Editorial: Randomized Algorithms for Training Neural Networks

Article in Information Sciences · May 2016

DOI: 10.1016/j.ins.2016.05.021

1 author:

Dianhui Wang
La Trobe University

Information Sciences 000 (2016) 1–3

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Editorial

Editorial: Randomized algorithms for training neural networks

Randomness is useful in numerical computations and it has received considerable attention in the past decades [9,10,18]. Randomised algorithms (RAs) offer fast solutions for problem solving with statistical characterization. In machine learning and computational intelligence, to the best of our knowledge, research on randomized algorithms for training neural networks can be tracked back to 80s. In [1], Broomhead and Lowe explored the possibility and limits of randomised scheme for training radial basis function (RBF) networks, where centers of RBFs were either randomly assigned without dependence on the training dataset or randomly selected from the training dataset whereas the output weights were obtained by least squares method (i.e., computing the pseudo inverse of the output matrix at the hidden layer). However, for some cases this solution is unsatisfactory and the resulted learner may perform poorly in generalization. From model perspectives, Pao and Takefji proposed random vector functional-link (RVFL) nets in [11], where the hidden layer parameters were randomly generated and then fixed during learning process. Distinguished from the single hidden layer feedforward neural networks (SLFNs), RVFL nets have a direct link from the input to the output, its output weights have a close form solution and can be calculated by using the pseudoinverse method done in a similar fashion as in [1] for RBF networks. Later on, Pao and his co-workers looked into the learning and generalization aspects for this type of randomised learner model in [12]. Indeed, a similar idea of such a randomized approach for neural networks was also proposed in [16], where Schmidt et al. suggested a random assignment of input weights and biases in $[-1, 1]$ with the uniform distribution for SLFNs without any direct link from the input to the output nodes.

A common nature of these pioneering works reported in [1,11,12,16] is that all of them are non-iterative solutions for training SLFNs. By doing this, one expected to overcome some shortcomings from gradient-based training algorithm through random assignment on the input weights and biases and determination of the output weights by least squares methods. However, these studies were carried out at the algorithmic level and lacked theoretical justification on the universal approximation capability of the randomised learner models until 1995. Some new developments on randomized algorithms for neural networks can be found in [2,13,14,17], where statistical properties of learning processes, kernel method and model's feasibility for data modelling are addressed. In [6,8], recurrent randomised learner models, termed as echo state networks (ESNs) and liquid state machines (LSMs), were developed for resolving real-time data modelling problems with uncertain dynamic orders. In this framework, the input weight matrix and feedback matrix can be randomly assigned subjected to stability constraint, and the output weights can be obtained by least squares methods, as done in [1,11,12,16].

Theoretical results on randomised learner models and algorithms mainly focused on the universal approximation capability [4,5], stability [3,7] and learnability [15]. In 1995, Igelnik and Pao established some fundamental results on the universal approximation capability of RVFL nets [5], where their mathematical proofs are constructive and clearly indicated that the setting of random input weights and biases of hidden nodes must be specified. In [4], Husmeier revisited the universal approximation theorem given in [5] and showed that RVFL nets can approximate any Lipschitz functions in the sense of probability, where the scope of random input weights and biases is characterised by a symmetric interval. In [3,7], various stabilities of randomised learning algorithms have been researched from statistical learning perspectives, aiming to develop a better understanding on the generalization of randomized models. In [15], Ross and Bagnell investigated the stability conditions for online learnability, which is an important concept for streaming data modelling.

This special issue is composed of eight contributions, including a survey paper, a theoretical work on feasibility of randomised learner models, four papers related to randomized algorithms for training neural networks, and two application papers. All papers published in this special issue have gone through formal and rigorous reviews, and eventually got accepted for publications after extensive reviews and revisions. These papers show some promising advancements of randomized approaches for neural networks, however it should be pointed out that there are many interesting and challenging

http://dx.doi.org/10.1016/j.ins.2016.05.021

problems uncovered by this special issue. As a special issue editor, I tried my very best to make this special issue informative, useful and valuable to readers who have experienced this method in the past or will be exploring the randomised learning techniques in the future for data processing, in particular, large scale or streaming data modelling. The following briefly introduces these papers.

