심층 신경망 기반 대화처리 기술 동향

Trends in Deep-Neural-Network-Based Dialogue Systems

권오욱 (O.W. Kwon, ohwoog@etri.re.kr) 언어지능연구실 책임연구원
홍택규 (T.G. Hong, tghong@etri.re.kr) 언어지능연구실 연구원
황금하 (J.X. Huang, hgh@etri.re.kr) 언어지능연구실 선임연구원
노윤형 (Y.H. Roh, yhroh@etri.re.kr) 언어지능연구실 책임연구원
최승권 (S.K. Choi, choisk@etri.re.kr) 언어지능연구실 책임연구원
김화연 (H.Y. Kim, cind0605@etri.re.kr) 언어지능연구실 UST학생연구원
김영길 (Y.K. Kim, kimyk@etri.re.kr) 언어지능연구실 책임연구원/실장
이윤근 (Y.K. Lee, yklee@etri.re.kr) 인공지능연구소 책임연구원/소장

ABSTRACT

In this study, we introduce trends in neural-network-based deep learning research applied to dialogue systems. Recently, end-to-end trainable goal-oriented dialogue systems using long short-term memory, sequence-to-sequence models, among others, have been studied to overcome the difficulties of domain adaptation and error recognition and recovery in traditional pipeline goal-oriented dialogue systems. In addition, some research has been conducted on applying reinforcement learning to end-to-end trainable goal-oriented dialogue systems to learn dialogue strategies that do not appear in training corpora. Recent neural network models for end-to-end trainable chit-chat systems have been improved using dialogue context as well as personal and topic information to produce a more natural human conversation. Unlike previous studies that have applied different approaches to goal-oriented dialogue systems and chit-chat systems respectively, recent studies have attempted to apply end-to-end trainable approaches based on deep neural networks in common to them. Acquiring dialogue corpora for training is now necessary. Therefore, future research will focus on easily and cheaply acquiring dialogue corpora and training with small annotated dialogue corpora and/or large raw dialogues.

KEYWORDS 대화처리, 종단형 대화처리, 강화학습, 목적지향 대화처리, 대화이해, 신경망 학습

Ⅰ. 서론

인간의 말을 기계가 이해하여 다양한 업무를 처리하고 정보를 제공하는 대화처리 기술이 다양한 스마트 기기의 가상 개인비서 또는 대화인터페이스에 적용되어 사용되고 있다.

* DOI: https://doi.org/10.22648/ETRI.2019.J.340406
* 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[2019-0-0004, 준지도학습형 언어지능 원천기술 및 이에 기반한 외국인 지원용 한국어 튜터링 서비스 개발].

대화처리 기술은 사용자의 말을 기계가 수행할 수 있는 명령어들 중에서 정확히 이해하는 기술 수준에서 사용자의 말을 이해하고 공감하며 기계가 보유한 전문분야 지식으로 사용자 요구를 만족시키는 기술 수준까지를 포함하고 있다. 기계가 인간의 대화 능력과 지식을 가져서, 인간의 다양한 역할인 업무 대리인, 친구, 선생님, 전문상담사 등이 가능하도록 하는 기술을 대화처리 기술로 보기도 한다.

대화처리 기술 연구는 인간이 대화로 수행하는 다양한 기능 및 능력을 모두 처리 가능한 연구 범위로 설정하는 데 어려움이 있어 세분화된 특정 기능 및 능력에 집중하는 기술 연구로 발전되어 왔다. 대표적으로, 특정 업무에서 사용자의 요구 목적을 대화로 처리하기 위한 목적지향 대화시스템(goal-oriented dialogue system), 인간처럼 일상 대화를 하지만 재미로 다양한 대화를 수행하게 하는 chit-chat 대화시스템, 사용자의 특정 질문에 전문지식으로 대답하기 위한 질의응답시스템(question answering system) 등이 있다. 상기 분류는 각 기술의 대화 목표에 따라 필요한 요소 기술과 구성 모듈이 다르기 때문에 구분되었다. 스마트 폰이나 스마트 스피커 등에 탑재된 상용화 개인비서 또는 대화처리 기술은 상기 여러 가지 대화처리시스템들이 같이 결합되어 다양한 영역의 인간 대화를 처리하도록 하고 있다.

최근 음성인식, 영상인식, 기계번역 등의 다양한 인공지능 분야에서 신경망(neural network) 심층학습(deep learning)으로 인해 큰 성능 향상을 가져왔다. 특히, 신경망 학습을 통해 원문 문장을 번역문 문장으로 자동으로 생성하는 신경망 기반 자동번역(neural machine translation)의 성공은 사용자 요구문장을 시스템 응답문장으로 자동 생성하는 신경망 기반 대화처리 기술 연구를 활발하게 하였다.

본 동향지에서는 최근 심층 신경망 기술을 적용한 대화처리 기술 사례를 기술하고, 그 전망을 기술하였다. 주로, 목적지향 대화처리와 chit-chat 대화처리 기술에 적용한 연구 사례들을 기술하였다. 또한 심층학습으로 대화처리 기술이 발전됨에 따라, 학습에 필요한 학습용 대화문장들의 양과 질이 매우 중요하게 되었다. 하지만 현실에서는 특정 목적 및 서비스를 위한 대화문장들이 매우 부족하거나 없는 실정이다. 최근 이러한 대화학습코퍼스 구축 사례를 조사하고 전망을 기술하였다.

