Enhancements to the Deep Learning Signal Detection Model in Non-Orthogonal Multiple Access Receivers and Noisy Channels

Ali Hilal Ali
University of Kufa

Raed S. H. Al-Musawi
University of Babylon

Kadhum Al-Majdi
Ashur University College

Abstract: This paper presents an enhanced deep learning-based Non-Orthogonal Multiple Access (NOMA) receiver that can mainly be used in low signal-to-noise channels. We show how a better dataset generation strategy for training Deep Learning (DL) could result in better generalization capabilities. Then, we apply hyperparameter tuning using exhaustive search to optimize the DL network. A Long-Short-Term-Memory (LSTM) DL architecture is used. The results show superior Symbol Error Rate vs Signal-to-Noise Ratio performance compared to the state-of-the-art methods such as Maximum Likelihood, Minimum Mean Square Error, and Successive Interface Cancellation, even though the network is only half as complex as previously proposed DL networks in the literature.

Keywords: NOMA, Deep Learning, LSTM, hyperparameter tuning, SIC

Introduction

Next-generation wireless networking has promised extensive capacity and a wide variety of interconnected devices. Not only is it required to support a vast amount of multimedia-based data traffic, but it must also support explosively growing, large-scale Internet-of-things (IoT) networks and an ever-growing user base (Chin, Fan & Haines, 2014). Although current multiple-access techniques such as Orthogonal Frequency Division Multiple Access (OFDMA) have allowed multiple users to share common network resources, they are, in terms of spectral usage, still inefficient enough to support the demands of the next-generation wireless communication systems (Hasan *et al.*, 2020).

http://doi.org/10.18080/jtde.v10n1.471

Fortunately, the rise of Non-Orthogonal Multiple Access (NOMA) techniques has promised many advantages over OFDMA (Islam, Zeng & Dobre, 2017) and has drawn extensive research interest in recent years. These advantages include more spectral efficiency (Khan *et al.*, 2020), reduced latency (Ye, Y., *et al.*, 2020), diverse power management policies (Park, Truong & Nguyen, 2019), and a vast increase in the simultaneous number of users it serves (Shin *et al.*, 2016).

Unlike OFDMA, where guard intervals are used to reduce the effect of interference, NOMA uses no guards at all, thereby increasing spectral efficiency tremendously (Hasan *et al.*, 2020). NOMA works by superimposing user signals, typically with different power levels, in a non-orthogonal fashion, which are then transmitted (Liu *et al.*, 2017). Successive Interface Cancellation (SIC) is performed where the user with the stronger channel condition is detected first while treating all other user signals as noise. Next, it is subtracted from the original stream at the receiver. The result is considered the user's signal with the weaker channel conditions and detected next (Chen, Jia & Ng, 2018). Maintaining perfect or near-perfect Channel State Information (CSI) is vital for NOMA if its superiority over OFDMA is to be realized, which is a challenging task (Hasan *et al.*, 2020).

The rise of Deep Learning (DL) in communication systems has provided another potential solution to channel estimation and detection in NOMA. DL is a method of learning from massive data with superior performance than traditional machine learning in many fields (Goodfellow, Bengio & Courville, 2016). Neng Ye *et al.* (2020) have proposed using deep multi-task learning to train several modules for channel estimation, mapping, and detection purposes of NOMA streams. However, their approach was not purely data-driven as some domain knowledge was exploited, and no evidence of parameter sharing makes it hardly a multi-task learning approach (Ruder, 2017). The researchers in Lin, Chang & Li (2019) have proposed a seven-layer neural network that can analyse CSI and detect transmitted sequences. While a seven-layer neural network is technically called a deep neural network, it often needs to comprise far more layers with varieties of architecture to warrant the badge of deep learning. Other researchers have suggested deep learning for various tasks in NOMA communication systems, like power minimization (Luo *et al.*, 2019), long-term power allocation (Sun *et al.*, 2019), and even joint precoding and decoding optimization (Kang, Kim & Chun, 2019). The literature is far more detailed to be listed in this short paper. Hence, we suggest Andiappan & Ponnusamy (2021) for a more comprehensive survey.

