

Received April 7, 2020, accepted April 20, 2020, date of publication May 8, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993270

Research and Experiment of Radar Signal Support Vector Clustering Sorting Based on Feature Extraction and Feature Selection

SHIQIANG WANG¹, CAIYUN GAO², QIN ZHANG¹, VEERENDRA DAKULAGI³,
HUIYONG ZENG¹, GUIMEI ZHENG¹, JUAN BAI¹, YUWEI SONG¹,
JILIANG CAI¹, AND BINFENG ZONG¹

¹Air and Missile-Defence College, Air Force Engineering University, Xi'an 710051, China

²Department of Basic Sciences, Air Force Engineering University, Xi'an 710051, China

³Department of Electronics and Communication Engineering, Lincoln University College, Petaling Jaya 47301, Malaysia

Corresponding author: Huiyong Zeng (1048768838@qq.com)

This work was supported by in part by the National Natural Science Foundation of China under Grant 61601499, Grant 61701527, Grant 61601503, and Grant 61971438, in part by the Youth Talent Project under Grant 20180109, and in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JM-345, Grant 2020JQ-482, and Grant 2019JM-155.

ABSTRACT The result of radar signal sorting directly affects the performance of electronic reconnaissance equipment. Sorting method based on intra-pulse features has become a research focus in recent years. However, as the number of extracted features increases, the dimension of the feature vector becomes higher and higher. And too many dimensional feature vectors would make the complexity of the sorting algorithm increase geometrically. In this way, feature selection becomes more and more necessary. Combining the latest research on fuzzy rough sets, this paper proposes two feature selection methods, namely two-steps attribute reduction based on fuzzy dependency (TARFD) algorithm and fuzzy rough artificial bee colony (FRABC) algorithm. The TARFD method uses the candidate attribute set as starting point, according to the definition of the redundant attribute set. Then the less important attributes are successively eliminated. The FRABC method starts from the dependence degree of fuzzy rough set, and constructs a fitness function that reflects the importance of the attribute subset and the reduction rate. Based on this function, the artificial bee colony algorithm is used to reduce the attributes of the dataset. Using the TARFD and FRABC algorithms, the extracted feature sets, including entropy feature set, Zernike moment feature set, pseudo Zernike feature set, gray level co-occurrence matrix (GLCM) feature set, and Hu-invariant moment feature set are processed, then an optimal feature subset was obtained and a sorting test was performed. The results show the effectiveness of the extracted intra-pulse features and the efficiency of the feature selection algorithm.

INDEX TERMS Feature extraction, feature selection, feature set, support vector machine, support vector clustering.

I. INTRODUCTION

At present, with the widespread application of complex radars, the modern electronic countermeasure environment is becoming more and more complex. The signal density of the electromagnetic threat environment has reached millions of orders [1], which mean that the pulse streams from the same direction are also densely complex. In this case, the performance is greatly depressed when only the RF and PW

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman¹.

parameters are used for sorting. In view of this problem, it is necessary to extract new features for characterizing the intra-pulse data of the radiation source signal, that is, to augment the conventional parameter set, and to describe the intra-pulse modulation difference of the radar signal with new features to solve the sorting problem caused by complex radar system and the diverse waveform. Radar signal sorting is difficult with conventional parameters because these parameters are usually from the single direction and can be changed in each parameter domain, and even overlap. In recent years, many new intra-pulse feature extraction methods have

been proposed, such as wavelet analysis [2], [3], atomic decomposition [4], time-frequency analysis [5]–[8], and high-order statistical analysis methods [9]–[11], for example, extracts entropy feature set [11], Zernike moment feature set [12], pseudo Zernike feature set [13], [14], gray level co-occurrence matrix (GLCM) feature set [15] and so on.

However, not all the extracted intra-pulse features contain necessary information for classifying and sorting radar signals. Too many features may not only depress the sorting effect, but also may reduce the sorting efficiency. Therefore, in view of changing rule, distribution form, anti-noise performance and other aspects of the effectiveness of each feature, we need to study the feature selection methods, that is, select a set of reliable feature subsets to reduce the intra-pulse feature dimension effectively and maximize the classification effect meanwhile [16]. In recent years, the feature selection (attribute reduction) based on fuzzy rough set (FRS) has received much attention, and has achieved many research results [17]–[20]. These methods mainly focus on the use of distinguishing matrix, dependency and reduction methods with optimization algorithms.

In this paper, the theoretical and experimental analysis of the above problems is carried out, that is, the sorting method based on intra-pulse features is studied, and the minimum intra-pulse feature subset which could better characterize the radar signal is obtained, and the intra-pulse feature subset is utilized. The sorting experiment verifies the effectiveness of the extracted features and feature selection methods with the proposed sorting method of modification cone mapping support vector clustering (MCMSVC) with similitude entropy (SE) index (called SE-MSVC) in [21].

II. FEATURE EXTRACTION AND SIGNAL SORTING BASED ON FEATURE SET

Rough set theory is a mathematical tool capable of handling inaccurate and incomplete information. Let $I = (U, A \cup Q)$ be an information system, where U represents a non-empty finite set of objects, namely the domain. And A represents a non-empty finite set of conditional attributes. Unless otherwise specified, attributes refer to conditional attributes. In the information system, Q represents a non-empty finite set of decision attributes, that is, a category attribute set. The attribute set $P \subseteq A$ corresponds to an indistinguishable equivalent relationship. Without ambiguity, this equivalent relationship is abbreviated as P , and the following research is carried out on the premise of this information system.

A. FEATURE SELECTION METHOD

1) FUZZY DEPENDENCY ANALYSIS

In the domain U , objects x may belong to more than one fuzzy equivalent class determined by fuzzy attributes. Therefore, the fuzzy equivalent class determined by multiple fuzzy attributes, that is, the Cartesian product of fuzzy equivalent classes determined by a single fuzzy attribute.

For the distinct rough sets, U/P determines a cluster set consisting of different sets composed of all indistinguishable objects. Supposing that the attribute set P contains m attributes $\{a_1, a_2, \dots, a_m\}$, for the sets B and C , the operator \otimes is defined as:

$$B \otimes C = \{X \cap Y : \forall X \in B, \forall Y \in C, X \cap Y \neq \phi\} \quad (1)$$

Then U/P can be calculated by:

$$\begin{aligned} U/P &= \otimes \{a_i \in P : U/IND(a_i), i = 1, 2, \dots, m\} \\ &= U/\{a_1\} \otimes U/\{a_2\} \otimes \dots \otimes U/\{a_m\} \end{aligned} \quad (2)$$

Consider the situation at $m = 2$:

$$\begin{aligned} U/P &= U/\{a_1\} \otimes U/\{a_2\} \\ &= \{X_1^{a_1}, X_2^{a_1}, \dots\} \otimes \{X_1^{a_2}, X_2^{a_2}, \dots\} \end{aligned} \quad (3)$$

This means:

$$U/P = \{X_1^{a_1} \cap X_1^{a_2}, X_1^{a_1} \cap X_2^{a_2}, \dots, X_2^{a_1} \cap X_1^{a_2}, \dots\} \quad (4)$$

The sets $\{X_1^{a_1}, X_2^{a_1}, \dots\}$ and $\{X_1^{a_2}, X_2^{a_2}, \dots\}$ represent the divisions of U determined by the attributes $\{a_1\}$ and $\{a_2\}$ respectively. In the distinct rough set, distinct set is $X_j^{a_i} = [x_j]_{\{a_i\}}$. The divisions of U determined by P are all distinct sets $U/P = \{X_1, X_2, \dots, X_n\}$, where $\forall X_i \in X$, and the membership function is used to represent the relationship with $x \in U$ and X_i , then:

$$\mu_{X_i}(x) = \begin{cases} 1 & x \in X_i \\ 0 & x \notin X_i \end{cases} \quad (5)$$

When the continuous attribute interval is divided according to actual needs, in order to determine the fuzzy equivalent class of each fuzzy attribute, each sub-interval divided by the fuzzy condition attribute corresponds to a fuzzy equivalent class of the attribute.

In the fuzzy rough set, a single attribute $\{a_i\}$ still corresponds to the equivalent relationship of an indistinguishable relationship. It is still briefly described as $\{a_i\}$. The calculation of U/P in the fuzzy rough set is similar to the distinct rough set. The difference is that the sets $\{X_1^{a_1}, X_2^{a_1}, \dots\}$ and $\{X_1^{a_2}, X_2^{a_2}, \dots\}$ represent a fuzzy division of U determined by attributes $\{a_1\}$ and $\{a_2\}$ respectively.

The symbol $X_j^{a_i}$ represents the j th interval obtained by segmenting attribute a_i by some attribute interval segmentation algorithm. All objects in this interval are indistinguishable. To distinguish them from distinct rough sets, let $F_j^{a_i} = X_j^{a_i}$.

Let the fuzzy division of U determined by P is $U/P = \{F_1, F_2, \dots, F_n\}$, since $\forall F_i \in U/P$ are all fuzzy sets, the degree to which object x_k belongs to F_i can be obtained by calculating the degree to which object x_k belongs to the intersection of all fuzzy sets forming F_i .

