An SNR Unaware Large Margin Automatic Modulations Classifier in Variable SNR Environments

Hamidreza Hosseinzadeh and Farbod Razzazi
Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Iran

Abstract: *Automatic classification of modulation type in detected signals is an intermediate step between signal detection and demodulation, and is also an essential task for an intelligent receiver in various civil and military applications. In this paper, a new two-stage partially supervised classification method is proposed for Additive White Gaussian Noise (AWGN) channels with unknown signal to noise ratios, in which a system adaptation to the environment Signal-to-Noise Ratios (SNR) and signals classification are combined. System adaptation to the environment SNR enables us to construct a blind classifier to the SNR. In the classification phase of this algorithm, a passive-aggressive online learning algorithm is applied to identify the modulation type of input signals. Simulation results show that the accuracy of the proposed algorithm approaches to a well-trained system in the target SNR, even in low SNRs.*

Keywords: *Automatic modulation classification, pattern recognition, partially supervised classification,passive-aggressive classifier, SNR un-aware classification.*

Received January 27, 2015; accepted March 9, 2014

1. Introduction

Automatic Modulation Classification (AMC) is an important issue in cognitive telecommunication signal processing and its related fields, including blind investigating the frequency spectrum for civil and military applications. Today, telecommunication signals are propagated using various modulation types and in different frequency bands. In many commercial and military applications (e.g., commercial communication systems, spectrum monitoring, signal surveillance, interferer identification, universal demodulation and software defined radio), it is essential to detect and demodulate the signal without given priori information on that signal. This is in contrast to regular receiver's functionalities in which the prior information of the received signals, such as carrier frequencies, frequency bandwidth, bit rate, and modulation type is available. In other words, the receiver is blind. Furthermore, online techniques are required for real-time signal interception and processing, which are vital for decisions involving electronic warfare operations and other tactical actions.

Generally, designing a modulation classifier substantially consists of two steps; pre-processing of the received signal and choosing a proper classification algorithm. The pre-processing task might include noise reduction, carrier frequency estimation, and signal to noise ratio estimation. Depending on the classification algorithm in the second step, the pre-processing tasks should be provided. The focus of this study is on the second step.

In the classification step, the employed algorithms may be divided into two major categories: Decision Theory Based Approaches (DTBAs) and Feature Matching Based Approaches (FMBAs) [7, 11, 18, 22]. DTBA algorithms are based on the received signal likelihood function. In this category, the decision is made by comparing the likelihood function with a threshold. The complexity of DTBA is usually high. In addition, this classifier does not perform well when the sampled signal has random phase offsets, timing offsets, and timing jitters. In contrast, FMBA usually extracts one or more features form the received signal and the decision is performed based on their measured values by using a trainable classification algorithms. The calculation complexity of FMBA is often lower than DTBA. Our proposed algorithm may be categorized as an FMBA algorithm.

Feature matching based systems can be further divided into two main subsystems: the feature extraction subsystem and the classifier subsystem. Feature extraction subsystem, extracts the features from the received signal, and the classifier subsystem, determine the membership of the received signal to each class. This study proposes a blind large margin classifier using classic state of the art features. Therefore, the contribution of the paper is concentrated on the classifier subsystem. Large margin classifiers are a category of classifiers which minimizes the empirical risk of misclassification and shows the state of the art results in pattern classification problems [5, 21].

From the published works in AMC [1, 13, 15, 19, 20, 23], it appears clear that one of the main drawbacks with these studies is the Signal-to-Noise Ratios (SNR) aware assumption of the classifier in the most learning methods. In otherwords, the training and testing procedure are performed on equal SNRs which are far away from realistic conditions. Sengur [19] is an exception for this deduction whose results is not glancing for low Signal-to-Noise Ratios (SNRs). In fact, since the optimistic objective of AMC is blind detection of the modulated signals, the receiver should have no pre-knowledge on the sent signal's SNR. Consequently, by now, the proposed algorithms could not be considered as blind classifiers. Therefore, proposing a new method that could automatically match with the SNR of environment and has lower computational complexity is crucial.

