Accurate Household Occupant Behavior Modeling Based on Data Mining Techniques

Márcia Baptista, Anjie Fang, Helmut Prendinger
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
marcia.baptista@ist.utl.pt, af12331@my.bristol.ac.uk,helmut@nii.ac.jp

Rui Prada
INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
Av. Prof. Cavaco Silva, Taguspark Porto Salvo, Portugal
rui.prada@ist.utl.pt

Yohei Yamaguchi
Sustainable Energy and Environmental Engineering
Graduate School of Engineering, Osaka University
2-1 Yamada-oka, Suita, Osaka 565-0871, Japan
yohei@see.eng.osaka-u.ac.jp

Abstract

An important requirement of household energy simulation models is their accuracy in estimating energy demand and its fluctuations. Occupant behavior has a major impact upon energy demand. However, Markov chains, the traditional approach to model occupant behavior, (1) has limitations in accurately capturing the coordinated behavior of occupants and (2) is prone to over-fitting. To address these issues, we propose a novel approach that relies on a combination of data mining techniques. The core idea of our model is to determine the behavior of occupants based on nearest neighbor comparison over a database of sample data. Importantly, the model takes into account features related to the coordination of occupants' activities. We use a customized distance function suited for mixed categorical and numerical data. Further, association rule learning allows us to capture the coordination between occupants. Using real data from four households in Japan we are able to show that our model outperforms the traditional Markov chain model with respect to occupant coordination and generalization of behavior patterns.

Introduction

Human behavior is a major determinant of energy consumption in buildings, particularly in the residential sector (Lopes, Antunes, and Martins 2012; Karatasou, Laskari, and Santamouris 2013). The way household members perform their daily routines, share rooms or use electrical appliances, can greatly impact total energy use. Yu et al. (2011) found that occupant behavior was responsible for variations on the order of up to four times the mean annual energy consumption of buildings, based on the investigation of six residential districts in Japan that were clustered according to other influencing factors such as city climate and building characteristics.

An increasing number of energy demand models (Kavgic et al. 2010) simulate the influence of occupant behavior upon energy consumption. In these models, households are simulated as multi-agent systems, where occupants are represented as social agents that behave according to the rules prescribed by Markov chains. Here, Markov chain probability matrices are calculated from time-use (TU) data, i.e., sequential high-resolution information of the real-life daily routines of household occupants, such as sleeping or having meals.

Traditional Markov chains show two important limitations as a technique to accurately encode occupant behavior. First, they do not accurately capture the coordinated behavior of household members, such as joint activities. By *joint activity* we mean joint-in-purpose (having meal, having bath, etc.) rather than joint-in-time or joint-in-location, following the distinction of Gliebe and Koppelman (2002). Second, these models are prone to over fitting. Since they tend to accurately replicate the given, relatively sparse TU data, they fail to generalize from the input trends.

To address these shortcomings, we propose an alternative approach to the Markov chain approach, which is based on a combination of data mining techniques. The core idea of our approach is that we predict agents' behavior by searching the TU database for past household states that resemble the current one. To select a past similar state we use nearest neighbor comparison based on a normal distribution. This mechanism aims to generate a model that can generalize from the input TU data. In addition, the nearest neighbor comparison allows us to more accurately represent coordinated agent behavior.

The paper is structured as follows. The next section reviews major work in the fields of occupant behavior modeling and multi-agent systems that estimate energy demand. Then we detail our proposed model. Next, we compare the performance of traditional Markov chains to our model using real data of household occupant behavior. Finally, we discuss and conclude the paper.

Related Work

Energy demand models can be classified in two broad categories: 'top-down' and 'bottom-up' models (Swan and Ugursal 2009; Grandjean, Adnot, and Binet 2012). The two approaches differ in the concepts and principles used for the modeling of demand. The top-down approach follows standard macroeconomic modeling techniques, whereas the bottom-up approach is based on the concept of disintegration. Top-down models describe behavioral relations at an aggregated level. By contrast, bottom-up models calculate energy demand at the level of the individual household, occupant or electrical appliance.

