



# 소비자 문의 QnA 챗봇 및 감성분석

# 목차

1 프로젝트 목적

2 데이터 전처리

3 모델 및 결과

4 마치며

---

## 1 프로젝트 목적

### 1) 소비자 문의 QnA 챗봇

- 배송/교환, 구매/결제, 제품/포장 등 소비자 문의를 입력하면 적절한 답변을 해주는 챗봇 제작
- Transformer 기반
- 고객응대 서비스 자동화 활용 가능

### 2) 소비자 문의 감성분석

- 소비자 문의 내용에 대한 긍정/부정 감성분석
- BERT 기반
- 질문 카테고리별 긍정/부정 비율 분석 등 마케팅 활용 가능

## 2 데이터 전처리

### 원본 데이터셋

① AI Hub(음성/자연어) 소상공인 고객 주문 질의-응답 텍스트 약 473만 개

② 카테고리 14개

: 가구인테리어, 건강, 디지털가전, 병원, 뷰티, 생활잡화, 슈퍼,  
식품, 음식점, 의류, 출산육아, 카페, 패션

③ 질문유형 대분류 17개

: 주문, 배송, 교환, 구매, 결제, 제품, 포장, 매장, 웹사이트, 행사, AS, 멤버십,  
부가, 수납, 수술, 예약, 외래

Unnamed: 0			IDX	발화자	발화문	② 카테고리	QA번호	QA여부	감성	③ 인텐트	가격	수량	크기	장소	조직	사람	시간	날짜	상품명	상당번호	상당대순번		
0	0	1	c		각각 두 번 주문, 결제하여 하나는 취소요청했어요 나머지 하나 주문만 발송되는 거 맞죠?	출산육아	157383	q	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	157383	1	
1	1	2	s		주문해주시 리얼소프트 밴드 특대형 4랙 급일 오전 중 정상 출고된 것으로 확인됩니다.	출산육아	157383	a	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	리얼소프트 밴드 특대형	157383	2		
2	2	3	c		주문 취소 확인 안 되는데 요주문취소된 거 맞나요?	출산육아	92796	q	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	92796	1	
3	3	4	s		네,잘못 간 주소로 반품요청을 해드렸는데요	출산육아	92796	a	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	92796	2	
4	4	5	c		방문 수령해 도 택배비 부담해야 되나요?	출산육아	135	q	m	배송_비용_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	135	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		
4734626	271718	271719	s		네, 냉동식품은 최대한 빨리 배송해드리고 있습니다.	슈퍼	55931	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	55931	2	
4734627	271719	271720	c		오늘 시킨 꿀 배송되었는지 확인해 주세요.	슈퍼	46757	q	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	꿀	46757	1
4734628	271720	271721	s		네, 꿀 발송 여부 확인해 볼게요.	슈퍼	46757	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	꿀	46757	2
4734629	271721	271722	c		내일 배송받고 싶은데, 배송일 바꿀 수 있습니까?	슈퍼	28366	q	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28366	1	
4734630	271722	271723	s		네, 배송일 변경할 수 있습니다.	슈퍼	28366	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28366	2	

4734631 rows x 20 columns

1

## 2 데이터 전처리

### Data preprocessing

- 고객의 질문, 판매자의 대답만 추출하여 각각 하나의 column으로 만드는 전처리 모듈 적용

<전처리 전>

- 발화자 : c 는 고객, s는 판매자
- QA여부 : q는 질문, a는 대답

Unnamed: 0	ID	발화자	발화문	카테고리	QA번호	QA여부	감성	인텐트	가격	수량	크기	장소	조직	사람	시간	날짜	상품명	상당번호	상당내순번
0	0	c	각각 두 번 주문, 결제하여 하나는 취소요청했어요 나머지 하나 주문만 발송되는 거 맞죠?	출산육아	157383	q	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	157383	1
1	1	s	주문해주신 리얼소프트 밴드 특대형 4랙 급일 오전 중 정상 출고된 것으로 확인됩니다.	출산육아	157383	a	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	리얼소프트 밴드 특대형	157383	2
2	2	c	주문 취소 확인 안 되는데 요주문취소된 거 맞나요?	출산육아	92796	q	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	92796	1
3	3	s	네, 잘못 간 주소로 반품요청을 해드렸는데요.	출산육아	92796	a	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	92796	2
4	4	c	방문 수령해 도 택배비 부담해야 되나요?	출산육아	135	q	m	배송_비용_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	135	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4734626	271718	271718	s	네, 냉동식품은 최대한 빨리 배송해드리고 있습니다.	슈퍼	55931	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	55931	2
4734627	271719	271720	c	오늘 시킨 꿀 발송되었는지 확인해 주세요.	슈퍼	46757	q	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	꿀	46757	1
4734628	271720	271720	s	네, 꿀 발송 여부 확인해 볼게요.	슈퍼	46757	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	꿀	46757	2
4734629	271721	271721	c	내일 배송받고 싶은데, 배송일 바꿀 수 있습니까?	슈퍼	28366	q	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28366	1
4734630	271722	271722	s	네, 배송일 변경할 수 있습니다.	슈퍼	28366	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28366	2