Ⅱ. 목적지향 대화처리 동향

1. 기존 목적지향 대화시스템

목적지향 대화시스템은 사용자들이 자연어 형태의 음성이나 텍스트 입력을 통해 원하는 작업을 수행할 수 있도록 도와주는 시스템으로, 그 예시로는 개인 스케줄 관리, 호텔 예약, 길 찾기 시스템 등이 있다. 목적지향 대화시스템은 사용자와의 단순한 잡담을 목적으로 하는 chit-chat 대화시스템과는 다르게 사용자가 시스템과의 대화를 통해 달성하고자 하는 명확한 목표가 있고, 시스템과 연결되어 있는 대용량 database가 있어서, 이러한 database에서 사용자가 원하는 정보를 찾아 전달해 주는 것을 목적으로 한다. 최근 국내외 유수 대기업들이 스마트 스피커를 출시함에 따라 목적지향 대화시스템과 이를 활용한 어플리케이션에 대한 관심이 증대되고 있다.

전통적인 목적지향 대화시스템은 그림 1과 같이 대화 이해, 대화 관리, 대화 생성 모듈이 pipeline 형태로 연결된 구조를 가지고 있다[1,2]. 대화 이해 모듈은 사용자 발화를 입력받아, 사용자의 발화 의도와 목적에 해당하는 의미인 슬롯 값들을 예측

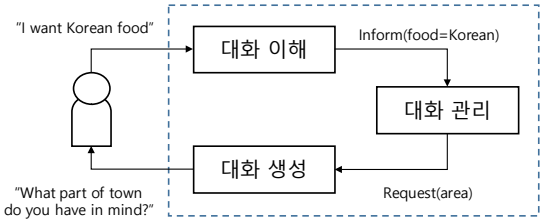

그림 1 Pipeline 기반 목적지향 대화시스템

한다. 대화 관리 모듈은 대화 이해 결과와 현재까지의 대화 이력을 바탕으로 시스템 응답 발화 의도를 선택한다. 마지막으로, 대화 생성 모듈에서 선택된 시스템 응답 발화 의도와 추적 중인 슬롯 값들로부터 자연어 형태의 시스템 응답을 생성한다.

그러나 이러한 pipeline 기반 목적지향 시스템은 다음과 같은 두 가지 단점을 가지고 있다[3]. 첫 번째는 오류가 발생했을 때 오류의 원인을 파악하기 힘들다. 한 모듈에서 발생한 오류가 다른 모듈로 전이되기 때문에, 문제가 발견되었을 때 해당 문제가 어떤 모듈에서 기인되었는지 알아내기 어렵다. 두 번째는 새로운 도메인에 적용하는 것이 어렵다. 모듈 간에 연관성이 있기 때문에 한 모듈이 변경된다면 그에 따라 다른 모듈들도 모두 재학습되어야 한다.

2. 종단형 목적지향 대화시스템

기존 목적지향 대화시스템의 한계를 극복하기 위해 심층 신경망 기반 종단형 방법이 도입되었다. 종단형 방법은 chit-chat 대화시스템에 먼저 적용되었고 가능성을 보여, 목적지향 대화시스템 연구 분야에서 2016년경에 적용되어 3년 사이에 핵심 연구 분야로 자리 잡았다.

Facebook AI Research 연구진은 종단형 메모리 네트워크를 이용하여 목적지향 대화를 수행하는 방법을 제시하였다[4]. Database 검색 결과를 포함하는 대화 이력을 메모리로, 최근 사용자 발화를 query로 설정하고, 시스템 응답 중 하나를 선택하는 방식으로 구성하였다.

Microsoft Research와 Brown 대학 공동 연구진은 도메인에 따라 직접 구현 가능한 슬롯 값 추적 및 대화 생성 부분과 시스템 발화 템플릿을 선택하는 학습 가능한 순환 신경망을 결합한 Hybrid Code Network(HCN)를 제안하였다[5]. 목적지향 대화시스템의 경우 대부분 도메인별로 개발되기 때문에, 이 모델은 도메인의 특성에 따라 간단하게 처리될 수 있는 부분은 개발자가 직접 코딩해서 처리하도록 하고, 나머지 도메인에 상관없는 공통적인 부분, 시스템 발화 템플릿을 선택하는 부분은 순환 신경망이 선택할 수 있도록 처리하였다. 목적지향 대화시스템 평가에 널리 사용되고 있는 식당 검색 도메인에서 HCN 모델이 시스템 발화 정확도 기준으로 현재 제일 높은 성능을 보이고 있다[4].

Carnegie Mellon 대학 연구진은 계층형 long short-term memory(LSTM)를 이용하여 목적지향 대화를 다중 작업 순차 라벨링 문제로 모델링하였다[6]. 문장 수준의 LSTM을 이용하여 사용자 발화를 인코딩하고, 이 값을 문맥 수준의 LSTM의 입력 값으로 주어 문맥 정보를 유지하여 시스템 발화 템플릿, 슬롯 값, database entity를 예측한 다음, 예측한 값들을 조합하여 최종적인 시스템 응답을 만든다.

홍콩과기대 연구진은 database 정보를 포함시켜서 시스템 응답을 생성할 수 있는 모델을 만들기 위해 종단형 메모리 네트워크와 sequence-to-sequence 모델을 결합하는 방법을 제시하였다[7]. 제안 방법에서 database 검색 결과를 포함하는 대화 이력을 메모리로 두는 것은 참고문헌 [4]와 비슷하지만, 여러 개의 시스템 응답 중 하나를 선택하는 것이 아니라, 이렇게 인코딩된 정보로부터 순환 신

경망을 이용하여 시스템 응답을 단어 단위로 생성한다.

