IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 4, 2018, accepted August 14, 2018, date of publication August 31, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2868224

Automatic Modulation Classification of Overlapped Sources Using Multi-Gene Genetic Programming With Structural Risk Minimization Principle

SAI HUANG[1,2], (Member, IEEE), YIZHOU JIANG[1], (Student Member, IEEE), XIAOQI QIN[1], (Member, IEEE), YUE GAO[3], (Senior Member, IEEE), ZHIYONG FENG[1], (Senior Member, IEEE), AND PING ZHANG[1], (Senior Member, IEEE)

[1]Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Ministry of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[3]School of Electrical Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.

Corresponding author: Sai Huang (huangsai@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801052 and Grant 61525101, in part by the National Key Research and Development Program of China under Grant 2018YFF0301202, in part by the National Key Technology R&D Program of China under Grant 2015ZX03002008, in part by the Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Ministry of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, under Grant KF20181901, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018RC02.

ABSTRACT As the spectrum environment becomes increasingly crowded and complicated, primary users may be interfered by secondary users and other illegal users. Automatic modulation classification (AMC) of a single source cannot recognize the overlapped sources. Consequently, the AMC of overlapped sources attracts much attention. In this paper, we propose a genetic programming-based modulation classification method for overlapped sources (GPOS). The proposed GPOS consists of two stages, the training stage, and the classification stage. In the training stage, multi-gene genetic programming (MGP)-based feature engineering transforms sample estimates of cumulants into highly discriminative MGP-features iteratively, until optimal MGP-features (OMGP-features) are obtained, where the structural risk minimization principle (SRMP) is employed to evaluate the classification performance of MGP-features and train the classifier. Moreover, a self-adaptive genetic operation is designed to accelerate the feature engineering process. In the classification stage, the classification decision is made by the trained classifier using the OMGP-features. Through simulation results, we demonstrate that the proposed scheme outperforms other existing methods in terms of classification performance and robustness in case of varying power ratios and fading channel.

INDEX TERMS Automatic modulation classification, cumulant, multi-gene genetic programming, overlapped signal classification, structural risk minimization principle.

I. INTRODUCTION

The rapid growth of the Internet of Things and mobile services leads to the large occupation of spectrum resources and aggravates the complication of spectrum environment [1], [2]. Cognitive radio (CR), in which radio senses the spectrum band and operates services on the best wireless channels in its vicinity to avoid user interference and congestion, is proposed to relieve the spectrum resource tension [3], [4]. In CR, simple signal detection techniques cannot satisfy the demand for spectrum sensing due to primary user emulation attacks [6]. Aiming at ensuring the proper function of CR networks, automatic modulation classification (AMC), which is an intermediate step between signal detection and demodulation, is viewed as a promising technology to identify the modulation formats of the target signals corrupted by noise or interference. AMC has many civilian and military applications including tracking the activities of specific users, such as interferer or jammer,

defeating primary user emulation attacks and supporting the demodulation of target signals [5].

Generally, AMC techniques are divided into two categories, i.e., likelihood based (LB) methods and feature based (FB) methods [7]. The former consider AMC as a hypothesis testing problem and yield the optimal solution in the Bayesian sense by minimizing the probability of misclassification [8]. However, calculating the probability of observed signals is computationally intensive especially when plenty of unknown parameters are considered. The latter extract a series of statistical features from received signals and achieve sub-optimal classification performance [9]. Due to the low computational complexity and strong robustness to specific transmission impairments, FB methods are widely exploited in practice and the sub-optimal classification performance can be compensated by increasing the signal symbol length. Higher order statistics [10] and cyclostationary spectrum [11], [12] are two commonly utilized features. Cumulants with proper orders are capable of distinguishing different classes of modulation formats, such as phase shift keying (PSK) and quadrature amplitude modulation (QAM), and they can also discriminate different orders of modulation within the same class, such as 16-QAM and 64-QAM. While cyclostationary spectrum is unable to classify the signal formats such as higher PSKs and QAMs. Moreover, cumulants possess stronger robustness to noise and model mismatches such as phase jitter, phase offset and frequency offset. Therefore, cumulants are utilized as the original classification features in this work.

Cumulant based AMC methods are widely investigated in existing literature [13]–[15], but most of them only utilize the cumulant features individually and directly [13], [14]. As Dandawate and Giannakis [16] verified that different cumulants maintain mutual independence although have a degree of statistical correlation, AMC using multiple cumulants may provide performance improvement. On this base, Swami *et al*. proposed a hierarchical modulation classification method using multiple cumulants in [14], which enlarges the theoretical discrepancy between different modulation formats and therefore improves the classification accuracy. However, minimum distance criterion leads to sub-optimal performance. Several very-high-order cumulants was utilized to recognize PSKs and QAMs with various orders by threshold comparing in [15]. The estimation variances of these very-high-order cumulants increase with theoretical distances, which limits the classification performance. As discussed above, on one hand, most multi-cumulant based classifiers directly utilize original cumulant features without feature engineering, which limits the classification performance. On the other hand, modulation classifiers need to be well designed to strengthen the robustness.

On this base, feature engineering is conducted to filter redundant features and construct highly discriminative features in a small amount of literature. An AMC system combining polynomial classifier and stepwise regression is proposed in [17], where stepwise regression is used to select key features from high-order cumulants and their second order polynomial expansions. However, this method does not produce satisfactory gain in classification accuracy due to the flawed feature engineering. Aslam *et al*. [18] proposed a feature engineering method named GPNN using single-gene genetic programming with k-nearest neighbors (KNN), which achieves acceptable classification accuracy. However, the performance gain carried by the feature engineering is limited due to the dimension reduction caused by integrating several cumulants into a single feature.

Recently, machine learning theory is viewed as a feasible approach which generates the classifier with high accuracy and strong robustness to model mismatches by exploiting the prior information. Zhou *et al*. [19] utilized support vector machine with Gaussian kernel (GSVM) for modulation classification, which possesses stronger robustness to noise, but it is only applicable to the binary classification problem. Moreover, A deep learning based method was proposed in [20], which exploits a stacked convolutional auto-encoder (SCAE) to extract high-level features from the cumulant patches of signals in fading channels. However, calculating the cumulant patches and training the auto-encoder are computationally complex.

In practice, modulation classification of overlapped sources attracts widespread attention due to the fact that communication signals are often interfered by the hostile signals or the out-of-band emission. However, most AMC researchers consider the identification of single source [12]–[20] and there is only a limited number of studies related to overlapped sources. Zaerin *et al*. proposed a cumulant based multiuser modulation classification method (MUMC) to determine the modulation formats of overlapped sources [21]. The computational complexity is low but the minimum distance criterion is utilized for decision making, which merely achieves sub-optimal classification. Huang *et al*. proposed a feature based AMC framework for multiple overlapped sources, which firstly separates the overlapped sources via blind channel estimation and then conducts a maximum likelihood based multi-cumulant classification for each source [22]. However, the necessary signal separation process aggravates the computational complexity and no feature engineering is considered.

In this paper, we propose a Genetic Programming based modulation classification method for Overlapped Sources named GPOS, which synchronously implements feature engineering and classifier modeling by using the improved multi-gene genetic programming (MGP) with structural risk minimization principle (SRMP). The innovation points of this paper are listed as follows.

1) A machine learning based AMC framework of overlapped sources without signal separation process and a novel modulation classification method named GPOS are proposed.
2) MGP based feature engineering is proposed to transform the cumulant features extracted from received

signals into highly discriminative MGP-features under the guidance of SRMP.

3) An improved MGP with the self-adaptive genetic operation is proposed to accelerate the training of GPOS.
4) Extensive simulations are conducted to verify the superiority of the proposed scheme and the robustness to signal-to-noise ratio (SNR), power ratio and fading channel is also explored.

The rest of this paper is organized as follows. Section II provides an overview of the signal model and the proposed AMC framework of overlapped sources. The definitions of the cumulant features and the foundation of MGP are introduced in Section III. Section IV presents GPOS method including the parameter settings and pseudo-code. Section V exhibits simulation results. Conclusion is drawn in Section VI.

