# 数据可视化

## Jupyter Notebook绘图配置显示

%matplotlib inline#jupyter中嵌入显示

%config InlineBackend.figure\_format = ‘retina’#呈现高分辨率图片

plt.rcParams[‘font.sans-serif’] = [‘SimHei’]#显示中文标签

plt.rcParams[‘axes.unicode\_minus’] = False#显示负号

from matplotlib.font\_manager import FontProperties#显示中文

font = FontProperties(fname = ‘路径 Sans GBW3.otf’, size = 14)#设置字体

## mpl

### 直方图hist

#### hist参数

n, bins, patches = plt.hist(arr, bins = 10, normed = 0, facecoloe = ‘balck’, edgecolor = ‘black’, alpha = 1. histtype = ‘bar’)

arr：直方图的一维数组

bins：柱数

normed：是否将得到的直方图向量归一化，默认0

facecolor：直方图颜色

edgecoloe：边框颜色

alpha：透明度

histtype：直方图类型，有bar，barstacked，step，stepfilled

n：直方图向量，是否归一化由normed设定

bins：返回各bin的区间范围

patches：返回每个bin里面包含的数据，是个list

#### 拟合曲线

y = mlab.normpdf(bins, mu, sigma)#返回关于数据的pdf值（概率密度函数）

plt.plot(bins, y, ‘r—‘)

### 等值线图contour

contour将三维图像在二维空间表示

clabel在每条线上显示数据值大小

### 三维曲面图plot\_surface

### 条形图

### 饼图pie

### 气泡图scatter

## Seabon

## pandas

## 文本数据可视化

## network网络图

## 地图folium

## Python图像处理基础

### PIL图库

### OpenCV图库

# 名词

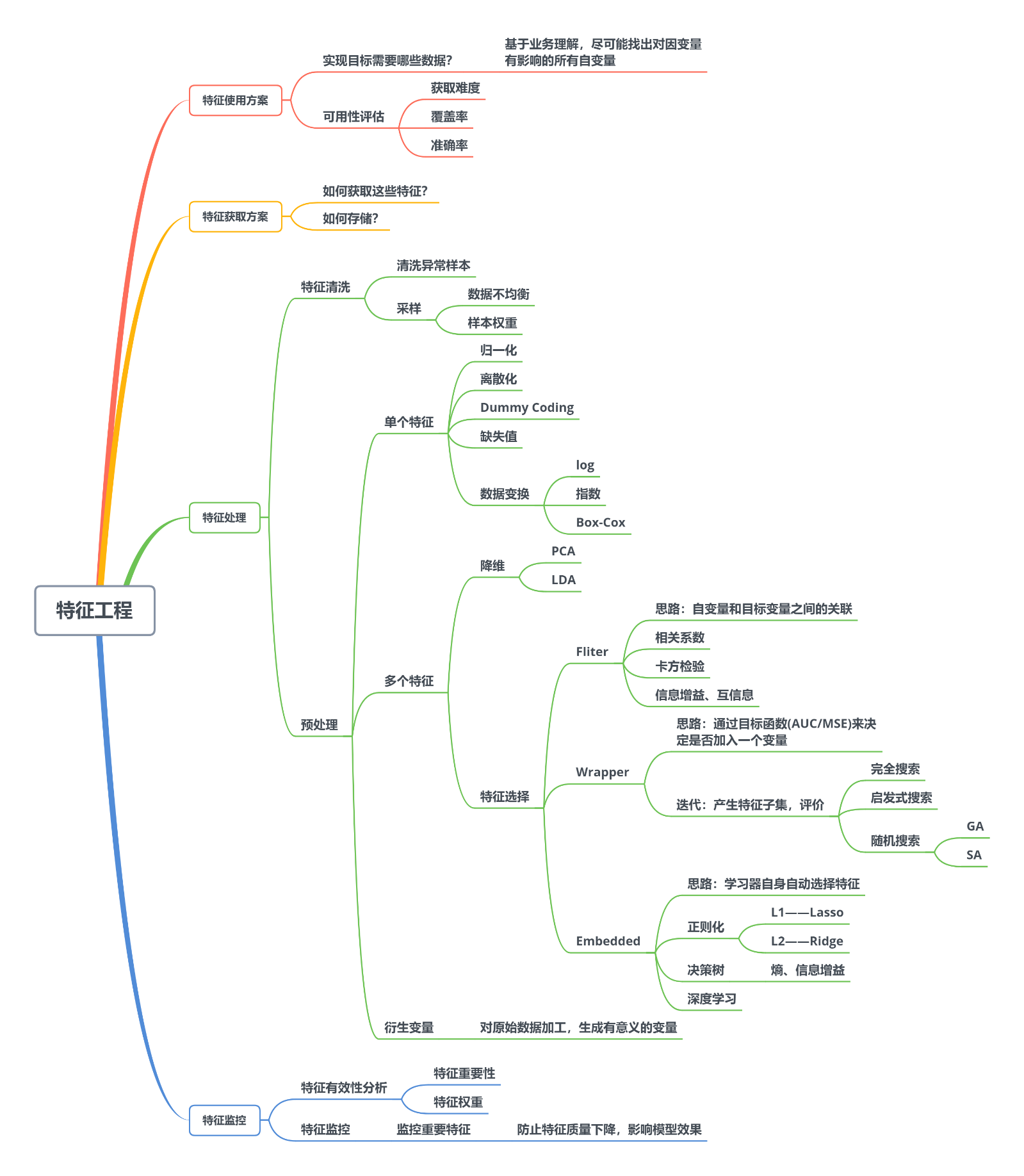
## 信号方面

## 信号自相关

信号ｘｔ的自相关函数Ｒｘ（τ）反映了信号在不同时刻ｔ１和ｔ２取值的相关程度，常用的是归一化自相关函数：

# 特征工程

特征处理是特征工程的核心部分



## 日期处理

### 日期转换to\_datetime(dateString, format)

### 日期格式化

### 日期抽取data\_dt.dt.properity

## 数据清洗

### 重复值处理duplicated

df.duplicated(self, subset = None, keep = ‘first’)

subset列标签的意思

### 缺失值处理

#### 删除多余维度

x = np.array([[[0], [1], [2]]])

y = np.squeeze(x) # 从数组的形状中删除单维条目，即把shape中为1的维度去掉

#### 矩阵转置

1. 数组转置data.T
2. 维度重置data.transpose(2,1,0)
3. 量轴对换data.swapaxes(0,2)

#### 缺失值识别isnull

df.isnull()

df.notnull()

#### 缺失值删除行dropna

df.dropna()

#### 缺失值填充fillna

df.fillna(‘???’)

