Extreme Learning Machines for Wireless Channel Prediction in Microcell and Picocell Environments

Nikola Sekulović[1], Miloš Stojanović[2], Aleksandra Panajotović[3] and Miloš Banđur[4]

Abstract – **In this paper, examination of possibility and effectiveness of extreme learning machines (ELM) application to predict wireless channel conditions for single-input single-output (SISO) systems in microcellular and picocellular environments is carried out. Normalized mean squared error (NMSE) and time consumption are used as performance indicators. The experimental results on measured values for signal-to-noise ratio (SNR) show high accuracy of the ELM prediction model and short execution time.**

Keywords – **Channel prediction, Extreme learning machines, Microcellular environment, Picocellular environment.**

I. Introduction

Knowledge of information about state of wireless channel is increasingly important. The reason of that trend lies in demands for high-data services and limited wireless spectrum. Unfortunately, a state of wireless channel changes very quickly, so channel state obtained by channel estimation can become outdated due to delay caused by processing and feedback phases. The system performance enhancement can be achieved using channel prediction based on channel states in previous moments rather than using channel estimation [1].

In the open technical literature, there are several papers dealing with channel states prediction. Autoregressive (AR) model, support vector machine (SVM), discrete wavelet transform (DWT) method in combination with AR and linear regression (LR) algorithm (DWT-AR-LR) and echo state network (ESN) are widely explored in [2]-[6].

Extreme learning machine (ELM) is a learning algorithm for feedforward artificial neural networks with one hidden layer. Compared with traditional artificial neural networks, ELM may achieve better generalization performance for regression and classification cases. ELM tends to minimize training error with the smallest norm of weights. In addition, ELM has faster learning speed, i.e. significantly low computational time required for training (up to thousands of times) [7]-[9]. It increases training speed by randomly assigning weights and biases in the hidden layer, instead of iteratively adjusting its parameters by gradient based methods.

[1]Nikola Sekulović is with the College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: nikola.sekulovic@vtsnis.edu.rs.

[2]Miloš Stojanović is with the College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia.

[3]Aleksandra Panajotović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[4]Miloš Banđur is with the Faculty of Technical Sciences, University of Priština, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia.

In this paper, the effectiveness of prediction scheme based on ELM is explored for microcellular and picocellular environments. Data sets used for training and testing contain measured signal-to-noise ratio (SNR) samples for scenarios described in details in [10]. Performance metrics used for analysis of the approach proposed is normalized mean squared error (NMSE) and time consumption.

The rest of the paper is organized as follows. Section II provides brief description of ELM. Section III describes communication scenario, data sets and ELM-based prediction algorithm. Experimental evaluation is presented in Section IV, while Section V concludes the paper.

II. Extreme Learning Machines

Let's denote N training samples as $(\mathbf{x}_j, \mathbf{y}_j)$, $j=1,\dots,N$, where $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in \mathbf{R}^n$ represents the j-th n-dimensional training instance and $\mathbf{y}_j = [y_{j1}, y_{j2}, \dots, y_{jm}]^T \in \mathbf{R}^m$ represents the j-th target value of the dimension m. ELM has the unified solutions for regression, binary and multiclass classification. In the case of regression, which is of interest for problem considered in this paper, it holds that m=1 [7]. Generally, the output of a standard single hidden layer feedforward network (SLFN) with L hidden neurons and activation function $h(x)$ is defined as

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i h(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{f}_j, \quad j = 1,\dots,N, \tag{1}$$

where $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$, $i=1,\dots,L$, is the weight vector connecting the i-th hidden neuron and all input neurons, $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ represents the weight vector connecting the i–th hidden neuron and all the output neurons, and b_i is the threshold of the i–th hidden neuron. According to ELM theory, $\mathbf{w}_i$ and b_i can be randomly and independently assigned a priori, i.e. without considering the input data [8].

