Performance evaluation of the correntropy coefficient in automatic modulation classification

Aluisio I.R. Fontes a,⇑, Allan de M. Martins a, Luiz F.Q. Silveira a, J.C. Principe b

a Department of Computer Engineering and Automation, Universidade Federal do Rio Grande do Norte, 59078-900 Natal, Brazil
b Computational Neuroengineering Laboratory (CNEL), Department of ECE, University of Florida, Gainesville, FL 32611, USA

Article info
Article history:
Available online 26 July 2014

Keywords:
Automatic modulation classification
Information theory
Correntropy coefficient

Abstract
Automatic modulation classification (AMC) techniques have applications in a variety of wireless communication scenarios, such as adaptive systems, cognitive radio, and surveillance systems. However, a common requirement to most of the AMC techniques proposed in the literature is the use of signal preprocessing modules, which can increase the computational cost and decrease the scalability of the AMC strategy. This work proposes the direct use of a similarity measure based on information theory for the automatic recognition of digital modulations, which is known as correntropy coefficient. The performance of correntropy in AMC applied to channels subject to additive white Gaussian noise (AWGN) is evaluated. Specifically, the influence of the kernel size on the classifier performance is analyzed, since it is the only free parameter in correntropy. Besides, a relationship between its respective value and the signal-to-noise ratio (SNR) of the channel is also proposed. Considering the investigated modulation techniques, numerical results obtained by simulation demonstrate that there are high accuracy rates in classification, even at low SNR values. By using correntropy, AMC task becomes simpler and more efficient.

1. Introduction

Automatic modulation classification is an intermediate step between signal detection and demodulation, aiming to identify the modulation format of the received signal with little prior knowledge (Zaerin, Masoud, & Nikoofar, 2012). It plays a key role in various applications such as software defined radio (Xu, Su, & Zhou, 2011), intelligent modems (Dobre, Bar-Ness, & Wei; Dobre, Abdi, Bar-Ness, & Su, 2007), civilian spectrum monitoring (Dobre et al., 2007), cognitive radio (Dobre et al., 2007; Ramkumar, 2009; Wang & Liu, 2011), electronic warfare, and threat analysis (Ramkumar, 2009).

AMC algorithms can be divided into two categories: likelihood-based ones (Luo, Li, Qian, & Lu, 2014; Ozdemir, Li, & Varshney, 2013; Choqueuse, Marazin, Collin, Yao, & Burel, 2010) and feature-based ones (Shrme, 2011; Türk & Ogras, 2011; Freitas, Cardoso, Muller, Costa, & Klautau, 2009; Muller, Cardoso, & Klautau, 2011; Ho, Prokopiw, & Chan, 2000; Aslam, Zhu, & Nandi, 2012; Zadeh, 2010; Avci & Avci, 2009). In the first category, there is a preprocessing step to estimate parameters such as the carrier phase, frequency, and symbol period of the signal. Then, classification is performed from a decision test based on maximum likelihood. Analogously, the second category also aggregates a preprocessing step, where the extraction of the signal features used by the classifier occurs.

Various techniques have been proposed for the extraction of signal features in order to improve the performance of the AMC classifier. Spectral correlation, entropy, cumulants, and wavelet transform are proposed in Freitas et al. (2009) for the extraction of features that can be classified by a support vector machine (SVM). Similarly, SVM classifiers have also been used in the features extraction based on higher order moments, higher order cumulants, spectral features, and discrete wavelet transform (Avci & Avci, 2009; Shrame, 2009; Zadeh, 2010). Wavelet transform has also been employed in features extraction, with subsequent classification by a neural network (Muller et al., 2011; Ho et al., 2000; Türk & Ogras, 2011). The standard deviation and other statistical measures of the received signal can be used to classify modulations by a neural network (Rube & El-Madany, 2010). Specifically, features extraction in Aslam et al. (2012) is achieved using fourth-order and sixth-order cumulants, while the KNN algorithm is used to evaluate the ability of GP individuals during the training phase.

