

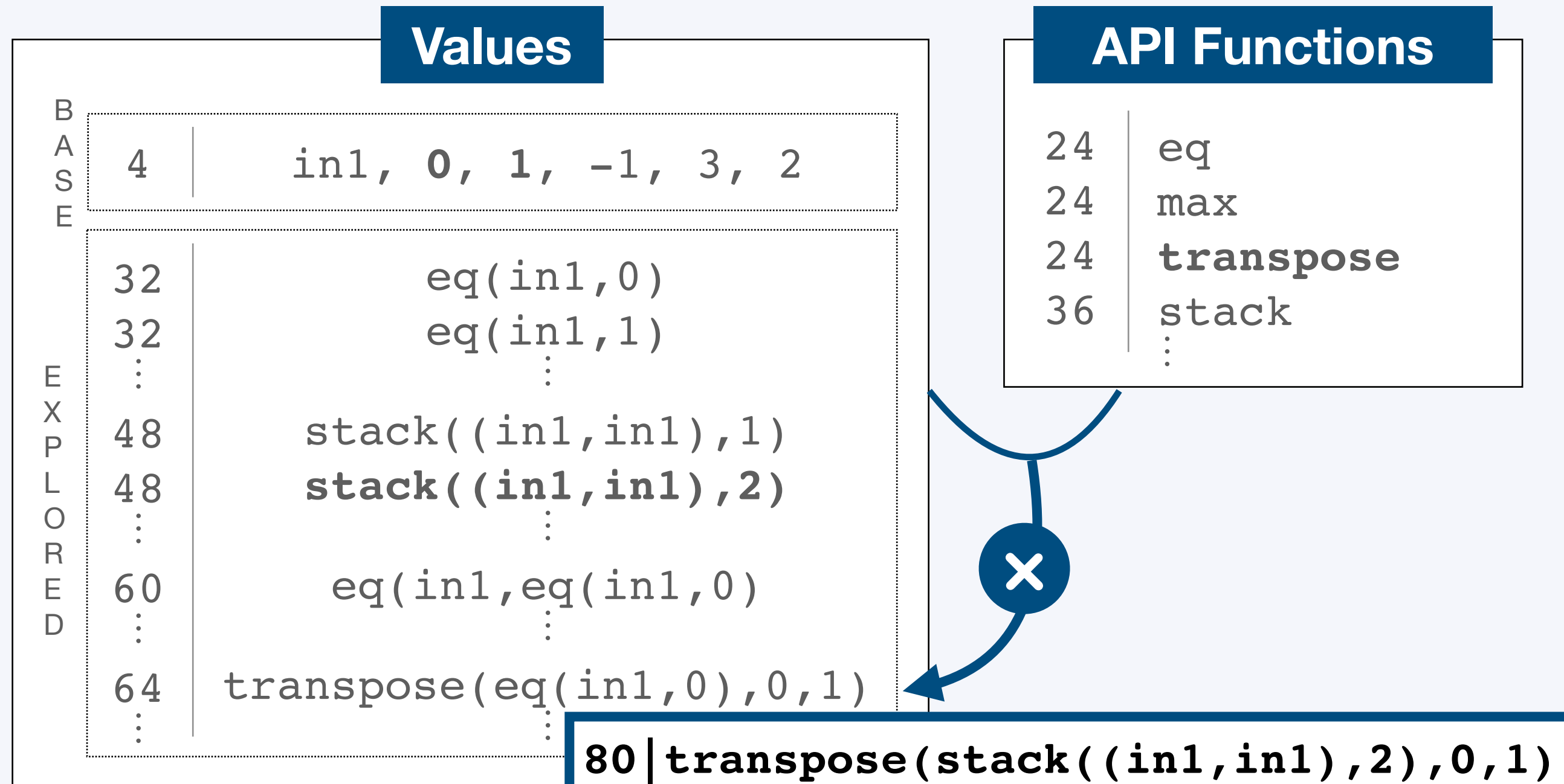
Predictive Synthesis of API-Centric Code

Daye Nam*, Baishakhi Ray*, Seohyun Kim, Xianshan Qu, Satish Chandra

TL;DR: We accelerate enumerative program synthesis by using predictions from our ML model, which is trained to compositionally predict the sequence of API functions, that are needed to transform a given input into a desired output.