도메인의 대화문장들로만 대화시스템을 학습하고 그 성능이 상용화 수준에 이르기 위한 종단형 목적지향 대화처리를 위한 최적의 심층 신경망 모델을 찾는 연구가 향후 계속 발표될 것으로 기대된다. 또한, 목적지향 대화에 필요한 여러 가지 도메인 지식 간의 관계를 심층 신경망에서 처리 가능한지에 대한 연구도 필요할 것이다.

3. 목적지향 대화시스템용 강화학습

지도 학습만으로 대화시스템을 학습시키면, 학습코퍼스에는 등장하지 않는 다양한 대화 흐름에 대처하지 못할 수 있게 된다. 그래서 보다 강건한 대화시스템을 구축하기 위해 강화학습을 이용하여 다양한 대화 흐름에도 대처할 수 있도록 하는 연구가 예전부터 진행되어 왔다.

강화학습을 적용하려면 환경에서의 보상을 정의해야 한다. 일반적으로, 목적지향 대화시스템에서는 대화가 성공했을 때 큰 양의 보상을, 대화가 실패했을 때 0 또는 음의 보상을, 그리고 매 턴마다 −1 또는 0의 보상을 준다. 여기서 대화 성공은 대화가 끝났을 때 사용자의 목표를 정확하게 예측했는지를 나타낸다. 여기서 사용자의 목표는 database 검색을 위해 사용자가 시스템에게 값을 알려주는 슬롯들과 사용자가 값을 물어보기 위한 슬롯들로 구성된다. 그리고 매 턴마다 −1의 보상을 주는 이유는 보다 적은 턴 안에 대화를 끝낼 수 있도록 하기 위함이다.

Carnegie Mellon 대학과 Microsoft Research, 그리고 국립 타이완 대학 공동 연구진은 database에 대한 사후 확률 분포를 계산하는 모델을 설계하고 해당 모델의 파라미터를 정책 경사 알고리즘 중 하나인 REINFORCE를 이용하여 종단형으로 학습하는 방법을 제시하였다[8]. 여기서는 시스템이 제안하는 database entity 목록 중 사용자가 원하는 entity가 있는 순위에 따라 보상을 주었다.

Microsoft Research와 Brown 대학 공동 연구진은 앞서 소개한 종단형 모델을 강화학습으로 학습시키는 방법도 소개하였다[5]. 시스템 발화 템플릿을 선택하는 순환 신경망을 정책으로 보고 정책 경사 알고리즘을 적용하였다.

Carnegie Mellon 대학 연구진도 참고문헌 [6]의 후속 연구로써 강화학습을 적용하였다[9]. 여기서는 문맥 정보를 인코딩하는 문맥 수준 LSTM의 내부 상태를 강화학습 환경에서의 상태로 보고, 시스템 발화 의도를 선택하는 계층을 정책으로 본 다음 REINFORCE 알고리즘을 이용하여 강화학습을 적용하였다. 여기서는 매 턴마다의 보상을 −1과 같은 형식으로 주는 것이 아니라 이전 턴 대비 사용자 목표를 얼마나 잘 예측하고 있는지에 대한 차이를 보상으로 주었다.

Microsoft Research와 국립 타이완 대학 공동 연구진은 대화 이해, 대화 관리, 대화 생성 모듈을 각각 종단형 방법으로 학습시키는 방법을 제안하였다[10]. 그리고 시스템 발화 의도를 선택하는 대화 관리 모듈의 경우 deep Q−network(DQN)로 표현하고 강화학습을 적용하는 방법을 제시하였다.

종단형 목적지향 대화시스템 연구에서 가장 큰 문제인 도메인 대화코퍼스 부재나 부족 문제를 해결하기 위해서는 대화처리를 위한 강화학습 방법론이 절실히 필요하다. 하지만 기존의 일반적인 강화학습과 달리, 특정 목적 환경에서 대화 문제의 강화학습을 위한 환경과 보상을 설정하는 문제가 단순하지 않기 때문에 현재의 단순한 강화학습 방법론을 적용하는 연구에서 벗어나 대화 문제를 가장 잘 해결하기 위한 강화학습 방법론 연구가 필요하다.

III. Chit-chat 대화처리 동향

재미나 잡담을 위한 목적으로 개발되는 대화시스템은 chatbot 또는 chit-chat 대화시스템이라고 불린다. 최근에는 대화 문맥, 주제, 지식까지 반영하는 지능형 자유 대화에 대한 연구가 활발하게 진행되고 있기 때문에 대화 에이전트(conversation agent)라고 부르는 경우가 많다.

자유 대화처리 방법으로 ChatterBot과 CleverBot과 같이 대용량 대화 예문을 이용한 검색기반 챗봇이나 AIML[11] 및 ChatScript[12]와 같이 스크립트 언어로 표현된 규칙 패턴 및 지식을 이용한 방법론들도 있지만, 신경망 기술의 발전에 따라 chit-chat 대화처리 연구는 대화코퍼스를 이용한 신경망 기반 대화 모델에 대한 연구로 옮겨지고 있다.

