Advanced Features Generation Algorithm for MPSK and MQAM Classification in Flat Fading Channel

Iyad KADOUN, Hossein KHALEGHI BIZAKI

Dept. of Electrical and Computer Engineering, Malek-Ashtar University of Technology, Tehran, Iran

idiyad@yahoo.com, bizaki@yahoo.com

Submitted December 9, 2021 / Accepted February 7, 2022

Abstract. *The Automatic Modulation Classification (AMC) performance depends on the selected features. Conventionally, Higher-Order Cumulants (HOCs) are the well-known features due to their discrimination ability under different channel conditions. HOCs have good performance under the Additive white Gaussian noise (AWGN) channel, but their performance degrades under fading channel. This paper proposes an Advanced Features Generation Algorithm (AFGA) that generates mathematical forms of new features based on the maximum discrimination between the digital modulation types to overcome this performance limitation. These features have similar complexity to HOCs but better performance accuracy. The simulation results show that the proposed AFGA improves the performance accuracy up to 4.5% for a Signal-to-noise ratio (SNR) value of 10 dB under fading channel conditions with respect to conventional methods.*

Keywords

Automatic modulation classification, Feature Selection Algorithms (FSA), higher-order cumulants, Mahalanobis distance (MD)

1. Introduction

Automatic Modulation Classification (AMC) is an important task in modern digital receivers. Two general types of AMC algorithms are Likelihood-Based (LB) and Feature-Based (FB) algorithms. Theoretically, LB algorithms have better performance than FB algorithms, but practically, FB algorithms were used due to their less computational complexity compared to LB.

Various studies were done to find good discriminative features for modulation classification like instantaneous, transformations, and Higher-Order Statistics [1–3]. Comparisons are made among them in [1, 3–5]. It was shown that HOCs have the greatest performance for classifying various digital modulation types under various channel characteristics, such as fading channels.

Various studies were done for AMC under fading channel conditions using HOCs. In [6], the author shows that the performance accuracy of M-array Phase Shift Keying (MPSK) and M-array Quadrature Amplitude shift Modulation (MQAM) classification by using HOMs and HOCs is 84.37%, for SNR value of 10 dB. In [7], the author shows that the performance accuracy of 4-Amplitude-Shift Keying (4-ASK), 8-ASK, Binary Phase-Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), 8-Phase-Shift Keying (8-PSK), 16-PSK, 16-Quadrature Amplitude Modulation (16-QAM), 32-QAM, and 64-QAM classification by using cyclic cumulants is 80%, for SNR value of 10 dB.

The performance accuracy worsens as the SNR value drops, as illustrated in [6], [7]. As a result, alternative features that perform better than cumulants under fading channel circumstances must be discovered. Our study focuses on MPSK (BPSK, QPSK, 8PSK, and 16PSK) and MQAM (8QAM, 16QAM, 32QAM, and 64QAM) digital modulation types. Since the cumulants have the best performance, our simulation results show that C_{40}, C_{61} and C_{80} cumulants have the best performance accuracy for the selected modulations under fading channel conditions.

Based on this opinion, our study proposes an Advanced Features Generation Algorithm (AFGA) that generates the mathematical forms of three features alternatives to C_{40}, C_{61} and C_{80}. Unlike the cumulants which have mathematical forms according to (2), the mathematical representations of the features are generated by AFGA using the maximum discrimination rule between the specified digital modulation types. As a result, they increase performance accuracy. To do this task, feature extraction and feature selection algorithms have been used. AGFA takes into account that the computational complexities of these new features are similar to the computational complexities of C_{40}, C_{61} and C_{80}. The three generated features have better classification accuracy than C_{40}, C_{61} and C_{80}. They could improve the performance accuracy up to 4.5% for SNR value of 10 dB with respect to conventional methods.

2. System Model

For eight selected digital modulation types in Sec. 1, the baseband waveform of the received signal under flat

DOI: 10.13164/re.2022.0127

fading conditions can be written as:

$$y_j(n) = h_l x_j(n) + g(n),\ j = 1,...,8 \tag{1}$$

where $x_j(n)$ is the transmitted symbol of the j digital modulation type, h_l is the complex fading coefficient of the channel and is considered as $h_l \in CN(0,\sigma_h^2)$, and $g(n)$ is a complex AWGN and is considered as $g(n) \in CN(0,\sigma_n^2)$. The general mathematical form of HOCs is[8]:

$$C_{pq} = \mathrm{Cum}\left[s_1,...,s_{p-q},s^*_{p-q+1},...,s^*_p\right] \tag{2}$$

where * denotes the complex conjugate, p is the order of the cumulant and q is the complex conjugate order of the cumulant and the Cum function is defined as [13]:

$$\mathrm{Cum}\left[s_1,...,s_n\right] = \sum_{\forall v}(-1)^{q-1}(q-1)!\,E\left[\prod_{j\in V_1} s_j\right].......E\left[\prod_{j\in V_q} s_j\right] \tag{3}$$

where the summation is being performed on all partitions $V = (V_1,V_2,\ldots,V_q)$ for the set of indexes $(1,2,\ldots,n)$. According to (2) and (3), the mathematical forms of C_{40}, C_{61} and C_{80} are shown in Tab. 1 [9] where $M_{pq} = E[y(k)^{p-q}y^*(k)^q]$ is the moment of the received signal $y(k)$, * denotes the complex conjugate, p is the order of the moment, and q is the complex conjugate order of the moment.

Selected cumulants	$C_{40} = M_{40} - 3M_{20}^2$
	$C_{61} = M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30M_{20}^2M_{21}$
	$C_{80} = M_{80} - 35M_{40}^2 - 28M_{60}M_{20} + 420M_{20}^2M_{40} - 630M_{20}^4$

Tab. 1. Mathematical forms of the selected HOCs.

