ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Federated learning vector quantization for dealing with drift between nodes

Valerie Vaquet[1*], Fabian Hinder[1*], Johannes Brinkrolf[1*], Patrick Menz[2], Udo Seiffert[3] and Barbara Hammer[1 †]

1- Machine Learning Group
Bielefeld University, Bielefeld - Germany

2- Cognitive Processes and Systems, Fraunhofer Institute of Factory Operation and Automation (IFF), Magdeburg - Germany

3- Compolytics GmbH, Barleben - Germany

Abstract. Federated learning is an efficient methodology to reduce the data transmissions to the server when working with large amounts of (sensor) data from diverse physical locations. When using data from different sensor devices concept drift between the single sensors poses an additional challenge. In this contribution we define a formal framework for federated learning with concept drift and propose a version of federated LVQ dealing with concept drift induced by different hyperspectral cameras. We evaluate this approach experimentally and demonstrate its robustness to class imbalance and missing classes.

1 Introduction

The application of machine learning systems proved successful in many areas in recent years. Especially, when coupling these algorithms with great amounts of data collected by sensors, processes can be either fully automated or improved as humans would not be able to interpret this extent of sensor data. One particular relevant type of sensor technology is hyperspectral imaging [1, 2]. It is measuring the reflectance of all kinds of substances in a wider range than humans can perceive. As this offers the opportunity to observe patterns which cannot be observed by humans, hyperspectral sensing is frequently used in quality control in food production and pharmaceutical applications, precision agriculture, environmental analysis, water resource management, medical diagnosis, and artwork and forensic document analysis [2, 3]. While the technology is promising in many areas, one core challenge is a deficiency in inter-operability, e. g. even if data is collected by sensors of the same model by the same manufacturer, machine learning models do not necessarily transfer from one to another without a serious decrease in accuracy. This especially hinders the broad usage of this sensor technology since updating or retraining models for new sensors is a cost and time consuming task, as new ground truth measurements need to be performed. Recent work [4] analyzed and categorized the sensor shifts in hyperspectral imaging data and proposed an efficient yet simple method to eliminate those given a balanced class distribution at each sensor node.

*Authors contributed equally.

†Funding in the frame of the BMBF project TiM, 05M20PBA and 05M20AFA, and BMWi project KI-Marktplatz, 01MK20007E is gratefully acknowledged.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Another more general challenge when working with extensive amounts of sensor data which are gathered at different places or institutions is that great loads of data need to be shared to create a model containing the information from all sensors. In recent years considerable research was conducted on federated learning: Instead of sharing the data, locally inferred models are shared. This way, the required communication bandwidth can be reduced and privacy increased. Yet, one challenge of such schemes is that it cannot be guaranteed that data are i.i.d. at the different nodes. This phenomenon is referred to as concept drift in between the nodes.

In this work, we adapt federated learning principles to manage concept drift: first, we propose a general framework for federated learning vector quantization (fedLVQ) with concept drift. This extends the formalism presented in [5] such that it also corrects drift when fusing the local models. We evaluate the framework in a setting where hyperspectral sensor data is collected by different sensors. In particular, we consider class imbalances and missing classes across the sensor nodes and show that the proposed method outperforms regular fedLVQ as well as a shift reducing preprocessing [4].

2 Formal Framework for Federated Learning with Concept Drift

Formally, federated learning constitutes a two-staged approach [6]: First models h_{θ_i}, parameterized by $\theta_i \in \Theta$, are trained locally for each dataset $S_1, \dots, S_n \in \cup_{N=1}^{\infty} (\mathcal{X} \times \mathcal{Y})^N$. Then these models are combined to obtain a global model h_θ with $\theta = C(\theta_1, \dots, \theta_n)$ using a fusion algorithm C. Only the parameters θ_i need to be communicated, which is beneficial if restrictions on available bandwidth or privacy concerns hold. Further, local training can easily be computed in parallel. A learning algorithm A allows an exact federated learning scheme if there exists a fusion algorithm $C : \cup_{n=1}^{\infty} \Theta^n \to \Theta$ such that the combination of trained models is the same as training on combined data, i.e. the following diagram commutes

$$\begin{array}{ccc} (S_i)_{i=1}^n & \longrightarrow & S_1 \times \cdots \times S_n \\ {\scriptstyle \prod_i A}\downarrow & & \downarrow{\scriptstyle A} \\ (\theta_i)_{i=1}^n & \xrightarrow[C]{} & \theta \end{array}$$

