페르소나를 활용한 대화 생성 방법 연구: 영화 '해리포터' 시리즈를 중심으로

임다영[a], 손건영[a] and 김미숙[a, *]

[a]세종대학교 소프트웨어융합대학 데이터사이언스학과
05006, 서울 특별시 광진구 능동로 209길
E-mail: flqhs5171@gmail.com, handgunzero2@gmail.com, misuk.kim@sejong.ac.kr

요약문

최근에는 특정 사용자 타겟을 위해 페르소나(Persona)를 활용한 연구가 활발히 이루어지고 있고, 페르소나 텍스트를 이용하면 모델에게 특정 성격 및 특징을 부여할 수 있다. 본 연구에서는, 장편 판타지소설이 원작인 '해리 포터' 영화 스크립트 및 크롤링을 활용하여 single-turn으로 구성된 대화 데이터셋과 페르소나 데이터셋을 구축하였다. 또한 오픈 도메인 대화 시스템의 응답 생성 과정에서, 주인공 캐릭터의 페르소나 텍스트를 활용하였다. 특정 캐릭터에 대한 페르소나 텍스트의 학습 적용 유무에 따른 성능 차이와 모델 학습 시에 페르소나 텍스트와 input 텍스트를 연결하는 과정에서 특수 토큰인 '[SEP]' 사용에 따른 성능 결과를 확인하였다. 성능을 평가하기 위해 정량적 평가와 정성적 평가를 수행하였으며, 실험 결과, 페르소나 텍스트를 사용하였을 때의 성능 향상을 확인할 수 있었다.

주제어:

Persona; Open-domain Dialogue System; Chatbot; Harry Potter Dataset

서론

최근 연구에서는 사용자에게 맞춤이 되는 서비스를 제공하고자 페르소나를 활용한 방법이 많이 제안되고 있다.[8] 사용자를 카테고리로 세분화하여 가상의 유저의 형태를 만드는 과정에서 사용하는 것이 페르소나(Persona)이다. 페르소나는 연극에서 쓰이는 탈(mask; character)에서 유래하였으며, 사회 역할이나 배우에 의해 연기되는 등장인물을 나타낸다. 현재는 가상의 인물을 설정하기 위해 페르소나를 매개체로 사용하기도 한다.[7] 챗봇(Chatbot) 분야에서의 페르소나는 인공지능 모델에 특정한 역할, 성격, 특징, 언어 스타일 등을 부여하는 것을 의미한다. 챗봇과 사용자가 상호작용을 하는 과정에서 페르소나는 특정한 개성을 가지고 응답을 생성하거나 대화 스타일을 조정하는 데 사용된다. 다시 말해, 페르소나는 챗봇을 더 생동감 있고 사용자 친화적으로 만들어서 더욱 편안하고, 자연스러운 대화 경험을 제공할 수 있다.

Open-domain dialogue system은 사람과 자동화된 시스템 간에 자연어를 이용한 대화가 가능하게 하는 인공지능 기반의 시스템[12]이다. 이는 크게 task-oriented dialogue system, open-domain dialogue system 두 가지로 분류된다. 그 중 Open-domain dialogue system은 답변을 데이터베이스에서 선택하는 것과 답변을 생성해내는 것, 이 두 가지 모델로 다시 나누어지게 된다. 전자의 경우 기존 구축된 데이터를 출력하기 때문에 주어진 쿼리에 대해 비슷한 데이터가 있어야 한다는 것과 많은 양의 데이터가 구축이 되어있어야 한다는 한계점이 있다. 후자의 경우, 답변 생성이 장점이자 단점이 되어 기존 데이터에 없더라도 생성할 수 있다는 장점과 동시에 일관성 있는 답변이 어려울 수 있다는 한계점이 존재한다. 따라서 기존 연구에서는 페르소나를 사용함으로써 생성하는 답변의 일관성을 유지하고자 하였다.

본 논문에서는 페르소나를 활용하여 캐릭터 챗봇을 구축하는 연구를 수행하였다. '해리 포터' 영화 시리즈 스크립트에서의 주인공 Harry의 응답으로 데이터셋을 구축하였고, 웹 크롤링을 통해 전체 스크립트 내용에서의 주인공 특징이 잘 표현된 문장을 추출하여 페르소나 데이터로 사용하였다. Dialogue generative pre-trained transformer(DialoGPT)[2] 모델을 구축한 데이터 기반으로 파인튜닝하여 실험을 진행하였고, 페르소나를 사용하면 캐릭터의 특성이 반영되어 모델이 유의미하게 학습되는 것을 확인하였다.