3. Optimization Problem Definition

By closely examining the mathematical forms of the chosen cumulants in Tab. 1, it is discovered that each cumulant is made up of a sum of terms, with each term being a multiplication of moments of various powers. It has the following general mathematical form:

$$\begin{aligned}
&f_L = \sum_{i=1..U_L} \alpha_{L_i} M_{A_{L_i},k_{L_i}}^{TA_{L_i}} . M_{B_{L_i},l_{L_i}}^{TB_{L_i}} ....M_{C_{L_i},m_{L_i}}^{TC_{L_i}} \\
&\alpha_{L_i} \text{ integers, } 0 < A_{L_i}, B_{L_i},...,C_{L_i} \le L \text{ integers} \\
&0 \le k_{L_i} \le A_{L_i}, 0 \le l_{L_i} \le B_{L_i},..,0 \le m_{L_i} \le C_{L_i} \text{ integers} \\
&TA_{L_i}, TB_{L_i},..,TC_{L_i} \ge 0, \text{ integers} \\
&TA_{L_i} + TB_{L_i} + .. + TC_{L_i} \le \frac{L}{2} \\
&L = 4, L = 6, L = 8,\ U_4 = 2,\ U_6 = 4,\ U_8 = 5
\end{aligned} \tag{4}$$

where A_{L_i}, B_{L_i}, …, C_{L_i} are the orders of moments, k_{L_i}, l_{L_i}, …, m_{L_i} are the orders of the conjugate of the moments, TA_{L_i}, TB_{L_i}, …, TC_{L_i} are the powers of the moments, and U_L is the number of terms of the feature f_L. We suppose the all unknowns are:

$$\zeta_L = \left\{\alpha_{L_i}, A_{L_i}, B_{L_i},..,C_{L_i}, k_{L_i}, l_{L_i},.,m_{L_i}, TA_{L_i}, TB_{L_i},.,TC_{L_i}, U_L\right\}.$$

As an example, for C_{40}, we find that

$$C_{40} = f_4 = \sum_{i=1..T_L} \alpha_{L_i} M_{A_{L_i},k_{L_i}}^{TA_{L_i}} . M_{B_{L_i},l_{L_i}}^{TB_{L_i}} ....M_{C_{L_i},m_{L_i}}^{TC_{L_i}} = M_{40} - 3M_{20}^2$$

with its variables summarized in Tab. 2.

α_1	A_1	k_1	TA_1	$TB_1,\ldots, TC_1$
1	4	0	1	0
α_2	A_2	k_2	TA_2	$TB_2,\ldots, TC_2$
–3	2	0	2	0

Tab. 2. Disassembly of C_{40} according to (4).

Hence, we define a mathematical optimization problem of finding new optimal mathematical forms of three features (4) $\{\hat{f}_4, \hat{f}_6, \hat{f}_8\}$, which are alternatives to C_{40}, C_{61}, C_{80}, based on maximum of the all minimum discrimination measurements among selected digital modulations as shown in (5):

$$\begin{aligned}
&\left\{\hat{f}_4, \hat{f}_6, \hat{f}_8\right\} = \\
&\underset{i}{\mathrm{argmax}}\left\{\min\left\{\mathrm{discrimination}\left\{\left(f_{4,i}, f_{6,i}, f_{8,i}\right)_{j,j=1}^{z}\right\}\right\}\right\}
\end{aligned} \tag{5}$$

where z is the number of classes, i all possible solutions. The discrimination function in (5) is the summation of all the statistical distances (Mahalanobis distances (MD)) between each two different classes as shown in Sec. 4.3 (18). It contains square root function, the covariance matrices, and the means vectors of the classes, which they are calculated for the generated features (4):

$$\begin{aligned}
&\mathrm{discrimination}\left\{\left(f_{4,i}, f_{6,i}, f_{8,i}\right)_{j,j=1}^{z}\right\} = \\
&\left[\mathrm{MD}\left(\left(f_{4,i}, f_{6,i}, f_{8,i}\right)_j, \left(f_{4,i}, f_{6,i}, f_{8,i}\right)_k\right)\right]_{\substack{j,k=1\\ j\ne k}}^{z}.
\end{aligned} \tag{6}$$

This makes it ultra-high complicated function. Optimization problem (5) means of calculating all optimum unknowns (4) for the three features as:

$$\begin{aligned}
&\left\{\hat{\zeta}_4, \hat{\zeta}_6, \hat{\zeta}_8\right\} = \\
&\underset{\zeta_{4,i},\zeta_{6,i},\zeta_{8,i}}{\mathrm{argmax}}\left\{\min\left\{\mathrm{discrimination}\left\{\left(f_{4,i}, f_{6,i}, f_{8,i}\right)_{j,j=1}^{z}\right\}\right\}\right\}.
\end{aligned} \tag{7}$$

This issue is a nonconvex general optimization problem with an extremely sophisticated discriminating function and a large number of unknown parameters. There is no mathematical solution. It is unrealistic to search for the best features by testing all conceivable solutions since it takes a long period. To solve this optimization problem, we propose an Advanced Feature Generation Algorithm (AFGA) that finds these optimum values of the unknown's parameters of each feature in (4) in which maximizes the discrimination ability, and generates their mathematical forms.

4. Mathematical Concepts for Study

4.1 Features Selection Algorithms (FSA)

In classification problems, when the feature space dimension increases, computational complexity increases, too. To overcome this challenge, we must minimize the feature space dimension by picking the best discrimination features and rejecting the rest in such a manner that as much class discrimination as feasible is preserved. For this reason, Feature Selection Algorithms (FSA) were studied [10]. Using FSA, we speed up the learning process, reduce the storage size and improve the learning performance.

