Control Engineering Practice 59 (2017) 194–203

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Optimal learning control of oxygen saturation using a policy iteration algorithm and a proof-of-concept in an interconnecting three-tank system

Anake Pomprapa*, Steffen Leonhardt, Berno J.E. Misgeld

Philips Chair for Medical Information Technology, RWTH Aachen University, Aachen, Germany

ARTICLE INFO

Article history:
Received 30 November 2015
Received in revised form
18 June 2016
Accepted 25 July 2016
Available online 17 August 2016

Keywords:
Policy iteration algorithm
Optimal control
Reinforcement learning
Control of oxygen saturation
Biomedical control system
Closed-loop ventilation

ABSTRACT

In this work, "policy iteration algorithm" (PIA) is applied for controlling arterial oxygen saturation that does not require mathematical models of the plant. This technique is based on nonlinear optimal control to solve the Hamilton–Jacobi–Bellman equation. The controller is synthesized using a state feedback configuration based on an unidentified model of complex pathophysiology of pulmonary system in order to control gas exchange in ventilated patients, as under some circumstances (like emergency situations), there may not be a proper and individualized model for designing and tuning controllers available in time. The simulation results demonstrate the optimal control of oxygenation based on the proposed PIA by iteratively evaluating the Hamiltonian cost functions and synthesizing the control actions until achieving the converged optimal criteria. Furthermore, as a practical example, we examined the performance of this control strategy using an interconnecting three-tank system as a real nonlinear system.

1. Introduction

Hypoxia, or oxygen deficiency, is a common consequence of respiratory insufficiency. If a patient with hypoxia is not properly treated on time, the prolonged state of impaired oxygenation can lead to potentially lethal conditions, such as cerebral hypoxia, cardiac malfunction, or multiple organ failure. The most effective therapy is to supply the patient with an increased oxygen fraction in the ventilated air, which is referred to as oxygen therapy (Tarpy & Celli, 1995). A proper control output of this therapy is the resetting oxygen content in the body, including peripheral oxygen saturation (SpO_2), arterial oxygen saturation (SaO_2) or partial pressure arterial oxygen (PaO_2), can be used in a negative feedback configuration. A single-input single-output (SISO) system can be formulated and a closed-loop control of oxygen therapy can be realized in clinical scenarios, which are unique, complex, and life-threatening.

To our knowledge, closed-loop ventilation for oxygen therapy developed gradually (Brunner, 2002) and was dependent on progress in control engineering. Early publications date back to 1975: this paper describes the control of SpO_2 (positioned at the ear) with a 'bang–bang' control of a controlling input fraction of inspired oxygen (FiO_2) and the results of limit cycles achieved in anesthetized dogs (Mitamura, Mikami, & Yamamoto, 1975). In 1985, a linear quadratic regulator (LQR) was shown for the dual control of oxygen and carbon dioxide (Giard, Bertrand, Robert, & Pernier, 1985). In 1987, a multiple-model adaptive control (MMAC) was applied for the control of SpO_2 in mongrel dogs and the results were compared with a proportional integral (PI) controller (Yu et al., 1987). In 1991, the multivariable inputs of FiO_2 and positive end-expiratory pressure (PEEP) were applied to control PaO_2 using a proportional–integral–derivative (PID) controller in four mongrel dogs (East, Tolle, MCJames, Farrell, & Brunner, 1991). In 2004, a non-linear adaptive neuro-fuzzy inference system (ANFIS) model was used to estimate shunt in combination with dynamic changes of blood gases for controlling PaO_2 (Kwok, Linkens, Mahfouf, & Mills, 2004). Further simulation of septic patients was carried out based on a hybrid knowledge/model-based advisory control. Recently, feedback-oriented oxygen therapy was presented in preterm infants, where the controlled variable was SpO_2 (Claure & Bancalari, 2009). In addition, based on the works in our research group, a proportional–integral (PI) controller with gain scheduling (Walter et al., 2009), a Smith predictor with an internal PI controller (Lüpschen, Zhu, & Leonhardt, 2009), a knowledge-based controller (Pomprapa, Misgeld, Lachmann, & Leonhardt, 2013), and, potentially, a self-tuning adaptive controller (Pomprapa, Pikkemaat, Lüpschen, Lachmann, & Leonhardt, 2010) and a funnel controller (Pomprapa, Alfocea, Göbel, Misgeld, & Leonhardt, 2014) can be used for this particular control problem. Moreover, in 2014, SpO_2 between 92% and 94% was targeted based on the