신경망 기반 자유대화 모델의 발전은 신경망 기술의 발전에 따라 함께 발전되어 왔으므로 초기 sequence-to-sequence 생성모델을 이용하거나 검색기반 방법과 하이브리드하여 보다 높은 정확도를 기록하였다[13-15]. 응답 생성에서 다음 단어 생성을 위한 중요한 단어를 찾기 위해 신경망 기반 자동번역을 위한 방법에서 많이 사용되는 attention 모델을 도입하기도 하였다[16]. 하지만, 번역과 달리 대화에서는 동일한 의미라고 하더라도 다양한 표현의 응답을 생성하는 것이 필요한 점에서 attention 모델 도입이 무조건 효과적이지 않을 것으로 보인다.

단일 생성모델을 이용한 대화의 경우 간단하고 일률적인 고빈도 응답만 생성하는 단점이 있었는데, 이를 극복하기 위하여 손실함수(loss function)를 수정하거나[17] Variational Autoencoders(VAE)를 도입하여 다양한 응답을 생성하는 데 관한 연구[18]도 진행되었다.

Chit-chat 대화시스템 또는 대화 에이전트 연구의 주요 이슈로 다양하고 자연스러운 응답 생성 외에 문맥 반영한 대화, 주제 및 주제 관련 지식을 반영한 대화, 개인 특성(persona)을 반영한 대화, 정확한 응답이 필요한 대화 등이 있다.

대화 흐름에 따라 보다 정확한 응답을 하기 위한 문맥을 반영한 대화를 생성하는 연구가 가장 먼저 제기되었다. 이를 해결하기 위해 인코더에서 하나의 재귀신경망 층으로 최근의 사용자 발화를 인코딩하고, 다른 한 층의 재귀신경망으로 지금까지의 대화들을 인코딩하는 계층 구조를 도입하기 시작하였다[19-21]. 문맥 정보에서도 보다 중요한 부분을 발굴하기 위해 사용자 발화를 인코딩한 회귀신경망 층에 단어 레벨의 attention을 가하고, 대화 문맥을 인코딩한 재귀신경망 층에 발화 레벨의 attention을 가지는 계층 재귀 집중 신경망 모델이 대화 문맥을 가진 대용량 대화코퍼스를 이용한 실험에서 좋은 성능을 보여주었다[22].

주제를 반영한 대화는 반영하고자 하는 정보가 지식인 경우 지식 기반 대화라고도 불리는데, 단순하고 일률적인 응답 대신 유익한 정보를 제공하고, 주어진 주제에 대한 보다 깊이 있는 대화를 생성하는 데 그 목적을 둔다[23-25]. 주제 반영 모델에서는 대화를 인코딩하는 것 외에 해당 대화와 관련된 주제관련 키워드[23] 또는 지식[24,25]을 따로 인코딩한 다음 attention을 더하는 방식으로 해당 발화와 가장 관련도 높은 지식에 가중치를 부여하여 디코더로 대화 문맥과 함께 입력하여 응답을 생성한다. 해당 대화와 관련된 주제 지식을 얻기 위해 트윗 데이터나 위키피디아 데이터 같은 대용량 지식에서 전통적인 키워드 기반 정보검색으로 대화 관련 주제지식을 검색한다[24,25].

개인 특성을 반영한 대화는 두 대화자의 나이, 직업, 가족, 취미 등 개인 정보가 정해진 상황에서 시스템의 발화가 부여된 개인 특성에 일치된 발화

를 생성하도록 하는 대화이다. 일반적인 지식기반 대화와 사용하는 지식의 양식이 다를 수 있지만 그 방법은 주제 대화와 비슷하다[26].

지식을 반영한 주제 대화 모델에서는 메모리 네트워크 구조를 많이 사용한다[24-26]. 마지막 상태 정보만 전달하는 재귀신경망으로 지식을 임베딩하는 경우를 단기메모리 방식이라고 본다면, 지식을 문장 단위로 메모리 벡터에 임베딩하되 여러 메모리 벡터로 여러 문장 정보를 독립적으로 저장하여 입력된 사용자 질의에서 참조할 수 있는 메모리 벡터 구조는 장기메모리 또는 글로벌 메모리 기능을 수행한다고 볼 수 있으므로 질의응답에서 많이 사용되어 왔던 구조이기도 하다[27-30].

Chit-chat 대화의 또 하나의 이슈는 평가방법이다. 목적지향 대화와 달리 자유대화에는 정답이라는 개념이 모호하므로 기존 생성모델 이용한 연구에서는 언어 모델을 평가할 때 사용하는 복잡도(perplexity) 점수, 번역 모델을 평가할 때 사용하는 정답과의 n-그램 유사도를 나타내는 BLEU, 정답 문장과의 중복된 단어수를 나타내는 F1 평가 기준 등을 사용해 왔고 검색 모델의 경우 질의응답 모델을 평가할 때 사용하는 hit@k이나 위의 F1 평가를 사용해왔다. 다만 복잡도를 제외한 다른 평가 기준은 모두 정해진 정답 또는 최소한 사람이 구축한 응답 등이 있어야 사용 가능하다[31]. 사람에 의한 평가로 정확한 정도를 점수로 표현하거나 재미있는 정도를 점수로 표현하는 방법 등도 있지만, 인간 대화에서 아주 중요한 대화의 일관성까지 반영할 수 있는 평가 지표는 아직 없는 실정이다[31].