The paper entitled "Approximation with Random Bases: Pro et Contra", by Alexander N. Gorban, Ivan Yu. Tyukin, Danil V. Prokhorov, and Konstantin I. Sofeikov, investigates a problem of selecting suitable approximators from families of parameterized basis functions that are known to be dense in a Hilbert space of functions. Both randomized and deterministic methods for selecting elements from these families are studied, and it has been shown that the rate of convergence in L_2 norm of order $O(1/N)$, where N is the number of elements. To ensure successful data modelling with a class of randomized approximators, additional conditions on the families of functions to be approximated are requested and provided. Theoretical analyses on the feasibility of randomized basis function approximators have been given, claiming that in the absence of such additional conditions one may observe exponential growth of the number of terms needed to approximate a nonlinear map, and/or the resulting learner model will be extreme sensitivity to the parameters. Implications of this study for applications of neural networks in modelling and control are demonstrated by some examples.

The paper entitled "A Survey of Randomized Algorithms for Training Neural Networks", by Le Zhang and Ponnuthurai N. Suganthan, presents an extensive survey on randomized methods for training neural networks, including multilayer perceptrons (MLPs) with a single hidden layer, random vector functional link (RVFL) nets, recurrent neural networks (RNNs, i.e., echo state networks (ESNs) and liquid state machines (LSMs)) and kernel-based learning machines. This paper offers some valuable insights into this important research topic and provides with some potential research directions. It is believed that this paper contributes to both literatures review and clarification on historical developments and trends of randomized approaches for neural networks. This survey paper summaries milestones about randomized learner models, theoretical studies on approximation property, various stabilities of learning with associated generalization capability, learning algorithms and applications. This survey paper provides readers with a clear picture on the history and progresses of randomized approaches for neural networks, and also encourages further advancements in this field.

The paper entitled "A Semi-supervised Random Vector Functional-Link Network based on the Transductive Framework", by Simone Scardapane, Danilo Comminiello, Michele Scarpiniti and Aurelio Uncini, considers a problem of semi-supervised learning (SSL) based on RVFL networks with transductive learning schemes. The proposed algorithm in this paper can achieve state-of-the-art performance in terms of recognition rate, and much efficient than other learner models such as SVMs. It has been shown that, thanks to the characteristics of RVFL networks, the resulting optimization problem can be safely approximated with a standard quadratic programming problem which is solvable in polynomial time. Comprehensive experiments with comparisons demonstrated the effectiveness of the proposed RVFL-based semi-supervised algorithm with the theory of manifold regularization.

The paper entitled "Dynamic neural modeling of fatigue crack growth process in ductile alloys", by Linsen Xie, Yi Yang, Zhenghua Zhou, Jinchuan Zheng, Mengqiu Tao, and Zhihong Man, develops a dynamic neural networks (DNNs) with random weights for modelling fatigue crack growth process in ductile alloys. It is shown that a fatigue crack growth process can be viewed as a virtual nonlinear dynamical system, which may be modelled through learning the dynamics of crack opening stress and crack length growth, respectively. The DNNs are constructed by adding the tapped-delay-line memories to both the input and the output layers of conventional single layered feed-forward neural networks (SLFNs). Since the delayed output feedback components are placed in parallel with the hidden nodes, a generalized hidden layer is formulated. Experimental results on data of 2025-T351 and 7075-T6 aluminium alloy specimens show that the well-trained DNN model with random weights is capable of capturing all dynamic characteristics of crack growth process.

