Peixi LIU, Jiamo JIANG, Guangxu ZHU, Lei CHENG, Wei JIANG, Wu LUO, Ying DU, Zhiqin WANG, 2022. Training time minimization for federated edge learning with optimized gradient quantization and bandwidth allocation. *Frontiers of Information Technology & Electronic Engineering*, 23(8):1247-1263. https://doi.org/10.1631/FITEE.2100538

Training time minimization for federated edge learning with optimized gradient quantization and bandwidth allocation

Key words: Federated edge learning; Quantization optimization; Bandwith allocation; Training time minimization

Corresponding authors: Jiamo JIANG, Guangxu ZHU
E-mail: Jiamo JIANG, jiangjiamo@caict.ac.cn;
Guangxu ZHU, gxzhu@sribd.cn
ORCID: Jiamo JIANG, https://orcid.org/0000-0002-4986-7081;
Guangxu ZHU, https://orcid.org/0000-0001-9532-9201

Motivation

1. In edge networks, computation resources of the edge devices and wireless resources of the network are limited. Therefore, training an artificial intelligence (AI) model by federated edge learning (FEEL) is usually time-consuming and expensive.
2. To reduce the total training time, we should not only bring down the number of communication rounds, but also shorten the per-round latency.
3. With the goal of minimizing the total training time, we focus on how to balance the number of communication rounds and the per-round latency via joint quantization level and bandwidth allocation optimization.

Main idea

1. We consider quantized FEEL, whereby a stochastic quantization scheme is adopted for updated gradient compression.
2. We make a comprehensive analysis of the total training time by taking into account the communication time, computation time, and number of communication rounds, based on which the intrinsic trade-off between the number of communication rounds and the per-round latency is characterized.
3. A joint quantization and bandwidth allocation optimization problem is formulated and solved.

Method

1. We propose a joint data-and-model-driven fitting method to yield an accurate estimate of the number of required communication rounds.
2. We adopt the alternating optimization technique to decompose the original problem into two sub-problems, and each optimizes one of the two control variables with the other variable fixed.

Simulation results

- **Experiment I**: ℓ_2 regularized logistic regression with synthetic dataset

- Total training time vs. quantization level

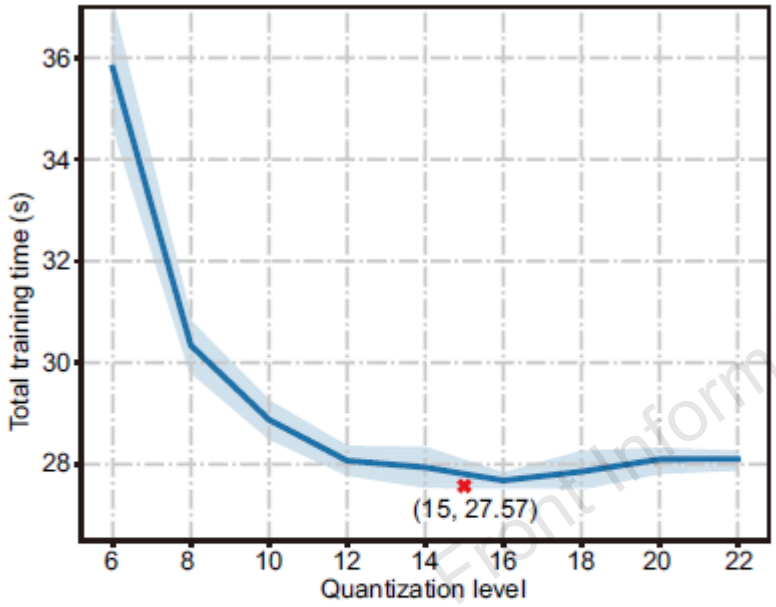


Fig. 4 Total training time vs. the quantization level q in simulation 1 when the bandwidth allocation is optimal (The point with the optimal quantization level and the corresponding training time from Algorithm 3 in theory is annotated by "x")

