Trends in Automatic Modulation Classification for Advanced Data Communication Networks

Praveen S Thakur\textsuperscript{1}, Sushila Madan\textsuperscript{2}, Mamta Madan\textsuperscript{3}

Abstract—Automatic Modulation Classification (AMC) in communication networks has been the hot area of research for last two decades. The initial interests were mainly confined to the military applications. With proliferation of Software Defined Radio and Cognitive Radio in advanced 3G and 4G data communication networks, the research area also spread up in commercial applications. Research so far evolved several multidisciplinary techniques offering solution to AMC problem under different conditions. However for a new researcher approaching the AMC problem for the first time there is no paper available currently that reconcile all available techniques till recent date. In this paper, we addressed this gap and took up an exhaustive survey starting from very early methods to most promising recent techniques that are being applied to AMC problem. Performance limitations of different techniques researched is discussed to identify further direction for future research work in this area.

Index Terms—Modulation, Neural Network, Genetic Algorithm, Fuzzy logic

I. INTRODUCTION

With the increasing demand for radio communications, the task of Automatic Modulation identification of transmitted signal in the RF spectrum has become a front end processing stage and critical requirements in communication systems employed in both military and commercial applications. In military systems, advanced techniques are required for real-time signal interception and processing, which are vital for decisions involving electronic warfare operations and other tactical actions. AMC in military applications is often a difficult task as it has to work in a non-cooperative environment where there is no apriori information about the signal, such as signal power, carrier frequency and timing parameters and this has to be extracted from the received signal. With the growing popularity of software defined radios (SDR) and cognitive radios (CR) in 3G and more advanced 4G communication networks, AMC has become an important enabler in commercial applications also. In software radio it allows the implementation of universal transceiver that enables SDR to dynamically adapt to the communications channel and user applications. Cognitive radios represent the use of artificial intelligence (AI) on flexible software configurable platforms to enable on-board, real-time optimization of frequency, time, power, space, and other parameters based on channel conditions. The purpose of AMC algorithms in a CR is to identify the existence of a signal in a particular frequency band at a given location, and then determine the modulation type being employed in the spectrum. The last available survey paper in this area is [32]. In this paper we addressed the gap since [32] and captured other evolved techniques further till date. The mathematical rigor found in [32] is also scaled down to large extent to arrive on important conclusions. Neural network has been most promising and came a long way since its first application [20] to AMC problem. This has been treated extensively to cover all possible aspects that were researched after [20] & [32]. In the end, comparative analysis of performance is done to identify merit and limitations in both decision theoretic and feature based approaches to AMC problem.

The paper is organized as follows. In section-II the signal and channel models are defined, section-III includes summary of algorithms used in AMC, section-IV gives focus on the performance bottlenecks and comparison of various algorithms and in section-V we conclude and try to identify some possible directions of research in future in AMC area research.

II. ASSUMPTIONS & CHANNEL MODEL

Let the received signal waveform be \( r(t) : 0 \leq t \leq T \) be described as \( r(t) = s(t) + n(t) \) where \( n(t) \) is the standard additive white gaussian noise (AWGN) with a two sided PSD of \( \frac{N_0}{2} \) W/Hz and \( s(t) \) is transmitted signal. The signal \( s(t) \) can be represented in the following ways.

\[
\begin{align*}
    s(t) &= s_\text{c}(t)\cos(\omega_\text{c}t+\Phi_\text{c}) - s_\text{e}(t)\sin(\omega_\text{c}t+\Phi_\text{e}) \\
    &= \sqrt{2S_\text{s}(t)}\cos(\omega_\text{c}t+\theta(t)+\Phi_\text{c}) - \text{ Polar (1b)}
\end{align*}
\]

Here, the PSK signal has power of \( S_\text{s} \), carrier frequency as \( \omega_\text{c} \), carrier phase as \( \theta(t) \) and PSK Phase stated as \( \Phi_\text{c} \) that assumes a value from set of \( K \) complex phases \( \{\Phi_1, \Phi_2, ..., \Phi_K\} \) which is called the constellation of the modulation, \( a(t) \) is

Praveen Singh Thakur, Department of Computer science, Banasthali university, Rajasthan, India.
Sushila Madan, Department of Computer, Delhi University
Mamta Madan, Department of Computer Science, Indraprastha University, India

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET
amplitude and equal to 1 for the constant amplitude modulation. The complex envelope \( \tilde{S}(t) \) of the transmitted signal \( s(t) \) can be expressed as

\[
\tilde{S}(t) = \sqrt{2S} \sum_{n=0}^{N-1} e^{j(\theta_n)} a_n(k) p(t - nT_s - \epsilon T_s)
\]

