NeurIPS 2025 SeDPO:
Learning to Rank for In-Context Example Retrieval

Yuwen Ji*, Luodan Zhang*, Ambyer Han*, Haoran Que, Lei Shi, Wang Chao, Yue Zhang†

Presenter: Luodan Zhang

2025/11/04

What is ICL and Why it Matters

In-Context Learning (ICL) Definition:

- **Parameter-efficient paradigm**
  Adapt to new tasks or generate task-specific outputs via **a few in-context examples (ICEs) in prompts**.
- **No model fine-tuning required**

ICL Advantages:

- **Reducing deployment/iteration costs**
  Empowers few-shot rapid adaptation.
- **No massive labeled data needed**
  Excels in **data-scarce, task-diverse, rapid prototyping** scenarios.

Current Limitation

Mainstream ICL Retriever Training:

- **Point-wise Paradigm**
  Classify ICEs into "**top-1 example**" and "**others**" based on LLM scores.
- **Critical Limitation**
  Training objective (classification) misaligns with **inference (ranking-based retrieval)**.

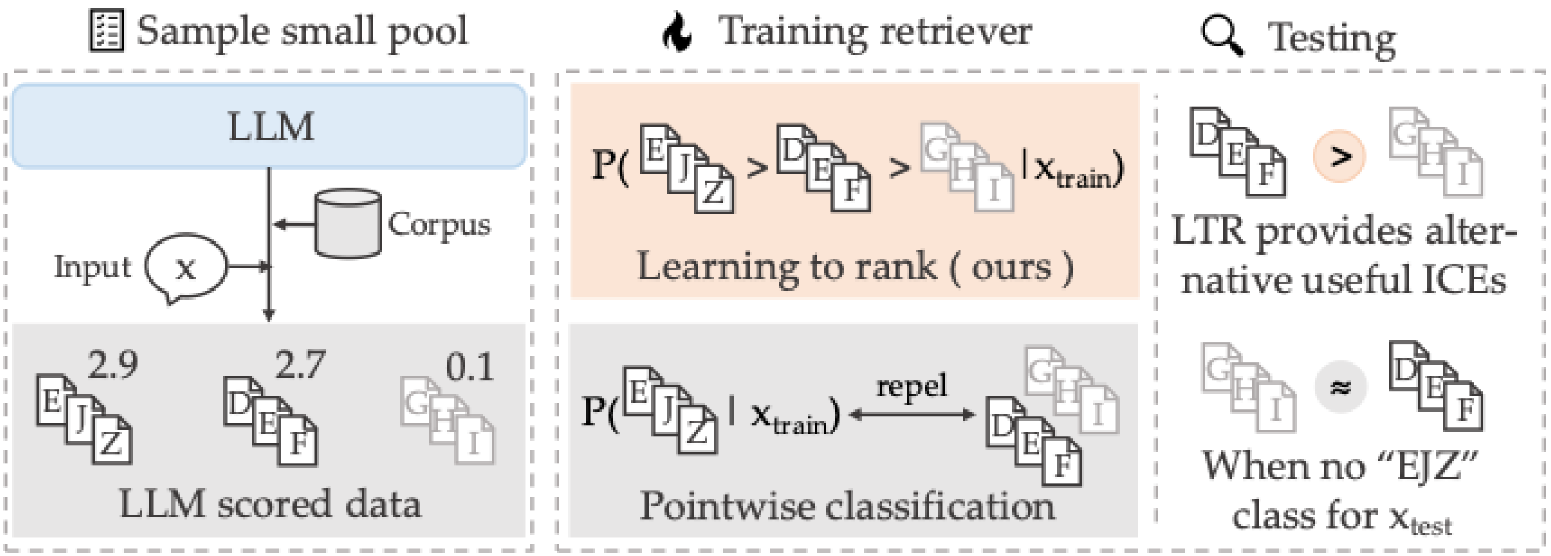

Fig.1. ICL Task Pipeline

SeDPO Introduction

Motivation

- **Learning inter-ICE preference rankings** outperforms **binary classification**.
- Even without exact matching ICEs, retrieves useful examples for better LLM performance.

Method: Train retrievers via a ranking task.

- ICEs's **Partial Order** by LLM's correct answer generation probability;
- Train retrievers with proposed **SeDPO loss**.