Considering the situation at $m = 2$ as well, let's set $F_i = F_r^{a_g} \cap F_t^{a_h}$, where $g, h, r, t = 1$ or 2 , and use the membership function to represent the relationship between x and F_i , which is:

$$\mu_{F_i}(x) = \mu_{F_r^{a_g} \cap F_t^{a_h}}(x) = \min(\mu_{F_r^{a_g}}(x), \mu_{F_t^{a_h}}(x)) \quad (6)$$

If $m > 2$, then the relationship between x and F_i is similar to the membership function obtained at $m = 2$.

The distinct positive field in classical rough set theory is defined as the lower approximation union set. This concept is extended to the fuzzy positive field to obtain the membership degree of the object x with respect to the fuzzy positive field:

$$\mu_{POS_P(Q)}(x) = \sup_{x \in U/Q} \mu_{PX}(x) \tag{7}$$

According to the meaning of membership and the definition of fuzzy positive domain, the dependence of decision attribute Q on conditional attribute set P under fuzzy rough set conditions can be found [17]:

$$k' = \gamma'_P(Q) = |\mu_{POS_P(Q)}(x)| / |U| = \sum_{x \in U} \mu_{POS_P(Q)}(x) / |U| \tag{8}$$

where $|U|$ represents the potential of U , that is, the number of objects in U . Similar to the classic rough set dependency, the above formula represents the proportion of distinguishable objects in the entire data set when only the information in P is used. According to the above formula, the potential of $\mu_{POS_P(Q)}$ is divided by the potential of the universe U . The above formula also indicates the degree to which Q depends on P . P determines the value of Q in proportion k' , that is, P determines the classification effect. The larger the value k' , the better the effect obtained when classifying signals based on attribute P .

The dependency described by formula (8) can be generalized to the dependency between attributes, taking P and Q as a single attribute set, that is, $P = a_i, Q = a_j, \forall a_i, a_j \in A$, then:

$$\begin{aligned} k' &= \gamma'_{a_i, a_j} = \gamma'_{a_i}(a_j) = |\mu_{POS_{a_i, a_j}}(x)| / |U| \\ &= \sum_{x \in U} \mu_{POS_{a_i, a_j}}(x) / |U| \end{aligned} \tag{9}$$

Equation (9) characterizes the degree to which the attribute a_j depends on the attribute a_i . And a_i determines the value of a_j in proportion. The larger the value k' , the greater the ability of a_i to determine the value of a_j .

The following description uses a simple decision table as an example to explain the calculation of the fuzzy dependency. Supposing the original decision table is shown in Table 1, where a_1, a_2 and a_3 denote the conditional attributes, and Q denote the decision attributes. Firstly, the fuzzy lower approximation of a_1, a_2, a_3 are calculated, and then the dependence degrees of each attribute by formula (9) are calculated.

TABLE 1. Decision table with continuous attribute values.

object	a_1	a_2	a_3	Q
1	-0.4	-0.3	-0.5	no
2	-0.4	0.2	-0.1	yes
3	-0.3	-0.4	-0.3	no
4	0.3	-0.3	0	yes
5	0.2	-0.3	0	yes
6	0.2	0	0	no

The attribute values in this table are continuous. Under the fuzzy similarity relationship, it is still required to divide the continuous attribute interval to determine the fuzzy equivalent class of each fuzzy attribute. Each sub-interval divided by the fuzzy condition attribute corresponds to a fuzzy equivalence class of the attribute. The membership function of the equivalence class generally includes a trapezoidal function, a trigonometric function, and a normal distribution function. After the decision table 1 is processed, the interval division results are obtained:

$$\begin{aligned} A' &= \{[-0.5, 0], [0, 0.5]\} = \{F_1^{a_1}, F_2^{a_1}\} \\ B' &= \{[-0.5, 0], [0, 0.5]\} = \{F_1^{a_2}, F_2^{a_2}\} \\ C' &= \{[-0.5, 0], [0, 0.5]\} = \{F_1^{a_3}, F_2^{a_3}\} \end{aligned} \tag{10}$$

It can be seen that in each two-dimensional Cartesian sub-interval determined by the above results, the objects have the same decision value. So, each condition attribute is divided into two fuzzy equivalence classes.

For convenience, the fuzzy subset $F_1^{a_1}$ of a_1 is treated the same as its membership function $\mu_{F_1^{a_1}}$, and the membership degree of element x with respect to $F_1^{a_1}$ is represented by $\mu_{F_1^{a_1}}$. For other fuzzy subsets, they are processed in the same way. Figure 1 shows the membership function corresponding to each fuzzy equivalent class of the attribute determined by the trigonometric function and the trapezoidal function.

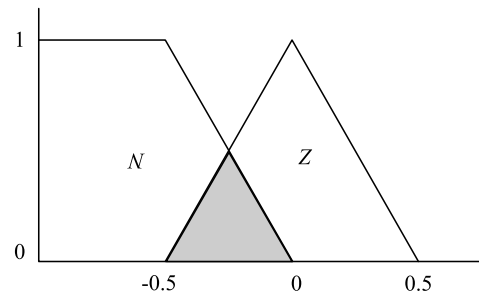


FIGURE 1. Membership function of fuzzy set corresponding to condition attribute.

The membership functions of fuzzy equivalent classes $F_1^{a_1}, F_2^{a_1}$ of attribute a_1 are defined as:

$$\begin{aligned} \mu_{F_1^{a_1}}(x) &= \begin{cases} 1, & x \leq 0.5 \\ \max\{-2x, 0\}, & x > 0.5 \end{cases} \\ \mu_{F_2^{a_1}}(x) &= \begin{cases} \max\{2x + 1, 0\}, & x \leq 0 \\ \max\{-2x + 1, 0\}, & x > 0 \end{cases} \end{aligned} \tag{11}$$

This gives:

$$\begin{aligned} \mu_{F_1^{a_1}}(x) &= \{0.8, 0.8, 0.6, 0, 0, 0\} \\ \mu_{F_2^{a_1}}(x) &= \{0.2, 0.2, 0.4, 0.4, 0.6, 0.6\} \end{aligned} \tag{12}$$

The attribute reduction algorithm based on fuzzy rough set dependency includes two steps. The first one is to determine the candidate attribute set, and the second one is to determine whether there are redundant attributes in the candidate

attribute set. Based on the above analysis, this paper gives two definitions. The attribute reduction based on fuzzy dependency is mainly carried out by these two definitions.

2) REDUCTION METHOD 1

Definition 1: Supposing $a_j \in A$, for a_i , it satisfies: $a_i = \max\{a_j | \gamma'_{a_j}(Q), 1, 2, \dots, m\}$. For a given threshold T_h , when calculating $\gamma'_{a_j}(Q)$ ($j \neq i$), if $\gamma'_{a_j}(Q) > T_h$, then a_j is called as the first candidate attribute. However, if $\gamma'_{a_j}(Q) \leq T_h$ and $\gamma'_{\{a_i, a_j\}}(Q) \leq \gamma'_{a_i}(Q)$, a_j is called a redundant attribute. But if $\gamma'_{a_j}(Q) \leq T_h$ and $\gamma'_{\{a_i, a_j\}}(Q) < \gamma'_{a_i}(Q)$, then a_j is called a second candidate attribute. The set consisting of the first and second candidate attributes is called a candidate attribute set.

Definition 2: Let the determined candidate attribute set be C , supposing $a_i \in C$, for $\forall a_j \in C$ ($i \neq j$), if $\gamma'_{a_i, a_j} = 1, \gamma'_{a_j, a_i} = 0$ or $\gamma'_{a_i, a_j} \approx 1, \gamma'_{a_j, a_i} \approx 0$, then the attribute a_j is called the first redundant attribute, otherwise a_j is called a dependent attribute. For the dependent attribute a_j , if $\gamma'_{a_j}(Q) \leq \gamma'_{a_i}(Q)$ and $\gamma'_{a_j, a_i} > \gamma'_{a_i, a_j}$, then a_j is a non-redundant attribute. And if $\gamma'_{a_j}(Q) \leq \gamma'_{a_i}(Q)$ and $\gamma'_{a_j, a_i} \leq \gamma'_{a_i, a_j}$, then a_j is called the second redundant attribute. The set of the first and second redundant attributes is called a redundant attribute set.

Based on the above analysis, a two-steps attribute reduction based on fuzzy dependency (TARFD) algorithm is proposed. The algorithm starts with the candidate attribute set and deletes the less important attributes one by one according to the redundant attributes obtained in Definition 2 until the termination condition is met. From this we can see that the feature reduction algorithm of TARFD belongs to filter method.

