Digital Signal Types Identification Using a Hierarchical SVM-Based Classifier and Efficient Features

Ataollah Ebrahimzadeh
Dep. ECE, Noshirvani Institute of Technology
ataebrahim@yahoo.com

Seyed Alireza. Seyedin
Dep. EE, University of Mashad
seyedin@fum.ac.ir

Abstract

Automatic digital signal type identification (ADSTI) is an important topic for both military and civilian communication applications. Most of proposed techniques (identifiers) can only recognize a few types of digital signal and usually need high levels of SNR. This paper presents a technique that includes a variety of digital signal types. In this technique a hierarchical support vector machine based structure is proposed for multi-class classification. Combination of higher order moments and higher order cumulants up to eighth are utilized as the effective features. Genetic algorithm is used to parameter selection in order to improve the performance of identifier. Simulation results show that proposed identifier has high performance even at low SNR values.

1. Introduction

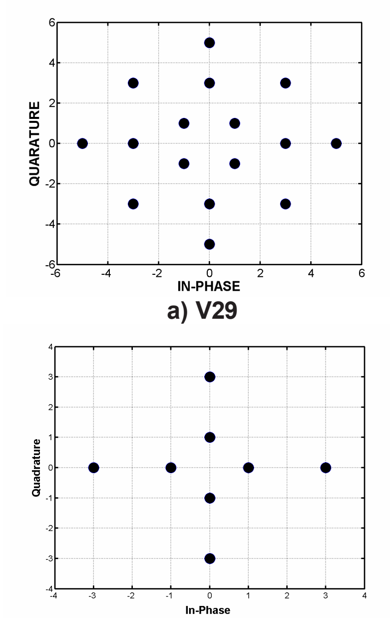
ADSTI plays an important role in military applications, such as electronic surveillance, and in civil applications, such as spectrum management, software radios, etc. With increasing usage of digital signals in many novel applications the need to find efficient techniques for their discrimination has become important. This paper focused on identifying of these signals. ADSTI techniques usually can be categorized in two main principles: the decision theoretic (DT) and the pattern recognition (PR). DT techniques use probabilistic and hypothesis testing arguments to formulate the recognition problem [1,2]. The major drawbacks of DT techniques are their too high computational complexity, lack of robustness to the model mismatch as well as careful analysis that are required to set the correct threshold values [3]. However, PR techniques don't need such careful treatment. They are easy to implement. In this paper we have proposed a PR based identifier. PR techniques can be further divided into two main subsystems: the feature extraction and the classifier. The former extracts the features and the latter determines the membership of signal [3-9].

Most of PR techniques can only recognize a few types of digital signal and usually need high SNR values. Those identifiers that use artificial neural networks (ANNs) as the classifier have higher performance than others [5-8]. However, with regard to effectiveness of ANNs, there are some problems. For example ANNs have limitations on generalization ability in low SNRs [9]. In the recent years support vector machines (SVMs), have used in area of pattern recognition because of excellent generalization capability [10,11]. Also it shows high classification accuracy in signal type identification [9]. Therefore, we have used this property of SVMs and proposed a hierarchical classifier that is based on SVMs. Suitable parameters selection of SVMs can improve the performance. We have used genetic algorithm (GA) in order to model selection. As the features, which have a vital role in identification, we have utilized the combination of higher order moments and cumulants.

The paper is organized as follows. Section 2 presents the feature extraction as well as the digital signal types (DST) that are considered in this paper. Section 3 describes the hierarchical classifier. Section 4 introduces the genetic algorithm, which is used for parameters selection. Section 5 shows some simulation results. Finally, Section 6 concludes the paper.

2. DST and Feature Extraction

Different types of digital signal have different characteristics. Thus finding the proper features for identification of them, particularly in case of higher order and/or non-square types, is a serious problem. In this paper the DST are: ASK4, ASK8, PSK2, PSK4, PSK8, Star-QAM8, V29, QAM128 and QAM64, that because of simplifying the indication, they are substituted with P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, and P_9, respectively. Figure 1 shows the constellation of two digital signal types, Star-QAM8 and V29.



a) V29

b) Star-QAM8

Figure 1. Constellations of: a) Star-QAM8, b) V29

Among the different features that we have computed and experimented, combination of the higher order moments and higher order cumulants (up to eighth) achieve the highest performances for DST.

