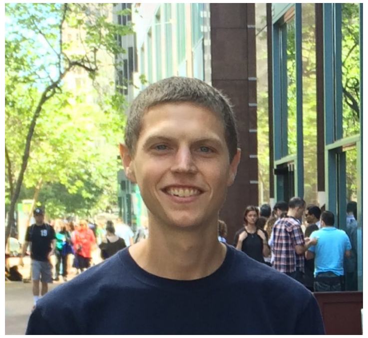
aws Microsoft University of Colorado Boulder Ai2 Allen Institute for AI

Penn PURDUE UNIVERSITY UNIVERSITY of WASHINGTON THE UNIVERSITY OF UTAH


Maria L. Pacheco


Sean Welleck


Yejin Choi


Vivek Srikumar


Dan Goldwasser


Dan Roth

October 2022

COLING Tutorial

NS4NLP: Neuro-Symbolic Modeling for NLP

COLING 2022

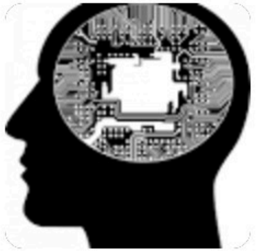
AI is in the News: The Success of Deep Learning

Phys.Org

Microsoft, Alibaba AI programs beat humans in a Stanford reading test

Now it's basic reading comprehension. ... as Stanford Question Answering Dataset or SQuAD for short, asks contestants—human and robot—to ...

Jan 19, 2018



GeekWire

Allen Institute's Aristo AI system finally passes an eighth-grade science test

The Aristo AI software has matched an eighth-grader's ability to pass a science test. (AI2 Illustration). Four years after the late Seattle billionaire ...

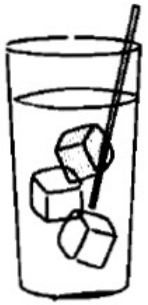
Sep 4, 2019

The Economist

A new AI language model generates poetry and prose

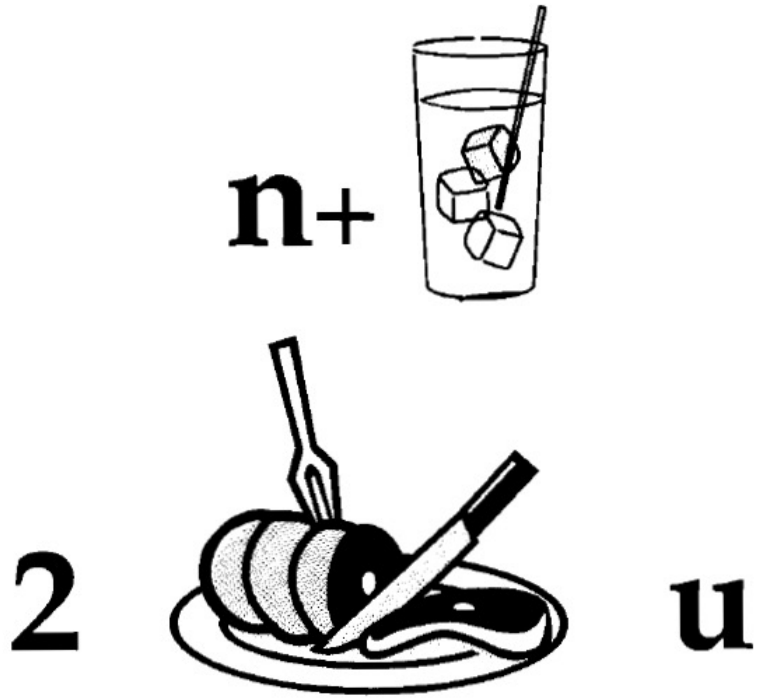
GPT-3, as it is more commonly known, was developed by OpenAI, an artificial-intelligence (AI) laboratory based in San Francisco, and which...

Aug 6, 2020

n+ 

2  **u**

Nice to Meet You



- Identify units
- Consider multiple representations & interpretations
  - Pictures, text, layout, spelling, phonetics
- Put it all together:
  - Determine "best" global interpretation
- Satisfy expectations
  - Slide; puzzle

Computational Problem:
Assigning values to multiple variables, accounting for interdependencies among them

Natural Language Understanding

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions, so they satisfy our expectations

- Of course, our programs need a lot more in order to understand and communicate in natural language
  - but it exemplifies some important aspects – "discrete reasoning";
    – "understanding ~~ best interpretation"

