Kernel Extreme Learning Machine Based Cluster Sparse Proportionate Affine Projection Algorithm for Echo Cancellation

Senthil Murugan.B[1], Sundaravanan. J[2]

[1,2]Assistant Professor/Electronics and Communication Engineering/Thanthai Periyar Government Institute of Technology/Vellore/Tamil Nadu/India

Abstract

The performance of kernel extreme learning machine based CSPAPA (KELM-CSPAPA) is proposed and studied. An adaptive density-based clustering algorithm (ADBC) is used to decompose the training dataset into multiple subsets. Furthermore, the improved flower pollination optimization algorithm (IFPA) is used to optimize the parameters of the KELM model. Initially, the performance of the ADBC algorithm is studied to estimate the optimal number of clusters using three different inputs. Then the KELM model is assessed with different machine learning models using various statistical indices. The time consumption of KELM for the colored noise and speech signal inputs are less compared to the white Gaussian noise input. Finally, a performance comparison of KELM-CSPAPA is presented with three benchmark algorithms: PAPA, block sparse PAPA(BS-PAPA), and CS-PAPA. The result shows that the proposed method outperformed existing algorithms. The simulation results demonstrate that KELM-CSPAPA fully exploits the sparsity and handled the clustered-sparse signal.

Keywords: KELM, Adaptive desnsity based, convergence rate Misalginment, cluster-sparse PAPA

1. Introduction

In recent years, the extreme learning machine algorithm (ELMA) attracted great attention due to its stability, high efficiency, and simple structure. It is widely used in various real-time applications including signal analysis [1, 2]. A combined model using the empirical model, artificial neural networks (ANN), and ELMA has been used to study the Thar Desert in India. It is found that ELMA is better than empirical and ANN models [3]. Recently, an improved ELMA model (IELMA) is proposed by Sebbar *et al*. to estimate monthly evaporation. IELMA is a robust tool to improve the available machine learning models [4]. Besides, various researchers combine other technologies with machine learning such as meta-heuristic optimization algorithms [4, 5], wavelet decomposition [6], K-means clustering [7], weighted chaotic salp swarm kernel ELMA [8], and Multi-kernel support vector regression [9].

The PAPA algorithm is improved further by including cluster sparse characteristics to exactly estimate these sparse channels. In this paper, we propose kernel extreme learning machine based cluster sparse proportionate affine projection algorithm (KELM-CSPAPA) for network echo cancellation. An adaptive density based clustering algorithm (ADBC) is used for PAPA. And decompose the training dataset into multiple subsets using the ADBC clustering algorithm. The IFPA algorithm was used to optimize the parameters of the KELM model. The IFPA is used to assign a threshold and adjust the weight of each neuron of the neural network. The KELM model is compared with four advanced machine learning models: k-nearest neighbors algorithm (KNN), social media optimization (SMO), support vector machine (SVM), and random vector functional link (RVFL). The performance of KELM-CSPAPA is studied and compared with PAPA, BS-PAPA [10], and CS-PAPA [10,11] algorithms. The contributions of the proposed work are as follows:

- It is the first time the adaptive density-based clustering algorithm and flower pollination algorithm have been used in the KELM model for faster and accurate estimation.
- Compare the performance of the proposed KELM-CSPAPA with other widely-used models - PAPA, BS-PAPA, and CS-PAPA.
- Assess the performance of the proposed method using five different statistical indices.

The remainder of this article is structured as follows: In Section 2, the conventional PAPA and PNLMS algorithms are briefly reviewed. In Section 3, the proposed KELM-CSPAPA including density-based clustering for PAPA and the IFPA optimization algorithm, is described. Section 4 explains the performance measurement of KELM. Section 5 describes the simulation results of the performance of the proposed KELM-CSPAPA and comparison results. Finally, the conclusion is presented in Section 6.

2. Proposed Algorithm and Methodology

In this section, an adaptive density based clustering algorithm is introduced for PAPA. The flower pollination algorithm is used to assign a threshold and adjust the weight of each neuron of the neural network. The kernel extreme learning machine algorithm is used in the echo cancellation problem.

