



Childbot: A Conversational Assistant for Childcare

Hwiyeol Jo

Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
hyjo@bi.snu.ac.kr

Woo-Young Kang

Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
wykang@bi.snu.ac.kr

Dong-Sig Han

Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
dshan@bi.snu.ac.kr

Byoung-Tak Zhang

Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
btzhang@bi.snu.ac.kr

Abstract— In this paper, we used the RNN Encoder-Decoder (*seq2seq*) model in Korean conversational robot (Konvbot). As a first step of being a general conversational model, we restricted our domain to childcare situation which includes six circumstances: 1) waking up, 2) morning exercise, 3) having breakfast, 4) taking a shower, 5) wearing clothes, and 6) going to school. We collected about 10,000 dialogue pairs for this scenario from more than 30 people. With the data, we implemented the base conversational model, which is used for collecting more dialogues in real environments (e.g. lab tour, conference demo). We present experiments in our expected scenario as well as general conversations which are out-of-script, and finally, real conversation with children. The result showed that our model could catch slight different expressions in the similar context, but it can cover only specific domain and has low vocabulary due to the small amount of training data. With more real experiments, we can collect data from experimenters, the better conversation the model will generate.

Keywords—chatbot; conversational model; *seq2seq*;

I. INTRODUCTION

Deep neural network models based on the recurrent neural networks (RNN) have shown great success in many application such as speech recognition [1] and natural language processing (NLP) tasks [2][3]. The RNN Encoder-Decoder (*seq2seq*) model has shown excellent performance in the machine translation task [4]. The *seq2seq* model has an advantage of sentence based machine translation, which is generating a target sentence from a source sentence directly. Not only that, research showed the *seq2seq* model can be used as a conversational model with little modification [5]. So, we adapted the *seq2seq* model in our Korean conversational robot whose goal is for childcare with a sentence to sentence conversations. After building our conversation model using the *seq2seq*, we embedded it into robot platform to interact with people directly through voices using the speech recognition technique and the text to speech technique. As a first step of being a general conversational model, we restricted our target domain task to childcare situation at home environment.

II. METHOD

A. Scenario

We restricted the global domain of natural language conversation problems to the situation of childcare at every morning, which is smaller and simpler (we expected), but commonly desired for dialogues with physically interacting robots. We categorized this scenario into six circumstances: 1) waking up, 2) morning exercise, 3) having breakfast, 4) taking a shower, 5) wearing clothes, and 6) going to the school. For qualitatively testing seamlessness of our implementation, we used an actual robot and a single room which was adorned like a home environment and attempted to talk with the robot in given scripts using a non-commercial Korean speech recognition and text-to-speech software.

B. Data

For collecting dialogue data for the scenario, more than 30 people produced synthetically generated sentences; each person was assigned to make 300 pairs of dialogues for the scenario using his/her imagination, and 300 instruction-operation pairs for constructing a personality of the robot and future research. Also, we made the data that the speaker in the first line and last line are always different, and the latter is for the model. We collected about 10,000 sentences pairs for the experiments. With the data, we implemented the base conversational model, which is used for collecting more dialogues in real environment. For example, we brought our model in lab tours and conferences, and got more dialogues from visitors, leaving conversation logs.

C. Sequence-to-Sequence Frameworks

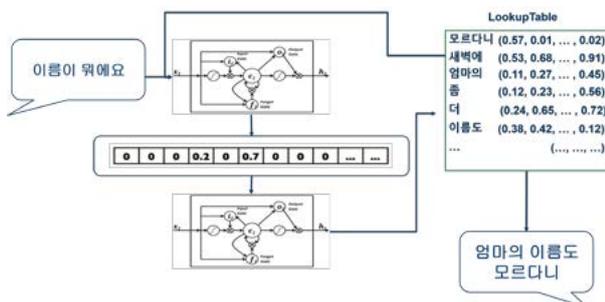
Our approach is based on sequence-to-sequence (*seq2seq*) framework described in [5]. The model consists of two recurrent neural networks. The encoder network reads one word in the input sequence at a time and also predicts the output sequence one word at a time. It is trained by maximizing



the probability output sequence for estimating the true output in the data.

Specifically, at the bottom, we have a word embedding matrix for Korean words initialized arbitrarily, which could be pre-trained word representation in the English setting. The encoding LSTM (Long-Short Term Memory) [6] just receives input words successively and does not produce any other outputs except hidden states. The decoding network uses the final hidden state of the encoder as its previous hidden values, and also use its last output as current input. The probability of output of the decoder is calculated by additional fully-connected layer with softmax activation. Fig. 1 summarized the architecture.

Fig. 1. Illustration of proposed architecture.



III. EXPERIMENT & RESULTS

A. Childcare Scenario

The followings are expected scenario in our childcare domain. We present simplified version.

Bot: Good morning, wake up now.
Child: Good morning.
Bot: Yeah, did you sleep well?
Child: I couldn't sleep well because of hot weather.
Bot: Then, shall we exercise?
Child: Sure.
Child: Wow, sausage! I love it.
Bot: Fried eggs are also there.
Child: Which clothes would I wear?
Bot: How about short pants? It's hot in these days.
Child: Where is my football?
Bot: Let's find together.
Bot: If finish to prepare, you should go to school.
Child: Goodbye.
Bot: Be careful.

B. General Conversation

After fitting the scenario, we tested our conversation model in out-of-script conversation. Green color means reasonable answer; Blue means ambiguous but not bad; Red means it does not make sense.

Child: I'm hungry
Bot: Don't you have lunch?
Child: Give me a dish

Bot: Won't you have it by yourself? I'm tired.

Child: When will we have a meal?

Bot: Why not?

Child: How old are you?

Bot: None of your business!

Child: What is for lunch?

Bot: It is in closet

Child: Where is my glasses?

Bot: Isn't it in your room?

The result showed that our model can catch slight different expressions in the similar context, but it can cover only specific domain and has small vocabulary size due to the small amount of training data.

C. Real Conversation with Child

Child: Do you like a white radish?

Bot: No, give me more.

Child: Where are you come from?

Bot: Yes.

Child: Stand up.

Bot: In this way, I will fall.

Child: What are you good at?

Bot: Just sitting

Child: You, stupid

Bot: Better than you

Child: What is 1+1?

Bot: What?

In this experiment, children asked unsuspected questions to our model. Although our model tried to find the best answer in its vocabulary, the answer was usually not good.

IV. CONCLUSION

In this paper, we proposed the conversational robot using the Korean language in restricted domain. Using the RNN Encoder-Decoder model (*seq2seq*) [5], we could implement sentence based conversation model. We trained our model with relatively small amount of data, so the model cannot generate appropriate answers in many cases. However, with more real experiments, we can collect data from experimenters, the better conversation the model will generate.

REFERENCES

- [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] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078*. 2014.
- [5] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems*. 2014.
- [6] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.