北京航空航天大學 BEIHANG UNIVERSITY

西安交通大學 XI'AN JIAOTONG UNIVERSITY

Compositional Generalization by Learning Analytical Expressions

Microsoft

Qian Liu (qian.liu@buaa.edu.cn)

On Behalf of **MSRA DKI Team**

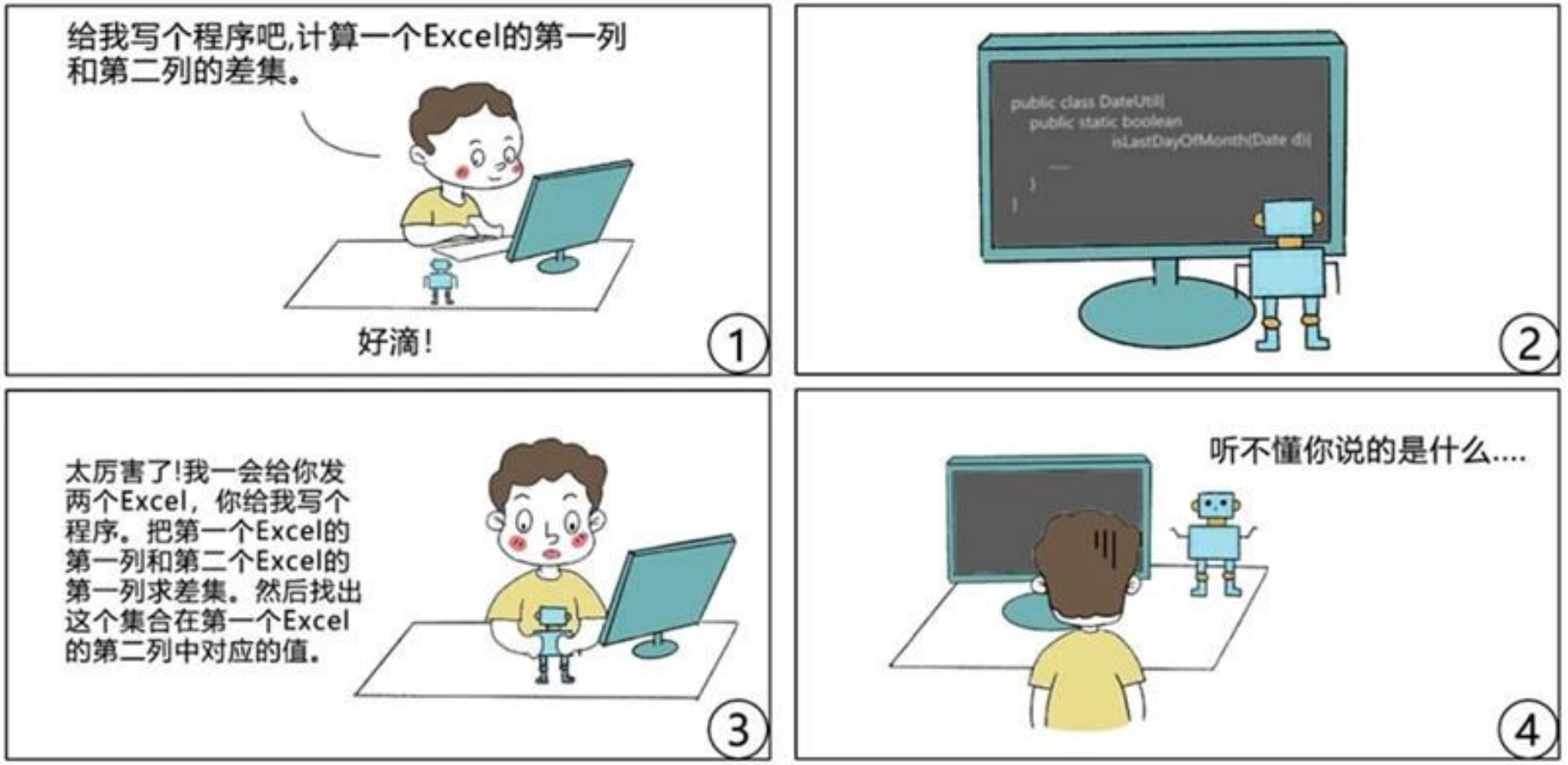
AI TIME @ 2020.12.11

We're Hiring

Data, Knowledge and Intelligence（数据, 知识与智能）

- In the past few years, we have published 20+ papers on top conferences.
- Join us to work on world leading research projects, exploring state-of-the-art algorithms in NLP and Software (NL2Code, Compositional Generalization).
- We're hiring FTEs（正式员工）which will graduate in 2021. Please send your CV to di-recruit@microsoft.com for job opportunity.

The current state of AI programmers



Compositional Generalization

- The compositionality of programs $\Longrightarrow$ huge search space of programs

- Compositional Generalization: human intelligence exhibits the algebraic capacity to dynamically recombine existing atoms.

