ENSEIRB-MATMECA

Apprentissage profond efficace

École / Prépa
ENSEIRB-MATMECA

Présentation

Code interne : EI9IS324

Description

Ce cours aborde les défis posés par les réseaux neuronaux modernes, qui nécessitent une mémoire et une puissance de calcul importantes, rendant difficile leur déploiement sur des dispositifs mobiles et périphériques. De plus, l'échelle et la complexité croissantes des réseaux neuronaux rendent l'entraînement très exigeant en ressources, créant souvent des goulets d'étranglement qui ralentissent les progrès des applications d'IA. Le cours est divisé en deux parties principales : l'amélioration de l'efficacité de l'inférence et l'optimisation du processus d'entraînement.
Dans la première partie, les étudiants se concentreront sur l'amélioration de l'efficacité de l'inférence en évaluant l'efficacité des réseaux neuronaux et en appliquant diverses techniques de compression, telles que le pruning, la factorisation tensorielle et la quantization, afin de créer des modèles plus petits et plus rapides sans perte de précision. La deuxième partie du cours est consacrée à l'optimisation du processus d'entraînement, en abordant les défis liés à la mise à l'échelle et à la complexité de l'entraînement des modèles d'IA modernes. Les étudiants exploreront des techniques telles que les méthodes d'économie de mémoire, y compris la re-materialization (activation checkpointing) et le offloading, ainsi que différents types de parallélisme—data, tensor, model, et pipeline parallelism—qui sont essentiels pour un entraînement efficace.
Le profilage des réseaux neuronaux pour identifier les goulets d'étranglement est souligné tout au long du cours, aidant les étudiants à comprendre et à résoudre les problèmes de performance tant pour l'inférence que pour l'entraînement. À la fin du cours, les étudiants auront acquis les compétences nécessaires pour optimiser la performance des réseaux neuronaux et réussir le déploiement et l'entraînement de modèles d'IA avancés dans des scénarios réels.

Heures d'enseignement

CI	Cours Intégrés	9,33h
TD	Travaux Dirigés	9,33h

Informations complémentaires

Outils pour l'apprentissage profond

Modalités de contrôle des connaissances

Évaluation initiale / Session principale

Type d'évaluation	Nature de l'évaluation	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'évaluation	Note éliminatoire de l'évaluation	Remarques
Projet	Contrôle Continu			1		

Seconde chance / Session de rattrapage

Type d'évaluation	Nature de l'évaluation	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'évaluation	Note éliminatoire de l'évaluation	Remarques
Projet	Rapport			1		

Infos pratiques

Contacts

Yulia Gusak
Yulia.Gusak@bordeaux-inp.fr