Digital Modulation Recognition Using Support Vector Machine Classifier

Hussam Mustafa and Miloš Doroslovački
Department of Electrical and Computer Engineering
The George Washington University
Washington, DC 20052

Abstract-**We propose four features to classify amplitude shift keying with two levels and four levels, binary phase shift keying, quadrature phase keying, frequency shift keying with two carriers and four carriers. After that we present a new method of classification based on support vector machine (SVM) that uses the four proposed features. We study the performance of SVM classifier and compare it to the previous work done in the literature on the digital modulation classification problem.**

I. Introduction

Recognition of modulation in received signals is important for many applications, such as signal interception, interference identification, electronic warfare, enforcement of civilian spectrum compliance, radar and intelligent modems. The modulation recognition methods can be divided into two categories. The first is modulation recognition with prior information available. The information provides knowledge of signal parameters such as amplitude, carrier frequency, symbol rate, pulse shape, initial phase, channel characteristic and noise power. The second, and more challenging, is modulation recognition without any prior information about signal parameters.

In the past years there have been different approaches to solve the modulation recognition problem. These approaches can be classified in three groups. The first group includes approaches that use memoryless nonlinearities and detect the spectrum lines occurring for specific modulation types [1]. The second group includes the feature based approaches, where the recognition is divided into two stages. The first stage maps the signal into a smaller feature domain; usually the feature domain is independent of the signal's parameters. The second stage does the classification of the signal by comparing the measured values of features to a priori collocation of the feature values for each modulation type [2], [3] and [4]. And the third are the decision theoretic approaches, where in [5,6] all the signal parameters are assumed known to the receiver. However in [7] the classifier does not need to know the initial phase. These approaches use the likelihood function to do recognition. They are optimal in the sense of the minimum probability of misclassification.

In this paper we present the signal model we assume (Section 2), as well as the new proposed features (Section 3) for modulation recognition. Further, we describe briefly the support vector machine (SVM) algorithm (Section 4) and discuss the construction of SVM classifier (Section 5). Finally, we present and comment on simulation results (Section 6).

II. Signal Model

First we consider the following complex baseband signal

$$r(k) = x(k) + n(k) \tag{1}$$

where $x(k)$ is the transmitted signal

$$x(k) = \sum_{n=0}^{N-1} a_n e^{j(\theta_n + \theta_c)} p(k - nT) \tag{2}$$

and (a_n, θ_n) are the magnitude and phase of a modulation constellation point. θ_c is the initial phase. $p(k - nT)$ is the pulse shape function and T is the symbol rate. $n(k)$ is assumed to be complex white Gaussian noise with power σ^2.

III. Classification Features

The features used in this paper are based on two main processing steps. The first step is the multiplication of two consecutive signal values. The second step is the statistical characterization of the quantity obtained in the first step. Based on these steps we choose the following features to distinguish between modulations:

1) $\frac{1}{M}\sum_{k=0}^{M-1} \mathrm{Im}(s(k)s^*(k-1)) \rightarrow$ ASK and PSK/FSK where M is the number of samples in the realization and * represents the conjugate operator.

2) $\text{Kurtosis}(\text{Re}(s(k)s^*(k-1))) \rightarrow$ ASK2/ASK4/PSK2

3) $\begin{bmatrix} \text{Kurtosis}(\text{Re}(s(k)s(k-1))) \\ \text{Kurtosis}(\text{Im}(s(k)s(k-1))) \end{bmatrix} \rightarrow$ PSK2/PSK4

4) $\dfrac{\text{Second maximum}(\, FFT(s(k))}{\text{Third maximum}(\, FFT(s(k))} \rightarrow$ FSK2/FSK4

Based on these features we constructed the classification tree shown in Fig.1.

IV. Support Vector Machine (SVM)

SVM is an empirical modeling algorithm that can be applied in classification problems. The first objective of the Support Vector Classification (SVC) is the maximization of the margin between the two nearest data points belonging to two separate classes. The second objective is to constrain that all data points belong to the right class. It is a two-class solution which can use multi-dimensions features. The two objectives of the SVC problem are then incorporated into an optimization problem. This is done by constructing the dual and primal problem of the classical Lagrangian problem with transferring the constraint of the second objective to become constraints on the Lagrange variables. The complete derivation of SVC is given in [8-9].

