

Logics with concrete domains: an introduction

Stéphane Demri
CNRS, LMF

NCL'22, Łódź + on-line, March 2022



Stéphane Demri, Karin Quaas:

Concrete domains in logics: a survey. ACM SIGLOG News 8(3): 6-29 (2021)

Concrete domains in TCS

- Constraint satisfaction problems (CSP).
- Satisfiability Modulo Theory (SMT) solvers.
String theories, arithmetical theories, array theories, etc.
See e.g. [Barrett & Tinelli, Handbook 2018]
- Description logics with concrete domains.
[Baader & Hanschke, IJCAI'91, Lutz, PhD 2002]
- Temporal logics with arithmetical constraints.
See e.g. [Bouajjani et al., LICS 95; Comon & Cortier, CSL'00]
- Verification of database-driven systems.
[Deutsch & Hull & Vianu, SIGMOD 2014]

Concrete domains and constraints

- Concrete domain $\mathcal{D} = (\mathbb{D}, R_1, R_2, \dots)$: fixed non-empty domain with a family of relations.
- $(\mathbb{Z}, +, <, =, 0, 1)$, $(\mathbb{N}, <, +1)$, $(\mathbb{R}, <, =)$, $(\mathbb{D}, \equiv)$.
- Terms are built from variables x and expressions $X^i x$.
- Constraint C : Boolean combination of atomic constraints of the form $R(t_1, \dots, t_d)$.

$$(Xx_1 = x_2 + XXXx_3) \vee (x_1 > Xx_4)$$

- Constraints are interpreted on valuations $\mathfrak{v}$ that assign elements from $\mathbb{D}$ to the terms and

$$\mathfrak{v} \models R(t_1, \dots, t_d) \text{ iff } (\mathfrak{v}(t_1), \dots, \mathfrak{v}(t_d)) \in R^{\mathcal{D}}.$$

- A constraint C over $\mathcal{D}$ is satisfiable $\stackrel{\text{def}}{\iff}$ there is a valuation $\mathfrak{v}$ such that $\mathfrak{v} \models C$.

More examples

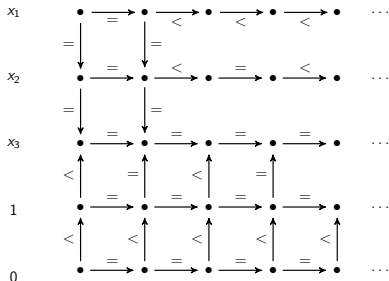
- $(\mathbb{Q}, <, =)$, $(\mathbb{R}, <, =)$, $(\mathbb{Z}, <, =)$, $(\mathbb{N}, <, =)$.
- $(\{0, 1\}^*, \preceq_{pre})$ with binary strings.
- Temporal concrete domain $\mathcal{D}_A = (I_{\mathbb{Q}}; (R_i)_{i \in [1, 13]})$ with
 - $I_{\mathbb{Q}}$: set of closed intervals $[r, r'] \subseteq \mathbb{Q}$
 - $(R_i)_{i \in [1, 13]}$ is the family of 13 Allen's relations.

[Allen83; CACM 1983]
- Concrete domain RCC8 with space regions in $\mathbb{R}^2$ contains topological relations between spatial regions.

See e.g. [Wolter & Zakharyashev, KR'00]

Symbolic models – the linear case

- AC_k : set of atomic constraints built over $\{x_1, \dots, x_k\}$ and $\{Xx_1, \dots, Xx_k\}$. (' Xx ' refers to the next value of x .)
- Symbolic model $w : \mathbb{N} \rightarrow \mathcal{P}(AC_k)$. (w -sequence)



- w is $\mathcal{D}$ -satisfiable $\stackrel{\text{def}}{\iff}$ there is $v : \mathbb{N} \times \{x_1, \dots, x_k\} \rightarrow \mathbb{D}$ such that for all i , $\{c \in AC_k \mid v, i \models c\} = w(i)$.
- $v, i \models x = Xy$ iff $v(i, x) = v(i + 1, y)$.

A selection of problems

- Given a concrete domain $\mathcal{D}$, how to characterise the class of $\mathcal{D}$ -satisfiable symbolic models?

