Automatic Modulation Recognition in OFDM Systems using Cepstral Analysis and Support Vector Machines

Rasha M. Al-Makhlasawy1, Mustafa M. Abd Elnaby1, Heba A. El-Khobby1, S. El-Rabaie2 and Fathi. E. Abd. El-samie2

1Faculty of Engineering, Tanta University, Tanta, Egypt
2Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt

Abstract

This paper discusses the modulation recognition for OFDM signals in different Signal to Noise Ratio (SNR) and multipath channels. In this paper, the Mel Frequency Cepstral Coefficients (MFCCs) used for feature extraction and the Support Vector Machine (SVM) as classifier or Artificial Neural Network (ANN). Simulation results indicate that the proposed feature classifier have good performances in different SNR and multipath channels for both recognition rate and CPU time from the Artificial Neural Network (ANN), and the SVM classifier’s generalizing ability proves to be good.

Keywords: ADMR; MFCCs; SVM; OFDM; LMMSE; ZF

Introduction

Automatic modulation recognition of digital communication signals is an important signal processing problem in communications and communication related fields. It is an intermediate step between signal interception and information recovery, which automatically identifies the modulation type of the received signal for further demodulation and other tasks. One problem is how to receive the signals modulated by various modulation types optimally, which can be done by using the automatic modulation recognition technology to identify the modulation type and its order, and then to make the system switch to the suitable demodulation system.

Yun Shi et al. [1], discusses Automatic Modulation Classification (AMC) of analog schemes. Histograms of instantaneous frequency are used as classification features and Support Vector Machines (SVMs) are then applied to classify the unknown modulation schemes. This novel machine-learning based method can ensure robustness in a wide range of SNR. Extensive simulation has demonstrated the validity of the AMC algorithm. It is a practical algorithm in blind AMC environments.

Punchihewa et al. [2], studies the cyclostationarity of an Orthogonal Frequency Division Multiplexing (OFDM) with a view to recognizing OFDM against Single Carrier Linear Digital (SCLD) modulations. The analytical expressions for the nth-order Cyclic Cumulants (CCs) and cycle frequencies of an OFDM signal embedded in additive white Gaussian noise and subject to phase, frequency and timing offsets are derived. An algorithm based on a second-order CC is proposed to recognize OFDM against SCLD modulations.

Dobr et al. [3], has summarised the two main approaches to Automatic Modulation Classification (AMC) that is, the Likelihood-Based (LB) and the Feature-Based (FB) methods, and has lightened their advantages and drawbacks.

Support Vector Machines (SVMs) are a useful technique for data classification. Although, SVM is considered easier to use than Neural Networks, the overall idea for the Support Vector Machine was put forth by Vapnik [4] in the 1990s. Using labelled training data as a basis, the SVM methodology attempts to calculate the optimal separating hyperplane between the two classes of data under consideration.

The classifier is an algorithm that performs the actual recognition on the basis of the feature vectors extracted from the Mel Frequency Cepstral Coefficients (MFCCs). Consequently, classifications together with feature extraction are essential stages of the recognition process that have the largest influence on the performance and robustness of the system.

In this paper we use the MFCCs as feature vectors that extracted from various modulated signals and SVM is used to classify the feature vectors that extracted from the received signals to determine which demodulation scheme is used to recover the original signal. This method is examined on different SNR and on multipath channels with equalization. The performance is determined by the recognition rate and the CPU time. And we compared the results with the method that uses the ANN as classifier to determine the best methods for Automatic Digital Modulation Recognition (ADMR).

Proposed ADMR System

In the proposed ADMR system signal identification involves three stages; feature extraction to represent the signal characteristics, modeling of the signal features, and decision making to complete the identification task. The main task in a signal identification system is to extract features capable of representing the information present in the modulated signal. Once a proper set of feature vectors is obtained, the next task is to develop a model for each modulated signal. Feature vectors representing the modulated signal characteristics are extracted and used for building the reference models. The final stage is the decision making to either accept or reject the claim of the signal. The decision is made based on the result of the matching technique used.

The signal identification process consists of two modes; a training mode and recognition or testing mode as shown in figure 1 [5]. In the training mode, a new modulated signal with known identity is enrolled into the system database. In the recognition mode, an unknown signal is inputted into the system and a decision is made about the signal identity.

