Automata Learning with an Incomplete Teacher

Seminar Presentation

Jasper Laumen

January 22, 2025

The Papers

First paper:

- Automata Learning with an Incomplete Teacher
- ECOOP 2023
- Mark Moeller, Thomas Wiener, Alaia Solko-Breslin, Caleb Koch, Nate Foster, Alexandra Silva

Second paper:

- Learning Minimal Deterministic Automata from Inexperienced Teachers
- ISoLA 2012
- Martin Leucker, Daniel Neider

Automata Learning

- Closed box inference of DFAs
- Active learning
- MAT framework, iMAT framework
- L^*, $L^*_\square$ algorithms

Applications of automata learning

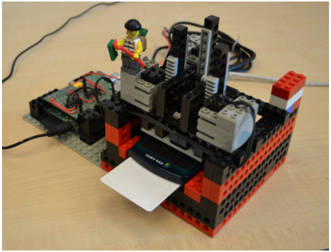
Software Verification

- Regression testing of telecommunication systems (Siemens)
- Testing requirements of a brake-by-wire system (Volvo)

Security

- Smartcards, network protocols

Legacy software (ASML)



MAT Framework

- Minimally adequate teacher / oracle
- Teacher has a regular language $L \subseteq \Sigma^*$
- Active learning based on queries
  - Membership ("yes" / "no")
  - Equality ("correct" / "counterexample")
- Sufficient to determine correct and minimal DFA
- L^* algorithm

iMAT framework

- In practice, oracle is not perfect
  - How to validate equivalence queries?
- Incomplete minimally adequate teacher
- Teacher has sets $L^+ \subseteq \Sigma^*$ and $L^- \subseteq \Sigma^*$, with $L^+ \cap L^- = \varnothing$
- New membership query: "yes" / "no" / "don't care"
- $L^*_\square$ algorithm
- Main subject of the paper

Observation Tables

Given a language $L \subseteq \Sigma^*$, an observation table is a tuple (S, E, T), where

- $S \subseteq \Sigma^*$ is a prefix-closed set of words
- $E \subseteq \Sigma^*$ is a suffix-closed set of words
- $T : (S \cup S \times \Sigma) \times E \to \{+, -\}$ is a map on words

Example with $S = \{\varepsilon, b, a\}$ and $E = \{\varepsilon, ab, b\}$

	ε	ab	b
ε	-	-	+
b	+	+	+
a	-	-	-
ba	+	+	+
bb	+	+	+
aa	-	-	-
ab	-	-	-

Closedness

An observation table (S, E, T) is closed if for every word $w \in S$ and letter $a \in \Sigma$, we have $\text{row}(wa) \in \text{row}(S)$

Closed

	ε	ab	b
ε	-	-	+
b	+	+	+
a	-	-	-
ba	+	+	+
bb	+	+	+
aa	-	-	-
ab	-	-	-

Not Closed

	ε	ab	b
ε	-	-	+
b	+	-	+
a	-	-	-
ba	-	+	-
bb	+	+	+
aa	-	-	-
ab	-	-	-

Distinctness

An observation table (S, E, T) is distinct if for every pair of words $w, v \in S$, we have $\text{row}(w) \neq \text{row}(v)$

Distinct

	ε	ab	b
ε	-	-	+
b	+	+	+
a	-	-	-
ba	+	+	+
bb	+	+	+
aa	-	-	-
ab	-	-	-

Not Distinct

	ε	ab	b
ε	-	-	-
b	-	+	+
a	-	-	-
ba	+	+	+
bb	+	+	+
aa	-	-	-
ab	-	-	-

DFA associated with observation table

Let (S, E, T) be an observation table with respect to $L \subseteq \Sigma^*$ that is closed and distinct. Then, there exists a DFA $(Q, \Sigma, \delta, q_0, F)$ that agrees with T (Myhill–Nerode, 1957), given by:

DFA associated with observation table

Let (S, E, T) be an observation table with respect to $L \subseteq \Sigma^*$ that is closed and distinct. Then, there exists a DFA $(Q, \Sigma, \delta, q_0, F)$ that agrees with T (Myhill–Nerode, 1957), given by:

- States $Q = \text{row}(S)$;

DFA associated with observation table

Let (S, E, T) be an observation table with respect to $L \subseteq \Sigma^*$ that is closed and distinct. Then, there exists a DFA $(Q, \Sigma, \delta, q_0, F)$ that agrees with T (Myhill–Nerode, 1957), given by:

- States $Q = \text{row}(S)$;
- Transitions $\delta(\text{row}(w), a) = row(wa)$;

