Team： DSEC-COER

Zhejiang University

DSEC

COER

Siyu Huang
CV, NLP, DL

Ming Chen
NLP, DL

Lei Wang
AI, DL

http://www.isee.zju.edu.cn/dsec/

http://coer.zju.edu.cn/

Speech Recognition Post-processing & Machine Translation

Train Set

A good bride and groom - that's it. You're done.
一对好的新娘新郎- 到此为止。你该说完了。

A pair of freaks... Who's the other freak?
一对怪咖...另一个怪咖是谁?

On the final tonight the most talk - about man in American
在节目最后我们请到了时下最受争议的

On the ferry? Have they....
在港口？他们有没有…

Valid Set

Women are in politics That 's not news Shouldn 't be surprising That fact is not even a unique phenomenon
女性从政已经不是什么新鲜事了。 这没什么可惊讶的。甚至不能称之为一件稀罕事。

So leadership today are facing some major obstacles so think about global financial turmoil so that 's just you know markets fluctuating exchange rates changing trade changing There 's just a lot going on with the global financial turmoil
所以当今的领导层面临着一些重大的阻碍，想想全球金融风暴，就这样，如你所知，市场波动的汇率变化，贸易变化。全球金融风暴正在开启。

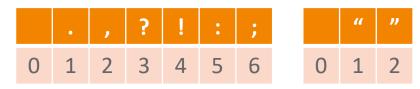|	N	d_{model}	d_{ff}	h	d_k	d_v	P_{drop}	ϵ_{ls}	word embed	train steps	PPL (dev)	BLEU (dev)	params $\times 10^6$
base	6	512	2048	8	64	64	0.1	0.1	32K	100K	4.92	25.8	65
(A)				1	512	512					5.29	24.9	
				4	128	128					5.00	25.5	
				16	32	32					4.91	25.8	
				32	16	16					5.01	25.4	
(B)				16							5.16	25.1	58
				32							5.01	25.4	60
(C)	2										6.11	23.7	36
	4										5.19	25.3	50
	8										4.88	25.5	80
		256			32	32					5.75	24.5	28
		1024			128	128					4.66	26.0	168
			1024								5.12	25.4	53
			4096								4.75	26.2	90
(D)							0.0				5.77	24.6	
							0.2				4.95	25.5	
								0.0			4.67	25.3	
								0.2			5.47	25.7	
(E)		positional embedding instead of sinusoids									4.92	25.7	
big	6	1024	4096	16			0.3			300K	4.33	26.4	213
Ours	**6**	**512**	**4096**	**8**	**64**	**64**	**0.3**	**0**	**60K**	**200K**			**136**

Hyper-Parameters Setting

Model	Training Cost (FLOPs)	
	EN-DE	EN-FR
ByteNet [18]		
Deep-Att + PosUnk [39]		$1.0 \cdot 10^{20}$
GNMT + RL [38]	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	$3.3 \cdot 10^{18}$	
Transformer (big)	$2.3 \cdot 10^{19}$	
Transformer (Ours)	$7.8 \cdot 10^{18}$	

- 4×GTX 1080 Ti (10.8 TFLOPs)
- 50 hours of training

3-1 Experimental Results

Baseline:

- Transformer (base)
- Single 1080Ti gpu
- 200k iterations (~1 day)
- Batch size ≈ 4096 tokens
- Without punctuation prediction and sentence segmentation

Model	BLEU (Valid)	Improvement
Baseline	25.68	-
Sentence segmentation	28.18	+ 2.50
Beam size=16 & α=1.0	28.54	+ 0.36
Add conv layer	28.74	+ 0.20
Bigger model (4 gpus)	30.22	+ 1.48
Checkpoints average	31.22	+ 1.00
Remove short sentences	31.60	+ 0.38
Independent models ensemble	32.76	+ 1.16
Reranking (beam size 100)	33.27	+ 0.51

Eng

I had the opportunity to meet a young man with a rough past And upon meeting me he exclaimed Me you we have nothing in common you 're young you 're white you 're from a family of privilege you can 't possibly understand what 's happened to me

Punc (Ours)

I had the opportunity to meet a young man with a rough past**.** And upon meeting me**,** he exclaimed**.** **"**Me you**,** we have nothing in common**,** you're young**,** you're white**,** you're from a family of privilege**,** you can't possibly understand what's happened to me**.**

CHN (GT)

一次我碰到一个有着艰辛过往的年轻人。一见面他就冲我喊道：你，我，我们没有共同点 你年轻，你是白人，你生在条件好的家庭，你不可能理解发生在我身上的事情。

CHN (Ours)

我有机会见到一个有着艰难过去的年轻人，在遇见我的时候，他惊呼："我你，我们没有共同点，你年轻，你是白人，你来自一个特权家庭，你不可能理解我发生了什么事。"

Eng

This research for me has been time well spent as it afforded me the opportunity to attend the International Science and Engineering Fair as a finalist The pivotal moment for me was seeing all of my peers from around the world reshaping science as we know it and that is what motivates me to be excellent every day

Punc (Ours)

This research for me has been time well spent as it afforded me the opportunity to attend the International Science and Engineering Fair as a finalist. The pivotal moment for me was seeing all of my peers from around the world reshaping science as we know it, and that is what motivates me to be excellent every day.

CHN (GT)

这项研究对我来说时间用得其所，因为它使我机会作为决赛选手参加国际科学与工程博览会。对我来说，关键时刻是看到来自世界各地的同龄人重塑我们已知的科学，这就是促使我每天都变得优秀的动力。

CHN (Ours)

对我来说，这项研究花费了很多时间，因为它让我有机会参加国际科学和工程博览会的决赛。对我来说，最关键的时刻是看到来自世界各地的同行们重塑我们所熟知的科学，而这正是促使我每天都变得优秀的原因。