A Novel Modulation Classification Method in Cognitive Radios using Higher-Order Cumulants and Denoising Stacked Sparse Autoencoder

Xu ZHU* and Takeo FUJII†
Advanced Wireless and Communication Research Center
The University of Electro-Communications
Tokyo, 182-8585 Japan
* Email: zhuxu@awcc.uec.ac.jp
† Email: fujii@awcc.uec.ac.jp

***Abstract*—In this paper, we propose a novel modulation classification method based on deep network as well as higher-order cumulants. The proposed algorithm uses the higher-order cumulants as the features, and thus achieves impressive noise suppression. We use Stacked Denoising Sparse Autoencoder as a classifier for single-carrier modulation classification. This classifier can classify different modulated signals by cumulants automatically, and omit the decision of feature thresholds. A very different aspect from conventional neural network is its stacked structure, which simplifies an exponentially large number of hidden units by a multi-layer construction. Moreover, the better performance of backpropagation and network tune can be achieved while using Stacked Sparse Autoencoder. In addition, Denoising process improves the performance of noise suppression by training the network with a corrupted database. The performance of the multi-classes classification is given by simulations, which indicates that there is a significant performance advantage over the conventional methods.**

I. INTRODUCTION

Recently, the scarcity of spectrum has not been able to satisfy the spectrum demand for future mobile communications. In order to solve the spectrum availability problem, cognitive radio (CR) has attracted much attention in recent days. CR requires communication devices that can transmit at low power, and that change the transmitting frequency and modulation format on the fly [1]. Nonetheless, the recent surge in CR applications has been accompanied by an essential component, i.e., spectrum sensing, to avoid interference with a primary user. The spectrum-sensing performance can be improved using a reliable modulation classification (MC) scheme, e.g., to monitor the changes in interference, which is a challenging task. The demand of spectrum sensing can be met by a reliable MC scheme, for instance, to monitor how interference develops. Therefore, modulation classification methods have been extensively studied over the past three decades [2], and this is mainly because of its applicability to many practical problems.

Methods for identifying the modulation scheme of an intercepted signal broadly fall into two categories [3]: maximum likelihood (ML) and feature based method. Feature based method can be free from the parameter estimations [4]. It can be generally decomposed into two moves: feature extraction and classifier, which determine the modulation scheme according to those features acquired above [5].

The higher-order cumulants have been used for modulation classification by many papers. Reference [6] proposed a method based on the fourth-order cumulants, which was estimated from the sample estimates of the corresponding moments. It made decisions on the threshold of C_{42} and C_{40} for the hierarchical classification structure. However, the threshold-choosing process is not optimal because of the assumption on noise variance. Moreover, it is unable to classify two different modulation types completely when their features are so close to each other, e.g. the C_{42} of 16 quadrature amplitude modulation (QAM) and 64QAM. A related extension is given by [7] to classify linear modulations in frequency-selective channels. Reference [8] proposed a sixth-order moment based method using m_{63} to identify the order of QAM signals. It made decision threshold by minimizing the gap between the estimated values and the empirical ones.

All the papers we mentioned above, used hierarchical structures as classifiers, which compared the estimates of higher-order statistical features with certain thresholds to make decisions. An unavoidable problem is that using one single cumulant to distinguish some modulation types from others extremely restricts the classification performance. This is because that the linearity condition is required in selecting thresholds [9].

To achieve performance improvement, we use the stacked sparse autoencoder instead of conventional artificial neural network (ANN) [10] which suffers from some generality problems, e.g. ending up with over fitting. Furthermore, the network with a stacked multi-layer structure has better performance than ANN, due to the advantages of tune-up etc. Moreover, denoising autoencoder extends its performance by reconstructing the data from a corrupted version to extract more robust feature [11]. Due to the advantages we mentioned above, in just the past years, preliminary interest and discussions about deep learning have evolved into a full-fledged conversation that has captured the attention and imagination of researchers and engineers around the world.