DFA associated with observation table

Let (S, E, T) be an observation table with respect to $L \subseteq \Sigma^*$ that is closed and distinct. Then, there exists a DFA $(Q, \Sigma, \delta, q_0, F)$ that agrees with T (Myhill–Nerode, 1957), given by:

- States $Q = \text{row}(S)$;
- Transitions $\delta(\text{row}(w), a) = row(wa)$;
- Initial state $q_0 = \text{row}(\varepsilon)$;

DFA associated with observation table

Let (S, E, T) be an observation table with respect to $L \subseteq \Sigma^*$ that is closed and distinct. Then, there exists a DFA $(Q, \Sigma, \delta, q_0, F)$ that agrees with T (Myhill–Nerode, 1957), given by:

- States $Q = \text{row}(S)$;
- Transitions $\delta(\text{row}(w), a) = row(wa)$;
- Initial state $q_0 = \text{row}(\varepsilon)$;
- Final states $F = \{\text{row}(w) \mid T(w, \varepsilon) = +\}$

DFA associated with observation table

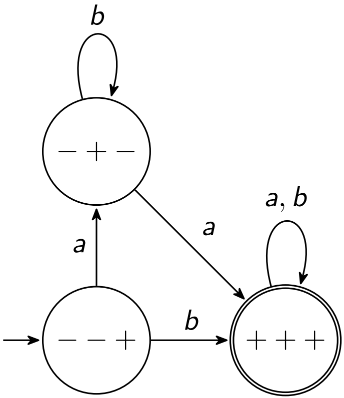
Let (S, E, T) be an observation table with respect to $L \subseteq \Sigma^*$ that is closed and distinct. Then, there exists a DFA $(Q, \Sigma, \delta, q_0, F)$ that agrees with T (Myhill–Nerode, 1957), given by:

- States $Q = \text{row}(S)$;
- Transitions $\delta(\text{row}(w), a) = row(wa)$;
- Initial state $q_0 = \text{row}(\varepsilon)$;
- Final states $F = \{\text{row}(w) \mid T(w, \varepsilon) = +\}$

We denote this DFA with $\mathcal{D}(S, E, T)$

Example

Let $S = \{\varepsilon, b, a\}$, $E = \{\varepsilon, ab, b\}$, and T as shown in the table. Then, $\mathcal{D}(S, E, T)$ is given by the following DFA:



	ε	ab	b
ε	-	-	+
b	+	+	+
a	-	+	-
ba	+	+	+
bb	+	+	+
aa	+	+	+
ab	-	+	-

L^* Learner

- Incrementally build DFA by querying MAT

High level overview:

1. Start with an empty observation table
2. Fill the observation table with membership queries
3. Expand S until the observation table is closed
4. Perform an equivalence query with $\mathcal{D}(S, E, T)$
5. Expand E with suffixes of counterexample

L^* Example

iMAT Framework

- Teacher can respond with "don't care"
  - How to build a DFA with incomplete information?
- Goal: Find a DFA that agrees with a set of positive examples L^+ and negative examples L^- from the teacher

Incomplete Observation Tables

- T is now a map $(S \cup S \times \Sigma) \times E \to \{+, -, \square\}$
- Given a table containing $\square$, can we fill it in such that it is closed and distinct?
- NP Complete (Gold, 1978)
- We will use SMT solvers

	ε	ab	b
ε	-	-	$\square$
b	$\square$	$\square$	$+$
a	-	-	$\square$
ba	$+$	$\square$	$\square$
bb	$+$	$+$	$+$
aa	-	-	-
ab	$\square$	-	-

SMT Formulas

SMT Formulas

1. Construct a table of boolean variables b_{wv} indexed by $(S \cup (S \times \Sigma)) \times E$

SMT Formulas

1. Construct a table of boolean variables b_{wv} indexed by $(S \cup (S \times \Sigma)) \times E$
2. For each w, v such that $T(w, v) \neq \square$, add a constraint

$$b_{wv} = T(w, v)$$

SMT Formulas

1. Construct a table of boolean variables b_{wv} indexed by $(S \cup (S \times \Sigma)) \times E$
2. For each w, v such that $T(w, v) \neq \square$, add a constraint

$$b_{wv} = T(w, v)$$

3. Closedness (every bottom row appears in the top)

$$\bigwedge_{w \in S \times \Sigma \setminus S} \left(\bigvee_{w' \in S} \left(\bigwedge_{v \in E} b_{wv} = b_{w'v} \right) \right)$$

