Sustainable Energy, Grids and Networks 6 (2016) 81–90

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

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

Experimental analysis of data-driven control for a building heating system

G.T. Costanzo [a], S. Iacovella [b,d], F. Ruelens [b,d], T. Leurs [b], B.J. Claessens [c,d,*]

[a] *Danish Technical University, Department of Electrical Engineering, Frederiksborgvej 399, 4000 Roskilde, Denmark*
[b] *Electrical Engineering of KU Leuven, Kasteelpark Arenberg 10, bus 2445, 3001 Leuven, Belgium*
[c] *Flemish Institute for Technological Research (VITO), Boeretang 200, B-2400 Mol, Belgium*
[d] *EnergyVille, Thor Park Poort Genk 8130, 3600 Genk, Belgium*

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form
3 December 2015
Accepted 1 February 2016
Available online 18 February 2016

Keywords:
Thermostatically controlled load
Batch reinforcement learning
Demand response
Data-driven modeling
Fitted Q-iteration

ABSTRACT

Driven by the opportunity to harvest the flexibility related to building climate control for demand response applications, this work presents a data-driven control approach building upon recent advancements in reinforcement learning. More specifically, model-assisted batch reinforcement learning is applied to the setting of building climate control subjected to dynamic pricing. The underlying sequential decision making problem is cast into a Markov decision problem, after which the control algorithm is detailed. In this work, fitted Q-iteration is used to construct a policy from a batch of experimental tuples. In those regions of the state space where the experimental sample density is low, virtual support tuples are added using an artificial neural network. Finally, the resulting policy is shaped using domain knowledge. The control approach has been evaluated quantitatively using a simulation and qualitatively in a living lab. From the quantitative analysis it has been found that the control approach converges in approximately 20 days to obtain a control policy with a performance within 90% of the mathematical optimum. The experimental analysis confirms that within 10 to 20 days sensible policies are obtained that can be used for different outside temperature regimes.

1. Introduction

Perez et al. estimate that 20 to 40% of the global energy is consumed in buildings [1]. About half of this energy is used for HVAC [2]. As a consequence, control strategies for HVAC have received considerable academic attention in recent years. A popular class of control strategies is that of model-based strategies, such as Model Predictive Control (MPC) [3]. MPC for HVAC systems has been largely investigated in the recent literature [4–8], in both of its main aspects, modeling [9,10], and control [11]. In MPC, at regular time intervals, a control action is selected by solving an optimization problem over a finite time horizon, which is typically a day for HVAC control. In MPC the impact of future disturbances, such as internal heating and meteorological conditions, is taken into account using forecasts. Predictive control allows using the load flexibility related to thermal storage, e.g. through the thermal inertia of the building or through direct heat storage [12].

This flexibility can be harvested to enable demand response and provide load control services, which value has been increasing together with the share of renewable energy in the production mix. Examples of services are peak shaving and valley filling for a distribution system operator [13], ancillary services towards a transmission system operator [14,15] or energy arbitrage [16]. However, deploying MPC can be a challenging task. The most significant challenge is to derive an accurate model which, in the case of thermal control, has to include the thermal dynamics and the actuation model. In [17], Širokỳ et al. give a detailed report on implementation issues of MPC controllers for building heating systems.

In this context, completely data-driven approaches are deemed interesting, sacrificing performance for practicality. One possible embodiment uses data-driven model in combination with an optimization algorithm to obtain a control policy [18]. Alternatively, it is possible to learn directly the control policy by estimating a state-action value function through interaction with the system. For example in [19], Reinforcement Learning (RL), a model-free control approach is applied to building thermal storage. In RL, the policy is updated online, i.e. at each time step. In Batch Reinforcement Learning (BRL), on the other hand, the policy is calculated offline

[*] Corresponding author. Tel.: +32 14335910.
E-mail address: bert.claessens@vito.be (B.J. Claessens).

http://dx.doi.org/10.1016/j.segan.2016.02.002
2352-4677/© 2016 Elsevier Ltd. All rights reserved.

using a batch of historical data. Even though (B)RL is getting more mature [20], as discussed in [21], combining techniques of RL with prior (domain) knowledge is a logical control paradigm. It is towards this direction that this paper is positioned, i.e. in applying BRL in combination with prior knowledge to the operation of a building climate control system for demand response applications.

The basis of our approach is BRL with Fitted Q-Iteration (FQI) [22,23], where the learning of an optimal control policy is enhanced by *virtual* data coming from a model. For this reason, such approach is called Model-Assisted Batch Reinforcement Learning (MABRL) as discussed in [24].

In Section 2, an overview of the related literature is provided and the contribution of this work is explained. Following the approach presented in [25], in Section 3 formulizes the building thermal scheduling is formalized as a sequential decision making problem under uncertainty. In Section 4 MABRL is detailed, while Section 5 presents a quantitative and qualitative assessment of the performance of the controller. Finally, Section 6 outlines the conclusions and discusses future research directions.

