## IEEE Transaction:

# Clustered Federated Learning: Model-Agnostic Distributed Multi-Task Optimization under Privacy Constraints

## Clustered federated learning is a model-agnostic distributed multitask optimization technique that addresses the challenges of unbalanced and non-i.i.d. data distribution among clients. It leverages a similarity metric to group clients with similar data distributions and provides specialized models for each group. To improve convergence speed and capture all data distributions, a two-phased client selection and scheduling approach is introduced. This approach ensures correct clustering and fairness between clients by leveraging bandwidth reuse and exploiting device heterogeneity. The server performs clustering based on predetermined thresholds and stopping criteria, employing a greedy selection for clusters that approximate a stopping point. Extensive simulations demonstrate that the proposed algorithms reduce training time and improve convergence speed by up to 50%, while equipping every user with a customized model tailored to its data distribution [1] [2] [3].

## <https://github.com/felisat/clustered-federated-learning>

## FedGroup: Efficient Federated Learning via Decomposed Similarity-Based Clustering

F. Sattler, K.-R. M ¨uller, and W. Samek, “Clustered federated learning:

Model-agnostic distributed multitask optimization under privacy constraints,” IEEE Transactions on Neural Networks and Learning Systems

(TNNLS), pp. 1–13, 2020.

[14] F. Sattler, K.-R. M ¨uller, T. Wiegand, and W. Samek, “On the byzantine

robustness of clustered federated learning,” in Proceedings of the IEEE

International Conference on Acoustics, Speech and Signal Processing

(ICASSP). IEEE, 2020, pp. 8861–8865.

[15] M. Xie, G. Long, T. Shen, T. Zhou, X. Wang, and J. Jiang, “Multi-center

federated learning,” arXiv preprint arXiv:2005.01026, 2020.

[16] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, “An efficient

framework for clustered federated learning,” in Advances in Neural

Information Processing Systems, vol. 33. Curran Associates, Inc., 2020,

pp. 19 586–19 597.

[17] C. Briggs, Z. Fan, and P. Andras, “Federated learning with hierarchical

clustering of local updates to improve training on non-IID data,” in

Proceedings of the IEEE International Joint Conference on Neural

Networks (IJCNN), 2020, pp. 1–9.

[18] Y. Zhang, M. Duan, D. Liu, L. Li, A. Ren, X. Chen, Y. Tan, and

C. Wang, “CSAFL: A clustered semi-asynchronous federated learning

framework,” arXiv preprint arXiv:2104.08184, 2021.

[19] S. Sarkar and A. K. Ghosh, “On perfect clustering of high dimension,

low sample size data,” IEEE transactions on pattern analysis and

machine intelligence (TPAMI), vol. 42, no. 9, pp. 2257–2272, 2019.

Recently, Sattler et al. [13], [14] propose a novel federated multi-task learning framework Cluster Federated Learning (CFL), which exploits geometric properties of FL loss surface to cluster the learning processes of clients based on their optimization direction, provides a new way of thinking about the statistical heterogeneity challenge. Many researchers follow up on CFL-based framework [15]–[18] and confirm CFL is more accurate than traditional FL with the consensus global model. However, above CFL-based frameworks are inefficient in the large-scale federated training systems or ignore the presence of newcomer devices. In this paper, we propose an efficient and accurate clustered federated learning framework FedGroup, which clusters clients into multiple groups based on a new decomposed data-driven measure between their parameter updates. In each communication round, each active client only contributes its local optimization result to the corresponding group model. The framework still maintains an auxiliary server to address the cold start issues of new devices. To improve the performance of high-dimension low-sample size (HDLSS) parameter updates clustering, we use a novel data-driven measure of cosine dissimilarity called Euclidean distance of Decomposed Cosine similarity (EDC), which can also avoid the concentration phenomenon of ℓp distances in high dimensional data clustering [19]. Furthermore, by combining FedGroup with the federated optimizer FedProx [20], FedGroup can be revised as FedGrouProx, which be explored in our experiments. With the above methods, FedGroup can significantly improve test accuracy by +6.9% on MNIST [21], +26.9% on FEMNIST [22], +5.3% on Sentiment140 [23] compared to FedSEM. We show that FedGroup has superior performance than FedProx and FeSEM [15]. Although FedGroup achieves performance improvements similar to IFCA [16], the latter has more communication and time overhead. The ablation studies of FedGroup are provided to demonstrate the usefulness of our clustering and cold start strategies.

* **A Systematic Literature Review on Federated Learning: From A Model Quality Perspective**

|  |  |
| --- | --- |
| **Item Type** | Preprint |
| **Author** | Yi Liu |
| **Author** | Li Zhang |
| **Author** | Ning Ge |
| **Author** | Guanghao Li |
| **Abstract** | As an emerging technique, Federated Learning (FL) can jointly train a global model with the data remaining locally, which effectively solves the problem of data privacy protection through the encryption mechanism. The clients train their local model, and the server aggregates models until convergence. In this process, the server uses an incentive mechanism to encourage clients to contribute high-quality and large-volume data to improve the global model. Although some works have applied FL to the Internet of Things (IoT), medicine, manufacturing, etc., the application of FL is still in its infancy, and many related issues need to be solved. Improving the quality of FL models is one of the current research hotspots and challenging tasks. This paper systematically reviews and objectively analyzes the approaches to improving the quality of FL models. We are also interested in the research and application trends of FL and the effect comparison between FL and non-FL because the practitioners usually worry that achieving privacy protection needs compromising learning quality. We use a systematic review method to analyze 147 latest articles related to FL. This review provides useful information and insights to both academia and practitioners from the industry. We investigate research questions about academic research and industrial application trends of FL, essential factors affecting the quality of FL models, and compare FL and non-FL algorithms in terms of learning quality. Based on our review's conclusion, we give some suggestions for improving the FL model quality. Finally, we propose an FL application framework for practitioners. |
| **Date** | 2020-12-01 |
| **Short Title** | A Systematic Literature Review on Federated Learning |
| **Library Catalog** | arXiv.org |
| **URL** | <http://arxiv.org/abs/2012.01973> |
| **Accessed** | 12/31/2023, 9:31:50 PM |
| **Extra** | arXiv:2012.01973 [cs] |
| **DOI** | [10.48550/arXiv.2012.01973](http://doi.org/10.48550/arXiv.2012.01973) |
| **Repository** | arXiv |
| **Archive ID** | arXiv:2012.01973 |
| **Date Added** | 12/31/2023, 9:31:50 PM |
| **Modified** | 12/31/2023, 9:31:50 PM |

* **Tags:**
  + Computer Science - Machine Learning
  + Computer Science - Cryptography and Security

**Attachments**

* + arXiv Fulltext PDF
  + arXiv.org Snapshot
* **A Survey on Federated Learning: The Journey From Centralized to Distributed On-Site Learning and Beyond**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Sawsan Abdulrahman |
| **Author** | Hanine Tout |
| **Author** | Hakima Ould-Slimane |
| **Author** | Azzam Mourad |
| **Author** | Chamseddine Talhi |
| **Author** | Mohsen Guizani |
| **Abstract** | Driven by privacy concerns and the visions of deep learning, the last four years have witnessed a paradigm shift in the applicability mechanism of machine learning (ML). An emerging model, called federated learning (FL), is rising above both centralized systems and on-site analysis, to be a new fashioned design for ML implementation. It is a privacy-preserving decentralized approach, which keeps raw data on devices and involves local ML training while eliminating data communication overhead. A federation of the learned and shared models is then performed on a central server to aggregate and share the built knowledge among participants. This article starts by examining and comparing different ML-based deployment architectures, followed by in-depth and in-breadth investigation on FL. Compared to the existing reviews in the field, we provide in this survey a new classification of FL topics and research fields based on thorough analysis of the main technical challenges and current related work. In this context, we elaborate comprehensive taxonomies covering various challenging aspects, contributions, and trends in the literature, including core system models and designs, application areas, privacy and security, and resource management. Furthermore, we discuss important challenges and open research directions toward more robust FL systems. |
| **Date** | 2021-04 |
| **Short Title** | A Survey on Federated Learning |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/9220780> |
| **Accessed** | 12/31/2023, 9:26:13 PM |
| **Extra** | Conference Name: IEEE Internet of Things Journal |
| **Volume** | 8 |
| **Pages** | 5476-5497 |
| **Publication** | IEEE Internet of Things Journal |
| **DOI** | [10.1109/JIOT.2020.3030072](http://doi.org/10.1109/JIOT.2020.3030072) |
| **Issue** | 7 |
| **ISSN** | 2327-4662 |
| **Date Added** | 12/31/2023, 9:26:13 PM |
| **Modified** | 12/31/2023, 9:26:13 PM |

