Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning

—— Gavin Xing

Content

- Motivation
- Outline
- MDP
- RL model
- Experiment

- Narasimhan, Karthik, Adam Yala, and Regina Barzilay. "Improving information extraction by acquiring external evidence with reinforcement learning." arXiv preprint arXiv:1603.07954 (2016)

Motivation

ShooterName: Scott Westerhuis
NumKilled: 6

A couple and four children found dead in their burning South Dakota home had been shot in an apparent murder-suicide, officials said Monday.

...

Scott Westerhuis's cause of death was "shotgun wound with manner of death as suspected suicide," it added in a statement.

Example database: shooting incidents

- Difficult to extract such expression
- A large annotated training set may not cover

- Quiz:
  - Can you give an approach to solve the "difficult expression" when extracting information?

Motivation

The **six** members of a South Dakota family found dead in the ruins of their burned home were fatally shot, with one death believed to be a suicide, authorities said Monday.

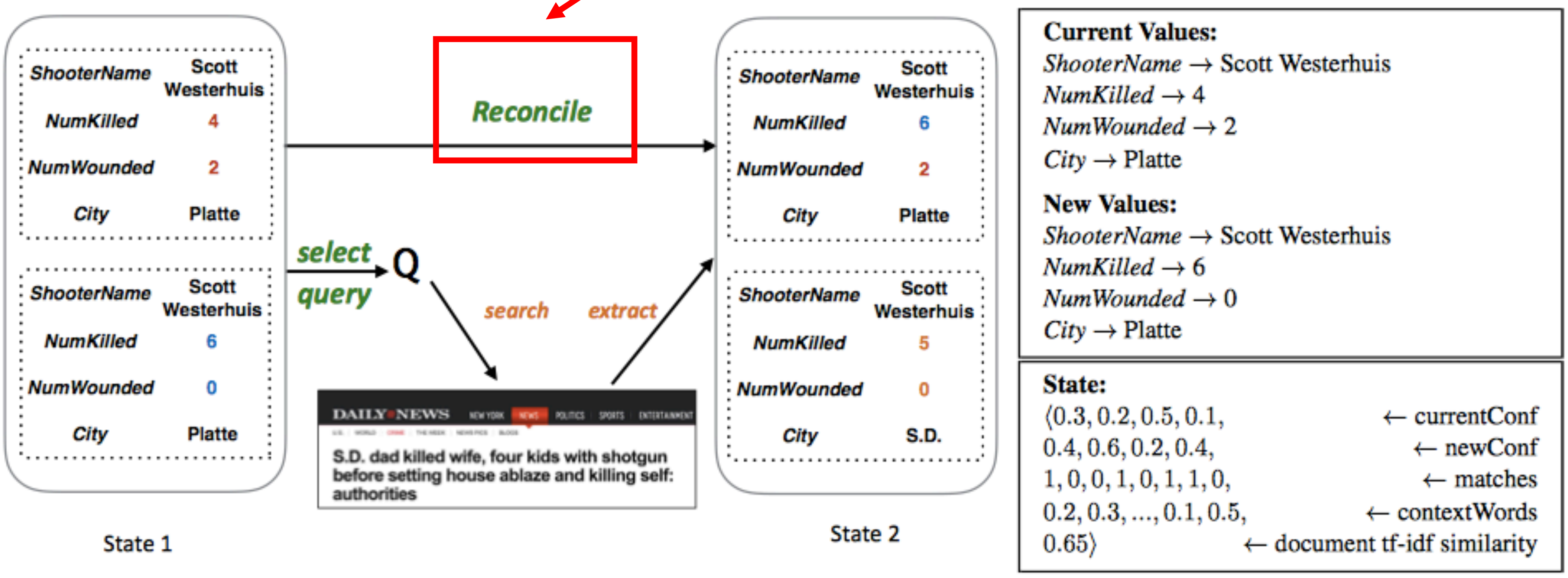
AG Jackley says all evidence supports the story he told based on preliminary findings back in September: **Scott Westerhuis** shot his wife and children with a shotgun, lit his house on fire with an accelerant, then shot himself with his shotgun.

Two other articles on the same shooting case.

- Another 2 articles describing same event
  - The target of extraction is expressed explicitly
- Challenges
  - performing event co-reference
  - reconciling the entities extracted from these different documents

Outline

RL model provides action



FRAMEWORK—MDP

- markov decision process
- Tuple <S, A, T, R>
- S = {s} is the space of all possible states
- A = {a = (d, q)} is the set of all actions
- R(s, a) is the reward function
- T(s 0 |s, a) is the transition function.

