Choosing the Best Strategy for Energy Aware Building System: an SVM-based Approach

Yuanyang Wang, Xiaohong Chen*, Haiying Sun, Mingsong Chen

Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, 200062, China

*corresponding author. xhchen@sei.ecnu.edu.cn

Abstract—For many old buildings in the world, due to the legacy devices problem, it is hard to supply appropriate energy for them. In order to reduce the energy consumption of buildings under the premise of satisfying user requirements, we use software control systems whose core part is the scheduling strategy, to reconstruct them. It is time consuming to choose a good scheduling strategy due to many uncertain factors, among which user actions are of the most influence. In this paper, we propose an Support Vector Machine (SVM) based approach to explore the relation between user action and the best scheduling strategy of a control system. The main contributions include: (1) obtaining the sample set by collecting data at the model level using Statistical Model Checking (SMC) based method; (2) using SVM algorithm to learn the relation model between user actions and the best scheduling strategies; and (3) applying the relation model to predict a best scheduling strategy. Finally a real case study is conducted showing the efficiency of our approach.

Keywords—Energy Aware Building; User Actions; Scheduling Strategy; Support Vector Machine; Statistical Model Checking.

I. INTRODUCTION

Although there are many new buildings which embed intellectual technology to minimize energy consumption, there are still a huge number of old buildings left. They do not have sensors, or enough devices, i.e., heaters to supply energy for each room at the same time. It is difficult to manually schedule heaters of the whole building to satisfy every user with least energy consumption. To financially reconstruct these buildings for minimizing energy consumption, designing a software system which provides a scheduling strategy to schedule heaters and control the energy consumption of the whole building is a good choice.

In order to design such a system easily for the system designer, we make the stable part of the software system to be a *Controller*, and the changeable part a *Scheduling strategy* [1]. Many factors should be considered when designing a scheduling strategy, for example, the rooms, the users and the weather [1] [2]. Among all these factors, we argue the users factor is a very important one because the user actions such as arriving the building and leaving the building finally determine the total energy the users need. But the actions of users are difficult to study, because they are uncertain, which means they are changing and can not be controlled. Designers can only monitor them, and use monitoring results to design proper scheduling strategies. To facilitate the system designer, we propose to define a few scheduling strategies beforehand. The designers could choose one from them such as the one that consumes the least energy. Our previous work

presents an Statistic Model Checking (SMC) based framework to evaluate the energy consumption [1] of each strategy. In that framework, user action is also one factor that is modelled by Stochastic Hybrid Automata (SHA). The energy evaluation is conducted by the probabilistic model checker UPPAAL-SMC [3]. However, each time the evaluation takes about 2.5 hours. Given 100 strategies, to choose one needs 250 hours. How to quickly choose a best strategy which consumes the least energy under the premise of satisfying user requirements, if only the users are changed, is an important problem that needs solving.

In this paper we intend to explore the relation between user actions and the scheduling strategy chosen in order to improve the efficiency of design procedure. The machine learning based approaches [4] give us the inspiration that we do not need to do the evaluation every time. We could select some samples, and learn the relation using machine learning algorithms. We regard the best strategy choosing problem as a classification problem as the user actions which have the same best scheduling strategy could be divided into the same class. Therefore, in this paper we propose an Support Vector Machine (SVM) [5] based approach to learn the relation between user actions and the best scheduling strategy, and use the relation learnt to predict the best scheduling strategy for specific building automatically. The sample set is acquired by applying our SMC based evaluation framework [1]. This is also because old buildings usually do not have any sensors to monitor its own running status. When we do the application, the user actions in terms of arriving and leaving a building are expressed quantitatively by data fitting tool of MATLAB [6].

The rest of the paper is organized as follows. Section II briefly introduces the SMC-based evaluation framework. Section III presents our approach. Section IV conducts a real case study, and finally section V concludes the paper, and put forwards our future work.

II. INTRODUCTION TO SMC-BASED EVALUATION FRAMEWORK

In our previous work [1], the SMC based evaluation framework includes two parts, i.e., the system modeling, and the energy evolution. The system is modelled by a Network of Stochastic Hybrid Automata (NSHA), while each component of the system is modelled by an SHA. Using the energy aware building system as an example, we introduce the system modeling especially the user modelling.

The energy aware building system is composed by *Controller* and four interactive components, i.e., *Weather*, *User*, *Room* and *Heater*, where *Controller* is the software to-be built. The scheduling strategy is embedded in the *Controller*. Each of them is modled by an SHA. Each SHA model has some parameters which may represent the domain knowledge such as the probability of rain, the power of heater and the room amount of a building. They need to be configured before getting a concrete system. In the following, we introduce the *User* model in detail for further use.