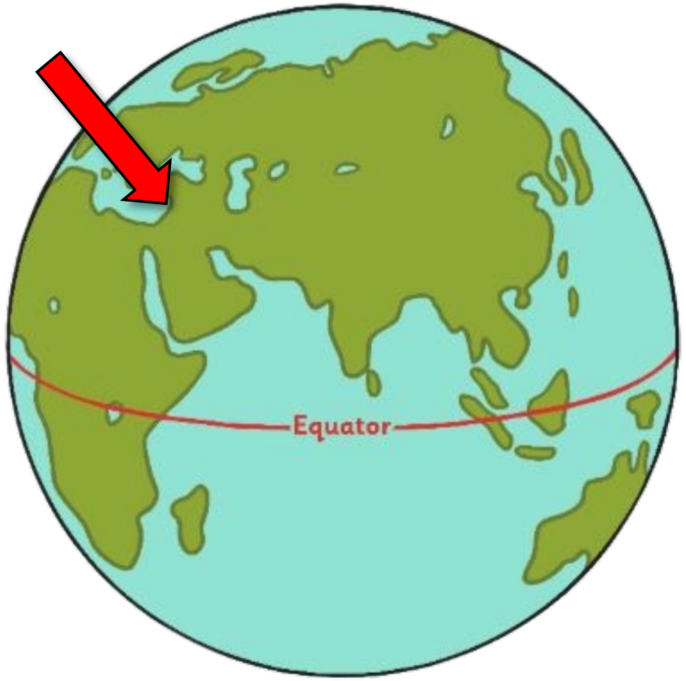
- And this example brings up another important question:
  - How do we **train** for these kinds of tasks?

Knowledge is Key

- In Yokneam, in what month of the year is the longest day?

You probably don't know the answer

But you have a plan



- In New York State, in what month of the year is the longest day?
- In Boston, in what month of the year is the longest day?
- In Melbourne, in what month of the year is the longest day?

- How do we **express** this plan?
- And how do we **train** for it? Or use it?

Learning to Reason over Natural Language

- Making decisions that depend on natural language understanding requires **reasoning abilities**, that depend on multiple, interdependent, models.
  - Sometimes it is useful to think about it as "symbols"

A lot of what we face is new and sparse

"Northern Hemisphere" is a symbol

- It cannot be accomplished by "evaluating" a single model nor can we **train directly** to accomplish it.

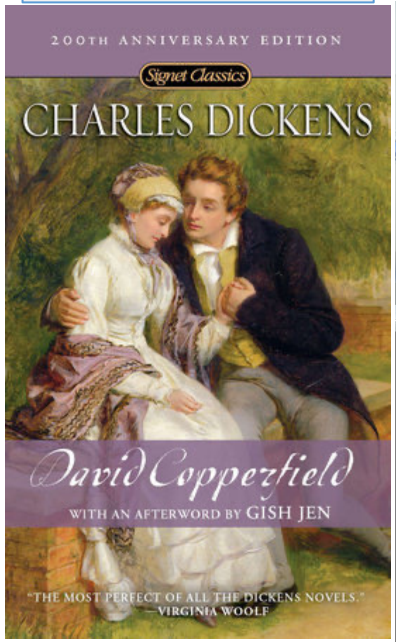
- At the heart of it is a **planning process** that determines what **modules** are relevant and what knowledge needs to be accessed to support the decision.
  - We need to **decompose, compose,** and **plan**

Exploit compositionality

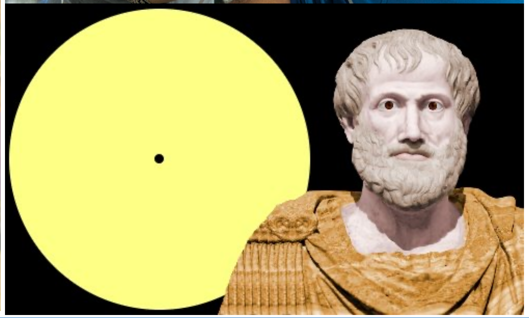
Putting things together

Supervise accordingly

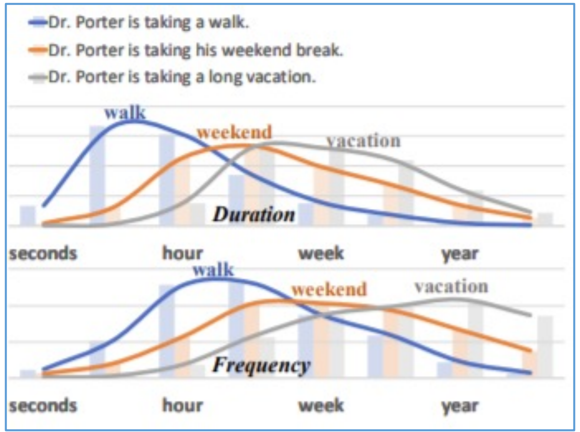
Facts & Fiction



When and Where?



Temporal Commonsense



Did Aristotle have a laptop?

Will we make it to dinner before the movie?

