Relation Extraction with Temporal Reasoning Based on Memory Augmented Distant Supervision

Jianhao Yan
Tsinghua University
Background

- Distantly Supervised (DS) Relation Extraction
Motivation

**Jolie** had fallen in love with **Pitt** during filming of *Mr. & Mrs. Smith* (2005)

**Pitt** and **Jolie** announced their engagement in April 2012 after seven years together.

**Jolie** and **Pitt** were married on August 23, 2014, in a private ceremony in Château Miraval, France.

On September 19, 2016, **Jolie** filed for divorce from **Pitt**, citing irreconcilable differences.
Motivation

**Dating**

Jolie had fallen in love with Pitt during filming of *Mr. & Mrs. Smith* (2005)

**Engagement**

Pitt and Jolie announced their engagement in April 2012 after seven years together.

**Marriage**

Jolie and Pitt were married on August 23, 2014, in a private ceremony in Château Miraval, France.

**Divorce**

On September 19, 2016, Jolie filed for divorce from Pitt, citing irreconcilable differences.
Motivation

Jolie had fallen in love with Pitt during filming of Mr. & Mrs. Smith (2005). Pitt and Jolie announced their engagement in April 2012 after seven years together. Jolie and Pitt were married on August 23, 2014, in a private ceremony in Château Miraval, France.

On September 19, 2016, Jolie filed for divorce from Pitt, citing irreconcilable differences.

Model has no clue how to predict relation !!!
Jolie had fallen in love with Pitt during filming of *Mr. & Mrs. Smith* (2005).

Pitt and Jolie announced their engagement in April 2012 after seven years together.

Jolie and Pitt were married on August 23, 2014, in a private ceremony in Château Miraval, France.

On September 19, 2016, Jolie filed for divorce from Pitt, citing irreconcilable differences.
Task Definition

• Introduce temporal information in both the relation labeling and the mention set.

Traditional DS:

$$P(r \mid S = \{s_1, s_2, \ldots, s_T\})$$

DS with Temporal Reasoning:

$$P(r_{t_i} \mid S = \{(s_1, t_1), (s_2, t_2), \ldots, (s_T, t_T)\}, t_i)$$

Where $t_1 \leq t_2 \leq \ldots \leq t_T$
Task Definition

- Introduce temporal information in both the relation labeling and the mention set.

Traditional DS:
\[ P(r \mid S = \{s_1, s_2, \ldots, s_T\}) \]

DS with Temporal Reasoning:
\[ P(r_{t_i} \mid S = \{(s_1, t_1), (s_2, t_2), \ldots, (s_T, t_T)\}) \]
Where \( t_1 \leq t_2 \leq \ldots \leq t_T \)

New dataset WIKI-TIME!
Method

- The task is similar to the classical sequence labeling problem.

So LSTM or LSTM + CRF?
Method

- The task is similar to the classical sequence labeling task.

- So LSTM or LSTM + CRF?

- No! Noisy input will destroy sequence model!
Method

$$P(r_t | S = \{(s_1, t_1), (s_2, t_2), \ldots, (s_T, t_T)\}, t_i)$$

Where $$t_1 \leq t_2 \leq \ldots \leq t_T$$

**Seq**

**Our**

mem - mem - mem - mem - mem
Motivation

Sentence-level memory nets.

- Make use of *supporting instances*.
  - In traditional DS, models always use attention and other techniques to denoise.
  - However, there are sentences that are not direct positive examples for the given relation, but can provide supporting evidence.
Method

Alleviate the **hard** order dependency effect in sequence model.

- Construct a query sequence on each mention time spot of mention set.
- Use Memory Nets to introduce sentence-level temporal reasoning.
Method

Alleviate the hard order dependency effect in sequence model.

- **Construct a query sequence on each mention time spot of mention set.**
- **Use Memory Nets to introduce sentence-level temporal reasoning.**
Method

For $h \in [1, H]$

$K_j \quad p_{ij} \quad V_j$

Inner Product

Relation Embedding

$Pr$

$C1$

Conv

Jobs and Wozniak co-founded Apple in 1976 to ...

