Learning Computation Tree Logic A SAT-based approach

Adrien Pommellet, Daniel Stan, Simon Scatton, LRE



January 16, 2025



What? Computing a theoretical model compatible with a labelled dataset (e.g. desirable and undesirable behaviours).

¹Daniel Neider and Ivan Gavran. "Learning Linear Temporal Properties". In: 2018 Formal Methods in Computer Aided Design, FMCAD 2018, Austin, TX, USA, October 30 - November 2, 2018. Ed. by Nikolaj S. Bjørner and Arie Gurfinkel. IEEE, 2018, pp. 1–10. DOI: 10.23919/FMCAD.2018.8603016. URL: https://doi.org/10.23919/FMCAD.2018.8603016.



- What? Computing a theoretical model compatible with a labelled dataset (e.g. desirable and undesirable behaviours).
 - Why? Explainability, extrapolate tests, system design.

¹Daniel Neider and Ivan Gavran. "Learning Linear Temporal Properties". In: 2018 Formal Methods in Computer Aided Design, FMCAD 2018, Austin, TX, USA, October 30 - November 2, 2018. Ed. by Nikolaj S. Bjørner and Arie Gurfinkel. IEEE, 2018, pp. 1–10. DOI: 10.23919/FMCAD.2018.8603016. URL: https://doi.org/10.23919/FMCAD.2018.8603016.



- What? Computing a theoretical model compatible with a labelled dataset (e.g. desirable and undesirable behaviours).
 - Why? Explainability, extrapolate tests, system design.
 - How? Automata, logical formulas¹.

¹Daniel Neider and Ivan Gavran. "Learning Linear Temporal Properties". In: 2018 Formal Methods in Computer Aided Design, FMCAD 2018, Austin, TX, USA, October 30 - November 2, 2018. Ed. by Nikolaj S. Bjørner and Arie Gurfinkel. IEEE, 2018, pp. 1–10. DOI: 10.23919/FMCAD.2018.8603016. URL: https://doi.org/10.23919/FMCAD.2018.8603016.









Kripke structures and execution trees



Kripke structures and execution trees



Universal and existential quantifiers

$\forall a \cup b$



Universal and existential quantifiers



Incomparable logics

$$\forall \mathsf{X} \left[(\exists \mathsf{X} \ b) \land (\exists \mathsf{X} \ c) \right]$$

Incomparable logics

$$\forall \mathsf{X} \left[(\exists \mathsf{X} \ b) \land (\exists \mathsf{X} \ c) \right]$$



Incomparable logics

$$\forall \mathsf{X} \left[(\exists \mathsf{X} \ b) \land (\exists \mathsf{X} \ c) \right]$$



The Learning Problem Sample of structures

A sample made of **dissimilar** positive and negative structures.



Formalizing various problems

The learning problem. Find a CTL formula φ verified by the positive sample and unanimously rejected by the negative sample.

Formalizing various problems

The learning problem. Find a CTL formula φ verified by the positive sample and unanimously rejected by the negative sample.

 ${\sf L}_{\sf CTL}.$ Any compatible formula φ will do.

Formalizing various problems

The learning problem. Find a CTL formula φ verified by the positive sample and unanimously rejected by the negative sample.

L_{CTL}. Any compatible formula φ will do. L_{CTL}^{$\leq n$}. Formula φ must be of size $\leq n$.

Formalizing various problems

The learning problem. Find a CTL formula φ verified by the positive sample and unanimously rejected by the negative sample.

L_{CTL}. Any compatible formula
$$\varphi$$
 will do.
L ^{$\leq n$} _{CTL}. Formula φ must be of size $\leq n$.
ML_{CTL}. Formula φ must be of minimal size.

Formalizing various problems

Theorem

L_{CTL} (thus ML_{CTL}) always admits a solution.

²M.C. Browne, E.M. Clarke, and O. Grümberg. "Characterizing finite Kripke structures in propositional temporal logic". In: *Theoretical Computer Science* 59.1 (1988), pp. 115–131. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(88)90098-9. URL: https://www.sciencedirect.com/science/article/pii/0304397588900989. Pommellet, Stan, Scatton (EPITA) January 16, 2025

9/33

Formalizing various problems

Theorem

*L*_{CTL} (thus *ML*_{CTL}) always admits a solution.

