Lecture Notes in Networks and Systems 1042

Adam Martowicz
Michał Mańka
Krzysztof Mendrok *Editors*

Mechatronics – Industry-Inspired Advances

Springer

Lecture Notes in Networks and Systems 1042

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Adam Martowicz · Michał Mańka · Krzysztof Mendrok
Editors

Mechatronics – Industry-Inspired Advances

Springer

Editors

Adam Martowicz
AGH University of Krakow, Faculty of Mechanical Engineering and Robotics, Department of Robotics and Mechatronics
Krakow, Poland

Michał Mańka
AGH University of Krakow, Faculty of Mechanical Engineering and Robotics, Department of Robotics and Mechatronics
Krakow, Poland

Krzysztof Mendrok
AGH University of Krakow, Faculty of Mechanical Engineering and Robotics, Department of Robotics and Mechatronics
Krakow, Poland

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-63443-7 ISBN 978-3-031-63444-4 (eBook)
https://doi.org/10.1007/978-3-031-63444-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This monograph deals with the selected recent achievements in the rapidly growing field of mechatronics. Specifically, it focuses on the aspects that make a link between academia and industry. A continuous development of mechatronics requires successful cooperation between the researchers and professionals in the scope of design, prototyping, manufacturing and testing methods and tools. The new ideas emerging in the scientific centers combined with a practical experience found in the industrial partners may lead to an advantageous collaboration and launching new products on the market. At this point, it is also worth mentioning about the crucial and invaluable contribution of mechatronics in the development of the approach Industry 4.0 and effective implementation of its requirements and assumptions. In fact, mechatronics provides a variety of robust hardware and software means ready to be applied in industry. This all explains why this specific scientific and technical field is found of the particular interest, for both scholars and engineers.

The following chapters of the monograph report the selected and peer-reviewed works presented during the 6th International Conference Mechatronics 2023: Ideas for Industrial Applications that was held from 11 to 13 December 2023 in Krakow, Poland. The conference was organized by the Faculty of Mechanical Engineering and Robotics at the AGH University of Krakow. The event was a successful continuation of the conferences previously held in Warsaw (2012), Łódź (2014), Gdańsk (2015), Wisła (2017) and Szczecin (2021). The conference gathered over 65 participants representing both academia and industry and providing 36 presentations. Over 10 attending companies presented the results of the conducted research and offers of cooperation, primarily focusing on their recent technological achievements in the field of mechatronics. Apart from regular technical sessions, the event also covered a discussion panel "Mechatronics: The Road Towards Industry 4.0" making the floor for ideas exchange between the attendees representing various professional backgrounds.

In this monograph, the following aspects of mechatronics and its practical implementations are particularly addressed: artificial intelligence, robotics and navigation methods, actuation and sensing, image processing and measurement techniques, developments of software and control systems, space applications, modeling and identification of the properties of mechatronic systems, use of smart materials as well as physical prototype construction employing 3D printing manufacturing.

March 2024

Adam Martowicz
Michał Mańka
Krzysztof Mendrok

Scientific Committee Members

- Eugeniusz Świtoński, Silesian University of Technology, Poland
- Anna Timfiejczuk, Silesian University of Technology, Poland
- Tadeusz Uhl, AGH University of Krakow, Poland
- Daniel H. Wagner, Weizmann Institute of Science, Israel
- Roman Szewczyk, Warsaw University of Technology, Poland
- Michał Wieczorowski, Poznań University of Technology, Poland

Contents

Comparative Analysis of RL-Based Algorithms for Complex Systems Control

Patryk Bałazy(✉), Paweł Knap, and Szymon Podlasek

AGH University of Krakow, Faculty of Mechanical Engineering and Robotics, Al. Mickiewicza 30, 30-059 Kraków, Poland
balazy@agh.edu.pl

Abstract. The increasing complexity and nonlinearityof modern control systems pose serious challenges to achieving effective control. While differential equations have traditionally been the foundation of control theory and optimization with constraints, they may not be sufficient to accurately model chaotic and unbalanced systems. In response to these challenges, newer approaches gaining prominence include the use of policy iteration and Reinforcement Learning (RL). These techniques revolve around the concepts of sequences of actions and rewards, providing a dynamic and adaptable framework for controllers. Learning with reinforcement offers a promising avenue for addressing control theory, enabling systems to robustly adapt to dynamic environments. This paradigm shift away from traditional approaches such as linear–quadratic regulator (LQR) controllers to a more robust reward-based control mechanism is particularly noteworthy. Through a comprehensive analysis of RL-based algorithms, including Deep Deterministic Policy Gradient (DDPG), Twin Delayed DDPG (TD3), Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC) and Trust Region Policy Optimization (TRPO), we examine their suitability and performance in crane control scenarios. This study aims to shed light on the advantages and limitations of each algorithm in optimizing crane operations, offering valuable insights for future applications of RL in complex control systems.