The SLFN defined with (1) has approximation capabilities with zero error means $\sum_{i=1}^{L} \|\mathbf{f}_i - \mathbf{y}_i\| = 0$, i.e., there exist $\boldsymbol{\beta}_i$, $\mathbf{w}_i$ and b_i such that

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i h(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{y}_j, \quad j = 1,\dots,N. \tag{2}$$

The previous equation can be expressed in matrix form resulting in

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}, \tag{3}$$

with

$$\mathbf{H} = \begin{bmatrix} h(\mathbf{w}_1\mathbf{x}_1 + b_1) & \cdots & h(\mathbf{w}_L\mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ h(\mathbf{w}_1\mathbf{x}_N + b_1) & \cdots & h(\mathbf{w}_L\mathbf{x}_N + b_L) \end{bmatrix}_{N\times L}, \quad (4)$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L\times m}, \quad (5)$$

and

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix}_{N\times m}. \quad (6)$$

The matrix **H** represents the hidden layer output matrix of the neural network where the *i*-th column of **H** represents the *i*–th hidden neuron's output vector in regard to inputs $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$. The output weights can be analytically determined by finding the unique smallest norm least-squares solution of the linear system described by (3). In order to improve the performance, the constrained optimization problem can be formed for ELM, as shown in [7]:

$$Minimize:\ L_P = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\frac{1}{2}\sum_{j=1}^{N}\|\boldsymbol{\xi}_j\|^2$$

$$Subject\ to:\ h(\mathbf{x}_j)\boldsymbol{\beta} = \mathbf{y}_j^T - \boldsymbol{\xi}_j^T,\ j = 1,...,N \quad (7)$$

where $\xi_j = \left[\xi_{j1},...,\xi_{jm}\right]^T$ is the training error vector of the m output nodes with respect to the training sample x_i, while C represents tradeoff parameter between model complexity and allowed errors ξ_j during training. Based on *Karush-Kuhn-Tucker* (KKT) theorem, the optimization problem previously defined is equivalent of solving the dual optimization problem

$$L_D = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\frac{1}{2}\sum_{j=1}^{N}\|\boldsymbol{\xi}_i\|^2 - \sum_{j=1}^{N}\sum_{i=1}^{m}\alpha_{ji}(h(\mathbf{x}_j)\boldsymbol{\beta}_i - \mathbf{y}_{ji} + \xi_{ji}), \quad (8)$$

where $\alpha_j = \left[\alpha_{j1},...,\alpha_{jm}\right]^T$ are Lagrange multipliers.

After solving (8) based on *KKT* conditions, which can be found in detail in [7], the following solution is obtained:

$$\boldsymbol{\beta} = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{Y} \quad (9)$$

and the function of ELM is:

$$f(\mathbf{x}) = h(\mathbf{x})\boldsymbol{\beta} = h(\mathbf{x})\left(\frac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{Y}. \quad (10)$$

III. Application of ELM for Wireless Channel Prediction

A. Channel Description

The efficiency of the proposed machine learning technique for SNR prediction in wireless communication system employing a single transmit antenna and a single receive antenna is investigated. Namely, single-input single-output (SISO) channel in two different environments is considered:

1) B channel model represents a microcell environment where distance between mobile station (MS) and base station (BS) is in the order of 30 m. It assumes indoor-to-outdoor propagation with BS located outside and indoor environment usually consisted of several small offices.

2) E channel model refers to indoor-to-indoor scenario. It represents a picocell environment in modern open office with windows metallically shielded.

B. Data Sets

Data sets used for analysis in this work contain SNR channel values obtained based on measurement campaigns described in details in [10]. A series of SNR samples $x(k) = x(kT)$, $k = \overline{1,N}$, from [10], are used for network training and testing. Parameter T denotes sampling interval and parameter N is the total number of samples.

C. Prediction Algorithm

In general, for a given training set with N instances of n features, the sigmoid activation function $g(x)$ and L hidden neurons number, the ELM algorithm for regression can be summarized as follows:

Training procedure

(a) Assign random input weights $\mathbf{w}_i$ and biases b_i, $i = 1,..., L$;
(b) Compute the hidden layer output matrix **H** using (4);
(c) Compute the output weights $\boldsymbol{\beta}$ using (9);

Testing procedure

(a) Compute the hidden layer output vector $h(\mathbf{x})$ for current instance from the test set using (4);
(b) Compute the output $f(\mathbf{x})$ according to (10) using the $\boldsymbol{\beta}$ obtained in step (c) of the training.

