

DEMANDE D'ALLOCATION DOCTORALE DE RECHERCHE DE L'ED SIE
Année universitaire 2023/2024
SUJET DE THESE

<p>1. LABORATOIRE</p> <p><i>Nom ou sigle : LISTIC</i> <i>Statut : Unité de recherche USMB</i></p>	<p>2. DIRECTION DE THÈSE</p> <p><i>Directeur de thèse (HDR) : Monnet Sébastien</i> <i>Codirecteur éventuel : Plassart Stéphan</i></p>
<p>Laboratoire partenaire ou collaborations éventuels :</p> 	
<p>3. SUJET DE THÈSE</p> <p>Titre : Vers une IA résiliente et soutenable : Plateformes d'apprentissage automatique distribuées, frugales et soucieuses de l'énergie</p> <p>Title: Towards sustainable and resilient AI: frugal and energy-aware distributed machine learning platforms</p>	
<p style="text-align: center;">4. RESUME <i>(Français et Anglais)</i></p> <p>Aujourd'hui, l'intelligence artificielle (IA) transforme l'informatique et notre vie quotidienne. Dans le cadre des méthodes d'apprentissage automatique pour l'IA, la quantité considérable de données et de calculs requis ne peut plus être gérée par un seul ordinateur. L'apprentissage automatique est donc souvent distribué sur de nombreuses machines. Toutefois, ce type d'apprentissage consomme beaucoup d'énergie. Cette thèse s'attachera à explorer différentes stratégies de répartition des tâches et à étudier leur impact sur la consommation d'énergie et la résilience du système dans son ensemble, le but étant de proposer des approches à la fois frugales et résilientes. De nombreux paramètres seront pris en compte dans ces stratégies, comme le modèle d'allocation de tâches, la topologie du réseau ou du degré de synchronisation. Nous prévoyons également d'anticiper les consommations d'énergie futures et de privilégier, lorsque cela est possible, l'utilisation d'énergie renouvelable.</p> <p>Mots clés : apprentissage distribué, ordonnancement, optimisation d'énergie, résilience</p> <p>Nowadays, artificial intelligence brings a revolution in computer science and in everyone's daily life. While using machine learning (ML) techniques, the huge amount of data and computation required can no longer be managed by a single computer. Machine learning is therefore often distributed across many computers. However, this type of learning consumes a lot of energy. This thesis will explore different task distribution strategies and study their impact on energy consumption and the system's resilience. Our goal is to propose approaches that are both frugal and resilient. Many parameters will be considered in these strategies, such as the task allocation model, the network topology or the synchronization level. We also plan to anticipate future energy consumption and, where possible, to favour the use of renewable energy.</p> <p>Keywords: distributed learning, scheduling, energy optimization, resilience</p>	

5. PROJET DE RECHERCHE DETAILLE

(2 pages environ)

De nos jours, l'intelligence artificielle (IA) révolutionne l'informatique et la vie quotidienne de chacun. Une grande partie de l'IA repose sur l'apprentissage automatique. Cela requiert l'entraînement de modèles sur d'importantes quantités de données, c'est d'ailleurs la quantité phénoménale de données disponibles aujourd'hui qui a rendu l'avènement de l'IA possible. Cependant, l'entraînement d'un modèle d'IA, implique également une quantité de calculs très importante et gourmande en énergie. La gestion de la masse de donnée et des calculs nécessaires à l'entraînement dépassant les capacités de calcul d'une machine, un engouement pour l'apprentissage automatique *distribué* s'est développé. *Distribué* signifie que les traitements s'effectuent sur plusieurs machines (un grand nombre en pratique) interconnectées via un réseau informatique.

Les plateformes sur lesquelles on entraîne des modèles d'intelligence artificielle via de l'apprentissage automatique sont très gourmandes en énergie. Or, il existe de nombreuses façons de répartir efficacement la charge de travail. L'une des familles d'approches est l'apprentissage machine distribué [1], qui décompose les tâches en parties plus petites réparties sur plusieurs machines. Au sein de cette famille, il existe de nombreuses manières de décomposer les tâches. Par exemple : (i) une sous famille consiste à répartir les données entre les machines, chaque machine travaillant alors sur un sous-ensemble des données ; (ii) une autre sous famille consiste à répartir le modèle d'IA entre les machines, le traitement de chaque donnée fera alors intervenir plusieurs machines dans une chaîne de traitement... Une autre façon de répartir le travail est l'apprentissage fédéré [2]. Il s'agit d'une stratégie prometteuse, dans laquelle l'apprentissage du modèle s'effectue localement sur des appareils ou des serveurs périphériques, là où sont présentes les données, ce qui réduit la nécessité d'un transfert de données important et permet d'économiser de l'énergie, tout en préservant la vie privée.

