
進捗報告

1 今週やったこと

- ラベル数の削減
- BERT 分類モデルの使用

2 ラベル削減

4300 漢字ラベルから 222 漢字部首ラベルに換えて, 出力ラベルを削減した.

3 BERT 分類モデルの使用

今回の実験は CKIP が訓練した BERT 分類モデルを使用した.

4 実験結果

実験結果は図 1 のように示す.

実験結果はよくないので, 現在 100 エポックで実験している.

次は seq2seq モデルで実験する.

具体的に, 入力の問題, 出力は漢字の成分 (部首と残った部分) のように, seq2seq モデルを構築する. 対照として, 入力のトークンは漢字だけ, 単語だけ, 漢字の成分だけで実験する.

5 来週目標

- seq2seq モデルで実験すること

| | |
|---|---|
| -----Epoch: 0 ----- | -----Epoch: 5 ----- |
| epoth: 0, iter_num: 100, loss: 4.3849, 49.02% | epoth: 5, iter_num: 100, loss: 4.5786, 49.02% |
| epoth: 0, iter_num: 200, loss: 4.5512, 98.04% | epoth: 5, iter_num: 200, loss: 4.4154, 98.04% |
| Epoch: 0, Average training loss: 4.4599 | Epoch: 5, Average training loss: 4.4619 |
| Accuracy: 0.1003 | Accuracy: 0.1001 |
| Average testing loss: 4.4427 | Average testing loss: 4.4433 |
| -----Epoch: 1 ----- | -----Epoch: 6 ----- |
| epoth: 1, iter_num: 100, loss: 4.2858, 49.02% | epoth: 6, iter_num: 100, loss: 4.5607, 49.02% |
| epoth: 1, iter_num: 200, loss: 4.6567, 98.04% | epoth: 6, iter_num: 200, loss: 4.5478, 98.04% |
| Epoch: 1, Average training loss: 4.4599 | Epoch: 6, Average training loss: 4.4583 |
| Accuracy: 0.1003 | Accuracy: 0.1002 |
| Average testing loss: 4.4430 | Average testing loss: 4.4429 |
| -----Epoch: 2 ----- | -----Epoch: 7 ----- |
| epoth: 2, iter_num: 100, loss: 4.7532, 49.02% | epoth: 7, iter_num: 100, loss: 4.3607, 49.02% |
| epoth: 2, iter_num: 200, loss: 4.4786, 98.04% | epoth: 7, iter_num: 200, loss: 4.2446, 98.04% |
| Epoch: 2, Average training loss: 4.4600 | Epoch: 7, Average training loss: 4.4609 |
| Accuracy: 0.1001 | Accuracy: 0.1001 |
| Average testing loss: 4.4431 | Average testing loss: 4.4431 |
| -----Epoch: 3 ----- | -----Epoch: 8 ----- |
| epoth: 3, iter_num: 100, loss: 4.4531, 49.02% | epoth: 8, iter_num: 100, loss: 4.3965, 49.02% |
| epoth: 3, iter_num: 200, loss: 4.5304, 98.04% | epoth: 8, iter_num: 200, loss: 4.4057, 98.04% |
| Epoch: 3, Average training loss: 4.4595 | Epoch: 8, Average training loss: 4.4622 |
| Accuracy: 0.1003 | Accuracy: 0.1003 |
| Average testing loss: 4.4430 | Average testing loss: 4.4427 |
| -----Epoch: 4 ----- | -----Epoch: 9 ----- |
| epoth: 4, iter_num: 100, loss: 4.4035, 49.02% | epoth: 9, iter_num: 100, loss: 4.5496, 49.02% |
| epoth: 4, iter_num: 200, loss: 4.5949, 98.04% | epoth: 9, iter_num: 200, loss: 4.0925, 98.04% |
| Epoch: 4, Average training loss: 4.4565 | Epoch: 9, Average training loss: 4.4616 |
| Accuracy: 0.1001 | Accuracy: 0.1003 |
| Average testing loss: 4.4431 | Average testing loss: 4.4420 |

図 1: 10 エポックの実験結果