df.fillna(method = ‘pad’)#用前一个列数据替代

df.fillna(method = ‘nfill’) #用后一个列数据替代

df.fillna(df.mean()) #用均值替代

df.fillna(df.mean(‘被填补的列名’:’计算均值的列名’)) #用均值替代

#### 清除字符串首尾指定字符，默认空格strip

df[‘name’].str.rstrip()

df[‘name’].str.lstrip(‘n’)

### 数据抽取

#### 选取列

data['w'] #选择表格中的'w'列，使用类字典属性,返回的是Series类型

data.w  #选择表格中的'w'列，使用点属性,返回的是Series类型

data[['w']] #选择表格中的'w'列，返回的是DataFrame属性

data[['w','z']] #选择表格中的'w'、'z'列

data[0:2] #返回第1行到第2行的所有行，前闭后开，包括前不包括后

data[1:2] #返回第2行，从0计，返回的是单行，通过有前后值的索引形式，

    #如果采用data[1]则报错

data.ix[1:2] #返回第2行的第三种方法，返回的是DataFrame，跟data[1:2]同

data['a':'b'] #利用index值进行切片，返回的是\*\*前闭后闭\*\*的DataFrame,

    #即末端是包含的

data.irow(0)  #取data的第一行

data.icol(0)  #取data的第一列

data.head() #返回data的前几行数据，默认为前五行，需要前十行则dta.head(10)

data.tail() #返回data的后几行数据，默认为后五行，需要后十行则data.tail(10)

ser.iget\_value(0) #选取ser序列中的第一个

ser.iget\_value(-1) #选取ser序列中的最后一个，这种轴索引包含索引器的series不能采用ser[-1]去获取最后一个，这回引起歧义。

data.iloc[-1]  #选取DataFrame最后一行，返回的是Series

data.iloc[-1:]  #选取DataFrame最后一行，返回的是DataFrame

data.loc['a',['w','x']]  #返回‘a'行'w'、'x'列，这种用于选取行索引列索引已知

data.iat[1,1]  #选取第二行第二列，用于已知行、列位置的选取。

#### 字段抽取slice(start, stop)

#### 字段拆分split(sep, n, expand = False)

#### 重置索引df.set\_index(‘列名’)

|  |  |
| --- | --- |
| loc | 通过索引抽取行数据 |
| iloc | 通过索引号抽取行数据 |
| ix | loc和iloc的混合 |

#### 记录抽取

根据一定条件进行抽取df[condition]

#### 随机抽样

numpy.random.randint(start, end, num)

#### 通过索引抽取数据

df.loc[行标签, 列标签]

df.iloc[行索引号, 列索引号]

ix

### 插入记录

#### 修改记录

#### 整体替换

#### 单值替换

#### 指定列单值替换

#### 多值替换

### 交换行或列df.reindex

### 排名索引

#### 重新排序df.sort\_index

#### 重新索引df.reindex

#### 重置索引df.set\_index

#### 索引还原df.reser\_index

### 数据合并

#### 记录合并

两个结构相同的数据框合并成一个数据框

pd.concat([DataFrame1, DataFrame2], …)

np.histack((a, b))#水平组合

np.vstack((a, b))#垂直组合

np.concatenate((a, b), axis = 1)

深度组合np.dstack((a, b))

列组合column\_stack((a, b))#对一维数组按列方向组合，对二维数组和hstack效果相同

行组合row\_stack((a, b))

#### 字段合并df[‘new’] = pd.DataFrame({‘A’: [1, 2, 3]})

df[‘序列名’]=序列值

同一个数据框中不同列合并

#### 字段匹配

## 数据预处理

|  |  |  |
| --- | --- | --- |
| **类** | **功能** | **说明** |
| StandardScaler | 无量纲化 | 标准化，基于特征矩阵的列，将特征值转换至服从标准正态分布 |
| MinMaxScaler | 无量纲化 | 区间缩放，基于最大最小值，将特征值转换到[0, 1]区间上 |
| Normalizer | 归一化 | 基于特征矩阵的行，将样本向量转换为“单位向量” |
| Binarizer | 二值化 | 基于给定阈值，将定量特征按阈值划分 |
| OneHotEncoder | 哑编码 | 将定性数据编码为定量数据 |
| Imputer | 缺失值计算 | 计算缺失值，缺失值可填充为均值等 |
| PolynomialFeatures | 多项式数据转换 | 多项式数据转换 |
| FunctionTransformer | 自定义单元数据转换 | 使用单变元的函数来转换数据 |

### 无量纲化——特征缩放

同比例缩放所有属性常用的两种方法是：归一化和标准化

标准化是依照特征矩阵的列处理数据，其通过求z-score的方法，将样本的特征值转换到同一量纲下。归一化是依照特征矩阵的行处理数据，其目的在于样本向量在点乘运算或其他核函数计算相似性时，拥有统一的标准，也就是说都转化为“单位向量”，以解决数据指标之间的可比性问题。

#### min-max归一化MinMaxScaler Normalizer



将值减去最小值并处以最大最小值的差。

Scikit-Learn提供了MinMaxScaler的转换器

from sklearn.preprocessing import MinMaxScaler

#区间缩放，返回值为缩放到[0, 1]区间的数据

MinMaxScaler().fit\_transform(iris.data)

规则为l2的归一化公式如下：



from sklearn.preprocessing import Normalizer

#归一化，返回值为归一化后的数据

Normalizer().fit\_transform(iris.data)

#### Z-Score标准化StandadCscler



减去平均值（所以标准化值的均值总为0），然后除以方差，使结果的分布具备单位方差，

标准化的方法受异常值的影响小

Scikit-Learn提供了StandadCscler的转换器。其可以保存训练集中的参数（均值、标准差）

from sklearn.preprocessing import StandardScaler

StandardScaler().fit\_transform(data)

### 二值化

定量特征二值化的核心在于设定一个阈值，大于阈值的赋值为1，小于等于阈值的赋值为0

### 对定性特征哑编码

### 缺失值计算

## 特征选择

当数据预处理完成后，我们需要选择有意义的特征输入机器学习的算法和模型进行训练。通常来说，从两个方面考虑来选择特征：

* 特征是否发散：如果一个特征不发散，例如方差接近于0，也就是说样本在这个特征上基本上没有差异，这个特征对于样本的区分并没有什么用。
* 特征与目标的相关性：这点比较显见，与目标相关性高的特征，应当优选选择。除方差法外，本文介绍的其他方法均从相关性考虑。

　　根据特征选择的形式又可以将特征选择方法分为3种：

1. Filter：过滤法，按照发散性或者相关性对各个特征进行评分，设定阈值或者待选择阈值的个数，选择特征。
2. Wrapper：包装法，根据目标函数（通常是预测效果评分），每次选择若干特征，或者排除若干特征。
3. Embedded：嵌入法，先使用某些机器学习的算法和模型进行训练，得到各个特征的权值系数，根据系数从大到小选择特征。类似于Filter方法，但是是通过训练来确定特征的优劣。

