Temporal Adaptive Clustering for Heterogeneous Clients in Federated Learning

Syed Saqib Ali, Ajit Kumar, Mazhar Ali, Ankit Kumar Singh, and Bong Jun Choi
School of Computer Science and Engineering
Soongsil University Seoul, Korea
{kazmi, ankit, mazhar, davidchoi}@soongsil.ac.kr, ajitkumar.pu@gmail.com

***Abstract*—Federated learning has emerged as a highly promising approach for training machine learning models across a decentralized network of clients, with a key focus on maintaining the privacy of data. Nevertheless, the management of system heterogeneity and the handling of time-varying interests continue to pose hurdles for conventional federated learning methodologies. This work presents temporal-based adaptive clustered federated learning as a viable solution to the difficulties mentioned above. The evaluation of clusterability is conducted by calculating the Silhouette score following each iteration of federated training. The process of model aggregation is performed at the cluster level, resulting in enhanced convergence efficiency and improved accuracy of predictions. The inclusion of temporal-based adaptiveness in clustered federated learning for time-varying environments enables the system to dynamically modify cluster configurations in response to clients joining or leaving the network. The experimental results on a real-world dataset of an electric vehicle charging station network illustrate the efficacy of the suggested approach in terms of model correctness, convergence, and adaptability. The temporal-based adaptive clustered federated learning framework has demonstrated significant advancements compared to the current state-of-the-art clustered federated learning approaches.**

***Index Terms*—Federated Learning, Energy Demand Prediction, Clustered Federated Learning, Adaptive Clustered Federated Learning, Temporal based CFL**

I. Introduction

Federated Learning (FL) [1], a cutting-edge approach in machine learning (ML), redefines the way of model training and processing. It uses the strength of distributed devices to cooperatively improve their model performance while assuring the locality of their data, in contrast to conventional centralized methods where data is gathered in a single location. With the help of this ground-breaking paradigm, a variety of devices, including smartphones, edge devices, and IoT sensors, can enhance their model's capabilities without the need to provide raw data. FL has become an essential mechanism to enable privacy-preserving collaborative learning to improve a single global model and deliver better-personalized models tailored to the end-users' local data and context [2; 3; 4]. However, it poses some challenges, such as data heterogeneity, data imbalances, time-varying data distribution, and diversified user behaviors while training a global model. Hence, extensive research works have been underway in many domains to mitigate these challenges and find global solutions.

Clustered federated learning (CFL) is a significant advancement in the domain of distributed machine learning since it effectively combines the benefits of federated learning and cluster analysis. This innovative approach, however, efficiently addresses these difficulties within the framework of collaborative model training. The utilization of CFL enhances the performance of the client's models by expediting the convergence time, guaranteeing cost minimization, and clustering them with similar data/model characteristics. The CFL technique enhances the trade-off between accuracy and local adaptability by facilitating the creation of tailored models for individual clients within a cluster.

Clients participate in the training process individually in classical FL, and it is assumed that their data distributions would remain static over time. The distribution of client data, however, may alter over time in real-world circumstances due to a variety of variables, including user behavior, seasonal patterns, or changing user preferences. Therefore, to increase the effectiveness and performance of the federated learning process, we propose a temporal-based adaptive clustered federated learning (TACFL), where clients are grouped into clusters according to the temporal patterns in their data distributions. TACFL focuses on identifying and utilizing these temporal patterns in clients' data distributions in the federated clustering process. By utilizing adaptiveness in the clustering approach, it overcomes the limitations of traditional clustering techniques. By tackling the complexities of time series and streaming data networks, the proposed algorithm can provide a detailed understanding of the clustering dynamics in these circumstances. In sum, TACFL can help to overcome the challenges of time-varying data distribution and improve the accuracy of model training in such federated settings.

The rest of the paper is organized as follows. Section II discusses some related works in the field. The system model with the energy demand learning and proposed clustering model is highlighted in Section III. Section IV describes the experimental setup and performance of the proposed approach. The paper is summarized with the conclusion in Section V.

