A Novel Hierarchical Method for Digital Signal Type Classification

[1]ATAOLLAH EBRAHIMZADEH, [2]SEYED ALIREZA SEYEDIN
Department of Electrical and Computer Engineering
[1]Nousirvani Institute of Technology   [1,2]Ferdowsi University of Mashad
IRAN

Abstract: - Automatic signals type classification (ASTC) is a technique for recognizing the modulation scheme of an intercepted signal and has seen increasing demand in both military and civilian. Most of previous techniques can identify only a few kinds of signal types. Also, usually, they request high SNR to achieve a minimum acceptable performance and don't include higher order and new types of signals (e.g. QAM types).The work presented here proposes novel high performance method for identification of digital signal types. In this method it is proposed a hierarchical multiclass classifier based on support vector machine (SVM). The inputs of this classifier are higher order moments and cumulants (up to eight). Genetic algorithm (GA) is used to select the best model of SVMs in order to improve the performance. Simulation results show that proposed method can recognize a lot of digital signal types and achieves a highly probability of classification even at low signal to noise ratios (SNRs).

Key-Words: - Pattern recognition, modulation, machine learning, support vector machine, model selection, eight order statistics.

1 Introduction

Automatic signals type classification is a technique for recognizing the modulation scheme of an intercepted signal. It plays an important role for various applications ad purposes. For example, in military application, modulation classification can be employed for electronic surveillance, interference identification, monitoring; in civil applications, it can be used for spectrum management, network traffic administration, different data rate allocation, signal confirmation, interference identification, software radios, multidrop networks, intelligent modem, etc.

In the past, signal type recognition relied mostly on operators scanning the radio frequency spectrum with a wide-band receiver and checking it visually on some sort of display. Clearly, these methods relied very much on the operators' skills and abilities. These limitations then led to the development of more automated modulation recognizers. One semi-automatic approach was to run the received signal through a number of demodulators and then have an operator determine the modulation format by listening to the output of each demodulator. This approach is however not very practical anymore due to the new digital techniques that transfer both voice and data. Then techniques for automatic signal type classification (ASTC) started to emerge. Whilst early researches ASTC were concentrated on analogue modulations and lower orders of some types of digital modulations, such as frequency shift keying (FSK), amplitude shift keying (ASK), the recent contributions in the subject focus more on higher order and new types of digital signals like phase shift keying (PSK), amplitude shift keying (QAM). Primarily, this is due to increasing usage of such modulations in many novel applications.

ASTC techniques usually can be categorized in two main principles: the decision theoretic (DT) and the pattern recognition (PR). DT approaches use probabilistic and hypothesis testing arguments to formulate the recognition problem and to obtain the classification rule [1-3]. The major drawbacks of these approaches are their very high computational complexity, difficulty within the implementation of decision rule, and lack of robustness to model mismatch. Also, these methods have difficulties to set the correct threshold values. PR approaches, however, do not need such careful treatment. These methods are simple to implement .They can be further divided into two subsystems: the feature extraction subsystem and the classifier subsystem. The former extracts prominent characteristics from the raw data, which are called features, and the classifier identifies the type of signal based on the extracted features [4-12]. It is showed that in modulation identification area, ANNs especially multilayer perceptron (MLP) outperform other classifiers [8-12]. However, with regard to effectiveness of ANNs, there are some problems worth mentioning. The traditional ANNs have limitations on generalization giving rise to models that can overfit to the training data. ANNs show

poor performances in low SNRs and higher number of modulation schemes usually requires a more complex neural network and higher training time [13, 14].This deficiencies are due to the optimization algorithms used in ANNs for selection of parameters and the statistical measures used to select the model.

Recently, support vector machines (SVMs), based on statistical learning theory are gaining applications in the area of machine learning and pattern recognition because of high accuracy and excellent generalization capability. In [13, 14], the authors showed that using SVM in content of signal type identification achieves highly success rate. Thus in this paper we use SVM as a classifier. Selection the proper model of classifier can improve a substantial amount of is performance. In this paper we have used the GA for SVM's model selection.

Plenty of ASTC methods are able to recognize only a few kinds of signals types and/or lower order of modulations; also, usually, they request high SNR for having a minimum acceptable performance. This is due to facts: classifier and features. Although the classifier has an important role in classification, it should be mentioned that the features have vital role. In this work we have used higher order moments and cumulants as features. The work presented here, proposes a high efficient ASTC method, which is able to recognize different type of received signals.

