

Federated Learning for Future Intelligent Wireless Networks

Edited by

Yao Sun • Chaoqun You
Gang Feng • Lei Zhang




IEEE PRESS

WILEY

Federated Learning for Future Intelligent Wireless Networks

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

IEEE Press Editorial Board
Sarah Spurgeon, *Editor in Chief*

Jón Atli Benediktsson
Anjan Bose
James Duncan
Amin Moeness
Desineni Subbaram Naidu

Behzad Razavi
Jim Lyke
Hai Li
Brian Johnson

Jeffrey Reed
Diomidis Spinellis
Adam Drobot
Tom Robertazzi
Ahmet Murat Tekalp

Federated Learning for Future Intelligent Wireless Networks

Edited by

Yao Sun

University of Glasgow, UK

Chaoqun You

Singapore University of Technology and Design, Singapore

Gang Feng

University of Electronic Science and Technology of China, China

Lei Zhang

University of Glasgow, UK



IEEE PRESS
WILEY

Copyright © 2024 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data Applied for:

[Hardback ISBN: 9781119913894]

Cover Design: Wiley

Cover Image: © Blue Planet Studio/Shutterstock

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

Contents

About the Editors xv

Preface xvii

1	Federated Learning with Unreliable Transmission in Mobile Edge Computing Systems	1
	<i>Chenyuan Feng, Daquan Feng, Zhongyuan Zhao, Howard H. Yang, and Tony Q. S. Quek</i>	
1.1	System Model	1
1.1.1	Local Model Training	1
1.1.2	Update Result Feedback via the Wireless Channels	2
1.1.3	Global Model Averaging	3
1.2	Problem Formulation	4
1.2.1	Model Accuracy Loss	4
1.2.2	Communication Loss $\epsilon_{t,m}^C$	6
1.2.3	Sample Selection Loss $\epsilon_{t,m}^S$	6
1.2.4	Model Training Loss $\epsilon_{t,m}^M$	7
1.2.5	Problem Formulation	8
1.2.5.1	Objective Function	8
1.2.5.2	Energy Consumption Constraint	9
1.2.5.3	User Selection Constraint	9
1.2.5.4	Data Volume Constraint of Local Training Datasets	9
1.3	A Joint Optimization Algorithm	10
1.3.1	Compression Optimization	10
1.3.1.1	Optimization of $A_{t,m}$	10
1.3.1.2	Optimization of $D_{t,m}$	12
1.3.2	Joint Optimization of $A_{t,m}$ and $D_{t,m}$	13
1.3.3	Optimization of Sample Selection	13
1.3.4	Optimization of User Selection	15
1.3.5	A Joint Optimization Algorithm	15

1.4	Simulation and Experiment Results	16
	Bibliography	21
2	Federated Learning with non-IID data in Mobile Edge Computing Systems	23
	<i>Chenyuan Feng, Daquan Feng, Zhongyuan Zhao, Geyong Min, and Hancong Duan</i>	
2.1	System Model	23
2.1.1	Local Model Training	23
2.1.2	Federated Averaging	24
2.2	Performance Analysis and Averaging Design	24
2.2.1	The Analysis of Expected Weight Divergence	26
2.2.1.1	The Analysis of Expected Data Distribution Divergence $\mathbb{E}\{\mathcal{L}_m\}$	27
2.2.1.2	An Upper Bound of $\delta(tK)$	28
2.2.2	Rethinking the Settings of Federated Averaging Weights	28
2.3	Data Sharing Scheme	30
2.3.1	Data Sharing	30
2.3.2	Problem Formation	31
2.3.2.1	Objective Function	32
2.3.3	Optimization Constraints	33
2.3.4	A Joint Optimization Algorithm	34
2.3.4.1	CPU Cycle Frequency Optimization Subproblem	34
2.3.4.2	Transmit Power Allocation Subproblem	35
2.3.4.3	Sharing Dataset Optimization Subproblem	37
2.3.4.4	User Selection Optimization Subproblem	39
2.3.4.5	A Joint Optimization Algorithm	41
2.4	Simulation Results	42
	Bibliography	47
3	How Many Resources Are Needed to Support Wireless Edge Networks	49
	<i>Yi-Jing Liu, Gang Feng, Yao Sun, and Shuang Qin</i>	
3.1	Introduction	49
3.2	System Model	50
3.2.1	FL Model	51
3.2.1.1	Loss Function	51
3.2.1.2	Updating Model	52
3.2.2	Computing Resource Consumption Model	52
3.2.3	Communication Resource Consumption Model	53
3.2.3.1	Uplink	53
3.2.3.2	DownLink	53

3.3	Wireless Bandwidth and Computing Resources Consumed for Supporting FL-Enabled Wireless Edge Networks	54
3.3.1	SINR Analysis (Uplink Direction)	54
3.3.1.1	Probability Density Function (PDF) of SINR	54
3.3.1.2	Transmission Success Probability of Local Models	55
3.3.2	SNR Analysis (Downlink Direction)	57
3.3.3	Wireless Bandwidth Needed for Transmitting Local/Global Models	57
3.3.4	Computing Resources Needed for Training Local Models	58
3.4	The Relationship between FL Performance and Consumed Resources	59
3.4.1	Local Model Accuracy	59
3.4.2	Global Model Accuracy	60
3.5	Discussions of Three Cases	62
3.5.1	Case 1: Sufficient Communication Resources and Computing Resources	62
3.5.2	Case 2: Sufficient Computing Resources and Insufficient Communication Resources	63
3.5.3	Case 3: Sufficient Communication Resources and Insufficient Computing Resources	64
3.6	Numerical Results and Discussion	67
3.6.1	Simulation Setting	67
3.6.2	Simulation Results	68
3.6.2.1	Verifying Analytical Results	68
3.6.2.2	Measuring the Performance of FL Settings	71
3.6.2.3	Examining the Trade-Off between the Computing and Communication Resources under FL Framework	73
3.7	Conclusion	75
3.8	Proof of Corollary 3.2	76
3.9	Proof of Corollary 3.3	77
	References	81
4	Device Association Based on Federated Deep Reinforcement Learning for Radio Access Network Slicing	85
	<i>Yi-Jing Liu, Gang Feng, Yao Sun, and Shuang Qin</i>	
4.1	Introduction	85
4.2	System Model	87
4.2.1	Network Model	87
4.2.2	RAN Slicing	88
4.2.3	Service Requirements	89
4.2.4	Handoff Cost	89

