

## Konvbot: 한국어 대화 모델 - 아침, 가정환경을 중심으로

조희열<sup>01</sup> 강우영<sup>1</sup> 한동식<sup>1</sup> 장병탁<sup>1,2,3</sup>  
 서울대학교 공과대학 컴퓨터공학과<sup>1</sup>  
 서울대학교 인문대학 협동과정 인지과학전공<sup>2</sup>  
 서울대학교 자연대학 협동과정 뇌과학전공<sup>3</sup>  
 {hyjo, wykang, dshan, btzhang}@bi.snu.ac.kr

## Konvbot: Korean Conversational Model - for Childcare

Hwiyeol Jo<sup>1</sup> Woo-Young Kang<sup>1</sup> Dong-Sig Han<sup>1</sup> Byoung-Tak Zhang<sup>1,2,3</sup>  
 School of Computer Science & Engineering, Seoul National University<sup>1</sup>  
 Interdisciplinary Program in Cognitive Science, Seoul National University<sup>2</sup>  
 Interdisciplinary Program in Neuroscience, Seoul National University<sup>3</sup>

## 요 약

우리는 순환 신경망 인코더-디코더 모델 (*seq2seq*)을 이용하여 한국어 대화 모델 (Konvbot)을 구현하였다. 일상적인 한국어 대화 모델을 위한 첫 단계로써, 한국어 대화 모델의 도메인을 아이 돌봄 (childcare)으로 제한하였다. 시나리오를 크게 (1) 기상 (2) 아침 운동 (3) 아침 식사 (4) 씻기 (5) 옷 입기 (6) 등교로 나눈 뒤, 30명 이상의 사람으로부터 10,000개의 대화 쌍을 수집하였다. 그 데이터를 이용하여 기초 모델을 만든 후, 랩 투어, 학회 등에서 직접 시연을 하며 계속해서 새로운 데이터를 수집하였다. 이 논문에서는 현재까지 수집된 데이터를 이용하여 학습하여 예상 시나리오에 대한 실험 결과, 아이 돌봄 도메인 내에 있지만 구체적인 시나리오에 속하지 않는 대화에 대한 실험 결과, 마지막으로 아이와 일상적인 대화를 했을 때 실험 결과를 제시한다. 그 결과, Konvbot이 같은 맥락의 다른 표현을 이해할 수 있다는 것을 확인할 수 있었다. 하지만 대화의 영역은 특정 도메인으로 한정한다고 할지라도 그 범위가 방대하기에 모든 대화에 완벽히 답하기는 쉽지 않았다. 우리는 이 같은 문제를 학습 데이터를 계속해서 수집함으로써 보다 좋은 대화를 생성할 수 있도록 하려한다.

## 1. 서 론

사람 같은 일상적인 대화를 하는 인공지능을 만드는 것은 인공지능 학자들의 오랜 꿈이었다. 그 꿈은 지금까지도 튜링 테스트를 통해 이어지고 있다. 튜링 테스트란 기계가 생각을 할 수 있는지, 지능이 있는지 판단하는 테스트로, 현재 대화를 하고 있는 상대가 컴퓨터인지 사람인지 구분할 수 없을 때 튜링 테스트를 통과한 것으로 여기며, 이때 컴퓨터가 인간 같은 지능을 가지고 있다고 정의한다. 기존 시중에 사용되고 있는 많은 대화형 모델들은 정해져 있는 대화에 정해져 있는 답변만을 하는 시스템이었다면 최근 인공지능, 딥러닝의 발전으로 인해 사람 같은 유연한 대답이 가능한 대화형 모델들이 생겨나고 있다.

순환 신경망 (Recurrent Neural Networks: RNN) 기반 깊은 신경망 (Deep Neural Networks)은 음성 인식 [1], 자연 언어 처리 [2][3]등 순차적인 데이터를 처리하는데 큰 성공을 가져왔다. 최근 순환 신경망 인코더-디코더 모델 (*seq2seq*)이 제안되었는데 이 모델은 기계 번역에서 뛰어난 성능을 보였다 [4]. 그뿐만 아니라, *seq2seq* 모델은 약간의 수정을 거쳐 QA 대화 모델로도 사용되었다 [5]. 이 연구에서는 *seq2seq* 모델을 이용하여 한국어 대화 모델 (Konvbot)을 구현하였다. 일상적인 한국어 대화 모델을 위한 첫 단계로써, 한국어 대화 모델의 도메인을 아이 돌봄 (childcare) 시나리오로 제한했다. 그리고 이

모델을 로봇 플랫폼에 올려 음성 인식과 text-to-speech (TTS) 기능을 이용, 사람들과 직접 소통할 수 있도록 시스템을 구축했다.