IV. Experimental Evaluation

In this section of the work, the accuracy of the ELM network for the time series prediction of SNR in the SISO system is evaluated using NMSE as a prediction error metrics. NMSE is defined as

$$NMSE = \frac{\sum_k \left(x(k) - f\left(x(k)\right)\right)^2}{\sum_k \left(x(k)\right)^2}. \quad (11)$$

Data sets containing the measured instantaneous SNR values at the receiver side for the case when SNR at the transmitter side is 20 dB for both B and E channel model are used to test the proposed method. Analysis is carried out using N=4000 samples. The data sets are divided into two equal sets for training and testing (N_{tr}=N_{te}=2000). It is determined by simulation that there is no need to use more than 3 neurons in the input layer. Sigmoid function is used as activation. For the tests, the ELM is implemented in MATLAB. As an illustrative example, Fig. 1 shows target signal and prediction curve for the case of E channel.

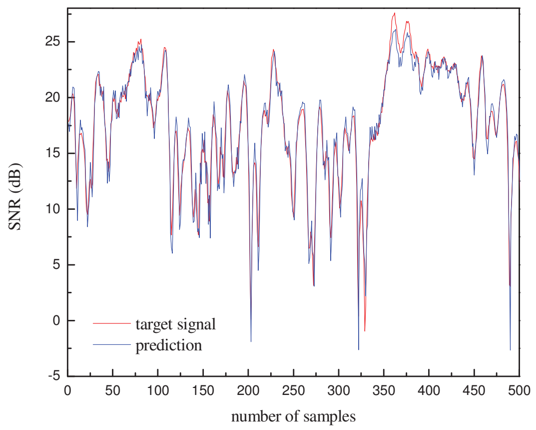

Fig. 1. Target signal and prediction curve for E channel

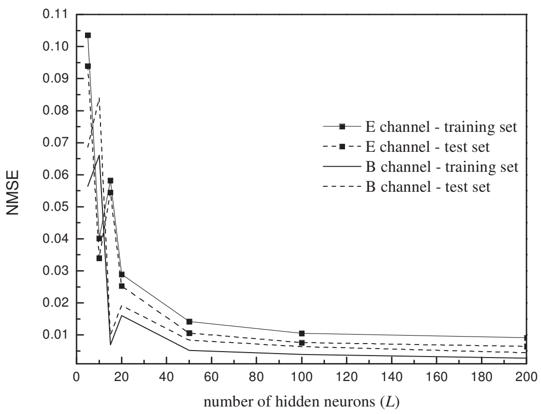

Fig. 2. NMSE versus number of hidden neurons

The NMSE is evaluated for both training and test set, with training and test times, measured in seconds on an *Intel Core i5* computer. Results for the NMSE as a function of the number of neurons in hidden layer are presented graphically in Fig. 2. It is notable that in the case of E channel, the NMSE on the test set remains in range from 0.1 to 0.006 for the number of neurons in hidden layer in range from 5 to 200. In the case of B channel, for the same prediction network, the NMSE values are slightly lower. They are in the range from 0.08 to 0.004. Furthermore, the analysis shows that increasing number of neurons above a certain value does not improve prediction results significantly. We can also observe that the NMSE on the training set follows trend of the NMSE on training set regardless of number of hidden neurons, which implies that models are not overfitted. The obtained values of the NMSE for all ELMs, regardless the number of neurons in the hidden layer are within the expected range of precision and comparable to the results obtained in other studies [5, 6].

