IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 25, 2019, accepted September 1, 2019, date of publication September 6, 2019, date of current version September 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939865

Detection of Array Signal Number With Multiple Sensors Based on Transfer Component Analysis

JIAQI ZHEN, (Member, IEEE)
College of Electronic Engineering, Heilongjiang University, Harbin 150080, China
e-mail: zhenjiaqi2011@163.com

This work was supported in part by the National Natural Science Foundation of China under Grant 61501176, in part by the Natural Science Foundation of Heilongjiang Province under Grant F2018025, and in part by the Postdoctoral Scientific Research Developmental Fund of Heilongjiang Province under Grant LBH-Q17149.

ABSTRACT The conventional algorithms for estimating number of array signals are only suitable for the background of Gaussian white noise, and need many snapshots, but their performance will reduce seriously in the circumstance of impulse noise and small samples. Therefore, a new method of detecting array signal number with multiple sensors based on transfer component analysis is proposed in this paper. First, the array signals in Gaussian white and impulse noise are respectively modeled. Then the received array data are transformed into a common hidden space by the mapping function, thus, data in the hidden space have the same distribution, and most initial characteristics are retained. Finally, a support vector machine or K-means clustering are used for classifying the mapped data into two categories, on this basis, the array signal number can be estimated.

INDEX TERMS Array signal number, transfer component analysis, impulse noise, support vector machine, K-means clustering.

I. INTRODUCTION

Signal number estimation is always one of the hottest topics in array signal processing, it is often the precondition of further processing [1]–[8], the research dates from the late 1950s. At that time, a factitious detecting threshold is needed to set to compare with the likelihood test statistics [9], so it is easy to be influenced by the subjective factors. The information theory criteria, such as modeling by shortest description(MDL) [10] and Akaike information criterion(AIC) [11] are widely concerned with their good estimation performance, but they are only suitable for the background of Gaussian white noise (GWN), in order to solve this problem, gerschgorin disk method [12] was proposed in the colored noise, then some modified techniques are presented successively. Another kind of method is based on Bootstrap, its essence is the resample process to the received signals, in 2000, Brcich *et al.* [13] first put forward the method to estimate source number through constructing hypothesis test statistics by Bootstrap. Combining with clustering, Zhang *et al.* [14] acquired number of signals under the circumstances of low signal to noise ratio(SNR) and small snapshots with Bootstrap, the performance was improved by the full use of eigenvalues and eigenvectors.

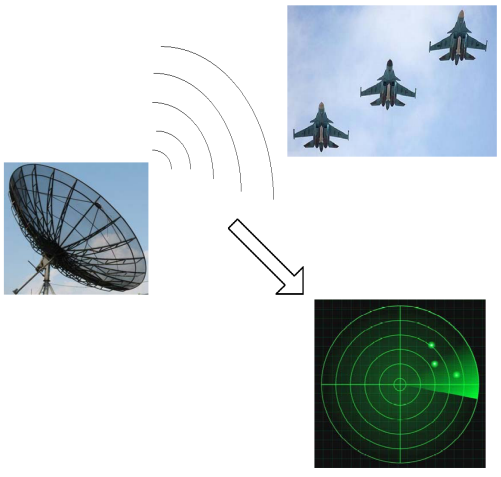

FIGURE 1. Source number detection.

In numerous colored noise, impulse noise is a kind of actual noise that occasionally appears impulsive characteristic, such as atmospherics [15], [16], ocean noise [17]–[20] and automobile engine [21]–[24], they manifest GWN characteristic

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

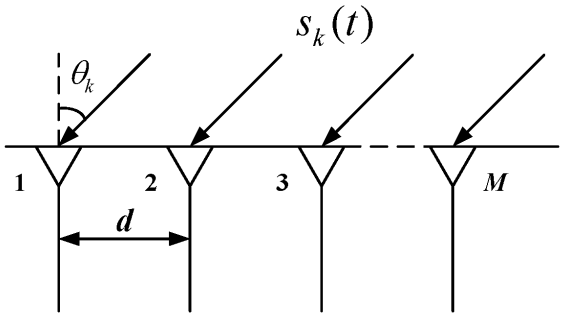

FIGURE 2. Signal model.

in most cases, but there are some very large amplitudes once in a while [25], so it is not proper to express it with GWN characteristic, its distribution of probability density has longer stretching than Gaussian noise. In many impulse noise models, α stable distribution is relatively simple, it can briefly be used for digital signal processing and has some impulse characteristic, so it is a proper hypothesis [26]–[32]. Gaussian noise is a special case of stable distribution, the difference between impulse and GWN is the probability density function. Some scholars applied Bootstrap to the background of impulse noise [33], and minimum covariance estimator was employed to weaken the effect of the noise, it needs a large sum of data via thousands of resamples and rearranges to the signals, so it requires a lot of time. The other effective technique for impulse noise is fractional lower order statistics(FLOS) [34]–[37] which is very commonly used in signal processing, but it demands many snapshots.

This paper borrows the idea of transfer learning [38], [39] and proposes a method for signal number estimation in the circumstance of impulse noise. The array signals in Gaussian white and impulse noise are separately modeled. Then they are transformed into a common hidden space by the mapping function, thus, data in the hidden space have the same distribution, and most initial characteristics are retained. Finally, a support vector machine(SVM) is obtained through training with the mapped data, on this basis, the array signal number can be estimated by classifying the source and impulse noise. Besides, we can also use K-means clustering for the classification.

