

HLM-Cite: Hybrid Language Model Workflow for Text-based Scientific Citation Prediction

Qianyue Hao, Jingyang Fan, Fengli Xu*, Jian Yuan, Yong Li*
Tsinghua University

Speaker: Qianyue Hao

Backgrounds

Citation network

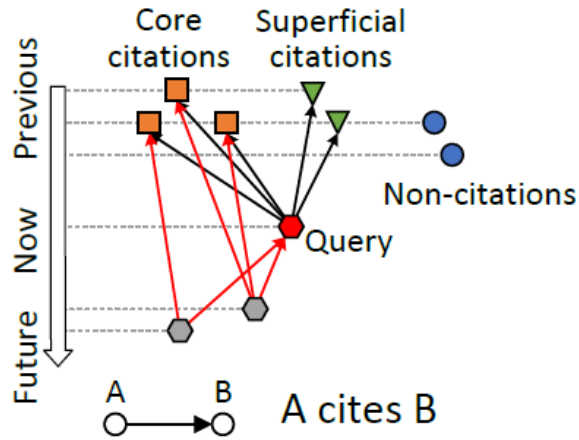
- Critical infrastructure of modern science
- Enabling researchers to navigate among the past knowledge space

Citation prediction

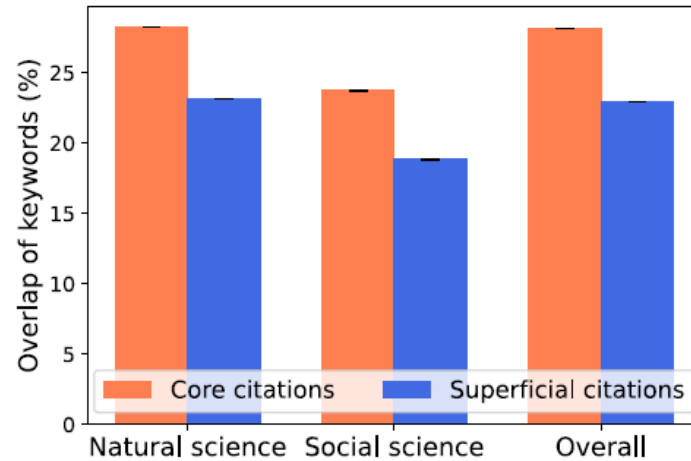
- A query paper q and a candidate set C_q , predicting which papers in C_q will q cite
- Applications: recommending citations for drafts, supporting science of science studies