\( p(t) \) is spectrum shaping pulse assumed to be normalized i.e \( \int |p(t)|^2 dt = T_s \). Carrier Phase \( \theta_c \) is assumed to be constant for the period of the observation and \( \epsilon \) is symbol timing offset of signal \( s(t) \), \( a_n(k) \) is sequence of i.i.d complex random variables such that \( E|a_n(k)|^2 = 1 \), SNR \( \rho \) of the signal with the signal duration \( T_s \) can be represented as \( \rho = ST_s/N \). The M-PSK modulation process can be considered as the process in which sequence of complex phases are drawn from process \( \tilde{S}(t) \) which are independent of each other in time duration \( T_s \). The problem of Automatic Modulation Classification is to identify the transmitted constellation or phases based on the received signal \( r(t) \) corrupted with noise.

\[ r(t) = s(t) + n(t) \]

There are three environments that are normally considered in AMC study. The first is the Coherent and Synchronous environment where the carrier phase \( \theta_c \) and the symbol timing offset \( \epsilon \) are assumed known. In the Non coherent and Asynchronous environment both these parameters are modeled as random variable uniformly distributed in \( (0,2\pi) \) and \( (0,1) \) respectively. The Non coherent and Synchronous is intermediate state where one parameter is assumed to be known and other as random variable.

Most of the study reported in literature is made with assumption that \( S, \theta_c, p(t), \epsilon, \) and \( \rho \) are known and in addition it is assumed that the data symbols are independent and pulse shape used is rectangular only. Although these conditions do not address the practical problem of AMC but still they help in establishing performance upper bound for the evolved technique. With these assumptions the signal available at demodulated matched filter output is

\[
\tilde{r}(n) = \int_0^{NT_s} r(t) e^{j\int_0^t \theta(t) dt} \sqrt{2S} p(t-nT_s-\epsilon T_s) dt
\]

Where \( r_{ln} \) is In phase component and \( r_{Qn} \) Quadrature component

III. ALGORITHMS IN AUTOMATIC MODULATION CLASSIFICATION

The most popular techniques that are employed in AMC are shown in Figure-1. The first among the techniques developed were identification for AM, FM, PM modulations. The focus eventually shifted to identification of digital modulations and their order of modulation. In general the automatic modulation classification framework can be categorized as decision theoretic approach [1-35] and Feature based approach [36-93]. In decision theoretic approach for classification, probabilistic and hypothesis testing arguments are used to formulate the classification problem and then to obtain classification rule. In feature based approach the classification problem is formulated in two stages. The first is feature extraction stage where appropriate feature are identified and useful information is extracted from the raw data. This can be viewed as mapping of input signal in to a selected feature space. The second stage compromises of the pattern recognizer stage where features are mapped to the modulation type. This essentially is mapping of feature space to decision space. There are many other techniques [76 - 93] overlapping feature based approach that has been used in the AMC area and which are still evolving.

A. Preliminary Methods

The earliest techniques [1-9] were based on measurement of time domain parameters like amplitude, instantaneous frequency and phase of the signal. This exploits the histogram characteristics of selected parameters that are unique for different modulation schemes. In [1-2] histogram of instantaneous phase is estimated. It formulated automatic histogram separation method by an optimal filter weighting function that separates two histogram from each other with minimal error probability as follows.

\[
f(x_0,x_1,x_2 \ldots ,xm-1) > TL.
\]

B. Decision Theoretic Approach

The Kim and Polydorous [7-8] were first to publish the technique based on decision theoretic approach to identify the BPSK and QPSK signals. The concept here states that the Likelihood functional (LF) or more specifically the Log-Likelihood (LLR) of the observed waveform conditioned on the embedded digitally modulated signal consist all useful information that can be used for various
Threshold comparison

<table>
<thead>
<tr>
<th>AMC Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary methods</td>
</tr>
<tr>
<td>Decision theoretic</td>
</tr>
<tr>
<td>Statistical methods</td>
</tr>
<tr>
<td>Feature based methods</td>
</tr>
<tr>
<td>Miscellaneous methods</td>
</tr>
<tr>
<td>Time domain [1,2,6]</td>
</tr>
<tr>
<td>Envelope characteristics [4,5]</td>
</tr>
<tr>
<td>Zero crossing [9]</td>
</tr>
<tr>
<td>qLLR [7,8,10,11]</td>
</tr>
<tr>
<td>Maximum Likelihood [12,13,14,15,17]</td>
</tr>
<tr>
<td>GLRT &amp; ADMP [16]</td>
</tr>
<tr>
<td>Higher order correlation [21,22,23]</td>
</tr>
<tr>
<td>Higher order moments [24,25]</td>
</tr>
<tr>
<td>Higher order cumulant [26,27,28,29,30,31,33,34,35]</td>
</tr>
<tr>
<td>Neural networks [36-47,49-55,62-65]</td>
</tr>
<tr>
<td>Wavelet transform [47,56,58,66-72,93]</td>
</tr>
<tr>
<td>Fuzzy rough theory [78,79,81]</td>
</tr>
<tr>
<td>Genetic Algorithms [46,54,73,74,75,76,83]</td>
</tr>
<tr>
<td>Constellation [80,82,90]</td>
</tr>
<tr>
<td>SVM &amp; SPRT [69,76,77,84,85,88,89,92]</td>
</tr>
<tr>
<td>Phase detector reliability [86]</td>
</tr>
</tbody>
</table>

Fig-1 : Popular techniques used in Automatic Modulation Classification problem.