4.3	Problem Formulation	90
4.3.1	Problem Statement	90
4.3.2	Markov Decision Process Modeling for Device Association	92
4.3.2.1	State	92
4.3.2.2	Action	93
4.3.2.3	Transition Probability	93
4.3.2.4	Reward	93
4.4	Hybrid Federated Deep Reinforcement Learning for Device Association	94
4.4.1	Framework of HDRL	94
4.4.1.1	DRL on Smart Devices	94
4.4.1.2	Horizontal Model Aggregation (hDRL) Level	95
4.4.1.3	Vertical Model Aggregation (vDRL) Level	95
4.4.2	Algorithm of Horizontal Model Aggregation	97
4.4.2.1	DDQN for Training Local Model	97
4.4.2.2	Update Models	98
4.4.3	Algorithm of Vertical Model Aggregation	98
4.4.4	HDRL Algorithm for Device Association	100
4.4.5	Convergence Analysis	102
4.5	Numerical Results	103
4.5.1	Simulation Settings	104
4.5.2	Numerical Results and Discussions	105
4.6	Conclusion	109
	Acknowledgment	110
	References	110
5	Deep Federated Learning Based on Knowledge Distillation and Differential Privacy	113
	<i>Hui Lin, Feng Yu, and Xiaoding Wang</i>	
5.1	Introduction	113
5.2	Related Work	115
5.3	System Model	118
5.3.1	Security Model	118
5.4	The Implementation Details of the Proposed Strategy	119
5.4.1	Security Analysis	119
5.5	Performance Evaluation	120
5.5.1	Experimental Environment	120
5.5.2	Experimental Results	120
5.6	Conclusions	122
	Bibliography	123

6	Federated Learning-Based Beam Management in Dense Millimeter Wave Communication Systems	127
	<i>Qing Xue and Liu Yang</i>	
6.1	Introduction	127
6.1.1	Prior Work	128
6.1.2	Contributions	129
6.2	System Model	130
6.3	Problem Formulation and Analysis	133
6.4	FL-Based Beam Management in UDmmN	135
6.4.1	Markov Decision Process Model	135
6.4.2	FL-Based Beam Management	136
6.4.2.1	Data Cleaning	136
6.4.2.2	Model Training	138
6.4.3	User Association in UDmmN	141
6.4.3.1	Downlink Measurements	141
6.4.3.2	User Perception	141
6.4.3.3	Multiple Association	142
6.5	Performance Evaluation	143
6.5.1	Comparison with BFS and EDB	146
6.5.2	Comparison with BMDL and BMCL	148
6.6	Conclusions	150
	Bibliography	150
7	Blockchain-Empowered Federated Learning Approach for An Intelligent and Reliable D2D Caching Scheme	155
	<i>Runze Cheng, Yao Sun, Yijing Liu, Le Xia, Daquan Feng, and Muhammad Imran</i>	
7.1	Introduction	155
7.2	Related Work	157
7.2.1	Learning-Based D2D Caching Schemes	157
7.2.2	Blockchain-Enabled D2D Caching Schemes	158
7.3	System Model	159
7.3.1	D2D Network	159
7.3.2	Content Caching Schemes	159
7.3.3	Transmission Latency	160
7.4	Problem Formulation and DRL-Based Model Training	160
7.4.1	Problem Formulation	161
7.4.1.1	Action	161
7.4.1.2	State	161
7.4.1.3	Reward and Return	161
7.4.2	DRL-Based Local Model Training	163

7.5	Privacy-Preserved and Secure BDRFL Caching Scheme Design	165
7.5.1	Task and Requirements Publication	166
7.5.2	Appropriate UE Selection	166
7.5.3	Local Model Training	166
7.5.4	Area Model Update and Recording	169
7.5.5	Global Model Update and Recording	169
7.6	Consensus Mechanism and Federated Learning Model Update	170
7.6.1	Double-Layer Blockchain Consensus Mechanism	170
7.6.2	FL Area Model Update in Subchain Layer	171
7.6.3	FL Global Model Update in Mainchain Layer	172
7.7	Simulation Results and Discussions	173
7.7.1	Simulation Setting	174
7.7.2	Numerical Results	175
7.8	Conclusion	177
	References	178

8 Heterogeneity-Aware Dynamic Scheduling for Federated Edge Learning 181

Kun Guo, Zihan Chen, Howard H. Yang, and Tony Q. S. Quek

8.1	Introduction	181
8.2	Related Works	184
8.3	System Model for FEEL	185
8.3.1	Flow of FEEL with Scheduling	186
8.3.2	Delay and Energy Model in FEEL	187
8.3.2.1	Delay Model	187
8.3.2.2	Energy Model	188
8.4	Heterogeneity-Aware Dynamic Scheduling Problem Formulation	189
8.4.1	Convergence of FEEL with Scheduling	189
8.4.2	Scheduling Policy with Sequential Transmission	190
8.4.3	Problem Formulation	191
8.5	Dynamic Scheduling Algorithm Design and Analysis	192
8.5.1	Benchmark: R -Round Lookahead Algorithm	192
8.5.2	DISCO: Dynamic Scheduling Algorithm	193
8.5.3	Algorithm Analysis, Complexity Reduction, and Implementation Discussion	196
8.5.3.1	Algorithm Analysis	196
8.5.3.2	Complexity Reduction	197
8.5.3.3	Implementation Discussion	197
8.6	Evaluation Results	197
8.6.1	Parameter Settings	198
8.6.2	Numerical Results	200