The paper organized as follows. Section 2 presents feature extraction. Also the considered modulation set is introduced in this section. Section 3, describes the classifier. Section 4, introduced the GA that is used for model selection of SVMs. Section 5, shows some simulation studies. Finally, in section 6 conclusions are presented.

2   Feature extraction

In modulation identification problem, finding the proper features is very important. For example QAM modulation schemes contain information in both amplitude and phase (that are regarded as complex signals), thus finding the proper feature that could be able to identify them (especially in case of higher order and/or non-square) is difficult. Based on our researches, the higher order moments and higher order cumulants up to eighth achieve the most highly performances to discriminating of digital modulations such as considered set in this paper. These features have many advantages e.g. they provide a good way to describe the shape of the probability density function.

Probability distribution moments are a generalization of concept of the expected value [13]. The auto-moment of the random variable may be defined as follows:

$$M_{pq} = E[s^{p-q}(s^*)^q] \tag{1}$$

where p called moment order. Assume a zero-mean discrete based-band signal sequence of the form $s_k = a_k + jb_k$. Using the (5), the expressions for different order may be easily derived.

The symbolism for p^{th} order of cumulant is defined as:

$$C_{pq} = Cum[\underbrace{s,...,s}_{(p-q)terms}, \underbrace{s^*,...,s^*}_{(q)terms}] \tag{2}$$

There are relations between moments and cumulants. The nth order cumulant is a function of the moments of orders up to including n. Moments is expressed in terms of cumulants as:

$$M[s_1,..,s_n] = \sum_{\forall v} Cum\left[\{s_j\}_{j\in v_1}\right]..um\left[\{s_j\}_{j\in v_1}\right] \tag{3}$$

where the summation index is over all partitions $v = (v_1,...,v_q)$ for the set of indexes $(1,2,...,n)$, and q is the number of elements in a given partition. Cumulants may be also be derived in terms of moments. The nth order cumulant of a discrete signal $s(n)$ is given by:

$$Cum[s_1,..,s_n] = \sum_{\forall v} (-1)^{q-1}(q-1)! E[\prod_{j\in v_1} s_j]..E[\prod_{j\in v_q} s_j] \tag{4}$$

In this paper the considered modulations set is {ASK4, ASK8, PSK2, PSK4, PSK8, Star-QAM8, V29, QAM16, QAM64, and QAM128}. We have computed the features of considered digital modulation set.

3   Classifier

As said we have used a hierarchical SVM-based structure as classifier. Support Vector Machine (SVM) is a supervised machine learning technique that can be applied as robust tool for binary and multiclass classifications [15]. In case of SVMs, structural risk minimization (SRM) principle is used minimizing an upper bound on the expected risk whereas in ANNs, empirical risk minimization (ERM) is used minimizing the error on the training data. The difference in RM leads to better generalization performance for SVMs than ANNs. The SVM performs classification tasks by constructing optimal separating hyperplanes (OSH). The idea of binary SVM can be expressed as follows [16].

3.1 Binary SVM

Suppose the training set, $(x_i, y_i), i = 1,2,....,l, x \in R^d, y \in \{-1,+1\}$ can be separated by the hyperplane $w^T x + b = 0$, where $\bar{w}$ is

weight vector and b is bias. If this hyperplane maximize the margin, then the following inequality is valid for all input data:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \text{for all } \mathbf{x}_i \quad i = 1,2,...,l \quad (5)$$

The margin of the hyper-plane is $2/\|\vec{w}\|$, thus, the problem is: maximizing the margin by minimizing of $\|w\|^2$ subject to constraints (5), that is a convex quadratic programming (QP) problem. This problem has a global optimum, and Lagrange multipliers are used to solve it:

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i \left[y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \right] \quad (6)$$

where $a_i, i = 1,...,l$ are the Lagrange multipliers ($\alpha_i \geq 0$).The solution to this QP problem is given by minimizing L_P with respect to w and b. After differentiating L_P with respect to w and b and setting the derivatives equal to 0, yields:

$$\mathbf{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \mathbf{x}_i \quad (7)$$

It can obtain the dual variables Lagrangian by imposing the Karush-Kuhn-Tucker (KKT) conditions:

$$L_d = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\|\mathbf{w}\|^2$$
$$= \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (8)$$

To find the OSH, it must maximize L_d under the constraints of $\sum_{i=1}^{l} \alpha_i y_i = 0$, and $\alpha_i \geq 0$. Note that the Lagrange multipliers are only non-zero ($\alpha_i \succ 0$) when $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$. Those training points for which the equality in (5) holds are called support vectors (SV) that can satisfy $\alpha_i \succ 0$. The optimal weights are given by (7) and the bias is given by:

$$b^* = y_i - \mathbf{w}^{*T}\mathbf{x}_i \quad (9)$$

for any support vector x_i. The optimal decision function (ODF) is then given by:

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{l} y_i \alpha_i^* \mathbf{x}^T \mathbf{x}_i + b^*) \quad (10)$$

where α_i^*'s are optimal Lagrange multipliers.

