

RESEARCH

Open Access



# A new method to solve the problem of facing less learning samples in signal modulation recognition

Sibao Fu<sup>1</sup> and Xiaokai Liu<sup>2\*</sup>

## Abstract

In machine learning method, the number of training samples is an exceedingly important factor determining the learning system's robustness. In our previous researches (Liu et al., *J. Syst. Eng. Electron.* 27.2:333–342, 2016; Liu et al., *IET Commun.* 11.7:1000–1007, 2017), the extreme learning machines (ELMs) have proven to be an effective and time-saving learning method for pattern classification and the signal modulation recognition. ELMs are utilized to supervised learning issues principally on signal modulation recognition. In this thesis, ELMs are extended for semi-supervised tasks that are based on the manifold regularization, therefore greatly enlarging ELMs' applicability. This article evolves countermeasures to the less training samples which mitigate the modulation recognition efficacy and demonstrates the robustness of semi-supervised learning for signal classification in AWGN and Rayleigh-fading channels.

**Keywords:** Modulation recognition, Machine learning, Less training data, Semi-supervised ELM

## 1 Introduction

With the evolution of the modernistic communication technologies, the automatic recognition of modulation signal is an significant issue in fast-growing diversified applications like satellite communications, scientific probes, and military communications [1–4]. From former study consequences [5–8], automatic modulation classification (AMC) is an implement to identify unknown or partially known signals by applying a brief sequence of the signal. Early AMC facilities worked as an ordinary demodulator to recover AM and FM automatically, but the algorithms became more sophisticated with the digital waveforms' emergence. there are two popular and effective statistical implements for sorting digital modulation schemes: the higher-order statistics (HOS) method and the maximum likelihood method. So far, the signal sorting algorithms are generally parted into two categories: feature-based methods and likelihood-based methods.

For the feature-based approaches, the classifier is the most vital part of the modulation recognition process.

And the classification algorithm has received widespread attention of many scholars and has been applied to many practical engineering, such as the single-layer feed-forward networks (SLFNs), which was intensively scrutinized during the past times. The most existent algorithms for training SLFNs adopt gradient methods for optimizing the weights in the training process, the two famous ways are back-propagation algorithm [8] and the Levenberg–Marquardt algorithm [9]. During the training process [10], the method takes the backward elimination or forward selection mechanisms dynamically to generate the network. The support vector machines (SVMs) [11, 12], supposed to be one of the most successful means for training the SLFNs network, is a maximum margin classifier derived under the framework of structural risk minimization (SRM). On account of its balanced generalization and simplicity of performance, the method SVMs was widely studied and utilized to various kinds of fields. Huang et al. [13, 14] recently raise a new way to train the SLFNs network, this method is called extreme learning machines (ELMs). In the ELM process, the output weights between the output layer and the hidden layer were exclusively modernized, while the parameters of the input weights and biases of the hidden layer were randomly created.

\*Correspondence: [fusibao@bupt.edu.cn](mailto:fusibao@bupt.edu.cn)

<sup>2</sup>School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUP), Beijing, China  
Full list of author information is available at the end of the article

By setting the prediction error squared loss, output training weights go into a regularized least square (or ridge regression) issue which can efficiently be solved in closed form. It has been illustrated that yet without modernizing the hidden layer parameters, the SLFN with randomly generated hidden neurons and tunable output weights maintains its universal approximation power [15–17].

All the methods mentioned above are verified in many fields for the modulation recognition. But they all need enough training data, which is called supervised learning classifiers. In the signal modulation recognition, getting tags for completely supervised learning is time-consuming and expensive, while a lot of unlabeled data are easy and tinny to collect. In this paper, ELMs are extended to deal with the modulation recognition when the training data is not adequate. Unlike from the existing works, we cover signal classification challenges in following parts:

- In comparison with the other modulation approaches, this paper concentrates on less sample data situation. The superiority of this algorithm lies in hidden nodes' random choices and analytically ascertains output weights, which leads to lower complexity.
- Unlike the existing works, this paper pays more attention to the channel environment. In the time-varying channel, frequency departure and time delay are significant parameters affecting the communication system performance. In this article, all study consequences are acquired in low SNR (–10 dB–10 dB) and Rayleigh fading channel: the maximal time delay was set  $10^{-3}$  s. The shift of the maximum Doppler frequency is 25 Hz and multipath fading conditions with complex channel environments.
- Feature-based means are more time-saving means for modulation recognition. The higher order moments (HOMs) and cumulants (HOCs) are utilized as the extraction features in this paper. All the picked features can guarantee that the later classifier can get the data in the real-time.

The remainder of this thesis is arranged as follows: Section defines model of the system and provides the relevant works. Part portrays the proposed modulation identification algorithm, section details the analysis process of the performance. Viewpoints and conclusions for the research work are eventually presented in part.

