



# Non-iterative approaches in training feed-forward neural networks and their applications

Xizhao Wang<sup>1</sup> · Weipeng Cao<sup>1</sup>

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Focusing on non-iterative approaches in training feed-forward neural networks, this special issue includes 12 papers to share the latest progress, current challenges, and potential applications of this topic. This editorial presents a background of the special issue and a brief introduction to the 12 contributions.

**Keywords** Non-iterative approaches · Feed-forward neural networks · Neural networks with random weights · Deep learning

In recent years, artificial neural networks have been widely studied and applied in many areas, especially in deep neural networks which have made many breakthroughs in image processing, speech recognition, natural language processing, etc. The training mechanism of traditional neural networks is that all the internal parameters need to be iteratively fine-tuned by using a gradient descent technique (Hinton and Salakhutdinov 2006; Bengio et al. 2013). During the training process, the derivatives of the loss function are continually back propagated to each hidden layer to guide the parameter adjustment until the difference between the model prediction and real observation is small enough. With the increase in the number of hidden layers and the number of hidden layer nodes, this method may cause problems such as slow convergence rate, time-consuming, local minima problem, and excessive resource consumption.

In order to alleviate the above problem, many novel algorithms have been proposed to optimize the training process of traditional neural networks. In this special issue, we focus on non-iterative approaches in training feed-forward neural networks, which include extreme learning machine (ELM) (Huang et al. 2006), random vector functional link network (RVFL) (Pao and Takefuji 1992), neural networks with random weights (NNRW) (Schmidt et al. 1992), etc.

The biggest difference between this type of neural network training and traditional BP-based training is that the hidden nodes in the former neural networks are randomly selected and kept frozen throughout the training process while the output weights are determined analytically. In recent years, the non-iterative-based feed-forward neural networks have received considerable attention, and many advanced variants have been proposed to solve practical problems (Ding et al. 2017; Mao et al. 2017; Liu et al. 2017; Zhai et al. 2017). Lots of experiments and applications have shown that these algorithms can provide good generalization performance with extremely fast speed. Among these non-iterative-based feed-forward neural network training such as ELM and RVFL, the core ideas and training mechanisms are same but some details are slightly different (Cao et al. 2018).

Although the non-iterative-based feed-forward neural networks have achieved many interesting and promising results, there are still some critical issues need to be further studied.

(1) *Rigorous theoretical analysis on the universal approximation and generalization ability of non-iterative-based feed-forward neural networks.* The speed advantage of these algorithms comes from the random generation of hidden nodes. However, there is no such thing as a free lunch, without the iterative tuning process, the hidden nodes may not be able to obtain rich feature information from data. We need to further study from a theoretical point of view and make sure that whether the non-iterative-based feed-forward neural networks have the universal approximation and generalization ability.

(2) *Model stability analysis.* The random feature mapping is the core idea of the non-iterative-based feed-forward

✉ Xizhao Wang  
xizhaowang@ieee.org

Weipeng Cao  
caoweipeng123@gmail.com

<sup>1</sup> College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

neural networks. However, there is still no good way to choose proper hidden nodes and activation function to ensure the high quality of random feature mapping. The applicable scenarios of these algorithms are also needed to be further studied.

(3) *Non-iterative-based deep learning.* One of the reasons for deep learning success is that the deep network structure and its iterative fine-tuning process can make the weights gradually learn high-level abstractions from data. However, the fundamental assumption of non-iterative learning is that not all weights need to be tuned iteratively. The key to building a non-iterative-based deep learning model is to identify the hidden nodes that do not need to be fine-tuned and then use the non-iterative learning strategy to optimize them. However, so far, there has not been an effective way to solve this problem.

This special issue is devoted to revealing the essence of non-iterative-based feed-forward neural network training and sharing the latest progress, current challenges, and potential applications of these algorithms. After a strict peer-review process, 12 papers are selected for publication from 50 submissions (35 online and 15 offline). The accepted papers cover a wide range of topics including the theoretical analysis and algorithm improvement on ELM, the application of non-iterative learning in specific domains, non-iterative deep neural networks, ELM-based semi-supervised learning, and the ensemble learning of ELM. The following is a brief introduction to the 12 accepted papers.

