Supplementary Material of Improving Few-Shot Entity Resolution with Large Language Models

Anonymous submission

1 Overview

In the supplementary material, we mainly provide (1) a running example of entity resolution; (2) an example of data enrichment using LLMs along with our observation; (3) detailed hyper-parameter configurations used in our experiment; (4) the notation table from our full paper; (5) comprehensive descriptions of the datasets used in our experiments; (6) experimental results for entity blocking across all datasets, comparing with other baselines in various settings; (7) a detailed comparison with online model, *e.g.,* ChatGPT, and (8) detailed information of prompts used in PUER.

2 Examples of Entity Resolution

We present an example of entity resolution in Table 1 (left table, denoted as $\mathcal{R}_l$) and Table 2 (right table, denoted as $\mathcal{R}_r$). Both tables contain multiple tuples with three fundamental attributes, specifically $\bar{A} = \{\mathsf{Manufacturer}, \mathsf{price}, \mathsf{title}\}$. Among these, the pairs l_{1162} and r_{2109}, as well as l_{587} and r_{2816}, represent the same real-world entities. The objective of the entity resolution task is to efficiently and effectively identify all matching tuple pairs between $\mathcal{R}_l$ and $\mathcal{R}_r$.

3 Examples of Data Enrichment

Schema Enrichment

In Table 3 and Table 4, we first use LLMs to enrich the data with 6 additional attributes, denoted as $\bar{B}$, where $\bar{B} = \{\mathsf{category}, \mathsf{subcategory}, \mathsf{platform}, \mathsf{edition}, \mathsf{type}, \mathsf{modelno}\}$. We then query LLMs for each tuple using the attributes in $\bar{B}$. The values of the columns highlighted in blue are imputed by LLMs. These imputed values provide PUER more valuable information to identify matched tuple pairs.

Our Observation

In Table 3, we demonstrate that l_{1162} is imputed with different values for $\bar{B}$, specifically l^1_{1162} and l^3_{1162}, depending on whether it is paired with r_{2109} or r_{2816} (in Table 4). Based on these observations, we can enhance the data quality of training data in our Positive-Unlabeled (PU) setting. Such **pairwise enrich** problem can activate the ability of LLM generating different information from various perspective and context.

4 Hyper-parameter Configuration

Please check Table 10 for detailed hyper-parameter configuration and corresponding explanations. Our code is available at https://anonymous.4open.science/r/PUER-CB71.

We apply LLaMA-Factory (Zheng et al. 2024) for training, and apply vLLM (Kwon et al. 2023) for efficient inference. To stabilize the output result for Matcher and Selector, during inference with vLLM, for querying LLM, we set the temperature to 0, and top-p to 1 for deterministic output.

We also incorporate outlines (Willard and Louf 2023) to fix the output of LLMs to JSON format.

5 Notation Table

In Table 5, we provide notation table and their corresponding descriptions.

6 The dataset descriptions

In Table 7, we provide descriptions of all benchmark datasets used in this paper. Following our PU learning setting, for each dataset we only use 50 random sampled positive samples for PUER.

- The column *# Dataset* lists all datasets used in this paper along with their abbreviation. For the WDC dataset (*w.r.t.* WS, COM, CA, SH, WAT), we sampled 50 positive tuple pairs within the small size (1/20 of all pairs, following (Mudgal et al. 2018; Li et al. 2020)) of each dataset.
- The column *# All* provides the total number of labeled examples for each dataset, and the column *Match* specifies the number of matched examples for each dataset, including the train/valid/test splits.
- The column *# of Original Attr* shows the number of attributes in each original dataset, and the column *# of Enriched Attr* displays the number of attributes for enrichment by PUER, *i.e.,* $|\bar{A}| + |\bar{B}|$. A detailed example for dataset is provided in Figure 3 and 4.
- The column *# of* $|\mathcal{R}_l|, |\mathcal{R}_r|$ indicates the sizes of left and right tables for each dataset, respectively.
- The column *Proportion of PU* represents the ratio of PU positive sample(*e.g.,* 50) to all labeled training samples in benchmark datasets, while the column *Proportion of*

id	title	Manufacturer	price
l_{1162}	motu digital performer 5 digital audio software competitive upgrade (mac only)	motu	395.0
l_{587}	microsoft word 2007 version upgrade	microsoft	109.95

Table 1: Examples of Amazon dataset with basic attributes $\bar{A}$

id	title	Manufacturer	price
r_{2816}	microsoft word 2007 upgrade (pc)	null	109.95
r_{2109}	motu digital performer dp5 software music production software	null	319.95

Table 2: Examples of Google dataset with basic attributes $\bar{A}$

Positive Samples represents the ratio of PU positive sample(*e.g.,* 50) to all labeled positive training samples.

7 Full Version of Blocking Result

Table 8 shows full version of blocking result in all datasets. PUEL shows the superior performance, *i.e.,* highest values of PC and PQ and the smallest value of K in most cases.