## 2 Preliminaries

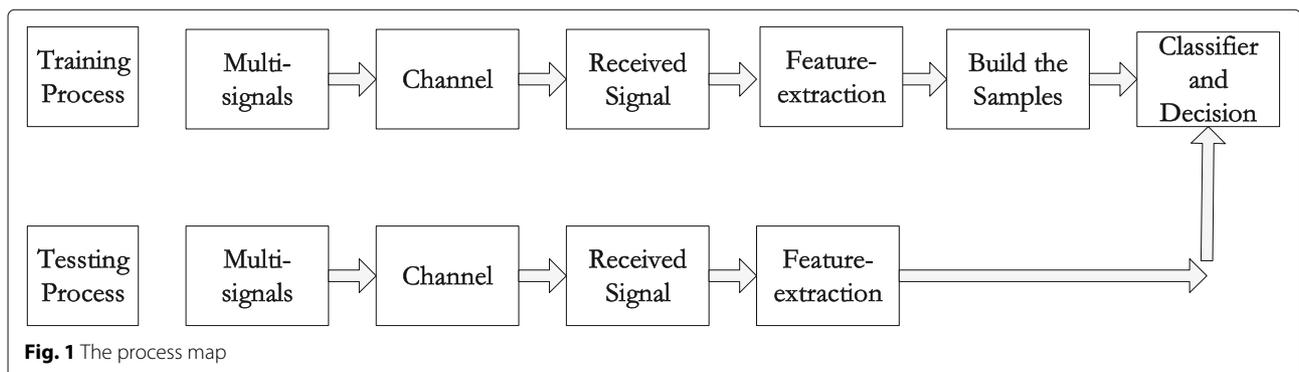
In this section, first of all, we define the time-varying Rayleigh fading channel model. In the channel model, the Gaussian noise and Rayleigh fading channel are considerable in our analysis. We add zero-mean white Gaussian interference to the transmitted signal. To demonstrate the robustness and dependability of the method, we set the lowest SNR to –10 dB and the time delay configuration up to  $10^{-3}$  s in multipath environment associated with Rayleigh fading. The environment channel model is supplied through the impulse response:

$$h(t; \tau) = \sum_{l=1}^N a_l(t) e^{j\theta_l(t)} \delta(\tau - \tau_l). \quad (1)$$

$N$  stands for the number of path, and the path delay is  $\tau_l$ .  $a_l(t)$  satisfies with the distribution of Rayleigh, and the  $\theta_l(t)$  is  $[0, 2\pi]$  followed uniform random phase over, respectively. As the received signal will suffer a small-scale Doppler shift exclusively for the terrestrial personal and mobile communication condition,  $\tau_l$  is assumed to be fixed over code acquisition time.

## 3 Methods

Our plan is to separate signals of the networks based on the pattern recognition and machine learning. This method is divided into two subsystems, namely, the feature extraction part and the classification part. The proposed signal classification method, Process Map, is illustrated in Fig. 1. The multisignal flows reach the receiver through the Rayleigh fading or Gaussian channel. In the receiver, varying frequency signals are reduced to medium frequency or less, resulting in overlapping of the spectrum



[18]. The receiver should have the signal separated without any prior information. The machine learning part is with the duty to intelligently draw a lesson from the signal attributes. In the part of machine learning process, the ELM algorithm is applied for sorting.

### 3.1 Features extraction

One important part of modulation identification was how to choose the suited identification features. Previous works have revealed the following rule: higher-order cumulants (HOCs) and higher-order moments (HOMs) of the received signal are one of the most fantastic applicants for signal identification, assuming the signal  $x$  with  $N$  samples; the HOM with order  $k$  is defined by the following equation.

$$M_{km}(x) = E[x^{k-m}(x^*)^m] \quad (2)$$

$E[\bullet]$  is the total expectation formula. Also, if the signal  $x$  is zero-mean, the cumulants of order  $k$  is followed by:

$$C_{km}(x) = \text{Cum} \left[ \underbrace{x, \dots, x}_{k-m}, \underbrace{x^*, \dots, x^*}_m \right] \quad (3)$$

The relations between the higher moment and cumulants can be expressed as follows:

$$\text{Cum}[x_1, \dots, x_n] = \sum_{\phi} (\alpha - 1)! (-1)^\alpha \prod_{\nu \in \phi} E(\prod_{i \in \nu} x_i) \quad (4)$$

where  $\phi$  covers all list of all partitions of  $1, \dots, n$ ;  $\nu$  runs through the list of all blocks of the partition  $\phi$ . Taking the fourth order for example, if the signal  $x, y, z$ , and  $w$  are zero-mean, the cumulant is defined:

$$\begin{aligned} \text{Cum}[x, y, z, w] &= E(xyzw) - E(xy)E(zw) \\ &\quad - E(xz)E(yw) - E(xw)E(yz) \end{aligned} \quad (5)$$

Based on Eqs. (3–5), if signal  $y$  is zero-mean with  $N$  samples, the moments and the cumulant can be expressed as in [19].

