52. IWK

Internationales Wissenschaftliches Kolloquium

International Scientific Colloquium

PROCCEDINGS | 10 - 13 September 2007

FACULTY OF COMPUTER SCIENCE AND AUTOMATION

COMPUTER SCIENCE MEETS AUTOMATION

VOLUME I

Session 1 - Systems Engineering and Intelligent Systems

Session 2 - Advances in Control Theory and Control Engineering

Session 3 - Optimisation and Management of Complex Systems and Networked Systems

Session 4 - Intelligent Vehicles and Mobile Systems

Session 5 - Robotics and Motion Systems

TECHNISCHE UNIVERSITÄT ILMENAU

Bibliografische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbiografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.ddb.de** abrufbar.

ISBN 978-3-939473-17-6

Impressum

Herausgeber:	Der Rektor der Technischen Universität Ilmenau Univ.-Prof. Dr. rer. nat. habil. Peter Scharff
Redaktion:	Referat Marketing und Studentische Angelegenheiten Kongressorganisation Andrea Schneider Tel.: +49 3677 69-2520 Fax: +49 3677 69-1743 e-mail: kongressorganisation@tu-ilmenau.de
Redaktionsschluss:	Juli 2007
Verlag:	Technische Universität Ilmenau/Universitätsbibliothek Universitätsverlag Ilmenau Postfach 10 05 65 98684 Ilmenau www.tu-ilmenau.de/universitaetsverlag
Herstellung und Auslieferung:	Verlagshaus Monsenstein und Vannerdat OHG Am Hawerkamp 31 48155 Münster www.mv-verlag.de
Layout Cover:	www.cey-x.de
Bezugsmöglichkeiten:	Universitätsbibliothek der TU Ilmenau Tel.: +49 3677 69-4615 Fax: +49 3677 69-4602

Preface

Dear Participants,

Confronted with the ever-increasing complexity of technical processes and the growing demands on their efficiency, security and flexibility, the scientific world needs to establish new methods of engineering design and new methods of systems operation. The factors likely to affect the design of the smart systems of the future will doubtless include the following:

- As computational costs decrease, it will be possible to apply more complex algorithms, even in real time. These algorithms will take into account system nonlinearities or provide online optimisation of the system’s performance.
- New fields of application will be addressed. Interest is now being expressed, beyond that in “classical” technical systems and processes, in environmental systems or medical and bioengineering applications.
- The boundaries between software and hardware design are being eroded. New design methods will include co-design of software and hardware and even of sensor and actuator components.
- Automation will not only replace human operators but will assist, support and supervise humans so that their work is safe and even more effective.
- Networked systems or swarms will be crucial, requiring improvement of the communication within them and study of how their behaviour can be made globally consistent.
- The issues of security and safety, not only during the operation of systems but also in the course of their design, will continue to increase in importance.

The title “Computer Science meets Automation”, borne by the 52nd International Scientific Colloquium (IWK) at the Technische Universität Ilmenau, Germany, expresses the desire of scientists and engineers to rise to these challenges, cooperating closely on innovative methods in the two disciplines of computer science and automation.

The IWK has a long tradition going back as far as 1953. In the years before 1989, a major function of the colloquium was to bring together scientists from both sides of the Iron Curtain. Naturally, bonds were also deepened between the countries from the East. Today, the objective of the colloquium is still to bring researchers together. They come from the eastern and western member states of the European Union, and, indeed, from all over the world. All who wish to share their ideas on the points where “Computer Science meets Automation” are addressed by this colloquium at the Technische Universität Ilmenau.

All the University’s Faculties have joined forces to ensure that nothing is left out. Control engineering, information science, cybernetics, communication technology and systems engineering – for all of these and their applications (ranging from biological systems to heavy engineering), the issues are being covered.

Together with all the organizers I should like to thank you for your contributions to the conference, ensuring, as they do, a most interesting colloquium programme of an interdisciplinary nature.
I am looking forward to an inspiring colloquium. It promises to be a fine platform for you to present your research, to address new concepts and to meet colleagues in Ilmenau.

