中文信息处理实验室-让机器理解中文
Chinese Information Processing Laboratory

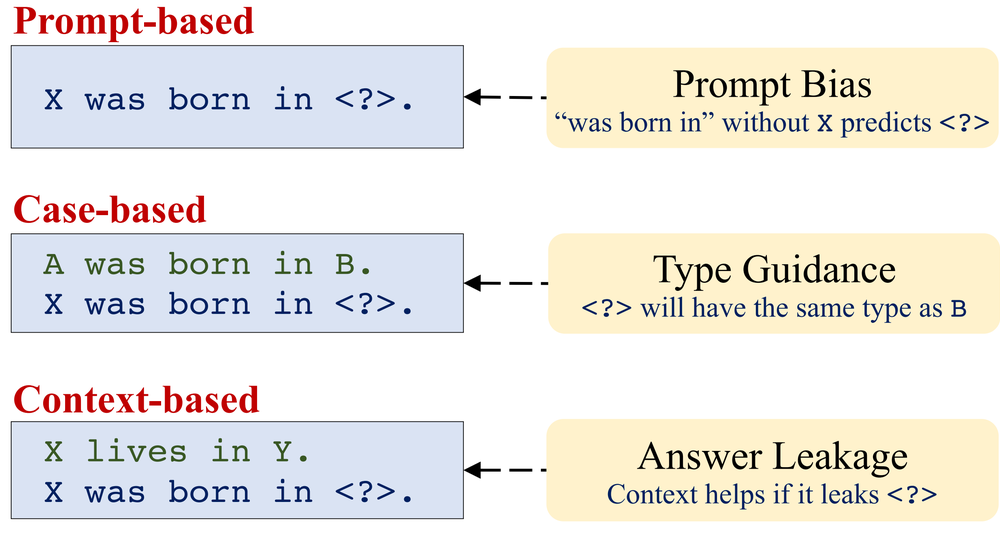
ISCAS

Knowledgeable or Educated Guess? Revisiting Language Models as Knowledge Bases

Boxi Cao[1], Hongyu Lin[1], Xianpei Han[1], Le Sun[1], Lingyong Yan[1], Meng Liao[2], Tong Xue[2], Jin Xu[2]
[1]Institute of Software, Chinese Academy of Sciences, Beijing, China
[2]Data Quality Team, WeChat, Tencent Inc., China
{boxi2020,hongyu,xianpei,sunle,lingyong2014}@iscas.ac.cn {maricoliao,xavierxue,jinxxu}@tencent.com

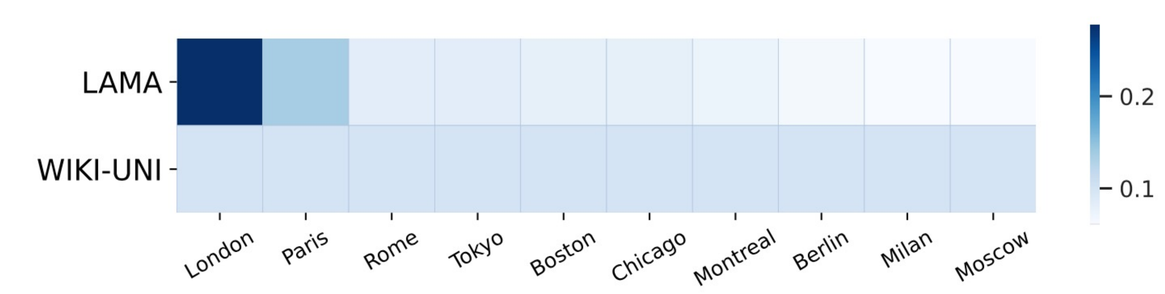
Introduction

- The underlying mechanisms behind PLMs' knowledge extraction achievements remain to be studied.
- We systematically investigate knowledge extraction from PLMs over three representative paradigms:
  - Prompt-based Retrieval
  - Case-based Analogy
  - Context-based Inference

