URSI GASS 2021, Rome, Italy, 28 August - 4 September 2021

Radio Waveforms Classification via Deep Q Learning Network

Siqi Lai [(1)], Mingliang Tao[(1)]*, Xiang Zhang[(2)], Ling Wang[(1)]
(1) Northwestern Polytechnical University, Xi'an 710072, China
(2) Shanghai Institute of Satellite Engineering, Shanghai 200090, China

Abstract

Radio waveforms classification plays a foundation role in cognitive radio, which promises a broad prospect in spectrum monitoring and management. In this paper, a radio waveforms classification via deep Q learning is proposed, in which a deep reinforcement learning agent is trained to classify signal modulation type. Differ from the widely applied deep learning strategy, the proposed method has strong self-learning decision-making ability, which can find the optimal strategy by trial and error. The simulation results show that it can realize classification of radio signal modulation type with high accuracy.

1 Introduction

With the proliferation of radio technology, the electromagnetic spectrum is gradually getting crowded. Accurate identification of radio waveforms is an important prerequisite for spectrum awareness [1]. In traditional radio waveforms identification method, the sampled signal is generally transformed into representative domains and the distinctive features are extracted. Then the classification is realized by applying machine learning classifiers including support vector machine (SVM), k-nearest neighbors (KNN) and artificial neural networks (ANN) [2]. However, these hand-engineered feature extraction and selection rely on professional experience, which is difficult to reflect the deep features of signals and poor in adaptability to complex scenarios.

As a branch of machine learning, deep learning can actively learn the intrinsic correlation of radio signals, extract the hidden features of signals through multi-layer neural networks, and realize the signal classification in an end-to-end learning mode. Typical network architectures have been successfully applied for classifying modulated signals, such as sparse autoencoders and convolutional neural network (CNN) [3]. However, the limitation of existing deep learning-based modulation methods is that the promising results are obtained under the condition of large amount of training samples and high SNRs.

Due to the limitation of deep learning, reinforcement learning can produce fully autonomous agents that interact with their environments to learn optimal behaviors and improve over time through trial and error [4]. In [5], a deep Q-Learning network (DQN) is proposed, which combines the perception ability of deep learning with reinforcement learning to realize the perception and decision-making of complex environment state. In [6], DQN is combined with the traditional exploration approach, making a robot explore unknown cluttered environment autonomously.

Considering the superiority of the unique learning mechanism, DQN is applied in radio waveforms classification in this paper. In DQN, agent selects a modulation type based on the radio signal in the environment, then gets reward. The task of the agent is to learn a strategy which directs the model to classify the signal as the correct modulation type by maximizing the reward value.

2 Problem Formulation

A typical received radio signal can be expressed as

$$r(t) = h(t) * x(t) + n(t) \tag{1}$$

where $r(t)$ is the received signal, $h(t)$ is the channel impulse response, $x(t)$ is modulated signal and $n(t)$ is the thermal noise usually regarded as additive Gaussian white noise. The goal of modulation classification is managed to identify the modulation type of $x(t)$ only by the received $r(t)$ [7].

Modulation classification can be regarded as an N-class decision problem where input is a complex base-band time series representation of the received signal. A radio signal is sampled in-phased (I_l) and quadrature (Q_l) components at discrete time steps, obtaining a $1\times N$ complex valued vector r_l.

$$r_l = [I_l, Q_l]^T \tag{2}$$

where I_l and Q_l are the real value and imaginary value respectively. The classifier based on DQN directly input r_l,which is treated as an input dimension of 2 real valued inputs. Then r_l is fed into the DQN network as a $2\times N$ vector.

3 Methodology

Reinforcement learning is a learning mechanism to learn how to map state to action in order to obtain the maximum reward [4]. It obtains reward through the interaction between agent and environment, learns experience from it, and constantly improves the strategy. The agent obtains the corresponding behavior value through constantly try various actions. The optimal strategy will be obtained by maximizing the cumulative reward.

3.1 Markov Decision Process

Generally, reinforcement learning can be described as a Markov decision process (MDP) [4]. MDP is consisted of four elements (S, A, P, R). S is the set of environment states, and A is the set of actions that the agent may choose. $P(s_{next} \mid s,a)$ is the probability that the agent selects action a to make the environment state transfer to s_{next} when the environment is in state s. $R(s_{next} \mid s,a)$ is the reward value that agent chooses action a to make the environment state s transfer to s_{next}.

