

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. 実世界の知識とそれらの関係を人手ですべて網羅して Knowledge Graph を作成するには多大なコストがかかる. そこで, Knowledge Graph 内の関係を基に含まれていない関係を自動的に補完する手法が求められている. 従来の Knowledge Graph 補完手法である TransE [2] や ComplEx [3] などの問題点として, 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点がある.

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深層言語モデルである BERT による Masked Language Modeling を用いた Knowledge Graph 自動補完手法を提案して, 評価実験によりその有効性を検証した.

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 の表現方法として (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, 2018 年に Google が発表した 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

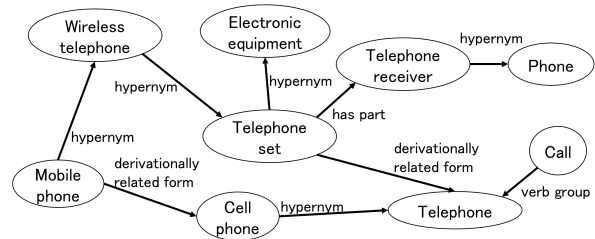


図 1: Knowledge Graph の例

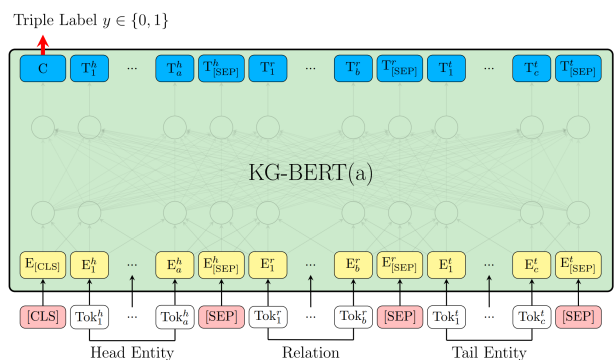


図 2: KG-BERT (文献 [5] 図 1 より参照)

Prediction (NSP) という 2 つの手法で学習している. このように事前学習したモデルは fine-tuning することでさまざまなタスクに対応することができる.

2.3 KG-BERT

Knowledge Graph BERT (KG-BERT) [5] は, BERT を用いた Knowledge Graph 補完手法の 1 つであり, Yao らによって提案された. 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.

2.4 WN18RR

WN18RR [6] は, 英語の大規模な語彙データベースである WordNet [7] から triple を自動抽出して得られたデータセット WN18 [2] を基に作成された triple のデータセットである. WN18RR では entity は “見出し語, その説明文” の形で表される. 表 1, 2

表 1: WN18RR における entity の例

caption	description
poodle dog	an intelligent dog with a heavy curly solid-colored ...
domestic dog	a member of the genus Canis (probably descended ...
telephone	get or try to get into communication (with ...
	...

表 2: WN18RR の例

head	relation	tail
telephone	hypernym	telecommunicate
telephone system	has part	telephone set
telephone system	has part	telephone exchange
telephone system	hypernym	communication system
...

に WN18RR における entity の例と WN18RR の例をそれぞれ示す. なお, 表 2 では entity の説明文を省略して記述している.

3 提案手法

Knowledge Graph の triple における head, relation, tail をそれぞれ h, r, t とすると, Knowledge Graph 補完では triple (h, r, t) に対して “t” に入る tail を回答することで entity 間の関係性を予測する.

本研究では, entity 自体の意味情報を効果的に捉えるために BERT を MLM で fine-tuning したモデルを用いて, 1 つのシーケンス (head, relation, [MASK]) を入力文として tail を予測することによる Knowledge Graph 補完手法を提案する. 図 3 に提案手法のモデル概略図を示す. BERT を MLM で fine-tuning するとき, BERT が提案された論文 [4] と同様に入力トークンの 15% のうち, 80% を “[MASK]” トークンに置き換え, 10% をランダムなトークンに置き換え, 10% をそのままのトークンにする. さらに, 入力トークンの 15% に tail の見出し語を必ず含むように設定する. 図 4 に BERT の MLM による fine-tuning の概略図を示す. 図 3, 4 の Tok はトークンを, E, T は埋め込みベクトルとモデルの出力のベクトルを示す. l, m, n は head, relation, tail のトークン数を, h, r, t は head, relation, tail をそれぞれ表す.

