Data-driven Methods for Smart Building AHU Subsystem Modelling

Grigore Stamatescu, Iulia Stamatescu, Nicoleta Arghira, Cristian Drăgana, Ioana Făgărașan
Department of Automatic Control and Industrial Informatics
University "Politehnica" of Bucharest, Romania
e-mail: grigore.stamatescu@upb.ro

Abstract – Modern, densely instrumented, smart buildings generate large amounts of raw data. This poses significant challenges from both the data management perspective as well as leveraging the associated information for enabling advanced energy management, fault detection and control strategies. Networks of intelligent sensors, controllers and actuators currently allow fine grained monitoring of the building state but shift the challenge to exploiting these large quantities of data in an efficient manner. We discuss methods for black-box modelling of input-output data stemming from buildings. Using exploratory analysis it is argued that data mining inspired approaches allow for fast and effective assessment of building state and associated predictions. These are illustrated using a case study on real data collected from commercial-grade air handling units of a research building. Conclusions point out to the feasibility of this approach as well as potential for data mining techniques in smart building control applications.

I. INTRODUCTION

An often quoted figure assigns 40% of the primary energy consumption in recent years to the built environment. Growing tendencies for urbanisation stand only to increase this value over the coming years with significant economic, environmental and social impact. By means of intelligent devices and algorithms the complexity of large scale systems belonging to the built environment can be efficiently managed and provide the tools to the responsible persons for operational improvements.

In the particular case of smart buildings, these are defined by being equipped with sensors, controllers and actuators which are able of advanced functionality for data pre-processing and communication over dedicated protocols. The buildings can be either newly constructed with state-of-the-art technology relating to the Building Automation System (BAS), Building Management System (BMS) or Building Energy Management System (BEMS) but also modernised with mostly wireless sensor and actuator networks (WSAN) thereby reducing significantly the installation and refurbishment costs. We associate the smart terminology in practice with technological and research advances which enable real-time optimal control of the system directly impacting the energy efficiency and occupant comfort for residential, commercial and industrial constructions alike.

Once the required exists or it is put in place, advanced control strategies in the form of model predictive control (MPC) can be implemented. This technique initially appeared in the chemical sector but has since been applied to other control areas as well. Based on the predicted evolution on the plant and associated disturbance and optimal sequence of control inputs is computed over a prediction horizon. At the next sampling step only the first control input is applied to the plant and the computation is repeated. The challenge in the case of MPC is mainly related to accurate but tractable mathematical plant/system models to allow reliable predictions over the receding horizon and managing the overall complexity of the optimisation problem.

Finally, big data, statistical learning and artificial intelligence can be implemented in smart buildings and represents a key area of research in order to build decision support and intelligent systems which assist the building operator to manage it in an energy efficient and comfortable manner.

The main contributions of the paper are summarised next:

- we argue the necessity of deriving black-box models of smart building subsystem for handling the data deluge in such highly complex systems;
- we review data mining inspired techniques applicable to HVAC components;
- we present a case study of exploratory data analysis and modelling for air handling units in a modern research building.

Further on, the paper is organised as follows. Section II discusses related work in the field of smart building monitoring and control with reference to relevant system frameworks and strategies for data management. Section III lists several methods for modelling of systems and subsystems at the building level. Section IV discusses a case study on a modern research building from our campus. We analyse the half-year data from two air handling units (AHU), parts of the Heating Ventilation and Air Conditioning (HVAC) and carry out modelling based on the discussed methods. Section V concludes the paper with outlook on future work.
II. RELATED WORK

In order to reliably collect and store data from across the smart building subsystems, a dedicated hardware and software system architecture has to be put in place. A framework for occupant feedback and environmental learning is presented in [1]. The authors design an end-to-end system to collect data from users and relay it to a central server. At the server level the integration of this data with the building control system is enabled by calculating optimal temperature set-point at the zone level. The system architecture required to enable data collection and control over pervasive networked devices is also presented in [2]. Emphasis is put on heterogeneous communication protocols and distributed computing. This includes generic communication interfaces like wired Ethernet and wireless but also the specific protocols implemented by local and centralized controllers: LonWorks, ModBus, BacNET etc. The security and vulnerabilities associated to Building Automation Systems (BAS) are reviewed in [3], when probabilistic models have been applied to assess the availability of instrumentation and control systems as part of the smart BAS.