2.1 Review of Proportionate affine projection algorithm

Fig.1 shows a model of an echo canceller. In the adaptive filtering algorithm, the impulse response of the unknown system is represented by **h** of length N. The far-end signal is represented by $\boldsymbol{x}(k) = [x(k), x(k-1), x(k-2), \ldots, x(k-N+1)]^T$. The desired signal $d(k)$ can be expressed based on [27] as:

$$d(k) = \boldsymbol{x}^T(k)\boldsymbol{h}(k) + \eta(k) \tag{1}$$

where $\boldsymbol{x}^T(k)$ and $\eta(k)$ denote respectively the transpose operation of $\boldsymbol{x}\,(k)$ and the additive noise. Then, the adaptive filter output is expressed as:

$$y(k) = \boldsymbol{x}^T(k)\mathbf{w}(k-1) \tag{2}$$

The estimated error is expressed as:

$$e(k) = d(k) - y(k) \tag{3}$$

To exploit the sparsity and enhance the speed of convergence PAPA initiates a method that recycles the input signal. The input data matrix is expressed based on [28] as:

$$\boldsymbol{X}(k) = [\boldsymbol{x}(k), \boldsymbol{x}(k-1), \boldsymbol{x}(k-2), \ldots, \boldsymbol{x}(k-P+1)] \tag{4}$$

where P denotes the order of the projection of PAPA. The output vector, the desired signal vector, and the error vector of PAPA is expressed as:

$$\boldsymbol{y}(k) = \boldsymbol{X}^T(k)\mathbf{w}(k-1), \tag{5}$$

$$\boldsymbol{d}(k) = [d(k), \boldsymbol{d}(k-1), \ldots, \boldsymbol{d}(k-P+1)]^T, \tag{6}$$

$$\boldsymbol{e}(k) = \boldsymbol{d}(k) - \boldsymbol{y}(k) \tag{7}$$

where $\boldsymbol{y}(k), \boldsymbol{d}(k)$ *and* $\boldsymbol{e}(k)$ denote respectively the output vector, desired signal vector, and the error vector. Then the updated equation of the PAPA is expressed based on [28] as:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \mu\boldsymbol{G}(k)\boldsymbol{X}(k)\boldsymbol{e}(k) \times (\boldsymbol{X}^T(k)\boldsymbol{X}(k)\boldsymbol{G}(k) + \delta_{PAPA}I_P)^{-1} \tag{8}$$

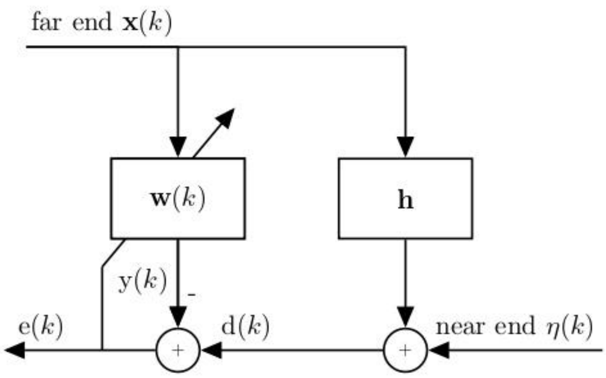

Fig 1. Echo canceller

2.2 Adaptive searching for core points

The core points of ADBC are identified through the global and fixed threshold values. Diversity of local minima leads to incorrect identification. In this paper, adaptive core point estimation is used. On several occasions, the point y_j is high density and nearest to other points and is noted. It is calculated as:

$$K_j = \sum_{i=1}^{n} Y\left(j - \begin{matrix} \arg\min(d_{ij}) \\ j: \gamma_j > \gamma_i \end{matrix}\right), \begin{matrix} if\ y = 0, Y(y) = 1 \\ else\ Y(y) = 0 \end{matrix} \tag{9}$$

It is noted that the point on the cluster border is hard to judge as being the nearest to other points. So a point has $K_j = 0$ are considered within the cluster border. Core points can be identified by comparing the mean density of the points to the high-density points. Now the core region consisting of core points for the t^{th} initial cluster is expressed as:

$$\rho_{core}^t = \{y_i | \gamma_i > M(y_i), y_i \in \rho^t, y_j \in \rho^t, \&\, K_j = 0 \tag{10}$$

where ρ_{core}^t, $M(y_i)$, and ρ^t denote, respectively, the core region consisting of core points, mean density of the points for which K=0, and t^{th} initial cluster. Algorithm I shows the pseudo code of the proposed ADBC-PAPA.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

2.3.1 Performance test of ADBC

The performance of the proposed ADBC has been evaluated and validated using the datasets (speech, colored noise, and white Gaussian noise (WGN)) [11]. The analysis of the proposed KELM-CSPAPA is carried out using Matlab 2020 on a computer with Intel Core i7–4700 CPU and 32 GB RAM. The training dataset was decomposed into multiple subsets based on the ADBC method. Finally, the sub-models were merged into a new model, based on the proposed method. In this study, the training and testing datasets were divided into subsets of 5, 10, 15, 20, 25, 30, and 40 folds. The proposed method and several other methods like Calinski–Harabasz Criterion (CHC) [12], Davies Bounding Criterion (BDC) [13,14], and Gap Value (GV) proposed by Tibshirani et al. [15,16,17] are used to estimate the optimal number of clusters in a given dataset.