⇑ Corresponding author.
E-mail addresses: aluisio@dca.ufrn.br (A.I.R. Fontes), allan@dca.ufrn.br (A. de M. Martins), lfelipe@dca.ufrn.br (L.F.Q. Silveira), principe@cnel.ufl.edu (J.C. Principe).
Both aforementioned AMC categories rely strongly on the preprocessing step of the signal. In addition, techniques based on likelihood have high computational cost (Eldemerdash, Marey, Dobre, Karagiannidis, & Inkol, 2012). Despite having lower computational cost (Zaerin et al., 2012), feature-based techniques employ classifiers with complex architectures e.g. artificial neural network (ANN), SVM, and hidden Markov model (HMM), which can restrict their application or even make the technique unfeasible in a real communication environment, with consequent decrease of scalability.

This work introduces a novel AMC approach based on a measure of information theory called correntropy coefficient using the adaptive adjustment of kernel size, which is the only free parameter of the method. The proposed featured-based method stands out both by the absence of the preprocessing step for feature extraction and simple classification step, thus reducing the complexity of the proposed architecture. Previous works with binary modulation (Fontes, Cavalcante, & Silveira, 2012b; Fontes et al., 2012a) have adjusted the kernel size using the empirical Silverman rule method (Silverman, 1986), which indicates the viability for the removal of the preprocessing step in AMC. This work investigates the influence of the kernel size on the performance of the classifier, and proposes a method to select it as a function of the measured data.

This paper is organized as follows: Section 2 provides a definition of the correntropy coefficient, which is the statistical similarity measure employed in the proposed classification architecture. Section 3 presents the proposed classifier, while Section 4 provides an analysis of its performance. The conclusions are summarized in Section 5.

2. Correntropy coefficient

The use of metrics based on information theory applied to problems of machine learning was introduced in Principie, thus defining a methodology known as Information Theoretic Learning (ITL). Formerly, such problems were solved primarily using measurements of the mean square error. The main ITL concept consist in using metric based on nonparametric estimates of Renyi’s quadratic entropy as cost functions for the design of adaptive systems.

Renyis definition of entropy is a generalization of the Shannon entropy, and is given by (Rényi, 1961):

\[ H_a(X) = \frac{1}{1-a} \log \left( \sum_{i=1}^{m} p_i^a \right), \]

where \( p_i \) is defined from the probability distribution of random variable (r.v.) \( X \) and \( a \) is a parameter that specifies the order of Renyi’s entropy. When the probability distribution of \( X \) is estimated from a finite set of \( n \) measured data \( \{x_i\}_{i=1}^{n} \) by a Gaussian kernel \( G(\sigma) \cdot (\cdot) \) with standard deviation \( \sigma \), the Renyi’s quadratic entropy (for \( a = 2 \)) can be estimated as (Principe, 2010):

\[ \hat{H}_2(X) = - \log \left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{\sigma/2}(x_i - x_j) \right), \]

where,

\[ G_{\sigma}(x_i - x_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[ -\frac{(x_i - x_j)^2}{2\sigma^2} \right]. \]

The argument of Renyi’s quadratic entropy in (2) is called Information Potential (IP) of the r.v. \( X \), given by (Information theoretic learning, 2000):

\[ \hat{V}_a(X) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{\sigma/2}(x_i - x_j). \]

The information potential can be generalized to two random variables, giving rise to the Cross-Information Potential (CIP) defined by:

\[ \hat{V}_a(X, Y) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{\sigma/2}(x_i - y_j). \]

For independent random variables, the following statement can be demonstrated (Information theoretic learning, 2000):

\[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{\sigma/2}(x_i - y_j) \approx \frac{1}{n} \sum_{i=1}^{n} G_{\sigma/2}(x_i - y_i), \]

which is an approximation of CIP with algorithmic complexity \( O(n) \).

From the viewpoint of kernel methods, the right-hand side of Eq. (6) can be seen as a sample estimator of a generalized similarity measure between two arbitrary scalar random variables \( X \) and \( Y \), which is called cross-correntropy, and defined as (Santamaria, Pokharel, & Principie, 2006):

\[ \nu_a(X, Y) = E_{X,Y}[G_a(X - Y)] = \int \int G_a(x - y) p_X(x) p_Y(y) dx dy, \]

where \( E[\cdot] \) is the expectation operator and, \( \sigma \) is the kernel size. By estimating the joint probability density function (pdf) of Eq. (7) by Parzen method from a finite number of data \( \{(x_i, y_i)\}_{i=1}^{n} \), the right-hand side of Equating (6) is obtained:

\[ \hat{\nu}_a(X, Y) = \frac{1}{n} \sum_{i=1}^{n} G_{\sigma/2}(x_i - y_i). \]

It is interesting to notice the relationship between cross-correntropy estimators and information-theoretic ones when both variables \( X \) and \( Y \) are independent.

