

報告書

1 今週の進捗

- テスト数 10, 100, 250, 500 とした実験
- head, relation が同じトリプルを考慮した実験
- MLM を用いた実験

2 KG-BERT [1]

2.1 モデルの説明

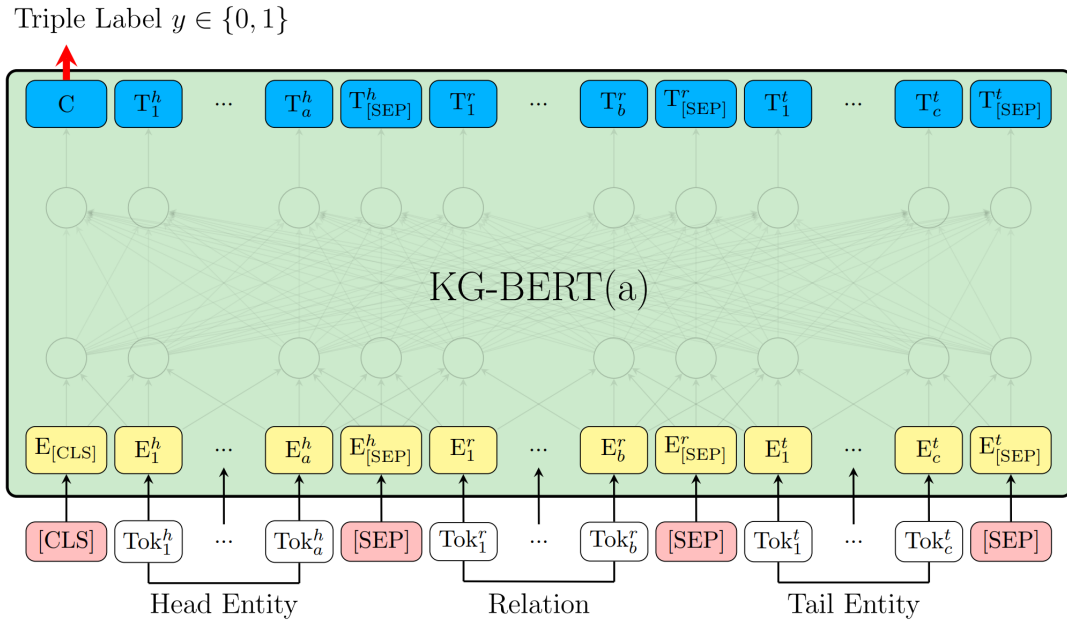


図 1: KG-BERT model [1]

表 1: データセット

Dataset	Entity	Relation	Train	Validation	Test
WN18RR	40,943	11	86,835	3,034	3,134

評価指標として Mean Rank (MR), Mean Reciprocal Rank (MRR), Hits@ k を使用する. MR とは, 予測したエンティティのランクの平均を指す. MRR とは, 予測したエンティティのランクの逆数をスコアとしており, こうして得たスコアの平均をとったものを指す. Hits@ k とは, 予測したエンティティを順位付けしたときに, 上位 k 個以内に正解が含まれている割合のことを指す.

表 2: パラメータ (WN18RR)

パラメータ	値 (default)
学習率	5e-5
epoch	5
dropout rate	0.1
batch size	32
eval batch size	128 (5000)
max seq length	32 (50)

$$\text{MR} = \frac{1}{|E|} \sum_{i=1}^{|E|} \text{rank}_i \quad (1)$$

$$\text{MRR} = \frac{1}{|E|} \sum_{i=1}^{|E|} \frac{1}{\text{rank}_i} \quad (2)$$

$|E|$ はエンティティ数, rank_i は予測したエンティティのランクを表している.

MR は値が小さいとき, MMR, Hits@ k はともに値が大きいとき推定精度が良いと判断される.

2.2 テスト数 10, 100, 250, 500 の実験結果の比較

テストデータからランダムにテスト数 {10, 100, 250, 500} だけトリプルを取ってきてテストする. 表 3 にそれぞれの結果を示す. 以下のテスト数 {10, 100, 250, 500} の実験はすべて同じ学習済みモデルを使用している.

