Neural Lattice Search for Domain Adaptation in Machine Translation

Huda Khayrallah, Gaurav Kumar
Kevin Duh, Matt Post, Philipp Koehn

This talk was presented at IJCNLP 2017
It is based on this paper:
http://aclweb.org/anthology/I17-2004
bib:
Neural Lattice Search for Domain Adaptation in Machine Translation

Huda Khayrallah, Gaurav Kumar
Kevin Duh, Matt Post, Philipp Koehn
combine adequacy of PBMT with fluency of NMT
use PBMT to constrain the search space of NMT
die brötchen sind warm
Neural Lattice Search

Source

die brötchen sind warm

Lattice

Target

the buns are warm
die brötchen sind warm  the buns are warm
die brötchen sind warm

bread is

buns are

warm

the

buns are

is

are

warm

bread
die brötchen sind warm

bread is

buns are

buns

are

warm

the

bread is

buns

are

warm

0

1

2

3

4

Khayrallah, Kumar, Duh, Post, Koehn
Die Brötchen sind warm

The bread is warm.

The buns are warm.

The bread is buns.

The buns are warm.
die brötchen sind warm
die brötchen sind warm

bread is buns are warm

bread are buns warm

the buns are warm

bread is buns are
the buns are bread is warm

bread is buns are warm
die brötchen sind warm
die brötchen sind warm

bread is buns are warm
buns are bread is warm

the buns are warm
the bread is warm

Khayrallah, Kumar, Duh, Post, Koehn
die brötchen sind warm
die brötchen sind warm
die brötchen sind warm
Experiments
Setting: Domain adaptation

Small **in-domain**

IT, Medical, Koran, Subtitles

PBMT outperforms NMT

Large **out-of-domain**

parliamentary proceedings (WMT)

NMT outperforms PBMT
## Setting: Domain adaptation

<table>
<thead>
<tr>
<th>in-domain</th>
<th>out-of-domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PBMT</strong></td>
<td></td>
</tr>
<tr>
<td>in-domain</td>
<td></td>
</tr>
<tr>
<td></td>
<td>out-of-domain</td>
</tr>
</tbody>
</table>
IT Results

Comparison of BLEU scores:
- PBMT (in)
- NMT (out)
- n-best
- Lattice Search

Lattice Search shows a +5.0 improvement over PBMT (in).
Results

<table>
<thead>
<tr>
<th></th>
<th>IT</th>
<th>Koran</th>
<th>Medical</th>
<th>Subtitle</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBMT (in)</td>
<td>+5.0</td>
<td>+0.4</td>
<td>+0.2</td>
<td>+1.6</td>
</tr>
<tr>
<td>NMT (out)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-best</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lattice search</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PBMT (in) vs. NMT (out): +5.0, +0.4, +0.2, +1.6
Conclusion

• Lattice search > \( n \)-best rescoring
• Use in-domain PBMT to constrain search space
• NMT can be in- or out-of-domain

Code:

github.com/khayrallah/nematus-lattice-search
Thanks!

This material is based upon work supported in part by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-15-C-0113.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA).
Neural Lattice Search for Domain Adaptation in Machine Translation

Huda Khayrallah, Gaurav Kumar
Kevin Duh, Matt Post, Philipp Koehn
{huda, gkumar, kevinduh, post, phi}@cs.jhu.edu

code:
github.com/khayrallah/nematus-lattice-search
## Corpus Sizes

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Words</th>
<th>Sentences</th>
<th>W/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical</td>
<td>14,301,472</td>
<td>1,104,752</td>
<td>13</td>
</tr>
<tr>
<td>IT</td>
<td>3,041,677</td>
<td>337,817</td>
<td>9</td>
</tr>
<tr>
<td>Koran</td>
<td>9,848,539</td>
<td>480,421</td>
<td>21</td>
</tr>
<tr>
<td>Subtitles</td>
<td>114,371,754</td>
<td>13,873,398</td>
<td>8</td>
</tr>
<tr>
<td>EuroParl</td>
<td>113,165,079</td>
<td>4,562,102</td>
<td>25</td>
</tr>
</tbody>
</table>
How much text do we have?

![Graph showing BLEU scores vs. corpus size](image.png)

- **Phrase-Based with Big LM**
- **Phrase-Based**
- **Neural**

**Corpus size (English words)**

**BLEU**

- **Medical**
- **Subtitles & WMT**
- **Koran**
- **IT**

To illustrate this, see Figure 4: Translations of the first sentence of a sample text:

```
Una estrategia republicana para contrarrestar la reelección de Obama.
```

- With 1.6 million words of training data, the output is completely unrelated to the input. Some key words are properly translated, but the translations become respectable.

**Phrase-Based** systems trained on each subset, respectively. In addition to a NMT system trained on the Spanish part of the data, outperforming SMT 25.7 vs. 16.4 for a big language model in contrastive SMT systems.

