

進捗報告

1 今週取り組んでいること

図 1 に提案モデルの概要図を示す。図 1 のモデルにおいて、CLS トークンの次元圧縮をして精度向上を試みている。また、BERT のパラメータチューニングと、チューニング後のマルチラベル分類モデルと mpm+T モデルに同じテストデータを入力して正解データの比較をする。今週中に結果が出せそうなので再度アップロードします。

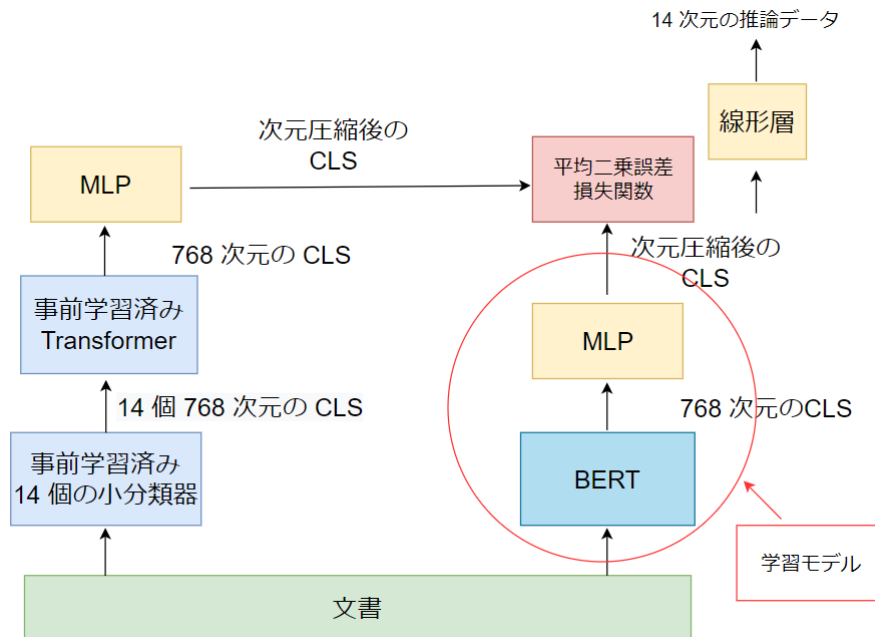


図 1: mpm+T を含めた実験モデルの概要図

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。総データ数は 50211 である。

*1 <https://www.nii.ac.jp/dsc/idr/rakuten/>