언어모델 기반 오픈 도메인 챗봇 구현

김승태*, 구자환*, 김응모*
*성균관대학교 소프트웨어대학

tae1303@skku.edu, jhkoo@skku.edu, ukim@skku.edu

Building Open Domain Chatbot based Language Model

Seung-Tae Kim* Jahwan Koo*, Ung-Mo Kim*
*College of Software, Sungkyunkwan University

요 약

자연어 처리는 인공지능의 핵심기술 중 하나이다. 그 중 오픈 도메인 챗봇 구현은 NLP 에서 어려운 태스크로 꼽힌다. 명확한 목표, FAQ 가 존재하는 기능형 챗봇과 달리 오픈 도메인 챗봇은 연속적 대화, 방대한 양의 상식 등 구현에 어려움이 많았다. 짧은 질문과 대답으로 이루어진 데이터로 학습한 모델을 대화 데이터로 학습시켜 좀더 자연스러운 챗봇을 구현해보고자 한다.

1. 서론

자연어 처리(NLP: Natural Language Processing)는 전 산업 부문에 걸쳐 혁신을 불러 일으키고 있는 기술이다. 은행, 보험 등 여러 분야에서 프로세스 자동화로 비용 절감 및 위험 최소화를 통한 생산성 향상 및 이익 증대 등 전 분야에서 성장하고 있다. 그 중 챗봇은 즉각적인 반응, 시간에 구애받지 않는 점 등 많은 장점으로 인해 대부분의 기업에서 도입을 고려하고 있는 AI 기반 컴퓨터 프로그램이다. 과거에는 간단한 문의에 정해진 답변을 출력하는 수준이었다면, 최근에는 AI, 딥러닝 등을 통해 콜센터의 기능을 일부 대신하거나, 정보를 수집하고 처리해 맞춤 서비스를 제공하는 등 심화된 수준까지 발전했다.

오픈 도메인 챗봇은 더 발전된 형태로, 특정 기업, 특정 도메인에서만 활용 가능하던 챗봇이 아닌 일반 상식을 가지고 실제 사람처럼 대화가 가능한 챗봇이다. 최근 방대한 양의 학습 데이터, 거대한 모델 사이즈를 기반으로 한 높은 성능의 오픈 도메인 챗봇이 발표되고 있는데, 한국어 챗봇은 부족한 학습 데이터로 인해 모델을 충분히 활용하지 못하고 있다. 그래서 본 논문에서는 학습 데이터 수정 및 미세 조정(fine tuning)이 성능에 미치는 영향을 제시하고자 한다. 질문-대답 쌍으로만 학습된 GPT-2 기반 한국어 챗봇을 대화 데이터로 미세 조정해 보다 자연스러운 한국어 챗봇을 구현해보고자 한다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련 연구로 여러가지 언어 모델의 구조를 기술하고 3 장에서는 GPT-2 의 구조 및 데이터 학습 방법에 대해, 4 장에서는 결론을 기술한다.

2. 관련 연구

본 장에서는 챗봇의 분류, 챗봇의 기반이 되는 언어 모델(Language Model)과 국내외 챗봇 서비스의 현황에 대해 살펴보도록 한다.

2.1 챗봇

챗봇은 용도가 특정 주제에 한정되어있는지 여부에 따라 open-domain 챗봇과 closed domain 챗봇으로 나눌 수 있다.

Closed-domain 챗봇은 의도된 특정영역에 대한 질문에만 응답하는 챗봇이다. 상품 예약 서비스, 은행 서비스 등 용도에 따라 도메인이 한정되어있고, 사용자도 그 사실을 인지하고 있기 때문에 도메인을 벗어난 입력은 거의 없다.

Open-domain 챗봇은 특정 도메인에 한정되지 않고 사람과 대화하는 것처럼 유저와 소통을 할 수 있는 챗봇이다.