Keywords: RL · control · complex · nonlinear · system · agent · algorithms · analysis

1 Introduction

Today's control systems, due to their increasing complexity and nonlinearity, pose serious challenges to achieving effective control. Traditionally, differential equations have been the foundation of control theory and constrained optimization; however, they may not be sufficient to accurately model chaotic and unstable systems. The complexity of control systems is due to many factors, such as the large number of state variables that describe the system, the combination of different subsystems that work together to achieve a common goal, uncertainty in the parameters and behavior of the system. The nonlinearity of control systems means that their behavior cannot be fully described

A. Martowicz et al. (Eds.): MECHATRONICS 2023, LNNS 1042, pp. 1–14, 2024.
https://doi.org/10.1007/978-3-031-63444-4_1

by linear equations. For linear systems, changes in system states and actions are proportional to each other. For nonlinear systems, changes in the states and actions of the system may not be proportional to each other. The complexity and nonlinear nature of today's control systems pose significant challenges in achieving effective control. Traditional control approaches based on differential equations may not be sufficient to properly model chaotic and asymmetric systems. Examples of complex and nonlinear control systems include [1–4]: industrial robots, autonomous cars, public transportation systems, power grid systems.

In these systems, traditional control approaches may not be able to provide stability and performance. In such cases, RL may offer a promising alternative [5]. RL is a branch of machine learning that addresses decision-making problems in dynamic and uncertain environments. In the context of systems control, RL can be used to learn policies that determine what actions should be performed in a given state to achieve a specific goal. For complex and nonlinear control systems, RL can have the following advantages:

– Allows adaptation to dynamic environments
– Is immune to disturbances
– Can be used to control systems whose behavior is not fully known.

In RL, the agent acts in the environment and receives rewards for its actions. The agent's goal is to maximize the aggregate reward it will receive while interacting with the environment. The process of RL can be divided into the following stages:

1. Exploration - the agent tries different actions to learn about the environment and understand what actions lead to rewards.
2. Learning - the agent uses the information gathered during exploration to improve its policy.
3. active learning - the agent learns by focusing on those states and actions that lead to the greatest reward.

RL has several advantages in the context of systems control, including [6]:

– Adaptation - RL enables the agent to adapt to dynamic environments. As the environment changes, the agent can update its policies to continue to achieve satisfactory performance levels.
– Resilience - RL is resilient to disruptions. If there is a disruption in the environment, the agent can continue to perform optimally.
– Uncertainty - RL can be used to control systems whose behavior is not fully known. In the absence of full knowledge of the system, the agent can learn a policy that provides a satisfactory level of performance.

RL is a promising direction for solving control problems of complex systems. RL allows adaptation to dynamic environments, is robust to disturbances and can be used to control systems whose behavior is not fully known. RL also has some limitations [7], including:

– Computational requirements - RL can be computationally demanding, especially for complex systems.
– Complex environment - RL can be difficult to apply to complex environments with many states and actions.

– Agent behavior - the behavior of an agent learned using RL can be unpredictable.

The purpose of this article is to comprehensively analyze RL-based algorithms in the context of crane control. To date, there has been no work focusing on comparing the performance quality of algorithms available for controlling a nonlinear multidimensional object through RL-based systems. Hence, it was decided to compare the following algorithms and select the best performing one for controlling an object such as an overhead crane. The selected algorithms are some of the most popular and effective RL algorithms available today. Each algorithm has its own advantages and disadvantages.

– TD3 [8] learns two Q-functions instead of one. It uses the smaller Q-value to form the targets in the Bellman error loss functions.
– DDPG [9] concurrently learns a Q-function and a policy. It uses off-policy data and the Bellman equation to learn the Q-function, and uses the Q-function to learn the policy
– PPO [10] is a policy-based algorithm that uses policy gradients to update an agent's policy with confidence in policy stability.
– SAC [11] is an algorithm that optimizes a stochastic policy in an off-policy way, forming a bridge between stochastic policy optimization and DDPG-style approaches.
– TRPO [12] updates policies by taking the largest step possible to improve performance, while satisfying a special constraint on how close the new and old policies are allowed to be. The constraint is expressed in terms of KL-Divergence, a measure of distance between probability distributions.

The results of the experiments, which were conducted to compare the performance of RL algorithms in the context of crane control, showed that all algorithms were able to achieve satisfactory performance levels. However, the TD3 algorithm proved to be the most efficient in terms of operation time and load placement accuracy. The simulation study took into account the variable dynamic properties of the crane. This means that the crane was not a perfect system, and its behavior was affected by factors such as varying sling lengths. The effect of these factors on the performance of the RL algorithms varied. The DDPG and PPO algorithms were more sensitive to the varying dynamic properties of the crane than the SAC and TRPO algorithms. In order to reduce the duration of the learning process, conditions were imposed that the system must meet in order for the control to be considered satisfactory. These conditions included:

– The load must be placed in the designated area with a certain accuracy
– The sway angle of the sling must not exceed the permissible value
– The crane must not exceed a certain speed or acceleration.