Threshold comparison [18,19,20]

Purpose like detection of signal, estimation of its transmitting parameters and or classifying the signal types as well. In decision theoretic approach AMC problem is viewed as multiple-hypothesis testing problem where modulation is selected from a candidate list by forming and maximising likelihood ratio. A simple binary test is constructed for testing a hypothesis $H_1$ corresponding to say MPSK$_1$ and hypothesis $H_2$ for another specific modulation say MPSK$_2$. The optimal classification rule for this binary hypothesis is simply then

$$L_{MPSK_1} - L_{MPSK_2} > \text{Threshold} \quad \text{(6)}$$

In [10,11] the decision theoretic approach was extended to classification of any M-ary Signals. The problem of classifying between M possible modulations given N phase rotated symbols in AWGN is resolved by forming following multiple hypothesis problem and optimising likelihood ratio

$$H_k : r = \sqrt{S} e^{j\theta} a^{(k)} + k = 1,2 \ldots M \quad \text{(7)}$$

This problem is solved by considering it as pair wise hypothesis problem and modeling the unknown parameters as random variables. The likelihood ratio test denoted by symbol $\mathcal{L}$ for multiple hypotheses problem for $M=2$ and 4 is formulated as below and realized as in figure-2

$$\mathcal{L} \equiv \frac{\text{qLLR}}{\text{GLRT}} > (\text{Threshold}) \quad \text{(8)}$$

The maximum likelihood methods [12-15] formulates rule for choosing among the candidate modulation type such that above pdf, $F$ governing $r$ in equation (4) is maximized. However, the pre-condition for ML approach is that the pdf of parameter is supposed to be known. Both qLLR and maximum likelihood classifier suffers from limitation that it works well in many cases but there performance is not good with constant envelope modulation like QAM signals. The Average Likelihood Ratio Test (ALRT) and the Generalised Likelihood Ratio Test (GLRT) removes these limitations.

1) Average Likelihood Ratio Test (ALRT)

ALRT [16] is applied when there is possibility to model the unknown parameters of complex signal as random variables and attach a probability distribution function (pdf) to them. It gives optimal solution only in this condition. The ALRT is expressed as following mathematically \[16,32\].

$$\mathcal{L} \equiv \frac{\text{qLLR}}{\text{GLRT}} > (\text{Threshold}) \quad \text{(9)}$$

here, $p$ is the vector of the unknown parameters of the signal. $f(r \mid p, H_i)$, $f(r \mid p, H_j)$ is the pdf of vector $r$ with Hypothesis $H_i$ and $H_j$, $f(p \mid H_i, p \mid H_j)$ are prior known pdf.

2) Generalised Likelihood Ratio Test (GLRT)

If we cannot attach the pdf to the parameters of interest in signal then they are modeled as fixed but unknown variables

$$f(x,y) = \exp \left\{ \frac{-z}{2M} \sum_{m=0}^{M-1} \exp \left[ \sqrt{2\rho} \right] \right\} \quad \text{(10)}$$
and this is GLRT. Since GLRT does not requires any assumptions of signal this classifier can be applied to number of conditions. The major limitation [16] of the GLRT is that it requires nonlinear, multidimensional, maximization computations

\[
\mathcal{g}_{\text{GLRT}}(\tau) = \frac{\max \left\{ x(t) \right\}}{\max \left\{ y(t) \right\}} \quad \mathcal{H}_0: \mathbf{Threshold} - (11)
\]

3) Hybrid Likelihood Ratio Test (HLRT)

The ALRT needs a multidimensional integration whereas GLRT requires multidimensional maximization. The ALRT turns out to be an impractical solution due to requirement of performing of multidimensional integration for large number of unknown quantities and need of knowing apriori pdf of the signal. The limitation in GLRT is that the maximization process may lead to same value of likelihood functional if signal is having nested constellation scheme. Both these problems are removed in HLRT which is combination of ALRT and GLRT. There are number of parameters p, used under the rigorous framework of decision theoretic approach. Most popular [18-20] parameters are either instantaneous phase , instantaneous amplitude, instantaneous phase or there derivation like standard deviation of normalized centered value of the normalized instantaneous frequency, standard deviation of absolute value of the normalized instantaneous amplitude , kurtosis of the normalized instantaneous phase, spectrum , maximum value of the spectral power density , standard deviation of the absolute value of normalized instantaneous frequency , kurtosis of the normalized instantaneous amplitude , kurtosis of the normalized instantaneous frequency

\[ C_{m}(\tau) = \int_{0}^{T_{s}-\tau} x(\tau) x(\tau + \tau) d\tau \quad \text{for } m = 1, 2, 3, \ldots \]

Where Ts is symbol duration. The \( C_{m}(\tau) \) is defined for \( m = 2, 3, \ldots \) as the energy of the \( m \)th order correlation function. The \( m \)th order correlation is defined by

\[ C_{m}(\tau) = \int_{-T_{s}}^{T_{s}} x(t) x(t + \tau) d\tau \quad \text{for } m = 1, 2, 3, \ldots \]

Where Ts is symbol duration. The \( C_{m}(\tau) \) is called HOC of \( x(t) \). The HOC statistics is used to approximate the Average Likelihood function (ALF) and to any desired level of accuracy. This is possible because the ALF series and HOC is Fourier transform pair [21].

