REINFORCEMENT LEARNING FOR THE CONTROL OF BLOOD PRESSURE IN POST CARDIAC SURGERY PATIENTS

Ciprian SANDU[1], Dumitru POPESCU[2]

This paper addresses the problem of automatic control of the blood pressure in post cardiac surgery patients. The focus of this study is on showing that reinforcement learning may be suitable for the blood pressure regulation in post cardiac surgery patients, using directly the clinical data, without the need of an accurate patient model. A detailed and realistic model of a hypertensive patient is successfully developed and implemented. Firstly, we use this model in order to create closed loop control with a classical PID controller. Secondly, we use the model only as a source of pseudo-clinical data with reinforcement learning.

Keywords: blood pressure regulation, reinforcement learning, mathematical model, PID

1. Introduction

Since some of the natural control systems of the body are disrupted during a cardiac surgical intervention, the patients need drug administration in order to keep their mean arterial pressure within safe limits. There are many benefits to including automatic control in the regulation of mean arterial pressure (MAP) in post cardiac surgery patients. In such patients, we have the blood pressure measured and the infusion of the fast-acting vasodilator - sodium nitroprusside (SNP) is adjusted as necessary.

One of the benefits is that the precision of the amount and of the rate of the administrated vasodilator is significantly increased.

As a result, the goals of safety, reducing costs and human effort are met more rapidly and naturally.

The human body involves many complex feedback control systems that are collectively called as homeostasis, some of which may be disrupted during surgery (the term homeostasis is used by physiologists to mean maintenance of nearly constant conditions in the internal environment.

[1]Eng., PhD, Faculty of Automatic Control and Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: ciprian.sandu.cs@gmail.com.
[2]Prof., PhD, Faculty of Automatic Control and Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: popescu_upb@yahoo.com.

Essentially all organs and tissues of the body perform functions that help maintain these constant conditions. For instance, the lungs provide oxygen to the extracellular fluid to replenish the oxygen used by the cells, the kidneys maintain constant ion concentrations, and the gastrointestinal system provides nutrients).

Hence, postsurgery patients require intravenous administration of suitable drugs to maintain key physiological variables such as blood pressure within desired limits. These allowable limits are rather narrow and small excursions beyond them may lead to undesirable outcomes. Hence, introducing automatic control of key physiological variables is beneficial for better patient care and reducing workload of healthcare staff. The simplest and potent type of automatic controllers is based on the feedback concept. Automatic feedback control systems are designed to control crucial variables by adjusting manipulated variables such as drug infusion rate based on the measured feedback signal.

The standard approach in such a control problem is the use of classical PID control. The main disadvantage is that it requires a mathematical model. The system identification for accurately modeling human functions is far from easy. This paper focuses on presenting an alternative to the classical PID control, namely reinforcement learning. The main advantage is that a mathematical model of the involved human functions is not required. We verify that the important parameters or performances of the system are within satisfying ranges - when reinforcement learning is used in a closed loop (feedback based control). The regulation of mean arterial pressure (MAP) in post cardiac surgery patients is an example where automation is particularly attractive. In such patients, patient's blood pressure is measured and the infusion of the fast-acting vasodilator - sodium nitroprusside (SNP) is adjusted as necessary.

The paper is structured as follows: Section 2 presents the motivation of using automatic control for blood pressure regulation and reinforcement learning, the problem formulation, as well as the algorithm chosen.

Section 3 presents: how the clinical data is generated (instead of actual clinical data we have used synthetic data obtained with a simulated model of the patient), without the need of an accurate patient model for the regulation itself and the simulation results obtained by using the reinforcement learning algorithm for the control of the blood pressure, based on the clinical data. The model as a source of artificial data (that we use here) is not the same thing with the model as a part of the control strategy (which reinforcement learning makes unnecessary – in this context). Section 4 presents the conclusions.

2. Problem formulation

In this section we first show the motivation for using automatic control for blood pressure regulation.

Then we present the problem formulation and the way to overcome the inherent difficulties of the classical approach. Finally, we detail the solution in terms of the algorithms we used.

