Compositional Data and Task Augmentation for Instruction Following

SpLU-RoboNLP 2021

Soham Dan* Xinran Han* Dan Roth

{sohamdan, hxinran, danroth}@seas.upenn.edu

Penn University of Pennsylvania

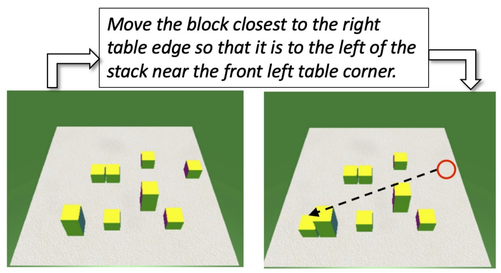
HARVARD UNIVERSITY

COGNITIVE COMPUTATION GROUP

ABSTRACT

Executing natural language instructions in a physically grounded domain requires a model that understands both spatial concepts such as *left of* and *above*, and the compositional language used to identify landmarks and articulate instructions relative to them. In this paper, we study instruction understanding in the blocks world domain. Given an initial arrangement of blocks and a natural language instruction, the system executes the instruction by manipulating selected blocks. The highly compositional instructions are composed of atomic components and understanding these components is a necessary step to executing the instruction. We show that while end-to-end training (supervised only by the correct block location) fails to address the challenges of this task and performs poorly on instructions involving a single atomic component, knowledge-free auxiliary signals can be used to significantly improve performance by providing supervision for the instruction's components. Specifically, we generate signals that aim at helping the model gradually understand components of the compositional instructions, as well as those that help it better understand spatial concepts, and show their benefit to the overall task, especially when the training data is limited---which is usual in such tasks.

TASK: INSTRUCTION FOLLOWING



- Given a configuration of blocks (world state) and an instruction, we have two subtasks to execute the instruction:
- Source subtask: Predict the source block to move.
- Target subtask: Predict the target location to place the source block.

CHALLENGES IN THIS TASK

- Instructions involve multiple spatial concepts
- Instructions involve a high degree of compositionality
- End-to-end learning is not sufficient.
- We use (1) Data Augmentation and (2) Auxiliary Tasks to address this.

DATA AUGMENTATION

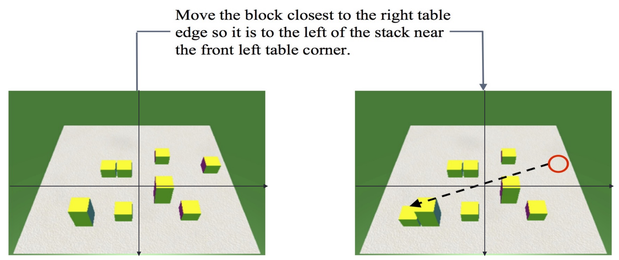
Data Augmentation: The data augmentation templates are designed to:

- teach the model the individual spatial concepts which form components of the compositional instructions.
- test if existing models do reasoning by evaluating them on the simpler, generated instructions.

Template	Example Sentence
TA	Move the northwest corner block to the center.
TR	Move the leftmost block to the center.
SA	Move the center block to the top left corner.
SR	Move the center block two spaces to the left.
Labeled	Move the BMW block above the Shell block.

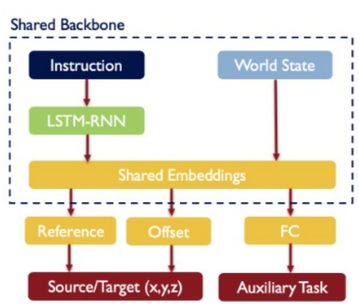
AUXILIARY TASKS

- Auxiliary Tasks: provides explicit signals to the model regarding components of the compositional instruction.
- Designed in a knowledge-free manner .
- The auxiliary tasks are trained jointly along with the main task.
- Two types considered: Quadrant and Anchor auxiliary tasks.



COMPARED MODELS

- We consider the following models from prior work: Bisk et al. (2016)
- B_U : Baseline model trained on the unlabeled dataset.
- B_L : Baseline model trained on the labeled dataset.
- For each of them, we add data augmentation (B_U + Aug., B_L + Aug.), auxiliary tasks, and a combination of both.



EXPERIMENTS

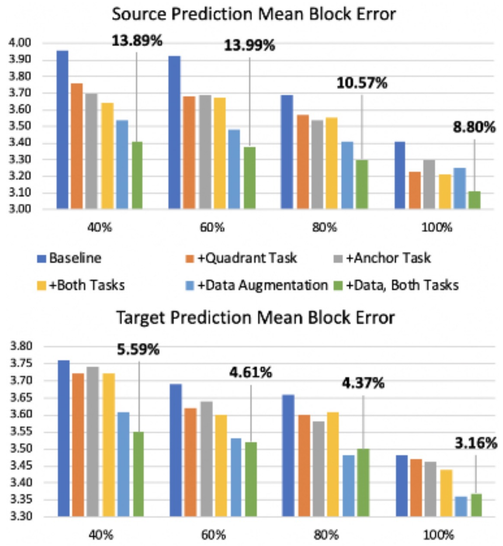


Figure 3: Ablation study of our approach against the baseline (B_U) for different percentages of training data. B_U is the model trained on the un-labeled blocks data (Bisk et al., 2016) with our batching scheme. The percentage above the green bar shows the relative improvement w.r.t. the dark blue (baseline) bar.

WHY DOES AUGMENTATION HELP ?

Model	Source		Target	
	BD	RI%	BD	RI%
B_U	3.12	—	3.59	—
B_U+ Q	2.80	10.26	3.48	3.06
B_U+ Q+ Aug.	**2.76**	11.54	**3.35**	6.69

Table 3: Ablated gains for mean block distance (BD) on the diagnostic subset. B_U: baseline model, Q: quadrant auxiliary task, *Aug* denotes the corresponding data augmentation: TA for Source, SA for Target.

REFERENCES

[1] Bisk, Yonatan, Deniz Yuret, and Daniel Marcu. "Natural language communication with robots." *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016.