Complex Query Answering on Neural Knowledge Graphs with Rich Semantics

Yangqiu Song

Department of CSE, HKUST

Slides Credit: Jiaxin Bai and Zihao Wang

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

KnowComp Group

Understanding the World by Computational Knowledge

Knowledge Graphs

- Large-scale knowledge graphs about entities and their attributes (property-of) and relations (thousands of different predicates)
- Developed since Google released its knowledge graph in 2012
  - Millions of entities and concepts
  - Billions of relationships

Google Knowledge Graph (2012)
570 million entities and 18 billion facts

Why is it still Important?

- Large language models (LLMs) tend to better memorize head (popular, more frequent) knowledge

Prompt	Frequent Emails (88)				Infrequent Emails (100)			
	# parsed	# correct	Acc (%)	Hit@5 (%)	# parsed	# correct	Acc (%)	Hit@5 (%)
DP	0	0	0.00	7.95	1	0	0.00	0.00
JP	46	26	29.55	61.36	50	0	0.00	0.00
MJP	85	37	42.04	79.55	97	0	0.00	0.00

Table 1: Email address recovery results on sampled emails from the Enron Email Dataset.

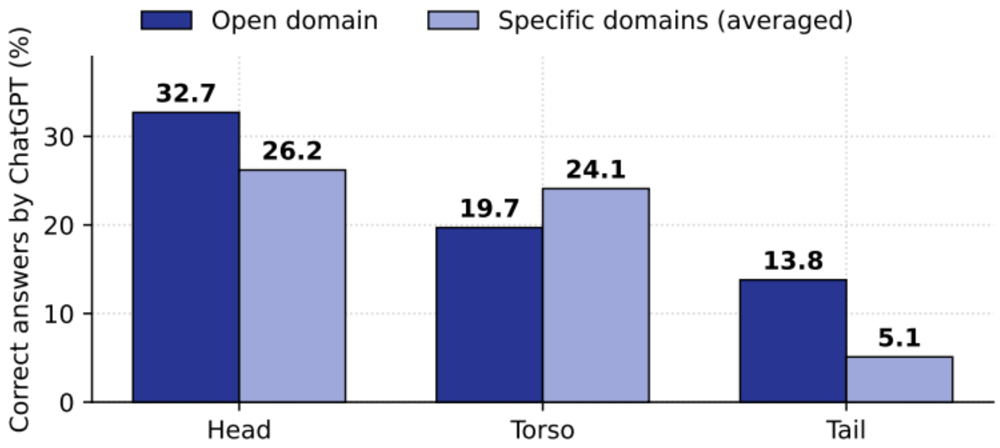
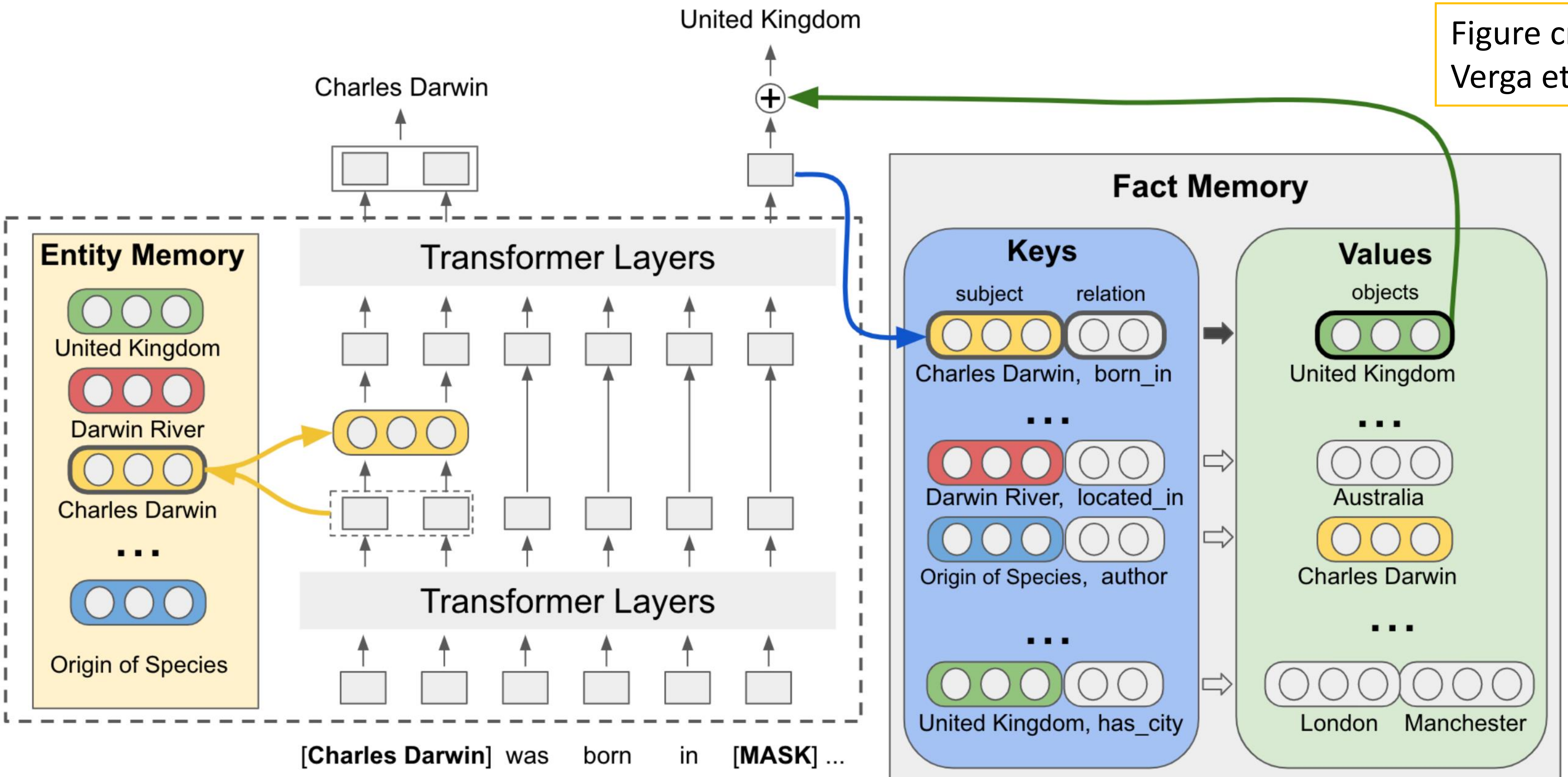


Figure credit: Sun et al., 2023

Haoran Li*, Dadi Guo*, Wei Fan, Mingshi Xu, Jie Huang, Fanpu Meng, Yangqiu Song. Multi-step Jailbreaking Privacy Attacks on ChatGPT. EMNLP (Findings), 2023.
Kai Sun, Yifan Ethan Xu, Hanwen Zha, Yue Liu, Xin Luna Dong. Head-to-Tail: How knowledgeable are Large Language Models? A.K.A. Will LLMs replace knowledge graphs? In arXiv2023.

Entity/Facts as Memories

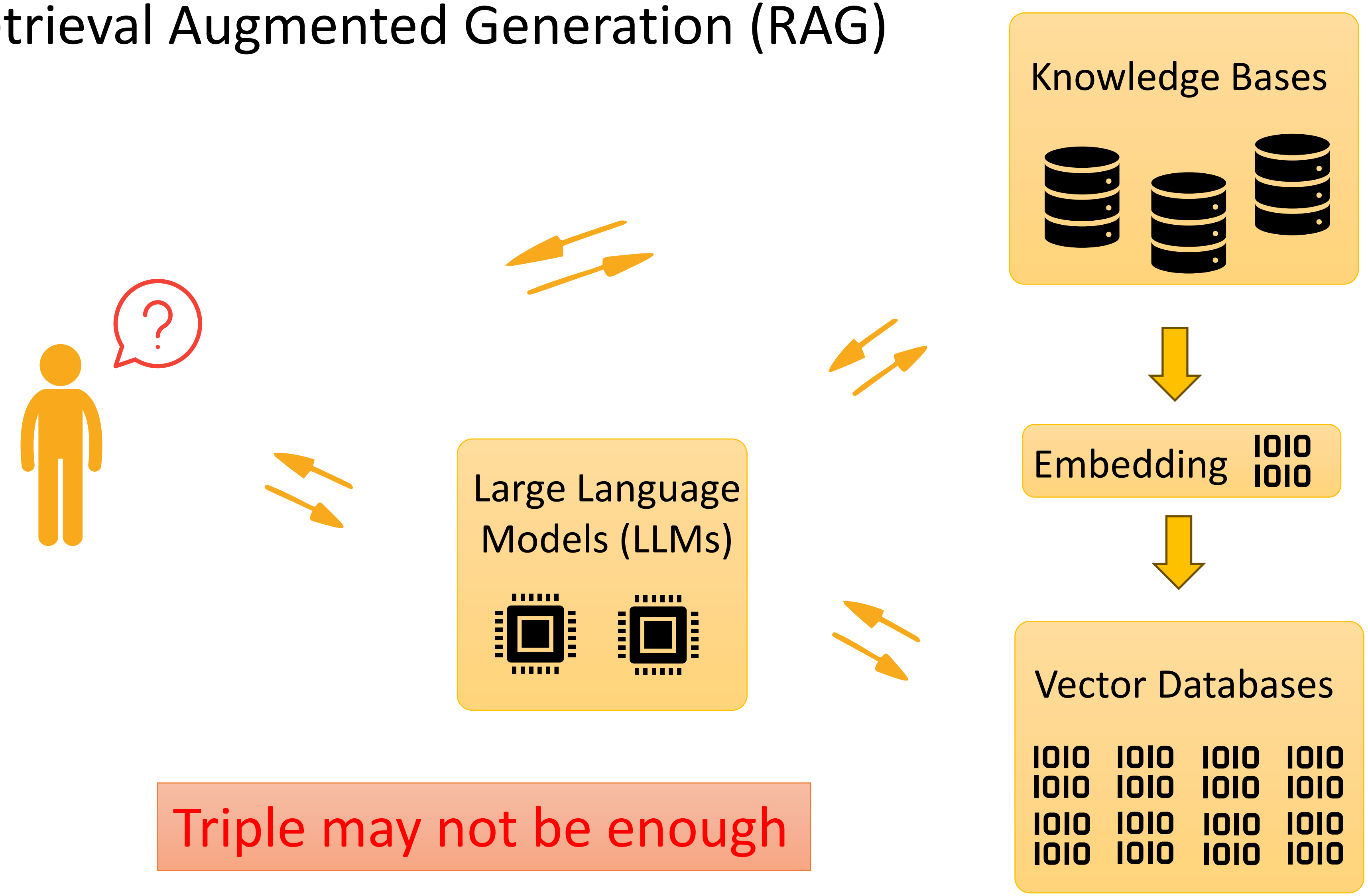
Figure credit:
Verga et al., 2021



Pat Verga, Haitian Sun, Livio Baldini Soares, William W. Cohen: Adaptable and Interpretable Neural Memory Over Symbolic Knowledge. NAACL-HLT 2021: 3678-3691

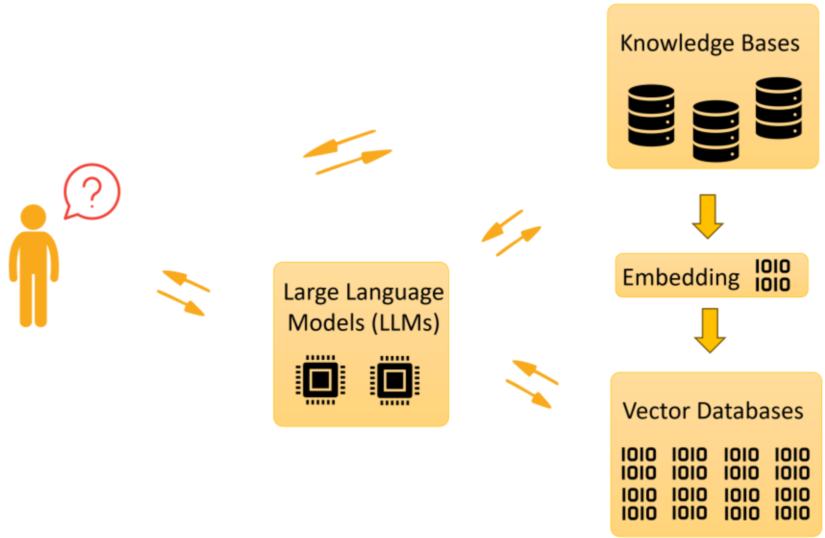
What is Missing?

- Retrieval Augmented Generation (RAG)



Triple may not be enough

Retrieval Augmented Generation



Q: What is the effect of the Fed raising the interest rate?

https://en.wikipedia.org/wiki/Monetary_policy_of_the_United_States
The Federal Reserve's main monetary policy instrument is its Federal funds rate target. By adjusting this target, the Fed affects a wide range of market interest rates and in turn indirectly affects stock prices, wealth and currency exchange rates. …

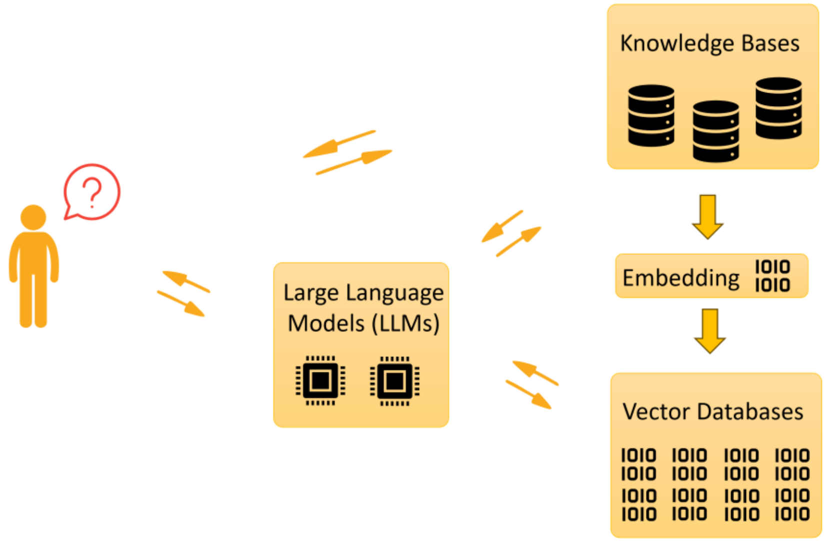
https://en.wikipedia.org/wiki/Federal_funds_rate
Interbank borrowing is essentially a way for banks to quickly raise money. …
Raising the federal funds rate will dissuade banks from taking out such inter-bank loans, which in turn will make cash that much harder to procure.