Both the training and the recognition modes include a feature...
The IDFT is defined as:

\[
s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{j2\pi nk/N}, 0 \leq n \leq N-1
\]  

The Mel filter bank

In the MFCCs method, the main advantage is that it uses Mel-frequency scaling, which approximates quite well the human auditory system. The Mel scale is defined as [6]:

\[
Mel(f) = \frac{2595 \log(1 + \frac{f}{700})}{\pi - 0.5}
\]  

Where Mel is the Mel-frequency scale and \( f \) is the frequency on the linear frequency scale.

The DCT

The final stage involves performing a DCT on the log of the Mel spectrum. If the output of the \( m \)th Mel filter is \( S(m) \), then the MFCCs are given as [9]:

\[
c_g = \sqrt{\frac{2}{N}} \sum_{m=0}^{N-1} \log(S(m)) \cos \left( \frac{g\pi (m - 0.5)}{N_f} \right)
\]  

Where \( g = 0, 1, \ldots, G-1 \), \( G \) is the number of MFCCs, \( N \) is the number of Mel filters and \( c_g \) is the \( g \)th MFCC. The number of the resulting MFCCs is chosen between 12 and 20, since most of the signal information is represented by the first few coefficients. The 0th coefficient represents the mean value of the input signal.

Discrete transforms can be used for extraction of robust MFCCs in modulation identification systems. The Discrete Wavelet Transform (DWT), the Discrete Cosine Transform (DCT), and the Discrete Sine Transform (DST) have been investigated in the literature for this purpose [10]. Figure 3 illustrates the utilization of discrete transforms in modulation identification systems.

Classification

Classification using ANN

Artificial Neural Networks (ANNs) are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. An ANN usually
We describe the basic theory of the SVM first, and then present the recognition using SVMs. Since the SVM is basically a binary classifier, been applied widely and it has been reported that SVM can perform very well. The actual weight update follows a very simple gradient descent algorithm.

The weight values are updated by a simple gradient descent algorithm. There are a number of efficient training algorithms for ANNs. Among the most famous is the back-propagation algorithm. An alternative is the back-propagation algorithm with momentum and learning rate to speed up the training. The training algorithm is mostly defined by the learning rule, that is, the weights update in each training epoch. There are a number of efficient training algorithms for ANNs. Among the most famous is the back-propagation algorithm. An alternative is the back-propagation algorithm with momentum and learning rate to speed up the training.

The training rate, \( \epsilon \), scales the derivative and it has a great influence on the training speed.

In this paper, we consider the resilient back-propagation algorithm [12]. This algorithm performs a direct adaptation of the weight update based on local gradient information.

The actual weight update follows a very simple rule as shown in the following equations:

\[
\Delta w\text{i}j(t) = \begin{cases} 
-\Delta ij(t), & \text{if } \frac{\delta E}{\delta w\text{i}j}(t) > 0 \\
+\Delta ij(t), & \text{if } \frac{\delta E}{\delta w\text{i}j}(t) < 0 \\
\Delta ij(t-1), & \text{else}
\end{cases}
\] (7)

\( w\text{i}j(t + 1) = w\text{i}j(t) + \Delta w\text{i}j(t) \) (8)

**Classification using SVM**

SVM is a powerful method for pattern recognition. The method has been applied widely and it has been reported that SVM can perform quite well in many pattern recognition problems [13].

In this paper, we study the method for multiclass pattern recognition using SVMs. Since the SVM is basically a binary classifier, it is not straightforward to apply it to multiclass recognition problem. We describe the basic theory of the SVM first, and then present the multiclass pattern recognition method based on SVM.

The Support Vector Machine formulation for pattern classification will be described. SVMs were initially developed to perform binary classification; though, applications of binary classification are very limited. Most of the practical applications involve multiclass classification, especially in remote sensing land cover classification. A number of methods have been proposed to implement SVMs to produce multiclass classification. A number of methods to generate multiclass SVMs from binary SVMs have been proposed by researchers and is still a continuing research topic. SVM are based on statistical learning theory and have the aim of determining the location of decision boundaries that provide the optimal separation of classes. In the case of a two-class pattern recognition problem in which the classes are linearly separate the SVM selects from among the infinite number of linear decision boundaries the one that minimizes the generalization error. Thus, the selected decision boundary will be one that leaves the greatest margin between the two classes, where margin is defined as the sum of the distances to the hyperplane from the closest points of the two classes. If the two classes are not linearly separate, the SVM tries to find the hyperplane that maximizes the margin while, at the same time, minimizing a quantity proportional to the number of misclassification errors.

