

# BERT 모델을 활용한 고객 질의 응답 챗봇 구현 Implementing a Customer O&A Chatbot Using the BERT Model

사수진\*

\* 소속: 충북대학교, 산업인공지능학과 e-mail: sasuiin@gmail.com

### 요약

본 논문은 BERT(Bidirectional Encoder Representations from Transformers) 모델을 활용하여 고객 질의응답 챗봇을 개발하는 과정을 설명한다. 본 연구의 주요 목적은 고객 지원 팀의 업무 부담을 경감하고, 24시간 연중무휴로 지원을 제공하며, 일관되고 정확한 응답을 보장하는 데 있다. BERT 모델을 파인튜닝(Fine-tuning)하여 고객의 문의를 효과적으로 이해하도록 함으로써, 챗봇은 신속하고 개인화된 응답을 제공하여 사용자 만족도를 높인 다. 또한, 한국어 문의 처리를 위한 KoBERT와 같은 특화된 모델이 고객 서비스에서 얼마나 효과적인지를 강조한다.

# 서론 & 시스템 구현 방법

#### 1. 서론

고객 문의의 증가로 인해 효율적인 고객 지원이 조직에 점점 더 어려운 과제가 되고 있습니다. 최근 1년간 고객 문의 건수 가 30% 증가하였으며, 이로 인해 평균 응답 시간이 20% 늘어 났습니다. 이러한 증가는 기존 고객 서비스 지원팀의 업무 부 안이 시급합니다. 대규모 고객 지원 팀을 유지하는 것은 비용 이 많이 들며, 사람에 의한 응답은 종종 일관성이 부족합니다 자연어 처리(NLP)의 발전을 활용한 챗봇은 고객 지원 인력의 업무 부담을 줄이고, 효율성을 높이며, 응답 시간을 개선할 수 있는 잠재적인 해결책을 제공한다.

본 논문은 BERT 모델을 사용하여 제품 매뉴얼 내용을 활용한 다 보고 있다. 그들을 되는 사용자 집의응답을 자동화하는 첫봇을 개발하는 것을 목표로 합니다. 기존의 고객 서비스 지원은 주로 고객의 요구에 맞춘 인력 배치와 비대면 상담에 의존해 왔습니다. 그러나 디지털 전환과 비대면 서비스의 수요가 증가함에 따라, 이러한 전통적 인 방법은 한계에 부딪히고 있습니다. 이에 따라, 더 나은 고객 경험을 제공하기 위해 AI 기반 챗봇의 도입이 점차 중요해지고 있습니다. 본 연구는 BERT 모델을 활용한 챗봇 개발이 이러한 문제 해결에 중요한 역할을 할 수 있음을 시사한다.

#### 2. 데이터셋 구축 및 모델 학습 평가

•질의응답 학습에 사용된 데이터셋은 제품 매뉴얼에서 추 출한 질문과 답변으로 구성된다. 총 1,200개의 Q&A 쌍 을 JSON 형식으로 생성하였으며, 이 데이터셋은 학습용 과 검증용으로 8:2 비율로 분할하여 모델 학습에 사용하 였다.

"context": "문서나 글의 일부, 질문의 배경이 되는 내용.".

"question": "질문 내용",

"id": "고유한 id값",

"is\_impossible": "문맥 내에서 질문에 대한 답변이 존재 유무(true/false)",

"answers": {

"text": ["답변 내용"],

"answer\_start": [문서 내에서 답변이 시작되는 위치 (index)]

그림 1 데이터셋 포맷 예시(KorQuad 한국어 질의응답 데이터셋)

#### •데이터 분석 및 전처리



그림. 2 데이터셋 지문, 질문, 답변 길이 분석

# 실험 및 분석 결과

•모델 학습 곡선 및 성능 평과 비교 • BERT 다국어 모델

그림, 3 모델별 학습 곡선

BERT 다국어 모델

그림, 4 모델별 평가 결과 비교

# 3. 모델별 추론 결과 비교

output\_1\_loss
output\_2\_loss

본 연구에서는 다국어 BERT 모델은 종종 [UNK] (unknown token)을 출력하는데, 이로 인해 다국어 BERT는 한국어 처리에 한계를 보였다. 반면, KoBERT는 한국어 문법과 도메인 특화 어휘를 더 잘 이해하여 고객 질의응답에서 높은 정확도를 기록하였다.

- 질문 : 현조문씨는 90구문에 아침에 [LDMS]? 정답 : [LDMS] 정조 [UDMS] 로 프시 매축 : [LDMS] 정조 [UDMS] 로 프시키고 [전다본째] 등에 등조하던 비문이 나타났니다.
- 지본 : 참조 기업관시의 인간이 참위와 관련이 있을 공우 배당 10대를 WIT-FI 값은 : 참조한 무있을 의치합니다? 정답 : 기업문서의 인간에 합투어 관련이 있을 경우 제당 인간을 순압하는 행위 배축 : 기업문서의 인간에 합투어 관련이 있을 경우 제당 연간을 순압하는 행위
- #### 구시 전체에 생각하여 열업한 후, 설정 사장을 모든 경수 문서에 동집하게 작용하여 점수합니다. 하게 되는 문서를 선택하여 열업한 후, 설정 사장을 모든 경수 문서에 동집하게 작용하여 점수합니다.

그림. 5 BERT 다국어 모델 추론

지는 : 육교 성이기간 변경에게 전혀 중인 목표의 성이 기 집한 : 육표 음료는 누가 할 수 있습니까 점당 : 목표 작성자 또는 목표 석제 관련이 있는 서울자 대축 : 목표 작성자 또는 목표 석제 관련이 있는 서울자

도한 등록하기 모든 주지를 등록하여 목중 인간에 대한 다른 사진들의 제안을 들을 수 있습니다. 등록한 도한 주제는 주시 공개하기 않고 중요의 근리자이 승인을 받은 후에 공개합니다. 1등속의 에는에서 중요의 홈 회전 상당 동료한 조조 주제가 공개교기 전에 대한 집에가 발전합니다.

변경 : 점 위하는 전혀 속이 있다. 있는 현수 시에 참석으로 설명 수 있는지 위한 제한 제한 전략적 전혀서 간접 하셔서 제신이 만든 회에서요 2 점 사용을 입어서요 3 제신이 등 전다 당 : 전혀 위한 시에 참석으로 설명하는 전혀 보다면 있다. 당 : 전혀 위한 시에 참석으로 설명하는 전혀 보다면 있다. 당 : 전혀 위한 시에 참석으로 설명하는 전혀 보다면 있다.

그림. 6 KoBERT 모델 추론

# 결론

본 논문에서는 다국어 BERT 모델이 종종 [UNK](unknown token)을 출력하여 한국어 처리에 한계가 있음을 보였다. 특히, KoBERT를 적용한 결과 정확도가 향상되며, 한국어 질의응답에서 KoBERT의 우수성이 두드러졌다. 이는 KoBERT와 같은 도메인 특화 언어 모델이 한국어 고객 서비스의 품질 향상에 중요한 역할을 한다는 점을 시 사한다.