大模型时代下的知识图谱推理

吴天星
东南大学认知智能研究所
2024年9月22日

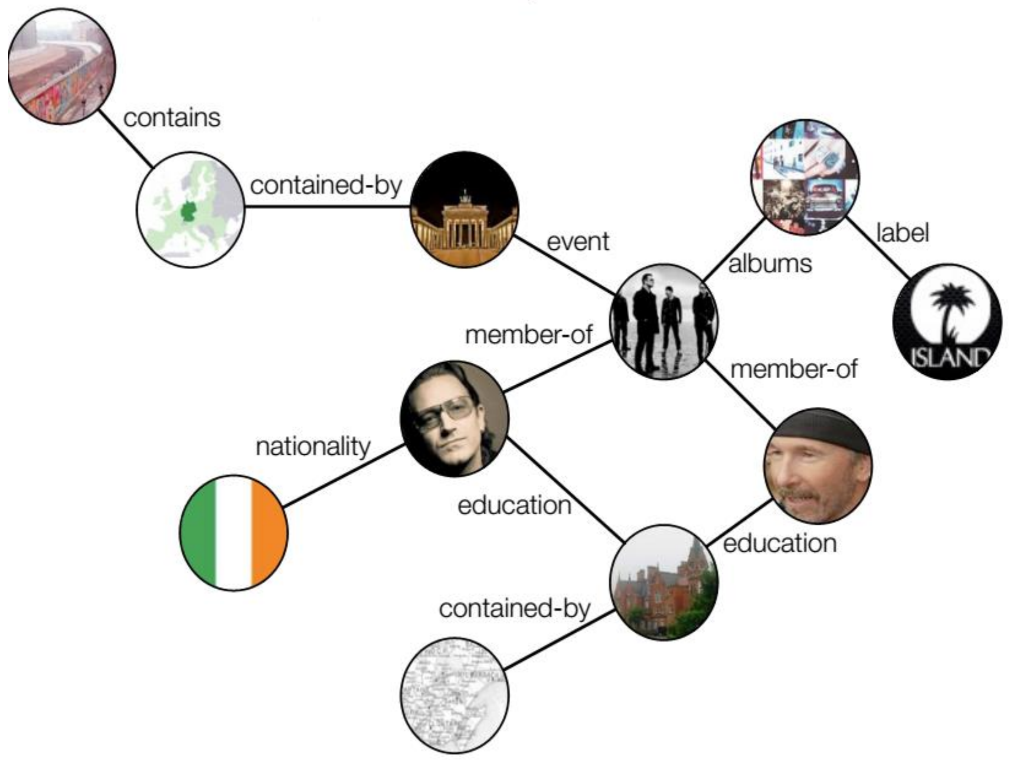
知识工程发展历史

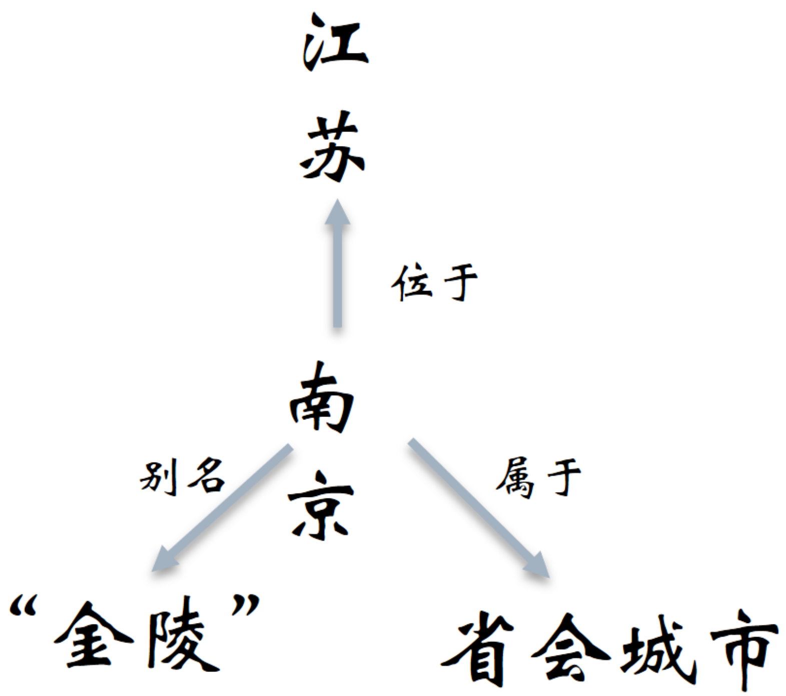
"the power of an AI program came to be seen as largely in its knowledge base"

Edward Feigenbaum, 1994 ACM Turing Award

1960s-1970s	1980s-1990s	1990s-2000s	2000s-2010s	2010s-now
Semantic Networks Quillian	Expert Systems Feigenbaum	KL-One&CLASSIC Brachman	Semantic Web Tim Berners-Lee	Knowledge Graph Google

知识图谱定义

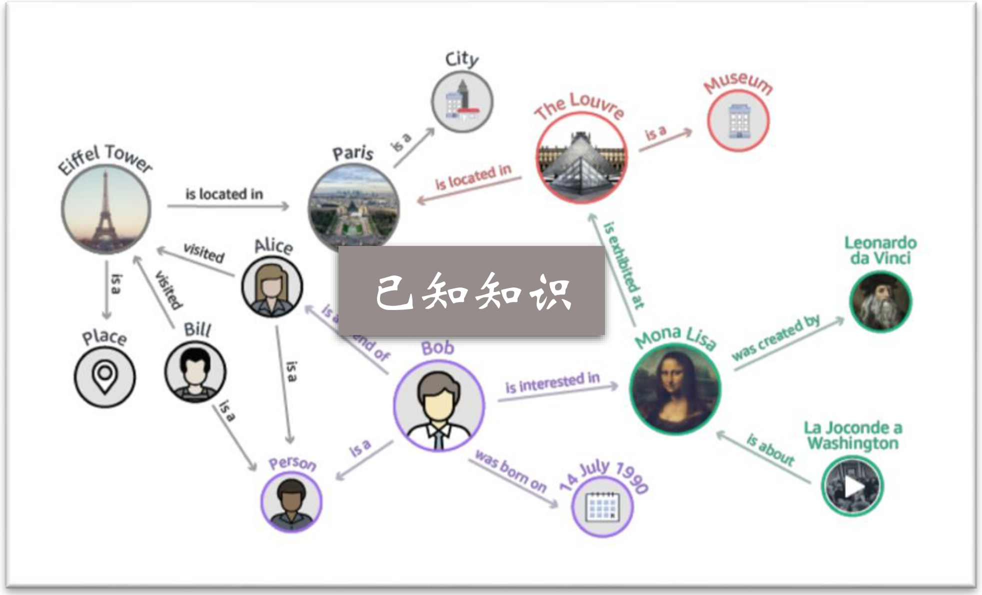




知识图谱是一个有向图 G=<V, E>， V 是实体(Entity)与字面量(Literal)的集合，E 是带标记的有向边，表示关系或属性，知识图谱的基本单位是三元组，如：<江苏，位于，南京>。

知识图谱推理

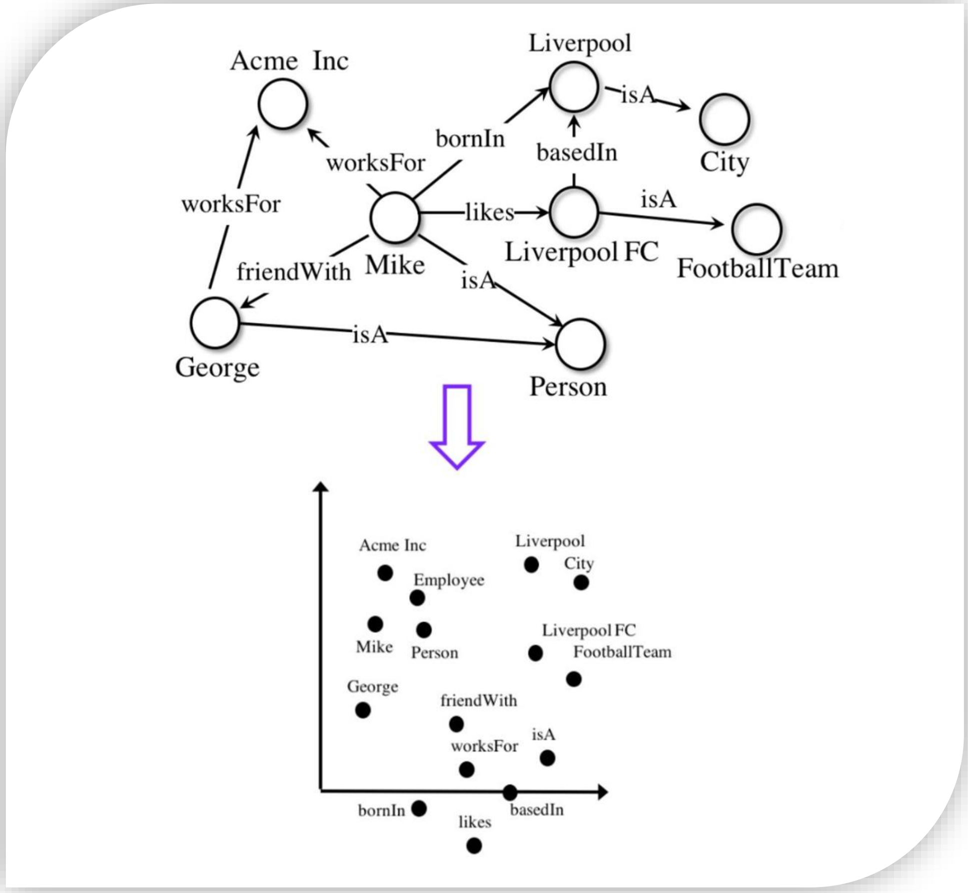
知识图谱推理旨在利用已有的知识，通过逻辑推理或统计模型等方法，从知识图谱中推导出**新的知识**。