Filter methods in feature selection techniques [11] are independent of any learning and classification algorithms, computationally inexpensive, simple, and faster than the other techniques. Hence, these methods were focused on in our study. Filter methods evaluate the score of each feature according to its discrimination ability and choose the features which have the highest scores. Here we mention some of the important filter methods and their mathematical score form:

Fisher Score (FS) [12]: It is calculated as the ratio of between scatter ($\sum_{j=1}^{z} n_j\left(\mu_{x_l,j}-\mu_{x_l}\right)^2$, where z is the number of classes, n_j is the dataset size of class j, μ_{x_l} is the total mean of the feature f_l, and $\mu_{x_l j}$ is the mean of the feature f_l and class j), and within scatter ($\sum_{j=1}^{z} n_j\sigma_{x_l,j}^2$, where $\sigma^2_{x_l j}$ is the variance of feature f_l and class j) as follows:

$$\mathrm{FS}(\boldsymbol{f}_l)=\frac{\sum_{j=1}^{z} n_j\left(\mu_{x_l,j}-\mu_{x_l}\right)^2}{\sum_{j=1}^{z} n_j\sigma_{x_l,j}^2},\ 1\le l\le d \tag{8}$$

where d is the number of features.

Relief-F (RF) [10, 13]: It is an iterative approach that estimates the score of each feature according to the differentiation of data samples which are near to each other. For each point i of class l, i.e. $\boldsymbol{x}_{i,l}$, we define $NH(i)$ are the nearest data of $\boldsymbol{x}_i$ in the same class with size h_i, $NM(i,k)$ are the nearest data of $\boldsymbol{x}_i$ in class k ($k\neq k_i$) with size h_{ik} and probability $p(k)$. The Relief-F Score can be calculated as [10]:

$$\mathrm{RF}(\boldsymbol{f}_l)=\frac{1}{u}\sum_{i=1}^{u}\left(\begin{array}{l}-\frac{1}{h_i}\sum_{x_r\in NH(i)}\mathrm{dm}\left(\boldsymbol{x}_{i,l},\boldsymbol{x}_{r,l}\right)\\ +\sum_{k\neq k_i}\frac{1}{h_{ik}}\frac{p(k)}{1-p(k)}\sum_{x_r\in NM(i,k)}\mathrm{dm}\left(\boldsymbol{x}_{i,l},\boldsymbol{x}_{r,l}\right)\end{array}\right) \tag{9}$$

where dm is defined as [10]:

$$\mathrm{dm}\left(\boldsymbol{x}_{i,l},\boldsymbol{x}_{r,l}\right)=\begin{cases}0 \text{ if } \boldsymbol{x}_{i,l}=\boldsymbol{x}_{r,l}\\ 1 \text{ if } \boldsymbol{x}_{i,l}\neq\boldsymbol{x}_{r,l}\end{cases} \tag{10}$$

and u is data instances that are randomly selected among all m instances.

Pearson Correlation Coefficient (PCC) [14]: It measures the similarity between data samples and its labels. PCC score can be calculated as:

$$\mathrm{PCC}(\boldsymbol{f}_l)=\left|\frac{\sum_{j=1}^{m}\left(x_{j,l}-\mu_{x_l}\right)\left(y_{j,l}-\mu_{y_l}\right)}{\sqrt{\sigma_{x_l}^2\sigma_{y_l}^2}}\right| \tag{11}$$

where μ_{x_l} is the mean value of the feature l and $\sigma^2_{x_l}$ its variance, and μ_{y_l} is the mean value of labels of feature l and $\sigma^2_{y_l}$ its variance.

Laplacian Score (LS) [15]: It selects features which can better preserve the manifold structure of the data (which is a set of points, along with a set of neighborhoods for each point). For each i-th feature, LS score can be calculated as:

$$\mathrm{LS}(\boldsymbol{f}_l)=\frac{\tilde{\boldsymbol{f}}_l^{\mathrm{T}}\mathbf{L}\tilde{\boldsymbol{f}}_l}{\tilde{\boldsymbol{f}}_l^{\mathrm{T}}\mathbf{D}\tilde{\boldsymbol{f}}_l} \tag{12}$$

where $\tilde{\boldsymbol{f}}_l=\boldsymbol{f}_l-\frac{\boldsymbol{f}_l^{\mathrm{T}}\mathbf{D1}}{\mathbf{1}^{\mathrm{T}}\mathbf{D1}}\mathbf{1}$, $\mathbf{D}=\mathrm{diag}(\mathbf{S.1})$, L = D – S, and

$$[S]_{jk}=\begin{cases}\mathrm{e}^{-\frac{\|x_j-x_k\|^2}{t}}; & \text{if } x_k \text{ is among } k \text{ nearest neighbors of } x_j\\ 0 \quad ; & \text{elsewhere}\end{cases}.$$

Term Variance (TV) [16]: It calculates the variance of each feature as:

$$\mathrm{TV}(\boldsymbol{f}_l)=\left(\boldsymbol{f}_l-\boldsymbol{\mu}_l\right)^{\mathrm{T}}*\left(\boldsymbol{f}_l-\boldsymbol{\mu}_l\right). \tag{13}$$

4.2 Optimum Features Weights Calculation

The weights that optimize discrimination across classes are known as optimal weights. The projection vector that optimizes discrimination between various classes is calculated via Fisher Discriminant Analysis (FDA) [17], and the values of this vector represent the optimal weights of the related features. Suppose the input features of p–instance of the j-class as $\boldsymbol{x}_{j,p}=\{x_{j,p,l}\}^d_{l=1}$. By the assumption that $\boldsymbol{w}$ is the projection vector, the output feature can be calculated as:

$$\boldsymbol{x}'_{j,p}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{j,p}. \tag{14}$$

Fisher criterion function is defined as [17]:

$$J=\frac{\sum_{j=1}^{z}\left(\mu'_j-\mu'\right)^2}{\sum_{j=1}^{z}\sigma'^2_j} \tag{15}$$

where μ'_j is the mean of the output feature of class j, μ' is the total mean of the output features of all classes, and σ'^2_j is the variance of the output features of class j. By using (14), (15) can be written as[17] :

$$J = \frac{\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_B\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_W\boldsymbol{w}} \quad (16)$$

where $\mathbf{S}_B = \sum_{j=1}^{z}\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}\right)^{\mathrm{T}} \in \mathbb{R}^{d\times d}$,