8 Detailed Comparison with Online Model

Table 6 shows the comparison of PUER and other offline and online models without SFT (Supervised Fine-Tuning). The results demonstrate the effectiveness of co-training of Matcher and Selector subtasks in our proposed method PUER. All performances are evaluated under the same In-Context Learning settings, using an equal number of positive and negative samples for demonstration.

In Table 6, the upper section includes our methods PUER and PUER without SFT, while the lower section follows the setting from (Wang et al. 2024a).

9 PC/PQ Curve for Blocking Experiment

Figure 5 shows the performance of our RAG blocker in PUER in terms of PC with respect to Top-K Recall for different values of K. A curve approaching the **upper left corner** of the figure indicates better performance. The results show that the RAG blocker of PUER is highly effective, capable of retrieving the smallest number of candidate tuples while achieving the highest recall.

10 CSSR Curve for Blocking Experiment varying K

Figure 6 provide the Blocker performance(in PC, *w.r.t.* Top-K Recall) under different K. The curve approaching the **lower right corner** of the figure indicates better performance. PUER also performs the best among all baseliens in most datasets.

11 Example of the Schema Enrichment Prompt $\mathsf{pt}_{\mathsf{SE}}$

In Prompt Template 1, **Entity 1** l_{1162} is from Table 3(left table *Amazon*), and **Entity 2** r_{2109} is from Table 4 (right table *Google*).

For each dataset, we query LLM using the same prompt $\mathsf{pt}_{\mathsf{SE}}$ with varying different **Entity 1** and **Entity 2** multiple times. We then apply majority voting to the different generated attributes to determine $\bar{B}$.

12 Example of the Data Enrichment Prompt $\mathsf{pt}_{\mathsf{enr}}$

Prompt Template 3 provides an example of $\mathsf{pt}_{\mathsf{enr}}$ using the Amazon-Google dataset. The enriched attribute set $\bar{B}$ is obtained from the previous step using $\mathsf{pt}_{\mathsf{SE}}$.

$\mathsf{pt}_{\mathsf{enr}}$ is queried with various combination of entity pairs, *e.g.,* $(l_{1162}, l_{587}), (l_{1162}, r_{2109}), (l_{587}, r_{2816}), (l_{587}, r_{2109})$ to generate different values of $\bar{B}$.

13 Example of the Subtask Matcher Prompt pt_{m}

Prompt Template 2 provides an example for the Matcher subtask using the DBLP-Scholar (DS) dataset. **Paper 1** and **Paper 2** both contain enriched attributes that are extracted in the previous step using the prompt $\mathsf{pt}_{\mathsf{enr}}$.

14 Example of the Subtask Selector Prompt pt_{s}

Prompt Template 4 provides an example for the Selector subtask using the Amazon-Google (AG) dataset. **Entity 1** and **Candidate** already contain enriched attributes extracted in the previous step using the enrichment prompt $\mathsf{pt}_{\mathsf{enr}}$. Additionally, **Candidate** entities are also retrieved and ranked using the preceding Blocker component, *i.e.,* $\mathcal{F}_{\mathsf{RAG}}$.

id	title	Manufacturer	price	category	sub-category	platform	edition	type	modelno
l^1_{1162}	motu digital performer 5 digital audio software competitive upgrade (mac only)	motu	395.0	Audio Production	DAWs	Mac	Competitive Upgrade	Software	DP5
l^2_{1162}	motu digital performer 5 digital audio software competitive upgrade (mac only)	motu	395.0	Audio & Music Software	Audio Editing & Production	Mac	Standard	Software	5
l^3_{1162}	motu digital performer 5 digital audio software competitive upgrade (mac only)	motu	395.0	Audio Editing Software	DAW (Digital Audio Workstation)	Mac	Upgrade	Software	5
l^1_{587}	microsoft word 2007 version upgrade	microsoft	109.95	Productivity Software	Office Suites	Windows	Standard	Upgrade	2007
l^2_{587}	microsoft word 2007 version upgrade	microsoft	109.95	Productivity Software	Word Processing	Windows	home	Upgrade	2007
l^3_{587}	microsoft word 2007 version upgrade	microsoft	109.95	software	office	Windows	ultimate	Upgrade	2007

Table 3: Examples for Amazon dataset (left table for Amazon-Google dataset). Grey columns are original attributes(*w.r.t.* $\bar{A}$), and blue columns are enriched attributes(*w.r.t.* $\bar{B}$)). For each entity (*e.g.,* l_{1162}, l_{587}), we report three different enrichment outputs, to demonstrate the uncertainty of our proposed data enrichment methods.

