

進捗報告

1 あらすじ

テーマは「自然言語処理と深層学習に基づいた 4 コマ漫画のセリフの感情推定」前回からのタスクは BERT[1] の fine-tuning バグ探しとやり直し, 及び作成した Doc2Vec モデルとの比較.

表 1: MLP パラメータ

	MLP
(in,hidden,out)	(300,30,2)
dropout rate	0.5
activation function	tanh

表 2: 学習パラメータ

	実験
epoch	200
batch size	16
loss function	Cross Entropy Loss
optimizer	Adam

2 進捗

- BERT の fine-tuning やり直し, 及び作成した Doc2Vec モデルとの比較

3 BERT fine-tuning

3.1 実装の手直し

先週, 実装したものにおいて, optuna でハイパーパラメータを最適化する際に, ネットワークのパラメータが引き継がれていたままトライアルが進んでいたため, Train 関数が呼ばれるたびにネットワークを宣言することで解決した.

3.2 optuna 目的関数

検証用データセットにおける正例の F1 値 が最大を取るときの値を, そのトライアルの評価値とし, この評価値が最大となるトライアルにおけるハイパーパラメータを探索結果とした. ハイパーパラメータをチューニングしたのちに, その値でネットワークを学習させ, 検証用データセットにおける正例の F1 値 が最大となるエポックを採用モデルとし, テストデータセットを用いて性能を確かめた.

4 実験設定

本実験では, 各タッチについて感情推定を行った. 本実験で使用するデータセットは全 7 種類の感情ラベル (ニュートラル, 驚愕, 喜楽, 恐怖, 悲哀, 憤怒, 嫌悪) を持っているが, データ数と解析の難しさの問題から, 今回は喜楽のみを正例, その他を負例とする 2 クラスに分類した.

訓練用及び検証用データは各タッチの前半 1 話から 5 話までの拡張されたセリフを用い, 訓練用データの内 20 % をサンプリングして検証用データとした. 評価用データは後半 6 話から 10 話におけるオリジナルのセリフのみを用いた.

分散表現化には以下の 4 パターンを用い, 識別器としては Doc2Vec においては 3 層 MLP, BERT では 1 層の全結合層を用いた. MLP のパラメータを図 1 に示す.

- Doc2Vec (manga109)
- Doc2Vec (wiki)
- Doc2Vec (manga109 + wiki)
- BERT pretrained model¹

入力 は 1 つのセリフ文, 出力は感情ラベル (0: '喜楽', 1: 'その他') であり, またクラス重みとしては訓練用データセットにおける正規化されたラベル数逆比を用いた. 実行ソースコードは "lg-hub/takayama/src/ex_2020_5_12.py" 及び "ex_2020_5_12.2.py" である.

学習パラメータを表 2 に示す. 学習率は optuna によって最適化されたものを用いた.

¹<http://nlp.ist.i.kyoto-u.ac.jp>

5 結果

実験の結果を表 3 に示す。表より、BERT fine-tuning は正例に偏るか負例に偏ってしまっており、どの d2v モデルよりも劣った結果となった。またギャグタッチ以外では loss の減少は見られず、学習が進んでいるとは言えなかった。この原因としては、全結合層だけでは表現しきれなかった可能性や optuna のパラメータ候補の範囲が不適切である可能性等が考えられる。d2v モデルの比較では平均は manga109 より wiki が若干勝っているといえるがタッチごとに見ると一概にそうとは言えない。manga109 + wiki で性能が上がったタッチもあったが、こちらも一概には言えない。

計算時間は BERT が約 15 sec / epoch, d2v が約 1.5 sec / epoch であった。

6 今後の実験予定

- 識別器をそろえて再実験。
- optuna でのハイパーパラメータの値の範囲や試行回数について考察。
- 正例に指定する感情ラベルを変えて実験。(ニュートラル, 驚愕, 喜楽)
- 直前 $n - 1$ 文を考慮した n 文を入力して末尾入力の感情推定をする。モデル図は 図 1

