## 5.5 Insights into Methodology

In this section, we summarize our experience of applying rational metareasoning to search problems. One should treat the recommendations provided here as general guidelines to improving search algorithms rather than as strict instructions. Designing better algorithms is an art, even when based on the solid foundation of rational metareasoning.

### 5.5.1 Assessing Applicability of Rational Metareasoning

The first question raised by a metareasoning researcher facing a new problem is whether a solution *would benefit at all* from applying rational metareasoning. In most cases, an existing algorithm involving heuristic evaluation of search states justifies an attempt to perform the heuristic computations selectively, based on the value of information. However, there are two important exceptions:

- On the one hand, if **the heuristic relies on pre-computed data**, and online evaluation is cheap at the expense of intensive offline computation performed ahead of time for a wide range of problem instances, computing the heuristic selectively is unlikely to save the total search time. For example, pattern databases [CS98] proved to be an efficient approach for building informative and fast heuristics. 15-puzzle is one of domains in which pattern databases are particularly powerful [FHK11]; at the same time evaluating a state is just a small number of table lookups, so it never makes sense to try and compute the heuristic selectively.
- On the other hand, in the case of an **informative but very expensive heuristic** evaluation of a state may actually make the search algorithm slower in most cases, and the algorithm should rely on domain-specific knowledge to identify the states in which computing the heuristic is beneficial. Such fine domain-specific knowledge is usually hard to derive from a general notion of value of information. One example is high-accuracy solution counting algorithms, mentioned in Section 5.5.3, which are too time-consuming to be used in a value-ordering heuristic, unless a fine tuned higher-level heuristic is involved.

As a rule of thumb, rational metareasoning is beneficial for optimizing heuristic evaluation in algorithms where:

1. Ubiquitous heuristic evaluation of the search space *decreases* the total search time.

2. The heuristic computation time constitutes *a significant part* of the total search time.

Of course, this criterion does not cover all possible situations where rational metareasoning is helpful, but provides a good starting point for assessing applicability of metareasoning approach to a given problem domain.

### 5.5.2 Identifying the Metareasoning Decision

For a successful realization of the rational metareasoning layer it is important to identify the metareasoning decision. Determining the choices available at the decision points, and benefits and costs associated with each of the choices is not always simple. For example, in Section 5.3, "VOI-aware MCTS", the originally considered decision was whether to continue sampling or to commit to a move. This model facilitated derivation of VOI estimates for simple regret in Multi-armed bandits (Section 5.3.3). However, we could improve the Go agent only after realizing that for sampling in trees the alternative is to either continue sampling or to commit to a move *and possibly spend the unused samples at the selected child* of the current state (Section 5.3.4).

Similarly, in $RLA^*$ (Section 5.4) the decision is whether to compute the expensive heuristic $h_2$ in the current node $n$ or to bypass the evaluation *and compute* $h_2$ *in all children of* $n$, as though the algorithm resorts to plain $LA^*$. Earlier attempts to derive a metareasoning rule that compares the benefit of pruning the subtree due to computing $h_2$ with the cost of traversing the subtree did not result in a faster algorithm.

One way to ensure robustness of an algorithm $\mathcal{A}^{\mathrm{M}R}$ with the metareasoning level is to start with an efficient non-metareasoning algorithm $\mathcal{A}$ and implement metareasoning decisions in $\mathcal{A}^{\mathrm{M}R}$ as choices between either behaving provably better than $\mathcal{A}$ or resorting to $\mathcal{A}$. $RLA^*$ in Section 5.4 which is based on $LA^*$, and MAC with VOI-driven solution-counting value-ordering heuristic in Section 5.5.3 which improves on MAC with ubiquitous solution counting are examples of this technique.

### 5.5.3 Formulating an Utility and Information Model

Decision-making of the rational metareasoning approach is based on the notions of state utility and value of information of computational and base-level actions. A more complicated utility function does not necessarily bring additional benefits:

- A complicated function is more expensive to compute, and metareasoning computations tend to increase the total search time. Chapter 4, "Rational Computation of Value of Information" suggests some ways to leverage the costs of computations at the metareasoning level when an expensive utility function is absolutely necessary, but, as the case studies in Sections 5.5.3– 5.4 show, even simple utility functions result in considerable improvements.

- A utility model is often overly complicated because the algorithm designer tried to squeeze into the model domain-specific heuristic knowledge. However, metareasoning decisions that implicitly depend on heuristic knowledge impair the idea of rational metareasoning as an abstraction and hinder reasoning about the algorithm.

