Data-Driven Compressed Sensing for Massive Wireless Access

Yanna Bai, Wei Chen, Feifei Sun, Bo Ai, and Petar Popovski

ABSTRACT

The central challenge in massive machine-type communications (MTC) is to connect a large number of uncoordinated devices through a limited spectrum. The typical MTC communication pattern is sporadic, with short packets. This could be exploited in grant-free random access in which the activity detection, channel estimation, and data recovery are formulated as a sparse recovery problem and solved via compressed sensing algorithms. This approach results in new challenges in terms of high computational complexity and latency. We present how data-driven methods can be applied in grant-free random access and demonstrate the performance gains. Variations of neural networks for the problem are discussed, as well as future challenges and potential directions.

INTRODUCTION

Massive machine-type communication (MTC) is one of three core services of 5G, characterized by sporadic uplink communication and low rates with small data packets. Already in 5G, it was considered that the grant-based access channel procedures of LTE need to be refined for mMTC. In grant-based access, active users send a randomly selected orthogonal preamble to the base station (BS) and wait for the response. The users that receive a response send connection requests for uplink resources. If the BS successfully decodes the contention request, it sends a contention resolution message. Otherwise, the users have to repeat the handshaking procedure. The limited number of preambles is a bottleneck on massive batch arrivals of packets, and complicated signaling exchanges result in high overhead/inefficiency for small data packets. Grant-free random access has been widely seen as an approach that can overcome these drawbacks [1]. In grant-free random access, active users directly transmit the predefined preambles and data to the BS without waiting for the grant from the BS. This greatly reduces the signaling overhead and access delay, but brings additional uncertainty and complexity due to the unknown user activity. By exploiting the natural sparsity of sporadic communication in mMTC, compressed sensing (CS) [2] is applied in grant-free random access [1], where the user activity detection and channel estimation can be addressed jointly by solving a sparse recovery problem.

In recent years, applying data-driven methods to solve the problem in massive wireless access has become a promising direction. For example, in data-driven methods, neural networks (NNs) can learn the hard-to-model activation patterns of devices from the training data, such that it can predict frequently activated devices more accurately. In addition, data-driven methods achieve a significant reduction in computational complexity, which is important to latency-sensitive applications.

In this article, we review the data-driven CS methods for user activity detection and channel estimation in grant-free access for mMTC. First, we briefly introduce the CS-based massive access and traditional model-driven optimization methods. Then, we investigate various aspects of data-driven CS methods for massive access. Last, we analyze the NN design problem, which can provide a guide for designing customized NN in practical scenarios. In addition, we discuss the main challenges in data-driven CS, present potential solutions, and indicate future directions.

CS-BASED GRANT-FREE RANDOM ACCESS

In grant-free access, each user is assigned a unique preamble, also referred to as the pilot sequence [1]. As shown in Fig. 1, in each coherence time, active users directly transmit their pilots and data to the BS, and the BS first detects the active users according to their pilots and estimates their channel, and then decodes their data. As the data transmission mostly is quasi-periodic or event-driven, only a small subset of devices is active at a given access instant. Thus, the activity of all users is a sparse signal, and the user activity detection leads naturally to solving a CS problem. With known pilots of all users, the BS performs joint activity detection and channel estimation (JADCE) by recovering a sparse signal. However, it is very expensive to assign an orthogonal pilot for each user due to the limited spectrum resource and massive number of users. Thus, non-orthogonal pilot sequences are needed to support a vast number of users with limited pilot resources. This renders the massive access problem underdetermined and equivalent to the classical CS problem.

Model-driven methods for this CS-based massive access problem explicitly utilize the characteristics (e.g., sparsity) to design customized optimization algorithms. However, those methods are usually...
In this section, we first introduce different sparse recovery problem formulations in communication systems. We consider the mMTC has sporadic communication. This study is limited to contention-free access where the non-zero elements of the output of the network indicate the detected active users and estimating their channel with one measurement [9, 11]. For a BS equipped with multiple antennas, we have multiple measurements, and activity detection and data recovery is constructed as a row-sparse matrix recovery problem [7, 12, 14]. Furthermore, with some appropriate transform, the users’ channel becomes sparse in the virtual angular domain in massive multiple-input multiple-output (MIMO), especially in the mmWave domain. Those structures, beyond the simple sparsity model, are desired to be exploited in the design of NNs for performance enhancement and parameter reduction.