The number of clusters identified by the proposed method and three different algorithms for the WGN dataset are shown in Fig. 2. Fig. 2 is created based on the Gaussian kernel function. It shows ADBC exactly identified the number of clusters from the datasets. The number of clusters identified by the proposed method and three different algorithms for the colored noise and speech dataset is shown in Fig. 3 and Fig. 4 respectively. From Fig. 3 and 4, one can easily understand that the proposed ADBC exactly identified the number of clusters present in the dataset with WGN, colored noise, and speech dataset compared to other methods. Moreover, CHC and BDC are not feasible because they detected fewer clusters. Therefore, we used ADBC in the proposed KELM-CSPAPA to find an optimal number of clusters. The GV method also performed better to identify the number of clusters. But, due to its huge time consumption, it is not considered.

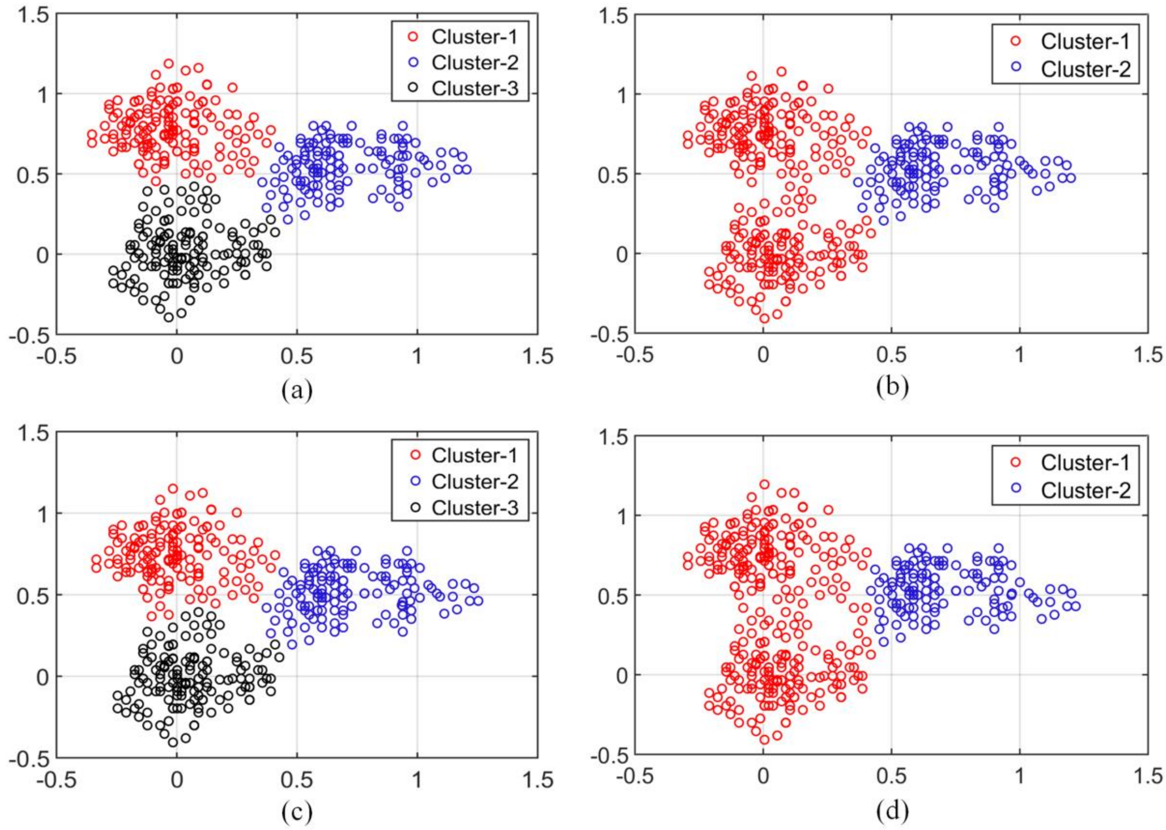

Fig. 2. Number of clusters identified by various algorithms with WGN dataset (a) ADBC (b) CHC (c) GV (d) BDC

