Lethe: Saturation-Based Reasoning for Non-Standard Reasoning Tasks

Patrick Koopmann, Renate A. Schmidt

LETHE

- “River of Forgetfulness”
- Usage from command line, as Java library, or via GUI
- Non-standard reasoning services relative to signatures
  - Forgetting / Uniform Interpolation
  - TBox Abduction
  - Logical Difference
- Support for expressive description logics (up to $\mathcal{SHQ}$)
- Problems reduced to forgetting, uses saturation-based reasoning

Uniform Interpolation/Forgetting

- Core Functionality of LETHE
- Restrict vocabulary in set of axioms
- Preserve entailments over that signature

Input Ontology

$$\textbf{Margherita} \sqsubseteq \forall\textbf{topping}.(\textit{Tomato} \sqcup \textit{Mozarella})$$
$$\textbf{American} \sqsubseteq \exists\textbf{topping}.\textit{Tomato}$$
$$\textbf{American} \sqsubseteq \exists\textbf{topping}.\textit{Mozarella}$$
$$\textbf{American} \sqsubseteq \exists\textbf{topping}.\textit{Pepperoni}$$
$$\textit{Tomato} \sqcup \textit{Mozarella} \sqsubseteq \textbf{VegTopping}$$
$$\textit{Pepperoni} \sqsubseteq \textbf{MeatTopping}$$

Uniform Interpolant

$$\textbf{Margherita} \sqsubseteq \forall\textbf{topping}.\textbf{VegTopping}$$
$$\textbf{American} \sqsubseteq \exists\textbf{topping}.\textbf{MeatTopping}$$

Applications of Forgetting

- Exhibit hidden concept relations
- Information hiding
- Ontology reuse
- Ontology summary
- Obfuscation
- ...

TBox Abduction

- Given TBox $\mathcal{T}$, axioms O, find axioms H with $\mathcal{T} \cup H \models O$
- "Complete" ontology such that given set of axioms is entailed
- *Abducibles* Σ specify concepts and roles allowed in solution
- Reducible to uniform interpolation:
  - $\mathcal{T} \cup \neg O \models \neg H$
  - Express $\neg(C \sqsubseteq D)$ as $\exists r^*.(C \sqcap D)$
  - Interpolate to set of abducibles
- Optimisations for large TBoxes and small inputs

Logical Difference

- "Semantical Diff"
- Analyse ontology changes, compare ontologies
- Look for *differing* entailments in specified signature Σ
- Compute new entailments in $\mathcal{O}_2$:
  - $LD(\mathcal{O}_1, \mathcal{O}_2, \Sigma) = \{\alpha \mid \alpha \in \mathcal{O}_2^\Sigma, \mathcal{O}_1 \not\models \alpha\}$
  - $\mathcal{O}_1^\Sigma$: Uniform interpolant of $\mathcal{O}_1$ for Σ
- Optimised for two use cases:
  1. Bigger changes, computation in minutes acceptable
  2. Small changes, computation in seconds required

Challenges Uniform Interpolation

1. Need for new reasoning methods
2. Cyclic TBoxes

   $A \sqsubseteq B, \quad B \sqsubseteq \exists r.B$
   $\mathcal{S} = \{A, r\}$

   - Uniform Interpolant in $\mathcal{ALC}$:
     - $A \sqsubseteq \exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\exists r.\ldots$
   - Solutions:

     Fixpoints: $A \sqsubseteq \nu X.(\exists r.X)$
     Approximate: $A \sqsubseteq \exists r.\exists r.\exists r.\top$
     Helper concepts: $A \sqsubseteq \exists r.D, \quad D \sqsubseteq \exists r.D$
3. High Complexity
   - $\mathcal{ALC}$ with fixpoints: 2^{2^n}, where n is size of input
   - Goal-oriented approach necessary

Normal form, $\mathcal{ALC}$

$\mathcal{ALC}$-Clause

$$\top \sqsubseteq L_1 \sqcup \ldots \sqcup L_n$$

L_i: $\mathcal{ALC}$-literal

$\mathcal{ALC}$-Literal

$$A \mid \neg A \mid \exists r.D \mid \forall r.D$$

A: any concept symbol, D: definer symbol

- Definer symbols: Special concept symbols, not part of signature
- Invariant: max 1 neg. definer symbol per clause
  $\Rightarrow$ $\neg D_1 \sqcup \exists r.D_2 \sqcup \neg B$, $\quad \cancel{\neg D_1 \sqcup \neg D_2 \sqcup A}$