**Problem Formulation and Data-Driven CS Solutions**

As shown in Fig. 1, different scenarios lead to different sparse structures in the massive access problem. For a BS equipped with a single antenna, the massive access problem is detecting active users and estimating their channel with one measurement [9, 11]. For a BS equipped with multiple antennas, we have multiple measurements, and activity detection and data recovery is constructed as a row-sparse matrix recovery problem [7, 12, 14]. Furthermore, with some appropriate transform, the users’ channel becomes sparse in the virtual angular domain in massive multiple-input multiple-output (MIMO), especially in the mmWave domain. Those structures, beyond the simple sparsity model, are desired to be exploited in the design of NNs for performance enhancement and parameter reduction.

Data-driven methods have been applied for individual user activity detection [7, 8], joint user activity detection and channel estimation [8–13], joint pilot design and activity detection [14], and joint pilot design and activity detection and channel estimation [14]. Using data-driven methods for activity detection leads to a multi-label classification problem, where each label is a one-hot vector that corresponds to a distinct subset of active users. With estimated user activity, the BS can recover the channel coefficients of active users easily via least squares or minimum mean square estimator. In JADCE, data-driven methods directly learn the mapping between the received pilot signal and the product of user activity indicator and user channel vector. In the inference phase, the received signal at the BS is fed into the trained network. The non-zero elements of the output of the network indicate the detected active users and their channel coefficients.

Auto-encoders are used to design pilot sequences [14]. As shown in Fig. 2, the input is the sparse channel vector, and the output is the user activity indicator or sparse channel vector. In the training phase, the encoder and the decoder are jointly trained. The pilot matrix design can be generated. Furthermore, real data can be used for training if available, which is more beneficial as the NN may learn more features from the real data, such as user activity patterns.

Data-driven methods for CS have been widely studied in the literature. As confirmed in [3], the trainable version of Li’s and Osher’s coordinate descent method has the potential to achieve the target reconstruction, performance 10 times faster than the original algorithm. This provides a considerable gain in access latency for time-sensitive applications. Data-driven methods also improve the CS estimation accuracy as observed in [3–15]. The performance gain of data-driven methods arises from several shortcomings of traditional CS methods, including the model error due to imperfect modeling, the structure error due to relaxation of the original problem, and the convergence error due to local optima.

**Key assumptions:** The performance of activity detection and channel estimation assumes that the mMTC has sporadic communication. This study is limited to contention-free access where each user has a unique pilot sequence and the system is assumed to be synchronized.

**Data-Driven Compressed Sensing for mMTC**

In this section, we first introduce different sparse problem formulations in communication systems and data-driven solutions. Table 1 summarizes several typical networks, outlining the problems they address in network design. We also show how to overcome the challenges of adapting standard NNs to the problem at hand.

**FIGURE 1.** An illustration of the CS-based grant-free random access scheme. The channel signal of all users can be thought of as a sparse signal due to sporadic communication. Different rows of the sparse signal correspond to different users. Different colors of the sparse signal represent channel coefficients, and blank represents zero.

The performance gain of data-driven methods is observed in [3–15]. Data-driven methods use NNs to solve the sparse recovery problem for activity detection or/and channel estimation. While model-driven methods rely on mathematical models in the formulation of each iterative step, data-driven methods learn the parameters of NN from available training data. In the training phase of data-driven methods, the network adjusts its parameters to minimize the loss function, such as the mean square error between the estimated user activity vector and the ground truth. When inference is carried out, the received pilot signal at the BS is fed into the well-trained network, and the output of the network is the detected active user vector or estimated channel. As the pilot sequences are known, plenty of training data via simulators can be generated. Furthermore, real data can be used for training if available, which is more beneficial as the NN may learn more features from the real data, such as user activity patterns.

Data-driven methods have been applied for individual user activity detection [7, 8], joint user activity detection and channel estimation [8–13], joint pilot design and activity detection [14], and joint pilot design and activity detection and channel estimation [14]. Using data-driven methods for activity detection leads to a multi-label classification problem, where each label is a one-hot vector that corresponds to a distinct subset of active users. With estimated user activity, the BS can recover the channel coefficients of active users easily via least squares or minimum mean square estimator. In JADCE, data-driven methods directly learn the mapping between the received pilot signal and the product of user activity indicator and user channel vector. In the inference phase, the received signal at the BS is fed into the trained network. The non-zero elements of the output of the network indicate the detected active users and their channel coefficients.

