Uplink NOMA signal transmission with convolutional neural networks approach

LIN Chuan¹, CHANG Qing¹,*, and LI Xianxu²

¹. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China;
². State Grid Information and Telecommunication Branch, Beijing 100761, China

Abstract: Non-orthogonal multiple access (NOMA), featuring high spectrum efficiency, massive connectivity and low latency, holds immense potential to be a novel multi-access technique in fifth-generation (5G) communication. Successive interference cancellation (SIC) is proved to be an effective method to detect the NOMA signal by ordering the power of received signals and then decoding them. However, the error accumulation effect referred to as error propagation is an inevitable problem. In this paper, we propose a convolutional neural networks (CNNs) approach to restore the desired signal impaired by the multiple input multiple output (MIMO) channel. Especially in the uplink NOMA scenario, the proposed method can decode multiple users’ information in a cluster instantaneously without any traditional communication signal processing steps. Simulation experiments are conducted in the Rayleigh channel and the results demonstrate that the error performance of the proposed learning system outperforms that of the classic SIC detection. Consequently, deep learning has disruptive potential to replace the conventional signal detection method.

Keywords: non-orthogonal multiple access (NOMA), deep learning (DL), convolutional neural networks (CNNs), signal detection.

DOI: 10.23919/JSEE.2020.000068

1. Introduction

With the growth of users’ demand for higher communication speed and Internet of Everything, the need for massive connection and high throughput has been put forward. Unlike the conventional orthogonal multiple access (OMA) commonly used in the fourth-generation (4G) communication system, the non-orthogonal multiple access (NOMA) technique can improve the spectrum efficiency effectively via the non-orthogonal superposition of the signal power. Hitherto, the power-domain NOMA has evolved into code-domain and hybrid-domain NOMA, such as multi-user shared access (MUSA), sparse code multiple access (SCMA), and pattern division multiple access (PDMA). Here, we focus on uplink power-domain NOMA systems for its representativeness and universality.

Successive interference cancellation (SIC) is a classic and effective method for signal detecting. For both uplink and downlink, SIC for NOMA signal detection provides an optimal scheme to realize the maximization of the multi-user capacity [1]. The fundamental of SIC is that user equipment (UE) with worse channel state should be allocated with higher power and it should be decoded with higher priority at the receiver while other UE signals are treated as interference. Then the decoded signal needs to be modulated again and removed from the received signal so that the next highest priority UE signal can be restored. According to the descending order of power, the decoding procedure of other UE signals will be carried out as mentioned above until the desired UE signal is obtained. The NOMA-SIC receiver has its shortcomings. On one hand, the high throughput is obtained at the expense of system complexity which increases with the UE numbers. On the other hand, the error will be accumulated during the process of SIC.

Inspired by the power of deep learning (DL) changing the world profoundly, we explore its power in the NOMA signal detection. So far DL technology has had great success in many areas and diversities of new structures are proposed. Deep neural networks (DNNs) are the most classic DL forms to enhance the neural networks (NNs) performance by deepening the number of layers and neurons, and many successful applications can be seen in pattern classification and recognition. For natural language processing (NLP), recurrent neural networks (RNNs) are also formidable tools to handle signals among which sequences are related with each other. Another famous type of DL is conventional neural networks (CNNs). Its unique convolutional layer and pooling layer reduce sharply the computation complexity of image processing.

Successful commercial applications using the learning
approach in wireless communication are much fewer [2]. However, steps of related scientific study never stop. Many successful examples have been reported in various parts of communication, such as channel decoding [3–5], compressed sensing [6], detection for multiple input multiple output (MIMO) [7–9] and orthogonal frequency division multiplexing (OFDM) system [10], and modulation classification and identification [11]. Preliminary explorations on NOMA with the DL method have made progress in recent years. DL can optimize the power allocation of NOMA UEs. Saetan et al. completed the power allocation for downlink NOMA system with DL and results verify the near optimal performance with low computational complexity [12]. Zhang et al. investigated the energy-efficient resource allocation under joint subcarrier assignment and power allocation for uplink NOMA with reinforcement learning [13]. For multi-carrier NOMA, the problem of power allocation and channel assignment with the reinforcement learning framework were studied in [14]. The power allocation scheme in [15] was ruled by sum rate maximization in NOMA-SIC with DNN and its sum rate performance was close to the optimal scheme but with lower complexity. The NOMA transmission can also be designed as an end-to-end DL system. Ye et al. regarded the overlapped NOMA transmissions as multiple distinctive but correlated learning tasks. An end-to-end NOMA with the DNN form was proposed in [16]. An end-to-end transmission system with long short term memory (LSTM) network for uplink NOMA was reported in [17]. Kang et al. utilized the DNN for precoding at the transmitter and decoding at the receiver [18]. Several deep feed forward neural networks are adopted to fit the SIC processing. Although the autoencoder system shows excellent performance for its global optimization, it is hard to train the parameters of transmitting and receiving ends simultaneously in most realistic scenarios.