* **DQRE-SCnet: A novel hybrid approach for selecting users in Federated Learning with Deep-Q-Reinforcement Learning based on Spectral Clustering**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Mohsen Ahmadi |
| **Author** | Ali Taghavirashidizadeh |
| **Author** | Danial Javaheri |
| **Author** | Armin Masoumian |
| **Author** | Saeid Jafarzadeh Ghoushchi |
| **Author** | Yaghoub Pourasad |
| **Abstract** | Machine learning models based on sensitive data in the real-world promise advances in areas ranging from medical screening to disease outbreaks, agriculture, industry, defense science, and more. In many applications, learning participant communication rounds benefit from collecting their own private data sets, teaching detailed machine learning models on the real data, and sharing the benefits of using these models. Due to existing privacy and security concerns, most people avoid sensitive data sharing for training. Without each user demonstrating their local data to a central server, Federated Learning allows various parties to train a machine learning algorithm on their shared data jointly. This method of collective privacy learning results in the expense of important communication during training. Most large-scale machine learning applications require decentralized learning based on data sets generated on various devices and places. Such datasets represent an essential obstacle to decentralized learning, as their diverse contexts contribute to significant differences in the delivery of data across devices and locations. Researchers have proposed several ways to achieve data privacy in Federated Learning systems. However, there are still challenges with homogeneous local data. This research’s approach is to select nodes (users) to share their data in Federated Learning for independent data-based equilibrium to improve accuracy, reduce training time, and increase convergence. Therefore, this research presents a combined Deep-Q-Reinforcement Learning Ensemble based on Spectral Clustering called DQRE-SCnet to choose a subset of devices in each communication round. Based on the results, it has been displayed that it is possible to decrease the number of communication rounds needed in Federated Learning. The realized reduction in the communication rounds are 51%, 25%, and 44% on the three datasets MNIST, Fashion MNIST, and CIFAR-10, respectively. |
| **Date** | 2022-10-01 |
| **Short Title** | DQRE-SCnet |
| **Library Catalog** | ScienceDirect |
| **URL** | <https://www.sciencedirect.com/science/article/pii/S1319157821002226> |
| **Accessed** | 12/31/2023, 9:29:15 PM |
| **Volume** | 34 |
| **Pages** | 7445-7458 |
| **Publication** | Journal of King Saud University - Computer and Information Sciences |
| **DOI** | [10.1016/j.jksuci.2021.08.019](http://doi.org/10.1016/j.jksuci.2021.08.019) |
| **Issue** | 9 |
| **Journal Abbr** | Journal of King Saud University - Computer and Information Sciences |
| **ISSN** | 1319-1578 |
| **Date Added** | 12/31/2023, 9:29:16 PM |
| **Modified** | 12/31/2023, 9:29:16 PM |

* **Tags:**
  + Deep learning
  + Ensemble model
  + Federated Learning
  + Independent heterogeneous data
  + Reinforcement Learning
  + Spectral Clustering

**Attachments**

* + Submitted Version
* **FedCO: Communication-Efficient Federated Learning via Clustering Optimization**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Ahmed A. Al-Saedi |
| **Author** | Veselka Boeva |
| **Author** | Emiliano Casalicchio |
| **Abstract** | Federated Learning (FL) provides a promising solution for preserving privacy in learning shared models on distributed devices without sharing local data on a central server. However, most existing work shows that FL incurs high communication costs. To address this challenge, we propose a clustering-based federated solution, entitled Federated Learning via Clustering Optimization (FedCO), which optimizes model aggregation and reduces communication costs. In order to reduce the communication costs, we first divide the participating workers into groups based on the similarity of their model parameters and then select only one representative, the best performing worker, from each group to communicate with the central server. Then, in each successive round, we apply the Silhouette validation technique to check whether each representative is still made tight with its current cluster. If not, the representative is either moved into a more appropriate cluster or forms a cluster singleton. Finally, we use split optimization to update and improve the whole clustering solution. The updated clustering is used to select new cluster representatives. In that way, the proposed FedCO approach updates clusters by repeatedly evaluating and splitting clusters if doing so is necessary to improve the workers’ partitioning. The potential of the proposed method is demonstrated on publicly available datasets and LEAF datasets under the IID and Non-IID data distribution settings. The experimental results indicate that our proposed FedCO approach is superior to the state-of-the-art FL approaches, i.e., FedAvg, FedProx, and CMFL, in reducing communication costs and achieving a better accuracy in both the IID and Non-IID cases. |
| **Date** | 2022/12 |
| **Language** | en |
| **Short Title** | FedCO |
| **Library Catalog** | www.mdpi.com |
| **URL** | <https://www.mdpi.com/1999-5903/14/12/377> |
| **Accessed** | 12/31/2023, 9:20:07 PM |
| **Rights** | http://creativecommons.org/licenses/by/3.0/ |
| **Extra** | Number: 12 Publisher: Multidisciplinary Digital Publishing Institute |
| **Volume** | 14 |
| **Pages** | 377 |
| **Publication** | Future Internet |
| **DOI** | [10.3390/fi14120377](http://doi.org/10.3390/fi14120377) |
| **Issue** | 12 |
| **ISSN** | 1999-5903 |
| **Date Added** | 12/31/2023, 9:20:07 PM |
| **Modified** | 12/31/2023, 9:20:07 PM |

* **Tags:**
  + convolutional neural network
  + federated learning
  + clustering
  + communication efficiency
  + Internet of Things

**Attachments**

* + Full Text PDF
* **FedCML: Federated Clustering Mutual Learning with non-IID Data**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Zekai Chen |
| **Author** | Fuyi Wang |
| **Author** | Shengxing Yu |
| **Author** | Ximeng Liu |
| **Author** | Zhiwei Zheng |
| **Editor** | José Cano |
| **Editor** | Marios D. Dikaiakos |
| **Editor** | George A. Papadopoulos |
| **Editor** | Miquel Pericàs |
| **Editor** | Rizos Sakellariou |
| **Abstract** | Federated learning (FL) enables multiple clients to collaboratively train deep learning models under the supervision of a centralized aggregator. Communicating or collecting the local private datasets from multiple edge clients is unauthorized and more vulnerable to training heterogeneity data threats. Despite the fact that numerous studies have been presented to solve this issue, we discover that deep learning models fail to attain good performance in specific tasks or scenarios. In this paper, we revisit the challenge and propose an efficient federated clustering mutual learning framework (FedCML) with an semi-supervised strategy that can avoid the need for the specific empirical parameter to be restricted. We conduct extensive experimental evaluations on two benchmark datasets, and thoroughly compare them to state-of-the-art studies. The results demonstrate the promising performance from FedCML, the accuracy of MNIST and CIFAR10 can be improved by $$0.53\%$$0.53%and $$1.58\%$$1.58%for non-IID to the utmost extent while ensuring optimal bandwidth efficiency ($$4.69\times $$4.69×and $$4.73\times $$4.73×less than FedAvg/FeSem for the two datasets). |
| **Date** | 2023 |
| **Language** | en |
| **Short Title** | FedCML |
| **Library Catalog** | Springer Link |
| **Place** | Cham |
| **Publisher** | Springer Nature Switzerland |
| **ISBN** | 978-3-031-39698-4 |
| **Pages** | 623-636 |
| **Series** | Lecture Notes in Computer Science |
| **Proceedings Title** | Euro-Par 2023: Parallel Processing |
| **DOI** | [10.1007/978-3-031-39698-4\_42](http://doi.org/10.1007/978-3-031-39698-4_42) |
| **Date Added** | 12/31/2023, 9:35:34 PM |
| **Modified** | 12/31/2023, 9:35:34 PM |

