Technical Report:
Empirical Bayes Assisted Probabilistic Search

Kevin Seppi
Computer Science Department
Brigham Young University
Provo, Utah 84602

Abstract

This report presents results obtained using an *Empirical Bayes* approach to improve cost estimates used in probabilistic, heuristic search algorithms. The details of this approach, which we call the Bayesian heuristic search ("BHS") algorithm, are given in (Seppi et al., 2004). In probabilistic heuristic search, a heuristic is used to create a probabilistic estimate of solution cost at each step of the search. These estimates guide the search algorithm to focus on areas most likely to yield good solutions. The BHS approach employed here combines information from heuristic estimates from other nodes in the search tree to improve neighboring estimates. The resultant BHS estimates have significantly lower mean squared error. The purpose of this report is to demonstrate the advantage of BHS in simulated environments especially an environment that does not match the model assumptions. In a search with a large branching factor, using the BHS estimates allows the search to find solutions while visiting as few as 1/5 of the nodes visited when the search is guided by a conventional heuristic.

1 Introduction

Search algorithms are fundamental to the solution to many problems in artificial intelligence and to many problems in other areas of computer science. Algorithms for searching large problem spaces can benefit greatly by using heuristic assessments of the best total solution cost to guide the search. These heuristic functions estimate and/or bound the cost of the best solution assuming the search proceeds through the node under consideration. Heuristics which give a lower bound on the total solution cost are used in algorithms like A* to find the best possible solution. Unfortunately, in many cases a tight lower bound is difficult to find. A poor lower bound results in a large effective search space, since the bound is unable to "rule out" very much of the search space. In such cases, a probabilistic estimate (an estimate which gives a probability density over the possible total solution cost through a particular node) may be effective(Russell & Norvig, 2003) or (Hansson et al., 1992). A probabilistic heuristic is not guarantee to lead the search to *the optimal* solution. For many problems tight probabilistic estimates are easier to find than a tight bound and may significantly reduce the search space, focusing the search on regions that are most likely to yield a good answer. We assert that heuristics which have lower variance will generally allow the search algorithm to spend less time searching parts of the search space which are unlikely to yield a good solution while focusing on regions most likely to contain good solutions.

In this report an Empirical Bayes method is used to *improve* the heuristic estimates; that is, Empirical Bayes is used on top of whatever heuristic naturally applies to the search problem. Empirical Bayes methods are a family of techniques that allow information from related estimates (heuristic assessments in this case) to be used to improve the quality of each individual estimate. The approach described in this paper uses estimates from sibling nodes in the search tree to provide the needed, related estimates. When the branching factor of the tree is medium or large (20 in our experiments), the BHS estimates have less than half the squared error of the conventional estimates. This improvement allows the search to find the same quality of solution while visiting only half as many nodes as a conventional search. For larger branching factors, the search visits a much smaller percentage of nodes.

The remainder of this paper is structured as follows:

Section 2 provides further background and introduces the notation. Section 3 very briefly revisits the algorithm of (Seppi et al., 2004). Section 4 demonstrates the superiority of the approach experimentally. Section 5 summarizes the results, and concludes the paper.

2 Background and Notation

This section gives further background on probabilistic search as implemented for the purpose of capturing the benefit of Empirical Bayes. It is assumed that the reader is familiar with heuristic search algorithms and with (Seppi et al., 2004).

2.1 Probabilistic Search

For the purposes of this paper, a *Search* is a search of a discrete decision space (a graph) for the lowest cost solution (destination or destinations). A *Probabilistic Search* is a search of the space in which a queue of nodes remaining to be expanded (nodes who's parents have been expanded but who's children have not yet been explored) is sorted in ascending order of a probabilistic estimate of the solution cost of the best possible solution (path) that passes through each node, n, in the queue(O. Hansson, 1990). Figure 1 shows a node n with its best possible descendent node n*.

Let θ_{n*} be the true cost of the best possible solution on a path through the node n. Let ν_n be an imperfect, probabilistic, heuristic estimate of θ_{n*}. To the extent that ν_n accurately estimates the cost of the best solution through n (θ_{n*}) this function will cause the search algorithm to expand and find the path to the lowest cost solution. Table 1 shows this algorithm in greater detail.

Basic Probabilistic Search Algorithm

- Create a prioritized queue ("p-queue") of nodes such that the nodes are maintained in ascending order based ν_n.
- Add the root node to "p-queue".
- Iterate over the following until a solution (leaf node) is found at the head of the queue.
  - Remove the node ("current") from the head of "p-queue". This node has the lowest estimated solution cost.
  - Expand all possible children from "current". For each child node:
    - If the child is a solution node (a leaf in the tree):
      - Set ν_{child} to θ_{child}, the true solution cost.
    - else:
      - Compute ν_{child}
      using a heuristic.
    - Add the child to the priority queue.
- Return the solution which was found at the head of "p-queue" or return "failure" if no solution exists.

Table 1: algorithm.

