Attention-based Neural Networks for Chemical Protein Relation Extraction

Sijia Liu¹,², Feichen Shen, Yanshan Wang¹, Majid Rastegar-Mojarad¹,³, Ravikumar Komandur Elayavilli¹, Vipin Chaudhary², Hongfang Liu¹

¹ Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
² Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA
³ University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Oct. 20th, 2017
Outline

• Background
• Materials
• Methods
• Results
• Conclusion
Background

• Need for automatic data curation from biomedical literatures
• Recent related share tasks
  • ScienceIE
    • SemEval 2017 Task 10
    • Material/Process/Task from scientific literatures
    • Hyponym and synonym between entities
  • Chemical Disease Relation (CDR)
    • Biocreative V Track 3 (2015)
    • Chemical and disease
  • DDI-Extraction
    • SemEval 2011 and 2013
    • Drug-Drug interaction from knowledge base and literatures
• Neural models for relation extraction tasks
  • CNN, RNN (LSTM, GRU)
  • Attention mechanism
Outline

• Introduction
• Materials
• Methods
• Results
• Conclusion
Materials

- In this study, we used CHEMPROT dataset provided by task organizers
- Corpus statistics
  - 4966 PubMed abstracts
  - 126,457 annotated entities (chemical, gene and protein)
  - 6573 positive relations (training + development)

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of docs</th>
<th>Average # of entities</th>
<th># of positive relations *</th>
<th># of all potential relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1020</td>
<td>25.247</td>
<td>4157</td>
<td>15842</td>
</tr>
<tr>
<td>Development</td>
<td>612</td>
<td>25.436</td>
<td>2416</td>
<td>9759</td>
</tr>
<tr>
<td>Test</td>
<td>3334</td>
<td>25.536</td>
<td>-</td>
<td>53457</td>
</tr>
</tbody>
</table>

* CPR 3, 4, 5, 6 and 9
Outline

• Introduction
• Materials
• Methods
• Results
• Conclusion
Methods

- We used deep neural networks to classify relation instances of entities into relation types
  - Entities are provided by the task organizers
- As a relation classification problem
  - NA and other relation labels
  - Only evaluated relations
- We only consider the chemical-protein relation pairs within one sentence.
  - Few positive relations across of sentence boundary
Relation Instance Generation

- Enumerate all potential chemical protein pair in the sentence
  - Gold standard labels for positive pairs
  - “NA” for negative pairs (including CPR 1, 2, 7, 8 and 10)

- Example:

  “Here, we compared the effects of a dimeric [**PSD-95**] inhibitor, [**UCCB01-125**], and the [**NMDAR**] antagonist, [**MK-801**], …”

<table>
<thead>
<tr>
<th>Label</th>
<th>Entity 1</th>
<th>Entity 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR:4</td>
<td>T13</td>
<td>T38</td>
</tr>
<tr>
<td>NA</td>
<td>T13</td>
<td>T39</td>
</tr>
<tr>
<td>NA</td>
<td>T14</td>
<td>T38</td>
</tr>
<tr>
<td>CPR:6</td>
<td>T14</td>
<td>T39</td>
</tr>
</tbody>
</table>
Input Representation

- Vector representation for neural network models
- Each relation is
  - Sentence consists of words and entities
- Words
  - Word embeddings
  - 300-dimensional Glove-6B
- Entities
  - Position embeddings follows Zeng et. al
  - Relative distance with a constant shift as index
  - Position embeddings are trained jointly
  - Index example:

[Cyclopentenone prostaglandins] were potent inhibitors of [iNOS] induction …

[e1]
[e2]

1 https://nlp.stanford.edu/projects/glove/
Input Representation (cont.)

- Weakness of previous representation
  - WE of out-of-vocabulary tokens
  - Difficult to handle phrases
- Solution
  - Replace tokens with the annotated entity type
    - CHEMICAL to “chemical”
    - GENE-N, GENE-Y to “gene”

\[
\begin{align*}
3 + 25 &= 28 \\
-2 + 25 &= 23 \\
\text{[Chemical]} &\text{ were potent inhibitors of [gene] induction} \ldots \\
\end{align*}
\]
Convolutional Neural Networks

