A Personalized Federated Learning Algorithm for One-Class Support Vector Machine: an Application in Anomaly Detection

Ali Anaissi*[0000−0002−8864−0314], Basem Suleiman[0000−0003−2674−0253], and Widad Alyassine

School of Computer Science, The University of Sydney, Australia

Abstract. Federated Learning (FL) has recently emerged as a promising method that employs a distributed learning model structure to overcome data privacy and transmission issues paused by central machine learning models. In FL, datasets collected from different devices or sensors are used to train local models (clients) each of which shares its learning with a centralized model (server). However, this distributed learning approach presents unique learning challenges as the data used at local clients can be non-IID (Independent and Identically Distributed) and statistically diverse which decrease learning accuracy in the central model. In this paper, we overcome this problem by proposing a novel personalized federated learning method based One-Class Support Vector Machine (FedP-OCSVM) to personalize the resulting support vectors at each client. Our experimental validation showed that our FedP-OCSVM precisely constructed generalized clients' models and thus achieved higher accuracy compared to other state-of-the-art methods.

1 Introduction

The emerging Federated Learning (FL) concept was initially proposed by Google for improving security and preventing data leakages in distributed environments [9]. FL allows the central machine learning model to build its learning from a broad range of data sets located at different locations. This innovative machine learning approach can train a centralized model on data generated and located on multiple clients without compromising the privacy and security of the collected data. Also, it does not require transmitting large amount of data which can be a major performance challenge especially for real-time applications. A good application of FL is in the civil infrastructures domain specifically in Structural Health Monitoring (SHM) applications where smart sensors are utilized to continuously monitor the health status of complex structures such as bridges to generate actionable insights such as damage detection.

This motivates for developing a more intelligent model that utilizes the centralized learning model but without the need to transmit the frequently-measured data to one central model for processing unit. In this sense, we propose a federated learning approach for anomaly detection using One-class support vector machine (OCSVM)[12] which has been widely applied to anomaly

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31

detection and become more popular in recent years [2–4]. OCSVM has been successfully applied in many application domains such as civil engineer, biomedical, and networking [1, 8], and produced promising results. Although our approach results in reducing data transmission and improving data security, it also raises significant challenges in how to deal with non-IID (Independent and Identically Distributed) data distribution and statistical diversity. Therefore, to address the non-IID challenge in our proposed FL approach, we developed a novel method to personalize the resulting support vectors from the FL process. The rationale idea of personalizing support vectors is to leverage the central model in optimizing the clients' models not only by using FL, but also by personalizing it w.r.t its local data distribution. The contribution of the work in this study is twofold.

- A novel method of learning OCVSM model in FL settings.
- A novel method to personalize the resulting support vectors to addresses the problem of non-IID distribution of data in FL.

2 Related Work

Federated Learning (FL) has gained a lot of interest in recent years and as a result, it has attracted AI researchers as a new and promising machine learning approaches. This FL approach attracts several well-suited practical problems and application areas due to its intrinsic settings where data needs to be decentralized and privacy to be preserved. For instance, McMahan *et al.*[11] proposed the first FL-based algorithm named *FedAvg*. It uses the local Stochastic Gradient Descent (SGD) updates to build a global model by taking average model coefficients from a subset of clients with non-IID data. This algorithm is controlled by three key parameters: C, the proportion of clients that are selected to perform computation on each round; E, the number of training passes each client makes over its local dataset on each round; and B, the local mini-batch size used for the client updates. Selected clients perform SGD locally for E epochs with mini-batch size B. Any clients which, at the start of the update round, have not completed E epochs (stragglers), will simply not be considered during aggregation. Subsequently, Li *et al.*[10] introduced the *FedProx* algorithm, which is similar to FedAvg. However, FedProx makes two simple yet critical modifications that demonstrated performance improvements. FedProx would still consider stragglers (clients which have not completed E epochs at aggregation time) and it adds a *proximal term* to the objective function to address the issue of statistical heterogeneity. Similarly, Manoj *et al.*[6] addressed the effects of statistical heterogeneity problem using a *personalization-based approach (FedPer)*. In their approach, a model is viewed as base besides personalization layers. The base layers will be aggregated as in the standard FL approach with any aggregation function, whereas the personalized layers will not be aggregated.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31

