Data Agent: Learning to Select Data via End-to-End Dynamic Optimization

Suorong Yang, Fangjian Su, Hai Gan, Ziqi Ye, Jie Li, Baile Xu, Furao Shen, Soujanya Poria
Nanjing University, National University of Singapore, CNRS@CREATE, Nanyang Technological University

Paper Code

ICML International Conference On Machine Learning

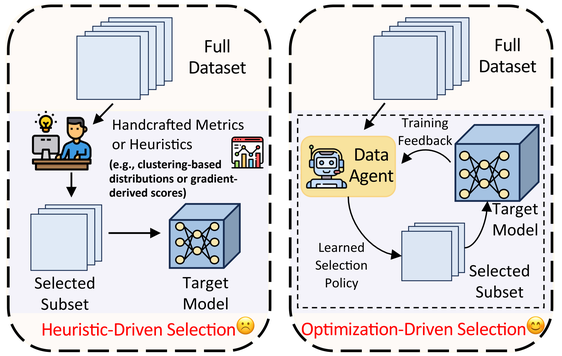
Background & Motivation

Problem

Existing data selection often depends on task-specific handcrafted scores or stale snapshot criteria, so sample utility cannot co-evolve with the model.

Closed-loop, training-aware selection.

Core Idea



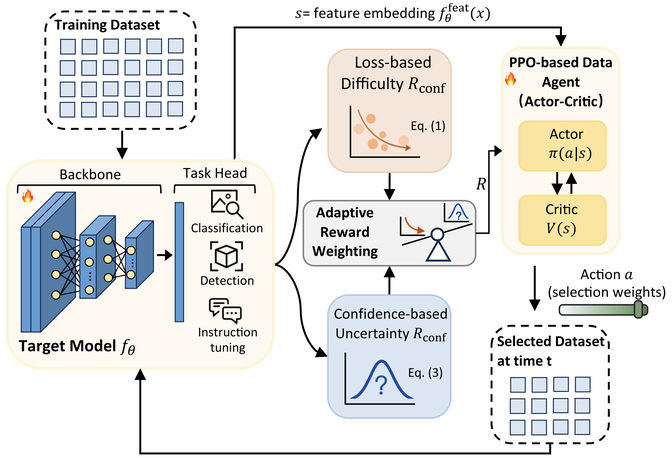
- Formulate dynamic selection as MDP;
- Agent observes model embeddings;
- Selection policy evolves online with training;

Training Costs: > 2× → No Acc Drop

Methodology

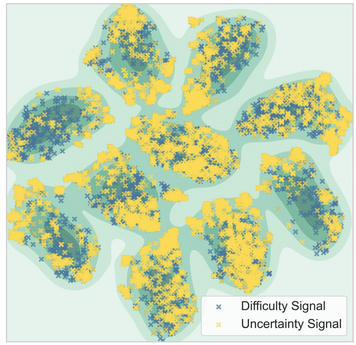
Method

Let AI Learn to Train AI.



Training-aware reward

- Difficulty reward: per-sample loss estimate optimization impact;
- Uncertainty reward: confidence signal targets information gains;
- Adaptive weighting shifts emphasis between difficulty and uncertainty;
- PPO actor-critic learns a coherent sample-wise selection policy.



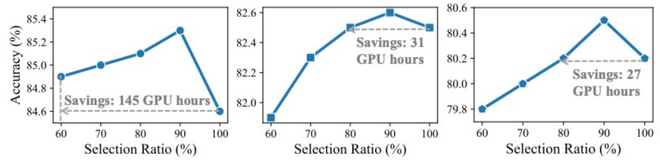
Training cost ~ 50% Lossless reduction

Noise ~ 8% Robustness gain

Cross-task > 10% Efficiency gain

Experiment

Dataset	CIFAR-10			CIFAR-100			Tiny-ImageNet		
Whole Dataset	95.6			78.2			45.0		
Selection Ratio (%)	30	50	70	30	50	70	30	50	70
Static selection methods									
Random	90.2	92.3	93.9	69.7	72.1	73.8	29.8	37.2	42.2
EL2N (Paul et al., 2021)	91.6	95.0	95.2	69.5	72.1	77.2	26.6	37.1	44.0
GraNd (Paul et al., 2021)	91.2	94.6	95.3	68.8	71.4	74.6	29.7	36.3	43.2
Forgetting (Toneva et al., 2018)	91.7	94.1	94.7	69.9	73.1	75.3	28.7	33.0	41.2
RL-Selector (Yang et al., 2025b)	91.8	-	95.4	71.1	-	77.6	31.1	-	44.5
Herding (Welling, 2009)	80.1	88.0	92.2	69.6	71.8	73.1	29.4	31.6	39.8
Moderate (Xia et al., 2023)	91.5	94.1	95.2	70.2	73.4	77.3	30.6	38.2	42.8
Glister (Killamsetty et al., 2021b)	90.9	94.0	95.2	70.4	73.2	76.6	30.1	39.5	43.9
DP (Yang et al., 2023b)	90.8	93.8	94.9	-	73.1	77.2	-	-	-
Self-sup. prototypes (Sorscher et al., 2022)	91.0	94.0	95.2	70.0	71.7	76.8	27.7	37.9	43.4
MoSo (Tan et al., 2024)	91.1	94.2	95.3	70.9	73.6	77.5	31.2	38.5	43.4
CLIP-Sel (Yang et al., 2024)	91.9	94.5	95.0	70.8	73.7	77.0	31.7	40.0	46.0
Dynamic pruning methods									
Random*	93.0	94.5	94.8	74.4	75.3	77.3	41.5	42.8	43.1
UCB (Hong et al., 2024)	93.9	94.7	95.3	-	75.3	77.3	-	-	-
ϵ-Greedy (Hong et al., 2024)	94.1	94.9	95.2	-	74.8	76.4	-	-	-
InfoBatch (Qin et al., 2023)	94.7	95.1	95.6	76.5	78.1	78.2	42.2	43.2	43.8
Ours	**95.0**	**95.3**	**96.0**	**77.6**	**78.9**	**79.5**	**44.9**	**47.0**	**49.4**



Method	Noisy Dataset		Corrupted Dataset	
	20%	30%	20%	30%
Random	17.8	23.9	20.0	25.9
Random*	32.5	36.1	36.5	38.3
CLIP-Sel	26.1	33.1	26.1	32.1
Ours	**38.1**	**41.9**	**41.9**	**46.7**

R_{diff}	R_{conf}	r	30%	50%	70%
			41.5	42.8	43.1
✓			42.1	45.0	48.2
	✓		42.2	45.8	48.4
✓	✓		42.5	46.1	48.6
✓	✓	✓	**44.9**	**47.0**	**49.4**