# 제주도 관광지 정보 상담 업무 자동화 서비스 챗봇 개발 발표 자료

안녕하십니까, 두둥탁팀 김호경입니다. 저희 팀이 직접 개발한 제주도 관광지 정보를 제공하는 챗봇 서비스를 발표하겠습니다.

[목차 10초 알아서 설명]

심심이. 다들 아시죠. 스마트폰 보급이 가속화될 때쯤 한 번쯤은 보셨을 인공지능 대화 엔진입니다. 저도 오래전에 사용해보았는데 전혀 사람 같다는 느낌이 들지 않고, 대화가 뚝뚝 끊기는 느낌이라 금세 실증이 났던 기억이 납니다. 화면을 보시면 비트 코인에 대해 물으니 자살을 유도하고, 질문과 응답이 전혀 매칭이 되지 않고 있습니다. 이렇게 챗봇은 나쁜 말 데이터를 학습할 수도 있고, 질문의 요점을 파악하지 못하기도 합니다. 추가적으로 한 가지 주제로 대화를 유지하는 것이 어려운데, 이는 인공지능 대화 업계에서 최고라 불리는 애플의 시리도 제대로 구현하지 못한다고 합니다. 때문에 심심이는 인공지능 개발보다 데이터 수집에만 집중한 것이 약점이라는 평을 받았지만, 현재는 꾸준한 딥러닝 개발로 나쁜 말을 제어하고 있고, 올해부터 네이버 클로바를 통해 음성대화챗봇 서비스도 시작했다고 합니다. 이렇게 챗봇은 많은 양의 좋은 데이터와 인공지능 기술의 적절한 조화가 중요하다고 생각합니다.

요즘 제주도가 많이 조용해졌지만 제주도는 여전히 한국인이 가장 사랑하는 우리나라 대표 관광지입니다. 하지만 저희가 알기로는 도내 관광지를 통합 관리하는 곳이 없었습니다. 도내에 오프라인 관광안내센터가 여러 군데 있지만 제주시청, 서귀포시청, 관광진흥과, 제주관광공사, 제주관광협회, 한국관광공사 등 담당부서가 상이하여 관광지 데이터 통합관리가 이루어지지 않고 있습니다. 코로나19 사태로 축제가 취소되고 운영이 중단되는 관광지들이 있지만, 종합적인 안내가 부족해 도민과 관광객에게 혼선을 주고 있다고 합니다.

둘째로 도내에 관광지 정보 문의 관련 자동화 서비스가 없었습니다. 현재 제주관광공사가 운영하는 제주관광정보센터에서 카카오톡 채널을 통해 1:1 질문이 가능하나 상담 인력에 의존하고 있습니다. 저희 팀은 이런 점을 착안하여 제주도 관광지 정보 상담 업무를 자동화하는 챗봇을 개발해보았습니다.

챗봇은 대화를 의미하는 Chatter와 로봇의 합성어입니다.

사람이 사용하는 언어, 즉 자연어를 이해하여 문장에서 의도와 핵심 키워드를 파악해야 합니다. 사용자가 ‘오늘 날씨 어때?”라고 묻는다면 오늘이 며칠인지, 위치가 어디인지 파악해야 하는데 추가 정보를 구하는 과정을 Slot Filling이라합니다. 챗봇은 보통 특정 목적으로, 또는 기업의 고객에게 특화된 서비스를 구축하기 위해서 고객이 필요로 하는 분야의 정보를 파악해 질문을 답변을 세팅해야 합니다. 영어에서 ‘비즈니스 영어’와 ‘생활 영어’가 나뉘어 있듯, 모든 것을 학습할 수 없어 필요한 상황에 맞는 문장만 배우는 것을 ‘대화 모델’이라고 합니다. 이렇게 챗봇에게 필요한 특성은 사실 이보다 더 많습니다.

다음은 챗봇의 타입입니다. 크게 검색 모델과 생성 모델로 나뉩니다. 한국 기업이 사용해온 대부분의 챗봇은 특정 주제를 다루는 검색 모델, 즉 미리 만들어진 답변 중 답을 고르는 시나리오 타입의 챗봇이었습니다. 사실상 메뉴를 채팅 형태로 도식화한 것이라 정해진 주제 말고는 대화가 불가능했지만, 점점 인공지능을 활용한 생성 모델로 진화하고 있습니다. 생성 모델은 대화 상대가 이런 질문을 했을 때는 이런 식의 답변을 하는 것이 최선이다라는 판단을 내려 답변을 생성합니다. 정확도를 높이기 위해서는 엄청난 양의 데이터가 필요합니다. 저희 챗봇 모델은 이 두가지를 섞은 형태라고 할 수 있을 것 같습니다.