Reasoning in Natural Language

end time = start time + duration
Zhou et al. NAACL'21, EMNLP'22

- Will we make it to the movie?
  - Time now + [time to get to the movie] < start of movie

- Will we make it to dinner before the movie?
  - Time now + [time to get to dinner] + [duration of dinner] < start of the movie

How about parking?

A fancy Japanese or Chipotle?

- Will we make it to the movie after the game?
  - Start time of the game + [duration of the game] < start of the movie

Identifying **compositional components (symbols)** is essential to facilitate training, and the **planning** process.

- Will we make it to a movie after the game?
  - Start time of the game + [duration of the game] < start of any movie

Reasoning: Is the **end time** of the game/dinner before the **start time** of the/any movie?

Let’s Talk about Dinner

- → Let’s talk about dinner.
  - **A:** Where do you want to go?
- → I really enjoy Mexican food, but not when it is spicy.
  - **A:** How about a Mexican restaurant with **plenty** of **non-spicy options**?
- → Yep, is there one in Philadelphia?
  - **A:** Here are a couple of good options
  - …….

$\exists\ r \in$ [Restaurants] | [Cuisine](r, Mexican)$\wedge$ [InCity](PH)

Constraints:

$\forall\ r \in$ [Restaurants] → [Contains](r, m) $\wedge$ [Menus](m)

$\forall\ m \in$ [Menus] → [Contains](d,m) $\wedge$ [Dishes](d)

$\exists\ o \in$ [Dishes] → [Tastes](o, $\neg$spicy)

- Don’t worry, you can still use neural embeddings for predicates and functions that are learned concepts
- But this abstraction shows that we need to put some learned components together, along with incorporating declarative constraints…

[Example due to Tania Berdax-Weiss]

Symbolic Reasoning?

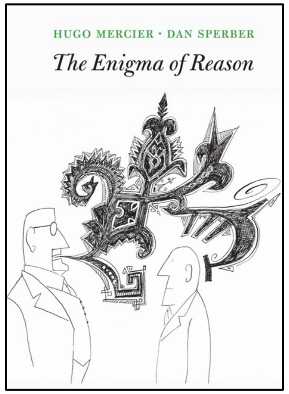
- Some people think that **symbols** are an evil invention of old AI people.
- It's not.
- Language is a symbolic system

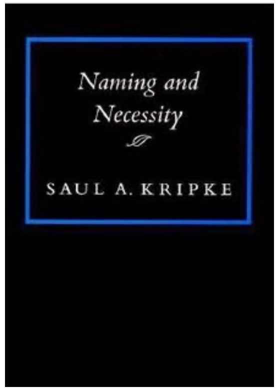
Rodgers finished 23-of-36 for 296 yards and two touchdowns. His numbers could've been even better had his receivers not dropped a couple of his passes. One dropped ball was a potential score to Allen Lazard. Despite the drop, Lazard made up for it by leading the Packers in receiving. With Davante Adams tied up with Jalen Ramsey, Lazard was able to snatch four balls for 96 yards and a touchdown. Adams still had a great game despite Ramsey's coverage, hauling in nine of his 10 targets for 66 yards and a touchdown. The score frustrated Ramsey because another defensive back was supposed to pick up Adams, who was in motion.

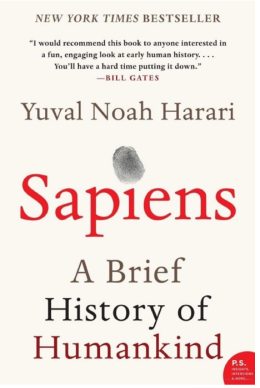
Mayor Rahm Emanuel now has raised more than $10 million toward his bid for a third term – more than five times the total raised by his 10 challengers combined, campaign finance records show.

- Even though we communicate via speech, gestures, writing, which are continuous, symbols are the invariants of this communication.

  - Harari: Language – the ability to assign symbols to "things" and "reason" about them is key to human cognitive revolution
  - Kripke: "Naming" things is key to communication and to cognition
  - The Enigma of Reason: "Reasoning is about giving reasons"