Word PF

Convolution Features

Temporal Encoding

Memory: \{m_1, m_2, ..., m_T\}
Method
Temporal Encoding (TE)

★ Several constraints for TE.

- Should comply with the chronological order of instances.
- Encoding similarity is only decided by the difference between two time spots.

\[
PE(j) = \begin{cases} 
\sin(j/10000^{d/d_m}) & \text{if } d \% 2 = 0 \\
\cos(j/10000^{(d-1)/d_m}) & \text{if } d \% 2 = 1 
\end{cases}
\]  

[Vaswani et al., 2017]
Query Construction

- 4 key variables for RE. \((relation, e_1, e_2, t)\)
Query Construction

- 4 key variables for RE. \((relation, e_1, e_2, t)\)

- We split each query into content and temporal encoding.

\[
q_r = R_r + (E_{e_1} + E_{e_2}) \ast \Phi_q
\]

\[
q_{r,i} = [q_r; \lambda \cdot TE(i)]
\]
Method

Alleviate the hard order dependency effect in sequence model.

- Construct a query sequence on each mention time spot of mention set.
- Use Memory Nets to introduce sentence-level temporal reasoning.
Memory Encoding

Jobs and Wozniak co-founded Apple in 1976 to...

Word PF

Convolution Features

max(C31) max(C32) max(C33)

Memory: \{m_1, m_2, ..., m_T\}

Temporal Encoding
Method

[Sukhbaatar et al., 2015]
Optimization

- Query-level Cross Entropy by SGD.

\[
J(\theta) = \sum_{s=1}^{N_s} \sum_{i=1}^{T} y_t \cdot \log p(\hat{y}_t|S_s, \theta, t_i)
\]
Experiment Results (WIKI-TIME)

PR curves on WIKI-TIME

<table>
<thead>
<tr>
<th>Method</th>
<th>P@N_100</th>
<th>P@N_200</th>
<th>P@N_300</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN_ATT</td>
<td>67.33</td>
<td>67.66</td>
<td>66.45</td>
</tr>
<tr>
<td>CNN_ONE</td>
<td>70.3</td>
<td>68.66</td>
<td>65.78</td>
</tr>
<tr>
<td>TempMEM</td>
<td>81.18</td>
<td>82.09</td>
<td>78.41</td>
</tr>
<tr>
<td>TempMEM+R</td>
<td>79.21</td>
<td>78.61</td>
<td>75.42</td>
</tr>
<tr>
<td>TempMEM+P</td>
<td>81.19</td>
<td>79.1</td>
<td>77.41</td>
</tr>
</tbody>
</table>

Automatic results on WIKI-TIME

<table>
<thead>
<tr>
<th>Method</th>
<th>Bag-level F1</th>
<th>Query-level F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN_ATT</td>
<td>39.66</td>
<td>-</td>
</tr>
<tr>
<td>CNN_ONE</td>
<td>40.15</td>
<td>-</td>
</tr>
<tr>
<td>TempMEM</td>
<td>47.88</td>
<td>54.75</td>
</tr>
<tr>
<td>TempMEM+R</td>
<td>46.76</td>
<td>47.83</td>
</tr>
<tr>
<td>TempMEM+P</td>
<td><strong>54.86</strong></td>
<td><strong>60.01</strong></td>
</tr>
</tbody>
</table>

Manual results on WIKI-TIME
Experiment Results (NYT-10)

NO TEMPORAL INFO!

PR curves with CNN

PR curves with PCNN
To Sum Up

- Temporal reasoning task
  - Newly developed dataset WIKI-TIME

- TempMEM model with temporal encoding and sentence-level reasoning

- Experimental results on WIKI-TIME & NYT-10 prove our model achieves better performance.
Thanks.
elliottyan37@gmail.com