Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method

Hai-xiang ZHAO and Frédéric MAGOULÈS
Ecole Centrale Paris, Grande Voie des Vignes
Applied Mathematics and Systems Laboratory
92295 Châtenay-Malabry Cedex, France

ABSTRACT

Machine learning methods are widely studied and applied to predict building energy consumption. Since the factors associated with building energy behaviors are quite abundant and complex, this paper investigates for the first time how the selection of subsets of features influence the model performance when statistical learning method is adopted to derive the model. In this paper the optimal features are selected based on the feasibility of obtaining them and on the scores they provide under the evaluation of some filter methods. The selected subset is then evaluated on three data sets by support vector regression involving two kernel functions: radial basis function and polynomial function. Experimental results confirm the validity of the selected subset and show that the proposed feature selection method can guarantee the prediction accuracy and reduces the computational time for data analyzing.

Keywords: Feature selection, support vector regression, prediction, building, energy consumption

1. INTRODUCTION

Supervised learning methods such as Artificial Neural Networks (ANNs), Bayesian network, regression analysis, Support Vector Machines (SVMs) have shown obvious advantages in the prediction of building energy consumption.

*Corresponding author. haixiang.zhao@ecp.fr, frederic.magoules@hotmail.com

The basic idea of these methods is to extract mathematical models from historical performances data and to apply these models to predict the energy behaviors in the unknown future. In comparison with other predictive methods such as engineering calculations, measuring and monitoring, physical simulation, and thermal dynamics, the advantages of supervised learning methods mainly consist in the following ways: first, the prediction is based on solid and realistic historical data, so the prediction is accurate as long as the derived model could represent correctly the historical performance. Second, the models are dynamic and can be retrained with new inputs of historical data. Third, it is comparatively easy to apply the models for prediction, i.e., without tedious human work on the operations such as adjusting input parameters or performing calibration. Finally, several applicable modeling techniques have been well developed during the last decades and have been shown capability for solving such problems appropriately.

Essentially, the purpose of model learning is to find out the proper correspondence between the features and the target. In our application, the features could be the factors influencing building's energy dynamics such as outdoor air temperatures or building structures. The target could be the energy requirements of a building, such as electricity consumption or district heating requirement. In fact, there are a great number of factors which probably have impact on energy dynamics of a building. How to reasonably choose a subset of appropriate features to be used in model learning is one of the key issues for machine learning methods on such applications. On one hand, using different sets of features would probably change the performance of the models in accuracy and learning speed. On the other hand, the optimal set of features would make the prediction models more practical.

To the best of our knowledge, there is little work concerning the Feature Selection (FS) of building energy consumptions with regard to machine learning methods. Most of the existing works derive their own model using previously established sets of features. Madsen et al. [1] derived their continuous-time models on five variables, which are room air temperature, surface temperature, ambient dry bulb temperature, energy input from the electrical heaters and solar radiation on southern surface. Neto et al. [2] built their neural network based on the input of daily average values of dry bulb temperature, relative humidity, global solar radiation and diffuse solar radiation. Azadeh et al. [3] and Maia et al. [4] forecasted electrical energy consumption through analyzing the varying inner targets without any contributory variable involved. Yokoyama et al. [5] considered only two features, air temperature and relative humidity in their neural network model. Tso et al. [6] used more than 15 features in their assessment of traditional regression analysis, decision tree and neural networks. Similar approaches can be found in [7–11].

This paper is so far the first attempt to discuss how to select subsets of features for statistical models applied to the prediction of building energy consumption. We present here an heuristic approach for selecting subset of features, and systematically analyze how it will influence the models performance. The models are trained by support vector regression (SVR) with different kernel methods based on three data sets. The FS method is evaluated by comparing the models' performances before and after FS is performed.

This paper is organized as follows. Section 2 briefly introduces support vector regression method. Section 3 introduces the energy data sets that will be used to train the models and how the data is obtained. Section 4 discusses general FS methods and in particular the one introduced in this work. Section 5 illustrates with several numerical experiments the robustness and efficiency of the proposed method. Finally, section 6 concludes this paper.

2. PRINCIPLES OF SUPPORT VECTOR REGRESSION

SVMs are based on the principle of structural risk minimization which aims at restricting the generalization error at the lowest level [12]. They have been proved to be excellent models in solving non-linear problems. Recently, they have been widely studied and applied in various researches, such as image processing, text mining, and biological analysis. Two main applications of SVMs are classification and regression, which are called Support Vector Classification (SVC) and Support Vector Regression (SVR), respectively. The latter has been introduced to energy load forecasting and has shown remarkable abilities [7, 10, 11, 13].

