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で, 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,入力は問題,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

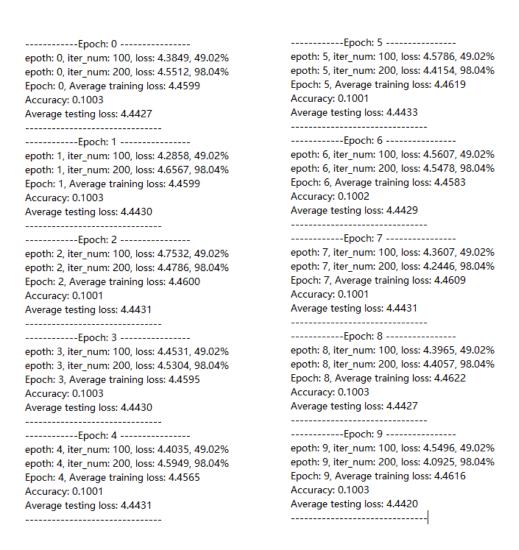


図 1: 10 エポックの実験結果