Several bottom-up models have recently been proposed in the energy field (Dounis and Caraiscos 2009; Widén and Wäckelgård 2010; Shimoda et al. 2010), which are based on the concepts of 'agent' and multi-agent system (as defined in Jennings, Sycara, and Wooldridge 1998). In those models, the occupant is modeled as a situated agent that can receive sensory input from the household environment, such as information on electrical appliance usage. The agent can perform actions in that environment, such as moving between rooms or using electrical appliances. Each agent decides its actions autonomously, in each simulation step, showing responsive and pro-active behavior.

Bottom-up models based on multi-agent systems are used for two distinct purposes: (i) real-time energy demand prediction and (ii) simulation. In the first case, models generate real-time predictions of occupant behavior. For instance, Mamidi, Chang, and Maheswaran (2012) proposed an approach to improve the energy efficiency of commercial buildings by controlling the HVAC (heating, ventilation, and air conditioning) systems according to an occupancy prediction model that estimates the presence or absence of persons in a room.

In the second case, models aim at the simulation of energy demand by generating probability densities of occupants' behavior. These models target the simulation the energy consumption of residential areas for extended periods of time. They can be used to investigate how new energy policies or smart grid technological advancements might impact energy consumption of an urban area. For instance, Shimoda et al. (2010) used an multi-agent approach to simulate the energy consumption and CO_2 emissions of the residential sector of Japan by 2025. Several CO_2 mitigation measures were evaluated using the simulation model as a test-bed.

The techniques used to model occupant behavior differ greatly depending on the purpose of the multi-agent system. Real-time prediction relies on sensor data and employs machine learning methods (Mamidi, Chang, and Maheswaran 2012) or genetic algorithms (Yu 2010) to predict e.g. room occupancy. In contrast, multi-agent systems rely on TU data and typically use Markov chains to simulate the behavior of occupants (Yamaguchi, Tanaka, and Shimoda 2012).

To date, most multi-agent systems aiming at energy demand simulation have modeled households as a set of *non-interacting* individuals (Munkhammar and Widén 2012; Baptista et al. 2014). This means that current systems are based on the assumption that the probabilities of the activities of a household member are not dependent on the activities of other occupants. Accordingly, these systems often fail to capture the coordination between occupants, such as the shared use of lighting and other electrical appliances. This in turn translates into inaccuracies in energy demand estimation.

Current energy demand simulations based on Markov chains present additional limitations, such as the inaccurate modeling of the duration and temporal order of activities. Another constraint of current energy demand simulations relates to over-fitting of simulated activities to the original TU data. This results in Markov chain based models being often unable to generalize beyond TU data to new data. Some authors (Tanimoto, Hagishima, and Sagara 2008; Richardson et al. 2010; Yamaguchi, Tanaka, and Shimoda 2012) have attempted to address these points using alternative approaches to Markov chains, with some success.

Nearest Neighbor (NN) Model

In this section, we introduce our approach to modeling occupant behavior in multi-agent systems. Our approach is based on data mining techniques that allow us to model the behavior of occupants in a household as a set of coordinating agents that learn from historical real-world data to decide their daily routines.

Our model assumes the existence of n agents that represent the occupants of a household. It also assumes the existence of m simulation time-steps. In each time-step, each agent in the household must decide its next behavior from a set of possible alternatives such as sleeping or having bath. The agents decide their behaviors taking into account the behaviors of others, as in real life.