Definer symbols

Invariant: max 1 neg. definer symbol per clause

- Allows easy translation to clausal form and back:

$$C_1 \sqcup \mathsf{Q}r.C_2 \iff C_1 \sqcup \mathsf{Q}r.D_1, \neg D_1 \sqcup C_2$$
$$C_1 \sqcup \nu X.C_2[X] \iff C_1 \sqcup \mathsf{Q}r.D_1, \neg D_1 \sqcup C_2[D]$$

$\Rightarrow$ Any set of clauses can be converted into an $\mathcal{ALC}\mu$-ontology ($\mathcal{ALC}$ with fixpoints)

- New definer symbols introduced by calculus
  - Number finitely bounded

Calculus

Resolution + *Combination rules*

- Resolution rule:
  - Direct inference on concept symbol to forget
  - Resolvent has to obey invariant

$$\frac{C_1 \sqcup A \quad C_2 \sqcup \neg A}{C_1 \sqcup C_2}$$

- Combination rules:
  - Combine context of nested definer symbols
  - Introduce new definer symbols
    - Representing conjunctions of definers
    - Max. 2^n new definer symbols
  - Make further inferences possible

Combination Rules

$\neg D_1 \sqcup A$
$C_1 \sqcup \exists r.D_1$

$\neg D_2 \sqcup B \sqcup \neg A$
$C_2 \sqcup \forall r.D_2$

Combination Rules

Cannot resolve due invariant

$\neg D_1 \sqcup A$

$C_1 \sqcup \exists r.D_1$

$\neg D_2 \sqcup B \sqcup \neg A$

$C_2 \sqcup \forall r.D_2$

Combination Rules

Cannot resolve due invariant

$\neg D_1 \sqcup A$

$C_1 \sqcup \exists r.D_1$

$\neg D_2 \sqcup B \sqcup \neg A$

$C_2 \sqcup \forall r.D_2$

combine

$C_1 \sqcup C_2 \sqcup \exists r.D_{12}$

$\neg D_{12} \sqcup A$

$\neg D_{12} \sqcup B \sqcup \neg A$

Combination Rules

Cannot resolve due invariant

$\neg D_1 \sqcup A$
$C_1 \sqcup \exists r.D_1$

$\neg D_2 \sqcup B \sqcup \neg A$
$C_2 \sqcup \forall r.D_2$

combine

$C_1 \sqcup C_2 \sqcup \exists r.D_{12}$
$\neg D_{12} \sqcup A$
$\neg D_{12} \sqcup B \sqcup \neg A$

Resolves to $\neg D_{12} \sqcup B$

Combination Rules $\mathcal{ALC}$

∀∃-Combination

$$\frac{C_1 \sqcup \forall r.D_1 \qquad C_2 \sqcup \exists r.D_2}{C_1 \sqcup C_2 \sqcup \exists r.D_{12}}$$

∀∀-Combination

$$\frac{C_1 \sqcup \forall r.D_1 \qquad C_2 \sqcup \forall r.D_2}{C_1 \sqcup C_2 \sqcup \forall r.D_{12}}$$

Combination Rules $\mathcal{SHQ}$

$\leq\leq$-Combination:

$$\frac{C_1 \sqcup \leq n_1 r_1.\neg D_1 \qquad C_2 \sqcup \leq n_2 r_2.\neg D_2 \qquad r \sqsubseteq r_1 \qquad r \sqsubseteq r_2}{C_1 \sqcup C_2 \sqcup \leq (n_1 + n_2) r.\neg D_{12}}$$

$\geq\leq$-Combination:

$$\frac{C_1 \sqcup \geq n_1 r_1.(D_1 \sqcup \ldots \sqcup D_m) \qquad C_2 \sqcup \leq n_2 r_2.\neg D_a \qquad r_1 \sqsubseteq_{\mathcal{R}} r_2}{C_1 \sqcup C_2 \sqcup \geq (n_1 - n_2) r_1.(D_{1a} \sqcup \ldots \sqcup D_{ma})}$$