8.6.3	Experimental Results	203
8.7	Conclusions	208
8.A.1	Proof of Theorem 8.2	208
8.A.2	Proof of Theorem 8.3	209
8.A.2.1	Feasibility Proof	210
8.A.2.2	Optimality Proof	211
	References	212

9 Robust Federated Learning with Real-World Noisy Data 215

Jingyi Xu, Zihan Chen, Tony Q. S. Quek, and Kai Fong Ernest Chong

9.1	Introduction	215
9.1.1	Work Prior to FedCorr	215
9.2	Related Work	217
9.2.1	Federated Methods	217
9.2.2	Local Intrinsic Dimension (LID)	218
9.2.2.1	Estimation of LID	219
9.3	FedCorr	219
9.3.1	Preliminaries	221
9.3.1.1	Data Partition	221
9.3.1.2	Noise Model	222
9.3.1.3	LID Scores for Local Models	222
9.3.2	Federated Preprocessing Stage	222
9.3.2.1	Client Iteration and Fraction Scheduling	223
9.3.2.2	Mixup and Local Proximal Regularization	223
9.3.2.3	Identification of Noisy Clients and Noisy Samples	224
9.3.3	Federated Fine-Tuning Stage	225
9.3.4	Federated Usual Training Stage	226
9.4	Experiments	226
9.4.1	Experimental Setup	227
9.4.1.1	Baselines	227
9.4.1.2	Implementation Details	227
9.4.2	Comparison with State-of-the-Art Methods	228
9.4.2.1	IID Settings	228
9.4.2.2	Non-IID Settings	228
9.4.2.3	Combination with Other FL Methods	231
9.4.2.4	Comparison of Communication Efficiency	231
9.4.3	Ablation Study	232
9.5	Further Remarks	232
	Bibliography	234

10	Analog Over-the-Air Federated Learning: Design and Analysis	239
	<i>Howard H. Yang, Zihan Chen, and Tony Q. S. Quek</i>	
10.1	Introduction	239
10.2	System Model	241
10.3	Analog Over-the-Air Model Training	242
10.3.1	Salient Features	243
10.3.2	Heavy-Tailed Interference	244
10.4	Convergence Analysis	245
10.4.1	Preliminaries	245
10.4.2	Convergence Rate of AirFL	246
10.4.3	Key Observations	250
10.5	Numerical Results	250
10.6	Conclusion	253
	Bibliography	253
11	Federated Edge Learning for Massive MIMO CSI Feedback	257
	<i>Shi Jin, Yiming Cui, and Jiajia Guo</i>	
11.1	Introduction	257
11.2	System Model	259
11.2.1	Channel Model and Signal Model	259
11.2.2	DL-Based CSI Feedback	259
11.3	FEEL for DL-Based CSI Feedback	260
11.3.1	Basic Autoencoder Architecture	260
11.3.2	FEEL-Based Training Framework	261
11.3.2.1	Motivation	261
11.3.2.2	Training Framework	261
11.3.3	Parameter Quantization in the FEEL-Based Training Framework	263
11.3.3.1	Key Idea	263
11.3.3.2	Details	264
11.4	Simulation Results	264
11.4.1	Simulation Settings	264
11.4.1.1	Channel Generation	264
11.4.1.2	Training Settings	264
11.4.1.3	Performance of FEEL-Based Training Framework	265
11.5	Conclusion	268
	Bibliography	269

12	User-Centric Decentralized Federated Learning for Autoencoder-Based CSI Feedback	273
	<i>Shi Jin, Jijia Guo, Yan Lv, and Yiming Cui</i>	
12.1	Autoencoder-Based CSI Feedback	273
12.1.1	CSI Feedback in Massive MIMO	273
12.1.2	System Model	274
12.1.3	AE-Based Feedback Framework	275
12.2	User-Centric Online Training for AE-Based CSI Feedback	275
12.2.1	Motivation	275
12.2.2	Key Idea of User-Centric Online Training	276
12.2.3	Three Online Training Frameworks	277
12.2.3.1	Edit Before Encoder	277
12.2.3.2	Edit After Encoder	278
12.2.3.3	Finetuning Encoder	279
12.3	Multiuser Online Training Using Decentralized Federated Learning	279
12.3.1	Motivation	279
12.3.2	Decentralized Federated Learning Framework	280
12.3.2.1	Key Idea of Decentralized Federated Learning	280
12.3.2.2	Merge Function	281
12.3.2.3	Network Connectivity Architecture	282
12.4	Numerical Results	283
12.4.1	Simulation Setting	283
12.4.1.1	Channel Generation	283
12.4.1.2	Training Details	283
12.4.2	Performance of User-Centric Online Training	283
12.4.3	Performance of Multiuser Online Training Using DFL	284
12.4.3.1	Performance Evaluation of DFL framework	284
12.4.3.2	NMSE Under Different Gossip Numbers	286
12.4.3.3	NMSE Under Different Numbers of UE Participation	286
12.4.3.4	NMSE Under Mismatched Channels	286
12.5	Conclusion	287
	Bibliography	287

Index 291

About the Editors

Yao Sun is currently a lecturer with James Watt School of Engineering, the University of Glasgow, Glasgow, United Kingdom. He has extensive research experience and has published widely in wireless networking research. He has won the IEEE Communication Society of TAOS Best Paper Award in 2019 ICC, IEEE IoT Journal Best Paper Award in 2022, and Best Paper Award in 22nd ICCT. He has been the guest editor for special issues of several international journals. He has served as TPC Chair for UCET 2021 and TPC member for a number of international flagship conferences, including ICC 2022, VTC Spring 2022, GLOBECOM 2020, and WCNC 2019. His research interests include intelligent wireless networking, semantic communications, blockchain systems, and resource management in next-generation mobile networks. He is a senior member of IEEE.

