Neurocomputing
Volume 350, 20 July 2019, Pages 70-90

Towards a more efficient and cost-sensitive extreme learning machine: A state-of-the-art review of recent trend

Author links open overlay panelPeter Adeniyi Alaba [a], Segun Isaiah Popoola [b], Lanre Olatomiwa [c] [d], Mathew Boladele Akanle [b], Olayinka S. Ohunakin [e] [h], Emmanuel Adetiba [b] [f], Opeoluwa David Alex [g], Aderemi A.A. Atayero [b], Wan Mohd Ashri Wan Daud [a]
https://doi.org/10.1016/j.neucom.2019.03.086Get rights and content

Highlights

- Despite the prominence of ELM, it is still limited in certain aspects.
- ELM architecture can be determined using Incremental and/or pruning techniques.
- Embedded dissimilarity can solve prediction instability and imbalance distribution.
- RO projection technique improved the capability of SSP for any training dataset.
- ELM can accommodate lateral inhibition with the use of sparse coding method.

Abstract

In spite of the prominence of extreme learning machine model, as well as its excellent features such as insignificant intervention for learning and model tuning, the simplicity of implementation, and high learning speed, which makes it a fascinating alternative method for Artificial Intelligence, including Big Data Analytics, it is still limited in certain aspects. These aspects must be treated to achieve an effective and cost-sensitive model. This review discussed the major drawbacks of ELM, which include difficulty in determination of hidden layer structure, prediction instability and Imbalanced data distributions, the poor capability of sample structure preserving (SSP), and difficulty in accommodating lateral inhibition by direct random feature mapping. Other drawbacks include multi-graph complexity, global memory size, one-by-one or chuck-by-chuck (a block of data), global memory size limitation, and challenges with big data. The recent trend proposed by experts for each drawback is discussed in detail towards achieving an effective and cost-sensitive model.

Introduction

Extreme learning machine (ELM) is a kind of neural network (NN) characterized by biologically inspired single-hidden-layer feedforward network (SLFN), using biological learning techniques rather than artificial learning techniques. It is a biological learning technique that involves the use of kernels, random neurons (with or without unknown modeling/shape), and optimization constraint. ELM is more effective in terms of speed, generalization performance, simplicity and efficiency than the traditional NN in practical applications. The word “extreme” implies beyond conventional artificial learning methods, towards brain-like learning [1]. ELM helps to fill the gap between biological learning and machine learning mechanism [1,2]. Rather than using known activation function such as sigmoid, ELM uses unknown nonlinear piecewise continuous functions h(x) being the real activation functions of most living brain neurons [1]. Theoretically, ELM somehow combines brain learning features, matrix theory, control theory, neural network theory, and linear system theory, which were previously regarded to be isolated with big gaps.

Due to the capability for a wide range of activation functions h(x), ELM exhibits universal classification capability and universal approximation capability [1]. ELM can be used in solving problems pertaining to regression, classification, representational learning, feature selection, clustering, and several other learning tasks. Successful applications of ELM have been reported in several domains, such as output power forecasting [3], system identification [2], function approximation [4,5], biomedical engineering [2], biological information processing, data classification [6], computer vision, pattern recognition [7,8], robotics and control [2]. ELM generates the input layer (sensory layer) weights and the hidden nodes biases randomly and determines the output layer weights rationally by solving a generalized inverse matrix. The study of Huang et al. [9,10], substantiated that SLFNs with randomly generated hidden node parameters and with radial or additive basis function hidden nodes could function as universal approximators. This is achievable by simply computing the output weights, which link the hidden layer (associator layer) to the output nodes, thereby substantiating its wide application in solving regression and classification problems [11]. Huang, Zhu and Siew [12] and Liang, Huang, Saratchandran and Sundararajan [13] also established that iterative methods are not needed at all for adjustment of SLFNs parameters in ELM, unlike NN, whose network parameters require iterative methods due to the extensive usage of gradient-based learning algorithms [14]. In comparison with other learning techniques, such as back-propagation algorithm (BP), various variants of BP and deep learning algorithms, support vector machine (SVM), the key superiority of ELM is that it does not involve iterative tuning of the parameters [15].

