

사용자 질의 자가학습형 인공지능 챗봇 시스템

박성현, 홍석훈, 황수현, 나스리디노프 아지즈, 류관희, 홍장의*
*충북대학교 소프트웨어학과
e-mail : jehong@chungbuk.ac.kr

A Study of Artificial Chatbot System for User Query Self-Learning

Seong-Hyeon Park, Seok-Hun Hong, Su-Hyeon Hwang, Aziz Nasridinov,
Kwan Hee Yoo, Jang-Eui Hong*
*Dept of Computer Science, Chungbuk National University

요 약

인공지능에 대한 연구가 최근 이슈가 되면서, 딥러닝 기술의 비약적인 발전 덕분에 대화형 에이전트가 인터페이스의 역할을 하고 있다. 이 중에서 최근 여러 대학에서 서비스로 지원하는 챗봇 시스템의 문제점에 대하여 개선된 시스템을 제안하고, 이를 구현하여 실험을 통해 연구하고자 한다. 기존 챗봇 시스템이 가진 문제점을 보완한 시스템은 서비스 사용자가 질의하는 의도에 더 알맞은 응답을 제공하여 서비스 사용자의 불편함을 최소화하고, 사용성과 편의성을 최대화 하는 것을 목적으로 한다.

1. 서론

최근 ‘4차산업혁명’, ‘알파고 쇼크’ 등 인공지능과 관련된 기술들이 발표되면서 인공지능에 대한 많은 관심과 혁신이 일어나고 있다. 특히 관심을 끄는 주요 기술은 딥러닝 기술이다. 이러한 딥러닝 기술이 발전하게 되면서 동시에 대화형 에이전트 시스템도 같이 빠르게 발전하였다.

대화형 에이전트(Conversational Agent)란 음성 또는 문자를 통하여 인간과 기계가 정보를 주고받는 시스템을 의미한다. 초창기 컴퓨터공학이 태동하면서 동시대적으로 연구된 분야이지만 최근 음성 인식과 처리 효율성의 비약적인 발전으로 실용적인 서비스 제공단계에 접어들고 있다. 최초의 대화형 시스템은 1966년 독일계 미국인 컴퓨터 과학자 Joseph Weizenbaum이 이끄는 MIT Artificial Intelligence Laboratory에 의해 개발된 자연어 처리 컴퓨터 프로그램인 ELIZA이다. 인간과 기계사이의 의사소통 피상성을 보여주기 위해 만들어졌다. 패턴 매칭과 대체 방법론을 사용하여 대화를 구성하였고, 이후 다수의 대화형 에이전트가 개발되었지만 인공지능의 연구와 깊은 관련이 있는 대화형 에이전트 기술은 최근 인공지능 기술의 급격한 발전으로 Siri, Contana, Alexa와 같은 서비스가 보급되면서 인간과 시스템 사이의 인터페이스로 자리를 잡고 있다[1].

대화형 에이전트의 분야 중에서 챗봇(chat bot)은 채팅과 로봇의 합성어로, 사람과의 문자 대화를 통해서 사람의 질의에 대해서 답변을 제공하는 대화형 인터페이스를 말한다. 기존의 챗봇 시스템은 특정 주제에 대해 미리 만들어 놓은 답변으로 대답하는 규칙 기반의 시스템이다. 답변을 자동으로 생성하여 사람의 질의 의도에 더 가까운 답변을 생성하게 해주는 것이 인공지능 기반 시스템이다[2].

챗봇의 활용 분야가 넓어지고 있으며, 그 기술 수준 또한 날이 갈수록 향상 되어가고 있다. 하지만 아직까지 챗봇이 사용자의 질의에 대하여 학습해서 추후에 다시 응답이 가능하도록 하는 시스템에 대해선 미비한 상태이다. 이러한 문제점에서 착안하여 챗봇이 사용자 질의에 대해 학습하여 능동적인 답변을 제공할 수 있는 챗봇 시스템에 관한 연구를 진행하였다.

2. 기존 챗봇 관련 연구

한글을 지원하는 챗봇의 경우, 한국어 자연어처리를 위해 형태소 분석과 CNN, RNN 등 인공 신경망을 이용하거나 Naivebayes, Hidden Markov Model 등의 분류기를 사용하기도 한다[3]. 현재 국내에는 다양한 챗봇이 존재하며, 특정 주제에 대한 챗봇 질의응답 서비스를 제공 중이다. 최근 국내 대학들 중 챗봇을 활용한 서비스를 도입하는 대학들이 늘고 있으며, 보통 교내 식당의 식단을 알려주거나, 도서관 좌석 정보 등을 서비스한다. 기업들도 챗봇을 적극적으로 도입하여 서비스 중이며 챗봇에 대해 사용자들이 익숙해지면서 챗봇 시스템의 사용은 계속해서 활발해지고, 이를 통한 시장도 성장할 것으로 전망된다.