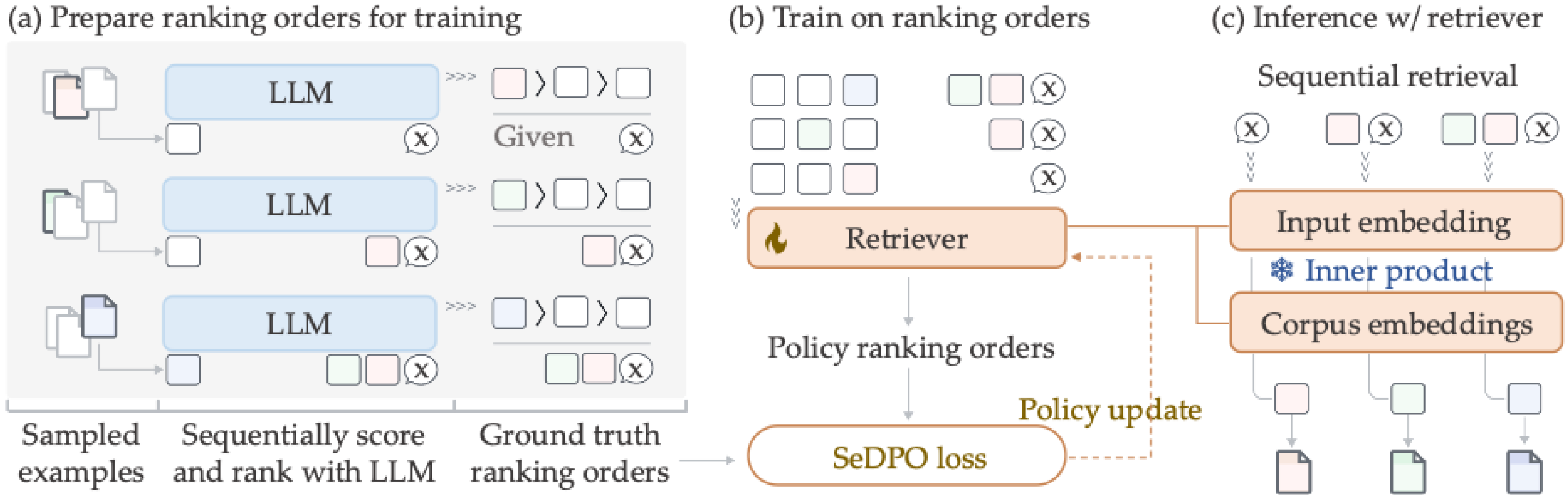

Fig.2. SeDPO Framework

How to Learn Ranking?

- **ICL Score Φ_{LLM} :**
  Given input x, K-shot ICE sets $\{e^w\}$ and $\{e^l\}$.

  Φ_{LLM} : The normalized log-likelihood of true label y after $x + \{e\} \rightarrow$ LLM.

  The **partial order** relationships of ICEs:

  $$\Phi_{LLM}(\{e^w\}, x, y) > \Phi_{LLM}(\{e^l\}, x, y)$$

  means $\{e^w\}$ is more optimal.

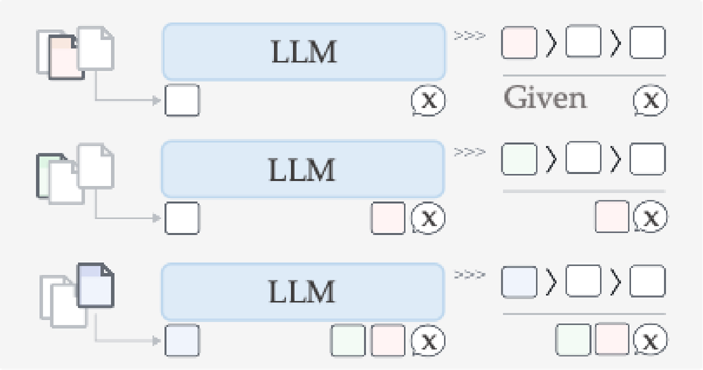

Fig.3. ICL Score

- **Retriever Score $\Phi_{retrieval}$**
  $\Phi_{retrieval}$: Similarity between input x and example e_k given condition $\{e_i\}_{i<k}$.

  $$\Phi_{retrieval} = sim(e_k, x|\{e_i\}_{i<k})$$

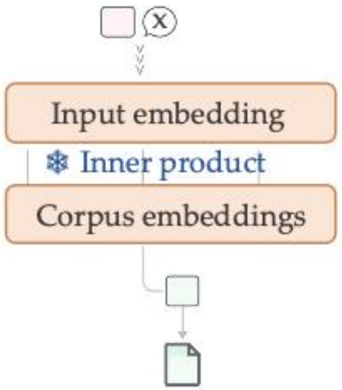

Fig.4. Retriever Score

How to Learn Ranking?