II. SIGNAL MODEL

As is shown in Figure 2, there are K far-field narrow-band sources from $\theta_k(k = 1, 2, \cdots K)$ arriving at the uniform linear array (ULA) formed by M sensors, these sources are Gaussian distribution, and there are no array perturbations, then output of array at time t in GWN can be written

$$\begin{aligned} \boldsymbol{X}_{\mathrm{S}}(t) &= \boldsymbol{A}(\theta)\boldsymbol{S}(t) + \boldsymbol{N}_{\mathrm{G}}(t) \\ &= [x_{\mathrm{S}_1}(t) \cdots x_{\mathrm{S}_m}(t) \cdots x_{\mathrm{S}_M}(t)]^{\mathrm{T}}. \end{aligned} \tag{1}$$

where $\boldsymbol{A}(\theta) = [\boldsymbol{a}(\theta_1) \cdots \boldsymbol{a}(\theta_k) \cdots \boldsymbol{a}(\theta_K)]$ is array manifold, $\boldsymbol{a}(\theta_k)$ is steering vector of the kth signal, $\boldsymbol{S}(t)$ and $\boldsymbol{N}_{\mathrm{G}}(t)$ are respectively the signal and GWN matrix, $x_{\mathrm{S}_m}(t)$ is the output of the mth sensor in GWN. As α stable distribution has no specific expression of probability density, eigenfunction is usually used for its description, it is

$$\varphi(v) = \begin{cases} \exp\{\mathrm{j}\zeta v - \delta|v|^{\alpha}[1 + \mathrm{j}\varepsilon\mathrm{sgn}(v)\tan(\frac{\alpha\pi}{2})]\}, & \alpha \neq 1 \\ \exp\{\mathrm{j}\zeta v - \delta|v|^{\alpha}[1 + \mathrm{j}\varepsilon\mathrm{sgn}(v)\frac{2}{\pi}\log|v|]\}, & \alpha = 1 \end{cases} \tag{2}$$

where $\alpha \in (0, 2]$ is characteristic index which decides the impulsive degree of the distribution, $\delta \geq 0$ is dispersion coefficient denoting degree of dispersion, $-1 \leq \varepsilon \leq 1$ is symmetric distribution, ζ is a real number which determines the location of the distribution on the coordinate, these parameters determine the concrete distribution, the noise will be impulsive when it is $\alpha(0<\alpha<2)$ stable distribution, then the output on this occasion can be expressed

$$\begin{aligned} \boldsymbol{X}_{\mathrm{T}}(t) &= \boldsymbol{A}(\theta)\boldsymbol{S}(t) + \boldsymbol{N}_{S\alpha S}(t) \\ &= [x_{\mathrm{T}_1}(t) \cdots x_{\mathrm{T}_m}(t) \cdots x_{\mathrm{T}_M}(t)]^{\mathrm{T}}. \end{aligned} \tag{3}$$

where $x_{\mathrm{T}_m}(t)$ is the output of the mth sensor in impulse noise, in circumstance of $0 < \alpha \leq 2$, the smaller α is, the stronger the impulse is, if $\alpha = 2$, it will be Gaussian distribution.

Here, we put the array output in GWN $\boldsymbol{X}_{\mathrm{S}}(1), \cdots \boldsymbol{X}_{\mathrm{S}}(T_1)$ (Source domain data) and that in impulse noise $\boldsymbol{X}_{\mathrm{T}}(1), \cdots \boldsymbol{X}_{\mathrm{T}}(T_2)$ (Target domain data) together, then the mixed matrix is acquired

$$\boldsymbol{X} = [\boldsymbol{X}_{\mathrm{S}}(1) \cdots \boldsymbol{X}_{\mathrm{S}}(T_1), \boldsymbol{X}_{\mathrm{T}}(1) \cdots \boldsymbol{X}_{\mathrm{T}}(T_2)]_{M \times (T_1 + T_2)}. \tag{4}$$

III. THE ALGORITHM BASED ON TRANSFER COMPONENT ANALYSIS

A. INFORMATION TRANSFER

Transfer component analysis(TCA) gives us an effective approach for solving domain matching, we can use this technique to train array data in white and impulse noise to transform the two kinds of signals into a common hidden space through a mapping function, the data in this space have the same distribution and most characteristics of original array signals are retained. Then train the classifier with the mapped data set based on the traditional machine learning. First, a nonlinear function $\chi(\cdot)$ is used to transform the data in GWN and impulse noise into W dimensional common hidden space, then the distance between the two kinds of probability distribution functions is

$$D(\boldsymbol{X}'_{\mathrm{S}}, \boldsymbol{X}'_{\mathrm{T}}) = \left\| \frac{\sum_{t=1}^{T_1} \chi(\boldsymbol{X}_{\mathrm{S}}(t))}{T_1} - \frac{\sum_{t=1}^{T_2} \chi(\boldsymbol{X}_{\mathrm{T}}(t))}{T_2} \right\|^2 \tag{5}$$

where $\boldsymbol{X}'_{\mathrm{S}}$ is the matrix after mapping the source domain data $[\boldsymbol{X}_{\mathrm{S}}(1) \cdots \boldsymbol{X}_{\mathrm{S}}(T_1)]$ to the hidden space, $\boldsymbol{X}'_{\mathrm{T}}$ is the matrix after mapping the target domain data $[\boldsymbol{X}_{\mathrm{T}}(1) \cdots \boldsymbol{X}_{\mathrm{T}}(T_2)]$ to the hidden space, then the kernel matrix $\boldsymbol{B}$ can be obtained via solving semidefinite programming [38], we separately define source domain, target domain and cross domain as $\boldsymbol{B}_{\mathrm{S,S}}, \boldsymbol{B}_{\mathrm{T,T}}$ and $\boldsymbol{B}_{\mathrm{S,T}} = \boldsymbol{B}_{\mathrm{T,S}}$, thus $\boldsymbol{B}$ is expressed

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{\mathrm{S,S}} & \boldsymbol{B}_{\mathrm{S,T}} \\ \boldsymbol{B}_{\mathrm{T,S}} & \boldsymbol{B}_{\mathrm{T,T}} \end{bmatrix} \tag{6}$$

then (5) can be written through kernel embedding

$$D(\boldsymbol{X}'_{\mathrm{S}}, \boldsymbol{X}'_{\mathrm{T}}) = \left\| \frac{\sum_{t=1}^{T_1} \chi(\boldsymbol{X}_{\mathrm{S}}(t))}{T_1} - \frac{\sum_{t=1}^{T_2} \chi(\boldsymbol{X}_{\mathrm{T}}(t))}{T_2} \right\|^2 = \mathrm{tr}(\boldsymbol{BC}) \tag{7}$$

where $\mathrm{tr}(\cdot)$ denotes trace of the matrix, $\boldsymbol{C}$ is a middle variable, and

$$\boldsymbol{C}_{i,j} = \begin{cases} \frac{1}{T_1}, & \boldsymbol{X}_i, \boldsymbol{X}_j \in \boldsymbol{X}_{\mathrm{S}} \\ \frac{1}{T_2}, & \boldsymbol{X}_i, \boldsymbol{X}_j \in \boldsymbol{X}_{\mathrm{T}} \\ -\frac{1}{T_1 T_2}, & \text{others} \end{cases} \tag{8}$$

here, $\boldsymbol{X}_i$ is the element on the ith line of $\boldsymbol{X}$, define the transforming matrix that maps $\boldsymbol{X}$ to W dimensional common hidden space as $\tilde{\boldsymbol{Z}}$, and the middle variable as [40], [41]