3 Personalized Federated Learning for OCSVM: FedP-OCSVM

3.1 OCSVM-FedAvg

In FL setting, learning is modeled as a set of C clients and a central server S, where each client learns based on its local data, and is connected to S for solving the following problem:

$$\min_{w \in \mathbb{R}^d} f(w) := \frac{1}{C} \sum_{c=1}^{C} f_c(w_c) \tag{1}$$

where f_c is the loss function corresponding to a client c that is defined as follows:

$$f_c(w_c) := \mathbb{E}[\mathcal{L}_c(w_c; x_{c,i})] \tag{2}$$

where $\mathcal{L}_c(w_c; x_{c,i}$ measures the error of the model w_c (e.g. OCSVM) given the input x_i. The Sequential Minimal Optimization (SMO) is often used in the support vector machine. However, in the case of the nonlinear kernel model as in OCSVM, SMO does not suit the FL settings well. Therefore, we propose a new method for solving the OCSVM problem in FL setting using the SGD algorithm.

The SGD method solves the above problem defined in Equation 2 by repeatedly updating w to minimize $\mathcal{L}(w; x_i)$. It starts with some initial value of $w^{(t)}$ and then repeatedly performs the update as follows:

$$w^{(t+1)} := w^{(t)} + \eta \frac{\partial \mathcal{L}}{\partial w}(x_i^{(t)}, w^{(t)}) \tag{3}$$

In fact, the SGD algorithm in OCSVM focuses on optimizing the Lagrange multiplier $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]$ for all patterns x_i where $x_i : i \in [n], \alpha_i > 0$ are called support vectors. Thus, exchanging gradient updates in FL for averaging purposes is not applicable. Consequently, we modified the training process of SGD to share the coefficients of the features in the kernel space under the constraints of sharing an equal number of samples across each client C. In this sense, our SGD training process computes the kernel matrix $K = \phi(x_i, x_j)_{i,j=1,...,n}$ before looping through the samples. Then it computes the coefficients w after performing a number of epochs as follows:

$$w^{(t+1)} = \alpha K; \tag{4}$$

$$s.t \qquad \alpha = \alpha + \eta(1 - \sum_{i=1}^{n} w)$$

Each client performs a number of E epochs at each round to compute the gradient of the loss over its local data and to send the model parameters w^{t+1}

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31

to the central server S along with their local loss. The server then aggregates the gradients of the clients and applies the global model parameters update by computing the average value of all the selected clients model's parameters as follows:

$$w^{(t+1)} := \frac{1}{C} \sum_{i=1}^{C} w_C^{(t+1)}; \tag{5}$$

where C is the number of selected clients.

The server then share the $w^{(t+1)}$ to all selected clients in which each one performs another iteration to update $w^{(t+1)}$ but with setting $w_i^{(t)} = w^{(t+1)}$ as defined in the traditional FedAvg method.

3.2 Personalized Support Vectors

Our proposed approach may work well when clients have similar IID data. However, it is unrealistic to assume that since data may come from different environments or contexts in FL settings, thus it can have non-IID. Therefore, it is essential to decouple our model optimization from the global model learning in a bi-level problem depicted for personalized FL so the global model optimization is embedded within the local (personalized) models. Geometrically, the global model can be considered as a "*central point*", where all clients agree to meet, and the personalized models are the points in different directions that clients follow according to their heterogeneous data distributions. In this context, once the learning process by the central model is converged and the support vectors are identified for each client, we perform a personalized step to optimize the support vectors on each client. Intuitively, to generate a personalized client model, its support vectors must reside on the boundaries of the local training data (i.e. edged support vector). Thus, we propose a new algorithm to inspect the spatial locations of the selected support vector samples in the context of the FL settings explained above. It is intuitive that an edge support vector x_e will have all or most of its neighbors located at one side of a hyper-plane passing through x_e. Therefore, our edge pattern selection method constructs a tangent plane for each selected support vector $x_i : i \in [n], \alpha_i > 0$ with its k-nearest neighbors data points. The method initially selects the k-nearest data points to each support vector x_s, and then centralizes it around x_s by computing the norm vector v_i^n of the tangent plane at x_s . If all or most of the vectors are located at one side of the tangent plane), we consider x_s as an edge support vector denoted by x_e, otherwise, it is considered as an interior support vector and it is excluded from the selected original set of support vectors.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31

4 Experimental Results and Discussions

We validate our FedP-OCSVM method based on a real dataset collected from a Cable-Stayed Bridge in Australia [1] to detect potential damage. In all experiments, we used the default value of the Gaussian kernel parameter σ and $\nu = 0.05$.