The reasoning of an agent for a time-step t involves three stages. First, in the **initialization stage**, the agent collects all the necessary data about the current state of the household (x_t). Second, in the **prediction stage**, the current state is matched against a structured database of real-world household states representing time-use data (TU data). The database is classified according to the resemblance of states to the current state, using a customized measure of similarity. A normal distribution ($\mathcal{N}(0, \sigma^2)$) determines the maximum allowed similarity distance i between the current state and the states in the database. This forms the set of *predictive* states ($\{w_0, \ldots, w_k\}$). By predictive states we mean the set of states in the database that, due to their similarity to the current state, are likely to predict what occupants will do next. We assume that if two household states are similar then there is a high probability that the next activity of one state is also the next activity for the other state. We assume that all predictive states have the same probability to accurately determine the next behavior of the agent. Hence, we use a random function to select the preferred predictive state (w) from the set of predictive states.

We introduce the notion of 'maximum allowed similarity distance' (i) to allow the agent to perform less restricted matches of the sample database while predicting its next behavior. In turn, the normal distribution mechanism ensures some degree of restriction in the selection of the set of predictive states. The effectiveness of the selection of this set greatly determines how well our model can fit the sample data and generalize from it. With this approach we aim to achieve a balance between over-fitting and under-fitting.

We assume that all predictive states have an equal probability of being selected and determine the agent's next behavior accordingly. Note that the normal distribution mechanism already implements an approach that favors states with similarity distances close to zero.

The third stage of the reasoning process consists in the **execution stage** of the selected behavior. In the execution the agent mimics the occupant behavior predicted by the state chosen in the prediction stage. The agent's reasoning cycle continues until the simulation ends.

Initialization

At each time-step t the agent senses the household environment according to an array of characteristics (features):

1. Activity, room and appliances used by agent at previous time-step ($t-1$);
2. Length in time-steps (duration) of the on-going activity (l);
3. Activities, rooms and appliances used by one other occupant at current time-step (t). This other (leader) occupant is selected as the agent whose behavior is more strongly related to the current behavior of the agent. Relations between the behaviors of agents (two agents mainly) are generated through association rules learning (Agrawal, Imieliński, and Swami 1993) in the original TU data with the Apriori algorithm (Agrawal et al. 1996). The Apriori algorithm is applied to a collection of activities performed by pairs of occupants in each time-step. The algorithm runs on this collection to find the most frequent relations. Strong relations are required to satisfy a minimum confidence value and are prioritized according to their support value. Leader agents are identified with basis on this strong relations according to the actual context of the simulation.

Finally, the feature vector representing the household current state (x_t) is passed to a prediction system.

Prediction

Having the feature vector as input (x_t), the agent must decide its next behavior based on a database of TU data using a fixed-radius near neighbors (FRNN) comparison (Bentley 1975). Given a feature vector and a distance i defined by a normal distribution $\mathcal{N}(0,\, \sigma^2)$, the FRNN search reports the set of predictive states $\{w_0, \ldots, w_k\}$ from the database that are within distance i of the feature vector. The search is optimized with a kd-tree approach. Prediction is complete when a random state w is chosen from the set of states returned by the FRNN search.

Since the set of features that characterize a household state include categorical attributes (such as activities, rooms, appliances) and numerical attributes (such as length of ongoing activity), we use the method of Ahmad and Dey (2007) to compute the distance between two data elements.

Our distance function uses a weighted squared value of the Euclidean distance. All numeric attributes are normalized to be on the same scale. This function is specially suited to compute similarity between categorical data objects. For instance, consider the categorical attribute of activity for the following data objects: 'Bathing', 'Showering' and 'Studying'. Our customized distance measure can recognize that the activities 'Bathing' and 'Showering' are more similar than the activity of 'Studying' to either of them.

Our method works well in capturing similarity because it considers the distribution of values in the dataset while computing the distance between two attribute values. For instance, consider the way the algorithm computes the distance ($distance(A, B)$) between two activities A, B. Let N be the number of activities and act represent an activity, $1 \leq act \leq N$. As shown in Algorithm 1, distance between activities is computed as a function of activities distribution in the overall dataset taking into account co-occurrence with other activities.

