Predictive Synthesis of API-Centric Code

Daye Nam*
CMU

Baishakhi Ray*
Columbia University

Seohyun Kim
Meta

Xianshan Qu
Meta

Satish Chandra
Meta

Motivation

How do we (humans) program?

in1

1	4	3
2	5	1

in2

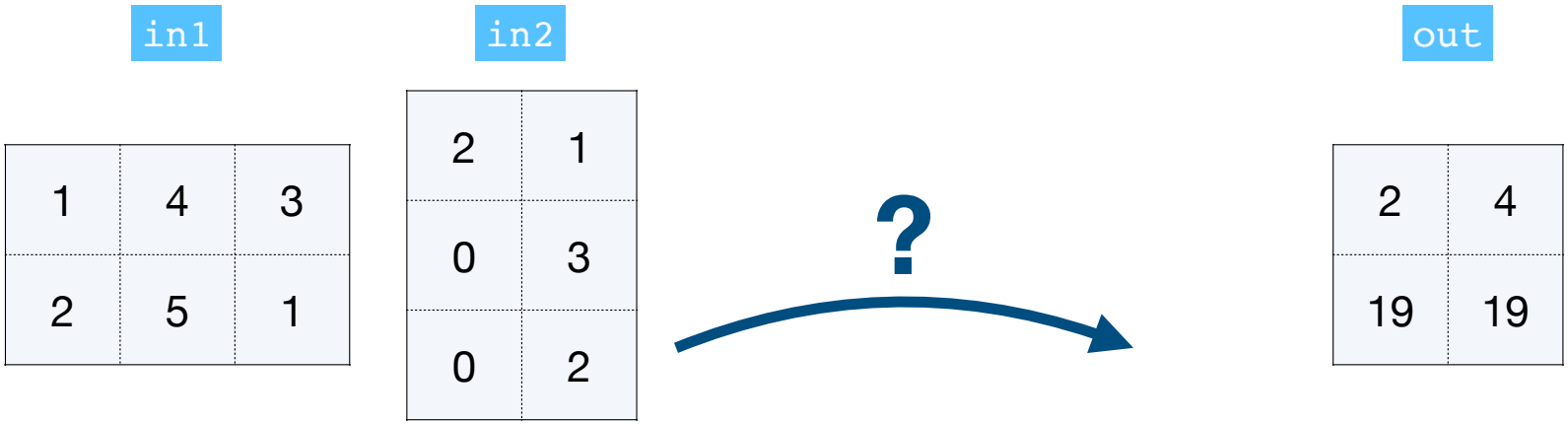
2	1
0	3
0	2

out

2	4
19	19

Motivation

How do we (humans) program?



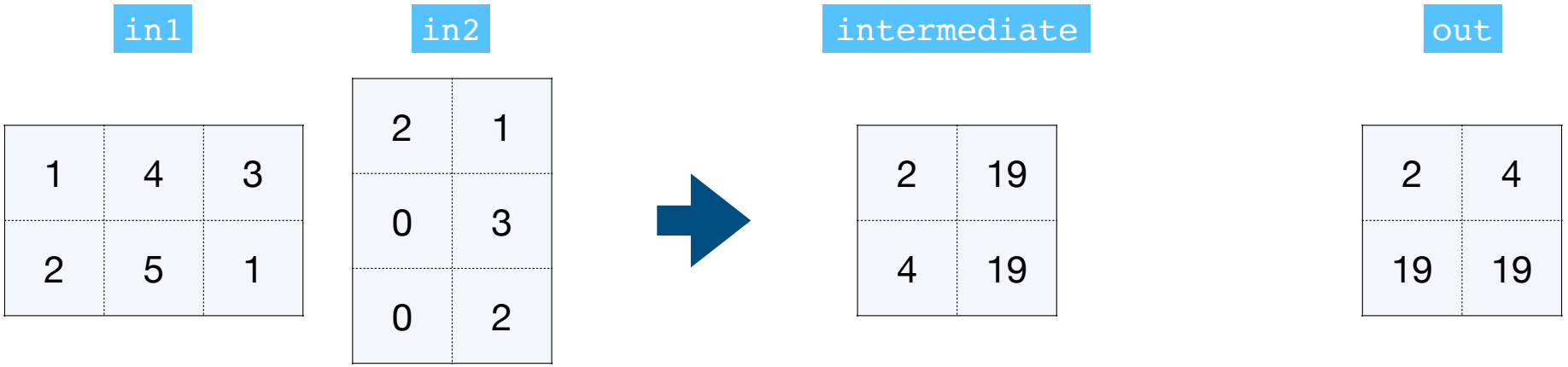
1: Choose a probable operation

Matrix multiplication

`torch.matmul(in1,in2)`

Motivation

How do we (humans) program?



