Self-Learned Formula Synthesis in Set Theory

Chad E. Brown Thibault Gauthier

Czech Technical University in Prague.

This work has been supported by the European Research Council (ERC) grant AI4REASON no. 649043 under the EU-H2020 programme.

September 14, 2020

Motivation

Understanding formulas is important for theorem proving.

How do we figure out what a formula means?

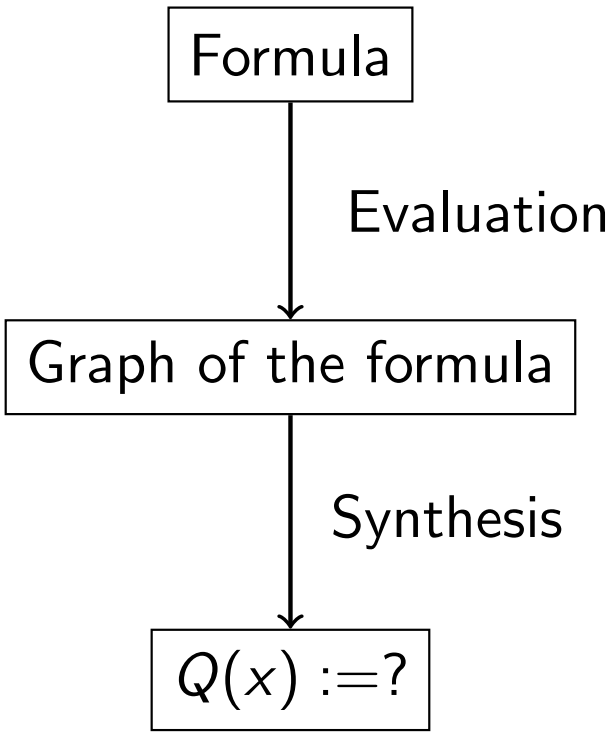
Example formula

What is the meaning of this set-theoretical formula?

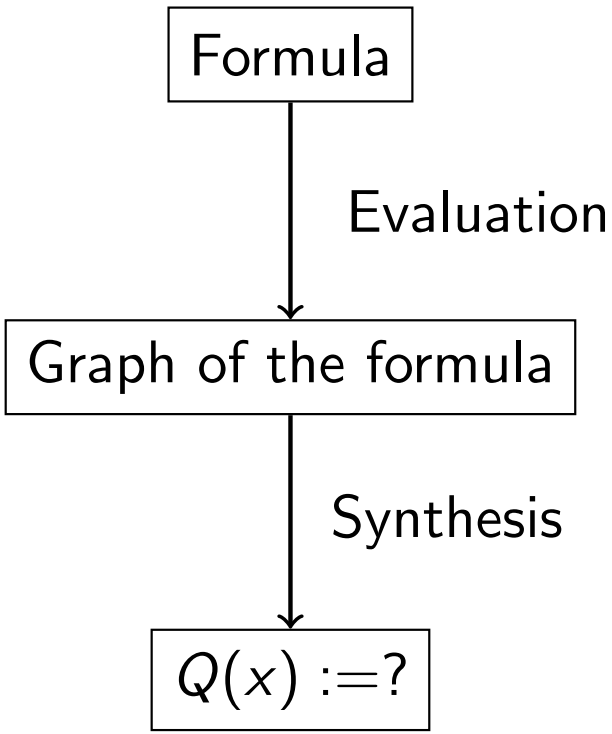
$$\exists y \in x.\ x \not\subseteq \wp(y)$$

There exists a set y member of x, such that
x is not a subset of the power set of y.

Outline



Outline



$$P(x) := \exists y \in x.\ x \not\subseteq \wp(y)$$

$$P(a_0), P(a_1), \ldots, P(a_{63})$$

Vocabulary

Terms s, t :

$$x, y, z, \ldots \qquad \varnothing \qquad \wp(t) \qquad \{t\} \qquad s \cup t$$

Atomic formulas φ, ψ :

$$s \in t \qquad s \notin t \qquad s \subseteq t \qquad s \not\subseteq t \qquad s = t \qquad s \neq t$$

Composite formulas φ, ψ :

$$\varphi \Rightarrow \psi \qquad \varphi \wedge \psi$$

$$\forall x \in s.\ \varphi \quad \forall x \subseteq s.\ \varphi \qquad \exists x \in s.\ \varphi \quad \exists x \subseteq s.\ \varphi$$

Example formula: understanding a special case

$$P(x) := \exists y \in x.\ x \not\subseteq \wp(y)$$

$$P(\varnothing) := \exists y \in \varnothing.\ \varnothing \not\subseteq \wp(y)$$

$P(\varnothing)$ is false.

What is a finite set?

