A “Tough Nuts” Track for the IPC

Jörg Hoffmann
University of Innsbruck, DERI
Innsbruck, Austria
joerg.hoffmann@deri.at

Abstract

In the development of the IPC, there is an ongoing conflict between pushing the frontiers, and consolidating our technology. With growing PDDL, it becomes ever more difficult to achieve both. We propose to introduce a separate instrument for consolidation. A “tough nut” is a domain that can be addressed, language-wise, by existing planning techniques, but that cannot be solved efficiently. In a tough nuts track, domains would be submitted to the IPC committee, and a web repository of accepted domains and results would be maintained. A domain can be removed if it is considered “solved”. Awards can be given to the first technique that solves a domain, and to the domains that survive longest. This simple mechanism naturally keeps track of the challenges remaining behind the moving frontier of PDDL. It explicits open questions and encourages research on subtle issues.

Motivation

In the development of the IPC, there is an ongoing conflict between pushing the frontiers, and consolidating our technology. The former aims at broadening the scope and expressivity of PDDL, and thereby directing planning research into new and practically relevant areas. The latter aims at maintaining challenges in the more established areas of PDDL. Obviously, the two goals are in conflict. This conflict manifests itself in the prolonged discussions, traditionally repeated for every IPC, whether to significantly extend PDDL or not. It manifests itself in the pains of the IPC organizer trying to design domains that are challenging in STRIPS and other established language subsets, while at the same time usefully explore the new language features.

The author has no illusions that every reader will agree on the above. Some people will say that the IPC does not actually have the goal of “consolidating technology”, and that the only purpose of such a thing as a STRIPS track can be to solicit a large enough number of participants. Be that as it may, certainly most people will find it hard to not agree that *consolidating our technology is important*. To name but a few examples:

- Numeric planning was introduced in IPC 2002. An extremely powerful language was defined (plan existence becomes undecidable even in very small subsets of the language (Helmert 2002)). Long and Fox provided some straightforward numeric versions of STRIPS benchmarks, and one more novel domain (Settlers) based entirely on numeric variables. The competitors came up with planning algorithms extended for (subsets of) the numeric language, and some results were obtained. In the five years since then, PDDL has moved on considerably, and not much has happened in the numeric area. Does that mean that numeric planning is solved? Certainly not! To name just one detail, Hoffmann et al. 2007 design a simple transportation domain that allows to control how tightly the resources are constrained. It turns out that heuristic planners fail quickly as the resources become tighter. While this is unsurprising, it also turns out that the heuristic planners are outperformed by an optimal SAT-based planner. This strongly indicates that numeric planning is not quite like other planning situations, and doing it efficiently may require different algorithms.

- IPC 2002 introduced almost arbitrary optimization functions. Although these are of paramount importance in practical planning, they have hardly been addressed so far.

- IPC 2002 introduced temporal planning. This has been addressed broadly and has since several years been widely considered one of the best solved areas beyond STRIPS. Still, temporal planning techniques are largely extensions of STRIPS machinery, and may be biased towards domains obtained in the same fashion. In fact, Cushing et al. 2007 show that current technology is mostly limited to domains that have non-temporal solutions.

- Last not least, even in propositional (STRIPS and ADL) planning there remain many challenges left to solve. In particular, there exist various domains that can be solved efficiently in principle, but that cannot be solved efficiently by planners. We are still waiting for a domain-independent planner that solves Blocksworld in polynomial time. A less disagreeable example is PSR.

It seems clear that the IPC, in its current form, is not suitable to address the above challenges; a more drastic formulation is that it has failed to address them. The moving target of IPC PDDL extensions is not helpful to encourage in-depth research on old language features. At the same time, pushing the frontier of PDDL is important. So how can we resolve this conflict?

Tough Nuts

The author believes that there actually exists an "eierlegende Wollmilchsau";[1] the author even believes he has found that animal. While neither seems particularly likely, perhaps this text can convince the reader that there is a potentially useful direction that we should at least consider exploring. The main idea is to introduce a separate mechanism, in parallel to the IPC, which is better suited to help consolidating planning technology:

Solicit "tough nuts" domains from the community: domains that fall into existing PDDL, but are challenging for existing planners. Maintain those domains in a web repository. Keep track of the best results achieved on them.