2 Object of Study

This chapter will focus on presenting a specific object on which a detailed analysis of RL algorithms will be carried out. A bridge crane, which is an example of a multidimensional system, has been chosen as this object. The study will be carried out using a simplified mathematical model of the crane, which will enable the identification of the most appropriate RL algorithm. The simplified mathematical model of the overhead crane will serve as a platform for evaluating the effectiveness of various RL algorithms.

The goal of this research stage is to understand exactly which of the analyzed algorithms best meets the requirements of controlling the crane in a dynamic environment.

2.1 Object Presentation

A bridge crane is a type of crane, consisting of a cart that moves in two axes and a sling that moves in one axis and swings in two planes. The following is an overview diagram of the crane (Fig. 1). The model of a bridge crane can be described by the state variables shown in Table 1.

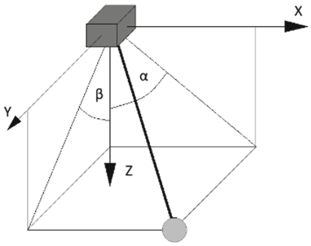

Fig. 1. Crane schematic drawing

Model inputs u were taken as the forces acting on the carriage in each of the two axes and the force acting on the sling:

$$u = \left[F_x, F_y, F_l\right] \tag{1}$$

The state variables x of the model were assumed:

$$x = [x, v_x, y, v_y, \alpha, \omega_\alpha, \beta, \omega_\beta, l, v_l] \tag{2}$$

The equations of state were determined assuming zero initial conditions:

$$\dot{x}_1 = \dot{x} = x_2 \tag{3}$$

$$\dot{x}_2 = \dot{v_x} = \frac{1}{m_c + m_r + m_l} F_x + \frac{m_l g}{m_c + m_r + m_l} \alpha \tag{4}$$

$$\dot{x}_3 = \dot{y} = x_4 \tag{5}$$

$$\dot{x}_4 = \dot{v_y} = \frac{1}{m_c + m_l} F_y + \frac{m_l g}{m_c + m_l} \beta \tag{6}$$

$$\dot{x}_5 = \dot{\alpha} = x_6 \tag{7}$$

$$\dot{x}_6 = \dot{\omega_\alpha} = \frac{-1}{l(m_c + m_r + m_l)} F_x - \frac{g(m_c + m_r + 2m_l)}{l(m_c + m_r + m_l)} \alpha \tag{8}$$

Table 1. Model parameters.

Symbol	Description	Value
m_c	Mass of the cart	1,2 kg
m_l	Mass of the payload	0,6 kg
m_r	Mass of the moving rail	2 kg
g	Gravitational constant	9,81 $\frac{m}{s^2}$
F_x	X-axis force acting on the carriage	-
F_y	Y- axis force acting on the carriage	-
F_l	Z-axis force acting on the payload	-
x	X-axis position of the cart	$0 \div 1{,}5$ m
v_x	X-axis velocity of the cart	$-0{,}8 \div 0{,}8$ $\frac{m}{s}$
y	Y-axis position of the cart	$0 \div 1{,}5$ m
v_y	Y-axis velocity of the cart	$-0{,}8 \div 0{,}8$ $\frac{m}{s}$
α	Angle between the lift-line and its projection on YZ	$-90° \div 90°$
ω_α	Angular velocity, rate of change of α angle	-
β	Angle between the lift-line and its projection on YZ	$-90° \div 90°$
ω_β	Angular velocity, rate of change of β angle	-
l	Length of the lifted load	$0 \div 1$ m
v_l	Lowering/lifting velocity of the load	$-0{,}7 \div 0{,}7$ $\frac{m}{s}$

$$\dot{x}_7 = \dot{\beta} = x_8 \tag{9}$$

$$\dot{x}_8 = \dot{\omega_\beta} = \frac{-1}{l(m_c + m_l)} F_y - \frac{g(m_c + 2m_l)}{l(m_c + m_l)} \beta \tag{10}$$

$$\dot{x}_9 = \dot{l} = x_{10} \tag{11}$$

$$\dot{x}_{10} = \dot{v_l} = \frac{1}{2m_l} F_z \tag{12}$$

This is how the crane object is presented.