The contributions of this paper are:

- It works for most major modulation types in cognitive radio system. Indeed, there are much more modulation types in wireless communication. However, the method we proposed here is for cognitive radio system, and the common modulation schemes are digital ones.
- It has better performance than conventional cumulants based methods and likelihood based methods.
- The linearity condition is not required since all the features are used simultaneously for the classification.

The rest of this paper is organised as follows. A brief primer about higher-order cumulants is given in Section 2. Classifier details are discussed in Section 3. Simulation results and conclusions are discussed in Section 4 and Section 5, respectively.

II. Higher-order Cumulants Feature of Signals

To better understand what can be done by deep networks, we first consider the feature extraction of relevant modulation. The nth-order cumulants of signal $x(k)$ can be represented as

$$C_{nr} = cum[x(k), x(k), ..., x(k), x^*(k), x^*(k), ..., x^*(k)], \tag{1}$$

where n indicates the total number of $x(k)$ and $x^*(k)$, r stands for the number of $x^*(k)$, and $cum[\cdot]$ represents cumulants operation. We use I to represent the set $[x(k), x(k), ..., x(k), x^*(k), x^*(k), ..., x^*(k)]$. Then, the nth-order cumulants operation can be represented as [12]

$$cum(I) = \sum_{\bigcup_{p=1}^{q} I_p = I} (-1)^{q-1}(q-1)! \prod_{p=1}^{q} E(I_p), \tag{2}$$

where $\bigcup_{p=1}^{q} I_p = I$ indicates an additive operation among all the division of I. We have $q = \{1, 2, ..., n\}$, and $p = \{1, ..., q\}$. E is the mathematical expectation. I_p is the division of the set I. I_p satisfies $\bigcup_p I_p = I$.

The purpose of a higher-order statistic is to build a more friendly feature space for the classifier [13]. We have calculated the normalized cumulants for the following modulation schemes: BPSK, QPSK, 8PSK, 16QAM, and 64QAM. We use following cumulants: C_{20}, C_{21}, C_{40}, C_{41}, C_{42}, C_{60}, C_{61}, C_{62}, C_{63}, C_{80}. The theoretical higher-order cumulants of MPSK and MQAM signals are listed in Table 1.

III. Classifier

In order to appreciate the advantages of stacked denoising sparse autoencoder (SDAE), let us briefly review its structure and function. The SDAE is a kind of multi-layer structure, whose activation (output value) of each layer is transmitted to the next layer forward, as shown in Fig. 1. The input layer represents input array, whose amount of input units is identical to the amount of input features. The hidden layers, whose values are not observed in the training set, contain multiple layers, whose parameters are obtained by the greedy layer-wise training [14]. The output layer consists of a softmax classifier, which is capable of classifying the modulation as desired. Moreover, the amount of output units is identical to the modulation schemes we desire to classify.

A. Classifier Training

The autoencoder tries to learn a function which aims to transform its input x into output $\hat{x}$, which is similar to x [15]. It values its distortion by the cost function and figures out the optimized activation of each node. We assume a set of unlabeled training examples $\{x^1, ..., x^m\}$. We define the network outputs and the cost function for a single training example by $h_{W,b}(x)$ and $J(W, b; x)$, respectively. Then the cost function of the single example can be given by

$$J(W, b; x) = \frac{1}{2} \left\| h_{W,b}(x) - x \right\|^2, \tag{3}$$

where W and b stand for the weight and the bias parameters, respectively. For a training set of m examples, the average error term of the overall cost function can be defined by

$$J(W, b) = \frac{1}{m} \sum_{i=1}^{m} J(W, b; x^{(i)}), \tag{4}$$

where m is the amount of examples, $x^{(i)}$ stands for the input vectors. To compress the weights magnitude and prevent overfitting, a decay has to be applied to the weight terms, which transforms (4) into

$$J(W,b) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \left\| h_{W,b}\left(x^{(i)}\right) - x^{(i)} \right\|^2 \right) + \frac{\lambda}{2} \sum_{l=1}^{n_l - 1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_l + 1} \left(W_{ji}^{(l)} \right)^2, \tag{5}$$

where λ is the weight decay parameter to control the relative weight of the two components. n_l stands for the amount of layers.

