Age and Gender Estimation from Speech using various Deep Learning and Dimensionality Reduction Techniques.

Laxmi Kantham Durgam*[1], Ravi Kumar Jatoth[2], Daniel Hladek[3], Stanislav Ondas[4], Matúš Pleva[5] and Jozef Juhar[5]

[1,2] Department of Electronics and Communication Engineering, National Institute of Technology Warangal, Telangana, India.

[3,4,5,6] Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Slovakia.

***Corresponding author(s). E-mail(s):[*1] ld712103@student.nitw.ac.in, [2] ravikumar@nitw.ac.in, [3]daniel.hladek@tuke.sk, [4]stanislav.ondas@tuke.sk, [5]matus.pleva@tuke.sk, [6]jozef.juhar@tuke.sk.**

Abstract:

Identifying a person's age and gender from speech signal characteristics poses a significant challenge in personal identity recognition systems, particularly when security considerations are involved. In signal processing applications such as speaker recognition, biometric identification, human-machine interface (HMI), and telecommunication, age and gender estimation from voice is a crucial and demanding problem. In several signal processing domains, deep learning models have demonstrated remarkable effectiveness. In this paper, we propose a new deep learning system for the identification of speakers age and gender from speech using various speech features. We tested modified convolutional neural networks (CNN), and recurrent neural networks (RNN), including machine learning modals like support vector classification (SVC), decision trees (DT). The deep learning-based modified CNN and RNN, along with dimensionality reduction and cross-validation, proposed for age and gender recognition from speech. We applied different dimensionality reduction techniques like principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), and T-distributed stochastic neighbourhood estimation (TSNE) along with various sets of cross-validations. In this study, we used the open-air Children Speech Recognition dataset, the Biometric Visions and Computing Dataset (BVC) and the Mozilla Common Voice speech datasets for age and gender estimation from speech. The proposed modal performs better compared with existing deep learning modals. The dimensionality reduction, selection of speech features, and cross-validation played a major role in age and gender identification from the speech signal. The evaluation metrics, like accuracy, loss, sensitivity, precision, recall, and F1 score, were evaluated and compared with existing methods.

Keywords: Age and Gender Estimation, Speech Recognition, MFCC, Modified CNN and RNN, Dimensionality Reduction, Cross Validation.