II. SIGNAL MODEL AND AMC FRAMEWORK

A. SIGNAL MODEL

Assume that K transmitters transmit signals simultaneously at the same carrier frequency and only one antenna is utilized at the receiving side. The received overlapped signal (with N symbols) can be expressed as

$$y(n) = \sum_{k=1}^{K} h_k A_k e^{j(2\pi \Delta f_k n + \theta_k + \varphi_{k,n})} x_k(n) + \omega(n), \qquad (1)$$

where $x_k(n)$ denotes the n-th (complex) symbol with the candidate constellation, which comes from the k-th source. h_k denotes the channel coefficient of the k-th source following the Rayleigh distribution and A_k is the corresponding signal amplitude. Δf_k is the frequency offset of the k-th source. $\varphi_{k,n}$ and θ_k are the phase jitter and phase offset between the k-th source and the receiver, both of which have a limited influence on the the magnitude of cumulants due to their robustness to constellation rotation [23]. $\omega(n)$ is assumed as the additive white Gaussian noise (AWGN) with zero mean and variance σ_ω^2. The average power of each symbol is normalized without loss of generality such that the SNR is given by

$$\gamma = \sum_{k=1}^{K} \frac{|h_k A_k|^2}{\sigma_\omega^2}. \qquad (2)$$

Moreover, we assume that the target overlapped sources are mutually independent and each of them is formed by independent symbol belonging to its constellation with equal probability.

The underlying modulation combination of the received signal can be represented as $\mathcal{H}_s = \{X_1, \cdots, X_k, \cdots, X_K\} \in \mathcal{U}$, where X_k denotes the modulation format of the k-th source and is selected from a predefined modulation format set, i.e., $\mathcal{M} = \{X_1, \cdots, X_u, \cdots, X_U\}$. $\mathcal{U} = \{\mathcal{H}_1, \cdots, \mathcal{H}_s, \cdots, \mathcal{H}_S\}$ is the universal set of candidate modulation combinations and S denotes the total number of modulation combinations. Note that the prior probability $P(\mathcal{H}_s)$ is assumed to be equal without loss of generality.

B. AMC FRAMEWORK OF OVERLAPPED SOURCES

Fig. 1 illustrates a GPOS based AMC framework of overlapped sources, which consists of two stages, i.e., the training stage and the classification stage.

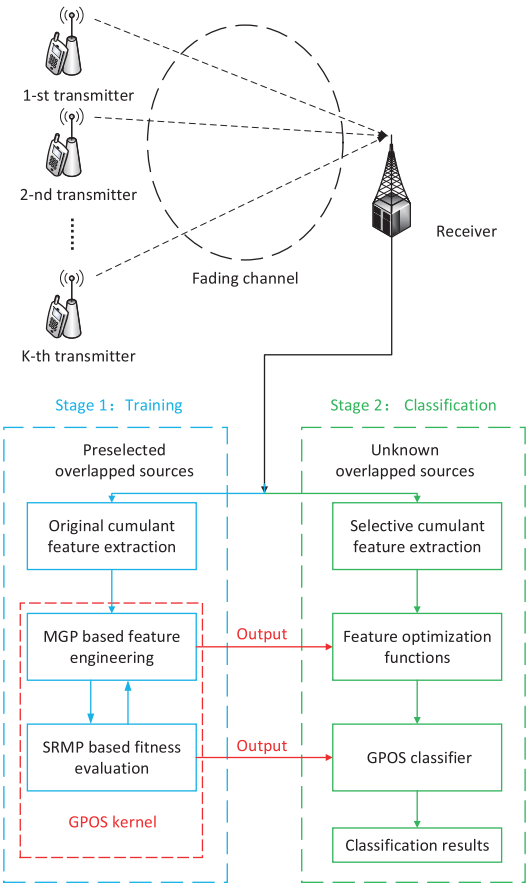


FIGURE 1. The diagram of the proposed AMC framework.

In the training stage, the original cumulant features are firstly extracted from the received signals which are overlapped by K preselected sources with specific modulation formats. Secondly, GPOS kernel conducts feature engineering and classifier modeling synchronously. Specifically, MGP based feature engineering is conducted to transform the cumulant features into a series of more discriminative MGP-features and SRMP based fitness evaluation is proposed to evaluate the classification performance of MGP-features. In the end of the training process, GPOS kernel outputs a group of feature optimization functions (FOFs) which can produce the optimal MGP-features (OMGP-features) with the best evaluation results. Meanwhile, a trained GPOS classifier that matches the FOFs is also output as the final classifier. In the classification stage, the original cumulant features extracted from unknown overlapped signals are transformed into the OMGP-features according to FOFs obtained in the training stage. Then, the trained GPOS classifier recognizes the modulation combination of the overlapped sources using the OMGP-features.

III. BACKGROUND

A. CUMULANT

The general definition of cumulant for a complex stationary random process $y(n)$ is given by

$$C_{\alpha,\beta}(y) = \sum_{q} \left[(-1)^{P-1}(P-1)! \prod_{p=1}^{P} M_{\alpha_p,\beta_p}(y) \right], \quad (3)$$

where $M_{\alpha,\beta}(y) = E((y(n))^{\alpha-\beta} \cdot (y^*(n))^{\beta})$ denotes the α-th-order moment of $y(n)$ with β conjugates and $E(\cdot)$ is the expectation operator. The outmost summation extends over all partitions of the element $\{1, \cdots, \alpha\}$ and each partition has P moments of order α_p and conjugate β_p. In practice, the moment can be estimated using N received symbols, which is given by

$$\hat{M}_{\alpha,\beta}(y) = \frac{1}{N} \sum_{n=1}^{N} (y(n))^{\alpha-\beta} \cdot (y^*(n))^{\beta}. \quad (4)$$

According to expressions of cumulants and moments mentioned above, the sample estimates of cumulants used in our method can be calculated by substituting sample estimates of moments into (3).

Without loss of generality, the constellations are normalized to the unit energy, which means $\hat{C}_{2,1}(y) = 1$, and the normalized cumulants can be given by $\bar{C}_{\alpha,\beta}(y) = \hat{C}_{\alpha,\beta}(y)/(\hat{C}_{2,1}(y))^{\alpha/2}$. Cumulants are utilized in the proposed GPOS method on account of their two advantageous properties. Firstly, the second order and higher order cumulants of Gaussian process are equal to zero, which means that AWGN has a limited influence on them [24]. Secondly, the cumulants of the sum of two statistically independent random processes equals the sum of the cumulants of the independent random processes [25]. This independence property is beneficial to modulation classification of overlapped sources especially. In this work, L-dimensional original feature vector is constructed by L normalized cumulants in various orders, which can be expressed by

$$\mathbf{C}(y) = [\bar{C}_{\alpha_1,\beta_1}(y), \bar{C}_{\alpha_2,\beta_2}(y), \cdots, \bar{C}_{\alpha_L,\beta_L}(y)]^T \in \mathcal{R}^L. \quad (5)$$

It is noted that using multiple cumulants may enlarge the discrepancy between different constellations to facilitate modulation classification and yield stronger robustness to AWGN. Furthermore, the computational complexity of calculating multiple cumulants is still acceptable, since moments are reused for various cumulants. Subsequently, original cumulant features are transformed into more discriminative features for modulation classification by MGP based feature engineering.

B. MULTI-GENE GENETIC PROGRAMMING

Generally speaking, MGP belongs to the category of evolutionary algorithms and originates from genetic algorithms (GAs), which are widely utilized to search high-quality solutions to optimization problems by emulating Darwinian model of natural evolution [26]. As a robust variant of GAs, MGP is designed to be a complex mathematical model with outstanding adaptability and versatility [27]. In MGP, the optimization object is a mathematical formula named individual ($\mathcal{I}$).

Initially, MGP randomly generates a group of individuals to form the population and each individual contains several independent mappings. A mapping in the individual can be characterized in the form of tree, which consists of terminal nodes (inputs) and internal nodes (operators). The elements forming terminal nodes and internal nodes can be respectively selected from two user-defined sets, i.e., terminal set and function set. In GPOS, the former includes original cumulant features and randomly generated constants. The latter determines the characteristics of mappings and nonlinear operators are generally preferred since nonlinear transformation may yield performance gain compared with linear ones. Moreover, it is assumed that all the mappings are mathematically legal and the randomly generated constants follow the uniform distribution between $[-1, 1]$.

