

Speech Emotion Recognition

Nouhayla Zakaria-Wiam Hamadi

Décrivez le projet en quelques phrases. Il ne doit PAS s'agir d'une revue, mais simplement d'une description du problème abordé par le projet, de sa solution et d'un résumé des résultats obtenus.

Le projet consiste à détecter les émotions à partir de sons préenregistrés. La source de ces sons est le dataset Ravdess (audio vocal étiquetés avec des classes d'émotions). Ce dataset contient des échantillons de paroles et de musique. L'équipe utilise du data augmentation en ajoutant du bruit dans les vocalisations initiales de Ravdess. L'objectif du data augmentation, fait en prétraitement, est de créer un modèle plus généraliste. Les données sont étiquetées. L'équipe propose d'utilisation de différents outils d'analyse comme: FFT, spectrogramme, spectral contrast. L'utilisation de ces techniques d'analyse permet de mieux comprendre les données utilisées. L'équipe propose l'utilisation de différents modèles existants pour la détection d'émotions comme : wave2vec, hubert et yamnet_model. Ces modèles sont combinés pour former des modèles plus complexes.

Décoder les dimensions humaines, technique de fine tuning.

Utilisation de modèle préentraîné.

Dataset Ravdess, parole et parole avec musique. Étiquetter, 8 classes d'étiquettes.

Data augmentation, création de bruit pour améliorer la généralisation.

Citez et justifiez au moins deux aspects forts du projet.

- **Idée d'ajouter du bruit au dataset original afin de mieux généraliser. En effet, l'ajout de bruit permet d'incorporer des éléments aléatoires dans le jeu de données. Ceci permettra une meilleure généralisation.**
- **L'équipe implémente une bonne diversité des modèles ré-entraînés. Entre autres, wave2vec, hubert et yamnet_model.**
- **Beaucoup d'analyses acoustiques (FFT, spectrogramme, spectral contrast pour comparer l'intensité du voisinage, etc) L'utilisation de ces techniques d'analyse permet de mieux comprendre les données utilisées. Cette meilleure compréhension permet de mieux designer les solutions à appliquer.**

Énumérez et justifiez au moins deux aspects manquants du projet.

- Il manque un comparatif des résultats avant/après noising. Le but est de voir l'effet de l'ajout du bruit sur les résultats.
- Il n'y a pas vraiment eu développement de nouveau modèle, l'essentiel de l'apport est sur le noising du dataset original. Il pourrait être intéressant de développer des nouveaux modèles dans le but de pousser plus loin l'état de l'art.
- Manque les résultats des expérimentations. Ces résultats permettraient de mieux apprécier la performance des solutions proposées.
- Si possible essayer plusieurs jeux de données. En utilisant plusieurs jeux de données, il serait possible de mieux entraîner les modèles et d'augmenter la capacité de généralisation des solutions. On pourrait aussi voir s'il y a des différences de performance en fonction des jeux de données.
- Préciser explicitement la façon, quand et comment le data augmentation se fait.

Énumérez au moins trois questions à l'équipe du projet.

- Je n'ai pas très bien compris quel(s) bruit(s) ont été insérés dans le dataset? Si les bruits ne sont pas divers, on a un risque davantage d'overfitting alors qu'on tente de faire l'inverse en rendant le modèle plus généraliste.
- Avez vous à regarder si les distributions de vos étiquettes sont skewed ou uniforme pour voir si vous pouvez appliquer du upsampling ou downsampling.
- Inclure plus d'informations sur votre processus de fine tuning, comment est-ce que vous procéder pour chaque modèle.