Transferability Evaluation of Speech Emotion Recognition Between Different Languages

Iosifov, I.[a], Iosifova, O.[a], Romanovskyi, O.[a], Sokolov, V.[b], Sukaylo, I.[b]

[a]Ukraine Ender Turing OÜ, Tallinn, Estonia
[b]Borys Grinchenko Kyiv University, Kiev, Ukraine

Abstract

Advances in automated speech recognition significantly accelerated the automation of contact centers, thus creating a need for robust Speech Emotion Recognition (SER) as an integral part of customer net promoter score measuring. However, to train a specific language, a specifically labeled dataset of emotions should be available, a significant limitation. Emotion detection datasets cover only English, German, Mandarin, and Indian. We have shown by results difference between predicting two and four emotions, which leads us to narrow down datasets to particular practical use cases rather than train the model on the whole given dataset. We identified that if emotion transfers good enough from source language to target language, it reflects the same quality of transferability in vice verse direction between languages. Hence engineers can not expect the same transferability in the mirror direction. Chinese language and datasets are the hardest to transfer to other languages for transferability purposes. English dataset transferability is one of the lowest, hence for a production environment, engineers cannot rely on a training model on English for their language. This paper conducted more than 140 experiments for seven languages to evaluate and show the transferability of speech recognition models trained on different languages to have a clear framework which starting dataset to use to achieve good accuracy for practical implementation. The novelty of this study lies in the fact that models for different languages have not yet been compared with each other. © 2022, The Author(s), under exclusive license to Springer Nature Switzerland AG.

Author keywords

Emotion detection; Engagement analysis; Sentiment analysis; Speech emotion recognition

About this paper

https://link.springer.com/chapter/10.1007/978-3-031-04812-8_35