Auto-encoders are used to design pilot sequences [14]. As shown in Fig. 2, the input is the sparse channel vector, and the output is the user activity indicator or sparse channel vector. In the training phase, the encoder and the decoder are jointly trained. The pilot matrix design can be generated. Furthermore, real data can be used for training if available, which is more beneficial as the NN may learn more features from the real data, such as user activity patterns.

Data-driven methods for CS have been widely studied in the literature. As confirmed in [3], the trainable version of Li’s and Osher’s coordinate descent method has the potential to achieve the target reconstruction, performance 10 times faster than the original algorithm. This provides a considerable gain in access latency for time-sensitive applications. Data-driven methods also improve the CS estimation accuracy as observed in [3–15]. The performance gain of data-driven methods arises from several shortcomings of traditional CS methods, including the model error due to imperfect modeling, the structure error due to relaxation of the original problem, and the convergence error due to local optima.

**Key assumptions:** The performance of activity detection and channel estimation assumes that the mMTC has sporadic communication. This study is limited to contention-free access where each user has a unique pilot sequence and the system is assumed to be synchronized.

**Data-Driven Compressed Sensing for mMTC**

In this section, we first introduce different sparse problem formulations in communication systems and data-driven solutions. Table 1 summarizes several typical networks, outlining the problems they address in network design. We also show how to overcome the challenges of adapting standard NNs to the problem at hand.
TABLE 1. The classification of mentioned data-driven methods.

| Category            | Ref. | Network                              | Input/output                               | Description                                                                                                                                 |
|---------------------|------|--------------------------------------|--------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| CS + data-driven    | [3]  | Fully connected network (FNN)        | The measurement/sparse signal              | The network is obtained via unfolding the iterative soft-thresholding algorithm (ISTA), named learned ISTA (LISTA).                          |
|                     | [4]  | Stacked denoising autoencoder        | The sparse signal/sparse signal            | The network can be used for linear and mildly nonlinear measurements.                                                                          |
|                     | [5]  | FNN                                  | The measurement/sparse signal              | The network is obtained via unfolding approximate message passing (AMP), named learned AMP (LAMP).                                           |
|                     | [6]  | FNN                                  | The measurement and measurement matrix/sparse signal | The network is adaptive to varying model scenarios by combining the measurement matrix into network.                                         |
|                     | [7]  | FNN                                  | The measurement/non-zero indices or signal | Authors explores the low-rank property in massive MIMO and utilize it in network design to reduce the number of network parameters.          |
|                     | [8]  | Long short-term memory (LSTM)        | The measurement/non-zero indices or signal | The network is adaptive to varying sparsity by controlling the number of LSTM units.                                                          |
|                     | [10] | FNN                                  | The measurement/sparse signal              | Authors combine the prior information from checking into network to enhance the performance.                                                  |
|                     | [11] | FNN                                  | The measurement/sparse signal              | Authors extend the LAMP into multiple measurements and propose a new activation layer to utilize asynchronous information.                 |
|                     | [12] | FNN                                  | The measurement/sparse signal              | The network has adaptive depth for varying sparsity.                                                                                         |
|                     | [13] | FNN                                  | The measurement/sparse signal              | The network is obtained via unfolding the alternative direction method of multipliers algorithm.                                            |
|                     | [13] | FNN                                  | The measurement/sparse signal              | Authors reduce redundant parameters of the LISTA for efficiency.                                                                               |
|                     | [14] | Autoencoder                          | The sparse signal/non-zero indices or signal | Authors propose to use autoencoder jointly learn the pilot matrix and receiver.                                                              |

Utilizing Domain Knowledge in Communication Systems

A promising direction in data-driven CS for mMTC is opened by utilizing various features of communication systems to design domain-specific NNs. In computer vision, the convolutional NN that considers the image structure information outperforms the fully connected NN. However, most existing networks are not designed for communication systems. There are various features of mMTC that can be utilized to enhance the NN design, including the low-rank property in massive MIMO [7], asynchronous information [10], various structure sparsity [12], nd so on. These communication-specific designs can be combined with various parts of NNs. For example, to use the row-sparse property in MIMO, we can use a modified activation function to enforce row-sparsity of estimated signal [12]. Or we can design the network structure, such as constructing multiple networks in a distributed way, where each NN works on the detection at one antenna, and various NNs can exchange the detected activity information to enhance the performance [10]. By utilizing the special structure in different communication systems, those carefully designed networks show improved performance gain for activity detection and channel estimation [7, 10, 12].