In this paper, we propose to employ the idea of passive-aggressive algorithm [9] as an online large margin classifier to present novel partially supervised classification architecture for Additive White Gaussian Noise (AWGN) channels with unknown or variable signal to noise ratios. In this paper, it is assumed that signal to noise ratio is unknown. However, the proposed idea is well applicable to variable SNR case with some generalizations in mathematical formulation and architecture.

The rest of the paper is organized as follows: in section 2, we present the proposed architecture and discuss our algorithm. This section also reviews the online passive-aggressive learning algorithms as the basis of the proposed architecture. Section 3 provides a description of the experimental setup and simulation results. Finally, section 4 concludes the paper and comments on how this algorithm can be further expanded.

2. Proposed Classifier Architecture

Large margin classifiers are assumed as state of the art classifiers. The most commonly used large margin classifiers are support vector machines [8]. A support vector machine is a batch learning algorithm that finds out the separating hyper-plane with the maximum possible margin. Our Approach is constructed based on online version of support vector machine. Online algorithms, such as passive-aggressive, can be used to adapt the classifier system within terms of input signals. In addition, they can be adapted to have a partially supervised classifier. Extensions of batch learning algorithms to online settings have practically proven to be successful in many civil and military applications. In this section, the mathematical formulation of classic offline and online large margin classifiers are briefly discussed and then the proposed idea is presented both in architecture and mathematical analysis aspects.

2.1. Large Margin Classifier

2.1.1. Support Vector Machine

Suppose to have a binary classification problem and training set of 1 labeledsamples $\{x_i, y_i\}_{i=1}^{l}$ where $x_i \in \Re^d$ is an input vector describing i^{th} samples and $y_i \in \{-1,1\}$ is its labels. We attempt to learn the function $f(x) = w.\varphi(x) + b$ which assigns the correct label to an unseen test sample. $\varphi(x)$ is a feature mapping induced by the kernel $k(x,x') = \varphi(x).\varphi(x')$, and b is the bias.

The goal is to find parameters (w,b) suchthat the following equation is minimized,

$$\min_{w,b} \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \quad (1)$$
$$s.t. \; y_i\left[w^T\varphi(x)+b\right] \geq 1 - \xi_i \,,\; \xi_i \geq 0 \;\; \forall i = 1,\ldots,l$$

Where the components of vector ξ are slack variables, and C is a penalty parameter.

2.1.2. Online Passive-Aggressive Classifier

Passive-Aggressive (PA) algorithms [9] are a family of maximum margin based, online learning algorithms that update the model parameters by each observation. Online PA algorithm, on one hand, modifies the current classifier $w_t \times \varphi(x) + b$ in order to correctly classify the current example X_i, by updating the weight vector from w_t to w_{t+1}; and on the other hand, the new classifier $w_{t+1} \times \varphi(x) + b$ should be as close as possible to the current classifier $w_t \times \varphi(x) + b$.

Our classifier architecture for modulations classification was based on PA algorithm, and we pursued both of above ideas at the same time. The vector w_1 is initialized to (0,…,0). In round t, the new weight vector w_{t+1} was determined by solving the following optimization problem,

$$w_{t+1} = \min_{w \in R^n} \tfrac{1}{2}\|w - w_t\|^2 + C\xi \quad (2)$$
$$s.t. \; \ell_t = \max\{0, 1 - y_t(w_t \times x_t)\} \leq \xi, \xi \geq 0$$

Where C is a positive parameter which controls the influence of the slack term ξ on the objective function. ℓ_t is a loss suffered by the algorithm on round t. In order to confidently make a prediction, ℓ_t claims that the margin should be larger than 1 or otherwise, the algorithm should suffer a positive loss.

We can see that there is no need to update the weight vectors whenever $\ell_t = 0$, that is, the resulting algorithm is passive. Otherwise, the algorithm aggressively forces the new weight vector w_{t+1} to satisfy the constraint $\ell_t = 0$. Therefore the algorithm has been named as passive-aggressive.The solution to the optimization problem in the Equation 2 can be rewritten as a simple closed form solution,

$$w_{t+1} = w_t + \mu_t y_t x_t \; ; \; \mu_t = min\{ C, \frac{\ell_t}{\|x_t\|^2} \} \tag{3}$$

Where parameter μ_t is called the learning rate and regulates the smoothness and the speed of convergence. This update is derived using standard convex analysis tools e.g., [4].