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}(\boldsymbol{B}^{-1/2}\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^{\mathrm{T}}\boldsymbol{B}^{-1/2})\boldsymbol{B} \tag{9}$$

let $\boldsymbol{B}^{-1/2}\tilde{\boldsymbol{Z}} = \boldsymbol{Q}$, then the maximum mean difference distance between source and target domain after mapping is

$$\begin{aligned} D(\boldsymbol{X}'_{\mathrm{S}}, \boldsymbol{X}'_{\mathrm{T}}) &= \mathrm{tr}\Big((\boldsymbol{B}(\boldsymbol{B}^{-1/2}\tilde{\boldsymbol{Z}})(\boldsymbol{B}^{-1/2}\tilde{\boldsymbol{Z}})^{\mathrm{T}}\boldsymbol{B})\boldsymbol{C}\Big) \\ &= \mathrm{tr}\Big((\boldsymbol{B}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B})\boldsymbol{C}\Big) \end{aligned} \tag{10}$$

on one hand, we need to minimize this distance, on the other hand, the mapping function $\chi(\cdot)$ needs to retain the useful characteristic for training target classifier, we can achieve the goal by maximizing the variance of the data after mapping according to principal component analysis, so the sample covariance matrix after mapping is $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{BPBQ}$, here, $\boldsymbol{P}$ is the centering matrix

$$\boldsymbol{P} = \boldsymbol{I}_{T_1+T_2} - \Big(\frac{1}{T_1+T_2}\Big)\boldsymbol{I}_{T_1+T_2} \tag{11}$$

where $\boldsymbol{I}_{T_1+T_2}$ is the unitary matrix with the dimension $(T_1+T_2)\times(T_1+T_2)$, so the domain matching equals to solve the following problem

$$\begin{aligned} &\min_{\boldsymbol{B}^{-1/2}\tilde{\boldsymbol{Z}}} \mathrm{tr}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{BCBQ}) + \gamma\,\mathrm{tr}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}) \\ &\text{s.t. } \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{BCBQ} = \boldsymbol{I}_W \end{aligned} \tag{12}$$

where, γ is a compromise parameter, in order to solve the constraint of the non-convex norm above, we transform (12) into the following optimization problem

$$\max_{\boldsymbol{B}^{-1/2}\tilde{\boldsymbol{Z}}} \mathrm{tr}\Big((\boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{BCB} + \gamma\boldsymbol{I}_W)\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{BPBQ}\Big) \tag{13}$$

in the process of disposing the problem above, we can decompose the kernel matrix $(\boldsymbol{BCB} + \gamma\boldsymbol{I})^{-1}\boldsymbol{BPB}$ according to Fisher discriminant analysis [42], then the eigenvector of corresponding W eigenvalues can be solved as $\boldsymbol{Q}$, then

$$\boldsymbol{X}' = \chi(\boldsymbol{X}) = \tilde{\boldsymbol{Z}}\boldsymbol{X} = \boldsymbol{B}^{1/2}\boldsymbol{Q}\boldsymbol{X} \tag{14}$$

thus, the new sample set $\boldsymbol{X}'$ in the hidden space is evaluated, and background noise of the data in $\boldsymbol{X}'$ will approximately obey the GWN distribution if source domain data occupies the majority in $\boldsymbol{X}$, then $\boldsymbol{X}'$ can be disposed as the data in GWN. On one hand, the problem of low precision caused by small samples in impulse noise will be settled, on the other hand, the training time is greatly shortened on account of referencing the experience in GWN, then number of array signals can be determined by SVM or clustering in the hidden space.

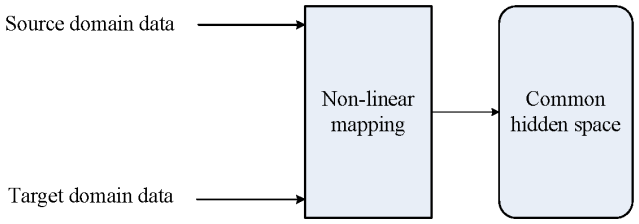

FIGURE 3. Map the source and target domain data into common hidden space.

B. CONSTRUCT CLASSIFICATION CHARACTERISTIC

The covariance of the data in hidden space is

$$\boldsymbol{R}' = E\Big[\boldsymbol{X}'(\boldsymbol{X}')^{\mathrm{H}}\Big] \tag{15}$$

then the eigenvalues $\lambda_1 > \cdots > \lambda_K > \lambda_{K+1} = \cdots = \lambda_M$ and corresponding eigenvectors $[\boldsymbol{u}_1, \cdots \boldsymbol{u}_K, \boldsymbol{u}_{K+1}, \cdots \boldsymbol{u}_M]$ can be acquired too, thus $\boldsymbol{a}(\theta)$ can be written

$$\boldsymbol{a}(\theta_k) = \beta_k \boldsymbol{u}_k, \quad k = 1, 2, \cdots M \tag{16}$$

where β_k is the weighting coefficient, as $\boldsymbol{a}(\theta)$ is orthogonal to $[\boldsymbol{u}_1, \cdots \boldsymbol{u}_K, \boldsymbol{u}_{K+1}, \cdots \boldsymbol{u}_M]$, so we have

$$\boldsymbol{a}^{\mathrm{H}}(\theta_k)\boldsymbol{u}_k = 0, \quad k = K+1, \cdots M \tag{17}$$

define weighting expression of eigenvectors as

$$\mu_m = |\boldsymbol{a}^{\mathrm{H}}(\theta_k)\boldsymbol{u}_m|, \quad m = 1, 2, \cdots M \tag{18}$$

according to (16) and (17), equation (18) can be expressed

$$\begin{aligned} \mu_m &= |(\beta_k \boldsymbol{u}_k)^{\mathrm{H}}\boldsymbol{u}_m| \\ &= \begin{cases} |\beta_m|, & m = 1, 2, \cdots K \\ 0, & m = K+1, \cdots M \end{cases} \end{aligned} \tag{19}$$

weight μ_m with ς_m, then

$$\xi_m = {\varsigma_m}^{1/2}\mu_m, \quad m = 1, 2, \cdots M \tag{20}$$

where ξ_m can be deemed as classification characteristic, ς_m can be selected as [43]

$$\varsigma_m = \frac{(\lambda_m - \sigma^2)^2}{\lambda_m}, \quad m = 1, 2, \cdots M-1 \tag{21}$$

here, σ^2 can be replaced by the minimum eigenvalue.