This paper improves the research effort in Thompson (2019), where a single DL network was used for channel estimation and detection. We chose this research for the following reasons. Firstly, their approach is end-to-end, where no pre-processing and/or domain knowledge are required. Secondly, it does not require channel estimation, which is a drawback of the

http://doi.org/10.18080/jtde.v10n1.471

traditional SIC detector. Thirdly, user signals are detected simultaneously in a single-shot fashion, rather than successively. However, their approach suffers from overfitting. Overfitting is when a DL fails to generalize from the training data, leading to poor classification performance. In addition, we have not, to the best of our knowledge, found any research, in the context of NOMA deep-learning-based receivers, where bias/overfitting and optimization are discussed and/or proved. We will show how a better-crafted dataset, training scheme, and simplified architecture can significantly reduce overfitting and increase performance.

The rest of the paper is divided into four sections. We will elaborate on the basic assumptions and scenario setting throughout the simulation. In the next section, we discuss the modelling of the DL network and our workflow. Then, we discuss the modelling procedure and hyperparameter tuning phase. Finally, we report on the simulation result and compare the performance of our approach to other state-of-the-art methods.

Basic Assumptions

This paper is essentially an improvement over the work presented in Thompson (2019). We will show how a better dataset generation strategy and careful fine-tuning of hyperparameters would improve accuracy and reduce overfitting. We will adhere to the same OFDM model assumptions. We repeat the details here for the reader's convenience.

For a fair comparison, we will also consider the case of two user terminals connected to a base station via a NOMA uplink, as illustrated in Figure 1. Each user will have a single antenna for transmitting data. The base station will receive superimposed signals from user terminals 1 and 2, and added noise. Since the mechanism for determining CSI is not considered in this paper, we will assume perfect CSI knowledge at the transmitters, i.e., the users, and the receiver, i.e., the base station.

With the assumption of $\mathcal{M}$-subcarriers and $\mathcal{N}$ users, we can write the amount of received signal in the frequency domain on subcarrier $\mathcal{K}$ as (Thompson, 2019):

$$\mathcal{Y}(\mathcal{K}) = \sum_{i=1}^{\mathcal{N}} \sqrt{\mathcal{P}_i(\mathcal{K})}\mathcal{H}_i(\mathcal{K})\mathcal{X}_i(\mathcal{K}) + \mathcal{W}(\mathcal{K}) \tag{1}$$

where $\mathcal{Y}(\mathcal{K})$, is the received signal, $\mathcal{P}_i(\mathcal{K})$ is the amount of transmitted power by user (i) on subcarrier $\mathcal{K}$, $\mathcal{X}_i(\mathcal{K})$ represents the transmitted symbol by user (i), and $\mathcal{W}(\mathcal{K})$ models additive white Gaussian noise. It is assumed by Thompson (2019) that the total power is $(\mathcal{P})$, whereas the power allocation coefficient for user (i) is $\sigma_i(\mathcal{K}) = \mathcal{P}_i(\mathcal{K})/\mathcal{P}$ and is constrained to have a unity sum. Finally, $\mathcal{H}_i(\mathcal{K})$ is the discrete Fourier transform of the multi-path channel $\hbar_i(t)$ given by (Thompson, 2019):

$$\hbar_i(t) = \sum_{\ell=1}^{\mathcal{L}} \vartheta_{i,\ell}\, \delta(t - \tau_{i,\ell}) \qquad (2)$$

where $\vartheta_{i,\ell}$ is the complex channel gain and $\delta()$ is the impulse function with a $\tau_{i,\ell}$ delay for user (i) along path ℓ. We will assume a Rayleigh fading channel and total paths ($\mathcal{L}$) of 20.

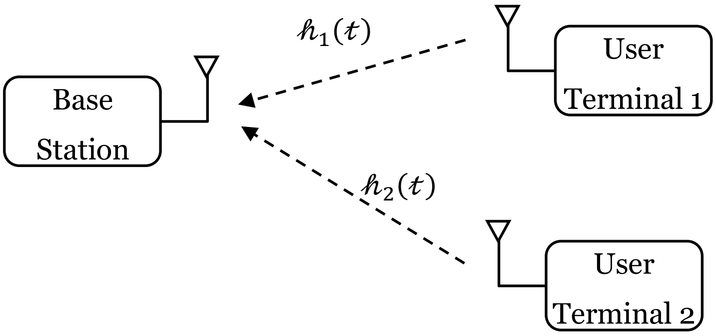

Figure 1. NOMA scenario of Thompson (2019) showing an uplink scenario of user terminal 1 & 2 sending packets to base station.

