Intelligent Control of Closed-Loop Sedation in Simulated ICU Patients

Brett L. Moore, Eric D. Sinzinger, Todd M. Quasny, and Larry D. Pyeatt

Texas Tech University
Computer Science Department
Box 43104
Lubbock, TX 79409-3104
{moore,sinzinge,quasny,pyeatt}@cs.ttu.edu

Abstract

The intensive care unit is a challenging environment to both patient and caregiver. Continued shortages in staffing, principally in nursing, increase risk to patient and healthcare workers. To evaluate the use of intelligent systems in the improvement of patient care, an agent was developed to regulate ICU patient sedation. A temporal differencing form of reinforcement learning was used to train the agent in the administration of intravenous propofol in simulated ICU patients. The agent utilized the well-studied Marsh-Schnider pharmacokinetic model to estimate the distribution of drug within the patient. A pharmacodynamic model then estimated drug effect. A processed form of electroencephalogram, the bispectral index, served as the system control variable. The agent demonstrated satisfactory control of the simulated patient's consciousness level in static and dynamic setpoint conditions. The agent demonstrated superior stability and responsiveness when compared to a well-tuned PID controller, the control method of choice in closed-loop sedation control literature.

Introduction

The Intensive Care Unit (ICU)[1] represents a challenging environment to patient and staff alike. ICU patients may experience high anxiety levels from the general environment, and effective sedation[2] is necessary to soothe the patient and to obliterate asynchronous breathing or movement that might interfere with adequate oxygenation (Kowalski & Rayfield 1999). In recent years, a shortage of ICU nurses has resulted in patient mortality (Aiken *et al.* 2002; Lasalandra 2001) among other negative outcomes. Norrie (1997) observed that the ICU nurse's greatest time expense was "direct nursing care," including the administration of intravenous sedating drugs. It is thus reasonable to conclude that automating some aspects of the intensive care environment (like patient sedation) can positively impact overall ICU patient care.

Reinforcement learning (RL) represents a relatively new framework for constructing and applying intelligent agents. RL merges ideas from stochastic approximation and optimal control theory with the traditional concept of the intelligent agent. Much of the current research in reinforcement learning has been dedicated to autonomous robotic applications (Russell & Norvig 1995; Sutton & Barto 1998). RL has demonstrated favorable results in the associated problem domains; however, the extent of RL's aptitude for other specialized planning tasks remains incompletely explored. Several encouraging works exist: Guallapalli demonstrated that RL could successfully perform closed-loop control of the benchmark peg-in-hole task (Gullapalli 1993), and Hu applied some of the founding principles of reinforcement learning to anesthesia control with favorable results (Hu, Lovejoy, & Shafer 1994).

[1] "Intensive Care Unit" is a general label for several critical care environments including the Medical, Surgical, Cardiac, and Burn ICUs.

[2] Sedation is a drug-induced depression of consciousness (Task Force on Sedation and Analgesia by Non-Anesthesiologists 1996).

Background

Target Controlled Infusion (TCI) systems were originally deployed in the operating room to aid in the intraoperative management of general anesthesia. TCI devices allow the clinician to select a target blood level of the drug; the system then dispenses the drug at a combination of bolus and maintenance infusion rates to achieve and maintain the desired level. The TCI system selects an infusion rate based on a precomputed drug-patient interaction model, and the clinician may be expected to provide a variety of model-specific patient parameters, which can include gender, height, weight, and age. TCI systems were initially evaluated in the clinical setting and found satisfactory (Swinhoe *et al.* 1998), safety and efficacy did not differ statistically from manually-controlled intravenous infusion systems (Hunt-Smith *et al.* 1999). Since target-controlled drug delivery systems have been introduced into the operating theater, clinicians have observed favorable patient outcomes, such as decreased intraoperative drug administrations and shortened postoperative arousal times (Theil *et al.* 1993; Servin 1998). TCI has since migrated from the surgical theater to the critical care arenas.