Another characteristic is considering the correlation embedded in the real and imaginary parts of the complex signal. As NNs cannot directly deal with complex signals, a common method is to flatten a complex signal into a real concatenated signal [9, 12, 13]. However, this neglects the correlation between the imaginary and real parts, and results in some performance degeneration. One method is to use different weights to process the real and imaginary parts of the complex signal separately, and combine the output to obtain the input of the next layer [4], as shown in Fig. 2.

In [9], the checking mechanism is used to obtain the indices of successfully detected active users. This information is then used in the activity detection again, which corresponds to a sparse recovery problem with a known partial non-zero indices set. Those known non-zero indices are constructed as a one-hot vector that is used as additional input of the network and assist in the selection of the remaining non-zero indices in the activation layer.

Enhanced Data-Driven Methods

Although the above data-driven methods show outstanding performance, they also have drawbacks. A network trained with a specific pilot matrix cannot be applied to a different scenario associated with a distinct pilot matrix. The time-consuming training process hinders the application of data-driven methods.

To address the problem of retraining, model-agnostic meta learning (MAML) is used in [15] to speed up the training phase. While different situations with different pilot matrices can be considered as different tasks, MAML at first reuses the structural similarity across the scenarios to find a good initialization for the network in those
tasks. Then, for a given pilot matrix, the network only needs a small number of iterations to adapt to new tasks. In [6], Aberdam et al. directly combine the measurement matrix into the network and learn a universal architecture for different models. The measurement matrix and the measurement signal are jointly sent to the learned network for recovery.

Another method is improving the training efficiency via decreasing the network parameters. For example, relying on rigorous analysis, Shi et al. reduce redundant training parameters to simplify the unfolding-based network [13]. In [7], Shao et al. reduce the network parameters by utilizing the low-rank property of the device state matrix that indicates the user activity and channel in MIMO, where the received pilot signal is mapped into a low-dimensional space, and activity detection and channel estimation are performed in the low-dimension space.

Another issue in data-driven methods is the fixed number of layers. While traditional algorithms for massive access can flexibly adjust the number of iterations for a varying number of active users, the trained network in data-driven methods usually has a fixed number of layers. This directly influences the performance of activity detection in real applications, where the number of active users varies with time. In addition to the accuracy degradation, such a network is also inefficient in terms of computation as it has fixed access latency for different numbers of active users. To solve this problem, a novel network is proposed in [11], using adaptive depths that can dynamically adjust the number of layers in executing specific tasks. In the proposed network, a mapping matrix is used to characterize the distribution of the reconstruction error of each layer; then a halting score is calculated to determine whether to stop at the current layer, and the weighted sum of the results of all layers is used as the output of the network. Another method to control the network depth is using the recurrent NN, such as long short-term memory (LSTM) networks. By controlling the number of LSTM cells in the network, Ahn et al. propose a network to support various scenarios with sparsity [11]. In each cell, the input and forget gates choose the active and inactive users, respectively. Thus, the output of a late stage traverses more cells and can thus estimate a signal with higher sparsity.

**EXPERIMENTAL RESULTS**

To show the performance gain of data-driven methods, we construct a simple CS-based grant-free access scenario as described earlier. We assume 500 users, and each has a unique pilot sequence of length 250. Elements in the pilot sequences are randomly generated following the independent and identically distributed Gaussian distribution. In practice, some finite alphabet symbols are used as pilots. The probability of activation is set to 0.1 and is identical for all devices. The signal-to-noise ratio is set to 20 dB. We compare the normalized mean square error (NMSE) of ISTA, LISTA [3], LISTA-H (which has an adaptive number of layers [11]), AMP, and LAMP [5].

