

知識グラフに基づく応答文生成におけるエンティティ名制約付きデコーディング

佐良和孝 滝口哲也 有木康雄

神戸大学大学院システム情報学研究科

203x110x@egsuite.kobe-u.ac.jp {takigu,ariki}@kobe-u.ac.jp

概要

近年、生成ベースの雑談対話システムにおいて、対話履歴のみではなく、検索してきた外部知識を参照することで、情報に富んだ応答文の生成が可能となっている。しかしながら、既存の生成手法では、検索してきた知識を利用せず応答文を生成してしまう場合がある。本研究では、外部知識として知識グラフを用いた応答文生成モデルにおいて、検索してきた知識の利用を促す手法として、エンティティ名制約付きデコーディングを提案する。デコーディングの際に、知識に関する語彙を生成する場合は、検索した知識に含まれるエンティティ名のみを生成するように制約をかけることで、検索した知識の利用を促す。自動評価の結果、提案手法は既存手法と比較し、検索してきた知識をより多く利用できていることが分かった。

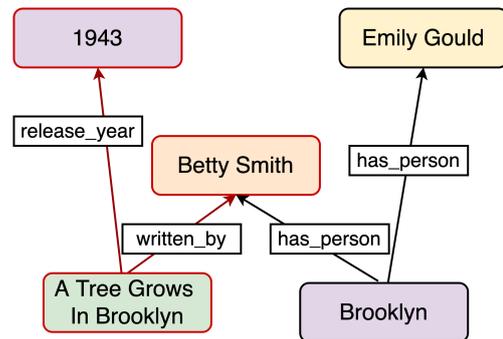
1 はじめに

雑談対話 AI の画期的な課題は、ユーザーに対して単に一般的な文章を応答するのではなく、より情報や知識に富んだ応答を行うことである。その際に、生成する応答文に含まれる情報や知識が正しいことが求められる。近年では、応答生成を学習する際に、対話履歴のみではなく、外部から検索してきた知識を入力しながら学習を行うことで、正しい知識を含んだ、情報に富んだ応答の生成が可能になっている [1]。このような形で学習を行うことで、知識自体をモデルのパラメータに保持するのではなく、検索してきた知識を利用しながら応答文を生成する仕組みをパラメータに保持することができる。これにより、事実知識に反する尤もらしい文を生成してしまう「Hallucination」といった問題を軽減することができ [2]、さらには、学習データに含まれていない、

表1 知識グラフに基づく対話の例

対話履歴	What can you tell me about Betty Smith ?
知識	[A Tree Grows In Brooklyn, release_year, 1943] [A Tree Grows In ..., written_by, Betty Smith]
応答文	she is an author who released a book in 1943 called A Tree Grows In Brooklyn

知識グラフ



新たな知識を用いた応答文の生成が、モデルの再学習を必要とせず可能になるといった利点がある。

しかしながら、既存の応答文生成モデルは、参照すべき知識が与えられているにも関わらず、その知識を用いることができない、もしくは、間違った使い方をしてしまうといった現象があることが報告されている [3]。

本研究では、知識グラフを外部知識として用いる応答生成モデルが、検索してきた知識をより利用するよう促すための手法として、エンティティ名制約付きデコーディング (ENCD: **Entity Name Constrained Decoding**) を提案する。ENCD は、デコーディング時に、モデルが知識に関する語彙を生成しようとした場合、参照している知識に含まれるエンティティ名のみを生成するように、モデルが出力する確率分布