我们使用sklearn中的feature\_selection库来进行特征选择。

#### 特征选取

声发射特征选取

|  |  |  |  |
| --- | --- | --- | --- |
|  | **评价指标** | **计算公式** | **评价的意义** |
| 时域特征 | 均值 |  | 描述信号的稳定分量，又称直流分量  机械振动的平衡点位置  信号的一届矩统计平均 |
|  | **均方根值** |  | 信号幅值在时间上的平均值，能够反映在时间域上膜厚幅值的变化曲线  机械故障诊断中判别运转状态是否正常的重要指标  描述信号的能量 |
|  | 峰态系数 |  | 可以反映该时刻信号与其平均值的差异大小 |
|  | 峰值 |  | 波形的最大单峰值，稳定性差，对有磨损又有表面损伤的状态比较敏感 |
|  | 峰值因子  峰值指标 |  | 反映零件磨损或损伤  检测信号中是否存在冲击的统计指标（无量纲相对值） |
|  | 脉冲因子 |  | 峰值与信号均值的比值，  检测信号中是否存在冲击的统计指标（无量纲相对值）  Xp的稳定性不好，对冲击的敏感度较差，故障诊断系统中应用逐渐减少，被峭度指标所取代 |
|  | 波形因子 |  | 时间周期上均方根值与平均值之比 |
|  | 振铃计数 |  | 反映信号强度和频度 |
|  | 事件 |  | 反映信号发射现象的次数 |
|  | 峭度系数 |  | 表示信号振幅的规则性，当过大时则意味着有故障形成的大幅值脉冲  反映型号的冲击特性  正常值在3左右，如果接近或超过4则存在冲击性振动 |
|  | 裕度系数 |  | 部件磨损情况评价，衡量信号中冲击的程度 |
|  | **偏态系数/歪度** |  | 反映了信号相对中心的非对称性  信号的三阶矩统计平均  振动波性不对称，则歪度指标增大 |
| 频域特征 | **重心频率** |  | 描述功率谱重心位置的变化，反映各频率成分的能量比变化。切削过程中某些频率的振动幅值会因为刀具磨损状态而产生很大变化 |
|  | **均方频率** |  | 显示功率谱主频带的位置变化，描述功率谱能量分布的分散程度，反映信号频率成分的多少 |
|  | **频率方差** |  | 描述功率谱能量分布与功率谱重心位置之间的关系，衡量信号能量的分散程度 |
|  | 有效值电压RMS |  | 与信号大小有关，不受门槛影响，适用于连续型信号 |
|  | 信号能量 |  | 反映事件的相对能量或强度，能显示各频率域信号的能量总和。  信号检波包络线下的面积 |

### Filter

#### 方差选择法

#### 相关系数法

#### 卡方检验

#### 互信息法

### Wrapper

#### 递归特征消除法

### Embedded

#### 基于惩罚项的特征选择法

#### 基于树模型的特征选择法

## 降维

维度的诅咒

数据降维会丢失一些信息，降低系统性能，让流水线复杂，维护难度上升。

数据降维可能会滤除不必要的噪声和细节，导致性能更好（通常不会，只会加速训练）

数据降维有助于数据可视化。

高维数据集很大可能是非常稀疏的，训练集维度越高，过度拟合风险越大。

数据降维主要方法：投影、流形学习

决策边界不总是维度越低越简单。

### 相关性分析.corr

DataFrame.corr(data, *method ='pearson'*，'kendall'，'spearman')

部分变量之间存在较强的线性关系，如xx和yy的相关系数为zzz，这说明我们可以对数据进行变量压缩。

连续变量通常使用Pearson相关系数，有序变量通常使用Spearman相关系数，Kendall相关系数主要用于探索两连续变量之间的非线性相关关系。

### 主成分分析PCA

pca\_IMF0\_new.explained\_variance\_ratio\_

>>array([0.68142319, 0.13481074])

可见仅用2个主成分就能解释原始变量的87%的变异，因此保留两个主成分

pca\_IMF0\_new.components\_

每个主成分在原始变量上的权重

### 自编码器

# 机器学习

表 1机器学习算法列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **算法** | **学习类型** | **类** | **限定偏置** | **优选偏置** |
| K-最近邻 | 有监督学习 | 基于实例的 | 适合度量基于距离的逼近，高维表现不佳 | 适合基于距离的问题 |
| 朴素贝叶斯 | 有监督学习 | 基于概率的 | 适用输入数据互相独立的问题 | 适用于每个类别的概率恒大于0的问题 |
| 决策树/随机森林 | 有监督学习 | 类 | 对协方差低的问题不适用 | 适合分类数据问题 |
| 支持向量机 | 有监督学习 | 决策边界 | 两种类别有明确边界的情况下有用 | 适合二进制分类问题 |
| 神经网络 | 有监督学习 | 非线性函数逼近 | 几乎没有限制倾向 | 适合二进制输入 |
| 隐马尔可夫模型 | 有/无 | 马尔科夫过程的 | 对马尔科夫假设成立的系统信息都比较有效 | 适合时间序列数据和无记忆性的信息 |
| 群集 | 无监督学习 | 群集 | 无限制 | 数据可分类且类别之间可用某种距离描述时适用 |
| 特征选择 | 无监督学习 | 矩阵分解 | 无限制 | 视具体算法而定，可能更适合有大量共有信息的数据 |
| 特征转换 | 无监督学习 | 矩阵分解 | 必须是非退化的矩阵 | 对于没有倒置问题的矩阵，效果好很多 |
| Bagging算法 | 元启发式算法 | 元启发式算法 | 对任意类型数据适用 | 更适用于变化量不大的数据 |

## 分类

## 训练模型

## SVM支持向量机

SVM分类器可视为在类别之间你和可能的最宽街道，也叫作大间隔分类。

与Logic回归分类器不同，SVM分类器不会输出每个类别的概率。

软间隔分类、硬间隔分类

C值越小，街道间隔越宽，违例会越多

SVM对特征缩放非常敏感，训练集不经缩放，SVM将趋于忽略值较小的特征。

训练非线性数据集时，记得使用核函数。

支持向量机训练完成后，处于“街道”上的实例成为支持向量，包括处于边界上的实例。决策边界完全由支持向量决定。街道之外的实例完全没有影响，计算预测结果只会涉及支持向量，而不涉及整个训练集。