$\mathbf{S}_W = \sum_{j=1}^{z}\mathbf{S}_j = \sum_{j=1}^{z}\sum_{p=1}^{n_j}\left(\boldsymbol{x}_{j,p} - \boldsymbol{\mu}_j\right)\left(\boldsymbol{x}_{j,p} - \boldsymbol{\mu}_j\right)^{\mathrm{T}} \in \mathbb{R}^{d\times d}$, $\boldsymbol{x}_{j,p}$ is the p-instance of class j, $\boldsymbol{\mu}_j$ is the mean of features of class j, and $\boldsymbol{\mu}$ is the total mean. The problem here is to find the optimum projection vector $\boldsymbol{w}$ which maximizes Fisher criterion function (16). One of the solutions for this problem is using the Lagrange multiplier where the solution is [17]:

$$\boldsymbol{w} = \mathrm{eig}\left(\mathbf{S}_W^{-1}\mathbf{S}_B\right). \quad (17)$$

4.3 Statistical Distance Calculation

One of the main criterion to calculate the statistical distance between two multivariate random variables is Mahalanobis distance (MD) which is defined as [18]:

$$\mathrm{MD}(\boldsymbol{u}_1,\boldsymbol{u}_2) = \sqrt{\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^{\mathrm{T}}\left(\mathbf{S}_1 + \mathbf{S}_2\right)^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)} \in \mathbb{R} \quad (18)$$

where $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ are the means vectors of the first and the second random variables, and $\mathbf{S}_1$, $\mathbf{S}_2$ are their covariance's matrices, respectively. MD is used as a discrimination measurement between each two classes.

5. Advanced Features Generation Algorithm (AFGA)

The new features must have similar complexity to C_{40}, C_{61}, C_{80} cumulants (have the same number of elements, terms, and maximum power of the C_{40}, C_{61}, C_{80}) as shown in Tab. 3.

cum	Number of elements	Elements	Number of terms	Terms	Max power
C_{40}	2	M_{20}, M_{40}	2	M_{40}, M_{20}^2	2
C_{61}	5	$M_{20}, M_{21}, M_{40}, M_{41}, M_{61}$	4	$M_{61}, M_{21}M_{40}, M_{20}M_{41}, M_{20}^2 M_{21}$	3
C_{80}	4	$M_{20}, M_{40}, M_{80}, M_{60}$	5	$M_{80}, M_{40}^2, M_{60}M_{20}, M_{20}^2 M_{40}, M_{20}^4$	4

Tab. 3. Disassembly of the selected HOCs.

To find three features alternative to C_{40}, C_{61}, C_{80} cumulants, AFGA has four main steps:

1- Determine the most discriminative elements (moments) of each feature.

2- Determine the most discriminative terms based on the selected elements in the previous step.

3- Determine the most optimum weights of the selected terms in the previous step.

4- Extract the final mathematical form.

We apply these steps as the following:

<u>First step</u>: Selection of moments $M_{A_{L_i},k_{L_i}}, M_{B_{L_i},l_{L_i}}, \ldots, M_{C_{L_i},m_{L_i}}$ in (4). In this step, we have to search about the most discriminative moments. To simplify the calculations, the absolute values of the moments can be used. They should have the same number of the elements of selected cumulants as follows:

- To find the elements of the first feature alternative to C_{40}, we have to search about the most two discriminative moments (according to Tab. 3) within $M_4^{th} = \{M_{20}, M_{21}, M_{22}, M_{40}, M_{41}, M_{42}, M_{43}, M_{44}\}$. The vector of the selected moments is called $\chi_4^{'} = \{m_{4,1}, m_{4,2}\}$ where $\chi_4^{'} \in M_4^{th}$.

- To find the elements of the second feature alternative to C_{61}, we have to search about the most five discriminative moments (according to Tab. 3) within $M_6^{th} = \{M_{20}, M_{21}, M_{22}, M_{40}, M_{41}, M_{42}, M_{43}, M_{44}, M_{60}, M_{61}, M_{62}, M_{63}\}$. The vector of the selected moments is called $\chi_6^{'} = \{m_{6,1}, m_{6,2}, m_{6,3}, m_{6,4}, m_{6,5}\}$ where $\chi_6^{'} \in M_6^{th}$.

- To find the elements of the third feature alternative to C_{80}, we have to search about the most four discriminative moments (according to Tab. 3) within $M_8^{th} = \{M_{20}, M_{21}, M_{22}, M_{40}, M_{41}, M_{42}, M_{43}, M_{44}, M_{60}, M_{61}, M_{62}, M_{63}, M_{80}, M_{81}, M_{82}, M_{83}, M_{84}\}$. The vector of selected moments is called $\chi_8^{'} = \{m_{8,1}, m_{8,2}, m_{8,3}, m_{8,4}\}$ where $\chi_8^{'} \in M_8^{th}$.

Searching about the most discriminative moments for each new feature is done by using feature selection algorithms (FSA) in Sec. 4.1 (FS, RF, PCC, LS, and TV) for the 8 selected digital modulation types in Sec. 1. First, we apply the five FSA-mentioned algorithms in Sec. 4.1 (FS, RF, PCC, LS, and TV) ($FSA_k, k = 1..5$) and select the moments that have the highest scores grouped in ($\chi_{L,FSA_k} \in M_L^{th}$, $L = 4,6,8, k = 1,..,5$). Then we choose the moments that have the maximum of the minimum Mahalanobis Distances (MD) (among the different digital modulation classes) among the different feature selection algorithms as:

$$\underset{L=4,6,8}{\chi_L^{'}} = \underset{k=1,..,5}{\mathrm{argmax}}\left(\underset{FSA_k}{\min}\left(\mathrm{MD}\left(\chi_{L,FSA_k,i}, \chi_{L,FSA_k,j}\right)\right)_{i\neq j,\text{classes},i,j=1,..,8}\right) \quad (19)$$

as shown in Fig. 1.