Use **DPO algorithm** to align the Φ_{LLM} and $\Phi_{retrieval}$ to learn the **preference ranking order** between different ICEs.

- **Retriever as Policy Model**

$$\pi(e_{[1:K]}|x) = \prod_{k=1}^{K} \frac{\exp(\mathrm{sim}(e_k, x|\{e_i\}_{i<k}))}{\sum_{e_* \in \mathcal{C} \setminus \{e_i\}_{i<k}} \exp(\mathrm{sim}(e_*, x|\{e_i\}_{i<k}))}$$

- **DPO Loss Formula**

$$\mathcal{L}_{\mathrm{DPO}}(\pi_\theta; \pi_{\mathrm{ref}}) =$$

$$-E_{(x, e^w_{[1:K]}, e^l_{[1:K]})}\left[\log \sigma\left(\beta \log \frac{\pi_\theta(e^w_{[1:K]}|x)}{\pi_{\mathrm{ref}}(e^w_{[1:K]}|x)} - \beta \log \frac{\pi_\theta(e^l_{[1:K]}|x)}{\pi_{\mathrm{ref}}(e^l_{[1:K]}|x)}\right)\right]$$

Improving Ranking Learning

- **Standard DPO Loss Problem in Practice**

$$\mathcal{L}_{\text{DPO}}(\pi_\theta;\pi_{ref}) = -\mathbb{E}_{(x,e^w_{[1:k]},e^l_{[1:k]})\sim\mathcal{D}}\left[\log\sigma\left(\beta\cdot f_\theta(x,e^w_{[1:K]}) - \beta\cdot f_\theta(x,e^l_{[1:K]}) - \beta\cdot(\gamma_w-\gamma_l)\right)\right]$$

$$f_\theta(x,e^j_{[1:K]}) = \sum_{k=1}^{K}\left[\text{sim}_\theta(e^j_k,x|\{e^j_i\}_{i<k}) - \text{sim}_{\text{ref}}(e^j_k,x|\{e^j_i\}_{i<k})\right], \qquad j\in\{w,l\}$$

$$\boxed{\gamma_j = \sum_{k=1}^{K}\log\frac{\sum_{e_*\in\mathcal{C}\setminus\{e^j_i\}_{i<k}}\exp(\text{sim}_\theta(e_*,x|\{e^j_i\}_{i<k}))}{\sum_{e_*\in\mathcal{C}\setminus\{e^j_i\}_{i<k}}\exp(\text{sim}_{\text{ref}}(e_*,x|\{e^j_i\}_{i<k}))}, \qquad j\in\{w,l\}}$$

Expensive computation!

Improving Ranking Learning

- **Sequential Relaxation**

$$\pi(e_{[1:K]}|x) = \prod_{k=1}^{K} \frac{\exp(\text{sim}(e_k, x|\{e_i\}_{i<k}))}{\sum_{e_* \in \mathcal{C}\setminus\{e_i\}_{i<k}} \exp(\text{sim}(e_*, x|\{e_i\}_{i<k}))}$$

$$\pi(e_k|\tilde{x}_{k-1}) = \frac{\exp(\text{sim}(e_k, \tilde{x}_{k-1}))}{\sum_{e_* \in \mathcal{C}\setminus\{e_i\}_{i<k}} \exp(\text{sim}(e_*, \tilde{x}_{k-1}))}$$