The basic idea of SVR is to find a decision function $f(x)$ (also called model) to represent the relationship between the features and the target [12, 14]. We represent the whole training data as (x_1, y_1) (x_2, y_2),... (x_l, y_l), where vector x_i is the *ith* sample of the features, y_i is the target value corresponding to x_i, l is the number of samples. If there are n features in the training data, then the dimensionality of x_i is n.

In practice, it is difficult to find out a linear function $f(x)$ for problems involving large data sets. But fortunately, we can solve this problem by mapping the 1-dimensional problem into a higher dimensional feature space where a linear function similar to $f(x)$ can be found. Moreover, it is not necessary to express explicitly the mapping during the computation. After introducing the kernel function K, $f(x)$ can be represented as:

$$f(x) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) K(x_i \cdot x) + b$$

then eventually, the optimization problem reduces to maximize the quadratic function:

$$W(\alpha_i^*, \alpha_i) = \sum_{i=1}^{l} y_i (\alpha_i^* - \alpha_i) - \varepsilon \sum_{i=1}^{l} (\alpha_i^* + \alpha_i) - \frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i \cdot x_j)$$

under the constraints $\sum_{i=1}^{l} \alpha_i^* = \sum_{i=1}^{l} \alpha_i$ and $0 \le \alpha_i^*, \alpha_i \le C$, $i = 1, 2, \ldots, l$.

Usually, only certain parts of the samples can satisfy the property: $\alpha_i^* - \alpha_i \neq 0$, they are called support vectors (SVs).

There are four commonly used kernel functions, which are linear function, polynomial function, Radial Basis Function (RBF) and sigmoid function. They represent different mappings of a problem from lower dimensional space to higher feature space. The linear kernel and sigmoid kernel would perform like RBF with some special parameters settings as described in [7, 12]. And the sigmoid kernel is even invalid under particular parameters. In our data training process, we have a large number of samples with dozens of features, therefore we need the kernels with the ability to solve non-linear problems. For this purpose, RBF and polynomial kernels are selected in this work to combine with SVR for model training. This selection is identical with the one utilized by Joachims [15] in his well-known text categorization work.

In the evaluation step, Mean Squared Error (MSE) and Squared Correlation Coefficient (SCC) are adopted to evaluate the performance of the predictor. MSE gives the average deviation of the predicted values to the measured one. It is a non-negative value, the lower the value, the better the performance of the prediction. SCC determines the ratio of successfully predicted number of target values on total number of target values. It lies between [0, 1]. The higher the SCC, the stronger the prediction ability.

3. DATA GENERATION AND TEST BUILDING

Hourly electricity consumption together with hourly values of several influencing factors is recorded as the building energy consumption data. This data is generated by *EnergyPlus*, which is a widely applied, the state-of-the-art building energy simulation tool developed by U.S. Department of Energy and some other collaborators [16]. It has been proved to have the capability of well reflecting the energy performance of actual buildings as long as calibration is properly performed [17]. Highly detailed energy variations like hourly loss through windows can be recorded while carrying out the simulation. These data would be significantly helpful in interpreting the sources of the total consumption. The aim of our work is to test the statistical models on some smooth consumption behaviors without concerning sudden variations which happen frequently in the real world. For this reason, the data sets obtained from simulation is considered to be adequate to test the models.

The test buildings are single-story mass-built buildings for office use located in five major cities of France, which are Paris-Orly, Marseilles, Strasbourg, Bordeaux and Lyon. They represent the office energy requirements under five typical weather conditions in France. The ambient dry bulb temperature is always considered to be the most relevant factor to the total energy consumption on regression analysis [18]. Thus we depict here this variable of these five places in a typical month of a year to obtain a glance of their variations. As Figure 1 shows, the five sets of ambient dry bulb temperature

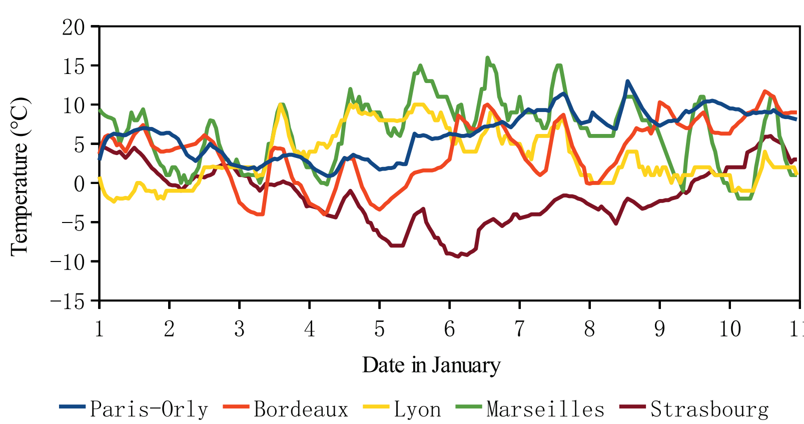

Figure 1. Dry bulb temperature in the first 11 days of January.

deviate from each other remarkably at any point, which implies that the energy behaviors of the buildings in the five cities would be quite different even in the case that the buildings have the same structural characteristics.

