Sequential Decision Making for Intelligent Agents
Papers from the AAAI 2015 Fall Symposium

Open Questions for Building Optimal Operation Policies for Dam Management Using Factored Markov Decision Processess

Alberto Reyes, Pablo H. Ibargüengoytia, Inés Romero, David Pech, Mónica Borunda
Electrical Research Institute
Reforma 113, Col. Palmira, CP. 62490
Cuernavaca, Mexico

Abstract

In this paper, we present the conceptual model of a real-world application of Markov Decision Processes to dam management. The idea is to demonstrate that it is possible to efficiently automate the construction of operation policies by modelling the problem as a sequential decision problem that can be easily solved using stochastic dynamic programming. We will explain the problem domain and provide an analysis of the resulting value and policy functions. We will also present a useful discussion about the issues that will appear when the conceptual model to be extended into a real-world application.

1 Introduction

The construction of operation policies for dam management is a complex and time-consuming task that requires multidisciplinary expert knowledge. Usually, a group conformed of specialists such as meteorologists, hydrologists, civil engineers, and others are encouraged of performing this task. However, one of the main challenges is how to represent the uncertainty of the rain behavior to change the water level of the big storage container and the significance of keeping the dam safe.

A reservoir used solely for hydropower or water supply is better able to meet its objectives when it is full of water, rather than when it is empty. On the other hand, a reservoir used solely for downstream flood control is best left empty, until the flood comes of course. A single reservoir serving all three purposes introduces conflicts over how much water to store in it and how it should be operated. In basins where diversion demands exceed the available supplies, conflicts will exist over water allocations. Finding the best way to manage, if not resolve, these conflicts that occur over time and space are other reasons for planning.

In general, water resources planning and management activities are usually motivated by the realization that there are both problems to solve and opportunities to obtain increased benefits from the use of water and related land resources. However, the uncertain and intermittent nature of this resources make them hard to solve.

Among the most traditional techniques to deal with a planning and decision making problems under uncertainty the decision trees approach (Quinlan 1986) can be found. A decision tree represents a problem in such a way that all the options and consequences can be reviewed. They allow to quantify the costs of all possible results before making a decision. They also quantify the probability of occurrence of each event. The problem with this technique is that it only can be applicable when the number of actions is small and not all combinations of them are not possible. Other approaches such as influence diagrams or decision networks (Howard and Matheson 1984; Pearl 1988) allow representing a situation with many variables involved, identifying the source of the information required to make a decision, and modelling dynamic decisions in time. A limitation with this technique is that it exploits spatial and temporally as the problem grows up. The approaches based on classical planners like DRIPS (Haddawy and Suwandy 1994), WEAVER (Veloso et al. 1995) or MAXPLAN (Majercik and Littman 1998) introduce a non-deterministic representation of actions, represent conditional planning to estimate the maximum utility of a plan, explicitly represent exogenous events, or profit the main concepts under decision-theory and constraint logic programming. The problem with this techniques is that given that each planner solves different parts of the planning problem and that each one visualizes the problem from a different perspective, their integration results hard to implement.

Due to its compactness, ability to exploit the domain structure, and feasibility to integrate the features of other logic-based planners, the factored Markov Decision Processes (MDP) approach (Boutilier, Dean, and Hanks 1999) is used in this work. This approach introduces a series of methods under the concept of intentional (or factored) representations through which a combinational problem can be solved to make tractable a planning under uncertainty problem computationally speaking. Another feature is that it concentrates the main features of other planners in just one. Some related work to deal with a water resources management and planning problems be found in (Feinberg and Shwartz 2002; Loucks and van Beek 2005).

This paper is organized as follows: Section 2 presents a formalization of a simplified dam management problem. Section 3 provides a brief background about factored MDPs. Section 4 formalizes the dam problem in terms of a sequential decision problem represented as a factored MDP. Sec-

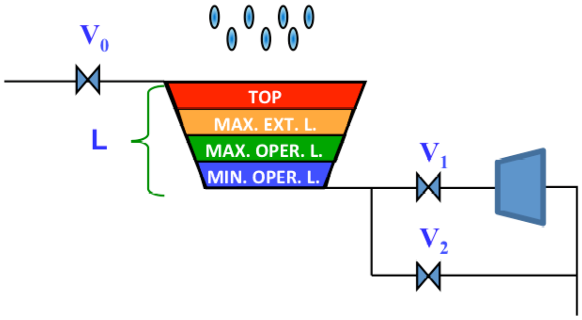

Figure 1: Simplified hydroelectric system.

tion 5 provides an analysis of the resulting value and policy functions for to test scenarios. Finally, conclusion and future directions are established in section 6.

