Transformers Learn Chain-of-Thought Reasoning with Length Generalization

Yu Huang[1,*] Zixin Wen[2,*] Aarti Singh[2] Yuejie Chi[3] Yuxin Chen[1]

[1]University of Pennsylvania [2] Carnegie Mellon University [3]Yale University

Mechanisms of Long CoT Reasoning

when transformers length-generalize, when they stall, and how recursive self-training pushes the length boundary?

Background

Standard

Q: There were 10 friends playing a video game online when 7 players quit. If each player left had 8 lives, how many lives did they have total?

A: The answer is

(Output) 80. ✗

CoT

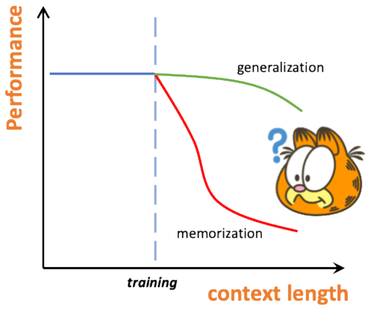
Q: There were 10 friends playing a video game online when 7 players quit. If each player left had 8 lives, how many lives did they have total?

A: Let's think step by step.

(Output) There were 10 friends playing a video game online. This means that, at the start, there were 10 x 8 = 80 lives in total. Then, 7 players quit. This means that 7 x 8 = 56 lives were lost. Therefore, the total number of lives remaining is 80 - 56 = 24. The answer is 24. ✓

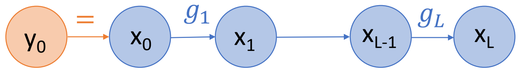
Question / Rationale / Answer

< Question → Answer > < Question → **Rationale** → Answer >



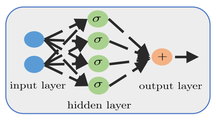
Length generalization: whether models can **extrapolate** to longer CoT beyond training?

State-tracking task



Variable $x \in \mathcal{X}$
Action $g_\ell \in \mathcal{G}$
Value: $y_0 \in \mathcal{Y}$

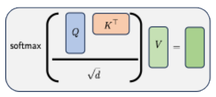
$$x_1 = g_1(x_0), \cdots x_{m+1} = g_{m+1}(x_m) \cdots, x_L = g_L(x_{L-1}), \quad x_0 = y_0, \cdots\cdots, x_m = y_m$$



MLP

1.Retrieve: $g_{m+1},\ y_m$

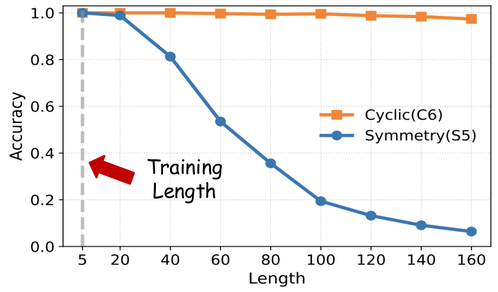
2. Apply $g_{m+1}(y_m)$



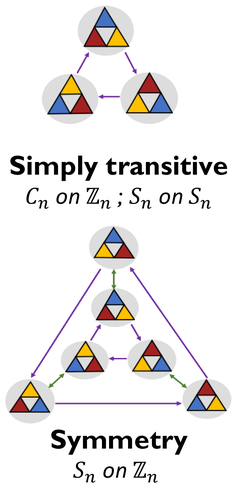
Attention

$x_{m+1} = y_{m+1}$

Does the learned CoT generalize ?



(a) Length generalization results of cyclic (C_6) vs. symmetry (S_5) tasks.



Theorem (informal)

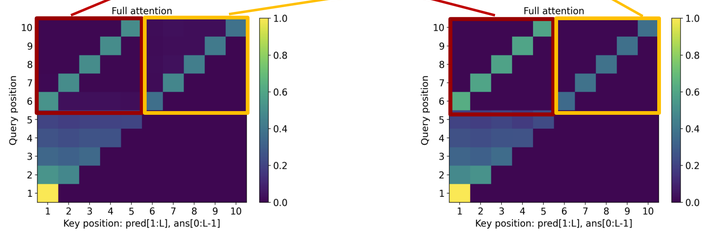
- **Simply transitive:** CoT training on constant-length yields generalization to ***significantly longer*** $d^{\Omega(1)}$.
- **Symmetry:** CoT training only generalizes to a ***constant factor*** of training length.

Message 1: **algebraic structure** of actions dictates how far the reasoning length generalizes

When does length generalization fail?

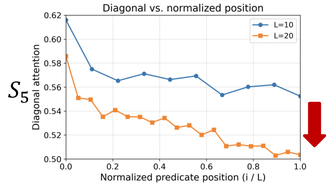
Attention Concentration Mechanism

focus on relevant group context — focus on current answer



(a) Cyclic (C_6) task (b) Symmetry (S_5) task

Message 2: as length increasing, more *irrelevant* context will **dilute** the attention concentration



S_5

Relevant attention dilutes **quickly** due to **many distractors**

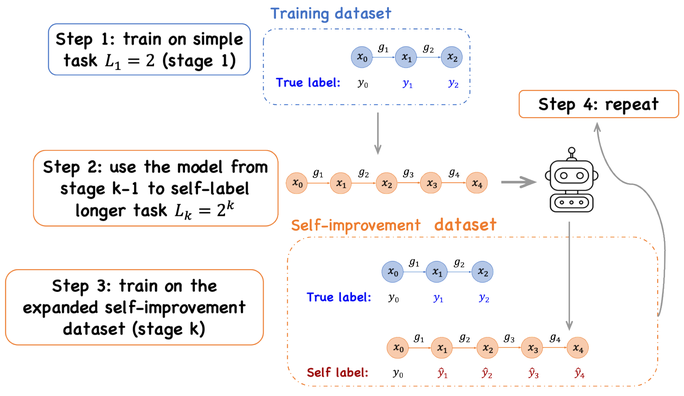
$$\ldots x_n = g_n(x_{n-1}), \cdots x_{m+1} = g_{m+1}(x_m), \cdots\cdots, x_m = y_m$$

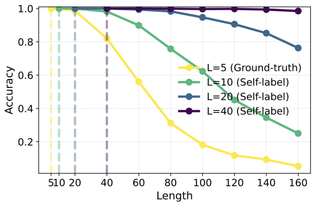
distractors: $g_n \circ y_m = g_{m+1} \circ y_m$

Loss convergence ≠ robust attention in longer context

Recursive self-improvement

If the transformer fails to length-generalize, can it **self-improve** (Lee et al., 2025) its reasoning length?





Theorem (informal)

Transformers trained on self-labeled CoTs of length 2^{k-1} **generalize** to solve task of 2^k

Message 3: recursive self-training provably extends the solvable reasoning length

Theoretical contribution: the **first optimization** guarantee showing constant-depth transformers learn to solve problems beyond TC^0