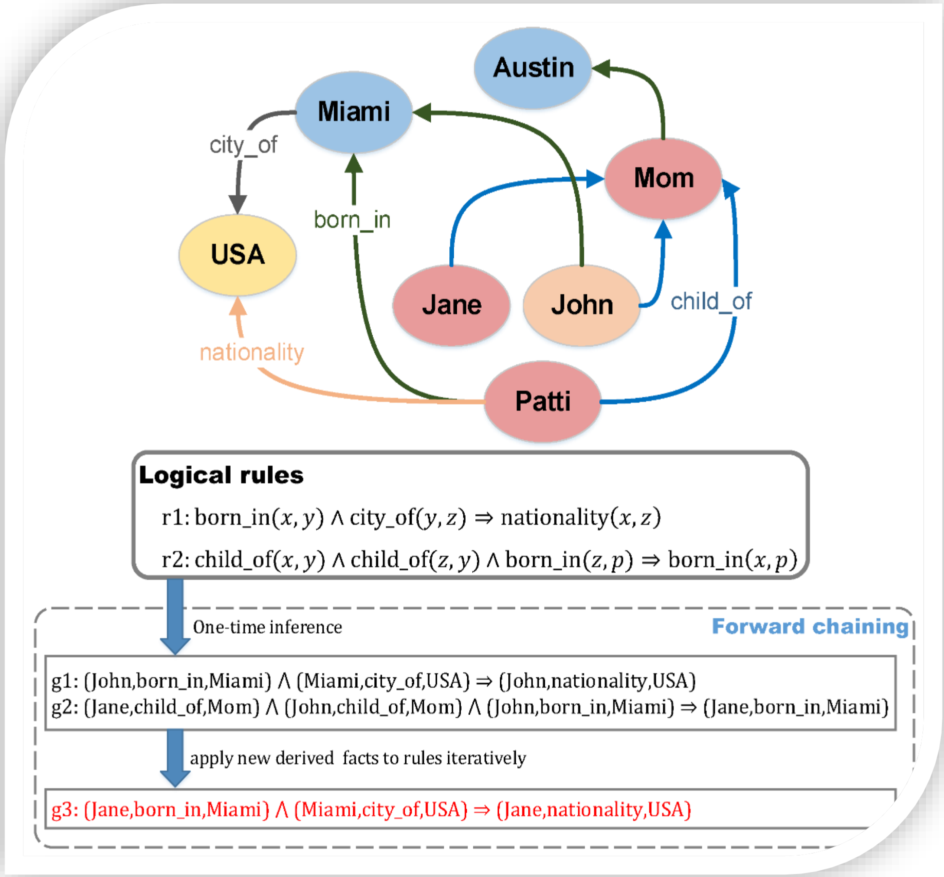
新的事实
新的公理
新的规则

- 属性补全
- 关系预测
- 错误检测
- 问题回答
- 语义理解

知识图谱统计推理分类



端到端统计推理



可解释统计推理

大模型赋能知识图谱

Why LLMs for KGs?

- 丰富的世界知识
- 任务泛化性高
- 文本理解和文本生成
- 小样本学习 (ICL)
- 复杂推理 (CoT)

相关工作汇总

東南大學 SOUTHEAST UNIVERSITY

大模型赋能知识图谱推理

预训练模型 + KG

判别式

- KG-BERT (Yao et al. 2019)
- MTL-KGC (Kim et al. 2020)
- KEPLER (Wang et al. 2021)
- StAR (Wang et al. 2021)
- MLMLM (Clouâtre et al. 2021)
- LP-BERT (Li et al. 2022)
- SimKGC (Wang et al. 2022)
- CSProm-KG (Chen et al. 2022)

生成式

- GenKGC (Wang et al. 2022)
- KGT5 (Saxena et al. 2022)
- KG-S2S (Chen et al. 2022)

大模型 + KG

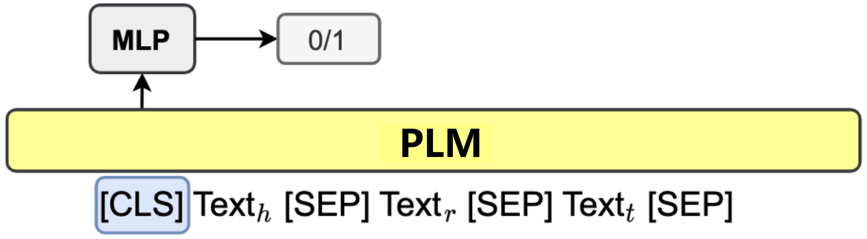
- KG-LLM (Yao et al. 2023)
- KICGPT (Wei et al. 2023)
- ChatRule (Luo et al. 2024)
- MindMap (Wen et al. 2024)
- KoPA (Zhang et al. 2024)
- KG-LLM (Shu et al. 2024)

基于预训练模型的知识图谱推理

借助预训练语言模型的上下文编码能力，利用知识图谱中的文本信息增强知识图谱的表示。

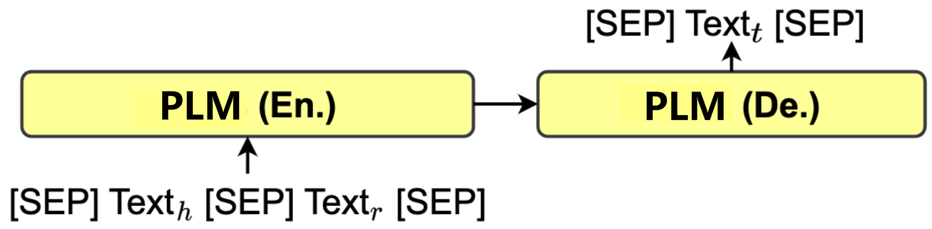
判别式方法：

- 使用encoder-only PLMs (如BERT)；
- 使用预训练模型编码文本信息与知识图谱中的事实，然后将获得的表示通过MLP或者传统的KG评分函数 (如TransE) 预测三元组的合理性。



生成式方法：

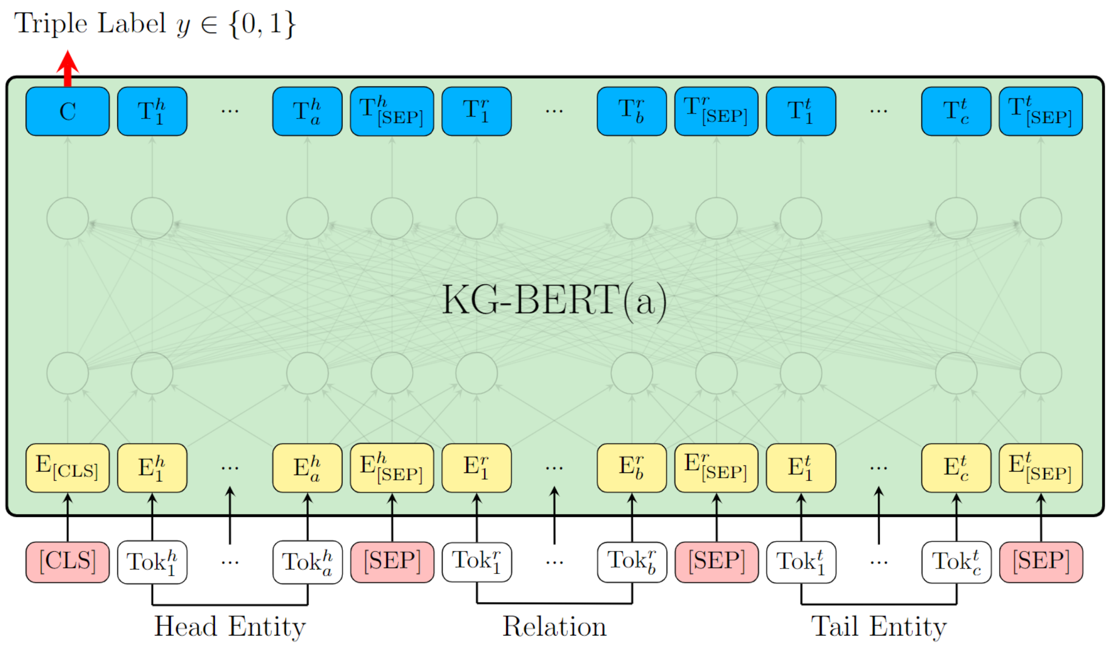
- 使用encoder-decoder或者decoder-only PLMs；
- 将知识图谱推理任务建模为sequence-to-sequence形式：将(h, r)输入PLM，直接生成尾实体t。