A large body of work handles MPC application in buildings with a concise review of the benefits and factors hindering deployment done by [4]. The competing objective of energy savings and comfort which are traded-off within the optimisation problem formulation are described in [5]. The authors propose a lexicographic approach to determine the comfort deviation band around the temperature comfort zone. In [6] the various modelling techniques used for describing the thermal dynamics and associated disturbances to the building thermal zone network are surveyed. The discussion of MPC is highly relevant in the context of this paper as it represents the final goal of the derived models.

As model derivation effort, complexity and accuracy is of key issue to MPC deployment in practical applications, several papers discuss the application of data mining (DM) techniques for achieving input-output black-box models. These are able to trade-off accuracy with a vastly reduced modelling time, given the availability of large quantities of collected data over operationally relevant periods which should cover: seasonal variations for heating and cooling modi and occupancy and building usage patterns. In [7] an unsupervised learning approach is presented that uses the k-means clustering algorithm to infer system states of chillers. Results of in depth statistical analysis of chilled water profiles for air handling units are similarly introduced by [8], who emphasise the potential for accurate predictions of the daily load profiles. Fault detection and diagnosis (FDD) applied to AHU units is presented in [9] using measured deviation from previously determined plant models. A more advanced method is proposed by [10] through ARX models and SVM classification techniques. Precision, recall and F-measure performance of the method is significantly improved over other FDD alternatives applied to the raw data. [11] introduce a rule-based system for energy-efficiency anomaly detection in HVAC system components. The authors present both the logic behind the rule derivation mechanism along a suitable application software implementation for online monitoring.

This paper also builds upon previous work where consensus algorithms where discussed as means of achieving local agreement upon temperature set-points over the network of thermal zones in a smart building [12], as well as carrying out MPC simulation scenarios for grey-box building models [13].

III. THEORETICAL BACKGROUND

This section represents a brief theoretical overview of the MPC control and DM strategies - for data preprocessing and prediction, applied to the particular case study on AHU monitoring.

A. Model Predictive Control

The basic formulation for MPC can be described as follows:

\[
\begin{align*}
\min_{x} & \quad \sum_{j=0}^{N-1} Q^2_{i,j} + \lambda \sum_{j=1}^{N} (T_{i,j} - T_{ref})^2 \\
\text{s.t.} & \quad x_{i,j} = A x_{i,j-1} + B u_{i,j-1} + B_d d_{i,j-1} \\
& \quad Q_{i} \leq Q_{i,j} \leq Q_{in} \\
& \quad T_{in} \leq T_{i,j} \leq T_{in} \\
& \quad j = 1, ..., N
\end{align*}
\]

where the net heat flow \( Q_{i} \) to the thermal zones is determined while keeping the zone temperature \( T_{i} \) close to a given reference \( T_{ref} \) as follows with an underlying state space model of the thermal zone network. Several constraints can be added relating to energy cost and mix, physical and mechanical HVAC system restrictions, occupancy and scheduling. Support of such control and optimisation problem is the main end goal of the black-box modelling tasks.

B. Building Data Mining

One first step in preliminary and exploratory data analysis consists of analysing the correlation in order to quantify the redundancy in the data set and potential for new information. Pearson’s correlation coefficient \( r \in [-1, 1] \) is computed as follows:

\[
r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}
\]

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]
Figure 1: Collected data from AHU 1 over the winter and summer seasons

with $|r| > 0.7$ indicating strong correlation among variables (or predictors), $0.4 < |r| < 0.7$ general correlation and $|r| < 0.4$ weak correlation. $x_i$ and $y_i$ represent the observations $i$ of variables $x$ and $y$ respectively, with means $\bar{x}$ and $\bar{y}$. Principal Component Analysis (PCA) is a popular technique which allows ranking of newly created components based on the original variables by contribution to the variance explained of the input data set. In our case we analyse the correlations between the available measured variables.

Also for DM purposes the variables have to be normalised which assumes an initial pre-processing step such that:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma}$$

the resulting data having $\mu = 0$ and $\sigma = 1$ such that the ranges and absolute values of the original variables do not have a disproportionate influence over classifier decisions.

With regard to black-box predictive modelling, decision trees are a powerful and intuitive technique for predictive modelling, in both classification i.e. assigning observations to discrete outcome classes, as well as regression, outputting numerical values. Several methods are described in the literature to enhance their predictive accuracy by bagging and boosting, random forests. Decision trees are usually applied as baseline for more complex prediction methods. Support Vector Machines (SVM) [14] offer a more powerful alternative which allows the partitioning of the observation space using complex functions for discrimination among classes, albeit with increased computational load.