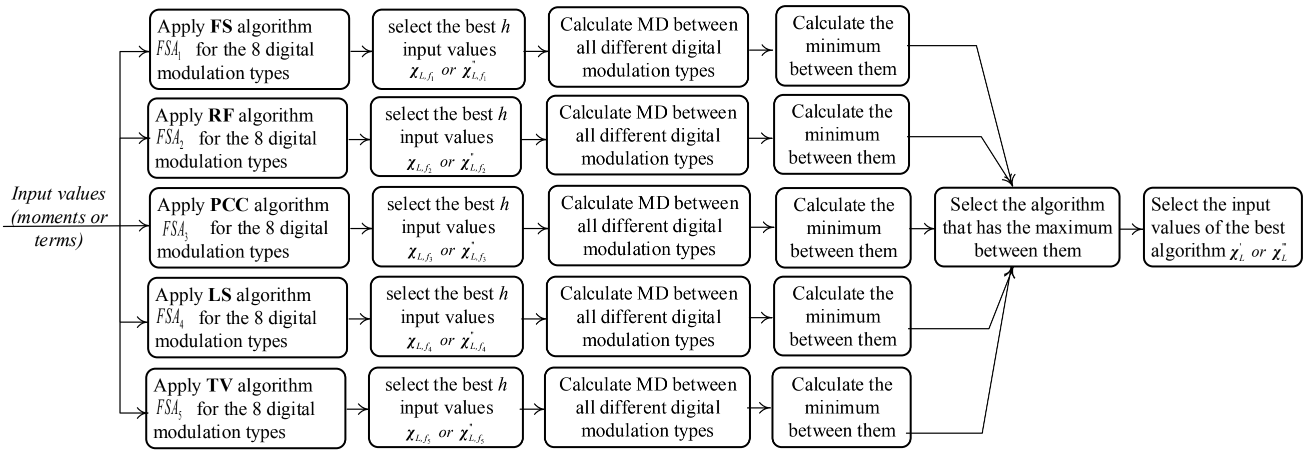

Fig. 1. Most discriminative moments/terms search algorithm.

Second step: Calculate the terms of the selected moments in the previous step $M^{TA_{L_i}}_{A_{L_i},k_{L_i}} . M^{TB_{L_i}}_{B_{L_i},l_{L_i}} \ldots M^{TC_{L_i}}_{C_{L_i},m_{L_i}}$ in (4). To do this step, first, we have to calculate all possible terms for the three alternative features according to its maximum power as shown in Tab. 3, and using the selected elements in previous step as:

- To find the terms of the first feature alternative to C_{40}, we have to calculate all terms of the form $T_4^{th} = \left\{ \prod_{0 \le i,j \text{ integers } 0 < i+j \le 2} m_{4,1}^i m_{4,2}^j \right\}$. Here we have 5 terms. Within these terms, we have to search about the most two discriminative terms ($h = 2$) (according to Tab. 3). The vector of the selected terms is $\boldsymbol{\chi}_4^{'''} = \{t_{4,1}, t_{4,2}\}$ where $\boldsymbol{\chi}_4^{'''} \in T_4^{th}$.

- To find the terms of the second feature alternative to C_{61}, we have to calculate all terms of form $T_6^{th} = \left\{ \prod_{0 \le i,j,k,l,t \text{ integers } 0 < i+j+k+l+t \le 3} m_{6,1}^i m_{6,2}^j m_{6,3}^k m_{6,4}^l m_{6,5}^t \right\}$. Here we have 55 terms. Within these terms, we have to search about the most four discriminative terms ($h = 4$) (according to Tab. 3). The vector of the selected terms is $\boldsymbol{\chi}_6^{'''} = \{t_{6,1}, t_{6,2}, t_{6,3}, t_{6,4}\}$ where $\boldsymbol{\chi}_6^{'''} \in T_6^{th}$.

- To find the terms of the third feature alternative to C_{80}, we have to calculate all terms of the form $T_8^{th} = \left\{ \prod_{0 \le i,j,k,l \text{ integers } 0 < i+j+k+l \le 4} m_{8,1}^i m_{8,2}^j m_{8,3}^k m_{8,4}^l \right\}$. Here we have 69 terms. Within these terms, we have to search about the most five discriminative terms ($h = 5$) (according to Tab. 3). The vector of the selected terms is $\boldsymbol{\chi}_8^{'''} = \{t_{8,1}, t_{8,2}, t_{8,3}, t_{8,4}, t_{8,5}\}$ where $\boldsymbol{\chi}_8^{'''} \in T_8^{th}$.

As previously, searching about the most discriminative terms for each new feature is done using feature selection algorithms (FSA) in Sec. 4.1 (FS, RF, PCC, LS, and TV) as shown in Fig. 1. First, we apply the five FSA-mentioned algorithms in Sec. 4.1 (FS, RF, PCC, LS, and TV) ($FSA_k, k = 1..5$) and select the terms that have h highest scores ($\boldsymbol{\chi}^{"}_{L,FSA_k} \in T_L^{th}$, $L = 4,6,8, k = 1,..5$). Then, we choose the one that gives the maximum of minimum Mahalanobis Distances (MD) among the different FSA algorithms like (19):

$$\underset{L=4,6,8}{\boldsymbol{\chi}_L^{'''}} = \underset{k=1,..,5}{\operatorname{argmax}} \left(\min_{FSA_k} \left(\underset{i \ne j, \text{classes } i,j=1,\ldots,8}{\text{MD}\left(\boldsymbol{\chi}^{"}_{L,FSA_k,i}, \boldsymbol{\chi}^{"}_{L,FSA_k,j} \right)} \right) \right). \quad (20)$$

Third step: Calculate the optimum weights $\boldsymbol{w}_L = \{\alpha_{L_i}\}_{i=1}^{T_L}$, $L = 4,6,8$, $T_4 = 2, T_6 = 4, T_8 = 5$ of the selected terms in previous step for each new feature using (17). According to [17], $\|\boldsymbol{w}_L\|_2 = 1$, this means $-1 < \alpha_{L_i} < 1, \forall i$, to make it integer like cumulants as shown in Tab. 1, mathematical round operation is necessary as:

$$\boldsymbol{w}_L = \text{round}(\boldsymbol{w}_L * \beta) \quad (21)$$

where β is any number under the condition of nonzero $\boldsymbol{w}_L$ elements (as a result of mathematical round operation).