SMT Formulas

1. Construct a table of boolean variables b_{wv} indexed by $(S \cup (S \times \Sigma)) \times E$
2. For each w, v such that $T(w, v) \neq \square$, add a constraint

$$b_{wv} = T(w, v)$$

3. Closedness (every bottom row appears in the top)

$$\bigwedge_{w \in S \times \Sigma \setminus S} \left(\bigvee_{w' \in S} \left(\bigwedge_{v \in E} b_{wv} = b_{w'v} \right) \right)$$

4. Distinctness (the top rows are unique)

$$\bigwedge_{w \in S} \left(\bigwedge_{w' \in S \setminus \{w\}} \left(\bigvee_{v \in E} b_{wv} \neq b_{w'v} \right) \right)$$

SMT Solvers

- Programmes like Z3 can solve these constraints, and provide a model if it exists
- SMT solvers are highly optimized and can give good performance in practice

Modifying L^*

- The goal is to find a DFA that is minimal and consistent with L^+ and L^-
- If the solver returns a model, we can construct a DFA and query it as in the L^* algorithm
- If the solver returns `unsat`, we have to do more work
- Not known which row of the bottom part of the table causes the `unsat`
- Try all rows!

$L^*_\square$ Learner

- We maintain a worklist of observation tables

1. Start with a worklist containing just an empty observation table
2. Pop the head of the worklist
3. Fill the observation table with membership queries
4. Check if the table can be closed with an SMT solver
5. Add all different expansions of S to the worklist
6. Perform an equivalence query with $\mathcal{D}(S, E, T)$
7. Extend E with suffixes of counterexample

$L^*_\square$ Example

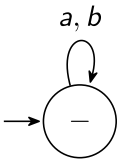
- Assume we have a teacher with

$$L^+ = \{ab, aab, bab, aaab, abab, baab, bbab\}$$
$$L^- = \{aa, ba, bb, aaa, baa, aba, bba, abb, bbb\}$$

- Worklist: $\{(S = \{\varepsilon\}, E = \{\varepsilon\})\}$

	ε
ε	$\square$
a	$\square$
b	$\square$

	ε
ε	$\boxminus$
a	$\boxminus$
b	$\boxminus$



$L^*_\square$ Example

- Receive counterexample *baab*
- Worklist: $\{(S = \{\varepsilon\}, E = \{\varepsilon, b, ab, aab, baab\})\}$

	ε	b	ab	aab	$baab$
ε	$\square$	$\square$	$+$	$+$	$+$
a	$\square$	$+$	$+$	$+$	$\square$
b	$\square$	$-$	$+$	$+$	$\square$

- This table is `unsat`
- Worklist:

$$\{(S = \{\varepsilon, a\}, E = \{\varepsilon, b, ab, aab, baab\}),$$
$$(S = \{\varepsilon, b\}, E = \{\varepsilon, b, ab, aab, baab\})\}$$

$L^*_\Box$ Example

- Pop $(S = \{\varepsilon, a\}, E = \{\varepsilon, b, ab, aab, baab\})$ from head of worklist

	ε	b	ab	aab	$baab$
ε	$\Box$	$\Box$	$+$	$+$	$+$
a	$\Box$	$+$	$+$	$+$	$\Box$
b	$\Box$	$-$	$+$	$+$	$\Box$
aa	$-$	$+$	$+$	$\Box$	$\Box$
ab	$+$	$-$	$+$	$\Box$	$\Box$

- `unsat` again
- Worklist:

$$\{(S = \{\varepsilon, b\}, E = \{\varepsilon, b, ab, aab, baab\}),$$
$$(S = \{\varepsilon, a, aa\}, E = \{\varepsilon, b, ab, aab, baab\}),$$
$$(S = \{\varepsilon, a, ab\}, E = \{\varepsilon, b, ab, aab, baab\})\}$$

$L^*_\Box$ Example

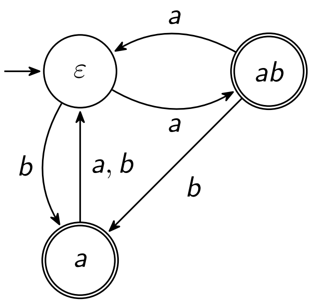
- Next two tables are again `unsat`
- Worklist:

$$
\begin{aligned}
\{&(S = \{\varepsilon, a, ab\}, E = \{\varepsilon, b, ab, aab, baab\}),\\
&(S = \{\varepsilon, b, ba\}, E = \{\varepsilon, b, ab, aab, baab\}),\\
&(S = \{\varepsilon, b, bb\}, E = \{\varepsilon, b, ab, aab, baab\}),\\
&\ldots,\\
&(S = \{\varepsilon, a, aa, aaa\}, E = \{\varepsilon, b, ab, aab, baab\}),\\
&(S = \{\varepsilon, a, aa, aab\}, E = \{\varepsilon, b, ab, aab, baab\})\}
\end{aligned}
$$

