Compositional Reasoning over

Natural Language Leveraging Neuro-Symbolic AI

Parisa Kordjamshidi
kordjams@msu.edu

Michigan State University, USA

Feb 26th, 2024

Nuclear Workshop

Deep Learning Challenges

Example from GQA (Hudson, and Manning, CVPR-2019)

Visual Question Answering

Question 1: Is the red bike to the right or to the left of the Asian people?

GroundTruth: right

Prediction (a SOTA deep model in 2019): right (correct)

Learning Reasoning
Heterogeneous

Deep Learning Challenges

Example from GQA (Hudson, and Manning, CVPR-2019)

Visual Question Answering

Question 1: Is the red bike to the right or to the left of the Asian people?
GroundTruth: right
Prediction (a SOTA deep model in 2019): right (correct)

Question 2: Is the red bike to the left of the Asian people?
GroundTruth: no
Prediction (a SOTA deep model in 2019): yes (wrong)

Learning Reasoning
Heterogeneous

Deep Learning Challenges

Example from GQA (Hudson, and Manning, CVPR-2019)

Visual Question Answering

Does the model understand?

Objects? Attributes? Relations?

Question 1: Is the red bike to the right or to the left of the Asian people?
GroundTruth: right
Prediction (a SOTA deep model in 2019): right (correct)

Question 2: Is the red bike to the left of the Asian people?
GroundTruth: no
Prediction (a SOTA deep model in 2019): yes (wrong)

Deep Learning Challenges

Example from GQA (Hudson, and Manning, CVPR-2019)

Visual Question Answering

Does the model understand?

Objects? Attributes? Relations?

Old times…

Question 1: Is the red bike to the right or to the left of the Asian people?
GroundTruth: right
Prediction (a SOTA deep model in 2019): right (correct)

Question 2: Is the red bike to the left of the Asian people?
GroundTruth: no
Prediction (a SOTA deep model in 2019): yes (wrong)

Deep Learning Challenges

Example from GQA (Hudson, and Manning, CVPR-2019)

Visual Question Answering

Does the model understand?

Objects? Attributes? Relations?

- Larger VL models?
- Increase data?
- Augment Examples?
- ...

Old times...

Question 1: Is the red bike to the right or to the left of the Asian people?
GroundTruth: right
Prediction (a SOTA deep model in 2019): right (correct)

Question 2: Is the red bike to the left of the Asian people?
GroundTruth: no
Prediction (a SOTA deep model in 2019): yes (wrong)

Learning Reasoning
Heterogeneous

Deep Learning Challenges

A woman with an apple in her left hand standing to the right of a car.

DALL·E

DALL-E is a multimodal implementation of GPT-3 with 12 billion parameters.
(Example from April 2023)

Learning Reasoning Heterogeneous

Deep Learning Challenges

A woman with an apple in her left hand standing to the right of a car.

DALL·E

DALL-E is a multimodal implementation of GPT-3 with 12 billion parameters.
(Example from April 2023)

Still old?

Learning Reasoning
Heterogeneous

Key Point?

Learning Reasoning

Heterogeneous

Key Point?

+ These models are improving as the observations increase and models enlarge.

Learning Reasoning
Heterogeneous

Key Point?

+ These models are improving as the observations increase and models enlarge.

– Mostly, Neural Solutions suffer from:
  - Lack of Generalizability in Basic Reasoning Skills
  - Lack of Interpretability
  - Need for huge amounts of data

Learning Reasoning
Heterogeneous

Key Point?

+ These models are improving as the observations increase and models enlarge.

– Mostly, Neural Solutions suffer from:
  - Lack of Generalizability in Basic Reasoning Skills
  - Lack of Interpretability
  - Need for huge amounts of data

Problematic in many applications of natural language processing and computer vision for which these models have been considered, *in some sense*, very successful.

Contributions of HLR team

Contributions of HLR team

- **Evaluation and benchmarking** of Complex/Compositional Reasoning

Learning Reasoning
Heterogeneous

Contributions of HLR team

- **Evaluation and benchmarking** of Complex/Compositional Reasoning
- **Integration of explicit reasoning** in neural architectures

Contributions of HLR team

- **Evaluation and benchmarking** of Complex/Compositional Reasoning
- **Integration of explicit reasoning** in neural architectures
- **Developing software tools** to facilitate integration of symbolic knowledge (DomiKnowS)

Contributions of HLR team

- **Evaluation and benchmarking** of Complex/Compositional Reasoning
- **Integration of explicit reasoning** in neural architectures
- **Developing software tools** to facilitate integration of symbolic knowledge (DomiKnowS)
- **Grounding language** into visual modality, instruction following and spatial understanding

Learning Reasoning
Heterogeneous

Contributions of HLR team

- **Evaluation and benchmarking** of Complex/Compositional Reasoning
- **Integration of explicit reasoning** in neural architectures
- **Developing software tools** to facilitate integration of symbolic knowledge (DomiKnowS)
- **Grounding language** into visual modality, instruction following and spatial understanding
- **Application problems:** *Information extraction by spatial/temporal reasoning over procedural text, Spatial QA, Causal QA, Long-document QA, Commonsense QA, Visual QA, KB-Visual QA, Probabilistic QA, VLN-navigation, Grounded concept learning.*

Learning Reasoning
Heterogeneous

Outline

- **Evaluation and Integration of explicit Spatial Reasoning in LLMs**
- Evaluation and Integration of explicit Probabilistic Reasoning in LLMs
- DomiKnowS framework for Integration of symbolic reasoning in Neural Models

Learning Reasoning
Heterogeneous

SpartQA, SpaRTUN, ReSQ

- Spatial Question Answering
- Spatial Role Labeling
    - Spatial concepts (landmarks, trajectors, …)
    - Spatial relationships (topological relation, directional relations, …)
- Chain of reasoning

Learning Reasoning Heterogeneous

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

SpartQA, SpaRTUN, ReSQ

- Spatial Question Answering
- Spatial Role Labeling
  - Spatial concepts (landmarks, trajectors, …)
  - Spatial relationships (topological relation, directional relations, …)
- Chain of reasoning

Three boxes called one, two and three exist in an image. Box one contains a big yellow melon and a small orange watermelon. **Box two has a small yellow apple**. A small orange apple is inside and touching this box. Box one is in box three. **Box two** is to the **south** of, **far from** and to the **west** of **box three**. A **small yellow watermelon** is **inside box three**.

Q: Is **the small yellow apple** to the **west** of the **small yellow watermelon**? Yes

Q: Where is **box two** relative to the **small orange watermelon**? Left, Below, Far

(a) SPARTUN - A synthetic large dataset provided as source of supervision

A **grey car** is parking **in front of** a **grey house with brown window frames** and **plants on** the **balcony**.