Algorithm 1 Calculating distance of activities

$distance(A, B) \leftarrow 0$
$act \leftarrow 1$
for $act \leq N, act++$ **do**
  if $P(act|A) \leq P(act|B)$ **then**
    $distance(A, B) \leftarrow distance(A, B) + P(act|B)$
  else
    $distance(A, B) \leftarrow distance(A, B) + P(act|A)$
  end if
end for
$distance(A, B) \leftarrow distance(A, B) - 1$

Execution

After selecting a predictive state w from the TU data, the next behavior of the agent is executed. This may involve the selection of the agent's next activity such as sleeping or having meals and also its location in the household as well as the electrical appliances used. This execution changes the state of the agent and household.

Experiment

The aim of this section is to compare the accuracy of two approaches to the modeling of occupant behavior in multi-agent systems of energy demand estimation: the standard Markov chain (MC) model and our Nearest Neighbor (NN) model.

Markov Chain (MC) Model

To compare the performance of our NN model with the state-of-the-art approach to modeling occupant behavior, we have developed a model based on discrete non-homogeneous Markov chains, following the implementation of Widén, Nilsson, and Wäckelgård (2009). In our model, time-step transition probabilities are estimated from TU data with the following equation:

$$\tau_{ij}(n) = \begin{cases} 0, \text{ if } \sum_{j=1}^{|S|} \eta_{ij}(n) = 0 \\ \frac{\eta_{ij}(n)}{\sum_{j=1}^{|S|} \eta_{ij}(n)}, \text{ otherwise} \end{cases}$$

where $\tau_{ij}(n)[0,1]$ defines the time-step transition probability that the system will evolve from state i to state j (assume the existence of S states) from time n to $n+1$ and $\eta_{ij}(n)$ stands for the number of occurrences of transitions between states i and j from time n to $n+1$.

Time-step transition probabilities define the likelihood of an agent to change state in a given time-step. States in the model correspond to the daily routines of the occupant, such as 'Sleeping', 'Being outside', 'Having meal', and so on.

Indicators and Hypotheses

We assess the accuracy of the MC and NN models according to some important indicators as presented in Table 1. Three of the four indicators are taken from the literature, the first indicator (I1) is our original indicator. This indicator measures if the models accurately reproduce how often an occupant performs an activity as a joint activity, or sole activity. This indicator is important since the simultaneous performance of an activity, such as watching TV or cooking, implies the sharing of electrical appliances. As a result, the accurate estimation of the time that occupants perform an activity simultaneously with others or individually may allow us to more accurately estimate the time when electrical appliances are shared.

Indicator I2 relates to variations in timing of performing some activity. Such variation is important when peak demand of large residential areas is estimated (Tanimoto, Hagishima, and Sagara 2008). In the energy field, this indicator is often used to evaluate the diversity of agents' behaviors (Tanimoto, Hagishima, and Sagara 2008; Yamaguchi, Tanaka, and Shimoda 2012).

Based on the most prominent indicators from the literature (see Table 1), we formulate our hypotheses as follows:

- *Coordinated Behavior Hypothesis*: our NN model is more accurate than the MC model in estimating the time an occupant performs an activity as a joint or sole activity (Indicator I1).
- *Patterns of Behavior Hypothesis*: our NN model can generate a higher number of different patterns of behavior transitions than the MC model (Indicator I2).

The hypotheses have been tested with real TU data. Regarding the Indicators I3–I4, we do not claim that our NN based approach outperforms the MC based approach.

	Indicators
I1	Time performing an activity jointly with others
I2	# of different patterns of occupant behavior transitions
I3	Probability distribution showing percentage of activity at each time step
I4	Number of activity transitions per day

Table 1: Indicator I1 is our original indicator; Indicators I2-I4 are proposed in (Yamaguchi, Tanaka, and Shimoda 2012).

Data

The data used in our experiment was collected from four households in the region of Osaka, Japan, in late 2011 and beginning of 2012. The surveyed household members agreed to document their daily routines in diaries during 14 days, on a minute by minute basis. Behaviors are classified in thirty categories (see Table 2). Table 3 describes the household composition, dwelling type and surveyed period for each household in the data set.

