

## 第一周报告 - 于建国

1. 特征提取
2. 预处理
3. 模型选择
  - 3.1. 选择策略
  - 3.2. 浅层模型结果
    - 3.2.1. 单特征集
    - 3.2.1. 多特征集
4. 结论
5. 附：导师评语
  - 5.1. 导师1
  - 5.2. 导师2
  - 5.3. 评分

# 第一周报告 - 于建国

## 1. 特征提取

特征	维度
IS09_emotion.conf	384
IS10_paraling.conf	1582
IS13_ComParE.conf	6373
ComParE_2016.conf	6373

extractor.py

- 作用：批量提取特征
- 格式：.wav → .txt

data\_maker.py

- 作用：与标签对齐后转成 pandas.DataFrame
- 格式：.txt → .csv

## 2. 预处理

标准化:

- $x$  (特征):  $x = (x - x.mean) / x.std$
- $y$  (标签):  $y = (y - y.mean) / y.std$

## 3. 模型选择

### 3.1. 选择策略

对 4 组特征集分别进行 7 折交叉验证, 找出最好的特征和模型

### 3.2. 浅层模型结果

shallow.py

标准化后的结果 (没有反标准化)

#### 3.2.1. 单特征集

特征	算法	$mse$	$r^2$
IS09	Support Vector Regression	0.5666	0.4324
IS09	Gradient Boosting Decision Tree	0.5402	0.4586
IS09	AdaBoost	0.6396	0.3678
IS09	Ridge Regression	0.6698	0.3272
IS09	Bayesian Ridge Regression	0.6082	0.3901
IS10	Support Vector Regression	0.4821	0.5164
IS10	Gradient Boosting Decision Tree	0.4645	0.5340
IS10	AdaBoost	0.5586	0.4452
IS10	Ridge Regression	2.550	1.562
IS10	Bayesian Ridge Regression	0.5484	0.4491
IS13	Support Vector Regression	0.5060	0.4923
IS13	Gradient Boosting Decision Tree	<b>0.4591</b>	<b>0.5392</b>
IS13	AdaBoost	0.5422	0.4498
IS13	Ridge Regression	0.8717	0.1241

IS13	Bayesian Ridge Regression	0.8723	0.1235
IS16	Support Vector Regression	0.5060	0.4937
IS16	Gradient Boosting Decision Tree	0.4693	0.5292
IS16	AdaBoost	0.5390	0.4592
IS16	Ridge Regression	0.8786	0.1175
IS16	Bayesian Ridge Regression	0.8791	0.1170

### 3.2.1. 多特征集

特征	维度	<i>mse</i>	<i>r</i> <sup>2</sup>
IS10,13	Gradient Boosting Decision Tree	0.4484	0.5484
IS09,10,13	Gradient Boosting Decision Tree	0.4446	0.5517
IS09,10,13,16	Gradient Boosting Decision Tree	<b>0.4428</b>	<b>0.5534</b>

## 4. 结论

- 四个特征集全部使用效果最好
- 浅层模型 Gradient Boosting Decision Tree 效果最好
  - `num_leaves =15`
  - `learning_rate=0.02`
  - `n_estimators=400`
- 预测代码: `predict.py`
- 预测值为 GBDT 两次 ( `n_estimators=400` 和 `n_estimators=800` ) 的平均数

## 5. 附：导师评语

### 5.1. 导师1

- 项目完成情况很好，并对其过程进行了详细的说明。语音情感特征提取部分完成了 `IS09_emotion.conf`、`IS10_paraling.conf`、`IS13_ComParE.conf`、`ComParE_2016.conf` 四个特征提取工作。而且完成了批量的特征提取。
- 数据预处理部分，完成了特征和标签的预处理。
- 对 4 组特征集分别进行 7 折交叉验证，找出最好的特征和模型。尝试了机器学习所有的几乎所有的模型，并且用多个指标进行验证，背后具有很多的工作量。

- 于建国同学在短短一周之内做出了很多模型和特征的尝试工作，其工程量庞大且很好完成任务。值得每位同学学习。

## 5.2. 导师2

- 该学员的学习报告对特征提取，预处理，模型选择等模块进行详细说明。文中多处使用表格，条理非常清晰。
- 用表格列举并对比了不同的特征，并说明了特征提取的过程，简述了特征提取后的格式转换过程。
- 预处理部分用到了均值方差归一化的小策略。
- 模型部分，对不同的特征集进行了详细的实验和说明。分别在单特征集和多特征集上进行多次实验，并详细列举不同特征集的表现。
- 最后根据不同特征集和不同模型的表现选出最好的性能，效果比较理想。

## 5.3. 评分