All of the buildings are located in urban area. Each building has similar structures, e.g., one rectangle room with attic roof and four windows in four directions without shading. Electrical equipments including lighting system, fans, water heaters, are scheduled as common office use. In winter season (from November 1st to March 31st), district heating is applied as the heating system in order to keep the room temperature at a constant level. At the same time, ventilation is adopted for indoor thermal comfort. The number of occupants depends on the housing space, with the average of 0.2 people per zone floor area. During the simulation, some input variables are set to different values to achieve diversity among different buildings. Detailed process of simulation is presented in section 5.

4. FEATURE SELECTION

FS aims at selecting the most useful feature set to establish a good predictor for the concerned learning algorithm. The irrelevant and unimportant features are discarded in order to reduce the input dimensionality. Several advantages will be achieved if we wisely select the best subset of features. The first one is the simplification of the calculation while keeping the dimensionality minimized, which could contribute to avoid the curse of dimensionality. The second one is the possible improvement of accuracy of the developed model. The third one is the improved interpretability of the models. The last one is the feasibility of obtaining accurate feature samples, especially for some time series problems in practice.

The methods for FS can be classified into three categories: filter, wrapper and imbedded method. The filter method aims at ranking features with some correlation or mutual information criteria, and selecting the features with the highest ranks. It can be regarded as a pre-processing step without the model training algorithm involved, which means it is independent of the predictor designed in the learning step. Wrapper method assesses the subsets of features according to the degree of accuracy they contribute to a given predictor. Embedded method evaluates the usefulness of feature sets in the same way as wrapper method does, while the selection proceeds directly in the training process and thus can consequently avoid multiple training for each candidate subset. More detailed information can be found in the work [19].

Two FS methods are used in our model training. The first one ranks the features individually by correlation coefficient between each feature and the target. We use CC to stand for this method. The correlation coefficient between two vectors is defined as:

$$\mathrm{CC}(f) = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right]\left[N\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

The other one is called Regression, Gradient guided feature selection (RGS), which is developed by A. Navot et al. in the application of brain neural activities [20]. We choose this method since it is designed specially for regression and has shown competitive ability to handle a complicated dependency of the target function on groups of features. The basic idea is to assign a weight to each feature and evaluate the weight vector of all the features simultaneously by gradient ascent. The non-linear function K-Nearest-Neighbor (KNN) is applied as the predictor to evaluate the dependency of the target on the features. The estimated target of sample x under KNN is defined as:

$$\hat{f}_w(x) = \frac{1}{Z}\sum_{x' \in N(x)} f(x')e^{-d(x,x')/\beta}$$

where $N(x)$ is the set of K nearest neighbors of sample x. $d(x,x') = \sum_{i=1}^{n}(x_i - x_i')^2 w_i^2$ is the distance between sample x and one of its nearest neighbors x', n is the number of features, w is the weight vector and w_i is the specific weight assigned to ith feature. $Z = \sum_{x' \in N(x)} e^{-d(x',x)/\beta}$ is a normalization factor and β is a Gaussian decay factor. Then the optimal w can be found by maximize the following evaluation function:

$$e(w) = -\frac{1}{2}\sum_{x \in S}(f(x) - \hat{f}_w(x))^2$$

where S is the samples for model training. Since $e(w)$ is smooth almost everywhere in a continuous domain, one can solve the extremum seeking problem by gradient ascent method. More details can be found in [20].

5. NUMERICAL EXPERIMENTS

5.1 Model Testing on One Building's Data before Feature Selection

In this section, we test SVR model to the prediction of the time series consumption of one building. To collect the historical data, a single office building located in Paris-Orly is simulated in *EnergyPlus* and the hourly electric demands together with hourly behaviors of 23 variables are recorded through one year. Table 1 shows the recorded variables and their units. It is known that building structures like room space, thickness of the walls, area of the windows, play important roles in the total energy consumption of a building. However for one particular building, these variables have constant values through the

Table 1. The 23 features for the model training and testing on one building's consumption.

