

감정 챗봇을 위한 새로운 접근: BERT 와 GPT 의 파인튜닝 및 파이프라이닝을 이용한 감성 대화 생성

김형준 박정민 유재형 이성주
경희대학교 컴퓨터공학과

gudwns4985@khu.ac.kr jungmin59@khu.ac.kr dbwogud227@khu.ac.kr juscomputer@khu.ac.kr

New Approach for Emotion Chatbot: Emotion-aware Conversation Generation using Fine-tuning and Pipelining of BERT and GPT

Hyungjun Kim Jungmin Park Jaehyeong Ruy Seongju Lee
School of Computer Science and Engineering, Kyung Hee University

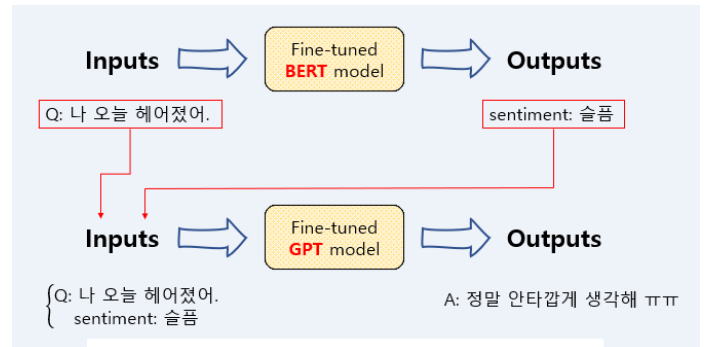
요 약

챗봇 서비스는 사용자의 입력 문장에 대해 적절한 응답을 지속적으로 반환하는 서비스로 GPT 등의 생성모델을 사용하여 문장을 생성한다. 현재의 챗봇 서비스는 사용자 입력 문장의 감정을 파악하고 그에 맞춰 응답 문장에 감정을 표현하지 못한다는 문제점이 있다. 이에 본 논문에서는 감정이 담긴 답변을 반환하는 BERT-GPT 파인튜닝 & 파이프라이닝 모델을 제안한다. BERT 모델을 <입력 문장-출력>으로 이루어진 학습데이터로 파인튜닝하여 사용자의 입력 문장으로부터 감정을 분석하고, GPT 모델을 <입력 문장+감정-출력 문장>으로 이루어진 학습 데이터로 파인튜닝하여 감정이 담긴 문장을 생성할 수 있게 한다. 최종적으로 사용자의 입력 문장에 대해 BERT 모델을 통해 감정을 추출하고, 입력문장과 감정을 레이블링한 결과를 GPT 모델의 입력으로 활용하여 감정이 담긴 답변을 출력값으로써 획득한다.

1. 서 론

챗봇 서비스는 사용자의 입력 문장에 대해 대화의 흐름에 따라 알맞은 답변을 반환하는 기능을 가진다. 현대의 챗봇 서비스는 사용자의 응답에 단순히 공감하는 응답 생성에서 사람의 감정을 비슷하게 표현할 수 있는 수준이다. 하지만 사용자의 감정을 부정확하게 해석하여 만들어낸 응답은 챗봇 서비스의 사용자 경험을 저하시키고 서비스의 효과성을 떨어뜨린다. 또한 사용자의 감정을 파악하고 그에 해당하는 감정을 담지 않은 응답은 사용자의 서비스 이용 간에 부자연스러움을 유발한다..

따라서 본 논문에서는 BERT-GPT 파인튜닝 & 파이프라이닝을 통한 향상된 감정 챗봇 서비스를 제안한다. 사람이 느낄 수 있는 감정을 세밀하게 분류한 58 개의 감정 소분류 데이터를 활용하여 BERT 모델을 파인튜닝함으로써 더욱 세분화된 사용자 입력 문장의 감정에 대응하며, 감정과 함께 레이블링된 <입력+감정-출력>형태의 데이터를 GPT 모델 파인튜닝에 활용함으로써 GPT 모델이 생성하는 응답에 입력 문장과 동일한 감정을 반영할 수 있다. 본 논문에서 제안하는 서비스의 파이프라이닝 과정은 [그림 1]과 같다.



