Algorithm Of Radar Pattern Clustering In Passive Radar Systems

Oleg V. Glukhov
dept. Radio Systems
National Research University MPEI
Moscow, Russia
glukhov.oleg.v@yandex.ru

Natalia V. Masalkova
dept. Radio Systems
National Research University MPEI
Moscow, Russia
natasha.masalckova@yandex.ru

Elena V. Silaeva
dept. Radio Systems
National Research University MPEI
Moscow, Russia
e.v.zakharova_work@mail.ru

Nikita I. Petukhov
dept. Radio Systems
National Research University MPEI
Moscow, Russia
nekitpetuhov@yandex.ru

Alexander A. Chugunov
dept. Radio Systems
National Research University MPEI
Moscow, Russia
iamchugunov@gmail.com

Ilya V. Korogodin
dept. Radio Systems
National Research University MPEI
Moscow, Russia
korogodin@srns.ru

***Abstract*—This article discusses the application of signal pattern clustering algorithms to increase the efficiency of passive radar complexes at the stage of detection of marks from targets. The solution of such a problem will allow in the future to select more effective algorithms for secondary (trajectory) processing of targets. Implementation of effective clustering of marks will allow in the future to adequately form the trajectories of target groups, which will lead to energy gain and reduce the number of false runs by an order of magnitude. To solve the problem of radar pattern clustering, we developed the cDBSCAN clustering algorithm, a density-based spatial clustering algorithm with the presence of noise, using a cascade of heuristic rules. The results on the resulting simulation model showed that the algorithm correctly detects patterns in the sample with high accuracy, more than 98%. At the same time, the value of false detection of patterns is about 2%.**

Keywords—AL, ML, machine learning, cDBSCAN, DBSCAN, clustering, algorithm, radar

I. Introduction

Target recognition in airspace is one of the priority areas of radar. Radar information processing is an important range of tasks solved with the help of radar stations (radars) or complexes. The purpose of processing - to prepare for delivery in the required form complete, reliable and up-to-date information for the user about the state of air situation, appearance and location of air objects, parameters of their movement, possible variants of development of dynamics of change of air-interference situation. A typical example of radar information processing is the task of aircraft detection and estimation of their belonging to a certain class or type. The problem is similar to the well-known task of clustering and subsequent classification [1, 10], it is necessary to select from a set of single marks a few grouping centers, which correspond to the targets to be detected.

In most systems, clustering is performed using semi-empirical methods, the efficiency of which is low. This leads, on the one hand, to missing some of the marks and reducing the energy in deciding on the presence of targets, as well as multiplying marks from large targets and producing false targets. On the other hand, if the clustering threshold is too high, the reflections from different nearby targets may be combined into one cluster, resulting in missed targets and poor target trajectory accuracy. The implementation of effective mark clustering will further allow adequate formation of target group trajectories, which will lead to energy gains and reduce the number of false traces by an order of magnitude [2].

At present, the number of works, where the problem of radar pulse clustering is solved using different methods, is growing. For example, the recognition of dynamic objects in radar space using k-means and fuzzy c-means clustering algorithms on navigation parameters [3], the scale mixture model of normal distributions for classification and clustering of radar emitters [4], automatic classification method using p-value calculation network to test hypotheses about the types of emitters, where the clustering algorithm is based on a trained vector clustering method [5], density-based radar scanning clustering algorithms [6], apply and neural network apparatus, namely deep recurrent neural networks (RNN) for classification and coarse clustering of different groups of pulses hierarchically with respect to their sequential structures [7].

The aim of the work is to develop and study an algorithm for clustering the sequence of radar pulses in passive radar complexes.

In this work considers the possibility of improving the efficiency of information processing in passive radar systems by identifying signals with targets by clustering the received radio pulses with an algorithm based on DBSCAN [8, 9]. In turn, this algorithm does not require to determine the number of clusters in advance and takes into account the emissions (noise) in the radar data. It is important to note, in this paper the clustering problem is not set for single pulses, but for the processing of a sequence of pulses, called a pattern. Such a formulation complicates the classical clustering problem because it is necessary to isolate in one cluster not simple single pulses, but a sequence (pattern or complex signal) from the observed target. In this regard, in addition to the DBSCAN algorithm, heuristic rules for screening out false signals were implemented.