Modelling and Workflow

Since this research effort is based upon a data-driven modelling approach, we have followed a standard machine-learning modelling procedure, as shown in Figure 2(a). Figure 2(a) is a typical supervised classification scenario. For it to work, we need pairs of input-output examples. Firstly, we artificially generate samples of received OFDM data packets. To that end, a 64-subcarrier OFDM system is assumed. Each received packet has three symbols per user in which the first two packets are fixed pilots, while the third is a data symbol. The output of the training examples is one of 16 different combinations of the symbols that could be transmitted by users 1 and 2 in a QPSK baseband modulation. The reader is encouraged to refer to Thompson (2019) for more details.

However, our improvements compared to Thompson (2019) are that the training dataset is generated at different E_s/N_o levels ranging from 5 to 40 dB. The reasoning behind that is that the training dataset should be sampled *uniformly* from the probability distribution function of the data model, rather than sampled at only the highest modelled E_s/N_o value. The latter would result in a DL bias towards cases with high signal-to-noise levels. In contrast, if the generated training dataset were at only low E_s/N_o levels, then the resulting DL model would have difficulties learning the underlying structure of the mapping between input-output pairs, as much of the training data comprises high uncorrelated noise. In perfect scenarios, the training dataset should have a balanced number of samples per output label/class and samples that depict all the "extremes" and typical cases the input is expected to have.

Next, we construct the DL model using Long-Short-Term-Memory (LSTM) layers. LSTMs are types of Recurrent Neural Networks used widely to model data sequences (Goodfellow, Bengio

http://doi.org/10.18080/jtde.v10n1.471

& Courville, 2016). Each LSTM layer comprises hidden units representing the amount of information that should be remembered between time steps. The number of LSTM layers and the number of units within each layer are hyperparameters that we have tuned using exhaustive search to optimize the model's performance. The LSTM layer is often preceded by an input layer, which acts as a buffer and data preparation for the LSTM layer. On the other hand, the LSTM layers are often followed by fully connected neural-network layers to linearly separate the features obtained by the LSTM layer, and a Softmax layer to output the probability of the predicted class for each label. Finally, a classification layer converts these probabilities into a verdict representing the predicted label (see Figure 2(b)). MATLAB Deep Neural Network is used for both modelling and simulation.