id	title	Manufacturer	price	category	sub-category	platform	edition	type	modelno
r^1_{2816}	microsoft word 2007 upgrade (pc)	null	109.95	Productivity Software	Office Suites	Windows	Standard	Upgrade	2007
r^2_{2816}	microsoft word 2007 upgrade (pc)	null	109.95	Software	Office Suites	Windows	Upgrade	Desktop Software	2007
r^3_{2816}	microsoft word 2007 upgrade (pc)	null	109.95	Productivity Software	Word Processors	PC	Upgrade	Desktop Software	WORD2007UPG
r^1_{2109}	motu digital performer dp5 software music production software	null	319.95	Audio Production	DAWs			Software	DP5
r^2_{2109}	motu digital performer dp5 software music production software	null	319.95	Audio Production	DAWs	Mac	Pro	Software	DP5

Table 4: Examples of Google dataset (right table for Amazon-Google dataset). Grey columns are original attributes(*w.r.t.* $\bar{A}$), and blue columns are enriched attributes(*w.r.t.* $\bar{B}$)). For each entity (*e.g.,* r_{2816}, r_{2109}), we report three different enrichment outputs, to demonstrate the uncertainty of our proposed data enrichment methods.

Symbol	Description
$t, \{A_1, \cdots, A_m\}$	tuple t with multi-attributes $\{A_1, \cdots, A_m\}$
$\mathcal{P}, \mathcal{P}_{\mathsf{enr}}$	the labeled positive training dataset, and its enriched version
$\mathcal{P}_{\mathsf{RAG}}, \mathcal{N}_{\mathsf{RAG}}$	the set of potentially positive and negative tuple pairs by the RAG blocker
$\mathcal{R}_l, \mathcal{R}_r$	the left and right relational tables of multi-attribute tuples
B, m	the set of enriched attributes, the number of enriched attributes
K	the top-K most similar tuples to retrieve by the blocker
$\mathsf{NN}_K(t)$	the set of top-K most similar tuples with the tuple t
$\mathcal{F}_{\mathsf{RAG}}$	the entity blocking model of PUER
$\mathcal{F}_{\mathsf{EM}}$	the entity matching model of PUER
$\mathcal{F}_{\mathsf{EM}}^{\mathsf{M}}$	the Matcher subtask in $\mathcal{F}_{\mathsf{EM}}$
$\mathcal{F}_{\mathsf{EM}}^{\mathsf{S}}$	the Selector subtask in $\mathcal{F}_{\mathsf{EM}}$
$\mathcal{C}_s(t)$	the candidate list of the tuple t in $\mathcal{F}_{\mathsf{EM}}^{\mathsf{S}}$
$\mathcal{F}_{\mathsf{label}}$	the labeler of the Selector
$\mathcal{D}_{\mathsf{train}}$	the generated training data to fine-tune $\mathcal{F}_{\mathsf{EM}}$, including labeled and pseudo-labeled training instances
$\mathsf{pt}_m, \mathsf{pt}_s$	the prompts of Matcher and Selector
$\mathsf{pt}_{\mathsf{enr}}$	the prompt of data enrichment by LLMs
$\mathsf{pt}_{\mathsf{SE}}$	the prompt of enriching more attributes by LLMs
λ	the warmup iteration
$\mathcal{M}_{\mathsf{embed}}$	Embedding model for Blocker
S_t	pairwise enriched tuple set in right table $\mathcal{R}_r$ for t
$p_m(s, t)$	query for LLM-based Matcher, to determine whether tuple pair s, t is match or mismatch

Table 5: General notations with corresponding descriptions.

Methods/Model	AB	AG	DA	DS	WA
PUER	**90.29**	**75.06**	**97.51**	**96.52**	**88.20**
PUER w.o. SFT	39.41	66.81	89.25	77.04	49.62
Mistral-7B	40.70	37.77	24.68	28.89	55.96
Qwen2-7B	72.39	61.03	81.49	76.57	72.96
LLAMA3-8B	74.37	49.50	78.91	68.79	42.33
Mixtral-8×7B	77.67	34.76	67.20	60.09	50.57
GPT-3.5-turbo-0613	87.62	69.63	90.85	84.68	86.37

Table 6: Comparison with Online Model (F1 Score)

Dataset	Domain	# All	# Match	# of Original Attr	# of Enriched Attr	# $\lvert\mathcal{R}_l\rvert, \lvert\mathcal{R}_r\rvert$	Proportion of PU	Proportion of Positive Samples
Abt-Buy (AB)	Product	9,575	1,028	3	8	1081, 1092	0.87%	8.11%
Walmart-Amazon (WA)	Electronic	10,242	962	5	9	2554, 22074	0.81%	8.68%
Amazon-Google (AG)	Electronic	11,460	1,300	3	9	1363, 3226	0.72%	7.15%
DBLP-ACM (DA)	Citation	12,363	2,224	4	6	2616, 2294	0.67%	3.75%
DBLP-Scholar (DS)	Citation	28,707	5,347	4	6	2616, 64263	0.29%	1.56%
Company(CO)	Company	112,632	28,200	1	3	28200, 28200	0.07%	0.29%
WDC-All-Small(WS)	Product	13,436	3,516	1	6	7437, 8091	0.77%	2.69%
Computer(COM)	Electronic	3,865	1,005	1	7	2204, 2443	2.24%	8.98%
Camera(CA)	Product	2,858	752	1	7	1561, 1743	3.54%	13.62%
Shoes(SH)	Product	3,099	812	1	8	1600, 1767	3.10%	11.85%
Watch(WAT)	Product	3,181	831	1	9	1821, 1991	2.84%	10.98%