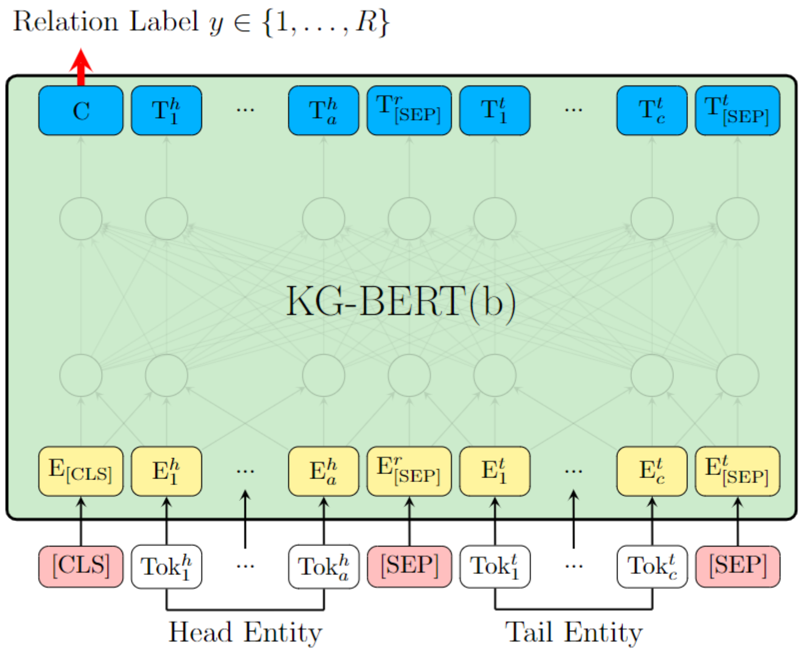
图片引用自Pan et al. Unifying Large Language Models and Knowledge Graphs: A Roadmap. IEEE TKDE 2024.

KG-BERT

将(htext, rtext, ttext)拼接为文本序列输入BERT，把知识图谱补全任务转换为序列分类任务微调BERT。



KG-BERT(a)：三元组分类



KG-BERT(b)：关系分类

Yao et al. KG-BERT: BERT for Knowledge Graph. arXiv'19

KG-BERT

Method	WN11	FB13	Avg.
NTN (Socher et al. 2013)	86.2	90.0	88.1
TransE (Wang et al. 2014b)	75.9	81.5	78.7
TransH (Wang et al. 2014b)	78.8	83.3	81.1
TransR (Lin et al. 2015b)	85.9	82.5	84.2
TransD (Ji et al. 2015)	86.4	89.1	87.8
TEKE (Wang and Li 2016)	86.1	84.2	85.2
TransG (Xiao, Huang, and Zhu 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al. 2016)	86.4	88.2	87.3
DistMult (Zhang et al. 2018)	87.1	86.2	86.7
DistMult-HRS (Zhang et al. 2018)	88.9	89.0	89.0
AATE (An et al. 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al. 2018a)	87.6	88.8	88.2
DOLORES (Wang, Kulkarni, and Wang 2018)	87.5	89.3	88.4
KG-BERT(a)	**93.5**	**90.4**	**91.9**

三元组分类结果
优于传统的KGE方法

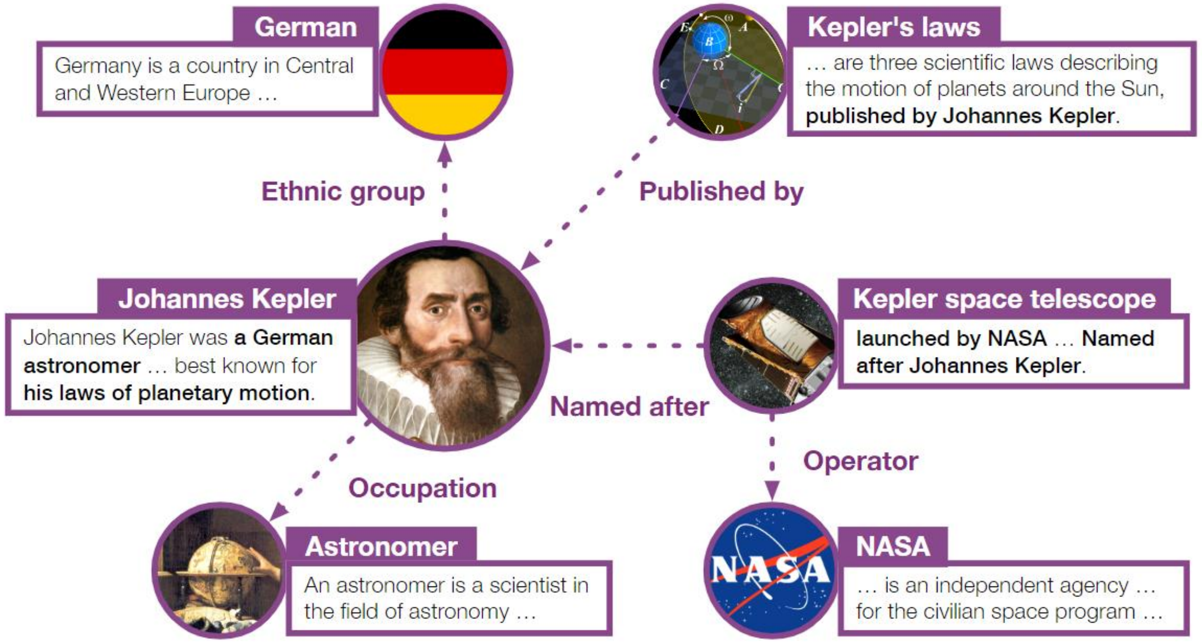
Method	WN18RR		FB15k-237		UMLS	
	MR	Hits@10	MR	Hits@10	MR	Hits@10
TransE (our results)	2365	50.5	223	47.4	1.84	98.9
TransH (our results)	2524	50.3	255	48.6	1.80	**99.5**
TransR (our results)	3166	50.7	237	51.1	1.81	99.4
TransD (our results)	2768	50.7	246	48.4	1.71	99.3
DistMult (our results)	3704	47.7	411	41.9	5.52	84.6
ComplEx (our results)	3921	48.3	508	43.4	2.59	96.7
ConvE (Dettmers et al. 2018)	5277	48	246	49.1	–	–
ConvKB (Nguyen et al. 2018a)	2554	52.5	257	51.7	–	–
R-GCN (Schlichtkrull et al. 2018)	–	–	–	41.7	–	–
KBGAN (Cai and Wang 2018)	–	48.1	–	45.8	–	–
RotatE (Sun et al. 2019)	3340	**57.1**	177	**53.3**	–	–
KG-BERT(a)	**97**	52.4	**153**	42.0	**1.47**	99.0