ALF coefficient \( \hat{C}_{m} \) is defined by

\[ \hat{C}_{m} \equiv \text{FT} (f(t)) \]

For Eg. 4th order likelihood term equals the energy of the 1st order correlation function

\[ \hat{C}_{1}(\tau) = \int_{-T_{s}}^{T_{s}} x(t) x(t + \tau) d\tau \quad \text{for } m = 1, 2, 3, \ldots \]

The 6th order correlation term equals inner product of the 1st and 2nd order correlation function

\[ \hat{C}_{2}(\tau) = \int_{-T_{s}}^{T_{s}} x(t) x(t + \tau) d\tau \quad \text{for } m = 1, 2, 3, \ldots \]

In general

\[ \hat{C}_{p} = \int_{-T_{s}}^{T_{s}} x(t) x(t + \tau) d\tau \quad \text{for } p > 1 \]

2) Signal Moment

In [24-25] different non linearity in the complex envelope of the signal are used to exploit differences in the higher order moment spaces. The \( n \)th order moment [25] of \( x(t) \) is defined as

\[ M_{n}(t, \tau, n, m) = E \left[ \prod_{j=1}^{n} x^{n}(t + \tau_j) \right] \quad (18) \]
Where $\tau$ is delay vector, $n$ is the order, $m$ is the number of conjugated factor, $(^*)$ is optional conjugation. In [25] it is shown that when moments of instantaneous phase of MPSK signal is taken, the even order moment are monotonic function of the order of modulation and the type of modulation can be determined by setting up appropriate thresholds.

3) Signal Cumulant

In Cumulant method are used more widely over moment and other methods due to its advantages that it is robust against phase rotation errors, gaussian noise and are easy to compute. They are mostly used as regular cumulant [27, 31, 32] or used as cyclic cumulant [26, 28-30]. For the signal $x(t)$ the time varying $n^{th}$ order cumulant $C_x$ and $n^{th}$ order moment $M_x$ are associated as [28]

$$C_x(t, \tau) = \sum_{p=1}^{P} K(p) \prod_{j=1}^{p} M_{x}(t: \tau_{jp})$$  \hspace{1cm} (19)

Where $P_n$ is the set of partitions, $\tau$ is delay vector, $K_p$ is number (-1) $(p-1)!$, $l=1(1,2 ...n)$ set of indices, $n_{ip}$ is the number of element of the subset. In terms of simple expected values expression, Let $X_i$ be a signal vector = $x_1^n, x_2^n, x_3^n, ... , x_n^n$ and $<>$ denotes the statistical expectation. We have following expression of second, third and fourth cumulants at zero lag

$$C_{X1,X2} = <X1, X2>$$
$$= \frac{1}{N} \sum_{n=1}^{N} x_1^n x_2^n$$  \hspace{1cm} (20)

$$C_{X1,X2,X3} = <X1, X2, X3>$$
$$= \frac{1}{N} \sum_{n=1}^{N} x_1^n x_2^n x_3^n$$  \hspace{1cm} (21)

$$C_{X1,X2,X3,X4} = <X1, X2, X3, X4> = <X1, X2, X3, X4> = <X1, X2, X3, X4> = <X1, X2, X3, X4> = <X1, X2, X3, X4>$$
$$= \frac{1}{N} \sum_{n=1}^{N} x_1^n x_2^n x_3^n x_4^n$$

Cumulant statistics are calculated and are combined to form the cumulant value. Since cumulant above order four are not affected by the gaussian noise they are more popular in AMC study. Once the cumulant statistic is available there are two popular approaches followed. The simplest case is one shot method where estimated cumulant value is compared with the ideal value. Table [1] list ideal theoretical cumulant value of $n^{th}$ order /q conjugate for few modulations and a more exhaustive table can be found in [32]. In the second method hierarchical schemes are used to isolate the family of modulations M-ASK, M-FSK, M-PSK, M-QAM more accurately than isolating a specific modulation.

