

# A Chatter Bot for a Task-Oriented Dialogue System

Jin-Xia Huang<sup>†</sup> · Oh-Woog Kwon<sup>\*\*</sup> · Kyung-Soon Lee<sup>\*\*\*</sup> · Young-Kil Kim<sup>\*\*\*\*</sup>

## ABSTRACT

Chatter bots are normally used in task-oriented dialogue systems to support free conversations. However, there is not much research on how chatter bots as auxiliary system should be different from independent ones. In this paper, we have developed a chatter bot for a dialogue-based computer assisted language learning (DB-CALL) system. We compared the chatter bot in two different cases: as an independent bot, and as an auxiliary system. The results showed that, the chatter bot as an auxiliary system showed much lower satisfaction than the independent one. A discussion is held about the difference between an auxiliary chatter bot and an independent bot. In addition, we evaluated a search-based chatter bot and a deep learning based chatter bot. The advantages and disadvantages of both methods are discussed.

**Keywords :** Dialogue System, Task-Oriented Dialogue System, Chatter Bot, Computer-Assisted Language Learning System

## 목적지향 대화 시스템을 위한 챗봇 연구

황금하<sup>†</sup> · 권오욱<sup>\*\*</sup> · 이경순<sup>\*\*\*</sup> · 김영길<sup>\*\*\*\*</sup>

### 요약

목적 지향 대화 시스템에서 자유대화를 지원하기 위해 챗봇이 활용되고 있다. 그러나 목적지향 대화시스템을 위한 챗봇과 독립 챗봇에 대한 사용자 기대와 평가가 같은지에 대한 연구는 거의 없는 상황이다. 본 논문에서는 목적지향 대화시스템으로 구현한 영어 교육용 대화시스템에서, 대화의 자유도를 높이기 위하여 주제의 사용자 발화를 허용하고, 이에 대응하기 위한 챗봇을 개발하였다. 독립 챗봇과 보조 시스템으로서의 챗봇에 대하여 비교 평가함으로써, 서로 다른 시스템에 대한 사용자의 서로 다른 기대를 살펴보았다. 또한 검색 기반 챗봇과 신경망 기술을 이용한 생성 기반 챗봇에 대한 비교 평가를 통해 이들의 장단점과 향후 활용 방안에 대하여 살펴보았다.

**키워드 :** 대화 시스템, 목적지향 대화 시스템, 챗봇, 언어 교육 시스템

### 1. 서론

대화 처리 시스템은 그 목표하는 바에 따라, 목적지향 대화시스템(task-oriented dialog system, goal-oriented dialog system)과 재미를 위한 대화시스템으로 나뉘며, 후자를 채팅용 로봇의 약어로 챗봇(chatter bot, chatbot)이라고 한다. 목적지향 대화시스템은 시스템과의 대화를 통해 일정 관리, 호

텔/식당/항공권 등의 예약, 음악 듣기 등 명령 전달, 콜센터 상담 등과 같은 특정된 목적을 달성해야 한다. 이와 달리, 챗봇은 사람과의 대화에서 시스템이 가능한 인간다운 대화를 하는데 그 목적을 둔다. Apple Siri[1], Amazon의 Alexa[2], SoundHound의 Hound[3] 등과 같이, 스마트 폰 또는 스마트 스피커를 이용한 가상 개인비서 서비스에서는 목적지향 대화 시스템을 챗봇과 결합하여 특정된 목적을 달성하면서 일반 대화도 수행 가능하도록 하였다. 이런 가상 개인비서 서비스는 향후 몇 년 사이 가장 주목할 인공지능 기술 중의 하나로 간주되고 있다[4]. 목적지향 대화시스템과 챗봇의 결합이 추세로 되고 있지만, 목적지향 대화시스템을 위한 챗봇에 대한 사용자 기대와 평가가 독립적인 챗봇과 어떻게 다른지에 대한 연구는 아직 활발하게 이루어지지 않고 있다.

본 논문에서는 목적지향 대화시스템을 이용하여 구현한 영어 교육시스템에서 사용자가 주제 밖의 대화를 시도할 때

\* 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(R0126-15-1117, 언어학습을 위한 자유대화형 음성대화 처리 원천기술 개발)과 한국전자통신연구원 연구운영비지원사업의 일환(17ZS1210, 언어장벽없는 국가구원을 위한 자동통번역산업 경쟁력 강화사업)으로 수행된 연구임.

† 정 회 원 : 한국전자통신연구원 언어지능연구그룹 선임연구원

\*\* 비 회 원 : 한국전자통신연구원 언어지능연구그룹 책임연구원

\*\*\* 정 회 원 : 전북대학교 컴퓨터공학부 교수

\*\*\*\* 비 회 원 : 한국전자통신연구원 언어지능연구그룹 그룹장

Manuscript Received: August 1, 2017

Accepted: August 22, 2017

\* Corresponding Author: Jin-Xia Huang(hgh@etri.re.kr)

시스템 응답을 담당하는 챗봇을 개발하였다. 또한 이런 보조용 챗봇과 독립된 챗봇에 대하여 사용자 평가를 진행하여 그 차이를 고찰하고자 하였다. 이외 검색 기반 챗봇과 신경망 기술을 이용한 생성 기반 챗봇에 대한 비교를 통해 그 장단점을 살펴보았다.

