Advanced Extreme Learning Machine Modeling using Radial Basis Function Network and Context Clustering

Junbeom Kim*, Wonjo Lee, KyoJoong Oh, Sung-Suk Kim and Ho-Jin Choi
Department of Computer Science, Korea Advanced Institute of Science and Technology
{silenteye, mochagold, aomaru, powerkimss, hojinc} @kaist.ac.kr

Abstract

In this paper, we propose a new hybrid intelligent modeling using context clustering and Extreme Learning Machine (ELM) mechanism. It has been a sensitive issue that the ELM mechanism assigns initial parameters randomly, despite of its superior performance. The proposed approach focuses on initial parameters determination of the modeling to improve the accuracy of the ELM mechanism, through removing randomness of assignment. To accomplish it, a context clustering based on Gaussian Mixture Model (GMM), considering a relationship between input-output spaces will be adopted to a Radial Basis Function Network (RBFN) of the ELM. In addition, the proposed approach will reduce the randomness of results from the original ELM. Simulations and the results show usefulness of the proposed approach with improved performance accuracy.

Keywords: *ELM, context clustering, Gaussian mixture model, initial parameter determination, RBFN*

1. Introduction

In mathematical modeling, an optimization of a non-linear system is difficult to be approached with the linear mathematical basis. Intelligent systems, [1, 2] including neural networks and fuzzy systems, are one of the methods to model the complex environment replacing the conventional linear systems. A dominant characteristic of the intelligent systems are optimizing a given system based on iterative learning strategies. Moreover, the intelligent systems use learning methods avoiding the mathematical dependency including linear system constraints.

General neural networks [1, 2] have iterative learning procedure that cause heavy computational loads and time consumptions in spite of obtaining the good performance. An Extreme Learning Machine (ELM) [3-5] is a simple and powerful learning mechanism avoiding the conventional neural network learning problems.

A clustering method [1, 2] is successfully applied for improving the performance of the neural network by determining initial parameters. Additionally, a context clustering method can induce good results, which an algorithm adopts input-output relationships in the learning process. The context clustering method [6-9] extends clustering performance from an unsupervised learning to a supervised learning.

The ELM mechanism randomly determines initial parameters of the system. The randomness is an advantage of the model. It causes, however, different output results on the same condition. This characteristic needs to be reinforced to improve the performance. In this paper, we propose an advanced ELM using the context clustering to reduce the randomness and assign the initial parameters of the ELM mechanism. Through proposed approach, overall improvements of both the performance and the randomness are intended.

In section 2, related works, which are GMM, context clustering, RBFN, and ELM, are described. In section 3, we denote the proposed approach to improve the overall system. Section 4 displays the simulations and discuss about the results. Finally, we conclude in section 5.

2. Related Works

In section 2, we briefly describe background knowledge of methods used in this paper.

2.1. Context Clustering using Gaussian Mixture Model

A Gaussian Mixture Model (GMM) [10][11] is a parametric probability model represented as a weighted sum of Gaussian component densities. The probability function p of d-dimensional Gaussian component densities is following.

$$p_l(x|\mu_l, \Sigma_l) = \frac{1}{(2\pi)^{d/2}|\Sigma_l|^{1/2}} e^{-\frac{1}{2}(x-\mu_l)^T \Sigma_l^{-1}(x-\mu_l)} \tag{1}$$

where $\mathbf{x}$ is input data vector, μ is mean of Gaussian model, Σ is relevant variance. Each parameter can be estimated using an Expectation Maximization (EM) algorithm iteratively.

$$\mu_l = \frac{\sum_{i=1}^{N} x_i p(l|x_i, \Theta^g)}{\sum_{i=1}^{N} p(l|x_i, \Theta^g)} \tag{2}$$

$$\Sigma_l = \frac{\sum_{i=1}^{N} p(l|x_i, \Theta^g)(x_i-\mu_l)(x_i-\mu_l)^T}{\sum_{i=1}^{N} p(l|x_i, \Theta^g)} \tag{3}$$

$$\alpha_l = \frac{1}{N}\sum_{i=1}^{N} p(l|x_i, \Theta^g) \tag{4}$$

where Θ^g is parameter that maximize the likelihood which consists of the mean and the variance.

A context clustering [6-9] is a clustering method considering a relationship between input-output. The output clusters consider a context that effect on connected input clusters. Compare to the general clustering approach using only given input data, the context clustering can achieve a local subspace detailed clustering.