Fig. 1 shows the SHA model of *User*. There are four basic locations, *Start*, *Wait*, *In* and *Out*. They have a time attribute *t_day* which increases linearly as the time goes on. For each room, when the time is later than arrive time *arrive_t*, it transfers from *Wait* to *In*. When the time is later than leave time *leave_t*, it transfers from *In* to *Out*. The *arrive_t* and *leave_t* are allocated by the function *time_init*(). The parameters of *User* include the range and distribution of *arrive_t* and *leave_t*.
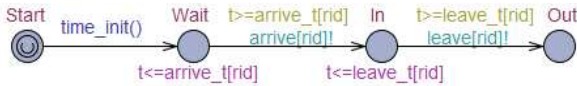

Fig. 1. *User* model SHA

After getting the concrete system, we conduct the energy evaluation on the probabilistic model checker UPPAAL-SMC [3] by property queries. Based on this, different scheduling strategies could be compared easily.

III. OUR APPROACH

The framework of our approach is shown in Fig.2. There are two parts, i.e., the SVM-based relation model training and checking, and the application. The former part uses SVM algorithm to train the relation model, and check the predicting accuracy of the model. The latter part is used to predict a best scheduling strategy for new user actions.

A. Relation model training and checking

The input of this part is the instances of the parameters of SHA models, and the output is a relation model. This part includes for steps: collecting original data, acquiring sample set, learning relation, and checking accuracy.

Step 1. Collecting original data

The original data means the user actions, and the best scheduling strategy at that time. This could be done by using the SMC-based evaluation framework. To be exact, this step includes three sub-steps, i.e., specific system configuration under certain situations, energy consumption evaluation, and original data extraction.

Firstly we conduct the system configuration by setting values to the parameters of *Weather*, *Room*, *Heater* and *Controller* according to domain knowledge. Here we focus on the *User* and *Controller* The parameters *Arrive_range* and *Leave_range* are easy to set according to actual situation. The distribution of *arrive_t* and *leave_t* can be expressed by
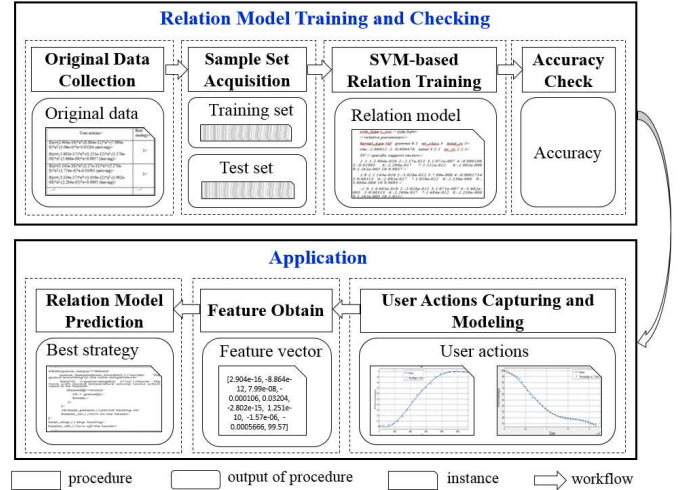

Fig. 2. The framework of our approach

the relation function between time and Percentage of Users (*PU*) in the building which can be calculated as formula (1). The data fitting result shows that polynomial function is suitable for being the fitting function. This is also reasonable in theory according to the characteristic of user actions.

$$PU = \frac{Number\ of\ users\ in\ the\ building}{Number\ of\ total\ users} \quad (1)$$

The controller configuration means to select a scheduling strategy from existing ones or design a new scheduling strategy for system according to user requirements.

Then we conduct the query $Pr[<= day](<> energy >= 20000000)$ which intends to explore the probability of energy consumption exceed 20000000 energy units within a day. The best scheduling strategy can be obtained by directly comparing the cumulative probability of each strategy. The lower the cumulative probability is, the better the strategy is. The corresponding user actions and its best scheduling strategy is recorded. In order to get more original data, we need to find the best strategies under different user actions PU_A and PU_L while other configuration is fixed. We change the parameters of user action expressions randomly in a reasonable scope. Based on the evaluation framework, we get the best strategies under different user actions which called original data of our approach.

Step 2. Acquiring sample set

We extract the sample set from the original data. A sample is composed by features and a label. In our approach, the parameters of user action expressions are regarded as features. And the label is just the number of best scheduling strategy. For example, if the best strategy under some specific user actions is Strategy 2 which has been predefined, the label of this sample is just 2. The sample set is divided into two subsets: the training set and the test set. The training set is used by SVM algorithm to learn the relation model we need. And the test set is used to check accuracy of relation model predictions.

Step 3. Training relation model

Based on the training set, the relation between user actions and energy aware building system designs can be obtained by applying SVM algorithm. In this paper, we call the LIBSVM [7] to implement the SVM classification.