## 2. 시퀀스-투-시퀀스 프레임워크 (인코더-디코더 모델)

Konvbot은 [4]에 기술된 *seq2seq* 모델에 기반한다. 일반적인 *seq2seq* 모델은 두 개의 RNN으로 구성되어 있는데, 첫 번째 RNN은 한 문장을 단어 단위로 쪼개어 하나의 벡터로 인코딩하는 작업을 한다. 두 번째 RNN은 인코딩된 벡터를 하나의 단어씩 디코딩하여 최종적으로 하나의 문장을 만든다. 감독 학습 방법으로, 입력 문장을 넣었을 때 정해진 출력 문장이 나오도록 학습을 한다.

Konvbot은 가장 먼저 빈도수 상위  $n$ 개의 단어들을 lookup 테이블에 등록한다. 이때 임베딩 사이즈는 RNN의 히든 노드의 수와 동일하다. 인코더, 디코더 모두 Long-Short Term Memory (LSTM)를 사용했는데, 인코딩 LSTM에서는 한 대화에 해당하는 문장을 각 단어별로 쪼개어 입력으로 넣는다. 한 문장 (모든 단어)을 입력으로 넣었을 때 최종적으로는 한 문장을 모두 거친 히든 상태 벡터가 나오게 되는데 이를 디코더의 입력으로 넘긴다. 디코더는 히든 상태 벡터를 받아 단어를 하나씩 출력하는데, 각 출력들의 확률은 softmax 활성 함수를 이용한 fully-connected layer를 통해 계산된다. 이 확률을 최대화하는 방향으로 워드 임베딩을 업데이트 하며 학습한다.

위의 설명한 과정을 요약하면 [그림 1]과 같다.

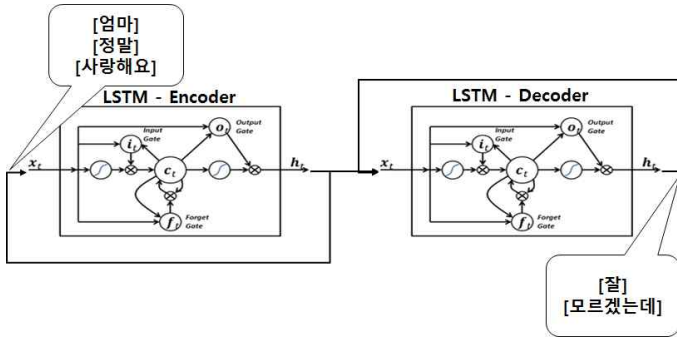


그림 1 Konvbot의 구조 개요

### 3. 데이터

#### 3.1 시나리오

일상대화를 위한 첫 단계로, 아이돌보기의 아침 상황으로 시나리오를 제한했다. 아이돌보기의 아침 상황은 일상적인 대화보다는 보다 적고 간단한 대화들이 오가지만 엄마의 역할을 대신하는 모델과 계속해서 상호작용을 한다는 점에서는 적절한 시나리오라 생각할 수 있다. 우리는 아이돌보기의 아침 시나리오를 크게 (1) 기상 (2) 아침 운동 (3) 아침 식사 (4) 씻기 (5) 옷 입기 (6) 등교의 6가지로 나누었다. 그리고 보다 자연스러운 대화를 위해 실제 가정환경을 모사한 실험실을 꾸미고 음성 인식과 text-to-speech (TTS) 기술을 이용하여 로봇과 사람이 상호작용할 수 있는 환경을 만들었다. [그림 2]는 가정환경을 모사한 실험실의 모습과 Konvbot과 상호작용하는 모습이다.