In the case of multiclass classification, the algorithm prediction is a vector in R^K where each element in the vector corresponds to a score which is assigned to thescore calculations, have been devised in [10]. The prediction of the PA algorithm is set to the label with the highest score.

2.2. Proposed Classifier

In this section, we introduce a partially supervised algorithm that is proposed in this paper for AMC application. The reason why we use the word "partially supervised" here is that the algorithm is capable to classify the input samples correctly, using a reasonable set of correctly recognized samples. In this study, it is assumed that the signal to noise ratio is unknown. Thus, the development of the self-adaptation of classifier with the SNR of environment is a problem demanding attention. In addition, it is assumed that the label of the all training samples is not determined.

It is worth mentioning that the novelty of this strategy mainly lies in the idea of the unique combination of the two stages; system adaptation and input signal classification. To illustrate the novelty of the proposed method, we hereinafter present theoverall block diagram of this classifier as shown in Figure 1.

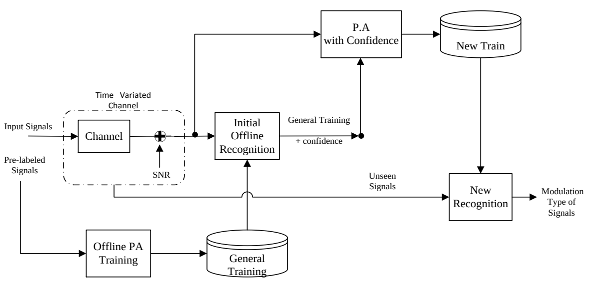

Figure 1. The overall block diagram of presented classifier.

The proposed classifier is conducted by the following steps:

- *Step* 0. Initial Training a general classifier was previously trained using a set of pre-labeled signals, which is to be expected have different SNRs. The results are stored in 'the general training' block.
- *Step* 1. Adaptation the input signals were introduced to the system to classify as well as its confidence. In this step, we consider a confidence measure to collect the well-recognized samples and to use them to adapt the new classifier through a self-training approach. The obtained results from the simulation show that a simple thresholding on the classifier output value could not be used for this purpose. The proposed procedure to extract a reliable confidence measure is discussed in the next subsection. The system is adapted to the high-confidence unlabeled signals to have a new adapted trained classification system. It should be noted that the weight vector of confident samples for new training did not start from the clean signal initial point, but started from the obtained weight vector of the general training.
- *Step* 2. Classification In the classification phase, this new adapted classifier is used to classify the unknown input signals.

2.3. Proposed Confidence Measure

In a partially supervised large margin algorithm, the system may miss the accuracy or even miss the convergence unless the new labeled data would be reliable. In large margin classifier, samples with greater distances from the decision boundary are assumed to be more reliable and have been called as confident samples.

Therefore, a class dependent threshold has been determined as the minimum distance of reliable samples for each class. This threshold is determined using the training input samples without prior knowledge of the label of this sample or the SNR of environment. However, the experiments show that this threshold varies in different signal to noise ratios, which is reported in the results section which is obviously depended on the predicted class. The histogram of distance to decision boundary (discriminant value) as the confidence measure for different classes in SNR= 4 dB are showed in Figure 2 which is obviously depended on the predicted class.

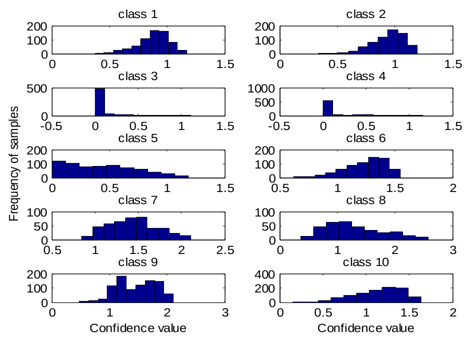

Figure 2. The histogram of distance to decision boundary for different classes (horizontal axis and vertical axis represents the confidence value and frequency of samples respectively).

Therefore, a constant threshold on discriminant values could not be used for separating confident samples in different predicted classes.