Q: Are the **plants in front of** the **car**? No

Q: Are the **plants in** the **house**? Yes

(b) RESQ - A human-generated dataset for probing the models on realistic spatial problems

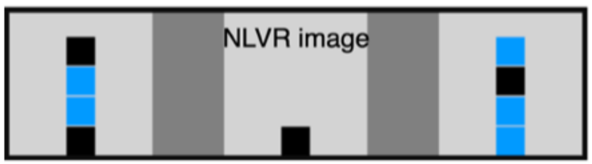
Learning Reasoning Heterogeneous

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

SpartQA, SpaRTUN, ReSQ

- Spatial Question Answering
- Spatial Role Labeling
  - Spatial concepts (landmarks, trajectors, …)
  - Spatial relationships (topological relation, directional relations, …)
- Chain of reasoning



Three boxes called one, two and three exist in an image. Box one contains a big yellow melon and a small orange watermelon. **Box two has a small yellow apple**. A small orange apple is inside and touching this box. Box one is in box three. **Box two** is to the **south** of, **far from** and to the **west** of **box three**. A **small yellow watermelon** is **inside box three**.

Q: Is **the small yellow apple** to the **west** of the **small yellow watermelon**? Yes

Q: Where is **box two** relative to the **small orange watermelon**? Left, Below, Far

(a) SpaRTUN - A synthetic large dataset provided as source of supervision

A **grey car** is parking in **front** of a **grey house with brown window frames** and **plants on** the **balcony**.

Q: Are the **plants in front of** the **car**? No

Q: Are the **plants in** the **house**? Yes

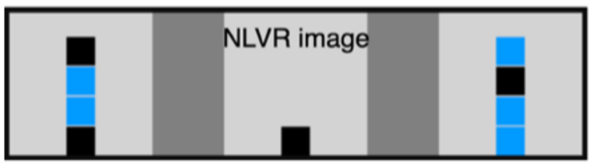
(b) ReSQ - A human-generated dataset for probing the models on realistic spatial problems

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

SpartQA, SpaRTUN, ReSQ

- Spatial Question Answering
- Spatial Role Labeling
  - Spatial concepts (landmarks, trajectors, …)
  - Spatial relationships (topological relation, directional relations, …)
- Chain of reasoning



Three boxes called one, two and three exist in an image. Box one contains a big yellow melon and a small orange watermelon. **Box two has a small yellow apple**. A small orange apple is inside and touching this box. Box one is in box three. **Box two** is to the **south** of, **far from** and to the **west** of **box three**. A **small yellow watermelon** is **inside box three**.

Q: Is **the small yellow apple** to the **west** of the **small yellow watermelon**? Yes

Q: Where is **box two** relative to the **small orange watermelon**? Left, Below, Far

(a) SPARTUN - A synthetic large dataset provided as source of supervision

A **grey car** is parking in **front** of a **grey house with brown window frames** and **plants on** the **balcony**.

Q: Are the **plants in front of** the **car**? No

Q: Are the **plants in** the **house**? Yes

(b) RESQ - A human-generated dataset for probing the models on realistic spatial problems

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

Spatial Reasoning over Text

SpartQA

#	Models	FB			FR			CO			YN		
		Seen	Unseen	Human*	Seen	Unseen	Human*	Seen	Unseen	Human*	Seen	Unseen	Human*
1	Majority	48.70	48.70	28.84	40.81	40.81	24.52	20.59	20.38	40.18	49.94	49.91	**53.60**
2	BERT	87.13	69.38	62.5	85.68	73.71	46.66	71.44	61.09	32.71	78.29	76.81	47.42
3	ALBERT	97.66	83.53	56.73	91.61	83.70	44.76	95.20	84.55	49.53	79.38	75.05	41.75
4	XLNet	**98.00**	**84.85**	**73.07**	**94.60**	**91.63**	**57.14**	**97.11**	**90.88**	**50.46**	**79.91**	**78.54**	39.69
5	Human	85		91.66	90		95.23	94.44		91.66	90		90.69

- Find Blocks (FB), Find Relations (FR), Choose Objects (CO), Yes/No (YN)

Concluding message

Learning Reasoning Heterogeneous

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

Spatial Reasoning over Text

SpartQA

#	Models	FB			FR			CO			YN		
		Seen	Unseen	Human*	Seen	Unseen	Human*	Seen	Unseen	Human*	Seen	Unseen	Human*
1	Majority	48.70	48.70	28.84	40.81	40.81	24.52	20.59	20.38	40.18	49.94	49.91	**53.60**
2	BERT	87.13	69.38	62.5	85.68	73.71	46.66	71.44	61.09	32.71	78.29	76.81	47.42
3	ALBERT	97.66	83.53	56.73	91.61	83.70	44.76	95.20	84.55	49.53	79.38	75.05	41.75
4	XLNet	**98.00**	**84.85**	**73.07**	**94.60**	**91.63**	**57.14**	**97.11**	**90.88**	**50.46**	**79.91**	**78.54**	39.69
5	Human	85		91.66	90		95.23	94.44		91.66	90		90.69

- Find Blocks (FB), Find Relations (FR), Choose Objects (CO), Yes/No (YN)

Concluding message

-Fine-tuning with SpartQA improved tests on different domains such as BoolQA, bAbI.

-Fine-tuning with the new extension SPARTUN, improved results on complex reasoning: StepGame.

-Fine-tuning with the new extension SPARTUN, improved results on realistic domain ResQ.

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

Learning Reasoning
Heterogeneous

Spatial Reasoning over Text

SpartQA

#	Models	FB			FR			CO			YN		
		Seen	Unseen	Human*	Seen	Unseen	Human*	Seen	Unseen	Human*	Seen	Unseen	Human*
1	Majority	48.70	48.70	28.84	40.81	40.81	24.52	20.59	20.38	40.18	49.94	49.91	**53.60**
2	BERT	87.13	69.38	62.5	85.68	73.71	46.66	71.44	61.09	32.71	78.29	76.81	47.42
3	ALBERT	97.66	83.53	56.73	91.61	83.70	44.76	95.20	84.55	49.53	79.38	75.05	41.75
4	XLNet	**98.00**	**84.85**	**73.07**	**94.60**	**91.63**	**57.14**	**97.11**	**90.88**	**50.46**	**79.91**	**78.54**	39.69
5	Human	85		91.66	90		95.23	94.44		91.66	90		90.69

- Find Blocks (FB), Find Relations (FR), Choose Objects (CO), Yes/No (YN)

Concluding message

-Fine-tuning with SpartQA improved tests on different domains such as BoolQA, bAbI.

-Fine-tuning with the new extension SPARTUN, improved results on complex reasoning: StepGame.

-Fine-tuning with the new extension SPARTUN, improved results on realistic domain ResQ.

SpartQA and SPARTUN datasets help transferring spatial reasoning knowledge.

Learning Reasoning
Heterogeneous

[**R. Mirzaee**, P. Kordjamshidi. **Transfer Learning with Synthetic Corpora for Spatial Role Labeling and Reasoning,** EMNLP-2022]

[**R. Mirzaee,** H. Faghihi, Q. Ning and P. Kordjamshidi, ***SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning***, NAACL-2021]

Integration of Spatial Logic in Training

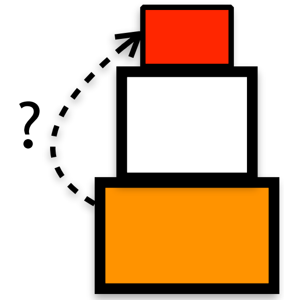
Context:
The white rectangle is on an orange rectangle. There is also the red rectangle, which is above the white rectangle.
Question:
Is the orange object below the red rectangle?

