Multimodal Analogical Reasoning over Knowledge Graph

ICLR 2023

Ningyu Zhang*, Lei Li*, Xiang Chen*, et al.

Zhejiang University, Alibaba Group

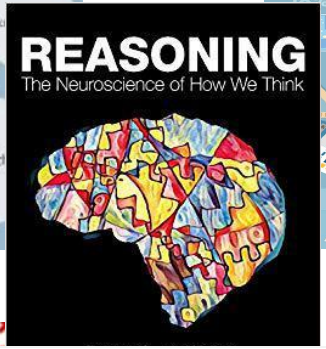
code: https://github.com/zjunlp/MKG_Analogy

leaderboard: https://zjunlp.github.io/project/MKG_Analogy/

Analogical Reasoning

Analogical reasoning – the ability to perceive and use relational similarity between two situations or events – is a fundamental aspect of human cognition. Indeed, some researchers suggest that it is the crucial cognitive mechanism that most distinguishes human cognition from that of other intelligent species.

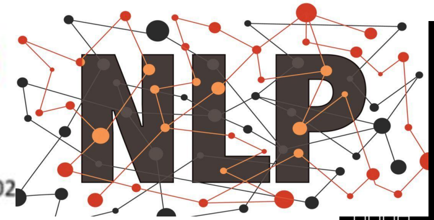
--Gentner, D. & Smith, L. (2012). Analogical reasoning. In V. S. Ramachandran (Ed.)