2.2 Bayesian Notation

In order to introduce Empirical Bayes we will need some basic Bayesian notation. Suppose that we wish to make a decision based on an unknown, random parameter θ. We use a *Prior Distribution*, (*Prior*), to characterize our understanding of θ before we observe additional data:

$$\theta \sim g(\theta) \tag{1}$$

Suppose also that we are permitted to observe data, y based on θ. The *Sampling Distribution* describes the data as if we knew the parameter:

$$y|\theta \sim f(y|\theta) \tag{2}$$

Bayes Law allows us to create a *Posterior Distribution* which combines our prior understanding of θ with the

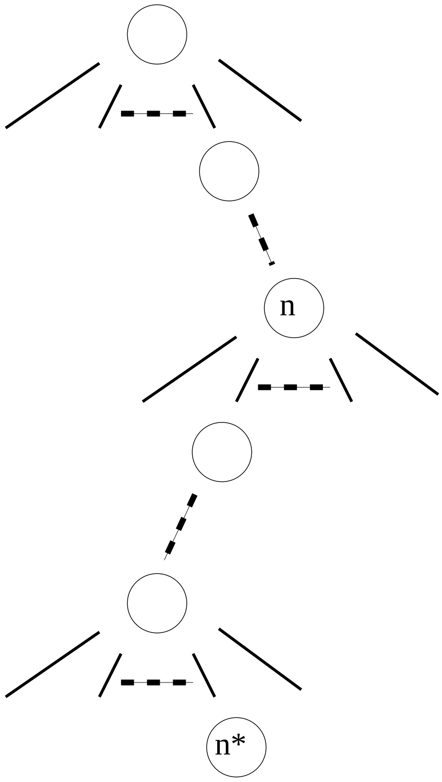

Figure 1: The node n*: The optimal solution under the node n.

understand of θ that we can infer from our observation of y:

$$\theta|y \sim \bar{g}(\theta|y) = \frac{f(y|\theta)g(\theta)}{\bar{f}(y)} \qquad (3)$$

where

$$y \sim \bar{f}(y) = \int_{\Theta} f(y|\theta)g(\theta)d\theta \qquad (4)$$

is the *Marginal* or *unconditional distribution* of y.

Note that for the Normal distribution we let the prior be

$$\theta \sim Normal(\theta_0, \sigma_0^2) \qquad (5)$$

and the sampling distribution be

$$y|\theta \sim Normal(\theta, \sigma_y^2) \qquad (6)$$

which yields the following posterior distribution:

$$\theta|y \sim Normal((1-B)y + B\theta_0, \sigma_y^2(1-B)) \qquad (7)$$

where:

$$B = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_0^2} \qquad (8)$$

and the posterior mean is:

$$\theta^* = (1-B)y + B\theta_0 \qquad (9)$$

The marginal is:

$$y \sim Normal(\theta_0, \sigma_y^2 + \sigma_0^2) \qquad (10)$$

See (DeGroot, 1970).

2.3 Empirical Bayes

In Empirical Bayes we use the marginal distribution 4 or 10 in the Normal case, as a basis for determining an unknown parameters for the Prior 1, θ_0 and σ_0^2 in the case of 5. Note that refGenericMarginal is the distribution one would observe if a sample of $y-i$'s were drawn unconditionally on θ, That is, draw a θ from $\theta|\alpha$ then draw a y from $y|\theta, \alpha$. We will refer to this as the Empirical Bayes Sample. If we are trying to estimate a vector of parameters, $\tilde{\theta}$ given a corresponding vector of observations, $\tilde{y}$ rather than a single value θ, then $\tilde{y}$ is exactly such an Empirical Bayes sample. Using $\tilde{y}$ we can compute a Maximum Likelihood Estimators of the mean and variance of 10. If σ_y^2 is known we can solve for the unknown θ_0 and σ_0^2. We can use θ_0 and σ_0^2 to define a common Prior Distribution, $g(\theta)$ and use Bayes law to compute a θ_i^* for each y_i in the usual fashion as shown in 7 through 9. This approach effectively pools the observations, $\tilde{y}$, to create an empirically defined Prior. In the case where all the distribution are Normal, this is equivalent (Morris & Efron, 1972) to Stein's estimator (Stein, 1955). Stein's estimator can be shown to yield lower expected total mean squared error (Morris, 1983).

3 Empirical Bayes Assisted Search Algorithm

The BHS uses the heuristic values from sibling search nodes as $\tilde{y}$ and computes improved heuristic values (θ_i^*) for using the Empirical Bayes approach described above. The lower mean Squared error of θ_i^* (vs. y_i, the normal heuristic value) more quickly guides the search to its objective.

Note that in this formulation $g(\theta)$ represents the hidden process that created the sibling nodes.