In MDP, action strategy is determined by action value, and $Q(s,a)$ is defined as the expected return for selecting action a in state s and then following a policy π, which can be expressed by Bellman equation [8].

$$Q(s,a) = R(s,a) + \gamma \sum P(s_{next} \mid s,a) Q(s_{next}, \pi(s_{next})) \quad (3)$$

where γ is the discount factor.

Under the condition of known action value, the optimal strategy can be found by maximizing the action value, according to the greedy decision-making method. The optimal strategy can be found if the optimal action value is known.

3.2 Q-learning

Q-learning is a common classical reinforcement learning algorithm. Q-learning stores action value $Q(s,a)$ in the form of a table, which approximates the correct action value function in an iterative way. The renewal formula is as follows,

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[R(s,a) + \gamma \max_a Q(s_{next}, a) - Q(s,a) \right] \quad (4)$$

where $\max_a Q(s_{next}, a)$ is the maximum Q value of s_{next}.

3.3 Deep Q Network

Q-learning is a good method to solve discrete problems. However, when solving complex problem, the table created by Q-learning will be very large, increasing the difficulty of training. In order to overcome the limitation, the DQN utilizes a neural network is used to approximate the Q-value function from the given input of state.

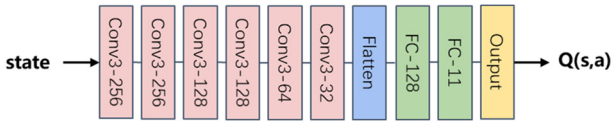


Figure 1. Neural Network of DQN.

Neural network is widely used in modeling, prediction and other areas due to its strong fitting ability [9]. DQN combines neural network and Q learning by using neural network to approximate the action value function [10]. In this paper, a six-layer CNN is utilized in DQN, as shown in Figure 1. The number of neurons for each layer is 256, 256, 128, 128, 64 and 32, and the value function is then output through two dense layers, in which the number of neurons is 128 and 11, respectively.

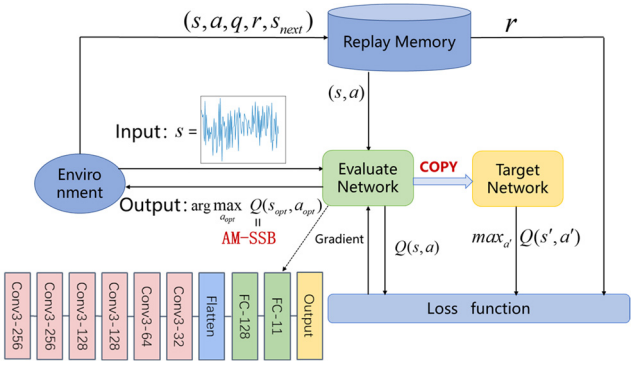


Figure 2. The framework of proposed DQN network for radio waveform classification.

Figure 2 shows the framework of proposed DQN network for radio waveform classification framework. DQN algorithm consists of two neural networks, i.e., *the evaluate network* and *the target network*, which have the same initial parameters. *The evaluate network* takes the environmental state s as the network input, calculates the estimated action value, and then selects the action with the highest value $Q(s,a,\theta)$. *The evaluate network* stores a group of data (s,a,r,s_{next}) composed of current state and reward information into *the replay memory*. The objective value of *the evaluate network* training is given by the target network combined with the reward r.

$$Q_{target} = R(s,a) + \gamma \max_a Q(s_{next}, a, \theta^{-}) \quad (5)$$

The latest parameters of *the evaluation network* will be copied to *the target network* after every certain number of iteration times.

4 Simulation Results and Analysis

In this part, a simulation experiment is realized to verify

the performance of the proposed radio waveform classification via DQN.

4.1 Experiment Settings

The radio signals are extracted from the public dataset RML2016.04C [11]. The dataset has 11 types of modulation, 8 digital modulation and 3 analog modulation, including digital modulation includes BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK and PAM4, while analog modulation includes WB-FM, AM-SSB and AM-DSB. These samples are uniformly distributed in SNR from -6dB to +18dB. Each sample has two data channels, i.e., real part and imaginary part, and each signal contains 128 sampling points. The data is represented as a matrix of 2×128 and the size of the whole data set is $105339\times2\times128$. In this experiment, totally 73737 samples are used for training, and 31602 samples are selected for testing and validation. Figure 3 shows spectrum of some particular samples with high SNR.

