Reconstruction of Radar Pulse Repetition Pattern via Semantic Coding of Intercepted Pulse Trains

SHUO YUAN
SHI-QIAN KANG
WEN-XIU SHANG
National University of Defense Technology, Changsha, China

ZHANG-MENG LIU
National University of Defense Technology, Changsha, China
Tianjin Institute of Advanced Technology, Tianjin, China

The timing structure of multiple successive radar pulses constructs a high-dimensional pattern for intercepted pulse trains, which is called as pulse repetition interval (PRI) pattern. By treating the radar as a machine that uses differently permuted pulses to communicate with surroundings, this pattern acts as the grammatical structure of its language. Compared with the discrete PRI set that is conventionally used for pulse train description, PRI pattern contains richer and more condensed structural information about radar pulse train, which is helpful for pulse deinterleaving and radar recognition. This article introduces the semantic coding theory to reveal and reconstruct PRI pattern from the intercepted radar pulse train. We first define the coding complexity of a pulse train, which is divided into two parts: 1) the complexity of encoding a PRI pattern dictionary with each element consisting of several successive PRIs; and 2) encoding the intercepted pulse train based on this dictionary. The coding complexity is then minimized by optimizing the components in dictionary, and the PRI timing patterns are finally obtained from the dictionary when the minimization is reached. The effectiveness of semantic coding model and PRI pattern reconstruction method is verified in the simulation part.

Manuscript received 15 November 2021; revised 4 May 2022; accepted 26 June 2022. Date of publication 30 June 2022; date of current version 10 February 2023.

DOI. No. 10.1109/TAES.2022.3187385

Refereeing of this contribution was handled by W. P. du Plessis.

This work was supported in part by the Hunan Provincial Natural Science Foundation of China under Grant 2020JJ2037 and in part by the Youth Talent Support Programme of Hunan Province under Grant 2019RS2026.

Authors' address: Shuo Yuan, Shi-Qian Kang, and Wen-Xiu Shang are with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), National University of Defense Technology, Changsha 410073, China, E-mail: (ysyuanshuo@126.com; sqkang@163.com; 851506204@qq.com), Zhang-Meng Liu is with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), National University of Defense Technology, Changsha 410073, China, and also with the Tianjin Institute of Advanced Technology, Tianjin, China, E-mail: (liuzhangmeng@nudt.edu.cn). *(Corresponding author: Zhang-Meng Liu.)*

I. INTRODUCTION

Pulse repetition interval (PRI) pattern of radar refers to a permutation of differential time-of-arrival (DTOA) extracted from several successive pulses with time-domain repetition regularity. It is a structural characteristic with high dimension in the pulse train, including the particular values, ordering, and switching law of PRIs. Among them, PRI values reflect the detection distance and platform of emitters to a certain extent, the ordering reflect signals' timing structure, which is required by online deinterleaving [1] and recognition [2], and the switching law, also known as PRI modulation type, indicates working modes of radar. For conventional radars using fixed and staggered PRIs [3], a complete PRI pattern is the basic temporal structure of pulses that repeats periodically all through the pulse train. For multifunction radars, PRI patterns are defined according to the frequently-used pulse episodes having determined interpulse DTOA sequences. In contrast, conventional radar signals account for a large proportion of the intercepted data because regular PRI patterns are widely used in radar systems to reduce the cost. Therefore, it is necessary to reconstruct PRI patterns of the conventional radar by analyzing its pulse train.

PRI pattern reconstruction comes from the previous tasks of PRI estimation [4]–[6] and PRI modulation recognition [7], [8]. PRI values and their modulation types are important characteristics of emitters, which are obtained noncooperatively for deinterleaving and recognition in electronic support. However, due to the serious parameter overlaps in intercepted pulse train, only obtaining discrete PRI values cannot meet the requirements of rapid and accurate deinterleaving. In addition, agile, fixed, stagger, and other PRI modulations are identified to judge radar working modes in the past. However, some new radars [9] can work under fine-grained parameters with the same modulation type but different parameters to meet different working modes. Compared with the discrete PRI values and several PRI modulation types, PRI pattern is a higher-dimensional characteristic. For instance, under the radar's PRI pattern of $[PRI_1, PRI_2, PRI_3]$, the discrete values $\{PRI_1, PRI_2, PRI_3\}$ from conventional statistic methods is a 1-D description of the intercepted pulse train. In contrast, PRI pattern reconstruction can obtain $[PRI_1, PRI_2, PRI_3]$, including PRI values and their structural relationship. This high-dimensional feature can make up for the shortcomings of the above PRI values and modulation type, and eliminate the influence of parameter overlap and data noise on online deinterleaving and accurate identification. Therefore, we expect to automatically reconstruct the PRI pattern from the actual pulse train containing various data noises.

PRI pattern reconstruction refers to extracting the high-dimensional timing feature from the intercepted pulse train without prior information. Its output has wide adaptability for signal-processing tasks such as deinterleaving [1], [4], [5] and recognition [2], [7], [8], [10], [11]. In the field of pulse deinterleaving, common methods [4]–[6] exploit discrete PRI values obtained mainly by statistics to deinterleave pulse trains. However, the statistical methods can only adapt to simple PRI signals and have poor robustness to parameter overlap. In [1], an automata model deinterleaves pulse trains online and in parallel through fine modeling of radar PRI patterns. Following this model, the online and parallel structure of the automata improves the deinterleaving efficiency, but the needed PRI pattern is hard to obtain. PRI pattern reconstruction proposed in this article could act as the previous operation to provide *a priori* PRI pattern for the automata. Besides, intrapulse and interpulse modulations [7], [8], [10]–[14] support the recognition of radar working modes and types. However, the research on interpulse modulation mainly lies in the recognition of several modulation types with little research on complete modulation. The PRI pattern, including the values, ordering, and switching laws of PRIs, is a higher-dimensional feature and can widely reflect the information of emitter's type, working mode, platform, and so on. Therefore, PRI pattern reconstruction can provide essential support for various recognition problems of radar.

The high-dimensional characteristics of PRI pattern make it more difficult to reconstruct, and there are few studies regarding the same. One of the typical and latest methods [15] introduces frequent items and expands them to reconstruct the PRI pattern. Although such a method realizes the reconstruction of PRI patterns, it is impractical since it sets many subjective parameters and does not apply to complex PRI modulation types. In addition, neural networks have been extensively adopted to learn the PRI pattern from intercepted pulse train. Patterns stored in the neural network have achieved excellent performance in deinterleaving [16] and identifying [17]. However, the neural-network methods require considerable time and workforce to label the data, and the obtained PRI pattern is not interpretable.