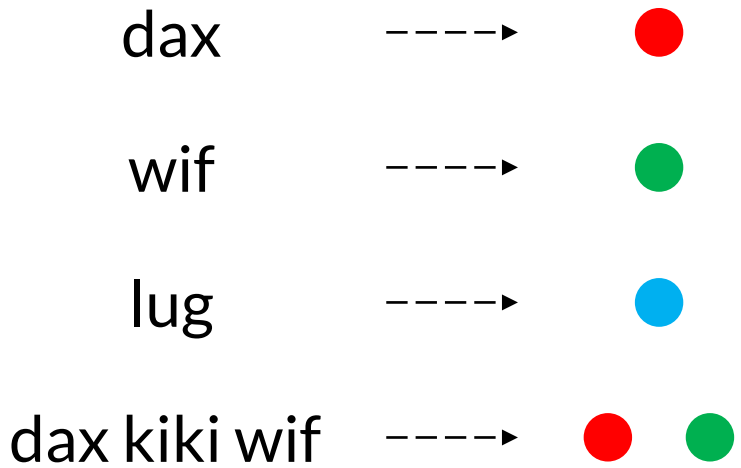
Infinite use of finite means.

—— Noam Chomsky

Compositional Generalization in Cognition

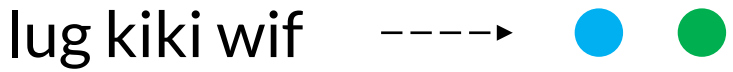
Compositional generalization is an ability to recombine known parts to understand novel sentences which have never been encountered before.

Observed Examples



Credit: Lake et al. 2019

A Novel Example



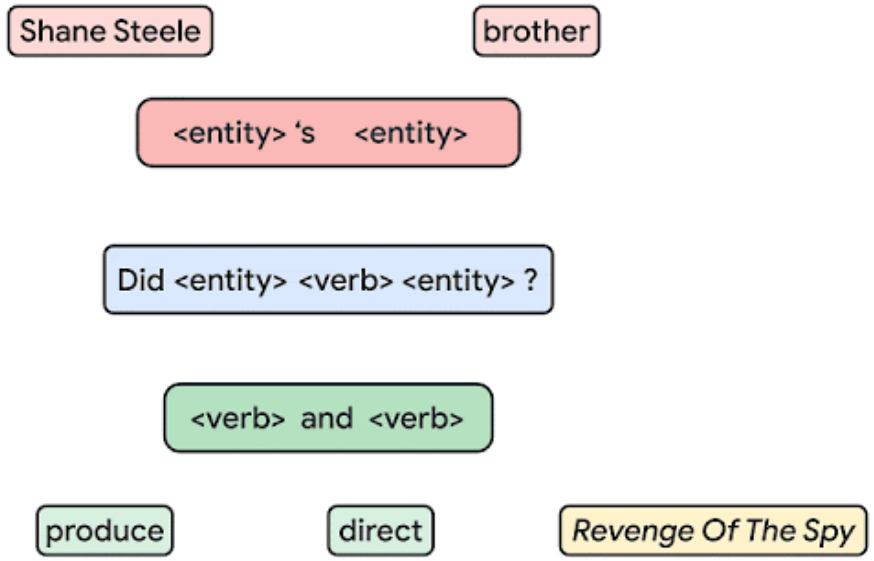
Compositional Generalization in NL2Code

The Simplified version of the CommAI Navigation (SCAN) is a **synthetic** benchmark (Lake & Baroni. 2018) with navigation commands and action sequences.

	Natural Language		Programming Language
Train	run twice	⟹	RUN RUN
	jump and walk	⟹	JUMP WALK
Test	run and jump twice	⟹	RUN JUMP JUMP

Compositional Generalization in NL2Code

CFQ (Compositional Freebase Questions) is a **realistic** benchmark (Keysers et al. 2020) that comprehensively measure compositional generalization on KBQA.



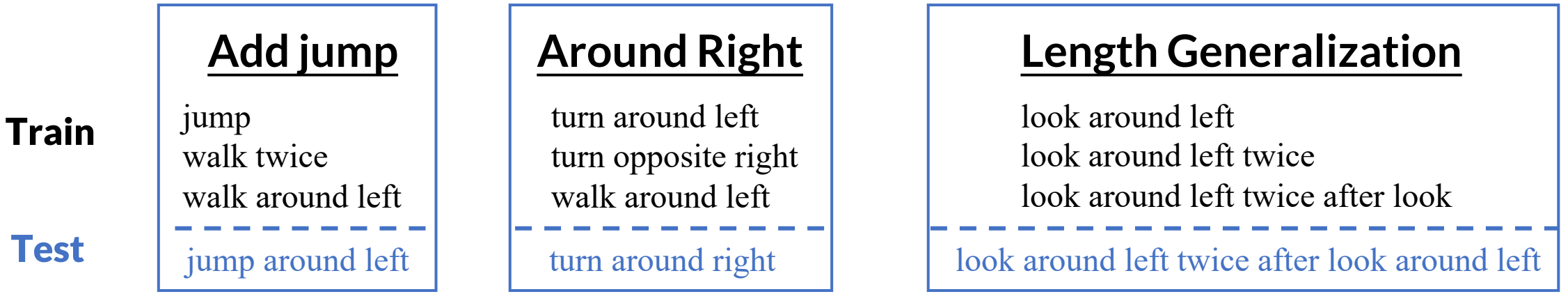
Credit: Keysers et al. 2020

Which Swedish founder of [M0] produced [M2] ?

```
SELECT DISTINCT ?x0 WHERE {
    ?x0 ns:film.producer.film|ns:film.production_company.films [M2] .
    ?x0 ns:organization.organization_founder.organizations_founded [M0] .
    ?x0 ns:people.person.nationality ns:m.0d0vqn
}
```

Measuring Compositional Generalization

The SCAN benchmark is split in **handcraft ways** to form the challenges:



No complex command of jump in training

"around right" is held out from the training set

Train: length of the action sequence is shorter than 24 actions; Test: all action sequences longer than or equal to 24 actions.