## 2. 관련 연구

### 2.1 목적 지향 대화 시스템

목적 지향 대화 시스템에서 핵심 모듈은 대화 이해 모듈과 대화 관리 모듈이다. 대화 이해에서는 사용자 발화 의도를 이해해야 하며, 대화 관리에서는 사용자 발화 의도와 현재의 대화 상태(state)에 따라 목적 달성 위한 시스템 발화를 결정한다. 대화 관리는 크게 유한상태 기계(finite-state machine) 기반 방법[5]과 양식기반(frame-based) 방법[6]으로 나뉜다[4]. 유한상태 기계 기반 방법에서는 시스템 주도로 대화를 진행한다. 목적 달성까지 필요한 상태를 이미 정의한 순서대로 발화되도록 사용자 응답이 적합해야만 다음 상태로 넘어간다. 예를 들면, 항공권 예약 위하여 우선 출발지를 물어보고, 다음 목적지, 날짜와 시간, 마지막으로 왕복 여부를 확인하는, 순차적인 시스템 주도 대화로 목적을 달성한다. 유한상태 기계 기반 방법은 구현하기 쉽고, 성능이 안정적이며, 단순한 목적에 적합한 장점이 있다. 단점은 인간 사이의 양방향 대화에 가깝기 보다는 사용자의 직접적인 대답을 요구하는 일방적이고 제한된 대화만 가능하다.

양식기반 대화 관리는 양식 채우기(form-filling) 방식이라고도 하는데, 목적 달성 위한 필수 정보를 슬롯(slot)으로 정의하고, 여러 슬롯으로 구성된 양식(form)이 채워질 때까지 대화를 통해 사용자에게 정보를 요구한다. 주로 시스템 주도형으로 대화를 진행하지만, 사용자가 그 대화 순서에서 일정한 자유도를 누릴 수 있다. 예를 들면 항공권 예약에서, 사용자가 출발지와 목적지 및 왕복 여부를 동시에 발화하면, 시스템은 언어 분석 및 슬롯 인식을 통하여 서식을 채워 나가고, 추가로 필요한 날짜와 시간을 사용자에게 물어보게 된다. 대화 자유도가 상대적으로 높은 장점이 있는 대신, 대화 이해 정확도에 대한 요구가 높다.

### 2.2 챗봇

챗봇은 사람 사이의 대화를 모방하여 재미있는 대화서비스를 제공하는 것을 목적으로 하는 대화 시스템이다. 대화 처리에서 사용하는 방법으로는 규칙 및 패턴 기반 방법[7], 검색 기반 방법[8-11], 생성 기반 방법[12, 13]으로 크게 나뉠 수 있다. 여기에서 생성기반 방법은 통계기반 자동번역 또는 신경망 기반 자동번역 방법을 주로 사용한다. 패턴 기반 방법에서 일부 연구에서는 사용자 의도 이해를 필요로 하기도 하지만[14] 대부분 엄격한 자연언어 분석이나 사용자

의도 이해과정을 필요로 하지 않는다. 사용하는 패턴 포맷은 시스템에 따라 다양하지만, 사용자 발화를 기 구축 패턴과 매칭 하여 대응되는 시스템 발화를 출력하는 방법은 동일하다. 패턴 지식 외에 대부분 제약 규칙과 유의어 및 관련어 지식을 함께 사용한다.

최근에는 대용량 대화 코퍼스의 활용이 가능해짐에 따라 실제로 사용되고 있는 대부분의 챗봇은 규칙 및 패턴기반 방법을 검색 기반 방법과 함께 사용하고 있으며, 복잡한 지식 구조를 표현하기 위해 스크립트 언어(script language)를 사용하는 챗봇이 많다[7, 15].

최근 활발한 연구가 진행되고 있는 신경망 기반 대화 생성 방법은 어떤 입력에 대해서도 출력을 생성할 수 있는 장점이 있다. 다만 생성 응답이 짧고 단순하며[13], 동일한 사용자 발화에 항상 동일한 시스템 응답을 출력하는 등, 재미를 요하는 챗봇 대화에 맞지 않는 단점도 있다.

## 3. 목적지향 대화시스템을 이용한 영어 교육시스템

본 연구의 선행 연구로, 목적지향 대화처리 기술을 이용한 외국어 교육 시스템에 대한 연구와 내비게이션을 위한 한국어 정보서비스에 대한 연구가 수행되었다[4, 16, 17]. 본 연구는 실험의 편의를 위해 이중 대화기반 영어 교육 시스템을 대상으로 진행하였다. 이 대화 시스템은 양식기반 대화 관리 방법을 사용하고 있으며, 하나의 태스크(task)를 여러 개의 작은 태스크로 나누었다. 작은 태스크들은 순서에 무관한 경우, 순차적으로 실행해야 하는 경우, 또한 건너뛰어도 되는 경우로 나누어 대화맵의 형식으로 구축하였다[18]. 예를 들면 음식 주문 주제(태스크)는 “[인사] > 메인메뉴 선택, 사이드메뉴 선택 > 계산 > [인사]”의 소주제로 나눌 수 있고, 여기에서 계산은 메뉴 선택이 끝나야 가능하지만, 메인메뉴 선택과 사이드메뉴 선택은 그 순서가 자유로우며, 인사는 건너뛰어도 된다.

