Selected Topics in Machine Learning Applied to NLP

Lili Mou

Dept. Computing Science, University of Alberta
Alberta Machine Intelligence Institute (Amii)

lmou@ualberta.ca

CMPUT466/566 Lab Session

Disclaimer

- The focus of CMPUT466/566 is ***NOT*** the mini-project
- The non-trivial part is optional. Even for grads, you may recover the marks by claiming the syllabus bonus.
- If you're mainly looking for a project, you are in a wrong course
  - The correct course: Russ's AI Capstone course
  - Individual Study (both undergrads and grads)
- The talk is a whirl-wind tour in selected topics
  - It gives an overview, not detailed derivations.

amii UNIVERSITY OF ALBERTA

Advertisements

- Lili Mou is accepting all-level students
  - URA, MSc, PhD, postdoc
  - A typical path for URAs/MSc

    (466 course project →) Individual Study 499 → RAship
- Previous achievements
  - CMPUT499 (F20) → URA (W+S21) → CIKM'21
  - CMPUT499 (W21) → URA (S21) → EMNLP'21 (Findings)

A fading memory …

- Of how I learned natural language processing (NLP):

$$NLP = \underbrace{NLU}_{\textbf{Understanding}} + \underbrace{NLG}_{\textbf{Generation}}$$

- NLU: text -> meaning
- NLG: meaning -> text

Selected Topics

- Natural Language Understanding
  - **Weakly supervised reasoning**
- Natural Language Generation
  - Search-based unsupervised text generation

Different Schools of NLU

- **Connectionism:** NLU is defined by tasks and achieved by neural networks
  - Relation extraction
  - Natural language inference
  - Machine comprehension
  - Commonsense QA

- **Symbolism:** NLU is defined by formal meaning representations
  - Semantic dependency parsing
  - Abstract meaning representations for text

amii UNIVERSITY OF ALBERTA

SQuaD

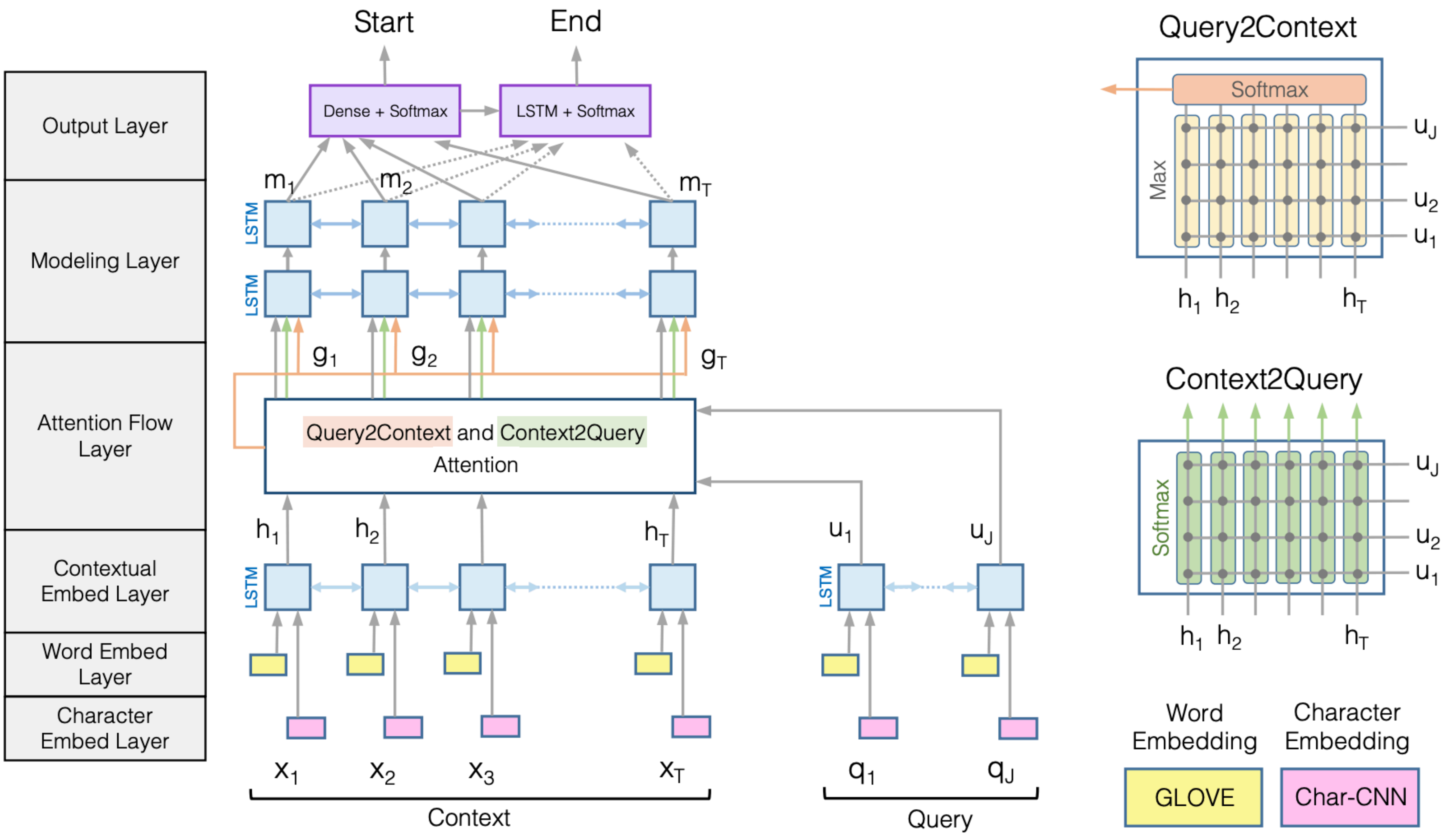


Figure 1: BiDirectional Attention Flow Model *(best viewed in color)* [Seo+ICLR'17]

Abstract Meaning Representation

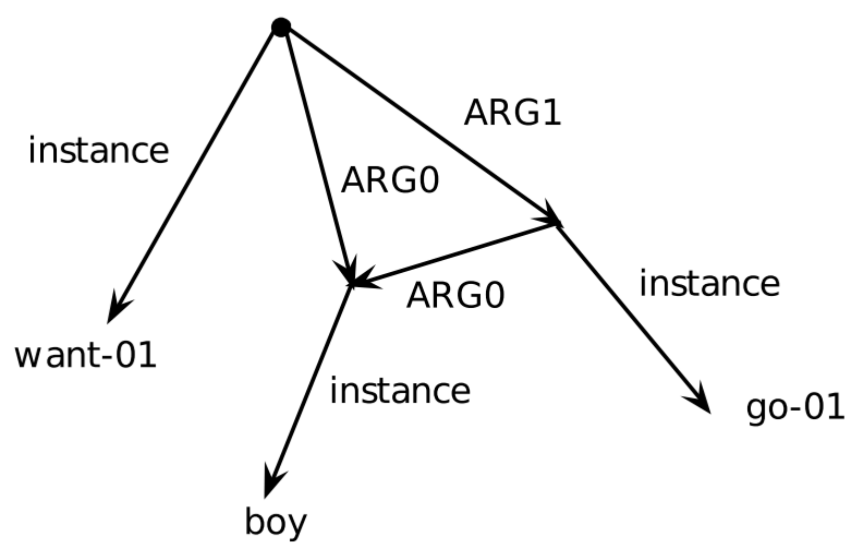
LOGIC format:

