잠재 변수 모델링 기반 잠재 가중치 어텐션 계산을 통한 문맥적 답변 생성 기법

이종원 [1], 조인휘 [2]
[1] 한양대학교 컴퓨터소프트웨어학과 석사과정
[2] 한양대학교 컴퓨터공학과 교수

Ljw4735@hanyang.ac.kr, iwjoe@hanyang.ac.kr

Generating Contextual Answers Through Latent Weight Attention Calculations based on Latent Variable Modeling

Jong-won Lee[1], In-whee Joe[2]
[1]Dept. of Computer Software, Han-Yang University
[2]Dept. of Computer Science, Han-Yang University

요 약

최근 많은 분야에서 인공지능을 사용한 산업이 각광을 받고 있고 그중 챗-GPT 로 인하여 챗봇에 관한 관심도가 높아져 관련 연구가 많이 진행되고 있다. 특히 질문에 대한 답변을 생성해주는 분야에 대한 연구가 많이 이루어지고 있는데, 질문-답변의 데이터 셋에 대한 학습 방식보다는 질문-답변-배경지식으로 이루어진 데이터 셋에 대한 학습 방식이 많이 연구가 되고 있다. 그러다 보니 배경지식을 어떤 방식으로 모델에게 이해를 해줄 지가 모델 성능에 큰 부분 차지한다. 그리고 최근 연구에 따르면 이러한 배경지식 정보를 이해시키기 위해 잠재 변수 모델링 기법을 활용하는 것이 높은 성능을 갖는다고 하고 트랜스포머 기반 모델 중 생성 문제에서 강점을 보이는 BART(Bidirectional Auto-Regressive Transformer)[1]도 주로 활용된다고 한다. 본 논문에서는 BART 모델에 잠재 변수 모델링 기법 중 잠재 변수를 어텐션에 곱하는 방식을 이용한 모델을 통해 답변 생성 문제에 관한 해결법을 제시하고 그에 대한 결과로 배경지식 정보를 담은 답변을 보인다. 생성된 답변에 대한 평가는 기존에 사용되는 BLEU 방식과 배경지식을 고려한 방식의 BLEU 로 평가한다.

1. 서론

챗-GPT 의 등장으로 챗봇에 관한 관심도가 올라가면서 딥러닝을 이용한 답변 생성 모델들에 대한 연구가 활발하게 이루어지고 있다. 이러한 연구에서 데이터 셋은 크게 질문-답변, 질문-답변-배경지식 이렇게 두 가지 종류가 있다. 질문-답변 데이터 셋의 경우에는 질문에 대한 답변 정보만을 학습하기 때문에 새로운 답변을 생성하기에는 어려움이 있지만, 질문-답변-배경지식 데이터 셋의 경우에는 답변 생성에 있어서 배경지식에 어느 정도의 해답이 담겨 있기 때문에 최근 답변 생성 문제에서는 질문-답변-배경지식으로 이루어진 데이터 셋을 주로 활용한다. 그리고 배경지식이 있다는 강점을 갖는 데이터 셋을 활용하기 위해 배경지식 정보를 어떤 방식으로 모델에 적용시킬 지에 대한 연구가 이루어졌는데 그 중 잠재 변수를 활용한 모델링을 통해 모델을 구성하는 것이 좋은 성능을 갖는다고 하고, 최근 자연어 처리 분야에서 강세를 이어가는 트랜스포머 기반의 BART 모델은 이러한 답변 생성 문제에서 주로 활용되고 있다. 잠재 변수 모델링을 하는 방법은 여러가지가 있지만 본 논문에서는 인풋 데이터를 잠재 변수로 이용하여 BART 모델의 디코더 어텐션 계산에서 관여하여 생성 부분의 가중치를 부여할 수 있는 기법을 도입하여 답변 생성에서의 이점을 갖을 수 있는 모델링 방법을 소개한다.

2. 관련 연구

(1) 잠재 변수 모델링

답변 생성 문제에서, 답변 생성에 가장 큰 어려움은 대화 맥락과 답변을 연관시키는 것이다. 그리고 이전의 연구들에서 대화의 잠재 공간을 모델링하는 것이 이를 해결하는 데에 도움이 될 수 있고, 한가지 답변이 아닌 여러 답변을 생성하는 데에도 큰 도움이 된다고 설명한다.

