Bayesian Weight Enhancement with Steady-State Adaptation

for Test-time Adaptation in Dynamic Environments

Jae-Hong Lee

Division of Language & AI, Hankuk University of Foreign Studies, Seoul, Republic of Korea.

In dynamic environments, weight evolution of TTA is unstable; the proposed regularization induces weights toward a steady-state.

Hankuk University of Foreign Studies

ICML International Conference On Machine Learning

1. Problem of TTA

Test-time adaptation (TTA)

- Unsupervised online adaptation
- Dynamic environments

Weight degradation problem

- Explicit gradient noise caused by unsupervised online learning in dynamic environments ➡ Model collapse

Existing methods leverage source model weight to recover the adapted weight.

2. Probabilistic Perspective of Existing Method

2.1 Existing method: weight-based TTA

$$\boxed{\hat{\boldsymbol{w}} = \alpha\hat{\boldsymbol{w}}_k + (1-\alpha)\hat{\boldsymbol{w}}_0}$$

2.2 Probabilistic framework: Bayesian weight enhancement

Bayesian weight averaging and plug-in approximation

$$p(y|\boldsymbol{x}_k, \boldsymbol{w}_k, \boldsymbol{w}_{k+1}) = \int p(y|\boldsymbol{x}_k, \boldsymbol{u})p(\boldsymbol{u}|\boldsymbol{w}_k, \boldsymbol{w}_{k+1})d\boldsymbol{u}$$

$$\approx p(y|\boldsymbol{x}_k, \hat{\boldsymbol{u}}_k)$$

$$\hat{\boldsymbol{u}}_k = \operatorname{argmax}_{\boldsymbol{u}}\, p(\boldsymbol{u}|\boldsymbol{w}_k, \boldsymbol{w}_{k+1})$$

Weight distribution

$$p(\boldsymbol{w}_{k+1}|\boldsymbol{u}) = \mathcal{N}(\boldsymbol{w}_{k+1}|\hat{\boldsymbol{w}}_0, R),\ p(\boldsymbol{u}|\boldsymbol{w}_k) = \mathcal{N}(\boldsymbol{u}|\hat{\boldsymbol{w}}_k, Q)$$

