Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics, Istanbul, Turkey, May 27-29, 2006 (pp409-414)

Automatic Digital Modulation Identification in Dispersive Channels

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

Abstract: - Automatic modulation type identification (AMTI) has seen increasing demand for both military and civilian, nowadays. Most of previous methods have been proposed on classification of modulations in additive white Gaussian noise (AWGN) channels. However in real world scenarios, communication channels suffer from dispersion (fading). This paper proposes a novel automatic digital modulation types identifier (ADMTI) in dispersive environment. In the ADMTI's structure, undesired effects of channel are mitigated by an equalizer. Higher order cumulants and moments (up to eighth) are used as features and classification is performed by a multiclass SVM-based classifier. Simulation results show that ADMTI is able to identify different types of modulations (e.g. QAM64, V.29, and ASK8) with high accuracy even at low SNRs.

Key-Words: - Statistical pattern recognition, modulation, support vector machine, dispersive environment.

1 Introduction

Automatic modulation type identifier is a system that recognizes the modulation type of received signal automatically, and has many applications such as electronic surveillance, threat evaluation, signal confirmation, spectrum management, software radio, etc. Whilst, early researches were concentrated on analog modulations, the recent contributions in the subject focus more on digital communications due to increasing usage of digital modulations in many novel applications.

Generally, AMTI methods can be categorized in two main categories: decision theoretic (DT) and pattern recognition (PR). DT approaches use probabilistic and hypothesis testing arguments to formulate the recognition problem and to obtain the classification rule [1-11]. The major drawbacks of these approaches are their very high computational complexity, difficulty within the implementation and lack of robustness to model mismatch. PR approaches, however, do not need such careful treatment. PR approaches mainly divided into two subsystems: the feature extraction subsystem and the recognition 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 [12-26]. PR approaches are simple to implement; however, selections of two subsystems are serious problems.

Most of previous methods have been proposed on classification of modulations in AWGNchannels [1-8], [12-24]. However, in real world communication channels, such as wireless communication environments, suffer from dispersion (fading) and most of recognizers that are designed for AWGN do not preserve their performance under impairment conditions. Research on AMTI, over fading channels has been only performed in a few works [9-11], [25-26]. In [9-10], classification between PSK2 and PSK4 in a flat Rayleigh fading are proposed. In [11] a quasi-optimal solution based on the approximation of the log-likelihood function is proposed. In [25], the modulations were identified by applying the nearest neighbor rule in a two-dimensional feature space. In [26], an identifier using neural network based on combinations of different order of moments is proposed to discriminate digital modulations in multipath fading.

This paper proposes a novel ADMTI in time dispersion channels.Figure1 shows the scheme of ADMTI. In this structure, Pre-processing module performs: rejection of noise outside of signal bandwidth, normalization, carrier frequency estimation, recovery of complex envelope, etc. Equalization module mitigates the channel that is presented in section2. Section 3 describes, feature extraction module. Section 4, presents the classifier. In section 5, some experimental results are shown for considered digital modulation set {PSK2, PSK4, PSK8, ASK8, QAM32, V29, Star-QAM8, and QAM64}. Finally in section 6 conclusions are presented.

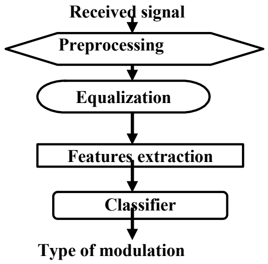

Figure1: Structure of ADMTI

2 Channel equalization

In digital communications, according to the changes in the message frequency, message amplitude, message phase, or changes in amplitude and phase, we have four main digital modulation techniques, frequency shift keying (FSK), amplitude shift keying (ASK), phase shift keying (PSK) and quadrature amplitude modulation (QAM), respectively. Most of them are applied in M-ary form [27]. In real world situations the transmission channel is a critical factor that may cause unrecoverable distortions on the signal, especially in higher order digital modulations, where the effect of channel may corrupt the signal constellation. In order to mitigate the dispersion effects of the propagation channel, an equalization stage is employed in the receiver. In AMTI applications, the training sequence that is needed for adjusting equalizer coefficients is not available. Hence, the equalization must be done blindly. When the type of modulation is unknown, usually, the Fractionally Spaced Equalizer- Constant Modulus Algorithm (FSE-CMA) is one of the commonly used blind equalization algorithms, which are designed to undo the channel effect without any knowledge of the channel itself [28]. The FSE-CMA is the integration of two different parts: the constant modulus algorithms (CMA) and the fractional spaced equalizer (FSE).

