Abstracting Temporal ABoxes in *TDL-Lite* (Extended Abstract)⋆ ⋆⋆

Sabiha Tahrat[1], German Braun[2], Alessandro Artale[3], and Ana Ozaki[4]

[1] Université de Paris, France sa.tahrat@gmail.com
[2] Universidad Nacional del Comahue, Argentina german.braun@fi.uncoma.edu.ar
[3] Free Univ. of Bolzano, Italy artale@inf.unibz.it
[4] University of Bergen, Norway ana.ozaki@uib.no

We propose an approach for abstracting temporal ABoxes over temporal *DL-Lite* knowledge bases (KBs) [2,3,1,4,12] to improve the reasoning performance when dealing with large ABoxes. We consider here the logic *TDL-Lite*($\mathbb{N}$), i.e., the fragment of the logic T_{FPX}*DL-Lite*$^{\mathcal{N}}_{bool}$ [3], combining *DL-Lite*$^{\mathcal{N}}_{bool}$ with *LTL*, but with just future temporal operators and thus interpreted over the natural numbers. Let $\mathsf{N_C}, \mathsf{N_I}, \mathsf{N_G}, \mathsf{N_L}$ be sets of *concept*, *individual*, *global role*, and *local role names*, respectively. The union $\mathsf{N_G} \cup \mathsf{N_L}$ is the set $\mathsf{N_R}$ of *role names*. *TDL-Lite roles* R, *basic concepts* B, and *(temporal) concepts* C are given by the following grammar where $L \in \mathsf{N_L}$, $G \in \mathsf{N_G}$, $A \in \mathsf{N_C}$, and $q \in \mathbb{N}$:

$$R ::= L \mid L^- \mid G \mid G^-, \qquad B ::= \bot \mid A \mid \geq qR,$$
$$C ::= B \mid \neg C \mid C_1 \sqcap C_2 \mid \Diamond_F C \mid \bigcirc_F C.$$

A TBox, $\mathcal{T}$, is a set of *general concept inclusions (GCI)* of the form $C_1 \sqsubseteq C_2$, where $C_1, C_2 \in \mathsf{N_C}$. An ABox. $\mathcal{A}$, is a set of *concept assertions* of the form $\bigcirc^n A(a)$ or $\bigcirc^n \neg A(a)$, or *role assertions* of the form $\bigcirc^n R(a,b)$ or $\bigcirc^n \neg R(a,b)$, where $a, b \in \mathsf{N_I}$, and $n \in \mathbb{N}$ (with, $\bigcirc^0 A(a) = A(a)$, $\bigcirc^{n+1} A(a) = \bigcirc \bigcirc^n A(a)$). A *TDL-Lite* KB is a pair, $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. The key idea is to map a *TDL-Lite* KB into an equisatisfiable *LTL* formula by applying the translation described in [3], and then abstracting large temporal ABoxes by adapting the technique presented in [9]. Given a *TDL-Lite* ABox we first shift all role assertions associated to global roles to time point 0, and call the resulting ABox $\mathcal{A}_G$. We then map $\mathcal{A}_G$ into $\mathcal{A}_G^\dagger$ a *first-order temporal formula with unary predicates*, $\mathcal{QTL}_1(\mathbb{N})$:

$$\mathcal{A}_G^\dagger = \varPhi_{\mathcal{A}_G} \quad \wedge \bigwedge_{R \in \mathsf{role}_{\mathcal{T}}, \bigcirc^n R(a,b) \in \mathcal{A}_G} \bigcirc^n (E_q\, R)(a), \tag{1}$$

where $\mathsf{role}_{\mathcal{T}}$ is the set of (global and local) role names occurring in $\mathcal{T}$ including their inverses, $\varPhi_{\mathcal{A}_G}$ is the conjunction of all concept assertions in $\mathcal{A}_G$, and the unary predicates $E_q\, R(x)$, $E_q\, R^-(x)$, for $R \in \mathsf{N_R}$, capture domain and range of role R, respectively, with q encoding the number of R-successors of a given individual. We abstract individuals according to their types. For each individual $a \in \mathcal{A}_G^\dagger$, its type $\tau(a) = \{\bigcirc^n P \mid \bigcirc^n P(a) \in \mathcal{A}_G^\dagger\}$. With each type, τ, we associate a fresh new individual, v_τ, the representative of all the individuals in $\mathcal{A}_G^\dagger$ with type τ, and denote with $\varGamma_{\mathcal{A}_G^\dagger}$ the set of all such representatives. Thus, by using

