Method for the Automatic Modulation Classification for the Sums of Phase-Shift-Keying Signals Based on Polynomial Features

Vasyl Semenov
Laboratory of Data Science and Machine Learning
Kyiv Academic University,
American University Kyiv,
Delta SPE LLC
Kyiv, Ukraine
vasyl.delta@gmail.com

Yevheniya V. Semenova
Laboratory of Data Science and Machine Learning
Kyiv Academic University,
Department of Numerical Mathematics
Institute of Mathematics of NANU
Kyiv, Ukraine
semenovaevgen@gmail.com

***Abstract*—This paper considers the task of automatic modulation classification for the sums of phase-shift keying (PSK) signals which occupy the same frequency bandwidth and have the same modulation. This problem is an important step which precedes the subsequent processing of such signals. For the solution of this classification problem we implement the method based on two machine learning tools: logistic regression or two-layer feedforward neural network. It allows to determine the class to which the sum of two PSK signals belongs. As the features used for the input of these classifications tools, we use second-order polynomial features calculated based on higher order cumulants of the received signal. The results of numerical experiments using different channel parameters and noise levels show the advantage of implementing second-order features over the raw features both in the case of logistic regression and two-layer feedforward neural network.**

Keywords—modulation classification, cumulants, BPSK, QPSK, logistic regression, neural network.

I. Introduction

In this paper we consider the problem of Automatic modulation classification (AMC) which is an important part for the processing of wireless signals when the characteristics of received signals are unknown a priori [1-3]. This task also finds its applications in non-cooperative communication, software defined radio etc. In practical systems, radio signals are modulated according to the capacities of transmitter taking also into account the signal impairments which occur over the channel.

While contemporary error-correction methods coding operate at the edge of theoretical Shannon limit, channel capacity can still be doubled by using technologies in which multiple transmitted signals occupy the same bandwidth [4]. While a lot of attention was paid in current literature for the AMC of one signal-per-band signals [1-3], less (if any) attention was given to the modulation detection when two or more modulated signals are transmitted in the same band.

Thus, AMC can be defined as the signal type detection for the sake of subsequent demodulation or signals separation. Usually AMC methods are subdivided into two categories: likelihood-based/probabilistic-based approaches and feature-based approaches. Though the likelihood-based approaches could theoretically provide optimal classification accuracy (assuming obtaining perfect information about signal and channel parameters which usually is not realizable in practice), they request high computational resources [1]. On the opposite, feature-based methods are usually more practical allowing easy implementation and low complexity. Feature-based AMC methods mainly include support vector machines, decision trees, different clustering methods, regression methods and deep learning architectures [1-3]. All these methods consider the situation when only one signal occupies some bandwidth. However, the determination of modulation type for the *sum* of two or more signals is a challenging problem which, to the authors' knowledge, was not addressed in the current scientific literature.

That is why, in this investigation we consider the binary classification of signals belonging either to the *sum* of two Binary Phase Shift Keying (BPSK) signals or to the *sum* of two Quaternary Phase Shift Keying (QPSK) signals. These two classes of signals are two simplest alternatives for the concurrent signals using the same bandwidth, although higher-order modulations have also already received practical implementation. Earlier, in paper [5] we proposed the classification method for such signals which was based on the logistic regression and have shown its advantage over the linear regression approach. In this paper we extend the proposed approach by incorporating two aspects. Firstly, we introduce the second-order polynomial features which allows for better separation between the two analyzed classes of signals. Secondly, we apply two-layer neural network (NN) to further enhance classification accuracy.

