Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6 July 2021: 5477 – 5486

Research Article

Relief based Optimized Feature Selection for Online Sequential Extreme Learning Machine

***Archana P. Kale,**

Modern Education Society's College of Engineering, Wadia Campus, 19, Late V.K.Joag Path, Department of Computer Engineering, SPPU Pune, P.O. Box: 411001, Pune, Maharashtra, India; Tel/Fax:; E-mails: archana.mahantakale@gmail.com

Abstract

Extreme learning machine (ELM) is a rapid classifier, evolved for batch learning mode which is not suitable for sequential input. As retrieving of data from new inventory which is leads to time extended process. Therefore, Online sequential ELM (OSELM) algorithm is progressed by Liang et. al. which is able to handle the sequential input in which data is read 1 by 1 or chunk by chunk mode. The overall system generalization performance may devalue because of the amalgamation of random initialization of OS-ELM and the presence of redundant and irrelevant features. To resolve the said problem, this paper proposes a correspondence multimodal genetic optimized feature selection paradigm for sequential input (RF-OSELM) for radial basis or function by using clinical datasets. For performance comparison, the proposed paradigm experimented and evaluated for ELM, multimodal genetic optimized for ELM classifier (RF-ELM), OS-ELM, RF-OSELM. Experimental results are calculated and analysed accordingly. The comparative results analysis illustrates that RF-ELM provides 10.94% improved accuracy with 43.25% features as compared to ELM.

Keywords: Online sequential Extreme Learning Machine, Genetic Algorithm, Feature Selection problem, Classification Problem

1. INTRODUCTION

Now a days Artificial intelligence is the growing and critical area [1]. Feature subset selection (FSS) is an intricate procedure in the fields of artificial intelligence and machine learning. The prime objective of feature selection is to adopt optimal features for further evaluation. The features which are relevant, non-redundant are called as optimal features. It is entangled to decide the significance of features [2]. To enhance the system generalization performance it is necessary to search and finalize only optimal features.

Various FSS algorithms like Half selection, Neural Network for threshold, Mean Selection etc. are used for FSS. Random search based Genetic algorithm is the random searched optimization techniques which uses to select optimal feature subset [3]. The big search space is handled by GA very effectively [4] and has maximum chances of a global optimal solution.

Extreme Learning Machine is rapid classifiers which have various advantages like good generalization performance, high speed, require less training time etc. ELM is primarily designed for batch mode in which all data is available before training. However, it is not suitable for sequential input [5]. Therefore, OS-ELM is designed by Liang et. al. for sequential input. Zhu et.al developed Evolutionary ELM [6] and Han et. al. developed the particle swarm optimization based Evolutionary ELM [7]. In many papers, for ELM sig activation function is used [8] [9]. Huang et. al. designed Incremental – ELM [12]. ELM is also used to solve the real time applications like medical data classification [13], universal approximation [14], big data [17].

The original Extreme learning machine (ELM) is primarily designed for batch mode. Nan-Ying et. al. [5] emerged online sequential – ELM (OS-ELM) for linear, incremental or sequential input. As ELM and OS-ELM calculates the input to hidden layer neurons by randomly assigning the specified input weights and biases and the target output is calculated [10] [11] by analytically evaluation the weights in-between hidden layer and resulted layer as shown in figure 1. So, the generalization performance of the system may deteriorate due to the random initialization. One of the most significant steps is required i.e. optimal feature subset selection.

The key intent of paper is the innovative use of a genetic algorithm with multimodal optimization approach for OS-ELM (RF-OSELM) for clinical datasets. In various papers, authors are evaluated OSELM only by changing hidden nodes, but an extensive literature break down to recognize the changes in the inceptive training data (block) according to the quantities of hidden nodes.

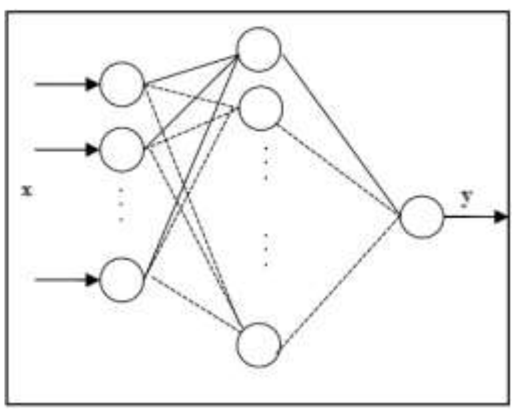

Figure 1: Basic ELM Architecture

The structure of the paper is organised as follows: The detail structure of the proposed methodology of RF-OSELM with the aid of paradigm is detailed in Section 2. Innovative results and comparison of results are mentioned in section 3. Finally, the future work in combination with conclusion are described in section 4.

