Deep Genetic Programming

Lino Alberto Rodríguez Coayahuitl, Hugo Jair Escalante, Alicia Morales Reyes

Computer Science Department
National Institute of Astrophysics, Optics and Electronics
Luis Enrique Erro # 1, Santa María Tonantzintla, Puebla, 72840, México

Abstract

We propose to develop a Deep Learning (DL) framework based on the paradigm of Genetic Programming (GP). The hypothesis is that GP non-parametric and non-differentiable learning units (abstract syntax trees) have the same learning and representation capacity to Artificial Neural Networks (ANN). In an analogy to the traditional ANN/Gradient Descend/Backpropagation DL approach, the proposed framework aims at building a DL alike model fully based on GP. Preliminary results when approaching a number of application domains, suggest that GP is able to deal with large amounts of training data, such as those required in DL tasks. However, extensive research is still required regarding the construction of a multi-layered learning architecture, another hallmark of DL.

1. Introduction and Background

In recent years, Artificial Neural Networks (ANN) have automatically produced *representations* (LeCun *et al.*, 2015) that significantly boosted classification systems accuracy (Ciregan *et al.*, 2012; Krizhevsky *et al.*, 2012; Sermanet *et al.*, 2012). These ANN consist of several stacked layers of non-linear transformations, called Deep Neural Networks (DNN), which were previously thought to be too slow to train or converged to very poor solutions when composed of more than two or three layers (Hinton & Salakhutdinov, 2006). Recent advances in the area together with fast technological development now allow to train DNN of several hundred layers (Huang *et al.*, 2016). DNN work by generating a more compact and more abstract representation of the input samples in each forward layer, thus at the final layer data should be clear enough for discrimination by a relatively simple classifier (LeCun *et al.*, 2015). The idea of developing complex, DL-alike, learning models based on structures other than ANN has recently began to be explored. ZH. Zhou and J. Feng proposed a DL system based on Decision Trees (DT) and Random Forests (RF) instead of ANN (Zhou & Feng, 2017, 2018; Feng & Zhou, 2018). They mention different reasons on why it would be desirable to develop powerful learning systems that do not depend on ANN structures, such as the huge amounts of training data required by DNN, the number of hyperparameters that need to be tuned in such networks, and the lack of interpretability (black-box model) of ANN.

In the same vein, we propose in this research a novel DL-alike framework based on Genetic Programming (GP), a bio-inspired algorithmic technique (Koza, 1992). GP is considered a trial and error algorithmic approach that belongs to the class of Evolutionary Algorithms (EA), where a mathematical model or a computer program is built by means of processes that mimic natural evolution such as small mutations of candidate solutions (*individuals*), crossover of good performing individuals and elimination of poor candidate solutions. The fundamental learning structures in GP are abstract syntax trees, which are non-linear, non-parametric, and non-differentiable (just as DT and RF).

Previous efforts to generate representation learning systems that boosted classification or regression algorithms through GP have *hit wall* when faced with high dimensionality problems and large training datasets (Limón *et al.*, 2015; Parkins & Nandi, 2004). Meanwhile, GP systems that are competitive with DNN require implicit use of domain experts' knowledge (Yan *et al.*, 2014; Khmag *et al.*, 2017). In order to tackle such disadvantages among others, we draw inspiration from deep architectures in their successive layers deployment; the proposed GP scheme aims at evolving intermediate data representations to, for example, gradually reduce an initial input representation up to a point where a useful one is achieved, or to clean noisy images through a cascade of sequentially evolved GP image filters. We call this approach *Deep Genetic Programming*; to the authors' best knowledge, no similar approach has been documented before.

2. Current Developments

We consider that some of the main aspects that DNN owe their success to, are: (1) a multi-layered sequential processing topology; (2) an ability to deal with very large training datasets; (3) an efficient implementation in massively parallel architectures such as those found in Graphic Processing Units (GPU). For this work, we focus on the first two aspects. We propose to develop a deep learning framework based on GP capable of: (1) automatically generate necessary internal data representations to boost classification and regression layers/algorithms performance, by transforming raw input training samples into more abstract representations through a cascade processing architecture; (2) dealing with large training datasets, such as those used in DL tasks.