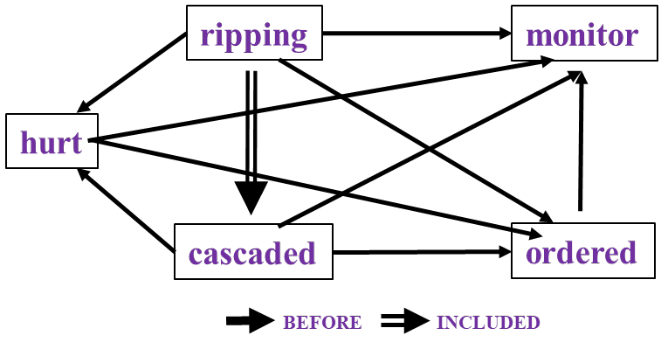
Putting Things Together

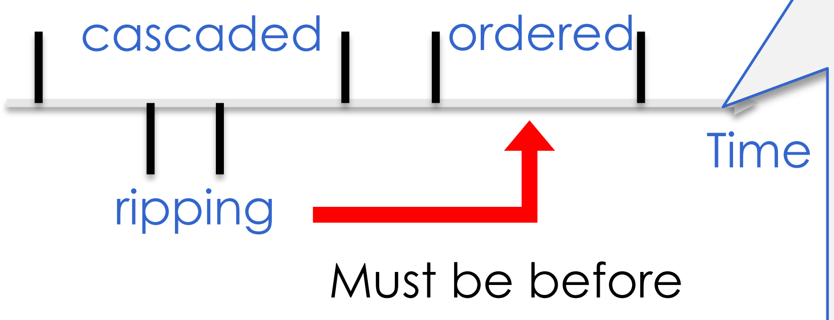
How to think about “reasoning” with learned modules?

Reasoning about Time and Events

[Ning et al. *SEM'2018; ACL'18, EMNLP'18, EMNLP'19; Wang et al. EMNLP'20]

- In Los Angeles that lesson was brought home today when tons of earth **cascaded** down a hillside, **ripping** two houses from their foundations. No one was **hurt**, but firefighters **ordered** the evacuation of nearby homes and said they'll **monitor** the shifting ground until March 23rd.





1. Reasoning: How to exploit these [declarative & statistical] "expectations"?
2. How/why does it impact generalization & supervision?

- Very difficult task— hinders exhaustive annotation (O(N^2) edges)
- But, it's rather easy to get partial annotation (and partial predictions)
- And, we have **strong expectations** from the output
  - Transitivity
  - Some events tend to precede others, or follow others

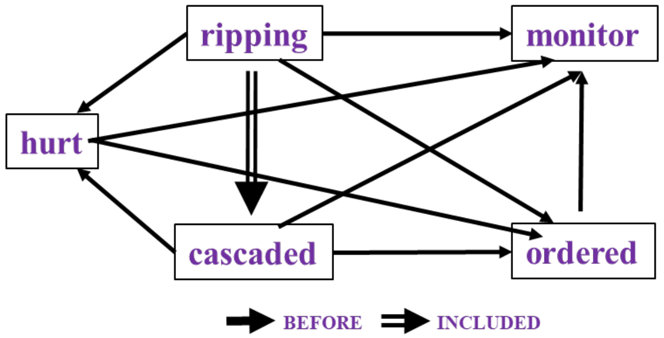
More than 10 people have (**event1: died**), police said.
A car (**event2: exploded**) on Friday in a group of men.

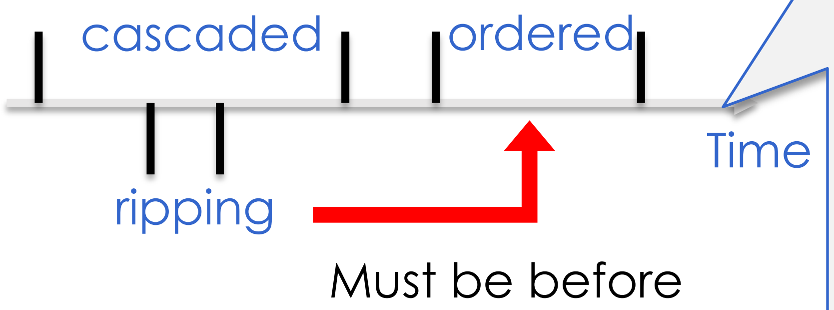
Reasoning about Time and Events

[Ning et al. *SEM’2018; ACL’18, EMNLP’18, EMNLP’19; Wang et al. EMNLP’20]

Hard Constraints: If event A happens before event B, and B happens before C, then A happens before C

was brought home today when tons of earth **cascaded** down a hillside, their foundations. No one was **hurt**, but firefighters **ordered** the evacuation they'll **monitor** the shifting ground until March 23rd.





1. Reasoning: How to exploit these [declarative & statistical] “expectations”?
2. How/why does it impact generalization & supervision?

- Very d icult task— hinders exhaustive annotation ($O(N^2)$ edges)
- But, it’s ather easy to get partial annotation (and partial predictions)
- And, we have **strong expectations** from the output
  - Transitivity
  - Some events tend to precede others, or follow others

More than 10 people have (**event1: died**), police said.
A car (**event2: exploded**) on Friday in a group of men.

