BERT を用いた深層言語処理における品詞推定の調査

1　はじめに

近年，人工知能や深層学習による創作に関する研究が盛んになっている．自然言語処理の分野においても小説の生成を目指した文や文章の生成などが研究されているが，これには画像などの生成とは異なる課題がある．その一つが，生成される文の文法情報や品詞情報が正しく得られるかどうかという課題である．文を生成するには言葉を正しくならべる必要があり，それには正しい文法や品詞の理解が欠かせない．

本実験では，最近注目を集めている深層言語モデル BERT を用いて，文生成の基礎である文の品詞の情報がどの程度獲得できているかを実験をもとに調査した．

2　要素技術

2.1　BERT

BERT (Bidirectional Encoder Representations from Transformers)[1] は，2018 年 10 月に Google の Jacob Devlin らが発表した，Transformer による双方向のエンコーダーを用いた言語モデルである．文の入力に対して，入力された文および，文に含まれる各単語に対応する分散表現を出力する．従来の言語モデルは特定の自然言語処理タスクにのみ対応しており，タスクに応じてモデルの修正が必要であった．しかし BERT はファインチューニングをすることで，モデルの修正をせずに様々なタスクに応用でき，汎化性に優れている．また当時 11 個のタスクに対して state-of-the-art を達成しており精度が高い．

BERT は事前学習モデルであり，入力文の一部の単語を ”[MASK]” 記号に置き換えてその元単語を予測する「masked language model」に基づいたタスクと，2 つの入力文に対して文の連続性を識別する「next sentence prediction」のタスクによって学習する．

本実験では，東北大学から公開されている，日本語 Wikipedia の文章を用いて事前学習されたモデル[1]を使用した．このモデルでは，入力文を MeCab[2] により形態素解析し，WordPiece という手法により subword 化している．

[1]https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/
[2]https://taku910.github.io/mecab/

3　使用データセット

本実験では，叙述的な文を対象に品詞を推定するため，毎日新聞データセット[3]の新聞記事を用いた．2008 年の 1 面に掲載された記事の中から文として成立しているものを抜き出し，それぞれの文章に対して ”□”，”×” などの記号，”＜＞” などの間に書かれた注釈等を取り除いた．

4　数値実験

4.1　実験の前処理

本実験では，文中の単語を ”[MASK]” 記号に置き換え，元単語が各品詞（名詞，動詞，形容詞）であるかどうかを正しく出力できるかを調査した．

前処理として，データセットの各文に対して分かち書きをし，単語に分割した．一文の単語の中から 1 つをランダムに選択して ”[MASK]” 記号に置き換えた．置き換えた箇所の元単語が名詞である文にラベル 1 を，そうでない文にラベル 0 を付与し，実験に用いるデータとした．また，置き換えた箇所の元単語が動詞である文にラベル 1 を，そうでない文にラベル 0 を付与し直し，実験に用いるデータとした．同様に，形容詞についてもラベルを付与し直して実験に用いるデータとした．以上の手順で名詞，動詞，形容詞の推定用のデータセットをそれぞれ 1 つずつ用意した．

4.2　実験

まず，データセットの各文に対して，データの偏りによる精度の差を避けるため，ラベル 0 とラベル 1 のデータ数を同じ 5000 個に揃えた．この各文を BERT に入力して，得られた各単語の分散表現を多層パーセプトロン（MLP）に入力して分類した．表 1 に本実験で用いた MLP のパラメータを示す．学習率は，BERT の初出論文 [1] で推薦されている値を使用した．また，データ数が少ないため，訓練データに対して 5 分割交差検証による平均値を確認して評価した．

[3]https://www.nichigai.co.jp/sales/mainichi/mainichi-data.html

表 1: 実験で用いた MLP のパラメータ

パラメータ	値
入力層の次元数	分散表現の次元数
隠れ層の次元数	768
出力層の次元数	2
エポック数	10
バッチサイズ	10
損失関数	cross entropy
最適化関数	Adam
学習率	5×10^{-5}

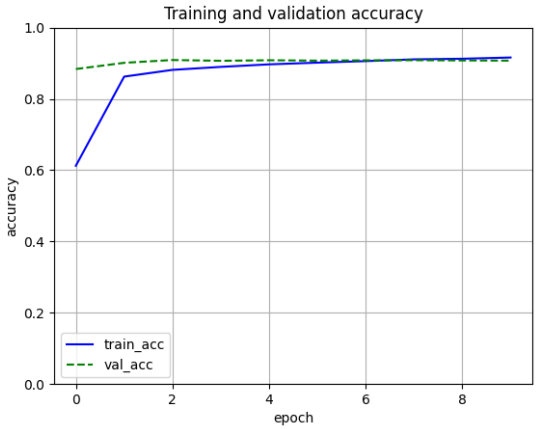

図 1: 名詞推定時の accuracy の推移

4.3 実験結果

図 1 に，学習時の accuracy と loss の推移を示す．横軸はどちらも epoch 数を，縦軸はそれぞれ accuraacy と loss を表している．表 2 に品詞の推定結果を示す．