…
…
…

GPT-3.5-Turbo-Instruct Poe

The effect of the Fed raising interest rates can lead to higher borrowing costs for consumers and businesses, which can slow down economic growth. It can also make it more expensive for people to take out loans, such as mortgages and car loans. However, it can also help to control inflation and stabilize the economy in the long term.

Retrieval Augmented Generation (Cont'd)



Q: What is the effect of the Fed raising the interest rate?

Embedding

Embedding

…
…
…

$$p(a_t | a_{t-1}, \dots, a_1, \boldsymbol{q}, \boldsymbol{c}_1, \dots, \boldsymbol{c}_M)$$

or

$$\sum_{\boldsymbol{c}} p(\boldsymbol{a}|\boldsymbol{q}, \boldsymbol{c}) p(\boldsymbol{c}|\boldsymbol{q})$$

GPT-3.5-Turbo-Instruct Poe

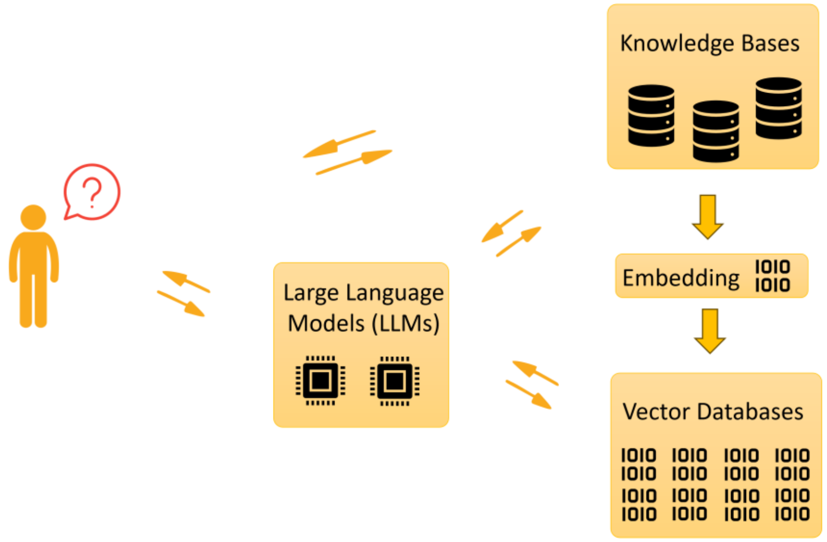
The effect of the Fed raising interest rates can lead to higher borrowing costs for consumers and businesses, which can slow down economic growth. It can also make it more expensive for people to take out loans, such as mortgages and car loans. However, it can also help to control inflation and stabilize the economy in the long term.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, Ming-Wei Chang: Retrieval Augmented Language Model Pre-Training. ICML 2020

Retrieval Augmented Generation (Cont'd)

	Texts	Embeddings
Match	Exact match	Semantic match
Search space	Sparse vectors	Dense vectors
In-context learning	Symbolic	Neural

Retrieval Augmented Generation (Cont'd)



What is the effect of the Fed raising the interest rate?

What is the effect of the Fed raising the interest rate while China dropping the rate?

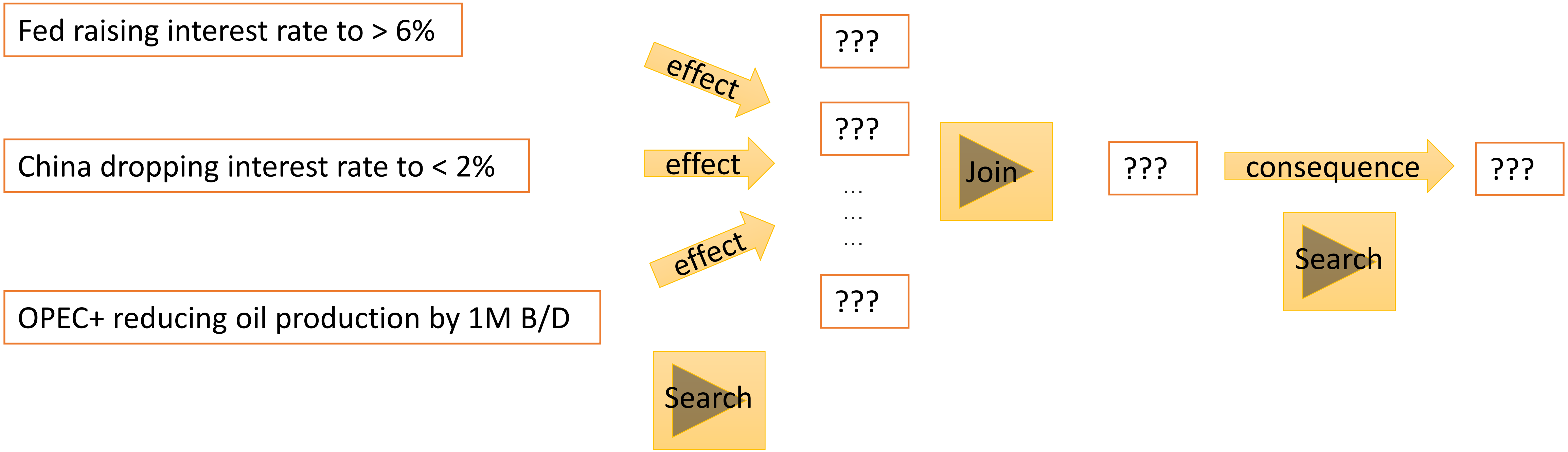
What is the effect of the Fed raising the interest rate while China dropping the rate and OPEC+ reduces oil production?

What is the effect of the Fed raising the interest rate while China dropping the rate and OPEC+ reduces oil production, and the consequence of this effect?

What is the effect of the Fed raising the interest rate to more than 6% while China dropping the rate to less than 2% and OPEC+ reduces oil production by 1 million barrels a day, and the consequence of this effect?

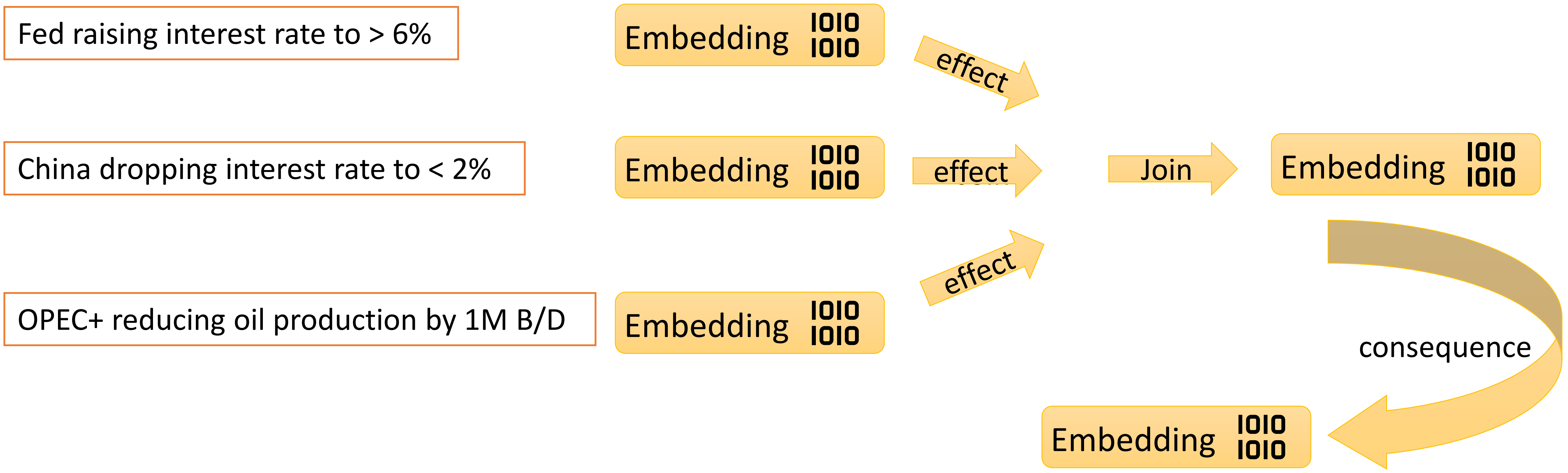
Retrieval Augmented Generation (Cont’d)

What is the effect of the Fed raising the interest rate to more than 6% while China dropping the rate to less than 2% and OPEC+ reduces oil production by 1 million barrels a day, and the consequence of this effect?



Retrieval Augmented Generation (Cont'd)

What is the effect of the Fed raising the interest rate to more than 6% while China dropping the rate to less than 2% and OPEC+ reduces oil production by 1 million barrels a day, and the consequence of this effect?

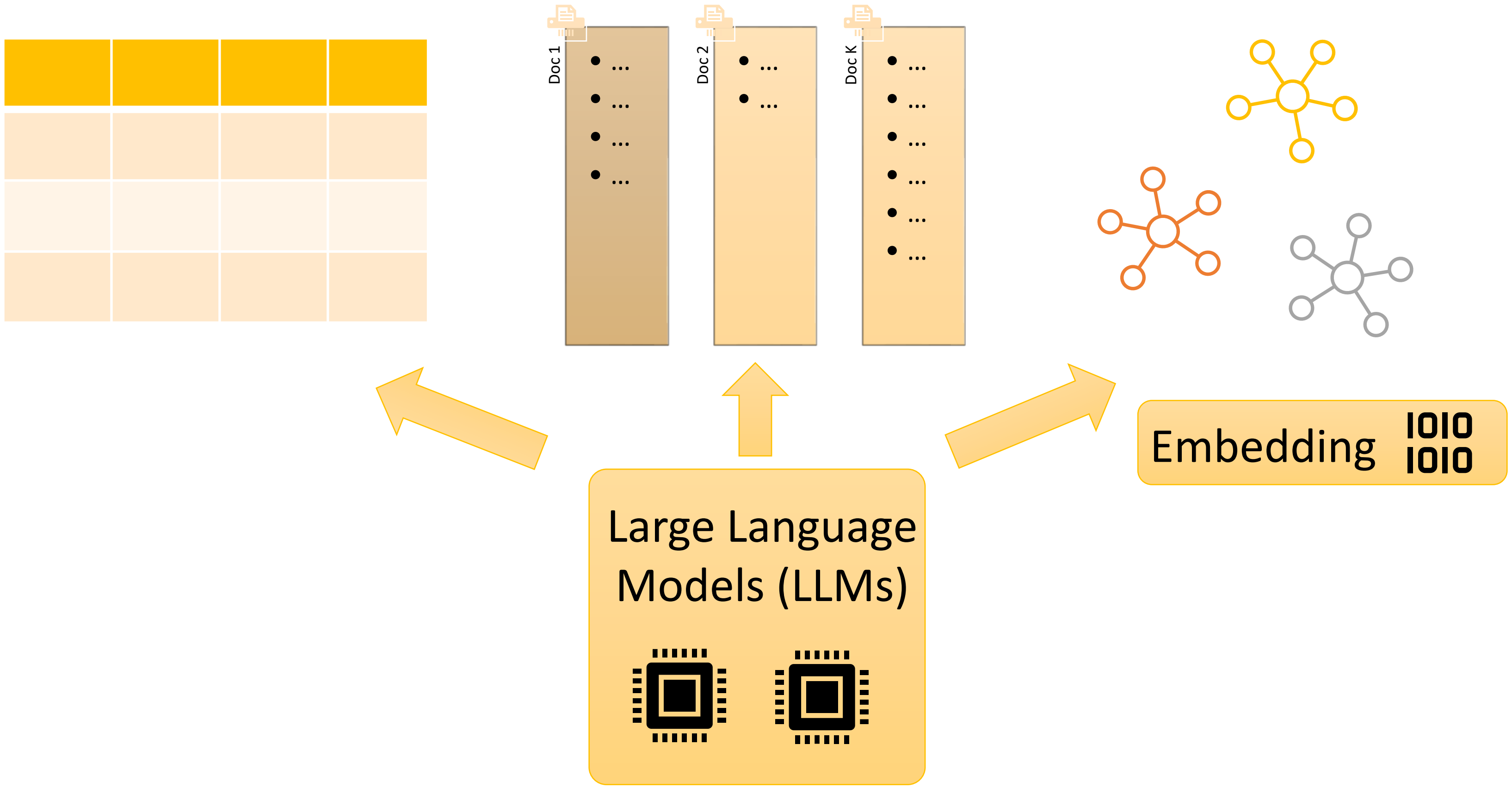


$$p(y_t | y_{t-1}, \dots, y_1, \boldsymbol{q}, \boldsymbol{c}_1, \dots, \boldsymbol{c}_M)$$

More Generally

- Traditional NoSQL or Graph databases
- Semantic match of strings

Can we support complex queries such as join, intersection, counting, etc., on top of that?



