

進捗報告

1 今週やったこと

- BERT による人間らしい文章の判定
- GPT-2 の学習データセットの影響

2 BERT による人間らしい文章の判定

表 1 に rinna のパラメータで生成した文章を 100 文生成し, BERT によって分類した出力から最も漫画データに近い結果と遠い結果を示した.

この結果から "ブログ" や "「」『』" などの漫画にはない表現から分類していることが考えられる.

3 GPT-2 の学習データセットの影響

追加学習するデータセットによる結果やデータセットの順番の影響を調査する実験をした. データセットは, 新聞データセットと 4 コマ漫画セリフデータセットの 2 つ. ベースとして rinna のパラメータを使用した.

1. news
2. manga
3. news → manga
4. manga → news

の 4 条件で実験した. 表 2, 3, 4, 5 に結果を示した. 表 2, 3 の結果から学習した直前のデータセットに強く影響を受けることがわかった. 直前に新聞データセットを学習した表 2, 5 の文章は新聞らしく長くなっている. 表 3 はテストデータセットが同じであるためか, 複数回の結果でほとんど同じ出力となっていた. 表 4 は新聞データセットから漫画セリフデータセットを学習した結果であるが, 新聞データを学習したことで多様性のある内容となりながら, 漫画の表現を模倣することができている.

4 今後の予定

-

表 1: 最も漫画データに近い結果 (=1) と遠い結果 (=0)

(index)	serif	serif_gen	score
98	# 1 大悪魔のように	ととはにのさでとなのや? とはないでとなと	0.7606968879699707
55	# 1 大悪魔のように	_ ホーム _ 「イニシャル・ワン」で検索したらすぐに出てきたわ..... : 今日もブログを読んでくださって、ありがとうございました	0.0094498405233025
123	エロエムエッサイム我は求め訴えたり ..	女 僕は女さんが一番ですよ	0.9308531880378724
120	エロエムエッサイム我は求め訴えたり ..	」 『エロエムエッサイム我』って言葉を聞いたのは、もうずいぶん昔のような気がするんだがなあ	0.0082009891048073
286	我 梢つつじの名において命じる!	我 己を愛するのぞやな	0.9493744373321532
208	我 梢つつじの名において命じる!	いやな思いをしないで帰って来ることね」 「あのね、私はただ、愛することだけを願ってるわけじゃないのよ	0.0100914966315031
375	いでよ! 悪魔っ	な どう	0.9812219738960266
378	いでよ! 悪魔っ	俺は何もしてないよ!」 というわけで、このブログを見てくださっている方々にはご心配おかけして申し訳ございません	0.0080730495974421
499	我を呼んだのは貴様か.....	と	0.937634825706482
428	我を呼んだのは貴様か.....	> 「私達も、ここに来るのがはじめてですねえ.....」 > 「私達みたいな、どこにでもいる普通の人間なんだろうなあ」 > 「それなら	0.0065410989336669
598	あれ?	なんだろう	0.9760836362838744
550	あれ?	最近、ここを見ていないけど、どうしたのかな? なんか、いろいろあったけど、おめでとうございませう~ ところで、今日、ブログで「東京では、もう桜が咲いている	0.0058337575756013
624	禁句	自分の本気を他人に見せるのが怖い	0.947410762310028
675	禁句	私はこの話を聞いて、まだ自分の子どもの頃の記憶を思い出し、涙があふれてきました	0.0125555805861949
719	我は悪魔 ..	今日も元気にやってます~	0.9658565521240234
713	我は悪魔 ..	』... この絵を見ただけで、そう思います (^ ^ > 「お前は悪魔だとかいう、言われて嫌な気持ちになる人は、この絵を見ない方が」...	0.0081411600112915
849	証拠	浮気不倫調査は探偵はバレるものと東京都北区に知っておきましょう	0.90934956073761
837	証拠	#501: もしかしたら、私が悪かったのかもしれませんが、私は、この前 #349 先生と電話したばかりなのに、もう少し頑張ります	0.0096263475716114
963	ホントに悪魔なの?	俺は悪魔じゃない	0.9663339853286744
992	ホントに悪魔なの?	こんな人が、何万年もの時空を超えて、今ここに、おられるんだ~って、ただただ感謝するしかないです	0.0067421859130263

表 2: 学習データ news

	serif	serif_gen
0	# 1 大悪魔のように	」..... 神戸市出身
1	# 1 大悪魔のように	弾力性ある国に未来はない
2	# 1 大悪魔のように 埼玉県出身
10	エロエムエッサイム我は求め訴えたり ..	陵墓の百合耐えて ... 百合苦しんで ...
11	エロエムエッサイム我は求め訴えたり ..	だが、営々と交わした契約書には上限があって「月額 60 万円」「年収 200 万円」など、運営資金に充てるための数値が並んだ
12	エロエムエッサイム我は求め訴えたり ..	熟れた豊乳と豊顔、巨体を今度はどうしようか
20	我 梢つつじの名において命じる！	我 つずつ山ほたる名において命じる！ 我 ずつ山ほたる名において命じる！ 我思う
21	我 梢つつじの名において命じる！	我 瘍 つ つ じ の 名 に お い て 命 じ る!..... 埼玉県生まれ
22	我 梢つつじの名において命じる！	我 ブックマーキング 我慢強く踏ん張れ！ 我 銘なる 我 候 我われればこそ 買ってはならない」
30	いでよ！ 悪魔っ	野太い声が響く
31	いでよ！ 悪魔っ	でよっ
32	いでよ！ 悪魔っ	でよっ
40	我を呼んだのは貴様か.....	貴様こそ敵に塩を送るのだ、と声を掛けると、薄っすらと涙をにじませた
41	我を呼んだのは貴様か.....	私を貴様大学に入れた貴様らは私の敵だ
42	我を呼んだのは貴様か.....	我を呼ぶのは貴様か.....
50	あれ？	これも京都の風景だったか、と思う
51	あれ？	あれ？ あれ、これも？ いまさらながらに、そうねえかもしれませんが
52	あれ？	3 月 19 日、宮城県南三陸町を訪ねた