In this article, we propose a PRI pattern reconstruction method based on semantic coding. A coding complexity is initially defined as the sum of the coding complexity of dictionary and that of encoding the intercepted pulse train based on the dictionary. Then, a semantic coding model is established and the first-order DTOA set of the pulse train is encoded by distinguishing single pulses and pulse groups. The coding model is optimized to minimize the coding complexity, and finally, the PRI pattern is obtained from the optimized dictionary set. This method can process pulse trains automatically and involve relatively few subjective parameters to regulate the entire coding process, which shows increased applicability and usefulness.

The rest of this article is organized as follows. Section II formulates the PRI pattern reconstruction problem. Section III establishes the coding model and defines the coding complexity. In Section IV, rules for optimizing the coding are made. Numerous simulations are conducted in Section V. The results are provided together with analyses. Finally, Section VI concludes the article.

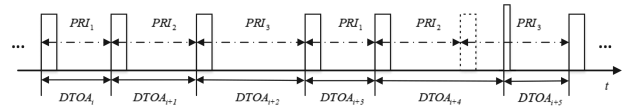


Fig. 1. Schematic diagram of pulse train.

II. FORMULATION OF PRI PATTERN ANALYSIS PROBLEM

Time-of-arrival (TOA) is one of the pulse description words directly measured from the intercepted pulse train. DTOA set can replace TOA sequence to represent pulse train with the complete information of TOA and less redundancy. Therefore, it is convenient for PRI pattern reconstruction. The pulse train in the following refers to DTOA sequence.

A. Concept and Meaning of High-Dimensional Repetition Features

As mentioned above, PRI pattern reconstruction aims to obtain a combination of multiple successive PRIs that can reflect the temporal structure stored in the intercepted pulse train. For example, we may intercept such a complete DTOA set as $\{a, b, c, a, b, c, \ldots, a, b, c\}$ under stagger PRIs $[a, b, c]$. If the number of repetitions is n, it can be described as $[a, b, c]^n$ and the repeated unit $[a, b, c]$ is targeted PRI pattern, i.e.,

$$\{a, b, c, a, b, c, \ldots, a, b, c\} \rightarrow [a, b, c]^n. \qquad (1)$$

Intuitively, this representation of high-dimensional feature is more compact and accurate than other representation methods (e.g., $[a, b, c, a, b, c]^{\frac{n}{2}}$ and $\{a, b, c\}$). Such an expression is often applied in compression coding issues.

The high-dimensional feature is the timing structure of the intercepted pulse train. Compared with the typical PRI set obtained through conventional statistical methods, the PRI pattern can distinguish more details in applications such as pulse train analysis and emitter recognition. For instance, when pulse trains of $\{a, b, c, a, b, c, \ldots, a, b, c\}$ and $\{a, c, b, a, c, b, \ldots, a, c, b\}$ from two radars are intercepted, the typical PRI value set $\{a, b, c\}$ easily misjudges the two radars as one, while the reconstructed PRI patterns $[a, b, c]$ and $[a, c, b]$ with high dimensions can directly distinguish the two radars. Furthermore, the high-dimensional temporal characteristic can act as valuable information for deinterleaving complex signals with complex PRI modulation [1] and pulse group recognition [2] of multifunction radars.

B. Model of Intercepted Pulse Train

The PRI pattern to be reconstructed is retained in the DTOA sequence of intercepted radar signals. If pulse trains are as pure as (1), PRI patterns can be reconstructed intuitively. However, this is not as simple as it seems. Data noises such as missing and interferential pulses cause the directly computed DTOAs to vary from true PRIs. Fig. 1 illustrates the DTOAs of intercepted radar pulse train and PRIs of radar transmitted pulses. In this figure, solid rectangles represent

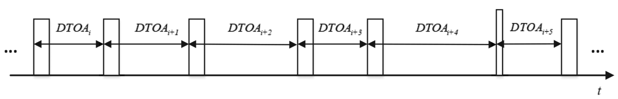


Fig. 2. Schematic diagram of intercepted pulse train.

pulses of the radar with third-order stagger PRIs, while dotted rectangles represent missing radar pulses caused by antenna scanning, signal propagation, and other factors. Furthermore, interferential pulses, involving pulses from other emitters or noises, are represented by higher solid rectangles.

Fig. 2 shows the real situation of intercepted pulse train in Fig. 1 without indicating missing or interferential pulses. Formula (2) is the digital representation of the intercepted pulse train. Without prior knowledge, the PRI pattern cannot be directly obtained

$$\{DTOA_1, DTOA_2 \cdots, DTOA_i, \ldots, DTOA_{\text{end}}\}. \quad (2)$$

The DTOA set is very lengthy with a large amount of pulses in the intercepted pulse train. We hope to express the pulse train in an intuitive form as in (1). If a simple symbol can represent the PRI pattern, i.e., $\boldsymbol{a} = [a_1, a_2, a_3]$ (a_i is the repetition value), the complete pulse train $\boldsymbol{y}$ can be expressed as (3). While in Fig. 2, the intercepted pulse train with noises is incomplete and can only be represented as a mathematical (4)

$$\boldsymbol{y} = |a_1, a_2, a_3|a_1, a_2, a_3| \cdots = \boldsymbol{aa} \cdots \boldsymbol{a} \quad (3)$$

$$\boldsymbol{y} = a_1, a_2, a_3, a_1, \Delta_1, \Delta_2, \cdots = \boldsymbol{a}a_1\Delta_1\Delta_2 \cdots \quad (4)$$

where Δ indicates a DTOA that is not equal to any PRI value. In the absence of prior information, the pulse train can only be represented by the left part of (4). In that case, each pulse must be coded individually, resulting in an extremely long coding length of the pulse train. While the pulse train is compressed to the right form of (4), the coding structure is simpler, and we can get the existing PRI pattern intuitively. It is a compression coding process from left to right in (4).

It should be mentioned that this article does not consider errors in the theoretical portion to clarify the model. Still, measurement errors in simulation section are comprehensively considered to simulate the real scene accurately.

As indicated from the above model of an intercepted radar pulse train, a compact and efficient repeating pattern with finite order is implied in the lengthy DTOA sequence. Once the pulse train is compressed properly, the PRI pattern could be reconstructed intuitively, and its redundancy can be greatly simplified.

