Control of Computational Effort through Robustness Analysis: A Study of Visual Position Estimation under Time-Critical Constraints

Fabio Cozman
Robotics Institute, Carnegie Mellon University
fgcozman@cs.cmu.edu — http://www.cs.cmu.edu/~fgcozman/home.html

Abstract

This work analyzes the relationship between flexible computation and robustness analysis. We start from the assumption that computational effort must be spent only when it can realistically affect decisions. We use robustness analysis to detect when decisions are (or are not) affected by changes in the agent state. As computational effort is aimed at changing the agent state, robustness analysis is the right tool for evaluating the impact of computational effort. We present models and methods that lead to efficient robustness analysis, particularly in the domain of maximum a posteriori estimation. We demonstrate this approach in an image processing system that estimates position from outdoor images.

Introduction

This work analyzes the relationship between flexible computation and robustness analysis. Decisions are *robust* when they are acceptable across variations in the decision maker state. The idea pursued here is that effort must halt when decisions cannot be reversed or defeated by changes in the agent state: computational effort is spent only when it can realistically affect decisions. As robustness analysis is concerned with the impact of changes in the agent state, it is the right tool for evaluating computational effort.

The key observation is that we can conduct robustness analysis efficiently. We benefit from recent breakthroughs in the theory of robust Bayesian Statistics (Berger 1985), which demonstrate the great convenience of the *bounded ratio family* for robustness analysis (DeRobertis & Hartigan 1981; Wasserman 1992).

Our approach grew out of experience with an image processing system that estimates position from outdoor images; we use this system as a case study. We model measurements through bounded ratio families. Position estimation begins with simplified measurements, which are used to narrow the search to a small number of possible estimates. Robustness of the estimates is then verified; if robustness fails, a refined model is used in a new estimation round. The estimation procedure terminates when a single robust estimate emerges, or no more refinement is possible for the image measurements. This procedure uses robustness analysis to suggest the most efficient use of computation.

Robustness and Flexible Computation

This work proposes an approach for flexible computation which is based on robustness as a measure of value of computation. A classic approach for evaluation of computational value is to obtain the expected value of computation and proceed if expected gains are larger than expected costs (Heckerman & Jimison 1989; Horvitz 1989; Matheson 1968; Russell & Wefald 1991). In the robustness approach, the gains from computation are judged with respect to computations already performed, not expected results.

A simplified example will clarify the robustness approach. Consider a decision problem with two possible actions, a_1 and a_2. Suppose the utility of each action depends on a state θ. Utilities are defined through functions $u(a_1, \theta)$ for a_1 and $u(a_2, \theta)$ for a_2. Beliefs about the state θ are translated to a distribution $p(\theta)$. Judgements about actions are made with respect to expected utility, *i.e.*, the values

$$E[a_i] = \int u(a_i, \theta) p(\theta) d\theta.$$

The best action is the action with largest expected utility.

For this example, suppose the distribution $p(\theta)$ is coded into a complex Bayesian network. There are methods that allow progressive construction of bounds for $p(\theta)$ (Draper & Hanks 1994; Jaakkola & Jordan 1996). For a particular set of bounds, consider a credal set defined by all consistent distributions[1]. Suppose this credal set can be indexed by $\alpha \in [0, 1]$, and call a member of the credal set by $p_\alpha(\theta)$. For each $p_\alpha(\theta)$ we can obtain $E_\alpha[a_i]$, the expected value of action a_i with distribution $p_\alpha(\theta)$.

Consider a graph of α vs. $E_\alpha[a_i]$. Each interval of α values is a credal set. One such graph is depicted in Figure 1, for which we analyze three possible credal sets.

- For any value in the interval $[\alpha_3, \alpha_4]$, action a_2 is best. Robustness is present and no further computation is necessary.
- In the interval $[\alpha_1, \alpha_2]$, differences in expected utility between a_1 and a_2 are small. Both actions are robust and satisfactory without further computation.
- In the interval $[\alpha_1, \alpha_3]$, differences of expected utility are too high: robustness fails and further computation is advisable.

