Incremental Joint Extraction of Entity Mentions and Relations

Qi Li and Heng Ji

{liq7,jih}@rpi.edu

Rensselaer Polytechnic Institute
End-to-End Relation Extraction

Baltimore is the largest city in the U.S. state of Maryland.
Baseline System

• Typical pipelined approach

The tire maker still employs 1,400

| arg-1 | ORG | tire maker |
|-------|-----|-------------|
| arg-2 | PER | 1,400       |

Entity Mention Boundaries + Types

Relation Extraction

Error Propagation
• Jointly extract and improve both subtasks

the tire maker still employs 1,400

• Exploit global features in the joint search space
Joint Extraction of Entity Mentions and relations

The tire maker still employs 1,400

Joint search space is exponentially large
Global features make inference even harder
Exact inference is expensive

The tire maker still employs 1,400
Learning Framework

- In each training iteration:
  - For each $(x, y) \in$ training set:
    - **Weights update:**
      
      $w \leftarrow w + f(x, y_{1:|z|}) - f(x, z)$

      (Collins and Roark 2004, Huang et al. 2012)
Search Algorithm

• Joint search framework
  o beam search
    • flexible and efficient
  o segment–based decoding
    • “segment” -- subsequence of input sentence
    • each segment is a hypothesis a entity mention or NIL

| The tire maker still employs 1,400 |
|------------------------------------|
| O       B-ORG L-ORG O        O      U-PER |

token-based

vs.

segment-based

| The tire maker still employs 1,400 |
|------------------------------------|
| ORG                 PER              |
Joint Search Algorithm

• Token-based decoder doesn’t work

- unfair to compare mentions with different boundaries
  • Complete mention is biased by the model

- difficult to synchronize relation links
  • (New$_{B-FAC}$ York$_{I-FAC}$) is not yet a complete mention
    no link can be made at this step
Joint Search Algorithm

• Mention-step
  o propose various segments at the current token
  o append to previous assignments
  o get best-k new assignments

...  

O

ORG

PER

The tire maker still employs 1,400
Joint Search Algorithm

• Mention-step
  o propose various segments at the current token
  o append to previous assignments
  o get best-k new assignments

The tire maker still employs 1,400
Joint Search Algorithm

- Mention-step (cont.)
  - propose various segments at the current token
  - append to previous assignments
  - get best-k new assignments

```
The tire maker still employs 1,400 people...
```

Diagram showing token assignment with ORG and PER tags.
Joint Search Algorithm

• Relation-step
  o link each new node to previous ones
  o following type constraints

Prune relations incompatible w/ entity types
Physical, Person-Social are ruled out in this example

The tire maker still employs 1,400

EMP-ORG

O ORG O O

The tire maker still employs

O

1,400

O

O

O

O

O

O
Search Algorithm

- Final structure
  - return top-ranked configuration in the beam
Features

• Segment-based features
  o Based on the entire mention instead of individual tokens
  o Gazetteer features
    • “New York City” is a city
    • “New York” is a state or city
  o Word case features
    • case information about all tokens contained
    • all-capitalized “Lusaka”
    • all-lowercase “magistrate”
    • mixture “Lusaka magistrate” -- a bad mention
Features

• Segment-based features (cont.)
  o Contextual features
    • neighbor unigrams and bigrams
  o Parsing features
    • phrase label of common ancestor (NP)
    • depth of common ancestor (2)
    • whether the segment matches a base phrase (true)
      or is a suffix of a base phrase
    • head word of the segment (maker)
Global Features

• Involve multiple local decisions
  ○ dynamically created during the search
  ○ capture long-distance dependencies

  ○ entity mentions are inter-dependent
  ○ a relation may indicate or contradict other ones
Global Entity Mention Features

• Co-referential mentions should be assigned the same label

thousands of **Muslims** marched to **their** main mosque

the senior Moscow **official, who** was ..
Global Entity Mention Features

• Neighbor entity mentions should have coherent types

\[ \text{pre}_\text{p_from} \]
Barbara Starr was reporting from the Pentagon
“PER–prep_from–PER” will receive negative weights

\[ \text{conj_and} \]
Syria, China and Germany all opposing
“GPE–conj_and–GPE” will receive positive weights
Global Entity Mention Features

• If an entity mention is semantically part of another mention, they should be assigned the same entity type

• Examples:
  o some of Iraq’s exiles
  o one of the town’s two meat-packing plants
  o the rest of America
  o ...