Initial facts: (red-r, white-r, above), (white-r, orange-r, above)

Rules:

- Converse
  (a, b, right) => (b, a, left)
- Symmetric
  (b, c, near) => (c, b, near)
- Transitive
  (x, b, above), (b, c, above) => (x, c, above)
  -Transitive + topological
  (x, y, inside), (h, z, inside), (y, z, front) => (x, z, front)

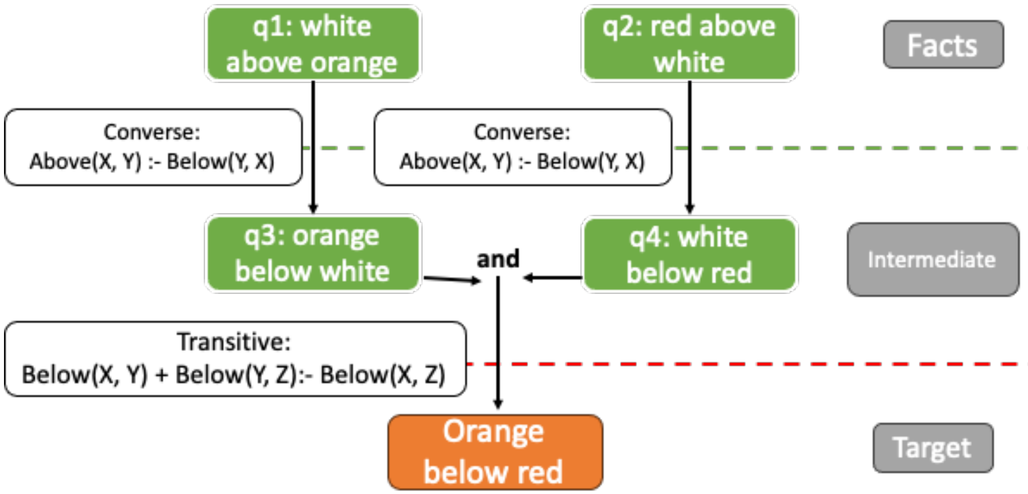
Target Query: (orange-r, red-r, below)?



[**Tanawan Premsri,** P. Kordjamshidi, **Transferring Spatial Reasoning Knowledge by Neuro-symbolic Training, 2024,** under review.]

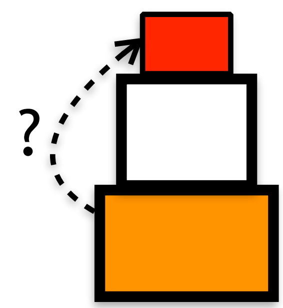
Learning Reasoning
Heterogeneous

Integration of Spatial Logic in Training

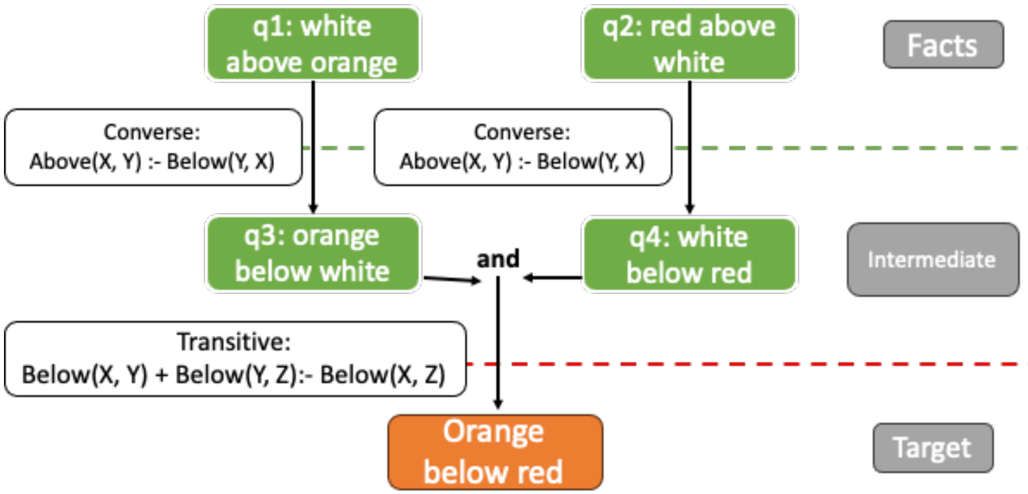


- Converse
  (a, b, right) => (b, a, left)
- Symmetric
  (b, c, near) => (c, b, near)
- Transitive
  (x, b, above), (b, c, above) => (x, c, above)

  -Transitive + topological
  (x, y, inside), (h, z, inside), (y, z, front) => (x, z, front)



Integration of Spatial Logic in Training



- Converse
  (a, b, right) => (b, a, left)
- Symmetric
  (b, c, near) => (c, b, near)
- Transitive
  (x, b, above), (b, c, above) => (x, c, above)

  -Transitive + topological

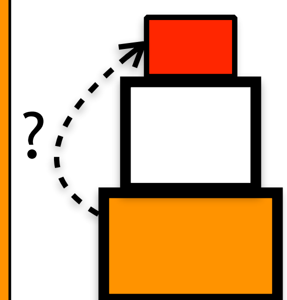
  (x, y, inside), (h, z, inside), (y, z, front) => (x, z, front)

- **Constraints for this example:**

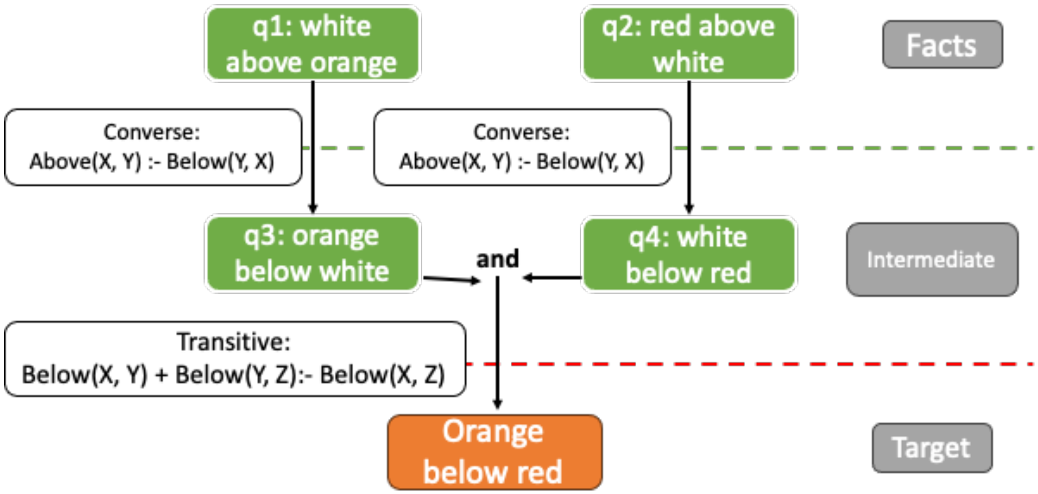
  C1: (q1=> q3)

  C2: (q2 => q4)

  C3: (q3 and q4 => target)



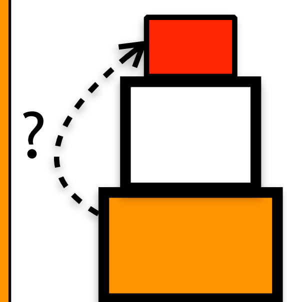
Integration of Spatial Logic in Training



The intermediate annotations are not needed for inference, the entailment rules are translated to consistency constraints for every specific situation. The constraints are used only during training.