1: Choose a probable operation

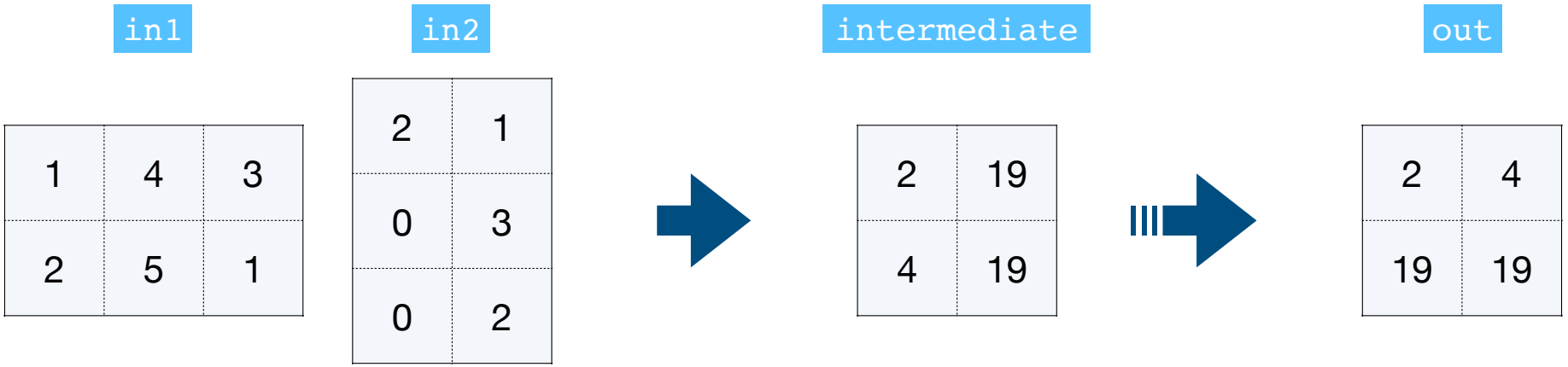
Matrix multiplication

`torch.matmul(in1,in2)`

2: Compute and check

Motivation

How do we (humans) program?



1: Choose a probable operation

Matrix multiplication

`torch.matmul(in1,in2)`

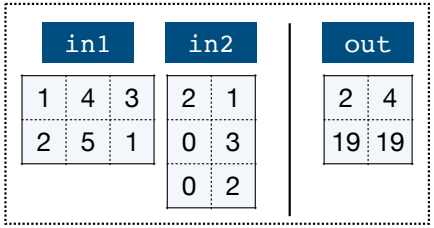
2: Compute and check **3: Repeat**

Transposition

`torch.transpose(intermediate,0,1)`

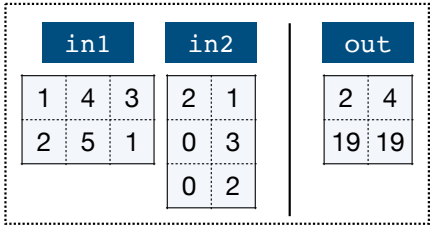
Program synthesis

Enumerative search based synthesis



Program synthesis

Enumerative search based synthesis



Base Values

in1

1	4	3
2	5	1

in2

2	1
0	3
0	2

0 1 -1 2 ...

API Functions

`add, arange, eq, matmul, transpose, eye, one_hot, reshape, gt, sum, where, …`

Values

`add(in1, 0)`

1	4	3
2	5	1

`add(in1, 1)`

2	5	4
3	6	2

`add(in1, -1)`

0	3	2
1	4	0

...

`add(add(in1,0),1)`

2	5	4
3	6	2

`add(add(in1,1),1)`

3	6	5
4	7	3

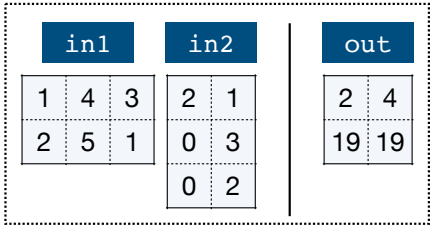
...

`?`

2	4
19	19

Program synthesis

Enumerative search like how we program



Base Values

in1

1	4	3
2	5	1

in2

2	1
0	3
0	2

`0` `1` `-1` `2` ...

Likely API Function Sequences

`API_1` matmul, tensordot, …

`API_2` transpose, reshape, …

Values

`matmul(in1, in2)`

2	19
4	19

`matmul(in2, in1)`

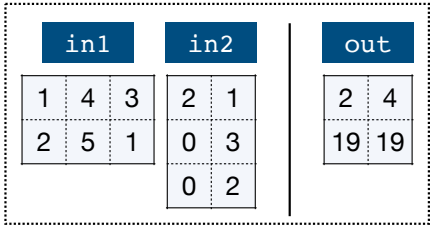
4	13	7
6	15	3
4	10	6

`transpose(matmul(in1, in2), 0, 1)` ...

2	4
19	19

Program synthesis

Enumerative search like how we program



Base Values

in1

1	4	3
2	5	1

in2

2	1
0	3
0	2

0 1 -1 2 ...