The proposed confidence measure is calculated as follows: the histogram of distance to classification boundary for predicted training samples is extracted in each predicted class separately and the threshold was set to select 90% in the collected samples.

To select a more reliable subset of this 90% samples set, we collected the samples that have positive discriminant value for the predicted class and negative discriminant value for all other classes. These final collected samples were chosen as confident samples. Consequently, according to the confidence measure extraction procedure, a confidence threshold is determined for each predicted class.

3. Experiments and Results

3.1. Evaluation Benchmarks

According to emerging development of digital systems and the trend towards digital telecommunications instead of analog telecommunication, digital signals are mostly put to use today. Considering the changes in message parameters, there are four general digital signal types, M-ary Amplitude Shift Keying (M-ASK), M-ary Phase Shift Keying (M-PSK), M-ary Frequency Shift Keying (M-FSK) and M-ary Quadrature Amplitude Modulation (M-QAM) [16]. In this study, the set of input digital signal types was considered as follows: 2FSK, 4FSK, 2ASK, 4ASK, 2PSK, 4PSK, 8PSK, 16QAM, 32QAM, and 64QAM. To simplify the notation, these signals were substituted with S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9 and S_{10}, respectively.

The carrier frequency (f_c) was assumed to be 150 kHz. The sampling rate (f_s)was 1200 kHz. The symbol rate (r_s) was 12.5 kHz and the number of samples in a symbol sequence was 4096.

The RBF kernel function of the form $k(x,x') = exp(-\gamma\|x - x'\|^2)$ was employed for this experiment as the kernel function of PA classifier. A grid search technique was used to find the optimal values of kernel parameters. In practice, the standard method to determine the optimum value for RBF kernel parameter, γ , and misclassification penalty parameter, C, is through grid search method [5]. We used a 3 fold cross-validation approach to evaluate the classifier generalization performance and to divide these generated dataset into training and testing dataset. The performance of the algorithms was compared on the basis of the classification accuracy. Classification accuracy assessments of different classes were provided by the confusion matrix and accuracy analysis of different classes in percentage. The value of learning rate for PA classifier was set to 0.01. By choosing this learning rate value, the system was able to learn well and the algorithm execution time for a real-time system would be acceptable. Here we assumed that the carrier frequency has previously been correctly estimated or it was known. Therefore, we considered complex baseband signals. In addition, it was assumed that the simulated signals were bandwidth limited. The Gaussian white noise was added according to SNRs in 0 dB, 4 dB, 8dB, and 12 dB. Each signal type has 100 realizations which were generated randomly for each trial to ensure independent results.

3.2. Feature Extraction

Since the choice of features highly affects the performance of the classifier, feature extraction is the determinant part of a pattern recognition system where its aim is to reveal the distinctive properties of an object to be recognized. In this study, a suitable set of features was considered as a combination of high order statistics and instantaneous characteristics of digital signal types. The rest of this section, briefly describe these features.

3.2.1. Instantaneous Feature

Instantaneous features are suitable for signals which contain instantaneous phase or instantaneous frequency [14]. In this work, the instantaneous features for classification were selected from the proposed features by Azzouz and Nandi [2, 3]. These features were derived from the instantaneous properties of the received signals. Therefore, these features are called as instantaneous features. The instantaneous key features which were used for the proposed tracking algorithm were derived from the instantaneous amplitude $a(t)$, and the instantaneous frequency $f(t)$, of the signal under consideration.

The first feature is the maximum value of the power spectral density of the normalized-centered instantaneous amplitude of the intercepted signal which is formulated as follows:

$$\gamma_{max} = max\left(\frac{|DFT(a_{cn}(i))|^2}{N_s}\right) \tag{4}$$

Where $N_s$is the number of the sample in the range and $a_{cn}(i)$ is value of centralized normalized instantaneous amplitude that is defined by

$$a_{cn}(i) \overset{\Delta}{=} a_n(i) - 1, a_n(i) \overset{\Delta}{=} \frac{a(i)}{m_a} \tag{5}$$

And m_a is the average value of instantaneous amplitude over one frame, i.e.,

$$m_a = \frac{1}{N_s}\sum_{i=1}^{N_s} a(i) \tag{6}$$

This feature is designed to discriminative between Constant Envelopes (CE) signals (e.g., FSK and PSK) and non-CE signals (e.g., ASK).