II. Literature Review

Federated learning enables distributed clients to jointly train machine learning models in a privacy-preserving way without sharing their local data. In federated settings, it is common to have several clients and a central server. The primary role of the server is to aggregate the local models from all clients in order to generate a global model. On the other hand, clients are responsible for managing the collection and storage of local data, as well as training their respective local models. However, it still poses numerous challenges due to clients' data heterogeneity, user preferences, streaming data, and communication cost [5].

CFL is one of the potential solutions recently proposed in numerous works [2; 3], where clients with jointly trainable data distribution are grouped and trained accordingly and the cluster-specific models are distributed back. In this way, the communication cost is reduced, with a fast convergence rate and more accurate results. Most of the existing CFL techniques utilize three common approaches: local model updates, local model weights, and losses-based collaborative training to find the similarity among clients [6]. However, there is still scope for improvement in the existing heuristic approaches in finding optimal numbers of clusters in CFL. Adaptive CFL, without specifying the number of clusters, groups clients into appropriate clusters based on learning from the model features. Wang et al. [7] have successfully introduced the adaptive clustered FL to clients with time-varying interests and for streaming nature of data. Saputra et al. [8] and Perry et al. [9] have applied FL and CFL for the energy demand of electric vehicle CSs network and successfully trained the clustered models for the charging stations using the conventional clustering methods, such as K-means clustering and hierarchical clustering with Euclidean and DTW measures. Tun et al. [10] have applied federated clustering for residential energy demand prediction by leveraging the RNN model.

The temporal nature of data, like time series data generation, is the core part of the modern world of IoTs. To deal with such real-world problems, we propose TACFL, where we introduce a dynamic clustering method with varying number of clusters based on Silhouette threshold to adapt with time-variant clients better.

TABLE I
LIST OF SYMBOLS USED

Symbols	**Description**	**Symbols**	**Description**
S_i	Silhouette coefficient	n_k	Number of samples of each client
α_i	Learning rate	L_{init}	Initial cluster label
L_{new}	New cluster label	E	Number of epochs
m	Number of clients	n	Number of neurons
ω	Trained model	Θ	Cumulative adaptivity
$\mathcal{D}$	Client dataset	g_i	Initial cluster number
δ	Decay rate	C_i	Final cluster number
$\mathcal{K}$	Number of clusters	ϵ	Silhouette threshold
θ_i	Adaptive feature	T	Number of rounds

III. System Model

A. System Environment

This study examines an EV charging station (CS) network in a typical urban region, Boulder City, USA, consisting of a set of 39 CSs across 5 postal code regions in the city. Each CS is equipped with computation and storing capabilities over its transaction log that records EV charging sessions occurring at the stations. This log file comprises information such as the CS ID, EV ID, charging date, charging time, energy consumption over time, etc. Each charging station does the centralized and federated training on its local dataset and establishes communication with servers and other CSs within the network for sharing and learning the model. The comprehensive mechanism of our technique from centralized to cluster-specific training is elucidated in Figure 1, and the list of symbols used in this work is shown in Table I.

B. Energy demand Prediction Model

Since the energy demand prediction for clients based on a temporal series of past usage patterns is a time-series regression task, the electric vehicle CSs have limited datasets (transactions) and computational capabilities to execute it locally. Therefore, in the centralized approach, all the clients send their data to the central server, which then makes an accumulated log file before training for the global model. Then, leveraging the based model, the server calculates the energy demand for the next period in a number of rounds and calculates the prediction error till it converges to the satisfied value. Then, the final global model is sent back to each client for local learning.

Since, in the above approach, the charging stations are sharing their raw data with the central server, there are privacy concerns and huge data communication overhead. To deal with these issues, an FL-based energy demand prediction model is proposed in this work, where the server only needs the collect the trained models from all the CSs. By leveraging the FL approach, the server aggregates the local models and shares the learned global model back to them. Each CS trains the global model on their local data to learn the aggregated demand as

$$w_k^{(t)} = w_k^{(t)} - w^{(t)}, \tag{1}$$

$$w^{(t)} = \frac{\sum_{k=1}^{K} n_k w_k^{(t)}}{\sum_{k=1}^{K} n_k}, \tag{2}$$

with a learning rate of

$$\alpha_i = \frac{\alpha_0}{1 + \delta t}, \tag{3}$$

where α_i is the learning rate, α_0 is the initial learning rate, δ is the decay rate, and t is the current epoch or iteration.

