ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

Research on Federated Learning for Non-IID Data

Canming Cao [a], Zhengyi Chai [b]

Tianjin University of Technology, Tianjin 300387, China

[a] ccm09162022@163.com, [b] super_chai@126.com

Abstract

In recent years, federated learning has attracted widespread attention as a technology to solve the problem of data islands, and has begun to be applied in fields such as finance, healthcare, and smart cities. Introduce federated learning from three levels. First introduce the concept of federated learning, and explain the concept of federated learning through the definition, architecture, classification and comparison with traditional distributed learning; then from the perspective of machine learning and deep learning, various current federated learning algorithms are analyzed Classification comparison and in-depth analysis; finally, in-depth classification of federated learning optimization algorithms from the perspectives of communication cost, client selection, and aggregation mode optimization, summarizes the research status of federated learning, and puts forward the communication and system differences faced by federated learning. Data heterogeneity and data heterogeneity are the three major problems and solutions, as well as the outlook for the future.

Keywords

Communication Efficiency; Deep Learning; Federated Learning; Ternary Coding.

1. Introduction

With the vigorous development of big data-driven artificial intelligence (AI) technology [1-3], AI algorithms have shown far superior performance in image processing [4-6], speech recognition [7, 8] and other fields. The accuracy and efficiency of the method are widely used in various industries. However, in the process of the performance shown by AI algorithms being recognized by the public, the effective acquisition and maintenance of data has become a bottleneck restricting its development. In actual scenarios, except for a few enterprises and institutions, most fields have limited data and poor quality, which makes the performance of AI models unsatisfactory, and it is difficult to apply them in more industries. As a result, many researchers began to seek cross-industry, cross-organization, and cross-domain data distribution integration solutions [9-11], trying to promote model performance improvement on the basis of data sharing. It is worth noting that in general, the data processed by AI involves various industries, and there will be natural barriers between the data types of different industries, and it is almost impossible to integrate the hashed data, which requires a lot of financial resources and time cost. At the same time, due to inter-industry competition and users' concerns about privacy protection [12, 13], different industries have gradually increased their emphasis on data privacy and security, so data in various industries basically exists in the form of isolated islands. Therefore, how to effectively utilize isolated data and mine new machine learning paradigms to solve the problem of data fragmentation while meeting data privacy, security, and regulatory requirements is an important topic in the current AI development. With the prosperity and development of Internet of Things (IoT) [14], cloud computing [15] and other technologies, Federated Learning (Federated Learning, FL) [16, 17] can effectively break the data barrier, and thus gradually become a An emerging distributed machine learning paradigm has set off the trend and trend of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

Internet of Everything. In the FL scenario, the data of all parties is kept locally, without revealing privacy or violating the General Data Protection Regulation (GDPR) [18]; Establish a virtual global model, form a system of common interests, and collaboratively train and share data values instead of sharing original data. This not only protects the privacy of local data, but also solves the problems of poor model generalization ability and unsatisfactory model performance caused by insufficient local data volume and insufficient data types.

1.1 Federated Learning Taxonomy

The island data of federated learning have different distribution characteristics. For each participant, the data it owns can be represented by a matrix. Each row represents each user or an independent research object, and each column represents a feature of the user or research object. At the same time, each row of data will have a label. For each user, one hopes to learn a model to predict his label Y based on his features X. In reality, different parties may be different companies or institutions. People do not want their data X to be known to others, but people hope to jointly train a more powerful model to predict label Y.

According to the data characteristics of federated learning, federated learning can be divided into horizontal federated learning, vertical federated learning, and migration federated learning, and federated learning is classified according to the degree of data overlap between different participants.

When the overlapping parts of users of the two participants are small, but the overlapping parts of user characteristics of the two data sets are relatively large, the federated learning in this scenario is called horizontal federated learning. For example, the headquarters of a banking system in Shenzhen and Shanghai are participants, and the businesses of both sides are similar, so the characteristics of the collected user data are relatively similar, but most of the users of the two branches are local residents.There is relatively little user overlap. When two branches need to do a federated model to classify users, it belongs to horizontal federated learning.