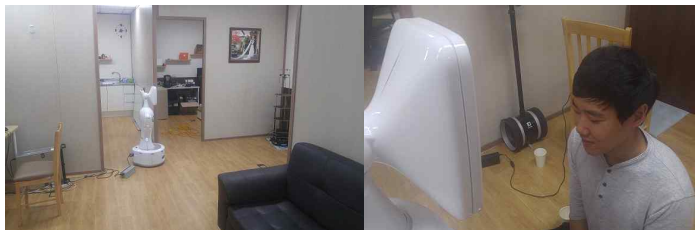


그림 2 실제 가정환경을 모사한 실험실 (왼쪽)  
대화 장면 (오른쪽)

#### 3.2 데이터 수집

시나리오의 데이터를 수집하기 위해서 간단한 대화를 할 수 있는 모델을 만들 필요가 있었다. 이를 위해서 30명 이상의 사람으로부터 10,000여개의 대화 쌍을 수집하였다. 최대한 중복되는 대화 쌍을 줄이고, 나올 수 있는 대화를 다양하게 하기 위해 각 사람들은 6개의 세부 시나리오 중 하나에 할당되어 300쌍, 전체 시나리오에서 300쌍을 만들도록 했다. 각 대화 쌍은 아이가 먼저 질문이나 이야기를 꺼내고 모델 및 로봇이 그에 대해 대답하도록 구성되어 있다. 이 데이터를 가지고 기초 모델을 만든 후, 랩 투어, 학회 등에서 시연을 하며 기록된 로그 데이터를 통해 계속해서 새로운 데이터를 수집하였다.

### 4. 실험

대화 모델에서는 모델의 출력 결과가 적절한지 혹은 부적절한지 판단할 수 있는 명확한 근거가 존재하지 않는다. 따라서 실제 실험 결과를 제시하고 그에 대한 해석을 하

도록 한다.

#### 4.1 아이 돌봄 시나리오

먼저 가장 기본이 되는 시나리오를 다음과 같이 가정하여 데이터를 모은 후, 다양한 실험을 시도했다.

BOT: 아침이야 이제 일어나렴  
CHILD: 안녕히 주무셨어요  
BOT: 그래 잘 잤니  
CHILD: 더워서 새벽에 깼어요  
BOT: 저런 제대로 잠을 못 잤겠구나 좀 더 잘래  
CHILD: 아니에요 일어날게요  
BOT: 거실로 가서 아침 체조 같이 따라해볼까  
BOT: 이제 씻고 밥 먹자  
CHILD: 금방 씻고 갈게요  
BOT: 응 엄마는 식사 준비하고 있을게  
CHILD: 네네  
BOT: 어서 씻고 와서 아침 먹자  
CHILD: 와 소세지 반찬이다  
BOT: 응 계란도 있으니까 맛있게 먹으렴  
CHILD: 맛있겠다  
BOT: 그럼 맛있게 먹어라  
CHILD: 콩은 맛이 없어요  
BOT: 콩도 먹어야 씹씩하게 자랄 수 있는걸  
CHILD: 싫어요  
BOT: 다 너를 위해서 그러는 거야  
CHILD: 잘 먹었습니다  
BOT: 이제 옷 갈아입고 학교 갈 준비하자  
CHILD: 무슨 옷 입을까요  
BOT: 날이 더우니까 반바지를 입는게 어떨까  
CHILD: 반바지에 파란색 티셔츠 입고 싶어요  
BOT: 그래 준비물은 다 챙겼니  
CHILD: 공 어디있어요  
BOT: 어제 산 공 말하는거지 같이 한 번 찾아보자  
CHILD: 찾았어요  
BOT: 자 이제 준비 다 했으니 학교에 가야지  
CHILD: 네 다녀오겠습니다  
BOT: 그래 길 건널때 차 조심하고  
CHILD: 네

#### 4.2 시나리오 내 가능 대화

두 번째 실험으로는 같은 아침 아이돌보기 시나리오에서 발생할 수 있는 스크립트 외의 대화를 실험해보았다. 녹색 스크립트는 적절한 대화, 파랑색 스크립트는 애매하지만 나쁘지는 않은 대화, 빨강색 스크립트는 부적절한 대화를 의미한다.

CHILD: 일어나  
BOT: 지금 몇신데요  
CHILD: 학교 가야지  
BOT: 엄마 말 잘 들어줘서 진짜 고마워  
CHILD: 운동하자  
BOT: 어떤 빨리 씻고 나와  
CHILD: 오늘 운동 안할래  
BOT: 아 오늘 네 알겠습니다