II. Problem Statement

Let the passive radar system receives simple signals (pulses or marks) in a discrete observation period equal to N cycles. The state vector describing a simple signal at the k-th time moment is written in the following form:

$$s_k = \begin{bmatrix} t_k \\ f_k \\ \delta t_k \\ T_k \end{bmatrix}, k = 1 \dots N, \qquad (1)$$

where t_k - time of pulse arrival, f_k - pulse carrier frequency, f_k - pulse duration, T_k - period between k -th and $(k-1)$ th pulses.

At the same time in the observed realization of pulses the following signal sequences (patterns or complex signals), described by the following state matrices, are encountered:

$$s_k^A = \begin{bmatrix} t_1^A & t_2^A & t_3^A \\ f_1^A & f_2^A & f_3^A \\ \delta t_1^A & \delta t_1^A & \delta t_1^A \\ T_1^A & T_1^A & T_1^A \end{bmatrix}, s_k^A = \begin{bmatrix} t_1^B & \dots & t_7^B \\ f_1^B & \dots & f_7^B \\ \delta t_1^B & \dots & \delta t_7^B \\ T_1^B & \dots & T_7^B \end{bmatrix}, \qquad (2)$$

where the number of columns in both matrices shows the size of these signal sequences.

The other signals in the implementation will be considered interference, and their state vector will be written as that of a single pulse:

$$s_j = \begin{bmatrix} t_j \\ f_j \\ \delta t_j \\ T_j \end{bmatrix}. \qquad (3)$$

The entire received pulse realization on an observation period equal to N is written by the following equation:

$$\boldsymbol{R}^{N\times 4} = \boldsymbol{S}^{N\times 4} + \boldsymbol{E}^{N\times 4}, \qquad (4)$$

where $\boldsymbol{S}^{N\times 4}$ is the state matrix of the whole pulse realization of dimension $N \times 4$, $\boldsymbol{E}^{N\times 4}$ is the matrix of additive white Gaussian noises with zero mean and finite dispersion of dimension $N \times 4$.

The task is to cluster, i.e. to separate into separate groups (clusters), all occurring patterns of types A and B on the background of the received pulse realization $\boldsymbol{R}^{N\times 4}$.

As a criterion of solution efficiency we will use the following metrics:

1) Pattern recognition accuracy (in %):

$$accuracy = \frac{\sum_{i=1}^{m} \frac{\hat{p}}{p}}{m}, \qquad (5)$$

where m - the number of true clusters containing the sought patterns, p - the number of true patterns m in a cluster, $\hat{p}$ - an estimate of the number of true patterns in a cluster.

2) False detection of patterns (in %):

$$false\_alarm = \frac{\sum_{i=1}^{\hat{m}} \hat{n}}{\sum_{i=1}^{\hat{m}} (\hat{p} + \hat{n})}, \qquad (6)$$

where $\hat{m}$ - estimate of the number of true clusters containing the patterns sought, $\hat{p}$ - estimate of the number of true patterns in the cluster, $\hat{n}$ - estimate of the number of incorrect patterns in the cluster.

3) Skip (in %):

$$skipping = \frac{\sum_{i=1}^{m} skip}{m}, \qquad (7)$$

where skip takes discrete values of 0 and 1 (1 - skip cluster with sought patterns, 0 - cluster with sought patterns recognized), m - number of true clusters containing sought patterns.

4) Estimation of number of true clusters $\hat{m}$.

III. Algorithm Development

The developed radar data processing cascade based on the DBSCAN algorithm, shown in Fig.1.

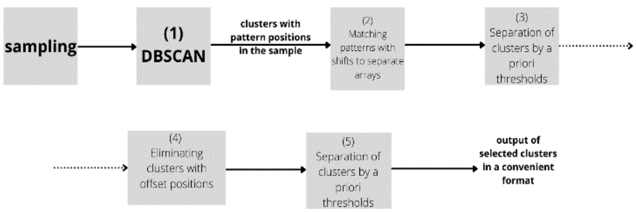

Fig. 1. cDBSCAN algorithm flowchart.

For convenience we call it cascade DBSCAN or cDBSCAN for short.

Block (1) is characterized directly by DBSCAN clustering algorithm.

The concept of the DBSCAN algorithm is to select high-density regions, which are separated from each other by low-density regions. And in order to apply this clustering algorithm correctly to solve the problem we need to define the input parameters and characteristics of the algorithm in detail. The input of the algorithm is a proximity matrix and two parameters - radius $epsilon$-neighborhood and $minpts$-minimum number of neighbors. To define $epsilon$ and $minpts$ several definitions are required. Let some symmetric distance function $\rho(x, x')$ and constants ε and m be given.

