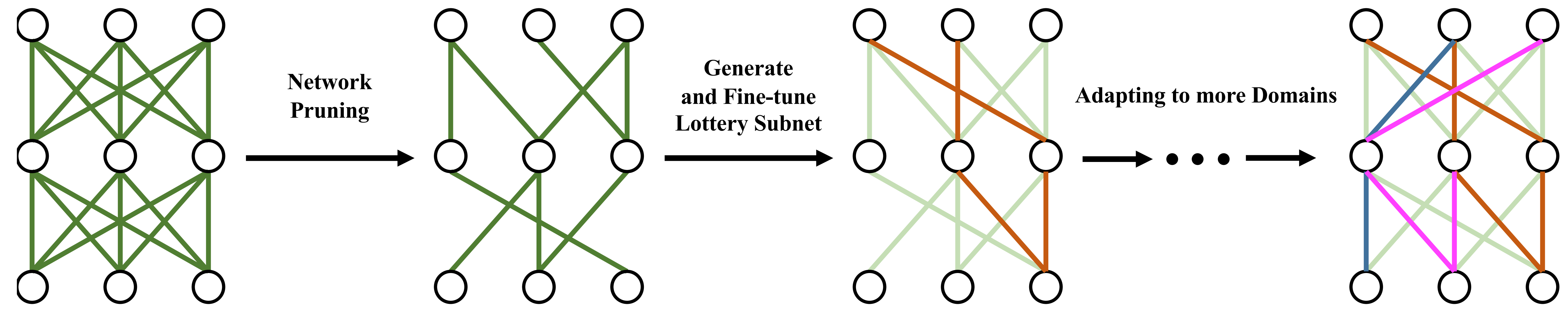
国际人工智能会议
AAAI 2021 论文北京预讲会

Finding Sparse Structure for Domain Specific Neural Machine Translation

ByteDance AI Lab
字节跳动人工智能实验室

Jianze Liang, [1,2*] Chengqi Zhao, [2] Mingxuan Wang, [2] Xipeng Qiu, [1] Lei Li [2]
1 Fudan University 2 ByteDance AI Lab