As Sparse Autoencoder embraces a spare constraint, the most of hidden units may keep zero if a sigmoid activation function is applied here. To add such penalty to the cost function, firstly, ρ, which is the sparsity parameter and typically a small value around zero, has to be introduced. Then an enforcement can be made to the average activation of hidden unit j, which is

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} \left[a_j^l \left(x^{(i)} \right) \right], \tag{6}$$

TABLE I
Higher-order cumulants of MPSK and MQAM signals

	BPSK	QPSK	8PSK	16QAM	64QAM
$c20$	1	0	0	0	0
$c21$	1	1	1	1	1
$c40$	-2	1	0	-0.68	-0.619
$c41$	-2	0	0	0	0
$c42$	-2	-1	-1	-0.68	-0.619
$c60$	16	0	0	0	0
$c61$	16	-4	0	2.08	1.7972
$c62$	16	0	0	0	0
$c63$	16	4	4	2.08	1.7972
$c80$	-272	-34	1	-13.9808	-11.5022

where l is the index of the layer. Then the the penalty term can be expressed by

$$KL\left(\rho \parallel \hat{\rho}_j\right) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j}, \tag{7}$$

which is based on the Kullback-Leibler divergence [11], a function to measure the difference between two distributions. With this in mind, it is apparent that the overall cost function with the weighted spare penalty term is given by

$$J_{sparse}\left(W,b\right) = J\left(W,b\right) + \beta \sum_{j=1}^{s_2} KL\left(\rho \parallel \hat{\rho}_j\right), \tag{8}$$

where s_2 represents the amount of hidden units. To optimize the activations of each layer, $J(W,b)$ should be minimized as a function of W and b. Although $J(W,b)$ is a non-convex function, gradient descent is still a practical algorithm to apply. We define $\delta_i^{(n_l)}$ as the difference between the activation and the true value by, then it can be expressed by

$$\delta_i^{(L)} = \left(\left(\sum_{j=1}^{s_2} W_{ji}^{(L)} \delta_j^{(L+1)}\right) + \beta\left(-\frac{\rho}{\hat{\rho}} + \frac{1-\rho}{1-\hat{\rho}_i}\right)\right) f'\left(Z_i^{(L)}\right), \tag{9}$$

where $f(z) = 1/(1+\exp(-z))$ is the sigmoid function. $f'\left(Z_i^{(L)}\right)$ is given by

$$f'\left(Z_i^{(L)}\right) = a_i^{(L)}\left(1 - a_i^{(L)}\right). \tag{10}$$

To compute the derivatives, we use following equations of partial derivatives of the cost function that are corresponding to single example (x,y) :

$$\begin{aligned} \frac{\partial}{\partial W_{ij}^{(L)}} J(W,b,x) &= a_j^{(L)} \delta_i^{(L+1)}, \\ \frac{\partial}{\partial b_{ij}^{(L)}} J(W,b,x) &= \delta_i^{(L+1)}. \end{aligned} \tag{11}$$

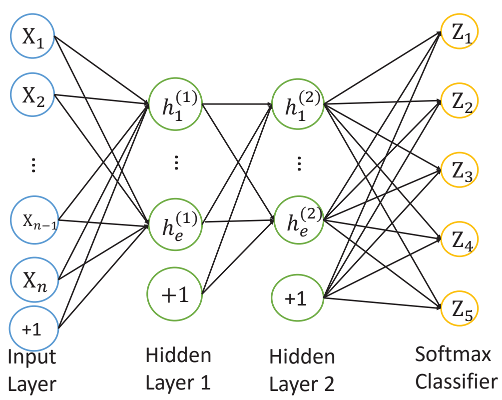

Fig. 1. Network structure of Denoising Stacked Sparse Autoencoder that contains 2 hidden layers

