

# ITERATIVE DOCUMENT-LEVEL INFORMATION EXTRACTION VIA IMITATION LEARNING



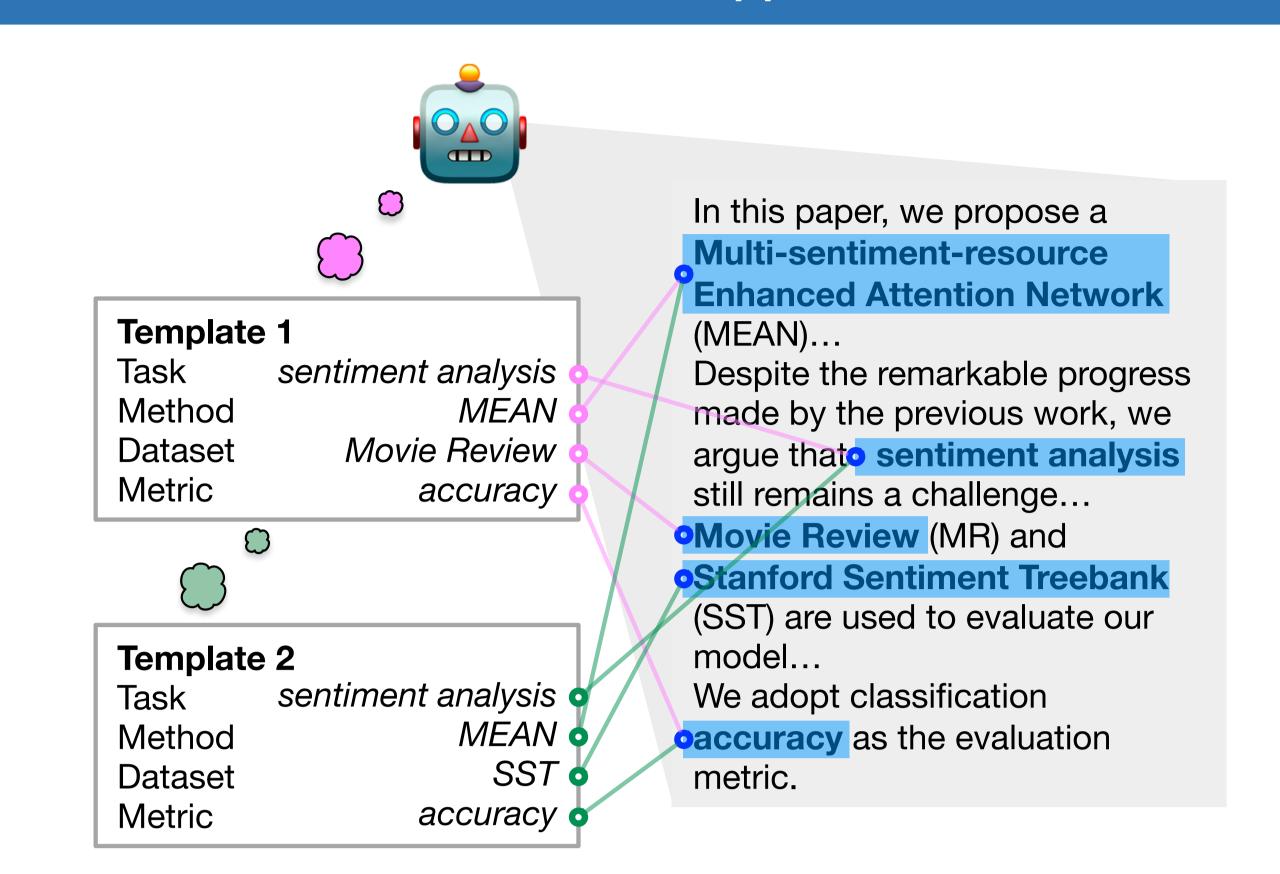
Yunmo Chen<sup>1</sup>, William Gantt<sup>2</sup>, Weiwei Gu<sup>2</sup>, Tongfei Chen<sup>3</sup>, Aaron White<sup>2</sup>, Benjamin Van Durme<sup>1</sup>

<sup>1</sup>Johns Hopkins University, <sup>2</sup>University of Rochester, <sup>3</sup>Microsoft Semantic Machines