Measuring Compositional Generalization

The CFQ benchmark is split based on **automatic algorithms** which highlight properties that intuitively correlate with compositional structure:

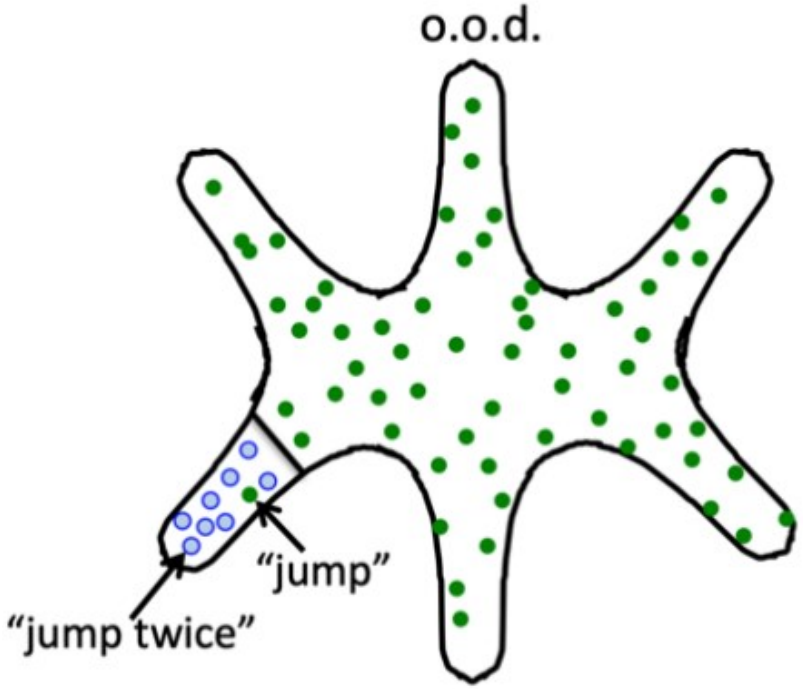
(1) Similar atom distribution: All test atoms occur in train, and Distribution of atoms is similar between train and test.

(2) Different compound distribution: Distribution of compounds is different between train and test.

Train	**Test**
Who **directed** Inception?	Who **produced** Inception?
Did Greta Gerwig **produce** Goldfinger?	Did Greta Gerwig **direct** Goldfinger?

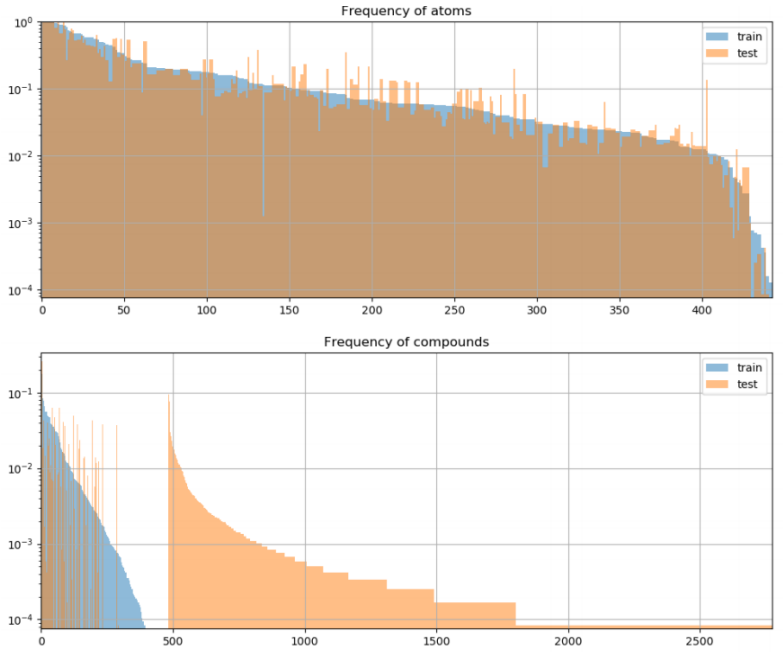
Measuring Compositional Generalization

The SCAN split distribution



Credit: Russin et al. 2020

The CFQ split distribution



Credit: Keysers et al. 2020

A Promising Direction

- Datasets
  - ✓ **SCAN** (Lake & Baroni, ICML'18)
  - ✓ **CFQ** (Keysers et al, ICLR'20)
  - ✓ **COGS** (Kim & Linzen, EMNLP'20)
  - ✓ **Grounded SCAN** (Ruis et al, NeurIPS'20)
- Methods
  - ✓ CGPS (Li et al, EMNLP'19)
  - ✓ Meta Seq2Seq (Brenden M. Lake, NeurIPS'19)
  - ✓ Permutation Equivariant Seq2Seq (Gordon et al, ICLR'20)
  - ✓ GECA (Jacob Andreas, ACL'20)

  …

Far From Compositional Generalization

No model can successfully solve compositional challenges on SCAN!

