Signal Sorting Using Teaching-Learning-Based Optimization and Random Forest

Zhang Mengmeng
School of IoT Engineering
Jiangnan University
Wuxi, China
E-mail: mengzhangm@qq.com

Liu Yian
School of IoT Engineering
Jiangnan University
Wuxi, China
E-mail: Lya_wx@jiangnan.edu.cn

***Abstract*—In this paper, based on the teaching-learning-based optimization (TLBO) algorithm, a new TCTLBO algorithm is formed by introducing twice chaotic searches, and then the TCTLBO algorithm is used to optimize the weighted random forest (WRF) algorithm, and the TCTLBO-WRF sorting model is established to improve the accuracy of radar signal sorting under low signal-to-noise ratio (SNR) conditions. In order to verify the validity of the algorithm, it is compared with commonly used sorting algorithms. Simulation experiments show that the proposed model can effectively improve the radar signal sorting accuracy under low SNR conditions.**

Keywords-radar signal sorting; teaching-learning-based optimization(TLBO) algorithm; chaotic search; random forest(RF);

I. INTRODUCTION

With the wide application of electronic information technology in the military field, electronic countermeasures have gradually switched from traditional mechanization to information technology, and information superiority has become the main competitive advantage. Signal sorting, as an important component of modern electronic warfare has become the focus of research on electronic countermeasures. However, as the battlefield environment continues to be complex and diversified, it is difficult to achieve favorable results by traditional sorting technology which relies on radar external parameters to match signals.

In order to improve the accuracy of radar signal sorting, through analyzing the internal effective features of radar signal [1], studying the corresponding sorting algorithms [2~3], scholars found that selecting radar features with strong stability and separability, or choosing the sorting algorithm with better performance can effectively improve the signal sorting efficiency. Random forest [4], as an integrated machine learning algorithm, can improve the problem that the accuracy of the single-model classification method is not high and is easy to overfit. However, not all decision trees in random forests have a positive effect on the accuracy of the classifier. How to select a part of the decision trees in random forests to form a better integrated classifier with lower complexity becomes the key issue. In order to solve this issue, this paper proposes an improved teaching-learning-based optimization (TCTLBO) algorithm to conduct an evolutionary search on decision trees in random forests. Considering that the classification ability of each decision tree in random forest is different, given corresponding weight according to the classification ability of each decision tree to form a weighted random forest (WRF). Then a classification model based on TCTLBO optimization WRF (TCTLBO-WRF) is constructed, and the optimal decision tree combination is selected by using this mode, then using this optimal decision tree combination to improve the accuracy of radar signal sorting.

II. THEORETICAL KNOWLEDGE

A. Teaching-Learning-Based Optimization Algorithm

The TLBO algorithm [5] is group intelligent optimization algorithm, which seeks the optimal solution by simulating the process of teacher "Teaching" and students "Learning" in the class. In the TLBO algorithm, the class is the population in the search space, the teacher and the student are individuals in the population, and the score is the fitness calculated by the fitness function. Specifically described as follows:

1) Teaching: The stage of "teaching" is that teacher impart knowledge to students and improve students' achievement through teaching. The teacher is the best student in the class, Students are updated according to the class average and the differences between teachers.

a) Calculate the difference between the average score of class and the teacher:

$$difference = r\left(X_{teacher} - T_f X_{mean}\right) \qquad (1)$$

$r = \text{rand}(0,1)$ is the learning step, $X_{teacher}$ is the teacher, $T_f = \text{round}\left(1+\text{rand}\left(0,1\right)\right)$ is a teaching factor, X_{mean} is the class average score.

b) Teaching through difference:

$$X_{new}\left(i\right) = X\left(i\right) + difference \qquad (2)$$

$X_{new}\left(i\right)$ is a new individual after individual $X\left(i\right)$ uses differential learning.

c) Individual substitution:

If $f\left(X_{new}\left(i\right)\right) > f\left(X\left(i\right)\right)$, then $X\left(i\right)=X_{new}\left(i\right)$

2) Learning: The "learning" phase means that after the teacher completes teaching, each student can use the differences between himself and another randomly selected student to learn from each other.

