

## 進捗報告

### 1 今週やったこと

- PyTorch を用いて多値分類タスクに取り組みました.

### 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, データセットは先週岡田先生に頂いた「楽天トラベルレビュー: アスペクト・センチメントタグ付きコーパス」データを用いました.
- プログラムは先週までに実装したものとほとんど同じで, 主な変更点は出力クラス数や損失関数を CrossEntropy-Loss に切り替えたことです.
- DataFrame で開いたデータは表 1 のようになっており, 旅行のレビュー文とその文章が朝食, サービス, 施設などといった対象に対してポジティブなことを言っているのか, ネガティブなことを言っているのかが記されている. しかし, 学習時に朝食のラベルの数字がどうしても読み込めなかったのでそれ以外の 12 のラベルを用いることにしました. データが入っていない NaN の部分は全て 0 に置き換えました.
- 表 2 は今回の実験で用いたパラメータです.
- 訓練データ数を 61300, テストデータ数を 15324 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 0.8851 でした.

表 1

テキスト	ポジティブラベル	ネガティブラベル
日本語レビュー	NaN or 1	NaN or 1

表 2

出力クラス	バッチサイズ	エポック数
12	128	5