Likely API Function Sequences

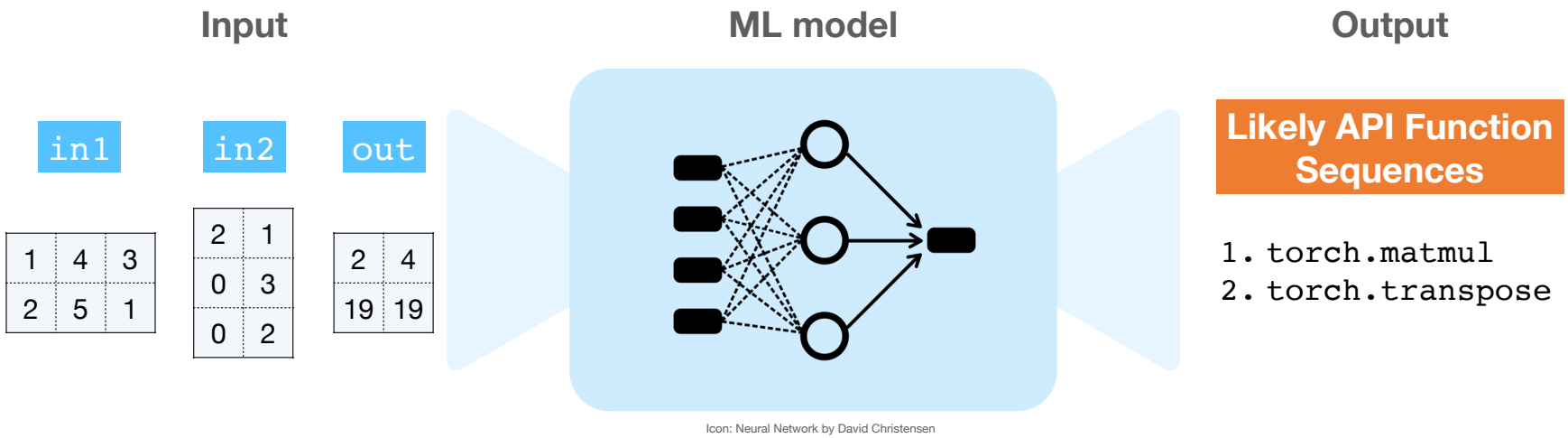
```
[matmul, transpose],
[matmul, reshape],
...
```

Values

`transpose(matmul(in1, in2), 0, 1)` `transpose(matmul(in2, in1), 0, 1)` ...

2	4
19	19

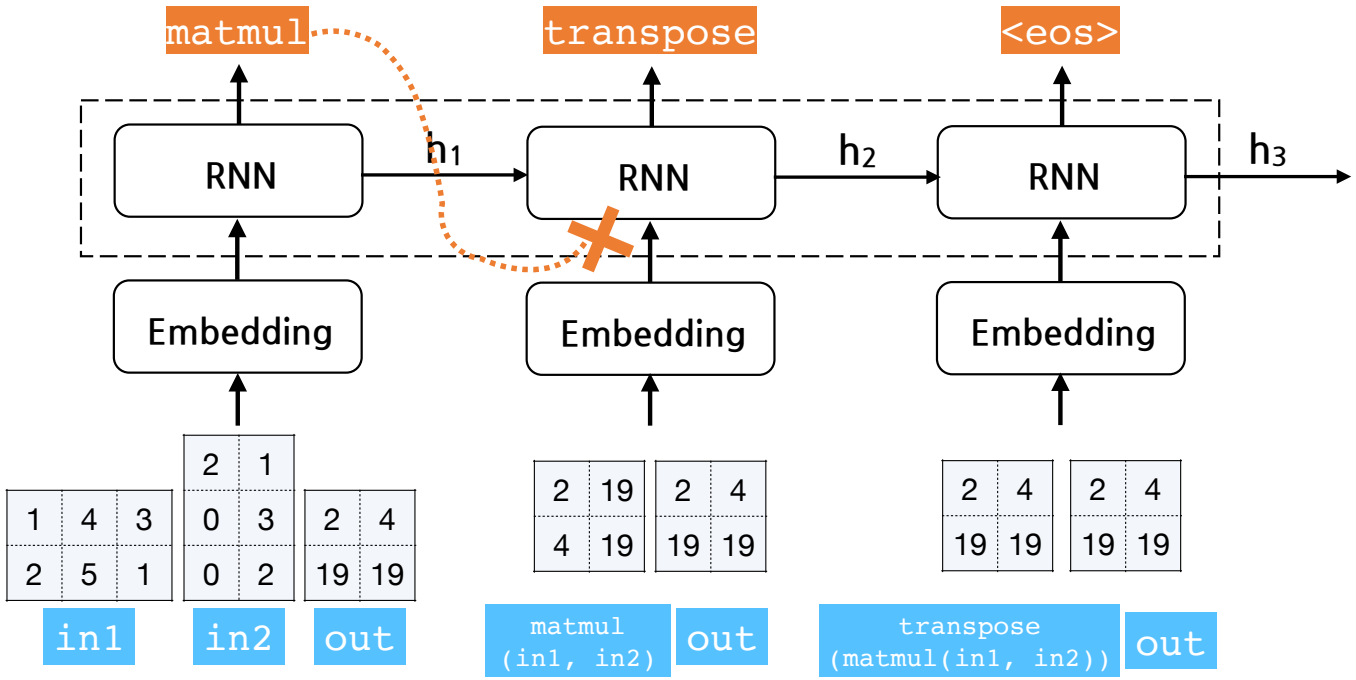
Goal



Build a ML model that is able to predict needed operations ***compositionally***, given task specifications

