ONLINE BLACK-BOX PROMPT OPTIMIZATION WITH REGRET GUARANTEES UNDER NOISY FEEDBACK

Jinjie Fang[1] **Runwen You**[1] **Wanli Shi**[1] **Wenkang Wang**[1]
Ganyu Wang[2] **Haozhen Zhang**[1] **Yi Chang**[1,3,4†] **Bin Gu**[1†]

[1]School of Artificial Intelligence, Jilin University, China
[2]Western University, Canada
[3]International Center of Future Science, Jilin University, China
[4]Engineering Research Center of Knowledge-Driven Human-Machine Intelligence, MOE, China
{wanli_sh, yichang, gubin}@jlu.edu.cn {gwang382}@uwo.ca
{fangjj24, runwen25, wangwk25, haozhen23}@mails.jlu.edu.cn

Background

1. **Impact of Generative AI** : Generative AI has demonstrated exceptional performance across a variety of tasks, including financial analysis, medical diagnosis support, and automated content creation.

2. **Prompt Tuning (PT) vs. Fine-Tuning (FT):** While FT adjusts all model weights—demanding massive computational resources and potentially reducing generalization—PT updates only a small subset of parameters. This approach preserves the model's inherent knowledge and adaptability while significantly reducing data requirements.

3. **The Necessity of Black-Box Optimization:** In many commercial platforms, intermediate model representations (like gradients) are inaccessible, making black-box prompt tuning essential for optimizing inputs without knowing internal model mechanisms.

Challenges

Limitations of Offline Learning: Most existing research on black-box prompt optimization focuses on offline scenarios using pre-established datasets. These methods lack the ability to adapt to dynamic data changes or real-time user interactions.

Inherent Randomness and Noise: GAI models are non-deterministic due to random sampling and seed initialization. This inherent randomness is often perceived as "noise," which complicates the optimization process in real-world online scenarios.

High Variance in Zeroth-Order (ZO) Estimates: While ZO optimization is a robust framework for black-box settings, it relies on a limited number of function evaluations to approximate gradients. This lead to high variance during the search process, further exacerbating the uncertainty in optimization.

Motivation

Online black-box prompt tuning: In an online learning scenario, a stream sample ξ^t is received at each round $t = 0, ..., T-1$, comprising an input sentence x^t and its corresponding true label y^t, i.e., $\xi^t = (x^t, y^t)$. Let $\mathcal{G}$ represent the black-box generative model and ℓ denote the loss function. The online black-box prompt tuning task involves minimizing the objective function f^t by optimizing the prompt ϕ:

$$f^t\left(\phi^t\right) \triangleq \ell\left(\mathcal{G}\left(\phi^t; x^t\right), y^t\right). \tag{1}$$

Based on the preceding discussion, mainstream black-box optimization methods, such as Bayesian and evolutionary algorithms, are impractical in online learning scenarios, necessitating gradient-based methods. However, directly applying gradient-based methods to optimize ϕ presents challenges, as ϕ represents a natural language sentence involving numerous discrete structures, rendering gradient-based methods unsuitable.

$$\hat{\nabla}_z f_\delta^t\left(z^t\right) = \frac{f_\delta^t\left(z^t + \mu u^t\right) - f_\delta^t\left(z^t - \mu u^t\right)}{2\mu} u^t,$$

$$\mathbf{m}_t = \frac{1}{W} \sum_{i=0}^{w-1} \alpha^i \cdot \hat{\nabla}_z f_\delta^{t-i}\left(z^{t-i}\right)$$

$$\mathbf{v}_t = \frac{1}{M} \sum_{i=0}^{w-1} \beta^i \cdot \left[\hat{\nabla}_z f_\delta^{t-i}\left(z^{t-i}\right)\right]^2$$