Neural Graph Databases

- Graph structure + vector storage
  - Leveraging the power of LLMs for textual data
- Query executor to support complex queries
  - Query encoder is trainable
  - Inductive method to be robust with insertion, deletion, and modification
  - Fuzzy semantic search
  - Generalizable to incomplete data

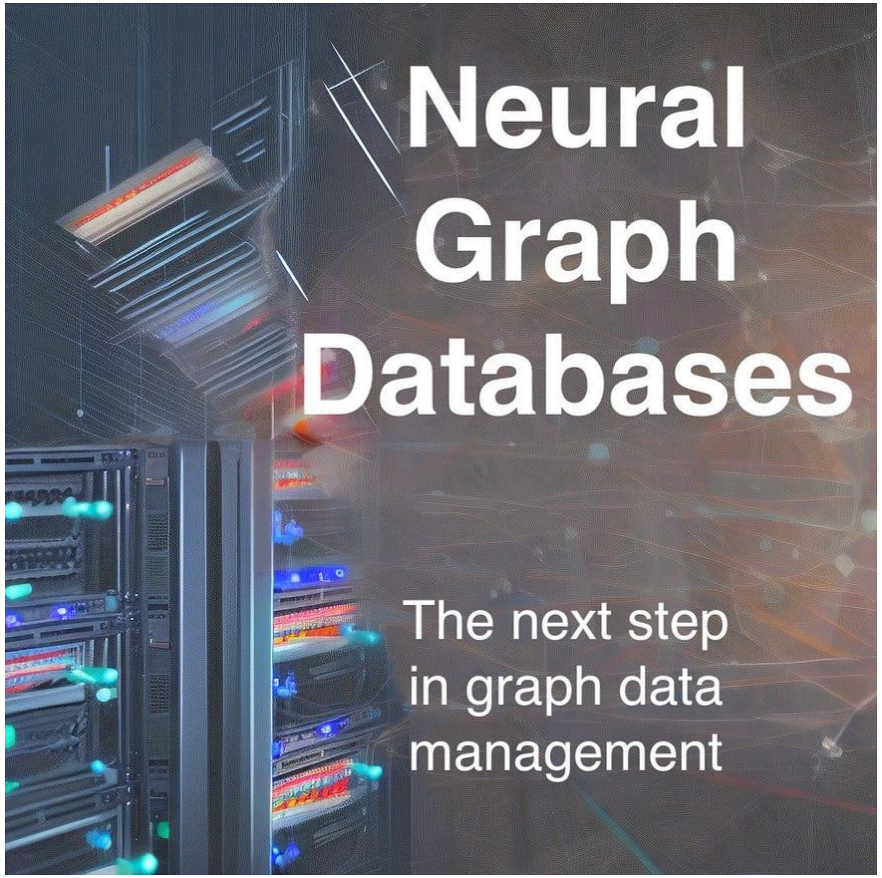
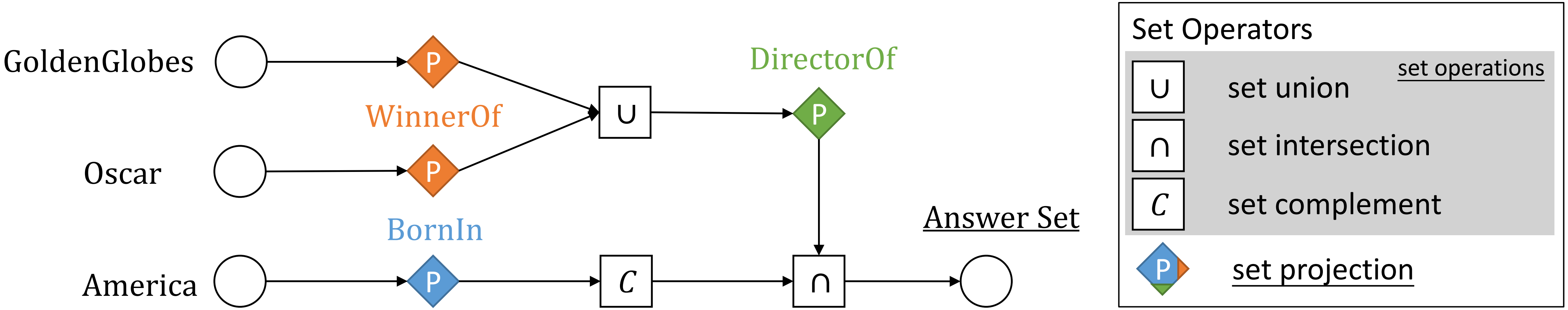


Figure taken from: https://towardsdatascience.com/neural-graph-databases-cc35c9e1d04f

Complex Queries on Neuralized Knowledge Graphs

- A working example: Tree-Formed Queries (TFQ):
  - Tree-form query family contains the queries that can be converted into the computational tree

Natural Language: Find non-American directors whose movie won Golden Globes or Oscar?
Logical Formula: $q = V_? \exists V_1 . (\mathrm{Won}(V_1, \mathrm{GoldenGlobes}) \vee \mathrm{Won}(V_1, \mathrm{Oscar})) \wedge \neg \mathrm{BornIn}(V_?, \mathrm{America}) \wedge \mathrm{Direct}(V_?, V_1)$
Set Operator Tree: $\mathrm{DirectorOf}(\mathrm{WinnerOf}(\mathrm{GoldenGlobes}) \cup \mathrm{WinnerOf}(\mathrm{Oscar})) \cap \mathrm{BornIn}(\mathrm{America})^C$



Example from: Zihao Wang, Weizhi Fei, Hang Yin, Yangqiu Song, Ginny Y Wong, and Simon See . Wasserstein-Fisher-Rao Embedding: Logical Query Embeddings with Local Comparison and Global Transport In Findings of ACL 2023

The Design Space of Neural TFQ Answering

Concept	Definition	Comment
Entity set	$\mathcal{E}$	The entity set in KG
Relation set	$\mathcal{R}$	The relation set in KG
Set embedding space	$\mathcal{X}$	Embedding space
Set embedding lookup	$E_{\mathcal{X}}: \mathcal{E} \mapsto \mathcal{X}$	Singleton set embedding
Entity embedding space	$\mathcal{Y}$	Embedding space
Entity embedding lookup	$E_{\mathcal{Y}}: \mathcal{E} \mapsto \mathcal{Y}$	Entity embedding
Set intersection	$I: \mathcal{X} \times \cdots \times \mathcal{X} \mapsto \mathcal{X}$	Binary or N-ary
Set union	$U: \mathcal{X} \times \cdots \times \mathcal{X} \mapsto \mathcal{X}$	Binary or N-ary
Set complement	$C: \mathcal{X} \mapsto \mathcal{X}$	Replaceable with set difference
Set projection	$P: \mathcal{X} \times \mathcal{R} \mapsto \mathcal{X}$	One-hop link prediction
Scoring function	$s: \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$	How much an entity is in a set

Converting to computational tree makes it possible to model set operations with neural networks

Set Operators

set operations

- $\cup$ set union
- $\cap$ set intersection
- C set complement
- P set projection

Wang, Z., Yin, H., & Song, Y. (2021). Benchmarking the combinatorial generalizability of complex query answering on knowledge graphs. *arXiv preprint arXiv:2109.08925*.

The Design Space of Neural TFQ Answering

KnowComp

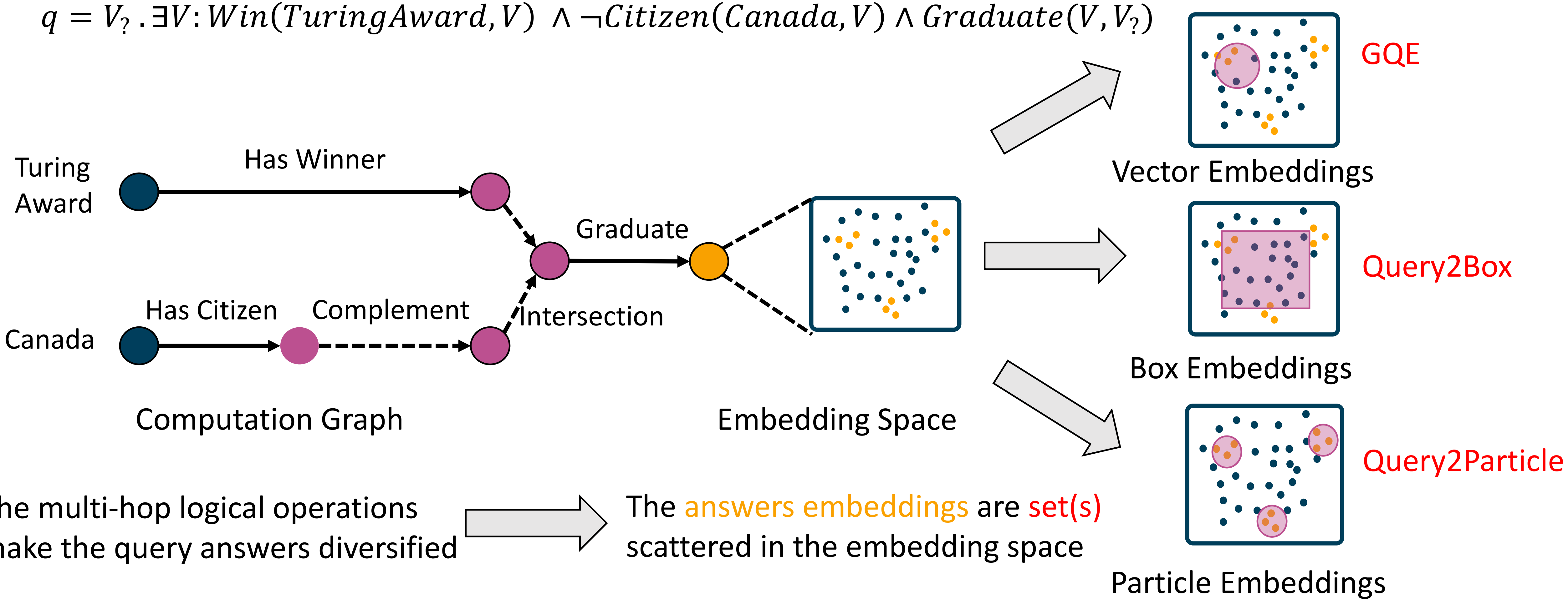
Understanding the World by Computational Knowledge

Concept	Definition	Comment
~~Entity set~~	$\mathcal{E}$	Known notation
~~Relation set~~	$\mathcal{R}$	Known notation
Set embedding space	$\mathcal{X}$	[Query Embedding: Slot 1]
~~Set embedding lookup~~	$\cancel{E_{\mathcal{X}}: \mathcal{E} \mapsto \mathcal{X}}$	Simplified
Entity embedding space	$\mathcal{Y}$	[Entity embedding: Slot 2]
~~Entity embedding lookup~~	$\cancel{E_{\mathcal{Y}}: \mathcal{E} \mapsto \mathcal{Y}}$	Simplified
Set intersection	$I: \mathcal{X} \times \cdots \times \mathcal{X} \mapsto \mathcal{X}$	[Slot 3]
Set union	$U: \mathcal{X} \times \cdots \times \mathcal{X} \mapsto \mathcal{X}$	[Slot 4]
Set complement	$C: \mathcal{X} \mapsto \mathcal{X}$	[Slot 5]
Set projection	$P: \mathcal{X} \times \mathcal{R} \mapsto \mathcal{X}$	[Slot 6]
Scoring function	$s: \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$	[Slot 7]

Each method will be introduced by filling 7 slots

Embedding Space and Set Representations

$$q = V_? \,.\, \exists V : Win(TuringAward, V) \wedge \neg Citizen(Canada, V) \wedge Graduate(V, V_?)$$

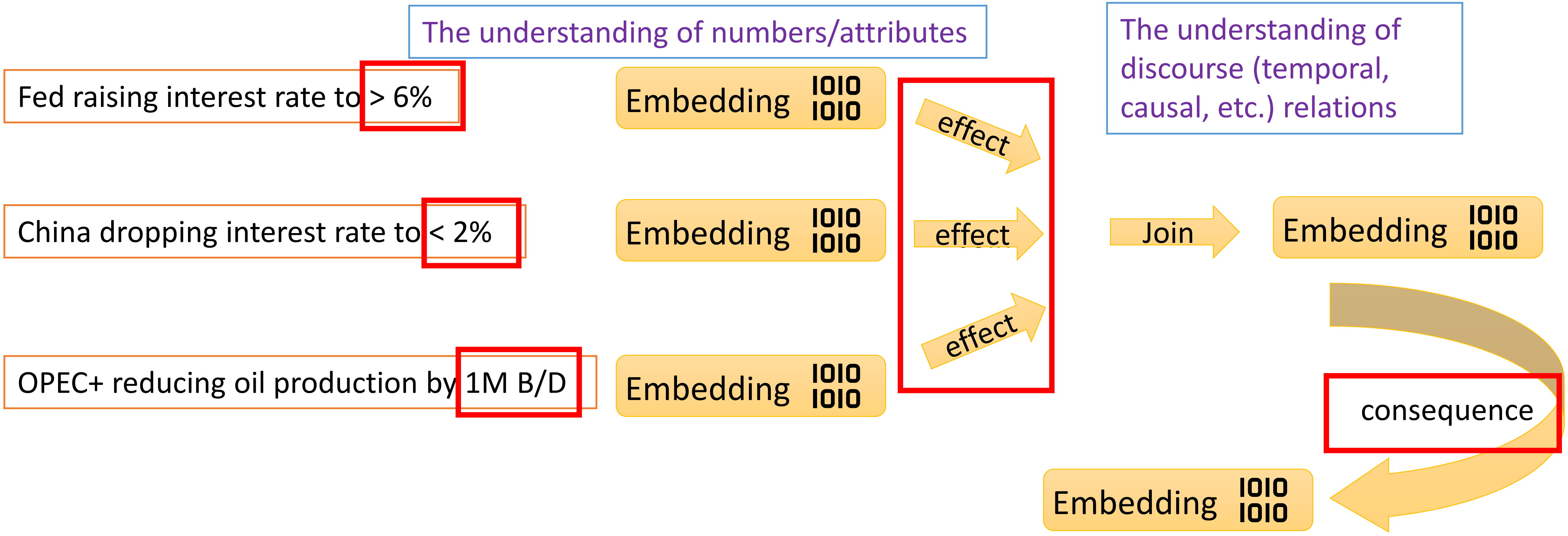


The multi-hop logical operations make the query answers diversified → The answers embeddings are set(s) scattered in the embedding space

William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, Jure Leskovec. Embedding Logical Queries on Knowledge Graphs. NeurIPS 2018.
Hongyu Ren, Weihua Hu, Jure Leskovec. Query2box: Reasoning over Knowledge Graphs in Vector Space using Box Embeddings. ICLR 2020.
Example from: Jiaxin Bai, Zihao Wang, Hongming Zhang, Yangqiu Song: Query2Particles: Knowledge Graph Reasoning with Particle Embeddings. NAACL-HLT (Findings) 2022.