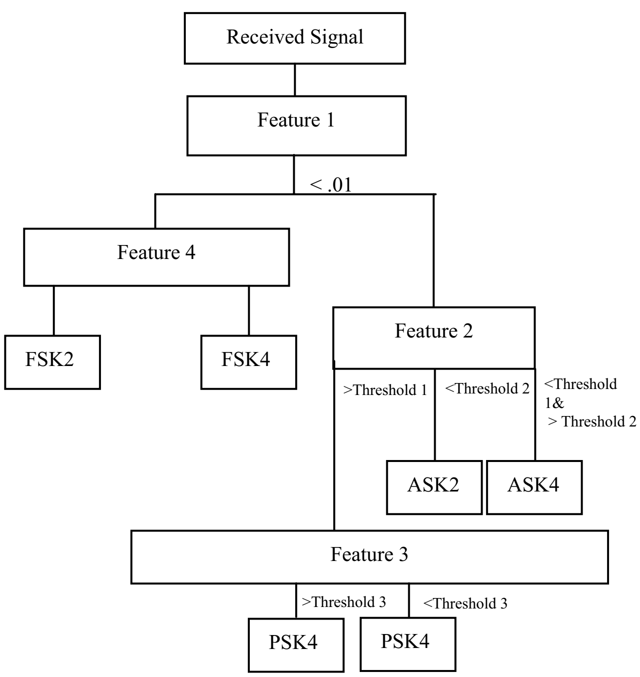

Fig. 1. Proposed recognition tree

SVC can be applied to separable and non-separable data points. In the non-separable case the algorithm adds one more design parameter. The parameter is the weight of the error caused by the points present in the wrong class region. In our application we face this issue in the low SNR cases. On the other hand, in the high SNR cases, the algorithm takes its simplest separable case version.

Another degree of freedom in the SVC is the kernel function used. In our application, since we are dealing with one and two dimensional features, we used linear and polynomial-of-power-2 kernels. Finally the number of data points used in the training procedure is also another parameter that needs to be determined before constructing the SVM classifier.

V. Classification Using Svm

Fig.2 presents the probability of correct classification of 2000 binary phase shift keying (PSK2) and quadrature phase shift keying (PSK4) signals using different numbers of training points. In this figure we present two cases. The first case is the separable data case (SNR=5dB) where all the data points are separated completely and there are no misclassified data points. In this case we see that as we increase the number of training points, the probability of correct classification converge towards 1. It should be noted that since we are dealing with a two-dimensional feature, the minimum number of training points needed to determine the SVM classifier is 3 [8-9]. The second case is the nonseparable case (SNR=0dB) where some of the data points are not separated from the data points corresponding to the other class. In this case again, as we increase the number of training points, we achieve better probability of correct classification. However, as we continue increasing the number of training points, we do not converge towards a specific value; instead we oscillate

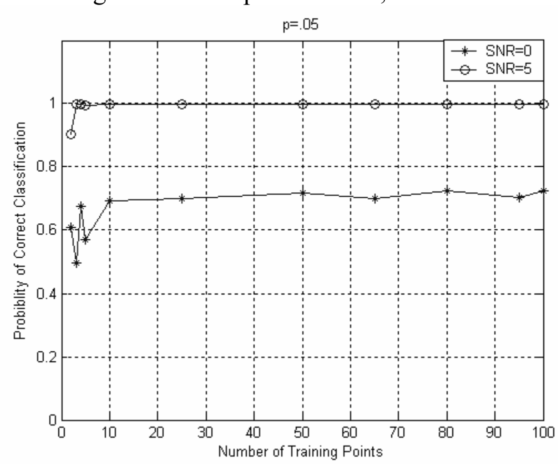

Fig. 2. The probability of correct classification of 2000 PSK2 and PSK4 signals using different numbers of training points. p=.05 and $\theta_c \in [0, 2\pi]$.

around it. This is due to the fact that as we increase the number of training points, we also increase the number of misclassified data points which affect the determination of the discriminating curve. From the simulation results we choose for the SVM classifier 25 training points as good candidate. At 25 training points we achieve the convergence in the separable case and an acceptable performance in the nonseparable case.

Due to the simplicity of the data structure we have, the value of C did not affect the classifier structure. From the simulation results we did not find much of a difference when we changed the value of C from [1,100]. In our simulation we choose for C the value of 1.

