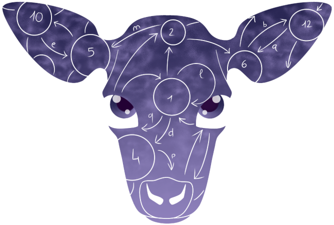
CALF: Categorical Automata Learning Framework

Gerco van Heerdt Matteo Sammartino Alexandra Silva

University College London

August 23, 2017



The L* algorithm (Angluin, 1987)

Finite alphabet A

System behaviour captured by a **regular language** $\mathcal{L} \subseteq A^*$

L* learns *minimal* DFA for $\mathcal{L}$ assuming an *oracle* that answers

- **Membership queries**

$$\boxed{w \in \mathcal{L}?}$$

- **Equivalence queries**

$$\boxed{\mathcal{L}(H) = \mathcal{L}?}$$

Negative result $\implies$ *counterexample*

Applications of L*

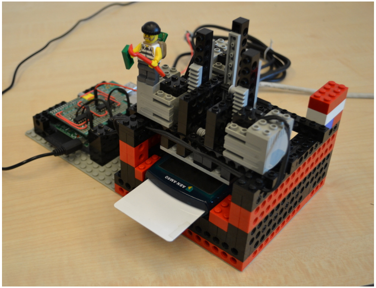
Through learning, verification methods for automata become available for black box systems

- Network protocols
- Devices such as smartcard readers
- Legacy software
- ...

Applications of L*

Through learning, verification methods for automata become available for black box systems

- Network protocols
- Devices such as smartcard readers
- Legacy software
- ...



Source: *Automated Reverse Engineering using Lego®*
Chalupar et al., WOOT 2014

Problem: ad hoc adaptations

$\boxed{\text{weighted}}$ $\boxed{\text{Büchi}}$ $\boxed{\text{non-deterministic}}$ $\boxed{\text{tree}}$ $\boxed{\text{universal}}$ $\boxed{\text{Mealy}}$ $\boxed{\text{alternating}}$ $\boxed{\text{probabilistic}}$ $\boxed{\text{register}}$

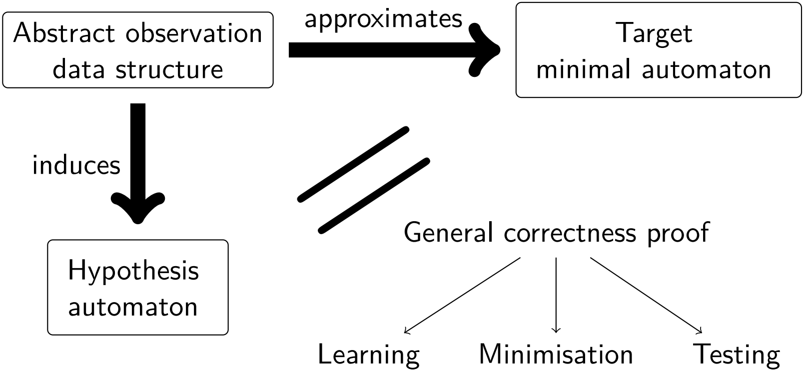
...

Complex automata $\Longrightarrow$

- Hard to adapt algorithm
- Complex correctness proofs

Solution: **category theory**

Contributions



Deterministic automaton

Deterministic automaton: **set** Q with **functions**

$$\begin{tikzcd}
1 \arrow[dr, "\mathsf{init}"] & & 2 \\
& Q \arrow[ur, "\mathsf{out}"] \arrow[d, "\delta"] & \\
& Q^A &
\end{tikzcd}$$

Other automata

Nominal automaton:[1] **nominal set** Q with **equivariant functions**

$$\begin{tikzcd} 1 \arrow[dr, "\mathrm{init}"] & & 2 \\ & Q \arrow[ur, "\mathrm{out}"] \arrow[d, "\delta"] & \\ & Q^A & \end{tikzcd}$$

Linear weighted automaton: **vector space** Q with **linear maps**

$$\begin{tikzcd} \mathbb{F} \arrow[dr, "\mathrm{init}"] & & \mathbb{F} \\ & Q \arrow[ur, "\mathrm{out}"] \arrow[d, "\delta"] & \\ & Q^A & \end{tikzcd}$$

[1] **Learning Nominal Automata (POPL 2017);** Joshua Moerman, Matteo Sammartino, Alexandra Silva, Bartek Klin, Michał Szynwelski