Robustness can be either an additional figure of merit for assessments of value of computation or a direct indicator of value of computation. The point is that robustness is an important measure for assessment of value of computation. The particular use of robustness analysis is problem dependent.

[1] All distributions with probabilities between the lower and upper bounds.

In general, we can think of a base level decision maker, who tries to make decisions with coarse models. Sometimes these coarse, abstracted models are sharp enough for a robust decision, sometimes they are not. A meta level decision maker looks at the robustness of the base level decisions and suggests possible refinements that require more computation. The important question for the meta level system is to detect when the abstracted details can be useful (Chrisman 1992), exactly the question that robustness analysis is suited to answer.

Robustness Analysis: the Bounded Ratio Family

Robust Bayesian tools form the basis of our analysis (Berger 1985). Such tools have been formalized and axiomatized; the name *Quasi-Bayesian* (Giron & Rios 1980; Walley 1991) refers to an axiomatic theory of decision that uses convex sets of distributions for reasoning about uncertainty[2]. The set of distributions is called *credal set* (Levi 1980) and represents the various distributions the agent deems possible at a given point. To study robustness of a decision, we analyze how much a decision is affected by variations within the credal set. Notice that, compared to the Bayesian framework for expected value of computations (Horvitz 1989), the prescription above does not require "meta-probabilities" over expected probabilities. "Meta-probabilities" may be be hard to specify and elicit in a real problem.

The main question is: how can we quickly conduct robustness analysis? We present here a model that fulfills this requirement.

Recent results in the theory of robust Statistics allow us to perform robustness analysis with efficient models. In particular, it has been

[2] An introduction to technical aspects of Quasi-Bayesian theory, with a larger list of references, is available at http://www.cs.cmu.edu/~fgcozman/qBayes.html.

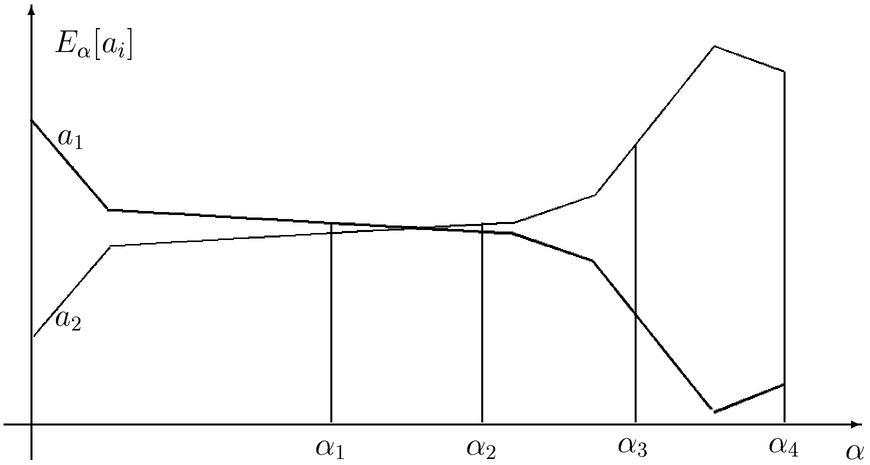

Figure 1: Expected utility $E_\alpha[a_i]$ for actions a_1 and a_2

proven that the *bounded ratio* family of distributions can be manipulated quickly and still produce all necessary information for robustness analysis (DeRobertis & Hartigan 1981). A *bounded ratio* family is defined by two measures, $l(A)$ and $u(A)$, where A is an event. The family consists of all probability densities $p(A)$ so that

$$l(A) \leq \alpha p(A) \leq u(A) \quad \text{for some } \alpha.$$

In this family of distributions, we can obtain posterior and marginal densities by using summations and Bayes rule on the measures $l(A)$ and $u(A)$ (Wasserman 1992). For any event A the minimum and maximum values of the probability of A are:

$$\underline{p}(A) = \frac{l(A)}{l(A) + u(A^c)} \quad \overline{p}(A) = \frac{u(A)}{u(A) + l(A^c)}. \tag{1}$$