We built English-Spanish systems on WMT training corpus sizes of a few million words or even beating the SMT system with a big language model in domain. However, both NMT and SMT systems continue to have difficulty translating some rare words, (Koehn & Knowles 2017).

**Neural** training corpus sizes of a few million words or even beating the SMT system with a big language model in domain. However, both NMT and SMT systems continue to have difficulty translating some rare words, (Koehn & Knowles 2017).
Corpus Sizes

words

Medical  IT  Koran  Subtitles  EuroParl

words
IT Baselines

![Bar chart showing BLEU scores for PBMT (in), PBMT (out), NMT (in), and NMT (out).]
<table>
<thead>
<tr>
<th>Test Domain</th>
<th>Training Configuration</th>
<th>PBMT 1-best</th>
<th>NMT Standard Search</th>
<th>NMT Rescoring</th>
<th>NMT Lattice Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>PBMT_{out} \times NMT_{out}</td>
<td>25.1 (-0.3)</td>
<td>22.5 (-2.9)</td>
<td>22.2 (-3.2)</td>
<td>25.4</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{in}</td>
<td>47.4 (-4.2)</td>
<td>34.2 (-17.4)</td>
<td>47.6 (-4.0)</td>
<td>51.6</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{out}</td>
<td>47.4 (-5.2)</td>
<td>22.5 (-30.1)</td>
<td>47.6 (-5.0)</td>
<td>52.6*</td>
</tr>
<tr>
<td></td>
<td>PBMT_{out} \times NMT_{in}</td>
<td>25.1 (-2.2)</td>
<td>34.2 (6.9)</td>
<td>22.4 (-4.9)</td>
<td>27.3</td>
</tr>
<tr>
<td>Medical</td>
<td>PBMT_{out} \times NMT_{out}</td>
<td>33.3 (-0.9)</td>
<td>32.9 (-1.3)</td>
<td>30.8 (-3.4)</td>
<td>34.2</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{in}</td>
<td>47.4 (-0.7)</td>
<td>37.8 (-10.3)</td>
<td>40.2 (-7.9)</td>
<td>48.1*</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{out}</td>
<td>47.4 (-0.4)</td>
<td>32.9 (-14.9)</td>
<td>39.7 (-8.1)</td>
<td>47.8</td>
</tr>
<tr>
<td></td>
<td>PBMT_{out} \times NMT_{in}</td>
<td>33.3 (-2.7)</td>
<td>37.8 (1.8)</td>
<td>31.2 (-4.8)</td>
<td>36.0</td>
</tr>
<tr>
<td>Koran</td>
<td>PBMT_{out} \times NMT_{out}</td>
<td>14.7 (-0.2)</td>
<td>10.8 (-4.1)</td>
<td>13.9 (-1.0)</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{in}</td>
<td>20.6 (-0.1)</td>
<td>15.9 (-4.8)</td>
<td>19.3 (-1.4)</td>
<td>20.7</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{out}</td>
<td>20.6 (-0.2)</td>
<td>10.8 (-10.0)</td>
<td>19.4 (-1.4)</td>
<td>20.8*</td>
</tr>
<tr>
<td></td>
<td>PBMT_{out} \times NMT_{in}</td>
<td>14.7 (-1.4)</td>
<td>15.9 (-0.2)</td>
<td>13.9 (-2.2)</td>
<td>16.1</td>
</tr>
<tr>
<td>Subtitle</td>
<td>PBMT_{out} \times NMT_{out}</td>
<td>26.6 (-0.9)</td>
<td>25.3 (-2.2)</td>
<td>19.7 (-7.8)</td>
<td>27.5</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{in}</td>
<td>26.8 (-1.1)</td>
<td>24.9 (-3.0)</td>
<td>17.8 (-10.1)</td>
<td>27.9</td>
</tr>
<tr>
<td></td>
<td>PBMT_{in} \times NMT_{out}</td>
<td>26.8 (-1.6)</td>
<td>25.3 (-3.1)</td>
<td>17.1 (-11.3)</td>
<td>28.4*</td>
</tr>
<tr>
<td></td>
<td>PBMT_{out} \times NMT_{in}</td>
<td>26.6 (-1.0)</td>
<td>24.9 (-2.7)</td>
<td>19.8 (-7.8)</td>
<td>27.6</td>
</tr>
<tr>
<td>Source</td>
<td>Versionsinformationen ausgegeben und beenden</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>output version information and exit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBMT</td>
<td>Spend version information and end</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMT</td>
<td>Spend and end versionary information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lattice</td>
<td>Print version information and exit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>Language</th>
<th>PBMT (in)</th>
<th>NMT (in)</th>
<th>N-best</th>
<th>Hybrid Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>+4.2</td>
<td>+0.7</td>
<td>+0.1</td>
<td>+1.1</td>
</tr>
<tr>
<td>Medical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koran</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtitle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Khayrallah, Kumar, Duh, Post, Koehn
Stack Based Decoding

- Stacks based on number of target words translated
- Keep track of:
  - Score
  - Current lattice node
  - Current neural state
  - incoming arc
  - length