REFEREE'S REPORT ON BACHELOR'S THESIS

Název: On the lattice of multi-sorted relational clones on a two-element domain
Autor: Vojtěch David

Summary of the content

The thesis re-proves Post's result that classifies Boolean clones, it gives an alternative proof for Taimanov's result showing that there are only countably many multi-sorted Boolean clones, and derives several novel results toward classifying these clones.

Evaluation

Topic. The topic is beyond the curriculum of the bachelor level. The goal was to provide proofs of the above results using a relational approach developed by Zhuk. This goal was clearly achieved.

Original contribution. Taiman's proof appeared in print only recently in a series of papers. It is complex and his approach is mainly via operations. The approach in the thesis is substantially different, mostly via relations. It seems that the presented proof is largely (or fully?) independent and therefore the majority of the thesis is an original contribution of the student.

Mathematical level. The results and proofs are very complex. This is apparent from the fact that Post's result can be seen as the first induction step (and, indeed, the thesis has about 70 dense pages). It is clear from the thesis that Vojtěch David has mastered the techniques and that he has enough skills and dedication to become a successful mathematician. Indeed, I am confident that the thesis, after little additional scientific work and substantial improvements in presentation, would be a successful PhD thesis.

References. The sources are correctly cited.

Presentation. Overall organization is good, the language, statements, and proofs can be understood with some effort. Taking into account the complexity of the results and the career stage of Vojtěch David, I rate the presentation as good, but please see my comments below.

Question for the defense

- You say that Appendix A provides some intuition about 2-sorted clones, while e.g. Theorem A.2 appears to be a full list. What remains to be done?
- To what extent is the proof independent to Taiman's proof? Do you see some intersections?
- Theorem C.3 (and the corresponding special case in the main part) seems like a basic fact about posets. Have you tried to find it in the literature?

Conclusion

This is an excellent thesis. I recommend to accept it as a Bachelor's thesis.

Libor Barto
7 June 2024

Comments to Vojtěch David

You do not need to react to these comments during the defense. I am happy to discuss the thesis in person at some point after the defense, if interested.

I repeat that this is an excellent work. Keeping this in mind, if I got the thesis as a paper to review, then I would suggest a *major* revision or reject.

Apart from interesting mathematical content, papers are ideally both precise and pleasant to read. Your presentation is simultaneously imprecise and hard to read. Below are some more detailed comments to illustrate where I see possible future improvements. This is by far not a comprehensive list of issues I spotted while reading the thesis.

- Mathematical imprecisions.
    - A lot of confusion can be caused by not distinguishing predicates and atomic formulas. Predicate does not have variables (atomic formula does), atomic formulas do not determine a predicate (one needs to know the order of variables and which ones should be dummy). It is, e.g., not clear what [S]_\wedge should be: closure under intersection or under conjunctive definitions? This confusion is one of the factors that makes statements such as Lemma 2.10 very hard to understand.
    - In the definition of k-sorted setting above Theorem 2.1 (btw definitions should be numbered so that one can refer to them): "A predicate is k-sorted if ..." is confusing. The formulation clearly suggests that k-sorted predicate is a special case of a predicate, which is not the case: k-sorted predicate is a predicate and an additional structure. The next paragraph is not understandable. E.g., what "identified" refers to? One can identify some variables if we already have some formula, but we do not have any. The explanation that follows is confusing because of the first item. For me, the two paragraphs altogether have the same information value as saying "k-sorted predicates and formulae are defined naturally".

      An example of a piece of text from the beginning, which is I believe impossible to understand unless one is already acquainted with the subject, is the example explaining CSPs.
    - Another type of sloppy writing is seen e.g. in the second paragraph in the proof of Lemma 2.3: „Now by setting .... In the equation (1) ..." does not make sense to me (should I replace the existentially quantified occurences of y? There are no others, the rest is hidden in the z...). Precise and correct would be e.g. to say that \rho'(\alpha) is witnessed by y_i=a_i.
    - Words such as e.g. „equivalent" and „corresponding" are often abused. For instance, „composition of these operations... is equivalent to composition of (k-1)-operations" on page 55 is too vague.
    - One more concrete example where the phrasing does not make sense to me. After definition of Inv and Pol you say „It is easy to see that... Pol and Inv gives us a certain Galois connection..." (I mostly agree). Then Theorem 1.2 states that Inv and Pol form a Galois connection..., and refer to this as a fundamental fact (I disagree, the fundamental fact is a diffent one, e.g., that closed sets in that Galois connection are clones and co-clones, which implies that Pol and Inv are mutually inverse order reversing bijections between clones and co-clones).
- Notation, terminology.
    - The same sort of objects (domain elements, tuples, variables, predicates, sets of predicates) should be denoted by the same sort of symbols (font, character set, close

in the alphabet) and different sorts of objects by different symbols (font and/or character set and/or far in the alphabet...). This is perhaps not always 100% achievable but the notation definitely needs to be well thought through and made consistent.