Algorithm TARFD

Step 1 Let $T_h = 0$, and determine the candidate attribute set C by definition 1. For attribute $a_i \in C$, calculate the dependency $\gamma'_{a_i}(Q)$ of decision attribute Q on a_i ;

Step 2 Sort $\{\gamma'_{a_i}(Q) | i = 1, 2, \dots, m, m = |C|\}$ in descending order to get a set of attributes, denoted as T , $T = \{\gamma'_{a_{i1}}(Q) | \gamma'_{a_{i1}}(Q) \geq \gamma'_{a_{i2}}(Q) \dots \geq \gamma'_{a_{i1}}(Q) \dots \geq \gamma'_{a_{im}}(Q)\}$, where $\{t1, t2, \dots, tl, \dots, tm\}$ is one arrangement of i ;

Step 3 Let $k = 1$;

Step 4 Let $j = 1$, if $j \geq m - k$, terminate the algorithm and go to step 7;

Step 5 Calculate the dependency between attributes. If $\gamma'_{a_{ij}}(a_{t(j+k)}) \geq \gamma'_{a_{t(j+k)}}(a_{ij})$, remove the attribute at $a_{t(j+k)}$, that is, $T = T \setminus \{a_{t(j+k)}\}$;

Step 6 Let $j = j + 1$, if $j < m - k$, go to step 5, otherwise let $k = k + 1$, go to step 4;

Step 7 Finally the T is a relative reduction of conditional attribute P relative to Q .

3) COMPLEXITY ANALYSIS OF TARFD

As we known, Jensen proposed a quick reduction algorithm for attribute reduction based on fuzzy rough dependency, and

achieved good reduction results with a small dataset dimension [17]. However, the computational complexity of the algorithm increases sharply with the increase of the condition attribute set. Assuming that the attribute set P contains m attributes, and the partition of U determined by each attribute includes n fuzzy classes, then the partition of U determined by attribute set P includes n^m fuzzy classes. The approximate computational complexity of the dependency of decision attribute set P can be expressed as $O((n^m n)) = O(N(n^{m+1}))$ in the worst case, which means the algorithm does not terminate until the last attribute is calculated.

For example, suppose $P = \{a, b\}$, and the partition determined by each attribute contains 3 fuzzy classes, then the dependency of Q with respect to P needs to calculate 9 fuzzy classes. In fact, it is normal for the number of attribute sets to be greater than 10. Therefore, if we set $|P| = 15$, It is needed to calculate 3^{15} fuzzy classes for the dependency of Q with respect to P . It can be seen that the computational complexity of this algorithm is quite large.

For the TARFD algorithm, it only needs to calculate the single attribute dependency in the attribute set and the dependency of any two attribute combinations. That is, the complexity of the dependency of the attribute set P in the worst case can be expressed as $O(2n^2 C_m^2 + n C_m^1) \approx O(2n^2 C_m^2)$. If $|P| = 15$ is set as mentioned above, then it needs to calculate 1935 fuzzy classes. So, compared with quick reduction algorithm, TARFD algorithm improves the execution efficiency by $3^{15}/1935 = 7415$ times in the worst case, indicating that the algorithm is efficient.

4) ANOTHER CALCULATION OF FUZZY ROUGH APPROXIMATION

Let (U, P) be the fuzzy approximation space, that is, P is the fuzzy equivalence relation on the universe U . Assuming I represents the edge implication operator and T represents the t -module, then (I, T) -fuzzy rough approximation on the fuzzy approximation space (U, P) represents such a mapping:

$$\text{Apr}^{I, T} : F(U) \rightarrow F(U) \times F(U) \quad (13)$$

For $\forall X \in F(U)$, $\text{Apr}^{I, T}(X) = (P_I X, \bar{P}^T X)$ is holded. Where $F(U)$ represents the whole of the fuzzy subset on U .

The fuzzy set $P_I X$ and $\bar{P}^T X$ are respectively called the I - low fuzzy rough approximation and the T - up fuzzy rough approximation for X in the fuzzy approximation space (U, P) , and their membership functions are determined by equations (14) and (15) respectively:

$$\mu_{P_I X}(x) = \inf_{y \in U} I(\mu_P(x, y), \mu_X(y)), \quad \forall x \in U \quad (14)$$

$$\mu_{\bar{P}^T X}(x) = \sup_{y \in U} T(\mu_P(x, y), \mu_X(y)), \quad \forall x \in U \quad (15)$$

For $\forall y \in U$, if $y \in X$, then $\mu_X(y)$ is 1, otherwise it is 0. $\mu_P(x, y)$ indicates the similarity between the objects x and y determined by the fuzzy relationship P . The implication operator and t -module are taken respectively:

$$I(x, y) = \min(1 - x + y, 1) \quad (16)$$

$$T(x, y) = \max(x + y - 1, 0) \quad (17)$$

Based on the concept of fuzzy equivalence relation P , the fuzzy equivalence class of an object can be defined.

Then $\mu_P(x, y)$ can be determined by the iterative formula (18):

$$\mu_P(x, y) = \bigcap_T (\mu_a(x, y)), a \in P \quad (18)$$

For example, if $P = \{a_1, a_2, a_3\}$, then $\mu_P(x, y) = (\mu_{\{a_1, a_2\}} \cap_T \mu_{\{a_3\}})(x, y)$, where

$$\begin{aligned} \mu_{\{a_1, a_2\}}(x, y) &= (\mu_{\{a_1\}} \cap_T \mu_{\{a_2\}})(x, y) \\ &= T(\mu_{\{a_1\}}, \mu_{\{a_2\}}) \end{aligned} \quad (19)$$

For attribute $a \in P$, $\mu_a(x, y)$ could be determined by the fuzzy similarity relationship, which is expressed by formula (20), (21) and (22):

$$\mu_a(x, y) = 1 - \frac{|a(x) - a(y)|}{a_{\max} - a_{\min}} \quad (20)$$

$$\mu_a(x, y) = \exp\left(-\frac{(a(x) - a(y))^2}{2\sigma_a^2}\right) \quad (21)$$

$$\mu_a(x, y) = \max\left(\min\left(\frac{a(x) - a(y) + \sigma_a}{\sigma_a}, \frac{a(y) - a(x) + \sigma_a}{\sigma_a}\right), 0\right) \quad (22)$$

In the formula, $a(x)$ and $a(y)$ represent the values of objects x and y on feature a respectively, and σ_a represents the standard deviation of the values of all objects on feature a . After many experiments, the equation (22) is used to calculate the fuzzy similarity relationship between the objects.

5) ARTIFICIAL BEE COLONY ALGORITHM

The artificial bee colony (ABC) algorithm is an intelligent optimization algorithm for the actual parameter clusters [22], [23]. This algorithm is based on the foraging behavior of bee swarms. In the bee swarms, the employed, onlookers and scouts could maximize the nectar through different functions. Based on this principle, the optimal solution of the objective function can be solved. Compared with ant colony algorithm (ACA) [24], [25], ABC algorithm has faster convergence speed and higher estimation accuracy. In the ABC algorithm, the position of the solution in the optimization problem is equivalent to the food source, and its value is determined by the fitness value. To indicate that the initial solution is equal to the number of leading bees or following bees, of which the leading bee and the following bees each account for 50% of the colony, and the detection bee is converted by the leading bee after abandoning a food source. The main search process is as follows:

Firstly, randomly generate N initialization solutions, each solution x_i is composed of a d -dimensional vector, denoted as $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]$, where d represents the number of parameters included in the solution for the optimization problem, $i = 1, 2, \dots, N$;

Secondly, set the maximum number of cycles of the ABC algorithm C_{\max} , and set the maximum searchable number of feasible solutions at a location t . If the updated solution at this

position still cannot improve the fitness value after searching t times, then the feasible solution at this position would be discarded. Let the number of leading bees and following bees in each cycle be N , the number of cycles $C = 1$, start the cycle according to the following steps:

Step 1 Let $i = 1$, and perform a search for lead bee E_i according to step 2 until $i = N$;

Step 2 Correct the food source solution x_i according to the randomly updated solution v_i . If v_i improves the fitness value, replace the original solution with v_i , and also record the replaced solution as x_i ; otherwise, record the position of solution x_i which is one of the N initialized solutions, and the solution at this position is called candidate drop solution. v_i is determined by (23):

$$v_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) \quad (23)$$

where x_{ij} represents the j th parameter randomly selected in the solution x_i , v_{ij} represents the parameter of the candidate solution v_i generated by modifying x_{ij} , x_{kj} represents the j th parameter in the randomly selected solution x_k , k is in the neighborhood of i , and $k \neq i$, Φ_{ij} is a random number related to x_{ij} , $\Phi_{ij} \in [-1, 1]$.

Step 3 Calculate the probability p_i of the following bee selecting the food source x_i . The larger the value of p_i is, the better the fitness value could be obtained by x_i , and the greater the probability of being selected by the following bee is. p_i is determined by (24):

$$p_i = \frac{fit_i}{\sum_{n=1}^N fit_n} \quad (24)$$

where fit_i represents the fitness value proportional to the amount of nectar of the food source x_i , that is, the value of the objective function, it is meaning the fitness value of the i th solution. In the formula, N denotes the number of solutions.