7 その他 課題

- 森先生と大工大の上野先生に 30 話までであるらしい追加データをお願いする。
- Data Augmentation の手法の改善案。

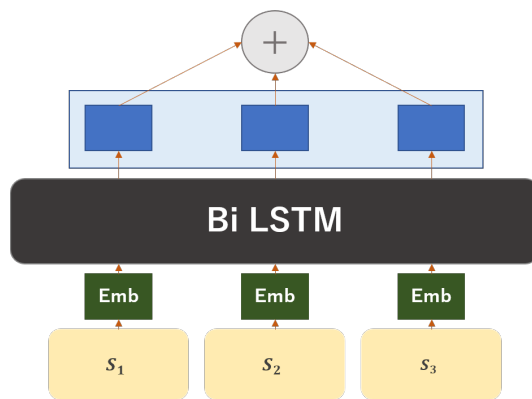


図 1: seq experiment net

参考文献

- [1] Chang M.-W. Lee K. Devlin, J. and K Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*, 2018.

表 3: result

model	ギャグ			少女漫画			少年漫画			青年漫画			萌え系			5 タッチ平均		
	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1
d2v (manga109)	0.652	0.300	<u>0.207</u>	0.537	0.632	0.608	0.625	0.667	0.400	0.662	0.357	0.312	0.563	0.364	0.364	0.608	0.464	0.378
d2v (wiki)	0.652	0.200	0.148	0.433	0.447	0.472	0.781	0.500	<u>0.462</u>	0.815	0.643	<u>0.600</u>	0.531	0.409	0.375	0.642	0.440	<u>0.411</u>
d2v (manga109 + wiki)	0.621	0.200	0.138	0.612	0.579	<u>0.629</u>	0.750	0.250	0.273	0.708	0.429	0.387	0.531	0.455	0.400	<u>0.644</u>	0.382	0.365
BERT fine-tuning	0.803	0.000	0.000	0.433	0.000	0.000	0.188	1.000	0.316	0.215	1.000	0.354	0.344	1.000	<u>0.512</u>	0.397	0.600	0.236
ベースライン	0.85	0	0	0.43	0	0	0.81	0	0	0.78	0	0	0.66	0	0	0.71	0	0

