Model Comparison Games for Horn Description Logics: A Summary*

Jean C. Jung[1], Fabio Papacchini[2], Frank Wolter[2], and Michael Zakharyaschev[3]

[1] Universität Bremen, Germany, jeanjung@uni-bremen.de
[2] University of Liverpool, UK, {papacchf,wolter}@liverpool.ac.uk
[3] Birkbeck, University of London, UK, michael@dcs.bbk.ac.uk

Horn DLs have been introduced as syntactically defined fragments of standard DLs that fall within the Horn fragment of first-order logic (henceforth Horn FO) and for which ontology-mediated query answering is in PTime for data complexity [7, 8]. One possible way of defining the concepts C of the basic fragment *horn*$\mathcal{ALC}$ of $\mathcal{ALC}$ is by the following rule:

$$C, C' ::= \bot \mid \top \mid A \mid C \sqcap C' \mid L \to C \mid \exists r.C \mid \forall r.C,$$

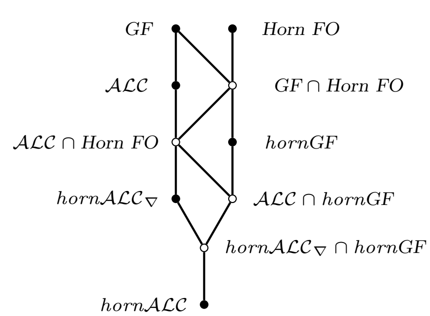
where A ranges over concept names and L over $\mathcal{EL}$ concepts. The modal logic corresponding to *horn*$\mathcal{ALC}$ was introduced independently with the aim of capturing the intersection of Horn FO and modal logic [11]. In this note we summarize the results obtained in [9], which aims to lay the foundations for a model-theoretic understanding of *horn*$\mathcal{ALC}$ and also the Horn fragment of the guarded fragment of FO by introducing model comparison games that characterize their expressive power. In a first application of these results, it is shown that concept learning and model indistinguishability in *horn*$\mathcal{ALC}$ are ExpTime-complete and that *horn*$\mathcal{ALC}$ does not capture the intersection of $\mathcal{ALC}$ and Horn FO.

A *Horn simulation game* is a two player game on the disjoint union of two interpretations that differs from the standard bisimulation or Ehrenfeucht-Fraïssé game in the following respects: (1) positions in the game consist of pairs (X, b) with a *set* X of nodes and a node b (which reflects that Horn languages are not closed under disjunction); (2) a Horn simulation game uses as a subgame the basic simulation game for checking indistinguishability by $\mathcal{EL}$. Both conditions have important consequences. The latter means that Horn simulation games are *modular* as far as the characterization of the left-hand side of implications is concerned. For example, this modularity is used to also characterize a proper extension, *horn*$\mathcal{ALC}_\nabla$, of *horn*$\mathcal{ALC}$ with the operators $\nabla R.C = \exists R.\top \sqcap \forall R.C$ (or $\nabla p = \Diamond\top \wedge \Box p$ in modal logic) on the left-hand side of *horn*$\mathcal{ALC}$ implications, which also lies in Horn FO. The consequences of (1) are three-fold. First, using sets rather than nodes in positions implies that the obvious algorithm checking the existence of Horn simulations containing a pair $(\{a\}, b)$ of nodes runs in exponential time. Thus, using Horn simulation games to check whether two nodes a and b satisfy the same *horn*$\mathcal{ALC}$-concepts or whether two models satisfy the same TBox axioms yields exponential time algorithms. We show that this is unavoidable by proving corresponding ExpTime lower bounds. Second, as player 2 does *not* have a winning strategy in position (X, b) in the Horn simulation game if, and only if, there exists a *horn*$\mathcal{ALC}$-concept that is true at all nodes in X