Toutefois, toutes ces stratégies sont gourmandes en données et en calculs, elles sont par conséquent gourmandes en termes de consommation d'énergie, mais, pas de manière équivalente. De plus, pour chaque stratégie, des dizaines de paramètres interviennent dans la vitesse d'apprentissage, la précision (qualité) du modèle produit et la consommation d'énergie. En fonction de la manière dont la charge de travail est distribuée et des paramètres de configuration de la plateforme (le nombre de machine, degré de synchronisation, distribution des calculs, topologie réseau, caractéristiques des systèmes considérés [4], comme la vitesse des processeurs [5] ...), les besoins en énergie peuvent varier considérablement. De plus, selon les machines qui exécutent les traitements, la source d'énergie peut être différente (charbon, nucléaire, solaire, éolien), elle peut également varier dans le temps. En conséquence, l'impact carbone sur la répartition des données peut devenir très important en fonction des modes énergétiques utilisés pour alimenter les différentes machines.

Dans [1], les auteurs ont montré l'impact que pouvait avoir la configuration système de la plateforme de calcul sur les performances de l'apprentissage. Mais ils ne se sont pas intéressés à l'impact sur la consommation d'énergie, ni à la résilience. La résilience est en effet un aspect important : à grande échelle (l'apprentissage automatique s'exécutant potentiellement sur un très grand nombre de machines) les défaillances ne sont plus une exception, mais une caractéristique du système (on parle du "temps moyen entre fautes" d'une plateforme à grande échelle). Un système non tolérant aux défaillances (i.e., non résilient), nécessitant de relancer des calculs et/ou menant à un modèle non précis, et/ou étant beaucoup plus long à entraîner n'est pas une solution acceptable.

Cependant, la modélisation et l'optimisation de la consommation d'énergie dans ces systèmes est une tâche difficile, nécessitant de prendre en compte un grand nombre de paramètres à des niveaux différents (systèmes, réseaux, modèles d'IA, ...). En outre, diminuer la consommation d'énergie tout en maintenant/améliorant la résilience du système aux défaillances (impliquant souvent l'ajout de redondance dans le système) complique encore cette tâche. Certains auteurs, comme [3], se sont intéressés à la fois à la consommation d'énergie mais également à l'impact carbone, cependant ils se sont focalisés uniquement sur le cas du Federated Learning, nous souhaiterions dans cette thèse avoir une analyse plus générale sur l'ensemble des modèles de répartition de tâches.

Cette proposition de recherche vise à explorer différentes approches pour distribuer efficacement les tâches d'apprentissage automatique tout en minimisant la consommation d'énergie, en optimisant la provenance de l'énergie et en améliorant la résilience.

À cette fin, le doctorant devra développer une expertise à la fois en distribution des calculs, en tolérance aux fautes, en ordonnancement de tâches et en apprentissage automatique distribué et fédéré. Il explorera les nombreuses familles de distribution de l'apprentissage automatique, effectuera de nombreuses expérimentations afin d'identifier et de mesurer l'impact des principaux paramètres sur la consommation d'énergie. Nous prévoyons de modéliser la consommation énergétique des mécanismes d'apprentissage automatique distribués en fonction de nombreux paramètres, d'en identifier les principaux et de proposer des bonnes pratiques pour concevoir des approches frugales et résilientes.

Nous prévoyons d'utiliser un simulateur basé sur SimGrid [6], en cours de développement au sein de notre équipe, mais également des expérimentations en conditions réelles sur de grandes plateformes nationales telles Grid'5000 [7] ou Jean Zay [8]. Le simulateur offrira un outil d'estimation de consommation d'énergie de configuration, les expérimentations sur plateformes réelles permettront de le valider et de le paramétrer.

Étant donnée la part importante de l'énergie consommée par les technologies de l'information - TIC (entre 5 et 9% de la consommation totale d'électricité au niveau français et en croissance importante [12,13]) et de l'IA en particulier, de l'impact des TIC sur l'environnement (2,5% de l'empreinte carbone nationale actuellement avec des projections d'augmentation de 45% d'ici 2030 [13]), et étant données les crises énergétiques et climatiques qui se profilent, même une faible amélioration de la consommation des mécanismes d'apprentissage automatique peut avoir un impact économique et sociétal considérable.