Chaoqun You is a postdoctoral research fellow at the Singapore University of Technology and Design (SUTD). She received the BS degree in communication engineering and the PhD degree in communication and information systems from the University of Electronic Science and Technology of China (UESTC) in 2013 and 2020, respectively. She was a visiting student at the University of Toronto from 2015 to 2017. Her current research interests include mobile edge computing, network virtualization, O-RAN, federated learning, and 6G.

Gang Feng received his BEng and MEng degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC) in 1986 and 1989, respectively, and the PhD degree in Information Engineering from the Chinese University of Hong Kong in 1998. At present, he is a professor with the National Key Laboratory of Wireless Communications, UESTC of China. His research interests include resource management in wireless networks, next-generation cellular networks, etc. Dr. Feng is a senior member of IEEE.

Lei Zhang is a Professor of Trustworthy Systems at the University of Glasgow. He has combined academia and industry research experience on wireless communications and networks and distributed systems for IoT, blockchain,

and autonomous systems. His 20 patents have been granted/filed in 30+ countries/regions. He published 3 books and 150+ papers in peer-reviewed journals, conferences, and edited books. He received the IEEE Internet of Things Journal Best Paper Award 2022, IEEE ComSoc TAOS Technical Committee Best Paper Award 2019, and IEEE ICEICT'21 Best Paper Award.

Preface

It has been considered one of the key missing components in the existing 5G network and is widely recognized as one of the most sought-after functions for next-generation 6G communication systems. Nowadays, there are more than 10 billion Internet-of-Things (IoT) equipment and 5 billion smartphones that are equipped with artificial intelligence (AI)-empowered computing modules such as AI chips and GPU. On the one hand, the user equipment (UE) can be potentially deployed as computing nodes to process certain emerging service tasks such as crowdsensing tasks and collaborative tasks, which paves the way for applying AI in edge networks. On the other hand, in the paradigm of machine learning (ML), the powerful computing capability on these UEs can decouple ML from acquiring, storing, and training data in data centers as conventional methods.

Federated learning (FL) has been widely acknowledged as one of the most essential enablers to bring network edge intelligence into reality, as it can enable collaborative training of ML models while enhancing individual user privacy and data security. Empowered by the growing computing capabilities of UEs, FL trains ML models locally on each device where the raw data never leaves the device. Specifically, FL uses an iterative approach that requires a number of global iterations to achieve a global model accuracy. In each global iteration, UEs take a number of local iterations up to a local model accuracy. As a result, the implementation of FL at edge networks can also decrease the costs of transmitting raw data, relieve the burden on backbone networks, reduce the latency for real-time decisions.

This book would explore recent advances in the theory and practice of FL, especially when it is applied to wireless communication systems. In detail, the book covers the following aspects:

- 1) principles and fundamentals of FL;
- 2) performance analysis of FL in wireless communication systems;

- 3) how future wireless networks (say 6G networks) enable FL as well as how FL frameworks/algorithms can be optimized when applying to wireless networks (6G);
- 4) FL applications to vertical industries and some typical communication scenarios.

Chapter 1 investigates the optimization design of FL in the edge network. First, an optimization problem is formulated to manage the trade-off between model accuracy and training cost. Second, a joint optimization algorithm is designed to optimize the model compression, sample selection, and user selection strategies, which can approach a stationary optimal solution in a computationally efficient way. Finally, the performance of the proposed optimization scheme is evaluated by numerical simulation and experiment results, which show that both the accuracy loss and the cost of FL in the edge network can be reduced significantly by employing the proposed algorithm.

Chapter 2 studies non-IID data model for FL, derives a theoretical upper bound, and redesigns the federated averaging scheme to reduce the weight difference. To further mitigate the impact of non-IID data, a data-sharing scheme is designed to jointly minimize the accuracy loss, the energy consumption, and latency with constrained resource of edge systems. Then a computation-efficient algorithm is proposed to approach the optimal solution and provide the experiment results to evaluate our proposed schemes.

Chapter 3 theoretically analyzes the performance and cost of running FL, which is imperative to deeply understand the relationship between FL performance and multiple-dimensional resources. In this chapter, we construct an analytical model to investigate the relationship between the FL model accuracy and consumed resources in FL-enabled wireless edge networks. Based on the analytical model, we explicitly quantify the model accuracy, computing resources, and communication resources. Numerical results validate the effectiveness of our theoretical modeling and analysis and demonstrate the trade-off between the communication and computing resources for achieving a certain model accuracy.

Chapter 4 proposes an efficient device association scheme for radio access network (RAN) slicing by exploiting a federated reinforcement learning framework, with the aim to improve network throughput, while guaranteeing user privacy and data security. Specially, we use deep reinforcement learning to train local models on UEs under a hybrid FL framework, where horizontally FL is employed for parameter aggregation on BS, while vertically FL is employed for access selection aggregation on the encrypted party. Numerical results show that our proposed scheme can achieve significant performance gains in terms of network throughput and communication efficiency in comparison with some known state-of-the-art solutions.

Chapter 5 proposes a deep FL algorithm that utilizes knowledge distillation and differential privacy to safeguard privacy during the data fusion process. Our approach involves adding Gaussian noise at different stages of knowledge distillation-based FL to ensure privacy protection. Our experimental results demonstrate that this strategy provides better privacy preservation while achieving high-precision IoT data fusion.

