Representation Learning of Logical Words via Seq2seq Learning from Linguistic Instructions to Robot Actions

Tatsuro Yamada
Shingo Murta
Hiroaki Arie
Tetsuya Ogata
Waseda University
Tokyo, Japan
yamadat@idr.ias.sci.waseda.ac.jp
{murata,arie}@sugano.mech.waseda.ac.jp
ogata@waseda.jp

ABSTRACT

Existing robot experiments have investigated how machine learning models learn and represent symbol-grounding relationships from co-occurence of linguistic and sensorimotor sequences. Such investigations have dealt mainly with grounded words, such as verbs, objectives, adjectives and adverbs that directly correspond to objects in the environment, to robot motions, or to their certain features. In contrast, this study includes logical words, which are not grounded in the world directly but contribute to the construction of meaning as logical operators (“true", “false", “and", “or", etc.), alongside grounded words. In our experiment, we built and trained a sequence-to-sequence (seq2seq) learning model that translates linguistic instructions to robot actions. We report how the model learns to represent logical operations from its experience.

HAI'17, October 2017, CITEC Bielefeld, Germany

CCS CONCEPTS

• **Computing methodologies → Neural networks**; **Learning latent representations**; *Cognitive robotics*;

KEYWORDS

symbol grounding; recurrent neural networks; sequence-to-sequence learning; human–robot interaction; logical operations

ACM Reference Format:

Tatsuro Yamada Shingo Murta Hiroaki Arie Tetsuya Ogata. 2017. Representation Learning of Logical Words via Seq2seq Learning from Linguistic Instructions to Robot Actions. In *Proceedings of The 5th annual International Conference on Human-Agent Interaction (HAI'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

INTRODUCTION

Investigating possible representations of symbol grounding relations [2] in a bottom-up constructive manner is important for understanding how human languages work in practice and for building intelligent communicative agents [1, 6]. From this point of view, many studies have conducted robot experiments in which machine learning architectures (e.g., probabilistic models, neural networks) develop various representations of relations between language and its referents in the robot's world via iterative experiences [4, 5, 8]. So far, such studies have mainly considered only grounded words, such as verbs, objectives, adjectives and adverbs that are directly grounded in objects in the environment, in robot motions, or in their certain features. However, language expressions also include logical words, such as "not", "and", and "or", which are not directly grounded in the real world but instead act as logical operators in the construction of the meaning of sentences. This study handles logical words simultaneously with grounded words. We build a recurrent neural network (RNN) model and train it by sequence-to-sequence learning [7]. We report how the model learns to represent logical operations as its functional dynamics from experiences of responding to linguistic instructions by generating robot actions in response.

METHOD

Model

We built a three-layer LSTM-RNN model [3]. At each time step, the model receives a word, visual information, the robot's current joint angles, and the context encoded in its internal memory. From these, the model generates the joint angles at the next time step. Figure 1 shows an overview of the model working after learning. After encoding a sentence with sensorimotor information over several

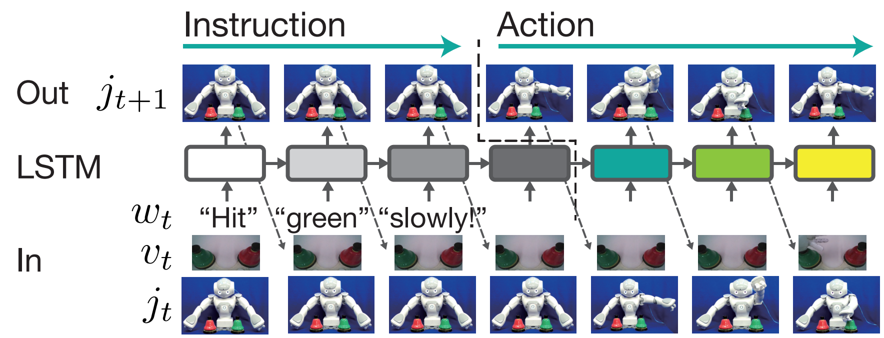

Figure 1: Overview of working of the RNN model, which translates linguistic instructions to robot action sequences.

steps, the model generates an appropriate motion sequence. The model is trained in a supervised manner.

Tasks

We designed two tasks, the flag task and the bell task. Both require the robot to understand instructions that include both grounded and logical words. For example, in the flag task, the robot might receive a sentence "do raise red" or "don't raise red". In this context "red", a grounded word, refers to a robot hand that is grasping a red flag. In contrast, "do" and "don't" are ungrounded logical words. If the other parts in these sentences are the same, the meanings is reversed by swapping "do" and "don't". The robot can also receive sentences such as "do raise red and green" and "do lower green or blue" that include other logical words ("and" and "or"). We train the model with these tasks and investigate the learned representations of these logical words.