Undoubtedly, DL technology will not succeed without the rapid development of the graphic processing unit (GPU). Google has announced three generations tensor processing units (TPUs) from the first generation in May 2016 to now. The cost could be lowered by 80% for the third-generation TPUs. The newest TPU Pod can be configured with more than 1 000 TPUs and its computational speed reaches 100 petaflops. Although GPU resources remain expensive for users, the price continues to decrease in accordance with the Moore’s law.

As mentioned above, the traditional SIC has the problems of error propagation and high complexity. Besides, the long delay is inevitable for the process of SIC, especially in situations of large numbers of UEs. To handle these problems, we proposed a deep learning approach for uplink NOMA detecting. In our early work, we have explored the power of DNN in downlink MIMO-NOMA signal detection [19]. To fully tap into the potential of the DL method, we propose a learning approach based on CNN to detect the uplink NOMA signal without any traditional communication signal processing steps. Unlike the downlink NOMA based on DNN, the uplink NOMA system with CNN transfers the cost of hardware into base station (BS), which is much stronger than UEs in terms of signal processing. The structure of groups output can deal with multiple UEs’ detection very well at BS. Results reveal its excellent detection performance and signal processing ability.

The main contributions of this paper are summarized as follows:

(i) We design the first CNN detection system for the uplink NOMA system. The received NOMA signal can be processed by CNN completely instead of the traditional SIC method.

(ii) We make the utmost of CNNs power to handle high-dimensional information. The proposed system is able to fix the impairment from MIMO Rayleigh fading channel intelligently and accomplish the complex procedure of multiple UEs demodulation and decoding.

(iii) To evaluate the proposed system, we provide comparable simulations to SIC. The CNN method has a better error performance and lower computation complexity.

The structure of remainder is organized as follows. The preliminaries of NOMA and CNN are reviewed in Section 2. Then, the NOMA with the CNN approach called “DenseNOMA” is depicted in Section 3. Numerical simulation is provided in Section 4. Finally, we summarize the research and look to the future work in Section 5.

2. Preliminaries

The basic knowledge of proposing DL system for NOMA detection mainly involves two parts: uplink NOMA transmission and CNNs. Here, we introduce briefly both parts. For simplicity, the bandwidth is assumed to be 1 Hz and the channel is composed of additive white Gaussian noise (AWGN) and Rayleigh fading channels.

2.1 Uplink NOMA transmission

In this subsection, we describe the basic principle of uplink NOMA. Unlike the traditional OMA where different UEs occupy orthogonal time-frequency resources, NOMA can further enhance the spectrum efficiency by superposition coding to implement the message overload transmission. In the uplink, suppose the antenna numbers of UEs and BS are 1 and \(N_r\) respectively, and \(N_t\) is the number of UEs. \(s_i\) and \(p_i\) express the \(i\)th UE signal and its power respectively. Thus the \(i\)th NOMA signal can be expressed as

\[ x_i = \sqrt{p_i}s_i. \]
At the BS, the received signal of the $j$th antenna can be represented as

$$y_j = \sum_{i=1}^{N_t} h_{ij} x_i + w_j$$  \hspace{1cm} (2)$$

where $h_{ij}$ is the channel gain from the $i$th UE to UE $j$. The matrix form of (2) is

$$y = H' x + w$$  \hspace{1cm} (3)$$

where $x = [x_1, x_2, \ldots, x_{N_t}]$ is the UE signal, $H' = [h'_1, h'_2, \ldots, h'_{N_t}]$ is the channel matrix, $h'_i = [h'_{i1}, h'_{i2}, \ldots, h'_{iN_t}]^T$ is the channel gain from the $i$th UE signal to the $j$th BS antenna and $w = [w_1, w_2, \ldots, w_{N_t}]^T$ is the AWGN channel.