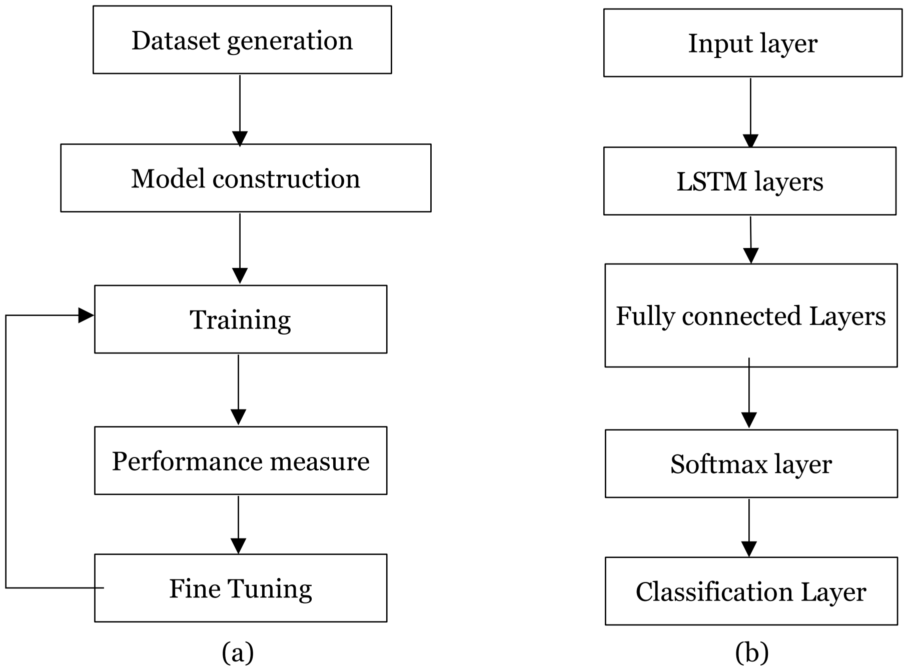

Figure 2. (a) Workgraph showing the major steps from modelling to LSTM training, testing, and fine-tuning; (b) layers details of the LSTM deep learning network

After testing its performance following the training phase, performance and fine-tuning are carried out in the DL model. The accuracy of both the training and testing data subsets is registered and compared to reduce bias and overfitting. For fine-tuning, we have considered: the number of LSTM layers, the number of hidden units, and the number of neurons in the fully connected layer.

Modelling Results

In this section, we report the simulation results of the scenarios presented in the previous section. Firstly, we have generated a training dataset at six E_s/N_o levels, namely: 5, 12, 19, 26, 33, and 40 dB. Each level had 1,000 samples per class for a total of 96,000 samples. This is only one-fifth of the training set size used by previous research efforts. While having a larger

http://doi.org/10.18080/jtde.v10n1.471

dataset size could help overcome potential overfitting, it substantially increases training time. The shorter the training time, the more simulation runs can be performed to fine-tune the model's hyperparameters. The training dataset has been divided into 90% training and 10% validation. Table 1 shows how we have varied some of the model hyperparameters and the number of layers for that end.

Table 1. Hyperparameters tuning by exhaustive search

Run No.	No. of LSTM layers	No. of LSTM units per layer	No. of Neurons in the fully connected layer	Training accuracy (%)	Validation accuracy (%)
1	1	16	16	48.18	43.81
2	2	16	16	53.95	49.66
3	3	16	16	47.48	42.8
4	1	32	16	78.11	70.73
5	2	32	16	79.91	70.73
6	**1**	**64**	**16**	**90.28**	**74.8**
7	1	128	16	93.09	65.4
8	1	64	two layers [4 16]	70.4	49.3
9	1	64	two layers [16 16]	88.1	66.45
10	1	64	two layers [32 16]	89.03	68.46

Our procedure was to change only one hyperparameter while keeping all others fixed. This hyperparameter was increased gradually until no further enhancement to the accuracy was possible. Then, we changed to another hyperparameter and so on. From Table 1, we started with a straightforward LSTM layer that has 16 units only. We noticed that increasing the number of layers in runs number 2 and 3 did not substantially increase accuracy but, rather, a slight decrease in validation accuracy was noticed. Furthermore, increasing the number of the fully connected layers or the number of neurons beyond 16 did not significantly impact accuracy. The setup of Thompson (2019), shown here as run number 7, suffered from overfitting, as there is a high difference of about 27 percentage points between training and validation accuracy.

The highest validation accuracy was registered in run 6 while, at the same time, keeping overfitting to a minimum. Overfitting is problematic in machine learning as it cripples the model's ability to generalize and, therefore, "learn" from data. However, we expect to have some generalization error because, when noise becomes dominant in low E_s/N_o levels, there would not be anything for the model to learn from, as the underlying signal structure is deeply buried inside the highly random nature of noise. We, therefore, settled with a value of hyperparameters obtained in run number 6, as it is a compromise between simplicity, higher validation accuracy, and "acceptable" overfitting. However, this issue could be investigated in future research.