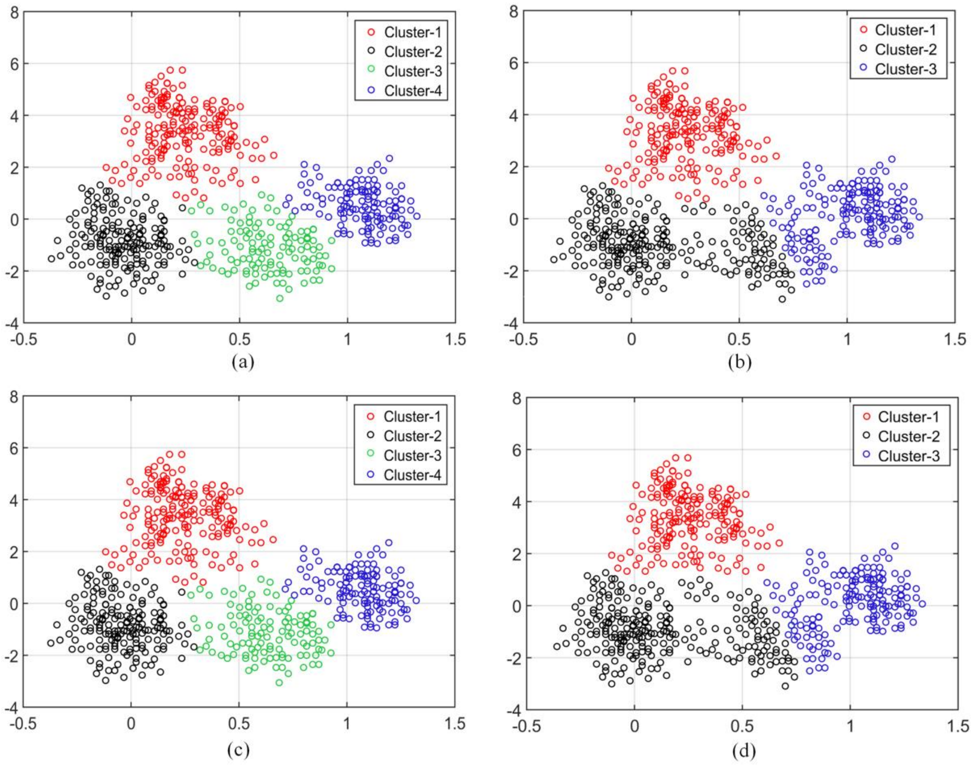

Fig. 3. Number of clusters identified by various algorithms with colored noise dataset (a) ADBC (b) CHC (c) GV (d) BDC

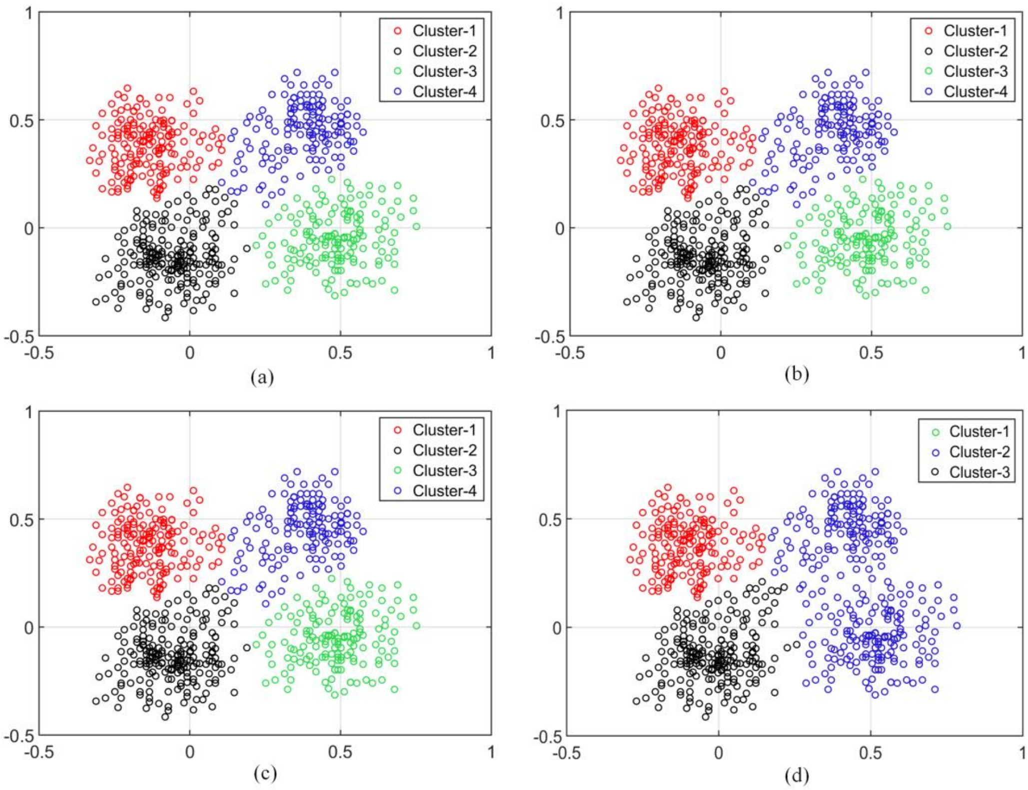

Fig. 4. Number of clusters identified by various algorithms with a speech dataset (a) ADBC (b) CHC (c) GV (d) BDC