The second feature is the standard deviation of absolute value of normalized-centered instantaneous frequency over non-weak segments of the intercepted signal which is calculated as:

$$\sigma_{af} \overset{\Delta}{=} \sqrt{\frac{1}{L}\left[\sum_{a_{cn}(i)>a_t} f_{cn}^2(i)\right] - \left[\frac{1}{L}\sum_{a_{cn}(i)>a_t} |f_{cn}(i)|\right]^2} \tag{7}$$

Where$f_{cn}(i)$ is the centralized normalized instantaneous frequency and it is defined by:

$$f_{cn}(i) = \frac{f_c(i)}{r_b},\ f_c(i) = f(i) - m_f,\ m_f = \frac{1}{N}\sum_{i=1}^{N} f(i) \quad (8)$$

Where$r_b$is the bit rate, and $a_t$is a preset threshold for detecting non-weak samples because instantaneous frequency is very noise sensitive. In this paper, the threshold for detection of non-weak samples is chosen asa_t=0.95 [2].

3.2.2. Higher Order Statistics

The first set of employed statistical features is moments. A moment of a random variable may be defined as:

$$M_{p,q} = E\left[s^{p-q}\left(s^*\right)^q\right] \quad (9)$$

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

The second set of employed statistical features is cumulants which is the most widely used feature in this area. The symbolism for p^{th} order cumulants is similar to that of the pth order moment.

$$C_{p,q} = Cum\left[s,...,s,s^*,...,s^*\right] \quad (10)$$

The mentioned expression has *(p-q)* terms of *s*, in addition to *q* terms of s^*. Cumulants may be expressed in term of moments as

$$Cum\left[s_1,...,s_n\right] = \sum_{\forall v}(-1)^{q-1}(q-1)!\,E\left[\prod_{i\in v_1} s_i\right]...E\left[\prod_{i\in v_q} s_i\right] \quad (11)$$

Where the summation index is over all partitions $v=(v_1,...,v_q)$ for the set of indices $(1,...,n)$, and q is the number of elements in a given partition.

Based on Fisher Discriminant Analysis (FDA) [6, 17], we selected a proper set of higher order moment and cumulants as below. FDA represents the capability of the selected features for separation of two predefined classes and is defined by

$$d_{ij} = \frac{(\mu_i - \mu_j)_2}{\sigma_1{}^2 + \sigma_1{}^2}\ i \neq j \quad (12)$$

Where μ and σ are mean and variance of these two classes.The important selected statistical features are M_{41}, M_{61}, M_{84}, C_{40}, C_{61}, C_{63}, C_{80}, C_{82}, and M_{84}. Unfortunately, these characteristics are noise dependent. Therefore, a strategy must be devised to decrease the effect of this dependency, as far as possible. The proposed classifier perfectly solves this problem. Figure 3 shows the variation of higher order statistics in different SNRs, and for the selected modulations set.

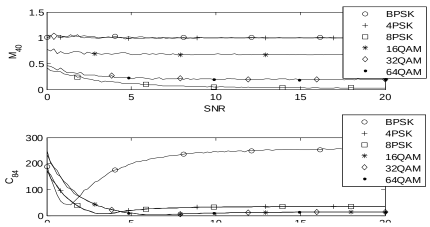

Figure 3. Variation of statistical featuresin different SNR.

3.3. Performance Evaluation of Proposed Classifier

In this section, we evaluated the performance of the proposed classifier at different SNRs. The classifier training was done on a subset of 100 symbols, out of the total 1000 symbols dataset.

The classification rates in percentage for the proposed adaptable PA Classifier in 0, 4, 8 and 12 dB SNRs is shown in Table 1. From the mentioned results in Table 1, it can be deduced that the performance of the classifier in different SNRs are generally good. Because the samples that were used to train the classifier had a high confidence. Of course, the performance is slightly degraded in lower SNRs. This indicates that these features may not be able to tolerate high noise.

Table 1. Classification rate of proposed classifier for each class in different SNR.