Proof. Consider a CTL formula that encodes the dissimilarity of the positive and the negative states.

²M.C. Browne, E.M. Clarke, and O. Grümberg. "Characterizing finite Kripke structures in propositional temporal logic". In: *Theoretical Computer Science* 59.1 (1988), pp. 115–131. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(88)90098-9. URL: https://www.sciencedirect.com/science/article/pii/0304397588900989. Pommellet, Stan, Scatton (EPITA)

9/33

Formalizing various problems

Theorem

*L*_{CTL} (thus *ML*_{CTL}) always admits a solution.

Proof. Consider a CTL formula that encodes the dissimilarity of the positive and the negative states.

This explicit answer is however of size $\mathcal{O}(|S^+| \cdot |S^-| \cdot k^c)$ where k is the degree of \mathcal{K} and c its characteristic number².

²M.C. Browne, E.M. Clarke, and O. Grümberg. "Characterizing finite Kripke structures in propositional temporal logic". In: *Theoretical Computer Science* 59.1 (1988), pp. 115-131. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(88)90098-9. URL: https://www.sciencedirect.com/science/article/pii/0304397588900989. Pommellet, Stan, Scatton (EPITA)

9/33

Bordais et al³ proved that:

Theorem

 $L_{CTL}^{\leq n}$ is NP-complete (assuming an unary encoding of n).

³Benjamin Bordais, Daniel Neider, and Rajarshi Roy. *Learning Temporal Properties is NP-hard*. 2023. arXiv: 2312.11403 [cs.L0].

Bordais et al³ proved that:

Theorem

 $L_{CTL}^{\leq n}$ is NP-complete (assuming an unary encoding of n).

We will now design a SAT instance Φ_n equivalent to $L_{CTL}^{\leq n}$.

³Benjamin Bordais, Daniel Neider, and Rajarshi Roy. *Learning Temporal Properties is NP-hard*. 2023. arXiv: 2312.11403 [cs.L0].

Solving the minimal learning problem

```
Input: a KS \mathcal{K} and a sample S.

Output: the smallest CTL formula \varphi consistent with S.

n \leftarrow 0;

repeat

\left|\begin{array}{c} n \leftarrow n+1; \\ \text{compute } \Phi_n; \end{array}\right|

until \Phi_n is satisfiable by some valuation v;

from v build and return \varphi.
```

Algorithm 1: Solving $ML_{CTL}(\mathcal{K}, S)$.

Representing CTL formulas

$\neg a \land \forall X a$

Representing CTL formulas

$$\neg a \land \forall X a$$



Representing CTL formulas





Representing CTL formulas





A binary syntactic DAG



Encoding the syntactic tree

Variables. • Node *i* has label *o*.

 τ_i^o

Encoding the syntactic tree

Variables.

Node *i* has label *o*. τ^o_i
 The left (resp. right) children of *i* is *j*. l_{i,i}, r_{i,i}

Encoding the syntactic tree

Variables.

Node *i* has label *o*.
The left (resp. right) children of *i* is *j*.

 au_i^o $I_{i,j}, \mathbf{r}_{i,j}$

Clauses. • Each node has a single label.

Encoding the syntactic tree

Variables.

• Node *i* has label *o*. $I_{i,i}, r_{i,i}$ • The left (resp. right) children of *i* is *j*.

Clauses. Each node has a single label.

• Each node has at most one left (resp. right) child.

 τ_i^o

Encoding the syntactic tree

Variables.

Node *i* has label *o*. τ^o_i
 The left (resp. right) children of *i* is *j*. l_{i,i}, r_{i,i}

Clauses.

- Each node has a single label.
 - Each node has at most one left (resp. right) child.
 - Binary (resp. unary) operators have 2 (resp. 1) children.
Encoding the semantics

Variables. • State q verifies sub-formula φ_i rooted in node i.

Encoding the semantics

- Variables. State q verifies sub-formula φ_i rooted in node i.
 - Clauses. Each state in the positive sample verifies φ .

Encoding the semantics

Variables. • State q verifies sub-formula φ_i rooted in node i.

Clauses. • Each state in the positive sample verifies φ .

• Each state in the negative sample does not verify φ .

Encoding the semantics

Variables. • State q verifies sub-formula φ_i rooted in node i.

