Siyu WANG
Chen SUN
Shuai GAO

Institut national des langues et civilisations orientales

Compte-rendu du projet
Reconnaissance des langues peu annotées

Objectifs

- Appliquer les méthodes étudiées dans la partie keras et réseaux de neurones à un nouveau jeu de données
- Effectuer une reconnaissance des langues peu annotées (classification)

Données

Langues choisies : mongol, tatar, estonien et chinois

Volume du corpus :

- Train : 16 727 fichiers .png au total, environs 4 000 fichiers par langue
- Test : 400 fichiers .png au total soit 100 fichiers par langue

Méthodologie

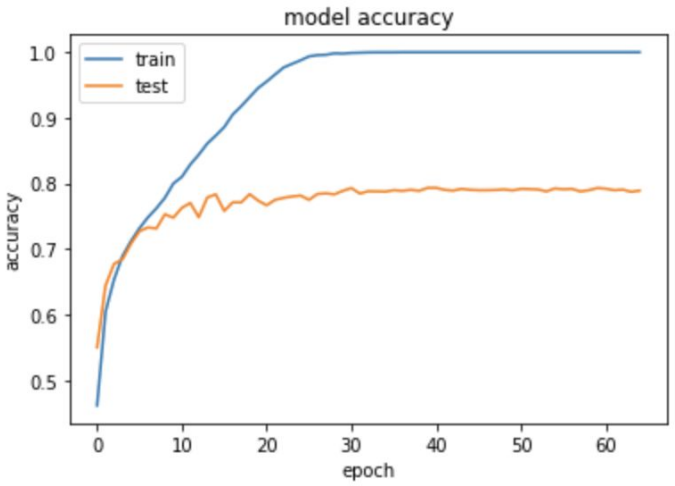
Étapes du projet :

- Télécharger les corpus des quatre langues depuis le site Common Voice.
- Prétraitement :
  - Sélectionner les fichiers valides (script python)
  - Supprimer les parties silencieuses (script praat)
  - Couper les fichiers audios en deux secondes (script python)
  - Supprimer les fichiers wav avec une durée moins de 0.8s (script python) puisqu'il sont moins pertinents
- Transformation:
  - Générer les spectrogrammes de format png à partir des fichiers audios (script praat)
- Apprentissage :
  - Google Colaboratory (script keras tensorflow)

Résultats et discussion

- Meilleurs taux d'*accuracy* avec les hyperparamètres ci-dessous :
  - batch_size = `50`
  - epoch = `100`
  - learning_rate = `0.01`

- Évolution des taux d'*accuracy* sur le corpus de test :



Malgré nos maintes tentatives en variant les hyperparamètres, nous ne sommes pas arrivés à augmenter davantage le taux d'*accuracy*, lequel demeure toujours inférieur à 0.8. Le taux le plus haut que l'on a obtenu est **0,793**.

Nos réflexions faites, les raisons sont peut-être les suivantes :

➢ Relevant de la même famille linguistique altaïque, la langue tatare et la langue mongole ont des caractéristiques communes sur lesquelles s'accordent les linguistes. D'où l'ambiguïté des données qui influerait le taux d'*accuracy* du modèle.

➢ Le volume de notre corpus est peut-être trop grand. Nous avons eu les taux autour de 0.855 sur un petit corpus de 344 fichiers .png de train et 40 fichiers .png de test, alors que quand on passe au grand corpus le taux d'*accuracy* a diminué.

➢ Pour le prétraitement, nous avons supprimé les silences et coupé les fichiers audio en 2 secondes, mais nous n'avons pas pris en compte les différences entre les sons voyelles et les sons consonnes. Peut-être il y aurait un changement de résultat si nous n'avions traité que les sons de voyelles de ces 4 langues. Par contre, nous ne sommes pas sûrs si le problème d'ambiguïté entre le mongol et le tatar soit résolu avec ce genre de traitement.