Table 7: Datasets used in our experiments, # means Number of, # Attr provide the original/enriched attribute number, Proportion of PU means the number of labeled samples divide **all train samples** in benchmark; Proportion of Positive Samples means the number of labeled positive samples in PU settings divide **all positive samples** in training samples.

	AG	AB	WA	DA	DS	WS	COM	CA	WAT	SH
DeepBlocker	85.69 / 3.67 / 20	75.19 / 3.57 / 20	90.12 / 3.39 / 10	97.21 / 82.49 / 1	90.14 / 18.43 / 10	55.00 / 0.72 / 20	61.59 / 0.80 / 20	60.37 / 0.79 / 20	28.03 / 0.30 / 20	25.86 / 0.27 / 20
Sudowoodo	90.06 / 9.80 / 8	90.37 / 28.73 / 3	90.54 / 8.53 / 4	98.92 / 84.92 / 1	90.24 / 12.30 / 15	53.04 / 0.92 / 20	68.55 / 1.12 / 20	61.96 / 1.01 / 20	26.83 / 0.35 / 20	26.23 / 0.34 / 20
STransformer	91.60 / 15.69 / 5	74.32 / 3.53 / 20	86.38 / 1.63 / 20	97.03 / 82.34 / 1	91.17 / 26.62 / 7	57.39 / 0.65 / 20	52.73 / 0.68 / 20	71.41 / 0.94 / 20	59.08 / 0.77 / 20	49.51 / 0.65 / 20
CLER	90.59 / 21.25 / 4	**94.96** / 48.88 / 2	92.14 / 13.47 / 3	98.04 / 84.40 / 1	90.72 / 30.14 / 6	63.68 / 0.91 / 20	74.91 / 1.07 / 20	60.00 / 0.95 / 20	33.21 / 0.47 / 20	30.84 / 0.44 / 20
PUER	**95.80 / 27.34 / 3**	94.06 / **89.45 / 1**	**93.76 / 17.65 / 2**	**99.72 / 84.64 / 1**	**92.79 / 31.61 / 6**	**90.35 / 1.39 / 17**	**90.84 / 2.15 / 11**	**90.55 / 2.97 / 8**	**90.49 / 1.68 / 14**	**90.51 / 1.39 / 17**

Table 8: Performance Evaluation. Following UniBlocker (Wang et al. 2024b), we report the first results (in order of PC/PQ/K) of baselines when their PC exceeds the threshold (90%). If both methods have larger PC than the threshold, we evaluate K, otherwise we evaluate their PC. If their K are the same, we evaluate their PC and PQ.

Dataset	Original Attribute $\bar{A}$	Enriched Attribute $\bar{B}$
Amazon-Google (AG)	title, manufacturer, price	category, subcategory, platform, edition, type, modelno
Abt-Buy (AB)	name, description, price	category, sku, brand, modelno, key$_f$eatures
Walmart-Amazon (WA)	title, category, brand, modelno, price	subcategory, key-features, sku, color
DBLP-ACM (DA)	title, authors, venue, year	keywords
DBLP-Scholar (DS)	title, authors, venue, year	keywords, research-area
Company	Description	CompanyName,CompanyType,ShortDescription,LongDescription
WDC-All-Small (WS)	title	category, subcategory, brand, modelno, key-features
WDC-Computer (COM)	title	category, subcategory, brand, modelno, sku, edition
WDC-Camera (CA)	title	category, subcategory, brand, modelno, sku, key-features
WDC-Shoes (SH)	title	category, sku, brand, modelno, colorway, type, edition
WDC-Watch (WAT)	title	brand, sku, gender, modelno, diameter, type, colorway , price

