On Limited Conjunctions in Polynomial Feature Logics, with Applications in OBDA — *Extended Abstract*

Stephanie McIntyre
Cheriton School of CS
University of Waterloo, Canada
srmcinty@uwaterloo.ca

Alexander Borgida
Department of CS
Rutgers University, NJ, U.S.A.
borgida@cs.rutgers.edu

David Toman, Grant Weddell
Cheriton School of CS
University of Waterloo, Canada
{david,gweddell}@uwaterloo.ca

Abstract

Standard reasoning problems are complete for EXPTIME in common feature-based description logics—ones in which all roles are restricted to being functional. We show how to control conjunctions on left-hand-sides of subsumptions in such a way so as to ensure polynomial time complexity. In particular, we present a PTIME algorithm for reasoning about knowledge base consistency. We then show how the resulting description logic allows features to be partial, not just total functions. Algorithms for polynomial-time query answering are presented. The above, in combination with referring expressions, provide a richer capability for ontology-based data access to relational data sources.

Introduction

The $\mathcal{CFD}$ family of feature-based DLs has been designed primarily to support PTIME reasoning in accessing relational data sources. A distinguishing feature of this family is support for expressing complex functional dependencies, which are the most widely used way to capture domain semantics in relational databases, in addition to foreign keys. One dialect, called $\mathcal{CFDI}_{nc}^{\forall-}$ (St. Jacques, Toman, and Weddell 2016), supported OBDA to relational databases, and was able to do so without "loading" the relational database into an ABox. In addition, it was capable of emulating DL-Lite$_{core}^{\mathcal{F}}$.

Our main contribution is a new member of this family called $\mathcal{CFDI}_{kc}^{\forall-}$, which starts by adding the ability for limited use of conjunctions on the left-hand side of subsumptions in $\mathcal{CFDI}_{nc}^{\forall-}$. Further contributions outline how this new capability can be utilized to define a variant of $\mathcal{CFDI}_{kc}^{\forall-}$ that supports *partial features* and subsequently how OBDA-style query answering in this setting can be efficiently implemented.

This clearly enhances the modeling capacity of $\mathcal{CFDI}_{kc}^{\forall-}$, since in a university context, we can now not only specify that StudentWorkers are both Students and Employees, but actually *define* StudentWorker as anyone who is both, by adding the axiom $(Student \sqcap Employee) \sqsubseteq StudentWorker$. This distinction between "primitive concepts" ("phones, which happen to all be black") and "defined

concepts" ("black phones") was one of the key insights that drove Brachman (Brachman 1977) to the development of KL-ONE, the progenitor of Description Logics, and was an important missing ingredient in Semantic Data Models (Hull and King 1987), such as Taxis (Mylopoulos, Bernstein, and Wong 1980) and GEM (Zaniolo 1983), as well as UML.

Main Result

The first and main contribution of this work is showing a parametric tractability bound for reasoning in $\mathcal{CFDI}_{kc}^{\forall-}$.

Definition 1 ($\mathcal{CFDI}^{\forall-}$) *A $\mathcal{CFDI}^{\forall-}$ KB consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. $\mathcal{T}$ consists of subsumption constraints of the form $C \sqsubseteq D$, where the structure of concepts C and D are given by the following respective grammars:*

$$
\begin{aligned}
C &::= A \mid \forall f.A \mid A_1 \sqcap A_2 \\
D &::= A \mid \bot \mid \forall f.A \mid \exists f^{-1} \mid A : \mathsf{Pf}_1, \ldots, \mathsf{Pf}_k \to \mathsf{Pf}
\end{aligned}
$$

where all concepts D conforming to the last concept constructor, called a path functional dependency *(PFD), adhere to additional restrictions given in (Toman and Weddell 2014). ABox $\mathcal{A}$ consists of assertions of the form "$A(a)$", "$a.f = b$", and "$a = b$".* □

KB satisfiability in the above is complete for EXPTIME (Toman and Weddell 2005). The following restriction confines the *exponentiality* of reasoning via a parameter k:

Definition 2 (Restricted Conjunction) *Let $k > 0$ be a constant. We say that TBox $\mathcal{T}$ is a $\mathcal{CFDI}_{kc}^{\forall-}$ TBox if, whenever $\mathcal{T} \models (A_1 \sqcap \cdots \sqcap A_n) \sqsubseteq B$ for some set of primitive concepts $\{A_1, \ldots, A_n\} \cup \{B\}$, with $n > k$, then $\mathcal{T} \models (A_{i_1} \sqcap \cdots \sqcap A_{i_k}) \sqsubseteq B$ for some k-sized subset $\{A_{i_1}, \ldots, A_{i_k}\}$ of the primitive concepts $\{A_1, \ldots, A_n\}$.* □