II. Model of Signal

Consider the sum of two signals $z_p(t),\ p=1,2$ which belong to different sources [4, 5]:

$$z(t)=z_1(t)+z_2(t)+w(t)\,,$$

where $z_p(t),\ p=1,2$ are the signals from two sources having the following form:

$$z_p(t)=a_p e^{j(\phi_p+\omega_p t)}\sum_{n=-\infty}^{\infty}s_p(n)h(t-nT-\tau_p),\ p=1,2 \qquad (1)$$

and $s_p(n),\ p=1,2$ are original sequences (from the alphabet of 2 symbols for BPSK and 4 symbols for QPSK

979-8-3503-6817-8/24/$31.00 ©2024 IEEE

respectively); T is the symbol period; a_p, ϕ_p, τ_p, ω_p are the signals' amplitudes, phases, time shifts and carrier frequencies respectively; $h(t)$ is a channel response; $w(t)$ is a background noise.

Note that the signals $s_p(n), p=1,2$ have the same modulation (both BPSK or both QPSK). The task is to determine the modulation type on the base of received signal (1).

III. Feature Extraction

Since in this paper we follow the feature-based approach, higher order cumulants (HOC) [3,5] are used as informative features. The HOCs are defined through the (p,q) moments which are defined as:

$$M_{pq} = E\{z^{p-q}(z^*)^q\},$$

where z^* is a complex conjugate of z.

For instance, HOCs C_{42} and C_{63} may be computed in the following way:

$$\begin{aligned}
&C_{42} = M_{42} - |M_{20}|^2 - 2M_{21}^2,\\
&C_{63} = M_{63} - 9M_{42}M_{21} + 12M_{21}^3 - 3M_{43}M_{20} -\\
&-3M_{41}M_{22} + 18M_{20}M_{21}M_{22}.
\end{aligned}$$

Note that all features must be normalized to have approximately the same range to facilitate training of the regression or neural network model.

The example of mutual distribution of features C_{42} and C_{63} is shown in Fig. 1.

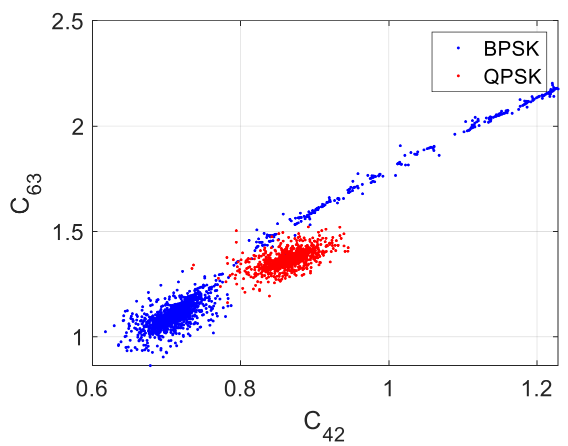

Fig. 1. Distribution of cumulants C_{42} and C_{63}.

IV. Polynomial Features

While the direct application of linear regression and logistic regression can provide only linear boundaries, such separating boundaries in fact can be highly non-linear, as in the case of features presented above in Fig. 1. However, polynomial features are allowing to adjust the classification boundary in higher-dimensional space to enable the proper linear distinction between features which belong to different classes [3].

The dimension of this space is defined by the order of polynomial used. While the higher order polynomial classifiers could be used, for the sake of implementation simplicity a second order polynomial classifier is most commonly used so that the initial vector of features

$$\mathbf{x} = [x_1, x_2, ..., x_p]$$

is replaced by the expanded feature vector consisting of the initial features, their cross-products and squared values:

$$\mathbf{x}' = [x_1, x_2, ..., x_p, x_1x_2, ..., x_{p-1}x_p, x_1^2, x_2^2, ..., x_p^2] \quad (2)$$

(so that the total number of coordinates in feature vector is now $p + p(p-1)/2 + p = p(p+3)/2$).

Example of such transformation is shown in Fig. 2.