대화이해 모듈에서 사용자 발화가 해당 상태에서의 적절한 발화로 인식되면 시스템 대화는 다음 상태로 넘어가는 반면, 적절한 발화가 아니라고 인식되는 경우 실패 피드백을 제공하거나, 종전의 시스템 발화를 반복하여 사용자가 다시 발화하도록 한다.

다음은 “city tour” 중 “downtown tour” 주제에 관한 대화 시나리오 중의 일부이다.

System: How many tickets for you today?

User: Four tickets please.

System: Your total comes to 160 dollars.

User: Here you go.

위 대화에서 사용자가 시스템이 금액 제시할 때 “Here you go.” 대신 “I have no money.”과 같이 목적 달성과 무

관한 주제의 발화할 경우 시스템은 사용자 발화를 무시하고 재발화를 요구한다.

System: How many tickets for you today?  
 User: I need 4 tickets.  
 System: Your total comes to 160 dollars.  
 User: I have no money.  
 System: Your total comes to 160 dollars.  
 User: Here you go.

본 영어 교육 시스템을 30명의 영어 학습자에게 사용하도록 하고, 사용자당 2-4개의 주제에 대하여 대화를 완성하게 한 다음 진행한 사용자 요구사항 조사에서, 주제 밖의 자유로운 대화 기능을 요구하는 사용자는 66%로, 필요 없다고 답한 사용자인 23%보다 훨씬 높게 나타났다. 이런 사용자 요구에 응하기 위해 본 연구에서는 챗봇을 도입하기로 하였다.

#### 4. 목적지향 대화 시스템을 위한 챗봇 시스템

##### 4.1 챗봇 지식

챗봇 지식은 주로 대용량의 대화 예문으로 구성된다. 대화 예문은 발화와 이에 대응하는 응답으로 구성되며 이러한 대화쌍 또는 한 대화턴(turn)이라고 한다. 대화 예문은 텍스트와 패턴 두 가지 형식으로 기술 가능하며 일반 분야에서는 가능한 텍스트로 된 대화 예문만 사용하고, 특정 주제에서 필요한 경우에만 패턴을 사용하였다(Table 1). 본 연구에서는 텍스트 형식의 대화 예문을 주로 사용하였는데, 이는 패턴 구축을 위한 사람의 수작업을 최소화할 뿐만 아니라, 패턴 매칭을 위한 슬롯 인식, 응답문 생성 위한 슬롯 생성 과정을 최소화하기 위함이다.

Table 1. Knowledge Base for Chatbot

Type	Utterance	Response	Topic
Example	How old are you?	One year old.	General
Example	How old are you?	Older than you.	General
Example	How old are you?	I am <system.age> years old.	Ordering food
Profile	system:age	23	Ordering food
Profile	system:job	waiter	Ordering food
Profile	system:job	pollster	Future currency
Example	Have you bought anything with your <cellphone>?	I bought <items> with my <cellphone>	Future currency
Synonym	cellphone	cell phone, smartphone, mobile	Future currency

챗봇 지식에는 이 외에도 시스템 개인정보와 유의어 등 지식이 있고, 주제가 표기된다. 특정 주제 지식은 해당 주제 대화에서만 사용되는 반면 일반 분야 지식은 특정 주제와 일반 대화에서 모두 사용된다. 동일한 지식이 일반분야와 특정주제에 모두 있는 경우, 해당 주제 지식 우선으로 채택한다.

특정주제 대화지식에는 대화시스템이 담당하는 역할(예: 직업, 나이, 성별)과 관련된 개인정보 지식, 감사 칭찬 불만 등 고빈도 일상 대화 예문도 포함시켰다. 이런 예문은 일반 분야 대화 코퍼스에도 대부분 포함되어 있으나, 해당 발화들이 주제 대화의 흐름과 맞지 않는 경우가 발생할 수 있기에 특정주제 위해 따로 구축한다.

일반분야 대화예문은 총 41만 쌍을 포함하고 있다. 이중 1.8만 턴은 사람이 챗봇과의 대화를 가정하고 챗봇 응답을 구축한 것이고, 영화 대본 데이터인 MovieDic[19]에서 긴 발화나 인명 등 고유명사를 많이 포함한 대화를 필터링하여 23.7만 대화쌍을 추출하였으며, 여행 등 다양한 상황에서의 영어대화 문장 15.5만 쌍을 추가로 포함시켰다.