SVM were initially designed for binary (two-class) problems. When dealing with multiple classes, an appropriate multiclass method is needed. Vapnik [4] suggested comparing one class with the others taken together.

**SVM for binary classification:** The traditional learning method uses the empirical risk minimization principle, while the statistical learning theory uses the structural risk minimization principle. A learning problem is often defined as follows: suppose a probability measure \( F(z) \) in the space \( Z \), consider the function collection \( Q(z, \alpha) \), \( \alpha \in \Lambda \) is parameters set. The goal of learning is to minimize the risk function [14]:

\[
R(\alpha) = \int Q(z, \alpha) dF(z), \alpha \in \Lambda \quad (9)
\]

where the probability measure \( F(z) \) is unknown, but a certain number of independence samples with the same distribution are given:

\[
Z_1, Z_2, ..., Z_k \quad (10)
\]

The goal is to minimize the risk function (5) on the basis of empirical data (6), where \( Q(z, \alpha) \) represents a certain loss function. In
order to minimize risk function (5) under the unknown distribution function \( F(z) \), the following principle is adopted: according to the large number theorem in probability theory, we use arithmetic average as the substitute of the expectation risk \( R(\alpha) \).

\[
R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^{l} Q(z, \alpha)
\]

where the first part is called empirical risk of the training samples, the other called confidence interval. The confidence interval is the function of the sets \((Q(z, \alpha))\) VC-dimension \( h \) and the number of the training samples, which monotonic decreases with \( n/h \) increasing. The method that implements the Structural Risk Minimization is the Support Vector Machine, and the following is the idea of SVM. In the linear SVM by the kernel function \( K(x, y) \), the parameters that minimize the risk function \( R(\alpha) \), as is known as the Empirical Risk Minimization (ERM). It has been proved by Swami and Sadler [15] that the minimum empirical risk does not mean the minimum expectation risk. The theorem about the generalizing boundary in statistical learning theory tells us that under the empirical risk minimization principle, the actual risk of learning machines is composed of two parts:

\[
R(\alpha) \leq R_{\text{emp}}(\alpha) + \Phi\left(\frac{n}{h}\right)
\]

maps the data from the input space \( \mathbb{R}^N \) to a high dimensional feature space \( H \) by \( x \rightarrow \psi(x) \subset H \), such that the mapped data points of the two classes are linearly separable in the feature space. Assuming there exist a kernel function \( K \) such that \( K(x, y) = \psi(x) \cdot \psi(y) \), then a nonlinear SVM can be constructed by replacing the inner product \( x \cdot y \) in the linear SVM by the kernel function \( K(x, y) \), the corresponding classification function is:

\[
f(x) = \text{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^* \right\}
\]

This corresponds to constructing an optimal separating hyperplane in the feature space [16].

**SVM for multiclass classification:** Originally, SVMs were developed to perform binary classification. However, applications of binary classification are very limited especially in remote sensing land cover classification where most of the classification problems involve

### Table 1: Common kernel functions for SVM.

<table>
<thead>
<tr>
<th>Function ( K(x, y) )</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear ( x \cdot y )</td>
<td>none</td>
</tr>
<tr>
<td>Polynomial ( (\gamma \cdot xy + b)^n )</td>
<td>( \gamma, b, n )</td>
</tr>
<tr>
<td>Radial basis function ( e^{-\gamma |x-y|^2} )</td>
<td>( \gamma )</td>
</tr>
<tr>
<td>Sigmoid ( \tanh(\gamma \cdot xy + b) )</td>
<td>( \gamma, b )</td>
</tr>
<tr>
<td>Chi-square ( e^{-\gamma x^2} )</td>
<td>( \gamma )</td>
</tr>
<tr>
<td>Gaussian ( e^{-\gamma |x-a-y|^2} )</td>
<td>( \gamma, a, b )</td>
</tr>
</tbody>
</table>

### Table 2: System specifications.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol rate</td>
<td>2500000 symbol/sec</td>
</tr>
<tr>
<td>FTT length</td>
<td>256</td>
</tr>
<tr>
<td>No. of carrier</td>
<td>128</td>
</tr>
<tr>
<td>No. of parallel channel</td>
<td>256</td>
</tr>
<tr>
<td>No. of data</td>
<td>6</td>
</tr>
<tr>
<td>Length of guard interval</td>
<td>32</td>
</tr>
<tr>
<td>Modulation types</td>
<td>QAM, FSK, PSK, MSK</td>
</tr>
</tbody>
</table>
more than two classes. This section provides a brief description of some methods implemented to solve multiclass classification problem with SVM in present study [17].