a) Learning through difference between students and students:

$$X_{new}\left(i\right) = \begin{cases} X\left(i\right)+r\left(i\right)\left(X\left(i\right)-X\left(j\right)\right), f\left(X\left(i\right)\right) > f\left(X\left(j\right)\right) \\ X\left(i\right)+r\left(i\right)\left(X\left(j\right)-X\left(i\right)\right), f\left(X\left(j\right)\right) > f\left(X\left(i\right)\right) \end{cases} \qquad (3)$$

b) Individual substitution:

978-1-5386-7445-1/18/$31.00 ©2018 IEEE
DOI 10.1109/DCABES.2018.00073

If $f\left(X_{new}\left(i\right)\right) > f\left(X\left(i\right)\right)$, then $X\left(i\right)=X_{new}\left(i\right)$

B. Random Forest Algorithm

Random forest is an algorithm that composed of a number of randomly generated decision trees. The basic principle is to generate multiple new training sample sets through bagging technology. Each training sample set constructs C4.5 classification decision tree [6] as the basic component unit by repeated dichotomized data. In each tree split, calculates the information gain-ratio of the randomly selected feature set and selects the feature with the largest information gain-ratio as the node for splitting.

III. REALIZATION OF THE TCTLBO AND WRF

A. TCTLBO Algorithm

TLBO algorithm has the characteristics of less parameter setting and strong convergence ability. However, in the iterative process, both the "Teaching" stage and the "Learning" stage retain the better individuals to form a new population for the next iteration, losing the diversity of the population, and making the algorithm easily fall into a local optimum. In order to solve this problem, this paper makes use of the randomness and traversal of chaos search [7], carries out chaos optimization searches for the initial population of TLBO algorithm and the individual iteration updated population, which balances the global and local search ability of the TLBO algorithm. The improvements are described as follows.

1) Optimizing initial population: The introduction of chaotic sequences in population initialization can improve the quality of the initial population.

a) Randomly initialize a sequence y as the initial value of chaotic Logistic, y satisfied conditions: $y\left(i\right)\in\left(0,1\right)$, $y\left(i\right)\neq 0.25, 0.5, 0.7$

b) Obtaining chaotic sequences y_L based on the form of Logistic mapping:

$$y_L = \mu y\left(1-y\right) \qquad (4)$$

μ is a control parameter, When μ = 4 in full chaos.

c) Mapping the chaotic sequence to the range of [a, b] by mapping formula, get the individual x :

$$x = y_L\left(b-a\right)+a \qquad (5)$$

d) Iterative NP times, generate initial population.

2) Optimizing population in individual iterations: After each iteration is completed,choose the best individual x_{best}, *make* $z=x_{best}$

a) Randomly initialize a sequence h , perform a search in the optimal individual nearby according to equation (6), then chaotic disturbed of search points according to equation (7).

$$z = z+\beta h \qquad (6)$$

$$z = \mu z\left(1-z\right) \qquad (7)$$

μ is a control parameter, When μ = 4 in full chaos.

b) Individual substitution: Mapping the chaotic sequence y to the range of [a, b] by mapping formula, get the individual cx, Calculate the fitness, if $f\left(cx\right) > f\left(z\right)$, renew the individual $z = cx$.

c) Population regeneration: Return z at the end of chaotic search, calculated the fitness, if $f\left(z\right) > f\left(x_{best}\right)$, replace x_{best} with z , and update the population.

B. Performance Analysis of TCTLBO Algorithm

In order to verify the optimization performance of TCTLBO algorithm, four different test functions were selected to test the algorithm and compared with DE algorithm, PSO algorithm, TLBO algorithm. The test function is shown in table 1. In table 1, f1 is a unimodal function and f2~f4 are multimodal functions.

