Satisfiability Checking and Conjunctive Query Answering in Description Logics with Global and Local Cardinality Constraints

Franz Baader
Bartosz Bednarczyk
Sebastian Rudolph

Technische Universität Dresden
and University of Wrocław

Description Logic Workshop 2019
Oslo, June 18th, 2019

Agenda

- Introduction to the logic $\mathcal{ALCSCC}^{++}$
- Expressivity examples
- Satisfiability is NExpTime-complete ...
- but conjunctive query entailment is undecidable
- Nice sub-fragment $\mathcal{ALCSCC}$ with decidable finite query answering

QFBAPA

- We recall quantifier-free fragment of Boolean Algebra with Presburger Arithmetic (QFBAPA)
- Set terms, boolean operations $\cap, \cup, \cdot^c$ on them, constants $\varnothing, \mathcal{U}$
- Set terms can be used to state set constraints, $s = t, s \subseteq t$
- Presburger arithmetic (PA) expressions are build from:
  - Integer constants. $\ldots, -2, -1, 0, 1, 2, \ldots$
  - set cardinalities $|s|$, for s being a set terms
- We can express Cardinality constraints $k = l, k < l, N \operatorname{dvd} l$

QBFBAPA formula = boolean comb. of set and cardinality constr.
Sat of QFBAPA is NP-complete [Kuncak&Rinard CADE 2007]

The definition of $\mathcal{ALCSCC}^{++}$

$\mathcal{ALCHQ}$ + concepts of the form sat(c) for a QFBAPA set/cardinality constaint c and c uses role names and $\mathcal{ALCSCC}^{++}$ concept descriptions in place of set variables

- $(\geq n r.C)^{\mathcal{I}} = \text{sat}(|C \cap r| \geq n)^{\mathcal{I}}$ - number restrictions
- $\text{sat}(|r| \, \text{dvd} \, 2)$ - even number of r successors
- $\text{sat}(|\top| \, \text{dvd} \, 2)$ - the total number of elements is even
- $\text{sat}(|A| = 1)$ - nominals
- $\text{sat}(\top \subseteq \text{sat}(r \cap s \subseteq \emptyset))$ - role disjointness
- $\text{sat}(\top \subseteq \text{sat}(|r| + |r^c| = |\mathcal{U}|))$ - role complementation
- $\text{sat}\big(\top \subseteq \text{sat}(|r| = |\mathcal{U}|)\big)$ - universal role

$\mathcal{ALCSCC}^{++}$ is NExpTime-complete

- We provide an exponential reduction to QFBAPA
- Matching lower bound from previous work [Baader&Ecke, GCAI'17]
- We define a notion of types and write a formula describing them
- M = set of all subconcepts from the input concept E
- If M is a set of concepts, then $t \subseteq M$ is a type if:
  - If $\neg C \in M$ then $C \in t \vee \neg C \in t$
  - If $C \sqcap D \in M$ then $C \sqcap D \in t$ iff $C \in t$ and $D \in t$
  - If $C \sqcup D \in M$ then $C \sqcup D \in t$ iff $C \in t$ or $D \in t$
- Such a type t can also be seen as a concept description C_t, which is the conjunction of all the elements of t.

Encoding types in QFBAPA

- Given a type t, we replace concepts C with X_C and roles r with $\mathrm{X}^\mathrm{t}_\mathrm{r}$. The resulting formula is ψ_t.

Encoding types in QFBAPA

- Given a type t, we replace concepts C with $\mathrm{X_C}$ and roles r with $\mathrm{X_r^t}$. The resulting formula is ψ_{t}.
- We can ensure boolean structure of types with $\beta :=$

$$\bigwedge_{\mathrm{C \sqcap D \in M}} \mathrm{X_{C \sqcap D}} = \mathrm{X_C} \cap \mathrm{X_D} \land \bigwedge_{\mathrm{C \sqcup D \in M}} \mathrm{X_{C \sqcup D}} = \mathrm{X_C} \cup \mathrm{X_D} \land \bigwedge_{\mathrm{\neg C \in M}} \mathrm{X_{\neg C}} = (\mathrm{X_C})^{\mathrm{c}}$$

