

ゲート制御付き Source-Target Attention を用いた複数知識文に基づく応答文の生成

Response generation based on multiple knowledge sentences using Source-Target Attention with Gate

佐良和孝^{1*} 滝口哲也¹ 有木康雄¹
Sara Kazutaka¹ Takiguchi Tetsuya¹ Ariki Yasuo¹

¹ 神戸大学システム情報学研究科

¹ Graduate School of System Informatics, Kobe University, Department of Information Science

Abstract: In recent years, Transformer model have been widely used in interactive systems to generate response sentences based on knowledge sentences. However, there is a limit to the number of knowledge sentences that can be referred to simultaneously in the Transformer due to the computational complexity of the Attention mechanism, since multiple knowledge sentences are concatenated and treated as one long sentence. To address this problem, we propose Source-Target Attention with Gate, which can feed multiple encoder outputs to the decoder simultaneously. This method does not require concatenation of knowledge sentences. We trained the proposed model using Persona-Chat dataset, and compare with the general Transformer model.

1 はじめに

近年、知識文に基づいた応答文を生成するモデルとして、Transformer モデル [1] が多く用いられている [2] [3]。複数の知識文に基づく応答文の生成を行う場合、それらの知識文を繋げて 1 つの系列としてモデルに入力することが一般的である。Transformer で用いられる Attention 機構の構造上、入力する系列が長くなると計算量が大幅に増加する問題がある。よって、知識文を繋ぎ合わせる手法では参照できる知識文の数に制限ができる。そこで、我々は複数の知識文をそれぞれ独立に Encoder へ入力し、複数の Encoder からの出力を Decoder に入力することが可能な、ゲート制御付き Source-Target Attention を用いる手法を提案する。この提案手法では、知識文を繋ぎ合わせることなくモデルへ入力することができるため、入力する知識文量に制限がない。

2 ゲート制御付き Source-Target Attention

従来手法: n 個の知識文を $S = (s_1, s_2, \dots, s_n)$ と表す。従来手法では、複数の知識文を Transformer モデルに入力する際、(1) 式のように知識文を繋ぎ合わせたものを 1 つの文として扱う。

$$\hat{s} = [s_1; s_2; \dots; s_n] \quad (1)$$

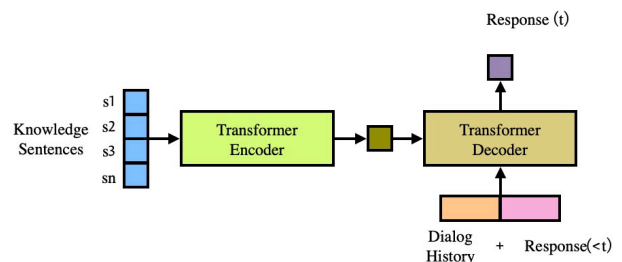


図 1: 従来手法

従来手法の概略図を図 1 に示す。従来手法の場合、知識文を繋ぎ合わせるため、参照する知識文量が増加するとモデルへの入力系列が長くなる。Transformer の Attention 機構は入力系列長の 2 次の計算量があるため、参照できる知識文量には制限が出来る。

提案手法: 提案手法では以上で述べた問題点を解決するために、知識文は繋ぎ合わせずに独立してエンコー

*連絡先: 神戸大学システム情報学研究科
〒657-8501 兵庫県神戸市灘区六甲台町 1-1 自然科学
総合研究棟 3 号館 805 号室
E-mail: 203x110x@gsuite.kobe-u.ac.jp

ドを行う。しかしながら、通常の Transformer Decoder は Encoder 側の出力を 1 つしか入力することができない。そこで [4] で用いられている手法を基に、複数の Encoder 側の出力を Decoder に入力することを可能にする、ゲート制御付き Source-Target Attention を提案する。知識文集合 S を Encoder を用いてエンコードしたものを $H = (h_1, h_2, \dots, h_n)$ とする。Decoder への入力を y とすると提案手法の Decoder 内での Source-Target Attention は式 (2)(3) で計算される。

$$Attention(H, y) = \sum_{i=1}^n G(h_i, y) \odot Attention(h_i, y) \quad (2)$$

$$G(h_i, y) = \sigma(W[h_i, Attention(h_i, y)] + b) \quad (3)$$

ここで W, b は学習可能なパラメータである。提案手法の概略図を図 2 に示す。

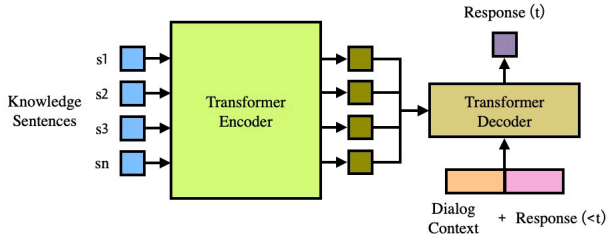


図 2: 提案手法

3 実験設定

本研究では Persona-Chat[5] を用い、提案手法を検証した。ベースラインとして、通常の Source-Target Attention を用いた Transformer を用意し比較を行った。ベースラインでは複数の知識文をつなげて、1 つの文として処理を行う。近年は Transformer モデルを用いる際、大量のコーパスを用いて事前学習を行う手法 [2][3][6] が主流であるが、今回の実験では事前学習は行っていない。

3.1 データセット

本研究では、知識に基づく会話データセットとして、Persona-Chat[5] を用いる。このデータセットは、それぞれの話者が 4, 5 文のプロフィール文 (I like to ski. , I like to eat cheetos. など) を与えられ、話者はそのプロフィール文に基づきながら会話を行ったオープンドメインの会話データである。本実験では生成する発話以前の 5 発話を対話履歴として用いた。