∃ w, b, g:
instance(w, want-01) ∧ instance(g, go-01) ∧
instance(b, boy) ∧ arg0(w, b) ∧
arg1(w, g) ∧ arg0(g, b)

AMR format (based on PENMAN):

```
(w / want-01
   :arg0 (b / boy)
   :arg1 (g / go-01
            :arg0 b))
```

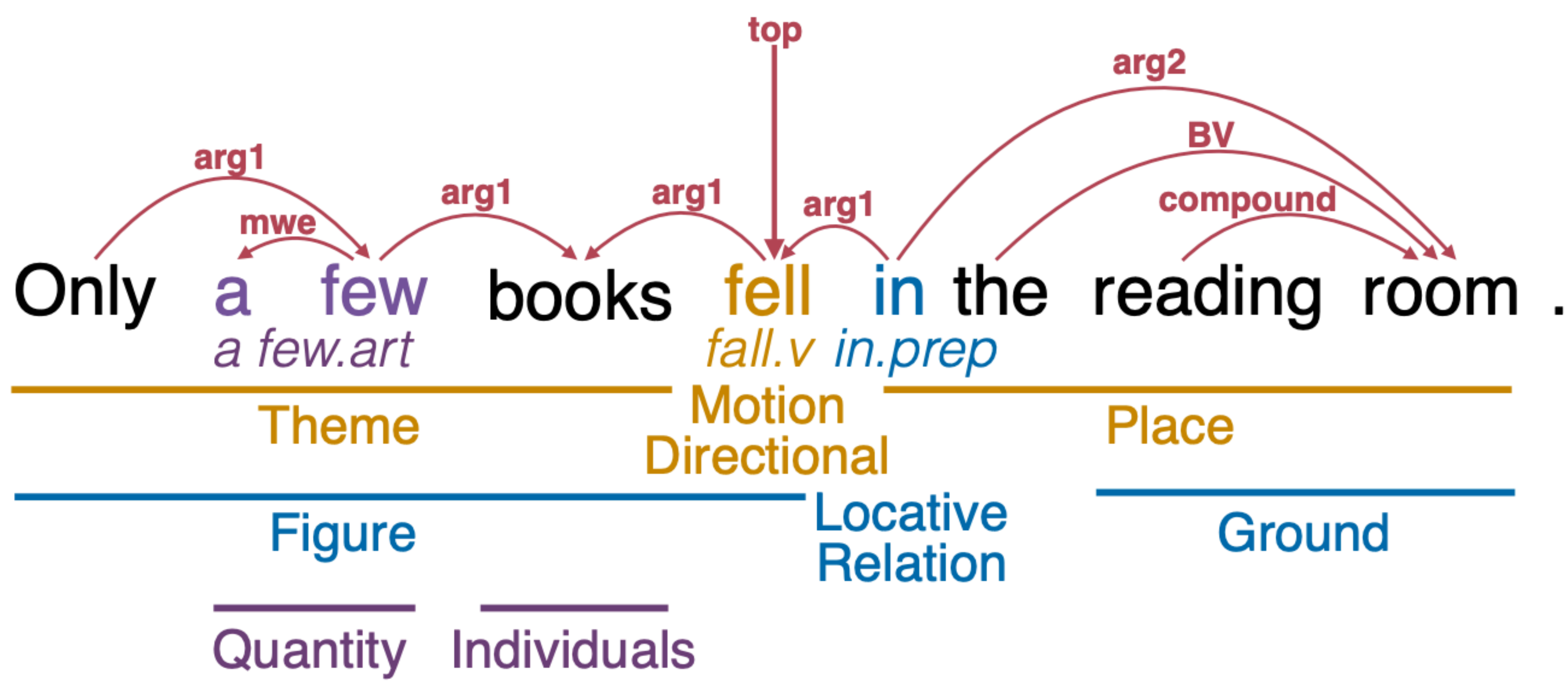
GRAPH format:



[Banarescu+2013]

Figure 1: Equivalent formats for representating the meaning of "The boy wants to go".

General Semantic Parsing



[Peng+NAACL'19]

Criticisms

- Pure NN approaches
  - Black-box machinery
  - Not truly “understanding”
- Formal methods for meaning representations
  - Not quite useful in real applications
    - Not complete in representation power
    - Semantics are not grounded
    - Undecidable or NP-hard
  - Usually fully supervised ⇒Still black-box machinery

Domain-Specific NLU

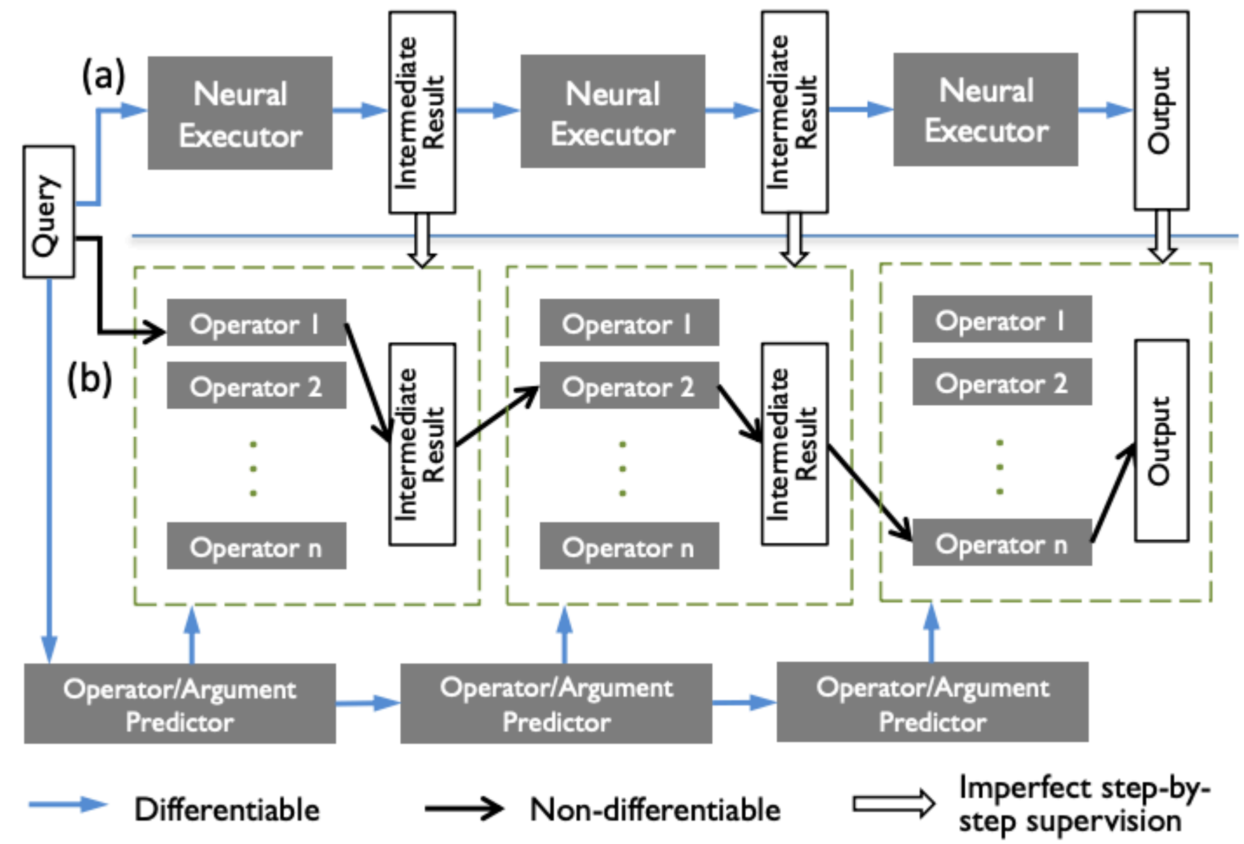
- General NLU is beyond state-of-the-art
- NLU in a certain domain can be done better
  - Let NN learn a little "**symbolic**" knowledge
  - Defined by the domain requirement
  - Weakly supervised ⇒ NN has to reason by itself
- Neuro-symbolic reasoning

Example: Semantic Parsing

Query:
How long is the game with the largest host country size?
Knowledge base (table):

Year	City	···	Area	···	Duration
···					
2000	Sydney	···	200	···	30
2004	Athens	···	250	···	20
2008	Beijing	···	350	···	25
2012	London	···	300	···	35
2016	Rio de Janeiro	···	200	···	40
···					

Table 1. An example of a natural language query and a knowledge base (table).



