Logics with concrete domains: an introduction

Stéphane Demri
CNRS, LMF

NCL’22, Łódź + on-line, March 2022
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 $D = (\mathbb{D}, R_1, R_2, \ldots)$: 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 $R(t_1, \ldots, t_d)$.

  $$(X x_1 = x_2 + XXX x_3) \lor (x_1 > X x_4)$$

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

  $$\nu \models R(t_1, \ldots, t_d) \iff (\nu(t_1), \ldots, \nu(t_d)) \in R^D.$$

- A constraint $C$ over $D$ is satisfiable \( \iff \) there is a valuation $\nu$ such that $\nu \models C$. 
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 & Zakharyaschev, KR’00]
Symbolic models – the linear case

• $AC_k$: set of atomic constraints built over $\{x_1, \ldots, x_k\}$ and $\{Xx_1, \ldots, Xx_k\}$. (‘Xx’ refers to the next value of $x$.)

• Symbolic model $w : \mathbb{N} \rightarrow \mathcal{P}(AC_k)$. ($\omega$-sequence)

• $w$ is $\mathcal{D}$-satisfiable $\iff$ there is $\nu : \mathbb{N} \times \{x_1, \ldots, x_k\} \rightarrow \mathcal{D}$ such that for all $i$, $\{c \in AC_k \mid \nu, i \models c\} = w(i)$.

• $\nu, i \models x = Xy$ iff $\nu(i, x) = \nu(i + 1, y)$. 
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

\[ Xx = x - 1 \]

\[ Xx = x + 1 \]

\[ x = 0 \land Xx = x \]

\[ q_0 \rightarrow q_1 \rightarrow q_2 \rightarrow \ldots \]

\[ A = (S, \delta, I, F) \] 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,
- \( \delta \) is a finite subset of \( S \times C_k \times S \), where \( C_k \) is the set of \( D \)-constraints built over \( \{x_1, \ldots, x_k\} \cup \{Xx_1, \ldots, Xx_k\} \).

[Revesz, Book 2002]

- \( v_0v_1\cdots \in L(A) \overset{\text{def}}{\iff} \) there is \( q_0 \xrightarrow{C_0} q_1 \xrightarrow{C_1} \cdots \) such that
  - \( q_0 \in I \) and \( q \in F \) occurs infinitely often in \( q_0q_1q_2 \cdots \).
  - for all \( i \in \mathbb{N} \),
    \[ q_i \xrightarrow{C_i} q_{i+1} \in \delta \) 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 $A$ and asks whether $L(A) \neq \emptyset$.

- $L(A) \neq \emptyset$ iff for some symbolic model $w : \mathbb{N} \rightarrow \mathcal{P}(AC_k)$,
  - there is an infinite run $q_0 \xrightarrow{C_0} q_1 \xrightarrow{C_1} \cdots$ such that for all $i \in \mathbb{N}$, validity of
    \[
    \left( \bigwedge_{c \in w(i)} c \right) \land \left( \bigwedge_{c \in (AC_k \setminus w(i))} \neg c \right) \Rightarrow C_i
    \]
  - $w$ is $\mathcal{D}$-satisfiable,
LTL(\mathcal{D}): LTL with concrete domain \mathcal{D}

\phi ::= R(t_1, \ldots, t_d) \mid \phi \land \phi \mid \neg \phi \mid X\phi \mid \phi U \phi

(the t_i's are terms of the form \textit{X}^jx)

• LTL(\mathcal{D}) model \nu : \mathbb{N} \times \text{VAR} \to \mathcal{D}.

Satisfaction relation

• \nu, i \models R(t_1, \ldots, t_d) \iff (\nu(i, t_1), \ldots, \nu(i, t_d)) \in R^{\mathcal{D}}

• \nu, i \models X\phi \iff \nu, i + 1 \models \phi

• 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 $(\mathcal{D}, \mathcal{W}, R, l, v)$ such that
  - Concrete domain $\mathcal{D} = (\mathbb{D}, \sigma)$, Kripke structure $(\mathcal{W}, R, l)$
  - $v : \mathcal{W} \times \text{VAR} \rightarrow \mathbb{D}$ is a valuation function.

- CTL*($\mathcal{D}$) formulae

  $\phi ::= \neg \phi \mid \phi \land \phi \mid E\Phi \quad \Phi ::= \phi \mid R(t_1, \ldots, t_d) \mid \neg \Phi \mid \Phi \land \Phi \mid X\Phi \mid \Phi U \Phi$

- Satisfaction relation

  - $\mathcal{K}, w \models E\Phi$ iff there is an infinite path $\pi$ starting from $w$ such that $\mathcal{K}, \pi \models \Phi$,
  - $\mathcal{K}, \pi \models R(t_1, \ldots, t_d)$ iff
    $(v(\pi(0), t_1), \ldots, v(\pi(0), t_d)) \in R^\mathcal{D} \& v(\pi(0), Xjx) \overset{\text{def}}{=} v(\pi(j), x)$

\[ x = 5 \quad x = 2 \quad x = 3 \]

\[ a_0 \quad a_1 \quad a_2 \quad a_3 \]

\[ a_0 \models E(x = X^2x + X^3x) \]
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 $N_R = \{r, s, \ldots\}$ and role path $P = r_1 \cdots r_n$. 