[그림 1]

2. 관련 연구

2.1 BERT

BERT(Bidirectional Encoder Representations from Transformers)는 2018 년에 구글에서 개발해서 공개한 사전 학습 언어 모델로, 트랜스포머 구조의 인코더 모델을 기반으로 구현되었으며, 위키피디아 (25 억 단어)와 BooksCorpus(8 억 단어)와 같은 레이블이 없는 텍스트 데이터로 사전학습 되었다. 기존의 단방향 언어 모델과 달리, BERT 는 양방향으로 문맥을 고려하여 문맥에 따른 텍스트의 표현을 학습한다. 이를 통해 BERT 는 맥락에 민감한 단어 임

베딩을 생성하여 다양한 자연어 처리 태스크에서 우수한 성능을 발휘할 수 있다. 이러한 사전 학습 모델은 특정한 태스크를 수행하기 위한 출력을 제공하지 않기 때문에, 작은 규모의 데이터를 사용하여 모델을 조정하는 파인튜닝 과정이 필요하다. BERT 모델을 파인튜닝하면, 질문 응답, 감정 분석, 개체명 인식 등 다양한 자연어 처리 태스크에 맞는 모델을 생성할 수 있다.[5]

2.2 GPT

GPT(Generative Pre-trained Transformer)는 OpenAI 에서 개발한 자연어 처리 모델로, 트랜스포머 구조를 기반으로 한다[1]. GPT 는 대량의 텍스트 데이터를 사전학습시켜, 문맥과 의미를 이해해야 하는 다양한 자연어 처리 작업에 사용될 수 있는 다목적 언어 모델이다. GPT 의 파인튜닝은 사전 학습된 GPT 모델을 특정한 자연어 처리 작업에 맞게 추가적인 학습을 진행하는 과정을 말한다. 파인튜닝은 GPT 모델을 더 특화된 작업에 적용하기 위해 필요하며, 작은 규모의 태스크 관련 데이터를 사용하여, GPT 모델을 조정하고 해당 작업에 최적화된 모델을 생성한다. GPT 모델은 파인튜닝을 통해 특정 작업에 대한 성능을 향상시키고, 더 나은 자연어 처리 결과를 얻을 수 있다[2].

3. 문제 정의

본 연구에서 해결하고자 하는 문제는 챗봇이 감정을 담긴 응답을 사용자에게 반환하고 있지 않다는 점이다. 감정이 담긴 챗봇의 응답 반환이라는 목표 관점에서 기존 챗봇에서는 사용자의 입력에 대해 감정이 담긴 응답을 생성해주기 보다는 상황에 알맞은 최적의 답변 및 공감형 답변을 제공하는 한계가 존재한다. 해당 공감형 답변은 감정이 정확하게 담기지 않은 상황에 맞게 충분히 내놓을 수 있는 답변이다. 이러한 문제 상황을 해결하기 위하여 사용자의 입력 문장에 따라 세분화된 감정으로 해당 문장을 분류하여 출력으로 충분한 감정이 담긴 문장을 내놓도록 하는 챗봇 모델의 설계가 필요하다.

4. 해결 방안

4.1 감정 소분류를 활용한 BERT 모델 파인튜닝

BERT 모델 파인튜닝 단계에서는 58 개의 감정 소분류(감사하는, 걱정스러운, 고립된, 기쁨, 노여워하는 등)를 통해 BERT 모델을 파인튜닝하여 세분화된 감정을 분류하는 BERT 모델을 생성한다.[그림 2]는 BERT 모델을 파인튜닝하기 위해 사용된 데이터이다. 파인튜닝에 활용된 데이터의 형식은 <문장-감정 소분류>로 구성되어 있으며 중복값과 결측값