Chapter 6 presents a novel systematic beam control scheme to tackle the formulated beam management problem, which is difficult due to the nonconvex objective function. The double deep Q-network (DDQN) under a FL framework is employed to solve the above optimization problem, thereby fulfilling adaptive and intelligent beam management in mmwave networks. In the proposed beam management scheme based on federated learning (BMFL), the non-raw-data aggregation can theoretically protect user privacy while reducing handoff costs. Moreover, a data cleaning technique is used before the local model training, with the aim to further strengthen the privacy protection while improving the learning convergence speed. Simulation results demonstrate the performance gain of the proposed BMFL scheme.

Chapter 7 proposes a double-layer blockchain-based deep reinforcement federated learning (BDRFL) scheme to ensure privacy-preserved and caching-efficient D2D networks. In BDRFL, a double-layer blockchain is utilized to further enhance data security. Simulation results first verify the convergence of BDRFL-based algorithm and then demonstrate that the download latency of the BDRFL-based caching scheme can be significantly reduced under different types of attacks when compared with some existing caching policies.

Chapter 8 aims to design a dynamic scheduling policy to explore the spectrum flexibility for heterogeneous federated edge learning (FEEL) so as to facilitate the distributed intelligence in edge networks. This chapter proposes a heterogeneity-aware dynamic scheduling problem to minimize the global loss function, with consideration of straggler and limited device energy issues. By solving the formulated problem, we propose a dynamic scheduling algorithm (DISCO), to make an intelligent decision on the set and order of scheduled devices in each communication round. Theoretical analysis reveals that under certain conditions, learning performance and energy constraints can be guaranteed in the DISCO. Finally, we demonstrate the superiority of the DISCO through numerical and experimental results, respectively.

Chapter 9 discusses FedCorr, a general multistage framework to tackle heterogeneous label noise in FL, which does not make any assumptions on the noise models of local clients while still maintaining client data privacy. Both theoretical analysis and experiment results demonstrate the performance gain of this novel FL framework.

Chapter 10 provides a general overview of the analog over-the-air federated learning (AirFL) system. Specially, we illustrate the general system architecture and highlight the salient feature of AirFL that adopts analog transmissions for fast (but noisy) aggregation of intermediate parameters. Then, we establish a new convergence analysis framework that takes into account the effects of fading and interference noise. Our analysis unveils the impacts from the intrinsic properties of wireless transmissions on the convergence performance of AirFL. The theoretical findings are corroborated by extensive simulations.

Chapter 11 investigates a FEEL-based training framework to DL-based channel state information (CSI) feedback. In FEEL, each UE trains an autoencoder network locally and exchanges model parameters via the base station. Therefore, data privacy is better protected compared with centralized learning because the local CSI datasets are not required to be uploaded. Neural network parameter quantization is then introduced to the FEEL-based training framework to reduce communication overhead. The simulation results indicate that the proposed FEEL-based training framework can achieve comparable performance with centralized learning.

Chapter 12 proposes a user-centric online training strategy in which the UE can collect CSI samples in the stable area and adjust the pretrained encoder online to further improve CSI reconstruction accuracy. Moreover, the proposed online training framework is extended to the multiuser scenario to improve performance sequentially. The key idea is to adopt decentralized FL without BS participation to combine the sharing of channel knowledge among UEs, which is called crowd intelligence. Simulation results show that the decentralized FL-aided framework has higher feedback accuracy than the AE without online training.

November 2023

*Yao Sun
Chaoqun You
Gang Feng
Lei Zhang*

1

Federated Learning with Unreliable Transmission in Mobile Edge Computing Systems

Chenyuan Feng¹, Daquan Feng¹, Zhongyuan Zhao², Howard H. Yang³, and Tony Q. S. Quek⁴

¹Shenzhen Key Laboratory of Digital Creative Technology, The Guangdong Province Engineering Laboratory for Digital Creative Technology, The Guangdong-Hong Kong Joint Laboratory for Big Data Imaging and Communication, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, Guangdong, China

²State Key Laboratory of Networking and Switching Technology, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

³Zhejiang University/University of Illinois at Urbana-Champaign Institute, Zhejiang University, The College of Information Science and Electronic Engineering, Haining, Zhejiang, China

⁴Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore

1.1 System Model

Consider the deployment of FL in an MEC scenario, which consists of an edge access point E and multiple users U_1, \dots, U_M . An edge computing server S_E is equipped with E , while a local computing unit S_m is equipped with U_m , $m = 1, \dots, M$. As shown in Figure 1.1, the edge computing server S_E and local computing units S_1, \dots, S_M can act as the computing server and the clients, respectively, which can interact with each other via the wireless channels between E and U_1, \dots, U_M .

As introduced previously, FL can be implemented by updating the local models and the global model iteratively. In particular, we focus on the t th iteration, which can be introduced as follows.

1.1.1 Local Model Training

In this phase, each user updates the local model independently based on its local collected data. Without loss of generality, we focus on a specific user U_m , the local model of U_m can be updated as follows:

$$\mathbf{w}_{t,m} = \mathbf{w}_{t-1,m} - \eta_t \nabla F(\mathbf{w}_{t-1,m}, D_{t,m}), \quad (1.1)$$

Federated Learning for Future Intelligent Wireless Networks, First Edition.

Edited by Yao Sun, Chaoqun You, Gang Feng, and Lei Zhang.

© 2024 The Institute of Electrical and Electronics Engineers, Inc. Published 2024 by John Wiley & Sons, Inc.