IV. HVAC SUBSYSTEM MODELLING - CASE STUDY

We collect data from a newly built 7-story facility for research from our campus. It leverages a BMS solution based on Honeywell Centraline technology that provides access to logged tags i.e. data points. A reference BMS screen for chiller monitoring and operation is shown in Figure 2. Our analysis is based on AHU-level data collected in bulk over half a year, from January to July 2017. The AHU main functionality is to assure proper air quality in the building through ventilation while conserving energy. The ratio of outside and recirculated air inserted into the building is set manually or controlled automatically. The AHU can also cool or heat the input air according to demand, for limited thermal load requests. The current challenge of the work is analysing the limited set of available data, without access to some historical parameters such as fan speed and temperature set points as control inputs that we aim to infer from the analysis.

Figure 1 presents the raw data for AHU1 over two periods: two months of winter, from January 15th to March 15th, and two months of summer, from May 15th to July 15th. The plot shows the exhaust and inserted air temperature by the AHU, relative to the external air temperature. The output air temperature reflects the average thermal zone temperature inside the building while the input air temperature is dependant on the recirculated and outside air as well as the control input to the heating or cooling coils of the AHU. In our case heating energy is provided through a district heating network while cooling energy is produced locally through electric chillers. Costs are not currently taken into consideration but should be accounted for given the different rate tiers.
Subsequently we analyse the distribution of the exhaust-input air temperature delta. This can be used to determine energy efficient operation of the AHU in heating or cooling modes by controlling the recirculation ratio. Normal probability distribution functions are mapped to both histograms. This has proven useful to validate normal operation and detect outliers resulting from misconfiguration or faults in the subsystem using typical statistical tests implemented online. The building operator establishes the respective balance.

![Figure 3: Winter and Summer comparison](image)

The variable of interest for the subsequent assessment is the input air temperature for AHU1 as this determines partly the indoor conditions experienced by the building occupants. Table 1 lists the correlation coefficient among this variable and the other measured variables. We include in the analysis also what we define as shoulder season, the period between March 15th and May 15th where it is assumed that only limited amount of energy is needed to condition the building given comfortable indoor temperature and outdoor conditions.

![Table I: Correlation analysis for input temperature](image)

<table>
<thead>
<tr>
<th>r</th>
<th>Winter</th>
<th>Summer</th>
<th>Shoulder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhaust temperature</td>
<td>0.5770</td>
<td>0.574</td>
<td>0.7526</td>
</tr>
<tr>
<td>Recirculation temperature</td>
<td>-0.0385</td>
<td>0.2148</td>
<td>0.7009</td>
</tr>
<tr>
<td>External temperature</td>
<td>-0.1070</td>
<td>0.1198</td>
<td>0.5475</td>
</tr>
<tr>
<td>External humidity</td>
<td>-0.1778</td>
<td>-0.0172</td>
<td>-0.1691</td>
</tr>
</tbody>
</table>

It can be seen how there is some positive correlation between the input air temperature and the exhaust air temperature for the winter and summer seasons respectively. The input temperature is uncorrelated with the other parameters given the effect of the local control loops, manual operator settings and thermal loads. For the shoulder season we observe increased correlation values as the system is balanced.

Finally, Figure 4 illustrates the comparative behaviour in summer conditions for the two air handling units: AHU1 and AHU2 that are placed on top of the building and are responsible for supplying conditioned air to the top half floors. Primary observation is the higher variability of operating modi of AHU1 in comparison to AHU2. This can result from and be used to infer different occupancy and usage level in the associated thermal zones.

![Figure 4: AHU1 and AHU2 summer comparison](image)

V. CONCLUSIONS

We discussed the development of data-driven methods for thermal simulation and control of HVAC systems in residential and office buildings. The main novelty stems from the black-box modelling useful for MPC control of the network of thermal zones, their interactions as well as the perturbations generated by the external environment and variable internal loads. The final aim will results in combined cost-comfort optimisation. It is foreseen the development of a small scale demonstrative system, highlighting the benefits of the new methods in a laboratory evaluation setting. Implementation is based on a distributed networked of open embedded platforms with building specific communication protocols and simulation run-time support capable of on-line querying the BMS infrastructure.

The approaches discussed in this paper are promising in incorporating expert built environment knowledge into predictive models that can yield fast assessments for intelligent system control of indoor conditions both both energy efficiency and comfort.

ACKNOWLEDGMENT

This work has been partially funded by University "Politehnica" of Bucharest, through the "Excellence Research Grants" Program, UPB-GEX. Identifiers: UPB-EXCELENTA-2016 "New Dynamic Consensus Algorithms for Building Thermal Energy Management (CONSENSYS), 20/26.09.2016.

We would like to thank all the referees for their constructive comments and helpful suggestions.
REFERENCES


