

VoiceTransformer pour la détection précoce des maladies neurodégénératives

Ahmad Tay, Mohamed Djallel Dilmi, Faten Chaieb-Chakchouk

Efrei Research Lab, Paris-Panthéon-Assas University, Paris 2,
30 Av. de la République, Villejuif, 94800, France
ahmad.tay@efrei.fr djallel.dilmi@efrei.fr
faten.chakchouk@efrei.fr

1 Introduction

Les maladies neurodégénératives chroniques telles que l'Alzheimer et la maladie de Parkinson sont caractérisées par une détérioration progressive du système nerveux. Leur détection précoce est essentielle pour ralentir leur progression et améliorer la qualité de vie des patients. Les recherches récentes suggèrent que la voix, en tant que biomarqueur non invasif, est un outil prometteur pour diagnostiquer ces pathologies avant l'apparition des symptômes cliniques visibles. Les analyses vocales permettent de détecter des changements subtils dans les caractéristiques acoustiques, liées à des déficiences cognitives et motrices.

Cependant, l'identification automatique de ces caractéristiques reste un défi. Les voix, en tant que signaux temporels, sont complexes, comportant des variations à la fois à court et à long terme. Les approches traditionnelles, basées sur des descripteurs acoustiques tels que la fréquence fondamentale (F0), les formants, et les mesures d'intensité, bien qu'efficaces dans certaines tâches de classification vocale, manquent souvent de flexibilité pour capturer des relations temporelles plus longues et non linéaires, importante pour détecter des signatures de pathologies neurodégénératives.

Avec l'essor de Transformers dans le traitement de la parole (Dong et al., 2018), une nouvelle voie s'est ouverte pour l'analyse des signaux vocaux, surtout leur efficacité dans la capture de dépendances à long terme dans les séquences temporelles. Leur base : les mécanismes d'attentions leur permet de pondérer différemment chaque élément d'une séquence, accordant ainsi plus d'importance aux segments pertinents pour la tâche de classification. Ces modèles, à travers les "cartes d'attention", fournissent une explicabilité alignée avec l'importance relative des différents segments de la voix, ce qui peut révéler des indices cruciaux pour la détection des maladies neurodégénératives.

2 Approche méthodologique

Dans ce travail, nous proposons d'utiliser un encodeur basé sur l'attention (Transformer encoders) pour extraire automatiquement des caractéristiques vocales pertinentes pour la dé-

tection précoce de maladies neurodégénératives. Nous partons d'un modèle d'encodeur pré-entraîné sur des grandes bases de données publiques de voix (Zhao et al., 2019). Le modèle est ensuite affiné (fine-tuning) sur un jeu de données spécifiques de patients (Pitt-corpus, Adresse Dataset). Ce fine-tuning permet d'adapter les paramètres du modèle aux spécificités des pathologies neurodégénératives, telles que les altérations subtiles de la prosodie, de la fluidité, et des pauses dans le discours.

Les cartes d'attention générées par les Transformers permettent une interprétation plus claire des parties de la voix qui contribuent à la détection des anomalies. Cela représente une avancée par rapport aux modèles classiques où l'interprétabilité des décisions est souvent limitée.

3 Expérimentations

Nous comparons la performance des modèles basés sur l'attention avec des approches classiques, telles que les modèles de régression logistique, les machines à vecteurs de support (SVM) et les réseaux de neurones multi-couches (MLP) utilisant des descripteurs acoustiques préconstruits (MFCC, énergie, F0, formants). Les résultats préliminaires (proof of concept) montrent une amélioration en termes d'accuracy de classification sur le jeu de données Dementia Bank Dataset (Becker et al., 1994). L'intégration de ce modèle avec des systèmes de suivi vocal dans les environnements de soins pourrait permettre une détection continue et en temps réel des premiers signes de déclin cognitif.

Références

- Becker, J. T., F. Boiler, O. L. Lopez, J. Saxton, et K. L. McGonigle (1994). The natural history of alzheimer's disease : description of study cohort and accuracy of diagnosis. *Archives of neurology* 51(6), 585–594.
- Dong, L., S. Xu, et B. Xu (2018). Speech-transformer : A no-recurrence sequence-to-sequence model for speech recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5884–5888, doi: 10.1109/ICASSP.2018.8462506.
- Zhao, Y., J. Li, X. Wang, et Y. Li (2019). The speechtransformer for large-scale mandarin chinese speech recognition. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.