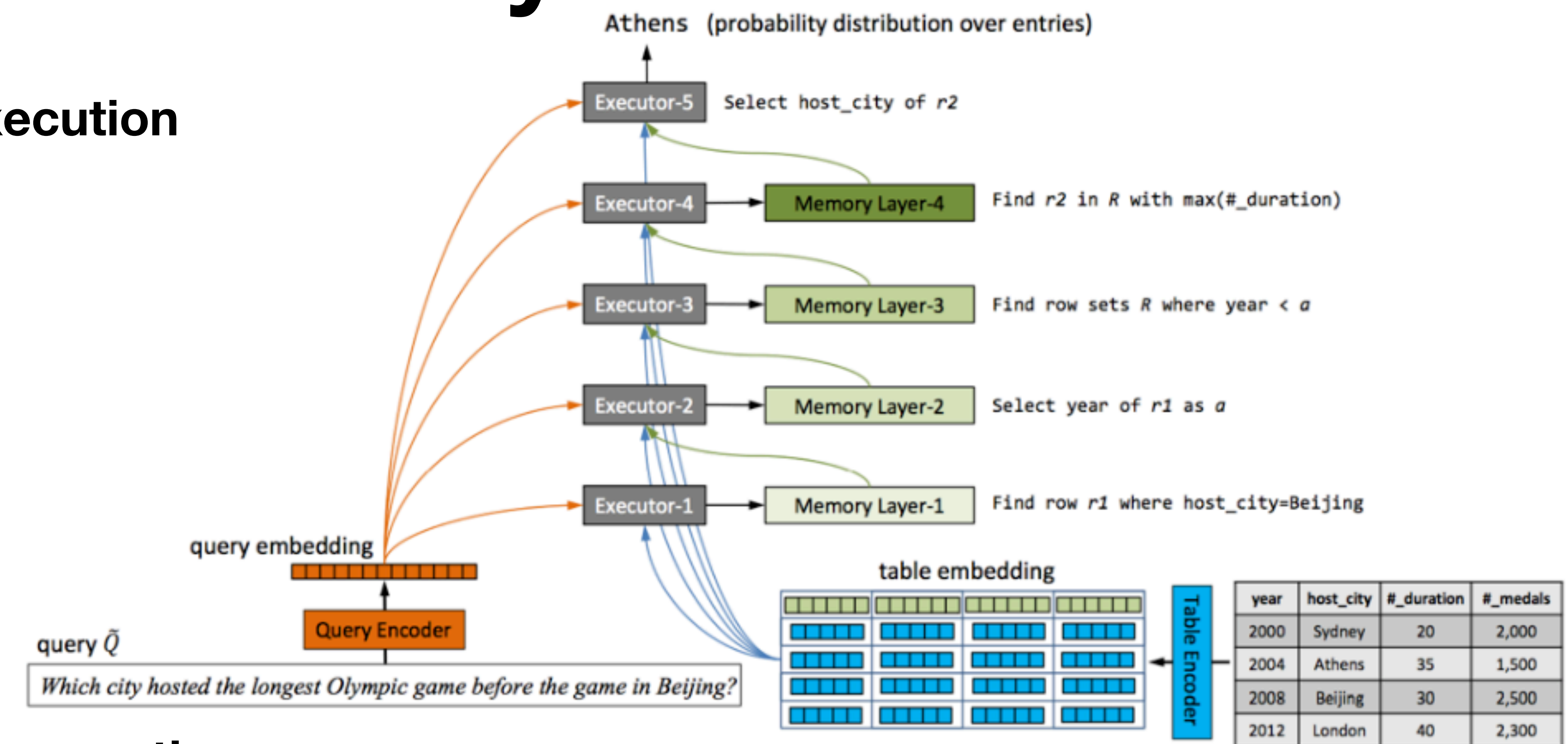
Goal: Answer + SQL-like execution

Coupling Distributed and Symbolic Execution for Natural Language Queries [Mou, Lu, Li, Jin, ICML'2017]

amii UNIVERSITY OF ALBERTA

Neuro-Symbolic Execution

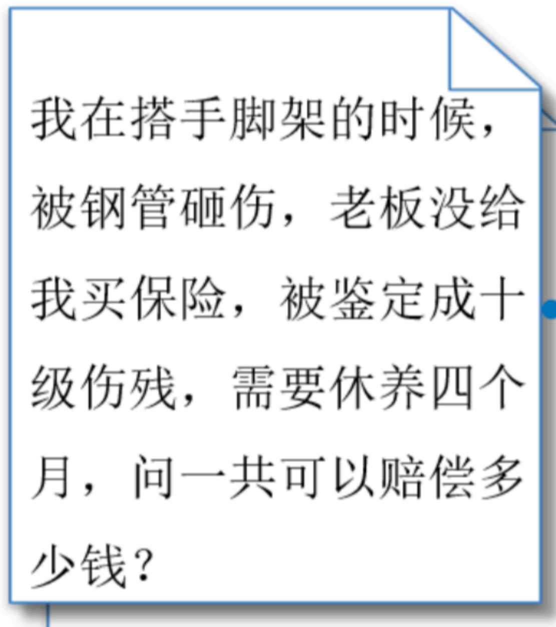
Neural execution



Symbolic execution

Operator	Explanation
`select_row`	Choose a row whose value of a particular column is mentioned in the query
`argmin`	Choose the row from previously selected candidate rows with the minimum value in a particular column
`argmax`	Choose the row from previously selected candidate rows with the maximum value in a particular column
`greater_than`	Choose rows whose value in a particular column is greater than a previously selected row
`less_than`	Choose rows whose value in a particular column is less than a previously selected row
`select_value`	Choose the value of a particular column and of the previously selected row
`EOE`	Terminate, indicating the end of execution

Example: Information Extraction



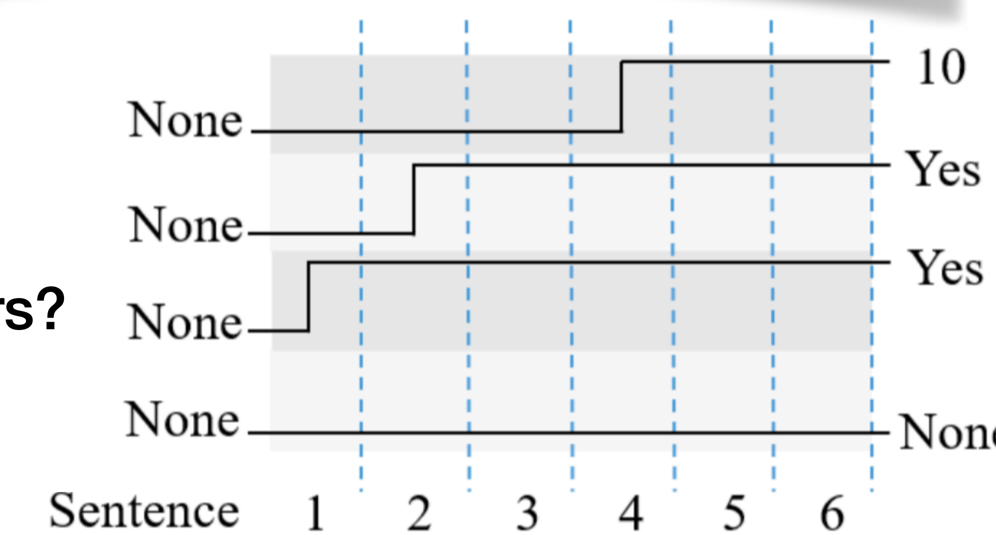
When I was building the scaffolding, I was hit by a steel tube, but my boss hadn't bought insurance for me. I was identified as Level-10 disabled, and I will have a sick leave for 4 months. How much money can I be compensated?