Modulation classes	**SNR**			
	0 dB	**4 dB**	**8 dB**	**12 dB**
S_1	99.9	100	100	100
S_2	99.8	100	100	100
S_3	92.2	96.1	100	100
S_4	92.5	96.7	100	100
S_5	100	100	100	100
S_6	97.4	100	100	100
S_7	87.6	67.4	100	100
S_8	35.2	64.3	88.6	100
S_9	68.4	99.4	100	100
S_{10}	40.6	96.1	94.2	100
Mean	81.36	92.00	98.28	100

As a sample, the confusion matrix was extracted at SNR=4 dB to analyse in the confusion of different classes. These results are presented in Table 2.

Table 2. Confusion matrix of proposed algorithm in SNR=4 dB (%).

True modulations	**Predicted modulations**									
	S_1	**S_2**	**S_3**	**S_4**	**S_5**	**S_6**	**S_7**	**S_8**	**S_9**	**S_{10}**
S_1	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S_2	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S_3	0.0	0.0	96.1	3.9	0.0	0.0	0.0	0.0	0.0	0.0
S_4	0.0	0.0	3.3	96.7	0.0	0.0	0.0	0.0	0.0	0.0
S_5	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0
S_6	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0
S_7	0.0	0.0	0.0	0.0	0.0	0.0	67.4	0.0	32.6	0.0
S_8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	64.3	0.3	35.4
S_9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.6
S_{10}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.9	96.1

As it is observed in Table 2, the recognition accuracy for all classes except class 7 and 8 was good even at low SNR. It can be seen that there is a tendency for S_7 modulation to be mostly confused with S_9 modulation, S_8 modulation withS_{10} modulation. This issue may be explained by the fact that the constellation shapes of these classes as very close together and there is a tendency for S_7 modulation to be mostly confused with S_8 modulation. Consequently, the selectedfeatures to discriminate these two classes in low SNRs do not perform well. In addition, the considered statistical features for S_7, S_8, S_9 and S_{10} are very close and heavily noise dependent. Therefore, the recognition accuracies of these modulated signals in low SNRs were not good.

As it is mentioned in section 2.2.a confidence threshold should be determined to collect well confident samples in each class based on the explained method in section 2.3. The extracted values of confidence thresholds for each class in different SNRs are presented in Table 3. As it is observed in Table 3, the confidence threshold is very low level for SNR=12 dB. In this case, all of the samples are assumed as high confident and there is no need to apply the confidence threshold.

Table 3. Confidence threshold for each class in different SNRs.

Modulation classes	**SNR**			
	0 dB	**4 dB**	**8 dB**	**12 dB**
S_1	0.8100	0.8910	1.1310	0.0750
S_2	0.7300	0.9810	1.4650	0.2190
S_3	0.7800	0.0550	0.0002	0.0004
S_4	1.6000	0.0570	0.0640	0.0006
S_5	1.6050	0.0590	0.4890	0.0002
S_6	0.0600	1.3140	1.5530	0.1490
S_7	0.0201	1.5400	1.3620	0.3270
S_8	0.0264	1.1370	1.2850	0.1080
S_9	1.4000	1.2110	1.2760	0.2440
S_{10}	1.4500	1.2620	1.3087	0.1490

In the next step, we evaluated the reliability of the proposed confidence measure. For this purpose, we found the number of high-confident samples in different classes and then determine the number of samples that were correctly predicted among them. This parameter, known as reliability, shows the performance of the proposed confidence measure. The result of this analysis is shown in Table 4.

Table 4. Reliability evaluation of the confidence measure.