Table 9: Original and enriched attribute for all datasets

Hyper-Parameter	Value	Description(Optional)
Backbone model of $\mathcal{F}_{\mathsf{EM}}$	Mistral-7B-Instruct-0.2 (Jiang et al. 2023)	Applied for both Enrichment, Matcher and Selector
Backbone Model of $\mathcal{F}_{\mathsf{RAG}}$	bge-large-en-1.5 (Zhang et al. 2023)	Applied for Blocker $\mathcal{M}_{\mathsf{embed}}$
Learning Rate for $\mathcal{F}_{\mathsf{EM}}$	1e-4	
Learning Rate for $\mathcal{F}_{\mathsf{RAG}}$	1e-5	
τ	0.02	Temperature parameter for contrastive learning of $\mathcal{M}_{\mathsf{embed}}$
K	20	Range of default NN search for Blocker, controlled by pointer ptr_s, ptr_e
δ	5	Step length for each iteration of pointer ptr_s, ptr_e
λ	2	iteration of co-training
n	6	number of candidate set for Selector during DPO phase
Max Input Length of $\mathcal{F}_{\mathsf{EM}}$	2048	
Max Input Length of $\mathcal{F}_{\mathsf{RAG}}$	256	
Lora-rank	16	Lora-Rank for fine-tune $\mathcal{F}_{\mathsf{EM}}$
Training epoch	3	Epoch for fine-tune $\mathcal{F}_{\mathsf{EM}}$, $\mathcal{F}_{\mathsf{RAG}}$

Table 10: Hyper-Parameter List

Instruction for $\mathsf{pt}_{\mathsf{SE}}$

(system message) You are an AI assistant that follows instruction extremely well. User will give you a question. Your task is to answer as faithfully as you can.
(task description) Your task is to determine additional attributes for dataset Amazon-Google. By adding these attributes, you will be leaded to a more clear justification on whether Entity 1 and Entity 2 are the same entity or not.
(instruction) Your output should be in JSON format, only contain the set of enriched attributes. You should take the following Incomplete Entity 1 and Entity 2 as reference.
(input)
Entity 1: {'title': 'motu digital performer dp5 software music production software', 'manufacturer': '', 'price': 319.95}
Entity 2: {'title': 'motu digital performer 5 digital audio software competitive upgrade (mac only)', 'manufacturer': 'motu', 'price': 395.0}
(output format) Enriched Attributes:
{Attribute 1:'',Attribute 2:''}

Figure 1: Schema Enrichment Prompt $\mathsf{pt}_{\mathsf{SE}}$

Instruction for pt_{m}

(system message) You are an AI assistant that follows instruction extremely well. User will give you a question. Your task is to answer as faithfully as you can.
(task description) You are an expert in computer science and database.
Judge whether record Paper 1 from DBLP, and record Paper 2 from Google Scholar are match or mismatch (refer to the same paper or not), and choose within the given Options.
(input)
Paper 1:
{title: fast algorithms for mining association rules in large databases, authors: R Agrawal, R Srikant, venue: VLDB, year: 1994, keywords: [association rules, large databases, data mining, algorithms, Apriori algorithm, FP-growth algorithm]}

Paper 2:
{title: an efficient algorithm for mining association rules in large databases, authors: a savasere , e omiecinski , s navathe, venue: , year: 1995}
(output format)
Options: [match,mismatch]

Output format example:{Output: }

Figure 2: The Prompt pt_{m} for the Matcher Subtask

Instruction for $\mathsf{pt}_{\mathsf{enr}}$

(task description) You are an expert in e-commerce, and you are well known to various goods in Amazon platform. Enrich Entity 1 and Entity 2 with attributes: category/subcategory/platform/edition/type/modelno.
(instruction) Your output should be in JSON format, only contain the value of enriched attributes. You should take the following Incomplete Entity 1 and Entity 2 as reference.
(input)

```
Entity 1:{\’title\’: \’microsoft visio standard 2007 version upgrade\’, \’manufacturer\’: \’microsoft\’, \’price\’: 129.95}\n
Entity 2:{\’title\’: \’adobe cs3 design standard upgrade\’, \’manufacturer\’: \’\’, \’price\’: 413.99}
```

(output format)

```
{"Entity 1": {"title": "", "manufacturer": "", "price": "", "category": "", "subcategory": "", "platform": "", "edition": "", "type": "", "modelno": ""},
"Entity 2": {"title": "", "manufacturer": "", "price": "", "category": "", "subcategory": "", "platform": "", "edition": "", "type": "", "modelno": ""}}
```

Figure 3: Data Imputation (Enrichment) Prompt $\mathsf{pt}_{\mathsf{enr}}$

Instruction for pt_{s}

Task: Entity Matching.
Objective: For the given Entity 1, determine which of the numbered Entity 2 candidates refer to the same real-world entity.
Instructions for your response:
1. Specify the id of candidates that match Entity 1 within <positive>...</positive> tags, return in list format.
2. Specify the id of candidates that DO NOT match Entity 1 within <negative>...</negative> tags, return in list format.
3. Ensure all candidate indices are covered in either the positive or negative set.

