Learning LTL$_f$ formulas from examples is hard

Guillaume Lagarde, IRIF 12/04/24

joint work with

Nathanaël Fijalkow Corto Mascle

LTL ? Linear Temporal Logic

$$\varphi : \underset{\Sigma}{\overset{\in}{a}} \;|\; F\varphi \;|\; X\varphi \;|\; \varphi U \varphi \;|\; G\varphi \;|\; \varphi \vee \varphi \;|\; \varphi \wedge \varphi$$

$F\varphi$  eventually

$G\varphi$  always

$X\varphi$  next

$\varphi U \psi$  until

Formalism to talk about traces $\approx$ words

Example $\varphi = F(a \wedge X^5(b \wedge F X^3 c))$

$w_1 = b\,a\,c\,a\,\underline{a}\,a\,b\,c\,a\,\underline{b}\,a\,a\,a\,a\,\underline{c}\,b\,a\,c$

$w_2 = b\,a\,c\,a\,a\,a\,b\,c\,a\,b\,a\,a\,c\,a\,a\,b\,a\,c$

$w_1 \models \varphi$ $\qquad$ $w_2 \not\models \varphi$

LTL LEARNING

Input $\begin{cases} u_1, \ldots, u_n \in \Sigma^* \text{ (positive)} \\ v_1, \ldots, v_m \in \Sigma^* \text{ (negative)} \\ k \in \mathbb{N} \end{cases}$

Output $\begin{cases} \exists ?\, \varphi \in LTL,\ |\varphi| \leq k \\ \text{s.t.} \quad \forall i \quad u_i \models \varphi \\ \qquad\quad \forall i \quad v_i \not\models \varphi \end{cases}$

↳ φ is a separating formula

LTL learning appeared more in applied studies:

- Software engineering
- cyber physical systems
- A.I

Related Work

Theorem (Pitt, Warmuth 93)

$\forall$ poly, unless P = NP, there is no PTIME algorithm constructing a separating DFA $\mathcal{A}$ s.t.

$$|\mathcal{A}| \leq \text{poly}(\text{min. sep. DFA})$$

↳ Does not apply to LTC

LTL ≡ First order, but more succint

Kamp's theorem

Motivation

EXPLANABLE MODEL

LTL formulas are "nice" to read.

Our work

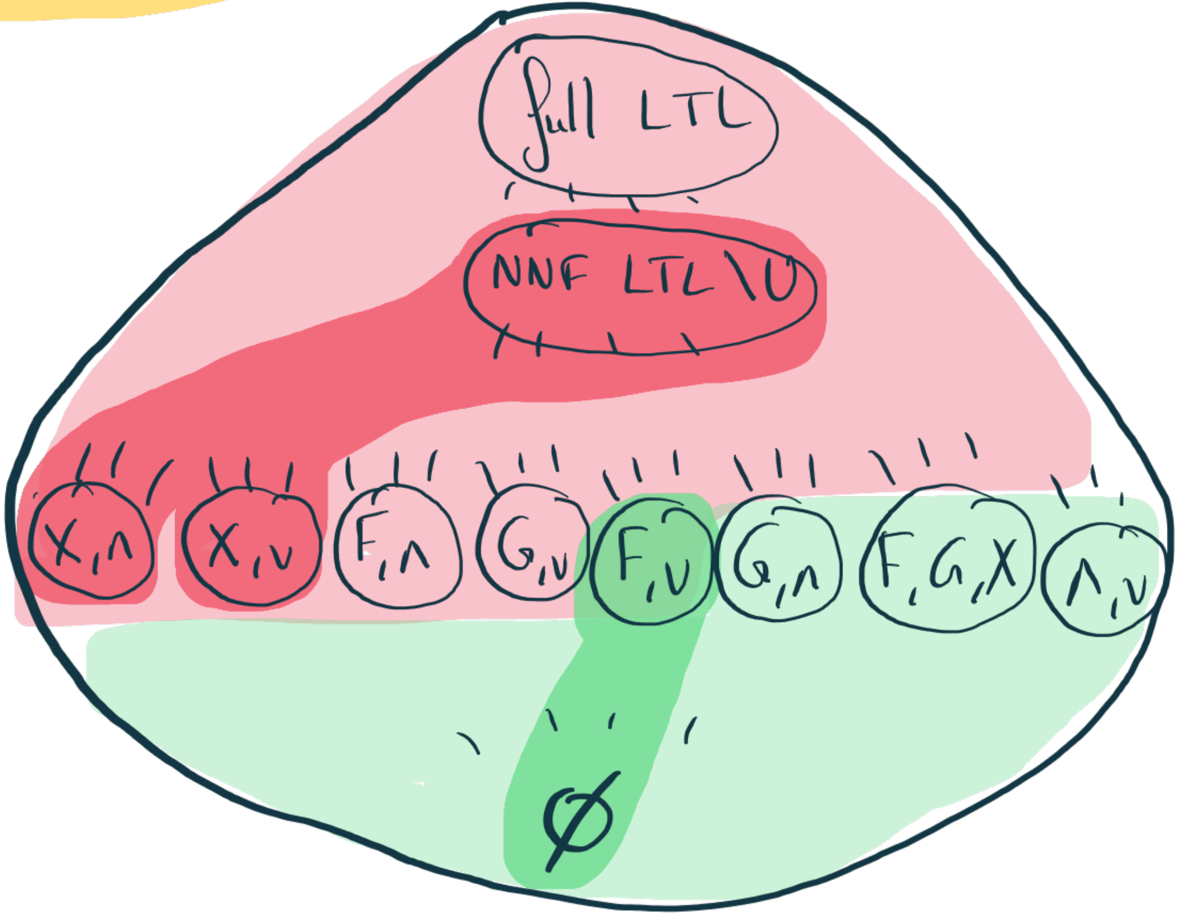
→ initiate the study of the complexity of this problem