그럼에도 불구하고 아직까지는 응답 가능한 질의에 대해서만 서비스를 하고, 응답불가능한 질의의 경우에는 대기 중인 상담원을 연결해주거나, 응답할 수 없음을 사용자에게 보낸다. 아직까지 국내 챗봇의 경우에도 이 수준에서 크게 벗어나지 못하였기 때문에 본 논문에서는 이와 같은 문제점을 해결할 수 있는 챗봇 시스템을 제안하고자 한다. 위 시스템을 사용자에게 제공한다면 높은 편의성과 사용성을 제공할 수 있을 것으로 기대된다.

3. 제안하는 학습형 인공지능 챗봇 시스템

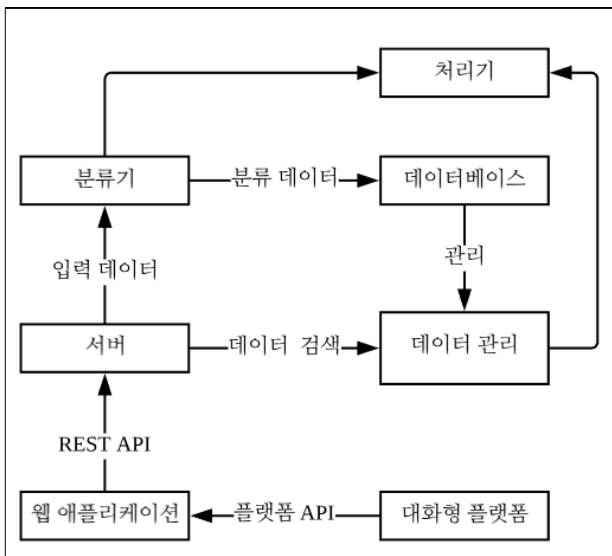
3.1. 기존 시스템의 문제

기존의 챗봇 시스템은 사전에 준비된 키워드가 아닌 경우에는 사용자의 질의에 답변할 수 없다는 문제가 있다. 특정 주제를 다루는 챗봇 시스템은 특정 주제에 대한 키워드를 미리 테이블에 저장한다. 그러나 특정 주제를 칭하는 단어는 비슷한 의미를 가지면서도 전혀 다른 형태를 지니는 경우가 굉장히 많다. 예컨대, ‘학교 식당’과 ‘식당’, ‘학식’, ‘밥집’, ‘음식점’, ‘레스토랑’은 모두 다른 형태를 지니지만 같은 의미를 지니고 있다. 즉, 챗봇의 개발자가 생각하지 못한 키워드에 대해서 사용자가 질의를 요청한 경우는 시스템 내에서 처리할 수 있는 질의임에도 불구하고, 알아듣지 못하여 지원할 수 없는 기능이라고 답변해야 하는 것이다. 한 가지의 키워드도 동의어가 많은데, 데이터 셋(Data Set)이 매우 많은 챗봇 시스템의 경우에는 준비된 기능들이 사용자의 요청에 전부 대응하지 못할 수 있다는 단점이 있다. 본 논문이 제안하는 시스템으로 이러한 단점을 보완하고자 한다.

기존 챗봇의 문제점을 해결하기 위해서는 자발적으로 학습을 실시할 수 있어야 한다. 본 논문에서는 이러한 학습 방식을 동적 학습 훈련이라고 칭한다. 기존의 챗봇 시스템은 사전에 정의된 키워드 학습 모델에 의존적이므로 정적인 학습 모델을 사용한다고 할 수 있다. 이와 반대로 제안하는 챗봇은 사전 학습 모델을 사용 후 계속해서 스스로 예외 키워드에 대한 학습으로 사용자 의도에 부합하고 높은 만족도를 얻을 답변을 제공하는 것이다.

3.3. 제안하는 시스템의 설계 및 구현

본 논문의 시스템은 파이썬 언어 기반의 분류기 모델을 이용하며, 카카오톡 API를 통해서 구현한다. 시스템의 구성도는 다음에 나오는 그림1과 같다.