表 3: テスト数ごとの実験結果

テスト数	WN18RR					実験時間
	MR	MRR	Hits@1	Hits@3	Hits@10	
3,134 (文献値)	97	-	-	-	52.4	
3,134 (再現実験)	117.77	0.25	12.41	29.44	51.85	約 10 日
10	54.7	0.0958	0.0	0.0	50.0	約 0.5 日
100	20.84	0.1238	0.0	9.68	51.61	約 2 日
250	85.528	0.146	5.2	12.4	38.0	約 4 日
500	81.332	0.131	3.80	12.60	33.00	約 8 日

テスト数が少なくなると文献値や再現実験の結果とは異なる結果が得られるため, テスト数を極端に減らすことは不正確な結果につながると考えられる. しかし, テスト数を増やすと実験時間がかかりすぎてしまう.

1つのテストトリプルに対して約 40,000 個の存在しないトリプルを作成してテストしているが, これを 10,000 個まで減らして実験することで実験時間の短縮が望まれる. 10,000 個の存在しないトリプルの選択方法としてはランダム, もしくはテストトリプルにある tail と似た tail を用いて作成したトリプルを選択する方法が考えられる.

2.3 エンティティ数 (40,943) 分類

2.3.1 存在するトリプル数の分だけ上位にあるトリプルの割合を指標とした実験

あるトリプルの tail を予測するとき, その head, relation をもつ他のトリプル数を N , 予測した結果の上位 N のうち存在するトリプル数を n とすると, 指標は以下のように表すことができる.

$$\frac{n}{N} \quad (3)$$

図 2 に縦軸を上記指標 Hits@ k の値, 横軸をテストトリプルと同じ head, relation をもつトリプル数 N とした棒グラフを示す. 時間の都合上, 修正できていないため横軸の値は見づらくなっている. 右から, 473, 244, 176, 123, 82, 80, 77, 76, ... となっている.

