

科大讯飞 - 婴儿啼哭声识别挑战赛

Cat.ai 团队 @ 2020



团队经历

- 1.2019 雪浪工业数据智能挑战赛 **Top1**
- 2.2019 智创工程 AI 挑战赛暨人工智能与轨道安全数据竞赛 **Top1**
- 3.2018 平安产险 - 数据建模大赛 - 驾驶行为预测驾驶风险 **Top2**
- 4.2018 卓尔智联研究院 -Z-HACK 算法创新大赛 - 农贸市场的价格预测 **Top2**
- 5.2018 CCF BDCI 大数据与计算智能大赛 - 基金相关性预测 **Top3**
- 6.2018 工业大数据创新竞赛 **Top3**
- 7.2018 创客中国 - 国家电投 - 基于大数据的汽轮机瞬态应力分析与预测 **Top3**

赛题背景

对婴儿来说，啼哭声是一种通讯的方式，也是一种生物报警器，向外界传达着婴儿生理和心理的需求。基于啼哭声声波携带的信息，婴儿的身体状况才能被确定，疾病才能被检测出来。

婴儿啼哭声识别挑战赛旨在判别婴儿啼哭声所传递的信息。

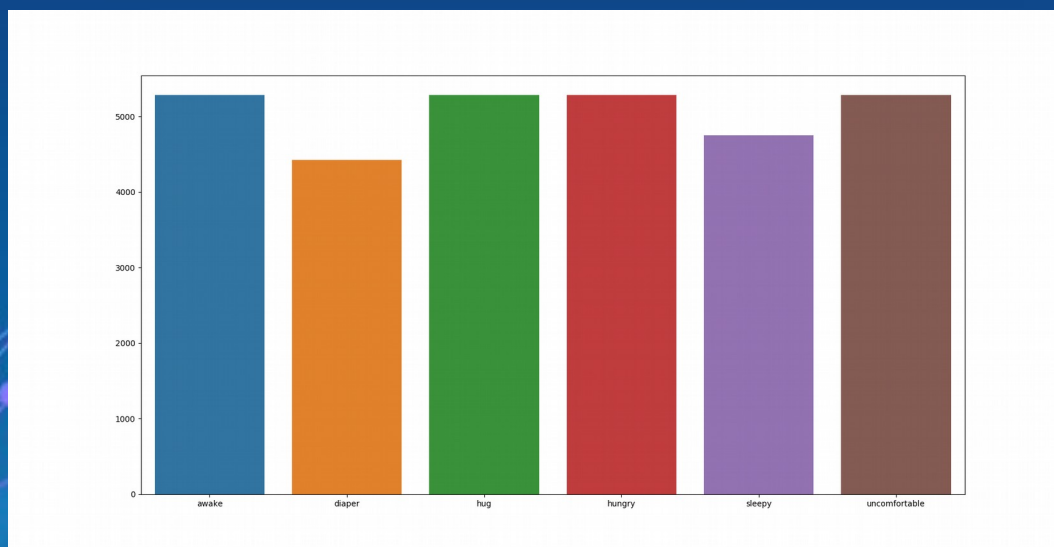


数据探索 EDA (Exploratory Data Analysis)

音频文件个数 train(30294) test(7524)

train(30294): awake(5280), diaper(4422), hug(5280), hungry(5280), sleepy(4752), uncomfortable(5280)

