RESEARCH

How to evaluate heuristics in planning? By time or by number of nodes considered? Are these results optimal? How to benchmark them?

The jury is still out (AIMA, 3rd edition), but there are now some interesting comparisons of the various approaches to planning. Helmert (2001) analyzes several classes of planning problems, and shows that constraint-based approaches such as GRAPHPLAN and SATPLAN are best for NP-hard domains, while search-based approaches do better in domains where feasible solutions can be found without backtracking. GRAPHPLAN and SATPLAN have trouble in domains with many objects because that means they must create many actions. In some cases the problem can be delayed or avoided by generating the propositionalized actions dynamically, only as needed, rather than instantiating them all before the search begins.

As Helmert (2001) explains, the efficiency of AI planning systems is usually evaluated empirically. The planning domains used in the competitions of the 1998 and 2000 AIPS conferences are of particular importance in this context. Many of these domains share a common theme of transporting portables, making use of mobiles traversing a map of locations and roads. In this contribution, we embed these benchmarks into a well-structured hierarchy of transportation problems and study the computational complexity of optimal and non-optimal planning in this domain family. We identify the key features that make transportation tasks hard and try to shed some light on the recent success of planning systems based on heuristic local search, as observed in the AIPS 2000 competition.

Running time on problems from classical planning domains such as LOGISTICS and BLOCKSWORLD has often been (and still is) used for comparing the relative merits of planning systems. However, this kind of comparison is always difficult. If no planning system performs well in a given domain, does that mean that they are all poor, or is that domain intrinsically hard? If they all perform well, is this because of their strength or because of the simplicity of the task? On a related issue, should planning systems be preferred that generate shorter plans but need more time? While there is no general answer to that question, theoretical results can contribute to the discussion, for example in cases where generating plans is easy but generating optimal plans is infeasible.
For addressing these issues, domain-specific complexity results for planning tasks appear to be useful. Pondering which domains to analyze, the ones that immediately spring to mind are the competition benchmarks from AIPS 1998 and AIPS 2000, considering their general importance for the planning community and the wealth of empirical performance data available.

While it would be possible to investigate each competition domain in isolation, it seems more worthwhile to identify commonly reoccurring concepts and prove more general results that apply to domain families rather than individual domains. Not only does this help present the results in a more structured way, it also allows to shed some light on the sources of hardness in these benchmarks.

Conclusions by Helmert (2001)
For fairly general transportation tasks, we have shown NP-completeness of non-optimal planning in the restricted fuel case and NP-completeness of optimal planning in all cases. Just finding some plan in tasks where fuel is abundant was shown to be a polynomial problem. This is detailed in Figure 3. For some domains, even some severe restrictions are still sufficient to get NP-hardness. Specifically, all NP-hardness results in the multi-agent competition domains still hold if there is only one agent, and the NP-hardness result for GRID still holds if there are no doors at all. For convenience, we repeat the results for the competition domains:

Domain name	PLANEX	PLANLEN
GRID	polynomial	**NP**-complete
GRIPPER	polynomial	polynomial
LOGISTICS	polynomial	**NP**-complete
MICONIC-10 (STRIPS)	polynomial	**NP**-complete
MICONIC-10 (simple)	polynomial	**NP**-complete
MICONIC-10 (full)	**NP**-complete	**NP**-complete
MYSTERY, MYSTERY'	**NP**-complete	**NP**-complete

It is interesting to observe that in those domains where heuristic local search planners such as FF (Hoffmann and Nebel 2001) excel, the table lists different results for plan existence and bounded plan existence. Because all hardness proofs only use a single agent, they carry over to optimal parallel planning, which implies that in these domains planners like Graphplan (Blum and Furst 1997) or IPP (Koehler et al. 1997) try to solve provably hard subproblems that local search planners do not have to care about. When optimal plans are not required, local search has a conceptual advantage here, and we cannot hope for similar performance from any planner striving for optimality.
Greedy local search is less appropriate, however, if additional constraints can lead to dead ends in the search space. We have faced this problem when dealing with fuel constraints and in the full MICONIC-10 domain, where it may be unwise to have people board the elevator who restrict its movement too much. In fact, the competition domains with NP-hard plan existence problems are exactly the ones for which current planners based on heuristic local search encounter unrecognized dead ends (Hoffmann 2001).
While the observation that non-optimal planning is often easier than optimal planning is by no means surprising or new, we consider it important to point

out. While there has been significant recent progress on non-optimal planning, optimal planners tend to get less attention than they deserve, maybe due to the fact that they are often compared to their non-optimal counterparts in terms of the size of problems they can handle. This kind of comparison is hardly fair.
We also observe that all discussed decision problems are in NP. We do not consider this a weakness of the benchmark set, as in STRIPS planning, NP membership is guaranteed as soon as plan lengths are polynomially bounded, which is a reasonable restriction from a plan execution point of view.

Bibliography:

- Helmert, M. (2001). On the complexity of planning in transportation domains. In ECP-01.
- AIMA 3rd edition