COMMENTS ON THE ORIGIN AND APPLICATION OF MARKOV DECISION PROCESSES

RONALD A. HOWARD

Stanford University, Palo Alto, California 94305-4026, rhoward@stanford.edu

Recently, a European mathematician who had spent many years on the theoretical study of Markov decision processes visited me and inquired about the range of application. I replied that I knew of very few practical applications and that I had found only one in the course of my career that I considered really successful. He was somewhat taken aback until we discussed the problem of the very large data requirements imposed by the Markov decision process formulation. Of course, he then asked me how these data requirements were met in my one successful application, and we spent the balance of his visit in discussing this application.

I was reminded of this episode when Dr. Martin Puterman asked me a few months ago if I would comment on the origins of Markov decision processes. It occurred to me that the story of my one successful application might suit his purposes and be as interesting to the conference participants as it was to my European visitor.

You see, my one successful application was the original application that sparked my interest in this whole research area. The story begins about twenty years ago, when I was a graduate student in electrical engineering at M.I.T., working part time with the Operations Research Group of Arthur D. Little, Inc. It was a period of great excitement in the Operations Research Group, for the first commercial applications were being performed with success, and there was considerable intellectual stimulation arising from many discussions of the fundamentals of the subject. Our mentor was Dr. George Kimball, who taught me and many of my colleagues the most valuable lessons I ever learned on the subject of operations research.

The group had an important relationship with Sears, Roebuck and Company and helped that firm with a number of its problems. When I became involved, Sears' management was becoming increasingly concerned with the effectiveness of the operation by which they sent catalogs to present and prospective customers. There were many options. A customer could receive up to fourteen individual mailings a year, ranging from the general catalog to individual sales fliers. Or, he might receive any subset of these mailings. The cost of any particular mailing was easily determined, but what was the benefit?

To answer that question, let me take you back with me to a grey Chicago day, twenty years ago, when I first saw the Sears' catalog information system. It was an unforgettable sight. Imagine, if you will, two or three acres of green steel filing cabinets. Each cabinet contained steel Addressograph plates, about four inches square. Each plate had a stencil for printing the customer's name and address, and, as inserts, three small cards with several punched holes. These holes provided a limited summary over three seasons of the customer's purchasing history as a Sears mail order client. About 100 young women continually circulated among the filing cabinets. They were supplied with one copy of the customer's latest mail order, and it was their job to update the punches on the cards to incorporate the effect of the order. The information recorded was highly quantized in terms of the number and amount of the orders. The key to the system was the machine that used the steel plates to print labels for the catalog to be mailed. A drawer of plates was stacked into the machine. The machine examined the punched holes in the cards on each plate and determined, according to wired-board logic, whether this pattern of holes qualified the customer to receive the particular catalog being distributed at that time. If the pattern was favorable, a label was printed; otherwise, not. Thus, for example, the management could decide to send the general catalog only to customers who purchased more than $20 during the present season. By wiring the machine appropriately, this decision was implemented simply by passing every drawer of cards through the labeling machine.

In making this decision, the management had traditionally looked at the direct profit to be expected from a customer as measured by the difference between the marginal profit on his purchases and the cost of sending him the catalog. As we examined the operation, we began to wonder whether it might be profitable to send catalogs, not just on the basis of the profit they might produce during the present season, but also for the impact they might have in moving the customer into more profitable categories in the future.

We decided to model this system by what has come to be known as a Markovian decision process. Each customer's state was described by his purchase history; there were a

Subject classification: Professional: comments on.
Area of review: Anniversary Issue (Special).

total of about 50 states. Transitions occurred each season. The reward for a transition, under a given catalog mailing policy, was simply the marginal profit from the transition less the cost of the catalogs mailed. Finally, the transition probabilities were computed by special runs from the Addressograph system. In fact, it was the existence of this system that made the entire approach feasible.