Relation Prediction

Non-linear Layer

Global Max Pooling

Convolution

Word Embedding

Pst Embedding 1

Pst Embedding 2

$argmax$
RNN for Relation Extraction

Non-linear Layer

Flatten

1 by (Sent len * RNN dim)

argmax

Relation Label

Word Embedding

Pos Embedding 1

Pos Embedding 2

RNN → RNN → RNN → ... → RNN
RNN for Relation Extraction

Non-linear Layer

Flatten

Hard to handle

Word Embedding

Pos Embedding 1

Pos Embedding 2

1 by (Sent len * RNN dim)

argmax

Relation Label

RNN → RNN → RNN → ... → RNN
Attention Mechanism

• Aims to emphasize the contribution of the informative neural units
• Has been applied to multiple NLP tasks
  • Machine translation
  • Question answering
  • Relation extraction
• Additional attention layer on top of RNN
  • Is a relational sentence encoder for relation classification
  • Overlooks all the RNN units of the sequence
  • Assigns attention weights according to the importance
Attention based RNN

Non-linear Layer

Attention Weights

Word Embedding
Pos Embedding 1
Pos Embedding 2

RNN

RNN

RNN

…

RNN

argmax

Weighted sum (1 by RNN dim)

Relation Label
Attention Layer

Hidden representation of RNN units:
\[ u_t = \tanh(W_wh_t + b_w) \]

Attention weights:
\[ \alpha_t = \frac{\exp(u_t^Tu_w)}{\sum_t \exp(u_t^Tu_w)} \]

Relation representation:
\[ s = \sum_t \alpha_t h_t \]

Where:
- \( t \): token index in the sentence
- \( W_w, b_w \): trainable weights and bias
- \( h_t \): RNN output
- \( u_w \): trainable word context vector
Outline

• Introduction
• Materials
• Methods
• Results
• Conclusion
Experiment Settings

- Implementation
  - Punkt sentence detector in NLTK
  - Keras 2.0.5 with Tensorflow backend

- Details
  - Fixed sentence length of 170 (zero padding or trimming)
  - Adam optimizer
  - Sparse categorical cross entropy as loss function
  - Dropout rate of 0.5
  - Class weight of 5.0 for positive labels to balance the precision and recall

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch size</td>
<td>64</td>
</tr>
<tr>
<td>Number of CNN filters</td>
<td>100</td>
</tr>
<tr>
<td>Filter length</td>
<td>3</td>
</tr>
<tr>
<td>RNN dimension</td>
<td>128</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
</tbody>
</table>

1 https://gist.github.com/cbaziotis/7ef97ccf71cbc14366835198c09809d2
# Results of Submitted Runs

<table>
<thead>
<tr>
<th>ID</th>
<th>Model</th>
<th>Development Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>1</td>
<td>CNN token</td>
<td>0.459</td>
<td>0.456</td>
</tr>
<tr>
<td>3</td>
<td>CNN entity</td>
<td>0.497</td>
<td>0.448</td>
</tr>
<tr>
<td>2</td>
<td>ATT GRU token</td>
<td>0.470</td>
<td>0.522</td>
</tr>
<tr>
<td>4</td>
<td>ATT GRU entity</td>
<td>0.512</td>
<td>0.501</td>
</tr>
</tbody>
</table>
Results: Relation Types (Development Set)

<table>
<thead>
<tr>
<th>Label</th>
<th>Support *</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR:3</td>
<td>498</td>
<td>0.473</td>
<td>0.388</td>
<td>0.426</td>
</tr>
<tr>
<td>CPR:4</td>
<td>990</td>
<td>0.569</td>
<td>0.663</td>
<td>0.613</td>
</tr>
<tr>
<td>CPR:5</td>
<td>112</td>
<td>0.357</td>
<td>0.634</td>
<td>0.457</td>
</tr>
<tr>
<td>CPR:6</td>
<td>184</td>
<td>0.505</td>
<td>0.609</td>
<td>0.552</td>
</tr>
<tr>
<td>CPR:9</td>
<td>407</td>
<td>0.468</td>
<td>0.442</td>
<td>0.455</td>
</tr>
<tr>
<td>Total *</td>
<td>2191</td>
<td>0.512</td>
<td>0.553</td>
<td>0.528</td>
</tr>
</tbody>
</table>

Note:
- “Total” is weighted F1-score by support counts
  - Different from but proportional to the micro-F1 score
  - Informative in the multiclass classification context
- Support numbers vary from gold standard annotation counts
  - Cross-sentence relations
  - Class with multiple labels
Outline

• Introduction
• Materials
• Methods
• Results
• Conclusion
Conclusion

• We developed deep neural network models
  • Word embeddings and position embeddings
  • Raw token vs. entity label
  • CNN
  • Attention-based RNN

• Attention-based GRU using entity labels is our best model
  • Micro-F1 score of 0.494 on test set
Future Work

• Our future work includes
  • Use external knowledge base
  • Use the token-level weights from the attention activations
    • Pattern mining
    • Cue generation
Acknowledgement

• We thank the support of
  • NIH R01 LM011829
  • NSF IPA

• Opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health and National Science Foundation.
Thank you!