

BERT を用いた深層言語処理における文法情報の影響の調査

1 はじめに

近年、機械学習は目覚ましい発展を遂げ、自然言語の分野においても大きな成果を上げている。特に自然言語処理モデルの一つである BERT は、多様な自然言語処理タスクで好成績を残しており、注目を集めている。

本実験では、BERT を用いて深層言語処理（とは？）を行い、文法情報（多分これは品詞のこと）が処理に（？）どのような影響を与えるかを確かめる（調べる？）

2 要素技術

2.1 BERT

BERT [1] は Transformer による双方向のエンコーダーである。2018 年 10 月に Google の Jacob Devlin らの論文で発表された。

本実験では、東北大学が公開している訓練済み日本語 BERT モデルを使用する。

2.2 MeCab

MeCab は、オープンソース形態素解析エンジンであり、京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクトを通して開発された。

日本語は、文中で単語が区切りなく並べられているため、BERT を使う前に文を単語に区切る必要がある。

3 実験手法

[MASK] がかかっているところの単語の品詞が当てられるかを調べたい。

- 名詞に [MASK] をかけた場合の精度を見る
- 動詞に [MASK] をかけた場合の精度を見る
- 形容詞に [MASK] をかけた場合の精度を見る
- 機能語に [MASK] をかけた場合の精度を見る
- 上記の結果を比較する

4 数値実験

4.1 実験 1

4.2 実験結果

4.3 考察

- どの品詞がうまく予測できたか
- なぜうまくいったと考えられるのか
- どの品詞がうまく予測できなかったか
- なぜうまくいかなかったのか
- じゃあどうすればうまくいくのか

っていうのを何かしら定量的にみないといけない？はずなので、比較や評価の方法を調べておく。

5 まとめと今後の課題

これからも頑張りたいです、まる。

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.