2 Problem domain

Consider the problem of creating the best operation policies for the hydroelectric system described in Figure 1.

The system consists of: a reservoir; an inflow conduit, regulated by V0, which can either be a river or a spillway from another dam; and two spillways for outflow: the first penstock, V1, which is connected to the turbine and thus generates electricity, and the second penstock, V2, allowing direct water evacuation without electricity generation. In this way the reservoir has two inflow sources coming either from the inflow conduit or the rainfall and two outflow sources namely the two spillways. We quantize all flows to a unit of flow, L, and consider them as multiples of this unit. We consider the four reservoir levels MinOperL, MaxOperL, MaxExtL and Top and consider the transition from one level to the other above the bottom of the reservoir.

The unit L is the required amount of displaced water required to move from one level in the reservoir to another one and it is defined by

$$L = \frac{Q}{S}\Delta t\,, \tag{1}$$

where $Q = [m^3/s]$ is a unit of flow, $S = [m^2]$ is the surface of the reservoir and $\Delta t = [s]$ is a unit of time. Therefore the rainfall, LL, and the inflow and outflows are multiples of L,

$$LL = n_{LL}L\,, \tag{2}$$
$$Q_i = n_i L\,, \tag{3}$$

where the subindex $i = 0, 1, 2$ and Q_0 is the inflow at V_0 and Q_1 and Q_2 are the outflows at V_1 and V_2 and $n_{LL}, n_i \in (0, N)$. Given this, we classify the rainfall as follows

$$LL = \begin{cases} \text{No rain;} & LL = 0\,, \\ \text{Moderate rain;} & LL = L\,, \\ \text{Heavy rain;} & LL \geq L\,. \end{cases} \tag{4}$$

The aim of the optimization process is to control V_0, V_1 and V_2 such that the water volume in the reservoir is as much as possible in the optimum level, namely the MaxOperL level, given the rainfall conditions. This optimization process creates the optimal operation conditions of the dam and becomes a decision maker depending on the meteorological and hydrological conditions of the site.

With the idea of having four interconnected dams, now the system is modeled by the interconnection of the previous set-up and three more copies as shown in Figure 2.

This process is more complex since the decision maker takes into account the inflows and outflows of each dam, the rainfall conditions, which may be different from one dam to another because they are located at different sites, in addition to keep consulting the operation policies of each of them to maintain the four dams as close as possible to the MaxOperL levels.

3 Factored Markov decision processes

A Markov decision process (MDP) (Puterman 1994) models a sequential decision problem, in which a system evolves over time and is controlled by an agent. At discrete time intervals the agent observes the state of the system and chooses an action. The system dynamics are governed by a probabilistic transition function Φ that maps states **S** and actions **A** (both at time t) to new states **S'** (at time $t + 1$). At each time, an agent receives a scalar reward signal R that depends on the current state s and the applied action a. The performance criterion that the agent should maximize considers the discounted sum of expected future rewards, or value V: $E[\sum_{t=0}^{\infty} \gamma^t R(s_t)]$, where $0 \leq \gamma < 1$ is a discount rate. The main problem is to find a control strategy or *policy* π that maximizes the expected reward V over time.

For the discounted infinite-horizon case with any given discount factor γ, there is a policy π^* that is optimal regardless of the starting state and that satisfies the *Bellman* equation (Bellman 1957):

$$V^*(s) = max_a\{R(s,a) + \gamma \sum_{s \in \mathbf{S}} \Phi(a, s, s')V^*(s')\} \tag{5}$$

Two methods for solving this equation and finding an optimal policy for an MDP are: (a) dynamic programming (Puterman 1994) and (b) linear programming.