To ensure fairness, the parameters of the four algorithms are set as follows: initial population size Np=50, variable dimension Dim=10, maximum number of iterations Nmax=200, number of runs Nrun=50. In DE algorithm, crossover probability CR=0.9, mutation operator is set to adaptive mutation. In PSO algorithm, the learning factor is C1=C2=1.49. The experimental results are shown in table 2 (in table 2, F: benchmark function)

From Table 2, we can see that in addition to the function f3, the optimal value of TCTLBO is equal to the TLBO (the Worst and standard deviations of TCTLBO are better than those of TLBO.). On the other test functions, the performance of TCTLBO algorithm is better than that of DE algorithm, PSO algorithm and TLBO algorithm. TCTLBO algorithm can achieve better search results than others, and the good performance of TCTLBO algorithm laid a foundation for the signal sorting model based on TCTLBO-WRF.

TABLE I. BENCHMARK FUNCTIONS

Number	Function name	Optimal value	Search space
f1	Sphere	0	(-100,100)
f2	Schwefel 2.22	0	(-10,10)
f3	Rastrigin	0	(-5.12,5.12)
f4	Quartic	0	(-1.28,1.28)

TABLE II. PERFORMANCE COMPARISON OF 4 ALGORITHMS

F	Algorithm	Best	Worst	Std
f1	DE	5.5510	35.0317	8.4681
f1	PSO	0.0031	0.0833	0.0338
f1	TLBO	2.6897×10^{-18}	1.8830×10^{-17}	1.8830×10^{-17}
f1	TCTLBO	1.1245×10^{-19}	9.8343×10^{-19}	1.0241×10^{-19}
f2	DE	142.7199	9.6579×10^{3}	2.35212×10^{3}
f2	PSO	1.4070×10^{-14}	1.6372×10^{-11}	6.2429×10^{-12}
f2	TLBO	1.3034×10^{-36}	1.3955×10^{-33}	2.3312×10^{-34}
f2	TCTLBO	3.3233×10^{-41}	3.9183×10^{-38}	8.1499×10^{-39}
f3	DE	35.119	70.478	8.3685
f3	PSO	3.9798	14.929	2.9737
f3	TLBO	0	23.181	5.9611
f3	TCTLBO	0	14.2889	2.7025
f4	DE	0.0447	3.1122	0.9037
f4	PSO	0.2806	0.9815	0.2380
f4	TLBO	2.9751×10^{-4}	0.0038	7.0816×10^{-4}
f4	TCTLBO	1.0758×10^{-4}	0.0016	4.6225×10^{-4}

C. WRF Algorithm

In random forest, each decision tree is independent of each other; different decision trees have different classification effects. However, when voting, the decision trees with different classification effects have the same weight, which results in some decision trees with poor classification effect easily cast the wrong votes, which affects the overall performance of the classifier. In order to improve the overall classification effect of the classifier, this paper sets the corresponding weight according to the different classification ability, and counts the voting result of each weighted decision tree as the classification result. The specific improvements are as follows:

1) Calculating the weight of decision tree:

$$w_t = \frac{\sum (W \times X_{\text{corr}})}{\sum W} \tag{8}$$

w_t is the weight of the t decision tree. W is the weight of each sample taken. X_{corr} is the classification result of the t decision tree.

2) New voting mechanism:

$$f_{RF}(S) = \underset{i=1,2,3,\ldots,c}{\arg\max} \left\{ I\left(\left(f_t^{tree}(S) == i \right) + w_t \right) \right\} \tag{9}$$

$f_{RF}(S)$ is the result of vote of sample S. c is the number of categories in the sample. $f_t^{tree}(S)$ is the output of the sample S in the t decision tree. i is a class in class c. $I(\bullet)$ is the weight of each category.

IV. Signal Sorting Based on TCTLBO-WRF

Based on the above analysis, the specific steps of the signal sorting model based on TCTLBO-WRF are given.

STEP 1. Data preparation

In the signal-to-noise ratio (SNR) range of -10dB~ 10dB, the autocorrelation operations of 7 commonly used radar signals are performed, and Shannon entropy, norm entropy [8], singular value entropy [9] of signal autocorrelation function [10] are extracted as radar data sample sets. Every 2 dB intervals, 100 samples per signal are generated, total of 7700 samples. 330 samples randomly sampled in the data sample set at each SNR as training samples. 300 samples randomly sampled in the data sample set at each SNR as test samples. Radar signal parameter settings are shown in table 3.