3.2 評価指標

真の応答文との関連性、多様性、知識文との関連性の 3 つの観点から評価を行う。

真の応答文との関連性: 生成した応答文と真の応答文の間で評価を行う。評価指標には PPL[7], BLEU[8], NIST[9], ROUGE[10] を用いた。

多様性: 生成した応答文内で評価を行う。評価指標には Distinct-N[11] を用いた。

知識文との関連性: 生成した応答文とプロフィール文の間で評価を行う。評価指標には BLEU, NIST, ROUGE を用いた。

3.3 ベースライン

Transformer モデルをベースラインとして用いる。対話履歴、知識文の入力形式の違いによる複数のベースラインを用意した。

model1: 知識文、対話履歴を繋ぎ合わせたものを Encoder、応答文を Decoder に入力する。

model2: 知識文を Encoder、対話履歴、応答文を繋ぎ合わせたものを Decoder に入力する。

model3: 対話履歴を Encoder、知識文と応答文を繋ぎ合わせたものを Decoder に入力する。

model4: 知識文、対話履歴、応答文を繋ぎ合わせたものを Decoder に入力する。

model5: 対話履歴を Encoder、応答文を Decoder に入力する。

model6: 対話履歴、応答文を繋ぎ合わせたものを Decoder に入力する。

ベースラインの入力形式についてのまとめを表 1 に示す。Encoder, Decoder 共に次元数は 256, Attention Heads 数は 8, 層数は 6 とした。

表 1: ベースラインモデルの入力形式 (k: 知識文, h: 対話履歴, r: 応答文)

Model	Encoder-Input	Decoder-Input
model 1	k + h	r
model 2	k	h + r
model 3	h	k + r
model 4	none	k + h + r
model 5	h	r
model 6	none	h + r

3.4 学習設定

全てのモデルを、optimizer は AdamW[12], 学習率 $1e-4$, バッチサイズ 16, Epoch 数 15 で学習した。

4 実験結果と考察

実験結果を表 2, 3, 4 に示す。提案手法と同様に、知識文のみを Encoder に入力する (知識文はつなぎ合わせて 1 つの文として扱う) model2 と提案手法を比較すると、PPL のみが向上しそれ以外の指標では劣化が見られた。Pesona-Chat は知識文の長さ、量が比較的少ないデータセットであるため、今後は Persona-Chat 以外の知識に基づくデータセットにおいても検証を行い、提案手法の分析を行う予定である。

ベースライン間での比較を行うと、対話履歴と応答文をつなげて入力しているモデル (model2, model4, model6) がそうではないモデル (model1, model3, model5) と比較して、多くの指標で優れていることが分かる。対話履歴と応答文はどちらも同じ発話文に分類されるため、Self-Attention を行うことが効果的であると考えられる。多様性について注目すると、知識文を Encoder へ、対話履歴と応答文を Decoder へ入力する model2 が他のモデルと比較して大きく優れていることが分かる。しかしながら、同様に知識を分離して入力する提案手法の場合は model2 と比較して大きく劣化しているため、その原因について今後調査を行っていききたい。

また、本実験では Transformer モデルを事前学習することなく用いたが、事前学習を行った場合についても同様の実験を行う予定である。

謝辞

本研究の一部は、JSPS 科研費 JP17H01995 の助成を受けたものである。

表 4 多様性

Method	Dist-1	Dist-2
model 1	0.0070	0.0227
model 2	0.0124	0.0440
model 3	0.0073	0.0235
model 4	0.0089	0.0325
model 5	0.0061	0.0193
model 6	0.0071	0.0246
Ours	0.0080	0.0285

表 2 真の応答文との関連度

Method	PPL	BLEU-4	NIST-4	ROUGE-1	ROUGE-2	ROUGE-L
model 1	39.96	0.0204	1.1357	0.1239	0.0192	0.1174
model 2	41.18	0.0217	1.1926	0.1304	0.0221	0.1240
model 3	42.33	0.0176	1.0889	0.1242	0.0193	0.1169
model 4	39.14	0.0234	1.2044	0.1359	0.0258	0.1293
model 5	40.45	0.0229	1.1325	0.1236	0.0184	0.1158
model 6	39.55	0.0177	1.0902	0.1264	0.0205	0.1195
Ours	39.28	0.0181	1.1430	0.1255	0.0206	0.1203

表 3 知識文との関連度

Method	BLEU-4	NIST-4	ROUGE-1	ROUGE-2	ROUGE-L
model 1	0.0242	1.3013	0.1399	0.0232	0.1249
model 2	0.0352	1.4109	0.1524	0.0261	0.1360
model 3	0.0227	1.4189	0.1381	0.0226	0.1242
model 4	0.0329	1.4094	0.1571	0.0319	0.1403
model 5	0.0239	1.2962	0.1401	0.0216	0.1242
model 6	0.0232	1.2184	0.1469	0.0241	0.1303
Ours	0.0261	1.2908	0.1481	0.0283	0.1355

参考文献

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [2] Thomas Wolf, Victor Sanh, Julien Chaumond, and Clement Delangue. Transfertransfo: A transfer learning approach for neural network based conversational agents, 2019.
- [3] Kurt Shuster, Da Ju, Stephen Roller, Emily Dinan, Y-Lan Boureau, and Jason Weston. The dialogue dodecathlon: Open-domain knowledge and image grounded conversational agents, 2020.
- [4] Marcella Cornia, Matteo Stefanini, Lorenzo Baraldi, and Rita Cucchiara. Meshed-memory transformer for image captioning, 2020.
- [5] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2204–2213, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [6] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation, 2020.
- [7] Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models, 2016.
- [8] 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 of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [9] George Doddington. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. pp. 138–145, 01 2002.
- [10] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [11] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 110–119, San Diego, California, June 2016. Association for Computational Linguistics.
- [12] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.