Dimensionality reduction and improving the performance of automatic modulation classification using genetic programming

Karim Hessampour Yazd University k.hessampour@stu.yazduni.ac.ir
AliMohammad Latif Assistant Professor Yazd University alatif@yazduni.ac.ir
Mohammad Ghasemzadeh Assistant Professor Yazd university m.ghasemzadeh@yazduni.ac.ir

Paper Reference Number: 07-02-7070
Name of the Presenter: Karim Hessampour

Abstract

Modulation recognition is one of the main parts of modern communication receivers. The automatic modulation recognition of received signal is considered as the intermediate step between signal detection and its demodulation. In most military and communication systems, modulation detection is considered as a part of whole system. The existing methods for this purpose are based on modulated signal's components. In this paper, the features of modulation classification were decreased using genetic programming. The performance of modulation classification can be improved using genetic programming. The performance improvement was compared between Multi-layer neural network perceptron, and decision tree. The purposed method was designed for 16QAM‹ 64QAM ‹2FSK ‹4FSK ‹2PSK ‹4PSK modulations. The simulation results show that this method has high accuracy and good convergence rate in presence of noise.

Key words: modulation automatic detection, genetic programming, entropy, Multi-layer neural network perceptron, and decision tree

1. Introduction

Assuming that making connection between two separate points is one of the basic goals of communication science, so modulation is one of its inseparable parts. Modulation is defined as process in which one or more features of carrier signal will be changed in accordance with the message signal [1].

Modulation recognition has attracted the attentions of communication systems. Accurate modulation recognition can be used to integrate some receivers. Routing is one of the other applications in a network environment, so each signal is sent to a suitable receiver. Modulation classification is also used in military applications. In electronic war systems, this classification is used as an information source for receiving enemies signals [2]. In civilian application, this is used to frequency management, transmitter monitoring and unauthorized transmitter detection.

Modulation recognition is an important step in transmission signals modulation and demodulation. In non-cooperative environments, the input signal detection is difficult since the lack of input signal information. The design of automatic modulation classification has two main steps: pre processing of input signal, and selection of classification method. The second step is divided into two groups as follows:

- Maximum likelihood approach: this approach considers automatic modulation classification as a hypothesis test problem. Therefore, the computation of maximum likelihood between the received signal and hypothesis test design, detect the received signal modulation.
- Statistical Pattern recognition approach: in this approach, fist some special features of the signal are obtained using different mathematical techniques, then decision making will be performed using this features. This has more diversity than the first solution. It should be noted that most modulation methods are not general. These methods have lower computational difficulties and so more practicable. On the other hand, the signal features are selected so those are resistant against non-ideal parameters.

Nandi and Wong have used artificial neural network and genetic algorithm (GA) for automatic modulation recognition [4]. Zhu, Aslam and Nandi have used genetic programming and K-nearest Neighborhood for modulations classification [6]. Neshatian and Zhang have used GP for dimensionality reduction and performance improvement of standard data mining datasets [7]. In this paper, it is tried to use GP intelligence selection of input features. However, this selection is done in such a way that the features number will be equal to output modulation number. Entropy concept was used for compute fitness function. In section 2, signal model and signal statistical specifications was have described. The main structure of algorithm producing initial features of GP and GP structure have introduced in section 3. In section 4, performance evaluation and results analyzing have presented. Finally, the conclusions have shown in section 5.

2. Data and Material

Signal model and features

For most communication systems The baseband wave form in receivers is considered as Equ.1 [2]:

$$y(n) = Ae^{j(2\pi f_o nT+\theta_n)} \sum_{l=-\infty}^{\infty} x(l)h(nT - lT + \epsilon_T T) + \mathrm{g(n)} \qquad (1)$$

Where, x(l) is symbols sequence, A is variable amplitude coefficient, T is symbols interval, $\epsilon_T$is time regulation error or sampling phase error in filter output, $f_o$is deviation of receiver oscillator frequency from carrier signal , h(.) baseband channel specification, θ_n is carrier signal phase deviation from receiver oscillator , g(n) is Additive white Gaussian Noise (AWGN). In this paper, it is assumed that the working conditions is ideal and only Gaussian nose is present, the channel compensation was done, and the channel effects can be neglected.