- Bandwidth allocation & CPU frequency

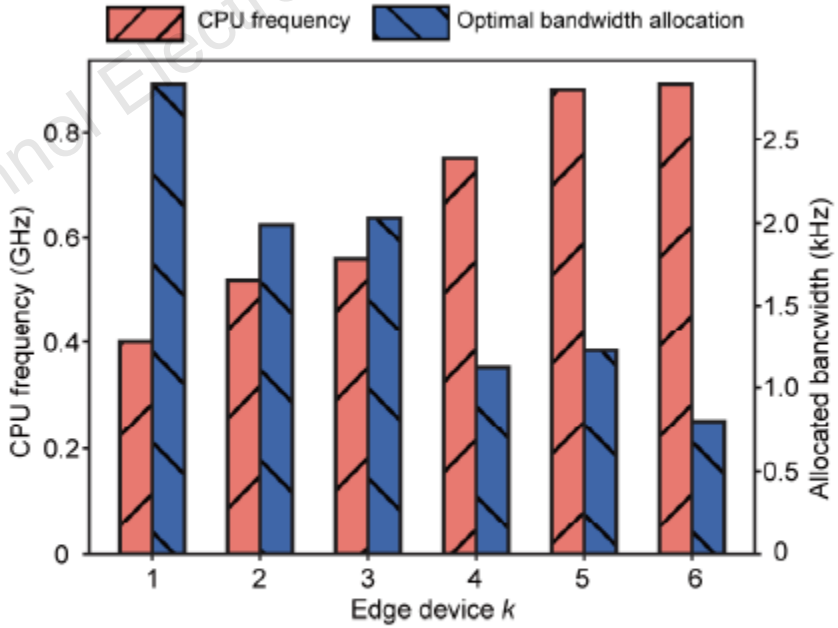


Fig. 5 Optimal bandwidth allocation (bars on the right) and CPU frequency (bars on the left) of each edge device in simulation 1

Simulation results

- **Experiment I**: ℓ_2 regularized logistic regression with synthetic dataset
  - Optimality gap & test accuracy vs. training time

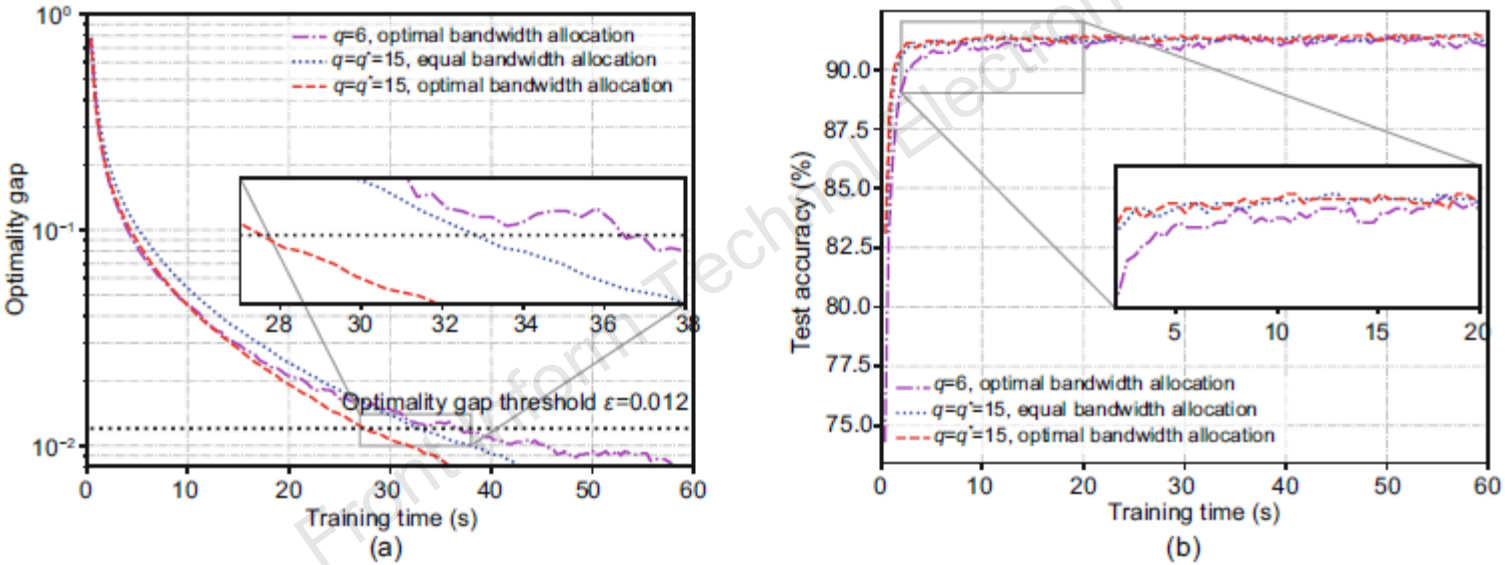


Fig. S1 Optimality gap (a) and test accuracy (b) in simulation 1

Major results

- **Experiment II**: CIFAR-10 dataset image classification with ResNet20

  - Simulation setup

Edge devices	Radius of network area	r	500m
	Number of devices	K	6
	Transmit power	p_k	1 dBm
	CPU frequency	f_k	[100MHz, 1GHz]
	Cycles for executing a batch of sample	$\mathcal{V}$	10^8
Wireless propagation	Path loss	$[PL]_{dB}$	$128.1 + 37.6\log_{10} l$ (l in Km)
	Shadow fading variance	σ_ζ^2	8 dB
	Noise power spectral density	N_0	-174 dBm/Hz
	Total bandwidth	B	10 KHz
Learning	Learning rate	η_n	$5/(n+10)$, $100/(n+1000)$
	Batch size	m_b	100
	Model size	d	1024
	Threshold of optimality gap	ϵ	0.015

