

아이디어 고민

1. 주제 방향성에 대한 고민

<기술적>

QA 형태

- KoGPT2 Text Generation(Auto Regressive)

지도학습 후, rule-base 형태로 제공

- KoELECTAR & KoBERT Text Classification
- 라벨을 달 때, 0 1 2 3... 등의 지도학습으로 만들고, 결과 값은 0은 2~3가지 / 1은 2~3가지 / 2는 2~3가지 / 3은 2~3가지로 문장을 만들어서 랜덤값으로 챗봇이 선정하여 return 하도록

<도메인>

- 심리 상담처럼 위로 형태로 제공
- 소상공인 분야 맞춤형 인공지능 데이터 활용 및 챗봇 서비스 구축에 활용
- 디지털 소외계층도 함께 누릴 수 있는 챗봇 생태계 조성을 위한 기반 데이터로 활용
- 공공 분야의 민원 상담 데이터를 기반으로 인공지능 민원 상담 시스템의 고도화 및 실시간 민원 접수에 활용

⇒

<https://drive.google.com/drive/u/0/folders/1u4XfkK5FimAlOUXR4r1sZCpwTUvu5NG2>

2. 데이터 셋

ChatbotData

Q	A	label
12시 땡!	하루가 또 가네요.	0
1지망 학교 떨어졌어	위로해 드립니다.	0
3박4일 놀러가고 싶다	여행은 언제나 좋죠.	0
3박4일 정도 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

- Label이 붙었기 때문에 KoELECTAR & KoBERT Text Classification 를 사용
- https://github.com/songys/Chatbot_data

SPEAKER	SENTENCE	DATAID	DOMAINID	DOMAIN	CATEGORY	SPEAKERID	SENTENCEID	MAIN	SUB	QA	QACNCT	MQ	SQ	UA	SA	개체명	용어사전	지식베이스
고객	첫 방문 알림	16321031	E	생원서비스	미용실	1	1	알림 문의		Q			첫 방문 알림하나요?			첫, 방문, 알림		
직원	네, 30프로	16321031	E	생원서비스	미용실	0	2	알림 문의		A						네, 30프로		30프로/알림
고객	몇 시까지가	16321031	E	생원서비스	미용실	1	3	원하는 시간에 예약이 가능한 문의		Q			몇 시까지 가야 할 수 있나요?			몇 시, 원		원/시
직원	오후 6시까지	16321031	E	생원서비스	미용실	0	4	원하는 시간에 예약이 가능한 문의		A						오후 6시까지		오후 6시/시
고객	두피 케어	16321031	E	생원서비스	미용실	1	5	모발 케어 서비스 가격 문의		Q			두피 케어 서비스는 얼마인가요?			두피 케어 서비스		두피 케어 시
직원	제품마다	16321031	E	생원서비스	미용실	0	6	모발 케어 서비스 가격 문의		A						제품마다	제품, 가격, 샴	
고객	임색약은	16321031	E	생원서비스	미용실	1	7	임색 제품 문의		Q			임색약은 뭐 쓰시나요?			임색약		임색약/제품
직원	로레알 사용	16321031	E	생원서비스	미용실	0	8	임색 제품 문의		A						로레알 사용	로레알	로레알/쓰
고객	키보라고	16321031	E	생원서비스	미용실	1	9	상투 서비스 문의		Q			키보라고 머리 잡게 주나요?			키보, 머리		키보/시
직원	네, 잡게	16321031	E	생원서비스	미용실	0	10	상투 서비스 문의		A						네, 잡게		

- QA 형태로 KoGPT2 Text Generation(Auto Regressive)
- <https://drive.google.com/drive/u/0/folders/1u4XfkK5FimAlOUXR4r1sZCpwTUvu5NG2>

3. 오픈 소스

[https://github.com/nawnoes/WellnessConversationAI?](https://github.com/nawnoes/WellnessConversationAI?fbclid=IwAR3ZhXYW_Dwl2RXP1mbHzvafGXF80QWERa4t6TTz_m2NQug5QwjOwQt6Hvw)
[fbclid=IwAR3ZhXYW_Dwl2RXP1mbHzvafGXF80QWERa4t6TTz_m2NQug5QwjOwQt6Hvw](https://github.com/nawnoes/WellnessConversationAI?fbclid=IwAR3ZhXYW_Dwl2RXP1mbHzvafGXF80QWERa4t6TTz_m2NQug5QwjOwQt6Hvw)
 (WellnessConversationAI)

- KoELECTAR & KoBERT Text Classification
- KoGPT2 Text Generation(Auto Regressive)

<https://github.com/ko-nlp/Korpora>

(이기창씨 - 일반대화 챗봇)

4. 궁금증

- QA 방법으로 하면 학습이 다 가능한가?
 - 라벨링을 해야될 것 같아서...

5. 현재 내 상황

- Text Classification 형태로 라벨링이 돌아가도록 오픈소스를 실행하고 있음.

참고