### 3.2 Extreme learning machines

As in the paper [19, 20], signal modulation recognition using ELM method includes the following steps:

- Step1: Given a training set  $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$ , activation function  $g(x)$  and hidden node number  $\tilde{N}$ .
- Step2: Randomly assign input weight  $\mathbf{w}_i$  and bias  $b_i, i = 1, \dots, N$ .
- Step3: Calculate the hidden layer output matrix  $\mathbf{H}$ .
- Step4: Calculate the output weight  $\boldsymbol{\beta} = \mathbf{H}^T \mathbf{T}$ , where  $\mathbf{T} = [t_1, \dots, t_N]^T$ .
- Step 5: The network is applied in test learning process to get the performance.

### 3.3 Semi-supervised extreme learning machines

Semi-supervised machine learning is found in the following two assumptions:

- Both the labeled data  $\mathbf{X}_l$  and the unlabeled data  $\mathbf{X}_u$  are drawn from the same marginal distribution  $\rho_x$ .
- If two signal samples  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are similar to each other, then the conditional probabilities  $P(\mathbf{y}|\mathbf{x}_1)$  and  $P(\mathbf{y}|\mathbf{x}_2)$  should be close as well.

To compel this presumption on the signal samples, the framework of the manifold regularization was proposed to make the following cost function minimum:

$$L_m = \frac{1}{2} \sum_{i,j} w_{ij} \|P(\mathbf{y}|\mathbf{x}_i) - P(\mathbf{y}|\mathbf{x}_j)\|^2 \quad (6)$$

$w_{ij}$  is the similarity between two patterns  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ;  $\mathbf{W} = [w_{ij}]$  is usually sparse. Because of the difficulties to get the conditional possibility, the following approach Eq. (6) should be adapted as the expression:

$$\hat{L}_m = \frac{1}{2} \sum_{i,j} w_{ij} \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\|^2 \quad (7)$$

where  $\hat{\mathbf{y}}_i$  and  $\hat{\mathbf{y}}_j$  are the predictions with respect to pattern  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , respectively. It is direct to simplify the above expression in a matrix shape:

$$\hat{L}_m = \text{Tr}(\hat{\mathbf{Y}}^T \mathbf{L} \hat{\mathbf{Y}}) \quad (8)$$

where  $\text{Tr}(\cdot)$  stands for the trace of a matrix,  $\mathbf{D}$  is a diagonal matrix with its diagonal elements  $D_{ii} = \sum_{j=1}^{l+u} w_{ij}$ , and  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  is graph Laplacian.

In the semi-supervised setting, we have few labeled data and a lot of unlabeled data. The labeled data is denoted in the training set as  $\{\mathbf{X}_l, \mathbf{Y}_l\} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^l$ , and unlabeled data is set as  $\mathbf{X}_u = \{\mathbf{x}_i\}_{i=1}^u$ ,  $l$  and  $u$  stand for the number of labeled and unlabeled data, respectively.

The suggested SS-ELM involves the manifold regularization to leverage unlabeled data to improve the classification precision while labeled signal samples are scarce. By changing the general ELM formulation [20], we furnish formulation of SS-ELM as:

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^{n_h \times n_0}} \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} \sum_{i=1}^l \mathbf{C}_i \|\mathbf{e}_i\|^2 + \frac{\lambda}{2} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}) \quad (9)$$

$$\mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} = \mathbf{y}_i^T - \mathbf{e}_i^T, \quad i = 1, \dots, l \quad (10)$$

$$\mathbf{f}_i = \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta}, \quad i = 1, \dots, l+u \quad (11)$$

where  $\mathbf{L} \in \mathbf{R}^{(l+u) \times (l+u)}$  is the graph Laplacian came from both labeled and unlabeled signal samples, and  $\mathbf{F} \in \mathbf{R}^{(l+u) \times n_0}$  is the output network matrix, which its  $i$  row is equal to  $\mathbf{f}(\mathbf{x}_i)$  and  $\lambda$  is the trade-off parameter.

Similar to the weighted ELM algorithm (W-ELM) presented in [18], in this paper, consider different punishment coefficient  $C_i$  on the forecast errors with regard to patterns from different categories. Assume that  $x_i$  is close to class  $t_i$ , which has  $N_{t_i}$  training kinds, then we associate  $e_i$  with a punishment of  $C_i = C_0/N_{t_i}$ , where  $C_0$  as in traditional ELMs is a user-defined parameter.

We replace the restrictions into the objective function and re-write the above formulation in the matrix state:

$$\min_{\beta \in \mathbf{R}^{n_h \times n_0}} \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \|\mathbf{C}^{\frac{1}{2}}(\hat{\mathbf{Y}} - \mathbf{H}\beta)\|^2 + \frac{\lambda}{2} \text{Tr}(\beta^T \mathbf{H}^T \mathbf{L} \mathbf{H} \beta) \quad (12)$$

where  $\hat{\mathbf{Y}} \in \mathbf{R}^{(l+u) \times n_0}$  is the augmented training samples and its first  $l$  rows are equal to  $\mathbf{Y}_l$  and the rest equal to 0, and  $\mathbf{C}$  is a  $(l+u) \times (l+u)$  diagonal matrix with its first  $l$  diagonal elements  $[\mathbf{C}]_{ii} = C_i$  and the rest equals to 0.