2.2 언어 모델(language model)

언어 모델이란 문장의 각 단어에 확률을 할당하는 모델으로, 가장 자연스러운, 가장 확률이 높은 단어들, 즉 가장 확률이 높은 단어 시퀀스를 찾아내는 모델이다. 단어 시퀀스의 확률을 $P(W)$라고 할 때, $P(W)$는 다음과 같이 나타낼 수 있다.

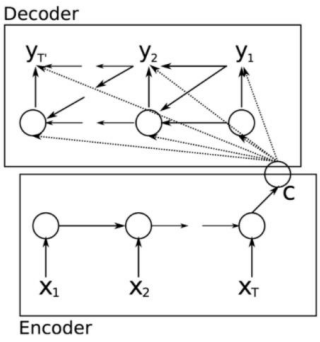
$$P(W) = P(w_1, w_2, \cdots, w_n)$$

보편적으로는 확률을 할당하기 위해 이전 단어들로부터 확률이 가장 높은 단어를 이어 붙이는 방식을 택한다. 단어 시퀀스 확률을 P(W)라고 할 때, P(W)는

$$P(W) = \prod_{i=1}^{n}(w_1, w_2 \cdots, w_{n-1})$$

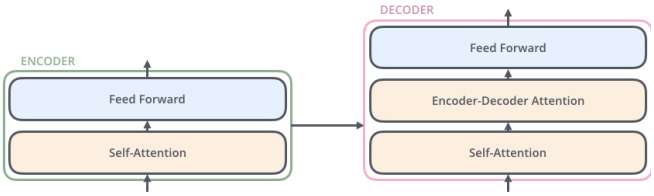
으로 나타낼 수 있다. [1].

언어 모델은 encoder 와 decoder 로 구성된다. encoder 와 decoder 는 RNN 으로 구축하거나 어텐션(Attention)으로 구축할 수 있다. RNN 으로 구축한 encoder 는 입력 시퀀스를 단어 단위로 쪼개 고정 길이의 벡터로 매핑하고, decoder 는 벡터를 다시 단어를 decoding 해 하나의 문장을 생성한다. encoder 에서 입력 시퀀스를 벡터로 압축하는 과정에서 정보가 손실되는 단점을 해결하기 위해 어텐션을 사용한다.



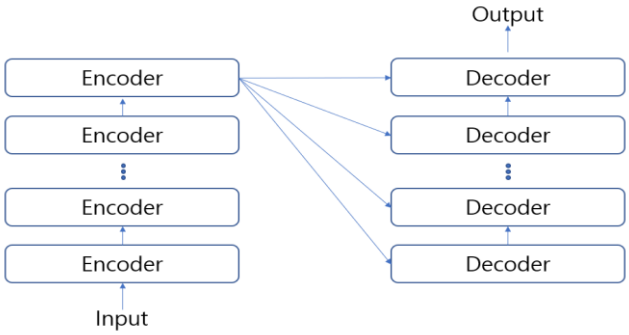
(그림 1) RNN 기반 encoder, decoder

어텐션을 RNN 의 보정 용도가 아닌 어텐션으로 encoder 와 decoder 를 구축할 수 있다. 어텐션을 이용해 구축한 encoder 와 decoder 에는 대표적으로 transformer 가 있다.



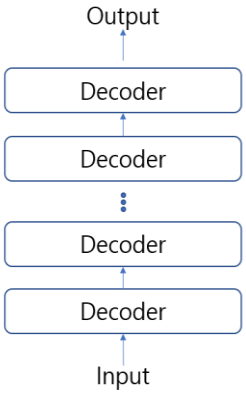
(그림 2) Transformer encoder, decoder

언어 모델은 크게 encoder-decoder 구조, decoder only 구조로 구분할 수 있다. Encoder 와 decoder 를 여러 층 쌓아 문장을 생성하는 방식을 Encoder-decoder 구조라고 한다. 대표적으로 transformer, meena 가 있다.



(그림 3) Encoder-decoder 구조

Decoder only 구조는 encoder 없이 문장이 바로 decoder 의 입력값으로 주어진다. 대표적으로 GPT 모델이 있다.