이외, 신경망 기반 챗봇에 대한 추가 평가 실험을 위해, 앞에서 설명한 일반분야 대화 41만 쌍 외에, Movie-Dic[19] 대화 37.7만 쌍, BNC 코퍼스[20] 대화 45.4만 쌍과 트위터 및 드라마 대본 등에서 추출한 대화 19.6만 쌍을 추가하여 총 140만 쌍의 대용량 코퍼스를 추가로 구성되었다.

##### 4.2 검색 기반 챗봇 시스템

목적지향 대화 시스템의 대화 관리 모듈에서 사용자 의도 인식 실패 시, 즉 해당 주제에서는 “모름”으로 분류된 사용자 발화는 챗봇에 입력된다. 이때 사용자 발화에 슬롯이 있을 수도 있으며, 슬롯이 없거나 독립 챗봇으로 사용되는 경우, 일반 텍스트로 된 사용자 발화가 챗봇으로 입력된다(Fig. 1).

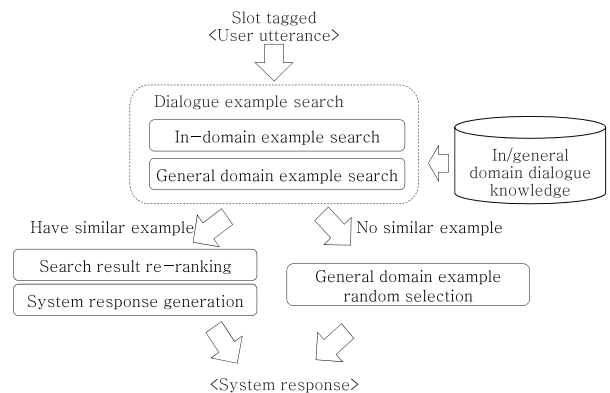


Fig. 1. System Architecture of the Search-Based Chatbot

검색기반 챗봇에서 사용자 발화는 질의로 간주되고, 대용량 대화 예문의 발화는 가상의 문서에 해당된다. 사용자 질의와 가장 유사한 발화가 검색되면, 이의 응답을 시스템 응

답으로 출력한다. 본 연구에서는 대화 예문의 색인 및 검색 위해 Indri 검색엔진[21]을 사용하였고, 검색 결과에 대하여 재순위를 통해 최적의 예문을 선택하였다. 주제대화 예문과 일반분야 대화 예문을 모두 검색하여, 유사도가 일정한 경계값 이상인 대화 예문이 없으면, 임의로 하나의 시스템 발화를 출력하였는데, 이는 사람의 대화에서 화제 전환에 해당된다고 가정하였다. 동일한 유사도를 가진 두 개 이상의 예문이 있는 경우, 임의로 그 중 하나를 발화한다.

일반 정보 검색에서 사용하는 TF-IDF기반 유사도 검색은 질의에서의 어휘가 가장 문서인 대화 예문에 많이 나타날수록 해당 문서가 더 상위 랭킹에 위치하게 되는데[22], 이는 가능한 정확한 매칭을 요구하는 대화 예문 검색에는 맞지 않다. 이런 문제를 해결하기 위해, 본 연구에서는 수정된 코사인 유사도를 사용하였다.

Equation (1)에서,  $q$ 는 질의 벡터,  $d$ 는 가상문서 벡터,  $s$ 는 이들의 유사도 점수,  $tf_q(x_i)$ 는 질의에서 어휘  $t_i$ 의 출현 빈도,  $tf_d(y_i)$ 는 가상문서에서의 어휘  $t_i$ 의 빈도이다.

$$s(q,d) = \frac{\sum_{i=1}^n \min(tf_q(x_i), tf_d(y_i))^2}{\sqrt{\sum_{i=1}^n tf_q(x_i)^2 \cdot \sum_{i=1}^n tf_d(y_i)^2}} \quad (1)$$

대화 예문 검색에서, 질의와 가상의 문서 모두 짧은 문장인 경우가 대부분이고 한 두 단어의 차이가 의미에서 큰 차이를 야기 할 수 있기에, 정확한 매칭을 위해 소문자처리를 하되, 불용어(stop-words)를 제거하지 않았고 어휘의 표층형(surface form)을 그대로 사용하였다.

#### 4.3 챗봇 이용한 영어 교육용 대화시스템

학습자가 선택 학습 주제에 대하여 시스템과 대화하던 중, 주제를 벗어나 발화할 경우, 끊기지 않고 대화하며 영어 학습을 계속 진행하기 위하여[15] 챗봇을 사용하되, 사용자가 학습 주제에서 오래 벗어나지 않고 학습 목적을 조기 달성하도록 도와주기 위해, 대화시스템은 챗봇 응답을 시스템 발화로 출력 한 다음, 해당 상태에서의 시스템 발화를 다시 출력하여, 학습자가 원 주제 대화로 돌아오도록 하였다.

