Incremental Joint Extraction of Entity Mentions and Relations

Qi Li and Heng Ji
{liq7,jih}@rpi.edu

Rensselaer Polytechnic Institute
Baltimore is the largest city in the U.S. state of Maryland.

City to distinguish it from surrounding Baltimore County. Founded seaport in the Mid-Atlantic United States and is situated closer to major seaport on the East Coast.\textsuperscript{[17]} Baltimore's Inner Harbor was for immigrants to the United States and a major manufacturing center. Baltimore shifted to a service-oriented economy, with the Johns Hopkins University serving as the city's top two employers.\textsuperscript{[19]}

Location in the contiguous United States

Coordinates: 39°17′N 76°37′W

- **Country**: United States of America
- **State**: Maryland

- **Founded**: 1729
- **Incorporation**: 1797
- **Named for**: Cecilius Calvert, 2nd Baron Baltimore
Baseline System

• Typical pipelined approach

The tire maker still employs 1,400

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORG</td>
<td>tire maker</td>
</tr>
<tr>
<td>PER</td>
<td>1,400</td>
</tr>
</tbody>
</table>

Error Propagation
Problem Statement

• Jointly extract and improve both subtasks

• Exploit global features in the joint search space
The tire maker still employs 1,400.

**Joint Extraction of Entity Mentions and relations**

Joint search algorithm

The tire maker still employs 1,400.

Joint search space is exponentially large.

Global features make inference even harder.

Exact inference is expensive.

beam

search space

EMP-ORG

ORG

PER
Learning Framework

- In each training iteration:
  - For each \((x, y) \in\) training set:
    - Beam Search
    - 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

<table>
<thead>
<tr>
<th>The tire maker still employs 1,400</th>
</tr>
</thead>
<tbody>
<tr>
<td>O  B-ORG L-ORG O O U-PER</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The tire maker still employs 1,400</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORG ___ PER</td>
</tr>
</tbody>
</table>

token-based vs. segment-based
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
    - \((\text{New}_{B\text{-FAC}} \text{ York}_{I\text{-FAC}})\) 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

...
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.)
  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

• 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

o iteratively update the beam
Search Algorithm

- Final structure

  - return top-ranked configuration in the beam

The tire maker still employs 1,400
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

\[
\text{prep\_from}
\]

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

![Diagram showing entity role coherence]

US forces in Somalia, Haiti and Kosovo

- entity mentions should play coherent roles
- a person mention is unlikely to have two employer
- 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

<table>
<thead>
<tr>
<th>Data Set</th>
<th># sentences</th>
<th># mentions</th>
<th># relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE’05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>7,273</td>
<td>26,470</td>
<td>4,779</td>
</tr>
<tr>
<td>Dev</td>
<td>1,765</td>
<td>6,421</td>
<td>1,179</td>
</tr>
<tr>
<td>Test</td>
<td>1,535</td>
<td>5,476</td>
<td>1,147</td>
</tr>
<tr>
<td>ACE’04</td>
<td>6,789</td>
<td>22,740</td>
<td>4,368</td>
</tr>
</tbody>
</table>

• 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

- Pipeline
- Joint w/ Local
- Joint w/ Global
- Chan & Roth (2011)
Experiments

• Real Example

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a marcher from Florida</td>
</tr>
<tr>
<td>2</td>
<td>a marcher from Florida per</td>
</tr>
</tbody>
</table>

• the correct hypothesis is ranked lower
# Experiments

- **Real Example**

<table>
<thead>
<tr>
<th>a marcher from Florida</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>o o o o o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>a marcher from Florida</td>
<td>2-&gt;4</td>
</tr>
<tr>
<td>o o per o</td>
<td></td>
</tr>
</tbody>
</table>

- correct one is ranked lower 😞
Experiments

• Real Example

\[\text{a marcher from Florida} \quad 4 \rightarrow 1\]
\[\text{o o per o gpe} \]

\[\text{a marcher from Florida} \quad 1 \rightarrow 2\]
\[\text{o o o o gpe} \]

• global entity feature of (\text{per-prep_from-gpe}) pushed the correct assignment to the top 😊
Experiments

• Real Example

- a marcher from Florida
  - o per o gpe
  - a marcher from Florida
  - 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 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