2. Related work

This section gives a non-exhaustive overview of related work regarding MPC and RL for building climate control, after which the main contributions of this work are explained.

2.1. Model predictive control

When considering building climate control, MPC has received considerable attention in the recent literature [6–8,26]. The overview of practical issues related to the implementation of an MPC controller can be found in [27]. The key elements of MPC comprise: mathematical model(s) of the building dynamics, comfort requirements and exogenous information such as user behavior and outdoor temperature. This information is used to cast an optimization problem that is solved to define optimal control actions with respect to a defined objective function, subject to constraints provided by the model.

In typical embodiments of MPC one tries to formalize the problem as a mixed-integer problem to allow using fast solvers with performance guarantees. Therefore, a Linear Time Invariant model (LTI) of the system under control is to be identified. If no domain knowledge is available, black-box identification techniques are used, such as subspace identification methods [28,29]. Alternatively, gray-box models can be used, where the model structure is defined and the parameters are estimated using experimental data [9]. In the context of thermal modeling a number of studies use thermal circuits [30–34].

Advanced climate control allows, besides efficient use of energy and comfort management, integration within aggregation schemes to provide ancillary services and portfolio management in demand side management [35]. For example, in [36] the aggregated flexibility of a cluster of buildings is used to provide balancing services using an aggregate-and-dispatch approach.

An alternative for LTI modeling is to use non-linear data-driven models, such as Artificial Neural Networks (ANNs) [18,37], in combination with Dynamic Programming (DP) [38] to compute a control policy. This form of control can be seen as a form of RL [39].

2.2. Reinforcement learning

As discussed in Section 1 RL is a model-free control technique whereby a control policy is learned from interactions with the environment. A well established reinforcement learning method is Q-learning [40] where the state-action value function, or Q-function, is learned. Compared to techniques provided in the previous section, RL mitigates the risk of model-bias [24] as a policy is built around the data. When considering Q-learning and its applications to demand response, mainly traditional Q-learning has been used [19,41,42]. More recently BRL [43,44] in the form FQI [21] has been investigated. The main advantage of BRL is the practical learning time required for convergence (20–40 days in [43,44]) which comes at the cost of an increased computational complexity. Although BRL can rival the performance of MPC techniques, as indicated in [21], the context of demand response allows to add *prior knowledge* to the optimal control problem that can result in faster convergence. A first approach uses prior knowledge by shaping the policy, obtained with FQI, by means of constrained regression [22]. A second approach is described by Lampe and Riedmiller in [24]. Here virtual data from a model is used together with experimental data to obtain an approximation of the Q-function (state-action value function).

Building upon [22,24,43], this work has the following contributions:

- BRL, in the form of FQI, [21] in combination with virtual trajectories [24] and policy shaping is applied to a HVAC system for a typical objective of dynamic pricing [45]. This effectively results in a data-driven solution for building climate control systems, combining state-of-the-art BRL with domain knowledge;
- Quantitative and qualitative performance assessment of MABRL in a simulated and experimental environment, where the operation of an air conditioner is subject to dynamic energy pricing.

3. Problem formulation

Before presenting the control approach in Section 4, this section formulates the decision-making process as a Markov Decision Process (MDP) [38,46]. An MDP is defined by its state space X, its action space U, and a transition function f:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k), \tag{1}$$

which describes the dynamics from $\mathbf{x}_k \in X$ to $\mathbf{x}_{k+1}$, under the control action $\mathbf{u}_k \in U$, and subject to a random process $\mathbf{w}_k \in W$, with probability distribution $p_w(\cdot, \mathbf{x}_k)$. The reward accompanying each state transition is r_k:

$$r_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{x}_{k+1}) = \rho(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \tag{2}$$

which is here considered as a cost, since it accounts for the energy price. Therefore, the objective is to find a control policy $h : X \to U$ that minimizes the T-stage cost starting from state $\mathbf{x}_1$, denoted by $J^h(\mathbf{x}_1)$:

$$J^h(\mathbf{x}_1) = \mathbb{E}\left(R^h(\mathbf{x}_1, \mathbf{w}_1, \dots, \mathbf{w}_T)\right), \tag{3}$$

with:

$$R^h(\mathbf{x}_1, \mathbf{w}_1, \dots, \mathbf{w}_T) = \sum_{k=1}^{T} \rho(\mathbf{x}_k, h(\mathbf{x}_k), \mathbf{w}_k). \tag{4}$$

It is worth remarking that an optimal control policy, here denoted by h^*, satisfies the Bellman optimality equation:

$$J^{h^*}(\mathbf{x}) = \min_{\mathbf{u}} \mathop{\mathbb{E}}_{\mathbf{w} \sim P_w(.|\mathbf{x})} \{\rho(\mathbf{x}, \mathbf{u}, \mathbf{w}) + J^{h^*}(f(\mathbf{x}, \mathbf{u}, \mathbf{w}))\}. \tag{5}$$

Typical techniques to find policies in an MDP framework are value iteration, policy iteration, and policy search [22]. As mentioned earlier, in this work MABRL (related to value iteration) is considered.