When the overlapping parts of users of the two participants are large, but the overlapping parts of user characteristics of the two data sets are relatively small, the federated learning in this scenario is called vertical federated learning. For example, two institutions in the same area, one institution has the user's consumption records, and the other institution has the user's bank records. The two institutions have many overlapping users, but the data characteristics of the records are different. The two institutions want to aggregate through encryption Different features of users are used to jointly train a more powerful federated learning model. This type of machine learning model is called vertical federated learning.

When the overlapping parts of users of the two parties are small, the overlapping parts of user features of the two data sets are also relatively small, and some data still has missing labels, the federated learning in this scenario is called transfer federated learning. For example, two institutions in different regions, one institution has the user consumption records in the region, and the other institution has the bank records in the region, the two institutions have different users, and the data characteristics are also different. In this case, the joint The trained machine learning model is transfer federated learning.

Most of the current research is based on horizontal federated learning and vertical federated learning. There is not much research in the field of transfer learning. Therefore, this article will focus on the algorithm types of horizontal federated learning and vertical federated learning. In horizontal federated learning, there are many overlapping dimensions of data features. Alignment is performed according to the overlapping dimensions, and the part of the participant data with the same features but not identical users is taken out for joint training; in vertical federated learning, users overlap more, so matching is performed based on user ID. Take out the part of the participant data that has the same user but not exactly the same features for joint training.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

1.2 The Background of the Rise of Federated Learning

In the current wave of artificial intelligence and deep learning, there are two prominent and acute problems-the "data island" problem and the privacy security problem. What's more serious is that there was a certain degree of checks and balances between the two.

(1)The “data island” conundrum

Most practical application fields have problems of limited data volume and poor quality, and it is even more difficult to obtain labeled data sufficient to support artificial intelligence and machine learning models for training in some highly specialized subdivision fields; at the same time, in different data There are barriers between sources that are difficult to break, and it is difficult to share and exchange across domains, which leads to the current big data becoming more and more a general term for "data islands" in a certain sense.

(2)Privacy protection problem

Emphasis on data privacy and security has become a worldwide consensus and trend, represented by the General Data Protection Regulation (GDPR) promulgated by the European Union in May 2018. The regulations that restrict the processing process have increased the difficulty of data acquisition, sharing and exchange, and brought unprecedented challenges to the implementation of many artificial intelligence technologies and applications. Facing the above challenges, federated learning technology emerged as the times require, and it has become a new attempt to solve the "data island" and privacy security problems faced by traditional machine learning and artificial intelligence methods in the process of obtaining labeled data for implementation. Federated learning is a new fundamental technology in the field of artificial intelligence. Its foundation is to protect data privacy and meet legal and regulatory requirements. On this basis, it can perform efficient machine learning among multiple participants or computing nodes. In addition, federated learning provides a "closed-loop" learning mechanism, the effectiveness of which depends on the contributions of data providers to themselves and others, which helps to motivate more participants to join the entire data "federated" ecosystem. federated learning It has the following advantages and characteristics:

(1) The data is stored locally on the terminal device to avoid data leakage and meet the needs of user privacy protection and data security.

(2) All participants have an equal status, which can achieve fair cooperation; ensure that participants can exchange information and model parameters in an encrypted manner while maintaining independence, and can grow at the same time.

(3) The model modeling effect is similar to traditional deep learning algorithms, especially in the process of federated transfer learning, which can achieve maximum non-destructive training and avoid negative transfer of performance loss in transfer learning.

1.3 Features of Federated Learning Algorithms

Based on the above introduction to federated learning, the following characteristics of federated learning algorithms are summarized.

Support for non-IID data: This is a very important feature of federated learning algorithms. Federated learning algorithms must perform well on non-IID data. In the actual use of federated learning, the data quality and distribution of data holders are uncontrollable, and the data of data holders cannot be required to be independent and identically distributed. Therefore, federated learning algorithms need to support non-independent and identically distributed data.

Efficient communication: Federated learning algorithms need to consider the system heterogeneity of data holders, and improve communication efficiency and reduce communication loss without loss of accuracy or with little loss.