4 Results

(Morris, 1983) shows that Empirical Bayes estimators can have significantly lower variance than conventional estimators. Using a simulated search problem the observed variance of the Empirical Bayes estimates is much lower than the variance of conventional estimates, as one would expect. The reduced variance of the Empirical Bayes estimator profoundly affects the performance of the search. Unlike (Seppi et al., 2004) we have structured the search to continue until a solution is found that is better than a given confidence interval. In the case of our experiments we subtracted one standard deviation from the posterior mean, that is we used $\theta^* - \sqrt{\sigma_y^2(1-B)}$ rather than just θ^*. Likewise we used $y - sigma_y$ instead of just y (although this has no impact on the order of the search since σ_y is a constant. Also, unlike (Seppi et al., 2004) we ran the search until a minimal cost path was found which was better (lower cost) that any heuristic estimate less one standard deviation. Since the both the conventional and /ebayes/ search proceeded until it achieved this confidence, both achieved essentially identical solutions. Both searches found solutions within 0.2% of the true best solution (which is know by construction in our simulation study). The impact of the Empirical Bayes estimator is manifest in the number of nodes searched to achieve the solution.

Figure 2 shows the average (over 100 repetitions) number of nodes searched in a tree with depth 4 and at various branching factors. The true solution costs are obtained from the sum of randomly generated edge costs (Normal(100,1)). The error is also randomly generated (Normal(0,10)).

As the branching factor rises there is progressively more information in the pool of sibling nodes. This significantly slows the exponential increase in the number of nodes searched, although both curves demonstrate the naturally exponential shape we would expect from a tree search. The Empirical Bayes approach reduces the *effective* branching factor by approximately 50%.

The results in Figure 2 are based on an environment where the model assumptions hold, including Normally distributed solutions and Normally distributed error. Figure 3 shows the results from the same set of experiments but in which the model assumptions are violated. In these experiments the solution costs and the errors are Uniformly distributed but with the same mean and variance as in experiments shown in 2, to allow comparison. Violating the model assumptions reduces the effect of the Empirical Bayes approach, (the effective branching improves by only 40%), but the results are still compelling.

5 Conclusions

These simulation results show that there is a large benefit to be gained from the use of BHS in the context of search problems where a probabilistic estimate is used to guide the search. These simulation results demonstrate that even when the model assumptions are violated, BHS retains most of its advantage over conventional probabilistic search heuristics.

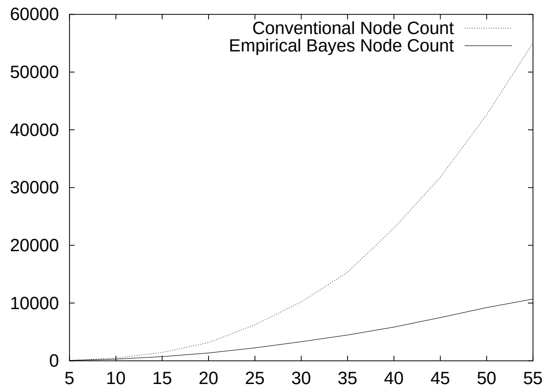

Figure 2: Nodes searched to find equivalent solutions, Conventional vs. Empirical Bayes, Model assumptions hold.

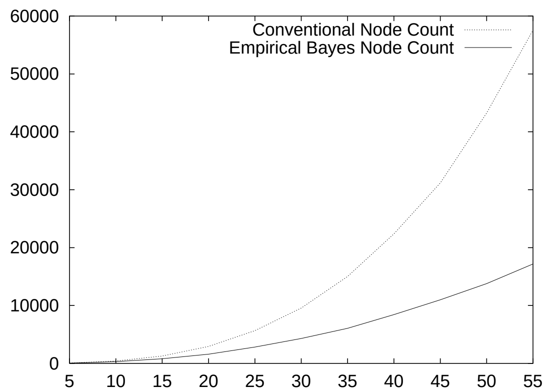

Figure 3: Nodes searched to find equivalent solutions, Conventional vs. Empirical Bayes, Model assumptions violated.

This approach has also been applied to real world problems, specifically guided model checking (Seppi et al., 2004). In this context the search is used to find errors in software or hardware systems given formal definitions of these systems. In this context, errors in models were found using the empirical bayes search where conventional techniques had to search *3 orders of magnitude more* states (nodes) to find the error.

Acknowledgements

This work was supported by a grant from the Brigham Young University College of Physical and Mathematical Sciences.

References

DeGroot, M. H. (1970). *Optimal statistical decisions.* McGraw-Hill.

Hansson, O., Mayer, A., & Valtorta, M. (1992). A new result on the complexity of heuristic estimates for the a* algorithm. *Artificial Intelligence*, *55*, 129–143.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. with discussion. *Journal of the American Statistical Association*, *78*, 47–65.

Morris, C. N., & Efron, B. (1972). Empirical Bayes estimators on vector observations - an extension of Stein's method. *Biometrika*, *59*, 335–347.

O. Hansson, A. M. (1990). Probabilistic heuristic estimates. *Annals of Mathematics and Artificial Intelligence*, *2*, 209–220.

Russell, S. J., & Norvig, P. (2003). *Artificial intelligence: A modern apporach.* Pearson Education. Second edition.

Seppi, K., Jones, M., & Lam, P. (2004). ??? *Submitted to Tacas 2004.*

Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 1* (pp. 197–206). University of California Press.