c.c. / c.$^{+}$	**SNR**			
	0 dB	**4 dB**	**8 dB**	**12 dB**
S_1	$^{82}/_{82}$	$^{358}/_{358}$	$^{82}/_{82}$	$^{567}/_{567}$
S_2	$^{355}/_{355}$	$^{287}/_{287}$	$^{293}/_{293}$	$^{155}/_{155}$
S_3	$^{87}/_{99}$	$^{117}/_{117}$	$^{19}/_{19}$	$^{30}/_{30}$
S_4	$^{75}/_{80}$	$^{205}/_{205}$	$^{106}/_{106}$	$^{145}/_{145}$
S_5	$^{590}/_{590}$	$^{619}/_{619}$	$^{329}/_{329}$	$^{184}/_{184}$
S_6	$^{267}/_{267}$	$^{282}/_{282}$	$^{258}/_{258}$	$^{296}/_{296}$
S_7	$^{188}/_{188}$	$^{161}/_{161}$	$^{78}/_{78}$	$^{443}/_{443}$
S_8	$^{254}/_{475}$	$^{184}/_{184}$	$^{278}/_{278}$	$^{492}/_{492}$
S_9	$^{158}/_{209}$	$^{583}/_{645}$	$^{294}/_{294}$	$^{373}/_{373}$
S_{10}	$^{127}/_{278}$	$^{319}/_{382}$	$^{318}/_{318}$	$^{301}/_{302}$

$^{+}$: c.c. / c. indicate the ratio of correct high-confidence samples to high-confidence samples.

3.4. Performance Comparison

In this section, the accuracies which were obtained from the proposed classifier were compared to the system that was trained by supervised PA algorithm in the matched SNR (SNR aware mode), a clean trained system (a system which was trained by clean samples instead of a general training set) and a general trained system with no adaptation. In addition, to evaluate the effectiveness of the confidence measure, the proposed algorithm is performed in the no confidence mode.

Furthermore, we simulated previous classifiers [12, 13] in similar situations, to show the superiority of our proposed idea. The results are indicated in Table 5 in percentage.

Table 5. Performance Comparison (%).

Methods	**0 dB**	**4 dB**	**8 dB**	**12 dB**	**Noise free**
Proposed Classifier using clean training samples	11.18	28.14	62.34	63.00	100
PA classifier with labeled data & General Training (No Adaptation)	79.35	88.86	96.21	98.91	100
Supervised PA classifier in SNR aware case with labeled data	85.09	97.21	99.84	100	100
Proposed classifier without confidence evaluation	80.06	88.27	97.88	99.90	100
[12]	71.91	86.22	92.87	98.26	100
[13]	79.88	90.17	97.23	100	100
Proposed classifier (SNR unaware case, training samples are not labeled)	81.36	92.00	98.28	100	100

As it is observed in Table 5, all of the evaluated algorithms have similar performance in noise-free mode. Simulation results show that the proposed algorithm generally performs well in low SNRs and its accuracy is close to the SNR aware case. If the initial training in the proposed method performs with clean signals, the recognition accuracy was greatly reduced in very low SNR. Therefore, the first algorithm cannot track SNR changes in new existing conditions. Simulation results indicate that the errors were relatively 40.22% reduced with applying the confidence idea. The fair comparison with other state of art competing algorithms shows that the proposed algorithm is well outperforms the others. If the initial training in the proposed method performs with clean signals, the recognition accuracy was greatly reduced in very low SNR.

Therefore, the first algorithm cannot track SNR changes in new existing conditions. Simulation results indicate that the errors were relatively 40.22% reduced with applying the confidence idea. The fair comparison with other state of art competing algorithms shows

thatthe proposed algorithm is well outperforms the others.

4. Conclusions

Automatic modulations classification plays a significant role in civil and military applications. In this paper, we have presented, implemented and tested a new partially supervised classifier for AMC application in which system adaptation to the SNR of environment and classification of modulated signal are efficiently combined.System adaptation and classification is performed according to the online passive-aggressive algorithm. In training procedure, the idea of general training is used. The experimental results revealed that the performance of proposed classifier, even in low SNRs, is comparable to batch trained system in SNR aware case. In future works, we attempt to make this algorithm online so that the evaluation results for each sample would be determined at the time of its entering.

References

[1] Azarmanesh O. and Bilen S., “New Results on Two-Stage Novel Modulation Classification Technique for Cognitive Radio Application,” *in Proceeding of IEEE MILCOM*, Maryland, pp. 266-271, 2011.

[2] Azzouz E. and Nandi A., “Automatic Identification of Digital Modulation Types,” *Signal Processing*, vol. 47, no. 1, pp. 55-69, 1995.