What is the injury level?

Is this an occupational injury?

Did the injury occur in working hours?

Has the contract terminated?



Goal: Answer + Evidence (which sentence)

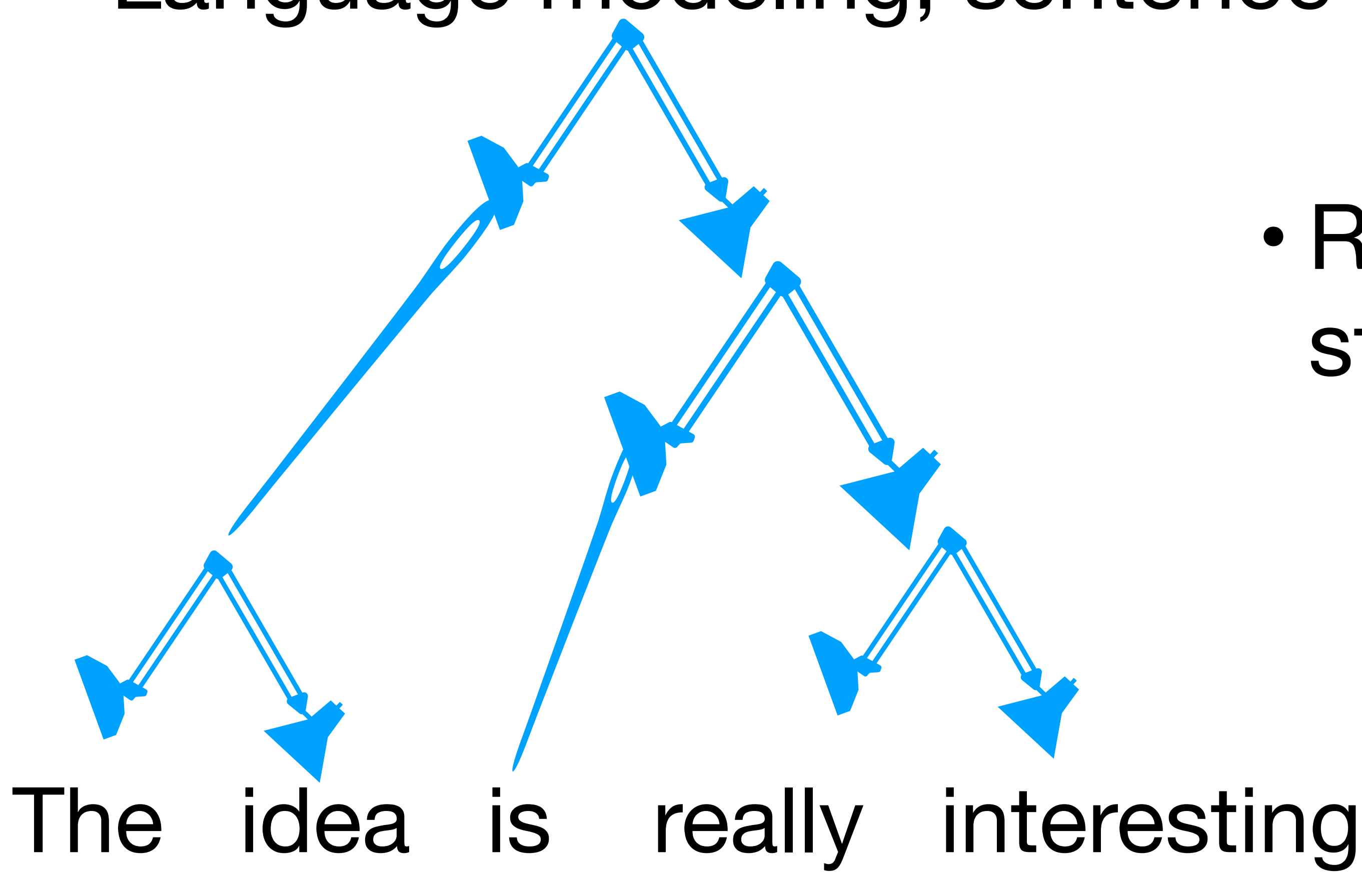
JUMPER: Learning When to Make Classification Decisions in Reading [Liu, Mou, Lu, Song, IJCAI'18]

Example: Grammar Induction

[Li, Mou, Keller, ACL'19]

Goal:

- Learn to parse a sentence without parsing labels
- Training objective: Downstream tasks
  - Language modeling, sentence classification, etc



- Relax constituents by structured attention

Example: Fuzzy Logic Reasoning

Input:
Premise: Several men helping each other pull in a fishing net.
Hypothesis: There is one man holding the net.

Sentence-Level Prediction:
[ ] Entailment [X] Contradiction [ ] Neutral

Phrase-Level Reasoning:
`Entailment:` *pull in a fishing net* VS *holding the net*
`Contradiction:` *several men* VS *one man*
`Neutral:` (None)
`Unaligned phrase(s):` *helping each other*

Why reasoning? Interpretability and explainability.