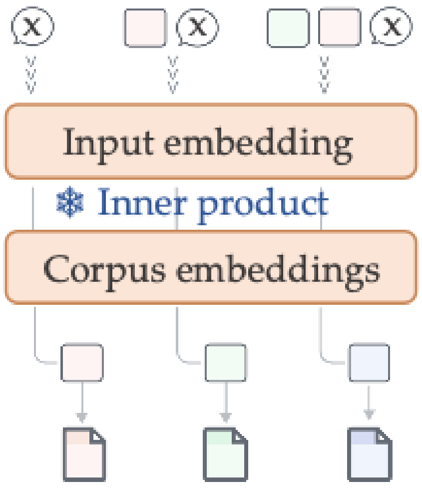

Fig.5. Sequential Retrieval

- **Improved SeDPO Loss with Sequential Relaxation**

$$\mathcal{L}_{\text{SE-DPO}}(\pi_\theta; \pi_{ref}) =$$

$$-E_{(\tilde{x}_k, e_k^w, e_k^l)\sim\tilde{\mathcal{D}}}\left[\log \sigma\left(\beta \cdot f_\theta(\tilde{x}_k, e_k^w) - \beta \cdot f_\theta(\tilde{x}_k, e_k^l) - \beta \cdot 0\right)\right],$$

$$f_\theta(\tilde{x}_k, e_k^j) = \text{sim}_\theta(e_k^j, \tilde{x}_k) - \text{sim}_{\text{ref}}(e_k^j, \tilde{x}_k)$$

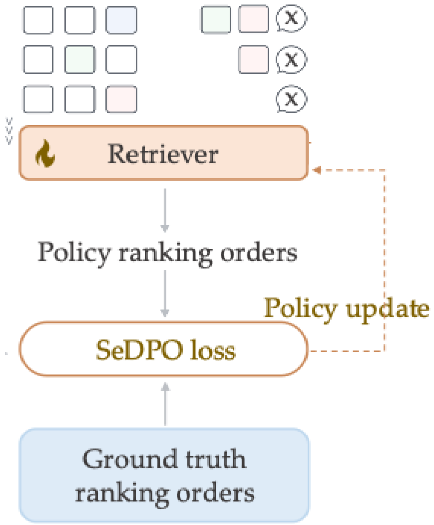

Fig.6. SeDPO Loss

Main Results

Table 1: Main results on various tasks. The **best results** and the second-best are highlighted. The *Avg.* of all metrics for tasks within the same category with significant improvements is marked by ↑.

	Paraphrase						*Coreference*
	MRPC		**PAWS**	**QQP**		*Avg.*	**WSC**
	acc	f1	acc	acc	f1		acc/*Avg.*
Zeroshot	46.1±0.0	45.3±0.0	51.8±0.0	48.4±0.0	42.1±0.0	46.7±0.0	59.6±0.0
Random	66.8±3.0	79.5±4.1	50.1±3.8	40.6±4.8	50.9±7.8	57.6±3.8	48.3±8.2
BM25	57.8±0.0	69.1±0.0	48.9±0.0	54.8±0.0	55.4±0.0	57.2±0.0	52.4±0.0
SBERT	56.4±0.0	66.9±0.0	49.4±0.0	51.2±0.0	56.2±0.0	56.0±0.0	46.2±0.0
UDR	65.9±4.6	75.4±3.5	51.8±1.2	74.1±1.9	67.9±2.4	67.0±1.2	52.0±4.7
UPRISE	74.0±0.8	83.3±0.1	49.1±0.0	71.0±1.0	69.8±0.1	69.4±0.2	46.5±2.2
Se^2	77.6±0.4	85.4±0.3	54.7±0.1	75.5±0.1	72.8±0.0	73.2±0.2	55.1±0.9
SE-DPO	**77.9±0.9**	**85.6±0.2**	**73.0±2.9**	**77.6±0.6**	**75.0±0.2**	**77.9±0.6**↑	**62.5±0.2**↑