Remarkably, similar mechanisms – selection of tough instances and keeping track of efforts to solve them – are already being successfully employed in other communities. In SAT, Theorem Proving, and OR, there is a highly successful tradition of using particularly tough benchmarks (CNF formulas, theorems, TSP instances) to drive the field. Planning is somewhat different in that it is after good performance over a broad range of examples, rather than after trying to solve a particular example. Still, highlighting core open issues would be just as useful in planning as in SAT, Theorem Proving, and OR.

Note that the above mechanism differs from the traditional IPC not only in its emphasis on "old" PDDL. Further key differences are:

- The tough nuts provide a long-term maintenance of particular domains, with particular sets of problem instances, and the long-term maintenance of results. This is much more suitable for a long-term consolidation than the fast-growing benchmark repositories, and distributed results, produced by the current IPCs.
- The careful selection of a particular set of domains is more suitable to highlight specific critical open points.
- Soliciting the domains from the community – rather than from the IPC organizers – is more suitable to not overlook any issues (400 eyes see more than 4 eyes).

Of course, the "traditional" IPCs have already made attempts at providing benchmarks for consolidation. Domains from previous IPCs have been re-used, and considerable efforts have been made to lay emphasis on open issues and to design interesting benchmarks even for the simpler PDDL dialects. The author does not wish to imply that these efforts were useless;[2] rather, the tough nuts track should be seen as a more consequent framework to contain such efforts. At the same time, moving these issues into a separate framework eases the burden of the traditional IPC to come up with challenges for old PDDL. Granted, the IPC must still be careful about letting go of old PDDL, so as to not make it prohibitively hard for newcomers to enter. But if someone else cares about the challenges for STRIPS, then it is much easier to come up with *new* domains that make sense for the *new* language features.[3] This provides more freedom in the design of new domains, and hence may ameliorate the difficulties experienced in the past, where new language features (e.g. durations) were more often than not put to use in the form of extended "old" domains, with the known consequences (e.g. (Cushing *et al.* 2007)).

The author is aware that this text is getting increasingly unstructured in the attempt to prove the existence of the eierlegende Wollmilchsau. Let us just consider one more point before thinking about what a "tough nuts track for the IPC" would actually look like. The critical reader may argue that the task of the IPC should be to keep moving the PDDL frontier, and that the task of many individual researchers writing many individual papers should be to consolidate the technology.[4] But even such a reader should agree that, for this kind of research, it would be quite useful to have a forum keeping track of open and solved challenges. For young researchers new to the field, this would definitely be a premium source of inspiration. For the field as a whole, it would help to focus, and to encourage progress on subtle issues – doing "temporally expressive planning" (Cushing *et al.* 2007) is nice, but if one can win a prize with it then it is even nicer.

The author hopes that, up to here, the main idea behind the "tough nuts track" was sufficiently clear to make sense of the discussion. To the author's defense, it should be said that there is a reason for not dwelling too much on actual organizational details here: the purpose of this text is to spark the reader's imagination and enthusiasm for the idea; defining all details would be futile since realizing the idea will involve extensive negotiations anyway. Here is a list of key rules that, or so the author believes, a tough nuts track should follow:

1. Domains can be submitted by anyone. The domains are reviewed under supervision by the IPC committee. Acceptance or rejection is decided by that committee. The criterion for acceptance should be that the domain is indeed a tough nut, and provides a useful challenge. An upper limit on the number of open domains at any time might be useful.
2. Results can be submitted by anyone. The results should be easy to verify, and accompanied by a 1-page abstract outlining what the underlying technique is. The results are reviewed under supervision by the IPC committee. Acceptance or rejection is decided by that committee. The criterion for acceptance should be whether the results are valid, and indeed constitute progress. Submission of invalid or questionable results should be actively discouraged. In particular, results should only be valid if obtained with domain-independent techniques.
3. Both the domains and the results are maintained in a web repository so everybody can access them. The repos-

[1] Animal combining the advantages (and none of the disadvantages) of chicken, sheep, cow, and pig.

[2] In particular, doing so would mean to imply that my own efforts were useless.

[3] At least that's what the author would have felt in the preparations for IPC 2004.

[4] At this point the author imagines an even more critical reader who asks what all this AI Planning is about anyway and where its place is in the larger context of life, the universe, and everything; but let's not get into that for now.

itory should be simple, to keep the maintenance effort low. Each domain could provide links to the PDDL files, and, for each result, a link to the abstract and to the IPC-formatted results files.