To facilitate reasoning over a $\mathcal{CFDI}_{kc}^{\forall-}$ TBox $\mathcal{T}$, we define $\mathsf{Clos}(\mathcal{T})$ to be a set of all "small" subsumptions entailed by $\mathcal{T}$, subsumptions of the form $\hat{E} \sqsubseteq \hat{F}$ where $\hat{E}$ and $\hat{F}$ are sets of primitive concepts, including $\bot$, of size at most k, or sets of value restrictions involving a common feature (f) over such sets of concepts of the form $\forall f.\hat{E}$. Then:

Theorem 3 (Primitive Concept Satisfiability) *Let $\mathcal{T}$ be a $\mathcal{CFDI}_{kc}^{\forall-}$ TBox in normal form and A be a primitive concept. Then A is satisfiable with respect to $\mathcal{T}$ if and only if $A \sqsubseteq \bot \notin \mathsf{Clos}(\mathcal{T})$.* □

We show that $\mathsf{Clos}(\mathcal{T})$ can be constructed from $\mathcal{T}$ in time polynomial in $|\mathcal{T}|$ and exponential in k. The above theorem transfers the complexity bound to reasoning in $\mathcal{CFDI}_{kc}^{\forall-}$. To extend this result to KB satisfiability, we need the notion of an *ABox closure*, $\mathsf{Completion}_{\mathcal{T}}(\mathcal{A})$, from (St. Jacques, Toman, and Weddell 2016; McIntyre et al. 2018) that fully saturates the ABox with respect to concept membership and PFD induced equalities. Theorem 3 is then applied to individuals in $\mathsf{Completion}_{\mathcal{T}}(\mathcal{A})$ to determine KB satisfiability:

Theorem 4 ($\mathcal{CFDI}_{kc}^{\forall-}$ KB Consistency) *Let* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ *be a* $\mathcal{CFDI}_{kc}^{\forall-}$ *KB. Then* $\mathcal{K}$ *is consistent if and only if* $\{A \mid A(a) \in \mathsf{Completion}_{\mathcal{T}}(\mathcal{A})\}$ *is satisfiable with respect to* $\mathcal{T}$ *for every "a" appearing in* $\mathsf{Completion}_{\mathcal{T}}(\mathcal{A})$. $\qquad\square$

Further standard reasoning tasks are reduced to KB satisfiability in the usual way.

On determining k: a pay as you go approach. The above development assumes that the value of k is known in advance. However, the above TBox closure procedure also allows for testing if a particular value of k is sufficient for a $\mathcal{CFDI}^{\forall-}$ TBox $\mathcal{T}$:

Theorem 5 (Test for k) *Let* $\mathcal{T}$ *be a* $\mathcal{CFDI}^{\forall-}$ *TBox. Then* $\mathcal{T}$ *is* not *a* $\mathcal{CFDI}_{kc}^{\forall-}$ *TBox if and only if there are* $\hat{E}, \hat{F}, \hat{G}, D$ *such that (1)* $\hat{E} \sqsubseteq \hat{F} \in \mathsf{Clos}(\mathcal{T})$, *(2)* $\hat{G} \sqsubseteq D \in \mathsf{Clos}(\mathcal{T})$, *(3)* $\hat{F} \subseteq \hat{G}$, *(4)* $|\hat{E} \cup (\hat{G} - \hat{F})| > k$, *and (5) for all* $\hat{H} \sqsubseteq D \in \mathsf{Clos}(\mathcal{T})$ *we have* $\hat{H} \not\subset \hat{E} \cup (\hat{G} - \hat{F})$. $\qquad\square$

This idea can be used in an *iterative-deepening*-style algorithm to determine the value of k for a given TBox $\mathcal{T}$ in time exponential in k but polynomial in $|\mathcal{T}|$.

