"Rational Metareasoning in Problem-Solving Search" by David Tolpin

This work explores the use of metareasoning in algorithms for decision-making and combinatorial search. This is a fascinating topic that holds great promise for the principled design of intelligent systems. Decision-making often benefits from improved information, yet gathering improved information consumes resources, so knowing how much information to gather before taking action is a key question. Metareasoning, and more specifically the value of information (VOI) concept explored in this work, shows how to evaluate improved information against the costs of collection and the delay in acting.

Metareasoning was a popular topic of research in the late 1980s and early 1990s, but due to difficulties in achieving practical results, it received little attention in the late 90s and early 2000s. The work in this thesis has played a significant role in buoying interest in this topic over the last several years, by showing practical benefits of metareasoning in a variety of settings. Showing progress in a single setting would have been a nice advance, and showing progress in four gives this work outstanding breadth and strengthens the author's argument for the practical benefits of metareasoning. There is no question that this is very strong work and represents a significant contribution to the field. I would rank it among the top quintile of doctoral theses in the discipline.

The main weakness of the thesis is that the work is not always described clearly enough to make the motivation of the equations and the justification of the experimental results completely convincing and to allow easy independent replication by others. While this would have increased the length of the thesis, this reviewer would have preferred increased clarity even at the cost of dropping one of the chapters. Furthermore, it would increase the impact and uptake of this important work by others. A list of minor comments follows on the next page. None of them are serious enough to warrant significantly delaying acceptance of the thesis.

Minor comments and suggestions

None of these are serious and none but the most trivial should be required to be fixed before final approval.

page ii: efficiency →efficiently

p4: personally, I prefer reserving the term 'utility' for the metric on the final net outcome, rather than a primitive cost term

p7: Besides →In addition

p8: cite Dechter and Pearl instead of RN03 for the optimality of A*.

calling RTA* an 'approximation version' of A* gives the incorrect impression that it is an approximation algorithm (in the sense of obeying an approximation bound)

hosts →costs

should C_n really refer to n? How about a new variable like m?

p9: need a citation for solving MAB using local search

p10: not sure what you mean by 'real-time A* is formulated as an offline search algorithm'. Korf introduces it as a real-time method for interleaving planning and acting.

RTA* also tries to choose the 'apparently best' action

p14: I found this very unclear, in part because the discussion is entirely abstract without concrete examples. By 'computational operators', I believe you mean actions of the search algorithm. Is a 'transition to a neighbor node' an expansion? It is confusing to think of the meta-level MDP's policy as giving base-level actions instead of meta-level actions.

Are you in the on-line setting here?

I think it would be clearer to consider each action as having multiple possible transitions (eg, many s' for which $P(s, a, s') > 0$).

Why is 'best' in scare quotes?

p15: I wish U took the same number of arguments all the time. Some of the notation is not formally introduced (eg A^T_α).

customary - citation

p19: not clear what a recurring meta-level action is.

gains →gain. This discussion is too brief.

p23: the utility of a base-level action is not the value of the evaluation function - that is just an estimate

I don't understand the last sentence on the page

p25: this is awfully abstract. I assume the second bullet is talking about sampling? But it's not clear at all.

p28: reward is used before being defined?

"a factor reflecting the time constraints" - do you mean the constraint C?

p29 - the s_{α^j} is after measurement, I assume?

p66 whether the goal node is expanded when it is generated depends on the f of the parent, doesn't it?

p67 the to time

p69 why not model a constant time bucket list?

we a heuristic

p76 I thought that ranking a state into a PDB index was expensive. Felner often talks about trying to limit the number of lookups.

the argument about expensive value ordering heuristics requiring domain-specific knowledge is unsupported. why shouldn't your technique work?

p77 why was it necessary to pretend that the algorithm would be LA* in the future? is there a lesson here?

The work of Othar Hansson and Andrew Mayer (students of Russell) is not cited.

do you mean "is not negligible" on the last line?

p32: was c_V defined?

This formula and derivation needs more explanation

p33: "the total cost ...may serve..." - why?

"can be approximated" - why?

p34: by gradually updating - how?

intensity - do you mean frequency?

"let η be the intensity..." - is this for a specific $Bel(V_j)$?

not sure how you got c/η

p35: ideally these plots would have confidence intervals around the means

can any of this discussion be in the terms of section 4.1?

it would be nice to show the Ackley function and $A(x, y)$

p36: The results should be discussed.

p37: second line: isn't the computational cost already included? what about showing both separately?

both the original *and*

p42: "expected remaining search time" - in the subtree, right?

wlog - expand acronym

p43: I didn't understand the equation just above 5.8

p44: *we* obtain

p45: you use $\propto$ in eq 5.14 and then in line 4 of algorithm 9 you compare against a specific value. Either I misunderstood or this requires some justification.

p46 "optimized" - how?

"gave no significant improvement" - why not? This is unusual! Is something fishy going on?

p47: you don't discuss using the VOI procedure with MC or pAC. Why not? (You explain a bit at the end of the thesis but the reader wonders here.)

you show the mean, but some CSPs are much harder than others. are a few instances driving the trend?

p50 "all stages" - what about all depths? perhaps just near the top of the tree?

p53: are we given a fixed number of pulls? Or a cost budget?

p54: I don't think the $\overline{X}_{\alpha}^{n_{\alpha}}$ was defined before use?

p58 confidence intervals would have been ideal

p59 an real-world

p60 "is the VOI" - needs more explanation

p61 there is no Figure 6.3.2

p64 "returns 0" - doesn't seem at all equivalent to me

p65 line 9 of algorithm 10 has problems