C. CONSTRUCTION OF CLASSIFIER

After obtaining the classification characteristic $\xi_1, \xi_2, \cdots \xi_{M-1}$, we can employ two classification means to determine signal number, SVM or K-means classifier, and they are described respectively below:

1) SVM CLASSIFIER

SVM can be described as a binary classification problem [44], namely to determine a hyperplane $(\boldsymbol{\omega}, b)$ which can classify two training samples in plane $\boldsymbol{c}$ correctly: if $f(\boldsymbol{c}) = \boldsymbol{\omega}^{\mathrm{T}}\phi(\boldsymbol{c}) + b \geq 1$, then $y = +1$; if $f(\boldsymbol{c}) = \boldsymbol{\omega}^{\mathrm{T}}\phi(\boldsymbol{c}) + b \leq -1$, then $y = -1$, where $f(\boldsymbol{c})$ is a linear function, $\boldsymbol{\omega}$ is the coefficient, b is the intercept, $\phi(\boldsymbol{c})$ is the mapping function of $\boldsymbol{c}$, y is the categorical variable, then corresponding maximum classification distance is resolved by

$$\begin{cases} \min\limits_{\boldsymbol{\omega},b} \dfrac{1}{2}\|\boldsymbol{\omega}\|^2 \\ y_i\Big(\boldsymbol{\omega}^{\mathrm{T}}\phi(\boldsymbol{c}_i) + b\Big) \geq 1, \quad i = 1, 2, \cdots M-1 \end{cases} \tag{22}$$

where $\boldsymbol{c}_i$ is the ith element in plane $\boldsymbol{c}$, y_i is the ith categorical variable, then the Lagrangian function of the problem can be written as

$$L(\boldsymbol{\omega}, b, \boldsymbol{\eta}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{M-1} \eta_i\Big(1 - y_i\Big(\boldsymbol{\omega}^{\mathrm{T}}\phi(\boldsymbol{c}_i) + b\Big)\Big) \tag{23}$$

here, $\boldsymbol{\eta} = (\eta_1, \cdots \eta_{M-1})$ is Lagrangian multiplier vector, $\eta_i \geq 0$ is the ith multiplier, let the partial derivatives of $L(\boldsymbol{\omega}, b, \boldsymbol{\eta})$ with respect to $\boldsymbol{\omega}$ and b equal zero, we separately have

$$\boldsymbol{\omega} = \sum_{i=1}^{M-1} \eta_i y_i \phi(\boldsymbol{c}_i) \tag{24}$$

$$\sum_{i=1}^{M-1} \eta_i y_i = 0 \tag{25}$$

take (24) into $f(\boldsymbol{c})$, then

$$f(\boldsymbol{c}) = \sum_{i=1}^{M-1} \eta_i y_i \phi^{\mathrm{T}}(\boldsymbol{c}_i)\phi(\boldsymbol{c}) + b \tag{26}$$

the dual problem of (22) is acquired through Lagrangian multiplier method

$$\begin{cases} \max_{\boldsymbol{\eta}} \sum\limits_{i=1}^{M-1} \eta_i - \dfrac{1}{2}\sum\limits_{i=1}^{M-1}\sum\limits_{i=1}^{M-1} \eta_i\eta_j y_i y_j \phi^{\mathrm{T}}(\boldsymbol{c}_i)\phi(\boldsymbol{c}) \\ \sum\limits_{i=1}^{M-1} \eta_i y_i = 0 \\ \eta_i \geq 0, \quad i = 1, 2, \cdots M-1 \end{cases} \tag{27}$$

next, we can select a proper kernel function $\kappa(\boldsymbol{c}_i, \boldsymbol{c}) = \phi^{\mathrm{T}}(\boldsymbol{c}_i)\phi(\boldsymbol{c})$ to transform the nonlinear problem into a linear one locating in a higher dimension space, and realize it in original space, thus, the computational complexity depends on number of samples, not the space dimensionality, so (27) is equivalent to

$$\begin{cases} \max_{\boldsymbol{\eta}} \sum\limits_{i=1}^{M-1} \eta_i - \dfrac{1}{2}\sum\limits_{i=1}^{M-1}\sum\limits_{i=1}^{M-1} \eta_i\eta_j y_i y_j \kappa(\boldsymbol{c}_i, \boldsymbol{c}) \\ \sum\limits_{i=1}^{M-1} \eta_i y_i = 0 \\ \eta_i \geq 0, \quad i = 1, 2, \cdots M-1 \end{cases} \tag{28}$$

then $\boldsymbol{\eta} = (\eta_1, \cdots \eta_{M-1})$ can be obtained by solving (28), consequently, the final decision function is determined according to $\kappa(\boldsymbol{c}_i, \boldsymbol{c})$ and (26)

$$f(\boldsymbol{c}) = \sum_{i=1}^{M-1} \eta_i y_i \kappa(\boldsymbol{c}_i, \boldsymbol{c}) + b \tag{29}$$

equation (29) can be used for the binary classification between signal and noise. Generally speaking, Gaussian kernel function is flexible, and we had better choose it for the solution if the distribution of the data can not be determined in advance, that is

$$\kappa(\boldsymbol{c}_i, \boldsymbol{c}) = \exp\Big(-\frac{\|\boldsymbol{c} - \boldsymbol{c}_i\|^2}{2\sigma^2}\Big) \tag{30}$$

where σ is the bandwidth of Gaussian kernel, then $\xi_1, \cdots, \xi_m, \cdots, \xi_{M-1}$ are separately taken into (29) as input data $\boldsymbol{c}$ and implement the classification, namely if $f(\xi_m) \geq 1$, it will be signal, else it will be noise, so signal number will be m who satisfies $f(\xi_m) \geq 1$ and $f(\xi_{m+1}) \leq -1$, then we summarize the algorithm as follows:

Step 1: Change signal number from 0 to $M-1$, generate array received data in GWN $[\boldsymbol{X}_{\mathrm{S}}(1), \cdots, \boldsymbol{X}_{\mathrm{S}}(T_1)]$ with (1) and in impulse noise $[\boldsymbol{X}_{\mathrm{T}}(1), \cdots, \boldsymbol{X}_{\mathrm{T}}(T_2)]$ with (3) as the training data;