The optimum policy, for both discounted present value and average reward criteria, was found by value iteration. This all took place in the days when computers still had vacuum tubes. And so the runs were fairly time-consuming, but still economical. The optimum policy was different from the policy that had previously been used. The optimum policy was not the policy that maximized expected immediate return, but rather a policy that balanced this return with the effect on future state transitions. The net result was a predicted few percent increase in the profitability of the catalog operation, which, however, amounted to several million dollars per year.

The optimum policy was confirmed by applying it to the test index, a selected set of customers whose purchases were very carefully monitored. When the policy was later implemented on the full customer set, the results closely confirmed the model predictions.

The experience left me with the suspicion that there might be a way to go directly to the best policy without the need for value iteration. I worked on the problem for about six months under Dr. Kimball's supervision and was able to develop a policy iteration method (Howard 1960). I would like to have described the motivating problem at that time, but the proprietary nature of the work with Sears, Roebuck and Company made that impossible. Perhaps, the cause of application might now be further advanced if this work had been presented in terms of its original application rather than by means of artificial examples (Howard 1960, 1971). Of course, on a broader scale this story makes one painfully aware of how thinly our professional journals cover significant applications.

The Markov decision process and its extensions have now become principally the province of mathematicians. That is fitting because the process is a structure within which a host of interesting mathematical problems can be posed and answered. But I feel a sense of loss that this quite useful and general decision model has not seen a wider range of application.

APPENDIX

The following questions and answers were in the spirit of the discussions following the presentation:

Question 1. Concerning the sparsity of significant applications in the operations research literature, I feel that the main cause is the proprietary or classified nature of such applications and not the unwillingness of the journals to publish such work. Do you agree?

Answer 1. I feel that there are many reasons why applied work is not more frequently published in journals:

(a) Most editors are academics who can neither appreciate nor evaluate applications.
(b) The best ideas will be proprietary, at least for some time.
(c) There is little incentive for applied management scientists to publish their work in most firms.

Question 2. Would you speculate on why there have not been other significant applications of M.D.P.s? Is it because of the immense data requirements, or are there other reasons? Also, are you aware of any other significant applications?

Answer 2. The reasons for the lack of use of M.D.P.s are the same as the reasons for the lack of use of M.P.s. Speaking in generalities, the people who face the potential problems are not well-enough versed in the models to apply them. Conversely, the people who know the models seldom confront practical problems. I have discussed the general issue of application in a talk called "The Practicality Gap" (*Management Science*, **14** (7), March 1968, 503–507).

Finally, the only other significant application I know of M.D.P.s is concerned with metals futures and is also proprietary.

Question 3. Bellman's work on dynamic programming and iteration in policy space had already appeared in 1957. Were you aware of or motivated by any of his material?

Answer 3. I was not aware of Bellman's ideas on policy iteration; I was already writing my thesis in 1957. It is possible, however, that my adviser, Dr. Kimball, was aware of this work and that I may have been indirectly influenced.

Question 4. Are the Sears data and study still proprietary? If not, how can the results be obtained?

Answer 4. The Sears data have probably been buried in the archives of Sears and Arthur D. Little, Inc., for the last 20 years. Because I am not currently in touch with either organization, I cannot answer your question.

REFERENCES

Howard, R. 1960. *Dynamic Programming and Markov Processes*. Technology Press–Wiley, Cambridge, MA.

Howard, R. 1971. *Dynamic Probabilistic Systems* (two volumes). John F. Wiley & Sons, Inc., New York.

Originally published in Dynamic Programming & Its Applications. 1978. Martin L. Puterman, ed. Academic Press, Orlando, FL. Reprinted with permission.

ADDENDUM

Reflecting on this article after the passage of more than two decades, I can add a few remarks that might serve to bring it up to date. The explosion in data made available by the

general digital recording of consumer behavior such as the automatic recording of point-of-sale information in retail stores and in online purchasing now permits the application of Markov decision processes to a wide variety of problems on a routine basis. A modern database can produce in instants the transition probabilities that were so laboriously derived from Addressograph plates more than 40 years ago. It is very gratifying to see how technological progress has enhanced the practicality of problem solution in this area.