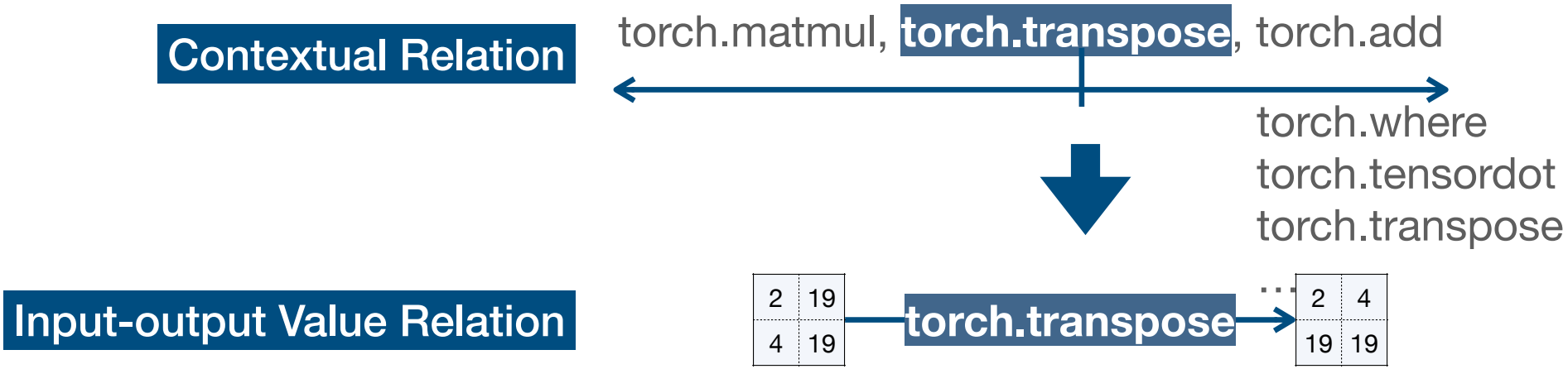
Compositional Model

Train a model to compositionally predict sequence



Compositional Model

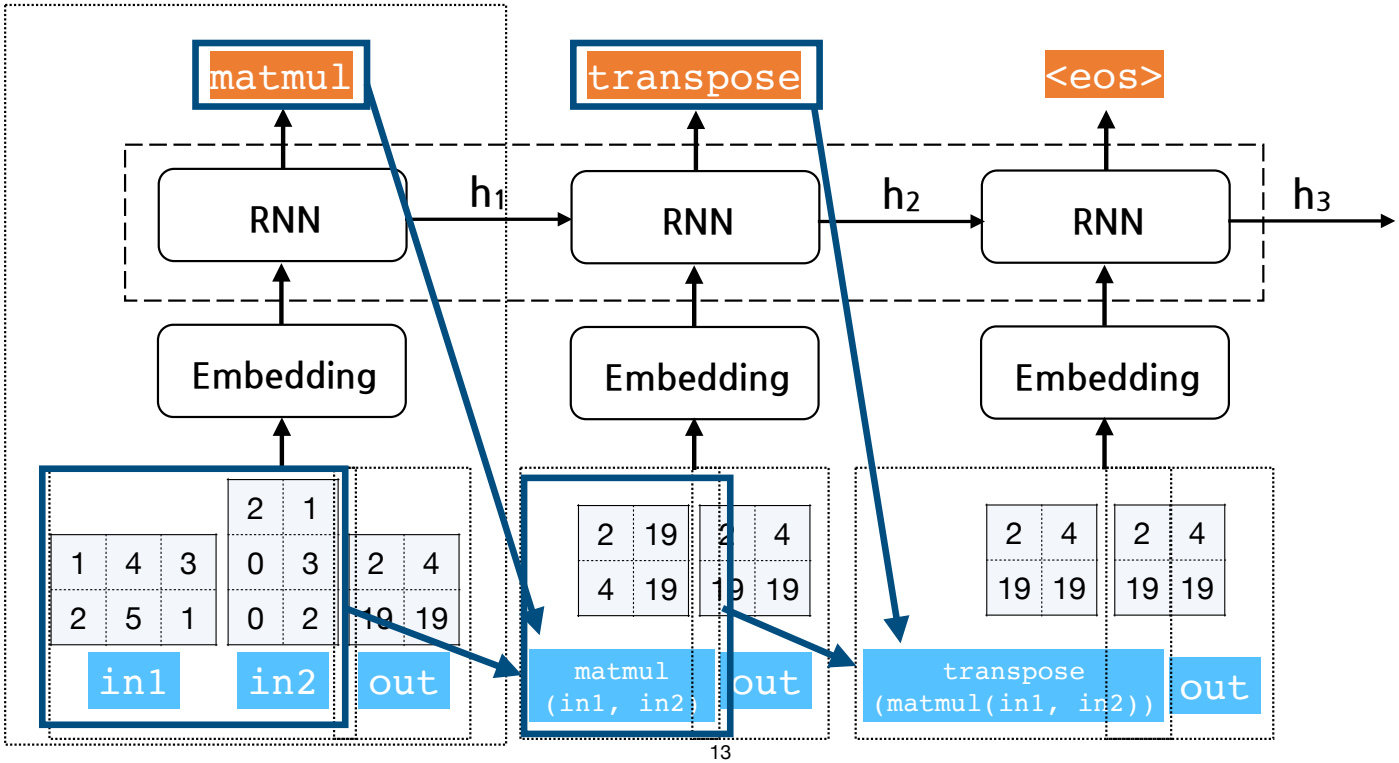
Train a model to compositionally predict sequence



