Unsupervised Pivot Translation for Distant Languages
Supplementary Materials

Yichong Leng[*]
University of Science and Technology of China
lyc123go@mail.ustc.edu.cn

Xu Tan
Microsoft Research
xuta@microsoft.com

Tao Qin[†]
Microsoft Research
taoqin@microsoft.com

Xiang-Yang Li[†]
University of Science and Technology of China
xiangyangli@ustc.edu.cn

Tie-Yan Liu
Microsoft Research
tyliu@microsoft.com

1 Language Name and Code

The language names and their corresponding language codes according to ISO 639-1 standard[1] are listed in Table 1.

2 Languages in Each Branch

There are 20 languages in total involved in our experiments, which can be divided in to 4 language branches: Balto-Slavic branch, Germanic branch, Italic branch and Uralic branch[2]. The detailed languages in each branch are listed in Table 2.

3 Training Data Size

We use monolingual data from Wikipidia[3] to train the unsupervised model and monolingual data from TED talks[4] to finish the fine-tune process. The training data size of each language is listed in Table 3.

4 BLEU Scores in Development and Test Set

For the direct translation (*DT*) and best multi-hop path (*GT*) performance of all language pairs in test and development set, we list their BLEU as well as relative BLEU gap in Table 4.

5 BLEU Scores in Development and Test Set with Supervised Pivoting

We leverage the existing bilingual data between pivot languages (English, German, Spanish, French, Finish and Russian) to train supervised NMT models, replacing their counterpart unsupervised models. Table 5 shows the ground-truth best path with and without supervised pivoting. In general, leveraging bilingual data can increase the performance of ground-truth best path. However, there still exists some language pairs whose ground-truth best path cannot benefit from the supervised pivoting like Nl-Ro and It-Et.

---

[*] The work was done when the first author was an intern at Microsoft Research Asia.

[†] Corresponding author

[1] https://www.loc.gov/standards/iso639-2/php/code_list.php

[2] The first three branches belong to Indo-European family while the last branch is actually a language family. We do not further split the 3 languages in Uralic family into different branches.

[3] For example, we download English Wikipedia contents from https://dumps.wikimedia.org/enwiki.

[4] https://github.com/ajinkyakulkarni14/TED-Multilingual-Parallel-Corpus/tree/master/Monolingual_data

Language	Code	Language	Code	Language	Code	Language	Code
Belarusian	Be	Bulgarian	Bg	Danish	Da	German	De
English	En	Spanish	Es	Estonian	Et	Finnish	Fi
French	Fr	Galician	Gl	Croatian	Hr	Hungarian	Hu
Italian	It	Macedonian	Mk	Dutch	Nl	Portuguese	Pt
Romanian	Ro	Russian	Ru	Swedish	Sv	Ukrainian	Uk

Table 1: The ISO 639-1 code of each language in our experiments.

Branch	Language
Balto-Slavic	Russian, Belorussian, Ukrainian, Bulgarian, Macedonian, Croatia
Germanic	English, German, Dutch, Swedish, Danish
Italic	Spanish, French, Romanian, Italian, Portuguese, Galician
Uralic	Finnish, Hungarian, Estonian

Table 2: Languages in each language branch. Uralic is actually a language family.

Language	Be	Bg	Da	De	En	Es	Et	Fi	Fr	Gl
Wiki Data	1.07M	1.93M	1.87M	21.00M	52.94M	18.61M	1.50M	3.29M	18.61M	1.18M
TED Talks Data	11K	476K	129K	472K	214K	521K	33K	62K	493K	22K

Language	Hr	Hu	It	Mk	Nl	Pt	Ro	Ru	Sv	Uk
Wiki Data	1.54M	3.71M	12.43M	0.83M	11.45M	7.80M	2.61M	14.85M	19.32M	7.12M
TED Talks Data	327K	398K	502K	69K	433K	251K	454K	610K	121K	282K

Table 3: The size of the monolingual data from Wikipedia and TED talks for each language.