Clauses. • Each state in the positive sample verifies φ .

- Each state in the negative sample does not verify φ .
- If state q verifies φ_i, define φ_i's semantics according to the label of node i.

Encoding $\forall X$'s semantics

$$(\tau_i^{\forall \mathsf{X}} \wedge \mathsf{r}_{i,j})$$

If node *i* is labelled by operator $\forall X$ and has right child *j*

Pommellet, Stan, Scatton (EPITA)

January 16, 2025 16 / 33

Encoding $\forall X$'s semantics

$$(au_i^{orall \mathsf{X}} \wedge \mathsf{r}_{i,j}) \implies \bigwedge_{q \in Q}$$

If node *i* is labelled by operator $\forall X$ and has right child *j* then for every state *q*,

Encoding $\forall X$'s semantics

$$(au_i^{orall \mathsf{X}} \wedge \mathsf{r}_{i,j}) \implies \bigwedge_{q \in Q} (\varphi_i^q)$$

If node *i* is labelled by operator $\forall X$ and has right child *j* then for every state *q*, if state *q* verifies formula us

if state q verifies formula φ_i ,

Encoding $\forall X$'s semantics

$$(au_i^{\forall \mathsf{X}} \wedge \mathsf{r}_{i,j}) \Longrightarrow \bigwedge_{q \in Q} (\varphi_i^q \iff \bigwedge_{q' \in \delta(q)} \varphi_j^{q'})$$

If node *i* is labelled by operator $\forall X$ and has right child *j* then for every state *q*, if state *q* verifies formula φ_i ,

then every successor q' of q must verify φ_j .

What about temporal operators?

$(q \models \forall \mathsf{G} a) \iff (q \models a) \land \bigwedge_{q' \in \delta(q)} (q' \models \forall \mathsf{G} a)$

What about temporal operators?

$$(\tau_2^{\forall \mathsf{G}} \land \mathsf{r}_{2,1}) \implies \bigwedge_{q \in Q} \left(\varphi_2^q \iff \left[\varphi_1^q \land \bigwedge_{q' \in \delta(q)} \varphi_2^{q'} \right] \right)$$



What about temporal operators?

$$(\tau_{2}^{\forall \mathsf{G}} \land \mathsf{r}_{2,1}) \Longrightarrow \bigwedge_{q \in Q} \left(\varphi_{2}^{q} \iff \left[\varphi_{1}^{q} \land \bigwedge_{q' \in \delta(q)} \varphi_{2}^{q'} \right] \right)$$



How can we encode CTL's semantics?

Another form of model-checking

Bounded model-checking (BMC): a CTL property holds for all computation trees of depth k.



An upper bound for finite systems

Recurrence diameter $\alpha(q)$: the length $(\leq |Q|)$ of the longest non-repeating sequence of states starting from q.



Equivalence with model-checking

Model-checking on Kripke structures is equivalent to BMC up to the system's recurrence diameter.

Defining bounded semantics

Base case. $(q \models_{\mathcal{K}} \forall \mathsf{F}^0 \varphi) \iff (q \models_{\mathcal{K}} \varphi)$

Defining bounded semantics

Base case.
$$(q \models_{\mathcal{K}} \forall \mathsf{F}^{0} \varphi) \iff (q \models_{\mathcal{K}} \varphi)$$

Inductive case. $(q \models_{\mathcal{K}} \forall \mathsf{F}^{u+1} \varphi) \iff (q \models_{\mathcal{K}} \varphi) \lor (\bigwedge_{q' \in \delta(q)} q' \models_{\mathcal{K}} \forall \mathsf{F}^{u} \varphi)$

Defining bounded semantics

Base case.
$$(q \models_{\mathcal{K}} \forall \mathsf{F}^{0} \varphi) \iff (q \models_{\mathcal{K}} \varphi)$$

Inductive case. $(q \models_{\mathcal{K}} \forall \mathsf{F}^{u+1} \varphi) \iff (q \models_{\mathcal{K}} \varphi) \lor (\bigwedge_{q' \in \delta(q)} q' \models_{\mathcal{K}} \forall \mathsf{F}^{u} \varphi)$
Equivalence with MC. $(q \models_{\mathcal{K}} \forall \mathsf{F} \varphi) \iff (q \models_{\mathcal{K}} \forall \mathsf{F}^{\alpha(q)} \varphi)$