Then the overall cost functions $J(W,b)$ can be derived by

$$\begin{aligned} \frac{\partial}{\partial W_{ij}^{(L)}} J(W,b) &= [\frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial W_{ij}^{(L)}} J(W,b,x^{(i)})] + \lambda W_{ij}^{(L)}, \\ \frac{\partial}{\partial b_{ij}^{(L)}} J(W,b) &= [\frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial b_{ij}^{(L)}} J(W,b,x^{(i)})]. \end{aligned} \tag{12}$$

The lower equation is different from the upper one because there is no weight decay applied to b. Then backpropagation can be used via parameters update which is

$$\begin{aligned} W_{ij}^{(L)} &= W_{ij}^{(L)} - \alpha \frac{\partial}{\partial W_{ij}^{(L)}} J\left(W,b\right), \\ b_{ij}^{(L)} &= b_{ij}^{(L)} - \alpha \frac{\partial}{\partial b_{ij}^{(L)}} J\left(W,b\right), \end{aligned} \tag{13}$$

where α represents the learning rate. A large learning rate benefits a fast descent but may suffer from a divergence. An accurate solution, however, can hardly be obtained unless a tolerance is given.

Set $\Delta W^{(l)}$ and $\Delta b^{(l)}$ a matrix of the same dimension of $W^{(l)}$ and vector of the same dimension of $b^{(l)}$, respectively. Then in pseudo-code, the algorithm can be expressed as follows:

input: Intercepted the modulated signal.

output: Classification result of the intercepted signal.

method:

1.set $\Delta W^{(l)} = 0$ and $\Delta b^{(l)} = 0$ for all layer.

2.for $i = 1 : m$

- calculate partial derivatives by (10) .
- set $\Delta W^{(l)} = \Delta W^{(l)} + \frac{\partial}{\partial W_{ij}^{(L)}} J(W,b,x)$.
- $\Delta b^{(l)} = \Delta b^{(l)} + \frac{\partial}{\partial b_{ij}^{(L)}} J(W,b,x)$.

end

3.paramters update by (13).

IV. Simulation and Results

We assume an environment that is synchronous and coherent. In addition, timing, carrier, and waveform recovery have been accomplished. All results are given under Additive White Gaussian Noise (AWGN) channel.

A. *The Training Phase of Network*

To train the classifier, a training database, where 10000 sets of each modulation scheme are included, has to be built in advance. Labels, which are unnecessary to self-taught learning, are compulsory in training samples due to softmax classifier, the last layer of network. The softmax classifier is a typical utilization of softmax regression, which generalizes the logistic regression to classification problems where label can takes on any values [16].

B. *Simulation Results*

Assume that the modulation types are obtained from a set of N possible modulations, where $M = M_1, M_2, ..., M_N$. We let (P_c) donate the probability that the classification result is identical to the transmitted signal. Then, using the conditional probability, (P_c) can be expressed as

$$P_c^{n|n} = P(D = M_n \mid M_n), \tag{14}$$

where $D = M_n$ represents the case where the classification result is M_n. Then, the average probability of obtaining a correct classification is given by

$$P_{cc} = N^{-1} \sum_{n=1}^{N} P_c^{n|n}. \tag{15}$$

We firstly concentrate on binary classes classifications. The performance at different SNRs for BPSK and QPSK is given by Fig. 2. The performance at different SNRs for 8PSK and 16QAM is given by Fig. 3. The performance at different SNRs for 16QAM and 64QAM is given by Fig. 4. The parameter N is the length of samples. We simulated 3 different lengths of samples. Apparently, a longer sample gives better performance.

