Self-Supervised Theorem Discovery in a Formal Axiomatic System

Kazuki Ota[1,2] Takayuki Osa[2] Tatsuya Harada[1,2] / [1]The University of Tokyo [2]RIKEN AIP

Motivation

Learning approaches without human-provided examples have brought new insights to humans.

Can we discover useful theorems from axioms alone?

Preliminaries

A minimal Hilbert system defines our testbed with three axioms and Modus Ponens.

Axioms in Hilbert System

Ax1: $A \to (B \to A)$

Ax2: $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

Ax3: $(\neg A \to \neg B) \to (B \to A)$

Inference Rule in Hilbert System

Modus Ponens (MP) $\dfrac{X \quad X \to Y}{Y}$

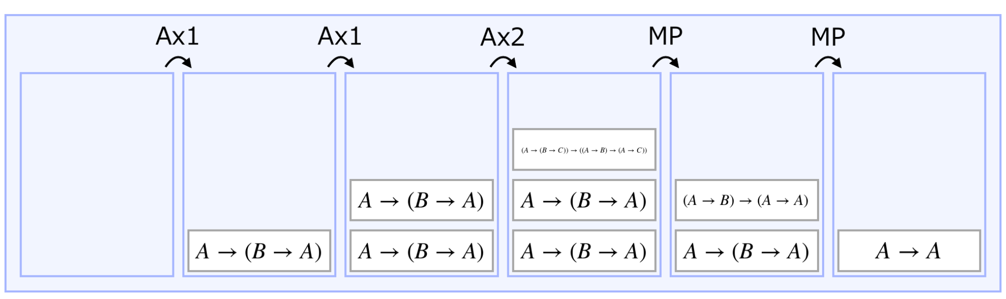
Example proof tree deriving $A \to A$.

$$\dfrac{\overset{\text{Ax1}}{A \to (B \to A)} \qquad \dfrac{\overset{\text{Ax1}}{A \to (B \to A)} \qquad \overset{\text{Ax2}}{(A \to (B \to C)) \to ((A \to B) \to (A \to C))}}{(A \to B) \to (A \to A)}\,\text{MP}}{A \to A}\,\text{MP}$$

Our Formulation

Proof trees become stack-machine decisions by flattening proof construction into action sequences.

A stack-machine formulation of the proof tree for $A \to A$.

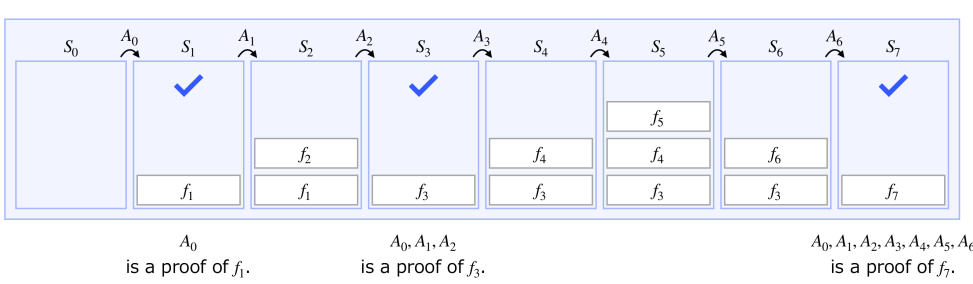


Our Algorithm

Self-Supervised Policy Learning

During proof search, each single-formula stack is treated as a reached theorem, even if it differs from the original goal.

Single-formula stacks become new goals, and their action prefixes provide training examples.



Reached formulas become training goals when successful proof prefixes are reused as supervision.

Algorithm 1 Policy Learning in a Fixed Action Space

1: **Input:** action space $\mathcal{A}$.
2: Initialize a new policy $\pi_\theta(A|S, g)$ for $\mathcal{A}$.
3: Bootstrap the goal buffer $\mathcal{G}$ with random rollouts.
4: **for** each iteration :
5: Initialize current training set $\mathcal{D}_{\text{train}} \leftarrow \emptyset$.
6: **for** each episode :
7: Sample a goal formula g uniformly from $\mathcal{G}$.
8: Roll out the current policy $\pi_\theta(\cdot|S, g)$ for at most T steps.
9: **for** each time t at which S_t is a single-formula stack containing f_t :
10: Add the reached theorem f_t to the goal buffer $\mathcal{G}$.
11: Add the found proof path $\{(S_{t'}, A_{t'}, f_t) \mid 0 \le t' < t\}$ to $\mathcal{D}_{\text{train}}$.
12: **end for**
13: **end for**
14: Update π_θ by minimizing cross-entropy loss on $\mathcal{D}_{\text{train}}$.
15: **end for**
16: **return** goal buffer $\mathcal{G}$.