As shown in Fig. 3, LISTA only contains matrix multiplication and thresholding function. By unfolding ISTA, we can obtain LISTA, where the matrix multiplication is realized via the fully connected layer, and the activation function is the soft-thresholding function with learned thresholding parameters. Compared to LISTA, LISTA-P adds the prior knowledge related to known non-zero indices sets to the activation function to help the sparse support selection. LISTA-H adds a halting function in each layer to evaluate whether to stop at this layer by comparing the received pilot signal to the product of the output at this layer and the pilot matrix. LAMP is the unfolding-based network of AMP.

To compare the computation complexity of ISTA, AMP, and LISTA, we evaluate them on the same test dataset, using a computer with a quad-core 2.3 GHz CPU and 16 GB RAM. AMP costs 1.06 s (30 iterations), and its NMSE reaches –17.45 dB. The NMSE performance of ISTA is –17.49 dB, which costs 40.9 s (1000 iterations). LISTA, whose NMSE reaches –18.31 dB with 16 layers, only uses 0.24 s.

In Fig. 4, we can see that LISTA and LAMP converge faster and perform better than ISTA and AMP. This is because the original activity detection and channel estimation is an NP-hard problem, and ISTA solves a convex relaxed optimization problem, which leads to reconstruction errors when the number of measurements is smaller than some threshold. Data-driven methods directly learn the mapping between the received pilot signal and channel of active users and provides the possibility to learn the optimal estimation result. As LISTA-H can flexibly adjust the number of layers when varying the sparsity, it has a lower average number of layers.

![Network Structure](image-url)
We also measure the training time of networks with different depths. We observe that the network with more layers requires more training time. For example, in comparison to a network with 10 layers, the network with 20 layers consumes 47.77 percent more training time, and its NMSE reduces 13.71 percent, which illustrates the trade-off between computation/training time and performance.

In Fig. 5, we compare the performance under different numbers of active users. LISTA-ACT30 is trained under sparsity level 30, and thus has the best test performance on sparsity level 30 and decreased performance on other sparsity levels. LISTA-MIX, which is trained under mixed sparsity levels, is more robust than LISTA-ACT30. LISTA is trained for each sparsity level, and thus has better performance. However, training networks for all settings is time-consuming. Training a network with data of mixed settings would be less time-consuming, while its performance is usually limited. Compared to LISTA, LISTA-P [9] utilizes the error-checking information as the prior information of the network and achieves outstanding performance.

Challenges and Future Directions
Data-driven methods are indispensable tools for 6G systems, but there are still significant challenges for efficient usage in various applications. In this section, we present the main challenges and possible research directions for data-driven CS in mMTC.

Constructing the Training Data
A key issue in the data-driven method for massive access is where the training data come from. Most existing research on data-driven methods for mMTC mainly use generated data from simulators according to some channel model. Nevertheless, the existing model-driven access schemes fail to consider complex cell-specific channel characteristics in mMTC. With data-driven methods, it is possible to learn more scenario-based characteristics by exploiting the historical channel state information stored in the BS. In situations where historical data is not available, one can rely on an unmanned aerial vehicle to scan the whole scenario and construct a three-dimensional map, and then use techniques such as ray tracing to generate channel data. In addition to the channel information, the training data may contain more information, such as the underlying traffic pattern of users. According to the historical access data, data-driven methods can learn the access probability of users and predict frequently activated users with higher probability. If only a small part of historical data is available, we can train the network with simulated data and fine-tune the trained model on the limited real data. In practice, the trained model can be frequently updated by using the newly available data. Another promising venue is the creation of synthetic data using generative models, such as variational autoencoders and generative adversarial networks.

Designing Specialized Networks
Most well-known NNs in data-driven methods are originally designed for processing visual signals and natural language. To fully exploit the potential of data-driven methods, it is therefore necessary to study dedicated NNs for communication systems (e.g., NNs working with complex signals).
in massive access, there are a lot of features that can be taken into account in the NN design, such as the user activity patterns. We can use some recurrent NN to capture the temporal correlation of user activity. In recent years, supporting mMTC with massive MIMO is attractive. However, this setup significantly increases the signal dimension, which puts forward a training challenge for high-dimensional networks. A possible method to deal with the prohibitive computational complexity due to large-scale antennas is designing light networks to reduce the number of parameters of the network, e.g., by exploring the low-rank property in massive MIMO [7]. Besides, in 6G, massive MIMO further extends to cell-free massive MIMO, which has different characteristics than massive MIMO and urges improvements in access system design. In cell-free massive MIMO, channel estimation is more difficult as signals at different antennas have different large-scale fading. Specific features, such as the spatial sparsity of signals, can inspire additional network optimization. For instance, we can use a new sparsity-enforcing activation function in the original sparse recovery network in order to attain spatial sparsity.