2. PROPOSED RF-OSELM APPROACHES

The paradigm of the proposed RF-OSELM approach is as shown in figure 2 which is categorized into threefold subsystems – a. Pre-processing subsystem, b. FSS subsystem 3. Classification Subsystem.

2.1 Datasets

The various datasets like Pima Indian Diabetes (PID), Statlog heart disease (SHD), Breast Cancer (BC), Australian (AS) [15] [16] are used. The dimensional scope of these datasets is from 8 to 36. Most of the considered datasets are clinical datasets which are standard UCI repository datasets. PID, SHD, BC and AS datasets contain 8, 13, 10, 14 attributes and 768, 270, 699, 690 instances respectively.

2.2 Preprocessing Subsystem

Data Normalization is used in pre-processing subsystem. Data normalization is an intricate pre-processing method which is used for various artificial intelligence and machine learning algorithms [17]. The features present in the dataset are of different scale. So it becomes very critical to handle such type of vector space as it contains the maximum range account of, a vital task to convert all vector spaced feature to unit space feature which result lies between zero and one.

After normalization, all datasets are further divided into duplet parts i.e. training set in which 70% instances are considered and remaining 30% instances are considered for testing set. For example, total number of instances present in PID dataset are 768, which are further divided into 538 instances are used as training set and 230 instances are used for testing set. Same for SHD dataset, total number of instances present in SHD dataset are 270, which are further divided into 189 instances are used as training set and 81 instances are used for testing set. For BC dataset, total number of instances present are 699, which are further divided into 490 instances are used as training set and 209 instances are used for testing set. For AS dataset, total number of instances present are 690, which are further divided into 483 instances are used as training set and 207 instances are used for testing set.

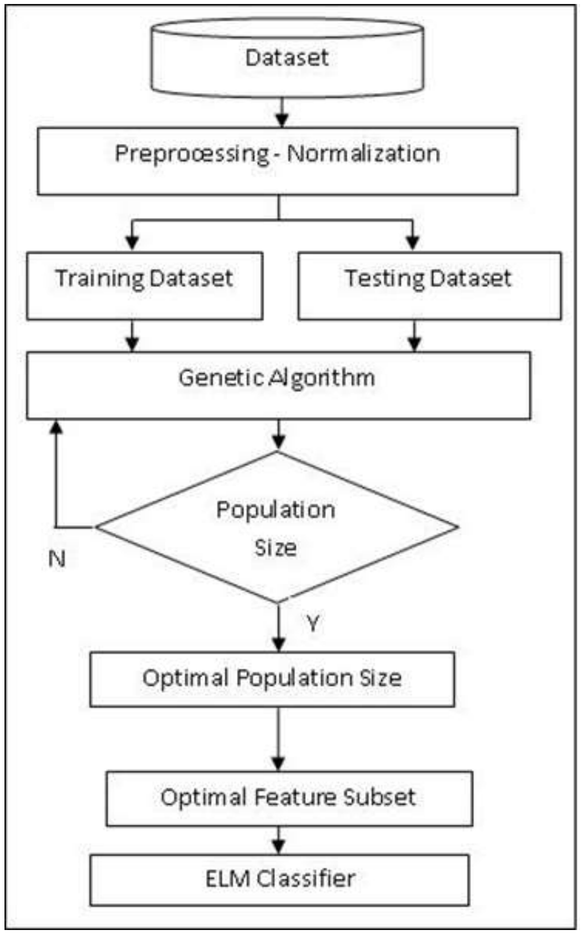

Figure 2: Proposed RF-OSELM Paradigm

2.3 Feature Subset Selection Subsystem

Thousands of features are present in high dimensional dataset. For classification, all features are not required as it may be the presence of non-optimal features i.e. irrelevant and redundant features. These features act as a noise which degrades the predictive accuracy.

Genetic Algorithm (GA) is relevant techniques for selecting ideal features. GA contains the population which is a collection of a set of possible solutions to solve the problem [18]. Three steps like selection, evaluation and recombination are executed in every iterative step. Selection, Crossover and Mutation are the genetic operators mainly used in GA. The number of iteration is depended on the condition of termination. Based on the quality, the fitness function is evaluated. And based on the evaluated fitness value, the strings are selected for new generation which have comparatively the super power than other strings. From the population, the points are eliminated in which moderate fitness value is present. For exploration, especially mutation and crossover are utilized in order to obtain the new solutions [19]. Mutation is major contribute to change a part genetic randomly. Crossover is used to incorporate the fittest members of genetic material from population.