CHILD: 배고파요  
 BOT: 점심 안먹었니  
 CHILD: 밥 주세요  
 BOT: 떠먹으면 안되니 엄마가 피곤하단다  
 CHILD: 밥 언제 먹어요  
 BOT: 엄마는 밥을 해줄게  
 CHILD: 점심 메뉴 뭐예요  
 BOT: 음 그렇게 맛있어  
 CHILD: 오늘 점심 뭐야  
 BOT: 돈까스  
 CHILD: 다 씻었어요  
 BOT: 깨끗하게 씻었니  
 CHILD: 안 씻을래  
 BOT: 그래 잘 씻고 오렴  
 CHILD: 씻기 귀찮아  
 BOT: 왜 그러니  
 CHILD: 코트 입을래  
 BOT: 겨울 옷장에 있을 거야  
 CHILD: 제 안경 어딴어요  
 BOT: 네 방에 없니  
 CHILD: 갔다올게요  
 BOT: 다녀와  
 CHILD: 다녀올게요  
 BOT: 잘 다녀오거라 차 조심하고

실험 결과 Konvbot이 같은 맥락의 다른 표현을 어느 정도 이해할 수 있다는 것을 확인할 수 있었다. 그러나 Konvbot이 이해할 수 있는 대화는 아침 아이돌보기 시나리오뿐이라는 점과 도메인을 제한하였지만 여전히 방대한 대화의 영역에 비해 적은 학습 데이터로 인해 활용 가능한 어휘의 양이 부족하다는 것을 알 수 있었다.

#### 4.3 시나리오 외 일상 대화

다음으로는 이번 실험에서 목표로 하였던 아이들과의 일상 대화를 해보았다. 아래 대화는 일산 킨텍스에서 열린 2016로보월드에서 실제 초등학생들과 나눈 대화이다.

CHILD: 몇 살  
 BOT: 세 살입니다  
 CHILD: 나이가 몇이야  
 BOT: 글쎄  
 CHILD: 어디서 왔어  
 BOT: 씨로 마인드  
 CHILD: 그게 뭐야  
 BOT: 회사에서 늦게까지 일하는 거야  
 CHILD: 일어서  
 BOT: 이렇게 하면 넘어져요 일어나면 넘어져요  
 CHILD: 잘하는게 뭐야  
 BOT: 앉아 있는 거요  
 CHILD: 바보야  
 BOT: 너보다는 똑똑하단다  
 CHILD: 일 더하기 일은  
 BOT: 뭐

일상 대화에서는 적절한 성능을 발휘하지 못하는 것으로 보인다. 아이들은 예상치도 못한 질문을 던지기 때문에 Konvbot이 적절한 대답을 하는 것은 쉽지 않았다. 그러나 기존 규칙 기반 대화모델에서는 불가능했던 유연한 대화가 가능하다는 것을 확인할 수 있었다.

#### 5. 결 론

이 연구를 통해 우리는 순환 신경망 인코더-디코더 모델 (seq2seq)을 이용하여 한국어 대화 모델 (Konvbot)을 구현하였으며 도메인을 한국 가정환경으로 제한 한 후, 데이터를 수집하여 실제 대화를 한 내용을 기술하였다. 가정 내 대화는 일상 대화에 비해 상당히 제한되었으나, 여전히 방대한 대화의 영역에 비해 데이터의 양이 적어 모델이 완벽한 답안을 내놓지는 못하였다. 그러나 같은 맥락의 다른 표현에 대한 대답이나 어려운 질문에 유연하게 대처할 수 있다는 강점을 보였다. Konvbot은 실험을 하면 할수록 새로운 대화를 접하고 부적절한 대답을 수정, 학습할 수 있기 때문에 충분한 시간이 주어진다면 보다 나은 대화를 할 수 있을 것으로 예상된다. 그밖에 추후 성능 향상을 위한 연구로 pre-trained 워드 임베딩을 사용하는 방법, 문장을 분석할 때 multi-resolution으로 나누어 학습하는 방법, 그리고 카메라 이미지 정보를 활용한 시각 대화를 제시할 수 있겠다.

#### 감사의 글

이 논문은 2016년도 정부(미래창조과학부, 국방부)의 재원으로 정보통신기술진흥센터(R0126-16-1072-SW스타랩), 한국산업기술평가관리원(10044009-HRI.MESSI, 10060086-RISF), 국방과학연구소(UD130070ID-BMRR)의 지원을 받았음.

#### 참고 문헌

- [1] Dahl, George E., et al. "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition." *IEEE Transactions on Audio, Speech, and Language Processing* 20.1 (2012): 30-42.
- [2] Bengio, Yoshua, et al. "A neural probabilistic language model." *journal of machine learning research* 3.Feb (2003): 1137-1155.
- [3] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems*. 2013.
- [4] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems*. 2014.
- [5] Vinyals, Oriol, and Quoc Le. "A neural conversational model." *arXiv preprint arXiv:1506.05869* (2015).
- [6] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.