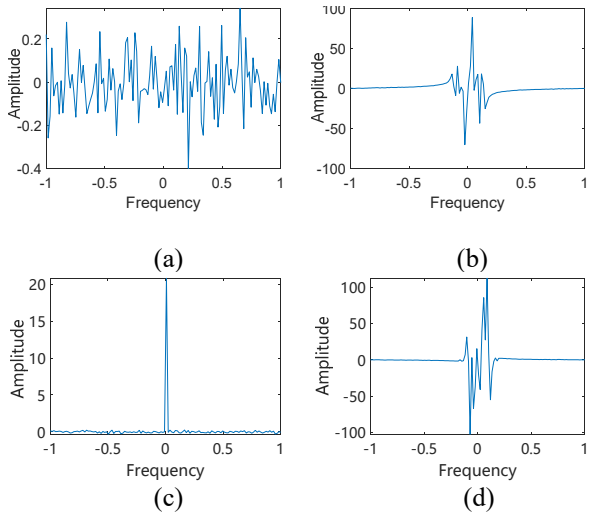


Figure 3. Spectrum of several typical samples with high SNR. (a) 8PSK, (b) BPSK, (c)AM-DSB, (d) PAM4.

In the course of the simulation, the agent action is chosen using the ε-greedy strategy. The ε in training can be expressed as,

$$\varepsilon = \max\left\{\varepsilon_{\min}, 1-\frac{1-\varepsilon_{\min} * step}{total}\right\} \tag{6}$$

where $\varepsilon_{\min}$=0.001, $step$ is the current iteration time and $total$ is the total iteration time.

Since the radio signals are all independently identically distributed, the discount factor γ of Q is set as 0. The learning rate α is set as 0.5. The capacity of replay memory is 1000 and batch size is 64. Use the Adam optimizer with a learning rate of $1e-6$ to optimize the deviation between the predicted Q value and target Q value. The maximal iteration times for convergence are set as 3000.

4.3 Results and Analysis

Generally, the change of reward can reflect the convergence rate of the DQN algorithm. Reward can guide DQN through training. In the training process, the maximum reward is set to 500. Reward gradually increased from -500 and close to 500, with the maximum reward of 482. Figure 4 shows the variation of reward with iterations. The reward of 3000 iteration times is plotted by average of every five iterations.

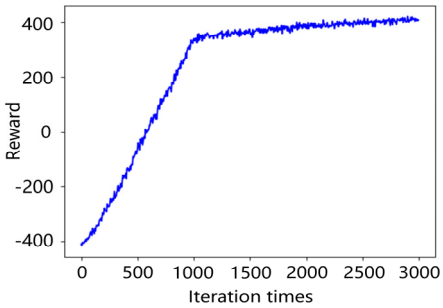


Figure 4. The variation of reward with iterations.

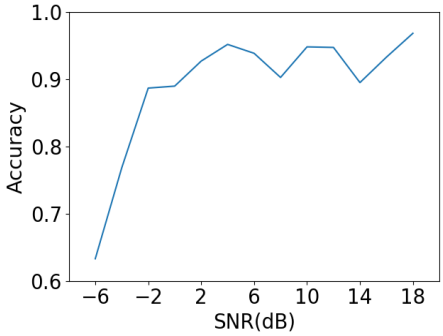
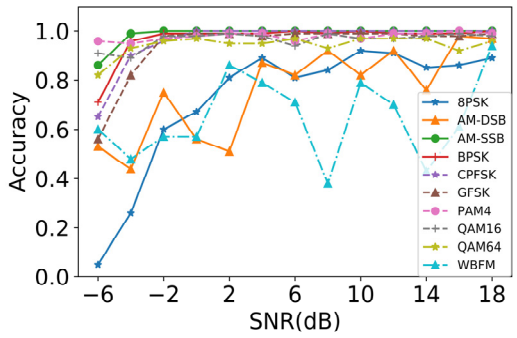


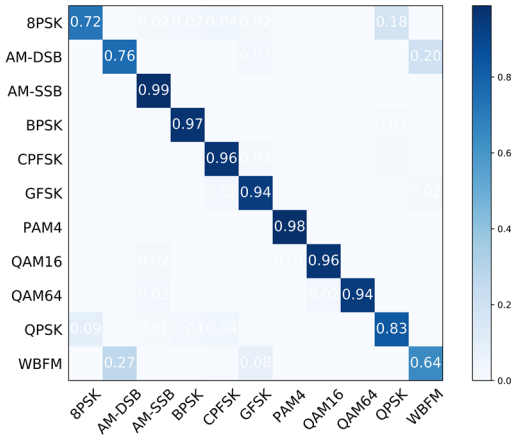
Figure 5 Average classification accuracy under different SNRs.