Despite the favorable outcomes associated with the use of target-controlled drug infusion systems, inherent limitations exist. The drug-patient interaction models, known as pharmacokinetic/pharmacodynamic (PK/PD) models, characterize the distribution of the drug within the body (pharmacokinetics), as well as the effect of the drug (pharmacodynamics). The PK/PD models are highly specialized and are usu-

ally derived from a small number of relatively healthy patients (Vuyk *et al.* 1995). These models are thus challenged by disease pathologies, interactions with other pharmaceuticals, and other variabilities (even simple demographics) encountered in an actual patient.

Target-controlled infusion systems generally perform open-loop control. The standard TCI system is not equipped with a feedback mechanism and simply assumes the patient responds in "average" fashion. Modeling errors resulting from the above influences remain uncorrected. Albrecht concluded that closed-loop sedation in the ICU is a rational component of sound patient care (Albrecht *et al.* 1999). This conclusion is motivated in part by the rigors of the ICU environment; sedation in the intensive care unit is not always a static process. Recent research indicates that waking patients once a day may improve outcomes and shorten the duration of mechanical ventilation (Kress *et al.* 2000). Furthermore, long-term ICU patients may develop a tolerance to the drug and require greater doses to maintain sedative effect (Fulton & Sorkin 1995). As Barr reports, the optimal dosing of sedatives in the ICU is a dynamic, multivariate process (Barr *et al.* 2001).

Most of the current research in closed-loop sedation uses some form of electroencephalogram (EEG) as the control variable (Leslie, Absalom, & Kenny 2002; Absalom, Sutcliffe, & Kenny 2002; Sakai, Matsuki, & Giesecke 2000; Struys *et al.* 2001; Mortier *et al.* 1998). EEG is a well-studied indicator of the state of the central nervous system, and its analysis has been used in the diagnosis of neurological disorders, as well as the intraoperative monitoring of anesthetic efficacy. The bispectral index (a processed form of EEG) has been identified as a reliable indicator of sedation for some drugs. Bispectral index, or BIS™ [3], is a statistically-derived measure of consciousness in which a value of 100 indicates complete wakefulness, and 0 indicates an isoelectric brain state.

Modeling the ICU Patient

The first choice in modeling the patient is drug selection. From a purely engineering perspective, the choice of sedative agent is not particularly relevant. However, clinical concerns drive drug selection, and the choice of drug determines the patient interaction model. Ideal sedative characteristics for the ICU include titrability, stable hemodynamic responses, fast onset, and fast offset. Of the sedative agents meeting these criteria, midazolam and propofol are the frontrunners; propofol is more generally favored for ICU sedation (Fulton & Sorkin 1995; Ronan *et al.* 1995). Mortier observed that the bispectral index is a suitable control variable for control of propofol sedation (Mortier *et al.* 1998).

Several PK/PD models exist for propofol interaction. The Marsh model is widely studied (Marsh *et al.* 1991) and represents the patient as a collection of three compartments: *central*, *rapid*, and *slow*. The central compartment corresponds to the patient's apparent volume of blood (and the site of drug infusion); the rapid and slow compartments represent different collections of tissue, fat, and bone. The Marsh model is defined by the following diffusion constants:

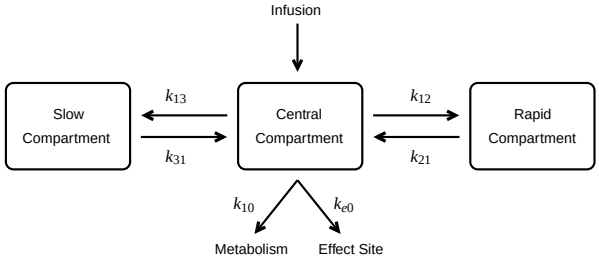

Figure 1: A four-compartment drug distribution model

- k_{10} metabolic clearance, $0.1190/\text{min}$,
- k_{12} central to rapid clearance, $0.1120/\text{min}$,
- k_{21} rapid to the central clearance, $0.0550/\text{min}$,
- k_{13} central to rapid clearance,$0.0419/\text{min}$,
- k_{31} rapid to central clearance, $0.0033/\text{min}$.