#### Summary

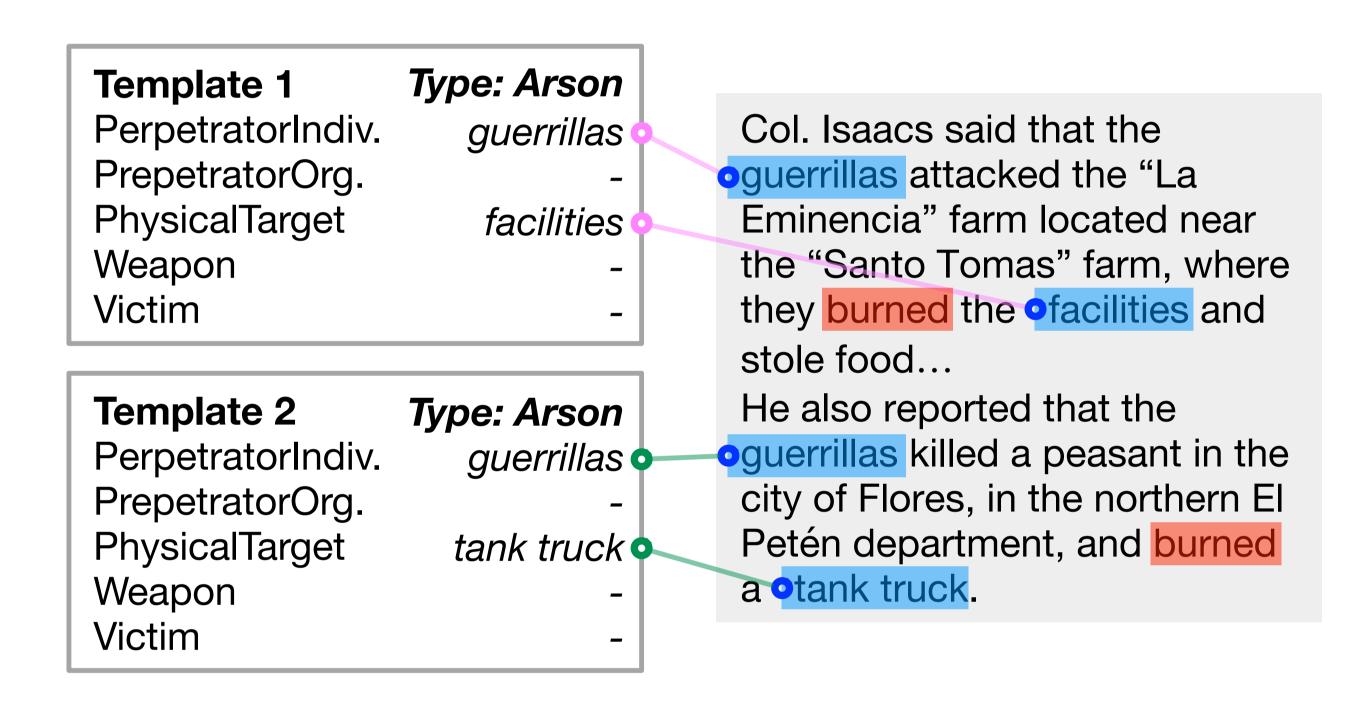
- (Generalized) template extraction is the task of extracting
   dictionary-like structures from text documents
- Template extraction can be formulated as a Markov Decision
   Process (MDP), where an agent iteratively extracts templates
- Imitation learning with an interactive expert can provide concrete (supervised) learning signals using dynamic oracles, enabling effective training of an agent
- The approach outperforms previous models on the SciREX and MUC 4 datasets and presents a strong baseline on the new Granular dataset