**Optimizing the Training Process**

Data-driven methods in the inference phase show improved computation complexity in various literature works, which is important for latency-sensitive applications such as real-time monitoring. However, this comparison happens in the inference phase, and the time-consuming training phase is still a considerable issue. Specifically, in most cases, we have to re-train a new network if some settings in the scenario are changed (e.g., the number of access users). Thus, optimizing the training process of data-driven methods to fit the flexible applications in 6G is a challenging problem. A promising direction is seen in meta-learning [15], as it enables the network to learn a new task quickly. To obtain a good meta-learning model with powerful adaptive and generalization capabilities, we need to train the network under different access scenarios, i.e., different pilot matrix. Then, for a new scenario with new pilot matrix, the network can learn fast, using only a few access data. Thus, we expect meta-learning to amend the deficiency of the existing data-driven methods, especially when only a limited amount of historical access data is available for training.

**Applying the Trained Network**

Data-driven methods for massive access have shown outstanding performance in plenty of the literature, but we still have a long way to go to use the data-driven methods in practice. The massive access scenarios in practical applications are often more complex than the models in the literature (dynamic access densities, different types of devices, different sizes of data packets, etc.). The practical scenario with dynamic user population can be addressed by contention-based CS access, where active users randomly select the pilots from the common pilot pool. Then we only need to train one network for each pilot matrix. The diversity between the distribution of the network input and training data results in performance degradation. In existing works, the network trained under certain configurations (e.g., fixed signal-to-noise ratio or fixed active probability) has obvious performance degradation under different settings. A mixed training dataset with different settings may help to improve the robustness of data-driven methods. However, the limited learning capability of the network also results in performance degradation, while networks with more parameters face the problem of increased training time and computation complexity issues. Alternatively, we can train multiple networks for different SNRs, and choose the suitable one at the inference time. Or we can use transfer learning to quickly fit the trained network to a specific SNR. Another direction is integrating the SNR into network design, like the network in [11].

![Figure 4](image1.png)  
**FIGURE 4.** The convergence rate and performance of the JADEC for the CS method and data-driven methods, respectively. The horizontal axis represents the number of iterations of ISTA and the number of layers of networks, which is related to the computational complexity. The vertical axis represents the normalized mean square error of the reconstructed sparse signal.

![Figure 5](image2.png)  
**FIGURE 5.** Performance of the JADEC for the CS method and data-driven methods, respectively, under different numbers of active users.
Conclusions

In this article, we discuss the data-driven CS for grant-free random access, which performs the activity detection and channel estimation using an NN to solve the sparse recovery problem. Data-driven methods decrease the computational complexity of the receiver and enhance performance, mainly due to leveraging domain knowledge from communication systems into the NN design. We show how data-driven methods can improve performance, and show several variations of NNs for CS-based grant-free random access. Finally, we summarize the challenges and future directions of data-driven CS methods for mMTC.