(그림 1) 시스템 구성도

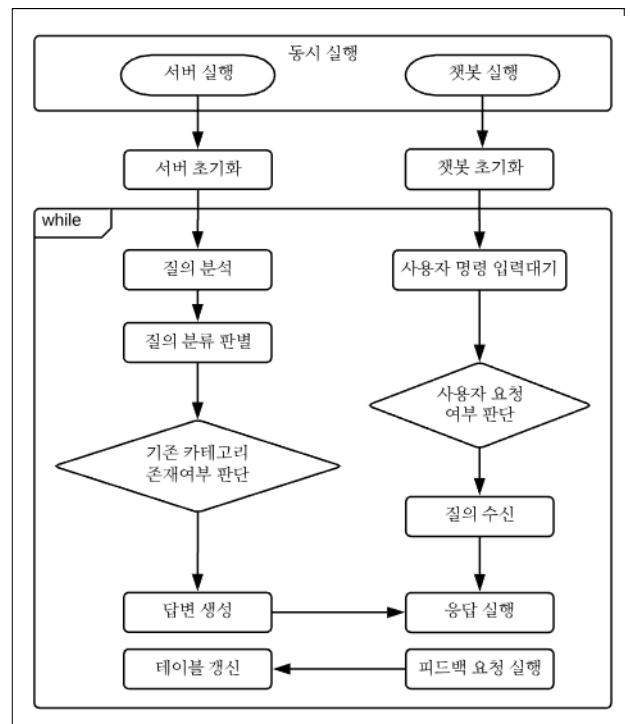
구현한 형태는 그림 1에서 보는 바와 같이 카카오톡의 플러스친구를 통해서 유저 인터페이스를 제공하고 있다. 챗

봇은 응답 후 피드백을 사용자에게 요청하고, 사용자는 챗봇의 피드백 요청에 대해 질의를 전송하게 된다. 사용자의 질의가 전송되면 서버와 분류기를 통해서 질의에 대한 데이터가 전송된다. 데이터베이스와 데이터 관리를 거친 데이터들은 최종적으로 처리기에 도달하여 사용자의 질의에 대한 응답을 하게 된다.

챗봇 시스템에서 질의를 처리하는 분류기는 나이브 베이즈 분류 모델을 적용한다. 나이브 베이즈 분류 모델은 베이즈 정리를 기반으로 스팸 메일이나 문서 분류하는데 사용되는 분류기이다. 여기서는 분류기는 학습시키기 위한 말뭉치 적재 모듈, 말뭉치에서 단어를 분리하고 단어 개수를 세는 모듈, 단어 개수를 통해 가중치를 계산하는 모듈, 각 단어의 가중치를 저장하는 모듈과 이 학습과정을 통제하는 학습모듈로 이루어져 있고, 학습된 정보를 통해 분류하는 분류모듈이 있다. 분류모델이 학습하는 말뭉치 테이블은 문장과 그 문장의 분류 값으로 구성되어 있다. 따라서 질의가 분류기의 분류모듈을 거치면 결과 값으로 질의가 어떤 분류인지 분류값을 반환한다.

3.3. 학습형 인공지능 챗봇의 알고리즘

제안하는 챗봇의 알고리즘은 다음에 나오는 (그림 2)와 같으며 핵심적인 동적 학습과 관련된 알고리즘은 (그림 3)에서 의사코드로 표현했다. 사용자가 챗봇을 사용함과 동시에 서버와 챗봇은 초기화 상태가 된다. 이후에 사용자의 질의를 분석하여 카테고리를 판별하고, 기존 카테고리에 존재하는 것인지에 대한 판단한다. 존재하면, 답변을 생성하고 사용자 질의에 대하여 응답한다. 카테고리에 존재하지 않는 질의에 대해서는 피드백을 요청하여 기존 테이블을 갱신하여 추후에 사용자 질의와 많은 연관성을 가진 응답을 하도록 한다.



(그림 2) 제안하는 챗봇 시스템 알고리즘

```

IF Synonym_Array == NULL
    Synonym_Array = synonym_creator( );
existence = SynonymExistenceChecker( unknown_word );
IF existence == TRUE
    inTable = keywords_existence_checker( keywords );
IF inTable == TRUE
    return_word = unknown_word;
ELSE
    return_word = 'ERROR'
IF return_word != 'ERROR'
    csv_updater( return_word );
    trainClassifier( return_word );
    
```

(그림 3) 챗봇의 동적학습 알고리즘 의사코드

(그림 3)과 같은 흐름은 단어가 분류가 안 되는 경우에도 오는 흐름으로 의미망에서 동의어 리스트를 가져와서 분류되지 않는 단어와 비교하고, 있다면 동의어와 의미가 같은 키워드를 검사해서 없다면 에러, 있다면 동적으로 학습시킨 후 테이블과 데이터베이스를 갱신한다.