Features	**Unit**
Outdoor Dry Bulb	C
Outdoor Relative Humidity	%
Wind Speed	m/s
Direct Solar	W/m^2
Ground Temperature	C
Outdoor Air Density	kg/m^3
Water Mains Temperature	C
Zone Total Internal Total Heat Gain	J
People Number Of Occupants	-
People Total Heat Gain	J
Lights Total Heat Gain	J
Electric Equipment Total Heat Gain	J
Window Heat Gain for each wall	W
Window Heat Loss for each wall	W
Zone Mean Air Temperature	C
Zone Infiltration Volume	m^3
District Heating Outlet Temp	C

simulation period, making them have no contributions to SVR model learning. Therefore it is practical to discard these variables in this test without losing accuracy in the model. Later in this work, we will take these factors into consideration when we carry out the testing on multiple buildings. Concerning the data analyzing process, the data for model training is the first 10 months of one year's consumption (from January 1st to October 31st) and for model testing is the remaining two months (from November 1st to December 31st). RBF kernel is chosen in this experiment. Three parameters of the learning algorithm, C, γ and p are optimized by stepwise 5-fold cross validation as described in the technical document of Libsvm [21]. The initial searching spaces are $\{2^{-3},2^{-2},...,2^{8}\}$, $\{2^{-10},2^{-9},...,2^{2}\}$ and $\{2^{-10},2^{-9},...,2^{-5}\}$ for C, γ and p respectively. In order to avoid numerical problems in the calculations, before training the model, we scale the training set into the range [0, 1] and then apply the scaling function to scale the testing set.

Since usually people do not work in weekends and holidays, the energy requirement in these days is quite small compared to normal working days. That means weekends and holidays have totally different energy behaviors from working days. Take the 56th day as an example, it is a Saturday, the energy consumption for that day is 0.18, compared to other working days which have a normal consumption of more than 4, it can be safely ignored. It is proved that if we distinguish these two types of days, when train the model by predictive models like neural networks, considerable performance improvements could be achieved [22, 23]. Therefore, to simplify the model in our practice, we only use the consumption data of working days for model training and testing. Consequently, the number of samples for training is 5064 and for testing is 1008.

As the result shows, MSE is $4.8e - 4$ and SCC is 0.97. To illustrate a global view of the results, we accumulate the hourly consumption into daily consumption and plot the measured and predicted daily values in Figure 2. The two curves are quite close to each other which indicates a high accuracy in the prediction. The relative errors for all days are within (–10%, 15%) as Figure 3 shows.

5.2 Model Testing on One Building's Data after Feature Selection

In this part, we experimentally analyze one approach aiming at selecting the best subset of features for training statistical models on building energy consumption data. Our method of FS is based on two criteria. The first is that the selected features should be potentially the most important ones to the

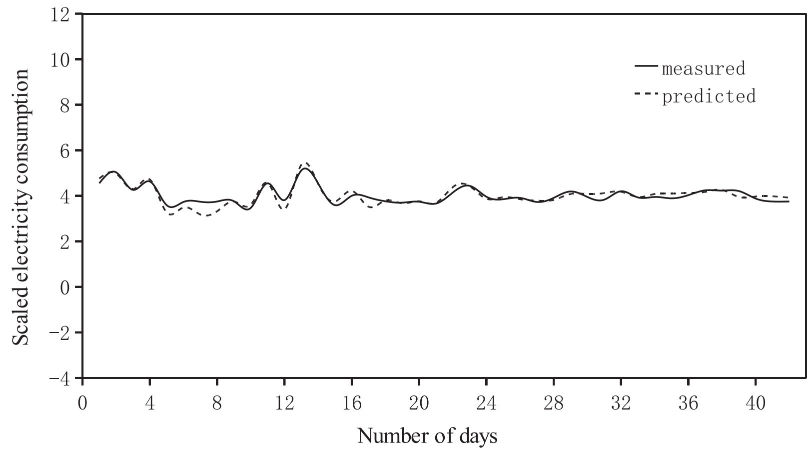

Figure 2. The comparison of measured and predicted daily electricity consumptions for a particular building in working days, from October to December.

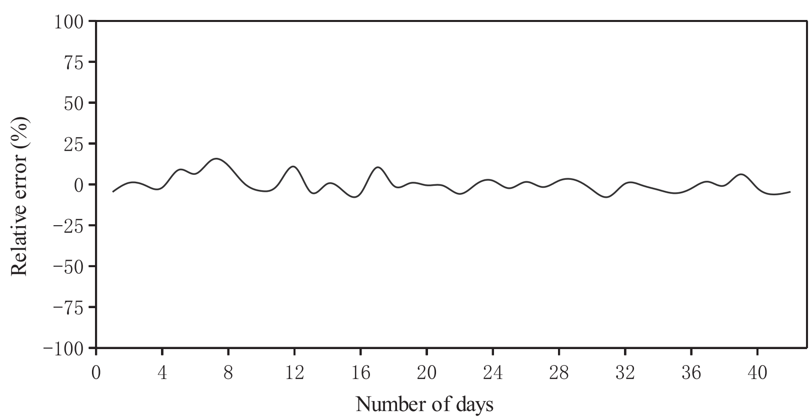

Figure 3. The relative error for the prediction.

predictor as required by a general FS method. For this purpose, we have to involve some FS algorithms in the object and choose the features with highest rankings or scores. The second one is to make sure that the selected features can be easily obtained in practice. In this work, the values of the chosen features normally could be collected practically from measurements, surveys, related documents like building plans and so on.