In a factored MDP, the set of states is described via a set of random variables **S** = $\{X_1, ..., X_n\}$, where each X_i takes on values in some finite domain *Dom*(X_i). A state **x** defines a value $x_i \in$ *Dom*(X_i) for each variable X_i. Thus, as the set of states **S** = *Dom*(X_i) is exponentially large, it results impractical to represent the transition model explicitly as matrices. Fortunately, the framework of dynamic Bayesian networks (DBN) (Dean and Kanazawa 1989; Darwiche and M. 1994) gives us the tools to describe the transition model concisely. In these representations, the post-action nodes (at the time $t+1$) contain smaller matrices with the probabilities of their values given their parents' values under the effects of an action. For a more detailed description of factored MDPs see (Boutilier, Dean, and Hanks 1999).

4 Factored MDP Problem specification

The MDP problem specification consists in establishing the set of states, set of actions, immediate reward function, and an state transition function. A simplified space state

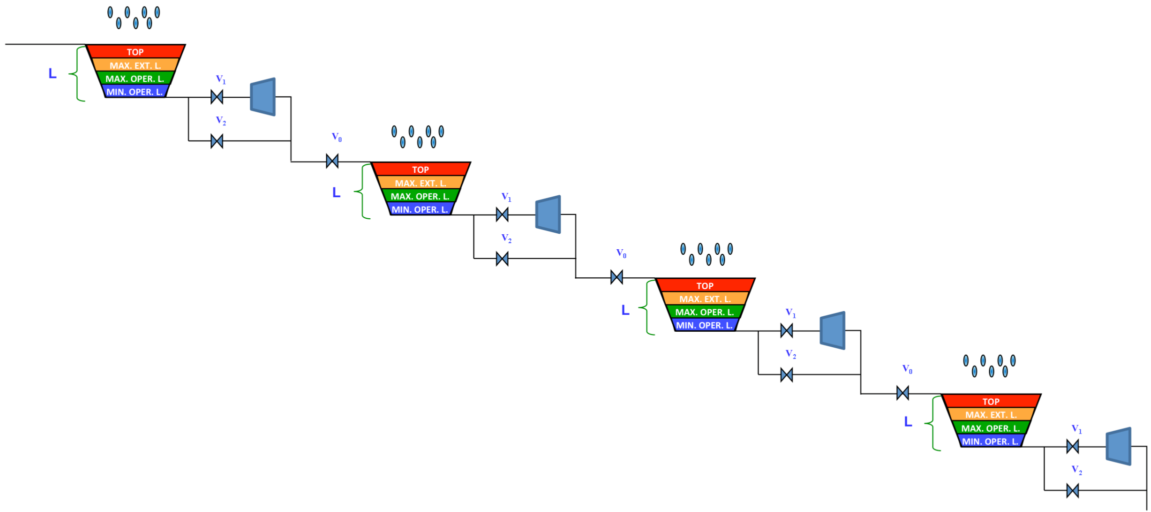

Figure 2: Multiple dam system.

is composed of the possible values for the variables rain intensity (Rain) and dam level (Level). The variable Rain can take three different nominal values: Null, Moderate and Intense, and Level can take eight values MinOperL1, MinOperL2, MaxOperL1, MaxOperL2, MaxExtL1, MaxExtL2, Top1 and Top2. As a consequence, the state space dimension will be 24 ($3^1 * 8^1$) with the values combination shown in Table 1.

The possible actions are given in terms of the operations permitted on the control elements (valves or gates) V0, V1 and V2. V0 is the inflow conduit valve, V1 is the spillway to the hydraulic turbine, and V2 is the direct evacuation gate or turbine bypass valve (see section 2 for details.). For this simple example, the actions could be close or open a control element. The possible actions combination are shown in Table 2.

In order to set a reward function for this problem, consider that keeping a dam level of MaxOperL1 or MaxOperL2 represents 100 economic units (best case). If the level is MaxExtL1 or MaxExtL2 the immediate reward value is -0 (irrelevant), if the level is Top1 or Top2 the reward value is -100 (worst case), and finally if the level is MinOperL the reward received is -50 (bad). In general terms, the dam levels around the MaxOperL value are awared while the levels nearby the top limits are penalized significantly. Notice that the reward rate is independent of the rainfall intensity. The decision tree of Figure 3 shows graphically the reward distribution as a function of the dam levels.