We instrumented the Cable-Stayed Bridge with 24 uni-axial accelerometers and 28 strain gauges. We used accelerations data collected from sensors Ai with $i \in \{1, 2, \ldots, 24\}$. Figure 1 shows the locations of these 24 sensors on the bridge deck. Each set of sensors on the bridge along with one line (e.g A1: A4) is connected to one client node and fused in a tensor node $\mathcal{T}$ to represent one client in our FL network, which results in six tensor nodes $\mathcal{T}$ (clients).

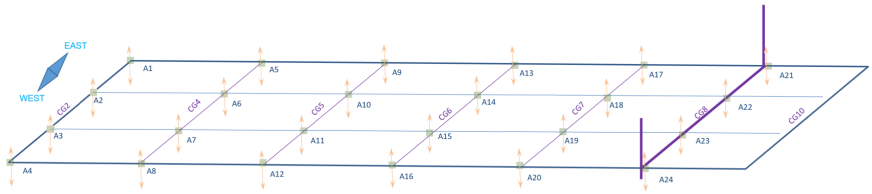


Fig. 1: The locations on the bridge's deck of the 24 Ai accelerometers used in this study. The cross girder j of the bridge is displayed as CGj [5].

This experiment generates 262 samples (a.k.a events) each of which consists of acceleration data for 2 seconds at a sampling rate of 600 Hz. We separated the 262 data instances into two groups, 125 samples related to the healthy state and 137 samples for the damage state.

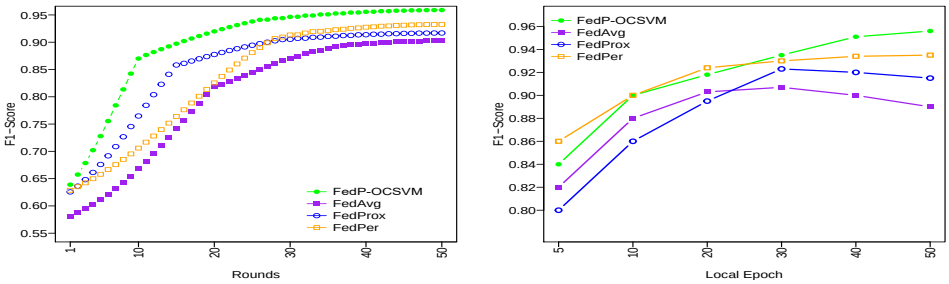
For each reading of the uni-axial accelerometer, we normalized its magnitude to have a zero mean and one standard deviation. The fast Fourier transform (FFT) is then used to represent the generated data in the frequency domain. Each event now has a feature vector of 600 attributes representing its frequencies. The resultant data at each sensor node $\mathcal{T}$ has a structure of 4 sensors $\times$ 600 features $\times$ 262 events.

We randomly selected 80% of the healthy events (100 samples) from each tensor node $\mathcal{T}$ for training multi-way of $\mathcal{X} \in \mathbb{R}^{4\times600\times100}$ (i.e. *training* set). The 137 examples related to the two damage cases were added to the remaining 20% of the healthy data to form a *testing* set, which was later used for the model evaluation.

We initially study the effect of the number of local training epochs E on the performance of the four experimented federated learning methods as suggested in previous works [11, 7]. The candidate local epochs we consider are

[1] The two bridges are operational and the companies which monitor them requested to keep the bridge name and the collected data about its health confidential.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31



(a) The effect on the number of commuication rounds.

(b) The effect of number of local training epochs.

Fig. 2: Convergence rates of various methods in federated learning applied on Cable-Stayed Bridge with $\mathcal{T} = 6$ clients.