Enumerative Search

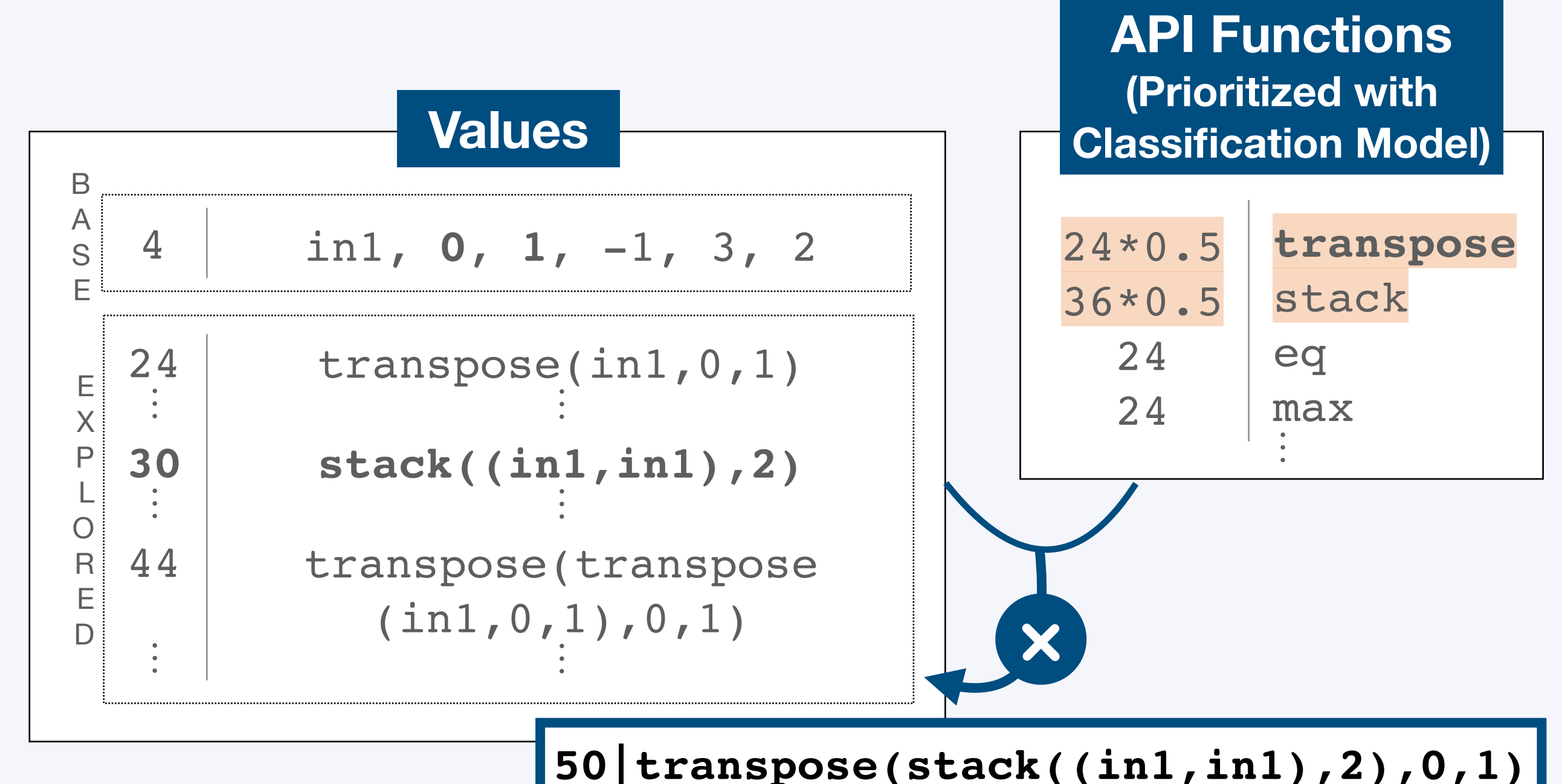
Weighted enumerative synthesis without ML model incorporation



