

Masked Language Modeling を用いた Knowledge Graph 補完手法の検討

1 はじめに

近年, 人工知能技術は急速な発展を遂げている. その中で人間の知識をグラフ構造で表現する Knowledge Graph [1] が注目を集めており, 人工知能の基盤技術としてさまざまな分野で活用されている. Knowledge Graph にはさまざまな知識が含まれているが, そこに新たに知識を付与したいという要求が生じる場合がある. そこで, Knowledge Graph を自動補完する手法が求められている. 代表的な補完手法として TransE [2] や ComplEx [3] などがある. しかし, これらの従来手法は知識の関係性を重視して学習しており, 知識自体の意味情報を効果的に捉えていない.

本研究では, 知識自体の意味情報を効果的に捉えるために言語モデルである BERT の Masked Language Modeling を用いて Knowledge Graph を補完する手法を検討する.

2 要素技術

2.1 Knowledge Graph

Knowledge Graph [1] とは, さまざまな知識を体系的に連結し, その関係をグラフ構造で表した知識ネットワークのことである. 図 1 に Knowledge Graph の例を示す. Knowledge Graph は, head, tail を要素に持つ entity 集合と, その entity 間の関係を表現する relation を要素に持つ relation 集合によって構成されており, 図 1 のように entity をノード, relation をエッジとする有向グラフとして表される. また, Knowledge Graph を表す表現方法として (head, relation, tail) という 3 組構造 (triple) の集合で表すこともできる.

2.2 BERT

Bidirectional Encoder Representations from Transformers (BERT) [4] は, Transformer をベースとして双方向エンコーダで構成された言語モデルであり, 2018 年に Google が発表した. 特徴として文脈を考慮した処理が可能であり, 翻訳や文書分類, 質問応

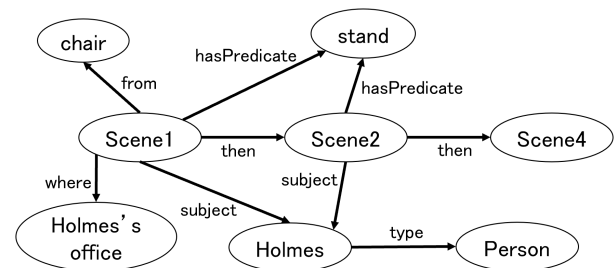


図 1: Knowledge Graph の例

答などのさまざまなタスクにおいて公開当時の最高スコアを超える結果が報告されている. BERT はマスクされた単語を予測する Masked Language Modeling (MLM) と 2 つの文が連続するかを分類する Next Sentence Prediction (NSP) という 2 つの手法で学習しており, このように事前学習したモデルに転移学習およびファインチューニングすることでさまざまなタスクに対応できる.

2.3 KG-BERT

Knowledge Graph BERT (KG-BERT) [5] は, Yao らによって提案された Knowledge Graph 補完手法の 1 つである. 図 2 に KG-BERT モデルの概略図を示す. KG-BERT では, Knowledge Graph の head, relation, tail を [SEP] トークンで区切ったものを BERT の入力として, その triple が存在するか否かを [CLS] トークンのみを利用した 2 値分類で判定している. 学習の際は 1 つの triple に対してその tail を他の entity と置換した triple を 5 つ生成してそれらが存在しない triple だと判定するように学習する. テストの際は 1 つのテスト triple に対してその tail を他の entity と置換した triple を entity の数だけ生成し, それらの triple の分類スコアをランク付けすることで評価する.

3 実験

3.1 データセット

WN18RR [6] は, 英語の大規模な語彙データベースである WordNet から triple を自動抽出して得られたデータセット WN18 から作られており, WN18 か

Triple Label $y \in \{0, 1\}$

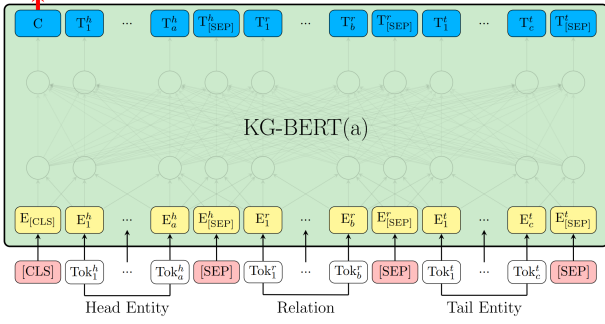


図 2: KG-BERT (文献 [5] 図 1 より参照)

表 1: WN18 から逆関係の triple を除去する例

	head	relation	tail
	dog, (description)	hypernym	poodle dog, (description)
除去	poodle dog, (description)	hyponym	dog, (description)
	⋮	⋮	⋮

ら head と tail を入れ替えて得られる逆関係の triple を除去したデータセットとなっている。WN18RR では, entity 情報を「見出し語, その説明文」の形で持っている。表 1, 2 に WN18 から逆関係の triple を除去する例と WN18RR の例をそれぞれ示す。本研究では WN18RR を訓練データとテストデータにそれぞれ 9:1 に分割して用いた。表 3 に WN18RR における entity, relation, triple の数, 訓練データ, テストデータの数を示す。