### D. Feature based Approach

The preliminary methods and decision theoretic approach

<table>
<thead>
<tr>
<th>Cumulant</th>
<th>BPSK</th>
<th>QPSK</th>
<th>8PSK</th>
<th>16-PSK</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_{2,0}</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C_{2,1}</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C_{4,0}</td>
<td>-2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C_{4,1}</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C_{4,2}</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The decision theoretic approach major drawback were to formulate the right hypothesis, determination of the threshold value, not computation friendly, not fail proof in noisy environment, not robust enough for phase, frequency offsets and synchronization errors. However they lay robust mathematical framework in selection of communication parameters that characterize the type of modulation accurately where some apriori information like pdf of some communication parameters is known. Feature based approach [91] is less complex and implementation friendly but are not optimal. They can work under different conditions and does not require communication parameters information apriori. Figure (3) depicts generic steps involved in a feature based approach. It consists of three stages. The first step is a pre processing step which is a signal processing step to remove noise, unnecessary information to emphasis only discriminatory features from the raw communication signal. The second stage is called the feature extraction step which extract useful information corresponding to exact feature from the processed data to map the information on chosen feature space. The third is the pattern classifier stage which finds out membership of the extracted feature to a family of modulation class. The performance of this approach depends heavily on the choice of the feature set in the second stage and the choice of classifier in third stage. In this section more focus is given to third steps and in particular to Neural Network that holds lot of promise to offer practical and efficient classifiers for modulation classification. The other less researched techniques are also covered and can be explored further as part of future research work.

1) Neural Network Approach

Neural network techniques in Automatic modulation classification mainly involve application of different architectures, trying different learning techniques and the identification of optimal parameters used to train the network. There is a tradeoff involved between getting accurate classification results and practical realizable network with minimum training time. Conventional Neural Tree network and K-nearest neighbor network were evaluated as early part of research in [36, 37, 500]
The performance of the neural network depends highly on the choice of the feature set used as input vector. A good feature set should yield limited training data, memory and computation power. It should be insensitive to the transmission effects on the communication parameter and should remain sensitive to only the class of modulation to which it belongs. The quality of feature set is closely associated with the kind of the classifier used. Therefore in literature plenty of research is seen to be done to study impact of various feature sets with different neural network structures in classifier stage. Some of the most popularly used feature sets that are tried out with different network architecture are Time domain parameters [4,5,36,42,44], Spectral features [20,39,40,41] , Signal moments [24,25,43] and Signal Cumulant [ 26-31,33]. Other techniques like Genetic Algorithm [73] , Fuzzy logic [78] were investigated to aid in optimal selection of features from multiple feature sets thereby reducing the complexity of the network architecture.

2) Wavelet transform

(WT) are used effectively in [47,56,58,66,67,69,70,72,93] by exploiting transients in amplitude, frequency or phase of cyclostationary characteristics of the modulated digital signal which are exhibited as distinctive peaks. Since the computation of wavelet transform is possible in real time it is a preferred approach for real time systems. In [69] , wavelet transform is combined with wavelet support vector machine to classify unstable signals buried in noise and in [93] wavelet cyclic feature is used to calculate cyclic spectrum feature which is a powerful feature to classify modulation for signals in noisy environment. The method requires only sub-nyquist rate to calculate spectrum feature.

3) Support vector machines ( SVM)

SVM [69,71,77,88,89,92] is emerging as a strong contender against neural network mainly due to its advantages that it avoids local minima trapping problem of neural network , has simpler geometric interpretations and its computational complexity does not depends on the dimension of the input vector. SVM is used as hierarchical structure as multiclass classifier and constructing an optimal separating hyper plane that has ability to maximize the boundaries of two nearest data points from two separate classes. SVM inherently being the binary classifier, strategies like one-against-all and one-against-one to get solution for multi class problem in AMC

4) Fuzzy Logic

Fuzzy logic based modulation classification has been employed in [78, 79, 81]. Fuzzy logic batch processing architecture is utilized in [78] where each fuzzy module in it produces fuzzy modulation as its output. The Fuzzy decisions from all Fuzzy modules are combined to get a fuzzy decision which is finally converted to a hard decision by a defuzzifier module. When fuzzy rules are derived from maximum likelihood principles [55] the performance matches with ML classifier in an ideal condition. Fuzzy logic classifier offers additional advantage that the classifier can be used in non-ideal condition also in which Maximum likelihood classifier fails since it’s easy to model non ideal conditions in

Fig-3 Feature based modulation identification process

38,59] that uses tree architecture to implement sequential linear decision strategy where feature vector evaluation ends at the leaf node which corresponds to the modulation of the feature vector. Most popular architecture involved multi-layer perceptron networks [39-64] with varying number of nodes to identify modulation with different accuracy under different conditions. Nodes in the architecture depend on the number of modulation that needs to be identified and the features considered in the input. The number of nodes determines the rate at which the network learning converges. The principal problem with multilayered network is that if the network is very small it cannot learn and if the network is large it cannot generalize. In [39] Nandi and Azzouz in their pioneering work suggested a 4 node input layer, 25 nodes hidden layer and 7 nodes output layer to identify analog modulation. To identify digital modulation network structure they tried 5 input nodes, one hidden layer with 10 nodes and 6 nodes to get good results. Some architecture [40, 41] utilizes no hidden layer or may utilize even more than one hidden layer. In some cases [43] even up to 50 nodes has been tried. Same results from [39] are also replicated by [44] by employing only 5 input node, 25 nodes in hidden layer and only 3 nodes in output layer thereby giving less training epochs than as compared to their predecessors. Multi-layer perceptron are by far the most widely network structure used to modulation classification. However it has limitation that its training procedure get stuck up at local optimum of the cost function .This is avoided by using Radial basis function network (RBFN) in which the activation function of the hidden nodes are radial basis function [ 45,48,53,57].