If we can combine the models in a reasonable way, we say that $S_1, \dots, S_n$ are *compatible*. One challenge is given if the data sets are incompatible, because their underlying distributions $P_i(X)$ or their posterior $P_i(Y|X)$ are very distinct, such that it is not clear how to combine the model parameters θ_i in a reasonable way. We refer to setups where $P_i(Y|X) \neq P_j(Y|X)$ as *concept drift* (or drift for short).

We suggest to counter such effects by applying transformations $T_1, \dots, T_n$ to the models, which remove the effect of the drift and allow fusing of otherwise incompatible models. The fused model for the i-th node can then be computed by applying the inverse transformation to the fused transformed model, i.e.

$$T_i^{-1} \circ C(T_1(\theta_1), \dots, T_n(\theta_n)).$$

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Parameterizing the transformation $T : \Psi \times \Theta \to \Theta$, i.e. $T_i = T_{\psi_i}$, we end up with the following learning problem for the transformations, given a model loss $\mathcal{L}$:

$$\arg\min_{\psi_i \in \Psi} \sum_{i=1}^{n} \frac{|S_i|}{\sum_{j=1}^{n} |S_j|} \mathcal{L}_{S_i} \left(T_{\psi_i}^{-1} \circ C(T_{\psi_1}(\theta_1), \ldots, T_{\psi_n}(\theta_n)) \right).$$

We refer to datasets as Ψ-*compatible* if there exist $\psi_1, \ldots, \psi_n$ that make the models compatible. Further, we refer to them as Ψ-*unique*, if those transformations ψ_i are uniquely determined (modulo global operations).

3 Federated Learning Vector Quantization for Hyperspectral Data

Learning Vector Quantization (LVQ) models [7] constitute a prominent class of (multi-class)-classification models on $\mathbb{R}^d$ that are particularly well compatible with horizontal federated learning [5], i.e., all features but not necessarily all classes are available. A LVQ-model h is parameterized by a set of w labeled prototypes $\mathcal{W} = \{(\boldsymbol{w}_j, c(\boldsymbol{w}_j)) \in \mathbb{R}^d \times \{1, \ldots, c\} \mid j \in \{1, \ldots, w\}\}$, which induce a winner takes all classification, i.e., $h(\boldsymbol{x}) = c(\boldsymbol{w}_l)$ with $l = \arg\min_{j \in \{1, \ldots, c\}} d(\boldsymbol{w}_j, \boldsymbol{x})$. A generalization is obtained by combining the classical LVQ with (local) *metric learning*, leading to *Local generalized matrix LVQ* (LGMLVQ), which uses a distinct matrix Λ_j for each prototype $\boldsymbol{w}_j$ which induces prototype specific distance function: $d_{\Lambda_j}(\boldsymbol{x}, \boldsymbol{w}_j) = (\boldsymbol{x} - \boldsymbol{w}_j)^T \Lambda_j (\boldsymbol{x} - \boldsymbol{w}_j)$. Positive semi-definiteness of Λ_j is guaranteed by choosing $\Lambda_j = {\Omega_j}^T \Omega_j$ [7, 8].