Fig. 2는 챗봇을 탑재한 영어 교육 시스템과 사용자간 대화의 예이다. 사용자가 돈을 지불해야 하는 상태에서 “I have no money.”라고 발화하면, 대화 이해에서 주제의 발화로 분류되어 챗봇으로 입력된다. 검색기반 챗봇에서 “What a pity!”를 응답으로 받아 출력한 다음 다시 주제 대화로 돌아와 “Your total comes to 160 dollars.”를 재발화 하는 것을 볼 수 있다. 챗봇 응답은 오류를 포함할 수 있기에 사용자가 인지 할 수 있도록 빨간색으로 출력된다. “Four ticket please.”와 같이 문법오류가 있는 사용자 발화는 교육 시스템이 수정하여 파란색으로 발화 해준다.

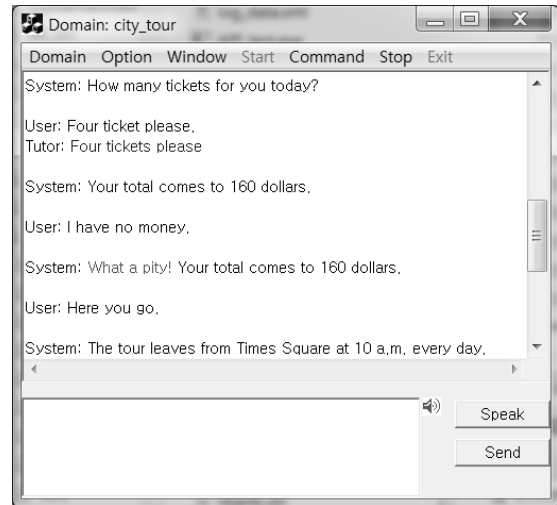


Fig. 2. A Dialog Between an English Learner and the DB-CALL System with Chatbot

#### 4.4 신경망 기반 사용자 발화 생성

검색 기반 챗봇이 기존 대화예문에서 응답을 가져오는 반면, 신경망 기반 챗봇은 sequence-to-sequence 생성 모델을 이용하여 입력된 사용자 발화에 대하여 대량의 대화 코퍼스로부터 새로운 응답을 생성한다. 대화 코퍼스에서 한 대화턴은 발화와 응답으로 구성되기에, 자동번역 학습코퍼스에서의 양국어 문장 쌍에 해당된다.

본 연구에서는 OpenNMT[23]의 양방향 순환(recurrent) 신경망 모델에 주의(attention)기반 인코딩-디코딩 구조를 적용하였다. 학습에서 41만 코퍼스와 140만 코퍼스에서 모두 고빈도 5만 어휘를 사용하였고, 각 1000개의 노드를 가진 4개의 층(layers)으로 20번 반복(epoch) 학습하여, 41만 대화 코퍼스에서 18.08의 혼잡도(perplexity)를 얻었고, 140만 코퍼스에서는 47.14의 혼잡도에서 수렴하였다.

### 5. 평가 및 실험

우선 목적지향 대화시스템 평가와 함께, 이를 위한 챗봇 평가를 함께 진행하였다(5.1절). 다음 독립된 챗봇에 대한 사용자 평가를 통해, 목적지향 대화시스템을 위한 챗봇과의 차이를 비교 토론하였다(5.2절). 마지막으로, 신경망 기반 챗봇과 검색기반 챗봇에 대한 비교 평가를 진행하여 챗봇 성능 향상을 위한 대응 방안을 모색하고자 하였다(5.3절)

#### 5.1 목적지향 대화시스템 및 이를 위한 챗봇 평가

영어교육 시스템에서 선택 주제에 대하여 사용자가 시스템과 대화하되 자유 발화를 허용하였다. 사용자가 대화맵의 지정된 태스크를 완성해야 해당 대화가 끝나며, 주제와 무관하게 발화하거나, 또는 실제 주제 발화하였으나 사용자 의도 인식 오류로 “모름”으로 분류된 경우, 챗봇이 응답하였

다. 모든 대화가 끝난 다음 사용자가 시스템 발화에 대해 0-1점을 부여하되, 해당 사용자 발화에 대하여 적합한 시스템 응답은 대화된 성공으로 간주하여 1점 부여하고, 그렇지 않은 경우 실패로 0점 부여한다. 20명의 영어 학습자가 평가에 참여하였으며, 4개의 주제에 대하여 평가하였다(Table 2)[24]. Table 2에서, 주제의 대화된 성공률은 주제대화 내에서 챗봇이 출력한 응답의 턴 성공률이다.

Table 2. Experimental Results on Chatbot for DB-CALL System

Topic	Turn success rate	Percentage of out domain utterance	Turn success rate of out domain dialogue
City tour tickets	71.86%	8.38%	50.00%
Ordering food	71.06%	9.56%	27.03%
Healthy habits	85.41%	13.30%	32.26%
Future currency	95.09%	10.53%	16.67%
Total	80.86%	10.44%	31.49%

5.2 독립 챗봇과의 비교 평가 및 토론

챗봇에 대한 독립 평가에서는 사용자가 독립된 챗봇을 이용하여 20~40 턴 발화하도록 하였다. 전체 대화가 종료된 후, 시스템 응답에 대하여 0~1점으로 시스템 응답 적합도를 평가하였다. 적합률은 목적지향 대화 평가에서의 턴 성공률에 해당된다. 사용자 15명이 실험에 임하여 평균 26.4턴 대화하였다. Table 3에서 볼 수 있듯이, 독립 챗봇에 대한 평가에서 1점 이상을 획득한 시스템 응답은 전체 응답의 52.78%이다. 여기에서 전체 응답의 26.27%를 차지하는 임의 발화 적합률은 25.96%로 챗봇 응답 적합률을 끌어내리는 요인으로 분석된다.