Then:

1. Let us call the region (x) for which $\forall x: \rho(x, x') \le \varepsilon$, where ε is the neighborhood of the object x.

2. A central object or a nuclear object of degree m is an object whose ε-neighborhood contains at least m objects $|E(x)| \geq m$.

3. An object p is directly densely reachable from an object q if $p \in E(q)$ and q is the root object.

4. An object p is densely reachable from object q if $\exists p_1, p_2 \dots p_n, p_1 = q, p_n = p$ such that $\forall i \in 1 \dots n - 1$: $p_i + 1$ is directly densely reachable from pi.

Following the definition of a dense domain, a point can be classified as a central point if $|E(x)| \geq m$. The central points, as the name implies, are inside the cluster. A boundary point (a reachable density point) has less than m in its $E(x)$ region but lies in the neighborhood of another central point. A noise (dropout point) is any data point that is neither a center point nor a boundary point. Reachability is not a symmetric relation because, by definition, no point can be reached from a non-main point, regardless of distance (so a non-main point can be reachable, but nothing can be reached from it). Therefore, a further notion of connectivity is needed to formally define the area of clusters found by the DBSCAN algorithm. Two points p and q are density connected if there is a point o such that both p and q are reachable from o. Density connectivity is a symmetric relation. Then the cluster satisfies two properties:

1. All points in the cluster are pairwise connected in density.

2. If a point is density reachable from some point in the cluster, it also belongs to the cluster

DBSCAN works well on dense, well separated clusters. The shape of the cluster does not matter at all. The algorithm is excellent at detecting clusters of small size. Successfully applied to a large dataset $N = 10^{6\dots8}$, and the complexity of the dataset elements does not matter. The number of elements in a cluster may vary, the number of outliers does not matter, if they are scattered over a large volume.

Before presenting the data to the clustering algorithm, they need to be normalized, or standardized. In this paper, a Z-standardization of the observation matrix parameters was chosen. Where Z-standardization is such a transformation of the data, which allows to translate the scale to Z scale, where the mean value will be equal to zero, and the standard deviation is equal to one.

The pulse arrival time parameter is not used in the clustering algorithm, because it is not informative. Therefore, the input parameters for the clustering algorithm are the following three parameters: pulse carrier frequency, period between pulses and pulse duration.

Our task was to find patterns from a sample of pulses. Therefore we need to transform the input data so that the clustering algorithm searches exactly for the set of pulses grouped into a pattern. For this purpose the observation matrix R must be modified, extending it to the dimension $[N \times (3 \min\_size\_pattern)]$. In our case, the minimum size of the pattern is equal to 3 pulses. Let us call this operation a "displacement window.

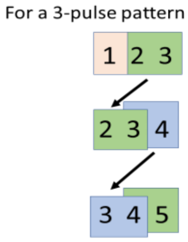

Fig. 2. Offset window operation.

The DBSCAN clustering algorithm requires the following parameters: the number of minimum neighbors minpts and search radius of neighbors epsilon.

The number of minimal neighbors minpts is chosen heuristically and should be one more than the number of features in the processed data. In our case we process three parameters, so we take the number of neighbors to be 4. Epsilon is chosen based on an ascending sorted matrix of pairwise distances with Euclidean metric.

We present the selection of ε-neighborhood in more detail in Fig. 3.

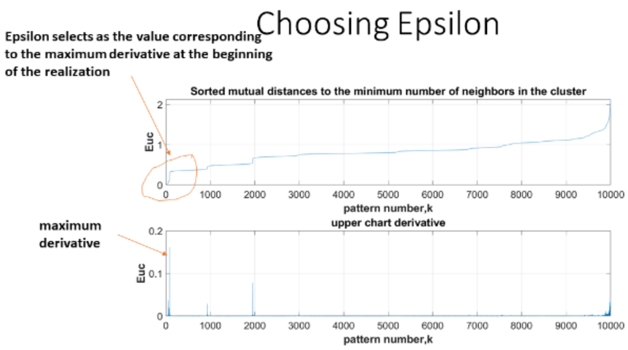

Fig. 3. Selection of the ε-neighborhood.

As a result, after selecting all the necessary parameters of the DBSCAN algorithm, its output is the evaluated clusters with the positions of the patterns detected in them from the fed sample. Due to the search for patterns of different lengths, namely a pattern of 3 pulses and a pattern of 7 pulses, and the "windowing" method used when forming the desired data format for the algorithm, its output can produce side shifts of pattern positions of a larger dimension. That is, the DBSCAN algorithm cannot unambiguously evaluate a cluster with patterns of 7 pulses. For this purpose it is necessary to augment the resulting algorithm with a number of heuristic rules, shown in the Fig. 1. with the radar information processing flowchart. This is implemented in blocks (2) and (5). In this way, the new cDBSCAN clustering algorithm is implemented.