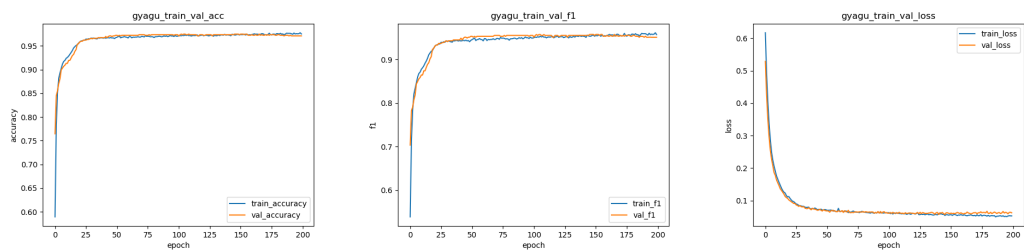


図 2: `gyagu.d2v_manga109`

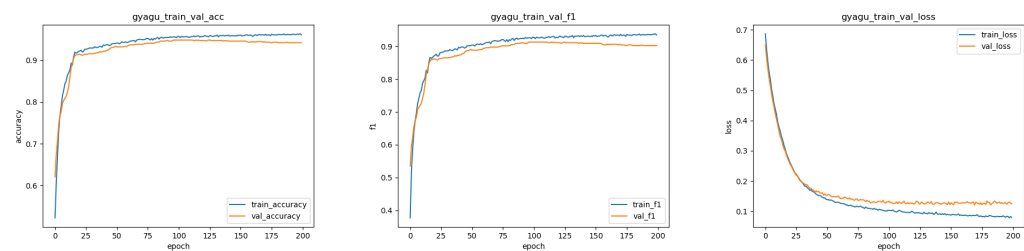


図 3: `gyagu.d2v_wiki`

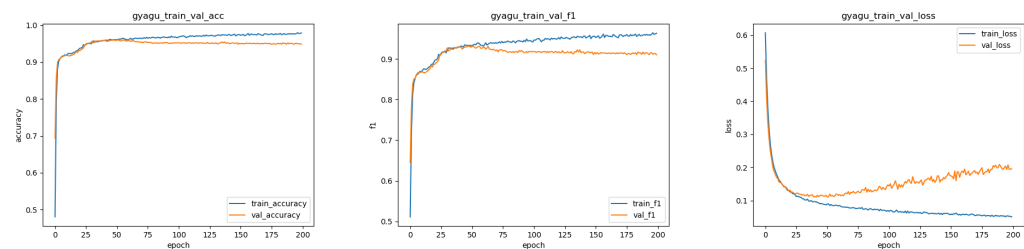


図 4: `gyagu.d2v_manga109_wiki`

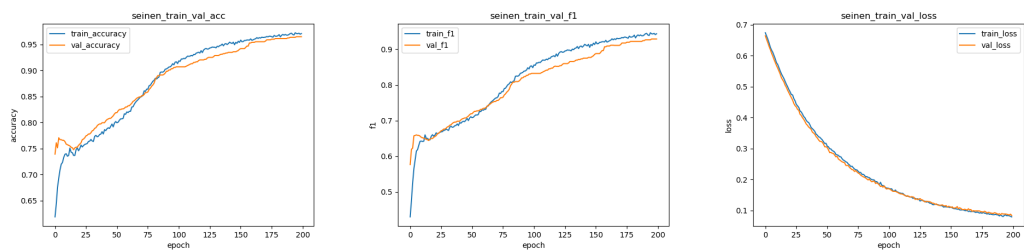


図 5: seinen_d2v_manga109

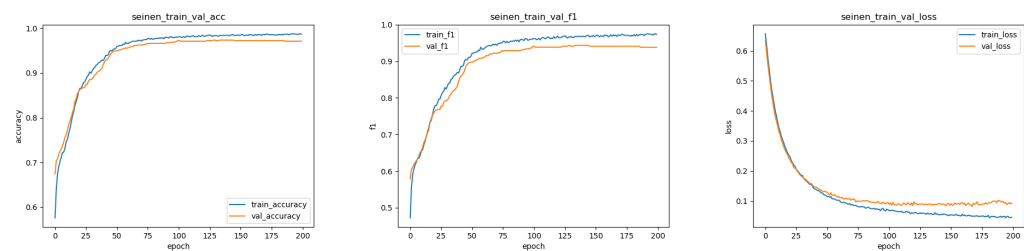


図 6: seinen_d2v_wiki