Probability distribution moments are a generalization of concept of the expected value. Recall that the general expression for the i^{th} moment of a random variable is given by [12]:

$$\mu_i = \int_{-\infty}^{\infty} (s-\mu)^i f(s)ds \tag{1}$$

where μ is the mean of the random variable. In this study signals are assumed to be zero mean. Thus, the definition for the i^{th} moment for a finite length (N) discrete signal is given by:

$$\mu_i = \sum_{k=1}^{N} {s_k}^i f(s_k) \tag{2}$$

Next, the auto-moment of the random variable is: may be defined as follows:

$$M_{pq} = E[s^{p-q}(s^*)^q] \tag{3}$$

where p is called the moment order and s^* stands for complex conjugation of s.

Assume a zero-mean discrete based-band signal sequence of the form $s_k = a_k + jb_k$. Using the definition of the auto-moments, the expressions for different orders may be easily derived.

Consider a scalar zero mean random variable s with characteristic function:

$$\hat{f}(t) = E\{e^{jts}\} \tag{4}$$

Expanding the logarithm of the characteristic function as a Taylor series, one obtains:

$$\log \hat{f}(t) = k_1(jt) + \ldots + \frac{k_r(jt)^r}{r!} + \cdots \tag{5}$$

The constants k_r in (5) are called the cumulants (of the distribution) of s. The symbolism for p^{th} order of cumulant is similar to that of the p^{th} order moment. More specially:

$$C_{pq} = Cum[\underbrace{s,\ldots,s}_{(p-q)terms}, \underbrace{s^*,\ldots,s^*}_{(q)terms}] \tag{6}$$

We have computed, all of the features for DST. Table1 shows some of these features for a number of considered digital signal types. These values are computed under the constraint of unit variance in noise free.

Table 1. Some of the features for a number of digital signal types

	P_1	P_3	P_4	P_6	P_9
M_{41}	1.64	1	0	0	0
M_{61}	2.92	1	-1	2.92	-1.3
C_{63}	8.32	16	4	.160	1.79
M_{84}	5.24	1	1	5.25	3.96
C_{80}	-30.1	-244	34	-88.9	-11.5
C_{82}	-30.1	-244	-46	63.31	-27.1

3. Classifier

We have proposed a multi-class SVM based classifier that has a hierarchical structure. SVMs were introduced on the foundation of statistical learning theory. The basic SVM deals with two-class problems; however, with some methods it can develop for multi-class classification [10]. Binary-SVM performs classification tasks by constructing the optimal separating hyper-plane (OSH). OSH maximizes the margin between the two nearest data points belonging to the two separate classes.

Suppose the training set, $(x_i, y_i), i = 1,2,\ldots,l$ $x \in R^d, y \in \{-1,+1\}$ can be separated by the $w^T x + b = 0$, where w is the weight vector and b is the bias. If this hyper-plane maximizes the margin, then:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \text{for all } \mathbf{x}_i \quad i = 1,2,\ldots,l \tag{7}$$

Those training points, for which the equality in (6) holds, are called support vectors (SV). By using Lagrange multipliers $(\alpha_i, i = 1,\ldots,l; \alpha_i \geq 0)$, after some computations, the optimal decision function (ODF) is then given [10]:

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{l} y_i \alpha_i^* \mathbf{x}^T \mathbf{x}_i + b^*) \tag{8}$$

where α_i^* 's are optimal Lagrange multipliers.

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

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad for \quad i = 1,2,...,l \tag{9}$$

OSH is achieved by minimizing the $\Phi = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i^k$ subject to (9), where C is the penalty parameter [11].

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. On of the kernel functions is Gaussian radial basis function (GRBF) given by:

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \tag{10}$$

where σ is the width of the RBF kernel.

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

$$f(\mathbf{x}) = \mathrm{sgn}(\sum_{i=1}^{l} y_i \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*) \tag{11}$$

The performance of SVM depends on penalty parameter (*C)* and the kernel parameter, which are called hyper-parameters. In this paper we have used the GRBF, because it shows better performance than other kernels. Thus hyper-parameters are: C and σ.