* **Tags:**
  + Cosine similarity
  + Distributed computing
  + Federate learning
  + Inter-clustering learning
  + non-IID data
* **A dynamic adaptive iterative clustered federated learning scheme**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Run Du |
| **Author** | Shuo Xu |
| **Author** | Rui Zhang |
| **Author** | Lijuan Xu |
| **Author** | Hui Xia |
| **Abstract** | Clustered federated learning (CFL), as an important research branch of personalized federated learning (FL), can better cope with the highly statistically heterogeneous federated learning environment and provide higher quality services to clients. However, existing CFL schemes have difficulties in adapting to real-time data distribution changes due to disadvantages such as relatively fixed cluster structure. This poses a great challenge to the practical deployment of CFL schemes. To address the common problems of existing CFL schemes, we propose a more flexible dynamic adaptive cluster federated learning scheme (AICFL). AICFL uses the mutual sensitivity between models and data as intuition to perform cluster identity estimation, cluster addition, cluster model updating, and cluster deletion in the early iterations of FL to find the optimal client cluster partitioning. Firstly, this process does not require a priori estimation of the number of clusters and does not require the online participation of all clients. Secondly, during cluster partitioning, AICFL is able to adjust the cluster structure in real time based on the overall data distribution. Moreover, AICFL has the same ability to adapt to changes in the system environment in the middle and late stages of FL. The experimental results show that our scheme gives the most reasonable cluster partitioning results in all cases which indicates that AICFL is able to cope with the above-mentioned distribution changes well, and has better adaptability and better flexibility than other schemes. |
| **Date** | 2023-09-27 |
| **Library Catalog** | ScienceDirect |
| **URL** | <https://www.sciencedirect.com/science/article/pii/S0950705123004914> |
| **Accessed** | 12/31/2023, 9:34:13 PM |
| **Volume** | 276 |
| **Pages** | 110741 |
| **Publication** | Knowledge-Based Systems |
| **DOI** | [10.1016/j.knosys.2023.110741](http://doi.org/10.1016/j.knosys.2023.110741) |
| **Journal Abbr** | Knowledge-Based Systems |
| **ISSN** | 0950-7051 |
| **Date Added** | 12/31/2023, 9:34:13 PM |
| **Modified** | 12/31/2023, 9:34:13 PM |

* **Tags:**
  + Distributed machine learning
  + Federated learning
* **FedGroup: Efficient Clustered Federated Learning via Decomposed Data-Driven Measure**

|  |  |
| --- | --- |
| **Item Type** | Preprint |
| **Author** | Moming Duan |
| **Author** | Duo Liu |
| **Author** | Xinyuan Ji |
| **Author** | Renping Liu |
| **Author** | Liang Liang |
| **Author** | Xianzhang Chen |
| **Author** | Yujuan Tan |
| **Abstract** | Federated Learning (FL) enables the multiple participating devices to collaboratively contribute to a global neural network model while keeping the training data locally. Unlike the centralized training setting, the non-IID and imbalanced (statistical heterogeneity) training data of FL is distributed in the federated network, which will increase the divergences between the local models and global model, further degrading performance. In this paper, we propose a novel clustered federated learning (CFL) framework FedGroup, in which we 1) group the training of clients based on the similarities between the clients' optimization directions for high training performance; 2) construct a new data-driven distance measure to improve the efficiency of the client clustering procedure. 3) implement a newcomer device cold start mechanism based on the auxiliary global model for framework scalability and practicality. FedGroup can achieve improvements by dividing joint optimization into groups of sub-optimization and can be combined with FL optimizer FedProx. The convergence and complexity are analyzed to demonstrate the efficiency of our proposed framework. We also evaluate FedGroup and FedGrouProx (combined with FedProx) on several open datasets and made comparisons with related CFL frameworks. The results show that FedGroup can significantly improve absolute test accuracy by +14.1% on FEMNIST compared to FedAvg. +3.4% on Sentiment140 compared to FedProx, +6.9% on MNIST compared to FeSEM. |
| **Date** | 2021-07-27 |
| **Short Title** | FedGroup |
| **Library Catalog** | arXiv.org |
| **URL** | <http://arxiv.org/abs/2010.06870> |
| **Accessed** | 1/4/2024, 12:20:28 AM |
| **Extra** | arXiv:2010.06870 [cs] |
| **DOI** | [10.48550/arXiv.2010.06870](http://doi.org/10.48550/arXiv.2010.06870) |
| **Repository** | arXiv |
| **Archive ID** | arXiv:2010.06870 |
| **Date Added** | 1/4/2024, 12:20:28 AM |
| **Modified** | 1/4/2024, 12:20:28 AM |

* **Tags:**
  + Computer Science - Machine Learning
  + Computer Science - Distributed, Parallel, and Cluster Computing

**Notes:**

* + Comment: This work will be presented at IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA) 2021. NOTE: This revision contains a crucial correction of the client cold start mechanism, please discard all previous manuscripts

**Attachments**

* + arXiv Fulltext PDF
  + arXiv.org Snapshot
* **FedGroup: Efficient Federated Learning via Decomposed Similarity-Based Clustering**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Moming Duan |
| **Author** | Duo Liu |
| **Author** | Xinyuan Ji |
| **Author** | Renping Liu |
| **Author** | Liang Liang |
| **Author** | Xianzhang Chen |
| **Author** | Yujuan Tan |
| **Abstract** | Federated Learning (FL) enables the multiple participating devices to collaboratively contribute to a global neural network model while keeping the training data locally. Unlike the centralized training setting, the non-IID and imbalanced (statistical heterogeneity) training data of FL is distributed in the federated network, which will increase the divergences between the local models and the global model, further degrading performance. In this paper, we propose a novel clustered federated learning (CFL) framework FedGroup, in which we 1) group the training of clients based on the similarities between the clients’ optimization directions for high training performance; 2) construct a new data-driven distance measure to improve the efficiency of the client clustering procedure. 3) implement a newcomer device cold start mechanism based on the auxiliary global model for framework scalability and practicality.FedGroup can achieve improvements by dividing joint optimization into groups of sub-optimization and can be combined with FL optimizer FedProx. The convergence and complexity are analyzed to demonstrate the efficiency of our proposed framework. We also evaluate FedGroup and FedGrouProx (combined with FedProx) on several open datasets and made comparisons with related CFL frameworks. The results show that FedGroup can significantly improve absolute test accuracy by +14.1% on FEMNIST compared to FedAvg, +3.4% on Sentiment140 compared to FedProx, +6.9% on MNIST compared to FeSEM. |
| **Date** | 2021-09 |
| **Short Title** | FedGroup |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/document/9644782> |
| **Accessed** | 1/4/2024, 12:16:31 AM |
| **Pages** | 228-237 |
| **Proceedings Title** | 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom) |
| **Conference Name** | 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom) |
| **DOI** | [10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00042](http://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00042) |
| **Date Added** | 1/4/2024, 12:16:31 AM |
| **Modified** | 1/4/2024, 12:16:31 AM |

* **Federated learning with incremental clustering for heterogeneous data**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Fabiola Espinoza Castellon |
| **Author** | Aurélien Mayoue |
| **Author** | Jacques-Henri Sublemontier |
| **Author** | Cédric Gouy-Pailler |
| **Abstract** | Federated learning enables different parties to collaboratively build a global model under the orchestration of a server while keeping the training data on clients' devices. However, performance is affected when clients have heterogeneous data. To cope with this problem, we assume that despite data heterogeneity, there are groups of clients who have similar data distributions that can be clustered. In previous approaches, in order to cluster clients the server requires clients to send their parameters simultaneously. However, this can be problematic in a context where there is a significant number of participants that may have limited availability. To prevent such a bottleneck, we propose FLIC (Federated Learning with Incremental Clustering), in which the server exploits the updates sent by clients during federated training instead of asking them to send their parameters simultaneously. Hence no additional communications between the server and the clients are necessary other than what classical federated learning requires. We empirically demonstrate for various non-IID cases that our approach successfully splits clients into groups following the same data distributions. We also identify the limitations of FLIC by studying its capability to partition clients at the early stages of the federated learning process efficiently. We further address attacks on models as a form of data heterogeneity and empirically show that FLIC is a robust defense against poisoning attacks even when the proportion of malicious clients is higher than 50%. |
| **Date** | 2022-07 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/document/9892653> |
| **Accessed** | 1/4/2024, 12:20:43 AM |
| **Extra** | ISSN: 2161-4407 |
| **Pages** | 1-8 |
| **Proceedings Title** | 2022 International Joint Conference on Neural Networks (IJCNN) |
| **Conference Name** | 2022 International Joint Conference on Neural Networks (IJCNN) |
| **DOI** | [10.1109/IJCNN55064.2022.9892653](http://doi.org/10.1109/IJCNN55064.2022.9892653) |
| **Date Added** | 1/4/2024, 12:20:44 AM |
| **Modified** | 1/4/2024, 12:20:44 AM |