GA provides different result of feature subset as per the changes in population size. In literature survey, various authors use the 50 and 70 as the population size. However, the vast literature survey has disadvantages that proper selection of population size is absent. Therefore multimodal genetic optimized feature selection paradigm is proposed for batch input as well as for sequential input in context of this paper. Here, feature subset is finalised by considering the various population size from 10:10:90. Thus, the total number of feature subset are evaluated are 9. One feature subset is selected for further experimentation as an ideal feature subset which provides maximum accuracy.

2.4 Classification Subsystem

By using optimal features OS-ELM is evaluated. Initialization and sequential learning phase are major two phases of OS-ELM. In initialization phase, various parameters executed like number of data required to fill up, the number of hidden nodes, defining of chunk size etc. Target class is decided based on initialized data and new arrived chunk data in sequential learning phase [21] [22].

3. EXPERIMENTAL RESULTS

Related experimentation has been conducted in MATLAB© R2014a. The two activation functions sigmoidal and radial basis activation functions (rbf) are used for simulation. For evaluation, in the literature survey, various authors consider only the quantity of hidden nodes in OS-ELM [5] [10] [11], but the initial block size is not considered for evaluation. By the virtue of experiment, it is observed that the initial training data is much important. So every step of hidden layer, training data (n) is also changing with the number of hidden node (j) like j to n with incremental I value. The output results are calculated by using fixed chunk size (1or 20) or randomly changing the chunk size between 10 to 30.

For evaluation, the accuracy measure is utilised an evaluation measure. The performance metrics are computed by evaluating the values of false negatives and positives both as well as true negatives and positives also [20]. The equation 1 shows the formula of the calculation of accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad .....(1)$$

As experimental results of ELM and OS-ELM by using all features for binary classification problem are given in Table 1. The evaluation performance is compared in thrice different ways - 1. RF-ELM and ELM 2. RF-OSELM and OS-ELM.

3.1 RF-ELM and ELM

As enhance efficiency and effectiveness of RF-ELM paradigm, the clinical datasets are used for experimentation by using ELM classifier. Table 2 indicates GA results changing with size of population. Total 9 feature subsets are evaluated by varying population size. Out of all these subsets, one subset is finalized which has maximum occurrence as an optimal subset as shown in the second last column in the Table 2. For example, for PID dataset the accuracy is calculated by differentiating the value of population. For each subset, the accuracy is shown. From all these subset one subset need to be finalize which produced maximum accuracy. For PID dataset optimal feature subset is { 2,5,6 } with accuracy 77.82 %. ELM classifier is source to computation of the classification accuracy. Comparative result analysis of ELM and RF-ELM is as shown in Table 3. With the result analysis, it is observed that RF-ELM has success to achieve 10.94% improved classification accuracy over 56.75% reduction in features as in ELM.

3.2 RF-OSELM and OSELM

The experimental results are calculated by using optimal features for ELM and OS-ELM is as shown in Table 4.The comparative performance between OS-ELM and RF-OSELM is as shown in Table 4 by calculating the average of sequential mode (1-by-1, 20-by-20 and [10,30]). Table 4 indicates the detailed comparative analysis of OSELM and RF-OSELM by using both activity function like sig and rbf.

Table 1 Experimental Results of ELM and OSELM on Binary classification application

Dataset	Act. Fun.	Algorithm	Learning Mode	Accuracy		Node	Initial Block Size
PID	Sig	ELM	Batch	87.91	81.30	50	-
		OSELM	1-by-1	88.29	82.60	115	465
			20-by-20	87.36	82.17	25	175
			10,30	83.82	83.04	15	65
	Rbf	ELM	Batch	88.66	81.73	157	-
		OSELM	1-by-1	87.17	82.17	80	180
			20-by-20	87.91	82.17	80	180
			10,30	81.59	83.47	45	245
SHD	Sig	ELM	Batch	99.47	88.88	17	
		OSELM	1-by-1	99.47	88.88	25	125
			20-by-20	99.47	88.88	20	70
			10,30	99.47	91.35	30	30
	Rbf	ELM	Batch	99.47	86.41	89	
		OSELM	1-by-1	96.29	87.65	50	100
			20-by-20	98.41	88.88	30	30
			10,30	98.94	90.12	20	20

