

進捗報告

1 今週やったこと

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。

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 4800、検証データを 1200 として 5 分割交差検証をした。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。

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	10

3.2 実験結果

現在計算途中であるため結果を記載することは出来ないが、計算が完了しているモデルのテスト時の正解率は高く出ており、混合行列を見ても多値分類時のように予測が 0 に寄っているケースも見受けられない。これらについては実験が完了次第結果と共に再記述する。

4 今後について

- BERT の最終層の Attention weight の可視化
- Transformer Encoder で得られる Attention weight の可視化

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