The training method used in the structure chosen has important impact on the performance of the classifier. The objective [39] of training the network is to determine the weights and biases of the various neurons in the network. The training algorithm updates the weight such that it tries to bring the output of the neural network close to the theoretical output in target vector. There are many methods of learning studied [40 , 65, 87] like Back error propagation , the Hebbian learning , competitive learning , Boltzman learning , Back propagation with momentum , Quick Prop , SuperSB , Conjugate gradient , Extended delta-bar-delta algorithms. The learning can be supervised or unsupervised. In supervised learning the training data consists of both input and output pairs whereas in unsupervised case the output is not presented to the network. Unsupervised learning is the characteristic of self-organizing networks [55] in which the network learn to adapt based on experience collected from previous training pattern. The most popular activation functions that are used to train the network are log-sigmoid [39, 40, 42, 47], tan-sigmoid [41, 52] which normally leads to a smaller network size and Radial basis functions [45, 48, 53, 57] which leads to rapid convergence apart from avoiding network to get stuck up in local optima.
fuzzy model than in probabilistic models. In [79] Fuzzy logic driven adaptive network inference system is constructed that analyses 20 feature set to classify digital modulations using wavelet adaptive entropy and construct network that uses to evolve its fuzzy if-then rules and an hybrid knowledge base as learning system. In another approaches fuzzy rules are created to find out feature level distance of parameters used by Nandi and Azzouz in [39–40]

5) Genetic Algorithm

GA approach [46, 54, 73, 74, 75, 76, 83] has also been tried by many researchers in the AMC problem. It has been used in two ways. The most successful usage of GA is to select the best feature set from the available feature sets where Artificial Neural Networks or Support Vector Machine is employed as classifiers. Its chief usage in AMC area is to enhance the performance of the ANN /SVN by aiding in best feature selection [46, 54, 74, 76] since performance of these classifiers are well known to be dependent on the features chosen. GA also has been used directly as classifiers in some cases [73, 75, 83] directly because of the facts that if more time is given to GA evolution its performance improves justifying its usage in classifier stage. When GA is used in classifier stage several techniques are used. It classifies the unique constellation by separating the forming and distinguishing constellation clusters hierarchically. GA efficiency and simplicity of being a binary classifier is exploited by converting it in to multi class classification problems where fitness functions is evaluated by setting class threshold criterion or using K nearest neighbor techniques.

6) Miscellaneous Algorithms

Under miscellaneous algorithms we have consolidated three algorithms which shows good potential and hence are good candidates for further research to check their performance against other more researched approaches. Constellation [80,82,90] has been tried in AMC that exploits the fact that constellation of different digital modulation are unique. In this approach the constellation points are assembled in one quadrant as clusters. The clusters are further analysed using fuzzy C-means [82] and centroids are calculated using TTSAS algorithm [80] as improved step that is implemented using 2 layer neural network to generate score and to choose best score. The resultant cluster centroids are matched to an ideal template to get an accurate classification result. The advantage of constellation approaches is that apriori information about signal is not necessary. Sequential Probability Ratio[84,85] test principles (SPRT) are used for classifying PSK & QAM signals. The sequential tests comprises of set of stopping rules and set of decision rules. In identification of PSK and QAM signals using SPRT principle the decision boundaries are based on pdf of phase of the signal. The received signal is divided in to number of windows and ML of phase is calculated for these windows. The final test statistics used in setting decision boundaries are the product of statistics data from individual windows Phase Detector Reliability [86] is used to classify the modulation scheme. The reliability is maximum in the neighborhood of the equilibrium point of the particular modulation. This is used as indicator of the underlying modulation scheme in the signal. In [93] , modulation is classified using phase entropy of the signal. A hypothesis test is used to classify the type of modulation of M-ary signal.

IV. PERFORMANCE ANALYSIS

In Automatic modulation classification problem, comparing the performance of the different published classifiers is not a straightforward task. This is because that there are number of methods tried to reach to the solution of the problem and their conditions are not same. First, performance of different classifiers cannot be compared, if the candidate modulations are not same. Second, most of the classifiers are designed to handle specific unknown parameters. So, the classifiers proposed algorithm should be tested in the same conditions for reliable comparison of their performance. We have outlined the performance of few sample classifiers taken from the studied techniques in this paper in table-II and table-III. The benchmark of comparison taken is signal-to-noise ratio (SNR) which is a very common criterion of performance measure in a communication system. There can be any other criterion as well chosen that impact the performance of the modulation classifier under design. Some of them can be, performance in particular channel condition, apriori knowledge or no knowledge of transmission parameters, accuracy to identify specific modulation from inter class, accuracy to identify specific modulation from intra class, number of samples used in classification etc. However they are presently ignored to get a comparative assessment of studied techniques in this paper but can be always selected as another benchmarking criterion to rate AMC algorithm.