The schematic of the uplink NOMA system can be seen in Fig. 1.

---

At the BS, the SIC mechanism is operated. The receiver decodes signals in multiple stages according to the desired UE power. In the first stage, signals with the highest power will be decoded while the others are treated as interference. The throughput of the $i$th UE can be expressed as

$$R_i = \log_2 \left( 1 + \frac{p_1 |h'_1|^2}{\sum_{i=1}^{K} p_i |h'_i|^2 + N_0} \right)$$  \hspace{1cm} (4)$$

Then the restored message should be remodulated and subtracted from the received signal so that the second stage can be carried out to decode the signal of the second highest UE power. This process continues until the $k$th desired signal has been decoded. The throughput of the $k$th UE is represented as

$$R_k = \log_2 \left( 1 + \frac{p_k |h'_k|^2}{\sum_{i=k+1}^{K} p_i |h'_i|^2 + N_0} \right)$$  \hspace{1cm} (5)$$

The diagram of minimum mean square error (MMSE) with SIC can be seen in Fig. 2. Undoubtedly, the error of the higher decoding priority signal has a tremendous impact on the lower priority signal, and these errors continue propagating with the SIC process. The complicated workflow of SIC increases the complexity of the receiver.
2.2 CNNs

Recently, DL technology has become one of the hottest issues. Most of DL frameworks are originated by NN, which is constituted by many hierarchical neurons. Different from the DNN which merely increases the number of neurons and the depth of hidden layers to improve the learning ability, CNN provides a novel structure for massive data processing. In general, a CNN consists of three types of layers. These layers can be used repeatedly to form diversity of networks.

The first layer is the convolutional layer. Feature extraction of the input data is completed in this layer. Many convolution kernels can be used to extract various features by scanning regularly the local input data.

The second layer is the pooling layer. Its function is to reduce the dimension of data and decrease the overfitting. The input data could be divided into several rectangular areas. Each area outputs the maximum value or the average value according to the layer type.

The third layer is the fully connected layer. After finishing feature extraction and dimensional reduction, this layer will realize the final step: data classification. Benefited from the previous layers which have reduced sharply the number of data, a shallow NN is good enough for this job.

3. Design of DenseNOMA

A novel NOMA-CNN system called “DenseNOMA” is proposed in this section. The received signal can be processed directly by this network without any traditional digital signal processing. The powerful CNN could be used adequately to consider the receiver as a whole instead of separate modules.

3.1 Feasibility analysis

The feasibility of using the DL method to replace the SIC method should be verified first. Supposing that the number of UEs is K and the numbers of transmitting and receiving antennas are respectively N_t and N_r. For the kth UE, its power is denoted by \( p_k \) and the transmitting signal in the matrix form can be expressed as

\[
x = (x^1, x^2, \ldots, x^K)^T
\]

where \( x^k \) \((k \in [1, K])\) is the signal of the kth UE and the nth transmitting antenna. It can be expressed as

\[
x^k = \sqrt{p_k} S^k
\]

where \( x^k = [x^k_1, x^k_2, \ldots, x^k_{N_t}]^T \), \( S^k = [s^k_1, s^k_2, \ldots, s^k_{N_t}]^T \) and \( s^k_n \) is the signal of the kth UE and the nth transmitting antenna.