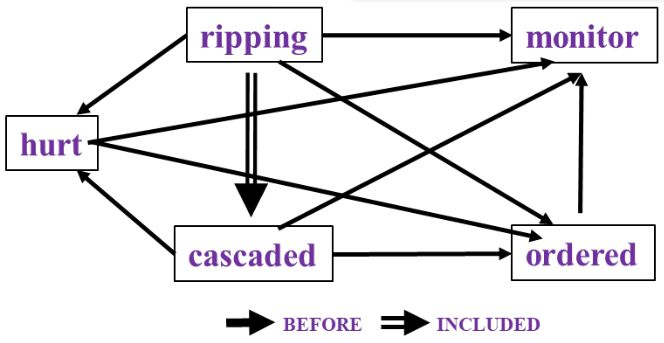
Reasoning about Time and Events

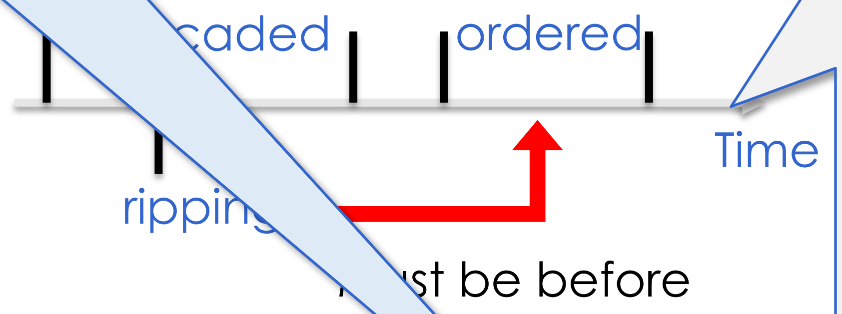
[Ning et al. *SEM'2018; ACL'18, EMNLP'18, EMNLP'19; Wang et al. EMNLP'20]

Hard Constraints: If event A happens before event B, and B happens before C, then A happens before C

Soft Constraints: Some events tent to happen before others

was … tons of earth **cascaded** down a hillside, … their … rt, but firefighters **ordered** the evacuation … the … nd until March 23rd.





1. Reasoning: How to exploit these [declarative & statistical] "expectations"?
2. How/why does it impact generalization & supervision?

- Very d…icult task— hinders exhaustive annotation (O(N^2) edges)
- But, it's …ather easy to get partial annotation (and partial predictions)
- And, we have **strong expectations** from the output
  - Transitivity
  - Some events tend to precede others, or follow others

More than 10 people have (**event1: died**), police said.
A car (**event2: exploded**) on Friday in a group of men.

Integer Linear Programming Inference for NLP

Variables are outputs of learned models

Statistical “soft” constrains regularize the objective

Hard constrains

real variable

$$\hat{I} = arg \max_{I} \sum_{i<j} \sum_{r} (f_r(ij) + \boxed{\boldsymbol{h_r(ij)}}) I_r(ij)$$

s.t. $\forall i, j, k$

$$\sum_r I_r(ij) = 1, \qquad I_{r1}(ij) + I_{r2}(jk) - I_{r3}(ik) \leq 1$$

Uniqueness

Transitivity (no loops)

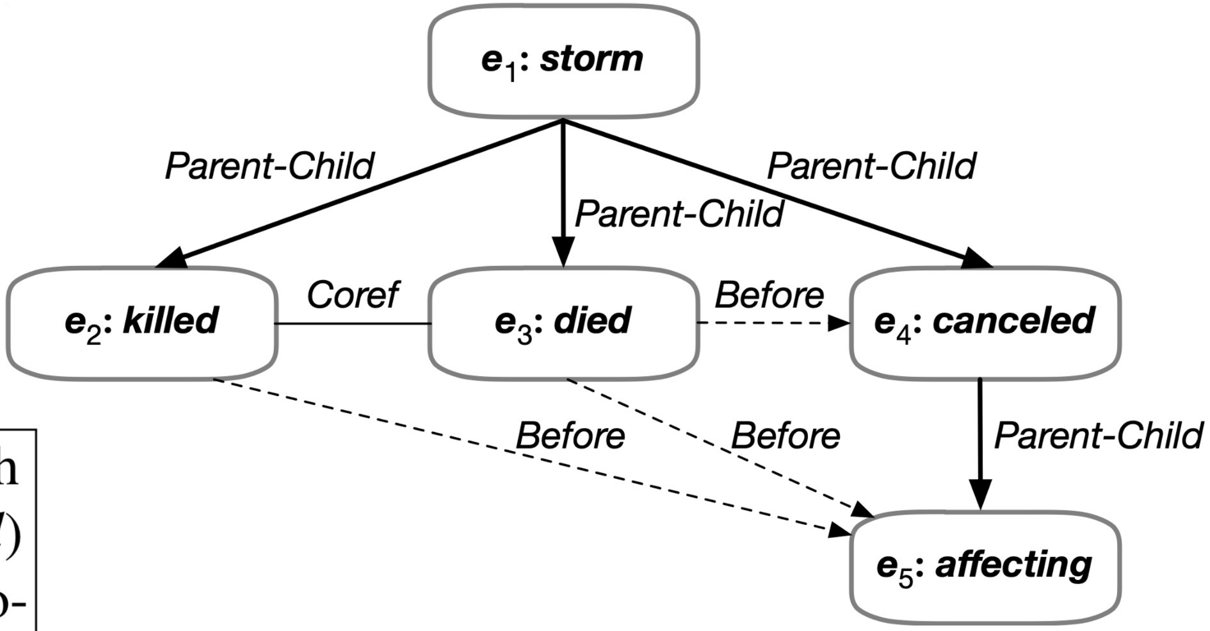
Maximizing the score of the entire graph while enforcing declarative (e.g., transitivity) and statistical constraints.