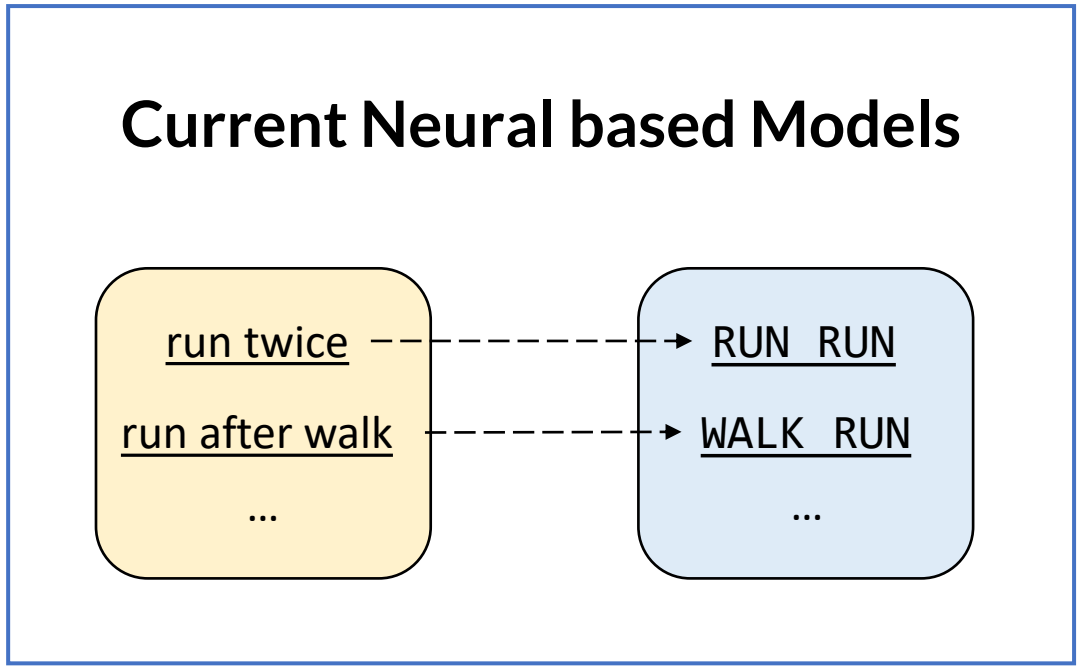
Model	Add Jump	Around Right	Length
Seq2Seq	1.2	2.5	13.8
CNN	69.2	56.7	0.0
Syntactic Attention (Russin et al. 2019)	91.0	28.9	15.2
CGPS (Li et al. 2019)	98.8	83.2	20.3
GECA (Jacob Andreas. 2020)	86.0	82.0	-
Meta Seq2Seq (Brenden M. Lake. 2019)	99.9	99.9	16.6
Equivariant Seq2Seq (Gordon et al. 2020)	99.1	92.0	15.9

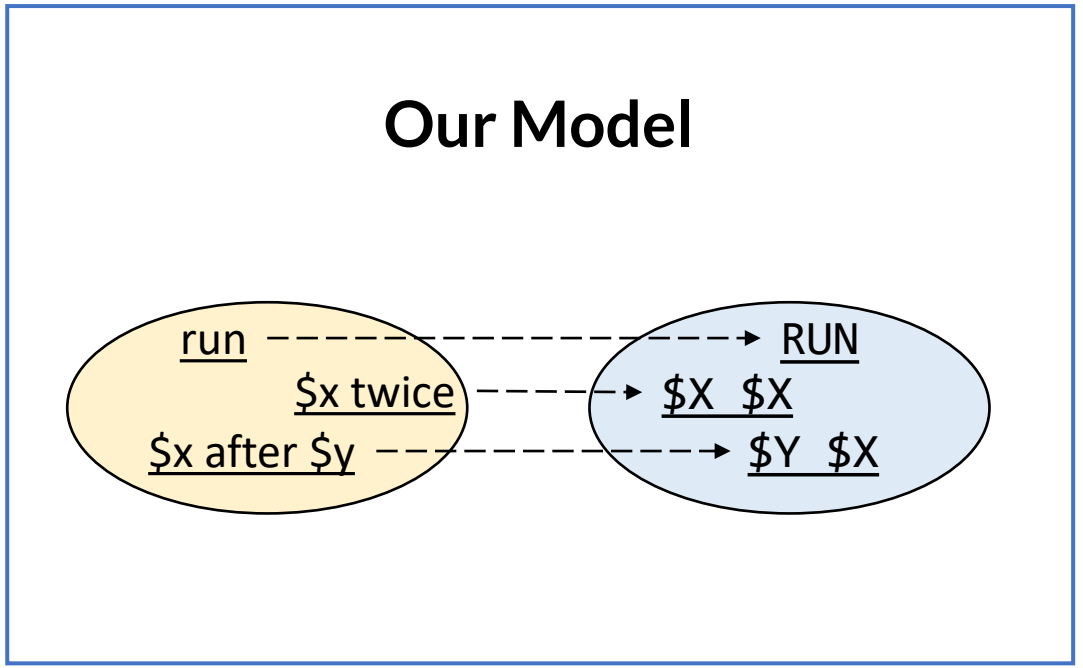
*green models trained w/o extra resources