What's Still Missing to Support RAG?

What is the effect of the Fed raising the interest rate to more than 6% while China dropping the rate to less than 2% and OPEC+ reduces oil production by 1 million barrels a day, and the consequence of this effect?



$$p(y_t | y_{t-1}, \dots, y_1, \boldsymbol{q}, \boldsymbol{c}_1, \dots, \boldsymbol{c}_M)$$

This Talk

- Neural KG CQA on Entities and Numerical Values
- Neural KG CQA on Eventuality Knowledge Graphs

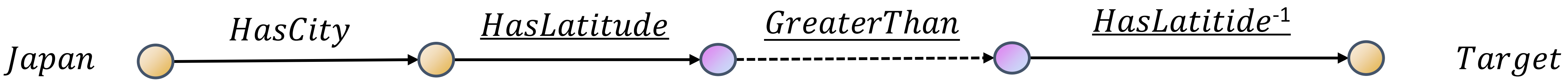
Numerical Complex Query Answering

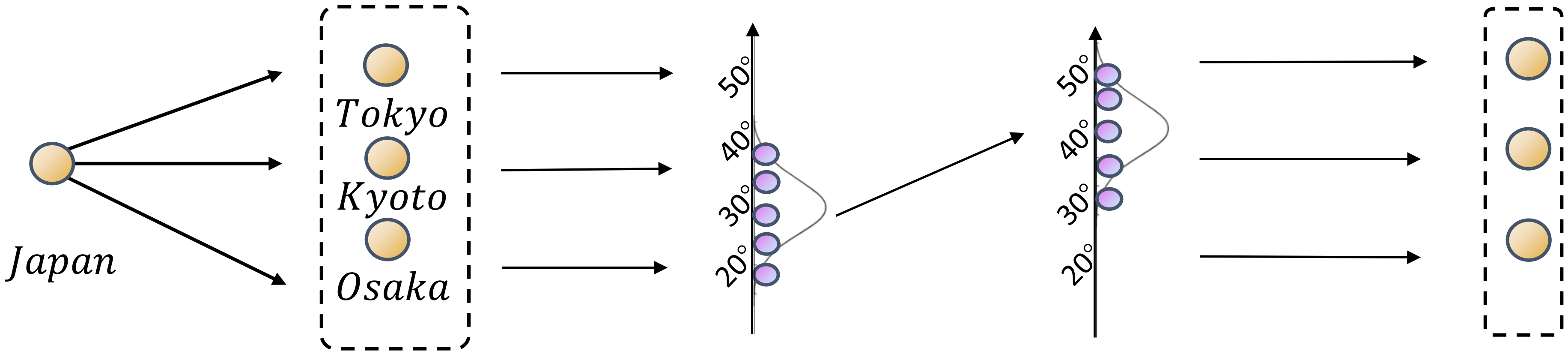
Category	Complex Queries	Interpretations
Numerical CQA	$q_2 = V_? \,.\, \exists X_1, X_2 \colon Win(V_?, TuringAward) \wedge \underline{GreaterThan}(1927, X_2) \wedge \underline{BornIn}\,(V_?, X_2)$	Find the Turing award winners that is born before the year of 1927.
Numerical CQA	$q_3 = V_? \,.\, \exists X_1, X_2 \colon LocatedIn(V_?, UnitedStates) \wedge \underline{HasLatitude}(V_?, X_1) \wedge \underline{GreaterThan}(X_1, X_2) \wedge \underline{HasLatitude}(Beijing, X_2)$	Find the states in US that have a higher latitudes than Beijing.
Numerical CQA	$q_4 = V_? \,.\, \exists X_1, X_2, X_3 \colon LocatedIn(V_?, UnitedStates) \wedge \underline{HasPopulation}(V_?, X_1) \wedge \underline{SmallerThan}(X_1, X_2) \wedge \underline{TimesByTwo}(X_2, X_3) \wedge \underline{HasPopulation}(California, X_3)$	Find the states in US that have a twice smaller population than California?

Number Reasoning Network

Find the cities that have a higher latitudes than Japanese cities.

$q = V_? \,.\, \exists V_1, X_1, X_2 : \underline{HasLatitude}(V_?, X_2) \wedge \underline{GreaterThan}(X_2, X_1) \wedge \underline{HasLatitude}(V_1, X_1) \wedge LocatedIn(V_1, Japan)$





(1) Relational Projection

(2) Attribute Projection

(3) Numerical Projection

(4) Reverse Attribute Projection

Jiaxin Bai, Chen Luo, Zheng Li, Qingyu Yin, Bing Yin, Yangqiu Song: Knowledge Graph Reasoning over Entities and Numerical Values. KDD 2023: 57-68

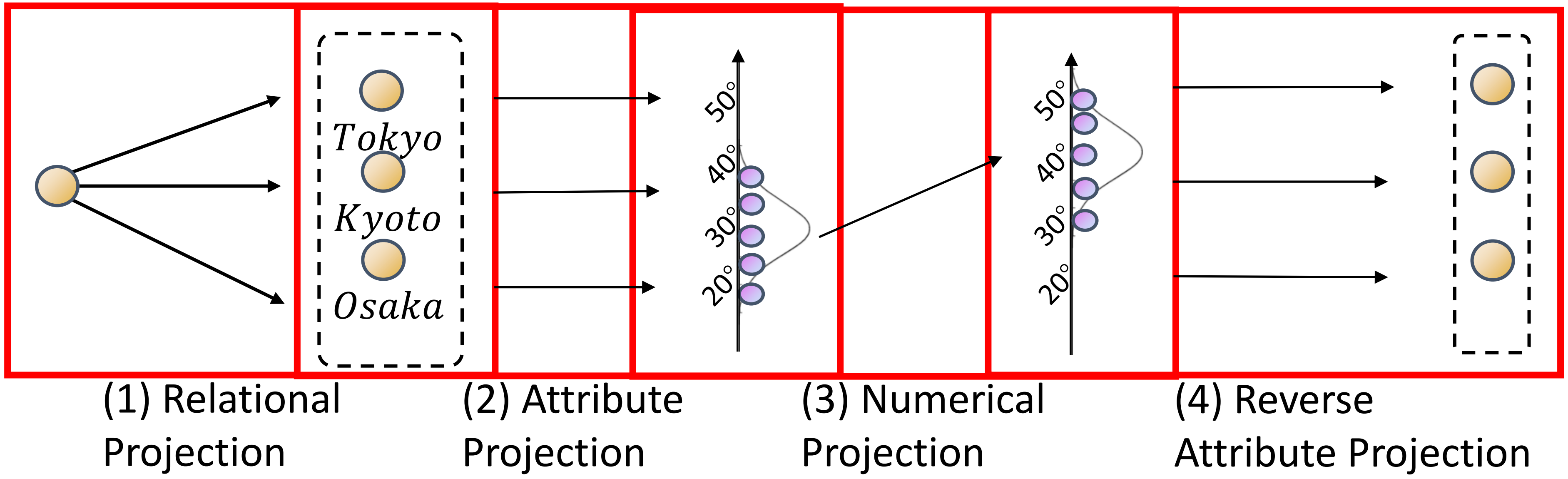
Number Reasoning Network

(1) Relational Projection (rp):
Query Embedding → Entity Set

(2) Attribute Projection (ap):
Query/Set Embedding → Value Distribution

(3) Numerical Projection (np):
Value Distribution → Value Distribution

(4) Reverse Attribute Projection (rap):
Value Distribution → Query Embedding



Number Reasoning Network

(1) Relational Projection:
Adopted from the backbones:
GQE, Query2Box, Query2Particles.

(2) (3) (4) Other Projections: Gated Transitions

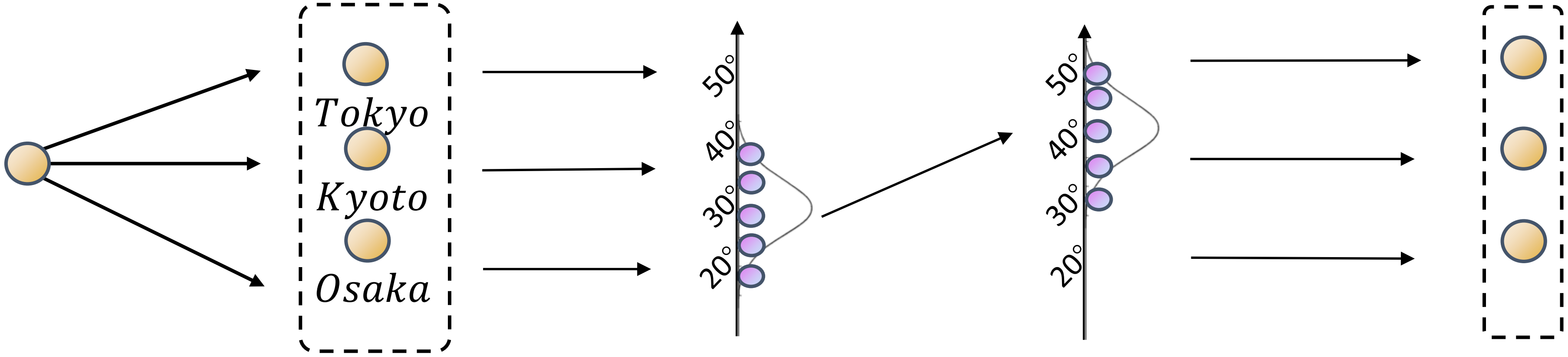
$$p_i = W_p^p q^i + b_p^p$$ Linear projection

$$z_i = \sigma\,(W_z^p e_a + U_z^p p_i + b_z^p)$$

$$r_i = \sigma\,(W_r^p e_a + U_r^p p_i + b_r^p)$$

$$t_i = \varphi\,(W_h^p e_a + U_h^p (r_i \odot p_i) + b_h^p)$$ MLP

$$\theta_{i+1} = (1 - z_i) \odot p_i + z_i \odot t_i$$ Gate selection



(1) Relational Projection

(2) Attribute Projection

(3) Numerical Projection

(4) Reverse Attribute Projection

Number Reasoning Network

Entity embeddings:
Adopted from the backbones:
GQE, Query2Box, Query2Particles.

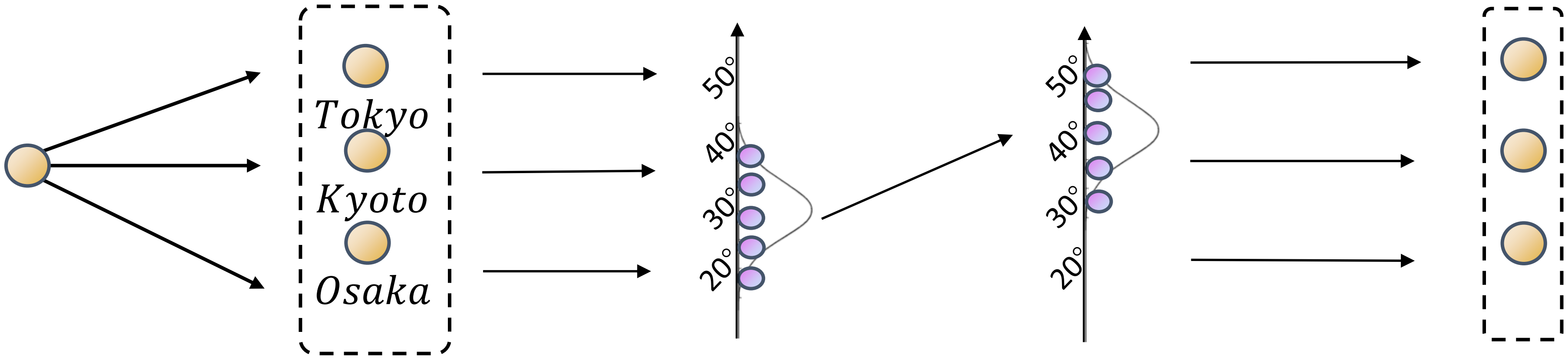
Input number embeddings

- DICE

$$\psi(v)_d = \begin{cases} \sin^{d-1}(\alpha)\cos(\alpha) \\ \sin^{D}(\alpha) \end{cases}$$

- Sinusoidal

$$\psi(v)_d = \begin{cases} \sin\dfrac{v}{v^{d/D}}, & d \equiv 0 \ (mod\ 2) \\ \cos\dfrac{v}{v^{(d-1)/D}}, & d \equiv 1 \ (mod\ 2) \end{cases}$$