$\leq\geq$-Combination:

$$\frac{C_1 \sqcup \leq n_1 r_1.\neg D_1 \qquad C_2 \sqcup \geq n_2 r_2.D_2 \qquad r_2 \sqsubseteq_{\mathcal{R}} r_1 \qquad n_1 \geq n_2}{\begin{array}{c} C_1 \sqcup C_2 \sqcup \leq (n_1 - n_2) r_1.\neg(D_1 \sqcup D_2) \sqcup \geq 1 r_1.D_{12} \\ \vdots \\ C_1 \sqcup C_2 \sqcup \leq (n_1 - 1) r_1.\neg(D_1 \sqcup D_2) \sqcup \geq n_2 r_1.D_{12} \end{array}}$$

$\geq\geq$-Combination:

$$\frac{C_1 \sqcup \geq n_1 r_1.\mathcal{D}_1 \qquad C_2 \sqcup \geq n_2 r_2.\mathcal{D}_2 \qquad r_1 \sqsubseteq_{\mathcal{R}} r \qquad r_2 \sqsubseteq_{\mathcal{R}} r}{\begin{array}{c} C_1 \sqcup C_2 \sqcup \geq (n_1 + n_2) r.(\mathcal{D}_1 \sqcup \mathcal{D}_2) \sqcup \geq 1 r.\mathcal{D}_{12} \\ \vdots \\ C_1 \sqcup C_2 \sqcup \geq (n_1 + 1) r.(\mathcal{D}_1 \sqcup \mathcal{D}_2) \sqcup \geq n_2 r.\mathcal{D}_{12} \end{array}}$$

Transitivity:

$$\frac{C \sqcup \leq 0 r_1.\neg D \qquad \mathrm{trans}(r_2) \in \mathcal{R} \qquad r_2 \sqsubseteq_{\mathcal{R}} r_1}{C \sqcup \leq 0 r_2.\neg D' \qquad \neg D' \sqcup D \qquad \neg D' \sqcup \leq 0 r_2.\neg D'}$$

Algorithm

- Compute all inferences on symbol to forget
- Use resolvents breaking invariant to choose combination rules
- Filter out all occurrences of symbol to forget
- Eliminate introduced symbols

Evaluation of Uniform Interpolation

$\mathcal{ALCH}$, forget 50 symbols	
Success Rate:	91.10%
Without Fixpoints:	95.29%
Duration Mean:	7.68 sec.
Duration Median:	2.74 sec.
Duration 90th percentile:	12.45 sec.

$\mathcal{ALCH}$, forget 100 symbols	
Success Rate:	88.10%
Without Fixpoints:	93.27%
Duration Mean:	18.03 sec.
Duration Median:	3.81 sec.
Duration 90th percentile:	21.17 sec.

$\mathcal{ALC}$ w. ABoxes, forget 50 symbols	
Success Rate:	94.79%
Without Fixpoints:	92.91%
Duration Mean:	23.94 sec.
Duration Median:	3.01 sec.
Duration 90th percentile:	29.00 sec.

$\mathcal{ALC}$ w. ABoxes, forget 100 symbols	
Success Rate:	91.37%
Fixpoints:	92.48%
Duration Mean:	57.87 sec.
Duration Median:	6.43 sec.
Duration 90th percentile:	99.26 sec.

$\mathcal{SHQ}$, forget 50 concept symbols	
Success Rate:	95.83%
Without Fixpoints:	93.40%
Duration Mean:	7.62 sec.
Duration Median:	1.04 sec.
Duration 90th percentile:	4.89 sec.

$\mathcal{SHQ}$, forget 100 concept symbols	
Timeouts:	90.77%
Fixpoints:	91.99%
Duration Mean:	13.51 sec.
Duration Median:	1.60 sec.
Duration 90th percentile:	11.65 sec.

Corpus Respective fragments of 306 ontologies from BioPortal having at most 100,000 axioms.

Timeout 30 minutes

Conclusion

- LETHE supports different non-classical reasoning methods via reduction to forgetting
- Usage as library, command line tool or via simple front end
- Available at `http://cs.man.ac.uk/~koopmanp/lethe`

- Future work
  - Better evaluation on abduction and logical difference
  - Use saturation-based approach for other non-classical reasoning problems such as approximation and ABox abduction
  - Investigate more expressive description logics