New variables and clauses for $\forall \mathsf{F}$

New variables and clauses for $\forall \mathsf{F}$

Base case.
$$\rho_{i,q}^{0} \iff \varphi_{j}^{q}$$

New variables and clauses for $\forall \mathsf{F}$

Base case.
$$\rho_{i,q}^{0} \iff \varphi_{j}^{q}$$

Inductive case. $\rho_{i,q}^{u+1} \iff \left(\varphi_{j}^{q} \lor \bigwedge_{q' \in \delta(q)} \rho_{i,q'}^{u}\right)$

New variables and clauses for $\forall \mathsf{F}$

Base case.
$$\rho_{i,q}^{0} \iff \varphi_{j}^{q}$$

Inductive case. $\rho_{i,q}^{u+1} \iff \left(\varphi_{j}^{q} \lor \bigwedge_{q' \in \delta(q)} \rho_{i,q'}^{u}\right)$
Equivalence with MC. $\tau_{i}^{\forall \mathsf{F}} \implies \left(\varphi_{i}^{q} \iff \rho_{i,q}^{\alpha(q)}\right)$

Estimating the instance's complexity

Variables.
$$\mathcal{O}(n^2 + n \cdot |AP| + n \cdot |Q| \cdot d)$$

Clauses. $\mathcal{O}(n \cdot |AP| + n^3 \cdot |Q| \cdot d + n \cdot |AP| \cdot |Q|)$

Approximating the recurrence diameter



$\alpha(q_0) = 3$	$\alpha(\alpha) = 1$
$\alpha(q_1) = 2$	$\alpha(q_4) = 1$
$\alpha(q_2) = 1$	$\alpha(q_5) = 2$
$\alpha(q_3) = 3$	$\alpha(q_6) \equiv 0$

Approximating the recurrence diameter



$\alpha(q_0) = 3$	
$\alpha(q_1) = 2$	
$\alpha(q_2) = 1$	
$\alpha(q_3) = 3$	

 $egin{aligned} &lpha(q_4)=1 \ &lpha(q_5)=2 \ &lpha(q_6)=0 \end{aligned}$



Approximating the recurrence diameter



Embedding negations



Embedding negations



Testing various fragments

$\mathsf{CTL} = \ \{\neg, \land, \lor, \forall X, \exists X, \forall \mathsf{F}, \exists \mathsf{F}, \forall \mathsf{G}, \exists \mathsf{G}, \forall \mathsf{U}, \exists \mathsf{U}\}.$

Testing various fragments

$$\begin{split} & \mathsf{CTL} = \; \{\neg, \land, \lor, \forall \mathsf{X}, \exists \mathsf{X}, \forall \mathsf{F}, \exists \mathsf{F}, \forall \mathsf{G}, \exists \mathsf{G}, \forall \mathsf{U}, \exists \mathsf{U}\}.\\ & \mathsf{CTL}_\forall = \; \{\neg, \land, \lor, \forall \mathsf{X}, \forall \mathsf{F}, \forall \mathsf{G}, \forall \mathsf{U}\}. \end{split}$$

Testing various fragments

$$\begin{split} & \mathsf{CTL} = \ \{\neg, \land, \lor, \forall \mathsf{X}, \exists \mathsf{X}, \forall \mathsf{F}, \exists \mathsf{F}, \forall \mathsf{G}, \exists \mathsf{G}, \forall \mathsf{U}, \exists \mathsf{U}\}.\\ & \mathsf{CTL}_{\forall} = \ \{\neg, \land, \lor, \forall \mathsf{X}, \forall \mathsf{F}, \forall \mathsf{G}, \forall \mathsf{U}\}.\\ & \mathsf{CTL}_{\mathsf{U}} = \ \{\neg, \lor, \exists \mathsf{X}, \exists \mathsf{G}, \exists \mathsf{U}\}. \end{split}$$

