Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-1 (2016) 089–094

MODEL-FREE PREDICTIVE CONTROL OF NONLINEAR PROCESSES BASED ON REINFORCEMENT LEARNING

Dr. Hitesh Shah*, Dr. M. Gopal**

***Professor, Department of Electronics & Communication Engineering,**

G H Patel College of Engineering & Technology

Vallabh Vidyanagar (Gujarat) 388120, INDIA

E-mail: iitd.hitesh@ gmail.com

**Director, *School of Engineering*

Shiv Nadar University, Noida (Uttar Pradesh), INDIA

E-mail: mgopal@snu.edu.in

Abstract: Model predictive control (MPC) is a model-based control philosophy in which the current control action is obtained by on-line optimization of objective function. MPC is, by now, considered to be a mature technology owing to the plethora of research and industrial process control applications. The model under consideration is either linear or piece-wise linear. However, turning to the nonlinear processes, the difficulties are in obtaining a good nonlinear model, and the excessive computational burden associated with the control optimization. Proposed framework, named as model-free predictive control (MFPC), takes care of both the issues of conventional MPC. Model-free reinforcement learning formulates predictive control problem with a control horizon of only length one, but takes a decision based on infinite horizon information. In order to facilitate generalization in continuous state and action spaces, fuzzy inference system is used as a function approximator in conjunction with *Q*-learning. Empirical study on a continuous stirred tank reactor shows that the MFPC reinforcement learning framework is efficient, and strongly robust.

Keywords: Model predictive control, Reinforcement learning, *Q*-learning.

1. INTRODUCTION

Model predictive control (MPC) is the most popular advanced control technique in the process industry [1]. The essence of MPC is to optimize, over the manipulable inputs, forecasts of process behavior. The forecasting is accomplished with process model, and therefore, the model is the essential element of the MPC controller.

Whereas the dynamic behavior of most chemical processes is nonlinear, linear models have predominantly been used for process control in practice because of the difficulty associated with building an accurate nonlinear model, either by first principles or by system identification techniques [1]. MPC approach determines a sequence of actions based on predictions using the system model that guarantee stability and certain optimal properties of the system in terms of the desired behavior. A model is, however, always an approximation of the system under consideration. Predictions about the behavior of the system become more and more inaccurate when considered further into the future. To deal with this, MPC techniques use a *rolling horizon* to increase robustness. The rolling horizon principle consists of synchronizing the state of the model with the state of the true system at every decision step. At every decision step, the MPC agent observes the state of the true system and synchronizes the estimate that it has of the state of the system with this, and tries to find the best sequence of actions, given the updated state. Typically, the agent only executes the first action of this sequence. It then observes the system state again and finds a new sequence of actions. However, the rolling horizon increases computational costs if the horizon over which the agent has to determine actions is infinite, since at each decision step the MPC agent has to find a sequence of actions. A finite horizon is therefore assumed. However, because of the limited horizon over which actions are considered, the resulting policy may be suboptimal. The smaller the control horizon used to reduce on-line computation, the more suboptimal the resulting solution may become. The MPC algorithm therefore might suffer from the dilemma of very high computational requirements *vs* suboptimality.

However, both the difficulties — obtaining a good model of the nonlinear process and the excessive computational burden associated with the control optimization, have been serious obstacles to wide spread use of MPC in industrial implementations. To deal with the issue of modeling difficulties, and the dilemma of very high computational cost versus suboptimality, a control approach based on learning is more adequate. Different model-free learning control approaches, useful for process control problems, have been proposed in the literature. Anders Stenman [2] presented the concept of model-free predictive control, which combines the idea of model-on-demand with MPC techniques. They estimate the process dynamics locally and on-line, using process data stored in the database. However, a drawback of their approach is that the performance of the controller depends on the quality of the data base, and the controller normally requires large computational resources due to the nature of the underlying estimation procedure. Lee and Lee [3] & Lee and Wong [4] have suggested an MPC approach based on approximate dynamic programming (ADP). This approach has its roots in the artificial intelligence (AI) literature and closely follows the ideas of reinforcement learning (RL) [5] and neuro-dynamic programming [6]. This

10.1016/j.ifacol.2016.03.034

approach attempts to solve the stochastic optimal control problem through dynamic programming (DP) but only approximately and within limited regions of state space. The usual barrier of *curse-of-dimensionality* is alleviated by employing closed-loop simulations and function approximation.