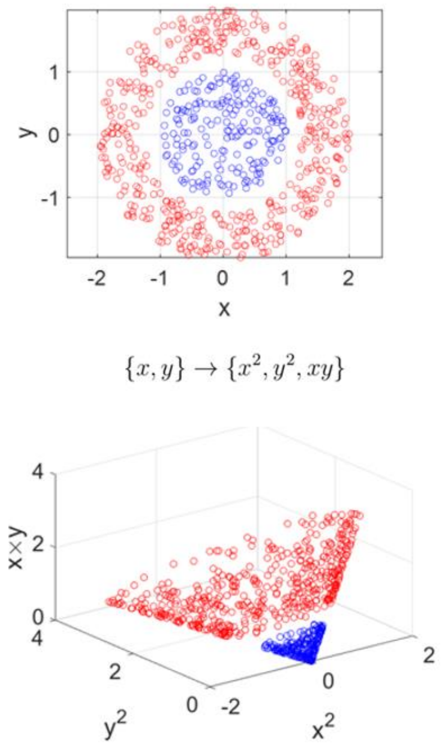

Fig. 2. Example of using second-order features.

V. Application of Logistic Regression for Classification

Supervised learning methods subdivide into regression and classification. Classification is different to regression in the sense that it identifies the class of the variable out of the several given classes while the latter has a continuous scale.

Suppose we have the set of features $X = X_1 \cup X_2$, $X_1 \cap X_2 = \varnothing$ and the task it to find mapping $y(x)$ to map each vector $\mathbf{x}_i \in X$ to one of the classes:

$$y_i = 0 \text{ if } \mathbf{x}_i \in X_1 \text{ and } y_i = 1 \text{ if } \mathbf{x}_i \in X_2.$$

Logistic regression consists of the estimation of coefficients $\mathbf{w}, b$ so that $g((\mathbf{w}, \mathbf{x}) + b)$ produces probability for feature vector $\mathbf{x}$ to belong either to class X_1 or X_2. Function $g(t)$ is usually called activation function and in the scope of logistic regression is supposed to be sigmoid function:

$$g(t) = \frac{1}{1+e^{-t}} . \qquad (3)$$

For example, $g((\mathbf{w},\mathbf{x})+b) = 0.7$ means that vector $\mathbf{x}$ belongs to class X_2 with probability 0.7 while $g((\mathbf{w},\mathbf{x})+b) = 0.2$ means that vector $\mathbf{x}$ belongs to class X_1 with probability 0.8. The coefficients $\mathbf{w}$, b can be estimated by the gradient descent method using binary cross-entropy loss function.

So, the essence of the proposed classification method based on logistic regression is as folllows.

1. Having the received signal (1), calculate the feature vector $\mathbf{x}$ as explained above.

2. Make classification based on the sigmoid function value: $g((\mathbf{w},\mathbf{x})+b)$.

Another possible approach described in [3] is based on simple linear regression to find the coefficients $\mathbf{w}$, b so that linear combinations $(\mathbf{w},\mathbf{x})+b$ approximate the binary values y_i in the least-squares sense. In this case the coefficients $\mathbf{w}$, b can be estimated by the traditional gradient descent method based on mean-square error loss function or by solving normal equation as was demonstrated in work [3].

An example of classification by logistic regression for the data of Fig.1 is shown in Fig. 3.

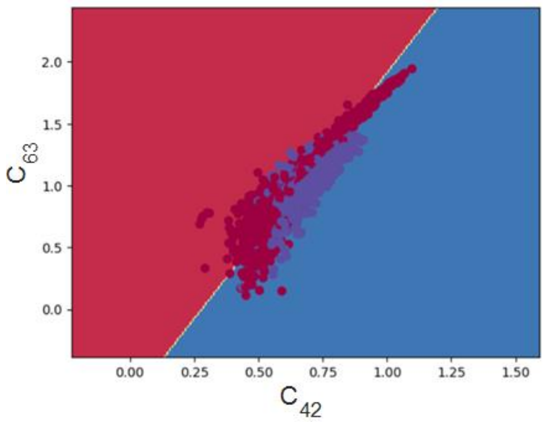

Fig. 3. Example of decision boundary obtained by logistic regression.

As it can be seen from Fig. 3, direct application of logistic regression provides linear decision boundary defined by the equation $(\mathbf{w},\mathbf{x})+b = 0$.