2. 관련 연구

2-1. Persona-based Dialogues

기존 Open-domain dialogue system 분야에서의 생성 모델은 모델이 chit-chat 시스템에서 같은 질문에 매번

* 교신 저자

다른 대답을 하는 것과 같이 일관성을 유지하지 못하는 모습을 보였다. Li et al.[1] 와 Song et al.[6] 은 생성 모델이 일관성을 유지하기 어렵기 때문에 이러한 화자 일관성 문제를 처리하기 위한 페르소나 기반 모델을 제시하였다.

Zhang et al. [4] 은 chit-chat 모델의 일관성 유지의 한계점을 해결하기 위해 대화 상대의 프로필 정보를 예측하는 데 사용할 수 있는 페르소나 데이터셋을 구축하였다. 크라우드소싱을 통해 페르소나를 구축하였고, 이후 수정된 페르소나로 대화를 하도록 하여 페르소나와 대화 데이터가 함께 있는 데이터셋을 완성시켰다.

최근의 연구에서는 부여된 성격의 일관성을 유지하며 답변을 생성하도록 하기 위해 Persona-Chat 데이터를 사용한 방법론과 모델이 제안되고 있다. 그 중 Liu et al.[5] 은 open-domain dialogue system에서 다양한 발화자에 대해 학습하기 때문에 모델이 페르소나를 반영하기 어려우며, 이러한 한계점을 해결하기 위해 사전 정의된 페르소나를 가진 chit-chat system을 만들어 생성된 답변의 일관성을 유지하는 모델을 제안하였다.

3. 데이터셋

3-1. Harry Potter 대화 데이터셋 구축

본 연구에서는 페르소나에 따른 결과를 확인하기 위해 기존 캐릭터의 대화 데이터를 사용하였다. '해리 포터'는 영국의 작가 J. K. 롤링의 판타지 소설 시리즈이다. 소설을 원작으로 영화 역시 시리즈로 개봉이 되어 있으며, 총 8개 시리즈로 구성되어 있고, 메인 주인공의 대사가 많은 비율을 차지한다. 해리 포터의 대화 데이터를 사용하기 위해 시나리오 창작을 위한 리소스를 제공하는 BULLETPROOF SCREENWRITING[2]에서 영화 스크립트를 다운받아 데이터셋으로 구축하였다.

영화 스크립트는 인물의 대사뿐만 아니라 소재, 주제, 장르, 작품의 배경과 촬영 설정 등의 내용이 있어 원문으로는 데이터 사용이 어렵기 때문에 전처리 과정을 거쳤다. 등장인물의 이름은 전부 대문자로 표기되어 있으며, 이름 바로 아래 위치한 텍스트인 대사를 사용하여 데이터로 구축하였다. 구축된 데이터의 예시는 Table 1과 같다.

Table 1 - 대화 데이터 구성 예시

Name	Question	Answer
Harry	Happy Birthday, Harry.	For me?... Really?... He's Mine?...
Harry	Indeed yes, sir. Dobby hid and watched for Harry Potter and sealed the gateway.	You nearly got Ron and me expelled!
Harry	At least I warned you about the dragons!	Hagrid warned me about the dragons!
Harry	The Order seriously doubts it. She's no picnic, but she's no Death Eater either.	But she's evil.

대사 전 인물의 행동을 위해 서술된 부분은 제거하였다. 파일의 상태에 따라 Optical Character Recognition(OCR) 이 적용되지 않는 대사는 직접 엑셀에 작성하였고, 스크립트 특성상 대사 중간에 표시되는 기호나 특정 인물 대사에서의 오탈자 등은 캐릭터의 언어적 특징이므로 추가 전처리(preprocessing) 없이 사용하였다. 주인공을 제외한 모든 인물의 대사에 대해 주인공이 대답한 대화를 'Question-Answer' pair 구조로 설계하였으며, 전체 1,296개의 데이터로 구축하였다.

3-2. 페르소나 데이터셋 구축

대화 데이터뿐만 아니라 텍스트 생성시에 '해리 포터'의 특징 반영도를 높이기 위하여 추가 페르소나 데이터셋을 구축하였다. 전체 시리즈에서 캐릭터를 대표할 수 있는 특징을 추출하기 위해 나무위키[3]와 해리포터 관련 웹 사이트[4]에서 웹 크롤링을 진행하였고, 총 16개의 페르소나를 구축하였다.