The most kernel part of the context clustering can be explained by $\mathrm{f_k}$ of the equation (5) compared with the equation (1).

$$p_l(x|\mu_l, \Sigma_l) = \frac{f_k}{(2\pi)^{d/2}|\Sigma_l|^{1/2}} e^{-\frac{1}{2}(x-\mu_l)^T \Sigma_l^{-1}(x-\mu_l)} \tag{5}$$

As shown in the Figure 1, the context clustering has a context space separated by context clusters. The conventional clustering has same value of the f_k as 1 in the view of the context clustering concept. However, the f_k of the context clustering changes according to the each cluster membership. The membership grades of each context cluster have different probabilities with the f_k. This information influences clustering of the input spaces.

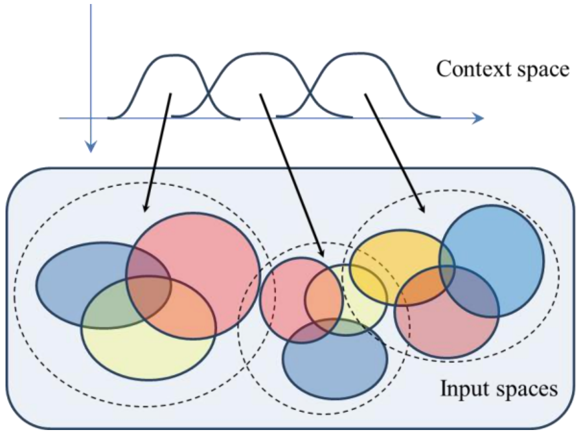


Figure 1. Concept of the Context Clustering

2.2. Extreme Learning Machine with Radial Basis Function Network

A Radial Basis Function Network (RBFN) [5, 7] is a neural network consisted with linear combination of radial basis functions. The output of the RBFN with N kernels hidden layer for an input vector $x \in R^n$ is following.

$$f_N(x) = \sum_{i=1}^{N} \beta_i \phi_i(x) \tag{6}$$

where β_i is the weight vector and $\phi_i(x)$ is the output of ith kernel.

$$\phi_i(x) = \exp\left(\frac{\|x-\mu_i\|^2}{\sigma_i}\right) \tag{7}$$

where μ_i is center and σ_i is impact width of ith kernel.

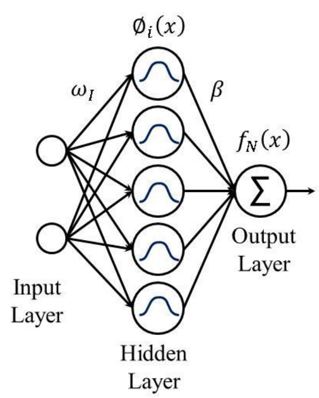


Figure 2. Concept of the RBFN

As shown in the Figure 2, the RBFN has an only single hidden layer. The RBFN has activation functions as radial basis function and the output classes of the RBFN are separated by hyper-spheres. These are two different characteristics compared to a Multi-Layer Perceptron (MLP) which has different activation functions and separate the output classes by hyper-planes.

An Extreme Learning Machine (ELM) [3-5] is iterative tuning-free learning algorithm that having extremely fast learning speed and the good generalization performance with random initial parameters assignment. The ELM with the RBFN algorithm consists of three steps which are following.

1) Assign randomly hidden node parameters, the kernel centers μ_i and the impact width σ_i,

2) Calculate the hidden layer output matrix H, and

3) Calculate the output weights β of RBFNs.

$$\beta = \mathrm{H}^{\dagger}\mathrm{T} \tag{8}$$

where $\mathrm{H}^{\dagger}$ is pseudo-inverse of the hidden layer output matrix H and T is the output vector of the system.

3. Proposed Approach

Although the ELM has characteristics which are fast learning and good performance, random initial parameter determination cause different output results from learning with same condition. In addition, the performance of the ELM is not the best result since it doesn't consider about the relationship between input-output data while assigning the initial parameters randomly.

In proposed approach, the context clustering is used to resolve these problems. At first, a clustering method is used to determine initial parameters rather than random determination, to reduce the randomness. For expecting better performance, the context clustering is used to assign initial parameters by perform a clustering with the relationship between the input-output spaces.

The context clustering considers the relationship between input-output data which means that it is a supervised learning upgraded from the typical cluster learning method which is an unsupervised learning. Thus, better clustering results are predictable and this is the reason why the context clustering is used for assigning initial parameters of the ELM.

The proposed algorithm is following.

1) Determine initial parameters of the context clustering using the Gaussian Mixture Model (CGMM)

- Initialize the CGMM parameters before clustering using a Fuzzy C-Means (FCM)
- Learning optimal parameters of the CGMM

2) Generate the hidden layer output matrix H of the ELM using the clustering results of the CGMM

- The center of the RBFN obtained from the inferred center of the CGMM
- The variance of the RBFN obtained from the related variance of the CGMM

3) Calculate output weights β

4) Infer model output

Through proposed algorithm, the Advanced ELM method reduces the randomness of the ELM, since the random selection of the initial parameters step is replaced by the CGMM.