Step 4 Let $j = 1$, remember that the food source selected by the follower O_j is x_i , and perform a search on O_j according to step 2 until $j = N$, where $i \in \{1, 2, \dots, N\}$;

Step 5 The detection bee compares the relationship between the number of records h_i and t of each candidate discarded solution in step 2. If $h_i > t$, the solution would be discarded, also let $h_i = 0$ and a new random feasible solution is generated according to formula (25):

$$r_{ij} = x_i^{\min} + \varphi_{ij}(x_i^{\max} - x_i^{\min}) \quad (25)$$

where r_{ij} represents the j th replacement parameter in the solution x_i , x_i^{\max} and x_i^{\min} represent the maximum and minimum values of x_i , respectively, and φ_{ij} denotes the random number associated to r_{ij} , $\varphi_{ij} \in [0, 1]$, $i \in \{1, 2, \dots, N\}$, $j = 1, 2, \dots, N$.

Step 6 Record the optimal solution at this time;

Step 7 Determine whether the algorithm has reached the maximum number of cycles or meets the termination conditions. If any of the conditions are met, the algorithm is terminated, otherwise $C = C + 1$ and the cycle is restarted from step 1.

Finally, when the algorithm terminates, the solution obtained in step 6 is the global optimal solution.

6) FITNESS FUNCTION DESIGN

The fitness function is the most important element in the ABC algorithm. A food source, that is, the fitness value of the solution can be understood as the degree to which the solution is close to the objective function. For function optimization problems, the fitness function is the objective function. For attribute reduction problem, there is no ready-made objective function available, so it is necessary to design a fitness function suitable for the problem.

The purpose of attribute reduction is to find the smallest set of conditional attribute whose dependency is equivalent to the original conditional attribute set, which means the maximum dependence and the smallest number of attributes are required. Therefore, the increase of the dependency in the fitness function or the decrease of the number of attributes should be beneficial to the increase of the overall fitness. In order to make the contributions to the fitness function of the dependency of the attribute set and the number of attributes in the solution equal, it is necessary to constrain the contributions of these two conditions to the fitness to the same order of magnitude. The fitness function in the design attribute reduction problem is as follows:

$$fit_i = 1/(1 - f_i) \quad (26)$$

where

$$f_i = (1 - \alpha) \left(\frac{|A| - |P_i|}{|A|} \right) + \alpha \frac{\gamma_{P_i}(Q)}{\gamma_A(Q)} \quad (27)$$

where fit_i represents the fitness value of the attribute subset P_i determined by the leading bee E_i , f_i represents the target value determined by the subset P_i , and according to $P_i \neq \emptyset$, it can be known that $0 < f_i < 1$. In formula (27) mentioned above, $(|A| - |P_i|)/|A|$ represents the reduction rate, and $\gamma_{P_i}(Q)/\gamma_A(Q)$ represents the normalization dependency of decision attribute Q to the subset P_i . And where α represents the weight factor, is used to adjust the flexibility of the fitness function.

Equation (27) shows that reduction rate and normalized dependency are positively related to f_i , that is, as the reduction rate and the normalized dependency increase, the f_i increases. However, there is not a positive correlation between the reduction rate and the normalized dependency. So how to obtain the maximum dependency and the minimum number of attributes on the premise of maximizing the target value is exactly the problem that this article is to solve.

7) REDUCTION METHOD 2

Based on the above analysis, this paper uses the ABC algorithm to complete the attribute reduction of the dataset. This method is based on the bee colony randomly selecting a subset of attributes. By calculating its fitness function through fuzzy rough sets, finally the best subset could be found during the iteration process. It is called the fuzzy rough artificial bee colony (FRABC) algorithm. By its processing we can see that the feature reduction algorithm of FRABC belongs to filter method. The specific steps of the algorithm are as follows.

Step 1 The hired bee randomly generates a subset of attributes. Supposing the attribute set P contains d attribute values $\{a_1, a_2, \dots, a_d\}$, the number of hired bees is N , and the number of randomly generated food source is N , and also the food source is considered as the initial solutions. x_i is composed of a d -dimensional vector, that is, $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]$, d represents the number of attributes included in the attribute set P , and $i = 1, 2, \dots, N$.

Step 2 Perform the operations of rounding and merging the same items for the generated solutions. For example, if $d = 8$, the randomly generated solutions are: $\{5.75, 6.30, 6.20, 3.75, 5.59, 2.20, 5.94, 1.22\}$. The same items are combined into: $\{6, 4, 2, 1\}$, which represents the 6th, 4th, 2nd, and 1st attributes in the attribute set P are selected to form a subset of attributes.

Step 3 Calculate the dependency of the N attribute subsets on the decision attributes according to the formula (8), that is, the importance of the attribute subsets, so as to obtain the fitness function of the ABC algorithm.

Step 4 Run the ABC algorithm based on the fitness function and N subsets of attributes.

The global optimal solution is the reduced attribute subset of the data set.

B. FEATURE EXTRACTION METHOD

In paper [15], [26] and [27], we extracted several feature sets, including gray level co-occurrence matrix (GLCM) feature set, entropy feature set, Zernike moment feature set and pseudo Zernike feature set, they are introduced as follows.

For the entropy feature set, the sample entropy S_e and fuzzy entropy F_e are extracted as follows [26]:

(1) Pre-processing the intra-pulse data of the radiation source signal to reduce the entropy features affected by the carrier frequency and noise;

(2) To avoid the influence of the difference in signal length on the entropy feature extraction, the signal is normalized and resampled;

(3) Determine the parameters used in estimating the sample entropy S_e , that is, take the scale parameter $m = 2$ and the tolerance parameter $r = 0.2\sqrt{2} \cdot \sigma$;

(4) Obtain the optimal parameters a, b, c of the fuzzy entropy F_e and the value of the sum by obtaining the entropy maximum of the fuzzy set;

(5) Calculate S_e and F_e use the sample entropy estimation algorithm and the fuzzy entropy formula, respectively.

The normalized energy entropy features P_e can be extracted as follows [27]:

(1) Endpoint extension of the data sequence according to the mirror closure extension method and resampling;

(2) Determine an EMD component of the signal sequence according to a convergence criterion;

(3) Find the normalized energy of each IMF component according to the formula;

(4) Calculate the normalized energy entropy P_e of the radiation source signal using the Shannon formula.

Finally, the entropy feature set $F_E = [S_e, F_e, P_e]$ is obtained.

The bispectral quadratic feature extraction algorithm is mainly performed as follows [15]:

- (1) Resample the intra-pulse data to maintain the same length of each signal;
- (2) Use the bispectrum estimation direct method to solve the bispectrum of the intra-pulse sequence of the radiation source signal;
- (3) Use the symmetry of the real-signal bispectrum, determine the following regions containing all the information of the bispectrum:

$$w_2 \geq 0, w_1 \geq w_2, w_1 + w_2 \leq \pi \quad (28)$$

In the condition, w_1 and w_2 represent the frequency axis of the bispectrum, the subsequent algorithm of the bispectral quadratic feature extraction is performed in this region;

- (4) Convert the bispectrum in the above region into a grayscale image;

- (5) Calculate the feature sets, including Zernike moment feature set $F_Z^{[12]}$, pseudo Zernike feature set $F_P^{[16]}$, gray level co-occurrence matrix (GLCM) feature set [15], and Hu-Invariant moment feature set F_φ for comparison, respectively, based on the calculation formulas of Zernike moments, pseudo-Zernike moments, gray-scale co-occurrence matrices and Hu-invariant moments.

It is worth noting that when calculating the features of the gray level co-occurrence matrix, the co-occurrence matrix of the four main directions is first obtained, and then the smoothing process is performed to extract the features, and the step size is taken as 1.

Before the sorting experiment, the data set is needed to be pre-processed, that is, the attributes of the data set would be blurred. For the sake of brevity, the standard ambiguity technique is used to obtain the TARFD fuzzy equivalent class determined by a single attribute, and the trapezoidal function and the trigonometric function are used to determine the membership of the objects in each data set with respect to the single attribute. Thus the similarity between objects of a single attribute are obtained, and then the membership degree of multiple fuzzy attributes in the TARFD and FRABC algorithms are determined by formulas (6) and (18), respectively. Based on the above analysis, the block diagram of radar signal sorting method based on feature extraction and feature selection is shown in Figure 2.

