TOWARDS RELIABLE STOCHASTIC DATA-DRIVEN MODELS APPLIED TO THE ENERGY SAVING IN BUILDINGS

A.Ridi[1,2], N.Zarkadis[3], G.Bovet[1,4], N.Morel[3], J.Hennebert[1,2]

[1] *University of Applied Sciences Western Switzerland (HES-SO), Fribourg, Switzerland*
[2] *University of Fribourg, Department of Informatics, Fribourg, Switzerland*
[3] *Laboratoire d'Energie Solaire et de Physique du Bâtiment (LESO-PB), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
[4] *Telecom ParisTech ENST, Paris, France*

ABSTRACT

We aim at the elaboration of Information Systems able to optimize energy consumption in buildings while preserving human comfort. Our focus is in the use of state-based stochastic modeling applied to temporal signals acquired from heterogeneous sources such as distributed sensors, weather web services, calendar information and user triggered events. Our general scientific objectives are: (1) global instead of local optimization of building automation sub-systems (heating, ventilation, cooling, solar shadings, electric lightings), (2) generalization to unseen building configuration or usage through self-learning data-driven algorithms and (3) inclusion of stochastic state-based modeling to better cope with seasonal and building activity patterns. We leverage on state-based models such as Hidden Markov Models (HMMs) to be able to capture the spatial (states) and temporal (sequence of states) characteristics of the signals. We envision several application layers as per the intrinsic nature of the signals to be modeled. We also envision room-level systems able to leverage on a set of distributed sensors (temperature, presence, electricity consumption, etc.). A typical example of room-level system is to infer room occupancy information or activities done in the rooms as a function of time. Finally, building-level systems can be composed to infer global usage and to propose optimization strategies for the building as a whole. In our approach, each layer may be fed by the output of the previous layers.

More specifically in this paper, we report on the design, conception and validation of several machine learning applications. We present three different applications of state-based modeling. In the first case we report on the identification of consumer appliances through an analysis of their electric loads. In the second case we perform the activity recognition task, representing human activities through state-based models. The third case concerns the season prediction using building data, building characteristic parameters and meteorological data.

Keywords: State-based modeling, Gaussian Mixture Models (GMMs), Hidden Markov Models (HMMs)

1. INTRODUCTION

In developed countries, the buildings energy demand represents one of the major source of consumption. As a matter of fact, the energy consumption of buildings in EU and USA is comprised between 20% and 40% of the total energy consumption, above industry and transportation. Systems like heating, ventilation, air conditioning (HVAC) represent the major source of consumption in buildings, reaching about an half of the consumption. Also electric lighting and consumer appliances represent an important part. In offices for example, HVAC, lighting and appliances, reach together about 85% of the total energy consumption [1]. Interestingly, HVAC systems represent one forth of total energy consumption in developed countries and an increase of these values is anticipated. This is mainly due to growth in population, increasing demand for building services and comfort levels.

Better controlling and automation procedures are therefore needed to optimize the energy consumption in buildings. A typical approach is to use mathematical models that are derived from a priori knowledge about the physics of the building. The approach we follow in our research is to use data driven approaches that involve mathematical equations not derived from physical processes but learnt from the analysis of time series data. Such approaches present several advantages. First, data-driven techniques and self-learning algorithms do not require a comprehension of the underlying physical process. Second, new or unseen building configuration can be handled as soon as observation data are available, making the installation of such systems easier. Third, the models will be able to handle not only the physics of the building but also the patterns of user behaviours.

In most applications the specific building physics are not directly examined. However when they are, they bring some added value, as in the case of the season variable described below in Section 2. Data-driven techniques typically need a large amount of data, because they use statistical properties of a data time series for characterizing the behaviour of a specific system. Data-driven techniques have been the subject of recent research and many projects can be found in the literature [2].