One-against-all: The one-against-all method employs \( c \) classifiers. The \( i \)-th classifier is trained to discriminate between the class \( w_i \) and all the other classes. During the test phase, the sample is classified using all classifiers and the final decision is made on the basis of the values of the discriminant functions as follows:

\[
\text{max} \{ \text{arg} \sum_{i=1}^{c} \frac{1}{\gamma} \} \quad (19)
\]

One-against-one: In the case of the one-against-one approach, \( c(c-1)/2 \) classifiers are trained to discriminate between each pair of classes. In order to make the final decision, each classifier votes on one class depending on the sign of its discriminant function. Consequently, the class which collects the highest number of votes is selected.

The kernel is a distance measure in the feature space. For SVM purposes some common kernels are listed in table 1 [17]. For all experiments in this paper we have found that the polynomial kernel function is the best function compared to the others. The \( \gamma \)-selection could be optimized by fitting a quadratic function on the recognition rates and calculating the \( \gamma_{\max} \) corresponding to the empirical

<table>
<thead>
<tr>
<th>SNR(dB)</th>
<th>Features from signal</th>
<th>Features from DWT</th>
<th>Features from signal plus DWT</th>
<th>Features from DCT</th>
<th>Features from signal plus DCT</th>
<th>Features from DST</th>
<th>Features from signal plus DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>73.75</td>
<td>86.33</td>
<td>75.58</td>
<td>88.83</td>
<td>95.5</td>
<td>89.58</td>
<td>85.41</td>
</tr>
<tr>
<td>-5dB</td>
<td>83.33</td>
<td>95.25</td>
<td>87.91</td>
<td>98</td>
<td>99.25</td>
<td>98.16</td>
<td>96.91</td>
</tr>
<tr>
<td>0</td>
<td>90.08</td>
<td>99.66</td>
<td>98.58</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.83</td>
</tr>
<tr>
<td>5dB</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3: Recognition rate vs. SNR (dB) for OFDM with AWGN using ANN.

<table>
<thead>
<tr>
<th>SNR(dB)</th>
<th>Features from signal</th>
<th>Features from DWT</th>
<th>Features from signal plus DWT</th>
<th>Features from DCT</th>
<th>Features from signal plus DCT</th>
<th>Features from DST</th>
<th>Features from signal plus DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>80.41</td>
<td>72.08</td>
<td>74.25</td>
<td>80.08</td>
<td>80.83</td>
<td>85</td>
<td>84.91</td>
</tr>
<tr>
<td>-5dB</td>
<td>92.66</td>
<td>81.58</td>
<td>79.25</td>
<td>89.75</td>
<td>88.33</td>
<td>94.25</td>
<td>93.08</td>
</tr>
<tr>
<td>0</td>
<td>98.75</td>
<td>89.33</td>
<td>93.16</td>
<td>98.91</td>
<td>97.41</td>
<td>99.58</td>
<td>99.16</td>
</tr>
<tr>
<td>5dB</td>
<td>100</td>
<td>99.58</td>
<td>99.5</td>
<td>100</td>
<td>99.91</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4: Recognition rate vs. SNR (dB) for OFDM with LMMSE using ANN.