Table-II summaries results of decision theoretic approaches and the features used by different authors. Decision theoretic approaches lay down robust mathematical explanation to support the chosen parameters, if appropriate thresholds are chosen. The limitation is however is that they do not convert to easily realizable systems. The early methods report good identification accuracy at higher SNR. Quassi Likelihood methods by Huang and Polydorous [7, 8, 10] for the first time reported high accuracy for SNR< 0dB albeit there modulation problem is two class problem. Nandi & Azzouz [18, 19, 20] used decision theoretic approaches to identify multiple modulation and at lower SNR than reported by traditional methods. Further techniques of higher order correlation methods, moment based approaches and higher order cumulant based studies yields classifiers that work at much lower SNR and offers additional immunity to noisy channels.

Table –III summaries the performance of feature based classifiers employing neural network in classification stage and it shows improvement over decision based approaches. Nandi classifier achieves more than 93% success rate at 10dB and 100% at 20dB. This was improved to 97% at 10dB in [41] along with improvement in training time. With further use of combination of spectral and cumulant
### Table 1: Performance analysis of decision theoretic approaches

<table>
<thead>
<tr>
<th>Modulation Type</th>
<th>SNR range</th>
<th>Accuracy at lowest SNR</th>
<th>Feature used</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK2,PSK2,PSK4</td>
<td>10dB-20dB</td>
<td>95.4%-100% depending on modulation.</td>
<td>Histogram of instantaneous amplitude, phase and freq.</td>
<td>Liedke, Jondral [1,2]</td>
</tr>
<tr>
<td>AM,FM,SSB,DSB</td>
<td>7dB-10.4 dB</td>
<td>100%-80% depending on modulation.</td>
<td>R - Ratio of variance of envelope to square of mean</td>
<td>Chan [5]</td>
</tr>
<tr>
<td>AM,DSB,SSB,VSB</td>
<td>10dB</td>
<td>91%-100% Accuracy in confusion matrix</td>
<td>Four Parameters deduced from Instantaneous Amplitude &amp; Phase</td>
<td>Nandi [18,19,20]</td>
</tr>
<tr>
<td>BPSK, QPSK</td>
<td>-2dB - 8dB</td>
<td>100% probability of correct classification</td>
<td>Quassi Likelihood ratio of Phase</td>
<td>Huang [10]</td>
</tr>
<tr>
<td>PSK16 Vs 16QAM</td>
<td>25dB</td>
<td>100% probability of correct classification</td>
<td>Quassi Likelihood ratio of Phase</td>
<td>Long[11]</td>
</tr>
<tr>
<td>16QAM,32QAM ,64</td>
<td>10dB-15dB</td>
<td>100% probability of correct classification</td>
<td>Maximum Likelihood of discriminating statistics of I-Q data</td>
<td>Wei[17]</td>
</tr>
<tr>
<td>BPSK, QPSK,PSK8</td>
<td>5dB</td>
<td>Probability of mis classification $10^{-4}$</td>
<td>Moment of phase of signal</td>
<td>Samir [25]</td>
</tr>
<tr>
<td>BPSK,PAM4,QAM4,</td>
<td>5dB-10dB</td>
<td>100% accuracy in confusion matrix</td>
<td>Second order cumulant</td>
<td>Swami[27]</td>
</tr>
<tr>
<td>PSK8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPSK,PAM,PSK4,PSK8,AS,PK4,PSK4,FSK2,FSK4</td>
<td>10dB-20dB</td>
<td>93%</td>
<td>Four Parameters deduced from Instantaneous Amplitude &amp; Phase</td>
<td>Nandi[39,40]</td>
</tr>
<tr>
<td>BPSK,QPSK,PSK8,PK4,AM16,64QAM ,FSK2</td>
<td>5dB-10dB</td>
<td>80%-95% Average probability of correct classification for two class</td>
<td>Fourth, Sixth, Eighth order cumulant</td>
<td>Dobre [28]</td>
</tr>
</tbody>
</table>