Then, the server updates the global model after multiple rounds till the error converges to a minimum. The prediction error is calculated and based on it readjust the local model weights for further clustering process.

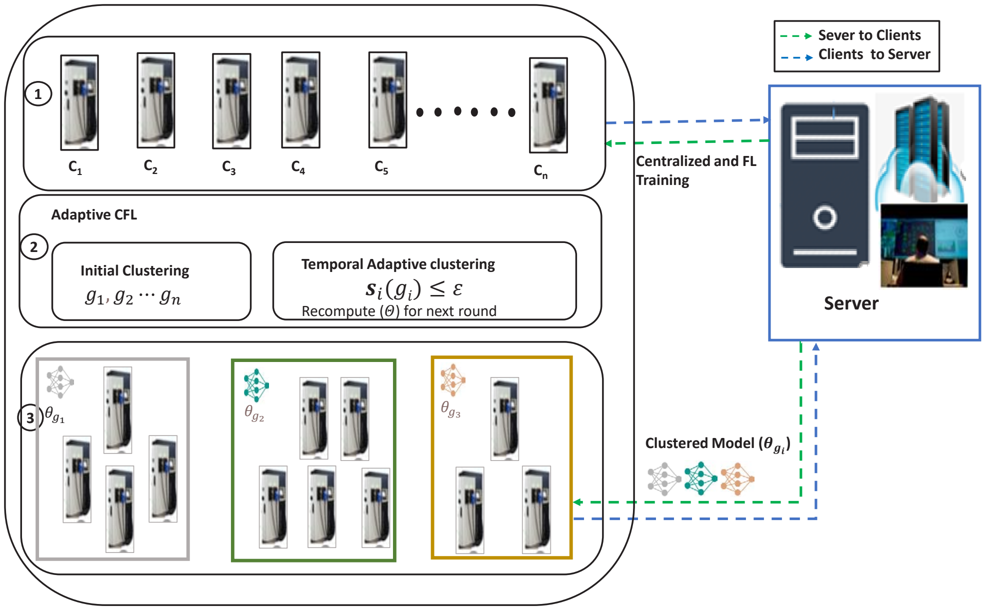

Fig. 1. Overview of TACFL system model: (1) centralized and federated training of clients dataset, (2) adaptive clustering based on Silhouette threshold, (3) clustered specific model training

C. Temporal-based Adaptive Clustering Model

A complex data analysis technique known as *temporal-based adaptive clustering* makes use of temporal information to divide and organize data points dynamically into groups throughout time. Temporal-based adaptive clustering recognizes the changing nature of data. It adapts clusters to take into account the changing patterns and trends, in contrast to typical static clustering approaches, which presume that data remains constant during the training. When dealing with time series data, where observations are made across a range of periods, this method is especially critical. Temporal-based adaptive clustering provides more precise insights into complicated datasets. It encourages the detection of dynamic links and trends that could otherwise go undetected by continuously updating clusters to reflect changing patterns.

1) Cluster Modeling: In a temporal-based adaptive clustering scenario, the focus shifts to the implementation of the K-means algorithm for dynamic cluster adaptation. Initially, historical temporal data from the charging station network is segmented into time intervals, creating subsets that capture temporal dynamics. These subsets are individually processed through the K-means algorithm, forming initial clusters. The crucial aspect is where the centroids of these initial clusters are leveraged as references for future adaptation. As new data becomes available, it is evaluated against these centroids, and clusters are updated accordingly. This iterative process ensures that the clusters remain attuned to the evolving patterns in the data. The re-evaluation of centroids and cluster assignments occurs periodically, preventing clusters from stagnating. This re-evaluation is carried out by computing stability analyses like Jaccard and Silhouette coefficients after every round. This dynamic adaptation over time enhances the clustering accuracy, allowing the algorithm to capture nuanced changes and trends. Ultimately, the adaptive clusters facilitate insightful data interpretation, revealing temporal patterns that offer deeper insights into the charging station network's behavior and usage dynamics. This adaptive K-means mechanism serves as a robust solution for harnessing temporal information, unveiling valuable knowledge embedded within the network.