Fast convergence: In the process of joint modeling, it is first necessary to ensure the convergence of the model, and at the same time, it is necessary to increase the convergence speed. Security and privacy: Data privacy security is an important feature of federated learning, so security and privacy

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

are important features of federated gradient updates necessary requirements. Security and privacy can be carried out in the aggregation process through encryption and other methods, and can also be reflected in the process of stand-alone optimization.

Support complex users: complex users refer to a large number of users, and user data is uneven or offset. This is very possible in the practical application of federated learning, and the federated optimization algorithm needs to have good compatibility with this situation.

1.4 Classification of non-IID in Federated Learning Scenarios

In the actual federated learning environment, the data among the participants is highly heterogeneous and the data volume gap is large [19]. Therefore, the data between participants in different environments may be completely different, that is, non-IID data. Suppose the data sample is (x, y), where x is the input attribute or feature and y is the label. For non-IID data, it is assumed that the local data distribution of terminal k is p(x, y), that is, p(x, y) is different from that of other terminal devices. Literature [16] divides non-IID data into the following five types according to the common way that data deviates from the same distribution:

(1) Feature distribution skew—covariate shift: The same feature has different manifestations on different clients. For example, the same number may be written differently by different people (such as the width and inclination of the strokes are different).

(2) Label distribution skew—prior probability shift: the distribution of the same label depends on the client. Such as kangaroos are only in Australia or zoos; certain emoji are only used by a certain group of people and so on.

(3) Same label but different features (Same label, different features—concept drift): For different clients, the same label has different features. For example, there are great differences in buildings in different regions; the same clothing brand will have great differences at different times, and so on.

(4) Same features but different labels (Same features, different label—concept shift): Due to personal preference, the same feature vector in the training data can have different labels. Labels such as reflecting sentiment or predicting the next word vary by individual and region.

(5) Data skew: Different clients can store a large amount of different data.

The above five types of non-IID data are the results of the division of the literature according to the various scenarios that may involve deviation from the same distribution in the real-world federated learning dataset.

2. Research Status of Various Aspects Involved in Non-IID Data in Federated Learning

2.1 Performance Optimization

When the client data has highly skewed non-IID data, the performance of federated learning (such as: model accuracy, learning efficiency, etc.) will be significantly reduced. Aiming at the problem of efficiency decline caused by non-IID data, many domestic and foreign experts and scholars have made unremitting efforts.

Li et al. proposed a federated learning framework FedProx to solve the data heterogeneity problem in federated learning. This scheme believes that although the parameter update of the client is very small, it has a very important impact on the experiment. At the same time, considering the heterogeneity of data between terminal devices, it is proposed to add a proximal term to each local objective function to make the algorithm more effective for the local client. It is more robust to heterogeneity between endpoints. Experiments show that the proposed FedProx algorithm improves the performance of federated learning [19]. Karimireddy et al. proved that when the data is heterogeneous, there will be client drift phenomenon, that is, update offset, which will lead to instability and slow down the convergence speed. In this regard, an algorithm SCAFFOLD is proposed in the article to solve the above problems. The algorithm corrects for client drift in local

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

updates by using a control variable (variance reduction) [20]. In addition, many experts and scholars have conducted a lot of research and discussion. For example, Huang et al. proposed a method FedAMP to facilitate pairwise collaboration between clients with similar data [21]. Ruan et al. proposed a novel federated learning aggregation method that allows more flexible devices to participate in the convergence [22]. And Zhang et al. also proposed an algorithm CSFedAvg to select a model with a higher frequency to alleviate the problem of accuracy drop caused by non-IID data [23].

The training efficiency of federated learning is largely related to non-IID data. If we start with non-IID data, it will be of great help to improve the efficiency of federated learning.

2.2 Algorithm Optimization

The federated averaging algorithm FedAvg is one of the most widely used algorithms in federated learning. However, the federated averaging algorithm updates the global model by using weight updates, which only considers the size of the client data, and does not take into account the impact of the client data quality on the overall model. Therefore, proposing an algorithm that can be applied to different data qualities is a very meaningful and worth exploring direction.

For independent and identically distributed data sets, Wang et al. proposed an adaptive communication strategy by observing the relationship between the convergence speed and the lower limit of the error when the number of local updates changes. Through this adaptive strategy, fast convergence and low error can be achieved. Lower bound for small microcomputer systems[24]. McMahan et al. proposed to train the local model with weighted average according to the size of the client's local data set [25], but the weighted average of the model only for the size of the data set will ignore the influence of the factor of data quality.