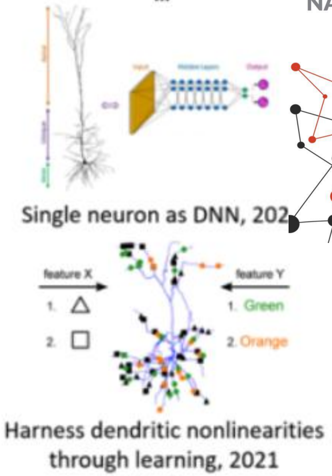


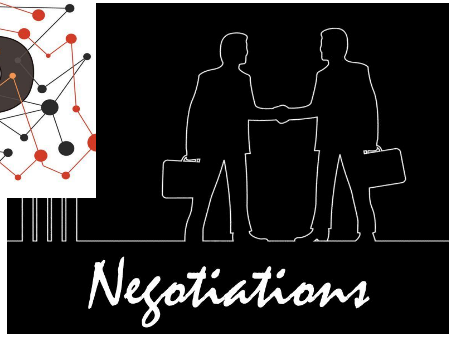
Cognitive Science

Education

NATURAL LANGUAGE PROCESSING



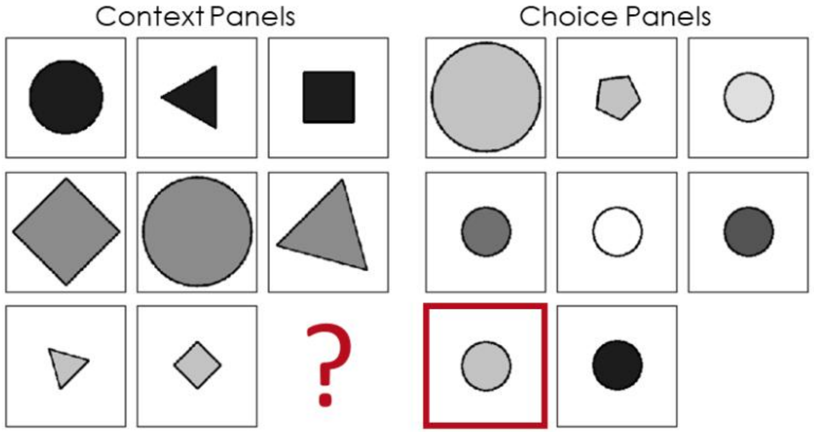


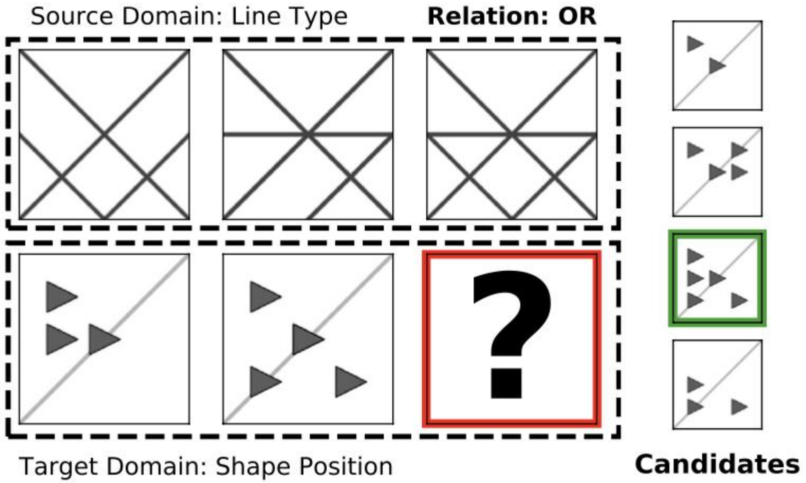


Negotiation

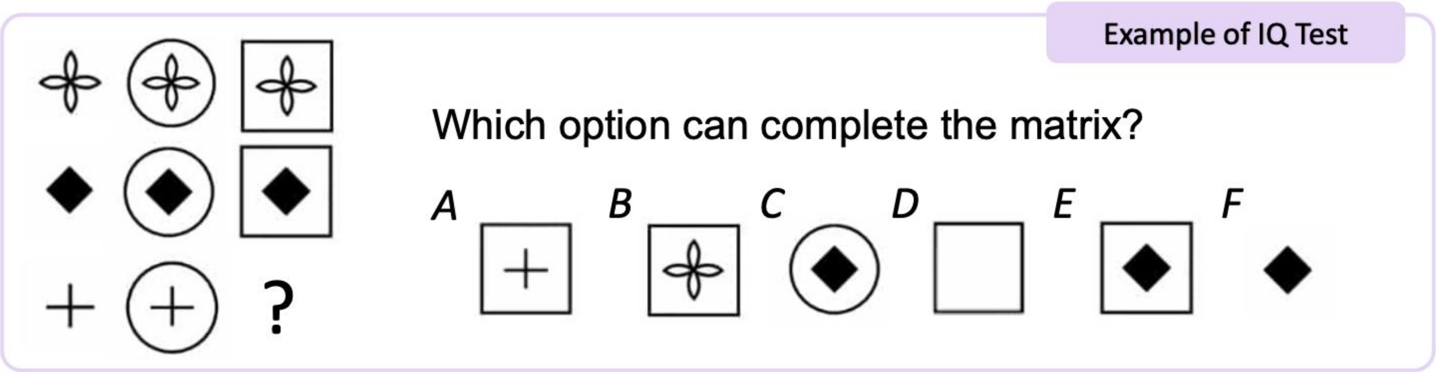
Brain Science

Analogical in CV





RAVEN dataset

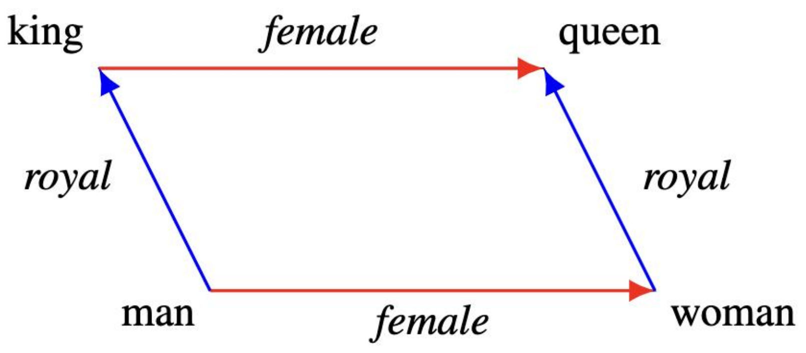


KOSMOS-1

Hill F, Santoro A, Barrett D G T, et al. Learning to make analogies by contrasting abstract relational structure. ICLR 2019.
Hayes T L, Kanan C. Selective replay enhances learning in online continual analogical reasoning. CVPR 2021.
Shaohan H, Li D et al. Language Is Not All You Need: Aligning Perception with Language Models. 2023.

Natural Language Processing (NLP)

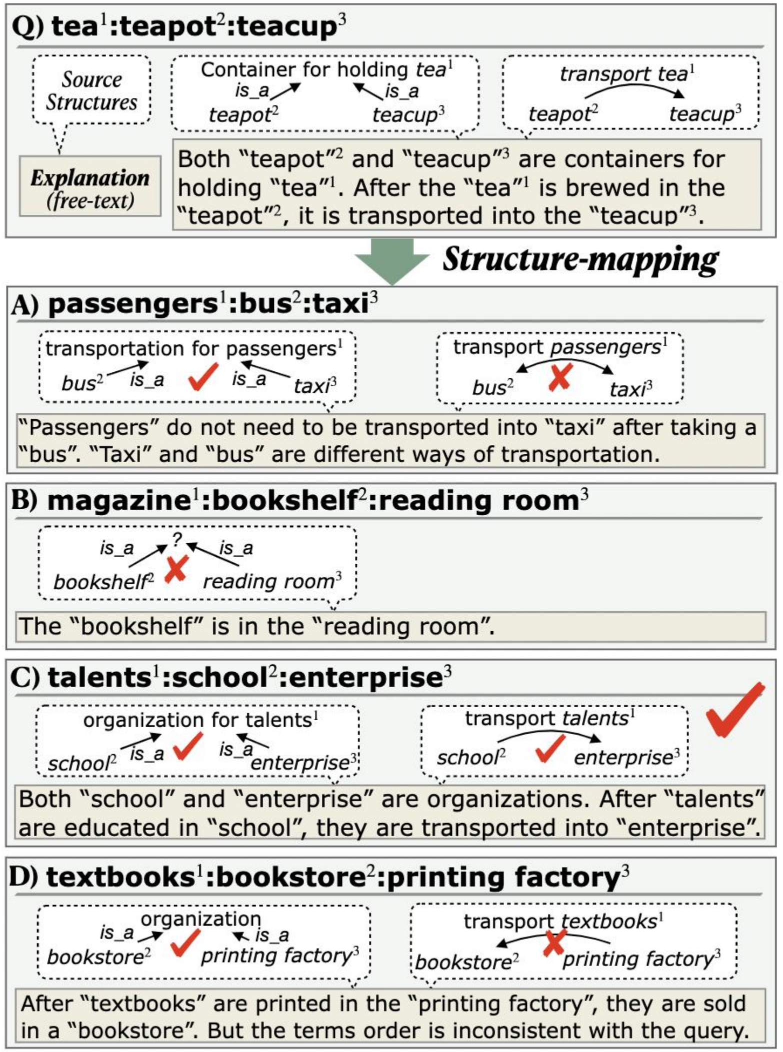
Word Analogy



Distributed representations of words and phrases and their compositionality. NeurIPS 2013

Ethayarajh K, et al. Towards understanding linear word analogies. ACL 2019.

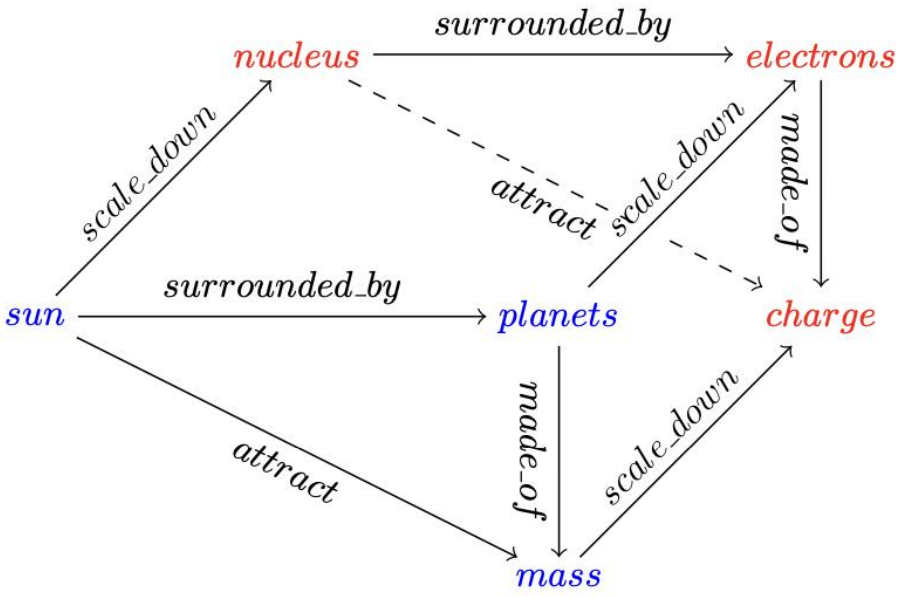
Chen J, Xu R, Fu Z, et al. E-KAR: A benchmark for rationalizing natural language analogical reasoning. ACL 2022 (Findings).



Multiple-Choice QA

Analogical in KG

Knowledge Graph (KG)



$$W_{r \circ r'} = W_r W_{r'} = W_{r'} W_r$$

$$\begin{tikzcd} a \arrow[r, "r"] \arrow[d, "r'"] & b \arrow[d, "r'"] \\ c \arrow[r, "r"] & d \end{tikzcd}$$

$$\begin{aligned} \phi_{v,W}(s,r,o) &= v_s^\top W_r v_o = v_s^\top Q B_r Q^\top v_o \\ &= u_s^\top B_r u_o = \phi_{u,B}(s,r,o) \end{aligned}$$

Liu H, Wu Y, Yang Y. Analogical inference for multi-relational embeddings[C], ICML, 2017.