2.2 Non-linearity of the Crane

A bridge crane is a nonlinear system, which means that its behavior cannot be described by linear equations. The nonlinearity of a bridge crane is due to a couple of factors. The first is the nonlinear dependence of force on velocity. The force that acts on the crane boom depends on the velocity of the boom. This force is proportional to the velocity of the boom, but its proportionality coefficient is not constant. As the velocity of the boom increases, this coefficient decreases. This means that the faster the boom moves, the smaller the force acting on it. For high boom velocity, the force can even reach zero. Another factor is the nonlinear dependence of torque on angular velocity. The torque that acts on the crane boom depends on the angular velocity of the boom. This torque is proportional to the angular velocity of the boom, but its proportionality coefficient is not constant. As the angular velocity of the boom increases, this coefficient decreases. This means that the faster the boom rotates, the smaller the torque acting on it. At high boom velocity, the torque can even reach zero. The next factor is the nonlinear dependence of the friction force on the velocity and angle of rotation. The frictional force that acts on the crane boom depends on the velocity and angle of rotation of the boom. This force is proportional to the velocity and angle of rotation of the boom, but its proportionality coefficient is not constant. This means that the friction force can vary depending on the operating conditions of the crane. For example, if the boom is moving at a high velocity, the friction force may be greater than if the boom is moving at a low velocity. The varying length of the sling creates additional nonlinearity in the behavior of the crane. Sling length affects the forces and moments acting on the crane boom. As the sling length increases, these forces and moments decrease. This means that an overhead crane with a longer sling will have less ability to lift large loads or move at high velocity. The varying weight of the load also creates additional nonlinearity in the behavior of the crane. The mass of the load affects the forces and moments acting on the crane boom. As the weight of the load increases, these forces and moments increase. This means that an overhead crane with a heavier load will have higher motor power requirements and will be more susceptible to instability. For the simulation study, the sling was assumed to be interpreted as rigid. A rigid sling is not able to bend under load. This means that the forces and moments acting on the crane boom can be calculated by assuming that the sling is perfectly straight. In addition, the load sway angles were assumed to be close to zero, eliminating the need to take into account complex calculations related to gravity and friction. Assumptions about a rigid sling and small sway angles lead to a simplified mathematical model of the crane. The simplified model is less complex and easier to analyze, but may be less accurate than the full model. The accuracy of the simulation depends on the operating conditions of the crane. For cranes operating with light loads and under stable conditions, the assumed simplifications may be sufficient. For cranes operating with heavy loads or under unstable conditions, the assumed simplifications can lead to significant errors in the simulation. In future work, the pre-learned artificial neural network algorithm will be able to adapt to the omitted nonlinearities by retraining on the real object. The nonlinearity of the simulated object is the variable length of the lifted load l. This can be seen in Eqs. 8 and 10. As the length of the lifted load changes, the dynamics of the system changes. This is done by calculating Eqs. 3–12 anew at each time sample.

3 Methodology

This chapter will focus on adapting the RL solution to an overhead crane control system. RL is a machine learning method in which an agent learns to make decisions in a complex environment to maximize its reward. An agent is an entity that makes decisions. In the case of an overhead crane, the agent could be the controller that controls the boom feed rate and the boom rotation velocity. The environment is the setting in which the agent operates. In the case of an overhead crane, the environment is the crane with its load. The reward is the value the agent receives for making certain decisions. In the case of an overhead crane, the reward may be the correct placement of the load in the designated area. In the RL learning process, the agent interacts with the environment and receives a reward or punishment for its actions. Based on this information, the agent learns to make decisions that maximize its reward.

3.1 Using the RL Algorithm as a Regulator

In a typical control system with a feedback controller, the controller observes the object's state and, using pre-calculated gains, generates a feedback signal to adjust the object's behavior [15]. The controller receives feedback from the object in the form of the object's response to the action. This feedback is used to refine the controller's behavior over time. In a RL system, the controller is replaced by an agent. The agent interacts with the environment, which is the controlled object, and receives a reward signal from the environment in addition to the object's response. The agent uses the reward signal to learn to take actions that maximize the reward. This is how the control system using the RL method was created (Fig. 2).

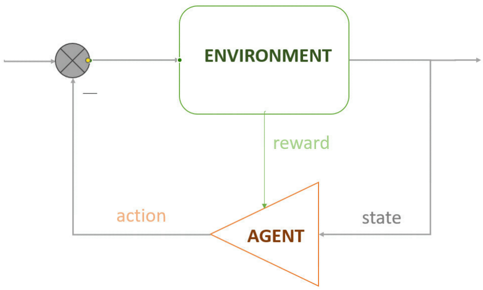

Fig. 2. Control system

This system works as follows:

1. The agent observes the state of the object
2. The agent takes an action to generate feedback gains
3. The object receives a control signal which is setpoint minus action
4. The environment provides the agent with a reward signal based on the object's response
5. The agent updates its policy based on the reward signal
6. The process repeats.

The main difference between a traditional control system with a feedback controller and an RL system is that the RL system does not require a model of the object [13]. The agent learns to control the object by interacting with it directly, and the reward signal provides the agent with the information it needs to learn.