As a priori thresholds in blocks (3) and (5), used for additional cluster separation and false pattern filtering are applied frequency thresholds equal to 240 MHz and 200 ns for pulse duration. These values are taken after analysis of real measurements from the radar passive complex.

IV. Modeling

To solve the problem and to qualitatively evaluate the effectiveness of the cDBSCAN clustering algorithm on the data, which are elementary radio pulses, it is required to describe the parameters of the simulation model of the signal.

In order to describe what are the elementary pulses, we analyzed the actual records of detectable elementary pulses from the passive radar complex. Single pulses have the following parameters: pulse arrival time T_{rec}, pulse duration τ_{pulse}, pulse period T_{pulse}, which is the time difference between the current pulse and the previous pulse, carrier frequency F. Analyzing real recordings, it was found that pulses with some parameters occur most frequently. Namely: the minimum value of period $T_{\min\_pulse} = 2$ μs, the most frequently occurring signal carrier frequencies are the following discrete set $F = [1.09; 1.5; 5.48; 9.8; 16]\ GHz$, the most frequently occurring pulse durations are the discrete set $\tau_{pulse} = [50; 100; 500; 20000; 65000]\ ns$. The parameters of these pulses are taken as the basis of the signals in the simulation model.

Real recordings always contain observation noises, so this should be taken into account for the simulation model as well. As a result, we chose the following observation noises: for pulse arrival time $\Delta T_{rec} = 50\ ns$, for carrier frequency $\Delta F = 1\ kHz$, for pulse duration $\Delta\tau_{pulse} = 10\ ns$.

In addition to observation noise there are also deviations from instantaneous values. Let's assume the following: frequency deviation in fractions of carrier frequency takes equal probability values in the following range of integer numbers $\Delta F_d = 0.001 \dots 0.005$ $(0.1 \dots 0.5\%)$, pulse duration deviation in fractions $\Delta\tau_{p.d} = 0.001 \dots 0.005$ $(0.1 \dots 0.5\%)$.

The task is as follows. We need to form a sample of pulses containing repeating signals with the same parameters - patterns. It is these patterns that should be selected by cDBSCAN algorithm during clustering.

The formation of a sample of pulses is as follows. The pulse parameters are set:

Formed randomly, the period between pulses:

$$T = T_{min\_pulse} \cdot f([1,10]) + g(0, \Delta T_{rec}), \tag{8}$$

where $f([1,10])$ is continuous uniform distribution of integers, $g(0, \Delta T_{rec})$ is normal distribution law with zero mean and variance T_{rec}.

The pulse carrier frequency is formed:

$$f = f(F) \cdot \left(1 + (-1) \cdot f([0,1]) \cdot f(\Delta F_d)\right). \tag{9}$$

Shapes the pulse duration randomly:

$$\delta t = f(\tau) \cdot \left(1 + (-1) \cdot f([0,1]) \cdot f\left(\Delta\tau_{p.d}\right)\right), \tag{10}$$

then the pattern length L is set.

Forms a sample of pulses of $N = 10000$ elements, where the patterns are randomly distributed. To the parameters of pulses, observation noise is added, distributed according to the normal law.

As a result, the observation model is described as follows:

$$\boldsymbol{R} = \boldsymbol{S} + \boldsymbol{E}, \tag{11}$$

where $\boldsymbol{R}$ is an observation noise matrix of dimension $(Nx4)$. $\boldsymbol{S}$ is a state matrix of dimension $(Nx4)$. R is an observation matrix of dimension $(Nx4)$.

It is required to form a sample of pulses and add to this sample two patterns at random. The size of the first pattern is equal to three pulses, and it was assumed that three pulses is the minimum size of the pattern. The size of the 2nd pattern is equal to seven pulses.

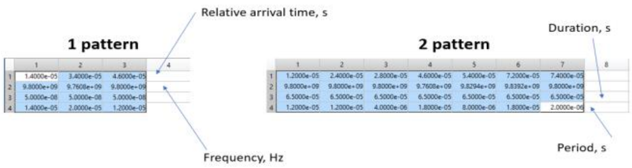

Fig. 4. Example patterns.