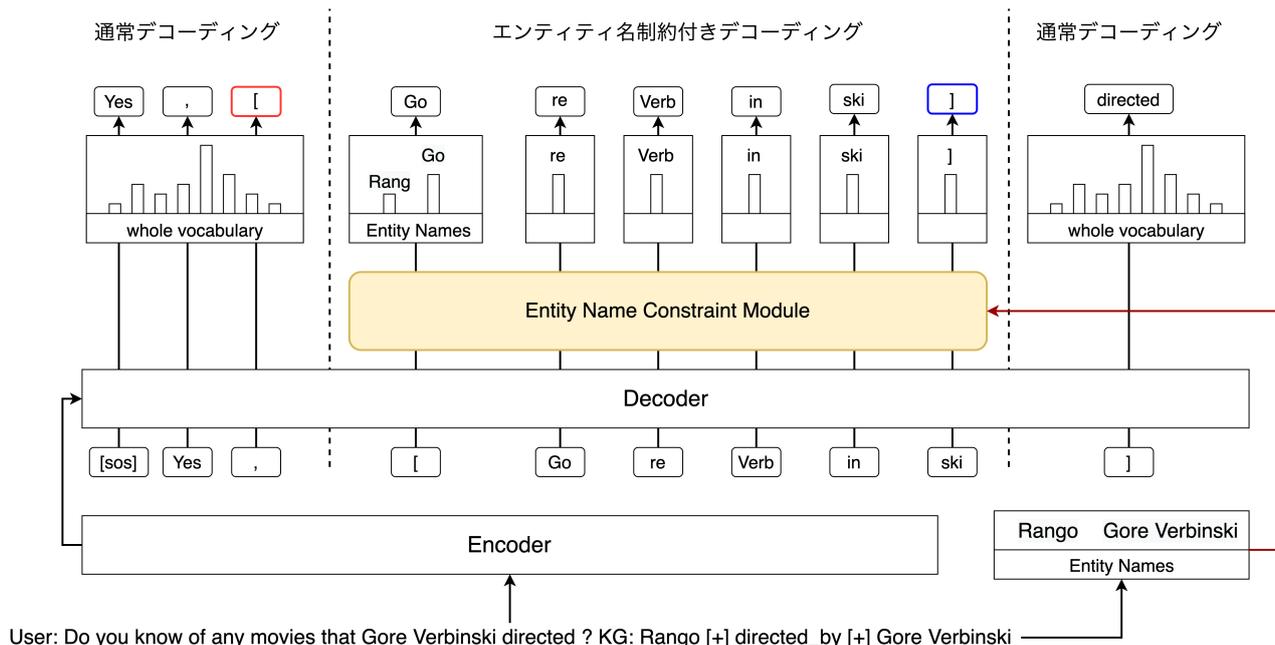


図 1 Entity Name Constrained Decoding

に対し、制約を行う。提案手法は、学習データに対する簡単な処理と、デコーディング時の処理のみを含んでいるため、多くのモデルに適用可能である。

自動評価を用いた比較実験の結果、提案手法は既存の手法と比較し、より多くの参照知識を利用できていることが分かった。また、提案手法は、既存手法と比較して、より正しい知識の使い方を学習できている可能性があることを示した。

2 知識グラフに基づく応答文生成

知識グラフに基づく対話例を表 1 に示す。知識グラフに基づく応答文生成は、知識グラフから検索された、応答生成の際に参照する知識トリプル $T = (t_1, t_2, \dots, t_n)$ 、対話履歴 $D = (d_1, d_2, \dots, d_m)$ 、応答文 r を用いて式 1 のように定義できる。 θ は学習されるパラメータである。

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

近年では、事前学習済み言語モデルに対して、対話履歴と知識トリプルを繋ぎ合わせ 1 つのテキストとして表現したものを入力し、応答文を生成するようにファインチューニングすることでモデル化を行う手法が広く用いられている。[4, 5]。このようにして学習されたモデルは、与えられた知識トリプルを参照しながら応答文を生成することが可能であるため、間違った知識を生成するような現象を防いだ

り、学習データに含まれないような新たな知識を用いた応答文の生成が可能になる。

3 提案手法

提案手法である ENCD は「エンティティタグ付き学習」と「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。1 つ目は学習段階で適用されるものであり、2 つ目はデコーディング段階で適用されるものである。それぞれの要素について詳しく述べる。

3.1 エンティティタグ付き学習

ENCD はデコーディングの段階で、モデルが知識に関する語彙を生成する場合、参照する知識トリプルに含まれるエンティティの名称のみを出力するように制限する手法である。しかしながら、従来の学習方法で学習されたモデルは、デコーディングのどのタイムステップで知識に関する語彙を生成しようとしているのかを明示的に知ることは難しい。そこで、本研究では、この問題に対処する手法として、エンティティタグ付き学習を提案する。

この手法は、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 "[" と "]" で囲うようにタグを挿入する。アノテーションの具体例を表 2 に示す。学習の際は、挿入したタグも含めて生成を行うように学習を行う。これによって、モデルは知識を参照する際には "[" トーク