(input)

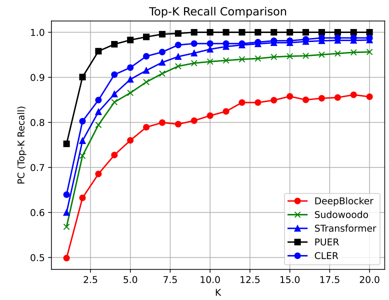
```
Entity 1:    {’id’: 574, ’title’: ’microsoft mappoint 2006 with gps’, ’manufacturer’: ’microsoft’, ’price’: 349.0}
```

Candidate Options:

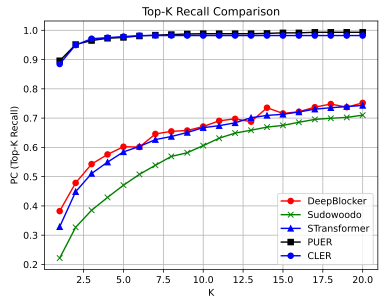
```
Entity 2 Candidates:
{’id’: 3029, ’title’: ’microsoft mappoint 2006 with gps locator ( pc )’, ’manufacturer’: ’’, ’price’: 349.99}
{’id’: 3190, ’title’: ’microsoft ( r ) mappoint ( r ) 2006’, ’manufacturer’: ’’, ’price’: 249.99}\n
{’id’: 2480, ’title’: ’microsoft b21-00806 ae mappoint 2006 cd’, ’manufacturer’: ’’, ’price’: 50.39}
{’id’: 1623, ’title’: ’language guide for nuvi 350’, ’manufacturer’: ’’, ’price’: 79.95}
```

(output format)
<think> $\cdots$ </think><positive> [1,2] </positive><negative> [3] </negative>

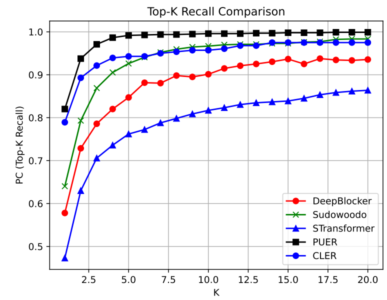
Figure 4: The Prompt pt_{s} for the Selector Subtask



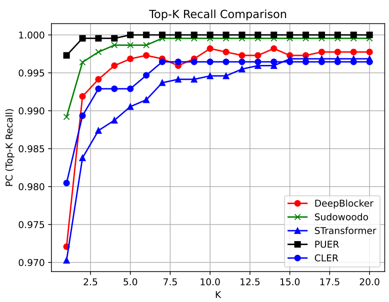
(a) Amazon-Google (varying K)



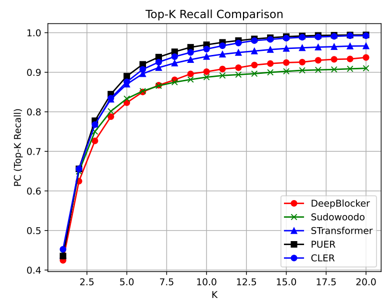
(b) Abt-Buy (varying K)



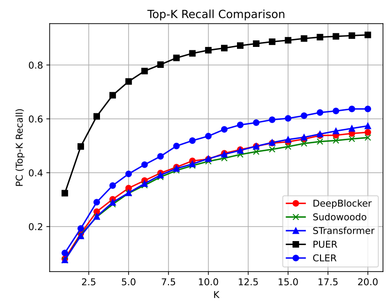
(c) Walmart-Amazon (varying K)



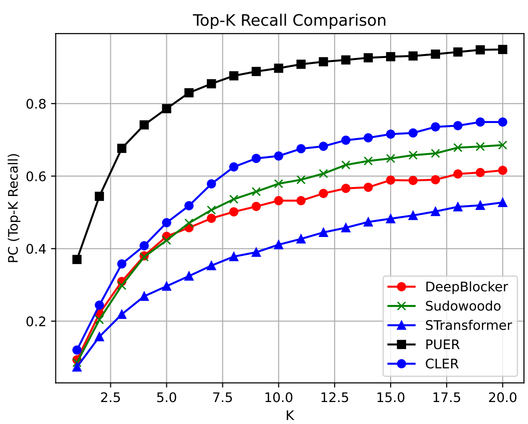
(d) DBLP-ACM (varying K)



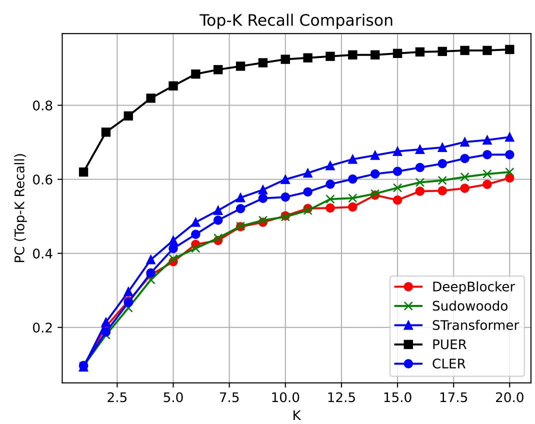
(e) DBLP-Scholar (varying K)