The MIMO-NOMA technique can improve the channel capacity by introducing the power dimension. Thus the channel matrix can be expressed by a three order tensor \( H \in \mathbb{R}^{N_t \times N_r \times K} \), where \( h_{n,t}^k \) \((k \in [1, K], n_r \in [1, N_r])\) and \( n_t \in [1, N_t])\) stands for the channel gain from the transmitting antenna \( n_t \) of UE \( k \) to the receiving antenna \( n_r \).

\[
H(:, :, k) = \begin{bmatrix}
h^k_{11} & h^k_{12} & \cdots & h^k_{1N_t}

h^k_{21} & h^k_{22} & \cdots & h^k_{2N_t}

\vdots & \vdots & \ddots & \vdots 

h^k_{N_t1} & h^k_{N_t2} & \cdots & h^k_{N_tN_t}
\end{bmatrix}
\]

Then the tensor \( H \) is transformed to a matrix \( H_{(1)} \) by mode-1 matricization [20] and it can be denoted as

\[
H_{(1)} = \begin{bmatrix}
h^k_{11} & \cdots & h^k_{1N_t} & h^K_{11} & \cdots & h^K_{1N_t}

h^k_{21} & \cdots & h^k_{2N_t} & h^K_{21} & \cdots & h^K_{2N_t}

\vdots & \vdots & \ddots & \vdots & \ddots & \vdots 

h^k_{N_t1} & \cdots & h^k_{N_tN_t} & h^K_{N_t1} & \cdots & h^K_{N_tN_t}
\end{bmatrix}
\]

The received signal can be expressed as

\[
Y = H_{(1)} x + N
\]

where the vector \( N \) is the AWGN.

Let \( l_k \) express the estimation output vector of the kth UE signal, the channel estimation value of the kth UE is denoted by

\[
\tilde{H}^k = \begin{bmatrix}
\tilde{h}^k_{11} & \tilde{h}^k_{12} & \cdots & \tilde{h}^k_{1N_t}

\tilde{h}^k_{21} & \tilde{h}^k_{22} & \cdots & \tilde{h}^k_{2N_t}

\vdots & \vdots & \ddots & \vdots 

\tilde{h}^k_{N_t1} & \tilde{h}^k_{N_t2} & \cdots & \tilde{h}^k_{N_tN_t}
\end{bmatrix}
\]

The UE1 signal should be decoded first with the MMSE process as follows:

\[
Y^1 = Y,
\]

\[
l_1 = \arg \min_l \| \tilde{H}_1 Y^1 - x^1 \|,
\]

\[
\tilde{H}_1 = (\tilde{H}^1)^H (\tilde{H}^1 (\tilde{H}^1)^H + \sigma^2 I)^{-1}.
\]

And then, the signal UE1 denoted by \( \tilde{Y}^1 \) should be reconstructed and it will be subtracted from the received signal.

\[
\tilde{Y}^1 = \tilde{H}^1 l_1
\]

\[
Y^2 = Y^1 - \tilde{Y}^1
\]

For UE2, the decoding signal is

\[
l_2 = \arg \min_l \| \tilde{H}_2 Y^2 - x^2 \|,
\]

\[
\tilde{H}_2 = (\tilde{H}^2)^H (\tilde{H}^2 (\tilde{H}^2)^H + \sigma^2 I)^{-1}.
\]
By following this procedure, the detection for UE $K$ can be shown as

$$
Y^K = Y^{(K-1)} - \tilde{Y}^{(K-1)},
$$

(19)

$$
l_K = \arg\min_i \| \bar{H} K Y^K - x^K \|,
$$

(20)

$$
\bar{H}_K = (\bar{H}^H K (\bar{H}^H K)^H + \sigma^2 K I)^{-1}.
$$

(21)

From (13), (17) and (20), it can be seen that the final classification result can be expressed in the form

$$
f_k(\cdots f_3(Y_3(f_2(Y_2 f_1(Y_1 A + b_1) + b_2) + b_3) \cdots) + b_k
$$

where $Y_k$ is a constant matrix, $b_k$ is a constant vector, and $f_i(\cdot)$ represents some form of the nonlinear function.

Obviously, the DL approach can be capable of replacing the traditional SIC method in the uplink MIMO-NOMA system and is even more powerful because it can search for the optimal solution in the way of data-driven.

### 3.2 DenseNOMA

Instead of designing a complex receiver for NOMA transmission, we propose a novel DenseNOMA system to detect the uplink NOMA signal. The construction of the DenseNOMA system can be seen in Fig. 3.