2.2. Used features

In this paper, statistical pattern recognition approach was used for modulation automatic detection. So a set of features was considered as GP search input. These features will be presented later.

2.2.1. Key features γ_{max} [8]

The first key features which is shown as γ_{max} presents the maximum normal and concentrated spectrum power density for received signal instant amplitude. In accordance with Parseval's theorem, the power spectrum density of a signal equals to its DFT SEQUENCE MAGNITUDE SQUARE. Therefore , γ_{max} is as follows:

$$\gamma_{\max} = \frac{\max |DFT(a_{cn}(i))|^2}{N_s} \qquad (2)$$

So that:

$$a_{cn}(i) = a_n(i) - 1 \qquad (3)$$

$$a_n(i) = \frac{a(i)}{m_a} \qquad (4)$$

$$m_a = \frac{1}{N_s}\sum_{i=1}^{N_s} a(i) \qquad (5)$$

N_s Is the samples in each part of signal and $a_{cn}(k)$ is the normalized instant amplitude and centered related to ($t = \frac{k}{fs}$). So that *k=1, 2… And* f_s is the sampling frequency of each part of the signal and m_a is the mean value of instant amplitude for each part of the signal.

2.2.2. Key features σ_{ap} [8]

σ_{ap} is the second key feature which is equals to standard deviation of absolute number of non-linear and concentrated component (displaced) of received signal instant phase and defined as follows. The instant phase absolute standard deviation when the received signal is higher than threshold value.

$$\sigma_{ap} = \sqrt{\frac{1}{C}\left(\sum_{a_n(i)\rangle a_t} \phi_{NL}^2(i)\right) - \left(\frac{1}{C}\sum_{a_n(i)\rangle a_t} |\phi_{NL}(i)|\right)^2} \qquad (6)$$

Where $\varphi_{NL}(i)$ is the instant phase concentrated non-linear component for t so that $(t = \frac{i}{f_s})$ and *i=1,2,…., Ns* , *C* equals to total samples related to instant phase $\varphi_{NL}(i)$ which their instant amplitude is higher than the threshold value. If the $a_n(i)$ is the amplitude related to phase $\varphi_{NL}(i)$ and a_t is the threshold value, C is the total samples which $a_n(i) > a_t$ conditions is satisfied.

2.2.3. Key features σ_{dp}

σ_{dp} is the third key feature which is defined as follows:

$$\sigma_{ap} = \sqrt{\frac{1}{C}\left(\sum_{a_n(i)\rangle a_t} \phi_{NL}^2(i)\right) - \left(\frac{1}{C}\sum_{a_n(i)\rangle a_t} \phi_{NL}(i)\right)^2} \qquad (7)$$

This is obtained in a similar way with the second feature and equals to standard deviation of real part of the concentrated non-linear component or the instant phase displacement.

2.2.4. Key features σ_{aa}

σ_{aa} is the forth key feature which is equal to standard deviation of absolute value of normalized instant amplitude when the received signal is higher than the threshold level.

$$\sigma_{aa} = \sqrt{\frac{1}{N_s}\left(\sum_{i=1}^{N_s} a_{cn}^2\right) - \left(\frac{1}{N_s}\sum_{i=1}^{N_s} | a_{cn} |\right)^2} \qquad (8)$$

2.2.5. Key features σ_{af}

σ_{af} is the fifth key feature which is equal to standard deviation of normalized instant frequency absolute value when the received signal is higher than the threshold level.

$$\sigma_{af} = \sqrt{\frac{1}{C}\left(\sum_{a_n(i)\rangle a_t} f_N^2(i)\right) - \left(\frac{1}{C}\sum_{a_n(i)\rangle a_t} | f_N(i) |\right)^2} \qquad (9)$$

σ_{af}

So that

$$f_N(i) = \frac{f(i) - m_f}{r_s} \qquad (10)$$

$$m_f = \frac{1}{N_s}\sum_{i=1}^{N_s} f(i) \qquad (11)$$

And r_s is the input signal symbol rate.

