

■ 연구과제 요약문

과제명(기간)	Coherent Knowledge Grounding for Open-Domain Dialogue System (2019.05.27. ~ 2019.12.20)
연구책임자	김 건 희 (gunhee@snu.ac.kr)
개요	<p>- 본 연구에서는 지식 기반 대화 시스템 (knowledge-grounded dialogue system)의 성능을 높이는 것을 목표로 한다. 그 중에서도 Facebook Research에서 제안한 현존하는 가장 큰 데이터셋인 Wizard of Wikipedia에서 향상된 성능을 달성하는 것을 목표로 한다. Wizard of Wikipedia 벤치마크에서의 대화는 두 사람이 참여하며, 한명은 wizard, 한명은 apprentice 역할을 맡는다. Wizard는 본인이 대답할 차례에 추가로 retrieved된 문서를 볼 수 있으며, apprentice는 wizard로부터 적극적으로 정보를 얻어내야 한다. 본 데이터셋에서는 wizard를 모델링하는게 최종 목표이며, 이는 다음 두 문제로 나눌 수 있다. (i) 검색된 문서들에서 관련된 문서 및 문장 고르기 (Knowledge Selection), (ii) 대화 이력과 선택된 문장으로 대답 완성하기 (Utterance Prediction). 따라서 본 데이터셋에서 모델의 성능을 향상시키기 위해서는 각 태스크를 푸는 모듈의 성능을 개선시키거나, 두 모듈을 효율적이게 end-to-end로 결합시켜야 한다.</p>
연구개발 결과	<p>- 본 연구에서는 Knowledge Selection 모듈의 성능을 높이는 것을 목표로 한다. 이는 Wizard of Wikipedia 데이터셋에서 모델의 성능이 Knowledge Selection 모듈의 성능에 상당히 의존적이기 때문이다. 이를 위해 처음으로 시퀀스 잠재 변수 모델 (sequential latent variable model)을 최초로 제안하였다. Sequential knowledge transformer (SKT)라 불리는 새로운 모델은 지식 선택에 대한 사전 및 사후 분포를 추적할 수 있다. 따라서, 대화에서의 지식 선택에서 필연적으로 동반하는 모호성을 크게 줄일 수 있으며, 또한 이전 턴의 응답 정보를 더 잘 활용할 수 있게 된다. 실험을 통해 본 방법은 Wizard of Wikipedia 데이터셋에서 지식 선택 및 응답 생성 정확도를 기존 방법들보다 크게 상회함을 보였다. 뿐만 아니라 또 다른 지식 기반 벤치마크인 Holl-E 데이터셋에서도 최고의 성능을 보였다.</p> <p>- 본 연구 결과는 기계학습 최우수 학회 중 하나인 ICLR 2020에서 spotlight paper로 게재 확정되었다. 총 2594편의 제출 논문 중 오직 156편만이 spotlight으로 선택되어 게재율은 6.0%에 불과하였다. 이는 본 연구의 수월성을 잘 보여주고 있다.</p>
활용분야 및 기대효과	<p>- 본 연구를 통해 고안한 모델을 통하여 국제 자연어/인공지능 최우수 학회에 논문을 발표함으로써 딥러닝 및 자연어 분야에서 world-leading research를 하고 있음을 대외적으로 홍보할 수 있다. 또한 knowledge-grounded dialogue system은 고객 응대 시스템, 대화형 검색엔진 등 다양한 산업군에 활용 가능하다.</p>