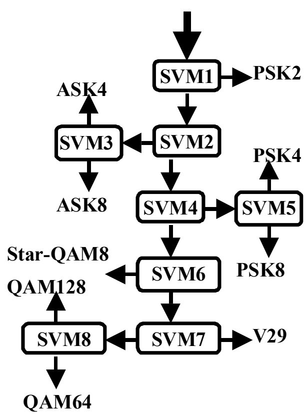


Figure 2. Hierarchical SVM-Based Classifier

There are two widely used methods to extend binary SVMs to multi-class problems: one-against-all (OAA) method and one-against-one (OAO) method [13]. In this paper we have proposed a hierarchical SVM-based classifier. Figure 2 shows the scheme of this classifier. One of the advantages this structure is that the number of SVMs is less than in cases of OAO and OAA.

4. Parameters Selection Using GA

Selection of the optimal values of hyper-parameters can improve the performance of SVMs. In this paper we have used GA for parameters selection of SVMs. To apply Gas, one has to specify it basic issues.

Real-encoded scheme is selected as the representation of the parameters. The research space of these parameters is $C \in [1:5:51]$, $\sigma \in [0.1:2]$. Because the real-coded scheme is used, the solution space coincides with the chromosome space. The size of population pop_size is choosing 16 in ordering to avoiding the convergence of the population becomes difficult. 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 is proposed as the fitness function [14]. 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:

$$P(a_i) = \frac{fit(a_i)}{\sum_{j=1}^{pop\_size} fit(a_j)} \times pop\_size \tag{12}$$

The means of crossover implement is closely integrated to the encoding scheme. The crossover operator in this paper can be defined as [15]. The adaptive mutation probability is adopted in this paper to solve the above two problems as follows:

$$P_m = \frac{\exp(-b \times t/2)}{pop\_size \times \sqrt{L}} \tag{13}$$

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. Terminate the program when the best fitness has not changed more than a very small value, i.e. 10^{-6} over the last generations.

5. Simulation results

The All of the considered digital signal types are simulated in MATLAB environment. The simulated signals were also band-limited and Gaussian noise was added according to SNR values –3, 0, 3, 6, 9, and 18 dB. Each digital signal type has 1200 realizations. Among the features that have we have mentioned in Section 2, Table2 shows the chosen features that achieve the best results for identification of DST.

Table 2. Chosen features for each SVM

Number of SVM	Chosen features
SVM 1	C_{83}
SVM 2	M_{41}
SVM 3	M_{41}, C_{81}
SVM 4	M_{63}, C_{63}
SVM 5	C_{61}
SVM 6	C_{63}
SVM 7	M_{82}, C_{80}
SVM 8	M_{61}, C_{80}

5.1. Performance without optimization

Based on some experiments, the values $\sigma=1$ and C=10 are selected for all SVMs. Table 3 shows the identification results (performances) on DST in different SNR values. These are the average of the values that appearance in the diagonal matrix. It can be seen that performance is generally very good even with low SNRs. This is due the two facts: chosen novel features and novel classifier. The chosen features have effective properties in signal representation. On the other hand, the SVM based classifier has high generalization ability such that enables it to classify the non-separable data (low SNR) with high accuracy.

Table 3. Performances of identifier without optimization

SNR	Training	Testing
-3	86.12	85.65
0	91.74	91.25
3	94.54	94.12
6	96.12	95.87
9	97.56	97.14
18	98.36	98.24

In order to compare the performance of hierarchical SVM-based classifier with another classifier, we have considered a hierarchical MLP-based classifier that SVMs are replaced with MLP neural networks. The simulation setups are the same. We name this technique as TECH2. Figure 3 shows the performances of two identifiers in different SNR values that term P_C means the percentage of correct classification. It can be seen that our proposed technique (PROTECH) that uses SVM in the structure of classifier has a better performances than of TECH2, particularly for low SNR values. When the SNR is low, TECH2 shows poor performance while in higher SNRs the percentage of correct classification is high. The construction of neural network in low SNRs is not proper, which results in low generalization ability. In higher SNRs the features are proper and closer to the noiseless state and it is easier to construct the neural network and results in high identification percentage.

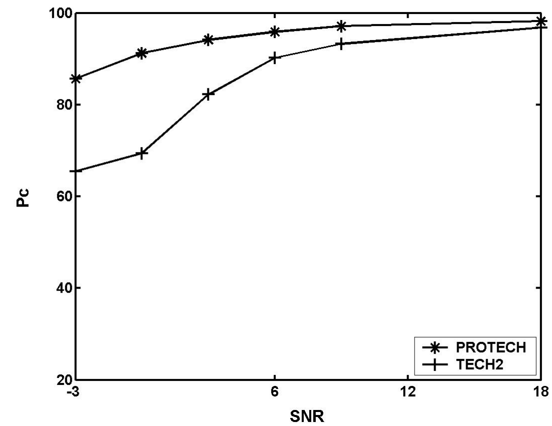


Figure3. Comparison between the performances of PROTECH and TECH2

5.2. Performance with applying GA

We apply GA for parameters selection of SVMs. Table 4 shows the performances of the optimized identifier for various SNRs. It can be seen that the optimization improves the performances of identifier for all SNRs; especially in lower SNRs.

