A propos du centre ou de la direction fonctionnelle

The PhD candidate will spend about 2/3 of his time at EDF Lab Paris-Saclay and 1/3 at Inria Grenoble (a full year at Inria Grenoble and 2 at EDF Lab Saclay, or regular long stays at Inria Grenoble).

EDF Lab Paris-Saclay is EDF RD&D main research center (see https://www.edf.fr/en/the-edf-group/who-we-are/activities/research-and-development).

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 39 research teams and 8 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Contexte et atouts du poste

The PhD candidate will spend about 2/3 of his time at EDF Lab Paris-Saclay and 1/3 at Inria Grenoble (a full year at Inria Grenoble and 2 at EDF Lab Saclay, or regular long stays at Inria Grenoble). EDF Lab Paris-Saclay is EDF RD&D main research center (see https://www.edf.fr/en/the-edf-group/who-we-are/activities/research-and-development).

Hiring date is flexible. We expect to hire the candidate in September 2020, but we have the possibility to start the contract sooner if we find a good candidate or even later to accommodate some specific situations.

Mission confiée

See next section

Principales activités

PhD subject

Numerical simulations are today commonly used for modeling complex phenomena or systems in different fields such as physics, chemistry, biology or industrial engineering. Some of these numerical simulations require supercomputers to run high-resolution models. In general, a numerical simulation needs a set of input parameters in order to produce the simulation outputs. The input parameters and the often complex internal model produce outputs that can be very large. Analyzing these outputs, tuning the input parameters, or adjusting the internal simulation state based on observation data provided by scientific instruments is still very challenging. Deep learning is emerging as a potentially disruptive approach to address some of these issues. The goal of this PhD is to investigate such approaches in the specific context of on-line learning from ensemble runs.

Very large scale supercomputers have the capacity to support the execution of many instances of these numerical simulations, usually called an ensemble run. Having a large sample of executions is classically used for statistical evaluation of the simulation quality, called a sensitivity analysis. But ensemble runs can also be used to support learning processes.

But storing the results of a large ensemble runs is becoming a strong bottleneck on supercomputers, requiring to look for on-line, also called in transit, data processing solutions. EDF and INRIA developed an original solution for computing on-line statistics for sensitivity analysis from ensemble runs. The framework, called Melissa, is elastic and fault tolerant leading to a very efficient use of available resources.

The objective of this PhD thesis is to extend this approach to on-line deep learning from simulations. Deep Neural Networks (DNN) are inherently on-line data processing machines that easily fit into the Melissa framework. The base ideas is to run many simulations producing data that are used on-line to train a DNN (with the proper parallelization solutions to process the incoming flow of data). We will focus on fast simulations, i.e. compute a NN based substitute of a classical large scale numerical model. A fast simulation, also called meta model, is an approximation of the simulation that is fast to compute. Classical approaches for computing meta-models rely on Kriging or Polynomial Chaos Expansions. But DNN could be used to learn a model of a simulation simply by training it to reproduce the outputs from a set of time indexed input parameters. We started to explore this approach. Melissa has been adapted to implement a DNN with TensorFlow into the Melissa server. The simulations executed provide the data to train the network on-line, without the performance bottleneck caused by intermediate storage. We worked with fluid simulations and managed to have a NN learn simple situations.

Informations générales

- Thème/Domaine : Calcul distribué et à haute performance
- Ville : Saclay & Grenoble
- Centre Inria : INRA Grenoble - Rhône-Alpes
- Date de prise de fonction souhaitée : 2020-09-01
- Durée de contrat : 3 ans
- Date limite pour postuler : 2020-05-31

Contacts

- Equipe Inria : DATAMOVE
- Directeur de thèse : Raffin Bruno / bruno.raffin@inria.fr

A propos d’Inria

Inria, est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 chercheurs et ingénieurs pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’appui à la recherche contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde.