Chit-chat 대화는 자동평가를 통하여 타 연구와의 비교평가가 어렵고 주어진 대화코퍼스에 따라 모델이나 성능이 많은 차이를 보이기 때문에 최근에는 여러 팀이 함께 참여하여 직접 비교 평가받는 경쟁 대화를 통하여 모델 비교에 대한 시도가 이루어지고 있다[32,33].

대화는 자연언어 이해와 함께 인공지능 분야에서 가장 각광받는 열린 문제(open problem) 중의 하나이다. 잡담 대화 또는 자유대화에 대한 연구는 향후 기계독해(machine reading comprehension), 질의 생성 등 문제에 대한 연구와 함께 결합하며 유용한 정보도 주고받는 보다 인간다운 지능형 대화에 대한 연구로 발전할 것으로 기대한다. 또한 목적지향 대화처리를 포함하는 모델로의 발전을 기대하고 있다.

IV. 신경망 기반 대화이해 동향

대화이해 기술은 사용자의 입력 발화를 기계가 이해할 수 있는 의미적인 표현으로 매핑하는 기술을 의미한다. 예를 들면 “내일 모레 서울 날씨 어때”라는 발화는 표 1과 같이 표현된다.

이렇게 표현된 대화이해 결과는 대화관리 모듈에서 사용자의 의도에 따라 적절한 응답을 주거나, 특정한 명령을 수행하는 용도로 사용된다.

따라서 대화이해는 사용자 발화가 들어오면, 주어진 도메인이나 서비스에서 사용되는 사용자 의도 중에서 하나로 사용자 의도를 분류하고, 슬롯을 인식하는 과정으로 이루어진다.

최근 신경망 기반 학습 기술의 발전으로 종래 규칙 기반이나 전통적인 기계학습 방법에 대해 규칙이나 자질을 기술할 필요가 없는 신경망 기반 대화이해 방법이 활발히 연구되고 있다. 신경

표 1 사용자 발화 의미 표현 예

발화	내일 모레 서울 날씨 어때	
사용자 의도	request(Weather.info)	
슬롯	date	내일 모레
	city	서울

표 2 BIO 방식 슬롯 태깅 예

	내일	모레	서울	날씨	어떻	어
BIO 태깅	B-date	I-date	B-city	O	O	O

망 기반 방법에서 사용자 의도를 분류하기 위한 방법은 Deep Belief Network(DBN), Deep Convex Network(DCN)과 같은 단순한 피드포워드 네트워크에서, 이전의 hidden state를 계속 반영하여 새로운 hidden state를 생성함으로써 sequence 처리에 적합한 Recurrent Neural Network(RNN)[34], 지역적인 문맥을 반영할 수 있는 Convolutional Neural Network(CNN)[35] 등을 사용하는 방법으로 발전하였다.

슬롯을 인식하는 과정은 주로 BIO 방식의 순차적 태깅 방법에 의해 이루어진다. 예를 들면 표 2와 같다.

슬롯 태깅 방법의 가장 기본적인 형태는 RNN인데, sequence가 길어지는 경우 나타나는 원거리 의존 문제를 처리하기 위해 Long Short-Term Memory(LSTM), Gated Recurrent Unit(GRU) 등을 이용한 방법들이 도입되었다. LSTM은 포겟 게이트, 입력 게이트, 출력 게이트 등을 이용하여 이전 문맥과 입력 정보 중에 어떤 정보를 얼마나 반영할지를 제어함으로써 원거리 의존 문제를 처리하는 방식이다. 또한 뒤에서부터 인코딩을 해오는 레이어를 추가하여 문장 뒤에서부터 인코딩을 해오는 레이어를 추가하여 뒤쪽 문맥을 반영하도록 한 bi-directional LSTM-RNN이 좋은 성능을 보이고 있고 널리 쓰이고 있다. 여기에, 앞뒤 슬롯 태그 간에는 의존성을 고려하기 위한 CRF 레이어가 추가된 형태[36], RNN으로 입력문장을 인코딩한 후 다시 RNN으로 결과를 생성하는 encoder-decoder 네트워크, 입력 단어 중에서 어떤 단어의 정보를 많이 반영할지를 고려하도록 attention이 추가된 모델[37] 등이 사용되었다.

최근에는 사용자 의도와 슬롯 태깅을 동시에 수행하는 방법이 주를 이루고 있다. 사용자 의도와 슬롯은 서로 의존 관계에 있고, 서로 간에 정보를 공유하면 성능향상에 도움이 될 수 있기 때문이다. 일부 CNN을 기반으로 한 방법이 있었지만, 주로 RNN을 기반으로 한 방법을 사용하고 있다. RNN에서 마지막 단어에 'EOS'를 추가해서 사용자 의도를 분류하는 방식[38]이나, encoding 후 사용자 의도 분류와 decoding을 통해 슬롯 태깅을 하는 방식[39] 등이 있다.

그 외에 Slot-gated 모델[40], Bi-model 방법[41] 등을 통해 성능개선을 나타낸 연구들이 있는데, 모두 사용자 의도와 슬롯 간의 영향력을 강화하고자 하는 데 초점을 둔 것이다. 최근 비지도 방식으로 언어모델을 학습한 결과를 실제 응용 태스크에 fine-tuning하는 multi-task 방식이 다양한 언어처리 태스크에 대해 좋은 성능을 내고 있고, 구글의 Transformer 모델 기반의 pre-trained 언어모델인 BERT를 기반으로 대화이해를 수행한 모델이 크게 개선된 성능이 보였고 현재 최신 성능을 나타내고 있다[42].