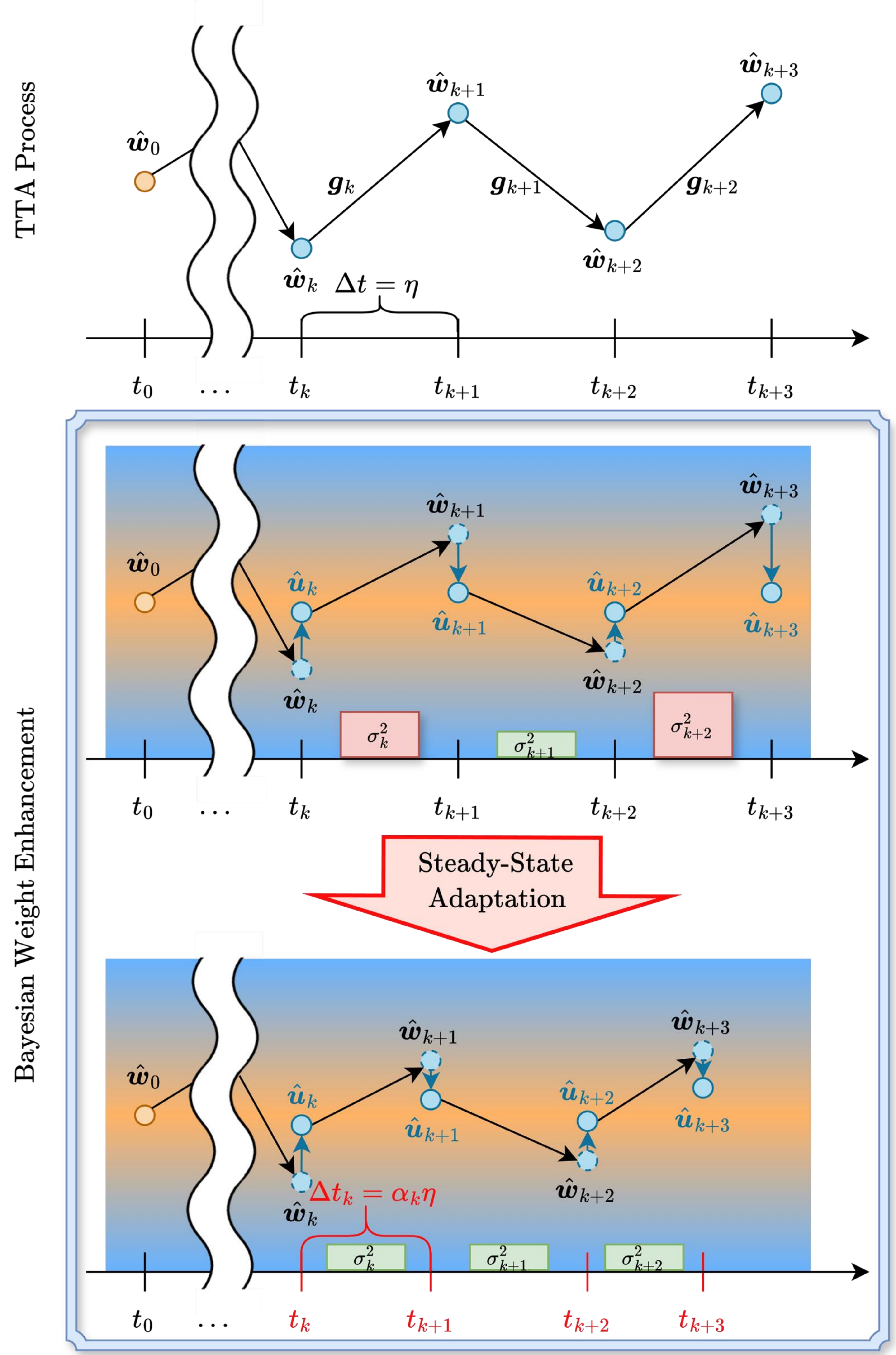
Bayesian inference

$$p(\boldsymbol{u}|\boldsymbol{w}_k, \boldsymbol{w}_{k+1}) \propto p(\boldsymbol{w}_{k+1}|\boldsymbol{u})p(\boldsymbol{u}|\boldsymbol{w}_k) = \mathcal{N}(\boldsymbol{u}_k|\boldsymbol{m}, \mathbf{P})$$

$$\mathbf{P} = AR,\ \ A = Q(Q+R)^{-1}$$

$$\boxed{\hat{\boldsymbol{u}}_k = \boldsymbol{m} = (\mathbf{I} - A)\hat{\boldsymbol{w}}_k + A\hat{\boldsymbol{w}}_0}$$

Can a static Q indicate the unstable weight evolution❓



3. Steady-State Adaptation

3.1 Transition distribution

$$p(\boldsymbol{u}_{k+1}|\boldsymbol{u}_k) \approx \mathcal{N}(\boldsymbol{u}_{k+1}|\boldsymbol{m}_{k+1|k}, \mathbf{P}_{\mathbf{k+1|k}})$$

$$\boldsymbol{m}_{k+1|k} = \boldsymbol{m}_k - g_k\Delta t, \quad \mathbf{P}_{k+1|k} = \sigma_k^2\Delta t^2\mathbf{I}$$

3.2 Bayesian filtering in Bayesian weight enhancement

$$p(\boldsymbol{u}_{k+1}|\boldsymbol{w}_{0:k+1}) = \mathcal{N}(\boldsymbol{u}_{k+1}|\boldsymbol{m}_{k+1}, \mathbf{P}_{k+1})$$

$$\mathbf{P}_{k+1} = (\mathbf{I} - \mathrm{A_k})\mathbf{P}^+_{k+1|k},\ \ A_k = \mathbf{P}^+_{k+1|k}(\mathbf{P}^+_{k+1|k} + R)^{-1}$$

$$\boxed{\boldsymbol{m}_{k+1} = \boldsymbol{m}^+_{k+1|k} + A_k(\hat{\boldsymbol{w}}_0 - \boldsymbol{m}^+_{k+1|k})}$$