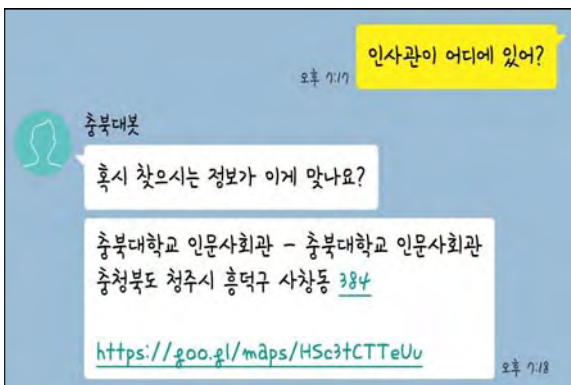
4. 실험과 결과 데이터

본 논문에서는 구현한 시스템이 올바르게 작동하는지 확인하기 위한 실험을 진행하고, 그 실험 데이터와 결과로 나타나는 데이터를 다음과 같이 나타낸다. 일반적인 사용자들의 질의에 대한 것들을 입력 데이터셋으로 구성하여 챗봇이 사용자의 질의에 대해 응답할 수 있도록 하였다.

<표 1> 입력 데이터 세트 예시(Data Set)

질의	분리	핵심키워드	보조키워드
N건물	N + 건물	건물	N
S건물	S + 건물	건물	S
E건물	E + 건물	건물	E
식단	식단		식단
도서관 위치	도서관+위치	위치	도서관
근처 화장실	근처 + 화장실	근처	화장실

여기서 더 나아가 본 논문에서 제시한 시스템에 의해 기존 테이블에 존재하지 않는 카테고리에 대해 응답하는 것에 대한 예시를 (그림 4)에서 확인 할 수 있다.



(그림 4) 카테고리에 없던 질의에 대한 응답 화면

5. 시스템 구현 환경 및 사용 기술

본 논문에서 제안한 시스템의 구현 환경과 사용기술은 사용자에게 보다 편안한 접근환경을 제공하기 위하여 대화형 채팅 플랫폼을 사용한다. 시스템 구현은 Linux 계열의 운영체제인 Ubuntu 16.04 상에서 진행 되었으며, Python 언어를 사용하였다. 마이크로프레임워크인 Flask를 이용하여 웹 서버를 구축하고, AWS EC2를 사용하여 할당받은 IP를 사용하여 웹서버에 접속이 가능하도록 구현하였다. AWS EC2 상에서 구현하기 위해 PuTTY를 통해 접속한 후 시스템 구축을 하였다. GitHub를 이용하여 시스템 버전관리를 진행 하였으며, 업데이트가 이루어 질 때마다 업로드 할 예정이다.

6. 결론 및 향후 계획

본 논문에서는 기존 챗봇 시스템이 가지는 한계점이 챗봇의 정적인 학습 모델에 있다는 것에 착안하여 동적으로 사용자의 질의에 대해 학습을 진행하는 시스템을 설계하고 구현해보았다. 기존에 존재하던 질의에 대한 응답뿐만 아니라 동적인 학습을 통해 최대한 사용자가 원하는 정보를 얻을 수 있도록 하였다. 아직은 많은 부분을 보완하고, 사용자 관점에서 더욱 편하게 느낄 수 있도록 개선해야할 부분이 많이 남아있는 상태이다.

향후에는 특정 주제만이 아닌 열린 주제에 대해서 즉, 일상생활에서 도움을 줄 수 있을만한 정도의 데이터셋을 어떻게 학습시키는 것이 효율적인지, 그리고 동적으로 학습하기 위한 방식을 다른 방식으로의 구현가능 여부와 효율적으로 탐색할 수 있는 방안에 대해 연구할 예정이다.

ACKNOWLEDGEMENT

“본 논문은 교육부가 지원하고 충북대학교가 수행하는 지역선도대학육성사업의 지원을 받아서 수행되었습니다.”

참고문헌

[1] 박동아, “인공지능 기반 대화형 공공 행정 챗봇 서비스에 관한 연구”, 멀티미디어학회 논문지, 2017.8, 1347p
 [2] 임형준 외, “지능검색 챗봇 시스템을 위한 대화형 인터페이스”, 한국정보과학회, 2017.12 pp 1665-1665
 [3] 이세훈 외, “입시 상담 챗봇 설계”, 한국컴퓨터정보학회, 2017.7, pp 430-431
 [4][1] 박동아, “인공지능 기반 대화형 공공 행정 챗봇 서비스에 관한 연구”, 멀티미디어학회 논문지, 2017. 8, 1347p
 [5] 박현길, “나를 읽는다 챗봇(Chatbot)!” , 마케팅, 51(5), pp 40-50
 [6] 서교리 외, “인공지능 기반의 공공 지능형 가상 비서 서비스 발전 모델 연구”, 한국통신학회, 2017.6 pp 890-891