We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Reinforcement Learning for Building Environmental Control

Konstantinos Dalamagkidis[1] and Dionysia Kolokotsa[2]
[1]University of South Florida,[2] Technical Educational Institute of Crete
[1]USA,
[2]Greece

1. Introduction

During the last few decades, significant changes have been made in the area of building construction and management, especially regarding climate control and energy conservation.

A significant turning point was reached in the early 70s with the oil crisis driving a movement for airtight buildings to minimize energy consumption. Unfortunately this turn resulted in a significant deterioration of the indoor air quality, raising health concerns around the world. This started a more involved study of human comfort with respect to air quality, lighting and temperature among other factors.

There has been a drive in recent years to enhance current Building Environmental Management Systems (BEMS) with decision logic that takes into account all of the aforementioned issues namely thermal comfort, visual comfort, air quality and energy consumption. In order to maximize performance on all of the above indexes, the BEMS controller can use among other things the mechanical HVAC system, natural ventilation through windows, artificial lighting and shading devices.

There are several aspects of the problem that make it attractive to intelligent control implementations. First of all the knowledge of the state of the indoor environment is imprecise due to several reasons. Localized phenomena can affect parameters like temperature or air velocity making it impossible to measure them accurately. Building environments are also characterized by changing dynamics due to human activity as well as equipment and building aging. Some parameters like clothing and activity type that are normally required to accurately estimate thermal comfort are difficult or even impossible to measure. Finally it should also be noted that despite the existence of mathematical models, thermal comfort remains a subjective measure and thus any such model is characterized by some error. On the other hand the action space is discrete and of small dimensionality.

The nature of the problem therefore indicates that controllers that are able to generalize can offer a good performance. This is also suggested by the numerous controllers proposed in the literature ranging from classic PD/PID to fuzzy, neural networks and their combinations.

For the reasons mentioned above, reinforcement learning is also suited for this problem, but it also has some unique features that make this approach of particular interest. Although a

specific setting of the HVAC system has an immediate impact on energy consumption, its effects on the indoor climate can endure for longer periods of time. Using reinforcement learning, it is straightforward to capture this effect through discounting, while the same would be more involved for other types of controllers. Additionally, its on-line learning capabilities allow the controller to follow gradually changing dynamics and minimize the adverse effects of abrupt changes, without need for significant reconfiguration.

This chapter will begin with a detailed presentation of the problem and prior art in the field. Controller design issues and alternatives will be discussed, although rigorous theoretical analysis will be omitted for the sake of brevity. A simple case study will be presented and the performance of the reinforcement learning controller will be compared to that of a fuzzy-PD and a typical on/off controller under various scenarios. The chapter concludes with current issues and suggestions for future research.

2. Comfort

After the emergence of the sick building syndrome and the realization that sealed indoor environments can have adverse effects on health and productivity, significant attention is now given to the comfort of buildings' occupants. Modern bioclimatic architecture dictates an exploitation of local climatic and geographic characteristics to provide a comfortable environment while minimizing energy consumption. On the other hand urban construction poses some limitations in the application of bioclimatic architecture and of course there are millions of buildings already constructed and in use. As a result the need arises for the introduction of sophisticated control systems. In order for such systems to be designed, comfort must be defined and quantified first.

When we refer to the comfort level of a building occupant, we have to consider several factors like thermal comfort, indoor air quality as well as light and noise levels. There are already published standards in the area of comfort and there are others in development. Thermal comfort is addressed in ISO 7730:1994 and ASHRAE 55. ISO 8995:1989 describes lighting demands of indoor work environments and ISO 1996-3:1987 and 1999:1990 describe noise limits and the impact of noise on human hearing. The Centre Europeen de Normalisation is also preparing prEN 15251:2006 under the title "Criteria for the indoor environment including thermal, indoor air quality, light and noise" (Olesen et al, 2006)

2.1 Thermal comfort

Thermal comfort is defined in the ISO 7730 standard as being "That condition of mind which expresses satisfaction with the thermal environment". Although this definition does not directly provide the means to measure thermal comfort, the standard proposes the use of the Fanger model. This model is based on the following equilibrium of the human body (Fanger, 1970):

$$M - W = H + E_c + C_{res} + E_{res} \qquad (1)$$

Where M is the metabolic rate (provided in tables as a function of activity), W is the external work (usually considered to be zero), H is the dry heat loss, E_c the evaporative heat exchange at the skin during thermal neutrality, C_{res} the convective respiratory heat exchange and E_{res} the evaporative respiratory heat exchange. The latter four require the knowledge of

the air temperature, the mean radiant temperature, the air velocity, the humidity and the clothing type in order to be evaluated.