## 决策树

需要的数据准备工作非常小

完全不需要特征缩放或集中

决策树是非常直观的，它的决策也容易解释，属于白盒模型。相反的随机森林和神经网络是黑盒模型。

无约束的决策树深度通常为log2（n）

数据集小于数千个实例时，预处理训练集可以加速训练。否则设置presort = True会减慢训练

## 集成学习和随机森林

#### 投票分类器

采用多个分类器

预测器尽可能互相独立时，集成方法的效果最优。使用不同算法训练会增加它们犯不同类型错误的机会，提升准确率。

在相同数据集上训练可能犯相同的错误。

硬投票法<软投票法

#### Bagging和pasting

使用相同的算法在不同训练集随机子集上进行。

样本放回叫bagging（bootstrap aggregating），样本不放回叫pasting。

每个预测器单独的偏差都高于原始训练集上的偏差，但是通过聚合，同时降低了偏差和方差。

包外评估：未被采样的训练实例称为包外实例

Bagging Classifier支持对特征抽样

Random Patches对训练实例和特征都抽样

#### 随机森林

分裂节点时不再搜索最好的特征，而是在随机生成的特征子集中搜索最好的特征。用更大的多样性，用更高的偏差换取更低的方差，产生整体更优的模型。

#### 极端随机数

对每个特征采用随机阈值，而不是搜索得出的最佳阈值，可能让决策树生长得更随机。

重要特征更可能出现在靠近根节点的位置。

## 朴素贝叶斯分类

## HMM隐马尔可夫模型

擅长使用观察结果找出给定系统的隐藏状态，以及做出这些状态如何工作的假设。

适合用于为分类添加本地化的上下文。

可预测随时间变化的模型状态。

# 神经网络和深度学习

## TensorFlow

## 人工神经网络

## 深度神经网络

## 跨设备和服务器的分布式TensorFlow

## 卷积神经网络

## 循环神经网络

## 自动编码器

## 强化学习

# 基于多传感器的实验方案

刀具磨损状态变化过程往往比较复杂和多变，具体的试验条件和加工环境也会对刀具的磨损状态变化产生较大影响，所以在试验中需要通过分析具体情况选择并设计一个合适的信号监测方法。而且，不同类型的信号反映出的信息有一定的差异，单一的信号所反映出来的信息往往比较片面而且存在较大误差。所以在试验中拟采取多传感器信号监测以及信息融合的方法，以提高刀具状态监测系统的监测能力、准确性和可靠性。

## 监测信号的选择

本文根据铣削加工的具体特点，主要分析了切削力信号、振动信号以及声发射信号等典型直观的信号与刀具磨损状态之间的关联，研究分析并设计了以上述几种信号为间接识别依据的刀具磨损状态监测系统，为后续试验设计研究打下坚实基础。

### 切削力信号

切削力是指在金属切削时刀具使加工材料变形成为切屑所需的力。切削力来自切削层金属、切屑和工件表面层金属的弹性、塑性变形所产生的抗力,以及刀具与切屑、工件表面间的摩擦阻力，广大研究工作者认为切削力可以最直接地反映刀具磨损状况，是能够反映加工状态最直接的因素，首先切削力和振动信号测量简单，易于在线监测，其次，在状态监测中相对于其他传感信号比较特殊，由于其信号直接来源于切削工作区，因此灵敏度更高，干扰信号更小，响应也更快。

Ertunc.H.M 等人通过研究显示了钻削加工中的进给力与刀具状态的直接关系，结果是磨损刀具至少超过正常刀具 3 倍以上,而且因磨损量成一定相关变化。Li.G.S 等人发现在刀具的严重磨损的情况下会使切削力振幅提高 50%-100%。

因此可以总结出,切削力是刀具磨损中表现优秀的监测信号。

在车削加工中,切削力主要由主切削力Fc,进给力Ff和径向力Fp三个分力组成,如图所示。

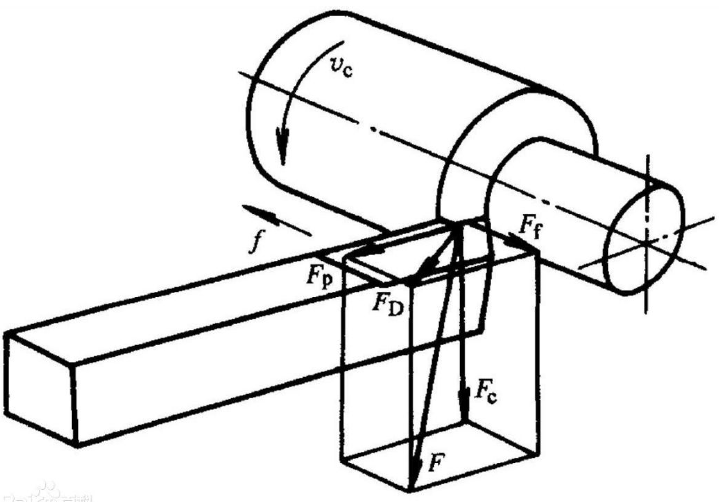


图 1车削加工的切削力

切削力信号是由安装于刀柄上的测力计传感器取得，由于实际切削条件和环境对切削力信号的影响较大，Thangaraj 等人为了尽可能消除切削条件和环境对切削力信号的影响，采用了进给力的变化率来反映刀具磨损状态并取得较好成果。

### 振动信号

同样的，振动信号测量简单，易于在线监测，采用加速度传感器即可进行检测，其安装简单易于操作和实现，也能够直接地反映刀具磨损状况的信号，由于其信号直接来源于切削工作区，因此灵敏度更高，干扰信号更小，响应也更快。

振动是在机械加工过程中，因多种原因机床、工件和刀具会产生周期性的跳动。切削加工中产生的振动信号的原因多种多样而且非常复杂，在此不一一进行分析，但是另一方面往往在振动信号里也包含着与切削状态相关的大量信息可以进行利用，研究表明振动信号在刀具状态监测中有重要作用,而且该信号的监测方法能够有效运用于各种切削加工场合中去。

### 声发射信号

声发射(AE)信号是某种物质在受到外力作用造成自身相对变形、断裂时释放的瞬态弹性应力波。在本文涉及的切削过程中,AE信号主要有连续型和突变型两种时域形态。切削表面和切屑产生的塑性形变,车刀后刀面与加工工件表面的摩擦以及前刀面与切屑之间的摩擦产生的主要是连续型AE信号,而突变型 AE 信号则一般是切削刀具破损断裂或者切屑断裂产生的。声发射信号频率较高，一般在 100kHz-1MHz 范围内，也是切削区直接产生的,声发射的频率远远超过切削过程中频率较低的振动信号以及环境噪声信号，因此切削条件以及刀具参数对声发射信号的影响较低,其具有抗干扰能力强,灵敏度极高的优点，能够充分反映切削刀具结构损伤和内部缺陷变化情况。