Symbol Error Rate of User 1 and 2

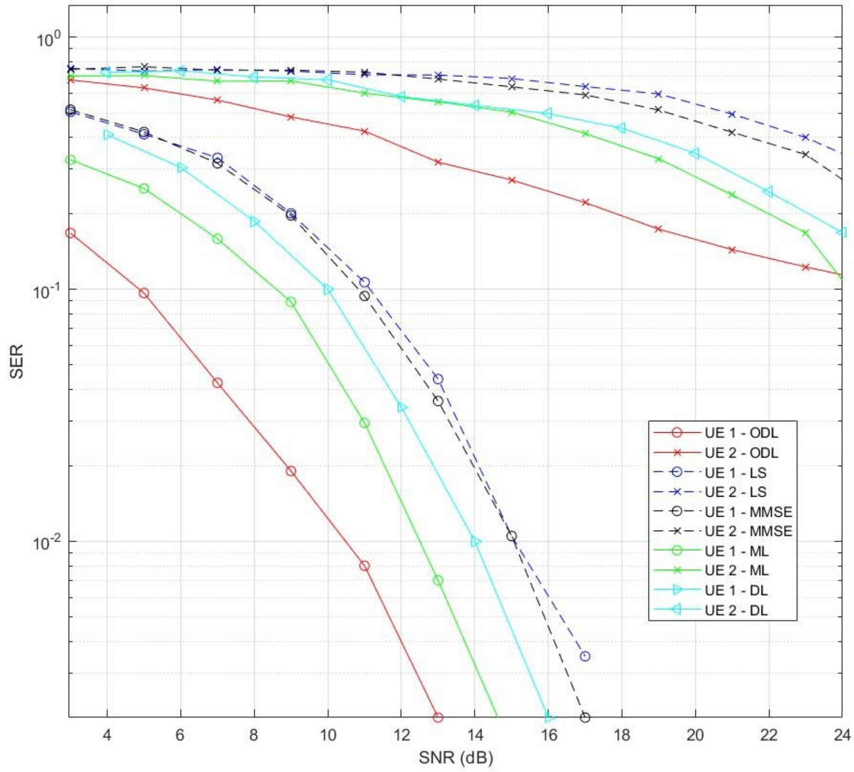

Figure 3. SER Curves for users 1 and 2 vs SNR. Our proposed optimized receiver is shown in red. LSTM of (Thompson, 2019) shown in sky blue. Other traditional methods are displayed in solid green (ML), dashed blue (LS) and dashed black (MMSE).

This section reports the results of using the trained deep-learning network on some test datasets and compares the results to that of Thompson (2019) and other standard methods. We will consider these standard methods: Maximum Likelihood (ML), Least Squares (LS), and Minimum Mean Square Error (MMSE) based SIC receivers. A typical procedure is followed for these methods, starting by estimating the channel using Least Squares (LS) and Minimum Mean Square Error (MMSE). We have varied the Signal-to-Noise Ratio (SNR) while registering the Symbol Error Rate (SER) value per-channel basis. To keep a fair comparison, we used a fixed phase shift and a cyclic prefix (CP) length of 20. Figure 3 shows the simulation results.

Our method is shown in red, marked with circles and crosses for users 1 and 2, respectively, and named ODL. In contrast, Thompson (2019) is shown in cyan marked with triangles for users 1 and 2 and named DL. The other curves are that of ML, LS, and MMSE. Figure 3 shows that the performance of our method exceeds that of all others, especially at low SNR. It proves how a better training strategy and careful optimization procedure can result in powerful and robust networks. However, as the SNR increases, the gap between the DL approach and the ML receiver becomes smaller, particularly for the weaker user 2. Hence, our approach is more suitable for channels with low SNR levels.

Conclusion

NOMA has become a de facto method for modern mobile communication systems because it promises superior spectral efficiency over OFDMA. With NOMA, the issue of reliably detecting users sharing the same resources in non-orthogonal settings is an ongoing research problem. One method to that end is to use deep learning and its capabilities to generalize from examples.

We have shown how we can improve one of the state-of-the-art deep learning approaches by generating a training dataset at different SNR levels, registering both the training and validation accuracy and optimizing the hyperparameters of the network. We have also shown how to decrease the impact of overfitting on the network capability of generalization.

The results prove our approach's superiority over state-of-the-art DL, ML, and SIC approaches, in particular at low signal-to-noise ratios. However, we recommend more network testing using various CP lengths, channel phase fading, and pilot symbols, though the robustness of DL in such situations has already been proven by other researchers.