→ NP-hardness results

+

Hardness of approximation

Lattice



strong ⇓ weak

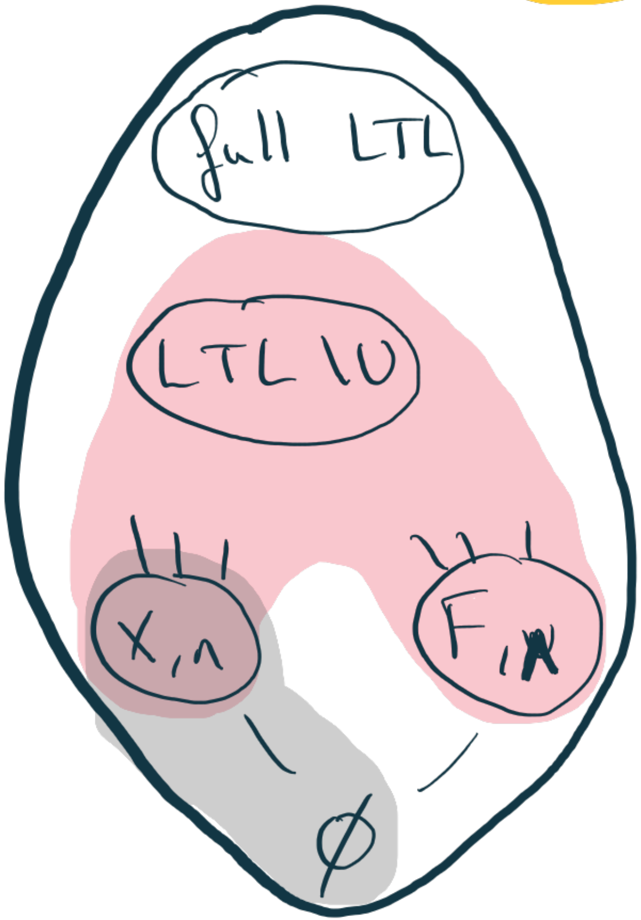
- NP-hard with fixed alphabet
- NP-hard

strong ⇓ weak

- ∈ P
- ∈ P with fixed alphabet

OPEN: Full LTL with fixed alphabet?

Approximations



hard to $\Theta(\log n)$-approx in polytime with fixed alphabet, unless P = NP

log n - approx in poly time

- (F, ∧) log n - approx ?
- beyond LTL \ U ?

