JMLR: Workshop and Conference Proceedings 1–2

Supplementary Material for Canonical Soft Time Warping

Keisuke Kawano KAWANO@MOSK.TYTLABS.CO.JP
Satoshi Koide KOIDE@MOSK.TYTLABS.CO.JP
Takuro Kutsuna KUTSUNA@MOSK.TYTLABS.CO.JP
Toyota Central R&D Labs., 41-1 Yokomichi, Nagakute, Aichi, Japan

1. CSTW as a generalization of CTW

Let S^* be a set of optimal solutions of (CTW). In the following situation, we say *CTW has a unique optimal alignment*.

$$\forall(U^*, Y^*, A^*) \in S^*, \forall(\tilde{U}^*, \tilde{V}^*, \tilde{A}^*) \in S^*,$$
$$\text{it holds that } A^* = \tilde{A}^* \text{ if } U^* = \tilde{U}^* \text{ and } V^* = \tilde{V}^*.$$

Given projection matrices (U, V) in CSTW, we define $\bar{\mathcal{A}}(U, V)$ by $\bar{\mathcal{A}}(U, V) := \{A \in \mathcal{A} \mid \forall A' \in \mathcal{A}, \langle A, \Delta(U^\top X, V^\top Y)\rangle \leq \langle A', \Delta(U^\top X, V^\top Y)\rangle\}$, which corresponds to the set of alignments that achieve lowest cost.

Proposition 1 *Let (U^*, V^*) be an optimal solution of (CSTW) and $A^* \in \bar{\mathcal{A}}(U^*, V^*)$. Then, (U^*, V^*, A^*) is also a solution of (CTW) if $\gamma \to 0$, $|\bar{\mathcal{A}}(U^*, V^*)| = 1$, and CTW has unique optimal alignment.*

Proof When $|\bar{\mathcal{A}}(U^*, V^*)| = 1$ and $\gamma \to 0$, $P_\gamma(A; U, V) \to \delta_{A^*}$, where δ_{A^*} is Dirac delta at A^*. Therefore, for arbitrary function f, the expectation of f is $\mathbb{E}_{A\sim P_\gamma(A;U,V)}[f(A)] = f(A^*)$. Then, (U^*, V^*, A^*) satisfies the following from the constraints of (CSTW).

$$\begin{aligned}
&{U^*}^\top X W_x(A^*)^\top W_x(A^*) X^\top U^* = \mathbb{I}_d,\\
&{V^*}^\top Y W_y(A^*)^\top W_y(A^*) Y^\top V^* = \mathbb{I}_d,\\
&{U^*}^\top X A^* Y^\top V^* \in \mathcal{D}_d,\\
&X W_x(A^*)^\top \mathbf{1}_{l_w(A^*)} = \mathbf{0}_{d_x},\\
&Y W_y(A^*)^\top \mathbf{1}_{l_w(A^*)} = \mathbf{0}_{d_y}.
\end{aligned}$$

Thus, (U^*, V^*, A^*) also satisfies the constraints of (CTW), i.e., (U^*, V^*, A^*) is a feasible solution for (CTW).

Here, we assume that (U^*, V^*, A^*) is not an optimal solution of CTW and CTW has an optimal solution $(\tilde{U}^*, \tilde{V}^*, \tilde{A}^*)$, which satisfies $\mathcal{L}_{\text{CTW}}(U^*, V^*, A^*) > \mathcal{L}_{\text{CTW}}(\tilde{U}^*, \tilde{V}^*, \tilde{A}^*)$, where $\mathcal{L}_{\text{CTW}}(U, V, A) = \langle A, \Delta(U^\top X, V^\top Y)\rangle$. When CTW has the unique optimal alignment, $(\tilde{U}^*, \tilde{V}^*, \tilde{A}^*)$ satisfies the constraints of (CSTW), i.e., it is a feasible solution of CSTW. The objective functions of (CSTW) and (CTW) are the same from the definition of $\min^\gamma$ when $\gamma \to 0$. This means that $(\tilde{U}^*, \tilde{V}^*)$ is an optimal solution but not (U^*, V^*). However, this is contrary to the definition that (U^*, V^*) is an optimal solution of (CSTW). Therefore, (U^*, V^*, A^*) is an optimal solution of (CTW). ■

2. Dependency on Annealing Parameter

ACTW has an annealing parameter α, which controls how to decrease γ in each step of the optimization. In this section, we investigate the dependency of the ACTW performance on the annealing parameter using the same experiments in Section 5.2. We set candidates of α {0.8, 0.85, 0.9, and 0.95}. We initialized γ to 100 and then multiplied it by α in each step. Unlike Section 5.2, we set the number of iterations to 500 to make γ small enough even with large α. The resulting alignment similarities were 0.331, 0.390, 0.422, 0.427, respectively. These results indicate that slow annealing process would lead to better solutions.