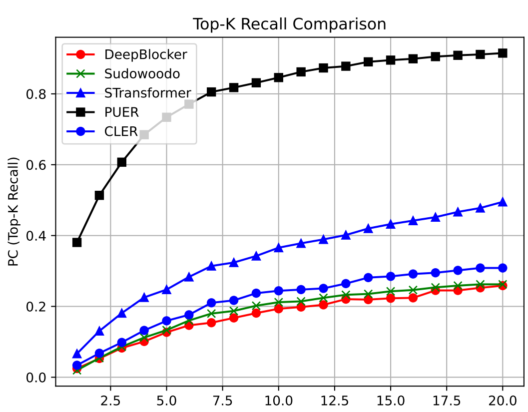
(f) WDC-All-Small (varying K)



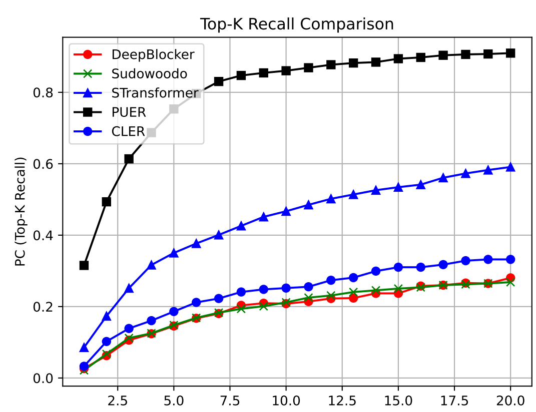
(g) WDC-Computer (varying K)



(h) WDC-Camera (varying K)

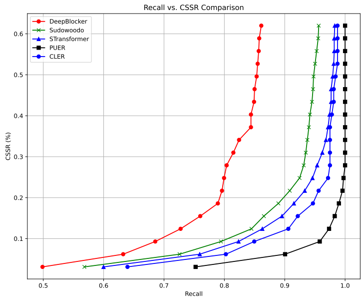


(i) WDC-Shoes (varying K)

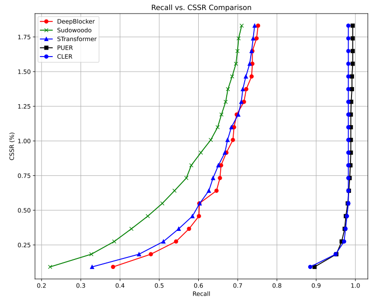


(j) WDC-Watch (varying K)

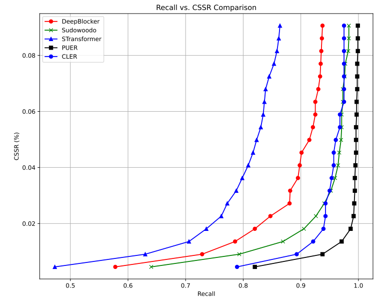
Figure 5: Effectiveness evaluation for Blocker vary K. The curve approaching the **upper left corner** of the figure indicates better performance



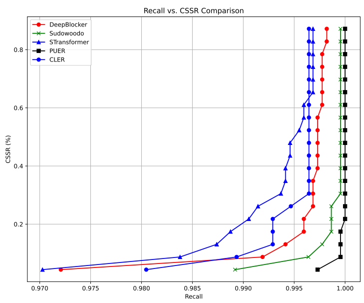
(a) Amazon-Google (varying PC)



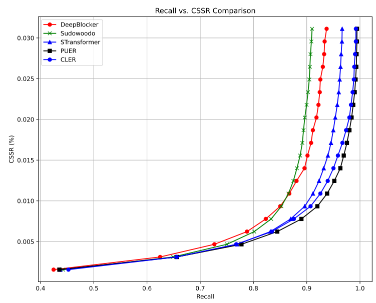
(b) Abt-Buy (varying PC)



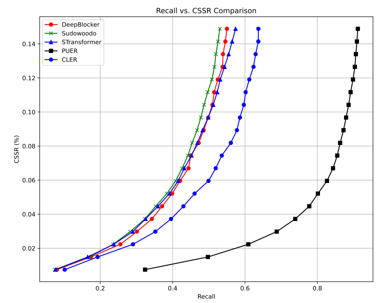
(c) Walmart-Amazon (varying PC)



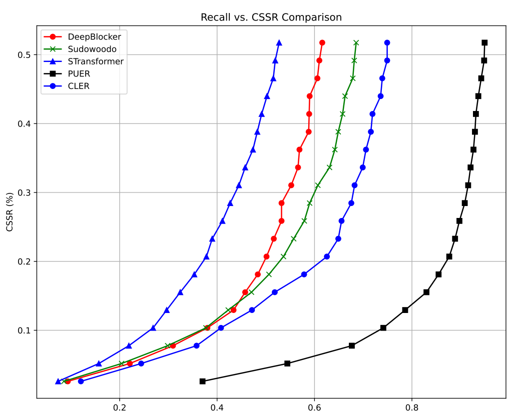
(d) DBLP-ACM (varying PC)