In this paper we concentrate on maximum a posteriori estimation. Estimation is viewed as decision making, where the actions are the estimates. Our purpose is to estimate variables by the maximum value of their posterior distribution. In this setting, robustness is attained when posterior probability ranges for different estimates do not overlap. For example, consider two estimates A and B. We have a robust situation when the posterior probability of A is inside $[0.1, 0.2]$ and the posterior probability of B is inside $[0.7, 0.8]$. We have a situation that deserves further computational effort when the posterior probability intervals overlap.

Suppose the agent is trying to decide between two estimates, A and B. Applying Bayes rule and marginalization on $l(\cdot)$ and $u(\cdot)$, the agent obtains $l(A)$, $u(A)$, $l(B)$ and $u(B)$. Four quantities can be computed:

$$\overline{p}(A) = \frac{u(A)}{u(A) + l(A^c)} \quad \underline{p}(A) = \frac{l(A)}{l(A) + u(A^c)},$$

$$\overline{p}(B) = \frac{u(B)}{u(B) + l(B^c)} \quad \underline{p}(B) = \frac{l(B)}{l(B) + u(B^c)}.$$

If $\overline{p}(B) \leq \underline{p}(A)$ or $\overline{p}(A) \leq \underline{p}(B)$, A or B are respectively robust estimates. If the ranges of posterior distributions overlap, further computation is necessary.

A Case Study: Visual Position Estimation in Remote Driving

We apply the theory above to an estimation activity that arises in planetary robotic exploration. Figure 2 illustrates the idea we pursue. A rover in a planetary mission sends a stream of images to mission control. An interface receives the images and produces estimates for the position of the rover. A large map of the rover's environment is available; the operator can relate the estimates produced by the system with information in the map[3]. The system automatically segments the images, extracts the horizon and searches for mountain structures in the horizon. These steps require a significant amount of computation, taking 5 to 6 seconds per image in a Silicon Graphics Impact workstation[4]. Figure 3 shows the output of the image processing system for a sequence of images from Pittsburgh; the boxes indicate the structures detected by the image processing system.

The key aspect we focus on is the search for a maximum a posteriori position estimate given the results of the image processing system. We have time constraints for this task as the operators must interact smoothly with the robot. After image processing, the system has to search for the best matches between image mountains and structures in the map. The search space is equal to the number of possible positions on the map. Maps have approximately 36km^2, quantized at 30 meter intervals. This gives us more than 30,000 positions to be checked and evaluated. A position (x, y) is evaluated by considering differences between image structures and map structures that are visible from the point (x, y).

The search is much simplified if we look at a reduced description of the "prominent" features in the image (larger mountains and peaks). Here is the trade-off: we have to decrease the number of image features for efficiency purposes, *but* increase the number of image features to improve the reliability of the estimates. We must not consider all possible image features at once. Instead, we must use a small number of features and analyze the robustness of the resulting estimates. If we have a single robust estimate, we stop. If not, we must continue looking at more features, spending computation only on the estimates that show the potential to be the best. We now outline this robustness based approach.

The Estimation Model

To obtain a posterior distribution over positions (x, y), we need a prior distribution and likelihoods for the visual measurements. As we have no a priori information about position, we define a single uniform distribution over positions as our prior.

The formulation of a complete likelihood function relating the position and images (measurements) is extremely costly because it involves comparing the images from the robot to images obtained from the map (Thompson *et al.* 1990). We have to simplify the description of image features. So, instead of specifying the complete likelihood to begin with, the estimation procedure uses an abstracted description of the images. Each image is converted to a vector of point features containing information about location of "prominent" image structures. To indicate the abstracted nature of this model, we represent the likelihood function by a bounded ratio family.