Unfortunately, none of previous studies can achieve phrasal reasoning for NLI

Example: Fuzzy Logic Reasoning

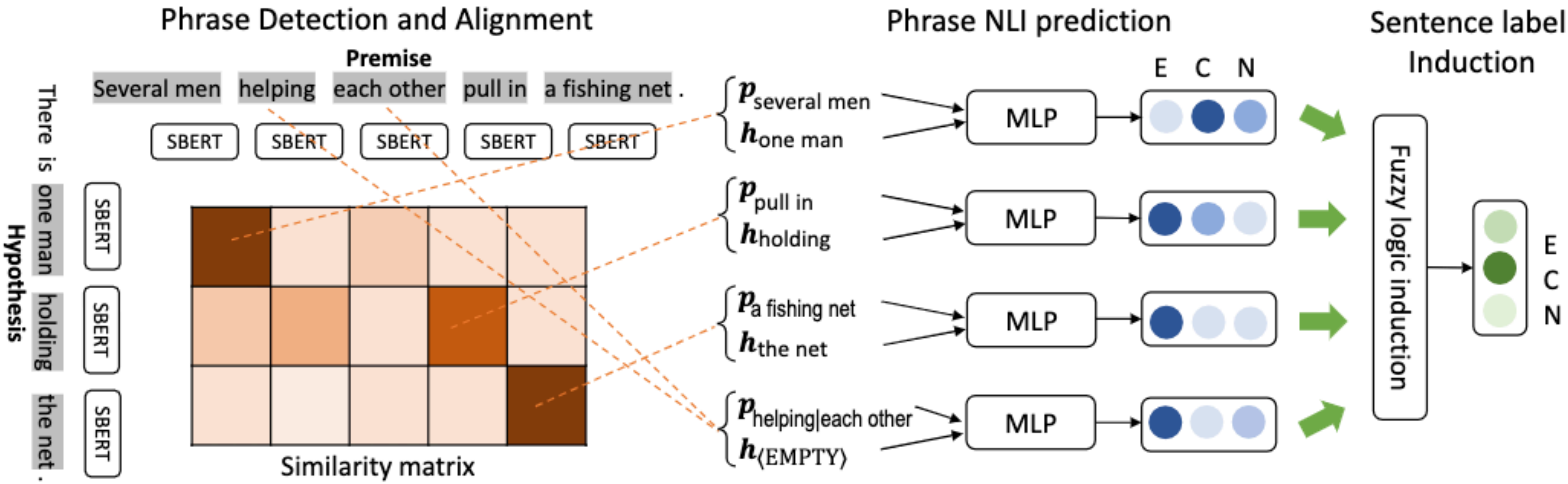


Figure 1: An overview of our Explainable Phrasal Reasoning (EPR) model.

Entailment Rule

$$s_{\text{sentence}}(\mathrm{E}|\mathrm{P},\mathrm{H}) = \left[\prod_{k=1}^{K'} P_{\text{phrase}}(\mathrm{E}|\mathrm{p}^{(k)},\mathrm{h}^{(k)})\right]^{\frac{1}{K'}}$$

Contradiction Rule

$$s_{\text{sentence}}(\mathrm{C}|\mathrm{P},\mathrm{H}) = \max_{k=1,\cdots,K} P_{\text{phrase}}(\mathrm{C}|\mathrm{p}^{(k)},\mathrm{h}^{(k)})$$

Rule of Neutral

$$s_{\text{sentence}}(\mathrm{N}|\mathrm{P},\mathrm{H}) = \left[\max_{k=1,\cdots,K'} P_{\text{phrase}}(\mathrm{N}|\mathrm{p}^{(k)},\mathrm{h}^{(k)})\right] \cdot \left[1 - s_{\text{sentence}}(\mathrm{C}|\mathrm{P},\mathrm{H})\right]$$

Example: Fuzzy Logic Reasoning

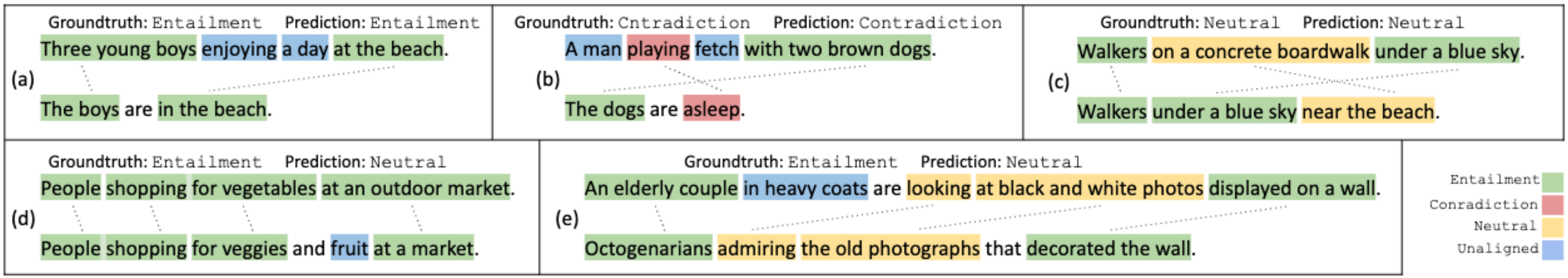


Figure 2: Examples of explainable phrasal reasoning predicted by our EPR model. Words in one color block are a detected phrase; a dotted line shows the alignment of two phrases; and the color represents the predicted phrasal NLI label. In Examples (d) and (e), EPR's prediction suggests several provided labels in SNLI are incorrect.

Summary

- Design a reasoning schema in an end task
- End-to-end training by
  - Reinforcement learning, or
  - Differentiable learning
- Challenges:
  - The reasoning schema has to be defined by humans.
  - No guarantee that machine performs the same reasoning as humans

Selected Topics

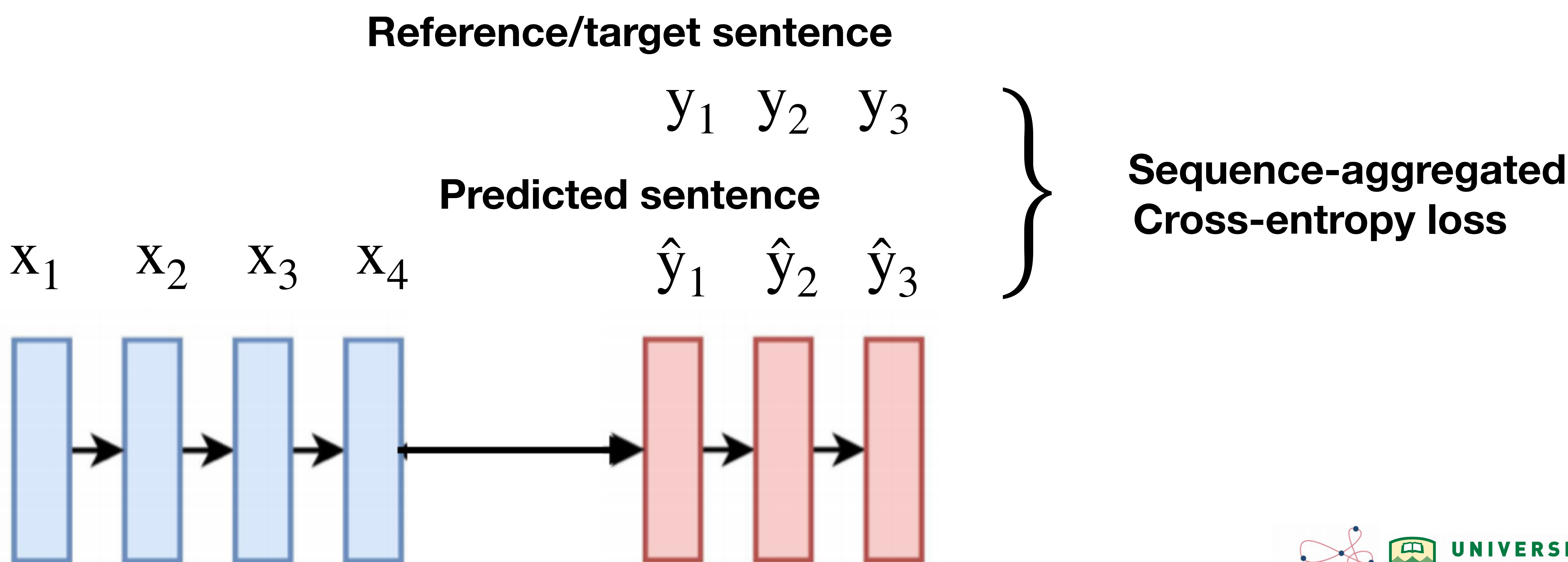
- Natural Language Understanding
  - Weakly supervised reasoning
- Natural Language Generation
  - **Search-based unsupervised text generation**