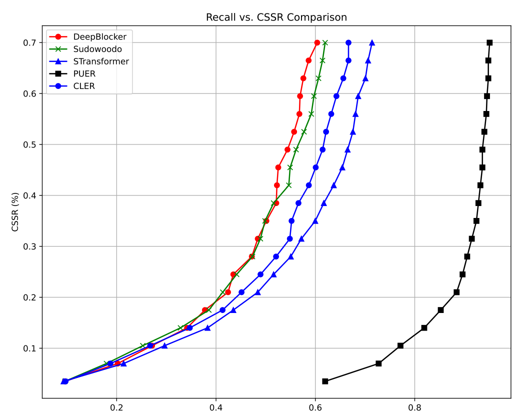
(e) DBLP-Scholar (varying PC)



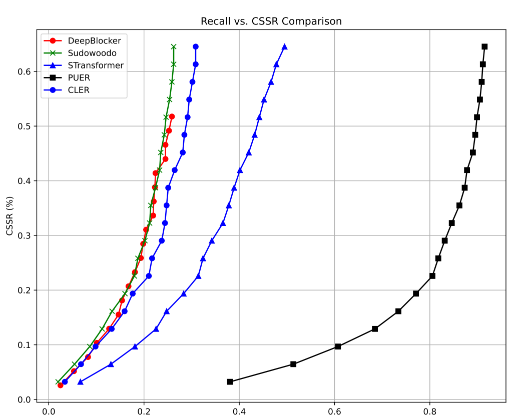
(f) WDC-All-Small (varying PC)



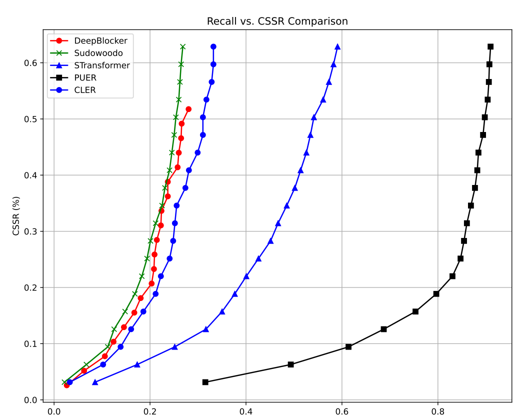
(g) WDC-Computer (varying PC)



(h) WDC-Camera (varying PC)



(i) WDC-Shoes (varying PC)



(j) WDC-Watch (varying PC)

Figure 6: Effectiveness evaluation for Blocker vary PC(*w.r.t.* Recall in figure). The curve approaching the **lower right corner** of the figure indicates better performance

References

Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; de Las Casas, D.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; Lavaud, L. R.; Lachaux, M.; Stock, P.; Scao, T. L.; Lavril, T.; Wang, T.; Lacroix, T.; and Sayed, W. E. 2023. Mistral 7B. *CoRR*, abs/2310.06825.

Kwon, W.; Li, Z.; Zhuang, S.; Sheng, Y.; Zheng, L.; Yu, C. H.; Gonzalez, J. E.; Zhang, H.; and Stoica, I. 2023. Efficient Memory Management for Large Language Model Serving with PagedAttention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.

Li, Y.; Li, J.; Suhara, Y.; Doan, A.; and Tan, W. 2020. Deep Entity Matching with Pre-Trained Language Models. *PVLDB*, 14(1): 50–60.

Mudgal, S.; Li, H.; Rekatsinas, T.; Doan, A.; Park, Y.; Krishnan, G.; Deep, R.; Arcaute, E.; and Raghavendra, V. 2018. Deep learning for entity matching: A design space exploration. In *SIGMOD*, 19–34.

Wang, T.; Lin, H.; Chen, X.; Han, X.; Wang, H.; Zeng, Z.; and Sun, L. 2024a. Match, Compare, or Select? An Investigation of Large Language Models for Entity Matching. *CoRR*, abs/2405.16884.

Wang, T.; Lin, H.; Han, X.; Chen, X.; Cao, B.; and Sun, L. 2024b. Towards Universal Dense Blocking for Entity Resolution. *CoRR*, abs/2404.14831.

Willard, B. T.; and Louf, R. 2023. Efficient Guided Generation for LLMs. *arXiv preprint arXiv:2307.09702*.

Zhang, P.; Xiao, S.; Liu, Z.; Dou, Z.; and Nie, J. 2023. Retrieve Anything To Augment Large Language Models. *CoRR*.

Zheng, Y.; Zhang, R.; Zhang, J.; Ye, Y.; Luo, Z.; Feng, Z.; and Ma, Y. 2024. LlamaFactory: Unified Efficient Fine-Tuning of 100+ Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*. Bangkok, Thailand: Association for Computational Linguistics.