In the thesis different objects are often denoted in a very similar way (e.g., in Lemma 2.10 x_i is a variable--a concrete element of a variable set--while x^i_j is a „variable for a variable"--x^i_j \in \{x_1,....\}), and the same kind of objects in a different way (e.g., caligraphic P for a set of predicates, plain font for different sets of predicates).

- The notation should be efficient and intuitive (while not causing ambiguity of course). E.g., there is no need for „0,1" in the subscript of P, no other domain is ever used. Only introduce symbolic notation when it is helpful, it is hard for the reader to keep in mind the meaning of too many symbols.

  If the meaning of some term differs from the standard one, this should be well justified and emphasized. For instance, in the standard terminology, primitive positive formula can use equalities. The standard term for what you use would be positive primitive equality free. (Moreover, I learned why equality-free pp definitions are considered only from a footnote on page 62; perhaps I missed something earlier.)

  An example of an extremely confusing piece of notation in the thesis is the following: \bar{T} (or similar) is defined first without defining T, which already is confusing. Then a remark afterwards says that T is the closure of \bar{T} under... As \bar is often used for closure, it would be hard to find a more chaotic notation than this.

- Heavily symbolic expressions and formulations are hard to read. There is a large room for improvements in the thesis in this respect. Micro-scale examples: „f \in Clo(F) \cap O^n_k" is not a nice way to write that f is a k-ary operation in Clo(F). It is easier to read „R is a pp-closed set of k-sorted predicates", than „R \subseteq P^k_{0,1} such that [R]_pp = R".

  Introducing more meaningful and intuitive terminology and less symbols (e.g. perhaps something like pp-closed set relative to DL for S \subseteq DL and [S]_pp \cap DL = S) would help the readability greatly, e.g., one can say that every pp-closed set R is equal to the closure of R \cap DL under intersections instead of a part of Theorem 2.1. This heavy use of symbols concerns both formulations (e.g.Theorem 2.1) and proofs, like the same theorem or Lemma 2.5. The latter one is a crucial fact, the proof is close to unreadable.

- Organization.
  - In several places, proofs are referenced rather than theorems or lemmas – you say something like „We have already seen in the proof of ....". In such cases, the fact perhaps should be stated explicitly. The need for such references often indicates that it's time to rethink the overal organization (i.e., what and when should be explicitly stated).

- Language and other presentation issues.

  - Avoid exaggeration, self-evaluation, and unsubstantiated claims. I think „...clone..., which also finds numerous applications beyond the scope of the subject itself" is a bit exaggerated and definitely unsubstantiated. It would be much better and stronger to write something like „the concept finds applications outside of universal algebra, for instance in....". „We prove two very important result..." does not sound nice. Much better would be something like „the two results will be important in ...".

  - Some important pieces of information, e.g., without which the particular claim is incorrect, are in parantheses. This requires a different phrasing.

- Some words that are incorrectly used, or at least used in a nonstandard way, are „propose" (e.g., instead of claim), „discuss" (e.g. instead of prove, argue,...), „upon"; I believe „followingly" does not exist.

- Some phrases that are incorrect or do not sound nice are „almost surprisingly powerful properties", „somewhat neat pairing", „By a k-operation, we define", „As Theorem... suggests, we initially describe...", „Let us proceed by induction, as the claim is trivial for n=1", „As was studied in ..., we may generalize...", „Due to the repeated use of..., we may rewrite..."

Returning to a positive feedback, I think the thesis makes a good job in delivering the high level message (which for me is: there is a calculus that makes classifying multi-sorted Boolean clones, to some extent, a mechanical job). If a reader, however, wants to understand the details, then he or she would need to decipher the thesis, rather than read it.

My last but one remark is about the emphasized „elementary and purely relational" proof. By elementary you seem to mean „not using the bijection between clones and coclones". This is strange to me, since it is a fundamental (and simple) insight. What is the value of not using some fundamental tool or insight to solve a problem? I believe that in this specific situation, there actually may be some value, but the value is not in being elementary or purely relational, and it would need to be explained better.

As a side remark, you say that it is quite easy to see that Pol Inv (F) = Clo(F) while the other direction is more interesting. In my opinion, both directions are of similar interest and complexity (e.g., both proofs are via the same object).

Finally, clearly communicating results is a very important part of scientist's job. Nobody is perfect, but we should all pay enough attention to this: think, re-think, and try to improve. Spending more energy on this aspect means saving way more total energy of our colleagues and readers, and bigger impact of the work.