

Manipulating Multilingual Models for Neural Machine Translation (NMT)

Fengjiao Sun
Jordan Meyer
Leon Rafael Gutierrez Angulo

INTRODUCTION

KPIS:

**PERFORMANT TRANSLATION
PORTABLE MODEL**

LANGUAGES

ENGLISH

Germanic
Latin¹



17%²

CHINESE

Chinese
Chinese



15%

SPANISH

Romance
Latin



6%

1. Angela Fan et al., "Beyond English-Centric Multilingual Machine Translation." arXiv, Oct. 21, 2020 [Online]. Available: <http://arxiv.org/abs/2010.11125>

2. The World Factbook 2021. Washington, DC: Central Intelligence Agency, 2021.

TECHNICAL BACKGROUND

METRIC:

BLEURT

³

DATASET:

NEWS COMMENTARY

⁴

Bitest Pair	Training	Evaluation	Test
En ↔ Zh	70% (49k)	15% (10k)	5k
Zh ↔ Es	70% (46k)	15% (10k)	5k
Es ↔ En	70% (167k)	15% (36k)	5k

3. Thibault Shellam, "BLEURT: Learning Robust Metrics for Text Generation." arXiv, May, 21, 2020 <https://arxiv.org/pdf/2004.04696>

4. Jörg Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation, (LREC 2012)

METHODOLOGY



EXPERIMENTS

SPECIFICITY

$N \leftrightarrow N$
 $1 \leftrightarrow 1$

SIZE

Parameters
Saved model size

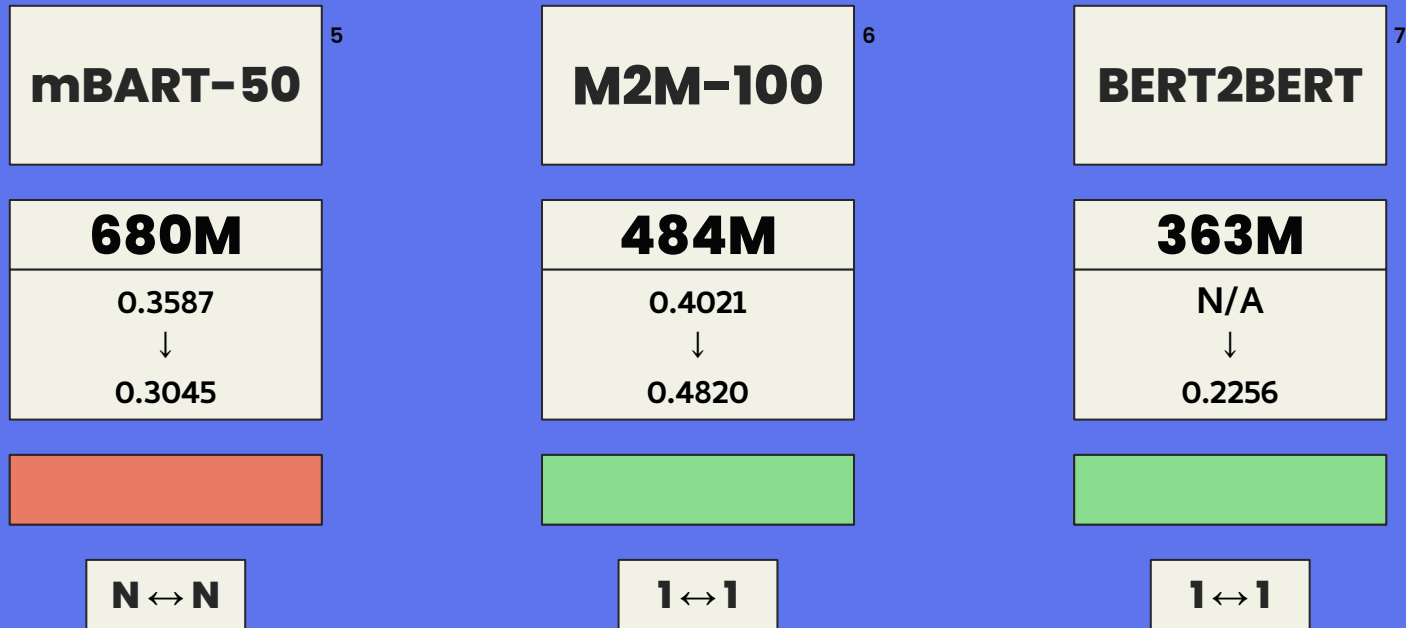
TRAIN ORDER

Language cycle
Shuffled training

OPTIMIZATION

Maximum length
Minimum length
Ngram repetition

MODELS



5. Yuqing Tang et al., "Multilingual Translation with Extensible Multilingual Pretraining and Finetuning." arXiv, Aug. 02, 2020 <http://arxiv.org/abs/2008.00401>

6. Angela Fan et al., "Beyond English-Centric Multilingual Machine Translation." arXiv, Oct. 21, 2020 <http://arxiv.org/abs/2010.11125>