In table 3, FS: frequency of sampling. FC: carrier frequency. PW: pulse width. TW: tape width.

TABLE III. Radar Parameter Setting

Signal Type	FS (MHZ)	FC (MHZ)	Encoded Mode	PW(μs)	TW (MHZ)
LFM	60	30	-	10	20
NLFM	60	30	-	10	20
LFMCW	60	30	-	10	-
BPSK	60	-	13 Barker code	13	-
FSK	60	-	13 Barker code	10	-
MPSK	60	-	Frank code	10	-
QPSK	60	-	Frank code	10	-

STEP 2. Calculate the weight of decision tree

According to the equation (10), using the training sample set to generate decision trees are combined into a random forest, and the training set is used as the input of the random forest to obtain the training set classification results. The weight of each decision tree in the forest is calculated according to equation (8) by using the classification results.

$$GR(S,A) = \frac{G(S,A)}{\sum_{i=1}^{c} \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|}} \tag{10}$$

$GR(S,A)$ is the Information gain-ratio of the A attribute in the sample S. $G(S,A)$ is the information gain of the A attribute. $|S|$ is the number of the sample S.

STEP 3. TCTLBO -WRF algorithm optimizes decision tree combination

1) Chaos initialization: According to the chaotic mapping definition of equation (4) and equation (5), initialize the population, that is, initializing different decision tree combinations.

2) Fitness function: Calculate the classification accuracy of the test sample under different weighted decision tree combinations according to equation (9), and use the classification accuracy as the fitness of the TCTLBO algorithm.

3) "Teaching" and "Learning": Select the individual with the highest degree of fitness in the initial population as the teacher, follow the equation (2) to teach students, and update the individual after the "Teaching" is completed. The students learn from each other according to equation (3), and "Learning" completes update the individual..

4) Chaos search: Using equation (6) and equation (7), chaos disturbance is performed on the updated optimal individual. If the fitness of the chaotic individual is better than that of the optimal individual, the optimal individual is renewed.

STEP 4. Output optimal decision tree combination

Determine whether the program has reached the maximum number of iterations. If it does not, continue to STEP 3, and if so, output the optimal individual.

STEP 5. Signal sorting

Use the selected optimal decision tree combination to sorting test samples and output test sample sorting results.

V. Simulation and Analysis

A. TCTLBO-WRF Performance Analysis

In order to verify the validity of the proposed algorithm, the test set is used as the input of the RF algorithm, WRF algorithm, TCTLBO-WRF algorithm, and SVM algorithm and the output sorting results are shown in Fig.1.

According to Fig.1, we can see that the sorting accuracy of signal increases with the increase of SNR. When the SNR < 0dB, the accuracy of the TCTLBO-WRF algorithm has a large gap with the other algorithms, shows the TCTLBO-WRF algorithm has better performance in the low SNR condition. When the SNR > 0dB, the effect of noise on the signal is reduced, the accuracy of the analysis increases slowly, and the gap between the algorithms decreases, but the accuracy of the TCTLBO-WRF is still better than the other algorithms, and the validity of the TCTLBO-WRF algorithm is verified.

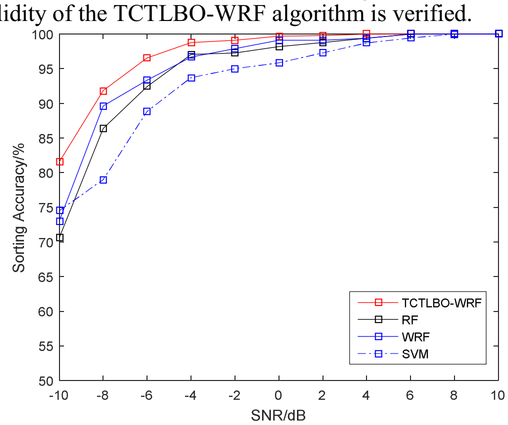

Figure 1. Signal sorting accuracy of 4 algorithms.

B. Radar Signal Sorting

Fig.2 shows the sorting results of each signal.