In actual conditions several of the above parameters (air velocity, clothing type, activity) vary from occupant to occupant and as a result the task of measuring and achieving thermal comfort for all occupants is almost impossible (Olesen & Parsons, 2002). On the other hand people have some ability for self-regulation by adjusting their clothes or opening a window. The Fanger model provides two indexes of the thermal comfort. The first is the PMV (Predicted Mean Vote) that corresponds to the average vote of a large volume of people about their thermal sensation and ranges between -3 (very cold) to +3 (very hot). The second is the PPD (Percent of People Dissatisfied) and is derived from the PMV using the following relationship (Memarzadeh & Manning, 2000):

$$PPD = 100 - 95 \exp(-0.03353\ PMV^4 - 0.2179\ PMV^2) \tag{2}$$

It should be noted that the PMV, PPD indexes are based on North American and European healthy adults in sedentary activity and the ISO standard warns against applying it to different groups.

The last decade the Fanger model has seen some criticism especially because of its inaccurate predictions for naturally ventilated buildings. This is because the model doesn't take into account the psychological adaptability that people exhibit, i.e. people living in naturally ventilated buildings are used to the large diversity and exhibit different preferences and wider tolerances (de Dear & Brager, 2002).

As an alternative to the Fanger model for fully air-conditioned buildings, the ACS or Adaptive Comfort Standard was proposed for naturally ventilated buildings. More specifically (de Dear & Brager, 2002) hypothesized that the thermal comfort temperature in the latter type buildings is a function of only the outdoor temperature:

$$T_{comfort} = 0.31\ T_{out} + 17.8 \tag{3}$$

In a similar study (Nicol & Humphreys, 2002) developed their comfort equation as a function of the monthly average temperature:

$$T_{comfort} = 0.54\ T_{out,avg} + 13.5 \tag{4}$$

Using the same methodology (McCartney & Nicol, 2002) developed comfort equations of the same type for 5 different European countries using the running mean temperature. These efforts where continued with studies for other regions like Singapore (Wong et al, 2002) and Indonesia (Feriadi & Wong, 2004)

Although it might seem that the ACS models differ significantly from study to study, its significance is still high because of its simplicity and lack of better alternatives for naturally ventilated buildings.

2.2 Air quality

It is a common misconception that the polluted air is outside, when the truth is that indoor air concentrations of various irritating, carcinogenic and mutagenic compounds can be higher than their corresponding outdoor concentrations even in industrial areas. Some of the most common pollutants found in indoor environments are radon (a carcinogenic compound that originates from the ground), CO and other pollutants from cigarette smoke,

volatile organic compounds from detergents, disinfectants, glues, paints etc and ozone from copying machines and air cleaners.

The only way to control the concentration of these pollutants is by introducing as much fresh air into the building as possible, although that is usually at some cost because the fresh air normally needs to be cooled or heated depending on the climate. As an indicator of indoor air quality it is common to use the concentration of CO_2 where such information is available. This is under the assumption that the concentrations of the other pollutants will follow similar trends, which in some cases may not be accurate.

2.3 Light requirements

Sufficient light is important in establishing a feeling of comfort and maintaining high productivity. Light comfort depends on illuminance or the adequacy of light, glare and light color (Serra, 1998). Therefore to enable people to perform visual tasks, adequate light without side glare and blinding must be provided. The required luminance levels can be reached by means of daylight, artificial light or a combination of both. For reasons of health, comfort and energy in most cases the use of daylight is preferred over the use of artificial light (Serra, 1998). The use of daylight though, depends on many factors like occupancy hours, autonomy, building location, daylight hours during summer and winter, window openings and orientation.

To make sure that for a reasonable amount of occupancy time daylight can be used, demands on the daylight penetration in the spaces meant for human occupancy have been set (CEN, 2002). Despite the fact that lighting can be controlled also by controlling shading devices, in commercial BEMS it is common to control only the artificial lighting to complement natural light when required.

2.3 Noise

Although noise can significantly affect the feeling of comfort and in some cases even have temporary or permanent health effects, unfortunately it is impossible to have any control over the noise levels via the BEMS. As a result this issue is dealt independently during the construction and operation of the building.

3. State of the art

Recent developments in BEMS have been influenced by the wider adaptation of intelligent control techniques like fuzzy systems, genetic algorithms and neural networks. Many of the controllers proposed in the literature have some provisions for thermal comfort and almost all have the reduction of energy consumption as objective function.