Step 4. Checking Accuracy

We use the test set to check the accuracy of relation model prediction results. Its defined as the percentage of correct predictions in total predictions as shown in formula (2). To some extent, the higher the accuracy, the better the relation model is.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Number\ of\ total\ predictions} \quad (2)$$

B. Application

The application uses the obtained relation model to predict the best scheduling strategy for specific to-be reconstructed building without modeling the whole building. The input of the application is the user actions of specific building and the obtained relation model, and the output is the predicted best scheduling of this building.

Step 1. Capturing and modeling user actions

The designers are expected to capture user actions at first. And then modeling them by data fitting tools in order to get user action expressions. Considering the convenience of extracting features, the user action expressions should be in the format required by the prediction model.

Step 2. Obtaining features

The parameter value of functions extracted from user actions will be regarded as the feature value of this sample. They will be stored in a feature vector which is regarded as the input of relation model.

Step 3. Predicting best strategy

After the feature vector obtained, software system designers just need to input it into relation model. The relation model will output the predicted best scheduling strategy automatically from given ones.

IV. CASE STUDY

We take the Science Building in East China Normal University (ECNU) as an example to show the efficiency of our approach, which is also used in [1]. We use the three predefined scheduling strategies in that paper, named Strategy 1, Strategy 2, and Strategy 3.

A. Relation model training and checking

1) Collecting original data: The system configuration is fixed (omitted due to the limited pages which can be found in [1] except for the user actions and strategy. We use the attendance data of Science Building in March of 2015 to get the distribution of $arrive\_t$ and $leave\_t$. Each parameters of fitted curve given by MATLAB has an interval. We select 100 kinds of user actions randomly to conduct the evaluation on UPPAAL-4.1.19. That is to say, we get 100 original data.

TABLE I
THE USER ACTIONS IN SCIENCE BUILDING: EXAMPLES

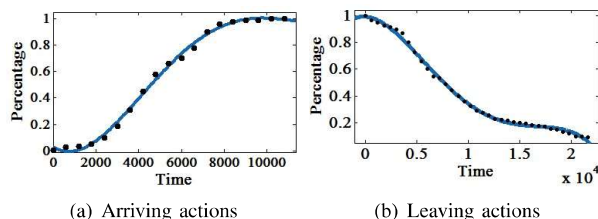| Time	Users	PU	Time	Users	PU
7:00	4	0.55%	16:00	733	100.00%
7:10	19	2.59%	16:20	699	95.36%
7:20	25	3.41%	16:40	674	91.95%
...	...	...	...	...	...
9:50	731	99.73%	21:40	70	9.55%
10:00	733	100.00%	22:00	67	9.14%

(a) Arriving actions   (b) Leaving actions

Fig. 3. User actions

2) Acquiring sample set: The sample set has 100 samples. We select 70 of them by stratified sampling method [8] to be the training data, and the remaining ones to be the test data.

3) Training relation model: We call the libsvm-3.20 with default parameters to training the 70 samples.

4) Checking Accuracy: The 30 test samples are used to test the prediction accuracy of relation model. The experiment result is that 26 predictions are correct. The accuracy is 86.7%. The testing results shows that our relation model is relatively good, and can be applied to practical use for predicting a best scheduling strategy under a specific kind of user actions.

B. Application

1) Capturing and modeling user actions: The attendance data of Science Building in April and May of 2015 is applied to conduct the case study. The data of each day is added together in order to obtain a more reliable distribution. For the arriving actions, we count the arrived users in each time interval which is 10 minutes here between 7:00 am and 10:00 am. As for the leaving actions of Science Building, we collect the number of leaving users every 20 minutes between 16:00 pm and 22:00 pm. Then we get the number of remained users by using number of users to subtract number of leaved users. The results are showed in Table I.

We use MATLAB7.1 to fit the curve of data in Table I (left). The time points are transformed into seconds. In order to simplify the calculation, we change the time 7:00 into 0, 7:10 into 600 as 10 minutes equals to 600 seconds. The rest can be done in the same manner. The result is shown in Fig.3(a). The X position is the time whose unit is second. The Y position is the PU in the building. The leaving action is processed in the same manner. And the result is shown in Fig.3(b). The expressions of user actions are shown in formula (3) (4) which correspond to Fig.3(a) and Fig.3(b) respectively.

$$PU\_A = 2.904e - 16 * x^4 - 8.864e - 12 * x^3$$
$$+ 7.99e - 08 * x^2 - 0.000106 * x + 0.03204 \quad (3)$$

$$PU\_L = - 2.802e - 15 * x^4 + 1.251e - 10 * x^3$$
$$- 1.57e - 6 * x^2 - 5.666e - 6 * x + 0.9957 \qquad (4)$$

2) Obtaining features: After the expressions of user actions obtained, we extract the parameters of user actions, and put them into a feature vector stored in a txt file shown as follows:[2.904e-16, -8.864e-12, 7.99e-08, -0.000106, 0.03204, -2.802e-15, 1.251e-10, -1.57e-06, -0.0005666, 0.9957] where they are values of parameter P_1 to P_{10}.