4. If the results for a domain are good enough, the domain is considered closed. This is decided by the IPC committee. The responsible technique/author obtains an award.
5. If a domain survives – stays open – particularly long, or if the domain has turned out particularly useful in some other way, then the responsible author obtains an award. This is decided by the IPC committee.

Here you go . . .

At this point, even the well-meaning reader will jump up from her chair/sofa/motorcycle, and shout "Stop! Shouldn't we put B onto A instead, and draw a line separating X from Y?"

Since the author cannot possibly foresee, let alone answer, all possible questions that may be asked, the asking and answering is instead put into the hands of the readers themselves. In the unlikely case that you can't think of any questions, here's a selection; in the equally unlikely case that you can't think of any answers, some candidates are also provided for your convenience:

1. **Q:** How to ensure that this process indeed becomes the intended focused exploration of a small number of key challenges – rather than a fast growing mess incurring huge maintenance overheads?

   A: The author believes that this could be accomplished with upper limits on the number of open domains, and with strict rules regarding re-submission of material (domains or results), such as allowing at most one revision, or even allowing no major revisions at all.
2. **Q:** How exactly will this relate to the traditional IPC: Are the tough nuts run alongside the other IPC domains, in each IPC? Are awards handed out along with the IPC awards? Is submission of results allowed anytime, or only according to certain deadlines (like, at the time of an IPC)? Are some IPC domains (those that could not be solved well by the competitors) automatically included as tough nuts?

   A: These are four questions, not one.
3. **Q:** What about distinctions between different kinds of planners?

   A: Well, we will have to make them.
4. **Q:** Who will maintain the web repository?

   A: Of course the person who couldn't keep his mouth shut.
5. **Q:** Should we be strict about the "old" PDDL, or should we allow domains with new features, given those make sense as a challenge?

   A: Never say never?
6. **Q:** Should we give awards not for the solution of one particular domain, but for the solution of an entire subset of the domains?

   A: A definite "maybe". (Pro: more in line with domain-independent planning; Contra: *which* subset?)

The above should illustrate that it'll be a long way from this draft to an implemented "Tough Nuts" track liked by enough planning people to survive. The author believes . . .

. . . well, the author *hopes* that a compromise can be reached in discussions and negotiations. Maybe we can just apply Occam's Razor and start with something simple.[5]

Conclusion

In conclusion, the IPC in its traditional form is not a useful mechanism for consolidating planning techniques behind the moving frontier of PDDL. A potentially much better mechanism is a "tough nuts" track, highlighting the open issues and encouraging their solution. This would also ease the burden of the traditional IPC, hence serve to improve the adequate exploration of new language features, and hence make everybody's life happier.

Acknowledgements.

I thank Blai Bonet, Patrik Haslum, Hector Geffner, and the entire IPC committee, for discussions on the tough nuts track. In particular, while I originally envisioned a separate track in the traditional IPC, it was Blai's idea to instead organize this track in a more permanent way.

I thank the anonymous reviewers, one of whom contributed to my English Grammar, and another of whom contributed questions 5 and 6 in "Here you go . . . ".

Finally, of course the sloppy and provocative style of this text is a badly done copy of various artworks by Rao Kambhampati; may he view it as a compliment.

References

Cushing, W.; Kambhampati, S.; Mausam; and Weld, D. 2007. When is temporal planning *really* temporal? In Veloso, M., ed., *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07)*, 1852–1859. Hyderabad, India: Morgan Kaufmann.

Helmert, M. 2002. Decidability and undecidability results for planning with numerical state variables. In Ghallab, M.; Hertzberg, J.; and Traverso, P., eds., *Proceedings of the 6th International Conference on Artificial Intelligence Planning and Scheduling (AIPS-02)*, 44–53. Toulouse, France: Morgan Kaufmann.

Hoffmann, J.; Kautz, H.; Gomes, C.; and Selman, B. 2007. SAT encodings of state-space reachability problems in numeric domains. In Veloso, M., ed., *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07)*, 1918–1923. Hyderabad, India: Morgan Kaufmann.

[5] Then we let it evolve gradually, until it becomes sufficiently complicated to be entertaining.