The transition model is represented using a two steps dynamic bayesian network for each action. Figure 4 shows the action A1 in three different scenarios. In all three scenarios the level of the dam is established in the **MaxOperL level** (red mark on interval 3-4). In the left case, Rain is instantiated in **Null** value (blue mark in interval 1). In this scenario, the level of tha dam has no change. Level_1 is maintained in interval 3-4 (orange mark). In the center scenario, Rain is **Moderate** (interval 1-2) and the dam level increments to reach interval 4-5 with 80% probability. Finally, in the rigth scenario, the Rain is intense (interval 2-3) and the dam level is incremented to interval 5-6 with 80% probability. The model and the inferences are visualized using the Hugin package (Andersen et al. 1989).

5 Experimental results

Given the reward and transition functions, we solved the factored MDP to obtain the policy and expected utility functions. As factored MDP solver we used the SPI (Planning under Uncertainty System in spanish) tool which managed to estimate 14 different values with a value iteration implementation. The minimum value obtained for this MDP model was 136.5961 and the maximum value was 771.2321. These values are represented with labels and colors in Figure 5 (left). The utility values for each state are shaded with light green color when the value is optimal and with darker colors as the value decreases. In the same figure, we show the resulting policy (recommended action) using labels with the action id. In all cases we used a discount value=1.

For example, the effect of the action A1, framed with red in Figure 5 (left) and represented by the symbol $\oplus$ (pointed with a red asterisk) in Figure 5 (right), means that the policy effect on the level is null due that the recommended action will keep the level. This is because the rain has no influence on the system. In the case of having a low level of the dam, independently of the rain condition, the policy function will recommend action A5 with the effect of increasing the level $\uparrow$. In this case, the influence of the rain could increase the level in one step or two depending on the rain intensity. In the opposite case when the dam has a high level the action effect of A4 will decrease $\downarrow\downarrow$ the level of the dam according to the rain intensity. The policy effects on the dam level are shown in table 3.

In order to show the system behavior, we followed the recommended actions from an random initial state until achieving the goal state under two different scenarios: 1-dam pro-

Table 1: State space for a single-dam system. The description is in terms of the values for the Level and Rain variables.

State ID	Description	State ID	Description	State ID	Description
1	MinOperL1, Null	9	MinOperL1, Moderate	17	MinOperL1, Intense
2	MinOperL2, Null	10	MinOperL2, Moderate	18	MinOperL2, Intense
3	MaxOperL1, Null	11	MaxOperL1, Moderate	19	MaxOperL1, Intense
4	MaxOperL2, Null	12	MaxOperL2, Moderate	20	MaxOperL2, Intense
5	MaxExtL1, Null	13	MaxExtL1, Moderate	21	MaxExtL1, Intense
6	MaxExtL2, Null	14	MaxExtL2, Moderate	22	MaxExtL2, Intense
7	Top1, Null	15	Top1, Moderate	23	Top1, Intense
8	Top2, Null	16	Top2, Moderate	24	Top2, Intense

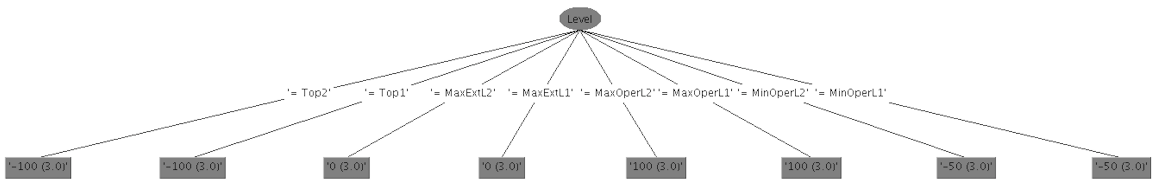

Figure 3: Reward function.

Table 2: Action space. The recommended actions are open or close the valves V0, V1 or V2.

Action ID	V0	V1	V2
A1	Close_Valve	Close_Valve	Close_Valve
A2	Close_Valve	Close_Valve	Open_Valve
A3	Close_Valve	Open_Valve	Close_Valve
A4	Close_Valve	Open_Valve	Open_Valve
A5	Open_Valve	Close_Valve	Close_Valve
A6	Open_Valve	Close_Valve	Open_Valve
A7	Open_Valve	Open_Valve	Close_Valve
A8	Open_Valve	Open_Valve	Open_Valve

cess and 4-dam process. In this demonstration we obtained each next state from the transition state function with the maximum probability of occurrence.

