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:

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
Source

die brötchen sind warm

Target

the buns are warm

Lattice

bread is
buns are
bread
buns
is
are
warm
warm

Neural Lattice Search
die brötchen sind warm  the buns are warm
die brötchen sind warm

bread is

buns are

warm

buns are

warm

bread

is

are

warm
die brötchen sind warm

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

bread is
buns are
warm
are

the
0
1
2
3
4

0
1
2
3
4

Khayrallah, Kumar, Duh, Post, Koehn
die brötchen sind warm
die brötchen sind warm
die brötchen sind warm
die brötchen sind warm

bread is buns are warm

bread is buns are

the bread are warm

0 1 2 3 4
die brötchen sind warm

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

bread is
buns are
buns are warm
bread is
is warm
are
are

Khayrallah, Kumar, Duh, Post, Koehn
die brötchen sind warm

The bread is warm.
die brötchen sind warm

bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
buns are warm
bread is warm
bun
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

PBMT

| in-domain | out-of-domain |
|-----------|--------------|
| in-domain |              |
| out-of-domain |          |

NMT
Results

BLEU

|        | PBMT (in) | NMT (out) | n-best | lattice search |
|--------|-----------|-----------|--------|----------------|
| IT     | +5.0      |           |        |                |
| Koran  | +0.4      |           |        |                |
| Medical| +0.2      |           |        |                |
| Subtitle| +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

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

![Graph showing BLEU scores for different text sizes.](image)

**Figure 3:** BLEU scores for English-Spanish systems.

We built English-Spanish systems on WMT 2016. For SMT, the translation models perform particularly poorly on translation of very infrequent words. However, both NMT and SMT systems perform much better than naive baseline systems. NMT systems outperform SMT at about 15 million English words.

**Figure 4:** Translations of the first sentence of the test set using NMT system trained on varying amounts of training data. Under low-resource conditions, the output is completely unrelated to the input. Some key words are properly translated with increasing amounts of training data, the output becomes respectable.

**Figure 5:** Corpus size (English words) vs. BLEU score for SMT, SMT+BigLM, NMT, and NMT+BigLM.

To illustrate this, see Figure 5. The contrast between the NMT and SMT learning curves is quite striking. While NMT is able to exploit increasing amounts of training data more effectively, it is unable to get off the ground with high-resource conditions. Under low-resource conditions, the output is completely unrelated to the input. Some key words are properly translated with increasing amounts of training data, the output becomes respectable.

**Figure 6:** Results are shown in Figure 6.

We use a publicly available model trained on 0.4 million to 385.7 million English words of parallel data. Quality for NMT starts much lower, outperforms SMT at about 15 million English words, and even beats a SMT system with a big language model trained on the Spanish part of the data and all additional provided monolingual data.

**Figure 7:** BLEU Scores with Varying Amounts of Training Data.

| Corpus Size (English Words) | Phrase-Based with Big LM | Phrase-Based | Neural |
|-----------------------------|--------------------------|-------------|--------|
| 10^6                        | 21.8                     | 23.4        | 16.4   |
| 10^7                        | 26.2                     | 26.9        | 19.6   |
| 10^8                        | 28.6                     | 29.2        | 21.2   |

**Figure 8:** Corpus size (English words) vs. BLEU score for various tasks.

- **IT:** 21.8, 26.2, 28.6
- **Koran:** 16.4, 19.6, 21.2
- **Subtitles & WMT:** 24.9, 26.9, 28.6
- **Medical:** 22.2, 24.7, 29.2

**Figure 9:** BLEU scores for SMT, SMT+BigLM, NMT, and NMT+BigLM.

**Table 1:** BLEU scores for various tasks.
Corpus Sizes

words

0 20000000 40000000 60000000 80000000 100000000 120000000 140000000

Medical IT Koran Subtitles EuroParl
IT Baselines

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

|                | PBMT (in) | NMT (in) | N-best | Hybrid Lattice |
|----------------|-----------|----------|--------|----------------|
| IT             | +4.2      |          |        |                |
| Medical        | +0.7      | +0.1     |        |                |
| Koran          |           |          | +0.1   | +1.1           |
| Subtitle       |           | +4.2     |        |                |
Stack Based Decoding

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