References

[1] L. Liu et al., “Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things,” IEEE Signal Processing Mag., vol. 35, no. 5, 2018, pp. 88–99.
[2] D. Donoho, “Compressed Sensing,” IEEE Trans. Info. Theory, vol. 52, no. 4, 2006, pp. 1289–1306.
[3] K. Gregor and Y. LeCun, “Learning Fast Approximations of Sparse Coding,” Proc. 27th Int’l Conf. Machine Learning, 2010, pp. 399–406.
[4] A. Mousavi, A. B. Patel, and R. G. Baraniuk, “A Deep Learning Approach to Structured Signal Recovery,” 2015 53rd Annual Allerton Conf. Commun., Control, and Computing, 2015, pp. 1336–43.
[5] M. Bogerdin, P. Schioler, and S. Rangan, “AMP-Inspired Deep Networks for Sparse Linear Inverse Problems,” IEEE Trans. Signal Processing, vol. 65, no. 16, 2017, pp. 4293–4308.
[6] A. Aberdam, A. Golts, and M. Elad, “Ada-Lista: Learned Solvers Adaptive to Varying Models,” IEEE Trans. Pattern Analysis and Machine Intelligence, 2021.
[7] X. Shao et al., “Feature-Aided Adaptive-Tuning Deep Learning for Massive Device Detection,” IEEE JSAC, vol. 39, no. 7, 2021, pp. 1899–1914.
[8] Y. Abe, W. Kim, and B. Shim, “Active User Detection and Channel Estimation for Massive Machine-Type Communication: Deep Learning Approach,” IEEE IoT J., vol. 9, no. 2, 2022, pp. 11904–17.
[9] Y. Bai et al., “Prior Information Aided Deep Learning Method for Grant-Free NOMA in mMTC,” IEEE JSAC, vol. 40, no. 1, 2022, pp. 112–26.
[10] W. Zhu et al., “Deep-Learned Approximate Message Passing for Asynchronous Massive Connectivity,” IEEE Trans. Wireless Commun., 2021.
[11] W. Chen et al., “Solving Sparse Linear Inverse Problems in Communication Systems: A Deep Learning Approach with Adaptive Depth,” IEEE JSAC, vol. 39, no. 1, 2021, pp. 4–17.
[12] Z. Mao et al., “Joint Channel Estimation and Active-User Detection for Massive Access in Internet of Things — A Deep Learning Approach,” IEEE IoT J., vol. 9, no. 4, 2022, pp. 2870–81.
[13] Y. Shi et al., “Algorithm Unrolling for Massive Access via Deep Neural Networks with Theoretical Guarantee,” IEEE Trans. Wireless Commun., vol. 21, no. 2, 2022, pp. 945–59.
[14] Y. Cui, S. Li, and W. Zhang, “Jointly Sparse Signal Recovery and Support Recovery via Deep Learning with Applications in MIMO-Based Grant-Free Random Access,” IEEE JSAC, vol. 39, no. 3, 2021, pp. 788–803.
[15] Y. Zou, Z. Qin, and Y. Liu, “Joint User Activity and Data Detection in Grant-Free Noma Using Generative Neural Networks,” IEEE Access, 2021.

Biographies

YANNA BAI (yanbabai@bjtu.edu.cn) received her B.Eng. and M.Eng. degrees from Beijing Jiaotong University, China, in 2016 and 2018, respectively, and is pursuing a Ph.D. degree at the same university. Her current research interests include AI-assisted communications and massive machine-type communications.

WEI CHEN (M’13, SM’18) (weich@bjtu.edu.cn) received his B.Eng. and M.Eng. degrees from Beijing University of Posts and Telecommunications, China, in 2006 and 2009, respectively, and his Ph.D. degree in computer science from the University of Cambridge, United Kingdom, in 2013. Later, he was a research associate with the Computer Laboratory, University of Cambridge from 2013 to 2016. He is currently a professor with Beijing Jiaotong University. His current research interests include intelligent wireless communication systems and multimedia processing.

FEIFEI SUN (feifei.sun@samsung.com) received her Ph.D. in communication and information engineering from Beijing University of Posts and Telecommunications. She is now a principal engineer at Samsung Research China — Beijing (SRC-B). Her research interests are in B5G/6G wireless communications standardization on IoT-related technologies such as NB-IoT, eMTC, RedCap, URLLC, as well as AI/ML in the physical layer.

BO AI (M’00, SM’10) (boai@bjtu.edu.cn) received his M.S. and Ph.D. degrees from Xidian University, Xian, China, in 2002 and 2004, respectively. He is currently a professor at Beijing Jiaotong University, where he is also the deputy director of the State Key Laboratory of Rail Traffic Control and Safety. His interests include high-power amplifier linearization techniques, radio propagation and channel modeling, global systems for mobile communications for railway systems, and LTE for railway systems.

PETAR POPOVSKI (petarp@es.aau.dk) is a professor at Aalborg University, where he heads the Connectivity section, and a visiting Excellence Chair at the University of Bremen. He received Dipl.-Ing and M. Sc. degrees in communication engineering from the University of Sts. Cyril and Methodius, Skopje, and his Ph.D. degree from Aalborg University (2004). He is currently an Editor-in-Chief of the IEEE Journal on Selected Areas in Communications. His research interests include wireless communications and communication theory.