Secondly, MGP iteratively optimizes all the mappings in each individual by genetic operation under the guidance of the fitness function. The fitness function is the most pivotal parameter of MGP, which is user-defined depending on the given problem. According to the fitness function, fitness value is calculated to quantify the capability of each individual to solve the given problem. Then, individuals with larger fitness values are retained and further optimized. In this paper, we propose a SRMP based fitness function to promote the optimization process, since SRMP facilitates improving the generalization capability of the classifier to achieve robust classification [28]. Moreover, genetic operation performed by genetic operators produces new individuals by altering the number and the structures of mappings in current individuals. Three kinds of genetic operators are utilized in GPOS, i.e., crossover, mutation and reproduction. The first row in Table 1 shows that mappings in different individuals swap their subtrees to generate the new individuals, which is called crossover. Mutation operator generates new mappings by altering the number of mappings or the structures of the mapping trees randomly, which is shown in the second row in Table 1. Reproduction operator copies the individual from the current population to the next iteration without any changes. At the end of MGP, the best individual is output as the optimal solution to the given problem.

The notations used in this paper are summarized in Table 2.

IV. PROPOSED METHOD

In this section, we introduce GPOS method including its training stage and classification stage in detail. Then, the parameter settings and the pseudo-code of GPOS are exhibited.

A. TRAINING STAGE OF GPOS

In the training stage of GPOS, the feature engineering and classifier modeling are synchronously conducted using MGP with SRMP. In each iteration, MGP based feature

TABLE 1. The diagram of crossover and mutation in MGP.

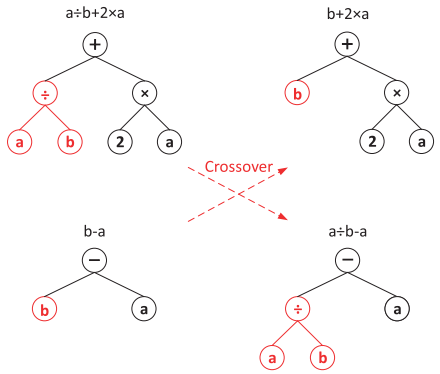
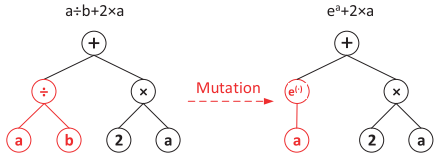
	Original mappings in the individual	New mappings in the individual
Crossover		
Mutation		

TABLE 2. Symbol table.

Symbol	Definition
K	The total number of transmitting sources
N	The signal symbol length
h_k	The channel effect on the power of the k-th source
A_k	The signal amplitude of the k-th source
Δf_k	The frequency offset of the k-th source
$\varphi_{k,n}$	The phase jitter of the k-th source
θ_k	The phase offset of the k-th source
γ	Signal-to-noise ratio (SNR)
X_k	The modulation format of the k-th source
$\mathcal{H}_s$	The modulation combination of the overlapped signal
S	The total number of modulation combinations
M	The number of training datapoints under hypothesis $\mathcal{H}_s$
$\mathcal{Q}$	The population in MGP
$\mathcal{I}^{(z)}$	The z-th individual in the population
$\mathcal{D}$	The training set with cumulant features
$\mathcal{T}^{(z)}$	The training set with MGP-features transformed by $\mathcal{I}^{(z)}$
ϕ	Feature optimization functions (FOFs)
$\mathbf{\Lambda}$	The Lagrange multiplier set
$\alpha_i^{r,v}$	The Lagrange multiplier of the i-th training datapoint for distinguishing $\mathcal{H}_r$ and $\mathcal{H}_v$
$\mathbf{\Omega}$	The parameter set of GPOS classifier
$\mathbf{W}_{r,v}$	The normal vector of the separating hyperplane for distinguishing $\mathcal{H}_r$ and $\mathcal{H}_v$
$b_{r,v}$	The bias of the separating hyperplane for distinguishing $\mathcal{H}_r$ and $\mathcal{H}_v$

engineering transforms the original cumulant features into the MGP-features and optimizes the MGP-features using SRMP based fitness function. It is noted that SRMP based fitness function evaluates the classification performance of MGP-features and models the modulation classifiers at a meanwhile. Moreover, the improved MGP with the self-adaptive genetic operation is proposed to speed up the training process.

At the beginning of training stage, we first generate Z individuals to initialize a population which can be given by $\mathcal{Q} = \{\mathcal{I}^{(1)}, \cdots, \mathcal{I}^{(z)}, \cdots, \mathcal{I}^{(Z)}\}$, where $\mathcal{I}^{(z)}$ is the z-th individual that represents a mapping set and can be expressed as $\mathcal{I}^{(z)} = \{f_1^{(z)}, \cdots, f_d^{(z)}, \cdots, f_{D_z}^{(z)}\}, \forall z \in \{1, \cdots, Z\}$. $f_d^{(z)}$ represents the d-th mapping and D_z denotes the total number of mappings in the z-th individual. Note that D_z varies dynamically during the iterations in the training stage. In each iteration, these mappings are utilized to conduct the feature engineering by transforming the original training set into the MGP-feature space from the cumulant feature space. This process can be given by $\mathcal{T}^{(z)} = \mathcal{I}^{(z)}(\mathcal{D}), \forall z \in \{1, \cdots, Z\}$, where $\mathcal{D} = \{\mathbf{C}_m|\mathcal{H}_s, \forall m \in \{1, \cdots, M\}, \forall s \in \{1, \cdots, S\}\}$ is the original training set and $\mathcal{T}^{(z)} = \{\mathbf{V}_m^{(z)}|\mathcal{H}_s, \forall m \in \{1, \cdots, M\}, \forall s \in \{1, \cdots, S\}\}$ is the transformed training set. $\mathbf{C}_m|\mathcal{H}_s$ represents a datapoint which is the cumulant feature vector extracted from the m-th overlapped signal under hypothesis $\mathcal{H}_s$. $\mathbf{V}_m^{(z)}|\mathcal{H}_s$ is the MGP-feature vector expressed as

$$\begin{aligned}&\mathbf{V}_m^{(z)}|\mathcal{H}_s\\&\quad= \mathcal{I}^{(z)}(\mathbf{C}_m|\mathcal{H}_s)\\&\quad= [f_1^{(z)}(\mathbf{C}_m|\mathcal{H}_s), \cdots, f_d^{(z)}(\mathbf{C}_m|\mathcal{H}_s), \cdots, f_{D_z}^{(z)}(\mathbf{C}_m|\mathcal{H}_s)]^T\\&\quad\in \mathcal{R}^{D_z},\end{aligned} \tag{6}$$

where the superscript T denotes the transpose operator. Moreover, M is the number of training datapoints under hypothesis $\mathcal{H}_s$. After all individuals in the current population are executed, a group of transformed training sets with different MGP-features are obtained.

In GPOS, we define the classification performance of MGP-features transformed by the z-th individual as the fitness value of the z-th individual. SRMP based fitness function is proposed to evaluate the fitness value of each individual using the transformed training sets, which can be derived as

$$\begin{aligned}&J(\mathcal{I}^{(z)})\\&\quad= -\min_{\mathbf{\Omega},\,\Xi} \sum_{r=1}^{S-1}\sum_{v=r+1}^{S}\left(\frac{||\mathbf{W}_{r,v}||^2}{2} + \beta\sum_{m=1}^{2M}\xi_m^{r,v}\right)\\&\qquad s.t.\ (\mathbf{W}_{r,v}\cdot\mathcal{I}^{(z)}(\mathbf{C}_m|\mathcal{H}_s) + b_{r,v})\Upsilon(\mathbf{C}_m|\mathcal{H}_s, \mathcal{H}_r, \mathcal{H}_v)\\&\qquad\quad + \xi_m^{r,v} \geq 1,\\&\qquad\quad \xi_m^{r,v} \geq 0, \quad \forall \mathbf{C}_m|\mathcal{H}_s \in \mathcal{D}_r \cup \mathcal{D}_v,\\&\qquad\quad \forall r, v \in \{1, \cdots, S\}, r < v, \quad \forall m \in \{1, \cdots, 2M\},\end{aligned} \tag{7}$$

where the item in the accumulation indicates the structural risk of the binary classification problem, which consists of the margin loss and regularization loss of the separating hyperplane. It is clear that the lower the structural risk is, the greater the fitness value of the individual is. Moreover, $\mathbf{\Omega} = \{(\mathbf{W}_{r,v}, b_{r,v}), \forall r, v \in \{1, \cdots, S\}, r < v\}$ represents the parameter set of GPOS classifier, where $\mathbf{W}_{r,v}$ and $b_{r,v}$ are the normal vector and the bias respectively.