2.1. Motivation

The reasons for using automatic control of blood pressure in post cardiac surgery patients are obvious and range from reducing costs and work effort to enhancing the safety level for the patients: an automatic controller keeps the arterial pressure between desired limits much longer even than an experienced clinician. Our final goal is to control the arterial pressure of a patient who suffered cardiac surgery. For reaching this goal we use a regular feedback loop control. The standard approach in such a control problem is the use of classical PID control. The main disadvantage is that it requires a mathematical model. The system identification for accurately modeling human functions is far from easy.

We will present an example of such mathematical model. We will propose an improvement to it, we will use it in a closed feedback loop with a PID controller. Then we will formulate the problem in such way that will allow the use of reinforcement learning; the mathematical model will serve only as a source of artificial clinical data.

2.2. PID Control

A realistic model of patient in the context of blood pressure regulation has the following components: a drug response model, models for internal reflexes, measurement dynamics, random noise due to respiration, patient movements etc.

2.2.1. Drug response model

In this study the drug response model of Slate et al. (1980) is used:

$$\frac{\Delta P_d(s)}{I(s)} = \frac{Ke^{-T_i s}(1+\alpha e^{-T_c s})}{\tau s + 1}, \tag{1}$$

where: $\Delta Pd(s)$ is the change in MAP (mmHg)
$I(s)$ is the infusion rate of SNP (ml h^{-1}),
K gives the patient's sensitivity (a high value of K->sensitive patient),
α is a recirculation index,
τ is a time constant,
T_i is the initial transport delay and
T_c is the recirculation time.

2.2.2. Models for internal reflexes

There are internal reflexes in the human body to regulate blood pressure. Lee et al. (2005) identified that RAS (renin-angiotensin system) and BRS (baroreceptor reflex system) are necessary for a complete model of a patient for good blood pressure control. The former, as shown by Lee et al. (2005), "is an internal blood pressure buffering system that is activated when MAP drops below a threshold value. Through a series of chemical reactions from renin to angiotensin II, RAS can alter the total peripheral resistance of arterioles and hence increase blood pressure".

The typical range of threshold for activation of RAS is between 70 to 75 mmHg for the general population. Hahn et al. (2002) noted the existence of the threshold and range for RAS, but he did not use them. Lee's model used the threshold value 72 mmHg and the range of 50-110: the "switch" simulates the RAS threshold of 72 mmHg whereas the saturation block represents the range of MAP (50 - 110 mmHg) where RAS remains effective.

The model for RAS proposed by Lee is shown in Fig. 1:

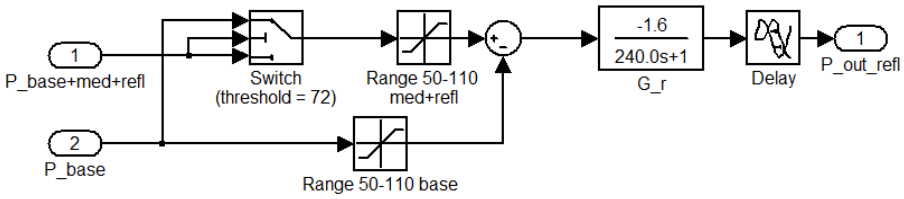

Fig. 1. Lee's model for RAS

The model for RAS proposed in a previous paper that will be used here in the feedback control loop for blood pressure regulation is shown in Fig. 2.

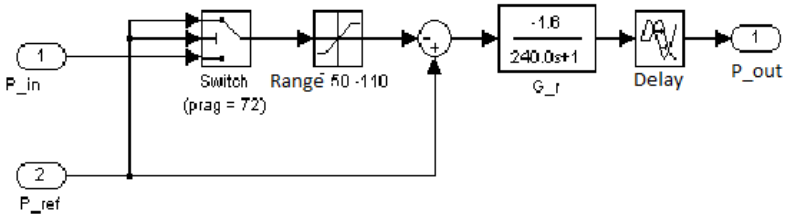

Fig. 2: The improved model for RAS

The closed loop used for this study is shown in Fig. 3:

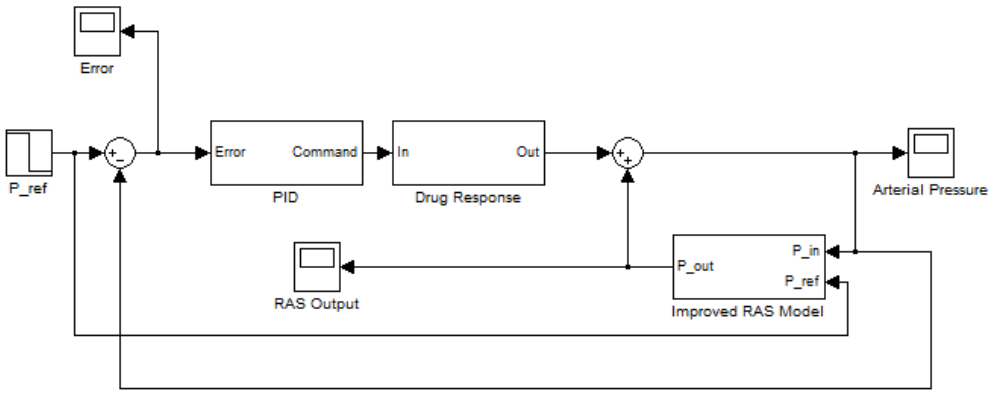

Fig. 3: The feedback control loop including the improved RAS model

The classical PID controller is shown in Fig. 4:

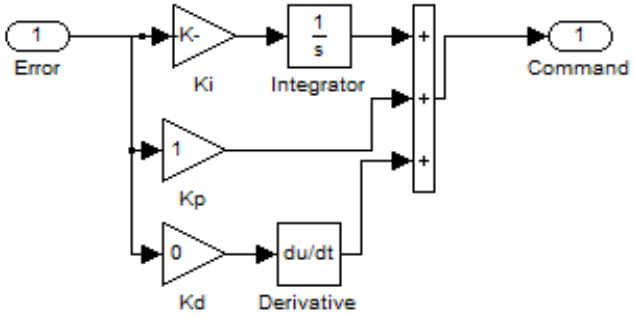

Fig. 4: The PI(D) controller used in the feedback control loop

In this example we will use a simpler PI controller, so the derivative part of the controller will be 0. This simplification does not affect in any way the relevance of our point. For the proportional part of the controller we will use the value 1 and for the integrator, 0.01.

When we plot the structure shown in Fig. 3, with the RAS block disconnected from the system, we obtain the result a detail of which is shown in Fig. 5 – we focused on the relevant part, the transition of the MAP (Mean Arterial Pressure) when the setpoint P_ref changes from 130 to 60 mmHg.

When we plot the structure shown in Fig. 3, since the 72 mmHg threshold is reached, the RAS system will become active – the new behavior is shown in Fig. 6.

In both responses one can notice that the stability of the system is not affected.

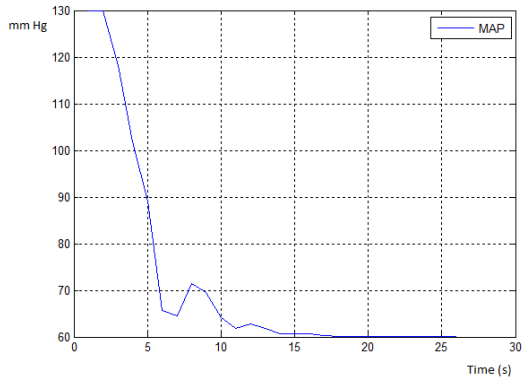

Fig. 5: The response of the system when the RAS block is disconnected

When the RAS block is connected, one can notice a slight negative overshoot of about 5 mmHg (is acceptable from a medical point of view: Lee argued that even a variation of 10 mmHg is acceptable from a medical point of view).

The settling time is about the same in both cases, therefore the presence of the RAS block does not affect this particular parameter.

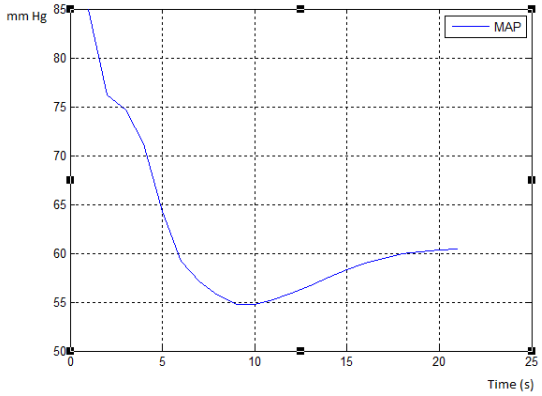

Fig. 6: The response of the system when the RAS block is connected (just as shown in Fig. 3)

2.3. Problem formulation

The control problem is easy to formulate as being defined by a set of states in which the environment (in this case the patient) may be observed, a set of actions that can be taken in order to influence the environment (the patient), a transition function linking actions to changes in state and a reward function (which evaluates the immediate control performance). The purpose is to make

sure the system reaches certain optimum states – with an acceptable error and overshoot.