III. SEMANTIC CODING MODEL OF PULSE TRAIN

Compression coding is a technique for removing redundancy in both space and time, and it can make the data's structure clearer by composing frequently-occurring pulse groups into a dictionary set. As a long sentence of radar communicating with surroundings, the pulse train could also be expressed by coding. If pulse groups with frequent occurrence in pulse train are encoded as dictionary elements and represented by simple symbols, the redundancy of the radar pulse train can be significantly decreased. The radar PRI pattern happens to be the most structural and frequent pulse group in the pulse train. According to Occam's razor principle [18] (also known as simplicity principle), when the dictionary set can be well geared into the PRI pattern, the coding of the pulse train should exhibit minimal complexity. In other words, the process of compression coding is just that of PRI pattern reconstruction. This process was termed semantic coding to distinguish common pragmatic coding. In another finding of authors, the semantic model is utilized to analyze the structure of multifunctional radar pulse trains in basic scenarios [19]. The semantic coding model for conventional radar pulse trains with missing and interferential pulses will be presented in this section.

The semantic coding model of the intercepted pulse train can fall into two phases depending on the content of data compression: 1) Building a coding dictionary; and 2) encoding the pulse train based on the dictionary. The model is expressed as follows:

$$\Psi = \{\mathbf{D}, S(\boldsymbol{y}; \mathbf{D})\} \quad (5)$$

where $\mathbf{D}$ denotes the dictionary set.

$$\mathbf{D} = \left\{ \begin{array}{c} \boldsymbol{a} = a_1, a_2, \ldots, a_{\text{end}} \left\{ \begin{array}{c} \underline{a_1 + a_2}, a_3, \ldots, a_{\text{end}} \\ \underline{a_1 + a_2 + a_3}, \ldots, a_{\text{end}} \\ \vdots \\ a_1, \ldots, \underline{a_{\text{end}-1} + a_{\text{end}}} \end{array} \right\} \\ \boldsymbol{b} = a_2, a_3, \ldots, a_{\text{end}} \\ \boldsymbol{c} = a_1, \ldots, a_{\text{end}}, a_{\text{end}+1} \\ \vdots \end{array} \right\}. \quad (6)$$

Assuming that the intercepted pulse train is complete and noise free, the sole element in its encoding dictionary set should be the PRI pattern according to the simplicity principle. However, due to missing and interferential pulses, the encoding dictionary set also covers partial and extended pulse groups of the radar's PRI pattern, as indicated in (6). To introduce more single pulses to form pulse groups, the semantic coding model encodes incomplete pulse groups with missing pulses on the base of there complete structure. In other words, after a certain pulse in the complete pulse group is lost, a subgroup is formed, as shown in Fig. 3. If an n-order pulse group is coded as $\boldsymbol{a}$, n-1 bits are added after $\boldsymbol{a}$ to indicate whether the n-1 internal pulses are lost (0 means lost, 1 means not lost). For a three-order PRI pulse group demonstrated in Fig. 3, there are three distinct combinations after adding two supplementary bits to $\boldsymbol{a}$ (the right part of Fig. 3). Three subgroups and one complete group are all represented by the identical element $\boldsymbol{a}$ in the dictionary set, i.e., $\boldsymbol{a} \Rightarrow \{\boldsymbol{a}, \boldsymbol{a}\underline{01}, \boldsymbol{a}\underline{10}, \boldsymbol{a}\underline{00}\}$.

In (5), $S(\boldsymbol{y}; \mathbf{D})$ indicates the coding of the pulse train based on the dictionary set $\mathbf{D}$, e.g., the left form of (4) is expressed as a simpler right form according to the dictionary set (6).

The coding complexity criteria [20] is used to assess the complexity of encoding a pulse train in this work. It consists

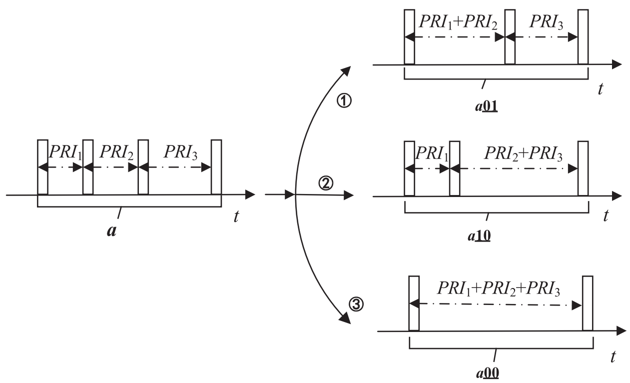


Fig. 3. Missing pulse coding rules.

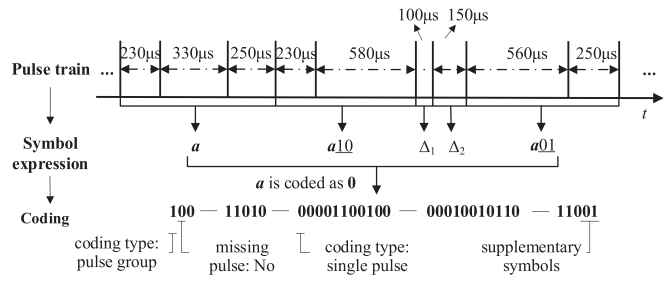


Fig. 4. Semantic coding of pulse train.

of two parts: 1) the complexity of forming a dictionary set; and 2) that of encoding the pulse train based on dictionary, as expressed below

$$c = c_D + c_S \tag{7}$$

where the complexity c_D of encoding the dictionary depends on the number of PRI values contained in it, and the calculation formula is

$$c_D = \left(\sum_{j=1}^{J} I_j\right) \times c_0 \tag{8}$$

where J represents the number of elements in the dictionary set; I_j is the number of DTOAs covered in the jth dictionary element; c_0 denotes the code length of a single DTOA, and

$$c_0 = \left\lceil \log_2 \left(DTOA_{\max}/\delta\right)\right\rceil + 1 \tag{9}$$

where $\lceil a \rceil$ means to take the smallest integer not less than a; δ is the quantization unit of the pulse interval, which is determined by the resolution requirement of DTOA; $DTOA_{\max}$ is the maximal setting value of the DTOA. When the pulse interval exceeds this value, the pulse train is split into two segments to encode in case of DTOA with very high order. "1" represents the coding type prefix symbol of the DTOA code hereafter to distinguish the single pulses and pulse groups in the coded pulse train (single pulse is 0, coded pulse group is 1).