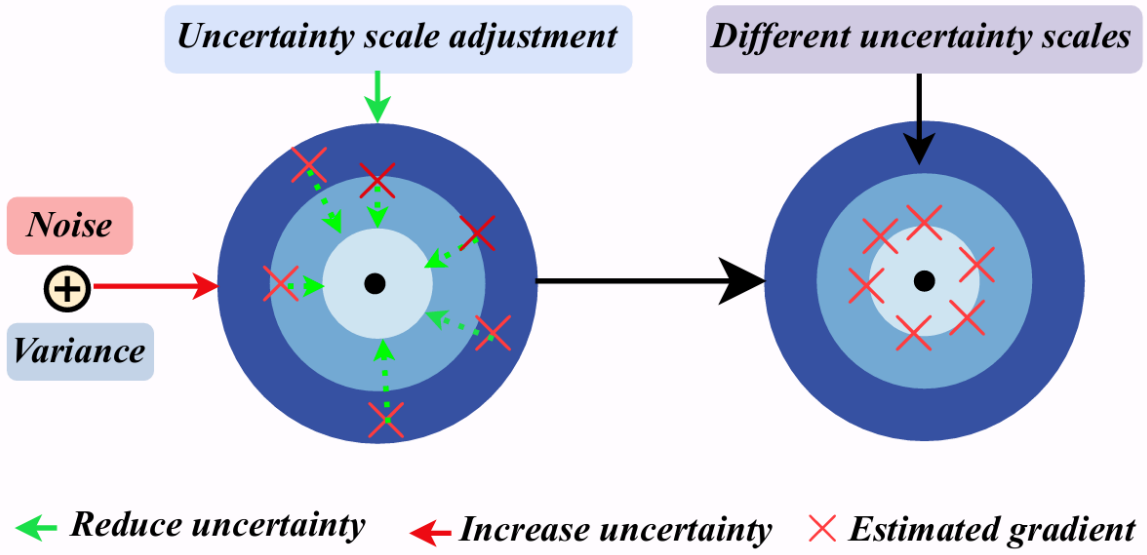

Figure 1: The adaptive uncertainty scale adjustment mechanism.

$$z^{t+1} \leftarrow z^t - \eta \cdot \frac{\mathbf{m}_t}{\sqrt{\mathbf{v}_t + \epsilon}}$$

Method

The AOZPT approach optimizes prompts in online black-box scenario (Figure 2). During the prompt generation phase, we utilize a frozen open-source LLM for instance optimization to refine the prompt tuning. This approach capitalizes on the LLM's robust capabilities in contextual learning and language comprehension. Specifically, we leverage the model's deep understanding of linguistic patterns and context to generate high-quality, semantically rich prompts by optimizing its soft prompts. In the prompt update phase, we introduce perturbations to the soft prompts to compute the differential of the output loss function, thereby approximating the gradient using zeroth-order gradient estimation. Additionally, we incorporate an adaptive uncertainty scale adjustment mechanism to address the uncertainty of online black-box prompt tuning (Algorithm 1).

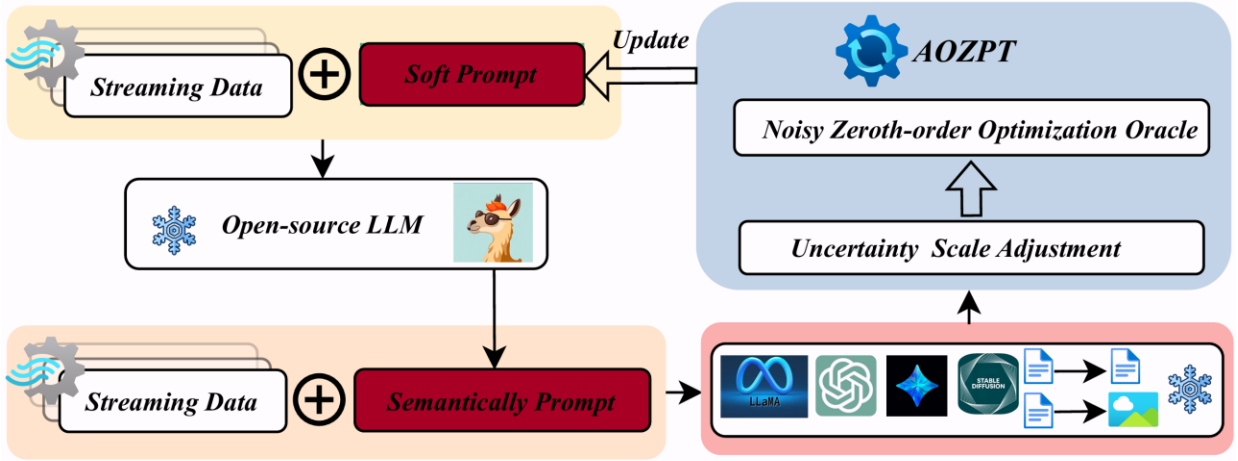

Figure 2: The architecture diagram of AOZPT model.

Algorithm 1 AOZPT

Input: learning rate η, smooth parameter μ, the length of the sliding window w, weighting parameter α and β, normalization parameter W and M, a small constant ϵ, initialize w-dimensional zero-initialized gradient vector Λ.

Output: $\{z^t\}_{t=1}^{T}$.

Initialize soft prompt z^0.

for $t = 0$ **to** $T-1$ **do**

Receive $\xi^t = \{x^t, y^t\}$.

Get u^t by sampled from unit sphere $\mathcal{S}^d$.

Compute: $\phi_+^t = \mathcal{F}\left(\mathrm{A}\left(z^t + \mu u^t\right) + \phi_0; \xi^t\right)$ and $\phi_-^t = \mathcal{F}\left(\mathrm{A}\left(z^t - \mu u^t\right) + \phi_0; \xi^t\right)$.