In our proposed method, we suggest that the temporal usage similarity of CSs is more critical than spatial locations of CSs as in previous work [8] for FL training. It highlights the problem that nearby CSs could have vastly different usage patterns. For instance, a CS close to a transit hub or shopping mall can see significantly more demand than other CSs in the area. In the existing literature, different distance measuring techniques are used for similarity calculation for clustering, such as Euclidean distance, dynamic time warping (DTW), and cosine similarity. Since our approach is adaptive clustering, we have introduced the Silhouette coefficient measurement in every round of FL to adjust the similarity between the clients in the cluster. We have compared the existing methods with our approach in Section IV.

2) Adaptive Clustering Model: Cluster stability analysis is greatly influenced by the data set, particularly by how clearly distinct and homogeneous the clusters are. The majority of clustering techniques rely on a particular cluster model or prototype, which may work for some types of data but not for others. Analyzing the stability and consistency of cluster assignments becomes crucial in time-variant environments where data distributions across clients change over time. In this work, we propose the Silhouette coefficient that measures the stability and adjusts the cluster memberships of the charging stations when the data distributions change significantly in rounds.

The term *Silhouette coefficient*, which incorporates the terms individual silhouette coefficient and cluster silhouette coefficient, was proposed by Kaufman et al. [11]. The individual silhouette coefficient is expressed as

$$a(i) = \frac{1}{n_c - 1} \sum_{i \neq j, j \in C_c} d(i,j), \tag{4}$$

$$b(i) = \arg\min_{i,j} \frac{1}{n_c} \sum_{i \in C_j} d(i,j), \tag{5}$$

$$S_i = \frac{b(i) - a(i)}{max[a(i), b(i)]}, \tag{6}$$

where $a(i)$ is the average distance between data point i and all other points in the same cluster, whereas $b(i)$ is the minimal average distance between point i and every other point in any cluster that does not contain the data point i. Finally, the silhouette coefficient for clusters is calculated as

$$S_c = \frac{1}{n} \sum_{i=1}^{n} S_i, \tag{7}$$

where n is the number of data points in the dataset. The Silhouette coefficient for the cluster efficiency is calculated after every round of federated clustering, and if the Silhouette score is less than 0.5, then relocation of the charging stations is processed for the next round. This makes the stability of clusters according to the varying data distribution across the stations.

Since the data distribution of clients changes after each training cycle, it might affect the cluster membership to which those clients belong to. In addition to increasing the cluster stability throughout federated learning rounds, it makes sure that the changing patterns of energy consumption in the charging stations are consistently and precisely recorded, enabling the best possible resource allocation and decision-making. Based on the temporal behavior change, the algorithm learns the model parameters by calculating the S_c to perform the cluster assignment adaptively. This approach delves into the varying hidden parameters of the model and, instead of the obvious model information, recomputes the centroids and assigns new cluster labels to the clients. The approach is, to the best of our knowledge, unexplored in the existing literature, where the cluster memberships of the clients are fixed throughout the training. The whole flow of our Approach is described in Algorithm 1.