사용자 발화를 이해하기 위해서는 문장만으로는 어려운 경우가 있기 때문에, 이전 문장 또는 대화 흐름을 반영하는 문맥기반 대화이해 방법이 연구되고 있다. 문맥기반 방법으로는 RNN으로 이전 문장들을 인코딩한 결과를 사용하거나[43], 메모리 네트워크를 통해 문맥을 반영하는 방법[44] 등이 있다.

또한 앞서 언급한 BERT기반 대화이해 방법에서도 이전 문장들을 문맥으로 같이 학습할 수 있는 구조를 가지고 있어서 문맥기반 대화이해 방법으로 쉽게 확장될 수 있을 것으로 여겨진다.

한편, 음성인식 모듈과 대화이해 모듈이 통합되

어 아예 음성신호로부터 대화이해 결과를 생성하는 종단 간(end-to-End) 음성대화이해 시스템에 대한 연구가 시도되고 있다[45]. 그러한 방법은 억양, 피치, 휴지(pause) 등의 음성 정보를 이용할 수 있고, 각각의 모듈을 고려하여 전체적인 성능을 최적화할 수 있는 장점이 있다. 하지만 각 도메인마다 음성 데이터를 구축하는 것은 쉽지 않은 일이기 때문에 실용적으로 적용되기에는 한계가 있을 것으로 생각된다.

현재 대화처리 기술의 추세가 종단 간 대화처리 방식을 지향하기 때문에, 대화이해 기술은 대화관리 기술과 결합하여 하나의 모델로 통합될 가능성이 있고, 궁극적으로 사용자 의도나 슬롯이 태깅이 필요 없는 원시 대화 시나리오로부터 자동으로 대화이해를 학습하는 방향으로 발전될 것으로 보인다.

V. 대화처리 데이터셋 구축 동향

대화시스템에서 대화 데이터셋의 구축량과 구축 품질은 대화시스템의 성능을 결정하는 중요한 요소 중 하나이다. 특히 신경망 기반 대화시스템에서는 학습용 데이터셋의 양과 질에 따라 대화의 품질이 결정되기 때문에 고품질의 대화 데이터셋의 구축양이 매우 중요하다고 할 수 있다.

대화처리 데이터셋의 언어와 관련하여, 대화처리 데이터셋은 영어를 중심으로 구축되고 있으며, 최근에는 중국어 데이터셋도 대량으로 구축되고 있으나 한국어는 연구 차원의 수준에서 구축되고 있다[46]. 영어를 중심으로 한 활용 가능한 대화처리 데이터셋에 대한 자세한 소개는 참고문헌[47,48]에 나타나 있다.

대화처리 데이터셋의 포맷과 관련하여, 대화처리 데이터셋은 slot-filling task를 수행하기 위한 DSTC(Dialog State Tracking Challenge)와 같은 구조화된(structured) 대화코퍼스로부터 Twitter로부터 추출한 비구조화된(unstructured) 대화코퍼스까지 다양하게 구축되고 있다[49].

대화처리 데이터셋의 도메인과 관련하여, 대화처리 데이터셋은 단일 도메인(예, bus timetable, restaurant booking 등) 대화코퍼스로부터 여러 개의 도메인(예, restaurant information과 tourist information)을 연계하는 멀티도메인으로 구축되고 있다[50].

대화처리 데이터셋의 구축 방법과 관련하여, 대화처리 데이터셋은 대부분 크라우드소싱(crowdsourcing)을 이용한 Wizard-of-Oz(WOZ) 방식으로 구축되고 있다[50-52]. 크라우드소싱을 이용하는 WOZ 방식으로 Amazon Mechanical Turk 플랫폼을 활용하고 있는 실정이다[50-52].

대화시스템의 활용도를 높이기 위해서는 대화 데이터셋을 도메인에 따라 쉽게 구축하고 다양한 사람들의 대화를 손쉽게 수집할 수 있는 환경과 수집한 대화데이터를 대화 학습에 필요한 정보를 빠르고 일관성 있게 부착할 수 있는 방법론이 요구되고 있다.

VI. 결론

살펴본 바와 같이 최근 대화처리 연구 동향은 목적지향 및 chit-chat 대화처리 연구만 아니라, 질의응답 연구에서도 심층 신경망 기반의 종단형 대화처리 방법론들이 활발히 연구되고 있다. 신경망 기반 종단형 대화처리 기술은 종래의 pipeline 방법론들과 달리, 목적지향, chit-chat 및 질의응답 대화처리에 모두 동일하게 학습대화코퍼스와 도메인(또는 전문) 지식 정보로 학습하고 학습되는 방식이다. 따라서, 기존 대화 형식이나 목표에 따라 다

른 방법론으로 연구되는 방향에서 동일한 방법론 및 기술로 연구될 것으로 기대된다.

신경망 기반 대화처리 기술은 도메인 지식에 기반한 대화에 적합한 종단형 신경망 모델 연구와 부족한 학습데이터 문제를 극복하는 연구가 활발히 이루어질 것으로 보인다. 또한, 새로운 도메인에 대한 학습대화를 쉽게 수집하고, 대화 및 도메인에 대한 학습 정보를 부착하지 않은 상태로 학습하는 연구도 활발해질 것으로 기대한다.