Prune-Tune: An Effective and Flexible Schema for Domain Adaptation in NMT



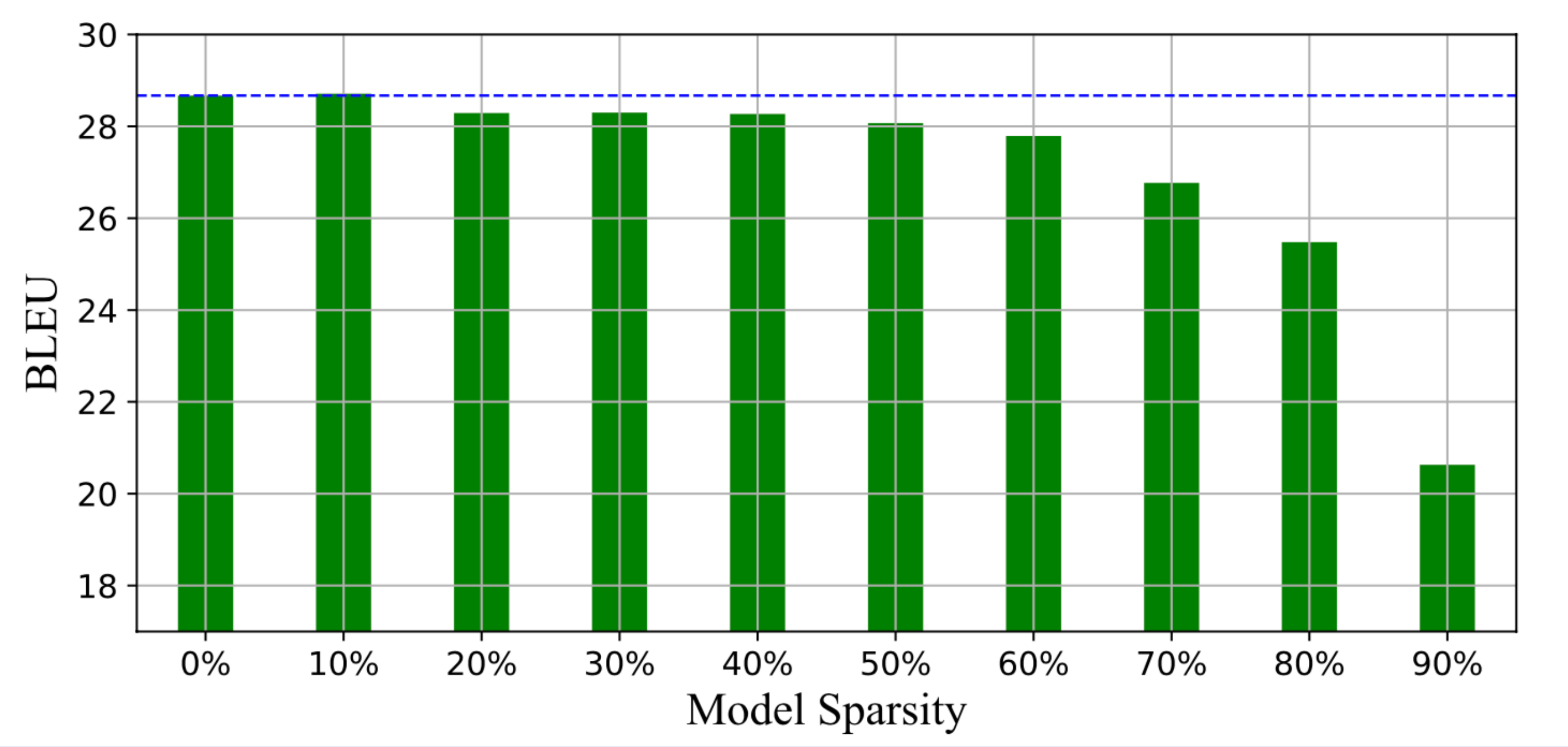
a) General Model

b) Informative General Subnetwork

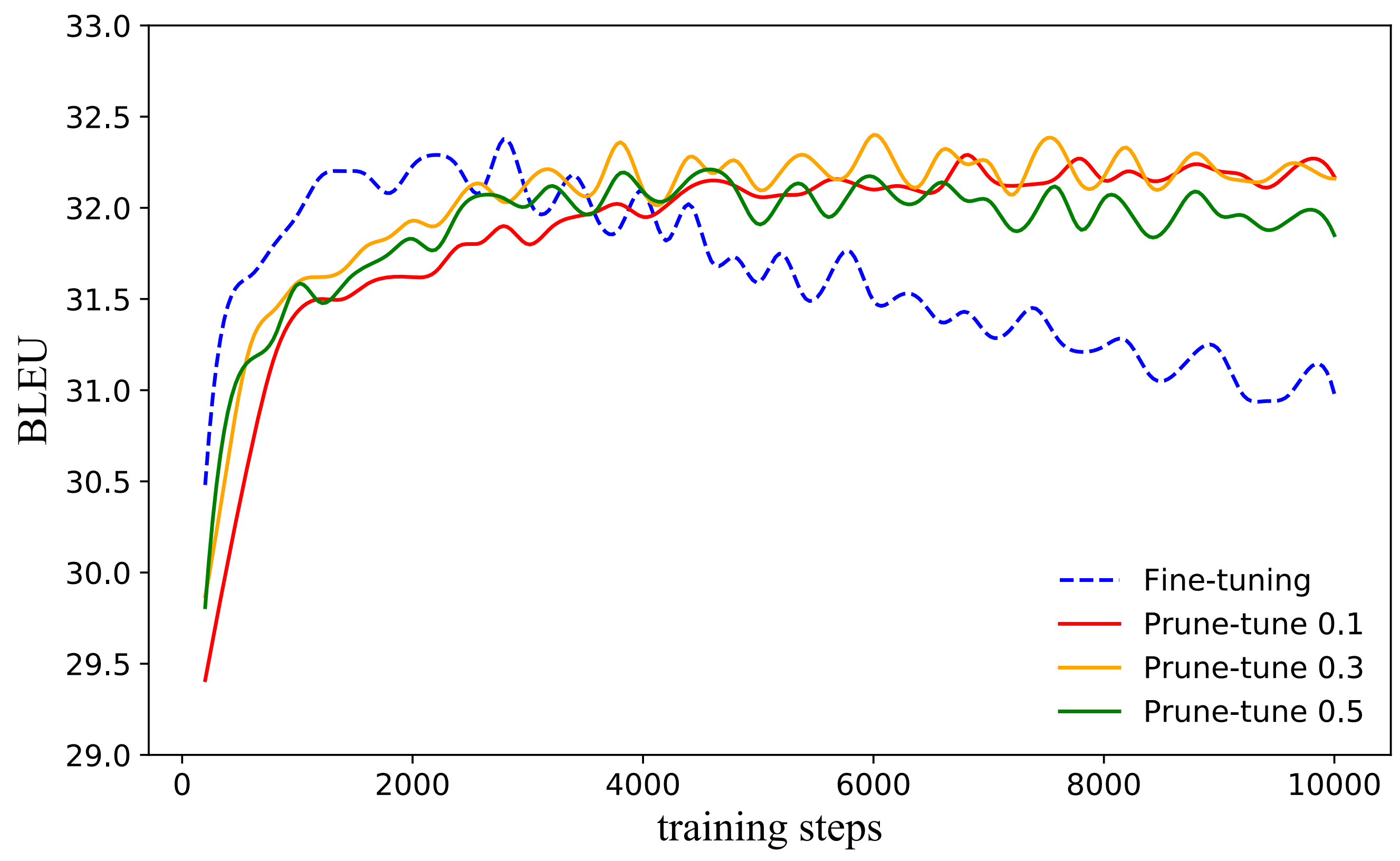
c) Lottery Domain Subnetwork

d) Multi-domain Model

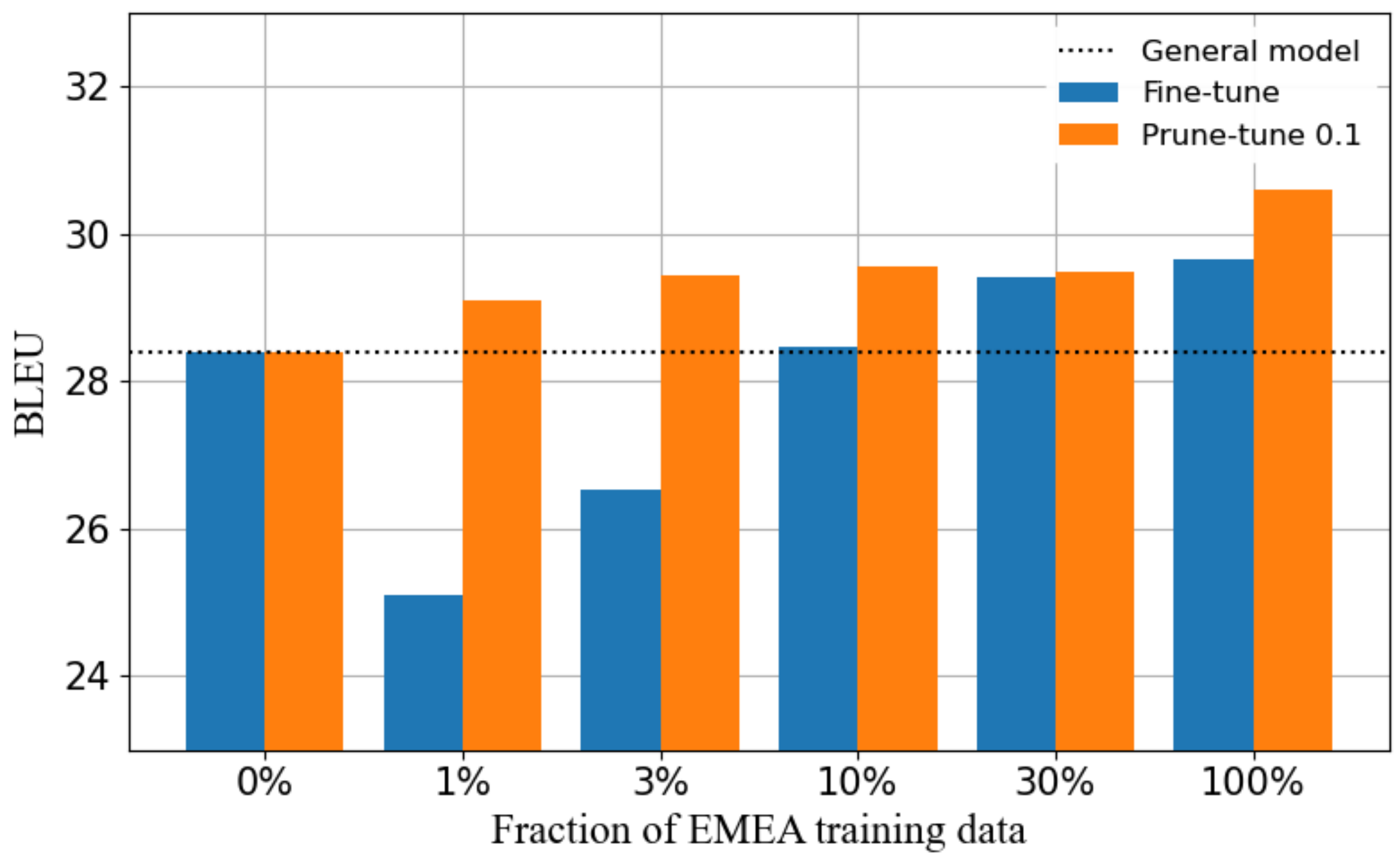
Effective Pruning for Transformer



Robust Training



Effective for Low-resource Domain Adaptation



Keeping General Knowledge to better Learn the Target Domain

Model	IWSLT (190k)		EMEA (587k)		Novel (50k)		#Tuning Params
	general	target	general	target	general	target	
Mixed Domain Model	27.9	31.3	27.9	**32.0**	27.9	21.2	273M
Target Domain Model	N/A	24.0	N/A	23.9	N/A	12.3	273M
General Domain Model	28.7	28.5	28.7	28.4	28.7	14.5	273M
+ Fine-tuning (Luong and Manning 2015)	27.0	31.5	17.1	29.7	12.1	23.4	273M
+ EWC-regularized (Thompson et al. 2019)	28.0	31.5	27.1	30.5	23.5	23.1	273M
+ Model Distillation (Khayrallah et al. 2018)	26.3	31.5	16.3	30.0	11.6	23.1	273M
+ Layer Freeze (Thompson et al. 2018)	28.6	31.3	26.9	29.8	23.0	23.0	29M