Starting with fuzzy systems, (Hamdi & Lachiver, 1998) designed a controller consisting of two separate fuzzy modules, one to determine the comfort zone and the other to provide the actual control. In (Salgado et al., 1997) fuzzy on/off and fuzzy-PID controllers where proposed with improved performance over their non-fuzzy counterparts. A PMV-based fuzzy controller was chosen by (Dounis & Manolakis, 2001) while (Kolokotsa et al. 2001) presented a family of fuzzy controllers that regulate also air quality and visual comfort. In (Alcala et al., 2005) the use of weighted linguistic fuzzy rule sets was proposed for controlling heating, cooling and ventilation systems. A

review of artificial intelligence in buildings with a focus on fuzzy systems is assessed by (Kolokotsa, 2006).

Genetic algorithms have been used to optimize the parameters of control systems in (Huang & Lam, 1997) as well as in (Kolokotsa et al., 2002). Similarly (Egilegor et al., 1997) used neural networks to adapt the parameters of their fuzzy-PI controller.

Neural networks have also been used independently for control as in the case of (Karatasou et al., 2006) or cooperatively with fuzzy systems as in (Yamada et al., 1999).

Some other approaches include the neurobat project by (Morel et al., 2001) for predictive control, empirical models used in (Yao et al, 2004) and decision support systems using rule sets proposed by (Doukas et al., 2007).

4. Designing a reinforcement learning controller

Despite the fact that a BEMS controller is also responsible for the artificial lighting, it is possible to delegate the lighting control to a separate slave controller, under the assumption that the two controllers are independent. This is because the output of the controller depends exclusively on the light conditions inside the building and that output only has an almost negligible effect on the thermal environment. As a result the state-action dimensionality is kept low and the controller can learn faster.

The choice of the learning algorithm is based on several factors. Since the state space is continuous and only a relatively small number of discrete actions are available, algorithms based on function approximation are particularly suitable. The authors have used with success the recursive least-squares (RLS-TD) algorithm proposed by (Xu et al., 2002) and presented in Fig. 1. Nevertheless other approaches are also available.

A very important aspect of the design of a controller that is based on reinforcement learning is the definition of a reward function. In the case of building environmental control, the reward can be based on more than one factors. To make the quantities comparable it is prudent to scale everything, so that they take values in the same range - usually [0,1].

The energy consumption can be scaled using the ratio of the current consumption that includes the heating, air conditioning and ventilation, to the maximum possible consumption. For the thermal comfort the PPD index can be chosen, since it is already in the desired range. Care should be taken to use the PPD only in the cases where it accurately models the reality, according with the guidelines of the ISO standard. Finally in order to create the indoor air quality index based on the CO_2 concentration, it is possible to use an appropriately chosen sigmoid function of the following form:

$$f(x) = 1/(1+\exp(-\alpha x + \beta)) \tag{5}$$

The constant parameters α, β for the sigmoid function are chosen empirically based on the fact that for CO_2 concentrations close to the average outdoor concentrations, the index should take near zero values. Similarly double the average outdoor concentration should yield an index close to 1.

Since all the rewards are in the same range with zero being the best value and one the worst, it is possible to create a simple reward function based on the weighted average of all the indexes:

$$r = -(w_1 r_{energy} + w_2 r_{comfort} + w_3 r_{air\_quality}) \tag{6}$$

where the w_i are weights chosen empirically and the minus sign is because learning tries to maximize the reward.

Using the sigmoid function described above it is also possible to scale the inputs to the controller to the same range. This approach can help with simplifying the coding scheme that precedes the feature vector generation. Besides the usual inputs (temperatures, CO_2 concentrations, etc.) an additional input was given to the controller regarding the time of the year and the time of the day. Although this choice increases the dimensionality of the state space, it gives the controller sufficient information to anticipate the changes in the outdoor environment and choose better actions.

The learning parameters are chosen based primarily on experience and trial and error. Special care should be taken in the choice of the decay and discount factors. This is because their effect depends also on the timestep chosen. As an example for a timestep of 10 minutes a decay factor of 0.5-0.6 should be used so that the eligibility trace reduces significantly after 30-40 minutes. If the timestep is lowered to 5 minutes, then the decay factor should be increased to 0.7-0.8 to get the same effect and comparable results.