* **Attachments**
  + IEEE Xplore Abstract Record
  + Submitted Version
* **Federated Learning for Big Data: A Survey on Opportunities, Applications, and Future Directions**

|  |  |
| --- | --- |
| **Item Type** | Preprint |
| **Author** | Thippa Reddy Gadekallu |
| **Author** | Quoc-Viet Pham |
| **Author** | Thien Huynh-The |
| **Author** | Sweta Bhattacharya |
| **Author** | Praveen Kumar Reddy Maddikunta |
| **Author** | Madhusanka Liyanage |
| **Abstract** | Big data has remarkably evolved over the last few years to realize an enormous volume of data generated from newly emerging services and applications and a massive number of Internet-of-Things (IoT) devices. The potential of big data can be realized via analytic and learning techniques, in which the data from various sources is transferred to a central cloud for central storage, processing, and training. However, this conventional approach faces critical issues in terms of data privacy as the data may include sensitive data such as personal information, governments, banking accounts. To overcome this challenge, federated learning (FL) appeared to be a promising learning technique. However, a gap exists in the literature that a comprehensive survey on FL for big data services and applications is yet to be conducted. In this article, we present a survey on the use of FL for big data services and applications, aiming to provide general readers with an overview of FL, big data, and the motivations behind the use of FL for big data. In particular, we extensively review the use of FL for key big data services, including big data acquisition, big data storage, big data analytics, and big data privacy preservation. Subsequently, we review the potential of FL for big data applications, such as smart city, smart healthcare, smart transportation, smart grid, and social media. Further, we summarize a number of important projects on FL-big data and discuss key challenges of this interesting topic along with several promising solutions and directions. |
| **Date** | 2021-10-17 |
| **Short Title** | Federated Learning for Big Data |
| **Library Catalog** | arXiv.org |
| **URL** | <http://arxiv.org/abs/2110.04160> |
| **Accessed** | 12/31/2023, 9:35:42 PM |
| **Extra** | arXiv:2110.04160 [cs] |
| **DOI** | [10.48550/arXiv.2110.04160](http://doi.org/10.48550/arXiv.2110.04160) |
| **Repository** | arXiv |
| **Archive ID** | arXiv:2110.04160 |
| **Date Added** | 12/31/2023, 9:35:42 PM |
| **Modified** | 12/31/2023, 9:35:42 PM |

* **Tags:**
  + Computer Science - Artificial Intelligence
  + Computer Science - Machine Learning

**Notes:**

* + Comment: Submitted for peer review in a journal

**Attachments**

* + arXiv Fulltext PDF
  + arXiv.org Snapshot
* **CFedPer: Clustered Federated Learning with Two-Stages Optimization for Personalization**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Zhipeng Gao |
| **Author** | Yan Yang |
| **Author** | Chen Zhao |
| **Author** | Zijia Mo |
| **Abstract** | Federated learning(FL) is a privacy-preserving dis-tributed learning paradigm in which clients cooperate with each other to train a global model. It is becoming progressively prevalent with the rapid development of edge devices. A critical challenge in federated learning is the data heterogeneity among clients, resulting in the global model generated by standard federated learning being unable to be adapted to all clients. To tackle this problem, we propose the CFedPer for personalized FL, which generates a personalized model for each cluster after clustering to address the deficiency of standard federated learning. Our algorithm is organized into two optimization phases. The pre-start phase clusters clients by our proposed similarity-based clustering model using distribution vector and similarity matrix. In the in-training phase, we represent the neural network as the base layer and personalization layer and propose a novel optimization objective with a regularization term for the personalization layer to achieve a balance between per-sonalization and generalization, preventing over-personalization. Extensive experiments on various datasets and data distributions indicate that the performance of our algorithm is superior to the existing algorithms in terms of average local accuracy and variance among clients. |
| **Date** | 2022-12 |
| **Short Title** | CFedPer |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/document/10076618> |
| **Accessed** | 1/4/2024, 12:20:24 AM |
| **Pages** | 171-177 |
| **Proceedings Title** | 2022 18th International Conference on Mobility, Sensing and Networking (MSN) |
| **Conference Name** | 2022 18th International Conference on Mobility, Sensing and Networking (MSN) |
| **DOI** | [10.1109/MSN57253.2022.00039](http://doi.org/10.1109/MSN57253.2022.00039) |
| **Date Added** | 1/4/2024, 12:20:24 AM |
| **Modified** | 1/4/2024, 12:20:24 AM |

* **Attachments**
  + IEEE Xplore Abstract Record
* **Federated learning with hyperparameter-based clustering for electrical load forecasting**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Nastaran Gholizadeh |
| **Author** | Petr Musilek |
| **Abstract** | Electrical load prediction has become an integral part of power system operation. Deep learning models have found popularity for this purpose. However, to achieve a desired prediction accuracy, they require huge amounts of data for training. Sharing electricity consumption data of individual households for load prediction may compromise user privacy and can be expensive in terms of communication resources. Therefore, edge computing methods, such as federated learning, are gaining more importance for this purpose. These methods can take advantage of the data without centrally storing it. This paper evaluates the performance of federated learning for short-term forecasting of individual house loads as well as the aggregate load. It discusses the advantages and disadvantages of this method by comparing it to centralized and local learning schemes. Moreover, a new client clustering method is proposed to reduce the convergence time of federated learning. The results show that federated learning has a good performance with a minimum root mean squared error (RMSE) of 0.117 kWh for individual load forecasting. |
| **Date** | 2022-03-01 |
| **Library Catalog** | ScienceDirect |
| **URL** | <https://www.sciencedirect.com/science/article/pii/S2542660521001104> |
| **Accessed** | 12/31/2023, 9:20:28 PM |
| **Volume** | 17 |
| **Pages** | 100470 |
| **Publication** | Internet of Things |
| **DOI** | [10.1016/j.iot.2021.100470](http://doi.org/10.1016/j.iot.2021.100470) |
| **Journal Abbr** | Internet of Things |
| **ISSN** | 2542-6605 |
| **Date Added** | 12/31/2023, 9:20:28 PM |
| **Modified** | 12/31/2023, 9:20:28 PM |

* **Tags:**
  + LSTM
  + Federated learning
  + Decentralized learning
  + Edge computing
  + Electricity load forecasting

**Attachments**

* + Submitted Version
* **Adaptive Client Clustering for Efficient Federated Learning over Non-IID and Imbalanced Data**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Biyao Gong |
| **Author** | Tianzhang Xing |
| **Author** | Zhidan Liu |
| **Author** | Wei Xi |
| **Author** | Xiaojiang Chen |
| **Abstract** | Federated learning (FL) is an emerging distributed and privacy-preserving machine learning framework. However, the performance of traditional FL methods is seriously impaired by the real-world data, which appear to be non-IID. The recent clustered federated learning (CFL) methods eliminate the impact of non-IID data by grouping clients with similar data distribution into the same cluster. Unfortunately, existing CFL methods heavily rely on the pre-setting of the cluster number, failing to achieve adaptive client clustering. We also experimentally observe that imbalanced data largely degrade their correctness of client clustering. In this paper, we present a novel CFL method without manual intervention, named AutoCFL, which can eliminate both effects of non-IID and imbalanced data simultaneously. To deal with imbalanced data, the local training adjustment strategy adaptively adjusts the number of local training epochs for each client. To further improve the clustering correctness and adaptability, the weighted voting-based client clustering strategy automatically groups each client into an appropriate cluster. Extensive experiments are conducted to evaluate the design of AutoCFL with three popular datasets under various data settings. Experimental results demonstrate that AutoCFL outperforms state-of-the-art methods, e.g., on average improving model accuracy by 9.24%, while reducing communication costs by 4.67 in an adaptive manner. |
| **Date** | 2022 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/9760109> |
| **Accessed** | 12/31/2023, 9:26:28 PM |
| **Extra** | Conference Name: IEEE Transactions on Big Data |
| **Pages** | 1-1 |
| **Publication** | IEEE Transactions on Big Data |
| **DOI** | [10.1109/TBDATA.2022.3167994](http://doi.org/10.1109/TBDATA.2022.3167994) |
| **ISSN** | 2332-7790 |
| **Date Added** | 12/31/2023, 9:26:28 PM |
| **Modified** | 12/31/2023, 9:26:28 PM |