*blue models trained with extra resource

Model on Compositionality

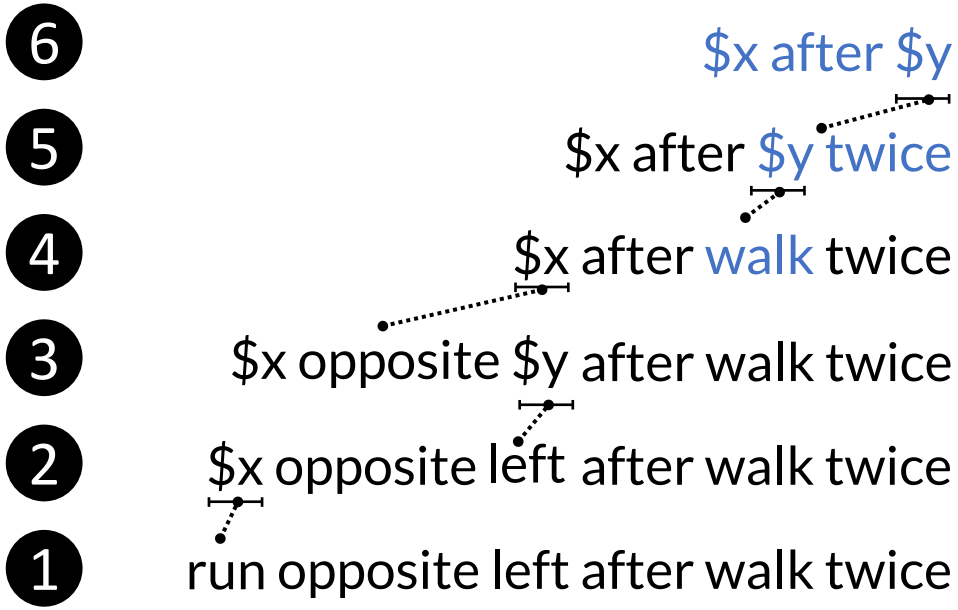
The compositionality of language constitutes an algebraic system, of the sort that can be captured by symbolic functions with variable slots (M. Baroni, 2019).





Learn Analytical Expressions

The understanding of “run opposite left after walk twice” can be regarded as a hierarchical application of symbolic functions.

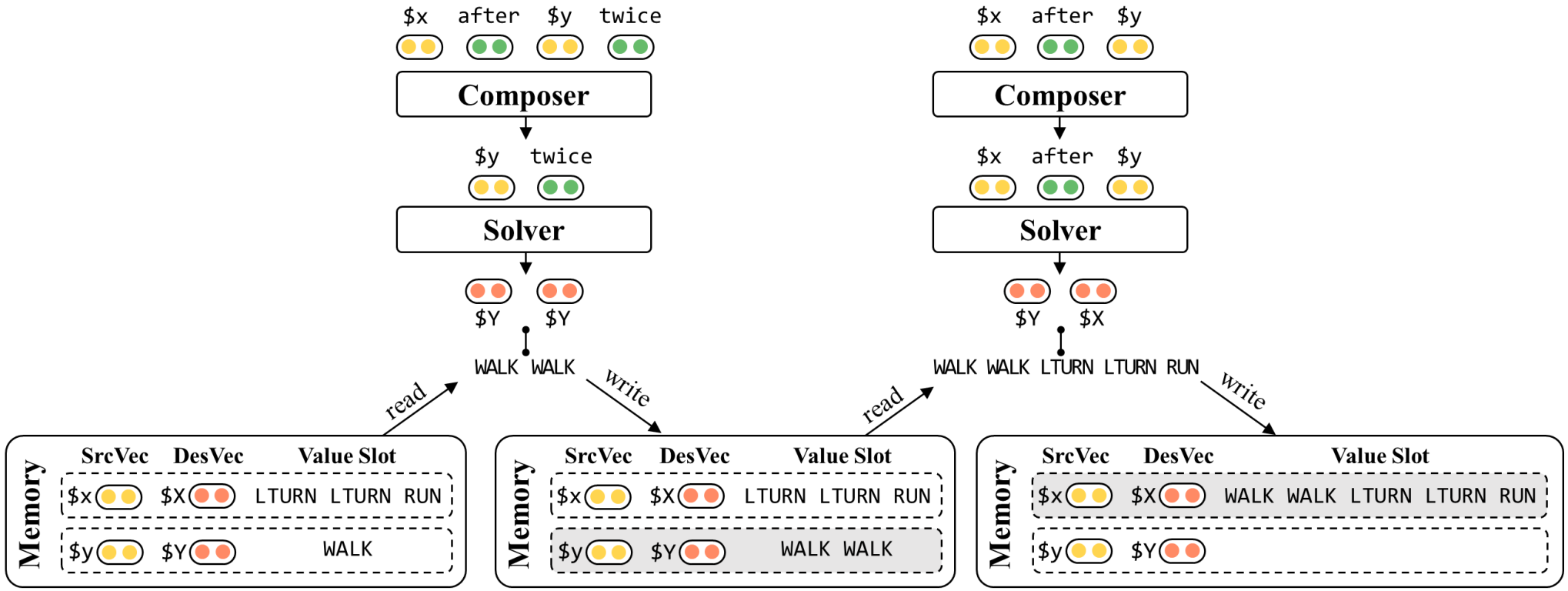


Symbolic Function

$x after $y	→	$Y $X
$y twice	→	$Y $Y
walk	→	WALK
$x opposite $y	→	$Y $Y $X
left	→	LTURN
run	→	RUN

