

# 数智创新 声至未来

DEEP IN DIALECTS, FOR FUTURE WAVE

第八届信也科技杯算法大赛

THE 8TH FINVOLUTION DATA SCIENCE COMPETITION

**best2fit**

丁攀，北京航空航天大学，算法工程师；  
朴奕，武汉理工大学，算法工程师



# 数据理解

1. 不同于一般的声纹识别任务，如说话人辨认，倾向于方言发音方式，不同音素占比等方面的区分，而非个人音调模式的匹配；
2. 同一种方言之间的距离为0，这类似于聚类任务；不同方言间却有确定的距离，并非越远越好，这又类似于回归任务；



# 目录

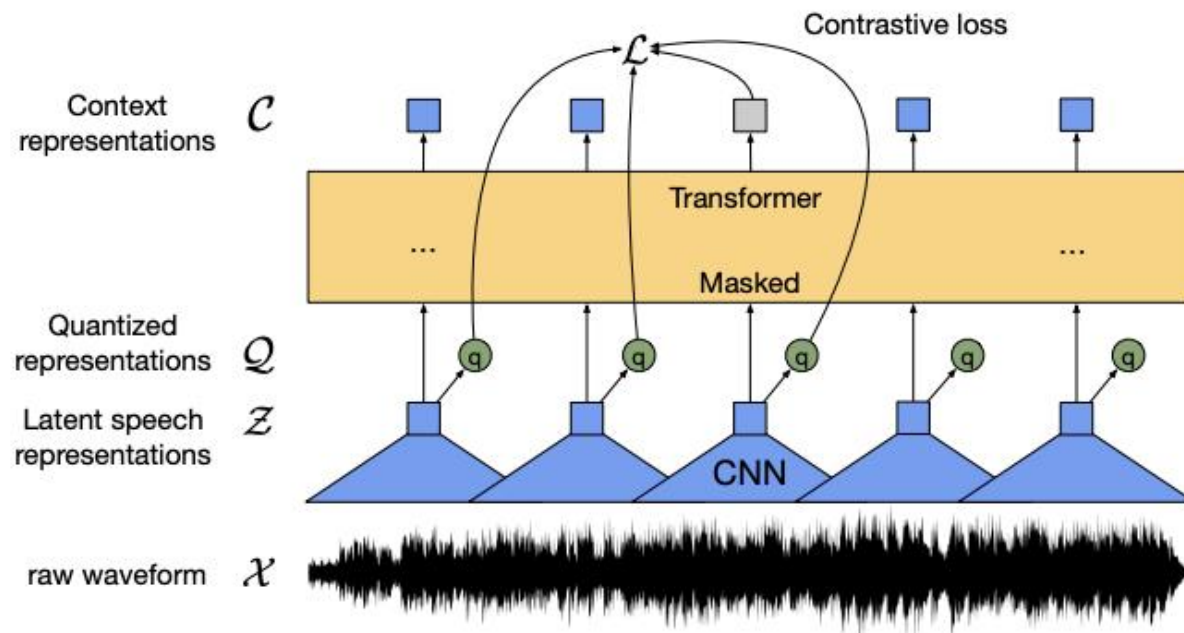
1. 数据理解
2. 方案介绍
3. 总结展望

CONTENTS



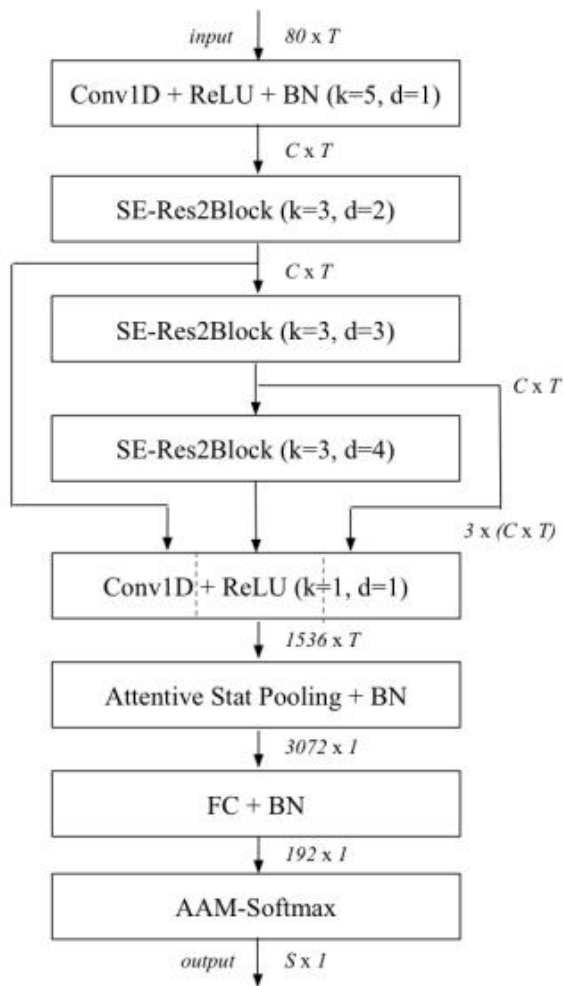
# 方案 - 特征

模型采用语音预训练模型作为backbone,选用了wave2vec模型，用于特征提取



# 方案 - 模型结构

## ECAPA - TDNN



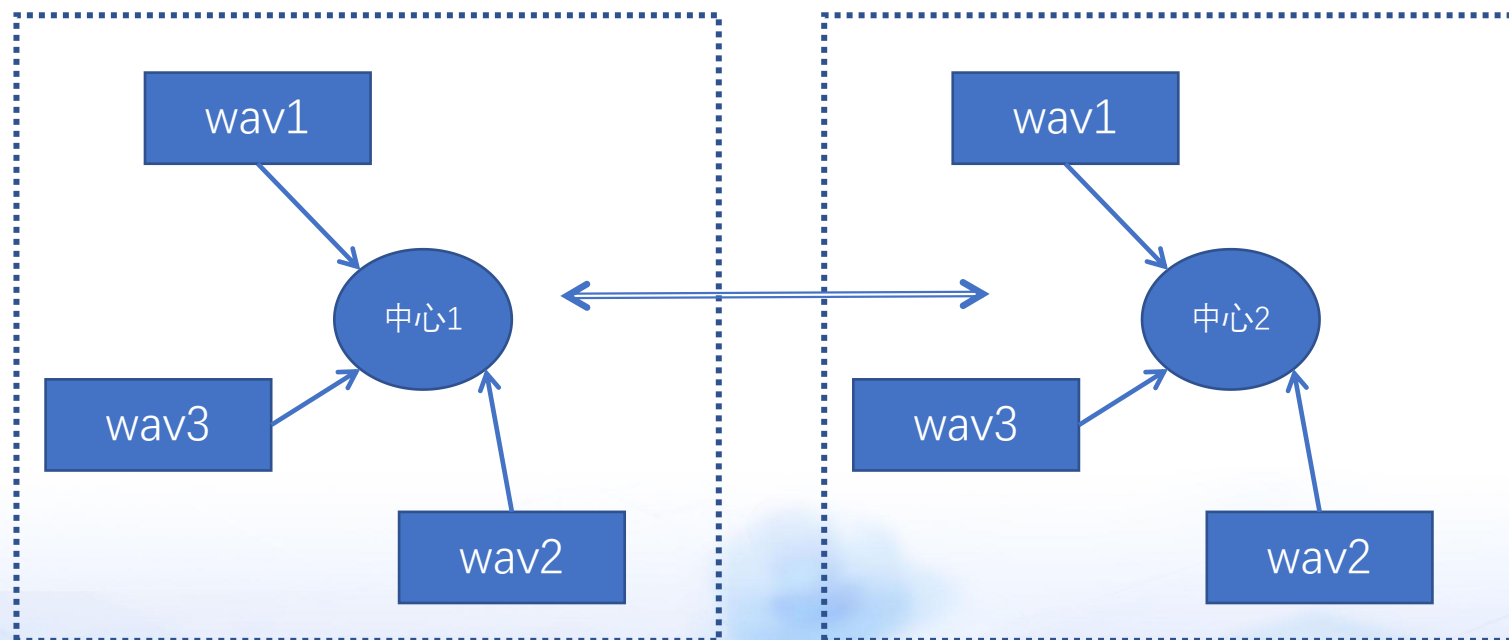
# 方案 - loss

$\text{Loss}_{\text{total}} = \text{Cross Entropy Loss} + \text{Embedding Distance Loss}$

