

Intelligent Modulation Type Identification Using GA-SVM Based on WPA

¹ATAOLLAH EBRAHIMZADEH, ²SEYED ALIREZA SEYEDIN

Department of Electrical and Computer Engineering

¹Noshirvani Institute of Technology, ^{1,2}Ferdowsi University of Mashad
IRAN

Abstract: - Automatic digital modulation identification through signal processing has found many applications such wireless communication system, nowadays. This paper proposes an Intelligent Digital Modulation Identifier (IDMI) that is new in this area. IDMI includes three main parts: wavelet packet analysis (WPA), support vector machines (SVM) and genetic algorithm (GA). WPA extracts features from received signal. SVM classifies according to extracted features. GA finds the best kernel parameter for SVM also does features selection. IDMTI is used to separate PSK2 and PSK4. Simulation result show IDMI has high performance comparable other methods even in low SNR.

Key-Words: - Genetic algorithm, modulation identification, support vector machine, wavelet packet analysis.

1 Introduction

Automatic modulation type identifier is a system which recognizes the modulation type of received signal automatically and has many applications such as: spectrum surveillance, signal confirmation, multi-drop networks, etc. With increasing usage of digital modulations such as phase shift keying (PSK) in many novel applications the need to find efficient methods for their discrimination in presence of noise has become increasingly important.

Generally, modulation type identification methods can be categorized into two main categories: decision theoretic (DT) or maximum likelihood (ML) and pattern recognition (PR). DT approaches use probabilistic and hypothesis testing arguments to formulate the recognition problem and to obtain the classification rule [1-2]. The major drawbacks of these approaches are their too high computational complexity and difficulty within the implementation decision rule. They need precise probability density function of received signal. Also, these methods have difficulties to set the correct threshold values. PR approaches, however, do not need such careful treatment. PR approaches mainly divided into two subsystems: the feature extraction subsystem and the classifier subsystem. The former subsystem is responsible for extracting prominent characteristics from received signal which are called features and the latter, classifier, is employed to indicate the membership of modulation type based on the features extracted [3-6]. PR approaches are more tractable, simple to implement and have more practical structure; however, selection of two subsystems is most serious problems.

It is showed modulation identification using ANNs outperforms other techniques [4-6]. However, the ANN techniques have limitations on generalization giving rise to models that can overfit to the training data. Recently, support vector machines (SVMs), based on statistical learning theory are gaining applications in the areas of machine learning and pattern recognition because of high accuracy and excellent generalization capability [7]. The main difference between ANNs and SVMs is in the principle of risk minimization (RM). The difference in RM leads to better generalization performance for SVMs than ANNs. In this paper we have used SVM as a classifier.

As mentioned the features have a vital rule in modulation identification. In this work we have used the WPA that has been introduced [5]. However there are large amount of redundant features that employed in [5]. There is a need to make the identification process faster and accurate using the minimum number of features which primarily characterize the type of modulation with an optimized structure of SVM. In this paper a genetic algorithm (GA) is used for optimization.

The paper is organized as follows. Section 2 presents the WPA and features. Section 3 introduces the SVM classifier. Section 4 presents the GA. In section 5 simulation results are shown. Finally section 6 concludes the paper.

2 WPA and features extraction

It is most important to find a set of features which could be used to differentiate members of PSK

modulation class; specially PSK2 and PSK4. Among the different types of features that we have evaluated WPA shows the best performances to extracting the effective features. WPA preserves the time, frequency and phase information. WPA can be considered as natural extensions of the wavelet transform [8]. The WPA technique calls for selecting a wavelet to be used in the decomposition of the input signal. The full WPA is displayed as a tree with a discrete sequence (coefficient) at every branch. Each branch is referred to as bin vector. Figure 1 shows the WPA for a signal of length eight. The S in the first row represents the entire signal, with no decomposition. The second row represents the first level of decomposition. There, node A holds the approximation coefficients and D holds the detail coefficients.