4 実験

BERT を MLM で fine-tuning したモデルに対して, 2 種類の入力文を用いて実験した. 以下では 2 種類の実験について説明する.

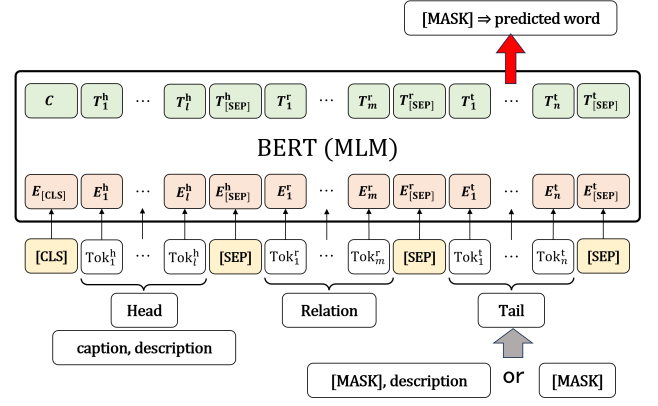


図 3: 提案手法のモデル概略図

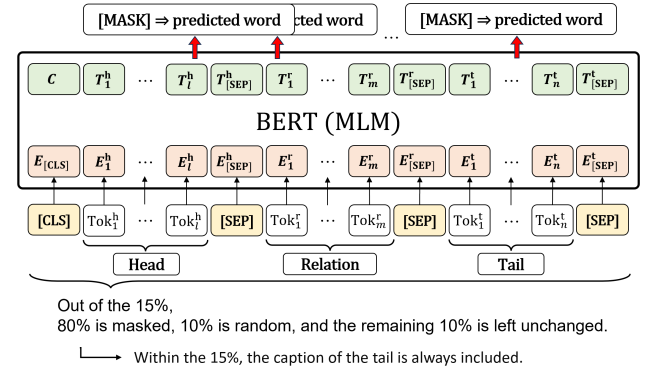


図 4: BERT の MLM による fine-tuning の概略図

4.1 実験 1

実験 1 では, BERT を MLM で fine-tuning したモデルに対して, triple “[CLS] head [SEP] relation [SEP] tail [SEP]” における tail の見出し語を “[MASK]” トークンに置き換えた文 “[CLS] head [SEP] relation [SEP] [MASK], tail の説明文 [SEP]” を 1 つのシーケンスとして入力する. 図 3 において “Tail” を “[MASK], description” とした場合である. これにより, “[MASK]” トークンとなった tail の見出し語を予測する.

4.2 実験 2

実験 2 では, 実験 1 の入力文 “[CLS] head [SEP] relation [SEP] [MASK], tail の説明文 [SEP]” における tail の説明文を除去した文 “[CLS] head [SEP] relation [SEP] [MASK] [SEP]” を 1 つのシーケンスとして入力する. この場合, 図 3 における “Tail” は “[MASK]” のみとなる. これにより, “[MASK]” トークンとなった tail の見出し語を予測する. なお, 予測する tail の見出し語の説明文は入力していない

表 3: Filtered rank の例

triple	rank	tail 予測結果	Filtered rank
存在	1	word 1	1
	2	word 2 除去	-
	3	word 3	2
	\vdots	\vdots	\vdots
正解	r	word r	$r - 1$

め, head と relation の情報のみを用いて tail を予測する形となる.

4.3 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 r 番目に正解があるとき, その順位 r のことを rank と呼ぶ. $|T|$ を triple 数, r_i を triple_i における正解 tail の rank とすると, MRR は (1) 式で表される.