The paper entitled "Modelling Neural Plasticity in Echo State Networks for Classification and Regression", by Mohd-Hanif Yusoff, Joseph Chrol-Cannon, and Yaochu Jin, investigates the influence of neural plasticity applied to the weights inside the reservoir on the learning performance of echo state networks (ESNs). This paper examines the influence of two plasticity rules, anti-Oja's learning rule and the Bienenstock–Cooper–Munro (BCM) learning rule on the prediction and classification performance when either offline or online supervised learning algorithms are employed for training the read-out connections. Empirical studies are conducted on two widely used classification tasks and two time-series prediction problems. Experimental results demonstrate that neural plasticity can effectively enhance the learning performance when offline learning is applied. Their results also indicate that the BCM rule outperforms the anti-Oja rule in improving the learning performance of the ENS in the offline learning mode.

The paper entitled "Randomized Algorithms for Nonlinear System Identification with Deep Learning Modification", by Erick de la Rosa and Wen Yu, presents a hybrid training scheme, incorporating randomized algorithms into deep learning techniques, for data regression and classification problems using neural networks. The distributions of the hidden weights are obtained by the restricted Boltzmann machines. This deep learning method uses input data to construct the statistical features of the hidden weights. The output weights of the neural model are computed by solving a least squares problem. This work successfully combines the deep learning and the randomized algorithm, and take advantages from both of them. Experimental results on three benchmark datasets are reported to support the effectiveness of the proposed deep learning with the randomized algorithms.

The paper entitled "SOM: stochastic initialization versus principal components", by Ayodeji A. Akinduko, Evgeny M. Mirkes, and Alexander N. Gorban, investigates the impact of initialization of weights in Self-Organising Maps (SOMs) on the learning performance. Indeed, this is a common problem for all iterative methods of data approximation. The quality of the resulting data approximation depends on the initial approximation. This study reveals that results obtained from the principal component initialization are controversial, although such a method is popular and widely used as an initial approximation for nonlinear dimensionality reduction because its convenience and exact reproducibility of the results. Concretely, for their quasilinear datasets the principal component initialization of the self-organizing maps is systematically better than the random initialization, whereas for the essentially nonlinear datasets the random initialization may perform better.

The paper entitled "Fuzzy Nonlinear Regression Analysis Using a Random Weight Network", by Yu-Lin He, Xi-Zhao Wang, Joshua Zhexue Huang, addresses a problem of fuzzy nonlinear regression (FNR) using random weight networks. Random Weight Networks (RWNs) are employed to develop a FNR model named $FNRRWN_{2015}$ in which both inputs and outputs are triangular fuzzy numbers. The input-layer weights and hidden-layer biases of $FNRRWN_{2015}$ are randomly assigned and fixed during learning processing, and $FNRRWN_{2015}$'s output-layer weights are analytically calculated based on the derived updating rule which is to minimize the integrated squared error between the predicted fuzzy outputs and target fuzzy outputs, which can be approximately resolved by means of Riemann integral theory. Experimental results indicate that the proposed $FNRRWN_{2015}$ can effectively approximate a fuzzy-in fuzzy-out system with favourable performances in terms of prediction accuracy and computational burden, compared against these FNR models based on MLP with back-propagation algorithm and RBF networks.

This special issue aims to promote randomized approaches for neural networks, providing both historical developments with milestone results and new trends in this research topic. I truly respect advancements of knowledge on randomized learner models and randomised learning algorithms, theory and applications. Due to the use of randomness in the learner model building, this type of neural networks with random weights must be viewed as random variables, but not deterministic models. Thus, performance evaluations on learning ability and predictability (i.e., generalization) will have to be carefully considered because they are closely associated with the distribution of random parameters of the model [4,5]. The selection of random parameters, including both the scope and type of distribution of random parameters, plays a key role for successful applications in practice. Therefore, randomized algorithms for training neural networks have very clear objectives, that is, how to define constraints on these random parameters in a specific range (unknown as well) so that the resulting randomized learner can perform properly in terms of both learning and generalization. Theoretically, stability concepts proposed and studied in [3,7,15] have links to measuring the generalization power of randomised learner models. Unfortunately, these established results have less practical value because of the assumption on infinite number of samples. Therefore, it will be more meaningful and useful to develop some practical criteria on these stabilities with limited training samples in the further studies.