* **Attachments**
  + IEEE Xplore Abstract Record
* **A Cluster-Driven Adaptive Training Approach for Federated Learning**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Younghwan Jeong |
| **Author** | Taeyoon Kim |
| **Abstract** | Federated learning (FL) is a promising collaborative learning approach in edge computing, reducing communication costs and addressing the data privacy concerns of traditional cloud-based training. Owing to this, diverse studies have been conducted to distribute FL into industry. However, there still remain the practical issues of FL to be solved (e.g., handling non-IID data and stragglers) for an actual implementation of FL. To address these issues, in this paper, we propose a cluster-driven adaptive training approach (CATA-Fed) to enhance the performance of FL training in a practical environment. CATA-Fed employs adaptive training during the local model updates to enhance the efficiency of training, reducing the waste of time and resources due to the presence of the stragglers and also provides a straggler mitigating scheme, which can reduce the workload of straggling clients. In addition to this, CATA-Fed clusters the clients considering the data size and selects the training participants within a cluster to reduce the magnitude differences of local gradients collected in the global model update under a statistical heterogeneous condition (e.g., non-IID data). During this client selection process, a proportional fair scheduling is employed for securing the data diversity as well as balancing the load of clients. We conduct extensive experiments using three benchmark datasets (MNIST, Fashion-MNIST, and CIFAR-10), and the results show that CATA-Fed outperforms the previous FL schemes (FedAVG, FedProx, and TiFL) with regard to the training speed and test accuracy under the diverse FL conditions. |
| **Date** | 2022/1 |
| **Language** | en |
| **Library Catalog** | www.mdpi.com |
| **URL** | <https://www.mdpi.com/1424-8220/22/18/7061> |
| **Accessed** | 12/31/2023, 9:27:09 PM |
| **Rights** | http://creativecommons.org/licenses/by/3.0/ |
| **Extra** | Number: 18 Publisher: Multidisciplinary Digital Publishing Institute |
| **Volume** | 22 |
| **Pages** | 7061 |
| **Publication** | Sensors |
| **DOI** | [10.3390/s22187061](http://doi.org/10.3390/s22187061) |
| **Issue** | 18 |
| **ISSN** | 1424-8220 |
| **Date Added** | 12/31/2023, 9:27:09 PM |
| **Modified** | 12/31/2023, 9:27:09 PM |

* **Tags:**
  + federated learning
  + adaptive training
  + clustering
  + non-IID
  + proportional fairness
  + straggler

**Attachments**

* + Full Text PDF
* **Efficient one-off clustering for personalized federated learning**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Tingting Liang |
| **Author** | Cheng Yuan |
| **Author** | Cheng Lu |
| **Author** | Youhuizi Li |
| **Author** | Junfeng Yuan |
| **Author** | Yuyu Yin |
| **Abstract** | In traditional federated learning such as FedAvg, the associations among clients are often ignored when executing on non-independent or heterogeneously distributed datasets, resulting in unsatisfactory accuracy. Although some previous works on clustered federated learning have been proposed to address such problems, most of them have a polarized problem. When the number of clustering is small, the model performs poorly and fails to accurately capture the distinction between clients. While a large number of clustering times tends to lead to higher communication costs. Therefore, a critical need is to design an efficient clustered federated solution that can both better capture the diversity between local clients and minimize the communication and computation costs. To this end, we propose an efficient one-off clustered federated learning framework called FedEOC. FedEOC exploits the “learning-to-learn” characteristic of meta-learning to enhance the generalization of the model across different clients so that only a small number of iterations are needed for each client to quickly obtain locally adapted weights. Based on the well-initially trained weights on all clients, we can cluster the clients only once to achieve the effect of one-off clustering and multiple-round applying. Additionally, to alleviate the issue of cluster imbalance, FedEOC is equipped with a Decomposition and Consolidation (Dec-Con) mechanism to decompose the clients from the extreme clusters and consolidate them into the most similar ones. The comprehensive experiments conducted on two real-world datasets demonstrate the superior capability of FedEOC from both aspects of accuracy and efficiency. |
| **Date** | 2023-10-09 |
| **Library Catalog** | ScienceDirect |
| **URL** | <https://www.sciencedirect.com/science/article/pii/S0950705123005634> |
| **Accessed** | 12/31/2023, 9:33:55 PM |
| **Volume** | 277 |
| **Pages** | 110813 |
| **Publication** | Knowledge-Based Systems |
| **DOI** | [10.1016/j.knosys.2023.110813](http://doi.org/10.1016/j.knosys.2023.110813) |
| **Journal Abbr** | Knowledge-Based Systems |
| **ISSN** | 0950-7051 |
| **Date Added** | 12/31/2023, 9:33:55 PM |
| **Modified** | 12/31/2023, 9:33:55 PM |

* **Tags:**
  + Federated learning
  + Decomposition and consolidation mechanism
  + Meta learning
  + One-off clustering
* **Federated Learning and Meta Learning: Approaches, Applications, and Directions**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Xiaonan Liu |
| **Author** | Yansha Deng |
| **Author** | Arumugam Nallanathan |
| **Author** | Mehdi Bennis |
| **Abstract** | Over the past few years, significant advancements have been made in the field of machine learning (ML) to address resource management, interference management, autonomy, and decision-making in wireless networks. Traditional ML approaches rely on centralized methods, where data is collected at a central server for training. However, this approach poses a challenge in terms of preserving the data privacy of devices. To address this issue, federated learning (FL) has emerged as an effective solution that allows edge devices to collaboratively train ML models without compromising data privacy. In FL, local datasets are not shared, and the focus is on learning a global model for a specific task involving all devices. However, FL has limitations when it comes to adapting the model to devices with different data distributions. In such cases, meta learning is considered, as it enables the adaptation of learning models to different data distributions using only a few data samples. In this tutorial, we present a comprehensive review of FL, meta learning, and federated meta learning (FedMeta). Unlike other tutorial papers, our objective is to explore how FL, meta learning, and FedMeta methodologies can be designed, optimized, and evolved, and their applications over wireless networks. We also analyze the relationships among these learning algorithms and examine their advantages and disadvantages in real-world applications. |
| **Date** | 2023 |
| **Short Title** | Federated Learning and Meta Learning |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/10310213> |
| **Accessed** | 12/31/2023, 9:31:40 PM |
| **Extra** | Conference Name: IEEE Communications Surveys & Tutorials |
| **Pages** | 1-1 |
| **Publication** | IEEE Communications Surveys & Tutorials |
| **DOI** | [10.1109/COMST.2023.3330910](http://doi.org/10.1109/COMST.2023.3330910) |
| **ISSN** | 1553-877X |
| **Date Added** | 12/31/2023, 9:31:41 PM |
| **Modified** | 12/31/2023, 9:31:41 PM |

* **Attachments**
  + Submitted Version
* **A Systematic Literature Review on Federated Machine Learning: From a Software Engineering Perspective**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Sin Kit Lo |
| **Author** | Qinghua Lu |
| **Author** | Chen Wang |
| **Author** | Hye-Young Paik |
| **Author** | Liming Zhu |
| **Abstract** | Federated learning is an emerging machine learning paradigm where clients train models locally and formulate a global model based on the local model updates. To identify the state-of-the-art in federated learning and explore how to develop federated learning systems, we perform a systematic literature review from a software engineering perspective, based on 231 primary studies. Our data synthesis covers the lifecycle of federated learning system development that includes background understanding, requirement analysis, architecture design, implementation, and evaluation. We highlight and summarise the findings from the results and identify future trends to encourage researchers to advance their current work. |
| **Date** | May 25, 2021 |
| **Short Title** | A Systematic Literature Review on Federated Machine Learning |
| **Library Catalog** | ACM Digital Library |
| **URL** | <https://doi.org/10.1145/3450288> |
| **Accessed** | 12/31/2023, 9:34:04 PM |
| **Volume** | 54 |
| **Pages** | 95:1–95:39 |
| **Publication** | ACM Computing Surveys |
| **DOI** | [10.1145/3450288](http://doi.org/10.1145/3450288) |
| **Issue** | 5 |
| **Journal Abbr** | ACM Comput. Surv. |
| **ISSN** | 0360-0300 |
| **Date Added** | 12/31/2023, 9:34:04 PM |
| **Modified** | 12/31/2023, 9:34:04 PM |

* **Tags:**
  + systematic literature review
  + Federated learning
  + distributed learning
  + edge learning
  + privacy
  + software engineering