In our approach, we focus on a layered architecture of systems based on state-based models able to capture the temporal and spatial nature of the signals that correspond to modes of use of the building, house or equipment. More specifically, we propose to model these signals with state-based stochastic approaches such as Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs). Such models allow detecting automatically probable states linked to modes of use that will, in turn, be used as input for the parameterization of smart control algorithms. The Figure 1 illustrates our approach.

More specifically in this paper, we report on the design, conception and validation of several machine learning applications corresponding to the different levels illustrated in Figure 1. These applications aim at showing the interest of the new data-driven modeling of occupancy-related characteristics of buildings. The concepts can be included in advanced control algorithms and allow at the same time to reduce the energy consumption and to improve the user comfort and the adaptation to user requirements.

The next Section give more details about the different applications that have been developed in this context. Section 3 presents results and discussions.

2. METHOD

In this Section we present three different state-based applications: electricity signature identification, human activity recognition and seasonal modeling.

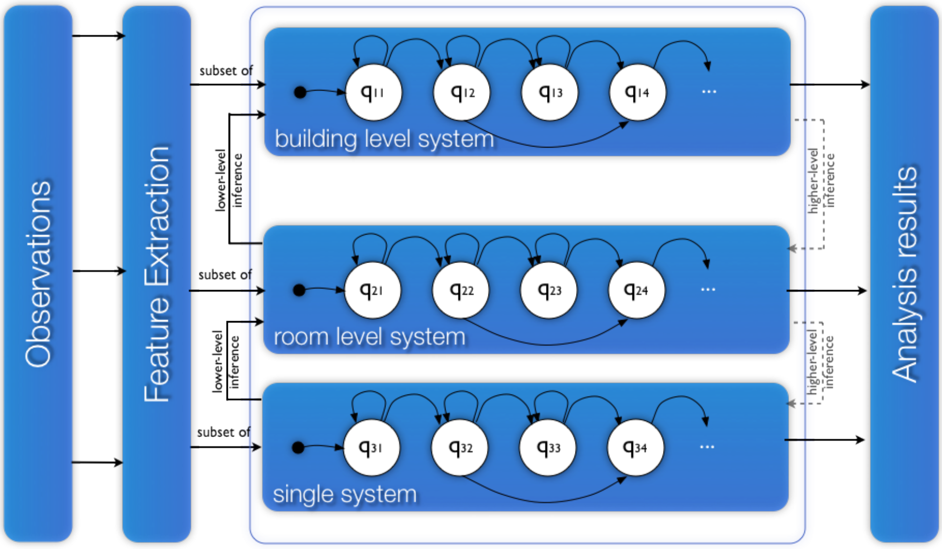

Figure 1: State based modelling.

Electric signature identification

The first application is in the domain of automatic analysis of electricity load in households. The system is based on low-cost smart-plugs measuring every 10 seconds the appliance consumption and producing time series of electric measurements. For the evaluation of the algorithms, we use a database of electric trace *Appliance Consumption Signature Fribourg 1* that includes two acquisition sessions of one hour for 100 appliances spread uniformly into 10 categories [3]. The scientific question here is about the feasibility of modeling, in a data-driven way, the load of single appliance and their automatic identification with generative models such as GMMs and HMMs. As features input of the model, we compute a sequence of vectors using a sliding window procedure on top of the time series. More details about the modelling scheme are provided in [4] and [5]. We applied the two evaluation protocols proposed in [3] to benchmark our models. In the first, called *intersession*, the train set is constituted by instances of the first session, while the testing by those of the second session. Using this paradigm, all the testing signatures come from appliances already seen in the training phase. In the second protocol, called *unseen instance*, the classifiers work with instances coming from appliances not seen before and a 10-cross fold procedure is used to smooth the results. This protocol potentially reveals the system capability of generalizing to new brands or models.