References

Andiappan, V., & Ponnusamy, V. (2021). Deep Learning Enhanced NOMA System: A Survey on Future Scope and Challenges. *Wireless Personal Communications*, *123*, 839–877. https://doi.org/10.1007/s11277-021-09160-1

Chen, X., Jia, R., & Ng, D. W. K. (2018). On the design of massive non-orthogonal multiple access with imperfect successive interference cancellation. *IEEE Transactions on Communications*, *67*(3), 2539–2551. https://doi.org/10.1109/TCOMM.2018.2884476

Chin, W. H., Fan, Z., & Haines, R. (2014). Emerging technologies and research challenges for 5G wireless networks. *IEEE Wireless Communications*, *21*(2), 106–112. https://doi.org/10.1109/MWC.2014.6812298

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Hasan, M. K., Shahjalal, M., Islam, M. M., Alam, M. M., Ahmed, M. F., & Jang, Y. M. (2020). The role of deep learning in NOMA for 5G and beyond communications. Paper presented at the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).

Islam, S., Zeng, M., & Dobre, O. A. (2017). NOMA in 5G systems: Exciting possibilities for enhancing spectral efficiency. arXiv preprint arXiv:1706.08215. https://doi.org/10.48550/arXiv.1706.08215

Kang, J.-M., Kim, I.-M., & Chun, C.-J. (2019). Deep learning-based MIMO-NOMA with imperfect SIC decoding. *IEEE Systems Journal*, *14*(3), 3414–3417. https://doi.org/10.1109/JSYST.2019.2937463

Khan, W. U., Liu, J., Jameel, F., Sharma, V., Jäntti, R., & Han, Z. (2020). Spectral efficiency optimization for next generation NOMA-enabled IoT networks. *IEEE Transactions on*

Vehicular Technology, *69*(12), 15284–15297. https://doi.org/10.1109/TVT.2020.3038387

Lin, C., Chang, Q., & Li, X. (2019). A Deep Learning Approach for MIMO-NOMA Downlink Signal Detection. *Sensors*, *19*(11), 2526. Retrieved from https://www.mdpi.com/1424-8220/19/11/2526

Liu, Y., Qin, Z., Elkashlan, M., Ding, Z., Nallanathan, A., & Hanzo, L. (2017). Non-orthogonal multiple access for 5G and beyond. *Proceedings of the IEEE*, *105*(12), 2347–2381. https://doi.org/10.1109/JPROC.2017.2768666

Luo, J., Tang, J., So, D. K., Chen, G., Cumanan, K., & Chambers, J. A. (2019). A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT. *IEEE Access*, *7*, 17450–17460. https://doi.org/10.1109/ACCESS.2019.2895201

Park, S., Truong, A. Q., & Nguyen, T. H. (2019). Power control for sum spectral efficiency optimization in MIMO-NOMA systems with linear beamforming. I*EEE Access*, *7*, 10593–10605. https://doi.org/10.1109/ACCESS.2018.2890441

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098. https://doi.org/10.48550/arXiv.1706.05098

Shin, W., Vaezi, M., Lee, J., & Poor, H. V. (2016). On the number of users served in MIMO-NOMA cellular networks. Paper presented at the 2016 International Symposium on Wireless Communication Systems (ISWCS).

Sun, Y., Wang, Y., Jiao, J., Wu, S., & Zhang, Q. (2019). Deep learning-based long-term power allocation scheme for NOMA downlink system in S-IoT. *IEEE Access*, *7*, 86288–86296. https://doi.org/10.1109/ACCESS.2019.2926426

Thompson, J. (2019). Deep learning for signal detection in non-orthogonal multiple access wireless systems. Paper presented at the 2019 UK/China Emerging Technologies (UCET). https://doi.org/10.1109/UCET.2019.8881888

Ye, N., Li, X., Yu, H., Zhao, L., Liu, W., & Hou, X. (2020). DeepNOMA: A unified framework for NOMA using deep multi-task learning. *IEEE Transactions on Wireless Communications*, *19*(4), 2208–2225. https://doi.org/10.1109/TWC.2019.2963185

Ye, Y., Hu, R. Q., Lu, G., & Shi, L. (2020). Enhance latency-constrained computation in MEC networks using uplink NOMA. *IEEE Transactions on Communications*, *68*(4), 2409–2425. https://doi.org/10.1109/TCOMM.2020.2969666

http://doi.org/10.18080/jtde.v10n1.471