**Attachments**

* + Submitted Version
* **Energy-efficient Clustering to Address Data Heterogeneity in Federated Learning**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Yibo Luo |
| **Author** | Xuefeng Liu |
| **Author** | Jianwei Xiu |
| **Abstract** | Federated Learning (FL) is a promising distributed learning paradigm and has gained recent attention from both academia and industry. One challenge in FL is that when local data across different devices are not independent and identically distributed (non-IID), models trained using FL generally have degraded performance. To address the problem, one natural approach is clustering: clients with similar data distributions are grouped into the same clusters and each cluster trains a specialized model. However, features utilized for clustering generally rely on a single global model trained during FL, whose convergence usually incurs high communication cost. In this paper, we propose CAFL, an energy-efficient clustering method in FL. In CAFL, clustering features of a client are not based on a collaboratively trained global model by FL, but a tensor of gradient vectors computed on local data. With this approach, the communication overhead for clustering is greatly reduced. We validated CAFL on simulated datasets include Fashion-MNIST and CIFAR-10, and the results show that compared with existing clustering methods in FL, CAFL has much lower communication cost while still ensuring a high clustering accuracy. |
| **Date** | 2021-06 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/document/9500901> |
| **Accessed** | 1/4/2024, 12:15:57 AM |
| **Extra** | ISSN: 1938-1883 |
| **Pages** | 1-6 |
| **Proceedings Title** | ICC 2021 - IEEE International Conference on Communications |
| **Conference Name** | ICC 2021 - IEEE International Conference on Communications |
| **DOI** | [10.1109/ICC42927.2021.9500901](http://doi.org/10.1109/ICC42927.2021.9500901) |
| **Date Added** | 1/4/2024, 12:15:57 AM |
| **Modified** | 1/4/2024, 12:15:57 AM |

* **Attachments**
  + IEEE Xplore Abstract Record
* **A state-of-the-art survey on solving non-IID data in Federated Learning**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Xiaodong Ma |
| **Author** | Jia Zhu |
| **Author** | Zhihao Lin |
| **Author** | Shanxuan Chen |
| **Author** | Yangjie Qin |
| **Abstract** | Federated Learning (FL) proposed in recent years has received significant attention from researchers in that it can enable multiple clients to cooperatively train global models without revealing private data. This training mode protects users’ privacy without violating the supervision, and aggregates scattered data to exert great potential. However, the data samples on each participating device of FL are usually not independent and identically distributed (IID), which leads to serious statistical heterogeneity challenges for FL. In this article, we analyze and establish the definition of non-IID data problems, and put forward a series of challenges that this problem may bring to FL. We classify existing methods to solve this problem from the researcher’s entry point and subsequent sub-methods, aiming to provide a comprehensive study for solving the problem of non-IID data in FL. Our research shows that non-IID data will not only reduce the performance of the FL model, but also damage the active participation of users in the FL process. Compared with methods based on data-side sharing, enhancement, and selection, it is more common for researchers to improve federated learning algorithms from models, algorithms, and frameworks to solve non-IID problems. To the best of our knowledge, although many efforts have been made to address the problem of non-IID data, there are currently few authoritative systematic reviews in this field and are not up-to-date. In this article, we will fill the gaps in FL and provide researchers with the state-of-the-art research results to solve non-IID problems in FL and promote the further implementation of FL. |
| **Date** | 2022-10-01 |
| **Library Catalog** | ScienceDirect |
| **URL** | <https://www.sciencedirect.com/science/article/pii/S0167739X22001686> |
| **Accessed** | 12/31/2023, 9:21:22 PM |
| **Volume** | 135 |
| **Pages** | 244-258 |
| **Publication** | Future Generation Computer Systems |
| **DOI** | [10.1016/j.future.2022.05.003](http://doi.org/10.1016/j.future.2022.05.003) |
| **Journal Abbr** | Future Generation Computer Systems |
| **ISSN** | 0167-739X |
| **Date Added** | 12/31/2023, 9:21:22 PM |
| **Modified** | 12/31/2023, 9:21:22 PM |

* **Tags:**
  + Machine learning
  + Federated Learning
  + Non-IID data
  + Statistical heterogeneity

**Attachments**

* + ScienceDirect Snapshot
* **CADIS: Handling Cluster-skewed Non-IID Data in Federated Learning with Clustered Aggregation and Knowledge DIStilled Regularization**

|  |  |
| --- | --- |
| **Item Type** | Preprint |
| **Author** | Nang Hung Nguyen |
| **Author** | Duc Long Nguyen |
| **Author** | Trong Bang Nguyen |
| **Author** | Thanh-Hung Nguyen |
| **Author** | Huy Hieu Pham |
| **Author** | Truong Thao Nguyen |
| **Author** | Phi Le Nguyen |
| **Abstract** | Federated learning enables edge devices to train a global model collaboratively without exposing their data. Despite achieving outstanding advantages in computing efficiency and privacy protection, federated learning faces a significant challenge when dealing with non-IID data, i.e., data generated by clients that are typically not independent and identically distributed. In this paper, we tackle a new type of Non-IID data, called cluster-skewed non-IID, discovered in actual data sets. The cluster-skewed non-IID is a phenomenon in which clients can be grouped into clusters with similar data distributions. By performing an in-depth analysis of the behavior of a classification model's penultimate layer, we introduce a metric that quantifies the similarity between two clients' data distributions without violating their privacy. We then propose an aggregation scheme that guarantees equality between clusters. In addition, we offer a novel local training regularization based on the knowledge-distillation technique that reduces the overfitting problem at clients and dramatically boosts the training scheme's performance. We theoretically prove the superiority of the proposed aggregation over the benchmark FedAvg. Extensive experimental results on both standard public datasets and our in-house real-world dataset demonstrate that the proposed approach improves accuracy by up to 16% compared to the FedAvg algorithm. |
| **Date** | 2023-04-15 |
| **Short Title** | CADIS |
| **Library Catalog** | arXiv.org |
| **URL** | <http://arxiv.org/abs/2302.10413> |
| **Accessed** | 1/4/2024, 12:20:32 AM |
| **Extra** | arXiv:2302.10413 [cs] |
| **DOI** | [10.48550/arXiv.2302.10413](http://doi.org/10.48550/arXiv.2302.10413) |
| **Repository** | arXiv |
| **Archive ID** | arXiv:2302.10413 |
| **Date Added** | 1/4/2024, 12:20:32 AM |
| **Modified** | 1/4/2024, 12:20:32 AM |

* **Tags:**
  + Computer Science - Machine Learning
  + Computer Science - Computer Vision and Pattern Recognition

**Notes:**

* + Comment: Accepted for presentation at the 23rd International Symposium on Cluster, Cloud and Internet Computing (CCGrid 2023)

**Attachments**

* + arXiv Fulltext PDF
  + arXiv.org Snapshot
* **Federated Learning for Internet of Things: A Comprehensive Survey**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Dinh C. Nguyen |
| **Author** | Ming Ding |
| **Author** | Pubudu N. Pathirana |
| **Author** | Aruna Seneviratne |
| **Author** | Jun Li |
| **Author** | H. Vincent Poor |
| **Abstract** | The Internet of Things (IoT) is penetrating many facets of our daily life with the proliferation of intelligent services and applications empowered by artificial intelligence (AI). Traditionally, AI techniques require centralized data collection and processing that may not be feasible in realistic application scenarios due to the high scalability of modern IoT networks and growing data privacy concerns. Federated Learning (FL) has emerged as a distributed collaborative AI approach that can enable many intelligent IoT applications, by allowing for AI training at distributed IoT devices without the need for data sharing. In this article, we provide a comprehensive survey of the emerging applications of FL in IoT networks, beginning from an introduction to the recent advances in FL and IoT to a discussion of their integration. Particularly, we explore and analyze the potential of FL for enabling a wide range of IoT services, including IoT data sharing, data offloading and caching, attack detection, localization, mobile crowdsensing, and IoT privacy and security. We then provide an extensive survey of the use of FL in various key IoT applications such as smart healthcare, smart transportation, Unmanned Aerial Vehicles (UAVs), smart cities, and smart industry. The important lessons learned from this review of the FL-IoT services and applications are also highlighted. We complete this survey by highlighting the current challenges and possible directions for future research in this booming area. |
| **Date** | 2021 |
| **Short Title** | Federated Learning for Internet of Things |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/9415623> |
| **Accessed** | 12/31/2023, 9:21:27 PM |
| **Extra** | Conference Name: IEEE Communications Surveys & Tutorials |
| **Volume** | 23 |
| **Pages** | 1622-1658 |
| **Publication** | IEEE Communications Surveys & Tutorials |
| **DOI** | [10.1109/COMST.2021.3075439](http://doi.org/10.1109/COMST.2021.3075439) |
| **Issue** | 3 |
| **ISSN** | 1553-877X |
| **Date Added** | 12/31/2023, 9:21:28 PM |
| **Modified** | 12/31/2023, 9:21:28 PM |