It is assumed that the clustering algorithm DBSCAN will be able to separate the added patterns into separate clusters, thus forming two clusters with similar patterns. The rest of the pulses it will consider as noise and noises (outliers).

Let us analyze the simulated sample in detail. For this purpose we plot the probability distribution of parameters of the observation matrix (Fig. 5) and the representation of pulses in the sample (Fig. 6-7).

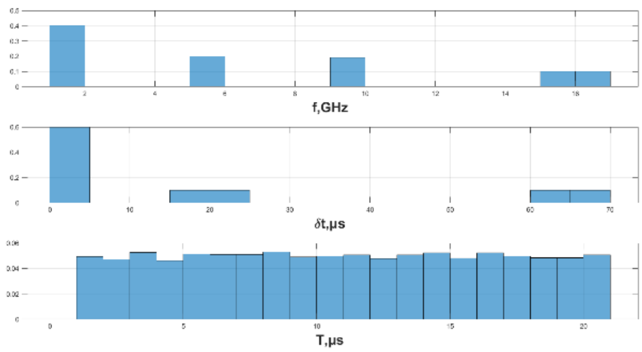

Fig. 5. Probability distributions of the parameters of the observation matrix (without arrival times).

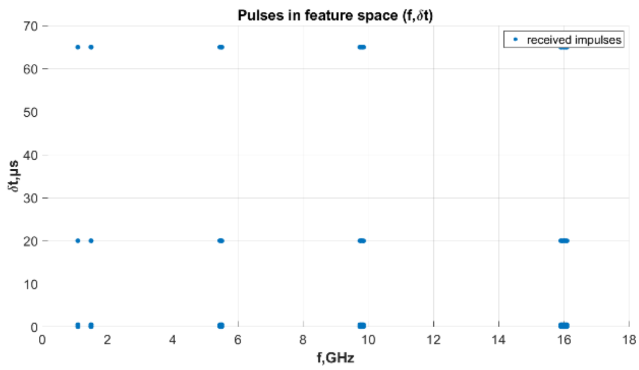

Fig. 6. Two-dimensional representation of pulses.

The following presents the results of the cDBSCAN clustering algorithm for this sample, which contains two patterns of different lengths.

The cDBSCAN algorithm has explicitly allocated two clusters. Now we need to check the accuracy of clustering quality estimation on the set of implementations.

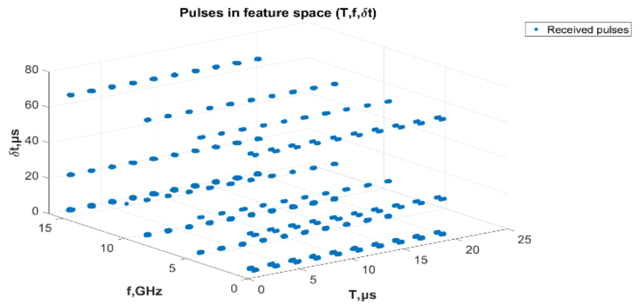

Fig. 7. Three-dimensional representation of pulses with the axis of period values.

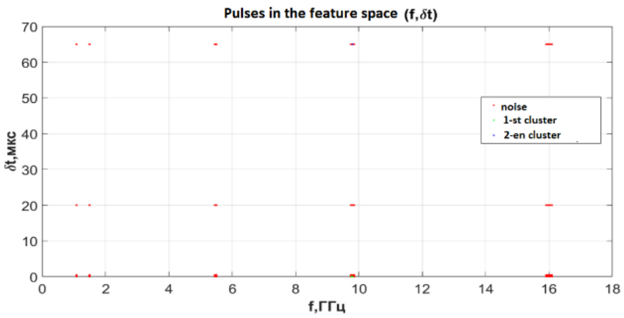

Fig. 8. Pulses in the feature space of pulse frequency and duration.

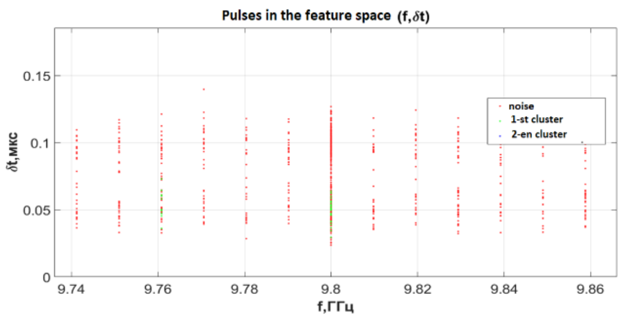

Fig. 9. Pulses in the feature space of the frequency and duration of the pulse in the enlarged scale of the first cluster.