链接预测结果
KG-BERT可以达到较高的平均排名(MR)，
但是Hits@10指标不能达到最优。

Yao et al. KG-BERT: BERT for Knowledge Graph. arXiv’19

KEPLER

通过实体的文本描述在预训练模型和知识嵌入之间建立联系，将文本的语义空间和KG的符号空间对齐。



1) 知识嵌入方法可以为PLM提供事实知识；

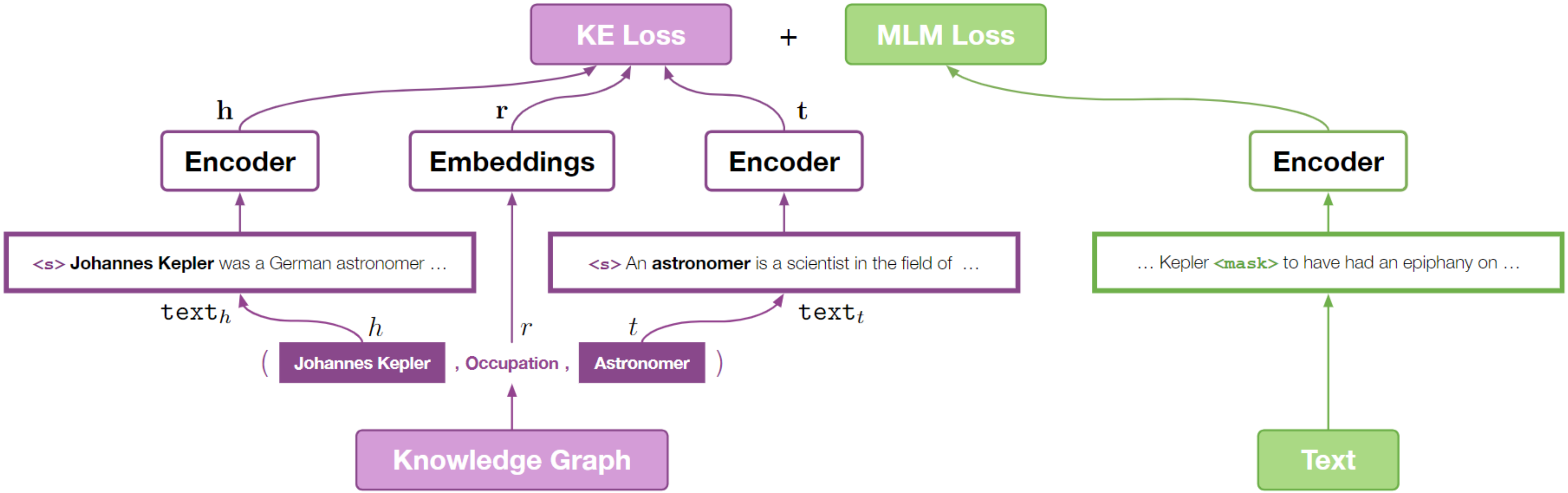
2) PLM通过处理文本知识(实体的文本描述)帮助关系事实表示。

二者可以互相促进

Wang et al. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language. TACL’21

KEPLER

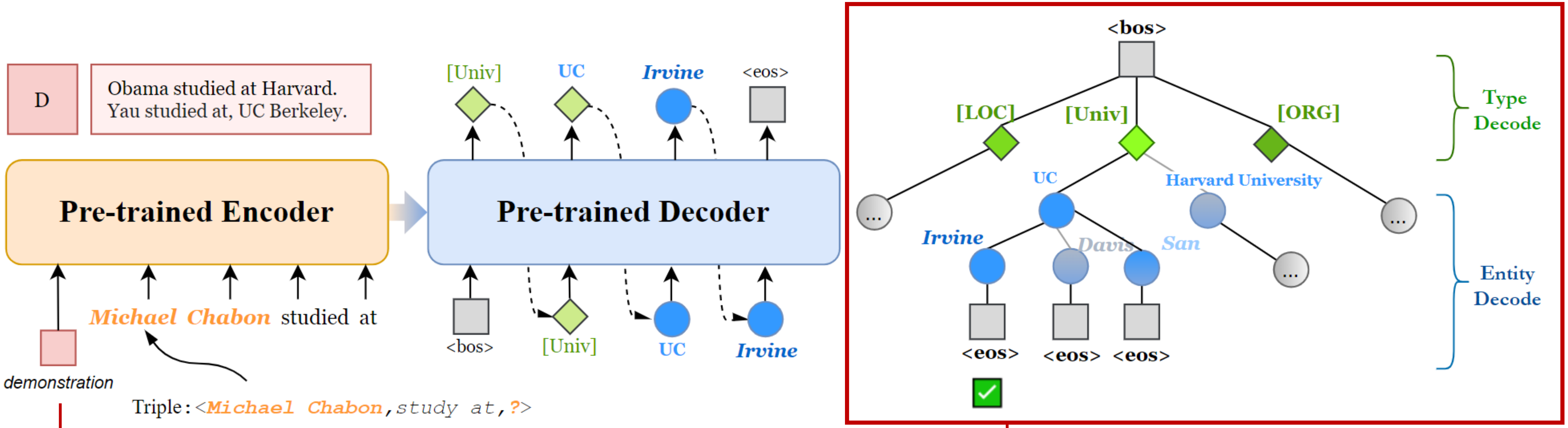
使用PLM编码实体的文本描述，联合优化知识嵌入模型(TransE)与语言模型(RoBERTa)。



Wang et al. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language. TACL’21

GenKGC

将知识图谱补全任务转换为seq2seq的生成任务，避免负采样效率低下问题。



关系引导示例：
在prompt添加相同关系的事实
引导模型进行上下文学习

实体感知层次解码：
束搜索获取top-k个实体；
通过prefix tree限制解码过程

Xie et al. From Discrimination to Generation: Knowledge Graph Completion with Generative Transformer. WWW’22

GenKGC

Method	WN18RR			FB15k-237			OpenBG500		
	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10
Graph embedding approach									
TransE [2] ◇	0.043	0.441	0.532	0.198	0.376	0.441	0.207	0.340	0.513
DistMult [13] ◇	0.412	0.470	0.504	0.199	0.301	0.446	0.049	0.088	0.216
ComplEx [11] ◇	0.409	0.469	0.530	0.194	0.297	0.450	0.053	0.120	0.266
RotatE [9]	0.428	0.492	0.571	0.241	0.375	0.533	-	-	-
TuckER [1]	0.443	0.482	0.526	0.226	0.394	0.544	-	-	-
ATTH [4]	0.443	0.499	0.486	0.252	0.384	0.549	-	-	-
Textual encoding approach									
KG-BERT [14]	0.041	0.302	0.524	-	-	0.420	0.023	0.049	0.241
StAR [12]	0.243	0.491	0.709	0.205	0.322	0.482	-	-	-
GenKGC	0.287	0.403	0.535	0.192	0.355	0.439	0.203	0.280	0.351

链接预测结果：

相比KG-BERT提高了Hits@1，但无法超越基于结构的KGE方法。

Xie et al. From Discrimination to Generation: Knowledge Graph Completion with Generative Transformer. WWW'22

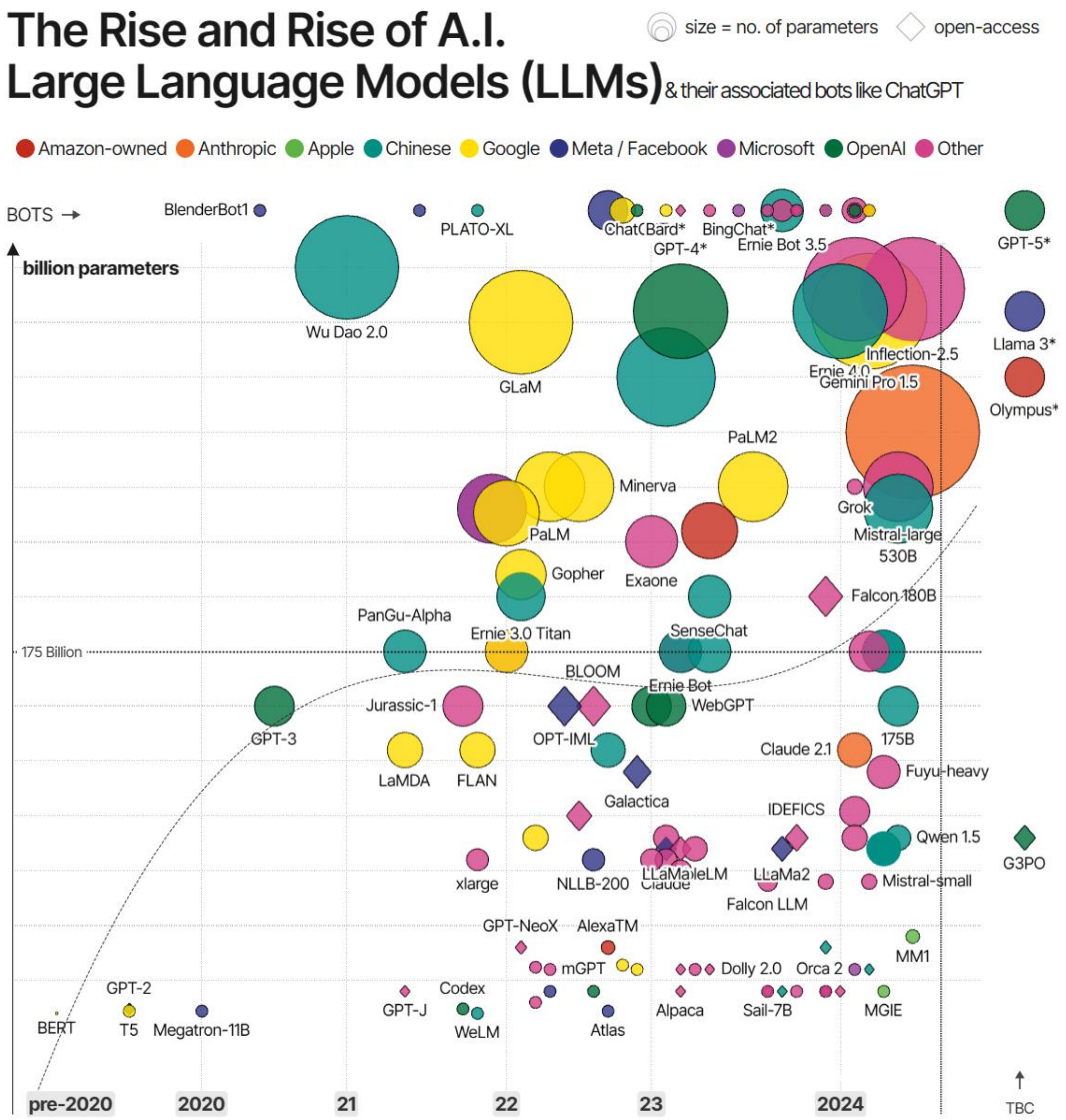
总结：

1、仅利用文本信息是不够的，还需结合图谱结构信息才能更好地执行推理任务。因此现有的方法往往同时使用预训练语言模型与传统的知识图谱嵌入方法。

2、判别式模型更容易微调，也可以与传统的知识图谱嵌入方法融合。但是在推理阶段，需要计算每个候选三元组的评分，因此计算成本高昂。而且不能推广到unseen实体。仅适用于开源PLMs。

