

知識グラフに基づく応答文生成における 記号化されたエンティティの利用

梶井裕貴¹ 滝口哲也² 有木康雄²

¹ 神戸大学工学部情報知能工学科 ² 神戸大学大学院システム情報学研究科
1935070t@gsuite.kobe-u.ac.jp {takigu,ariki}@kobe-u.ac.jp

概要

近年では、生成ベースの応答文生成において、対話履歴のみではなく、外部のデータベースから検索した外部知識を利用することによって、情報に富んだ応答文生成が可能となることがわかっている。一方、既存の生成手法では、応答文生成モデルの学習に知識を具体的な名称（エンティティ名）のまま用いている。外部から必要な知識を得るのであれば、具体的な知識を学習せず、知識の用い方のみを学習すべきであると本研究では考え、具体的な名称を記号化し、知識の用い方を学習しやすくする手法を提案する。実験により、提案手法は、既存手法よりも、多くの知識を応答文に取り入れることが可能であることがわかった。

1 はじめに

雑談対話システムの課題として、一般的な応答ではなく、情報に富んだ応答生成が挙げられる。生成ベースの手法においては、学習時に対話履歴だけではなく、外部から検索した知識を組み込むことで、情報の多い応答文生成が可能となる [1]。しかし、既存の応答文生成モデルでは、参照すべき知識が与えられているにも関わらず、知識を用いずに応答文を生成してしまう現象、あるいは誤った使い方をしてしまう現象があることもわかっている [2]。

この問題に対処する方法として、具体的に知識を学習するのではなく、知識の用い方、すなわち応答文中のどこに知識を含めるかを学習する方法が考えられる。[3]では知識グラフを用いた場合を対象として、Entity Name Constrained Decoding(ENCD)という手法が提案されている (3.1 節参照)。しかし、この手法では、知識を具体的な名称のまま学習に用いており、具体的な知識が学習されてしまう。この結果、知識の用い方だけではなく、具体的な知識もパ

ラメータ中に保持されてしまい、学習できる用い方が減少する問題が発生する。その場合、与えられた外部知識を利用できない可能性が増してしまう。

本研究では、ENCDにおけるこの問題を解消する手法として、参照している知識のエンティティ名を記号化する手法を提案する。エンティティ名を特殊トークンに置き換えることで、具体的な名称での学習を避け、知識の用い方のみを学習させることを目的とする。自動評価による比較実験の結果、提案手法は、既存の手法と比べて、応答文としての適切性を保ちつつ、より多くの知識を利用できることがわかった。

2 関連研究

2.1 外部知識の利用

対話システムが参照する外部知識としては、文章のような非構造的な知識から、テーブルデータベースのような構造的な知識も存在している。提案手法では、外部知識として構造的な知識である知識グラフを使用する。知識グラフは知識の追加や削除、更新が行いやすく、大規模な知識グラフとして、代表的なものに Wikidata[4]がある。知識グラフの例を図1に示す。

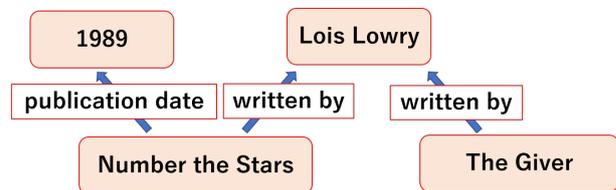


図1 知識グラフの例

2.2 知識グラフに基づく応答文生成

知識グラフでは、知識はトリプル (s,r,o) で表現されている。s,o がエンティティ、r が s,o の関係を表すリレーションである。知識グラフに基づく対話例

を表 1 に示す。

知識グラフに基づいて、応答文生成を行う方法として、いくつか提案されているものがある。一つは、検索してきた知識トリプルを、あらかじめ作成したテンプレートに当てはめて生成するものである [5]。別の方法は、知識グラフを GNN に、対話履歴を LSTM などのエンコーダに入力してそれぞれの分散表現を獲得し、それらを結合して応答文生成モデルに入力する方法である [6]。さらに、対話履歴と知識トリプルを繋ぎ合わせて一つのテキストとして表現し、生成モデルに入力して応答文を生成する手法もある [7]。この場合は、対話履歴 $D = (d_1, d_2, \dots, d_m)$ 、知識トリプル $T = (t_1, t_2, \dots, t_n)$ 、応答文 r を用いて、以下の式 1 のように表される。 θ は学習パラメータである。

$$P_{\theta}(r|D, T) = P(r|d_1, d_2, \dots, d_m, t_1, t_2, \dots, t_n) \quad (1)$$

