

챗봇을 활용한 영화정보 서비스 개발

김유리*

*한성대학교 산업경영공학과

e-mail : dbflek620@gmail.com

The development of cinema information service using chatbot

Yu-Ri Kim*

*Dept. of Industrial & Management Engineering, Han-Sung University

요 약

인공지능 기술이 발달하면서 챗봇 플랫폼이 주목받고 있다. 챗봇이란 규칙 또는 인공지능(AI)을 이용해 사용자와 상호작용을 하는 대화형 인터페이스다. 챗봇에서 대화를 처리하는 방법은 규칙기반 대화 시스템, 검색기반 대화 시스템, 생성기반 대화 시스템이 있다. 본 논문에서는 규칙 기반 대화 시스템을 바탕으로 하는 모바일 영화 챗봇 서비스를 개발하였다. 이를 통하여 사용자는 더 편리하게 영화 관련 정보를 제공받을 수 있다.

1. 서론

인공지능 기술이 발달하면서 챗봇이 IT 기업들에 새로운 플랫폼으로 주목받고 있다. 챗봇이란 대화형 인터페이스 상에서 규칙 또는 인공지능(AI)을 이용해 사용자와 상호작용을 하는 서비스를 지칭한다. [1, 2] 챗봇 서비스가 주목받으면서 카카오톡, 텔레그램 등 다양한 모바일 메신저 기업들이 클라우드 기반의 오픈소스 플랫폼을 제공하여 개발자들이 쉽게 챗봇을 개발할 수 있도록 돕고 있다.

챗봇에서 대화를 처리하는 방법은 크게 규칙기반 대화 시스템, 검색 기반 대화 시스템, 생성 기반 대화 시스템이 있다. 규칙 기반 대화 시스템은 사용자 입력 발화 문장과 미리 구축한 패턴을 비교하여 가장 알맞은 패턴 발화 문장을 출력하는 것이다. 검색 기반 대화 시스템은 예제 데이터에서 가장 유사한 대화 쌍을 검색해 그 대화 쌍의 답변 문장을 출력하는 것이다. 마지막으로, 생성 기반 대화 시스템은 통계 기반 자동번역, 신경망 기반 자동번역 방법을 주로 사용한다. [3, 4]

본 연구에서는 규칙 기반 대화 시스템을 이용해 챗봇을 활용한 영화정보 서비스를 개발하고자 한다.

2. 관련 연구

챗봇은 대화 시스템에서 고객 서비스 또는 정보 제공 등 다양한 목적을 위해 설계된 플랫폼이다. 최근, 의료, 음식, 학습 등 다양한 분야에서 챗봇 서비스에 관한 연구가 진행되고 있다.

규칙 기반 대화 시스템을 바탕으로 하는 연구는 대부분 규칙 기반 대화 시스템과 검색 기반 대화 시스템을 함께 사용하고 있다. [3, 4, 5] 사용자 입력 발화에 따라 통사/구문분석과 의미 분석을 진행한 후 예제 데이터베이스에

서 가장 유사한 대화 쌍의 답변을 출력해주거나 미리 저장해 둔 규칙에 따라 서비스를 제공하는 방식이다. [5]

다양한 관련 연구에서 대화 시스템의 규칙을 설정한 뒤 시스템 응답을 제공할 때에는 Facebook Messenger API, Kakaotalk 자동응답 API 등 오픈소스 플랫폼을 사용하는 것을 볼 수 있었다. [6, 7, 8, 9]

3. 챗봇을 활용한 영화정보 서비스

본 논문에서 제안하는 챗봇을 활용한 영화정보 서비스는 규칙 기반 대화 시스템을 바탕으로 한다. 규칙 기반 대화 시스템 구축을 위해 먼저 변수 설정, 질의·응답 상황 설정, 예제 데이터 생성, 예제 문장 분석, 키워드 설정을 진행했다. 본 논문의 대화 시스템은 크게 형태소 분석 및 구문 분석, 의미 분석, 응답 생성 단계로 구성된다.

3.1 변수 설정

분석을 위한 예제 질의·응답 데이터 구축과 사용자 입력 발화 문장 키워드 추출 시 사용할 변수를 설정하였다. 변수는 영화/영화관 정보를 묻는데 필요한 정보로 구성된다. 하단의 <표 1>은 본 논문에서 설정한 변수이다.

