
Commande vocale

Olivier Schwander olivier.schwander@sorbonne-universite.fr

2024

Malgré les progrès récents en apprentissage profond, la classification de données audio reste un problème difficile et non résolu de façon satisfaisante. On s'intéresse ici à une tâche de commande vocale¹, les exemples sont des mots très courts, tels que ceux utilisés pour des enceintes connectées ou des interfaces de voitures autonomes. Au-delà de la classification proprement dite, plusieurs questions spécifiques se posent, comme par exemple :

- variabilité entre les locuteurs,
- inférence en temps réel,
- taille des modèles selon si la classification a lieu sur un serveur ou dans un système embarqué.

Les étapes de ce projet seront les suivantes :

- mise au point d'un protocole expérimental
- études des différents pré-traitements possibles pour les données audio
- construction de baselines à l'aide d'algorithmes non-deep learning (sklearn)
- prise en main d'une plateforme deep learning (pytorch)
- réflexion sur les contraintes spécifiques

1. https://github.com/DAGsHub/audio-datasets/tree/main/Speech_Commands_Dataset