Table 2 RF-ELM for binary classification problem for clinical dataset

	Population Size										Optimal Subset	ELM Accu-Racy
Data Set	All	10	20	30	40	50	60	70	80	90	-	-

PID	69.56	2,4,8	2,5,6 2,5,6		2,5,8	2,5,6	2,5,6	2,5,6	2,5,6	2,5,6	2,5,6	77.82
SHD	**77.77**	3,8,9, 10,13	2,3,10, 12,13	1,2,3, 7,12,13 9,12	1,2,3,	3,8,9, 10,13	1,2,3, 9,12	1,2,3, 9,12	1,2,3, 9,12	1,2,3, 9,12	1,2,3, 9,12	83.95
BC	85.16	2,7,8, 9	2,7,8, 9	2,7,8, 9	2,7,8, 9	2,3,4, 5,6,7, 10	2,3,4, 5,6,7, 10	2,3,4, 5,6,7, 10	2,3,4, 5,6,7, 10	2,3,4, 5,6,7, 10	2,3,4, 5,6,7, 10	99.52
AS	73.91	3,4,8, 9,11	3,5,8, 9	5,7,8, 10,11	3,5,8, 9	3,5,8, 9	3,5,8, 9	3,5,8, 9	3,5,8, 9	3,5,8, 9	3,5,8, 9	88.88

Table 3 Comparative analysis of ELM and RF-ELM

Dataset	ELM (%)	RF-ELM(%)	Accu-racy(%)	All Features	Feature Subset	Reduction %	Used Features
PID	69.56	77.82	8.26	8	3	63	37
SHD	77.77	83.95	6.18	13	5	62	38
BC	85.16	99.52	14.36	10	7	30	70
AS	73.91	88.88	14.97	14	4	72	28
Average	76.6	87.54	10.94	-	4.75	56.75	43.25

Table 4 Experimental Results of RF-ELM and RF-OSELM on classification

Dataset	Act. Fun.	Algorithm	Learning Mode	Accuracy		Node	Initial Block Size
				Training	Testing		
PID	Sig	RF -ELM	Batch	87.36	81.73	43	
		RF -OSELM	1-by-1	83.27	81.73	20	20

			20-by-20	82.71	81.73	55	205
			[10,30]	81.41	82.60	55	55
	Rbf	RF -ELM	Batch	82.89	81.73	12	
		RF - OSELM	1-by-1	83.64	81.30	40	290
			20-by-20	83.64	81.30	45	395
			[10,30]	81.41	82.60	55	305
SHD	Sig	RF -ELM	Batch	99.47	83.95	91	
		RF - OSELM	1-by-1	87.83	83.95	10	10
			20-by-20	89.41	83.95	10	40
			[10,30]	89.41	86.41	10	14
	Rbf	RF -ELM	Batch	99.47	85.18	36	
		RF - OSELM	1-by-1	87.30	85.18	20	30
			20-by-20	87.30	85.18	15	33
			[10,30]	88.88	86.41	15	29

5. CONCLUSION

Genetic Algorithm is top priority based optimization algorithm which to classify best of optimal feature subset. However, GA varies its result as per the changes in the population. To solve this problem, in this paper multimodal genetic optimized feature selection paradigm is proposed for sequential input (RF-OSELM) by using clinical datasets. The proposed paradigm is accomplished to handle the dimensionality reduction and optimization problems for sequential input. OS-ELM algorithm is used for sequential input and it trains only new arrival data instead of the whole training dataset which saves the computational cost. In order to prove the importance and strength of the RF-OSELM, comparative study of results for ELM, OS-ELM, RF-ELM are carried out. Here, the RF-OSELM paradigm is evaluated for binary classification problem. The work can be extended by using archetype for multiclass classification problem and improved shuffled frog leaping algorithm [23] which may support inclusive of clear insight and direction regarding future improvements.

ACKNOWLEDGEMENTS

I am gratefully acknowledge Modern Education Society's College of Engineering (MESCoE), Pune for the support. I am extremely thankful to Dr. S. S. Sarawade, Principal, MESCOE for his moral support. I am indebted to Dr. A. J. Hake, Secretary, MESCoE, Pune for his enriching advices and strong support. I am thankful to Dr. Sampada Joshi, Chief Finance Officer, MESCoE, Pune for her encouragement. I also wanted to thank all the persons who helped us in direct or indirect way.