Further Results

We define *partial-$\mathcal{CFDI}_{kc}^{\forall-}$*, a variant of $\mathcal{CFDI}_{kc}^{\forall-}$ with *partial features*. This entails changing the interpretation of features and path functions to be *partial* and adding an additional concept constructor, $\exists f$, interpreted as the *set of objects for which f is defined*, to the syntax of subsumptions. We show that reasoning in *partial-$\mathcal{CFDI}_{kc}^{\forall-}$* can be reduced to our main result as follows:

Theorem 6 *Let* $\mathcal{K}$ *be a partial-$\mathcal{CFDI}_{kc}^{\forall-}$ KB. Then there is a* $\mathcal{CFDI}_{(k+1)c}^{\forall-}$ *KB* $\mathcal{K}'$ *that can be constructed from* $\mathcal{K}$ *in linear time such that* $\mathcal{K}$ *is consistent if and only if* $\mathcal{K}'$ *is consistent.* $\qquad\square$

This theorem shows that partiality in the $\mathcal{CFDI}^{\forall-}$ family can be accommodated at the price of increasing the size of conjunctions (and consequently at the cost of increasing the leading coefficient of the polynomial upper-bound) by one.

To define OBDA-style query answering for conjunctive queries we need to rewrite the queries to account for anonymous objects entailed by a TBox. Note that unlike (Lutz, Toman, and Wolter 2009; Kontchakov et al. 2010) these *cannot* be part of any poly-sized ABox completion. We define $\mathsf{Fold}_{\mathcal{T}}(Q)$, a rewriting of the query to compensate for the anonymous objects. However, unlike (Calvanese et al. 2007), we no longer need to rewrite the query with respect to concept hierarchies, yielding much smaller rewritings (for details see (McIntyre et al. 2018)). Alltogether we have:

Theorem 7 *Let* Q *be a CQ and* $\mathcal{K}$ *a partial-$\mathcal{CFDI}_{kc}^{\forall-}$ KB. Then* $\bar{a}$ *is a certain answer to* Q *over* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ *if and only if* $\bar{a}$ *is an answer for some* $\{\bar{x} \mid \psi\} \in \mathsf{Fold}_{\mathcal{T}}(Q)$ *over* $\mathsf{Completion}_{\mathcal{T}}(\mathcal{A})$.

In addition, we have combined the ABox closure with the idea of *referring expressions* (Borgida, Toman, and Weddell 2016) to facilitate access to relational back-ends. The use of referring expressions sidesteps the need for inventing opaque Skolem constants/functions commonly used to describe objects implied by the knowledge base, but not present explicitly in the data sources.

Future research will consider a range of topics from enhancements to the expressiveness of $\mathcal{CFDI}_{kc}^{\forall-}$ to implementation issues, in particular to optimizing the ABox completion $\mathsf{Completion}_{\mathcal{T}}(\mathcal{A})$ in the presence of constraints enforced in the original data sources.

References

Borgida, A.; Toman, D.; and Weddell, G. 2016. On referring expressions in query answering over first order knowledge bases. In *Proc. KR*, 319–328.

Brachman, R. J. 1977. What's in a concept: structural foundations for semantic networks. *International journal of man-machine studies* 9(2):127–152.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. Autom. Reasoning* 39(3):385–429.

Hull, R., and King, R. 1987. Semantic database modeling: Survey, applications, and research issues. *ACM Comput. Surv.* 19(3):201–260.

Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyaschev, M. 2010. The combined approach to query answering in DL-Lite. In *Proc. KR*, 247–257.

Lutz, C.; Toman, D.; and Wolter, F. 2009. Conjunctive query answering in the description logic EL using a relational database system. In *Proc. IJCAI*, 2070–2075.

McIntyre, S.; Borgida, A.; Toman, D.; and Weddell, G. 2018. On limited conjunctions and partial features in parameter-tractable feature logics. (submitted).

Mylopoulos, J.; Bernstein, P. A.; and Wong, H. K. T. 1980. A language facility for designing database-intensive applications. *ACM Trans. Database Syst.* 5(2):185–207.

St. Jacques, J.; Toman, D.; and Weddell, G. E. 2016. Object-relational queries over $\mathcal{CFDI}_{nc}$ knowledge bases: OBDA for the SQL-Literate. In *Proc. IJCAI*, 1258–1264.

Toman, D., and Weddell, G. E. 2005. On reasoning about structural equality in XML: a description logic approach. *Theor. Comput. Sci.* 336(1):181–203.

Toman, D., and Weddell, G. E. 2014. On adding inverse features to the description logic $\mathcal{CFD}_{nc}^{\forall}$. In *PRICAI 2014*, 587–599.

Zaniolo, C. 1983. The database language GEM. In *ACM SIGMOD Int. Conf. on Management of Data*, 207–218.