Strategy

$\{X, n\}$ NP-hard

$$\Downarrow$$

$$\{X, n\} \leq OP \leq \{F, G, X, n, v\}$$

with instances where F, G, v useless

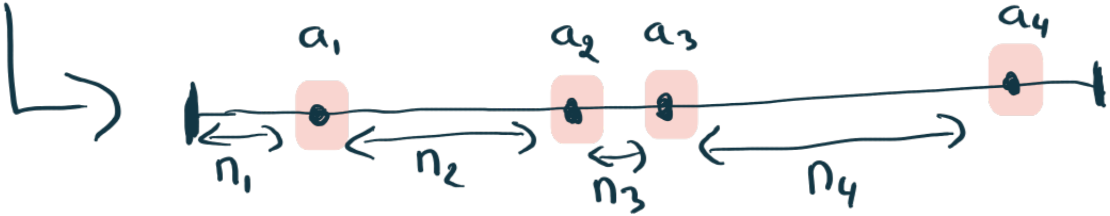
Warm-up

$(X, \wedge)$

- NP-hard
- $\exists$ polytime $\log n$-approx
- $\nexists$ polytime $(1-o(1))$-$\log n$ approx

Formulas in $LTL(X, \wedge)$ of min size:

$$X^{n_1}(a_1 \wedge X^{n_2}(a_2 \wedge X^{n_3}(a_3 \wedge \cdots\cdots) \cdots)$$



SUBSET-COVER

Input

$U = [1, n]$ univers

$\{S_i\}_{1 \le i \le m} \subseteq 2^U$

Output

$\min \mathcal{F} \subset \{S_i\}$

s.t $\bigcup_{S \in \mathcal{F}} S = U$

log n - approx using a greedy algorithm

Theorem (Dinur, Steurer '14)

NP-hard to do a $o(\log n)$-approx

$|\Sigma| = 2$

Example

$S_1 = \{1, 4\}$ $\quad$ $S_3 = \{2, 4\}$

$S_2 = \{1, 2\}$ $\quad$ $S_4 = \{1, 3\}$

Univers $= \{1, 2, 3, 4\}$

A separating φ

$x^3(a \wedge X^1 a)$

S_3 and $S_{4=3+1}$

		S_1	S_2	S_3	S_4	S_∞ ← little hack
	$v =$	a	a	a	a	a
1	$v_1 =$	b	b	a	b	a
2	$v_2 =$	a	b	b	a	a
3	$v_3 =$	a	a	a	b	a
4	$v_4 =$	b	a	b	a	a
	$v_\infty =$	a	a	a	a	b

↑ little hack

Extension

$$(X, \wedge) \subseteq Op \subseteq (F, G, X, \wedge, \vee)$$

$\exists!$ positive word

Idea $\cdot$ $LTL(OP) \leq^{!} LTL(X, \wedge)$ (over the same alphabet)

$\cdot$ Instances where $F, G, \vee$ are useless

Observation

$\vee$ is uless for free since $\exists!$ positive word.

PROOF'S IDEA

$LTL^{!}(O_p) \leq LTL(X, \wedge)$

* $u, v_1, \ldots, v_n \in \Sigma$
* M upper bound of min size of a separating formula.

$$U = \left(u a^M v_1 a^M v_2 a^M \ldots\right)^{M+1}$$

$$V_i = \left(v_i a^M v_{i+1} a^M \ldots v_n a^M\right)\left(u a^M v_1 a^M \ldots\right)^M$$

Lemmas

* if $|\varphi| \leq N$ then $w_1 w_2^N \models \varphi \iff w_1 w_2^{N+1} \models \varphi$
* $\varphi \in LTL(X, \wedge, \vee)$, φ has at most N operators X $\Rightarrow \left(w_1 \models \varphi \iff w_1^{\leq N} \models \varphi\right)$

