

タイトル

1 はじめに

文章の生成をしたいのでとりあえず二値分類から始める。BERT が賢すぎて嬉しい。

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自然言語処理モデルである。

使うのはたぶん東北大 BERT だと思います。[MASK] をかけることで、その単語を予測してくれます。そもそも BERT は二種類の事前学習をしていて、単語の予測と文の関係性の予測の二つを勉強していて、それによってとても賢いです。今回はその単語の予測の部分を使うのですが、事前学習とタスクが同じなので、ファインチューニングが不要ですか...? (何もわかっていない顔)

2.2 MeCab

分かち書きをしてくれました。ありがとう。形態素解析で助詞とか名詞とかを抽出して [MASK] をかけるのに使います。

3 実験手法

今回は、助詞に [MASK] をかけて、そこに入る単語が主格かどうかを判定してもらいます。他には、特定の品詞に [MASK] をかけて、それがその品詞かどうかを判定できたら嬉しいです。

4 数値実験

4.1 実験 1

パラメータがからころちん

4.2 実験結果

実験結果が出ればここに書けますが、でなければ書けません。

4.3 考察

うーん、ばなな！

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.