From Fig.2, we can see that the accuracy of all signals increases with the increase of SNR. Among all the signals, the CW signal and the BPSK signal have the best sorting effect, the accuracy rate of SNR= - 8dB can reach 100%. The NLFM signal sorting effect is the worst, but the rising speed is the fastest, which indicates that the NLFM signal is greatly influenced by SNR. And the sorting effect of other signals is a slight fluctuating, but in SNR > 2dB, the sorting accuracy tends to 100%, which indicates that the TCTLBO-WRF algorithm can effectively sorting the radar signals under the condition of low SNR.

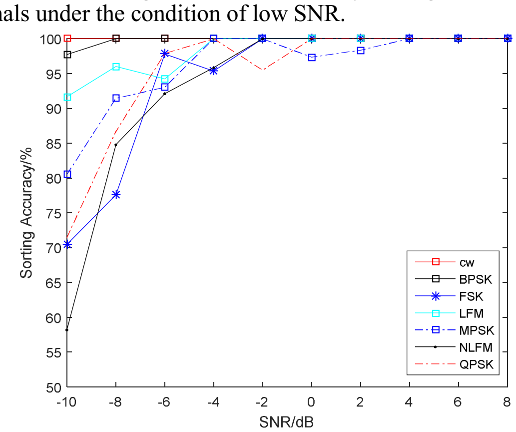

Figure 2. Accuracy of each signal sorting under TCTLBO-WRF algorithm.

VI. Conclusion

With the increasingly complex battlefield environment and the widespread use of new radars, how to accurately classify radar signals becomes a key technology for electronic warfare. Aiming at the problem that the accuracy of radar signal sorting is not high under the condition of low signal-to-noise ratio (SNR). Proposed to use improved TLBO algorithms and weighted random forests, establishes a sorting model that based on the improved TLBO algorithm to optimize weighted random forest. In the SNR range of -10dB to 10dB, the three-dimensional entropy of seven common radar signals after auto-correlation operation is extracted as a radar data sample set. The experimental results show that, under the same conditions, the accuracy of radar signal sorting by TCTLBO-WRF algorithm is better than two other commonly used signal sorting algorithms.

References

[1] Wang. W, Hua. W. U, and Wang. J and Zhan. Q, "Subtle intrapulse feature extraction based on CEEMDAN for radar signals," Journal of Beijing University of Aeronautics & Astronautics,vol.42,no.10,pp.2532-2539, 2016.

[2] Zhang. W, Sun. F, and Wang. B, “Radar signal intra-pulse feature extraction based on improved wavelet transform algorithm,” International Journal of Communications Network & System Sciences, vol.8, no.10, pp.118-127, 2017.

[3] Zhao. G. .X, Sun. Y, and Wang .K, “Radar Signal Sorting Algorithm Based on Artificial Bee Colony,” Electronic Information Warfare Technology, vol .30, no.4, pp.4-7, 2015.

[4] L. Breiman, "Random forest," Machine Learning,vol.45, pp.5-32, 2001.

[5] Rao. R. V, Savsani. V. J, and Vakharia. D. P, “Teaching–Learning-Based Optimization: An optimization method for continuous non-linear large scale problems,” Information Sciences, vol.183, no.1, pp.1-15, 2012.

[6] Xu. P and Lin. S, “Internet Traffic Classification Using C4.5 Decision Tree,” Journal of Software, vol.20, no.10, pp.2692-2704, 2009.

[7] Wang. Hai and Hu. Jiandong. “Logistic-Map Chaotic Spread-Spectrum Sequence,”.Acta Electronica Sinica, 1997.

[8] Zhang. G. X, “Radar emitter signal recognition based on entropy features,” Chinese Journal of Radio Science, vol.4, pp.30-35, 2005.

[9] Li. J and Ying. Y, “Radar signal recognition algorithm based on entropy theory,” International Conference on Systems and Informatics IEEE, pp.718-723, January. 2015, doi: 10.1109/ICSAI.2014.7009379.

[10] Chen. T, Li. C, and Sha. L, “Feature extraction using autocorrelation function for radar emitter signals,” Cross Strait Quad-Regional Radio Science and Wireless Technology Conference. IEEE, pp.1371-1374, October. 2011,doi: 10.1109/CSQRWC.2011.6037219

.