상위	하위	세부
영화관	특정영화관	
	불특정영화관	
영화	특정영화	영화명
		감독
		배우
		개봉일
		장르
		포스터

		줄거리
		평점
시간	불특정영화	
	현재시간	
	특정 시간	
위치	현재 위치	
	특정 위치	
날짜	당일	
	특정일	
영업 여부	영업	
	휴무	
추천		

<표 1> 사용 변수

3.2 질의·응답 상황 설정 및 예제 데이터 생성

예제 질의·응답 데이터 구축 전 3.1에서 설정한 변수들을 조합하여 상황을 설정하고, 그에 해당하는 예제 질의·응답 문장 데이터를 구축했다. 예제 질의·응답 문장 데이터 구축 시에는 상황 설정 시 포함된 변수를 질문 문장에 포함하고 해당 질문에 맞는 답변을 답변 문장으로 생성했다. 설정된 상황의 개수는 총 2,361개, 질의·응답 문장 데이터는 총 4,772개(질문 2,361개, 답변 2,351개)이다.

번호	변수 개수	포함된 변수
730	3	평점, 당일, 휴무
731	3	평점, 특정일, 영업
732	3	평점, 특정일, 휴무
733	4	영화명, 특정영화관, 현재시간, 현재 위치
734	4	영화명, 특정영화관, 현재시간, 특정 위치

<표 2> 상황 설정 예시

구분	예시
번호	941
변수 개수	4
변수 종류	장르, 불특정영화관, 현재시간, 영업
질문 문장	지금 영업 중인 영화관에서 볼 수 있는 액션 장르 영화 알려줘.
답변 문장	현재 근처 영화관에서 상영하는 액션 장르 영화를 보여 드릴게요.

<표 3> 질의·응답 문장 데이터 예시

3.3 예제 문장 분석 및 키워드 설정

앞서(3.2 참고) 생성한 예제 질의·응답 문장 데이터를 KoNLPy(KOMORAN)을 사용해 형태소 분석 및 품사 부착을 진행하였다. [13]

구분	예시
번호	941
변수 개수	4
변수 종류	장르, 불특정영화관, 현재시간, 영업
질문 문장	지금 영업 중인 영화관에서 볼 수 있는 액션 장르 영화 알려줘.
답변 문장	현재 근처 영화관에서 상영하는 액션 장르 영화를 보여 드릴게요.
질문 문장 형태소 분석 결과	(‘지금’, ‘MAG’), (‘영업’, ‘NNG’), (‘중’, ‘NNB’), (‘이’, ‘VCP’), (‘ㄴ’, ‘ETM’), (‘영화관’, ‘NNG’), (‘에서’, ‘JKB’),

	(‘보’, ‘VV’), (‘르’, ‘ETM’), (‘수’, ‘NNB’), (‘있’, ‘VV’), (‘는’, ‘ETM’), (‘액션’, ‘NNG’), (‘장르’, ‘NNG’), (‘영화’, ‘NNG’), (‘알리’, ‘VV’), (‘어’, ‘EC’), (‘주’, ‘VX’), (‘어’, ‘EC’), (‘.’, ‘SF’)
답변 문장 형태소 분석 결과	(‘현재’, ‘MAG’), (‘근처’, ‘NNG’), (‘영화관’, ‘NNG’), (‘에서’, ‘JKB’), (‘상영’, ‘NNG’), (‘하’, ‘XSV’), (‘는’, ‘ETM’), (‘액션’, ‘NNG’), (‘장르’, ‘NNG’), (‘영화’, ‘NNG’), (‘를’, ‘JKO’), (‘보이’, ‘VV’), (‘어’, ‘EC’), (‘드리’, ‘VV’), (‘르’, ‘ETM’), (‘게요’, ‘EF’), (‘.’, ‘SF’)

<표 4> 질의·응답 문장 형태소 분석 예시

품사 부착 결과를 분석하여 사용자 입력 발화 문장의 미 분석 시 사용할 품사를 지정했다. 사용할 품사는 총 20개이며, 그 결과는 아래 <표 5>와 같다.

Tag	Description	Tag	Description
NNG	일반명사	EP	선어말어미
NNP	고유명사	EF	종결어미
NNB	의존명사	EC	연결어미
NR	수사	ETN	명사형 전성어미
NP	대명사	ETM	관형형 전성어미
VV	동사	XPN	체언 접두사
VA	형용사	XSN	명사 파생접미사
VX	보조용언	XSV	동사 파생접미사
VCP	긍정 지정사	XSA	형용사파생접미사
VCN	부정 지정사	XR	어근
MM	관형사	SF	마침표, 물음표, 느낌표
MAG	일반부사	SE	줄임표
MAJ	접속부사	SS	마음표, 괄호표, 줄표
IC	감탄사	SP	쉼표, 가운뎃점, 콜론, 빗금
JKS	주격조사	SO	불임표
JKC	보격조사	SW	기타 기호
JKG	관형격조사	SH	한자
JKO	목적격조사	SL	외국어
JKB	부사격조사	SN	숫자
JKV	호격조사	NF	명사추정범주
JKQ	인용격조사	NV	용언추정범주
JC	접속조사	NA	분석불능범주
JX	보조사		