용어해설

목적지향 대화시스템 대화시스템 구현 의도가 특정 목적(업무나 명령 수행 등)을 이루기 위해 구축된 대화시스템

chit-chat 대화시스템 대화시스템 구현 의도가 재미 위주의 잡담을 하기 위해 구축된 대화시스템

종단형(End-to-end) 시스템 시스템의 기능이 하나의 모델로 이루어져 있는 형태의 시스템

약어 정리

CNN	Convolutional Neural Network
DBN	Deep Belief Network
DCN	Deep Convex Network
DSTC	Dialog State Tracking Challenge
DQN	Deep Q-Network
GRU	Gated Recurrent Unit
HCN	Hybrid Cod Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
WOZ	Wizard-of-Oz

참고문헌

[1] J.D. Williams and S. Young, "Partially Observable Markov Decision Processes for Spoken Dialog Systems," *Comput. Speech Language*, vol. 21, no. 2, 2007, pp. 393-422, doi: 10.1016/j.csl.2006.06.008.

[2] S. Young et al., "Pomdp-Based Statistical Spoken Dialog Systems: A Review," *Proc. IEEE*, vol. 101, no. 5, 2013, pp. 1160-1179, doi: 10.1109/JPROC.2012.2225812.

[3] T. Zhao and M. Eskenazi, "Towards End-to-End Learning for Dialog State Tracking and Management Using Deep Reinforcement Learning," in *Proc. Annu. Meeting Special Interest Group Discourse Dialogue (SIGDIAL)*, Los Angeles, CA, USA, Sept. 2016, pp. 1-10.

[4] A. Bordes et al., "Learning End-to-End Goal-Oriented Dialog," in *Proc. Int. Conf. Learning Representations (ICLR)*, Toulon, France, Apr. 2017, pp. 1-5.

[5] J.D. Williams et al., "Hybrid Code Networks: Practical and Efficient End-to-End Dialog Control with Supervised and Reinforcement Learning," in *Proc. Annu. Meeting Association Comput. Linguistics (ACL)*, Vancouver, Canada, 2017, pp. 665-677.

[6] B. Liu and I. Lane, "An End-to-End Trainable Neural Network Model with Belief Tracking for Task-Oriented Dialog," in *Proc. Annu. Conf. Int. Speech Commun. Association (INTERSPEECH)*, Stockholm, Sweden, Aug. 2017, pp. 2506-2510.

[7] A. Madotto et al., "Mem2seq: Effectively Incorporating Knowledge Bases into End-to-End Task-Oriented Dialog Systems," in *Proc. Annu. Meeting Association Comput. Linguistics (ACL)*, Melbourne, Australia, July 2018, pp. 1468-1478.

[8] B. Dhingra et al., "Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access," in *Proc. Annu. Meeting Association Comput. Linguistics (ACL)*, Vancouver, Canada, 2017, pp. 484-495.

[9] B. Liu and I. Lane, "Iterative Policy Learning in End-to-End Trainable Task-Oriented Neural Dialog Models," in *IEEE Autom. Speech Recogn. Understanding Workshop (ASRU)*, Okinawa, Japan, Dec. 2017, pp. 482-489.

[10] X. Li et al., "End-to-end Task-Completion Neural Dialogue Systems." in *Proc. Int. Joint Conf. Natural Language Process. (IJCNLP)*, Taipei, Taiwan, 2017, pp. 733-743.

[11] I. Ahmed and S. Singh, "AIML Based Voice Enabled Artificial Intelligent Chatterbot." *Int. J. u- e- Service, Sci. Technol.*, vol. 8, no. 2, Feb. 2015, pp. 375-384.

[12] B. Wilcox and S. Wilcox, "Winning the Loebner's," 2014, http://brilligunderstanding.com/Winning.pdf.

[13] O. Vinyals and Q. Le, "A Neural Conversational Model," in *Proc. ICML*, Lille, France, 2015, pp. 1-8.

[14] 황금하 외, "목적지향 대화시스템을 위한 챗봇 연구", 정보처리학회논문지, 제6권 제11호, 2017, pp. 499-507.

[15] J.X. Huang et al., "Improve the Chatbot Performance for the DB-CALL System Using a Hybrid Method and a Domain Corpus," in *Future-Proof CALL: Language Learning Exploration Encounters-Short Papers from EUROCALL*, 2018, pp. 100-105, doi: 10.14705/rpnet.2018.26.820.

[16] L. Shang et al., "Neural Responding Machine for Short-Text Conversation," in *Proc. ACL*, Beijing, China, July 2015, pp. 1577-1586.

[17] J. Li et al., "A Diversity-Promoting Objective Function for Neural Conversation Models," in *Proc. HLT-NAACL*, San Diego, CA, USA, June 2016, pp. 110-119.

[18] T. Zhao et al., "Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders," in *Proc.*

Annu. Meeting Association Comput. Linguistics (ACL), Vancouver, Canada, 2017, pp. 654-664.

[19] A. Sordoni et al., "A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion," in *Proc. Conf. Inf. Knowl. Manag.*, Melbourne, Australia, Oct. 2015, pp. 553-562.

[20] I.V. Serban et al., "Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models," in *Proc. AAAI*, Phoenix, AZ, USA, Feb. 2016, pp. 3776-3783.

[21] I.V. Serban et al., "A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues," in *Proc. AAAI Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 3295-3301.

[22] C. Xing et al., "Hierarchical Recurrent Attention Network for Response Generation," in *Proc. AAAI Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 5610-5617.