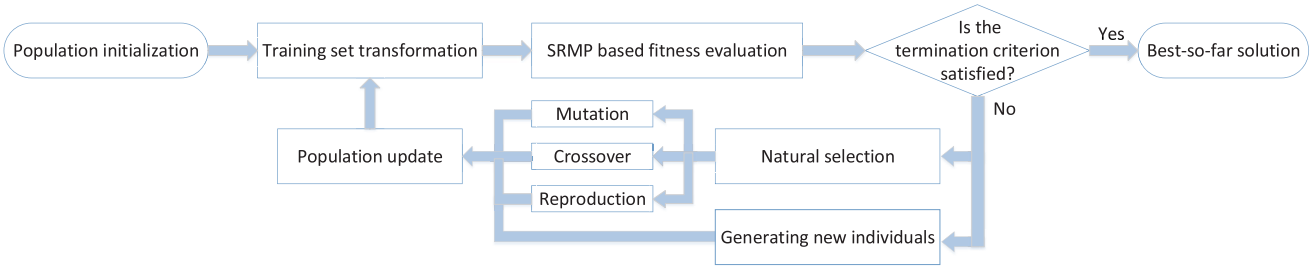


FIGURE 2. The diagram of the training stage of GPOS.

These two parameters jointly determine the optimal separating hyperplane, which distinguishes the overlapped signals under hypotheses $\mathcal{H}_r$ and $\mathcal{H}_v$ in the MGP-feature space. $\Xi = \{\xi_m^{r,v}, \forall r, v \in \{1, \cdots, S\}, r < v, \forall m \in \{1, \cdots, 2M\}\}$ is the set of slack variables, which strengthens the robustness to linearly non-separable classification problems. $\beta > 0$ is the penalty factor for adjusting the regularization, which can prevent the classifier from overfitting. $\mathcal{D}_r = \{\mathbf{C}_m|\mathcal{H}_r, \forall m \in \{1, \cdots, M\}\}$ is a subset of the original training set, which contains all cumulant feature vectors under the hypothesis $\mathcal{H}_r$. Υ is the indicator function which indicates the underlying hypothesis of the overlapped signal, which is defined as

$$\Upsilon(\mathbf{C}_m|\mathcal{H}_s, \mathcal{H}_r, \mathcal{H}_v) = \begin{cases} 1 & \mathcal{H}_s = \mathcal{H}_r \\ -1 & \mathcal{H}_s = \mathcal{H}_v. \end{cases} \tag{8}$$

Note that $\Upsilon_{m,s,r,v}$ is short for $\Upsilon(\mathbf{C}_m|\mathcal{H}_s, \mathcal{H}_r, \mathcal{H}_v)$. Since the Lagrange multiplier method [29] can be utilized to search the maxima or minima of this related problem, the fitness function in the dual form is given by (9), as shown at the bottom of this page, where $\mathbf{\Lambda} = \{\alpha_m^{r,v}, \forall r, v \in \{1, \cdots, S\}, r < v, \forall m \in \{1, \cdots, 2M\}\}$ represents the Lagrange multiplier set and $\cdot$ is the operator of inner product. The fitness value of the z-th individual and the corresponding Lagrange multiplier set can be obtained by using sequential minimal optimization method [30].

According to the fitness function described above, fitness values of all individuals are calculated and form a set, i.e., $\mathcal{F} = \{F^{(1)}, \cdots, F^{(z)}, \cdots, F^{(Z)}\}$, where $F^{(z)} = J(\mathcal{I}^{(z)})$. The individual with the largest fitness value is selected as the best-so-far solution of GPOS, which can be expressed as

$$\mathcal{I}_{best} = \mathcal{I}^{\left(\arg\max\limits_{z} F^{(z)}\right)}, \quad s.t.\ F^{(z)} \in \mathcal{F}. \tag{10}$$

The training process of GPOS comes to an end when the fitness value of the best-so-far solution is above the given threshold or stays static for continuous several iterations. The mappings in the best-so-far solution are output as FOFs (ϕ) and the corresponding Lagrange multipliers are retained to construct GPOS classifier. Otherwise, the population goes into the next iteration, which can be conducted in the following three steps.

1) A proportion of individuals with larger fitness values are retained and the others are eliminated, which is named natural selection.
2) The proposed self-adaptive genetic operation (Γ) is performed to generate highly fit individuals based on the retained individuals.
3) To keep the size of the population consistent, these highly fit individuals and other randomly generated individuals make up new population for the next iteration.

The diagram of the training process of GPOS is shown in Fig. 2.

B. SELF-ADAPTIVE GENETIC OPERATION

Traditional genetic operation in MGP applies the genetic operators with a fixed probability, resulting in low efficiency of the training process [31]. In GPOS, we propose the self-adaptive genetic operation which balances the tradeoff between convergence and efficiency of the training process. This mechanism can contribute to seek appropriate probability of using each genetic operator at different phases of the training process. The self-adaptive genetic operation is described as

$$P_c^{(z)} = \begin{cases} \left(P_{c_1} - \dfrac{(P_{c_1} - P_{c_2})(F^{(z)} - F_a)}{F_m - F_a}\right) \cdot \tau, & F^{(z)} > F_a \\ P_{c_1} \cdot \tau, & F^{(z)} \le F_a \end{cases}$$

$$J(\mathcal{I}^{(z)}) = \max_{\Lambda} \sum_{r=1}^{S-1} \sum_{v=r+1}^{S} \left\{ \sum_{m=1}^{2M} \alpha_m^{r,v} - \frac{1}{2} \sum_{i=1}^{2M} \sum_{j=1}^{2M} \alpha_i^{r,v} \alpha_j^{r,v} \Upsilon_{i,s,r,v} \Upsilon_{j,s,r,v} \left(\mathcal{I}^{(z)}(\mathbf{C}_i|\mathcal{H}_s) \cdot \mathcal{I}^{(z)}(\mathbf{C}_j|\mathcal{H}_s) \right) \right\}$$

$$s.t. \sum_{m=1}^{2M} \alpha_m^{r,v} \Upsilon_{m,s,r,v} = 0, 0 \le \alpha_m^{r,v} \le \beta, \forall \mathbf{C}_m|\mathcal{H}_s, \mathbf{C}_i|\mathcal{H}_s, \mathbf{C}_j|\mathcal{H}_s \in \mathcal{D}_r \cup \mathcal{D}_v,$$
$$\forall r, v \in \{1, 2, \cdots, S\}, r < v, \quad \forall m, i, j \in \{1, 2, \cdots, 2M\}, \tag{9}$$

$$P_m^{(z)} = \begin{cases} \left(P_{m_1} - \dfrac{(P_{m_1} - P_{m_2})(F^{(z)} - F_a)}{F_m - F_a}\right) \cdot \tau, & F^{(z)} > F_a \\ P_{m_1} \cdot \tau, & F^{(z)} \leq F_a \end{cases}$$
$$P_r^{(z)} = 1 - P_c^{(z)} - P_m^{(z)}, \tag{11}$$

where $P_c^{(z)}$, $P_m^{(z)}$ and $P_r^{(z)}$ are respectively the probability of crossover, mutation and reproduction operated on the z-th individual. F_m and F_a are the largest and average fitness values of the current population respectively. τ is the variation factor and $\tau = \tau' e^{(F_a' - F_a)}$, where τ' and F_a' are the variation factor and average fitness value in the last iteration. In addition, P_{c_1}, P_{c_2}, P_{m_1} and P_{m_2} are user-defined parameters, where $0 < P_{c_2} < P_{c_1} < 1$, $0 < P_{m_2} < P_{m_1} < 1$.

According to the self-adaptive genetic operation, the probability that the individual is changed by using some genetic operator depends on its own fitness value and the average fitness values of the population. The advantages of the proposed self-adaptive genetic operation can be explained from two aspects. Firstly, the smaller the fitness value of the individual in the current population is, the greater the probability of change is. It can accelerate the optimization of all the individuals and reduce the probability of breaking the structures of highly fit individuals. Secondly, the probability of change is greater in the early phase of training, which prevents the training process from prematurely converging to local optimum. As the population evolves, the average fitness value increases, indicating that the probability of change gradually decreases, which can protect the prominent individuals generated in the long-term training process.