The coding complexity c_S of pulse train $\boldsymbol{y}$ based on the dictionary set $\mathbf{D}$ is

$$c_S = \left(\sum_{j=1}^{J} N_j\right) \times \left(\sum_{j=1}^{J} \frac{N_j}{N_{\text{sum}}} \log_2(N_j/N_{\text{sum}}) + 2\right) + \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij} \cdot b_j + N_{res} \times c_0 \tag{10}$$

where

N_j represents the frequency of the jth dictionary element in the pulse train $\boldsymbol{y}$. The more frequent the element, the shorter its code length will be;

$N_{\text{sum}} = \sum_{j=1}^{J} N_j$ refers to the total occurrence times of all dictionary elements in the pulse train;

$\sum_{j=1}^{J} (N_j/N_{\text{sum}}) \log_2(N_j/N_{\text{sum}})$ is the average code length of a pulse group symbol [21];

"2" can be expressed as "1+1," respectively, corresponding to the coding type prefix symbol (to distinguish the pulse group and single pulse) as well as the missing pulse indication symbol (to distinguish whether it is a group with missing pulses);

the average code length of pulse groups multiplied by the total number of pulse groups is the coding complexity of all pulse groups in the pulse train;

n_{ij} is the frequency of the ith pattern in the jth dictionary element appearing in the pulse train;

b_j is the number of supplementary symbols in the ith dictionary element;

N_{res} denotes the number of uncoded pulses which need to be encoded in the form of single pulses in the pulse train. N_{res} is multiplied by the code length c_0 of a single pulse to determine the code length of remaining pulses;

by adding the code length of pulse groups and that of single pulses, the complexity c_S of encoding the pulse train $\boldsymbol{y}$ based on the dictionary set $\mathbf{D}$ is determined.

Subsequently, the sum of c_S and coding complexity c_D of dictionary set is the total complexity c of the intercepted pulse train, corresponding to (7).

Here is a toy example to illustrate the above radar pulse coding rules. Set the maximal allowable DTOA is 1 ms and the quantization unit length is 1 μs, then, the code length of a single DTOA is $c_0 = \lceil \log_2(1\,\text{ms}\ /1\,\mu\text{s})\rceil + 1 = 11$. It is assumed that the dictionary set only contains a dictionary element $\boldsymbol{a} = [230, 330, 250]$ with its three subgroups: $\boldsymbol{a}\underline{01} = [230, 580]$, $\boldsymbol{a}\underline{10} = [560, 250]$, and $\boldsymbol{a}\underline{00} = [810]$. The dictionary set is encoded as $\boldsymbol{a} \Rightarrow 0$. Then, a hypothetical pulse train with nine DTOAs $\{230, 330, 250, 230, 580, 100, 150, 560, 250\}$ is encoded, and its symbol expression is $\boldsymbol{aa}\underline{01}\Delta_1\Delta_2\boldsymbol{a}\underline{10}$ including three pulse groups and two single pulses. Its encoding process is shown in Fig. 4. Among them, two single pulses ($\Delta_1 = 100\,\mu$s , $\Delta_2 = 150\,\mu$s) are encoded as 00001100100 and 00010010110, respectively, and the first "0" indicates the single pulse type. In addition, the complete pulse group $\boldsymbol{a}$ is encoded as "100" with the first bit "1" indicating that the coding type is pulse group, and the second bit "0" indicating no missing pulse. Other incomplete pulse groups have two supplementary symbols, indicating the position of pulse

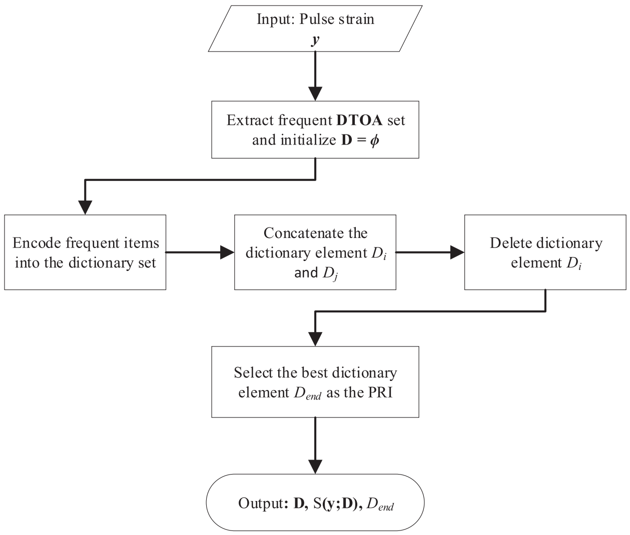


Fig. 5. Semantic coding optimization flowchart.

loss, which are 11010 and 11001, respectively. Lastly, the coding result of the pulse train is 100 | 11010 | 00001100100 | 00010010110 | 11001, and the original pulse train can be uniquely determined by such coding. Before encoding the pulse train with dictionary, its complexity is $9 \times 11 = \mathbf{99}$, while its total complexity turns out to be **68** (c_S is 35, and c_D is 33) based on dictionary. Although establishing dictionary set increases a small amount of complexity, the three repetitions of the dictionary element still reduce the overall complexity. As suggested from the above example, the semantic coding model can remove the redundancy of the pulse train to reconstruct its PRI pattern. Furthermore, the longer the intercepted pulse train, the better the compression effect will be.

IV. CODING MODEL OPTIMIZATION FOR PRI PATTERN RESTRUCTION

Following Occam's razor principle, the dictionary set should be gradually established to minimize the coding complexity. PRI pattern can be obtained in the final dictionary. Specific optimization rules are recommended in this section as shown in Fig. 5.

As shown in Fig. 5, PRI pattern reconstruction from the intercepted pulse train requires the following five processes.

1) Extract frequent items to provide processing objects and eliminate the influence of interferential pulses.
2) Frequent items are encoded into the dictionary set, and the dictionary set with single-order DTOA elements is established.
3) The elements of the dictionary set are concatenated to extend the element order.
4) Residual elements in the dictionary set are deleted to keep the dictionary simplest.
5) PRI pattern extraction.

Specifically, the steps 2), 3), and 4) can be collectively referred to as the update process of the dictionary set, and all should calculate the total coding complexity change Δc as the operating criterion.

A. Frequent DTOA Set Extraction

Frequent item refers to the DTOA whose frequency exceeds the preset threshold [19]. The common frequent items involve radar PRI values and high-order DTOA values generated by missing pulses.

For a DTOA set of the pulse train $\boldsymbol{y}$, the DTOAs in it are clustered at a certain error limit and a frequent itemset can be obtained as follows:

$$\left\{DTOA_i, M_i, \left\{h_m^{(i)}\right\}_{m=1,\cdots,M_i}, \left\{t_m^{(i)}\right\}_{m=1,\cdots,M_i}\right\}_{i=1,\cdots,I} \quad (11)$$

where

I represents the number of frequent DTOA items, and each subscript i corresponds to a frequent item;

$DTOA_i$ is the differential time-of-arrival corresponding to the ith frequent item appearing M_i ($M_i > M_0$) times in the pulse train. The value is obtained by numerical clustering and calculating the average value;

the head pulse sequence numbers of pulse pairs that combined into the frequent item $DTOA_i$ constitute the set $\{h_m^{(i)}\}_{m=1,\cdots,M_i}$, and the tail pulse sequence numbers constitute the set $\{t_m^{(i)}\}_{m=1,\cdots,M_i}$ [6].