For a given set of basis functions $\varphi(x)$
Initialization:
  Initialize weight vector W_t randomly
  Variance matrix $P_o = \delta I$
  Eligibility traces $z_o = 0$
Loop:
  For current state x_t and chosen action a_t, observe new state x_{t+1} and reward r_t
  $K_{t+1} = P_t z_t \ / \ (1+(\varphi^T(x_t) - \gamma\,\varphi^T(x_{t+1}))\,P_t z_t\,)$
  $W_{t+1} = W_t + K_{t+1}(r_t - (\varphi^T(x_t) - \gamma\,\varphi^T(x_{t+1}))W_t)$
  $P_{t+1} = P_t - K_{t+1}(\varphi^T(x_t) - \gamma\,\varphi^T(x_{t+1}))P_t$
  $z_{t+1} = \gamma\lambda z_t + \varphi(x_t)$

Fig. 1. RLS-TD(λ) algorithm from (Xu et al., 2002)

5. Case study and sample results

As an example the case study of a simulated simple building will be investigated. The building is comprised of a single 35m² room, with one window to the north (1m²) and one to the west (1m²). All the exterior walls are insulated concrete with a total thickness of 22.5cm. The sensor information available is the indoor and outdoor temperatures, the current PPD and the date and time. All these variables are used as the state input to the controller with the exception of the PPD which is used only in determining the reward function.

The inputs where scaled using sigmoid functions and then encoded using radial basis functions, to create the feature vector.

The environmental control equipment is composed of a heat pump that has a cooling and heating mode and 3 levels for each mode (low, medium, high) with specifications similar to modern AC inverter units and a ventilation unit that has a low and high setting. A window actuator is also available that can open or close the windows. As a result, all the possible combinations give a total of 42 different actions.

The reward function is given by equation 7, where the weights were chosen after trial and error to fairly balance the importance of energy conservation and thermal comfort.

$$r = -(0.4\ r_{energy} + 0.6\ r_{comfort}) \tag{7}$$

The decay chosen was 0.5 and the discount factor was 0.85, nevertheless it should be noted that small perturbations around those values did not significantly alter the performance of the controller. The ε parameter was chosen to be 0.05 for the first two years and 0 afterwards.

The results of the first 5 years of simulation are summarized in Table 1, along with comparative results from a single year simulation of an on/off and a fuzzy-PD controller. It should be noted that the parameters of the fuzzy-PD were chosen empirically and where not optimized with training data.

	Reinforcement Learning					Fuzzy-PD	On/off
Year	1	2	3	4	5	5	5
Energy	19.47%	17.16%	15.82%	12.73%	12.78%	11.17%	12.89%
Avg. PPD	21.1%	15.3%	12.9%	13.7%	13.6%	15.8%	14.44%
Max PPD	100%	79.8%	65.7%	58.3%	56.4%	60.3%	62.2%

Table 1. Results from 5 years of simulated building response to the reinforcement learning controller. The energy index is given as the ratio of the integrated power consumption to the maximum possible consumption in the same period.
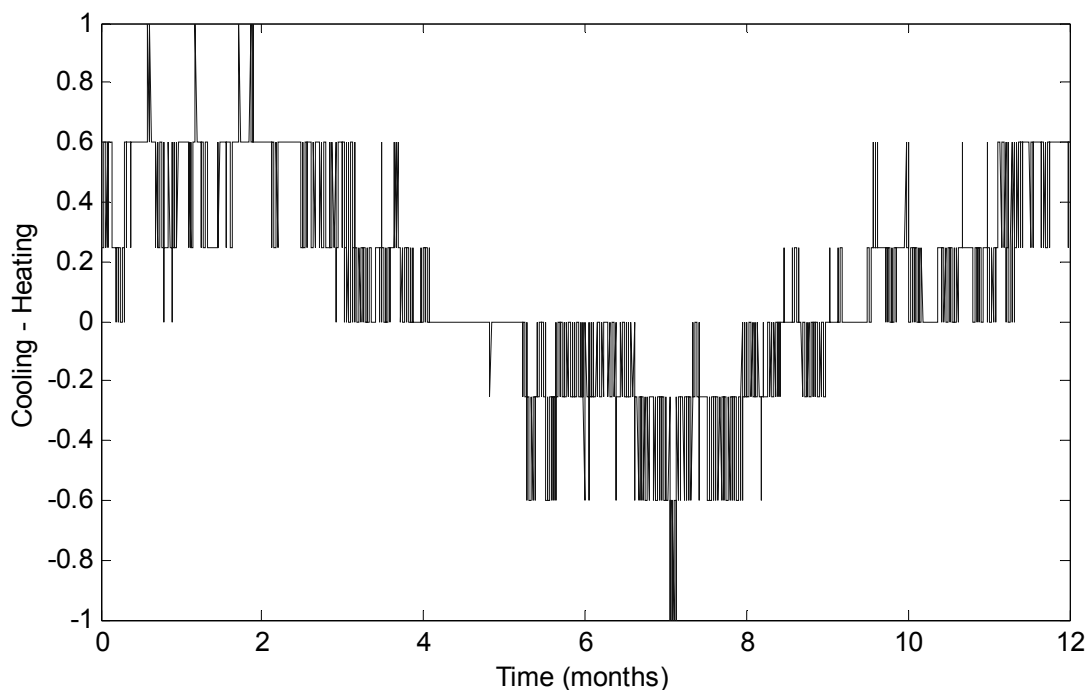

Fig. 2. A plot of cooling/heating controller signal after 4 years of simulated training (1 in 10 points are shown for better clarity).