S							
A				D			
AA		DA		AD		DD	
AAA	DAA	ADA	DDA	AAD	DAD	ADD	DDD

Figure1: Tree of WPA for a signal of length eight

Careful choices of the wavelet family and zero-crossing order led to a decomposition that extracts unique characteristics of each signal. Therefore we should determine the appropriate wavelet to use. The complex envelopes of biorthogonal wavelet proved to be the key. Biorthogonals are compactly supported spline wavelets where symmetry and exact reconstruction are possible [9]. One of the unique features of biorthogonal wavelets is preservation of phase linearity during decomposition [8-9]. It would seem desirable to preserve phase coherence when decomposing signals with phase modulation. Biorthogonal 3.1 (reconstruction order of 3 and a decomposition order of 1), provided the best separability of the PSK2 and PSK4 signals.

After WPA, Shannon entropy function acts as an entropy measure [9]. This entropy function was used for characterizing each of the sets of coefficients produced by the wavelet decomposition. Other entropy functions, such as kurtosis, were tried but failed to produce the consistent results of Shannon's function. The resulting values were used to form the input vector for the classifier.

3 Classifier

Support Vector Machine (SVM) is a supervised machine learning technique that can be applied as the robust tool classifier. The SVM performs classifica-

tion tasks by constructing separating hyperplanes [7]. The idea of SVM can be expressed as follows. Suppose the training set, $(x_i, y_i), i = 1, 2, \dots, l, x \in R^d, y \in \{-1, +1\}$ can be separated by the hyperplane $w^T x + b = 0$, where w weight vector and b is bias. If this hyper-plane maximizes the margin, then the following inequality is valid for all input data:

$$y_i(w^T x_i + b) \geq 1, \text{ for all } x_i \quad i = 1, 2, \dots, l \quad (1)$$

The margin of the hyper-plane is $2/\|w\|$. The hyper-plane which maximizes $\|w\|^2$ and is subject to constraints (1) is called optimal separating hyperplanes (OSH). Those training points for which the equality in (1) holds are called support vectors. For a training set, in order to find OSH, we need to solve the convex quadratic programming (QP) problem. This problem has a global optimum, and Lagrange multipliers are used to solve it. After solution the QP, the optimal weights and the bias are given by:

$$w^* = \sum_{i=1}^l \lambda_i^* y_i x_i \quad (2)$$

$$b^* = y_i - w^{*T} x_i \quad (3)$$

where $\lambda_i, i = 1, \dots, l$ are the Lagrange multipliers ($\lambda_i \geq 0$) for any support vector x_i . The optimal decision function is then given by:

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \lambda_i^* x^T x_i + b^*) \quad (4)$$

For input data with a high noise level, SVM uses soft margins can be expressed as follows with the introduction of the non-negative slack variables $\xi_i, i = 1, \dots, l$:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \text{ for } i = 1, 2, \dots, l \quad (5)$$

To obtain the OSH, it should be minimizing the

$$\Phi = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i^k \text{ subject to constraints (5).}$$

where C is a regularization constant that controls how heavily training errors are penalized.

In the nonlinearly separable cases, the SVMs map the training points, non-linearly, to a high-dimensional feature space using kernel function $K(\vec{x}_i, \vec{x}_j)$, where linear separation may be possible.

The most famous kernel functions are Linear, Polynomial, RBF, and Sigmoid. First a kernel function is selected. After training, the following, the decision function, becomes:

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \lambda_i^* K(x, x_i) + b^*) \quad (6)$$

In the present work, Gaussian RBF (GRBF) kernel

function, given by (7) is used. It shows better performances in modulation identification in comparison other kernels.

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (7)$$

The generalization capability of SVMs can be controlled through a few free parameters like the term C and the kernel parameter like RBF width (σ). In the present work, C is chosen as 10 and the kernel parameter (σ) and features are selected using GA-based approach.

4 Genetic algorithm

From the section 2 it is found that the total number of features was a lot. This number of features causes the high computational complexity in the identifier. On the other hand, although some of these features may carry good classification information when treated separately, there is little gain if they are combined together (due to sharing the same information content). Thus it is needed to perform the features selection to increase the execution speed of identifier. Having a favorable set of features, it is also crucial that the classifier (SVM) is optimized so as to have best model. Both of the features selection and the model selection could be done with an optimization problem. Among the methods that exist, the GA is chosen. GA, not only has ability to efficiently search large spaces about which little is known, but also is easy to implementation [10].