3、生成式模型不需要获取PLM的中间表示，在推理时可以直接生成尾实体，提高了推理效率。但生成的实体可能不在KG中。

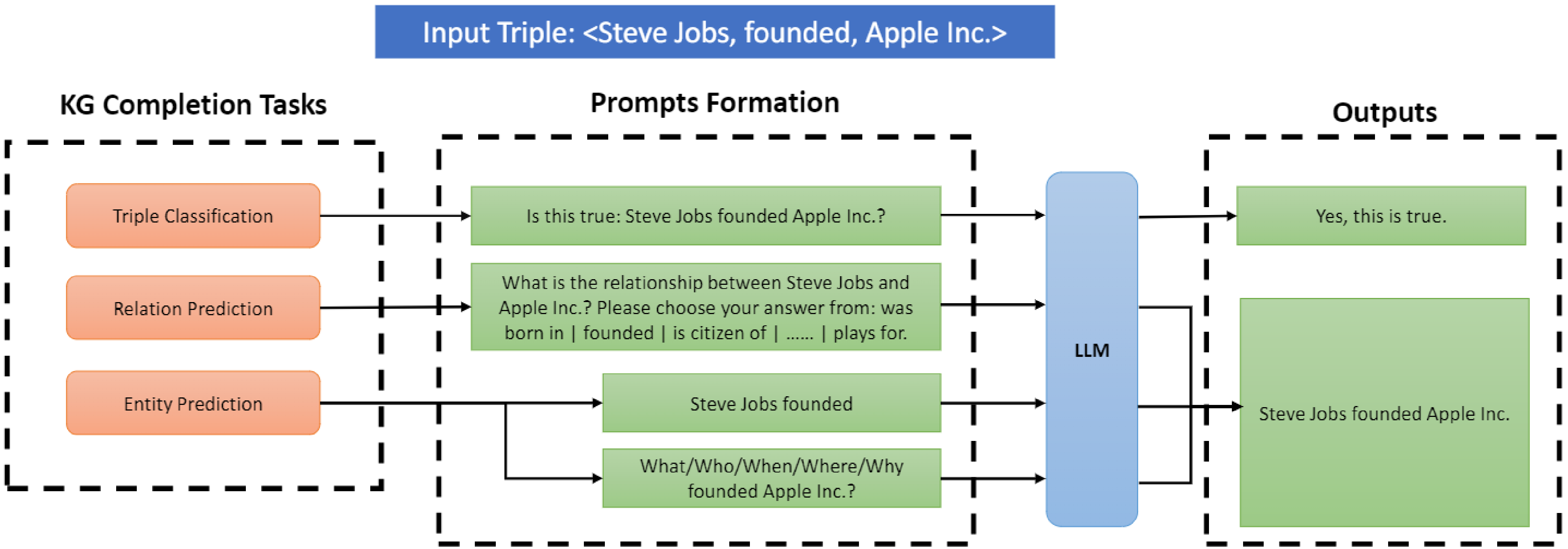
大模型的发展



图源：Information is Beautiful

Knowledge Graph LLM

针对三元组分类、关系预测、链接预测三个任务分别设计prompts，对LLM进行指令微调



Yao et al. Exploring Large Language Models for Knowledge Graph Completion. arXiv'23

Knowledge Graph LLM

Method	WN11	FB13	Avg.
NTN (Socher et al., 2013)	86.2	90.0	88.1
TransE (Wang et al., 2014b)	75.9	81.5	78.7
TransH (Wang et al., 2014b)	78.8	83.3	81.1
TransR (Lin et al., 2015)	85.9	82.5	84.2
TransD (Ji et al., 2015)	86.4	89.1	87.8
TEKE (Wang and Li, 2016)	86.1	84.2	85.2
TransG (Xiao et al., 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al., 2016)	86.4	88.2	87.3
DistMult (Zhang et al., 2018)	87.1	86.2	86.7
DistMult-HRS (Zhang et al., 2018)	88.9	89.0	89.0
AATE (An et al., 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al., 2018)	87.6	88.8	88.2
DOLORES (Wang et al., 2018)	87.5	89.3	88.4
DKRL (BERT)	87.3	79.8	83.6
KG-BERT(a) (Yao et al., 2019)	93.5	**90.4**	91.9
KGT5	72.8	66.3	69.6
LLaMA-7B	21.1	9.1	15.1
LLaMA-13B	28.1	17.6	22.9
KG-LLaMA-7B	95.5	89.2	92.4
KG-LLaMA-13B	**95.6**	90.2	**92.9**

三元组分类结果：
可以超过传统的KGE方法与PLM-based方法， 相比于未经微调的LLaMA结果有很大提升。

Method	WN18RR	YAGO3-10	YAGO3-10-100
KG-BERT(a)	0.1102	–	–
StAR (Wang et al., 2021)	0.2430	–	–
KGT5	0.1011	0.0484	0.12
ChatGPT	–	–	0.22
GPT-4	–	–	**0.24**
KG-ChatGLM-6B	0.1613	0.0455	0.11
LLaMA-7B	0.0849	0.0254	0.03
LLaMA-13B	0.0991	0.0276	0.01
KG-LLaMA-7B	0.2415	0.0782	0.16
KG-LLaMA-13B	**0.2559**	**0.0872**	0.13

链接预测结果：
指令微调提升了语言模型的表现。

Yao et al. Exploring Large Language Models for Knowledge Graph Completion. arXiv’23

Knowledge Graph LLM

Input	Label	Output	
Is this true: Everett T Moore profession Librarian?	True	GPT-4	I cannot verify specific personal information about individuals who are not public figures.
		ChatGPT	I'm sorry, but I don't have enough information to confirm whether Everett T Moore is a librarian or not.
		LLaMA-7B	I was wondering if anyone could tell me if this is true.
		LLaMA-13B	I'm a librarian at Everett T Moore Middle School in San Antonio, TX.
		KG-LLaMA-7B	Yes, this is true.
		KG-LLaMA-13B	Yes, this is true.
		ChatGLM-6B	Yes, Everett T Moore is a profession Librarian.
		KG-ChatGLM-6B	Yes, this is true.
Josip Škorić has gender	male	GPT-4	Josip Škorić is a male name. Josip is a Croatian form of the name Joseph, which is a masculine name.
		ChatGPT	Josip Škorić is a human being and therefore has a gender. However, I couldn't find any information about the gender of Josip Škorić.
		LLaMA-7B	Josip Škorić has gender issues. I'm a 22-year-old guy from Croatia, who has gender issues.
		LLaMA-13B	Josip Škorić has undergone gender reassignment surgery.
		KG-LLaMA-7B	male
		KG-LLaMA-13B	male
		ChatGLM-6B	As an AI language model, I do not have access to information about specific individuals' genders or their personal identity, including JosipŠkorić's.
		KG-ChatGLM-6B	male

