Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 45(1), 2018, Pages 151–155
ISSN: 1223-6934

Improving extreme learning machine performance using ant colony optimization feature selection. Application to automated medical diagnosis

Robert Berglund and Smaranda Belciug

Abstract. Extreme Learning Machine (ELM) is a single-hidden layer feedforward neural network, where the weights between the input and hidden layer are randomly generated and never updated, whereas the hidden-output weights are analytically computed. Theoretical studies have shown that ELM maintains the universal approximation capability. Artificial Intelligence applied in automated medical diagnosis is problematic due to the high risk of overfitting the data, because of the large number of attributes. The goal of this paper is to propose a feature selection (FS) mechanism based on Ant Colony Optimization (ACO), in order to speed up the computational process of the ELM. The proposed model has been tested on three publicly available high-dimensional datasets.

2010 Mathematics Subject Classification. Primary 60J05; Secondary 60J20.
Key words and phrases. neural networks, extreme learning machine, feature selection, ant colony optimization.

1. Introduction

In the last decade, machine learning (ML) and medicine go hand in hand [17]. Neural networks (NNs) are one of the most powerful ML methods. They represent "massively parralled distributed processors" [9] that can be used in modelling non-linear data. NNs mimic the way the human brain works. NNs have been used in breast cancer detection and recurrence [2], [5], [8], [11], cardovascular diseases [13], colon cancer [15], Alzheimer's disease [3], liver fibrosis stadialization [6], [7].

There are two important issues that need to be addressed in order to unlimit the use of ML techniques as classifiers. The first is the "curse of dimensionality" due to the large number of features that describe the data, and the second is the "curse of dataset sparsity" due to the limited number of instances.

ELM has been proposed as a new learning algorithm for NNs, causing an intense dispute in the academic community. Despite this, there have been proposed a numerous of variants of ELM [18], [19], [20]. ELM has been successfully used in medical applications including different types of cancer (breast, colon, leukemia, lung).

Fully connected NN can lead to high complexity or even redundancy. In order to boost ELM computational performance, one can perform feature selection (FS) on the database, thus selecting the most relevant attributes. [12] have used association rules as FS, [14] have used particle swarm optimization as FS, [16] have used hill climbing as FS, [1] have used statistical tools as FS. The current work proposes an

Received March 18, 2018.

ACO algorithm in order to improve the performance and speed of the ELM. The remainder of this paper is organized as it follows. Section 2 presents both the design and implementation of the tandem ACO-ELM. Section 3 presents the benchmarking datasets. Section 4 presents the experimental results and corresponding discussions. The last Section deals with conclusions and future work.

2. Materials and Methods

This paper proposes a FS mechanism that is based on an ACO algorithm. The ACO-ELM is presented in two steps: firstly, the traditional ELM, and, secondly, the attribute filtering module using ACO.

2.1. Traditional ELM algorith. ELM is a single layer neural network, with randomly initialized weights between the input and the hidden layer. The weights between the hidden layer and the output are computed using a Moore-Penrose generalized inverse [10]. We present below the traditional ELM algorithm.

ELM algorithm

(1) Randomly assign the input weights w_i.
(2) Compute the hidden layer output matrix G.
(3) Compute the hidden layer output weight vector $h = G^+ * o$, where G is the hidden output matrix computed at step 2, G^+ is the Moore-Penrose generalized inverse of G, and o is the output vector.

2.2. The ACO FS module. The ACO FS algorithm is a method inspired by nature, simulating the behavior of ants through their cooperation and adaptation to the environment. In real colonies, the ants use an odorous substance, pheromone, in order to indirect communicate with each other. The pheromone is laid by an ant when a source of food is found. The quantity of the laid pheromone is strictly correlated to the distance between the nest and the food source. The other ants move randomly through space, but, when one detects a certain amount of pheromone, it is very likely that it will decide to follow that path. Thus, this ant itself will lay a certain amount of pheromone to that specific path. In this way, the path that has been used by more ants will become more appealing to follow. This means that each time an ant follows that path, it lays pheromone and that increases the probability of that path to be chosen again and again. [4] proposed this method in order to solve combinatorial optimization problems. The first use of ACO was on the solving of the classical traveling salesman problem. We propose the following ACO algorithm in order to perform FS.