	Activities		
1	Being outside	16	Other housework
2	Sleeping	17	Communication
3	Having meal	18	Hobbies
4	Face washing	19	TV in living room
5	Bathing	20	TV in bedroom
6	Changing clothes	21	Stereo living room
7	No activity living room	22	Reading living room
8	Working at home	23	Reading bedroom
9	Studying	24	Stereo bedroom
10	Cooking	25	Video
11	Cleaning dishes	26	No activity bedroom
12	Cleaning living room	27	Showering
13	Cleaning other rooms	28	Having breakfast
14	Clothes washing	29	Having lunch
15	Ironing	30	Having dinner

Table 2: Available categories of activities in a household.

Methodology

Our NN model was compared with the traditional MC model using the following methodology:

- *Simulation*: Each of the four households is simulated using the MC model and the NN model ($\sigma^2 = 0.002$). In each simulation of a household, the daily activities of the occupants are reproduced for a period of 500 days.
- *Validation*: The accuracy of the models is assessed by comparing their output with the original TU data.

This validation method is the standard approach to validating occupant behavior models in the energy field (Widén, Nilsson, and Wäckelgård 2009; Widén and Wäckelgård 2010; Tanimoto, Hagishima, and Sagara 2008; Yamaguchi, Tanaka, and Shimoda 2012). It is important to note that this approach is different from a method like 10-fold cross-validation, which is used to validate models generated by

	Dwelling	Period	Occupant	Abbr
HA	Detached house 3LDK	30 Oct 12 11 Nov 11	Working male	M(A)
			Working female	F(A)
			High-school	H(A)
			Junior high-school	J(A)
HB	Appartment 4LDK	16 Jan 12 29 Jan 12	Working male	M(B)
			Working female	F(B)
			Primary school	P(B)
			Junior high-school	J(B)
HC	Appartment 4LDK	7 Dec 11 21 Dec 11	Working male	M(C)
			Housewife	F(C)
			Primary school	P(C)
HD	Appartment 4LDK	8 Dec 11 22 Dec 11	Working male	M(D)
			Housewife	F(D)
			Primary school	P(D)

Table 3: Household composition of the TU dataset with 3 or 4 rooms and LDK (Living, Dining, Kitchen).

machine learning techniques (Mamidi, Chang, and Maheswaran 2012). Please note that the models we study output probability densities rather than specific predictions.

Results

We first present results regarding the Coordinated Behavior Hypothesis (Indicator I1) and the Patterns of Behavior Hypothesis (Indicator I2). Then we present results for Indicators I3 and I4.

Coordinated Behavior Hypothesis

Table 4 reports the average error (in minutes) in estimating the time each occupant spent performing an activity jointly with others (I1) for both MC and NN models. It also shows the results of Student t-tests. We note that these estimation errors were calculated only from activities that were performed jointly by the occupants.

Interestingly, we obtain different results for 3- and 4-person households. In 4-person households (A and B), we found that our NN model is better than the MC model, mostly with statistical significance (or near-significance), except for one case. In 3-person households (C and D), the MC model is better in the majority of cases. Hence, our results only partially support the Coordinated Behavior Hypothesis.

With a larger number of household members it is more likely that an occupant performs some activity jointly with another occupant. Since our NN model explicitly models joint activity, it performs better in 4-person households than in 3-person households.

We found differences in the performance of NN model related to the joint duration of activities. Here, joint duration is time of activity performing together with others. In activities with large joint duration such as ‘Sleeping’, ‘Being outside’, ‘TV in living room’ and ‘Hobbies’, the better performance of NN is not as pronounced as for remaining activities. For these four activities, Student t-tests could only reveal a statistically significant difference between NN and MC for 3 cases. The average difference of error was of 1% (35.6 to 36.1min). For remaining activities, difference was the double: 2% (1.2 to 1.19min) and t-tests proved difference between models for six cases. These differences can be explained by NN overcoming the limitations of MC in modeling the coordination of agents in short joint time activities (Yamaguchi, Tanaka, and Shimoda 2012).