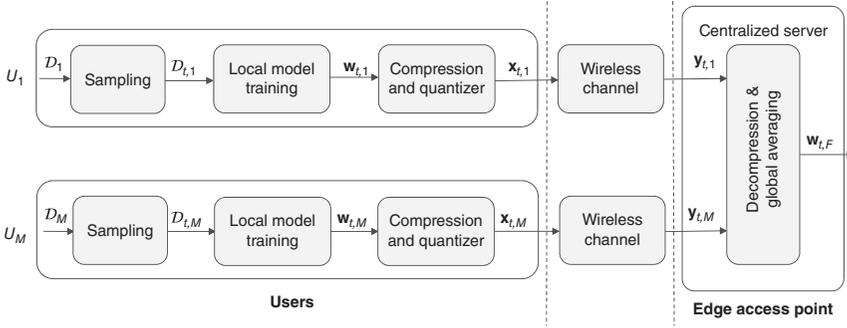


Figure 1.1 The system model of wireless FL.

where $\mathbf{w}_{t,m}$ and $\mathbf{w}_{t-1,m}$ denote the update results of U_m 's local model during the t th and $(t-1)$ -th iterations, respectively, $D_{t,m}$ denotes the training dataset for updating $\mathbf{w}_{t,m}$, which is randomly selected from D_m , $D_{t,m} \subseteq D_m$, D_m denotes the local dataset located at U_m , η_t is the learning rate of the t th iteration, and $\nabla F(\mathbf{w}_{t-1,m}, D_{t,m})$ is the gradient of loss function $F(\mathbf{w}_{t-1,m}, D_{t,m})$ with respect to $\mathbf{w}_{t-1,m}$. In this chapter, the loss function is defined as the empirical risk with respect to $\mathbf{w}_{t,m}$, which can be defined as follows:

$$F(\mathbf{w}_{t,m}, D_{t,m}) = \frac{1}{N_{t,m}} \sum_{\mathbf{x} \in D_{t,m}} l(\mathbf{w}_{t,m}; \mathbf{x}), \quad (1.2)$$

where $l(\mathbf{w}_{t,m}; \mathbf{x})$ denotes the loss function of the data element \mathbf{x} , and $N_{t,m}$ denotes the size of $D_{t,m}$.

1.1.2 Update Result Feedback via the Wireless Channels

When the local model training procedure is accomplished, U_m should transmit its update result $\mathbf{w}_{t,m}$ to E via the wireless channel. In the existing works, the server randomly selects the users since it is assumed that the communications between the computing server and the clients are ideal. However, it cannot be ensured in the MEC systems due to the unreliable wireless transmission circumstances, which will cause accuracy loss of FL models. Therefore, in this chapter, only the users with high communication reliability and low model accuracy loss are scheduled to participate in each iteration of global model averaging. In particular, the scheduling status of U_m for the t th iteration of global averaging is characterized by a Boolean variable $z_{t,m}$, i.e.,

$$z_{t,m} = \begin{cases} 1, & \text{if } U_m \text{ is scheduled} \\ 0, & \text{if } U_m \text{ is not scheduled.} \end{cases} \quad (1.3)$$

If U_m is scheduled, its update result $\mathbf{w}_{t,m}$ can be modeled as a $d \times 1$ vector, which is usually with high dimension, especially for the deep neural network models. Therefore, to improve the efficiency of update result feedback, **model sparsification** and **parameter quantization** techniques should be employed to compress $\mathbf{w}_{t,m}$. As introduced previously, $\mathbf{w}_{t,m}$ can be transformed into a sparse form via model sparsification, which can be expressed as follows:

$$\mathbf{s}_{t,m} = \mathbf{A}_{t,m} \mathbf{w}_{t,m}, \quad (1.4)$$

where $\mathbf{A}_{t,m}$ denotes a $d \times d$ sparsification matrix for $\mathbf{w}_{t,m}$.

Next, each element of $\mathbf{s}_{t,m}$ is quantized independently by employing uniform quantization. The quantization error can be approximated as an additive Gaussian noise, which is independent with $\mathbf{w}_{t,m}$. Then the quantized parameter vector can be expressed as

$$\mathbf{x}_{t,m} = \mathbf{s}_{t,m} + \mathbf{q}_{t,m} = \mathbf{A}_{t,m} \mathbf{w}_{t,m} + \mathbf{q}_{t,m}, \quad (1.5)$$

where $\mathbf{q}_{t,m}$ denotes a $d \times 1$ quantization noise vector, i.e., $\mathbf{q}_{t,m} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{t,m})$, and $\mathbf{\Omega}_{t,m}$ denotes the covariance matrix. Due to the implementation of independent quantization, each element of $\mathbf{q}_{t,m}$ is independent with each other, i.e., $\mathbb{E}\{q_i^{t,m}(q_j^{t,m})^H\} = \omega_i^{t,m}$, $\mathbb{E}\{q_i^{t,m}(q_j^{t,m})^H\} = 0$, $i \neq j$, where $q_i^{t,m}$ and $q_j^{t,m}$ denote the i th and j th elements of $\mathbf{q}_{t,m}$, respectively. Therefore, $\mathbf{\Omega}_{t,m}$ is a diagonal matrix, which can be denoted as $\mathbf{\Omega}_{t,m} = \text{diag}\{q_1^{t,m}, \dots, q_d^{t,m}\}$.

After model sparsification and parameter quantization, $\mathbf{x}_{t,m}$ is suitable for base-band processing and wireless transmissions. In this chapter, the flat fading model is employed to characterize the wireless channels between U_m and E . Therefore, the channel fading can be assumed to be unchanged during the transmission of $\mathbf{x}_{t,m}$. Then the observation of $\mathbf{x}_{t,m}$ at E can be expressed as

$$\mathbf{y}_{t,m} = h_{t,m} \mathbf{x}_{t,m} + \mathbf{n}_{t,m}, \quad (1.6)$$

where $h_{t,m}$ captures the flat channel fading of the wireless link between U_m and E , $\mathbf{n}_{t,m}$ denotes the additive white Gaussian noise at E , i.e., $\mathbf{n}_{t,m} \sim \mathcal{CN}(\mathbf{0}, \sigma_t^2 \mathbb{I}_L)$, \mathbb{I}_L denotes a $d \times d$ identity matrix, and σ_t^2 is the power of noise.

