
12/7 ~ 12/13 報告書

1 進捗

- 夏目漱石の小説『吾輩は猫である』の文章 (neko.txt) をデータとして例として使用した
- データに対して形態素解析することで形容詞を含む文を検出し、その形容詞部分を [MASK] トークンに変換して、文頭に [CLE] トークン、文末に [SEP] トークンを付加した文をデータセットとして用いた
- 東北大学の研究室が公開している事前学習済み BERT モデルを用いて [MASK] 部分の推定を行い、上位 100 件の候補の中から形容詞のみを最大 5 個抽出して表示した
- 抽出した候補の中に [MASK] に変換される前の形容詞と一致するものがあれば 1、なければ 0 として、1 を出力する割合を出すと 0.366 であった

2 次に取り組むこと

- 元となる文に対して係り受け解析をする
- 形容詞以外の形容語 (形容動詞、副詞など) についても同様の処理を行ってみる
- 直喩法 (～のような) についても同様の処理を行ってみる
- もう少し大きなデータセットを用いる

参考文献

- [1] 鈴木勘太, 杉本徹. Encoder-decoder モデルを用いた文章表現を豊かにする執筆支援システム. 言語処理学会 第 27 回年次大会 発表論文集, pp. 1862–1866, 2021.