2026-2027
Sujet de stage M1/M2
Titre: Conception et Implémentation d'une Arithmétique "Loi des Grands Nombres"

Contexte

Tout ouvrage mathématique d'introduction à l'Intelligence Artificielle (IA) commence par une discussion de la loi des grands nombres et son exploitation pour entraîner des systèmes d'IA, en particulier les perceptrons multicouche (MLP) et réseaux de neurones artificiels profonds (DNN) avec des données expérimentales bruitées [2, 3]. Les bruits entâchant les données sont souvent supposés distribués selon une loi gaussienne, garantissant ainsi une convergence théorique des algorithmes d'entraînement [2, 3].
Une fois l'introduction mathématique faite, les mêmes ouvrages d'IA se passent aussitôt de ces modèles probabilistes. L'entraînement d'un MLP ou DNN devient, en pratique, plus un problème d'optimisation continue qui essaie de minimiser une fonction "de perte" (loss function). Cette optimisation se réalise avec des algorithmes d'analyse numérique classique, comme la descente de gradient [2, 4]. L'implémentation de ces algorithmes d'optimisation numérique se fait en utilisant l'arithmétique en virgule flottante [1] disponibles sur les accélérateurs de calculs (GPUs), dont la performance est le moteur principal du développement rapide de l'IA. Cette arithmétique est une autre source de bruits et d'erreurs, dont la modélisation fait objet de seulement peu de publications [5].
Une modélisation de la nature probabiliste des données bruitées, influençant *in fine* les poids des différents neurones des MLPs, une fois que ceux-ci sont entraînés, est absente dans la littérature. Pourtant, la possibilité de connaître non seulement une valeur (moyenne) d'un poids mais aussi sa distribution, un intervalle d'encadrement ou d'autres moments statistiques ouvrirait la voie à des recherches plus avancées, comme par exemple au contrôle de la collecte de données dans le but d'optimiser l'impact d'un réentraînement.

Objectifs

L'objectif principal du stage est de concevoir, de pratiquement implémenter, intégrer à des suites logicielles existantes et de tester une telle arithmétique "loi des grands nombres". Le stage passera par plusieures phases:

- Compréhension de la littérature existante ainsi que des suites logicielles pour MLPs/DNNs.
- Expérimentation avec des logiciels permettant de remplacer l'arithmétique (flottante) de base par une arithmétique quelconque pour les passes "avant" et "arrière" des DNNs.
- Concéption et implémentation d'un arithmétique "loi des grands nombres" proprement dite.
- Intégration dans une suite logicielle pour entraînement de MLPs/DNNs.
- Tests de performance en termes d'expressivité, de convergence et de rapidité des calculs.
- Rédaction d'un rapport de conception logicielle ou d'une publication scientifique.

Les travaux du/de la stagiaire pourront donner lieu à des publications dans des conférences et revues nationales et internationales, ainsi qu'à des présentations lors des séminaires de l'équipe. La soutenance de stage aura lieu à l'institution d'origine de l'étudiant(e), avec une présence des encadrant(e)s de stage en visioconférence. Les encadrant(e)s américain(e)s délivreront un certificat de réussite de stage détaillé, ainsi que des lettres de recommandation pour la suite des études du/de la stagiaire. Les stages M2 peuvent donner lieu à des propositions de continuation en thèse, sous condition d'adéquation des compétences, des financements disponibles et des règles d'admission de l'École doctorale locale.

Etablissement d'accueil

L'Université du Texas à El Paso (UTEP) est une université publique américaine classée **R1** (à activité de recherche de haute catégorie) par la classification Carnegie, qui distingue les établissements ayant une très forte activité de recherche. Située au Texas, à la frontière entre les États-Unis et le Mexique, elle accueille plus de 26000 étudiants et propose 169 formations de licence, de master et de doctorat. Elle se distingue par son engagement en faveur de l'égalité des chances et de l'accès à un enseignement supérieur de qualité. L'UTEP figure parmi les 5% d'universités américaines les mieux classées en matière de recherche. Son département d'informatique, rattaché au *Miguel A. Loya College of Engineering*, propose des formations en informatique du niveau licence au doctorat, et collabore avec des partenaires académiques et industriels. Ses principaux axes de recherche couvrent le génie logiciel, la cybersécurité, l'intelligence artificielle et les fondements théoriques de l'informatique.