1.1.3 Global Model Averaging

To recover the update results of local models, $\mathbf{y}_{t,m}$ should be first decompressed by E . In this chapter, the minimum mean-square error (MMSE) criterion is employed, and the decompression result can be written as

$$\bar{\mathbf{w}}_{t,m} = \arg \min_{\mathbf{w}^q \in \mathcal{C}} \|\mathbf{D}_{t,m} \mathbf{y}_{t,m} - \mathbf{w}^q\|^2, \quad (1.7)$$

where $\mathbf{D}_{t,m}$ is a $d \times d$ decompression matrix of $\mathbf{y}_{t,m}$, \mathcal{C} denotes a set that consists of all the possible quantized parameter vectors, i.e., $\mathbf{w}^q \in \mathcal{C}$. Since each element of

the quantized model parameter vector can be detected individually, recalling the computational complexity of MMSE, the complexity of this detection is a linear function of the vector dimension.¹

Then the global model can be updated by averaging the decompressed results of local models. As introduced in Konecný et al. [2016], the update result of global model can be expressed as

$$\begin{aligned}\mathbf{w}_{t,F} &= \mathbf{w}_{t-1,F} + \sum_{m=1}^M \frac{N_m}{N} z_{t,m} (\bar{\mathbf{w}}_{t,m} - \mathbf{w}_{t-1,F}) \\ &= \sum_{m=1}^M \frac{N_m}{N} [(1 - z_{t,m})\mathbf{w}_{t-1,F} + z_{t,m}\bar{\mathbf{w}}_{t,m}],\end{aligned}\quad (1.8)$$

where $\mathbf{w}_{t,F}$ and $\mathbf{w}_{t-1,F}$ denote the global model for the t th and $(t - 1)$ -th iterations, respectively, $z_{t,m}$ is defined by (1.3), $N = \sum_{m=1}^M N_m$.

After global model averaging, $\mathbf{w}_{t,F}$ is sent back to the users. Since $\mathbf{w}_{t,F}$ are transmitted via downlink transmissions, which can acquire more radio resource and higher transmit power than the local model update phase. Therefore, it can be assumed that $\mathbf{w}_{t,F}$ is received successfully by all the users. Then the local model of each user can be updated as $\mathbf{w}_{t,m} = \mathbf{w}_{t,F}$, $m = 1, \dots, M$.

1.2 Problem Formulation

The performance of existing learning techniques is mainly determined by the accuracy of generated learning models. It is difficult to be modeled in a tractable form, and thus cannot be optimized by employing the existing resource management schemes in the MEC systems. In this section, we first derived a closed-form upper bound of model accuracy loss, which is an efficient metric to evaluate the quality of FL models. Then an optimization problem is formulated to improve the model accuracy and training efficiency of FL with limited budget of computation and communication resources.

1.2.1 Model Accuracy Loss

As introduced in Konecný et al. [2016], the objective of model training is to minimize the expected risk, which can be estimated by employing the the empirical risk given by (1.2). Therefore, the model accuracy loss, which can be defined as the

¹ To implement (1.7) at edge server, it requires to feedback sparsification matrix $\mathbf{A}_{t,m}$, and this communication overhead can be reduced significantly when $\mathbf{w}_{t,m} \in \mathbb{R}^d$ is divided into multiple segments. Moreover, the detection can be done with channel estimation, each device can feedback $\mathbf{D}_{t,m}$ instead of channel state information.

Index

a

access control 85, 111
 access point(s) (AP(s)) 1, 2, 85, 241, 242
 actor-critic (AC) 52
 additive white Gaussian noise 3
 aggregated model 32, 223
 alternating direction methods of
 multipliers (ADMM) 240
 artificial intelligence (AI) 49, 94

b

base stations (BS(s)) 85
 batch size(s) 27, 30, 32, 104, 223, 228,
 264, 283
 beam management 127–142, 145–148,
 150
 Beta distribution 58
 blockchain-based deep reinforcement FL
 (BDRFL) 156, 157
 Brute-force search (BFS) 145

c

centralized learning 17, 23, 25, 26, 42,
 45, 47, 135, 146, 215, 227, 258, 279
 centralized unit (CU) 88
 central limit theorem 27, 56
 channel fading 3, 32, 132, 133, 199,
 239–244, 250, 253

channel state information (CSI) 4, 21,
 129, 243, 257, 273
 communication error 6
 communication round 51–53, 57–65,
 70–74, 93, 95–101, 103–105,
 136–141, 181, 183–186, 189–191,
 194, 197–199, 202–208, 221,
 223, 232, 239, 240, 243, 247, 251,
 252
 compressive sensing (CS) 257, 273
 computational complexity 4, 141,
 145, 150, 171, 195, 196, 208,
 258
 computation and communication costs
 8
 consensus mechanism 157, 158, 166,
 170, 171
 convex/nonconvex problem(s) 10
 convolutional neural network (CNN)
 17, 198
 cumulative distribution function 57,
 219

d

data encryption 114
 data sharing 30–33, 35, 37, 39, 41, 45,
 46, 111, 114, 156, 176, 179
 deep reinforcement learning (DRL) 85,
 86, 88, 90, 94, 156

device association 85–94, 100, 103, 104, 109
 device-to-device (D2D) 155, 280
 distributed unit (DU) 88
 distribution divergence 24, 26–32, 42, 43, 45
 double deep Q-Network (DDQN) 95
 downlink transmissions 4
 dynamic programming 15
 dynamic scheduling algorithm (DISCO) 184, 192, 193, 195, 208