In Fig. 2 the control signal for the heat pump is presented. It can be observed that the controller does not make significant errors and rarely chooses to use the heat pump at the high settings. Nevertheless the fact remains that the controller is still learning and its

behavior still hasn't converged. Similarly Fig. 3 shows the PMV and PPD indexes in the building for the same period. Here it is more evident that the controller is still learning to maintain an acceptable PPD level throughout the simulated year. Nevertheless it is also obvious that it is able to keep the PPD under 25% more than 96% of the time. It should be noted that the PPD cannot drop below 5% by definition.
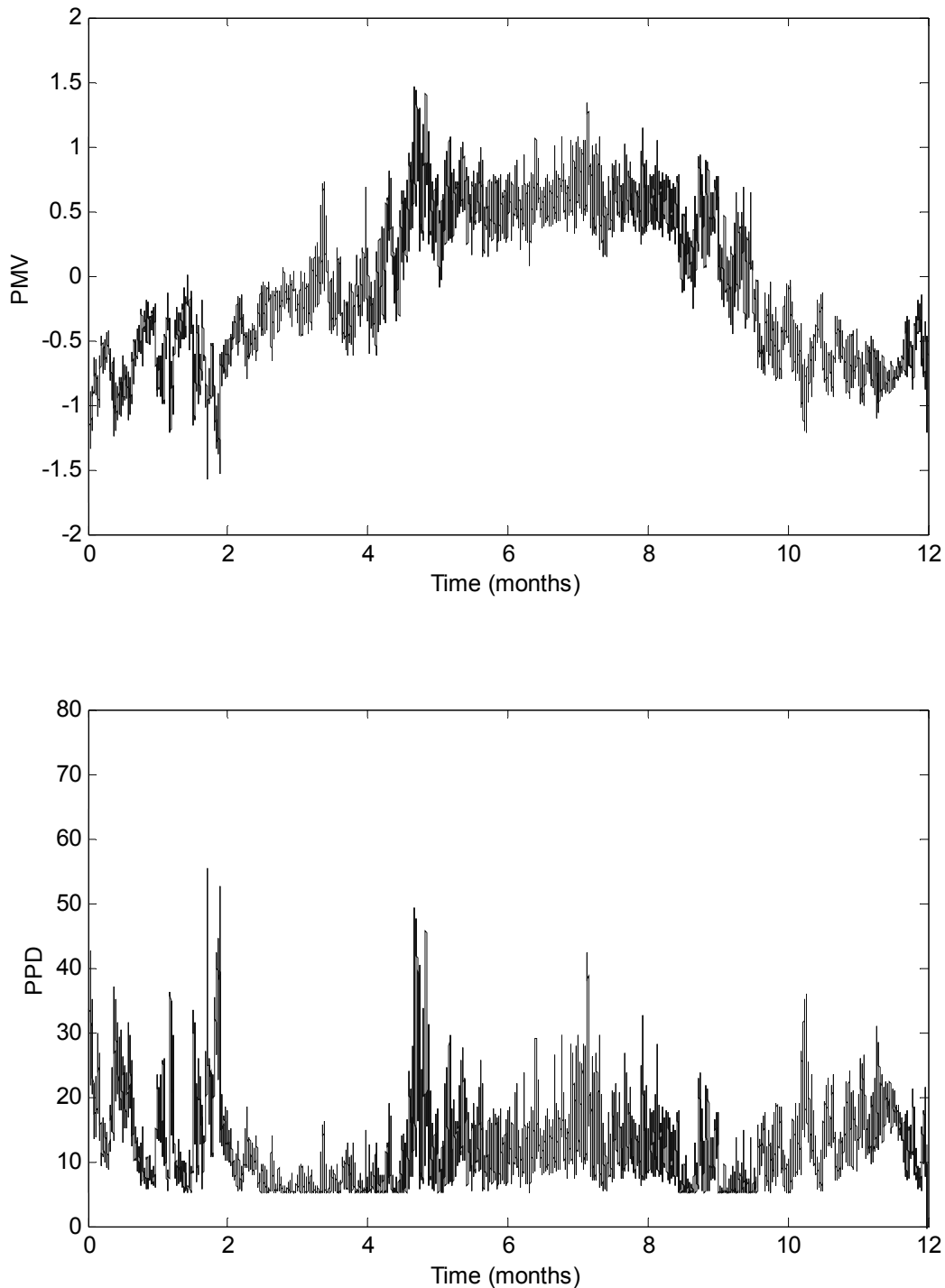

Fig. 3. The PMV and PPD indexes as a function of time, after 4 years of simulated time.