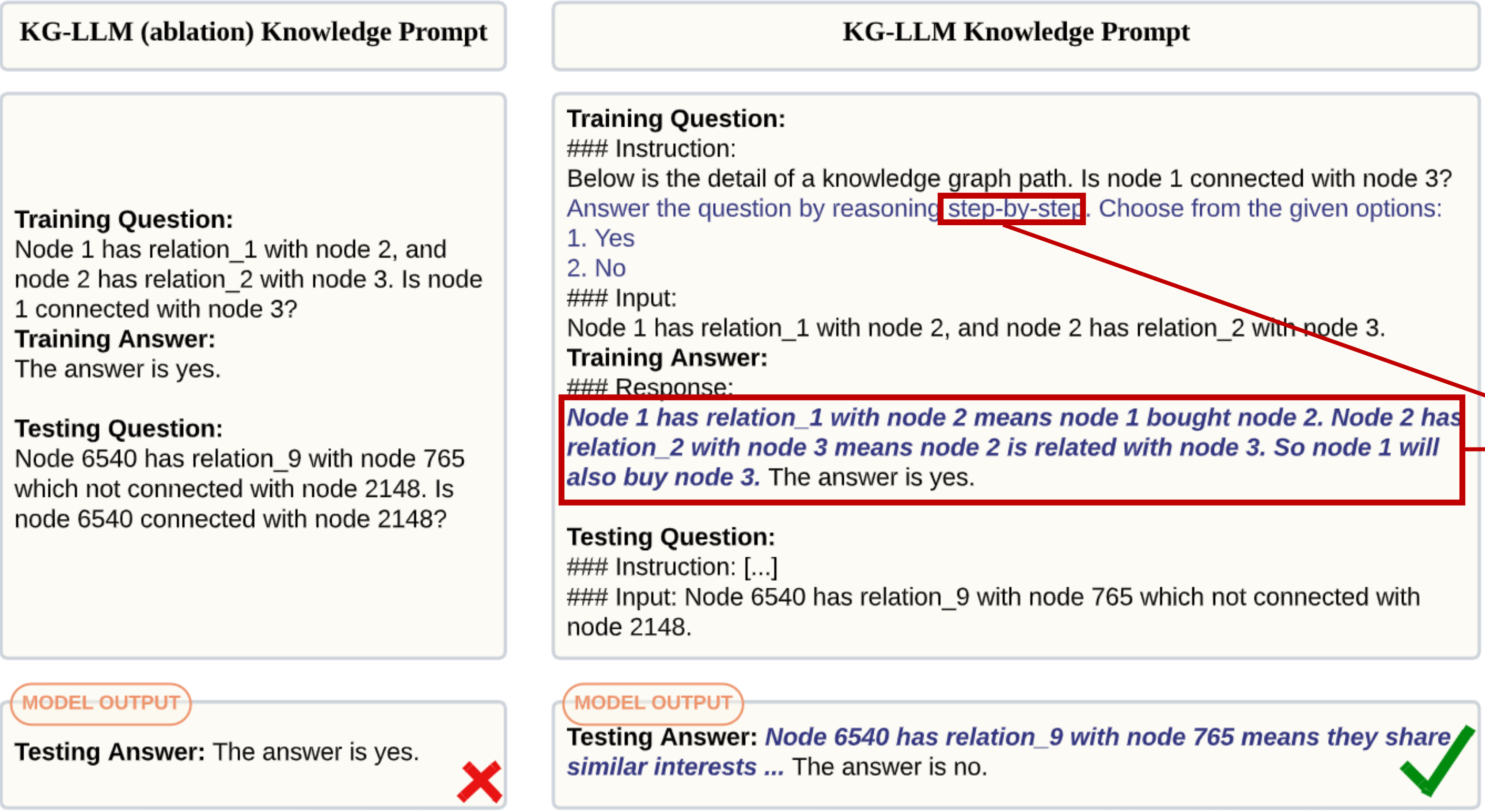
微调后的LLM
回复质量更高

Yao et al. Exploring Large Language Models for Knowledge Graph Completion. arXiv'23

KG-LLM

东南大学 SOUTHEAST UNIVERSITY

将知识图谱中的路径转换为自然语言，结合思维链提示作为LLM的输入，通过指令微调增强大模型的多跳链接预测能力。



Chain-of-Thought

简洁prompt vs 思维链prompt

Shu et al. Knowledge Graph Large Language Model (KG-LLM) for Link Prediction. arXiv'24

KG-LLM

Datasets	WN18RR		NELL-995	
Metrics	F1	AUC	F1	AUC
TransE	0.37	0.37	0.26	0.13
Analogy	0.61	0.52	0.29	0.20
CompleX	0.60	0.51	0.29	0.19
DistMult	0.56	0.48	0.25	0.14
Rescal	0.61	0.50	0.59	0.53
Flan-T5-Large (ablation)	0.63	0.67	0.60	0.66
LlaMa2-7B (ablation)	0.74	0.72	0.71	0.73
Gemma-7B (ablation)	0.76	0.73	0.72	0.71
Flan-T5-Large (KG-LLM)	0.73	0.71	0.70	0.72
LlaMa2-7B (KG-LLM)	<u>0.82</u>	**0.83**	<u>0.81</u>	<u>0.80</u>
Gemma-7B (KG-LLM)	**0.84**	<u>0.81</u>	**0.82**	**0.83**

Datasets	WN18RR	NELL-995
Metric	Accuracy	Accuracy
Flan-T5-Large (ablation)	0.14	0.06
LlaMa2-7B (ablation)	0.23	0.17
Gemma-7B (ablation)	0.22	0.19
Flan-T5-Large (KG-LLM)	0.29	0.21
LlaMa2-7B (KG-LLM)	<u>0.33</u>	<u>0.25</u>
Gemma-7B (KG-LLM)	**0.36**	**0.27**

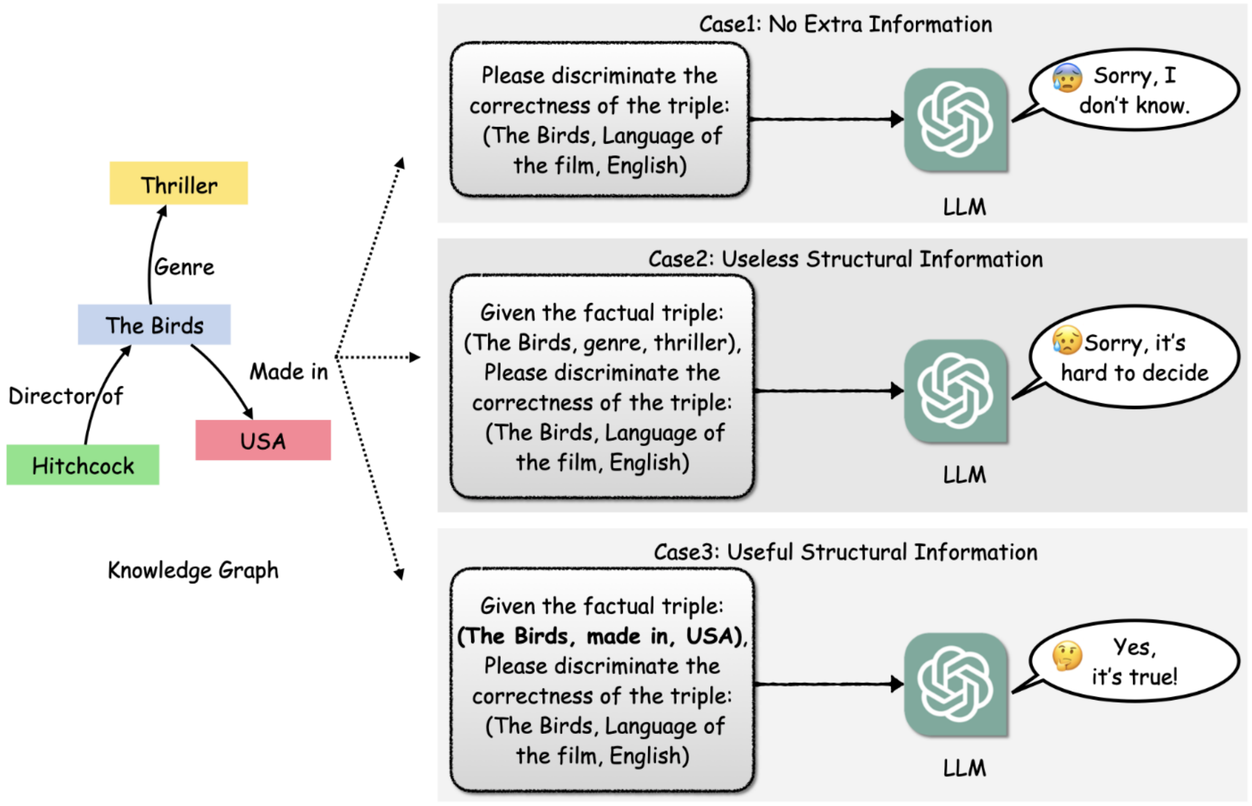
多跳关系预测结果：
CoT提示可以增强的LLM的表现。

多跳链接预测结果：
1) 相比于传统的方法，LLM处理复杂任务的能力更强；
2) CoT的引入进一步提高的LLM的表现。

Shu et al. Knowledge Graph Large Language Model (KG-LLM) for Link Prediction. arXiv’24

KoPA

知识图谱拥有错综复杂的结构信息，如何借助这些信息提高LLM推理效果，实现structural-aware reasoning?