Motivation

1. The perspective of cognitive reasoning:

The ability to quickly map multiple modalities allows humans to realize multimodal analogical reasoning.

2. The perspective of data model:

Hallucination problem and statistical distribution difference in the unimodal data.

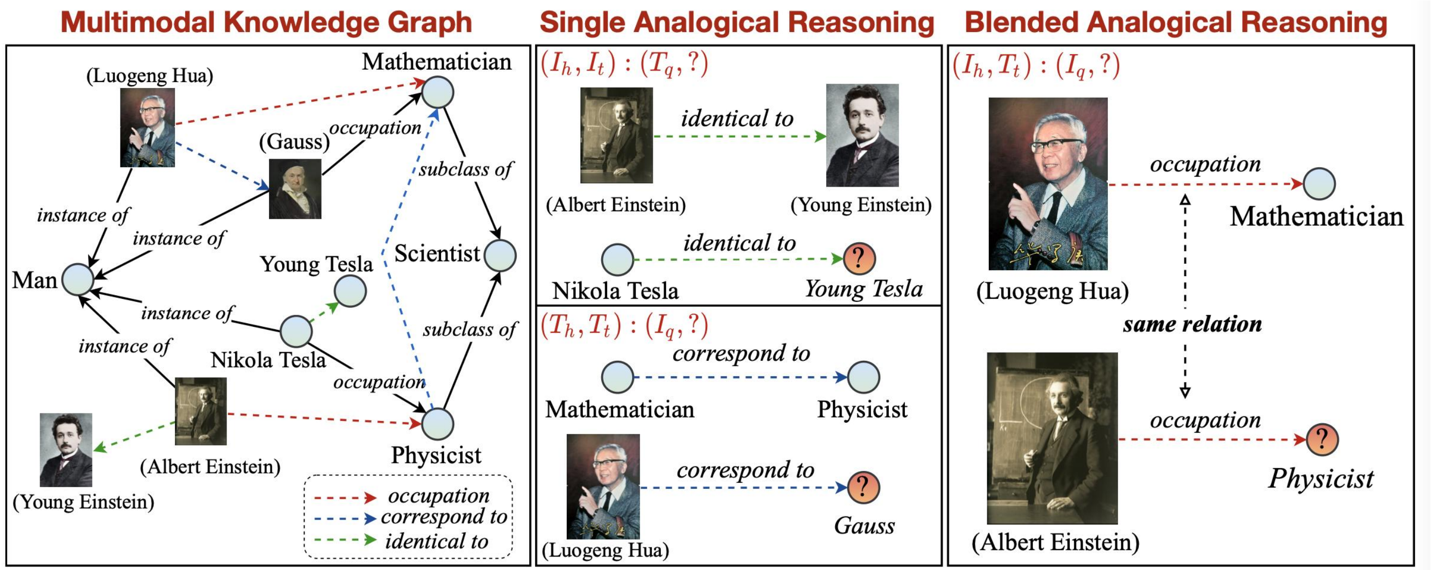
Why Multimodal?

Mayer's theory: Evidence from psychology and learning sciences that human learners often perform better on tests of recall and transfer when they have learned from **multimodal sources** than single modality sources (e.g. (Mayer, 2001; Hegarty & Just, 1993))

Contribution

Propose the **Multimodal Analogical Reasoning task** with a multimodal analogical reasoning dataset (**MARS**), Multimodal KG (**MarKG**) and some baseline models.

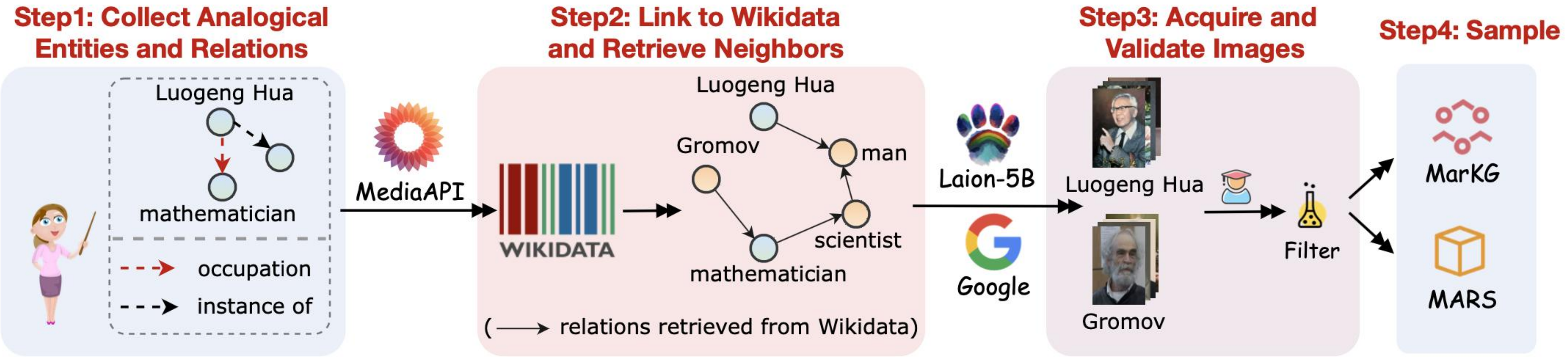
Task Definition



Given a background multimodal graph $G = (E, R, I, T)$, the analogical reasoning task can be formulized as $(e_h, e_t) : (e_q, ?)$, where $e_h, e_t, e_q \in E$.

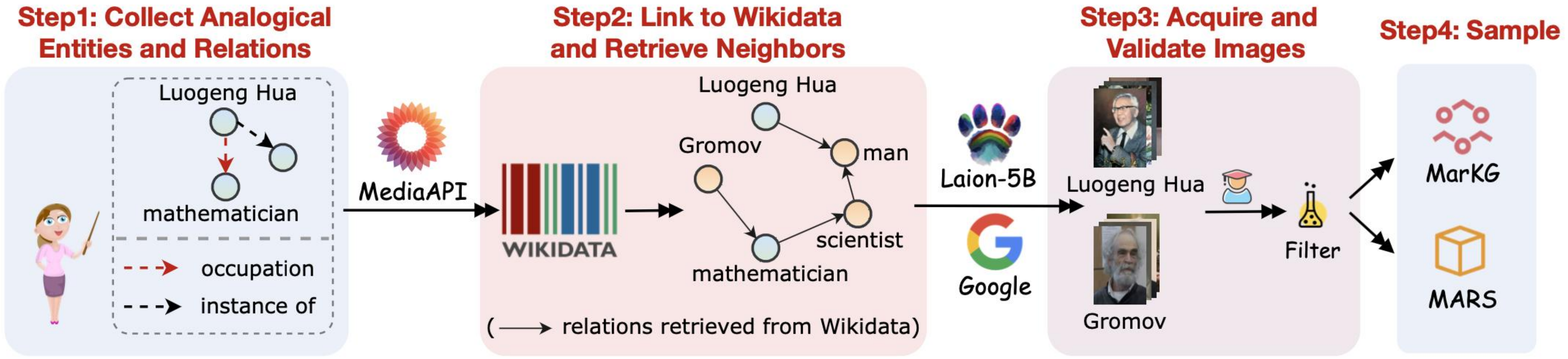
- Single Analogical Reasoning: $(I_h, I_t) : (T_q, ?)$ or $(T_h, T_t) : (I_q, ?)$
- Blended Analogical Reasoning: $(I_h, T_t) : (I_q, ?)$

