Chase Variants & $k$-Boundedness

Reunion GRAPHIK 2018

Stathis Delivorias, Marie-Laure Mugnier, Michel Leclère, Federico Ulliana

17 December, 2018
Contents:

- Redundancy & Retraction
- Chase Variants
- $k$-Boundedness
- Chase Graphs & Ancestry
- Summary
Redundancy & Retraction

\[ \exists z_1 \exists z_2 \exists z_3 \; p(a, b) \land p(a, z_1) \land p(a, z_2) \land p(a, z_3) \]

\[ F = \{ p(a, b), p(a, z_1), p(a, z_2), p(a, z_3) \} \]
Redundancy & Retraction

$$\exists z_1 \exists z_2 \exists z_3 \ p(a, b) \land p(a, z_1) \land p(a, z_2) \land p(a, z_3)$$

$$F = \{ p(a, b), p(a, z_1), p(a, z_2), p(a, z_3) \}$$

Diagram:

- $z_1 \mapsto b$
- $z_2 \mapsto b$
- $z_3 \mapsto b$

Retraction $\sigma$
Redundancy & Retraction

$$\exists z_1 \exists z_2 \exists z_3 \ p(a, b) \land p(a, z_1) \land p(a, z_2) \land p(a, z_3)$$

$$F = \{p(a, b), p(a, z_1), p(a, z_2), p(a, z_3)\}$$

$$F \equiv \{p(a, b)\}$$
Redundancy & Retraction .2

\[ F = \{ p(a, b), p(b, z_1), p(z_1, a), p(a, z_3), p(z_3, z_2), p(z_2, a) \} \]
Redundancy & Retraction

\[ F = \{ p(a, b), p(b, z_1), p(z_1, a), p(a, z_3), p(z_3, z_2), p(z_2, a) \} \]
\[ F = \{ p(a, b), p(b, z_1), p(z_1, a), p(a, z_3), p(z_3, z_2), p(z_2, a) \} \]

\[ F \equiv \{ p(a, b), p(b, z_1), p(z_1, a) \} \]
Existential Rules:

\[ \forall \bar{x} \forall \bar{y} (\text{Body}(\bar{x}, \bar{y}) \rightarrow \exists \bar{z} \text{ Head}(\bar{x}, \bar{z})) \]

Body(\bar{x}, \bar{y}), \text{Head}(\bar{x}, \bar{z})$: conjunctions of atoms with variables from the sets \( \bar{x}, \bar{y} \) and \( \bar{x}, \bar{z} \) respectively.

Without loss of clarity we can omit the quantifiers:

\[ \text{Body}(\bar{x}, \bar{y}) \rightarrow \text{Head}(\bar{x}, \bar{z}) \]

for example

\[ p(x, y) \land q(x, z) \rightarrow p(x, w) \land s(z, y) \]
Rule Application

for example the rule

\[ R = \text{mammal}(x) \rightarrow \text{motherof}(w, x) \land \text{mammal}(w) \]

applied to the factbase

\[ F = \{ \text{mammal}(\text{Ringo}) \} \]

gives

\[ F' = \{ \text{mammal}(\text{Ringo}), \text{motherof}(w_0, \text{Ringo}), \text{mammal}(w_0) \} \]
the trigger

\((R, \{x \mapsto \text{Ringo}\})\)

applied to

\(F = \{\text{mammal}(\text{Ringo})\}\)

gives

\(F' = \{\text{mammal}(\text{Ringo}), \text{motherof}(w_0, \text{Ringo}), \text{mammal}(w_0)\}\)
Forward Chaining

is the process of using a set of rules $\mathcal{R}$ to infer information from a factbase $F$. 
The sequential application of rules on an evolving factbase is called a *derivation*. A derivation is not necessarily terminating...

\[
\{ \text{mammal}(Ringo), \text{motherof}(w_0, Ringo), \text{mammal}(w_0), \text{motherof}(w_1, w_0), \\
\text{mammal}(w_1), \text{motherof}(w_2, w_1), \text{mammal}(w_2), \text{motherof}(w_3, w_2), \text{mammal}(w_3), \ldots \}
\]
Chase Variants: Classes of Derivations

In existential rules, several *chase variants* handle differently the possible redundancies caused by the introduction of *nulls* to the factbase:

The **oblivious chase** performs all possible rule applications without repeating triggers:

\[
R = p(x, y) \rightarrow p(x, z) \land r(z)
\]

\[
F = \{p(a, b)\}
\]
Chase Variants: Classes of Derivations

In existential rules, several *chase variants* handle differently the possible redundancies caused by the introduction of *nulls* to the factbase:

The **semi-oblivious chase** does not repeat triggers which map the *frontier* variables similarly:

\[ R = p(x, y) \rightarrow p(x, z) \land r(z) \]

\[ F = \{ p(a, b) \} \]
Chase Variants: Classes of Derivations

In existential rules, several *chase variants* handle differently the possible redundancies caused by the introduction of *nulls* to the factbase:

The **restricted chase** does not apply triggers when their output can be retracted back to the existing factbase:

\[ R = p(x, y) \rightarrow p(y, z) \land p(z, x) \]

\[ F = \{ p(a, b) \} \]
Chase Variants: Classes of Derivations

In existential rules, several chase variants handle differently the possible redundancies caused by the introduction of nulls to the factbase:

The equivalent chase\(^1\) applies a trigger only if the resulting factbase is not logically equivalent to the existing factbase:

\[ R = p(x, y) \rightarrow p(x, x) \land p(y, z) \]

\[ F = \{ p(a, b) \} \]

\(^1\)the equivalent chase is equivalent with the core chase in terms of termination
Boundedness

Even if a rule(set) is such that forward chaining always terminates, like

\[ p(x, y) \land q(y, z) \rightarrow p(x, z) \]

it might not be bounded:

- \[ F = \{ p(a, b), q(b, c) \} \]
- \[ F = \{ p(a, b), q(b, c), q(c, d) \} \]
- \[ F = \{ p(a, b), q(b, c), q(c, d), q(d, e) \} \]
- \[ F = \{ p(a, b), q(b, c), q(c, d), q(d, e), q(e, f) \} \]
Boundedness

In particular a ruleset is *bounded* if there is a bound to the depth of derivations for every factbase.

\[ p(a, b) \quad q(b, c) \quad q(c, d) \quad q(d, e) \]

- \( p(a, c) \) depth=1
- \( p(a, d) \) depth=2
- \( p(a, e) \) depth=3
Boundedness

- Several semantic properties are ensured when a ruleset is bounded (chase termination, query rewriting termination, non-recursivity).

- There has been extensive research around boundedness in *datalog*, where it has been shown to be *undecidable*.

- We parametrize boundedness by chase variants, and introduce *\(k\)-boundedness*. 
X-\(k\)-boundedness

\(X \in \{O, SO, R, E\}\) is used as a parameter for each (X-)chase variant.

A ruleset \(\mathcal{R}\) is \textbf{X-\(k\)-bounded} if for every factbase \(F\), every X-derivation from \((F, \mathcal{R})\) is of depth at most \(k\).

For example

\[
\mathcal{R} = \{p(x, y) \rightarrow p(y, z) \land p(z, z) \land p(z, x)\}
\]

\(\mathcal{R}\) is \textbf{R-3-bounded}

but

\(\mathcal{R}\) is \textbf{SO-unbounded}
Chase derivations can be represented using the concept of a *chase graph*:

\[ R = \{ R_1, R_2, R_3, R_4 \} \]
Chase derivations can be represented using the concept of a *chase graph*:

$$\mathcal{R} = \{R_1, R_2, R_3, R_4\}$$
Preservation of Ancestry

The X-chase preserves ancestry if for every atom $A$ in an X-derivation, there is an X-derivation from its ancestors (& same ruleset) that produces $A$ at the same rank.

$$\mathcal{R} = \{ R_1, R_2, R_3, R_4 \}$$

$F' \rightarrow F$
Some Results

I. Determining if a set of rules is $X$-$\kappa$-bounded is decidable if the $X$-chase preserves ancestry.

II. The $X$-chase preserves ancestry for $X \in \{O, SO, R\}$.

III. The $E$-chase does not preserve ancestry.
E-chase Does Not Preserve Ancestry

\[ R_1 = s(x, y) \land p(y, y) \rightarrow \exists z \ p(x, z) \]
\[ R_2 = s(x, y) \rightarrow p(x, y) \]
\[ R_3 = t(x) \land p(x, y) \rightarrow q(y) \]
\[ R_4 = p(x, y) \rightarrow \exists w \ p(y, w) \]
E-chase Does Not Preserve Ancestry

\[ R_1 = s(x, y) \land p(y, y) \rightarrow \exists z \ p(x, z) \]
\[ R_2 = s(x, y) \rightarrow p(x, y) \]
\[ R_3 = t(x) \land p(x, y) \rightarrow q(y) \]
\[ R_4 = p(x, y) \rightarrow \exists w \ p(y, w) \]

\[ F' \]

\[ p(b, b) \quad s(a, b) \quad t(a) \]
Summary of thesis work

▶ Extensive study of the notion of chase variant and its potential properties, new framework that incorporates many chase variants.
▶ Defined two new chase variants, the vacuum chase and the local core chase, that detect more redundancies than the restricted chase.
▶ Decidability of $k$-boundedness shown for many chase variants (oblivious, semi-oblivious, restricted, breadth-first oblivious, breadth-first semi-oblivious, breadth-first restricted, parallel, local core chase).
▶ Open Question: (Un)Decidability of $k$-boundedness for the equivalent chase, the core chase, the frugal chase and the vacuum chase to be researched.
▶ Open Question: (Un)Decidability of $\exists$-boundedness of restricted chase.

THANK YOU !!