[3] Azzouz E. and Nandi A., “Algorithms for Automatic Modulation Recognition of Communication Signals,” *IEEE Transaction Communication*, vol. 46, no. 4, pp. 431-436, 1998.

[4] Boyd S. and Vandenberghe L., *Convex Optimization*, Cambridge University Press, 2004.

[5] Burges C., “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

[6] Chandrashekar G. and Sahin F., “A Survey on Feature Selection Methods,” *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16-28, 2014.

[7] Chergui L., Mammar K., and Salim C., “Combining Neural Networks for Arabic Handwriting Recognition,” *The International Arab Journal of Information Technology*, vol. 9, no. 6, pp. 588-595, 2012.

[8] Cortes C. and Vapnik V., “Support Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[9] Crammer K., Dekel O., Keshet J., Shalev-shwartz S., and Singer Y., “Online Passive-Aggressive Algorithms,” *Journal of Machine Learning Research*, vol. 7, pp. 551-585, 2006.

[10] Crammer K. and Singer Y., “A New Family of Online Algorithms for Category Ranking,” *Journal of Machine Learning Research*, vol. 3, pp. 1025-1058, 2003.

[11] Dobre O., Abdi A., Bar-Ness Y., and Su W., “Survey of Automatic Modulation Classification Techniques: Classical Approaches and New Trends,” *IET Communications*, vol. 1, no. 2, pp. 137-156, 2007.

[12] Dulek B., Ozdemir O., Varshney P., and Su W., “A Novel Approach to Dictionary Construction for Automatic Modulation Classification,” *Journal of the Franklin Institute*, vol. 351, no. 5, pp. 299-3012, 2014.

[13] Ebrahimzadeh A., “A Novel Method for Automatic Modulation Recognition,” *Applied Soft Computing*, vol. 12, no. 1, pp. 453-461, 2012.

[14] Lopatka J. and Macrej P., “Automatic Modulation Classification using Statistical Moments and a Fuzzy Classifier,” *in Proceeding of International Conference on Signal Processing*, Beijing, pp. 1500-1506, 2000.

[15] Mobasseri B., “Digital Modulation Classification using Constellation Shape,” *Signal Processing*, vol. 80, no. 2, pp. 251-277, 2000.

[16] Proakis J., *Digital Communications*, McGraw-Hill, 2001.

[17] Rahman M., Kumar M., and Uddin M., “Optimum Threshold Parameter Estimation of Wavelet Coefficients using Fisher Discriminant Analysis for Speckle Noise Reduction,” *The International Arab Journal of Information Technology*, vol. 11, no. 6, pp. 573-581, 2014.

[18] Ramkumar B., “Automatic Modulation Classification for Cognitive Radios using Cyclic Feature Detection,” *IEEE Circuits and Systems*, vol. 9, no. 2, pp. 27-45, 2009.

[19] Sengur A., “Multiclass Least-Squares Support Vector Machines for Analog Modulation Classification,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 6681-6685, 2009.

[20] Swami A. and Sadler B., “Hierarchical Digital Modulation Classification using Cumulants,” *IEEE Transactions on Communications*, vol. 48 no. 3, pp. 416-429, 2000.

[21] Vapnik V., *Statistical Learning Theory*, John Wiley and Sons, 1998.

[22] Wu H., Saquib M., and Yun Z., “Novel Automatic Modulation Classification using Cumulant Features for Communications via Multipath Channels,” *IEEE Transactions on Wireless Communications*, vol. 7 no. 8, pp. 3089-3105, 2008.

[23] Yaqin Z., Guanghui R., Xuexia W., Zhilu W., and Xuemai Gu., “Automatic Digital Modulation

Recognition using Artificial Neural Networks," *in International Conference on Neural Networks and Signal Processing*, Cairo, pp. 257-260, 2003.

Hamidreza Hosseinzadeh is a Ph.D. candidate in Department of Electrical Engineering, in Islamic Azad University, Science and Research Branch, Tehran, Iran. His research interests are signal processing, and pattern recognition.

Farbod Razzazi is an Associate Professorin Department of Electrical Engineering, in Islamic Azad University, Science and Research Branch, Tehran, Iran. His research interests are pattern recognition methods.