Behavioral Indoor Navigation With Natural Language Directions

X. Zang[1], M.Vázquez[1], J.C. Niebles[1], A. Soto[2], S. Savarese[1]

[1]Stanford University

[2]P. Universidad Católica de Chile

Introduction

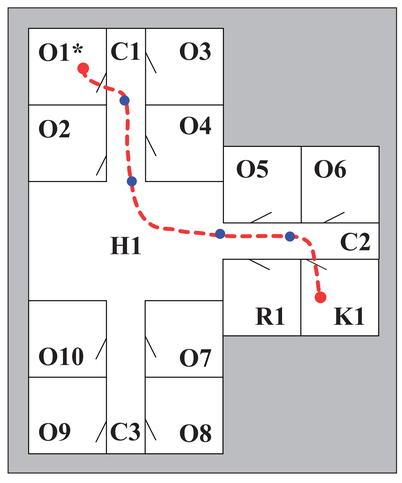
Man-made environments are composed of *navigational structures*, such as corridors or stairs, that in turn are intended to connect meaningful neighboring places, such as rooms or halls.

We hypothesize is that by providing robots with suitable abilities to understand the world at this semantic level, it is possible build robust navigational systems.

Behavioral Navigation Approach

The proposed robot navigation approach [1] revisits the ideas of behavioral robot control [2] and early topological map representations [3].

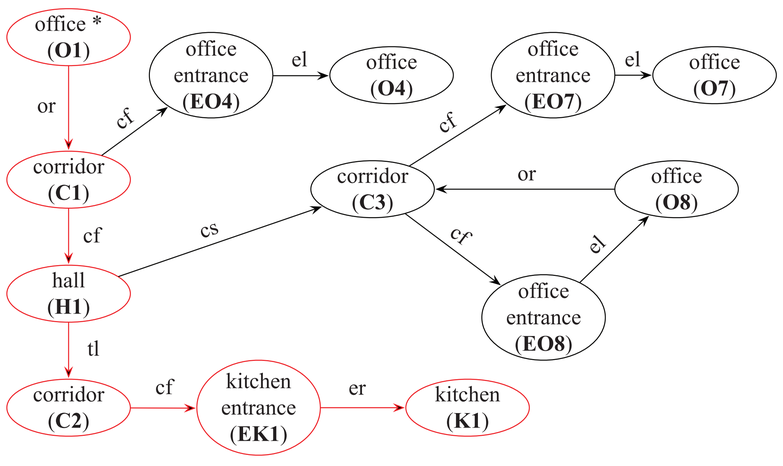
In our approach, complex navigation routes are composed of simple, parameterized visuo-motor behaviors that leverage the semantic structure of indoor environments. These routes are generated by planning over a behavioral graph that encodes relations among navigational structures.



O1* Start location

Route

Behavioral Graph Representation



Following Navigation Instructions

Inspired by prior work [4, 5], we cast the problem of following navigation instructions in natural language as a machine translation task. We perform end-to-end learning to translate commands into navigational routes.

Translating Natural Language

We estimate a function *f* that maps navigation commands to a sequence:

$$S = \langle p1, b1, R1, p2, b2, R2, \ldots, pn \rangle$$

of places p, behaviors b, and reference sequences R. These reference sequences:

$$R = \langle \text{“(”}, l1, r1, l2, r2, \ldots, \text{“)”} \rangle$$

are ordered collections of reference actions r grounded on landmarks l. These references guide the execution of the prior behavior b in S.

As an initial proof of concept, we implemented the function *f* as a differentiable sequence-to-sequence deep learning model with attention [6]. We trained the model using an adaptation of the learning environment DeepMind Lab.

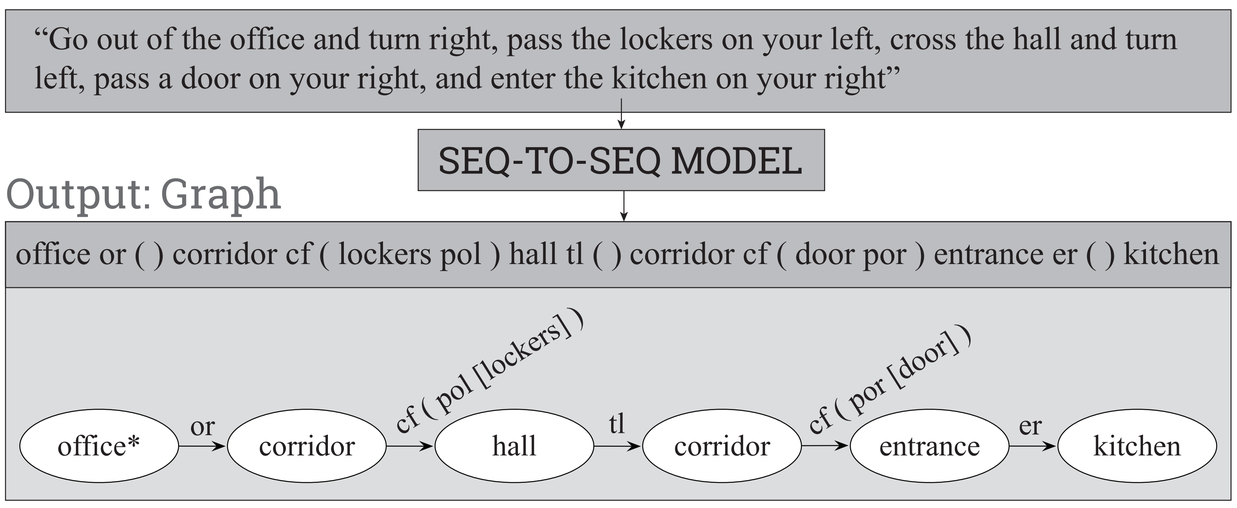
Experiments

We evaluated our approach on a dataset of 16, 370 unique examples using 5-fold cross-validation. The target sequences in the dataset were composed of 40.18 elements on average (STD = 9.75).

# GRU Units	Avg. Accuracy	Std. Dev.
64	99.53%	0.01
128	99.99%	0.0002
256	99.99%	0.0002

With less than 32 GRU units, the performance dropped signicantly and learning was unstable.

Input: Navigation Commands in Natural Language



Acknowledgements

The Toyota Research Institute (TRI) provided funds to assist with this research, but this paper solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity. We are also thankful to G. Sepúlveda for his assistance with data collection.

Contact

marynelv@stanford.edu

References

[1] G. Sepulveda, J. C. Niebles, and A. Soto. *A Deep Learning Based Behavioral Approach to Indoor Autonomous Navigation.* ICRA'18.

[2] R. Brooks. *A robust layered control system for a mobile robot.* IEEE J. Robot. Autom. 2, 2, 1 (1986), 14–23.

[3] B. Kuipers. *Modeling spatial knowledge.* Cogn. Sci., v.2, 1978

[4] C. Matuszek, D. Fox, and K. Koscher. *Following Directions Using Statistical Machine Translation*. HRI'10.

[5] H. Mei, M. Bansal, and M. R. Walter. *Listen, Attend, and Walk: Neural Mapping of Navigational Instructions to Action Sequences.* AAAI'16.

[6] O. Vinyals, L. Kaiser, T. Koo, S. Petrov, I. Sutskever, and G. E. Hinton. Grammar as a Foreign Language. CoRR abs/1412.7449, 2014.