(1) Relational Projection

(2) Attribute Projection

(3) Numerical Projection

(4) Reverse Attribute Projection

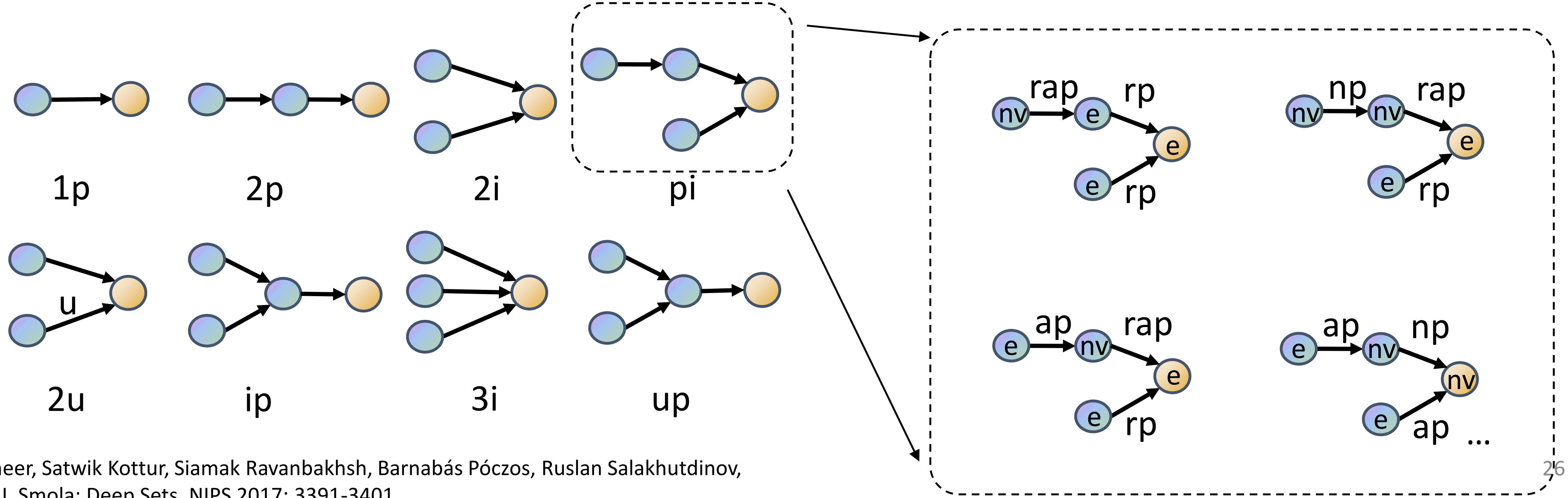
Dhanasekar Sundararaman, Shijing Si, Vivek Subramanian, Guoyin Wang, Devamanyu Hazarika, and Lawrence Carin. 2020. Methods for Numeracy-Preserving Word Embeddings. In EMNLP. 4742–4753
Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. NeurIPS 2017

Number Reasoning Network

- Logic Operators on Entities, adopted from the backbones:
  - GQE, Query2Box, Query2Particles.
- Logic Operator on Value Distribution:
  - Intersection and Union: DeepSet

$$a_i = Attn(W_q\theta_i^T, W_k\theta_i^T, W_v\theta_i^T)^T$$
$$\theta_{i+1} = MLP(a_i)$$

- Relational Projection (rp)
- Attribute Projection (ap)
- Numerical Projection (np)
- Reverse Attribute Projection (rap)



Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, Alexander J. Smola: Deep Sets. NIPS 2017: 3391-3401

Number Embeddings and Learning Objective

Use maximize a posteriori probability (MAP) estimation to derive an objective function for type-aware reasoning.

v is the positive answer value.
t is the type of this value like date, length, size etc.
θ_I is the distribution parameters in the last step.
$\psi(v)$ is number embeddings

$$\hat{\theta}_I(v,t) = \arg\max_{\theta_I} f(\theta_I|v,t)$$

$$= \arg\max_{\theta_I} \; f(\psi(v)|\theta_I,t)g(\theta_I|t)$$ (Bayes' Rule; Remove the denominator: a constant in argmax)

$$= \arg\min_{\theta_I} \; -\log f(\psi(v)|\theta_I) - \log g(\theta_I|t)$$ (Conditional Independent of v and t on θ_I)

Set the parameteration as: $f(\psi(v)|\theta_I) = p_{\theta_I}(\psi(v))$ and $g(\theta_I|t) = \phi_t(\theta_I)$

happenedOnDate	date_of_death
createdOnDate	date_of_birth
latitude	film_release_date
longitude	org.date_founded
person.height_mt	location.date_founded

$$L_A = \frac{1}{M}\sum_{j=1}^{M}\left(-\log p_{\theta_I^{(j)}}\left(\psi(v^{(j)})\right) - \log \phi_{t^{(j)}}(\theta_I^{(j)})\right)$$

Jiaxin Bai, Chen Luo, Zheng Li, Qingyu Yin, Bing Yin, Yangqiu Song: Knowledge Graph Reasoning over Entities and Numerical Values. KDD 2023: 57-68

Number Embeddings and Learning Objective

End-to-end training by Joint optimization of two losses:

$$L_A = \frac{1}{M}\sum_{j=1}^{M}\left(-\log \boxed{p_{\theta_I^{(j)}}\left(\psi(v^{(j)})\right)} - \log \boxed{\phi_{t^{(j)}}(\theta_I^{(j)})}\right)$$

The likelihood of the value $v^{(j)}$ sampled from distribution of $\theta_I^{(j)}$

The likelihood of the distribution parameter $\theta_I^{(j)}$ is of type $t^{(j)}$

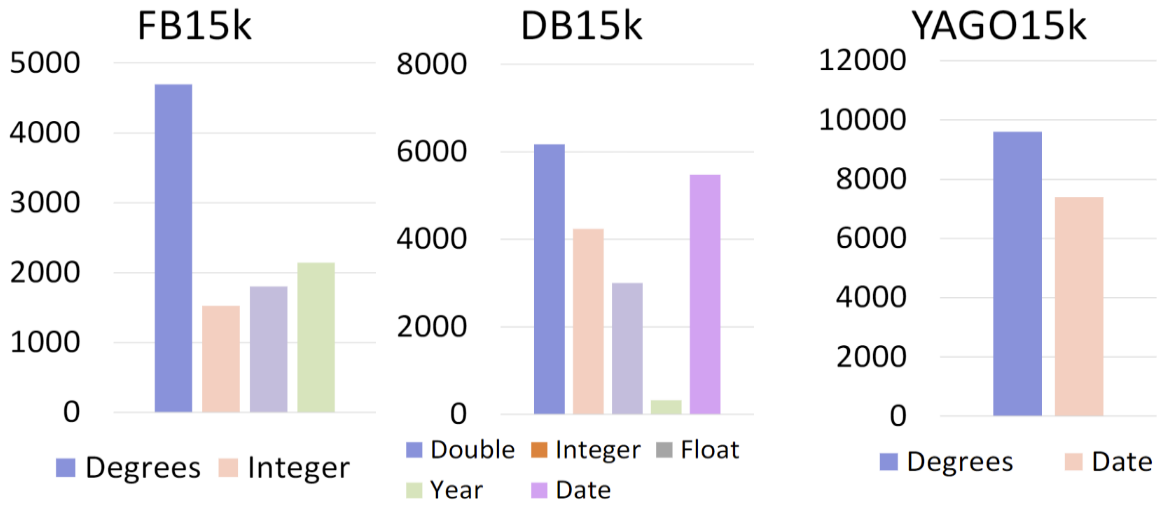
$$L_E = -\frac{1}{N}\sum_{j=1}^{N}\log \boxed{p(q_I^{(j)}, v^{(j)})}$$

The likelihood of the entity $v^{(j)}$ is the answer of the query encoding $q_I^{(j)}$.

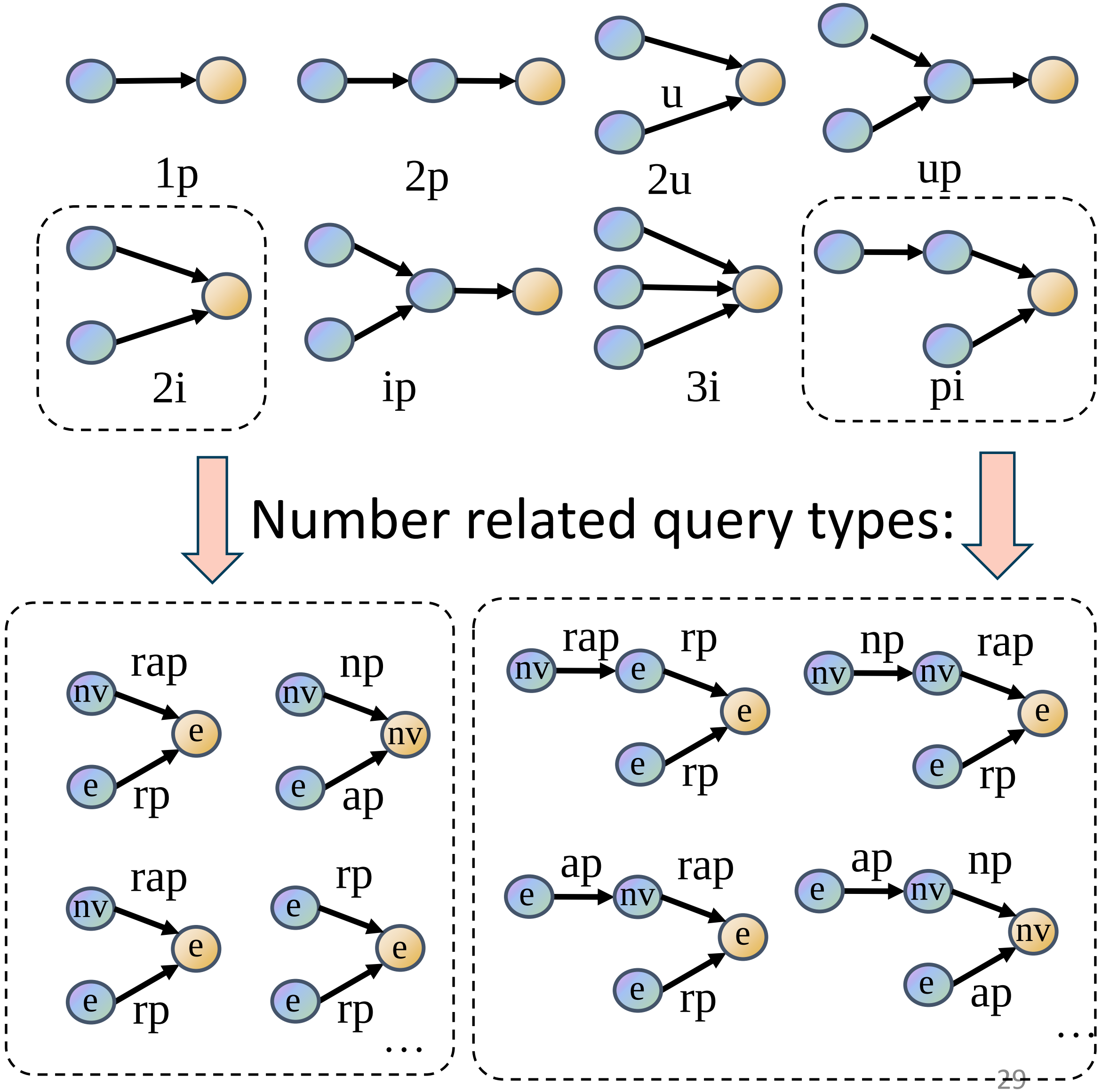
j is means the j-th sample, and I means the last step of distribution parameter encoding.

Sampling Data

Numerical Values Types on Different KGs:



General Query Types:



Data Statistics

Graphs	Data Split	#Nodes	#Rel.	# Attr.	#Rel. Edges	#Attr. Edges	#Rev. Attr. Edges	#Num. Edges	#Edges
FB15K	Training	25,106	1,345	15	947,540	20,248	20,248	27,020	1,015,056
	Validation	26,108	1,345	15	1,065,982	22,779	22,779	27,376	1,138,916
	Testing	27,144	1,345	15	1,184,426	25,311	25,311	27,389	1,262,437
DB15K	Training	31,980	279	30	145,262	33,131	33,131	25,495	237,019
	Validation	34,191	279	30	161,978	37,269	37,269	25,596	262,112
	Testing	36,358	279	30	178,394	41,411	41,411	25,680	286,896
YAGO15K	Training	32,112	32	7	196,616	21,732	21,732	26,616	266,696
	Validation	33,078	32	7	221,194	22,748	22,748	26,627	293,317
	Testing	33,610	32	7	245,772	23,520	23,520	26,631	319,443

Data Statistics

Graphs	Data Split	1p	2p	2i	3i	pi	ip	2u	up	All
FB15K	Training	304,633	138,192	226,729	288,874	260,057	233,834	284,301	284,931	2,021,551
	Validation	8,271	15,860	23,359	28,836	25,081	22,930	29,187	29,210	182,734
	Testing	7,969	15,431	23,346	28,865	24,810	22,232	29,212	29,274	181,139
DB15K	Training	124,851	99,698	140,427	190,413	171,353	163,687	190,364	194,244	1,275,037
	Validation	3,529	10,388	9,792	13,817	14,594	16,651	19,512	19,792	108,075
	Testing	3,387	10,047	9,914	14,603	14,642	15,897	19,504	19,773	107,767
YAGO15K	Training	84,014	76,238	136,282	183,850	162,712	145,994	183,963	183,459	1,156,512
	Validation	2,833	7,986	10,757	16,884	13,485	13,899	18,444	19,105	103,393
	Testing	2,713	7,949	10,935	17,171	13,481	13,526	18,433	18,997	103,205

Main Results on Three KGs

Query Encoding	Attribute	Hit@1	Hit@3	Hit@10	MRR
GQE	Baseline	10.33	18.19	27.91	16.29
	NRN + DICE	11.03	19.18	29.01	17.15
	NRN + Sinusoidal	**11.14**	**19.39**	**29.23**	**17.31**
Q2P	Baseline	10.22	17.35	26.61	15.81
	NRN + DICE	11.86	19.70	29.46	17.84
	NRN + Sinusoidal	**12.25**	**20.16**	**29.96**	**18.28**
Q2B	Baseline	11.81	20.93	31.19	18.41
	NRN + DICE	12.52	22.09	32.34	19.34
	NRN + Sinusoidal	**12.75**	**22.22**	**32.46**	**19.51**