有用的结构信息可以作为辅助提示，指导LLM做出正确的决策。

Zhang et al. Making Large Language Models Perform Better in Knowledge Graph Completion. arXiv'24

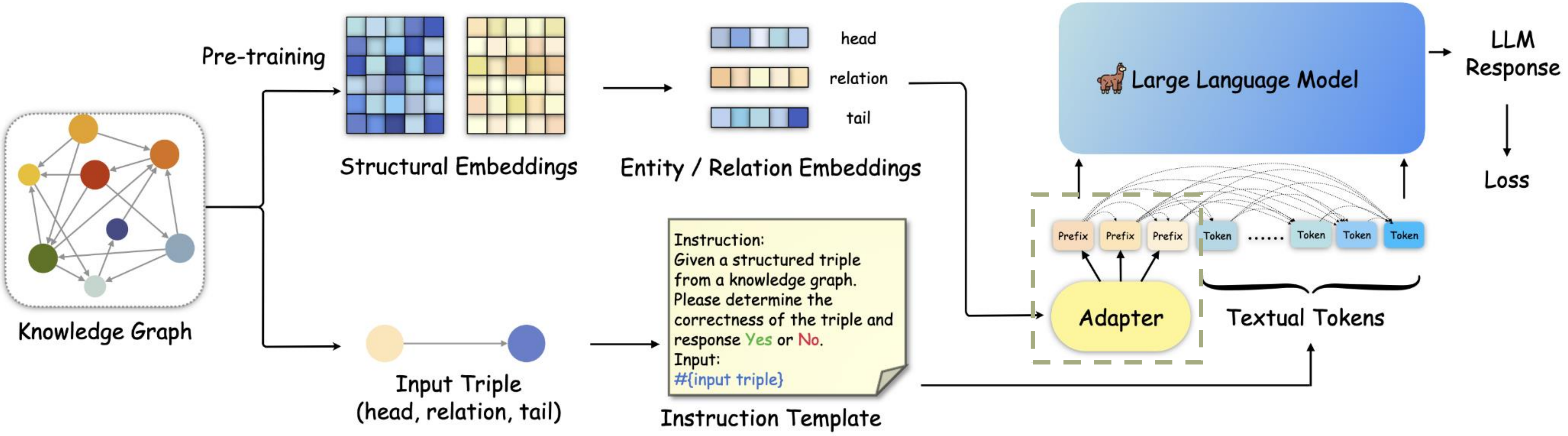
KoPA

知识图谱拥有错综复杂的结构信息，如何借助这些信息提高LLM推理效果，实现structural-aware reasoning?

提出knowledge prefix adapter (KoPA)

- 通过结构预训练获取实体和关系的结构嵌入；
- 将结构嵌入通过Adapter映射到LLM的文本空间中，作为input prompt的前缀输入LLM。

然后，通过三元组分类任务指令微调LLM。



Zhang et al. Making Large Language Models Perform Better in Knowledge Graph Completion. arXiv'24

KoPA

	Model	UMLS				CoDeX-S				FB15K-237N			
		Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
Embedding-based	TransE [3]	84.49	86.53	81.69	84.04	72.07	71.91	72.42	72.17	69.71	70.80	67.11	68.91
	DistMult [38]	86.38	87.06	86.53	86.79	66.79	69.67	59.46	64.16	58.66	58.98	56.84	57.90
	ComplEx [34]	90.77	89.92	91.83	90.87	67.64	67.84	67.06	67.45	65.70	66.46	63.38	64.88
	RotatE [31]	92.05	90.17	94.41	92.23	75.68	75.66	75.71	75.69	68.46	69.24	66.41	67.80
PLM-based	KG-BERT [40]	77.30	70.96	92.43	80.28	77.30	70.96	92.43	80.28	56.02	53.47	97.62	67.84
	PKGC [21]	-	-	-	-	-	-	-	-	79.60	-	-	79.50
LLM-based Training-free	Zero-shot(Alpaca)	52.64	51.55	87.69	64.91	50.62	50.31	99.83	66.91	56.06	53.32	97.37	68.91
	Zero-shot(GPT-3.5)	67.58	88.04	40.71	55.67	54.68	69.13	16.94	27.21	60.15	86.62	24.01	37.59
	ICL(1-shot)	50.37	50.25	75.34	60.29	49.86	49.86	50.59	50.17	54.54	53.67	66.35	59.34
	ICL(2-shot)	53.78	52.47	80.18	63.43	52.95	51.54	98.85	67.75	57.81	56.22	70.56	62.58
	ICL(4-shot)	53.18	52.26	73.22	60.99	51.14	50.58	99.83	67.14	59.29	57.49	71.37	63.68
	ICL(8-shot)	55.52	55.85	52.65	54.21	50.62	50.31	99.83	66.91	59.23	57.23	73.02	64.17
LLM-based Fine-tuning	KG-LLaMA [41]	85.77	87.84	83.05	85.38	79.43	78.67	80.74	79.69	74.81	67.37	96.23	79.25
	KG-Alpaca [41]	86.01	94.91	76.10	84.46	80.25	79.38	81.73	80.54	69.91	62.71	98.28	76.56
	Vanilla IT	86.91	95.18	77.76	85.59	81.18	77.01	88.89	82.52	73.50	65.87	97.53	78.63
	Structure-aware IT	89.93	93.27	86.08	89.54	81.27	77.14	88.40	82.58	76.42	69.56	93.95	79.94
KoPA		**92.58**	90.85	94.70	**92.70**	**82.74**	77.91	91.41	**84.11**	77.65	70.81	94.09	**80.81**

三元组分类结果：

1) KoPA具有较高的Accuracy和F1 score

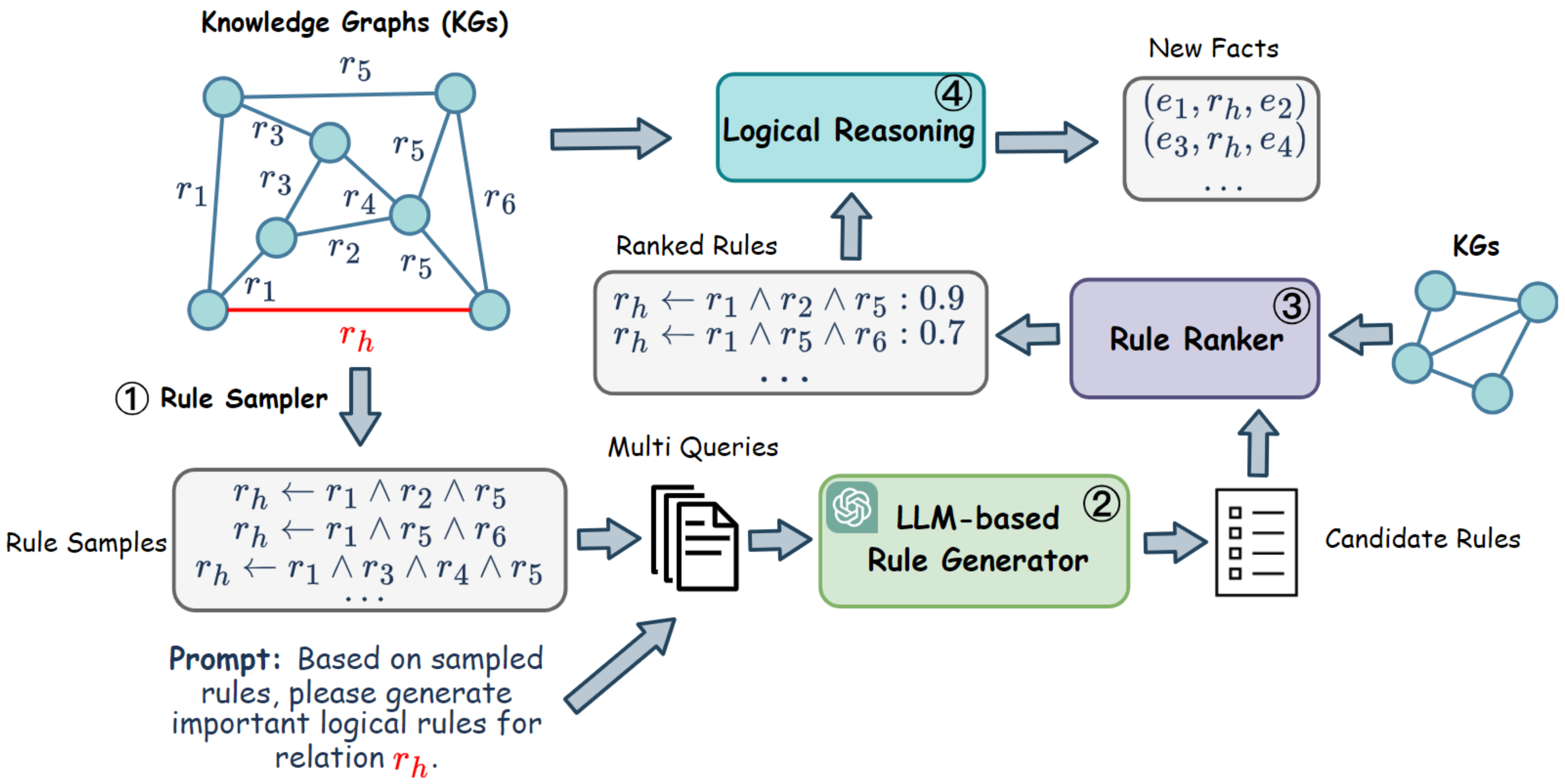
2) 未经微调的LLM不具备结构感知能力；

3) 与提供文本形式的结构信息相比，结构嵌入更易被LLM理解。

Zhang et al. Making Large Language Models Perform Better in Knowledge Graph Completion. arXiv'24