(1) *Four papers focus on the theoretical analysis of ELM and propose new non-iterative algorithms from different perspectives.*

In the paper “Discovering the impact of hidden layer parameters on non-iterative training of feed-forward neural networks” authored by Zhiqi Huang et al., the authors integrate the restricted Boltzmann machine (RBM) and ELM and propose the RBM–ELM algorithm to solve classification problem. In RBM–ELM, the RBM is used as an unsupervised pre-training tool for the input weights and bias of ELM. Compared with the ELM model with randomly assigned parameters, the RBM–ELM can achieve better generalization performance in some cases. The authors also point out that the RBM pre-training process could lead to low variance of the hidden layer matrix, which gives researchers a new direction to analysis the validity of ELM theory.

In the paper “Efficient extreme learning machine via very sparse random projection,” Chuanguan Chen et al. propose a new ELM network structure in which a compression layer is added to the middle of the hidden layer and output layer. The compression layer is used to compress the hidden layer output matrix based on the random sparse-Bernoulli (RSB) projection technique, which can efficiently eliminate the redundant and irrelevant information of the hidden layer and further improve the training speed of ELM.

To improve the robust performance of ELM model, in the paper “Training an extreme learning machine by localized generalization error model,” Hong Zhu et al. develop a novel algorithm named LGE2LM based on the localized generalization error model (LGEM) technique. The LGEM can provide an upper bound of generalization error by the stochastic sensitivity. The stochastic sensitivity plays a regularization role in the LGE2LM model and measures the sensitivity of the model with respect to the input perturbations. Several experiments show that the proposed method has better generalization performance than the original ELM algorithm.

Focusing on the online sequential learning scenarios, in the paper “Fuzziness based online sequential extreme learning machine for classification problems” authored by Weipeng Cao et al., the authors optimize the sequential learning phase of online sequential extreme learning machine algorithm (OS-ELM) from the perspective of fuzziness and propose a new online learning algorithm named FOS-ELM. In FOS-ELM, they first use the existing data to train an initial model and then modify it to be a fuzzy classifier. When new data arrive, the fuzzy classifier will calculate the new samples’ fuzziness value and then only pick out the samples with high-output fuzziness for sequential learning. Compared with the OS-ELM, FOS-ELM can achieve better generalization performance and higher computational efficiency.

(2) *There are two papers in this special issue that apply the non-iterative learning algorithm to specific domains.*

In the paper “Incremental multiple kernel extreme learning machine and its application in Robo-Advisors,” Jingming Xue et al. integrate the multiple kernel learning (MKL) and incremental extreme learning machine (IELM) into a unified framework and propose an incremental multiple kernel extreme learning machine (IMK-ELM) algorithm. The IMK-ELM algorithm automatically increases the hidden layer nodes one by one until the model satisfies the preset condition. The advantage of MKL can further improve the feature extraction capability of the model. The authors have shown the efficiency of IMK-ELM on several benchmark datasets and applied it to solve financial recommendation problem.

The paper “Model-aware categorical data embedding: a data-driven approach” authored by Wentao Zhao et al. proposes a data-driven model-aware embedding algorithm named DAME for categorical data representation, which can simultaneously reflect the data characteristics and optimize the fitness of the represented data for the follow-up learning models such as ELM. The experimental results show that the proposed method can efficiently improve the representation performance for categorical data.

(3) *The iterative-based deep neural networks suffer from some notorious problems such as time-consuming and slow convergence. This special issue includes two papers that use*

*the non-iterative learning strategy to optimize the traditional deep learning model, which effectively improves the training speed of the model.*

In the paper “ELM-based convolutional neural networks making move prediction in Go,” Xiangguo Zhao et al. replace the pooling layers of convolutional neural networks (CNNs) with ELM layers and propose a novel deep neural network model named ECNN. The newly added ELM layers can accelerate the convergence of weights in CNNs, and the model still maintains the efficient feature extraction ability of CNNs. In addition, they apply the ECNN to Go game and several experiments show that the proposed algorithm can effectively make move prediction.

The paper “Weakly-paired multi-modal fusion using deep extreme learning machine” authored by Xiaohong Wen et al. proposes a novel weakly paired multimodal learning framework based on multilayer extreme learning machine (ML-ELM) for object recognition. The ML-ELM can learn the nonlinear representations from weakly paired data and output the corresponding higher-level representations for each modality, which then are conducted joint dimension reduction by weakly paired maximum covariance analysis.

