Towards Uncertainty-Calibrated Structural Data Enrichment with Large Language Model for Few-Shot Entity Resolution

Mengyi Yan, Yaoshu Wang, Xiaohan Jiang, Haoyi Zhou, Jianxin Li

Frontiers of Computer Science, FCS-241143.R1

Problems & Ideas

- Problems of Conventional ER Approaches::
  - **Input Length Constraints**: Pre-trained language models (PLMs) are limited to 512 tokens, leading to truncation or summarization of long-text entities with information loss.
  - **Needle in the Haystack**: Critical information is scattered across unstructured long-text entities, making extraction challenging.
- Ideas: Combines Large Language Models (LLMs) for structural data enrichment (SDE), and uncertainty calibration module to filter hallucination generations of LLMs with its inherit parameter.

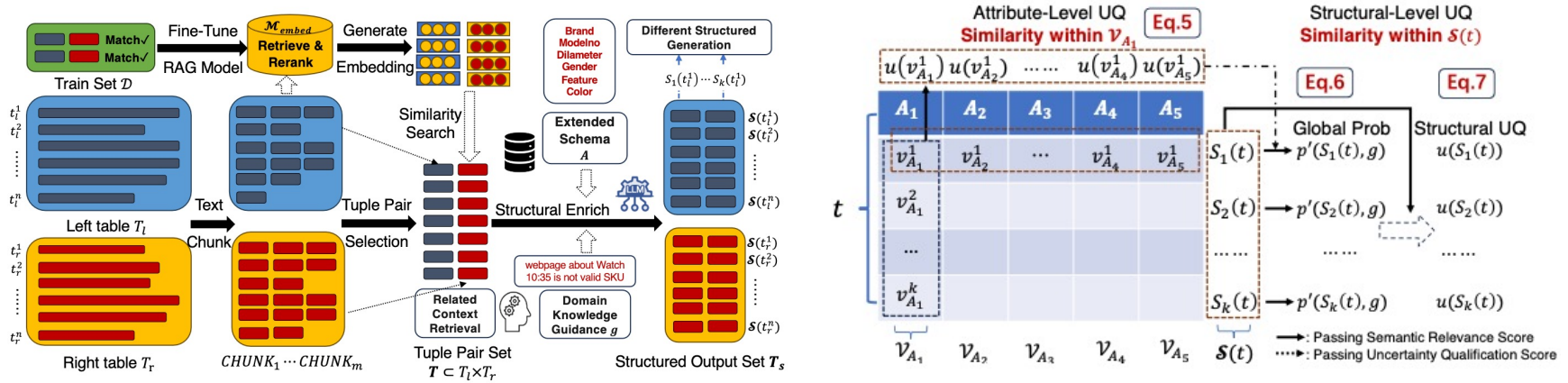

Illustration of FUSER's Structural Data Enrichment (SDE) and Uncertainty Calibration modules. The SDE framework(Left) uses a typical retrieval-augmented generation (RAG) pipeline to chunk and extract structured attributes from unstructured long-text entities, while the uncertainty calibration module (Right) evaluates attribute- and entity-level uncertainty to filter hallucinated LLM outputs.

Main Contributions

- Contributions:
  - **Structural Data Enrichment (SDE):** LLMs extract schema-aligned attributes from unstructured text via retrieval-augmented chunking, pairwise enrichment and domain knowledge injection;
  - **Uncertainty Calibration**: A two-tier module evaluates attribute- and entity-level uncertainty to filter hallucinations and ensure reliability;
  - **Lightweight ER Pipeline**: Combines enriched data with contrastive learning for blocking/matching tasks, achieving state-of-the-art performance with minimal labeled data.

Datasets	FUSER			Ditto [2]			Rotom [3]			Unicorn [47]			PromptEM [4]			JellyFish [15]		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
WA	87.36	86.01	**86.68**	78.43	20.72	32.78	11.80	73.10	20.30	89.99	60.62	72.44	93.55	30.05	45.49	80.09	93.78	86.39
AB	87.44	91.26	**89.31**	97.24	51.45	67.30	14.60	42.70	21.70	97.11	49.02	65.16	98.04	48.54	64.94	99.38	78.15	87.50
CO	98.52	78.01	**87.08**	25.06	100.00	40.08	n/a	n/a	n/a	77.60	8.84	15.88	n/a	n/a	n/a	96.43	22.07	35.92
WS	91.55	94.99	**93.24**	71.89	20.28	31.64	27.4	75.00	40.2	95.97	43.73	60.09	91.55	28.05	42.94	96.64	52.92	68.39
SW	92.85	61.61	**74.07**	11.43	100.00	20.51	21.70	4.70	7.80	90.93	35.06	51.72	11.69	17.06	13.87	49.41	60.19	54.27
SC	81.85	71.67	**76.42**	13.85	100.00	24.34	23.40	16.20	19.20	99.99	39.33	56.46	17.00	13.30	14.92	77.79	74.43	76.08

Entity matching performance in comparison to the baselines, we report the F1-score in 100 scale,