

Rapport SAM

Modèle multi-modaux

Cléa Han, Yanis Labeyrie et Adrien Zabban

janvier 2024

1 Introduction

2 Données

parler des données (d'où elles viennent) dire que les vidéos sont déjà pré-traitées (analyse des landmarks) parler (et redéfinir) les IPUs comment sont agencées les données parler des problèmes (temps de chargement du csv) et comment on l'a résolu : avec des skiprows

3 Traitement unimodale

3.1 Traitement du texte

on a utilisé Bert

3.2 Traitement de l'audio

on a utilisé wave2vect

3.3 Traitement de la vidéo

on utilise 1 voir 2 couches LSTM

4 Traitement multimodal

4.1 Maximum de vraisemblance

on prend la solution la plus prédite

4.2 Apprentissage basique

on enlève la dernière couche dense de chaque modèle et on concatène tout et on refait passer dans une couche dense (on utilise les entraînements déjà faits) en gelant les poids précédant

4.3 Entraînement du gros modèle

on ne gèle pas les poids

5 Conclusion