Supplementary Material for A Memory-Efficient Federated Kernel Support Vector Machine for Edge Devices

Xiaochen Zhou and Xudong Wang, Fellow, IEEE

I. PROOF OF THEOREM 1

Let $P_{m,l}(w_m[l])$ denote the objective function of problem $\Omega_l$, i.e.,

$$P_{m,l}(w_m[l]) = C \sum_{i=1}^{N_m} \max\{0, 1 - B_i - y_i w_m[l]^T a(x_i)[l]\} + \frac{1}{2}||w_m[l]||^2.$$ 

In the $q$-th iteration of block boosting, block $w_m[l]$ of the local parameter vector $w_m = [w_m[1]^T, ..., w_m[L]^T]^T$ is optimized by first solving the dual problem of problem $\Omega_l$:

$$\max_{\alpha_{m,l}} \left\{ -\frac{1}{2}||A_m[l]\alpha_m[l]||^2 + \sum_{i=1}^{N_m} (1 - B_i(q))\alpha_{m,l,i} \right\}$$

s.t. $0 \leq \alpha_{m,l,i} \leq C, i = 1, 2, ..., N_m.$

As the optimal solution $\alpha^*_m(q)$ to the dual problem is obtained, it can be transformed to the optimal solution $w_m[l](q)$ to problem $\Omega_l$ via $w_m[l](q) = A_m[l]\alpha^*_m(q)$. Let $w_m[l](q-1)$ denote the initial value of $w_m[l]$ in the $q$-th iteration, then we have

$$P_{m,l}(w_m[l](q-1)) \geq P_{m,l}(w_m[l](q)). \quad (1)$$

Let

$$w_m(q-1) = [w_m[1](q-1)^T, ..., w_m[l](q-1)^T, ..., w_m[L](q-1)^T]^T$$

and

$$w_m(q) = [w_m[1](q)^T, ..., w_m[l](q)^T, ..., w_m[L](q)^T]^T,$$

then we have

$$P_m(w_m(q-1)) = P_m(w_m(q))$$

Based on equation (1), we have

$$P_m(w_m(q-1)) \geq P_m(w_m(q)). \quad (2)$$

Equation (2) indicates that the sequence of local training loss $(P_m(w_m(0)), ..., P_m(w_m(q)))$ is non-increasing when the local parameter vector is optimized by block boosting. Since $P_m(\cdot)$ is a strongly convex function, if the training loss cannot be further reduced, then the optimal local parameter vector $w^*_m$ is obtained.

The authors are with UM-SJTU Joint Institute, Shanghai Jiao Tong University. Corresponding author: Xudong Wang. Email: wxudong@ieee.org

II. PROOF OF THEOREM 2

In the $(t+1)$-th iteration of Fed-KSVM, edge device $m$ initially holds a local parameter vector $w_m(t) = A_m\alpha_m(t)$, and it employs block boosting to obtain the optimal solution $w^*_m(t+1)$ to

$$\min_{w_m} P_m(w_m; w^*_m(t+1)) := C \sum_{i \in I_m} \max\{0, 1 - y_i \hat{f}_m(x_i)\} + \frac{1}{2}||w_m||^2,$$

where $\hat{f}_m(x_i) = \frac{1}{\sqrt{D}}(w_m(t+1) + w_m)^T a(x_i)$. Based on the duality, $w^*_m(t+1)$ can be also expressed by

$$w^*_m(t+1) = A_m\alpha^*_m(t+1),$$

where $\alpha^*_m(t+1)$ is the optimal solution to

$$\max_{\alpha_m} D_m(\alpha_m; w^*_m(t+1)) := -\frac{1}{2}||\tilde{w}_m(t+1) + A_m\alpha_m||^2 + \sum_{i \in I_m} \alpha_i + \frac{1}{2}||\tilde{w}_m(t+1)||^2$$