The constant modulus algorithm (CMA) is a stochastic gradient algorithm, designed to force the equalizer weight to keep a constant envelop on the received signals. Thus, it is designed for problems where the signal of interest has a constant envelope property. However, extensive simulations have shown that it can still be used in amplitude-phase modulation types, but the success of equalization is decreased with increasing of order. As a result, the CMA is expected to have better performance for FSK and PSK rather than QAM types. The CMA cost function is given by:

$$J(k) = E\{(|y(k)|^2 - \gamma)^2\} \tag{1}$$

where $y(k)$ is the equalizer output and γ called the dispersion constant defined by (2).

$$\gamma = \frac{M_{4,0}}{M_{2,0}} \tag{2}$$

where $M_{4,0}$ and $M_{2,0}$ are fourth and second order moment respectively. The cost function $J(k)$ is minimized iteratively using a gradient based algorithm with update equation.

In any standard CMA equalization system, the coefficient taps are baud-spaced. However, it is often desired to use an equalizer with taps spaced at a fraction of the data symbol period T. This configuration gives the extra degrees of freedom to perform additional filtering. Such a scheme is called fractional spaced equalization (FSE).

We assume that the received signal is:

$$x(t) = \sum_{k=-\infty}^{k=\infty} s(k)h(t - kT) + \upsilon(t) \tag{3}$$

where $h(t)$ is the channel impulse response, $s(k)$ the sequence of information and $\upsilon(t)$ is AWGN. The response $h(t)$ is assumed to be of finite length. Fractionally spaced channel output resulting from P times oversampling with respect to symbol rate may be written as:

$$x(k\frac{T}{P}) = \sum_{l=-\infty}^{\infty} s(l)h(k\frac{T}{P} - lT) + \nu(k\frac{Tt}{P}) \tag{4}$$

An equivalent representation may be formed using P-channel parallel filter bank model. Then the output of the i^{th} sub-channel $h_i(k)$ is given by:

$$x_i(k) = \sum_{l=-\infty}^{\infty} s(l)h_i(k - l) + \nu_i(k) : i = 0,..., P - 1 \tag{5}$$

Now, we assume an equalizer $w_i(k)$ which is used in cascade with each subchannel $h_i(k)$. The equalizer coefficients (taps) are adjusted using FSE-CMA algorithm:

$$w(k+1) = w(k) + \zeta x^*(k)y(k)(|y(k)|^2 - \gamma)^2 \tag{6}$$

where ζ is the step size parameter and:

$$\begin{aligned} x(k) = [&x^1(k),..., \quad x^1(k - (N-1)),..., \\ &x^P(k),..., \quad x^P(k - (N-1))]^T \end{aligned} \tag{7}$$

and superscript denotes the subchannel, i.e., fractionally sampled data are organized on subchannel basis. It should be mentioned because of fractional sampling in channel equalization stage, the symbol rate needs to be known or to be estimated prior to choosing the sampling rate

3 Features extraction

In AMTI it is most important to find a set of features which could be used to discriminate the members of considered modulation set. Among the different types of features that we have evaluated and ex-

Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics, Istanbul, Turkey, May 27-29, 2006 (pp409-414)

perimented, higher order moments and higher order cumulants up to eight, produced the most effective features. These features provide a good way to describe the shape of the constellation. Following subsections, briefly describe these features.

3.1 Moments (Mom.s)

Probability distribution moments are a generalization of concept of the expected value, and can be used to define the characteristics of a probability density function. Recall that the general expression i^{th} moment of random variable is given by [29]:

$$\mu_i = \int_{-\infty}^{\infty} (s-\mu)^i f(s) ds \quad (8)$$

where μ is the mean of the random variable. The definition for the i^{th} moment for a finite length discrete signal is given by:

$$\mu_i = \sum_{k=1}^{N} (s_k - \mu)^i f(s_k) \quad (9)$$

where N is the data length. In this study signals are assumed to be zero mean. Thus Eq. (9) becomes:

$$\mu_i = \sum_{k=1}^{N} s_k{}^i f(s_k) \quad (10)$$

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

$$M_{pq} = E[s^{p-q}(s^*)^q] \quad (11)$$

where p called moment order and s^* stands for complex conjugation. Assume a zero-mean discrete signal sequence of the form $s_k = a_k + jb_k$. Using (11), different orders of moment derived, e.g.:

$$\begin{aligned} M_{41} &= E[s^3(s^*)^1] = E[(a+jb)^3(a-jb)] \\ &= E[a^4 - b^4] \end{aligned} \quad (12)$$