To automate the coding optimization process and simplify the setting of subjective parameters, this article sets the proportional parameter k to make the threshold M_0 adaptive, i.e., $M_0 = k \cdot M_{\max}$ ($M_{\max}$ denotes the frequency of the most frequent DTOA).

Accordingly, high-order DTOAs (the difference of two pulse sequence numbers $\Delta > 1$) could also be obtained, which can help to eliminate the effect of interferential pulses. However, they lead the head and tail pulse sets to cover numerous repetitive pulses to form DTOAs of different orders. For example, $h_m^{(i)} = 1$ and $t_m^{(i)} = 2$ form $DTOA_i$ (one-order), $h_m^{(i)} = 1$ and $t_m^{(i)} = 3$ form $DTOA_j$ (two-order), and the same head pulse forms different DTOAs. $DTOA_j$ is the high-order DTOA value across one pulse. This high-order DTOA across pulses will upregulate the number of imcomplete pulse groups caused by missing pulses and affect the encoding. To control the DTOAs caused by missing pulses from being supplemented by the DTOAs across pulses, the frequent item set should satisfy

$$\prod_{\substack{i,j=1,\cdots,I \\ i \neq j}} \prod_{\substack{m=1,\cdots,M_i \\ n=1,\cdots,M_j}} (h_m^{(i)} - h_n^{(j)})(t_m^{(i)} - t_n^{(j)}) \neq 0 \quad (12)$$

where I denotes the total number of frequent items; M_i and N_j are the total number of pulse pairs of the ith and jth frequent items, respectively.

In other words, a principle is followed that two frequent DTOA items cannot have the identical head or tail pulse according to (12). Then, the repetitive pulses in the small DTOA set are retained, and those in the large DTOA set are removed.

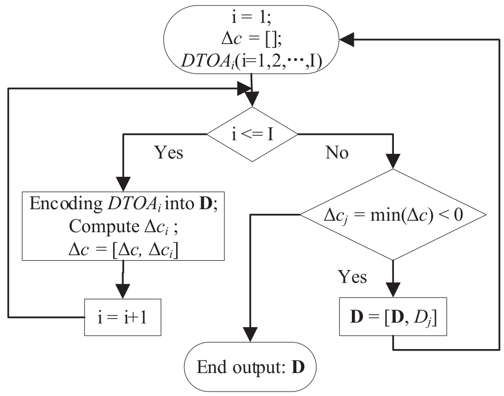


Fig. 6. Flowchart of encoding frequent items into dictionary.

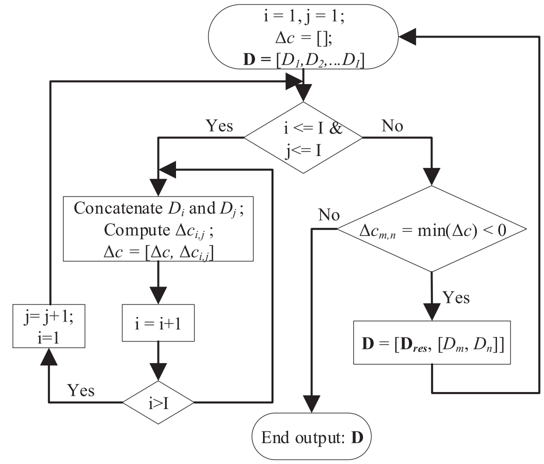


Fig. 7. Flowchart of concatenating dictionary elements.

Extracting frequent items can improve the efficiency of constructing the dictionary set by filtering processing objects. Moreover, DTOAs produced by interferential pulses are extremely difficult to cluster into frequent items, thus, eliminating the effect of the interferential pulse on encoding.

B. Updating of Dictionary Set

The encoding dictionary set records structurally strong and frequently-occurring pulse combinations in the pulse train. During the model optimization, a principle of minimizing the total coding complexity is followed and the dictionary set is updated in the order of elements addition, concatenation, and deletion as follows:

1) Encode every frequent item into the dictionary set and the complexity change Δc_i is computed. In each iteration, only the frequent item with the greatest complexity reduction is encoded, as expressed below

$$\begin{aligned} i_{\max} &= \arg\max_i \Delta c_i \\ \mathbf{D}_{\text{new}} &= \left\{\mathbf{D}_{\text{old}}, DTOA_{i_{\max}}\right\}. \end{aligned} \tag{13}$$

Repeat this process until the complexity is not decreased. A flowchart of encoding frequent items to the dictionary is shown in Fig. 6.

2) Concatenate the dictionary elements to extend them. Any two items D_i and D_j are traversed and concatenated according to the identical head and tail pulses. The total complexity reduction $\Delta c_{i,j}$ is computed after the concatenation. Finally, the two dictionary elements with the maximal complexity decrease are concatenated, referring to (14). It is noteworthy that the leftover pulses of the two dictionary items are retained in the dictionary set for subsequent connections

$$\begin{aligned} i_{\max}, j_{\max} &= \arg\max_{i,j} \Delta c_{i,j} \\ \mathbf{D}_{\text{new}} &= \{\mathbf{D}_{res}, [D_{i_{\max}}, D_{j_{\max}}]\} \end{aligned} \tag{14}$$

where $\mathbf{D}_{res}$ indicates the remaining dictionary set after removing the concatenated dictionary items. Such a procedure is repeated till the complexity stops decreasing as shown in Fig. 7.

3) Delete residual dictionary elements. There are many residual items in the dictionary set after concatenation. Traverse the dictionary element D_i to calculate the reduced complexity Δc_i after it is deleted. Lastly, the element with the maximal complexity reduction is deleted from the dictionary set. This operation is similar to encode frequent items into a dictionary and is designed to make the dictionary set more compact.

C. PRI Reconstruction

After establishing the dictionary set with minimal complexity, the PRI pattern should be selected from the dictionary set. According to Occam's razor principle[18], the dictionary element with the simplest structure and the most frequent occurrence in the dictionary set should match the PRI pattern, represented as

$$\begin{aligned} j_{\max} &= \arg\max_j \sum_i^{I_j} M_{i,j} \\ D_{\text{end}} &= D_{j_{\max}} \end{aligned} \tag{15}$$

where $M_{i,j}$ denotes the frequency of the ith subgroup in the jth dictionary element; I_j indicates the number of subgroups in the jth dictionary element, including both the missing pulse patterns and the complete pattern, as seen in Fig. 3.