을 제거한 58 개의 레이블링된 감정과 144723 개의 데이터를 활용하였다. (출처:Aihub “한국어 감정 정보가 포함된 연속적 대화 데이터셋”)

	A	B	C	D	E	F	G	H
1		연령	성별	상황키워드	신체질환	감정_대분류	감정_소분류	사람문장1
2	1	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	얼른 왜 해도 해도 끝이 없을까? 왜가 난다.
3	2	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	이번 달에 또 급료가 왔잖아! 물가는 오르는데 월급
4	3	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	회사에 신입이 들어왔는데 밑둥가 거슬러. 그런 애
5	4	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	직장에서 막내라는 이유로 나에게만 온갖 심부름을
6	5	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	얼마 전 입사한 신입사원이 나를 무시하는 것 같아
7	6	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	직장에 다니고 있지만 시간만 버리는 거 같아. 진짜
8	7	청년	여성	진로,취업,해당없음	분노	노여워하는	노여워하는	성인인데도 진료를 아직도 못 정했다고 부모님이!
9	8	청년	여성	진로,취업,해당없음	기쁨	느긋	느긋	회사한 지 얼마 안 됐지만 천천히 직장생활 구해보려
10	9	청년	여성	진로,취업,해당없음	불안	걱정스러운	걱정스러운	졸업반이어서 취업을 생각해야 하는데 지금 너무!
11	10	청년	여성	진로,취업,해당없음	기쁨	느긋	느긋	요즘 직장생활이 너무 편하고 좋은 것 같아!
12	11	청년	여성	진로,취업,해당없음	기쁨	느긋	느긋	취업해야 할 나이인데 취업하고 싶지가 않아.
13	12	청년	여성	진로,취업,해당없음	불안	당혹스러운	당혹스러운	면접에서 부모님 직업에 대한 질문이 들어왔어.
14	13	청년	여성	진로,취업,해당없음	불안	당혹스러운	당혹스러운	큰일이야. 부장님께 결재받아야 하는 서류가 사라졌
15	14	청년	여성	진로,취업,해당없음	불안	당혹스러운	당혹스러운	나 얼마 전에 면접 본 회사에서 면접 합격했다고 인
16	15	청년	여성	진로,취업,해당없음	당황	당황	당황	길을 가다가 우연히 마주친 동네 아주머니께서 취

```
print(f'총 데이터 개수: {len(dataset)}')
sub_sentiment_cnt = dataset['sub_sentiment'].nunique()
print(f'감정 라벨 개수: {sub_sentiment_cnt}')
```

총 데이터 개수: 144723

감정 라벨 개수: 58

[그림 2]

해당 데이터를 통해 파인튜닝된 BERT 모델은 GPT 모델 파인튜닝 과정에서 학습 데이터를 레이블링 하는 경우와 최종 챗봇 모델 구현 완료 시 사용자의 입력 문장에 대한 감정을 분석하여 GPT 모델로 입력 문장과 함께 전달할 때 활용된다.

4.2 감정 레이블을 활용한 GPT 모델 파인튜닝

GPT 모델 파인튜닝 단계에서는 파인튜닝된 BERT 모델을 활용하여 입력-출력 대화 쌍으로 이루어진 학습 데이터의 감정 레이블을 추출하여 입력과 함께 레이블링한 결과를 파인튜닝 과정에 활용한다. [그림 3]은 GPT 모델을 파인튜닝하기 위해 사용된 데이터이다. 파인튜닝 과정에서 활용되는