表 3: 学習データ manga

	serif	serif_gen
0	# 1 大悪魔のように	エロエムエッサイム我は求め訴えたり ..
1	# 1 大悪魔のように	エロエムエッサイム我は求め訴えたり ..
2	# 1 大悪魔のように	エロエムエッサイム我は求め訴えたり ..
10	エロエムエッサイム我は求め訴えたり ..	我 夢野アキラの名で命じる!
11	エロエムエッサイム我は求め訴えたり ..	我 夢野アキラの名で命じる!
12	エロエムエッサイム我は求め訴えたり ..	我 夢野アキラの名で命じる!
20	我 梢つつじの名において命じる!	いでよ! ハムル様ー!
21	我 梢つつじの名において命じる!	いでよ! 悪魔っ
22	我 梢つつじの名において命じる!	いでよ! 悪魔っ
30	いでよ! 悪魔っ	だから 1 年生の間に勉強しときゃよかったんだ
31	いでよ! 悪魔っ	ぼくを持ち上げろ
32	いでよ! 悪魔っ	っていうと
40	我を呼んだのは貴様か……	どうしたのだ?
41	我を呼んだのは貴様か……	あれ?
42	我を呼んだのは貴様か……	ち 違いますっ
50	あれ?	石焼きイモや!!
51	あれ?	もしかしてノストラダムス信じてます?
52	あれ?	すりりんごと桃缶とヨーグルトのせいじゃない?

表 4: 学習データ news → manga

	serif	serif_gen
0	# 1 大悪魔のように	正義の味方なるもの悪くないよね
1	# 1 大悪魔のように	つつじ大悪魔
2	# 1 大悪魔のように	正義の味方なるもの
10	エロエムエッサイム我は求め訴え たり ..	我 我 悪魔のように
11	エロエムエッサイム我は求め訴え たり ..	我 悪魔操り人形
12	エロエムエッサイム我は求め訴え たり ..	我 梢つつじの名において命じる!
20	我 梢つつじの名において命じる!	私の望みかなえてよ
21	我 梢つつじの名において命じる!	貴様の望みが!!わぬと!!
22	我 梢つつじの名において命じる!	いでよ!ハムちゃん
30	いでよ! 悪魔っ	大悪魔だもんねっ
31	いでよ! 悪魔っ	そりゃそーですよ
32	いでよ! 悪魔っ	風邪薬切れてた?
40	我を呼んだのは貴様か……	貴様の魂と引き換えに何でも望みを!!
41	我を呼んだのは貴様か……	断る
42	我を呼んだのは貴様か……	ハムル様と戦うなんてできません——～
50	あれ?	あたし用事あるんでえー
51	あれ?	原くんの希望が私と同じ高校だよー～
52	あれ?	風邪薬切れてた?

表 5: 学習データ manga → news

	serif	serif_gen
0	# 1 大悪魔のように	政治も経済も今のシステムを続ければ「バケツの水をざるに流す」のような生き方はできない
1	# 1 大悪魔のように 、20181 大団体戦のように
2	# 1 大悪魔のように	機械化 1 大趨のように
10	エロエムエッサイム我は求め訴えたり ..	陵代行さまは「こんな厳しい世界があるとは思わなかった
11	エロエムエッサイム我は求め訴えたり ..	ナンカシー、ナンカシー、エッチなアブデル・エルハジを思え、我は求めるべき我我を献ずり女性の怒りの琴
12	エロエムエッサイム我は求め訴えたり ..	官学タフネス鈴木があこがえた「東京の人びと」を、邦訳した一冊
20	我 梢つつじの名において命じる！	我 マーティンつつじの名において命じる！ 我われらの過ちを直せ!.....
21	我 梢つつじの名において命じる！	我 匡つつじの名において命じる！ 我われを苦しめる負債を断ち切り、無念の決別を果たさなければならない
22	我 梢つつじの名において命じる！	我 衣笠つつじの名において命じる！
30	いでよ！ 悪魔っ	で、なぜそんなことをするのか？ まあ、ご心配には及びません
31	いでよ！ 悪魔っ	ともあれ面白かった
32	いでよ！ 悪魔っ	でしたと、新聞が書けというのである
40	我を呼んだのは貴様か.....	だが貴様は今、貴様の手にそれを望むか、貴様の胸に手を差し伸べ、貴様が満足するように努めるか、まず選択を留保せねばならない
41	我を呼んだのは貴様か.....	貴様は去れ〜っしや」私「貴様もおっしやらず」貴様「私の従うべき道は、後に続く者よ」
42	我を呼んだのは貴様か.....	貴様も国に教にしておれば、この苦難を乗り越れると所を しておるのだ.....」 貴様も国に教にしておれば、この
50	あれ？	これはどういうことだろう
51	あれ？	あれ？ あれ？ あれ？ あれ？ ああ、これも、今の段階では、とんでもない失礼だろうか
52	あれ？	「うわさは聞くが.....」 東京電力福島第1 原発事故は収束の見通しが立たず、放射性物質による汚染が急速に拡大している