由于 AE 信号的传播随距离的衰减很大,为了能够充分获得信号中的信息就需要把声发射传感器安装在距离切削刀具和加工工件尽可能近的位置。在车削加工中,可以把AE 传感器安装在刀具后端较近处,但是在铣削、钻削和加工中心这些具有动态旋转刀具的应用设备中则不可以直接安装于切削刀具后端,而需要通过某些方法把 AE 信号传递到静止的传感器上，例如 Dolinsek Slavko 等人使用润滑冷却液传导声发射信号，然后把声发射传感器安装于润滑液管上,使声发射信号顺着润滑冷却液最终传播到安装好的采集 AE 信号的传感器上，而且因为液体只可以传播纵波,因此信号通过液体的传播后的信号失真更低、灵敏度更高。

## 实验平台搭建

### 切削力信号采集系统

试验中选择的测力仪型号为 Kistler9257B 如图 3-2 所示，该测力仪由瑞士奇石乐公司生产制造，每个传感器拥有三个石英板用来测量三个正交分量的力信号,具有很高的刚度以及固有频率，能够在信号的动态变化中满足最高的分辨率，表面集成了特殊隔热涂层,而且具有防水和冷却液的功能，可测力信号范围为-5kN-5kN。此型号测力仪可以同时实时、可靠、精确得完成切削过程中主切削力 Fc、进给力 Ff 和径向力 Fp三个方向力信号的采集工作。



图 2Kistler Type9257B 测力仪

并且在此基础上配套使用电荷放大器 Kistler5073A，如图 3-3 所示，该仪器广泛应用于工业控制过程中，能够同时完成最多四传感器的电荷放大，把输入量电荷信号呈正比地有效转换为电压或者电流信号，可输出范围为-10V-10V。



图 3电荷放大器 Kistler5073A

### 振动信号采集系统

试验中选择 8702B K-Shear 石英加速度计监测切削过程中刀具的振动信号，其灵敏度 50mV/g,频率范围 1Hz-10kHz,测量范围为±50g[32]。如图 3-4 所示，其功能是在监测到机械振动信号后转化输出一个和振动加速度成正比的电信号，三角剪切形式能够很大程度上降低温度等环境因素对其影响。可通过内置电荷放大器以及外接电荷放大

器的方式完成振动过程中加速度、速度、位移以及冲击信号的测量。因为其 90kHz 谐振频率确保准确监测高度重复冲击和振动信号，而且具有体积小巧、质量轻便、安装方便、设计坚固、测量范围广和抗干扰能力强等优点[33]

，因此在振动测量领域广泛应用。试验中 K-Shear 石英加速度计平行于主切削力 Fc 得安装在刀座上。



图 48702B K-Shear 石英加速度计

### 数字采集系统

试验中工控机采用 DEWETRON 公司的基于 PC 的虚拟仪表 DEWE 数字采集系统，该系统为方便多种类传感器的输入内置了插入式 8 到 80 路多通道信号调理插件以及 AD 卡，单通道采样频率为 100kHz-50MHz[35]。试验中采用了采样频率 100kHz，16 通道的，利伟创工控机如图 3-5 所示,而且具有 USB, IEEE-1394 (˜Firewire˜) 和 PCI-PCI/PCM-CIA等 3 种接口[36]，仅要把 DEWE 与 PC 连接就能够构成完整的数据采集系统，自带数据采集软件 DEDESoft 可以实现采集数据的设定与配置、在线显示、硬盘存储、多格式输出和报告打印等多种功能。



图 5DEWE-3021 数字采集系统

### 声发射信号采集系统

### 传感器安装位置选择

1. 力信号传感器的安装位置对力信号均值方差等主要特征影响比较小。
2. 加速度传感器信号的强弱与加工件距离有关，加速度传感器的安装位置越远整体信号等比例减弱,特征值也相对越小；
3. 声发射信号对刀具磨损最为敏感，理论和试验都显示了声发射传感器安装在车刀后刀面方向最为合适[39]

综合探索性试验得出合适的传感器安装位置如图 3-7 所示，试验中力信号的采集需要用测力仪通过刀架间接测量获得，试验中需要先将数控车床原刀架卸后,把测力仪装配于基座和刀架之间。提取振动信号的加速度传感器则安装于刀架和靠近车削刀具的各个方向，以使采集振动信号的信息更加完备，声发射传感器则安装在车刀后刀面方向较近处最为合适

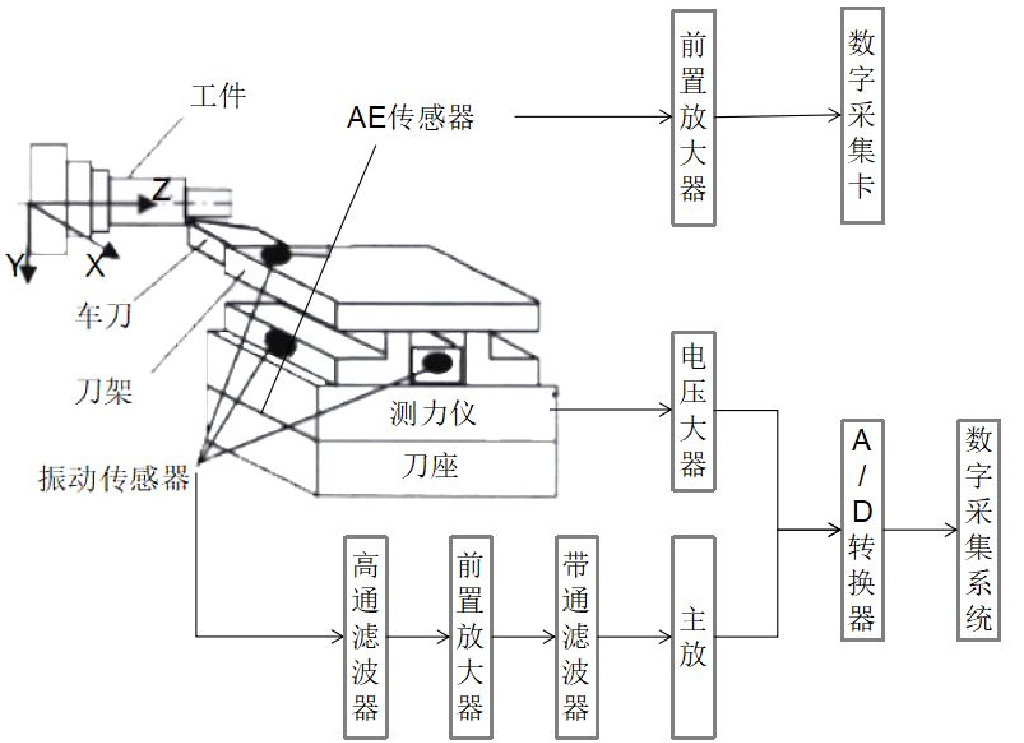


图 6信号采集原理图

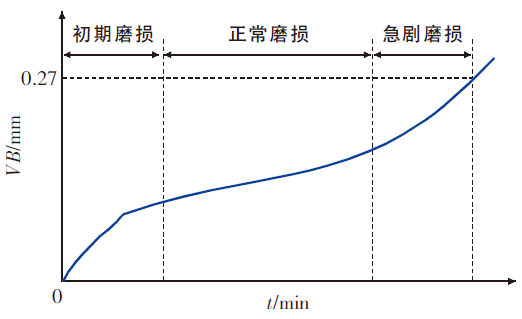
## 刀具磨损状态实验设计

### 切削三要素

|  |  |  |
| --- | --- | --- |
| νc /（m·min-1） | f/（mm·r-1） | ap/mm |
|  |  |  |

### 刀具磨损量

研究中进行了刀具全磨损实时监测试验，从新刀到失效的整个过程中信息量巨大，因此首先需要用科学的方法从中获取具有代表性的分析样本，将刀具从新刀到 ISO 磨钝标准的磨损量等分为表 3-3 中 6 个状态。