For a single mountain detected in the image, with image position b and height h, we have the lower and upper likelihood measures[5]:

$$l(b, h|x, y) \propto \underline{\alpha}(h - h(x, y)) \times \exp\left(-\frac{(b - b(x, y))^2}{2\sigma_b^2}\right)$$

$$u(b, h|x, y) \propto \overline{\alpha}(h - h(x, y)) \times \left(\exp\left(-\frac{(b - b(x, y))^2}{2\sigma_b^2}\right) + k\right),$$

where $b(x, y)$ and $h(x, y)$ are the ground truth

[3] A description of the project is available at http://www.cs.cmu.edu/~fgcozman/Research/VIPER/viper.html.

[4] A report of the computer vision aspects of this work can be found at (Cozman & Krotkov 1996).

[5] A collection of several image features is modeled by multiplying lower and upper measures for all features. This is one of many possible assumptions of independence related to sets of probability distributions (Walley 1991).

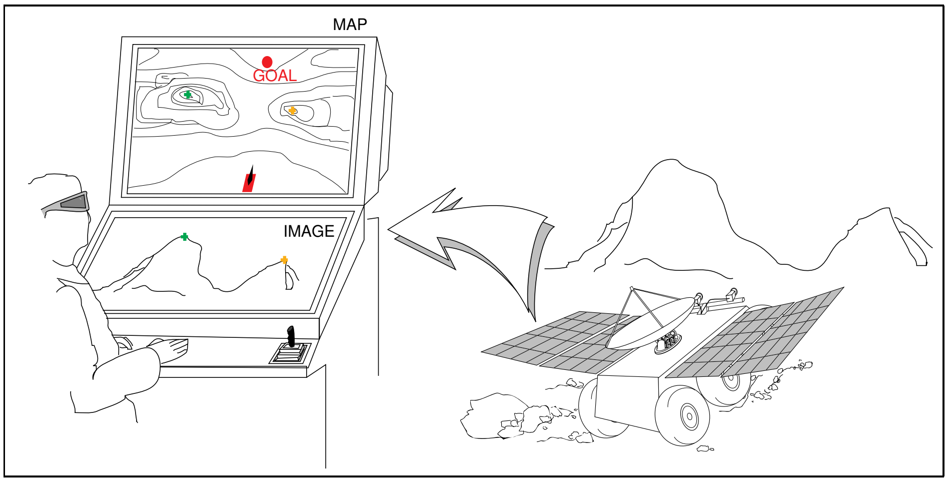

Figure 2: An Interface for Remote Teleoperation of Mobile Robots

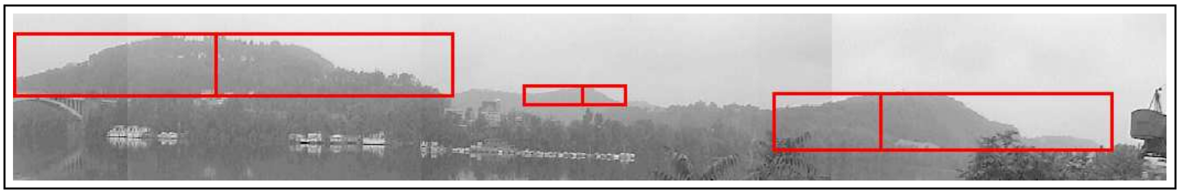

Figure 3: Result of Image Processing System

values for bearing and height (derived from the map); k, σ_b are constants.

The functions $\underline{\alpha}(\cdot) < \overline{\alpha}(\cdot)$ are defined by tables. In our implementation, the gap between these two functions increases as the height of an image feature increases. This gap represents the fact that larger image structures have potentially more information hidden in the abstracted model than small structures do. Tuning these various parameters is similar to assessing a Bayesian model, but here we can codify computational aspects that are absent in the single probability approach.