Table 4. Performances of identifier with GA

SNR (dB)	Training	Testing
-3	91.75	91.30
0	94.56	93.54
3	96.90	96.12
6	98.43	98.27
9	98.84	98.45
18	99.35	99.13

5.3. Comparison with some works

As mentioned in [4], direct comparison with other works is difficult in signal type identification. In [5], the authors reported a success rate 90% and 93% with SNR of 15–25 dB. However, the performance for lower SNRs is less than 80%. In [7], the authors show an average accuracy around 90% with SNR ranges between 5 and 25 dB. In [8], the authors developed a neural network classifier for identification. The results show 88% accuracy at 15 dB SNR.

This paper proposed an efficient identifier that includes a variety of digital signal types. In this identifier a hierarchical SVM based classifier is proposed. Utilizing SVMs causes that identifier would have great generalization ability. Higher order moments and higher order cumulants (up to eighth) have proposed as the features. These features have effective properties in signals representation. This identifier shows high accuracy even at very low levels of SNR. Utilizing GAs in order to optimization of the identifier, improves the performance of identifier.

6. Conclusions

ADSTI has seen increasing demand in different applications. This paper presents a high efficient technique for identification of digital signal types. In this technique a hierarchical multi-class classifier based on SVMs is proposed. The inputs of this classifier are higher order moments and higher order cumulants. Each SVM uses the features 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 avoids the over-fitting and local minimum. Chosen features of the higher order moments and the higher order cumulants have high ability to signal representation. Optimization using GA, improves the performance of system especially in lower SNRs. The proposed identifier includes different kinds of digital signal and can separate them with high accuracy even at low SNR values. For future work it can be developed to include a lot of signals with considering additional features such as introduced in [8].

Acknowledgements

The authors would like to thank Iran Telecommunication Research Center (ITRC) for funding this work.

References

[1] W. Wei, and J. M. Mendel, Maximum-likelihood classification for digital amplitude-phase modulations, *IEEE Trans. Comm.,* 2000, Vol. 48, pp. 189-193.
[2] P. Panagotiou, and A. Polydoros, Likelihood ratio tests for modulation classifications, *Proc. MILCOM,* 2000*,* pp. 670-674.
[3] A. Swami, and B. M. Sadler, Hierarchical digital modulation classification using cumulants, *IEEE Trans. Comm.,* Vol. 48, 2000, pp. 416–429.
[4] B. G. Mobasseri, Digital modulation classification using constellation shape, *Signal Processing*, vol. 80, 2000, pp. 251–277.
[5] C. L. P. Sehier, Automatic modulation recognition with a hierarchical neural network, *Proc. MILCOM,* 1993, pp. 111–115.
[6] A. Ebrahimzadeh, and S. A. Seyedin, Automatic psk identification using wpa and a modified mlp, *Proc. CIS,* 2005, pp. 250-255.
[7] L. Mingquan, X. Xianci, L. Leming, Cyclic spectral features based modulation recognition, *Proc. ICCT*, 1996, pp. 792–795.
[8] A.K. Nandi, E.E. Azzouz, Algorithms for automatic modulation recognition of communication signals, *IEEE Trans. Comm.*, vol. 46, 1998, pp. 431–436.
[9] A. Ebrahimzadeh, and S. A. Seyedin, Identification of signal type using svm in fading environments, *Proc. ICEE*, 2006.
[10] C. Cortes, and V. Vapnic, Support Vector Network, *Machine Learning*, vol. 20, 1995, pp.1-25.
[11] C. Burges A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, vol. 2, 1998, pp 121-167.
[12] P. McCullagh, *Tensor Methods in statistics,* Chapman & Hall, 1987.
[13] B. Scholkopf, C. Burges, and V. Vapnik, Extracting support data for a given task, *ICKDDM,* 1995, pp. 252-257.
[14] 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.
[15] 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.