3) Predicting best strategy: And then transform it into the format can be accepted by LIBSVM. Later, we use the relation model acquired above to check this sample just by running the model using the feature vector as an input. For this case, the predicted best scheduling strategy is Strategy 3.
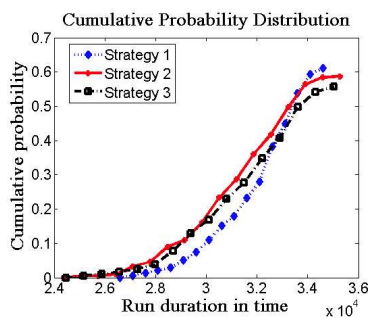

Fig. 4. Energy consumption cumulative probability of three strategies

C. Evaluation

In order to check whether Strategy 3 is the best strategy or not, we run the whole system model by importing the user actions. The simulation result given by query $Pr[<= day](<> energy >= 20000000)$ in Fig.4 shows that the cumulative probability of Strategy 3 is smallest. It means that the probability of energy consumption exceed 20000000 energy units within a day is smallest by using Strategy 3. The application result shows that our approach can correctly predict the best scheduling strategy for specific building. It will greatly reduce the workload of building system designers.

In this application, the average energy consumption under each scheduling strategy of Science Building is shown in Table II. We can see that the Strategy 3 is the best strategy with the lowest energy consumption, and the Strategy 1 is the worst one. If the designers choose Strategy 1 or Strategy 2 as the scheduling strategy unfortunately, the energy consumption will increase by 2100000 or 800000 energy units each day respectively. Maybe the value for one day is not very big. But for one month, one year, the value will be huge. These results show the effectiveness of our approach. They demonstrate that using the scheduling strategy our relation model recommended can reduce the building energy consumption effectively without modeling the whole system or evaluating each strategies. Assuming that there are 100 buildings need to be allocated scheduling strategy from three ones. Our approach can give out the predicted strategy automatically. We just

TABLE II
ANALYSIS OF A SIMULATION

Strategy	Energy consumption	Analysis	Energy saving
1	2.19e7	Worst	2100000
2	2.06e7	-	800000
3	1.98e7	Best	-

need a little time to model the user actions. Without our approach, you approximately need 2.5*3*100=750 hours to allocate scheduling strategies for the 100 buildings.

V. CONCLUSIONS

For old buildings reconstruction, choosing a good scheduling strategy for limited heaters is important for reducing energy consumption. We regard this issue as a classification problem in machine learning, and propose to use the SVM algorithm to learn a relation model between between user actions and best scheduling strategies. The SVM sample set is collected on the model level by applying an energy consumption evaluation framework.

Our experiment based on the data collected from Science Building of ECNU shows that our approach can correctly and quickly select a best scheduling strategy within a short time. Using the strategy recommended by our relation model can reduce considerable energy consumption in our application example. These all show the correctness and effectiveness of our approach. Of course, this is only an initial work. There are still much work left. For example, we should add more samples to improve the predicting accuracy, and add more aspects to measure the best scheduling strategy such as the comfort of users.

REFERENCES

[1] X. Chen, F. Gu, M. Chen, D. Du, J. Liu, and H. Sun, "Evaluating energy consumption for cyber-physical energy system: an environment ontology-based approach," in *Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual*, vol. 2. IEEE, 2015, pp. 5–14.
[2] A. David, D. Du, K. G. Larsen, M. Mikučionis, and A. Skou, "An evaluation framework for energy aware buildings using statistical model checking," *Science China information sciences*, vol. 55, no. 12, pp. 2694–2707, 2012.
[3] P. Bulychev, A. David, K. G. Larsen, A. Legay, M. Mikučionis, and D. B. Poulsen, "Checking and distributing statistical model checking," in *NASA Formal Methods*. Springer, 2012, pp. 449–463.
[4] L. Wu, G. Kaiser, D. Solomon, R. Winter, A. Boulanger, and R. Anderson, "Improving efficiency and reliability of building systems using machine learning and automated online evaluation," in *Systems, Applications and Technology Conference (LISAT)*. IEEE, 2012, pp. 1–6.
[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[6] "Matlab," http://www.mathwork.com.
[7] "Libsvm," http://www.csie.ntu.edu.tw/ cjlin/libsvm.
[8] P. S. Levy and S. Lemeshow, *Sampling of populations: methods and applications*. John Wiley & Sons, 2013.