Human activity recognition

In this second application, we show the potential of automatic activity recognition using temporal data captured from presence sensors. In our work, we propose to use HMMs to recognize such human activities. The states in the HMMs are related to the activities and to the expected locations of the activities. The signals we use are asynchronous sensor events from which we sample a sequence of feature vectors spaced in time with a constant interval. Each state includes transition probabilities representing the transitions between states and the corresonding probability to go from one state to the other. Each state also models the probability density function of observing a single feature vector in a given state, i.e. the emission probabilities. Regarding the HMM topologies, we use two types of states, one for the stable parts of the signals, one for the less-stable parts of the signals (see figure 2a). Using this paradigm, we considered as *unstable* the transition areas, e.g. the corridors or the doors. In the (*stable*) areas, e.g. the kitchen for cooking or the bedroom for sleeping, the transition probabilities are computed through the training and their values depend on the nature of the activities. An evaluation benchmark, results and discussions are presented in Section 3.

Season modelling

In a third application, we introduce a new model of season prediction based on state-based models such as HMMs able to detect more finely the change of seasons from meteorological data. When implementing an advanced control scheme for building services, typically heating, cooling, ventilation (HVAC) and solar shadings, for instance using Fuzzy Logic, a variable "season" (typically represented as a fuzzy variable) represents a good method to fit the behaviour of the control system to the weather conditions. A simple definition, for instance based only on calendar or on outside temperature, is not satisfactory, since building characteristics and use should also be considered.

For the purpose of the season definition, we involve the signals from HVAC, window opening, window blinds, external and internal temperature and solar irradiation. We also use a separate module to calculate the building's time constant (a simple 2-node model) which we include as parameter in our observation vector.

We consider that the "season" variable can have the value of the following 3 states (figure 2b):

- Heating season which is defined when heating is required to avoid that the inside temperature is getting lower than the optimal comfort temperature.
- Cooling season when cooling is needed, either with mechanical cooling or by measures such as passive night cooling or protection against passive solar gains.
- Intermediate season (mid-season) which is normally the most critical season. During this season the building might need sometimes heating and sometimes cooling.

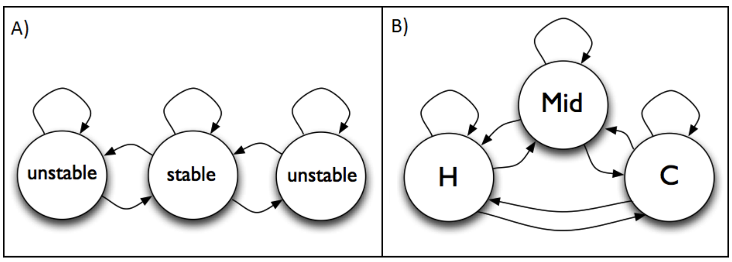

Figure 2: Two different state based modelling: A) Example of model for activity recognition with three states. Such topology is used for modelling Relaxing, Working, Sleeping, Out of Home, Going from Bed to Toilet. B) The three possible states of the ergodic variable "season" and the possible transitions from one state to the next: Heating season (H), cooling season (C) and intermediate season (Mid).

3. RESULTS AND DISCUSSION

Electricity signature identification

In order to evaluate the state-based model performances, two classifiers have been compared, namely k-Nearest Neighbor (k-NN) and GMM systems. Using the *intersession* protocol, we achieved respectively 88% and 93.8% correct category identification [5], while using the *unseen instance* protocol we achieved respectively 57% and 74% correct category identification [6]. Clearly the first protocol achieved better results than the second, given that in the latter case instances coming by appliances never seen before have to be classified. For evaluating the state-based model we used GMM, but we envisage to use the HMMs, which should be particularly suitable for electrical signatures of appliances, which in most cases are state-based machines.