Table 3. Experiment Results on Independent Chatbot

Response type	Turn number	Percentage	Turn success number	Turn success rate
Exact matching	98	24.75%	74	75.51%
Fuzzy matching	194	48.99%	108	55.67%
Random response	104	26.27%	27	25.96%
Total	396	100%	209	52.78%

독립 챗봇의 전체 응답 적합률이 52.78%인데 반해(Table 3), 주제대화에서의 챗봇 응답 적합률은 31.49%밖에 되지 않는다(Table 2). 이는 챗봇 대화에서는 문맥에서 조금 벗어난다고 하더라도 해당 턴의 사용자 발화에 유의미한 응답이면, 적절한 응답으로 평가하는 경향이 있는 반면, 주제대화에서는

문맥에 정확히 맞는 응답이 요구되기 때문이다. 특히 챗봇 대화에서 시스템은 그 역할이 제한 받지 않는 반면, 주제대화에서 시스템에게는 특정된 역할이 부여되는데, 예를 들면 “시티투어 티켓구매”에서 시스템은 티켓판매원 역할을 담당하고, “음식주문”에서 시스템은 주문을 받는 역할을 맡고 있기에, 사용자는 이런 역할에 맞는 응답을 기대하게 된다. 다음은, 동일한 시스템 응답이 달리 평가 받는 대화의 예이다.

사용자-챗봇 대화:

Chatbot: You agree with me about what?

User: Nothing.

Chatbot: It's a damned ugly nothing.

또는,

User: Nothing.

Chatbot: Come on. Your face tells me something is wrong.

위의 응답이 챗봇 대화에서는 모두 적절한 것으로 간주되는 반면, 주제대화에서는 그렇지 않다. 아래는 영어교육시스템의 식사주문 주제대화에서 사용자가 시나리오에 없는 발화를 했을 경우 시스템이 챗봇을 통해 응답한 경우이다. 모두 적절한 응답으로 평가 받지 못한다(밑줄 그은 부분이 챗봇 이용한 응답이다):

DB-CALL: Would you like something to drink?

User: Nothing.

DB-CALL: It's a damned ugly nothing. Would you like something to drink?

또는:

DB-CALL: Come on. Your face tells me something is wrong. Would you like something to drink?

현재 각 주제를 위한 주제의 대화지식과 예문이 시스템이 담당하는 역할의 연령, 직업 등 개인정보에 제한되어 있는데, 향후 주제대화에서의 자유대화를 지원하기 위해, 각 주제별 상황에 맞는 일반 자유대화 예문의 자동 추출 방법에 대하여 고민할 필요가 있어 보인다.

이외에도, 현재의 영어교육시스템에서는 사용자와의 대화를 다시 주제대화로 유도하기 위해 챗봇 응답에 이어 시나리오에서의 시스템 응답을 그대로 출력한다. 그러나 다음 두 예문에서 볼 수 있듯이, 챗봇과의 보다 유기적인 결합이 필요해 보인다.

DB-CALL: All right, how about a beverage?

User: What do you have?

DB-CALL: You're not gonna believe this. All right, would you like a drink?

User: I want coffee.

DB-CALL: All right! Would you like something to drink?

5.3 검색기반 및 신경망 기반 챗봇 비교 평가

Table 3에서 볼 수 있듯이 검색기반 챗봇에서 유사예문이 없는 경우 임의발화를 사용하며, 이에 대한 응답 적합률은 25.96%로 전체 응답 적합률의 절반밖에 안 된다. 성능 개선을 위하여 유사한 예문이 없는 경우 임의발화 대신 생성모델을 적용해 보고자 한다.

평가문장은 Table 3에서 임의발화에 해당되는 104개의 사용자 발화이다. 신경망 기반으로 생성된 응답이 짧고 단순하기에[13] 시스템 발화 문장의 평균길이와 비중복률을 함께 평가하였다(Equation (2)).

$$\text{비중복률} = \text{중복제거 발화 수} / \text{전체 발화 수} \quad (2)$$

Table 4. Comparison Between Search Based and Neural Network Based Chatbots

Type	Turn success rate	Avg. length	Non-redundant turn no. (104 in total)	Non-redundancy rate
User utterance	-	4.7 words	101	97.11%
Random response (search based)	25.96%	4.8 words	101	97.11%
DNN based response(0.41M)	68.27%	2.5 words	53	50.96%
DNN based response(1.4M)	78.85%	2.4 words	53	50.96%

