

# 脑肌电信号的分类识别

小组成员：安俊杰、伍毅、符鹏阳

2023年12月27日



浙江大学 控制科学与工程学院  
COLLEGE OF CONTROL SCIENCE AND ENGINEERING  
ZHEJIANG UNIVERSITY





# 目录

01

特征提取

02

分类算法

03

对比分析



# 01

## 特征提取



浙江大学 控制科学与工程学院  
COLLEGE OF CONTROL SCIENCE AND ENGINEERING  
ZHEJIANG UNIVERSITY



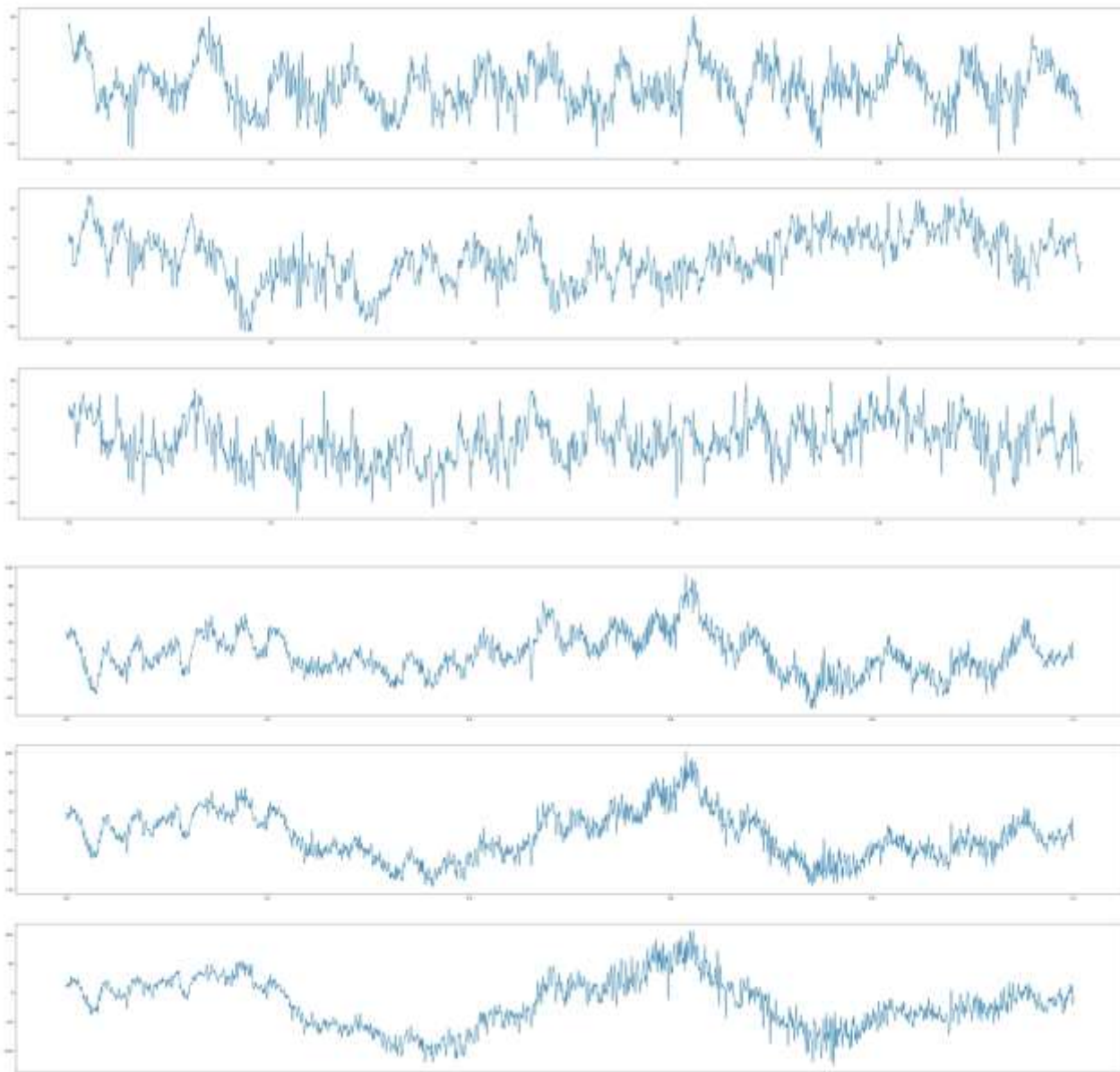
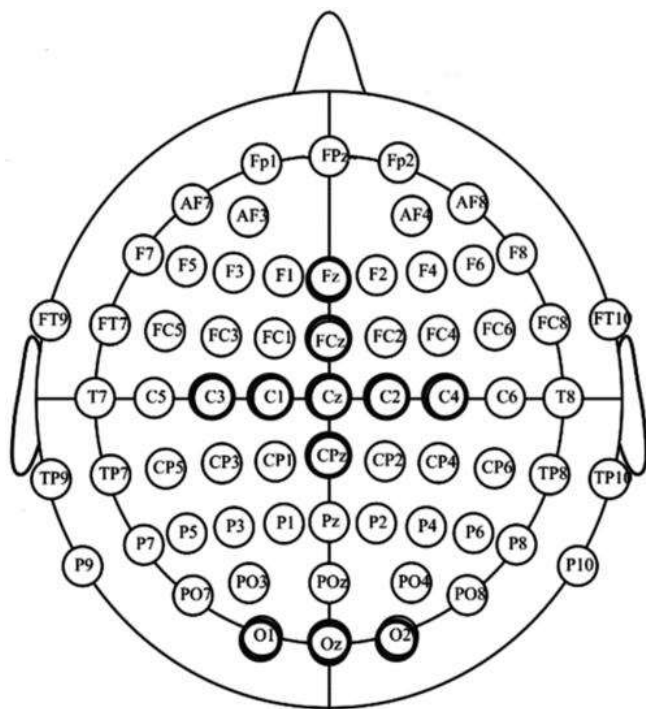