The two methods RGS and CC described in section 4 are applied to evaluate the usefulness of features. The object data set is the previous consumptions of working days. The scores for each feature are listed in Table 2 in columns 2 and 3. We can see that even the same feature could probably have totally different scores under the evaluation of two FS algorithms. For example, the outdoor dry bulb temperature is the most important feature under the judgment of RGS, while on the contrary, it is almost useless according to CC ranking method. As experimental results have shown, the features with the highest scores under RGS are generally more useful than those with the highest ranks according to CC. This indicates that RGS method is more applicable to SVR than CC method. However, since the feature subsets with low scores are still possibly useful for the learning algorithms [19], we take both RGS and CC into consideration while choosing the features.

The weather data can be recorded in site or gathered from metro department. We keep two weather features that have the highest scores under RGS, which are Dry Bulb Temperature and Outdoor Air Density. And at the same time, we discard Relative Humidity, Wind Speed, Direct Solar and Ground Temperature, no matter how their variations could contribute to energy requirement as we thought naturally. The Water Mains Temperature which gives water temperatures delivered by underground water main pipes, Electrical Equipment Heat Gain which is the heat gain of the room from electrical equipments such as lights, TVs, are determined by their power and occupants' schedule, they could probably be measured or assessed in actual buildings. We divide the room into several zones according to their thermal dynamics. The two features, Zone Mean Air Temperature which is the effective bulk air temperature of the zone and Zone Infiltration Volume which denotes hourly air infiltration of the zone, could also be measured or estimated in a normally operated building. All of the above selected features have scores not less than 1. A special case we have to consider is the People Number Of Occupants. This feature takes a middle place under RGS, but since it can be easily counted in real life and has a very high score under the evaluation of CC, we choose to keep it in the final subset. All other features will be discarded since they get low scores or are hard to be collected in actual buildings. For example, Zone Total Internal Total Heat Gain is difficult to be obtained directly and District Heating Outlet Temp is useless according to CC. The selected features are indicated with stars in column Case1 in Table 2.

New data sets for both training and testing are generated by eliminating useless features from the data sets used in previous experiment. Then, the model

Table 2. The scores of features evaluated by RGS and CC selection methods. The stars indicate selected features in that case.

Features	RGS	CC	Case1	Case2	Case3	Case4	Case5
Outdoor Dry Bulb	1.61	0.29	*		*		*
Outdoor Relative Humidity	0.62	0.26			*	*	
Wind Speed	0.52	0.01			*	*	
Direct Solar	0.54	0.47				*	
Ground Temperature	0.99	0.07				*	
Outdoor Air Density	1.26	0.20	*				*
Water Mains Temperature	1.30	0.07	*				*
Zone Total Internal Total Heat Gain	1.01	0.67					
People Number Of Occupants	0.93	0.68	*	*	*		
People Total Heat Gain	0.93	0.68		*		*	
Lights Total Heat Gain	1.13	0.05	*		*		*
Electric Equipment Total Heat Gain	1.06	0.69	*	*	*		*
Window Heat Gain for each wall	1.03	0.62		*		*	
Window Heat Loss for each wall	0.93	0.50		*		*	
Window Heat Gain for each wall	0.82	0.35				*	
Window Heat Loss for each wall	0.82	0.49				*	
Window Heat Gain for each wall	0.73	0.56		*		*	
Window Heat Loss for each wall	0.82	0.48				*	

Window Heat Gain for each wall	0.89	0.56		*		*	
Window Heat Loss for each wall	0.95	0.50		*		*	
Zone Mean Air Temperature	1.14	0.22	*				*
Zone Infiltration Volume	1.00	0.34	*		*		
District Heating Outlet Temp	0.95	7.35e–4			*	*	

is retrained from the new training data and after applying the model to predict on the testing data, our results are as follows: MSE is $6.19e-4$ and SCC is 0.97. To obtain a clear view of how the model performance changes before and after FS, we plot the measured and predicted daily consumptions in Figure 4. The relative errors are within (–16%, 12%) as shown in Figure 5. We note that after FS, the number of features is 8, which is only one third of the original set which has 23 features. However, compared to the results before FS, the model's prediction ability is still very high and the selected subset is therefore regarded acceptable.