By applying a Taylor series expansion to the Gaussian function in (7), and assuming that all the moments of the joint pdf are finite, Eq. (7) becomes

\[ \nu_a(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=0}^{\infty} \frac{(-1)^k}{2^{k-1}k!} E[(X - Y)^{2k}]. \]

This expression states that the cross-correntropy is sensitive to the sum of second-order and higher-order moments of the difference variable. Thus, cross-correntropy is considered a generalization of the correlation concept (Santamaria et al., 2006), and it can be also used as a measure of similarity between the random variable. In fact, since cross-correntropy is sensitive to the sum of all even moments of the random variables, it can be used in quantification with improved performance if compared with correntropy in many applications e.g. non-Gaussian and nonlinear problems (Santamaria et al., 2006).

The kernel size in Eq. (9) appears as a parameter that weighs the second-order moment and the higher-order moments. For sufficiently large values of \( \sigma \), the second-order moment is predominant and the measure approaches correlation.

Due to nonlinear transformations produced by the Gaussian kernel, cross-correntropy has no guarantee of zero mean, even when the input data are centered at zero. The definition of centered cross-correntropy is able to overcome such limitation (Principe, 2010):

\[ \hat{\nu}_a(X, Y) = \frac{1}{n} \sum_{i=1}^{n} G_{\sigma/2}(x_i - y_i) - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{\sigma/2}(x_i - y_j). \]
It is possible to notice that the centering term in (10) is numerically equal to the estimator of the cross-information potential in (5). Thus, from Eq. (6), centered cross-correntropy is reduced to zero if \( X \) and \( Y \) are independent random variables.

In applications involving signals with unknown amplitudes, some type of normalization may be required. In order to avoid this need, Xu, Bakardjian, Cichocki, and Principe (2007) have presented a new similarity measure called correntropy coefficient, which is estimated by:

\[
\tilde{\eta} = \frac{\tilde{u}_t(X, Y)}{\sqrt{\tilde{u}_t(X, X)\tilde{u}_t(Y, Y)}}.
\]  

(11)

The correntropy coefficient \( \tilde{\eta} \) can be considered a generalization of the known correlation coefficient. It can be proved that this parameter becomes null if two random variables \( X \) and \( Y \) are statistically independent, while its absolute value approaches 1 as they become statistically related.

This work uses the ability of the correntropy coefficient to measure high order statistical information for the accurate characterization of the dynamic interdependences between digital modulation signals. This enables AMC to be performed without the need for a preprocessing module to extract the features unlike other AMC approaches, thus reducing the computational complexity of the proposed method.

3. Proposed classifier

The proposed classification architecture has been investigated in a communication environment characterized by the presence of AWGN. In this model, the modulated signal \( s(t) \) can be expressed generically by (Sklar, 2001):

\[
s(t) = A(t)\cos(\theta(t)) \quad \text{if} \quad t \in [i\cdot T_s, (i + 1)\cdot T_s),
\]  

(12)

where \( A(t) \) and \( \theta(t) \) are the amplitude and angle of the carrier, respectively, defined according to the modulation format adopted for the transmitted signal, which in this work is restricted to modulations BPSK, OOK, 16-QAM, MSK, and QPSK. The received signal \( r(t) \) is given by:

\[
r(t) = s(t) + n(t) \quad \text{if} \quad t \in [i\cdot T_s, (i + 1)\cdot T_s),
\]  

(13)

where \( n(t) \) is the complex Gaussian noise with zero mean and power spectral density equal to \( N_0/2 \) per dimension. According to the proposed model, the signal contaminated with noise, \( r(t) \), whose modulation format is initially unknown (but still limited to the modulation schemes BPSK, OOK, 16-QAM, MSK, and QPSK), is sampled in the receiver as shown in Fig. 1, being represented by sample vector \( X \).

