



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





Uniform Interpolant

$Margherita \sqsubseteq \forall topping.VegTopping$
American 📃 🗄 topping. Meat Topping





Applications of Forgetting

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





TBox Abduction

- Given TBox \mathcal{T} , axioms \mathcal{O} , find axioms H with $\mathcal{T} \cup H \models \mathcal{O}$
- "Complete" ontology such that given set of axioms is entailed
- Abducibles $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\Sigma}$
- Compute new entailments in \mathcal{O}_2 :
 - $LD(\mathcal{O}_1, \mathcal{O}_2, \Sigma) = \{ \alpha \mid \alpha \in \mathcal{O}_2^{\Sigma}, \mathcal{O}_1 \not\models \alpha \}$
 - $\mathcal{O}_1^{\Sigma} \colon$ 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

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

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





Normal form, ALC

 $\mathcal{ALC}\text{-}\mathsf{Clause}$

 $\top \sqsubseteq L_1 \sqcup \ldots \sqcup L_n$ $L_i: \mathcal{ALC}\text{-literal}$

 $\mathcal{ALC}\text{-Literal}$

 $A \mid \neg A \mid \exists r.D \mid \forall r.D$ A: any concept symbol, D: definer symbol

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





Definer symbols

Invariant: max 1 neg. definer symbol per clause

• Allows easy translation to clausal form and back:

$$C_1 \sqcup Qr.C_2 \qquad \Longleftrightarrow C_1 \sqcup Qr.D_1, \neg D_1 \sqcup C_2 C_1 \sqcup \nu X.C_2[X] \qquad \Longleftrightarrow C_1 \sqcup Qr.D_1, \neg D_1 \sqcup C_2[D]$$

- $\Rightarrow \text{ Any set of clauses can be converted into an } \mathcal{ALC}\mu\text{-ontology} \\ (\mathcal{ALC} \text{ with fixpoints})$
- New definer symbols introduced by calculus
 - Number finitely bounded

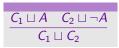




Calculus

Resolution + Combination rules

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



- 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



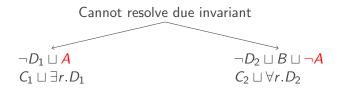


 $\neg D_1 \sqcup A$ $C_1 \sqcup \exists r.D_1$

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

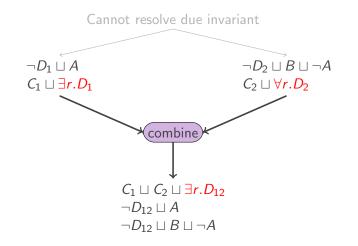






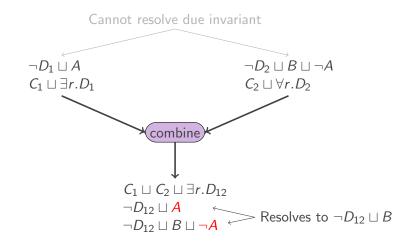








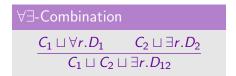








Combination Rules \mathcal{ALC}



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





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:

 $\frac{C_1\sqcup \ge n_1r_1.(D_1\sqcup \ldots \sqcup D_m) \qquad C_2\sqcup \le n_2r_2.\neg D_a \qquad r_1\sqsubseteq_{\mathcal{R}} r_2}{C_1\sqcup C_2\sqcup \ge (n_1-n_2)r_1.(D_{1a}\sqcup \ldots \sqcup D_{ma})}$

$\leq\geq$ -Combination:

 $\begin{array}{c} \underbrace{C_1 \sqcup \leq n_1 r_1 . \neg D_1 \quad C_2 \sqcup \geq n_2 r_2 . D_2 \quad r_2 \sqsubseteq_R r_1 \quad n_1 \geq n_2}_{C_1 \sqcup \sqcup C_2 \sqcup \leq (n_1 - n_2) r_1 . \neg (D_1 \sqcup D_2) \sqcup \geq 1 r_1 . D_{12}} \\ \vdots \\ C_1 \sqcup C_2 \sqcup \leq (n_1 - 1) r_1 . \neg (D_1 \sqcup D_2) \sqcup \geq n_2 r_1 . D_{12} \end{array}$

$\geq\geq$ -Combination:

Transitivity: $\frac{C \sqcup \leq 0_{f_1}, \neg D}{C \sqcup < 0_{f_2}, \neg D'} \xrightarrow{\text{trans}(r_2) \in \mathcal{R}} r_2 \subseteq_{\mathcal{R}} r_1}{\neg D' \sqcup Q}$





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

Duration Median: 2.74 sec. Duration Median:	88.10% 93.27% 18.03 sec. 3.81 sec. 21.17 sec.		
Without Fixpoints: 95.29% Without Fixpoints: Duration Mean: 7.68 sec. Duration Mean: Duration Median: 2.74 sec. Duration Median: Duration 90th percentile: 12.45 sec. Duration 90th percentile: ALC w. ABoxes, forget 50 symbols ALC w. ABoxes, forget 100	93.27% 18.03 sec. 3.81 sec.		
Duration Mean: 7.68 sec. Duration Median: Duration Median: 2.74 sec. Duration Median: Duration 90th percentile: 12.45 sec. Duration 90th percentile: ALC w. ABoxes, forget 50 symbols ALC w. ABoxes, forget 100	18.03 sec. 3.81 sec.		
Duration Median: 2.74 sec. Duration Median: Duration 90th percentile: 12.45 sec. Duration 90th percentile: ALC w. ABoxes, forget 50 symbols ALC w. ABoxes, forget 100	3.81 sec.		
Duration 90th percentile: 12.45 sec. Duration 90th percentile: ALC w. ABoxes, forget 50 symbols ALC w. ABoxes, forget 100			
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	21 17 sec		
Success Rate: 94.79% Success Rate:	symbols		
511570 Success rate.	91.37%		
Without Fixpoints: 92.91% Fixpoints:	92.48%		
Duration Mean: 23.94 sec. Duration Mean:	57.87 sec.		
Duration Median: 3.01 sec. Duration Median:	6.43 sec.		
Duration 90th percentile: 29.00 sec. Duration 90th percentile:	99.26 sec.		
SHQ, forget 50 concept symbols SHQ , forget 100 concept sy	SHQ, forget 100 concept symbols		
Success Rate: 95.83% Timeouts:	90.77%		
Without Fixpoints: 93.40% Fixpoints:	91.99%		
Duration Mean: 7.62 sec. Duration Mean:	13.51 sec.		
Duration Median: 1.04 sec. Duration Median:	1.60 sec.		
Duration 90th percentile: 4.89 sec. Duration 90th percentile:	1.00 360.		

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