2.2.6. Second order, forth order, sixth order and eight order cumulants

For a complex-valued stationary random process y(n) second order, forth order, sixth order and eight order cumulants based on moments can be calculated according to following equations[10].

$$
\begin{aligned}
&C_{20} = M_{20} \\
&C_{21} = M_{21} \\
&C_{40} = M_{40} - 3M_{20}^2 \\
&C_{41} = M_{40} - 3M_{20}M_{21} \\
&C_{42} = M_{42} - |M_{20}|^2 - 2M_{21}^2 \\
&C_{60} = M_{60} - 15M_{20}M_{40} + 30M_{20}^3 \\
&C_{61} = M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30M_{20}^2M_{21} \\
&C_{62} = M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40} + 6M_{20}^2M_{22} + 24M_{21}^2M_{20} \\
&C_{63} = M_{63} - 9M_{21}M_{42} + 12M_{21}^3 - 3M_{20}M_{43} - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22} \\
&C_{80} = M_{80} - 35M_{40}^2 - 630M_{20}^4 + 420M_{20}^2M_{40} \\
&C_{81} = M_{81} - 35M_{40}M_{41} + 630M_{22}^3M_{21} + 210M_{40}M_{20}M_{21} + 210M_{20}M_{41} \\
&C_{82} = M_{82} - 15M_{40}M_{42} - 20M_{41}^2 + 30M_{40}M_{20}^2 + 60M_{40}M_{21}^2 + 240M_{41}M_{21}M_{20} \\
&+90M_{42}M_{20}^2 - 90M_{20}^4 - 540M_{20}^2M_{21}^2 \\
&C_{83} = M_{83} - 5M_{40}M_{41} - 30M_{41}M_{42} + 90M_{41}M_{20}^2 + 120M_{41}M_{21}^2 + 180M_{42}M_{21}M_{20} \\
&+30M_{40}M_{20}M_{21} - 270M_{20}^3M_{21} - 360M_{21}^3M_{20} \\
&C_{84} = M_{84} - M_{40}^2 - 18M_{42}^2 - 16M_{41}^2 - 54M_{20}^4 - 144M_{21}^4 - 432M_{20}^2M_{21}^2 + 12M_{40}M_{20}^2 \\
&+96M_{41}M_{21}M_{20} - 144M_{41}M_{21}^2 + 72M_{42}M_{20}^2 + 96M_{41}M_{20}M_{21} \\
&M_{pq} = E\left[y(k)^{p-q}(y^*(k))^q\right]
\end{aligned}
\tag{12}
$$

3. Purposed method

3.1. Intelligence features selection

In this method, it is tried to reduce the dimensionality of modulation classification using GP in order to improve the classification performance. In pre processing step, a diagonal linear transform is applied on main features set. The steps of this operation are as follows:[7]

Let us to consider the input data set as D (features explained in section 2.2), $D = X \to C$ is a transform of main features set into modulation types. C is a set of different modulations $\{c_1, c_2, \ldots, c_m\}$, and m is the number of different modulations , then goal is to find a new space as $D' = Y \to C$ to make the relations between the input feature space and modulation types easier. For applying class wise transformation, the training data set should be divided into m parts depending on their modulation types [7].