* F. Papacchini was supported by the EPSRC UK projects EP/R026084 and EP/R026173

but not true at b, our complexity results are directly applicable to the *concept learning by example* (*CBE*) *problem*: given a data set, and sets P and N of positive and negative examples, does there exist a *horn$\mathcal{ALC}$*-concept C separating P from N over the data? The goal of this supervised learning problem is to automatically derive new concept descriptions from labelled data. It has been investigated before in DL [10, 2, 4] and for many logical languages, in particular in databases [3, 1]. Horn DLs are of particular interest as target languages for CBE as they can be regarded as 'maximal DLs without disjunction,' and the unlimited use of disjunction in derived concept descriptions is undesirable as it leads to *overfitting*: learnt concepts enumerate the positive examples rather than generalize from the examples. The complexity analysis for Horn simulation games shows that the CBE problem for *horn$\mathcal{ALC}$* is ExpTime-complete. Finally, the presence of sets in positions of the Horn simulation games has an impact on the standard infinitary saturated model approach to proving van Benthem style expressive completeness results [5, 6]. As subinterpretations of saturated interpretations are not always saturated, it seems that the only way to obtain expressive completeness results with an infinitary approach is to restrict the moves of players to 'saturated sets,' say sets definable as the intersection of FO-definable sets. Thus, in this paper, we prove van Benthem style expressive completeness results for *horn$\mathcal{ALC}$*-concepts and TBoxes via Horn simulation games by developing appropriate *finitary* methods which do not require saturated structures. In fact, our results hold both in the classical and the finite model theory setting, and without any restrictions on the moves of players.

In the second part of the paper, we consider the guarded fragment, GF, of FO and introduce its Horn fragment *hornGF*. *hornGF* captures *horn$\mathcal{ALC}$* and many popular extensions thereof, such as inverse roles or the universal role. We generalize Horn simulations to guarded Horn simulations, and show an Ehrenfeucht-Fraïssé type definability result for *hornGF*. This result is used to prove an ExpTime upper bound for model indistinguishability in *hornGF*. We also show that *hornGF* captures more of the intersection of $\mathcal{ALC}$ and Horn FO than *horn$\mathcal{ALC}$* but does not capture the intersection of GF and Horn FO. Finally, we show expressive completeness of *hornGF*: an FO-formula is equivalent to a *hornGF*-formula just in case it is preserved under guarded Horn simulations. Our proof uses infinitary methods and thus the moves of player 1 are restricted to intersections of FO-definable sets. It remains open whether the expressive completeness holds without this restriction and whether it holds in the finite model theory setting. The emerging landscape of the fragments of Horn FO and GF is depicted in the right lattice of languages and their intersections (modulo equivalence) where all inclusions are proper.



References

1. Arenas, M., Diaz, G.I.: The exact complexity of the first-order logic definability problem. ACM Trans. Database Syst. 41(2), 13:1–13:14 (2016)
2. Badea, L., Nienhuys-Cheng, S.: A refinement operator for description logics. In: Proceedings of ILP. pp. 40–59 (2000)
3. Barceló, P., Romero, M.: The complexity of reverse engineering problems for conjunctive queries. In: Proceedings of ICDT. pp. 7:1–7:17 (2017)
4. Funk, M., Jung, J.C., Lutz, C., Pulcini, H., Wolter, F.: Learning description logic concepts: When can positive and negative examples be separated? In: Proceedings of IJCAI (2019)
5. Goranko, V., Otto, M.: Model theory of modal logic. In: Handbook of Modal Logic, pp. 249–329. Elsevier (2007)
6. Grädel, E., Otto, M.: The freedoms of (guarded) bisimulation. In: Johan van Benthem on Logic and Information Dynamics, pp. 3–31. Springer (2014)
7. Hustadt, U., Motik, B., Sattler, U.: Data complexity of reasoning in very expressive description logics. In: Proceedings of IJCAI. pp. 466–471 (2005)
8. Hustadt, U., Motik, B., Sattler, U.: Reasoning in description logics by a reduction to disjunctive datalog. J. Autom. Reasoning 39(3), 351–384 (2007)
9. Jung, J.C., Papacchini, F., Wolter, F., Zakharyaschev, M.: Model comparison games for Horn description logics. In: Proceedings of LICS (2019)
10. Lehmann, J.: Dl-learner: Learning concepts in description logics. Journal of Machine Learning Research 10, 2639–2642 (2009)
11. Sturm, H.: Modal horn classes. Studia Logica 64(3), 301–313 (2000)