We developed a GP-based, 2-layer, autoencoder algorithm, aiming at dimensionality reduction. We tested this autoencoder with three different image datasets. Evolving through a on-learning mechanism resembling *stochastic gradient descend* is its most important characteristic. This mechanism accelerates the autoencoder evolution one order of magnitude, and thus allowing the use of GP with large training datasets. Detailed results can be reviewed at (Rodriguez-Coayahuitl *et al.*, 2018).

We are currently working on developing a multi-layered, convolutional GP architecture to tackle image denoising tasks, such as white noise filtering, single image super-resolution, and inpaiting. We propose to consider DNN basic architecture and replace neuron-filters on each layer with GP syntax trees. Current challenges we are facing with the proposed architecture are: (1) to develop partial objective functions to assess intermediate representation at internal, or "hidden" convolutional layers, (2) to develop GP operations for individuals fine tuning, since we suspect that most common GP operations are too coarse-grained in comparison to the tuning ability of gradient-descent alike optimization methods used in DNN, and (3) to determine an medium level representation for GP syntax trees in order to extend GP to be able to deal with high dimensionality problems.

Finally, we believe that developing new deep learning models fundamentally divergent from the classical ANN paradigm, might shed light on questions such as how DL really works, and why DL models are so succesful where other ML approaches fall short. Therefore the importance of this research.

References.

References

Ciregan, Dan, Meier, Ueli, & Schmidhuber, Jürgen. 2012. Multi-column deep neural networks for image classification. *Pages 3642–3649 of: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE.

Feng, Ji, & Zhou, Zhi-Hua. 2018. Autoencoder by forest. *In: Thirty-Second AAAI Conference on Artificial Intelligence*.

Hinton, G. E., & Salakhutdinov, R. R. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science*, **313**(5786), 504–507.

Huang, Gao, Sun, Yu, Liu, Zhuang, Sedra, Daniel, & Weinberger, Kilian Q. 2016. Deep networks with stochastic depth. *Pages 646–661 of: European Conference on Computer Vision*. Springer.

Khmag, Asem, Ramli, Abd Rahman, Al-Haddad, SAR, Yusoff, Suhaimi, & Kamarudin, NH. 2017. Denoising of natural images through robust wavelet thresholding and genetic programming. *The Visual Computer*, **33**(9), 1141–1154.

Koza, John R. 1992. *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press.

Krizhevsky, Alex, Sutskever, Ilya, & Hinton, Geoffrey E. 2012. Imagenet classification with deep convolutional neural networks. *Pages 1097–1105 of: Advances in neural information processing systems*.

LeCun, Yann, Bengio, Yoshua, & Hinton, Geoffrey. 2015. Deep learning. *Nature*, **521**(7553), 436–444.

Limón, Mauricio García, Escalante, Hugo Jair, Morales, Eduardo, & Pineda, Luis Villaseñor. 2015. Class-specific feature generation for 1NN through genetic programming. *Pages 1–6 of: Power, Electronics and Computing (ROPEC), 2015 IEEE International Autumn Meeting on*. IEEE.

Parkins, AD, & Nandi, Asoke K. 2004. Genetic programming techniques for hand written digit recognition. *Signal Processing*, **84**(12), 2345–2365.

Rodriguez-Coayahuitl, Lino, Morales-Reyes, Alicia, & Escalante, Hugo Jair. 2018. Structurally Layered Representation Learning: Towards Deep Learning Through Genetic Programming. *Pages 271–288 of: European Conference on Genetic Programming*. Springer.

Sermanet, Pierre, Chintala, Soumith, & LeCun, Yann. 2012. Convolutional neural networks applied to house numbers digit classification. *Pages 3288–3291 of: Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE.

Yan, Ruomei, Shao, Ling, Liu, Li, & Liu, Yan. 2014. Natural image denoising using evolved local adaptive filters. *Signal Processing*, **103**, 36–44.

Zhou, Zhi-Hua, & Feng, Ji. 2017. Deep forest: Towards an alternative to deep neural networks. *arXiv preprint arXiv:1702.08835*.

Zhou, Zhi-Hua, & Feng, Ji. 2018. Deep forest. *National Science Review*.