Supervised Text Generation

Sequence-to-sequence training

Training data = $\{(\mathbf{x}^{(m)}, \mathbf{y}^{(m)})\}_{m=1}^{M}$

known as a *parallel corpus*

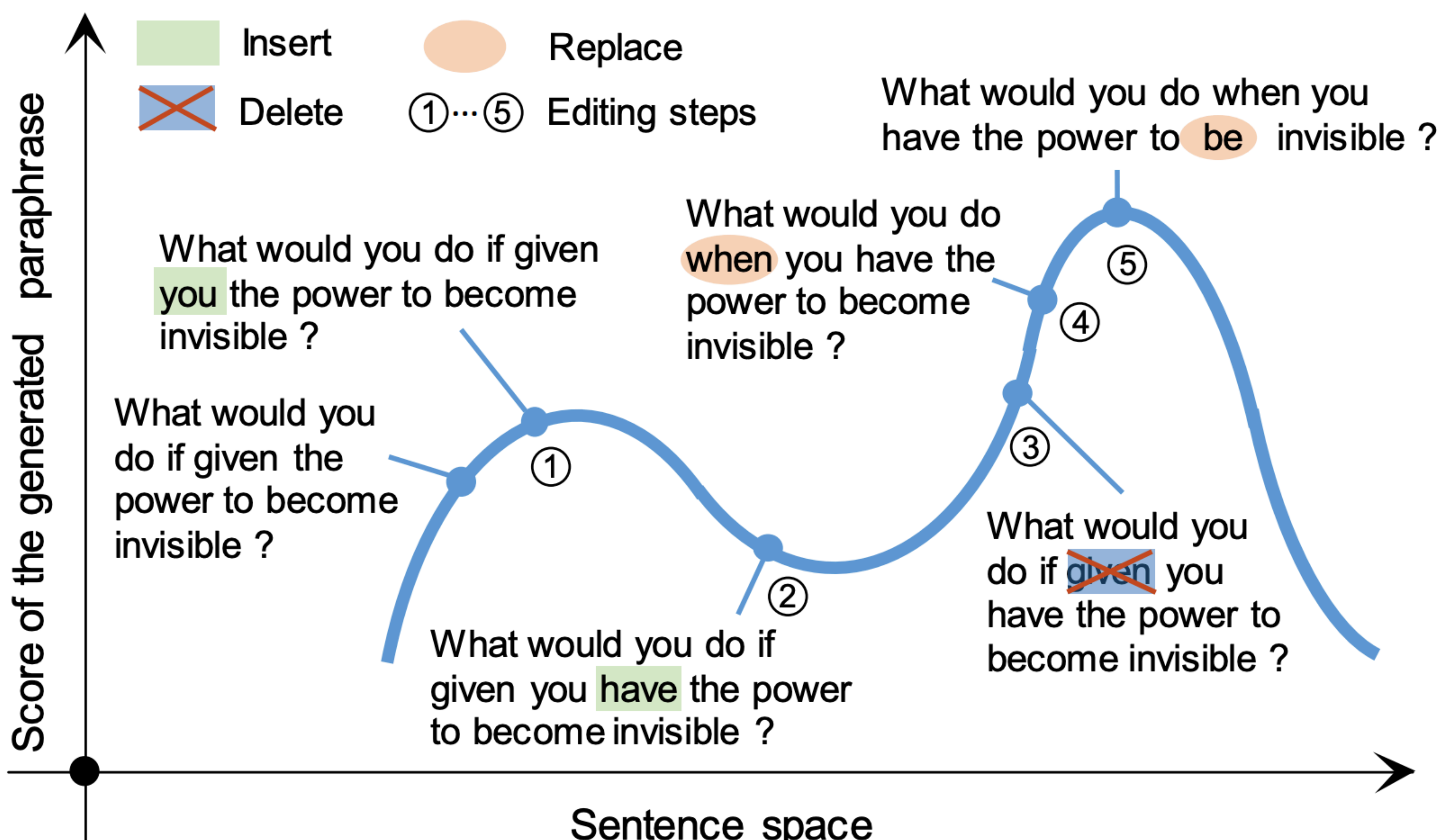


Unsupervised Text Generation

- Data = $\{\mathbf{x}^{(m)}\}_{m=1}^{M}$
- Important to **industrial applications**
  - Startup: No data
  - Minimum viable product
- Scientific interest
  - Unique sampling and search problems

General Framework

- **Search objective**
  - Scoring function measuring text quality
- **Search algorithm**
  - Currently we are using stochastic local search



Scoring Function

- **Search objective**
  - Scoring function measuring text quality

$$s(\mathbf{y}) = s_{LM}(\mathbf{y}) \cdot s_{Semantic}(\mathbf{y})^{\alpha} \cdot s_{Task}(\boldsymbol{y})^{\beta}$$

- Language fluency

- Semantic coherence

- Task-specific constraints

Scoring Function

- Search objective
  - Scoring function measuring text quality

$$s(\mathbf{y}) = s_{LM}(\mathbf{y}) \cdot s_{Semantic}(\mathbf{y})^{\alpha} \cdot s_{Task}(\mathbf{y})^{\beta}$$

- **Language fluency**
  - Language model estimates the "probability" of a sentence

$$\overleftrightarrow{\mathrm{PPL}}(\mathbf{y}) = \sqrt[2|\mathbf{y}|]{\prod_{i}^{|\mathbf{y}|} \frac{1}{p_{\overrightarrow{\mathrm{LM}}}(y_i|\mathbf{y}_{<i})} \prod_{i}^{|\mathbf{y}|} \frac{1}{p_{\overleftarrow{\mathrm{LM}}}(y_i|\mathbf{y}_{>i})}}.$$

$$s_{LM}(\mathbf{y}) = \mathrm{PPL}(\mathbf{y})^{-1}$$

- Semantic coherence

- Task-specific constraints

Scoring Function

- Search objective
  - Scoring function measuring text quality

$$s(\mathbf{y}) = s_{LM}(\mathbf{y}) \cdot s_{Semantic}(\mathbf{y})^{\alpha} \cdot s_{Task}(\boldsymbol{y})^{\beta}$$

- Language fluency

- **Semantic coherence**

$$s_{semantic} = \text{normalize}[\cos(\boldsymbol{e}(\mathbf{y}), \boldsymbol{e}(\mathbf{x}))]$$

- Task-specific constraints

Scoring Function

- Search objective
  - Scoring function measuring text quality

$$s(\mathbf{y}) = s_{LM}(\mathbf{y}) \cdot s_{Semantic}(\mathbf{y})^{\alpha} \cdot s_{Task}(\boldsymbol{y})^{\beta}$$

- Language fluency

- Semantic coherence

- **Task-specific constraints**
  - Paraphrasing: lexical dissimilarity with input
  - Summarization: length budget

Search Algorithm

- Observations:
  - The output closely resembles the input
  - Edits are mostly local
  - May have hard constraints

- Thus, we mainly used **local stochastic search**

Search Algorithm

(stochastic local search)

Start with $\mathbf{y}_0$ # an initial candidate sentence

Loop within budget at step t:

$\mathbf{y}' \sim \text{Neighbor}(\mathbf{y}_t)$ # a new candidate in the neighbor

Either reject or accept $\mathbf{y}'$

If accepted, $\mathbf{y}_t = \mathbf{y}'$, or otherwise $\mathbf{y}_t = \mathbf{y}_{t-1}$

Return the best scored $\mathbf{y}_*$

Search Algorithm

Local edits for $\mathbf{y}' \sim \text{Neighbor}(\mathbf{y}_t)$

- General edits
  - Word deletion
  - Word insertion
  - Word replacement

$$p(w_*|\cdot) = \frac{f_{\text{sim}}(\mathrm{x}_*, \mathrm{x}_0) \cdot f_{\text{exp}}(\mathrm{x}_*, \mathrm{x}_0) \cdot f_{\text{flu}}(\mathrm{x}_*)}{Z},$$

$$Z = \sum_{w_* \in \mathcal{W}} f_{\text{sim}}(\mathrm{x}_*, \mathrm{x}_0) \cdot f_{\text{exp}}(\mathrm{x}_*, \mathrm{x}_0) \cdot f_{\text{flu}}(\mathrm{x}_*),$$

Gibbs in Metropolis

- Task specific edits
  - Reordering, swap of word selection, etc.

