
進捗報告

1 今週やったこと

- 岡田先生にゴールデンウィーク前に貸していただいた参考書を読み進めながら基礎知識の理解に努めました。
- 感情分析タスクを PyTorch で実装しました。

2 内容について

- 実装環境は GoogleColaboratry です。
- 使用したのは東北大学の BERT で、データセットは岡田先生の github に挙げられている評判分析チェック用データを用いました。
- データセットを取り込み、ラベルデータと文章データに分けてから文章データを分かち書きし、tokenizer を用いて分かち書きしました。その際、単語長が 512 を超えるデータは除きました。
- 訓練データ数を 4511、テストデータ数を 1127 に分割し、日本語学習済みモデルとして bert-base-japanese-whole-word-masking を用いました。正答率は 0.8616 でした。
- LSTM での実装も行いましたが、正答率 0.8327 となりました。パラメータのチューニングがうまく行っていないと思われます。

3 次回までに行うこと

- 今回実装した 2 種類のプログラムでの実行結果の精度を上げる。
- 改めてもう一度参考書を読み返す。