Despite the prominence of ELM model, as well as its excellent features such as insignificant intervention for learning and model tuning, the simplicity of implementation, and high learning speed, which makes it a fascinating alternative method for Artificial Intelligence, including Big Data Analytics, it is still limited in certain aspects. Huang et al. [2] suggested further research on high dimensional data analysis. The network structure of ELM is more complex for large data since the hidden biases and input weights are randomly selected, making the traditional ELM to require more hidden nodes, thereby affecting the network generalization ability. The order of complexity of the algorithm for output weights estimation is $O(M^2K)$, where M is the number of hidden units and K is the number of training points [16]. Furthermore, some ELM could be time-consuming and ineffective due to the predetermination of their network size when trial by error is used before training [13,14]. ELM uses of batch training approach, indicating that the network is trained with all the dataset at once, thereby requiring large processing power and memory. Classical ELM is also plagued with prediction instability. It occurs because of random initialization of the hidden layer biases and the input weights, and imbalanced data distributions. ELM has no mechanism that takes care of imbalanced data distributions that may be encountered in many fields [17] since it is assumed that every single class size is comparatively balanced and the costs of misclassification are equal in the entire datasets [18]. The imbalanced class distribution could make the classification mechanisms learn very complex models, thereby over-fitting [19].

Moreover, despite the recent prominence of ELM due to remarkable learning speed, little or no manual intervention, and good generalization performance, the generalization performance could be worse than that of SVM algorithm for small sample cases or small network size. This is ascribed to the use of Monte Carlo (MC) sampling technique for generation of random input weights. ELM models with small network size exhibit random input weight with poor SSP capability, leading to poor generalization performance [20].

ELM cannot accommodate lateral inhibition by direct use of random feature mapping [21]. This is a serious challenge in learning of large-scale data since ELM was initially developed to run on a machine with a single processing unit. Handling a block of data, or incremental training samples (one-by-one or chunk-by-chunk) is another serious drawback in ELM. Incremental training samples (such as dynamic changes of tidal level) are presented chunk-by-chunk or one by one and they involve time-varying dynamics. Therefore, modeling time-varying process becomes challenging in real-time [22].

This review discussed the major drawbacks of ELM, which affect its efficiency and cost, as well as the current techniques used as the panacea. The drawbacks addressed here include difficulty in determination of hidden layer structure, prediction instability and Imbalanced data distributions, the poor capability of sample structure preserving (SSP), and difficulty in accommodating lateral inhibition by direct random feature mapping. Other drawbacks include multi-graph complexity, global memory size, one-by-one or chuck-by-chuck (a block of data), global memory size limitation, and challenges with big data. The recent trend proposed by experts for each drawback are discussed in detail towards achieving an effective and cost-sensitive model.

Access through your organization
Check access to the full text by signing in through your organization.
Access through **Covenant University**

Section snippets

Extreme learning machine

ELM developed by Huang at el [12,23] utilizes Single Layer Feedforward Neural Network (SLFN) Architecture [24]. ELM engages a random selection of input weights to rationally calculate the output weights of SLFN. The generalization performance of ELM is remarkable with a high learning speed. ELM does not require much human intervention and the learning speed is thousands time faster when compared with the conventional techniques. The rational determination of the network parameters is automatic,

Difficulty in the determination of network architecture

Due to the random selection of hidden biases and input weights, the traditional ELM certainly requires more hidden nodes, making the network structure more complex for large data, thereby affecting the network generalization ability. The order of complexity of the algorithm for output weights estimation is $O(M^2K)$, where M is the number of hidden units and K is the number of training points [16]. Furthermore, some ELM could be time-consuming and ineffective due to the predetermination of their

Prediction instability and imbalanced data distributions

Prediction accuracy is certainly a crucial measurement of models in risk analysis [19]. Prediction instability is one of the major limitations of ELM [19,60]. It occurs because of random initialization of the hidden layer biases and the input weights, and imbalanced data distributions. ELM has no mechanism that takes care of imbalanced data distributions that may be encountered in many fields [17] as it is assumed that every single class size is comparatively balanced and the costs of