In the first case, one dam was set up under the minimum operation level (level=MinOperL) and with no rain (rain=Null). The policy was applied for a time horizon of twenty steps to observe the utility trend through the states transited. Figure 6 (up) shows how the dam starts with a minimum operation level with an expected utility=427.68 units, in the next step the system reaches a state with utility=588.30 units, and in a third step it reaches the maximum utility value =771.23 units.

In a second case, we initialized multiple dams at the conditions shown in Table 4, the algorithm is executed to check the optimization twenty executions.

As shown in Figure 6 (down) the Dam1 (blue line), which started with a low level, reached the optimum value in 3 steps. The Dam2 (green line) started in a goal state (optimal value) and remains. The Dam3 (red line), which started with

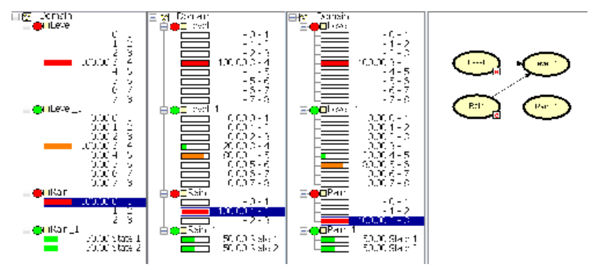

Figure 4: Transition model for action A1.

Table 3: Policy effects on the dam level according to the rain condition.

ID	Rain = Null	Rain = moderate	Rain = intense
A1	⊕	↑	↑↑
A2	↓↓	↓	⊕
A3	↓	⊕	↑
A4	↓↓↓	↓↓	↓
A5	↑	↑↑	↑↑↑
A6	↓	⊕	↑
A7	⊕	↑	↑↑
A8	↓↓	↓	⊕

max extraordinary level , and it achieves the maximum utility in 2 state transitions. Finally, the Dam4 (light blue line) that was initialized with a high level got its optimal value in 3 steps.

6 Discussion and future work

In this paper, we showed the conceptual model of a real-world application of Markov Decision Processes to dam management. The idea was to demonstrate that it is possible

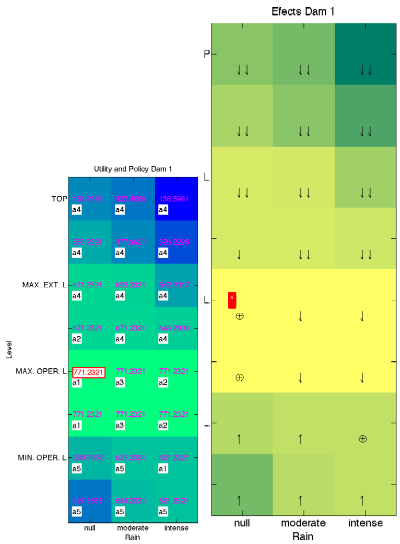

Figure 5: (left) Value and utility functions . Dark colors represent low expected utility values and light colors represent high utility values. (right) Effects of the policy on the dam level at different rain conditions. Refer table 3 for symbols interpretation.

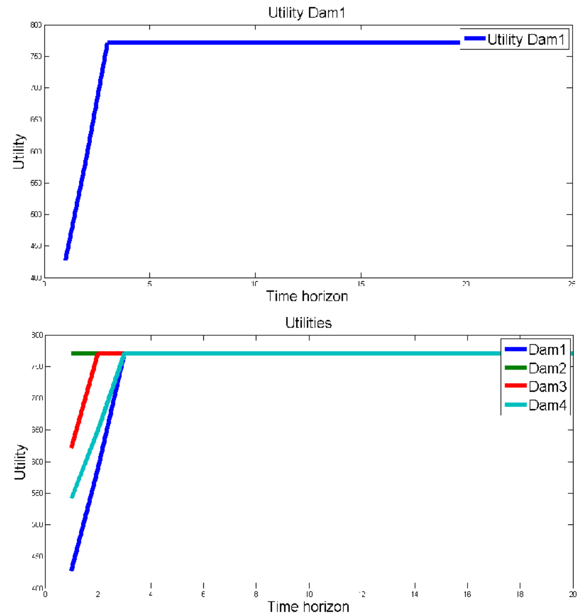

Figure 6: Utility plot. (up) Utility for a single-dam system. (down) Utility for multiple dams

Table 4: Initial states and expected utility values for a multiple dam system. The states variables are Level and Rain.