4734631 rows × 20 columns

## 2 데이터 전처리 Data preprocessing

```
def dataframe_handling_for_Q_and_A(DataFrame):
```

```
    df_need_col = DataFrame[['발화자', '발화문',  
                             'QA번호', 'QA여부',  
                             '상담번호', '상담내순번',  
                             '인텐트']) # 필요한 columns만 추출
```

```
    df_sorted = df_need_col.sort_values(['상담번호', 'QA번호']) # 대화 주제별 및 순서 정렬
```

```
    cond_customer_question= (df_sorted['발화자'] == 'c') & (df_sorted['QA여부'] == 'q') # consumer의 질문만 추출
```

```
    cond_seller_answer = (df_sorted['발화자'] == 's') & (df_sorted['QA여부'] == 'a') # seller의 답변만 추출
```

```
    df_filtered = df_sorted.loc[cond_seller_answer|cond_customer_question]
```

```
    list_talker = list(df_filtered['발화자'])
```

```
    list_sentence = list(df_filtered['발화문'])
```

```
    new_list_s, new_list_c = list_seperate_for_Q_and_A(list_talker, list_sentence)
```

```
    df_customer = pd.DataFrame(new_list_c)
```

```
    df_seller = pd.DataFrame(new_list_s)
```

## 2 데이터 전처리

### Data preprocessing

- 고객의 질문, 판매자의 대답만 추출하여 각각 하나의 column으로 만드는 전처리 모듈 적용

<전처리 후>

- 'Question' column : 고객(c)의 질문(q)
- 'Answer' column : 판매자(s)의 답변(a)

	Question	Answer
5	붕대 안 썼으면 환불 가능해요?	네, 붕대 안 쓰셨으면 환불 가능합니다.
6	무료 a/s 기간이 언제까지죠? 히팅 서포터 무릎용은 1개 구매하면 양 무릎용 2개...	수량을문의하신 상품은 2개로 구매해 주셔야 양쪽 무릎에 사용 가능해요.
7	하루에 몇 팩씩 먹을 수 있나여?	보통 하루 한 포 드시기도 하는데 혈당 조적을 위해서 하루 한 포씩 세 번까지 도드...
8	문제가 있는지 이 가습기 작동을 안하는데 확인해 주시겠어요?	네, 제가 열어보고 문제가 뭔지 알려드리겠습니다.
9	뚜껑 있나요?	문의는 테스터 아닌 본품으로 뚜껑이 있는 제품이 맞습니다. 네, 그렇습니다. 2만원...
...	...	...
2192402	세로가 183가로가 100인 가요?	가로세로는 고객님께서 사용하고자 하는 방향에 따라 달라집니다.
2192403	몇 년 제품인가여?	현재 저희 쪽에서 출고되고 있는 상품은 2021년형으로 본사에서 입고되자마자 바로바...
2192404	뉴청맥전기장판과 어떻게 다른 가요?	상판의 원단 제조사가 다릅니다. 각각 LG와 KCC 원단을 사용하고 있습니다.
2192405	15시간 후에는 자동 꺼지는지요?	안전을 위해 15시간 지나면 전원이 자동으로 차단됩니다.
2192406	전기요금은 어떻게 되나요?	해당 상품을 하루 8시간 30일 사용하신다면 한 달 9000원 정도 부담하시면 됩니다.

2192402 rows × 2 columns

## 2 데이터 전처리

### 원본 데이터셋 - BERT 감성분석 사전학습

- AI Hub(음성/자연어) 감성 대화 말뭉치 27만 문장
- 감정분류 60가지  
: 기본 감정 6개(분노, 슬픔, 불안, 상처, 당황, 기쁨) 및  
각 기본감정별 9개의 세부감정

60가지 감정 분류						
기분	분노	슬픔	불안	상처	당황	기쁨
1	툼툼대는	실망한	두려운	질투하는	고립된	감사하는
2	좌절한	비통한	스트레스 받는	배신당한	남의 시선 의식하는	사랑하는
3	짜증나는	후회되는	취약한	고립된	외로운	편안한
4	방어적인	우울한	혼란스러운	충격 받은	열등감	만족스러운
5	악의적인	마비된	당혹스러운	불우한	죄책감	흥분되는
6	안달하는	염세적인	회의적인	희생된	부끄러운	느긋한
7	구역질 나는	눈물이 나는	걱정스러운	억울한	혐오스러운	안도하는
8	노여워하는	낙담한	조심스러운	괴로워하는	한심한	신이 난
9	성가신	환멸을 느끼는	초조한	버려진	혼란스러운	자신하는



