

進捗報告

1 今週やったこと

- LSTM を用いた台詞および画像の入れ替え識別

1.1 タイトル推定

4 コマ漫画の 4 コマ目だけを入れ替えて入れ替えた or 入れ替えてないのクラス識別をした。用意したクラスは最後の 4 コマ目に対して (0) 正解 (入れ替えていない) (1) 同じタイトルの違う台詞 (2) 違うタイトルの違う台詞の 3 クラス。4 つめのクラスとしてランダムノイズをおいていたが識別率がさすがに 100 % だったため消去。この識別実験を (一) 台詞を入れ替える (二) 画像を入れ替える (三) 両方入れ替えるの 3 つについて実施した。表 1, 2, 3 にそれぞれの実験結果を示す (train:val=9:1)。エポックは 100, 損失関数は Class Entropy Loss を用いた。また、モデル構造は 2 つの LSTM の最終層を concat して MLP にかける形となっている (画像と台詞を concat して 1 つの LSTM 構造のモデルも試したがそんなに識別率は変わらなかった。)

表 1: 入れ替え推定結果 (台詞入れ替え)

クラス	Precision	Recall	F-1	support
0	0.40	0.75	0.52	68
1	0.45	0.13	0.20	68
2	0.54	0.44	0.48	68
-	-	-	-	-
accuracy			0.4412	204
macro avg	0.46	0.44	0.40	204

表 2: 入れ替え推定結果 (画像入れ替え)

クラス	Precision	Recall	F-1	support
0	0.34	0.32	0.33	68
1	0.38	0.38	0.38	68
2	0.53	0.56	0.54	68
-	-	-	-	-
accuracy			0.4200	204
macro avg	0.42	0.42	0.42	204

ベースライン 0.33 に対して 0.42-0.5 出てるのでまだなんとか... 画像と台詞の両方を変えた識別率が上がったのは狙い通りという感じ。

また、クラス 0(正解) とクラス 2(タイトル違うものから持ってくる) だけの 2 クラス識別も試してみたが、ほとんど精度出ず (0.6 くらい)。画像と台詞だけのタイトル識別が 5 クラス識別で 0.6 くらい出たしもう少し上がってもいいような気がするが、1 vs rest の構造とタイトルをランダムに入れ替えているところに問題があるのだろうか... ?

表 3: 入れ替え推定結果 (画像, 台詞入れ替え)

クラス	Precision	Recall	F-1	support
0	0.43	0.43	0.43	68
1	0.50	0.51	0.51	68
2	0.58	0.56	0.57	68
-	-	-	-	-
accuracy			0.5000	204
macro avg	0.50	0.50	0.50	204

2 今後の方針

ひとまず JSAI をこの 3 クラス識別の実験内容をベースに書きたい．本番までにまだまだやれることはありそう．現段階であと確認したい事項としては

- 画像, 台詞だけの LSTM での破綻識別
- 一番入れ替えると識別率が上がるコマはどれか調べる
- ランダムに入れ替えているところのタイトル間の偏りをなくす
- 誤識別したデータの確認 (この識別率の段階でやることではないような気も)
- end to end なモデルで識別タスクを行う (Future Work)

など．