今回は、対話履歴と知識トリプルを繋ぎ合わせて、式 1 のようにモデルに入力する方式を用いる。

表 1 知識グラフに基づく対話例

対話履歴	Do you like the book Number the Stars ?
知識トリプル	Number the Stars, written by, Lois Lowry
応答文	I do . Are you interested in other books by Lois Lowry ?

3 提案手法

3.1 Entity Name Constrained Decoding

提案手法のもとになった **Entity Name Constrained Decoding(ENCD)**[3] について述べる。これは、応答文生成時、与えた知識を用いるよう、モデルに促す手法である。ENCD は、学習時に適用される **エンティティタグ付き学習** と、推論時に適用される **エンティティ名制約モジュール (ENCM:Entity Name Constraint Module)** の 2 つの要素で構成される。

エンティティタグ付き学習 学習データの応答文において、参照する知識トリプルに含まれるエンティティ名の前後を” [“, ”] ” のタグで囲う。学習の際は、タグを含めて生成を行うように学習する。この学習により、応答文生成モデルは、知識を参照し始めると” [“を生成し、参照し終わると”] ” を生成するようになり、与えられた知識を参照するタイミングを学習する。表 1 の対話例に、タグ挿入を行った例を表 2 に示す。

エンティティ名制約モジュール (ENCM) エンティティタグ付き学習により、知識を参照するタイ

ミングがわかるようになるので、推論時に知識を参照する際には、ENCM を用いて、参照する知識トリプル中のエンティティ名のみを生成するよう、制約を加える。全体図を図 A(付録) に示す。

ENCD では、知識の使い方（応答文中のどこに知識を含めるか）を学習できていると考えられる。しかし、学習時にはエンティティ名は具体的なままであることから、知識の使い方と具体的な知識内容がともに学習されてしまうという問題がある。外部から知識を得るのであれば、具体的な知識内容の学習は排除し、使い方のみの学習にすべきであると考え、本研究では、エンティティを記号化することにより、ENCD を拡張する方法を提案する。

表 2 タグの挿入例

対話履歴	Do you like the book Number the Stars ?
知識トリプル	Number the Stars, written by, Lois Lowry
応答文+タグ	I do . Are you interested in other books by [Lois Lowry]?

3.2 エンティティの記号化

本論文では、エンティティを SKA, SKB といった特殊トークンに置換することで、エンティティを記号化する。具体例として、表 2 のタグ挿入を行った対話例に記号化を行ったものを表 3 に、提案手法を用いた応答文生成の全体図を図 2 に示す。記号化により、具体的な知識内容の学習を回避しながら、対話履歴、知識トリプル中のリレーションをもとにして、応答文中のどこに知識を挿入するかを学習していくことが可能となる。

なお、このエンティティ記号化は、学習時には対話履歴、知識トリプル、応答文に、推論時には対話履歴、知識トリプルに対して行う。モデルが生成する応答文は、SKA, SKB などを使ったものになるので、元のエンティティに戻して最終的な生成文とする。この際、記号化されたエンティティが元の知識トリプルでどのエンティティであったかは、SKA, SKB 中の A, B によって判断することができる。また、A, B によって知識トリプル中での関係を学習することも期待できる。

表 3 エンティティ記号化の例

対話履歴+記号化	Do you like the book SKA ?
知識トリプル+記号化	SKA, written by, SKB
応答文+タグ+記号化	I do . Are you interested in other books by [SKB]?