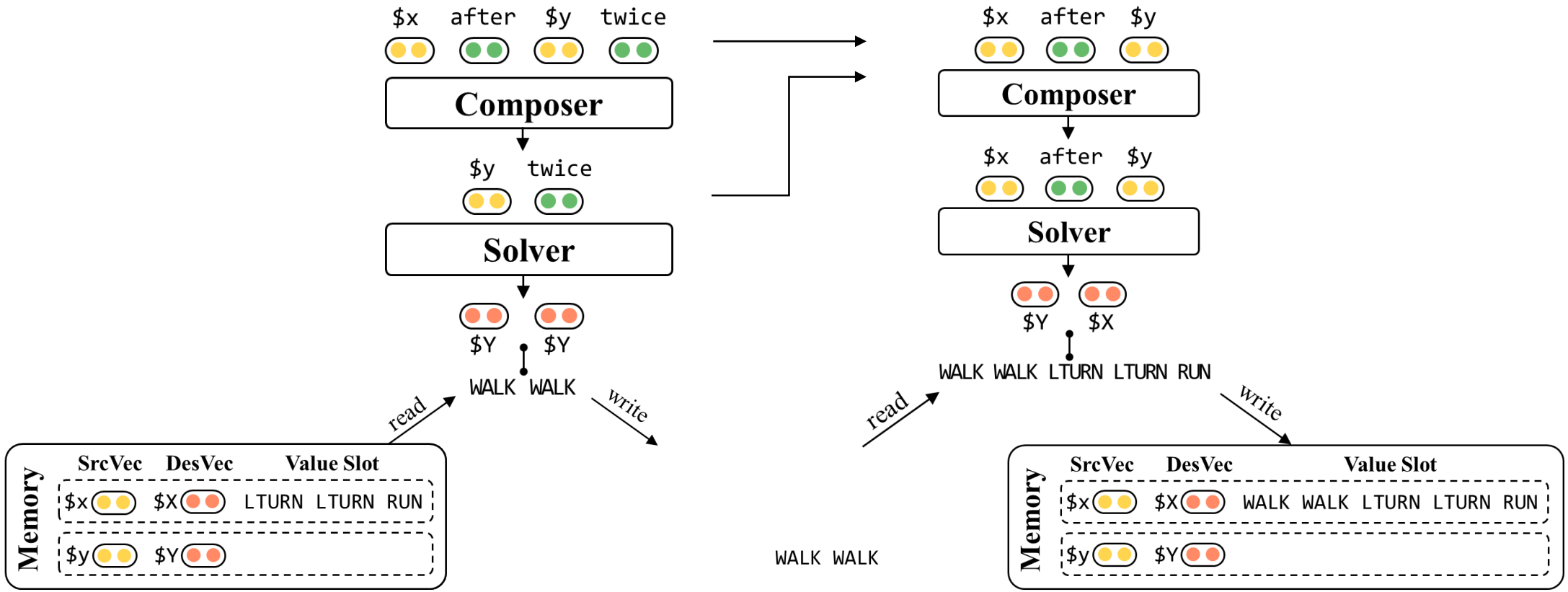
LANE: Memory-Augmented Model

We propose a memory-augmented neural model to achieve compositional generalization by automatically learning the above analytical expressions.



LANE: Memory-Augmented Model

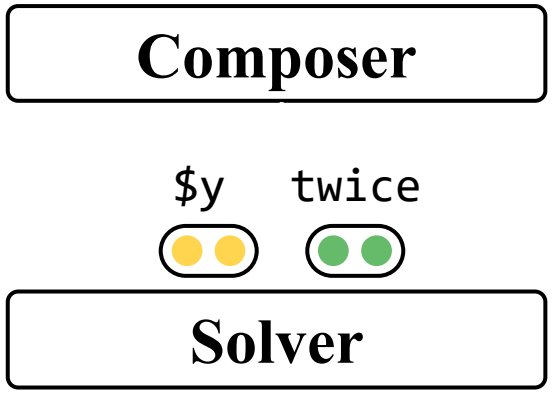
Our model understands via interaction between Composer, Solver and Memory.



The Training is Challenging!

Challenges

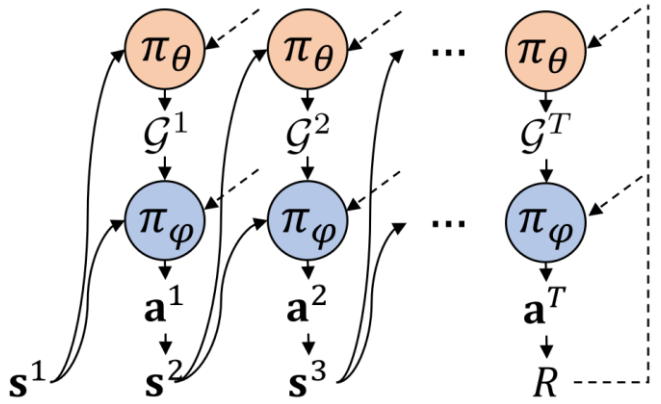
(i) Discrete Action, Non-differentiable.