Human activity recognition

We evaluated the proposed models using the WSU CASAS dataset [7]. The raw data is recorded from 31 motion sensors, 4 temperature sensors and 4 door sensors, from which we use the motion and door sensors. In the observation vector we added the time information computed through a 24h cyclic function in order to make it more suitable for the analysis. Among the activities labeled in this dataset, we selected those having an adequate number of data sequences, in order to have a sufficiently large balanced dataset: *Meal Preparation, Relaxing, Eating, Working, Sleeping, Going from Bed to Toilet, Out of Home*. The total time of data recording was 8 months for one person living dayly in a smart home and receiving visits on a regular basis. After several tuning, we measured an overall accuracy using a leave-one out procedure of 98.9% for the HMM presented in Section 2.

Season model

We developed and evaluated the proposed season prediction model using data recorded over a long period of time at the LESO-PB experimental building which is located in EPFL campus near Ecublens and Lausanne. The observation vector consisted of the signals of 8 actuators and sensors from a single-occupant office room including ambient and indoor temperatures, window states (open/closed), external blinds position and solar irradiation values. Hidden states (heating, cooling and intermediate season) are considered ergodic (any transition between various states is possible; see Figure 2b). We verify the developed model against the heating power: we consider the use of heating as an indication for the heating season. Results demonstrated that season was correctly identified with an accuracy of 69% to 91%. Higher accuracy is observed when identifying heating or cooling seasons, while it drops significantly for the intermediate season. This is expected behaviour as intermediate season is not a crisp state and it is far more hard to define and thus to correctly identify. Also, the absence of active cooling in LESO building adds a further difficulty in the validation of our model (it would have provided us with an indication of the cooling season).

4. CONCLUSION AND OUTLOOK

In this paper we showed different applications of state-base modeling in a smart building context. The underlying motivation is to build more precise models to better model and control the energy usage in buildings. In the first application we showed the possibility to automatically recognize appliance from their electric signatures. When the appliances are known to the database used to build the models, the identification accuracy can be up to 93.8%. The second application is about the use of HMM for the recognition of human activities in smart environments. Quite high accuracy rates have been measured, up to 98.9%. This performance is however to take cautiously as the procedure uses a large number of presence sensors and as the benchmark involves only one person. The third application is about season modelling where the developed HMM model predicted correctly the season with an accuracy of 69% to 91%. As expected, cooling and heating season prediction is more accurate than that of the intermediate season.

As a further step to our research, we aim to demonstrate in practice the interest of the new data-driven modelling of occupancy-related characteristics of buildings. We plan to carry out simulations as well as real life experiments where these concepts will be included in advanced control algorithms and allow at the same time to reduce the energy consumption and to improve the user comfort and the adaptation to user requirements.

ACKNOWLEDGEMENTS

This work was supported by the research grant of the Hasler Foundation project GreenMod.

REFERENCES

1. Pérez-Lombard, L., Ortiz, J., and Pout, C.: A review on buildings energy consumption information. *Energy and Building*, 40:394–398, 2008.
2. Chan, M., Estève, D., Escriba, C., and Campo, E.: A review of smart homes: Present state and future challenges. *Computer Methods and Programs in Biomedicine*, 91:55–81, 2008.
3. Gisler, C., Ridi, A., Zufferey, D., Khaled, O. A., and Hennebert, J.: Appliance consumption signature database and recognition test protocols. In *Proceedings of the 8th Internation Workshop on Systems, Signal Processing and their Applications (Wosspa '13), to appear*, 2013.
4. Zufferey, D., Gisler, C., Khaled, O. A., and Hennebert, J.: Machine learning approaches for electric appliance classification.
5. Ridi, A., Gisler, C., and Hennebert, J.: Automatic identification of electrical appliances using smart plugs. In *Proceedings of the 8th Internation Workshop on Systems, Signal Processing and their Applications (Wosspa '13), to appear*, 2013.
6. Ridi, A., Gisler, C., and Hennebert, J.: Unseen appliances identification. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13), submitted*, 2013.
7. Cook, D.: Learning setting-generalized activity models for smart spaces. *IEEE Intelligent Systems*, 27:32–38, Jan 2012.