This Talk

- Neural KG CQA on Entities and Numerical Values
- Neural KG CQA on Eventuality Knowledge Graphs

CQA on Eventuality Knowledge Graph

KnowComp
Understanding the World by Computational Knowledge

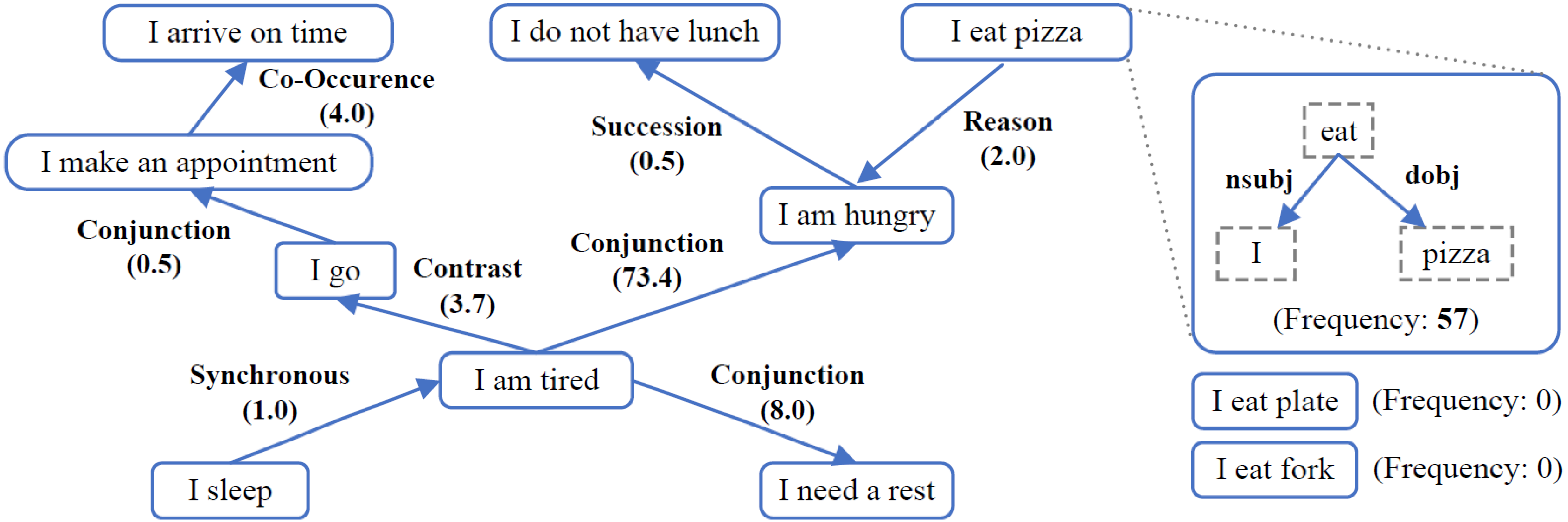
Complex query on eventuality graphs are **different** from the entity-relation graph

Whether and **when** the eventualities occur are important

Queries	Type	Interpretations
$q_1 = V_?.\exists V: \text{Interact}(V_?, V) \wedge \text{Assoc}(V, \mathit{Alzheimer}) \wedge \text{Assoc}(V, \mathit{MadCow})$	Entity	Find the substances that interact with the proteins associated with Alzheimer's and Mad cow disease.
$q_2 = V_?.\text{Precedence}(\mathit{Food\ is\ bad}, \mathit{PersonX\ add\ soy\ sauce}) \wedge \text{Reason}(\mathit{Food\ is\ bad}, V_?)$	Eventuality	Food is bad before PersonX add soy sauce. What is the reason for food being bad?
$q_3 = V_?.\text{Precedence}(V_?, \mathit{PersonX\ go\ home}) \wedge \text{ChosenAlternative}(\mathit{PersonX\ go\ home}, \mathit{PersonX\ buy\ an\ umbrella})$	Eventuality	Instead of buying an umbrella, PersonX go home. What happened before PersonX go home?

Jiaxin Bai, Xin Liu, Weiqi Wang, Chen Luo, Yangqiu Song: Complex Query Answering on Eventuality Knowledge Graph with Implicit Logical Constraints. NeurIPS, 2023

ASER (Activities, States, Events, and their Relations)



Principle 1: Comparing semantic meanings by fixing grammar (Katz and Fodor, 1963)

Principle 2: The need of language inference based on 'partial information' (Wilks, 1975)

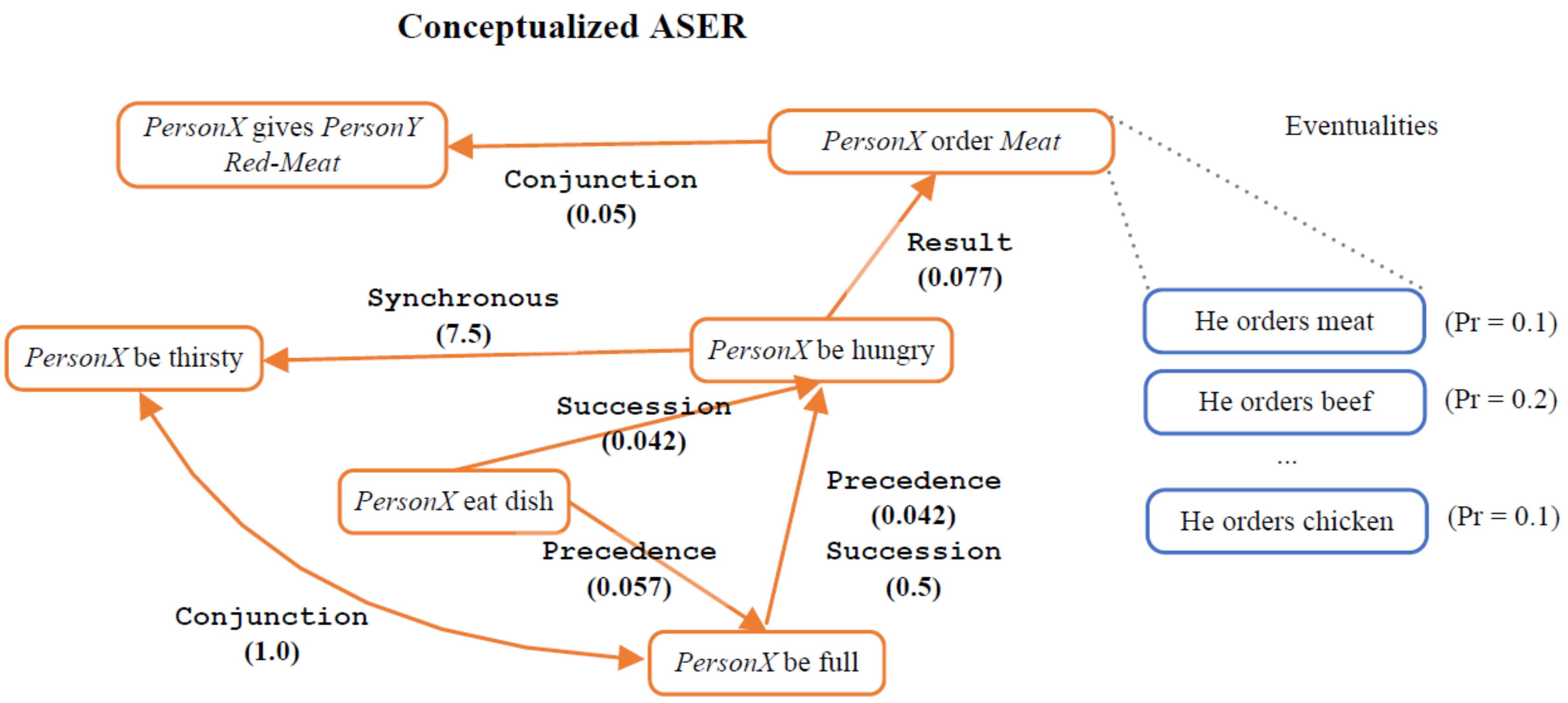
https://github.com/HKUST-KnowComp/ASER

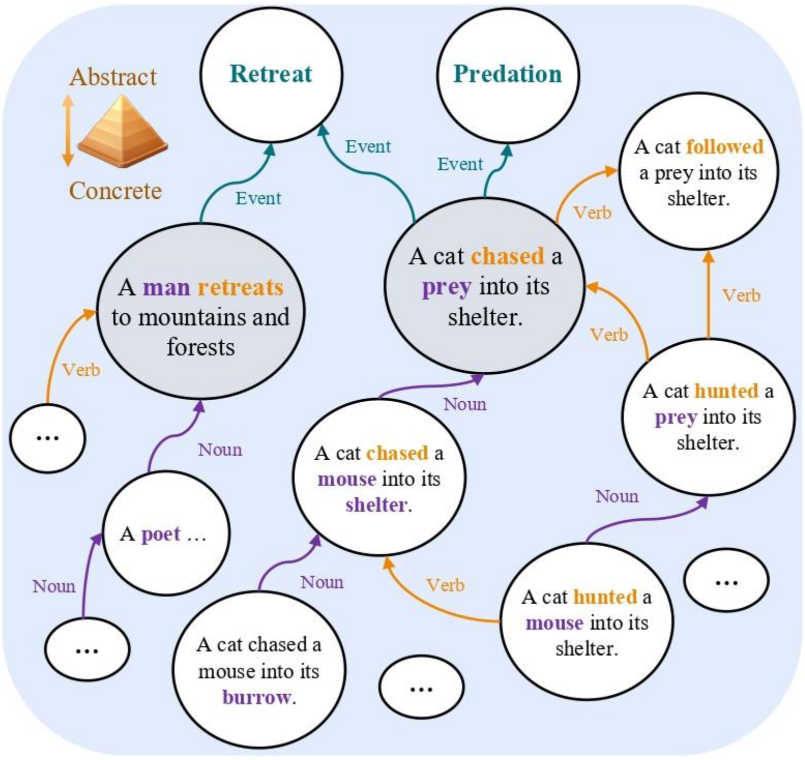
Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, Cane Wing-Ki Leung: ASER: A Large-scale Eventuality Knowledge Graph. WWW 2020: 201-211

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. Language, 39(2), 170–210.

Yorick Wilks. 1975. An intelligent analyzer and understander of English. Communications of the ACM, 18(5):264–274.

Conceptualization and Normalization





https://github.com/HKUST-KnowComp/ASER

Hongming Zhang, Xin Liu, Haojie Pan, Haowen Ke, Jiefu Ou, Tianqing Fang, Yangqiu Song: ASER: Towards large-scale commonsense knowledge acquisition via higher-order selectional preference over eventualities. Artif. Intell. 309: 103740 (2022)

https://github.com/HKUST-KnowComp/AbsPyramid

Zhaowei Wang, Haochen Shi, Weiqi Wang, Tianqing Fang, Hongming Zhang, Sehyun Choi, Xin Liu, Yangqiu Song: AbsPyramid: Benchmarking the Abstraction Ability of Language Models with a Unified Entailment Graph. CoRR abs/2311.09174 (2023)

Discourse Relations and Implicit Constraints

- PersonX did not eat anything because PersonX was full

Reason(*PersonX did not eat anything, PersonX was full*)

Occurrence Constraint

$\eta(\textit{PersonX did not eat anything}) \wedge \eta(\textit{PersonX was full})$ ^
$\wedge\ \eta(\textit{PersonX did not eat anything}) \leftarrow \eta(\textit{PersonX was full})$

Temporal Constraints

$\tau(\textit{PersonX did not eat anything}) \succ \tau(\textit{PersonX was full})$

$\eta(A)$ = 1 if and only if it occurs
$\tau(A) \succ \tau(\mathrm{B})$: A happens after B

Discourse Relations and Implicit Constraints

- Food is bad before PersonX add soy sauce

Precedence(*Food is bad, PersonX adds soy sauce*)

Occurrence Constraint

$\eta(\textit{Food is bad}) \wedge \eta(\textit{PersonX adds soy sauce})$

Temporal Constraints

$\tau(\textit{Food is bad}) \prec \tau(\textit{PersonX adds soy sauce})$

$\tau(A) \prec \tau(\mathrm{B})$: A happens before B
$\eta(A)$ = 1 if and only if it occurs

Discourse Relations and Implicit Constraints

- Instead of buying an umbrella, PersonX go home

ChosenAlternative($PersonX$ go home, $PersonX$ buy an umbrella)