Challenges

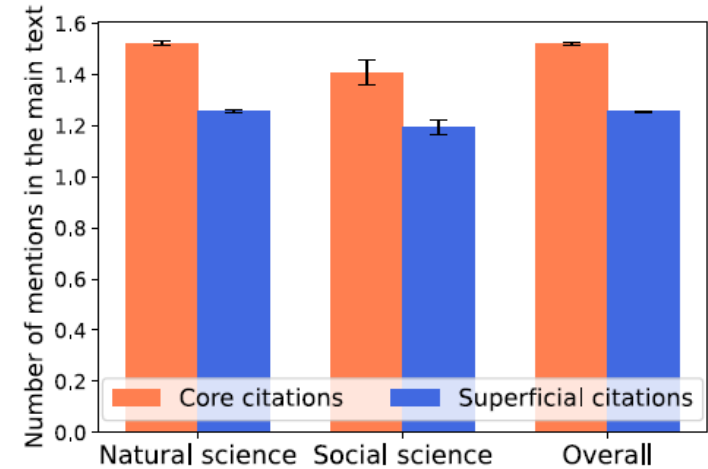
Various roles of citations



(a) Definition of **core citation**



(b) Overlap of keywords

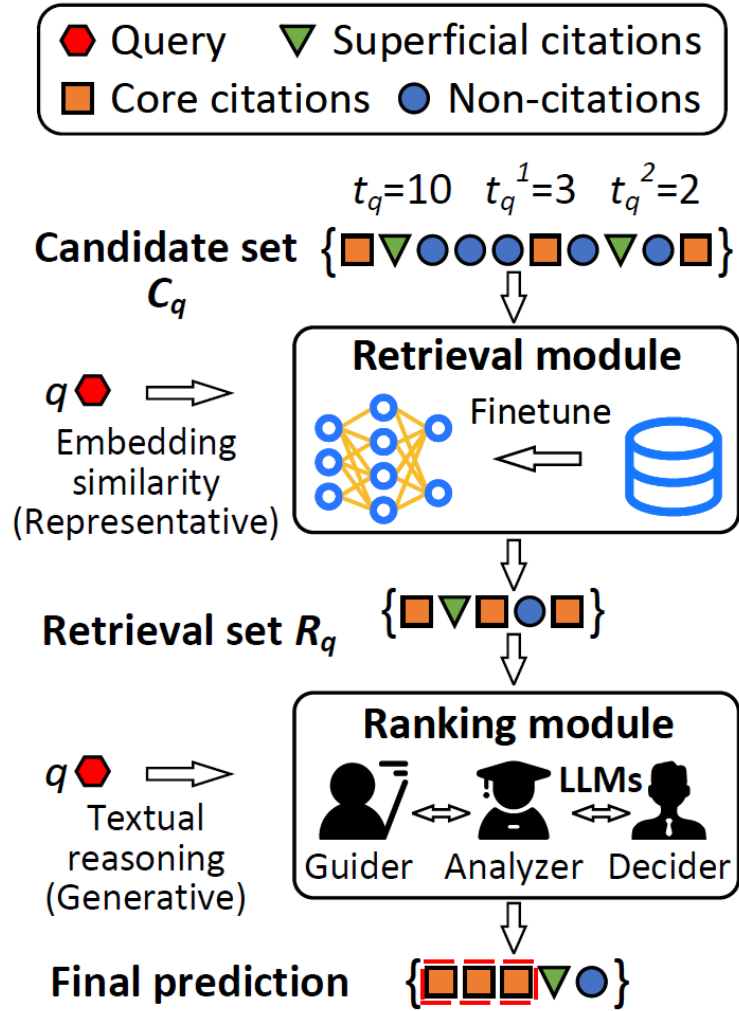


(c) Number of mentions in the main text

Vast size of the candidate set C_q

- Existing works: $10^1 \sim 10^2$ Our work: **$10^4 \sim 10^5$**

Methods



Phase 1-Retrieval

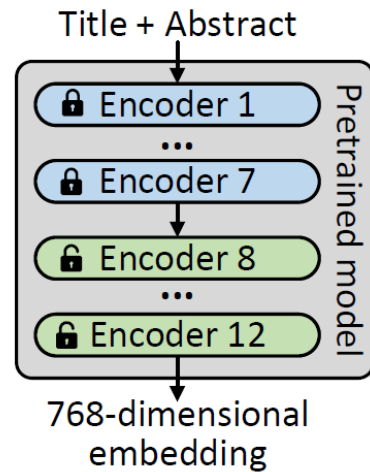
- Embedding model (<1B, finetuned)
- Candidate set ($10^4 \sim 10^5$) -> Retrieval set (10^1)

Phase 2-Ranking

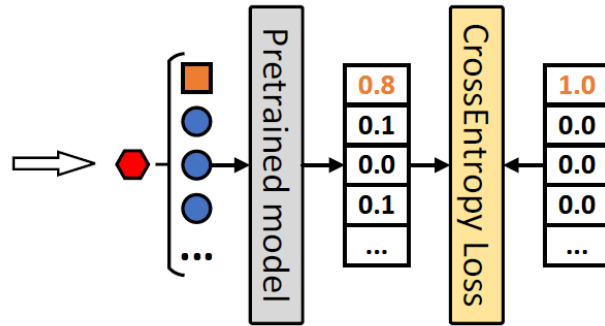
- LLM agents workflow (GPT-4 via API call)
- Retrieval set (10^1) -> Final results