LTL$(F, \wedge)$

$|\Sigma| = 2$ $\Sigma = \{a, b\}$

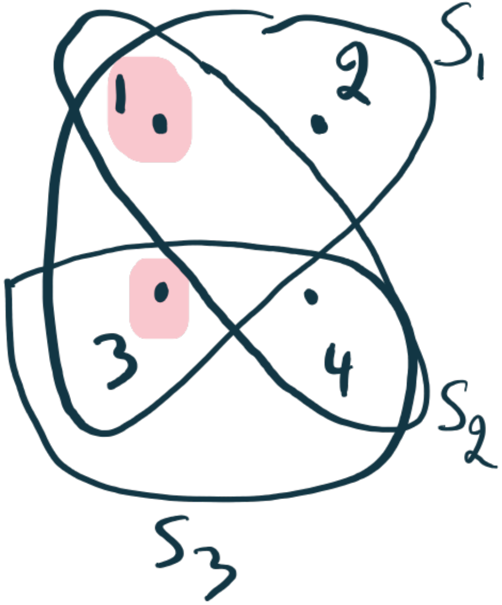
$\varphi = F(a \wedge F(b \wedge F(a \wedge \dots) \dots))$
or $= F(b \wedge F(a \wedge F(b \wedge \dots) \dots))$

$\rightarrow$ Talks about prefixes of ababa... / babab...

$\rightarrow$ $\in P$

Σ is part of the input $(|\Sigma| = |\text{Universl}|)$

Hitting set



$U = 0 \; 1 \; 2 \; 3 \; 4$

$(\overline{S_1}) \quad \vartheta_1 = 0 \; 4$

$(\overline{S_2}) \quad \vartheta_2 = 0 \; 2 \; 3$

$(\overline{S_3}) \quad \vartheta_3 = 0 \; 1 \; 2$

$$\varphi = F(1 \wedge F(3))$$

Most technical part : $|\Sigma| = 3$

$$(F, \wedge) \subseteq Op \subseteq (F, G, \wedge, \vee, \neg)$$

large enough

IDEA

$$u = \left((ab)^{M+1} c\right)^{m}$$ number of elements

$$v_i = c\, w_1^i\, c\, w_2^i\, c \cdots c\, w_m^i\, c$$

$$= \begin{cases} (ab)^{M} & \text{if } j \in S_i \\ (ab)^{M+1} & \text{otherwise} \end{cases}$$

$$u = \left(\left(ab\right)^{M+1} c\right)^m$$

$$v_i = c\, w_1^i\, c\, w_2^i\, c \cdots c\, w_m^i\, c$$

$$w_j^i = \begin{cases} (ab)^M & \text{if } j \in S_i \\ (ab)^{M+1} & \text{otherwise} \end{cases}$$

H hitting set

$$z_i = \begin{cases} (ab)^{M+1} & \text{if } i \in H \\ ab & \text{otherwise} \end{cases}$$

$$w = c \cdot z_1 \cdot c \cdot z_2 \cdots\cdots = x_1 \cdot x_2 \cdots\cdots$$

$$\varphi = F(x_1 \wedge F(x_2 \wedge F(x_3 \wedge \cdots) \cdots))$$

PROOF : SOMEWHAT LOOKS LIKE EF GAMES

(Ehrenfeucht - Fraïssé)

Duplicator → wants to prove $\{v_i\}$ v.s $\{v_i'\}$ are "equivalent" w.r.t small formulas

Spoiler → wants to show there are $\neq$.

- S plays a move from the syntax
- D points to position in the words

$\{(u_i, v_i)\}_{i\in I}, p \in \mathbb{N}$

↑ budget for S

Exemple

WANTED

$\exists\, \varphi$ small $\Longrightarrow$ Spoiler wins

$\neg(\exists\, \varphi$ small$)$ $\not\Longrightarrow$ Duplicator wins

S play ∧

- S chooses $\begin{cases} J \sqcup K = I \\ p_J + p_K = p-1 \end{cases}$

D chooses J OR K $= \boxed{}$

→ we move to $\{(u_i, v_i)\}_{i \in \boxed{}}, p_{\boxed{}}$

WHAT COMES NEXT

→ full LTL with fixed Σ? NP-complete?

→ Formalize clearly the link with EF GAMES

→ others approx / hardness of approx?