An improved model was obtained by adding a fourth compartment to model the drug's effect site in the brain: $k_{e0} = 1.2195/\text{min}$ (Struys *et al.* 2000; Marsh *et al.* 1991; Schnider *et al.* 1999). Drug transport between compartments is governed by a set of first-order differential equations. Using Ψ to represent the four-element vector of compartmental quantities, distribution of propofol can be modeled with the following equations:

$$\frac{\partial \psi_1}{\partial t} = \psi_2(t)k_{21} + \psi_3(t)k_{31} - \psi_1(t)\left(k_{10} + k_{12} + k_{13}\right) + I,$$

$$\frac{\partial \psi_2}{\partial t} = \psi_1(t)k_{12} - \psi_2(t)k_{21},$$

$$\frac{\partial \psi_3}{\partial t} = \psi_1(t)k_{13} - \psi_3(t)k_{31}, \text{ and}$$

$$\frac{\partial \psi_e}{\partial t} = \psi_1(t)k_{e0} - \psi_e(t)k_{e0}.$$

Figure 1 presents a block diagram of the model[4]. Figure 2 illustrates the model's response to a bolus of propofol at $t = 0$. The dynamic nature of the model, particularly the effect site's delay in following the central compartment, present interesting challenges for the controller.

Figures 1 and 2 illustrate pharmacokinetics, or drug distribution. The pharmacodynamics of propofol (the relation of effect-site concentration to bispectral index) were estimated using the following equation (Doi *et al.* 1997):

$$BIS_{measured} = -12.8 \cdot \upsilon_e + 93.6,$$

where υ_e was the effect site concentration of propofol.

[3] BIS™ is a trademark of Aspect Medical Systems, Newton, MA.

[4] Figure 1 also illustrates a physiological reality which limits controllers using intravenous infusions. Propofol cannot be mechanically removed from the patient's blood, and upward setpoint changes can only be accommodated by the controller choosing a "do-nothing" action.

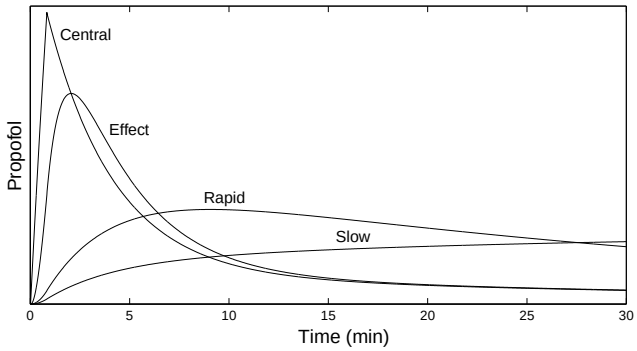

Figure 2: The model's response to a bolus of propofol

Methods

This section outlines the agent's architecture in terms of learning method and state space representation. These discussions also cover implementation specifics including the agent's inputs, actions, and other learning parameters. This section also describes the method of assessing agent performance.

Learning Technique

The objective was to assess the suitability of reinforcement learning to ICU sedation. Watkins' Q-learning technique (Watkins 1989) is appealing due to its mathematical soundness (Sutton 1987; Dayan 1992; Tsitsiklas & Van Roy 1996) and has become the *de facto* standard in the study of reinforcement learning. Q-learning is a temporal difference learning method characterized by model-free, off-policy learning. This technique learns an action-value function by iteratively advancing intermediate policies to optimality. The equation below, a Bellman expression of the Q-learning update rule, emphasizes an important aspect of the technique: the value of a state-action pair is expressed in terms of the value of successor states encountered under policy π.

$$Q^{\pi}(\boldsymbol{s},a) = \sum_{a} \pi(\boldsymbol{s},a) \sum_{s'} \mathcal{P}^{a}_{ss'} \left[\mathcal{R}^{a}_{ss'} + \gamma Q^{\pi}(\boldsymbol{s}',a) \right]$$

The basic Q-learning algorithm has since been improved to provide one of the strengths of dynamic programming, multi-step backups. The resulting algorithm, Q(λ), is shown in Figure 3. Q(λ) uses the *eligibility trace* and the parameter $\lambda \in (0,1)$ to vary temporal credit assignment for faster learning (Sutton & Barto 1998). After experimentation, one favorable reward propagation scheme was observed when $\lambda = 0.998$, while γ (the reward discounting parameter) equaled 0.80.