### Demo of the Approach



Example from the SciREX dataset

### **Examples of Other Template Extraction Datasets (Tasks)**



Example from the MUC-4 dataset

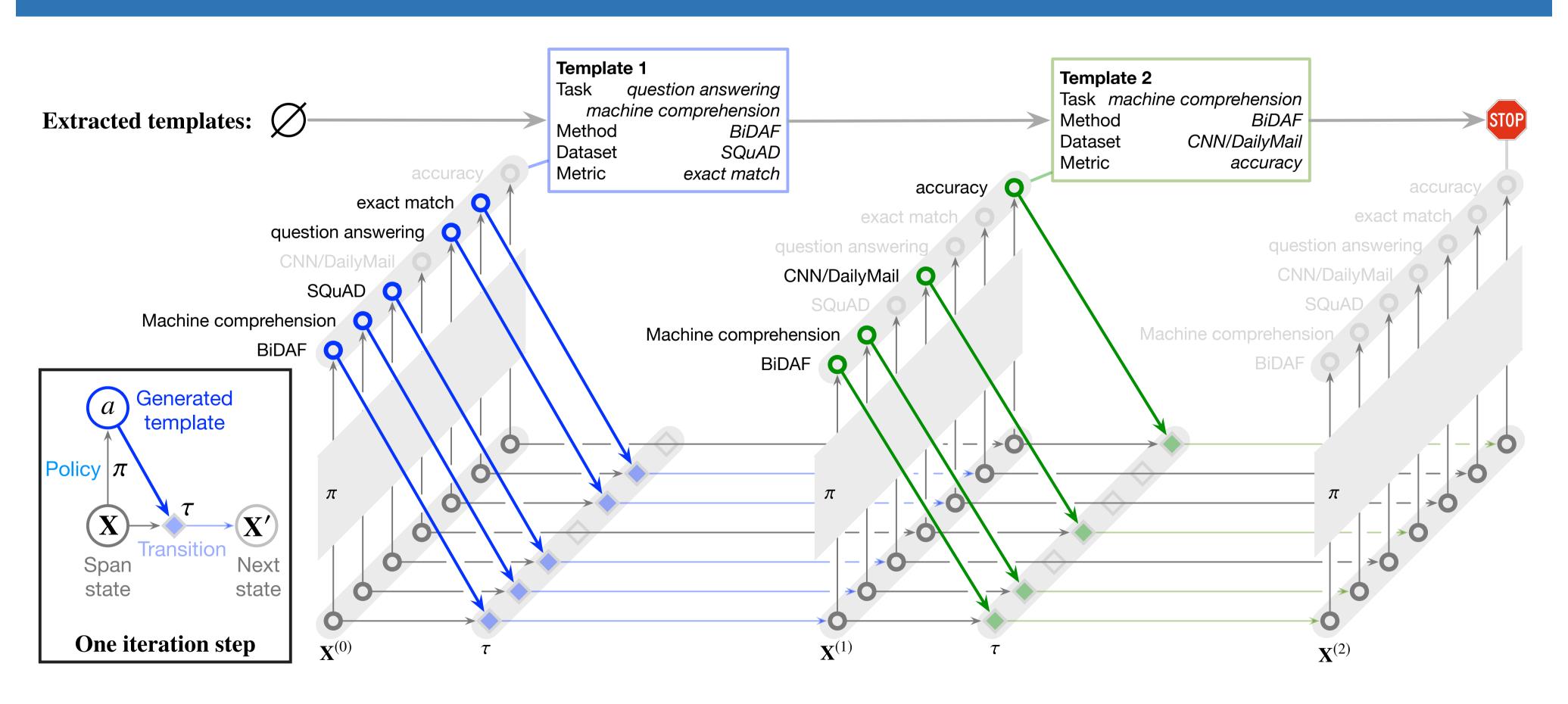
**Template 1** Type: Epidemic **NPI-Events** lockdown enter "wartime status" COVID-19 Disease InfectedCount two people OutbreakEvent pandemic Where Ruili at least eight prefectures 25 counties Type: Displacement **Template 2** Origin Myanmar 🔾 China 👌 Destination Ouctome sentenced to prison 👌 TotalDisplacedCount twenty people 🍼 When OverTime TRUE

In early September, illegal border crossings by two people infected with COVID-19, triggered a week-long lockdown of another Yunnan border city, Ruili, and prompted eat least eight prefectures and 25 counties to enter "wartime status." Following the incident, Yunnan vowed to strengthen border patrols.

Twenty people have been osentenced to prison in Southwest China's Yunnan Province for crimes relating to illegal immigration from oMyanmar to oChina during the oCOVID-19 opandemic, Yunnan's high people's court reported on September 28.

Example from the Granular dataset

## Markov Decision Process



## Metrics

- CEAF-REE is a standard metric for template extraction that measures entity-level slot filler precision, recall, and F1
- Since it is common in template extraction to identify representative mentions (rather than full coreference chains), we also present a mentionlevel variant of this metric called CEAF-RME