(그림 4) Decoder only 구조

2.3 챗봇 서비스 동향

챗봇은 단순한 질의 응답 자동화 프로그램이 아닌 금융권, 교육 등 전 분야에서 사용할 수 있는 기술이다. 온라인 서비스의 비중이 급격히 증가하고 있는 상황에서 챗봇이 어떻게 사업을 변화시키고 있는지 사례를 통해 알아보고자 한다.

2.3.1 우리은행

우리은행이 선보인 '위비봇'은 인공지능을 활용한 실시간 상담 서비스이다. 위비봇은 대화를 통해 예금 계좌조회, 상품 추천, 환율 안내 등 기존 은행원의 업무를 대신하고 있다. [2].

2.3.2 Meena

구글은 2020 년 오픈 도메인 챗봇인 Meena 를 공개했다. 단순하지만 거대한 모델을 학습시켜 거의 사람에 가까운 대화 능력을 선보였다. 복잡한 구조 설계 없이 명확한 한계가 있던 End-to-end 모델을 사용해 최고의 성능을 기록했다. Meena 는 26 억개의 parameter 를 가지고 있고, encoder-decoder 구조로 1 개의 Transformer encoder 와 13 개의 Transformer decoder 를 쌓아놓은 구조로 이루어져있다.

3. 모델 구축

3.1 GPT-2

GPT 는 Generative Pre-Training 의 약자로, 모델이 입력값으로 다음에 올 단어 또는 문장을 생성하는 모델이라는 뜻이다. Pre-training 은 GPT 모델을 만든 OpenAI 에서 모델을 사전학습시킨 후 각자의 목적에 맞게 미세 조정해서 사용할 수 있게 했다는 뜻이다.

GPT 는 encoder-decoder 구조의 seq2seq 모델인 Transformer 의 decoder 블록을 여러 개 쌓아 만든 구조이다. 입력으로 문장을 받아 다음에 올 단어를 예측한 후, 그 토큰을 포함해 다시 입력하는 방식으로 결과를 출력한다. [3].

3.2 KoGPT-2

KoGPT-2 는 SKT 에서 공개한 모델로, GPT-2 SMALL 모델을 약 20GB 의 한글 데이터로 학습시킨 모델이다. 데이터 가공 및 미세 조정이 성능에 긍정적인 영향을 미칠 수 있는 지를 알아보기 위해 기존에 존재하던 한국어 챗봇 모델을 이용한다.

3.3 학습 데이터

KoGPT-2 모델을 질문-대답-감정상태레이블으로 이루어진 단순한 데이터형식을 가지고 학습시킨다. 감정상태레이블은 0(중립), 1(긍정), 2(부정)으로 구분된다.

챗봇 구현을 위해 대화 데이터로 모델을 미세 조정한다. 데이터는 AI HUB 의 '한국어 감정 정보가 포함된 단발성 대화 데이터셋'을 사용하였다. [4]. 이 데이터는 약 3 만개의 문장으로 이루어진 대화 데이터로, 기쁨, 분노, 혐오, 중립, 놀람, 공포, 슬픔의 7 가지로 감정상태를 구분한다. 이를 중립, 긍정, 부정으로 분류해 학습시킨다. 대화는 질문-대답이 무조건 반복되는 패턴이 아니기 때문에 기존 데이터 형식에 바로 적용하기에는 무리가 있다. n 개의 문장으로 된 대화에서 첫 번째 문장부터 i-1 번째 문장을 질문으로, i 번째 문장을 대답으로 하는 데이터를 생성했다. 이를 통해 챗봇이 기존 챗봇보다 맥락있는 대화를 진행할 수 있는지 확인해보고자 한다.