The results presented above show that the reinforcement learning controller within 4 years of simulated time has achieved performance comparable if not better than other controllers. Nevertheless even after the 4 years the controller is still changing its policy and further improvements are to be expected by more training, optimization of the training parameters and more descriptive feature vectors.

The advantage that this controller has over the other controllers is that it is possible to continually adapt the weight vector, thus allowing it to follow gradual changes in the behavior of a building. These changes can stem from equipment aging, leakages or changes in the climatic patterns in the area. On the other hand since in a real building suboptimal action selection is usually not permitted, exploration during the operation of the controller should be avoided. As a result it is necessary to exhaustively train the controller beforehand in a similar simulation environment.

6. Future research opportunities

This chapter has demonstrated the utility of reinforcement learning in the development of controllers for BEMS. It is also evident that the potential benefits both in terms of energy conservation as well as in terms of comfort are significant. Nevertheless the results presented above and earlier in (Dalamagkidis et al, 2007) signify that there is still room for improvements. Besides better fine-tuning of the parameters and further training, the authors would like to propose some other ideas for future work.

On the first order of business is the development of a better reward assignment algorithm. Although only hinted at in previous sections, the reward mechanism described in this chapter does not represent the real phenomenon with accuracy. A separation of the energy consumption related reward from the rest is suggested. This is because any control action is associated with a very specific energy consumption that occurs only while this control action is in effect. On the other hand the impact of that control action on the other comfort related factors lingers for several minutes after that action has ceased to occur.

As an example the effect of the heat pump control on the indoor temperature is presented in Fig. 4. Although most of the effect on the temperature occurs within the time that the control action was active, there is significant after-effect, that lasts almost for 40 minutes after that control action ceased. The same is true for the indoor air quality index which represents a more complicated case since it depends on the rate of air exchanges and thus depends exclusively on the window and ventilator actions.

For the case study presented earlier, it was decided to represent the complete reward as a simple weighted average of the energy and thermal comfort indexes and a decay factor was chosen that reduces the weight of the reward after 30-40 minutes. The suggested alternative solution is an n-step backup that waits until the real reward is available, with the known disadvantages presented by (Sutton & Barto, 1998). A different scheme would involve a more complicated value-updating step, which would entail an immediate update based on the energy performance and eligibility traces (perhaps even separate) for the other indexes. This mixture of TD(0) and TD(λ) of course needs to be evaluated for its convergence properties. The authors expect that an improved reward assignment algorithm could benefit both the final performance of the controller as well as the convergence speed.
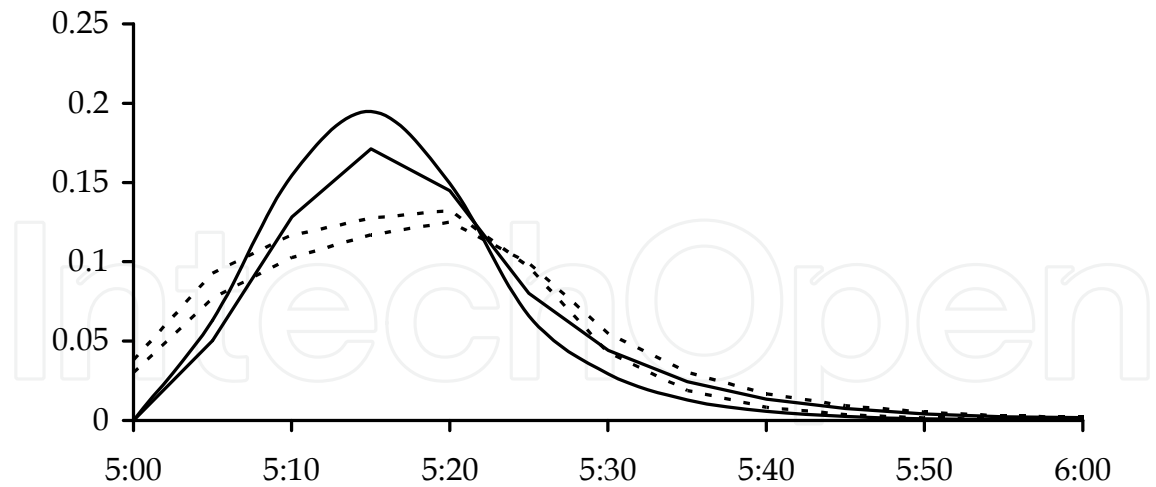
Fig. 4. A plot of the absolute temperature deviation after a control command that begun at 5:00 and lasted 20 minutes, with respect to no control. The two filled lines represent cooling for two different buildings and the dashed lines heating.