DialogVED[2]에서 인코더-디코더 사전 훈련 프레임 워

크에 잠재 공간을 이용한 잠재 변수를 활용하여 답변의 관련성을 높이고 다양한 답변 결과를 얻었고, CKL[3]에서는 배경지식과 문맥에 대해 독립적인 잠재 변수를 두어 배경지식에 문맥 정보를 추가하는 방식을 활용하여 답변의 질을 높였다. 마지막으로 본 논문에서 기본 모델이 된 LMEDR[4]은 두 번의 학습 과정을 통해 두 잠재 변수를 얻어 답변 생성에 활용하여 배경지식의 활용을 극대화하였다.

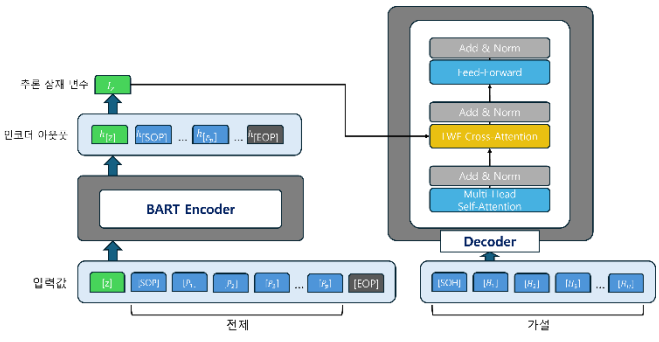
(2) LWE Attention (Latent Weight Enhanced Attention)

LWE Attention 은 기존 Attention 계산의 결과물에 잠재 변수를 곱하는 Attention 방식으로, CKL 은 기존의 Attention 계산은 두 문장 표현에 대한 단어 단위의 계산이라면 LWE Attention 은 문장 단위의 계산이라고 하였다. 그리고 이러한 계산 방식은 잠재 변수에 의한 정보를 더 의식하여 답변을 생성할 수 있는 계산 방식이기 때문에 답변 생성 시 사용되면 잠재 변수에 속한 배경지식과 질문에 대한 표현을 더 관여한 답변 생성이 가능하다.

$$LWE\ Attention(Q,K,V) = LW \times softmax\left(\frac{QK^T}{\sqrt{d}}\right)V$$

3. 모델

본 논문의 모델은 학습과 생성에서 end-to-end 방식으로 활용되는 BART 모델이 기본 모델로 활용되었고, 학습 방식은 LMEDR 의 방식을 차용하여 전제-가정으로 이루어져 있는 자연어 추론 데이터 셋을 통하여 먼저 학습하여 모델 전체의 파라미터를 조정하고 추론 잠재 변수를 얻고, 이후 질문-답변-배경지식의 데이터 셋으로 모델 전체를 학습하고 이때 발생하는 잠재 변수를 학습된 추론 잠재 변수와 더하여 디코더의 LWE Attention 에 활용하여 문맥을 이해한 답변 생성을 사전 훈련한다.



(그림 1) 추론 모델 구조

(1) 추론 학습

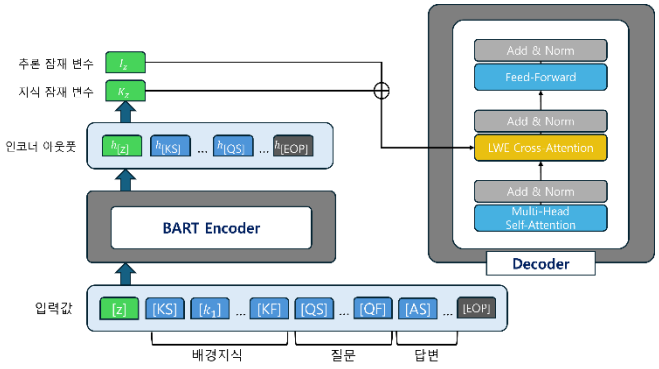
추론 학습은 주어진 정보의 일관성에 대한 수반 관계를 학습하고 저장하는 학습 단계로 주어진 전제에 대한 가정을 생성하도록 진행이 된다. 여기서 전제와 가정이 한 쌍의 수반 관계를 이룬다.

