Control Engineering Practice 18 (2010) 532–539

Contents lists available at ScienceDirect

Control Engineering Practice

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

Online tuning of a supervisory fuzzy controller for low-energy building system using reinforcement learning

Zhen Yu, Arthur Dexter *

Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

ARTICLE INFO

Article history:
Received 22 March 2009
Accepted 27 January 2010
Available online 18 February 2010

Keywords:
Online tuning
Fuzzy
Reinforcement learning
Supervisory control
Building energy system

ABSTRACT

This paper proposes a model-free method using reinforcement learning scheme to tune a supervisory controller for a low-energy building system online. The training time and computational demands are reduced by basing the supervisor on sets of fuzzy rules generated by off-line optimisation and by learning the optimal values of only one parameter, which selects the most appropriate set of rules. By carefully choosing the tuning targets, discretizing the state space, parameterizing the fuzzy rule base, using fuzzy trace-back, the proposed method can complete the training process in one season.

1. Introduction

Proper supervisory control of a building energy system has the potential to reduce energy consumption, increase occupant satisfaction and lower maintenance costs (Wang & Ma, 2008). In the last 20 years, many supervisory control schemes have been proposed by researchers. The rapid progress in this field is a result of increasing industrial experience, better understanding of building energy systems, great advances in the discipline of optimisation and artificial intelligence, and, most importantly, the wide-spread use of computer based building simulation. Among many approaches, model-based optimal control has attracted most attention. Intensive studies have been undertaken by many researchers and fruitful results have been reported (Braun, 1990; Henze, 2003; Lu, Cai, Chai, & Xie, 2004; Salsbury, 2005; Wang & Ma, 2008). However, despite theoretical advances, expert rule based control is still dominant in practice (Wang & Ma, 2008). Successful applications of advanced control approaches to the supervisory control of building energy systems are rarely reported because modern optimal control schemes are widely considered by practitioners as unreliable, overly complex and time-consuming in terms of their design, tuning and maintenance (Wang & Ma, 2008). Most model-based optimisation methods require accurate models of the building system, which are very difficult and time-consuming to derive. One solution is to identify the models and calculate the supervisory control commands online (Chen, 2001). Another, less computationally demanding approach is to tune a model-free supervisory controller online (Tan & Dexter, 2002).

However, as attractive as it looks, the online tuning of a supervisory controller is not an easy task because

1. The evaluation of a current decision cannot be determined in isolation at every time step but must be based on a summation of the costs across a time horizon. A good decision is a trade-off between short-term interests and long-term interests (the so called *delay in reward*). Therefore, a series of decisions over the time horizon must be evaluated together, as a whole policy.
2. Examples of the relationship between the overall performance and the control decisions cannot be observed directly or calculated explicitly, as is the case with supervised learning, because a sufficiently accurate simplified model is not available. The tuning algorithm must learn from its own experience or recorded previous experience.
3. Because the tuning is conducted online, the performance during the tuning process is also of concern. The overall performance may actually deteriorate if too long a period of time is spent searching for the optimal control strategy. However, the algorithm cannot tell whether the current control decisions are better than decisions that have never been tried, if it is not allowed to explore previously unseen territory. This problem is called a balance between exploration and exploitation.

Reinforcement learning (RL) is a powerful unsupervised learning scheme, which has been widely studied (Kaelbling, Littman, & Moore, 1996; Barto, Sutton, & Watkins, 1998; Sutton & Barto, 1998; Barto & Mahadevan, 2003). RL is the name given to a group of methods to deal with the problems when the learning agent needs to find the

* Corresponding author. Tel.: +44 1865 2 73007; fax: +44 1865 2 73902.
E-mail address: arthur.dexter@eng.ox.ac.uk (A. Dexter).

0967-0661/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2010.01.018

optimal behaviour by interacting with an unknown environment, usually involving a *delay in reward* (Sutton & Barto, 1998). Various reinforcement learning algorithms have been suggested depending on whether the state value function is learned or the state-action function is learned; how the state-action values are changed, whether a system model is used and how it is used. Among many other RL algorithms, Q-learning is regarded as "One of the most important breakthroughs in reinforcement learning" (Sutton & Barto, 1998). It is favoured by many researchers for practical applications because it is relatively simple, converges reliably and has a solid theoretical background.

In this paper, a reinforcement learning method is used to tune a fuzzy rule based supervisory controller for building energy system. A short review of the applications of RL in built environment control is first given. Following that, in Section 3, a simplified $Q(\lambda)$ algorithm, which uses fuzzy eligibility to reduce the learning time, is presented. Section 4 describes the complex low-energy building system used for the case study. Details about how to design the learning agent are also presented. Section 5 presents the results of the online tuning process and demonstrate the scheme's ability to learn and to improve the performance of a supervisory controller, originally designed using off-line simulation based on an inaccurate model of the building system. Finally, the conclusions and future work are summarized in Section 6.