Step 2: Map the two kinds of data into common hidden space with (14), if source domain data $[\boldsymbol{X}_{\mathrm{S}}(1), \cdots, \boldsymbol{X}_{\mathrm{S}}(T_1)]$ occupies the majority in $\boldsymbol{X}$, then $\boldsymbol{X}'$ can be disposed as the data in GWN;

Step 3: Calculate the classification characteristic $\xi_1, \cdots, \xi_m, \cdots, \xi_{M-1}$ according to (20);

Step 4: Construct SVM by (29);

Step 5: Take $\xi_1, \cdots, \xi_m, \cdots, \xi_{M-1}$ into SVM for training, then determine γ,η_i, y_i, c_i, σ and b;

Step 6: Map the test data into hidden space, then take them into trained SVM to estimate array signal number in impulse noise.

The proposed algorithm uses TCA and SVM for the estimation, so it can be called TCA-SVM for short. In fact, TCA is the critical first step for estimating signal number which transferring the knowledge of Gaussian white noise into impulse noise, after that, we can employ any criteria to determine signal number, where SVM/Cluster are very appropriate for the circumstances of small samples and low SNR.

2) K-MEANS CLASSIFIER

K-means clustering [45] is determined by distance similarity, that is to say the nearer the two samples is, the more similar

they are, and they can be classified into a same group, then some compact and independent clusters are acquired. First, k initial clustering centers are selected randomly from input data set, then based on close criterion, the distance between every data object and k clustering centers are calculated, the data are classified into the region where the nearest cluster center is located, thus, a category composed by the cluster center and all the allocated data is formed. After allocating all the data, compute the average of all data object in every cluster repeatedly, then a new clustering center is obtained, we do not iterate in turn until the termination conditions are met. All the data categories have been classified at this moment, k cluster is acquired.

The ξ_m denoting clustering center is expressed by the following equation

$$\xi_m = \frac{1}{N_j} \sum_{u \in Q_j} \xi_u \tag{31}$$

where ξ_u is the vector of attribute u, $\xi_1, \xi_2, \cdots, \xi_k$ are the vector sets corresponding k clustering centers, and they have been estimated by the process above, Q_j is the clustering region of the center, N_j is the number of the data in ξ_u. The vector distance is calculated by data and information of center, the popular algorithms include Euclidean and cosine distance, choosing different formulas will influence the distance calculation results to some extent, then affect the clustering. In this paper, Euclidean distance is adopted according to the extensive application of various formulas, that is

$$d_{ij} = |\xi_i - \xi_j|, \quad (i, j = 1, 2, \cdots, M-1) \tag{32}$$

Then we define two initial values of signal and noise: ${\xi_1}^{(1)} = \xi_1$, ${\xi_2}^{(1)} = \xi_{M-1}$, in the ith iteration, if $d_{1p} > d_{p2}$, $\xi_p \in {Q_2}^{(i)}$, ξ_p belongs to the noise; Else, it is corresponding to signal. This algorithm uses TCA and cluster for the estimation, so it can be called TCA-Cluster for short.

IV. RESULTS

In this section, several experiments are performed with matlab, Pentium 4 processors with dual cores, 4GHz memory. In the first experiment, signal model is shown as Figure 2, four normal sinusoidal sources are from 10°, 20°, 30° and 40°, number of sensors $M = 10$, impulse noise with $\alpha = 1.3$ is illustrated in Figure 4, we employ the conventional MDL and MUSIC algorithms for signal number and spatial spectrum estimations separately, Figure 5 and Figure 6 show their simulation results. We can see that the estimation result of MDL is not stable all the time, it still has a large error even if SNR is high, and corresponding MUSIC spectrum no longer relates to the DOA of actual signals.

Next, only white and impulse noise form the mixed matrix $\boldsymbol{X}$, the total number of two kinds of noises is 200, when GWN accounts for different proportions, the new sample $\boldsymbol{X}'$ is given in Figure 7 to Figure 10.

From Figure 7 to Figure 10 we know impulsive characteristic of $\boldsymbol{X}'$ is becoming less and less obvious with the increase

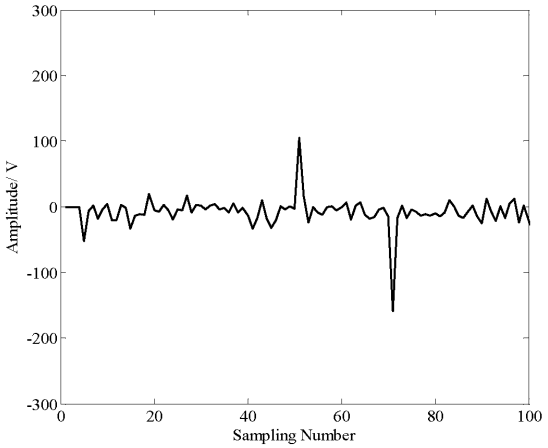

FIGURE 4. Impulse noise when $\alpha = 1.3$.

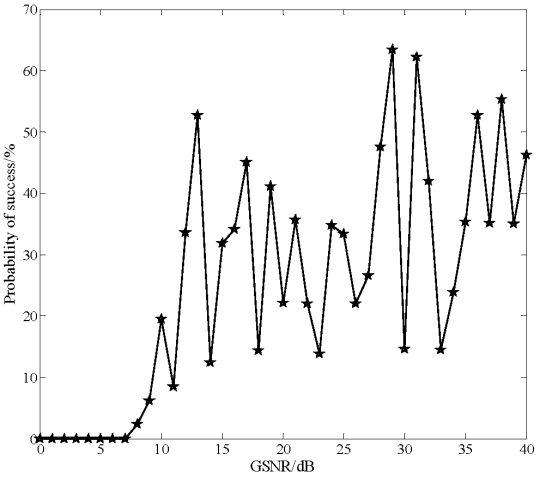

FIGURE 5. Estimation precision of MDL.