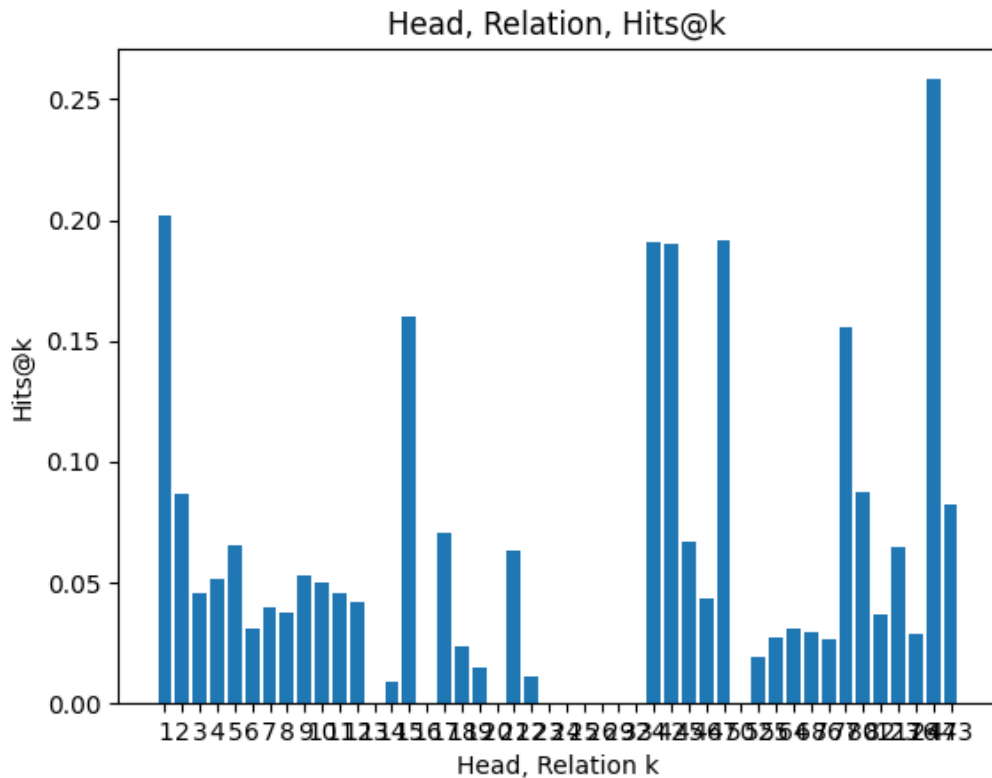


図 2: Exist Hits

テストトリプルと同じ head, relation をもつトリプル数が 244 のとき, Hits@244 は 0.2582 で最も高い値となっている. しかし, テストトリプルと同じ head, relation をもつトリプル数と Hits@ k の相関はないと考えられる.

2.4 MLM を用いた実験

現在, プログラムを書いており, まだ完成していない. 以下にコードと現在出ているエラー内容を示す.

Listing 1: MLM

```
1 def convert_examples_to_features(examples, label_list, max_seq_length, tokenizer, print_info =  
    True):  
2     features = []
```

```

3     for (ex_index, example) in enumerate(examples):
4         if ex_index % 10000 == 0 and print_info:
5             logger.info("Writing example {} of {}" % (ex_index, len(examples)))
6
7         tokens_a = tokenizer.tokenize(example.text_a)
8
9         tokens_b = None
10        tokens_c = None
11
12        if example.text_b and example.text_c:
13            tokens_b = tokenizer.tokenize(example.text_b)
14            tokens_c = tokenizer.tokenize(example.text_c)
15            _truncate_seq_triple(tokens_a, tokens_b, tokens_c, max_seq_length - 4)
16        else:
17            if len(tokens_a) > max_seq_length - 2:
18                tokens_a = tokens_a[: (max_seq_length - 2)]
19
20        tokens = ["[CLS]"] + tokens_a + ["[SEP]"]
21
22        if tokens_b:
23            tokens += tokens_b + ["[SEP]"]
24        if tokens_c:
25            tokens += tokens_c + ["[SEP]"]
26
27        input_ids = tokenizer.convert_tokens_to_ids(tokens)
28
29        padding = [0] * (max_seq_length - len(input_ids))
30        input_ids += padding
31
32        assert len(input_ids) == max_seq_length
33
34        if ex_index < 5 and print_info:
35            logger.info("*** Example ***")
36            logger.info("guid: {}" % (example.guid))
37            logger.info("tokens: {}" % " ".join(
38                [str(x) for x in tokens]))
39            logger.info("input_ids: {}" % " ".join([str(x) for x in input_ids]))
40
41        features.append(input_ids)
42    return features

```

Listing 2: Error Message

```

1  Traceback (most recent call last):
2    File "run_bert_mlm.py", line 1229, in <module>
3        main()
4    File "run_bert_mlm.py", line 846, in main
5        train_data = TensorDataset(train_features)
6    File "/usr/local/lib/python3.8/site-packages/torch/utils/data/dataset.py", line 192, in
7        __init__
8        assert all(tensors[0].size(0) == tensor.size(0) for tensor in tensors), "Size mismatch
9        between tensors"
10   File "/usr/local/lib/python3.8/site-packages/torch/utils/data/dataset.py", line 192, in <
11   genexpr>

```

```
9     assert all(tensors[0].size(0) == tensor.size(0) for tensor in tensors), "Size mismatch_
    between_tensors"
10 AttributeError: 'list' object has no attribute 'size'
```

エラーメッセージではサイズが違っているが, padding によってサイズを合わせている. padding 以外の理由でエラーが出ているのか, padding の方法が間違っているのか, デバッグしてコードを見直す必要がある.

3 今後したいこと

- KG-BERT に MLM を適用して実装
- エンティティ数 (40,943) 分類の改良
- KG-BERT の改良

参考文献

- [1] Liang Yao, Chengsheng Mao, and Yuan Luo. KG-BERT: BERT for knowledge graph completion. *CoRR*, Vol. abs/1909.03193, , 2019.