Inputs

The agent utilized two external inputs and four internal inputs to control the patient's level of consciousness. The desired bispectral index and the measured bispectral index were combined to form the control error, BIS_{error}. The agent also employed an internal instance of the Marsh-Schnider PK/PD model for estimates of the four compartmental propofol concentrations, $\Upsilon = [\upsilon_c, \upsilon_r, \upsilon_s, \upsilon_e]$.

Initialize $Q(\boldsymbol{s},a)$, $e(\boldsymbol{s},a) \forall \boldsymbol{s},a$ arbitrarily
Repeat (for each episode)
  Initialize $\boldsymbol{s},a$
  Repeat (for each step in the episode)
    Take action a, observe r, and transition to $\boldsymbol{s}'$
    Choose a' from $\boldsymbol{s}'$ using ε-greedy policy
    $a^* \leftarrow \text{argmax}_b\, Q(\boldsymbol{s}',b)$
    $\delta \leftarrow r + \gamma Q(\boldsymbol{s}',a^*) - Q(\boldsymbol{s},a)$
    $e(\boldsymbol{s},a) \leftarrow 1$
    $\forall \boldsymbol{s},a$
      $Q(\boldsymbol{s},a) \leftarrow Q(\boldsymbol{s},a) + \alpha\delta e(\boldsymbol{s},a)$
      If $a' = a^*$, $e(\boldsymbol{s},a) \leftarrow \gamma\lambda e(\boldsymbol{s},a)$
      else $e(\boldsymbol{s},a) = 0$
    $\boldsymbol{s} \leftarrow \boldsymbol{s}'$
    $a \leftarrow a'$
  Until $\boldsymbol{s}$ is terminal

Figure 3: A Q(λ) algorithm for learning optimal policies (Sutton & Barto 1998)

Actions

The agent could choose from the following set of propofol infusion rates $\mathcal{A} = \{0.0, 0.1, 0.5, 1.0, 2.0, 4.0\}$ (ml/min). To maintain consistency with existing precision syringe pumps, each action was considered atomic for a ten-second interval. Chosen actions were uninterruptible for this duration; once an action expired, the agent was free to choose another action from $\mathcal{A}$.

Reward Function

Reward is an immediate mapping from state to value and is the basis for agent's goal: "a reinforcement learning agent's sole objective is to maximize the total reward it receives in the long run (Sutton & Barto 1998)." Rewards may be viewed as positive reinforcements for favorable behavior or as negative reinforcements for unfavorable behavior. A successful agent was developed using the reward function below:

$$r = -\left|BIS_{measured} - BIS_{desired}\right| = -\left|BIS_{error}\right|.$$

This reward function was bounded to the interval $(-100, 0)$.

State Space Representation

Reinforcement learning tasks frequently use a function approximator to store the value function (the basis for making optimal decisions). Uniformly discretized tables are straightforward to implement and mathematically robust. Baird observes, "Algorithms such as Q-learning... are guaranteed to converge to the optimal answer when used with a lookup table (Baird 1995)." However, this approach is prone to tractability problems as the table size grows exponentially with increased dimension. The problem of concisely representing highly-dimensional state spaces is not new and has been the object of study for some time (Sutton & Barto 1998; Bellman, Kalaba, & Kotkin 1963). Regrettably, many alternatives fail to reliably converge to optimal policy when

Table 1: Input ranges and partitions

Input	Min	Max	Units	Knots
BIS_{error}	-20	20	BIS™	21
υ_c	0	1×10^5	μg	10
υ_r	0	1×10^5	μg	10
υ_s	0	1×10^5	μg	10
υ_e	0	8×10^4	μg	10

used with reinforcement learning (Thrun & Schwartz 1993; Boyan & Moore 1995; Sutton & Barto 1998).

