

Développement d'architectures de réseaux de neurones en graphes pour données massives et calcul parallèle

1 - Contexte

Les réseaux de neurones en graphes (Graph Neural Networks, ou GNN) sont des méthodes récentes visant à adapter les réseaux de neurones artificiels aux espaces non-euclidiens que décrivent les données structurées en graphes, afin de résoudre des problèmes propres à ce type de données (classification de nœuds ou de sous-graphes, etc.). L'application de ces méthodes émergentes aux très grandes données pose plusieurs défis au niveau de la définition de l'architecture et de la communication des différentes couches de neurones, et de leur combinaison pour conserver la localité des régions du graphe.

Le passage à l'échelle nécessite également d'utiliser des concepts et outils de calcul haute performance (code parallèle, mémoire distribuée, formats de stockage adaptés, etc.), dont une partie ont déjà été développés à la Maison de la Simulation. Il est donc attendu des travaux du stage proposé, d'utiliser ces outils afin de proposer une solution compréhensive à des problèmes de "machine learning" sur graphes.

2 - Objectifs

Dans ce cadre, les objectifs seront les suivants :

- explorer différentes méthodes d'algèbre linéaire, ou d'autres propriétés, permettant d'extraire des aspects distincts de la structure du graphe,
- utiliser ces méthodes pour proposer des architectures de GNN (ainsi que leurs formulations mathématiques associées),
- développer une implémentation parallèle, adaptable à de très larges données, d'un framework générique pour une telle méthode (en premier lieu en C++ sous MPI ; d'autres implémentations pourront être considérées par la suite, avec CUDA et XMP par exemple),
- implémenter plusieurs options pour chaque élément de la méthode, et réaliser des séries d'expérimentations sur un cluster d'accélérateurs permettant d'identifier l'impact des différents composants, à la fois sur l'exactitude des résultats pour des tâches de "machine learning" et sur le temps d'exécution.

3 - Prérequis



Sujet de stage de fin d'études/M2, 2021 (6 mois)

- des connaissances en algèbre linéaire, programmation parallèle et répartie, et génie logiciel sont souhaitées.

4 - Encadrement et environnement

- Il s'agit d'un stage de 6 mois, rémunéré selon les grilles du CNRS.
- Des accès à des superordinateurs et clusters sont prévus.
- Le stage se déroulera au sein du laboratoire mixte CNRS-CEA-Université Paris Saclay-Université de Versailles (USR 3441) à Saclay.
- Le directeur de stage est Serge Petiton

5 - Doctorat

Une poursuite en doctorat est envisageable.