Occurrence Constraint

$\eta(PersonX\ go\ home) \land \neg\, \eta(PersonX\ buy\ an\ umbrella)$

$\eta(A)$ = 1 if and only if it occurs

Logical Constraints behind Discourse Relations

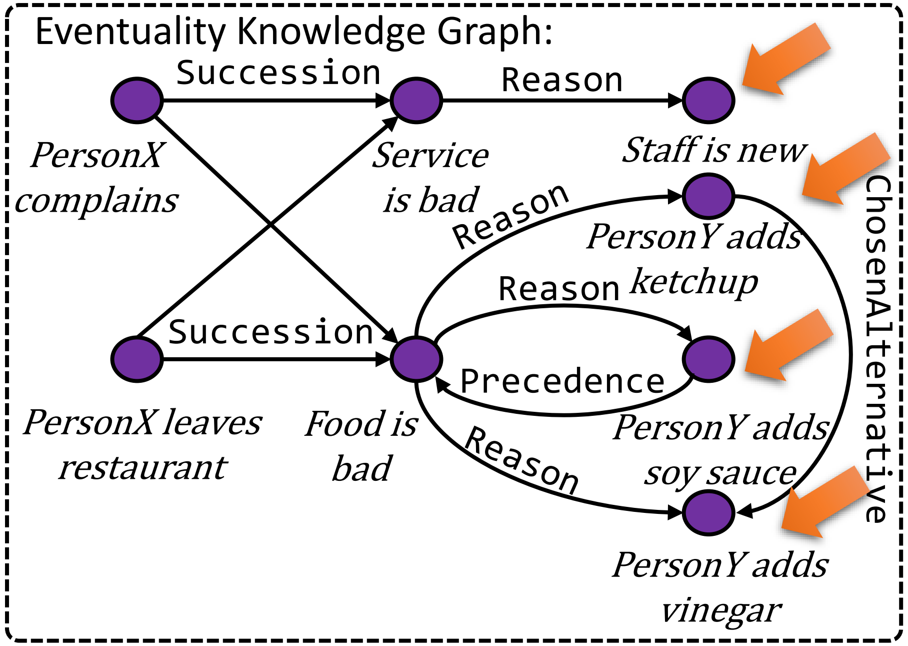
Discourse Relations	Semantics	Implicit Constraints	
		Occurrence Constraints	Temporal Constraints
Precedence(A, B)	A occurs before B.	$\eta(A) \wedge \eta(B)$	$\tau(A) \prec \tau(B)$
Succession(A, B)	A occurs after B happens.	$\eta(A) \wedge \eta(B)$	$\tau(A) \succ \tau(B)$
Synchronous(A, B)	A occurs at the same time as B.	$\eta(A) \wedge \eta(B)$	$\tau(A) = \tau(B)$
Reason(A, B)	A occurs because B.	$\eta(A) \wedge \eta(B) \wedge (\eta(A) \leftarrow \eta(B))$	$\tau(A) \succ \tau(B)$
Result(A, B)	A occurs as a result B.	$\eta(A) \wedge \eta(B) \wedge (\eta(A) \rightarrow \eta(B))$	$\tau(A) \prec \tau(B)$
Condition(A, B)	If B occurs, A.	$\eta(A) \rightarrow \eta(B)$	$\tau(A) \succ \tau(B)$
Concession(A, B)	B occurs, although A.	$\eta(A) \wedge \eta(B)$	-
Constrast(A, B)	B occurs, but A.	$\eta(A) \wedge \eta(B)$	-
Conjunction(A, B)	A and B both occur.	$\eta(A) \wedge \eta(B)$	-
Instantiation(A, B)	B is a more detailed description of A.	$\eta(A) \wedge \eta(B)$	-
Restatement(A, B)	A restates the semantics of B.	$\eta(A) \leftrightarrow \eta(B)$	-
Alternative(A, B)	A and B are alternative situations.	$\eta(A) \vee \eta(B)$	-
ChosenAlternative(A, B)	Instead of B occurs, A.	$\eta(A) \wedge \neg \eta(B)$	-
Exception(A, B)	A, except B.	$\neg \eta(A) \wedge \eta(B) \wedge (\neg \eta(B) \rightarrow \eta(A))$	-

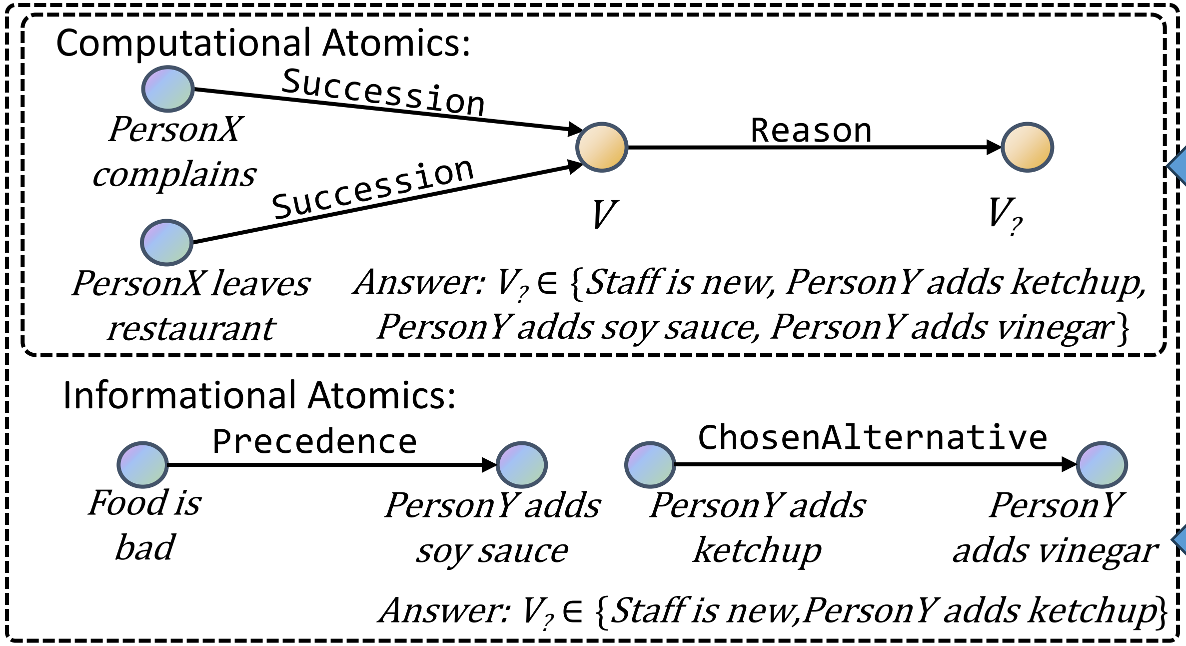
$\tau(A) \prec \tau(B)$: A happens before B, $\eta(A)$ = 1 if and only if it occurs

Logical Query with Implicit Constraints

Question: Food is bad before PersonY adds soy sauce. Instead of adding vinegar, PersonY adds ketchup. PersonX complains after V. PersonX leaves the restaurant after V. The reason V is V'?. What is V'?

Query on Graph: $q = V_?\,.\ \exists V$: Succession(*PersonX complains, V*) ∧ Succession(*PersonX leaves restaurant, V*) ∧ Reason(*V, $V_?$*) ∧ Precedence(*Food is bad, PersonY adds soy sauce*) ∧ ChosenAlternative(*PersonY adds ketchup, PersonY adds vinegar*)

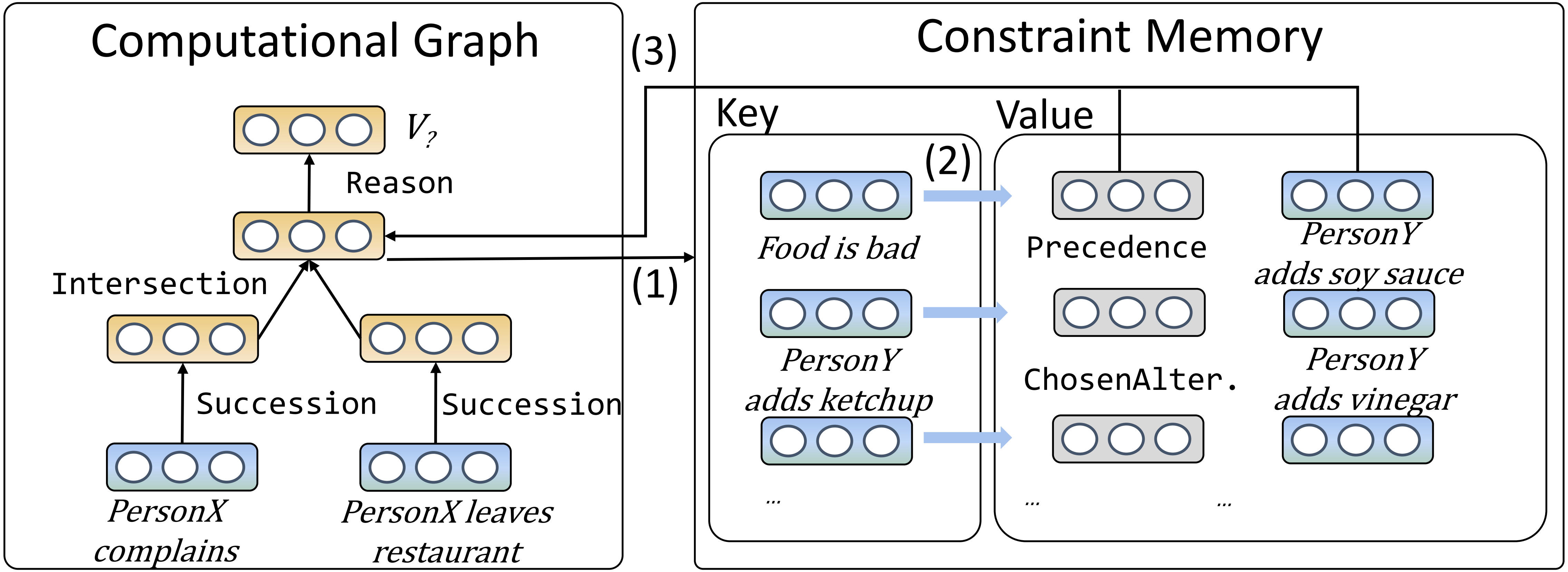




Without context and its constraints: 4 answers

With Implicit Constraints: Only 2 answers

Query Encoding with Constraint Memory



(1) $s_{i,m} = < q_i,\ c_h^{(m)} >$
Computes the relevance of query embedding to the head of the memory key at position m.

(2) $v_i = \sum_{m=1}^{M} s_{i,m} \left(c_r^{(m)} + c_t^{(m)}\right)$
Computes the aggregated memory values across M memory cells with the importance weighted by relevance scores.

(3) $q_i = q_i + MLP(v_i)$
Computes the query embedding with memory values with the help of a MLP layer.

Jiaxin Bai, Xin Liu, Weiqi Wang, Chen Luo, Yangqiu Song: Complex Query Answering on Eventuality Knowledge Graph with Implicit Logical Constraints. NeurIPS, 2023

Complex Eventuality Queries on ASER

Data Split	Types	Query with Occurrence Constraints			Query with Temporal Constraints		
		#Queries	#Answers	#Contr. Answers	#Queries	#Answers	# Contr. Answers
Train Queries	6	58,797	4.74	1.44	18,033	4.37	1.09
Validation Queries	15	22,320	7.20	1.63	19,637	8.85	1.41
Test Queries	15	24,466	7.93	1.68	20,788	10.88	1.46

- The tables shows the types and number queries;
- The number of answers on ASER;
- The number of logically contradictory answers.

The MEQE Combined with Various QE methods

Models	Occurrence Constraints			Temporal Constraints			Average		
	Hit@1	Hit@3	MRR	Hit@1	Hit@3	MRR	Hit@1	Hit@3	MRR
GQE	8.92	14.21	13.09	9.09	14.03	12.94	9.12	14.12	13.02
+ MEQE	**10.20**	**15.54**	**14.31**	**10.70**	**15.67**	**14.50**	**10.45**	**15.60**	**14.41**
Q2P	14.14	19.97	18.84	14.48	19.69	18.68	14.31	19.83	18.76
+ MEQE	**15.15**	**20.67**	**19.38**	**16.06**	**20.82**	**19.74**	**15.61**	**20.74**	**19.56**
Nerual MLP	13.03	19.21	17.75	13.45	19.06	17.68	13.24	19.14	17.71
+ MEQE	**15.26**	**20.69**	**19.32**	**15.91**	**20.63**	**19.47**	**15.58**	**20.66**	**19.40**
FuzzQE	11.68	18.64	17.07	11.68	17.97	16.53	11.68	18.31	16.80
+ MEQE	**14.76**	**21.12**	**19.45**	**15.31**	**21.01**	**19.49**	**15.03**	**21.06**	**19.47**

Conclusions

- Complex query answering on neural knowledge graphs/bases has great potential to support retrieval augmented generation (RAG)

- Two things are missing
  - Number and attribute understanding
  - Discourse relation modeling for logical queries

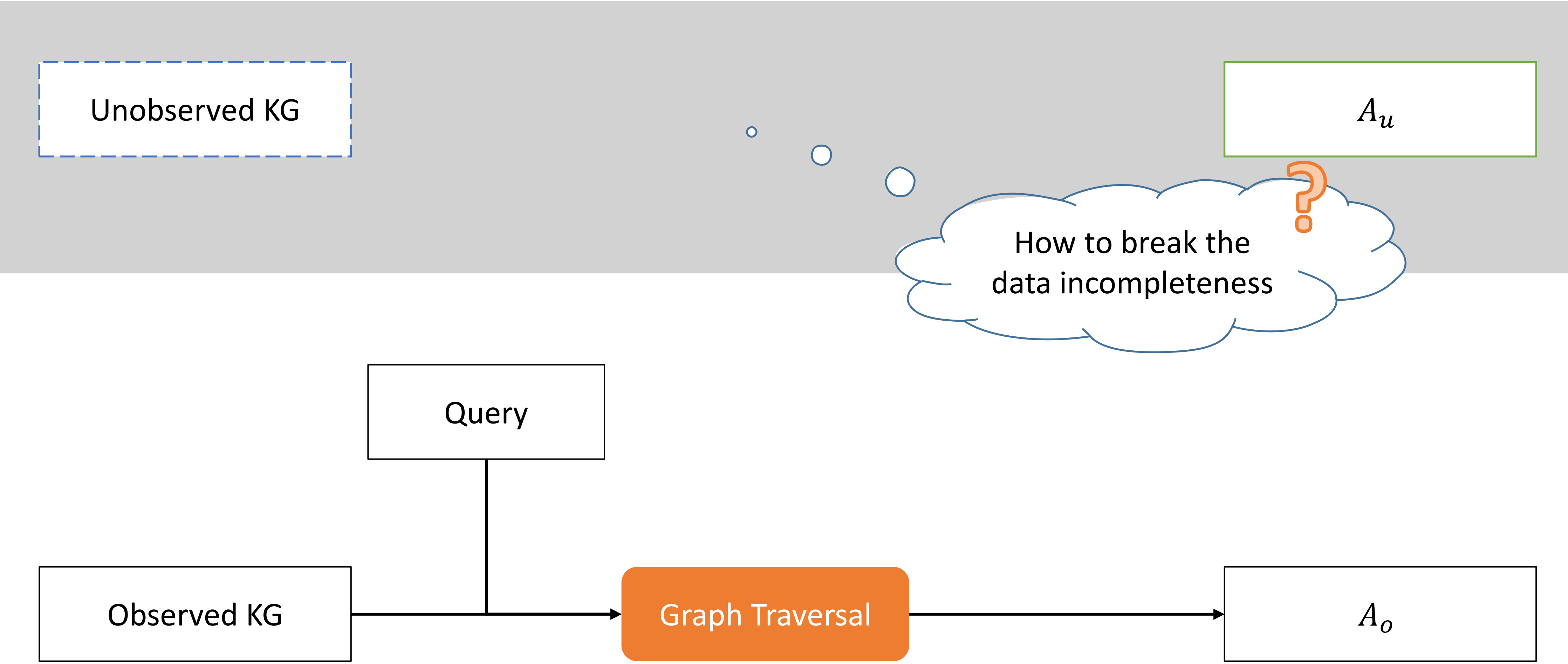
- We implemented models that can handle numbers and discourse relations