One alternative is to enhance the discretized table with linear interpolation. The interpolated representation partitions the space in a manner similar to the discretized table. However, the partition boundaries now serve as control points for a first-order spline approximation. This technique assumes that regions between the partition boundaries may be satisfactorily fit in piecewise-linear fashion; hence, linear interpolation represents the value function in continuous space. Davies (1997) observed good performance with linearly interpolated value functions, though the state space was coarsely partitioned. Gordon (1995) developed a proof of convergence for a class of fitted temporal difference algorithms that includes linear interpolation.

To generalize linear interpolation, first consider the three-dimensional case. Given a regular bounding region in 3-space, the interpolated value $\hat{f}$ may be determined by:

$$\hat{f} = \sum_{i=0}^{2^3-1} w_i \cdot f_i,$$

where w_i is the interpolation weight at the i^{th} vertex, and f_i is the function's value at vertex p_i. The interpolation weight vector W is constrained such that $\hat{f} = f_i$ when the interpolating point is coincident with a cell vertex.

To apply the interpolation method to an arbitrary dimension d, the bounding region must be translated and scaled to the unit hypercube. Given that X represents the coordinates of the interpolation point $[x_0, x_1, \ldots, x_{d-1}]$, F is the set of vertex values $\left[f_0, f_1, \ldots, f_{2^d-1}\right]$, and P^i represents the set of coordinates for vertex i, $\left[p_0^i, p_1^i, \ldots, p_{2^d-1}^i\right]$, the following equation may then be applied:

$$\hat{f} = \sum_{i=0}^{2^d-1} f_i \prod_{j=0}^{d-1} \left(1 - \left|x_j - p_j^i\right|\right).$$

This interpolation process is $O(2^d)$, but interpolating over simplicial meshes (rather than hypercubes) can reduce the complexity to $O(d)$ (Munos & Moore 1999).

Table 1 summarizes the RL agent inputs, as well as the number of partitions for each dimension. The value function approximation consisted of 1.26×10^6 entries (one five-dimensional table of 210,000 entries for each of the six possible actions).

Performance Evaluation

The root-mean-squared error (RMS error) metric was used to evaluate the performance of the agent over three different sedation profiles. These profiles, summarized in Table 2, assessed the agent's control capacity under a variety of transient and steady state conditions. While the exact timings and target BIS™ levels were somewhat arbitrarily chosen, they represent fair and reasonable dosing events. For comparison, a proportional-integral-derivative (PID) controller was also constructed. The equation below summarizes a basic parallel PID control law in which a controlling signal, u, is computed from the weighted proportional, integral, and derivative terms:

$$u = K_p e + \frac{1}{K_i} \int e dt + K_d \frac{\partial e}{\partial t}.$$

The PID controller performed well for constants $K_p = 0.1$, $K_i = 600$, and $K_d = 0.8$ where u was the prescribed infusion rate and $e = BIS_{error}$.

Table 2: Tested sedation profiles

Profile	Interval (min)	Target BIS™
1	0 - 120	50
2	0 - 20	50
	20 - 80	80
	80 - 130	40
3	0 - 80	40
	80 - 160	94
	160 - 240	40
	240 - 320	94