Algorithm 1 Temporal-based Adaptive CFL

Input: $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N\}$, $\mathcal{K}, T, \epsilon$.
Output: Updated global model ω^t.
Server:
Initialization: Set $\mathcal{K}$ Clusters to the clients L_{init}
ω_{k} models and hyper parameters
Adaptivity $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$
for $t = 1$ to T **do**:
Client Update:
  for $\mathcal{D}_i$ in $\mathcal{D}$ **do**:
    Compute $\alpha_i = \frac{\alpha_0}{1+\delta t}$, ; δ is a decay constant.
    Update ω using equation 2 and update the model:

$$\omega^{\text{t+1}} \leftarrow \omega^t - \alpha_i \nabla F(\omega; \mathcal{D}_i)$$

Adaptive Clustering:
    Compute

$$S_i = \frac{b_i - a_i}{\arg\max(a_i, b_i)}$$

  if $S_i < \epsilon$ **then**
    Update Θ and assign L_{new}
    Update L_{init} with L_{new}
    Update ω

IV. Experimental Setup and Results

A. Dataset

We evaluated our temporal-based Adaptive clustering approach on a real-world dataset of EV charging stations network datasets [12], Boulder City, USA. EV charging stations Network, which comprises 39 charging stations across 4 different postal codes of the city. Setting up a federated learning experiment for the EV charging stations network in Boulder City involves creating a distributed machine learning system that allows multiple charging stations to train a model collaboratively without sharing their raw data. The data include information about charging patterns, energy consumption, charging station ID, charging start and end times, transaction history, and other relevant features.

B. Data Preprocessing

The dataset comprises a variety of features, which include station ID, Charging start and end times, location information, and the transaction ID of each charging station, aiming to predict the energy consumption in each session better. Then, we calculated the daily, weekly, and monthly energy consumption of every individual charging station for the trend and seasonality analysis. The dataset lacks samples for all the stations; thus, only 10 clients have been chosen based on the two years (2019-2020) available data for our setup. Data preprocessing has been done, including data normalization and feature engineering, and for time series clustering, the lookback approach involved dependency of the next prediction on historical trends.

C. Experiment Setup

The PyTorch framework is used to simulate the LSTM model in a federated learning approach. The LSTM model has one LSTM input layer and a fully connected linear layer with 10 neurons ($n = 10$) in each layer. The hyper-parameters are set as follows: $T = 500$, $E = 2$, lookback equal to 5, $\epsilon = 0.5$, $m = 10$, and Adam optimizer have been used in evaluating the LSTM model. The model is evaluated in each round, and the loss of each cluster is also calculated based on the client's

TABLE II
RMSE COMPARISON OF TACFL AND CONVENTIONAL METHODS

Clusters	Saputra et al. [8]	Perry et al. [9]	**TACFL (Proposed)**
C_1	1.108	1.093	**0.954**
C_2	1.147	1.082	**0.935**
C_3	1.28	1.119	**0.926**
Average	1.1783	1.09	**0.938**

model in that cluster. We utilize the root mean squared error (RMSE) to demonstrate prediction accuracy, i.e., prediction error, because we are dealing with energy demand prediction, which is classified as a regression prediction model, i.e., when the mapping function gives continuous prediction outputs. In each round, the epsilon value of clients is calculated based on the given above equation 6. The client remains in the same cluster when the epsilon value is within the threshold constraints, otherwise, it will be ready to be reclustered in the preceding round. Such an approach is effective as the client adapts to the cluster based on variance within the cluster and across the cluster.

D. Results

Adaptive approach is effective in the federated clustering approach, as the clients adapt to the cluster with an increasing number of rounds. In each round, the client is evaluated to the relevant cluster membership based on the epsilon values. Figure 2 better illustrates the adaptive cluster membership of clients involved in federated training. From the figure, we can observe that clients change their cluster membership in rounds and try to find their optimal cluster centroid while likely optimizing the global cost function.

Comparative analysis has also been carried out with the state-of-the-art clustering approaches in EV charging station energy demand prediction. The TACFL approach was compared with Euclidean distance-based K-means and DTW method-based K-means clustering approaches adopted in [8] and [9] on the same LSTM predictive model and parameters. Results show that the TACFL approach gives better results than those methods, as shown in Table II. The aggregated RMSE of TACFL is 0.938, while Euclidean-based and DTW-based approaches are 1.1783 and 1.09, respectively. The third cluster in TAFCL has shown better performance, while for DTW, 2nd cluster shows less loss, i.e., 1.082. The cluster-based approaches are also better for personalization, thus our approach is well-defined for adaptive federated personalized learning.