Although experts and scholars at home and abroad have considered dealing with data heterogeneity from the algorithm, there are relatively few studies on this aspect. How to design an algorithm that can take heterogeneity into account and truly reflect the data situation of terminal devices is still an urgent problem to be solved.

2.3 Model Optimization

Existing research on non-IID data mainly involves performance optimization, communication cost, privacy protection, etc. It is not only data problems and algorithm problems that need to be solved, but also model optimization problems. In federated learning, since the training methods of parametric models and non-parametric models are inconsistent, non-IID data also has a great impact on the training performance of federated learning. In addition, due to the large number of participants in federated learning, the communication cost and communication link of the federated server are very limited. Therefore, how to efficiently train the model and ensure the robustness of the model is very important. In existing research, the general solution to a large number of models is model compression (sparseness). From the perspective of random sparsity, Shi et al. combined the training algorithm with local computing and gradient sparsity to propose a flexible sparsity method, that is, to provide error compensation for participants, allowing participants to upload only a small part of gradients with significant features, thereby reducing The communication load of each round [26]. In addition, in response to the imbalance of data distribution on mobile terminal devices, Duan et al. proposed a self-balancing framework Astraea. In this scheme, the global imbalance is alleviated through adaptive data amplification and downsampling, and at the same time, according to the client data The relative entropy of the distribution (Kullback–Leibler Divergence, KLD) creates edge layers to rearrange client training to alleviate model bias due to data imbalance [1]. Since Hinton et al. proposed the knowledge distillation method in 2015, Federated Distillation has also been proposed. Federated Distillation only exchanges local model outputs instead of exchanging traditional federated learning.

The model parameters of , since the size of these outputs is usually smaller than the model size, thus reducing the communication consumption. For example, a distributed deep reinforcement learning framework with high communication efficiency and privacy protection proposed in [27] - Federated

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

Reinforcement Distillation (FRD). Since in federated learning, the number of clients is huge, the federated learning environment may consist of millions of participants. Therefore, how to efficiently train the model and ensure the robustness of the model is very important. However, it is obvious that the existing FedAvg is not the optimal comprehensive solution. Many current studies have improved the existing model update and model aggregation. How to improve the robustness and effectiveness of the model through non-IID data is also a current research topic. A hot spot in federated learning. Compared with the existing methods, there is Experimental Study[6], but the current research is far from enough. For example, whether it can start from the perspective of model cleaning and filtering according to the literature [28], and then according to non-IID The data screens the differences of the model, such as filtering out the joint training of IID data and non-IID data The model with poor performance comes out [29].

2.4 Communication Cost

The communication problem of federated learning is mainly caused by connecting and transferring data (models, parameters) between the client and the central server via the network. Since a large number of user devices are sending local models, updates to the server, and mobile devices usually have limited data and slow network connections to the central server, reducing communication overhead is critical [30]. For the problem of communication overhead, many domestic and foreign scholars only train low-capacity models that occupy less communication space in the overall framework of federated learning at the expense of model accuracy. From this point of view, Rothchild et al. proposed FedSGD[31], which uses Count Sketch to compress client-side model updates, and FedPAQ[32], a cycle averaging and quantization method proposed by Reisizadeh et al., etc., but they all Little attention has been paid to the impact of non-IID data on communication costs. In this regard, some domestic and foreign scholars have started to reduce communication costs from the non-IID aspect. Among them, the scheme with a relatively large impact factor is the sparse ternary compression (STC) [7] proposed by Sattler et al. STC adopts the method described in [33] for federated learning settings, and uses the tok-k sparse algorithm as a starting point to construct an efficient communication protocol that compresses uplink and downlink communications through sparse, mutualization, and error correction. Solve three problems: quantify weight update and lossless encoding to further improve communication efficiency; design compressed communication for download; implement a cache mechanism to keep the client synchronous in the case of partial participation of the client. Itahara et al. proposed a distillation-based semi-supervised federated learning algorithm (DS-FL). In this approach, the communication cost only depends on the output dimension of the model instead of scaling according to the model size. Experimental results show that DS-FL reduces the communication cost by up to 99% compared to the FL baseline [34]. Chai et al. proposed FedAT, an asynchronous hierarchical federated learning method based on non-IID data. FedAT synergistically combines synchronous intra-layer training and asynchronous cross-layer training. Through layered bridging of synchronous and asynchronous training, the miniature microcomputer systematically reduces the straggling effect, improves the convergence speed and test accuracy, and uses a high-efficiency compression algorithm based on polyline coding to compress the uplink and downlink communication, thereby maximizing reduce communication costs. Experimental results show that compared with the current state-of-the-art FL method, the prediction performance of this method is improved by 21.09%, and the communication cost is reduced by 8.5 times [35].