Future Work

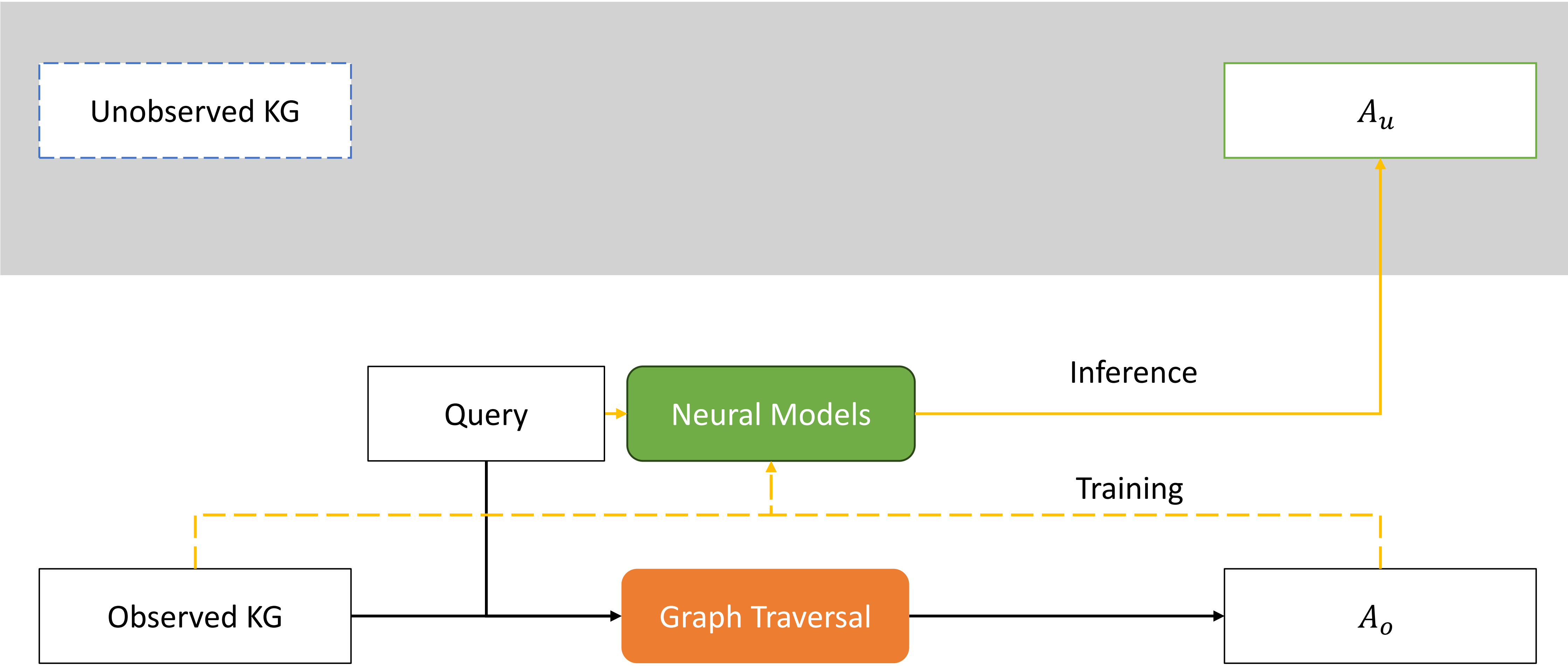
- Incorporating complex query answering into retrieval augmented generation

- Exploring more strategies for complex query answering with rich semantics to handle Open World Problems

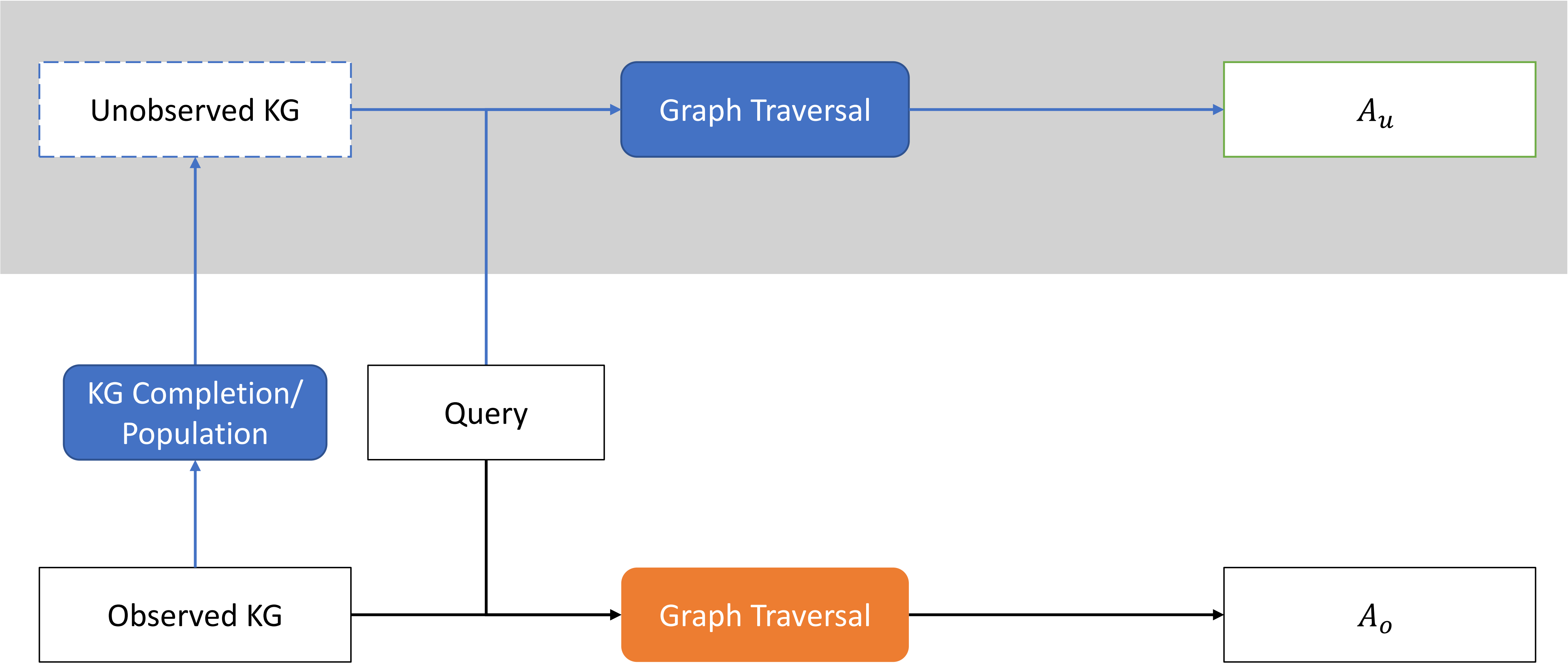
The Challenge of the Open World Problem



Neural strategy: End-to-end Training



Symbolic Strategy: Completion and Search



Other Works on CQA

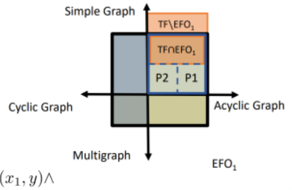
Data

Benchmarking EFO-1 (Existential First-Order Queries with Single Free Variable)

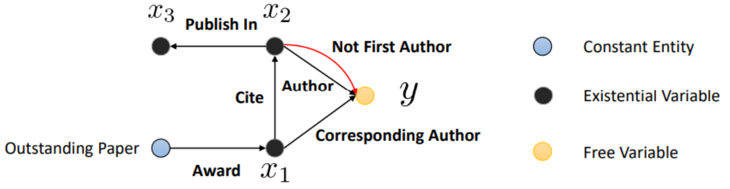
CQA Dataset	Support Operators									Support EPFO	Support EFO-1	Num. of Forms	Num. of Test Query Types
	e	p	i	I	u	U	n	d	D				
Q2B dataset [15]	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓*	✗	1	9
HypE dataset [6]	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓*	✗	1	9
BetaE dataset [16]	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓*	✓*	2	14
EFO-1-QA (ours)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	9	301

NeurIPS'21

EFO-1 queries with cycles



$\exists x_1 \exists x_2 \exists x_3$, Award(OutstandingPaper, x_1)∧CorrespondingAuthor(x_1, y)∧ Cite(x_1, x_2) ∧ PublishIn(x_2, x_3) ∧ Author(x_2, y) ∧ ¬FirstAuthor(x_2, y)

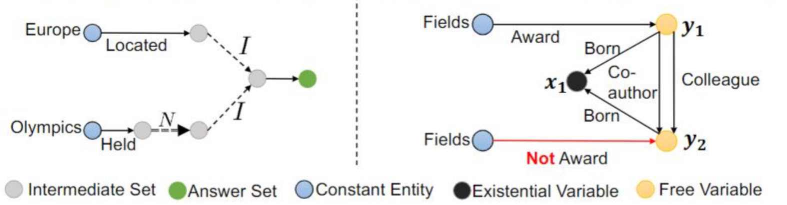


ICLR'24

EFO-K more than one variables

$\exists x_1$. Located(Europe,x_1) ∧ ¬Held(Olympics,x_1)∧Presdient(x_1, y_1)

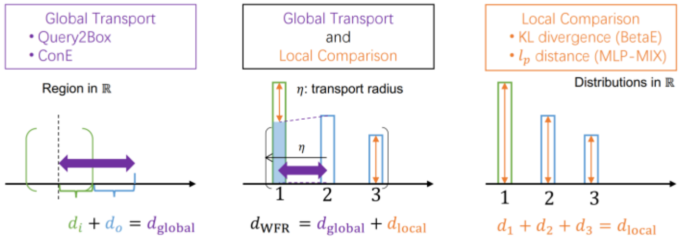
$\exists x_1$. Award(Fields,y_1)∧ ¬Award(Fields,y_2)∧ Born(y_1, x_1) ∧ Born(y_1, x_2) ∧Colleague(y_1, y_2) ∧Co-author(y_1, y_2)



Arxiv'23

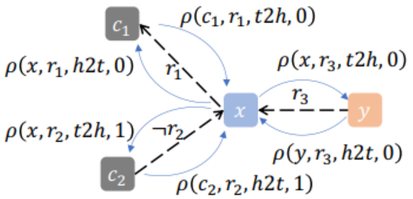
Models

Query encoder with OT



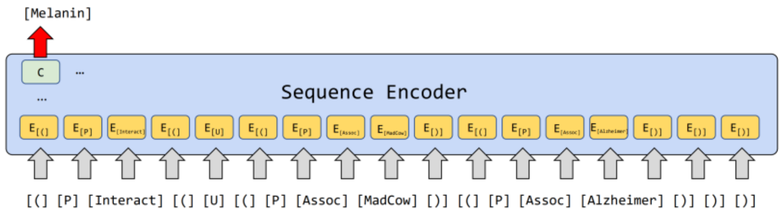
ACL Findings'23

Learning in the inference step as a GNN (one-hop logical inference based MPNN)



ICLR'23

Sequence encoder of queries



TMLR'23

Jiaxin Bai*, Tianshi Zheng*, Yangqiu Song. Sequential Query Encoding For Complex Query Answering on Knowledge Graphs. Transactions of Machine Learning Research. 2023

Zihao Wang, Yangqiu Song, Ginny Y. Wong, and Simon See. Logical Message Passing Networks with One-hop Inference on Atomic Formulas. In The Eleventh International Conference on Learning Representations, ICLR 2023

Zihao Wang, Weizhi Fei, Hang Yin, Yangqiu Song, Ginny Y Wong, and Simon See . Wasserstein-Fisher-Rao Embedding: Logical Query Embeddings with Local Comparison and Global Transport In Findings of ACL 2023

Hang Yin, Zihao Wang, and Yangqiu Song. EFO_k-CQA: Towards Knowledge Graph Complex Query Answering beyond Set Operation. Arxiv 2023

Hang Yin, Zihao Wang, and Yangqiu Song. Rethinking Existential First Order Queries and their Inference on Knowledge Graphs. Arxiv 2023

Zihao Wang, Hang Yin, and Yangqiu Song. Benchmarking the Combinatorial Generalizability of Complex Query Answering on Knowledge Graphs. In NeurIPS Datasets and Benchmarks Track, 2021

Thank you for your attention ☺

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

KnowComp Group
Understanding the World by Computational Knowledge