Fourth step: Extract the mathematical form as:

$$\hat{f}_L = \boldsymbol{w}_L^{\mathrm{T}} \boldsymbol{\chi}_L^{'''}, \ L = 4,6,8. \quad (22)$$

The flowchart of the proposed AFGA is shown in Fig. 2.

6. AFGA Implementation

AFGA algorithm generates new features according to the desired SNR region.

Table 4 shows the steps results of applying AFGA for high-SNR values as shown in Fig. 2 which are called here as features-group1 (FG-I).

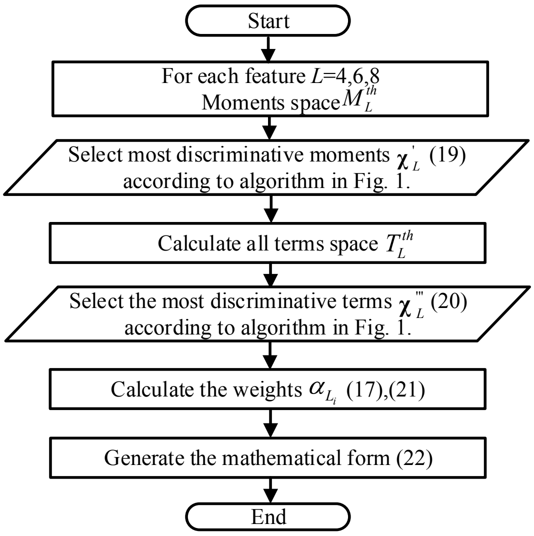

Fig. 2. Flowchart of AFGA.

First step	First feature	$\chi'_{h4} = [\lvert M_{40}\rvert, \lvert M_{44}\rvert]^T$
	Second feature	$\chi'_{h6} = [\lvert M_{40}\rvert, \lvert M_{44}\rvert, \lvert M_{61}\rvert, \lvert M_{42}\rvert, \lvert M_{41}\rvert]^T$
	Third feature	$\chi'_{h8} = [\lvert M_{84}\rvert, \lvert M_{83}\rvert, \lvert M_{82}\rvert, \lvert M_{80}\rvert]^T$
Second step	First feature	$\chi'''_{h4} = [\lvert M_{40}\rvert, \lvert M_{44}\rvert]^T$
	Second feature	$\chi'''_{h6} = [\lvert M_{40}\rvert\lvert M_{42}\rvert^2, \lvert M_{44}\rvert\lvert M_{42}\rvert^2, \lvert M_{40}\rvert\lvert M_{42}\rvert, \lvert M_{44}\rvert\lvert M_{42}\rvert]^T$
	Third feature	$\chi'''_{h8} = [\lvert M_{84}\rvert^4, \lvert M_{84}\rvert^3\lvert M_{83}\rvert, \lvert M_{84}\rvert^3\lvert M_{82}\rvert, \lvert M_{84}\rvert^3\lvert M_{80}\rvert, \lvert M_{84}\rvert^2\lvert M_{83}\rvert^2]^T$
Third step	First feature	$w_{h4} = [7 \quad 7]^T$
	Second feature	$w_{h6} = [-3 \quad -3 \quad 6 \quad 6]^T$
	Third feature	$w_{h8} = [1 \quad -4 \quad 6 \quad -6 \quad 4]^T$
Fourth step	First feature	$\hat{f}_4 = w_{h4}^T \chi'''_{h4} = 7\lvert M_{40}\rvert + 7\lvert M_{44}\rvert$
	Second feature	$\hat{f}_6 = w_{h6}^T \chi'''_{h6} = -3\lvert M_{40}\rvert\lvert M_{42}\rvert^2 - 3\lvert M_{44}\rvert\lvert M_{42}\rvert^2 + 6\lvert M_{40}\rvert\lvert M_{42}\rvert + 6\lvert M_{44}\rvert\lvert M_{42}\rvert$
	Third feature	$\hat{f}_8 = w_{h8}^T \chi'''_{h8} = \lvert M_{84}\rvert^4 - 4\lvert M_{84}\rvert^3\lvert M_{83}\rvert + 6\lvert M_{84}\rvert^3\lvert M_{82}\rvert - 6\lvert M_{84}\rvert^3\lvert M_{80}\rvert + 4\lvert M_{84}\rvert^2\lvert M_{83}\rvert^2$

Tab. 4. New generated features for high SNR values.

Table 5 shows the steps results of applying AFGA algorithm for all-SNR values as shown in Fig. 2 which are called here as features-group2 (FG-II).

When comparing Tabs. 4 and 5, it can be seen that FG-I and FG-II share the identical components, terms, and weights. This indicates that by substituting FG-II weights with their weights, the mathematical forms of high SNR (FG-I) may be deduced from the mathematical form of the general case (all SNR range, i.e. FG-II).