Poor capability of sample structure preserving (SSP)

Despite the recent prominence of ELM due to remarkable learning speed, little or no manual intervention, and good generalization performance, the generalization performance could be worse than that of SVM algorithm for small sample cases or small network size. This is ascribed to the use of Monte Carlo (MC) sampling technique for the generation of random input weights. ELM models with small network size exhibit random input weight with poor SSP capability, leading to poor generalization

Difficulty in accommodating lateral inhibition by direct random feature mapping

The random feature mapping involves the introduction of optimization constraint (Fig. 7), which is the fundamental features of ELM. Feature mapping in ELM enhances its capacity for universal approximation and efficiency in training. Furthermore, random feature mapping enhances generalization performances and reduces over-fitting [2]. Physiological research revealed that similar layer neurons are laterally inhibited to each other such that the outputs of each layer are sparse [86]. Meanwhile, it

Challenges with big data

The global interest in Big data, as well as its economic value and adoption, have continued to grow exponentially [88]. The advent of emerging technologies (such as Internet of Things (IoT) [89,90], Machine-to-Machine (M2M) communications [91], and cloud/fog computing [92], [93], [94]), and the high proliferation of smart mobile devices [95] has contributed immensely to the evolution of the Big data era. ELM has been widely and successfully applied to gain useful insights from data obtained in

More on big data: acceleration of ELM for practical big data tasks

To extend the capability of ELM for handling large scale or big data tasks (such as image classification, voice recognition, and object detection & tracking), authors in [145,146] proposed multilayer or hierarchical ELM (H-ELM) frameworks that are based on the universal approximation capability of the original ELM. Thus, H-ELM has extended the original ELM from shallow architecture to deep architecture, which has potentially provided a boost to the capability of ELM. Generally, deep

Difficulties in handling a block of data

Handling a block of data, or incremental training samples (one-by-one or chunk-by-chunk) is another serious drawback in ELM. Incremental training samples (such as dynamic changes of tidal level) are presented chunk-by-chunk or one by one and they involve time-varying dynamics. However, it is difficult to develop a model that is suitable for time-varying process, making identification and prediction challenging in real-time [22]. To obtain accurate predictions for time-varying systems in real

Conclusion

This work gives a state-of-the-art review of the recent trend towards achieving an efficient and cost-effective ELM model. Various drawbacks of ELM as a machine learning technique such as difficulty in determination of hidden layer structure, prediction instability and Imbalanced data distributions, the poor capability of sample structure preserving (SSP), and difficulty in accommodating lateral inhibition by direct random feature mapping, were addressed. Other drawbacks addressed include

Acknowledgments

The authors acknowledge the Fundamental Research Grant Scheme (FRGS) from the University of Malaya for funding this work through Project no. “FP046-2017A”.
Peter Adeniyi Alaba does research in Artificial Intelligence (AI), Renewable Energy and Advanced Environmental engineering with a focus on AI in chemical engineering, renewable and sustainable energy, wastewater treatment, solid waste treatment. He is a graduate of Chemical Engineering from the Federal University of Technology, Minna, building up with an M.Eng.Sc. in Catalysis and Reaction Engineering from University of Malaya, Malaysia. He is a Phd candidate of University of Malaya. Peter is

References (170)

- G.-B. Huang

  An insight into extreme learning machines: random neurons, random features and kernels
  Cogn. Comput.
  (2014)
- G. Huang *et al.*

  Trends in extreme learning machines: a review
  Neural Netw.
  (2015)
- M. Hossain *et al.*

  Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems
  J. Clean. Prod.
  (2017)
- S.Y. Wong *et al.*

  A constrained optimization based extreme learning machine for noisy data regression
  Neurocomputing
  (2016)
- G.-B. Huang *et al.*

  Extreme learning machine for regression and multiclass classification
  IEEE Trans. Syst. Man Cybern. Part B (Cybern.)