If each partition is considered as a spherical cloud, then its axial should be placed within modulations boundaries.

$$X_i = \{x \mid (x,c) \in D, c = c_i\} where \cup_{i=1}^{m} \tag{13}$$

For finding a transform with above mentioned specifications, the data covariance should be diagonal in each part. The ith covariance is computed using Equ.13:

$$\sum\nolimits_i = E\{X_i X_i^T\} - E\{X_i\}E\{X_i^T\} \tag{14}$$

Where $\sum_i$ is a $n\times n$ square matrix including feature covariance depending on samples observed in i^{th} partition. This covariance matrix can be diagonal using data transform in each partition:

$$Y_i = \phi_i X, where: \phi_i = eigen(\sum_i) \qquad (15)$$

This equation transforms all training dataset but the computation of covariance and eigenvector are based on the data of each partitions. If n is the input space of training dataset, then each transformation will produce n intermediate features. This processing is iterated for each partition (for each modulation of the problem) finally, $m\times n$ new features will be produced. Then the D' data space will use transformed Y as Y= $\{Y_1, Y_2, \ldots Y_m\}$[7]. In this step, the features of modulation recognition and consequently the problem dimensions is increased.

This can't be used without using a suitable selection mechanism. GP search combines these elements for constructing the features of higher levels and use a ranking function for selecting the features containing more information.

The process of terminals sets construction is performed in preprocessing step. Since we have m separate modulations, $m \times n$ temporary features will be constructed and add with the main features of reference problem. They constitute the variable set of terminal pool in GP search. The variables number in intermediate buffer is $n\times(m+1)$. Dimension is increased in this step but the GP is allowed to find better feature with higher level. During the search process, an entropy-based fitness function is used to select the best features. A suitable fitness function should be able to separate informative and non informative features. This function is used to classify all solutions in the population. Generally, entropy measurement on a modulation interval is an efficient tool for minimizing the possibility of other modulations occurrence in its interval. This measurement system has two parts: modulation limit and a metric for measuring uncertainty level in the intervals. The entropy of an interval's data is considered as a criterion for measuring the uncertainty level. Here, Shannon's entropy is used:

$$H_2 = \log_2 \sum_{c\in C} \frac{1}{p_I^2(c)} \qquad (16)$$

Where I is the estimated interval, C is the sum of all modulations, $P_1(c)$ is the possibility of modulation c in interval I which is computed using the occurrence frequency of modulation c in its related interval [7].

The new feature construction problem is formulated as an optimization problem for minimizing the previous equation. $\sum_{c\in C} p_I^2(c)$ should be maximized in order to minimize the $H_2(I)$.This suggest that a feature can be considered as suitable if the modulation interval only includes samples from the specific modulation. The steps for evaluating a member of population I for target modulation of t (a feature to separate the t modulation samples) on a training dataset S is as follows:

- i program is used to transform each sample in S dataset into a transformed set S'. The main features are extracted from the terminals pool and produce a floating digit for each sample. Therefore, S' is 1D.
- An interval in S' including %99 of t modulation samples is considered as suitable interval.

- All samples of S' within I_t are gathered. Then $\sum_{c \in C} \frac{1}{p_I^2(c)}$ is calculated on the data of that interval.

In most samples of an interval which belongs to a given modulation, the fitness function's value is low. In this design, the lower fitness function improves the program and consequently produces better features. Only one GP run should be performed for each modulation. In each run, the fitness function is only concentrated on a given modulation. At the end of each run, the best evaluated program is selected as the feature of that modulation [7]. Therefore, the features number equals to modulation types in automatic modulation recognition.

4. Results and Analysis

GPLAB tools were used in this study (http://GPlab.sourceforge.net). For evaluating the performance of the purposed method, the signals modulated with 10 db and 5db SNRs and 64QAMو 256QAM و 2FSKو 4FSK و 2PSK و 4PSK modulations were studied.

