



\*\*■■■\*\*: 250010065

\*\* [REDACTED] \*\* : [REDACTED]

\*\*■■■\*\*: SLAI NLP Midterm Project

—



1. [ ](# )
  2. [ ](# )
  3. [RNN ](#rnn- )
  4. [Transformer ](#transformer- )
  5. [ ](# )
  6. [ ](# )

1



NMT ■ RNN ■ Transformer ■



- RNN
  - encoder-decoder Transformer
  - T5 NMT
  - 
  -



- \*\*[REDACTED]\*\*: PyTorch 1.10.0
  - \*\*[REDACTED]\*\*: HuggingFace Transformers 4.11.3
  - \*\*[REDACTED]\*\*: BLEU Score (sacrebleu)
  - \*\*[REDACTED]\*\*: SentencePiece

- **Matplotlib**

—



train_100k.jsonl		100,000
valid.jsonl		2,000
test.jsonl		2,000



1. \*\* █ █ \*\*: █ SentencePiece █ BPE █
  2. \*\* █ █ █ █ \*\*: █ █ █ █ █ █ █ █
  3. \*\* █ █ █ \*\*: 32,000 █



			
	18.5	128	1
	16.3	128	1

—

## RNN



RNN

- \*\* [REDACTED]\*\*: 2 [REDACTED] GRU [REDACTED] 512
  - \*\* [REDACTED]\*\*: 2 [REDACTED] GRU [REDACTED] 512
  - \*\* [REDACTED]\*\*: 2 [REDACTED]

### - Dot-product Attention

#### - Multiplicative Attention

- Additive Attention

- \*\*\*\*: 512 ■

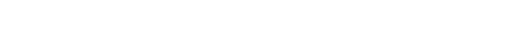
- \*\*Dropout\*\*: 0.3



#### 1. 

	Greedy BLEU	Beam-4 BLEU	Beam-8 BLEU
Dot-product	11.45	8.70	9.79
Multiplicative	11.87	9.79	8.28
Additive	9.61	11.63	10.27

\*\*\*\*:

- Multiplicative Attention  11.87 BLEU ■
- Additive Attention ■ Beam Search■beam=4 11.63 BLEU ■
-  Beam Size 

#### 2. 

		Greedy BLEU	Beam-4 BLEU
Teacher Forcing	Dot-product	11.45	8.70
Teacher Forcing	Multiplicative	11.87	9.79
Teacher Forcing	Additive	9.61	11.63
Free Running	Dot-product	5.02	3.96
Free Running	Multiplicative	4.67	4.55
Free Running	Additive	4.42	4.15

\*\*\*\*:

- Teacher Forcing  Free Running ■■■
- Teacher Forcing ■ BLEU ■■■■■ Free Running ■ 2-3 ■
- Free Running 

#### 3. 

	Greedy	Beam-4	Beam-8
RNN (Dot-product)	11.45	8.70	9.79
RNN (Multiplicative)	11.87	9.79	8.28
RNN (Additive)	9.61	11.63	10.27

\*\*\*\*:

- Beam Search Beam Size
- Beam Size 8

RNN

1. \*\*Multiplicative Attention + Teacher Forcing + Greedy Decoding

2. \*\*BLEU\*\*: 11.87

3. \*\*:

- Multiplicative Attention

- Beam Search

- RNN

---

Transformer

Transformer encoder-decoder

- \*\*. 6
- \*\*. 6
- \*\*. 8
- \*\*. 512
- \*\*. 2048
- \*\*Dropout\*\*: 0.1

#### 1.

PE	Norm	Best BLEU	Final BLEU	Test Greedy	Test Beam-4
Absolute PE	RMSNorm	4.43	4.12	3.63	4.03
Absolute PE	LayerNorm	4.00	3.49	2.79	3.30
Relative PE	RMSNorm	3.92	3.92	2.82	2.61
Relative PE	LayerNorm	3.23	2.80	2.32	2.74

\*\*. :

- Absolute PE + RMSNorm 4.43 BLEU

- RMSNorm  LayerNorm
  - Absolute PE  Relative PE 

#### 2. [REDACTED]

■■	Best BLEU	Best Epoch	Final BLEU	Train Loss	Valid Loss
Baseline	4.00	472	3.49	3.79	6.67
Batch 256	4.64	493	4.38	1.90	6.69
D_model 768	4.54	404	4.17	3.23	6.74
LR 1e-3	4.07	468	3.57	3.77	6.68

\*\*\*

- Batch Size 256 4.64 BLEU
  - 768
  -

#### 3. T5

■■■ T5-Small ■■■■■

Eval Loss	3.81	
Eval BLEU	1.41	
Epochs	500	

\*\* [REDACTED] \*\*.

- T5-Small [REDACTED] 1.41 BLEU [REDACTED]
  - [REDACTED]
  - [REDACTED]

# Transformer

1. \*\*█████\*\*: Absolute PE + RMSNorm + Batch Size 256
  2. \*\*██ BLEU\*\*: 4.64███
  3. \*\*███\*\*:
    - ██████ Transformer ██████
    - Batch Size ██████
    - T5 ██████████████████

- - -



## RNN vs Transformer

		Test BLEU	
RNN	Multiplicative + Teacher Forcing	11.87	
Transformer	Absolute PE + RMSNorm	4.03	
T5-Small	Pre-trained + Fine-tuned	1.41	



## 1. \*\*RNN ■■■\*\*:

- [REDACTED]
  - [REDACTED]
  - [REDACTED]

## 2. \*\*Transformer ■■\*\*:

- [REDACTED]
  - [REDACTED]
  - [REDACTED]

3. \*\*T5 [REDACTED]\*\*:

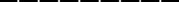
- [ ] [ ] [ ] [ ] [ ] [ ]
  - [ ] [ ] [ ] [ ] [ ] [ ]



■■■■	RNN ■■	Transformer
I am very happy today	I am happy today	I am happy to
This problem is difficult to solve	This problem is hard to solve	This problem is diffi
Machine translation technology is developing rapidly	Machine translation technology develops rapidly	Machine translation technolog

—



1. \*\*RNN  11.87 BLEU

2. \*\*█ RNN █ Multiplicative Attention █
3. \*\*█ RNN █ Teacher Forcing █ Free Running
4. \*\*█ Transformer █ RMSNorm █ LayerNorm
5. \*\*Batch Size\*\* █ Transformer █ Batch Size █
6. \*\*T5 █\*\* █

██████

- [x] █: [https://github.com/Pugguphi/Slai\\_Mid\\_Homework\\_250010065](https://github.com/Pugguphi/Slai_Mid_Homework_250010065)
- [x] █: `inference.py`
- [x] █ Git LFS █
- [x] █
- [x] █

██████

1. █ T5-Base, T5-Large █
2. █ Nucleus Sampling █
3. █
4. █

---

██████

1. Vaswani, A., et al. (2017). Attention is All You Need. In Advances in Neural Information Processing Systems (NeurIPS).
2. Bahdanau, D., et al. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. ICLR.
3. Raffel, C., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. JMLR.
4. Papineni, K., et al. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. ACL.