Dataset Collection



1. **Collect Analogy Entities and Relations**. Collect the analogy seed entities and relations from E-KAR and SATs datasets. Then merge similar and discard simply items.
2. **Link to Wikidata and Retrieve Neighbors.** Link the entities and relations to Wikidata and retrieve their one-hop neighbors to construct graph data.
3. **Acquire and Validate Images.** Collect images from Google and Laion-5B.
4. **Sample Analogical Reasoning Data.** Sample analogy data to construct MARS dataset.
5. **Quality Control of Datasets**. Formalization and normalization, image validation mechanism, control of text description.

Dataset Collection

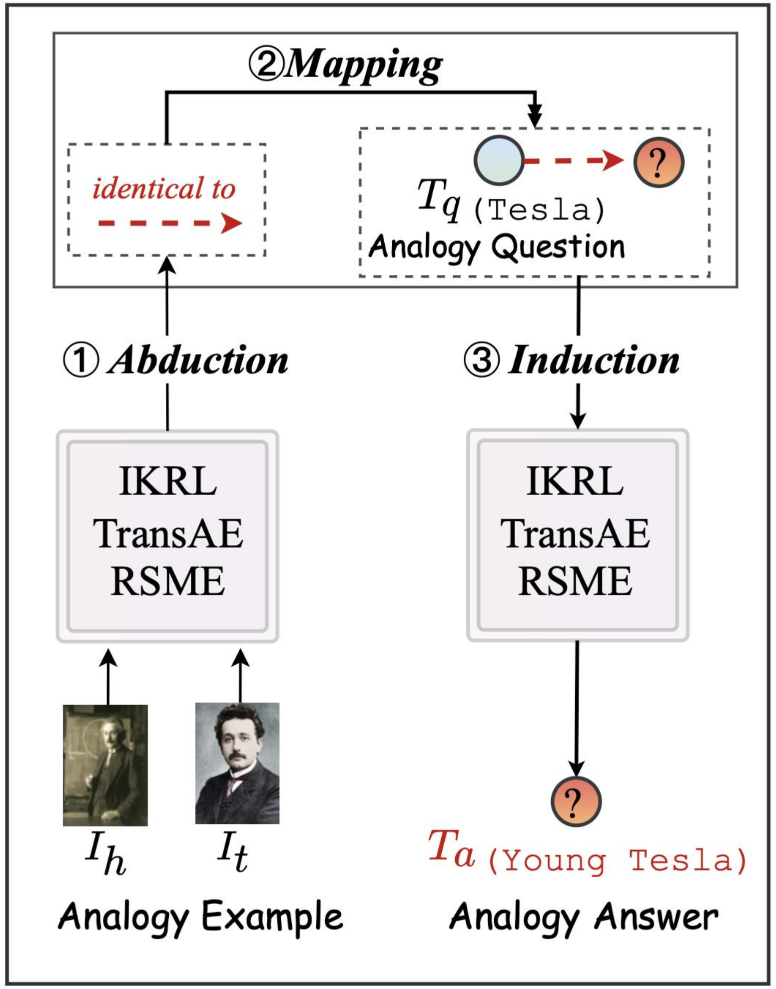


Dataset	Size (train / dev / test)	KB	Modality	# Entity	# Relation	# Images	Knowledge Intensive	Task Format
RAVEN	42,000 / 14,000 / 14,000	✘	Vision	-	8	1,120,000	✘	Classification
SAT	0 / 37 / 337	✘	Text	-	19	-	✘	Linear Word Analogy
Google	0 / 50 / 500	✘	Text	919	14	-	✘	Linear Word Analogy
BATs	0 / 199 / 1,799	✘	Text	6,218	40	-	✘	Linear Word Analogy
E-KAR	870 / 119 / 262	✘	Text	2,651	28	-	✔	Multiple Choice QA
MARS	10,685 / 1,228 / 1,415	MarKG	Vision+Text	2,063	27	13,398	✔	Entity Prediction