## 数据说明

脑电信号：样本数5500 电极数60

肌电信号：样本数5500 电极数3

信号长度3000





## 时域特征的提取

时域特征量作为一类典型的信号分析指标，可以直观反应脑电信号的信息。

常用的有量纲时域特征参数主要包括：最大值、最小值、峰值、均值、方差、标准差、绝对值的平均值、均方值、均方根值等。

针对本次实践我们选择了绝对值的平均值、均方根、过零率、短时平均幅度差和信号转折次数。

首先我们对一段信号进行分帧处理，以增强特征的时间关联性。

绝对值的平均值：每个时刻信号值的绝对值之和，他可以反映该信号的总能量

$$Av = \frac{1}{N} \sum_{i=1}^N (|X(i)|)$$

均方根：均方根值也称作有效值，将所有值平方求和，取其平均数再开方就得到均方根值，可用于描述信号的波动幅度，反映出信号中交流成分的强度

$$Rms = \sqrt{\frac{1}{N} \sum_{i=1}^N X^2(i)}$$

过零率：过零率表示一段时间内波形信号穿过零值的次数，表示了一个信号的符号变化的比率

$$Z = \sum_{i=1}^N \text{sgn}(X(i)X(i+1))$$

短时平均幅度差：相邻信号值之差的绝对值之和，它反映了信号变化的速度

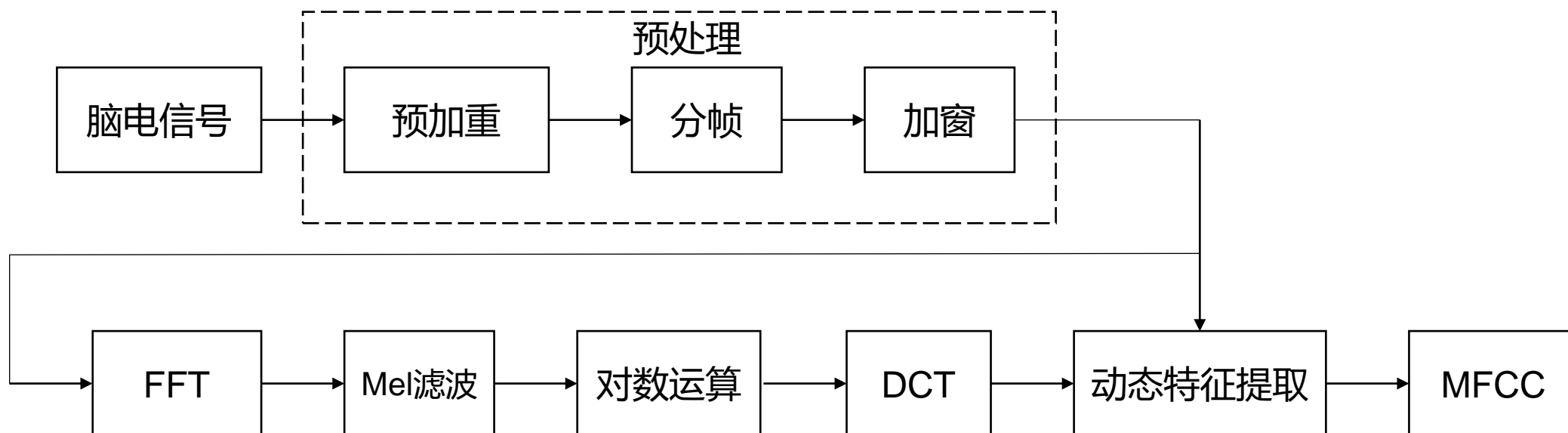
$$F = \sum_{i=1}^N |X(i) - X(i+1)|$$

信号转折次数：信号到达极值的次数，表示信号转折的频率

$$N = \sum_{i=2}^{N-1} \text{sgn}[(X(i) - X(i+1))(X(i-1) - X(i))]$$

## 频域特征的提取

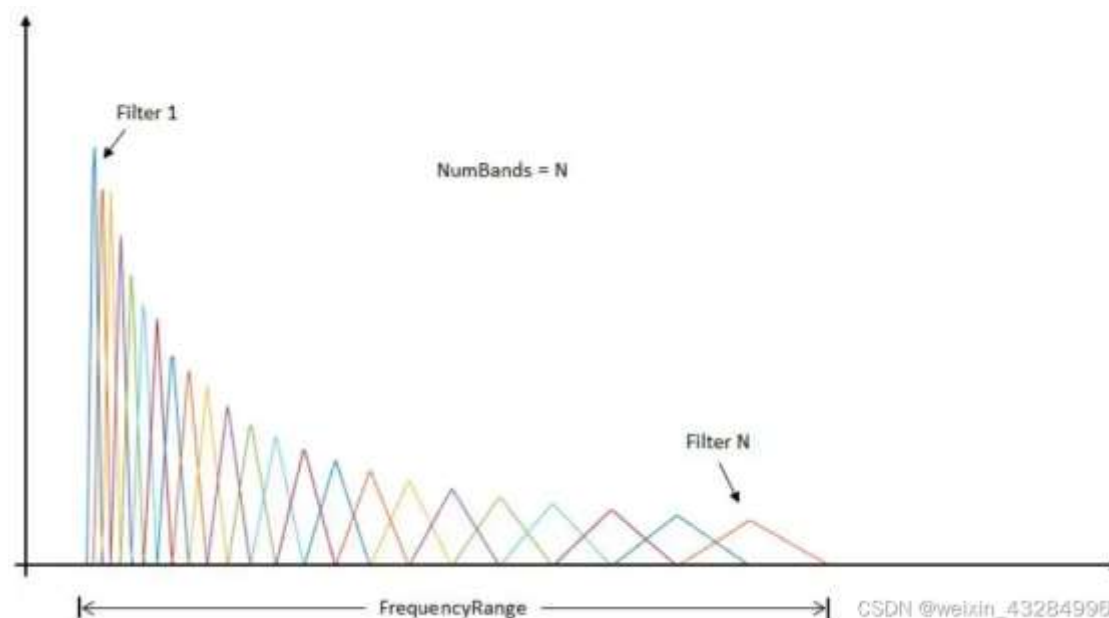
频域特征量采用梅尔倒谱系数（Mel-scale Frequency Cepstral Coefficients，简称MFCC），MFCC的提取过程包括预处理、快速傅里叶变换、Mel滤波器组、对数运算、离散余弦变换、动态特征提取等步骤。







## Mel滤波器组：



研究表明，人类对频率的感知并不是线性的，并且对低频信号的感知要比高频信号敏感。对1kHz以下，与频率成线性关系，对1kHz以上，与频率成对数关系。频率越高，感知能力就越差。为了模拟人耳的听觉机制。从而研制出来了Mel滤波器组。