As may be seen, relatively few subjective variables are involved in the above processes. When a pulse train is entered, the semantic coding model automatically compresses its coding and reconstructs the conventional radar's PRI pattern. In fact, the semantic coding theory can also be extended to radars with more complex PRI patterns, such as multifunction radar [19].

V. SIMULATION

In this section, simulations are performed to verify the performance of proposed semantic coding method for PRI pattern reconstruction. The influence of real-world factors and adaptability to complex PRI pattern are also analyzed. Due to the complexity of PRI pattern reconstruction, there are few studies that are consistent with the background of this article. The method of frequent item extraction and expansion proposed in the literature [15] (hereafter referred to as frequent term extension method) is the latest related study, and it has good results in relatively simple PRI patterns, which is very worthy of comparison.

The simulation falls into five parts: The first part introduces the setting of simulation parameters. The second part gives an example to illustrate the process of PRI pattern reconstruction. The third and fourth parts compare the performance of the proposed method (hereafter referred to as semantic coding method) and frequent term extension method [15] affected by real-world factors. The fifth part explores the adaptability of this method to more complex PRI modulation types and higher-order PRI patterns.

A. Parameter Settings

When exploring the influence of missing and interferential pulses, two distinct stagger PRI patterns are employed, i.e., a basic pattern [230,330,250] and a complex pattern [300,300,330,630]. The simpler one is third-order stagger repetition. If the complex PRI pattern is expressed as $[PRI_1, PRI_2, PRI_3, PRI_4]$, it exhibits the following properties:

$$PRI_1 = PRI_2 \tag{16}$$

$$PRI_2 + PRI_3 = PRI_4. \tag{17}$$

When frequent items are obtained through clustering, the recurrence of the identical repetition values causes two parts of pulses to fall into the same category, according to (16). In that case, these pulses should be distinguished by the pulses before and after. Since the second and third repetition values total equal the fourth repetition value according to (17), the high-order repetition value produced by missing pulses will be confused with the actual fourth repetition value, thereby complicating analysis.

PRI modulations of stagger, slide, dwell, and switch are set as follows when exploring the adaptability of proposed method.

1) Set the measurement error of TOA to comply with the normal distribution (error $\sim N(0, 0.2)$).
2) Maximal $DTOA_{\max}$ is 2 ms.
3) The time of the quantization unit δ is 1 μs.
4) The ratio threshold of frequent items is 0.1.

Reconstruction accuracy is utilized to measure the reconstructing ability, which is defined as

$$\text{Reconstruction accuracy} = \frac{\text{correct times}}{\text{simulation times}}. \tag{18}$$

Correct times add 1 when the reconstructed result matches the true PRI pattern in both values and their connection structure.

B. Verification of PRI Pattern Reconstruction Ability

The actual repetition pattern of the radar pulse train is [230, 330, 250] at this part. To present the semantic coding process more accurately, the number of missing pulses and interferential pulses is set to 0. The total number of intercepted pulses is 100. Table I lists the iterative changes made to the encoding dictionary set. The results maintain two decimal places.

Without data noise, the frequent items are extracted as three PRI values {230, 250, 330}, and they all appear 33 times in the pulse train. In the second column of Table I, frequent items are first encoded into the dictionary set one by one according to Fig. 6 and (13), and after three iterations, a dictionary set containing three frequent items gets formed. Subsequently, the dictionary elements in the dictionary set are concatenated according to Fig. 7 and (14). [330.00] is first connected to [250.00], and then, [230.00] is connected in front of them. A dictionary set with only one element is developed. Lastly, no residual dictionary element should be deleted. It is still emphasized that each action needs to reduce the coding complexity. The final dictionary element ([230.00, 330.00, 250.00]) is consistent with the radar PRI pattern.

C. Influence of Missing Pulse Rate and Pulse Number

The missing pulse rate is set to 0–0.7, the number of intercepted pulses is 100 and 300, and no interferential pulse is set. The semantic coding method and the frequent item extension method [15] are tested 100 times to calculate the reconstruction accuracy as shown in Fig. 8.

Fig. 8(a) shows that when the missing pulse rate does not exceed 50%, the semantic coding method can achieve an analysis accuracy approaching 100% even if only 100 pulses are intercepted. With the continuous increase of the missing pulse rate, the accuracies of the two methods drop. However, the semantic coding method constantly outperforms the frequent item extension method under the same conditions. Under a pulse missing rate of 0.7 and 100 intercepted pulses, the frequent term extension method loses its reconstruction ability, whereas the semantic coding method can still achieve an analysis accuracy of 55%.

It can be seen from the results in Fig. 8(b) that under a complex radar PRI pattern, the reconstruction capabilities of the two methods significantly decrease. However, the semantic coding method outperforms the frequent item extension method with a maximum lead of 40%.

Based on the above results, the proposed method is superior to the frequent item extension method under the influence of missing pulses. Especially in the case of less intercepted pulses and more complicated PRI pattern, the comparison is better.

D. Influence of Interferential Pulse Ratio and Pulse Number

100 and 300 effectively intercepted pulses (pulse train that removes interferential pulses) are set respectively in this section. The ratio of the interferential pulse is set to 0–1, the missing pulse rate is locked at 0.5, and the arrival time of interferential pulses is uniformly distributed. The reconstruction accuracies after 100 repeats are presented in Fig. 9.

TABLE I
Iterative Establishment of Dictionary

Iteration number	Encode frequent items →	Concatenate dictionary elements →	Delete dictionary elements
1	**E_1: [230.00], F_1 = 33**	**E_1: [230.00], F_1 = 33** **E_2: [250.00], F_2 = 33** **E_3: [330.00], F_3 = 33**	**End**
2	**E_1: [230.00], F_1 = 33** **E_2: [250.00], F_2 = 33**	**E_1: [230.00], F_1 = 33** **E_2: [330.00,250.00], F_2 = 33**	
3	**E_1: [230.00], F_1 = 33** **E_2: [250.00], F_2 = 33** **E_3: [330.00], F_3 = 33**	**E_1: [230.00,330.00,250.00], F_1 =33**	
4	**End**	**End**	

E_i represents the *i*th dictionary element.
F_i represents occurrence times of the *i* th dictionary element in pulse train.