![Fig. 3 Design of the DenseNOMA system](image)

The proposed system consists of three blocks. The first block is the training block. Original UE signals will be produced in this module and specific power should be allocated to UEs. Each UE signal will reach the BS through independent MIMO channels separately. Meanwhile, all the UE labels are also sent to CNN for the training process. The second block is the testing block. The main function of this block is to produce the original UE information as well. The difference is that the purpose of UE signals is used for real-time NOMA detection testing and labels are not provided to CNN as input. To prevent the perfect match between the training and the testing sets, the channel models and generated data in training and testing blocks are independent and identically distributed (i.i.d.) so that the CNN performs with a smaller difference between the training and the testing process. The third block is the CNN detecting block. All the processes including channel estimation, digital down conversion, decoding, filtering and decision are replaced by the unified CNN detecting. The details are discussed in the next subsection.

Overall, two steps should be taken in the workflow of the DenseNOMA system. In Step 1, the training block starts to work. It is connected to the CNN block while the testing block remains unconnected. The original UE signals and labels are provided to CNN for training in the way of supervised learning. In Step 2, the testing block is activated while the training block remains suspended. The real-time testing is carried out by CNN. System performance will be evaluated. Relevant parameters and results are presented in Section 4.

### 3.3 CNN design

Inspired by the DenseNet [21], we propose a novel DenseNOMA system for uplink NOMA detecting. The system can process the received signal directly. Considering the digitized received signal with the carrier contains a mass of sampling sequences, classical DNN is hard to handle the situation. The proposed system can reduce sharply the computational complexity by the convolutional and pooling operation, and the dense structure can improve the information transmission efficiency.

The input data is a 3-D matrix including antenna dimension, time dimension and data real/image dimension. Before the signal is sent to the DenseNOMA, a color preprocessing data is normalized in the three dimensions separately should be finished to speed up the convergence in the training step.

The DenseNOMA system is designed for the characteristics of uplink NOMA signals. Its architectures can be seen in Table 1. Nine layers including one convolution layer, four dense blocks, three transition layers and one classification layer constitute the system. Input data is delivered to the first $4 \times 7$ convolution layer for elementary feature extraction. The next is four continuous dense blocks among which three transition layers are embedded into every two dense blocks. It is the main body of the DenseNOMA to process and decode the NOMA signal. The last layer is a fully connected layer to realize the classification process.
Table 1: Architectures of DenseNOMA

| Layer          | Output size | DenseNOMA              |
|----------------|-------------|------------------------|
| Convolution    | 4×28        | (4×7) conv, (1,2) stride |
| Dense block 1  | 4×28        | 1×1, 3×3 conv, 2       |
| Transition layer 1 | 4×14     | (1×1) conv             |
| Dense block 2  | 4×14        | 1×1, 3×3 conv, 4       |
| Transition layer 2 | 4×7    | (1×1) conv             |
| Dense block 3  | 4×8         | 1×1, 3×3 conv, 8       |
| Transition layer 3 | 4×4    | (1×1) conv             |
| Dense block 4  | 4×4         | 1×1, 3×3 conv, 4       |
| Classification | 1×1, 16     | 228D fully-connected   |

Notably, the first dimension of the input stands for the received antenna. The information of all antennas should be used simultaneously throughout the whole process. Thus, the size of the antenna dimension is designed to remain unchanged until the last layer. Moreover, the initial size of the sampling in one symbol is 50, and it should be padded to 56 before it is sent to the DenseNOMA because of the proper size for the following layers’ output size. The size of the convolution kernel and stride is designed empirically after a large amount of test data.

The dense block consists of several bottleneck layers. The output of the bottleneck layers will be concatenated to the previous feature map as the input of the next layer. This structure of dense connectivity can improve the information flow between layers rather than by increasing the layer of CNN. It means that fewer parameters are needed to optimize the networks and the final classification can be executed based on all the mappings in the networks. A dense block with two bottleneck layers can be seen in Fig. 4(a). Fig. 4(b) shows the specific process of the bottleneck layer. This layer can be divided into two parts. The first part can be defined as a function including four operations: normalization, Relu, convolution and drop out. The kernel of convolution is 1×1. The second part is exactly the same as the first part except that the kernel of convolution is changed to 3×3. The bottle structure can obtain better representation of the input with reduced dimensionality to improve the computation efficiency. Fig. 4(c) shows the transition layer. As can be seen, it contains a batch normalization, a Relu function, a 1×1 convolution, a drop out layer and an average pool. This layer is used for connecting two dense blocks. The convolution layer and the average pool can halve the data and feature maps.