Table I contains training and test time in seconds for different number of neurons in hidden layer. We can see that training time for all 2000 instances in training set is only 0.001 seconds with 100 neurons in hidden layer, while prediction for all 2000 instances is done in 0.06 second. With increase of number of neurons up to 1000, these times slightly increase. For more than 2000 neurons in hidden layer training time increases significantly (greater than 1s), while test time increases slightly (but it is still less than 1s). These results demonstrate high performances in terms of training and test speed on this data set.

TABLE I
TIME CONSUMPTION OF THE ELM MODEL

Number of neurons	Training time (s)	Test time (s)
5	0.001	0.0312
10	0.001	0.0312
15	0.001	0.0468
20	0.001	0.0468
50	0.001	0.0468
100	0.001	0.0625
200	0.0468	0.0625
300	0.0781	0.0625
400	0.2031	0.0625
500	0.2187	0.0756
1000	0.8437	0.1406
2000	3.6875	0.3281
3000	8.9218	0.3593
4000	16.9063	0.4843
5000	29.1719	0. 5781

In order to compare the results of the ELM with other common classification techniques, we have measured accuracy of the Linear SVM and RBF SVM, on the same dataset. NMSE for Linear SMV was 0.0118, while NMSE for RBF SVM reached 0.0083. It can be noted that ELM outperforms Linear SVM in terms of NMSE, having the similar algorithm complexity. On the other side, ELM reaches results comparable to the RBF SVM, while operating significantly faster during the training and testing.

V. CONCLUSION

This paper has investigated the ELM-based prediction scheme for SISO systems in microcellular and picocellular environments. The effectiveness of the framework has been confirmed using NMSE as a performance measure along with training and test time. Simulation results have shown that no more than several hundred neurons in hidden layer should be used. Further increasing the number of neurons will not result in significant prediction accuracy gain. The NMSE of the order of 10^{-3} and training and test time less than 0.22 and 0.08 seconds, respectively, can be expected.

ACKNOWLEDGEMENT

This work has been funded by Serbian Ministry for Education and Science under the projects TR-32052 and III-44006.

REFERENCES

[1] A. Panajotović, F. Riera-Palou and G. Femenias, “Limited feedback MU-MIMO-OFDM system”, International ITG Workshop on Smart Antennas WSA 2016, pp. 348-355, Munich, Germany, 2016.

[2] P. Sharma and K. Chandra, “Prediction of state transitions in Rayleigh fading channels”, IEEE Trans. Veh. Technol., vol. 56, pp. 416-425, 2007.

[3] X. Zhao, C. Hou and Q. Wang, “A new SVM-based modeling method of cabin path loss prediction”, Int. J. Antennas and propagation, vol. 2013, pp. 1-7, April 2013.

[4] X. Long and B. Sikdar, “A wavelet based long range signal strength prediction in wireless networks”, in Proc. IEEE Int. Conf. Commun., pp. 2043-2047, 2008.

[5] Y. Zhao, H. Gao, N. C. Beaulieu and Z. Chen, “Echo State Network for Fast Channel Prediction in Rician Fading Scenarios”, IEEE Comm. Lett., vol. 21, pp. 672-675, 2017.

[6] N. Sekulović, A. Panajotović, M. Banđur, Đ. Banđur and D. Milović, “Channel Prediction in Wireless Microcell and Picocell Systems Using Echo State Network”, AJER, vol. 7, no. 12, pp. 134-138, 2018.

[7] G.-B. Huang, H. Zhou, X. Ding and R. Zhang, “Extreme Learning Machine for Regression and Multiclass Classification”, IEEE T. Syst. Man CY-S., vol. 42, no. 2, pp. 513-529, 2012.

[8] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, “Extreme Learning Machine: Theory and Applications”, Neurocomputing, vol. 70, no. 1-3, pp. 489-501, 2006.

[9] X. Sun, J. Xu, C. Jiang, J. Feng, S-S. Chen and F. He, “Extreme Learning Machine for Multi-Label Classification”, Entropy, vol. 18, pp. 225-236, 2016.

[10] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen and F. Frederiksen, “A Stochastic MIMO radio channel model with experimental validation”, IEEE J. Select. Areas Commun., vol. 20, pp. 1211-1226, 2002.