<table>
<thead>
<tr>
<th>SNR(dB)</th>
<th>Features from signal</th>
<th>Features from DWT</th>
<th>Features from signal plus DWT</th>
<th>Features from DCT</th>
<th>Features from signal plus DCT</th>
<th>Features from DST</th>
<th>Features from signal plus DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>99.5</td>
<td>98.83</td>
<td>99.33</td>
<td>98.83</td>
<td>99.33</td>
<td>99.66</td>
<td>99.16</td>
</tr>
<tr>
<td>-5</td>
<td>99.33</td>
<td>99.66</td>
<td>99.33</td>
<td>98.83</td>
<td>99.33</td>
<td>98.33</td>
<td>99.33</td>
</tr>
<tr>
<td>0</td>
<td>99.33</td>
<td>99.33</td>
<td>99.33</td>
<td>100</td>
<td>99.66</td>
<td>98.66</td>
<td>98.66</td>
</tr>
<tr>
<td>5</td>
<td>99</td>
<td>98.83</td>
<td>99.16</td>
<td>99.33</td>
<td>99</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>10</td>
<td>98.5</td>
<td>99.33</td>
<td>98.83</td>
<td>100</td>
<td>99.16</td>
<td>100</td>
<td>99.16</td>
</tr>
<tr>
<td>15</td>
<td>98.5</td>
<td>99.33</td>
<td>99.33</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5: Recognition rate vs. SNR (dB) for OFDM with ZF using ANN.

<table>
<thead>
<tr>
<th>SNR(dB)</th>
<th>Features from signal</th>
<th>Features from DWT</th>
<th>Features from signal plus DWT</th>
<th>Features from DCT</th>
<th>Features from signal plus DCT</th>
<th>Features from DST</th>
<th>Features from signal plus DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>99.5</td>
<td>98.83</td>
<td>99.33</td>
<td>98.83</td>
<td>99.33</td>
<td>99.66</td>
<td>99.16</td>
</tr>
<tr>
<td>-5</td>
<td>99.33</td>
<td>99.66</td>
<td>99.33</td>
<td>98.83</td>
<td>99.33</td>
<td>98.33</td>
<td>99.33</td>
</tr>
<tr>
<td>0</td>
<td>99.33</td>
<td>99.33</td>
<td>99.33</td>
<td>100</td>
<td>99.66</td>
<td>98.66</td>
<td>98.66</td>
</tr>
<tr>
<td>5</td>
<td>99</td>
<td>98.83</td>
<td>99.16</td>
<td>99.33</td>
<td>99</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>10</td>
<td>98.5</td>
<td>99.33</td>
<td>98.83</td>
<td>100</td>
<td>99.16</td>
<td>100</td>
<td>99.16</td>
</tr>
<tr>
<td>15</td>
<td>98.5</td>
<td>99.33</td>
<td>99.33</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 6: Recognition rate vs. SNR (dB) in OFDM with AWGN Using SVM.
Results and Discussion

In the training phase of the modulation identification system, a database is first composed for 6 modulation schemes. To generate this database, 6 modulated signals are used to generate MFCCs to form the feature vectors of the database. These features are used to train a SVM. In the testing phase, each one of these modulated signals is transmitted. Similar features to those used in the training are extracted from the noisy modulated signals and used for matching.

The features used in all experiments are 13 MFCCs forming a feature vector of the modulated signal. Seven methods for extracting features are adopted in the experiments. These methods are illustrated follows:

1. The MFCCs are extracted from the modulated signals, only.
2. The features are extracted from the DWT of the modulated signals.
3. The features are extracted from both the original modulated signals and the DWT of these signals and concatenated together.
4. The features are extracted from the DCT of the modulated signals.
5. The features are extracted from both the original modulated signals and the DCT of these signals and concatenated together.
6. The features are extracted from the DST of the modulated signals.
7. The features are extracted from both the original modulated signals and the DST of these signals and concatenated together.

The recognition rate is used as the performance evaluation metric in all experiments. It is defined as the ratio of the number of success identifications to the total number of identification trials.

Recognition rate =
The number of success identifications \* 100
The total number of identification trials

(20)

The proposed algorithm in this paper has been verified and validated for various orders of digital modulation types including PSK, MSK, FSK, and QAM. Table 2 illustrates the system specifications. The performance of the classifier has been examined for 100 realizations of each modulation type/order, and the results are presented in tables 6-9 for the different feature extraction methods. If we compare these results with results in tables 3,4 and 5 as in [1] it is clear that the performance of SVM achieves high recognition rate at low signal to noise ratio and CPU time is less than ANN using machine with specifications (CPU 2.53 GHz PC running with Matlab 7.1).

Conclusion

This paper has presented a suggested approach for automatic modulation recognition based on using MFCCs for feature extraction and SVM is used as a classifier. Simulation results illustrated that the modulation type and order can be determined by extracting cepstral features from the signals and from their transforms such as DWT, DCT and DST, and classifying these features. Signal classification from the SVM has achieved minimum CPU time and high recognition rate compared with ANN classifier.

References