令和 5 年 1 月 11 日 (水) ゼミ資料 M2 陳偉斉

進捗報告

1 今週やったこと

- baseline 実験

2 実験内容

SVM, LSTM, bert-base-chinese モデルに対して実験しました.

3 データセット

「問題文 + ヒント + 答え」,「正解」で構成されるの実験データ 72937 件集まりました. そして Levenshtein を利用し,「不正解だけと似てる漢字 (画と SUB 漢字両方も Levenshtein 距離が 1)」の条件でデータ数を 218811 件に増加しました. bert-base-chinese モデルのみに対して,「明らかに不正解 (画と SUB 漢字両方も Levenshtein 距離が 20 以上)」と「ヒントなし」の条件を設定し実験をしました. 他のモデルはまだ実験中です.

データセットの構成は表 1 のように示します.

表 1: データセットの構成

条件	訓練データ	テストデータ	合計
似てる不正解	175048	43763	218811
明らかに不正解	175048	43763	218811
ヒントなし	175048	43763	218811

4 実験結果

実験結果は表 2 , 表 3 に示します.

表 2: 各モデルの実験結果

結果	SVM(CountVectorizer)	SVM(TfidfVectorizer)	LSTM(bidirectional)	bert-base-chinese
訓練誤差			0.21	0.21
訓練精度	0.67	0.68	0.90	0.90
テスト精度	0.66	0.56	0.86	0.78

その中に最も表現がいいのは LSTM モデルですが, Bert の訓練は時間かかりますので, 毎回 20 Epoch を設定しました (コスト 6 時間). それに反して, LSTM は 50 Epoch d でした. 故に Bert モデルは訓練不足の可能性もあります.

結果として, 人類の判断に影響する「漢字の形」と「ヒント」が bert-base-chinese に与える影響は僅かですが, 今回の実験は全部漢字を最小単位として扱うため, 画と SUB 漢字は今回の実験に導入されてません. 次に導入します.

表 3: Bert と違うデータセットの実験結果

条件	似てる不正解	明らかに不正解	ヒントなし
訓練誤差	0.21	0.16	0.40
訓練精度	0.90	0.94	0.85
テスト精度	0.78	0.76	0.78

5 来週目標

- Bert を 100 epoch に増加し実験する
- 不正解データを増加し実験する
- 画と SUB 漢字を導入する (データの中に分けるか, 分散表現を生成するか)