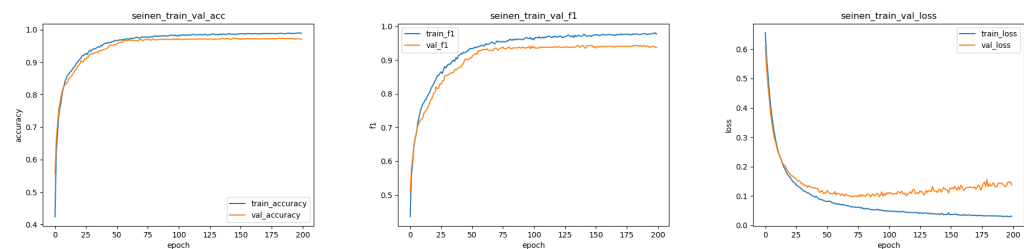


図 7: seinen_d2v_manga109_wiki

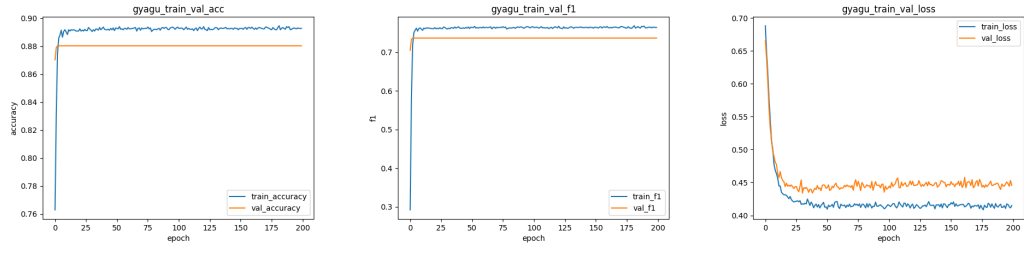


図 8: gyagu_bert

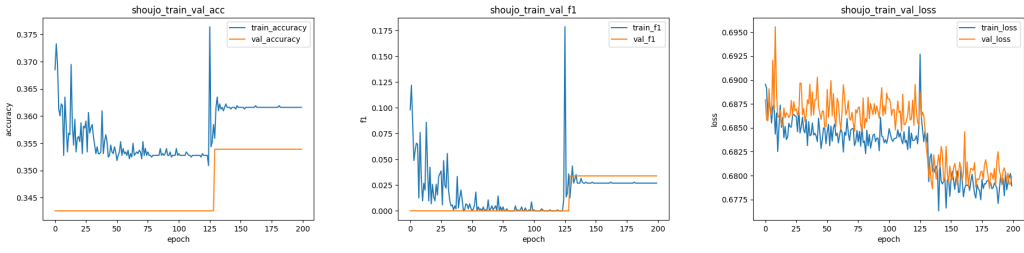


図 9: shoujo_bert

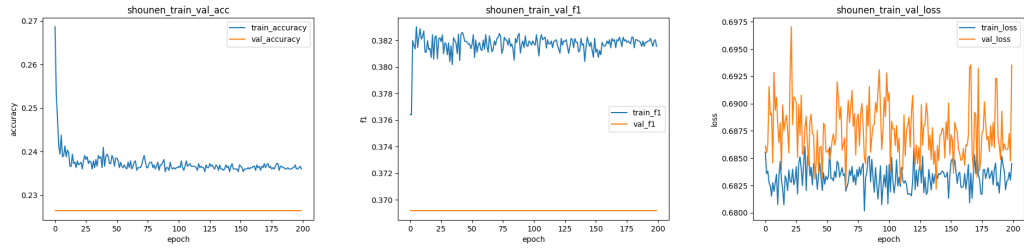


図 10: shounen_bert

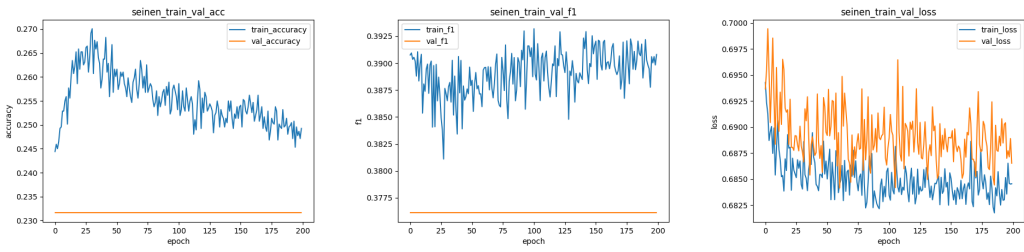


図 11: seinen_bert

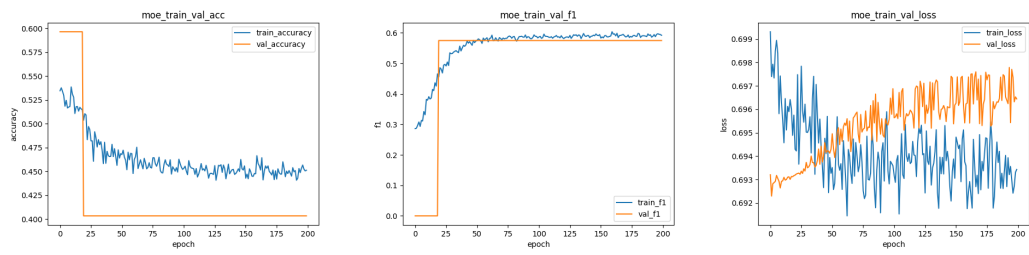


图 12: moe_bert