刀具各磨损状态的平均磨损量

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 状态 | 初期磨损1 | 初期磨损2 | 正常磨损3 | 正常磨损4 | 急剧磨损5 | 急剧磨损6 |
| VB/mm | 0-0.05 | 0.05-0.10 | 0.10-0.15 | 0.15-0.20 | 0.20-0.25 | 0.25-0.30 |

分别将这六种磨损状态依据 VB 值再等量的分成 5 个信号段，每个信号段中抽取 5 个由 2 万个连续采集的数据点。因此每组试验总共有 150 组样本。

试验样本数量选择表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 状态 | 初期磨损1 | 初期磨损2 | 正常磨损3 | 正常磨损4 | 急剧磨损5 | 急剧磨损6 |
| 抽取组数 | 25 | 25 | 25 | 25 | 25 | 25 |
| 总计组数 | 250 | | | | | |

### 试验设计方法

1. 对比试验方法
2. 正交试验方法
3. 均匀试验方法

# 振动

振动信号幅值围绕零值上下波动，其幅值随着刀具磨损量的增加而逐渐增大。然而，磨损量并不是影响信号能量的唯一因素。一般地，随着切削加工，工件逐渐变细，为了保证切削3 要素一致，须不断增加车床的主轴转速，由此也会给原信号的幅值造成影响。事实上，单从图中振动信号的幅值来看，前3 个状态几乎无法区分，因此难以从中直接看出刀具磨损程度的变化规律。类似地，切削力虽然在总体上随着刀具的磨损逐渐增大，但实验中因为更换工件，或其他因素的影响，测得的力信号有时也会出现不规则波动，因此也不能将之作为识别刀具磨损状态的直接依据。综上所述，须进一步挖掘信号中所蕴含的有用信息，削弱信号噪声的影响，提取信号特征，并找到特征与刀具磨损之间的内涵关系。

# 方案

## 特征选择

#### 主成分分析

传统的PCA 方法通过计算样本X 协方差矩阵A的特征值λ 和对应的特征向量v，计算其前r 个特征值的贡献度( 贡献度≥90%) ，求出样本在相应特征向量方向的投影结果作为新的样本，很好地将线性可分数据从高维降到低维。

但

#### 核主成分分析

### 马氏距离

马氏距离是衡量服从同一分布的且有着协方差矩阵的随机变量之间的差异程度。

### 机械密封寿命预测方法

针对传统声发射参数分析方法准确度较低、易受环境影响，模型预测准确度不高的问题，对机械密封声发射信号进行了如下处理:

首先对数据进行小波降噪处理，

再对降噪信号进行小波包分解，对各频带的子信号进行特征抽取，得到大量声发射特征，

再利用KPCA 对特征矩阵降维处理，得到贡献率高的少量特征，

再利用马氏距离进行特征融合，

最后利用QPSO算法优化SVＲ 支持向量回归进行寿命预测。

### 信号处理

#### 信号的特征提取

#### 寿命预测

## 信息融合

### 信息融合的分类

按照数据抽象的不同层次，融合的形式有像素级、特征级和决策级。

#### 像素级融合

预处理之前的原始数据基础的融合，这是最低层次的融合。融合后的数据拥有最多的信息量,这也往往造成计算处理时间成本高,实时性差等缺陷。

#### 特征级融合

属于中间层次，它是对原信号进行特征提取后进行综合分析和处理的方法。特征级融合可划分为两类：目标状态信息融合和目标特性融合。

1. 特征级目标状态信息融合主要用于多传感器目标跟踪领域。融合系统首先对传感器数据进行预处理以完成数据校准，然后主要实现参数相关和状态向量估计。
2. 特征级目标特性融合就是特征层联合识别，具体的融合方法仍是模式识别的相应技术，只是在融合前必须先对目标特征进行相关处理，把特征向量分类成有意义的组合。

#### 决策级融合

一种高层次融合，其结果为指挥控制决策提供依据。因此，决策级融合必须从具体决策问题的需求出发，充分利用特征融合所提取的测量对象的各类特征信息，采用适当的融合技术来实现。决策级融合是三级融合的最终结果，直接针对具体决策目标，融合结果直接影响决策水平。决策级融合具备更高传输速度和实时性，且数据处理量低，抗干扰能力强，但是决策级融合首先要对原传感器信息进行预处理以获得各自的判定结果，所以预处理代价高。

### 决策级融合

在多源信息融合技术中，信息融合算法的选择直接关系到融合效果的好坏，在实际运用中，决策级融合运用最为广泛。决策级融合主要方式和优缺点如下：

#### 物理模型法

物理模型法在基础研究中广泛运用，该方法通过所建立的模型精确得到可观察的或可计算的数据，用模型数据与实际数据比较，从而达到识别物体的目的。但因为其计算量大、实现比较困难等原因在实际系统中很少应用。

#### 统计学和信息论

不同于物理模型法，基于统计学和信息论的算法需要寻找一个建立在识别参数而非数学模型的方法。该方法通过参数和识别结果之间的变换、映射和采样分布得到决策误差。但其需要大量的先验知识。

#### 贝叶斯算法

贝叶斯算法准确率高速度快，其更新了先前所给假设的似然函数，解决了一些推理方法不能解决的问题，也需要以先验知识为基础，但贝叶斯算法需要属性之间相互独立，这在很多情况下是不成立的，而且其无法区分“不确定”和“不知道”两种情况。

#### D-S 证据理论

D-S 证据理论对贝叶斯理论进行了重大推广，使其能够处理大量的不确定系统估计问题，可解决满足比概率论更弱的条件，采用概率空间来描述假设的似然度，用区间估计未知信息或信息不全的造成的不确定性，能够区分贝叶斯算法所不能区分的“不确定”和“不知道”两种情况。但 D-S 理论没有完整严密得证据证明出其可靠性。另一方面，D-S 证据论需要样本证据具有独立性，而实际应用中遇到很多场合的独立证据是非常有限的，因此在实际的多数情况下由于条件的限制也使得 D-S 证据论无法广泛应用和推广。