모델의 인코더 입력 값은 학습을 위해 스페셜 토큰인 잠재 토큰과 전제를 앞뒤로 감싸주는 토큰을 추가하는 과정을 거쳐 구성하고 디코더 입력 값은 가설을 앞뒤로 감싸주는 토큰만을 추가하여 구성한다. 이렇게 구성된 인코더 입력 값은 BART 인코더에 넣어 전제에 대한 잠재 표현을 얻고, 이 잠재 표현 값은 디코더의 학습 과정에서 각 트랜스포머 레이어에 적용되어 전제(P)-가설(H)에 관한 정보의 일관성을 형성하는데 도움을 주도록 하였다.

정보의 일관성을 표현하기 위한 학습된 추론 잠재 변수(z)을 얻고 추론 잠재 표현(θ)와 모델의 파라미터(φ)을 학습하기 위해 손실 함수는 언어 모델링 손실 함수($L_{추론}$)을 활용하였다.

$$L_{추론} = -\mathbb{E}_{z\sim p_{\theta,\varphi}(z|P)} \log p_{\theta,\varphi}(H|z,P)$$

$$= -\mathbb{E}_{\boldsymbol{z}\sim \boldsymbol{p}_{\theta,\varphi}(\boldsymbol{z}|\boldsymbol{P})} \sum_{t=1}^{|H|} \log \boldsymbol{p}_{\theta,\varphi}(\boldsymbol{H_t}|\boldsymbol{z},\boldsymbol{P},\boldsymbol{H_{<t}})$$



(그림 2) 전체 모델 구조

(2) 답변 생성 학습

답변 생성 학습에서는 추론 학습을 통해 얻은 일관성을 유지할 수 있는 학습된 BART 와 추론 잠재 표현 값을 가지고 추가적인 학습을 통해 일관성을 유지하면서 배경지식을 통한 더 자세한 질문의 답변을 생성하도록 학습한다.

인코더에는 가장 앞에 잠재 토큰을 두고 뒤로는 배경지식(C), 질문(Q), 답변(R)을 각각 스페셜 토큰으로 감싸서 만들어 낸 입력 값을 넣어 추론 학습 때와 비슷하게 입력 값에 대한 잠재 표현을 얻어내고 이 잠재 변수는 지식 잠재 변수라 한다. 그리고 추론 잠재 변수와 지식 잠재 변수는 각각 의미하는 바가 다르기 때문에 독립적이라고 볼 수 있어, 두 잠재 변수의 코사인 유사도를 낮게 만들어 줄 수록 더 높은 효과를 볼 수 있다고 판단하여 L2 정규화를 활용한 코사인 유사도로 손실 함수를 구성한다.

$$L_{잠재\ 변수} = \sum_{i\le k, j\le l}\left(\frac{M_i N_j^T}{\|M_i\|_2 \|N_j\|_2}\right)^2$$

디코더에는 입력 값으로 답변을 감싸주는 토큰을 추가하여 구성하고 추론 잠재 변수와 지식 잠재 변수을 합한

값을 각 트랜스포머 레이어의 LWE Cross Attention 에 적용하여 질 높은 답변을 생성하도록 한다.

마지막으로 답변 생성에 사용되는 손실 함수는 추론 학습에서 얻은 추론 잠재 변수를 활용하여 지식 잠재 변수(z^d)를 얻고 지식 잠재 표현(ϕ)과 모델의 파라미터(φ)을 학습하기 위해 언어 모델링 손실 함수($L_{답변\ 생성}$)식을 활용한다.

$$L_{답변\ 생성} = -\mathbb{E}_{z\sim p_\varphi(z|C),z^d\sim P_{\phi,\varphi}(z^d|C,Q)} \log_{p_{\phi,\varphi}}(R|C,Q,z,z^d)$$

$$= -\mathbb{E}_{z\sim p_\varphi(z|C),z^d\sim P_{\phi,\varphi}(z^d|C,Q)} \sum_{t=1}^{|R|} \log_{p_{\phi,\varphi}}(R_t|R_{<t},C,Q,z,z^d)$$