C. CLASSIFICATION STAGE OF GPOS

In the classification stage, the feature transformation and modulation classification are implemented by utilizing the FOFs and the corresponding Lagrange multipliers provided by the training process. According to the majority voting rule [32], the modulation combinations of the unknown overlapped signals can be identified by GPOS classifier with $S(S-1)/2$ optimal separating hyperplanes, which is expressed as $\mathcal{H}_x = \Psi(\mathcal{G}_x)$, where Ψ is a function that chooses one from the set as the final decision made by the majority of optimal separating hyperplanes (not absolute majority). Moreover, $\mathcal{G}_x = \{g_x^{r,v} = h(\mathbf{C}_x|\mathbf{W}_{r,v}, b_{r,v}, \mathcal{I}_{best}), \forall r, v \in \{1, \cdots, S\}, r < v\}$ is the set of decisions made by $S(S-1)/2$ optimal separating hyperplanes, where $h(\mathbf{C}_x|\mathbf{W}_{r,v}, b_{r,v}, \mathcal{I}_{best})$ is the decision function that identifies whether the unknown overlapped signal is under the specific hypothesis, which can be written as follow

$$\begin{aligned} g_x^{r,v} &= h(\mathbf{C}_x|\mathbf{W}_{r,v}, b_{r,v}, \mathcal{I}_{best}) \\ &= \begin{cases} \mathcal{H}_r & \mathbf{W}_{r,v} \cdot \mathcal{I}_{best}(\mathbf{C}_x) + b_{r,v} > 0 \\ \mathcal{H}_v & \mathbf{W}_{r,v} \cdot \mathcal{I}_{best}(\mathbf{C}_x) + b_{r,v} < 0 \end{cases} \\ &\qquad \forall r, v \in \{1, \cdots, S\}, r < v. \end{aligned} \tag{12}$$

$\mathbf{W}_{r,v}$ and $b_{r,v}$ are the parameters of the optimal separating hyperplane in OMGP-feature space, which can be given by

$$\begin{aligned} \mathbf{W}_{r,v} &= \sum_{i=1}^{2M} \alpha_i^{r,v} \Upsilon_{i,s,r,v} \mathcal{I}_{best}(\mathbf{C}_i|\mathcal{H}_s) \\ b_{r,v} &= \Upsilon_{j,s,r,v} \\ &\quad - \sum_{i=1}^{2M} \alpha_i^{r,v} \Upsilon_{i,s,r,v} \left(\mathcal{I}_{best}(\mathbf{C}_i|\mathcal{H}_s) \cdot \mathcal{I}_{best}(\mathbf{C}_j|\mathcal{H}_s)\right) \\ &\quad s.t.\ \mathbf{C}_i|\mathcal{H}_s, \mathbf{C}_j|\mathcal{H}_s \in \mathcal{D}_r \cup \mathcal{D}_v, \forall r, v \in \{1, \cdots, S\}, \\ &\qquad r < v, \alpha_i^{r,v}, \quad \alpha_j^{r,v} \in \mathbf{\Lambda}_{best},\ 0 < \alpha_j^{r,v} < \beta, \end{aligned} \tag{13}$$

where $\mathcal{I}_{best}$ represents the individual containing FOFs and $\mathbf{\Lambda}_{best}$ is the Lagrange multiplier set corresponding to the FOFs.

D. PARAMETER SETTINGS

Table 3 shows the parameter settings of GPOS method. These parameters are selected according to some previously suggested values and plenty of trial-and-error experiments. The first five parameters are used to resize the breadth and depth of the searching space, which makes a major influence on the computational complexity and classification performance. Note that a wider searching space provided by the larger values of these five parameters may achieve better performance while aggravating the complexity of training process. Moreover, in order to avoid the non-convergence, GPOS is terminated and outputs the best-so-far solution when the training reaches the maximum number of iterations. The natural selection is a program, by which 95% individuals with larger fitness values are retained and the others are replaced by new individuals generated randomly. Similar to the mutation, this program strengthens the population diversity and enhances the possibility of getting rid of local optimum. The final solution of GPOS is chosen on the basis of a compromise between performance and complexity. The pseudo-code of GPOS is described in Algorithm 1.

TABLE 3. The parameter settings of GPOS method.

Parameters	Settings
Population size (Z)	200
Maximum number of iterations (C)	100
Maximum number of static iterations (T)	5
Maximum number of mappings in an individual (R)	20
Maximum tree depth	5
Function set	$\left\{ \begin{matrix} +, -, \times, \div, \lvert\cdot\rvert, \sqrt{\cdot}, \\ (\cdot)^2, (\cdot)^3, e^{(\cdot)}, \ln(\cdot) \end{matrix} \right\}$
Terminal set	The cumulant features and randomly generated constants
Natural selection (ϑ)	95% of population retained
Genetic operators	Crossover, mutation, reproduction
$P_{c_1}, P_{c_2}, P_{m_1}, P_{m_2}$	0.9, 0.5, 0.1, 0

Algorithm 1 GPOS Method

Require: the original training set with cumulant features ($\mathcal{D}$)
Ensure: FOFs (ϕ) and the Lagrange multiplier set ($\mathbf{\Lambda}_{best}$)

$a = 0$; $b = R$; $c = 0$; $t = 0$;
randomly initialize $\mathcal{Q} = \{\mathcal{I}^{(1)}, \cdots, \mathcal{I}^{(z)}, \cdots, \mathcal{I}^{(Z)}\}$;
while $(c \leq C) \wedge (t \leq T)$ **do**
  $\mathcal{F} = \varnothing$;
  for $z = 1 : Z$ **do**
    $\mathcal{T}^{(z)} = \mathcal{I}^{(z)}(\mathcal{D})$;
    $F^{(z)} = J(\mathcal{I}^{(z)})$;
    $\mathcal{F} = \mathcal{F} \cup \{F^{(z)}\}$;
  end for
  $\mathcal{I}_{best} = \mathcal{I}^{\left(\arg\max_z F^{(z)}\right)}, s.t.\ F^{(z)} \in \mathcal{F}$;
  $a' \leftarrow$ the fitness value of $\mathcal{I}_{best}$;
  $b' \leftarrow$ the number of mappings in $\mathcal{I}_{best}$;
  if $(a' > a) \vee ((a' == a) \wedge (b' < b))$ **then**
    $t = 0$; $a = a'$; $b = b'$;
    $\phi = \mathcal{I}_{best}$;
  else
    $t = t + 1$;
  end if
  randomly generate a set of new individuals, $\mathcal{Q}'$;
  $\mathcal{Q} = \Gamma(\vartheta(\mathcal{Q})) + \mathcal{Q}'$;
  $c = c + 1$;
end while
$\mathbf{\Lambda}_{best} = \arg J(\phi)$;

V. SIMULATIONS

In this section, extensive simulations are conducted to verify the performance of the proposed GPOS method. We first analyze the performance gain and convergence of GPOS and then demonstrate its superiority and robustness compared with other methods. We assume there are two transmitters and their modulation formats are selected from $\mathcal{M} = \{\text{BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM}\}$ with equal prior probability. Therefore, there are fifteen kinds of modulation combinations to form the universal set of overlapped sources, i.e., $\mathcal{U} = \{\{X_e, X_f\}, \forall X_e, X_f \in \mathcal{M}\}$, and $S = 15$. Unless specified otherwise, the power ratio is set as 0 dB, indicating that the power of each source is equal, and the channel effects are assumed to be known at the receiver. A set of cumulants $\mathcal{C} = \{C_{2,0}, C_{4,0}, C_{4,1}, C_{4,2}, C_{6,0}, C_{6,1}, C_{6,2}, C_{6,3}, C_{8,0}, C_{8,4}\}$ is chosen to form the ten-dimensional cumulant feature vector. On this base, 500 feature vectors under each hypothesis are obtained to form the training set. Ten thousand Monte Carlo trials are performed to obtain the probability of correct classification respectively in condition that the symbol lengths are 512, 1024, 2048, 4096 and the SNRs range from -5 dB to 15 dB with an interval of 2 dB. The probability of correct classification (P_{cc}) is utilized as the performance metric, which is defined as

$$P_{cc} = \sum_{s=1}^{S} P(\mathcal{H}_s|\mathcal{H}_s)P(\mathcal{H}_s), \mathcal{H}_s \in \mathcal{U}, \tag{14}$$

where $P(\mathcal{H}_s|\mathcal{H}_s)$ is the P_{cc} of overlapped signals under the hypothesis $\mathcal{H}_s$.