3.2 Cumulants (Cum.s)

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

$$\hat{f}(t) = E\{e^{jts}\} \quad (13)$$

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

$$\log \hat{f}(t) = k_1(jt) + \frac{k_2(jt)^2}{2} + \cdots + \frac{k_r(jt)^r}{r!} + \cdots \quad (14)$$

the constants k_i, in (14), called the cumulants. The symbolism for p^{th} order of cumulant is similar to the p^{th} order moment. More specially:

$$C_{pq} = Cum[\underbrace{s,...,s}_{(p-q)terms}, \underbrace{s^*,...,s^*}_{(q)terms}] \quad (15)$$

For example:

$$C_{81} = Cum(s,s,s,s,s,s,s,s^*) \quad (16)$$

It can be computed relation between moments and cumulants.

3.3 Relation between Cum.s and Mom.s

The n^{th} order cumulant is a function of the moments of orders up to including n. Moments may be expressed in terms of cumulants as:

$$M[s_1,..,s_n] = \sum_{\forall v} Cum[\{s_j\}_{j\in v_1}]..um[\{s_j\}_{j\in v_1}] \quad (17)$$

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 n^{th} order cumulant of a discrete signal $s(n)$ is given by:

$$Cum[s_1,..,s_n] = \sum_{\forall v} (-1)^{q-1}(q-1)! E[\prod_{j\in v_1} s_j]..E[\prod_{j\in v_q} s_j] \quad (18)$$

where the summation is being performed on all partitions $v = (v_1,...,v_q)$ for the set of indices $(1,2,...,n)$ .For example:

$$C_{63} = M_{63} - 9M_{41}M_{21} - 6M_{21}{}^3 \quad (19)$$

$$C_{80} = M_{80} - 35M_{40}{}^2 - 630M_{21}{}^4 + 420M_{20}{}^2 M_{40}{}^2 \quad (20)$$

Table1 shows chosen features for considered set (theoretical values under the constraint of unit variance). In this table, for simplifying, we substitute the modulations PSK2, PSK4, PSK8, ASK8, QAM32, V29, Star-QAM8 and QAM64 with P_1, P_2, P_3, P_4, P_5, P_6, P_7 and P_8 respectively.

Table1: Chosen features

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
M_{41}	1	0	0	1.76	0	0	0	0
M_{61}	1	-1	0	3.62	-.380	8.667	2.92	-1.30
C_{63}	16	4	4	7.19	2.11	-4.43	.160	2.11
M_{84}	1	1	1	7.92	2.89	28.75	5.25	3.96
C_{80}	-244	34	1	9.27	-1.99	-198	-88.9	-11.5
C_{82}	-244	-46	0	9.27	-8.41	74.04	63.31	-27.1
C_{83}	-244	0	0	9.27	0	0	0	0

4 Support Vectors Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning technique that can be applied for both binary and multi-class classification [30]. The SVM is based on structural risk minimization (SRM) principle that gives it to have highly generalization ability comparison other approaches (e.g. neural networks, etc.[31].

4.1 Binary SVM

The binary SVM performs classification tasks by constructing optimal separating hyperplanes (OSH).

Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics, Istanbul, Turkey, May 27-29, 2006 (pp409-414)

OSH maximizes the margin between the two nearest data points belonging to two separate classes. The idea of SVM can be expressed as follows.

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 $\vec{w}$ is weight vector and b is bias. If this hyperplane maximize the margin, then the following inequality is valid for all input data:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \text{for all } \mathbf{x}_i \quad i = 1,2,...,l \tag{21}$$

The margin of the hyper-plane is $2/\|\vec{w}\|$, thus, the problem is: maximizing the margin by minimizing of $\|w\|^2$ subject to (21), that is a convex quadratic programming (QP) problem that Lagrange multipliers are used to solve it:

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l}\lambda_i\left[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right] \tag{22}$$

where $\lambda_i, i = 1,...,l$ are the Lagrange multipliers ($\lambda_i \geq 0$).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:

$$\mathbf{w}^* = \sum_{i=1}^{l}\lambda_i^* y_i \mathbf{x}_i \tag{23}$$

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

$$L_d = \sum_{i=1}^{l}\lambda_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \mathbf{x}_i{}^T\mathbf{x}_j \tag{24}$$

To find the OSH, it must maximize L_d under the constraints of $\sum_{i=1}^{l}\lambda_i y_i = 0$, and $\lambda_i \geq 0$ .Those training points for which the equality in (21) holds are called support vectors (SV) that can satisfy $\lambda_i \succ 0$. The optimal bias is given by:

$$b^* = y_i - \mathbf{w}^{*T}\mathbf{x}_i \tag{25}$$

for any support vector x_i.

