Proposal for L3 Research Internship for ENS Lyon for 2023

Computability and Complexity Theory for Models of Very Deep Learning.

Olivier Bournez

Keywords: Very Deep Learning Models. Neural Ordinary Differential Equations. Continuous time models of computation. Analog machines and models. Ordinary Differential Equations. Computability. Complexity.

Superviser: Olivier Bournez.

Administrative Position: Professor (of Computer Science) at Ecole Polytechnique.

Location: Laboratoire d'Informatique de l'X, LIX, Ecole Polytechnique , 91128 Palaiseau Cedex

Phone: +33 (0)1 77 57 80 78

Email: olivier.bournez@lix.polytechnique.fr.

Www: http://www.lix.polytechnique.fr/~bournez.

Language: French or English. This proposal is intentionally written in English for Non-French speaking students who may be interested.

General context

With no contests, models and approaches from deep learning have revolutionized machine learning. It is well known that when the number of layers increases (so-called **very deep models**, with sometimes more that 100 or 1000 layers), the models become very hard to train. Among a plethora of options that have been considered, Residual Neural Networks (*ResNets*) [8] have very clearly emerged as an important subclass of models. They mitigate the gradient issues [1] arising when training the deep neural networks. The idea in these particular models is to add skip connections between the successive layers, an idea partially bio-inspired. Since residual neural network was used and won the ImageNet 2015 competition, this particular architecture become the most cited neural network of the 21st century according to some studies (see references in wikipedia). Up to this date, winners of this competition are variations of such models.

Some authors, such as [14], proved that **there is a mathematical explanation** for their performance in practice, as the discrete time process used in these models can be proved to be actually the Euler discretization of some continuous time Ordinary Differential Equation (ODE). The observed obtained robustness and training properties, comes then from the well-known robustness of ODEs with respect to perturbation and with respect to perturbation of their initial conditions.

It was later realized and proved mathematically that **various efficient models are actually nothing but reformulations of discretization schemes for ODEs**. For example, following [12], the architecture of *PolyNet* [15] can be viewed as an approximation to the backward Euler scheme solving the $\mathrm{ODE}\, u_t = f(u)$. *Fractalnet* [11] can be read as a well-known Runge-Kutta scheme in numerical analysis. *RevNet* [6] can be interpreted as a simple forward Euler approximation of some simple continuous dynamical system. All these models are very deep models, but this remains true for simpler models. For example, following [10], it transpires that the key features of well-known *GRU* [5] or an *LSTM* [9], over generic recurrent networks, are updates rules that look suspiciously like discretized differential equations.

This leaded to consider some models such as *neural ODE* [4], which can be seen as continuous versions of *ResNet*. While Neural ODEs do not necessarily improve upon the sheer predictive performance of ResNets, they offer the vast knowledge of ODE theory to be applied to deep learning research. For instance, the authors in [7] discovered that Neural ODEs are more robust for specific perturbations than convolutional neural networks. Moreover, inspired by the theoretical properties of the solution curves, they proposed a regularizer that improved the robustness of Neural ODE models even further. We do not intend to be exhaustive on the various applications of this new point of view on deep learning models.

Description of the work

We are expert of computability and complexity issues related to continuous time models of computation, and in particular models based on ordinary differential equations. In particular, we know how to program with ordinary differential equations, and how to measure complexity for such models: see e.g. [3, 2, 13] for surveys. We used this knowledge in various contexts to solve some open problems in bioinformatics, applied mathematics and other contexts. We propose here to develop this approach to above models of very deep learning.

The purpose of the internship is to review, and contribute to complexity and computability issues for models of very deep learnings. While most of the approaches in the context of deep learning try to learn models, without clear understanding of what is feasible and what is not, the fact that we can actually build on purpose particular ordinary differential solving a given problem do provide some lower and upper bounds on the hardness of the learning process.

The objective will be to develop such results, and provide the basis for a theory for models of very deep learning. Some contributions are indeed possible in the short period of the ENS Lyon internships.

Comment

The actual topic of the work is related to computability and complexity theory. This requires only common and basic knowledge in ordinary differential equations, or in computability or complexity theory, at the level of 1st year of ENS Lyon. Most of the intuitions of our today's constructions come from classical computability and complexity.

Possibilities of funding according to the administrative situation of candidates.

The subject can also be adapted according to requests, knowledge and skills of candidates. Please contact me if interested or in case of questions.

References

[1] David Balduzzi, Marcus Frean, Lennox Leary, JP Lewis, Kurt Wan-Duo Ma, and Brian McWilliams. The shattered gradients problem: If resnets are the answer, then what is the question? In *International Conference on Machine Learning*, pages 342–350. PMLR, 2017.

[2] Olivier Bournez and Manuel L. Campagnolo. New computational paradigms. changing conceptions of what is computable. chapter A Survey on Continuous Time Computations, pages 383–423. Springer-Verlag, New York, 2008.

[3] Olivier Bournez and Amaury Pouly. A survey on analog models of computation. In *Handbook of Computability and Complexity in Analysis*, pages 173–226. Springer, 2021.

[4] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[5] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[6] Aidan N Gomez, Mengye Ren, Raquel Urtasun, and Roger B Grosse. The reversible residual network: Backpropagation without storing activations. *Advances in neural information processing systems*, 30, 2017.

[7] YAN Hanshu, DU Jiawei, TAN Vincent, and FENG Jiashi. On robustness of neural ordinary differential equations. In *International Conference on Learning Representations*, 2019.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[10] Patrick Kidger. On neural differential equations. *arXiv preprint arXiv:2202.02435*, 2022.

[11] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *ICLR*, 2016.

[12] Yiping Lu, Aoxiao Zhong, Quanzheng Li, and Bin Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *International Conference on Machine Learning*, pages 3276–3285. PMLR, 2018.

[13] Pekka Orponen. A survey of continous-time computation theory. In D.-Z. Du and Ker-I Ko, editors, *Advances in Algorithms, Languages, and Complexity*, pages 209–224. Kluwer Academic Publishers, 1997.

[14] E Weinan. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 1(5):1–11, 2017.

[15] Xingcheng Zhang, Zhizhong Li, Chen Change Loy, and Dahua Lin. Polynet: A pursuit of structural diversity in very deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 718–726, 2017.