추가적으로, 잠재 변수를 촉진하기 위해 bag-of-words loss[5]를 $L_{잠재\ 변수}$로 활용하였다.

$$L_{잠재\ 변수} = -\mathbb{E}_{z\sim p_\varphi(z|C),z^d\sim P_{\phi,\varphi}(z^d|C,Q)} \sum_{t=1}^{|R|} \log_{p_{\phi,\varphi}}(R_t|C,Q,z,z^d)$$

$$= -\mathbb{E}_{z\sim p_\varphi(z|C),z^d\sim P_{\phi,\varphi}(z^d|C,Q)} \sum_{t=1}^{|R|} \log \frac{e^{f(R_t)}}{\sum_{v\in V} e^{f(v)}}$$

마지막으로, t 개의 답변이 생성되었을 때 모델이 올바른 답변을 선택하도록 훈련하는 방식인 (Wolf et al. 2019)[6]을 차용하여, 답변 생성 학습 시에 입력 값의 마지막 토큰을 통해 모델에서 생성한 답변 후보와 실제 값과의 cross-entropy loss 를 활용하였다.

$$\hat{y} = softmax(W_h h_{eos} + b_h)$$

$$L_{CEL} = -\sum_{i=1}^{t+1} \hat{y_i} \log(y_i)$$

그렇게 모델은 다음의 식을 최소화하도록 최적화를 진행하였다.

$$L(\phi,\varphi) = L_{추론} + L_{답변\ 생성} + L_{잠재\ 변수} + L_{CEL}$$

4. 실험

모델 학습 데이터로는 두번의 학습을 위한 두 가지 데이터 셋을 활용하였다. 첫번째 학습에는 MNLI[7] 데이터 셋을 활용하였는데, 이는 다국어 자연어 추론 말뭉치로 추론 표현을 학습할 수 있는 가장 큰 데이터 셋 중 하나이며 이번 학습에서는 영어 부분을 활용하여 모델을 학습하였다. 두번째 학습에는 DSTC7-AVSD[8] 데이터 셋을 활용하였다. 기존의 DSTC7-AVSD 데이터 셋은 이미지에 대한 주석과 요약이 주어지고 그에 대한 질문-답 형태의 대화가 존재하는 데이터 셋이다. 본 논문에서는 답변 생성 문제에 대한 모델을 다루기 때문에 이미지는 제외한 부분을 활용하였다. 그리고 비교 모델은 본 논문의 모델 아이디어가 된 LMEDR 과의 비교를 통해 실험 결과를 비교하였다.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
LMEDR	0.514	0.369	0.294	0.249
Our	**0.556**	**0.421**	**0.343**	**0.290**

<표 1> DSTC7-AVSD 의 배경지식과 bleu_score

먼저 모델이 생성한 답변은 배경지식인 주석과 요약을 기반하여 생성된 결과 값이기 때문에, 배경지식과 생성된 답변 간의 BLEU 점수를 확인해 보았다. 이 때는 기존의 모델인 LMEDR 보다 근소하지만 높은 점수를 얻어내는 것을 보였다.

하지만 다음 기존의 평가 방식인 DSTC7-AVSD 에서 제공하는 권장 답변 데이터 셋 두 가지 버전에 대하여 BLEU 점수를 확인해 보았을 때는 점수가 낮은 걸 확인하였다.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
LMEDR	**0.603**	**0.478**	**0.402**	**0.346**
Our	0.382	0.250	0.181	0.137

<표 2> DSTC7-AVSD 에서 제공한 데이터와 bleu_score

주석	a man sits on a couch reading a book. he closes the book and puts it on a table. he grabs a pillow from next to him, cuddles it, and seems to fall asleep with it.
요약	a man is reading a book while sitting on a couch. he closes the book and puts it down. then he cuddles a pillow and tries to take a nap.
질문	Q: are there any people in the video?
• **답변**	
Baseline	yes, there is one man in the video
LMEDR	there is one person in the video
Our	a man is sitting on a couch reading a book