GA has been considered with increasing interest in a wide variety of applications. The use of genetic algorithm needs consideration of six basic issues: chromosome (genome) representation, selection function, genetic operators like mutation and crossover for reproduction function, creation of initial population, termination criteria, and the evaluation function. Though the traditional genome representation has been in binary form, the interest in real-coded or floating-point genomes for multidimensional parameter optimization problems is on the rise because of the closeness of the second type of representation to the problem space, better average performance and more efficient numerical implementation.

In this paper, real-coded genomes were used. The genome X contains $N+1$ real numbers $X = \{x_1 x_2 \dots x_N x_{N+1}\}^T$. The first N numbers in the genome are the selected features from the total features (R) and constrained to be in the range $1 \leq x_i \leq R$. The last number (x_{N+1}) shows the RBF kernel width value. The selection of kernel width for

good classification accuracy is one of the major issues in SVMs.

A probabilistic selection, namely, normalized geometric ranking method [11] was used based on the individual's fitness such that the better individuals have higher chances of being selected. A non-uniform mutation function [12] using a random number for mutation based on current generation and the maximum generation number, among other parameters was adopted. Heuristic crossover [12] was used. This operator produces a linear extrapolation of two individuals using the fitness information. To start the solution process, the random generation of initial solutions for the population is used. The maximum number of generations was adopted as the termination criterion for the solution process. The fitness function used in the present work is defined the overall performance (OP) of identifier, which is an average performance of training and testing. The better classification results give rise to higher fitness index.

5 Simulation Results

This section presents the simulation results of the proposed identifier.

In this study, the received radio signal is assumed to have been corrupted with additive white Gaussian noise (AWGN). A numbers of PSK2 and PSK4 radio signals were generated under various noise conditions (different SNRs). The signals were processed every 512 samples. These are then divided into training and testing data sets. To obtain wavelet based Shannon entropy estimates, this data was segmented into cuts of length 64. The level of decomposition is chosen four. Thus full wavelet decomposition was performed on this data with the bi-orthogonal 3.1 wavelet. Shannon entropy analysis was performed on all 30 coefficient packets, plus the original signal sample, to provide 31 parameters to describe the signal [9]. The 31 entropy values were combined to form an input vector for classification. Target values were specified 1 and -1 representing PSK2 and PSK4, respectively.

Firstly, we evaluate the performance of proposed identifier without optimization (Straight Identifier). The full features are used. The constant parameter of kernel function was chosen $\sigma = 1$. Table 1 shows the overall performance (OP) of the identifier for various SNR values. Performance is generally good even with low SNRs.

Table1: Performances of straight identifier

SNR (dB)	OP (%)
-2	68.02
0	74.48
4	98.75
8	99.93
12	99.96

Now, we apply the GA for the optimization. In the GA, a population size of ten individuals was used starting with randomly generated genomes. This size of population was chosen to ensure relatively high interchange among different genomes within the population and to reduce the likelihood of convergence within the population. One has to specify the number of features that varies 3 to 31 (total number of features). Therefore, experiments were carried out to investigate the possibility of even smaller data set and the compromise in performance that one might observe. RBF kernel width was selected in the range 0.4-1.4 with a step size of 0.2. Table 2 indicates this performance using eight features selected by GA. Also at each SNR, the optimal kernel width (OKW) is identified. Results indicate that identifier records a performance degradation of about 0.5% only at SNR= -2dB. For other levels of SNR, the difference is negligible. Thus it can be said that the proposed method achieves high performance on most SNR values with only eight features has that selected using GA.