Assuming that we use exactly one prototype per class, i.e. $c(\boldsymbol{w}_k) = k$, and that there is no real drift, i.e., $P_i(Y \mid X) = P_j(Y \mid X)$, combining LVQ models becomes particularly easy [5]: Denoting by $\boldsymbol{w}_k^{(i)}$ the prototype of class k used by the model trained on S_i we obtain the fused prototypes as the weighted mean:

$$\boldsymbol{w}_k = \sum_{i=1}^{n} \frac{|S_i^{(k)}|}{\sum_{j=1}^{n} |S_j^{(k)}|} \boldsymbol{w}_k^{(i)}, \tag{1}$$

here $S_i^{(k)} \subset S_i$ is the sub-dataset of all datapoints of class k. Likewise, for LGMLVQ we obtain the fused relevance matrix as

$$\Lambda_k = \sum_{i=1}^{m} \frac{|S_i^{(k)}|}{\sum_{j=1}^{m} |S_j^{(k)}|} {\Omega_k^{(i)}}^T \Omega_k^{(i)} \quad \forall\, k \in \{1, \ldots, c\}. \tag{2}$$

Notice that this corresponds to the weighted mean of the local distance functions of the models. A computationally efficient realization to construct the fused model relies on Schur's Algorithm which allows us to find $\Lambda_j = {\Omega_j}^T \Omega_j$.

To deal with drift between the separate datasets we rely on a suitable transformation: We restrict to the case discussed in [4], namely hyperspectral data, where it is reported to be sufficient to consider translations of the input data. Thus, we choose $\Psi = \mathbb{R}^d$ and $T_v((\boldsymbol{w}_i, \Omega_i)_{i=1}^c) = (\boldsymbol{w}_i + v, \Omega_i)_{i=1}^c$. Further, we assume that there is one prototype per class. In this case Ψ-consistency can be characterized as follows:

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Theorem 1. *Let G be the graph with a node for each dataset S_i and an edge $(i,j) \in E(G)$ if and only if S_i and S_j share a class. The following holds: (i) Assume class wise relevance matrices, i.e., $\Omega_k^{(i)} = \Omega_k^{(j)}$ for all nodes i, j and prototypes k. If for every edge $(i,j) \in E(G)$ there exists vectors v_{ij} such that $\boldsymbol{w}_k^{(j)} - \boldsymbol{w}_k^{(i)} = v_{ij}$ all classes k observed in S_i and S_j, and the sum along every cycle $C \subset G$ vanishes, i.e. $\sum_{i=1}^{|C|-1} v_{C_i C_{i+1}} = -v_{C_{|C|} C_1}$, then there exists a global model $(\boldsymbol{w}_k, \Omega_k)_k$ and shifts $v_1, \ldots, v_n \in \mathbb{R}^d$ such that $\boldsymbol{w}_k^{(i)} = \boldsymbol{w}_k + v_i$ for all i and k. (ii) If G is connected, the global model (assuming it exists) and the shifts are uniquely determined up to a single, global shift. (iii) If G is connected, the global model (assuming it exists) can be computed based on any spanning tree (up to global shift). If G is a tree, the transfer v_i can be computed by choosing a root with $v_{\text{root}} = 0$ and then computing the remaining v_i along the paths.*

We omit the proof due to space limitations. The theorem yields an algorithmic solution as follows: first, we build the graph G and compute a spanning tree $T_G \subset G$. For each node i and each class k, we compute the mean vector $\mu_{i,k} = \mu(S_i^{(k)})$. If distributions are shifted versions of each other, i.e. parameters are Ψ-compatible, mean values provide sufficient information for a full model [9]. Similar to the data, means are shifted in accordance to the distribution shift, i.e., $\mu_{i,k} - \mu_{j,k} = v_i - v_j$ for all i, j, k. Thus, if we choose a root $r \in T_G$ and set $v_r = 0$ we can compute the v_i inductively, v_j denoting its parent

$$v_i = \frac{\sum_{k=1}^{N} \mathbf{1}_{\min\{|S_i^{(k)}|, |S_j^{(k)}|\} \geq 1} (\mu_{i,k} - \mu_{j,k})}{\sum_{k=1}^{N} \mathbf{1}_{\min\{|S_i^{(k)}|, |S_j^{(k)}|\} \geq 1}} + v_j. \tag{3}$$