Baselines

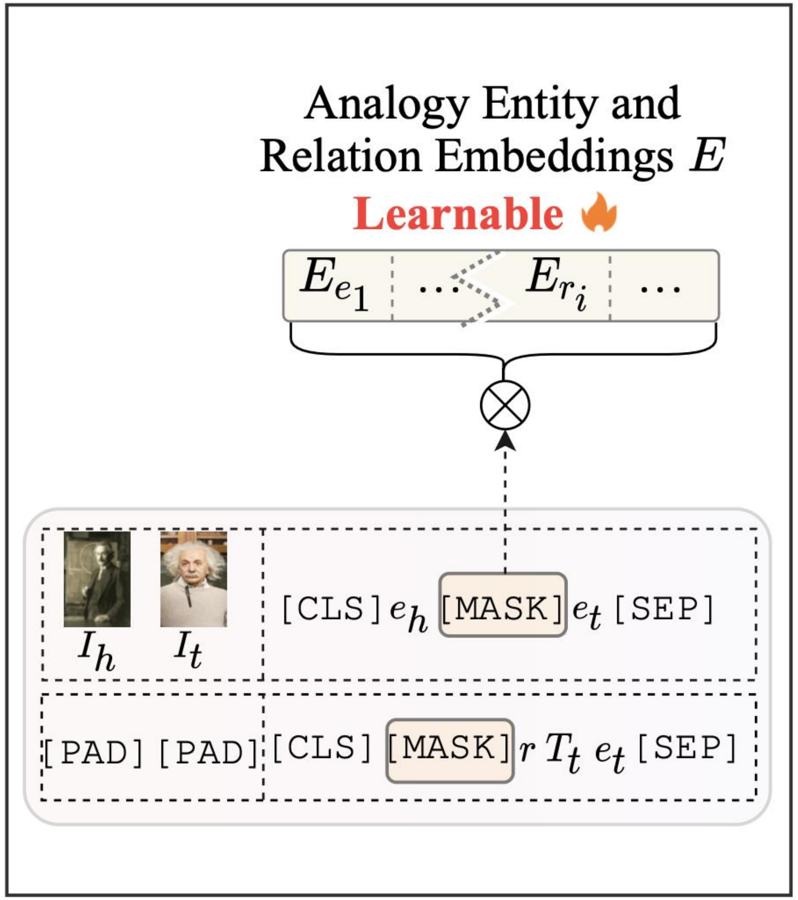
➢ MKGE Baselines - Pipeline

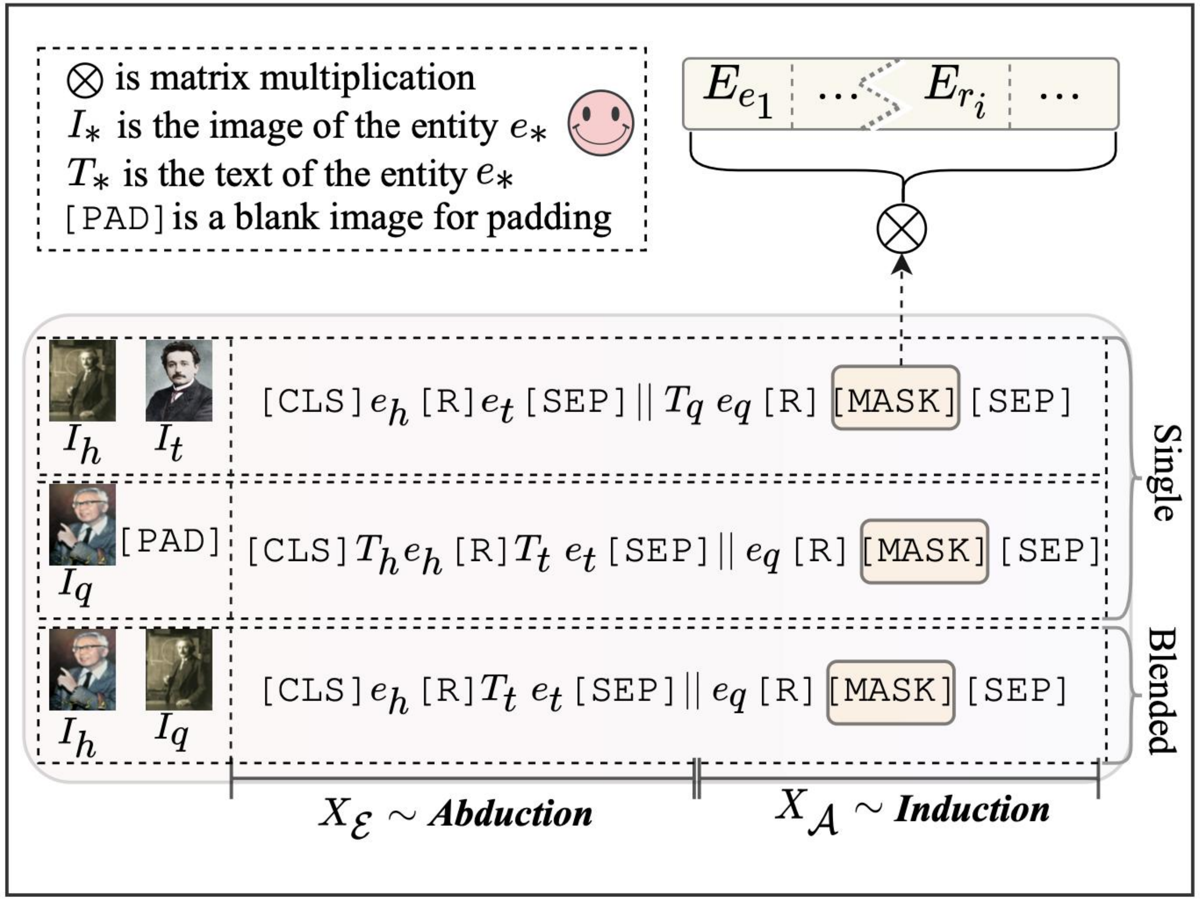
- Three multimodal knowledge embedding (MKGE) approaches as baselines：
  - IKRL, TransAE, and RSME.
- Typically based on TransE or ComplEx and combine with visual encoders to encode images for multimodal knowledge representation learning.
- Replace the backbone of MKGE methods with ANALOGY (Liu et al. 2017) that **models analogical structure explicitly** as baselines.



ZHEJIANG UNIVERSITY

- **Multimodal Pre-trained Transformer Baselines - End-to-End**

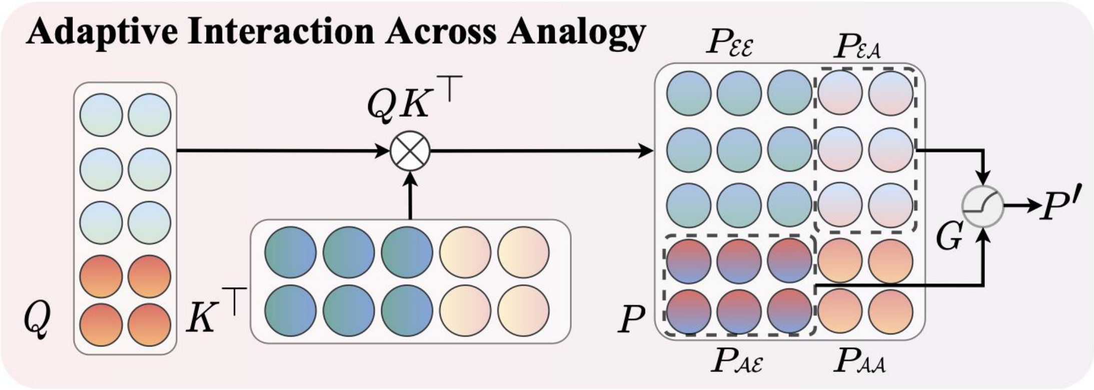




$$\mathcal{T}_{(I_h,T_t,I_q)} = \mathcal{T}_{\mathcal{E}} \parallel \mathcal{T}_{\mathcal{A}} = I_h\ I_q\ \texttt{[CLS]}\ e_h\ \texttt{[R]}\ T_t\ e_t\ \texttt{[SEP]} \parallel e_q\ \texttt{[R]}\ \texttt{[MASK]}\ \texttt{[SEP]}$$

浙江大学 ZHEJIANG UNIVERSITY

➢ MarT: A Multimodal Analogical Reasoning Framework with Transformer



- Diagonal of P are intra-analogy attentions
- **Anti-diagonal of P** are inter-analogy attentions