An additional method to increase the learning speed is to have two different speeds during learning. The first speed would correspond to the value function update speed which could be as frequent as every 2 or 3 minutes, while the second speed would correspond to the frequency with which a new action is selected and should be around 15 to 20 minutes. This change would require using different learning algorithms since now the controller would be operating off-policy. Additionally the learning parameters would need to be adjusted to correspond to the new temporal characteristics of the learning process.

Although the scientific literature of the last couple of decades is full with proposals of classic and intelligent controllers that claim significant savings in terms of energy and superior performance in terms of comfort, the technology mostly used today is still the basic thermostat. Besides being a simple and tried technology, the thermostat also has two features that help it retain its position. The first is that it requires only a temperature sensor that is usually incorporated in the device and the second is that it allows people to directly control their environment thus contributing to their feeling of comfort.

It is therefore the authors' opinion that in order for a modern controller to be successful and replace the installed base of controllers, it needs the aforementioned characteristics in some degree. Reducing the sensor demand of the controller can be difficult. Nonetheless it is a straightforward process. On the other hand finding ways for the occupants to interact with the controller is far more complicated. In (Dalamagkidis et al, 2007) an additional module was proposed that is trained by occupant input whenever the latter is available. This module is then used as another component in determining the reward function. Another idea is for the occupants to be able to override to a degree some of the controller's actions and the controller learning from that experience.

Regardless of the number of papers proposing new and more efficient technologies for building environmental control, the fact remains that the penetration of these technologies in the market is minimal if any. To open the doors for the introduction of more advanced controllers in BEMS, more work should be done on controllers that can efficiently replace what is already installed with little to no additional requirements.

7. References

Alcala, R.; Casillas, J.; Cordon, O.; Gonzalez, A. & Herrera, F. (2005). A genetic rule weighting and selection process for fuzzy control of heating, ventilating and air conditioning systems. *Engineering Applications of Artificial Intelligence*, Vol. 18, No. 3, April 2005, pp. 279-296

CEN, 2002. EN 12464-1:2002, Light and lighting. Lighting of work places. Indoor work places

Dalamagkidis, K.; Kolokotsa, D.; Kalaitzakis, K. & Stavrakakis, G. S. (2007). Reinforcement learning for energy conservation and comfort in buildings. *Building and Environment*, Vol. 42, No. 7, July 2007, pp. 2686-2698

De Dear, R. J. & Brager, G. S. (2002). Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55. *Energy and Buildings*, Vol. 34, No. 6, July 2002, pp.549-561

Doukas, H.; Patlitzianas, K. D.; Iatropoulos, K. & Psarras, J. (2007). Intelligent building energy management system using rule sets, *Building and* Environment, Vol. 42, No. 10, October 2007, pp.3562-3569

Dounis, A. I. & Manolakis, D. E. (2001). Design of a fuzzy system for living space thermal-comfort regulation. *Applied Energy*, Vol. 69, No. 2, pp. 119-144

Egilegor, B.; Uribe, J. P.; Arregi, G.; Pradilla, E. & Susperregi, L. (1997). A fuzzy control adapted by a neural network to maintain a dwelling within thermal comfort. *Proceedings of Building Simulation 1997*, Vol. 2, September 1997, pp. 87-94

Fanger, P. O. (1970). Thermal comfort analysis and applications in environmental engineering. Mc Graw Hill, New York

Feriadi, H. & Wong, N. H. (2004). Thermal comfort for naturally ventilated houses in Indonesia. *Energy and Buildings*, Vol. 36, No. 7, July 2004, pp.614-626

Hamdi, M. & Lachiver, G. (1998). A fuzzy control system based on the human sensation of thermal comfort, In: *Fuzzy Systems, The 1998 IEEE international conference on*, Vol. 1, May 1998, pp. 487-492

Huang, W. & Lam, H. N. (1997). Using genetic algorithms to optimize controller parameters for HVAC systems. *Energy and Buildings*, Vol. 26, No.3, pp. 277-282

Karatasou, S.; Santamouris M. & Geros V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, Vol. 38, No. 8, August 2006, pp. 949-958

Kolokotsa, D.; Tsiavos, D.; Stavrakakis, G. S.; Kalaitzakis, K. & Antonidakis, E. (2001). Advanced fuzzy logic controllers design and evaluation for buildings' occupants thermal – visual comfort and indoor air quality satisfaction. *Energy and Buildings*, Vol. 33, No. 6, July 2001, pp. 531-543

Kolokotsa, D.; Stavrakakis, G. S.; Kalaitzakis, K. & Agoris, D. (2002). Genetic algorithms for optimized fuzzy controller for the indoor environmental management in buildings implemented using PLC and local operating networks. *Engineering Applications of Artificial Intelligence*, Vol. 15, No. 5, September 2002, pp. 417-428.