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

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad for \quad i = 1,2,...,l \tag{26}$$

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 (26), where C is regularization constant that controls how heavily training errors are penalized. In the nonlinearly 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 famous kernel functions are linear, polynomial, radial basis function (RBF), and sigmoid. Having selected a kernel function, the QP problem is:

$$L_d = \sum_{i=1}^{l}\lambda_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{27}$$

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

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{l} y_i\lambda_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*) \tag{28}$$

4.2 Multiclass SVM-based classifier

In this research, we have derived at a novel, simple and effective solution for combining binary SVMs to construct a multi-class classifier. In our algorithm, we use an approach similar to the one reported in reported in [32]. Our approach can be described as follows:

Let $\{P_i : i = 1,2,..,N\}$ be N classesof signals.We construct N classifiers $\{f_i : i = 1,2,..,N\}$ and each classifier is trained by the method of one-c1ass-versus-the-rest; that is, the classifier f_i is trained for P_i versus the rest of the classes. Then in the signal classification phase, the classifiers perform according to the following decision rule:

$$x \Rightarrow P_i \quad if \quad f_i = \max\{f_k(x) \succ 0; k = 1,..., N-1\} \tag{29}$$

where the function $f_k(x)$ provides the distance of x to the decision surfaces.

5 Experimental results

In simulations we have used the channel model that has been introduced in [33]. ADMI was tested under conditions: typical urban propagation environment, mobile speed =85 km/h. SNR levels are considered 0-20 dB. The symbol rate is assumed to be known (or estimated). While classifying using our multi-class SVM-based, we used both Gaussian RBF (GRBF) and polynomial (POLY) kernel functions and obtained a little bit better performance in the case of POLY; however, the computational speed of classification was faster in the case of GRBF. Hence, we used GRBF in our experiments with $K(x, x_i) = \exp(-1/2 \times \|\vec{x} - \vec{x}_i\|^2)$ .Tables 2-4 show confusion matrix for three selected SNR levels, 2dB, 8dB and 17dB. As we see, the results imply how ADMTI can identify modulation type with the high accuracy in dispersive channels even at low SNR. This is due the two facts: chosen novel features and classifier. The chosen features, have highly effective properties in signal representation that enable the classifier to separate modulation set with high accuracy, on the other hand, the SVMs act excellent on non-separable data (low SNR) .

Table2: Confusion matrix at SNR=2dB (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	91							
P_2		84						
P_3			78					
P_4				86				
P_5					76			
P_6						78		
P_7							75	
P_8								71

Table3: Confusion matrix at SNR=8dB (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	100							
P_2		96						
P_3			93					
P_4				98				
P_5					90			
P_6						91		
P_7							88	
P_8								87

Table4: Confusion matrix at SNR=17dB (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	100							
P_2		100						
P_3			100					
P_4				100				
P_5					97			
P_6						98		
P_7							96	
P_8								94

As mentioned in section2, CMA shows good results for PSK types, but for QAM types, when the number of states increases (higher order) and/or the SNR level decreases, its performance degrades. However, simulations imply good results for these types of modulations. The reason for this property should be looked for in chosen features and classifier that cover the weakness of equalizer.

For comparison the performance of SVM with other classifier, we consider a radial basis function neural network (RBF-NN) [34]. Table5, show the performance of RBF-NN for similar situation in SNR=2dB (low SNR). It is found in low SNRs, the RBF-NN shows poor performance. The reason maybe that, in low SNRs, the construction of neural network is not proper. It should be mentioned, though, the success rate of RBF-based system is lower than SVM- based system, but it is higher than a system that uses other features and RBF-NN as a classifier. Table6 shows the performance of a system that utilizes RBF-NN and features in [19].

Table5: Confusion matrix at SNR=2dB for RBF-NN

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	69							
P_2		64						
P_3			61					
P_4				66				
P_5					55			
P_6						56		
P_7							51	
P_8								49

Table6: SNR=2dB for RBF-NN and features in [19] (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	51							
P_2		35						
P_3			33					
P_4				35				
P_5					29			
P_6						21		
P_7							17	
P_8								13

6 Conclusion

AMTI is an important issue in communication intelligence and electronic support measure systems. In this paper, we present ADMTI to identify digital modulations types in dispersive channels. ADMTI uses higher order moments and cumulants up to eight as features and a multiclass SVM-based classifier. Simulation results show ADMTI is able to discriminate different types of modulations with high accuracy even at low SNR.

