進捗報告

1 今週やったこと

- IDS で SUB 漢字で実験した
- IDS の内容の確認

2 実験

2.1 データセット

問題文-不正解データの生成をランダムに直した. データの数は表 1 のように示す.

表 1: Dateset

experiment	Real	Fake	Total	Training	Validation	Testing
LSTM(Char)	79725	79725	159450	127560	15945	15945
LSTM(SUBchar)	79725	79725	159450	127560	15945	15945

算式で解ける灯謎は確認中である.

2.2 実験用モデル

双方向二層 LSTM モデルを使用した.

2.3 実験結果

実験結果として,LSTM (char) に対し,Train Loss と Validation Loss は 0.550 と 0.874 に収束し,Train Accuracy と Validation Accuracy は各自 69.77 と 55.26 パーセントに収束した.

LSTM (subchar) に対し,Train Loss と Validation Loss は 0.347 と 0.524 に収束し,Train Accuracy と Validation Accuracy は各自 84.20 と 77.81 パーセントに収束した.

実験結果は図1,図2のように示す.

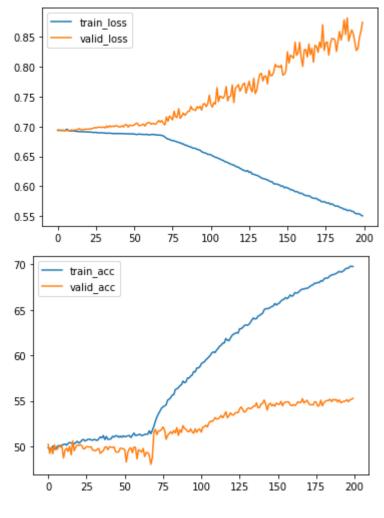
Testing データによる結果は表 2 のように示す.

表 2: Testing result

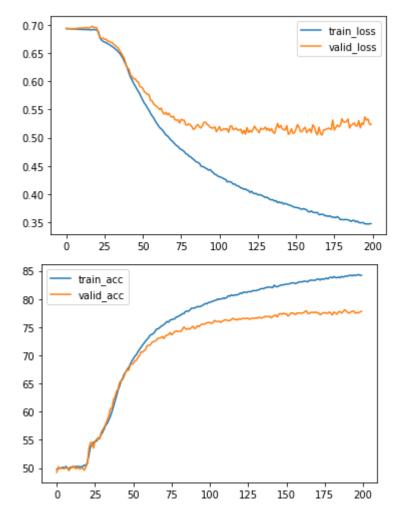
experiment	TP	TN	FP	FN	Accuracy	Precision	Recall
LSTM(char)	2609	5342	2596	5398	49.84	50.12	32.58
LSTM(subchar)	6559	5908	2030	1448	78.18	76.37	81.92

結果の例は図3,図4のように示す.

結論として、SUB 漢字の方が表現が良いである.



☑ 1: LstmChar



☑ 2: LstmSUBChar

	Sentence	answer	prediction
0	一湖新荷初露头?灞	1	0
1	有错更别一直吹?爽	0	1
2	首富得力留空中一下花完所有钱?隼	0	0
3	东跑西撞?抱	1	0
4	前头同心干点滴作奉献?巨	0	0
	老卢七十不显老?轳	1	0
6	卧看舞姬人纵横?广	0	0
	重阳国庆两相逢?畦	1	1
8	天地无情多是非?如	0	0
9	俺在找高手?百	0	0
10	日本曲调?洵	0	1
11	边陲事急心如割?茎	0	0

図 3: CharResult

	Sentence	answer	predictio
0	一》胡亲斤艹何衤刀雨路头?》霸	1	1
1	月年 昔一日乂另 1一十且一口欠?大	0	1
2	自宀富行ノ刀田穴エロトーートサ化宀元戶斤月针 考?隹十	0	0
	东包一扌童?扌包	1	1
	刖头口心干占灬氵商仁乍二丨南犬?巨	0	0
	耂ヒ尸し十一日业耂ヒ?车卢	1	1
6	臣卜目卌一舛女人纟从木黄?广	0	1
7	重阝日口玉广大一口从木目辶夆?田圭	1	1
8	一大土也一九十青夕夕日非?女口	0	0
9	亻奄土扌戈广口问手?一白	0	0
10	日本曲 : 周? 注 旬	0	0

☑ 4: SUBCharResult

3 IDS に関する調査

IDS の中国漢字に対して、部首は間違いがないことを確認した. しかし、成分が表示出来ない SUB 漢字成分がある. IDS の例は図 5 のように示す.

3.1 次のステップ

IDS に表示出来ない SUB 漢字部分を別の方法で表示して実験する. Baseline を確認するための SVM のプログラムは作成中で, 出来次第実験に使用する.

4 来週目標

- SVM 実験をする
- クイズに関する論文を読む

```
—3
U+4E0D
          不
U+4E0E
               ___(GTKV]
                              @@—[J]
          与
          丏
               丏
U+4E0F
          丐
               丐
U+4E10
U+4E11
          {\mathbb H}
               {\mathbb H}
U+4E12
          丒
               ⊟刃一
U+4E13
          专
               专
U+4E14
          且
               且
               ⊟不一
U+4E15
          丕
          世
U+4E16
               世
               ⊕廿Ⅰ
U+4E17
          丗
U+4E18
          丘
               Æ
               一内
U+4E19
          丙
               4—
          业
U+4E1A
               ⊟从一
U+4E1B
          丛
U+4E1C
          东
               东
U+4E1D
          <u>44</u>
               =(20-
U+4E1E
          丞
               ⊞承一
          丟
               三王厶
                         二一去
U+4E1F
U+4E20
          丠
               ⊟北一
               ⊞—□□山
U+4E21
          両
                         三壬厶
               ⊟丿去
U+4E22
          丢
U+4E23
          丣
               丣
               □─□□从
U+4E24
          两
          )
III
U+4E25
                 -④厂
```

図 5: IDS