	Reading				*Natural Language Inference (NLI)*		
	MultiRC	**BoolQ**	**AGNews**	*Avg.*	**MNLI-m**	**MNLI-mm**	*Avg.*
	f1	acc	acc		acc	acc	
Zeroshot	57.1±0.0	54.6±0.0	38.4±0.0	50.0±0.0	35.2±0.0	36.4±0.0	35.8±0.0
Random	57.7±2.5	54.8±6.7	25.8±1.1	46.1±1.2	34.2±3.0	34.9±3.9	34.6±1.6
BM25	46.5±0.0	60.3±0.0	81.7±0.0	62.8±0.0	35.3±0.0	35.6±0.0	35.5±0.0
SBERT	49.3±0.0	58.1±0.0	84.7±0.0	64.0±0.0	37.3±0.0	37.3±0.0	37.3±0.0
UDR	55.3±3.1	54.6±1.9	88.5±1.0	66.1±0.9	62.7±1.5	65.0±1.3	63.8±1.4
UPRISE	55.4±0.2	61.5±0.1	90.6±0.8	69.2±0.1	68.5±0.1	70.3±0.3	69.4±0.2
Se^2	47.1±3.3	64.1±2.2	90.7±0.3	67.3±0.7	69.4±0.2	70.4±0.1	69.9±0.2
SE-DPO	**61.6±0.4**	**66.2±1.7**	**90.7±0.2**	**72.8±0.6**↑	**70.6±0.1**	**72.0±0.3**	**71.3±0.2**↑

Demonstrating top-1 performance across 9 NLP tasks.

Diversity of Retrieved Examples

西湖大學 WESTLAKE UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Table 3: The average textual/semantic diversity of selected ICEs, as well as the average performance when the input order of ICEs is randomized. We take the main results on *Paraphrase* as our base.

	SE-DPO	Se^2	UPRISE	UDR	SBERT	BM25	Random
Textual Diversity	53.3%	49.0%	46.7%	54.3%	49.7%	46.0%	**61.4%**
Semantic Diversity	40.7%	39.0%	37.3%	40.3%	25.3%	29.0%	**46.0%**
Random order (Best of 5)	78.2%	73.5%	70.9%	68.4%	57.3%	58.1%	-
Random order (Worst of 5)	77.1%	72.5%	68.6%	66.3%	55.1%	57.0%	-

Better trades off diversity with ICL utility and successfully retrieves diverse yet useful ICEs.

Transferability

Table 4: Transferability on shot number and model size. The average performance of *Paraphrase*.

Inference Model	Method	1-shot	3-shot	6-shot	9-shot	12-shot	15-shot	Average
GPT-2-XL-1.5B (0-shot=39.6)	BM25	57.6	58.5	58.8	58.7	59.3	60.1	58.8
	SBERT	57.9	57.5	59.0	59.6	58.6	58.3	58.5
	UPRISE	69.2	69.4	69 .8	69.8	70.0	70.2	69.7
	Se^2	73.9	72.9	72.9	72.8	72.8	72.7	73.0
	SeDPO	75.0	78.9	79.5	79.2	79.0	79.2	**78.5**
GPT-Neo-2.7B (0-shot=46.7)	BM25	57.1	57.2	58.9	59.5	59.0	59.4	58.5
	SBERT	56.6	56.0	59.4	58.9	59.8	58.4	58.2
	UPRISE	69.4	69.7	69.5	69.2	69.2	69.3	69.4
	Se^2	73.5	73.2	73.1	73.0	72.8	72.6	73.0
	SeDPO	77.6	77.9	78.0	77.9	78.2	78.1	**78.0**
Llama3-8B-Instruct (0-shot=56.4)	BM25	68.6	73.2	74.7	75.1	75.6	76.6	74.0
	SBERT	68.3	73.0	73.4	75.1	75.4	76.1	73.5
	UPRISE	70.9	75.3	76.4	76.6	76.9	77.0	75.5
	Se^2	71.9	76.7	78.0	78.0	77.9	77.9	76.7
	SeDPO	71.9	77.4	78.5	79.3	80.2	80.3	**77.9**
Llama3.3-70B (0-shot=67.6)	BM25	78.4	80.7	82.2	81.7	81.8	82.2	81.2
	SBERT	78.3	80.3	81.2	81.7	81.7	82.7	81.1
	UPRISE	77.3	80.2	81.3	80.5	80.8	81.3	80.2
	Se^2	77.9	81.0	82.0	82.2	81.9	81.9	81.1
	SeDPO	78.6	81.0	82.3	82.9	83.2	83.2	**81.9**

Showing transferability on LLMs scales and shot number.

NEURAL INFORMATION PROCESSING SYSTEMS

ZHEJIANG UNIVERSITY

WESTLAKE UNIVERSITY

Thanks

2025/11/04