

A Neural Network Selection Approach of Constructing Ensemble Based on CS Algorithm

Jun Hua¹, Tiecheng Song^{1#}, Jing Hu¹, Yu Jiang¹, Shidong Liu², Jinghong Guo²

1.School of Information Science and Engineering, Southeast University, 211189, China

2.Research Institute of Information Technology & Communication, China Electric Power Research Institute, 210003, China

#Email: songtc@seu.edu.cn

Abstract

Selective neural network ensemble is at present a heated research issue in the field of neural network, aimed at improving classification results accompanied by boosting generalization ability. Based on existing techniques, a neural network selection algorithm of constructing ensemble named CSEN is presented, which, to some extent, tackles the problem of selecting component neural networks. Experiments were conducted to test the validity of CSEN on several datasets and results show strong stability and high running efficiency of the algorithm. Moreover, its misclassification rate turns out to be relatively lower compared with its counterparts and leads to a much better performance.

Keywords: *Neural Network, Cuckoo Search Algorithm, Selective Ensemble*

1 INTRODUCTION

As is known to all, neural network has been successfully applied in numerous fields; the working effects of neural network, however, tend to be greatly influenced by users' experience. In fact, selection of the architecture of neural network currently lacks rigorous guidance and the generalization ability of network usually struggles to improve. Accordingly, a novel notion named neural network ensemble was presented by Hasen^[1], being simple and easy to use and effectively improving the neural network's generalization ability, thus becoming a heated issue in various research fields^{[2][3]}.

One main aspect of research on neural network ensemble is how to employ individual networks to make up the ensemble. Selective ensemble picks a subset of individual networks from the set of candidate ones available^[4], revealing fine generalization ability with less space occupied according to existing research results. The research direction of selective neural network ensemble can be primarily divided into two categories: one is discovering appropriate individual ones to constitute the ensemble using searching algorithms, and the other is finding the ones of great diversities utilizing clustering approach.

Zhou et al.^[5] proposed the GASEN approach that selects the individual neural networks by Genetic Algorithm (GA) and the results showed that selective ensemble is superior to assembling all of the individual neural networks at hand. In paper [6], selective ensemble called CLUSEN based on clustering approach was presented with good experimental results. But for complex data distribution, traditional clustering approach is not good enough and exerts an adverse influence over the ensemble. In paper [7], a novel selective ensemble algorithm referred to as particle swarm optimization selective ensemble (PSOSEN) was proposed and showed satisfactory results.

Recently, a new method based on KPCA and selective neural network ensemble was developed to deal with fault diagnosis. Specifically, the ensemble method was on the basis of improved binary particle swarm optimization algorithm (IBPSOSEN)^[8]. As for the gait recognition problem, a hierarchical fair competition-based parallel genetic algorithm as well as a neural network ensemble was applied by Lee et al.^[9] and the effectiveness of the proposed method was illuminated. In paper [10], two different populations was taken to construct the algorithm, where one was trained by PSO and the other by Differential Evolution (DE) and experiments demonstrated its superiority.

The Bagging algorithm^[11] employs bootstrap sampling to generate several training sets from the original dataset and then trains each neural network using each training sets to acquire individual neural networks of great diversities to compose an ensemble. AdaBoost^[12] sequentially generates individual neural networks where those training instances wrongly predicted by previous neural networks will appear in the training set of later ones with a greater possibility and endeavors to deal with those instances comparably difficult for previous networks. As is illuminated above, AdaBoost relies on the outcome of previous networks and pays excessive attention to some “difficult” instances, resulting in poor stability and bad performance in some cases.

Based on the analysis of common neural network ensemble approaches mentioned above, a neural network selection approach of constructing ensemble named CSEN in the combination of Cuckoo Search (CS) Algorithm is proposed.

2 CSEN

A novel meta-heuristic algorithm called Cuckoo Search (CS) was formulated by Yang and Deb[13] based on interesting brood parasitism of certain species of cuckoos. This algorithm can explore search space more efficiently with fewer parameters, being easy to implement as well. Research findings indicate that CS is superior to GA, Particle Swarm Optimization (PSO) and Artificial Bees Colony (ABC) with higher efficiency and better optimization effects.

Three following idealized rules are assumed for CS[14]:

- [1] Each cuckoo lays only one egg at a time and dumps it in randomly chosen nest;
- [2] The best nests with high quality of eggs will carry over to next generations;
- [3] The sum of host nests is constant and the cuckoo egg is discovered by the host bird with a probability of $p_a \in [0,1]$.

