Master Thesis Opponent Report

Faculty of Mathematics and Physics, Charles University

Thesis author Jelena Glišić
Thesis title Non-standard representations of Boolean functions for knowledge compoilation
Submitted 2024
Program Computer Science **Specialization** Theoretical Computer Science

Review author RNDr. Petr Kučera, Ph.D. **Role** reviewer
Position Department of Theoretical Computer Science and Mathematical Logic

Review text:

The thesis studies a specific type of representation of boolean functions which was introduced by Hajnal, Liu, and Turán (2022) and which represents a boolean function $f(x_1, \ldots, x_n)$ as a pair of sets of prototypes (P, N). Each prototype is a boolean vector of length n and if we wish to evaluate the function f on a particular assignment $\mathbf{a}$, we look for a closest prototype measured by Hamming distance, if it belongs to P, then $f(\mathbf{a}) = 1$, if the prototype belongs to N, then $f(\mathbf{a}) = 0$, ties are not allowed. This representation is called BNN (Boolean Nearest Neighbor). A more general NN (Nearest Neighbor) representation differs from BNN by allowing real vectors as prototypes and measuring the distance as Euclidean.

The main interest of the thesis is in studying the BNN representation from the point of view of knowledge compilation. In particular, the main goal is to study the complexity of answering queries and performing transformation considered in the knowledge compilation map (Darwiche and Marquis, 2002), in addition, the thesis aims at positioning BNN among some of the other languages considered in knowledge compilation with respect to succinctness.

The thesis contains some new and nontrivial results regarding BNN. It turns out that BNN has some unusual properties which is demonstrated by the following two results: conditioning can lead to an exponential blowup, but on the other hand BNN still allows polynomial time clausal entailment. From other transformations, BNN supports negation and does not support forgetting. From queries, BNN support consistency and validity checking, clausal entailment and implicant checking. The complexity of the remaining queries and transformations was left unresolved in the thesis. With respect to succinctness, it was shown that a BNN representation can be used to construct a binary decision diagram (unrestricted in any way) in polynomial time. It was shown that BNN is thus strictly less succinct than BDDs, strictly more succinct than a list of models and incomparable with DNF, CNF, and a list of prime implicates or implicants.

The thesis is generally well written, the proofs are detailed and easy to follow. On the other hand, the presentation could be improved at some places which I note below.

Altogether, although the goal of the thesis may have not been fully reached since it leaves several open questions, the thesis contains enough theoretical results so that I can recommend it for defense. That said, I also have some remarks and notes to the thesis.

First of all, I was missing a more general introduction into knowledge compilation. The introduction is very short with no citations. Chapter 3 is a bit more detailed, but still, there is no general description of knowledge compilation giving its motivation and possibly cite some of its applications. Chapter 3.1 introduces concepts and languages derived from negation normal forms. However, it is just a list of definitions without giving any hint why they are included in the thesis. For instance, why smoothness is defined here? Why is it good to have decomposability or determinism in a NNF? The thesis also includes the definitions of languages such as PBC, CARD and SL, although these languages are not used in the thesis and BNN is not compared to them in any way. I think it would be enough to mention them with a citation.

I also think that there were some low hanging fruit to be picked up, such as the fact that clausal entailment implies model enumeration, or that the formula ϕ from the proof of Lemma 17 can be used to show that BNN does not admit $\bigwedge\mathbf{C}$ or $\bigvee\mathbf{C}$.

A few minor notes:

- The lines for d-DNNF in tables in figures 3.4 and 3.5 are incorrect, there are also mistakes at the FBDD line in Figure 3.4.
- The shortcuts for queries and transformations should be explained (i.e. mention that **CO** means "consistency" etc.).
- Figure 3.2 would be more illustrative, if the BDD would not be an OBDD or FBDD (i.e. some variable would actually repeat on some path and the order of variables would not be fixed).
- It would be nice to have BNN included in Figure 3.3 and the tables in Figures 3.4 and 3.5. It would make the presentation of the results in the thesis clearer. The meaning of the rectangles in around PBD+CARD and SL+$SL_<$ is also unclear.
- "However, since there are functions with exponentially many prime implicates but a small CNF, …" at page 13 — a citation or an example should be included.
- Proposition 22 could be easier shown with the parity function.

I recommend the thesis for defense.

I suggest to not consider the thesis for the annual award.

August 16, 2024

Signature:

References

Darwiche, Adnan and Pierre Marquis (2002). "A Knowledge Compilation Map". In: *Journal of Artificial Intelligence Research* 17, pp. 229–264.

Hajnal, Péter, Zhihao Liu, and György Turán (2022). "Nearest neighbor representations of Boolean functions". In: *Information and Computation* 285, p. 104879. ISSN: 0890-5401. DOI: `https://doi.org/10.1016/j.ic.2022.104879`.