3.2 評価指標

評価指標として Mean Reciprocal Rank (MRR), Hits@ k , Filtered MRR, Filtered Hits@ k を用いる。予測結果の k 番目に正解があるとき, その順位 k のことをランクと呼ぶ。|T| を triple 数, rank_i を triple_i における正解 tail のランクとすると, MRR は (1) 式で表される。

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

Hits@ k は, tail 推定において上位 k 個以内に正解の要素が出力されている割合を表す。このとき, あるテスト triple の正解 tail より予測結果の上位にそのテスト triple と同じ head, relation をもつ triple の tail が存在する場合, その分不当に低くランク付けされる。そこで, テスト triple と同じ head, relation をもつ triple の tail を予測結果から除去してランク付け

表 2: WN18RR の例

head	relation	tail
doggy do, (description)	hypernym	stool, (description)
rime, (description)	derivationally related form	verse, (description)
sphingidae, (description)	member meronym	genus acherontia, (description)
⋮	⋮	⋮

表 3: データセットの内訳

データセット	entity	relation	triple	訓練	テスト
WN18RR	40,943	11	93,003	89,869	3,134

をし, それによって得られた MRR, Hits@ k をそれぞれ Filtered MRR, Filtered Hits@ k とする。本研究では $k = 1, 3, 10$ で評価した。MMR, Hits@ k , Filtered MRR, Filtered Hits@ k はすべて値が大きいに推定精度が良いと判断される。

3.3 実験 1

triple の head, relation, tail を [SEP] トークンで区切ったものを入力として BERT の MLM をファインチューニングしたモデルを使用した。このモデルは, 15% の入力トークンを 80% の確率で [MASK] トークンに置き換え, 10% の確率でランダムに選んだトークンに置き換え, 10% の確率でそのままにすることで学習した。テスト時には tail の見出し語のトークンを [MASK] トークンに置き換えることで tail の見出し語を推定した。なお, tail の見出し語の説明文はそのまま入力している。図 3 にモデルの概略図を示す。このタスクは推定する tail の説明文も入力としているため, 通常想定される tail 推定タスクより tail の推定が容易である。

3.4 実験 2

実験 1 と同様に, triple の head, relation, tail を [SEP] トークンで区切ったものを入力として BERT の MLM をファインチューニングしたモデルを使用した。テスト時には, head, relation の入力方法は同じであるが, tail の見出し語のトークンを [MASK] トークンに置き換えたもののみで説明文のトークンを取り除いたものを入力とした。図 4 にモデルの概略図を示す。このタスクは実験 1 とは異なり, 入力が head, relation の情報のみであるため通常想定される tail 推定と同等のタスクとなっている。

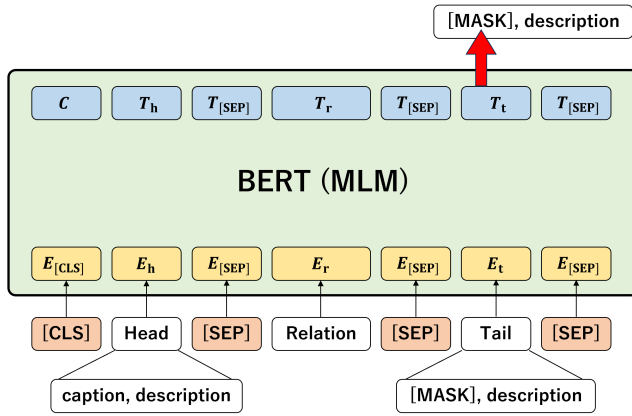


図 3: 実験 1 のモデル概略図

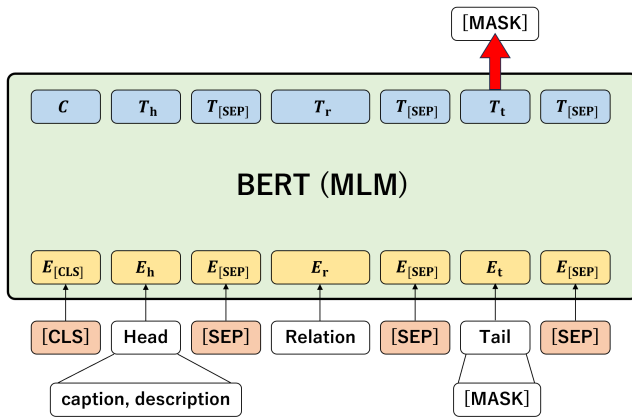


図 4: 実験 2 のモデル概略図

表 4 に比較のための KG-BERT と実験 1, 2 のパラメータを示す。なお, KG-BERT では訓練データ, 検証データ, テストデータは 8:1:1 で分割しており, eval batch size と max seq length はサーバのメモリの都合上文献値とは異なっている。また, 表中の「- (ハイフン)」はそのモデルには必要のないパラメータを表している。

4 実験結果

4.1 実験 1, 2

表 5 に KG-BERT と実験 1, 2 の結果を示す。KG-BERT は文献値と再現実験の 2 つの結果を示している。また, 実験 1, 2 の結果はそれぞれ 3 回試行した平均と標準偏差を示している。