Major results (Cont'd)

- **Experiment II**: CIFAR-10 dataset image classification with ResNet20
  - Total training time vs. quantization level

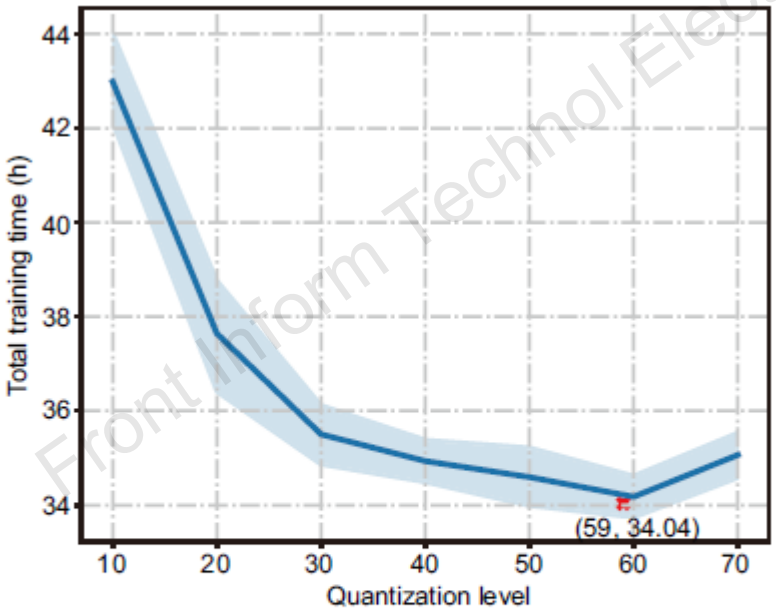


Fig. 6 Total training time vs. quantization level q in simulation 2 when the bandwidth allocation is optimal (The point with the optimal quantization level and the corresponding training time from Algorithm 3 in theory is annotated by "x")

Major results (Cont'd)

- **Experiment II**: CIFAR-10 dataset image classification with ResNet20
  - Optimality gap & test accuracy vs. training time

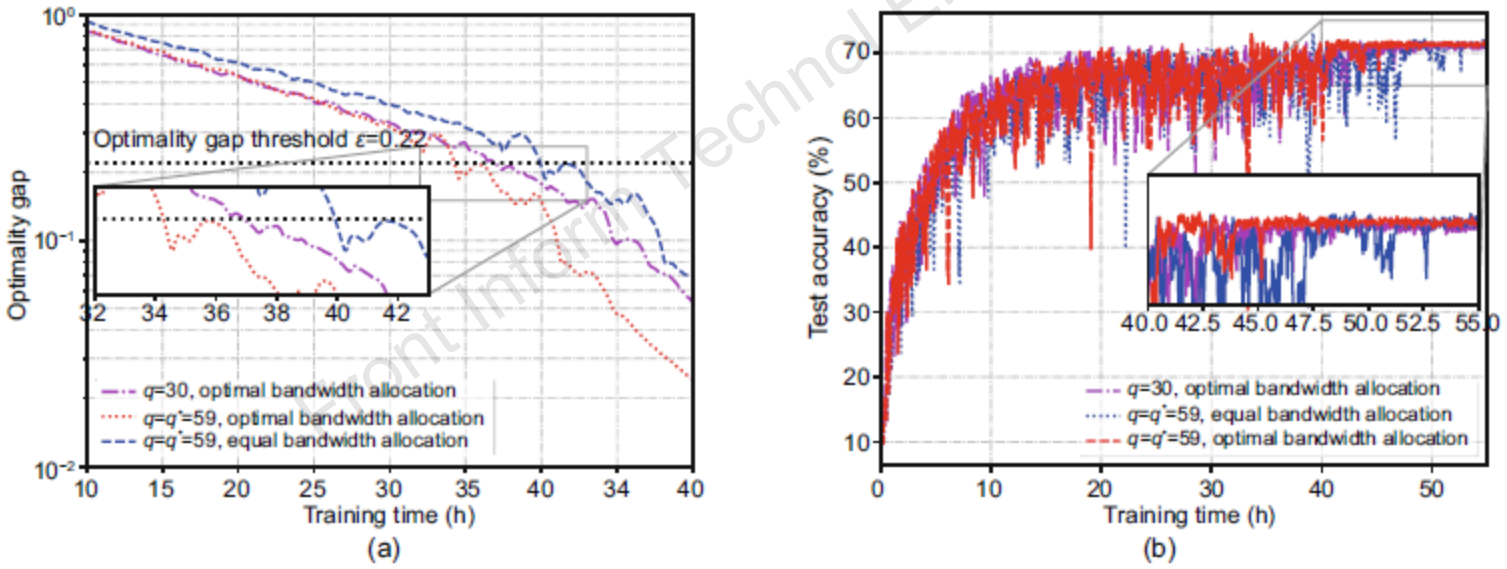


Fig. S2 Optimality gap (a) and test accuracy (b) in simulation 2

Conclusions

1. We studied the minimization of training time for quantized FEEL with optimized quantization level and bandwidth allocation.
2. The theoretical results developed can be used to guide system optimization and contribute to the understanding of how a wireless communication system can properly coordinate resources to accomplish learning tasks.