Again, the objective function gradient can be computed with regard to  $\beta$ :

$$\nabla_{\text{LSS-ELM}} = \beta + \mathbf{H}^T \mathbf{C}(\hat{\mathbf{Y}} - \mathbf{H}\beta) + \lambda \bullet \mathbf{H}^T \mathbf{L} \mathbf{H} \beta \quad (13)$$

By arranging the gradient close to zero, the solution for the SS-ELM is obtained by:

$$\beta^* = (\mathbf{I}_{n_h} + \mathbf{H}^T \mathbf{C} \mathbf{H} + \lambda \mathbf{H}^T \mathbf{L} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C} \hat{\mathbf{Y}} \quad (14)$$

on the condition that labeled samples are fewer than the hidden neurons; also, this is the ordinary situation in SSL, take the following alternative solution:

$$\beta^* = \mathbf{H}^T (\mathbf{I}_{l+u} + \mathbf{C} \mathbf{H} \mathbf{H}^T + \lambda \mathbf{L} \mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \hat{\mathbf{Y}} \quad (15)$$

where  $\mathbf{I}_{l+u}$  is the identity matrix with the dimension  $l+u$ .

In conclusion, signal recognition that is based on SS-ELM consists of the following steps:

Input:

Labeled patterns  $\{\mathbf{X}_l, \mathbf{Y}_l\} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^l$ ,

Unlabeled patterns  $\mathbf{X}_u = \{\mathbf{x}_i\}_{i=1}^u$ ,

Output: the mapping function of SS-ELM:  $\mathbf{f}: \mathbf{R}^{n_i} \rightarrow \mathbf{R}^{n_o}$

- Step 1: Construct the graph Laplacian  $\mathbf{L}$  from both  $\mathbf{X}_l$  and  $\mathbf{X}_u$ .
- Step 2: Initiate an ELM network of  $n_h$  hidden neurons with random input weights and biases and calculate the output matrix of the hidden neurons  $\mathbf{H} \in \mathbf{R}^{(l+u) \times n_h}$ .
- Step 3: Choose the trade-off parameter  $C_0$  and  $\lambda$ .
- Step 4: if  $n_h \leq N$   
 Compute the output weights  $\beta$  using 14  
 or else  
 compute the output weights  $\beta$  using 15

return the network function  $\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\beta$ .

## 4 Experimental result and discussion

In this part, we take the following configurations into account: signals transmit though AWGN or Rayleigh fading environments with the SNR varying from  $-10$  dB to  $10$  dB stepped by  $2$  dB in interval. For Rayleigh fading channel, the channel maximum Doppler frequency is  $15$  Hz, the time delay is  $10^{-4}$  s. The symbol rate was  $200$  bps. The carrier frequency was set  $20$  KHz, and sampling frequency is  $40$  KHz. The number of unlabeled sample varies in the different application scenarios. All inputs have been normalized to the range of  $[-1, 1]$ .

In the simulations configuration, ELM is composed of  $100$  hidden nodes, making the algorithm furnish brilliant generalization performance at fast learning speed.

In what comes along, we will present three modulation types in the simulation procedure, that is  $\phi_1 = \{\text{BPSK}, \text{4PSK}, \text{8PSK}\}$ ,  $\phi_2 = \{\text{4ASK}, \text{8ASK}\}$ ,  $\phi_3 = \{\text{16QAM}, \text{64QAM}\}$ , and  $\phi_4 = \{\phi_1, \phi_2, \phi_3\}$ . All the results are based on  $1000$  Monte Carlo trials for each modulation scheme. The probability of identification is given in percentage and estimated by  $\frac{N_t}{N_{\text{total}}} \times 100$ , where  $N_{\text{total}}$  is the total number of trials and  $N_t$  is the number of trials for which the modulation is correctly identified.

### 4.1 Performance results of HOC and HOM

In the SISO system, the HOMs and HOCs are employed by using the hierarchical modulation identification to attain wonderful results. It was illustrated that modulation identification probability is a nonlinear function that is been connected with the SNR, the number of symbols, and the modulation types. It is understandable that raising the number of symbols will increase the performance. Pay attention that identification probability is an increasing function of SNR while bringing a known modulation pool into consideration and an enough quantity of symbols for SISO system [21]. Theoretic values of a number of HOCs and HOMs are provided in Table 1. In Table 1, we can come to realize how they can discriminate the different modulation schemes. These theoretic values are figured for dissimilar digital modulation constellations below unit variance symbols' restraints and noise free instance.

