Accent neutralization for non-native speech using neural style transfer

Kacper Radzikowski

Waseda University, Graduate School of Information, Production and Systems
Warsaw University of Technology, Institute of Computer Science

Abstract:

Automatic speech recognition (ASR) has been an object of extensive research since the second half of the previous century. ASR systems achieve high accuracy rates, however, only when the system is used for recognizing the speech of native speakers. The score drops in case the ASR system is being used with a non-native speaker of the language to be recognized, as the pronunciation is affected by the patterns of the mother tongue. Traditional approaches for developing speech recognition classifiers are based on supervised learning, relying on the existence of large labeled datasets. In case of non-native speech such datasets do not always exist and even if they do, the number of samples is not always high enough to train accurate classifiers. We have dealt with the problem of the non-native speech in our previous research using different approach of dual-supervised learning [1]. This time, we try tackling the problem using the style transfer methodology. We designed a pipeline for modifying the non-native speech, so that it resembles the native one to a higher extent. In this research, we plan to tackle the problem of non-native accent, using style transfer methodology. We adjust style transfer to the domain of speech and sound, to create an algorithm for real-time accent modification. Such an approach could allow to modify non-native speaker's voice on-the-fly, so that the ASR system can recognize the speech with higher accuracy. Our methodology could potentially be used as a wrapper for existing ASR system, reducing the necessity of training new algorithms for non-native speech.

Biography:

Kacper Radzikowski is currently pursuing his PhD from Waseda University and Warsaw University of Technology (double degree). He is focusing on the area of machine learning and deep learning, especially targeted for speech recognition problem for non-native speakers. He has published several conference and journal papers. Currently he is employed at the Waseda University as a Research Associate.

Publication of speakers:

1. Radzikowski, K., Wang, L., Yoshie, O., Nowak, R.: Dual supervised learning for non-native speech recognition. EURASIP Journal on Audio, Speech and Music Processing 2019:3, 1–10 (2019), doi:10.1186/s13636-018-0146-4

2. Alhindi, Tariq, Savvas Petridis, and Smaranda Muresan. "Where is your Evidence: Improving Fact-checking by Justification Modeling." Proceedings of the First Workshop on Fact Extraction and verification (FEVER). 2018.

3. Yin, Wenpeng, and Dan Roth. "twowingos: a two-wing optimization strategy for evidential claim verification." arxiv preprint arxiv:1808.03465 (2018).

4. Mueller, Jonas, and Aditya Thyagarajan. "Siamese recurrent architectures for learning sentence similarity." thirtieth AAAI conference on artificial intelligence. 2016