<표 5> 의미 분석에 사용할 품사

대화 시스템에서 사용자 입력 발화 문장 분석 시 키워드를 추출한다. 질의·응답 문장 데이터의 품사 부착 결과를 분석하여 키워드를 지정하였다. 키워드는 변수 파악 키워드, 답변 문장 생성 시 사용할 키워드, 제거 키워드로 나누어 추출한다. 변수 파악 키워드는 형태소 분석 결과 하단의 <표 6> 범주에 포함되는 단어들이다. 답변 문장 생성 시 사용할 키워드는 변수 파악 키워드가 아니지만, 답변 문장을 생성할 때 사용되는 단어들이다. 마지막 제거 키워드는 품사 부착 결과가 의미 분석에 사용하는 품사에(<표 5> 참조) 해당하지 않는 단어들이다.

번호	0	1	2	3	4	5	6
변수	시간	날짜	영화	영화관	영업	휴무	위치
번호	7	8	9	10	11	12	13
변수	감독	배우	개봉일	장르	평점	줄거리	추천

<표 6> 사용자 입력 문장 의미 분석에 사용할 변수

3.4 대화 시스템

본 논문의 대화 시스템은 크게 형태소 분석 및 구문 분석, 의미 분석, 응답 생성 단계로 구성된다.

3.4.1 형태소 분석 및 구문 분석

사용자 입력 발화 문장 형태소 분석은 KoNLPy(KOMORAN)를 사용해 진행하였다. [13] 형태소 분석이란 형태소를 비롯하여 어근, 접두사/접미사, 품사(POS, part-of-speech) 등 언어적 속성 구조를 파악하는 것이다. [11]

형태소 분석을 통해 문장에 품사를 부착한 후에는 품사들을 조금 더 큰 단위의 묶음, 즉 구문으로 묶을 수 있다. 구문 분석은 KoNLPy(KOMORAN)과 NLTK 3.2.5 패키지를 이용하였다. [13, 14]

구분	예시
예시 문장	지금 영업 중인 영화관에서 볼 수 있는 액션 장르 영화 알려줘.
형태소 분석 결과	(‘현재’, ‘MAG’), (‘근처’, ‘NNG’), (‘영화관’, ‘NNG’), (‘에서’, ‘JKB’), (‘상영’, ‘NNG’), (‘하’, ‘XSV’), (‘는’, ‘ETM’), (‘액션’, ‘NNG’), (‘장르’, ‘NNG’), (‘영화’, ‘NNG’), (‘를’, ‘JKO’), (‘보이’, ‘VV’), (‘어’, ‘EC’), (‘드리’, ‘VV’), (‘르’, ‘ETM’), (‘게요’, ‘EF’), (‘.’, ‘SF’)
구문 분석 결과	(S 현재/MAG (NP 근처/NNG 영화관/NNG) 에서/JKB (NP 상영/NNG) 하/XSV 는/ETM (NP 액션/NNG 장르/NNG 영화/NNG) 를/JKO (VP 보이/VV) 어/EC (VP 드리/VV) 르/ETM 게요/EF ./SF)

<표 7> 형태소 분석, 구문 분석 결과 예시

3.4.2 의미 분석

의미 분석 단계는 키워드 추출, 문형인식, 긍/부정 인식, 논항 인식으로 구성된다.

키워드 추출은 3.3에서 언급한 방식으로 진행한다. 다음으로 문형인식 단계에서 시스템 응답을 해야 하는 문장인지 판단한다. 문형 인식은 술어부, 종결/비종결 어미 등을 이용하여 의문형과 평서문으로 구분한다. 의문형으로 구분된다면, 설정했던 의미 분석 변수 <표 6>을 이용해 시스템 응답을 하는 영화 관련 문장인지 판단한다. 그리고 질

문 긍/부정에 따라 문장 의도가 달라질 수 있으므로 긍/부정 인식을 수행한다. 긍/부정 인식은 3.4.1의 형태소 분석 결과 (VCP, 긍정 지정사), (VCP, 부정 지정사)로 분류된 단어와 (‘열리’, ‘VV’), (‘닫’, ‘VV’)과 같이 미리 저장한 긍/부정 단어를 이용해 판단한다. 마지막으로 논항 인식 단계에서는 <표 6>에서 언급한 변수를 통해 시간, 행위주, 대상, 장소, 도착점, 결과상태, 출발점, 목적으로 분류된 논항으로 인식한다. 논항 인식 결과에 따라 시스템 응답에 제공할 데이터가 달라진다.