The objective of the proposed architecture is to identify the modulation format employed by the received signal \( X \) through the correntropy coefficient defined in (11). The architecture uses a set of previously stored signals with known modulations called templates, which act as a reference during calculation of the correntropy coefficient. These templates are grouped in the architecture as a function of their modulation formats. It was observed experimentally that the amount of binary templates in each group associated with a given modulation type must be equal to the greatest number of symbols \( M \), amongst the M-ary investigated schemes. In addition, the size of each template must be equal to \( N \cdot \log(M) \), where \( N \) is the number of samples per bit of the system. Specifically in this work, each modulation format is associated with 16 different templates, representing the 16 possible nibbles (semioctets): \('0000'\), \('0001'\), \('0010'\)....\('1111'\). Each template, denoted by \( Y_i \), \( 1 \leq i \leq 16 \), is sampled at a rate of 100 samples per bit.

The classification process is performed in two steps. Firstly, the proposed classifier calculates the correntropy coefficient \( \tilde{\eta}_i \) given by Eq. (11), which involves the signal sample \( X \) and each template \( Y_i \) associated with a given modulation format, according to the block diagram in Fig. 1. Then, the correntropy coefficients are calculated, which are represented by the intersection among the following main blocks:

(a) \( \tilde{u}_t(X, X) \): auto-correntropy of the input signal.
(b) \( \tilde{u}_t(Y_i, Y_i) \): auto-correntropy of each template associated with a modulation format.
(c) \( \tilde{u}_t(X, Y_i) \): correntropy between the input signal \( X \) and each template \( Y_i \) associated with a modulation format.

The second step consists in selecting the largest \( \tilde{\eta}_i \) in each group of templates associated with a modulation format, with the subsequent comparison of the largest value with the distinct modulation techniques investigated in the architecture. The detection of the largest correntropy coefficient is represented by the block furthest to the right in Fig. 1. Finally, the algorithm determines the modulation of the received signal according to the group that contains the largest coefficient.

Since the proposed architecture uses a kernel size \( \sigma \) that is adjusted for each type of modulation, it is possible to refine the observation window within which similarity is evaluated, consequently maximizing the accuracy rate of the proposed method. The classification efficiency of the introduced architecture based on the correntropy coefficient has been evaluated comparing it with a reference architecture that is identical to the proposed one, except for the measure of similarity, which in the reference system is the correlation coefficient. This is a fair reference, since the correntropy coefficient can be considered a generalization of the correlation coefficient, with similar computation complexity.

4. Simulations and results

4.1. Simulation data

The proposed architecture has been evaluated by computational simulation carried out in MATLAB using modulation techniques BPSK, OOK, QPSK, 16-QAM, and MSK. The simulation parameters are presented in Table 1.

The energy of the constellations was normalized in order to avoid bias in the estimates, since the classifier is based on a statistical measure of similarity.

4.2. Simple case study

In this section, a simple case study is presented in order to demonstrate the applicability of the proposed AMC architecture. The parameters used in the simulation test are described in Table 1 considering that SNR is equal to 0 dB.

The correntropy coefficient \( \tilde{\eta}_i \) is calculated between the received signal and each template associated with a given modulation format (OOK, MSK, BPSK, QPSK, 16-QAM). In this example, the received signal has modulation format BPSK and each modulation is associated with a set of 15 templates, \( i = 0, \ldots, 15 \). It is possible to calculate 16 values of \( \tilde{\eta} \) for each modulation technique, which are given in Table 2.

Table 2 shows that the largest correntropy coefficient occurs for the template \( '0111' \) associated with BPSK. Therefore, the classifier is able to identify BPSK as the modulation format used by the received signal accurately.

It is important to observe that the two largest values of coefficients related the two largest values for the coefficients in 3 case...
are obtained for BPSK and QPSK i.e. 0.731 and 0.553, respectively. This is an expected result, since the received signal is in the BPSK format, which in the other hand is very similar to QPSK. Templates associated with other modulations lead to small values of $\eta$ because such modulation techniques belong to other distinct families.

### 4.3. Adaptive architecture

The only adjustable parameter in the architecture of the proposed AMC is the kernel size in Eq. (3). Many systems use a heuristic known as the Silverman rule to determine such value (Silverman, 1986). However, this section considers that the kernel size is an additional parameter of the AMC process and is supposed to show that it can be adjusted as a function of the SNR in the receiver.

