

報告書

1 今週の進捗

- テスト数 10 での実験
- 今後したい実験の提案

2 KG-BERT [1]

2.1 テスト数 10 での実験

テスト数が多いと時間がかかりすぎる (10 日ほど) ため, テスト数を 10 にして実験を回した. 4 日ほどで実験が完了する. しかし, 実験結果から得られるものはなかった.

2.2 モデルの説明

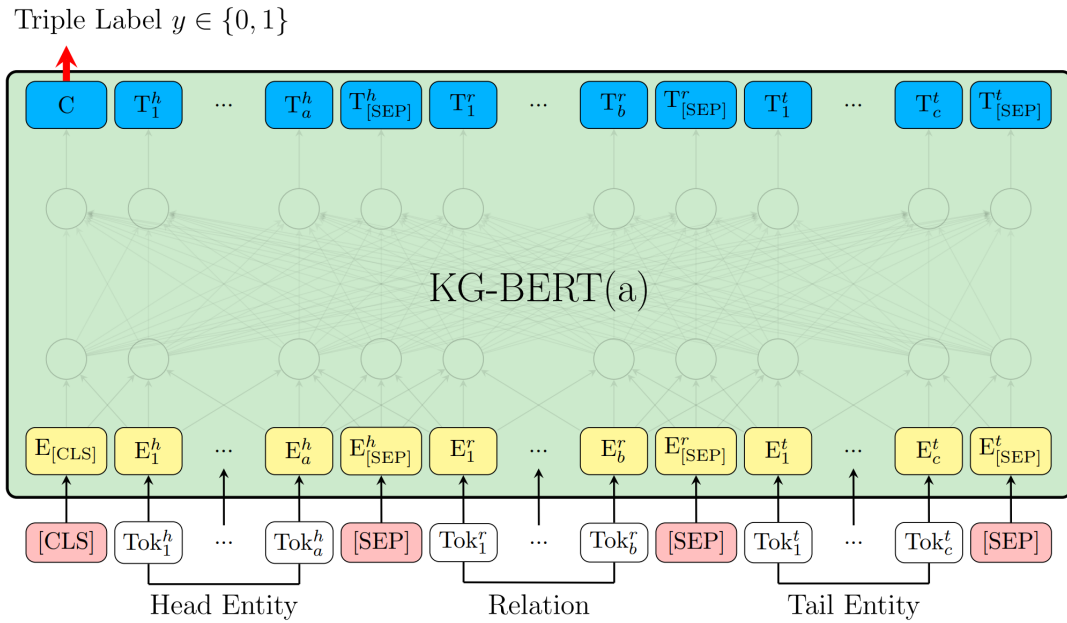


図 1: KG-BERT model [1]

表 1: データセット

Dataset	Entity	Relation	Train	Validation	Test
WN18RR	40,943	11	86,835	3,034	3,134 (10)

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)

2.3 今後したい実験の提案

- エンティティ数 - 分類

テストに時間がかかりすぎるため, 時間削減のためにテストのやり方をヘッドとリレーションに対して 1 つのエンティティを予測する分類にしたい. エンティティは 40,943 個あるため, 40,943- 分類をする.

- sentence-BERT の使用

KG-BERT の BERT の部分に sentence-BERT を導入する.

- リレーションごとに予測するモデルの変更

リレーションごとの予測成功確率によって使用するモデルを決定する. 例えば, wordnet のリレーション「hypernym」を含むトリプルの予測は訓練時において KG-BERT が最も精度が良い場合, テストでも KG-BERT を使用する. これにより, 精度は高くなる.

3 ナレッジグラフ推論チャレンジ¹

3.1 チャレンジタスク

大規模言語モデルを使ったナレッジグラフの構築

3.2 応募部門

- 推理小説部門 (ホームズの小説 8 つ)

タスク: 「シャーロック・ホームズの小説を対象としたナレッジグラフ」と同等のものを構築する.

- 一般部門 (対象領域を問わない任意のナレッジグラフ)

タスク: 対象領域を問わない任意のナレッジグラフを構築する.

3.3 応募締切

2023 年 12 月末

4 次週すること

•

¹<https://challenge.knowledge-graph.jp/2023/>

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.