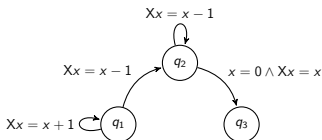
$(\{x > Xx\}^\omega \text{ not } \mathbb{N}\text{-satisfiable})$

- Given a formalism to define symbolic models (logics, automata, etc.), how to determine whether a recognized $\mathcal{D}$ -satisfiable symbolic model exists?
- Can the class of $\mathcal{D}$ -satisfiable symbolic models be expressed by a given formalism?

$(\omega\text{-regularity/Büchi automata?})$

- In this talk:
 - Concrete domains: $(\mathbb{Q}, <, =)$, $(\mathbb{N}, <, =)$.
 - Formalisms: constrained automata, constrained LTL, description logics, MSO-like logics.

Constrained automata



- $\mathcal{D}$ -automaton $\mathbb{A} = (S, \delta, I, F)$ with k variables:
 - S is a non-empty finite set of control states,
 - Set $I \subseteq S$ of initial states; set $F \subseteq S$ of final states,
 - δ is a finite subset of $S \times C_k \times S$, where C_k is the set of $\mathcal{D}$ -constraints built over $\{x_1, \dots, x_k\} \cup \{Xx_1, \dots, Xx_k\}$.

[Revesz, Book 2002]

- $v_0 v_1 \dots \in L(\mathbb{A}) \stackrel{\text{def}}{\iff}$ there is $q_0 \xrightarrow{C_0} q_1 \xrightarrow{C_1} \dots$ such that
 - $q_0 \in I$ and $q \in F$ occurs infinitely often in $q_0 q_1 q_2 \dots$.
 - for all $i \in \mathbb{N}$,

$$q_i \xrightarrow{C_i} q_{i+1} \in \delta \text{ and } v_i, v_{i+1} \models C_i.$$

Non-emptiness problem

- Non-emptiness problem for $\mathcal{D}$ -automata takes as input a $\mathcal{D}$ -automaton $\mathbb{A}$ and asks whether $L(\mathbb{A}) \neq \emptyset$.
- $L(\mathbb{A}) \neq \emptyset$ iff for some symbolic model $w : \mathbb{N} \rightarrow \mathcal{P}(AC_k)$,
 - there is an infinite run $q_0 \xrightarrow{C_0} q_1 \xrightarrow{C_1} \dots$ such that for all $i \in \mathbb{N}$, validity of

$$\left(\bigwedge_{c \in w(i)} c \right) \wedge \left(\bigwedge_{c \in (AC_k \setminus w(i))} \neg c \right) \Rightarrow C_i$$

- w is $\mathcal{D}$ -satisfiable,

LT $\mathcal{L}(\mathcal{D})$: LTL with concrete domain $\mathcal{D}$

$$\phi ::= R(t_1, \dots, t_d) \mid \phi \wedge \phi \mid \neg \phi \mid X\phi \mid \phi U \phi$$

(the t_i 's are terms of the form $X^j x$)

- LT $\mathcal{L}(\mathcal{D})$ model $\mathbf{v} : \mathbb{N} \times \text{VAR} \rightarrow \mathbb{D}$.

Satisfaction relation

- $\mathbf{v}, i \models R(t_1, \dots, t_d) \stackrel{\text{def}}{\iff} (\mathbf{v}(i, t_1), \dots, \mathbf{v}(i, t_d)) \in R^{\mathcal{D}}$

- $\mathbf{v}, i \models X\phi \stackrel{\text{def}}{\iff} \mathbf{v}, i+1 \models \phi$

$$x_1 \quad 0 \quad \frac{3}{8} \quad \frac{1}{9} \quad 3 \quad \dots$$

$$x_2 \quad \frac{1}{2} \quad \mathbf{0} \quad \frac{3}{4} \quad 2 \quad \dots$$

$$x_3 \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \mathbf{1} \quad \dots$$

$$x_4 \quad 1 \quad 2 \quad 3 \quad 4 \quad \dots$$

$$\models F(x_2 < X^2 x_3)$$

- Automata-based approach for temporal logics applies!