An ADP-based data driven control algorithm presented in [4], iteratively learns cost-to-go function and maps the state to the cost-to-go value. If we map all relevant state and action pairs to cost-to-go values, no model would be necessary at all. Such a model-less scheme is called *Q*-learning in the field of AI [5]. A model-free learning control (MFLC) approach, proposed by S. Syafiie, et. al. [7, 8], is based on reinforcement learning algorithms. The MFLC controller algorithm, based on the one-step ahead *Q*-learning look-up table, performs reasonably well. However, for continuous state space systems, such as in most process control problems, the conventional *Q*-learning framework designed for finite Markov Decision Problems (MDP) is not appropriate since it is unlikely that exactly same states would be visited multiple times. An alternative methodology is based on replacing the table with a function approximation [9]. This methodology has been quite successful in many practical cases, and does not need to use explicitly the table of *Q*-values.

We propose a novel intelligent control method that is based on reinforcement learning, where in the design and on-line learning is not based on a model; rather can be implemented only by evaluative feedback during interaction with the plant. Explicit and exact modeling of system dynamics is not required; and the machine learning algorithm realizes adaptivity to uncertainties, without requiring any prior knowledge. RL framework is, in fact, a means to deal with issues arising in MPC—system-model requirement, computational complexity, and suboptimality of actions due to limited horizon over which actions are considered. A model-free predictive control (MFPC) based on RL takes decisions on infinite-horizon information, with computational cost of a one-step MPC [10].

To obtain a MFPC controller, we run RL experiment on the plant under nominal conditions (no disturbances) with no prior knowledge on the ranges of the PID parameters (trial values are used). The controller is model-free, and it learns from interactions with the environment using the values of the state variables and the cost signal at each time step. Subsequently, controller finds the parameters corresponding to steady-state behavior under nominal conditions. Simulation studies demonstrate the setpoint tracking and disturbance rejection capability of the MFPC controller based on RL. Further, to add adaptivity to uncertainties, we redesign the RL control scheme with the prior knowledge of the expected ranges of parameters for good control performance. The redesigned RL provides on-line tuning of parameters.

The proposed MFPC based on RL scheme provides a flexible approach to the design of intelligent agents (here, PID Controllers) in situations for which both conventional and supervised learning methods are impractical or difficult to be employed. Unlike conventional methods, the RL scheme is not based on approximate model of the plant; it gives parameters by on-line interaction with the plant, thereby producing results which are expected to be close enough to the design if exact model of the plant were available. Unlike supervised learning methods, RL can be applied to problems where significant domain knowledge is either unavailable or costly to obtain.

We base our analysis and simulations on fuzzy inference system (FIS) as a function approximation in RL. The fuzzy *Q*-learning [11] requires some prior knowledge. In case the required knowledge is not available or available knowledge is not reliable, one can acquire the knowledge using dynamic fuzzy-*Q* [12] learning algorithm. The proposed dynamic fuzzy-*Q* adaptive MFPC controller is very effective for complex nonlinear systems, and it doesn't need any prior knowledge to find optimal PID parameters.

The paper is structured as follows. Section 2 describes basics behind model-free predictive control. Section 3 presents the proposed reinforcement learning controller framework, its architecture and the design steps of adaptive MFPC controller. Section 4 discusses basics of model predictive control. Section 5 gives details of a highly nonlinear process—the continuous stirred tank reactor (CSTR): its dynamics and controller learning. Section 6 compares and discusses the empirical performance of set point tracking and disturbance rejection for MPC and MFPC controllers on the basis of simulation results. Finally in Section 7, the conclusions are presented.

2. MODEL FREE- PREDICTIVE CONTROL

The basic idea behind the model-free predictive control (MFPC) philosophy is to learn action values directly, by trial and error, without building an explicit model of the environment, and thus it retains no explicit estimate of the probabilities that govern state transitions. Reinforcement learning (RL) provides a way to build learning agents (intelligent controllers) that optimize their behavior in unknown environments [13, 14]. In RL, experience is built up over time through interaction with the system the agent has to control, rather than assumed available a priori (system model). The experience is based on the performance indicators that give information about how well a certain action was in a certain state of the system. The experience is also based on the state transitions of the system under actions taken. The performance function (value function) is approximated by keeping track of the performance obtained at each decision step considering the system state, performed action, and the resulting system state. At each decision step, the value function of the previous decision step is updated with experience built up over that previous decision step. By accumulating sufficient experience, the agent may accurately estimate the true value function—an infinite horizon estimation.

Thus, once the value functions are known well, an RL problem reduces to an MPC problem with a control horizon of only length one. At the same time, decisions are based on infinite horizon information. This takes care of both the issues associated with conventional MPC.

2.1 Learning framework

RL is a promising technique for adaptive control problems, where learning and control are performed