So we have all the ingredients for formulating our problem as an optimal control problem, specifically a Markov Decision Process (a mathematic framework for optimal decision making in systems with uncertainty): a set of states of the environment, a set of actions to influence it, the transition function for linking the two and the reward function.

We have formulated the problem as a Markov Decision Process. We propose an alternative to the classical PID control, namely Reinforcement Learning (RL). The Markov property is essential in providing theoretical guarantees about RL algorithms.

Reinforcement Learning is an intelligent systems technique. It represents a mathematically robust method of achieving optimal control in systems challenged with noise, nonlinearity, time delay, and uncertainty. The main advantage is that a mathematical model of the involved human functions is not required.

The controller will be represented by a Reinforcement Learning agent.

The Reinforcement Learning controller applies its commands (actions) on a (simulated) patient. The policy within the agent receives 1 input: the control error, which is the difference between the measured arterial pressure and the pressure reference (or setpoint).

The states set x is defined by the error (the difference between the measured arterial pressure and the pressure reference) and it is represented as a discrete set of values:

$$X = [1,2,\ldots,200] \tag{2}$$

We provide the controller with a discrete set of SNP infusion rates (in ml per hour) u (actions); we call this set "the action set":

$$U = [1, 2, \ldots, 350] \tag{3}$$

These are all the theoretical possible values. In practice, the actual action values will be much below 100.

As a result of the action u_k, applied in the state x_k, the state changes to x_{k+1}, according to the transition function:

$$f : X \text{ x } U \rightarrow X :$$
$$x_k + 1 = f(x_k, u_k) \tag{4}$$

In order to train the RL controller, we need some measure of the utility of the states: the reward function provides the mechanism for deciding the state value. The controller receives the scalar reward signal r_{k+1}, according to the reward function r : X x U -> **R**

$$r_k + 1 = r(x_k, u_k) \tag{5}$$

The reward evaluates the immediate effect of action u_k, namely the transition from x_k to x_{k+1}, but in general does not say anything about its long-term effects.

The controller chooses the actions according to its *policy* $h : X \to U$, using:

$$u_k = h(x_k) \tag{6}$$

Given f and r, the current state x_k and the current action u_k are sufficient to determine both the next state x_{k+1} and the reward r_{k+1}. This is the Markov property, which is essential in providing theoretical guarantees about RL algorithms.

2.4. The solution

The main reason for using RL (reinforcement learning) control is that it uses the clinical data alone, without relying on the identification of an accurate model of the patient (which is a complex process).

The policy is the dependency between action and state and is defined (after an intense trial and error session to cover as many scenarios as possible) as follows:

$$u = \text{ref} - 145 + x^* \tag{7}$$

where ref is the pressure reference and x* is an artificial parameter that depends on the state x, that can take the following values:

[x/10], [x/5], x, 2x, 3x, 4x, 5x, 6x

where "[ ]" means the nearest integer upwards

To avoid the so-called "tunnel vision" problem in which the agent revisits a neighborhood of states and chooses the same actions repetitively, the action takes one of these values a random value 20% of the time and the value that maximizes the reward – 80% of the time.

The equation below summarizes the reward function we used for the closed loop arterial pressure control problem to penalize the agent when the measured arterial pressure was off target (the reasons for choosing this particular reward function are complex, they range from detailed calculations to empirical results observed for a large number of scenarios and they are not to be detailed here):

$$\begin{aligned} r(x, u) = {} & 120000 - 1000(x + xx'')/2 \\ & 1000 + 10u - xx'' - x \end{aligned} \tag{8}$$

where *x"* is the previous state.

The two values depend on how large the state we are in is. The first one is for small values (lower than 10), the second one for big values (10 or above).

We applied a SARSA algorithm to estimate the value function for each of the SNP infusion rates.