音频文件采样率 Sampling Rate: 大多数 16000 Hz, 部分 44100 Hz, 少部分 1600 Hz
模型采用 : 16000 Hz



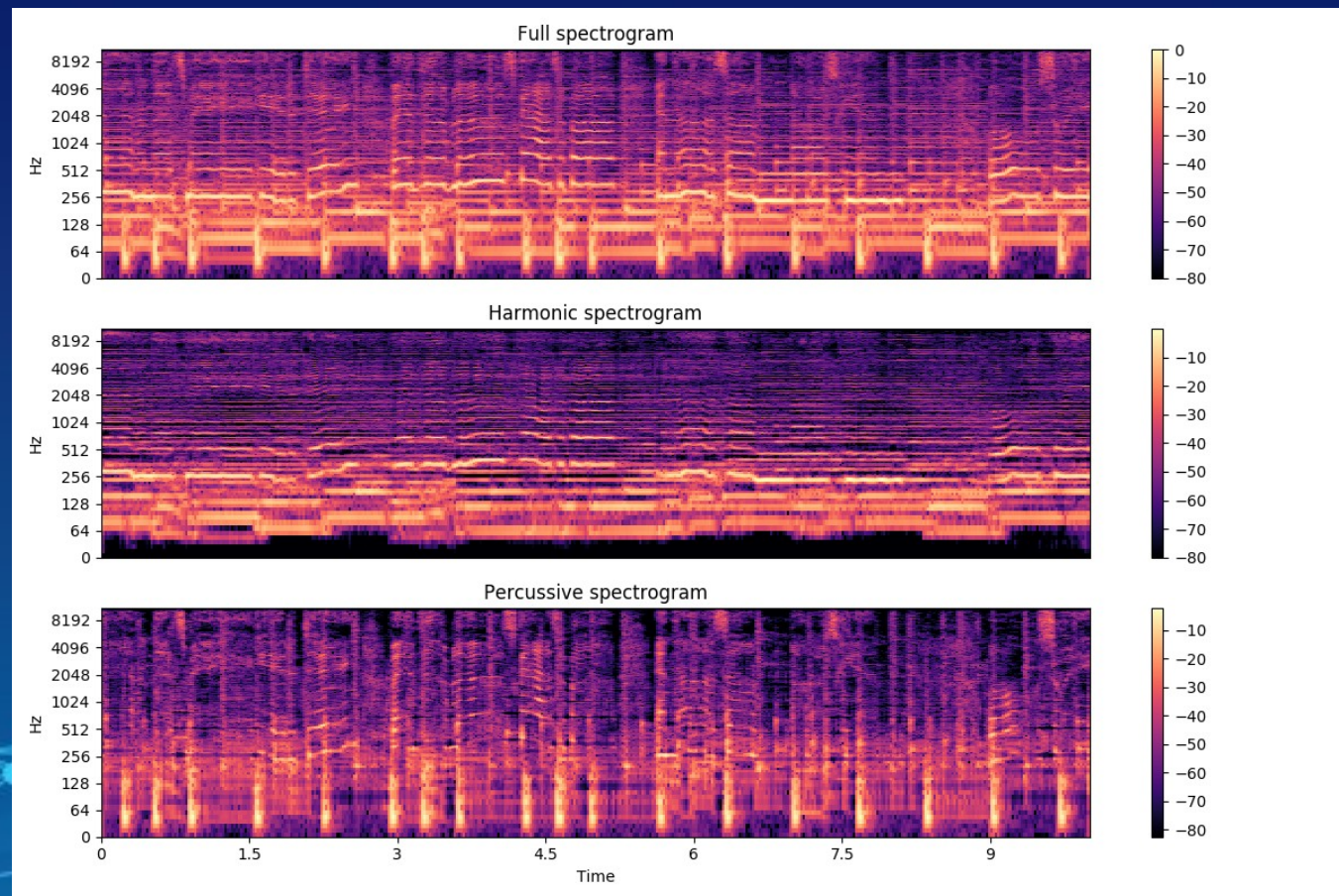
基本思路



数据预处理

1. Harmonic-Percussive Source Separation

2. Noise Reduce



特征工程 1

滑窗构建统计特征

滑窗大小的选择：尝试了 1.2s、2s、5s、10s 的划窗长度，最终选择 1.2s 的滑动窗口。



特征工程 1

滑窗构建统计特征

