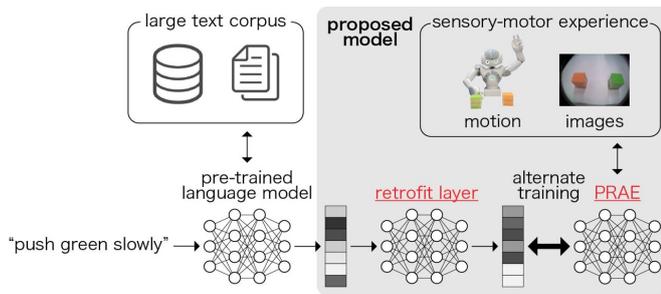# An artificial neural network to acquire grounded representations of robot actions and language

11 May 2021, by Ingrid Fadelli

The researchers' proposed framework transforms word embeddings by observing robot actions and generates appropriate actions for unseen words using the non-linear layer (retrofit layer) and bidirectional translation model for actions and descriptions (Paired Recurrent Autoencoder (PRAE)). Credit: Toyoda et al.

To best assist human users while they complete everyday tasks, robots should be able to understand their queries, answer them and perform actions accordingly. In other words, they should be able to flexibly generate and perform actions that are aligned with a user's verbal instructions.

To understand a user's instructions and act accordingly, robotic systems should be able to make associations between linguistic expressions, actions and environments. Deep neural networks have proved to be particularly good at acquiring representations of linguistic expressions, yet they typically need to be trained on large datasets including robot actions, linguistic descriptions and information about different environments.

Researchers at Waseda University in Tokyo recently developed a deep neural network that can acquire grounded representations of robot actions and linguistic descriptions of these actions. The

technique they created, presented in a paper published in *IEEE Robotics and Automation Letters*, could be used to enhance the ability of robots to perform actions aligned with a user's verbal instructions.

"We are tackling the problem of how to integrate symbols and the real world, the 'symbol grounding problem,'" Tetsuya Ogata, one of the researchers who carried out the study, told TechXplore. "We already published multiple papers related this problem with robots and neural networks."

The new deep neural network-based model can acquire vector representations of words, including descriptions of the meaning of actions. Using these representations, it can then generate adequate robot actions for individual words, even if these words are unknown (i.e., if they are not included in the initial training dataset).
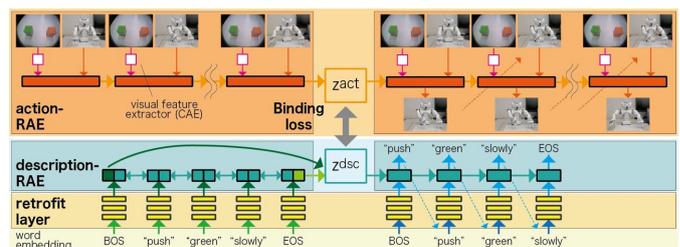
Model overview. The proposed model consists of two recurrent autoencoders and a retrofit layer. Credit: Toyoda et al.

"Specifically, we convert the word vectors of the deep learning model pre-trained with a text corpus into different word vectors that can be used to describe a robot's behaviors," Ogata explained. "In normal language-corpus learning, similarity vectors

are given to words that appear in similar contexts so the meaning of the appropriate action cannot be obtained. For example, 'fast' and 'slowly' have similar vector representations in the language, but they have opposite meanings in the actual action. Our method solves this problem."

Ogata and his colleagues trained their model's retrofit layer and its bidirectional translation model alternately. This training process allows their model to transform pre-trained word embeddings and adapt them to existing pairs of actions and associated descriptions.

"Our study suggests that the integration learning of language and action could enable vector representation acquisitions that reflect the real-world meanings of adverbs and verbs, including unknown words, which are difficult to acquire in deep learning models using only a large text corpus," Ogata said.

In initial evaluations, the deep learning technique achieved highly promising results, as it could generate robot actions from previously unseen words (i.e., words that were not paired with corresponding actions in the dataset used to train the model). In the future, the new model could enable the development of robots that are better at understanding human instructions and acting accordingly.

"This study was the first step of our research in this direction and there is still a lot of room for improvement in linking language and behavior," Ogata said. "For example, it is still difficult to convert some words. In this research, the number of robot motions was small, so we would like to increase the flexibility of the robot to handle more complex sentences in the future."

More information: Embodying pre-trained word embeddings through robot actions. *IEEE Robotics and Automation Letters*(2021). DOI: 10.1109/LRA.2021.3067862.

Paired Recurrent Autoencoders for Bidirectional Translation between Robot Actions and Linguistic Descriptions. *IEEE Robotics and Automation Letters (RA-L)*(2018). DOI: 10.1109/LRA.2018.2852838.

Representation Learning of Logic Words by an RNN: from Word Sequences to Robot Actions. *Frontiers in Neurorobotics*(2017) DOI: 10.3389/fnbot.2017.00070.

Two-way Translation of Compound Sentences and Arm Motions by Recurrent Neural Networks. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2007)* (2007). DOI: 10.1109/IROS.2007.4399265.

APA citation: An artificial neural network to acquire grounded representations of robot actions and language (2021, May 11) retrieved 4 December 2021 from https://techxplore.com/news/2021-05-artificial-neural-network-grounded-representations.html