Testing various fragments

$$\begin{split} \mathsf{CTL} &= \ \{\neg, \land, \lor, \forall \mathsf{X}, \exists \mathsf{X}, \forall \mathsf{F}, \exists \mathsf{F}, \forall \mathsf{G}, \exists \mathsf{G}, \forall \mathsf{U}, \exists \mathsf{U}\}.\\ \mathsf{CTL}_{\forall} &= \ \{\neg, \land, \lor, \forall \mathsf{X}, \forall \mathsf{F}, \forall \mathsf{G}, \forall \mathsf{U}\}.\\ \mathsf{CTL}_{\mathsf{U}} &= \ \{\neg, \lor, \exists \mathsf{X}, \exists \mathsf{G}, \exists \mathsf{U}\}. \end{split}$$

Succinctness vs Number of clauses and variables

A C++ implementation using Z3

	-	eta	β, \neg
CTL_\forall	50 46870	14 6493	4 2271
CTL	50 42658	8 5357	5 3370
CTL_U	46 31975	28 5064	4 1987

Timeouts (over 234 samples) | Arithmetic mean (ms)

Topology-guided solving. Parallel enumeration of DAG shapes⁴.

⁴Heinz Riener. "Exact Synthesis of LTL Properties from Traces". In: 2019 Forum for Specification and Design Languages (FDL). 2019, pp. 1–6. DOI: 10.1109/FDL.2019.8876900.

⁵Nathanaël Fijalkow and Guillaume Lagarde. "The complexity of learning linear temporal formulas from examples". In: *Proceedings of the 15th International Conference on Grammatical Inference, 23-27 August 2021, Virtual Event*. Ed. by Jane Chandlee et al. Vol. 153. Proceedings of Machine Learning Research. PMLR, 2021, pp. 237–250. URL: https://proceedings.mlr.press/v153/fijalkow21a.html.

Topology-guided solving. Parallel enumeration of DAG shapes⁴. Hardness bounds. What about the minimal learning problem⁵?

⁴Heinz Riener. "Exact Synthesis of LTL Properties from Traces". In: 2019 Forum for Specification and Design Languages (FDL). 2019, pp. 1–6. DOI: 10.1109/FDL.2019.8876900.

⁵Nathanaël Fijalkow and Guillaume Lagarde. "The complexity of learning linear temporal formulas from examples". In: *Proceedings of the 15th International Conference on Grammatical Inference, 23-27 August 2021, Virtual Event*. Ed. by Jane Chandlee et al. Vol. 153. Proceedings of Machine Learning Research. PMLR, 2021, pp. 237–250. URL: https://proceedings.mlr.press/v153/fijalkow21a.html.

Topology-guided solving. Parallel enumeration of DAG shapes⁴. Hardness bounds. What about the minimal learning problem⁵? State-of-the-art SAT solving. Looking at the SAT competition.

⁴Heinz Riener. "Exact Synthesis of LTL Properties from Traces". In: 2019 Forum for Specification and Design Languages (FDL). 2019, pp. 1–6. DOI: 10.1109/FDL.2019.8876900.

⁵Nathanaël Fijalkow and Guillaume Lagarde. "The complexity of learning linear temporal formulas from examples". In: *Proceedings of the 15th International Conference on Grammatical Inference, 23-27 August 2021, Virtual Event*. Ed. by Jane Chandlee et al. Vol. 153. Proceedings of Machine Learning Research. PMLR, 2021, pp. 237–250. URL: https://proceedings.mlr.press/v153/fijalkow21a.html.


Thank you!

An Example

Working on an exclusion protocol

```
bool turn, flag [2];
// t = "turn"; m = "ncrit>1"; c = "ncrit>0"
byte ncrit;
active [2] proctype user()
{
        assert (_pid == 0 || _pid == 1);
        again:
        flag[_pid] = 1;
        turn = _pid;
        (flag[1 - _pid] = 0 || turn = 1 - _pid);
        ncrit++;
        assert(ncrit == 1); /* critical section */
        ncrit ---:
        flag[_pid] = 0;
        goto again
```

An Example

Introducing mutations

```
bool turn, flag [2];
// t = "turn"; m = "ncrit>1"; c = "ncrit>0"
byte ncrit;
active [2] proctype user()
{
        assert(_pid == 0 || _pid == 1);
        again:
        flag[_pid] = 1; m1
        turn = \_pid; m2
        (flag[1 - _pid] = 0 || turn = 1 - _pid); m3
        ncrit++; m4
        assert(ncrit == 1); /* critical section */
        ncrit ---:
        flag[_pid] = 0; m5
        goto again m6
```

An Example

Characterizing errors

