Non-iterative Learning Approaches and Their Applications

Filippo Maria Bianchi[1] · Ponnuthurai Nagaratnam Suganthan[2]

Optimization, which plays a central role in learning, has received considerable attention from academics, researchers, and domain workers [3]. Many optimization problems in machine learning can be tackled with non-iterative approaches, which can be solved in closed-form manner [4]. Those methods are in general computationally faster than iterative solutions, such as the stochastic gradient descent, used in modern deep learning architectures [2]. Even though non-iterative methods, such as Echo State Networks [1], Extreme Learning Machines [5], and Random Vector Functional Link [6], have attracted much attention in recent years, there exists a performance gap when compared with older methods and other competing paradigms.

The main goal of this special issue is to present recent advances in non-iterative solutions in learning that can reduce such a gap. Secondly, the special issue focuses on showing the important advantages of non-iterative optimization compared with the iterative counterpart, such as gradient-based methods and derivative-free iterative optimization techniques. Besides the dissemination of the latest research results on non-iterative algorithms, the special issue covers practical applications, presents new methodological paradigms, and identifies directions for future studies.

We selected six papers to appear in this special issue. All of them have gone through at least two rounds of revision by two to four expert reviewers that have been carefully selected. One of the papers, coauthored by one of the guest editors, underwent an independent review process to guarantee fairness. The papers published in the special issue are summarized in the following.

✉ Filippo Maria Bianchi
  fibi@norceresearch.no

  Ponnuthurai Nagaratnam Suganthan
  EPNSugan@ntu.edu.sg

[1] NORCE the Norwegian Research Center, Bergen, Norway

[2] Nanyang Technological University, Singapore, Singapore

– In **AEKOC+: Kernel Ridge Regression-based Auto-Encoder for One-class Classification using Privileged Information**, C. Gautam, A. Tiwari, and M. Tanveer proposed AEKOC, a novel AutoEncoder (AE) rchitecture for One Class Classification (OCC). AEKOC uses privileged information, which is generally ignored by traditional machine learning techniques, but it is accounted by human learning processes. The authors combined learning using privileged information (LUPI) with AEKOC and proposed a classifier, which is referred to as AEKOC+. Privileged information is only available during training but not during testing. Therefore, AEKOC is unable to utilize this information for building the model. However, AEKOC+ can efficiently handle the privileged information due to the inclusion of the LUPI framework with AEKOC. Experiments conducted on several downstream tasks demonstrates the superiority of AEKOC+ over AEKOC. The authors concluded that AEKOC+ benefits from the privileged features differently from other available features and improved generalization performance of AEKOC. Moreover, AEKOC+ outperformed two LUPI framework-based one-class classifier, such as OCSVM+ and SSVDD+.

– In **How deep should be the depth of convolutional neural networks: a backyard dog case study**, A. N. Gorban, E. Mirkes, and I. Tyukin first observed that delivering the highest accuracy in a range of operational conditions may not necessarily be always needed, or even achievable due to the lack of data or technical constraints. In face recognition problem, the authors formulate an example of such a usecase, the 'backyard dog' problem. The 'backyard dog', implemented by a lean network, should correctly identify members from a limited group of individuals, a 'family', and should distinguish between them. At the same time, the network must produce an alarm to an image of an individual who is not in a member of the family, i.e. a 'stranger'. To produce such a lean network, a network shallowing algorithm is proposed, which takes an existing deep learning model on its

input and outputs a shallowed version of the model. The algorithm is non-iterative and is based on the Advanced Supervised Principal Component Analysis. Performance of the algorithm is assessed in exhaustive numerical experiments, revealing that the method is capable of drastically reducing the depth of deep learning neural networks, albeit at the cost of mild performance deterioration.