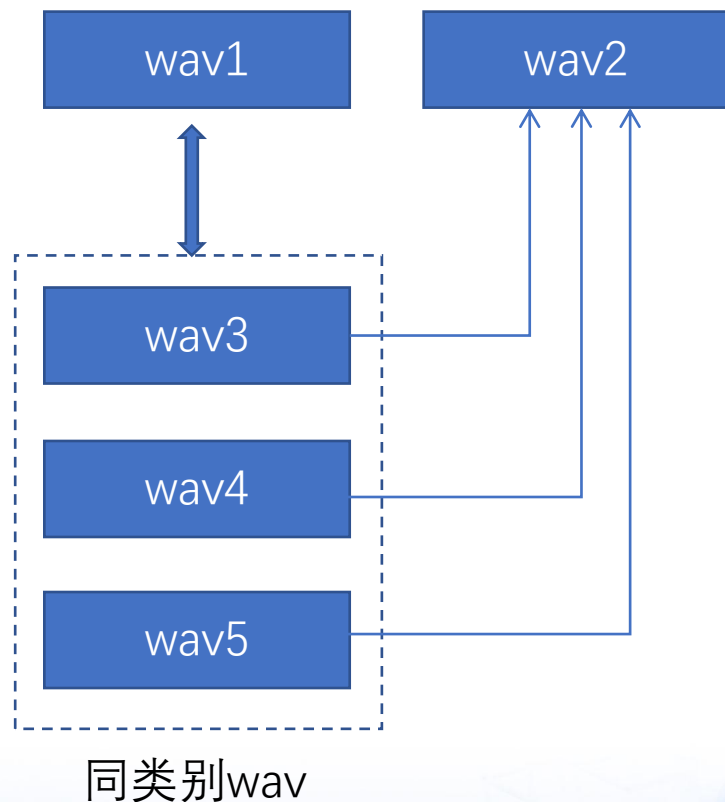
Cross Entropy Loss : 分类损失函数

Embedding Distance Loss : 方言距离距离损失函数

先进行分类任务, 逐渐加入距离损失函数



# 方案 - 后处理



“同”山之石，可以攻玉

为了求wav1 和wav2的距离，可以使用同类别其他音频和wav2的距离，多个距离取平均值



## 4:总结展望

技术层面:

- 1.是否可以引入LM-FT ( Large Margin Fine-tuning) 技术, 先针对分类任务进行优化;
- 2.合理的数据增强, 如语速, 音量, 音高等;
- 3.变更模型结构, 如ResNet101;
- 4.尝试采用训练数据进行预训练任务;

Zhengyang Chen<sup>1</sup>, Bing Han<sup>1</sup>, Xu Xiang<sup>2</sup>, Houjun Huang<sup>2</sup>, Yanmin Qian<sup>1</sup>, "Build a SRE Challenge System: Lessons from VoxSRC 2022 and CNSRC 2022," 2023, arXiv:2211.00815



谢谢