Methods

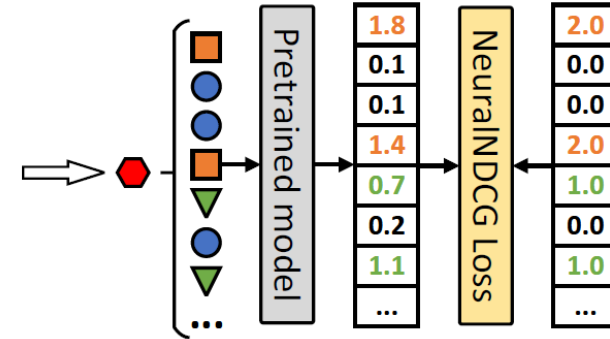
Phase 1-Retrieval



Stage 1: Classification task

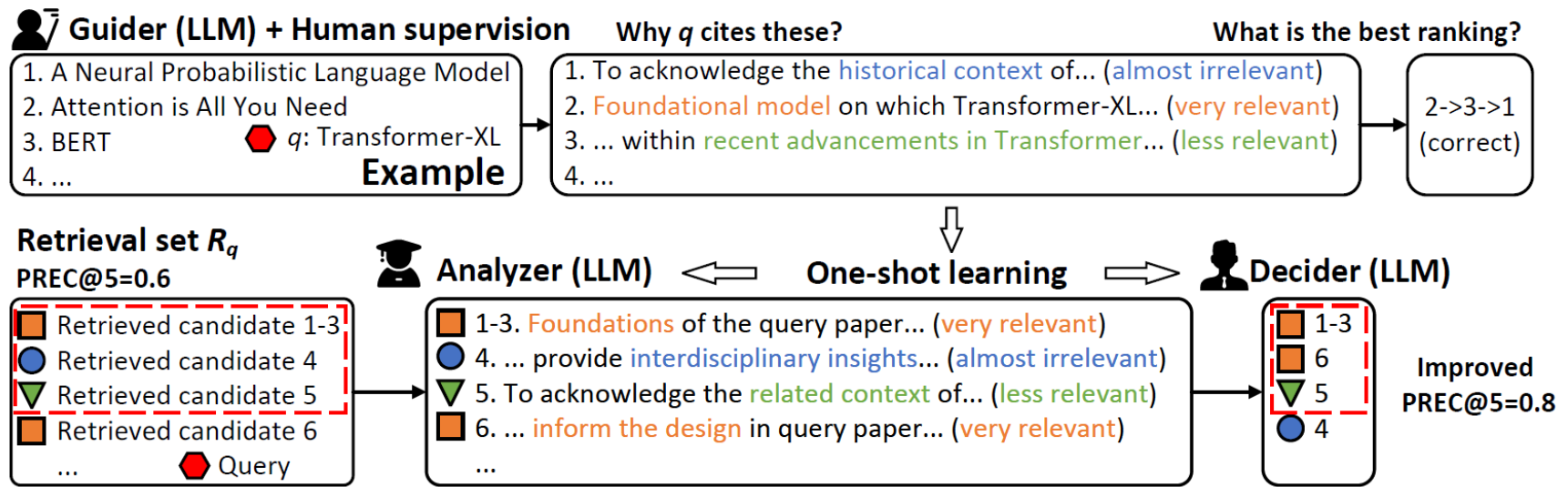


Stage 2: Ranking task



(b) Curriculum finetuning of the retrieval module

Phase 2-Ranking



(c) LLM agentic ranking module