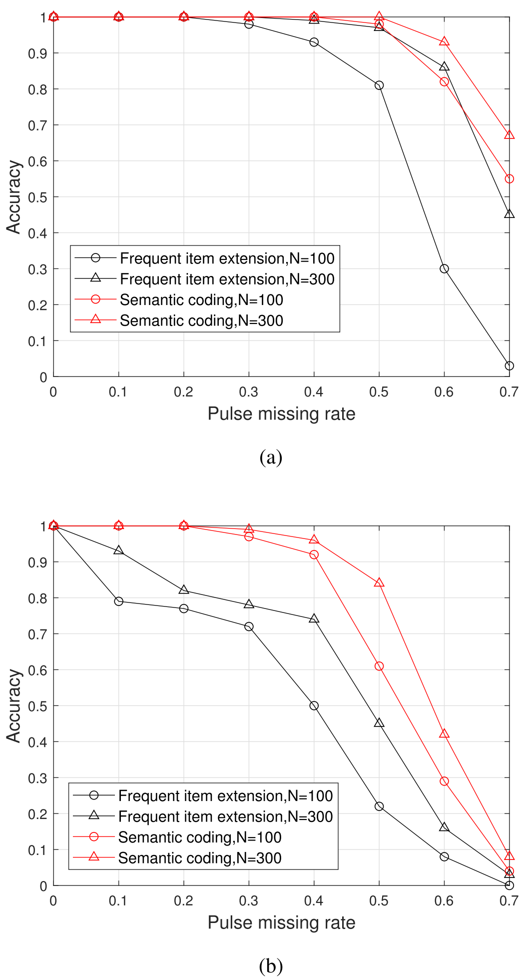


Fig. 8. Pattern reconstruction accuracies in case of different pulse missing rates. (a) PRI pattern [230, 330, 250]. (b) PRI pattern [300, 300, 330, 630].

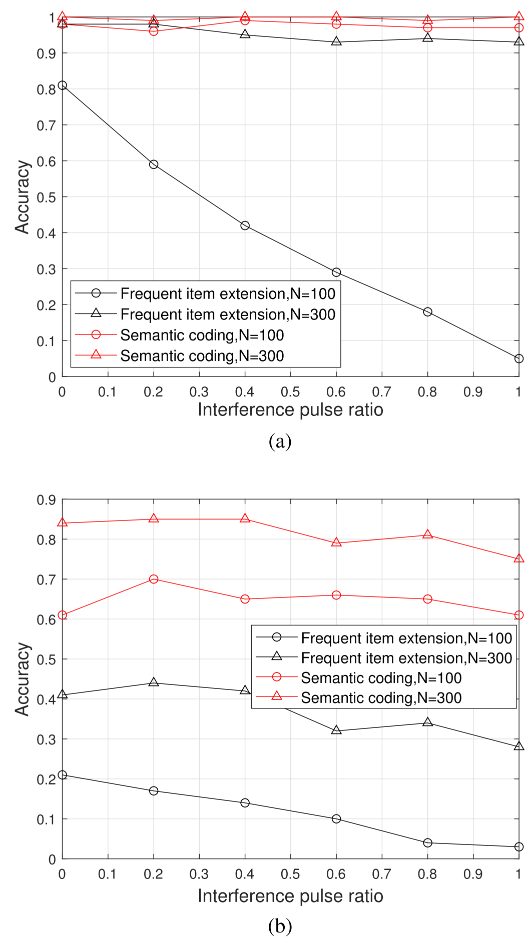


Fig. 9. Pattern reconstruction accuracies in case of different interferential pulse ratios. (a) PRI pattern [230, 330, 250]. (b) PRI pattern [300, 300, 330, 630].

According to Fig. 9, whether the PRI pattern is simple or complex, and whether the number of effectively intercepted pulses is more or less, the semantic coding method does not show sensitivity to interferential pulses. As the interferential pulse ratio increases from 0 to 1, the analysis accuracies constantly oscillate in a small amplitude. In contrast, when the number of intercepted pulses is 100, the reconstruction accuracies of the frequent item extension method significantly decline as the ratio of interferential pulses increases. In addition, the frequent term extension method can achieve similar results to those of the semantic encoding method only in the case of sufficient intercepted pulses (300 in this simulation) and the simple PRI.

E. Adaptability to PRI Modulation Types and PRI Values

To further demonstrate the adaptability of semantic coding method in this article, this part studies the influence of

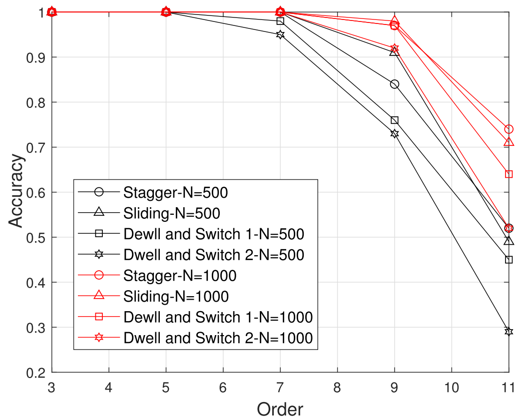


Fig. 10. Pattern reconstruction results under different orders and types.

different types and orders of PRI. In this part, no interferential pulse is set, and the missing pulse rate is 0.3. The results are shown in Fig. 10. In Dwell and Switch 1, there are 2–3 dwells and no limit on the number of switches in a cycle, such as PRI patterns of five-order [200,200,250,250,250] and six-order [200,200,250,250,300,300]; Dwell and Switch 2 means that there are only two PRI values and there is no limit to the number of dwells, such as the six-order PRI pattern [200,200,200,250,250,250].

As can be seen from Fig. 10, the semantic coding method is adaptive to more complex PRI modulations because they are periodic in pulse train. In comparison, the reconstruction effect of dwell and switch is poorer, which is due to the confusion caused by the repetition of PRI values, which is similar to the complex PRI pattern [300, 300, 330, 630] in parts C and D. The results of stagger and sliding are similar. When intercepting 500 pulses, the semantic coding method can obtain a reconstruction accuracy about 50% for the eleven-order PRI pattern. When 1000 pulses are intercepted, the reconstruction accuracies of stagger and slide modulations are more than 70%.

The semantic coding method also has good adaptability to high-order PRI patterns. When the number of intercepted pulses is 500, the reconstruction accuracy reaches more than 90% for PRI patterns with seven PRIs; when intercepting 1000 pulses, the proposed method can obtain an accuracy of more than 90% for the nine-order PRI patterns and improve about 20% for eleven-order PRI patterns. With the increase in the number of intercepted pulses, PRI pattern is more significant in the intercepted pulse train, and it is easier to reconstruct its real PRI pattern for the proposed semantic coding method.