References:

[1] C.- S. Long, K. M. Chugg, and A. Polydoros, Further results for digital quadrature modulation classification in AWGN, *Proc. MILCOM*,1994, pp. 55-61.

[2] A. Polydoros, and K. Kim, On the detection and classification of quadrature digital modulation in broadband noise, *IEEE Trans. Commun.*,Vol.38, No.8,1990, pp.1199-1211.

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

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

[5] E. E. Azzouz, and A. K. Nandi, Automatic identification of digital modulations, *Signal Processing*, Vol. 47, No.1, 1995, pp. 55-69.

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

[7] W. Wei and J. M. Mendel, Maximum-likelihood classification for digital amplitude-phase modulations, *IEEE Trans. Commun.,* 2000,Vol. 48, pp. 189-193.

[8] K. M. Chugg, C. S. Long and A. Polydoros,Combined likelihood power estimation and multiple hypothesis modulation classification, *Proc. Asilomar Conf.,* 1996, pp. 1137-1141.

[9] L. Hong and K. C. Ho, BPSK and QPSK modulation classification with unknown signal level, *Proc. MILCOM,* 2000, pp. 976-980.

[10] L. Hong and K. C. Ho, Classification of BPSK and QPSK signals with unknown signal level using the Bayes technique, *Proc. IEEE ISCAS,* 2003, pp. IV.I-IVA.

[11] A. E. El-Mahdy and N. M. Narnazi, Classification of multiple *M-ary* frequency-shift keying signals over a Rayleigh fading channel, *IEEE Trans. Commun.,* Vol. 50, 2002, pp. 967-974.

[12] D. S. Clair, G. Hagedorn, B. James, and C. Miller, III, Neural network recognition of signal modulation types, *Artificial Neural Networks In Engineering,* 1997, pp.561-567.

[13] 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, 1998.

[14] M. Richterova, Signal modulation recognition based on method of artificial neural networks, *Proc. Progress In Electromagnetic Research Symposium,* 2005, pp.101-104.

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

[16] P. Feng, L. Ge, Adaptive dbscan-based algorithm for constellation reconstruction and modulation identification, *Proc. ICASSP*, 2004, pp.177-180.

[17] A. O. Hero, and H. Hadinejad-Mehram, Digital modulation classification using power moment matrices, *Proc. ICASSP*, 1998, pp. 3285–3288.

[18] 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.

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

[20] S. Z. Hsue, and S. S. Soliman, Automatic modulation classification using Zero-Crossing, *Proc. IEE*, Vol. 137, No. 6, 1990, pp. 459-464.

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

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

[23] A. Ebrahimzadeh, and S. A. Seyedin, Automatic psk modulation identification using wpa and mlp, *Proc. CIS,* 2005, *pp*.156-161.

[24] Z. Yaqin, Automatic digital modulation recognition using artificial neural networks, *Proc. ICNNSP*, 2003, pp. 257-260.

[25] H. Yoshioka, A fast modulation recognition technique using nearest neighbor rules with optimized threshold for modulation classification in Rayleigh fading channels, *Proc. WPMC,* 2002, pp. 1049-1052.

[26] A. Ebrahimzadeh, and S. A. Seyedin, A prneural modulation identification in fading channels, *Proc. CIS,* 2005, *pp*.280-285.

[27] J. G. Proakis, *Digital Communications,* New York: McGraw-Hill, 2001.

[28] R. Johnson, P. Schniter, T. Endres, J. Behm, D. Brown and R. Casas, Blind equalization using the Constant Modulus criterion: A review, *Proc. IEEE,* Vol. 86, No. 10, 1998, pp. 1927-1950.

[29] P. McCullagh, *Tensor Methods in Statistics,* Chapman & Hall, 1987.

[30] C. Cortes, and V. Vapnic, Support vector network, *Machine Learning*, Vol. 20, 1995, pp.1-25.

[31] C. Burges, A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, Vol. 2, 1998, pp 121-167.

[32] B. E. Boser, and V. Vapnik, A training algorithm for optimal margin classifiers, *5th ACM,* 1996.

[33] P. Hoehor, A statistical disceret-time model for the wssus multipath channel, *IEEE Trans. Vehicular Tech.,* Vol. 41, No. 4, 1992, pp. 461-468.

[34] S. Chen, C. F. N. Cowan, and P. M. Grant, Orthogonal least squares learning algorithm for radial basis networks, *IEEE Trans. Neural Networks,* Vol.2, 1991, pp. 302–309.