Table 2 - Harry Potter의 페르소나 데이터 예시

Index	Persona
Persona 1	I am wearing glasses.
Persona 2	My parents died in Voldemort's attack.
Persona 3	I'm attending Hogwarts.
Persona 4	I have a lightning bolt scar on my forehead.

4. 방법론

4-1. 코사인 유사도

가장 유사도가 높은 페르소나를 선택하기 위해 코사인 유사도(cosine-similarity)를 사용했다. 코사인 유사도는 두 벡터 간의 코사인 각도를 이용하여 두 벡터가 얼마나 유사한지 수치로 나타내는 방법이다. 두 벡터의 방향이 동일해서 사잇각이 0° 일 경우 가장

[2] https://bulletproofscreenwriting.tv/

[3] https://namu.wiki/

[4] https://harrypotter.fandom.com/wiki/Main_Page

큰 값인 1을 갖게 된다. 같은 방법으로 두 벡터의 방향이 정반대인 경우 사잇각이 180°로 가장 작은 값인 -1을 갖게 된다. 두 벡터 A, B에 대한 코사인 유사도는 다음과 같은 식(1)로 표현할 수 있다.

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (1)$$

4-2. GPT-2

Generative Pre-trained Transformer(GPT)-2[3]는 OpenAI에서 발표한 자연어처리를 위한 인공지능 모델로 딥러닝 모델 중 하나인 Transformer 아키텍처를 사용하며, 텍스트 생성 및 이해 작업에 사용된다. GPT-2는 대규모 텍스트 데이터셋을 사전 훈련하여 다양한 언어 특성과 문맥을 학습하여 이를 통해 문장의 구조, 단어 간의 관계, 의미 등을 이해하고, 새로운 텍스트를 생성할 수 있다. GPT-2는 다양한 분야와 주제에 대한 대화, 문장 생성, 요약, 번역 등 다양한 NLP 작업에 적용된다. 비지도 학습 방식으로 사전 훈련되어 사전 정의된 작업이나 레이블이 필요하지 않으며, 대량의 텍스트 데이터만으로 모델을 학습할 수 있어 다양한 언어와 주제에 대해 확장성과 유연성이 제공된다.

4-3. DialoGPT

DialoGPT[2]는 Microsoft에서 개발한 뉴럴 대화 응답 생성 모델이다. open-domain dialogue system에서의 성능 향상을 위해 GPT-2의 아키텍처를 기반으로 설계가 되었으며, 다양한 주제와 도메인에서 자연스러운 대화를 생성하고 이해할 수 있다. DialoGPT 는 대화를 생성하기 위해 이전 대화 내용을 이해하고 기억하는 능력이 있어, 이전 대화의 맥락과 상황을 파악하여 응답을 생성하며, 상대방의 대화 스타일과 언어적 스타일을 모방할 수 있다는 특징이 있다. 또한 문장의 의미를 이해하고 그에 맞는 응답을 생성하는 것을 잘 수행해낸다. 학습을 위해 미국의 소셜 뉴스 집계, 콘텐츠 등급 및 토론 웹사이트인 Reddit에서 2005년부터 2017년까지의 글을 조건에 따라 크롤링하였고, 결과적으로 총 147M의 대화 데이터를 사용해 파인 튜닝 하였다. single-turn 대화 세팅에서 automatic과 human evaluation 모두에서 사람과 비슷한 성능을 냈으며 최고 수준의 성능을 달성하였다.

5. 실험 및 결과

5-1. 실험 셋팅

본 논문에서는 모델 학습 및 평가를 위해 TITAN RTX GPU 1대와 Pytorch 프레임워크를 사용하였다. DialoGPT-medium을 사용하였으며, DialoGPT 모델의 파라미터 수는 345M이고, 디코더 레이어 수는 24, model dimension 은 1024 이다. 전체 데이터(1,269) 중 90%(1,015)는 훈련 데이터로, 10%(254)는 테스트 데이터로 사용하였고, 배치 사이즈는 4, 블록 사이즈는 512, learning rate는 5e-5, weight_decay 는 0.0, warmup_steps 는 0으로 설정하여 실험을 진행하였다.

