

報告書

1 今週の進捗

- MLM を用いた実験
- 評価指標 ROUGE に関して

2 KG-BERT [1]

2.1 MLM を用いた実験

ナレッジグラフにおける tail 推定モデルとして, BERT の Masked Language Model を適用した実験をする. 入力として, head, relation, tail の説明文を用い, tail の見出し文を MASK として推定させる. 表 1 に入力と出力の例を示す. このとき, tail の見出し文の単語数は複数になる可能性がある.

表 1: MLM の入力と出力の例

triple	Head	Relation	Tail
入力	family crocodylidae, true crocodiles	member meronym	[MASK], a genus of Malayan crocodiles
出力			tomistoma

表 2 に本実験で用いるデータセットである WN18RR におけるデータ数, 表 3 に本実験のパラメータを示す.

表 2: データセット

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

表 3: パラメータ

パラメータ	値
学習率	5e-5
epoch	20
mlm probability	0.15
batch size	32
eval batch size	128
max seq length	128

評価指標として Hits@ k を使用する. 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し, 値が大きいとき推定精度が良いと判断される.

表 4 に実験結果を示す. 比較として KG-BERT における文献値と再現実験の結果も示している. 評価指標の MR と MRR については MLM を用いた実験では実装できていない.

KG-BERT における文献値と再現実験の結果と比較すると, Hits@ k において MLM を用いた実験のほうが良い精度となっていることがわかる. しかし, KG-BERT では見出し文と説明文を含めた tail を推定しているの

表 4: MLM を用いた実験結果

モデル	WN18RR				
	MR	MRR	Hits@1	Hits@3	Hits@10
KG-BERT (文献値)	97	-	-	-	52.4
KG-BERT (再現実験)	117.77	0.25	12.41	29.44	51.85
MLM	-	-	44.58	57.02	61.93

に対し, MLM を用いた実験では tail の見出し文のみを推定している. そのため, KG-BERT の実験結果と正確な比較はできていない.

2.1.1 評価指標 ROUGE に関して

ROUGE-N とは, 正解文中の N-gram のうち, どの程度が生成文に含まれているかを評価する指標のことである.

$$\text{ROUGE}_N = \frac{\text{Count}_{\text{match}}}{\text{Count}_{\text{reference}}} \quad (1)$$

$\text{Count}_{\text{match}}$, $\text{Count}_{\text{reference}}$ はそれぞれ正解文と生成文の間で一致する N-gram の数, 正解文に含まれる N-gram の数を指す.

tail の見出し文は単語数が 1, 2 のものがほとんどであるため, tail の見出し文の推定では適用しにくいと考えられる.

3 今後したいこと

- MLM を用いた実験の改良
- 評価指標の検討

4 KG-BERT のモデル図

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.

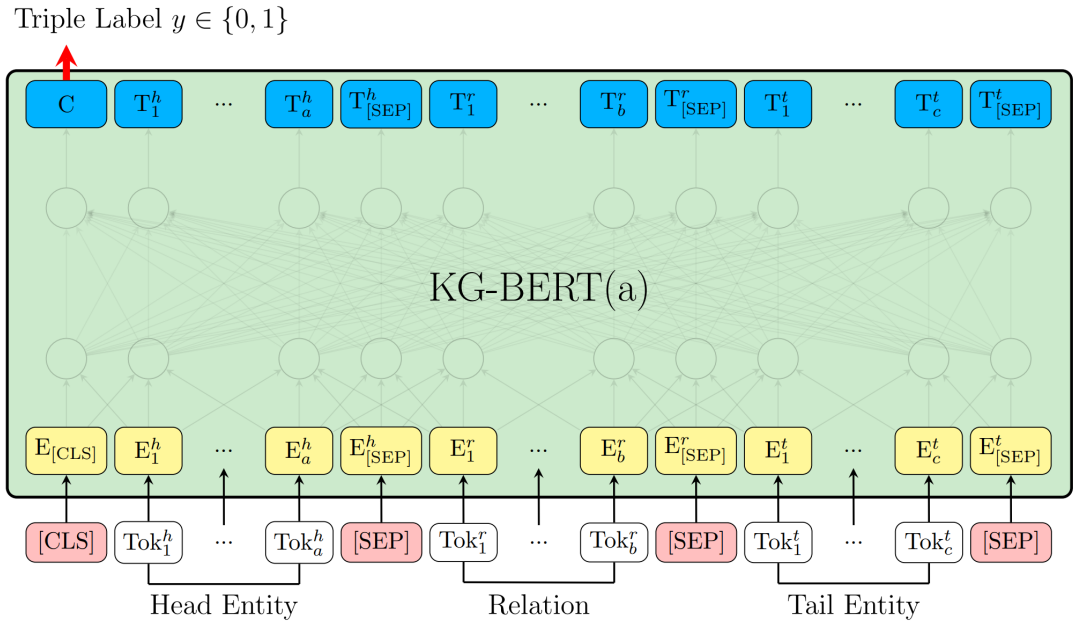


图 1: KG-BERT model [1]