Table 4에서 볼 수 있듯이, 신경망 기반 챗봇의 응답 적합률은 검색기반과 동일한 41만 코퍼스에서 68.27%로, 검색기반 임의발화의 25.96%보다 훨씬 높을 뿐만 아니라, 검색기반 챗봇의 유사기반 매칭인 55.67%보다도 높다(Table 3). 특히 140만 코퍼스를 사용 시, 적합률은 78.85%로 검색기반 챗봇의 정확매칭의 75.51%보다도 높아진다. 그러나 이러한 적합률은 짧은 응답 길이(평균 2.4~2.5단어), 단순 응답의 중복발화(비중복률 50.96%)를 대가로 얻어진 것이다. 실제로 “I don’t know.”, “I know.”, “Yes.”, “No.”, “Yes~!”, “Why?”, “Thank you.” 등 고빈도 응답의 출현 빈도가 검색 기반의 그것보다 훨씬 높았다.

신경망 기반 응답의 높은 중복도가 대화가 길어질수록 더 심각할 것으로 사료되어, 이를 확인하기 위해 다시 20명의 사용자에게 챗봇과 20턴 이상 대화하도록 하였다. 이번 평가에서는 평균 발화 길이와 중복도만 평가하였다. 사용자 발화수는 총 646 발화였고, 검색기반 챗봇의 결과에서 정확매칭 유사매칭 등 결과를 구분하지 않고 모두 포함시켰다.

Table 5. The Responses from Neural Network Based Chatbot Tend to be Simple, Short and Redundant

Type	Avg. length	Non-redundant turn no. (646 in total)	Non-redundancy rate
User utterance	3.55 words	553	85.60%
Search based response	4.39 words	586	90.71%
DNN based response(1.4M)	2.53 words	134	20.74%

Table 5에서 보듯이, 대화가 진행 될수록, 신경망 기반 응답의 중복도는 더 올라가고, 평균 발화 길이는 크게 변하지 않는다. 그 만큼 짧은 문장 위주로 단순 중복 발화가 많다는 뜻이며, 사용자 입장에서는 대화의 흥미도가 떨어 질 수 있다. 다만, 응답 적합률이 높은 장점을 가지고 있기에, 향후 검색기반 챗봇의 보조역할로 유사한 예문을 찾지 못할 경우 생성기반 방법으로 응답을 생성할 수 있을 것으로 기대된다.

6. 결론 및 향후 연구

본 논문은 목적지향 대화시스템에서 대화 자유도를 향상시키기 위해 챗봇을 활용하였다. 사용자 평가를 통해 주제 대화의 문맥에서 자유대화를 지원하기 위하여 챗봇을 단순히 추가하는 것만으로는 만족스러운 효과를 보지 못할 수 있음을 알 수 있었다. 주제대화의 문맥에 맞는 자유대화를 지원하기 위해, 주제대화의 상황에 맞는 소규모 대화 코퍼스의 자동 학습 및 추출에 대한 연구가 필요하다. 이외 일련의 사용자 평가를 통해 챗봇 성능 향상을 위해 검색기반 및 신경망 기반 챗봇의 하이브리드를 제안하였다. 챗봇 대화를 위한 신경망 모델에 관한 연구가 활발히 진행되고 있는 만큼, 향후 문맥을 반영하는 최적화된 대화 모델에 대한 연구도 필요할 것이다.

References

[1] Apple Siri [Internet], <https://www.apple.com/kr/ios/siri/>.  
 [2] Amazon Alexa[Internet], <https://developer.amazon.com/alexa>  
 [3] Houndify[Internet], <https://www.houndify.com/>.  
 [4] O.-W. Kwon, T.-K. Hong, J.-X. Huang, and Y.-K. Kim “Virtual personal assistant: dialogue processing technique and recent trends,” *Communications of the Korean Institute of Information Scientists and Engineers*, Vol.35, No.8, pp.19-27, 2017.  
 [5] A. Raux and M. Eskenazi, “A finite-state turn-taking model for spoken dialog systems,” *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, Boulder, Colorado, pp.629-637, 2009.