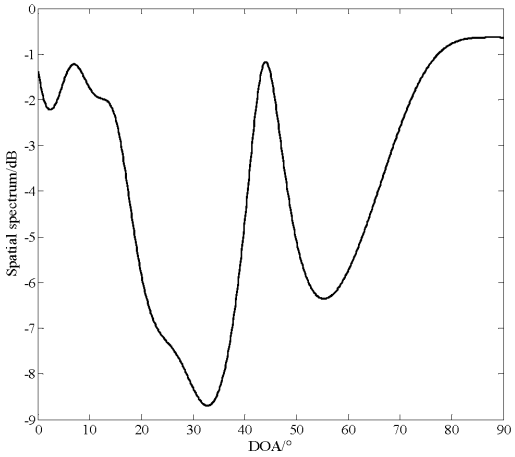

FIGURE 6. MUSIC spatial spectrum.

of proportion of GWN, when it is 80%, we can approximately regard $\boldsymbol{X}'$ as GWN, then some traditional algorithms can be used here.

In the third experiment, number of training samples in GWN $T_1 = 800$, that in impulse noise $T_2 = 200$; number

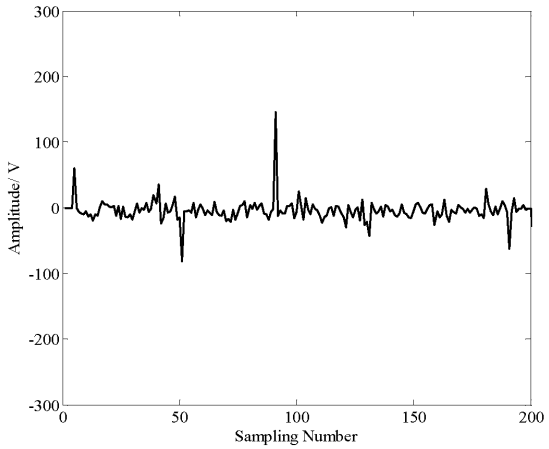

FIGURE 7. Converted data when GWN accounts for 20%.

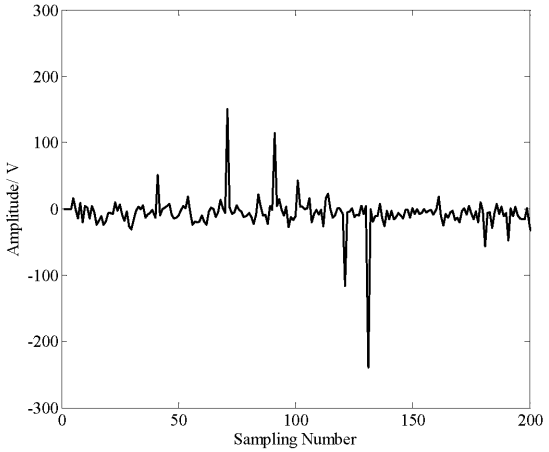

FIGURE 8. Converted data when GWN accounts for 40%.

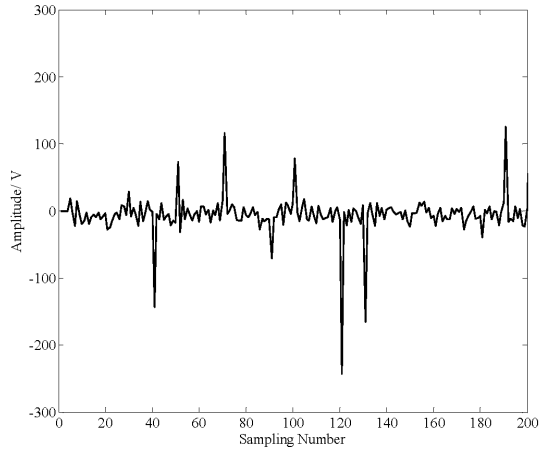

FIGURE 9. Converted data when GWN accounts for 60%.

of test samples in GWN $T_1 = 80$, and that in impulse noise $T_2 = 20$, $\alpha = 1.3$, $\delta = 1$, and generalized signal-to-noise ratio(GSNR) is defined as

$$\rho_{\mathrm{GSNR}} = 10\lg\Big(\frac{1}{\delta(T_1+T_2)}\sum_{t=1}^{T_1+T_2}|s(t)|^2\Big) \tag{33}$$

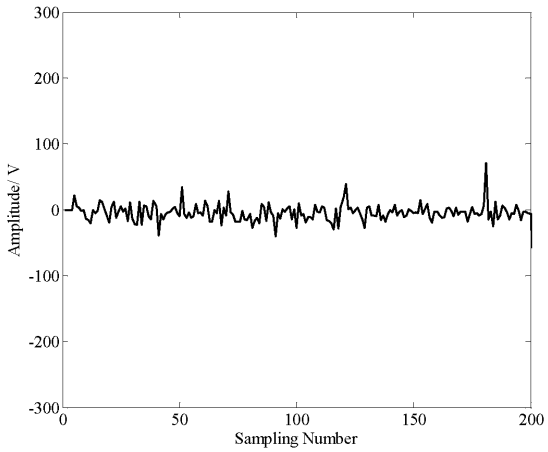

FIGURE 10. Converted data when GWN accounts for 80%.

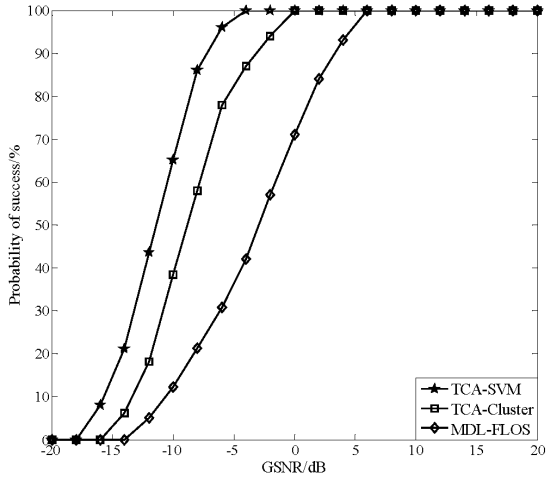

FIGURE 11. Probability of success versus GSNR.

TABLE 1. Average running time.

Algorithms	Training Times	Test Time
MDL-FLOS	0s	0.16s
TCA-SVM	0.319s	0.032s
TCA-Cluster	0s	0.025s

other conditions are the same with the first experiment, the proposed TCA-SVM and MDL based on FLOS(MDL-FLOS) are respectively employed for the estimation, 500 trials, the precision and computational time are given in Figure 11 and Table 1.