VI. CONCLUSION

This article presents a semantic coding method to reconstruct PRI pattern from intercepted radar pulse train with data noises. In the proposed method, semantic coding theory is introduced to establish a coding dictionary set for encoding the pulse train. According to the principle of Occam's razor, we compress pulse train coding and select the most frequent element in dictionary as the PRI pattern. At last, the proposed method realizes the high-accuracy PRI pattern reconstruction in the case of complex data. Simulation results are as follows:

1) First, the semantic coding method is adaptive to various periodic PRI patterns, including stagger, slide, dwell, and switch, etc.
2) Second, when the missing pulse rate does not exceed 50%, it has little effect on the reconstruction accuracy. Otherwise, the reconstruction accuracy decreases seriously. The reconstruction accuracy of simple PRI modulation is close to 100% when missing pulse rate is 0.5.
3) Third, the semantic coding method is robust to interferential pulses, thereby improving the tolerance of PRI pattern reconstruction in complex electromagnetic environment.
4) Finally, the semantic coding method is also adaptable to high-order PRI patterns, but it needs to intercept sufficient pulses. In order to make the reconstruction accuracy reach 90%, 500 pulses need to be intercepted for seventh-order PRI pattern and 1000 pulses are needed to be intercepted for ninth-order PRI patterns.

REFERENCES

[1] Z.-M. Liu, "Online pulse deinterleaving with finite automata," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 2, pp. 1139–1147, Apr. 2020, doi: 10.1109/TAES.2019.2925447.

[2] S.-Q. Kang et al., "Sequential extraction and recognition of pulse group structure for multi-function radar," *IET Radar Sonar Navig*, vol. 16, no. 4, pp. 678–691, 2022, doi: 10.1049/rsn2.12212.

[3] R. G. Wiley, *ELINT: The Interception and Analysis of Radar Signals*. Norwood, MA, USA: Artech House, 2006.

[4] H. Mardia, "New techniques for the deinterleaving of repetitive sequences," *IEE Proc. F. (Radar Signal Process.)*, vol. 136, no. 4, pp. 149–154, 1989.

[5] D. Milojevic and B. Popovic, "Improved algorithm for the deinterleaving of radar pulses," *IEE Proc. F. (Radar Signal Process.)*, vol. 139, no. 1, pp. 98–104, 1992.

[6] K. Nishiguchi and M. Kobayashi, "Improved algorithm for estimating pulse repetition intervals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 407–421, Apr. 2000.

[7] K. Gençol, A. Nuray, and K. Ali, "A wavelet-based feature set for recognizing pulse repetition interval modulation patterns," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 24, no. 4, pp. 3078–3090, 2016.

[8] Y. Liu and Q. Zhang, "An improved algorithm for PRI modulation recognition," in *Proc. IEEE Int. Conf. Signal Process. Commun. Comput.*, 2017, pp. 1–5, doi: 10.1109/ICSPCC.2017.8242587.

[9] M. S. Greco, F. Gini, P. Stinco, and K. Bell, "Cognitive radars: On the road to reality: Progress thus far and possibilities for the future," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 112–125, Jul. 2018, doi: 10.1109/MSP.2018.2822847.

[10] J. Matuszewski, "The radar signature in recognition system database," in *Proc. Conf. Proc. 19th Int. Conf. Microw. Radar Wireless Commun.*, 2012, vol. 2, pp. 617–622.

[11] J. Matuszewski, "The analysis of modern radar signals parameters in electronic intelligence system," in *Proc. 13th Int. Conf. Modern Problems Radio Eng. Telecommun. Comput. Sci.*, Lviv-Slavske, Ukraine, 2016, Art. no. 7452040, pp. 298–302.

[12] M. Zhang et al., "Neural networks for radar waveform recognition," *Symmetry*, vol. 9, no. 5, pp. 75, 2017, doi: 10.3390/sym9050075.

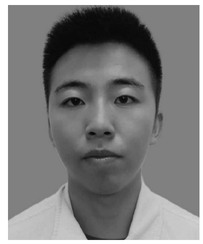
[13] Z. Zhou et al., “Automatic radar waveform recognition based on deep convolutional denoising auto-encoders,” *Circuits, Syst., Signal Process.*, vol. 37, no. 9, pp. 4034–4048, Jan. 2018, doi: 10.1007/s00034-018-0757-0.
[14] M. Zhu, Y. Li, and S. Wang, “Model-based time series clustering and interpulse modulation parameter estimation of multifunction radar pulse sequences,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 6, pp. 3673–3690, Dec. 2021, doi: 10.1109/TAES.2021.3082660.
[15] Z. M. Liu, S. Q. Kang, and X. M. Chai, “Automatic pulse repetition pattern reconstruction of conventional radars,” *IET Radar Sonar Navig*, vol. 15, no. 5, pp. 500–509, 2021.
[16] Z. Liu and P. S. Yu, “Classification, denoising, and deinterleaving of pulse streams with recurrent neural networks,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 4, pp. 1624–1639, Aug. 2019.
[17] Y. Li, M. Zhu, Y. Ma, and J. Yang, “Work modes recognition and boundary identification of MFR pulse sequences with a hierarchical seq2seq LSTM,” *IET Radar Sonar Navig.*, vol. 14, no. 9, pp. 1343–1353, 2020.
[18] P. D. Grunwald, *The Minimum Description Length Principle*. Cambridge, MA, USA: MIT Press, 2007.
[19] L. Zhangmeng, Y. Shuo, and K. Shiqian, “Semantic coding and model reconstruction of multifunctional radar pulse train,” *J. Radars*, vol. 10, no. 4, pp. 559–570, 2021, doi: 10.12000/JR21031.
[20] K. Sayood, *Introduction to Data Compression*, 4th ed., Burlington, MA, USA: Morgan Kaufmann, 2012.
[21] F. Zuyun, *Information theory–basic theory and application* Beijing, China: Publishing House of Electronics Industry, 2015.

Shuo Yuan was born in 1997. He received the B.E. degree in electronic engineering in 2020 from the National University of Defense Technology, Changsha, China, where he is currently working toward the M.E. degree in communication and information engineering.

His research interests include radar signal processing and machine learning.

Shi-Qian Kang was born in 1996. He received the B.E. degree in information engineering in 2018 from the National University of Defense Technology, Changsha, China, where he is currently working toward the M.E. degree in communication and information engineering.

His research interests include radar signal processing and machine learning.



Wen-Xiu Shang was born in 1998. He received the B.E. degree in information engineering in 2021 from the National University of Defense Technology, Changsha, China, where he is currently working toward the M.E. degree in communication and information engineering.

His research interests include radar signal processing.

Zhang-Meng Liu received the Ph.D. degree in statistical signal processing from the National University of Defense Technology (NUDT), Changsha, China, in 2012.

He is currently a Professor of the interdiscipline of electronics engineering and computer science, especially electronic data mining with NUDT.