ACO algorithm

(1) randomly initialize feature subsets for each of the n ants of the colony, with the constrain that no ant has the same feature set.
(2) initialize the pheromone matrix T with 0.
(3) for each ant, compute the pheromone level. Transform the pheromone values into probabilities for the features that will be chosen by the colony in the next iteration. Each feature has the probability p_i, given by:

$$p_i = \frac{T_i}{\sum_{i=0}^{n} T_i}$$

to be picked by the ant S_j, , taking into consideration the fact that each feature set cannot have duplicates.

(4) the number of selected features is determined by m, where m is decreased by 1 at each step to pinpoint the most significant feature. If $m < 2$, we generate m according to a random selection procedure similar to the classical *roulette* method.

(5) Solution evaluation: evaluate each ants feature subset using the ELM algorithm.

(6) Pheromone update: we iterate over the k best ants, and update the pheromone map using the MSE (mean squared error) computed in the evaluation step of the solution, given by:

$$\Delta T_i = \begin{cases} \frac{(max(MSE)-MSE_j)}{max_{g=1:k}(max(MSE)-MSE_g)}, & \text{if } f_i \in S_j \\ 0, & \text{otherwise} \end{cases}$$

$$T_i = (1-p) * T_i + \Delta T_i$$

where T_i is the pheromone associated to the feature f_i, and MSE_j is the estimated MSE of the classification result obtained from the ELM for the features of S_j.

3. Benchmarking datasets

Two of the three medical datasets used in the benchmarking process refer to gene expression used in cancer detection from the publically available Machine Learning Data Set Repository (http://mldata.org/). They refer to the DNA micro-array technology in breast cancer and lung cancer. The last one refers to thyroid disease and contains bio-chemical analysis. It is noteworthy that all datasets are wide datasets, ranging from 21 to 24881 attributes, and from 62 to 7200 instances. They are chosen intentionally unbalanced in order to see the natural performance of the ACO-ELM algorithm in real-world conditions.

(1) Breast cancer Kent Ridge (BCKR). The datasets contains 97 instances with 24481 attributes and two-class decision (46 relapse vs. 51-non-relapse). Available at: http://mldata.org/repository/data/viewslug/breast-cancer-kent-ridge-2/

(2) Lung cancer Michingan (LCM). The dataset contains 96 instances with 7129 attributes and two-class decision (86-tumor vs. 10 -normal). Available at: http://mldata.org/repository/data/viewslug/lung-cancer-michigan/

(3) Thyroid disease (TD). The dataset contains 7200 instances with 21 attributes and three decision classes, that is the stages of the disease (1, 2 or 3).

To assess the performance of this ACO-ELM model we have used the 10-fold cross-validation as testing method, since in this case, two out of three datasets have not a large enough size. Hence, the classification accuracy for the training/testing phases has been computed 10 times, each time leaving out one of the sub-samples and using that sub-sample as a test sample for cross-validation. Therefore, each sub-sample is used 9 times as training sample and just once as testing sample. If the model performs as well in the test sample as in the training sample, we can say that it cross-validates well.

4. Results

Since one of most important goals of this paper is to compare the classification performances of the ACO-ELM model with the traditional ELM, an *a priori* statistical power analysis (two-tailed type of null hypothesis) has been performed to determine the appropriate sample size in order to achieve adequate statistical power. Accordingly, we have considered a sample of 100 different computer runs (of a complete cross-validation cycle) for each model, providing a statistical power equaling 95% with type I error $\alpha = 0.05$ for the comparison test subsequently used.
The results of the classification performances of the ACO-ELM, in terms of average, standard deviation (SD) and 95% confidence intervals (CI), averaged over 100 computer runs of a complete cross-validation cycle are displayed in Table 1.

Datasets	BCWD	LC	THY
Performance measures (%) ACO-ELM	52.00	93.70	92.59
Performance measures (%) 3-MLP	47.50	69.17	91.00

We can see from above the table that the new model outperforms the classic MLP on these specific databases.

5. Conclusion

This paper presents a FS model using ACO along with on an ELM algorithm, and its performance to a standard ELM. The model was tested on five large publically available medical datasets in order to validate our conclusions. The computational results suggest the suitability of this new model for the classification of different diseases.