It is important to note that the NN model not only replicates but also generalizes from the coordination behavior patterns found in TU data. The MC model tries to predict the future based on samples of TU data that exactly match the current state in the dimensions of time and activity (Baptista et al. 2014). By contrast, our NN model is not limited to perfect matching with TU data, and thus is less prone to over-fitting. The NN model occasionally picks a state from the past TU data that is slightly dissimilar to the current state with regard to some dimension (time or activity). Simply put, the NN model trades accuracy for generality.

Occupant	Estimation Error MC	NN	p
M(A)	7.55	7.2	0.18
F(A)	9.05	8	<0.05
H(A)	9.3	8.3	<0.05
J(A)	7.8	7.35	0.12
M(B)	11.1	11.15	0.43
F(B)	8.35	7.75	0.06
P(B)	10.35	7.15	<0.05
J(B)	10.75	9.3	<0.05
M(C)	6.3	6.5	0.25
F(C)	6	6	0.45
P(C)	5.95	6.4	0.09
M(D)	7	7.7	0.09
F(D)	6.65	8.55	<0.05
P(D)	8.7	8.8	0.37

Table 4: Average daily estimation error (in minutes) of time performing an activity jointly with others (I1) for NN and MC models.

Patterns of Behavior Hypothesis

Fig. 1 compares the number of different patterns of occupant behavior transitions per day (I2) found in TU data, MC and NN model. We calculate this indicator by first counting, for each time-step, the number of distinct transitions registered in all the simulated days. Here, transition means a change from an activity to a different activity. Then, we sum the distinct transitions of each time-step for all time-steps.

As illustrated in Fig. 1, the number of different daily behavior transition patterns registered in all simulations runs is very different in MC and NN. The MC model generated as much distinct patterns as the ones found in TU data. By contrast, the NN model was able to generate up to 394 new patterns averaging 228 new patterns per occupant. Hence, our results strongly support the Patterns of Behavior Hypothesis.

Fig. 2 shows, for each time step, the number of distinct

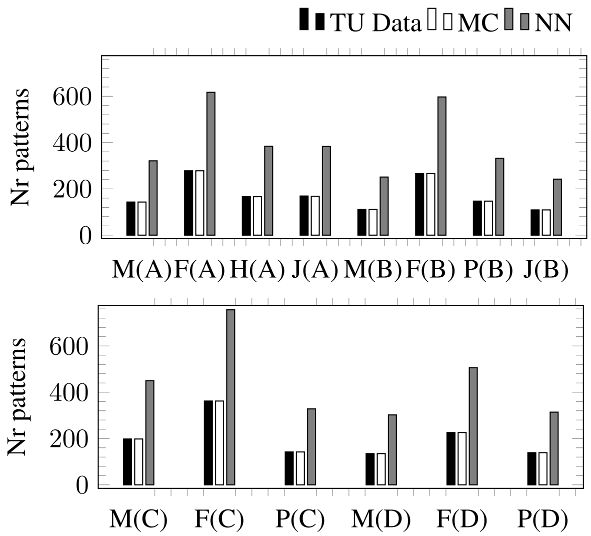

Figure 1: Comparison of number of different patterns of occupant behavior transitions (I2) per day.

transitions (transition patterns) registered during the 500 simulations runs of the MC model (top chart) and NN model (bottom chart) for occupant M(A). The figure illustrates that the NN model generates far more patterns than TU data and MC model. Additionally, the NN model generates transition patterns in time ranges near the ones of the original patterns. In other words, NN agents are able to delay and advance, the transitions found in the TU data.