In order to verify the effectiveness of the extracted features and the algorithms of TARFD and FRABC, 6 typical radar signals widely used in international literature are selected for simulation experiments. These signals are: conventional wave radar signals (CW), linear frequency modulated radar signals (LFM), nonlinear frequency-modulated radar signals (NLFM), two-phase coded radar signals (BPSK), four-phase coded radar signals (QPSK), and frequency-coded radar signals (FSK). The signal carrier frequency is 850MHz, and its pulse width and sampling frequency are $10.8\mu s$ and 2.4GHz respectively. The frequency offset of the linear frequency

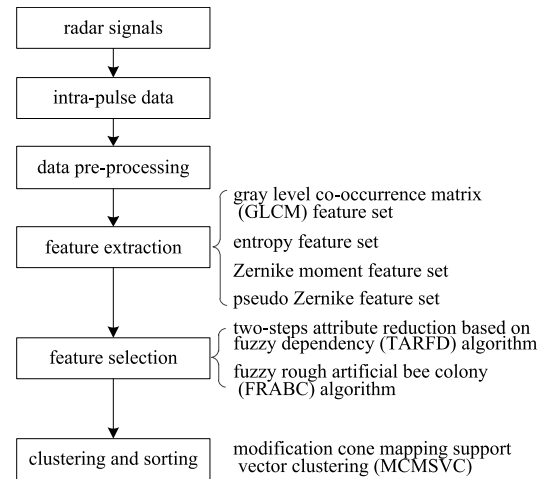


FIGURE 2. Radar signal sorting based on feature extraction and feature selection.

modulated radar signal (LFM) is set to 45MHz. The frequency of the non-linear frequency modulated radar signal (NLFM) is sinusoidally modulated. The coded radar signal (BPSK) uses a 31-bit pseudorandom code, the four-phase coded radar signal (QPSK) uses a Huffman code, and the FSK signal uses a Barker code. For these 6 type radar signals, 100 samples were generated within a typical SNR range (SNR = 15dB), for a total of 600 samples. The 200 samples generated were randomly selected for feature extraction and used for classifier training, and the remaining 400 samples were used for feature extraction and used as a test set for signal classification. The sample distribution is shown in Table 2.

TABLE 2. Distribution of the 6 typical radar signals.

signal type	SNR (dB)	number of samples	samples randomly selected for training	samples for testing
CW	15	100	30	70
LFM	15	100	32	68
NLFM	15	100	36	64
BPSK	15	100	40	60
QPSK	15	100	32	68
FSK	15	100	30	70

The experiments were performed on a Pentium (R) dual-core E5300 personal computer. The computer configuration is: CPU frequency is 2.6GHz, memory is 2GB, and hard disk is 250GB.

In order to ensure the data is comparable, each dimension feature sequence needs to be normalized. Here, it is processed according to the interval value method represented by formula (29).

$$f(i) = \frac{f(i) - \min(f(i))}{\max(f(i)) - \min(f(i))}, \quad i = 1, 2, \dots, N \quad (29)$$

In the formula, $f = [f(1), f(2), \dots, f(N)]$ represents a feature vector, N indicates the length of the vector.

In view of the fact that the radar radiation source training database is usually not complete and applied in the actual

battlefield environment, it is necessary to adopt a classifier that is suitable for training with fewer training samples, and has a faster training and classification speed, and is not easy to fall into local minimums. The support vector machine (SVM) has these advantages. Therefore, the SVM is selected here as a learning algorithm for classification and recognition. The SVM can not only solve many problems existing in previous machine learning methods, such as practical problems of non-linearity, small samples, high dimensions, local minimum points, over-learning, etc. Compared with the neural network learning (NNL) algorithm based on the principle of empirical risk minimization (ERM), the support vector machine (SVM) has better generalization ability and stronger theoretical basis. The SVM uses a kernel function to complete the mapping of the input feature vector from a low-dimensional feature space to a high-dimensional feature space. With this method, the nonlinear separable problem of the original space can be transformed into the linear separable problem in a high-dimensional space. Through this process, the purpose of classification and identification of the radiation source signal can be achieved.

The comparison of classification accuracy is shown in Table 3. In Table 3, the AN represents the number of condition attributes, the Acc represents the classification accuracy, and the ET represents the execution time of reduction method.

TABLE 3. Statistical results of classification of different feature subsets.

reduction method	origin	TARFD	FRABC
	AN	7	2
F ϕ	Acc(%)	76.47	76.15
	ET(s)	—	6.387
	AN	3	3
FE	Acc(%)	97.17	97.17
	ET(s)	—	0.271
	AN	12	2
FZ	Acc(%)	89.18	88.50
	ET(s)	—	11.437
	AN	15	3
FP	Acc(%)	92.59	89.33
	ET(s)	—	14.659
	AN	20	3
FG	Acc(%)	86.04	85.03
	ET(s)	—	17.519
	AN	11.4	2.6
Average	Acc(%)	88.36	87.20
	ET(s)	—	9.293

The results in Table 3 show that the TARFD algorithm can obtain the smallest average feature subset without significantly reducing the average classification accuracy rate. The average number of features in the subset is 2.2. For the FRABC algorithm, the average number is 2.6. In addition, compared with the TARFD reduction algorithm, the FRABC algorithm has a higher average accuracy rate, reaching 87.20%. For the execution time of the algorithms, the average execution time of the TARFD algorithm is 9.293s and of the

FRABC algorithm is 10.055s. Compared with the FRABC algorithm, the TARFD is more efficient.

In addition, from the comparison of the classification accuracy of different intra-pulse feature subsets in Table 3, it can be seen that the classification results of the reduced subsets obtained by the TARFD and FRABC algorithms are close to the classification results before the reduction, indicating that the feature subsets selected by these two algorithms can obtain most of the classification information of the original feature set. Therefore, the subsets of these feature sets are used for cluster sorting in the following, and the relationship between the accuracy rate and the classification precision is comprehensively considered.

The FRABC algorithm obtains the following feature subsets to characterize the intra-pulse information of the radiation source signal: sample entropy, fuzzy entropy and normalized energy entropy features in the entropy feature, that is, $F'_E = [S_e, F_e, P_e]$, Zernike moment feature subset $F'_Z = [Z_{42}, Z_{62}]$, and pseudo Zernike moment feature subset $F'_P = [P_{21}, P_{53}, P_{54}]$, the bi-spectrum smooth gray level co-occurrence matrix feature subset $F'_G = [f_4, f_{14}]$ and the first-order and second-order moments $F'_\phi = [\phi_1, \phi_2]$ of Hu-invariant moments, of which F'_ϕ is still used for the experimental comparison later.

After extracting the intra-pulse features by the above method and pre-treating, the feature distribution of each feature subset of the experiment is shown in Figure 3.

III. SIGNAL SORTING BASED ON FEATURE SET

A. RADIATION SOURCE SIGNAL PARAMETER SETTING

Assume that the parameters of these 6 radar emitter signals used in the paper are shown in Table 4, and the pulse flows are assumed to be in the same direction.

TABLE 4. Radiation source signal parameters.

radar number	PRI(μ s)	RF(MHz)	PW(μ s)
Rd1	200~330	3200~3400	15
	Repeated frequency slip	Single pulse agility 3000~3300	Fixed frequency
Rd2	550/680/850	Three-pulse frequency hopping, 12 frequency points	10~20 Agile frequency
	Three staggered	2850~3150	32
Rd3	780 Repetitive jitter amount of 15%	Three-pulse agility	Fixed frequency
	250~450	2900~3100	10
Rd4	Repetitive frequency agile (synchronized with RF)	Three-pulse pulse group agility	Fixed frequency
	960/1050/1160/1290/1440/1610	3250~3550	20~32
Rd5	Six staggered	Single pulse agility	Agile frequency
	900 Repetitive jitter amount of 15%	3300~3500	25
Rd6		Double pulse agility	Fixed frequency

In the intra-pulse modulation mode, the frequency offset of the LFM requires $B \cdot \tau$ keeping at 100MHz μ s, where B

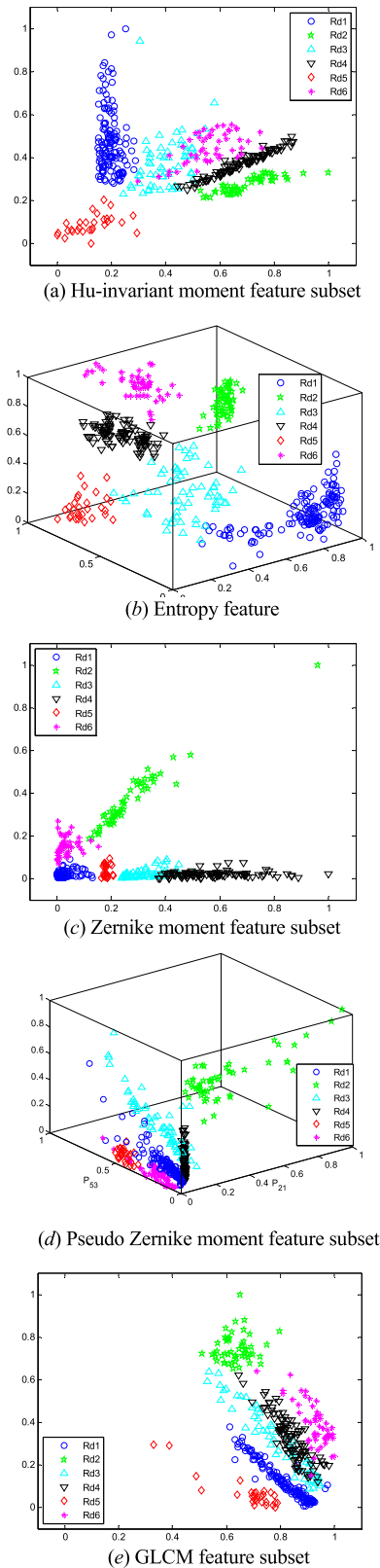


FIGURE 3. Feature distribution of each feature subset.

is bandwidth and τ is the pulse width; BPSK and FSK use 13 bit Barker code, QPSK uses 16 bit Frank code, and NLFM bandwidth is 6~10MHz, using sine frequency modulation.