■ **Attention values and keys**

$$Q = XW^Q = \begin{pmatrix} X_{\mathcal{E}} \\ X_{\mathcal{A}} \end{pmatrix} W^Q = \begin{pmatrix} Q_{\mathcal{E}} \\ Q_{\mathcal{A}} \end{pmatrix}, K = XW^K = \begin{pmatrix} X_{\mathcal{E}} \\ X_{\mathcal{A}} \end{pmatrix} W^K = \begin{pmatrix} K_{\mathcal{E}} \\ K_{\mathcal{A}} \end{pmatrix}$$

■ **Decompose attention scores into intra-analogy and inter-analogy**

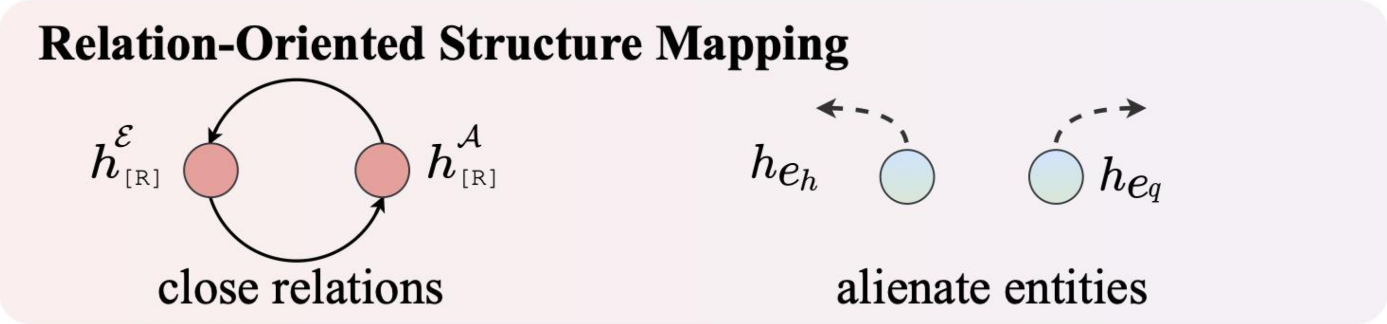
$$P = QK^\top = \begin{pmatrix} Q_{\mathcal{E}} \\ Q_{\mathcal{A}} \end{pmatrix} (K_{\mathcal{E}}^\top, K_{\mathcal{A}}^\top) = \begin{pmatrix} Q_{\mathcal{E}}K_{\mathcal{E}}^\top & Q_{\mathcal{E}}K_{\mathcal{A}}^\top \\ Q_{\mathcal{A}}K_{\mathcal{E}}^\top & Q_{\mathcal{A}}K_{\mathcal{A}}^\top \end{pmatrix} = \begin{pmatrix} P_{\mathcal{E}\mathcal{E}} & P_{\mathcal{E}\mathcal{A}} \\ P_{\mathcal{A}\mathcal{E}} & P_{\mathcal{A}\mathcal{A}} \end{pmatrix}$$

■ **Do inter-analogy interactions adaptively**

$$P' = G \odot P = \begin{pmatrix} 1 & g_{\mathcal{E}\mathcal{A}} \\ g_{\mathcal{A}\mathcal{E}} & 1 \end{pmatrix} \odot \begin{pmatrix} P_{\mathcal{E}\mathcal{E}} & P_{\mathcal{E}\mathcal{A}} \\ P_{\mathcal{A}\mathcal{E}} & P_{\mathcal{A}\mathcal{A}} \end{pmatrix} = \begin{pmatrix} P_{\mathcal{E}\mathcal{E}} & g_{\mathcal{E}\mathcal{A}}P_{\mathcal{E}\mathcal{A}} \\ g_{\mathcal{A}\mathcal{E}}P_{\mathcal{A}\mathcal{E}} & P_{\mathcal{A}\mathcal{A}} \end{pmatrix}$$

Baselines

➢ **MarT: A Multimodal Analogical Reasoning Framework with Transformer**



*"**relations** between objects, rather than attributes of **objects**, are mopped from base to target."* -- Structure Mapping Theory

■ **Bring the relations closer and alienate the entities**

$$\mathcal{L}_{\text{rel}} = \frac{1}{|\mathcal{S}|}\sum_{i}^{|\mathcal{S}|}(\underbrace{1 - \text{sim}(h^{\mathcal{E}}_{[\text{R}]}, h^{\mathcal{A}}_{[\text{R}]})}_{\text{close relations}} + \underbrace{\max(0, \text{sim}(h_{e_h}, h_{e_q}))}_{\text{alienate entities}})$$

■ **Cross-entropy loss in masked entity prediction**

$$\mathcal{L}_{\text{mem}} = -\frac{1}{|\mathcal{S}|}\sum_{(e_h,e_t,e_q,e_a)\in\mathcal{S}} \log(p(\texttt{[MASK]} = e_a)|\mathcal{T}_{(e_h,e_t,e_q)})$$

■ **Final loss function**

$$\mathcal{L} = \lambda\mathcal{L}_{\text{rel}} + (1-\lambda)\mathcal{L}_{\text{mem}}$$

Experiment

Method	Baselines	Backbone	Hits@1	Hits@3	Hits@5	Hits@10	MRR
MKGE	IKRL	TransE	0.254	0.285	0.290	0.304	0.274
	TransAE	TransE	0.203	0.233	0.241	0.253	0.223
	RSME	ComplEx	0.255	0.274	0.282	0.291	0.268
	IKRL	ANALOGY	0.266	0.294	0.301	0.310	0.283
	TransAE	ANALOGY	0.261	0.285	0.289	0.293	0.276
	RSME	ANALOGY	0.266	0.298	0.307	0.311	0.285
MPT	VisualBERT	Single-Stream	0.247	0.281	0.289	0.303	0.269
	ViLT	Single-Stream	0.235	0.266	0.274	0.286	0.257
	ViLBERT	Dual-Stream	0.252	0.308	0.320	0.338	0.287
	FLAVA	Mixed-Stream	0.257	0.299	0.312	0.325	0.284
	MKGformer	Mixed-Stream	0.293	0.335	0.344	0.367	0.321
	MarT_VisualBERT	Single-Stream	0.261	0.292	0.308	0.321	0.284
	MarT_ViLT	Single-Stream	0.245	0.275	0.287	0.303	0.266
	MarT_ViLBERT	Dual-Stream	0.256	0.312	0.327	0.347	0.292
	MarT_FLAVA	Mixed-Stream	0.264	0.303	0.309	0.319	0.288
	MarT_MKGformer	Mixed-Stream	**0.301**	**0.367**	**0.380**	**0.408**	**0.341**

- The performance of MKGE and MPT methods are **comparable**.
- The **analogical structures** significantly improve performance.
- MarT_MKGformer perform best.