Categorical automaton

An automaton in a category $\mathbf{C}$ is an **object** Q with **morphisms**

$$\begin{tikzcd}
I \arrow[dr, "\mathsf{init}"] & & O \\
& Q \arrow[ur, "\mathsf{out}"] \arrow[d, "\delta"] & \\
& FQ &
\end{tikzcd}$$

initial state object

output object

dynamics functor $\mathbf{C} \to \mathbf{C}$

$\mathtt{L}^\star$ observation table

$\mathtt{L}^\star$ maintains $S, E \subseteq A^*$ inducing a table

$$
\begin{array}{r|cc}
 & \overbrace{\varepsilon \quad \mathsf{a}}^{E} & \\
\end{array}
$$

		ε	a
S	ε	1	0
	a	0	1
	aa	1	0
$S \cdot A$	aaa	0	1

$\mathtt{L}^\star$ observation table

$\mathtt{L}^\star$ maintains $S, E \subseteq A^*$ inducing a table

$$
\begin{array}{rr|cc}
 & & \multicolumn{2}{c}{\overbrace{\qquad\qquad}^{E}} \\
 & & \varepsilon & \mathsf{a} \\
\hline
\multirow{3}{*}{$S \left\{ \right.$} & \varepsilon & 1 & 0 \\
 & \mathsf{a} & 0 & 1 \\
 & \mathsf{aa} & 1 & 0 \\
\hline
S \cdot A & \mathsf{aaa} & 0 & 1
\end{array}
$$

Prepend *row label* to *column label* and pose **membership query**

$$
(s, e) \mapsto \begin{cases} 1 & \text{if } se \in \mathcal{L} \\ 0 & \text{if } se \notin \mathcal{L} \end{cases}
$$

L* observation table

L* maintains $S, E \subseteq A^*$ inducing a table

$$\begin{array}{rl|cc}
 & & \multicolumn{2}{c}{\overbrace{\qquad\quad}^{E}} \\
 & & \varepsilon & \mathsf{a} \\ \hline
 & \varepsilon & 1 & 0 \\
S & \mathsf{a} & 0 & 1 \\
 & \mathsf{aa} & 1 & 0 \\ \hline
S \cdot A & \mathsf{aaa} & 0 & 1
\end{array}$$

$\mathcal{L} = \{a^n \mid n \text{ is even}\}$

$aa \cdot a \notin \mathcal{L}$

Prepend *row label* to *column label* and pose **membership query**

$$(s, e) \mapsto \begin{cases} 1 & \text{if } se \in \mathcal{L} \\ 0 & \text{if } se \notin \mathcal{L} \end{cases}$$

L* hypothesis DFA

Hypothesis states are upper rows of the table; transitions append symbols to row labels

	ε	a
ε	1	0
a	0	1
aa	1	0
aaa	0	1

Requires properties **closedness** and **consistency** to be well-defined

L* hypothesis DFA

Hypothesis states are upper rows of the table; transitions append symbols to row labels

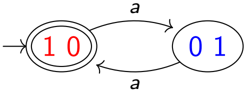
	ε	a
ε	1	0
a	0	1
aa	1	0
aaa	0	1

Requires properties **closedness** and **consistency** to be well-defined

L* hypothesis DFA

Hypothesis states are upper rows of the table; transitions append symbols to row labels

	ε	a
ε	1	0
a	0	1
aa	1	0
aaa	0	1



Requires properties **closedness** and **consistency** to be well-defined

L* algorithm overview

table updated using
membership queries

1. Initialise $S = E = \{\varepsilon\}$
2. Satisfy closedness and consistency (by augmenting S and E)
3. Construct hypothesis
4. Pose equivalence query
5. On a counterexample, add its prefixes to S and repeat from 2

Main observation

The state space of the hypothesis is the image of the composition

$$S \xrightarrow{\text{inclusion}} A^* \xrightarrow{\text{state reached}} \underbrace{M}_{\boxed{\text{state space of minimal DFA accepting } \mathcal{L}}} \xrightarrow{\text{language accepted}} 2^{A^*} \xrightarrow{\text{restriction}} 2^E$$