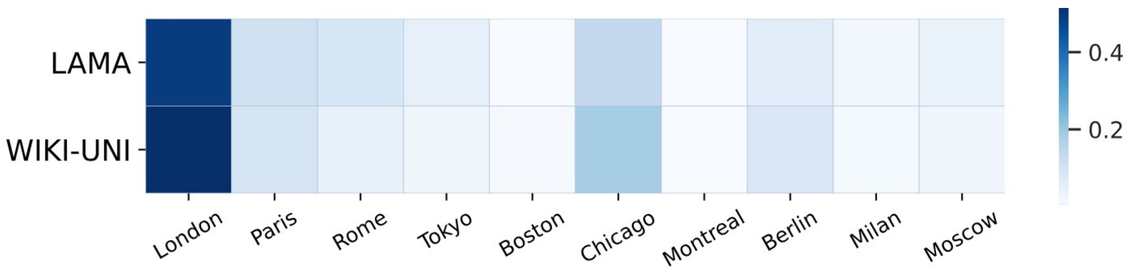


Prompt-based Retrieval

- **Prompt Bias:** The prediction distribution is severely prompt-biased.



(a) The true answer distributions are very different between LAMA and WIKI-UNI.



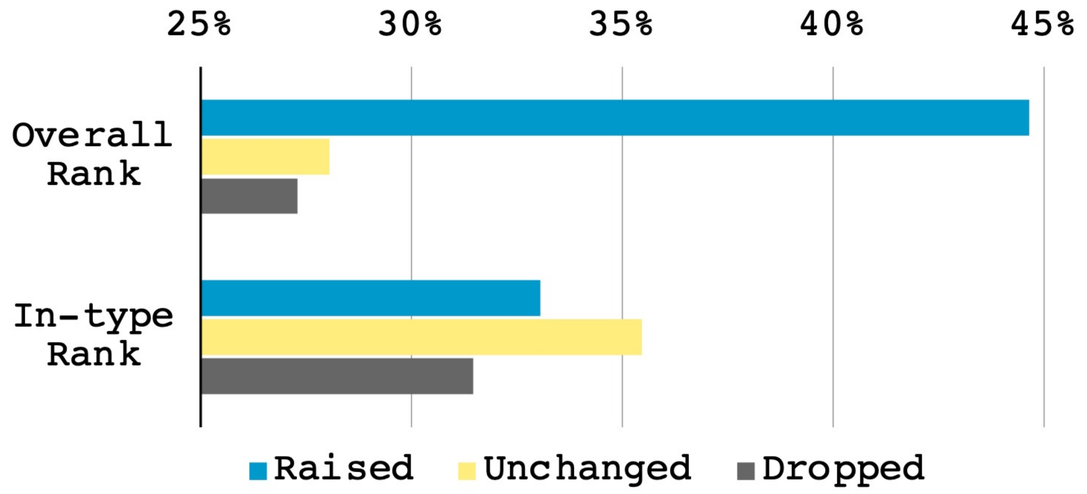
(b) However, the prediction distribution made by MLMs on them are still very similar.

- "Better" prompts are the prompts fitting the answer distribution better, rather than the prompts with better retrieval ability.

Prompt	Prec.	KL div.
T_{man}	30.36	12.27
T_{mine}	39.49	10.40
T_{auto}	40.36	10.27

Case-based Analogy

- **Type Guidance:** Illustrative cases guide MLMs to better recognizing object type, rather than better predicting facts.



Context-based Inference

- **Answer Leakage:** Additional contexts help MLMs to predict the answer because they contain the answer, explicitly or implicitly.

Answer in context	Prompt-based	Context-based	Δ
Present (45.30%)	34.83	64.13	+29.30
Absent (54.70 %)	25.37	23.26	-2.11

Answer Reconstructable	Prompt-based	Context-based	Δ
Reconstructable (60.23%)	39.58	60.82	+21.24
Not-reconstructable (39.77 %)	28.84	35.83	+6.99

Conclusion

- Previous decent performance mainly owes to the prompt bias, type guidance and answer leakage, rather than PLMs' knowledge extraction ability.