3.4.3 응답 생성

응답 생성 단계에서는 의미 분석 결과를 토대로 시스템 응답을 생성한다. 시스템 응답은 답변 문장과 영화 관련 데이터로 이루어진다. 응답 생성 단계는 응답 데이터 검색, 답변 문장 생성, 사용자에게 시스템 응답 전달, 시스템 응답 시각화로 이루어진다.

먼저 영화 관련 데이터를 답변 문장과 함께 제공하기 위해 현재 상영작, 상영 예정작, 상영 시간표, 영화관 정보 데이터베이스를 구축하였다. 이 데이터는 매일 업데이트된다. 구축한 데이터베이스에서 답변에 해당하는 데이터를 검색한다.

의미 분석에서 추출한 키워드와 검색한 영화 데이터를 이용해 답변 문장을 생성한다. 이때, 데이터 검색 결과가 없다면, “해당 검색 결과가 없다.”는 내용의 문장을 답변으로 생성한다.

답변 문장 생성까지 완료되면, 영화데이터와 답변 문장을 챗봇을 통해 사용자에게 전달한다. 시스템 응답을 받은 사용자는 시각화된 응답 데이터를 보게 된다. 모바일 환경에서 효과적으로 응답을 전달하기 위해서 응답 데이터 종류에 따라 화면 구성을 달리하였다. 다양한 화면 구성은 후면 (그림 1)에서 확인할 수 있다.

4. 서비스 구조

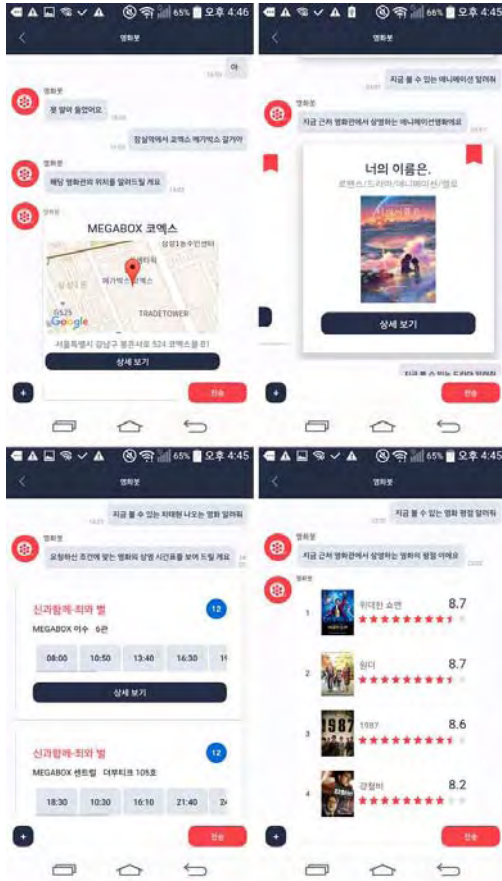
본 서비스는 서버와 안드로이드 애플리케이션으로 구성된다. 서버는 LINUX 운영체제(Ubuntu 16.04) 환경에서 웹서버 Nginx 1.10.3, RDBMS PostgreSQL 9.5.7, NoSQL MongoDB 3.4.10을 사용했으며, 서버 언어로는 PHP 7.0.2, Python 3.5.2, JAVA를 사용하였다. [15, 16, 17, 18] 채팅 서버는 오픈소스 프레임워크인 Netty 4.1을 이용해 구축하였다. [19] 서비스 구조도는 후면 (그림 2)와 같다.

5. 결론

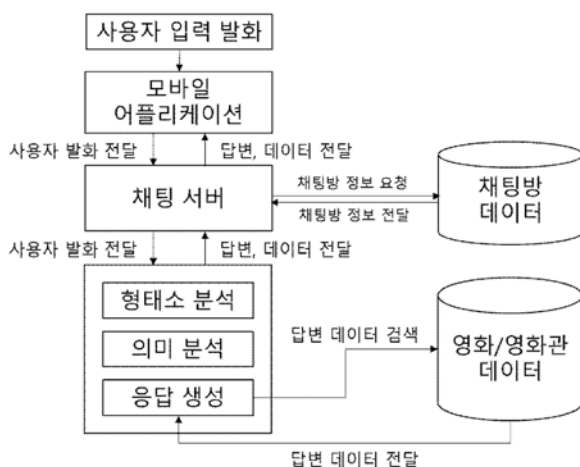
본 논문에서는 규칙 기반 대화 시스템을 활용하여 챗봇을 활용한 영화정보 서비스를 개발하였다. 외부 챗봇 API를 사용하지 않고 사용자 입력 발화에 따라 형태소 분석, 의미 분석, 응답 생성 과정을 거쳐 답변할 수 있도록 직접 개발 하였다. 이를 통해 사용자는 채팅 인터페이스에 질문을 작성하기만 하면, 원하는 영화 정보가 포함된 시스

템 응답을 받을 수 있다.