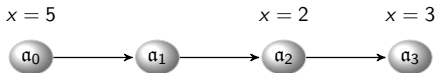
[Vardi & Wolper, IC 1994]

Branching-time temporal logics

- $\mathcal{D}$ -decorated Kripke structure $\mathcal{K}$ is a structure of the form $(\mathcal{D}, \mathcal{W}, \mathcal{R}, I, \mathbf{v})$ such that
 - Concrete domain $\mathcal{D} = (\mathbb{D}, \sigma)$, Kripke structure $(\mathcal{W}, \mathcal{R}, I)$
 - $\mathbf{v} : \mathcal{W} \times \text{VAR} \rightarrow \mathbb{D}$ is a valuation function.
- $\text{CTL}^*(\mathcal{D})$ formulae

$$\phi := \neg\phi \mid \phi \wedge \phi \mid \text{E}\Phi \quad \Phi := \phi \mid R(\mathbf{t}_1, \dots, \mathbf{t}_d) \mid \neg\Phi \mid \Phi \wedge \Phi \mid \text{X}\Phi \mid \Phi \text{U}\Phi$$

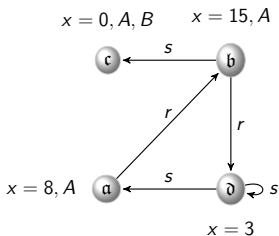
- Satisfaction relation
 - $\mathcal{K}, w \models \text{E}\Phi$ iff there is an infinite path π starting from w such that $\mathcal{K}, \pi \models \Phi$,
 - $\mathcal{K}, \pi \models R(\mathbf{t}_1, \dots, \mathbf{t}_d)$ iff $(\mathbf{v}(\pi(0), \mathbf{t}_1), \dots, \mathbf{v}(\pi(0), \mathbf{t}_d)) \in R^{\mathcal{D}}$ & $\mathbf{v}(\pi(0), \text{X}^j x) \stackrel{\text{def}}{=} \mathbf{v}(\pi(j), x)$



$$a_0 \models \text{E}(x = \text{X}^2 x + \text{X}^3 x)$$

Description logics with concrete domains

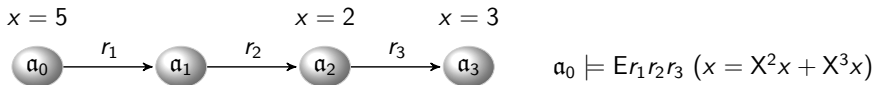
- Description logics are well-known logical formalisms for knowledge representation. [Baader et al., Book 2017]
- Concrete domains in DLs to refer to concrete objects and built-in predicates on these objects for designing concepts. [Baader & Hanschke, IJCAI'91, Lutz, PhD 2002]
- Role names $N_R = \{r, s, \dots\}$ and role path $P = r_1 \cdots r_n$.



$\mathcal{D}$ -decorated interpretations
 $(\mathcal{D}, \mathcal{W}, (\mathcal{R}_r)_{r \in N_R}, I, \mathfrak{v})$ with
 $\mathfrak{v} : \mathcal{W} \times \text{VAR} \rightarrow \mathbb{D}$.
(often partial in the literature)

ALC^ℓ($\mathcal{D}$) (with “linear-path constraints”)

- ALC^ℓ($\mathcal{D}$)-formulae *(unorthodox pre)*
 $\phi ::= p \mid Er_1 \cdots r_n R(t_1, \dots, t_d) \mid \phi \wedge \phi \mid \neg \phi \mid EX_r \phi$
- $\mathcal{K}, w \models EX_r \phi \stackrel{\text{def}}{\iff}$ there is $w' \in \mathcal{R}_r(w)$ s.t. $\mathcal{K}, w' \models \phi$,



- Logics of the form ALC^ℓ($\mathcal{D}$) considered in
[Carapelle & Turhan, ECAI'16; Labai & Ortiz & Simkus, KR'20]
- Conditions on $\mathcal{D}$ for decidability/low complexity studied
in [Lutz & Milićić, JAR 2007; Baader & Rydval, IJCAR'20]
... but this excludes domains such as $(\mathbb{N}, <, =)$.

What's next?

- ① Characterisations of satisfiable symbolic models.
 - Characterisation for $(\mathbb{R}, <, =)$ (and $(\mathbb{Q}, <, =)$).
 - Characterisation of $\mathcal{D}$ -satisfiable symbolic models for $\mathcal{D} = (\mathbb{N}, <, =)$.