#### Table 1
Simulation parameters for the transmitter.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling frequency $F_s$</td>
<td>500 kHz</td>
</tr>
<tr>
<td>Carrier frequency $F_c$</td>
<td>20 kHz</td>
</tr>
<tr>
<td>Symbol rate</td>
<td>5 kbps</td>
</tr>
<tr>
<td>SNR range</td>
<td>-10 dB to 15 dB</td>
</tr>
<tr>
<td>Constellation energy</td>
<td>5</td>
</tr>
</tbody>
</table>

#### Table 2
Values of the correntropy coefficient $\eta_{x,y}$ given by Eq. (11) between the received signal and each template associated a modulation format.

<table>
<thead>
<tr>
<th>Templates</th>
<th>OOK</th>
<th>BPSK</th>
<th>MSK</th>
<th>QPSK</th>
<th>16-QAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>0.000</td>
<td>0.398</td>
<td>0.074</td>
<td>0.246</td>
<td>0.246</td>
</tr>
<tr>
<td>0001</td>
<td>0.018</td>
<td>0.050</td>
<td>0.128</td>
<td>0.000</td>
<td>0.360</td>
</tr>
<tr>
<td>0010</td>
<td>0.071</td>
<td>0.015</td>
<td>0.046</td>
<td>0.024</td>
<td>0.077</td>
</tr>
<tr>
<td>0011</td>
<td>0.037</td>
<td>0.332</td>
<td>0.010</td>
<td>0.270</td>
<td>0.244</td>
</tr>
<tr>
<td>0100</td>
<td>0.028</td>
<td>0.000</td>
<td>0.029</td>
<td>0.036</td>
<td>0.279</td>
</tr>
<tr>
<td>0101</td>
<td>0.032</td>
<td>0.348</td>
<td>0.086</td>
<td>0.283</td>
<td>0.177</td>
</tr>
<tr>
<td>0110</td>
<td>0.030</td>
<td>0.383</td>
<td>0.002</td>
<td>0.306</td>
<td>0.408</td>
</tr>
<tr>
<td>0111</td>
<td>0.014</td>
<td><strong>0.731</strong></td>
<td>0.057</td>
<td><strong>0.553</strong></td>
<td>0.327</td>
</tr>
<tr>
<td>1000</td>
<td>0.070</td>
<td>0.717</td>
<td>0.012</td>
<td>0.471</td>
<td>0.329</td>
</tr>
<tr>
<td>1001</td>
<td>0.036</td>
<td>0.369</td>
<td>0.069</td>
<td>0.224</td>
<td>0.188</td>
</tr>
<tr>
<td>1010</td>
<td><strong>0.100</strong></td>
<td>0.334</td>
<td>0.019</td>
<td>0.200</td>
<td>0.398</td>
</tr>
<tr>
<td>1011</td>
<td>0.071</td>
<td>0.013</td>
<td>0.074</td>
<td>0.046</td>
<td>0.328</td>
</tr>
<tr>
<td>1100</td>
<td>0.030</td>
<td>0.319</td>
<td>0.087</td>
<td>0.188</td>
<td>0.246</td>
</tr>
<tr>
<td>1101</td>
<td>0.014</td>
<td>0.029</td>
<td><strong>0.142</strong></td>
<td>0.058</td>
<td>0.372</td>
</tr>
<tr>
<td>1110</td>
<td>0.066</td>
<td>0.063</td>
<td>0.059</td>
<td>0.082</td>
<td>0.067</td>
</tr>
</tbody>
</table>

Bold value shows the largest correntropy coefficient associated to modulation format.

Fig. 1. Detailed view of the proposed classifier.
Specifically, the performance of the correntropy classification architecture is analyzed in an adaptive communication system, which is shown in Fig. 2, whose receiver is able to provide SNR estimates for the AMC. The correntropy classifier is considered capable to adjusting the kernel size according to the estimated SNR range, consequently improving its classification performance. It is possible to obtain in practice kernel sizes 15 times larger than the ones given by Silverman’s rule.

The experimental results are given in Figs. 3–7. The results were used to determine suboptimal values for the width of the kernel associated with each SNR.

The architecture performance has been evaluated according to Monte Carlos Method. A minimum number of 100 trials was considered for each experiment, with a window size of 400 samples for each value of kernel size and SNR. The experimental results are given in Figs. 3–7. The results were used to determine suboptimal values for the width of the kernel associated with each SNR.