Dam	Level	Rain	Utility
1	MinOperL1	Null	427.68
2	MaxOperL2	Moderate	771.23
3	MinOperL2	Intense	621.23
4	MaxExtL2	Intense	542.20

to efficiently automate the construction of operation policies by modelling the problem as a sequential decision problem that can be easily solved using stochastic dynamic programming. We provided an analysis of the resulting value and policy functions in a conceptual hydroelectric process and showed how a single-dam system or a multiple-dam system can easily achieve an optimal operation state.

Due that this is a demonstration of the feasibility to use the framework of factored MDPs to solve problems in a hydroelectric domain and that several assumptions were made, there are still many challenges to face, particularly when the problem grows to a real-world application. Some open questions that could lead to new directions towards the solution of the water planning and management problems stated here are:

- How useful could result the use of problem abstractions or domain transforms in combination with the use of compact representations from the AI community?
- How to optimize jointly a multiple-dam system without losing compactness in the representation (remind that here we have solved four dams in an isolated way).
- Would it be possible to jointly optimize a set of n dams using local optimizations?
- Could reinforcement learning algorithms based on past experiences improve the trade-off exploration vs exploitation in this problem?
- What benefits can be obtained formulating this problem in terms of a decentralized MDP?

We are currently involved in a project for the Grijalva hydroelectric system in Mexico where we have started facing these and other challenges. Answering these questions would help us give steps a-head towards a very efficient AI approach to deal with water management problems and related domains.

7 Acknowledgments

Authors are especially grateful to the Electrical Research Institute-Mexico (IIE), the National Institution of Higher Education, Science, Technology and Innovation-Ecuador (SENESCYT) under the PROMETEO program, and the National Institute of Energetic Efficiency and Renewable Energies-Ecuador (INER) for the support and sponsorship in this research.

References

Andersen, S. K.; Olesen, K. G.; Jensen, F. V.; and Jensen, F. 1989. Hugin: a shell for building bayesian belief universes for expert systems. In *Proc. Eleventh Joint Conference on Artificial Intelligence, IJCAI*, 1080–1085.

Bellman, R. 1957. *Dynamic Programming*. Princeton, N.J.: Princeton U. Press.

Boutilier, C.; Dean, T.; and Hanks, S. 1999. Decision-theoretic planning: structural assumptions and computational leverage. *Journal of AI Research* 11:1–94.

Darwiche, A., and M., G. 1994. Action networks: A framework for reasoning about actions and change under understanding. In *Proceedings of the Tenth Conf. on Uncertainty in AI, UAI-94*, 136–144.

Dean, T., and Kanazawa, K. 1989. A model for reasoning about persistence and causation. *Computational Intelligence* 5:142–150.

Feinberg, E. A., and Shwartz, A. 2002. *Handbook of Markov Decision Processes, Methods and Applications*. Kluwer. chapter Water Reservoir Applications of Markov Decision Processes.

Haddawy, P., and Suwandy, M. 1994. Decision-theoretic refinement planning using inheritance abstraction. In ed K. Hammond., ed., *Proceedings of the Second International Conference on Artificial Intelligence Planning Systems*. Menlo Park, CA: AAAI Press.

Howard, R. A., and Matheson, J. E. 1984. *Influence Diagrams*. Principles and Applications of Decision Analysis. Menlo Park, CA: Howard, R. A. and Matheson J. E. Editors.

Loucks, D. P., and van Beek, E. 2005. *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*. United Nations Educational, Scientific and Cultural Organization (UNESCO). chapter Dynamic Programming.

Majercik, S. M., and Littman, M. 1998. Maxplan: A new approach to probabilistic planning. In eds. R. Simmons, and Smith, S., eds., *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems*, 86–93. Menlo Park, Calif.: AAAI Press.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, San Francisco, CA.

Puterman, M. 1994. *Markov Decision Processes*. New York: Wiley.

Quinlan, J. 1986. Induction of decision trees. *Machine Learning* 1(1):81–106.

Veloso, M.; Carbonel, J.; Perez, A.; Borrajo, D.; Fink, E.; and Blythe, J. 1995. Integrating planning and learning: The prodigy architecture. *Journal of Experimental and Theoretical AI* 1:81–120.