Selecting Useful Theorems

Useful theorems are selected by favoring candidates that are general and difficult to reprove.

Theorem Ranking

	No.	Formula	Proof	Sampled	Solved	Ratio	General	Selected
	1	$(A \to (B \to C)) \to (D \to ((A \to B) \to (A \to C)))$	Ax2 Ax1 Ax1 MP Ax2 MP MP	2443	0	0.04%	✓	✓
Special	2	$A \to ((B \to (C \to D)) \to (B \to (C \to D)))$	Ax2 Ax1 Ax2 MP MP Ax1 MP	2435	0	0.04%		
	3	$(\neg A \to \neg((B \to (C \to D)) \to ((B \to C) \to (B \to D)))) \to A$	Ax2 Ax1 MP Ax3 Ax2 MP MP	2416	0	0.04%	✓	✓
	4	$((A \to B) \to A) \to ((A \to B) \to B)$	Ax1 Ax1 Ax2 MP MP Ax2 MP	2398	0	0.04%	✓	✓
	5	$((A \to (B \to C)) \to (((A \to B) \to (A \to C)) \to D)) \to ((A \to (B \to C)) \to D)$	Ax2 Ax1 MP Ax2 Ax2 MP MP	2379	0	0.04%	✓	✓
	6	$(A \to \neg\bot) \to (A \to A)$	Ax2 Ax3 Ax1 MP Ax2 MP MP	2362	0	0.04%		
	7	$(\neg A \to \neg(B \to (C \to B))) \to A$	Ax1 Ax1 MP Ax3 Ax2 MP MP	2334	0	0.04%	✓	✓
	8	$(\neg(A \to B) \to \neg C) \to ((C \to A) \to (C \to B))$	Ax3 Ax2 Ax1 MP Ax2 MP MP	2492	1	0.08%	✓	✓
	9	$(\neg\bot \to \neg\bot) \to \neg\bot$	Ax3 Ax1 Ax2 MP Ax2 MP MP	2441	1	0.08%		
	10	$(\neg A \to \neg B) \to (C \to (B \to A))$	Ax3 Ax1 Ax1 MP Ax2 MP MP	2370	1	0.08%	✓	✓
	11	$\neg A \to (A \to B)$	Ax1 Ax3 Ax1 MP Ax2 MP MP	2449	2	0.12%	✓	✓
	12	$(A \to B) \to ((C \to A) \to (C \to B))$	Ax1 Ax2 Ax1 MP Ax2 MP MP	2613	3	0.15%	✓	✓
	13	$(A \to B) \to (A \to B)$	Ax1 Ax2 MP Ax1 Ax2 MP MP	2399	3	0.17%		
	14	$(A \to (B \to C)) \to (A \to (B \to C))$	Ax2 Ax1 Ax2 MP MP	2357	4	0.21%		
	15	$A \to (B \to (C \to A))$	Ax1 Ax1 Ax1 MP Ax2 MP MP	2343	4	0.21%	✓	✓
	16	$(\neg\neg A \to \neg\neg B) \to (A \to B)$	Ax3 Ax3 Ax1 MP Ax2 MP MP	2313	4	0.22%	✓	✓
	17	$((A \to (B \to C)) \to A) \to ((A \to (B \to C)) \to (B \to C))$	Ax2 Ax1 Ax2 MP MP Ax2 MP	2375	5	0.25%		
	18	$((\neg A \to \neg B) \to B) \to ((\neg A \to \neg B) \to B)$	Ax3 Ax2 MP Ax1 Ax2 MP MP	2361	7	0.34%		
	19	$((A \to (B \to C)) \to (A \to B)) \to ((A \to (B \to C)) \to (A \to B))$	Ax2 Ax2 MP Ax1 Ax2 MP MP	2409	17	0.75%		
	20	$((\neg A \to \neg B) \to ((B \to A) \to C)) \to ((\neg A \to \neg B) \to C)$	Ax3 Ax1 MP Ax2 Ax2 MP MP	2377	283	11.94%	✓	✓
	21	$(\neg A \to \neg(\neg A \to \neg A)) \to A$	Ax1 Ax2 MP Ax3 Ax2 MP MP	2395	682	28.49%	✓	✓
	22	$(A \to (B \to C)) \to (((A \to B) \to A) \to ((A \to B) \to C))$	Ax2 Ax2 Ax1 MP Ax2 MP MP	2381	1051	44.15%	✓	✓
	23	$A \to ((\neg\bot \to \neg\bot) \to (\neg\bot \to \neg\bot))$	Ax3 Ax2 MP Ax3 MP Ax1 MP	2337	1312	56.14%		
	24	$(A \to ((B \to A) \to C)) \to (A \to C)$	Ax1 Ax1 MP Ax2 Ax2 MP MP	2493	1607	64.45%	✓	✓
General	25	$A \to (B \to B)$	Ax1 Ax1 Ax2 MP MP Ax1 MP	2429	1635	67.30%	✓	✓
	26	$A \to \neg\bot$	Ax1 Ax2 MP Ax3 MP Ax1 MP	2407	1655	68.74%		