4. Simulation

To show effectiveness of the proposed approach we will simulate two examples which are representative benchmark datasets.

4.1. Simulation Environments

We use datasets which are IRIS classification and Auto MPG prediction problems from the UCI machine learning repository site [12]. Simulation environments are following. CPU specification is Intel Core i7 3.4GHz, RAM is 12GB, and simulation language is MATLAB version R2011b.

4.2. Simulation and Results

Simulation is performed a hundred times repeat in same parameter setting and the best and average results are displayed. To avoid overfitting problem, abnormal testing results, which of the accuracy is more than two times of the training accuracy, are removed

Three simulation results, from the ELM with the RBFN (RBFN-ELM), the ELM initialized using the FCM (FCM-ELM), and the proposed method (CGMM-ELM), are compared.

First simulation is the IRIS classification problem. This data consists of 4 features which are length and width of sepal and petal respectively. The outputs are classified into 3 classes which are Setosa, Versicolour, and Virginica.

Numerical comparisons of the best and average results are denoted in Table 1, 2, and 3. The attribute ‘Remark’ means a number of the context and a number of the cluster. The error of the proposed approach, the CGMM-ELM, is the smallest among three models, as shown in the Table 1. On the other hand, in case of the average, the FCM-ELM obtained the best result, as shown in the Table 2. The Table 3 denotes the variance of the variances from a hundred repeated simulation results, for estimating the changes of variances. The FCM-ELM obtained the smallest variance and the CGMM-ELM obtained higher variance because of the random determination of the initial center of the CGMM. However, it is smaller than the ELM-RBFN.

Table 1. The Best Performances of the IRIS Classification

	trn_err	chk_err	Remark
ELM-RBFN	0.1754	0.2130	3, 3
FCM-ELM	0.1882	0.2119	3, 3
CGMM-ELM	0.1655	0.1906	3, 3

Table 2. The Average Performances of the IRIS Classification

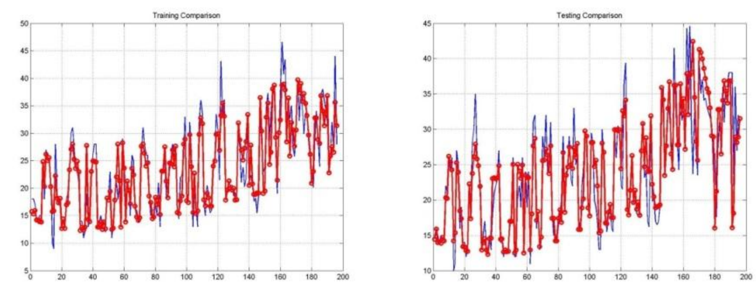
	trn_err	chk_err	Remark
ELM-RBFN	0.3669	0.4361	3, 3
FCM-ELM	0.1890	0.2286	3, 3
CGMM-ELM	0.2662	0.2651	2, 3

Table 3. The Randomness of the IRIS Classification

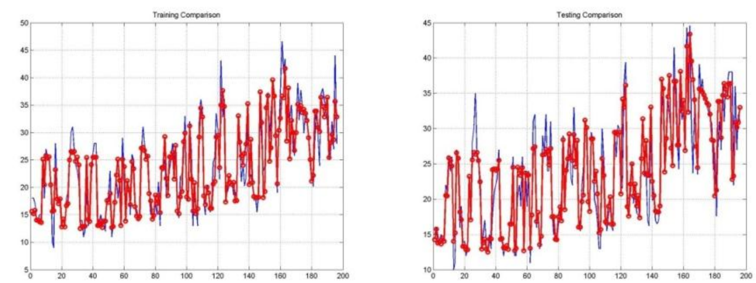
	trn_err	chk_err	Remark
ELM-RBFN	0.0124	0.0201	3, 3
FCM-ELM	1.2428e-05	5.2630e-05	3, 3
CGMM-ELM	0.0049	0.0064	2, 3

Second simulation is the Auto MPG prediction problem. This data consists of 8 attributes and simulation uses only two attributes. The outputs are fuel consumption prediction.

Figure 3 display the training and testing error of the RBFN-ELM and the CGMM-ELM.