Adjustment of the kernel size provides an effective mechanism of eliminating noise. There is clearly a kernel size appropriate for each type of modulation and range of SNR, as shown in Table 3. This adjustment of the kernel acts as a zoom controlling the observation window in which the statistical similarity is evaluated. The correct selection of its size can substantially increase the accuracy rate, as shown in Fig. 3 for the MSK modulation, where the correct classification rate increases from 20% to 100% at a signal-to-noise ratio (SNR) of −10 dB.

The remaining energy in one of the symbols of the OOK modulation made it difficult to adjust the kernel size used by the architecture in classification. However, it is still possible to notice in Table 3 and Fig. 4 that $\sigma = 4.64$ provides a correct classification rate of about 96% at an SNR of −10 dB. In the case of 16-QAM modulation (Fig. 5), the proposed architecture achieved correct classification rates of 93% at an SNR of −10 dB, and 100% at an SNR of 0 dB, with kernel sizes equal to 21.56 and 2.15, respectively (Table 3). At a SNR of 0 dB, the correct classification rate is 100% for most of the kernel sizes regarding QPSK modulation according to (Fig. 6). At an SNR of −5 dB the correct classification rate is about 96% with $\sigma = 10$. These results indicate that the adjustment of the kernel size according to the estimated SNR of the system improves the performance of the classifier.
The relation between the kernel size and the SNR Table 3 effectively shows that the smaller the SNR of the system, the larger the kernel size in the proposed classification architecture. Thus, in channels with high levels of additive white Gaussian noise, the correntropy AMC architecture is supposed to have the same performance as a correlation classifier, since second-order statistical moments should prevail in the modulated signals. On the other hand, when the level of Gaussian noise is low, the kernel size is small and higher-order moments predominate in the received modulated signal, which favors its classification by correntropy. Finally, it is worth to mention that the smallest value of \( \sigma \) is chosen order to favor the significance higher-order statistical moments in the measure of correntropy whenever there are distinct kernel sizes that maximize the accuracy rate of the classifier.

4.4. Nonadaptive architecture

This section analyzes the correntropy classification when the SNR estimator is absent in the system receiver. Under this condition, a given value is attributed to the kernel size previously to the correntropy coefficient associated with each modulation technique. The aforementioned values listed in Table 4 have been selected from the analysis of results shown in Figs. 3–7. The performance of the method in this new communication scenario is evaluated in comparison with two reference architectures. The first one adjusts the kernel size of the correntropy coefficients by means of the Silverman rule, while the second one replaces the correntropy coefficient by the correlation coefficient.

Regarding the first reference architecture, the kernel size of each correntropy coefficient is estimated using the Silverman rule, given by (Silverman, 1986):

\[
\sigma = \sigma_0 \left( 4N^{-1/2}d^{-1} \right)^{1/4},
\]

(14)

where \( d \) corresponds to the dimension of the data (\( d = 1 \) in the studied cases), \( N \) corresponds to the number of samples, and \( \sigma_0 \) corresponds to the plot of the autocovariance matrix of \( X \). The classification rates obtained in this experiment are presented in Fig. 8.

It can be seen from Fig. 8 that adjusting the kernel size as proposed in Section 4.3 is decisive in increasing the accuracy rate of the classification architecture. Besides, it is possible to state that estimation of the kernel size using the Silverman rule is not satisfactory, especially at low SNR values. By using the preset kernel size values obtained by simulation in Table 4, a comparison of hit rates for classification is presented in Table 5 using the correntropy coefficient and the correlation coefficient.

It can be seen from in Table 5 that the correlation coefficient is unable to achieve the correct classification rate in 100% of the performed tests using 16-QAM, QPSK, and OOK modulation techniques for any of the SNR values. Even at very low values of SNR, which favor second-order moments when Gaussian noise exists, correlation is unable to achieve an accuracy rate greater than the one obtained using correntropy. However, correlation has achieved the same accuracy rate as correntropy for the MSK modulation technique.

