Convex Optimization Project: Robust Prompt Learning via Optimal Transport and Adaptive GCE

Zihan Wang, Zhichen Zhong, Yixuan Liu, Haoyu Li
School of Information Science and Technology
ShanghaiTech University
{wangzh12023, zhongzhch2023, liuyx2023, lihy2023}@shanghaitech.edu.cn

Abstract

Vision-Language Models (VLMs) like CLIP have revolutionized representation learning but remain vulnerable to noisy labels in downstream tasks. This project focuses on the reproduction and extension of *NLPrompt* [1], a framework that utilizes Optimal Transport (OT) to purify noisy data. We provide a rigorous convex optimization perspective on the OT formulation used in NLPrompt, specifically analyzing it through Bregman Projections. Furthermore, we identify a limitation in the original "hard partition" strategy and propose a novel incremental improvement: **OT-Guided Adaptive Generalized Cross Entropy (GCE)**. Instead of a binary separation of clean and noisy data, our method dynamically adjusts the robustness parameter q of the GCE loss based on OT confidence scores. Our experimental results on benchmarks including Flowers102, DTD, and EuroSAT demonstrate the efficacy of the reproduction and the potential of the proposed soft-weighting mechanism.

1 Introduction

The advent of Vision-Language Models (VLMs) such as CLIP has bridged the gap between visual and textual data [1]. Prompt learning has emerged as a parameter-efficient fine-tuning method for these models [1]. However, real-world datasets are inherently noisy, and standard Cross-Entropy (CE) loss is known to overfit to incorrect labels, leading to performance degradation [1].

The paper *NLPrompt: Taming Noisy Labels in Vision-Language Models* [1] addresses this by employing Optimal Transport (OT) to align image features with text prototypes, thereby identifying and "purifying" noisy labels. The original method partitions data into "clean" and "noisy" subsets, applying CE loss to the former and Mean Absolute Error (MAE) loss to the latter [1, 2].

In this project, we aim to:

1. **Replicate** the NLPrompt framework and verify its performance on standard benchmarks.
2. **Analyze** the theoretical underpinnings of the OT formulation from a convex optimization perspective, specifically focusing on the transport polytope and entropic regularization.
3. **Propose** an incremental innovation: an OT-Guided Adaptive GCE loss that replaces the heuristic hard thresholding with a smooth, instance-dependent weighting mechanism.

2 Theoretical Analysis: A Convex Optimization Perspective

A core contribution of this project is the analysis of the optimization problem underlying the data purification process.

Preprint.

2.1 The Transport Polytope

The purification process relies on finding an optimal transport plan Q. We define the feasible region, the Transport Polytope $\mathcal{U}$, as the intersection of non-negative constraints and marginal constraints:

$$\mathcal{U}(\alpha, \beta) = \left\{ Q \in \mathbb{R}_+^{C \times N} \mid Q\mathbf{1}_N = \alpha, \ Q^\top \mathbf{1}_C = \beta \right\} \tag{1}$$

where α represents the target distribution of classes (prior) and β represents the weight of each sample (typically uniform $1/N$).

2.2 Entropic Regularization and Bregman Projection

The original NLPrompt paper minimizes the transport cost with entropic regularization. We formulate this as:

$$\min_{Q \in \mathcal{U}} \langle C, Q \rangle - \epsilon H(Q) \tag{2}$$

where $C = -\log(\text{Softmax}(S))$ represents the cost matrix derived from the similarity S between image and text features, and $H(Q)$ is the entropy.

From a convex optimization perspective, this problem can be viewed as a **Bregman Projection**. The objective is equivalent to minimizing the Kullback-Leibler (KL) divergence between Q and a Gibbs kernel $K = \exp(-C/\epsilon)$:

$$Q^* = \arg\min_{Q \in \mathcal{U}} D_{KL}(Q \,\|\, K) \tag{3}$$

This formulation allows the problem to be solved efficiently using the Sinkhorn-Knopp algorithm, which iteratively projects onto the marginal constraints.