e

edge computing server 1
 edge network intelligence 49, 50
 empirical risk 2, 4, 5, 7, 17, 24, 25, 27, 237, 241, 248
 energy consumption 9, 10, 19–21, 32, 33, 45, 46, 47, 94, 30, 135, 182, 183, 185, 188, 191–193, 196, 201, 202
 expected mean square error (MSE) 5

f

federated averaging 24, 28, 29, 30, 32, 43, 44
 federated distillation 82, 119, 124
 federated edge learning (FEEL) 22
 federated learning (FL) 181, 208, 257, 258
 federated transfer learning (FTL) 86
 first coming and first scheduling (FCFS) 183
 fractional programming problem 13, 39
 frequency division duplex (FDD) 257

g

Gamma distribution 58
 generated learning model(s) 4
 generative adversarial network(s) (GAN(s)) 31, 115, 125
 global model updating 51, 52, 138

gradient descent (GD) 52, 68, 97, 99, 103, 119, 139–141, 145, 164, 165, 228, 253, 255, 262, 280
 graphic processing unit(s) (GPU(s)) 49

h

handoff management 85
 heterogeneous network(s) (HetNet(s)) 127, 215
 heuristic algorithm 157
 horizontally FL (hFL) 86, 98
 hybrid federated deep reinforcement learning (HDRL) 86, 94
 hybrid FL 86, 109

i

independent and identically distributed (IID) 182, 243
 industrial internet of things (IIoT) 181
 intelligent reflecting surface (IRS) 127
 Internet of Things (IoT) 113, 127, 181

k

Karush–Kuhn–Tucker (KKT) condition(s) 34
 knowledge distillation-based federated 114

l

Lagrangian function 35
 learning (KDFL) 114
 linear programming 41
 line-of-sight (LoS) 132
 local intrinsic dimensionality (LID) 216
 local model training 1, 2, 8, 9, 17, 23, 24, 32, 33, 56, 59, 95, 99, 115, 118–120, 136, 138, 160, 163, 166–168
 local model updating 51, 52, 98, 136, 138, 141
 long short-term memory (LSTM) 158

loss function 2, 21, 25, 51, 52, 60, 72, 95, 97, 98, 103, 116, 119, 138, 139, 141, 143, 145, 164, 185, 186, 189, 190, 192, 218, 220, 223, 241, 250, 260, 263, 275, 278

m

machine learning (ML) 49, 93, 113, 114, 117, 128, 156, 239, 240, 245, 250, 253, 280

macro base station (MBS) 130

management based on federated learning (BMFL) 129

Markov decision process (MDP) 157

massive machine-type communications (mMTC) 85

millimeter wave (mmWave) 127

minimum mean-square error (MMSE) 3

mmWave SBS (mSBS) 128

mobile edge computing (MEC) 1, 23, 96

mobility management function (AMF) 88

model accuracy loss 115

model compression loss 6

model sparsification 3

model training loss 6, 7, 17

multilayer perceptron (MLP) 17, 250

multiple choice multidimensional knapsack problem (MMKP) 92

multiple-input multiple-output (MIMO) 127

mutual information 8

n

network slice(s) (NS(s)) 85, 88

network slice selection function (NSSF) 88

new radio (NR) 181

non-deterministic polynomial (NP)-hard 92

Non-LoS (NLoS) 132

normal distribution 27, 56–58, 240

o

orthogonal channel 32

orthogonal frequency division multiplexing (OFDM) 54

orthogonal frequency division multiplexing access (OFDMA) 190

over-the-air federated learning (AirFL) 239

p

parameter quantization 3, 258–260, 263, 268

Poisson cluster process (PCP) 50

Poisson distribution 50, 54

polynomial function 40

power of noise 3, 32

practical Byzantine fault tolerance (PBFT) 157

probability density function (PDF) 24, 54, 226, 244, 245

proof of stake (PoS) 170

proof of work (PoW) 158

q

quadrature amplitude modulation (QAM) 240

quality of service (QoS) 85

r

radio access network(s) (RAN(s)) 85, 95

radio resource allocation 31

radio unit (RU) 88

Rayleigh fading 53, 199, 251

reference signal received power (RSRP) 89

reinforcement learning (RL) 85, 86, 94, 135, 156, 163, 185
 resource consumption 50, 52, 53, 62, 64, 66, 70, 71, 75
 Rician fading 53

S

sample selection loss 6, 9
 Shapley value(s) 87, 94, 96, 100–102
 signal-noise ratio (SNR) 50
 signal-to-interference-plus-noise ratio (SINR) 53, 132
 single network slice selection assistance information (S-NSSAI) 88
 small base station(s) (SBS(s)) 127
 stochastic gradient descent (SGD) 119, 164, 228, 280
 supervised multiclassification task(s) 8

t

training dataset(s) 2, 6, 8, 9, 17, 23–25, 29, 30, 42, 45, 135, 285–287
 transition probability 92, 93, 135, 136
 transmission success probability 50, 55, 67, 68, 73, 184
 transmit power 4, 19, 20, 32, 33, 35, 41, 53, 54, 57, 67, 104, 132, 133, 144, 150, 160, 165, 172, 187, 242, 243
 transmit power allocation 35, 41

U

ultradense mmWave network(s) (UDmmN(s)) 129
 ultradense network (UDN) 127
 ultrareliable and low-latency communications (URLLC) 85
 uniform quantization 3
 user equipment(s) (UE(s)) 49, 155, 258, 269, 273
 user plane function (UPF) 88
 user privacy 29, 31, 49, 113, 128
 user selection 9, 15, 17, 21, 31, 32, 39, 40, 42, 45

V

vertically FL (vFL) 86

W

wireless bandwidth 54, 55, 57, 89, 91–93, 10, 155
 wireless edge networks 49–52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74–76, 78, 80, 82

Z

Zipf distribution 174