#### 模糊推理

模糊推理理论是 L.A.Zadeh 提出的，通过综合规则进行演绎推理，并在推理中使用模糊概率，最终建立了模糊逻辑，通过模糊化的方法可以给出证据函数的表示以及在限定条件内的信任函数修正模型。虽然可以解决证据论的一些不足，但也存在很多不足：一方面在模糊系统中，专家经验总结的主观性太强，增加了对专家知识的依赖，另一方面运算量庞大，当模糊规则较多时难以实施。

#### 支持向量机

它基本上不涉及概率测度及大数定律等,因此不同于现有的统计方法。从本质上看, 它避开了从归纳到演绎的传统过程,实现了高效的从训练样本到预报样本的“转导推理”,大大简化了通常的分类和回归等问题；SVM 的最终决策函数只由少数的支持向量所确定,计算的复杂性取决于支持向量的数目,而不是样本空间的维数,这在某种意义上避免了“维数灾难”。少数支持向量决定了最终结果,这不但可以帮助我们抓住关键样本、“剔除”大量冗余样本,而且注定了该方法不但算法简单,而且具有较好的“鲁棒”性。

## 基于 SVM 的信息融合

### SVM 融合原理

分类是数据分析中的重要任务之一，其目标为建立分类函数或者模型，而 SVM 就是一类广泛使用于统计学习理论和回归分析中的机器学习方法。SVM 从上世纪 90 年代发展至今已逐渐成熟，其为了增强泛化能力，需要寻求最小结构风险和置信范围，通过 VC 维理论衡量结构风险，机器学习的目标是降低经验风险，要降低经验风险，就要提高分类函数的复杂度，导致 VC 维很高，置信风险相对就高，所以，结构风险也高，这是 SVM 比其他机器学习具有优势的地方。而且通过核函数的引用成功避免了“维数灾难”的问题，最终能够在样本数量非常少的条件下实现优秀的统计规律。

## 刀具状态预测技术

金属切削加工所加工产品的尺寸、形状、位置精度、表面形貌等都与刀具有关，刀具的性能、质量及其管理直接影响到能否顺利加工出所需要的合格产品，影响到加工节拍和生产效率，影响制造成本的控制和降低，关系到制造型企业的核心竞争力，掌握刀具使用寿命分布，对生产有指导意义。在现代化加工中，对刀具使用寿命需进

行科学管理，定时换刀，必须了解刀具使用寿命的分布情况。

在实际生产加工过程中，大部分企业严格按照加工时间和加工件数量进行切削刀具的更换依据，这么做虽然在一定程度上能够使产品质量和生产安全得到保障，可是另一方面也会导致很多刀具还没达到磨削标准便被更换抛弃的现象发生，同时也造成了大量刀具因提前被报废产生经济损失的现象。近年来的研究成果已经可以对切削刀具进行实时的状态监测，但假如可以通过更加科学有效的方法预测出刀具未来一定时间内所处的状态，便能够更加有效得避免刀具的提前被报废现象发生，减少了生产加工中的经济损失，预知刀具切削状态变化所产生的的影响。

### 刀具磨损状态预测方法的分类



图 7车削刀具预测

#### 基于数据的方法

此类预测方法适用领域广、使用成本低，需要从工作现场的数据中直接寻找出蕴藏的信息，但是拥有广阔的发展前景，所以该方法是故障诊断和预测领域的主要发展方向。

#### 基于模型的方法

此类方法能够在一定程度上揭示研究对象的本质，通过模型的建立能够较好地预测对象变化趋势，可是因为切削加工中切削环境通常较复杂，刀具磨损量和磨损状态也具有一定随机性，以及加工条件的多变性等多种原因，导致通常无法建立准确合适的模型。

#### 基于知识的方法

此类方法的不需要精确得数学模型，只需要利用专家知识和先验经验就可以预测研究对象的发展趋势，但是专家知识和先验经验总是较难获得的。

# EEMD降噪---BP&Elman模式识别----灰色隐马尔可夫预测模型预测

## EEMD

以一组冷切数据为例,对原始信号进 M-EEMD 分解并进行特征提取的说明：

(1)引入方差均为 0.5 的均值为 0 的一对白噪声。

(2)对原振动信号进行 FFT 变换和分解后 IMF 使用 FFT 变换,根据公式 4-17 得出两者能量值；

该信号 M-EEMD 分解后的前五个 IMF 信号分量如图 4-6 所示：

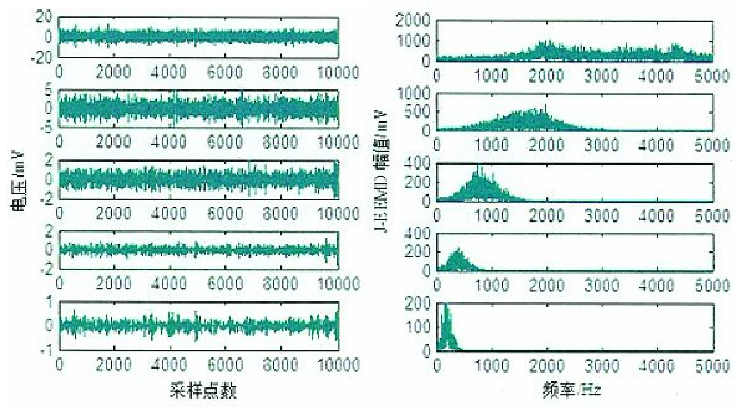


图 8M-EEMD 分解后的 IMF 分量

原振动信号进行 M-EEMD 分解后的 IMF 的频域分布差异很大。图中可以看出，IMF1 的能量主要集中在 2k-4.5kHz,IMF2 的能量主要集中在 1k-2kHz,IMF3的能量主要集中在 0.5-lkHz。而且随着 MEEMD 的不断分解,其分量能量频率集中范围越低越窄。

通过观察和能量计算结果显示,前三个 IMF 集中了约 80%以上的信号能量,也间接显示了前三个 IMF 分量具有原信号大部分信息,所以通过 MEEMD 计算分析可知，仅选取前三个 IMF 分量并对其提取出信号特征便可获得信号中大部分信息，从而在一定程度上提高了特征提取效率、简化了步骤、降低了计算成本。

## 基于 BP 和 Elman 神经网络的模式识别

从前文 150 组样本中每个磨损阶段抽取 1 组样本，最终得到 50 组数据的训练样本,并通过前文中关于信号特征筛选的分析研究最终分别选取振动信号以及 AE 信号的：均方根值、歪度指标、重心频率、频率方差、均方频率以及经过 MEEMD 分解后的前三个IMF 分量中的重心频率，最终归一化后得到如图 4-1 所示的训练样本，其中特征 1-8是振动信号特征，特征 9-16 是声发射信号特征。