Found	Not Found	Time	Max	Median
18	0	10.01	96.53	0.46

Enumerative Search + Prioritization

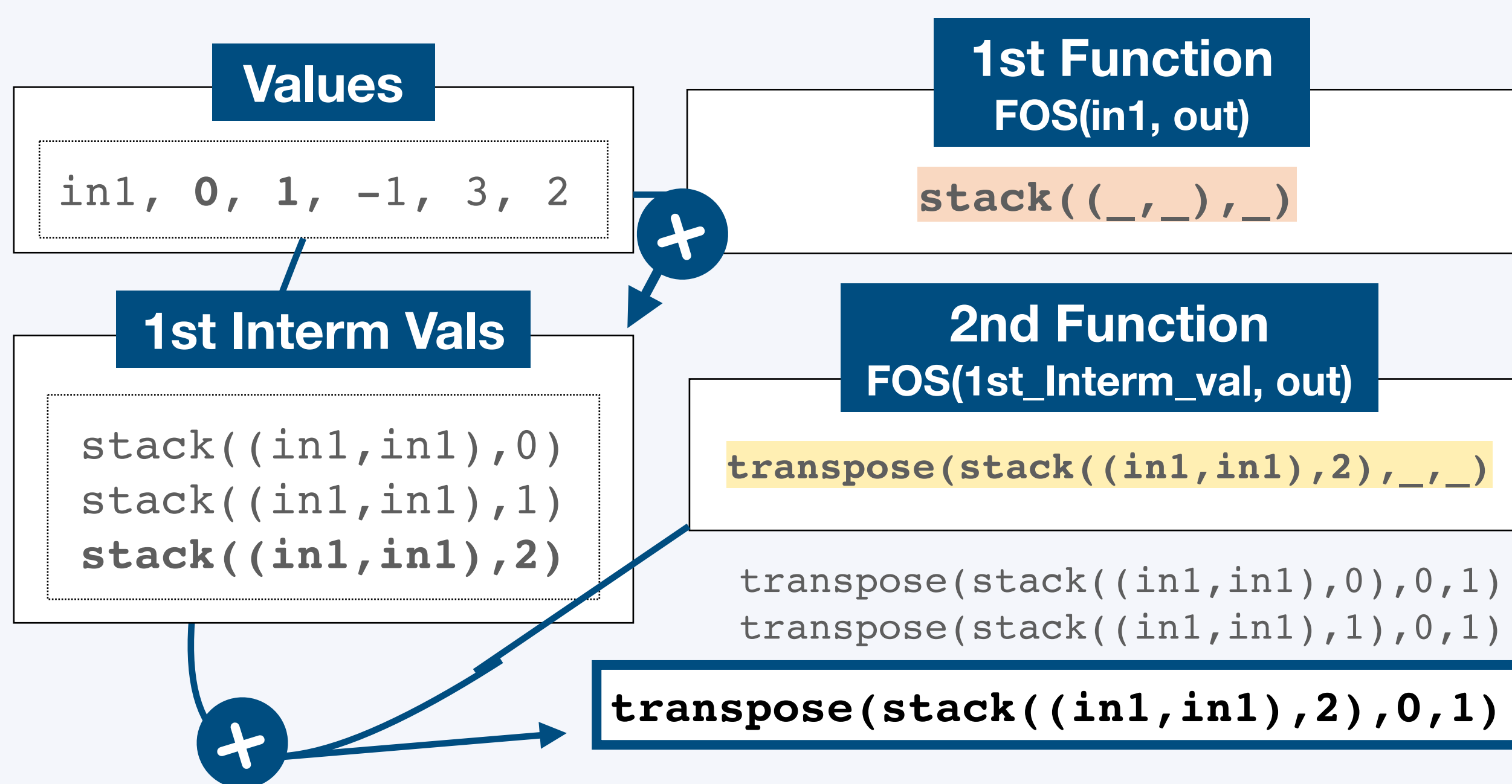
Weighted enumerative synthesis with one-time ML-based prioritization