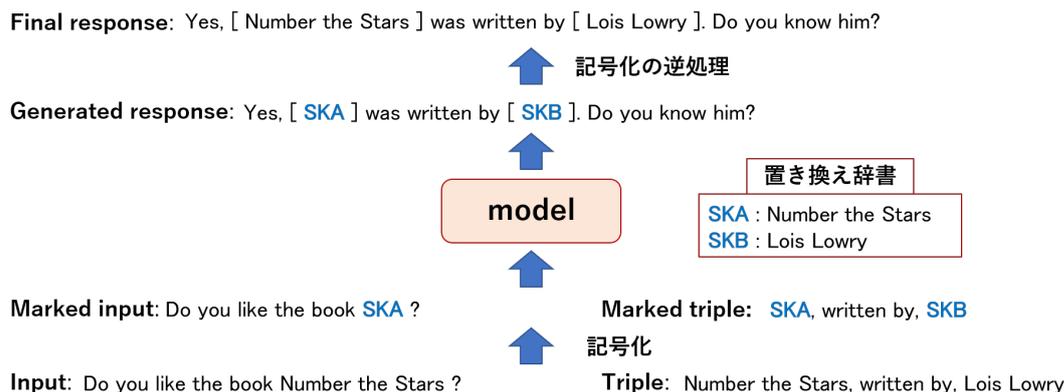


図2 エンティティ記号化を利用した応答文生成

4 実験設定

4.1 データセット

学習と評価には、OpenDialKG[8]を用いた。これは、各応答が知識グラフのどのトリプルを参照したかアノテーションされたマルチターンのデータセットである。知識グラフには、freebase[9]のサブセットが用いられている (freebaseは現在では、Wikidataに統合されている)。なお、知識を参照していない応答もあるため、それらは前処理として除外し、学習データ22589文、検証データ3535文、評価データ2832文が得られた。

4.2 評価指標

[10]に倣い、F1, KF1を用いる。F1は、モデルが生成した応答文と正解の応答文とのユニグラム一致度を測るものである。一方、KF1は生成した応答文と参照している知識とのユニグラム一致度を測るものである。生成した応答文、正解の応答文のユニグラムのリストをそれぞれ R_g , R_t とすると、F1は以下の式4-1, 4-2, 4-3で計算される。

$$Precision = \frac{len(set(R_g) \cap set(R_t))}{len(R_g)} \quad (4-1)$$

$$Recall = \frac{len(set(R_g) \cap set(R_t))}{len(R_t)} \quad (4-2)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4-3)$$

なお、KF1は、知識トリプルをテキスト表現にし、これを R_t として計算できる。F1が高いほど正解の応答文との類似度が高く、KF1が高いほど参照して

いる知識をより多く取り入れたことを示す。

また、知識の使い方が学習できているかを評価するため、推論時に参照する知識トリプル中のエンティティを、ランダムに選択した別のエンティティに置換した場合の評価も行う。全体の流れを図3に示す。使い方を学習できている場合、与える知識トリプル中のエンティティと該当する対話履歴中のエンティティを変更した場合、エンティティ部分のみが異なる応答文が生成される。図3の例だと、通常の場合は、「He wrote Les Fleurs de mal. Have you read it?」が生成され、ランダムなエンティティの場合では、「He wrote Apocalypto. Have you read it?」が生成される。そして、エンティティ部分を置換前のエンティティに戻せば、通常の知識トリプルと対話履歴を与えた場合と比較して、F1, KF1が同じ値になるはずである。今回の実験では、通常の場合とランダムなエンティティに置換した場合でF1, KF1がどれだけ変化するかを確認する。なお、ランダムに置換する際は、共通のリレーションを持つ他のトリプルを二つ選択し、それぞれからs, oを抜き出して置換した。

4.3 ベースライン

事前学習済み言語モデルであるT5[11]のsmallモデルに対して以下の2パターンでファインチューニングしたものを、ベースラインとして用いる。なお、入力には、2.2節で述べた、対話履歴と知識トリプルを繋ぎ合わせる入力方法を用いた。