First step	First feature	$\chi'_{t4} = [\lvert M_{40}\rvert, \lvert M_{44}\rvert]^T$
	Second feature	$\chi'_{t6} = [\lvert M_{40}\rvert, \lvert M_{44}\rvert, \lvert M_{61}\rvert, \lvert M_{42}\rvert, \lvert M_{41}\rvert]^T$
	Third feature	$\chi'_{t8} = [\lvert M_{84}\rvert, \lvert M_{83}\rvert, \lvert M_{82}\rvert, \lvert M_{80}\rvert]^T$
Second step	First feature	$\chi'''_{t4} = [\lvert M_{40}\rvert, \lvert M_{44}\rvert]^T$
	Second feature	$\chi'''_{t6} = [\lvert M_{40}\rvert\lvert M_{42}\rvert^2, \lvert M_{44}\rvert\lvert M_{42}\rvert^2, \lvert M_{40}\rvert\lvert M_{42}\rvert, \lvert M_{44}\rvert\lvert M_{42}\rvert]^T$
	Third feature	$\chi'''_{t8} = [\lvert M_{84}\rvert^4, \lvert M_{84}\rvert^3\lvert M_{83}\rvert, \lvert M_{84}\rvert^3\lvert M_{82}\rvert, \lvert M_{84}\rvert^3\lvert M_{80}\rvert, \lvert M_{84}\rvert^2\lvert M_{83}\rvert^2]^T$
Third step	First feature	$w_{t4} = [3 \quad 3]^T$
	Second feature	$w_{t6} = [-3 \quad -3 \quad 6 \quad 6]^T$
	Third feature	$w_{t8} = [1 \quad -4 \quad 4 \quad -3 \quad 3]^T$
Fourth step	First feature	$\hat{f}_4 = w_{t4}^T \chi'''_{t4} = 3\lvert M_{40}\rvert + 3\lvert M_{44}\rvert$
	Second feature	$\hat{f}_6 = w_{t6}^T \chi'''_{t6} = -3\lvert M_{40}\rvert\lvert M_{42}\rvert^2 - 3\lvert M_{44}\rvert\lvert M_{42}\rvert^2 + 6\lvert M_{40}\rvert\lvert M_{42}\rvert + 6\lvert M_{44}\rvert\lvert M_{42}\rvert$
	Third feature	$\hat{f}_8 = w_{t8}^T \chi'''_{t8} = 1\lvert M_{84}\rvert^4 - 4\lvert M_{84}\rvert^3\lvert M_{83}\rvert + 4\lvert M_{84}\rvert^3\lvert M_{82}\rvert - 3\lvert M_{84}\rvert^3\lvert M_{80}\rvert + 3\lvert M_{84}\rvert^2\lvert M_{83}\rvert^2$

Tab. 5. New generated features for all SNR values.

7. Performance Improvement of the New Features

The classification of chosen digital modulation types in Sec. 1 using the selected cumulants in Sec. 1 and the new generated alternatives features in Tabs. 4 and 5 is done. Performance accuracy and its improvement (subtract the first accuracy (using the selected cumulants) of the second accuracy (using the new features) are shown in Figs. 3 and 4.

As shown in Fig. 3, the new generated features could improve the performance accuracy up to 4.5% for SNR value of 10 dB. The improvement is in the case of SNR values of larger than 4 dB. No accuracy improvement is for SNR values larger than 16 dB because of the accuracy by using the selected cumulants and the new generated features is 100%, as shown in Fig. 3.

Another comparison should be done with some references in Sec. 1. In [6], it was shown that average performance accuracy is 84.37%, for SNR value of 10 dB. While our work shows that the performance accuracy using our new generated features is 98.78% as shown in Fig. 3. As a result, our improvement compared to [6] is 14.41%. In [7], it has been shown that the performance accuracy is 80%, for SNR value of 10 dB. As a result, our improvement compared to [7] is 18.78%.

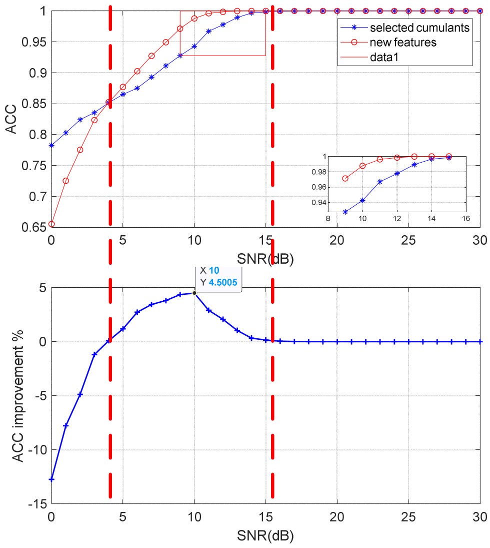

Fig. 3. Classification accuracy and improvements of the new features FG-I compared with the selected cumulants.

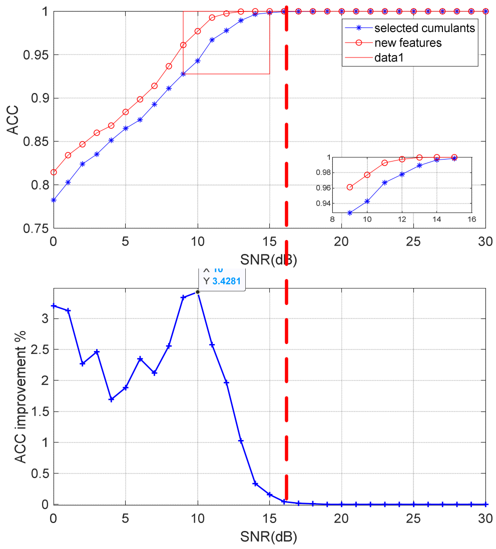

Fig. 4. Classification accuracy and improvements of the new features FG-II compared with the selected cumulants.

As shown in Fig. 4, the new generated features could improve the performance accuracy up to 3.428% for SNR value of 10 dB. Unlike the performance of the FG-I, it could improve the performance accuracy for all SNR values up to 16 dB.

8. Conclusion

New algorithm to generate more discriminative features than the cumulants, for MPSK and MQAM classification was proposed. Mathematical features forms have been generated. The new generated features could effectively improve the performance accuracy compared to the selected cumulants as summarized in Tab. 6.

AFGA is a very powerful algorithm and can be used for other classification tasks.

New features	Improvement range	Maximum Improvement
FG-I	5 dB–16 dB	4.5%
FG-II	0 dB–16 dB	3.428%

Tab. 6. Performance comparison between the new features.