(2012)
- H. Zhu *et al.*

  Monotonic classification extreme learning machine
  Neurocomputing
  (2017)
- A. Ekbal *et al.*

  Simultaneous feature and parameter selection using multiobjective optimization: application to named entity recognition
  Int. J. Mach. Learn. Cybern.
  (2016)
- S. Yu *et al.*

  Invariant feature extraction for gait recognition using only one uniform model
  Neurocomputing
  (2017)
- M.-B. Li *et al.*

  Fully complex extreme learning machine
  Neurocomputing
  (2005)
- G.-B. Huang *et al.*

  Real-time learning capability of neural networks
  IEEE Trans. Neural Netw.
  (2006)

View more references

Cited by (50)

- Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring
  2022, Mechanical Systems and Signal Processing

  Citation Excerpt :

  Thanks to random hidden nodes used in ELM, only output weights that are needed to be determined can be efficiently and analytically determined by using the Moore-Penrose generalized inverse. Remarkably, it has been theoretically and experimentally proven that ELM [23,24] satisfies the universal approximation capability and can provide good generalization performance. Due to random nodes in the single-hidden layer of ELM, ELM is still uninterpretable.
  Show abstract

- Binary imbalanced data classification based on diversity oversampling by generative models
  2022, Information Sciences

  Citation Excerpt :

  The novelty of the proposed method is a self-adaptive cost-sensitive-based support vector machine used for ensemble classification. Alaba et al. [40] reviewed state-of-the-art cost-sensitive ELM methods and presented recent trends in imbalanced data classification. This section presents the two methods BIDC1 and BIDC2 for binary imbalanced data classification.
  Show abstract

- Machine learning technology in biodiesel research: A review
  2021, Progress in Energy and Combustion Science

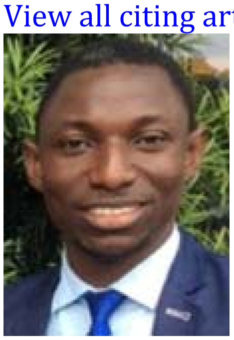
  Show abstract

- A hybrid method based on extreme learning machine and wavelet transform denoising for stock prediction
  2021, Entropy

- Deep learning in electron microscopy
  2021, Machine Learning: Science and Technology

- Thermal decomposition of rice husk: a comprehensive artificial intelligence predictive model
  2020, Journal of Thermal Analysis and Calorimetry

View all citing articles on Scopus



Peter Adeniyi Alaba does research in Artificial Intelligence (AI), Renewable Energy and Advanced Environmental engineering with a focus on AI in chemical engineering, renewable and sustainable energy, wastewater treatment, solid waste treatment. He is a graduate of Chemical Engineering from the Federal University of Technology, Minna, building up with an M.Eng.Sc. in Catalysis and Reaction Engineering from University of Malaya, Malaysia. He is a Phd candidate of University of Malaya. Peter is an author and has worked in collaboration with several scientists from Australia, Malaysia, Mexico, Egypt, Iraq, Pakistan, Canada, United Kingdom and Nigeria on projects such as AI in chemical engineering, Catalyst Development for Efficient Biodiesel Production, and

Carbon Dioxide Reduction, Electrocatalytic reactions, Animal Nutrition, Enzymes, Nanotechnology, Pipeline Corrosion and Wastewater Treatment. He also renders professional services to academic journals such as American Chemical Society (ACS) journals, Chemical Papers, Separation and Purification Technology, Industrial and Engineering Chemistry, Applied Water Science, Journal of Cleaner Production and African Journal of Biotechnology. He is currently a grant peer-review international expert for National Center of Science and Technology Evaluation, JSC, Almaty, Kazakhstan. He is a corporate member of COREN (Council for the Regulation of Engineering in Nigeria).

View full text

Recommended articles

- Probabilistic skyline queries on uncertain time series
  Neurocomputing, Volume 191, 2016, pp. 224-237
  Guoliang He, …, Guofu Zhou
- Probability prediction of short-term user-level load based on random forest and kernel density estimation
  Energy Reports, Volume 8, Supplement 5, 2022, pp. 1130-1138
  Lu Zhang, …, Yongqiang Cheng
- Extreme learning machine based transfer learning algorithms: A survey
  Neurocomputing, Volume 267, 2017, pp. 516-524
  Syed Moshfeq Salaken, …, Saeid Nahavandi

Show 3 more articles