[Roth & Yih 2004, many others]

- Involves reasoning with external knowledge on top of neural models that constitute components (here: temporal relations) ; can express a range of reasoning patters (e.g, abduction)
  - Declarative knowledge: **Transitivity**
  - Statistical knowledge: **Commonsense knowledge**

Reasoning about Event Relations [Wang et al. EMNLP'20]

- ❑ Temporal Relations
- ❑ Subevent Relations
- ❑ Event Coreference



On Tuesday, there was a typhoon-strength (***e_1:storm***) in Japan. One man got (***e_2:killed***) and thousands of people were left stranded. Police said an 81-year-old man (***e_3:died***) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines (***e_4:canceled***) 230 domestic flights, (***e_5:affecting***) 31,600 passengers.

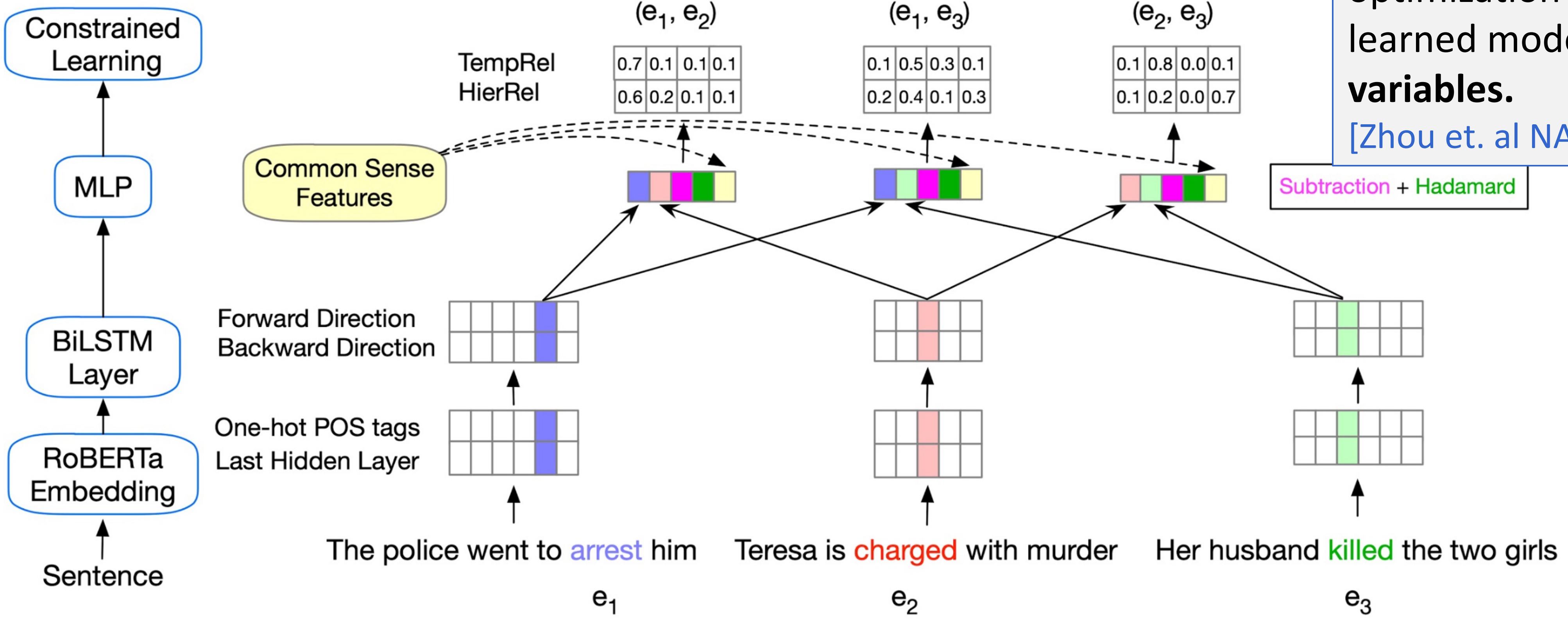
Reasoning about Event Relations [Wang et al. EMNLP'20]

- ❑ Temporal Relations
- ❑ Subevent Relations
- ❑ Event Coreference

Enforcing logical constraints: Temporal, Symmetry, and Conjunctive by **converting declarative constraints into differentiable learning objectives**