- Converse
  (a, b, right) => (b, a, left)
- Symmetric
  (b, c, near) => (c, b, near)
- Transitive
  (x, b, above), (b, c, above) => (x, c, above)
  -Transitive + topological
  (x, y, inside), (h, z, inside), (y, z, front) => (x, z, front)

- **Constraints for this example:**
  C1: (q1=> q3)
  C2: (q2 => q4)
  C3: (q3 and q4 => target)



Integration of Spatial Logic in Training

Training

$$Loss = TaskLoss + \sum_{i=1}^{m} \lambda_j * C_i$$

Constraint Loss

- This training objective can be used for tuning both generative and encoder-based language models when adapted to the classification task.

Learning Reasoning
Heterogeneous

Integration of Spatial Logic in Training

Overall performance on **Realistic ResQ** dataset (All). Performance per reasoning steps (k) required for the question to be answered. Null indicates the cases in which k was not clear.

Model	ResQ			
	k=1	k=2	null	All
BERT	70.67	56.85	60.66	60.98
BERT-T	76.00	54.79	61.18	61.15
BERT-T+Q-Chain	72.00	**58.90**	59.90	61.31
T5	74.67	56.16	61.44	61.80
T5-T	81.33	54.79	61.44	62.30
T5-T+Q-Chain	**81.33**	57.53	**63.75**	**64.43**
GPT-3 (zero-shot)	74.67	60.95	66.58	66.22
GPT-3 (few-shot)	84.00	68.49	68.12	70.16
GPT-3 (COT)	**86.67**	67.12	68.64	70.49
GPT-4 (zero-shot)	84.00	**73.97**	**76.86**	**77.05**

Learning Reasoning
Heterogeneous

[**Tanawan Premsri,** P. Kordjamshidi, **Transferring Spatial Reasoning Knowledge by Neuro-symbolic Training, 2024, in progress.**]

Integration of Spatial Logic in Training

A **grey car** is parking in **front** of a **grey house with brown window frames** and **plants on** the **balcony**.

Q: Are the **plants in front of** the **car**? No ; ChatGPT: No

Q: Are the **plants in** the **house**? Yes; ChatGPT: No

Overall performance on **Realistic ResQ** dataset (All). Performance per reasoning steps (k) required for the question to be answered. Null indicates the cases in which k was not clear.

Model	ResQ			
	k=1	k=2	null	All
BERT	70.67	56.85	60.66	60.98
BERT-T	76.00	54.79	61.18	61.15
BERT-T+Q-Chain	72.00	**58.90**	59.90	61.31
T5	74.67	56.16	61.44	61.80
T5-T	81.33	54.79	61.44	62.30
T5-T+Q-Chain	**81.33**	57.53	**63.75**	**64.43**
GPT-3 (zero-shot)	74.67	60.95	66.58	66.22
GPT-3 (few-shot)	84.00	68.49	68.12	70.16
GPT-3 (COT)	**86.67**	67.12	68.64	70.49
GPT-4 (zero-shot)	84.00	**73.97**	**76.86**	**77.05**

[**Tanawan Premsri,** P. Kordjamshidi, **Transferring Spatial Reasoning Knowledge by Neuro-symbolic Training, 2024, in progress.**]

Integration of Spatial Logic in Training

A **grey car** is parking in **front** of a **grey house with brown window frames** and **plants on** the **balcony**.

Q: Are the **plants in front of** the **car**? No ; ChatGPT: No

Q: Are the **plants in** the **house**? Yes; ChatGPT: No

Model	ResQ			
	k=1	k=2	null	All
BERT	70.67	56.85	60.66	60.98
BERT-T	76.00	54.79	61.18	61.15
BERT-T+Q-Chain	72.00	**58.90**	59.90	61.31
T5	74.67	56.16	61.44	61.80
T5-T	81.33	54.79	61.44	62.30
T5-T+Q-Chain	**81.33**	57.53	**63.75**	**64.43**
GPT-3 (zero-shot)	74.67	60.95	66.58	66.22
GPT-3 (few-shot)	84.00	68.49	68.12	70.16
GPT-3 (COT)	**86.67**	67.12	68.64	70.49
GPT-4 (zero-shot)	84.00	**73.97**	**76.86**	**77.05**

Overall performance on **Realistic ResQ** dataset (All). Performance per reasoning steps (k) required for the question to be answered. Null indicates the cases in which k was not clear.

Model	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
BERT	98.51	95.53	91.68	66.71	49.11	41.47	41.47	32.09	28.94	28.16
BERT-T	98.50	95.32	93.26	76.78	66.36	58.76	53.70	46.27	42.71	40.12
BERT-T+Q-Chain	98.70	96.45	93.03	74.58	64.95	59.04	54.38	49.23	45.36	44.05
T5	98.51	96.97	89.71	78.47	71.11	65.49	61.18	55.89	52.20	50.34
GPT3 (few-shot)	55.00	37.00	25.00	30.00	32.00	29.00	21.00	22.00	34.00	31.00
GPT3 (COT)	61.00	45.00	30.00	35.00	35.00	27.00	22.00	24.00	23.00	25.00

Accuracy of various models on STEPGAME including result of GPT3 model reported in

Zhun Y., et.al,. 2023. Coupling large language models with logic programming for robust and general reasoning from text.

Learning Reasoning Heterogeneous

[**Tanawan Premsri,** P. Kordjamshidi, **Transferring Spatial Reasoning Knowledge by Neuro-symbolic Training, 2024, in progress.**]

Spatial QA

Findings

- Pre-trained Large Language models comparatively weak in spatial reasoning
- We created benchmarks to evaluate them with spatial QA
- Synthetic data can help transfer learning to realistic domain.
- Logic-based training helped complex multi-hop reasoning
- GPT models are better in commonsense reasoning.

Outline

- Evaluation and Integration of explicit Spatial Reasoning in LLMs
- **Evaluation and Integration of explicit Probabilistic Reasoning in LLMs**
- Grounding language in visual perception and Navigation
- DomiKnowS framework for Integration of symbolic reasoning in Neural Models

Learning Reasoning Heterogeneous

Reasoning over Uncertain Text

Logical Reasoning	Probabilistic Logical Reasoning
Dave is big. If someone is big, then they are green. If someone is green, then they are round.	Dave is big. Usually, If someone is big, then they are green. Normally, If someone is green, then they are round.
Conclusion: David is round.	Conclusion: David is round with a probability of 72%.