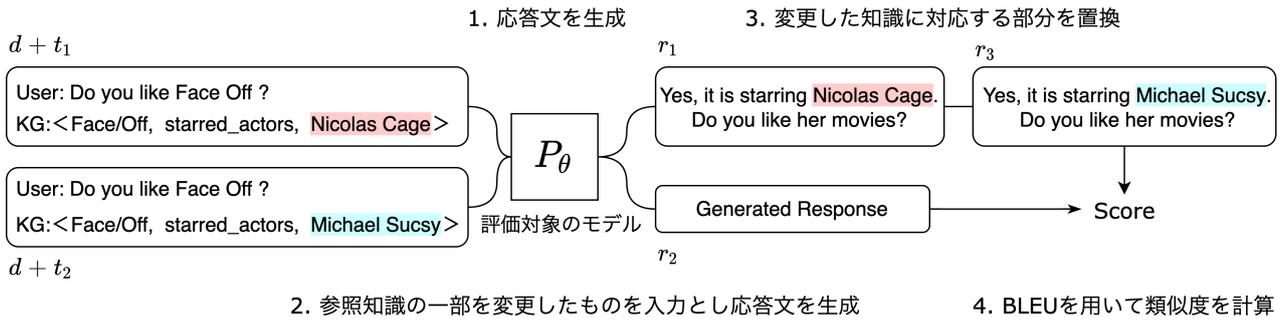


図2 Entity Replace Sentence Score の算出手順

表2 エンティティタグの挿入例

応答文	she is an author who released a book in 1943 called A Tree Grows In Brooklyn
知識	[A Tree Grows In Brooklyn, release_year, 1943]
タグ付き 応答文	she is an author who released a book in [1943] called [A Tree Grows In Brooklyn]

ンを、参照が終わると"]"トークンを生成するようにモデルを学習することが可能になり、デコーディング時に、どのタイムステップで知識に関する語彙を生成しようとしているかが判別できるようになる。今回の実験では、エンティティタグの挿入処理に、文字列マッチングを用いた。

3.2 エンティティ名制約モジュール (ENCM)

先ほど述べた、エンティティタグ付き学習を行うことで、デコーディングの際に、現在のステップで、モデルが知識トリプルを参照しているかどうかが明示的に分かるようになる。そこで、モデルが知識トリプルを参照している場合は、ENCMを用いて、トリプル内に含まれるエンティティ名のみを生成できるように、モデルが出力する確率分布に制約を加える。提案手法の全体図を図1に示す。

4 実験設定

4.1 データセット

学習と評価には、OpendialKG[6]を用いた。91Kのマルチターンの会話からなり、それぞれが、タスク指向(推薦)の会話、または与えられたトピックに関する雑談の会話である。それぞれの発話は、話者が発話をする際に、知識グラフ上のどのトリプルを参照したかがアノテーションされている。知識グラフにはfreebase[7]のサブセットが用いられてお

り、100,813個のエンティティ、1,358種類のエッジ、1,190,658個のトリプルで構成されている。データの中には、発話の際に知識を参照していないようなものがあるため、今回の実験ではそれらは前処理として取り除き、80%を学習データ、10%を開発データ、10%をテストデータとして用いた。

4.2 評価指標

モデルが知識を用いながら応答を生成できているか、3つの自動評価手法を用いて評価を行った。

BLEU[7] モデルが生成した応答文と、真の応答文との間で評価を行う。真の応答文は、モデルが参照する知識グラフと同じものを、人間が参照しながら応答したものである。

Entity Score モデルが生成した応答文の中に、参照した知識トリプルを構成するエンティティが、どれほど含まれているかで評価を行う。

Entity Replaced Sentence Score ただ単に知識を利用するのではなく、モデルがどれほど知識の用い方を学習しているかを評価するため、新たな評価指標としてERSS: **Entity Replacement Sentence Score**を提案する。モデルが知識をそのまま記憶しているのではなく、知識の利用方法を学習できている場合、与える対話履歴をそのままにして、与える知識トリプルの一部を変更してモデルに入力すると、変更された知識に対応する部分のみが変化した応答文が生成されるはずである。図2に提案する評価手法の全体図を示す。

はじめに、評価対象の学習済み応答文生成モデルに対して、対話履歴 d と参照する知識トリプル t_1 を入力し、応答文 r_1 を生成する。次に、 d と、知識トリプル t_1 の一部を、別のエンティティに置き換えた t_2 を入力とし、応答文 r_2 を生成する。そして、置換されたエンティティに対応する r_1 中のスパンを特定し、新たなエンティティに置き換えることで r_3 を

表3 応答文生成例. ENCM を用いることで、間違った知識を用いずに応答を生成することができている。

対話履歴	Assistant: That 's correct . John Heard also starred in The Guardian and The Great Debaters . Both are drama if you like that genre . User: No I do n't really like dramas . I liked John Heard in White Chicks though ! Who else was in that ?
知識トリプル	[White Chicks, starred_actors, John Heard] [White Chicks, starred_actors, Lochlyn Munro]
T5-small	John Heard also starred in White Chicks.
+ ENCD	He also starred in [White Chicks] with [Lochlyn Munro].
+ ENCD-w/o ENCM	He also starred in [White Chicks] with [Glenlyn Munro].