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 $\xi_i, i = 1,...,l$:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad for \quad i = 1,2,...,l \quad (11)$$

To obtain the OSH, it should be minimizing the $\Phi = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} \xi_i^k$ subject to constraints (11), where C is the penalty parameter, which controls the tradeoff between the complexity of the decision function and the number of training examples misclassified, i.e., controls the tradeoff between margin maximization and error minimization. This problem can be solved with similar method.

In the nonlinearly separable cases, the SVM map the training points, nonlinearly, 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, radial basis function (RBF), and sigmoid. For example RBF will be shown by:

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (12)$$

where σ is the width of the RBF kernel. After a kernel function is selected, the QP problem is:

$$L_d = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (13)$$

the α_i^* is derived by:

$$\alpha_i^* = \arg \max_{\alpha} L_d$$
$$0 \leq \alpha_i \leq C; i = 1,2,...,l; \sum_{j=1}^{l} \alpha_i y_i = 0 \quad (14)$$

After training, the following, the decision function, becomes:

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{l} y_i \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*) \quad (15)$$

The performance of SVM can be controlled through a few free parameters like the term C and the kernel parameter which called are hyperparameters.

3.2 Multiclass SVM

As said SVM was originally developed for binary classification problems, but it can be used for multiclass classification. There are several approaches available to extend binary SVMs to multi-class problems that fall into two categories: the one-against-others method and one-against-one method [17]. In this research, we have proposed a novel, simple and effective solution for mutilclass SVM-based that is hierarchical. In this scheme at first all of inputs fed to one SVM. This SVM separates them in two groups. Each of these groups fed to another SVMs and process will be continued till all of patterns are classified. The details of this structure introduced in section 5.

4 Model selection and GA

Driving the optimal values of SVMs parameters are important to achieve high generalization ability and performance. GAs with their characteristics of high efficiency and global optimization are widely applied in many areas. In this paper, according the determined appropriate fitness function for GA operation, an effective strategy for parameters selection for SVM is proposed by using the GA.

GA is a stochastic optimization algorithm which adopts Darwin's theory of survival of the fittest. To apply GAs to the SVM parameters selection problem, one has to consider the following issues: the encoding scheme, the methodology to produce the initial population, the fitness function and the genetic operators such as reproduction, crossover and mutation.

4.1 Encoding Scheme

Here real-encoded scheme is selected as the representation of the parameters in this paper. The research space of these parameters is $C \in [5:5:50]$, $\sigma \in [0.1:0.1:3.0]$

4.2 Produce the Initial Population

Because the real-coded scheme is used, the solution space coincides with the chromosome space. Considering the bigger population will enlarge the GA running time and disperse the conformation of the population, the size of population, pop_size is choose 16 in ordering to avoiding the convergence of the population becomes difficult. For producing the initial population, the initial values of the designed parameters are distributed in the solution space as even as possible.

4.3 Fitness Function

According to the aforementioned analysis, the average performance of the SVM classifier is depended on $E\{R^2/\gamma^2)$ and not simply on the large margin γ. The Radius-margin bound [18] is proposed as the fitness function

$$T = \frac{1}{l}\frac{R^2}{\gamma^2} \qquad (16)$$

where γ denotes the margin, l is the size of the training samples, R is the radius of the smallest sphere containing the training data, $R = 0.5$.

4.4 Genetic Operators

Genetic operator includes the following three basic operators such as selection, crossover and mutation. Here, a heuristic search strategy is adopted to realize the genetic optimization for automatic parameters selection. The adopted GA operators are briefly presented as follows.

4.4.1 Selection operator

Selection operators is composed of a copy selection operation and a survive selection operation. Here the method of survival of the fittest was used to select the next generation individual.

