Coach: A Coarse-to-Fine Approach for Cross-domain Slot Filling
ACL 2020

Zihan Liu
Genta Indra Winata
Peng Xu
Pascale Fung

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
CAiRE Center for Artificial Intelligence Research
Outline

● Background
● Coach Framework
● Experiment Setup
● Results and Analysis
● Conclusion
Background
Cross-domain Slot Filling

- Slot Filling Example

```
O O O O B-music item O B-playlist l-playlist l-playlist
Can you put this tune onto latin dance cardio
```

- Cross-domain Slot Filling
  - Use zero or only a few labeled training samples in target domains.
Challenge & Previous work

Challenge: Handle unseen slot types

- Some slot types in the target domain do not exist in the source domain

Previous work: Concept Tagger (CT) (Bapna, et al 2017)

- CT conducts slot filling for each slot type

Disadvantages

- Difficult to capture the whole slot entity
- An entity could have multiple predictions

Coach Framework
Concept Tagger vs. Coarse-to-Fine

Concept Tagger

Coarse-to-Fine (Coach)
Coarse-to-Fine

Step One (Coarse Step)

\[ [h_1, h_2, ..., h_n] = \text{BiLSTM}(E(w)), \]  
\[ [p_1, p_2, ..., p_n] = \text{CRF}([h_1, h_2, ..., h_n]), \]  

(1)  
(2)

Step Two (Fine Step)

\[ r_k = \text{BiLSTM}([h_i, h_{i+1}, ..., h_j]), \]  
\[ s_k = M_{\text{desc}} \cdot r_k, \]  

(3)  
(4)

We sum the embeddings of the N slot description tokens to obtain slot representation

\[ r^{\text{desc}} = \sum_{i=1}^{N} E(t_i), \]  

(5)

3-way (B-I-O) Classification Per Token

<table>
<thead>
<tr>
<th>O</th>
<th>O</th>
<th>...</th>
<th>B</th>
<th>...</th>
<th>B</th>
<th>I</th>
<th>I</th>
</tr>
</thead>
</table>

Conditional Random Field (CRF)

Encoder

Utterance Can you put this *tune* onto *latin dance cardio*

Step One

Music Item

Slot Type Description Matrix

Similarity comparison

Playlist

Representation

Encoder

...         ...         ...

tune        latin dance cardio

Step Two
Template Regularization

\[ e_t = h_t w_a, \quad \alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^{n} \exp(e_j)}, \quad R = \sum_{t=1}^{n} \alpha_t h_t, \quad (6) \]

\[ L^r = \text{MSE}(R^u, R^r), \quad (7) \]

\[ L^w = -\beta \times \text{MSE}(R^u, R^w), \quad (8) \]
Experiment Setup
Datasets

SNIPS (Slot Filling)

- SNIPS contains 39 slot types across seven domains
- Each time, we choose one domain as the target domain and the other six domains as the source domains.

CoNLL-2003 & SciTech News (NER)

- Source domain: CoNLL 2003 English NER dataset
- Target domain: CBS SciTech News NER dataset
- These two datasets have the same four types of entities, namely, PER (person), LOC (location), ORG (organization), and MISC (miscellaneous).
Baselines

Concept Tagger (CT)

● Bapna et al. (2017) proposed a slot filling framework that utilizes slot descriptions to cope with the unseen slot types in the target domain.

Robust Zero-shot Tagger (RZT)

● Based on CT, Shah et al. (2019) leveraged example values of slots to improve robustness of cross-domain adaptation.

BiLSTM-CRF

● BiLSTM-CRF uses the same label set for the source and target domains and casts it as an entity classification task for each token. This baseline is only for the cross-domain NER.


Results & Analysis
Zero/Few-shot Results on SNIPS

<table>
<thead>
<tr>
<th>Training Setting</th>
<th>Zero-shot</th>
<th>Few-shot on 20 (1%) samples</th>
<th>Few-shot on 50 (2.5%) samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CT</td>
<td>RZT</td>
<td>Coach</td>
</tr>
<tr>
<td>AddToPlaylist</td>
<td>38.82</td>
<td>42.77</td>
<td>45.23</td>
</tr>
<tr>
<td>BookRestaurant</td>
<td>27.54</td>
<td>30.68</td>
<td>33.45</td>
</tr>
<tr>
<td>GetWeather</td>
<td>46.45</td>
<td>50.28</td>
<td>47.93</td>
</tr>
<tr>
<td>PlayMusic</td>
<td>32.86</td>
<td>33.12</td>
<td>28.89</td>
</tr>
<tr>
<td>RateBook</td>
<td>14.54</td>
<td>16.43</td>
<td>25.67</td>
</tr>
<tr>
<td>SearchCreativeWork</td>
<td>39.79</td>
<td>44.45</td>
<td>43.91</td>
</tr>
<tr>
<td>FindScreeningEvent</td>
<td>13.83</td>
<td>12.25</td>
<td>25.64</td>
</tr>
<tr>
<td>Average F1</td>
<td>30.55</td>
<td>32.85</td>
<td>35.82</td>
</tr>
</tbody>
</table>

Table 1: Slot F1-scores based on standard BIO structure for SNIPS. Scores in each row represents the performance of the leftmost target domain, and TR denotes template regularization.

- Coach outperforms CT and RZT, especially in the few-shot scenario.
- Template Regularization (TR) further improves the cross-domain performance.
Both Coach and TR improve the performance on seen and unseen slots.

For unseen slot: Our models are better able to capture the unseen slots since they explicitly learn the general pattern of slot entities.

For seen slots: Baseline models might fail to recognize these seen slots in the target domain, while our approaches can adapt to the seen slot types more quickly in comparison.

Table 2: Averaged F1-scores for seen and unseen slots over all target domains. \(\dagger\) represent the number of training samples utilized for the target domain.
Zero/Few-shot Results on SciTech News

- Coach framework is also suitable for the case where there are no unseen labels in the target domain.

- Template regularization loses its effectiveness in this task, since the text in NER is relatively more open, which makes it hard to capture the templates for each label type.

<table>
<thead>
<tr>
<th>Target Samples</th>
<th>0</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT (Bapna et al. (2017))</td>
<td>61.43</td>
<td>65.85</td>
</tr>
<tr>
<td>RZT (Shah et al. (2019))</td>
<td>61.94</td>
<td>65.21</td>
</tr>
<tr>
<td>BiLSTM-CRF</td>
<td>61.77</td>
<td>66.57</td>
</tr>
<tr>
<td>Coach</td>
<td>64.08</td>
<td><strong>68.35</strong></td>
</tr>
<tr>
<td>Coach + TR</td>
<td><strong>64.54</strong></td>
<td>67.45</td>
</tr>
</tbody>
</table>

Table 3: F1-scores on the NER target domain (CBS SciTech News).
Conclusion

- We introduce a new cross-domain slot filling framework to handle the unseen slot type issue.
- Moreover, template regularization is proposed to improve the adaptation robustness further.
- Experiments show that our model significantly outperforms existing cross-domain slot filling approaches, and it also achieves better performance for the cross-domain NER task, where there is no unseen label type in the target domain.
Thank you
All questions are welcome

Check our code
https://github.com/zliucr/coach