It can be seen from Table 4 that the 6 radar signals to be sorted have complex and variable features in addition to the inter-pulse parameters with severe overlap.

Let the receiver have an intermediate frequency of 30 MHz, a bandwidth of 20 MHz, and an ADC sampling frequency of 150 MHz. The TOA of the starting pulse of a radiation source with a uniformly distributed random number from 0 to the total interception time T_{int} is simulated. When the total intercept time is $T_{int} = 50ms$, a typical experiment (SNR = 15dB) produces a total of 476 pulses for the 6 type radar signals in Table 4. The details of the generated pulses and residual pulses are shown in Table 5.

TABLE 5. Simulation generates pulse loss.

radiation source	Rd1	Rd2	Rd3	Rd4	Rd5	Rd6	total number	loss rate
generate pulse	152	59	65	112	32	152	476	9.24%
residual pulse	138	54	59	101	29	138	432	

It should be pointed out that in the experiment, each pulse contains many intermediate frequency (IF) sampling data. Feature extraction is performed on these sampling data of each pulse. According to different algorithms, feature sets of different dimensions are obtained. For example, for a pulse of Rd1 with a pulse width of $15\mu s$ and a sampling frequency of 150 MHz, the number of data obtained after this pulse is sampled is 2250. Entropy feature extraction is performed on these data to obtain a three-dimensional feature vector $F'_E = [S_e, F_e, P_e]$. Further, for the 152 pulses generated by Rd1, the total data generated is 152×2250 . After entropy feature extraction is performed, the resulting feature matrix row \times column is 152×3 , that is, there are 152 samples in Rd1 in total. Each sample generates three entropy features and composes the entropy feature vector F'_E . It's all like this for Rd2, Rd3, Rd4, Rd5, and Rd6.

B. EXPERIMENT OF SIGNAL SORTING

The following 5 kinds of feature subsets are used for representing signal sorting, and the MCMSVC algorithm is still used for clustering, and the SE index is used to adjust the clustering parameters. The initial value is set $C = 1$ during simulation, and the parameter adjustment of q is similar to the step of adjusting with the inter-pulse parameter sorting according to formula $q = 1/\max_{ij} \|g_i - g_j\|^2$. In addition, as can be seen from Table 5, the actual pulse numbers of the radiation sources Rd1-Rd6 for sorting are 138, 54, 59, 101, 29 and 51, respectively, and the radiation source signals represented by each feature subset are classified by SE-MSVC. The results are shown in Table 6. The results in the table are the statistical average of 20 sorting results, n_i indicating the actual number of pulses of i th radiation source signal, NCSP means number of correct sorting pulses, NMP means number of misselected pulses and TSA means total sorting accuracy.

The following conclusions can be drawn from Figure 3 and Table 6:

TABLE 6. Comparison of results when sorting with each feature subset.

feature subset	sorting result	Rd1 $n_1=138$	Rd2 $n_2=54$	Rd3 $n_3=59$	Rd4 $n_4=101$	Rd5 $n_5=29$	Rd6 $n_6=51$
F'_ϕ	NCS	132	49	25	75	26	32
	P						
	NMP	6	5	34	26	3	19
	TSA			78.47%			
F'_E	NCS	137	54	34	86	25	42
	P						
	NMP	1	0	25	15	4	9
	TSA			87.50%			
F'_Z	NCS	114	54	34	95	13	40
	P						
	NMP	24	0	25	16	16	11
	TSA			81.01%			
F'_P	NCS	118	54	39	53	49	45
	P						
	NMP	20	0	20	49	0	6
	TSA			82.87%			
F'_G	NCS	136	49	44	51	28	30
	P						
	NMP	2	5	15	47	1	21
	TSA			82.17%			

(1) The distribution of feature subsets shown in Figure 3(b) shows that the three-dimensional entropy feature of Rd1, Rd2 and Rd5 with CW, LFM and NLFM modulation modes is better, so when using entropy feature sorting, the three types of radiation source signals have fewer misselected pulses, especially the radiation source Rd2 is selected correctly, while Rd1 and Rd5 also have only 1 and 4 misselected pulses respectively. Relatively speaking, Rd3, Rd4 and Rd6 misselected more pulses. The result is determined by the modulation mode that the signal itself has. The analysis is corresponding to the items F'_E listed in Table 6.

(2) When the sorting method based on the bispectral quadratic feature, that is, using the feature subset F'_P , F'_Z , F'_G and F'_ϕ for sorting, since the bispectral gray image determines that each feature subset exhibits a characteristic distribution with a certain trend, the sorting accuracy rate is lower than the characteristic sorting result with 87.50% correct rate using F'_E , as shown in Table 6. However, the average sorting accuracy rate of more than 80% is also the same as the sorting result obtained by using the inter-pulse parameters, which means this method is effective. In addition, the results in the table also show the sorting effect of several feature subsets F'_P , F'_Z , F'_G are better than that obtained when the subset F'_ϕ is used.

(3) Comparing the results obtained by using the inter-pulse parameters in [28], it can be seen that both Rd2 and Rd5 have more misselected pulses. In addition to the corresponding results, Rd5 has more misselected pulses using F'_Z . And the remaining feature subsets are significantly less misselected pulses than the number of Rd2 and Rd5 caused by the inter-pulse parameter sorting, indicating that the feature

subsets obtained by these feature extraction methods effectively compensate for the disadvantage that some signals cannot be sorted efficiently with the inter-pulse parameters. At the same time, this point can be explained by comparing the distribution of the parameters of the inter-pulse parameters in [28]. It can be seen from [28] that the RF-PW feature distribution of the radiation source Rd1 and Rd2, Rd5 and Rd6 overlaps severely, so that the SE-MSVC clustering method with excellent sorting performance can not achieve satisfactory results. Among them, Rd2 and Rd5 have more misselected pulses. In general, the extracted features of the intra-pulse data supplement the new features for sorting except for the conventional parameters. Thus the characteristic representation of the signal is performed from a new view, and a better sorting effect could be obtained.

Although the Hu-invariant moment feature set and subset are not as good as other feature sets in classification and clustering, considering the effectiveness of the feature set in characterizing the bispectral information of the radiation source signal, the feature set is still added to the total feature vector for clustering experiments.

The following considers the comprehensive use of these several intra-pulse features for signal sorting, that is, the SE-MSVC sorting of the interleaved pulse stream is completed by using the feature vector $G = [S_e, F_e, P_e, \phi_1, \phi_2, f_4, f_{14}, Z_{42}, Z_{62}, P_{21}, P_{53}, P_{54}]$. At this time, the pulse data shown in Table 6 is still taken, and the clustering sorting correctly rate using G is shown in Table 7. The result in the table is the statistical average of 20 sorting results.

TABLE 7. Results when sorting using feature vectors G .

sorting result	Rd1 $n_1=138$	Rd2 $n_2=54$	Rd3 $n_3=59$	Rd4 $n_4=101$	Rd5 $n_5=29$	Rd6 $n_6=51$
NCSP	138	54	48	97	21	38
NMP	0	0	11	4	8	13
TSA			91.67%			

As can be seen from Table 7, at the same typical signal-to-noise ratio, the correct rate of 91.67% is only about 4% higher than the average correct rate obtained by using the sorting parameter F'_E . Compared with the results obtained by using the feature subsets shown in Table 6, except for the Zernike moment feature subset F'_Z , the other sorting results are less than the number of Rd5 misselected pulses obtained by using G . The reason for this phenomenon is that there is an interaction relationship between the features included in the vector G , so that the sorting effect of some of the radiation sources is rather reduced.

IV. FEATURE SELECTION AND SIGNAL SORTING BASED ON FEATURE SUBSET

Considering the above factors and the relatively low computational complexity of the TARFD algorithm, the feature set G is selected again by TARFD. Under the typical signal-to-noise

ratio, the experiment is performed according to the pulse data shown in Table 4, and the feature selection of the TARFD algorithm is adopted. The process is as follows.

According to the proposed algorithm of TARFD, first makes the decision attribute $Q \in \{1, 2, 3, 4, 5, 6\}$, for the convenience of the narrative, the candidate feature set is:

$$G = [S_e, F_e, P_e, \varphi_1, \varphi_2, f_4, f_{14}, Z_{42}, Z_{62}, P_{21}, P_{53}, P_{54}] \\ = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}] \quad (30)$$

The membership is then calculated from the fuzzy function. Through a large number of experiments, it is found that the characteristic membership degree of each radiation source signal calculated by using trapezoidal function and trigonometric function is more favorable for feature selection. Therefore, the membership function corresponding to the condition attribute is calculated by using the membership function.

Then, according to formula (9), the dependence $\gamma'_{a_i}(Q)$ of each feature is calculated, and a_i denotes the i th feature in the feature set $G, i = 1, 2, \dots, 12$. The obtained feature dependence is shown in Table 8.

