

## 進捗報告

### 1 今週やったこと

- BERT-Transformer モデルで多値分類をした.
- 14 クラス分類では精度が向上しなかったため、朝食、夕食、風呂、サービス、立地、施設、部屋のそれぞれのカテゴリにおいて、そのカテゴリに属するか否かの 2 値分類をした.
- BERT-Transformer モデルに、各ラベルと関連度の高いフレーズを抽出するターゲットフレーズラベリング器の実装中

### 2 データセット

#### 2.1 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

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とは、楽天グループ株式会社が提供しているデータセットである。日本語レビュー文章とそれぞれの文章について、立地、部屋、食事等の 7 項目のアスペクトに対するポジティブまたはネガティブのタグが付与されている。総データ数は 76624 で、朝食、夕食、風呂、サービス、施設、立地、のポジティブ、ネガティブの 14 クラスである。今回は 14 のいずれのラベルにも属さないデータを除くことで、少なくとも 1 つのラベルに属し、語彙数が 10 以下と 100 以上のデータを取り除いたデータ群にした。総データ数は 50211 である。

### 3 実験

#### 3.1 各カテゴリ 2 値分類

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を用いて、朝食、夕食、風呂、サービス、施設、立地のそれぞれのカテゴリにおいて、そのカテゴリに属するか否かの 2 値分類をした。表 1 にデータの具体例を示す。訓練データ数を 6000、検証データを 2000 として学習をした。表 2 に訓練データと検証データに含まれる正例の数について示す。表 3 に学習時のパラメータを示す。多値分類の

時と異なるのは Transformer 層の出力数や損失関数である。

表 3: 各クラス 2 値分類のパラメータ

パラメータ	値
BERT 層の入力次元数	768
BERT 層の出力次元数	768
Transformer 層の層数	1
Transformer 層の入力次元数	768
Transformer 層の出力次元数	2
バッチサイズ	32
最適化関数	Adam
学習率	0.0001
損失関数	CrossEntropyLoss
エポック数	20

#### 3.2 実験結果

図 1 から図 14 に 7 つのカテゴリでの 2 値分類モデルを訓練した時の訓練データと検証データの損失と正解率の推移を示す。

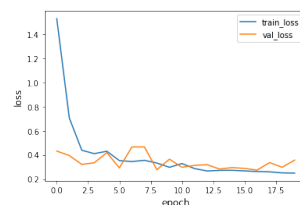


図 1: 朝食カテゴリでの損失の推移

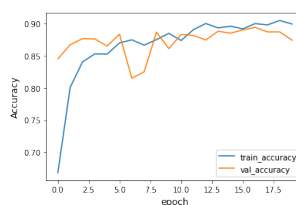


図 2: 朝食カテゴリでの正解率の推移

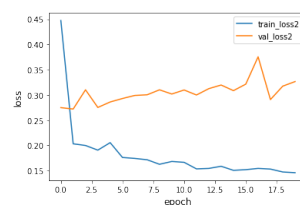


図 3: 夕食カテゴリでの損失の推移

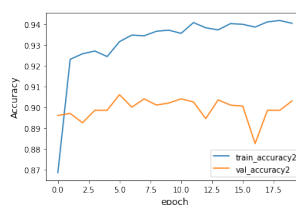


図 4: 夕食カテゴリでの正解率の推移

表 1: データの具体例

テキスト	朝食	夕食	風呂	サービス	立地	施設	部屋
お部屋も広くて、お料理もとても美味しく、部屋の露天風呂からは星がプラネタリウムのように広がっていて、とにかく最高でした。	1	1	1	1	0	1	1
部屋も綺麗で、対応もよく、朝食もおいしいのでとても満足しています。	1	0	0	1	0	1	1
気になるところは廊下の天井が低いのと、部屋数がたくさんあり、温泉が集中するとお風呂まちになるところくらいですかね。	0	0	0	0	0	0	1

表 2: 訓練データと検証データに含まれる正例の数

	朝食	夕食	風呂	サービス	立地	施設	部屋
訓練データ (6000) における正例	2832	2853	1780	3774	1313	2727	1953
検証データ (2000) における正例	1115	1557	703	1452	530	1124	771

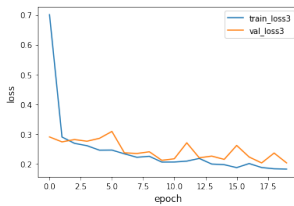


図 5: 風呂カテゴリでの損失の推移

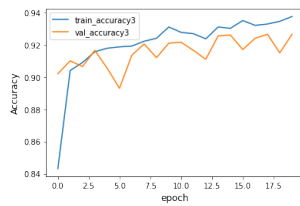


図 6: 風呂カテゴリでの正解率の推移

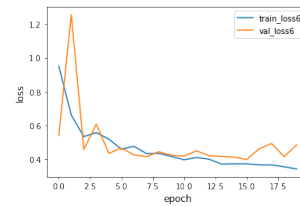


図 11: 施設カテゴリでの損失の推移

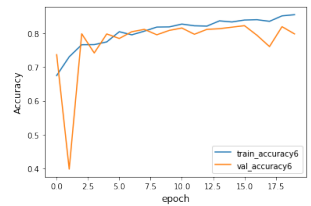


図 12: 施設カテゴリでの正解率の推移

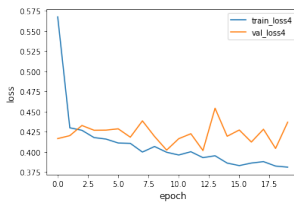


図 7: サービスカテゴリでの損失の推移

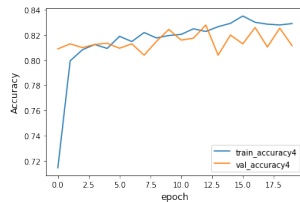


図 8: サービスカテゴリでの正解率の推移

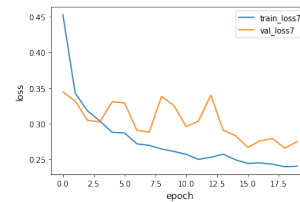


図 13: 部屋カテゴリでの損失の推移

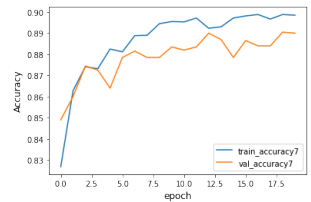


図 14: 部屋カテゴリでの正解率の推移

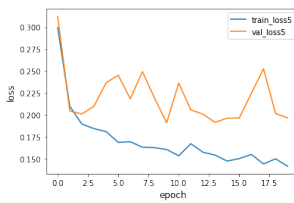


図 9: 立地カテゴリでの損失の推移

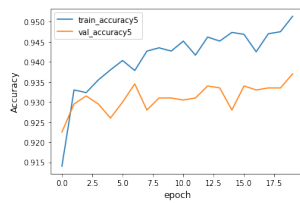


図 10: 立地カテゴリでの正解率の推移

全体的な結果から言えることは、損失の推移が依然うましくないことである。交差検証をしてどのような結果になるのかを確認する。