A. PERFORMANCE ANALYSIS OF MGP BASED FEATURE ENGINEERING

In order to demonstrate the superiority of MGP based feature engineering, the classification performance of GPOS with and without the feature engineering is shown in Fig. 3. All fifteen kinds of overlapped sources are taken into consideration. Note that GPOS without the feature engineering degenerates into a typical SRMP based classifier using original cumulant features. In Fig. 3, it is clear that MGP based feature engineering can achieve remarkable performance improvement and almost 3 dB gain is achieved at 90% P_{cc} when the symbol length is 2048. It is also found that GPOS can yield higher performance gain using fewer symbols at lower SNR. For example, the P_{cc} rises by 6.30% at -5 dB SNR when symbol length is 1024 while 3.71% improvement of P_{cc} is provided by GPOS at 11 dB SNR when symbol length is 4096. Moreover, the P_{cc} of GPOS increases but the gain gradually reduces with the growth of the signal symbols, proving the convergence of GPOS method.

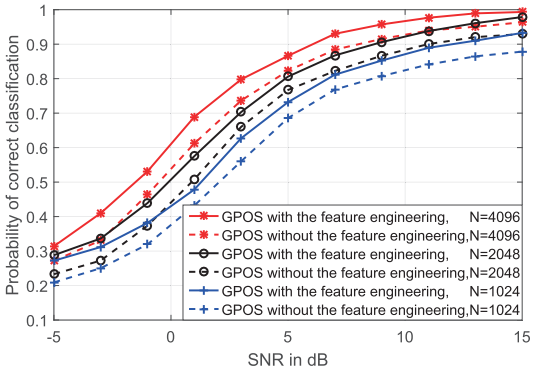


FIGURE 3. The performance comparison between GPOS with and without the feature optimization.

The performance gain yielded by MGP based feature engineering can be explained from two aspects, i.e., the generated features and the dimensionality increasing of feature space, as shown in Table 4. It can be observed that the OMGP-features contribute the major performance gain. Fig. 4 illustrates an example for explanation and shows the datapoint distributions of two kinds of overlapped sources, i.e., {QPSK, 16-QAM} and {QPSK, 64-QAM} in cumulant feature space and the OMGP-feature space respectively. Note that black dotted lines denote the optimal separating hyperplanes provided by GPOS classifier. It is observed that datapoints of each overlapped source gather into a cluster and these clusters overlap with each other slightly in Fig. 4(b), illustrating that feature engineering can enlarge the discrepancy between different modulation combinations and produce better classification performance. Besides, dimensionality increasing occurs and yields certain performance gain when the classification is performed using limited

TABLE 4. The performance gain of the feature engineering.

Symbol length	4096				1024			
SNR (dB)	15	gain	5	gain	15	gain	5	gain
P_{cc} of GPOS with the feature engineering	99.33%	2.98%	86.62%	4.32%	92.72%	4.33%	73.20%	4.64%
P_{cc} of GPOS without the feature engineering	96.35%		82.30%		88.39%		68.56%	
P_{cc} gain achieved by OMGP-features	2.98%		3.01%		3.05%		2.99%	
P_{cc} gain achieved by dimensionality increasing	0.00%		1.31%		1.28%		1.65%	
Dimensionality of the OMGP-feature space	8		14		13		16	

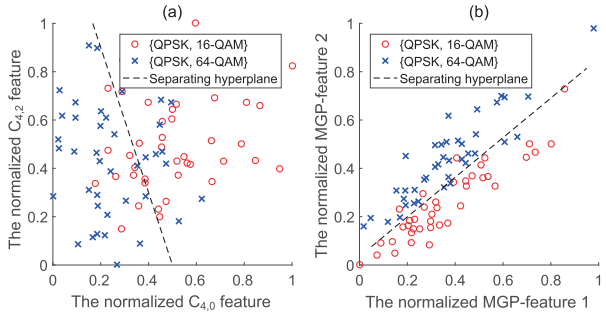


FIGURE 4. The datapoint distributions of {QPSK, 16-QAM} and {QPSK, 64-QAM} in the cumulant feature space (a) and the OMGP-feature space (b) respectively when symbol length is 2048 at 5 dB SNR.

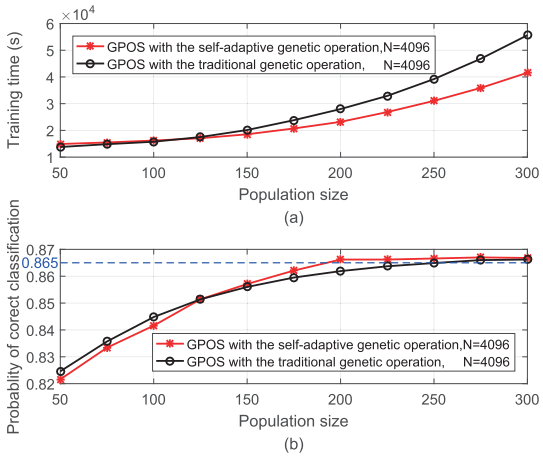


FIGURE 5. The performance comparison between GPOS with and without the self-adaptive genetic operation.

number of signal symbols at low SNRs. This nonlinear dimensionality increasing provides additional improvements for the classification performance of GPOS [33]. Moreover, dimensionality reduction occurs at high SNR scenarios, which reduces the computational complexity without any performance loss. It is noted that the dimensionality of the original cumulant feature space is 10.

B. PERFORMANCE ANALYSIS OF SELF-ADAPTIVE GENETIC OPERATION

In order to illustrate the effectiveness of the self-adaptive genetic operation in GPOS, the performance comparison between GPOS with and without the self-adaptive genetic operation is plotted in Fig. 5. The simulation is conducted at 5 dB SNR and the symbol length is 4096. The probabilities of crossover and mutation are fixed and set as 0.9 and 0.1, respectively in the traditional genetic operation. In terms of training time, Fig. 5(a) illustrates that the gap between the two methods significantly expands with the increase of population size. When the population size is set as 200, GPOS with the self-adaptive genetic operation ends at 4794 seconds(s) earlier than that with the traditional one using the same training set and equipment (a E3-1575 CPU). In Fig. 5(b), the self-adaptive genetic operation reduces over 50 individuals less than the traditional one at $P_{cc} = 86.5\%$. This further demonstrates the advantages of the proposed self-adaptive genetic operation in terms of efficiency and convergence. Moreover, it is found that the proposed method is not as good as we expected when population size is less than 125. This can be explained by the fact that it is difficult for the self-adaptive genetic operation to obtain enough highly fit individuals for further optimization when the search space is small.

C. COMPARISON WITH OTHER AMC METHODS

Fig. 6 reveals the performance comparison among GPOS and the other four cumulant based AMC methods, i.e., GPNN [18], GSVM [19], SCAE [20], and MUMC [21]. The symbol length is set as 4096. GPNN is selected due to its similarity with GPOS, which conducts feature engineering based on single-gene genetic programming with KNN. It is observed that the classification accuracy of GPOS is superior to that of GPNN, owing to the dimensionality increasing achieved by MGP. Moreover, SCAE, MUMC and GSVM perform the modulation classification without any feature engineering, where GSVM maps datapoints into a high-dimensional feature space and constructs separating hyperplanes according to SRMP [34]. It is found that GPOS obviously outperforms the above three methods. When $P_{cc} = 90\%$, GPOS yields 2.5 dB, 5 dB and 7 dB gains than SCAE, MUMC and GSVM respectively. This can be explained

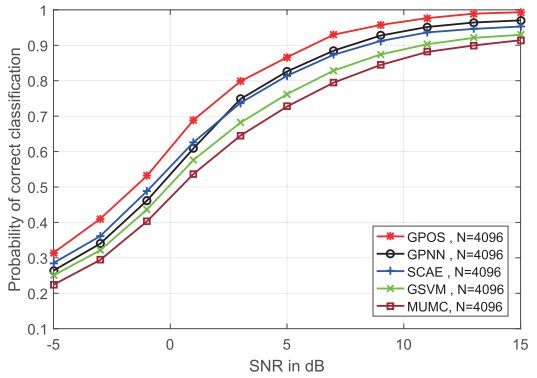


FIGURE 6. The performance comparison between GPOS and the other four AMC methods.

from two aspects. Firstly, MGP based feature engineering conducted by GPOS makes great contribution to the classification accuracy while SCAE and MUMC directly utilize the original cumulant features. Secondly, FOFs are more applicable to the modulation classification problem than the commonly used Gaussian kernel since the FOFs are generated by training in GPOS.