## 2 데이터 전처리

### Data preprocessing - BERT 감성분석 사전학습

- 긍정 9가지, 부정 14가지 선별

긍정 : '신이 난', '안도하는', '기쁨', '만족스러운', '자신하는', '편안한',  
'감사하는', '신뢰하는', '느긋한'

부정 : '스트레스 받는', '당황', '분노', '회의적인', '혼란스러운', '흥분되는', '불안',  
'걱정스러운', '충격 받은', '짜증내는', '실망한', '슬픔', '툭툭대는', '악의적인'

Unnamed: 0

사람문장 감정

0	0	난 내 가족을 꾸리는 게 이런 느낌일 줄 몰랐어.	긍정
1	1	저녁에 약속이 있었는데 야근을 안 해서 다행이야. 절대로 빠지면 안 되는 약속이었거든.	긍정
2	2	응. 어릴 적부터 그랬는데 노년이 되어서도 이러네.	부정
3	3	요즘 내가 치마만 입으면 남자 동료들이 내 다리만 쳐다보는 것 같아서 기분 나빠.	부정
4	4	이번 팀플레이는 팀원들이 적극적으로 임해서 크게 힘들지 않고 좋은 결과를 낼 수 있었어.	긍정

### 3 모델 및 결과

#### 1) 소비자 문의 QnA 챗봇

- Transformer 모델 사용

```
1 tf.keras.backend.clear_session()
2
3 # Hyper-parameters
4 NUM_LAYERS = 2
5 D_MODEL = 256
6 NUM_HEADS = 8
7 DFF = 512
8 DROPOUT = 0.2
9
10 model = transformer([
11     vocab_size=VOCAB_SIZE,
12     num_layers=NUM_LAYERS,
13     dff=DFF,
14     d_model=D_MODEL,
15     num_heads=NUM_HEADS,
16     dropout=DROPOUT])
```

```
[ ] output = predict('이거 택배 회사가 어디인가요?')
```

Input: 이거 택배 회사가 어디인가요?

Output: 편의점 포스트박스 고객센터로 가시면 됩니다 .

```
[ ] output = predict('지금도 배달 신청 되나요?')
```

Input: 지금도 배달 신청 되나요?

Output: 네 , 배달 가능합니다 .

```
[ ] output = predict('상한 제품 같아요')
```

Input: 상한 제품 같아요

Output: 죄송합니다 , 지금 바로 환불해 드리겠습니다 .

### 3 모델 및 결과

#### 2) 소비자 문의 감성분석

- BERT 모델 사용
- AI Hub 감성 대화 말뭉치 데이터로 사전학습



config



```
BertConfig {  
  "_name_or_path": "klue/bert-base",  
  "architectures": [  
    "BertForMaskedLM"  
  ],  
  "attention_probs_dropout_prob": 0.1,  
  "classifier_dropout": null,  
  "hidden_act": "gelu",  
  "hidden_dropout_prob": 0.1,  
  "hidden_size": 768,  
  "initializer_range": 0.02,  
  "intermediate_size": 3072,  
  "layer_norm_eps": 1e-12,  
  "max_position_embeddings": 512,  
  "model_type": "bert",  
  "num_attention_heads": 12,  
  "num_hidden_layers": 12,  
  "pad_token_id": 0,  
  "position_embedding_type": "absolute",  
  "transformers_version": "4.19.2",  
  "type_vocab_size": 2,  
  "use_cache": true,  
  "vocab_size": 32000  
}
```