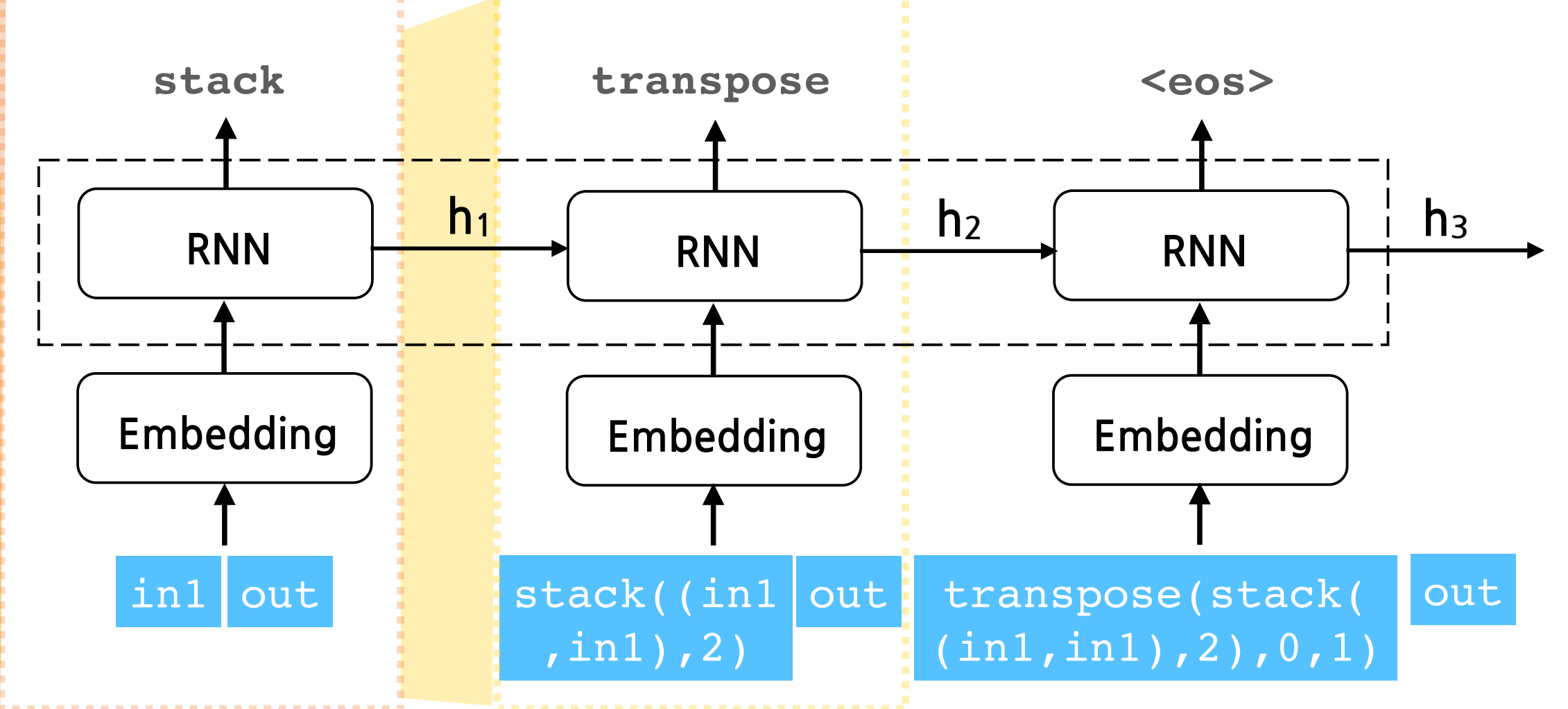
Found	Not Found	Time	Max	Median
18	0	7.44	77.00	0.32

Enumerative Search + First-Of-Sequence (FOS)

Given an input for the next API function and a final output, the compositional model predicts the most probable API function needs to come in the sequence. The synthesizer computes the intermediate values using the predicted API functions, and iteratively invokes the compositional model.