7. Jacob Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019 <http://arxiv.org/abs/1810.04805>

EXAMPLES

MBART SIZE

Example:
Input (English): “<s>en_XX [‘Companies tend not to invest in talent or develop new skills, and the quality of existing skills can suffer. In the longer run, flexible labor markets also increase structural unemployment and fuel the informal economy.</s>”

Prediction (Undirected): "~~</s><unk>~~ ['One cannot afford to take the table over human rights in one country, but to keep one's views in one another – a wholesome occurrence. In other words, one's views can be held back from</s>"]"

[illegible]

Predicted (Spanish):
</s>es_XX [In some cases, they involve border disputes (Bolivia and Venezuela, Bolivia, Bolivia, and Nicaragua, etc.). In others, they concern the armed factions (Colombia and Venezuela, Colombia and</s>

**REPEAT TO
FILL OUT MIN**

Example:
Input (English): The process resembles the selection of a standard in, say, consumer electronics.

Label (Spanish): El proceso se parece a la selección de una norma, digamos, entre productos electrónicos.

Predicted (Spanish): el proceso se parece a la seleccion de un estandar en el que se asemeja a la seleccion de un estandar en la industria de consumo, por ejemplo, la seleccion de un estandar en la industria de consumo de consumo de consumo, digamos, por ejemplo, la electronica de consumo de los consumidores.

NOBODY REALLY UNDERSTANDS CHINESE

Example: Input (English): With the stakes so high, the Russian-US initiative to organize a peace conference has not only been welcomed, but is seen as an indispensable initiative that must be planned with excruciating care. We are likely to hear of delay after delay in the coming days, as various sides try to choose their representatives.

Input (English): With the stakes so high, the Russian-US initiative to organize a peace conference has not only been welcomed, but is seen as an indispensable initiative that must be planned with excruciating care. We are likely to hear of delay after delay in the coming days, as various sides try to choose their representatives.

Label (Chinese): 因为利害攸关, 俄美组织和和平会议的
计划不仅颇受欢迎, 而且被视为不可或缺的行动
而受到慎重对待。因为各方都力争选定代表, 和平
会议的日程可能一推再推。但我们必须保持耐心,
因为会议一旦开始除成功外就没有别的选择。没有
结果的努力不会有任何用处。

Predicted (Chinese):
因此,俄罗斯-美国组织和和平会议的意图不但得到了欢迎,也被视为一个不可或缺的计划,必须以谨慎的态度进行规划。我们很可能会在未来几天中听到延迟之后的延迟,因为各方试图选择代表。

RESULTS

Model	Name / Modifier	Parameters	Encoder/ Decoder/ Hidden Layers	Status	en-zh	zh-en	en-es	es-en	es-zh	zh-es	Average
M2M-100	Baseline	484M	12 / 12 / 12	Baseline	0.3785	0.4018	0.6232	0.4421	0.2651	0.3021	0.4021
	Trained	484M	12 / 12 / 12	Fine-tuned	0.4154	0.4288	0.6482	0.6775	0.3906	0.3313	0.4820
mBART	Standard	610M	12 / 12 / 12	Baseline	0.3259	0.4138	0.5017	0.6128	0.1417	0.1566	0.3587
	Standard	610M	12 / 12 / 12	Fine-tuned	0.2174	0.3257	0.5861	0.3764	0.0965	0.2249	0.3045
	Mini	435M	6 / 6 / 6	Baseline	0.0714	0.1031	0.1244	0.1348	0.0948	0.0865	0.1025
	Mini	435M	6 / 6 / 6	Fine-tuned	0.2762	0.1220	0.2158	0.1833	0.2764	0.0846	0.1931
	Tiny	343M	3 / 3 / 3	Baseline	0.0712	0.0510	0.0304	0.0538	0.0720	0.0319	0.0517
	Tiny	343M	3 / 3 / 3	Fine-tuned	0.2337	0.1261	0.1213	0.1742	0.2454	0.0806	0.1635
	Single	288M	1 / 1 / 1	Baseline	0.0712	0.0422	0.0304	0.0356	0.0719	0.0319	0.0472
	Single	288M	1 / 1 / 1	Fine-tuned	0.0491	0.1470	0.4301	0.2448	0.0487	0.1233	0.1738
	Shuffled	288M	1 / 1 / 1	Fine-tuned	0.0713	0.1419	0.0758	0.3001	0.0645	0.0538	0.1179
BERT2BERT	Default ¹	363M	12 / 12 / 768	Baseline	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		363M	12 / 12 / 768	Fine-tuned	0.1479	0.2425	0.2963	0.4502	0.1340	0.0827	0.2256
	Max Length 100	363M	12 / 12 / 768	Fine-tuned	0.3267	0.1889	0.4951	0.6044	0.3076	0.1686	0.3485
	Min Length 50	363M	12 / 12 / 768	Fine-tuned	0.3320	0.3428	0.3304	0.5336	0.3185	0.1600	0.3362
	Max Length 100 N-Gram Repetition 3	363M	12 / 12 / 768	Fine-tuned	0.3269	0.3575	0.4937	0.6036	0.3077	0.1635	0.3755

THANK YOU

QUESTIONS?