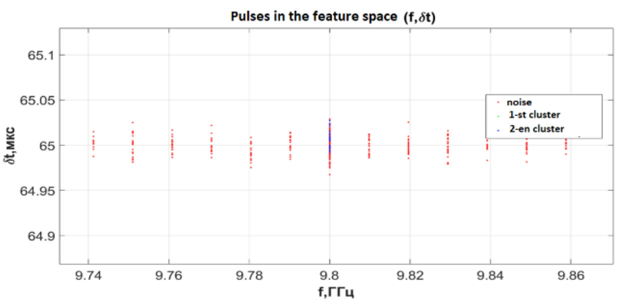

Fig. 10. Pulses in the feature space of the frequency and duration of the pulse in the enlarged scale of the second cluster.

As a result, we obtained the following values of the declared quality metrics.

The result of clustering on 100 cases, the average accuracy of pattern detection was - 98.2632%. The average percentage of false pattern detections was 2.0087%. The average percentage of missed clusters was - 1%. The average number of clusters detected was 1.99.

V. Conclusion

In this paper, we considered modern methods and algorithms for solving the problem of radar information clustering. It was shown that the application of algorithms of statistics and machine learning to radar information processing systems is considered relevant and promising direction. In this paper, it was proposed to apply algorithms for clustering signal patterns at the stage of detection of marks from targets. Solving such a problem will make it possible in the future to select more effective algorithms for secondary (trajectory) target processing.

To solve the problem of radar pattern clustering, we developed the cDBSCAN clustering algorithm, a density-based spatial clustering algorithm with the presence of noise, using a cascade of heuristic rules.

Also, a simulation model of elementary radio pulses based on the analysis of real records of detectable elementary pulses from a passive radar complex was simulated in this work. Applying the cDBSCAN clustering algorithm to the data generated by the model showed that the algorithm correctly identifies patterns in the sample with high accuracy, over 98%.

Further work in this direction is to cluster the real records from the passive radar complex in order to isolate recurring patterns and further analyze the records using cDBSCAN.

References

[1] S. A. Aivazyan, V. M. Buchbacker, L. S. Yenyukov, L. D. Megyalkin, “Applied Statistics: Classification and Dimensional Reduction [Prikladnya statistica: classificatciya i snizenie razmernosti” (In Russian). Reference Edition under Rev. S.A. Aivazyan. M.: Finance and statistics,1989. 607.

[2] A. L. Tatuzov. “Neural networks in radar tasks [Neuronnie seti v radiolocatsii]”. (In Russian). Kn. 28. - M.: Radio Engineering, 2009. - 432 pp.: il. (Scientific series Neurocomputers and their applications.

[3] N. L. Knyazev, L. A. Denisova, “Methods of recognition of dynamic objects in radar space [Metodi raspoznavania dinamicheskix obectov v radiolocatsionnom prostranstve”. (In Russian). UDC 621.396.96:004.021. Omsk State Technical University, t. Omsk, Russia.

[4] G. Revillon, A. Mohammad-Djafari and C. Enderli, “Radar emitters classification and clustering with a scale mixture of normal distributions ”, 2018 IEEE Radar Conference (RadarConf18), 2018, pp. 1371-1376, doi: 10.1109/RADAR.2018.8378764.

[5] L. S. Kim, R. M. Kil, C. H. Jo. “Radar Pattern Classification Based on Class Probability Output Networks”. In: Arik, S., Huang, T., Lai, W., Liu, Q. (eds) Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science, vol 9489, 2015. Springer, Cham.

[6] Ramirez-Figueroa, John A. et al. “A new principal component analysis by particle swarm optimization with an environmental application for data science.” Stochastic Environmental Research and Risk Assessment, 2021: 1-16.

[7] Z. M. Liu, “Recognition of Multifunction Radars Via Hierarchically Mining and Exploiting Pulse Group Patterns, ” in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 6, pp. 4659-4672, Dec. 2020, doi: 10.1109/TAES.2020.2999163.

[8] Jang, Jennifer and Heinrich Jiang. “DBSCAN++: Towards fast and scalable density clustering.” ArXiv abs/1810.13105, 2019: n. pag.

[9] Ester, Martin et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” KDD, 1996.

[10] S. Haikin. “Neural networks: a complete course, 2nd edition: Translated from English: Williams Publishing House”. Publishing House Williams, 2006. - 1104 p.: Ill. - Paral. tit. Eng.