+ Adapter (Bapna and Firat 2019)	27.0	31.6	26.7	30.1	19.8	24.3	13M
Prune-Tune Model	**28.8**	**31.9**	**28.9**	30.6	**28.8**	**24.3**	27M

Table 2: BLEU results of domain adaptation on EN→DE

Few parameters are needed to train most target domains

Pruning Rate	WMT	IWSLT	EMEA	Novel
10%	**28.7**	32.3	**30.6**	**24.3**
30%	28.3	**32.4**	30.3	23.9
50%	28.1	32.2	29.5	23.6
70%	26.8	31.8	28.9	23.1

Direction	Corpus	Train	Dev.	Test
EN→DE	WMT14	3.9M	3000	3003
	IWSLT14	170k	6750	1305
	EMEA	587k	500	1000
	Novel	50k	1015	1031
ZH→EN	WMT19	20M	3000	3981
	Laws	220k	800	456
	Thesis	300k	800	625
	Subtitles	300k	800	598
	Education	449K	800	791
	News	449K	800	1500
	Spoken	219k	800	456

Sequential Multi-Domain Adaptation: Learning without Forgetting

Model	Input domain	#M	WMT14 (**W**)	IWSLT (**I**)	EMEA (**E**)	Novel (**N**)
Mixed Domain Model	**W, I, E, N**	1	27.9	31.3	32.0	21.2
General Domain Model	**W**	1	28.7	28.5	28.4	14.5
+ Fine-tuning	**I, E, N**	3	N/A	31.5	29.7	23.4