To determine the kernel used to construct the SVM classifier in the simulation, we tested two kernels, the linear and second-order polynomial. In the linear kernel case the SVM classifier is a straight line separating the two classes. In the second-order polynomial case the SVM classifier is a parabolic curve separating the two classes.

VI. SIMULATION AND DISSCUSION

In this section we compare the proposed SVM classifier to three previously discussed classifiers: the maximum likelihood classifier proposed by [6-7]; the qLLR classifier proposed by [5] and finally the cumulant-based classifier proposed by [2]. Also we compare SVM classifier to another two proposed classifiers based on the classification tree in Fig.1: fixed threshold classifier and dynamic threshold classifier. The dynamic threshold is determined by the value of SNR. In order to compare fairly these classifiers with the SVM classifier, we need to determine the amount of information needed from the receiver point of view in order for the classifier to operate. In the case of the maximum likelihood classifier, the receiver needs to know all the signal parameters and the noise power. In the case of signal parameters, this includes the value of the constellation points and the random initial phase (θ_c). In the case of qLLR classifier, cumulant-based classifier, dynamic threshold and SVM, the noise power and all signal parameters (except the value of the constellation points and the random initial phase) must be known to the receiver. Finally in the case of fixed threshold classifier, the same scenario as for SVM classifier applies here, except that the receiver does not need to know the noise power. We now present the simulation examples in which we compare all six classifiers.

Fig.3-5 present the probability of misclassification as a function of SNR of 2000 PSK2 and PSK4 signals at p={.05,.1,.2} where p is the ratio of symbol rate to sampling rate. For each signal we choose sampling frequency of 500 samples/second; time duration of 4 seconds; random $\theta_c \in [0, 2\pi]$; and each constellation point has equal probability of occurrences. Tables 1-6 present the confusion matrix of the SVM classifier of the same simulation example presented in Fig.3-5. The results in the table are limited to SNR 0 dB and 5dB. Fig. 6 and 7 present the probability of misclassification of 3000 two-level amplitude shift keying (ASK2), four-level amplitude shift keying (ASK4), PSK2, PSK4, two-carrier frequency shift keying (FSK2) and four-carrier frequency shift keying (FSK4) signals for p={.05,.1}. Each signal has a sampling frequency of 10,000 and time duration of 4 seconds. In the case of FSK2 the two carrier frequencies are {2000,3000} samples/seconds. The center frequency for FSK4 is 2500 with frequency separation of 500 samples/seconds.

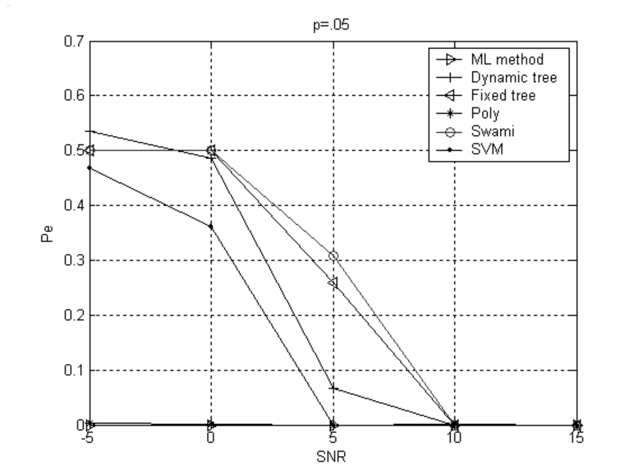

Fig. 3. Probability of classification error (P_e) for 2000 PSK2 and PSK4 signals for different SNRs. 'ML' represents the maximum likelihood classifier, 'Dynamic tree' is the proposed dynamic threshold classifier, 'Fixed tree' is the proposed fixed threshold classifier, 'Poly' is the qLLR classifier, 'Swami' is the cumulant-based classifier and 'SVM' is the proposed SVM classifier. p=.05.

Classification Output \ Actual Modulation	PSK2	PSK4
PSK2	573	197
PSK4	427	803

Table 1: Confusion Matrix of SVM algorithm for SNR=0dB and p=.05.

Classification Output \ Actual Modulation	PSK2	PSK4
PSK2	1000	0
PSK4	0	1000

Table 2: Confusion Matrix of SVM algorithm for SNR=5dB and p=.5.