Encoding types in QFBAPA

- Given a type t, we replace concepts C with X_C and roles r with X_r^t. The resulting formula is ψ_t.
- We can ensure boolean structure of types with $\beta :=$

$$\bigwedge_{C \sqcap D \in M} X_{C \sqcap D} = X_C \cap X_D \wedge \bigwedge_{C \sqcup D \in M} X_{C \sqcup D} = X_C \cup X_D \wedge \bigwedge_{\neg C \in M} X_{\neg C} = (X_C)^c$$

- Overall, we translate the concept E into the QFBAPA δ_E:

$$\delta_E := (|X_E| \geq 1) \wedge \beta \wedge \bigwedge_{t \in types(E)} (|\bigcap_{C \in t} X_C| = 0) \vee \psi_t.$$

Encoding types in QFBAPA

- Given a type t, we replace concepts C with X_C and roles r with X_r^t. The resulting formula is ψ_t.
- We can ensure boolean structure of types with $\beta :=$

$$\bigwedge_{C \sqcap D \in M} X_{C \sqcap D} = X_C \cap X_D \wedge \bigwedge_{C \sqcup D \in M} X_{C \sqcup D} = X_C \cup X_D \wedge \bigwedge_{\neg C \in M} X_{\neg C} = (X_C)^c$$

- Overall, we translate the concept E into the QFBAPA δ_E:

$$\delta_E := (|X_E| \geq 1) \wedge \beta \wedge \bigwedge_{t \in types(E)} (|\bigcap_{C \in t} X_C| = 0) \vee \psi_t.$$

- First conjunct = witness for E
- Last two conjuncts = for any type that is realized (i.e., has elements), the constraints of this type are satisfied

Conclusion: satisfiability for $\mathcal{ALCSCC}^{++}$

Lemma

The QFBAPA formula δ_{E} is of size at most exponential in the size of E, and it is satisfiable iff the $\mathcal{ALCSCC}^{++}$ concept description E is satisfiable.

Conclusion: satisfiability for $\mathcal{ALCSCC}^{++}$

Lemma

The QFBAPA formula δ_{E} is of size at most exponential in the size of E, and it is satisfiable iff the $\mathcal{ALCSCC}^{++}$ concept description E is satisfiable.

Satisfiablity of $\mathcal{ALCSCC}^{++}$

The sat problem for $\mathcal{ALCSCC}^{++}$ is NExpTime-complete.

Query answering is undecidable

- Source of undecidability: Universal role is expressible in $\mathcal{ALCSCC}^{++}$ (as we have seen before)
- Proof by [Pratt-Hartmann, Inf. Comput. 207] for $\mathcal{FO}^2$ without equality - sketchy, it is not clear whether it is correct
- So we prove it on our own! for $\mathcal{ALC}^{\mathrm{cov}}$, i.e., $\mathcal{ALC}$ extended by role cover axioms of the form $\mathrm{cov}(\mathrm{r}, \mathrm{s})$
- An interpretation $\mathcal{I}$ satisfies $\mathrm{cov}(\mathrm{r}, \mathrm{s})$ if $\mathrm{r}^{\mathcal{I}} \cup \mathrm{s}^{\mathcal{I}} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.
- Role cover axioms can be expressed in $\mathcal{ALCSCC}^{++}$ via

$$\mathrm{sat}\big(\top \subseteq \mathrm{sat}(|\mathrm{r} \cup \mathrm{s}| = |\mathcal{U}|)\big)$$

- Reduction from the looping Turing machines, i.e. it is undecidable whether DTM is looping

How decidability could be regain?

- We move to a less expressive, e.g., $\mathcal{ALCSCC}$ with RCBoxes.
- In $\mathcal{ALCSCC}$ all QFBAPA constraints must be local!
  It is equivalent to write:

$$C_{new} = C_{old} \cap \Big(\bigcup_{r \in N_R} r \Big)$$

- RCBoxes = finite sets of restricted cardinality constraints

$$N_1|C_1| + \ldots + N_k|C_k| \leq N_{k+1}|C_{k+1}| + \ldots + N_{k+l}|C_{k+l}|,$$

- Not able to express nominals!
  But still useful to express statistical knowledge-bases.