Q,A,label
12시 됐!,하루가 또 가네요.,0
1지말 학교 떨어졌어,위로해 드립니다.,0
3박4일 놀러가고 싶다,여행은 언제나 좋죠.,0
3박4일 절도 놀러가고 싶다,여행은 언제나 좋죠.,0
PPL 심하네,눈살이 찌푸려지죠.,0
SD카드 망가졌어,다시 새로 사는 게 마음 편해요.,0
SD카드 안돼,다시 새로 사는 게 마음 편해요.,0
SNS 맞팔 왜 안하지ㅠㅠ,잘 모르고 있을 수도 있어요.,0
SNS 시간날비인 거 아는데 매일 하는 중,시간을 절하고 해보세요.,0
SNS 시간날비인데 자주 보게됨,시간을 절하고 해보세요.,0
SNS보면 나만 빠고 다 행복해보여,자랑하는 자리니까요.,0
가끔 궁금해,그 사람도 그럴 거예요.,0
가끔 뭐하는지 궁금해,그 사람도 그럴 거예요.,0
가끔은 혼자있게 좋다,혼자를 즐기세요.,0
가난한 자의 설움,돈은 다시 들어올 거예요.,0
가만 있어도 땀난다,땀을 식혀주세요.,0
가상화폐 폭락 망할,어서 잊고 새출발 하세요
가스를 켜고 나갔어,빨리 집에 들어가서 끄고
가스를 켜놓고 나온거 같아,빨리 집에 들어가
가스비 너무 많이 나왔다.,다음 달에는 더 절
가스비 비한데 갑자기 걸리겠어,따뜻하게 사세요
가스비 잘난 아님,다음 달에는 더 절약해보

len(dataset)

11823

[그림 3]

데이터의 형식은 <Q+label-A>으로 구성되어 있으며 중복값과 결측값을 제거한 11823 개의 데이터를 활용하였다.

(출처: https://github.com/songys/Chatbot_data)

해당 데이터를 통해 파인튜닝된 GPT 모델은 최종 답변 생성 과정에서 사용자의 입력 문장과 BERT 모델을 통해 추출한 감정을 함께 받아서 출력 문장을 생성한다.

Word Masking The Journal of Korean Institute of Information Technology, 19(9), 29-40

[4] Won-Min Lee, Byung-Won On. (2021). Emotion sentence generation system for artificial intelligence chatbot. Proceedings of KIIT Conference, (), 461-464.

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

4.3 BERT-GPT 파이프라이닝 모델 설계

감정 소분류를 통해 파인튜닝된 BERT 모델과 감정 레이블을 통해 파인튜닝된 GPT 모델을 활용하여 감정이 담긴 응답을 반환하는 챗봇 서비스 모델을 구성한다.[그림 1]의 서비스 파이프라이닝 모델의 형태로 사용자의 입력 문장이 파인튜닝된 BERT 모델을 거쳐서 감정이 추출되고, 해당 감정을 사용자의 입력 문장과 함께 레이블링하여 GPT 모델의 입력으로 사용하고 최종적으로 감정이 담긴 출력 문장을 획득한다.

5. 결 론

본 논문에서 제시한 BERT-GPT 파인튜닝 & 파이프라이닝 방법은 사용자의 입력 문장으로부터 감정을 추출하여 해당 감정과 원래 입력 문장을 함께 레이블링하고 그것을 최종 문장 생성에 이용하는 방식이다.

58 개의 감정 소분류(감사하는, 걱정스러운, 고립된, 기쁨, 노여워하는 등)를 BERT 모델의 파인튜닝에 활용함으로써 사용자의 입력 문장으로부터 더욱 세분화된 감정을 추출한다. BERT 모델이 감정 분류를 더욱 구체적으로 수행함에 따라 GPT 모델 또한 더욱 세분화된 감정 분류를 가진 데이터를 기반으로 파인튜닝된다. 따라서 GPT 의 문장 생성 부분에서도 감정이 더 잘 반영된 문장을 생성된다. 따라서 사용자의 입력에 따른 감정을 분석하여 그에 맞는 감정이 담긴 응답을 반환하는 향상된 감정 챗봇 설계라는 목적을 달성할 수 있다.

참 고 문 헌

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention is all you need. In Advances in neural information processing systems, pp.5998-6008, 2017.

[2] Radford, A..et al. "Improving language understanding by generative pre-training." OpenAI, 2018.

[3] Won-Min Lee, Byung-Won On. (2021). Generating Emotional Sentences Through Sentiment and Emotion