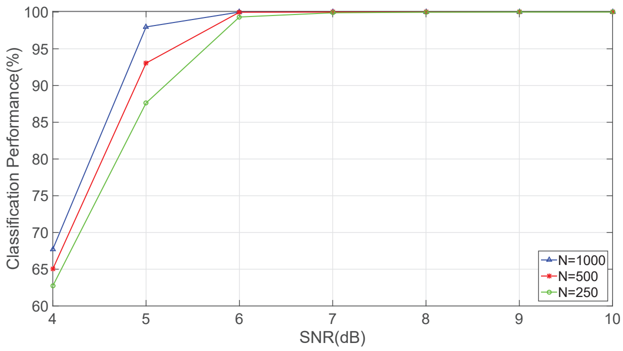

Fig. 2. The performance at different SNRs for BPSK and QPSK

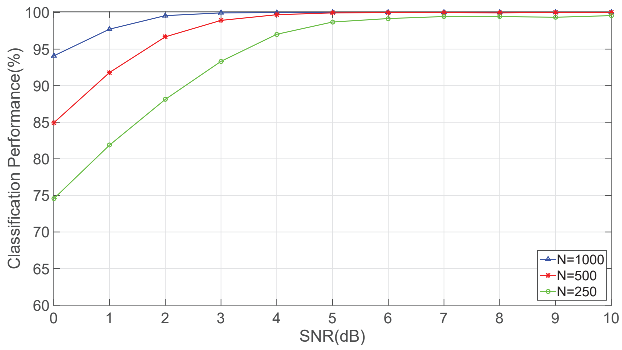

Fig. 3. The performance at different SNRs for 8PSK and 16QAM

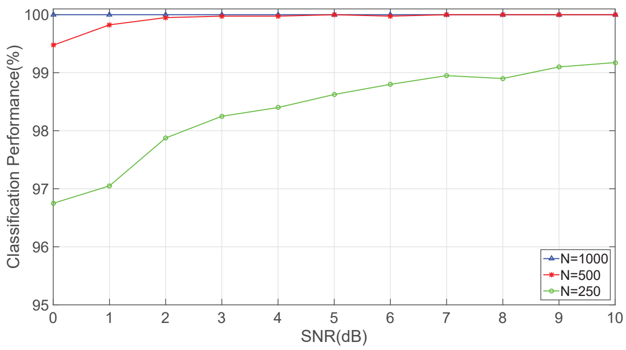

Fig. 4. The performance at different SNRs for 16QAM and 64QAM

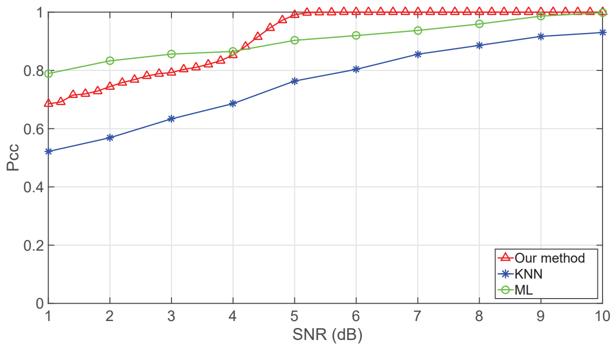

Fig. 5. Performance comparison with KNN and ML at different SNRs

We then compare the performance with the k-nearest neighbor (KNN) [17] and ML [18] methods for multi-classes classifications in Fig. 5. For each value of the SNR, 10,000 realizations of the test data have been produced. The ML shows a P_{cc} of $> 97\%$ at 9 dB. Our method shows a P_{cc} of $> 97\%$ at 5 dB. The superior performance of our method is reasonable because our method can be interpreted as an integration of the higher-order cumulants, spare autoencoders and a softmax classifier. The higher-order cumulants suppress Gaussian noise [19]. The spare autoencoder extracts features from the higher-order cumulants [15]. Then, the softmax classifier achieves maximum likelihood classifications [20]. Our method performs better than KNN, although KNN also utilizes higher-order cumulants.