- [1] Y. Djebrouni et al. Characterizing distributed machine learning workloads on apache spark: (experimentation and deployment paper). In *Middleware*, pp. 151–164. ACM, 2023.
- [2] Q. Yang et al. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019
- [3] Xinchu Qiu, Titouan Parcollet, Javier Fernández-Marqués, Pedro P. B. de Gusmao, Yan Gao, Daniel J. Beutel, Taner Topal, Akhil Mathur, Nicholas D. Lane: A First Look into the Carbon Footprint of Federated Learning.
- [4] Discover, model and combine energy leverages for large scale energy efficient infrastructures. (Découvrir, modéliser et combiner des leviers énergétiques pour des infrastructures de calculs basse consommation). University of Lyon, France, 2018
- [5] S. Plassart, Online optimization in dynamic real-time systems. Grenoble Alpes University, France, 2020
- [6] Henri Casanova, Arnaud Giersch, Arnaud Legrand, Martin Quinson, Frédéric Suter: SimGrid: a Sustained Effort for the Versatile Simulation of Large Scale Distributed Systems, 2013 (<https://simgrid.org/>)
- [7] Grid'5000 (<https://www.grid5000.fr/w/Grid5000:Home>)
- [8] Jean-Zay (<http://www.idris.fr/jean-zay/jean-zay-presentation.html>)
- [9] Etienne Mauffret, Flavien Vernier, Sébastien Monnet. The burden of time on a large-scale data management service. AINA-2023, Mar 2023, Juiz de Fora, Brazil.
- [10] Etienne Mauffret, Flavien Vernier, Sébastien Monnet. CAnDoR: Consistency Aware Dynamic data Replication. NCA 2019 - 18th IEEE International Symposium on Network Computing and Applications, Sep 2019, Cambridge, MA, United States. pp.1-5, (<http://www.ieee-nca.org/2019/>)
- [11] Wen Sun, Véronique Simon, Sébastien Monnet, Philippe Robert, Pierre Sens. Analysis of a Stochastic Model of Replication in Large Distributed Storage Systems: A Mean-Field Approach. ACM Sigmetrics 2017- International Conference on Measurement and Modeling of Computer Systems, Jun 2017, Urbana-Champaign, Illinois, United States. ACM, pp.51--51, 2017, (<http://www.sigmetrics.org/sigmetrics2017/>). (10.1145/3078505.3078531)
- [12] <https://www.enerdata.fr/publications/breves-energie/consommation-energie-mondiale-digitalisation.pdf>
- [13] https://www.arcep.fr/uploads/tx_gspublication/note-synthese-au-gouvernement-prospective-2030-2050_mars2023.pdf

6. CANDIDAT RECHERCHE : Détailler en quelques lignes vos besoins et les qualités du candidat recherché...

La/le candidat-e sera issu-e d'une formation de niveau master ou équivalent dans le domaine de l'informatique avec de solides bases en programmation, en système et en calcul distribué. Il sera apprécié que la/le candidat-e ait une appétence pour les considérations environnementales.

La/le candidat-e devra également avoir une bonne maîtrise (écrite et orale) de l'anglais.

7. FINANCEMENT DE LA THESE : *Le contrat doctoral fixe une rémunération principale, indexée sur l'évolution des rémunérations de la fonction publique. Arrêté du 26 décembre 2022 modifiant l'arrêté du 29 août 2016 fixant le montant de la rémunération du doctorant contractuel NOR : ESRH2235820A - Article 1 - L'article 1er de l'arrêté du 29 août 2016 susvisé est remplacé par les dispositions suivantes : «Art. 1er. – La rémunération mensuelle minimale des doctorants contractuels est fixée ainsi qu'il suit, sous réserve des dispositions prévues à l'article 2 du présent arrêté :« – à compter du 1er janvier 2023 : 2 044,12 euros brut*

Des heures d'enseignements peuvent être effectuées dans la limite de 64 heures équivalent TD par année universitaire après autorisation du président de l'université et rémunérées au taux fixé pour les travaux dirigés en vigueur. D'autres activités complémentaires au contrat doctoral sont prévues par l'article 5 du décret n° 2009-464 du 23 avril 2009 modifié. La durée totale des activités complémentaires aux activités de recherche confiées au doctorant dans le cadre du contrat doctoral ne peut excéder un sixième du temps de travail annuel.

8. CONTACT :

Nom prénom : Monnet Sébastien / Plassart Stéphane

Tél :

Email : sebastien.monnet@univ-smb.fr / stephan.plassart@univ-smb.fr