Algorithm-I : ADBC-PAPA Algorithm

Input:
A set of K objects in a spatial area, A={a_1, a_2, …, a_k}, with N neighbour list
Radius for non-spatial and spatial data objects (R_{sn})
Minimum number of existing points in the neighbourhood of R_{sn} (M_{pt})
Threshold value of the cluster (T_c)
Output:
Core objects of the cluster and noise points, C={c_1, c_2, …, C_k}
Method:
Initialize the layers of the cluster =0;
For selecting objects from the database form a loop
    For i=1 to n do
        If arbitrary object a_k not belong to any cluster, then
            Move to next point
        F=query for the neighbour (a_i, R_{sn});
            If sizeof M_{pt} > F then
                Take next point of a_i as noise
            Else
            C_Layer = C_Layer+1;
        For i=1 to sizeof(F)
        End
Expand cluster (push()) // pushing all points to F
While (A!=empty()) // Stop the process when database is empty
Object = pop(); //Enable pop operation
W=Estimate_neighbour()
    If R_{sn} > W then
        For j=1 to a_k
    If Tc < W then
            Add a with the present cluster
    End if
  End if
        Push(a);
End for
  End for
  End while
Return

3. Improved Flower pollination algorithm

The flower pollination algorithm (FPA) is based on the flower pollination process of a plant. It is a population-based optimization algorithm. Each solution has a set of pollens called flowers [18]. The performance of FPA is evaluated through the fitness value. Each flower of the FPA is evolving through the local pollination process (LPP) and global pollination process (GPP).

A random number with a range (0,1) is generated for each flower before the pollination process starts. For each flower, anyone pollination method is performed based on rand with the switching probability p. The GPP is performed if rand > p, else LPP is performed.

Kernel extreme learning machine

The extreme learning machine (ELM) is first proposed by Huang et al. [19] with a single hidden layer feed-forward neural network (SFNN). The ELM model used in this study. The generalized SFNN output function for the ELM is expressed as:

$$f_s(x) = \sum_{i=1}^{N} \mu_i \, h_i(x) = h(x)\mu \quad (11)$$

where $\mu = [\mu_1, \mu_2, \dots, \mu_n]^T$ represents the output weight vector between the output node to the hidden layer with N nodes. The output vector of the hidden layer concerning the input x is expressed as:

$$h(x) = [h_1(x), h_2(x), \dots, h_N(x)] \quad (12)$$

4. Performance measures

To predict the performance of KELM, four different machine learning algorithms - KNN, SMO, SVM, and RVFL - are used. The parameters of KNN are set as α = 10 and the number k of nearest neighbours is set to 5, for SMO the activation function is set Radial Basis Function (RBF), the parameters of SVM are set based on the original function, and for RVFL, the number of nodes in the hidden layer is set to 100 and the activation function is set to tribas.

The performance of the proposed KELM for clustered sparse affine projection algorithm is evaluated using five statistical error criteria: root mean square error (RMSE), coefficient of determination R^2, mean absolute error (MAE), efficiency coefficient (EC), and overall index (OI). The RMSE is the standard deviation of the differences between actual and predicted data, and it can be expressed as based on [16].

5. Simulation Results

The efficacy of the proposed KELM-CSPAPA algorithm is assessed concerning the effect of projection order, group size, and convergence speed. The obtained results of KELM-CSPAPA are compared with different techniques, such as PAPA, BS-PAPA, and CS-PAPA. The simulation parameters follow as: the step size μ=0.1 for WGN and μ=0.2 for colored noise speech signal; $\varepsilon = 0.01$; and a=5/N. In all the experiments, WGN, colored noise, and the speech signal is considered as an input and the adaptive filter length is L=1024. The sampling frequency of the speech and colored noise is 8 kHz. In this study, two clusters such as single-cluster and double-cluster have been taken. The single-cluster system is performed for 25,000 iterations and continues the double-cluster system for the next 25,000 iterations. The evaluation metrics employed for validation is the normalized misalignment (NMIS), which is defined as $10log(\|w-\hat{w}\|_2^2)/\|w\|_2^2$. Other parameters used in the simulation are shown in Table 2.

Table 2. Simulation parameters

	a	b	ρ	σ
PAPA	0.02	0.015	-	-
BS-PAPA	0.015	0.013	-	-
CS-PAPA	-	-	0.01	0.01

5.1 Performance of KELM-CSPAPA on projection order

Initially, the performance of the proposed method is studied by changing the values of different parameters of KELM-CSPAPA. The projection order is varied (M=2, 4, and 8) for the three inputs (WGN, colored noise, and speech signal). The scatter plot of the three inputs for different values of M is illustrated in Fig. 5. It is plotted against the experimental and predicted values of NMIS. From Fig. 5 it is observed that the experimental and predicted values are very close to the 1:1 line. The data points of the colored noise are more scattered than the WGN and Speech input. It should be noted that the proposed KELM-CSPAPA method distributed the data points more evenly on both sides of the 1:1 line.