\[ x = 0, A, B \quad x = 15, A \]

\[ x = 8, A \quad x = 3 \]

$D$-decorated interpretations $(D, W, (R_r)_{r \in N_R}, I, v)$ with $v : W \times \text{VAR} \to D$. (often partial in the literature)
\( \text{ALC}^{\ell}(D) \) (with “linear-path constraints”)

- \( \text{ALC}^{\ell}(D) \)-formulae (unorthodox presentation)
  \[
  \phi ::= p \mid E_{r_1} \cdots r_n R(t_1, \ldots, t_d) \mid \phi \land \phi \mid \neg \phi \mid \text{EX}_{r}\phi
  \]

- \( \mathcal{K}, w \models \text{EX}_{r}\phi \iff \text{there is } w' \in R_{r}(w) \text{ s.t. } \mathcal{K}, w' \models \phi, \)

\[
\begin{align*}
  x &= 5 & x &= 2 & x &= 3 \\
  a_0 &\xrightarrow{r_1} a_1 & a_1 &\xrightarrow{r_2} a_2 & a_2 &\xrightarrow{r_3} a_3 \\
  a_0 &\models E_{r_1}r_2r_3 (x = X^2x + X^3x)
\end{align*}
\]

- Logics of the form \( \text{ALC}^{\ell}(D) \) considered in
  [Carapelle & Turhan, ECAI’16; Labai & Ortiz & Simkus, KR’20]

- Conditions on \( D \) 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 \(i \in \mathbb{N}\),
  \[C_Q(1) : w(i) \text{ and } w(i + 1) \text{ are satisfiable},\]
  \[C_Q(2) : \{Xx_1, \ldots, Xx_k\} \text{ in } w(i) \text{ and } \{x_1, \ldots, x_k\} \text{ in } w(i + 1) \text{ are related in the same way}.\]

- The set of \(\mathbb{Q}\)-satisfiable symbolic models is \(\omega\)-regular.
  (good news to use Büchi automata)

- Sat. problem for \(\text{LTL}(\mathbb{Q}, <, =)\) is \(\text{PSpace}\)-complete.
  [Balbiani & Condotta, FroCoS’02]

- \(\text{LTL}(\mathcal{D}_A)\) \(\text{PSpace}\)-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 $w : \mathbb{N} \to \mathcal{P}(\mathbb{AC}_k)$ understood as an infinite labelled graph on $\{x_1, \ldots, x_k\} \times \mathbb{N}$.

- A simple non $\mathbb{N}$-satisfiable symbolic model.

- Strict length of the path $\pi$:
  \[ \text{slen}(\pi) \overset{\text{def}}{=} \text{number of edges labelled by } <. \]

- Strict length of $(x, i)$:
  \[ \text{slen}((x, i)) \overset{\text{def}}{=} \sup \{\text{slen}(\pi) : \text{path } \pi \text{ from } (x', i') \text{ to } (x, i)\} \]
\( \mathbb{N} \)-satisfiable symbolic models

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

\[ C_\mathbb{N}(1) : \text{local consistency between two consecutive positions and,} \]
\[ (C_\mathbb{Q}(1) \land C_\mathbb{Q}(2)) \]
\[ C_\mathbb{N}(2) : \text{any node has a finite strict length.} \]

[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 \( \omega \)-regular.
The EHD approach

- The set of $\mathbb{N}$-satisfiable symbolic models is not $\omega$-regular but can it be captured by decidable extensions of MSO?

  ($\text{MSO} = \text{monadic 2nd logic} \cong \text{Büchi automata}$)

- 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 $C_\mathbb{N}(2)$. 


Decidability status

• Satisfiability $\text{MSO}^+\mathbb{B}$ is undecidable over $\omega$-words.
  [Bojańczyk & Parys & Toruńczyk, STACS’16]

• Boolean combinations of $\text{MSO}$ and $\text{WMSO}^+\mathbb{B}$ (BMW) is decidable over infinite trees of finite branching degree.
  [Carapelle & Kartzow & Lohrey, CONCUR’13]

• Negation-closed $\mathcal{D}$ with EHD(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 $\mathcal{D}$.
  \[ (\neg(x = n) \iff \exists y (y = n) \land ((x < y) \lor (y < x))) \]

- EHD(BMW) property for symbolic models.
  There is $\phi_{\text{SAT}}$ in BMW for $\omega$-words such that $w$ is $\mathbb{N}$-satisfiable iff $w \models \phi_{\text{SAT}}$.