Baseline-original 応答文をそのまま用いて学習し、推論時は通常のデコーディングを行うもの。

Baseline-encd 3.1節で述べた、エンティティタグ付き学習を用いて学習をし、推論時はENCMを用いるもの。

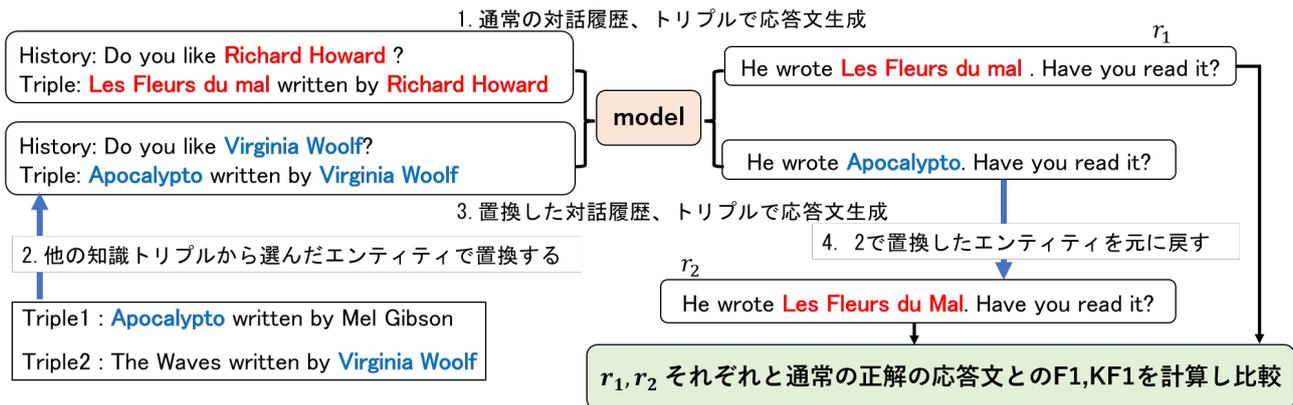


図3 エンティティをランダムに置換した知識トリプルを用いた評価の流れ

表4 生成された文章の例

対話履歴	I like the book The Hobbit or There and Back Again . Can you show me some books like it ?
知識トリプル	The Hobbit or There and Back Again, written by, J.R.R. Tolkien The Return of the King, written by, J.R.R. Tolkien
Baseline-original による応答文	Sure, The Return of the King is a good one. Have you read it?
Baseline-encd による応答文	[The Return of the King] is a good one. Have you read it?
提案手法による応答文	[The Hobbit or There and Back Again] is written by [J.R.R. Tolkien]. [J.R.R. Tolkien] also wrote [The Return of the King].

4.4 実装

実装は、[3] に倣い、ベースライン、提案手法ともに Python の OSS ライブラリ Transformers(<https://huggingface.co/google/t5-v1.1-small>) を用いて実装を行った。最適化手法は AdamW[12]、学習率 $1e-3$ 、バッチサイズ 16 で学習を行い、推論時にはビーム幅 3 でビームサーチを行った（これらの設定も [3] を参考にした）。

5 実験結果

実験結果を表 5 に示す。通常、ランダムないずれの場合でも、KF1 は提案手法で最も高いことがわかる。これから、具体的な知識を学習しなくなった分、知識の使い方を多く学習できたと考えられる。生成の例を表 4 に示す。Baseline-original と Baseline-encd では、与えられた知識トリプルのうち一つしか参照できていない。提案手法では、どちらも用いることができ、作者が共通しているその本を薦めているという内容の文が生成できている。

一方、表 5 の通常の場合、F1 については提案手法と Baseline-encd とでほとんど変化せず、Baseline-originsl を上回った。この結果から、応答文としての適切性は変化してないことがわかる。これは、知識以外の単語生成については、参照するエンティ

表5 実験結果

Model	通常		ランダム	
	F1	KF1	F1	KF1
Baseline-original	36.65	46.92	32.59	42.37
Baseline-encd	38.64	54.88	35.55	52.80
提案手法	38.14	59.23	38.14	59.23

ティ名ではなく、対話履歴やリレーションが大きく影響しているからであると考えられる。この点は、ENCND の知識を参照するステップとそうでないタイミングを切り替えて学習できるという利点だが、提案手法でも保たれているためであると考えられる。

また、ランダムなエンティティに知識を変更した場合、ベースラインの2つでは F1, KF1 のスコアが低下したが、提案手法では変化しなかった。

6 終わりに

本論文では、知識グラフに基づく応答文生成において、エンティティを記号化することにより、応答文としての適切性を保ちつつ、知識の使い方（応答文中のどこに知識を含めるか）が学習できるようになることを示した。自動評価の結果、従来手法よりも知識の使い方を多く学習できていることがわかった。これは、記号化により具体的な知識内容では学習を行わず、使い方といったメタな学習が可能になったためであると考えられる。