결과  
이미지

---

## 3 모델 및 결과

통합 시각화

빛남우와

## 4 마치며

### 1) 어려웠던 점

- 데이터 전처리

: AI hub 원본 데이터셋이 복잡

⇒ 단순화하는 과정 어려움

Unnamed: 0	IDX	발화자	발화문	카테고리	QA번호	QA여부	감성	인텐트	가격	수량	크기	장소	조직	사람	시간	날짜	상품명	상당번호	상당내순번
0	0	1	c	각각 두 번 주문, 결제하여 하나는 취소요청했어요 나머지 하나 주문만 발송되는 거 맞죠?	출산육아	157383	q	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	157383	1
1	1	2	s	주문해주신 리얼소프트 밴드 특대형 4랙 금일 오전 중 정상 출고된 것으로 확인됩니다.	출산육아	157383	a	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	리얼소프트 밴드 특대형	157383	2
2	2	3	c	주문 취소 확인 안 되는데 모주문취소된 거 맞나요?	출산육아	92796	q	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	92796	1
3	3	4	s	네, 잘못 간 주소로 반품요청을 해드렸는데요.	출산육아	92796	a	m	주문_취소_확인	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	92796	2
4	4	5	c	방문 수령해 도 택배비 부담해야 되나요?	출산육아	135	q	m	배송_비용_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	135	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4734626	271718	271719	s	네, 냉동식품은 최대한 빨리 배송해드리고 있습니다.	슈퍼	55931	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	55931	2
4734627	271719	271720	c	오늘 시킨 꿀 발송되었는지 확인해 주세요.	슈퍼	46757	q	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	꿀	46757	1
4734628	271720	271721	s	네, 꿀 발송 여부 확인해 볼게요.	슈퍼	46757	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	꿀	46757	2
4734629	271721	271722	c	내일 배송받고 싶은데, 배송일 바꿀 수 있습니까?	슈퍼	28366	q	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28366	1
4734630	271722	271723	s	네, 배송일 변경할 수 있습니다.	슈퍼	28366	a	m	배송_날짜_요청	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28366	2

4734631 rows × 20 columns

```
def dataframe_handling_for_Q_and_A(DataFrame):

    df_need_col = DataFrame[['발화자', '발화문',
                              'QA번호', 'QA여부',
                              '상당번호', '상당내순번',
                              '인텐트']) # 필요한 columns만 추출

    df_sorted = df_need_col.sort_values(['상당번호', 'QA번호']) # 대화 주제별 및 순서 정렬

    cond_customer_question= (df_sorted['발화자'] == 'c') & (df_sorted['QA여부'] == 'q') # consumer의 질문만 추출
    cond_seller_answer = (df_sorted['발화자'] == 's') & (df_sorted['QA여부'] == 'a') # seller의 답변만 추출

    df_filtered = df_sorted.loc[cond_seller_answer|cond_customer_question]

    list_talker = list(df_filtered['발화자'])
    list_sentence = list(df_filtered['발화문'])

    new_list_s, new_list_c = list_seperate_for_Q_and_A(list_talker, list_sentence)

    df_customer = pd.DataFrame(new_list_c)
    df_seller = pd.DataFrame(new_list_s)

    df_Q_A = pd.concat([df_customer, df_seller], axis=1)
    df_Q_A.columns = ['Question', 'Answer']
```

## 4 마치며

### 2) 아쉬웠던 점

- 다양한 카테고리 통합에 따른 구체적 답변의 한계  
: 카테고리별로 상황이 제각기 다르기 때문에 디테일한 답변 어려움

```
▶ set(df_train['카테고리'])
```

```
↳ {'가구인테리어',  
    '건강',  
    '기타',  
    '디지털가전',  
    '병원',  
    '뷰티',  
    '생활잡화',  
    '슈퍼',  
    '식품',  
    '음식점',  
    '의류',  
    '출산육아',  
    '카페',  
    '패션'}
```

```
▶ Q = list(df_train['인텐트'].unique())  
Qtype = []
```

```
for i in range(len(Q)):  
    Qtype.append(Q[i][:2])
```

```
set(Qtype)
```

```
↳ {'AS',  
    '결제',  
    '교환',  
    '구매',  
    '매장',  
    '멤버',  
    '배송',  
    '부가',  
    '수납',  
    '수술',  
    '예약',  
    '외래',  
    '웹사',  
    '제품',  
    '주문',  
    '포장',  
    '행사'}
```

## 4 마치며

### 2) 아쉬웠던 점

- 애매한 질문에 대한 안내원 연결 디폴트 답변 구현
  - : 챗봇이 답하기 애매한 질문에도 학습기반 답변을 반환
  - ⇒ 질문이 애매한 경우 안내원 연결 디폴트 답변을 하도록 구현하려 했으나, 이를 위한 또 하나의 모델 구축을 완성하지 못함

```
[ ] output = predict('배고파요')
```

```
Input: 배고파요  
Output: 네 삼만원 이상 사시면 배달됩니다
```

```
[ ] output = predict('사랑해요')
```

```
Input: 사랑해요  
Output: 배달은 하지않습니다
```

```
[ ] output = predict('이렇게 하면 널 가질 수 있을 거라고 생각했어!')
```

```
Input: 이렇게 하면 널 가질 수 있을 거라고 생각했어!  
Output: 지금 바로 다시 보내드리겠습니다 .
```

감사합니다.