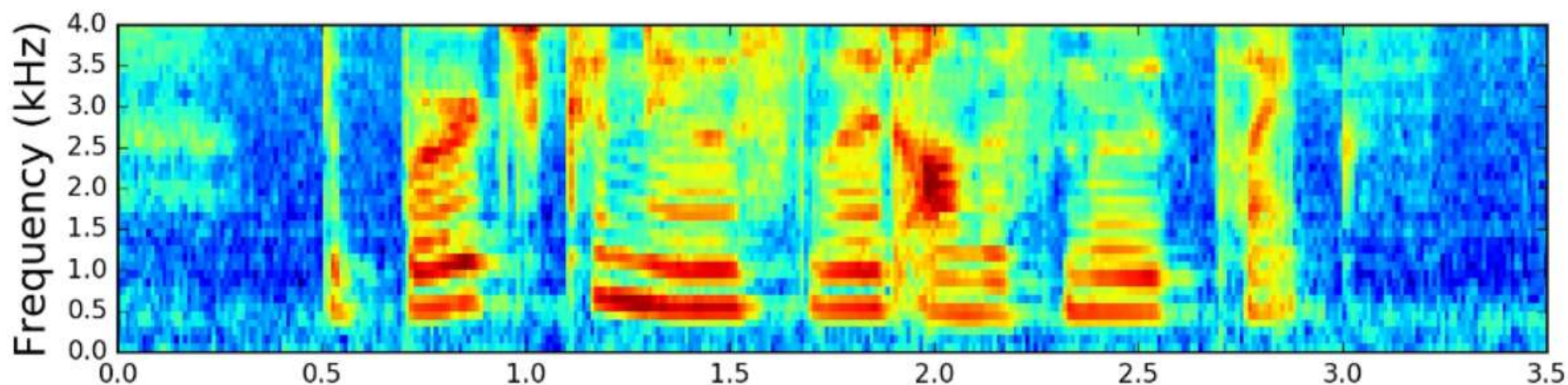


设置M个滤波器，每个滤波器的中心频率为

$$f(m) = \left( \frac{N}{f_s} \right) \text{Mel}^{-1} \left( \text{Mel}(f_{\text{low}}) + m \frac{\text{Mel}(f_{\text{low}}) - \text{Mel}(f_{\text{high}})}{M+1} \right)$$