Four other subsets are developed in order to further evaluate if the selected feature set is optimal. They are indicated by columns Case2, Case3, Case4 and Case5 in Table 2. In case 2, we select the top 8 features under the evaluation of CC alone. By doing this, we are aiming at demonstrating that if the single CC is sufficient to select the best feature set. The feature Zone Total Internal Total Heat Gain is also ignored in this case just as we do in case 1. In case 3, we change three of the selected features to other three unselected ones. Outdoor Air Density, Water Mains Temperature and Zone Mean Air Temperature which are selected in case 1 are substituted with Outdoor Relative Humidity, Wind Speed and District Heating Outlet Temp. In case 4, all of the selected features are substituted with other unselected ones except Zone Total Internal Total Heat Gain which is regarded can not be obtained directly in practice. In the last case, two features which gain lowest scores are removed from selected subset. They are People Number Of Occupants and Zone Infiltration Volume.

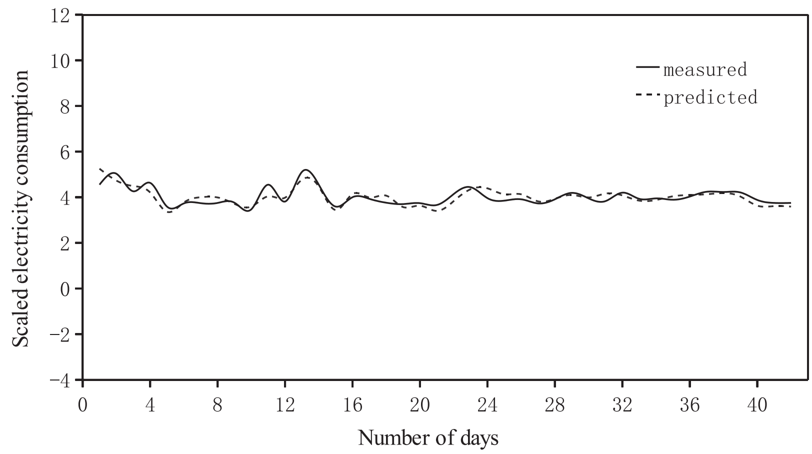

Figure 4. The comparison of measured and predicted daily electricity consumption for a particular building in working days, with FS performed.

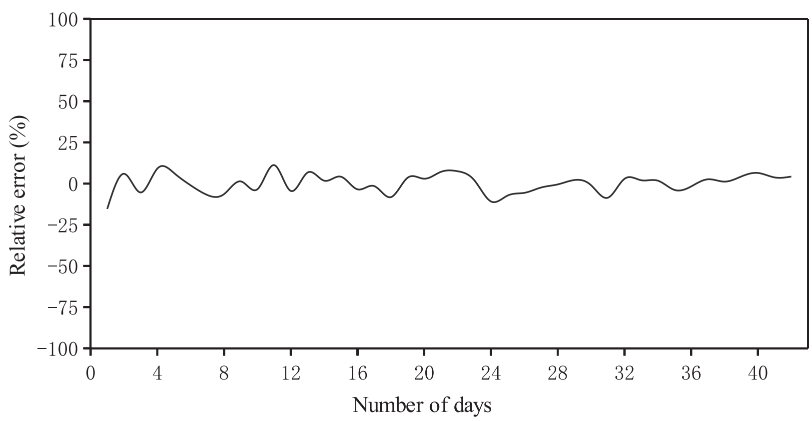

Figure 5. The relative error for the prediction.

Based on these considerations, four new data sets are generated for training and testing, and model is retrained for each case. We show the results of all five cases in Table 3. Two conclusions can be reached according to the results, the first is that the designed FS method is valid due to model performance in case 1 outperforms other three cases. The other is that SVR model with RBF kernel has stable performance since high prediction accuracy is always achieved on all of the four subsets.

Table 3. Comparison of model performance on different feature sets. NF: Number of features, MSE: Mean squared error, SCC: Squared correlation coefficient.

	Case1	**Case2**	**Case3**	**Case4**	**Case5**
NF	8	8	8	14	6
MSE	6.2e–4	1.9e–3	7.5e–4	2.1e–3	9.2e–4
SCC	0.97	0.93	0.96	0.90	0.96

5.3 Model Testing on More Data Sets

Previously, we have already tested the FS method on one particular building's consumption over a year. In this section, we investigate how the subset of features influences the model performance on multiple buildings' consumptions.