Similarly, to evaluate the communication efficiency of the TACFL model with baseline federated learning (FedAvg), different combinations of clusters have been analyzed in Figure 3. It can be seen that the TACFL with three clusters is giving the best results both on accuracy and communication rounds, outclassing other scenarios and vanilla FL. It is observed that the TACFL clustering approach converges faster during the initial rounds, but an increasing number of clusters have poor performance after adequate global rounds. FedAvg gives similar trends to TACFL, having 2 clusters after some communication rounds, and adapting becomes easy as there are only two clusters. Still, the performance is not better than the optimal cluster number. The optimal cluster has been evaluated by manually cluster number parameter; an approach to determine the optimal number of clusters will be considered as future work. Overall, the TACFL is effective for CFL approaches as it reduces the global rounds, making it suitable for FL having communication cost bottlenecks.

In Table II, the RMSE values with different conventional model distance criteria are compared with the TACFL approach for clustering accuracy measurements. TACFL outperforms all in the case of individual cluster losses and the overall average loss. When the number of clusters is increased, the RMSE loss decreases in TACFL, while in conventional methods, the trend is not well followed, which shows inefficient cluster stability. The average RMSE is calculated for the TACFL and previous works and plotted in Figure 4, which shows that the TACFL has a loss of 0.938 which is comparatively better than the previous works, whose average RMSE values are 1.178 and 1.09.

V. CONCLUSION

Time series regression is a task that requires a limited dataset (transactions) and computer resources to predict energy consumption for customers based on a temporal series of prior usage patterns. Recognizing the patterns and clustering them needs keen observation and accurate model design. In this work, we have introduced a temporal-based adaptive clustering mechanism to cluster the heterogeneous clients based on learning the time-varying data distribution and assigning them the most accurate clustering labels. Compared to conventional clustering algorithms, where the client's positions in the clusters are fixed across the whole training and learning process, TACFL introduces the cluster stability calculations by leveraging the Silhouette coefficient and introduces adaptivity in the conventional CFL methods. Regarding prediction accuracy and adapting client memberships in clusters, TACFL has shown great power with fast convergence. Cluster stability analysis like the Jaccard coefficient and rand statistics for the conventional and TACFL are expected to be carried out in future work with more real-world applications.

ACKNOWLEDGMENT

This research was supported by the MSIT Korea under the NRF Korea (NRF-2022R1A2C4001270) and the Information Technology Research Center (ITRC) support program (IITP-2022-2020-0-01602) supervised by the IITP.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.