VI. Application of Neural Network

While one of the guiding principles for communications systems is to split the signal processing into a chain of multiple independent blocks, it is not clear that individually optimized processing blocks achieve the best possible performance [6]. At the same time, it has been demonstrated that Neural networks (NN) are universal function approximators [7, 8] providing that the sufficient number of layers/units are included. Therefore, application of NN for signal processing allows the simultaneous optimization of the signal processing system parameters to achieve the optimal performance.

NN can be either shallow with only few hidden layers or deep with large number of hidden layers. The example of architecture of a simple feedforward NN with two hidden layers is shown in Fig. 4.

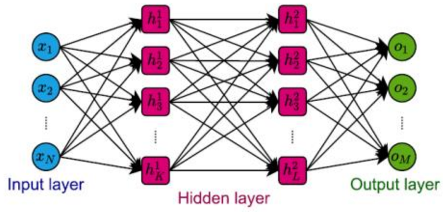

Fig. 4. Example feedforward NN with two hidden layers.

The nonlinearity in feedforward NN is represented by activation functions, which help the network to learn any relationship between the input and the output. Some typical include: sigmoid function (3), tangent hyperbolic (tanh) function and ReLU (rectified linear unit) function:

$$\mathrm{ReLU}(x) = \begin{cases} 0, x < 0 \\ x,\ x \geq 0, \end{cases}$$

which have output range (0,1), (-1,1) and $(-\infty,\infty)$ respectively.

Although feedforward NNs for existing AMC methods typically involve more than two hidden layers, in this investigation only two layers are used. This is done since we want just to show here the potential of this approach and also because we have to classify the sums of PSK signals in two classes. For example, in the case of NN with two hidden layers and respectively 1024 and 32 ReLU-activated units in each layer, we have the decision boundary shown in Fig. 5 for the same data as shown in Fig. 1 (note that sigmoid activation function was used at the output layer). Obviously, such flexible decision boundary results in higher classification accuracy as compared with logistic regression as shown in Fig 3.

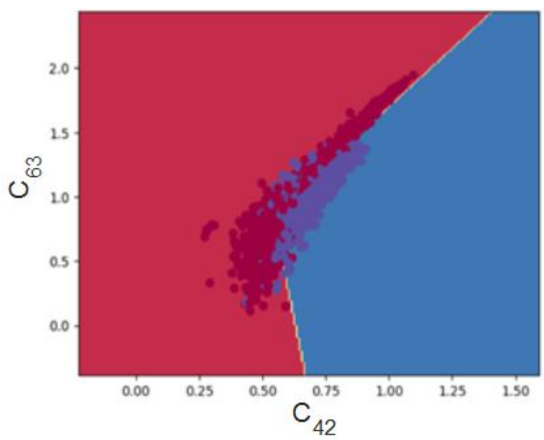

Fig. 5. Example of decision boundary obtained by NN with two layers.

VII. Experimental results.

We verified the effectiveness of the proposed method on the base of 17500 signals, of which 8750 were used for training and 8750 for testing purposes. Both training and validation tests were balanced, i.e. they contained equal number of signals from each of two classes.

These signals were generated according to model (1) using various random values for the parameters a_p, ϕ_p, τ_p, ω_p ($p = 1, 2$). As for the white noise $w(t)$, different signal-to-noise ratios (SNR) from 0 to 15 dB with step of 2.5 dB were modeled.

The examples of signal constellation for the signals from each of two classes are shown in Fig. 6 and Fig. 7.

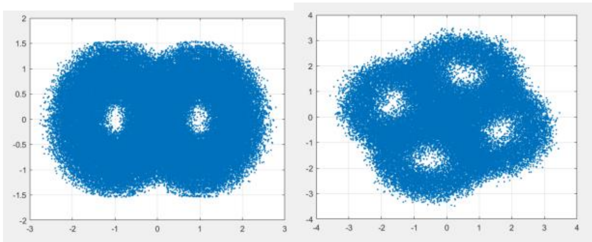

Fig. 6. Examples of signal constellation for the signals from the BPSK+BPSK class (left) and QPSK+QPSK class (right).

