



概要

複数話者が参加するビジネス会議の議事録の作成や同時通訳などへの応用を目指して、話者識別技術の研究に取り組んでいます。本展示は短い音声や未知データに対する話者識別技術を紹介します。

特徴

- 短い音声でも高精度検出のためのデータ処理やモデル構造の工夫
- 生成モデルと判別モデルの特性を利用した新たなモデル学習手法
- 国際大会コンテストで準優勝を獲得

ユースケース

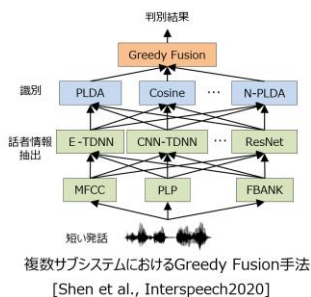
- 国際会議の議事録作成
- 同時通訳の話者ごとに字幕付与

今後の展開

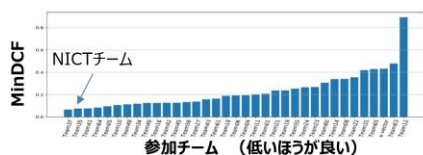
- 逐次処理システムを作成し、応用開発向けの技術の検証

■ 短い発話話者識別技術

短い発話の低識別精度の問題に対して、データの分布や多様なモデル構造などを考慮し、有効な話者情報を抽出するとともに、Greedy Fusion手法の開発により、システムの性能が大幅に改善した。



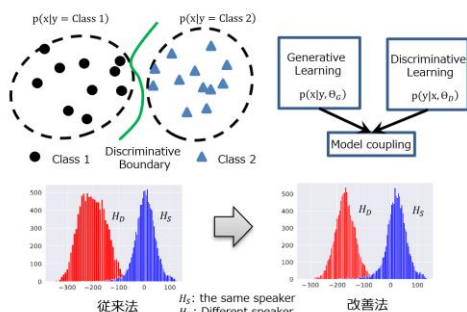
■ 高精度な短い発話話者識別システムの構築



*MinDCF: minimum detection cost function

■ 未知データに対する汎化能力を向上する技術

生成モデルと判別モデルの特性を利用し、新たなモデル学習手法を提案し、汎化能力がよい話者識別技術を開発した。



提案法: Coupling a generative model with a discriminative learning



国際大会INTERSPEECHにおける短い発話話者認識コンテストで3/4の国際研究チーム中準優勝を獲得した。

【お問合せ先】

国立研究開発法人情報通信研究機構 先進的音声翻訳研究開発推進センター 先進的音声技術研究室
astl-service@ml.nict.go.jp