

# TransformerでBERTを扱う -超入門-

2022年2月3日(木)

打ち合わせ用

太田博三

# 目次

1. ポジネガ分析(文書分類)
2. 穴埋め問題: 単語(<mask>)の予測

- 本スライド:
- Colab:
- [https://github.com/otanet/hannari\\_NLP\\_hands\\_on\\_2022/blob/main/Transformer%E3%81%A7BERT%E3%82%92%E6%89%B1%E3%81%86\\_%E8%B6%85%E5%85%A5%E9%96%80\\_ipynb\\_20220203.ipynb](https://github.com/otanet/hannari_NLP_hands_on_2022/blob/main/Transformer%E3%81%A7BERT%E3%82%92%E6%89%B1%E3%81%86_%E8%B6%85%E5%85%A5%E9%96%80_ipynb_20220203.ipynb)

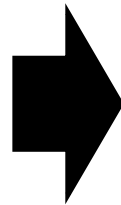
# 1. 感情分析(ポジネガによる文書分類)

1. “(Hugging Face )Transformers”のlibraryを用いて、感情分析を行います。
2. ここでの感情分析はポジティブかネガティブかの2値分類となります。

入力:

“I love you”

“I hate you”



出力:

POSITIVE, score:  
0.9998

NEGATIVE, score:  
0.9991

# 1. 感情分析(ポジネガによる文書分類)

1. “(Hugging Face )Transformers” のlibraryを用いて、感情分析を行います。
2. ここでの感情分析はポジティブかネガティブかの2値分類となります。

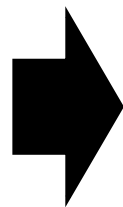
---

## 英語での感情分析の例

入力:

“I love you”

“I hate you”



出力:

POSITIVE, score:  
0.9998

NEGATIVE, score:  
0.9991

# 1. 感情分析(ポジネガによる文書分類)

## 日本語での感情分析の例

入力:

この商品を買って  
よかった。



出力:

ポジティブ, score:  
0.98984

この商品を買って  
失敗した。

ネガティブ, score:  
0.80741

# 1. 考察:感情分析(ポジネガによる文書分類)

## 日本語での感情分析の例

入力:

この商品を買ってよかった。

この商品を買って失敗した。



出力:

ポジティブ, score: 0.98984

ネガティブ, score: 0.80741

※考察

(応用: 拡張性)

ポジティブが0.7を超えたら、ポジティブ、0.3まではネガティブ、0.3-0.7はニュートラルなど、3つ以上に分類ができる。

(課題)

単体の1つの文での想定になる。元データ(学習データ)に依存する分、当てはまらないこともあり得る。

従来からある比較対象: ナイーブベイズ分類器やSVM

## 2. 穴埋め問題(単語の予測問題)

入力:  
朝食に[MASK]を  
食べる。



出力:

score: 0.09675

パン', 'sequence':

朝食にパンを食べる。

score: 0.0734

米', 'sequence': '朝食に米を食べる。

score: 0.063

朝食', 'sequence': '朝食に朝食を食べる。

score: 0.0531

カレー', 'sequence': '朝食にカレーを食べる。

score': 0.0523

ワイン', 'sequence': '朝食にワインを食べる。

score: 0.063 朝食', 'sequence': '朝食に朝食を食べる。

⇒日本語的に、おかしいのも出力される！！致命傷になりうるレベルも出力される。

# まとめ

1. Hugging FaceのlibraryのTransformerでBERTを体感してみた。

2. 2つのタスクを行った。

1)感情分析

2)穴埋め問題

3. 1)感情分析は確率の設定で2値分類を3つ以上に拡張できそう。

2)穴埋め問題は、候補の一覧から選択するところで、カスタマイズして拡張できそうだ。日本語としてあり得ない出力を抑える必要がある。