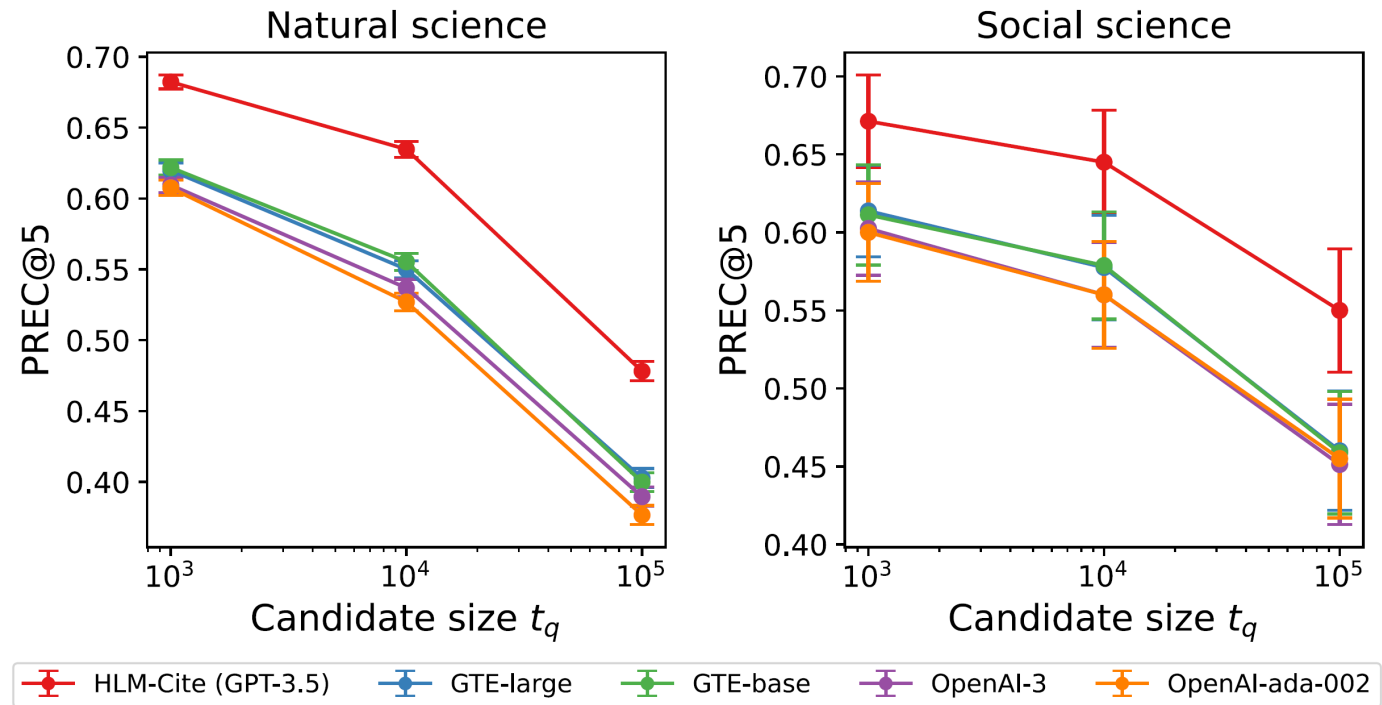
Results

Scientific domain	Training set		Testing set		Total
	Query	Candidate	Query	Candidate	
Natural science	386655	3830273	48388	479596	4744912
Social science	13345	169727	1612	20404	205088
Total	400000	4000000	50000	500000	4950000