Compositional Model - First of Sequence

Model input: Input tensors for each API call, final output tensor

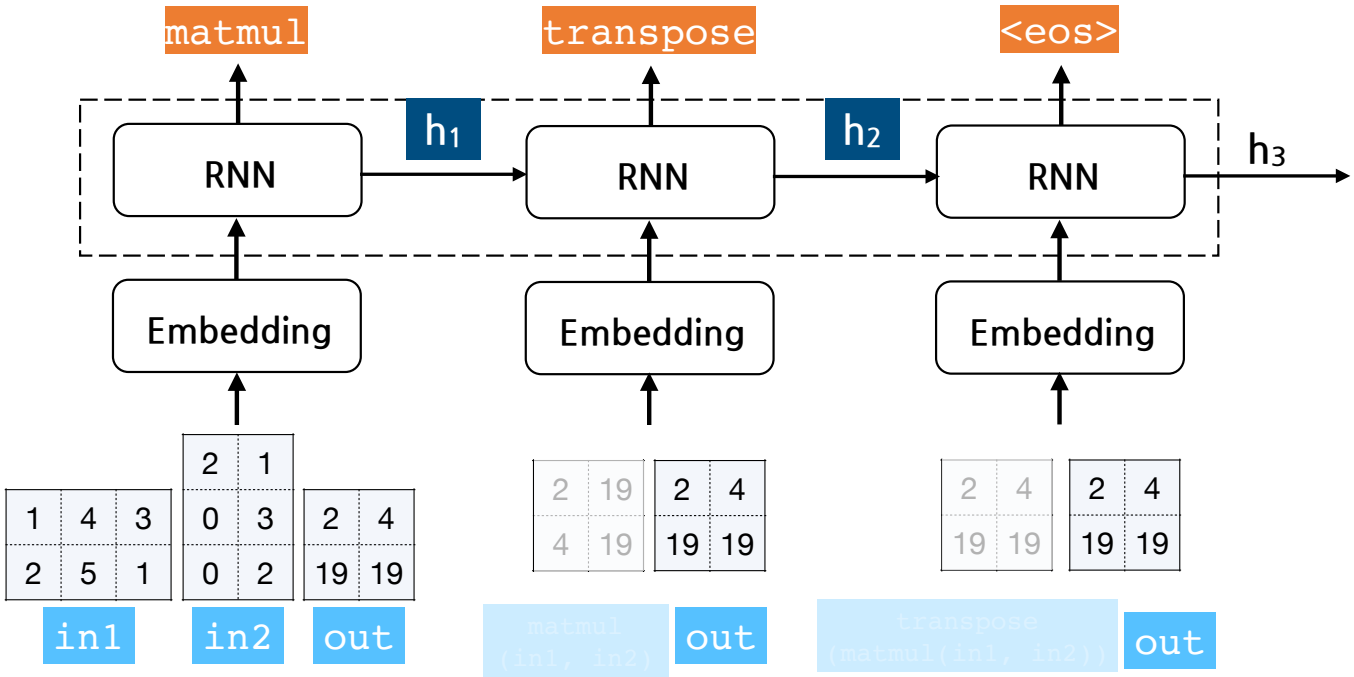
Model output: A sequence of API functions (one by one)



Compositional Model - Full Sequence

Model input: Input/output tensors of the task

Model output: A sequence of API functions



Dataset

Synthetic dataset & Stack Overflow Benchmarks

- **Synthetic Dataset**: Randomly generated input tensors, run the API functions to capture the corresponding output tensors in black-box manner

		Train	Valid	Test
Synthetic	# of unique seqs (Len)	16 (1) + 186 (2)		
	# of in/out values	5.5M	10K	10K

- **Stack Overflow Benchmarks**: Real-world tasks collected from Stack Overflow by TF-Coder authors

		Train	Valid	Test
Stack Overflow	# of unique seqs (Len)	Only used for evaluation		8 (1) + 7 (2)
	# of in/out values			18