Estimation Procedure

We start with a selection of a few image features, described as point features, and obtain ranges of posterior distribution from both the prior and likelihood measures described in the last section. We then check whether the ranges overlap for different estimates. If there are many overlapping candidate positions, additional image features and refined descriptions (such as roundness and spatial extent) are used to pin down the best estimate. Finally, the terrain is rendered as it would be seen from the candidate estimates, and the image from the rover is directly compared with the rendered images. The best candidate is the one that produces the best match between rover and rendered images. The procedure stops when a best estimate emerges or no image feature is available. Because we look only at complex features for a few overlapping estimates at any given step of the estimation procedure, the cost of the overall procedure is much smaller than a full Bayesian analysis.

Conclusion

This paper emphasized robustness as an effective measure for the value of computation. Direct robustness measures can either be incorporated into the usual expected value of computation metrics, or replace such metrics. The key challenge is to present techniques that allow a decision maker to quickly evaluate the robustness of a decision, which we met by proposing the bounded ratio model.

We focused on maximum a posteriori estimation. The robustness measure we used is the simplest: whether or not the posterior probabilities for two estimates overlap. As a case study, we described how we defined bounded ratio families for an image processing application, and how these models are used for flexible control of computation.

The approach here presented can be useful in a variety of circumstances. First, it may be easier to assess robustness than to elicit all quantities involved in the expected value of computation. Second, it may be natural to assess robustness when approximation algorithms provide probability bounds. Third, models like the bounded ratio family lead to very efficient methods for control of computational effort.

References

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag.

Chrisman, L. 1992. Abstract probabilistic modeling of action. *First International Conference on AI Planning Systems.*

Cozman, F. G., and Krotkov, E. 1996. Position estimation from outdoor visual landmarks for teleoperation of lunar rovers. *Proc. Third IEEE Workshop on Applications of Computer Vision.*

DeRobertis, L., and Hartigan, J. A. 1981. Bayesian inference using intervals of measures. *The Annals of Statistics* 9(2):235–244.

Draper, D., and Hanks, S. 1994. Localized partial evaluation of belief networks. *Tenth Conference on Uncertainty in Artificial Intelligence.*

Giron, F. J., and Rios, S. 1980. Quasi-bayesian behaviour: A more realistic approach to decision making? In Bernardo, J. M.; DeGroot, J. H.; Lindley, D. V.; and Smith, A. F. M., eds., *Bayesian Statistics.* Valencia, Spain: University Press. 17–38.

Heckerman, D., and Jimison, H. 1989. A bayesian perspective on confidence. In Kanal, L. N.; Levitt, T. S.; and Lemmer, J. F., eds., *Uncertainty in Artificial Intelligence 3.* North-Holland: Elsevier Science Publishers. 149–160.

Horvitz, E. J. 1989. Reasoning about beliefs and actions under computational resource contraints. In Kanal, L. N.; Levitt, T. S.; and Lemmer, J. F., eds., *Uncertainty in Artificial Intelligence 3.* North-Holland: Elsevier Science Publishers. 301–324.

Jaakkola, T. S., and Jordan, M. I. 1996. Computing upper and lower bounds on likelihoods in intractable networks. *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence* 340–348.

Levi, I. 1980. *The Enterprise of Knowledge.* Cambridge, Massachusetts: The MIT Press.

Matheson, J. E. 1968. The economic value of analysis and computation. *IEEE Trans. on Systems Science and Cybernetics* SSC-4(3):325–332.

Russell, S., and Wefald, E. 1991. *Do the Right Thing.* Cambridge, MA: The MIT Press.

Thompson, W.; Pick, H.; Bennett, B.; Heinrichs, M.; Savitt, S.; and Smith, K. 1990. Map-Based Localization: The "Drop-Off" Problem. In *Proc. Image Understanding Workshop*, 706–719.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities.* New York: Chapman and Hall.

Wasserman, L. 1992. Invariance properties of density ratio priors. *The Annals of Statistics* 20(4):2177–2182.