From Figure 11 we know that the precisions of TCA-SVM and TCA-Cluster are higher than that of MDL-FLOS at the same GSNR, the probabilities of success are 100% when their GSNR respectively reach -4dB, -1dB and 6dB. Table 1 shows that although the training time of TCA-SVM is long, its test time is short, and it has a higher probability of success, so we can effectively detect array signal number as long as the classifier is trained well.

In the fourth experiment, number of samples of training data is the same with experiment 1, in the test data, number of samples in GWN accounted for 80% of the total data, other conditions are the same with experiment 1, GSNR=10dB, the precision is illustrated in Figure 12. We know that the

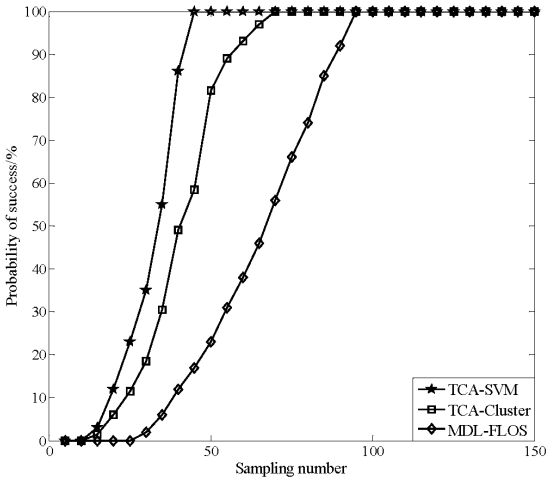

FIGURE 12. Probability of success versus sampling number.

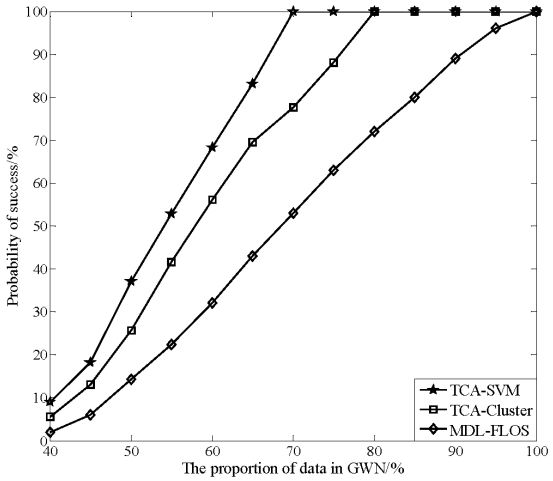

FIGURE 13. Effect of different proportions of data in GWN.

precisions of proposed TCA-SVM and TCA-Cluster are higher than that of MDL-FLOS in the same number of test samples, we can also completely estimate signal number accurately when sample number of MDL-FLOS is 95, then the two algorithms no longer have any differences.

In the fifth experiment, we will observe the effect of different proportions of data in GWN, number of samples of training data is the same with experiment 1, in the test data, number of samples in GWN accounted for 40% to 100% of the total data, other conditions are the same with experiment 1, GSNR = 0dB, the precision is illustrated in Figure 13.

We observe that the precision is getting higher with the ratio of GWN increasing, so the proposed TCA method are appropriate for the circumstances that array sensors ccumulate a sufficient proportion of data in GWN.

V. CONCLUSION

In order to dispose the problem of signal number estimation in impulse noise with small samples, the paper provided a new method based on transfer component analysis, the array data in GWN and impulse noise are transformed into a common hidden space by the mapping function acquired through training the two kinds of data, then the machine learning is employed to obtain a SVM with the mapped data, or we can use K-means clustering, on this basis, the array signal number can be estimated by classifying the source and impulse noise, simulation examples have shown that the proposed method performs better than that of MDL based on FLOS in the condition of low GSNR or a little sample data, meanwhile, it has a faster test speed. Thus, the problem of inaccurately estimation of signal number due to the small samples in impulse noise will be solved to a large extent. In fact, after transforming the data into common hidden space, more problems will also be disposed, so we are going to work on DOA and other sensor parameters estimation with TCA in future.

ACKNOWLEDGMENT

The author would like to thank professor Qun Ding, and all the editors and reviewers of this paper.

REFERENCES

[1] A. Alexandridis and A. Mouchtaris, "Multiple sound source location estimation in wireless acoustic sensor networks using DOA estimates: The data-association problem," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 2, pp. 342–356, Feb. 2018.

[2] N. H. Nguyen and K. Doğançay, "Instrumental variable based Kalman filter algorithm for three-dimensional AOA target tracking," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1605–1609, Oct. 2018.

[3] E. Kazikli and S. Gezici, "Hybrid TDOA/RSS based localization for visible light systems," *Digit. Signal Process.*, vol. 86, pp. 19–28, Mar. 2019.

[4] S. Tomic, M. Beko, and M. Tuba, "A linear estimator for network localization using integrated RSS and AOA measurements," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 405–409, Mar. 2019.

[5] Z. Na, Y. Wang, X. Li, J. Xia, X. Liu, M. Xiong, and W. Lu, "Subcarrier allocation based simultaneous wireless information and power transfer algorithm in 5G cooperative OFDM communication systems," *Phys. Commun.*, vol. 29, pp. 164–170, Aug. 2018.

[6] Z. Na, J. Lv, M. Zhang, B. Peng, M. Xiong, and M. Guan, "GFDM based wireless powered communication for cooperative relay system," *IEEE Access*, vol. 7, pp. 50971–50979, 2019.

[7] H. Yang, B. Li, G. Liu, X. Liu, and X. Peng, "DNF-SC-PNC: A new physical-layer network coding scheme for two-way relay channels with asymmetric data length," *Wireless Netw.*, pp. 1–8, Feb. 2018.

[8] A. Liu, G. Liu, and B. Li, "Parameters estimation of precession cone target based on micro-Doppler spectrum," *Wireless Netw.*, pp. 1–7, Feb. 2018.

[9] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP- 34, no. 3, pp. 276–280, Mar. 1986.

[10] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[11] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[12] H.-T. Wu, J.-F. Yang, and F.-K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1325–1333, Jun. 1995.

[13] R. F. Brcich, A. M. Zoubir, and P. Pelin, "Detection of sources using bootstrap techniques," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 206–215, Feb. 2002.

[14] Y. Zhang, N. Hu, and Z. Ye, "A source enumeration method based on subspace orthogonality and bootstrap technique," *Signal Process.*, vol. 93, no. 4, pp. 972–976, Apr. 2013.