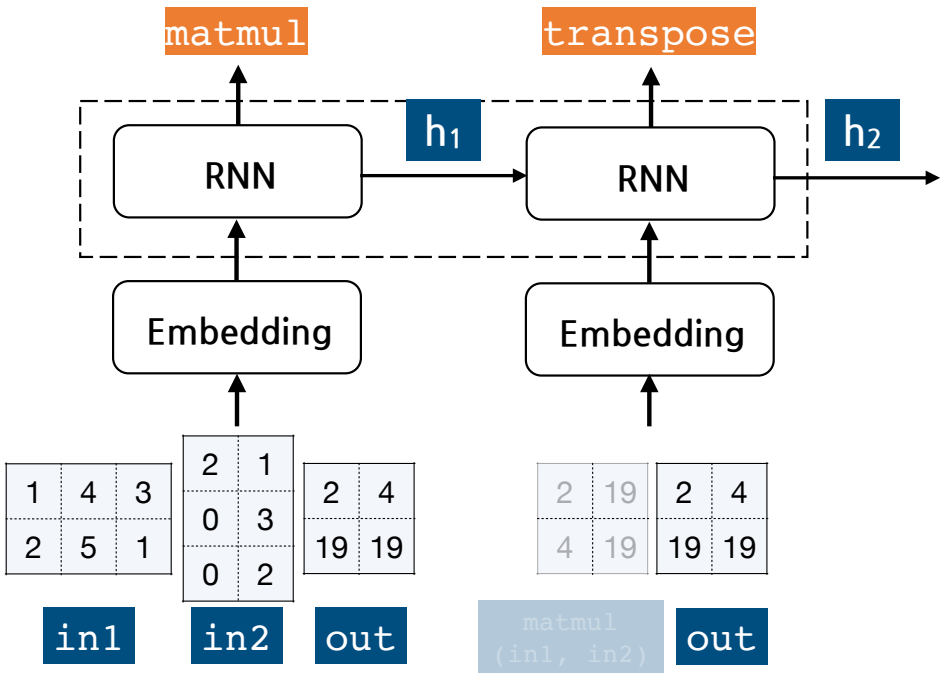
Compositional Model Results

Test Accuracy

	Synthetic Data	SO Benchmarks	
	Top-1	Top-1	Top-3
First of Seq	66.88	52.38	76.19
Full sequence	79.36	35.29	76.47

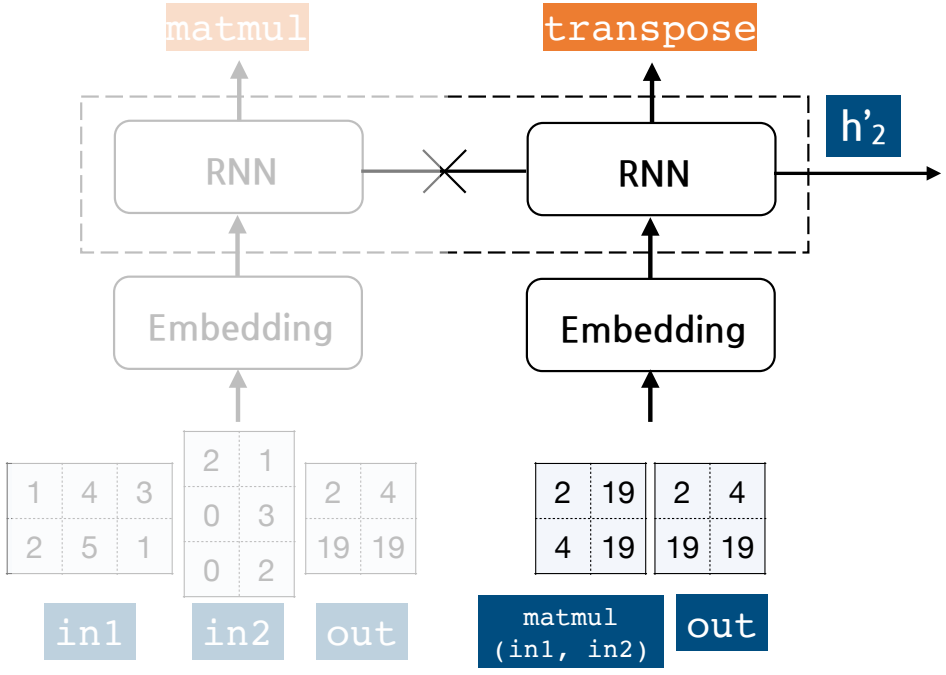
Compositional Model

Can model convey information of the intermediate values?