Equation 9 presents the SARSA update rule:

$$Q(x,u) = (1-\alpha)Q(x,u) + \alpha[r(x,u) + \gamma Q(x',u')] \quad (9)$$

where *u', x'* are the next action and respectively the next state.

Each time the agent completes one experiment with its environment (*i.e.*, chooses an action and receives a reward), the Q function is incrementally adjusted.

The duration of an action is set to 100 seconds (best results being obtained this way). We mention that the way the algorithm has been built, it permits total control over the duration of each action (so one can modify it freely if desired).

3. Results

In this section we present the results we have obtained by using the reinforcement learning approach on artificially generated data. Firstly we show how the data is generated and secondly we discuss the performances of the control loop.

3.1. Artificial generation of the clinical data

We use the same model of a patient as earlier, but this time with the only purpose of generating the input data (instead of gathering the data from a real patient). This is why we call the generation of the data "artificial".

The closed loop used for this study is shown in fig. 7:

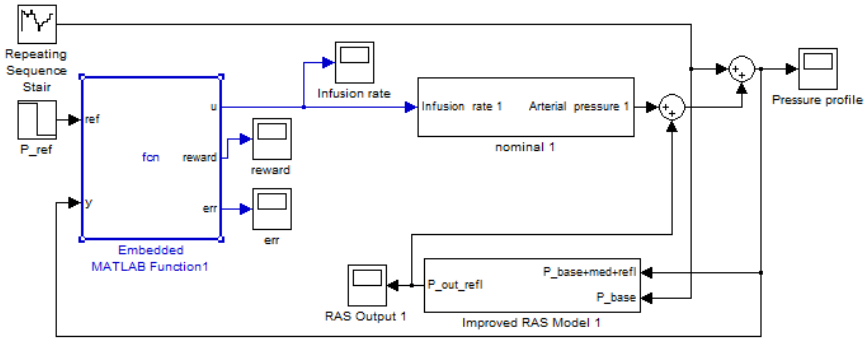

Fig. 7. The feedback control loop including the RAS model

3.2. Results

In Fig. 8 we show the behavior of the arterial pressure when controlled with the SARSA algorithm:

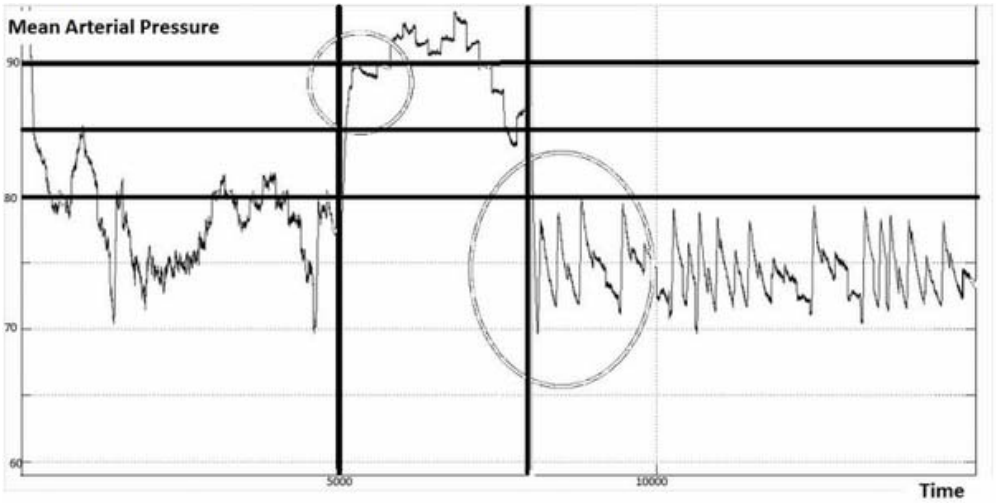

Fig. 8. The arterial pressure when controlled with the SARSA algorithm

On the vertical axis we have the arterial pressure.
On the horizontal axis we have the time.

We have highlighted 2 important moments with circles and with vertical lines:

5000 seconds – when the agent switching from learning (random actions) to applying what it learned (20% rand actions, 80% intelligent actions)

8000 seconds – when the pressure reference changes from 90 to 70 mmHg.