MDP—States

- Confidence scores of current and newly extracted entity values.
- One-hot encoding of matches between current and new values.
- Unigram/tf-idf counts of context words.
- tf-idf similarity between the original article and the new article

State:

$\langle 0.3, 0.2, 0.5, 0.1,$	$\leftarrow$ currentConf
$0.4, 0.6, 0.2, 0.4,$	$\leftarrow$ newConf
$1, 0, 0, 1, 0, 1, 1, 0,$	$\leftarrow$ matches
$0.2, 0.3, ..., 0.1, 0.5,$	$\leftarrow$ contextWords
$0.65\rangle$	$\leftarrow$ document tf-idf similarity

MDP—Actions

At each step, the agent is required to take two actions - a reconciliation decision d and a query choice q.

(1) accept a specific entity's value (one action per entity)
(2) accept all entity values
(3) reject all values
(4) stop

MDP—Rewards

- The reward function is chosen to maximize the final extraction accuracy while minimizing the number of queries.
- Embedded in DQN(Deep Q-Network)

$$R(s, a) = \sum_{\text{entity } j} \text{Acc}(e^j_{cur}) - \text{Acc}(e^j_{prev})$$

MDP—Queries

- The queries are based on automatically generated templates

⟨*title*⟩
⟨*title*⟩ + (police | identified | arrested | charged)
⟨*title*⟩ + (killed | shooting | injured | dead | people)
⟨*title*⟩ + (injured | wounded | victim)
⟨*title*⟩ + (city | county | area)

MDP—Transitions

- T(s 0 |s, a) incorporates the reconciliation decision d from the agent in state s along with the values from the next article retrieved using query q and produces the next state s 0 .

Reinforcement Learning

- Utilizes Q(s, a) to determine which action a to perform in state s
- Deep Q-learning model
- Iteratively updates Q(s, a) using the rewards obtained from episodes—R(s, a)

$$Q_{i+1}(s,a) = \mathrm{E}[r + \gamma \max_{a'} Q_i(s',a') \mid s,a]$$

Quiz:

what's the usage of γ (discounting factor) ?

Experiment

Number	Shootings			Adulteration		
	Train	**Test**	**Dev**	**Train**	**Test**	**Dev**
Source articles	306	292	66	292	148	42
Downloaded articles	8201	7904	1628	7686	5333	1537

- Datasets
  - Gun Violence archive
  - Foodshield EMA database

- For each source article ——download top 20 links using the Bing Search API 11 with different automatically generated queries.
- Train source articles -> base extractor
- Train whole articles -> DQN
- maximum entropy classifier as the base extraction system

Experiment—Result

System	Shootings				Adulteration		
	ShooterName	**NumKilled**	**NumWounded**	**City**	**Food**	**Adulterant**	**Location**
CRF extractor	9.5	65.4	64.5	47.9	41.2	28.3	51.7
Maxent extractor	45.2	69.7	68.6	53.7	56.0	52.7	67.8
Confidence Agg. (τ)	45.2 (0.6)	70.3 (0.6)	72.3 (0.6)	55.8 (0.6)	56.0 (0.8)	54.0 (0.8)	69.2 (0.6)
Majority Agg. (τ)	47.6 (0.6)	69.1 (0.9)	68.6 (0.9)	54.7 (0.7)	56.7 (0.5)	50.6 (0.95)	72.0 (0.4)
Meta-classifier	45.2	70.7	68.4	55.3	55.4	52.7	72.0
RL-Basic	45.2	71.2	70.1	54.0	57.0	55.1	76.1
RL-Query (conf)	39.6	66.6	69.4	44.4	39.4	35.9	66.4
RL-Extract	**50.0**	**77.6***	**74.6***	**65.6***	**59.6***	**58.9***	**79.3***
ORACLE	57.1	86.4	83.3	71.8	64.8	60.8	83.9

Accuracy of various baselines (italics), our system (DQN) and the Oracle on Shootings and Adulteration datasets

Experiment—Result

Reconciliation (RL-Extract)	Context	Reward	Accuracy				Steps
			S	K	W	C	
Confidence	tf-idf	Step	47.5	71.5	70.4	60.1	8.4
Majority	tf-idf	Step	43.6	71.8	69.0	59.2	9.9
Replace	*No context*	Step	44.4	77.1	72.5	63.4	8.0
Replace	*Unigram*	Step	48.9	76.8	74.0	63.2	10.0
Replace	tf-idf	*Episode*	42.6	62.3	68.9	52.7	6.8
Replace	tf-idf	Step	**50.0**	**77.6**	**74.6**	**65.6**	9.4

Effect of using ***different reconciliation schemes***, ***context-vectors***, and ***rewards*** in our RL framework (Shootings domain)

Conclusion

- Using external evidence to improve information extraction accuracy for domains with limited access to training data.
- Each step
  - search queries
  - extraction from new sources
  - reconciliation of extracted values
- Using a reinforcement learning framework and learn optimal action sequences to maximize extraction accuracy while penalizing extra effort

Thank You!