

1) Neural Machine Translation with RNNs

Screenshots from training process on Microsoft azure:

```
Microsoft Azure Search resources, services, and docs (G+) es-Youssef.Aly20@alexu... ALEXANDRIA UNIVERSITY
Home > MyMachine | Serial console > Serial console
MyMachine | Serial console Virtual machine
? Feedback Feedback
epoch 12, iter 75380, avg. loss 23.84, avg. ppl 3.25 cum. examples 44137, speed 4621.08 words/sec, time elapsed 11055.80 sec
epoch 12, iter 75390, avg. loss 25.45, avg. ppl 3.32 cum. examples 44457, speed 4540.48 words/sec, time elapsed 11057.30 sec
epoch 12, iter 75400, avg. loss 22.93, avg. ppl 3.01 cum. examples 44777, speed 4616.32 words/sec, time elapsed 11058.75 sec
epoch 12, iter 75410, avg. loss 24.08, avg. ppl 3.15 cum. examples 45097, speed 4762.69 words/sec, time elapsed 11060.16 sec
epoch 12, iter 75420, avg. loss 24.84, avg. ppl 3.26 cum. examples 45417, speed 4590.62 words/sec, time elapsed 11061.63 sec
epoch 12, iter 75430, avg. loss 25.43, avg. ppl 3.25 cum. examples 45737, speed 4684.63 words/sec, time elapsed 11063.11 sec
epoch 12, iter 75440, avg. loss 24.08, avg. ppl 3.06 cum. examples 46057, speed 4816.43 words/sec, time elapsed 11064.55 sec
epoch 12, iter 75450, avg. loss 23.60, avg. ppl 3.17 cum. examples 46377, speed 4647.36 words/sec, time elapsed 11065.96 sec
epoch 12, iter 75460, avg. loss 25.03, avg. ppl 3.23 cum. examples 46697, speed 4451.35 words/sec, time elapsed 11067.49 sec
epoch 12, iter 75470, avg. loss 21.95, avg. ppl 2.99 cum. examples 47017, speed 4543.82 words/sec, time elapsed 11068.91 sec
epoch 12, iter 75480, avg. loss 20.64, avg. ppl 3.04 cum. examples 47337, speed 4546.81 words/sec, time elapsed 11070.22 sec
epoch 12, iter 75490, avg. loss 26.54, avg. ppl 3.36 cum. examples 47657, speed 4684.08 words/sec, time elapsed 11071.72 sec
epoch 12, iter 75500, avg. loss 23.39, avg. ppl 3.12 cum. examples 47977, speed 4560.36 words/sec, time elapsed 11073.16 sec
epoch 12, iter 75510, avg. loss 22.67, avg. ppl 3.03 cum. examples 48297, speed 4795.08 words/sec, time elapsed 11074.53 sec
epoch 12, iter 75520, avg. loss 24.19, avg. ppl 3.20 cum. examples 48617, speed 4922.23 words/sec, time elapsed 11075.88 sec
epoch 12, iter 75530, avg. loss 23.56, avg. ppl 3.18 cum. examples 48937, speed 4749.59 words/sec, time elapsed 11077.25 sec
epoch 12, iter 75540, avg. loss 23.94, avg. ppl 3.27 cum. examples 49257, speed 4462.40 words/sec, time elapsed 11078.70 sec
epoch 12, iter 75550, avg. loss 26.87, avg. ppl 3.44 cum. examples 49577, speed 4462.85 words/sec, time elapsed 11080.27 sec
epoch 12, iter 75560, avg. loss 23.24, avg. ppl 3.18 cum. examples 49897, speed 4338.32 words/sec, time elapsed 11081.75 sec
epoch 12, iter 75570, avg. loss 25.23, avg. ppl 3.29 cum. examples 50217, speed 4453.90 words/sec, time elapsed 11083.27 sec
epoch 12, iter 75580, avg. loss 21.91, avg. ppl 2.83 cum. examples 50537, speed 4734.88 words/sec, time elapsed 11084.70 sec
epoch 12, iter 75590, avg. loss 24.37, avg. ppl 3.23 cum. examples 50857, speed 4818.50 words/sec, time elapsed 11086.08 sec
epoch 12, iter 75600, avg. loss 23.53, avg. ppl 3.21 cum. examples 51177, speed 4746.11 words/sec, time elapsed 11087.44 sec
epoch 12, iter 75610, avg. loss 21.21, avg. ppl 3.04 cum. examples 51497, speed 4596.64 words/sec, time elapsed 11088.77 sec
epoch 12, iter 75620, avg. loss 22.00, avg. ppl 3.13 cum. examples 51817, speed 4293.10 words/sec, time elapsed 11090.21 sec
epoch 12, iter 75630, avg. loss 25.19, avg. ppl 3.17 cum. examples 52137, speed 4806.47 words/sec, time elapsed 11091.67 sec