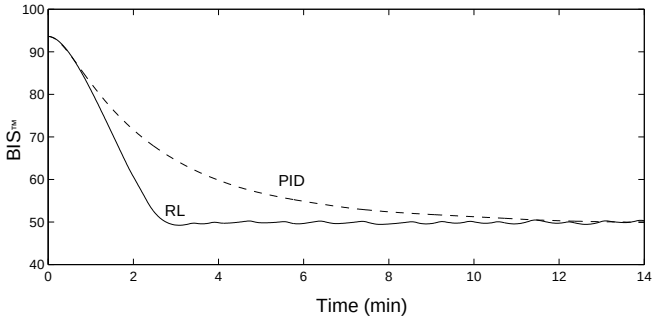

Figure 5: Observed responses to setpoint transition

Results and Discussion

Both the RL agent and the benchmark PID controller demonstrated good control characteristics. Figure 4 illustrates the behavior during one of the more challenging sedation profiles. As shown, the agent effected a quick transition to the initial setpoint, then maintained the setpoint with minimal deviation. At $t = 80$, a waking event was simulated and the agent correctly chose the "do-nothing" action allowing the patient to clear the drug and approach consciousness. At $t = 160$, deep sedation was targeted again. Although the patient's propofol load differed significantly from the initial transition, the agent achieved this setpoint equally well. Table 3 summarizes the RMS error for all tested profiles.

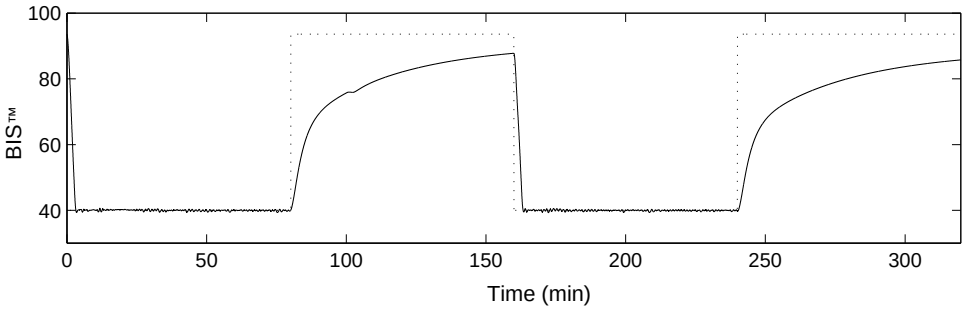

Figure 4: RL agent using linearly interpolated value function approximation (RMSE=13.72)

Table 3: RMS error observed in tested profiles

Profile	PID	RL
# 1	4.93	4.12
# 2	7.25	6.73
# 3	17.63	13.72

Benefits

The intelligent agent demonstrated an ability to reliably control the simulated patient's consciousness level through transition and steady-state periods (Figure 4). Table 3 illustrates satisfactory control over all tested scenarios. Figure 5 is particularly encouraging; short, controlled transitions are highly favored in sedation, but are a challenge for automation [5].

From an applied standpoint, provably optimal solutions (such as the RL agent) are appealing in life-critical applications. The bounded, tabular nature of the agent's world model assures that well-defined behaviors are associated with each state in the model. In addition, the temporal difference learning method deterministically yields the optimal control policy under reasonable assumptions. It is also important to note that constant-coefficient PID controllers (such as the benchmark controller and those used in current sedation research (Leslie, Absalom, & Kenny 2002; Absalom, Sutcliffe, & Kenny 2002)) are optimal only in the average sense. Undesirable control characteristics, like the oscillations reported by Leslie and Absalom, are not uncommon when PID control is applied to noisy, uncertain processes.

Limitations and Future Work

The RL agent studied in this work encountered three closely-related limitations. First, the agent assumed the patient would respond in the "average" fashion stipulated by the Marsh-Schnider model. Of course, the individual patient can be expected to vary from this ideal to some unknown degree. Second, the Marsh-Schnider PK model considers the patient weight to be an influential cofactor; however, this research was limited to simulated patients weighing 70 kg. Lastly, the agent's internal Marsh-Schnider model assumed no propofol was present in the patient's system prior to its administration of the drug. (Propofol is commonly used in the operating room, and the intensive care unit is a frequent destination for post-operative patients.) All of these limitations could be removed if the agent were equipped with online adaptivity: the capability to modify the learned value function in the presence of a systematic control bias.

The next step is to ready the agent for the rigors of control in the actual environment. Physiological systems are notorious for their uncertainty and noise, and the agent must be prepared to handle noisy observations, as well as biased patient responses. Once these measures are established, the agent can be evaluated outside of simulation. It is expected that the RL agent will regulate consciousness more effectively than existing techniques, and this hypothesis can be confirmed under controlled laboratory trials using human subjects (with appropriate review board approval and a physician's supervision). These trials would also present an opportunity to improve the agent and support enhancements, such as online learning.