V. Conclusion

The aim of this paper is to achieve a digital modulation classification in application of CR. A completed classification scheme has been proposed. Simulation results reveal that this method performs better than the KNN method and ML method.

Also, although training phase is necessary, process of classification is quite rapid once training phase is finished. Concretely, the classifier we used here consumes time for training phase, but the activation values of all layers are universal as long as the network and modulation pool are decided. This is a significant advantage over other methods when rapid processing is expected.

References

[1] B. Ramkumar, T. Bose, and M.S. Radenkovic. Combined blind equalization and automatic modulation classification for cognitive radios. In *IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop (DSP/SPE)*, pages 172–177, Marco Island, USA, 2009.

[2] X. Zhu and T. Fujii. Modulation classification in cognitive radios for satellite and terrestrial systems. In *IEEE International Conference on Communications Workshop*, pages 1612–1616, London, UK, 2015.

[3] X. Zhu and T. Fujii. A novel modulation classification method in cognitive radios based on features clustering of time-frequency. In *IEEE Radio and Wireless Symposium (RWS)*, pages 45–47, Austin, USA, 2016.

[4] L. Hong and KC. Ho. An antenna array likelihood modulation classifier for bpsk and qpsk signals. In *IEEE Conference on Military Communications*, pages 647–651, Anaheim, CA, 2002.

[5] A. Abdi, OA. Dobre, R. Choudhry, Y. Bar-Ness, and W. Su. Modulation classification in fading channels using antenna arrays. In *IEEE Conference on Military Communications*, pages 211–217, Monterey, CA, 2004.

[6] A. Swami and BM. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, 48(3):416–429, 2000.

[7] A. Swami, S. Barbarossaand, and BM. Sadler. Blind source separation and signal classification. In *the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers,*, pages 1187–1191, Pacific Grove, CA, 2000.

[8] W. Dai, Y. Wang, and J. Wang. Joint power estimation and modulation classification using second- and higher statistics. In *IEEE Wireless Communications and Networking Conference*, pages 155–158, Orlando, USA, 2002.

[9] A. Hazza, M. Shoaib, SA. Alshebeile, and A. Fahad. An overview of feature-based methods for digital modulation classification. In *1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, pages 1–6, Sharjah, UAE, 2013.

[10] A. Nandi and EE. Azzouz. Algorithms for automatic modulation recognition of communication signals. *IEEE Transactions on Communications*, 46(4):431–436, 1998.

[11] P. Vincent, H. Larochelle, Y. Bengio, and PA. Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1096–1103, Helsinki, Finland, 2008.

[12] WA. Gardner and CM. Spooner. The cumulant theory of cyclostationary time-series. i. foundation. *IEEE Transactions on Signal Processing*, 42(12):3387–3408, 1994.

[13] W. Su. Feature space analysis of modulation classification using very high-order statistics. *IEEE Communications Letters*, 17(9):1688–1691, 2013.

[14] Bengio, Yoshua, Lamblin, Pascal, Popovici, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[15] Bengio and Yoshua. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[16] C. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, USA, 2006.

[17] MW. Aslam and AK. Nandi. Automatic modulation classification using combination of genetic program and knn. *IEEE Transactions on Wireless Communications*, 11(8):2742–2750, 2012.

[18] W. Wei and J.M. Mendel. Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, 48(2):189C193, 2000.

[19] B. M. Sadler, G. B. Giannakis, and Keh-Shin Lii. Estimation and detection in non-gaussian noise using higher order statistics. *IEEE Transactions on Signal Processing*, 42(10):2729–2741, 1994.

[20] I. T. Nabney. Efficient training of rbf networks for classification. In *Ninth International Conference on Artificial Neural Networks*, pages 210–215, Edinburgh, UK, 1999.