阿尔茨海默综合症预测挑战赛

团队名称：1767923
答辩人：吴绍武

2019年10月24日 • 合肥

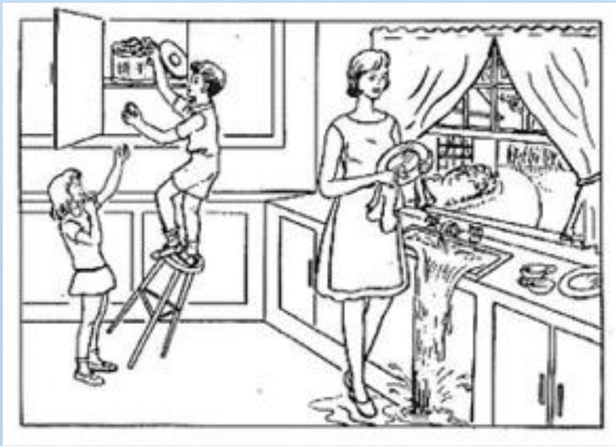
目录

问题简介

- 赛题目的：如何自动筛查出阿尔茨海默综合症患者
- 看图说话任务：取自波士顿失语症诊断。任务要求主试者先向被试展示指定图片，然后说“告诉我你在这幅图里看到的正在发生的一切”。允许主试在被试无法说出很多内容的时候鼓励被试。每个音频文件都先被采集然后人工转出文本。音频中出现不属于看图说话任务的对话没有被转写。

- 比赛数据：音频、文本、个人信息、整体统计量
- 数据大小：CTRL人数138，MCI人数179，AD人数84

其中：CTRL：健康
MCI：轻度认知障碍
AD：可能是阿尔茨海默综合症或其他种类的痴呆症



看图说话任务图

问题简介

- 分类问题：初赛是二分类，复赛是三分类
- 小样本：样本量小
- 评分标准：UAR(unweighted average recall，非加权平均召回率)

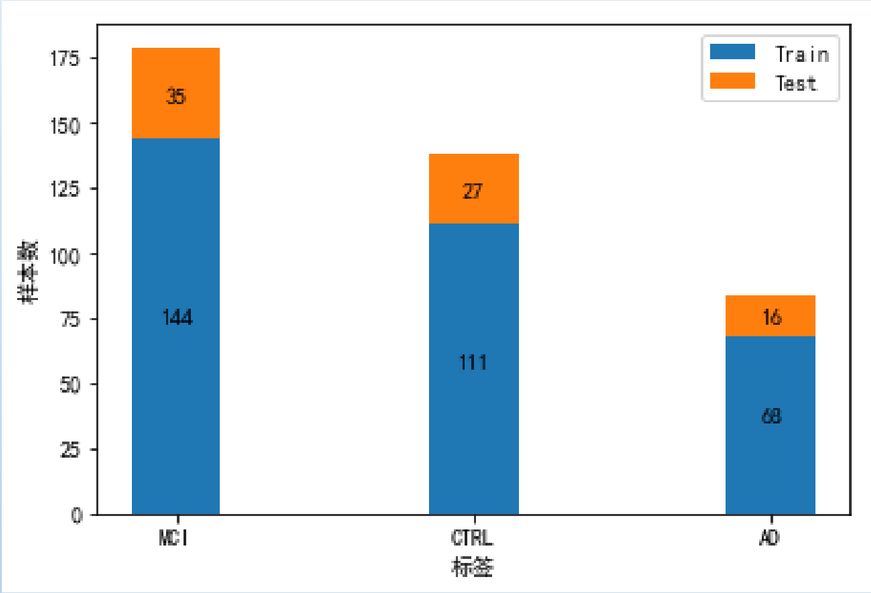
例如：对于3分类问题，各个类别的召回率分别为recall1，recall2和recall3，那么

$$UAR = (recall1 + recall2 + recall3)/3$$

分析：样本少的类别，比样本多的类别带来的收益要大！

特征工程

- 标签分布：
- 异常值处理：均值±3*标准差
- 特征提取：转写文本、音频（eGeMAPs特征，由openSMILE工具提取而得）



标签分布示意图

特征工程

一、转写文本：

1、说话时间

- 主试者提问时间；
- 被试者回答时间；
- 无人说话时间。

Index	no	start_time	end_time	speaker	value
0	1	0	0.02	sil	noise
1	2	0.02	4.77	<A>	图片上有哪些人他们在做什么，你…
2	3	4.77	5.05781	sil	noise
3	4	5.05781	10.05	<A>	把你看到的东西都告诉我，越多越…
4	5	10.05	11.3762	sil	noise
5	6	11.3762	11.7313	<A>	&噢。
6	7	11.7313	12.836	sil	noise
7	8	12.836	14.06	<DEAF>	noise

uuid为P0001_0017的被试者

2、说话内容

对无意义、听不懂的内容进行清洗：

['&哦','&啊','&嗯','&呃','&唉','&哎'] → '&噢'

【上海话】 → China

转写成短文本，把字段value的值进行拼接，

例如：

Index	uuid	conment
0	P0001_0017	noise,图片上有哪些人他们在做什么，你觉得他们在做什么告诉我，,noise,把你看到的东西都告诉我，越多…