3.3 Balancing covariance under steady-state condition

Steady-state condition

$$\mathrm{P_{k+1}} = \mathrm{P_k} = \mathrm{P_\infty} = \mathrm{p_\infty}\mathbf{I} \ \ ➡\ \ A_k = A$$

$$p(\boldsymbol{u}_{k+1}|\boldsymbol{w}_{0:k}) = \mathcal{N}(\boldsymbol{u}_{k+1}|\boldsymbol{m}^+_{k+1|k}, \mathbf{P}^+_{\mathbf{k+1|k}})$$

$$\mathbf{P}^+_{k+1|k} = \mathbf{P}_k + \alpha_k^2\sigma_k^2\Delta\mathrm{t}^2\mathbf{I} \qquad \alpha_k = \sqrt{\sigma_\lambda^2/(\eta^2\sigma_k^2)}$$

$$\boldsymbol{m}^+_{k+1|k} = \boldsymbol{m}_k - \alpha_k g_k\Delta t$$

"The results indicate that dynamic adjustment of the learning rate rule is required."

4. Benchmark

Table 3. Average error rates (%) and standard deviations in label shifts ($\gamma = 0.0$) on ImageNet-C. Red fonts indicate performance degradation with respect to Source.

Method	Adaptation Order (→) NOISE			BLUR				WEATHER				DiGiTAL				Avg.
	gaussian	shot	impulse	defocus	glass	motion	zoom	snow	frost	fog	bright	contrast	elastic	pixelate	jpeg	
Source	43.9	43.3	43.4	69.7	78.3	59.6	69.1	40.1	44.3	36.3	26.5	50.6	67.6	60.6	43.4	51.8
TENT	44.1	43.7	44.0	71.1	79.2	61.6	69.8	43.2	53.1	55.9	30.8	48.7	69.4	69.1	58.9	56.2±0.98
LAME	30.5	29.8	30.2	49.4	62.4	39.9	50.3	31.3	34.3	31.4	22.7	39.9	55.9	41.4	34.5	38.9±0.07
RoTTA	43.8	42.0	42.0	69.9	74.5	59.3	67.4	40.3	39.5	40.2	29.0	74.5	72.4	72.8	51.5	54.6±0.04
SAR	44.2	41.8	41.0	67.6	71.7	54.8	63.5	39.2	39.0	38.2	25.6	67.5	66.0	57.9	39.0	50.5±1.38
EATA	44.2	41.4	40.8	64.7	66.7	52.2	60.5	39.7	40.6	39.4	24.8	46.0	55.4	49.7	38.3	47.0±0.11
DeYO	41.5	38.9	38.9	61.7	61.3	51.8	72.0	42.2	41.6	39.7	26.5	56.4	57.1	47.3	41.4	47.9±0.57
ROID	**12.2**	11.8	11.6	32.5	33.5	18.4	30.1	12.4	11.6	9.8	7.3	12.6	25.1	15.3	13.0	17.1±0.32
+SSA	**12.2**	11.8	11.5	32.6	29.6	18.3	28.9	11.9	11.3	9.6	7.1	12.1	23.2	14.1	11.5	16.4±0.14
CMF	12.4	12.0	11.9	**28.8**	23.9	15.6	22.4	11.2	10.2	8.9	6.3	11.3	17.9	13.0	9.7	14.4±0.24
+SSA	12.3	**11.4**	**11.3**	29.2	**20.5**	**14.5**	**19.4**	**10.5**	**9.7**	**8.3**	**6.1**	**10.0**	**14.6**	**10.8**	**8.7**	**13.1±0.29**