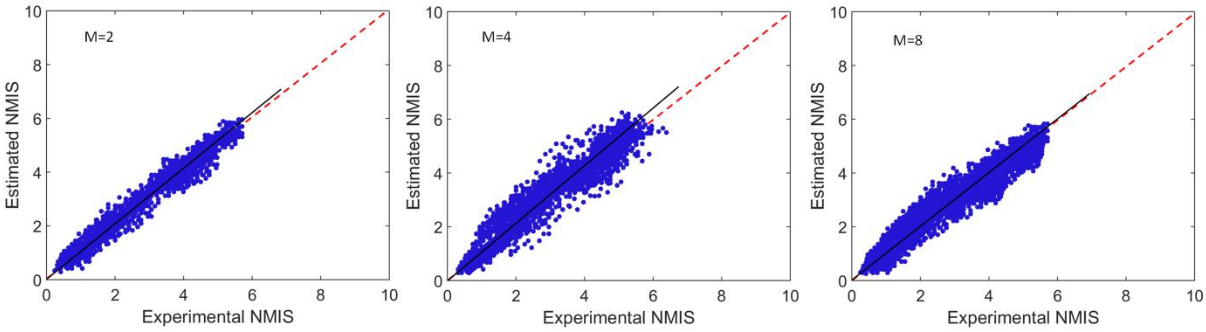

Fig. 5. Scatter plots of the performance of KELM-CSPAPA for M=2, 4, and 8 with WGN input

Similarly, the scatter plot of the three inputs for different values of M for the colored noise and the speech inputs are illustrated in Fig. 6 and 7 respectively. The fitting lines between the experimental and predicted values for M=4 and 8 deviate significantly from the 1:1 line than M=2 for the colored noise input. On the other hand, the line passes through the experimental and predicted values and are scattered in the vicinity of the line for all the M values for the speech signal.

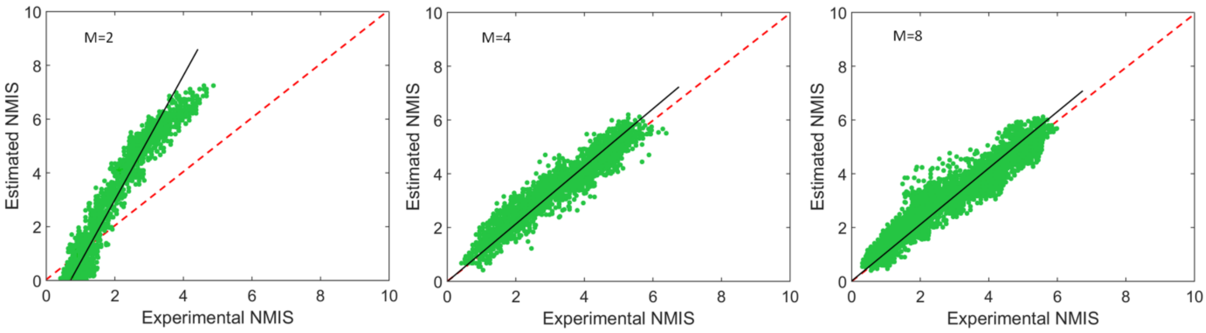

Fig. 6. Scatter plots of the performance of KELM-CSPAPA for M=2, 4, and 8 with colored noise input

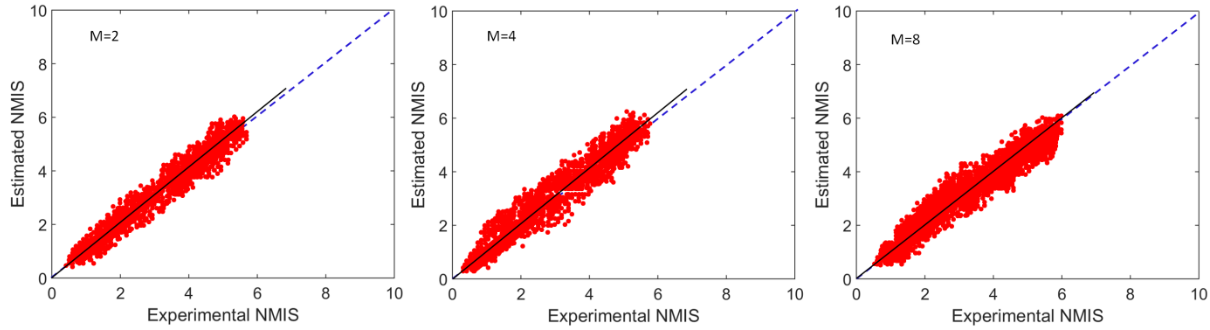

Fig. 7. Scatter plots of the performance of KELM-CSPAPA for M=2, 4, and 8 with speech input