- Pop $(S = \{\varepsilon, a, ab\}, E = \{\varepsilon, b, ab, aab, baab\})$ from the worklist

$L^*_\square$ Example

	ε	b	ab	aab	$baab$
ε	$\boxed{-}$	$\boxed{+}$	$+$	$+$	$+$
a	$\boxed{+}$	$+$	$+$	$+$	$\boxed{+}$
ab	$+$	$-$	$+$	$\boxed{+}$	$\boxed{+}$
b	$\boxed{+}$	$-$	$+$	$+$	$\boxed{+}$
aa	$-$	$+$	$+$	$\boxed{+}$	$\boxed{+}$
aba	$-$	$+$	$\boxed{+}$	$\boxed{+}$	$\boxed{+}$
abb	$-$	$\boxed{+}$	$\boxed{+}$	$\boxed{+}$	$\boxed{+}$



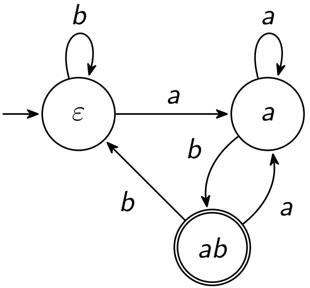
$L^*_\square$ Example

- Receive counterexample *bbb*
- New filled in table:

	ε	b	ab	aab	$baab$	bbb	bb
ε	$\boxed{-}$	$\boxed{-}$	$+$	$+$	$+$	$-$	$-$
a	$\boxed{-}$	$+$	$+$	$+$	$\boxed{+}$	$-$	$\boxed{-}$
ab	$+$	$-$	$+$	$\boxed{+}$	$\boxed{+}$	$\boxed{-}$	$\boxed{-}$
b	$\boxed{-}$	$-$	$+$	$+$	$\boxed{+}$	$-$	$\boxed{-}$
aa	$-$	$+$	$+$	$\boxed{+}$	$\boxed{+}$	$\boxed{-}$	$\boxed{-}$
aba	$-$	$+$	$\boxed{+}$	$\boxed{+}$	$\boxed{+}$	$\boxed{-}$	$\boxed{-}$
abb	$-$	$\boxed{-}$	$\boxed{+}$	$\boxed{+}$	$\boxed{+}$	$\boxed{-}$	$\boxed{-}$

$L^*_\square$ Example

- Query new DFA
- Agrees with L^+ and L^-



$$L^+ = \{ab, aab, bab, aaab, abab, baab, bbab\}$$
$$L^- = \{aa, ba, bb, aaa, baa, aba, bba, abb, bbb\}$$

Correctness

$L^*_\square$ returns the smallest DFA that agrees with L^+ and L^-, which can be shown with three main lemmas:

Correctness

$L^*_\square$ returns the smallest DFA that agrees with L^+ and L^-, which can be shown with three main lemmas:

1. The sizes of S of the tables in the worklist increase monotonically by at most 1

Correctness

$L^*_\square$ returns the smallest DFA that agrees with L^+ and L^-, which can be shown with three main lemmas:

1. The sizes of S of the tables in the worklist increase monotonically by at most 1
2. The worklist always contains at least one table that is compatible with a correct minimal DFA

Correctness

$L^*_\square$ returns the smallest DFA that agrees with L^+ and L^-, which can be shown with three main lemmas:

1. The sizes of S of the tables in the worklist increase monotonically by at most 1
2. The worklist always contains at least one table that is compatible with a correct minimal DFA
3. If a table (S, E, T) is compatible with a smallest correct DFA with states Q and $|S| = |Q|$, then (S, E, T) can be filled in to be closed and distinct

Second Paper

- Learning Minimal Deterministic Automata from Inexperienced Teachers
- Summary of research on "inexperienced" teachers and SAT/SMT approaches
- Encode DFA directly into the SMT formulas

Alternative Approach

- Different notion of "closedness" and "distinctness"
- Use SMT minimization to find DFA consistent with current table

Given a table (S, E, T), two rows $\mathrm{row}(w), \mathrm{row}(w')$ look similar, denoted $\mathrm{row}(w) \equiv \mathrm{row}(w')$, if the blanks can be filled in so the rows are the same