Search Algorithm

Example: Metropolis—Hastings sampling

Start with $\mathbf{y}_0$ # an initial candidate sentence

Loop within budget at step t:

$\mathbf{y}' \sim \mathrm{Neighbor}(\mathbf{y}_t)$ # a new candidate in the neighbor

Either reject or accept $\mathbf{y}'$

$$\boxed{\begin{aligned} A(\mathrm{x}'|\mathrm{x}_{t-1}) &= \min\{1, A^*(\mathrm{x}'|\mathrm{x}_{t-1})\} \\ A^*(\mathrm{x}'|\mathrm{x}_{t-1}) &= \frac{\pi(\mathrm{x}')g(\mathrm{x}_{\mathrm{t}-1}|\mathrm{x}')}{\pi(\mathrm{x}_{t-1})g(\mathrm{x}'|\mathrm{x}_{t-1})} \end{aligned}}$$

If accepted, $\mathbf{y}_t = \mathbf{y}'$, or otherwise $\mathbf{y}_t = \mathbf{y}_{t-1}$

Return the best scored $\mathbf{y}_*$

Ning Miao, Hao Zhou, Lili Mou, Rui Yan, Lei Li. CGMH: Constrained sentence generation by Metropolis-Hastings sampling. In AAAI, pages 6834--6942, 2019.

Search Algorithm

Example: Simulated annealing

Start with $\mathbf{y}_0$ # an initial candidate sentence

Loop within budget at step t:

$\mathbf{y}' \sim \text{Neighbor}(\mathbf{y}_t)$ # a new candidate in the neighbor

Either reject or accept $\mathbf{y}'$

$$\boxed{p(\text{accept}|\mathrm{x}_*, \mathrm{x}_t, T) = \min\left(1, e^{\frac{f(\mathrm{x}_*)-f(\mathrm{x}_t)}{T}}\right)}$$

If accepted, $\mathbf{y}_t = \mathbf{y}'$, or otherwise $\mathbf{y}_t = \mathbf{y}_{t-1}$

Return the best scored $\mathbf{y}_*$

Xianggen Liu, Lili Mou, Fandong Meng, Hao Zhou, Jie Zhou, Sen Song. Unsupervised paraphrasing by simulated annealing. In ACL, pages 302--312, 2020.

Search Algorithm

Example: Hill climbing

Start with $\mathbf{y}_0$ # an initial candidate sentence

Loop within budget at step t:

$\mathbf{y}' \sim \text{Neighbor}(\mathbf{y}_t)$ # a new candidate in the neighbor

Either reject or accept $\mathbf{y}'$

whenever $\mathbf{y}'$ is better than $\mathbf{y}_{t-1}$

If accepted, $\mathbf{y}_t = \mathbf{y}'$, or otherwise $\mathbf{y}_t = \mathbf{y}_{t-1}$

Return the best scored $\mathbf{y}_*$

Raphael Schumann, Lili Mou, Yao Lu, Olga Vechtomova, Katja Markert. Discrete optimization for unsupervised sentence summarization with word-level extraction. In ACL, 2020.

Paraphrase Generation

- Search objective
  - Fluency
  - Semantic preservation
  - Expression diversity
    - The paraphrase should be different from the input

$$s_{exp}(\mathbf{y}_*, \mathbf{x}) = 1 - \mathrm{BLEU}(\mathbf{y}_*, \mathbf{x})$$

BLEU here measures the *n*-gram overlapping

- Search algorithm
- Search space
- Search neighbors

Xianggen Liu, Lili Mou, Fandong Meng, Hao Zhou, Jie Zhou, Sen Song. Unsupervised paraphrasing by simulated annealing. In ACL, pages 302--312, 2020.

Paraphrase Generation

- Search objective
  - Fluency
  - Semantic preservation
  - Expression diversity
    - The paraphrase should be different from the input

$$s_{exp}(\mathbf{y}_*, \mathbf{x}) = 1 - \mathrm{BLEU}(\mathbf{y}_*, \mathbf{x})$$

BLEU here measures the *n*-gram overlapping

- Search algorithm: Simulated annealing
- Search space: the entire sentence space with $\mathbf{y}_0$ = input $\mathbf{x}$
- Search neighbors
  - Generic word deletion, insertion, and replacement
  - Copying words in the input sentence

Text Simplification

Input	**Reference**
In 2016 alone, American developers had spent 12 billion dollars on **constructing** theme parks, <u>according to a Seattle based reporter.</u>	American developers had spent 12 billion dollars in 2016 alone on **building** theme parks.

Could be useful for

- education purposes (e.g., kids, foreigners)
- for those with dyslexia

Key observations

- Dropping phrases and clauses
- Phrase re-ordering
- Dictionary-guided lexicon substitution

Text Simplification

Search objective

- Language model fluency (discounted by word frequency)
- Cosine similarity
- Entity matching
- Length penalty
- Flesh Reading Ease (FRE) score [Kincaid et al., 1975]

Search operations

Dhruv Kumar, Lili Mou, Lukasz Golab, Olga Vechtomova. Iterative edit-based unsupervised sentence simplification. In ACL, pages 7918--7928, 2020.

amii UNIVERSITY OF ALBERTA

Text Simplification

Search objective

- Language model fluency (discounted by word frequency)
- Cosine similarity
- Entity matching
- Length penalty
- Flesh Reading Ease (FRE) score [Kincaid et al., 1975]

Search operations

- Dictionary-guided substitution (e.g., WordNet)
- Phrase removal
- Re-ordering

with parse trees

Text Summarization

Input	Reference
The *world's biggest miner* ***bhp billiton*** *announced tuesday it was* ***dropping*** *its controversial hostile* ***takeover bid*** *for rival* ***rio tinto*** *due to the state of the global economy*	bhp billiton drops rio tinto takeover bid

Key observation

- Words in the summary mostly come from the input
- If we generate the summary by selecting words, we have

```
bhp billiton dropping hostile bid for rio tinto
```

Text Summarization

- Search objective
  - Fluency
  - Semantic preservation
  - A hard length constraint

$$f_{\mathrm{LEN}}(\mathbf{y}; s) = \begin{cases} 1, & \text{if } |\mathbf{y}| = s, \\ -\infty, & \text{otherwise.} \end{cases}$$

  (Explicitly controlling length is not feasible in previous work)

- Search space
- Search neighbor
- Search algorithm

Raphael Schumann, Lili Mou, Yao Lu, Olga Vechtomova, Katja Markert. Discrete optimization for unsupervised sentence summarization with word-level extraction. In ACL, pages 5032--5042, 2020.