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Teaching Probabilistic Logical Reasoning to Transformers,** EACL-2024 Findings**]**
[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi**, Probabilistic Reasoning in Generative Large Language Models,** under review, arxived**]**

Learning Reasoning
Heterogeneous

Reasoning over Uncertain Text

Logical Reasoning	Probabilistic Logical Reasoning
Dave is big. If someone is big, then they are green. If someone is green, then they are round.	Dave is big. Usually, If someone is big, then they are green. Normally, If someone is green, then they are round.
Conclusion: David is round.	Conclusion: David is round with a probability of 72%.

Motivation: Scientific text specially medical domain:
Combining uncertain evidences and applying uncertain rules.

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Teaching Probabilistic Logical Reasoning to Transformers,** EACL-2024 Findings**]**
[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi, **Probabilistic Reasoning in Generative Large Language Models,** under review, arxived**]**

Learning Reasoning
Heterogeneous

Reasoning over Uncertain Text

Logical Reasoning	Probabilistic Logical Reasoning
Dave is big. If someone is big, then they are green. If someone is green, then they are round.	Dave is big. Usually, If someone is big, then they are green. Normally, If someone is green, then they are round.
Conclusion: David is round.	Conclusion: David is round with a probability of 72%.

Motivation: Scientific text specially medical domain:
Combining uncertain evidences and applying uncertain rules.

People with diabetes often have many of the same risk factors associated with heart disease …

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Teaching Probabilistic Logical Reasoning to Transformers,** EACL-2024 Findings**]**
[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi**, Probabilistic Reasoning in Generative Large Language Models,** under review, arxived**]**

Learning Reasoning
Heterogeneous

Reasoning over Uncertain Text

new

RuleBERT	RuleTaker-pro
(Fact 1) David is a cousin of Ann. (Fact 2) Mike is a child of Ann. (Rule 1, 0.90) If A is a spouse of B and C is a child of B, then C is a child of A. (Rule 2, 0.15) If A is a cousin of B, then A is a spouse of B.	(Fact 1) Dave is big. (Fact 2) Erin is sad. (Rule 1) Usually, If someone is big then they are green. (Rule 2) Normally, If someone is green then they are round. (Rule 3) Seldom, If someone is sad then they are round.
(Query) Mike is a child of David.	(Query) Dave is round.
Required Steps of Reasoning to Answer	
Fact 1 (1.00) & Rule 2 (0.15) $\Longrightarrow$ Fact 3: David is a spouse of Ann. (0.15) (Inferred) Fact 3 (0.15) & Fact 2 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 4: Mike is a child of David. (0.135) (Inferred) **Answer: 0.135**	Fact 1 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 3: Dave is green. (0.90) (Inferred) Fact 3 (0.90) & Rule 2 (0.80) $\Longrightarrow$ Fact 4: Dave is round. (0.72) (Inferred) **Answer: 0.72**

Learning Reasoning
Heterogeneous

Reasoning over Uncertain Text

new

RuleBERT	RuleTaker-pro
(Fact 1) David is a cousin of Ann. (Fact 2) Mike is a child of Ann. (Rule 1, 0.90) If A is a spouse of B and C is a child of B, then C is a child of A. (Rule 2, 0.15) If A is a cousin of B, then A is a spouse of B.	(Fact 1) Dave is big. (Fact 2) Erin is sad. (Rule 1) Usually, If someone is big then they are green. (Rule 2) Normally, If someone is green then they are round. (Rule 3) Seldom, If someone is sad then they are round.
(Query) Mike is a child of David.	(Query) Dave is round.
Required Steps of Reasoning to Answer	
Fact 1 (1.00) & Rule 2 (0.15) $\Longrightarrow$ Fact 3: David is a spouse of Ann. (0.15) (Inferred) Fact 3 (0.15) & Fact 2 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 4: Mike is a child of David. (0.135) (Inferred) **Answer: 0.135**	Fact 1 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 3: Dave is green. (0.90) (Inferred) Fact 3 (0.90) & Rule 2 (0.80) $\Longrightarrow$ Fact 4: Dave is round. (0.72) (Inferred) **Answer: 0.72**

Approach: Converting Probabilistic Reasoning Steps to Equality Constraints	
Constraint 1: P(Fact 1) * 0.15 = P(Fact 3) Constraint 2: P(Fact 3) * P(Fact 2) * 0.90 = P(Fact 4)	Constraint 1: P(Fact 1) * 0.90 = P(Fact 3) Constraint 2: P(Fact 3) * 0.80 = P(Fact 4)

Learning Reasoning
Heterogeneous

Integration of Probabilistic Reasoning in Training

Training

$$Loss = TaskLoss + \sum_{i=1}^{m} \lambda_j * C_i$$

Approach: Converting Probabilistic Reasoning Steps to Equality Constraints	
Constraint 1: P(Fact 1) * 0.15 = P(Fact 3) Constraint 2: P(Fact 3) * P(Fact 2) * 0.90 = P(Fact 4)	Constraint 1: P(Fact 1) * 0.90 = P(Fact 3) Constraint 2: P(Fact 3) * 0.80 = P(Fact 4)

Examples of constrains

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Teaching Probabilistic Logical Reasoning to Transformers,** EACL-2024 Findings, https://arxiv.org/pdf/2305.13179.pdf]
[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi, **Probabilistic Reasoning in Generative Large Language Models,** https://arxiv.org/abs/2402.09614**]**

Integration of Probabilistic Reasoning in Training

Training

$$Loss = TaskLoss + \sum_{i=1}^{m} \lambda_j * C_i$$

Approach: Converting Probabilistic Reasoning Steps to Equality Constraints	
Constraint 1: P(Fact 1) * 0.15 = P(Fact 3) Constraint 2: P(Fact 3) * P(Fact 2) * 0.90 = P(Fact 4)	Constraint 1: P(Fact 1) * 0.90 = P(Fact 3) Constraint 2: P(Fact 3) * 0.80 = P(Fact 4)

Examples of constrains

- This training approach improves the PLMs, in particular for deeper reasoning networks.
- It can better transfer probabilistic reasoning from one domain to another.

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Teaching Probabilistic Logical Reasoning to Transformers,** EACL-2024 Findings, https://arxiv.org/pdf/2305.13179.pdf]
[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Probabilistic Reasoning in Generative Large Language Models,** https://arxiv.org/abs/2402.09614]

Learning Reasoning
Heterogeneous

Integration of Probabilistic Reasoning in Training

- Using probabilistic constrains during training helps in more complex and deeper reasoning steps.
- This result in on newly created Rule-take-pro with example-specific Rules.

	RoBERTa			
D/M	M1	M2	M3	Mmax
Total	38.2	38.3	20.4	33.8
D1	56.0	52.7	29.6	43.7
D2	36.4	38.2	20.3	32.8
D3	29.3	31.3	14.9	28.3
D4	27.4	28.5	14.0	27.1
D5	24.9	26.7	14.7	28.2
CS1	47.8	35.7	16.2	20.7
	RoBERTa + PCT			
Total	38.0	39.5	41.1	37.6
D1	53.3	50.8	50.5	46.9
D2	37.4	40.4	42.2	37.0
D3	26.4	32.9	36.0	32.4
D4	26.5	31.9	33.9	31.8
D5	23.3	30.4	33.4	31.4
CS1	44.9	42.6	34.5	35.2

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi, **Teaching Probabilistic Logical Reasoning to Transformers,** EACL-2024 Findings]

Integration of Probabilistic Reasoning in Inference

Context: Green event is True with probability of 80%. If Green event is True, then Pink event is True with probability of 10%.If Green event is True, then Pink event is False with probability of 90%. If Green event is False, then Pink event is True with probability of 20%. If Green event is False, then Pink event is False with probability of 80%.