To sum up, in terms of communication costs, most of them start from the model point of view, but there are also dimensions to reduce costs. The schemes proposed by scholars to reduce communication costs from non-IID all have high effects.

2.5 Privacy Protection

From traditional cloud computing and edge computing to today's federated learning, federated learning has shown strong advantages and development potential. Although federated learning largely guarantees the data privacy of terminal devices. However, some scholars have found that in a

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0006

federated learning environment, even if the terminal device only shares model information, a piece of private information will be leaked. Few of the current research on privacy protection for federated learning start with non-IID data, but there are a small number of scholars who want to use this as a breakthrough point to conduct related research. Yang et al. proposed a federated averaging algorithm for global model aggregation in order to ensure data privacy, solve data imbalance between different devices, and non-IID data. The algorithm is implemented by computing a weighted average of the local models on each selected device [36]. For the feature shift non-IID problem, Li et al. proposed to use local batch normalization before averaging the model to alleviate feature shift [37]. Xiong et al. made an innovative exploration of the privacy protection of federated learning from the aspect of non-IID data. First, the author conducts an in-depth analysis of the privacy leakage problem in FL, proves the upper bound of the performance of privacy reasoning attacks, and designs the 2DP-FL algorithm on this basis. The algorithm achieves differential privacy by adding noise when training the local model and the global model.

2.6 Personalized Federated Learning

In federated learning, due to the existence of non-IID data on the client side, statistical heterogeneity usually leads to different requirements for model performance on each client side. For some customers, the local model trained only based on private data may be better than the global model, so it is difficult to meet the needs of all participants with only a single global model. At this time, personalized research methods can be used to make the well-trained The global model is optimized for different users. Therefore, personalized federated learning is also a research method to solve non-IID data. Literature [38] proposed a scalable federated multi-task learning framework Ditto. Ditto can be viewed as a lightweight personalization add-on to the standard global FL, providing intrinsic fairness and robustness by solving the global model and the personalization model in an alternating manner. Literature [39] emphasizes the necessity of personalization of federated learning and summarizes the latest research on personalization of federated learning. Common approaches to personalize global models include augmenting user context, federated transfer learning, federated multi-task learning, federated meta-learning, federated knowledge distillation, mixing global and local models, and more. However, due to data heterogeneity, it is difficult for federated learning to train a single model suitable for all clients, and additional overhead will inevitably be added through personalized methods. Therefore, how to establish personalized federated learning for clients in a federated learning environment and reduce additional It is still a very promising research direction, and scholars can continue to conduct in-depth research based on the existing research of domestic and foreign scholars. The current state of research on various aspects of non-IID data in federated learning.

At present, many experts and scholars have noticed the advantages of personalized federated learning, and began to explore from this aspect to find a personalized model suitable for different participants. Personalized federated learning is an important direction to solve the non-IID data heterogeneity of federated learning. At present, there are relatively few research programs on the personalization of federated learning, but it is indeed a direction worthy of in-depth study.

3. Conclusion

With the advent of the era of big data, users are increasingly interested in protecting privacy and security, and traditional machine learning can no longer satisfy people's protection of privacy and security. The emergence of federated learning has broken the problem of data islands and ensured privacy security to a great extent, so it has been widely used. However, the existence of non-IID data not only leads to problems such as low efficiency and decreased accuracy of federated learning, but also brings new challenges to the privacy protection of federated learning, so the research on non-IID data is still a hot issue. This paper summarizes the above schemes and systematically analyzes related research. Aiming at the deficiencies of existing research programs, this paper proposes the future research direction of federated learning from the aspects of trust and incentive mechanism, user

association model and perception scenarios, etc., and provides reference for further research on non-IID problems in federated learning, thus providing a reference for related fields. The researchers provide investigation and assistance.