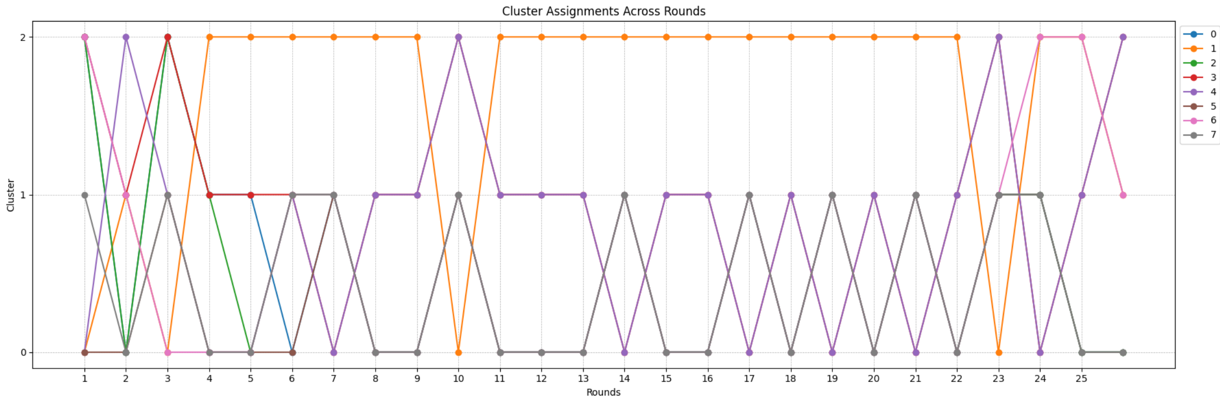

Fig. 2. Adaptive cluster assignments across rounds

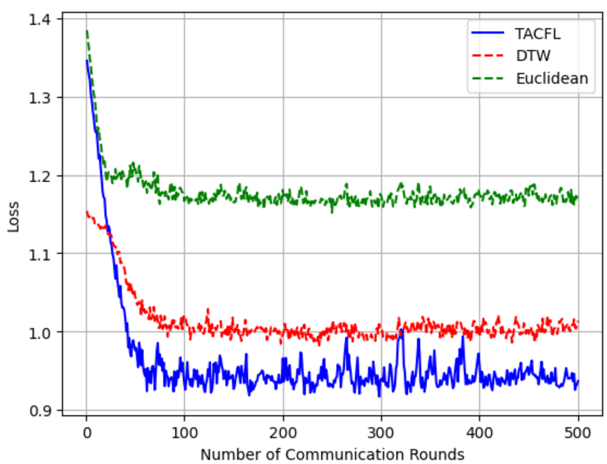

Fig. 3. Comparison of RMSE of TACFL and existing approaches

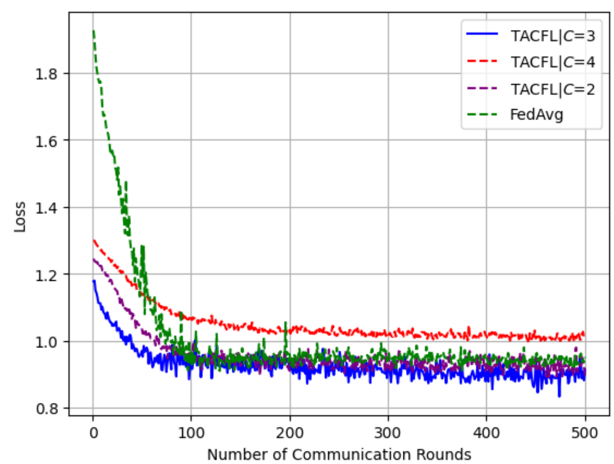

Fig. 4. RMSE of TACFL with varying number of clusters ($C = 2, 3, 4$) and FedAvg

[2] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, "Federated learning with personalization layers," *arXiv preprint arXiv:1912.00818*, 2019.

[3] A. Hilmkil, S. Callh, M. Barbieri, L. R. Sütfeld, E. L. Zec, and O. Mogren, "Scaling federated learning for fine-tuning of large language models," in *International Conference on Applications of Natural Language to Information Systems*, pp. 15–23, Springer, 2021.

[4] G. Cheng, K. Chadha, and J. Duchi, "Fine-tuning is fine in federated learning," *arXiv preprint arXiv:2108.07313*, vol. 3, 2021.

[5] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[6] B. Gong, T. Xing, Z. Liu, W. Xi, and X. Chen, "Adaptive client clustering for efficient federated learning over non-iid and imbalanced data," *IEEE Transactions on Big Data*, 2022.

[7] N. Wang, R. Zhou, L. Su, G. Fang, and Z. Li, "Adaptive clustered federated learning for clients with time-varying interests," in *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*, pp. 1–10, 2022.

[8] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, M. D. Mueck, and S. Srikanteswara, "Energy demand prediction with federated learning for electric vehicle networks," in *2019 IEEE global communications conference (GLOBECOM)*, pp. 1–6, IEEE, 2019.

[9] D. Perry, N. Wang, and S.-S. Ho, "Energy demand prediction with optimized clustering-based federated learning," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2021.

[10] Y. L. Tun, K. Thar, C. M. Thwal, and C. S. Hong, "Federated learning based energy demand prediction with clustered aggregation," in *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 164–167, IEEE, 2021.

[11] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.

[12] BoulderCO, "City of Boulder Open Data." accessed 2023-10-01.