The automatic modulation recognition involves two parts namely preprocessing and modulation classification. In the first one, the main features should be extracted first. For this purpose, 600 different signals which were modulated randomly with one of the given modulations, as well as 4000 sequent samples of each signal sample which were sampled with 80 kHz sampling frequency, were gathered as training data. GP was used to construct new features. The number of this feature equals to output modulations types. 6 modulations were studied in this paper. Tree structure was used to present GP output. Each tree produces a floating point as the output. The GP functions set contain four basic operators ÷ و × ، - ، +. The ramped half-and-half method was used to generate initial programs and the elitism policy was used to transfer the best parents to the next generation. The population size was 1024 and the maximum tree's depth when initializing was considered as 6. This can be increased to 12 when running. The evolution process is stopped after 25 generations.

The results obtained from signal classification with main and new features (extracted from GP) were compared with each other. Multi-layer neural network perceptron was used as classifier. The 10-fold cross validation method was used for classifiers test. Best results were shown as 'confusion matrix' in the following table.

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	64QAM	16QAM	**Simulated**
1	**1**	**0**	**0**	**1**	**99**	16QAM
3	**2**	**0**	**0**	**98**	**0**	64QAM
0	**0**	**0**	**100**	**0**	**0**	2FSK
1	**0**	**99**	**0**	**0**	**0**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
100	**0**	**0**	**0**	**0**	**0**	4PSK

Table1.modulation classification using Multi-layer neural network perceptron and extracted features using GP for SNR=10db

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	**64QAM**	**16QAM**	**Simulated**
1	**1**	**0**	**0**	**1**	**97**	16QAM
3	**2**	**0**	**0**	**95**	**0**	64QAM
0	**0**	**2**	**98**	**0**	**0**	2FSK
1	**0**	**97**	**1**	**1**	**0**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
99	**1**	**0**	**0**	**0**	**0**	4PSK

Table2.modulation classification using Multi-layer neural network perceptron and using main features for SNR=10db

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	64QAM	16QAM	**Simulated**
0	**1**	**0**	**0**	**0**	**99**	16QAM
1	**0**	**0**	**0**	**98**	**1**	64QAM
0	**0**	**0**	**100**	**0**	**0**	2FSK

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	**64QAM**	**16QAM**	**Simulated**
1	**1**	**0**	**0**	**0**	**98**	16QAM
2	**0**	**0**	**0**	**96**	**2**	64QAM
0	**0**	**3**	**97**	**0**	**0**	2FSK

0	**0**	**99**	**1**	**0**	**0**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
100	**0**	**0**	**0**	**0**	**0**	4PSK

Table3.modulation classification using Decision tree classifier and extracted features using GP for SNR=10db

0	**0**	**96**	**3**	**0**	**0**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
99	**1**	**0**	**0**	**0**	**0**	4PSK

Table4.modulation classification using Decision tree classifier and using main features for SNR=10db

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	64QAM	16QAM	**Simulated**
0	**1**	**0**	**0**	**1**	**98**	16QAM
1	**2**	**0**	**0**	**97**	**0**	64QAM
0	**0**	**1**	**99**	**0**	**0**	2FSK
1	**0**	**98**	**0**	**0**	**1**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
100	**0**	**0**	**0**	**0**	**0**	4PSK

Table5.modulation classification using Multi-layer neural network and extracted features using GP for SNR=5db

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	**64QAM**	**16QAM**	**Simulated**
1	**2**	**0**	**0**	**1**	**96**	16QAM
3	**2**	**0**	**0**	**95**	**0**	64QAM
0	**0**	**2**	**98**	**0**	**0**	2FSK
1	**0**	**97**	**1**	**1**	**0**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
99	**1**	**0**	**0**	**0**	**0**	4PSK

Table6.modulation classification using Multi-layer neural network perceptron and using main features for SNR=5db

Extracted modulation						
4PSK	**2PSK**	**4FSK**	**2FSK**	**256QAM**	16QAM	**Simulated**
0	**1**	**0**	**0**	**1**	**98**	16QAM
1	**2**	**0**	**0**	**97**	**0**	64QAM
0	**0**	**0**	**100**	**0**	**0**	2FSK
1	**0**	**98**	**0**	**0**	**1**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
100	**0**	**0**	**0**	**0**	**0**	4PSK