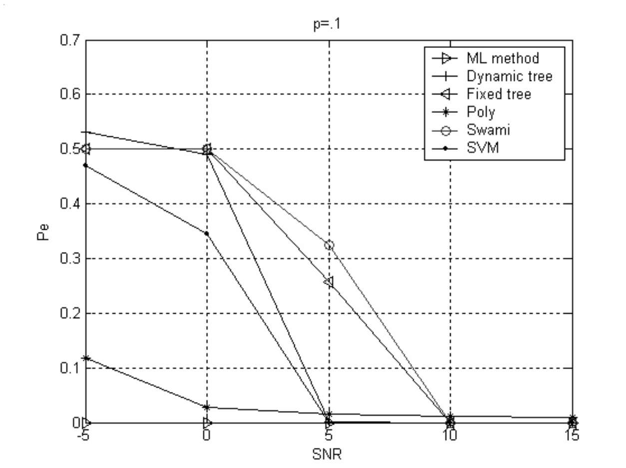

Fig. 4. Probability of classification error (P_e) for 2000 PSK2 and PSK4 signals for different SNRs. Acronyms are the same as for Figure 2. *p*= .1.

Classification Output \ Actual Modulation	PSK2	PSK4
PSK2	616	201
PSK4	384	799

Table 3: Confusion Matrix of SVM algorithm for SNR=0dB and *p*=.1.

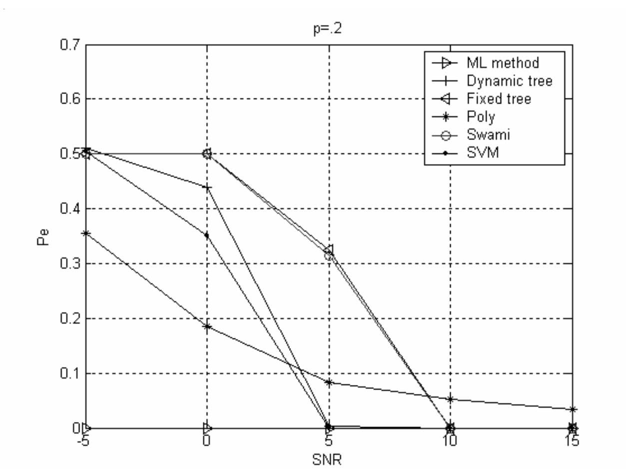

Fig. 5. Probability of classification error (P_e) for 2000 PSK2 and PSK4 signals for different SNRs. Acronyms are the same as for Figure 2. *p*= .2.

Classification Output \ Actual Modulation	PSK2	PSK4
PSK2	589	250
PSK4	411	750

Table 4: Confusion Matrix of SVM algorithm for SNR=0dB and *p*=.2.

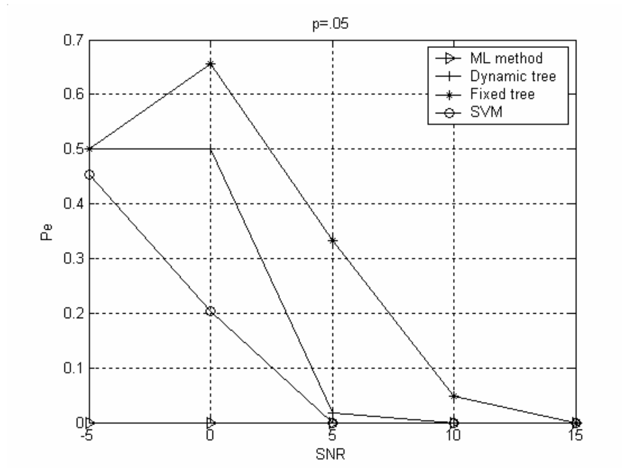

Fig. 6. Probability of classification error (P_e) for 3000 ASK2, ASK4, PSK2, PSK4, FSK2 and FSK4 signals for different SNRs. Acronyms are the same as for Figure 2. *p*= .05.