Consignes pour postuler

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.
fluids flows with very high precision. This is a very promising research direction that will be the starting point for this PhD. Mixing conventional numerical simulation and deep learning is a recent but trendy topic. Have a look to the 2019 Deep Fluids paper that investigates a related approach but with off-line learning and targeting computer graphics precisions. At CERN somehow similar approaches are considered using 3DGAN neural architectures as an alternative to compute intensive Monte Carlo Simulations. A second likely scenario that we are likely to investigate in a second phase is dimension reduction. Outputs from numerical simulation are huge and difficult to analyse. Dimension reduction is a classical approach to try to keep the most relevant data only. Auto-encoding deep neural networks are a promising approach for efficient data reduction. The network is trained to reproduce the inputs, but one intermediate layer is limited to a few neurons forcing the network to reduce the data flowing through the network accordingly but without losing the information necessary to reproduce the presented input at the output. Training such network requires sufficient data that can be provided through ensemble runs. CosmoFlow is one example of dimension reduction applied to a large scale cosmology simulation, but trained off-line.

The main common challenges that we will address during this PhD to support these scenarios include:

- DNN architecture design, try to propose generalized solutions. In particular Eulerian numerical simulations are based on a mesh to discretize the space in cells. This discretization is often irregular (cells of shapes) and the cell number very large (millions is common). Traditional NN approaches like convolution networks are designed for regular structures like images, and thus cannot be applied directly here. We will need to develop specific novel adapted approaches.
- Parallel learning strategies that scale. Data parallelism, based on the DNN replication has recently shown that it can scale. As an example, ResNet-50 training evolved from 29 hours on 8 GPUs in 2016 to 224 seconds on 2176 GPUs in 2018. Model parallelism that consists in splitting the NN across different compute nodes is still a challenge but is required when a single DNN is too large to fit in one node. This will likely be a point to be investigated as we target simulations with millions of cells.
- On-line learning differ from the traditional epoch based learning. This requires specific solutions, like replay memories, to avoid induced biases. But replay memories require storage that is not always available, calling for alternative approaches.
- Running a very high number of large simulations is not always feasible. Could we develop parsimonious strategies that require less simulation runs (acting on the sampling strategy for instance or using some GAN like strategies).

This PhD thesis will address these different scenarios within a close collaboration between EDF Lab and INRIA. This PhD thesis is based on a CIFRE contract supported by EDF Lab. The PhD candidate will spend about 2/3 of his time at EDF Lab Saclay and 1/3 at INRIA Grenoble (a full year at INRIA Grenoble and 2 at EDF Lab Saclay, or regular long stays at INRIA Grenoble). We will target high impact international publications at journals and conferences. For experiments we will have access to two EDF supercomputers that rank amongst the 500 most powerful in the world with respectively 41000 international publications at journals and conferences. For experiments we will have access to two EDF supercomputers that rank amongst the 500 most powerful in the world with respectively 41000 and 29000 cores and multiple GPUs (top500.org) as well as the Jean Zay national supercomputer equipped with 1000 GPUs.

To apply submit you CV, references, recent marks, and if available your Master Thesis manuscript.

References :
- Melissa: https://hal.inria.fr/hal-01607479v1 and https://melissa-ca.github.io/
- Towards Efficient Large-Scale Graph Neural Network Computing https://arxiv.org/abs/1810.08403
- End-to-End Differentiable Physics for Learning and Control https://pdfs.semanticscholar.org/093d/f58d33cf05d7e7d9ad38c96d9a492f0394.pdf
- TF-replicator: https://deepmind.com/blog/tf-replicator-distributed-machine-learning/
- TensorFlow: https://www.tensorflow.org/

**Compétences**

This PhD work is at the crossroad between Large Scale Numerical Simulation and Deep Learning. We expect candidates to have some good knowledge for at least one of these domains and the motivation to quickly acquire the missing complementary skills. This research work will involve theoretical developments with validation expriments on supercomputers. Candidates should have the quality required to pursue a successfull research work: technical skills, autonomy, scientific creativity, writing abilities (english), good oral communication skills (english and possibly french), and taste for teamwork.