[23] C. Xing, et al., "Topic Aware Neural Response Generation," in *Proc. AAAI Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 3351-3357.

[24] M. Ghazvininejad et al., "A Knowledge-Grounded Neural Conversation Model," in *Proc. AAAI Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 5510-5517.

[25] E. Dinan et al., "Wizard of Wikipedia-Knowledge-Powered Conversational Agents," in *Proc. Int. Conf. Learning Representations*, New Orleans, LA, USA, May 2019, pp. 1-8.

[26] S. Zhang et al., "Personalizing Dialogue Agents: I have a dog, do you have pets too?" in *Proc. Association Comput. Lignuistics*, Melbourne, Australia, July 2018, pp. 2204-2213.

[27] J. Weston, "Memory Networks," in *Proc. Int. Conf. Learning Representations*, San Diego, CA, USA, Dec. 2015.

[28] S. Sukhbaatar et al., "End-To-End Memory Networks," in *Proc. NIPS*, Montreal, Canada, Dec. 2015, pp. 1-9.

[29] A. Miller et al., "Key-Value Memory Networks for Directly Reading Documents," in *Proc. Conf. Empirical Methods Natural Language Process.*, Austin, TX, USA, Nov. 2016, pp. 1400-1409.

[30] A. Kumar et al., "Ask Me Anything: Dynamic Memory Networks for Natural Language Processing," in *Proc. Int. Conf. Mach. Learning*, New York, USA, June 2016.

[31] E. Dinan et al., "The Second Conversational Intelligence Challenge (ConvAI2)," 2019, arXiv: 1902.00098, https://arxiv.org/pdf/1902.00098.pdf.

[32] DSTC8, "The Eighth Dialog System Technology Challenge," https://sites.google.com/dstc.community/dstc8.

[33] Jason Weston, "ConvAI2 Competition: Future Work," http://convai.io/NeurIPSConvAI2FutureWork.pptx

[34] S. Ravuri and A. Stolcke, "Recurrent Neural Network and LSTM Models for Lexical Utterance Classification," in *Proc. Interspeech*, Dresden, Germana, Sept. 2015, pp. 135-139.

[35] J.Y. Lee and F. Dernoncourt, "Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks," in *Proc. NAACL-HLT*, San Diego, CA, USA, June, 2016, pp. 515-520.

[36] X. Ma and E. Hovy, "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF," arXiv: 1603.01354v5, 2016.

[37] E. Simonnet et al., "Exploring the use of attention-based recurrent neural networks for spoken language understanding," in *Proc. Mach. Learning Spoken Language Understanding Interactions*, Montreal, Canada, 2015, pp. 1-7.

[38] D. Hakkani-Tür et al., "Multi-Domain Joint Semantic Frame Parsing Using bi-Directional RNN-LSTM," in *Proc. Interspeech*, San Francisco, CA, USA, Sept. 2016, pp. 715-719.

[39] B. Liu and I. Lane, "Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling," in *Proc. Interspeech*, San Francisco, CA, USA, Sept. 2016, pp. 685-689.

[40] C.-W. Goo et al., "Slot-Gated Modeling for Joint Slot Filling and Intent Prediction," in *Proc. NAACL-HLT*, New Orleans, LA, USA, June 2018, pp. 753-757.

[41] Y. Wang et al., "A bi-Model Based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling," in *Proc. NAACL-HLT*, New Orleans, LA, USA, June 2018, pp. 309-314.

[42] Q. Chen et al., "BERT for Joint Intent Classification and Slot Filling," 2019, arXiv: 1902.10909v1.

[43] R. Gupta et al., "An Efficient Approach to Encoding Context for Spoken Language Understanding," 2018, arXiv: 1807.00267v1.

[44] Y.-N. Chen et al., "End-to-End Memory Networks with Knowledge Carryover for Multi-Turn Spoken Language Understanding," in *Proc. Interspeech*, San Francisco, CA, USA, Sept. 2016, pp. 3245-3249.

[45] D. Serdyuk et al., "Towards End-to-End Spoken Language Understanding," 2018, arXiv: 1802.08395v1.

[46] S.K. Choi et al., "Using a Dialogue System Based on Dialogue Maps for Computer Assisted Second Language Learning," in *EUROCALL*, Limassol, Cyprus, Aug. 2016, pp. 106-112.

[47] I.V. Serban et al., "A Survey of Available Corpora for Building Data-Driven Dialogue Systems," 2015, arXiv preprint arXiv: 1512.05742.

[48] I.V. Serban et al., "A Survey of Available Corpora for Building Data-Driven Dialogue Systems," https://breakend.github.io/Dialog-Datasets/.

[49] R. Lowe et al., "The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems," in *SIGDIAL*, Prague, Czech Republic, Spet. 2015. pp. 285-294.

[50] P Budzianowski et al., "Multiwoz-a Largescale Multi-domain Wizard-of-Oz Dataset for Taskoriented Dialogue Modelling," in *Proc. Conf. Empirical Methods Natural Language Process.*, Brussels, Belgium, 2018, pp. 5016-5026.

[51] W.S. Lasecki et al., "Conversations in the Crowd: Collecting Data for Task-Oriented Dialog Learning," in *AAAI Conf.*, Bellevue, WA, USA, 2013, pp. 2-5.

[52] M. Eric et al., "Key-Value Retrieval Networks for Task-Oriented Dialogue," in *SIGDIAL Conf.*, Saarbrucken, Germany, Aug. 2017, pp. 37-49.