把上述 16 个特征作为神经网络的输入训练样本,根据神经网络经验公式取隐含层节点数 25,学习速率取 0.01,训练误差取 0.0001,利用样本分别对 BP 神经网络和 Elman神经网络进行训练,网络输出为切削过程中的六种状态，然后按照 50 个训练样本取出另外 50 个不同的样本进行神经网络识别率检测，最终结果如表 4-2 所示。

表中显示，ELman 神经网络和 BP 神经网络都具备优秀的识别效果，但是在通过样本的训练过程中步数平均在 1000 步以上,时间成本和资料成本消耗巨大。扩大训练网络的训练误差确实能够减少很多训练时间,但是同时也会降低识别率和鲁棒性。

## 小结

在本章中,一种适用于车削信号特征提取预处理的改进型总体经验模态分解(Modified-EEMD)算法被研究并提出。此方法是基于信号固有特征,将目标信号分解成若干蕴涵模态函数(IMF),该方法区别于传统 EMD和EEMD方法在于通过对原信号添加一正一负两组均值为零的白噪声信号，从而减少添加白噪声的干扰，使得 EMD 分解更加具有完备性，并通过不断检验 IMF 分量是否异常决定是否继续 EMD 分解，从而减少了计算量和工作量。随后利用了 BP 网络和 Elman 网络对振动和声发射特征进行网络训练并进行识别率验证。试验结果表明：在允许误差较低的情况下单独使用 BP 神经网络或者 Elman 神经网络进行模式识别能够达到相对较高的识别率,但这种情况下需要严格限制训练误差，而且与此同时对时间成本需求太大，所以，为了保证结果准确性的基础上减少时间成本,本文进行了接下来的研究工作。

# EEMD-SVM

## 概念

#### 集合经模态分解

（ensemble empirical mode decomposition，EEMD）算法，即在分解过程中引入高斯白噪声，当附加的白噪声均匀分布在整个时频空间时，原信号将会按照不同的尺度投影到相应的时频空间。考虑到白噪声的随机及零均值性，认为当附加的噪声次数足够多时，便可通过求取全体信号的均值来消除噪声的影响，从而得到理想的结果。

#### 支持向量机

算法步骤如下：

1）通过非线性变换将输入向量映射到高维特征空间。

2）在约束条件为求解使目标函数最大化的αop，其目标函数为

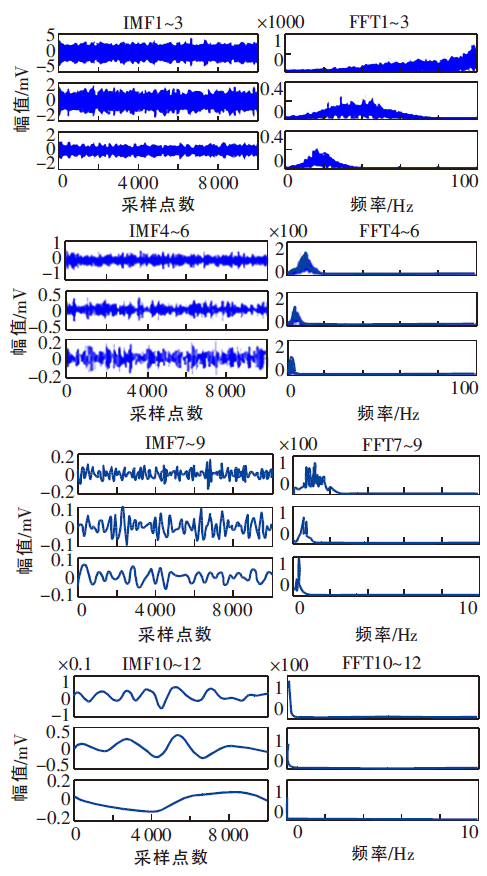
3）计算最优权值

4）计算分类判别函数

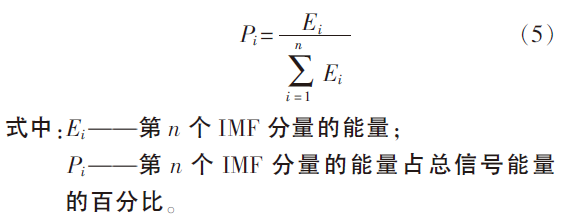
## EEMD-SVM 的刀具故障诊断

由于不同类型的信号所表征的状态信息不同，因此须对两种信号分别处理。对每把刀具，从每种状态中各随机选择10 个样本，其中每个样本包含连续采集的10 000 个振动信号，以及对应3 个方向上的6 000 个力信号（每个方向2 000 个采样点）。4种状态一共得到80 个样本，每个样本分别包含4 组信号。从总体上看，样本涵盖了刀具从新到旧各个时间段的状态信息，因此能很好地体现出刀具磨损的变化规律。

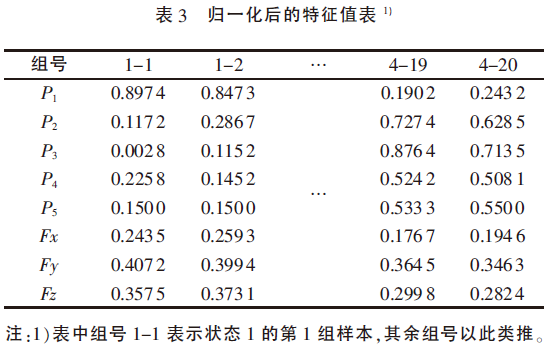
以状态2 为例， 使用EEMD 对振动信号进行分解，其中随机白噪声的标准差取0.4，NE 为EMD 分解次数取100，数据长度为10 000，按照公式计算，每组振动信号数据可以得到n 个有效IMF 分量和1 个残余分量，如图3 所示，为初期磨损阶段一组振动数据样本经EEMD 算法处理得到的12 个IMF 函数图及对应的FFT 频谱图。



图中，IMF1～3 表示第1～3 个IMF 分量函数图，FFT1～3 表示对应IMF 分量函数的FFT 图谱，其余编号依此类推。各IMF 分量的信号能量从高频到低频依次排列，且高频信号所占频带较宽，低频信号所占频带较窄，信号能量主要集中在前5 个IMF 分量中。由于刀具磨损程度不同，其振动信号的能量所占的频带也不同，且通过进一步比较发现，不同状态下，前5个IMF分量能量比值的数值变化差异较大。因此，选取振动信号的前5个IMF分量能量百分比来表征刀具的磨损特征，如下式所示：



此外，引入3 个轴向上的切削力信号的平均值Fx、Fy、Fz。这样，每个样本得到8 个特征，其中包含5 个振动特征，其归一化结果如表3 所示。



## 刀具的磨损状态识别