[5] A recent study in sedation relied on the clinician to manually effect the initial loading of propofol (Leslie, Absalom, & Kenny 2002).

Conclusions

ICU patient sedation proved to be an interesting experiment in intelligent system control. The domain was challenging, but well-defined: prior clinical research demonstrated the efficacy of the bispectral index as a control variable, and existing pharmacokinetic/pharmacodynamic models provided a workable simulated patient. The RL agent demonstrated an ability to regulate the simulated patient's consciousness within acceptable limits, and the agent learned to dose the patient with the characteristics of good process control: rapid, well-managed transitions with stable steady-state responses. The RL agent compared favorably with the PID controller, a conventional control technique currently being applied in closed-loop sedation research.

Acknowledgments

This work was supported in part by NASA grant NNJ04HC19G.

References

Absalom; Sutcliffe; and Kenny. 2002. Closed-loop control of anesthesia using bispectral index performance assessment in pa-

tients undergoing major orthopedic surgery under combined general and regional anesthesia. *Anesthesiology* 96(1):67–73.

Aiken; Clarke; Sloane; Sochalski; and Silber. 2002. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA* 288(16):1987–1993.

Albrecht; Frenkel; Ihmsen; and Schuttler. 1999. A rational approach to the control of sedation in intensive care unit patients based on closed-loop control. *Eur J Anaesthesiol* 16(10):678–687.

Baird. 1995. Residual algorithms: Reinforcement learning with function approximation. In *Proc. 12th International Conference on Machine Learning*, 30–37. Morgan Kaufmann.

Barr; Egan; Sandoval; Zomorodi; Cohane; Gambus; and Shafer. 2001. Propofol dosing regimens for ICU sedation based upon an integrated pharmacokinetic-pharmacodynamic model. *Anesthesiology* 95(2):324–333.

Bellman; Kalaba; and Kotkin. 1963. Polynomial approximation–A new computational technique in dynamic programming: Allocation processes. *J Math Comput* 17(82):155–161.

Boyan, and Moore. 1995. Generalization in reinforcement learning: Safely approximating the value function. In *Advances in Neural Information Processing Systems 7*, 369–376. The MIT Press.

Davies. 1997. Multidimensional triangulation and interpolation for reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 9, 1005–1011. The MIT Press.

Dayan. 1992. The convergence of TD(λ) for general λ. *Machine Learning* 8:341–362.

Doi; Gajraj; Mantzaridis; and Kenny. 1997. Relationship between calculated blood concentration of propofol and electrophysiological variables during emergence from anaesthesia: Comparison of bispectral index, spectral edge frequency, median frequency and auditory evoked potential index. *Br J Anaesth* 78(2):180–4.

Fulton, and Sorkin. 1995. Propofol. An overview of its pharmacology and a review of its clinical efficacy in intensive care sedation. *Drugs* 50(4):636–57.

Gordon. 1995. Stable function approximation in dynamic programming. In *Proceedings of the Twelfth International Conference on Machine Learning*, 261–268. Morgan Kaufmann.

Gullapalli. 1993. Learning control under extreme uncertainty. In *Advances in Neural Information Processing Systems*, volume 5, 327–334. Morgan Kaufmann, San Mateo, CA.

Hu; Lovejoy; and Shafer. 1994. Comparison of some control strategies for three-compartment PK/PD models. *Journal of Pharmacokinetics and Biopharmaceutics* 22(6):525–550.

Hunt-Smith; Donaghy; Leslie; Kluger; Gunn; and Warwick. 1999. Safety and efficacy of target controlled infusion (Diprifusor) vs manually controlled infusion of propofol for anaesthesia. *Anaesth Intensive Care* 27(3):260–264.

Kowalski, and Rayfield. 1999. A post hoc descriptive study of patients receiving propofol. *Am J Crit Care* 8(1):507–13.