M is the **target** of the algorithm

Select states of M using S

Classify states of M into 2^E

Wrapper

Wrapper for **target** T consists of objects S and P with morphisms

$$(S \xrightarrow{\sigma} T, T \xrightarrow{\pi} P)$$

- σ **selects** from T
- π **classifies** T

Define the (unstructured) *hypothesis* as the image of

$$S \xrightarrow{\quad\sigma\quad} T \xrightarrow{\quad\pi\quad} P$$

Categorical setting uses a *factorisation system*

Additional structure

If T comes with a coalgebra $T \xrightarrow{f} FT$, we can wrap that as well:

$$S \xrightarrow{\sigma} T \xrightarrow{f} FT \xrightarrow{F\pi} FP$$

$$\Downarrow$$

f-**closedness** and f-**consistency** properties

$$\Downarrow$$

compatible F-coalgebra on hypothesis

$\mathtt{L}^\star$ definitions recovered by $f = \delta\colon M \to M^A$

Recovering the target

$$S \xrightarrow{\sigma} T \xrightarrow{\pi} P$$

Conditions for a hypothesis isomorphic to the target:

- **Selecting everything:** σ surjective
- **Classifying faithfully:** π injective

Imply every notion of closedness and consistency

Isomorphism preserves resulting structures

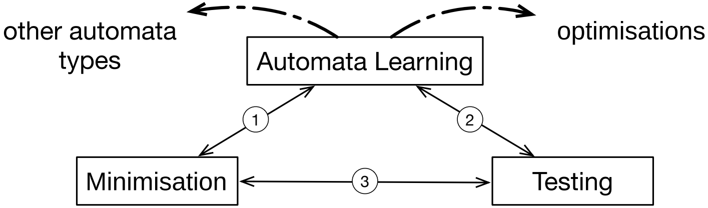
Categorically, surjective/injective defined by factorisation system

Main correctness theorem

For certain automata, either of the following is sufficient:

- selecting everything and consistency
- classifying faithfully and closedness

CALF



Project: `calf-project.org`

Learning Automata with Side-Effects:
`https://arxiv.org/abs/1704.08055`

Future work

- Describing non-trivial ad hoc automata as categorical ones
- Optimisations in categories other than **Set**
- Implementation
- Integrating testing into $\mathtt{L}^\star$

Computing wrapped morphisms

The composition

$$S \longrightarrow A^* \longrightarrow M \longrightarrow 2^{A^*} \longrightarrow 2^E$$

is known because the $A^* \to 2^{A^*}$ part depends only on $\mathcal{L}$

Reachability/language maps preserve transition structure:

symbol concatenation

$$\begin{tikzcd}
S \arrow[r] & A^* \arrow[r, "r"] \arrow[d] & M \arrow[r, "o"] \arrow[d, "\delta"] & 2^{A^*} \arrow[d] & \\
& (A^*)^A \arrow[r, "r^A"] & M^A \arrow[r, "o^A"] & (2^{A^*})^A \arrow[r] & (2^E)^A
\end{tikzcd}$$

Brzozowski derivative

Closedness and consistency

$$
\begin{tikzcd}
S \arrow[r, "\sigma"] \arrow[dr, bend right, two heads, "e"'] & T \arrow[r, "\pi"] & P \\
& H \arrow[ur, tail, bend right, "m"'] &
\end{tikzcd}
\qquad
S \xrightarrow{\sigma} T \xrightarrow{f} FT \xrightarrow{F\pi} FP
$$

The wrapper is *f-closed* if there is a morphism close making the left triangle commute

$$
\begin{tikzcd}
S \arrow[r, two heads, "e"] \arrow[d, dotted, "\text{close}"'] \arrow[dr] & H \arrow[d, dotted, "\text{cons}"] \\
FH \arrow[r, tail, "Fm"] & FP
\end{tikzcd}
$$

It is *f-consistent* if there is a morphism cons making the right triangle commute

Structured hypothesis

If f-closedness and f-consistency hold, we have a coalgebra

$$\begin{tikzcd} S \arrow[r, two heads, "e"] \arrow[d, "\mathsf{close}"'] & H \arrow[d, "\mathsf{cons}"] \arrow[dl, dotted, "\theta"'] \\ FH \arrow[r, tail, "Fm"] & FP \end{tikzcd}$$

which is compatible with f:

$$\begin{tikzcd} T \arrow[d, "f"'] & S \arrow[l, "\sigma"'] \arrow[r, "e"] & H \arrow[d, "\theta"] \\ FT \arrow[r, "F\pi"] & FP & FH \arrow[l, "Fm"'] \end{tikzcd}$$