Query: What is the probability that Green event is True and Pink event is True?

Learning Reasoning Heterogeneous

[**Aliakbar Nafar,** Kristen Brent Venable, Parisa Kordjamshidi**, Probabilistic Reasoning in Generative Large Language Models, 2024,** arxiv]

Integration of Probabilistic Reasoning in Inference

Context: Green event is True with probability of 80%. If Green event is True, then Pink event is True with probability of 10%.If Green event is True, then Pink event is False with probability of 90%. If Green event is False, then Pink event is True with probability of 20%. If Green event is False, then Pink event is False with probability of 80%.

Query: What is the probability that Green event is True and Pink event is True?

BQA

BQA Instruction: Solve the following probabilistic question and generate the probability of the answer by only providing a number from 0 to 100.

Context | Query

Answer: 8%

COT

COT Instruction: Solve the following probabilistic question and provide a detailed explanation of the mathematical reasoning. At the end, write a sentence that will give the final answer rounded to an integer.

Context | Query

Answer:
This probability can be found using the conditional probability:
P(g, p)=P(g)*P(p|g) =
0.80 * 0.10 = 0.08
The final answer is **8%**.

Learning Reasoning
Heterogeneous

[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi, **Probabilistic Reasoning in Generative Large Language Models, 2024,** arxiv**]**

Integration of Probabilistic Reasoning in Inference

Context: Green event is True with probability of 80%. If Green event is True, then Pink event is True with probability of 10%.If Green event is True, then Pink event is False with probability of 90%. If Green event is False, then Pink event is True with probability of 20%. If Green event is False, then Pink event is False with probability of 80%.

Query: What is the probability that Green event is True and Pink event is True?

BQA

BQA Instruction: Solve the following probabilistic question and generate the probability of the answer by only providing a number from 0 to 100.

Context | Query

Answer: 8%

COT

COT Instruction: Solve the following probabilistic question and provide a detailed explanation of the mathematical reasoning. At the end, write a sentence that will give the final answer rounded to an integer.

Context | Query

Answer:
This probability can be found using the conditional probability:
P(g, p)=P(g)*P(p|g) =
0.80 * 0.10 = 0.08
The final answer is **8%**.

ProbLog

ProbLog Instruction: Solve the following probabilistic question by writing a Problog code that represents the probabilities, and the query.

Context | Query

Answer:

```
0.8::green.
0.2::pink:-\+green.
0.1::pink:- green.

q1:- green, pink.
```

Learning | Reasoning
Heterogeneous

[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi, **Probabilistic Reasoning in Generative Large Language Models, 2024,** arxiv**]**

Integration of Probabilistic Reasoning in Inference

Context: Green event is True with probability of 80%. If Green event is True, then Pink event is True with probability of 10%.If Green event is True, then Pink event is False with probability of 90%. If Green event is False, then Pink event is True with probability of 20%. If Green event is False, then Pink event is False with probability of 80%.

Query: What is the probability that Green event is True and Pink event is True?

BQA

BQA Instruction: Solve the following probabilistic question and generate the probability of the answer by only providing a number from 0 to 100.

Context | Query

Answer: 8%

COT

COT Instruction: Solve the following probabilistic question and provide a detailed explanation of the mathematical reasoning. At the end, write a sentence that will give the final answer rounded to an integer.

Context | Query

Answer:
This probability can be found using the conditional probability:
P(g, p)=P(g)*P(p|g) =
0.80 * 0.10 = 0.08
The final answer is **8%**.

ProbLog

ProbLog Instruction: Solve the following probabilistic question by writing a Problog code that represents the probabilities, and the query.

Context | Query

Answer:

```
0.8::green.
0.2::pink:-\+green.
0.1::pink:- green.

q1:- green, pink.
```

- Mapping to code is always better here and the best results obtained from mapping to **probabilistic logical programming** code in ProbLog language.

[Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi, **Probabilistic Reasoning in Generative Large Language Models, 2024,** arxiv**]**

Learning Reasoning
Heterogeneous

Outline

- **Evaluation and Integration of explicit Spatial Reasoning in LLMs**
- Evaluation and Integration of explicit Probabilistic Reasoning in LLMs
- DomiKnowS framework for Integration of symbolic reasoning in Neural Models

Motivation

- In complex intelligent systems:
  - There are many models: basic concepts, relations, types of relations, compositional concepts…
  - Decisions depend on models and their interactions, we need to reason based on models.

This kind of modeling seems to be challenging with current tools.

Typical show case: Entity Mention Relation Extraction

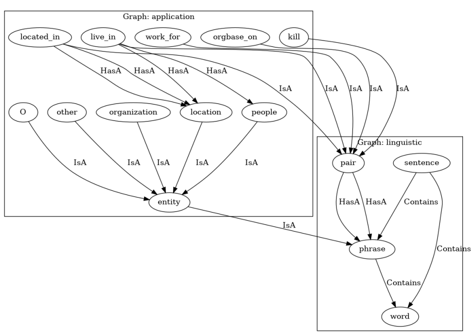
other	0.05
location	0.92
people	0.90

other	0.05
location	0.33
org	0.82

Inference helps to improve local predictions.

Models can be trained locally or globally using soft and hard constraints.

Washington covers Seattle for Associated Press.



irrelevant	0.10
lives_in	0.35
works_for	0.85

Relation	Entity 1	Entity 2
kill	people	people
works_for	people	org
live_in	people	location
orgbase_on	org	location
locate_in	location	location

Learning Reasoning
Heterogeneous

[Rajaby Faghihi, Guo, Uszok, Nafar, Raisi, Kordjamshidi, **DomiKnowS: A Library for Integration of Symbolic Domain Knowledge in Deep Learning,** Demo Track, EMNLP-2021]

[Guo, Rajaby Faghihi, Zhang, and Kordjamshidi. **Inference-masked loss for structured output learning**. IJCAI-2020]

Typical show case: Entity Mention Relation Extraction

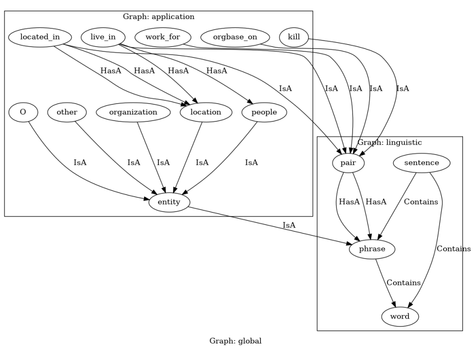
other	0.05
location	0.92
people	0.90

other	0.05
location	0.33
org	0.82

Washington covers Seattle for Associated Press.

Inference helps to improve local predictions.

