Title: Towards improved automatic speech recognition for children

Abstract

In recent years, Automatic Speech Recognition (ASR) technology has advanced significantly, opening avenues for novel applications targeting young speakers. However, the unique challenges presented in children's speech seriously affect general-purpose systems. This thesis explores strategies to overcome the inherent challenges associated with recognising children's speech, particularly focusing on the variability in acoustics and the limited availability of data.

First, this thesis delves into the development of hybrid ASR systems using knowledge transfer approaches. A novel multilingual transfer learning method is introduced, combining multi-task learning with transfer learning, which proves to be superior in low-resource scenarios.

Next, the role of the different components in Transformer-based architectures when fine-tuning for children's ASR is investigated. Additionally, a novel partial fine-tuning approach is shown to be superior to traditional entire model fine-tuning. Further investigation focuses on the use of Adapter modules for parameter-efficient transfer learning, showcasing their effectiveness over full model fine-tuning. Additionally, a novel unsupervised utterance clustering strategy is proposed to enhance Adapter performance, revealing its potential for group-specific adaptation.

Building upon the efficiency of Adapters, this thesis introduces the Double Way Adapter Tuning method, leveraging Text-to-speech data for data augmentation. This technique significantly reduces the gap between synthetic and real speech during fine-tuning, resulting in notable improvements.

Lastly, alternative parameter-efficient methods are investigated, ultimately proposing the concept of Shared-Adapters, where one Adapter is shared across all layers. Shared-Adapters offer superior parameter efficiency transfer compared to traditional methods, while maintaining the recognition performance, making them a compelling choice for children's ASR models.

Overall, this thesis offers comprehensive insights and innovative methodologies to tackle the challenges associated with children's ASR, thereby contributing significantly to the advancement of the field. The results of this thesis pave the way for improved human-computer interactions in educational, entertainment, health, and assistive technology applications specifically tailored for children.