![Fig. 4 Structure of dense block](image)

After finishing the feature extraction by all the dense blocks, the data is sent to a classification layer to accomplish the data classification. A 4×4 average pool reduces further the data dimension to 288 neurons. Benefiting from the strong learning ability in previous layers, the last fully connected layer with the sigmoid function can realize the final signal recovery with a few shallow layers.

Considering that all the UE signals should be decoded at the BS simultaneously, the output structure can be designed to form groups which can be seen in Fig. 5. For a
four antennas quadrature phase shift keying (QPSK) signal at the receiver, there are four groups and 16 neurons at the output layer after one-hot coding. In the form of soft-decision, each group outputs its UE signals with the grouping sigmoid function.

Fig. 5 Design of the output layer

4. Simulation

The proposed system works under the Windows 10 operating system. Numerical experiments of training and testing blocks are carried out with the simulation software Matlab. The CNN detecting block is completed with Python 3.6 and the DL framework TensorFlow with the GPU acceleration is employed to realize the DenseNOMA detecting algorithm. Suppose there are two UEs in one cluster. Each UE has only one antenna while BS has four. The MIMO channels with Rayleigh fading and the AWGN channels are considered in our simulation. The total power of the received signal is normalized to 1 W, and the signal from UE1 is supposed to be allocated with 80% of the power and UE2 20%. QPSK is adopted as the modulation scheme. Input data of the DenseNOMA receiver is the digital signal with the carrier and the number of sample points per symbol is 50. All the key parameter settings can be seen in Table 2.

| Parameter                  | Value                      |
|----------------------------|----------------------------|
| Operating system           | Windows 10                 |
| Programming               | Python + Matlab            |
| Framework                 | TensorFlow with GPU        |
| Channel                   | Rayleigh + AWGN            |
| UEs/cluster                | 2                          |
| Transmit antennas          | 1                          |
| Receive antennas           | 4                          |
| Modulation                | Phase shift keying         |
| Training symbol           | 20 480                     |
| Testing symbol            | 409 600                    |
| Sample/symbol             | 50                         |
| Power allocation factor   | 0.8                        |

The DL optimization algorithm is used for searching for the optimal hyper parameters efficiently. Here, we use the Adam algorithm for its rapid convergence of the network. Cross-entropy $H(P, Q)$ can measure the difference between two distributions from the view of information theory, and it is considered as the loss function which can be expressed as

$$H(P, D) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)} + H(P). \quad (22)$$

The regularization methods can be used to avoid the problem of network overfitting by adding the specific term to the loss function. Here, the $L_2$ regularization $\lambda \sum_i \|w_i\|_2$ is used. The detail of DenseNOMA can be depicted in Algorithm 1.

Algorithm 1 DenseNOMA detecting

Step 1 Initialize the CNN model.

Step 2 Generate training and testing data. The data can be expressed as a 3-D form (antenna $\times$ sample/symbol $\times$ real/imaginary).

Step 3 Set the key parameters, including convolutional kernel, pooling size, stride, mini-batch, learning rate, and output function.

Step 4 Execute the forward transmission process and acquire the final output $\hat{y}_i = \{\hat{y}_i[1], \hat{y}_i[2], ..., \hat{y}_i[N]\}$.

Step 5 Calculate the loss function $Loss(y, \hat{y})$:

$$Loss(y, \hat{y}) = \sum_i y_i \log_2 \frac{y_i}{\hat{y}_i} + H(y) + \lambda \sum_i \|w_i\|_2. \quad (23)$$

Step 6 Calculate and update the parameters with the Adam optimization algorithm [22].

Step 7 If the loss function cannot meet the requirement, return to Step 4 for retraining, otherwise go to next step.

Step 8 Test the trained DenseNOMA system. The symbol error rate (SER)-signal to noise ratio (SNR) curve is provided.

Fig. 6 shows the SER of the DenseNOMA system compared with the SIC method.