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

Hits@ k は, tail 予測において上位 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 と relation をもつ triple の tail が存在する場合, その分低くランク付けされる. そこで, テスト triple と同じ head と relation をもつ triple の tail を予測結果から除去してランク付けをする. この rank を Filtered rank とする. 表 3 に例を示す. 表 3 において, テスト triple が (head, relation, word r) で, データとして triple (head, relation, word 2) が存在するとき, rank を 1 つ繰り上げることで Filtered rank を得る. Filtered rank によって計算される 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. MRR, Hits@ k , Filtered MRR, Filtered Hits@ k はすべて値が大きいと推定精度が良いと判断される.

4.4 パラメータ

本研究では WN18RR を訓練データ, 検証データ, テストデータとしてそれぞれ 8 : 1 : 1 に分割して用いた. 表 4 に WN18RR における entity, relation, triple の数, 訓練データ, 検証データ, テストデータの数を示す.

表 5 に比較手法の KG-BERT と実験 1, 2 のパラメータを示す. なお, KG-BERT の eval batch size

表 4: WN18RR の内訳

entity	relation	triple	訓練	検証	テスト
40,943	11	93,003	86,835	3,034	3,134

表 5: KG-BERT と実験 1, 2 のパラメータ

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

と max seq length は実験環境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.

5 実験結果

5.1 実験 1, 2

表 6 に KG-BERT と実験 1, 2 の結果を示す. KG-BERT は本実験と tail の予測方法が異なるため単純な比較はできないが, 参考として文献値と再現実験の 2 つの結果を示している. 表中の「- (ハイフン)」は文献 [5] に記載されていなかったことを表している.

表 6 より, 実験 1 ではすべての評価指標において KG-BERT の結果を上回る結果が得られた. これにより, tail 予測に BERT の MLM を利用することの有効性が確認できた. また, Filtered MRR, Filtered Hits@ k について MRR, Hits@ k と比較するとその値の変化は微細な範囲にとどまった.

実験 2 では Hits@1 においてのみ KG-BERT と同等の結果が得られた. 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いなかった.

5.2 考察

実験 1, 2 における tail 予測結果について考察する.

実験 1 でテスト triple (“position, the particular portion of …”, “hypernym”, “point, the precise location of something; a spatially limited location; “she walked to a point where …”) の見出し語 “point” を “[MASK]” に置き換えて tail を予測すると, 提案手法では “point” を予測して正解したのに対し, KG-BERT では “situation” を予測していた. これは tail

表 6: KG-BERT と実験 1, 2 の結果

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	0.546	52.55	56.16	57.79
実験 1 (Filtered)	0.550	52.81	56.38	57.82
実験 2	0.168	10.94	19.37	27.95
実験 2 (Filtered)	0.169	11.04	19.66	28.05

の見出し語の説明文に“point”が含まれていることが予測に影響を与えたと考えられる。そのため、tail の説明文の情報が結果に大きく影響することがわかった。

実験 2 でテスト triple (“evidence, an indication that makes …”, “hypernym”, “indication, something that serves to indicate …”) の見出し語 “indication” とその説明文を “[MASK]” に置き換えて tail を予測すると、提案手法では “indication” を予測して正解したのに対し、KG-BERT では “avermint” を予測していた。これは head の見出し語の説明文に “indication” が含まれていることが予測に影響を与えたと考えられる。このような head の見出し語の説明文に tail の情報を含む triple に対して提案手法は有効であることがわかった。

6 まとめと今後の課題

本研究では、MLM を用いた 2 種類の入力文に対する tail 予測について KG-BERT と比較し、実験 1 において KG-BERT を上回る結果が得られた。また、実験 2 では提案手法が head の見出し語の説明文に tail の情報を含む triple に対して有効であることがわかった。BERT の MLM を用いた Knowledge Graph 補完において、学習していない見出し語でも head と relation と説明文のような前後の文章から適切に予測することができたため、Knowledge Graph 補完手法としての有効性が確認できた。

今後の課題として、MLM の出力候補を entity に限定したモデルの作成、head 予測や relation 予測への応用が挙げられる。

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.
- [7] Christiane Fellbaum. In theory and applications of ontology: computer applications. *Springer*, pp. 231–243, 2010.