Models can be trained locally or globally using soft and hard constraints.



irrelevant	0.10
lives_in	0.35
works_for	0.85

Relation	Entity 1	Entity 2
kill	people	people
works_for	people	org
live_in	people	location
orgbase_on	org	location
locate_in	location	location

Learning Reasoning
Heterogeneous

[Rajaby Faghihi, Guo, Uszok, Nafar, Raisi, Kordjamshidi, **DomiKnowS: A Library for Integration of Symbolic Domain Knowledge in Deep Learning,** Demo Track, EMNLP-2021]

[Guo, Rajaby Faghihi, Zhang, and Kordjamshidi. **Inference-masked loss for structured output learning**. IJCAI-2020]

Typical show case: Entity Mention Relation Extraction

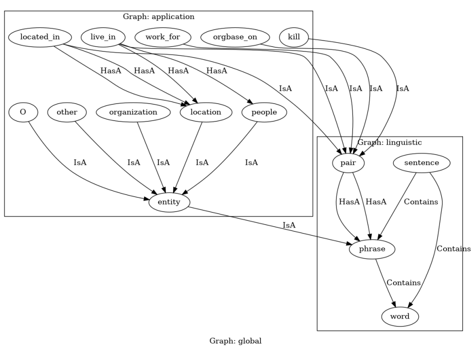
other	0.05
~~location~~	0.92
people	0.90

other	0.05
location	0.33
org	0.82

Inference helps to improve local predictions.

Models can be trained locally or globally using soft and hard constraints.

Washington covers Seattle for Associated Press.



irrelevant	0.10
lives_in	0.35
works_for	0.85

Relation	Entity 1	Entity 2
kill	people	people
works_for	people	org
live_in	people	location
orgbase_on	org	location
locate_in	location	location

Learning Reasoning
Heterogeneous

[Rajaby Faghihi, Guo, Uszok, Nafar, Raisi, Kordjamshidi, **DomiKnowS: A Library for Integration of Symbolic Domain Knowledge in Deep Learning,** Demo Track, EMNLP-2021]

[Guo, Rajaby Faghihi, Zhang, and Kordjamshidi. **Inference-masked loss for structured output learning**. IJCAI-2020]

Ideal Language Components for Declarative Modeling

- Data Modeling
- Knowledge Representation and Reasoning
- Learning from Data
  - Dealing with uncertainty in Data and Knowledge
  - Data Representations
- Designing Learning and Inference Paradigms

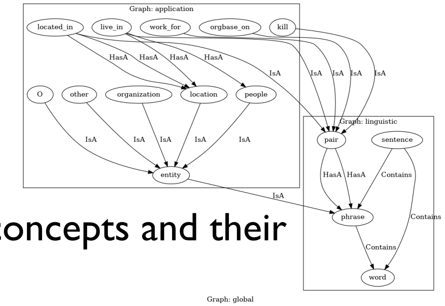
Finding the right abstraction and primitives for an easy integration of all of these components is a challenge!

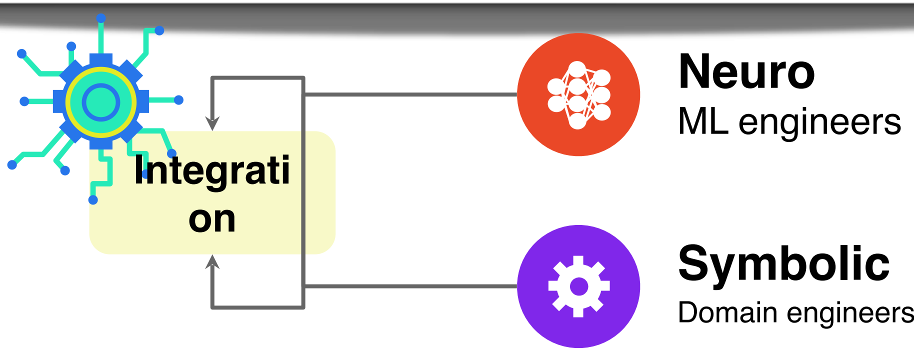
Learning Reasoning
Heterogeneous

[Kordjamshidi, Kersting and Roth, Declarative Learning Based Programming as an Interface to AI systems, Frontiers in Artificial Intelligence, 2022]

What is DomiKnowS?

- Current deep learning libraries ( ) are the backbone of DomiKnowS.
- The interface is a symbolic representation of concepts and relationships.
- Knowledge in the form of Logical expressions can be expressed over named concepts and their instantiations.
- Names/symbols are connected to deep learning modules.
- We can reason over symbols and integrate explicit reasoning over output structures or latent variables into Deep learning
- Knowledge is used in various types of reasoning and as hard or soft constraints on the output concepts of the deep models based on a variety of algorithms.

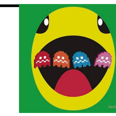




Learning Reasoning
Heterogeneous

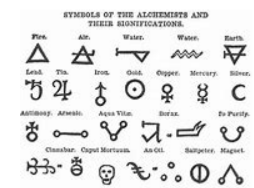
Why knowledge integration? Why symbols? Why DomiKnows?

- Deep Learning is data hungry and Labeled data is expensive.

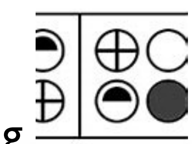


- Domain Knowledge (DK) can directly provide the level of abstraction needed for reasoning.

- DK potentially reduces the need for large amounts of data



- DK is often expressed symbolically & Symbolic reasoning is more interpretable



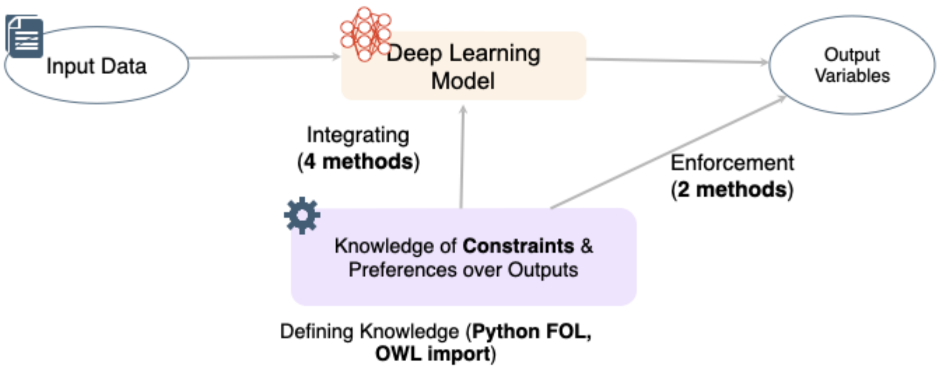
- Cover the gap between current deep learning tools and symbolic reasoning

Learning Reasoning
Heterogeneous

Program Components

- Declarative Problem Specification
  - Graph-based Data Modeling
  - Knowledge Representation
- Defining computational units
  - Wrapping feature generation modules
  - Wrapping deep learning architectures
  - Connecting the Symbolic and sub-symbolic layers
- Deep Constraint-Based Learning and Prediction
  - Various algorithms for applying soft and hard constrains during training
  - Using integer linear programming for inference during prediction [which will be extended further]

- Data Modeling
- Knowledge Representation and Reasoning
- Learning from Data
  - Dealing with uncertainty in Data and Knowledge
  - Data Representations
- Designing Learning and Inference Paradigm