ChatRule

借助LLM生成规则，对这些规则进行排名后在KG上进行推理。



Luo et al. ChatRule: Mining Logical Rules with Large Language Models for Knowledge Graph Reasoning. arXiv'24

基于大模型的知识图谱推理

总结：

1、以上工作借助LLM的文本生成能力直接生成答案。如何将知识图谱转换为LLM可以理解的文本序列是一个关键问题。

2、LLMs相比于PLMs微调的代价更大，通常采用指令微调方式对LLM参数进行更新。

3、未来可以继续探索LLMs在复杂任务场景下的表现，如多跳推理、低资源、零样本、多任务等，因为这些是LLMs相比于PLMs的优势。

不确定性知识图谱

确定性知识图谱

DBpedia　YAGO select knowledge

所有三元组都正确！

<AlphaGo,Produced_by,Google> ✓
<US, Synonym_for, America> ✓

不确定性知识图谱

NELL　ConceptNet An open, multilingual knowledge graph

每个三元组带有一个置信度，描述事实的不确定性。

<AlphaGo,Produced_by,Google>,0.56
<US, Synonym_for, America>, 0.98

不确定性知识图谱推理

难点与挑战：

- 如何在嵌入空间中保留事实的不确定性信息？
- 如何计算unseen事实的置信度？

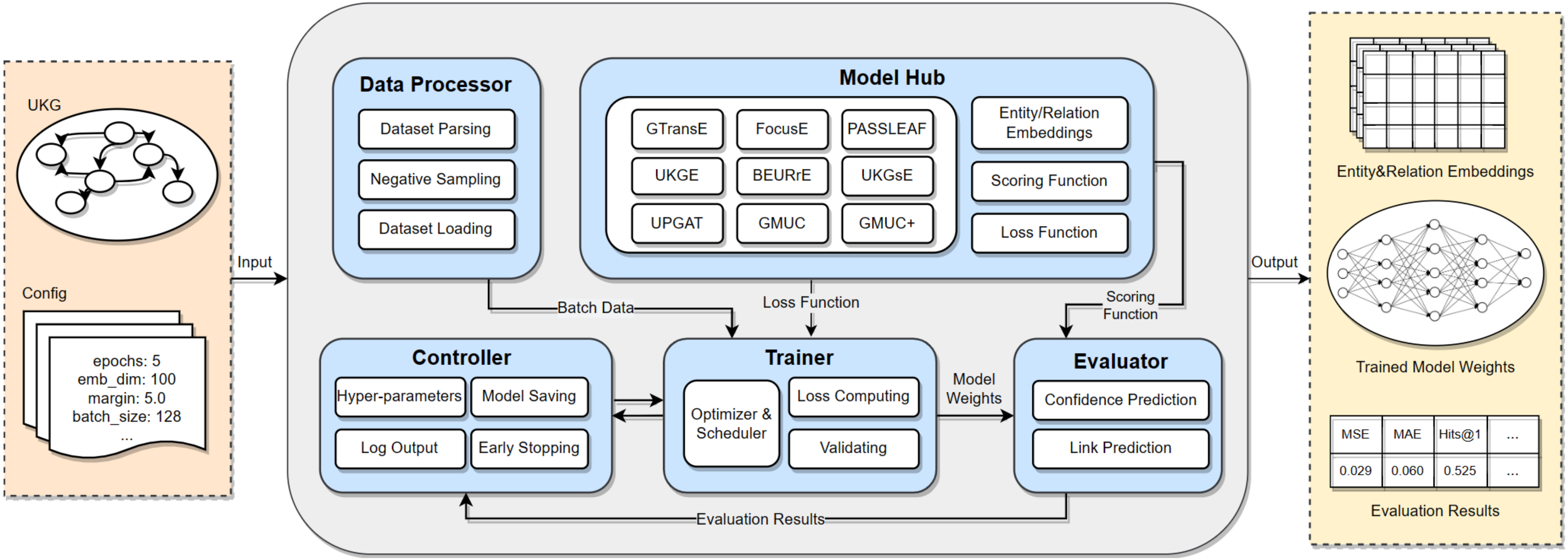
如何借助**LLM的零样本/小样本学习能力**，提高在不确定性知识图谱上的推理效果仍有待研究。

UKG推理工具库

https://github.com/seucoin/unKR

unKR

Uncertain Knowledge Reasoning



Wang et al. unKR: A Python Library for Uncertain Knowledge Graph Reasoning by Representation Learning. SIGIR’24.

unKR-LLM

SOUTHEAST UNIVERSITY 東南大學

利用参数高效微调(PEFT)技术，探索大模型在UKG推理任务上发挥的作用

- 使用LoRA微调Llama2-7B，基于置信度预测与链接预测任务微调。
- 评估结果：

Instruction	Response
置信度预测 Below is an instruction that describes a task. Write a response that appropriately completes the request. ### Instruction: **What is the probability of the following fact being true:** stone is related to diamond? ### Response:	0.709
链接预测 Below is an instruction that describes a task. Write a response that appropriately completes the request. ### Instruction: **Stone is related to** ? ### Response:	Diamond

Wang et al. unKR: A Python Library for Uncertain Knowledge Graph Reasoning by Representation Learning. SIGIR'24.

unKR-LLM

置信度预测与链接预测结果：

Dataset	MSE	MAE	Hits@1
CN15K	0.03110	0.08980	0.15767
NL27K	0.05848	0.13009	0.35599

LLM在常识类知识图谱（如：ConceptNet）上能达到更好的推理效果，而在事实类知识图谱上（如：NELL）的表现仍有待提高，因为事实类知识图谱推理更依赖于图谱结构。因此，如何将带置信度的结构信息注入到LLM中是值得探索的方向。

Wang et al. unKR: A Python Library for Uncertain Knowledge Graph Reasoning by Representation Learning. SIGIR'24.

参考文献

[1] Yao et al. KG-BERT: BERT for Knowledge Graph Completion. arXiv preprint 2019.

[2] Kim et al. Multi-Task Learning for Knowledge Graph Completion with Pre-trained Language. COLING 2020.

[3] Wang et al. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language. TACL 2021.

[4] Wang et al. Structure-Augmented Text Representation Learning for Efficient Knowledge Graph Completion. WWW 2021.

[5] Clouâtre et al. MLMLM: Link Prediction with Mean Likelihood Masked Language Model. Findings of ACL 2021.

[6] Wang et al. SimKGC: Simple Contrastive Knowledge Graph Completion with Pre-trained Language. ACL 2022.

[7] Li et al. Multi-task Pre-training Language Model for Semantic Network Completion. TALLIP 2022.

[8] Chen et al. Dipping PLMs Sauce: Bridging Structure and Text for Effective Knowledge Graph Completion via Conditional Soft Prompting. Findings of ACL 2023.

[9] Xie et al. From Discrimination to Generation: Knowledge Graph Completion with Generative Transformer. WWW 2022.

[10] Saxena et al. Sequence-to-Sequence Knowledge Graph Completion and Question Answering. ACL 2022.

[11] Chen et al. Knowledge Is Flat: A Seq2Seq Generative Framework for Various Knowledge Graph Completion. COLING 2022.

[12] Yao et al. Exploring Large Language Models for Knowledge Graph Completion. arXiv preprint 2023.

[13] Zhang et al. Making Large Language Models Perform Better in Knowledge Graph. arXiv preprint 2024.

[14] Wei et al. KICGPT: Large Language Model with Knowledge in Context for Knowledge Graph Completion. Findings of EMNLP 2023.

[15] Luo et al. ChatRule: Mining Logical Rules with Large Language Models for Knowledge Graph Reasoning. arXiv preprint 2024.

[16] Wen et al. MindMap: Knowledge Graph Prompting Sparks Graph of Thoughts in Large Language Models. arXiv preprint 2024.

[17] Shu et al. Knowledge Graph Large Language Model (KG-LLM) for Link Prediction. arXiv preprint 2024.

[18] Wang et al. unKR: A Python Library for Uncertain Knowledge Graph Reasoning by Representation Learning. SIGIR 2024.

语言与知识计算专委会

OpenKG

东南大学认知智能研究所

数据智能期刊

谢谢!