* Corresponding author.
E-mail address: pomprapa@hia.rwth-aachen.de (A. Pomprapa).

http://dx.doi.org/10.1016/j.conengprac.2016.07.014
0967-0661/

automatic adjustment of FiO_2 using a step change approach (Saihi et al., 2014). Based on a literature search, three different categories were found for the control system design and their advantages and disadvantages are given below.

A model-based approach is a traditional procedural design in the control community. Generally, it requires a mathematical description of the system behavior. The difficulty of this approach lies in the selection of an appropriate model structure for this particular problem and has to deal with model uncertainties based on varying conditions of the system. Therefore, self-tuning adaptive control was introduced for this control problem (Pomprapa et al., 2010). With this model-based approach, system performance and further corrective actions can be preanalyzed.

A model-free approach requires neither prior knowledge of the system nor system identification, which is a major advantage. Therefore, the time-consuming process for determining an individual model can be significantly reduced, which is certainly beneficial for a patient with severe hypoxia. In fact, the patient requires an immediate therapeutic action. Hence, this approach is beneficial in the real clinical scenarios. Some of the proposed control techniques were, for example, a knowledge-based controller (Pomprapa, Misgeld, Lachmann et al., 2013) or a funnel controller (Pomprapa, Alfocea, et al., 2014). Nevertheless, the drawback is the requirement for experience to reach a good performance.

An intelligent hybrid approach of combined model-based and model-free methods (Kwok et al., 2004) requires a knowledge and model-based advisory system for intensive care ventilation. The strength and the weakness have been mentioned for both model-based approach and model-free approach. The design of this approach was based on neuro-fuzzy system (ANFIS), which should closely resemble to mental operation. However, neither an animal experiment nor a clinical result has yet been reported.

For best patient benefits, we employ a model-free approach in this work. Therefore, we examine the application of a generally known control strategy to the control of oxygenation based on SaO_2 measurement. For the first time (first results on this new approach have been prepublished in Pomprapa, Mir Wais, Walter, Misgeld, & Leonhardt, 2015), we propose to use a reinforcement learning optimal control method, called the policy iteration algorithm (PIA) (Bhasin et al., 2013; Vamvoudakis & Lewis, 2010; Vrabie, Pastravanu, Abu-Khalaf, & Lewis, 2009), for this particular application. This control strategy should provide an optimal solution for not only treating hypoxia, but also avoiding hyperoxia (excess oxygen in the lungs) (Clark, 1974; Kallet & Matthay, 2013; Mach, Thimmesch, & Pierce, 2011). The PIA is classified as reinforcement learning with the actor–critic architecture framework, which conventionally learns behavior through interactions with a dynamic environment and a survey for the development of reinforcement learning can be found in Kaelbling, Littman, and Moore (1996). Based on this particular approach, it improves the effectiveness of learning power using a policy gradient for stochastic search for an optimal result. The technique is based on a recursive two-step iteration, namely policy improvement and policy evaluation. These steps are repeated iteratively until achieving the stopping criteria for the converged optimal solution (Vamvoudakis & Lewis, 2010; Vrabie et al., 2009). Practical implementations of PIA control in other areas include optimal-load-frequency control in a power plant (Wang, Zhou, & Wen, 1993), a DC–DC converter (Wernrud, 2007), adaptive steering control of a tanker ship (Xu, Hu, & Lu, 2007), and a hybrid system of a jumping robot (Suda & Yamakita, 2013). Therefore, it is motivated to apply this modern control strategy to biomedical applications and, specifically, to closed-loop ventilation therapy. Computer simulations are used to investigate the feasibility of this controller in oxygen therapy. Subsequently, a proof-of-concept is implemented in a real application for a nonlinear interconnecting three-tank system in order to control the water level in the last tank. In fact, the dynamics of this three-tank system resembles that of a patient with respiratory deficiency, i.e. nonlinear and with time-delay behavior.