Peixi LIU received his BE and ME degrees from Northwestern Polytechnical University and his PhD degree from Peking University in 2022. His research interests include edge intelligence, distributed machine learning, next-G technologies such as integrated sensing, computation, and communication (ISCC).

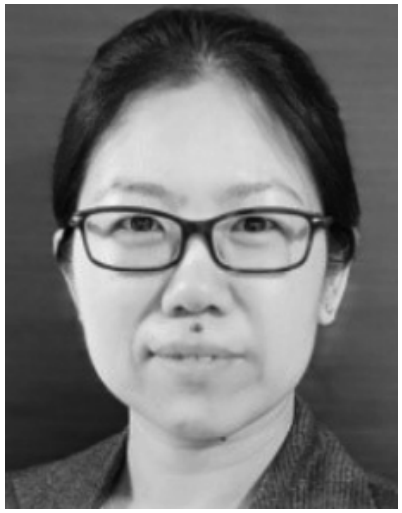
Jiamo JIANG received his BS degree in measurement and control technology and instrumentation and his PhD degree in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008 and 2014, respectively. He is currently the director engineer of the Mobile Communications Innovation Center (MCIC), China Academy of Information and Communications Technology (CAICT), Beijing. He has been engaged in research on technologies, standards, simulations, and experiments of 5G and 6G for seven years. His research interests include integrated sensing and communications, machining learning, and edge intelligence in wireless networks.

Guangxu ZHU received his BE and ME degrees from Zhejiang University and his PhD degree from The University of Hong Kong in 2019. He is currently a research scientist with Shenzhen Research Institute of Big Data. His research interests include edge intelligence, distributed machine learning, and 5G technologies such as massive MIMO, mmWave communication, and wirelessly powered communications.

Lei CHENG received his BE degree from Zhejiang University, Hangzhou, China, in 2013, and his PhD degree from The University of Hong Kong, Hong Kong, in 2018. He is currently an assistant professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His research interests include Bayesian machine learning for tensor data analytics and interpretable machine learning for information systems.

Wei JIANG received his BS and PhD degrees in electronics engineering from Peking University in 1997 and 2003, respectively. He is currently an associate professor with the Department of Electronics, Peking University. His research interests include communication algorithms, multiple antenna technologies, and satellite communications.

Wu LUO received his BS, MS, and PhD degrees in electronics engineering from Peking University, Beijing, China, in 1991, 1998, and 2006, respectively. He is currently a professor with the State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronics, Peking University. His research interests include wireless and satellite communication networks, digital signal processing, multiple-input-multiple-output networks, channel estimation, and channel coding techniques.



Ying DU is currently a professor-level senior engineer of the China Academy of Information and Communications Technology (CAICT). She is also the Vice Chair of the Standards and International Cooperation Working Group, IMT-2030 (6G) Promotion Group. She has contributed to the research, evaluation, and international standardization of IEEE 802.16, 4G, 5G, and 6G communication systems. She has authored or co-authored more than 30 research papers and over 40 patents in this area. She has hosted three national major projects on mobile broadband systems. Currently, she focuses on technology research, standardization, and implementation for 5G-advanced and 6G systems.

Zhiqin WANG is currently the Vice President with the China Academy of Information and Communications Technology (CAICT) and a professor-level senior engineer. She is the Chair of the Wireless Technical Committee of China Communications Standards Association (CCSA) and the Director of the Wireless and Mobile Technical Committee of China Institute of Communications (CIC). She is serving as the Chair for the IMT-2020 (5G) Promotion Group and the IMT-2030 (6G) Promotion Group in China. She has contributed to the design, standardization, and development of 3G (TD-SCDMA), 4G (TD-LTE), and 5G mobile communication systems. She has authored or co-authored more than 60 research papers and over 20 patents in this area. Her current research interests include the theoretical research, standardization, and industry development for 5G-advanced and 6G. Her achievements have received multiple top awards and honors, including the Grand Prize of the National Award for Scientific and Technological Progress in 2016 (the highest Prize in China), the First Prize of the National Award for Scientific and Technological Progress (once), the Second Prize of the National Award for Scientific and Technological Progress (twice), the National Innovation Award (once), and the Ministerial Science and Technology Awards (many times).