Visa et mobilité

La réalisation du stage à l'Université du Texas à El Paso (UTEP) nécessite l'obtention d'un visa J-1. L'établissement pourra accompagner le/la candidat(e) dans ses démarches administratives, mais les frais liés à la demande de visa et aux formalités associées resteront à sa charge. L'accueil en stage est conditionné à l'obtention du visa requis.

Durée, rémunération et frais de scolarité

Le stage, d'une durée de quatre à six mois, n'est pas rémunéré. Aucun frais de scolarité (tuition fees) ne sera demandé par l'UTEP au stagiaire dans le cadre de ce stage, et une attestation de stage sera délivrée à l'issue de ce dernier cf. ci-dessus). Les encadrant(e)s aideront le/la stagiaire pour toute démarche pour obtenir une bourse de mobilité internationale.

Encadrement

Le (la) stagiaire sera encadré(e) par :

- Christoph Lauter – Maître de conférences en informatique (Associate Professor), HDR, à l'Université de l'UTEP (`https://www.christoph-lauter.org/`)
- Hanane Benmaghnia – Post-Doctorante en validation numérique au LIP6, Sorbonne Université – CNRS, future Chargée de recherche à l'UTEP, (`https://hanane-benmaghnia.github.io/`)
- Martine Ceberio – Professeure en Informatique à l'UTEP (`https://expertise.utep.edu/profiles/mceberio`)

Profil recherché

Le stage s'adresse à des étudiant(e)s (M1 ou M2) en informatique, mathématiques appliquées, électronique ou dans un domaine connexe en premier lieu, mais est ouvert aussi aux étudiant(e)s d'autres disciplines s'ils/elles peuvent démontrer une forte capacité de travail interdisciplinaire. Les compétences recherchées sont :

- Modélisation de problèmes découlant des différentes disciplines des science à l'aide d'outils mathématiques et informatiques.
- Compréhension générale des approches modernes des sciences des données et de l'Intelligence Artificielle, y compris pour les modèles moins repandus commes les modèles de Kolmogorov-Arnold.
- Curiosité d'attaquer un problème de recherche interdisciplinaire dans un contexte international et ouverture d'esprit pour la découverte de la langue anglaise utilisé dans un contexte universitaire de recherche.
- Compétences basique en programmation et en génie logiciel basique.

La curiosité scientifique, l'autonomie, la rigueur et la capacité à analyser des résultats expérimentaux seront particulièrement appréciées.

Modalités de candidature

Les candidatures doivent être envoyées à Christoph Lauter (cqlauter@utep.edu), Hanane Benmaghnia (hanane.benmaghnia@lip6.fr) et Martine Ceberio (mceberio@utep.edu). Elles doivent inclure :

- un curriculum vitae,
- une lettre de motivation,
- des liens vers des projets personnels, professionnels ou universitaires représentatifs, par exemple des dépôts Git (si disponibles) ou publications antérieures.
- une ou deux personnes référentes à contacter.

References

[1] IEEE Standard for Floating-Point Arithmetic.

[2] Ethem Alpaydin. *Introduction to Machine Learning.* The MIT Press, 2nd edition, 2010.

[3] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452, 2015.

[4] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* MIT Press, 2016.

[5] Christoph Lauter and Anastasia Volkova. A framework for semi-automatic precision and accuracy analysis for fast and rigorous deep learning. In *2020 IEEE 27th Symposium on Computer Arithmetic (ARITH)*, pages 103–110, 2020.