### Table 2: Performance analysis of Feature based approaches

<table>
<thead>
<tr>
<th>Modulation Type</th>
<th>SNR range</th>
<th>Accuracy at lowest SNR</th>
<th>Feature used</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSK,QPSK,PSK8,PSK8,PK4,AM16,64QAM,F SK2</td>
<td>5dB-50dB</td>
<td>63 %</td>
<td>Mean and next three statistic moment of phase, amplitude &amp; frequency derived from complex samples</td>
<td>Louis [38]</td>
</tr>
<tr>
<td>AM,DSB,VSB,LSB,USB,FM,ASK2,ASK4,PSK4,PSK8,F SK2</td>
<td>10dB-20dB B</td>
<td>93%</td>
<td>Statistics of instantaneous amplitude, phase, frequency, spectrum symmetry.</td>
<td>Nandi[39,40]</td>
</tr>
<tr>
<td>ASK2,ASK4,PSK2,PSK4,FSK2,FSK4</td>
<td>10dB-20dB B</td>
<td>97%</td>
<td>Statistics of instantaneous amplitude, phase, frequency, spectrum symmetry.</td>
<td>Arulampalam [41]</td>
</tr>
<tr>
<td>ASK2,ASK4,PSK2,PSK4,FSK2,FSK4</td>
<td>10dB-20dB B</td>
<td>97%</td>
<td>Combination of spectral and higher order cumulant</td>
<td>Wong [46]</td>
</tr>
<tr>
<td>CW,FSK2,FSK4,FSK8,ASK4,PSK4,PSK8,PK4,AM16,64QAM ,PSK8</td>
<td>5dB</td>
<td>94.9%</td>
<td>Power spectral features, Cyclic Spectral features of IF signal, Higher order cumulant</td>
<td>Wu [49]</td>
</tr>
<tr>
<td>ASK4,ASK8,FSK4,BPSK,PSK8,PK4,AM16,29,32QAM,64</td>
<td>-5dB</td>
<td>89.76%</td>
<td>Combination of signal moment and cumulant</td>
<td>Ebrahimzadeh [53,54]</td>
</tr>
<tr>
<td>PAM4,PAM8,FSK2,FSK4,PSK4,PK8</td>
<td>2dB</td>
<td>97% .93%</td>
<td>Combination of spectral and statistical features with weights based on degree of dispersion.</td>
<td>Khurshid [57]</td>
</tr>
<tr>
<td>ASK4,ASK8,FSK4,BPSK,PSK4,PK4,PSK8,AM16,32,64</td>
<td>4dB</td>
<td>99% Higher order statistical moment of wavelet</td>
<td>Hassan [58]</td>
<td></td>
</tr>
<tr>
<td>ASK4,ASK8,FSK4,BPSK,PSK4,PSK8,AM16,32,64</td>
<td>-5dB</td>
<td>91.03%</td>
<td>Energy of signal, zero crossing rate, variance of wavelet packet transform, entropy of signal</td>
<td>Ebrahimzadeh [65]</td>
</tr>
</tbody>
</table>
Automatic modulation classification has been evolving continuously and harnessing latest techniques in pursuit of achieving fast, compact & accurate solutions that is applicable across all analog and digital modulation signals. Optimal features set selection is a first step for any classification algorithm and chosen feature set should characterize the underlying modulation type adequately. Decision theoretic approach gives sound mathematical justifications for features chosen for classification but are not practical system friendly. Pattern recognition approach and in particular classifiers employing artificial neural network are able to meet or exceed decision theoretic performance by use of various other techniques like Particle swarm optimisation, Genetic Algorithm aided feature filtering and Principal component analysis. Feature based approach thus has high potential against decision theoretic approaches and more classifiers can be investigated further. Robustness of AMC algorithms to all channel impairments is necessary for converting it to a practical solution since all real time communication signals inevitably suffer channel impairments. The challenge in it is to achieve to get accurate results at as low SNR as possible. Further, the algorithm should not depend on apriori information of communication parameters for it to work accurately. These blind classifiers have special significance where transmission parameters are not known. Real time implementable solution that adapts to any channel condition and can classify any type of modulation types without any prior information should be the end target of any AMC algorithm under research pursuit.

REFERENCES

[26] C. M. Spooner, “Classification of cochannel communication signal using cyclic cumulants”, in Proc ASILOMAR'95 , pp. 531-536, 1995


[60] Ibrahim Hafiz , Nour Madany , “Cognitive Digital modulation classifier using Artificial Neural network for NGN’s”, 7th International Conference on Wireless and Optical Communication Networks ( WOCN), pp 6-8, Sep 2010


[64] Ataollah Ebrahimzadeh , “Hybrid Intelligent technique for automatic communication signal recognition using Bees Algorithm and MLP neural network based on the efficient features” , Expert system with Applications ,38, 2011, 6000-6006


[80] Negar Ahmadi , “Using Fuzzy Clustering and TTSA algorithm for modulation classification based on constellation diagram”, Engineering Application and Artificial Intelligence , pp 145-151, 2010


[84] Yu Chuan , C Jay Kuo , “Classification of quadrature amplitude modulated (QAM) signals via sequential probability ratio test (SPRT)” , Signal Processing , 1997


[93] Lei zhou; Hong Man, “Wavelet Cyclic Feature Based Automatic Modulation Recognition Using Nonuniform Compressive Samples “, IEEE 78th Vehicular Technology Conference (VTC Fall), pp 1-6, 2013