

進捗報告

1 今週やったこと

- Transformer の実装
- ManyThings データセットの訓練

2 Transformer の実装

Transformer [1] とは,Google が Attention is All You Need に提出したモデルである. 従来のモデルと違い,Transformer は完全に Attention で実現するモデルであるため,データの時系列を考えなしに並行作業が可能になる.

多数の Sequence To Sequence(Seq2Seq) モデルと同じ, Transformer も Encoder - Decoder 構造なモデルである. 今回の実験は Harvard(ハーバード) 大学の Transformer モデル [2] で実験する.

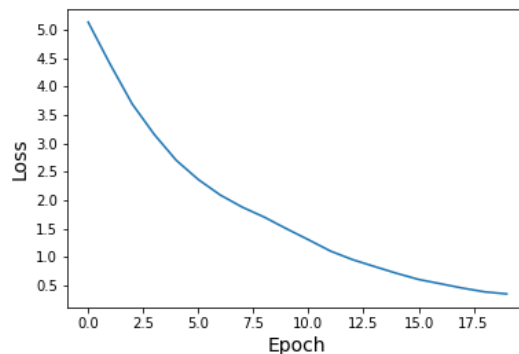


図 1: Testing Loss

3 データセット

データセットは ManyThings の英中データセットである. 今回の実験は, 中国語データセットの最小単位を漢字で試した. 表 1 に訓練データとテストデータの文章数を示す. 表 2 にデータセットのボキャブラリー数を示す.

表 1: DataSet

DataSet	Number of pairs
Train	14,533
Test	1,817

表 2: ManyThings_Words

ManyThings	ENG	CMN
Vocab	6,048	3,194

BOS Tom seems to be much UNK than me . EOS
 BOS 汤姆看来比我高兴。EOS
 translation: 汤姆看来比我高兴。

BOS There is a fork missing . EOS
 BOS 少一把叉子。EOS
 translation: 少一把叉子。

BOS He 's bound to forget . EOS
 BOS 他准会忘。EOS
 translation: 他必须拒绝。

BOS I 'm planning to study tonight . EOS
 BOS 我打算今天晚上讀書。EOS
 translation: 我打算今天晚上讀書。

BOS I was just talking about Tom . EOS
 BOS 我仅仅是在和Tom交谈。EOS
 translation: 我就是在汤姆旁边说话。

BOS Tom is a very good tennis player . EOS
 BOS 汤姆是一个非常好的網球選手。EOS
 translation: 汤姆是個很棒球選手。

4 実験結果

BatchSize 64,epoch 20 で訓練した結果, テストセットのロス は 0.35 に収束した. 図 1 にテストセットのロスの変化曲線を示す. 図 2 は翻訳の例を示す.

図 2: Example

5 実験用パラメータ

表 3 に実験用パラメータを示す.

表 3: parameters

parameter	Value
Batch_size	64
Epoch	20
n_layer	6
h_num	8
n_embedding	256
d_FF	1024
dropout	0.1
Max_length	60
Crit	LabelSmoothing

6 来週目標

- BLEU の導入
- モデルを理解する

参考文献

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, Vol. abs/1706.03762, , 2017.
- [2] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. Opennmt: Open-source toolkit for neural machine translation. In *Proc. ACL*, 2017.