

Commentary on Baum's "How a Bayesian ..."

Stuart Russell, Computer Science Division,
University of California, Berkeley, CA 94720.

Summary of the Paper

The paper divides the problem of game playing into two parts: growing a search tree and evaluating the possible moves on that basis. The evaluation process is based in part on the idea that leaf node evaluations should be probability distributions rather than point values, and should be propagated back to the root *before* taking expectations. The propagation algorithm is more sophisticated than the product formula derived by assuming independence of leaf values. It is, however, weaker than the method of Hansson and Mayer (1989), who allow for dependencies by formulating the search tree as a belief network. Hansson and Mayer use the probability distribution of the *true cost* of the leaf. Baum and Smith use the distribution of the backed up value after expansion (Russell & Wefald, 1988). Although this short paper does not describe the analysis, the BPIP algorithm uses this because it describes the information the algorithm expects the player to have when (or if) that node becomes the current state. The distribution is obtained empirically, thereby providing metalevel information on the results of computations. As stated, it assumes the expansion yields a point value, whereas in the Baum-Smith scheme (unlike the Russell-Wefald scheme) it actually yields a distribution. Presumably this can be fixed by a rigorous analysis.

The relevance of the paper to the symposium lies in its use of decision-theoretic metareasoning to select nodes to expand and to terminate search. A computation is relevant if it promises to provide utility in the form of improved decisions (either now or eventually). Although this idea goes back at least to I. J. Good (1968), it has proved difficult both in theory and in practice. Ultimately, the approach promises to eliminate the notion of algorithm in favour of the notion of adaptive, rational metalevel control of object-level computations.

Baum and Smith have made significant contributions to the understanding of how to find efficient approximations to rational metacontrol (which is itself intractable). They attempt to avoid myopic estimation, which was perhaps the principal problem with Russell and Wefald's MGSS* algorithm, and to improve the bookkeeping in order to cut search overhead per node. Their approximation is based on the idea that there is some total value to be obtained from *complete* expansion of the tree (to termination?). Each node expansion makes some contribution to this. They make a plausible claim to the effect that this is related to the expected absolute change in the completion utility when the node is expanded. However, the connection to the actual expansion utility of the node is unclear. The correct expansion utility can only be analysed by treating the metalevel problem as a stochastic sequential decision problem, but I could find no such analysis. Discussion of the *qualitative* properties of the expansion measure should prove interesting, and will certainly contribute to our understanding of the relevance of computations.

Relevance and the Value of Computation

The value of computation (VOC), which is strongly connected to information value theory, has some nice properties. For example, provided rational use is made of the results of computation, computations have nonnegative expected value. It can be quite confusing, however.

I. J. Good, for example, suggested that a computation is valuable if it increases the value estimate of the root (see also Rivest). This rules out computations that might reveal one's plan to be a blunder—OK for politicians, but not for game-playing programs.

Part of the difficulty lies in the formulation. $P(A|B)$ should be independent of the *form* of B —i.e., any logically equivalent expression should be treated the same way—yet computations do nothing but yield logically equivalent forms. No new information is generated. Hence one has to resort to arbitrary restrictions, such as that the metalevel may not do node expansions in order to find the value of a node expansion. While one can "partition" the agent, so that the metalevel has no access to the rules of chess, this makes it impossible to show rationality of the agent as a whole.

Ultimately, there is no solution within the approach to control that tries to calculate what calculations are rational. The theory of bounded optimality offers a way out, by insisting not on rational control (which would involve arbitrary amounts of metalevel computation) but simply on whatever control results in the best overall program.

One way to achieve this is through adaptation within some architectural parameters. The value of computation can be *learned* by doing computations and seeing whether they helped. Computation sequences can be handled by formulating a reinforcement learning problem at the metalevel. One benefit of the kind of analysis in Russell-Wefald and Baum-Smith is that it helps to identify the relevant (in the inductive sense) features of the computational state so that learning can proceed quickly.

This leads to another interesting question. Consider the problem of learning the value of (say) allocating 100 nodes of search to a given type of problem. The value depends on how those 100 nodes are used, which depends in turn on the metalevel control algorithm's method for estimating the value of computation. We therefore have a feedback loop in the learning process. As yet, little analysis has been done on the stability and convergence of such processes (Tash & Russell, 1994).

Finally, one can question whether the tree-searching approach to games can be taken much further. Restricting computations to generation and evaluation of concrete states in continuous sequences from the current state seems unnecessary, although it does make for a sufficiently uniform computational process that allows metalevel analysis to proceed. More complex architectures, including condition-action rules, goal-based planning, and so on, provide much more relevant computations for the metalevel to choose from.

References

- Good, I. J. (1968) A five year plan for automatic chess. *Machine Intelligence*, 2.
- Hansson, O., and Mayer, A. (1989) Heuristic search as evidential reasoning. In *Proceedings of the Fifth Workshop on Uncertainty in Artificial Intelligence*, Windsor, Ontario.
- Russell, S. J., and Wefald, E. H. (1988) Multi-level decision-theoretic search. In *Proceedings of the AAAI Spring Symposium Series on Computer Game-Playing*, Stanford, CA.
- Tash, J., and Russell, S. (1994) Control strategies for a stochastic planner. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Seattle, WA.