- ② 3 methods for handling $\mathbb{N}$ -satisfiable symbolic models.
 - EHD approach with BMW.
 - Nonemptiness problem for $\mathbb{N}$ -automata.
 - Approximating condition $\mathcal{UPM}_{\mathbb{N}}$ for $(\mathbb{N}, <, =)$ with ultimately periodic symbolic models.

Easy case with $(\mathbb{R}, <, =)$ and $(\mathbb{Q}, <, =)$

- Symbolic model w is $\mathbb{Q}$ -satisfiable iff for all $i \in \mathbb{N}$,

$\mathcal{C}_{\mathbb{Q}}(1)$: $w(i)$ and $w(i+1)$ are satisfiable,

$\mathcal{C}_{\mathbb{Q}}(2)$: $\{Xx_1, \dots, Xx_k\}$ in $w(i)$ and $\{x_1, \dots, x_k\}$ in $w(i+1)$ are related in the same way.

- The set of $\mathbb{Q}$ -satisfiable symbolic models is ω -regular.

(good news)

- Sat. problem for $\text{LTL}(\mathbb{Q}, <, =)$ is PSPACE-complete.

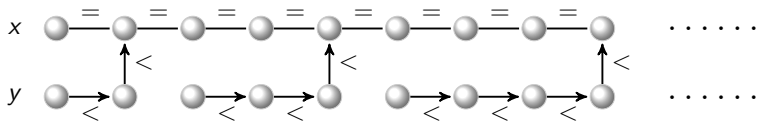
[Balbiani & Condotta, FroCoS'02]

- $\text{LTL}(\mathcal{D}_A)$ PSPACE-complete too with the temporal concrete domain $\mathcal{D}_A = (I_{\mathbb{Q}}; (R_i)_{i \in [1,13]})$.

[Balbiani & Condotta, FroCoS'02]

Characterisation for $(\mathbb{N}, <, =)$

- Symbolic model $w : \mathbb{N} \rightarrow \mathcal{P}(AC_k)$ understood as an infinite labelled graph on $\{x_1, \dots, x_k\} \times \mathbb{N}$.
- A simple non $\mathbb{N}$ -satisfiable symbolic model.



- Strict length of the path π :

$$\text{slen}(\pi) \stackrel{\text{def}}{=} \text{number of edges labelled by } <.$$

- Strict length of (x, i) :

$$\text{slen}((x, i)) \stackrel{\text{def}}{=} \sup \{ \text{slen}(\pi) : \text{path } \pi \text{ from } (x', i') \text{ to } (x, i) \}$$

$\mathbb{N}$ -satisfiable symbolic models

- Symbolic model w is $\mathbb{N}$ -satisfiable iff

$\mathcal{C}_{\mathbb{N}}(1)$: local consistency between two consecutive positions and,
 $(\mathcal{C}_{\mathbb{Q}}(1) \wedge \mathcal{C}_{\mathbb{Q}}(2))$

$\mathcal{C}_{\mathbb{N}}(2)$: any node has a finite strict length.

[Cerans, ICALP'94; Demri & D'Souza, IC 07; Carapelle & Kartzow & Lohrey, CONCUR'13; Exibard & Filiot & Khalimov, STACS'21]

- The set of $\mathbb{N}$ -satisfiable symbolic models is not ω -regular.

The EHD approach

- The set of $\mathbb{N}$ -satisfiable symbolic models is not ω -regular but can it be captured by decidable extensions of MSO?
(MSO = monadic 2nd logic $\approx$ Büchi automata)
- Starting point of the EHD approach with the bounding quantifier B. [Carapelle & Kartzow & Lohrey, CONCUR'13]
- Bounding quantifier B: $BX.\phi(X)$ expresses that there is a finite bound on the size of the sets that satisfy $\phi(X)$.
[Bojańczyk, CSL'04]
- B fits well to express the condition $\mathcal{C}_{\mathbb{N}}(2)$.

Decidability status

- Satisfiability $\text{MSO}+\text{B}$ is undecidable over ω -words.

[Bojańczyk & Parys & Toruńczyk, STACS'16]

- Boolean combinations of MSO and $\text{WMSO}+\text{B}$ (BMW) is decidable over infinite trees of finite branching degree.