### 4.2 Classification performance for each communication scheme in AWGN and Rayleigh fading channel

For  $\phi_1, \phi_2$ , and  $\phi_3$ , we show the performance results of the identification probability in the AWGN and Rayleigh fading channel in a single-path environment. For the intra-class recognition, we generate signal feature sets for the experiment. For  $\phi_1$ , we generate  $300$  groups of labeled samples and  $900$  groups of unlabeled samples to train the SEM-ELM neural network to get the parameters needed for the subsequent testing. In the testing process, we also randomly select  $1200$  groups to test the neural network. For  $\phi_2$  and  $\phi_3$  separately, we generate

**Table 1** Theoretical value of some HOMs and HOCs without noise

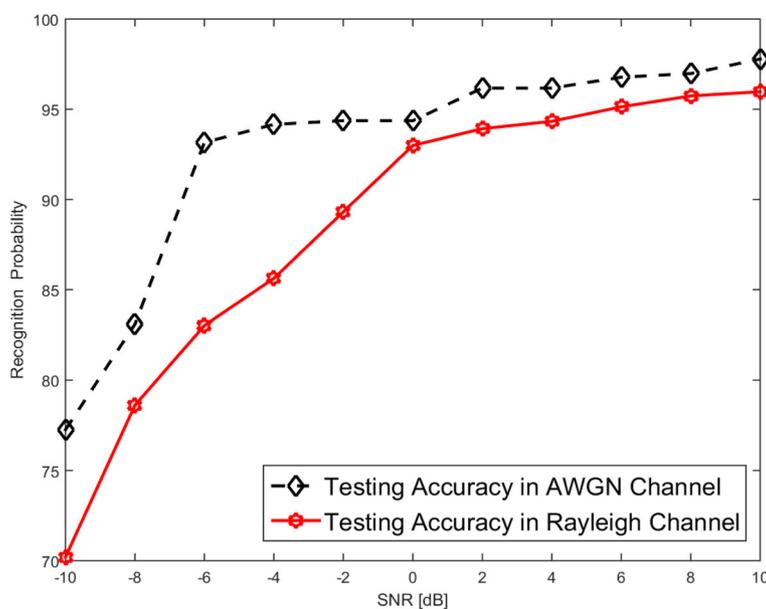
	BPSK	QPSK	8PSK	16QAM	64QAM	4ASK	8ASK
C21	1	1	1	1	1	0.58	0.53
M40	1	-1	0	0.68	0.619	0.89	0.73
M20	1	0	0	0	0	0.58	0.53
C40	2	1	0	0.68	0.619	0.12	0.13
M42	1	1	1	1.32	1.38	0.89	0.73
M21	1	1	1	1	1	0.58	0.53
C42	-2	-1	-1	-0.68	-0.62	-0.12	-0.13
C42/C21	-2	-1	-1	-0.68	-0.62	-0.12	-0.46
M63	1	1	1	1.96	2.23	1.77	1.28
C63	4	4	4	2.08	1.79	-0.53	-0.39
C63/C21	4	4	4	2.08	1.79	-2.73	-2.56

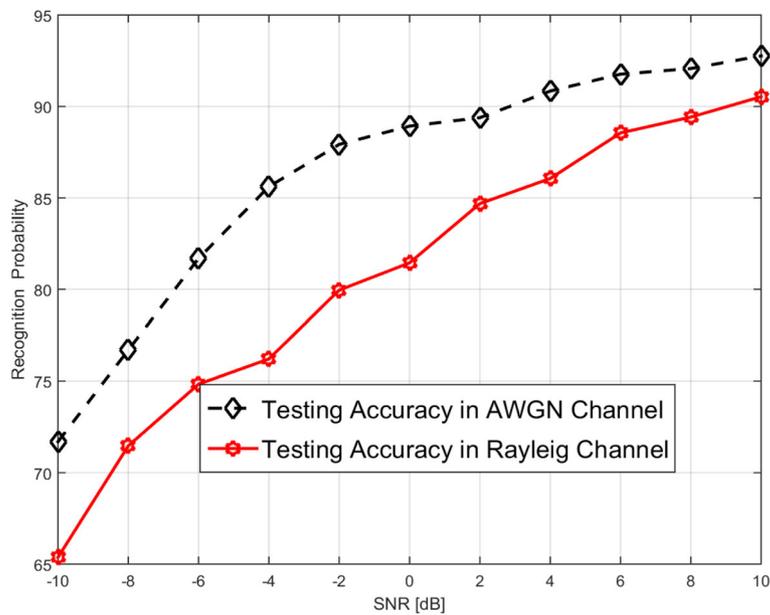
200 groups labeled samples and 600 groups of unlabeled samples to form the training data. Respectively, in the testing process, we choose 800 groups of samples to test the performance. Figures 2, 3, and 4 separately show the performance for the scenarios of  $\phi_1$ ,  $\phi_2$ , and  $\phi_3$ , respectively. As could be observed: (1) The SNR has a negative impact on the modulation recognition accuracy. As the SNR increases, modulation identification probability become more acceptable. For the PSKS, modulation identification probability reached its upper bound at 8 dB, more than 93%. While the FSKS and QAMS were slightly worse, that is owing to the channel environments destroyed the original signal structure. (2) The SS-ELM method, serving

as perspective method, shows a strong robustness when the SNR is more than 4 dB in all the signal modulation recognition.