Fig. 6 SER of NOMA-SIC and DenseNOMA with QPSK
The signal is modulated with QPSK. The channel state information (CSI) is supposed to be known at the receiver with SIC while it is unknown with the DenseNOMA system. Obviously, the DenseNOMA detecting has a better performance. At the $10^{-4}$ level, SER values of the proposed DL method reach 5.7 dB and 6 dB for UE1 and UE2 respectively, while those of the SIC method are 12.1 dB and 12.5 dB respectively. The performance improvement reaches approximately 6.5 dB with the data-driven approach.

We also provide the simulation of detection performance with different modulations. Fig. 7 shows the first scheme: the BPSK modulation for both UE1 and UE2. It can be seen that the SER curve with DenseNOMA of UE1 is nearly coincided with that of UE2. Both UEs reach 1 dB while those with the SIC method reach 9 dB and 11 dB at the $10^{-4}$ level. The excellent improvement not only results from the greater noise margin of BPSK, but also from the powerful ability to search for the optimal solution of the DenseNOMA system. Meanwhile, it can be proved that the error propagation problem of SIC can be eliminated with the DL method.

Fig. 7 SER curve of NOMA-SIC and DenseNOMA with BPSK

Fig. 8 depicts the second scheme: BPSK for UE1 and QPSK for UE2. To balance the throughput and performance of NOMA transmission, the power allocation factor is changed to 0.7. From the SER-SNR curve, we can see that the DenseNOMA system brings 4.8 dB and 6.7 dB gains for UE1 and UE2 respectively.

Fig. 9 shows the impact of the mini-batch size on the loss function. Restricted by the computer memory, we compare three different mini-batch sizes. It can be seen that size 320 can converge faster than others for its larger data size per iteration. The convergence speed of the loss function decreases gradually with the decrease of the mini-batch size. The minimum size 32 has the lowest convergence speed and the curve has been in a significant fluctuation. In the simulation, there is a small difference of time we take to train the DenseNOMA per iteration between these mini-batch sizes. Thus, we recommend using larger mini-batch sizes to reach faster convergence speed.

Fig. 8 SER curve of NOMA-SIC and DenseNOMA with B/QPSK

Fig. 9 Impact of mini-batch on the loss function

In addition, the SIC computational complexity can be expressed as $O(L)$ where $L$ stands for the number of UEs in one cluster. For the well-trained DenseNOMA system, however, its computational complexity is $O(1)$ with limited multiplication and addition operations. It means that, it can be realized to detect the NOMA signal in real time.

5. Conclusions

In this paper, we design a DenseNOMA system to detect the uplink NOMA signal. The proposed DL method can handle the NOMA signal with the carrier directly without any traditional signal processing. The optimal signal processing and detecting can be finished by the DL approach intelligently without the complex modularity receiver design. Numerical results show the powerful SER performance compared with the SIC. We can obtain up to 6.5 dB gain in our simulation. The problem of error propagates of SIC
can be restrained to some extent with the DL method. In order to further improve the thoroughness and accuracy of the NOMA transmission, learning-based power allocation and real-time channel estimation with low computation complexity can be considered as a topic for future research. Besides, exploring the LSTM network to deal with the effect of memory channels is important in many practical scenarios. Signal transmission with unsupervised learning is also an interesting direction to study.