(ii) Sparse Reward, Hard to Train.

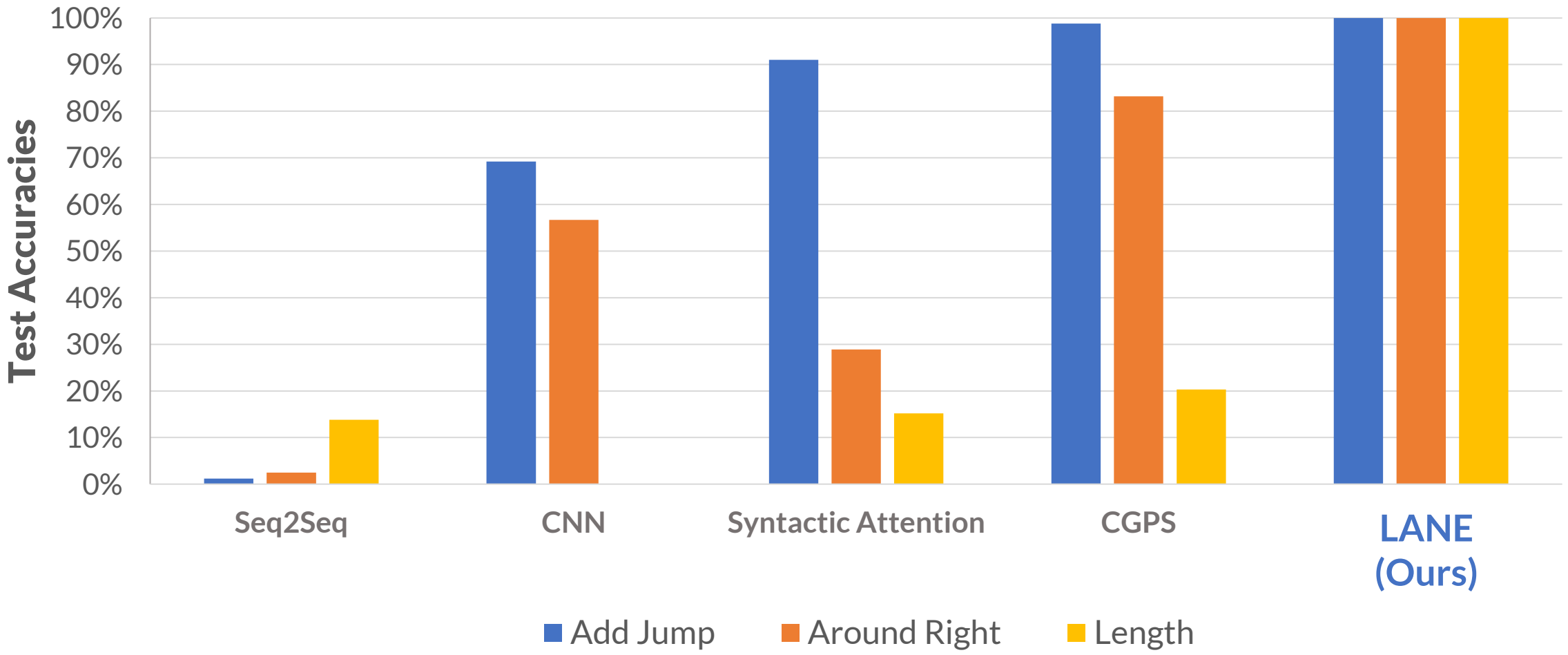
Solutions

(i) Hierarchical Reinforcement Learning.



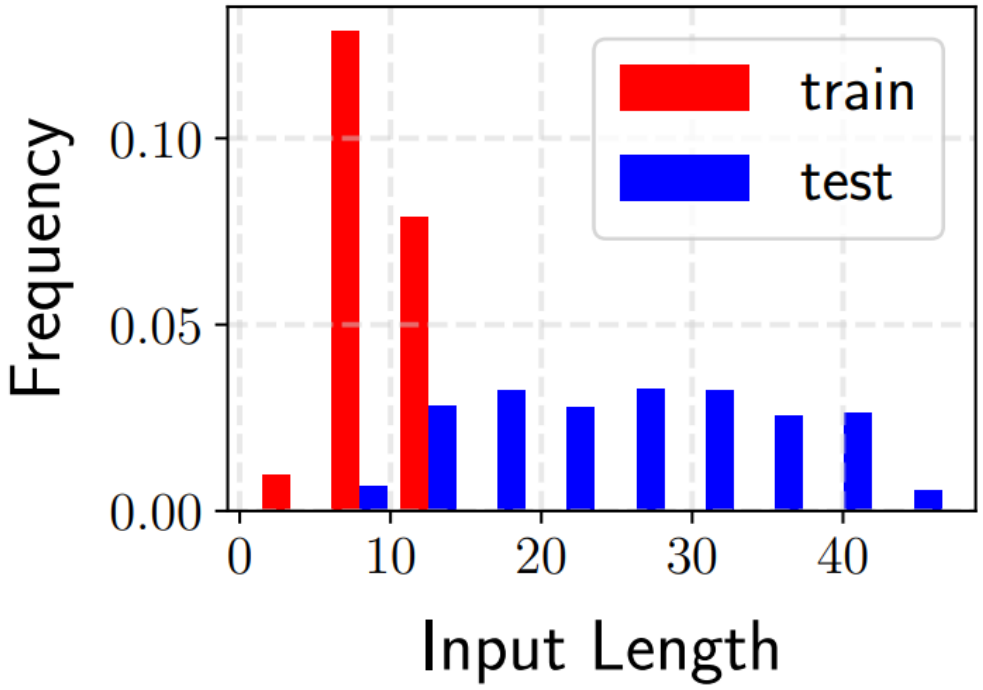
(ii) Curriculum Learning.