Table7.modulation classification using Decision tree classifier and extracted features using GP for SNR=5db

Extracted modulation						
4PSK 2PSK		**4FSK**	**2FSK**	**64QAM**	**16QAM**	**Simulated**
1	**1**	**0**	**0**	**1**	**97**	16QAM
2	**2**	**0**	**0**	**96**	**0**	64QAM
0	**0**	**2**	**98**	**0**	**0**	2FSK
1	**0**	**97**	**1**	**1**	**0**	4FSK
0	**100**	**0**	**0**	**0**	**0**	2PSK
99	**1**	**0**	**0**	**0**	**0**	4PSK

Table8.modulation classification using Decision tree classifier and using main features for SNR=5db

As it can be seen, the performance of Multi-layer neural network perceptron and Decision Tree Classifier in 5db and 10 db SNRs were improved considerably using features generated by GP. This is because that the genetic programming uses entropy fitness function to find features with higher classifying ability.

5. Conclusions

The aim of this paper was to use an evolutionary method to intelligently select the feature in such a way that the performance of Multi-layer Neural Network Perceptron and Decision Tree (j48) classifiers can be improved for different modulation classifications. This process has 4 important steps. The first step includes the extraction of ordinary features and establishment of training dataset. In the second step, class wise transform was used to transfer the main features space into a more classifiable space. In the third step, the new features space is used as the input of GP search terminal pool in order to make features with higher levels. Entropy concept was used to evaluate the features quality. In the forth step, the modulations were classified using Multi-layer neural network Perceptron.

The new features were studied in SNRs 5db and 10 db. The obtained results suggested that the purposed GP system can be used to improve the performance of Multi-layer neural network perceptron and Decision Tree classifiers. However, the process of new-features construction can be time consuming and difficult.

Acknowledgements

This research has been supported by the "IRAN's Ministry of Communications and Information Technology-Communications and Information Technology Research Institute"

References

[1]. B. Carlson, P. B. Crilly and J. C. Rutledge. (2001). "Communication system", McGraw Hill, Fourth edition.

[2]. Y. O. Al-Jalili. (1995). " Identification algorithm of upper sideband and lower sideband SSB signals", Signal Processing, Vol. 42, pp. 207-213,

[3]. M. L. D. Wong and A. K. Nandi, "Semi-blind algorithms for automatic classification of digital modulation types," Digital Signal Processing, vol. 18, pp. 209-227, 2008.

[4]. M. L. D. Wong and A. K. Nandi. (2004). "Automatic digital modulation recognition using artificial neural network and genetic algorithm," Sig. Proc., vol. 84, pp. 351-365.

[5]. Z. Zhu, M.W. Aslam and A.K. Nandi. (2010). "Augmented genetic programming for automatic digital modulation classification," in Proc.IEEE MLSP 2010, pp. 391-396,

[6]. L. Zhang, L. B. Jack and A. K. Nandi. (2005). "Extending genetic programming for multi-class classification by combining k-nearest neighbor," IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 349-352.

[7]. K.Neshatian. and M.Zhang. (2008). "Genetic programming for performance improvement and dimensionality reduction of classification problem" IEEE World Congress on Computational Intelligence(CEC), Hong Kong

[8]. Wong, M.L.D., and Nandi, A.K(2001). 'Automatic digital modulationrecognition using spectral and statistical features with multi-layer perceptrons'. Proc. Int. Symp. Signal Processing and its Applications, Kuala Lumpur, Malaysia, , pp. 390–393

*[9]. E. Azzouz and A.Nandi. (1998). "*Algorithms for Automatic Modulation Recognition of Communication Signals*," IEEE Trans. Commun., vol. 46, no. 4, pp. 431-436.*

[10]. *Swami, A., and Sadler, B.M. (2000) 'Hierarchical digital modulation classification using cumulants', IEEE Trans. Commun pp. 416–429*