

進捗報告

1 今週取り組んだこと

提案モデルで得た特徴量を正解データとして, BERT モデルと楽天トラベルレビューのデータセットを用いたマルチラベル分類をした. 結果をまとめる.

2 データセット

データセットは楽天グループ株式会社が公開している「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」[1]を使用した. 楽天トラベルの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 項目のカテゴリ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. 「朝食, 夕食, 風呂, サービス, 施設, 立地, 部屋」のポジティブ, ネガティブの 14 個のカテゴリに分類される. 今回は 14 のいずれのカテゴリ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. また, いずれのカテゴリにも属さないデータと, 1 つのカテゴリにポジティブとネガティブの両方が付与されたデータを取り除いた. 総データ数は 48354 である.

3 実験

1. BERT+MLP 1 (正解データは 14 クラスのラベル): 正解データをデータセットから得られる 0,1 の 14 クラスのラベルとする場合.

2. BERT+MLP 2 (正解データは提案モデルから得られた特徴量ベクトル): 正解データを, 事前学習済みのモデルにデータセットの文章を入力して得られる特徴量とする場合.

いずれの場合も訓練データ数 30000, テストデータ数 6000 で 5 分割交差検証をした.

表 1 に実験時のパラメータを示す. 表 2 に実験結果を示す.

表 1: 2 値分類の実験時のパラメータ

パラメータ	値
BERT 層の入力次元数	100
BERT 層の出力次元数	14
バッチサイズ	4
最適化関数	Adam
学習率	0.00001
BERT 1 の損失関数	BCELoss
BERT 2 の損失関数	MSELoss
エポック数	30

表 2: 実験結果の評価指標

評価指標	Precision	Recall	micro-F1
BERT+MLP 1 (正解データは 14 クラスのラベル)	0.579 \pm 0.0359	0.775 \pm 0.0235	0.679 \pm 0.0299
BERT+MLP 2 (正解データは提案モデルから得られた特徴量ベクトル)	0.703 \pm 0.0824	0.810 \pm 0.0691	0.751 \pm 0.0343

データ数を大幅に増加することで精度の向上は確認できた.

4 今後

作成途中の, 評価語が 1 つの各文章における係り受け語をまとめたデータセットを作成する.

5 ポスター発表のまとめ

ポスター発表では提案として, 1 つのカテゴリにポジネガのラベルが両方付与されているデータセットを含めたデータセットで小分類器か, Transformer Encoder を再度学習する事が挙げられた. また, Attention の可視化精度の定量的な指標についても考えながら今後の研究を進めるべきという意見もいただいた. 今後はマルチラベル分類タスクの評価指標として, 全てのクラスで正解した指標を Exact Match Ratio として, 一部正解データにおいて, precision, recall, micro-F1 を用いることなどを検討している.

参考文献

[1] 楽天グループ株式会社. 楽天データセット (コレクション) , aug 2010.