

基于加权因子自动编码器和潜在特定源生成因子学习的单通道语音分离

陈静静¹, 毛启容^{1,2}, 秦友才¹, 钱双庆¹, 郑志燊¹

¹江苏大学计算机科学与通信工程学院, 中国镇江市, 212013

²江苏省工业网络安全技术重点实验室, 中国镇江市, 212013

摘要: 通过一系列基于自动编码器的深度学习网络结构, 单通道语音分离方法最近取得诸多进展, 其使用编码器将输入信号压缩为中间特征, 再把这些特征送入解码器重构感兴趣的特定音频源。然而, 这些方法既无法为单通道语音分离学习原始输入的生成因子, 也无法构造混合语音中的所有音频源。本文提出一个新的加权因子自动编码器模型, 在目标函数中引入正则化损失以约束目标源, 摒除其他信号源。通过在分离层中引入潜在注意力机制和监督源构造器, 加权因子自动编码器可为每一个信号源习得特定于源的生成因子和一组鉴别性特征, 从而提升单通道语音分离性能。在基准数据集上的实验表明所提方法优于现有方法。就3个重要指标而言, 加权因子自动编码器在相对更具挑战性的任务(与说话人无关的单通道语音分离)上取得巨大成功。

关键词: 语音分离; 生成因子; 自动编码器; 深度学习

<https://doi.org/10.1631/FITEE.2000019>