- [6] D. Goddeau, H. Meng, J. Polifroni, S. Seneff, and S. Busayapongchai, "A form-based dialogue manager for spoken language applications," in *Proceedings of International Conference of Speech Language Processing*, Philadelphia, PA, 1996.
- [7] I. Ahmed and S. Singh, "AIML based voice enabled artificial intelligent chatterbot," *International Journal of u- and e-Service, Science and Technology* Vol.8, No.2, pp.375-384, 2015.
- [8] C. Lee, S. Jung, S. Kim, and G. Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Communications*, Vol.51, No.5, pp.466-484, 2009.
- [9] A. Ritter, C. Cherry, and W. B. Dolan, "Data-driven response generation in social media," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp.583-593, 2011.
- [10] S. Jung, S.-H. Na, "Refining sentence similarity with discourse information in dialog system," in *Proceedings of INTERSPEECH*, 2013.
- [11] R. Yan, Y. Song, and H. Wu, "Learning to respond with deep neural networks for retrieval-based human-computer conversation system," in *Proceedings of SIGIR'16*, Pisa, Italy, 2016.
- [12] I. Sutskever, O. Vinyals, and Q. Le, "Sequence to sequence learning with neural networks," in *Proceedings of NIPS*, 2014.
- [13] O. Vinyals and Q. Le, "A neural conversational model," *Proceedings of the 31st International Conference on Machine Learning, Deep Learning Workshop*, 2015.
- [14] J. Bang, H. Noh, Y. Kim, and G. Lee, "Example-based chat-oriented dialogue system," in *Proc. BigComp* pp.238-243. 2015.
- [15] ChatScript [Internet], <http://chatscript.sourceforge.net/>.
- [16] O.-W. Kwon, K. Lee, Y.-H. Roh, J.-X. Huang, S.-K. Choi, Y.-K. Kim, H.B. Jeon, Y. R. Oh, Y.-K. Lee, B.O. Kang, E. Chung, J.G. Park, and Y. Lee, "GenieTutor: a computer assisted second-language learning system based on spoken language understanding," in *Proceedings of the 2015 International Workshop on Spoken Dialogue Systems*, 2015.
- [17] J.-X. Huang, K.-S. Lee., O.-W. Kwon, and Y.-K. Kim, "Automatic dialogue scoring for a second language learning system," in *S. Papadima-Sophocleous, L. Bradley & S. Thouesny (Eds), CALL Communities and Culture - Short Papers from EUROCALL*, pp.1-6, 2016.
- [18] S.-K. Choi, , O.-W. Kwon, Y.-K. Kim, and Y. Lee, "Using a dialogue system based on dialogue maps for computer assisted second language learning," in *S. Papadima-Sophocleous, L. Bradley & S. Thouesny (Eds), CALL Communities and Culture. Short papers from EUROCALL*, pp.1-7, 2016.
- [19] R. E. Banchs, "MovieDic: a movie dialogue corpus for research and development," in *Proceedings of ACL*, pp.203-207, 2012.
- [20] BNC [Internet], <http://www.natcorp.ox.ac.uk/>.
- [21] T. Strohan, D. Metzler, H. Turtle, and W. B. Croft, "Indri: a language model-based search engine for complex queries," in *Proceedings of the International Conference on Intelligence Analysis*, 2005.
- [22] C. X. Zhai, "Notes on the Lemur TFIDF model" [Internet], <http://www.lemurproject.org/lemur/tfidf.pdf>, 2001.
- [23] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. "OpenNMT: open-source toolkit for neural machine translation" [Internet], *CoRR*, *abs/1701.02810*, <http://arxiv.org/abs/1701.02810>. 2017.
- [24] S.-K. Choi, O.-W. Kwon, K. Lee, Y.-H. Noh, J.-X. Huang, and Y.-K. Kim "English education system using a chatter bot and a dialogue system," in *Proceedings of Spring Conference of KIPS*, 2017.

### 황금하



e-mail : hgh@etri.re.kr

1991년 JILIN University, 물리학과(학사)

2000년 KAIST 전산학과(석사)

2008년 KAIST 전산학과(박사수료)

1994년~1996년 연변과학기술대학 직원

2001년~2003년 Microsoft Research Asia,  
Assistant Researcher

2008년~현 재 한국전자통신연구원 언어지능연구그룹  
선임연구원

2014년~현 재 전북대학교 컴퓨터공학부 박사과정  
관심분야: 대화시스템, 기계번역, 기계학습, 지식추출

### 권오욱



e-mail : ohwoog@etri.re.kr

1992년 경북대학교 컴퓨터공학과(학사)

1995년 KAIST 전산학과(석사)

2001년 포항공과대학교 컴퓨터공학과(박사)

1995년~2002년 포항공과대학교  
정보통신연구원 연구원

2002년~2004년 (주)유니소프트 NLP연구소 책임연구원  
2004년~현 재 한국전자통신연구원 언어지능연구그룹  
책임연구원

관심분야: 대화시스템, 기계번역, 기계학습, 언어처리, 문서분류



**이 경 순**

e-mail : selfsolee@chonbuk.ac.kr  
1997년 한국과학기술원 전산학과(석사)  
2001년 한국과학기술원 전산학과(박사)  
2001년~2003년 일본 국립정보학연구소  
(NII) 연구원  
2007년~2008년 미국 메사추세츠주립대학  
방문교수

2004년~현 재 전북대학교 컴퓨터공학부 교수  
관심분야: 정보검색, 정보마이닝, 소셜데이터분석



**김 영 길**

e-mail : kimyk@etri.re.kr  
1991년 한양대학교 전자통신학과(학사)  
1993년 한양대학교 전자통신학과(석사)  
1997년 한양대학교 전자통신학과(박사)  
1997년~현 재 한국전자통신연구원  
언어지능연구그룹 그룹장

관심분야: 기계번역, 대화처리, 언어처리, 기계학습