Basic steps of CS^[13] are shown as the pseudo code in FIG. 1:

Begin

Object function $f(\mathbf{x})$, $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$

Generate initial population of n host nests x_i ($i = 1, 2, \dots, n$)

while (not meet the stop criterion)

 Get a cuckoo randomly by Levy flights

 evaluate the corresponding quality F_i

 Choose a nest among n (say j) at random

 if $F_i > F_j$

 replace j by the new solution

 end

 Abandon worse nests and build new ones with a probability of p_a

 Keep the best solutions

end while

End

FIG. 1 PSEUDO CODE of CUCKOO SEARCH(CS)

The new solution is generated based on equation(1):

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \alpha \oplus \text{Levy}(\lambda), i = 1, 2, \dots, n \quad (1)$$

Where \mathbf{x}_i^t denotes the location of the i th nest in the t th generation, \oplus means entrywise multiplications and α is adjusting step size. The random step length is drawn from a Levy distribution as in equation(2).

$$Levy \sim u = t^{-\lambda} (1 < \lambda \leq 3) \quad (2)$$

Combined with CS, a neural network selection algorithm of constructing ensemble named CSEN is proposed to select individual neural networks for an ensemble. To be more specific, CSEN is built on the basis of CS algorithm and picks up appropriate individual ones from current neural networks available. The selection criterion is the ability of each neural network accurately judging on the instances. The main difficulty of selective ensemble is how to select a subset of neural networks from N total ones, already proven to be a NP-hard problem according to theoretical analysis. Nevertheless, it is essentially equivalent to an optimization problem. Besides, it is worthwhile to point out that CS behaves quite well in optimization problems as in [15][16][17][18], and herein, CSEN originated from CS is further studied.

Main procedures of this approach are summarized in FIG. 2:

Input: training set $TrainSet$, individual neural networks $Net_i (1 \leq i \leq n)$, the number of total neural networks n , threshold λ

Procedure:

for $i = 1$ to n
 train corresponding neural network Net_i utilizing $TrainSet$
 generate an initial row vector $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ of n dimensions
 obtain the optimal row vector by evolving the weight employing CS
for $i = 1$ to n
 if ω_i is greater than λ
 select the corresponding individual neural network Net_i

Output: a set of selected individual neural networks $\{Net_j\} (1 \leq j \leq n)$

FIG. 2 The CSEN APPROACH

In CSEN, a series of individual neural networks are firstly trained and then a row vector is assigned, each component indicating a random weight of each network. CS algorithm is later on applied to evolve the weights, characterizing the fitness of each individual neural network in building the ensemble. Eventually, neural networks meeting requirements are selected to join in the ensemble based on evolved weights. The main feature of this proposed approach is the way it picks up component networks using meta-heuristic algorithm, which is quite distinguished from current other algorithms.

3 EXPERIMENTAL RESULTS AND ANALYSIS

The datasets used in our experiments from UCI datasets[19] are shown in TABLE 1 and the performance of CSEN on real datasets is mainly investigated with its results analyzed.

TABLE 1 DATASETS for EXPERIMENTS

No.	Name	Features	Class	Samples
1	iris	4	3	150
2	wine	13	3	178
3	cancer	9	2	699
4	glass	9	2	214
5	simpleclass	2	4	1000
6	thyroid	21	3	7200
7	crab	6	2	200
8	ovarian	100	2	216

In our experiments, CSEN is applied respectively in datasets above, and for different algorithms, time-consumption to build the model is recorded in TABLE 2. Experiments were conducted repeatedly and misclassification rate (defined as equation(3)) calculated with the mean value of experimental results listed in TABLE 3. To make the

illustration of comparison more convenient, here relative performance is introduced and taken into consideration. Setting the misclassification rate of BP as the baseline, the relative error of other algorithms is defined as the ratio against the error of BP. FIG. 3 demonstrates the bar graph of relative error among Bagging, AdaBoost, GASEN and CSEN.

$$err = \frac{\text{the number of misclassified instances}}{\text{the total number of instances to be classified}} \quad (3)$$