s.t. $0 \leq \alpha_i \leq C, i \in I_m.$

By applying Lemma 1 in [1] to the dual objective function $D_m(\alpha_m; w^*_m(t+1))$, we have

$$\mathbb{E}[D_m(\alpha^*_m(t+1); w^*_m(t+1)) - D_m(\alpha_m(t); w^*_m(t+1))] \geq \frac{s_m}{N} \mathbb{E}[P_m(w_m(t); w^*_m(t+1)) - D_m(\alpha_m(t); w^*_m(t+1))],$$

where $s_m = \min_{w} \frac{\sum_{i \in I_m} ||\frac{1}{\sqrt{D}}a(x_i) - y_i||}{\sum_{i \in I_m} ||\frac{1}{\sqrt{D}}a(x_i)||^2 + ||\frac{1}{\sqrt{D}}w^T a(x_i) - y_i||}.$
Note that

\[
\sum_{m=1}^{M} P_m(\mathbf{w}_m(t); \mathbf{w}_m(t+1)) - D_m(\mathbf{a}_m(t); \mathbf{w}_m(t+1)) = C \sum_{i=1}^{N} \max\{0, 1 - y_i f(x_i)\} - \sum_{i} \alpha_i(t)
\]

\[
+ \sum_{m=1}^{M} \left( \frac{1}{2} ||\mathbf{w}_m(t)||^2 + \frac{1}{2} ||\mathbf{w}_m(t+1) + \mathbf{A}_m \alpha_m(t)||^2 \right)
\]

\[
- \frac{1}{2} ||\mathbf{w}_m(t+1)||^2
\]

\[
= C \sum_{i=1}^{N} \max\{0, 1 - y_i f(x_i)\} - \sum_{i} \alpha_i(t)
\]

\[
+ \sum_{m=1}^{M} \left( \frac{1}{2} ||\mathbf{w}_m(t)||^2 + \frac{1}{2} ||\mathbf{w}_m(t)||^2 - \frac{1}{2} ||\mathbf{w}_m(t+1)||^2 \right)
\]

\[
= C \sum_{i=1}^{N} \max\{0, 1 - y_i f(x_i)\} - \sum_{i} \alpha_i(t) + ||\mathbf{w}(t)||^2,
\]

and

\[
P(\mathbf{w}(t)) - D(\mathbf{a}(t))
\]

\[
= C \sum_{i=1}^{N} \max\{0, 1 - y_i f(x_i)\} + \frac{1}{2} ||\mathbf{w}(t)||^2
\]

\[
- \left(-\frac{1}{2} ||\mathbf{A} \mathbf{a}(t)||^2 + \sum_{i} \alpha_i(t) \right)
\]

\[
= C \sum_{i=1}^{N} \max\{0, 1 - y_i f(x_i)\} - \sum_{i} \alpha_i(t) + ||\mathbf{w}(t)||^2.
\]

Thus,

\[
\sum_{m=1}^{M} P_m(\mathbf{w}_m(t); \mathbf{w}_m(t+1)) - D_m(\mathbf{a}_m(t); \mathbf{w}_m(t+1)) = P(\mathbf{w}(t)) - D(\mathbf{a}(t)).
\]

By summing up equation [3] from 1 to \(M\), we have

\[
\sum_{m=1}^{M} \mathbb{E}[D_m(\mathbf{a}_m^*(t+1); \mathbf{w}_m(t+1)) - D_m(\mathbf{a}_m(t); \mathbf{w}_m(t+1))]
\]

\[
\geq \sum_{m=1}^{M} \frac{1}{N} \mathbb{E}[P_m(\mathbf{w}_m(t+1); \mathbf{w}_m(t)) - D_m(\mathbf{a}_m(t); \mathbf{w}_m(t+1))]
\]

\[
\geq \frac{s}{N} \mathbb{E}[P(\mathbf{w}(t)) - D(\mathbf{a}(t))],
\]

where \(s = \min_m s_m\).