Superior performance in multiple domains

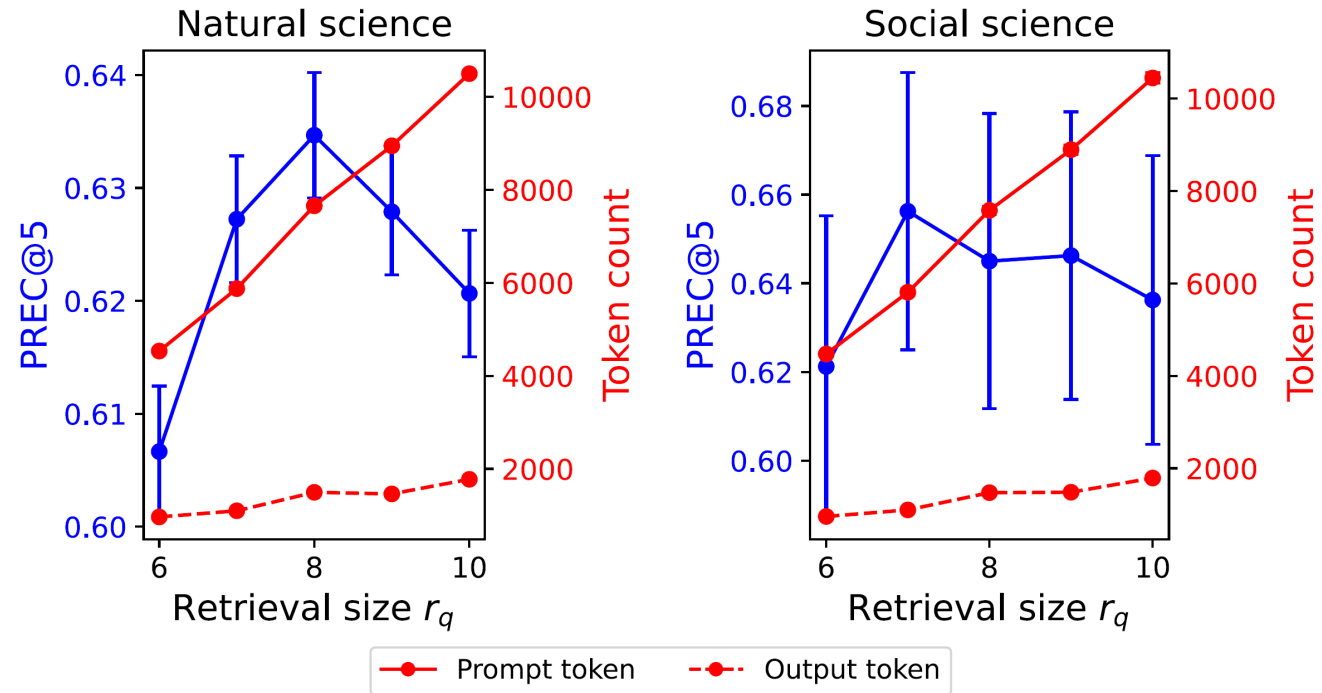
Model	Natural science				Social science				Overall			
	PREC@3/5	NDCG@3/5	PREC@3/5	NDCG@3/5	PREC@3/5	NDCG@3/5	PREC@3/5	NDCG@3/5	PREC@3/5	NDCG@3/5	PREC@3/5	NDCG@3/5
Keywords overlap	0.334	0.267	0.302	0.262	0.402	0.322	0.359	0.311	0.336	0.269	0.304	0.264
SciBERT [31]	0.053	0.046	0.056	0.050	0.083	0.069	0.087	0.076	0.054	0.046	0.057	0.051
METAG [6]	0.112	0.089	0.124	0.104	0.180	0.142	0.196	0.166	0.114	0.090	0.126	0.106
PATTON [5]	0.248	0.201	0.266	0.229	0.407	0.341	0.429	0.378	0.253	0.205	0.271	0.234
SciPATTON [5]	0.444	0.368	0.470	0.410	0.529	0.448	0.548	0.487	0.447	0.371	0.472	0.413
SPECTER [3]	0.542	0.457	0.567	0.502	0.620	0.537	0.641	0.579	0.545	0.460	0.570	0.504
SciNCL [4]	0.575	0.495	0.598	0.537	0.634	0.558	0.655	0.597	0.577	0.497	0.600	0.539
SciMult-vanilla [33]	0.568	0.483	0.591	0.527	0.623	0.547	0.644	0.586	0.569	0.485	0.593	0.529
SciMult-MoE [33]	0.578	0.493	0.601	0.537	0.637	0.558	0.658	0.598	0.579	0.496	0.603	0.539
SPECTER-2.0 [32]	0.600	0.512	0.625	0.558	0.654	0.579	0.674	0.617	0.602	0.515	0.627	0.560
BERT-base [18]	0.036	0.034	0.036	0.035	0.129	0.115	0.133	0.122	0.039	0.036	0.039	0.038
BERT-large [18]	0.025	0.027	0.024	0.026	0.055	0.062	0.051	0.057	0.026	0.029	0.025	0.027
OpenAI-ada-002	0.623	0.534	0.646	0.579	0.671	0.590	0.692	0.631	0.624	0.536	0.648	0.581
OpenAI-3	0.632	0.543	0.655	0.588	0.671	0.592	0.691	0.632	0.633	0.545	0.656	0.589
GTE-base [16]	0.638	0.555	0.659	0.596	0.669	0.596	0.688	0.633	0.639	0.556	0.659	0.597
GTE-base-v1.5 [34]	0.637	0.549	0.660	0.593	0.670	0.591	0.692	0.631	0.638	0.551	0.661	0.594
GTE-large [16]	0.640	0.556	0.661	0.597	0.669	0.593	0.690	0.632	0.641	0.557	0.662	0.599
GTE-large-v1.5 [34]	0.647	0.562	0.669	0.605	0.690	0.606	0.707	0.645	0.649	0.563	0.671	0.606
H-LM (GPT3.5)	0.725	0.644	0.734	0.677	0.743	0.661	0.751	0.693	0.725	0.644	0.735	0.677
H-LM (GPT4o)*	0.736	0.655	0.743	0.686	0.756	0.670	0.763	0.702	0.736	0.655	0.743	0.686

Results



- Superior performance on candidate sets with various sizes

Results



- Trade-off effect
- Optimal balancing point of small-LM and LLM

Poster Session 3, Thu 12 Dec 11 a.m. PST — 2 p.m. PST

Thank You!



arxiv



GitHub



Google Scholar



Lab Homepage