
Zihan Liu, Genta Indra Winata, Zhaojiang Lin, Peng Xu, Pascale Fung
Background

- Supervised neural-based approaches have shown the effectiveness for natural language processing.
Background

- Supervised neural-based approaches have shown the effectiveness for natural language processing.
- However, they heavily rely on large amounts of training data, which makes them not scalable to low-resource languages.
Background

- Supervised neural-based approaches have shown the effectiveness for natural language processing.
- However, they heavily rely on large amounts of training data, which makes them not scalable to low-resource languages.
- A straightforward idea is to adapt the model from the high-resource language into the low-resource languages.
Cross-lingual Task-oriented Dialogue Systems

- Dialogue State Tracking (DST)

**English**

Usr: are there any **eritrean restaurants** in town?
Sys: no, there are no **eritrean restaurants** in town. would you like a different restaurant?
Usr: how about **chinese food**?
Sys: there is a wide variety of **chinese restaurants**, do you have an area preference or a price preference to narrow it down?
Usr: i would like the east part of town.

**Italian**

Usr: "ci sono dei **ristoranti eritrei** in città?",
Sys: no, there are no **eritrean restaurants** in town. would you like a different restaurant?
Usr: che ne dici di cucina **cinese**?
Sys: there is a wide variety of **chinese restaurants**, do you have an area preference or a price preference to narrow it down?
Usr: vorrei nella parte **orientale** della città
Cross-lingual Task-oriented Dialogue Systems

- Natural Language Understanding (NLU)

**English**
Set an alarm for 9 pm tonight

**Spanish**
Configurar alarma para las 9 pm mañana
Straightforward solutions

1. Translate training set from source language to target language
2. Translate test samples
Straightforward solutions

1. Translate training set from source language to target language
2. Translate test samples

Problems

1. We need large amounts of resources to build machine translation systems.
2. Machine translation systems perform badly if the source language and target languages are unrelated languages (e.g., English and Chinese).
Cross-lingual Adaptation

Leverage the inter-connections among languages

English Systems
Cross-lingual Adaptation

Leverage the inter-connections among languages
Cross-lingual Adaptation

Leverage the inter-connections among languages

English Systems

Chinese Systems
Related work

- Chen et al (2018)[1] utilized large amounts of parallel data or bilingual dictionary to build zero-shot cross-lingual DST systems.
- Schuster et al (2019)[2] also leveraged extensive parallel data to build zero-shot cross-lingual NLU systems.
- Collecting bilingual resources is expensive and time-consuming, our work only utilizes very few word pairs as bilingual resources.

Hi, I am looking for a German restaurant

Food type
German food

Model

English

Hi, I am looking for a German restaurant

Attention score
Intuition of Mixed-Language Training

Food type: German food

Model

zero-shot

Food type: ???

Model

English

Hi, I am looking for a German restaurant

Italian

Ciao, sto cercando un bel ristorante Tedesco

Attention score
Intuition of Mixed-Language Training

Hi, I am looking for a Tedesco restaurant

(replace German with Tedesco)
Intuition of Mixed-Language Training

Hi, I am looking for a Tedesco restaurant

Ciao, sto cercando un bel ristorante Tedesco

(replace German with Tedesco)
Attention Layer to select keywords

Filtered words by frequency

Attention-selected words

mapping

German-Tedesco
......
expensive-caro

Attention Layer

selected

hi im looking for a nice German restaurant

English Training Set

EN: hi, im looking for a nice German restaurant.

......
Mixed-Language Training

Filtered words by frequency → mapping → German-Tedesco

Attention-selected words

Attention Layer

English Training Set

EN: hi, im looking for a nice German restaurant.

Food: Tedesco

Attention Model

Cross-lingual Embeddings

Mixed-language Sentence Generation

target word: Tedesco

Code-Switching Sentence Generator

source word: German

EN: hi, im looking for a nice German restaurant.
Dialogue State Tracking (DST)
Natural Language Understanding (NLU)
Zero-shot Results in DST Task

<table>
<thead>
<tr>
<th>Model</th>
<th>slot acc.</th>
<th>German</th>
<th>joint goal acc.</th>
<th>request acc.</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BASE</td>
<td>MLT&lt;sub&gt;O&lt;/sub&gt;</td>
<td>MLT&lt;sub&gt;A&lt;/sub&gt;</td>
<td>BASE</td>
<td>MLT&lt;sub&gt;O&lt;/sub&gt;</td>
</tr>
<tr>
<td>MUSE</td>
<td>60.69</td>
<td>68.58</td>
<td>71.38</td>
<td>21.57</td>
<td>30.61</td>
</tr>
<tr>
<td>XLM (MLM)&lt;sup&gt;*&lt;/sup&gt;</td>
<td>52.21</td>
<td>66.26</td>
<td>68.25</td>
<td>14.09</td>
<td>29.45</td>
</tr>
<tr>
<td>+ Transformer</td>
<td>53.81</td>
<td>65.81</td>
<td>68.55</td>
<td>13.97</td>
<td>30.87</td>
</tr>
<tr>
<td>XLM (MLM+TLM)&lt;sup&gt;*&lt;/sup&gt;</td>
<td>58.04</td>
<td>65.39</td>
<td>66.25</td>
<td>16.34</td>
<td>29.22</td>
</tr>
<tr>
<td>+ Transformer</td>
<td>56.52</td>
<td>66.81</td>
<td>68.88</td>
<td>16.59</td>
<td>31.76</td>
</tr>
<tr>
<td>Multi. BERT&lt;sup&gt;*&lt;/sup&gt;</td>
<td>57.61</td>
<td>67.49</td>
<td>69.48</td>
<td>14.95</td>
<td>30.69</td>
</tr>
<tr>
<td>+ Transformer</td>
<td>57.43</td>
<td>68.33</td>
<td>70.77</td>
<td>15.67</td>
<td>31.28</td>
</tr>
</tbody>
</table>

Zero-shot results for the target languages on Multilingual WOZ 2.0. MLT<sub>A</sub> denotes our approach (attention-informed MLT), which utilizes the same number of word pairs (90 word pairs) as MLT<sub>O</sub> (MLT based on ontology).
Zero-shot Results in NLU Task

Zero-shot results on multilingual NLU dataset (Schuster et al. 2019), and the number of word pairs on both $\text{MLT}_H$ and $\text{MLT}_A$ is 20.
Visualization

Attentions on words in both training and testing phases.
Zero-shot Results in NLU Task

The dynamics of the NLU task: intent and slot-filling results with different numbers of word pairs on Spanish test data.
Conclusion

- We propose attention-informed mixed-language training for cross-lingual task-oriented dialogue systems.
- Our approach utilizes very few task-related parallel word pairs base on the attention scores.
- The task-related words have a generalization ability to other words that have similar semantics in target languages.
Thanks!

Check our code
https://github.com/zliucr/mixed-language-training