表 5 より, 実験 1 の結果が MRR, Hits@1, Hits@3, Hits@10 において KG-BERT の結果を上回っていることがわかる。これにより, MLM における tail 推定

表 4: KG-BERT と実験 1, 2 のパラメータ

パラメータ	KG-BERT (default)	実験 1	実験 2
learning rate	5.0×10^{-5}	5.0×10^{-5}	5.0×10^{-5}
mask probability	-	0.15	0.15
batch size	32	32	32
eval batch size	128 (5000)	-	-
max seq length	32 (50)	128	128
epoch	5	50	20

の有効性を示すことができた。しかし, 通常想定される tail 推定タスクより tail の推定が容易であるため, KG-BERT と単純に比較することはできないと考えられる。また, Filtered MRR, Filtered Hits@ k について MRR, Hits@ k と比較してその値の変化は微細な範囲にとどまった。これは tail の説明文の情報が結果に大きく影響したためと考えられる。

実験 2 の結果は, Hits@1 において KG-BERT の結果を少し上回っており, KG-BERT より tail の推定が向上していることが示せた。しかし, MRR, Hits@3, Hits@10 においては KG-BERT の結果を下回っており, 入力 head, relation の情報をうまく捉えられていないと考えられる。また, Filtered MRR, Filtered Hits@ k については実験 1 と同様に MRR, Hits@ k より少し向上したがほとんど変化しておらず, 訓練データ内の triple の tail 推定もうまくできていないことがわかる。そのため, head, relation の情報だけでは tail の推定が難しいと考えられる。

4.2 考察

実験 1, 2 の結果より, 一部の結果において KG-BERT と同程度, もしくはそれ以上の結果が得られたため, MLM を用いて Knowledge Graph を補完できる可能性があることが示された。MLM の出力は BERT に登録されている単語すべてを候補にしているため, データセット内に entity として存在しない単語も出力されている。そこで今後の方針として, MLM の出力候補をデータセット内の entity のみに限定した場合を検証したいと考えている。

Knowledge Graph 補完は Knowledge Graph 内に存在しない entity に対して適切な triple を生成することを想定しているが, 本研究ではテストデータの triple 内の entity が訓練データの triple 内に存在しているため, テストは Knowledge Graph 補完を完全に再現できていない。そこで, ある entity を含む triple データを訓練データより除外してそれらをすべてテストデータとすることで Knowledge Graph

表 5: MLM を用いた実験結果

モデル	WN18RR			
	MRR	Hits@1	Hits@3	Hits@10
KG-BERT (文献値)	-	-	-	52.4
KG-BERT (再現実験)	0.25	12.41	29.44	51.85
実験 1	0.558 ± 0.002	53.37 ± 0.14	57.40 ± 0.20	58.80 ± 0.13
実験 1 (Filtered)	0.560 ± 0.002	53.51 ± 0.14	57.57 ± 0.18	58.87 ± 0.12
実験 2	0.186 ± 0.002	12.81 ± 0.16	21.55 ± 0.16	30.00 ± 0.72
実験 2 (Filtered)	0.188 ± 0.001	12.97 ± 0.12	21.68 ± 0.13	30.09 ± 0.69

補完を再現することも今後の課題である。

5 まとめと今後の課題

本研究では, MLM を用いて KG-BERT より tail の推定が容易なタスクと同等のタスクとして tail を推定し, 一部の結果において KG-BERT と同程度, もしくはそれを上回る結果が得られた. これにより, MLM を用いた Knowledge Graph 補完手法の有効性が確認できた.

今後の課題として, 先述した MLM の出力候補の限定, Knowledge Graph 補完の評価として適切なテストデータの作成が挙げられる. また, tail 推定タスクにおいてテスト時の入力における [MASK] トークンの位置が決まっているため, MLM の学習時の入力における [MASK] トークンの位置を tail の位置に固定したモデルを用いること, 同じ head, relation をもつ triple における tail のうち, 予測結果として出力された割合を評価指標として加えることを検討する.

参考文献

- [1] 川村隆浩, 江上周作, 田村光太郎, 外園康智, 鶴飼孝典, 小柳佑介, 西野文人, 岡嶋成司, 村上勝彦, 高松邦彦, 杉浦あおい, 白松俊, 張翔宇, 古崎晃司. 第 1 回ナレッジグラフ推論チャレンジ 2018 開催報告 — 説明性のある人工知能システムを目指して —. 人工知能 34 巻 3 号, 2019.
- [2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2013.
- [3] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, Vol. 48 of *Proceedings of Machine Learning Research*, pp. 2071–2080, New York, New York, USA, 20–22 Jun 2016. PMLR.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [5] Liang Yao, Chengsheng Mao, and Yuan Luo. KG-BERT: BERT for knowledge graph completion. Vol. abs/1909.03193, 2019.
- [6] Dettmers T., Minervini P., Stenetorp P., and Riedel S. Convolutional 2d knowledge graph embeddings. Vol. 32(1), 2018.