We record the energy consumption data in winter season for 50 buildings as indicated in section 3. The differences among these buildings mainly come from the weather conditions, building structures and number of occupants. The weather conditions are different in five cities as shown in Figure 1. The buildings have diverse characteristics with randomly generated length, width, height and window/wall area ratio. The number of occupants is determined by the ground area and people density of the buildings. The time series data of those buildings is combined together to form the training sets. One more building is simulated for model evaluation purpose.

Two sets of consumption data are designed. The first set has 20 buildings and the second one includes all the 50 buildings. To fully investigate how FS on these two data sets influence SVR models, two kernels are involved. Besides RBF kernel, we also test the performance of FS on SVR with polynomial kernel, which is also applicable on non-linear problems. The kernel parameters r is set to zero and d is estimated by 5-fold cross validation in a searching space $\{2, 3,....,7\}$. Selected features for representing multiple buildings are the feature sets for single building plus building structures. Therefore, FS for multiple buildings has reduced the number of features from 28 to 12. The changes of MSE and SCC on these data sets are shown in Table 4. For information, the results of one single building is indicated in the same table.

After FS, the accuracy of the prediction on 50 buildings' consumptions improves significantly. With regard to 20 building's consumptions, MSE increases to a certain extent, indicating a decrease in prediction accuracy.

Table 4. Prediction results of SVR with two kernel methods on three data sets. (BF: Before feature selection, AF: After feature selection, MSE: Mean squared error, SCC: Squared correlation coefficient).

			One building	**20 buildings**	**50 buildings**
RBF kernel	BF	MSE	4.8e–4	4.3e–4	4.4e–4
		SCC	0.97	0.97	0.97
	AF	MSE	6.2e–4	2.1e–3	3.7e–4
		SCC	0.97	0.96	0.97
Polynomial kernel	BF	MSE	8.0e–4	5.8e–4	5.9e–4
		SCC	0.96	0.96	0.96
	AF	MSE	2.1e–3	0.19	4.7e–4
		SCC	0.91	0.85	0.98

However from the standpoint of SCC, the performance of the model with RBF kernel involved is quite close to the situation without FS performed, as shown in Figure 6(a). With regard to polynomial kernel, when training on the original data sets, the prediction ability of the model is just as good as RBF kernel, indicating that polynomial kernel is also applicable on such problem. After adopting FS, the performance of the model becomes better for the case of 50 buildings. Unfortunately, it decreases largely for the cases of 20 buildings. It seems that polynomial kernel is not as stable as RBF kernel when applied on such problems. However we can see that it performs better for the case of 50 buildings than for that of 20 buildings. Same trend is also found for RBF kernel. These phenomena indicate that proposed FS approach is able to give better performance to the models when more training samples are involved.

Another advantage of FS for statistical models is the reduction of training time. We show the time consumed for training SVR models with RBF kernel in Figure 6(b) where the time is in the logarithm form. The training time after FS is obviously less than that before FS. But the reduction is not too much. This phenomenon can be explained by the different parameter values we assigned for the learning algorithm, which always have a great influence on the training speed. We note that the time for choosing parameters for the predictor via cross validation is too long to be ignored when evaluating a learning algorithm. While

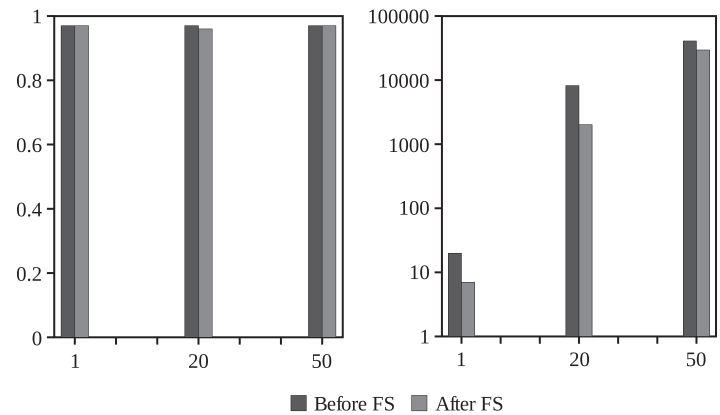

Figure 6. (a) The comparison of model performance in the standpoint of SCC before and after FS for RBF kernel. X-axis represents the number of features, Y-axis is the SCC. (b) The comparison of training time before and after FS for RBF kernel. Y-axis stands for training time (in seconds).

in this paper we mainly focus on the influences of FS on predictors, the labor and time for choosing model parameters are not considered since they are quite approximate before and after FS.

6. CONCLUSION

This paper introduces a new feature selection method for applying support vector regression to predict the energy consumptions of office buildings.