Table2: Performance with eight FS using GA

SNR	No. FS	OKW	OP with GA
-2	8	1	67.53
0	8	0.8	73.96
4	8	0.6	98.52
8	8	1	99.83
12	8	1	99.94

We have compared our novel technique with the technique that has been proposed in [3] (TECH2). Figure 2, shows the performances of these techniques with the same set ups. Our intelligent identifier outperforms TECH2. It has high performance at low levels of SNR. This is due the two facts: chosen novel features and chosen classifier. The chosen features have effective properties in signal representation. On the other hand, the SVM classifier has high generalization ability such that enables it to classify the non-separable data (low SNR) with high accuracy. The SVM classifier sufficiently uses the information of features and maps them nonlinearly into a high dimensional feature space and constructs the optimum separating hyperplane to make the data

linearly separable. Thus the probability is high in the whole SNR range and the algorithm has the excellent performance of identification. Cumulant-based technique only use the partial information of higher order cumulant, so in low SNR values it is poor.

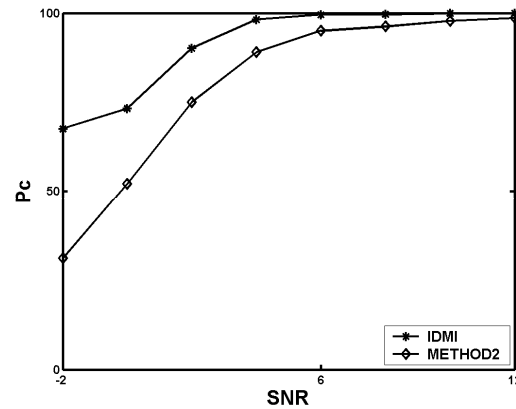


Figure2: Comparison of proposed identifier with TECH2.

6 Conclusion

Automatic modulation type identification is an important issue in communication intelligence and electronic support measure systems. This paper proposes a new efficient intelligent system uses high effective features and classifier that enables it to have high performance. The SVM classifier, use the feature vector and maps the input vectors non-linearity into high dimensional feature space and constructs the optimum separating hyper-plane in the space to realize signal recognition. This technique is robust and avoids over-fitting and local minimum. Optimization technique enables it to achieve a highly performance by only a few number of features. It might be to combine this technique with other techniques (such as the technique has been introduced in [4] and did not work as well with PSK modulation types) and construct a general identifier that identifies different types of modulation with high accuracy.

Acknowledgements

The authors would like to thank ITRC and TCKH.

References:

- [1] C. Le Martret, and D. Boitea, A general maximum likelihood classifier for modulation classification, *Proc. ICASSP*, Vol. 4, 1998, pp. 2165-2168..
- [2] A. Ebrahimzadeh, S. A. Seyedin, Identifying of digital modulation without prechannel estimation, *Proc. ICEE*, Vol.4, 2004, pp. 180-185.

- [3] A. Swami, and B.M. Sadler, Hierarchical digital modulation classification using cumulants, *IEEE Trans. Comm.*, Vol. 48, No. 3, 2000, pp. 416–429.
- [4] A. K. Nandi, and E. E. Azzouz, Algorithms for automatic modulation recognition of Communication Signals, *IEEE Trans. Commons.*, Vol. 46, No. 4, 1998, pp. 431-436.
- [5] A. Ebrahimzadeh, and S. A. Seyedin, A new method for automatic digital signal type identification, *Proc. CIS*, 2005, pp. 156-161
- [6] S. A. Seyedin, and, A. Ebrahimzadeh, Automatic signal type identification by neuralwavelet based method, *Proc. SICED*, 2005, pp. 350-355.
- [7] C. Burges “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Mining and Knowledge Discovery*, Vol. 2, 1998, pp 121-167.
- [8] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2002.
- [9] M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi, *Wavelet Toolbox User's Guide*, The MathWorks, Inc., 2002.
- [10] G. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, NY, USA, 1989.
- [11] C.R. Houk, J. Joines, M. Kay, *A genetic algorithm for function optimization: a matlab implementation*, North Carolina State University, Report no.: NCSU IE TR 95 09, 1995.
- [12] Z. Michalewicz, *Genetic algorithms+Data Structures=Evolution Programs*, 3rd Edition, Springer, NY, USA, 1999.