⋆ Extended version published as arXiv:2008.07463

TDL-Lite*($\mathbb{N}$) *ABox				$\mathcal{QTL}_1(\mathbb{N})$ ***ABox***			**Gain**	
ABox (Sample)	**# Concept Assert.**	**# Role Assert.**	**Global Roles Gain**	**ABox size**	**Abs. ABox size**	**Abs. Ind.**	**Abs. ABox**	**ABox size**
3 677 052 (7 027 000)	23 481	3 653 571	1 346 369	80 881	80 881	100	**0%**	**97,8%**
4 241 783 (10 500 000)	24 617	4 217 166	1 621 779	82 017	80 369	98	**2%**	**98,1%**
4 587 504 (15 500 000)	24 955	4 562 549	1 790 413	82 355	27 150	33	**67,03%**	**99,4%**

Table 1: Abstraction results for randomly generated ABoxes with number of individuals, $I = 100$, time span, $T = 5$ and $N = 50$. Each row represents ABox sizes ranging from 70% up to 90% of the whole ABox space (i.e., 5 025 000 distinct possible assertions). The first column corresponds to the number of *distinct* assertions generated from a sample size counting repetitions written below in parentheses. *Global Roles Gain* captures the number of gained assertions due to the shift of global role assertions to time point 0. The $\mathcal{QTL}_1(\mathbb{N})$ ABox columns include the size of the resulting $\mathcal{QTL}_1(\mathbb{N})$ ABox, the size of the abstracted ABox and the number of abstracted individuals. The last two columns report the percentages of the abstraction gain, and the global gain considering both the translation and the abstraction of the ABox.

just one representative for each type, we finally generate the (equisatisfiable) abstraction $\mathcal{B}_G$ of $\mathcal{A}_G^\dagger$:

$$\mathcal{B}_G = \bigcup_{v_\tau \in \Gamma_{\mathcal{A}_G^\dagger}} \mathcal{B}_\tau, \text{ with } \mathcal{B}_\tau = \{\bigcirc^n P(v_\tau) \mid \tau = \tau(a), \bigcirc^n P \in \tau\} \tag{2}$$

Experiments on Randomly Generated ABoxes. To evaluate the gain in size of the ABox after the abstraction, we randomly generate temporal ABoxes from a space of I individual names, N concept names, N global and N local role names, while the time point n is chosen within the time interval $[0, T-1]$ in accordance with the uniform distribution over the space of $N \cdot T \cdot I \cdot (1 + 2I)$ possible assertions. Table 1 reports abstraction results on sampled ABoxes with size varying from 70% up to 90% over the whole ABox space.

We observe a gain of more than 95% in the number of $\mathcal{QTL}_1(\mathbb{N})$ ABoxes assertions compared to the original *TDL-Lite*($\mathbb{N}$) ABoxes. This is due to the translation of roles into pairs of unary predicates, together with the shifting of global role assertions to the initial time point. Moreover, there is a clear gain in the number of individuals, specially when the ABox size is close to the ABox space, confirming experimental results obtained for non-temporal DLs [6,8,9].

Experiments on Randomly Generated TBoxes and ABoxes. To evaluate the runtime efficiency of reasoning over KBs with an abstracted ABox, we pair a randomly generated *TDL-Lite*($\mathbb{N}$) TBox with each of the ABoxes in Table 1. To generate random TBoxes, we extend the test method proposed by [5] in the context of propositional temporal logic to our case. We guess a TBox given a fixed number of N concept names, N global role names, and N local role names; a fixed maximum Q for the value q in basic concepts of the form $\geq qR$; a fixed

***TDL-Lite*(ℕ) ABox**		***LTL***					`BLACK`	
Space	**ABox (Sample)**	**# abs. prop.**	**# abs. ind.**	**tr. time**	**abs. time**	**abs tr. time**	**SAT time** *(runtime)*	**abs. SAT time** *(runtime)*
Prop.Var. 110 150	3 675 455 (7 027 000)	110 150	100	484,01	62,02	546,03	546,2 (1 030,21)	546,2 (1 092,23)
	4 241 783 (10 500 000)	109 334	98	400,08	56,27	457,38	541,12 (941,2)	546,81 (1 004,19)
ABox Space 5 025 000	4 588 504 (15 500 000)	82 814	33	432,68	3,1	79,65	545,35 (978,03)	273,3 (352,95)

Table 2: Reasoning results over KBs randomly generated with $I = 100$, $T = 5$ and $N = 50$, and with abstracted ABoxes. Each row represents ABox sizes ranging from 70% up to 90% of the whole ABox space. *LTL* columns include the resulting number of propositional variables after the abstraction, the size of the abstracted individuals, the translation time (from the original *TDL-Lite*(ℕ) KB to *LTL*), the time spent to calculate the abstraction, and the translation time on the abstracted KB. The last two columns show the runtime efficiency by comparing the runtime of `BLACK` on KBs without or with ABox abstraction. Time is in seconds and 'runtime' includes translation.