(a) Training error of RBFN-ELM (b)Testing error of RBFN-ELM



(c) Training error of CGMM-ELM (d) Testing error of CGMM-ELM

Figure 3. Comparison of RBFN-ELM and CGMM-ELM

Numerical comparisons of the best and average results are denoted in Table 4, 5, and 6. The attribute 'Remark' means the number of the context and the number of the cluster. The proposed approach, the CGMM-ELM, obtained the smallest error in both of the best and average case. Descriptions of the Table 6 are identical to that of the Table 3.

Table 4. The Best Performances of the Auto MPG Prediction

	trn_err	chk_err	Remark
ELM-RBFN	2.8918	3.5791	4 4
FCM-ELM	2.9165	3.3938	4 3
CGMM-ELM	2.8472	3.0812	4 5

Table 5. The Average Performances of the Auto MPG Prediction

	trn_err	chk_err	Remark
ELM-RBFN	2.8945	3.5236	4 4
FCM-ELM	2.9331	3.5504	4 3
CGMM-ELM	2.8419	3.6035	4 5

Table 6. The Randomness of the Auto MPG Prediction

	trn_err	chk_err	Remark
ELM-RBFN	0.0040	0.0276	3, 4
FCM-ELM	3.6924e-04	0.0033	4, 3
CGMM-ELM	0.0011	0.0311	4, 5

4.3. Discussion

As shown in figures and tables, the proposed approach has advanced performance including improved accuracy and reducing the randomness. The original ELM depends on random selection of the initial parameters for constructing the model. The ELM mechanism obtains good performance, but not best results. We tried to resolve the randomness problem and improve the performance using the context clustering method. In the proposed method, the initial parameters of the system include model structure and related information, through clustering with the input-output relationship.

5. Conclusion

In this paper, we showed advantages of the advanced ELM learning mechanism using the context clustering concept with GMM. The pre-determining initial parameters of the model achieve improving of the performance and reducing the randomness. To generate more advanced initial parameters, proposed approach used input-output relationship of the given system, using the context clustering method. The improvement of the performance verified usefulness and effectiveness of the proposed approach.

In spite of improved features, there are still a few considerations. An optimal structure determination step, before parameter optimization in the main progress, is needed to construct model frame. Good results can be induced by optimal structure and parameter learning. In addition, an initialization of the GMM is needed to be improved. We simply used FCM algorithm to obtain the context clusters and randomly initialized the center of the GMM, before learning main clustering. Improved results can be obtained, if it is possible to take better initial parameters.

Related future researches will include improving context clustering performance itself and obtaining better initial parameters of the clustering. Various researches, to improve the performance of the ELM while keeping dominant concepts, should be progressed at the same time.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant (No. 2012-0001001) of Ministry of Education, Science and Technology (MEST) of Korea.

References

[1] J. S. R. Jang, C. T. Sun and E. Mizutani, "Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence", Prentice Hall, **(2007)**.
[2] S. Haykin, "Neural Networks: A Comprehensive Foundation Second Edition", Prentice Hall, **(1999)**.
[3] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme learning maching: Theory and applications", Neurocomputing, vol. 70, Issue. 1-3, **(2006)**, pp. 489-501.
[4] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks", Proceeding of IEEE International Joint Conference on Neural Networks, vol. 2, **(2004)** July 25-29, pp. 985-990, Budapest, Hungary.
[5] G. B. Huang and C. K. Siew, "Extreme learning machine: RBFnetwork case", The 8th International Conference on Control, Automation, Robotics and Vision, vol. 2, **(2004)** December 6-9, pp. 1029-1036, Kuming, China.
[6] W. Pedrycz, "Conditional Clustering fuzzy C-Means", Pattern Recognition, vol. 17, Issue 6, **(1996)**, pp. 625-631.
[7] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks", IEEE Trans on Neural Networks, vol. 9, no. 4, **(1998)**, pp. 601-612.
[8] W. Pedrycz and K. C. Kwak, "Linguistic Models as a Framework for user-centric system modeling", IEEE Trans on Systems, Man, and Cybernetics-Part A, vol. 36, Issue 4, **(2006)**, pp. 727-745.
[9] W. Pedrycz, "Knowledge-Based Clustering", John Wiley & Sons, **(2005)**.
[10] J. A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models", [Internet], International Computer Science Institute; **(1997)**, Avilable from: http://crow.ee.washington.edu/people/bulyko/papers/em.pdf.
[11] D. Reynold, Gaussian Mixture Models [Internet], Encyclopedia of Biiometric Rcognition; **(2008)**, Available from: http://www.ll.mit.edu/mission/communications/ist/publications/0802_Reynolds_Biometrics-GMM.pdf.
[12] UCI Machine Learning Repository, Available from: http://archive.ics.uci.edu/ml.