* **Attachments**
  + IEEE Xplore Abstract Record
  + IEEE Xplore Full Text PDF
* **The Communication-Aware Clustered Federated Learning Problem**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Nir Shlezinger |
| **Author** | Stefano Rini |
| **Author** | Yonina C. Eldar |
| **Abstract** | Federated learning (FL) refers to the adaptation of a central model based on data sets available at multiple remote users. Two of the common challenges encountered in FL are the fact that training sets obtained by different users are commonly heterogeneous, i.e., arise from different sample distributions, and the need to communicate large amounts of data between the users and the central server over the typically expensive up-link channel. In this work we formulate the problem of FL in which different clusters of users observe labeled samples drawn from different distributions, while operating under constraints on the communication overhead. For such settings, we identify that the combination of statistical heterogeneity and communication constraints induces a tradeoff between the ability of the users of each cluster to learn a proper model and the accuracy in aggregating these models into a global inference rule. We propose an algorithm based on multi-source adaptation methods for such communication-aware clustered FL scenarios which allows to balance these performance measures, and demonstrate its ability to achieve improved inference over conventional federated averaging without inducing additional communication overhead. |
| **Date** | 2020-06 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/9174245> |
| **Accessed** | 12/31/2023, 9:26:01 PM |
| **Extra** | ISSN: 2157-8117 |
| **Pages** | 2610-2615 |
| **Proceedings Title** | 2020 IEEE International Symposium on Information Theory (ISIT) |
| **Conference Name** | 2020 IEEE International Symposium on Information Theory (ISIT) |
| **DOI** | [10.1109/ISIT44484.2020.9174245](http://doi.org/10.1109/ISIT44484.2020.9174245) |
| **Date Added** | 12/31/2023, 9:26:01 PM |
| **Modified** | 12/31/2023, 9:26:01 PM |

* **A survey on participant selection for federated learning in mobile networks**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Behnaz Soltani |
| **Author** | Venus Haghighi |
| **Author** | Adnan Mahmood |
| **Author** | Quan Z. Sheng |
| **Author** | Lina Yao |
| **Abstract** | Federated Learning (FL) is an efficient distributed machine learning paradigm that employs private datasets in a privacy-preserving manner. The main challenges of FL are that end devices usually possess various computation and communication capabilities and their training data are not independent and identically distributed (non-IID). Due to limited communication bandwidth and unstable availability of such devices in a mobile network, only a fraction of end devices (also referred to as the participants or clients in a FL process) can be selected in each round. Hence, it is of paramount importance to utilize an efficient participant selection scheme to maximize the performance of FL including final model accuracy and training time. In this paper, we provide a review of participant selection techniques for FL. First, we introduce FL and highlight the main challenges during participant selection. Then, we review the existing studies and categorize them based on their solutions. Finally, we provide some future directions on participant selection for FL based on our analysis of the state-of-the-art in this topic area. |
| **Date** | October 17, 2022 |
| **Library Catalog** | ACM Digital Library |
| **URL** | <https://doi.org/10.1145/3556548.3559633> |
| **Accessed** | 12/31/2023, 6:00:00 AM |
| **Place** | New York, NY, USA |
| **Publisher** | Association for Computing Machinery |
| **ISBN** | 978-1-4503-9518-2 |
| **Pages** | 19–24 |
| **Series** | MobiArch '22 |
| **Proceedings Title** | Proceedings of the 17th ACM Workshop on Mobility in the Evolving Internet Architecture |
| **DOI** | [10.1145/3556548.3559633](http://doi.org/10.1145/3556548.3559633) |
| **Date Added** | 12/31/2023, 9:33:40 PM |
| **Modified** | 12/31/2023, 9:33:40 PM |

* **Tags:**
  + machine learning
  + federated learning
  + participant selection

**Attachments**

* + Submitted Version
* **Towards Federated Clustering: A Federated Fuzzy $c$-Means Algorithm (FFCM)**

|  |  |
| --- | --- |
| **Item Type** | Preprint |
| **Author** | Morris Stallmann |
| **Author** | Anna Wilbik |
| **Abstract** | Federated Learning (FL) is a setting where multiple parties with distributed data collaborate in training a joint Machine Learning (ML) model while keeping all data local at the parties. Federated clustering is an area of research within FL that is concerned with grouping together data that is globally similar while keeping all data local. We describe how this area of research can be of interest in itself, or how it helps addressing issues like non-independently-identically-distributed (i.i.d.) data in supervised FL frameworks. The focus of this work, however, is an extension of the federated fuzzy $c$-means algorithm to the FL setting (FFCM) as a contribution towards federated clustering. We propose two methods to calculate global cluster centers and evaluate their behaviour through challenging numerical experiments. We observe that one of the methods is able to identify good global clusters even in challenging scenarios, but also acknowledge that many challenges remain open. |
| **Date** | 2022-01-18 |
| **Short Title** | Towards Federated Clustering |
| **Library Catalog** | arXiv.org |
| **URL** | <http://arxiv.org/abs/2201.07316> |
| **Accessed** | 12/31/2023, 9:32:51 PM |
| **Extra** | arXiv:2201.07316 [cs] |
| **DOI** | [10.48550/arXiv.2201.07316](http://doi.org/10.48550/arXiv.2201.07316) |
| **Repository** | arXiv |
| **Archive ID** | arXiv:2201.07316 |
| **Date Added** | 12/31/2023, 9:32:51 PM |
| **Modified** | 12/31/2023, 9:32:51 PM |

* **Tags:**
  + Computer Science - Machine Learning

**Notes:**

* + Comment: International Workshop on Trustable, Verifiable and Auditable Federated Learning in Conjunction with AAAI 2022 (FL-AAAI-22)

**Attachments**

* + arXiv Fulltext PDF
  + arXiv.org Snapshot
* **Personalized Federated Learning with Clustered Generalization**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Xueyang Tang |
| **Author** | Song Guo |
| **Author** | Jingcai Guo |
| **Abstract** | The prevalent personalized federated learning (PFL) usually pursues a trade-off between personalization and generalization by maintaining a shared global model to guide the training process of local models. However, the sole global model may easily transfer deviated knowledge (e.g., biased updates) to some local models when rich statistical diversity exists across the local datasets. Thus, we argue it is of crucial importance to maintain the diversity of generalization to provide each client with fine-grained common knowledge that can better fit the local data distributions and facilitate faster model convergence. In this paper, we propose a novel concept called clustered generalization (CG) to handle the challenge of statistical heterogeneity, and properly design a CG-based framework of PFL, dubbed CGPFL. Concretely, we maintain K global (i.e., generalized) models in the server and each local model is dynamically associated with the nearest global model to conduct ‘push’ and ‘pull’ operations during the iterative algorithm. We conduct detailed theoretical analysis, in which the convergence guarantee is presented and $\mathcal{O}(\sqrt{K})$ speedup over most existing methods is granted. To quantitatively study the generalization-personalization trade-off, we introduce the ‘generalization error’ measure and prove that the proposed CGPFL can achieve a better trade-off than existing solutions. Moreover, our theoretical analysis further inspires a heuristic algorithm to find a near-optimal trade-off in CGPFL. Experimental results on multiple real-world datasets show that our approach surpasses the state-of-the-art methods on test accuracy by a significant margin. |
| **Date** | 2021/10/06 |
| **Language** | en |
| **Library Catalog** | openreview.net |
| **URL** | <https://openreview.net/forum?id=dJk1vpEFYF0> |
| **Accessed** | 12/31/2023, 9:32:43 PM |
| **Date Added** | 12/31/2023, 9:32:43 PM |
| **Modified** | 12/31/2023, 9:32:43 PM |