Given the fitness function $fit(a_i)$ of the individual a_i, the probability of a_i selected as the next generation one is as follow:

$$P(a_i) = \frac{fit(a_i)}{\sum_{j=1}^{pop\_size} fit(a_j)} \times pop\_size \qquad (17)$$

4.4.2 Crossover operator

The means of crossover implement is closely integrated to the encoding scheme. Due to the real-encoding scheme is utilized; the crossover operator in this paper can be defined as [19]:

$$P' = aP_1 + (1-a)P_2 \qquad (18)$$

where P' is the offspring after crossover operation, P_1 and P_2 are two parents to be implemented the crossover operation, and a is a constant which belongs to (0,1). Here a =0.5.

4.4.3 Mutation operator

How the bigger value of the mutation operator is chose to maintain the diversity of the population in the early GA operation and avoid the precocity? The adaptive mutation probability is adopted in this paper to solve the above two problems as follows:

$$P_m = \frac{\exp(-b \times t/2)}{pop\_size \times \sqrt{L}} \qquad (19)$$

where t is the generation of the genetic iteration, pop_size is the size of the population, L is the length of the individual, b=1.5 is a preset parameter.

4.4.4 The stopping criteria

In general GA algorithm, terminate the program when the best fitness has not changed more than a very small value, i.e. 10^{-6} over the last generations. But we choose the average fitness rather than best fitness as the stopping criteria.

5   Simulation study

This section present some evaluation results of proposed identifier for modulation set that is defined in section 2. For simplifying the indication, we substitute the modulations ASK4, ASK8, PSK2, PSK4, PSK8, Star-QAM8, V29, QAM16, QAM64,

QAM128 with P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10} respectively. All signal are digitally modulated in MATLAB simulator. The simulated signals were also band-limited and Gaussian noise was added according to SNRs, –3, 0, 3, 6, 9, 12, and 18 dB. Each modulation type has 1200 realizations of 2048 samples. Figure 1 shows the hierarchal SVM-based classifier. Table 1 shows the chosen features for each SVM that discriminate signal types with them.

At the first, we evaluate the performance of system without GA. Based on some simulations, the value σ =1 is selected for all SVMs. The effect of noise on classifier performance is also studied with different SNRs. Random only 20% of data is used for training. Overall performance (OP) is shown in Table 2. As we see, the performance is generally good even with low SNR values.
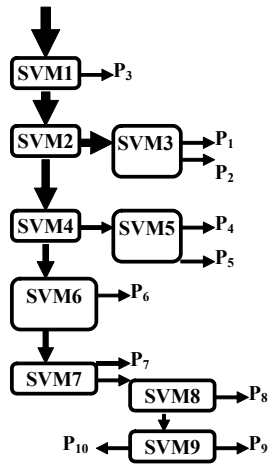

Figure1. Hierarchical SVM-Based Classifier

Table 1: Chosen features for each SVM

SVM's Number	Chosen features
SVM 1	C_{83}
SVM 2	M_{41}
SVM 3	M_{41}
SVM 4	C_{80}
SVM 5	C_{80} , M_{61}
SVM 6	C_{63}
SVM 7	C_{82}
SVM 8	C_{63}
SVM 9	C_{80} , M_{84}

Table 2: OP without SVMs model selection at different SNRs.

SNR (dB)	OP (%)
-3	84.23
0	91.67
3	92.11
6	94.12
9	96.88
18	98.05

Now, we apply the GA or model selection of SVMs. We have used the GA method for each of SVM separately. One of the advantages of hierarchical multiclass-SVM based in comparison others multi-class SVM-based methods (i.e. one-against-all and one-against-one) is that optimization of each SVM could be done separately. In the other methods even though each SVM is tuned very well for the binary problem, there is no guarantee that they work well together for the entire problem and the parameters of the kernels affect the structure of the feature space and the classification accuracy. Table 3, shows the overall performance of proposed identifier. As we see, model selection exactly improves the performances of system in all SNRs; especially in lower SNRs.

Table 3: OP with applying of GA for model selection at different SNRs.