REFERENCES

1. C a m b r i a, E r i k, G u a n g – B i n H u a n g, O t h e r. Extreme learning machines [trends & controversies]. IEEE Intelligent Systems 28(6). 2013, 30-59.
2. Y u, L e i, H u a n L i u, and I s a b e l l e G u y o n. Efficient feature selection via analysis of relevance and redundancy. Journal of Machine Learning Research 5, 2004, 1205-1224.
3. P a h u j a, G u n j a n and T. N. N a g a b h u s h a n. A novel ga-elm approach for Parkinson's disease detection using brain structural t1-weighted mri data. In IEEE Cognitive Computing and Information Processing (ccip), second international conference on 2016, 1-6.
4. M a o, L i, L i d o n g Z h a n g, O t h e r. Improved extreme learning machine and its application in image quality assessment. Mathematical Problems in Engineering. 2014.
5. L i a n g, N a n – Y i n g, G u a n g – B i n H u a n g, and O t h e r. A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Transactions on neural networks, , 2006, 17 (6): 1411-1423.
6. Z h u, Q i n – Y u, A K a i Q i n, P o n n u t h u r a i and O t h e r. Evolutionary extreme learning machine. Pattern recognition, 2005, 38 (10): 1759-1763.
7. H a n, F e i, H a i – F e n Y a o and Q i n g – H u a L i n g. An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing, 2013, 116: 87-93.
8. C h e n, Z h i x i a n g X and O t h e r. A modified extreme learning machine with sigmoidal activation functions. Neural Computing and Applications, 2013, 22 (3-4): 541-550.
9. N g u y e n, D e r r i c k, and B e r n a r d W i d r o w. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In IEEE IJCNN international joint conference on Neural networks, 1990, 21-26.
10. W o n g, S h e n Y u o n g, and O t h e r. A truly online learning algorithm using hybrid fuzzy artmap and online extreme learning machine for pattern classification. Neural Processing Letters, 2015 42 (3): 585-602.
11. L a n, Y u a n, Y e n g C h a i S o h, and G u a n g – B i n H u a n g. Ensemble of online sequential extreme learning machine. Neurocomputing, 2009, 72 (13): 3391-3395.
12. H u a n g, G u a n g – B i n, and L e i C h e n. Enhanced random search based incremental extreme learning machine. Neurocomputing, 2008, 71 (16): 3460-3468.
13. M, S e e r a , and L i m C P. A hybrid intelligent system for medical data classification extreme learning machine: theory and applications. Expert System Application, 2014, 41 (05): 2239-2249.
14. Z h a n g, R u i, Y u a n L a n, G u a n g – b i n H u a n g, and Z o n g – B e n X u. Universal approximation of extreme learning machine with adaptive growth of hidden nodes. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23 (2): 365-371.
15. A s u n c i o n, A r t h u r, and D a v i d N e w m a n. UCI machine learning repository. 2007.
16. Li J, Liu H e n t r i d g e. Bio-medical data set repository. School of Computer Engineering, Nanyang Technological University, Singapore. 2004.
17. A k u s o k, A n t o n, K a j – M i k a e l B j o r k, Y o a n M i c h e and A m a u r y L e n d a s s. High-performance extreme learning machines: a complete toolbox for big data applications. In: IEEE Access 3, 2015, 1011-1025.
18. Z h u o, L i, J i n g Z h e n g, and O t h e r. A genetic algorithm based wrapper feature selection method for classification of hyper spectral images using support vector machine. In Geoinformatics 2008 and joint conference on built environment: Classification of remote sensing images, 2008, 71471-71471.

19. Fei, Ye, and Han Min. Simultaneous feature with support vector selection and parameters optimization using ga-based svm solve the binary classification. IEEE international conference on Computer communication and the internet (iccci), 2016, 426-433.
20. Nahato, Kindie Biredagn, and Other. Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets. Informatics in Medicine Unlocked, 2016, 2: 1-11.
21. Huang, Guang – Bin, and Other. An efficient sequential learning algorithm for growing and pruning rbf (gap-rbf) networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2004, 34 (6): 2284-2292.
22. Huang, Guang – Bin, and Other. A generalized growing and pruning rbf (ggap-rbf) neural network for function approximation. IEEE Transactions on Neural Networks, 2005, 16 (1): 57-67.
23. Hu, Bin, Yongqiang Dai, Yun Su, and Other. Feature selection for optimized high-dimensional biomedical data using the improved shuffled frog leaping algorithm. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2016.