References

[1] S. Belciug, M. Lupsor, R. Badea, Feature selection approach for non-invasive evaluation of liver fibrosis, *Annals of the University of Craiova, Mathematics and Computer Science Series* **35** (2008) 15–20.
[2] S. Belciug, F. Gorunescu, M. Gorunescu, A. Salem, Assessing performances of unsupervised and supervised neural networks in breast cancer detection, *The 7th International Conference on Informatics and Systems (INFOS)*, (2010), 1–8.
[3] F. Copped, E. Grossi, M. Buscema, L. Migliore, Application of artificial neural networks to investigate one-carbon metabolism in Alzheimer's disease and healthy matched individuals, *PLoS One* **8** (2013), (8): e74012.
[4] M. Dorigo, V. Maniezzo, A. Colorni, Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics,* Part B **26** (1996), 29–41.
[5] F. Gorunescu, S. Belciug, Boosting backpropagation algorithm by stimulus-sampling: Application in computer-aided medical diagnosis, *Journal of Biomedical Informatics* **63** (2016), 74–81.
[6] F. Gorunescu, S. Belciug, M. Gorunescu, R. Badea, Intelligent decision-making for liver fibrosis stadialization based on tandem feature selection and evolutionary-driven neural network, *Expert Systems with Applications* **39** (2012), 12824–12833.
[7] F. Gorunescu, S. Belciug, Evolutionary strategy to develop learning-based decision systems. Application to breast cancer and liver fibrosis stadialization, *J. Biomed. Inform.* **49** (2014), 112–118.

[8] F. Gorunescu, M. Gorunescu, E. El-Darzi, S. Gorunescu, A statistical framework for evaluating neural networks to predict recurrent events in breast cancer, *International Journal of General Systems* **39** (2010), no. 5, 471–488.
[9] S. Haykin, *Neural Networks. A comprehensive foundation.* (2nd Ed.), Prentice Hall International, 1999.
[10] G.B. Huang, Q.C. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in *Proc. Intl. Joint Conf. Neur. Net* (2006), 985–990.
[11] A.A. Kalteh, P. Zarbakhsh, M. Jirabadi, J. Addeh, A research about breast cancer detection using different neural networks and K-MICA algorithm, *J. Cancer Res. Ther.* **9** (2013), no. 3, 456–466.
[12] M. Karabatak, M.C. Ince, An expert system for detection of breast cancer based on association rules and neural network, *Expert Syst. Appl.* **36** (2009), 3465–3469.
[13] K. Nakajima, S. Matsuo, H. Wakabayashi et al., Diagnostic Performance of Artificial Neural Network for Detecting Ischemia in Myocardial Perfusion Imaging, *Circ J.* **79** (2015), no. 7, 1549–56.
[14] Q. Shen, Z. Meia, B-X. Ye, Simultaneous genes and training samples selection by modified particle swarm optimization for gene expression data classification, *Comput. Biol. Med.* **39** (2009)(7), 646-649.
[15] L. Spelt, J. Nilsson, R. Andersson, B. Andersson, Artificial neural networks - a method for prediction of survival following liver resection for colorectal cancer metastases, *Eur. J. Surg. Oncol.* **39** (2013), no. 6, 648–654.
[16] C. Stoean, R. Stoean, M. Lupsor, H. Stefanescu, R. Badea, Feature selection for a cooperative coevolutionary classifier in liver fibrosis diagnosis, *Comput. Biol Med.* **41** (2011), 238–246.
[17] Z.R. Yang, *Machine learning approaches to Bioinformatics*, World Scientific, Science, Engineering, and Biology Informatics, 2010.
[18] L. Zhang, D. Zhang, Evolutionary Cost-Sensitive Extreme Learning Machine, *IEEE Transactions on Neural Networks and Learning Systems* **(99) (2016)**, 1–16.
[19] Y-P. Zhao, Parsimonious kernel extreme learning machine in primal via Cholesky factorization, *Neural Networks* **80** (2016), (C):95–109.
[20] G.-G. Wang, M. Lu, Y.-Q. Dong, et al., Self-adaptive extreme learning machine, *Neural Computing and Applications* **27** (2016), no. 2, 291–303.

(Robert Berglund, Smaranda Belciug) Department of Computer Science, University of Craiova, 13 A.I. Cuza Street, Craiova, 200585, Romania
E-mail address: rbrtberglund@gmail.com, smaranda.belciug@inf.ucv.ro