Kress; Pohlman; O'Connor; and Hall. 2000. Daily interruption of sedative infusions in critically ill patients undergoing mechanical ventilation. *N Engl J Med* 342(20):1471–7.

Lasalandra. 2001. *Norwood hospital staff shortage eyed in patient's death*. `http://www.bostonherald.com/news/local_regional/cari01152001.htm`: World Wide Web.

Leslie; Absalom; and Kenny. 2002. Closed loop control of sedation for colonoscopy using the bispectral index. *Anaesthesia* 57(7):690–709.

Marsh; White; Morton; and Kenny. 1991. Pharmacokinetic model driven infusion of propofol in children. *Br J Anaesth* 67(1):41–8.

Mortier; Struys; De Smet; Versichelen; and Rolly. 1998. Closed-loop controlled administration of propofol using bispectral analysis. *Anaesthesia* 53(8):749–754.

Munos, and Moore. 1999. Variable resolution discretization for high-accuracy solutions of optimal control problems. In *IJCAI*, 1348–1355.

Norrie. 1997. Nurses' time management in intensive care. *Nurs Crit Care* 2(3):121–125.

Ronan; Gallagher; George; and Hamby. 1995. Comparison of propofol and midazolam for sedation in intensive care unit patients. *Crit Care Med* 23(2):286–93.

Russell, and Norvig. 1995. *Artificial Intelligence*. Prentice-Hall.

Sakai; Matsuki; and Giesecke. 2000. Use of an EEG-bispectral closed-loop delivery system for administering propofol. *Acta Anesthesiologica Scandinavica* 44:1007–1010.

Schnider; Minto; Shafer; Gambus; Andresen; Goodale; and Youngs. 1999. The influence of age on propofol pharmacodynamics. *Anesthesiology* 90(6):1502–16.

Servin. 1998. TCI compared with manually controlled infusion of propofol: A multicentre study. *Anaesthesia* 53(Suppl 1):82–86.

Struys; De Smet; Depoorter; Versichelen; Mortier; Dumortier; Shafer; and Rolly. 2000. Comparison of plasma compartment versus two methods for effect compartment-controlled target-controlled infusion for propofol. *Anesthesiology* 92(2):399–406.

Struys; De Smet; Versichelen; Van De Velde; Van den Broecke; and Mortier. 2001. Closed-loop controlled administration of propofol using bispectral analysis. *Anesthesiology* 95(1):6–17.

Sutton, and Barto. 1998. *Reinforcement Learning: An Introduction*. MIT Press.

Sutton. 1987. Learning to predict by the method of temporal differences. Technical Report TR87-509.1, University of Massachusetts.

Swinhoe; Peacock; Glen; and Reilly. 1998. Evaluation of the predictive performance of a 'Diprifusor' TCI system. *Anaesthesia* 53(Suppl 1):61–67.

Task Force on Sedation and Analgesia by Non-Anesthesiologists. 1996. Practice guidelines for sedation and analgesia by non-anesthesiologists. *Anesthesiology* 84(2):459–471.

Theil; Stanley; White; Goodman; Glass; Bai; Jacobs; and Reves. 1993. Midazolam and fentanyl continuous infusion anesthesia for cardiac surgery: a comparison of computer-assisted versus manual infusion systems. *J Cardiothorac Vasc Anesth* 7(3):300–6.

Thrun, and Schwartz. 1993. Issues in Using Function Approximation for Reinforcement Learning. In *Proceedings of the 1993 Connectionist Models Summer School*. Lawrence Erlbaum.

Tsitsiklas, and Van Roy. 1996. An analysis of temporal difference learning with function approximation. Technical Report LIDS-P-2322, Massachusetts Institute of Technology.

Vuyk; Engbers; Burm; Vletter; and Bovill. 1995. Performance of computer-controlled infusion of propofol: An evaluation of five pharmacokinetic parameter sets. *Anesth Analg* 81(6):1275–1282.

Watkins. 1989. *Learning from Delayed Rewards*. PhD dissertation, Cambridge University, Computer Science Department.