5-2. 실험 설계

페르소나를 적용 유무에 대한 비교를 위해 페르소나를 사용하지 않고, 등장인물의 대사만 사용한 경우를 베이스라인(baseline)으로 설정하였다. 테스트 시에 페르소나를 적용하는 것의 유의미함을 확인하기 위해 모델 학습 시에만 사용하거나 혹은 학습과 테스트에서 모두 사용한 경우를 모두 실험하였다. 또한 페르소나와 텍스트 사이의 연결에서의 특수 토큰인 '[SEP]' 토큰의 유의미함을 확인하기 위해 모델을 학습할 때, '[SEP]' 토큰을 사용하지 않고, 띄어쓰기로 페르소나 텍스트를 연결한 경우에 대한 실험도 진행하였다. 모델 구조는 그림 1과 같다.

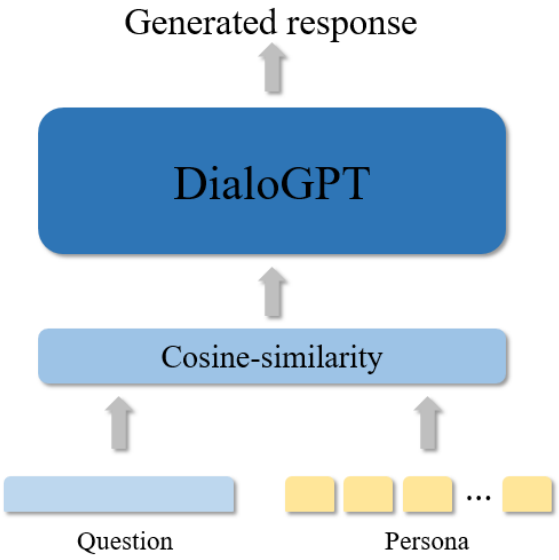

그림 1 – 모델 구조

5-3. 평가지표

생성된 결과가 얼마나 적합한지를 판단하기 위해서 기계 번역 평가 지표 중 하나인 METEOR[9] 를 사용하였다. METEOR는 주어진 번역 문장과 참조 문장(정답 문장)간의 유사성을 측정하여 번역의 품질을 평가하는데 본 연구에서는 정답 문장과 생성된 문장 간의 유사도를 확인하고자 하였다. BERTScore[10] 는 BERT[11] 모델을 기반으로 문장 수준에서 번역 또는 생성된 문장과 정답 문장 간의 유사성을 측정하는 방법으로 사전 훈련된 BERT 모델을 사용하여 두

문장을 벡터로 인코딩한 후 문장 유사성 점수를 계산한다.

5-4. 실험 결과

본 연구에서는 open-domain dialogue system에서 학습된 모델에 추가 학습 및 테스트 시에서의 페르소나 사용 여부에 따른 답변 생성의 성능을 비교 평가하였고, 그 결과 표 3과 같다.

Table 3. 대화 생성 정량 평가 결과

Method	METEOR	BERT score_F1	BERT score_Precision
Baseline	9.31	85.76	86.32
Persona concat (w Test)	11.02	86.15	86.98
Persona concat (w/o Test)	10.66	84.49	83.83
Persona concat (w/o '[SEP]', w Test)	**11.07**	**86.43**	**87.42**

세 평가 지표 모두 페르소나를 사용하지 않은 것과 비교하여 성능이 향상된 것을 확인할 수 있었으며, 페르소나와 대화 텍스트 사이에 특수 토큰인 '[SEP]' 없이 연결했을 때가 가장 효과적임을 확인했다. 페르소나를 모델 학습 시에 같이 적용하게 되면 모델 학습 과정에서만 적용하였을 때와 비교하여, METEOR 성능이 약 1.75% 향상되었고, BERTScore에서의 F1은 약 1.66%, Precision은 약 3.15% 정도 향상되어 모델 테스트시에도 페르소나를 사용하는 것의 유의미함을 확인하였다. 이 결과를 정성적으로 확인하기 위하여 모델의 답변을 표 4에 나타내었다. 정성 평가에서도 페르소나를 사용하지 않은 것과 사용한 방법에서의 차이가 나타난다. 페르소나 없이 실험을 진행하였을 때의 생성된 결과는 정답 문장과는 거리가 있다. 페르소나의 적용 여부에 대한 차이는 페르소나를 학습과 테스트 모두 적용시켰을 때와 Question과 페르소나 연결 시에 '[SEP]' 토큰을 사용하지 않은 방법의 결과가 유사하나, 후자의 경우 실제 답변에 더 가까움을 확인할 수 있었다.

