Multilingual Disfluency Removal using NMT

Eunah Cho, Jan Niehues, Thanh-Le Ha, Alex Waibel | December 9, 2016
Overview

- Motivation
- Speech disfluency
  - Impact of disfluency in MT performance
- Previous work
- Model
- Experiments and results
- Conclusion
Spoken Language

- Spontaneous speech
- Disfluency: hesitation, stutter, repetition, correction

"uh the one is the uhm is the position"

"uhm maybe I mean let's talk to some of the orga- to some of the orga- group leaders"
Motivation

- Disfluency annotation: expensive!
- Limited data

- Are there similarities of disfluency across languages?
- Potential of multilingual disfluency modeling
- Multilingual approach for speech disfluency removal
Speech Disfluency

(Reparandum)*<editing term> correction

I’d like to book a flight to Boston, sorry, to Denver.
Speech Disfluency

- Filler: *uh, uhm*

- Discourse marker: *you know, well*

- Repetition
  *uhm right, they don’t, uhm, they don’t actually go into ...*

- False starts
  *what if, oh, the, that’s what you want to do?*
Impact of Disfluency on BLEU

- English meeting data → French
- German lecture data → English
- 3.6 - 4.3 BLEU points of difference when removing disfluencies according to manual annotation
Related work: CRF-based approach

- Given the observed sequence, a hidden label sequence is modeled
  - Label: disfluency classes

![Diagram showing three labels: Non-copy, Filler, Clean, with features and their values]

- Features
  - Lexical features: word, POS, their patterns, etc
  - Clusters from word vectors
  - Phrase table information
Disfluency removal in NMT framework

Disfluency removal as a translation task from disfluent to clean language

What if, oh, the, that’s what you want to do?

⇓

that’s what you want to do?
Disfluency removal in NMT framework

Possible to extend to further tasks: reconstruction (reordering, replacement of words)

what if, oh, the, that’s what you want to do?

Is that what you want to do?
Colloquial expressions and ungrammatical phrases still remain

Surfen sie *mal* ein bisschen 'rum auf den Seiten, die ich Ihnen gegeben habe, vielleicht fällt Ihnen *was* auf, *was* sie gerne machen wollen.

Surfen sie *einmal* ein bisschen *auf den Seiten herum*, die ich Ihnen gegeben habe, vielleicht fällt Ihnen *etwas* auf, *das* sie gerne machen wollen.
Multilingual disfluency removal

- Source: disfluent language in English and German
- Target: cleaned-up English/German

what if, oh, the, that’s what you want to do?

that’s what you want to do?
Multilingual disfluency removal

- Source: disfluent language in English and German
- Target: cleaned-up English/German

es gab uh es gibt da drei Prinzipien ...

es gibt da drei Prinzipien ...
Multilingual learning

- Sub-word operation?

  *he bro+ ke his arm*.

- Rare-word problem of NMT

- Vocabulary sharing, language specification?

  *@en_he @en_bro+ @en_ke @en_his @en_arm @en_.
@de_er @de_ist @de_arm @de_.*

- Share parameters
- Disambiguation
Experimental Setup

- Two single language disfluency removal systems
  - EN-EN
  - DE-DE
- A multilingual disfluency removal systems
  - EN/DE - EN/DE
- Identical architecture

- Output of disfluency removal is translated into another language
  - English → French
  - German → English
  - Evaluate the impact of disfluency removal in a downstream application
Experimental Setup

- Attention-based encoder-decoder model for NMT \textit{nematus}
  - Source: disfluent transcripts
  - Target: clean transcripts

- Data
  - English: meetings, online lectures
  - German: university lectures
  - Manual annotation on disfluency
  - Almost identical guideline
    - Meeting: \textit{interruption} class

- ~100K tokens of training data for each language
- ~20K, ~30K tokens of test data
- Disfluency rate: 12%
Sub-word operation

- Evaluation of performance depending on different sub word operation size
- Multilingual system
- BLEU against the human-cleaned transcript

<table>
<thead>
<tr>
<th>Sub-word operation</th>
<th>No. tokens</th>
<th>Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character</td>
<td>971k</td>
<td>78.97</td>
</tr>
<tr>
<td>BPE 150</td>
<td>498k</td>
<td>92.59</td>
</tr>
<tr>
<td>BPE 200</td>
<td>465k</td>
<td>92.01</td>
</tr>
<tr>
<td>BPE 500</td>
<td>372k</td>
<td>92.37</td>
</tr>
</tbody>
</table>

- 42.4% tokens left unsplit
Sub-word operation

w_hat i_f o_h the th_at’s w_hat you w_an_t to d_o ?
Language specific representation

Would the language ID be helpful?