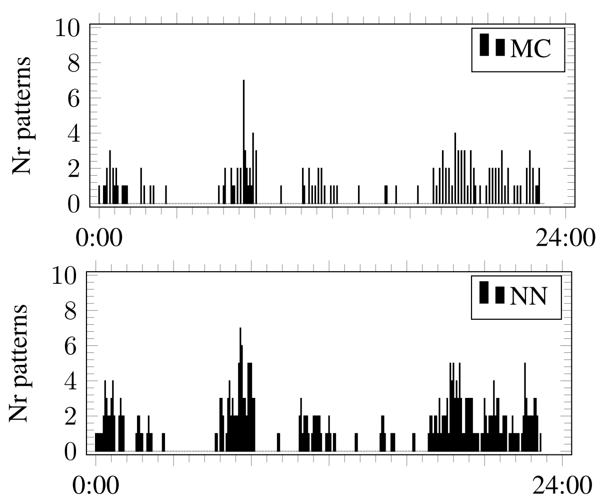

Figure 2: Comparison of average number of different behavior transitions (I2) generated by MC and NN models for occupant M(A).

Other Indicators

Table 5 presents the average, maximum and standard deviation of Indicators I3 and I4.

	MC			NN		
	Avg	Max	STD	Avg	Max	STD
I3	0.001	0.003	0.0005	0.003	0.004	0.0008
I4	3.55	5.17	0.88	3.52	4.87	0.87

Table 5: Estimation errors of indicators I3 and I4.

Student t-tests were used to compare the MC and NN models with respect to indicator I3 and I4. For Indicator I3, Student's t-test did not show a statistically significant difference for 13 of the 14 occupants. In the case of Indicator I4, Student's t-tests could reveal a statistically significant difference ($p < 0.05$) for 8 out of 14 occupants.

The results on Indicators I3 and I4 are important as they show that the NN model is comparable to the MC model in modeling (1) the way agents perform activities at each time-step (I3) and (2) how often agents change their activities in a day (I4). The second finding is noteworthy: even though NN generates more patterns of transitions (I2), this does not influence the accuracy of the number of transitions per day (I3). In other words, NN agents change activities during a day as often as MC agents despite the fact that changes of activities are being delayed or advanced in time. We note that we calibrated NN to delay and advance activities to a maximum of one time-step (5 minutes).

Conclusions

Energy demand simulations play an important role in implementing the smart grid and in reducing CO_2 emissions. Human behavior is at the core of these simulations; the way household occupants behave and interact with each other can greatly determine the accuracy of demand estimation. Our work attempts to present a new perspective on occupant behavior modeling by accommodating three factors: individual behavior, coordination of behavior among occupants, and diversity of behavior patterns.

We present a Nearest Neighbor (NN) occupant behavior model that (1) aims to replicate coordinated behavior more accurately than a Markov Chain (MC) model, the state-of-the-art approach, while at the same time (2) generalizes new patterns of behavior. Our results suggest that our solution is a good compromise between agreement of simulated data to TU data and generalization of patterns.

There are two main research directions for future work. The first line of research will focus on validating the quality of the new generalized patterns of data. As our model outputs density probabilities, we can investigate density estimation cross validation techniques (Marron 1987).

The second direction relates to further exploring coordination of agents. For instance, we may consider more than one leader in the household or coordinate all agents at the same time. The influence of these strategies upon variability of coordinated behavior is an important future topic.

Acknowledgments

This work was partially supported partly by the 'Global Lab' Grand Challenge grant from the National Institute of Informatics (NII), the NII Int'l Internship Program, and Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2013 and scholarship reference SFRH/BD/52347/2013.

References

Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen, H.; Verkamo, A. I.; et al. 1996. Fast discovery of association rules. *Advances in knowledge discovery and data mining* 12:307–328.

Agrawal, R.; Imieliński, T.; and Swami, A. 1993. Mining association rules between sets of items in large databases. *ACM SIGMOD Record* 22(2):207–216.

Ahmad, A., and Dey, L. 2007. A K-mean clustering algorithm for mixed numeric and categorical data. *Data & Knowledge Engineering* 63(2):503–527.