#### 4.3 Classification performance for mixed communication schemes in AWGN and Rayleigh fading channel

The SS-ELM algorithm performance for each kind of signals attains substantial results when the SNR is greater than 6 dB. These consequences are checked from our former performance analysis of Figs. 2, 3, and 4. In this subsection, we organized blind signals recognition for the mixed signals. Other than the procedure noted in the “Classification performance for each communication scheme in AWGN and Rayleigh fading channel” section, in the blind recognition method, the data sets are picked randomly from the feature sets. In machine learning training process, the labeled sample is 100 groups, and the unlabeled is 700 groups. Figure 5 reveals the full class modulation identification results in the AWGN channel, and Fig. 6 demonstrates the algorithm property in the Rayleigh fading channel. In Figs. 5 and 6, we can establish the following conclusions: (1) In comparison with the “Classification performance for each communication scheme in AWGN and Rayleigh fading channel” section, the blind recognition probability for the mixed schemes demonstrate performance degradation’s indications, specially while the SNR is below 4 dB. (2) In the mixed signal modulation recognition, the learning performance is connected with the SNR closely; with the SNR’s rise, the correct probability become better and better, when the SNR is 10 dB, the identification precision gets up to an

**Fig. 2**  $\phi_1$  in the AWGN and Rayleigh fading channel



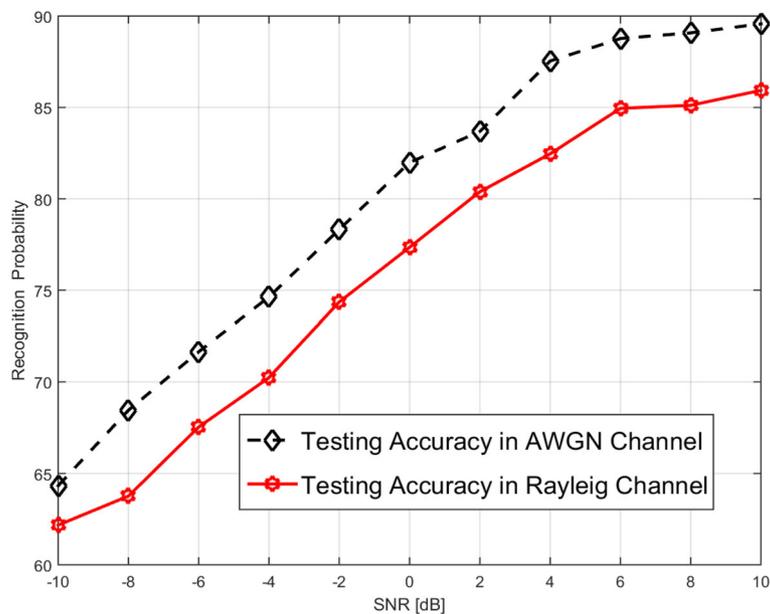
**Fig. 3**  $\phi_2$  in the AWGN and Rayleigh fading channel

satisfactory level, achieving 95% in AWGN and 92% in the Rayleigh fading channel.

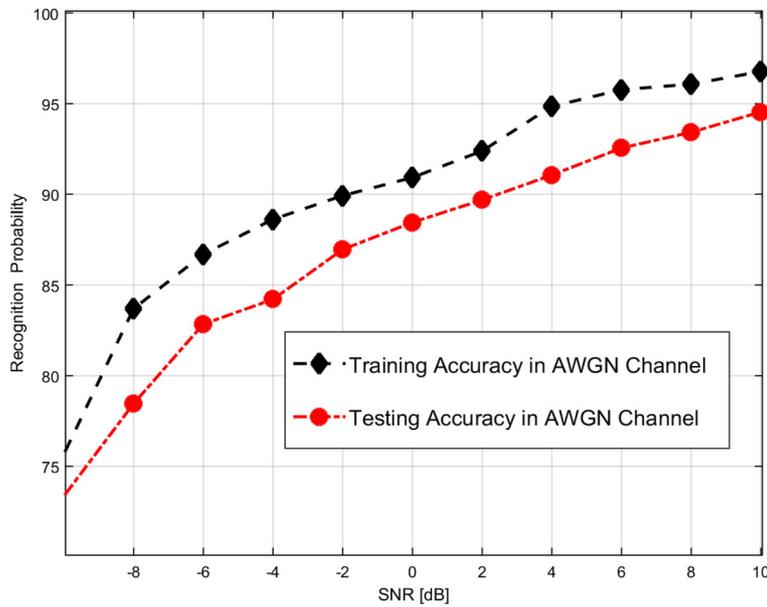
**4.4 The number of labeled samples effect for blind signal recognition**

In the semi-supervised system, the number of the labeled samples may influence the communication performance reliability greatly. In this section, we do check the SEMI-

ELM algorithm with varying number of the labeled data. The simulated environment include the objective signal is  $\phi_1 = \{BPSK, 4PSK, 8PSK\}$ ; set the labeled data NUM = 50, 100, 200, 250, 300, 350, and 400; the training data sets is 1200; the unlabeled data is 1200 NUM; in the testing process, we set the testing sample setting to 1200 to test the semi-supervised ELM classifier. The SNR is SNR = 0, and the results are presented in Fig. 7.