4.5. Comparison with existing methods

The use of GP combined with the KNN algorithm is investigated in Aslam et al. (2012). This algorithm has been used to evaluate the performance of the KNN-GP approach during the training phase. The features are fourth-order and sixth-order cumulants. When compared with Aslam et al. (2012), the performance of correntropy with adaptive kernel size is superior in terms of scalability, since the addition of a new modulation scheme is only performed by means of noise-free associated templates. Every GP needs to be readjusted in Aslam et al. (2012), as it is necessary to extract the features of each new modulation with the calculation of cumulants for the modulated signals. Besides it is possible to state that the proposed architecture is able to recognize five formats (16-QAM, QPSK, BPSK, MSK, and OOK), while the one in Aslam et al. (2012) does only recognized four formats (16-QAM, 64-QAM, QPSK, and BPSK). The architecture proposed in this paper considers that SNR is between \(-10 \text{ dB} \) and \(15 \text{ dB} \), while the one in Aslam et al. (2012) investigated SNR values between \(5 \text{ dB} \) and \(20 \text{ dB} \), which are too high in many realistic conditions. In terms of accuracy rates, the same performance equivalent to 100% is achieved by both architectures for BPSK and QPSK modulation techniques. In the case of the 16-QAM modulation, the novel architecture provides accuracy 100%, while in Aslam et al. (2012) an accuracy of 80% was obtained. The 64-QAM modulation has not been analyzed in this work, although MSK and OOK have not been implemented in Aslam et al. (2012).

Another method based on discriminative learning and SVM for AMC is proposed in Muller et al. (2011). The features are according to the phase and magnitude of the received symbols in order to train a neural network. The performance of correntropy with adaptive kernel size is superior if compared with Muller et al. (2011) in terms of scalability, because the aforementioned method requires the adjustment of the neural network in order to insert a new mod-
main characteristic of the proposed approach is the absence of the preprocessing step that is common to most of the AMC methods described in literature (Freitas et al., 2009; Muller et al., 2011; Ho et al., 2000; Rube & El-Madany, 2010; Aslam et al., 2012; Zadeh, 2010; Avci & Avci, 2009), what effectively reduces computational complexity of the technique. Therefore, its application to practical studies in real-time systems becomes quite interesting. The results obtained with Monte Carlo simulation have demonstrated that the introduced method can lead to significant improvement of AMC performance for a wide range of SNR values.

From the correntropy theory, it is known that the kernel size of the correntropy coefficient acts as a parameter weighting the second-order moment and the higher-order moments. The influence of the kernel size in the classification hit rate has been analyzed, and it has been demonstrated that information gathered within higher-order statistical moments is quite efficient in digital modulation recognition. This fact clearly evidences the importance of varying the kernel size as a function of the channel’s SNR. Finally, this work has established a numerical relationship between the most appropriated kernel size of the correntropy and the channel’s SNR for AMC purposes.

Experimental tests involving the proposed method can be carried out by using hardware platforms for Software Defined Radio (SDR). Future work aims at the development of an experimental setup to investigate synchronism effects between the transmitter and receiver on the proposed method performance, as well as its respective robustness in the presence of multi-path fading. Besides, the hit rate of the proposed classifier is also supposed to be evaluated in scenarios with high order modulation schemes (e.g. 8-PSK, 8-APSK, 32-QAM), as well as its related computational cost.

5. Conclusions

This work has presented a novel method based on the correntropy coefficient with kernel size adjustment for the AMC. The main characteristic of the proposed approach is the absence of the preprocessing step that is common to most of the AMC methods described in literature (Freitas et al., 2009; Muller et al., 2011; Ho et al., 2000; Rube & El-Madany, 2010; Aslam et al., 2012; Zadeh, 2010; Avci & Avci, 2009), what effectively reduces computational complexity of the technique. Therefore, its application to practical studies in real-time systems becomes quite interesting. The results obtained with Monte Carlo simulation have demonstrated that the introduced method can lead to significant improvement of AMC performance for a wide range of SNR values.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>MSK</th>
<th>OOK</th>
<th>BPSK</th>
<th>QPSK</th>
<th>16-QAM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_k$</td>
<td>$c_0$</td>
<td>$c_k$</td>
<td>$c_0$</td>
<td>$c_k$</td>
</tr>
<tr>
<td>−10</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
<td>0.92</td>
<td>0.97</td>
</tr>
<tr>
<td>−5</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.97</td>
<td>0.94</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.95</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.92</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5

Comparison of classification using the correntropy and correlation coefficients, with the values of $\sigma$ given in Table 4. ($c_k$ – correntropy, $c_0$ – correlation).

Fig. 8. Results for the architecture without a SNR estimator using a suboptimum kernel and the Silverman rule.

References