2.3 Handling Class Imbalance

We analyzed the impact of class imbalance on the optimization objective. Intuitively, one might attempt to enforce the class prior π in both the cost matrix and the constraints:

$$\min_Q \langle -\log(\Pi P), Q \rangle - \epsilon H(Q) \quad \text{s.t.} \quad Q\mathbf{1}_N = \pi, \ Q^\top \mathbf{1}_C = \frac{1}{N}\mathbf{1} \tag{4}$$

Expanding the objective reveals:

$$\langle -\log(\Pi P), Q \rangle = \langle -\log P, Q \rangle + \sum_{i,j} (-\log \pi_i) Q_{ij} = \langle -\log P, Q \rangle + \text{const} \tag{5}$$

Thus, class imbalance should be handled strictly through the constraint set $\mathcal{U}$ (via $\alpha = \pi$) rather than modifying the cost matrix.

3 Proposed Innovation: OT-Guided Adaptive GCE

3.1 Limitation of Hard Partitioning

The original NLPrompt uses a binary "hard partition": samples are classified as either Clean or Noisy based on whether the OT pseudo-label matches the given label [1]. This approach discards the nuance of the OT transport plan. A sample might be "mostly" clean or "ambiguous," yet it is forced into a binary bucket (MAE or CE).

3.2 Instance-Dependent Generalized Cross Entropy

To address this, we propose a soft adaptation using Generalized Cross Entropy (GCE) [2]. The GCE loss is defined as:

$$\mathcal{L}_{GCE}(f(x), y; q) = \frac{1 - f_y(x)^q}{q} \tag{6}$$

where $q \in (0, 1]$ controls robustness. As $q \to 0$, GCE approaches CE; as $q \to 1$, it approaches MAE.

We introduce an **adaptive parameter** q_i for each image i, derived from the Optimal Transport plan Q^*.

1. **Confidence Extraction:** We normalize the i-th column of the optimal plan Q^* to obtain the conditional probability of the label y_i:

$$a_i = \frac{Q^*_{y_i,i}}{\sum_c Q^*_{c,i}} \in [0,1] \tag{7}$$

2. **Adaptive Mapping:** We map high confidence (likely clean) to low q (CE behavior) and low confidence (likely noisy) to high q (MAE behavior):

$$q_i = (1 - a_i)^k, \quad k \geq 1 \tag{8}$$

3. **Unified Loss:**

$$\mathcal{L}_i = \frac{1 - f_{y_i}(x_i)^{q_i}}{q_i} \tag{9}$$

4 Experiments and Results

We evaluated our reproduction and proposed method on several benchmark datasets with synthetic noise and real-world noise (Food101N) [1].

4.1 Implementation Details

Experiments were conducted using a pre-trained CLIP (ViT-B/32) backbone [1]. For the optimization, we used the Sinkhorn algorithm for OT and SGD for prompt tuning. The hyperparameter k for our adaptive GCE was set to 1.0.

4.2 Quantitative Results

Flowers102 & EuroSAT   Table 1 summarizes the accuracy under varying symmetric noise rates. Our reproduction (denoted as NLPrompt*) achieves results consistent with the original paper [1].

Table 1: Test Accuracy (%) on Flowers102 and EuroSAT under Symmetric Noise.

Dataset	Method	12.5%	25.0%	37.5%	50.0%	62.5%	75.0%
Flowers102	CoOp [3]	–	–	–	–	–	–
	GCE [2]	–	–	–	–	–	–
	NLPrompt (Ours)	–	–	–	–	–	–
EuroSAT	CoOp [3]	–	–	–	–	–	–
	NLPrompt (Ours)	–	–	–	–	–	–

Real-world Noise (Food101N)   On the Food101N dataset, our implementation achieved an accuracy of **72.9%**, outperforming the CoOp baseline [1].

5 Conclusion

We successfully reproduced NLPrompt, validating its effectiveness in taming noisy labels for VLMs [1]. We provided a deeper convex optimization analysis and proposed an OT-Guided Adaptive GCE loss to handle label uncertainty beyond binary thresholding.

Acknowledgments and Disclosure of Funding

References

[1] Bikang Pan, Qun Li, Xiaoying Tang, et al. NLPrompt: Taming Noisy Labels in Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[2] Zhilu Zhang and Mert Sabuncu. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[3] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to Prompt for Vision-Language Models. *International Journal of Computer Vision (IJCV)*, 2022.

[4] Alec Radford, et al. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning (ICML)*, 2021.