

Démêlage de représentations neuronales pour l'audio et la voix

L'apprentissage de représentations, un autre nom pour l'apprentissage profond, désigne l'ensemble des méthodes d'apprentissage basées sur les réseaux de neurones et qui visent à transformer les représentations brutes des données (qu'elles soient images, sons, signaux) en de nouvelles représentations plus pertinentes pour des problématiques de classification, de clustering, d'explicabilité etc. Le démêlage de représentations est un sous domaine de l'apprentissage de représentations qui regroupe les méthodes capables d'apprendre des représentations de données dites démêlées, dans lesquelles les différents facteurs de variation sont isolés dans différentes composantes. Un des intérêts du calcul de représentations démêlées est de pouvoir « éditer » des données. Lorsque l'on travaille sur des données de photos de visages par exemple cela permet de modifier la couleur des cheveux d'un visage, de le féminiser, de le vieillir etc.

Le but de ce projet est de mettre en oeuvre certaines de ces stratégies sur des signaux audio et des signaux de parole et d'étudier leurs capacités pour l'édition de signaux vocaux.