RESULTS

We briefly report some of the results here[1]. The graphs shown in this section visualize the learned representations by using principal component analysis (PCA) to reduce the dimensionality of the internal state space of the RNN. Figure 2 shows the embedding of "do" and "don't". By nonlinear transformation of the RNN, sentences that share a meaning despite their elements being orthogonal to each other in the input space are embedded close to one another. For example, "do raise" and "don't lower" are both embedded in the left side of the area. In contrast, "don't raise" and "do lower" are in the right side of the area. Figure 3 shows the representation of "and" and "or". The word "or", which

[1] In the workshop, we will report the results of analysis in more detail with visualization.

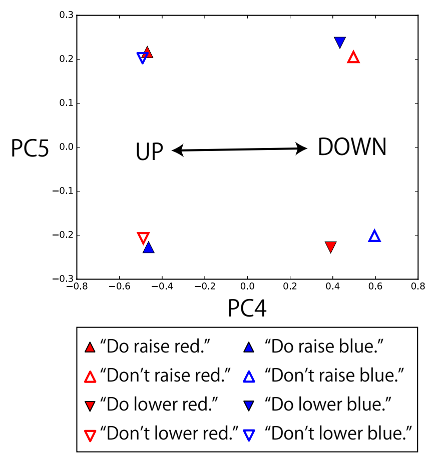

Figure 2: The embeddings of "do" and "don't".

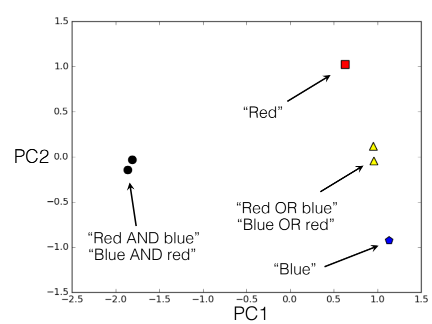

Figure 3: The embeddings of "and" and "or".

instructs the robot to raise (lower) only one of the arms is embedded in the middle area, between "red" and "blue". This suggests that "or" is represented as an unstable point of the RNN's dynamical system.

SUMMARY AND FUTURE WORK

We built and trained an RNN model to investigate how the model learns and represents logical operations from experiences of translating linguistic instructions into robot actions. In future work, we will confirm whether these representations can be achieved even when the task complexity is higher. We will also investigate a method for analyzing the dynamical aspects of the representations, particularly how the representation changes during receiving an instruction or generating an action, to address one shortcoming of using PCA, which is that PCA statically visualizes the states.

ACKNOWLEDGMENTS

This work was supported by JSPS Grant-in-Aid for Young Scientists (A) (No. 16H05878), JSPS Grant-in-Aid for JSPS Research Fellows (No. 17J10580), and by JST CREST Grant Number JPMJCR15E3.

REFERENCES

[1] Angelo Cangelosi, Giorgio Metta, Gerhard Sagerer, Stefano Nolfi, Chrystopher Nehaniv, Kerstin Fischer, Jun Tani, Tony Belpaeme, Giulio Sandini, Francesco Nori, Luciano Fadiga, Britta Wrede, Katharina Rohlfing, Elio Tuci, Kerstin Dautenhahn, Joe Saunders, and Arne Zeschel. 2010. Integration of Action and Language Knowledge : A Roadmap for Developmental Robotics. *IEEE Transactions on Autonomous Mental Development* 2, 3 (2010), 167–195.

[2] Stevan Harnad. 1990. The Symbol Grounding Problem. *Physica D:* 42, 1-3 (1990), 335–346.

[3] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[4] Joe Nishihara, Tomoaki Nakamura, and Takayuki Nagai. 2016. Online Algorithm for Robots to Learn Object Concepts and Language Model. 8920, c (2016), 1–15. https://doi.org/10.1109/TCDS.2016.2552579

[5] Francesca Stramandinoli, Davide Marocco, and Angelo Cangelosi. 2017. Making sense of words : a robotic model for language abstraction. *Autonomous Robots* 41, 2 (2017), 367–383. https://doi.org/10.1007/s10514-016-9587-8

[6] Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh. 2016. Symbol Emergence in Robotics: A Survey. *Advanced Robotics in press* (2016). arXiv:arXiv:1509.08973v1

[7] Oriol Vinyals and V. Quoc Le. 2015. A Neural Conversational Model. In *Proceedings of the 31st International Conference on Machine Learning*. arXiv:1506.0586 http://arxiv.org/abs/1506.0586

[8] Tatsuro Yamada, Shingo Murata, Hiroaki Arie, and Tetsuya Ogata. 2016. Dynamical Integration of Language and Behavior in a Recurrent Neural Network for HumanâĂŞRobot Interaction. *Frontiers in neurorobotics* 10, 5 (2016), 1–17. https://doi.org/10.3389/fnbot.2016.00005