Professor Peter Scharff
Rector, TU Ilmenau

Professor Christoph Ament
Head of Organisation

Table of Contents

C O N T E N T S

Paper submitted after copy deadline

2 Advances in Control Theory and Control Engineering

52nd Internationales Wissenschaftliches Kolloquium
Technische Universität Ilmenau
10 – 13 September 2007

René Güttler/Axel Schneider/Christoph Ament/Josef Schmitz

Using a reinforcement learning approach in a discrete event manufacturing system

Abstract

Up to date assembly lines and manufacturing systems use manufacturing cells and single machines which are loaded and unloaded by a gantry loader. The gantry loader is a rare resource that has to be shared by all manufacturing cells and machines (clients). In this context the loading and unloading strategy which is implemented in the controller of the gantry is essential to gain a high manufacturing efficiency. To our knowledge, the implemented control strategies are mostly based on fixed schedules that dictate the order in which the available clients are served. On the one hand a fixed schedule guarantees safe operation in the normal case. On the other hand a fixed schedule can not adapt to unknown situations. We propose a reinforcement learning approach to add the flexibility of lifelong learning to the classical controller's ability of keeping the process in well defined boundaries. The control approach introduced in this paper can also be used offline to train a new controller or it can be used online. To evaluate the adaptive properties of the flexible controller trained by means of reinforcement learning we present simulation data of a small setup consisting of n single machines and a gantry loader.

Introduction

Assembly and production lines consist of components like conveyors, machines for handling the work pieces, CNC-machines for drilling, milling and turning etc. The production processes are controlled by human workers who still, in some cases, fulfil production steps manually. In these processes, all members, be it humans or machines, have to interact in such a way that the work pieces are produced with a high efficiency at a desired quality. The work pieces have to pass different production steps in different machines. Each machine in this chain is specialized on certain production steps to minimize set-up time. Usually, there is a well defined number of machines at each production site in the line to process the desired operations for the work pieces in a desired cycle time. In a real world process, each machine is affected by different machine failures. In such a case, a human operator has to find the source of defect, eliminate it and restart the machine.

Especially when more than one machine accomplishes the same process (parallel) usually a gantry loader transports the work pieces from a storage area or a conveyor to the machines and back. In a well configured system this transfer time (including waiting time due to busy machines) is minimal because the process time of each machine and the arrival of new work pieces are synchronized. This balancing is disturbed by the above mentioned machine failures. A classical controller for the gantry loader orientates itself to the given state vector of the machines at time t and reacts according to its implemented strategy. However, in some cases it is more suitable to alter the implemented strategy, for example in unpredictable situations like the sudden shutdown of machines due to failures or configurationally changes. Therefore, an adaptive strategy for loading and unloading, even in the situations described above, would minimize the transport costs in error-prone real world scenarios. The loader should be able to decide on the action it has to take next for a given situation of the environment. This decision should consider past and current data respectively to adapt to changing situations. Strategies that use past data in current control actions can be regarded as learning strategies.

On the one hand, a controller with a live long learning algorithm is able to find time optimal solutions in new situations but stability can not always be guaranteed. On the other hand, a classical controller uses a robust strategy which is not time optimal in unpredictable situations. In this paper we propose a combination of these two control approaches to combine their advantages (see Fig. 1).

Different applications have already been introduced in the context of industrial manufacturing in which an optimal control strategy was designed by means of a reinforcement learning algorithm. Creighton and Nahavandi developed a reinforcement learning agent to determine the optimal operating policy in a multi-part serial line by using a discrete event simulation environment [3]. Ayedin and Oztemel successfully implemented a reinforcement learning agent for a job-shop scheduling problem [5]. Real world problems for elevator dispatching were presented by Crities and Barto [6, 7].

