2021-03418 - Post-Doctoral Research Visit F/M [Campagne Post-doctorant 2021 – CRI Lille] Parallel surrogate-assisted optimization for automatic hyper-parameter and neural architecture search in Deep Learning (M/F)

Type de contrat : CDD
Niveau de diplôme exigé - Thèse ou équivalent
Fonction : Post-Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 320, including 280 scientists working in fourteen research teams. Recognised for its outstanding contribution to the socle of the Inria research ecosystem, the Centre Inria Lille - Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

Contexte et atouts du poste

Over the last years, Deep Learning (DL) [1] has enabled significant progress in many application fields including image vision, speech recognition, language translation, computer vision, etc. Among the crucial contributing aspects for this progress are the design of novel neural architectures [2] and optimization of their associated hyper-parameters [3] and the advent of ultra-scale GPU-powered supercomputers [4]. Currently employed neural architectures have mostly been developed manually by human experts, which is a time-consuming and error-prone process. Consequently, there is growing interest in automated neural architecture search methods and hyper-parameter optimization. In addition, the rise of DL is continuing to be fuelled by the improvements in accelerators. Important efforts have been directed towards improving the DL methods to deal with neural architecture engineering, hyper-parameter optimization and their GPU-accelerated implementation. However, despite these efforts and the impressive growth of high-performance computing technologies [5], the practical impact of these methods fails to meet expectation due mainly to the huge computational complexity when it comes to deal with big networks [6]. Indeed, dealing with many network layers and millions of hyper-parameters is a tedious complex task.

On the road to the exascale era, the focus of this proposal is put on the design and implementation of ultra-scale approaches for the automatic hyper-parameter and neural architecture search in Deep Learning. The addressed topic is a major part of the research roadmap of the BONUS (Big Optimization and Ultra-Scale computing) team. Moreover, it is related to the CPER Conoëia (2021-2027) and the national ESR Equipee+ MessNet project (2021-2028).

Mission confiée

The success of DL can be attributed to the ability of DL algorithms to automatically extract features from data stored in various formats such as audio, image or text (commonly known as unstructured data). These algorithms allow one to shift from manual feature engineering requiring to spend time on the analysis of data sets to extract new meaningful features to automated process where the time is rather spent on building deep neural network architectures and optimizing their hyper-parameters. Although such automatic optimization allows one to improve the performance of DL it comes however with different optimization problems including mainly the neural architecture search (NAS) [7] and hyper-parameter optimization [3]. The objective of NAS problem is to find the architecture with the lowest validation error. Many methods have been proposed for NAS, including random/grid search, evolutionary algorithms, reinforcement learning, Bayesian optimization, and gradient descent. For the second problem, very often the neural networks are trained using simple heuristics based on gradient descent such as SGD, RMSProp or Adam.

Because of the huge complexity of the whole DL process, in many existing works the two above problems are considered separately. In this Post-Doc proposal, they will be tackled in a joint two-level nested way leading to a big discrete-continuous mixed optimization problem. The starting point will be the approach proposed in the PhD thesis of Léo Souquet [8] in which convolutional neural networks are considered. In this approach, the first level consists in a random exploration of a neural architecture space considering discrete parameters including the number of layers, the filter size, etc. at the second level is moderate-scale multi-core. To deal with large-scale networks with millions of hyper-parameters more advanced parallelism combining multi-core and GPU computing for GPU-powered clusters will be considered.

Another way to deal with the computational burden of DL is to use surrogate-assisted optimization. Bayesian optimization [9] with Gaussian Process surrogate is a leading state-of-the-art method for the optimization of computationally expensive functions, which is precisely the setting of DL model selection. BO is a sequential design strategy requiring mainly a covariance kernel function and an acquisition function (infill criterion). Although BO has seen great success in hyper-parameter optimization for deep models [10], several issues arise when it comes to use BO for NAS. Indeed, while there is a substantial body of work on BO over Euclidean space, using BO for NAS requires so far specifying a distance function between neural architectures, in order to define a surrogate model (kernel function). This is often a cumbersome task involving tuning hyper-parameters of the distance

Informations générales

- Thème/Domaine : Optimisation, apprentissage et méthodes statistiques Calcul Scientifique (BAP 8)
- Ville : Villeneuve d'Ascq
- Centre Inria : CRI Lille - Nord Europe
- Date de prise de fonction souhaitée : 2021-10-01
- Durée de contrat : 1 an, 6 mois
- Date limite pour postuler : 2021-04-25

Contacts

- Équipe Inria : BONUS
- Recruteur : Meliad Nouréddine / Nouréddine.Mela@inria.fr

A propos d’Inria

Inria is the national institute of research dedicated to sciences and technologies of the numerical. It employs 2600 persons. Two 200-employee projects agiles, in general communes with other partners academiques, imply plus of 3500 scientists for release the def of the numerical, pour n’autre d’autres disciplines.

Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-ups. L’institut Cévennes a contribué à la transformation numérique de la science, de la société et de l’économie.

L’essentiel pour réussir

- Being passionate about scientific research, curiosity, interacting with researchers, criticism...
- Feeling at ease in a dynamic scientific environment
- Enjoying learning and listening are essential qualities to succeed in this mission

Consignes pour postuler

CV + application letter + recommendation letters + List of publications

Sécurité défense :

Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention : Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures acceptées par d’autres canaux n’est pas garanti.
In addition, infill criteria will be defined (or redefined from existing ones such as Expected Improvement). Finally, massive parallelism will be investigated at different levels of the optimization process considering the different levels of parallelism provided in large GPU-powered clusters.

For validation, the proposed approaches will be evaluated and extensively experimented considering the optimization of hyper-parameters of Convolutional Neural Networks for computer vision problems using popular datasets such as NASBench dataset [11]. The experiments will be conducted on the hybrid cluster of the Grid'5000 testbed at Lille (CPER Comelia) and on the supercomputer that will acquired within the context of ESR Equipex+ MesoNET project. In addition, the Jean-Zay GPU-powered supercomputer located at IDRIS will also serve as a target experimentation platform. This supercomputer is much larger including more than 2,500 GPUs Nvidia V100 and having 28 PFlops.

**Bibliography**

5. [Top500 international ranking](https://www.top500.org/)

**Principales activités**

Principal activities:
- State of the art on the major topics related to the proposal (Neural Network architecture and hyper-parameters exploration, Parallelization, ...)
- Extension of the work proposed in the PhD thesis of Léo Souquet with parallel advanced metaheuristics for the NAS exploration.
- Design and implementation of parallel Bayesian optimization approach for DL using large GPU-powered clusters.
- Validation through extensive experimentation on extreme-scale supercomputers.
- Publications in high-ranked journals.

Other activities:
- Software packaging and publication on Gitlab
- Participation to scientific animation of the team (organisation of workshops/conferences)
- Participation to Inria Lille - level collective tasks

**Compétences**

Technical skills and required level:
- Being technically familiar with at least one of the topics: Parallel Optimization, Metaheuristics, Surrogate-assisted Optimization and Deep Learning
- Programming (C++/Python, parallel programming libraries)

Language: English

Social skills:
- Spirit of collaboration and sharing

**Avantages**

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

**Rémunération**

Gross monthly salary (before taxes): 2 653 €