Source	Target	DT	GT	Δ	Source	Target	DT	GT	Δ
Da	Gl	6.56	12.14	5.58	It	Et	4.05	6.46	2.41
Bg	Sv	4.72	9.92	5.20	Da	Mk	5.89	8.27	2.38
Gl	Sv	3.79	8.62	4.83	Pt	Mk	6.33	8.69	2.36
Sv	Gl	3.70	8.13	4.43	Gl	Bg	8.56	10.90	2.34
Be	It	2.11	6.40	4.29	Et	It	4.57	6.89	2.32
Pt	Be	4.76	8.86	4.10	Uk	Nl	6.22	8.05	1.83
Sv	It	8.67	12.77	4.10	Bg	Gl	8.84	9.95	1.11
Gl	Da	7.45	11.33	3.88	It	Mk	6.16	7.17	1.01
Mk	It	6.60	10.38	3.78	Mk	Da	7.26	8.20	0.94
Be	Pt	6.39	9.77	3.38	Et	Nl	6.65	7.54	0.89
Sv	Bg	5.05	8.12	3.07	Nl	Et	6.38	7.26	0.88
It	Be	2.24	5.19	2.95	Da	Hu	6.54	6.99	0.45
It	Sv	9.38	12.20	2.82	Hu	Da	6.75	6.66	-0.09
Nl	Uk	4.69	7.23	2.54	Mk	Pt	9.28	8.25	-1.03

Source	Target	DT	GT	Δ	Source	Target	DT	GT	Δ
Be	Nl	2.02	6.81	4.79	Ro	Nl	9.40	10.73	1.33
Et	Sv	6.30	9.48	3.18	Uk	Et	6.25	6.78	0.53
Nl	Be	1.93	5.11	3.18	Hr	Hu	6.20	6.68	0.48
Mk	Hu	1.47	4.62	3.15	Hu	Hr	5.32	5.78	0.46
Hu	Mk	1.22	3.67	2.45	Et	Uk	6.54	6.55	0.01
Sv	Et	6.72	8.98	2.26	Nl	Ro	9.44	9.25	-0.19

Table 4: The BLEU scores of all language pairs in the test and development set. The upper table is the test set BLEU score and the lower table is that of development set. Δ is the BLEU gap between direct translation (*DT*) and best path (*GT*).

Source	Target	*DT*	*GT*-unsup	*GT*-sup	Δ		Source	Target	*DT*	*GT*-unsup	*GT*-sup	Δ
Da	Gl	6.56	12.14	15.20	8.64		It	Et	4.05	6.46	6.39	2.34
Bg	Sv	4.72	9.92	9.92	5.20		Da	Mk	5.89	8.27	8.27	2.38
Gl	Sv	3.79	8.62	9.58	5.79		Pt	Mk	6.33	8.69	8.58	2.25
Sv	Gl	3.70	8.13	9.38	5.68		Gl	Bg	8.56	10.90	11.38	2.82
Be	It	2.11	6.40	9.26	7.15		Et	It	4.57	6.89	8.03	3.46
Pt	Be	4.76	8.86	13.03	8.27		Uk	Nl	6.22	8.05	9.09	2.87
Sv	It	8.67	12.77	12.77	4.10		Bg	Gl	8.84	9.95	12.55	3.71
Gl	Da	7.45	11.33	15.52	8.07		It	Mk	6.16	7.17	7.24	1.08
Mk	It	6.60	10.38	10.38	3.78		Mk	Da	7.26	8.20	8.20	0.94
Be	Pt	6.39	9.77	14.50	8.11		Et	Nl	6.65	7.54	7.54	0.89
Sv	Bg	5.05	8.12	8.12	3.07		Nl	Et	6.38	7.26	7.23	0.85
It	Be	2.24	5.19	8.60	6.36		Da	Hu	6.54	6.99	6.99	0.45
It	Sv	9.38	12.20	12.20	2.82		Hu	Da	6.75	6.66	6.66	-0.09
Nl	Uk	4.69	7.23	8.07	3.38		Mk	Pt	9.28	8.25	9.15	-0.13

Source	Target	*DT*	*GT*-unsup	*GT*-sup	Δ		Source	Target	*DT*	*GT*-unsup	*GT*-sup	Δ
Be	Nl	2.02	6.81	9.97	7.95		Ro	Nl	9.40	10.73	10.73	1.33
Et	Sv	6.30	9.48	9.48	3.18		Uk	Et	6.25	6.78	6.49	0.24
Nl	Be	1.93	5.11	8.29	6.36		Hr	Hu	6.20	6.68	6.68	0.48
Mk	Hu	1.47	4.62	4.07	2.60		Hu	Hr	5.32	5.78	5.78	0.46
Hu	Mk	1.22	3.67	3.67	2.45		Et	Uk	6.54	6.55	6.55	0.01
Sv	Et	6.72	8.98	8.98	2.26		Nl	Ro	9.44	9.25	9.21	-0.23

Table 5: The BLEU scores of all language pairs in the test and development set when enhanced with supervised pivoting. The upper table is the test set BLEU score and the lower table is that of development set. *GT*-sup and *GT*-unsup represent the ground-truth best path with and without supervised pivoting. Δ is the BLEU gap between *GT*-sup and *DT*.