1. Motivation

The Entity and Relation Extraction (ERE) task can be decomposed into two subtasks: named entity recognition (NER) and relation extraction (RE).

Two typical approaches of ERE are:

- Pipeline of NER model and RE model.
- Joint model of these two subtasks.

In common sense, joint models are better than pipeline approaches, but recent work do not agree with that.

We design eight pipeline and joint approaches with similar settings to fairly compare them.

2. Approaches design

<table>
<thead>
<tr>
<th>NER</th>
<th>RE</th>
<th>relation</th>
<th>relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>Ei</td>
<td>Ec</td>
<td>Ri</td>
</tr>
<tr>
<td>a2</td>
<td>Ei</td>
<td>Ec</td>
<td>Ri</td>
</tr>
<tr>
<td>a3</td>
<td>Ner</td>
<td>Ner</td>
<td>Ri</td>
</tr>
<tr>
<td>a4</td>
<td>Ner</td>
<td>EcRi</td>
<td>Ri</td>
</tr>
<tr>
<td>a5</td>
<td>Ei</td>
<td>EcRi</td>
<td>Ri</td>
</tr>
<tr>
<td>a6</td>
<td>Ei</td>
<td>EcRi</td>
<td>Ri</td>
</tr>
<tr>
<td>a7*</td>
<td>(with span pruner) EcRi</td>
<td>Ri</td>
<td>Rc</td>
</tr>
<tr>
<td>a8*</td>
<td>(with span pruner) NerRe</td>
<td>Ri</td>
<td>Rc</td>
</tr>
</tbody>
</table>

3. Our methods

Single task modules

With the span or span pair representations, we score the span \( s_i \) or span pair \((s_i, s_j)\) for each label.

\[
g_i = \text{Linear}_\text{ent}(h_i), \quad g_{ij} = \text{Linear}_\text{rel}(h_{ij})
\]

The predication is the label with the largest score.

Cross task modules

Unary score The unary score for span or span pair captures the prior distribution information and is computed solely based on the feature of the variable.

Ternary score It is defined cover a span pair that captures the three-way correlation between their entity labels and the label of the relation between them.

High order inference: Mean-field Variational Inference

With the unary scores and ternary scores of two spans \( s_i, s_j \) and the span pair \((s_i, s_j)\), we use Mean-field Variational Inference (MFVI) to get a factorized variational distribution \( Q \) to approximate the posterior label distributions.

Update message delivered for entity than relation types

\[
F_j^1(a) = \sum_{j} \sum_b Q_j^1(b) \sum_r (Q_j^1(r) f_j(a,b,r) + Q_j^1(r) f_j(a,b,r))
\]

Update the posterior distributions with the messages.

\[
Q_j^1(e) \propto \exp(g_j(e) + F_j^1(e))
\]

\[
Q_j^1(r) \propto \exp(g_j(r) + F_j^1(r))
\]

4. Experiment results

5. Analysis

- Different modules in an approach are pipelined with no shared parameters.
- Cross task modules share encoders and use high-order inference across sub-tasks from NER and RE.