Estimating the value of information of actions involves maintaining beliefs about probabilities of action outcomes. In some cases, such as applying $RLA^*$ to planning problems [TBS$^+$13], guessing prior beliefs and updating them based on obtained evidence is possible and even necessary. In other cases, such as the MAC algorithm for constraint satisfaction problems (Section ), the beliefs can be constructed from the analysis of the algorithm under assumptions about the structure of the search space. Yet in other applications, such as the selection problem (Section 5.3), prior beliefs are not easily obtainable. An alternative to maintaining beliefs is estimating the value of information based on distribution-independent bounds derived from on concentration inequalities; Section 5.3.3 serves as an example of this approach. An additional benefit of bound-based VOI estimates is robustness of metareasoning decisions which would be otherwise by inadequate prior beliefs.

### 5.5.4 Analyzing the Results of Empirical Evaluation

Empirical evaluation is an indispensable part of developing an application of the rational metareasoning approach to a given problem. Analysis of the empirical performance of the algorithm facilitates verifying and improving the metareasoning model. Still, misleading results are frequent in the assessment of efficiency of the rational metareasoning approach. Such assessment involves comparing real search times which are influenced by inevitable noise of the computer facilities, but even more by features of a particular problem domain or test set.

For example, a useful comparison of VOI-aware MCTS to UCT in the context of Computer Go (Section 5.3.4) required configuring restricted versions of both players which are strictly based on Monte Carlo sampling with minimum use of position-dependent rules. Otherwise, the evaluation would compare compatibility of the rules with the sampling scheme in each of the agents rather than the advantage of one sampling scheme over the other. Developing a Go agent which is both based on VOI-aware sampling and wins over other state-of-the-art Go agents is an ambitious objective. However, this objective is only loosely related to verification of the rational metareasoning approach to Monte Carlo tree search.

Another example is the analysis of $RLA^*$ (Section 5.4.5. Applying $RLA^*$to 15-puzzle and to planning problems commands different decision rules, because in 15-puzzle the time required to compute $h_1$ is negligible compared to the open list manipulation, while in planning both heuristics are expensive to compute. 15-puzzle could also be used, in addition to evaluation on established planning problem sets, to analyze efficiency of the approach in the planning domain using the appropriate decision rule, but in the latter case the empirical results had to be processed accordingly, by extracting just the heuristic computation times.

# Chapter 6

# Summary and Contribution

The thesis comprises two major topics:

1. Rational computation of value of information (Chapter 4).

2. Case studies of application of rational metareasoning to selection and application of heuristic computations (Chapter 5).

The first topic addressed cases in which estimating the value of information of a computational action is expensive, as well as cases of too many computational actions, such that although estimating the value of information of a single action is cheap, estimating the value of information of all actions at each step of the algorithm is expensive. The research, published in [TS12b], resulted in an improvement to a widely used class of VOI-based optimization algorithms that allows to decrease the computation time while only slightly affecting the reward. Theoretical analysis of the proposed approach to rational computation of the value of information was supported by empirical evaluation of several combinations of algorithms and search problems.

The second topic considered several search problems and improved some well-known algorithms for solving the problems using the rational metareasoning approach:

- Adaptive deployment of value-ordering heuristics in constraint satisfaction problems (Section 5.5.3) [TS11];

- Monte Carlo tree search based on simple regret (Section 5.3) [TS12a, HRTS12];

- Decreasing heuristic evaluation time in a variant of A* (Section 5.4) [TBS+13].

In addition to the algorithm improvements, the studies demonstrated a number of common rational metareasoning techniques which can be extended to other problem types. In particular,

- Section 5.3, "VOI-aware Monte-Carlo tree search" provided distribution-independent upper bounds for semi-myopic VOI estimates in Monte-Carlo sampling.

- Section 5.4, "Towards rational deployment of Multiple Heuristics in A*", introduced a novel area of application of rational metareasoning—optimal search in optimization problems.

As a whole, the research advanced the use of rational metareasoning in problem-solving search algorithms. Applications of rational metareasoning in the case studies serve as examples to help researchers employ the methodology in solutions for other problems. Advances in rational computation and estimation of VOI increase performance and applicability of existing and new search algorithms and alleviate dependence of algorithm performance on manual fine-tuning.

The field of rational metareasoning still poses serious challenges. An important yet unanswered question is the extent to which algorithm performance can be improved due to employment of the metareasoning approach. In the case studies presented in the thesis, the improvements, while approached the theoretical limits of each particular algorithm variant, were moderate. It is still not clearly understood **whether a dramatic breakthrough in performance is possible** if more elaborated models of utility and information are used, or the limitations are immanent to the approach itself. Another interesting field of research is rational metareasoning where **action costs and state utilities are not commensurable**. Instead of mapping between different measures, as suggested in Chapter 4, a more general combination function should probably be used for best results. Ways of choosing or deriving such a function still remain to be discovered.