epoch 12, iter 75640, avg. loss 22.88, avg. ppl 3.06 cum. examples 52457, speed 4363.35 words/sec, time elapsed 11093.17 sec
```

```
Microsoft Azure Search resources, services, and docs (G+) es-Youssef.Aly20@alexu... ALEXANDRIA UNIVERSITY
Home > MyMachine | Serial console > Serial console
MyMachine | Serial console Virtual machine
? Feedback Feedback
epoch 15, iter 95810, avg. loss 24.60, avg. ppl 3.15 cum. examples 57897, speed 4856.10 words/sec, time elapsed 14026.01 sec
epoch 15, iter 95820, avg. loss 25.01, avg. ppl 3.24 cum. examples 58217, speed 4835.95 words/sec, time elapsed 14027.42 sec
epoch 15, iter 95830, avg. loss 25.64, avg. ppl 3.47 cum. examples 58537, speed 4603.28 words/sec, time elapsed 14028.86 sec
epoch 15, iter 95840, avg. loss 23.65, avg. ppl 3.20 cum. examples 58857, speed 4333.67 words/sec, time elapsed 14030.36 sec
epoch 15, iter 95850, avg. loss 24.45, avg. ppl 3.22 cum. examples 59177, speed 4619.63 words/sec, time elapsed 14031.82 sec
epoch 15, iter 95860, avg. loss 24.63, avg. ppl 3.25 cum. examples 59497, speed 4517.37 words/sec, time elapsed 14033.30 sec
epoch 15, iter 95870, avg. loss 23.94, avg. ppl 3.18 cum. examples 59817, speed 4439.70 words/sec, time elapsed 14034.80 sec
epoch 15, iter 95880, avg. loss 24.67, avg. ppl 3.09 cum. examples 60137, speed 4895.74 words/sec, time elapsed 14036.23 sec
epoch 15, iter 95890, avg. loss 23.42, avg. ppl 3.33 cum. examples 60457, speed 4462.73 words/sec, time elapsed 14037.63 sec
epoch 15, iter 95900, avg. loss 24.91, avg. ppl 3.29 cum. examples 60777, speed 4681.73 words/sec, time elapsed 14039.06 sec
epoch 15, iter 95910, avg. loss 23.62, avg. ppl 3.28 cum. examples 61097, speed 4314.11 words/sec, time elapsed 14040.54 sec
epoch 15, iter 95920, avg. loss 26.36, avg. ppl 3.49 cum. examples 61417, speed 4593.76 words/sec, time elapsed 14042.01 sec
epoch 15, iter 95930, avg. loss 23.86, avg. ppl 3.23 cum. examples 61737, speed 4483.04 words/sec, time elapsed 14043.47 sec
epoch 15, iter 95940, avg. loss 23.27, avg. ppl 3.11 cum. examples 62057, speed 4466.20 words/sec, time elapsed 14044.94 sec
epoch 15, iter 95950, avg. loss 23.92, avg. ppl 3.24 cum. examples 62377, speed 4954.79 words/sec, time elapsed 14046.26 sec
epoch 15, iter 95960, avg. loss 25.35, avg. ppl 3.41 cum. examples 62697, speed 4811.30 words/sec, time elapsed 14047.64 sec
epoch 15, iter 95970, avg. loss 24.59, avg. ppl 3.16 cum. examples 63017, speed 4558.63 words/sec, time elapsed 14049.14 sec
epoch 15, iter 95980, avg. loss 25.25, avg. ppl 3.24 cum. examples 63337, speed 4559.85 words/sec, time elapsed 14050.65 sec
epoch 15, iter 95990, avg. loss 20.92, avg. ppl 2.86 cum. examples 63657, speed 4581.77 words/sec, time elapsed 14052.04 sec
epoch 15, iter 96000, avg. loss 23.93, avg. ppl 3.16 cum. examples 63977, speed 4581.80 words/sec, time elapsed 14053.50 sec
epoch 15, iter 96000, cum. loss 24.30, cum. ppl 3.26 cum. examples 63977
begin validation ...
validation: iter 96000, dev. ppl 7.421629
hit patience 5
hit #5 trial
early stop!
```

```
bin ./en_es_data/test.es ./en_es_data/test.en outputs/test_outputs.txt --cuda1.b
[nltk_data] Downloading package punkt to /home/ykishk/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
load test source sentences from [./en_es_data/test.es]
load test target sentences from [./en_es_data/test.en]
load model from model.bin
Decoding: 100%|██████████████████████████████████████| 8064/8064 [06:19<00:00, 21.25it/s]
Corpus BLEU: 36.004354648520106
(base) ykishk@MyMachine:~/a4$
```