We have highlighted
The upper horizontal line the 90 mmHg level,
The middle horizontal line the 85 mmHg level,
The lower horizontal line the 80 mmHg level.

Between the 5000 and the 8000 seconds moments, the error is greater than 5 mmHg only once, and not with much (it is acceptable from a medical point of view). The behavior after 8000 seconds is due to the RAS reflex: when the pressure goes below 72 mmHg, the reflex increases the pressure. Again, the error is acceptable and the overshoot within desired limits.

4. Conclusions

Control of blood pressure regulation in post cardiac surgery patients using reinforcement learning is studied. We used reinforcement learning algorithms for the controller in order to control the mean arterial pressure in post cardiac surgery patients. Such an approach does not require a mathematical model of the patient as part of the regulation approach. A detailed and realistic model for blood pressure regulation is used for generating clinical data by including the drug response model of Slate et al. (1980) and models for internal reflexes (namely RAS) of the body.

The purpose of this report is to highlight the state of art in automatic blood pressure regulation, to propose alternative techniques to do it and the performances obtained with these alternative techniques.

The state of art of blood pressure regulation consists in the classical feedback control loop using a PID controller. PID control is suitable for this context, as it meets the requirements of being simple and easy to implement. But it does require a mathematical model of the patient, which is rather difficult to implement. We showed that we can obtain good results using Reinforcement Learning, which does not require a mathematical model of the patient. Therefore, Reinforcement Learning is an elegant solution, as by adopting it, the difficult modelling is no longer necessary, but the desired performances are still met.

The research continues in the direction of refining the algorithms and of finding even more methods that have good results in controlling the blood pressure for patients who have suffered heart surgery.

R E F E R E N C E S

[1] *Brett L. Moore, PhD, Anthony G. Doufas, MD, PhD and Todd M. Quasny, MS,* Reinforcement Learning: A novel Method for Optimal Control of Propofol-Induced Hypnosis, International Anesthesia Research Society, **Anesthesia – Analgesia**, February 2011, Vol. 112, Number 2.

[2] *Brett L. Moore, PhD, Anthony G. Doufas, MD, PhD and Larry D. Pyeatt, PhD.* Reinforcement Learning: A novel Method for Optimal Control of Propofol-Induced Hypnosis. International Anesthesia Research Society, **Anesthesia – Analgesia**, February 2011, Vol. 112, Number 2.

[3] *Brett L. Moore, B.S., B.A., M.S.*, Reinforcement Learning for Patient-Specific Propofol Anesthesia: A Human Volonteer Study, 2010.

[4] *H.W. Lee, S. Lakshminarayanan, G.P. Rangaiah*. Models and Simple Controllers for Blood Pressure Regulation in Post Cardiac Surgery Patiens. **Journal of The Institution of Engineers, Singapore**. Vol. 45 Issue 6, 2005.

[5] *Slate, J. B., Sheppard, L. C., Rideout, V. C. and Blackstone, E. H*. Closed-loop nitroprusside infusion: Modeling and control theory for clinical application, 1980.

[6] *Hahn, J., Edison, T. and Edgar, T. F*. Adaptive IMC control for drug infusion for biological systems. Control Engineering Practice. Vol. 10. pp 45 – 56, 2002.

[7] *Lee, H. W*. Analysis of Models and Techniques for Blood Pressure Control in Patients. **B. Eng. Thesis**. National University of Singapore, 2004.
[8] *Castronovo, M., Maes, F., Fonteneau, R., Ernst, D*. Learning Exploration/Exploitation Strategies for Single Trajectory Reinforcement Learning. 10th European Workshop on Reinforcement Learning , JMLR: Workshop and Conference Proceedings vol:1-9, 2012
[9] *Arthur C. Guyton, John E. Hall*. Textbook of Medical Physiology – Eleventh Edition, 2006.
[10] *Tom De Smet, Michel M. R. F. Struys, Martine M. Neckebroek, Kristof Van den Hauwe, Sjoert Bonte, EricP.Mortier*, The Accuracy and Clinical Feasibility of a New Bayesian-Based Closed-Loop Control System for Propofol Administration Using the Bispectral Index as a ControlledVariable, 2008.
[11] *Lucian Busoniu, Robert Babuska, Bart De Schutter and Damien Ernst*. Reinforcement Learning and Dynamic Programming Using Function Approximators.