To evaluate the proposed feature selection method, three data sets are first generated by *EnergyPlus*. They are time series consumptions for respectively one, twenty and fifty buildings. We assume that the developed models are applied to predict the energy requirements of actual buildings, therefore, the features are selected according to their feasibility in practice. To support the selection, we adopt two filter methods: the gradient guided feature selection and the correlation coefficients, which can give each feature a score according to its usefulness to the predictor. Extensive experiments show that the selected subset is valid and can provide acceptable predictors. Performance improvement is achieved in some cases, e.g., accuracy enhanced remarkably for the models with either radial basis function or polynomial kernel on fifty buildings' data, the time for model learning decreases to a certain extent. We also identify that the performance becomes better when more training samples are involved.

Besides radial basis function kernel, we show that polynomial kernel is also applicable to our application. However it seems not as stable as radial basis function kernel. Furthermore, it requires more complicated pre-processing work since there are more kernel parameters need to be estimated.

This preliminary work on feature selection for building energy consumptions has paved a way for its further progress. It serves as the first guide for selecting an optimal subset of features when applying machine learning methods on the prediction of building energy consumption. Future work will be carried out on developing appropriate feature selection algorithms for other modeling methods.

REFERENCES

1. H. Madsen and J. Holst. Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings*, 22(1):67–79, 1995.
2. A.H. Neto and F.A.S. Fiorelli. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, 40(12):2169–2176, 2008.
3. A. Azadeh and S. Tarverdian. Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption. Energy Policy, 35(10):5229–5241, 2007.
4. C.A. Maia and M.M. Gonçalves. A methodology for short-term electric load forecasting based on specialized recursive digital filters. *Computers and Industrial Engineering*, 57(3):724–731, 2009.
5. R. Yokoyama, T.Wakui, and R. Satake. Prediction of energy demands using neural network with model identification by global optimization. *Energy Conversion and Management*, 50(2):319–327, 2009.
6. G.K.F. Tso and K.K.W. Yau. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768, 2007.
7. B. Dong, C. Cao, and S.E. Lee. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37(5):545–553, 2005.
8. A.E. Ben-Nakhi and M.A. Mahmoud. Cooling load prediction for buildings using general regression neural networks. *Energy Conversion and Management*, 45(13-14):2127–2141, 2004.
9. S.L. Wong, Kevin K.W. Wan, and Tony N.T. Lam. Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, 87(2):551–557, 2010.
10. H.X. Zhao and F. Magoulès. Parallel support vector machines applied to the prediction of multiple buildings energy consumption. *Journal of Algorithms & Computational Technology*, 4(2):231–249, 2010.

11. F. Lai, F. Magoulès, and F. Lherminier. Vapnik's learning theory applied to energy consumption forecasts in residential buildings. *International Journal of Computer Mathematics*, 85(10):1563–1588, 2008.
12. V.N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.
13. M. Espinoza, J. Suykens, and B. Moor. Fixed-size least squares support vector machines: a large scale application in electrical load forecasting. *Computational Management Science*, 3(2):113–129, April 2006.
14. A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2004.
15. T. Joachims. Text categorization with support vector machines: learning with many relevant features. In C. Nédellec and C. Rouveirol, editors, *ECML-98: Proceedings of the 10th European Conference on Machine Learning, Chemnitz, DE*, pages 137–142. Springer Verlag, Heidelberg, DE, 1998.
16. EnergyPlus, 2009. Available online at: http://www.EnergyPlus.gov (Accessed January 2010).
17. F.S. Westphal and R. Lamberts. Building simulation calibration using sensitivity analysis. In *IBPSA: Proceedings of the 9th International Building Performance Simulation Association Conference, Montréal, Canada*, pages 1331–1338, 2005.
18. J.R. Wu and E.D. Claridge. Statistical modeling of daily energy consumption in commercial buildings using multiple regression and principal component analysis. In *Proceedings of the 8th Symposium on Improving Building System in Hot and Humid Climates, Dallas, TX*, pages 155–164, 1992.
19. G. Isabelle and E. André. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
20. A. Navot, L. Shpigelman, N. Tishby, and E. Vaadia. Nearest neighbor based feature selection for regression and its application to neural activity. In Y. Weiss, B. Scholkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 995–1002. MIT Press, 2006.
21. C.C. Chang and C.J. Lin. *LIBSVM: a library for support vector machines*, 2001. Available online at http://www.csie.ntu.edu.tw/~cjlin/libsvm (Accessed January 2010).
22. S. Karatasou, M. Santamouris, and V. Geros. Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, 38(8):949–958, 2006.
23. J. Yang, H. Rivard, and R. Zmeureanu. On-line building energy prediction using adaptive artificial neural networks. *Energy and Buildings*, 37(12):1250–1259, 2005.