Multi-Task Transfer Learning for Fine-Grained Named Entity Recognition

Masato Hagiwara\textsuperscript{1}, Ryuji Tamaki\textsuperscript{2}, Ikuya Yamada\textsuperscript{2}

\textsuperscript{1}Octanove Labs \hspace{1cm} \textsuperscript{2}Studio Ousia
Named Entity Recognition (NER)

- Few systems deal with more than 100+ types
  - cf. FIGER 112 types (Ling and Weld, 2012)
- Entity typing
  - (Ren et al., 2016), (Shimaoka et al., 2016), (Yogatama et al., 2015)

Can we solve NER (detection and classification) with 7,000+ types in a generic fashion?
Challenge 1: Lack of Training Data

Lack of NER datasets annotated with AIDA

Silver-standard dataset with YAGO annotations

Transfer learning to AIDA
Challenge 2: Large Tag Set

Cost of CRF = $O(n^2)$ (n = # of types)
Challenge 3: Ambiguity in Types

House103544360 vs House107971449

WorldOrganization108294696 vs Alliance108293982

Plaza108619795 vs Plaza103965456

Hierarchical Multi-label Classification

The Statue of Liberty in New York
Challenge 4: Hierarchical Types

Hierarchy-aware soft loss
Hierarchy-Aware Soft Loss

Type confusion weight $W$

Cross entropy loss

Soft GOLD Labels

GOLD

PRED
Experiments

Datasets

1) Pre-training
   OntoNotes 5.0 (subset) for detection
   Silver-standard Wikipedia for classification
   Manually-annotated subset for dev.

2) Fine-tuning
   Manually-annotated Wikipedia
   Manually-fixed AIDA sample data (LDC2019E04)
   Manually-annotated OntoNotes 5.0 (subset)

Settings

- Embeddings
  bert-base-cased
  2-layer BiLSTM (200 hidden units)

- Type conversion
  2-layer feed-forward with ReLU

- Optimization
  Adam (lr = 0.001) for pre-training
  BertAdam (lr = 1e-5 with 2,500 warm-up)
Results

### Performance on validation set

<table>
<thead>
<tr>
<th>Method</th>
<th>Prec</th>
<th>Rec</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>0.45</td>
<td>0.42</td>
<td>0.43</td>
</tr>
<tr>
<td>Fine-tuned</td>
<td>0.65</td>
<td>0.57</td>
<td>0.61</td>
</tr>
<tr>
<td>Fine-tuned w/o loss</td>
<td>0.60</td>
<td>0.50</td>
<td>0.55</td>
</tr>
</tbody>
</table>

### Performance on test set

<table>
<thead>
<tr>
<th>Run</th>
<th>Prec</th>
<th>Rec</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st submission</td>
<td>0.504</td>
<td>0.468</td>
<td>0.485</td>
</tr>
<tr>
<td>After feedback</td>
<td>0.506</td>
<td>0.493</td>
<td>0.499</td>
</tr>
</tbody>
</table>
Error Analysis

- **Location vs GPE**
  - “Southern Maryland”
    - OK: `loc.position.region`, NG: `gpe.provincestate.provincestate`

- **Ethnic/national groups**
  - “Syrians”
    - OK: no annotation, NG: `gpe.country.country`

- **Type too specific**
  - “Obama”
    - OK: `per.politician`, NG: `per.politician.headofgovernment`

- **Type too generic**
  - “SANA news agency”
    - OK: `org.commercialorganization.newsagency`, NG: `org`
Conclusion

- Multi-task transfer learning approach for ultra fine-grained NER
  - Transfer learning from YAGO to AIDA
  - Multi-task learning of named entity detection and classification
  - Multi-label classification of named entity types
  - Hierarchy-aware soft loss
Improvement Ideas

- Using “type name” embeddings
  - e.g., `per.professionalposition.spokesperson`
  - e.g., `org.commercialorganization.newsagency`
- Gazetteers and handcrafted features
- Hierarchical model
  - BIO+loc/org/per/... -> more fine-grained types
- Ensemble
- Post-processing
- Finally... read the annotation guideline and examine the training data!
Thanks for listening!

Masato Hagiwara\textsuperscript{1}, Ryuji Tamaki\textsuperscript{2}, Ikuya Yamada\textsuperscript{2}

\textsuperscript{1}Octanove Labs
http://www.octanove.com/

\textsuperscript{2}Studio Ousia
http://www.ousia.jp/en/