根据中心频率得到滤波器函数值

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}$$





经过特征提取，我们将数据压缩，大大减少了后续分类算法的计算量

脑电信号特征数量：  $60 \times 3000 \rightarrow 1315$

肌电信号特征数量：  $3 \times 3000 \rightarrow 600$

02

# 分类算法



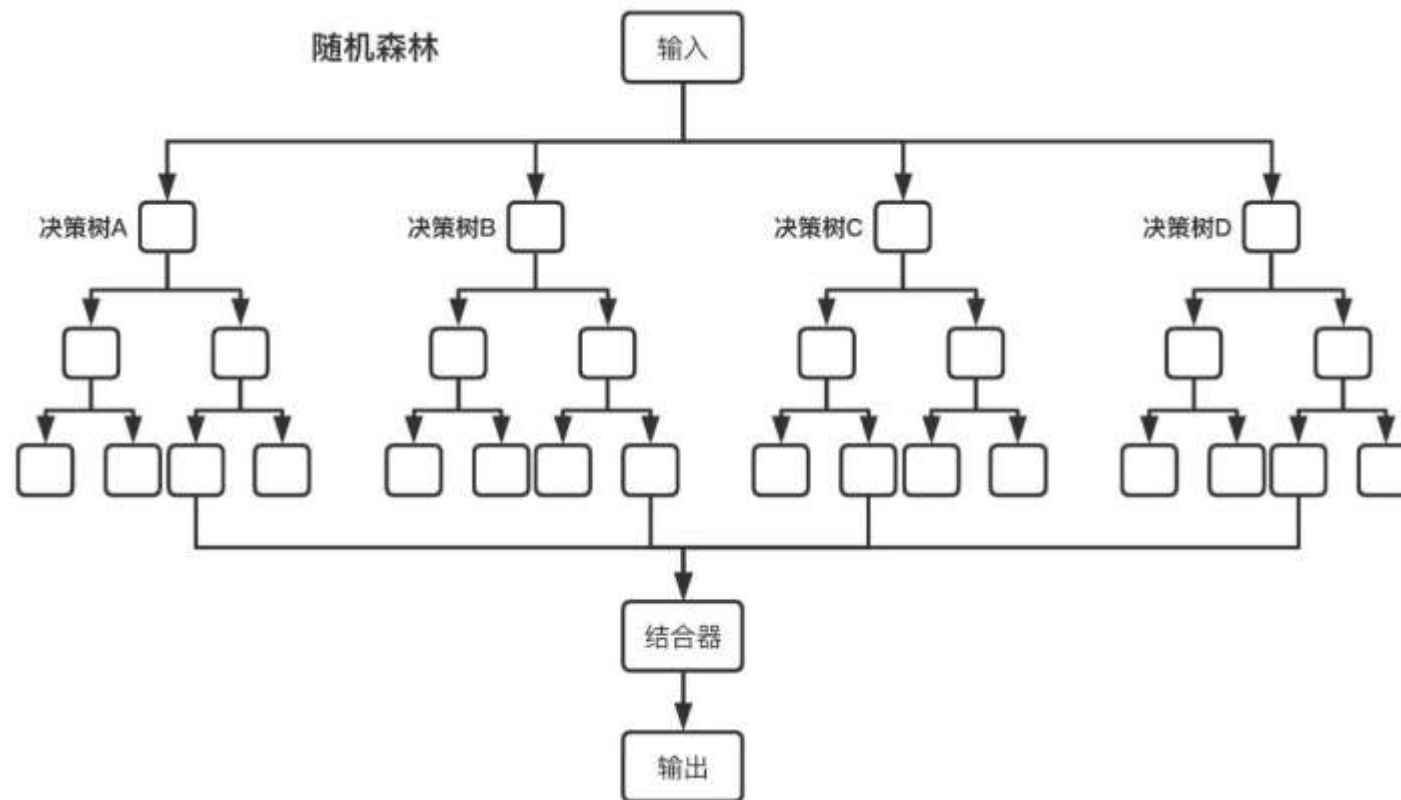
浙江大学 控制科学与工程学院  
COLLEGE OF CONTROL SCIENCE AND ENGINEERING  
ZHEJIANG UNIVERSITY







## 随机森林算法



将多个决策树结合在一起，每次数据集是随机有放回的选出，同时随机选出部分特征作为输入，所以该算法被称为随机森林算法。

输入参数 $n\_estimators=500$ ,  $max\_depth=40$ ，得到正确率0.871

## Lightgbm算法

GBDT(Gradient Boosting Decision Tree) 梯度提升决策树是一种学习策略，XGBoost是一种典型的boosting迭代型GBDT算法。它对所有特征都按照特征的数值进行预排序。其次，在遍历分割点的时候用 $O(\#data)$ 的代价找到一个特征上的最好分割点。最后，在找到一个特征的最好分割点后，将数据分裂成左右子节点。对于高维数据处理时空间时间消耗大，效果不好。

Lightgbm算法能够在不损害准确率的前提下加快GBDT模型的训练速度。采用了直方图算法将遍历样本转变为遍历直方图，极大的降低了时间复杂度；在训练过程中采用单边梯度算法过滤掉梯度小的样本，减少了大量的计算。

运行测试得到Lightgbm的准确率为0.967  
XGboost的准确率为0.949



## 集成学习

为进一步提升准确率，采用集成学习的方法，将多个机器学习算法融合，提升准确率和交叉准确率。

Stacking：两个阶段的模型。第一个阶段的模型是以原始训练集为输入的模型，叫做基模型（也叫 level-0 模型），可以选取多种基模型进行训练。第二个阶段的模型是以基模型在原始训练集上的预测作为训练集，以基模型在原始测试集上的预测作为测试集，叫做元模型（也叫 level-1 模型）。

Voting：投票法是一种简单且常用的集成学习方法。在投票法中，多个基学习器（如随机森林、XGBoost、LightGBM等）被训练并组合在一起，通过对每个基学习器的预测结果进行投票来确定最终的集成预测结果。投票法可以根据投票方式分为两种类型：硬投票和软投票。