- EHD(BMW) property (complete version).
  For every finite subsignature $\tau$, one can compute $\phi_{\tau}$ such that for every countable $\tau$-structure $\mathcal{S}$, there is an homomorphism from $\mathcal{S}$ to $\mathcal{D}$ iff $\mathcal{S} \models \phi_{\tau}$.

- EHD = “the Existence of a Homomorphism is Definable”.

19
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]
- **$\mathbb{N}$-automata**

  - EHD good for decidability, unsatisfactory for complexity!

  - Concrete domains $\mathcal{D} = (\mathbb{D}, <, P_1, \ldots, P_l, =_{d_1}, \ldots, =_{d_m})$, 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.
\(\mathbb{N}\text{-}\text{automata: extensible lassos}\)

A has an accepting run iff there are finite runs \(\pi, \lambda\) s.t.

1. \(\pi = (q_I, \vec{x}_0) \xrightarrow{\ast} (q_F, \vec{x})\) and \(\lambda = (q_F, \vec{x}) \xrightarrow{+} (q_F, \vec{y'})\)

2. “\(\text{type}(\vec{x}) = \text{type}(\vec{y})\), \(\vec{x} \leq \vec{y}\) and \(\text{dv}(\vec{x}) \leq \text{dv}(\vec{y})\).

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

\[\vec{x}\begin{array}{c}
\Downarrow \quad < \\
\Downarrow \quad <
\end{array}\begin{array}{c}
\Downarrow \quad < \\
\Downarrow \quad <
\end{array}
\quad \begin{array}{c}
\Downarrow \quad < \\
\Downarrow \quad <
\end{array}\quad \begin{array}{c}
\Downarrow \quad < \\
\Downarrow \quad <
\end{array}\quad . . . . . \quad \begin{array}{c}
\Downarrow \quad < \\
\Downarrow \quad <
\end{array}\]

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 \(\vec{x}'[j'] < \vec{x}[j]\). 

Conditions (2) and (3) allow us to repeat infinitely \(\lambda\).
\textbf{N-automata: lasso detection in PSpace}

- Existence of finite runs $\pi, \lambda$ can be checked in $\text{PSpace}$.

- The non-emptiness problem for $(\mathbb{N}, <)$-automata is $\text{PSpace}$-complete. \cite{SegoufinTorunczykSTACS11}

- A similar method used in \cite{KartzowWeidnerCoRR2015}.

- \text{PSpace}-completeness for the concrete domains
  - $D_{Q^*} = (Q^*; \preceq_{\text{pre}}, \preceq_{\text{lex}}, =_1, \ldots, =_m)$.
  - $D_{[1, \alpha]^*} = ([1, \alpha]^*; \preceq_{\text{pre}}, \preceq_{\text{lex}}, =_1, \ldots, =_m), \alpha \geq 2$. \cite{KartzowWeidnerCoRR2015}
Condition $\text{UPM}_\mathbb{N}$: ultimately periodic models

- $\mathbb{N}$-automata and $\text{LTL}(\mathbb{N}, <, =)$ define $\omega$-regular classes of symbolic models with uninterpreted constraints.

- A symbolic model $w$ is ultimately periodic iff $w$ of the form

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

- Characterisation for $\mathbb{N}$-satisfiable ultimately periodic models might be simpler than the general case.
Condition $\mathcal{UPM}_N$: definition

Symbolic model $w$ satisfies the condition $\mathcal{UPM}_N$ 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.
   \[ \{a_1, a_2, \ldots\} \subseteq \{=, >\} \] 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 $<$)

   \[ \star\star \xrightarrow{=} \bullet \xrightarrow{<} \bullet \xrightarrow{=} \bullet \xrightarrow{<} \bullet \xrightarrow{=} \bullet \xrightarrow{<} \bullet \xrightarrow{=} \bullet \xrightarrow{<} \bullet \cdots \]

   (finite amount of $>$)

\[ C_N(1) \wedge C_N(2) \Rightarrow \mathcal{UPM}_N \]
**Condition $\mathcal{UPM}_N$: properties**

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

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

- The class of symbolic models having $\mathcal{UPM}_N$ is $\omega$-regular.

- By-products:
  - Non-emptiness problem for $\mathbb{N}$-automata is in $\text{PSPACE}$.
  - Satisfiability problem for $\text{LTL}(\mathbb{N}, <, =)$ is in $\text{PSPACE}$.

- Remarkable generalisation to description logics:
  - $\mathcal{UPM}_N$ for regular tree symbolic models and regularity via Rabin tree automata.
  - Satisfiability problem w.r.t. TBoxes in $\text{ALC}^\ell(\mathbb{N}, <, =)$ is in $\text{EXPTIME}$.  
    [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 LTL(\( \{0,1\}^*, \preceq_{\text{pre}}, \preceq_{\text{suf}} \)).
  - Satisfiability w.r.t. TBoxes for \( \text{ALC}^\ell(\mathbb{Z}; <, =, (\equiv_n)_{n \in \mathbb{Z}}) \) in \( \mathsf{EXPTIME} \) with integers in binary.