- $(+, \Box, -) \equiv (+, -, -)$
- $(+, \Box, -) \not\equiv (-, \Box, \Box)$

Weak closedness

An observation table (S, E, T) is weakly closed if every bottom row looks similar to a top row

Weakly closed

	ε	ab	b
ε	-	-	$\square$
b	$\square$	$\square$	+
a	-	-	$\square$
ba	+	$\square$	$\square$
bb	+	+	+
aa	-	-	-
ab	$\square$	-	-

Not weakly Closed

	ε	ab	b
ε	-	-	$\square$
b	$\square$	$\square$	+
a	-	-	$\square$
ba	+	$\square$	$\square$
bb	+	+	+
aa	-	-	-
ab	$\square$	+	-

Weak consistency

A table (S, E, T) is weakly consistent if for every pair of words $w, w' \in S$ and letter $a \in \Sigma$ such that $\text{row}(w) \equiv \text{row}(w')$, we have $\text{row}(wa) \equiv \text{row}(w'a)$

- Distinctness implies consistency
- We use SMT solvers to find the smallest DFA that is consistent with a weakly closed and weakly consistent table

Biermann and Feldman

Let S_w be the state that is reached after reading the word w. To determine the DFA, we have to solve the following constraints:

1. If two words lead to the same state, then any next step must lead to the same state
2. If two words have a different acceptance, they must not lead to the same state

SMT Formulas

SMT Formulas

- Let n be the number of states. Define a table of boolean variables $b_{w,i}$, indexed by $W \times \{1, 2, \ldots, n\}$ with $W = (S \cup (S \times \Sigma)) \times E$

SMT Formulas

- Let n be the number of states. Define a table of boolean variables $b_{w,i}$, indexed by $W \times \{1, 2, \ldots, n\}$ with $W = (S \cup (S \times \Sigma)) \times E$
- Add constraints to ensure exactly one of $b_{w,i}$ is true for fixed w and each i.

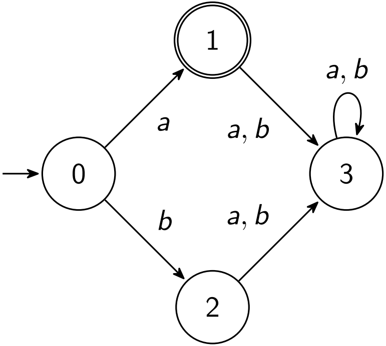
$$\bigwedge_{w \in W} \left(\bigvee_{1 \leq i \leq n} b_{w,i} \right)$$

$$\bigwedge_{w \in W} \left(\bigvee_{1 \leq i < i' \leq n} \neg b_{w,i} \vee \neg b_{w,i'} \right)$$

More SMT Formulas

1. If two words lead to the same state, then any next step must lead to the same state

$$b_{ab,3} = b_{ba,3} \rightarrow$$
$$\bigwedge_{0 \le i \le 3} (b_{aba,i} = b_{baa,i} \wedge b_{abb,i} = b_{bab,i})$$



More SMT Formulas

2. If two words have a different acceptance, they must not lead to the same state

- If $a \in L^+$ and $b \in L^-$, then

$$\bigwedge_{0 \le i < n} \neg\left(b_{a,i} \wedge b_{b,i}\right)$$

Heule and Verwer

- Additionally include transitions $d_{i,w,j}$ and final states f_i in encoding
- Reached states b have to conform to transitions d
- More variables, but sometimes faster in practice

Finalizing

- Use binary search to find smallest n that is satisfiable
- Upper bound of n is $|W|$
- Model directly gives a DFA, which we can query

Comparison to First Paper

- More variables (factor n)
- Harder to study and implement
- No implementation given
- No benchmarking of efficiency
- Comparison of efficiency is not given in the first paper

Practicality

DFA size	Mean learn time (s)	Mean worklist items
5	0.1237	8.6200
6	0.3803	13.6381
7	1.1251	20.0886
8	10.6307	44.1613
9	50.1672	96.8784
10	98.0573	176.5200
11	1498.4836	933.2857

Future Work

- Adapt work to modern algorithms
  - TTT, ADT
  - Discrimination trees
- Study incomplete teachers for more general automata
  - Mealy machines
  - Moore machines
  - Weighted automata

Summary

- Defined automata learning using the MAT framework and looked at the L^* algorithm
- Generalised the MAT framework using an incomplete teacher and studied that $L^*_\square$ algorithm
- Discuss older methods of learning with incomplete teachers

End

- Thank you for your attention!
- Any questions?