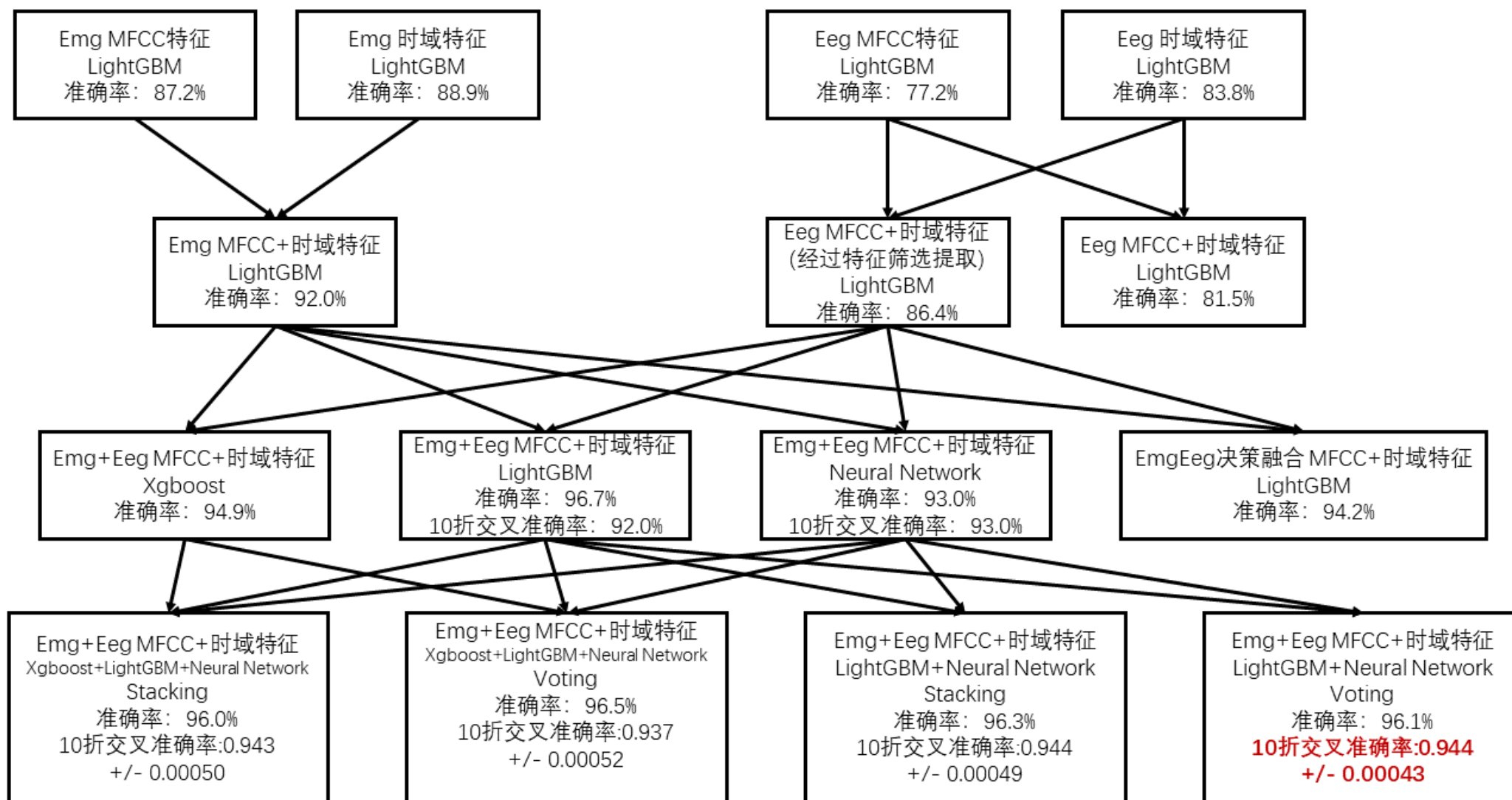
# 03 对比分析



浙江大学 控制科学与工程学院  
COLLEGE OF CONTROL SCIENCE AND ENGINEERING  
ZHEJIANG UNIVERSITY

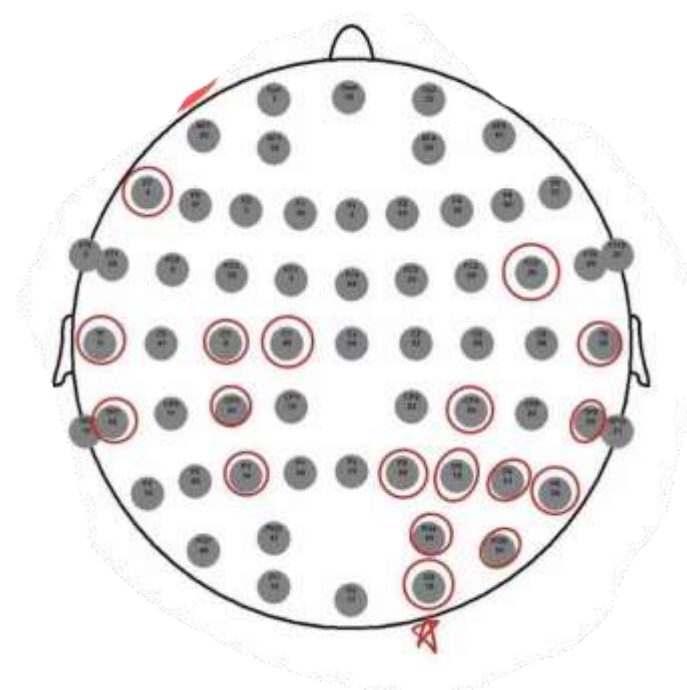