生成する。最後に r_2 と r_3 の一致度を BLEU スコアを用いて算出することで、最終的な評価値を得る。

今回の実験では、エンティティを置換する方法として、2通りの手法を用いる。1つ目は、同じ属性を持つエンティティに置き換えるものである。例えば、"Japan"というエンティティを置換する際は、同じ、国の属性を持つ、"America"や"Germany"といったエンティティに置き換える。2つ目は、エンティティの属性を考慮せず、ランダムに置き換える。本研究において、前者を ERSS-1、後者を ERSS-2 とする。

4.3 ベースライン

応答文生成モデルとして、事前学習済み言語モデル T5[8] の small モデルをファイン・チューニングしたものを用いて、通常のデコーディング手法で応答文を生成するものをベースラインとして用いる。

4.4 実装

ベースライン、提案手法ともに、Python の OSS ライブラリ、Transformers(<https://github.com/huggingface/transformers>)を用いて実装を行った。最適化手法は AdamW[9]、学習率 $1e-3$ 、バッチサイズ 16 で学習を行い、推論時はビーム幅 3 でビームサーチを行った。

5 実験結果

実験結果を表 4 に示す。表から分かる通り、従来手法 (T5-small) と比較し、全ての指標で提案手法が優れていることが分かる。アブレーション・スタディとして ENCM を用いない場合、すなわち、学習段階でエンティティタグを挿入するが、デコーディング時は通常のデコーディング手法を用いる場合についても評価を行った。その結果、ENCM を用い

表4 実験結果

Model	BLEU	Entity	ERSS-1	ERSS-2
T5-small	12.58	59.16	50.32	41.22
+ ENCD	12.88	66.58	56.84	44.73
+ ENCD-w/o ENCM	12.89	66.16	56.77	44.84

ない場合でも、従来手法と比較し優れていることが分かり、学習段階でのエンティティタグの挿入が、モデルの学習に大きな影響を与えていることが分かる。詳しい理由はまだ解析できていないが、エンティティタグを挿入することで、応答文生成の際に知識を参照するステップとそれ以外のステップの切り替えをうまく学習できていることが要因であると考えている。この原因特定については今後の課題である。ENCM を用いる場合と用いない場合の比較では、自動評価の観点から見ると、ほぼ同じであるが、表 3 に示す、実際の生成例を見てみると、ENCM を用いることで、エンティティ名を間違って出力することを防ぐことが出来ている。このことが ENCM を用いた場合の Entity Score に表れていると考えられる。

6 おわりに

本研究では、知識グラフに基づく応答文生成において、検索してきた知識の利用を促す手法として ENCD を提案した。自動評価の結果、従来手法と比較し、より多くの検索知識を利用できていることが分かった。また、提案手法は正しい知識の利用方法を学習できている可能性があることを示した。

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, 2021.
- [2] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. Retrieval augmentation reduces hallucination in conversation, 2021.
- [3] Nouha Dziri, Andrea Madotto, Osmar Zaiane, and Avishek Joey Bose. Neural path hunter: Reducing hallucination in dialogue systems via path grounding, 2021.
- [4] Yu Li, Baolin Peng, Yelong Shen, Yi Mao, Lars Liden, Zhou Yu, and Jianfeng Gao. Knowledge-grounded dialogue generation with a unified knowledge representation, 2021.
- [5] Andrea Madotto, Zhaojiang Lin, Yejin Bang, and Pascale Fung. The adapter-bot: All-in-one controllable conversational model, 2020.
- [6] 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**, July 2019.
- [7] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: A method for automatic evaluation of machine translation. In **Proceedings of the 40th Annual Meeting on Association for Computational Linguistics**, ACL '02, USA, 2002. Association for Computational Linguistics.
- [8] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2020.
- [9] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.