In addition, the computational complexity of classification stage has a major influence on the efficiency of implementing modulation classification and we compare the complexity among GPOS, GPNN and SCAE. Since the classification stage can be divided into two parts, the complexity consists of the complexity of the feature extraction (O_1)[1] and the complexity of the decision making (O_2). It is noted that the O_1 of GPOS is minimum among these methods because the MGP based feature engineering may eliminate redundant original features. Table 5 exhibits the O_2 of these methods, where D represents the dimension of feature vector and n_h is the total number of nodes in hidden layers of SCAE. In general, n_h and M are far larger than D and S. Hence, it is clear that the complexity of GPOS is smaller than that of the other two methods.

TABLE 5. The comparison of the classifier complexity (O_2).

	GPOS	GPNN	SCAE
O_2	$O(\frac{DS(S-1)}{2})$	$O(DSM)$	$O(n_h(D+S))$

D. ROBUSTNESS TO POWER RATIO AND FADING CHANNEL

In the second phase of simulations, we evaluate and compare the robustness of GPOS, SCAE, GPNN and MUMC to power ratio of overlapped sources and unwilling parameters, such as phase jitter, phase offset and frequency offset. In practice, the signals overlapped with each other are likely to possess different receiving power when different transmitting power and distances between the transceivers are considered. The fact that the stronger one overwhelms the weaker one may result in the misclassification result. Therefore, a good AMC method of overlapped sources should be able to achieve robust classification over a wide range of power ratios. The power ratio is defined as

$$\mu = 10\log_{10}\|h_1A_1\|/\|h_2A_2\|. \tag{15}$$

The robustness of these methods versus power ratio is plotted in Fig. 7. The symbol length is 2048 and SNR is set as 15 dB. Note that the SNR is defined as the total power of all signals over noise. To illustrate the influence of different transmitted sources, the five combinations of the different sources with the same modulation format are not taken into account. In Fig. 7, it can be firstly observed that the feasible region (where the P_{cc}s are higher than 90%) of GPOS is much wider than others and is symmetrically centered at 0 dB, illustrating that GPOS performs best when $\mu = 0$ dB and gives consistent classification accuracy as the discrepancy of power extends. Secondly, it can be found that the left side of the feasible region of SCAE is wider than its right side, indicating that SCAE is sensitive to the different combinations of candidate modulation formats. For example, QPSK is easier to be identified than 16-QAM at low SNR. Thirdly, GPOS shows a stronger robustness than MUMC. This can be explained by the fact that setting fixed thresholds is not appropriate for the modulation classification of overlapped sources, because the theoretical values of cumulants of overlapped sources associate with the power ratio.

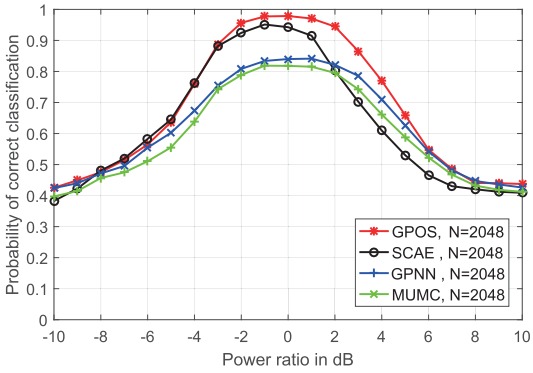


FIGURE 7. The robustness to power ratio of GPOS and the other three AMC methods.

Fig. 8 reveals the classification performance of these methods under the phase jitter. The SNR and symbol length are set as 15 dB and 2048 respectively. The phase of each source is assumed to be uniformly distributed over $[-\Phi, \Phi]$. It is observed that the classification performance degrades as the range of the phase jitter increases. Moreover, the proposed GPOS method is more robust to the phase jitter since the width range is increased by 7° compared with SCAE at 90% P_{cc}.

[1]The operator O represents the complexity of the algorithm, which quantifies the number of basic operation units, such as exponential, logarithmic and complex multiplication [35] .

TABLE 6. The classification performance versus phase offset.

(θ_1,θ_2) / $degree$	(0,0)	(0,15)	(15,15)	(15,-15)	(15,30)	(15,45)
GPOS	97.87%	96.14%	97.54%	89.47%	96.07%	91.74%
SCAE	92.90%	91.70%	91.83%	82.50%	91.37%	85.37%
GPNN	89.90%	87.30%	89.03%	80.83%	86.30%	83.83%
MUMC	85.03%	83.66%	81.57%	78.01%	79.99%	74.23%

TABLE 7. The classification performance versus frequency offset.

$(\Delta f_1,\Delta f_2)\times 10^{-4}$	(0,0)	(0,1)	(1,1)	(1,-1)	(1,2)	(1,3)
GPOS	97.87%	93.40%	91.47%	86.74%	86.60%	78.14%
SCAE	92.90%	84.23%	84.10%	81.90%	81.23%	74.90%
GPNN	89.90%	85.70%	80.77%	76.63%	74.37%	67.57%
MUMC	85.03%	81.74%	78.74%	75.21%	72.57%	63.66%

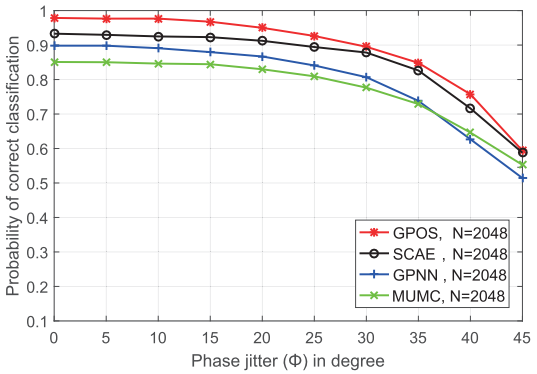


FIGURE 8. The robustness to phase jitter of GPOS and the other three AMC methods.

The phase and frequency offsets are considered separately, indicating that we only consider phase offset and assume the carrier frequency is perfectly estimated and vice versa. The classification performance under the phase and frequency offsets is illustrated and compared in Table 6 and Table 7 respectively. The parameter settings are the same as Fig. 8. In Table 6, it is found that GPOS is superior to the other three methods with some specific phase offset values since MGP based feature engineering generates discriminative and adaptable features. Moreover, it is concluded that the discrepancy between phase offsets aggregates the degradation in classification performance. In Table 7, the frequency offset is normalized by the sampling frequency, which varies from -0.0001 to 0.0003 (corresponding to a maximum rotation of 186°). As expected, we found GPOS performs best and the similar trends as the one concluded in the Table 6. However, the frequency offset severely degrades the classification performance. This can be associated to the dense constellation of the modulation combination of overlapped sources and only limited margin is left for any frequency offset with dense constellation scatter.

VI. CONCLUSIONS

In this paper, we demonstrate a novel GPOS method for identifying the modulation formats of overlapped sources, which can be divided in two stages to improve the classification performance. In the training stage, the feature engineering, which transforms the sample estimates of cumulants into highly discriminative OMGP-features, is conducted by MGP with SRMP. The self-adaptive genetic operation is designed to accelerate the feature engineering process. In the classification stage, the performance is deduced by GPOS using the OMGP-features. Extensive simulations demonstrate that the proposed method yields satisfactory performance gain and achieves stronger robustness to power ratios and fading channel compared to other recent methods.

REFERENCES

[1] Z. Feng, C. Qiu, Z. Feng, Z. Wei, W. Li, and P. Zhang, "An effective approach to 5G: Wireless network virtualization," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 53–59, Dec. 2015.

[2] Y. Gao, Z. Qin, Z. Feng, Q. Zhang, O. Holland, and M. Dohler, "Scalable and reliable IoT enabled by dynamic spectrum management for M2M in LTE-A," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1135–1145, Dec. 2016.

[3] Q. Wu, G. Ding, J. Wang, and Y. D. Yao, "Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: Two-dimensional sensing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 516–526, Feb. 2013.