Analysis

Model	Hits@1	Hits@3	Hits@5	Hits@10	MRR
TransAE	0.203	0.233	0.241	0.253	0.223
w/o MarKG	0.191	0.224	0.235	0.245	0.214
MarT_ViLBERT	0.256	0.312	0.327	0.347	0.292
w/o MarKG	0.253	0.292	0.297	0.310	0.270
MarT_MKGformer	**0.301**	**0.367**	**0.380**	**0.408**	**0.341**
w/o MarKG	0.270	0.305	0.309	0.315	0.289
w/o Relaxation loss	0.295	0.349	0.373	0.399	0.332
w/o Adaptive interaction	0.285	0.345	0.365	0.395	0.324
w/o MarT	0.293	0.335	0.344	0.367	0.321

Table 4: Ablation experiments on MARS. *w/o MarKG* refers to the model without pre-training on MarKG dataset. w/o MarT refers to ablate all components of MarT that equivalents to MKGformer.

Model	Novel Relation Transfer			
	Hits@1	Hits@3	Hits@10	MRR
MarT_MKGformer	0.254	0.285	0.292	0.273
w/o MarKG	0.217	0.228	0.231	0.224
w/ Full MARS	0.365	0.419	0.433	0.395

Table 3: Results of novel relation generalization, we test the most strong baseline MKGformer on novel target relations data. "w/ Full MARS" is the result trained with full data (upper bound).

- MarKG provides **good priors** for entities and relations representation learning
- The effectiveness of the analogical components
- MarT_MKGformer can indeed learn to make sense of unfamiliar relations

Analysis

Task Setting	Analogical Example	Question-Answer Pair	Model	Top-3 Entity			Gold Rank
$(I_h, I_t) \downarrow (T_q, ?)$	 *Qinghai Lake* —*correspond to*→  *Inland Lake*	campaign —*correspond to*→ ? battle	MKGformer	*film*	*increment*	*scheme*	207
			MKGformer*	***battle***	*war*	*siege*	1
			TransAE	*life*	*aircraft*	*ocean*	254
			TransAE*	***battle***	*reaction*	*court*	1
$(I_h, T_t) \downarrow (I_q, ?)$	 *Panax notoginseng* —*instance of*→ Traditional Chinese Medicine	 *Apple* —*instance of*→ ? fruit	MKGformer	*bread*	*capital*	*phone*	80
			MKGformer*	***fruit***	*dried fruit*	*citrus*	1
			TransAE	*Citrus*	*grain*	*shipping*	125
			TransAE*	*plant*	*Citrus*	*dessert*	6

- Top ranking entities of vanilla baselines are usually **irrelevant** to the question entity.
- **Analogical components** make the predictions more reasonable.
- Relaxation loss can alienate the entities and focus on **relation structures transfer**.

Conclusion

- We introduce a new task of multimodal analogy reasoning over KGs
- Baselines: MKGE Modes, Multimodal Pretrained Transformers
- The main reason for the unsatisfactory performance:
    - multimodal analogical reasoning is a difficult task
- **Implicit analogical components** benefits the performance

Potential Future Works

- Stronger models for multimodal analogy reasoning over KGs
- Addressing tasks via **multimodal analogy reasoning in NLP, CV etc.** including 1) knowledge graph completion with analogies, (2) transfer learning and zero-shot learning by analogy, (3) analogical question answering
- Theoretical principle of multimodal analogy reasoning
- Potential Applications
  - Education
  - Design
  - Psychology

Potential Future Works

Multimodal analogical reasoning refers to the ability to **recognize and reason about analogie**s across multiple modes of sensory perception or representation. In other words, it involves using **similarities and relationships** between different types of information, such as visual, auditory, and spatial information, to solve problems and make inferences.

For example, in a simple analogy, one might compare the relationship between a dog and a bone to the relationship between a cat and a mouse. However, in a multimodal analogy, one might compare the relationship between a **musical note and a chord** to the relationship between a **color and a painting**. The analogy involves different types of sensory information, such as sound and sight, but still requires **recognizing relationships** and **making connections** between them.

Multimodal analogical reasoning is an important cognitive ability that is related to creativity, problem-solving, and learning. It has applications in a variety of fields, including **education, computer science, and design.**

-- From **ChatGPT**

OpenAI

Code & Datasets & Checkpoints

Flow3rDown fix test bug d069599 2 days ago 24 commits

M-KGE	first commit	2 months ago
MarT	fix test bug	2 days ago
resource	fix bugs	last week
.gitignore	fix models path bug	last month
LICENSE	Initial commit	6 months ago
README.md	Update README.md	last week
requirements.txt	fix models path bug	last month

README.md

MKG_Analogy

Code and datasets for the ICLR2023 paper "Multimodal Analogical Reasoning over Knowledge Graphs"

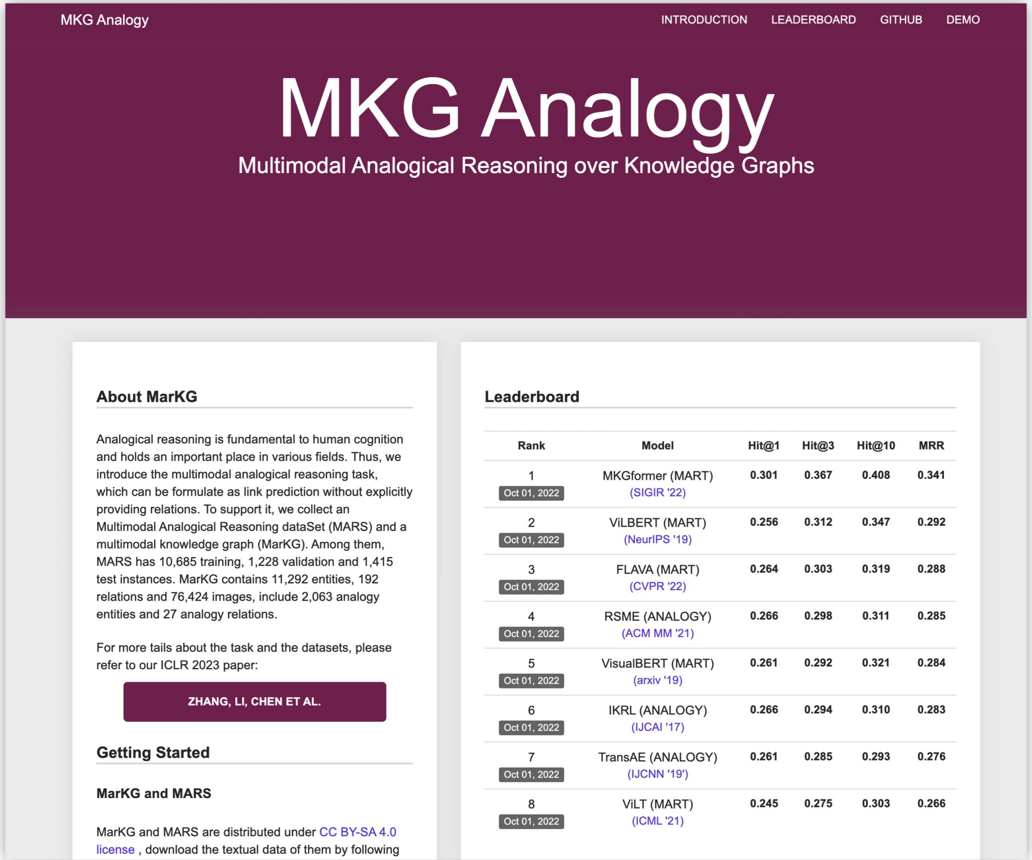
- ❗ New: We provide a Huggingface Demo at https://huggingface.co/spaces/zjunlp/MKG_Analogy, have fun!
- ❗ New: We have released the Checkpoints at Google Drive for reproducibility.

