フェイクニュースの検出プログラム

2-4-32　寶木　隆正

研究背景

フェイクニュースがSNSを通して拡散されている。

機械学習によって見分けることができるのか。

研究方法

日本語フェイクニュースデータセットを使用

https://github.com/tanreinama/Japanese-Fakenews-Dataset

データセットを学習用とテスト用に分け学習を行った後テストを行いモデルを評価する。

分類にはランダムフォレストを使用する。

データ概要

コラム名	意味
id	ユニークid
context	記事の文章
isfake	記事がフェイクであるかどうか 0: オリジナル, 1:部分的にフェイク, 2:完全にフェイク
nchar_real	記事の中の、人間が執筆した部分の文字数
nchar_fake	記事の中の、モデルが生成した部分の文字数

分析方法

- ランダムフォレストでの学習を行うためデータ中の文字列を数値データに変換する。

- isfakeの値が0,2の時と0,1,2の時とで精度を比較する。

結果

	precision	recall	F1-score	support
0	0.79	0.83	0.81	1085
2	0.87	0.83	0.85	1422
accuracy			0.83	2507
macro avg	0.83	0.83	0.83	2507
weighted avg	0.83	0.83	0.83	2507

表 1 . isfake 0, 2

- モデルは全体として83%のAccuracy(正解率)を達成した。
- Precision(適合率)とRecall(再現率)がどちらも高いため、誤分類が少ない

結果

	precision	recall	F1-score	support
0	0.63	0.63	0.62	1124
1	0.47	0.47	0.47	1405
2	0.64	0.66	0.65	1383
accuracy			0.58	3912
macro avg	0.58	0.58	0.58	3912
weighted avg	0.58	0.58	0.58	3912

表 2 . isfake 0, 1,2

- Accuracy(正解率)は58%
- クラス1（部分的にフェイク）の分類が難しい

- テストの割合を減らせば正解率を上げることはできた。

Transformer

- Transformerは、Googleが2017年に発表した自然言語処理のモデル

- Sentence Transformersを使用して文章をベクトル化し、学習させる

```
from sentence_transformers import SentenceTransformer
model = SentenceTransformer("all-MiniLM-L6-v2")
X_train_transformers = model.encode(X_train)
X_test_transformers = model.encode(X_test)
```

結果

	precision	recall	F1-score	support
0	0.47	0.30	0.36	391
1	0.37	0.41	0.39	436
2	0.57	0.68	0.62	477
accuracy			0.47	1304
macro avg	0.47	0.58	0.58	3912
weighted avg	0.58	0.58	0.58	3912

- Accuracy（正解率）：モデル全体の正解率は0.47
- クラス０、クラス１が再現率、適合率どちらも低い

- Isfakeが 0 と 2 のみの分類では精度が高い

- テキストデータをベクトル化する方法として『TfidfVectorizer』を用いた。Isfakeが0,1,2の時を判定すると精度が悪くなる

- 『SentenceTransforme』を用いて精度が上がると思ったが、下がった。

- 理解が浅いため、深いところまで追求できなかった。