[4] Y. Ma, Y. Gao, A. Cavallaro, C. G. Parini, W. Zhang, and Y.-C. Liang, "Sparsity independent sub-Nyquist rate wideband spectrum sensing on real-time TV white space," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8784–8794, Oct. 2017.

[5] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 428–445, 1st Quart., 2013.

[6] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.

[7] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[8] J. L. Xu, S. Wei, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 4, pp. 455–469, Jul. 2011.

[9] S. Majhi, R. Gupta, W. Xiang, and S. Glisic, "Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11057–11069, Dec. 2017.

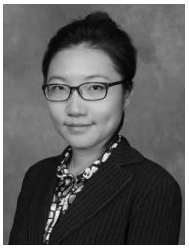
[10] D.-C. Chang and P.-K. Shih, "Cumulants-based modulation classification technique in multipath fading channels," *IET Commun.*, vol. 9, no. 6, pp. 828–835, Apr. 2015.

[11] E. Karami and O. A. Dobre, "Identification of SM-OFDM and AL-OFDM signals based on their second-order cyclostationarity," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 942–953, Mar. 2015.

[12] O. A. Dobre and M. Öner, S. Rajan, and R. Inkol, "Cyclostationarity-based robust algorithms for QAM signal identification," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 12–15, Jan. 2012.
[13] M. S. Mühlhaus, M. Öner, O. A. Dobre, H. U. Jkel, and F. K. Jondral, "Automatic modulation classification for MIMO systems using fourth-order cumulants," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2012, pp. 1–5.
[14] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[15] W. Su, "Feature space analysis of modulation classification using very high-order statistics," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1688–1691, Sep. 2013.
[16] A. V. Dandawate and G. B. Giannakis, "Asymptotic theory of mixed time averages and kth-order cyclic-moment and cumulant statistics," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 216–232, Jan. 1995.
[17] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification using hierarchical polynomial classifier and stepwise regression," in *Proc. IEEE Wireless Conf. Netw. Conf. (WCNC)*, Apr. 2016, pp. 1–5.
[18] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.
[19] X. Zhou, Y. Wu, and B. Yang, "Signal classification method based on support vector machine and high-order cumulants," *Wireless Sensor Netw.*, vol. 2, no. 1, pp. 48–57, 2010.
[20] Z. Zhang, Z. Hua, and Y. Liu, "Modulation classification in multipath fading channels using sixth-order cumulants and stacked convolutional auto-encoders," *IET Commun.*, vol. 11, no. 6, pp. 910–915, Apr. 2017.
[21] M. Zaerin and B. Seyfe, "Multiuser modulation classification based on cumulants in additive white Gaussian noise channel," *IET Signal Process.*, vol. 6, no. 9, pp. 815–823, Dec. 2012.
[22] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, "Automatic modulation classification of overlapped sources using multiple cumulants," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6089–6101, Jul. 2017.
[23] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *IEEE Sensors J.*, vol. 10, no. 11, pp. 1779–1785, Nov. 2010.
[24] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.
[25] D. R. Brillinger, *Time Series: Data Analysis and Theory*, vol. 36. Philadelphia, PA, USA: SIAM, 2001.
[26] J. H. Holland, "Adaptation in natural and artificial systems," in *Quarterly Review of Biology*, Cambridge, MA, USA: MIT Press, 1992, pp. 126–137.
[27] J. Koza and R. Poli, "Genetic programming," in *Search Methodologies*, E. K. Burke and G. Kendall, Eds. New York, NY, USA: Springer, 2005, pp. 127–164.
[28] V. N. Vapnik, "Statistical learning theory," in *Encyclopedia of the Sciences of Learning*. New York, NY, USA: Wiley, 2008, pp. 3185–3194.
[29] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 215–231.
[30] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods, Support Vector Learning*. Cambridge, MA, USA: MIT Press, 2008, pp. 185–208.
[31] C. Hamzaçebi, "Improving genetic algorithms' performance by local search for continuous function optimization," *Appl. Math. Comput.*, vol. 196, no. 1, pp. 309–317, Feb. 2008.
[32] G. B. Marković, "Cooperative modulation classification by using multiple sensors in dispersive fading channels," in *Proc. 22nd Telecommun. Forum Telfor (TELFOR)*, Nov. 2014, pp. 264–271.
[33] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. Electron. Comput.*, vol. 14, no. 3, pp. 326–334, Jun. 1965.
[34] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. ACM 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
[35] A. H. Gandomi and A. H. Alavi, "A new multi-gene genetic programming approach to nonlinear system modeling. Part I: Materials and structural engineering problems," *Neural Comput. Appl.*, vol. 21, no. 1, pp. 171–187, 2012.

SAI HUANG received the Ph.D. degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2017. Since 2017, he has been with the School of Information and Communication Engineering, BUPT, as a Lecturer. His research interests include universal signal detection and identification, millimeter wave signal processing, and energy harvesting networks.

YIZHOU JIANG is currently pursuing the M.S. degree with the Beijing University of Posts and Telecommunications. His research interests include dynamic spectrum management, universal signal detection and identification, and intelligent information processing.



XIAOQI QIN received the B.S., M.S., and Ph.D. degree from the Bradley Department of Electrical and Computer Engineering, Virginia Tech, in 2011, 2013, and 2016, respectively. She is currently a Lecturer with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China. Her research interests include algorithm design and cross-layer optimization in wireless networks, coexistence and spectrum sharing in cognitive radio networks, and intelligent Internet of Things.

YUE GAO (SM'13) received the Ph.D. degree from the Queen Mary University of London (QMUL), U.K., in 2007. He was a Research Assistant, a Lecturer (Assistant Professor), and a Senior Lecturer (Associate Professor) at QMUL. He is currently a Reader in antennas and signal processing, and also the Director of the Whitespace Machine Communication Lab, School of Electronic Engineering and Computer Science, QMUL. He is currently leading a team, developing theoretical research into practice in the interdisciplinary area among smart antennas, signal processing, spectrum sharing and Internet of Things (IoT) applications. He has published over 140 peer-reviewed journal and conference papers, two patents, and two book chapters. He is an Engineering and Physical Sciences Research Council Fellow from 2018 to 2023. He is a co-recipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016, and the Research Performance Award from the Faulty of Science and Engineering, QMUL, in 2017. He is serving as the Cognitive Radio Symposium Co-Chair for the IEEE GLOBECOM 2017. He has served as the Signal Processing for Communications Symposium Co-Chair for the IEEE ICCC 2016, the Publicity Co-Chair for the IEEE GLOBECOM 2016, and the General Chair of the IEEE WoWMoM and iWEM 2017. He is the Secretary of the IEEE Technical Committee on Cognitive Networks, and an IEEE Distinguished Lecturer of the Vehicular Technology Society. He is an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE WIRELESS COMMUNICATION LETTER AND CHINA COMMUNICATIONS.

ZHIYONG FENG (SM'10) received the B.S., M.S., and Ph.D. degrees in information and communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China. She is currently a Full Professor. She is also the Director of the Key Laboratory of Universal Wireless Communications, Ministry of Education. She is also the Technical Advisor of NGMN. Her main research interests include wireless network architecture design and radio resource management in the 5th Generation mobile networks (5G), spectrum sensing and dynamic spectrum management in cognitive wireless networks, universal signal detection and identification, and network information theory. She is active in ITU-R, IEEE, ETSI, and CCSA standards. She is an Editor of the *IET Communications*, and the *KSII Transactions on Internet and Information Systems*. She is a Reviewer of the IEEE TWC, IEEE TVT, and IEEE JSAC.

PING ZHANG received the M.S. degree in electrical engineering from Northwestern Polytechnical University, Xi'an, China, in 1986, and the Ph.D. degree in electric circuits and systems from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1990. He is currently a Professor at BUPT. His research interests include 5th Generation mobile networks (5G), communications factory test instrument, and universal wireless signal detection instrument. He was a recipient of the First and Second Prizes of the National Technology Invention and Technological Progress Awards, as well as the First Prize of the Outstanding Achievement Award of Scientific Research in College. He is the Executive Associate Editor-in-Chief on Information Sciences of the Chinese Science Bulletin, a Guest Editor of the *IEEE Wireless Communications Magazine*, and an Editor of the *China Communications*.