하지만 규칙 기반 대화 시스템은 규칙에 어긋난 새로운 발화 문장에 대해서 제대로 응답할 수 없다는 문제점이 있다. 따라서 충분한 학습 데이터를 구축한 후에 규칙 기반이 아닌 학습을 통해 답변을 생성할 수 있는 생성 기반 대화 시스템으로 발전시키고자 한다. [10, 11, 12]



(그림 1) 응답 데이터 종류에 따른 화면



(그림 2) 시스템 구조도

참고문헌

- [1] 민수진, 김현진, 송근혜 “통합기술수용이론(UTAUT)을 이용한 챗봇(chatbot)의 수용 결정요인에 대한 탐색적 연구” 한국기술혁신학회 2017년도 춘계학술대회 논문집
- [2] 강희주, 김승인 “메신저 기반의 모바일 챗봇 서비스 사용자 경험 평가 -구글(Allo)과 페이스북(M messenger)을 중심으로-” Journal of the Korea Convergence Society Vol. 8. No. 9
- [3] 황금하, 권오욱, 이경순, 김영길 “목적 지향 대화 시스템을 위한 챗봇 연구” KIPS Tr. Software and Data Eng. Vol.6, No.11
- [4] 홍금원, 이정훈, 신중휘, 이도길, 임해창 “대화시스템의 로그를 이용한 대화예제의 자동 확충에 관한 연구” 한국 HCI학회 2009년도 학술대회
- [5] 홍금원, 이연수, 김민정, 이승욱, 이주영, 임해창 “한국어 모바일 대화형 에이전트 시스템” 韓國 컴퓨터情報學會 論文誌 第13卷 第6號, 2008. 11.
- [6] 김주영, 한소희, 박현민 “카카오톡 챗봇 API를 이용한 요리정보 제공 서비스 개발 (카톡 API 사용)” 한국정보처리학회 2017년 추계학술발표대회 논문집 제24권 제2호
- [7] 최승권, 권오욱, 이기영, 노윤형, 황금하, 김영길 “챗봇과 대화시스템을 이용한 영어 교육 시스템” 한국정보처리학회 2017년 춘계학술발표대회 논문집 제24권 제1호
- [8] 김계영, 이현동, 조대수 “룰 엔진 기반의 음성 인식 챗봇 IFTTT 서비스 시스템” 한국정보처리학회 2017년 추계학술발표대회 논문집 제24권 제2호 (2017. 11)
- [9] 강정석, 권성재, 서원진, 장소진 “챗봇을 활용한 전문 의료 상담 및 예약플랫폼 “헬스챗”” 한국정보처리학회 2017년 추계학술발표대회 논문집 제24권 제2호 (2017. 11)
- [10] 양민철, 이연수, 임해창 “예제 기반 챗봇을 위한 기계 학습 기반의 발화 간 유사도 측정 방법” Journal of the Korea Academia-Industrial cooperation Society Vol. 11, No. 8
- [11] 이창기, 임수중, 김현기 “Structural SVM 기반의 한국어 의미역 결정” Journal of KIISE, Vol.42, No.2
- [12] 오교중, 박찬용, 이동건, 임채균, 최호진 “RNN 문장 임베딩과 ELM 알고리즘을 이용한 금융 도메인 고객상담 대화 도메인 및 화행분류 방법” 제29회 한글 및 한국어 정보처리 학술대회 논문집 (2017년)
- [13] “KoNLPy”, <http://konlpy-ko.readthedocs.io/ko/v0.4.3/>
- [14] “NLTK 3.2.5”, <http://www.nltk.org/>
- [15] “Ubuntu 16.04”, <https://www.ubuntu.com/>
- [16] “Nginx 1.10.3”, <https://nginx.org/en/>
- [17] “PostgreSQL 9.5.7”, <https://www.postgresql.org/>
- [18] “MongoDB 3.4.10”, <https://www.mongodb.com/>
- [19] “Netty 4.1”, <https://netty.io/>