주석	a man opens a window to another room. he then takes a photo of it on his smartphone then closes the window and sits down in a chair.
요약	he opens the window and takes some pictures. he then cleans the window and goes and sits down in the chair.
대화	Q: hello. did someone come to the door? R: no there is no one else in the room.
질문	Q: is he looking at something outside the window?
• **답변**	
Baseline	no he is just looking at it
LMEDR	no he is not looking at anything
Our	he is taking pictures outside the window

<표 3,4> 모델 결과 비교

표 3, 4 에서 Baseline 은 DSTC7-AVSD 에서 제공하는 권장 답변 데이터 셋을 의미한다. 모델의 결과를 비교해 보면 Baseline 과 LMEDR 의 경우에는 질문에 대한 정확하지만 단편적이라고 보일 수 있는 답변을 놓지만 본 논문의 모델의 경우에는 주석과 요약의 내용을 포함하는 답변을 내놓는 걸 볼 수 있다.

5. 결론

답변 생성 문제에서 문맥적인 정보를 갖고 올바른 답

변을 생성하는 것이 중요하다. 이를 위해 주변 정보를 잘 활용하여 답변을 생성하는 여러 연구가 이루어지고 있고, 그 과정에서 다양한 모델링 기법과 수학적인 계산식들이 파생되었다. 이 중 어떠한 기법과 수식을 사용할 지에 대하여 논리적인 판단이 필요하다.

본 논문에서는 여러가지 적절한 답변을 생성하고 대화의 맥락을 이용하여 답변을 생성하는 데 도움이 되는 잠재 변수 모델링과 잠재 변수에 들어있는 배경지식 정보를 답변 생성 과정에서 활용할 수 있도록 하는 LWE Attention 기법을 활용하여 기존의 연구보다 배경지식을 더 활용한 답변을 얻을 수 있음을 보인다.

하지만 문맥 정보를 활용한 답변 생성을 하였다는 것을 사람이 직접 평가하는 것 이외에 지표로서 판단하기 어렵다고 판단되며, 답변 생성 문제에 대한 적절한 평가 지표가 필요하다고 판단된다.

참고문헌

[1] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension" Meeting of the Association for Computational Linguistics (ACL 2020), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7871—7880, 2020.

[2] Wei Chen, Yeyun Gong, Song Wang, Bolun Yao, Weizhen Qi, Zhongyu Wei, Xiaowu Hu, Bartuer Zhou, Yi Mao, Weizhu Chen, Biao Cheng and Nan Duan "Dialog VED: A Pre-trained Latent Variable Encoder-Decoder Model for Bialog Response Generation" Meeting of the Association for Computational Linguistics (ACL 2022) Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 4852-4864, 2022.

[3] Wen Zheng, Natasa Milic-Frayling, Ke Zhou "Contextual Knowledge Learning For Dialogue Generation" Meeting of the Association for Computational Linguistics (ACL 2023), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 7822-7839, 2023.

[4] Ruijun Chen, Jin Wang, Liang-Chih Yu and Xuejie Zhang "Learning to Memorize Entailment and Discourse Relations for Persona-Consistent Dialogues" Association for the Advancement of Artificial Intelligence (AAAI 2023), The Thirty-Seventh AAAI Conference on Artificial Intelligence, 12653-12661, 2023.

[5] Tiancheng Zhao, Ran Zhao, Maxine Eskenazi, "Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders", Meeting of the Association for Computational Linguistics (ACL 2017), Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 654–664, 2017

[6] Thomas Wolf, Victor Sanh, Julien Chaumond, Clement Delangue "TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents" Conference on Neural Information Processing Systems (NeurIPS 2019), Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2019

[7] Adina Williams, Nikita Nangia, Samuel Bowman, "A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference" Meeting of the Association for Computational Linguistics (ACL 2018), Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1112-1122, 2018.

[8] Ramon Sanabria, Shruti Palaskar and Florian Metze "CMU Sinbad’s Submission for the DSTC7 AVSD Challenge" Association for the Advancement of Artificial Intelligence (AAAI 2019), 2019