있어?	아니.	0
있어?아니.	있을턱이 없지.	0
있어?아니.있을턱이 없지.	보증금 없고, 월세 싸고. 중고생 둘이 살만한...그런 곳.	0
있어?아니.있을턱이 없지. 보증금 없고, 월세 싸고. 중고생 둘이 살만한...그런 곳.	불가능이야.	0
있어?아니.있을턱이 없지. 보증금 없고, 월세 싸고. 중고생 둘이 살만한...그런 곳.불가능이야	그냥 우리집에 와서 나랑 같이 빙 쓰면 안돼?	0
있어?아니.있을턱이 없지. 보증금 없고, 월세 싸고. 중고생 둘이 살만한...그런 곳.불가능이야	우리 강산인 어쩌구?	0
있어?아니.있을턱이 없지. 보증금 없고, 월세 싸고. 중고생 둘이 살만한...그런 곳.불가능이야	참!! 금강산이 있었지. 어후- 답이 없네.	0

(그림 5) 학습 데이터

3.4 실험 결과 및 평가

Input	기존 모델의 output	추가학습한 모델의 output
가족이랑 싸웠어.	싸우면서 이겨보세요.	많이 힘들었을 거예요.
가슴이 답답해	답답하면 쉬어도 돼요.	가슴이 답답하면 약을 먹어보세요.
나이가 어떻게 돼?	나이가 중요하지 않아요.	서른셋.
오늘 너무 피곤하다	푹 쉬세요.	내일은 오늘보다 더 나은 내일을 위해 일찍 주무세요.
요즘 스트레스가 심해	스트레칭을 해보세요.	스트레스를 풀 수 있는 방법을 찾아보세요.
시험에 떨어졌어	열심히 공부해서 좋은 결과 있을 거예요.	좌절하지 마세요.

(그림 6) 실험 결과

기존 챗봇은 학습이 부족해 출력 문장의 자연스러움과 구체성 모두 부족한 경우가 대부분이었다. 입력 문장과 관계없는 문장을 출력하는 경우가 많고, 문장 전체가 아닌 단어 한 가지에 관련된 문장을 자주 출력했다. 또 자연스러운 대답이 출력된 경우에도 입력 문장을 그대로 반복하거나 대부분의 입력에 대응할 수 있는 '좋은 생각이에요'와 같은 챗봇의 역할을 잘 하지 못하는 결과가 많았다.

추가학습한 챗봇의 출력 문장은 기존 챗봇에 비해 더 자연스럽고 구체적인 응답을 출력했다. 의미없는 문장의 반복이 별로 없고 자연스러운 응답이 많아졌다.

4. 결론

유저의 질문에 학습된대로, 또는 추상적인 대답을 했던 기존의 챗봇에서 완벽하지는 않지만 보다 구체적인 수준의 대답을 내놓는 정도로 개선된 것처럼 보인다. 하지만 유저의 입력은 학습 데이터처럼 중첩되어 입력되지 않기 때문에 다른 seq2seq 모델이나 거대언어 모델처럼 뛰어난 성능을 보이지는 못했다. 단순히 학습 데이터의 양의 증가만으로 눈에 띄게 성능이 증가한 것을 알 수 있다.

전세계에서는 NLP 의 중요성을 깨닫고 예전부터 국가차원에서 NLP 학습 데이터를 구축하기 시작했다. 그 결과 미국, 중국 등 AI 분야에 강세를 보이는 국가에서는 엄청난 양의 데이터를 축적했지만, 우리나라는 그에 비하면 1% 정도의 데이터만을 가지고 있다. 챗봇을 위한 충분한 한국어 데이터를 확보한다면 우리나라도 충분히 사람 수준의 오픈 도메인 챗봇을 선보일 수 있다는 증거일 것이다.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2018R1D1A1B07049464).

참고문헌

[1] 지인영, 김희동. 사전학습 언어모델의 기술 분석 연구. 국제언어인문학회, 2020

[2] 서교리. 인공지능 기반 챗봇 서비스의 국내외 동향분석 및 발전 전망. 한국정보화진흥원, 2018

[3] Alec Radford, Jeffrey Wu, Rewon Child, David Lua, Dario Amodei, Ilya Sutskever. Language Models are Unsupervised Multitask Learners. OpenAI, 2018

[4] AI HUB Open 데이터, https://aihub.co.kr/keti_data_board/language_intelligence