Figure 5 shows the accuracy of DQN classifier under different SNRs. It is shown that the DQN classifier achieves an average of 89.13% classification accuracy across all signal to noise ratios on the test dataset.

Figure 6 shows the accuracy of 11 modulation types under different SNRs and a confusion matrix. There is a clear diagonal in the confusion matrix. The remaining error are that 8PSK is misclassified as QPSK, and WBFM is misclassified as AM-DSB. The constellation of 8PSK includes the constellation of QPSK, so it is difficult to distinguish. In the analog voice signal, there is a silent period in the process of silence, where only a single carrier exists, making it difficult to distinguish between AM-DSB and WBFM.



(a)



(b)

Figure 6. Experimental results (a) Average accuracy of 11 modulation types under different SNRs, (b) Average confusion matrix of classification accuracy under SNR ranging from -6dB to +18dB.

5 Conclusion

This paper has presented a novel approach for radio waveforms classification using deep Q learning network. With the ability of self-learning and decision-making, the agent in DQN interacts with the environment constantly to learn the optimal strategy, which can achieve the accurate classification of radio signals.

6 Acknowledgements

This work is supported by National Natural Science Foundation of China under Grant No. 61801390, 61901377. This work is also supported by National Postdoctoral Program for Innovative Talents under grant BX201700199.

7 References

1. H. Dong, G. C. Sobabe, C Zhang, X. Bai, Z. Wang, L. Shuai, B. Guo, "Spectrum sensing for cognitive radio based on convolution neural network," *2017 10th International Congress on Image and Signal Processing, Bio Medical Engineering and Informatics*, October 2017, pp. 1-6, doi: 10.1109/CISP-BMEI.2017.8302117.

2. M. Nabian, "A Comparative Study on Machine Learning Classification Models for Activity Recognition," *Journal of Information Technology & Software Engineering*, **07**, 04, January 2017, doi: 10.4172/2165-7866.1000209.

3. W. Wang, Y. Yang, X. Wang, W. Wang, and J. Li, "Development of convolutional neural network and its application in image classification: a survey," *Optical Engineering*, **58**, 4, April 2019, doi: 10.1117/1.OE.58.4.040901.

4. K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Processing Magazine*, **34**, 6, pp. 26-38, November 2017, doi: 10.1109/MSP.2017.2743240.

5. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," December 2013, *arXiv preprint* arXiv:1312.5602.

6. F. Niroui, K. Zhang, Z. Kashino and G. Nejat, "Deep Reinforcement Learning Robot for Search and Rescue Applications: Exploration in Unknown Cluttered Environments," *IEEE Robotics and Automation Letters*, **4**, 2, pp. 610-617, April 2019, doi: 10.1109/LRA.2019.2891991.

7. H. Zhang, M. Huang, J. Yang, and W. Sun, "A Data Preprocessing Method for Automatic Modulation Classification Based on CNN," *IEEE Communications Letters*, doi: 10.1109/LCOMM.2020.3044755.

8. B. O'Donoghue, I. Osband, R. Munos, and V. Mnih, "The uncertainty bellman equation and exploration," *International Conference on Machine Learning*, July 2018, pp. 3836-3845.

9. O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, **4**, 11, November 2018, doi: 10.1016/j.heliyon.2018.e00938.

10. F. Tan, P. Yan, and X. Guan, "Deep Reinforcement Learning: From Q-Learning to Deep Q-Learning," *Neural Information Processing*, October 2017, pp. 475-483, doi: 10.1007/978-3-319-70093-9_50.

11. T. J. O'Shea, J. Corgan, and T. C. Clancy. "Convolutional Radio Modulation Recognition Networks," *International Conference on Engineering Applications of Neural Networks*, August 2016, pp. 213-226, doi: 10.1007/978-3-319-44188-7_16.