

進捗報告

1 やったこと

BERT を用いて, セリフからキャラクターを分類するタスクに取り組んだ。

2 実験

村田君の JSAI の論文を参考に, 「キャラクターらしさ」を定量化する指標として BERT によるランク付けが有効だと考えた. 実験で対象とするキャラクターは主人公以外のメインキャラクター 8 人とした.

表 1: メインキャラクター 8 人の名前と特徴

キャラクター名	特徴
相沢 ちとせ	ノリが良く関西弁を話す
和泉 穂多流	物静かで思慮深い
河合 理佳	精神的に幼い
牧原 優紀子	内気で心優しい
御田 万理	勝ち気で意地っ張り
神条 芹華	クールでぶっきらぼう
橘 恵美	おしとやかで敬語を話す
渡井 かずみ	元気で明るい

2.1 データセット

train, valid, test をそれぞれ 8: 1: 1 の割合に分割してあるデータセットをもらった. 村田君の論文を参考に, シナリオデータに含まれる外字を表す「gaiji」, 「3000」, 「@」など学習の妨げになる文字列を除去した. また, データセットを見たところキャラクターのセリフ内に//を入れて開発の方の修正文が書かれているセリフがあったため, それらはデータセットから除外した.

データセット内の // が含まれているデータ例

え、えっと、えっと、んっと…。//00/07/26 修正そろそろロジに戻りましょ…;0,0,0,0,1,0,0,0,0

表 ?? に上記の前処理をした結果を示す.

2.2 BERT 分類器

事前学習済みモデルとして, 1G の研究で良く用いられている東北大学の乾研究室によって公開されているモデルを使用した. BERT の入力へは, 村田君と同様に以下の手順で BERT に入力できる形に変換して処理した.

1. BERT Tokenizer を用いて ID へ変換
2. 最後に SEP トークン, 初めに CLS トークンを追加し, special token を追加

表 2: データセット内のデータ数

キャラクター名	train	valid	test
相沢 ちとせ	4470	589	580
和泉 穂多流	3505	398	367
河合 理佳	4576	525	625
牧原 優紀子	4253	551	678
御田 万理	3801	546	640
神条 芹華	3678	527	426
橘 恵美	3827	562	429
渡井 かずみ	3008	395	360
合計	31118	4093	4145

3. データセット内の最長の文を基準としてトークン列の長さを固定長として, 足りない部分は Pad トークンで埋め合わせ

4. Pad トークン进行处理させるために Attention Mask を作成

また, 表 3 に学習時のパラメータを示す

表 3: BERT 分類器の学習パラメータ

パラメータ	値
最適化関数	AdamW
バッチサイズ	32
エポック数	5
損失関数	Cross Entropy Loss
学習率	1.00×10^{-4}

2.3 結果

図 1 に学習時の loss の推移を示す. また, 表 4 に各エポック終了時の valid データの accuracy を示す. test データでの正解率は 4145 データ中 3526 件分類に成功しており, accuracy は **0.8507** となった.

表 4: 各エポック終了時点の valid データの正解率

エポック数	valid データの正解率
1	0.7794
2	0.7933
3	0.8158
4	0.8283
5	0.8263

2.4 考察

表 5 に test データにおける各キャラごとの accuracy を示す.

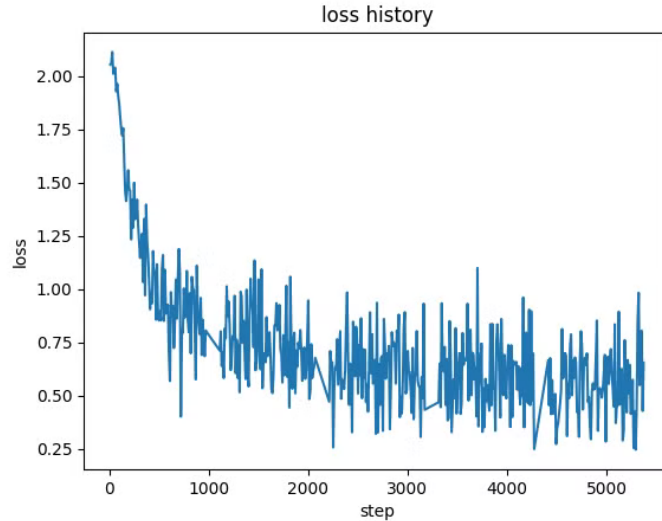


図 1: 学習時の loss の推移

表 5: test データにおける各キャラごとの分類正解率

キャラクター名	全てのデータ	分類に成功したデータ	accuracy
相沢 ちとせ	580	556	0.9586
和泉 穂多流	367	290	0.7902
河合 理佳	625	521	0.8336
牧原 優紀子	678	544	0.8324
御田 万理	640	525	0.8203
神条 芹華	426	388	0.9108
橘 恵美	469	462	0.9851
渡井 かずみ	360	240	0.6666

関西弁のセリフが多い相沢や敬語が多い橘, 男勝りな神条といったわかりやすくしゃべり方に特徴のあるキャラクターの accuracy は 0.9 を超える非常に高い結果となった。

相沢データ例 (分類成功) —

今日は思いっきり遊ぶで～～うち、バッチリ盛り上げるからな

橘データ例 (分類成功) —

実は..... 私も力及ばず補習になってしまいました

神条データ例 (分類成功) —

けど、お前も変なヤツだよ あたしなんかこうして話してて楽しいかい?

3 今後やること

- パラメータのチューニング

学習率を $1e-5$ とした際には, accuracy は 0.8191 となっており, パラメータによって精度がかなり変わって

くるタスクであると感じている。そのため, Optuna などを用いた最適化をすることでより満足できる精度の BERT 分類器が得られる可能性が高い。

- テキストデータを用いた LLM のファインチューニング

ChatHaruhi [1] などの先行研究やローカル LLM のファインチューニングの本を研究室の金で買っていたのでそれを参考に来週の研究会までに動かしてみる。

- BERT の重み保存, Attention の可視化まだファインチューニング済みの BERT の重みの保存ができてない, この後ファインチューニングした LLM のテキストを投げてどのキャラかの指標にする予定なので重みが保存できるように実装する。また, より深い考察のために田中 (寛人) さんが研究会の資料でやっているように Attention を可視化して表示できるようにする。

参考文献

- [1] Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, Linkang Zhan, Yaokai Jia, Pingyu Wu, and Haozhen Sun. Chathaviving anime large language model, 2023.