Using a Reinforcement Learning approach

The idea of Reinforcement Learning (RL) was derived from dynamic programming (DP) where a full model of the process has to be known (including all states and transition probabilities). The motivation for using a reinforcement learning approach for the control of a manufacturing process is that it does not need an explicit model or an a-priori strategy. In real world control problems a closed model of a complex process is difficult to derive, especially in manufacturing applications that consist of many components. A controller based on reinforcement learning develops its strategy by getting rewards for beneficial behaviour and punishments (or fewer rewards) for useless actions.

A control system that is based on reinforcement learning consists of the following components: A policy, a reward function, a value function and usually a model of the environment [1]. During a learning phase the agent makes a decision (policy) and controls the environment by an action a so that the environment changes its state from s to s'. As a result of the change in the state the agent receives a specified reward r ($r \in R$) for the action he has taken. A cumulative reward is calculated by integration of all current rewards over time. The cumulative reward function implicitly defines the goal of the RL while the agent tries to maximize the overall received reward. The decision for a certain action of the agent for state s depends on the value function which describes the long term experience of the agent [1, 10 and 11].

As opposed to classical controller design, in which the developer has to choose the control strategy based on a model, in RL-based controller design the difficulty lies in designing the reward and value functions. Current research is focussed on finding suitable reward and value functions for a given application because there is no general framework for the design process [7, 8 and 9].

Common methods for calculating the value-function are so called Q-values [1]. These values are stored in large tables or are estimated by neural networks when the state space is too large. The following updating equation (1) describes how to estimate a Value-Function that is derived by the Bellman Equation. It is independent of the transition probability to get from state s to state s',

$$Q_{k+1}(s,a) = (1-\alpha_{k+1})Q_k(s,a) + \alpha_{k+1}[r(s,a) + \gamma max_{a'} Q_k(s',a')] \tag{1}$$

where $s \in S$ are the states and $a \in A$ are the actions of the agent. $r(s,a)$ denotes the reward for a given state s and action a. The parameter α describes the learning rate and γ the discount rate for learning. Discounting the reward determines the present value for future rewards [1]. The states of the environment include the position of the gantry loader, the loading states of the gripper, the state of the storage area or conveyor and the states of the machines. Actions that are processed by the crane are wait, turn left, turn right and pick-and-place for both grippers.

Architecture of the proposed system

The architecture of the system proposed in this paper comprises a classical controller, a flexible controller, a comparator and the event discrete model of the material flow in the real process. Fig.1 shows the control scheme of this approach. The new, flexible controller to be trained is arranged in parallel to the classical controller. During a training period only the classical controller works on the real process whereas the flexible controller works on a model of the real process. Both the real and the model process receive work pieces at the input and release work pieces at the output after manufacturing. A second control input of the real and model process is connected to the output of the classical and flexible controller, respectively. Both controllers receive a state vector of their own process as input. Besides the states of the process, the state vector also contains a reward signal from a teacher that judges the efficiency of the real and the modelled process' actions. The classical controller does not benefit from the teaching signal. However, the flexible controller learns to improve its control strategy because controller actions which lead to undesirable results (e.g. low work piece throughput) are punished and actions that improve the work piece flow are rewarded. Therefore, the flexible controller can adapt to situations in which, for example, the productivity of manufacturing cells is decreased due to failures or is increased due to recovery of the cell. A comparator constantly monitors the reward signals from the process controlled by the classical controller and from the process controlled by the flexible controller. When the flexible controller achieves higher rewards, the comparator circuit can decide to swap the classical and the flexible controller (commutator swiches in Fig. 1). As a result, the flexible controller governs the real process and the classical controller works on the process model. If the quality of the flexible controller's actions on the process decreases, the comparator can also reverse the swapping process and use the classical controller again until the flexible controller achieves better rewards.