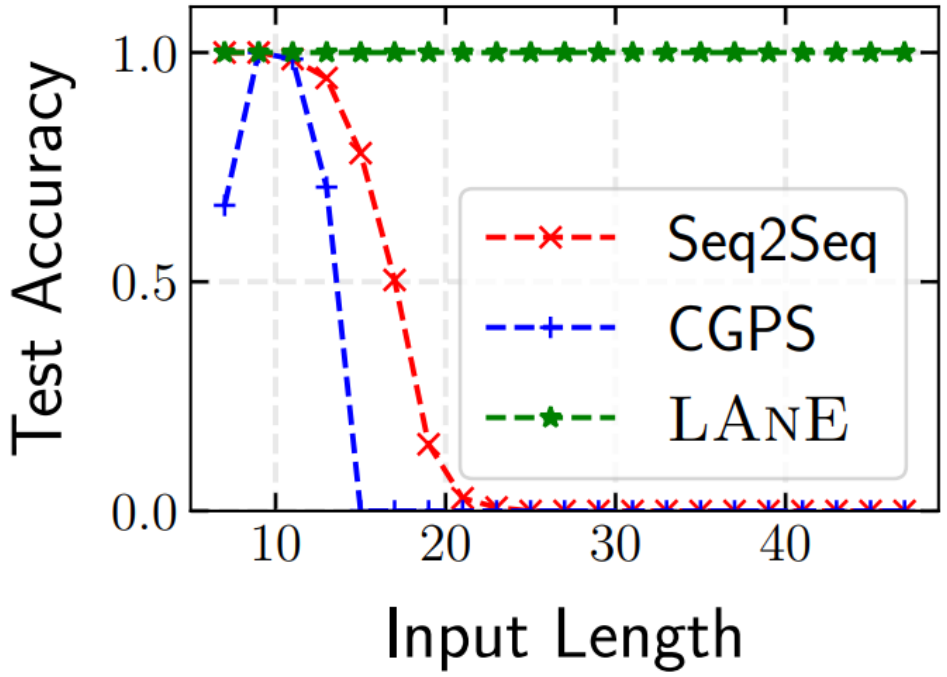
Evaluate on SCAN



Evaluate on Longer Inputs

Languages license a theoretically infinite set of sentences due to compositionality, and our model maintains a perfect trend as the input length increases.





Take Away

Compositional Generalization by Learning Analytical Expressions [NeurIPS'20]

- The key for compositionality is to regard language as an algebraic system, which be captured by analytical expressions.
- Learning analytical expressions can be modeled as the joint optimization of three cooperative modules Composer, Solver and Memory.
- Latent discrete actions between modules can be tackled by the combination of hierarchical reinforcement learning and curriculum learning.

Reference

[1]. Lake et al. Human few-shot learning of compositional instructions. In CogSci 2019.

[2]. Lake & Baroni. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In ICML 2018.

[3]. Keysers et al. Measuring Compositional Generalization: A Comprehensive Method on Realistic Data. In ICLR 2020.

[4]. Russin et al. Compositional generalization by factorizing alignment and translation. In EMNLP 2020 Student Workshop.

[5]. Kim & Linzen COGS: A compositional generalization challenge based on semantic interpretation. In EMNLP 2020.

[6]. Ruis et al. A Benchmark for Systematic Generalization in Grounded Language Understanding. In NeurIPS 2020.

[7]. Li et al. Compositional Generalization for Primitive Substitutions. In EMNLP 2019.

[8]. Brenden M. Lake. Compositional generalization through meta sequence-to-sequence learning. In NeurIPS 2019.

[9]. Gordon et al. Permutation Equivariant Models for Compositional Generalization in Language. In ICLR 2020.

[10]. Jacob Andreas. Good-Enough Compositional Data Augmentation. In ACL 2020.

[11]. M. Baroni. Linguistic generalization and compositionality in modern artificial neural networks. In Phil. Trans. R. Soc. B. 2019.

Thanks & QA

- Code: https://github.com/microsoft/ContextualSP
- Related papers from our team (MSRA DKI):
  - Hierarchical Poset Decoding for Compositional Generalization in Language
  - Revisiting Iterative Back-Translation from the Perspective of Compositional Generalization
  - Iterating Utterance Segmentation for Neural Semantic Parsing