*(4) Semi-supervised learning deals with the learning task by utilizing both the unlabeled samples and the labeled samples. In this special issue, there are two papers that focus on using the ELM to solve the semi-supervised learning problem.*

In the paper “FSELM: fusion semi-supervised extreme learning machine for indoor localization with Wi-Fi and bluetooth fingerprints,” Xinlong Jiang et al. propose a fusion semi-supervised extreme learning machine algorithm (FS-ELM) to solve the indoor localization problem. The FS-ELM combines the Wi-Fi and Bluetooth Low Energy (BLE) signals into a unified model and adopts the semi-supervised manifold regularization to reduce the human calibration effort. The non-iterative-based training mechanism of ELM can make the FS-ELM model have fast learning speed. The experiments on a practical indoor localization dataset show that the proposed algorithm can yield good results on the multi-signal-based semi-supervised learning problem.

The underlying geometrical structure of training data has a significant impact on the performance of semi-supervised extreme learning machine (SSELM). As for this issue, the paper “Adaptive multiple-graph regularized semi-supervised extreme learning machine” authored by Yugen Yi et al. introduces the multiple graph learning approach into SSELM and proposes a new algorithm named adaptive multiple graphs regularized semi-supervised ELM (AMGR-SSELM). AMGR-SSELM combines six different weighted undirected graphs (i.e., kNN-graph, LLE-graph, SG-graph, L2-graph, L1-graph, and LRR-graph) to exploit the intrinsic structure of the data. The underlying data distribution is well exploited through the combination of these different graphs. The semi-

supervised ELM is used as a fast classifier to solve the “out-of-sample” problem.

*(5) Ensemble learning is an efficient way to improve the model stability. This special issue includes two papers that focus on using the ensemble learning strategy to improve the performance of non-iterative learning algorithms.*

In the paper “Fuzzy integral based ELM ensemble for imbalanced big data classification,” Junhai Zhai et al. combine the oversampling technique, ELM algorithm, and fuzzy integral-based ensemble learning to propose a novel algorithm named MR-FI-ELM for imbalanced big data classification. The training process of MR-FI-ELM includes three stages: step1, oversample the positive instances within the hypersphere; step 2, construct  $K$  balanced data subsets and train  $K$  initial classifiers based on ELM algorithm; step 3, combine the  $K$  initial classifiers by fuzzy integral to obtain the final model. The experiments have shown that the MR-FI-ELM algorithm can effectively classify imbalanced big data.

The paper “Data-driven prediction model for adjusting burden distribution matrix of blast furnace based on improved multi-layer extreme learning machine” authored by Xiaoli Su et al. proposes an ensemble learning framework named EPLS-ML-ELM based on the partial least square multilayer extreme learning machine (PLS-ML-ELM). The EPLS-ML-ELM inherits the advantages of the PLS-ML-ELM and ensemble learning strategy, which not only can effectively deal with the multi-collinearity problem, but also can improve the model stability. For the regression problem, the final prediction result comes from the average of the  $K$  models. For the classification problem, the final result is determined by the maximum voting strategy.

We hope that the accepted papers in this special issue are helpful for readers to understand the essence of non-iterative approaches in training feed-forward neural networks.

**Acknowledgements** We would like to thank all the authors and reviewers for their contributions to this special issue. We also sincerely thank Prof. Antonio Di Nola, the Editor-in-Chief of Soft Computing, for his support to edit this special issue.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35(8):1798–1828
- Cao WP, Wang XZ, Ming Z, Gao JZ (2018) A review on neural networks with random weights. *Neurocomputing* 275:278–287
- Ding S, Zhang N, Zhang J, Xu X, Shi Z (2017) Unsupervised extreme learning machine with representational features. *Int J Mach Learn Cybern* 8(2):587–595
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- Huang GB, Chen L, Siew CK (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans Neural Netw* 17(4):879–892
- Liu M, Liu B, Zhang C, Wang W, Sun W (2017) Semi-supervised lowrank kernel learning algorithm via extreme learning machine. *Int J Mach Learn Cybern* 8(3):1039–1052
- Mao W, Wang J, Xue X (2017) An ELM-based model with sparse weighting strategy for sequential data imbalance problem. *Int J Mach Learn Cybern* 8(4):1333–1345
- Pao YH, Takefuji Y (1992) Functional-link net computing: theory, system architecture, and functionalities. *Computer* 25(5):76–79
- Schmidt WF, Kraaijveld MA, Duin RP (1992) Feedforward neural networks with random weights. In: *Proceedings of the 11th IAPR international conference on pattern recognition, 1992. Vol II. conference B: pattern recognition methodology and systems, IEEE*, pp 1–4
- Zhai J, Zhang S, Wang C (2017) The classification of imbalanced large data sets based on mapreduce and ensemble of elm classifiers. *Int J Mach Learn Cybern* 8(3):1009–1017

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.