In the end, I take the opportunity to express my sincere appreciation to the authors for their contributions to this special issue, and also to the reviewers for their tremendous effort during several rounds of paper reviews. Lastly, I would like to thank the Editor-in-Chief of Information Sciences, Professor Witold Pedrycz, for his great support throughout the development of this special issue.

References

[1] D.S. Broomhead and D. Lowe, Radial basis functions, multi-variable functional interpolation and adaptive networks. DTIC Document, 28 March, 1988.
[2] A. Elisseeff, H. Paugam-Moisy, JNN, a randomized algorithm for training multilayer networks in polynomial time, Neurocomputing 29 (1–3) (1999) 3–24.
[3] A. Elisseeff, T. Evgeniou, M. Pontil, Stability of randomized learning algorithms, J. Mach. Learn. Res. 6 (1) (2005) 55–79.
[4] D. Husmeier, Random vector functional link (RVFL) networks, Neural Networks for Conditional Probability Estimation: Forecasting Beyond Pint Predictions, Springer, 1999 (Chapter 6).
[5] B. Igelnik, Y.-H. Pao, Stochastic choice of basis functions in adaptive function approximation and the functional-link net, IEEE Trans. Neural Netw. 6 (6) (1995) 1320–1329.
[6] H. Jaeger, W. Maass, J. Principe, Special issue on echo state networks and liquid state machines: Editorial, Neural Netw. 20 (3) (2007) 287–289.
[7] K. Luo, Z.Y. Jia, W. Gao, Strong and weak stability of randomized learning algorithms, in: Proceedings of 2012 IEEE 14th International Conference on Communication Technology, Nov. 2012, 2012, pp. 9–11.
[8] W. Maass, T. Natschlager, H. Markram, Real-time computing without stable states: a new framework for neural computation based on perturbations, Neural Comput. 14 (11) (2002) 2531–2560.
[9] M.W. Mahoney, Randomized algorithms for matrices and data, Found. Trends Mach. Learn. 3 (2) (2011) 123–224.
[10] R. Motwani, P. Raghavan, Randomized Algorithms, Chapman & Hall/CRC, 2010.
[11] Y.-H. Pao, Y. Takefji, Functional-link net computing, IEEE Comput. J. 25 (5) (1992) 76–79.
[12] Y.-H. Pao, G.-H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional link net, Neurocomputing 6 (2) (1994) 163–180.
[13] A. Rahimi, B. Recht, Random features for large-scale kernel machines, in: Proceedings of International Conference on Advances in Neural Information Processing Systems, 2007, pp. 1177–1184.
[14] A. Rahimi, B. Recht, Weighted sums of random kitchen sinks: replacing minimization with randomization in learning, in: Proceedings of International Conference on Advances in Neural Information Processing Systems, 2009, pp. 1313–1320.
[15] S. Ross, J.A. Bagnell, Stability Conditions for Online Learnability, The Computing Research Repository (CoRR), August 2011.
[16] F. Schmidt, M.A. Kraaijveld, R.P. Duin, Feedforward neural networks with random weights, in: Proceedings of the 11th IAPR International Conference on Pattern Recognition, Vol. II. Conference B: Pattern Recognition Methodology and Systems, 1992, pp. 1–4.
[17] I.Y. Tyukin, D.V. Prokhorov, Feasibility of random basis function approximators for modeling and control, in: Proceedings of IEEE Conference on Control Applications & Intelligent Control, 2009, pp. 1391–1396.
[18] H. Zenil, Randomness Through Computation: Some Answers, More Questions (Ed.), World Scientific Publishing Company, February 11, 2011.

Dianhui Wang
School of Engineering and Mathematical Science, La Trobe University, Melbourne, Victoria 3086, Australia

E-mail address: dh.wang@latrobe.edu.au