6. 결론

본 연구에서는 '해리 포터' 영화 스크립트를 활용하여 대화 데이터셋을 구축하였고, 추가 크롤링을 통해 페르소나를 설정하였다. 해당 데이터들을 기반으로 Open-domain dialogue system에서 대화 생성 시에 페르소나 사용 유무에 따른 성능 차이를 확인하고자 하였다. 모델 학습 시에서의 페르소나의 적용은 정량, 정성적으로 성능을 높였으며, 대화 텍스트와 페르소나의 연결 방식에서 '[SEP]' 토큰의 사용 유무로 성능의 차이가 있음을 확인하였다. '[SEP]' 토큰으로 구분하지 않고, 띄어쓰기를 사용하는 방법이 정량적으로 가장 높은 성능을 보였으며, 정성적으로도 실제 정답과 가장 유사함을 확인할 수 있었다.

7. 한계점 및 향후 연구

영화 스크립트의 특성상 csv파일로 변환하는 과정에서의 작업에서 한계점이 존재하였다. 또한 총 8편의 영화 시리즈의 스크립트를 사용하였음에도 불구하고 한 캐릭터가 상대방의 대사에 대응하는 답변이 많지 않기 때문에 딥러닝 모델을 파인튜닝 하는 과정에서, 더 나은 결과를 도출하지 못했다는 한계점이 있다.
향후 연구에서는, 데이터 증강 기법을 활용하여 전체 데이터의 크기를 키워 실험을 진행하고자 한다. 또한, 코사인 유사도와 같은 기본적인 알고리즘이 아닌, 최신의 알고리즘들을 적용하여, 페르소나들을 적용하는 실험들을 진행할 예정이다.

Acknowledgments

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program-Knowledge service industry technology development project - service core technology development) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) (Project Number: 20018758).

Table 4. 대화 생성 정성 평가 결과

Ground Truth	페르소나 사용하지 않음 (baseline)	페르소나 학습 시에만 사용	페르소나 학습, 테스트 모두 사용	페르소나 연결 시 '[SEP]' 토큰 사용하지 않음
Blimey, Hermione!	I don't think you can!	I'm not your friend, Potter! [SEP] I am wearing glasses.	Hermione!	Hermione!
No, sir.	I'm not sure.	Np, I didn't think I would be able to put my name in…	No.	No, sir.
A message…?	You don't understand, do you?	I understand. But I have to ask, is there anything you'd like to tell me?	What message?	And that message is?
Yes, Aunt Petunia.	You mean you don't want to eat it?	I'll be careful. Don't want to get burned. But if you do, I'm sure you're going to want to share it with me.	No.	I'll try.

References

[1] Jiwei Li, Michel Galley, Chris Brockett, Georgios P. Spithourakis, Jianfeng Gao and Bill Dolan (2016) *A Persona-Based Neural Conversation Model*

[2] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu and Bill Dolan (2020*) DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation*

[3] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei,and Ilya Sutskever (2018) *Language Models are Unsupervised Multitask Learners*

[4] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela and Jason Weston (2018) *Personalizing Dialogue Agents: I have a dog, do you have pets too?*

[5] Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, Dongmei Zhang (2020) *You Impress Me: Dialogue Generation via Mutual Persona Perception*

[6] Haoyu Song, Yan Wang, Kaiyan Zhang, Wei-Nan Zhang and Ting Liu (2021) *BoB: BERT Over BERT for Training Persona-based Dialogue Models from Limited Personalized Data*

[7] Seungju Han, Beomsu Kim, Jin Yong Yoo, Seokjun Seo, Sangbum Kim, Enkhbayar Erdenee, Buru Chang (2022) *Meet Your Favorite Character: Open-domain Chatbot Mimicking Fictional Characters with only a Few Utterances*

[8] Ruijun Chen, Jin Wang, Liang-Chih Yu and Xuejie Zhang (2023) *Learning to Memorize Entailment and Discourse Relations for Persona-Consistent Dialogues*

[9] Satanjeev Banerjee and Alon Lavie (2005) *METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments*. In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics

[10] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger and Yoav Artzi (2020) *BERTScore: Evaluating Text Generation with BERT*

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018) *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*

[12] Haodong Yang, Wenge Rong and Zhang Xiong (2019) *Open-Domain Dialogue Generation: Presence, Limitation and Future Directions*