Decidable query answering

- We reduce query entailment to satisfiability
- We enrich our knowledge-base with ability to block all tree-shaped query matches (so-called rolling-up technique)
- Then we employ pumping technique to obtain models with arbitrary girth, while preserving satisfiability
- So if there is a counter-model, there is be a model without query tree-shaped matches

Blocking tree-shaped query matches

- We take an arbitrary CQ q.
- Consider all the possible ways q' how q can match as a tree
- We create a concept $C_{q',x}$, with the supposed meaning that $d \in C^{\mathcal{I}}_{q',x}$ if variable x from q' can be mapped to d in a query match represented by q'.
- We roll them into concepts in bottom-up way:
- $C_{q',x}$ equals $\sqcap_{C(x)\in q'} C$ if x is a leaf (i.e. $\prec$-minimal), otherwise:

$$\bigsqcap_{C(x)\in q'} C \sqcap \bigsqcap_{\substack{(x,y)\in E_{q'} \\ y\prec x}} \left(\exists \bigcap_{s(x,y)\in q'} s.C_{q',y} \right) \sqcap \bigsqcap_{\substack{(y,x)\in E_{q'} \\ y\prec x}} \left(\exists \bigcap_{s(y,x)\in q'} s^{-}.C_{q',y} \right)$$

Correctness

We define $\mathcal{R}^{\mathrm{Match_q}}$ as:

$$\bigsqcup_{\mathrm{q'} \in \mathrm{trees(q)}} \mathrm{C}_{\mathrm{q'}, \mathrm{x}^{\mathrm{r}}_{\mathrm{q'}}} \sqsubseteq \mathrm{Match_q} \tag{1}$$

Lemma

Assume that $\mathcal{R} \cup \mathcal{R}^{\mathrm{Match}}_{\mathrm{q}}$ has a model $\mathcal{I}$ such that $\mathrm{Match}^{\mathcal{I}}_{\mathrm{q}}$ is empty. Then $\mathcal{I}$ does not have any tree-shaped query matches.

Lemma

If there is a model $\mathcal{I}$ of $\mathcal{R}$ without any tree-shaped query matches, then $\mathcal{R}^* = \mathcal{R} \cup \mathcal{R}^{\mathrm{Match}}_{\mathrm{q}} \cup \{\top \sqsubseteq \neg \mathrm{Match_q}\}$ is satisfiable.

Pumping lemma for graphs

The girth of $\mathcal{I}$ is the length of a shortest (undirected) proper cycle contained in $\Delta^{\mathcal{I}}$. Below we present pumping method for graphs:

Pumping lemma for graphs

The girth of $\mathcal{I}$ is the length of a shortest (undirected) proper cycle contained in $\Delta^{\mathcal{I}}$. Below we present pumping method for graphs: Let $G = (V, E)$ be a graph, F be the set of functions $f : E \to \{0, 1\}$. Construct $H = (V', E')$ as follows:

- $V' = V \times F$
- E' be the set of edges $((u, f), (u', f'))$ such that $e = (u, u')$ is in edge of E and f' is the same as f except that the value at e is flipped: $f'(e) = 1 - f(e)$.

Then, the girth of H is at twice that of G.
Easy to generalize to structures.

Properties of pumping

Lemma

Let $\mathcal{I}$ be an interpretation with girth k.
Then the girth of pump($\mathcal{I}$) is at least 2k.

Lemma

Let $\mathcal{R}$ be an RCBox with a model $\mathcal{I}$.
Then pump($\mathcal{I}$) is a model of $\mathcal{R}$.

Decidable query answering

- We check enriched kb for satisfiability.
- If it is satisfiable, it does not have any tree-shaped query matches.
- We pump it at least $|q|$ times to obtain a countermodel.
- Thus satisfiable = query is not entailed.

Querying for $\mathcal{ALCSCC}$ RCBoxes

Conjunctive Querying is decidable for $\mathcal{ALCSCC}$ RCBoxes.

Conclusion

Satisfiablity and querying of $\mathcal{ALCSCC}^{++}$

The sat problem for $\mathcal{ALCSCC}^{++}$ is NExpTime-complete, but query-answering is undecidable

Querying for $\mathcal{ALCSCC}$ RCBoxes

Conjunctive Querying is decidable for $\mathcal{ALCSCC}$ RCBoxes.

We also know how to add ABoxes. We are working on improving the complexity (seems to be doable) of CQ entailment.
Open problems? The case with nominals!