Single P-Tune Model	**W, I, E, N**	3	N/A	31.9	30.6	24.3
Sequential P-Tune Model	#1 **W**	1	28.4	N/A	N/A	N/A
	#2 + **I**		28.4	31.9	N/A	N/A
	#3 + **E**		28.4	31.9	30.1	N/A
	#4 + **N**		28.4	31.9	30.1	23.6

Table 3: BLEU Results of Sequential Domain Adaptation on EN→DE. #M denotes the number of required models. **W**, **I**, **E**, **N** refer to dataset WMT14, IWSLT, EMEA, Novel, respectively. In our Sequential P-Tune Model, general domain occupied 50% parameters, and each target domain occupied 10%.

Model	#M	Laws	Thesis	Subtitles	Education	News	Spoken	Avg.
Mixed Domain Model	1	47.4	15.6	17	31.4	21.2	16.7	24.9
General Domain Model	1	44.9	13.8	16.1	30.8	21.4	16.7	23.9
+ Fine-tuning	6	55.9	17.9	20.8	29.2	22.1	14.8	26.7
Sequential P-Tune Model	1	50.3	16.2	17.2	31.2	21.3	14.6	25.1

Table 4: BLEU Results of Sequential Domain Adaptation on ZH→EN. #M denotes the number of required models. In our Sequential P-Tune Model, general domain occupied 50% parameters, and each target domain occupied 5%.

主办方：中国中文信息学会青年工作委员会
承办方：智源社区