Learning Reasoning
Heterogeneous

DomiKnowS Development

- Domain Declaration (Concepts, Relationships)

```
word = Concept(name='word')
phrase = Concept(name='phrase')
sentence = Concept(name='sentence')
sentence.contains(word)
sentence.contains(phrase)
phrase.contains(word)

pair = Concept(name='pair')
pair.has_a(arg1=word, arg2=word)

people = word(name='people')
organization = word(name='organization')
location = word(name='location')
other = word(name='other')
o = word(name='O')
```

DEMO

https://hlr.github.io/domiknows-nlp/

- Constraints Declaration

```
disjoint(people, organization, location, other, o)

nandL(people, organization)
```

+ Automatically compile OWL ontologies

Learning Reasoning Heterogeneous

[**DomiKnowS:** A library for integration of symbolic domain knowledge in deep learning. ***Hossein Rajaby Faghihi, Quan Guo, Andrzej Uszok, Darius Nafar, Elaheh Raisi, and Parisa Kordjamshidi.*** EMNLP-2021 Demo Track]

DomiKnowS Development

- ## Domain Declaration (Concepts, Relationships)

```
word = Concept(name='word')
phrase = Concept(name='phrase')
sentence = Concept(name='sentence')
sentence.contains(word)
sentence.contains(phrase)
phrase.contains(word)

pair = Concept(name='pair')
pair.has_a(arg1=word, arg2=word)

people = word(name='people')
organization = word(name='organization')
location = word(name='location')
other = word(name='other')
o = word(name='O')
```

```
paragraph = Concept(name='paragraph')
    question = Concept(name='question')
    para_quest_contains, = paragraph.contains(question)
…
question[is_more] = ModuleLearner("robert_emb",
module=RobertaClassificationHead(roberta_model.last_layer_size))
…
ifL(is_more, V(name='x'), is_less, V(name='y', v=('x',
symmetric, s_arg2.name)))
```

DEMO

https://hlr.github.io/domiknows-nlp/

- ## Constraints Declaration

```
disjoint(people, organization, location, other, o)

nandL(people, organization)
```

+ Automatically compile OWL ontologies

[**DomiKnowS:** A library for integration of symbolic domain knowledge in deep learning. ***Hossein Rajaby Faghihi, Quan Guo, Andrzej Uszok, Darius Nafar, Elaheh Raisi, and Parisa Kordjamshidi.*** EMNLP-2021 Demo Track]

Learning Reasoning
Heterogeneous

Underlying algorithms

- **Inference masked loss:** Using integer linear programming formulation of the inference to find the best output subject to constraints in the loop of training.
- **Sampling loss/Semantic loss:** sampling the outputs and reduce the probability of violations as a part of the loss based on the samples.
- **Dual formulation:** Forming a dual formulation of the objective and adding a differentiable surrogate of the constrains to the loss.
- **Prediction time Inference:** In all cases ILP can be used in the inference time to find the most probable output.

> If the constraint violation is a part of the loss, the model learns to respect the constraints, however, those can be still violated at the end. Using ILP at inference time ensures consistent results with domain knowledge. Constraints can be used for unsupervised learning too.

—Guo, Rajaby Faghihi, Zhang, and Kordjamshidi. Inference-masked loss for structured output learning. IJCAI-2020

—Hossein Rajaby Faghihi, Parisa Kordjamshidi, Consistent Joint Decision-Making with Heterogeneous Learning Models, The 2024 European conference in the field of computational linguistics (EACL-2024 findings)

—Yatin Nandwani, Abhishek Pathak, Mausam, and Parag Singla. 2019a. A primal dual formulation for deep learning with constraints. In NeurIPS.

—J. Xu, Z. Zhang, T. Friedman, Y. Liang, and G. Van den Broeck. 2018. A semantic loss function for deep learning with symbolic knowledge. Volume 80 of Proceedings of Machine Learning Research, pages 5502–5511. PMLR.

—Dan Roth and Wen-tau Yih, "A Linear Programming Formulation for Global Inference in Natural Language Tasks." CoNLL, (2004)

GLUECons

- The first collection of the benchmarks to evaluate constraint integration in Deep Models
- Providing a set of metrics for in-depth evaluation of constraint integration methods
- A set of baselines for such integration during both inference and training
- Releasing a boilerplate code and evaluation of all tasks to facilitate future research and comparisons

GLUECons: A Generic Benchmark for Learning Under Constraints, **Hossein Rajaby Faghihi, Aliakbar Nafar,** Chen Zheng, Roshanak Mirzaee, Yue Zhang, Andrzej Uszok, Alexander Wan, Tanawan Premsri, Dan Roth, and Parisa Kordjamshidi. The 37th Conference of Artificial Intelligence **(AAAI-2023)**

Selected Tasks

-Classification with label dependencies

- Mutual Exclusivity (MNIST)
- Hierarchical constraints (CIFAR 100)

-Self-Consistency in decisions

- What-If question answering (WIQA)
- Natural language inference (NLI)
- BeliefBank

-Consistency with external knowledge

- Entity and relation extraction (CONLL2003) (EMR)

-Constraints in (un/semi)supervised setting

- MNIST arithmetic
- Sudoku

-Structural Consistency

- BIO Tagging

Learning Reasoning
Heterogeneous

DomiKnowS’s Natural Language interface

- Explain the graph and constraints in natural language
- Interactively prompt GPT (using Langchain)
- Convert Natural language to declarative language!
- DomiKnowS release as an independent package

```
pip install DomiKnowS
```

https://hlr.github.io/domiknows/

Learning Reasoning

Heterogeneous

Prompt2DeModel: Declarative Neuro-Symbolic Modeling with Natural Language, **Hossein Rajaby Faghihi**, Aliakbar Nafar, Andrzej Uszok, Hamid Karimian, Parisa Kordjamshidi, 2024, under-review

Summary Results

- Evaluation of Reasoning Skills of Language Models
- Integration of Explicit Reasoning in form of Constraints in Training and Inference —spatial and uncertainty reasoning
- Building tools (DimiKnowS) and benchmarks (GLUECons) for integration of logic in learning.
- Natural language interface to declarative programming.

Acknowledgments and Thanks

Thanks to Andrzej Uszok from IHMC and all my students, post-docs at ***Heterogenous Learning and Reasoning (HLR)*** Lab who helped in developing DomiKnowS.

https://hlr.github.io.

kordajms@msu.edu

Michigan State University
Computer Science and Engineering Department

I am actively **seeking** motivated **PhD students and post-doctoral researchers.** my research focus within the fields of Natural Language Processing (NLP), Vision and Language (V&L) and Neuro-symbolic modeling,

Learning Reasoning
Heterogeneous

Acknowledgments and Thanks

Thanks to Andrzej Uszok from IHMC and all my students, post-docs at ***Heterogenous Learning and Reasoning (HLR)*** Lab who helped in developing DomiKnowS.

https://hlr.github.io.

kordajms@msu.edu

Thank you!

Michigan State University
Computer Science and Engineering Department

I am actively **seeking** motivated **PhD students and post-doctoral researchers.** my research focus within the fields of Natural Language Processing (NLP), Vision and Language (V&L) and Neuro-symbolic modeling,

Learning Reasoning
Heterogeneous