$$RNN_i(\mathbf{h}_{i-1}, [\![inp_i\#out]\!])$$

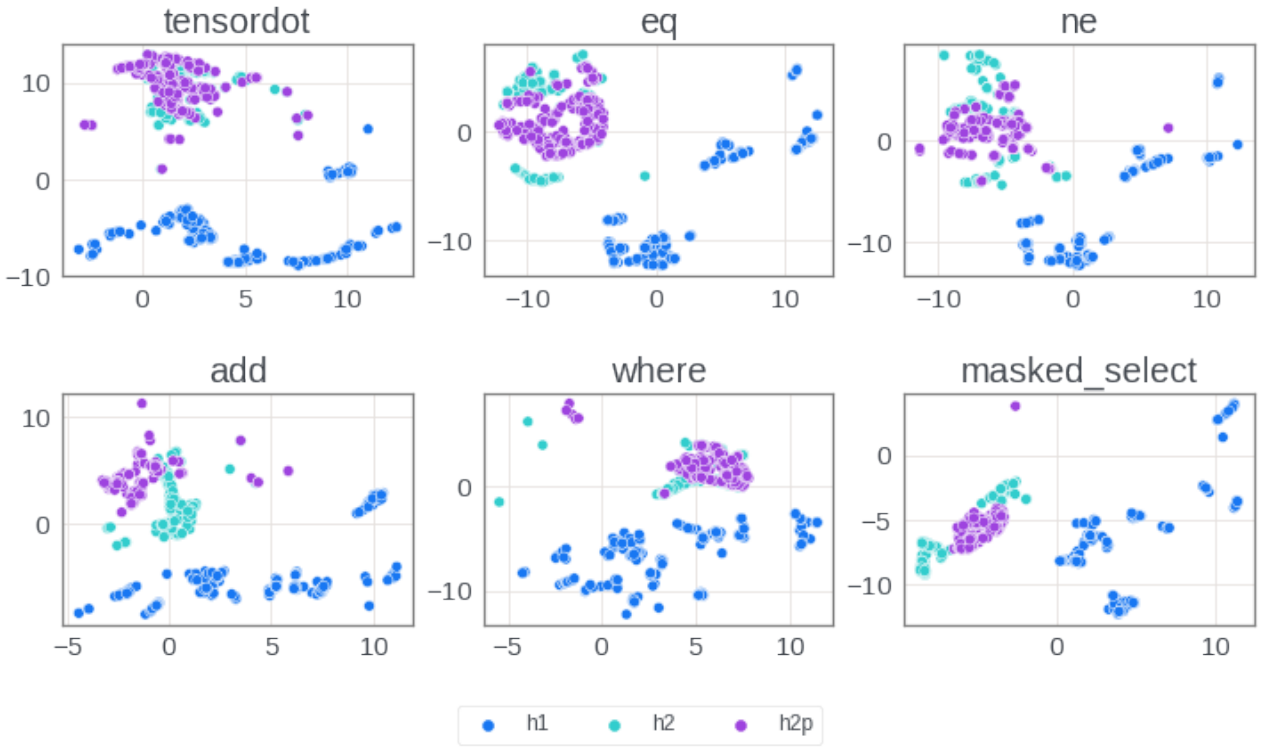
$\approx$



$$RNN_i([\![f_{i-1}(args_{i-1})\#out]\!], [\![inp_i\#out]\!])$$

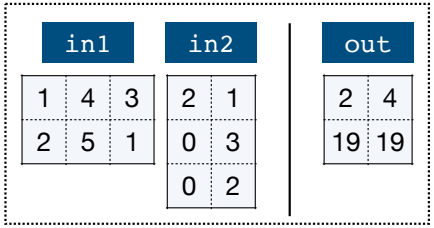
h2 vs h'2

tSNE plot of h1, h2, h’2



Compositional Model Incorporation

FOS Mode



Base Values

in1

1	4	3
2	5	1

in2

2	1
0	3
0	2

0 1 -1 2 ...

First of Sequence

API_1 **FOS_1(in1, in2, out)**
→ **matmul**

API_2 **FOS_2(** matmul(in1, in2) **, out)**
→ **transpose**

Values

matmul(in1, in2)

2	19
4	19

matmul(in2, in1)

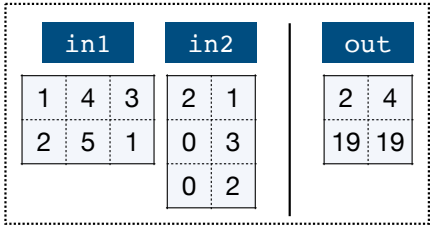
4	13	7
6	15	3
4	10	6

transpose(matmul(in1, in2), 0, 1)

2	4
19	19

Compositional Model Incorporation

FUS Mode



Base Values

in1

1	4	3
2	5	1

in2

2	1
0	3
0	2

0 1 -1 2 ...