划窗后对每段 1.2s 的音频提取音频特征，并对音频特征提取统计特征。最后得到 165 个特征列。

音频特征

mfcc

rms

centroid

bandwidth

spectral
contrast

flatness

spectral rolloff

zero_crossing
_rate

统计特征

mean

max

min

std

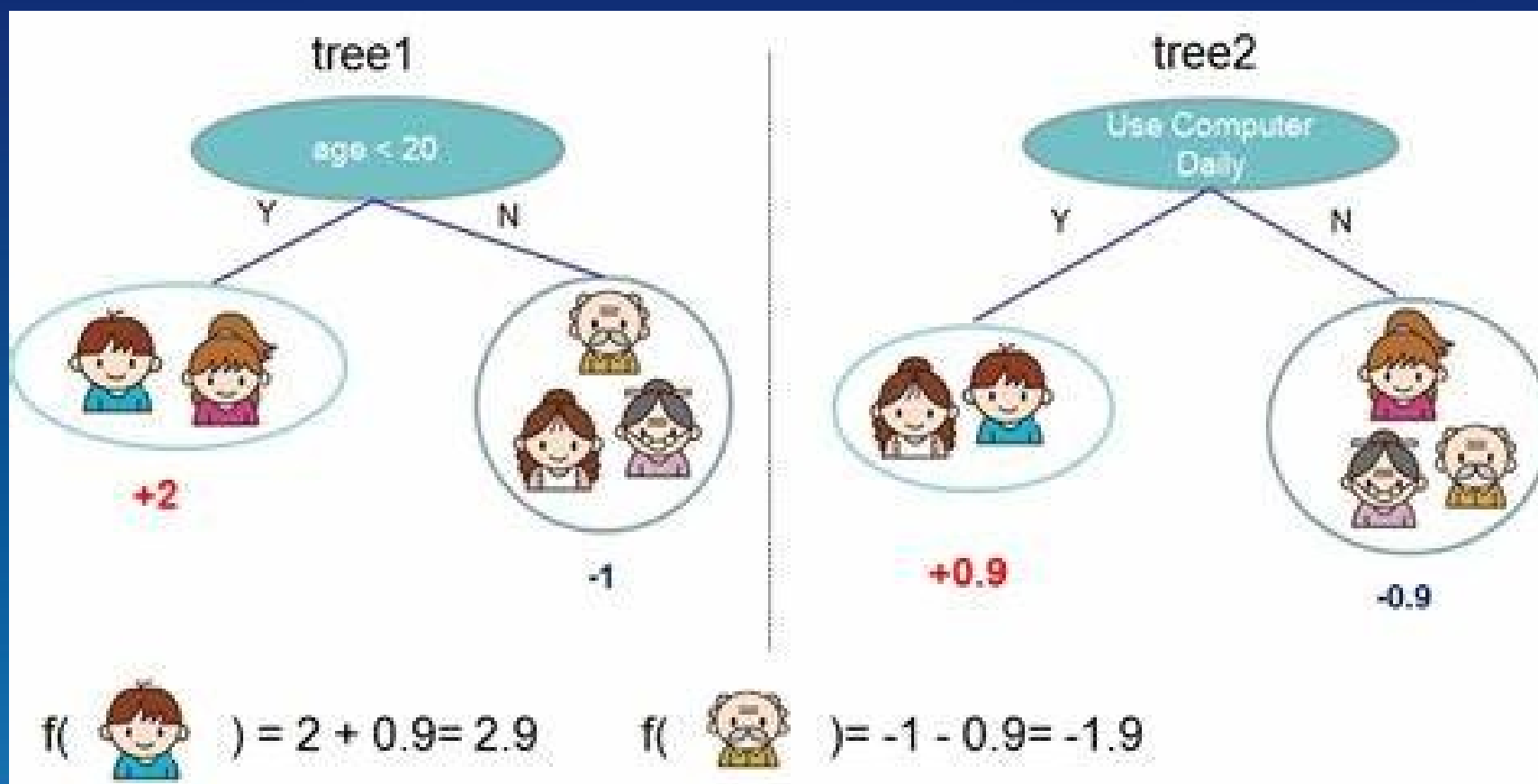
skew

165 列特征
的单个样本

算法模型 1

xgboost

训练提取音频特征的统计特征

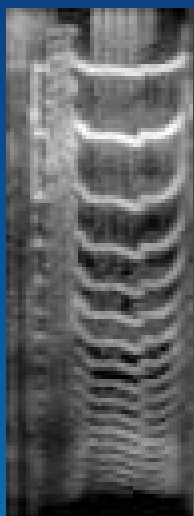


特征工程 2

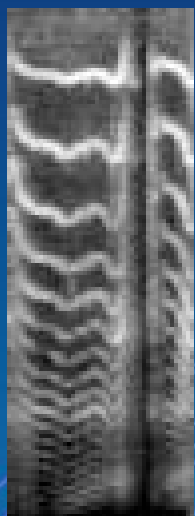
滑窗构建 mel_spectrogram 特征

使用 `librosa.feature.melspectrogram` 方法从音频中提取得到 melspectrogram 特征，划窗长度选择 2s。

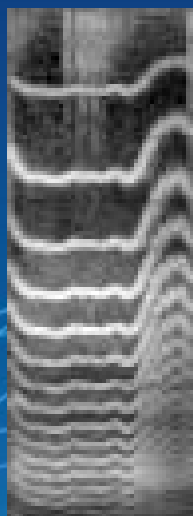
考虑到提取的特征矩阵中数值为 float 类型，存储和训练时占用很大的硬盘和内存空间，将之转换为 uint8 的图片格式，图片大小是 (47,128)，空间占用大大降低，也便于使用 resnet 等图片处理的模型训练。