⋮

Reusing Useful Theorems

Selected theorems become reusable lemma actions in later generations of proof search.

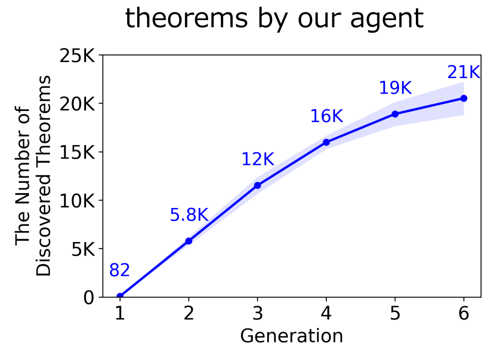
Algorithm 2 Multi-Generation Library Expansion

1: Initialize action space $\mathcal{A} \leftarrow \{\texttt{Ax1}, \texttt{Ax2}, \texttt{Ax3}, \texttt{MP}\}$.
2: **for** each generation :
3: Give action space $\mathcal{A}$ to Algorithm 1 and obtain theorem buffer $\mathcal{G}$.
4: Extract useful theorems $\{\texttt{Thm1}, \dots, \texttt{ThmN}\}$ from $\mathcal{G}$ using generality and reprovability criteria.
5: Expand action space $\mathcal{A} \leftarrow \mathcal{A} \cup \{\texttt{Thm1}, \dots, \texttt{ThmN}\}$ for the next generation.
6: **end for**

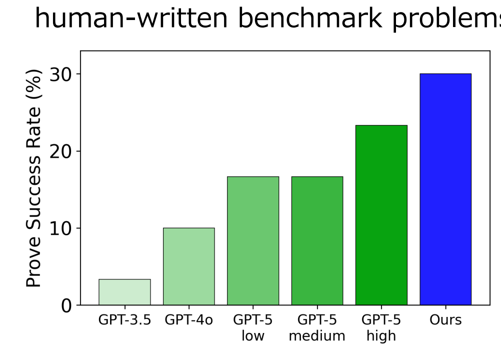
Experimental Results

The agent discovers thousands of theorems and solves human-written benchmark problems.

The number of discovered theorems by our agent

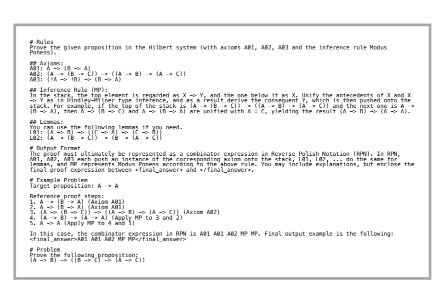


Proof success rates on human-written benchmark problems

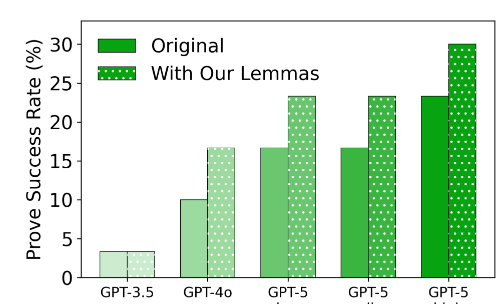


Extracted theorems help LLMs solve more Hilbert-system proof problems when provided in prompts.

Prompts for LLM agents



Success rates improvements with lemmas extracted by our agent



Discovered lemmas can be practically useful even when their role is not obvious from inspection.

Goal Formula: $g : A \to (\neg A \to B)$.

Lemmas extracted by our agent and utilized by GPT-5-high: $\texttt{Thm1} : \neg A \to (A \to B)$, $\texttt{Thm2} : (A \to (B \to C)) \to (D \to ((A \to B) \to (A \to C)))$.

Proof path which GPT-5-high found: Ax1, Thm1, Thm2, MP, Ax2, MP, MP.

Conclusion

Useful theorems can emerge from self-supervised proof search without human-provided theorem libraries.