product attention	multiplicative attention	additive attention
Memory efficient and efficient in computation	Efficient to compute	Slower to compute

a)

ii) **Error:** I'm probably the author for children,

Reason: Name **Bolingbroke** is not defined to the NMT system

v) **Error:** bathroom in the women's room.

Solution: Use more data from different sources of collection

Reason: Maybe because of lack of understanding of currency conversion.

c)

i) $\lambda_i = 0.5 \quad i \in \{1, 2\}, \lambda_i = 0 \quad i \in \{3, 4\}$

$c_1 \rightarrow$ the love can always do

$$p_1 = \frac{0+1+1+1+0}{5} = \frac{3}{5}$$

$$p_2 = \frac{0+1+1+0}{4} = \frac{1}{2} \quad (\text{len}(c)=5, \text{len}(r)=4 \text{ shorter})$$

$$BP = 1$$

$$\text{Bleu } c_1 = BP * \exp(\lambda_1 \log(p_1) + \lambda_2 \log(p_2))$$

$$= 1 * \exp(\frac{1}{2} \log(\frac{3}{5}) + \frac{1}{2} \log(\frac{1}{2})) = 0.77$$

$c_2 \rightarrow$ Love can make any thing possible

$$p_1 = \frac{4}{5}$$

$$p_2 = \frac{1}{2}$$

$$(\text{len}(c)=5, \text{len}(r)_{\text{shorter}}=5), BP = 1$$

$$\text{Bleu } c_2 = BP * \exp(\lambda_1 \log(p_1) + \lambda_2 \log(p_2))$$

$$= 0.82$$

$\text{Bleu } c_2 > \text{Bleu } c_1$, c_2 is considered better translation.

* yes I agree, c_2 seems better dot

ii) r_1 only

$C_1 \rightarrow$ the Tore can always do

$$P_1 = \frac{3}{5}, \quad P_2 = \frac{1}{2}$$

$$BP = 1$$

$$\text{Bleu}_{C_1} = 0.77$$

$C_2 \rightarrow$ Love can make anything possible

$$P_1 = \frac{2}{5}, \quad P_2 = \frac{1}{4}$$

$$BP = 1$$

$$\text{Bleu}_{C_2} = 0.6 \left(1^* \exp \left(\frac{1}{2} \log \left(\frac{2}{5} \right) + \frac{1}{2} \log \left(\frac{1}{4} \right) \right) \right)$$

now $\text{Bleu}_{C_1} > \text{Bleu}_{C_2}$, which is not true as actually C_2 is better

iv)

Bleu score Adv:

- Fast to compute compared to human eval.

Disadv:

- doesn't consider semantics of sentence while evaluating