Ground terms a,b:

$$\varnothing \qquad \wp(a) \qquad \{a\} \qquad a \cup b$$

The set $\{\varnothing, \{\varnothing\}\}$ can be constructed as $\{\varnothing\} \cup \{\{\varnothing\}\}$

Enumerating finite sets ($f : \mathbb{N} \to$ finite sets)

$$f(10) \to f(0101) \to \{f(1), f(3)\} \to \{f(1), f(11)\} \to$$

$$\{\{f(0)\}, \{f(0), f(1)\}\} \to \{\varnothing, \{\varnothing, \{\varnothing\}\}\}$$

Idea: position of the 1s in the inverted binary encoding.

Example formula: truth values on initial sets

$P(x) := \exists y \in x.\ x \not\subseteq \wp(y)$

$P(0)$ *is false*.
$P(1)$ *is false*.
$P(01)$ *is false*.
$P(11)$ *is true*.
$P(001)$ *is false*.
$P(101)$ *is true*.
$P(011)$ *is true*.
$P(111)$ *is true*.
$P(0001)$ *is false*.
$\ldots P(63)$

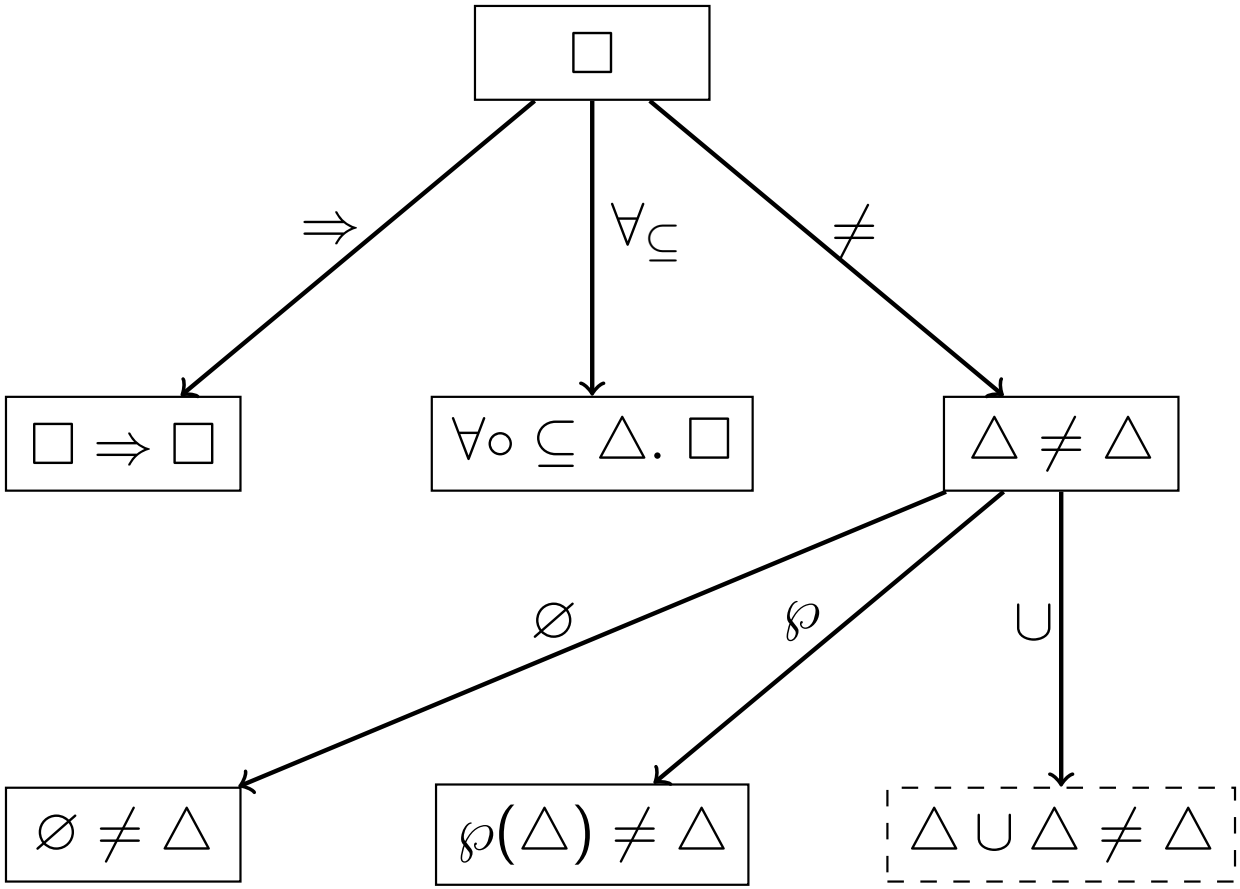
Graph of P : $FFFTFTTTF\ldots$

Synthesis problem

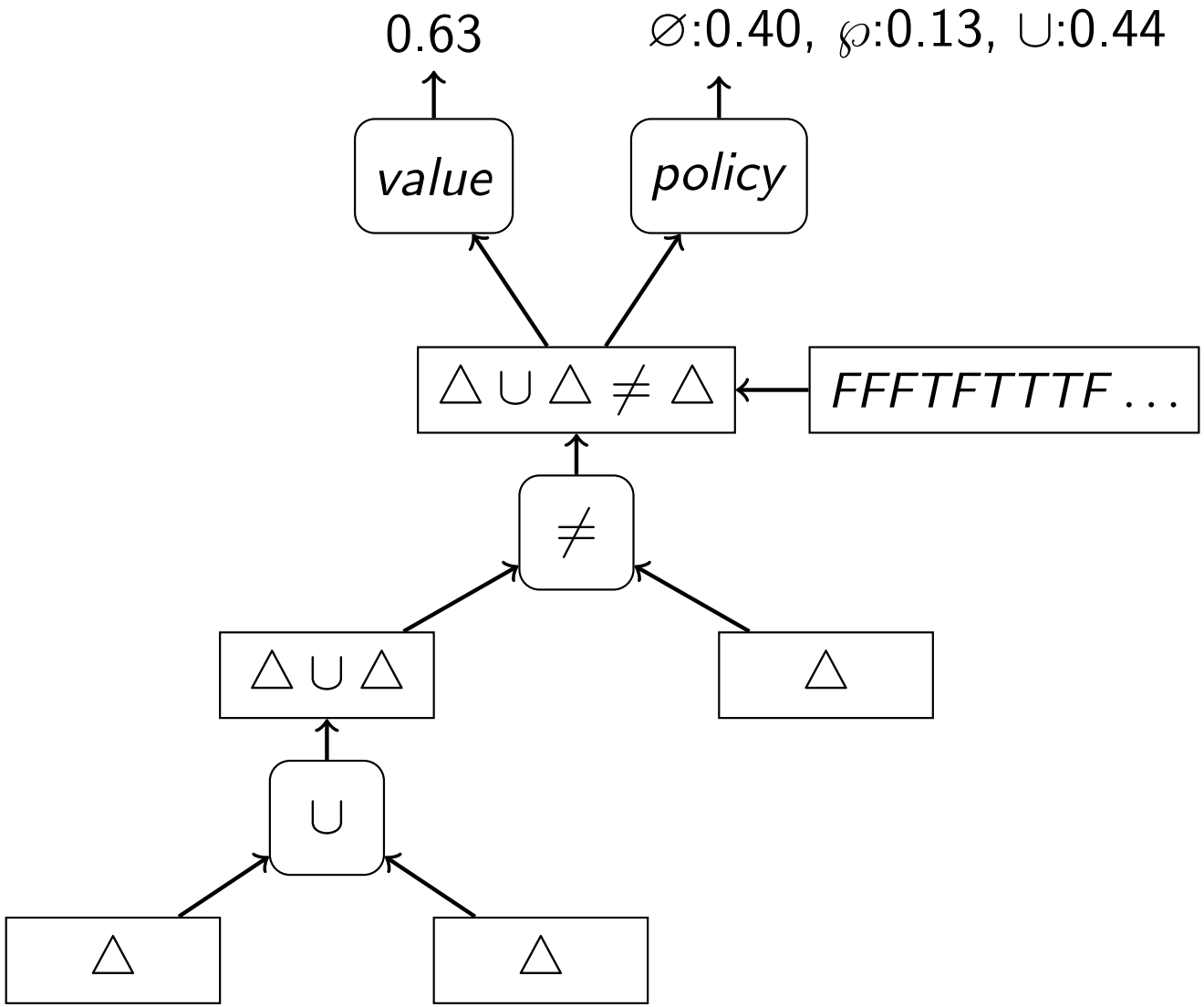
Given a graph (list of truth values), can we find a formula for it?