TABLE 2 MODEL BUILDING TIME of DIFFERENT ALGORITHMS

No.	Name	BP	Bagging	AdaBoost	GASEN	CSEN
1	iris	1.9156	8.1688	6.6375	9.6219	9.2133
2	wine	1.9844	11.3142	7.6711	12.5828	11.5436
3	cancer	2.4922	10.7281	7.5703	15.0203	14.6947
4	glass	2.5312	10.4461	7.2336	11.9531	11.5746
5	simpleclass	3.6125	25.9422	9.6922	30.4190	30.0172
6	thyroid	95.5188	1422.0891	120.6172	1438.1203	1418.2302
7	crab	2.1094	9.0110	7.1311	10.2750	10.0295
8	ovarian	3.1828	54.4578	20.1047	50.7610	37.7444

TABLE 3 MISCLASSIFICATION RATE of DIFFERENT ALGORITHMS

No.	Name	BP	Bagging	AdaBoost	GASEN	CSEN
1	iris	0.0600	0.0300	0.0333	0.0367	0.0283
2	wine	0.1052	0.0807	0.0626	0.1043	0.0786
3	cancer	0.0552	0.0348	0.0354	0.0334	0.0313
4	glass	0.0606	0.0455	0.0500	0.0534	0.0466
5	simpleclass	0.0040	0.0013	0	0.0042	0
6	thyroid	0.0252	0.0218	0.0674	0.0188	0.0187
7	crab	0.0567	0.0600	0.0370	0.0500	0.0367
8	ovarian	0.0500	0.0406	0.4438	0.1031	0.0348

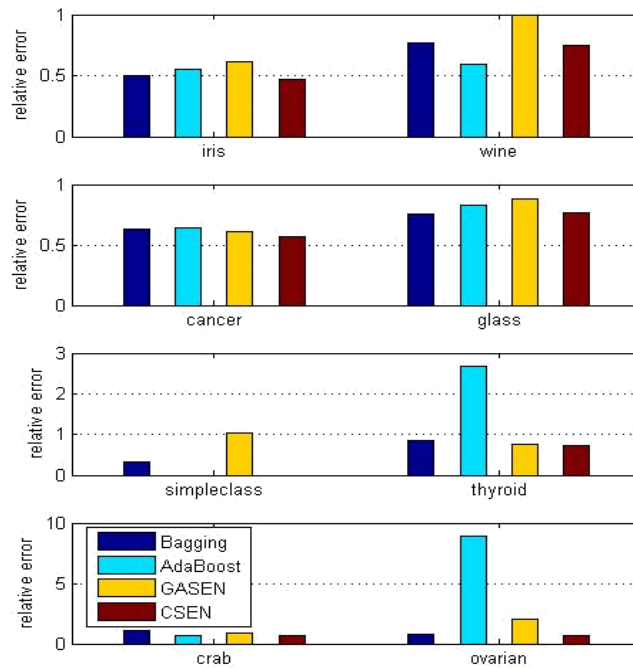


FIG. 3 COMPARISON OF RELATIVE ERROR AMONG DIFFERENT ALGORITHMS

As can be seen from TABLE 2, the model building time of CSEN is consistently shorter than that of GASEN, while Bagging and AdaBoost save much more time spent searching for better neural networks by exploiting different schemes without selecting individual networks. From TABLE 3 and FIG. 3, the misclassification rate of CSEN decreases more or less in 8 datasets, especially dropping sharply in ovarian and decreasing much in wine and crab. In contrast with Bagging, CSEN behaves better with lower misclassification rate in numerous datasets except in glass, where there is no great difference between them. As for AdaBoost, the performance appears to be unstable, for example, the misclassification rate in ovarian is far higher than any other algorithms. The probable reason is that AdaBoost is not valid for this dataset, bringing on high final misclassification rate.

4 CONCLUSION

In this paper, a neural network selection approach of constructing ensemble called CSEN is proposed. In the proposed algorithm, random initial values are assigned to candidate component neural networks and the weight vector optimized by CS algorithm, enabling it to indicate the performance of component networks to a certain extent. The result of CSEN provides us with the ones meeting the requirements to construct an ensemble on the basis of predetermined threshold. Moreover, extensive experiments were conducted on several datasets to test the performance of the approach. Results show that the proposed approach works stably with lower misclassification rate and higher running efficiency compared with other algorithms, thus featuring certain superiority over its counterparts.