Compute $f_\delta^t\left(z^t + \mu u^t\right)$ and $f_\delta^t\left(z^t - \mu u^t\right)$:

$$f_\delta^t\left(z^t + \mu u^t\right) = \ell\left(\mathcal{G}\left(\phi_+^t; x^t\right), y^t\right) + \delta\left(z^t + \mu u^t\right),$$
$$f_\delta^t\left(z^t - \mu u^t\right) = \ell\left(\mathcal{G}\left(\phi_-^t; x^t\right), y^t\right) + \delta\left(z^t - \mu u^t\right).$$

Compute the estimation gradient $\hat{\nabla}_z f_\delta^t\left(z^t\right)$:

$$\hat{\nabla}_z f_\delta^t\left(z^t\right) = \frac{f_\delta^t\left(z^t + \mu u^t\right) - f_\delta^t\left(z^t - \mu u^t\right)}{2\mu} u^t$$

Update gradient vector:

$$\Lambda = \left[\hat{\nabla}_z f_\delta^{t-w+1}\left(z^{t-w+1}\right), \hat{\nabla}_z f_\delta^{t-w+2}\left(z^{t-w+2}\right), ..., \hat{\nabla}_z f_\delta^t\left(z^t\right)\right]$$

Compute $\mathbf{m}_t \leftarrow \frac{1}{W}\sum_{i=0}^{w-1} \alpha^i \cdot \hat{\nabla}_z f_\delta^{t-i}\left(z^{t-i}\right)$ and $\mathbf{v}_t \leftarrow \frac{1}{M}\sum_{i=0}^{w-1} \beta^i \cdot \left[\hat{\nabla}_z f_\delta^{t-i}\left(z^{t-i}\right)\right]^2$.

Update $z^{t+1} \leftarrow z^t - \eta \cdot \frac{\mathbf{m}_t}{\sqrt{\mathbf{v}_t + \epsilon}}$.

end for

Convergence Analysis

Theorem 4.9. *Under Assumption 4.3 - Assumption 4.6, solving the online Black-box prompt learning problem with Algorithm 1. For $t = 1, \ldots, T$, we suppose $\gamma = \frac{\alpha}{\beta^{1/2}} \in (0, 1]$. The following inequality is satisfied:*

$$\mathfrak{R}(T) \leq \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3. \tag{17}$$

where

$$\mathcal{E}_1 = \frac{\left(4H + 2V^T\right) G_\infty}{\eta}, \quad \mathcal{E}_2 = \frac{TG_\infty}{W\epsilon^{\frac{1}{2}}} \left(\frac{2d\Delta^2}{\mu^2} + \frac{L^2\mu^2(d+3)^3}{2} \right),$$

$$\mathcal{E}_3 = \frac{LT\eta M^{\frac{1}{2}} d^{\frac{1}{2}} G_\infty}{2W(1-\gamma)\epsilon^{\frac{1}{2}}} \left(\frac{L\mu(d+3)^{\frac{3}{2}}}{2} + dG + \frac{d^{\frac{1}{2}}\Delta}{\mu} \right).$$

Futher, we can get:

$$\mathfrak{R}(T) = \mathcal{O}\left(\frac{T}{W} + \frac{TM^{\frac{1}{2}}}{W} \right). \tag{18}$$

Experiments

Table 1: The average cumulative F1 score / accuracy ± standard deviation using Llama-3.1-8B, GPT-3.5-turbo and Qwen2.5-14B models for CNN/DailyMail, GSM8K Datasets. Each result is reported based on three Monte Carlo experiments. The best results are in bold.

Dataset	**CNN/DailyMail**			**GSM8K**		
Method	**Llama-3.1-8B**	**GPT-3.5-turbo**	**Qwen2.5-14B**	**Llama-3.1-8B**	**GPT-3.5-turbo**	**Qwen2.5-14B**
MP	24.253±0.079	34.269±0.035	22.068±0.038	60.533±0.471	69.200±2.209	80.200±0.589
ICL	23.500±0.601	32.364±0.259	23.064±0.028	60.667±0.250	69.933±0.806	86.733±0.416
BDPL	23.885±0.280	35.372±0.098	21.700±3.909	37.667±14.055	36.406±1.765	89.000±0.748
RLPROMPT	23.618±0.175	34.681±0.031	20.098±0.579	66.867±0.471	63.800±2.168	81.867±0.094
ZO-OGD	24.667±0.027	34.682±0.291	22.034±0.651	65.067±5.705	69.533±2.532	92.533±0.929
AOZPT(Ours)	**24.707±0.047**	**35.399±0.297**	**24.767±0.502**	**69.733±1.514**	**78.133±3.583**	**92.933±0.822**

Table 2: The average cumulative aesthetic ± standard deviation using Dreamlike-photoreal-2.0 and Stable Diffusion v1.5 models for Anime, Painting Datasets. Each result is reported based on three Monte Carlo experiments. The best results are in bold.

Dataset	**Anime**		**Painting**	
Method	**Dreamlik-2.0**	**Stable Diffusion v1.5**	**Dreamlike-2.0**	**Stable Diffusion v1.5**
MP	5.785±0.002	5.336±0.010	6.364±0.008	5.858±0.011
ICL	6.133±0.008	5.710±0.021	6.521±0.016	6.074±0.015
SFT	6.117±0.004	5.621±0.025	6.645±0.004	6.103±0.023
Promptist	6.093±0.010	5.579±0.006	6.552±0.004	6.011±0.022
ZO-OGD	6.263±0.024	5.892±0.039	6.602±0.053	6.287±0.013
AOZPT (Ours)	**6.282±0.021**	**5.930±0.015**	**6.656±0.015**	**6.313±0.009**

Thanks