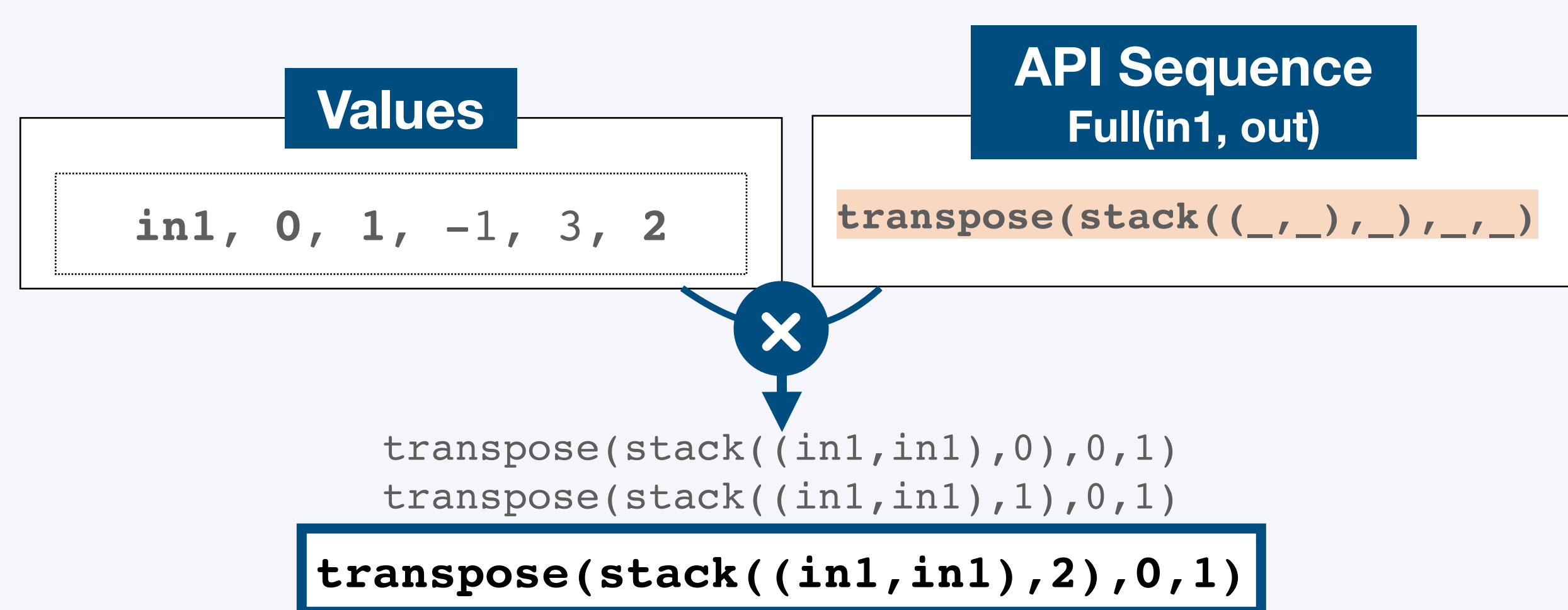
Found	Not Found	Time	Max	Median
17	1	5.87	59.93	0.39



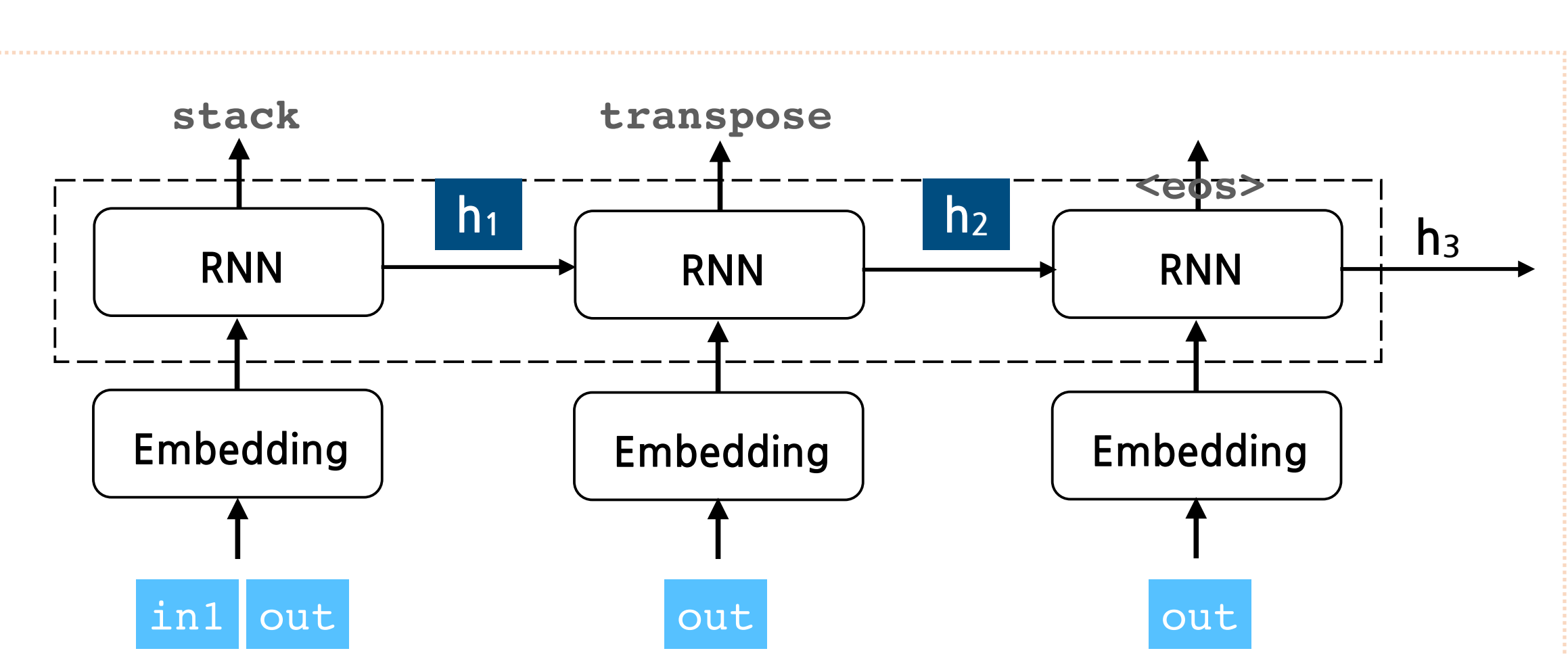
Synthetic (Top-1)	Stack Overflow
66.88	52.38 (Top-1) 76.19 (Top-3)