Text Summarization

- Search objective
  - Fluency
  - Semantic preservation
  - A hard length constraint $$f_{\text{LEN}}(\mathbf{y}; s) = \begin{cases} 1, & \text{if } |\mathbf{y}| = s, \\ -\infty, & \text{otherwise.} \end{cases}$$

  (Explicitly controlling length is not feasible in previous work)

- Search space with only feasible solutions

$$|\mathscr{V}|^{|\mathbf{y}|} \Longrightarrow \binom{|\mathbf{x}|}{s}$$

- Search neighbor: swap only

- Search algorithm: hill-climbing

General Performance

Paraphrase generation

	Model	Quora iBLEU	Quora BLEU	Quora Rouge1	Quora Rouge2	Wikianswers iBLEU	Wikianswers BLEU	Wikianswers Rouge1	Wikianswers Rouge2
Supervised	ResidualLSTM	12.67	17.57	59.22	32.40	22.94	27.36	48.52	18.71
	VAE-SVG-eq	15.17	20.04	59.98	33.30	26.35	32.98	50.93	19.11
	Pointer-generator	16.79	22.65	61.96	36.07	31.98	39.36	57.19	25.38
	Transformer	16.25	21.73	60.25	33.45	27.70	33.01	51.85	20.70
	Transformer+Copy	17.98	24.77	63.34	37.31	31.43	37.88	55.88	23.37
	DNPG	**18.01**	**25.03**	**63.73**	**37.75**	**34.15**	**41.64**	**57.32**	**25.88**
Supervised + Domain-adapted	Pointer-generator	5.04	6.96	41.89	12.77	21.87	27.94	53.99	20.85
	Transformer+Copy	6.17	8.15	44.89	14.79	23.25	29.22	53.33	21.02
	Shallow fusion	6.04	7.95	44.87	14.79	22.57	29.76	53.54	20.68
	MTL	4.90	6.37	37.64	11.83	18.34	23.65	48.19	17.53
	MTL+Copy	7.22	9.83	47.08	19.03	21.87	30.78	54.10	21.08
	DNPG	10.39	16.98	56.01	28.61	25.60	35.12	56.17	23.65
Unsupervised	VAE	8.16	13.96	44.55	22.64	17.92	24.13	31.87	12.08
	Lag VAE	8.73	15.52	49.20	26.07	18.38	25.08	35.65	13.21
	CGMH	9.94	15.73	48.73	26.12	20.05	26.45	43.31	16.53
	UPSA	12.03	18.21	59.51	32.63	24.84	32.39	54.12	21.45

BLEU and ROUGE scores are automatic evaluation metrics based on references

amii UNIVERSITY OF ALBERTA

General Performance

Text Summarization

	Model	Data			Len D	Rouge F1			Len O
		article	title	external		R-1	R-2	R-L	
	Lead-N-8	✓			8	21.39	7.42	20.03	7.9
A	*HC_article_8*	✓			8	23.09	7.50	21.29	7.9
	HC_title_8		✓		8	**26.32**	**9.63**	**24.19**	7.9
	Lead-N-10	✓			10	23.03	7.95	21.29	9.8
	Wang and Lee (2018)	✓	✓		-	27.29	10.01	24.59	10.8
	Zhou and Rush (2019)		✓	billion	-	26.48	10.05	24.41	9.3
B	*HC_article_10*	✓			10	24.44	8.01	22.21	9.8
	HC_title_10		✓		10	27.52	10.27	24.91	9.8
	HC_title+twitter_10		✓	twitter	10	28.26	10.42	25.43	9.8
	HC_title+billion_10		✓	billion	10	**28.80**	**10.66**	**25.82**	9.8
	Lead-P-50	✓			50%	24.97	8.65	22.43	14.6
	Fevry and Phang (2018)	✓		SNLI	50%	23.16	5.93	20.11	14.8
C	Baziotis et al. (2019)	✓			50%	24.70	7.97	22.14	15.1
	HC_article_50p	✓			50%	25.58	8.44	22.66	14.9
	HC_title_50p		✓		50%	**27.05**	**9.75**	**23.89**	14.9

General Performance

Text Simplification

Method	BLEU	SARI	Add	Delete	Keep	GM	FKGL	Len
Reference	100	70.13	-	-	-	83.74	3.20	12.75
Baselines								
Complex	21.30	2.82	-	-	-	7.75	8.62	23.06
Reduced-250	11.79	28.39	-	-	-	18.29	-0.23	14.48
Supervised Methods								
PBMT-R	18.1	15.77	3.07	38.34	5.90	16.89	7.59	23.06
Hybrid	14.46	28.61*	0.95*	78.86*	6.01*	20.34	4.03	12.41
EncDecA	21.68	24.12	2.73	62.66	6.98	22.87	5.11	16.96
Dress	23.2	27.37	**3.08**	71.61	7.43	25.2	4.11	14.2
Dress-Ls	24.25	26.63	3.21	69.28	7.4	25.41	4.21	14.37
DMass	11.92	31.06	1.25	84.12	7.82	19.24	3.60	15.07
S2S-All-FA	19.55	30.73	2.64	81.6	**7.97**	24.51	2.60	10.81
Edit-NTS	19.85	30.27*	2.71*	80.34*	7.76*	24.51	3.41	10.92
EncDecP	23.72	28.31	-	-	-	25.91	-	-
EntPar	11.14	**33.22**	2.42	**89.32**	7.92	19.24	1.34	7.88
Unsupervised Methods (Ours)								
Base	**27.22**	26.07	2.35	68.35	7.5	26.64	2.95	12.9
Base+LS	27.17	26.26	2.28	68.94	7.57	**26.71**	2.93	12.88
Base+RO	26.31	26.99	2.47	70.88	7.63	26.64	3.14	12.81
Base+LS+RO	26.21	27.11	2.40	71.26	7.67	26.66	3.12	12.81

amii UNIVERSITY OF ALBERTA

General Performance

Human evaluation on paraphrase generation

Model	Relevance		Fluency	
	Mean Score	Agreement	Mean Score	Agreement
VAE	2.65	0.41	3.23	0.51
Lag VAE	2.81	0.45	3.25	0.48
CGMH	3.08	0.36	3.51	0.49
UPSA	**3.78**	0.55	**3.66**	0.53

General Performance

Examples

Input	VAE	Lag VAE	CGMH	UPSA
where are best places for spring snowboarding in the us?	where are best places for running in the world? (3.33)	where are best places for honeymoon year near the us? (2.33)	Where is best store for the snowboarding in the US? (3.67)	Where can I find the best places in the US for snowboarding? (4.67)
how can i become good in studies?	how can i have a good android phone? (2.33)	how can i become good students? (4.33)	how can i become very rich in studies? (4.00)	how should i do to get better grades in my studies? (4.33)
what are the pluses and minuses about life as a foreigner in singapore?	what are the UNK and most interesting life as a foreigner in medieval greece? (2.33)	what are the UNK and interesting things about life as a foreigner? (2.33)	what are the misconception about UNK with life as a foreigner in western? (2.33)	what are the mistakes and pluses life as a foreigner in singapore? (2.67)

Main conclusion

- Search-based unsupervised text generation works in a variety of applications
- Surprisingly, it does yield **fluent sentences**.

Conclusion

- Natural Language Understanding
  - Weakly supervised reasoning
- Natural Language Generation
  - Search-based unsupervised text generation

A Different Perspective

- Deep learning works well for **continuous** features
- However, everything in NLP is **discrete**
  - Discrete input space
    - Tree-based convolution
  - Discrete latent space
    - Weakly supervised reasoning
  - Discrete output space
    - Continuous-to-discrete generation
    - Unsupervised text generation

Advertisements

- Lili Mou is accepting all-level students
  - URA, MSc, PhD, postdoc
  - A typical path

    466/566 course project → Individual Study → RAship
- Previous achievements
  - CMPUT499 (F20) → URA (W+S21) → CIKM'21
  - CMPUT499 (W21) → URA (S21) → EMNLP'21 (Findings)

Thanks for listening!