– In **Energy consumption forecasting for the non-ferrous metallurgy industry using hybrid support vector regression with an adaptive state transition algorithm**, Z. Huang, C. Yang, X. Zhou, and S. Yang propose a hybrid support vector regression (HSVR) with an adaptive state transition algorithm (ASTA) named ASTA-HSVR to forecast energy consumption in the non-ferrous metallurgy industry. The proposed support vector regression (SVR) model consists of a linear weighting of -SVR and ν-SVR. The ASTA was developed to optimize the parameters of the HSVR. Experiments are conducted on two cases of energy consumption from the non-ferrous metallurgy industry in China. The results indicate that the ASTA-HSVR method is superior to other state-of-the-art methods for energy consumption forecasting.

– In **Interpreting recurrent neural networks behaviour via excitable network attractors**, A. Ceni, P. Ashwin, and L. Livi aim to open the black-box of recurrent neural networks, an important family of neural networks used for processing sequential data. The authors propose a novel methodology that provides a mechanistic interpretation of behaviour when solving a computational task. The proposed methodology uses mathematical constructs called excitable network attractors, which are invariant sets in phase space composed of stable attractors and excitable connections between them. As the behaviour of recurrent neural networks depends both on training and on inputs to the system, the proposed algorithm extracts network attractors directly from the trajectory of a neural network while solving tasks. Simulations conducted on a controlled benchmark task confirm the relevance of these attractors for interpreting the behaviour of recurrent neural networks, at least for tasks that involve learning a finite number of stable states and transitions between them.

– In **Detecting Gas Turbine Combustor Anomalies Using Semi-Supervised Anomaly Detection with Deep Representation Learning**, W. Yan proposes a deep semi-supervised anomaly detection (deepSSAD) that (i) uses DL to learn representations or features from multivariate, time-series sensor measurements; (ii) uses one-class classification to model normality in the learned feature space, thus performing anomaly detection. Both steps use normal data only; thus, our anomaly detection falls into the semi-supervised anomaly detection category, which is advantageous for industrial asset condition monitoring where abnormal or faulty data is rare. The proposed approach achieves good detection performance on data collected from a real-world gas turbine combustion system. Furthermore, the proposed approach is compared against different designs, including different features (i.e. the deep learned, handcrafted, and PCA features) and different detection models (i.e. one-class ELM, one-class SVM, isolation forest, and Gaussian mixture model).

– In **A novel algorithm for online inexact string matching and its FPGA implementation**, A. Cinti, F. M. Bianchi, A. Martino, and A. Rizzi introduce a novel algorithm for Online Approximate String Matching (OASM), which can filter shadow hits on the fly, according to general purpose priority rules that greedily assign priorities to overlapping hits. An FPGA hardware implementation of OASM is proposed and compared with a serial software version. Even when implemented on entry level FPGAs, the proposed procedure can reach a high degree of parallelism and superior performance in time compared with the software implementation, while keeping low the usage of logic elements. This makes the developed architecture very competitive in terms of both performance and cost of the overall computing system.

References

1. Bianchi FM, Livi L, Alippi C. Investigating echo-state networks dynamics by means of recurrence analysis. IEEE Trans Neur Netw Learn Syst. 2016;29(2):427–439.

2. Bianchi FM, Maiorino E, Kampffmeyer MC, Rizzi A, Jenssen R. Recurrent neural networks for short-term load forecasting: an overview and comparative analysis. Springer. 2017.

3. Bottou L, Curtis FE, Nocedal J. Optimization methods for large-scale machine learning. Siam Rev. 2018;60(2):223–311.

4. Das S, Mullick SS, Suganthan PN. Recent advances in differential evolution–an updated survey. Swarm Evol Comput. 2016;27:1–30.

5. Scardapane S, Comminiello D, Scarpiniti M, Uncini A. Online sequential extreme learning machine with kernels. IEEE Trans Neur Net Learn Syst. 2014;26(9):2214–2220.

6. Zhang L, Suganthan PN. A comprehensive evaluation of random vector functional link networks. Inf Sci. 2016;367:1094–1105.