

進捗報告

1 今週取り組んでいること

3 値分類用の Attribute Classifiers の作成している.

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.

3 実験

表 1 に 5 分割交差検証時のパラメータを示す.

表 1: 朝食の Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	0.0001
損失関数	CrossEntropyLoss
エポック数	20

表 2: 夕食の Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	0.00001
損失関数	CrossEntropyLoss
エポック数	20

表 2 に 5 分割交差検証をした各評価指標の平均と分散を示す. いずれのアスペクトにおける 3 つのラベルの中でもニュートラル (2) のラベルが非常に多く, 2 値分類に比べて正解率は低下している. 損失関数の重みを訓練データに占める各ラベルの割合の逆数とした.

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

表 3: 風呂の Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	9×10^{-5}
損失関数	CrossEntropyLoss
エポック数	20

表 4: サービスの Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	1×10^{-3}
損失関数	CrossEntropyLoss
エポック数	20

表 5: 立地の Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	6×10^{-5}
損失関数	CrossEntropyLoss
エポック数	20

表 6: 立地の Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	1×10^{-4}
損失関数	CrossEntropyLoss
エポック数	20

表 7: 部屋の Attribute Classifier 3 値分類パラメータ

パラメータ	値
BERT の入力次元数	768
BERT の出力次元数	3
バッチサイズ	40
最適化関数	Adam
学習率	3×10^{-5}
損失関数	CrossEntropyLoss
エポック数	20

表 8: Entity CClassifiers 作成時に使用した正解率の平均とその標準偏差

朝食	夕食	風呂	サービス	立地	施設	部屋
0.827 ± 0.019	0.805 ± 0.023	0.848 ± 0.025	0.810 ± 0.004	0.850 ± 0.020	0.830 ± 0.007	0.813 ± 0.020