2. RL for the control of the built environment

Despite extensive research into RL by researchers working in the fields of artificial intelligence and machine learning, there have been very few applications of this powerful technique in the field of built environment or the control of building energy systems. Henze and his co-researchers have examined the use of RL to find optimal or near optimal control strategies for a building with active and passive thermal storage (Liu & Henze, 2006; Henze & Dodier, 2007), and for a photovoltaic system (Liu & Henze, 2006; Henze & Dodier, 2007). It was concluded that the performance of the RL agent is heavily reliant on the dimensions of the problem, the choice of input and output variables, and the choice of the user-defined parameters of the RL agent, and that direct application of the RL controller without prior knowledge will lead to unacceptably long training times before good performance can be achieved. Dalamagkidis and his co-workers have also investigated the use of RL to find the balance of energy and comfort in buildings (Dalamagkidis, Kolokotsa, Kalaitzakis, & Stavrakakis, 2007; Dalamagkidis & Kolokotsa, 2008). Even though a very small number of inputs and outputs were used, the training process took more than four years to converge. It was concluded that the controller should be well trained before installation and that it should not rely solely on online learning. Unlike previous applications of RL to environmental control in buildings, the method described in this paper uses fuzzy discretization, and an online learning scheme that is based on pre-generated fuzzy rule bases. As a result, the learning process is shortened and more complex problems can be tackled.

Table 1
The $Q(\lambda)$ algorithm used for the supervisory control of the building energy system.

Initialize $Q(s,a)$ arbitrarily or offline simulation

For *time*=0 to *n*
1. Observe state $s(t)$
2. Choose action $a(t)$ by using policy π: *S- > A using a strategy based on a greedy control parameter* (ε)
3. Use $a(t)$> for the next *m* hours
4. Sum all the costs that have occurred during the *m* hours, $SumCost_m$
5. After *m* hours, observe next state $s(t+1)$
6. Choose the best action $a^*(t+1)$ based on $s(t+1)$ for next step but do NOT use it
7. Calculate the cost difference: $\delta = SumCost_m + \gamma Q/(s(t+1), a^*(t+1)) - Q(s(t), a(t))$
8. Update the eligibility trace: $e(s(t), a(t))=1$
9. Update all the state-action pairs on the trace: $Q(s,a)=Q(s,a)+\beta e(s,a)\ \delta$
10. Decay the eligibility trace with λ: $e(s,a)= \lambda\ e(s,a)$

End

3. $Q(\lambda)$ learning with fuzzy discretization

Previous work has shown that the applications of RL for supervisory control of building systems suffer from the "curse of dimensionality". Building systems are complex, nonlinear, and dynamic systems with high input and output dimension, resulting in a controller with very large state and action spaces. The nonlinear character of the system also hinders the use of function generalization to reduce the learning space. Therefore, the learning time required to achieve satisfactory performance can easily become unacceptably long. As a result, the successful application of RL to the supervisory control of buildings is very much dependent on reducing the state space and action space of the

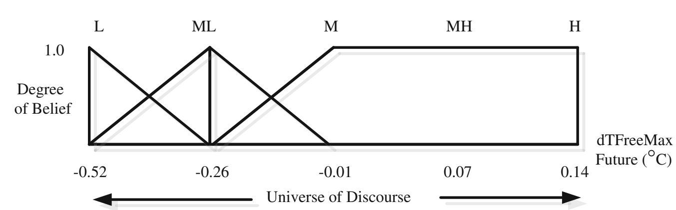

Fig. 2. Fuzzy sets used to describe DT_{future}.

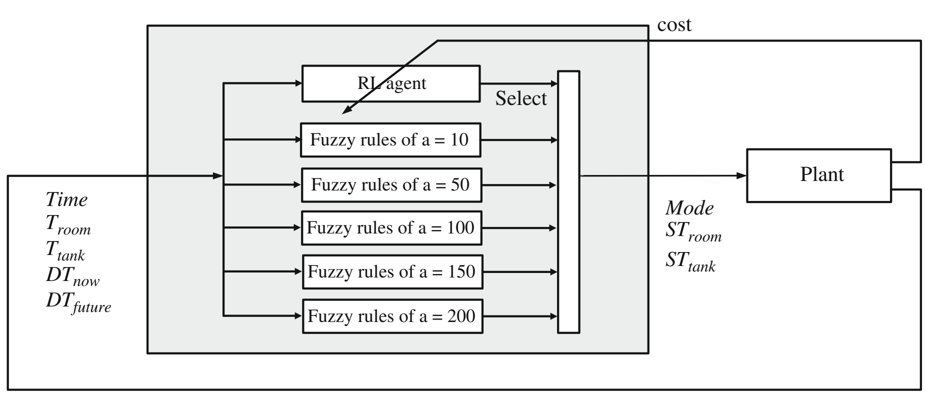

Fig. 1. RL online tuning process using an indirect learning variable.