特征工程

二、音频

以帧级别的Low-level descriptors (LLD)音频特征，包含25个字段：

- 统计学特征
- 一阶差分统计学特征

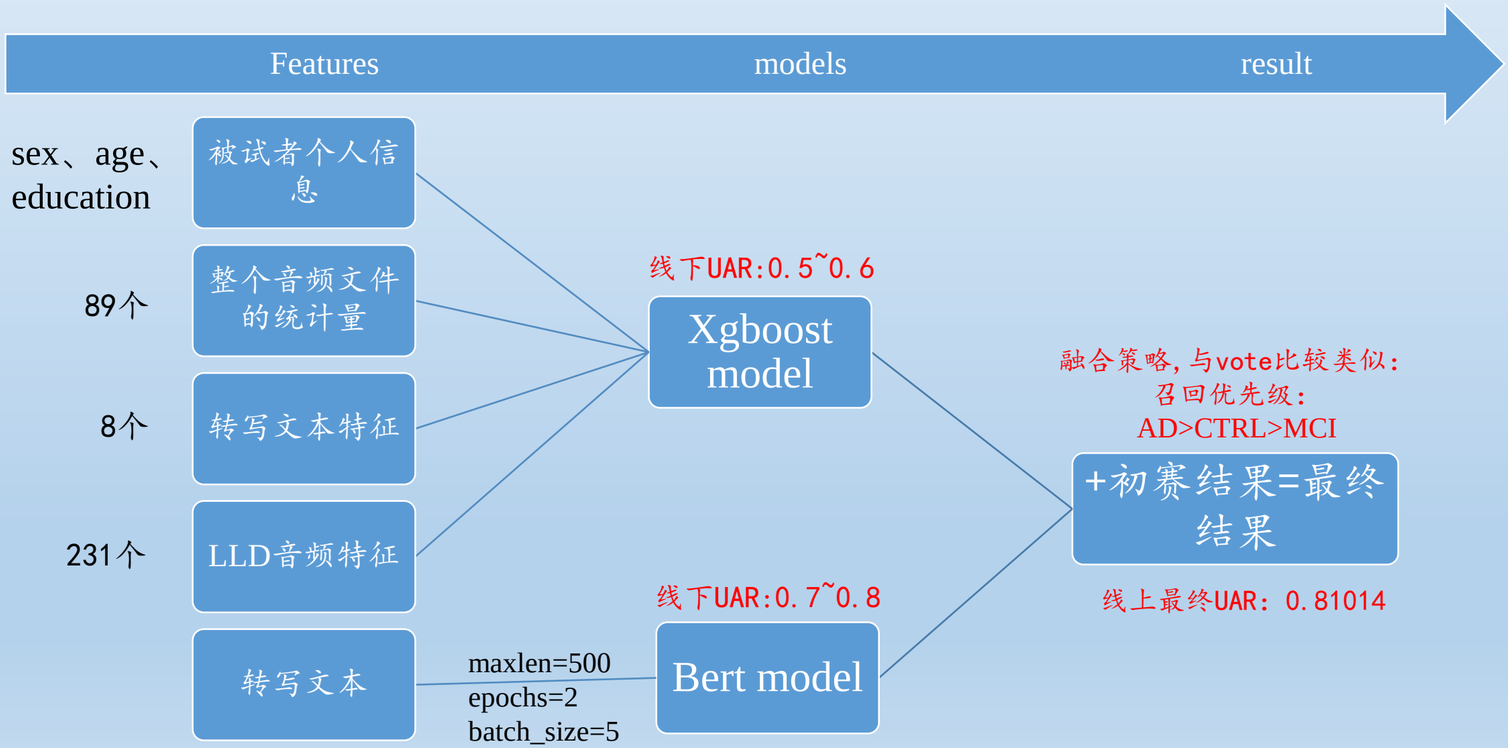
主要包括最大值max、最小值min、平均值mean、标准差std、中位数median等等

uuid为P0001_0017的被试者

Index	name	frameTime	Loudness_sma3	alphaRatio_sma3	marbergIndex_s
0	'P0001_0017'	0	0.0108697	-20.9425	30.544
1	'P0001_0017'	0.01	0.0101555	-19.0469	27.2792
2	'P0001_0017'	0.02	0.00969725	-18.5027	25.9852
3	'P0001_0017'	0.03	0.0089931	-14.3442	21.8435
4	'P0001_0017'	0.04	0.00897479	-13.8879	22.0608
5	'P0001_0017'	0.05	0.0128404	-8.42097	14.9312
6	'P0001_0017'	0.06	0.0158561	-7.52699	15.0117
7	'P0001_0017'	0.07	0.018833	-5.17018	12.5611

算法模型

两个模型分别进行5折交叉验证，最后是两个模型的融合：



分数
81.01411
78.54277
77.12963
73.85802
73.07099

其他尝试

- ✓构造词向量：tf-idf特征（维数达上万）
- ✓降维处理：pca降维/高相关性，包括整个音频文件的统计量、tf-idf特征、LLD音频特征
- ✓模型尝试：cat/lgb/其他
- ✓参数设置：bert模型参数（maxlen限制,batch_size,epochs等等）

致谢

谢谢聆听！

感谢主办方对此次竞赛的支持