5.2 Convergence Performance of Acoustic Echo Path and a Random Dispersive System

In this experiment, the convergence performance of the proposed KELM-CSPAPA for the Acoustic Echo Path and a Random Dispersive System was studied and compared with other algorithms such as PAPA, BS-PAPA, and CS-PAPA with G=16. Switching the 512 taps measured acoustic echo path to a random impulse response was done by changing the echo path at 30000 iterations. The simulation is carried out for WGN, colored noise, and speech input signals, and the results are shown in Fig. 8. Fig. 8(a) shows the simulated output for the WGN input with μ=0.2. It should be noted that the proposed KELM-CSPAPA outperformed the PAPA, BS-PAPA, and CS-PAPA for both the acoustic echo path and the random dispersive impulse response. Similarly, the obtained output of the proposed KELM-CSPAPA and the other algorithms for the colored noise and speech input are shown in Fig. 12(b) and (c) respectively. It is observed that the proposed KELM-CSPAPA works better than the family of PAPA algorithms for the random dispersive impulse responses.

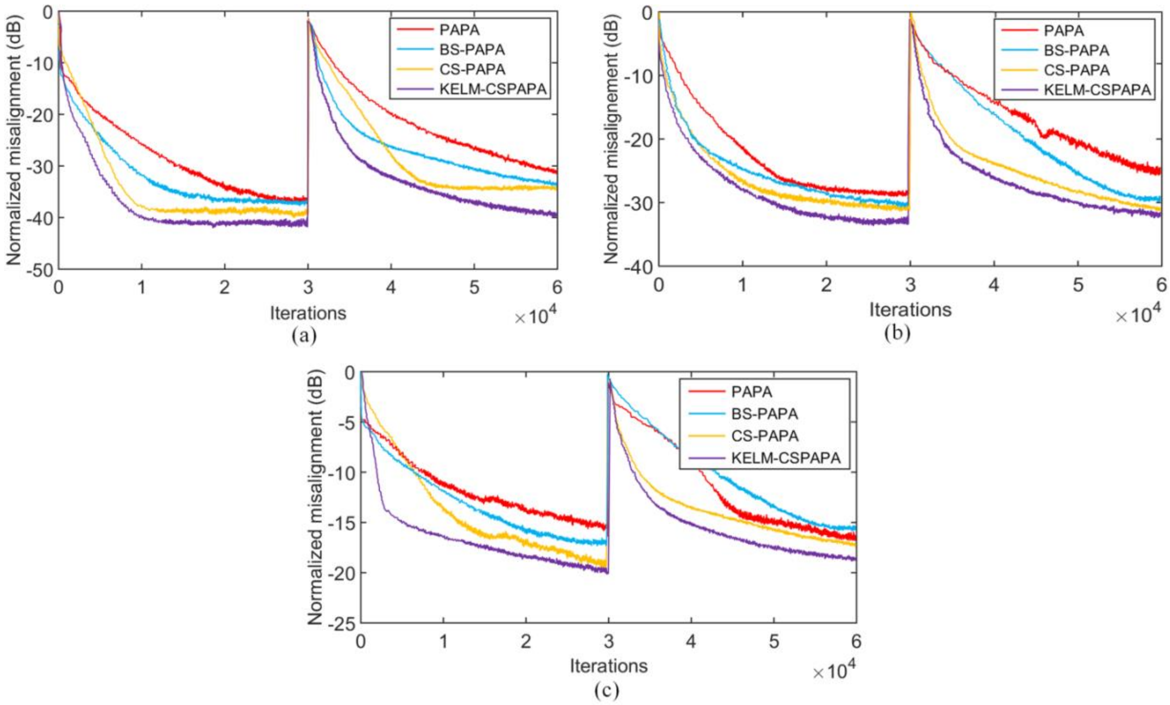

Fig. 8. Comparison of the proposed KELM-CSPAPA with other algorithms for acoustic echo path and dispersive system (a) WGN input (b) coloured noise input (c) speech input

6 Conclusion

In this paper, KELM-CSPAPA has been proposed for the network echo channel estimation problem. In this study, instead of using conventional CSPAPA, PAPA with adaptive density-based clustering algorithm (ADBC) is used to achieve CSPAPA. Three datasets such as WGN, colored noise, and speech signal are used to study the performance of ADBC. From all datasets, the ADBC method provides an exact number of clusters with lesser processing time. Besides, IFPA is used to optimize the parameters of the KELM model for adjusting the weight of each neuron of the neural network. Five different statistical indices have been used to study the effectiveness of KELM. The statistical performance of KELM is compared with other machine learning models such as KNN, SMO, SVM, and RVFL. The result shows that KELM performed better in almost all the input cases. The simulation is carried out to study the effectiveness of the convergence of clustered-sparse system and the convergence of a random dispersive system. The performance of the proposed KELM-CSPAPA is compared with three other algorithms - PAPA, BS-PAPA, and CS-PAPA. The results show the proposed KELM-CSPAPA is suitable for the echo cancellation problem.