Reference

[1] AL-NUAIMI, D. H., HASHIM, I. A., ZAINAL ABIDIN, I. S., et al. Performance of feature-based techniques for automatic digital modulation recognition and classification–A review. *Electronics,* 2019, vol. 8, no. 12, p. 1–25. DOI: 10.3390/electronics8121407

[2] HAZZA, A., SHOAIB, M., ALSHEBEILI, S. A., et al. An overview of feature-based methods for digital modulation classification. In *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*. Sharjah (United Arab Emirates), 2013, p. 1–6. DOI: 10.1109/ICCSPA.2013.6487244

[3] SOBOLEWSKI, S., ADAMS, W. L., SANKAR, R. Universal nonhierarchical automatic modulation recognition techniques for distinguishing bandpass modulated waveforms based on signal statistics, cumulant, cyclostationary, multifractal and Fourier-wavelet transforms features. In *2014 IEEE Military Communications Conference*. Baltimore (MD, USA), 2014, p. 748–753. DOI: 10.1109/MILCOM.2014.130

[4] GHAURI, S. A. *Automatic Modulation Classification Using Feature Based Approach.* Ph.D dissertation. ISRA University Hyderabad, Islamabad Campus, 2015.

[5] SIMIC, M., STANKOVIC, M., ORLIC, V. D. Automatic modulation classification of real signals in AWGN channel based on sixth-order cumulants. *Radioengineering*, 2021, vol. 30, no. 1, p. 204–214. DOI: 10.13164/re.2021.0204

[6] GHAURI, S. A., QURESHI, I. M., AZIZ, A., et al. Classification of digital modulated signals using linear discriminant analysis on faded channel. *World Applied Sciences Journal,* 2014, vol. 29, no. 10, p. 1220–1227. DOI: 10.5829/idosi.wasj.2014.29.10.1540

[7] DOBRE, O. A., ABDI, A., BAR-NESS, Y., et al. Cyclostationarity-based modulation classification of linear digital modulations in flat fading channels. *Wireless Personal Communications,* 2010, vol. 54, no. 4, p. 699–717. DOI: 10.1007/s11277-009-9776-2

[8] EL-KHAMY, S. E., ELSAYED, H. A. Classification of multi-user chirp modulation signals using wavelet higher-order-statistics features and artificial intelligence techniques. *International Journal of Communications, Network and System Sciences,* 2012, vol. 5, no. 9, p. 520–533. DOI: 10.4236/ijcns.2012.59063

[9] GHAURI, S. A., QURESHI, I. M., MALIK, A. N., et al. Higher order cummulants based digital modulation recognition scheme. *Research Journal of Applied Sciences Engineering & Technology,* 2013, vol. 6, no. 20, p. 3910–3915. DOI: 10.19026/rjaset.6.3609

[10] LI, J., CHENG, K., WANG, S., et al. Feature selection: A data perspective. *ACM Computing Surveys (CSUR),* 2017, vol. 50, no. 6, p. 1–45. [Online] Available at: https://arxiv.org/pdf/1601.07996.pdf

[11] STAŃCZYK, U. Feature evaluation by filter, wrapper, and embedded approaches. In *Feature Selection for Data and Pattern Recognition*. Springer, 2015, p. 29–44. DOI: 10.1007/978-3-662-45620-0_3

[12] PÉREZ-ORTIZ, M., TORRES-JIMÉNEZ, M., GUTIÉRREZ, P. A., et al. Fisher score-based feature selection for ordinal classification: A social survey on subjective well-being. In *International Conference on Hybrid Artificial Intelligence*

Systems. Seville (Spain), 2016, p. 597–608. DOI: 10.1007/978-3-319-32034-2_50

[13] URBANOWICZ, R. J., MEEKER, M., LA CAVA, W., et al. Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics,* 2018, vol. 85, p. 189–203. DOI: 10.1016/j.jbi.2018.07.014

[14] EID, H. F., HASSANIEN, A. E., KIM, T.-H., et al. Linear correlation-based feature selection for network intrusion detection model. In *International Conference on Security of Information and Communication Networks*. Cairo (Egypt), 2013, p. 240–248. DOI: 10.1007/978-3-642-40597-6_21

[15] ZHU, L., MIAO, L., ZHANG, D. Iterative Laplacian score for feature selection. In *Chinese Conference on Pattern Recognition*. Beijing (China), 2012, p. 80–87. DOI: 10.1007/978-3-642-33506-8_11

[16] SHAMSINEJADBABKI, P., SARAEE, M. A new unsupervised feature selection method for text clustering based on genetic algorithms. *Journal of Intelligent Information Systems,* 2012, vol. 38, no. 3, p. 669–684. DOI: 10.1007/s10844-011-0172-5

[17] GHOJOGH, B., KARRAY, F., CROWLEY, M. *Fisher and Kernel Fisher Discriminant Analysis: Tutorial.* arXiv Prepr. arXiv1906.09436, 2019, [Online] Available at: http://arxiv.org/abs/1906.09436

[18] LAHAV, A., TALMON, R., KLUGER, Y. Mahalanobis distance informed by clustering. *Information and Inference: A Journal of the IMA,* 2019, vol. 8, no. 2, p. 377–406. DOI: 10.1093/imaiai/iay011

About the Authors ...

Iyad KADOUN was born in Damascus, Syria, in 1979 and received his B.S. degree in Communication Engineering from HIAST in 2002, and his M.Sc. degree in Communication Engineering from Malek Ashtar University in 2012. His research interests include digital communications.

Hossein KHALEGHI BIZAKI received the Ph.D. degree in Electrical Engineering and Communication Systems from Iran University of Science and Technology, Tehran, Iran, in 2008. He is an author or coauthor of more than 50 publications. His research interests include information theory, coding theory, wireless communication, multiple-input-multiple-output systems, space-time processing, and other topics on communication system and signal processing.