Notice that this procedure also compensates for class imbalances: If $|S_i^{(k)}|/|S_i| \gg |S_j^{(k)}|/|S_j|$ then the influence of μ_{ik} on $\mu(S_i)$ is much stronger than the influence of μ_{jk} on $\mu(S_j)$ so that $\mu(S_i) - \mu(S_j) \neq \mu_{ik} - \mu_{jk}$. On the other hand, if we sample uniformly from each class we have $\mu_{ik} - \mu_{jk} = v_i - v_j$ assuming $P_i(Y|X + v_i) = P_j(Y|X + v_j)$.

4 Experiments

Data The models are evaluated on a dataset of hyperspectral signatures of Arabica, Robusta and immature Arabica coffee beans. The data was collected by three different hyperspectral cameras with slightly different measurement characteristics. The sensors measure 256 to 288 spectral bands in the wavelenghth area between 950 nm and 2500 nm. To ensure the same bands are considered for all nodes interpolation and subsampling to 50 features is performed [4]. Additionally to the three sensors, two additional artificial ones are simulated by adding an offset (as in described in [4]).

Set-Up We consider federated LVQ (FED) proposed in [5] and the weighted version (wFED) defined in Eqs. (1) and (2) as baselines not accounting for drift. Besides, we consider both models with offset elimination (FED+OE, wFED+OE)

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Sensor node S_i	S_0	S_1	S_2	S_2'	S_2''
Minority Classes	2	0	0,2	1,2	0,1
Missing Classes	0	1	2		

Table 1: Experimental setup: minority/missing classes for each sensor

sen.	raw data wFED	offset elim. wFED	drift fusion wFED
0	0.795 (0.0430)	0.777 (0.0367)	**0.826** (0.0506)
1	0.802 (0.0426)	0.751 (0.0274)	**0.819** (0.0508)
2	**0.822** (0.0440)	0.696 (0.0612)	0.816 (0.0505)

Table 2: Results of the experiment with missing classes

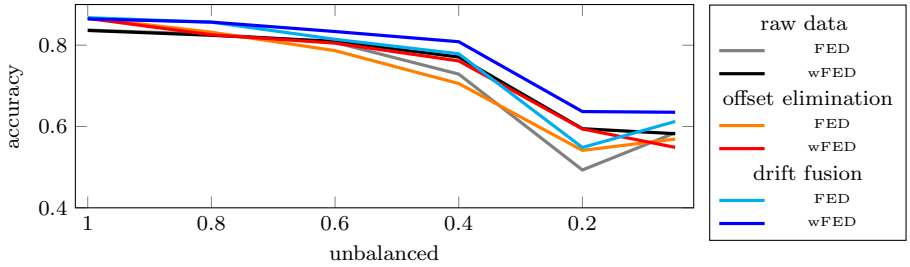

Fig. 1: Results of the experiment with unbalanced classes. Lighter colors are for FED and the darker ones for wFED and all three scenarios of offset elimination.

as proposed in [4] as a drift eliminating pre-processing. Additionally, we evaluate the proposed technology which accounts for drift by correcting the prototypes as defined in Eq. (3), on top of both models (drift-FED, drift-wFED).

We evaluate two different settings. First, we evaluate how well the different methods cope with imbalance in the different sensor nodes. For this purpose, we reduce the samples of one or two minority classes from all available samples to a fraction of 5 %. In a second experiment, one class is missing completely. The minority and missing classes for the different sensors are summarized in Table 1. We apply a 10-fold cross validation. Testing is done on a balanced test fold.

Results The results for the experiments with imbalanced data are documented in Fig. 1 and Table 3. In case the data is balanced (imbalance=1) all implementations accounting for the sensor shift perform better than the baselines FED and wFED. For increasing imbalance the accuracy for all variants decreases. However, we report a faster decline for the variants with offset elimination than for the novel method. Overall, as expected the weighted fusing function outperforms the standard fusing in presence of class imbalance.