<table>
<thead>
<tr>
<th>System</th>
<th>Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPE 150</td>
<td>92.59</td>
</tr>
<tr>
<td>+ LangID</td>
<td>91.05</td>
</tr>
<tr>
<td>BPE 200</td>
<td>92.01</td>
</tr>
<tr>
<td>+ LangID</td>
<td>91.75</td>
</tr>
</tbody>
</table>

No language specific representation, shared vocabulary
## Results - intrinsic

<table>
<thead>
<tr>
<th>System</th>
<th>English</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>74.37</td>
<td>78.03</td>
</tr>
<tr>
<td>+ no <em>uh</em></td>
<td>76.82</td>
<td>84.90</td>
</tr>
<tr>
<td>Single language NMT sys.</td>
<td>81.56</td>
<td>89.61</td>
</tr>
<tr>
<td><strong>Multilingual NMT sys.</strong></td>
<td><strong>83.57</strong></td>
<td><strong>90.75</strong></td>
</tr>
<tr>
<td>CRF-based single language sys.</td>
<td>78.78</td>
<td>-</td>
</tr>
</tbody>
</table>

- Evaluated against human-cleaned test set, in BLEU
- Baseline: all disfluencies kept
- no *uh*: remove *uh* and *uhm*
- Outperforms CRF-based English system
Results - extrinsic

<table>
<thead>
<tr>
<th>System</th>
<th>English</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>17.08</td>
<td>21.58</td>
</tr>
<tr>
<td>+ no <em>uh</em></td>
<td>17.75</td>
<td>23.46</td>
</tr>
<tr>
<td>Single language NMT sys.</td>
<td>19.36</td>
<td>24.34</td>
</tr>
<tr>
<td><strong>Multilingual NMT sys.</strong></td>
<td><strong>19.59</strong></td>
<td><strong>24.43</strong></td>
</tr>
<tr>
<td>CRF-based single language sys.</td>
<td>18.22</td>
<td>-</td>
</tr>
<tr>
<td>Oracle</td>
<td>21.38</td>
<td>25.22</td>
</tr>
</tbody>
</table>

- Translate the disfluency-cleaned test data
  - English → French
  - German → English
- 1-1.8 BLEU points of improvement over no *uh* baseline
- Oracle: human-cleaned test data
Example output of disfluency removal

| Input | and as we find more groups, we record more groups. |
| Ref   | and as we find more groups, we record more groups. |
| CRF   | as we record more groups. |
| MultiNMT | and as we find more groups, we record more groups. |

| Input | yeah but this is really I we we’re not uh record a really tight subset of meetings. |
| Ref   | we’re not record a really tight subset of meetings. |
| CRF   | we’re not record a really tight subset of meetings. |
| MultiNMT | but this is really tight subset of meetings. |
Example output of disfluency removal

- Test on ASR output
- Punctuation inserted using monolingual MT
- German and English mixed input

**Input:** ja, dann ja, und kann dann auch die Fragen stellen oh, okay, that is cool

**MultiNMT:** dann auch die Fragen stellen okay, that is cool

**Input:** oh, okay, okay, I mean, they got like like like federal und wir Informationen heraus suchen

**MultiNMT:** okay, I mean, they got like federal und wir Informationen heraus suchen
Conclusion

- Model disfluency removal in an NMT framework
- Motivated by data sparsity for disfluency removal
- Multilingual learning for English and German spontaneous data
- Outperformed CRF-based model and single language based models

- Extend to further tasks (reconstruction, reordering, etc)
Thank you!