6　使用SVM对AE两组根据小波系数cD1标准差进行分类