SNR (dB)	OP (%)
-3	89.8
0	93.17
3	94.93
6	97.35
9	98.43
18	98.76

For comparison our method with other methods as said there are a few researches on modulation identification that consider the number of modulations about ten or a little and most of methods that have been proposed have limitation to consider additional modulation. This is mainly because of their features and/or classifiers. In [10], the author reported a generalization rate of 90% and 93% of accuracy of data sets with SNR of 15–25 dB for considered modulation set. However, the performances for lower SNRs are reported to be less than 80% for a fully connected network, and about 85% for a hierarchical network. In [11], the authors developed and compared one decision tree and one neural network classifier. Half of the simulated signals used for training had a SNR value of 15 dB and the other half 20 dB. It used instantaneous features in addition spectral features. The neural network classifier consisted of three MLPs. The results showed 88% success rate at 15 dB SNR for considered modulations set. In [12], the researches show the average identification rate is 83%, and it reaches over 90% for SNR value of over 20 dB. However, if SNR is less than 10 dB, the performance drops to less than 70%. The identifier proposed in this work shows a steady performance with different SNR values and has a high performance at very low SNRs. Using an opti-

mization method, the performance of identifier increase. The structure of proposed method is very simple. The proposed method has a high generalization ability and accuracy.

6 Conclusions

Automatic signals type classification (ASTC) has seen increasing demand in both military and civilian. Most of previous techniques can identify only a few kinds of signal types and usually need high SNR to achieve a minimum acceptable performance and don't include higher order and new types of signals. The work presented here proposes novel high performance method for identification of digital signal types. In this method we have used a hierarchical multiclass classifier based on support vector machine (SVM). The inputs of this classifier are higher order moments and cumulants (up to eight). 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 method is robust and avoids over-fitting and local minimum. Optimization using GA, improves the performance of system especially in lower SNRs.

References:

[1] C. Le Martret, and D. Boitea, A general maxi mum likelihood classifier for modulation classification, *Proc.   ICASSP,* Vol. 4, 1998, pp. 2165-2168.

[2] J. A. Sills, Maximum-likelihood modulation classification for PSK/QAM, *Proc. MILCOM,* 1999, pp. 57-61.

[3] P. Panagotiou, and A. Polydoros, Likelihood ratio tests for modulation classifications, *Proc. MILCOM, 2000,* pp. 670-674.

[4] K.C. Ho, W. Prokopiw and Y.T. Chan, Modulation identification of digital signals by wavelet transform, *Proc. IEE ,Radar, Sonar Navig.,* Vol.147, No.4, 2000,pp. 169-176.

[5] B. G. Mobas, Digital modulation classification using constellation shape, *Signal Processing*, Vol. 80, 2000, pp. 251–277.

[6] J. Lopatka, and P. Macrej, Automatic modulation classification using statistical moments and a fuzzy classifier, *Proc. ICSP,* 2000, pp.121-127.

[7] A. Swami, and B. M. Sadler, Hierarchical digital modulation classification using cumulants," *IEEE Trans. Comm.,* Vol. 48, No. 3, 2000, pp. 416–429.

[8] S. A. Seyedin, and, A. Ebrahimzadeh, Automatic signal type identification by neuralwavelet based method, *Proc. SICED,* 2005, pp. 350-355.

[9] A. Ebrahimzadeh, S. A. Seyedin, A new method for identifying of signal type, *Proc. CIS,* 2005, 156-161.

[10] C. L. P. Sehier, Automatic modulation recognition with a hierarchical neural network, *Proc. MILCOM,* 1993, pp. 111–115.

[11] A.K. Nandi, E.E. Azzouz, Algorithms for automatic modulation recognition of communication signals, *IEEE Trans. Commun.,* Vol. 46, No. 4, 1998, pp. 431–436.

[12] L. Mingquan, X. Xianci, L. Leming, Cyclic spectral features based modulation recognition, *Proc. Comm. Tech.,* Vol.2, 1998, pp. 792–795.

[13] A. Ebrahimzadeh, and S. A. Seyedin, Identification of signal type using svm in fading environment, *Proc. ICEE*, 2006.

[14] A. Ebrahimzadeh, and S. A. Seyedin, An Intelligent Technique for Identification of Digital Modulations, *accepted to publish in WSEAS Trans Comp.,* 2006.

[15] C. Cortes, and V. Vapnic, Support Vector Network, *Machine Learning*, Vol. 20, 1995, pp.1-25.

[16] C. Burges A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, Vol. 2, 1998, pp 121-167.

[17] B. SchOlkopf, C. Burges, and V. Vapnik, Extracting support data for a given task, *ICKDDM,*1995, pp. 252-257.

[18] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, Choosing multiple parameters for support vector machines, *Machine Learning*, 2002, Vol. 46, No.1, pp. 131-159.

[19] Y. X. Su, B. Y. Duan, C. H. Zheng, Genetic design of kinematically optimal fine tuning Stewart platform for large spherical radio telescope, *Mechatronics*, 2001, Vol.11, No.7, pp. 821-835.