References

1. D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Trans. Speech Audio Process., vol. 8, no. 5, pp. 508–518, Sep. 2000
2. H. Deng and M. Doroslovacki, ''Improving convergence of the PNLMS algorithm for sparse impulse response identification,'' IEEE Signal Process. Lett., vol. 12, no. 3, pp. 181–184, Mar. 2005
3. C. Paleologu, S. Ciochina, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," IEEE Signal Process. Lett., vol. 17, no. 2, pp. 165–168, Feb. 2010.
4. Z. Wu, J. Shi, X. Zhang, W. Ma, B. Chen, Kernel recursive maximum correntropy, Signal Processing 117 (2015) 11–16.
5. P. Zhu, B. Chen, J.C. Principe, Learning nonlinear generative models of time series with a Kalman filter in RKHS, IEEE Transactions on Signal Processing 62 (2014) 141–155.
6. S. Zhao, B. Chen, P. Zhu, J.C. Principe, Fixed budget quantized kernel least-mean-square algorithm, Signal Processing 93 (9) (2013) 2759–2770.
7. B. Chen, X. Liu, H. Zhao, J.C. Principe, Maximum correntropy Kalman filter, Automatica 76 (2017) 70–77.
8. H. Qu, W. Ma, J. Zhao, B. Chen, Kernel least mean kurtosis based online chaotic time series prediction, Chinese Physics Letters 30 (11) (2013) 110505.
9. W. Ma, J. Duan, W. Man, H. Zhao, B. Chen, Robust kernel adaptive filters based on mean p-power error for noisy chaotic time series prediction, Engineering Applications of Artificial Intelligence 58 (2017) 101–110.
10. Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.W., 2018. An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. 569, 387–408.
11. Fan, J., Wu, L., Ma, X., Zhou, H., Zhang, F., 2020. Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. Renew. Energy 145, 2034–2045.

12. Patil, A.P., Deka, P.C., 2016. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. Comput. Electron. Agric. 121, 385–392.
13. Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM–a new approach. J. Hydrol. 572, 556–570.
14. Figlu Mohanty, Suvendu Rup, Bodhisattva Dash, Banshidhar Majhi, M.N.S.Swamy, An improved scheme for digital mammogram classification using weighted chaotic salp swarm algorithm-based kernel extreme learning machine, Applied Soft Computing, 91 (2020).
15. I. Boulkaibet, K. Belarbi, S. Bououden, M. Chadli, T. Marwala, An adaptive fuzzy predictive control of nonlinear processes based on Multi-Kernel least squares support vector regression, Applied Soft Computing 73 (2018) 572-590.
16. H. Zhao, Y. Yu, S. Gao, X. Zeng and Z. He, "Memory Proportionate APA with Individual Activation Factors for Acoustic Echo Cancellation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 6, pp. 1047-1055, June 2014
17. A. H. Sayed, Fundamentals of Adaptive Fltering. Hoboken, NJ, USA: Wiley, 2003.
18. T. Gansler, J. Benesty, S. L. Gay, and M. M. Sondhi, ''A robust proportionate affine projection algorithm for network echo cancellation,'' in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Istanbul, Turkey, Jun. 2000, pp. 5–9.
19. K. T. Wagner and M. I. Doroslovacki, ''Complex proportionate-type normalized least mean square algorithms,'' in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Kyoto, Japan, Jul. 2013, pp. 3285–3288.
20. Z. Aoying and Z. Shuigeng, "Approaches for scaling DBSCAN algorithm to large spatial databases," Journal of Computer Science and Technology, vol. 15, no. 6, pp. 509–526, 2000.

AUTHOR PROFILES

.

Senthil Murugan Boopalan received his B.E from Madras University, M.E from Anna University, and Ph.D from Anna University. He is working as Assistant Professor in the Department of Electronics and Communication Engineering at Thanthai Periyar Government Institute of Technology, Vellore. He has more than 20 years of teaching experience. He is currently pursuing his research in the field of Signal Processing, Communication and Sensor Networks.

Sundaravanan Jothiprakasam received his B.E from Bharathiyar University, M.E from Sathyabama University and Ph.D from Anna University. He is working as Assistant Professor in the Department of Electronics and Communication Engineering at Thanthai Periyar Government Institute of Technology, Vellore. He has more than 10 years of industry experience and 8 years of academic experience. His research interest includes Wireless Sensor Networks, Embedded Systems, and Signal Processing.