謝辞

本研究の一部は、JSPS 科研費 JP21H00906 の支援を受けたものである。

参考文献

- [1] Mojtaba Komeili, Kurt Shuster, and Jason Weston. Internet-augmented dialogue generation. **arXiv preprint arXiv:2107.07566**, 2021.
- [2] Nouha Dziri, Andrea Madotto, Osmar Zaiane, and Avishek Joey Bose. Neural path hunter: Reducing hallucination in dialogue systems via path grounding. **arXiv preprint arXiv:2104.08455**, 2021.
- [3] 佐良和孝, 滝口哲也, 有木康雄. 事前学習済み言語生成モデルを用いた知識グラフ検索に基づく対話システム. 人工知能学会研究会資料 言語・音声理解と対話処理研究会 93 回 (2021/11), pp. 44–49. 一般社団法人 人工知能学会, 2021.
- [4] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. **Communications of the ACM**, Vol. 57, No. 10, pp. 78–85, 2014.
- [5] 曾傑, 中野有紀子. 知識と話題の埋め込み表現に基づく質問生成と対話システムへの適用—料理嗜好インタビューシステムに向けて—. 自然言語処理, Vol. 28, No. 2, pp. 598–631, 2021.
- [6] Yueqing Sun, Qi Shi, Le Qi, and Yu Zhang. Jointlk: Joint reasoning with language models and knowledge graphs for commonsense question answering. **arXiv preprint arXiv:2112.02732**, 2021.
- [7] Yu Li, Baolin Peng, Yelong Shen, Yi Mao, Lars Liden, Zhou Yu, and Jianfeng Gao. Knowledge-grounded dialogue generation with a unified knowledge representation. **arXiv preprint arXiv:2112.07924**, 2021.
- [8] Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 845–854, 2019.
- [9] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In **Proceedings of the 2008 ACM SIGMOD international conference on Management of data**, pp. 1247–1250, 2008.
- [10] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. Retrieval augmentation reduces hallucination in conversation. **arXiv preprint arXiv:2104.07567**, 2021.
- [11] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. **J. Mach. Learn. Res.**, Vol. 21, No. 140, pp. 1–67, 2020.
- [12] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. **arXiv preprint arXiv:1711.05101**, 2017.

A Entity Name Constrained Decoding の全体図

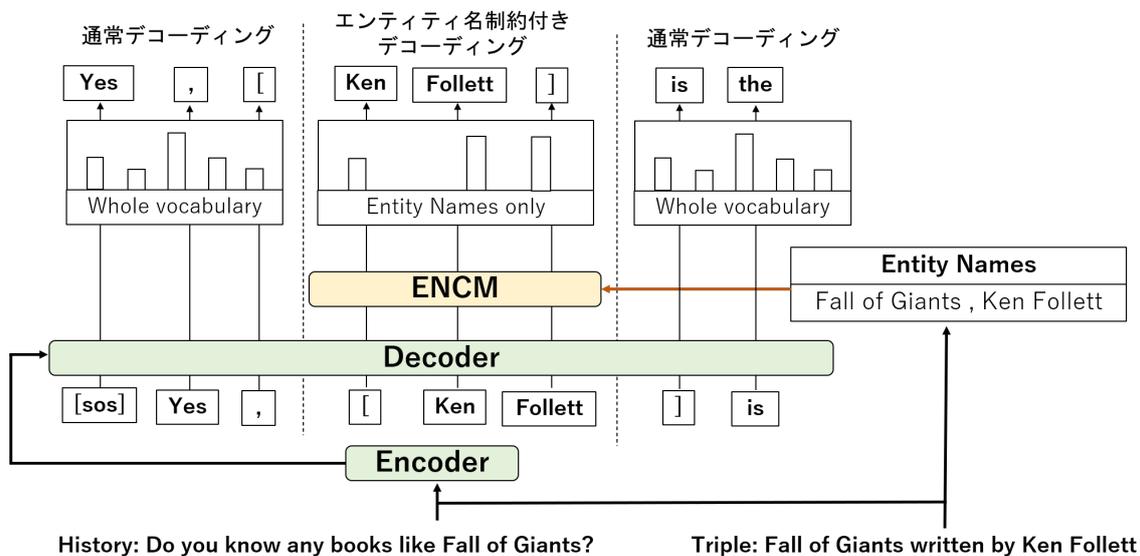


図 A Entity Name Constrained Decoding の全体図