[Carapelle & Kartzow & Lohrey, CONCUR'13]

- Negation-closed $\mathcal{D}$ with $\text{EHD}(\text{BMW})$ -property.
Satisfiability problem for $\text{CTL}^*(\mathcal{D})$ is decidable.

[Carapelle & Kartzow & Lohrey, JCSS 2016]

(tree model property + decidability of BMW)

EHD approach: two conditions

- $\mathcal{D}$ negation-closed if complements of relations definable by positive existential first-order formulae over $\mathcal{D}$.
 $(\neg(x = n) \Leftrightarrow \exists y (y = n) \wedge ((x < y) \vee (y < x)))$
- EHD(BMW) property for symbolic models.
There is ϕ_{SAT} in BMW for ω -words such that
 w is $\mathbb{N}$ -satisfiable iff $w \models \phi_{\text{SAT}}$.
- EHD(BMW) property (complete version).
For every finite subsignature τ , one can compute ϕ_τ such that for every countable τ -structure $\mathcal{S}$,
 $\underbrace{\text{there is an homomorphism from } \mathcal{S} \text{ to } \mathcal{D}}_{\approx \mathcal{D}\text{-satisfiability}} \text{ iff } \mathcal{S} \models \phi_\tau$.
- EHD = “the Existence of a Homomorphism is Definable”.

New decidability results

- $(\mathbb{Z}, <, =, (=_n)_{n \in \mathbb{Z}})$ has the EHD(BMW)-property.
- The satisfiability problem for $\text{CTL}^*(\mathbb{Z}, <, =, (=_n)_{n \in \mathbb{Z}})$ is decidable. [Carapelle & Kartzow & Lohrey, JCSS 2016]
- Satisfiability w.r.t. TBoxes for $\text{ALC}^\ell(\mathbb{Z}, <, =, (=_n)_{n \in \mathbb{Z}})$ is decidable [Carapelle & Turhan, ECAI'16]

N-automata

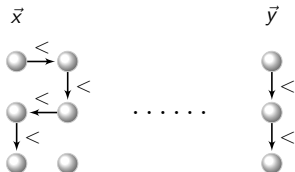
- EHD good for decidability, unsatisfactory for complexity!
- Concrete domains $\mathcal{D} = (\mathbb{D}, <, P_1, \dots, P_l, =_{\mathfrak{d}_1}, \dots, =_{\mathfrak{d}_m})$, where $(\mathbb{D}, <)$ is a linear ordering and the P_i 's are unary relations.
[Segoufin & Toruńczyk, STACS'11]
- Existence of accepting runs characterised by existence of extensible lassos.

N-automata: extensible lassos

$\mathbb{A}$ has an accepting run iff there are finite runs π, λ s.t.

- ① $\pi = (q_I, \vec{x}_0) \xrightarrow{*} (q_F, \vec{x})$ and $\lambda = (q_F, \vec{x}) \xrightarrow{+} (q_F, \vec{y})$
- ② “type($\vec{x}$) = type($\vec{y}$)”, $\vec{x} \leq \vec{y}$ and $\text{dv}(\vec{x}) \leq \text{dv}(\vec{y})$.

$$0^7 7^2 9^6 15 \quad \text{dv}\left(\begin{pmatrix} 15 \\ 9 \\ 7 \end{pmatrix}\right) = \begin{pmatrix} 7 \\ 2 \\ 6 \end{pmatrix}$$



Conditions (2) and (3) allow us to repeat infinitely λ .

- ③ For all $j \in [1, k]$ such that $\vec{x}[j] = \vec{y}[j]$, there is no j' such that $\vec{x}[j'] < \vec{y}[j']$ and $\vec{x}[j'] < \vec{x}[j]$.

$\mathbb{N}$ -automata: lasso detection in PSpace

- Existence of finite runs π, λ can be checked in PSPACE.
- The non-emptiness problem for $(\mathbb{N}, <)$ -automata is PSPACE-complete. [Segoufin & Toruńczyk, STACS'11]
- A similar method used in [Kartzow & Weidner, CoRR 2015].
- PSPACE-completeness for the concrete domains
 - $\mathcal{D}_{\mathbb{Q}^*} = (\mathbb{Q}^*; \preceq_{\text{pre}}, \preceq_{\text{lex}}, =_{\mathfrak{d}_1}, \dots, =_{\mathfrak{d}_m})$.
 - $\mathcal{D}_{[1, \alpha]^*} = ([1, \alpha]^*; \preceq_{\text{pre}}, \preceq_{\text{lex}}, =_{\mathfrak{d}_1}, \dots, =_{\mathfrak{d}_m}), \alpha \geq 2$.[Kartzow & Weidner, CoRR 2015]