TABLE 8. Dependence of each dimension feature.

feature	a_1	a_2	a_3	a_4	a_5	a_6
dependency	0.1693	0.0786	0.0906	0.0902	0.0300	0.0471
feature	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
dependency	0.0630	0.0089	0.0171	0.0166	0.0558	0.0105

Once the threshold T_h is determined, it is needed to judge according to the threshold whether the feature needs to be reduced in advance. In the experiment the T_h is set to $T_h = 0$, so the next step is directly performed.

Then calculate the degree of interdependence between features a_i and a_j according to formula (9) to obtain the interdependency relationship as shown in Table 9.

After obtaining the dependency degree between attributes, it is needed to judge whether the feature is redundant according to the condition: if $\gamma'_{a_j}(Q) \leq \gamma'_{a_i}(Q)$ and $\gamma'_{a_j, a_i} \leq \gamma'_{a_i, a_j}$, the redundant feature a_j is eliminated, wherein $i, j = 1, 2, \dots, m, m = |G|, i \neq j, |G|$ represents the dimension of the feature set, and γ'_{a_i, a_j} represents the interdependency of the i th row and j column in table 9.

For example, as can be seen from Table 7 and Table 8, for the features a_1 and a_2 , we have $\gamma'_{a_1}(Q) = 0.1693, \gamma'_{a_2}(Q) = 0.0786, \gamma'_{1,2} = 0.1212, \gamma'_{2,1} = 0.2361$, obviously, the redundancy condition is not satisfied. But for features a_3 and a_4 , there are $\gamma'_{a_3}(Q) = 0.0906, \gamma'_{a_4}(Q) = 0.0902, \gamma'_{3,4} = 0.1130, \gamma'_{4,3} = 0.0929$, and at this time we have $\gamma'_{a_4}(Q) < \gamma'_{a_3}(Q)$ and $\gamma'_{4,3} < \gamma'_{3,4}$ therefore a_4 is the redundant features, and it should be eliminated.

After the redundant features are removed, a subset of the feature sets G is obtained. Table 10 lists the different levels of feature subsets obtained during the reduction process, where the last result refers to the reduced subset obtained from the TARFD algorithm to the termination condition.

TABLE 9. Interdependence relationship between features.

feature	a_1	a_2	a_3	a_4	a_5	a_6
a_1	0.5004	0.1212	0.1242	0.1751	0.1425	0.0472
a_2	0.2361	0.5014	0.1085	0.2163	0.2470	0.0746
a_3	0.1344	0.0460	0.5026	0.1130	0.1777	0.0906
a_4	0.1014	0.0839	0.0929	0.5034	0.1052	0.1012
a_5	0.0354	0.0329	0.0252	0.0336	0.5030	0.0308
a_6	0.1011	0.0598	0.1678	0.3835	0.2206	0.5015
a_7	0.1010	0.1014	0.1496	0.3138	0.0630	0.1060
a_8	0.0466	0.0409	0.1344	0.3009	0.1861	0.0773
a_9	0.2216	0.0778	0.1634	0.1031	0.2271	0.0665
a_{10}	0.2937	0.0660	0.1920	0.2616	0.2800	0.0919
a_{11}	0.0448	0.0492	0.1084	0.1201	0.1180	0.0776
a_{12}	0.0472	0.0537	0.0363	0.2221	0.0867	0.0715
feature	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
a_1	0.0857	0.0393	0.1484	0.1671	0.0576	0.1601
a_2	0.1855	0.0620	0.0961	0.1013	0.1707	0.0766
a_3	0.1103	0.0573	0.2457	0.0906	0.1337	0.0952
a_4	0.0874	0.0747	0.0725	0.0886	0.1092	0.0689
a_5	0.0368	0.0382	0.0360	0.0331	0.0410	0.0370
a_6	0.1953	0.1080	0.0834	0.0856	0.1693	0.1443
a_7	0.5042	0.0683	0.0883	0.1309	0.0612	0.0793
a_8	0.0738	0.5034	0.0192	0.2133	0.0587	0.0273
a_9	0.0209	0.0258	0.5021	0.0230	0.2324	0.1165
a_{10}	0.2117	0.0225	0.0278	0.5129	0.0137	0.1592
a_{11}	0.0670	0.0558	0.1067	0.0567	0.5008	0.1029
a_{12}	0.1227	0.0116	0.0700	0.0954	0.1507	0.5018

TABLE 10. Feature subsets obtained from different reductions.

reduction order	reduction result	feature dimension
original feature set	{1,2,3,4,5,6,7,8,9,10,11,12}	12
4th reduction	{1,2,3,8,5,12,9,10,6,11}	10
7th reduction	{1,2,3,8,12,9,10,6,11}	9
10th reduction	{1,2,3,8,9,10,6,11}	8
12th reduction	{1,2,3,8,9,6,11}	7
13th reduction	{1,2,3,8,6,11}	6
Last result	{2,3,8,6,11}	5

Thus an optimal subset of the feature set is obtained:

$$G' = [F_e, P_e, f_4, Z_{42}, P_{53}] \quad (31)$$

The results in Table 10 show that the Hu-invariant moment feature subset belongs to the redundant feature set, while the fuzzy entropy feature F_e , the normalized energy entropy feature P_e , the correlation feature of the bispectral gray level co-occurrence matrix f_4 , and the Zernike moment feature Z_{42} with the repetition degree of 2 and 4 order, and the pseudo Zernike moment feature P_{53} with the repetition degree of 3 and 5 order belong to the optimal feature determined by the algorithm.

Using the feature subset G' for sorting, the results of 20 sorting statistical averages are shown in Table 11, also n_i indicating the actual number of pulses of i th radiation source signal.

The results in Table 11 show that the results obtained by using subsets G' are slightly lower than the results obtained by using subsets G , and the average correct rate differs by less than 1%.

Compared with the results in Table 7, the misselected pulses of Rd4 increase by using subsets G' , but the misselected pulses of Rd5 are obviously less. This shows that the

TABLE 11. Results when sorting using feature subsets G' .

Sorting result	Rd1 $n_1=138$	Rd2 $n_2=54$	Rd3 $n_3=59$	Rd4 $n_4=101$	Rd5 $n_5=29$	Rd6 $n_6=51$
NCSP	138	54	48	87	28	38
NMP	0	0	11	14	1	13
TSA	90.97%					

sorting results of the two feature sets are not much different, indicating that the method of sorting by using subsets G' is effective. In addition, considering the feature dimension, the feature subset G' only contains 5-dimensional features. Compared with the feature set G containing 12-dimensional feature vectors, the feature dimension is obviously reduced, and the sorting result with little difference from the feature set G is obtained. The result illustrates the effectiveness of the feature selection method proposed in the paper.

V. CONCLUSIONS

This paper studies and verifies the feature extraction and feature selection methods using for radar signal sorting. In the experiment, only the case of intra-pulse noise at typical signal-to-noise ratio (SNR = 15dB) is considered. The results show that with the proposed two-step attribute reduction (TARFD) method based on fuzzy dependence and the fuzzy rough artificial bee colony (FRABC) algorithm, the selected optimal subset of features can effectively sort the radar signals. At the same time, after reducing the dimension of the feature subset efficiently, the sorting result with little difference from the feature set G is obtained, and the correct rate is 90.97%.

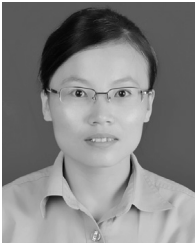
REFERENCES

- [1] Q. Guo. *Research on Radar Signal Sorting Theory*. Beijing, China: Science Press, 2010, pp. 78–96.
- [2] W. Zhang, F. Sun, and B. Wang, "Radar signal intra-pulse feature extraction based on improved wavelet transform algorithm," *Int. J. Commun., Netw. Syst. Sci.*, vol. 10, no. 08, pp. 118–127, 2017.
- [3] Y. Cao, J. L. Bai, H. Li, J. Zhao, S. Liu, and Z. Gao, "Deep representation method for radar emitter signal using wavelet packets decomposition," *J. Eng.*, vol. 2019, no. 19, pp. 6282–6286, Oct. 2019.
- [4] S. J. Quan, W. P. Qian, and J. H. Guo, "A rapid method of extracting unintentional modulation feature using Gabor atoms," in *Proc. CSTT&CT*, Guangzhou, China, 2015, pp. 455–463.
- [5] Z. T. Yang, W. Qiu, H. J. Sun, and A. Nallanathan, "Robust radar emitter recognition based on the three-dimensional distribution feature and transfer learning," *Sensors*, vol. 16, no. 3, pp. 289–302, Feb. 2016.
- [6] X. L. Fan, T. Li, and S. Su, "Intrapulse modulation type recognition for pulse compression radar signal," *J. Appl. Remote Sens.*, vol. 11, no. 3, pp. 1–19, Sep. 2017.
- [7] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2437–2441.
- [8] M. Zhang, L. Liu, and M. Diao, "LPI radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, pp. 1682–1702, Oct. 2016.
- [9] C. L. Yu, J. Han, and F. Wan, "Sorting radar signal based on the resemblance coefficient of bispectrum two dimensions characteristic," in *Proc. ICMIC&ICOEE*, Guilin, China, 2016, pp. 286–294.
- [10] R. Cao, J. Cao, J.-P. Mei, C. Yin, and X. Huang, "Radar emitter identification with bispectrum and hierarchical extreme learning machine," *Multimedia Tools Appl.*, vol. 78, no. 20, pp. 28953–28970, Oct. 2019.