References

[1] TORREY L, SHAVLIK J. Transfer learning. [M]//Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. [S.l.]:IGI global, 2010: 242-264.

[2] LIU Y, KANG Y, XING C, et al. A secure federated transfer learning framework.[J]. IEEE Intelligent Systems, 2020, 35(4): 70-82.

[3] LIU H, ZHANG X, SHEN X, et al. A fair and efficient hybrid federated learning framework based on xgboost for distributed power prediction. [J]. arXiv preprintarXiv:2201.02783, 2022.

[4] ZHANG S, ZHOU J, HE X. Learning implicit temporal alignment for few-shot video classification [M/OL]. arXiv, 2021. https://arxiv.org/abs/2105.04823. DOI:10.48550/ARXIV.2105.04823.

[5] SU N M, CRANDALL D J. The affective growth of computer vision [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2021: 9291-9300.

[6] PESAVENTO M, VOLINO M, HILTON A. Attention-based multi-reference learning for image super-resolution [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). [S.l.: s.n.], 2021: 14697-14706.

[7] JIANG Z, REN Y, LEI M, et al. FedSpeech: Federated text-to-speech with continual learning [C/OL]// Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2021: 3829-3835. https://doi.org/ 10. 24963/ ijcai. 2021/ 527.107.

[8] XU J, ZHOU H, GAN C, et al. Vocabulary learning via optimal transport for neural machine translation [M/OL]. arXiv, 2020. https://arxiv.org/abs/2012.15671. DOI:10.48550/ARXIV.2012.15671.

[9] YANG S, REN B, ZHOU X, et al. Parallel distributed logistic regression for vertical federated learning without third-party coordinator. [J]. arXiv preprint arXiv:1911.09824, 2019.

[10] TIAN Z, ZHANG R, HOU X, et al. Federboost: Private federated learning for gbdt.[J]. arXiv preprint arXiv:2011.02796, 2020.

[11] GU B, DANG Z, LI X, et al. Federated doubly stochastic kernel learning for vertically partitioned data. [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. [S.l.]: ACM, 2020: 2483-2493.

[12] WU Y, CAI S, XIAO X, et al. Privacy preserving vertical federated learning for tree-based models. [J]. arXiv preprint arXiv:2008.06170, 2020.

[13] GU B, XU A, HUO Z, et al. Privacy-preserving asynchronous federated learning algorithms for multi-party vertically collaborative learning. [J]. arXiv preprint arXiv:2008.06233, 2020.

[14] ATZORI L, IERA A, MORABITO G. The internet of things: A survey [J/OL].Computer Networks, 2010, 54(15): 2787-2805.

[15] SADIKU M N, MUSA S M, MOMOH O D. Cloud computing: Opportunities and challenges [J/OL]. IEEE Potentials, 2014, 33(1): 34-36. DOI: 10.1109/MPOT.2013.2279684.

[16] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical secure aggregation forprivacy-preserving machine learning. [C]//proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. [S.l.]: ACM, 2017: 1175-1191.

[17] HARD A, RAO K, MATHEWS R, et al. Federated learning for mobile keyboard prediction. [J]. arXiv preprint arXiv:1811.03604, 2018.

[18] REGULATION P. General data protection regulation. [J]. Intouch, 2018, 25.XU J, ZHOU H, GAN C, et al. Vocabulary learning via optimal transport for neural machine translation [M/OL]. arXiv, 2020. https:// arxiv. org/abs/2012.15671. DOI:10.48550/ARXIV.2012.15671.

[19] Kairouz P, McMahan H B, Avent B, et al. Advances and open problems in federated learning [J]. Foundations and Trends® in Machine Learning,2021,14(1–2):1-210.

[20] Li Tian, Sahu Anit Kumar, Zaheer Manzil, et al. Federated optimization in heterogeneous networks [J]. Proceedings of Machine Learning and Systems, 2020, 2:429-450.