프로젝트를 위해 필요했던 데이터들입니다. 딥러닝 활용을 위해 가장 중요한 데이터는 질문과 답변이 세트로 이루어진 텍스트 데이터였으나 결국에는 구하지 못해 직접 생성하고 증폭했습니다. 둘째로는 관광지 안내를 위한 도내 관광지의 세부 정보가 필요했습니다. 부가 기능으로 특정 관광지 질문 검색 시 할인 티켓 구매 주소를 전송해 주기 위해 관광협회에서 운영하는 제주관광상품 오픈마켓에서 할인 상품들을 크롤링했습니다.

모델 학습을 위해 직접 만든 질문-응답 데이터셋입니다. 관광지 주소, 소개, 주차 정보, 연락처, 입장료, 이용 시간, 소요 시간, 날씨, 날씨 기반 추천, 위치 기반 추천, 유모차 대여 가능 여부로 코드를 나누어 사용자의 관점에서 질문을 직접 작성하여 800개 이상의 질문을 만들었습니다. 그에 대응하는 응답 데이터는 텍스트 요약 형태로 ‘관광지’+’키워드’ 형식으로 만들었습니다.

이를 다시 저희가 가진 도내 관광지들의 이름과 결합시켜 총 38만 개의 질문-응답 데이터가 생성되었습니다.

이렇게 만들어진 데이터셋들은 모두 토큰화하고 형태소를 분석하여 질문자의 질문을 인식할 수 있도록 단어 사전을 만들어 전처리했습니다.

저희가 사용한 딥러닝 모델 트랜스포머입니다. 트랜스포머는 2017년에 구글이 발표한 신경망 모델로, RNN(순환신경망)이 가장 발전된 형태의 아키텍처인 Seq2Seq를 기반으로 Attention 신경망으로만 이루어져 있다는 것이 특징입니다. 셀프-어텐션만을 사용하면서 생기는 장점은 레이어당 전체 연산량이 줄어들고, 병렬화가 가능한 연산이 늘어나며, 긴 텍스트도 잘 학습하게 된다는 것입니다. 학습 속도가 빨라서 이런 단기 프로젝트에 적합했습니다.

텐서보드로 확인한 텐서 그래프의 구조입니다.

약 2만번의 학습을 거쳐 accuracy 0.\*\*\*, loss 값은 0.\*\*\*\*

이렇게 학습된 모델에 부가 기능을 첨가하기 위해 사용한 API들입니다. 앞의 두 가지 API를 활용하여 검색어에 따라 위도와 경도를 추출한 뒤 근처 가까운 관광지를 알려주도록 했습니다. 기상청 API는 오늘 날씨 예보와 오늘 날씨에 따른 관광지 추천을 하는데 활용했습니다. 마지막으로 텔레그램 봇 API로 사용자 환경을 손쉽게 구축했습니다.

그렇게 해서 만들어진 저희 챗봇의 기능들을 소개합니다.

크게 위치 기반 추천, 주차, ~~~ 등 총 11가지 종류의 질문에 응답할 수 있고요,

사용자가 관광지의 주소, 입장 요금, 소개 및 정보 관한 질문을 할 때 해당 관광지에 실제로 방문할 의사가 많다고 생각하여 이 세가지에 대한 응답 시 할인 예약을 할 수 있도록 URL을 전송하게 됩니다.

즉석에서 시연을 보여드리고 싶지만 DB 서버를 결제할 돈이 없어 동영상으로 대체하겠습니다.

아직 저희 챗봇의 성능이 많이 부족하지만 추가 개발을 한다면 여행객으로부터 데이터를 수집해서 여행객 트렌드 파악이 가능하고, 인력 없이 24시간 운용 가능하며, 각종 비용을 절감할 수 있습니다.

한계점으로는 데이터를 구하지 못해 한 달의 시간을 낭비한 점, 때문에 데이터의 단순화로 성능이 부족한 점, 각종 API를 결합했으나 최대한으로 활용하지 못한 점입니다.

참고 자료와 활용한 도구들 입니다

감사합니다.