Condition $\mathcal{UPM}_{\mathbb{N}}$: ultimately periodic models

- $\mathbb{N}$ -automata and $\text{LTL}(\mathbb{N}, <, =)$ define ω -regular classes of symbolic models with uninterpreted constraints.
- A symbolic model w is ultimately periodic iff w of the form

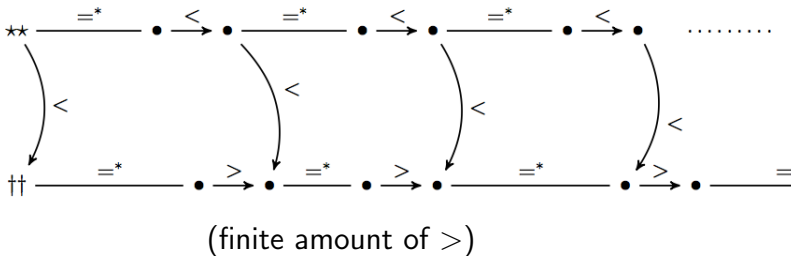
$$w(0) \cdots w(I-1) \cdot \left(w(I) \cdots w(I+J) \right)^\omega$$

- Characterisation for $\mathbb{N}$ -satisfiable ultimately periodic models might be simpler than the general case.

Condition $\mathcal{UPM}_{\mathbb{N}}$: definition

Symbolic model w satisfies the condition $\mathcal{UPM}_{\mathbb{N}}$ iff

- ① Local consistency btw. two consecutive positions holds.
- ② There is no infinite $(z_1, j_1) \xrightarrow{a_1} (z_2, j_2) \xrightarrow{a_2} (z_3, j_3) \cdots$ s.t. $\{a_1, a_2, \dots\} \subseteq \{=, >\}$ and an infinite amount of a_j 's are equal to $>$.
- ③ There do not exist nodes $\star\star$ and $\dagger\dagger$ such that
(infinite amount of $<$)



$$\mathcal{C}_{\mathbb{N}}(1) \wedge \mathcal{C}_{\mathbb{N}}(2) \Rightarrow \mathcal{UPM}_{\mathbb{N}}$$

Condition $UPM_{\mathbb{N}}$: properties

- Ultimately periodic symbolic model w . Equivalence btw.
 - w is $\mathbb{N}$ -satisfiable.
 - w satisfies the condition $UPM_{\mathbb{N}}$.

[Demri & D'Souza, IC 2007; Exibard & Filiot & Reynier, STACS'21]

- The class of symbolic models having $UPM_{\mathbb{N}}$ is ω -regular.
- By-products:
 - Non-emptiness problem for $\mathbb{N}$ -automata is in PSPACE.
 - Satisfiability problem for $LTL(\mathbb{N}, <, =)$ is in PSPACE.
- Remarkable generalisation to description logics:
 - $UPM_{\mathbb{N}}$ for regular tree symbolic models and regularity via Rabin tree automata.
 - Satisfiability problem w.r.t. TBoxes in $ALC^{\ell}(\mathbb{N}, <, =)$ is in EXPTIME. [Labai & Ortiz & Šimkus, KR'20]
- Results apply to $(\mathbb{Z}, <, =)$ with adequate adaptations.

Concluding remarks

- Presentation of three methods for handling $\mathbb{N}$ -satisfiable symbolic models.
 - ① MSO-like logics (EHD approach),
 - ② $\mathcal{D}$ -automata (for linear domains or strings)
 - ③ overapproximation (condition $\mathcal{UPM}_{\mathbb{N}}$)
- A selection of open problems.
 - Decidability status for $\text{LTL}(\{0, 1\}^*, \preceq_{pre}, \preceq_{suf})$.
 - Satisfiability w.r.t. TBoxes for $\text{ALC}^\ell(\mathbb{Z}; <, =, (=_n)_{n \in \mathbb{Z}})$ in EXPTIME with integers in binary.