**Résumé d’article**

**Apprentissage Profond**

**8INF892**

**Kévin Bouchard**

**Université du Québec à Chicoutimi**

**Département d’Informatique**

**Jean-Philippe Larouche**

**LARJ07089501**

**28/04/2024**

# Introduction

Le domaine de la reconnaissance d’activités humaines a un vaste potentiel d’application dans une multitude de contexte. Cependant, une des plus grandes difficultés de ce domaine sont les données (activités). Notamment, elles nécessitent un grand effort de traitement manuel afin d’entraîner des modèles performants. L’arrivée des *Large Language Model* (LLM) offre de nouvelles possibilités de traitement de données avec leur raisonnement étape par étape similaire à un humain. C’est dans cet optique que l’équipe de *Ji et Al.* présente l’étude *HARGPT : Are LLMs Zero-Shot Human Activity Recognizers*. L’étude tente de déterminer si les LLMs sont en mesure d’interpréter directement (sans modification) les données de capteurs, puis de correctement identifier l’activité qu’un humain effectue selon les données. Le tout sans *fine-tunning* ou préparation du modèle.

# Méthodologie

## Données

Les chercheurs font l’utilisation de deux bases de données pour l’étude. La première est Capture24, une base de données contenant des données d’accélérations provenant d’un capteur sur le poignet des participants. Ces données sont associées à quatre activités : faire du vélo, dormir, marcher et debout-lever. Ces quatre activités sont très différentes en terme de données et alors considérées « faciles ». La deuxième est HHAR, une base de données contenant que deux activités : monter et descendre des marches. Ces deux activités sont très similaires au niveau des données et alors considérées « difficiles ». L’objectif est d’évaluer séparément la capacité des LLMs de différencier entre des activités semblables en terme de données. Les deux ensembles sont séparés en ensemble de test, validation, vue et non-vue de manière 4 :1 :1 :2.

## Modèles

Afin d’évaluer l’efficacité des LLMs à traiter des données de capteurs brutes, les chercheurs commencent par établir quatre modèles « classiques » de comparaison. On y retrouve : Random Forest (RF), Support Vector Machine (SVM), Deep Convolutional Neural Network (DCNN) et LIMU-LSTM. Pour représenter les LLMs, les chercheurs décident d’évaluer le modèle GPT4. Ce modèle est, au moment de la rédaction, considéré comme étant le modèle de type LLM le plus performant.

## Invite de texte

La communication avec des LLMs se fait par invite de texte. Plusieurs recherches ont démontré que les LLMs performent le mieux lorsqu’on leur permet de tirer avantage de leur raisonnement étape par étape. Afin d’évaluer l’impact de cette approche, les chercheurs ont soumis deux types d’invite de texte au LLM. Une demande au LLM de faire l’analyse des données étape par étape, dénotée *CoT*, alors que l’autre lui demande une réponse immédiate sans analyse, dénotée *DO*.

L’invite de texte commence par une instruction suivie d’une question. Des éléments importants sont fournis au modèle à cette étape : type d’appareil, fréquence des données et type de données. Ensuite, les données brutes de capteurs sont fournies. Finalement, la liste des activités possibles à reconnaitre est donnée au LLM avant de lui demander sa réponse.

## Mesure de performance

La performance des différents modèles est évaluée à l’aide de trois métriques : Précision, Recall et F1-Score.

# Résultats

Les résultats démontrent que le LLM GPT4 a des performances remarquables de reconnaissance d’activités humaines lorsqu’on lui fournit des données brutes provenant de capteurs. Sur l’ensemble de données Capture24, lorsqu’on demande à GPT4 d’interpréter les données étape par étape, il démontre une précision de 0.818, recall de 0.793 et un score F1 de 0.795. Les modèles LIMU-LSTM et DCNN sont ceux qui se rapprochent le plus de ce dernier avec une précision de 0.595, recall de 0.600 et score F1 de 0.588. Lorsqu’on demande à GPT4 de fournir une réponse immédiate (sans analyse) les performances chutent de manière catastrophique, soit une précision de 0.498, recall de 0.468 et score F1 de 0.495. Sur l’ensemble de données HHAR, on observe une tendance similaire mais, un peu inférieure. GPT4, avec l’analyse étape par étape, démontre une précision de 0.790, recall de 0.795 et score F1 de 0.790.

# Discussion et conclusion

Pour commencer, l’étude a effectivement démontré que les LLMs performent beaucoup mieux lorsqu’on tire avantage de leurs architectures d’analyse étape par étape. On observe près de deux fois la performance versus demander une réponse immédiate. Ensuite, les LLMs démontrent une performance remarquable quant à leur capacité à reconnaitre des activités humaines lorsqu’on leur fournit des données de capteurs brutes, et ce, sans aucune manipulation des données au préalable. Cependant, des tests préliminaires sur d’autres LLMs, tel que Gemini et Llama2-70b, et même sur GPT4 ont démontré que les LLMs peuvent ne pas fournir une réponse définitive. Parfois, ils suggèrent que les données peuvent représenter deux différentes activités et présentent leurs résultats de manière incertaine.

# Opinion

À mon avis, l’étude HARGPT démontre un avenir prometteur pour LLMs dans le contexte de la reconnaissance d’activité humaine. Cependant, les tests furent effectués que sur un modèle et les tests préliminaires sur les autres modèles ont démontré des réponses ambiguës. Afin que ces modèles soient d’une utilité réelle, le problème de réponse multiples et ambiguës doit être résolu. De plus, la performance de ces modèles lorsqu’on leur fournit des données provenant de plusieurs capteurs en même temps reste à démontrer. Finalement, les ensembles de données simples et limités utilisés pour l’étude ne valident pas la performance de ces modèles dans un environnement complexe et multi-dimensionnel.

# Bibliographie

Ji S., Zheng X., Wu C. 2024. HARGPT: Are LLMs Zero-Shot Human Activity Recognizers. arXiv:2403.02727 [cs.CL]. <https://doi.org/10.48550/arXiv.2403.02727>