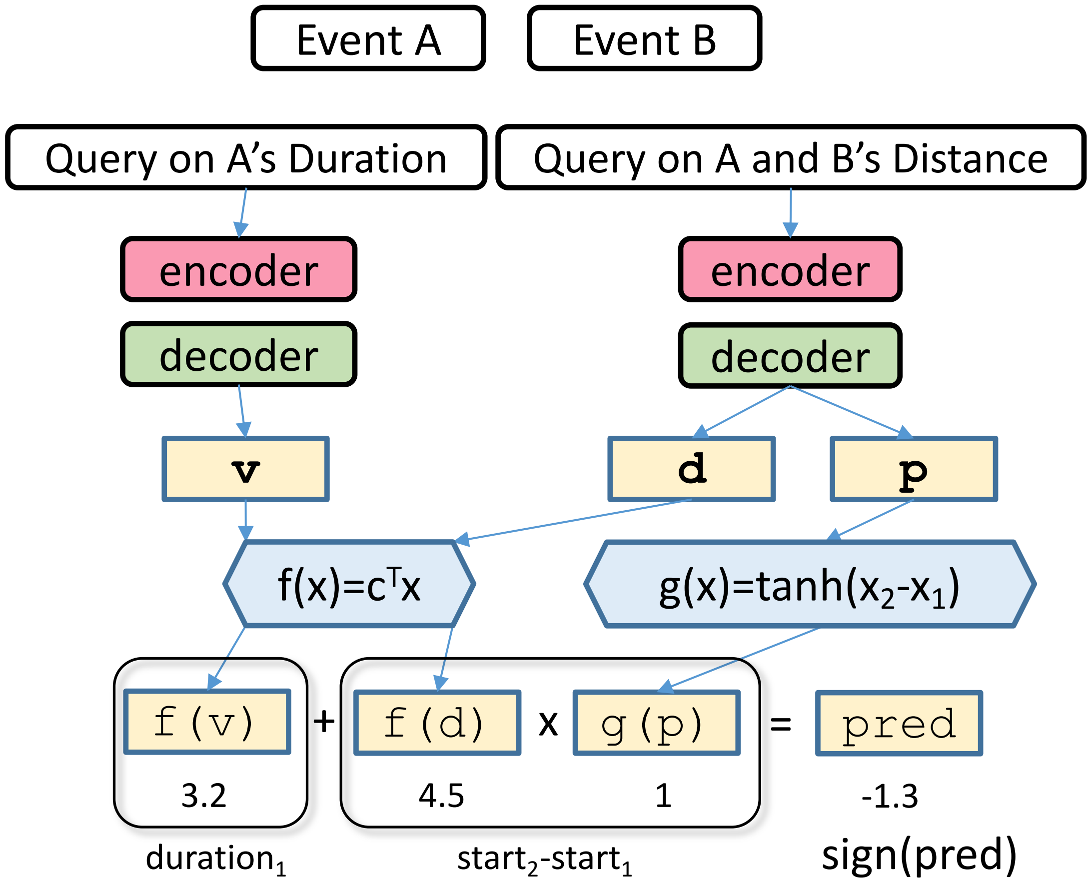
$$\top \rightarrow subevent(e_1, e_2)$$
$$subevent(e_1, e_2) \leftrightarrow superevent(e_2, e_1)$$
$$subevent(e_1, e_2) \wedge after(e_2, e_3) \rightarrow after(e_1, e_3)$$

There are other ways to reason over learned components; quite often these are done via constrained optimization with learned models as **variables.** [Zhou et. al NAACL'21]



Symbolic Computation

end time = start time + duration
Zhou et al. NAACL'21



- Given that End Time is much harder to predicts than Start Time we can also conduct symbolic computation on end times with start times and duration values

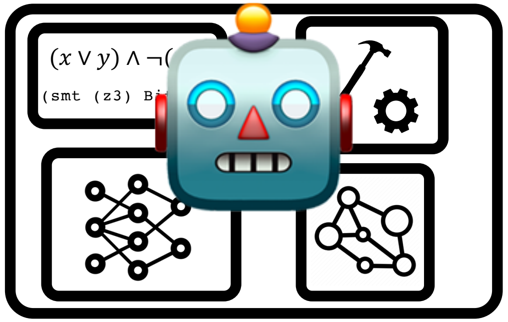
This Tutorial

Integrating
Symbolic Modules, Constraints,
and
Knowledge
Into Neural Language Models