Kolokotsa, D. (2006). Artificial Intelligence in Buildings: A Review on the Application of Fuzzy Logic, *Advances in Building Energy Research*, Vol.1, pp. 29-54, James and James Publisher.

McCartney, K. J. & Nicol, J. F. (2002). Developing an adaptive control algorithm for Europe. *Energy and buildings*, Vol. 34, No.6, July 2002, pp. 623-635

Memarzadeh, F. & Manning, A. (2000). Thermal comfort, uniformity and ventilation effectiveness in patient rooms: performance assessment using ventilation indices, ASHRAE Annual Meeting, MN-00-11-3

Morel, N.; Bauer, M.; El-Khoury, M. & Krauss, J. (2001). Neurobat, a predictive and adaptive heating control system using artificial neural networks. *International Journal of Solar Energy*. Vol. 21, April 2001, pp. 161-202

Nicol, J. F. & Humphreys, M. A. (2002). Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and Buildings*, Vol. 34, No. 6, July 2002, pp. 563-572

Olesen, B. W. & Parsons, K. C. (2002). Introduction to thermal comfort standards and to the proposed new version of EN ISO 7730. *Energy and Buildings*, Vol.34, No. 6, July 2002, pp.537-548

Olesen, B. W.; Seppanen, O. & Boerstra, A. (2006). Criteria for the indoor environment for energy performance of buildings: A new European standard. *Facilities*, Vol. 24, No 11/12, pp. 445-457

Salgado, P.; Cunha, J. B. & Couto, C. (1997). A computer-based fuzzy temperature controller for environmental chambers, In: *Industrial Electronics, IEEE 1997 international symposium on*, Vol. 3, July 1997, pp. 1151-1156

Serra, R. (1998). Daylighting, In: *Renewable and sustainable energy reviews*, Vol. 2, No.1-2, June 1998, pp. 115-155

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement learning: an introduction*, MIT Press, ISBN 0-262-19398-1, Cambridge, MA

Tsitsiklis, J. & Van Roy, B. (1997). An Analysis of Temporal-Difference Learning with Function Approximation. *IEEE Transactions on Automatic Control*, Vol. 42, No. 5, May 1997, pp.674-690

Wong, N. H.; Feriadi, H.; Lim, P. Y.; Tham, K. W., Sekhar, C. & Cheong K. W. (2002). Thermal Comfort evaluation of naturally ventilated public housing in Singapore. *Building and Environment*, Vol. 37, No. 12, December 2002, pp. 1267-1277

Xu, X.; He, H. G. & Hu, D. (2002). Efficient reinforcement learning using recursive least-squares methods. *Journal of Artificial Intelligence Research*, Vol. 16, pp.259-292

Yamada, F.; Yonezawa, K.; Sugarawa, S. & Nishimura, N. (1999). Development of air-conditioning control algorithm for building energy-saving, In: *Control applications, IEEE 1999 International conference on*, Vol. 2, pp. 1579-1584

Yao, Y.; Lian, Z.; Hou, Z. & Zhou, X. (2004). Optimal operation of a large cooling system based on an empirical model. *Applied Thermal Engineering*, Vol. 24, No. 16, November 2004, pp. 2303-2321

Reinforcement Learning

Edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer

Brains rule the world, and brain-like computation is increasingly used in computers and electronic devices. Brain-like computation is about processing and interpreting data or directly putting forward and performing actions. Learning is a very important aspect. This book is on reinforcement learning which involves performing actions to achieve a goal. The first 11 chapters of this book describe and extend the scope of reinforcement learning. The remaining 11 chapters show that there is already wide usage in numerous fields. Reinforcement learning can tackle control tasks that are too complex for traditional, hand-designed, non-learning controllers. As learning computers can deal with technical complexities, the tasks of human operators remain to specify goals on increasingly higher levels. This book shows that reinforcement learning is a very dynamic area in terms of theory and applications and it shall stimulate and encourage new research in this field.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Konstantinos Dalamagkidis and Dionysia Kolokotsa (2008). Reinforcement Learning for Building Environmental Control, Reinforcement Learning, Cornelius Weber, Mark Elshaw and Norbert Michael Mayer (Ed.), ISBN: 978-3-902613-14-1, InTech, Available from:
http://www.intechopen.com/books/reinforcement_learning/reinforcement_learning_for_building_environmental_control

INTECH
open science | open minds