Self Iterative Label Refinement via Robust Unlabeled Learning

Hikaru Asano[1], Tadashi Kozuno[2], Yukino Baba[1]
[1]The University of Tokyo, [2]OMRON SINIC X

NEURAL INFORMATION PROCESSING SYSTEMS

OpenReview

TL;DR

LLM self-refinement fails when the model's internal knowledge is insufficient. We resolve this issue using Robust Unlabeled-Unlabeled Learning, enabling for iterative performance gains even with minimal human supervision.

Introduction

Background

- **LLM Self-refinement**: a method for enhancing a model's ability by using the LLM as its own evaluator to iteratively gain performance.
- **Problem**: self-refinement fails when insufficient internal knowledge results in unreliable self-evaluation and performance degradation.

? Research Question

Can LLMs achieve self-refinement in domains where their internal knowledge is insufficient, using only minimal human supervision?

Our Findings

YES! LLM knowledge + weakly supervised learning achieves self-refinement even when recent reasoning models fail to improve.

Preliminaries

Focus: Binary classification, as it serves as a fundamental building block for LLM post-training (e.g., RLHF, DPO).

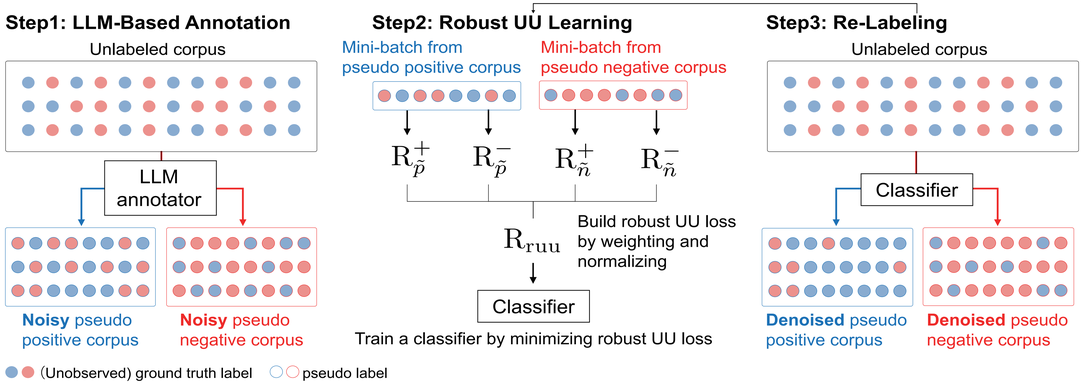
PN Learning: Standard supervised learning assumes access to ground-truth label $y \in \{\pm 1\}$, and positive C_p, negative C_n datasets.

$$R_{pn}(g) = \pi_+ \mathbb{E}_{x\sim C_p}[l(g(x), +1)] + (1-\pi_+)\mathbb{E}_{x\sim C_n}[l(g(x), -1)]$$

Pseudo Dataset: Assume two noisy datasets, pseudo positive $\tilde{C}_p$ pseudo negative $\tilde{C}_n$ datasets. Each dataset contains true positive samples but with different positive priors.

$$\theta_p = p(y=+1 \mid \tilde{x} \in \tilde{C}_p) \quad \text{and} \quad \theta_n = p(y=+1 \mid \tilde{x} \in \tilde{C}_n)$$

Method



Robust Unlabeled-Unlabeled Learning

Unlabeled-Unlabeled Learning (UU Learning) [1]

- Weakly supervised framework training a classifier using two unlabeled datasets $\tilde{C}_p$ and $\tilde{C}_n$ with different positive priors $\theta_p > \theta_n$. The resulting loss provides an unbiased estimator of the supervised risk R_{pn}

$$R_{\mathrm{uu}}(g) = \underbrace{a\mathbb{E}_{x\sim\tilde{C}_p}[l(g(x),+1)]}_{R^+_{\tilde{p}}} - \underbrace{b\mathbb{E}_{x\sim\tilde{C}_n}[l(g(x),+1)]}_{R^+_{\tilde{n}}} - \underbrace{c\mathbb{E}_{x\sim\tilde{C}_p}[l(g(x),-1)]}_{R^-_{\tilde{p}}} + \underbrace{d\mathbb{E}_{x\sim\tilde{C}_n}[l(g(x),-1)]}_{R^-_{\tilde{n}}}$$

Robust Unlabeled-Unlabeled Learning (Robust UU Learning) [2]

- An extension of UU learning that applies negative risk correction and mitigates overfitting.

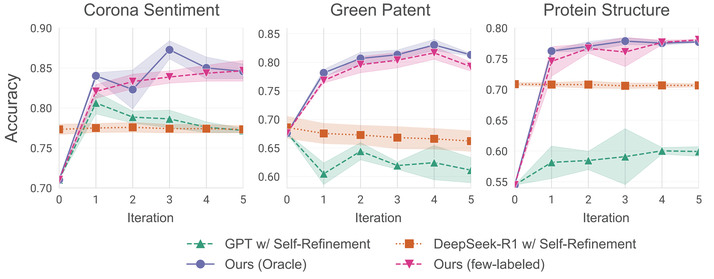
$$R_{\mathrm{ruu}}(g) = f\left(aR^+_{\tilde{p}}(g) - cR^+_{\tilde{n}}(g)\right) + f\left(dR^-_{\tilde{n}}(g) - bR^-_{\tilde{p}}(g)\right)$$

Reference

[1] Lu et al., "On the minimal supervision for training any binary classifier from only unlabeled data." ICLR, 2019
[2] Lu et al., "Mitigating overfitting in supervised classification from two unlabeled datasets: A consistent risk correction approach." AISTATS, 2020

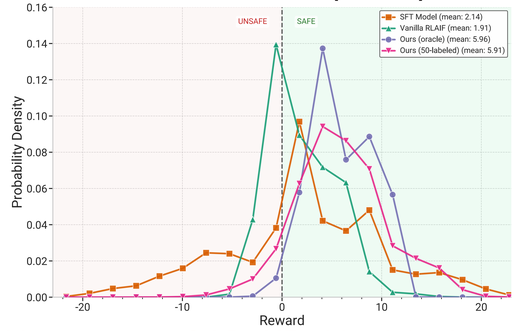
Experiment

Classification Tasks



Ours improves classification accuracy with only 50 labels, even in domains where reasoning models degrade performance.

Generative Task (RLHF)



Ours achieves successful RLAIF with minimal supervision, facilitating effective post-training alignment for generative task.