In this article, Sections 2.1 and 2.2 provide the statement of the medical perspective and the formulation of optimal control problem. Section 3 presents a control system design using the PIA for nonlinear optimal control based on the Hamilton–Jacobi–Bellman (HJB) equation. The system identification of the cardiopulmonary system to identify a mathematical model for further design of the PIA controllers, and a simulation of this control strategy is proposed in Sections 4 and 5, respectively. A practical example of this controller is presented in Section 6 for controlling the water level in a nonlinear interconnecting three-tank system. Section 7 presents a discussion, and our conclusions are presented in Section 8.

2. Problem formulation

2.1. Statement of the medical perspective

The overall transfer function from the settings of inspired oxygen fractions FiO_2 to the oxygen saturation SaO_2 measured in arterial blood depends on many factors, including lung function and blood transport (which then depends on cardiac output, state of circulation, etc.). Let $G(s)=SaO_2(s)/FiO_2(s)$ be the transfer function describing the system under investigation. $G(s)$ certainly is nonlinear (e.g. due to the nonlinear saturation function, see West, 2011) and depends on many unknown and time-variant conditions (like e.g. fluid status, heart function, or many diseases). Often, for the individual patient requiring artificial ventilation, this model is not available. Furthermore, as there usually is no time for individual model selection and parameter identification, this has been the motivation to investigate the performance of control algorithms which do not require explicit models.

As an illustrative example and in order to roughly understand the dynamics, we provide the transfer function $G(s)$ of a female domestic pig weighing 34 kg, which we obtained during a laboratory experiment (Lüpschen et al., 2009). In this special case, respiratory distress was induced by repeated lung lavage (ARDS-model, acute respiratory distress syndrome; Ashbaugh, Bigelow, Petty, & Levine, 1967; Lachmann, Robertson, & Vogel, 1980).

2.2. Optimal control problem

We consider a specific class of nonlinear systems, namely affine nonlinear systems of the following form:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}) + g(\mathbf{x})\cdot u(t), \tag{1}$$

where $\mathbf{x}(t) \in \chi \subseteq \mathbb{R}^n$ denotes the states of the system in a vector form of n dimension, $u(t) \in \upsilon \subseteq \mathbb{R}$ represents the control input or FiO_2, and $\mathcal{F}: \chi \times \upsilon \to \mathbb{R}^n$ is Lipschitz continuous on $\chi \times \upsilon$, such that the state vector $\mathbf{x}(t)$ is unique for a given initial condition $\mathbf{x_0}$. Initially, it is assumed that the system is stabilizable. This type of model has been used to describe the dynamics of various plants, for example a robot manipulator (Sun, Sun, Li, & Zhang, 1999), a continuous stirred-tank reactor (CSTR) (Kamalabady & Salahshoor, 2009) and a non-interconnecting three-tank system (Orani, Pisano, Franceschelli, Giua, & Usai, 2011).

Let us consider a cost function $V(\mathbf{x})$ given by Eq. (2), which is to be minimized.

$$V(\mathbf{x}) = \int_0^{T_\infty} r(\mathbf{x}(\tau), u(\tau))d\tau, \tag{2}$$

let $r(\mathbf{x}, u)$ be determined by $Q(x) + u^T\mathbf{R}u$. If $Q(\mathbf{x}) \in \mathbb{R}$ is positive-