References

[1] HIGUCHI K, BENJIEBBOUR A. Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access. IEICE Trans. on Communication, 2015, 98(3): 403 – 414.
[2] O’SHEA T J, HOYDIS J. An introduction to machine learning communications systems. http://arxiv.org/pdf/1702.00832.pdf.
[3] NACHMANI E, BE’ERY Y, BURSHTEIN D. Learning to decode linear codes using deep learning. Proc. of the 54th Annual Conference on Communication, Control and Computing, 2016. DOI: 10.1109/ALLERTON.2016.7852251.
[4] GRUBER T, CAMMERER S, HOYDIS J, et al. On deep learning-based channel decoding. Proc. of the 51st Annual Conference on Information Sciences and Systems, 2017: 1 – 6.
[5] LIANG F, SHEN C, WU F. An iterative BP-CNN architecture for channel decoding. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(1): 144 – 159.
[6] SUN B, FENG H. Efficient compressed sensing for wireless neural recording: a deep learning approach. IEEE Signal Processing Letters, 2017, 24(6): 863 – 867.
[7] JEON Y S, HONG S N, LEE N. Blind detection for MIMO systems with low-resolution ADCs using supervised learning. Proc. of the IEEE International Conference on Communications, 2017: 21 – 25.
[8] WANG C. Research and application of traffic sign detection and recognition based on deep learning. Proc. of the International Conference on Robotics & Intelligent System, 2018: 21 – 23.
[9] O’SHEA T, ERPEK T, CLANCY T C. Deep learning-based MIMO communications. http://arxiv.org/pdf/1707.07980.
[10] YE H, LI G Y, JUANG B H. Power of deep learning for channel estimation and signal detection in OFDM systems. IEEE Wireless Communication Letters, 2017, 7(1): 114 – 117.
[11] XU Y, LI D Z, WANG Z Y. A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals. Proc. of the 2nd EAI International Conference on Machine Learning & Intelligent Communications, 2017: 5 – 6.
[12] SAETAN W, THIPCHAKSURAT S. Application of deep learning to energy-efficient power allocation scheme for 5G SC-NOMA system with imperfect SIC. Proc. of the 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2019: 661 – 664.
[13] ZHANG Y H, WANG X M, XU Y Y. Energy-efficient resource allocation in uplink NOMA systems with deep reinforcement learning. Proc. of the 11th International Conference on Wireless Communications and Signal Processing, 2019: 1 – 6.
[14] HE C F, HU Y, CHEN Y, et al. Joint power allocation and channel assignment for NOMA with deep reinforcement learning. IEEE Journal on Selected Areas in Communications, 2019, 37(10): 2200 – 2210.
[15] SAETAN W, THIPCHAKSURAT S. Power allocation for sum rate maximization in 5G NOMA system with imperfect SIC: a deep learning approach. Proc. of the 4th International Conference on Information Technology, 2019: 195 – 198.
[16] YE N, LI X M, YU H X, et al. DeepNOMA: a unified framework for NOMA using deep multi-task learning. IEEE Trans. on Wireless Communications, 2020, 19(4): 2208 – 2225.
[17] GUI G, HUANG H, JUANG B H, et al. Deep learning for an effective non-orthogonal multiple access scheme. IEEE Trans. on Vehicular Technology, 2018, 67(9): 8440 – 8450.
[18] KANG J M, KIM I M, CHUN C J. Deep learning-based MIMO-NOMA with imperfect SIC decoding. IEEE Systems Journal, 2020, 14(3): 3414 – 3417.
[19] LIN C, CHANG Q, LI X X. A deep learning approach for MIMO-NOMA downlink signal detection. Sensors, 2019, 19(11): 2526.
[20] KOLDA T G, BADER B W. Tensor decompositions and applications. SIAM Review, 2009, 51(3): 455 – 500.
[21] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2261 – 2269.
[22] KINGMA D, BA J. Adam: a method for stochastic optimization. Proc. of the 3rd International Conference on Learning Representations, 2015: 7 – 9.

Biographies

LIN Chuan was born in 1988. He received his B.S. degree in communication engineering from Guilin University of Electronic Technology, Guilin, China in 2010 and M.S. degree in communication and information system from Xidian University, Xi’an, China. He is now studying toward his Ph.D. degree in the School of Electronic and Information Engineering, Beihang University. His research interests include wireless communication and deep learning.
E-mail: lclkzjp@hotmail.com

CHANG Qing was born in 1962. He received his B.S. degree in mathematics from Beijing Normal University in 1982 and M.S. degree in mathematics and Ph.D. degree in communication and information system from Beihang University in 1991 and 1995 respectively. Since 2000, he has been a professor with the School of Electronic and Information Engineering, Beihang University. He is the dean of the Teaching-Research Section of Communication and Information System in his school. He is the author of one book, over 30 articles, and over ten patents. His research interests include satellite communication and navigation, and wireless networks.
E-mail: changq@263.net

LI Xianxu was born in 1988. He received his B.S. degree in communication and information system from Beijing University of Posts and Telecommunications, Beijing, China in 2010 and Ph.D. degree in communication and information system from Beihang University, Beijing, China in 2017. He is now working for State Grid Information and Telecommunication Branch. His research interests include navigation, satellite networks and network coding.
E-mail: lixianxu@buaa.edu