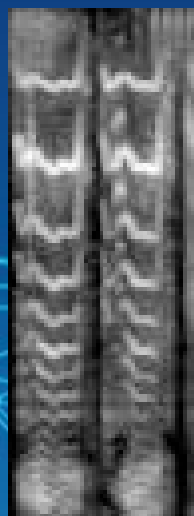
awake



diaper



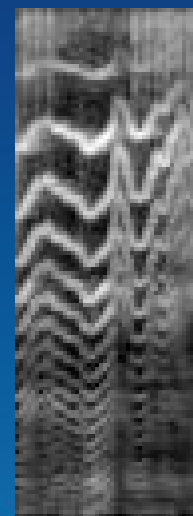
hug



hungry



sleepy



uncomfortable

算法模型 2

resnet 网络

训练提取的 melspectrogram 图片特征

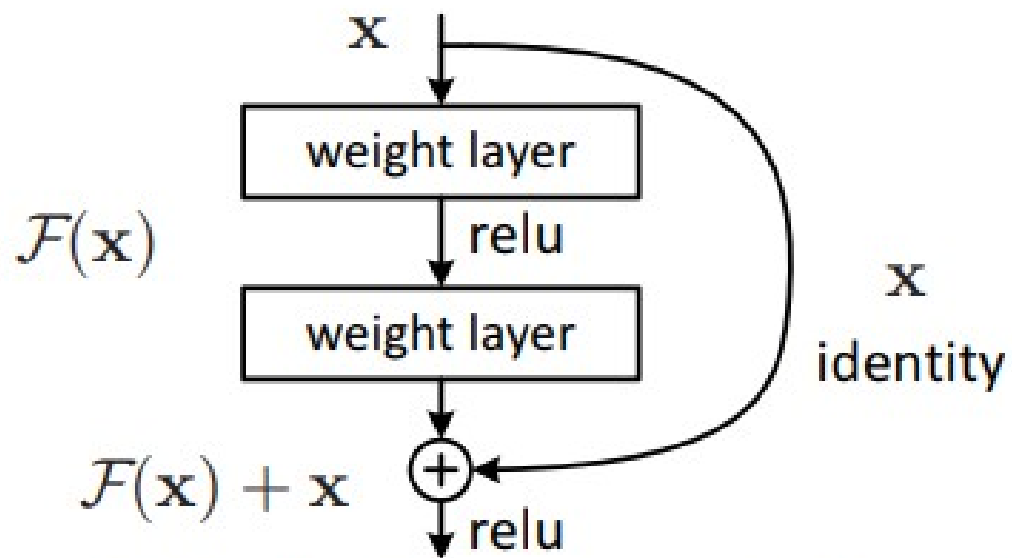
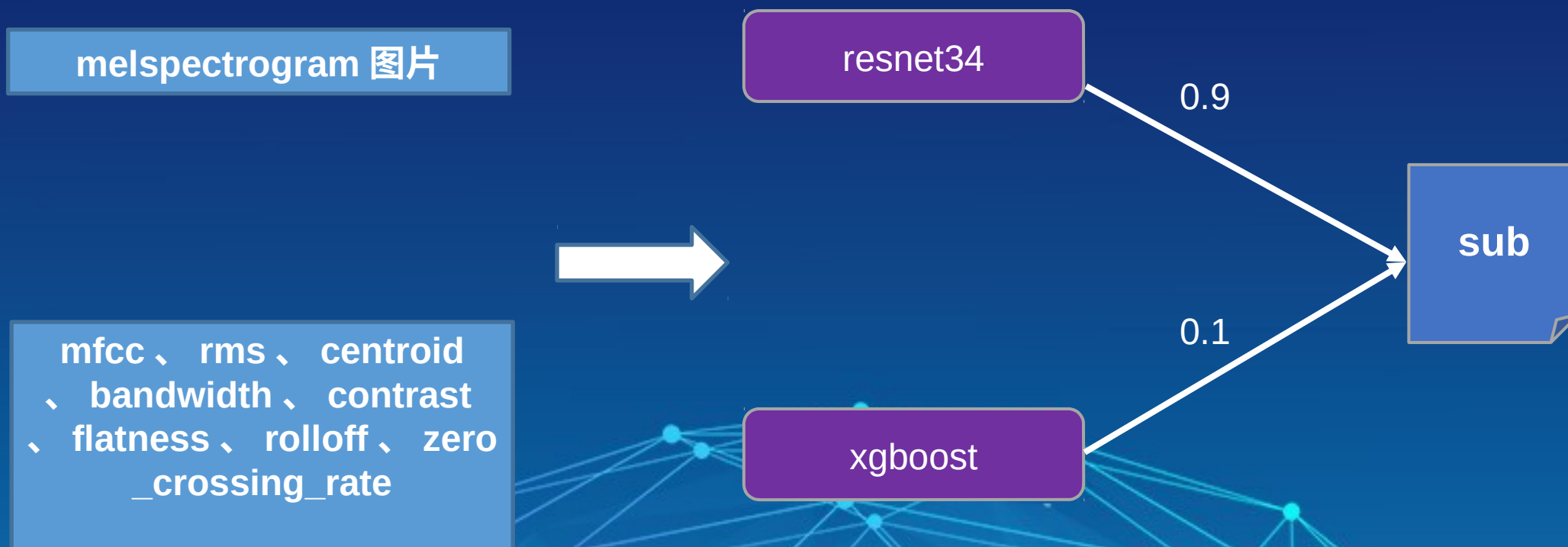


Figure 2. Residual learning: a building block.

模型融合



生成结果

sample1_1 proba

sample1_2 proba

sample1_3 proba

sample1_4 proba

sample1_5 proba

...

sample1_x proba

1. 取 mean
2. 取 argmax



sample1 result

- 1 提升采样率为 1600Hz 音频预测准确率
- 2 进一步改进 CNN 模型，考虑余弦退火训练、label smoothing 等 trick，还可考虑可视化神经网络结果，尝试更多模型，如 resnext50 等。
- 3 对树模型进行调优，使得结果更优



THANK
YOU!
Q&A