NAIST’s Machine Translation Systems for IWSLT 2020 Conversational Speech Translation Task

Ryo Fukuda¹, Katsuhito Sudoh¹, and Satoshi Nakamura¹,²

¹Nara Institute of Science and Technology
²AIP Center, RIKEN, Japan
**Brief Overview**

**Challenge track:** Conversational Speech Translation

Translation task from disfluent Spanish to fluent English
• Includes speech-to-text and text-to-text translation subtask

**Motivation:** Tackle two problems on text-to-text NMT
1. Low-resource translation
2. Noisy input sentences
   • fillers, hesitations, self-corrections, ASR errors, ...

**Proposal:** Domain adaptation using style transfer
• transfer the styles of out-of-domain data to be like in-domain data, and then performed domain adaptation
Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Summary
Motivation

The “style” of task data (in-domain):

→ Ideally, augment data by using large corpus same style

Large corpus available (out-of-domain):

→ Effects of training with them are limited
Motivation

Style transfer model: fluent to disfluent

- increase the similarity between out-of-domain and in-domain data

→ Enables effective domain adaptive training
Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Summary
Overview

Generate pseudo in-domain data and adapt it for NMT

(1) Style Transfer model
Fluent-to-Disfluent

(2) NMT model
Spanish-to-English
(1) Style Transfer model

Transfer fluent input sentences of out-of-domain parallel data into disfluent styles

Style Transfer model:
• based on Unsupervised NMT (Artetxe et al., 2018; Lample et al., 2018) with out-of-domain fluent data and in-domain disfluent data
Apply fine-tuning

• conventional domain adaptation methods of MT
• greatly improves the accuracy of low-resource domain-specific translation (Dakwale and Monz, 2017)

Learning steps for fine-tuning:

1. Pre-training steps
   - Pseudo in-domain parallel
   - pre-trained model
     Spanish-to-English

2. Fine-tuning steps
   - In-domain parallel
   - fine-tuned model
     Spanish-to-English
Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Conclusion
Datasets

- LDC Fisher Spanish speech with English translations (Fisher)
  - parallel in-domain data
  - disfluent Spanish to (fluent/disfluent) English

- United Nations Parallel Corpus (UNCorpus)
  - parallel out-of-domain data
  - fluent Spanish to fluent English

<table>
<thead>
<tr>
<th>Data statistics</th>
<th># sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fisher</strong> (in-domain)/Train</td>
<td>138,720</td>
</tr>
<tr>
<td>Dev</td>
<td>3,977</td>
</tr>
<tr>
<td>Test</td>
<td>3,641</td>
</tr>
<tr>
<td><strong>UNCorpus</strong> (out-of-domain)/Train</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Dev</td>
<td>4,000</td>
</tr>
<tr>
<td>Test</td>
<td>4,000</td>
</tr>
</tbody>
</table>
(1) Spanish Style Transfer

Data: Fisher (disfluent) and UNCorpus (fluent) Spanish data

Model: Unsupervised NMT (UNMT) based on Transformer

Evaluation:
• Estimate the similarity between domains by measuring the perplexity of 3-gram language model
(1) Spanish Style Transfer

Results
- reduced perplexity and number of unknown words by style transfer

<table>
<thead>
<tr>
<th>Training data</th>
<th>perplexity</th>
<th>unknown words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fisher</td>
<td>72.46</td>
<td>0</td>
</tr>
<tr>
<td>UNCorpus</td>
<td>589.81</td>
<td>5,173,539</td>
</tr>
<tr>
<td>Fisher-like UNCorpus</td>
<td>474.47</td>
<td>4,217,819</td>
</tr>
</tbody>
</table>

Examples of pseudo in-domain data (*Fisher-like UNCorpus*)

<table>
<thead>
<tr>
<th>UNCorpus</th>
<th>Fisher-like UNCorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>d conducta y disciplina</td>
<td>eh conducta y disciplina</td>
</tr>
<tr>
<td>c lista amplia de verificación para la autoevaluación</td>
<td>mhm lista amplia de verificación para la la tele</td>
</tr>
</tbody>
</table>

- Delete paragraph symbol
- Insert “Disfluency” (filler, repetition, missing words, ASR error, ..)
(2) NMT with Domain Adaptation

Data
- **in-domain**: 130K bilingual pairs of Fisher
- **out-of-domain**: 1M of UNCorpus or Fisher-like UNCorpus

Model: Transformer (almost follow the transformer_base settings)