Table 2 summarizes the results of the experiment with missing classes. As similar results for the weighted and regular version result if classes are balanced, we only report the results for the weighted version. In this scenario drift-wFED performs best again. We report worse results for the versions with offset elimination than for raw data. This is expected as the offset cannot be computed reliably if different classes are available at different nodes.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

	raw data		offset elimination		drift fusion	
imb	FED	wFED	FED	wFED	FED	wFED
1	0.837 (0.0442)	0.836 (0.0447)	0.867 (0.0129)	0.866 (0.0135)	**0.868** (0.0100)	0.865 (0.0112)
0.8	0.827 (0.0485)	0.824 (0.0504)	0.832 (0.0462)	0.826 (0.0496)	0.857 (0.0176)	**0.857** (0.0173)
0.6	0.807 (0.0532)	0.810 (0.0571)	0.787 (0.0678)	0.805 (0.0584)	0.814 (0.0404)	**0.834** (0.0345)
0.4	0.729 (0.0571)	0.771 (0.0493)	0.706 (0.0535)	0.761 (0.0590)	0.779 (0.0601)	**0.809** (0.0485)
0.2	0.493 (0.0567)	0.595 (0.0545)	0.541 (0.0526)	0.594 (0.0570)	0.548 (0.0486)	**0.637** (0.0481)
0.05	0.584 (0.0453)	0.582 (0.0425)	0.569 (0.0447)	0.549 (0.0482)	0.612 (0.0313)	**0.635** (0.0452)

Table 3: Results of the experiments with increasing class imbalance

5 Conclusion

We proposed a general formalism for federated learning with concept drift across nodes via adapting a transformation. As a special instantiation we considered federated LVQ and intensity shift as occur in hyperspectral data. In our experiments we found that while the federated LVQ models are already robust to sensor shifts, the proposed method can improve the performance. Besides, we demonstrated that it is more robust to class imbalance than other drift eliminating preprocessing schemes and showcased that it can handle the presence of different classes at different sensor nodes.

References

[1] Arjun Chennu, Paul Färber, Glenn De'ath, Dirk de Beer, and Katharina E. Fabricius. A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Scientific Reports*, 7(1):7122, Aug 2017.

[2] Muhammad Jaleed Khan, Hamid Saeed Khan, Adeel Yousaf, Khurram Khurshid, and Asad Abbas. Modern trends in hyperspectral image analysis: A review. *IEEE Access*, 6:14118–14129, 2018.

[3] Celio Pasquini. Near infrared spectroscopy: A mature analytical technique with new perspectives–a review. *Anal. Chim. Acta*, 1026:8–36, 2018.

[4] Valerie Vaquet, Patrick Menz, Udo Seiffert, and Barbara Hammer. Investigating intensity and transversal drift in hyperspectral imaging data. In *29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2021, Online event (Bruges, Belgium), October 6-8, 2021*, 2021.

[5] Johannes Brinkrolf and Barbara Hammer. Federated learning vector quantization. In *29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2021, Online event (Bruges, Belgium), October 6-8, 2021*, 2021.

[6] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol.*, 10(2):12:1–12:19, 2019.

[7] Petra Schneider, Michael Biehl, and Barbara Hammer. Adaptive relevance matrices in learning vector quantization. *Neural Computation*, 21(12):3532–3561, 2009.

[8] Kerstin Bunte, Petra Schneider, Barbara Hammer, Frank-Michael Schleif, Thomas Villmann, and Michael Biehl. Limited rank matrix learning, discriminative dimension reduction and visualization. *Neural Netw.*, 26:159–173, 2012.

[9] Valerie Vaquet, Patrick Menz, Udo Seiffert, and Barbara Hammer. Investigating intensity and transversal drift in hyperspectral imaging data. *Neurocomputing*, 2022.