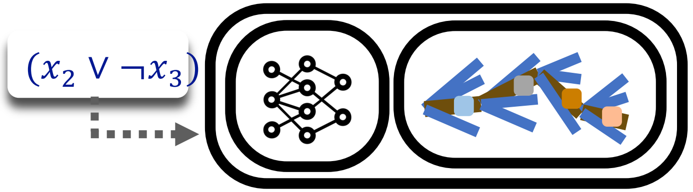
Sean Welleck and Yejin Choi

Framework I

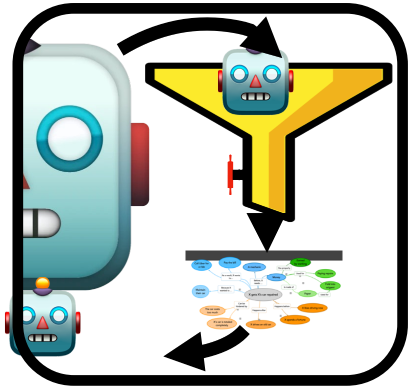
- **Modularity**
  - Single monolithic system ⟶ decomposed neural & symbolic modules



- **Constraints**
  - Discrete logical constraints



- **Knowledge**
  - Hand-crafted ⟶ *generated* and *distilled*



Augmenting Network Architectures and Loss Functions
Using Logic Rules

Vivek Srikumar

Framework II

- Tasks = contracts
  - We want models that do more than what the data says
- Learning from ~~examples~~ Knowledge
  - Relaxing logic and using relaxed logic to learn
- Three case studies

Augmentation Consistency Structured

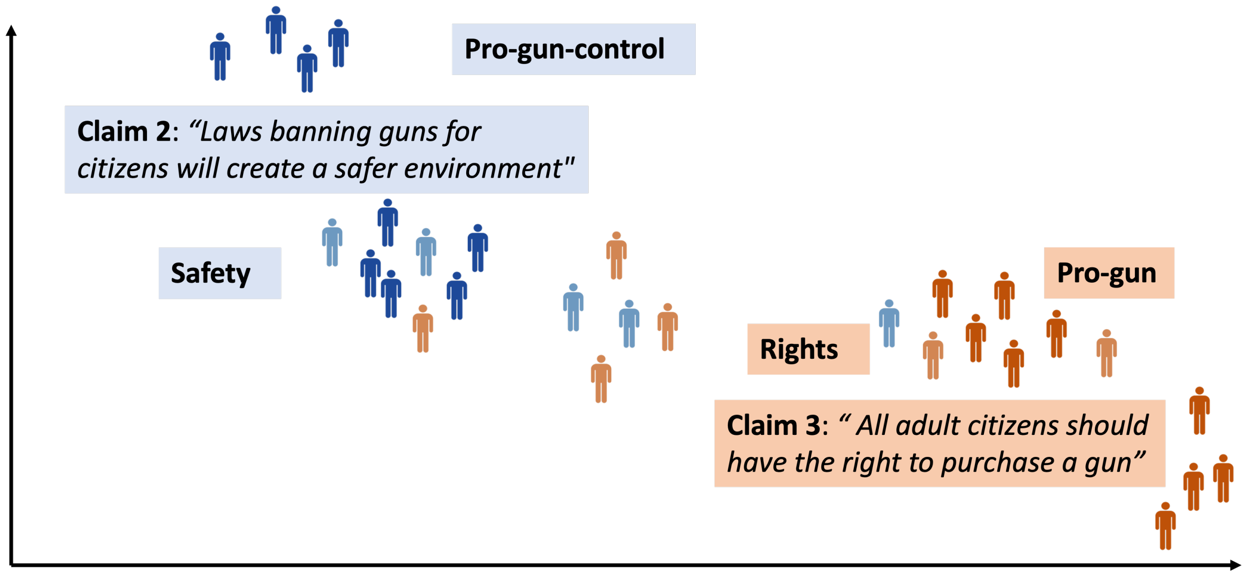
Tuning for SRL

For natural language inference

From Statistical to Deep Relational Learning

- Maria L. Pacheco and Dan Goldwasser

MakesClaim(user, claim) ∧
HasFrame(claim, safety) →
HasStance(user, pro-gun-control)



Framework III

- We know a lot about *Co-ref chains, Dep. Parsing, etc.*
- What about more abstract structures of interest outside of NLP?
  - E.g.: *Parsing the landscape of opinions and perspective*
- Declarative Deep Relational Learning:
  - Rules as context, using a graphical model
  - Representation as context, embedding symbols in a shared space
- Advanced scenarios: explanations, limited supervision, interaction

Tutorial Outline

- Introduction
  - Dan Roth 15 min.
- Algorithmic Frameworks and Applications
  - Framework 1 Sean Welleck & Yejin Choi 35 min.
- Break 15 min
  - Framework 2 Vivek Srikumar 35 min.
  - Framework 3 Maria Pacheco & Dan Goldwasser 35 min.
- Break 15 min
- Challenges and Opportunities
  - Maria Pacheco 10 min.
- Demo
  - Maria Pacheco 20 min.