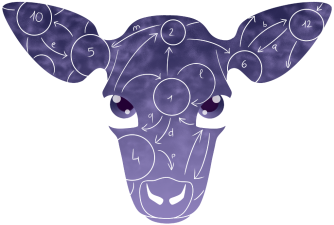
CALF: Categorical Automata Learning Framework

Gerco van Heerdt

July 15, 2020



The $\mathtt{L}^\star$ algorithm (Angluin, 1987)

Finite alphabet A

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

$\mathtt{L}^\star$ 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*

Adaptations

weighted

Büchi

non-deterministic

tree

universal

alternating

Mealy

register

...

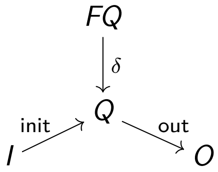
Complex automata $\Longrightarrow$

- Hard to adapt algorithm
- Complex correctness proofs

Solution: **category theory**

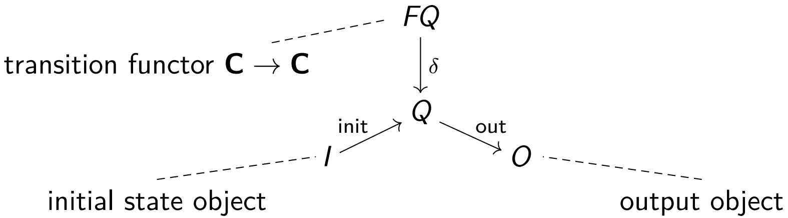
Categorical automaton

An automaton in a category **C** is an **object** Q with **morphisms**



Categorical automaton

An automaton in a category **C** is an **object** Q with **morphisms**



Previous work

Bart Jacobs and Alexandra Silva. "Automata Learning: A Categorical Perspective". In: Horizons of the Mind. A Tribute to Prakash Panangaden. LNCS. 2014.

- Abstract definitions of closedness, consistency, hypothesis
- Minimality proof

but

- No abstract data structure
- No correctness results

Previous work

Gerco van Heerdt. “An Abstract Automata Learning Framework”. MA thesis. Radboud University Nijmegen, 2016.

- Abstract data structure
- Correctness characterisations

but

- No algorithm
- No tree automata

This work

Main contributions:

- Abstract iterative **algorithm**
- Instantiation to generalised tree automata
- Instantiation to weighted automata
- Instantiation to automata with side-effects

Overview

Abstract minimisation (Chapter 3)

Abstract automata learning algorithm (Chapter 4)

Learning weighted automata (Chapter 5)

Learning automata with side-effects (Chapter 6)

Contributions, Chapter 3

Minimisation of automata

- Iterative minimisation (cobase)
- Nerode equivalence

Challenge: using algebras, no final sequence

Contributions, Chapter 4

Abstract automata learning

- Provably correct abstract version of $\mathtt{L}^\star$
- Instantiation to learning generalised tree automata

Challenge: using algebras, no final coalgebra

Contributions, Chapter 5

Learning weighted automata

- General algorithm over semiring, conditional termination proof
- Non-termination over naturals
- Termination over PIDs

Challenge: showing termination for PIDs

Contributions, Chapter 6

Learning automata with side-effects

- Algorithm in category of algebras for a monad
- Succinct hypotheses
- Optimised counterexample handling
- Implementation
- Experiments

Challenge: developing optimisations

Impact

Insight into existing algorithms

- Non-termination of NFA learning algorithm
- Applicability of counterexample optimisation to NFA/WFA/... learning

Development of new algorithms

- Generalised tree automata learning
- Learning weighted automata over PIDs
- Learning automata with side-effects

Future projects

Automata with infinite side-effects

- General conditions on the monad, subsuming PID setting
- Subsequential transducers

Automata accepting infinite words

Register automata

Recognisable languages over monads

- Pomset automata