

Techniques d'optimisation pour les Graph Neural Networks

Encadrement : Olivier Beaumont, Lionel Eyraud-Dubois, Julia Gusak (Topal, Centre Inria de l'Université de Bordeaux)

Ce sujet s'inscrit dans le cadre d'un projet en cours de finalisation, en collaboration avec Météo France, autour de l'utilisation efficace de machines hautes performances pour l'entraînement de modèles d'IA, en particulier appliqués à la prédiction météo. À la suite de ce stage, une poursuite en thèse dans le cadre de ce projet est très probable.

Les modèles d'IA du type Graph Neural Networks (GNNs) sont actuellement de très bons candidats pour des applications de prédiction météo, permettant une émulation efficace des modèles physiques utilisés traditionnellement, pour un coût significativement moindre. L'entraînement de ces modèles reste coûteux, en particulier pour des données de grande taille et/ou de grande précision.

Le travail portera sur l'étude des techniques d'optimisation pour améliorer l'entraînement de ces modèles, aussi bien du point de vue de l'utilisation mémoire que du volume de communications ou de l'efficacité parallèle. On combinera une étude bibliographique et des expériences pratiques pour bien comprendre le comportement et les performances de certains modèles cibles, comme le modèle AIFS du Centre Européen de Prévisions Météorologiques. En se basant sur les travaux autour de l'optimisation mémoire et parallèle de modèles classiques effectués dans l'équipe TOPAL, on pourra aussi éventuellement proposer et évaluer des nouvelles techniques de partitionnement ou de parallélisme adaptées aux modèles GNNs afin d'en optimiser l'exécution.

Ce travail se situe à la frontière entre le calcul parallèle intensif et l'IA, et nécessite un intérêt fort pour le HPC et/ou les techniques d'IA efficace.