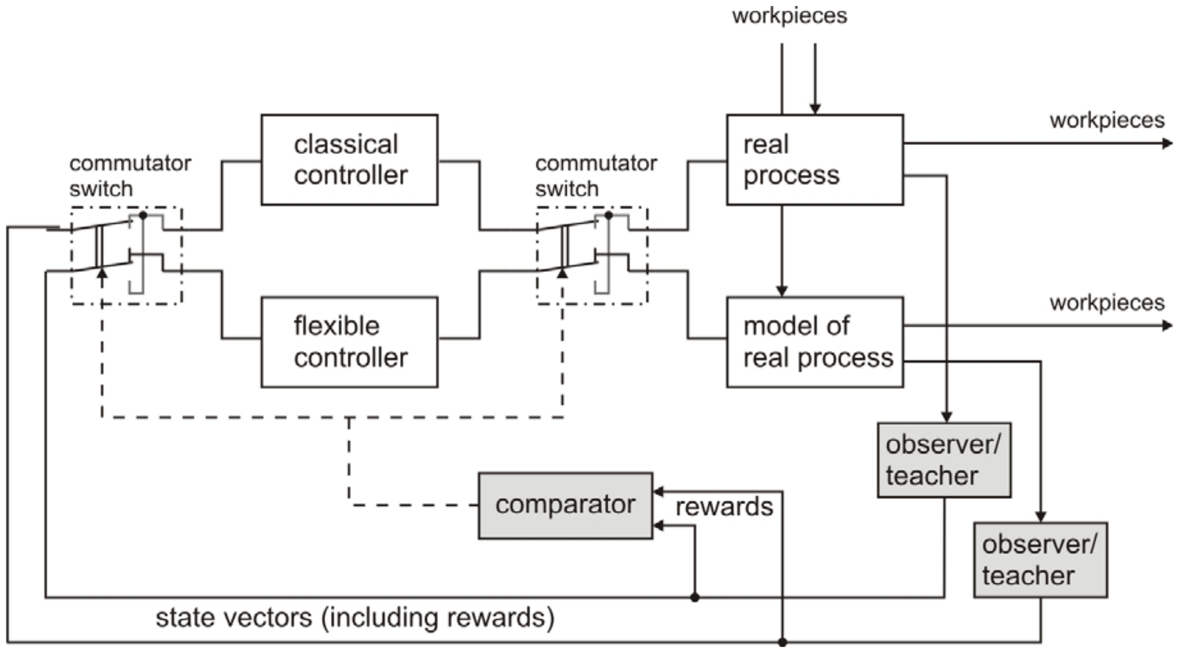

Fig. 1 Combined controller

Simulation data of a small configuration

In order to test our control approach we implement a model of a manufacturing system Matlab/Simulink (The MathWorks Inc., Natick, MA). We compared a simple FIFO (first in first out) strategy with a Reinforcement Learning strategy in throughput of workpieces and the cumultativ reward rate of the action of the gantry loader. The configuration has 4 machines in which work pieces are processed with a processing time of 120s, a gantry loader with two grippers, both for transporting work pieces from the machines to the storage area and back. Finally, there is a storage area where new work pieces arrive every 30s. The system dynamics are characterized as follows: The time to travel from one machine to another is 2.5s, the time for grabbing (pick or place) a work piece takes 6s. In an optimal case the gantry loader's position is above the machines or the storage area whenever the processing of a work piece is finished or a new work piece has arrived. Hence there is no waiting for transport and one can expect a maximal throughput. For each movement of the gantry loader the agent gets a static negativ Reward (penalty). In the successful state, where the crane places a processed work piece at the conveyor belt, the agent gets a reward of 1000 Points (see Fig. 2)

strategy	Output (workpieces)
FIFO	300
RL with exploration factor 0.1	350
RL with manual mode	453