Baptista, M.; Prendinger, H.; Prada, R.; and Yamaguchi, Y. 2014. A cooperative multi-agent system to accurately estimate residential energy demand. In *Proc. of the 13th Conf. Autonomous Agents and Multiagent Systems*. IFAAMS. Short paper.

Bentley, J. L. 1975. Survey of techniques for fixed radius near neighbor searching. Technical report, Stanford Linear Accelerator Center, Calif.(USA).

Dounis, A. I., and Caraiscos, C. 2009. Advanced control systems engineering for energy and comfort management in a building environment A review. *Renewable and Sustainable Energy Reviews* 13(6):1246–1261.

Gliebe, J. P., and Koppelman, F. S. 2002. A model of joint activity participation between household members. *Transportation* 29:49–72.

Grandjean, A.; Adnot, J.; and Binet, G. 2012. A review and an analysis of the residential electric load curve models. *Renewable and Sustainable Energy Reviews* 16(9):6539–6565.

Jennings, N. R.; Sycara, K.; and Wooldridge, M. 1998. A roadmap of agent research and development. *Autonomous agents and multi-agent systems* 1(1):7–38.

Karatasou, S.; Laskari, M.; and Santamouris, M. 2013. Models of behavior change and residential energy use: A review of research directions and findings for behavior-based energy efficiency. *Advances in Building Energy Research* 1–11.

Kavgic, M.; Mavrogianni, A.; Mumovic, D.; Summerfield, A.; Stevanovic, Z.; and Djurovic-Petrovic, M. 2010. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment* 45(7):1683–1697.

Lopes, M.; Antunes, C.; and Martins, N. 2012. Energy behaviours as promoters of energy efficiency: A 21st century review. *Renewable and Sustainable Energy Reviews* 16(6):4095–4104.

Mamidi, S.; Chang, Y.-H.; and Maheswaran, R. 2012. Improving building energy efficiency with a network of sensing, learning and prediction agents. In *The 11th Conf. Autonomous Agents and Multiagent Systems, AAMAS 2012*, 45–52. IFAAMS.

Marron, J. 1987. A comparison of cross-validation techniques in density estimation. *The Annals of Statistics* 152–162.

Munkhammar, J., and Widén, J. 2012. A stochastic model for collective resident activity patterns and energy use: preliminaries. In *Future technology press*, 1–4.

Richardson, I.; Thomson, M.; Infield, D.; and Clifford, C. 2010. Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings* 42(10):1878–1887.

Shimoda, Y.; Yamaguchi, Y.; Okamura, T.; Taniguchi, A.; and Yamaguchi, Y. 2010. Prediction of greenhouse gas reduction potential in Japanese residential sector by residential energy end-use model. *Applied Energy* 87(6):1944–1952.

Swan, L. G., and Ugursal, V. I. 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews* 13(8):1819–1835.

Tanimoto, J.; Hagishima, A.; and Sagara, H. 2008. A methodology for peak energy requirement considering actual variation of occupants behavior schedules. *Building and Environment* 43(4):610–619.

Widén, J., and Wäckelgård, E. 2010. A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy* 87(6):1880–1892.

Widén, J.; Nilsson, A. M.; and Wäckelgård, E. 2009. A combined markov-chain and bottom-up approach to modelling of domestic lighting demand. *Energy and Buildings* 41(10):1001–1012.

Yamaguchi, Y.; Tanaka, M.; and Shimoda, Y. 2012. Comparison of occupant behavior models applied to a household. In *The 1st Asia Conference on International Building (ASIM'12)*.

Yu, Z.; Fung, B.; Haghighat, F.; Yoshino, H.; and Morofsky, E. 2011. A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings* 43(6):1409–1417.

Yu, T. 2010. Modeling occupancy behavior for energy efficiency and occupants comfort management in intelligent buildings. In *9th International Conference on Machine Learning and Applications, 2010*, 726–731. IEEE.