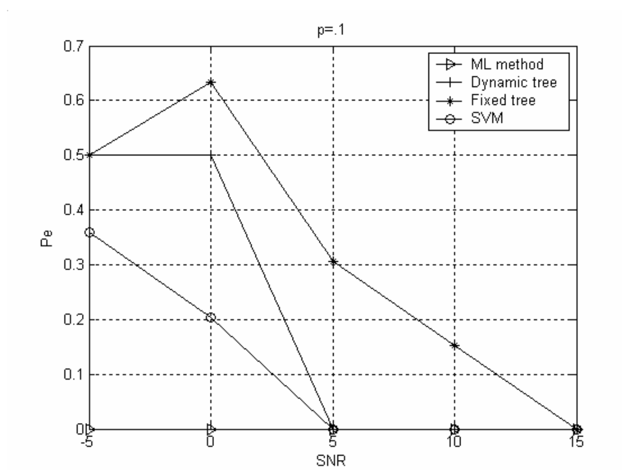

Fig. 7. Probability of classification error (P_e) for 3000 ASK2, ASK4, PSK2, PSK4, FSK2 and FSK4 signals for different SNRs. Acronyms are the same as for Figure 2. *p*= .1.

Classification Output \ Actual Modulation	ASK2	ASK4	PSK2	PSK4	FSK2	FSK4
ASK2	0	0	0	0	0	0
ASK4	10	490	14	4	0	0
PSK2	490	2	476	5	0	0
PSK4	0	0	10	491	0	0
FSK2	0	0	0	0	500	0
FSK4	0	0	0	0	0	500

Table 5: Confusion matrix of SVM algorithm for SNR 0dB.*p*=.05.

Actual Modulation / Classification Output	ASK2	ASK4	PSK2	PSK4	FSK2	FSK4
ASK2	500	0	0	0	0	0
ASK4	0	500	0	0	0	0
PSK2	0	0	500	0	0	0
PSK4	0	0	0	500	0	0
FSK2	0	0	0	0	500	0
FSK4	0	0	0	0	0	500

Table 6: Confusion matrix of SVM algorithm for SNR 5dB.*p*=.05.

From Fig. 3-7 it is clear that the maximum likelihood classifier has the best performance among the compared classifiers. Following the maximum likelihood classifier, the qLLR comes second for small values of p. However as p increases, the performance of qLLR classifier deteriorates. The reason for that is as we increase the value of p, we decrease the number of samples in the averaging process used in the qLLR classifier. This affects the approximation used in the algorithm.

The simulation results also show that dynamic threshold classifier outperforms the fixed threshold classifier. This is due to the curvature of the kurtosis curves at SNR<10dB. In the case of cumulant-based classifier, from the figures it is clear that the performance of the algorithm is independent of p.

Finally the SVM classifier shows robust performance over all simulations (whether distinguishing PSK2 from PSK4 or applied to the classification tree proposed in Fig.1) for different values of p. In the SNR=0dB area the SVM classifier outperforms the dynamic tree classifier and cumulant-based classifier. This is due to the fact that the SVM classifier is modified such that it can be used on nonseparable data.

REFERENCES

[1] J. Reichert. "Automatic classification of communication signals using higher order statistics", *ICASSP*, vol.5, New York, NY, USA. 1992 pp.221-4.

[2] A. Swami, B.M. Sadler. "Hierarchical digital modulation classification using cumulants", *IEEE Transactions on Communications,* vol.48, no.3, pp.416-29, March 2000.

[3] K.C. Ho, W. Prokopiw, YT. Chan. "Modulation identification of digital signals by the wavelet transform", *IEE Proceedings: Radar, Sonar & Navigation*, vol.147, no.4, pp.169-76, Aug. 2000.

[4] H. Deng, M. Doroslovacki, H. Mustafa, X. Jinghao, K. Sunggy. "Automatic digital modulation classification using instantaneous features", *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings IEEE,* vol.4, 2002, pp.IV4168.

[5] A. Polydoros, K. Kim. "On the detection and classification of quadrature digital modulations in broad-band noise", *IEEE Transactions on Communications*, vol.38, no.8, pp.1199-211 Aug. 1990.

[6] W. Wei, J. M. Mendel. "Maximum-likelihood classification for digital amplitude-phase modulations", *IEEE Transactions on Communications*, vol.48, no.2, pp.189-93, Feb. 2000.

[7] W. Wei "Classification of Digital Modulation Using Constellation Analyzes." Ph.D. dissertation. Univ. of Southern California. 1998.

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

[9] N. Cristianini and J. Shawe-Taylor *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods,* Cambridge University Press, 2000.