Full Sequence

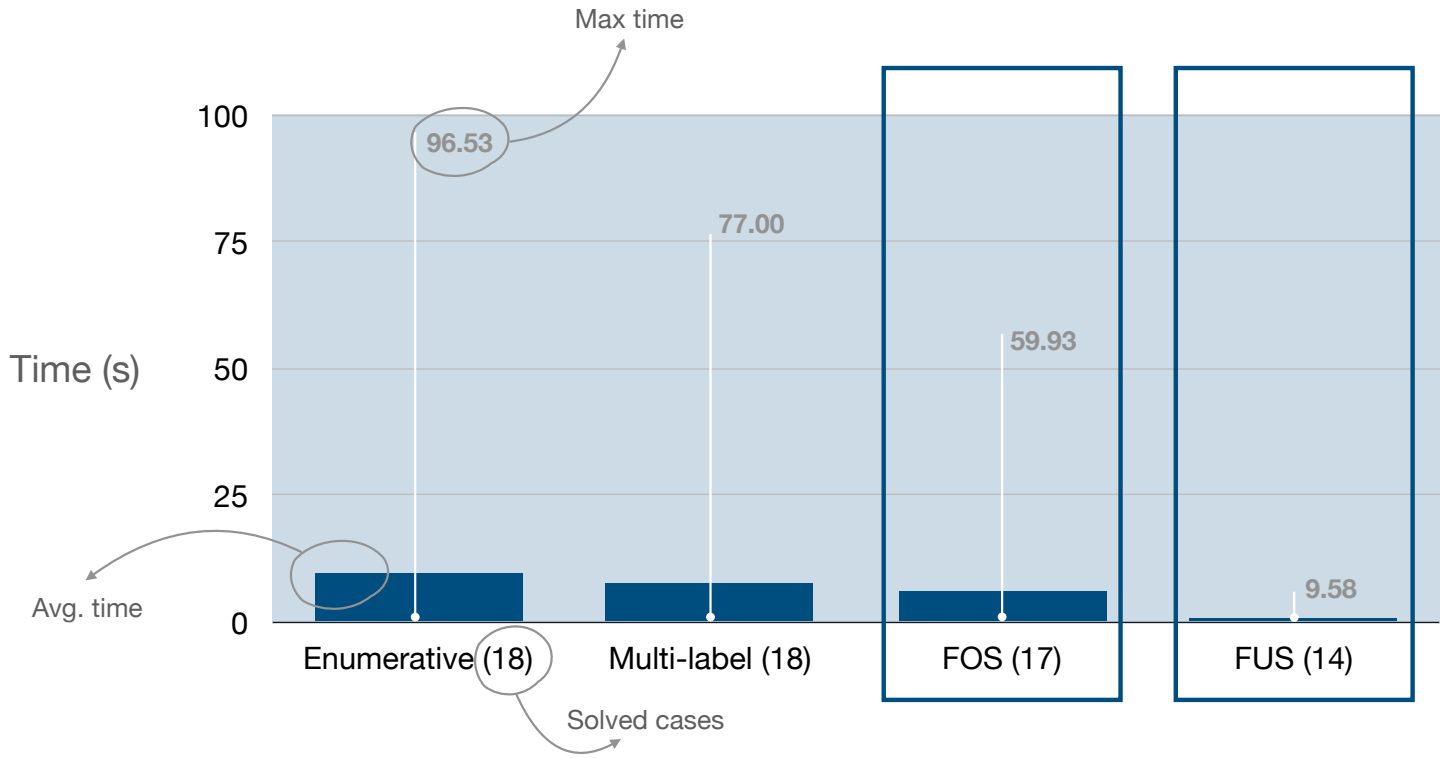
```
Full(in1, in2, out)
→ [matmul, transpose]
```

Values

`transpose(matmul(in1, in2), 0, 1)` `transpose(matmul(in2, in1), 0, 1)` ...

2	4
19	19

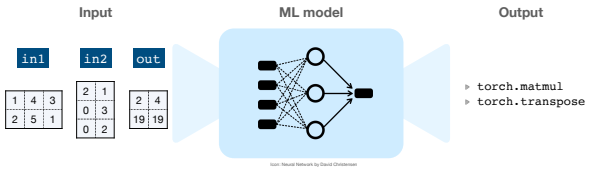
Results


Max time
100
75
50
25
0
Time (s)
96.53
77.00
59.93
9.58
Avg. time
Enumerative (18)
Multi-label (18)
FOS (17)
FUS (14)
Solved cases

Daye Nam (dayen@andrew.cmu.edu)

Compositional Prediction of API Functions from Inputs and Outputs

Goal

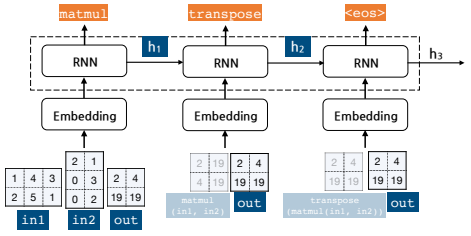


Build a ML model that is able to predict needed apis compositionally, given task specifications

7

Compositional Model - Full Sequence

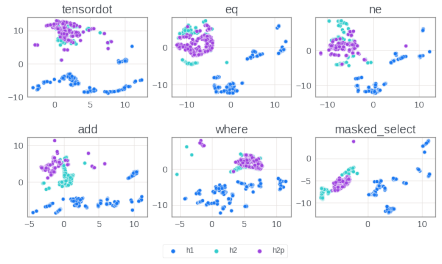
Model input: Input/output tensors of the task
Model output: A sequence of API functions



12

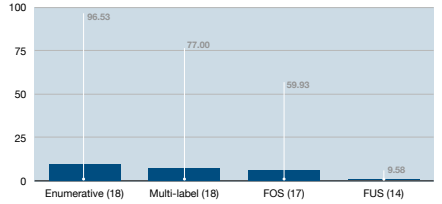
h2 vs h'2

tSNE plot of h1, h2, h'2



15

Results



23