* **Attachments**
  + Full Text PDF
* **Federated Learning based Energy Demand Prediction with Clustered Aggregation**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Ye Lin Tun |
| **Author** | Kyi Thar |
| **Author** | Chu Myaet Thwal |
| **Author** | Choong Seon Hong |
| **Abstract** | To reduce negative environmental impacts, power stations and energy grids need to optimize the resources required for power production. Thus, predicting the energy consumption of clients is becoming an important part of every energy management system. Energy usage information collected by the clients’ smart homes can be used to train a deep neural network to predict the future energy demand. Collecting data from a large number of distributed clients for centralized model training is expensive in terms of communication resources. To take advantage of distributed data in edge systems, centralized training can be replaced by federated learning where each client only needs to upload model updates produced by training on its local data. These model updates are aggregated into a single global model by the server. But since different clients can have different attributes, model updates can have diverse weights and as a result, it can take a long time for the aggregated global model to converge. To speed up the convergence process, we can apply clustering to group clients based on their properties and aggregate model updates from the same cluster together to produce a cluster specific global model. In this paper, we propose a recurrent neural network based energy demand predictor, trained with federated learning on clustered clients to take advantage of distributed data and speed up the convergence process. |
| **Date** | 2021-01 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/9373194> |
| **Accessed** | 12/31/2023, 9:32:28 PM |
| **Extra** | ISSN: 2375-9356 |
| **Pages** | 164-167 |
| **Proceedings Title** | 2021 IEEE International Conference on Big Data and Smart Computing (BigComp) |
| **Conference Name** | 2021 IEEE International Conference on Big Data and Smart Computing (BigComp) |
| **DOI** | [10.1109/BigComp51126.2021.00039](http://doi.org/10.1109/BigComp51126.2021.00039) |
| **Date Added** | 12/31/2023, 9:32:28 PM |
| **Modified** | 12/31/2023, 9:32:28 PM |

* **Attachments**
  + Submitted Version
* **Clustered federated learning with weighted model aggregation for imbalanced data**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Dong Wang |
| **Author** | Naifu Zhang |
| **Author** | Meixia Tao |
| **Abstract** | As a promising edge learning framework in future 6G networks, federated learning (FL) faces a number of technical challenges due to the heterogeneous network environment and diversified user behaviors. Data imbalance is one of these challenges that can significantly degrade the learning efficiency. To deal with data imbalance issue, this work proposes a new learning framework, called clustered federated learning with weighted model aggregation (weighted CFL). Compared with traditional FL, our weighted CFL adaptively clusters the participating edge devices based on the cosine similarity of their local gradients at each training iteration, and then performs weighted per-cluster model aggregation. Therein, the similarity threshold for clustering is adaptive over iterations in response to the time-varying divergence of local gradients. Moreover, the weights for per-cluster model aggregation are adjusted according to the data balance feature so as to speed up the convergence rate. Experimental results show that the proposed weighted CFL achieves a faster model convergence rate and greater learning accuracy than benchmark methods under the imbalanced data scenario. |
| **Date** | 2022-08 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/9861223> |
| **Accessed** | 12/31/2023, 9:34:46 PM |
| **Extra** | Conference Name: China Communications |
| **Volume** | 19 |
| **Pages** | 41-56 |
| **Publication** | China Communications |
| **DOI** | [10.23919/JCC.2022.08.004](http://doi.org/10.23919/JCC.2022.08.004) |
| **Issue** | 8 |
| **ISSN** | 1673-5447 |
| **Date Added** | 12/31/2023, 9:34:46 PM |
| **Modified** | 12/31/2023, 9:34:46 PM |

* **HACCS: Heterogeneity-Aware Clustered Client Selection for Accelerated Federated Learning**

|  |  |
| --- | --- |
| **Item Type** | Conference Paper |
| **Author** | Joel Wolfrath |
| **Author** | Nikhil Sreekumar |
| **Author** | Dhruv Kumar |
| **Author** | Yuanli Wang |
| **Author** | Abhishek Chandra |
| **Abstract** | Federated Learning is a machine learning paradigm where a global model is trained in-situ across a large number of distributed edge devices. While this technique avoids the cost of transferring data to a central location and achieves a strong degree of privacy, it presents additional challenges due to the heterogeneous hardware resources available for training. Furthermore, data is not independent and identically distributed (IID) across all edge devices, resulting in statistical heterogeneity across devices. Due to these constraints, client selection strategies play an important role for timely convergence during model training. Existing strategies ensure that each individual device is included, at least periodically, in the training process. In this work, we propose HACCS, a Heterogeneity-Aware Clustered Client Selection system that identifies and exploits the statistical heterogeneity by representing all distinguishable data distributions instead of individual devices in the training process. HACCS is robust to individual device dropout, provided other devices in the system have similar data distributions. We propose privacy-preserving methods for estimating these client distributions and clustering them. We also propose strategies for leveraging these clusters to make scheduling decisions in a federated learning system. Our evaluation on real-world datasets suggests that our framework can provide 18% −38% reduction in time to convergence compared to the state of the art without any compromise in accuracy. |
| **Date** | 2022-05 |
| **Short Title** | HACCS |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/document/9820684> |
| **Accessed** | 1/4/2024, 12:20:25 AM |
| **Extra** | ISSN: 1530-2075 |
| **Pages** | 985-995 |
| **Proceedings Title** | 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS) |
| **Conference Name** | 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS) |
| **DOI** | [10.1109/IPDPS53621.2022.00100](http://doi.org/10.1109/IPDPS53621.2022.00100) |
| **Date Added** | 1/4/2024, 12:20:25 AM |
| **Modified** | 1/4/2024, 12:20:25 AM |

* **Attachments**
  + IEEE Xplore Abstract Record
* **Hierarchical Federated Learning With Social Context Clustering-Based Participant Selection for Internet of Medical Things Applications**

|  |  |
| --- | --- |
| **Item Type** | Journal Article |
| **Author** | Xiaokang Zhou |
| **Author** | Xiaozhou Ye |
| **Author** | Kevin I-Kai Wang |
| **Author** | Wei Liang |
| **Author** | Nirmal Kumar C. Nair |
| **Author** | Shohei Shimizu |
| **Author** | Zheng Yan |
| **Author** | Qun Jin |
| **Abstract** | The proliferation in embedded and communication technologies made the concept of the Internet of Medical Things (IoMT) a reality. Individuals’ physical and physiological status can be constantly monitored, and numerous data can be collected through wearable and mobile devices. However, the silo of individual data brings limitations to existing machine learning approaches to correctly identify a user’s health status. Distributed machine learning paradigms, such as federated learning, offer a potential solution for privacy-preserving knowledge sharing without sending raw personal data. However, federated learning is vulnerable to harmful participants that can degrade the overall model quality by sharing low-quality data. Therefore, it is critical to select suitable participants to ensure the accuracy and efficiency of federated learning. In this article, a unique clustering-based approach is proposed to use social context data for participant selection. Different edge participant groups will be established, and group-specific federated learning will be performed. The models of various edge groups will be further aggregated to strengthen the robustness of the global model. The experimental results demonstrated that through participant selection, clustering-based hierarchical federated learning can achieve better results with less participants in two different IoMT applications for ECG and human motion monitoring. This shows the efficacy of the proposed method in improving federated learning performance and efficiency in various IoMT applications. |
| **Date** | 2023-08 |
| **Library Catalog** | IEEE Xplore |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/10091843> |
| **Accessed** | 12/31/2023, 9:33:04 PM |
| **Extra** | Conference Name: IEEE Transactions on Computational Social Systems |
| **Volume** | 10 |
| **Pages** | 1742-1751 |
| **Publication** | IEEE Transactions on Computational Social Systems |
| **DOI** | [10.1109/TCSS.2023.3259431](http://doi.org/10.1109/TCSS.2023.3259431) |
| **Issue** | 4 |
| **ISSN** | 2329-924X |
| **Date Added** | 12/31/2023, 9:33:04 PM |
| **Modified** | 12/31/2023, 9:33:04 PM |

* **Attachments**
  + IEEE Xplore Abstract Record
* **[2202.06187] On the Convergence of Clustered Federated Learning**

|  |  |
| --- | --- |
| **Item Type** | Web Page |
| **URL** | <https://arxiv.org/abs/2202.06187> |
| **Accessed** | 1/4/2024, 12:20:38 AM |
| **Date Added** | 1/4/2024, 12:20:38 AM |
| **Modified** | 1/4/2024, 12:20:38 AM |

* **Attachments**
  + [2202.06187] On the Convergence of Clustered Federated Learning
* **FedLC: Optimizing Federated Learning in Non-IID Data via Label-Wise Clustering | IEEE Journals & Magazine | IEEE Xplore**

|  |  |
| --- | --- |
| **Item Type** | Web Page |
| **URL** | <https://ieeexplore.ieee.org/abstract/document/10110974> |
| **Accessed** | 12/31/2023, 9:31:58 PM |
| **Date Added** | 12/31/2023, 9:31:58 PM |
| **Modified** | 12/31/2023, 9:31:58 PM |