2023 年度年次報告書社会変革に向けた ICT 基盤強化2023 年度採択研究代表者

Wang Xin

情報・システム研究機構 国立情報学研究所 特任准教授

プライバシー保護と偽音声検出を統合する音声データ処理基盤

研究成果の概要

This project is about creating a speech data processing framework that jointly conducts fake speech detection, speech anonymization, and speaker verification. These three tasks are implemented using deep-learning-based classification and regression models, but they are jointly designed and optimized based on log-likelihood ratios and related theory.

The main research activity follows the research plan and focuses on the joint of fake speech detection and speaker verification. While there are methods that directly sum the output scores from a fake speech detection model and a speaker verification model, there is no theory to explain why summation should be used, what the implicit requirement for the summation is, and whether summation is optimal. Using the compositional data analysis, this project gives a theoretic interpretation that links the summation of scores to the addition of log-likelihood ratios. It also points out the condition that the output scores from the two models should be well-calibrated before summation. Furthermore, this project shows that the summation is not optimal from the point of view of Bayes's decision theory. Instead, this project proposes a method to fuse the log-likelihood ratios produced by the two models and shows that this fusion method theoretically minimizes the Bayes decision cost. The performance was testified on both simulated and standard speech datasets.

This work has been submitted to Interspeech 2024 and is under double-blind review.