- [11] C. M. Jeong, Y. G. Jung, and S. J. Lee, "Neural network-based radar signal classification system using probability moment and ApEn," *Soft Comput.*, vol. 22, no. 13, pp. 4205–4219, Jul. 2017.
- [12] L. D. Zhang, T. Lu, D. F. Wang, and S. Q. Wang, "A new radar signal feature extraction algorithm based on Zernike moment bispectrum," *Shipboard Electron. Countermeasure*, vol. 35, no. 6, pp. 43–47, Nov. 2012.
- [13] J. Gao, L. Shen, and L. Gao, "Modulation recognition for radar emitter signals based on convolutional neural network and fusion features," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 12, pp. 3612–3632, Dec. 2019.
- [14] Z. L. Wang, D. F. Zhang, and S. Q. Wang, "Pseudo-Zernike moment bispectrum quadratic feature extraction of radar emitter signal," *Modern Radar*, vol. 35, no. 11, pp. 39–43, Dec. 2013.
- [15] Z. L. Wang, D. F. Zhang, and S. Q. Wang, "Two-spectrum quadratic feature extraction method for radar emitter signal," *J. Air Force Eng. Univ. (Natural Sci.)*, vol. 15, no. 1, pp. 48–52, Jan. 2014.
- [16] J. Dudczyk, "A method of feature selection in the aspect of specific identification of radar signals," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 65, no. 1, pp. 113–119, Feb. 2017.
- [17] R. Jensen and N. Mac Parthaláin, "Towards scalable fuzzy-rough feature selection," *Inf. Sci.*, vol. 323, pp. 1–15, Dec. 2015.
- [18] T. K. Sheeja and A. S. Kuriakose, "A novel feature selection method using fuzzy rough sets," *Comput. Ind.*, vol. 97, pp. 111–116, May 2018.
- [19] X. Zhang, C. Mei, D. Chen, and Y. Yang, "A fuzzy rough set-based feature selection method using representative instances," *Knowl.-Based Syst.*, vol. 151, pp. 216–229, Jul. 2018.
- [20] A. Mahanipour, H. Nezamabadi-pour, and B. Nikpour, "Using fuzzy-rough set feature selection for feature construction based on genetic programming," in *Proc. 3rd Conf. Swarm Intell. Evol. Comput. (CSIEC)*, Mar. 2018, pp. 1–6.
- [21] Z. Wang, D. Zhang, D. Bi, and S. Wang, "Multiple-parameter radar signal sorting using support vector clustering and similitude entropy index," *Circuits, Syst., Signal Process.*, vol. 33, no. 6, pp. 1985–1996, Jun. 2014.
- [22] X. Li and G. Yang, "Artificial bee colony algorithm with memory," *Appl. Soft Comput.*, vol. 41, pp. 362–372, Apr. 2016.
- [23] Y. Xue, J. Jiang, B. Zhao, and T. Ma, "A self-adaptive artificial bee colony algorithm based on global best for global optimization," *Soft Comput.*, vol. 22, no. 9, pp. 2935–2952, May 2018.
- [24] R. K. V. Penmatsa and P. Kakarlapudi, "Web phishing detection: Feature selection using rough sets and ant colony optimisation," *Int. J. Intell. Syst. Des. Comput.*, vol. 2, no. 2, pp. 102–113, Jan. 2018.
- [25] P. R. K. Varma, V. V. Kumari, and S. S. Kumar, "Hybrid computational intelligent attribute reduction system, based on fuzzy entropy and ant colony optimization," in *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications (Advances in Intelligent Systems and Computing)*, vol. 628. Springer, Jan. 2018, pp. 307–318.
- [26] S. Q. Wang, G. A. Zhou, B. J. Song, C. Y. Gao, and P. F. Wan, "Research on radar emitter signal feature extraction method based on fuzzy entropy," *Procedia Comput. Sci.*, vol. 154, pp. 508–513, Jan. 2019.
- [27] S. Q. Wang, C. Y. Wang, Q. Sun, C. Y. Gao, and S. G. Yang, "Radar emitter signal intra-pulse feature extraction based on empirical mode decomposition," *Procedia Comput. Sci.*, vol. 154, pp. 504–507, Jan. 2019.
- [28] S. Q. Wang, C. Y. Gao, X. C. Li, H. Y. Zeng, and J. Bai, "Comparison of radar signal sorting method between single and multi-parameter based on," in *Recent Trends in Intelligent Computing, Communication and Devices (Advances in Intelligent Systems and Computing)*, vol. 1006, Jan. 2020, pp. 851–857.



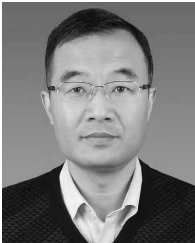
SHIQIANG WANG was born in Shanxi, China. He received the Ph.D. degree in information and communication engineering from Air Force Engineering University, China, in 2012. He is currently an Assistant Professor with the Air and Missile Defense College, Air Force Engineering University. His current works include intelligent information processing, radar signal processing, deep learning, and radar engineering.



CAIYUN GAO was born in Shanxi, China. She received the M.Eng. degree in aerospace manufacturing from Northwestern Polytechnical University, Xi'an, China, in 2009. She is currently an Instructor with the Air Force Engineering University, Xi'an. Her current works include signal processing, aerospace manufacturing, deep learning, and mechanical manufacturing.



GUIMEI ZHENG received the B.Eng. degree in biomedical engineering and the Ph.D. degree in signal and information processing from Xidian University, China, in 2009 and 2014, respectively. He is currently a full-time Postdoctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include MIMO radar and vector sensor array signal processing.



QIN ZHANG received the Ph.D. degree in information and communication engineering from Xidian University, Xi'an, China, in 2018. He is currently a Professor with the Air and Missile Defense College, Air Force Engineering University, China. His current work is array signal processing of MIMO radar.



JUAN BAI received the Ph.D. degree in telecommunication engineering from the Beijing University of Post and Telecommunications (BUPT), in 2015. She is currently a Faculty Member with Air Force Engineering University. Her major research is physical-layer security in stochastic wireless networks.



VEERENDRA DAKULAGI received the Ph.D. degree in array signal processing from Visvesvaraya Technological University, Belagavi, India. From 2007 to 2008, he was an Assistant Professor with the Sathagiri Engineering College, Bangalore. From 2009 to 2010, he was an Assistant Professor with BKIT, Bhalki, India. From 2010 to 2017, he was an Assistant Professor with the Department of E&CE, Guru Nanak Dev Engineering College, Bidar, India, where he was an Associate Professor, from 2017 to 2019. He is currently working with the Department of Electronics and Communication Engineering, Lincoln University College, Kuala Lumpur, Malaysia. His research interests are in the area of signal processing and communications that include statistical and array signal processing, adaptive beamforming, spatial diversity in wireless communications, multiuser, and MIMO communications. He has published over 40 technical articles (including the IEEE, Elsevier, Springer, and Taylor and Francis) and 02 patents in these areas. Dr. Dakulagi received many awards in education including, the Award for Research Publications by Vision Group on Science and Technology, Government of Karnataka, India under the leadership of Bharat Ratna Prof. C. N. R. Rao. He is currently serving as a Reviewer to various IEEE, Elsevier, Taylor & Francis, Springer journals, and International Conferences. He is the Editor-in-Chief of the *Journal of Advanced Research in Wireless, Mobile and Telecommunication* and the *Journal of Advanced Research in Signal Processing and Applications*. He also serves as an Associate Editor/Editorial Board for the *Journal of Communication Engineering and Systems* and the *Journal of Electronics Engineering*.



YUWEI SONG received the B.Eng. degree in biomedical engineering and the M.S. degree in circuits and systems from Xidian University, China, in 2009 and 2012, respectively. She is currently pursuing the Ph.D. degree with Air Force Engineering University, Xi'an, China. Her research interest includes array radar signal processing.



JILIANG CAI received the Ph.D. degree in electromagnetic field and microwave technology from Air Force Engineering University (AFEU), in 2013. He is currently a Faculty Member with Air Force Engineering University. His major researches include radar signal processing, deep learning, and radar engineering.



HUIYONG ZENG received the Ph.D. degree in electromagnetic field and microwave technology from Air Force Engineering University (AFEU), in 2013. He is currently a Faculty Member with Air Force Engineering University. His interests include electromagnetic field and microwave technology, radar signal processing, signal and information processing.



BINFENG ZONG received the bachelor's degree in radar engineering from Air Force Engineering University, Xi'an, China, in 2010, and the master's and Ph.D. degrees in electronic science and technology from Air Force Engineering University, Xi'an, China, in 2012 and 2016, respectively. He is currently an Instructor with the Air and Missile Defense College, Air Force Engineering University. His current research interests include aviation electronics, millimeter wave antennas and arrays, and MIMO radar.

...