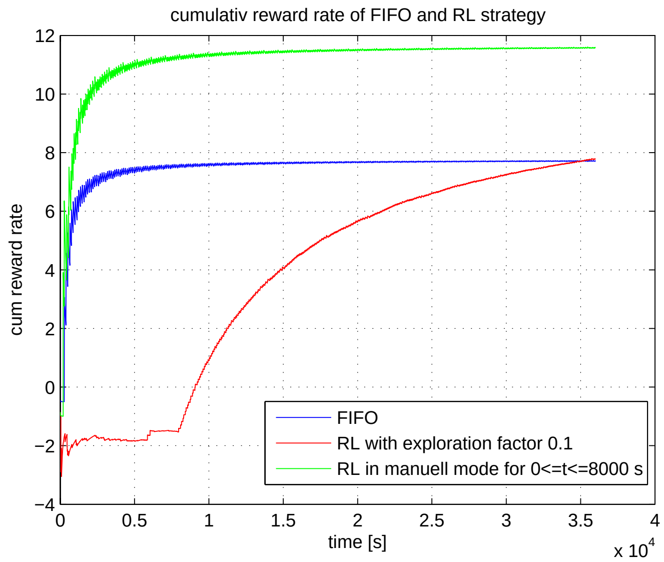

Fig. 2 Comparison between a simple FIFO strategy and a Reinforcement Learning approach in output and reward rate

For this architecture to become active in controlling the real production system the commutator switch is needed (Fig.1). At the branching point between the FIFO reward rate and the Reinforcement Learning reward rate the commutator have to switch from the classical strategy to the reinforcement learning strategy until the reward rate of the new strategy becomes worse. We suppose that for a well trained flexible controller with Reinforcement Learning the average throughput time for both cases, undisturbed and disturbed process, is less than for the classical controller (e.g. FIFO). Thus, one can expect more output during a given time window.

Conclusion and Future Work

We propose a new architecture to handle control problems in a discrete manufacturing environment by implementing a reinforcement learning approach. We tested this approach separately on a simulated machine configuration with a control task for the gantry loader. In this simulation we compared a classical FIFO strategy with a reinforcement learning strategy. We show that the new approach works in an undisturbed environment. In the present paper we did not consider failures of the machines. Consideration of stochastic failures will be part of our future work.

References

[1] Sutton, R.S.; Barto, A. G. Reinforcement learning: an introduction. Cambridge: MIT Press, 1998.

[2] Ghavamzadeh; M., Mahadevan, S. Hierarchically Optimal Average Reward Reinforcement Learning. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2002), pp.195-202. Sydney, Australia, 2002.

[3] Creighton, D. C.; Nahavandi S. Optimizing Discrete Event Simulation Models Using a Reinforcement Learning Agent. Proceedings of the 2002 Winter Simulation Conference Vol.2,p. 1945-1950 , Sch. of Eng. & Technol., Deakin Univ., Geelong, Vic., Australia, 8-11 Dec 2002.

[4] Mahadevan, S.; Theocharos, G.: Optimizing Production Manufacturing using Reinforcement Learning. Eleventh International FLAIRs Conference 372-377, 1998.

[5] Aydin, M. E.; Öztemel, E. Dynamic job-shop scheduling using reinforcement learning agents. Robotics and Autonomous Systems 33, 169-178. 2000.

[6] Crites, R. H.; Barto, A.G. Improving Elevator Performance Using Reinforcement Learning. Advances in Neural Information Proceedings Systems. Cambridge: MIT Press. 1996.

[7] Crites, R. H.; Barto, A.G. Elevator Group Control Using Multiple Reinforcement Learning Agents. Boston. Kluwer Academic Publishers. 1997.

[8] Hong, J.; Prabhu V. V. Distributed Reinforcement Learning Control for Batch Sequencing and Sizing in Just-In-Time Manufacturing Systems. United States. Kluwer Academic Publishers, 2004.

[9] Ghavamzadeh, M.; Mahadevan, S. Hierarchical Average Reward Reinforcement Learning, CMPSCI Technical Report 03-19, University of Massachusetts, 2003.

[10] Mitchell, T.M. 1997: Machine Learning. McGraw Hill.

[11] Gosavi, A: Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning. Kluwer Academic Publishers, Boston, 2003.

Authors:
Title René Güttler,
Title Axel Schneider,
Title Christoph Ament,
Title Josef Schmitz

Company, street, P.O.B.
98693, Ilmenau
Phone: 03677/691467, E-mail: rene.guettler@tu-ilmenau.de