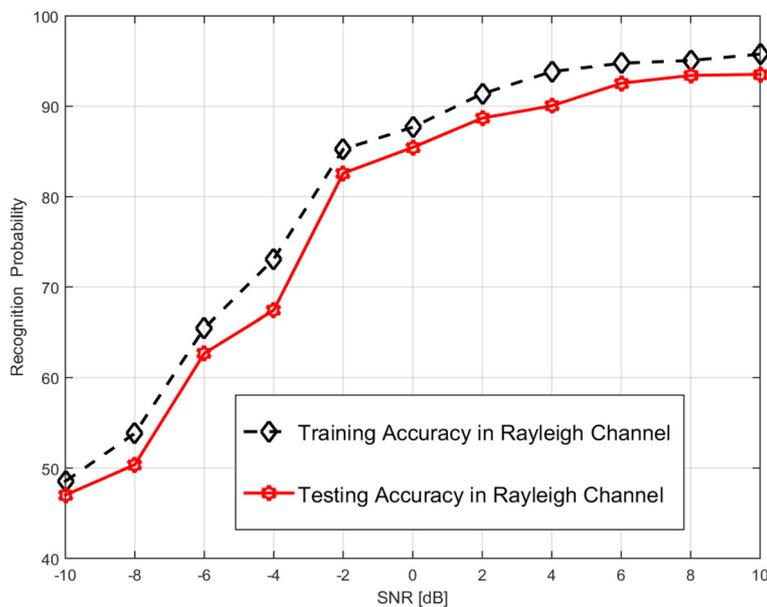
**Fig. 4**  $\phi_3$  in the AWGN and Rayleigh fading channel



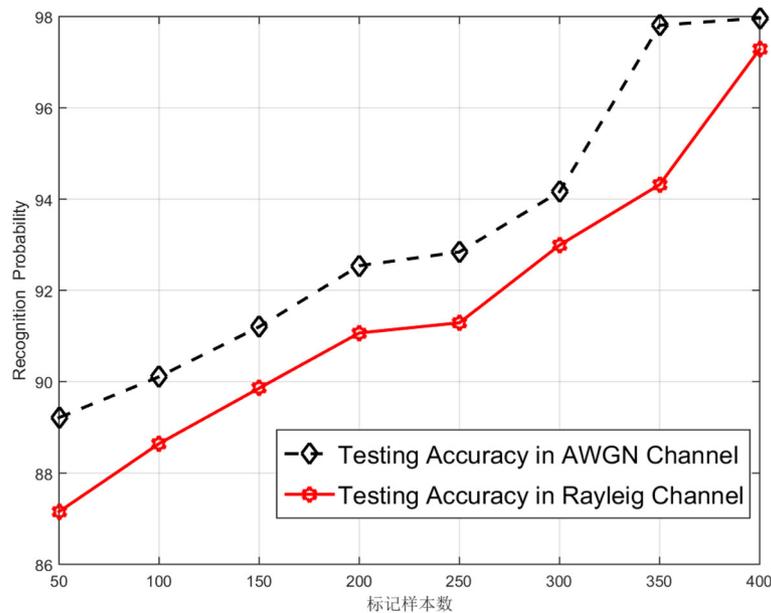
**Fig. 5**  $\phi_4$  in the AWGN channel

In Fig. 7, we can conclude (1) when SNR = 0 dB, for  $\phi_1 = \{BPSK, 4PSK, 8PSK\}$ , the SNR has a little effect on the signal modulation recognition, but the recognition accuracy has some relations with the number of labeled samples. When the labeled sample occupy 1/3 of all the training data, the probability of the system identification will reach 98%, which is an acceptable outcome. (2) When

the labeled data is less, such as 50, the identification accuracy is about 90%; the main reason is that the channel affects the unlabeled data structure. (3) Above all, the number of the labeled samples has some relation with the modulation recognition result, when have difficulty to get enough labeled training datas, the SS-ELM might be a good choice.



**Fig. 6**  $\phi_3$  in the Rayleigh fading channel



**Fig. 7** The number of labeled sample effect for blind signal recognition

## 5 Conclusions

In this paper, we explore the machine learning's utilization and pattern recognition for the signal classification when facing less training samples. The features extraction came from the HOCs and HOMs. The SS-ELM algorithm outperforms all researched scenarios with rapider convergence and a lower computational complexity with no iterative tuning. The proposed algorithms are examined through different number of labeled samples to ratify the reliability. Subsequently, the implied algorithm is examined through AWGN and Rayleigh fading channels, and they show a very impressive skill of recognizing different modulation schemes with high preciseness in lower SNR time-varying channel. The succeeding work will focus on the unsupervised extreme learning machines for more brilliant generalization in the signal identification.