• Part-whole relation is identified by prep_of dependency
Global Entity Mention Features

• Entity role coherence

(\text{PER forces})
\text{EMP-ORG} \quad \text{EMP-ORG}
(GPE Somalia) \quad (GPE US)

\textbf{US forces} in Somalia, Haiti and Kosovo

- entity mentions should play coherent roles
- a person mention is unlikely to have two employers
- a geo-political mention is likely to be physical locations for two other mentions
Global Entity Mention Features

- Penalize triangle structures

US forces in Somalia, Haiti and Kosovo

- multiple entity mentions are unlikely to be fully connected with the same relation type
- triangle structure will be penalized
Global Entity Mention Features

• Dependency compatibility

US forces in Somalia, Haiti and Kosovo

○ two dependent mentions should have compatible relations
Experiments

• Data
  o ACE’05 corpus: exclude genres cts and un
  o ACE’04 corpus: bnews and nwire subsets

| Data Set | # sentences | # mentions | # relations |
|----------|-------------|------------|-------------|
| ACE’05   |             |            |             |
| Train    | 7,273       | 26,470     | 4,779       |
| Dev      | 1,765       | 6,421      | 1,179       |
| Test     | 1,535       | 5,476      | 1,147       |
| ACE’04   | 6,789       | 22,740     | 4,368       |

• Evaluate Metric
  o precision/recall and f-measure for entity mention and relation
  o entity mention + relation: consider entity type
Experiments

- Performance on development set (beam size = 8)

- Global feature improves performance on both tasks
- Set training iteration as 22 for remaining experiments
Experiments

- Overall performance on ACE’05 corpus
Experiments

- Overall performance on ACE’04 corpus
Experiments

- Real Example

| Ranking | Sentence                                      |
|---------|-----------------------------------------------|
| 1       | a marcher from Florida                        |
| 2       | a marcher from Florida per                    |

- the correct hypothesis is ranked lower

Several thousand demonstrators also gathered outside the White House in Washington, accompanied by a major security presence. Bush went to his Camp David retreat for the weekend.

"I'm mourning because the 'shock and awe' started yesterday," Abigail Fletcher, a marcher from Florida, said outside the president's residence.

"They can say they're 'smart bombs,' but smart bombs aren't able to distinguish between military and human targets," she added.
Experiments

- Real Example

| Sentence                          | Ranking |
|----------------------------------|---------|
| a marcher from Florida o o o o    | 1       |
| a marcher from Florida o per o    | 2->4    |

- correct one is ranked lower 😞
Experiments

• Real Example

\[
\begin{align*}
\text{a marcher from Florida} & \quad \text{Ranking} \quad 4->1 \\
\text{o per o gpe} & \\
\text{a marcher from Florida} & \quad \text{1->2} \\
\text{o o o o gpe} & \\
\text{global entity feature of (per-prep_from-gpe)} & \\
\text{pushed the correct assignment to the top 😊}
\end{align*}
\]
Experiments

• Real Example

  a marcher from Florida 1
  o per o gpe

  a marcher from Florida 2->4
  o o o o gpe

• adding relation link makes the margin even larger 😊
Related Work

• ACE Entity Mention and Relation Extraction
  o Florian et al., 2006, Florian et al., 2010, Ohta et al., 2012 etc.
  o Zhou et al., 2007, Jiang & Zhai, 2007, Chan & Roth 2011, etc.
  o Pipelined methods, assumed entity mentions were given

• Joint Inference Methods for IE
  o Re-ranking: Ji & Grishman 2005. Parsing: Kate & Mooney, 2010
  o ILP-inference: Roth & Yih, 2004, Roth & Yih 2007, Yang & Cardie, 2013 etc.
  o Models are separately learned
  o Ours: single model + global features

• Joint Graphical Models
  o Singh et al., 2013, Yu & Lam, 2010 etc.
  o Computationally expensive
Conclusions & Future Work

- jointly model and extract mentions and relations is Possible, Advantageous, and Easy
- global inference is Intuitive and Important
- Future work: incorporate other IE components, such as Event, into the joint framework

The tire maker still employs 1,400.