4.4 考察

実験の結果から，名詞，動詞，形容詞のいずれにおいても，高い精度で品詞の推定ができた．

5 まとめと今後の課題

本実験では，文中の単語がある品詞かそうでないかの 2 クラスに分類し，BERT を用いた推定精度を確認した．結果として，名詞，動詞，形容詞のいずれにおいても，高い精度で品詞を推定できることが分かった．このことから BERT は，隠された単語であってもその

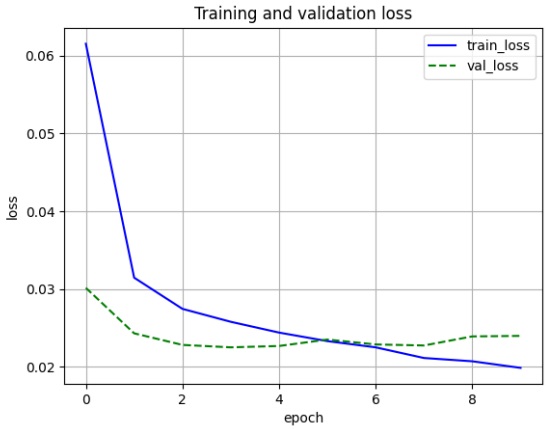

図 2: 名詞推定時の loss の推移

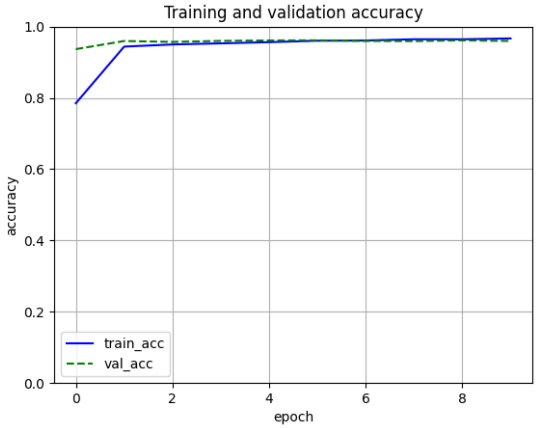

図 3: 動詞推定時の accuracy の推移

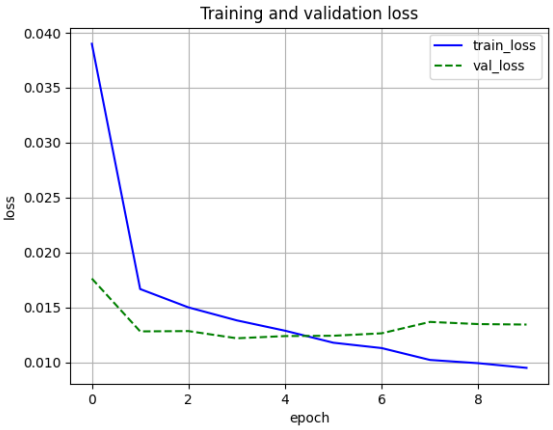

図 4: 動詞推定時の loss の推移

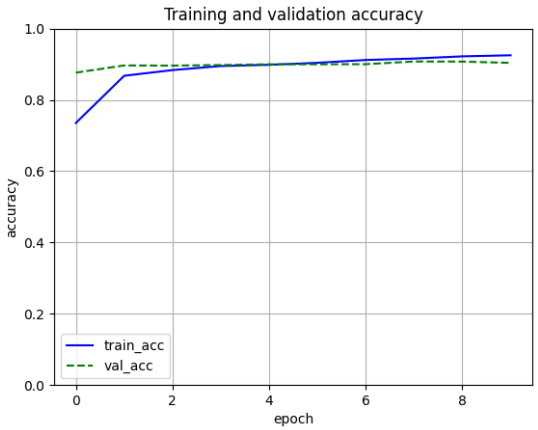

図 5: 形容詞推定時の accuracy の推移

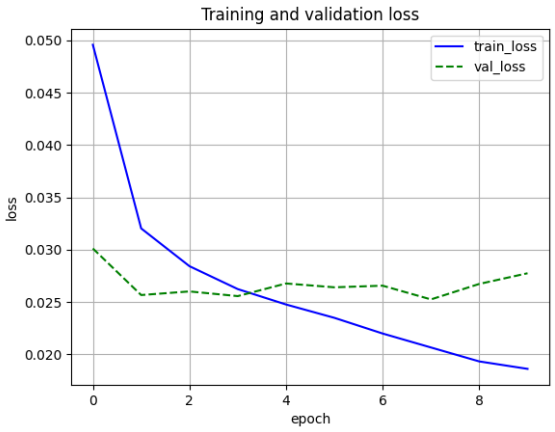

図 6: 形容詞推定時の loss の推移

表 2: 実験精度

品詞	平均値 (標準偏差)
名詞	0.914 (0.004)
動詞	0.953 (0.004)
形容詞	0.905 (0.004)
ベースライン	0.500

単語の品詞情報を獲得することができるため，適切な品詞の単語を補って文を生成することができると考えられる．

今後の課題としては以下のようなものが挙げられる．

- BERT の出力で得られる単語の分散表現を用いた，単語の意味情報を取得できているかの確認
- 一文の複数箇所を ”[MASK]” 記号に置き換えた場合にも品詞情報が正しく獲得できるかの確認

参考文献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.