Reinforcement learning solution

Search tree, policy and value



Tree neural networks



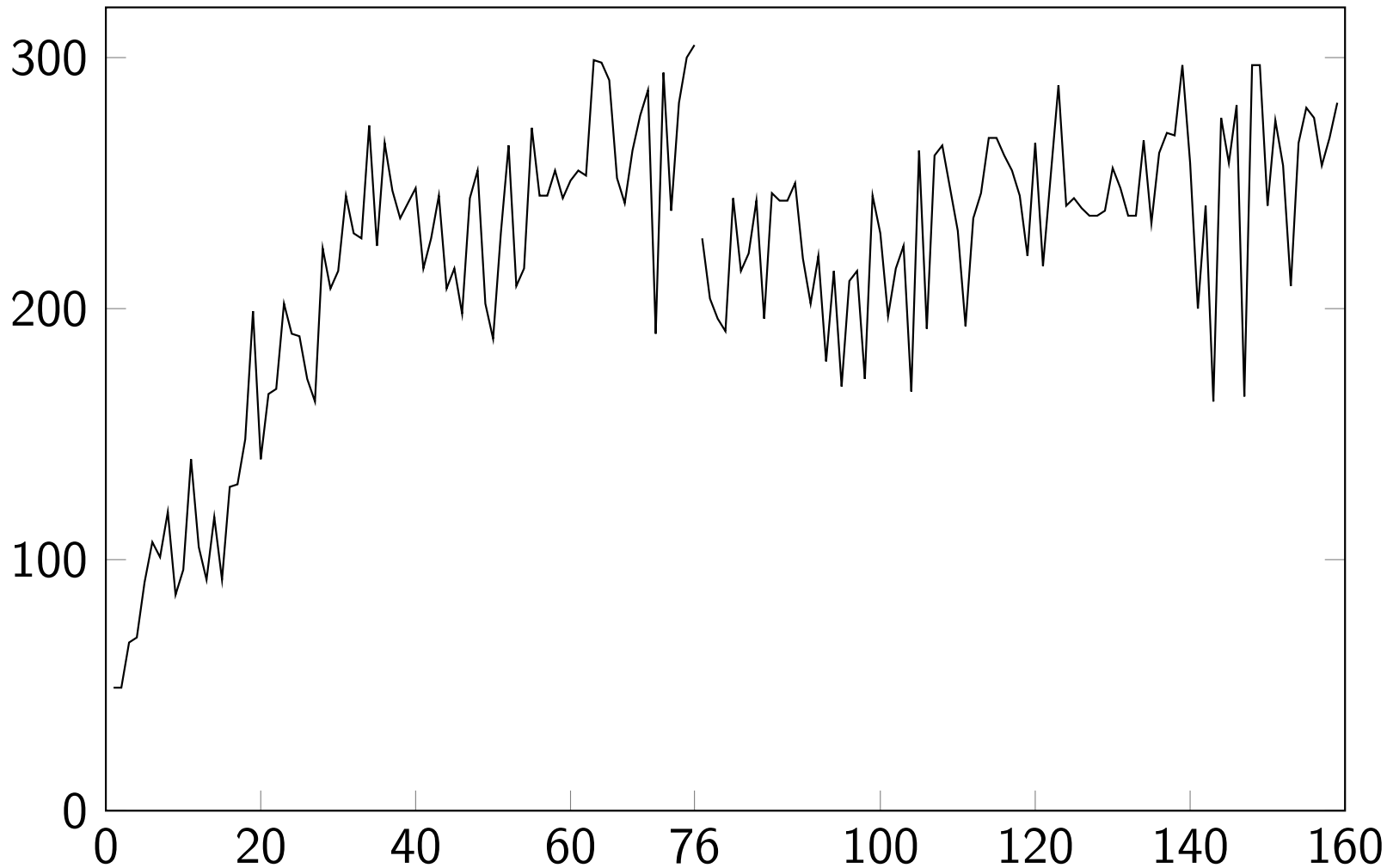
Training and testing datasets

3	4	5	6	7	8	9	10	11	12	13	14	15
6	8	22	60	88	260	472	960	638	992	1582	1056	606

Table: Number of generated graphs of each size

Level 1: 400 graphs
Level 2: 400 graphs
Level 3: 400 graphs
...

Progress of the training run



Number of successful formula synthesis (y) at generation (x)
Level 1 on the left, Level 2 on the right

Final evaluation

Abstract time limit of 50,000 search steps.

	Uniform search	Hidden-graph	Guided
Level 1, 2, 3	68, 0, 0	270, 126, 59	338, 240, 165

Example formula

$$P(x) := \exists y \in x.\ x \not\subseteq \wp(y)$$

Graph of P : $FFFTFTTTF\ldots$

Example formula

$$P(x) := \exists y \in x.\ x \not\subseteq \wp(y)$$

Graph of P : $FFFTFTTTF\ldots$

$$Q(x) := \exists y \in x.\ \{y\} \neq x$$

Summary

Teach an algorithm to synthesize formulas from graphs.

Procedure for understanding a formula with one free variable:

1) Create its graph.
2) Synthesize a (new) formula.
3) Is the new formula more meaningful?

Improvements

Synthesis of combinators and Diophantine equations
(published at LPAR 2020)

1) Self-determined levels.

2) Comparison with ATPs: Vampire, E-prover.