Table 1: *F1-score* of various methods.

	FedP-OCSVM	FedProx	FedPer	FedAvg
Cable-Stayed Bridge	0.96±0.02	0.92±0.01	0.93±0.03	0.90±0.04

$E \in \{5, 10, 20, 30, 40, 50\}$. For each of the candidate E, we run all the methods for 40 rounds and report the final *F1-score* accuracy generated by each method. The result is shown in Figure 2(b). We observe that conducting longer epochs on the clients improves the performance of FedP-OCSVM and FedPer, but it slightly deteriorates the performance of FedProx and FedAvg. The second experiment was to compare our method to FedAvg, FedPer and FedProx in terms of accuracy and the number of communication rounds needed for the global model to achieve good performance on the test data. We set the total number of epochs E for FedP-OCSVM and FedPer to 50, and 30 for FedProx and FedAvg as determined by the first experimental study related to the local training epochs E. The results showed that FedP-OCSVM outperforms FedAvg, FedProx and FedPer in terms of local training models and performance accuracy. Table 1 shows the accuracy results of all experiments using *F1-score*. Although no data from the damaged state has been employed to construct the central model, each personalized local client model was able to identify the damage events related to "Car-Damage" and "Bus-Damage" with an average *F1-score* accuracy of 0.96 ± 0.02.

5 Conclusions

In this paper, we present a novel machine learning approach for an effective and efficient anomaly detection model in such applications like SHM systems that

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31

require information derived from many spatially-distributed locations throughout large infrastructure covering various points in the monitored structure. Our method employs a Federated Learning (FL) approach to OCSVM as an anomaly detection model augmented with a method to personalize the resulting support vectors from the FL process. Our experimental evaluation on a real bridge structure dataset showed promising damage detection accuracy by considering different damage scenarios. In the "Cable-Stayed Bridge" dataset, our FedP-OCSVM method achieved an accuracy of 96%. The experimental results of this case study demonstrated the capability of our FL-based damage detection approach with the personalization algorithm to improve the damage detection accuracy.

References

1. Anaissi, A., Goyal, M., Catchpoole, D.R., Braytee, A., Kennedy, P.J.: Ensemble feature learning of genomic data using support vector machine. PloS one **11**(6), e0157330 (2016)
2. Anaissi, A., Khoa, N.L.D., Mustapha, S., Alamdari, M.M., Braytee, A., Wang, Y., Chen, F.: Adaptive one-class support vector machine for damage detection in structural health monitoring. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 42–57. Springer (2017)
3. Anaissi, A., Khoa, N.L.D., Rakotoarivelo, T., Alamdari, M.M., Wang, Y.: Self-advised incremental one-class support vector machines: An application in structural health monitoring. In: International Conference on Neural Information Processing. pp. 484–496. Springer (2017)
4. Anaissi, A., Khoa, N.L.D., Rakotoarivelo, T., Alamdari, M.M., Wang, Y.: Adaptive online one-class support vector machines with applications in structural health monitoring. ACM Transactions on Intelligent Systems and Technology (TIST) **9**(6), 1–20 (2018)
5. Anaissi, A., Makki Alamdari, M., Rakotoarivelo, T., Khoa, N.: A tensor-based structural damage identification and severity assessment. Sensors **18**(1), 111 (2018)
6. Arivazhagan, M.G., Aggarwal, V., Singh, A.K., Choudhary, S.: Federated learning with personalization layers. arXiv preprint arXiv:1912.00818 (2019)
7. Chen, F., Luo, M., Dong, Z., Li, Z., He, X.: Federated meta-learning with fast convergence and efficient communication. arXiv preprint arXiv:1802.07876 (2018)
8. Khoa, N.L.D., Anaissi, A., Wang, Y.: Smart infrastructure maintenance using incremental tensor analysis. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 959–967. ACM (2017)
9. Konečnỳ, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 (2016)
10. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. arXiv preprint arXiv:1812.06127 (2018)
11. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics. pp. 1273–1282. PMLR (2017)
12. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural computation **13**(7), 1443–1471 (2001)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08760-8_31