Much work can be done in the upcoming future. Firstly, influence of predetermined threshold on the final result remains to be investigated and examined so that appropriate values can be set differently in various situations. Secondly, it can be of help to introduce some other modifications to current approach CSEN and explore its potential in terms of performance. Another direction is to take advantage of proposed algorithm in the identification of network traffic, so as to improve the accuracy of network traffic identification.

ACKNOWLEDGMENT

This work is supported by the science and technology project of State Grid Corporation of China (SGCC) in 2014: Research and application of key technologies for traffic prediction and smart pipeline in power communication network.

REFERENCES

- [1] Hansen L K, Salamon P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12 (10): 993-1001
- [2] Chiu C Y, Verma B, and Li M. Impact of variability in data on accuracy and diversity of neural network based ensemble classifiers[C]//Neural Networks (IJCNN), The 2013 International Joint Conference on. IEEE, 2013: 1-5
- [3] Shao M, Zhu X J, Cao H F, and Shen H F. An artificial neural network ensemble method for fault diagnosis of proton exchange membrane fuel cell system[J]. *Energy*, 2014, 67: 268-275
- [4] Qi L, Yu H, and Chen P. Selective ensemble - mean technique for tropical cyclone track forecast by using ensemble prediction systems[J]. *Quarterly Journal of the Royal Meteorological Society*, 2014, 140(680): 805-813
- [5] Zhou Z H, Wu J, and Tang W. Ensembling neural networks: many could be better than all[J]. *Artificial intelligence*, 2002, 137(1): 239-263
- [6] Li G Z, Yang J, Kong A SH, and Chen N Y. Clustering algorithm based selective ensemble[J]. *Journal of Fudan University(Natural Science)*, 2004, 05: 689-691+695
- [7] Liu Y, He B, Dong D, et al. Robust OS-ELM with a novel selective ensemble based on particle swarm optimization[J]. *arXiv preprint arXiv: 1408.2890*, 2014
- [8] Pang Y Y, Zhu H P, and Liu F M. Fault Diagnosis Method Based on KPCA and Selective Neural Network Ensemble[J]. *Advanced Materials Research*, 2014, 915: 1272-1276
- [9] Lee H, Lee H, and Kim E. A new gait recognition system based on hierarchical fair competition-based parallel genetic algorithm and selective neural network ensemble[J]. *International Journal of Control, Automation and Systems*, 2014, 12(1): 202-207

- [10] Zhao Z S, Feng X, Wei F, et al. Optimized neural network ensemble by combination of particle swarm optimization and differential evolution[M]//Advances in Neural Networks–ISNN 2013. Springer Berlin Heidelberg, 2013: 367-374
- [11] Breiman L. Bagging predictors[J]. Machine learning, 1996, 24(2): 123-140
- [12] Schapire R E. The strength of weak learnability[J]. Machine learning, 1990, 5(2): 197-227
- [13] Yang X, Deb S. Cuckoo Search via Levy Flights[C].World Congress on Nature & Biologically Inspired Computing (NaBIC 2009). IEEE Publications, USA, 2009: 210-214
- [14] Yang X S, Deb S. Cuckoo search: recent advances and applications[J]. Neural Computing and Applications, 2014, 24(1): 169-174
- [15] Gandomi A H, Yang X S, and Alavi A H. Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems[J]. Engineering with Computers, 2013, 29(1): 17-35
- [16] Yildiz A R. Cuckoo search algorithm for the selection of optimal machining parameters in milling operations[J]. The International Journal of Advanced Manufacturing Technology, 2013, 64(1-4): 55-61
- [17] Burnwal S, Deb S. Scheduling optimization of flexible manufacturing system using cuckoo search-based approach[J]. The International Journal of Advanced Manufacturing Technology, 2013, 64(5-8): 951-959
- [18] Moravej Z, Akhlaghi A. A novel approach based on cuckoo search for DG allocation in distribution network[J]. International Journal of Electrical Power & Energy Systems, 2013, 44(1): 672-679
- [19] Asuncion A, Newman D J. UCI Machine Learning Repository [EB/OL], Irvine: School of Inf and Comp Sci, Univ of California, <http://www.ics.uci.edu/~mllearn/MLRepository.html>, 2010

AUTHORS

¹**Jun Hua** was born in Wuxi, Jiangsu Province, China in May 1990. He is primarily engaged in the study of traffic identification based on neural network.

²**Tiecheng Song**, Professor, Ph.D, was born in Zhangjiagang City, Jiang su Province, China in 1967.His area of research includes wireless communication and mobile communication.