Learning Latent Forests for Medical Relation Extraction

Zhijiang Guo*, Guoshun Nan*, Wei Lu and Shay Cohen
Medical Relation Extraction

• Input Sentence

Crystallographic analysis of the human *phenylalanine hydroxylase* catalytic domain with bound *catechol* inhibitors at 2.0 Å resolution.

• Output Relation

*Down regulator*
Related Work

• Sequence-based Methods

• Dependency-based Methods
Related Work

• Sequence-based Method

Operates only on the text sequences, **RNN** (Liu et al., 2017) or **Transformer** (Verga et al., 2018) is used as the *sequence encoder*. 
Related Work

• Dependency-based Method

Incorporates *dependency structures* into the model, **Graph LSTM** (Peng et al., 2017) or **GRN** (Song et al., 2019) is used as the encoder.
Related Work

• Dependency-based Method

Dependency-based models are able to *capture non-local relations* that are obscure from the surface form alone (Zhang et al., 2018).
Related Work

- Dependency Tree

**Graph LSTM** (Peng et al., 2017) is used to encode the *1-best dependency tree* generated by an off-the-shelf parser.

Crystallographic analysis ... phenylalanine hydroxylase catalytic domain ... catechol inhibitors ...
For the medical domain, *parsing accuracies drop significantly* (Lease and Charniak, 2005), which leads to *error propagation*.
Recent works (Song et al., 2019; Jin et al., 2020) use dependency forests that consist of arcs that a parser is relatively confident about.
Related Work

• Dependency Forest Construction

*Edgewise* (Song et al., 2019): keeps all the edges with *scores greater than a threshold*. Scores are given by a parser trained on *news domain*. 
Related Work

• Dependency Forest Construction

*KBests* (Song et al., 2019): finds *K highest-scored trees*, then merge them by *combining identical dependency edges* to make the forest.
Related Work

• Dependency Forest Construction

*FullForest* (Jin et al., 2020): defines the forest as a **3-dimensional tensor**.
Related Work

- Dependency Forest Construction

**3D Tensor**: each point representing the conditional probability $p(w_j, l|w_i)$ of one word $w_i$ modifying another word $w_j$ with a relation $l$. 

Related Work

• Dependency Forest Construction

**3D Tensor**: use the dependency parser trained on the news domain to obtain the tensor and *adjust the parser by relation extraction loss*. 

Deep bi-affine parser (Dozat and Manning, 2017)
Motivation

• **Research Question**: Can we build a model without relying on an out-of-domain parser?
Motivation

• Yes! We treat the dependency structure as a latent variable and induce it end-to-end.
Forest Inducer

- **Multi-head attention** is used to construct $N$ attention matrices, which will be fed into the next module to obtain $N$ latent forests in order to capture different dependencies in different representation subspaces.
• **Matrix-Tree Theorem** (Tutte, 1984; Koo et al., 2007) is used to induce $N$ latent structures, i.e., edge-weighted dependency forests.
Model

Forest Encoder

Classifier

Graph Neural Networks

Forest Inducer

Sentence

Matrix-Tree Theorem

Multi-Head Attention
Forest Encoder

- Given the generated latent forests, we use the \textit{densely-connected graph convolutional networks} (Guo et al., 2019) as the encoder.
## Experiments

- **Data Statistics**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Relations</th>
<th>Domain</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGR</td>
<td>2</td>
<td>Medical</td>
<td>11,780</td>
<td>N/A</td>
<td>219</td>
</tr>
<tr>
<td>CPR</td>
<td>5</td>
<td>Medical</td>
<td>16,107</td>
<td>10,030</td>
<td>14,269</td>
</tr>
<tr>
<td>Semeval</td>
<td>10</td>
<td>News</td>
<td>10,717</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>10-task8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiments

- Results on medical dataset: PGR
Experiments

- Results on medical dataset: CPR
# Experiments

- Results on news dataset: Semeval-2010 task 8

<table>
<thead>
<tr>
<th>Type</th>
<th>Model</th>
<th>w Parser</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree</td>
<td>Song et al., 2019</td>
<td>Y</td>
<td>84.6</td>
</tr>
<tr>
<td>Tree</td>
<td>Zhang et al., 2018</td>
<td>Y</td>
<td>84.8</td>
</tr>
<tr>
<td>Forest</td>
<td>Jin et al., 2020</td>
<td>Y</td>
<td>85.7</td>
</tr>
<tr>
<td>Forest</td>
<td>Guo et al., 2019</td>
<td>Y</td>
<td>85.7</td>
</tr>
<tr>
<td>Forest</td>
<td>Song et al., 2019</td>
<td>Y</td>
<td>86.3</td>
</tr>
<tr>
<td>Forest</td>
<td>Ours</td>
<td>N</td>
<td>85.7</td>
</tr>
</tbody>
</table>
Conclusion

• Instead of using an *out-of-domain* parser, we treat the *dependency structure* as a *latent variable* and induce the *task-specific forests*.

• Extensive experiments show that our proposed method is able to achieve state-of-the-art results on *medical relation extraction* tasks.
Thank you!

Code Available:
https://github.com/Cartus/Latent-Forests