number L_t of GCIs in a TBox; and a fixed value L_c of the length of concept. The guessed TBox for Table 2 has the following parameters: $N = 50$, $L_c = 5$, $Q = 3$, while the value of L_t is set close to $3 \cdot N$ so that all concept and role names are considered in guessing the TBox (when $L_c = 5$ three names are at most chosen from a space of $3 \cdot N$ and we have at most $2 \cdot 3 = 6$ names considering both the right and left hand sides of each GCI). The resulting KB is translated into *LTL* resulting in a formula with 110 150 propositional variables (this number depends only on the size of the TBox and on the number of individuals). Table 2 reports reasoning results on such abstracted KBs using the tableaux based *LTL* solver `BLACK` [7]. We can observe that there is a notable gain in the number of propositional variables after the abstraction (*# abs. prop.*), in particular when the gain on individuals is high. The total runtime, indicated in parentheses, and including both the translation and the satisfiability checking time, is around 10 minutes for each satisfiability check. The abstraction improves sensibly the reasoner performances. The translation time for processing the largest instance was reduced from 432 to 79 seconds due to the 67% gain on individuals, while the satisfiability checking time reduced from 545 to 273 seconds. For more details, see the full version of the paper [11].

Conclusions. We proposed an approach for abstracting temporal ABoxes which shows good results when increasing the number of ABox assertions in randomly generated KB to values close to the ones observed in realistic (static) ontologies, being able to process temporal KBs with millions of assertions. We confirmed the results of this approach by comparing the runtimes of reasoning with ABoxes vs abstracted ABoxes by showing significant savings. As a future work, one can consider the feasibility of pairing the abstraction technique with parallelised reasoning algorithms, splitting the parallel processes according to the abstracted individuals, in the lines of what has already been experimented in [10].

References

1. A. Artale, R. Kontchakov, A. Kovtunova, V. Ryzhikov, F. Wolter, and M. Zakharyaschev. Temporal ontology-mediated querying: A survey. In *24th International Symposium on Temporal Representation and Reasoning (TIME17)*, Mons, Belgium, October 2017.
2. A. Artale, R. Kontchakov, A. Kovtunova, F. Wolter, and M. Zakharyaschev. First-order rewritability of temporal ontology-mediated queries. In *Proc. of the 24th Int. Joint Conference on Artificial Intelligence (IJCAI-15)*, Buenos Aires, Argentina, 25-31 July, 2015. AAAI Press.
3. A. Artale, R. Kontchakov, V. Ryzhikov, and M. Zakharyaschev. A cookbook for temporal conceptual data modelling with description logics. *ACM Trans. Comput. Log.*, 15(3):25:1–25:50, 2014.
4. A. Artale, A. Mazzullo, and A. Ozaki. Temporal DL-Lite over finite traces (preliminary results). In *32nd Int. Workshop on Description Logics, (DL'19)*, Oslo, Norway, 18-21 June, 2019.
5. M. Daniele, F. Giunchiglia, and M. Y. Vardi. Improved automata generation for linear temporal logic. In *CAV*, 1999.
6. A. Fokoue, A. Kershenbaum, L. Ma, E. Schonberg, and K. Srinivas. The summary ABox: Cutting ontologies down to size. In *Proc. of the 5th Int. Semantic Web Conference, ISWC-06*, volume 4273 of *Lecture Notes in Computer Science*, pages 343–356. Springer, 2006.
7. L. Geatti, N. Gigante, and A. Montanari. A sat-based encoding of the one-pass and tree-shaped tableau system for LTL. In *Automated Reasoning with Analytic Tableaux and Related Methods - 28th International Conference, TABLEAUX 2019, Proceedings*, volume 11714 of *Lecture Notes in Computer Science*, pages 3–20. Springer, 2019.
8. B. Glimm, Y. Kazakov, T. Liebig, T. Tran, and V. Vialard. Abstraction refinement for ontology materialization. In *Proceedings of the International Semantic Web Conference (ISWC'14)*, volume 8797 of *Lecture Notes in Computer Science*. Springer, 2014.
9. B. Glimm, Y. Kazakov, and T. Tran. Scalable reasoning by abstraction beyond DL-Lite. In M. Ortiz and S. Schlobach, editors, *Web Reasoning and Rule Systems (RR) - Proceedings of the 10th International Conference*, volume 9898 of *Lecture Notes in Computer Science*, pages 77–93. Springer, 2016.
10. A. Steigmiller and B. Glimm. Parallelised ABox reasoning and query answering with expressive description logics (extended abstract). In S. Borgwardt and T. Meyer, editors, *Proc. of the 33rd Int. Workshop on Description Logics (DL 2020)*, volume 2663 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.
11. S. Tahrat, G. Braun, A. Artale, M. Gario, and A. Ozaki. Automated reasoning in temporal DLite. *arXiv*, arXiv:2008.07463, 2020.
12. S. Tahrat, G. A. Braun, A. Artale, M. Gario, and A. Ozaki. Automated reasoning in temporal DL-Lite (extended abstract). In *Proceedings of the 33rd International Workshop on Description Logics (DL 2020)*, volume 2663 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.