Enumerative Search + Full Sequence (FUS)

The compositional model predicts a sequence of API functions given the final output and the inputs to each functions. The synthesizer searches only the parameter values (e.g., dimension) that were not provided in the specification.



Found	Not Found	Time	Max	Median
14	3	1.04	9.58	0.25

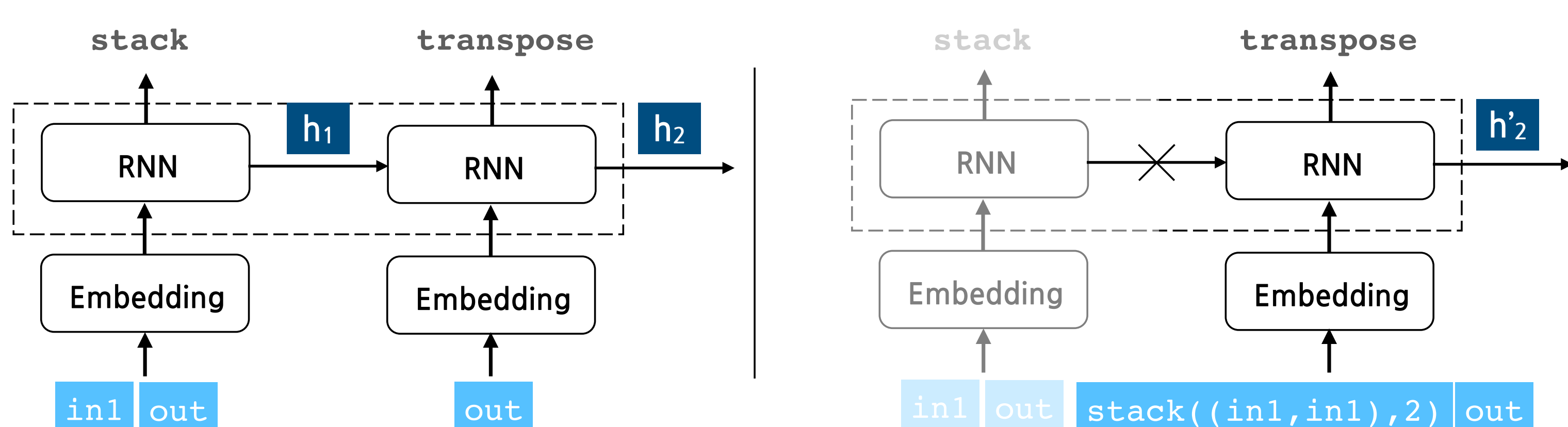


Synthetic (Top-1)	Stack Overflow
79.36	35.29 (Top-1) 76.47 (Top-3)

Why Composition Works?

The model learns to convert its incoming hidden vector to its outgoing hidden vector in a way consistent with the semantics of the API function it predicts, albeit in embedding space.

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



Dataset

The compositional model was trained with the Synthetic dataset. For the model evaluation, both synthetic and Stack Overflow benchmarks were used. To evaluate the program synthesis performance, we only used the Stack Overflow benchmarks.

		Train	Valid	Test
Synthetic	# of unique seqs (Len)	16 (1) + 186 (2)		
	# of in/out values	5.5M	10K	10K
Stack Overflow	# of unique seqs (Len)	Only used for evaluation		
	# of in/out values	8 (1) + 7 (2)		
		18		