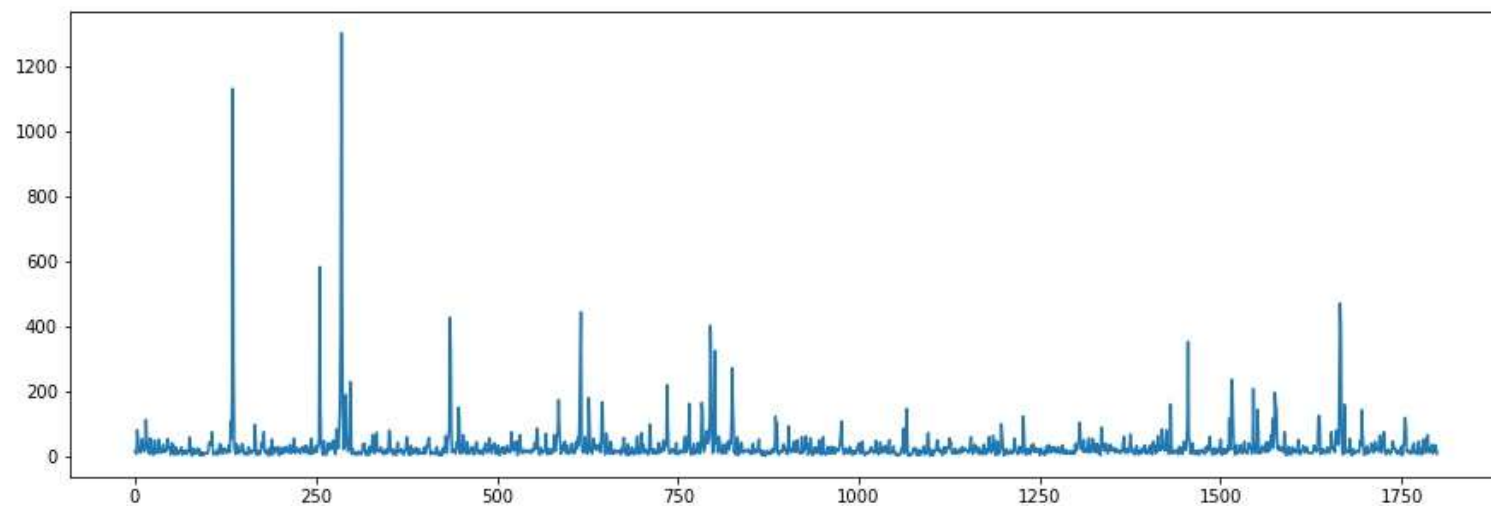








通过主成分分析哪个位置的电极作用最大



# 感谢观看



浙江大学 控制科学与工程学院  
COLLEGE OF CONTROL SCIENCE AND ENGINEERING  
ZHEJIANG UNIVERSITY