### Abbreviations

AM: Amplitude modulation; AMC: Automatic modulation classification; AWGN: Additive white Gaussian noise; ELM: Extreme learning machines; FM: Frequency modulation; HOCs: Higher-order cumulants; HOMs: Higher-order moments; HOS: Higher-order statistics; SLFNs: Single-layer feed-forward networks; SNR: Signal-to-noise ratio; SRM: Structural risk minimization; SS-ELM: Semi-supervised extreme learning machines; SVMs: Support vector machines

### Acknowledgements

Not applicable

### Authors' contributions

FSB and LXX conceived and designed the study. LXX performed the experiments. FSB provided the mutants. FSB and LXX wrote the paper. Both authors read and approved the manuscript.

### Funding

This work was supported by the Natural Science Foundation of China (NSFC) under Grant No. 61471061.

### Availability of data and materials

Data sharing is not applicable to this article as no data sets were generated or analysed during the current study.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Economics and Management, Beijing University of Posts and Telecommunications (BUPT), Beijing, China. <sup>2</sup>School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), Beijing, China.

Received: 20 June 2019 Accepted: 18 December 2019

Published online: 06 January 2020

### References

- W. Su, J. A. Kosinski, M. Yu, in *Systems, Applications and Technology Conference*. Dual-use of modulation recognition techniques for digital communication signals (IEEE, 2006), pp. 1–6
- Y. Jiang, Z. Y. Zhang, P. L. Qiu, in *Military Communications Conference, 2004*. Modulation classification of communication signals, Vol. 3 (Milcom IEEE, 2004), pp. 1470–1476
- O. A. Dobre, A. Abdi, Y. Bar-Ness, et al., Survey of automatic modulation classification techniques: classical approaches and new trends. *IET Commun.* **1**(2), 137–156 (2007)
- M. M. Shakra, E. M. Shaheen, H. A. Bakr, et al., in *Radio Science Conference*. C3. Automatic digital modulation recognition of satellite communication signals (IEEE, 2015), pp. 118–126
- J. L. Xu, W. Su, M. Zhou, Likelihood-ratio approaches to automatic modulation classification. *IEEE Trans. Syst. Man Cybern. Part C.* **41**(4), 455–469 (2011)
- O. A. Dobre, M. Oner, S. Rajan, et al., Cyclostationarity-based robust algorithms for QAM signal identification. *IEEE Commun. Lett.* **16**(1), 12–15 (2012)
- S. Wei, Feature space analysis of modulation classification using very high-order statistics. *IEEE Commun. Lett.* **17**(9), 1688–1691 (2013)
- D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors. *Read. Cogn. Sci.* **3**(23), 399–421 (1986)
- M. T. Hagan, M. B. Menhaj, Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural. Netw.* **5**(6), 989–993 (2002)

10. G. Huang, S. Song, C. Wu, Orthogonal least squares algorithm for training cascade neural networks. *IEEE Trans. Circ. Syst. I Regular Pap.* **59**(11), 2629–2637 (2012)
11. V. N. Vapnik, The nature of statistical learning theory. *IEEE Trans. Neural Netw.* **38**(4), 409–409 (1997)
12. C. Cortes, V. Vapnik, Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
13. G. B. Huang, Q. Y. Zhu, C. K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks. *Proc. Int. Joint Conf. Neural Netw.* **2**, 985–990 (2004)
14. G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications. *Neurocomputing.* **70**(1), 489–501 (2006)
15. G. B. Huang, L. Chen, C. K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Netw.* **17**(4), 879–892 (2006)
16. G. B. Huang, Y. Q. Chen, H. A. Babri, Classification ability of single hidden layer feedforward neural networks. *IEEE Trans. Neural Netw.* **11**(3), 799–801 (2000)
17. R. Zhang, Y. Lan, G. B. Huang, et al., Universal approximation of extreme learning machine with adaptive growth of hidden nodes. *IEEE Trans. Neural Netw. Learn. Syst.* **23**(2), 365 (2012)
18. (Publishing House of Electronics Industry; Pearson Education Asia, 2007)
19. X. Liu, et al., Blind modulation classification algorithm based on machine learning for spatially correlated MIMO system. *IET Commun.* **11.7**, 1000–1007 (2017)
20. X. Liu, et al., Robust signal recognition algorithm based on machine learning in heterogeneous networks. *J. Syst. Eng. Electron.* **27.2**, 333–342 (2016)
21. A. Swami, B. M. Sadler, Hierarchical digital modulation classification using cumulants. *IEEE Trans. Commun.* **48**(3), 416–429 (2000)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---