Quick links

github.com/zjunlp/MKG_Analogy

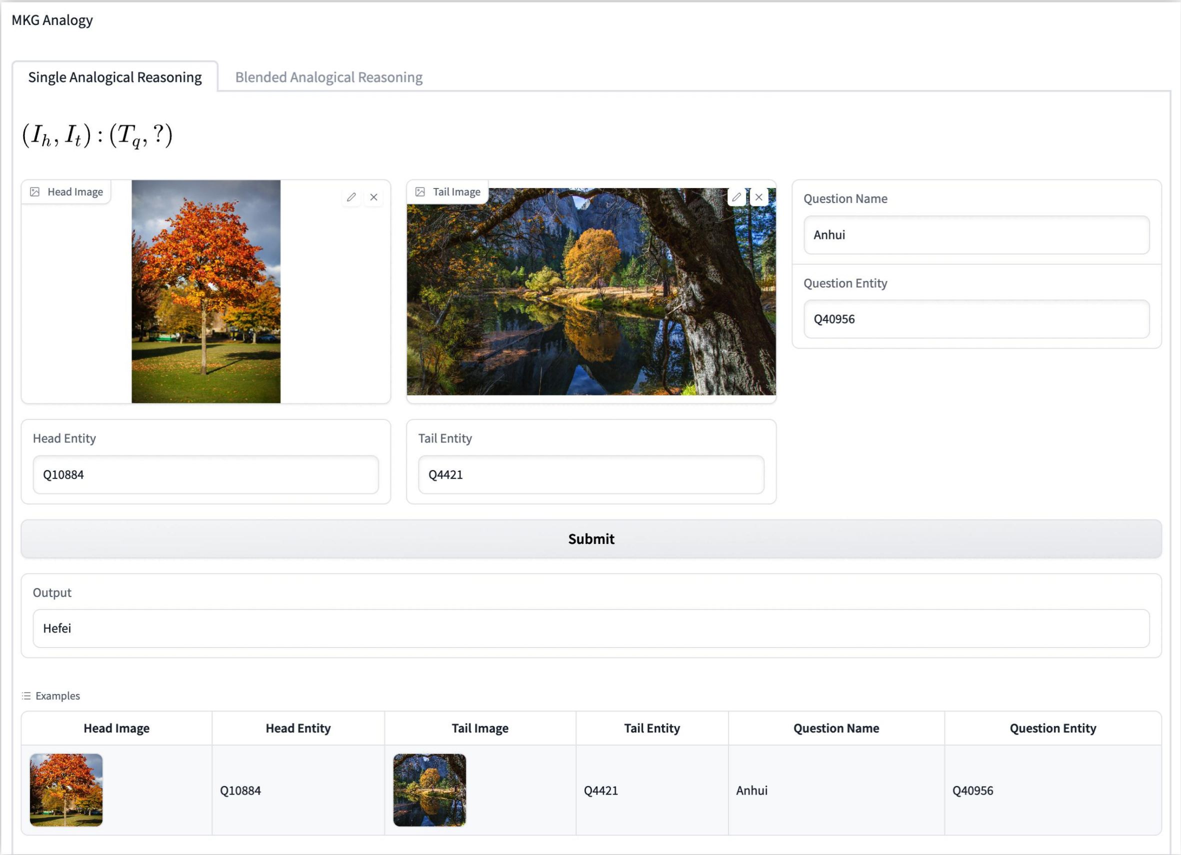
LeaderBoard & Demo

浙江大学 ZHEJIANG UNIVERSITY



Leaderboard

zjunlp.github.io/project/MKG_Analogy/



Demo

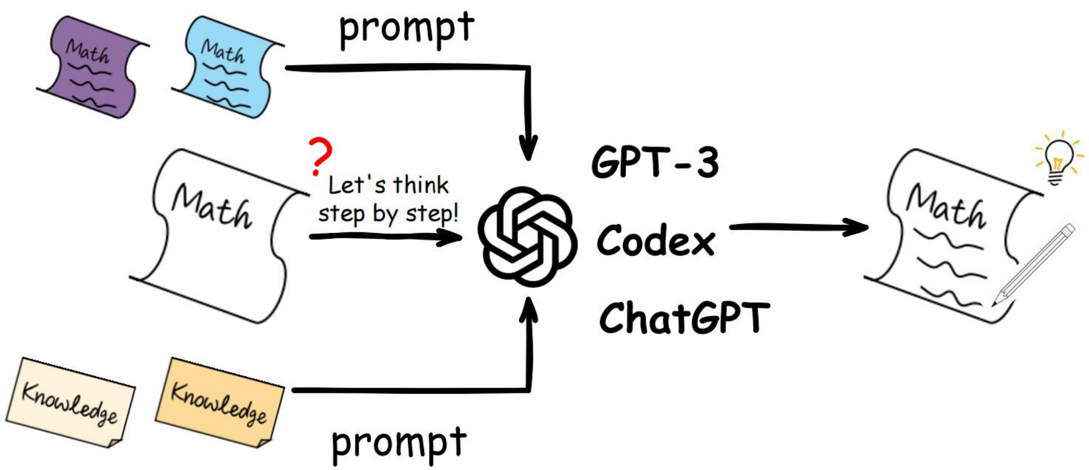
huggingface.co/spaces/zjunlp/MKG_Analogy

求是创新

More Related Works

Reasoning with Language Model Prompting: A Survey

We provide a **review of reasoning** with language model prompting, including comprehensive comparisons, and several research directions.



In the future, we envision **a more potent synergy between the methodologies from the NLP and other domains** and hope sophisticated and efficient LM prompting models will increasingly contribute to improving reasoning performance.

Github Paper-list:
https://github.com/zjunlp/Prompt4ReasoningPapers
Paper: https://arxiv.org/abs/2212.09597

浙江大學
ZHEJIANG UNIVERSITY

Thank You!

leili21@zju.edu.cn

ACCEPT MY ENDLESS GRATITUDE