Evaluation: calculated the BLEU scores with sacreBLEU
(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

<table>
<thead>
<tr>
<th>System</th>
<th>Fisher</th>
<th>Fisher-like UNCorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Training</strong></td>
<td>UNCorpus</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Fisher</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>Fisher-like UNCorpus</td>
<td>6.7</td>
</tr>
<tr>
<td><strong>Fine-tuning</strong></td>
<td>UNCorpus + Fisher</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>Fisher-like UNCorpus + Fisher</td>
<td>18.5</td>
</tr>
</tbody>
</table>

- Domain adaptation training outperformed the baseline
- slightly improved by using the pseudo in-domain data
(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

<table>
<thead>
<tr>
<th>System</th>
<th>Fisher</th>
<th>Fisher-like UNCorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Training</td>
<td>UNCorpus</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Fisher</td>
<td>14.8</td>
</tr>
<tr>
<td>Fine-tuning</td>
<td>UNCorpus + Fisher</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>Fisher-like UNCorpus + Fisher</td>
<td>18.5</td>
</tr>
</tbody>
</table>

- Domain adaptation training outperformed the baseline
- Slightly improved by using the pseudo in-domain data
(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

BLEU scores of trained NMT models for Disfluent Spanish to Fluent English

<table>
<thead>
<tr>
<th>System</th>
<th>Fisher</th>
<th>Fisher-like UNCorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Training</strong></td>
<td>Fisher</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>UNCorpus</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Fisher-like UNCorpus</td>
<td>6.7</td>
</tr>
<tr>
<td><strong>Fine-tuning</strong></td>
<td>UNCorpus + Fisher</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>Fisher-like UNCorpus + Fisher</td>
<td>18.5</td>
</tr>
</tbody>
</table>

- Domain adaptation training outperformed the baseline
- Slightly improved by using the pseudo in-domain data

$+0.2$
(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

**BLEU scores of trained NMT models for Disfluent Spanish to Fluent English**

<table>
<thead>
<tr>
<th>System</th>
<th>Fisher</th>
<th>Fisher-like UNCorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Training</td>
<td>14.8</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>18.3</td>
<td>18.5</td>
</tr>
<tr>
<td>Fine-tuning</td>
<td>18.3</td>
<td>18.5</td>
</tr>
</tbody>
</table>

- Domain adaptation training outperformed the baseline
- Slightly improved by using the pseudo in-domain data
(2) NMT with Domain Adaptation

Results (2/2) – Fluent vs Disfluent references

“Fisher (disfluent)” did not use Fisher’s fluent references but instead used disfluent references

<table>
<thead>
<tr>
<th>System</th>
<th>Fisher/test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fisher (fluent)</td>
<td>14.8</td>
</tr>
<tr>
<td>UNCorpus + Fisher (fluent)</td>
<td>18.3</td>
</tr>
<tr>
<td>Fisher-like UNCorpus + Fisher (fluent)</td>
<td>18.5</td>
</tr>
<tr>
<td>Fisher (disfluent)</td>
<td>11.6</td>
</tr>
<tr>
<td>UNCorpus + Fisher (disfluent)</td>
<td>15.2</td>
</tr>
<tr>
<td>Fisher-like UNCorpus + Fisher (disfluent)</td>
<td>15.6</td>
</tr>
</tbody>
</table>

-3.2
-3.1
-2.9

• models trained with Fisher’s original disfluent references had about 3 points lower BLEU
Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Summary
Effect of Style Transfer

The use of pseudo in-domain data improved accuracy, but
- there was no significant improvement
- was worse in the pre-training phase

An example of style transferred sentence:

nueva york 1 a 12 de junio de 2015 (original)
nueva york oh a mi eh de de de de (generated)

- some sentences lost the meaning of the sentence
- style transfer constrains may be too strong

→ This problem may be mitigated by a model that can control the trade-off between style transfer and content preservation
Fluent vs Disfluent References

The model trained using Fisher’s original disfluent data had a BLEU score of about 3 points lower than the model trained using fluent data.

→ by removing the disfluency of reference sentences improves the BLEU by about three points for all the learning strategies we tried

• the use of large out-of-domain data with fluent reference sentences did not mitigate this problem

Style of the sentence has an impact on the translation accuracy
Summary

Translation accuracy was improved

- by domain adaptation (+3.7)
- by style transfer of out-of-domain (+0.4)
  - effect was limited due to parallel data quality degradation

Future work

pursue a style transfer that does not reduce the quality of the parallel data