Then we evaluated classification accuracy for the following methods:

- linear regression,
- logistic regression,
- feedforward NN with two hidden layers activated by respectively 1024 and 32 ReLU-units, followed by sigmoid activation function at the output layer. This NN was trained by "Adam" algorithm using binary cross-entropy loss function. 10 epochs and batch size 16 were used for the training of NN.

Using feature selection approach proposed in [5], out of 9 first cumulants we selected the 3 most informative ones: features 1, 3, 5 (i.e. cumulants $_{20}, C_{40}, {}_{42}$). This was done by selecting three coordinates of vector $\mathbf{w}$ (calculated for the logistic regression task) which have the highest absolute values.

These three methods were applied for the abovementioned data using a) most informative features 1, 3, 5; b) features 1-9; c) second-order set of features 1-54 as given by (2). The corresponding results for the classification accuracy are presented in Table I. Table II presents the same results being distinguished for both types of signals (the numbers in each cell are the accuracies for the first type and second type of signals respectively).

TABLE I. Classification accuracy for the different methods

	Linear Regression	Logistic regression	Neural network
Features 1,3,5	0.963	0.972	0.980
Features 1-9	0.972	0.979	0.981
Features 1-54	0.982	0.984	0.984

TABLE II. Classification accuracy for the different methods (distinguished for both types of signals)

	Linear Regression	Logistic regression	Neural network
Features 1,3,5	0.975/0.949	0.973/0.970	0.977/0.983
Features 1-9	0.974/0.970	0.980/0.978	0.991/0.971
Features 1-54	0.979/0.986	0.983/0.985	0.987/0.981

The results presented in Table I and Table II show their consistent nature: the logistic regression has advantage over the linear regression while the two-layer NN performs not worse that both linear and logistic regression. At the same time, using polynomial features reduces the difference between all three approaches thus achieving the accuracy of 0.984.

The selection of the number of NN's layers and the number of nodes in each layer as well as the application of convolutional NNs could be the topic of future research.

VIII. Conclusions

In this paper we considered the task of automatic modulation classification for the sums of BPSK and QPSK signals. For this purpose, we used logistic regression and two-layer feedforward neural network based on second-order polynomial features calculated from higher order cumulants of the received signal. The simulation results for the classification with various channel parameters and noise levels shows the advantage of using second-order features over the raw features both in the case of logistic regression and two-layer feedforward neural network. It allowed to achieve 98.4% accuracy of classification over considered test database both for the logistic regression and the neural network approach. Investigation of more elaborate structures of neural network allowing to comprise more classes of signals is the topic for future research.

References

[1] Th. Huynh-The et al., "Automatic modulation classification: a deep architecture survey," IEEE Access, vol. 51, 2021, pp. 142950–142971.

[2] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," IEEE Journal of Selected Topics in Signal Processing, vol. 12, 2018, pp. 168–179.

[3] A. Abdelmutalab, Kh. Assaleh, and M. El-Tarhuni, "Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers," Physical communication, vol. 21, 2016, pp. 10–18.

[4] V. Semenov, P. Omelchenko, O. Kruhlyk, "Method for the detection of Carrier-in-Carrier Signals based on fourth order cumulants," Computer Science and Information Technology, Nagamalai Dh. et al., Eds., 2019, pp. 25–32.

[5] V. Semenov, "Method for the automatic modulation classification based on linear regression and feature selection," Physico-mathematical modelling and informational technologies, Vol. 36, 2023, pp. 22-26.

[6] T. O'Shea, "An Introduction to Deep Learning for the Physical Layer," IEEE Transactions on Cognitive Communications and Networking, Vol. 3, pp. 563-575.

[7] H. Mhaskar, Q. Liao, and T.Poggio, "Learning functions: when is deep better than shallow," https://arxiv.org/abs/1603.00988, 2016.

[8] M. Blanchard and A. Bennouna, "The representation power of neural networks: breaking the curse of dimensionality," https://arxiv.org/abs/2012.05451, 2021.