[15] S. A. Cummer and U. S. Inan, "Modeling ELF radio atmospheric propagation and extracting lightning currents from ELF observations," *Radio Sci.*, vol. 35, no. 2, pp. 385–394, Mar./Apr. 2000.

[16] R. Pieren, L. Bertsch, D. Lauper, and B. Schäffer, "Improving future low-noise aircraft technologies using experimental perception-based evaluation of synthetic flyovers," *Sci. Total Environ.*, vol. 692, pp. 68–81, Nov. 2019.

[17] S. Roy, T. M. Duman, and V. K. McDonald, "Error rate improvement in underwater MIMO communications using sparse partial response equalization," *IEEE J. Ocean. Eng.*, vol. 34, no. 2, pp. 181–201, Apr. 2009.
[18] L. Na, H. X. Sun, C. Ken, and X. Y. Kuai, "Impulse noise and large scale single frequency noise mitigation for underwater acoustic OFDM," in *Proc. OCEANS*, Taipei, Taiwan, Apr. 2014, pp. 1–4.
[19] M. A. Montes, A. K. Vuorenkoski, and F. R. Dalgleish, "An interpretation of underwater LiDAR waveforms based on a modified Weibull probability distribution function," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Beijing, China, Jul. 2016, pp. 3774–3776.
[20] L. Liu, X. Dong, W. Lin, J. Zhu, and D. Zhu, "Spatial resolution and precision properties of scatterometer reconstruction algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2372–2382, May 2017.
[21] G. R. Wang, H. G. Han, S. Y. Kim, and T. W. Kim, "Wireless vital sign monitoring using penetrating impulses," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 1, pp. 94–96, Jan. 2017.
[22] S. Chikhalikar, O. Khandekar, and C. Bhattacharya, "Design of real-time acquisition and filtering for MEMS-based accelerometer data in microcontroller," in *Proc. IEEE Electron Devices Kolkata Conf.*, Kolkata, India, Nov. 2018, pp. 15–18.
[23] J. Lin, Z. Song, N. Qi, W. Rhee, Z. Wang, and B. Chi, "A 77-GHz mixed-mode FMCW signal generator based on bang-bang phase detector," *IEEE J. Solid-State Circuits*, vol. 53, no. 10, pp. 2850–2863, Oct. 2018.
[24] A. Krasovsky, S. Vasyukov, and I. Murzin, "The use of power line communication technology for electronic relays of car alarms," in *Proc. Int. Conf. Ind. Eng., Appl. Manuf. (ICIEAM)*, Sochi, Russia, Mar. 2019, pp. 1–5.
[25] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*. Hoboken, NJ, USA: Wiley, 1995, pp. 70–96.
[26] J. Dai and H. C. So, "Sparse Bayesian learning approach for outlier-resistant direction-of-arrival estimation" *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 744–756, Feb. 2018.
[27] G. Pok and K. H. Ryu, "Efficient block matching for removing impulse noise," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1176–1180, Aug. 2018.
[28] E. Axell, P. Eliardsson, S. O. Tengstrand, and K. Wiklundh, "Power control in interference channels with class A impulse Noise," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 102–105, Feb. 2017.
[29] A. Awad, "Impulse noise reduction in audio signal through multi-stage technique," *Eng. Sci. Technol., Int. J.*, vol. 22, no. 2, pp. 629–636, Apr. 2019.
[30] F. Rouissi, A. J. H. Vinck, H. Gassara, and A. Ghazel, "Improved impulse noise modeling for indoor narrow-band power line communication," *Int. J. Electron. Commun.*, vol. 103, pp. 74–81, May 2019.
[31] L. Jin, Z. Zhu, E. Song, and X. Xu, "An effective vector filter for impulse noise reduction based on adaptive quaternion color distance mechanism," *Signal Process.*, vol. 155, pp. 334–345, Feb. 2019.
[32] Z. Zhang, D. Han, J. Dezert, and Y. Yang, "A new adaptive switching median filter for impulse noise reduction with pre-detection based on evidential reasoning," *Signal Process.*, vol. 147, pp. 173–189, Jun. 2018.
[33] Z. Lu, Y. Chakhchoukh, and A. M. Zoubir, "Source number estimation in impulsive noise environments using bootstrap techniques and robust statistics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2712–2715.
[34] L. Li, T.-S. Qiu, and D.-R. Song, "Parameter estimation based on fractional power spectrum under alpha-stable distribution noise environment in wideband bistatic MIMO radar system," *Int. J. Electron. Commun.*, vol. 67, no. 11, pp. 947–954, Nov. 2013.
[35] C. Shi, S. Zheng, Y. Lin, and W. Li, "Dynamic threshold-setting for RF-powered cognitive radio networks in non-Gaussian noise," *Phys. Commun.*, vol. 27, pp. 99–105, Apr. 2018.
[36] M. Rupi, P. Tsakalides, E. D. Re, and C. L. Nikias, "Constant modulus blind equalization based on fractional lower-order statistics," *Signal Process.*, vol. 84, no. 5, pp. 881–894, May 2004.
[37] Z. Chen, X. Geng, and F. Yin, "A harmonic suppression method based on fractional lower order statistics for power system," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3745–3755, Jun. 2016.
[38] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[39] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
[40] R. K. Ando and T. Zhang, "A framework for learning predictive structures from multiple tasks and unlabeled data," *J. Mach. Learn. Res.*, vol. 6, pp. 1817–1853, Nov. 2005.
[41] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[42] K.-R. Müller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.
[43] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.
[44] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995, pp. 63–66.
[45] J. B. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Berkeley, CA, USA, 1967, pp. 281–297.

JIAQI ZHEN (M'18) was born in Harbin, Heilongjiang, China, in 1981. He received the master's and Ph.D. degrees from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, in 2007 and 2011, respectively.

From 2011 to 2015, he was a Lecturer in the major of communication engineering with Heilongjiang University, Harbin. Since 2015, he has been an Associate Professor in the major of Internet of Things engineering with Heilongjiang University, and the Director of the Institute of Instrumentation. He has published an academic monograph, two textbooks, more than 30 articles, and 15 inventive patents. His research interests include spatial spectrum estimation, indoor positioning, array signal parameter estimation, and embedded system designs. He received the Best Paper at the 2nd EAI International Conference on Advanced Hybrid Information Processing, in 2018.