[21] Karimireddy Sai Praneeth, Kale Satyen, Mohri Mehryar, et al. Scaffold: stochastic controlled averaging for federated learning[C]//Proceedings of the 37th International Conference on Machine Learning (PMLR), 2020: 5132-5143.

[22] Huang Yu-tao,Chu Ling-yang,Zhou Zi-rui,et al. Personalized cross-silo federated learning on non-iid data [C]//Proceedings of the AAAI Conference on Artificial Intelligence,2021:7865-7873.

[23] Ruan Yi-chen, Zhang Xiao-xi, Liang Shu-che, et al. Towards flexible device participation in federated learning [C]// Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (PMLR), 2021:3403-3411.

[24] Zhang Chen, Xie Yu, Bai Hang, et al. A survey on federated learning[J].Knowledge-Based Systems, 2021, 216 (1):106775.

[25] Wang Jian-yu, Joshi Gauri. Adaptive communication strategies to achieve the best error-runtime trade-off in local-update SGD[J]. Proceedings of Machine Learning and Systems,2019,1:212-229.

[26] McMahan H Brendan, Moore Eider, Ramage Daniel, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (PMLR),2017:1273-1282.

[27] Shi Dian, Li Liang, Chen Rui, et al. Towards energy efficient federated learning over 5G+ mobile devices [J]. IEEE Signal Processing Magazine,2020,37(3):50-60.

[28] Cha Han, Park Jihong, Kim Hyesung, et al. Proxy experience replay: federated distillation for distributed reinforcement learning[J].IEEE Intelligent Systems,2020,35(4):94-101.

[29] Liu Yan, Wang Tian, Peng Shao-liang, et al. Edge-based model cleaning and device clustering in federated learning [J]. Chinese Journal of Computers,2021,44(12):2517-2530.

[30] Sun Bing, Liu Yan, Wang Tian, et al. Review of federated learning efficiency optimization in mobile edge networks [EB/OL].http://kns.cnki.net/kcms/detail/11.1777.TP.20211104.2019.024.html.2022-4-16.

[31] Wang Lu-ping,Wang Wei,Li Bo. CMFL: mitigating communication overhead for federated learning [C]// IEEE 39th International Conference on Distributed Computing Systems (ICDCS),2019:954-964.

[32] Rothchild Daniel,Panda Ashwinee,Ullah Enayat,et al.Fetchsgd:communication-efficient federated learning with sketching[C]//Proceedings of the 37th International Conference on Machine Learning, PMLR, 2020:8253-8265.

[33] Reisizadeh Amirhossein, Mokhtari Aryan, Hassani Hamed, et al. Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization[C]//Proceedings of the 23th International Conference on Artificial Intelligence and Statistics, PMLR,2020:2021-2031.

[34] Sattler Felix, Wiedemann Simon, Müller KlausRobert, et al. Sparse binary compression: Towards distributed deep learning with minimal communication[C]//2019 IEEE International Joint Conference on Neural Networks (IJCNN). 2019:1-8.

[35] Itahara Sohei,Nishio Takayuki,Koda Yusuke,et al. Distillation-Based Semi-Supervised Federated Learning for CommunicationEfficient Collaborative Training with Non-IID Private Data[J]. IEEE Transactions on Mobile Computing,2021,PP(99):1-1.

[36] Chai Zheng,Chen Yujing,Zhao Liang,et al. Fedat:A communication-efficient federated learning method with asynchronous tiers under non-IID data[EB/OL].https://par.nsf.gov/servlets/purl/10213680,2020.

[37] Yang Kai, Jiang Tao, Shi Yuanming, et al. Federated learning via over-the-air computation[J]. IEEE Transactions on Wireless Communications,2020,19(3):2022-2035.

[38] Li Xiao-xiao, Jiang Mei-rui, Zhang Xiao-fei, et al. FedBN: federated learning on Non-IID features via local batch normalization [EB/OL]. https://arxiv.org/abs/2102.07623,2021-2-15.

[39] Kulkarni Viraj, Kulkarni Miliand, Pant Aniruddha. Survey of personalization techniques for federated learning [C]// 4th World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020: 794-797.